Universidad Europea VALENCIA

GRADO EN CIENCIA DE DATOS

Visualización de Datos y Storytelling Digital: Diseño de Dashboards Interactivos y Automatización de Reporting para la Toma de Decisiones en Entornos de Negocio y Marketing

Presentado por:

LUIS BALLESTER TAMARIT

Dirigido por:

VICENTE CASTILLO FAULI

CURSO ACADÉMICO 2025-2026

RESUMEN

En el mercado actual sigue siendo una capacidad sin explotar la de transformar información compleja o difícil de analizar a simple vista en narrativas visuales claras debido a un contexto empresarial cada vez más orientado al dato. Este problema se ha abordado en el presente trabajo mediante el diseño y la evaluación de soluciones de visualización interactiva aplicada a entornos de negocio y marketing. El objetivo principal es elaborar un conjunto de dashboards (cuadros de mando) y desarrollar un sistema para crear informes automáticamente, que mejore la comprensión de los datos y la forma de comunicar los resultados. También se aborda una revisión teórica de las leyes de UX/UI (user experience y user interface) y storytelling (visualización narrativa) para su aplicación en la analítica de negocio mediante las herramientas adecuadas. Finalmente, el trabajo se evalúa, desde diferentes aspectos, tales como la eficacia en tareas de interpretación, y el rendimiento técnico del sistema (la frecuencia de actualización o latencia, y la reproducibilidad). También se ha realizado una evaluación de usabilidad y satisfacción del usuario (SUS, System Usability Scale). Se demuestra cómo una buena combinación entre diseño, narrativa y automatización puede contribuir significativamente al análisis del usuario y la toma de decisiones.

Palabras clave: visualización, storytelling, dashboards, UX/UI, evaluación.

ABSTRACT

In today's market, the ability to transform complex or difficult-to-analyze information into clear visual narratives remains largely untapped, due to an increasingly data-driven business environment. This problem has been addressed in this work through the design and evaluation of interactive visualization solutions applied to business and marketing environments. The main objective is to develop a set of dashboards and a system for automatically generating reports, thereby improving data comprehension and the communication of results. A theoretical review of UX/UI (user experience and user interface) principles and storytelling (narrative visualization) is also included, focusing on their application to business analytics using appropriate tools. Finally, the work is evaluated from various perspectives, such as the effectiveness of interpretation tasks and the system's technical performance (update frequency or latency, and reproducibility). A usability and user satisfaction assessment (SUS, System Usability Scale) was also conducted. The work demonstrates how a good combination of design, narrative, and automation can significantly contribute to user analysis and decision-making.

Key words: visualization, storytelling, dashboards, UX/UI, evaluation.

ÍNDICE DE CONTENIDOS

1. Introducción 9
2. Objetivos 11
   O1 - Diseñar la solución de visualización y storytelling 11
      O1.1 - Revisión de buenas prácticas UX/UI y storytelling 11
      O1.2 - Criterios de diseño 11
      O1.3 - Selección de KPIs 11
      O1.4 - Estructura narrativa de páginas 11
   O2 - Implementar el sistema (dashboards y automatización de reporting) 12
      O2.1 - Construcción de dashboards (Looker/Power BI) 12
      O2.2 - Integración de datos 12
      O2.3 - Pipeline ETL 12
      O2.4 - Actualización/automatización 12
      O2.5 - Opciones de despliegue 12
   O3 - Evaluar la solución y derivar recomendaciones 12
      O3.1 - Evaluación de usabilidad (SUS) y tareas 13
      O3.2 - Evaluación técnica (actualización, latencia, reproducibilidad) 13
      O3.4 - Análisis de resultados 13
      O3.5 - Limitaciones, recomendaciones y líneas futuras. 13
3. Marco teórico 13
   3.1 Visualización de datos en analítica de negocio 13
   3.2 KPIs, OKRs y métricas en entornos de negocio y marketing digital 15
      3.2.1. Métricas y su importancia en la toma de decisiones 15
      3.2.2. KPIs: indicadores clave de rendimiento 16
      3.2.3. OKRs: objetivos y resultados clave 16
      3.2.4. Vanity metrics vs. actionable metrics 17
   3.3 Fundamentos de percepción y comunicación visual 17
   3.4 Storytelling y narrativa visual con datos 18
   3.5 UX/UI aplicado a dashboards interactivos 19
   3.6 Sistemas de business intelligence y dashboards 20
   3.7 Automatización de reporting y pipelines de datos 21

3.8 Evaluación de sistemas de visualización 21
4. Metodología 23
4.1 Estrategia general del proyecto 23
4.2 Metodología de estudio 23
4.2.1 Revisión bibliográfica 23
4.2.2 Análisis comparativo de herramientas 24
4.3 Metodología de desarrollo 27
4.3.1 Diseño de dashboards y storytelling 27
4.3.2 Integración de datos y automatización de informes 27
4.4 Evaluación de la solución 30
4.5 Posibilidades de despliegue (deployment) 31
5. Desarrollo, resultados y discusión 35
5.1 Selección del dataset 35
5.2 Exploración y análisis de calidad del dataset 36
5.3 Limpieza, transformación y automatización ETL 38
5.4 Diseño y desarrollo del dashboard en Looker Studio 40
5.5 Diseño y desarrollo del dashboard en Power BI 45
5.6 Pipeline ETL automatizado 49
5.7 Evaluación del sistema 52
5.7.1 Criterios de evaluación basados en literatura 52
5.7.2 Walkthrough por tareas 53
5.7.3 Evaluación técnica del sistema 56
5.7.4 Síntesis de hallazgos y plan de mejoras 58
5.7.5 Validación con usuarios (SUS) 60
5.7.6 Resultados e insights obtenidos 63
6. Objetivos de desarrollo sostenible 66
ODS 8: Trabajo decente y crecimiento económico 66
ODS 9: Industria, innovación e infraestructura 66
ODS 10: Reducción de las desigualdades 67
7. Conclusiones 67
8. Referencias 71
9. Acrónimos 74
10. Anexos 76

Anexo A: Repositorio de código en GitHub 76
Anexo B: Dashboard interactivo en Looker Studio 76
Anexo C: Dashboard en Power BI 76
Anexo D: Formulario de evaluación SUS y resultados 76
Anexo E: Declaración de uso de Inteligencia Artificial: 77

ÍNDICE DE TABLAS

Tabla 1: Clasificación de métricas de marketing digital y visualización recomendada..... 16
Tabla 2: Comparativa de herramientas.....25
Tabla 3: Comparativa de métodos de automatización de la ejecución periódica del pipeline ETL.....29
Tabla 4: Comparativa de opciones de despliegue.....33
Tabla 5: Métricas técnicas de evaluación del pipeline ETL y sistema de reporting 57
Tabla 6: Cinco mejoras prioritarias identificadas a partir de los hallazgos de la evaluación.....59
Tabla 7: Resultados del cuestionario SUS..... 60
Tabla 8: Resultados de las tareas de interpretación..... 61
Tabla 9: Valoración general.....62
Tabla 10: Acrónimos.....74

ÍNDICE DE FIGURAS

Ilustración 1: Modelo de interacción de Shneiderman (1996)..... 20
Ilustración 2: Primeros 5 registros dataset.....36
Ilustración 3: Distribución de los tipos de engangement.....37
Ilustración 4: Distribución de los formatos de banner 37
Ilustración 5: Scorecards de KPIs con codificación cromática semántica 41
Ilustración 6: Página 1 dashboard Looker - Visión general..... 42
Ilustración 7: Página 2 dashboard Looker - Rendimiento por campaña..... 43
Ilustración 8: Página 3 dashboard Looker - Análisis de creatividades 44
Ilustración 9: Página 4 dashboard Looker - Análisis de conversiones y ventas
..... 45
Ilustración 10: Página 1 dashboard Power BI - Visión general.....47
Ilustración 11: Página 2 dashboard Power BI - Rendimiento por campaña..... 48
Ilustración 12: Página 3 dashboard Power BI - Análisis de creatividades 49

1. Introducción

Actualmente, todas las organizaciones generan y almacenan grandes volúmenes de datos que pueden ser convertidos en una ventaja competitiva si son bien analizados. Esto ha llamado a las empresas con una creciente disponibilidad de datos; los cuales pueden provenir de actividades empresariales como ventas de sus productos o servicios, interacciones digitales como visitas a una página web o procesos internos como la distribución interna de tareas y flujos de trabajo; la necesidad de implementar herramientas que ayuden a entenderlos y aprovechar el alto potencial significativo que tienen.

La visualización de datos combinada a principios de UX/UI y técnicas de storytelling digital es conocida como la herramienta más efectiva para facilitar la comprensión de esos datos, la comunicación de los hallazgos dentro de ellos y la toma de decisiones en áreas como negocio y marketing. Representar los datos de una forma coherente y visual hace posible que personas con poca formación técnica o nula puedan comparar datos y entender qué está pasando para encontrar problemas e implementar mejoras.

Las técnicas de narrativa visual generan a los datos una estructura y un sentido que nos ayuda a ponerles un contexto y guiar a la persona que está leyéndolos a unas conclusiones más relevantes. Nos permiten construir una estructura que explica que datos son importantes y porqué aparte de explicar su significado y como nos pueden influir a la hora de tomar una decisión. Al utilizar informes automáticos y paneles interactivos podemos actualizar la información al instante y sin esfuerzo, lo cual es estupendo cuando quieres disponer de datos recientes y fiables para la toma de decisiones.

En los ámbitos de negocio y marketing es muy importante comprender los datos y hacer que otros departamentos o dirigentes con menor capacidad técnica los comprendan ya que es fundamental comunicar resultados de una forma ágil y convincente. Estos elementos nos ayudan a convertir procesos lentos y costosos como la elaboración de informes en hojas de cálculo a mano o reuniones donde se explican datos complejos a personas no preparadas por procesos más ágiles y eficientes.

Este Trabajo de Fin de Grado está centrado en el estudio, diseño y aplicación de métodos de visualización interactiva y generación automática de informes con el objetivo de crear narrativas visuales que disminuyan la dificultad del análisis y hagan el análisis de datos más eficaz. Está basado en un enfoque mixto donde, por un lado, se

desarrolla un marco teórico basado en buenas prácticas de diseño y comunicación visual y, por el otro, una parte práctica basada en el diseño de dashboards interactivos. El caso de uso tiene como tema el ámbito de la publicidad digital. Se ha utilizado un conjunto de datos público proveniente de Kaggle, el cual recoge el rendimiento de tres campañas publicitarias durante el segundo trimestre de 2020. Los datos incluyen métricas clave del sector como impresiones, clics, coste, ingresos y conversiones, segmentadas por formato de banner, emplazamiento publicitario y nivel de engagement del usuario. Este dataset permitirá comprender cómo se deben diseñar visualizaciones útiles y claras, y demostrar el valor que aportan en un caso de uso real de analítica de campañas digitales.

El propósito final es ver como la narrativa y el diseño pueden aclarar la interpretación de los datos para aportar valor a la empresa/organización y acelerar la toma de decisiones. Se busca evidenciar que este enfoque, aparte de simplificar el análisis de datos, convierte a información que antes era muy difícil de analizar y solo expertos en el sector podían entender en un elemento estratégico para tanto la parte interna de la empresa como la externa. Permitiendo a las empresas encontrar nuevos retos y ser capaces de diferenciarse de la competencia analizando de una manera más eficaz el mercado y sus competidores.

2. Objetivos

El objetivo general de mi Trabajo de Fin de Grado es diseñar y evaluar un sistema de visualización interactiva y generación automática de informes para la mejora de comunicación de resultados y la toma de decisiones en entornos de negocio y marketing.

Identificando buenas prácticas y recomendaciones para futuros proyectos de análisis y visualización, se va a demostrar que, si implementamos unos buenos principios de diseño centrado en el usuario, a la narrativa visual con automatización para tener datos en tiempo real. Podremos mejorar la comunicación interna dentro del departamento y fuera, generando un valor estratégico dentro de la organización.

En el presente trabajo se proponen los siguientes objetivos principales y específicos:

O1 - Diseñar la solución de visualización y storytelling

Diseñar bien un dashboard requiere trabajo previo. Hay que revisar qué dice la literatura sobre visualización, UX/UI y storytelling, evaluar qué herramientas existen y definir con criterio qué gráficos usar y por qué. Sin esta base, las decisiones de diseño son arbitrarias.

O1.1 - Revisión de buenas prácticas UX/UI y storytelling

Revisar guías y documentos técnicos orientados en dashboards e informes automatizados. Evaluar herramientas como Power BI o Looker Studio e identificar buenas prácticas en jerarquía visual, colores y diseño. Transmitidas en el desarrollo de un dashboard interactivo estructurado en cuatro páginas, cada una orientada a responder preguntas de negocio concretas

O1.2 - Criterios de diseño

Definir que gráficos son más adecuados según los datos y el objetivo del análisis (series temporales, distribuciones, comparaciones de categorías).

O1.3 - Selección de KPIs

Identificar y seleccionar los indicadores clave de rendimiento (KPIs) más relevantes para este caso, diferenciando entre métricas de volumen, métricas de eficiencia y métricas de calidad. Definir para cada KPI su fórmula de cálculo, su interpretación en el contexto de negocio y el benchmark que permita al usuario contextualizar el dato y tomar decisiones a partir de él.

O1.4 - Estructura narrativa de páginas

Justificar qué elementos no incluimos y cuales sí.

O2 - Implementar el sistema (dashboards y automatización de reporting)

Con los criterios definidos, se construyen los dashboards en Looker Studio y Power BI, se integra el dataset mediante un pipeline ETL y se automatiza la actualización periódica de los datos. El resultado es un sistema funcional conectado a una fuente de datos real.

O2.1 - Construcción de dashboards (Looker/Power BI)

Implementar elementos de interacción como filtros o segmentaciones para tener más detalle a la hora de analizar los datos. Garantizar que los dashboards se pueden adaptar a distintos dispositivos y tamaños de pantalla y funcionan de la misma manera. Documentar la arquitectura de los dashboards para que sean replicables y escalables en un entorno empresarial real.

O2.2 - Integración de datos

Integrar de una forma correcta la información que proviene de un archivo CSV o una base de datos SQL y tener en cuenta su actualización y la consistencia de los datos.

O2.3 - Pipeline ETL

Diseño de un flujo de ETL para que consolide los datos de las distintas fuentes

O2.4 - Actualización/automatización

Automatizar la actualización de los datos y la generación de los informes con Python o Power BI para no tener que realizarlo de forma manual y poder causar errores

O2.5 - Opciones de despliegue

Analizar y comparar las posibles formas de publicación y distribución del dashboard disponibles en Looker Studio y Power BI. Evaluar cada opción en cuanto a coste, accesibilidad, permisos y adecuación a distintos perfiles de usuario final, con el objetivo de identificar la opción más adecuada para entornos de negocio reales.

O3 - Evaluar la solución y derivar recomendaciones

Construir el sistema no es suficiente. Hay que comprobar si funciona para quien lo usa, mediante walkthrough por tareas, métricas técnicas del pipeline y el cuestionario SUS con usuarios reales, y traducir los resultados en mejoras concretas y líneas de trabajo futuro.

O3.1 - Evaluación de usabilidad (SUS) y tareas

Para dicha evaluación, se van a utilizar dos tipos de métricas complementarias. Por un lado, el cuestionario SUS (System Usability Scale), el cual sirve para obtener una puntuación estandarizada de usabilidad evaluada en una escala de 0 a 100, permitiendo comparar el sistema desarrollado con puntos de referencia del sector. Por el otro lado, métricas de eficiencia basadas en tareas concretas sobre el dashboard, como el tiempo que se ha necesitado para identificar un patrón o responder una pregunta de negocio específica, junto con la tasa de acierto y error en la interpretación de los gráficos. La evaluación se realizará con una muestra de entre 5 y 8 participantes con perfil de negocio o marketing, pero sin conocimientos técnicos avanzados, para que represente el usuario objetivo del sistema.

O3.2 - Evaluación técnica (actualización, latencia, reproducibilidad)

Realizar pruebas para saber como de rápido interpreta los datos e identifica los problemas. Instaurar unas métricas para una mejor evaluación como la precisión en la interpretación de los gráficos o la actualización de los informes.

O3.4 - Análisis de resultados

Sintetizar y analizar de forma integrada los resultados obtenidos en las distintas fases de evaluación, para obtener una valoración global del sistema desarrollado.

O3.5 - Limitaciones, recomendaciones y líneas futuras.

Proponer mejoras en la técnica y la metodología basadas en los resultados que he obtenido en el desarrollo del sistema y su evaluación. Realizar un documento que sirva de guía para próximos proyectos.

3. Marco teórico

3.1 Visualización de datos en analítica de negocio

En los últimos años, la digitalización de muchos de los procesos y servicios de las organizaciones ha causado un aumento enorme en la generación de datos. Hay actividades empresariales que antes no existían, como las transacciones comerciales, el gran volumen de interacciones de los usuarios en plataformas, las redes sociales o las campañas de marketing digital. La visualización de datos surge como proceso fundamental para transformar esos datos en conocimiento útil.

La visualización de datos es mucho más que una mera representación gráfica, se comprende mejor como un proceso analítico que permite al dueño de los datos

explorarlos y comprenderlos, pudiendo comunicar información compleja. Munzner (2014) definió la visualización como “el uso de representaciones visuales interactivas para amplificar la cognición”. Esto implica que su objetivo no es simplemente mostrar los datos, sino facilitar el razonamiento humano.

En el entorno de los negocios y el marketing, la visualización es beneficiosa en la detección de patrones, tendencias, relaciones o anomalías, las cuales no serían tan sencillas de evidenciar en tablas numéricas. Uno de los mayores referentes en la visualización de datos, como lo es Stephen Few, destaca que un buen diseño visual disminuye la carga cognitiva y aumenta la capacidad de los usuarios para realizar tareas como la comparación de información o la priorización de indicadores (Few, 2006).

Un sistema de Business Intelligence (BI) es un conjunto de herramientas de software que permite la adquisición de datos de una organización, y que proporciona capacidades para la manipulación, transformación y representación gráfica de dichos datos con el objetivo de facilitar su análisis y apoyar la toma de decisiones (Kimball y Ross, 2013). Estos sistemas integran procesos de extracción, transformación y carga de datos (ETL), almacenamiento en repositorios estructurados y capas de visualización que permiten a los usuarios explorar la información de forma interactiva. Las herramientas de BI más actualizadas y potentes del mercado son Power BI, Looker Studio, Tableau o QlikView. Cada una tiene sus propias capacidades de integración, modelado y visualización adaptadas a diferentes perfiles y necesidades organizacionales (Gartner, 2023).

La integración de la visualización de datos en los sistemas de BI (Business Intelligence) ha aumentado la accesibilidad de los usuarios a la analítica y su posibilidad de entender/comprender la información, con unas barreras de entrada pequeñas, pero al mismo tiempo incluyendo grandes capacidades analíticas para los usuarios expertos. Power BI o Looker Studio (entre otras) ayudan a explotar los datos y crear gráficos, favoreciendo la cultura organizacional orientada al dato, donde las organizaciones se apoyan en evidencias visuales comprensibles para la toma de decisiones.

Al fin y al cabo, la visualización cumple también una función comunicativa, ya que, en los entornos empresariales, los resultados deben ser compartidos entre departamentos y equipos con diferentes conocimientos técnicos. Ser capaz de sintetizar información compleja en dashboards (paneles, o cuadros de mando) claros y coherentes es fundamental para la gestión.

3.2 KPIs, OKRs y métricas en entornos de negocio y marketing digital

La correcta definición y clasificación de las métricas en la analítica de negocio y marketing digital es un paso previo fundamental para el diseño de cualquier sistema de visualización. Sin una selección adecuada de los indicadores que se van a medir y comunicar, el dashboard pierde su capacidad narrativa y su utilidad para la toma de decisiones. Como señala Few (2006), un dashboard efectivo no muestra todos los datos disponibles, sino únicamente aquellos que son relevantes para los objetivos del usuario en un momento determinado.

3.2.1. Métricas y su importancia en la toma de decisiones

Una métrica es cualquier medida cuantitativa que permite evaluar el estado o la evolución de un proceso. Sin embargo, no todas las métricas tienen el mismo valor estratégico. Kaplan y Norton (1992), creadores del Balanced Scorecard, ya advertían que las organizaciones tienden a medir lo que es fácil de medir en lugar de lo que realmente importa. Esta distinción entre métricas operativas y métricas estratégicas es clave para diseñar dashboards que aporten valor real a la organización.

En el ámbito del marketing digital, las métricas pueden clasificarse en tres grandes categorías:

- Las métricas de volumen cuantifican la magnitud de una actividad publicitaria, como el número de impresiones, clics o conversiones totales.
- Las métricas de eficiencia relacionan dos magnitudes para evaluar el rendimiento relativo, como el CTR (Click-Through Rate), el CPC (Cost Per Click) o el CPM (Cost Per Mille).
- Las métricas de calidad evalúan el valor generado por la actividad, como el ROAS (Return On Ad Spend), el ingreso por conversión o la tasa de conversión post-clic.

Esta clasificación es muy relevante para el diseño del dashboard, ya que cada tipo de métrica requiere un tratamiento visual diferente. La Tabla 1 recoge la definición de cada categoría y el tipo de visualización más adecuado para comunicarla (Munzner, 2014):

Tabla 1: Clasificación de métricas de marketing digital y visualización recomendada

Tipo de métrica	Definición	Visualización recomendada
Volumen	Cuantifican la magnitud de una actividad publicitaria	Scorecards, gráficos de barras
Eficiencia	Relacionan dos magnitudes para evaluar el rendimiento relativo	Indicadores, gráficos comparativos
Calidad	Evalúan el valor generado por la actividad	Series temporales, gráficos de dispersión

3.2.2. KPIs: indicadores clave de rendimiento

Un KPI (Key Performance Indicator o indicador clave de rendimiento) es una métrica que mide el grado de consecución de un objetivo estratégico concreto. A diferencia de una métrica genérica, un KPI está siempre vinculado a una meta y a un umbral de referencia que permite evaluar si el rendimiento es satisfactorio o no (Parmenter, 2015). En el ámbito de la publicidad digital, algunos ejemplos de KPIs son el ROAS (Return On Ad Spend), la tasa de conversión, el coste por adquisición o el CTR (Click Through Rate), que mide el porcentaje de usuarios que hacen clic sobre un anuncio respecto al total de usuarios que lo han visto. La elección de los KPIs adecuados no es trivial. Seleccionar demasiados indicadores genera ruido visual y dificulta la interpretación, mientras que seleccionar muy pocos pueden ocultar problemas relevantes. Few (2006) recomienda limitar el número de KPIs visibles simultáneamente a un máximo de siete para no sobrecargar la capacidad cognitiva del usuario.

La correcta visualización de los KPIs en un dashboard requiere que estos estén acompañados de un valor de referencia o benchmark que permita al usuario contextualizar el dato. Un ROAS de 1,92 por sí solo no dice nada, su significado emerge cuando se compara con el objetivo fijado o con la media del sector. Heer y Shneiderman (2012) destacan que la comparación es una de las tareas cognitivas más frecuentes en el análisis visual de datos, y que el diseño del dashboard debe facilitarla explícitamente.

3.2.3. OKRs: objetivos y resultados clave

Los OKRs (Objectives and Key Results) son un marco de gestión por objetivos desarrollado originalmente en Intel por Andy Grove y popularizado posteriormente por

Google (Doerr, 2018). A diferencia de los KPIs, los OKRs definen objetivos ambiciosos a corto plazo y los resultados clave medibles que permiten verificar su consecución. La diferencia fundamental entre ambos marcos es que los KPIs responden a la pregunta "¿cómo estamos funcionando?" mientras que los OKRs responden a "¿hacia dónde queremos ir y cómo sabremos que hemos llegado?".

3.2.4. Vanity metrics vs. actionable metrics

Un concepto relevante en la literatura sobre analítica de negocio es la distinción entre vanity metrics y actionable metrics (Croll y Yoskovitz, 2013). Las vanity metrics son indicadores que aparentan mostrar un buen rendimiento pero que no están vinculados a decisiones concretas ni a resultados de negocio, como el número total de impresiones o los seguidores en redes sociales. Las actionable metrics, en cambio, son aquellas que permiten tomar decisiones directas. Por ejemplo: si el CTR de un banner cae, se puede decidir cambiar el formato; si el ROAS de una campaña no alcanza el objetivo, se puede reasignar el presupuesto.

Esta distinción tiene implicaciones directas para el diseño de dashboards. Un dashboard orientado a la toma de decisiones debe priorizar las métricas accionables sobre las de vanidad, estructurando la narrativa visual de forma que el usuario pueda identificar rápidamente qué está funcionando, qué no y qué acción debe tomar. Este principio guía el diseño del dashboard desarrollado en este trabajo, donde se han priorizado métricas como el ROAS, el CTR y la tasa de conversión sobre métricas de volumen puro como el total de impresiones.

3.3 Fundamentos de percepción y comunicación visual

Los principios de la percepción humana y comunicación visual son elementos fundamentales para el diseño de visualizaciones efectivas. La interpretación de las personas sobre los colores, las formas, las posiciones y los tamaños influyen más de lo que pensamos en la comprensión de los datos.

La teoría de la Gestalt, desarrollada por psicólogos como Wertheimer, Köhler y Koffka a principios del siglo XX, explica cómo el ser humano organiza la información visual a través de principios de proximidad, similitud, continuidad o cierre, los cuales permiten reconocer patrones rápidamente y agrupar elementos (Koffka, 1935). Aplicando esto a los dashboards, facilitaría a los usuarios la identificación de las relaciones entre indicadores y la comprensión de la estructura de la información.

En relación a esto, un uso adecuado de los colores es también importante. El color puede utilizarse para muchas estrategias, como resaltar información relevante, diferenciar categorías o representar magnitudes. Pero si es utilizado de forma excesiva

o inadecuada puede generar confusión. Stephen Few recomienda utilizar paletas limitadas, mantener una consistencia y evitar una combinación de colores que le resulte complicado leer o acceder.

Otro componente esencial sería la jerarquía visual ya que a través del tamaño, la posición y el contraste se establece qué información es prioritaria. En el ámbito de los dashboards, los KPIs (indicadores clave) se deberán destacar visualmente en comparación con los datos secundarios. Permitiendo al usuario comprender de forma rápida que ocurre antes de entrar en detalle.

En conclusión, una visualización bien diseñada no hace que el usuario se fuerce a realizar cálculos mentales innecesarios, sino que presenta los datos de una forma en la que la interpretación sea más sencilla.

3.4 Storytelling y narrativa visual con datos

El storytelling con datos es una combinación entre análisis, visualización y narrativa que tiene el objetivo de comunicar información de manera estructurada. Aparte de mostrar gráficos, construye un relato que guía al usuario a través de los datos aportándole contexto, significado y coherencia.

Segel y Heer (2010) introducen el concepto de visualización narrativa, donde se integran elementos narrativos como pueden ser anotaciones, transiciones o énfasis visual. Estos elementos ejercen de directores para captar la atención del usuario y construir una historia a partir de los datos.

Los investigadores Jessica Hullman y Nicholas Diakopoulos dicen que la visualización no es neutral, sino que siempre transmite una interpretación determinada (Hullman y Diakopoulos, 2011). Entonces, la narrativa visual influye en la percepción de los usuarios de la información y en las decisiones que tomar a partir de ella. Diseñar el storytelling conlleva ser consciente de los efectos comunicativos de cada elección visual.

En el ámbito del marketing y el negocio, permite que las métricas dispersas cambien a mensajes claros para directivos y equipos. Por añadir un ejemplo, no es lo mismo enseñar una tasa de conversión que explicar visualmente cómo un cambio en una campaña afecta al comportamiento del cliente a lo largo del tiempo.

3.5 UX/UI aplicado a dashboards interactivos

Tanto la experiencia del usuario (UX, User Experience) como la interfaz del usuario (UI, User Interface) son componentes esenciales para el diseño de sistemas de visualización interactiva. Aunque ambos conceptos están estrechamente relacionados, responden a preguntas diferentes: UX se centra en cómo el usuario percibe, siente y utiliza el sistema en su conjunto, mientras que UI se ocupa de la disposición visual y la organización de los elementos con los que el usuario interactúa directamente (Norman, 2013).

La usabilidad es el concepto central que articula ambas disciplinas. Nielsen (1994) define la usabilidad como la capacidad de un sistema para ser utilizado por usuarios específicos con el fin de alcanzar objetivos concretos con eficacia, eficiencia y satisfacción en un contexto de uso determinado. Aplicado al diseño de dashboards, un sistema usable es aquel que permite al usuario encontrar la información que necesita en el menor tiempo posible, con el menor número de errores y con una experiencia satisfactoria que fomente su uso continuado.

Si un dashboard tiene una mala usabilidad, puede generar errores de interpretación, hacer que los usuarios se frustren y abandonen su uso, limitando así el valor estratégico del sistema. Miya y Govender (2022) señalan que cuando el diseño está centrado en el usuario, mejora la eficiencia, la satisfacción y la calidad de la interacción, lo que se traduce directamente en una mayor adopción de la herramienta por parte de los equipos de negocio.

Desde la perspectiva de UX, diseñar un dashboard implica comprender en profundidad qué tareas realiza el usuario, qué preguntas de negocio necesita responder y en qué contexto consume la información. A partir de estos puntos se definen los flujos de interacción y navegación, la organización de los elementos visuales y la jerarquía de la información. Nielsen y Molich (1990) propusieron un conjunto de diez heurísticas de usabilidad que han servido de referencia en el diseño de interfaces durante décadas, entre las que destacan la visibilidad del estado del sistema, la consistencia y los estándares, la prevención de errores y la flexibilidad y eficiencia de uso. Estas heurísticas son especialmente relevantes en el diseño de dashboards interactivos, donde el usuario debe ser capaz de navegar entre páginas, aplicar filtros y explorar los datos de forma autónoma sin necesidad de formación previa.

Desde la perspectiva de UI, el diseño se ocupa de aspectos como la tipografía, el espaciado, el uso del color y la disposición de los elementos en el lienzo. Un diseño limpio y coherente reduce la carga cognitiva y mejora la legibilidad, tal como señala

Few (2006), quien recomienda eliminar todos los elementos visuales que no contribuyan directamente a la comprensión de los datos, lo que denomina chartjunk o ruido visual. La incorporación de elementos interactivos como filtros, segmentadores y botones de navegación permite además al usuario explorar los datos según sus intereses y necesidades, transformando el dashboard de un informe estático en una herramienta dinámica de análisis.

La integración de UX y UI con los principios de storytelling digital, descritos en la sección anterior, permite que el usuario no solo observe los datos, sino que pueda experimentarlos mediante una interacción intuitiva y orientada a objetivos concretos. Este enfoque, denominado data-driven UX en la literatura reciente, representa el estándar de diseño en sistemas de Business Intelligence modernos orientados a la toma de decisiones (Heer y Shneiderman, 2012).

3.6 Sistemas de business intelligence y dashboards

Los sistemas de business intelligence integran procesos, herramientas y metodologías predestinadas a convertir datos en información beneficiosa para la gestión de una empresa. Y dentro de dichos sistemas, los dashboards ocupan un lugar central como interfaz entre los datos y los usuarios, es la conexión.

Ben Shneiderman (1996) propone el modelo de interacción “overview first, zoom and filter, then details-on-demand”, el cual describe la secuencia más eficaz para la exploración visual. Primero es ofrecida una visión general, luego se permite filtrar y ampliar y finalmente se accede a los detalles. Un buen enfoque que se adapta muy bien al diseño de los dashboards interactivos.



Ilustración 1: Modelo de interacción de Shneiderman (1996)

En el marketing, los dashboards permiten a las empresas monitorizar sus campañas, analizar embudos de conversión, la segmentación de clientes y la evaluación del rendimiento de acciones comerciales. La capacidad de poder añadir varias fuentes de datos en una única vista mejora la coherencia y reduce la dependencia de informaes manuales.

3.7 Automatización de reporting y pipelines de datos

La automatización de reporting ayuda a la necesidad de disminuir el esfuerzo manual en la generación de informes usuales. La forma tradicional que aún emplean muchas organizaciones es elaborar reportes mediante hojas de cálculo y procesos repetitivos, lo único que consiguen es consumir mucho tiempo y, al ser manual, son propensos a errores.

Los procesos ETL (Extract, Transform, Load) permiten recoger datos de fuentes diferentes, después transformarlos para que sean de calidad y tengan coherencia y ya cargarlos en sistemas analíticos, facilitando la actualización continua de dashboards y reportes.

La automatización, aparte de mejorar la eficiencia, incrementa la fiabilidad de la información. Ya que, si minimizamos la intervención manual, reducimos las inconsistencias y nos aseguramos de que los datos trabajen con datos actualizados. Además, los pipelines pueden ejecutarse en entornos locales o en la nube, adaptándose a las necesidades del proyecto.

Entonces, la automatización también debe tener en cuenta aspectos de seguridad, privacidad y gobernanza del dato, y aún con más intensidad si se trabaja con información de clientes o campañas de marketing.

3.8 Evaluación de sistemas de visualización

El desarrollo de sistemas de visualización finaliza con la evaluación del dashboard, no con su diseño. Al evaluarlo nos permite comprobar si el sistema cumple con los objetivos de apoyo de la toma de decisiones y su usabilidad para los usuarios finales.

Antes de involucrar a ningún usuario, el sistema se puede contrastar con los criterios establecidos por la investigación académica en visualización de datos y diseño de dashboards. Nielsen (1994) formalizó este enfoque en sus heurísticas de usabilidad, que permiten identificar problemas de diseño sin necesidad de reclutar participantes. El proceso consiste en aplicar sistemáticamente esos criterios al sistema desarrollado y documentar las discrepancias entre lo implementado y lo recomendado por la literatura.

Los criterios de referencia más utilizados en el ámbito de los dashboards interactivos provienen de autores como Few (2006), Munzner (2014) y Shneiderman (1996), quienes establecen principios sobre jerarquía visual, uso del color, narrativa estructurada y modelos de interacción. Para sistemas que incorporan pipelines de datos automatizados, Kimball y Ross (2013) añaden otra dimensión al marco

evaluativo: reproducibilidad, robustez ante errores y latencia de actualización, factores que determinan directamente la confianza del usuario en los datos que visualiza.

El walkthrough por tareas consiste en definir un conjunto de acciones representativas de uso real y ejecutarlas paso a paso, adoptando el rol de usuario. En cada paso se documenta qué vería el usuario, qué acción esperaría realizar, si existe alguna ambigüedad que pueda generar confusión y qué mejora concreta resolvería el problema. Albers et al. (2014) defienden este método como el más útil para detectar problemas concretos de interacción que no emergen en evaluaciones puramente subjetivas, precisamente porque obliga al evaluador a recorrer el sistema con un objetivo definido en lugar de explorarlo libremente.

Munzner (2014) complementa este enfoque: la efectividad de una visualización se mide por la capacidad del usuario para completar tareas concretas sobre los datos, no por su aspecto visual. Un gráfico puede parecer bien diseñado y aun así fallar cuando el usuario intenta extraer un insight específico de él. El walkthrough permite detectar exactamente ese tipo de fallo antes de que lo encuentre un usuario real.

Los investigadores Albers, Correll y Gleicher proponen evaluar sistemas de visualización mediante tareas concretas que simulan el uso real, Albers et al. (2014). El planteamiento es simple: en lugar de preguntar al usuario si le gusta el sistema, se le pide que haga algo con él. Buscar el valor máximo de una métrica, identificar una tendencia, comparar dos categorías o filtrar por un segmento específico. Mientras el usuario completa la tarea, se registra si lo consigue, cuánto tarda y dónde comete errores. Este enfoque mide tres dimensiones:

- eficacia (¿completa la tarea?)
- eficiencia (¿cuánto tiempo necesita?)
- tasa de error (¿cuántas veces falla antes de conseguirlo?)

La ventaja frente a los cuestionarios subjetivos es que produce evidencia observable, no lo que el usuario cree que hace, sino lo que realmente hace. En el contexto de los dashboards interactivos, las tareas más representativas son aquellas que replican preguntas de negocio reales: identificar qué campaña tiene mejor ROAS, detectar en qué día de la semana se producen más conversiones o encontrar el formato de banner con mayor CTR.

El cuestionario SUS (System Usability Scale) complementa esta evaluación desde otro ángulo. Donde las tareas miden comportamiento observable, el SUS mide percepción

subjetiva, cómo siente el usuario la experiencia de uso, independientemente de si completó las tareas o no. Brooke (1996) lo desarrolló como una herramienta rápida y robusta: diez ítems con escala Likert de 1 a 5, alternando afirmaciones positivas ("Me resultó fácil usar el sistema") y negativas ("Encontré el sistema innecesariamente complejo"), cuya puntuación se calcula mediante una fórmula estándar que produce un valor entre 0 y 100. Por encima de 68 el sistema se considera usable. Por encima de 85 es excelente (Bangor et al., 2009). Su principal ventaja es la velocidad y su validez, demostrada en miles de estudios desde su publicación. Nielsen (1994) añade que con tan solo 5 participantes es posible detectar la mayoría de los problemas de usabilidad de un sistema, lo que lo convierte en un instrumento especialmente adecuado para estudios con muestras reducidas como el presente trabajo.

4. Metodología

4.1 Estrategia general del proyecto

Para el desarrollo de este TFG se va a adoptar una metodología mixta donde se combinará un estudio teórico y bibliográfico con un proceso práctico de desarrollo y evaluación de prototipos.

Este tipo de metodología permite analizar la evidencia existente sobre la visualización de datos, el storytelling, UX/UI y reporting automático, para después diseñar dashboards interactivos y sistemas de generación automática de informes basados en dicha evidencia. Y, por último, evaluar su usabilidad, eficacia y comprensión mediante métricas cuantitativas y cualitativas.

De esta manera, el proyecto se centra tanto en la implementación técnica como en la comunicación efectiva de la información y la mejora de la toma de decisiones en entornos empresariales.

4.2 Metodología de estudio

4.2.1 Revisión bibliográfica

La primera fase consiste en la revisión exhaustiva de la literatura sobre los temas:

- Automatización de informes y ETL: la integración de los datos que provienen de fuentes distintas y la generación automática de informes en la empresa.
- Visualización de datos: principios de diseño, jerarquía, percepción visual y claridad en los dashboards.
- Narrativa visual y storytelling: técnicas para guiar la interpretación del usuario y resaltar insights relevantes.

- UX/UI aplicado a dashboards: la importancia del diseño de la interfaz y su impacto en la eficiencia, satisfacción y comprensión del usuario.
- Evaluación de dasboards y reporting: métricas de usabilidad, interpretación de datos y eficiencia en la toma de decisiones.

El objetivo, como describimos en el apartado 3 (objetivo O.1), es identificar buenas prácticas, criterios de diseño y herramientas efectivas para guiar la fase práctica.

4.2.2 Análisis comparativo de herramientas

Realizar un análisis comparativo de las herramientas de visualización y reporting más usadas en el ámbito del negocio y marketing es una parte del objetivo O.2. Según el informe anual de Gartner (2023) sobre plataformas de analytics e inteligencia empresarial, ambas herramientas se encuentran entre las más adoptadas a nivel mundial, liderando sus respectivos segmentos de mercado.

Looker Studio:

Looker Studio (anteriormente Google Data Studio) es una herramienta gratuita basada en la nube desarrollada por Google. Su principal fortaleza es la integración nativa con el ecosistema de Google, incluyendo Google Analytics, Google Ads, Google Sheets y BigQuery, lo que la convierte en una opción especialmente adecuada para entornos de marketing digital (Google, s.f.). Permite crear dashboards interactivos con filtros, segmentaciones y controles dinámicos sin necesidad de conocimientos de programación.

Su modelo de datos es relativamente sencillo, trabaja directamente sobre las fuentes de datos sin una capa de modelado intermedia, lo que simplifica el proceso, pero limita la capacidad de realizar transformaciones complejas directamente en la herramienta. En cuanto a la automatización, Looker Studio permite la actualización automática de los datos cuando la fuente está conectada directamente, como es el caso de Google Sheets o BigQuery. Sin embargo, no dispone de un sistema nativo de envío automatizado de informes por correo electrónico, funcionalidad que sí ofrece Power BI (Google, s.f.).

En términos de costes, Looker Studio es completamente gratuito en su versión estándar, lo que lo convierte en la opción más accesible para proyectos académicos y pequeñas empresas. Existe una versión de pago denominada Looker Studio Pro, orientada a equipos empresariales, que añade funcionalidades de gestión de equipos y soporte técnico prioritario. Su principal limitación es la capacidad de procesamiento,

no está diseñado para gestionar volúmenes de datos muy grandes y puede presentar lentitud con datasets de varios millones de registros.

Power BI:

Power BI es la plataforma de inteligencia empresarial de Microsoft, ampliamente adoptada en entornos corporativos (Microsoft, s.f.). A diferencia de Looker Studio, Power BI incorpora una capa de modelado de datos propia denominada DAX (*Data Analysis Expressions*), un lenguaje de fórmulas avanzado que permite crear métricas calculadas, relaciones entre tablas y modelos de datos complejos directamente en la herramienta. Esto lo convierte en una opción más potente para análisis avanzados pero también más compleja de aprender.

En cuanto a la integración de datos, Power BI admite más de 100 conectores nativos, incluyendo bases de datos SQL, Azure, Salesforce, SAP y servicios web mediante APIs REST (Microsoft, s.f.). Su capacidad de automatización es superior a la de Looker Studio, permite programar actualizaciones de datos, configurar alertas automáticas y enviar informes por correo electrónico de forma periódica mediante Power Automate.

En términos de despliegue, Power BI puede utilizarse en la nube a través del servicio Power BI Service o de forma local mediante Power BI Report Server, lo que lo hace adecuado para organizaciones con requisitos estrictos de privacidad y gobernanza del dato. Su modelo de licencias es más complejo: dispone de una versión gratuita de escritorio (Power BI Desktop) con capacidades limitadas de compartición, y versiones de pago (Power BI Pro a 9,40€/usuario/mes y Power BI Premium desde 4.995€/mes) necesarias para la colaboración en equipo y el despliegue empresarial (Microsoft, s.f.). Su principal limitación es el coste para equipos grandes y la curva de aprendizaje asociada al lenguaje DAX.

Tabla comparativa resumen:

Tabla 2: Comparativa de herramientas

Criterio	**Looker Studio**	**Power BI**
Coste	Gratuito	Desde 9,40€/usuario/mes
Integración de datos	Alta (ecosistema Google)	Muy alta (+100 conectores)

Criterio	Looker Studio	Power BI
Modelado de datos	Básico	Avanzado (DAX)
Automatización de informes	Limitada	Alta (Power Automate)
Curva de aprendizaje	Baja	Media-Alta
Despliegue en la nube	Sí	Sí
Despliegue local	No	Sí (Report Server)
Límite de datos	Moderado	Alto
Gobernanza del dato	Básica	Avanzada
Adecuación para marketing digital	Muy alta	Alta

Justificación de la elección

Para el desarrollo de este trabajo se seleccionó Looker Studio como herramienta principal por varias razones.

- En primer lugar, su coste cero elimina barreras de acceso y lo hace replicable en cualquier organización sin necesidad de inversión en licencias.
- En segundo lugar, su integración nativa con Google Sheets permite conectar directamente el dataset procesado en Python sin necesidad de infraestructura adicional.
- En tercer lugar, su orientación específica al marketing digital y su compatibilidad con fuentes de datos como Google Analytics y Google Ads lo posicionan como la herramienta más adecuada para el caso de uso de este trabajo (Google, s.f.).

Finalmente, su curva de aprendizaje reducida es coherente con el objetivo del TFG de demostrar que herramientas accesibles pueden producir visualizaciones de alto valor narrativo sin requerir conocimientos técnicos avanzados.

Cabe señalar que cuando en este trabajo se hace referencia a Power BI, se contempla exclusivamente el uso de Power BI Desktop, la versión gratuita de escritorio. La versión en la nube, Power BI Service, requiere licencia de pago y depende de un ecosistema de herramientas de Microsoft que excede los propósitos de este trabajo académico.

4.3 Metodología de desarrollo

4.3.1 Diseño de dashboards y storytelling

El diseño de dashboards debe seguir principios de UX/UI y narrativa visual, aplicando conceptos de jerarquía visual y técnicas de storytelling.

Pasos principales:

1. Definir los objetivos de negocio y preguntas clave: establecer los KPIs y las métricas que los dashboards deben comunicar.
2. Estructuración de la narrativa visual: decidir el orden de presentación de la información, el uso de anotaciones y transiciones para guiar al usuario.
3. Diseño iterativo del dashboard: crear prototipos iniciales y probarlos internamente para luego ajustarlos según los criterios de usabilidad y claridad visual.
4. Implementar la interactividad: añadir filtros, segmentaciones y enlaces para una exploración más dinámica de la información.

4.3.2 Integración de datos y automatización de informes

Debe implementarse el flujo ETL (Extract, Transform, Load) para garantizar que los dahsboards siempre reflejen los datos actualizados y que los informes se puedan generar de forma automática.

1. Extracción de datos: se obtienen de archivos csv y bases de datos SQL.
2. Transformación y limpieza: manejar valores faltantes, outliers y normalizar variables.
3. Carga en dashboards y sistema de reporting: los datos ya limpios son integrados en Power BI para generar informes automáticos periódicos y bajo demanda.

Este proceso nos asegura la consistencia, fiabilidad y reproducibilidad de los datos que aparecen en los dashboards e informes.

Automatización de la ejecución periódica del pipeline ETL:

Una vez desarrollado el pipeline ETL, el siguiente paso es automatizar su ejecución para que el dashboard de Looker Studio se actualice periódicamente sin intervención manual. Existen varias alternativas según el entorno de despliegue, cada una con sus ventajas e inconvenientes.

- **Opción 1: Programador de tareas del sistema operativo (entorno local)**

Esta opción consiste en configurar el sistema operativo del equipo local para que ejecute el script Python de forma periódica y automática. En sistemas Windows se utiliza el Programador de tareas, mientras que en sistemas macOS y Linux se utiliza cron, permite programar la ejecución de comandos en intervalos definidos.

En Windows, la configuración se realiza a través de la interfaz gráfica del Programador de tareas, donde se define la frecuencia de ejecución (diaria, semanal, etc.) y la ruta al ejecutable de Python junto con el script ETL. En macOS y Linux, la configuración se realiza editando el archivo crontab mediante el comando `crontab -e` e introduciendo una línea con la expresión de programación y la ruta al script:

```
bash# Ejecutar el pipeline ETL cada día a las 8:00
0 8 * * * /usr/bin/python3 /ruta/al/script/etl_pipeline.py
```

La principal ventaja de esta opción es su simplicidad, no requiere infraestructura adicional ni costes. Su principal limitación es que el equipo local debe estar encendido en el momento de la ejecución, lo que la hace poco fiable para entornos de producción empresarial.

- **Opción 2: Google Cloud Scheduler + Cloud Functions**

Esta opción despliega el pipeline ETL en la infraestructura de Google Cloud, eliminando la dependencia del equipo local. Cloud Functions permite ejecutar el script Python como una función serverless en la nube, mientras que Cloud Scheduler la invoca periódicamente según una expresión cron configurada. El flujo completo es el siguiente: Cloud Scheduler lanza la función a la hora programada, Cloud Functions ejecuta el pipeline ETL en la nube, los datos se cargan en Google Sheets y Looker Studio los detecta automáticamente en tiempo real.

La configuración requiere una cuenta de Google Cloud, habilitar las APIs de Cloud Functions y Cloud Scheduler, desplegar el script como una función y configurar el trigger temporal. Google Cloud ofrece una capa gratuita que cubre un número limitado de ejecuciones mensuales, suficiente para un pipeline de actualización diaria en proyectos de tamaño medio. Su principal ventaja es la fiabilidad y disponibilidad

continua, el pipeline se ejecuta independientemente del estado del equipo local. Su principal limitación es la complejidad de configuración inicial y los costes asociados a volúmenes de datos muy grandes.

- **Opción 3: Google Apps Script**

Google Apps Script es un entorno de scripting basado en JavaScript integrado directamente en el ecosistema de Google. Permite automatizar tareas dentro de Google Workspace (incluyendo Google Sheets) mediante scripts que se ejecutan en los servidores de Google sin necesidad de infraestructura adicional. Para el caso de uso de este trabajo, Apps Script puede configurarse para invocar periódicamente una función que actualice los datos en Google Sheets, que Looker Studio consume en tiempo real. Su principal ventaja es que es completamente gratuito y no requiere configuración de infraestructura cloud. Su limitación es que está diseñado para operaciones dentro del ecosistema de Google y no es adecuado para transformaciones de datos complejas realizadas en Python.

Comparativa de opciones:

Tabla 3: Comparativa de métodos de automatización de la ejecución periódica del pipeline ETL

Criterio	**Tarea local**	**Google Cloud**	**Apps Script**
Coste	Gratuito	Gratuito (límites) / De pago	Gratuito
Disponibilidad	Depende del equipo	24/7	24/7
Complejidad de configuración	Baja	Alta	Media
Adecuación para producción	Baja	Muy alta	Media
Dependencia de infraestructura	Local	Google Cloud	Google Workspace
Lenguaje	Python	Python	JavaScript
Escalabilidad	Baja	Muy alta	Media

Opción óptima para un entorno empresarial:

Para un entorno empresarial real, la opción más adecuada es Google Cloud Scheduler combinado con Cloud Functions, porque esta arquitectura garantiza la disponibilidad continua del pipeline independientemente del estado de los equipos locales, escala automáticamente ante aumentos en el volumen de datos, ofrece registros de ejecución para monitorizar posibles errores y se integra de forma nativa con el resto del ecosistema de Google donde ya viven Looker Studio y Google Sheets. Además, su modelo de precios de pago por uso la hace económicamente eficiente para empresas que no requieren actualizaciones muy frecuentes.

Para el desarrollo de este TFG se ha implementado la automatización mediante un scheduler local desarrollado en Python utilizando la librería schedule. El archivo *ejecutor_periodico.py* ejecuta automáticamente el pipeline ETL cada 15 minutos mientras está en ejecución, sincronizándose con el intervalo mínimo de actualización de datos de Looker Studio desde Google Sheets. De esta manera se garantiza que el dashboard refleje siempre los datos más recientes con una latencia máxima de 15 minutos desde la actualización del CSV original. Es completamente gratuita, no requiere infraestructura adicional y es suficiente para demostrar el funcionamiento del sistema en un entorno académico, pudiendo ser utilizado por pequeñas empresas donde no se disponga de infraestructura cloud.

4.4 Evaluación de la solución

Construir el sistema no es suficiente, hay que comprobar que funciona para quien lo usa y que los datos que muestra son fiables. La evaluación se estructura en tres fases secuenciales, cada una con un objetivo distinto y complementario.

La primera fase (Walkthrough por tareas) consiste en recorrer el dashboard adoptando el rol de usuario y ejecutando un conjunto de tareas representativas de uso real. Para cada tarea se documenta qué ve el usuario, qué acción esperaría realizar, dónde aparece fricción y qué mejora concreta resolvería el problema. Este método no requiere reclutar participantes y permite identificar problemas de diseño antes de exponer el sistema a usuarios reales (Nielsen, 1994). Las tareas se organizan en dos bloques: cuatro orientadas al contenido de cada página del dashboard y cuatro orientadas a la interacción con filtros y elementos de navegación.

La segunda fase (Evaluación técnica del sistema) evalúa el rendimiento del pipeline ETL mediante métricas objetivas y medibles. Kimball y Ross (2013) establecen que un

sistema de reporting automatizado debe evaluarse en términos de reproducibilidad, robustez ante errores y latencia de actualización. En este trabajo se miden ocho métricas concretas: tiempo de ejecución del pipeline, frecuencia de actualización, latencia end-to-end entre la carga en Google Sheets y el refresco de Looker Studio, robustez ante errores mediante la verificación post-carga implementada, reproducibilidad desde el repositorio de GitHub, tamaño del dataset final, resumen del control de calidad aplicado y opciones de despliegue disponibles.

La tercera fase (Evaluación de usabilidad con usuarios) involucra usuarios reales. Se aplica el cuestionario SUS a una muestra de 8 participantes con perfil de negocio o marketing, sin conocimientos técnicos avanzados, que representan el usuario objetivo del sistema. El SUS produce una puntuación de 0 a 100 que permite comparar el sistema con benchmarks del sector (Brooke, 1996). Complementariamente, se plantean cinco tareas de interpretación que miden la tasa de acierto al responder preguntas de negocio concretas sobre el dashboard. Estas tareas aportan evidencia de desempeño real más allá de la percepción subjetiva recogida por el SUS.

El orden de las tres fases no es arbitrario. El walkthrough detecta problemas de diseño antes de involucrar a usuarios externos. La evaluación técnica verifica la fiabilidad de los datos que esos usuarios van a ver. Y el SUS con tareas de interpretación mide si, con el diseño corregido y los datos fiables, el sistema cumple su objetivo final: que cualquier persona con perfil de negocio pueda extraer insights accionables del dashboard sin necesidad de formación técnica previa.

4.5 Posibilidades de despliegue (deployment)

Una vez desarrollado el dashboard, es necesario definir cómo se va a publicar y compartir con los usuarios finales. Las opciones de despliegue varían entre Looker Studio y Power BI, tanto en términos de facilidad de configuración como de coste y permisos de acceso.

Despliegue en Looker Studio:

Looker Studio ofrece varias modalidades de publicación, todas son gratuitas y accesibles desde la propia interfaz de la herramienta.

La primera opción es el uso compartido mediante enlace. Desde el menú Compartir del informe, Looker Studio genera una URL única que puede enviarse a cualquier persona. El propietario directamente puede configurar los permisos de acceso en tres niveles: solo las personas añadidas explícitamente pueden ver el informe, cualquier persona con el enlace puede verlo, o cualquier persona en Internet puede acceder.

Esta opción es completamente gratuita y no requiere que el destinatario tenga una cuenta de Google para visualizar el informe si los permisos están configurados como públicos. La limitación es que el destinatario no puede editar el informe a menos que el propietario le otorgue permisos de edición explícitamente.

La segunda opción es la incrustación en una página web. Looker Studio permite generar un código HTML de incrustación que puede pegarse en cualquier página web o portal corporativo donde se visualiza el dashboard de forma interactiva dentro de la propia web. Esta opción es especialmente útil para empresas que quieren integrar el dashboard en su intranet o en un portal de clientes sin necesidad de que los usuarios accedan directamente a Looker Studio.

La tercera opción es la exportación a PDF. Para situaciones donde se requiere un informe estático, por ejemplo, para enviar por correo electrónico a directivos que no necesitan interactividad. Looker Studio permite exportar cualquier página del dashboard como PDF con un solo clic. Esta opción pierde la interactividad, pero es útil para reporting periódico formal.

La cuarta opción es la programación de entregas por correo electrónico. Looker Studio permite configurar el envío automático del informe en formato PDF a una lista de destinatarios con una frecuencia definida por el propietario. Esta funcionalidad convierte el dashboard en un sistema de reporting automatizado sin que los destinatarios accedan activamente a la herramienta.

En cuanto al control de acceso, Looker Studio gestiona los permisos a través de cuentas de Google. El propietario puede añadir colaboradores con permisos de visualización o edición, revocar accesos en cualquier momento y transferir la propiedad del informe a otro usuario. No existe un sistema de roles avanzado ni gobernanza corporativa en la versión gratuita, para esas funcionalidades si ya sería necesario contratar Looker Studio Pro.

Despliegue en Power BI:

Power BI ofrece opciones de despliegue más potentes, pero también más complejas y costosas que Looker Studio.

La primera opción es el archivo .pbix local. Power BI Desktop genera un archivo con extensión .pbix que contiene el modelo de datos, las transformaciones y las visualizaciones. Este archivo puede enviarse por correo electrónico o compartirse a través de una unidad de red. El destinatario puede abrirlo con Power BI Desktop de forma gratuita siempre que las fuentes de datos estén disponibles en las mismas rutas

locales de su equipo. Esta opción no constituye un despliegue en sentido estricto ya que no es accesible a través de una URL y requiere que el destinatario tenga Power BI Desktop instalado y acceso a los datos originales. Es útil para compartir informes internamente en equipos pequeños, pero no es escalable ni adecuada para entornos empresariales.

La segunda opción es la publicación en Power BI Service. Power BI Service es la plataforma en la nube de Microsoft que permite publicar informes y dashboards en una URL accesible desde cualquier navegador. Para publicar en Power BI Service es necesario disponer de una licencia Power BI Pro (9,40€/usuario/mes) tanto para el publicador como para cada usuario que quiera acceder al informe. Esta licencia incluye funcionalidades de colaboración, programación de actualizaciones automáticas de datos, suscripciones por correo electrónico y gestión de permisos por grupos. Es la opción estándar para despliegue empresarial en entornos Microsoft.

La tercera opción es Power BI Premium. Power BI Premium es la modalidad de licenciamiento orientada a grandes organizaciones, con un coste que parte de 4.995€/mes. Su principal ventaja es que permite compartir informes con usuarios que no disponen de licencia Pro, cualquier persona con una cuenta de Microsoft puede visualizar los informes publicados en un espacio de trabajo Premium sin coste adicional. Incluye además capacidades avanzadas de gobernanza del dato, mayor límite de tamaño de datasets y la posibilidad de despliegue en infraestructura local mediante Power BI Report Server.

La cuarta opción es la incrustación mediante Power BI Embedded. Power BI Embedded es una API que permite integrar informes de Power BI en aplicaciones web propias sin que el usuario final necesite una cuenta de Microsoft. Es la opción más flexible para desarrolladores que quieren integrar visualizaciones de Power BI en portales de clientes o aplicaciones SaaS. Su coste se basa en la capacidad computacional contratada en Azure y puede resultar elevado para proyectos de pequeña escala.

Comparativa de opciones de despliegue:

Tabla 4: Comparativa de opciones de despliegue

Criterio	**Looker Studio**	**Power BI**
Compartir por enlace URL	Gratuito	Requiere licencia Pro (9,40€/usuario/mes)

Criterio	Looker Studio	Power BI
Incrustación en web	Gratuito	Power BI Embedded (coste por capacidad Azure)
Exportación a PDF	Gratuito	Gratuito (Power BI Desktop)
Envío automático por email	Gratuito	Requiere licencia Pro
Acceso sin cuenta del proveedor	Sí (modo público)	No (requiere cuenta Microsoft)
Control de permisos avanzado	Looker Studio Pro	Power BI Pro / Premium
Despliegue local	No	Sí (Power BI Report Server)
Despliegue en la nube	Sí (nativo)	Sí (Power BI Service)
Coste mínimo para uso empresarial	Gratuito	9,40€/usuario/mes
Gobernanza corporativa avanzada	No (versión gratuita)	Sí (Premium)

Para el caso de uso de este TFG, Looker Studio presenta una ventaja clara en términos de accesibilidad y coste de despliegue. El dashboard desarrollado puede compartirse mediante un enlace público sin coste alguno, siendo accesible desde cualquier navegador sin necesidad de que el destinatario disponga de una cuenta de Google. Esta característica lo convierte en la opción más adecuada para organizaciones que necesitan comunicar resultados a perfiles no técnicos de forma rápida y sin barreras de acceso.

Power BI, en cambio, es más adecuado para entornos corporativos con infraestructura Microsoft ya establecida, donde la gobernanza del dato, la integración con Azure Active Directory y las capacidades avanzadas de modelado justifican el coste de las licencias. En ausencia de esa infraestructura previa, el coste y la complejidad de

despliegue de Power BI suponen una barrera significativa frente a la gratuidad y simplicidad de Looker Studio.

5. Desarrollo, resultados y discusión

5.1 Selección del dataset

Para desarrollar la parte práctica de este trabajo se ha seleccionado el conjunto de datos “Online Advertising Digital Marketing Data”, el cual está disponible públicamente en la plataforma Kaggle (Ruddhan, 2020).

En primer lugar, este dataset contiene múltiples canales de marketing digital, lo que permite comparar el rendimiento entre distintas plataformas y tipos de campaña de forma significativa. Esta variedad de canales es un requisito fundamental para crear el dashboard orientado a la toma de decisiones, Esta variedad de canales hace de éste un dataset muy adecuado para crear un dashboard orientado a la toma de decisiones, ya que permite analizar los canales para identificar cuáles generan mayor retorno y eficiencia en distintos segmentos y para distintos objetivos de negocio, tomando la decisión de qué en canales continuar invirtiendo o cuáles cancelar.

En segundo lugar, la estructura del dataset incluye múltiples métricas relacionadas con sector de marketing, como impresiones, clics, conversiones y coste, lo que permite calcular indicadores derivados ampliamente utilizados en analítica de marketing como el CTR, el CPC o el ROI. Estas variables facilitan el diseño de visualizaciones comparativas y la construcción de una narrativa visual coherente a lo largo del dashboard.

En tercer lugar, la variación real entre grupos, canales y tipologías de campaña presentes en los datos permite que las visualizaciones diseñadas generen información útil y diferenciada, lo cual es perfecto para demostrar el valor del storytelling digital aplicado a entornos de negocio. Un dataset con distribuciones uniformes entre grupos no serviría para mostrar el impacto del diseño narrativo sobre la comprensión de los datos.

Y, al ser un conjunto de datos de acceso libre y sin restricciones de uso, puede ser utilizado adecuadamente en un trabajo académico desde el punto de vista ético y legal.

Limitaciones del dataset:

El conjunto de datos abarca un periodo de tres meses, en específico los meses de abril, mayo y junio de 2020, lo que puede ser una limitación en cuanto al análisis de tendencias a largo plazo o patrones estacionales anuales.

No obstante, este periodo es suficiente para los objetivos del presente trabajo, ya que su foco principal no es el análisis predictivo sino el diseño de visualizaciones interactivas y la aplicación de técnicas de storytelling digital para la toma de decisiones.

La variación significativa entre campañas, niveles de engagement y formatos de banner presente en los datos deja claro que las visualizaciones diseñadas van a transmitir información diferenciada y útil, que es el requisito fundamental para demostrar el valor del sistema desarrollado.

5.2 Exploración y análisis de calidad del dataset

Con el objetivo de entender la estructura del conjunto de datos y eliminar posibles problemas de calidad antes de proceder a su transformación, se realizó una exploración inicial exhaustiva del dataset mediante scripts de Python alojados en el repositorio de GitHub del proyecto (Anexo A).

El dataset está compuesto de 15.408 registros y 14 columnas, distribuidas en cinco variables categóricas (month, campaign_number, user_engagement, banner y placement) y nueve variables numéricas (day, displays, cost, clicks, revenue, post_click_conversions, post_click_sales_amount y dos columnas sin nombre). Las observaciones corresponden a combinaciones diarias de campaña, formato de banner, placement y nivel de engagement del usuario, abarcando los meses de abril, mayo y junio de 2020 (Ruddhan, 2020).

	month	day	campaign_number	user_engagement	banner	placement	displays	cost	clicks	revenue
0	April	1	camp 1	High	160 x 600	abc	4	0.01	0	0.00
1	April	1	camp 1	High	160 x 600	def	20170	26.78	158	28.97
2	April	1	camp 1	High	160 x 600	ghi	14701	27.63	158	28.98
3	April	1	camp 1	High	160 x 600	mno	171259	216.88	1796	329.45
4	April	1	camp 1	Low	160 x 600	def	552	0.07	1	0.18

Ilustración 2: Primeros 5 registros dataset

Al analizar la distribución de registros por campaña, podemos ver que la campaña 3 concentra el mayor volumen con 6.917 registros, seguida de la campaña 1 con 6.877,

mientras que la campaña 2 presenta una presencia notablemente inferior con 1.614 registros. Respecto al nivel de engagement, la distribución esta equilibrada entre los tres niveles definidos: Medium (5.489 registros), Low (5.035) y High (4.884).

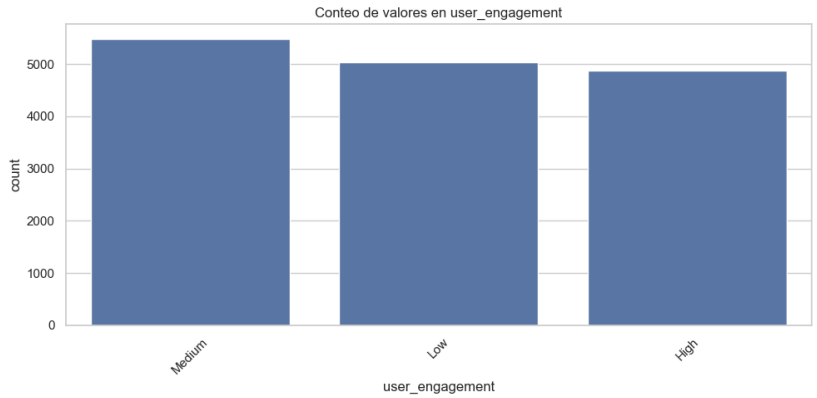

Ilustración 3: Distribución de los tipos de engangement

En cuanto a los formatos de banner, se identificaron ocho tamaños distintos, siendo el más frecuente el formato 728×90 con 3.142 registros, seguido del 300×250 con 2.820.

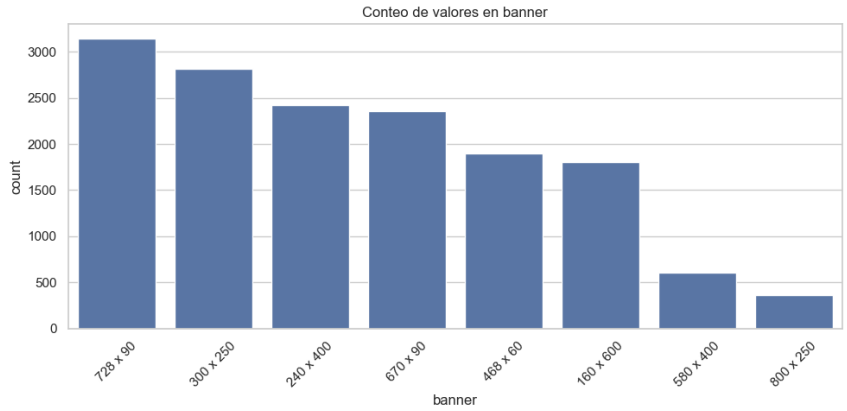

Ilustración 4: Distribución de los formatos de banner

En el análisis de calidad se encontraron tres anomalías. En primer lugar, se identificaron dos columnas completamente vacías (Unnamed: 12 y Unnamed: 13), sin ningún valor en sus registros. Asumiendo que son columnas superfluas creadas accidentalmente durante el proceso de exportación de la información, eliminamos estas columnas. En segundo lugar, en la columna placement se detectaron 413 valores nulos, equivalentes al 2,68% del total, siendo la única variable con datos faltantes. Y, en tercer lugar, se detectaron 5 filas duplicadas en el dataset.

Los valores atípicos se revelaron utilizando el método del rango intercuartílico (IQR), observando la presencia de outliers en todas las métricas numéricas principales, con porcentajes que oscilan entre el 14,65% en displays y el 19,40% en post_click_sales_amount. Estos valores atípicos no son errores, sino que reflejan la

naturaleza heterogénea de la publicidad digital, donde conviven registros de campañas con muy bajo alcance junto a otros con un volumen de impresiones y conversiones considerablemente superior. Se decidió conservarlos en el análisis, siguiendo la recomendación de Tukey (1977), quien advierte que la eliminación sistemática de outliers puede sesgar el análisis cuando estos reflejan variabilidad real del fenómeno estudiado.

Finalmente, se hizo un análisis de variación entre grupos, el cual confirmó que el dataset presenta diferencias significativas entre sus dimensiones. El coeficiente de variación es una medida estadística que expresa la dispersión de los datos con relación a su media, de forma que valores elevados indican que existen diferencias importantes entre los grupos analizados. En este caso, las conversiones post-clic presentaron un coeficiente de variación del 162% según el nivel de engagement del usuario, lo que significa que el número de conversiones varía de forma muy pronunciada entre los usuarios clasificados como High, Medium y Low. De forma similar, el coeficiente de variación entre campañas alcanzó el 150%, evidenciando que las tres campañas analizadas presentan rendimientos muy distintos entre sí en términos de conversiones generadas. Estos valores contrastan con el dataset previamente descartado, que presentaba una variación prácticamente nula entre grupos, y validan la idoneidad de este conjunto de datos para construir visualizaciones con narrativa diferenciada, ya que la variabilidad real entre grupos es precisamente lo que permite que los gráficos transmitan información útil y accionable al usuario.

5.3 Limpieza, transformación y automatización ETL

Una vez completado el análisis de calidad del dataset, se realizó su limpieza, transformación y automatización del pipeline ETL con el objetivo de preparar un conjunto de datos listo para su integración en Looker Studio, herramienta seleccionada como principal para el desarrollo del dashboard interactivo tal como se justificó en el apartado 4.2.2. Adicionalmente, el mismo dataset procesado fue utilizado para el desarrollo del dashboard comparativo en Power BI descrito en la sección 5.5, conectándose en ese caso mediante una URL pública de exportación CSV desde Google Sheets. Todo este proceso se desarrolló en scripts de Python alojados en el repositorio de GitHub del proyecto (Anexo A).

Limpieza:

Se realizaron tres correcciones observadas en el análisis previo. En primer lugar, se eliminaron las columnas Unnamed: 12 y Unnamed: 13 al estar completamente vacías, porque están generadas durante la exportación del archivo original. En segundo lugar,

se eliminaron las 5 filas duplicadas detectadas, bajando los registros del dataset a 15.403. En tercer lugar, los 413 valores nulos presentes en la columna placement fueron imputados con el valor 'Unknown', ya que al representar un porcentaje reducido (2,68%), su eliminación habría supuesto una pérdida de información innecesaria. Esta decisión se fundamenta en la recomendación de García et al. (2015), quienes indican que la imputación mediante un valor constante es adecuada cuando los valores faltantes no siguen un patrón sistemático y su proporción es inferior al 5%. Tras estas operaciones, el dataset quedó completamente libre de valores nulos.

Construcción de la fecha y componentes temporales:

El dataset original no incluía una columna de fecha completa, sino únicamente el mes en formato texto y el día como valor entero. Para poder realizar un análisis temporal en Looker Studio, se construyó una columna Date de tipo datetime combinando ambas variables con el año 2020, correspondiente al periodo real del dataset según la documentación del autor en Kaggle (Ruddhan, 2020). Como resultado se obtuvo un rango temporal que abarca del 1 de abril al 30 de junio de 2020. A partir de esta columna se extrajeron tres componentes adicionales: el número de semana (Week), el nombre del día de la semana (Weekday) y el número de mes (Month_Num), este último con el objetivo de garantizar el orden cronológico correcto en los gráficos de Looker Studio, ya que la herramienta ordena los valores de texto alfabéticamente.

Métricas derivadas:

Con el objetivo de enriquecer el dataset con indicadores estándar del sector del marketing digital, se calcularon seis métricas derivadas a partir de las variables originales. En todos los casos se aplicó protección contra divisiones por cero mediante la función np.where de la librería NumPy. Las métricas creadas fueron las siguientes:

- CTR (Click-Through Rate), calculado como el cociente entre clics e impresiones expresado en porcentaje, el cual mide la proporción de usuarios que hacen clic sobre el anuncio.
- CPC (Cost Per Click), obtenido dividiendo el coste entre el número de clics, que indica el coste medio de cada interacción.
- CPM (Cost Per Mille), calculado como el coste por cada mil impresiones, útil para evaluar la eficiencia del alcance.
- Tasa de conversión (Conv_Rate), que relaciona las conversiones post-clic con el número de clics.

- Ingreso medio por conversión (Revenue_per_Conv), obtenido dividiendo los ingresos entre las conversiones.
- -ROAS (Return On Ad Spend), que expresa el retorno generado por cada unidad monetaria invertida en publicidad (Farris et al., 2010).

Automatización del pipeline ETL:

Para garantizar la actualización continua del dashboard sin intervención manual, se implementó un pipeline ETL automatizado que integra Python con la API de Google Sheets. El pipeline ejecuta secuencialmente las tres fases del proceso ETL: la extracción de los datos desde el CSV original, la aplicación de todas las transformaciones descritas anteriormente y la carga directa de los datos limpios en una hoja de Google Sheets que Looker Studio consume en tiempo real.

Para la autenticación con la API de Google se utilizó OAuth 2.0 mediante la librería Google-auth-oauthlib, que abre una ventana de autorización en el navegador la primera vez que se ejecuta el script y almacena las credenciales localmente para ejecuciones posteriores. La carga de datos se realizó mediante el método Update con el parámetro `value_input_option='USER_ENTERED'`, que garantiza que Google Sheets interprete correctamente los tipos de datos numéricos y de fecha sin convertirlos a texto.

La ejecución periódica del pipeline se implementó mediante el script ejecutor_periodico.py, desarrollado con la librería Schedule de Python. Este script ejecuta automáticamente el pipeline ETL cada 15 minutos mientras está en ejecución, como el intervalo mínimo de actualización de datos configurado en Looker Studio para fuentes de Google Sheets. De esta forma el ciclo completo del sistema queda cerrado con una latencia máxima de 15 minutos y sin intervención manual.

Dataset final:

El dataset resultante tras todo el proceso cuenta con 15.403 registros y 23 columnas, incluyendo las variables originales, los componentes temporales derivados de la fecha y las seis métricas calculadas. Este dataset es consumido directamente por Looker Studio a través de Google Sheets, actualizándose automáticamente cada vez que se ejecuta el pipeline ETL.

5.4 Diseño y desarrollo del dashboard en Looker Studio

El dashboard interactivo fue desarrollado en Looker Studio, conectado directamente a la hoja de Google Sheets que alimenta el pipeline ETL descrito en la sección 5.3

(Anexo A). Está estructurado en cuatro páginas, cada una orientada a responder preguntas de negocio concretas, siguiendo el modelo de interacción "overview first, zoom and filter, then details-on-demand" propuesto por Shneiderman (1996) y recogido en el marco teórico.

Decisiones de diseño:

El diseño del dashboard se fundamenta en los principios de UX/UI y storytelling digital descritos en el marco teórico. Se adoptó una paleta de colores limitada y semántica basada en Google Material Design (Google, s.f.), utilizando el azul como color principal, el verde representa métricas de valor positivo como los ingresos, el rojo identifica el coste como métrica de gasto, y el amarillo se utiliza más para elementos de alerta o referencia. Esta distribución cromática permite al usuario interpretar el tipo de métrica de forma intuitiva sin necesidad de leer las etiquetas, reduciendo la carga cognitiva tal como recomienda Few (2006).

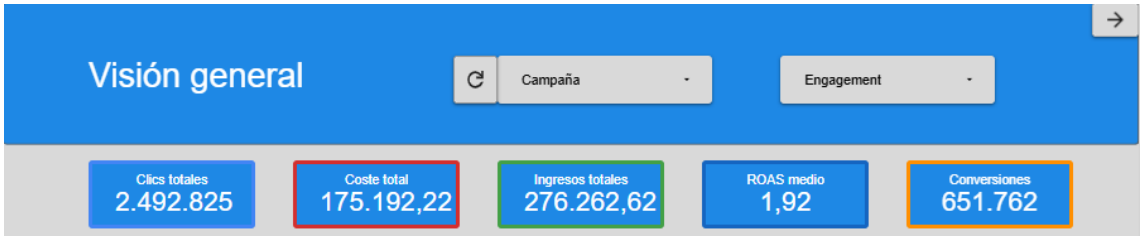

Ilustración 5: Scorecards de KPIs con codificación cromática semántica

Los scorecards de KPIs se situaron en la parte superior de cada página, siguiendo el principio de jerarquía visual que establece que la información más relevante debe ocupar las posiciones de mayor prominencia (Munzner, 2014). Los filtros interactivos por campaña y nivel de engagement se integraron en el header de cada página para que estén siempre accesibles sin interrumpir el flujo de lectura. También se pusieron botones de navegación entre páginas y un botón de restablecimiento de filtros para mejorar la experiencia de usuario.

El fondo de las páginas se configuró en gris muy claro con sombra en los gráficos para generar contraste visual y separar claramente cada elemento del lienzo, aplicando el principio de figura-fondo de la teoría de la Gestalt (Koffka, 1935).

Página 1: Visión general

La primera página genera una visión panorámica del rendimiento global de las campañas publicitarias durante el segundo trimestre de 2020. En la parte superior se presentan cinco scorecards con los KPIs principales: clics totales (2.492.825), coste total (175.192,22 USD), ingresos totales (276.262,62 USD), ROAS medio (1,92) y

conversiones totales (651.762). Cada scorecard utiliza un color de borde diferente según el tipo de métrica para facilitar su identificación visual.

En la zona central se incluyen cuatro gráficos complementarios. El gráfico combinado de barras y líneas "Clics e ingresos por campaña" permite comparar el volumen de tráfico y los ingresos generados por cada campaña, evidenciando que la campaña 1 domina en ambas métricas. El gráfico donut "Distribución de clics por nivel de engagement" muestra que el nivel High concentra el 52,8% de los clics, seguido del nivel Medium con el 43,1%. La serie temporal "Evolución de clics en el tiempo" revela una tendencia decreciente desde abril hasta junio. Finalmente, el gráfico de barras horizontales "ROAS por campaña" incluye una línea de referencia en ROAS=2 como objetivo del sector, permitiendo identificar de forma inmediata qué campañas superan el objetivo de rentabilidad.

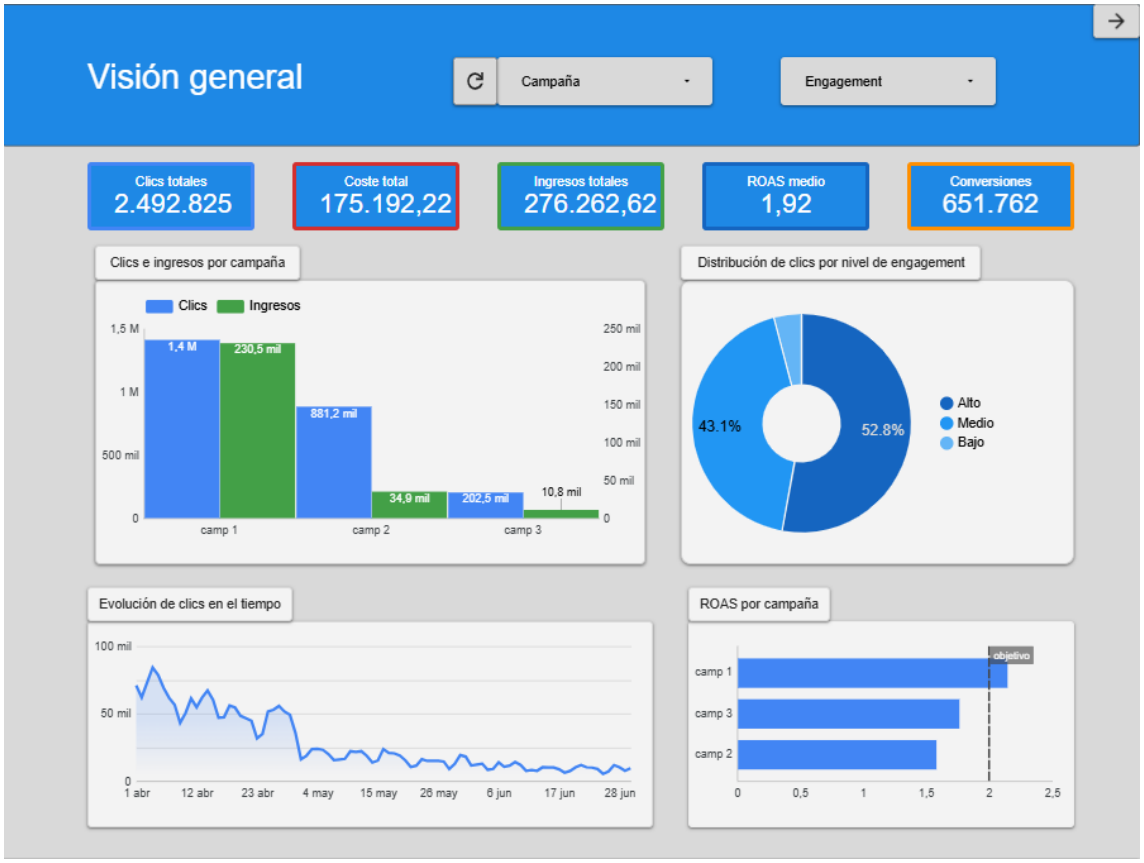

Ilustración 6: Página 1 dashboard Looker - Visión general

Página 2: Rendimiento por campaña

La segunda página profundiza en el análisis comparativo de las tres campañas. Dos indicadores de velocímetro muestran el ROAS medio (1,92) y el CTR medio (0,8) como métricas globales de referencia. El gráfico combinado "Coste e ingresos por campaña" utiliza colores semánticos (rojo para coste y verde para ingresos) para

facilitar la comparación entre inversión y retorno, mostrando que la campaña 1 es la única que genera ingresos claramente superiores al coste. El gráfico circular "Distribución de clics por campaña" revela que la campaña 1 concentra el 56,5% de los clics totales. La serie temporal "Evolución de ingresos por campaña" desglosa la evolución de cada campaña a lo largo del trimestre, donde la campaña 1 en azul muestra una actividad claramente superior y más volátil que las otras dos. Por último, el gráfico de dispersión "Coste vs conversiones por campaña" evidencia una correlación positiva entre el coste invertido y las conversiones obtenidas.

Página 3: Análisis de creatividades

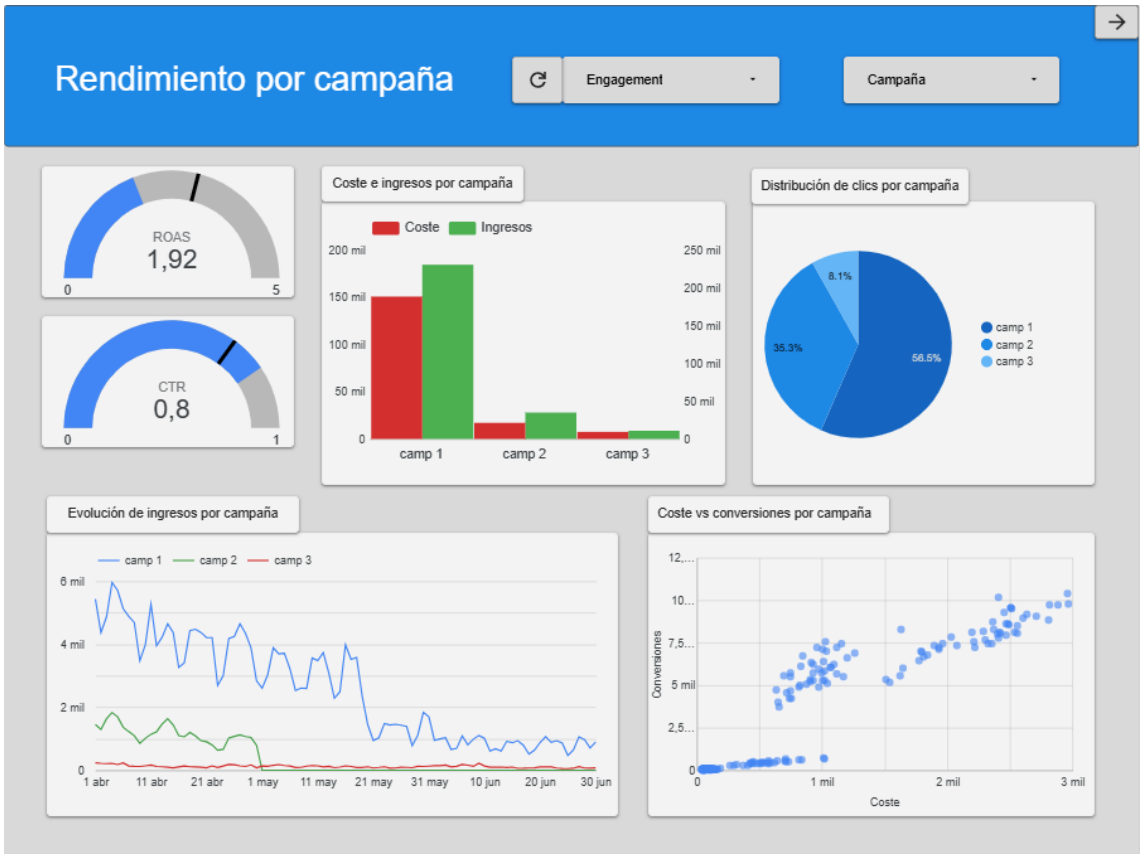

Ilustración 7: Página 2 dashboard Looker - Rendimiento por campaña

La tercera página se centra en el análisis del rendimiento por formato de banner y emplazamiento publicitario. Tres scorecards presentan las métricas específicas de esta dimensión: CTR medio (0,81%), CPC medio (0,06 USD) y CPM medio (0,58 USD). El gráfico de barras "CTR medio por formato de banner" muestra que el formato 580×400 obtiene el CTR más elevado (1,27%), mientras que el formato 468×60 presenta el valor más bajo (0,25%), lo que quiere decir que el tamaño del anuncio influye significativamente en la tasa de interacción. La serie temporal de áreas "Evolución de ingresos por placement" desglosa la contribución de cada

emplazamiento a lo largo del trimestre, con una paleta de colores diferenciada por placement. El diagrama de Sankey "Flujo de clics por banner y placement" es el elemento visualmente más diferenciador del dashboard. Muestra mediante líneas de grosor proporcional al volumen de clics cómo fluye el tráfico desde cada formato de banner hacia cada emplazamiento publicitario, revelando qué combinaciones concentran mayor actividad. Finalmente, el gráfico circular "Conversiones por formato de banner" muestra que el formato 240×400 genera el 41,4% de las conversiones totales.

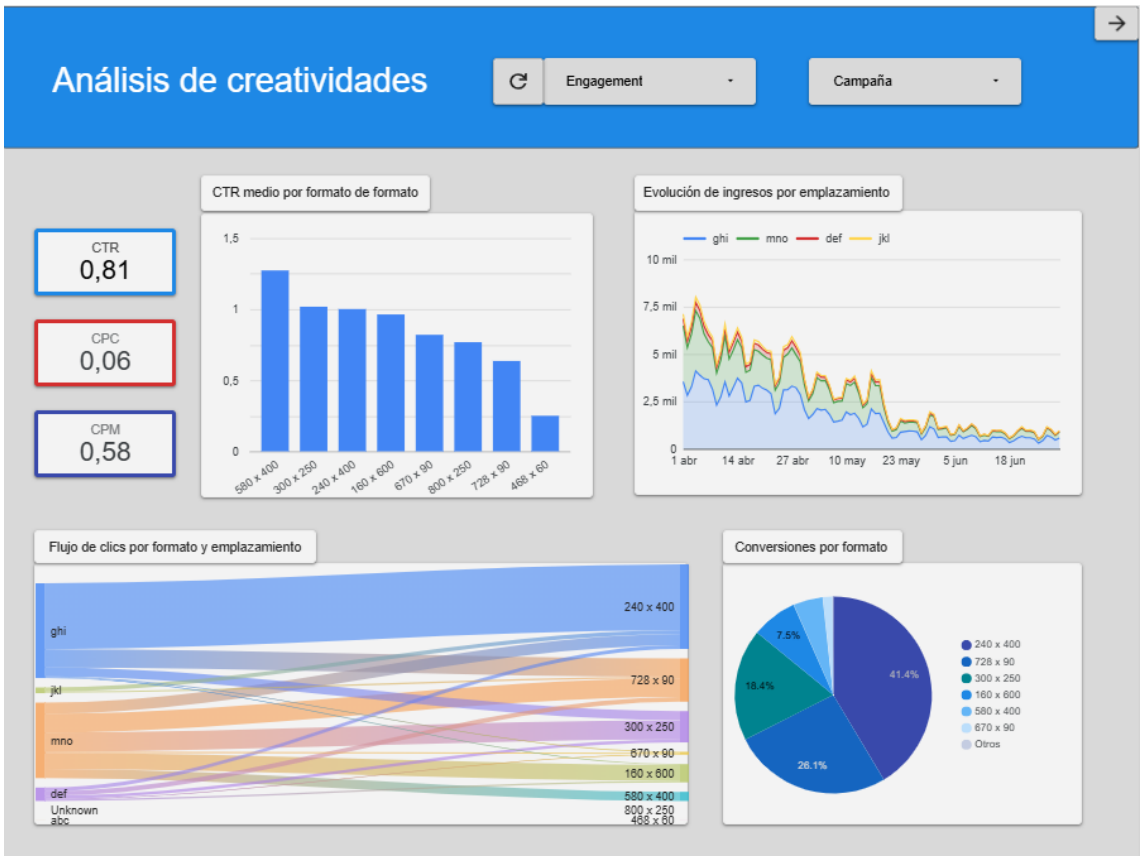

Ilustración 8: Página 3 dashboard Looker - Análisis de creatividades

Página 4: Análisis de conversiones y ventas

La cuarta página explota las variables de conversión y ventas post-clic que no habían sido analizadas en las páginas anteriores. El gráfico combinado "Evolución de conversiones y ventas" superpone ambas métricas con ejes secundarios diferenciados uno a cada lado, mostrando que siguen una evolución paralela con tendencia decreciente a lo largo del trimestre. El gráfico de barras "Conversiones por día de la semana" revela patrones de comportamiento temporal. Los lunes y martes concentran mayor volumen de conversiones que los fines de semana. El gráfico donut "Distribución de ventas por campaña" muestra el descubrimiento más relevante de

esta página, la campaña 1 concentra el 94,8% de las ventas totales, lo que la identifica como la campaña estratégicamente más importante del conjunto. El gráfico de dispersión "Relación entre alcance y conversiones" confirma una correlación positiva clara entre el número de impresiones y las conversiones obtenidas, validando la importancia de la inversión en alcance publicitario. Por último, la serie temporal "Evolución de la tasa de conversión" muestra un incremento progresivo del Conv_Rate a lo largo del trimestre, especialmente pronunciado en junio.

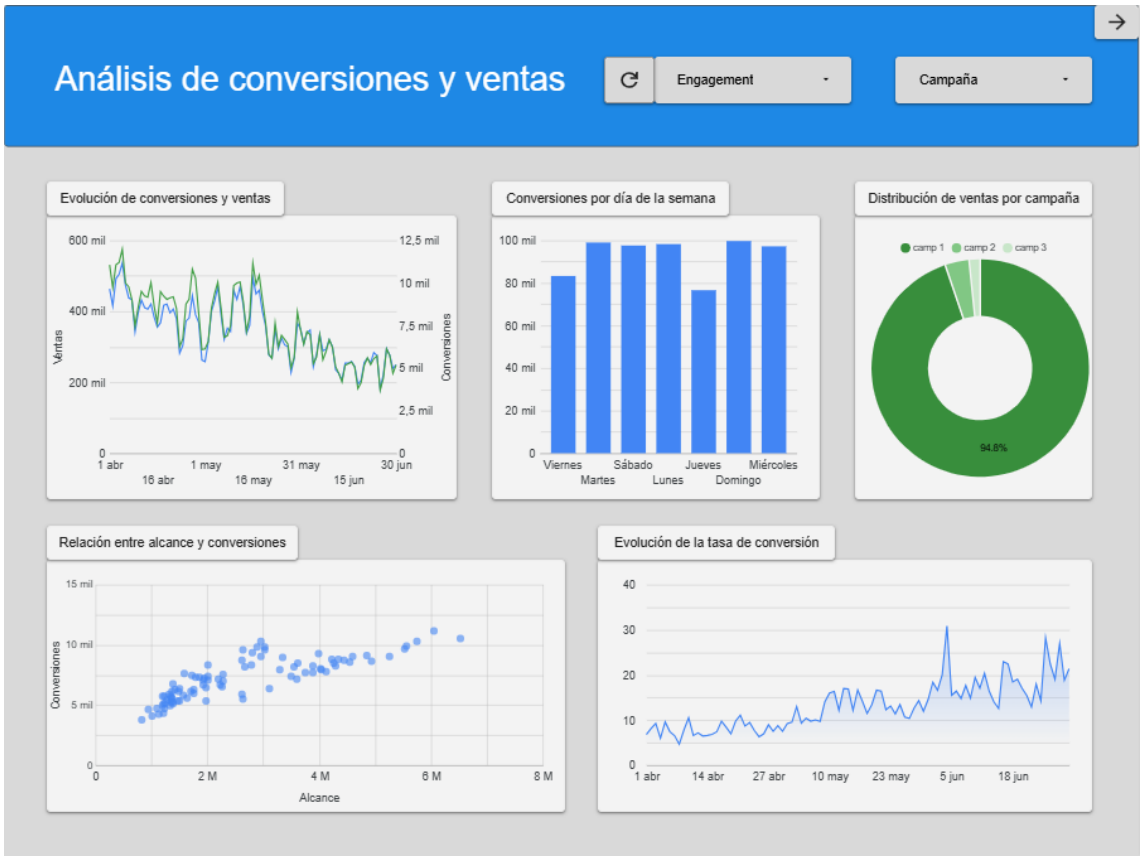

Ilustración 9: Página 4 dashboard Looker - Análisis de conversiones y ventas

5.5 Diseño y desarrollo del dashboard en Power BI

Con el objetivo de realizar una comparativa práctica entre herramientas de visualización, se desarrolló un segundo dashboard en Power BI Desktop replicando la misma estructura y contenido del dashboard de Looker Studio descrito en la sección 5.4 (Anexo C). La fuente de datos es la misma hoja de Google Sheets alimentada por el pipeline ETL, conectada a Power BI mediante una URL pública en formato CSV, lo que garantiza que ambos dashboards consumen siempre los mismos datos actualizados.

Diferencias de configuración respecto a Looker Studio:

Como se explicó en el apartado 4.2.2, cuando se hace referencia a Power BI en este trabajo, se refiere únicamente el uso de Power BI Desktop, la versión gratuita de escritorio de Microsoft. Esta decisión responde a que la alternativa en la nube (Power BI Service), la cual requiere una licencia Pro de 9,40€/usuario/mes para publicar y compartir informes, y su funcionamiento óptimo depende de un ecosistema de herramientas de Microsoft que incluye Azure Active Directory, Power Automate y Power BI Report Server, excediendo el alcance de este trabajo académico. En consecuencia, a diferencia de Looker Studio, que actualiza los datos en tiempo real de forma automática, Power BI Desktop requiere que el usuario haga clic manualmente en "Actualizar" para refrescar los datos, ya que la actualización automática programada solo está disponible en Power BI Service con licencia Pro.

El modelado de datos se realizó directamente en Power BI Desktop sin necesidad de una capa adicional de transformación, ya que el dataset cargado procedía del pipeline ETL ya ejecutado y contenía todas las métricas derivadas calculadas en Python. Los tipos de datos se verificaron manualmente en la vista de tabla, corrigiendo la columna last_updated que había sido interpretada incorrectamente como fecha en vez de texto.

Decisiones de diseño:

Se mantuvo la misma paleta de colores y estructura visual que en el dashboard de Looker Studio para facilitar la comparativa entre herramientas. El header azul con título en blanco se replicó mediante un rectángulo y un cuadro de texto superpuesto. Los filtros se implementaron mediante segmentadores en formato lista desplegable, el equivalente en Power BI a los filtros de Looker Studio. Los scorecards se construyeron con el componente Tarjeta de Power BI, configurando el formato numérico mediante la expresión #,##0 para mostrar los números completos sin abreviación.

Página 1: Visión general

La primera página replica la estructura de la página equivalente en Looker Studio, con cinco tarjetas de KPIs en la parte superior mostrando los mismos valores globales: clics totales (2.492.825), coste total (175.192 USD), ingresos totales (276.263 USD), conversiones totales (651.762) y ROAS medio (1,92). El gráfico combinado "Clics e ingresos por campaña" muestra las barras de clics en azul y la línea de ingresos en naranja con eje secundario, evidenciando la misma dominancia de la campaña 1 que en Looker Studio. El gráfico donut "Distribución de clics por engagement" confirma que el nivel High concentra el 52,85% de los clics. La serie temporal "Evolución de clics en el tiempo" y el gráfico de barras horizontales "ROAS por campaña" completan la página con la misma información que su equivalente en Looker.

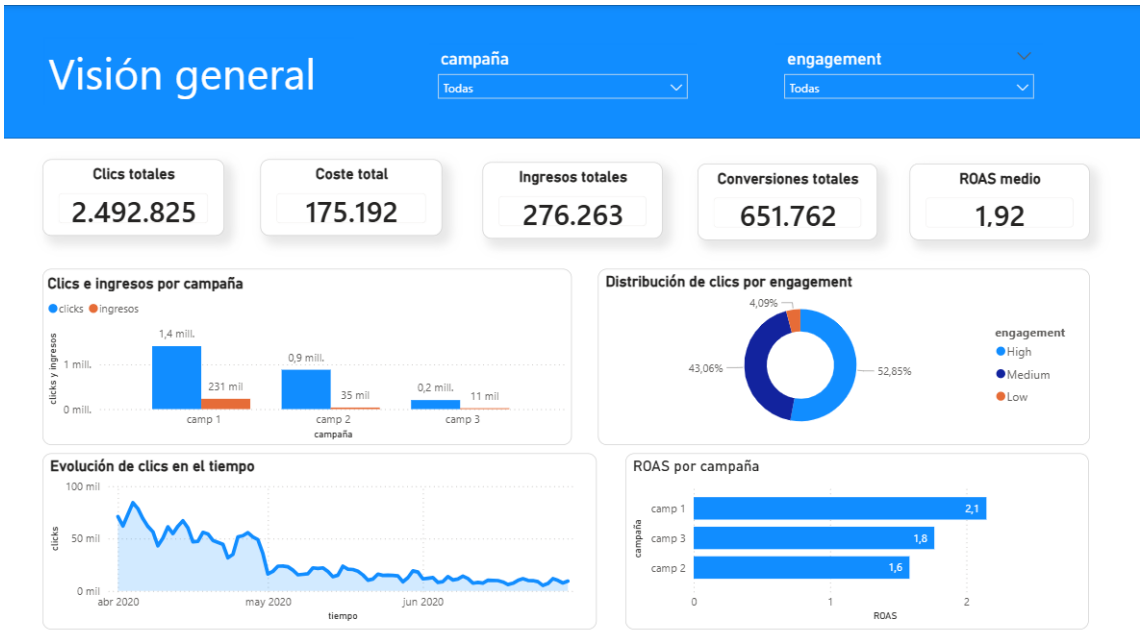

Ilustración 10: Página 1 dashboard Power BI - Visión general

Página 2: Rendimiento por campaña

La segunda página incorpora los dos indicadores de velocímetro (medidor en Power BI) mostrando el ROAS medio (1,92) y el CTR medio (0,81), el gráfico combinado de coste e ingresos por campaña con codificación cromática semántica (verde para ingresos y rojo para coste); el gráfico circular de distribución de clics por campaña, la serie temporal de evolución de ingresos por campaña desglosada por las tres campañas y el gráfico de dispersión de coste vs conversiones. Una diferencia notable respecto a Looker Studio es que el gráfico de dispersión de Power BI muestra todos los puntos individuales del dataset en lugar de agregar por campaña, lo que ofrece una visión más granular de la relación entre coste y conversiones, pero dificulta la comparación directa entre campañas.



Ilustración 11: Página 2 dashboard Power BI - Rendimiento por campaña

Página 3: Análisis de creatividades

La tercera página replica el análisis de formatos de banner y placements. Dos tarjetas muestran el CPC medio (0,06 USD) y el CPM medio (0,58 USD). El gráfico de barras "CTR medio por formato de banner" confirma que el formato 580×400 presenta el CTR más elevado (1,27%) y el formato 468×60 el más bajo (0,25%), resultados consistentes con los obtenidos en Looker Studio. El diagrama de Sankey "Flujo de clics por banner y placement" fue implementado mediante la visualización personalizada Sankey Chart disponible en el marketplace de Power BI, mostrando los mismos flujos de tráfico entre formatos y emplazamientos que en Looker Studio. La serie temporal de áreas "Evolución de ingresos por placement" y el gráfico circular de conversiones por formato de banner completan la página.

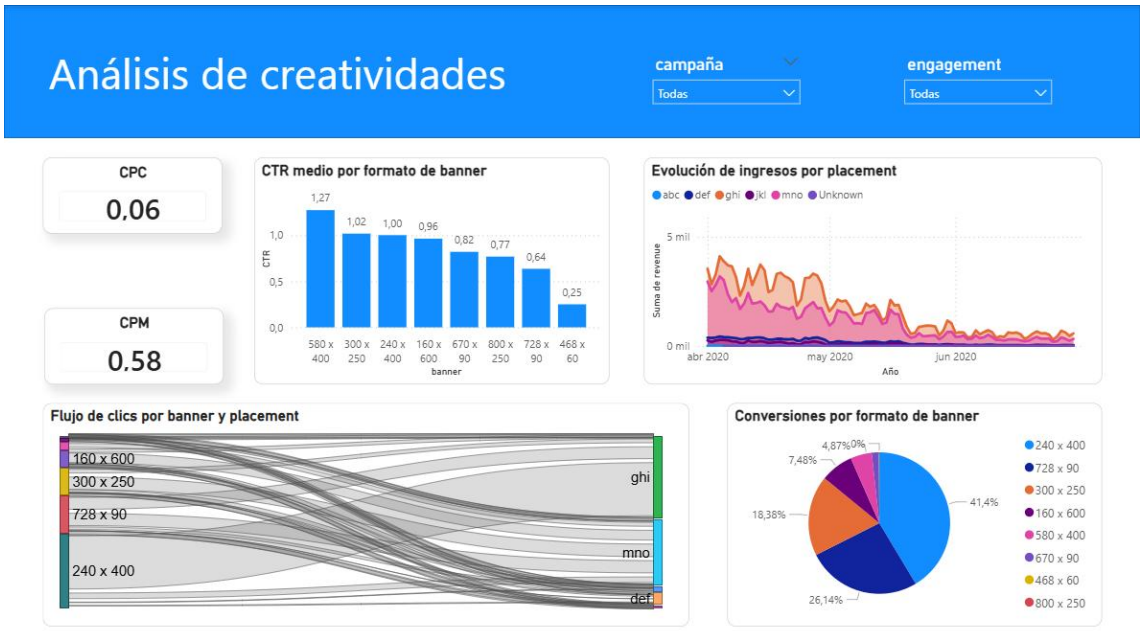

Ilustración 12: Página 3 dashboard Power BI - Análisis de creatividades

Comparativa con Looker Studio:

El desarrollo del dashboard en Power BI permitió identificar diferencias relevantes entre ambas herramientas más allá de las descritas en el análisis teórico de la sección 4.2.2. En términos de facilidad de uso, Looker Studio resultó ser más intuitivo para usuarios sin experiencia previa en herramientas de BI, mientras que Power BI ofrece mayor flexibilidad en la configuración de visualizaciones avanzadas como los medidores o las tablas dinámicas. En cuanto a la paleta de colores, Power BI aplica sus propios colores por defecto que requieren configuración manual para alcanzar la consistencia visual lograda en Looker Studio. La principal ventaja diferencial de Looker Studio fue la actualización automática en tiempo real de los datos, frente a la actualización manual requerida en Power BI Desktop. Looker Studio permite compartir el dashboard mediante un enlace público de forma completamente gratuita, mientras que Power BI requeriría licencia Pro para una publicación equivalente en la nube.

5.6 Pipeline ETL automatizado

La automatización del pipeline ETL es uno de los elementos diferenciadores del sistema desarrollado en este trabajo, ya que permite que el dashboard de Looker Studio muestre siempre los datos más recientes sin necesidad de intervención manual. El pipeline fue implementado en Python mediante mediante Visual Studio Code, con los scripts alojados en un repositorio público de GitHub, y se integra directamente con la API de Google Sheets (Anexo A).

Arquitectura del pipeline:

El pipeline sigue la arquitectura ETL estándar descrita en el marco teórico (3.6), estructurada en cuatro fases secuenciales.

- En la fase de extracción se carga el dataset original en formato CSV mediante la librería Pandas, registrando el timestamp de extracción para trazabilidad del proceso.
- En la fase de transformación se aplican secuencialmente todas las operaciones de limpieza y enriquecimiento descritas en la sección 5.3 como son la eliminación de columnas vacías, eliminación de duplicados, imputación de nulos, construcción de la columna de fecha, extracción de componentes temporales y cálculo de las seis métricas derivadas.
- En la fase de carga los datos transformados se escriben directamente en Google Sheets mediante la librería gspread, utilizando el parámetro `value_input_option='USER_ENTERED'` para garantizar que los tipos de datos numéricos y de fecha se interpretan correctamente.
- Finalmente, la fase de verificación comprueba que el número de filas cargadas en Google Sheets coincide exactamente con el del dataframe procesado, alertando al usuario en caso de discrepancia.

Vale la pena aclarar qué observa realmente el pipeline cuando se ejecuta. El script lee el CSV original almacenado en local, lo transforma y carga los datos resultantes en Google Sheets. Looker Studio nunca accede al CSV, solo lee Google Sheets. Son dos capas completamente separadas.

Esto tiene una implicación directa sobre la utilidad de la ejecución periódica. Si el CSV no cambia entre ejecuciones, el pipeline siempre cargará los mismos datos, lo que significa que el scheduler ejecuta trabajo innecesario. En el contexto de este trabajo eso es así, porque el dataset es estático. La ejecución periódica se implementa para demostrar que la arquitectura funciona y está lista para escalar, no porque el CSV se actualice solo.

En un entorno real el CSV sería sustituido por una fuente de datos dinámica, una base de datos SQL que recibe transacciones en tiempo real, una API de Google Ads que devuelve el rendimiento de las campañas del día, o un archivo que una plataforma publicitaria deposita automáticamente en un servidor cada noche. En ese escenario el scheduler cobra todo su sentido: cada vez que se ejecuta, encuentra datos nuevos, los procesa y los carga en Google Sheets, manteniendo el dashboard siempre actualizado sin que nadie intervenga.

Esta limitación está directamente relacionada con el dataset utilizado y se recoge también en las líneas de trabajo futuro de la sección 7, donde se propone conectar el pipeline a fuentes de datos dinámicas como Google Analytics o Meta Ads a través de sus APIs oficiales.

Autenticación con la API de Google:

La autenticación con la API de Google Sheets se gestiona mediante OAuth 2.0 a través de la librería Google-auth-oauthlib. La primera vez que se ejecuta el script, se abre automáticamente el navegador para que el usuario autorice el acceso a su cuenta de Google. Las credenciales obtenidas se almacenan localmente en un archivo token.pickle, de forma que las ejecuciones posteriores son completamente automáticas sin necesidad de volver a autorizar. Este enfoque es seguro ya que las credenciales se almacenan únicamente en el equipo local y nunca se suben al repositorio de GitHub, estando excluidas mediante el archivo `.gitignore`.

Integración con Looker Studio:

Una vez que el pipeline carga los datos actualizados en Google Sheets, Looker Studio los refleja automáticamente en el dashboard sin necesidad de intervención manual. Aunque, Looker Studio no detecta los cambios en tiempo real de forma instantánea, sino que funciona mediante un sistema de caché con un umbral de frescura de datos configurable.

Para fuentes de datos de Google Sheets, el intervalo mínimo de actualización es de 15 minutos, lo que significa que después de ejecutar el pipeline ETL y actualizar Google Sheets, el dashboard mostrará los datos nuevos en un máximo de 15 minutos sin que el usuario tenga que hacer nada. Si se necesita ver los cambios de forma inmediata, el editor del informe puede forzar una actualización manual desde el menú de Looker Studio.

Existe una opción adicional de auto refresh configurable disponible en Looker Studio Pro que permite recargar los datos del informe periódicamente de forma automática mientras el informe está abierto, pero se necesita la licencia de pago. En el contexto de este TFG, la configuración de frescura de datos por defecto de 15 minutos es suficiente para garantizar que el dashboard refleje siempre los datos más recientes tras cada ejecución del pipeline.

Hay dos intervalos que no hay que confundir. El primero es el del scheduler, cada 15 minutos ejecuta el pipeline y carga los datos en Google Sheets. El segundo es el de

Looker Studio, cada 15 minutos consulta Google Sheets para ver si hay datos nuevos. Son procesos independientes y no están sincronizados.

Configurar ambos intervalos al mismo valor puede generar un problema: si el pipeline tarda más de lo esperado en completarse y Looker Studio consulta Google Sheets justo en ese momento, leerá los mismos datos que antes o, peor, datos a medias. Para evitarlo, lo recomendable es que el intervalo de refresco de Looker sea siempre mayor que el del scheduler, por ejemplo, scheduler cada 7 minutos y Looker cada 15. Así se garantiza que cuando Looker consulta los datos, el pipeline ya ha terminado de cargarlos.

Automatización de la ejecución periódica

Para completar la automatización del sistema, se implementó el script ejecutor_periodico.py desarrollado con la librería Schedule de Python, que ejecuta el pipeline ETL automáticamente cada 15 minutos mientras está en ejecución. Este intervalo coincide con la frecuencia mínima de actualización de datos configurada en Looker Studio para fuentes de Google Sheets, garantizando que el dashboard refleje siempre los datos más recientes con una latencia máxima de 15 minutos desde la actualización del CSV original.

El script funciona manteniendo un bucle continuo que comprueba cada minuto si ha llegado el momento de ejecutar el pipeline. Cuando se cumple el intervalo programado, lanza automáticamente el script etl_pipeline.py como un subproceso independiente, captura su output y lo muestra en el terminal junto con un timestamp de ejecución. En caso de error, el script lo notifica en el terminal sin interrumpir el bucle, garantizando que el scheduler siga funcionando, aunque una ejecución concreta falle.

Esta solución es completamente gratuita y no requiere infraestructura cloud adicional, siendo adecuada para entornos empresariales de pequeña escala. Para entornos que requieran mayor fiabilidad y disponibilidad continua independiente del equipo local, la opción más adecuada seguiría siendo Google Cloud Scheduler combinado con Cloud Functions, tal como se describió en la sección 4.3.2.

5.7 Evaluación del sistema

5.7.1 Criterios de evaluación basados en literatura

La evaluación de un sistema de visualización interactiva no puede limitarse a la opinión subjetiva de sus usuarios ni a una única dimensión de análisis. Para garantizar una evaluación rigurosa y fundamentada, es necesario definir previamente qué se evalúa y por qué, apoyándose en la literatura académica especializada. En este

trabajo se han considerado cuatro dimensiones de evaluación complementarias: la usabilidad percibida, la claridad narrativa, la capacidad de apoyo a la toma de decisiones y la fiabilidad del sistema de reporting automatizado.

La primera dimensión es la usabilidad percibida. Brooke (1996) propone el cuestionario SUS como método estandarizado para medir la facilidad de uso de un sistema desde la perspectiva del usuario, permitiendo obtener una puntuación comparable entre sistemas y detectar problemas de interacción de forma rápida y económica. Complementariamente, Nielsen (1994) establece que la usabilidad se compone de cinco atributos: facilidad de aprendizaje, eficiencia, memorabilidad, tolerancia a errores y satisfacción; que deben tenerse en cuenta al diseñar y evaluar cualquier interfaz interactiva.

La segunda dimensión es la claridad narrativa. Few (2006) establece que un dashboard efectivo debe comunicar la información más relevante de forma inmediata, sin requerir que el usuario realice cálculos mentales adicionales. Munzner (2014) complementa esta idea señalando que la efectividad de una visualización debe medirse en términos de la capacidad del usuario para completar tareas concretas sobre los datos, lo que justifica el uso de tareas guiadas como método de evaluación. Segel y Heer (2010) añaden que la narrativa visual debe guiar al usuario desde una visión general hasta el detalle de forma progresiva y coherente, principio que estructura el diseño de las cuatro páginas del dashboard desarrollado en este trabajo.

La tercera dimensión evalúa la capacidad de apoyo a la toma de decisiones. Heer y Shneiderman (2012) argumentan que un sistema de visualización interactiva debe facilitar tareas cognitivas como la comparación, la detección de tendencias y la identificación de valores atípicos. La evaluación de esta dimensión se realiza mediante tareas de interpretación que simulan preguntas de negocio reales, midiendo la tasa de acierto y los puntos de fricción detectados durante la interacción.

Finalmente, la cuarta dimensión considera la fiabilidad del sistema de reporting automatizado. Kimball y Ross (2013) establecen que un pipeline ETL debe evaluarse en términos de reproducibilidad, robustez ante errores y latencia de actualización, ya que estos factores determinan la confianza del usuario en los datos que visualiza. Esta dimensión se evalúa mediante métricas técnicas objetivas descritas en la sección 5.7.3.

5.7.2 Walkthrough por tareas

El walkthrough por tareas es una técnica de evaluación de usabilidad que consiste en recorrer el sistema siguiendo una serie de tareas representativas de uso real,

documentando para cada una los pasos esperados, los puntos de fricción detectados y las mejoras propuestas (Nielsen, 1994). A diferencia de una prueba con usuarios, el walkthrough es realizado por el propio desarrollador adoptando el rol de usuario, lo que permite identificar problemas de diseño de forma sistemática sin necesidad de reclutar participantes externos. Se definieron ocho tareas distribuidas en dos bloques: cuatro orientadas al contenido de cada página y cuatro orientadas a la interacción con los elementos del dashboard.

Tareas orientadas al contenido:

- Tarea 1: Visión general: identificar la campaña con mejor ROAS y justificarla con dos KPIs adicionales

El usuario accede a la página 1 y localiza el gráfico de barras horizontales "ROAS por campaña" en la zona inferior derecha. La campaña 1 se identifica visualmente como la de mayor ROAS de forma inmediata gracias a la longitud de la barra y a la línea de referencia de objetivo en ROAS=2, que permite además contextualizar el dato. Para justificar la respuesta con dos KPIs adicionales, el usuario debe recurrir al gráfico "Clics e ingresos por campaña" y al scorecard de ingresos totales, lo que requiere combinar información de tres elementos distintos de la página. El punto de fricción detectado es que no existe un elemento visual que agrupe los KPIs de una campaña concreta en un único lugar, lo que obliga al usuario a hacer una lectura comparativa entre varios gráficos. Una mejora propuesta sería añadir un tooltip enriquecido sobre cada barra del gráfico de ROAS que muestre automáticamente los KPIs asociados a esa campaña al pasar el ratón.

- Tarea 2: Rendimiento por campaña: detectar si hay campañas con coste mayor que ingresos y qué implicación tiene

El usuario navega a la página 2 y localiza el gráfico combinado "Coste e ingresos por campaña". La codificación cromática semántica (rojo para coste y verde para ingresos) permite identificar de forma inmediata que en la campaña 1 los ingresos superan claramente al coste, mientras que en la campaña 2 y la campaña 3 la diferencia es mínima y visualmente difícil de cuantificar debido a la escala del eje, que está dominada por los valores de la campaña 1. El punto de fricción es que las barras de las campañas 2 y 3 son tan pequeñas en comparación con la campaña 1 que resulta complicado determinar con precisión si el coste supera o no a los ingresos. Una mejora propuesta sería añadir etiquetas de datos sobre cada barra o permitir al usuario filtrar por campaña para ver cada una con su propia escala.

- Tarea 3: Análisis de creatividades: encontrar el banner con mayor CTR y el placement más relevante

El usuario accede a la página 3 y localiza el gráfico de barras "CTR medio por formato de banner", donde el formato 580×400 se identifica como el de mayor CTR de forma inmediata al ser la barra más alta. Para identificar el placement más relevante, el usuario puede recurrir tanto a la serie temporal "Evolución de ingresos por placement" como al diagrama de Sankey "Flujo de clics por banner y placement". El Sankey es especialmente eficaz para esta tarea porque muestra visualmente qué placement recibe más flujo de clics. El punto de fricción detectado es que en el Sankey aparecen los valores "Unknown" y "abc" que no corresponden a placements reales y generan ruido visual. Una mejora propuesta sería filtrar estas categorías directamente en la fuente de datos para que no aparezcan en el diagrama.

- Tarea 4: Análisis de conversiones y ventas: detectar qué campaña concentra casi todas las ventas y qué día de la semana rinde mejor

El usuario accede a la página 4 y localiza el gráfico donut "Distribución de ventas por campaña", donde el 94,8% correspondiente a la campaña 1 se identifica de forma inmediata y contundente. Para identificar el día de mayor rendimiento, el usuario recurre al gráfico de barras "Conversiones por día de la semana". El punto de fricción detectado es que los días de la semana aparecen en inglés (Friday, Saturday, Monday...) en lugar de en español, lo que puede generar confusión en usuarios no familiarizados con el idioma. Adicionalmente, el orden de los días no es el convencional de lunes a domingo, sino que está ordenado por volumen de conversiones, lo que dificulta identificar patrones semanales. Una mejora propuesta sería traducir los días al español y ofrecer la opción de ordenarlos cronológicamente.

Tareas de interacción:

- Tarea 5: Aplicar filtro por campaña y comprobar qué elementos cambian

Al seleccionar una campaña concreta en el filtro desplegable del header, todos los gráficos y scorecards de la página se actualizan correctamente mostrando únicamente los datos de esa campaña. La respuesta del sistema es inmediata y coherente. El punto de fricción detectado es que el filtro afecta a todos los elementos de la página simultáneamente pero no existe ninguna indicación visual que confirme al usuario que el filtro está activo, más allá del valor seleccionado en el desplegable. Una mejora propuesta sería añadir un indicador visual en el header que muestre claramente los filtros activos en cada momento.

- Tarea 6: Aplicar filtro por engagement y extraer un insight accionable

Al filtrar por nivel de engagement High, todos los gráficos se actualizan mostrando únicamente los datos de usuarios con alto engagement. El insight más inmediato es que los valores de ROAS y conversiones aumentan significativamente respecto a la vista global, confirmando que los usuarios de alto engagement son los más rentables. El punto de fricción detectado es el mismo que en la tarea anterior (la ausencia de un indicador visual del filtro activo) y se añade que el nombre del campo en el filtro aparece como "Engangement" con una errata ortográfica en lugar de "Engagement".

- Tarea 7: Restablecer filtros y confirmar que el dashboard vuelve al estado inicial

Al hacer clic en el botón de restablecimiento de filtros ubicado en el header, todos los filtros vuelven a su estado inicial mostrando todos los datos. La funcionalidad es correcta y la respuesta es inmediata. No se detectaron puntos de fricción en esta tarea, el botón es visible y su función es intuitiva.

- Tarea 8: Exportar una página a PDF y comprobar la legibilidad

Looker Studio permite exportar cualquier página del dashboard a PDF desde el menú de opciones. Al exportar la página 1, el resultado es un documento estático que conserva la estructura visual del dashboard con todos los gráficos y scorecards. La legibilidad es correcta aunque los gráficos interactivos pierden su funcionalidad al convertirse en imágenes estáticas, lo cual es inherente al formato PDF y debe comunicarse al usuario final. No se detectaron problemas de maquetación ni pérdida de información en la exportación.

5.7.3 Evaluación técnica del sistema

Esta sección trata medir el rendimiento y la fiabilidad del pipeline ETL y del sistema de reporting automatizado mediante métricas objetivas y cuantificables. Kimball y Ross (2013) establecen que un pipeline de datos debe evaluarse en términos de reproducibilidad, robustez ante errores y latencia de actualización, ya que estos factores determinan directamente la confianza del usuario en los datos que visualiza. La tabla recoge las ocho métricas técnicas evaluadas, su definición, el método de medición empleado y el resultado obtenido.

Tabla 5: Métricas técnicas de evaluación del pipeline ETL y sistema de reporting

Métrica	**Definición**	**Resultado**
Tiempo de ejecución del ETL	Tiempo transcurrido desde el inicio del pipeline hasta la confirmación de carga en Google Sheets	Menos de 60 segundos
Frecuencia de actualización	Periodicidad con la que se ejecuta el pipeline	Automática cada 15 minutos mediante ejecutor_periodico.py
Latencia end-to-end	Tiempo desde la ejecución del pipeline hasta que los cambios son visibles en Looker Studio	Máximo 15 minutos
Robustez ante errores	Capacidad del sistema para detectar y notificar errores durante la ejecución	Verificación post-carga implementada (notificación en terminal en caso de discrepancia)
Reproducibilidad	Capacidad de ejecutar el pipeline desde cero siguiendo la documentación del repositorio	Total, esta documentado en GitHub on README e instrucciones de instalación
Tamaño del dataset final	Número de filas y columnas del dataset procesado	15.403 filas × 23 columnas
Control de calidad	Tratamiento de nulos, duplicados y outliers documentado	0 nulos, 5 duplicados eliminados, 413 nulos imputados
Portabilidad	Opciones de despliegue disponibles	Local (VS Code + scheduler) y nube (Google Cloud Scheduler)

El tiempo de ejecución del pipeline completo (incluyendo la extracción del CSV, la aplicación de todas las transformaciones y la carga en Google Sheets) es inferior a 60 segundos, lo que lo convierte en una solución eficiente para la actualización periódica

del dashboard en un entorno de reporting real. La latencia end-to-end máxima de 15 minutos corresponde al intervalo de frescura de datos configurado en Looker Studio para fuentes de Google Sheets, tal como se describe en la sección 5.6.

La robustez ante errores está garantizada mediante la fase de verificación post-carga implementada en el pipeline, que comprueba que el número de filas cargadas en Google Sheets coincide exactamente con el del dataframe procesado y notifica al usuario en caso de discrepancia. Esta medida reduce el riesgo de que el dashboard muestre datos incompletos o corruptos sin que el operador sea consciente de ello.

La reproducibilidad del sistema es total, el repositorio público de GitHub contiene los scripts documentados y el archivo README con las instrucciones necesarias para ejecutar el pipeline desde cero, lo que garantiza que cualquier persona con acceso al repositorio pueda replicar el proceso sin necesidad de conocimiento previo del proyecto.

La frecuencia de actualización automática de 15 minutos implementada mediante ejecutor_periodico.py garantiza que el dashboard refleje siempre los datos más recientes con una latencia máxima de 15 minutos, sincronizándose con el intervalo mínimo de actualización de Looker Studio para fuentes de Google Sheets. Esta solución elimina la necesidad de intervención manual y es completamente gratuita. Para entornos que requieran mayor fiabilidad y disponibilidad continua independiente del equipo local, la opción más adecuada seguiría siendo Google Cloud Scheduler combinado con Cloud Functions, tal como se describió en la sección 4.3.2.

5.7.4 Síntesis de hallazgos y plan de mejoras

A partir de los resultados obtenidos en las tres dimensiones de evaluación descritas en las secciones anteriores, criterios basados en literatura, walkthrough por tareas y evaluación técnica, se presenta una síntesis de los principales hallazgos del sistema y un plan de mejoras priorizado según su impacto y viabilidad de implementación.

Hallazgos principales:

El walkthrough por tareas permitió identificar cinco puntos de fricción concretos.

- En primer lugar, la ausencia de un indicador visual de filtros activos en el header dificulta que el usuario sepa en todo momento qué segmentación está aplicada.
- En segundo lugar, el gráfico de coste e ingresos por campaña presenta problemas de escala cuando se muestran las tres campañas

simultáneamente, ya que los valores de la campaña 1 dominan el eje y hacen que las diferencias entre las campañas 2 y 3 sean prácticamente imperceptibles.

- En tercer lugar, el diagrama de Sankey incluye las categorías "Unknown" y "abc" que no corresponden a placements reales y generan ruido visual.
- En cuarto lugar, los días de la semana aparecen en inglés y ordenados por volumen en lugar de cronológicamente, lo que dificulta la identificación de patrones semanales.
- En quinto lugar, existe una errata ortográfica en el filtro de engagement del header, donde aparece "Engangement" en lugar de "Engagement".

Desde el punto de vista técnico, el sistema presenta un rendimiento sólido con un tiempo de ejecución del pipeline inferior a 60 segundos y una latencia end-to-end máxima de 15 minutos. La principal limitación técnica es la frecuencia de actualización manual, que en un entorno empresarial real requeriría la implementación de una solución de ejecución periódica automática.

Plan de mejoras:

La tabla recoge las cinco mejoras prioritarias identificadas a partir de los hallazgos de la evaluación, ordenadas según la combinación de impacto esperado y facilidad de implementación.

Tabla 6: Cinco mejoras prioritarias identificadas a partir de los hallazgos de la evaluación

Prioridad	Mejora	Impacto	Viabilidad
1	Corregir la errata "Engangement" por "Engagement" en el filtro del header	Alto	Inmediata
2	Traducir los días de la semana al español y ordenarlos cronológicamente	Alto	Inmediata
3	Filtrar las categorías "Unknown" y "abc" del diagrama de Sankey	Alto	Inmediata
4	Añadir indicador visual de filtros activos en el header	Alto	Media

Prioridad	Mejora	Impacto	Viabilidad
5	Implementar la ejecución periódica automática del pipeline ETL mediante Google Cloud Scheduler	Alto	Media-Alta

Las tres primeras mejoras son de implementación inmediata y no requieren cambios estructurales en el dashboard ni en el pipeline. La cuarta mejora requiere crear un campo calculado en Looker Studio que muestre el estado de los filtros activos. La quinta mejora requiere configurar una cuenta de Google Cloud y desplegar el script ETL como función serverless, tal como se describió en la sección 4.3.2

5.7.5 Validación con usuarios (SUS)

Con el objetivo de evaluar la usabilidad y eficacia del dashboard desarrollado, se aplicó el cuestionario de evaluación descrito en la sección 4.4, compuesto por el cuestionario SUS (System Usability Scale), tareas de interpretación y preguntas de valoración general. La evaluación se realizó con una muestra de 8 participantes con perfil de negocio y marketing, sin conocimientos técnicos avanzados en ciencia de datos, que representan el usuario objetivo del sistema (Anexo D).

Resultados del cuestionario SUS:

El cuestionario SUS está compuesto de 10 ítems con escala Likert de 1 a 5, cuya puntuación se calcula mediante la fórmula estándar propuesta por Brooke (1996): para los ítems impares se resta 1 al valor dado, para los ítems pares se resta el valor dado a 5, y la suma de los 10 resultados se multiplica por 2,5, obteniendo una puntuación final entre 0 y 100. La tabla recoge las puntuaciones individuales obtenidas por cada participante.

Tabla 7: Resultados del cuestionario SUS

Participante	Puntuación SUS
P1	100,0
P2	75,0
P3	100,0

Participante	Puntuación SUS
P4	80,0
P5	90,0
P6	95,0
P7	92,5
P8	100,0
Media	**91,6**

La puntuación media obtenida es de 91,6 sobre 100, con un mínimo de 75,0 y un máximo de 100,0. Según la escala de interpretación propuesta por Bangor et al. (2009), una puntuación superior a 85 se considera excelente y equivale a una calificación de "Best Imaginable" en la escala de adjetivos. La puntuación de 91,6 supera ampliamente el umbral de 68 considerado como buena usabilidad y se sitúa en el rango de excelencia, lo que indica que el sistema es percibido como altamente usable por los participantes.

Resultados de las tareas de interpretación:

Las cinco tareas de interpretación evaluaron la capacidad del dashboard para comunicar información de negocio de forma efectiva, midiendo la tasa de acierto de cada participante al responder preguntas concretas sobre los datos visualizados. La tabla recoge los resultados obtenidos.

Tabla 8: Resultados de las tareas de interpretación

Tarea	Pregunta	Respuesta correcta	Aciertos	Tasa de acierto
T1	¿Qué campaña generó más ingresos?	Camp 1	5/8	62,5%
T2	¿Qué nivel de engagement concentra más clics?	High	7/8	87,5%

Tarea	Pregunta	Respuesta correcta	Aciertos	Tasa de acierto
T3	¿En qué mes fue mayor la actividad publicitaria?	Abril	8/8	100,0%
T4	¿Qué formato de banner tiene el CTR más alto?	580×400	7/8	87,5%
T5	¿Qué placement genera más conversiones?	jkl	6/8	75,0%

La tasa de acierto media global fue del 82,5%, lo que indica que el dashboard comunica la información de negocio de forma efectiva para la mayoría de los usuarios. La tarea 3 tiene un 100% de aciertos, lo que confirma que la serie temporal es el elemento más legible del dashboard. La tarea 1 obtuvo la tasa de acierto más baja (62,5%), varios participantes respondieron el nombre de la campaña en formatos diferentes ("campaña 1", "camp 1", "Camp 1") que fueron contabilizados como correctos, aunque uno de los participantes respondió erróneamente identificando el nivel de engagement en lugar de la campaña.

Valoración general:

Las tres preguntas de valoración general evaluaron el diseño visual, la claridad de la información y la utilidad de los filtros interactivos en una escala del 1 al 5. La tabla recoge los resultados.

Tabla 9: Valoración general

Dimensión	Media	Mínimo	Máximo
Diseño visual	4,62	4	5
Claridad de la información	4,38	4	5
Utilidad de los filtros	4,75	4	5

Los filtros interactivos obtuvieron la valoración más alta (4,75), lo que confirma que la interactividad del dashboard es acogida como uno de sus elementos más valiosos por los usuarios.

Análisis cualitativo:

Los comentarios cualitativos de los participantes refuerzan los resultados cuantitativos. Entre los aspectos más valorados destacan la facilidad de interpretación de los datos, la claridad de la estructura visual y la paleta de colores. En cuanto a las áreas de mejora, los comentarios más frecuentes fueron el deseo de aumentar el tamaño de los números en los scorecards, la inclusión de explicaciones o tooltips sobre las variables utilizadas y la posibilidad de reducir el número de gráficos por página para simplificar la visualización.

Conclusión de la evaluación:

Los resultados de la evaluación confirman que el sistema de visualización desarrollado cumple con los objetivos de usabilidad y eficacia establecidos en el objetivo 1 del trabajo. La puntuación SUS de 91,6 evidencia una usabilidad excelente, la tasa de acierto del 82,5% en las tareas de interpretación demuestra que el dashboard comunica la información de negocio de forma efectiva, y las valoraciones cualitativas identifican áreas de mejora concretas que podrían abordarse en iteraciones futuras del sistema.

5.7.6 Resultados e insights obtenidos

El análisis del conjunto de datos de publicidad digital a través del sistema de visualización desarrollado ha permitido identificar una serie de descubrimientos importantes para la toma de decisiones en entornos de marketing digital. Los insights presentados a continuación emergen directamente de las visualizaciones del dashboard y constituyen el valor analítico del sistema desarrollado.

Rendimiento global de las campañas:

Durante el segundo trimestre de 2020, las tres campañas analizadas generaron conjuntamente 2.492.825 clics, con un coste total de 175.192,22 USD y unos ingresos de 276.262,62 USD, lo que representa un ROAS medio de 1,92. Este valor nos dice que por cada euro invertido en publicidad se generaron 1,92 euros de ingreso, situándose por debajo del objetivo de referencia de ROAS=2 establecido. Este resultado sugiere que existe margen de mejora en la rentabilidad global de las campañas, especialmente en aquellas que presentan un ROAS inferior al objetivo.

Concentración del rendimiento en la campaña 1:

El hallazgo más relevante del análisis es la elevada concentración del rendimiento en la campaña 1. Esta campaña concentra el 56,5% de los clics totales, genera ingresos muy superiores a su coste; siendo la única de las tres que supera el objetivo de ROAS=2 con un valor de 2,1; y acapara el 94,8% de las ventas totales registradas durante el trimestre. La campaña 2 y la campaña 3 presentan rendimientos significativamente inferiores en todas las métricas, con ROAS de 1,8 y 1,6 respectivamente, ambos por debajo del objetivo. Desde el punto de vista de la toma de decisiones, este hallazgo sugiere que una reasignación del presupuesto hacia la campaña 1 podría incrementar el retorno global de la inversión publicitaria.

Impacto del nivel de engagement en el rendimiento:

El nivel de engagement del usuario se revela como un factor determinante del rendimiento de las campañas. Los usuarios clasificados como High engagement concentran el 52,8% de los clics totales y generan consistentemente un ROAS superior en todas las campañas respecto a los niveles Medium y Low. Este hallazgo es coherente con la literatura sobre segmentación en marketing digital, que establece que los usuarios más comprometidos con el contenido publicitario presentan tasas de conversión más elevadas (Farris et al., 2010). La implicación práctica es que las estrategias de segmentación orientadas a usuarios de alto engagement pueden mejorar la eficiencia del gasto publicitario sin necesidad de aumentar el presupuesto total.

Rendimiento por `formato de banner`:

El análisis de creatividades reveló diferencias significativas en el CTR según el formato del banner. El formato 580×400 obtuvo el CTR más elevado (1,27%), más del doble que la media global (0,81%), mientras que el formato 468×60 presentó el valor más bajo (0,25%). Adicionalmente, el formato 240×400 concentra el 41,4% de las conversiones totales, lo que lo posiciona como el formato más eficiente en términos de generación de acción por parte del usuario. Estos resultados sugieren que la selección del formato publicitario tiene un impacto directo y medible en la efectividad de las campañas, y que una estrategia orientada a priorizar los formatos 580×400 y 240×400 podría mejorar tanto el CTR como las conversiones.

Rendimiento por `placement`:

El diagrama de Sankey del dashboard evidenció que el placement ghi concentra el mayor volumen de clics procedentes de prácticamente todos los formatos de banner,

seguido del placement mno. El gráfico de conversiones por placement confirmó esta tendencia, con ghi generando el 51,8% de las conversiones totales. El placement abc presentó el rendimiento más bajo en todas las métricas analizadas. Desde el punto de vista de la optimización del presupuesto, estos resultados sugieren que concentrar la inversión en los placements ghi y mno podría maximizar el retorno publicitario.

Evolución temporal y tendencia decreciente:

La serie temporal de clics e ingresos mostró una tendencia decreciente a lo largo del trimestre, con una actividad publicitaria significativamente mayor en abril que fue reduciéndose progresivamente hacia junio. Esta caída podría responder a factores estacionales, a una reducción progresiva del presupuesto publicitario o a una saturación de la audiencia objetivo a lo largo del trimestre. Sin datos adicionales de trimestres anteriores o posteriores no es posible determinar si esta tendencia es estructural o puntual, lo que refuerza la limitación temporal del dataset señalada en la sección 5.1.

Patrones temporales por día de la semana:

El análisis de conversiones por día de la semana reveló que los lunes y martes concentran el mayor volumen de conversiones, mientras que los fines de semana presentan valores más bajos. Este patrón es coherente con el comportamiento típico de los usuarios en entornos de marketing B2B, donde la actividad comercial se concentra en los días laborables (Farris et al., 2010). Este hallazgo tiene implicaciones directas para la programación de las campañas publicitarias, concentrar las impresiones en los primeros días de la semana podría incrementar la tasa de conversión sin aumentar el coste total.

Correlación entre alcance y conversiones:

El gráfico de dispersión entre impresiones y conversiones mostró una correlación positiva clara entre ambas variables, a mayor número de impresiones, mayor número de conversiones. Este resultado, aunque esperado desde un punto de vista teórico, confirma que el modelo de atribución de las campañas analizadas responde de forma coherente al volumen de inversión en alcance publicitario, validando la calidad y consistencia del dataset utilizado.

6. Objetivos de desarrollo sostenible

Los Objetivos de Desarrollo Sostenible (ODS) fueron adoptados por la Organización de las Naciones Unidas en 2015 como parte de la Agenda 2030 para el Desarrollo Sostenible, constituyendo un marco global de 17 objetivos orientados a erradicar la pobreza, proteger el planeta y garantizar la prosperidad para todas las personas (ONU, 2015). Este TFG contribuye directamente a tres de estos objetivos, tal como se describe a continuación.

ODS 8: Trabajo decente y crecimiento económico

El ODS 8 promueve el crecimiento económico sostenido, inclusivo y sostenible, el empleo pleno y productivo y el trabajo decente para todas las personas. El sistema de visualización interactiva y automatización de reporting de este trabajo es beneficioso para este objetivo ya que mejora la eficiencia de los procesos de análisis de datos en las organizaciones. La automatización del pipeline ETL elimina tareas repetitivas y propensas a errores como la elaboración manual de informes en hojas de cálculo, liberando tiempo de los profesionales para dedicarlo a tareas de mayor valor añadido como el análisis estratégico y la toma de decisiones. Aparte, los dashboards interactivos desarrollados permiten que perfiles no técnicos accedan a información analítica compleja de forma autónoma, mejorando la productividad organizacional y reduciendo la dependencia de especialistas en datos para la generación de informes periódicos.

ODS 9: Industria, innovación e infraestructura

El ODS 9 promueve la construcción de infraestructuras resilientes, la promoción de la industrialización inclusiva y sostenible y el fomento de la innovación. Este trabajo contribuye a este objetivo mediante el desarrollo de una solución de inteligencia empresarial basada en herramientas de código abierto y plataformas gratuitas en la nube, que demuestra cómo la innovación en los procesos de análisis de datos puede implementarse sin necesidad de grandes inversiones en infraestructura tecnológica. La integración de Python, Google Sheets, Looker Studio y Power BI en un pipeline automatizado constituye una propuesta de arquitectura tecnológica replicable y escalable que puede servir de referencia para organizaciones que busquen modernizar sus procesos de reporting sin incurrir a grandes costes de licencias o infraestructura. La publicación del código en forma de scripts documentados facilita además la transferencia de conocimiento y la adopción de estas prácticas por parte de otros profesionales y organizaciones.

ODS 10: Reducción de las desigualdades

El ODS 10 busca reducir la desigualdad en y entre los países, promoviendo la inclusión económica y social. Este trabajo contribuye al objetivo ya que demuestra que herramientas de visualización de datos de alto valor analítico pueden implementarse de forma completamente gratuita mediante plataformas como Looker Studio, eliminando las barreras económicas que tradicionalmente han limitado el acceso a soluciones de Business Intelligence a grandes organizaciones con capacidad de invertir en licencias costosas como Power BI Premium. Este aumento de la facilidad al acceso al análisis de datos permite que pequeñas y medianas empresas, organizaciones sin ánimo de lucro y entidades públicas con recursos limitados puedan tomar decisiones basadas en evidencia con las mismas capacidades analíticas que las grandes corporaciones, contribuyendo a reducir la brecha digital y las desigualdades en el acceso a la información estratégica.

7. Conclusiones

En este TFG se ha abordado el diseño, desarrollo y evaluación de un sistema de visualización interactiva y storytelling digital aplicado a un caso de uso real de publicidad digital, combinando un marco teórico fundamentado en la literatura académica con una implementación práctica basada en herramientas accesibles y ampliamente utilizadas en el sector.

Conclusiones sobre los insights obtenidos:

Más allá de los objetivos metodológicos, el análisis del dataset de publicidad digital permitió identificar hallazgos de valor para la toma de decisiones. La campaña 1 concentra el 94,8% de las ventas totales y es la única que supera el objetivo de ROAS=2, lo que sugiere una clara oportunidad de reasignación presupuestaria. El nivel de engagement High genera consistentemente un retorno superior en todas las campañas, validando la importancia de la segmentación por comportamiento del usuario. El formato de banner 580×400 presenta el CTR más elevado y el placement ghi contiene el mayor volumen de conversiones, identificando las combinaciones creatividad con emplazamiento más eficientes del conjunto analizado.

Conclusiones respecto a los objetivos planteados:

El objetivo general del trabajo era diseñar y evaluar un sistema de visualización interactiva para la mejora de la comunicación de resultados y la toma de decisiones en entornos de negocio y marketing. Este ha sido alcanzado satisfactoriamente. El sistema desarrollado integra un pipeline ETL automatizado en Python, dos dashboards

interactivos en Looker Studio y Power BI, y una evaluación de usabilidad mediante el cuestionario SUS, demostrando que es posible construir una solución de inteligencia empresarial completa y funcional utilizando exclusivamente herramientas gratuitas y de código abierto.

En cuanto a los objetivos específicos, los resultados nos dejan sacar varias conclusiones:

El O1 se cumple. La revisión bibliográfica identificó los principios que guían el diseño del sistema (jerarquía visual, uso semántico del color, narrativa estructurada y el modelo de interacción de Shneiderman) y los tradujo en criterios de diseño concretos aplicados a cada página del dashboard. La comparativa entre Looker Studio y Power BI reveló diferencias relevantes en coste, automatización y facilidad de uso. Looker Studio ganó por su gratuidad, su integración nativa con Google Sheets y su actualización en tiempo real sin coste adicional. Power BI tiene más potencia de modelado y mejores capacidades de gobernanza del dato, pero requiere licencia Pro para desplegarlo y compartirlo correctamente, lo que lo hace menos accesible para organizaciones sin infraestructura Microsoft previa.

El O2 produjo un sistema funcional de principio a fin. Dos dashboards interactivos (uno en Looker Studio con cuatro páginas y otro en Power BI con tres), un pipeline ETL en Python que limpia y transforma el dataset, calcula seis métricas derivadas estándar del sector y carga los datos automáticamente en Google Sheets, y un scheduler que ejecuta todo el proceso cada 15 minutos sin que nadie tenga que hacer nada. La arquitectura completa funciona con herramientas gratuitas y está documentada en un repositorio público de GitHub para que cualquiera pueda replicarla.

El O3 confirma que el sistema funciona para quien lo usa. El walkthrough por tareas detectó cinco fricciones concretas (problemas de escala en el gráfico de campañas, categorías no válidas en el Sankey, días de la semana desordenados y una errata en el filtro de engagement), todas con mejoras propuestas y priorizadas. Las métricas técnicas muestran un pipeline robusto: tiempo de ejecución inferior a 60 segundos, latencia máxima de 15 minutos y verificación post-carga implementada. Los 8 participantes del cuestionario SUS dieron una puntuación media de 91,6 sobre 100, rango de excelencia según Bangor et al. (2009), con una tasa de acierto del 82,5% en las tareas de interpretación del dashboard.

Limitaciones del trabajo:

El trabajo presenta dos limitaciones principales que deben tenerse en cuenta al interpretar los resultados. En primer lugar, como ya se mencionó en el apartado 5.1, el dataset utilizado abarca únicamente tres meses de actividad publicitaria, lo que impide el análisis de tendencias estacionales anuales y limita la generalización de los insights obtenidos. Sin datos de periodos anteriores o posteriores no es posible determinar si las tendencias observadas son estructurales o puntuales.

Por otro lado, la evaluación realizada con participantes está compuesta por 8 participantes, lo que, aunque suficiente para un estudio exploratorio en el contexto de un TFG, no permite extraer conclusiones estadísticamente representativas sobre la usabilidad del sistema en poblaciones más amplias. Como ya se indicó en el apartado 5.7, para que dicha evaluación aporte resultados generalizables sería necesario ampliar el tamaño muestral e incorporar una mayor diversidad de perfiles de usuario en términos de sector, experiencia con herramientas de Business Intelligence y características demográficas.

Líneas de trabajo futuro:

Como continuación de este trabajo, la ampliación del dataset con datos de periodos más largos permitiría identificar patrones estacionales y tendencias a largo plazo que enriquecerían el análisis. Contener datos de al menos un año completo permitiría detectar variaciones estacionales relevantes para la planificación de campañas publicitarias, algo que el dataset actual de tres meses no permite.

La incorporación de fuentes de datos adicionales como Google Analytics o Meta Ads a través de sus APIs oficiales permitiría construir un dashboard multicanal más completo y representativo de la realidad empresarial.

La implementación del pipeline ETL en Google Cloud Scheduler permitiría automatizar completamente la actualización del dashboard sin depender de la ejecución manual del script.

Finalmente, una evaluación más amplia con un mayor número de participantes y perfiles más diversos permitiría obtener conclusiones Para que dicha validación sea generalizada, sería necesario diseñar un protocolo de evaluación que garantice la representatividad de la muestra en varias dimensiones:

- En cuanto al tamaño muestral, se recomienda contar con un mínimo de 30 participantes para poder aplicar análisis estadísticos con cierta robustez.

- En cuanto a la diversidad de perfiles, la muestra debería incluir participantes de distintos rangos de edad, géneros, departamentos y sectores de actividad, así como diferentes niveles de experiencia previa con herramientas de Business Intelligence, desde usuarios sin experiencia hasta usuarios avanzados. Esta diversidad permitiría detectar si el sistema presenta fricciones específicas para determinados grupos de usuarios y obtener recomendaciones de mejora más transferibles a contextos profesionales reales.

El protocolo de evaluación combinaría el cuestionario SUS con tareas guiadas de interpretación, entrevistas cualitativas posteriores y, opcionalmente, métricas de eye-tracking para analizar los patrones de atención visual sobre el dashboard.

8. Referencias

Albers, D., Correll, M., y Gleicher, M. (2014). Task-driven evaluation of aggregation in time series visualization. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 551–560. https://doi.org/10.1145/2556288.2557200

Bangor, A., Kortum, P., y Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. Journal of Usability Studies, 4(3), 114–123.

Brooke, J. (1996). SUS: A quick and dirty usability scale. En P. W. Jordan, B. Thomas, B. A. Weerdmeester y I. L. McClelland (Eds.), Usability evaluation in industry (pp. 189–194). Taylor & Francis.

Cesar, M. L. J. (2020). Sistema de inteligencia de negocios y la automatización de reportes en la empresa ADECCO [Tesis de grado, Universidad Nacional José Faustino Sánchez Carrión]. Repositorio institucional. https://repositorio.unjfsc.edu.pe/handle/20.500.14067/4175

Croll, A., y Yoskovitz, B. (2013). Lean analytics: Use data to build a better startup faster. O'Reilly Media.

Deloitte. (2022, 6 de mayo). The five benefits of data visualization. https://www.deloitte.com/nl/en/services/tax/perspectives/bps-the-five-benefits-of-data-visualization.html

Doerr, J. (2018). Measure what matters: How Google, Bono, and the Gates Foundation rock the world with OKRs. Portfolio/Penguin.

Farris, P. W., Bendle, N. T., Pfeifer, P. E., y Reibstein, D. J. (2010). Marketing metrics: The definitive guide to measuring marketing performance (2.ª ed.). Pearson Education.

Few, S. (2006). Information dashboard design: The effective visual communication of data. O'Reilly Media.

García, S., Luengo, J., y Herrera, F. (2015). Data preprocessing in data mining. Springer. https://doi.org/10.1007/978-3-319-10247-4

Gartner. (2023). Magic Quadrant for Analytics and Business Intelligence Platforms. Gartner Research.

Google. (s.f.-a). Looker Studio documentation. Google Cloud. https://docs.cloud.google.com/looker/docs/studio/new-studio-docs?hl=es

Google. (s.f.-b). Looker Studio — compartir y publicar informes. Google. https://support.google.com/looker-studio/answer/6313851

Google Cloud. (s.f.). Cloud Scheduler documentation. Google. https://cloud.google.com/scheduler/docs

Heer, J., y Shneiderman, B. (2012). Interactive dynamics for visual analysis. Queue, 10(2), 30–55. https://doi.org/10.1145/2133416.2146416

Hullman, J., y Diakopoulos, N. (2011). Visualization rhetoric: Framing effects in narrative visualization. IEEE Transactions on Visualization and Computer Graphics, 17(12), 2231–2240. https://doi.org/10.1109/TVCG.2011.255

Kaplan, R. S., y Norton, D. P. (1992). The balanced scorecard: Measures that drive performance. Harvard Business Review, 70(1), 71–79.

Kimball, R., y Ross, M. (2013). The data warehouse toolkit: The definitive guide to dimensional modeling (3.ª ed.). Wiley.

Kim, Y., Reinecke, K., y Hullman, J. (2017). Explaining the gap: Visualizing the unseen and uncertain. Proceedings of the ACM on Human-Computer Interaction, 1(CSCW), 1–19. https://doi.org/10.1145/3025453.3025592

Koffka, K. (1935). Principles of Gestalt psychology. Harcourt, Brace and Company.

Microsoft. (s.f.-a). Power BI guidance documentation. Microsoft Learn. https://learn.microsoft.com/es-es/power-bi/guidance/

Microsoft. (s.f.-b). Power BI pricing. Microsoft. https://powerbi.microsoft.com/es-es/pricing/

Microsoft. (s.f.-c). Power Query documentation. Microsoft Learn. https://learn.microsoft.com/es-es/power-query/

Microsoft. (s.f.-d). DAX reference. Microsoft Learn. https://learn.microsoft.com/es-es/dax/

Microsoft. (s.f.-e). Power BI Desktop documentation. Microsoft Learn. https://learn.microsoft.com/es-es/power-bi/fundamentals/desktop-what-is-desktop

Miya, T. K., y Govender, I. (2022). UX/UI design of online learning platforms and their impact on learning: A review. International Journal of Research in Business and Social Science, 11(10), 316–327. https://doi.org/10.20525/ijrbs.v11i10.2236

Munzner, T. (2014). Visualization analysis and design. A K Peters/CRC Press. https://doi.org/10.1201/b17511

Organización de las Naciones Unidas. (2015). Transformar nuestro mundo: la Agenda 2030 para el Desarrollo Sostenible. ONU. https://www.un.org/sustainabledevelopment/es/2015/09/la-asamblea-general-adopta-la-agenda-2030-para-el-desarrollo-sostenible/

Parmenter, D. (2015). Key performance indicators: Developing, implementing, and using winning KPIs (3.ª ed.). Wiley.

Ruddhan, N. (2020). Online Advertising Digital Marketing Data [Conjunto de datos]. Kaggle. https://www.kaggle.com/datasets/naniruddhan/online-advertising-digital-marketing-data

Segel, E., y Heer, J. (2010). Narrative visualization: Telling stories with data. IEEE Transactions on Visualization and Computer Graphics, 16(6), 1139–1148. https://doi.org/10.1109/TVCG.2010.179

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. Proceedings of the IEEE Symposium on Visual Languages, 336–343. https://doi.org/10.1109/VL.1996.545307

Tukey, J. W. (1977). Exploratory data analysis. Addison-Wesley.

Vande Moere, A., Tominski, C., y van Wijk, J. J. (2010). Declarative language design for interactive visualization. IEEE Computer Graphics and Applications, 30(5), 37–45. https://doi.org/10.1109/MCG.2010.91

Anthropic. (2025). Claude (claude-sonnet-4-6) [Modelo de lenguaje de gran escala]. https://www.anthropic.com

Nielsen, J. (1994). Usability engineering. Morgan Kaufmann.

Microsoft. (s.f.-f). Visual Studio Code documentation. Microsoft. https://code.visualstudio.com/docs

Burnashev, A. (2024). gspread: Google Spreadsheets Python API (versión 6.2.1) [Software]. https://github.com/burnash/gspread

Bader, D. (2024). schedule: Job scheduling for humans (versión 1.2.2) [Software]. https://github.com/dbader/schedule

Google Cloud Platform. (2024). google-auth-oauthlib: Google Authentication Library (versión 1.3.1) [Software]. https://github.com/googleapis/google-cloud-python/tree/main/packages/google-auth-oauthlib

GitHub. (s.f.). GitHub documentation. GitHub, Inc. https://docs.github.com

McKinney, W., y el equipo de desarrollo de pandas. (2024). pandas: Powerful data structures for data analysis, time series, and statistics (versión 2.3.3) [Software]. https://pandas.pydata.org

9. Acrónimos

Tabla 10: Acrónimos

Acrónimo	Significado
API	Application Programming Interface (Interfaz de Programación de Aplicaciones)
BI	Business Intelligence (Inteligencia Empresarial)
B2B	Business to Business (Negocio a Negocio)
CPC	Cost Per Click (Coste por Clic)
CPM	Cost Per Mille (Coste por Mil Impresiones)
CRM	Customer Relationship Management (Gestión de Relaciones con Clientes)
CTR	Click-Through Rate (Tasa de Clics)
CSV	Comma-Separated Values (Valores Separados por Comas)
DAX	Data Analysis Expressions (Expresiones de Análisis de Datos)
ETL	Extract, Transform, Load (Extraer, Transformar, Cargar)
IQR	Interquartile Range (Rango Intercuartílico)
KPI	Key Performance Indicator (Indicador Clave de Rendimiento)
ODS	Objetivos de Desarrollo Sostenible
OKR	Objectives and Key Results (Objetivos y Resultados Clave)
ONU	Organización de las Naciones Unidas
ROAS	Return On Ad Spend (Retorno sobre el Gasto Publicitario)

Acrónimo	Significado
ROI	Return On Investment (Retorno sobre la Inversión)
SaaS	Software as a Service (Software como Servicio)
SQL	Structured Query Language (Lenguaje de Consulta Estructurado)
SUS	System Usability Scale (Escala de Usabilidad del Sistema)
TFG	Trabajo de Fin de Grado
UX	User Experience (Experiencia de Usuario)
UI	User Interface (Interfaz de Usuario)
WCAG	Web Content Accessibility Guidelines (Pautas de Accesibilidad para el Contenido Web)
URL	Uniform Resource Locator (Localizador Uniforme de Recursos)

10. Anexos

Anexo A: Repositorio de código en GitHub

Se recoge el repositorio público de GitHub que contiene todos los scripts desarrollados para la parte práctica del TFG: el notebook de exploración y análisis de calidad, el pipeline ETL con autenticación OAuth, y el scheduler de ejecución periódica automática. El repositorio incluye además el archivo README.md con las instrucciones completas de instalación y ejecución.

El repositorio puede consultarse en el siguiente enlace:

https://github.com/luisballtam335/tfg-visualizacion-datos-marketing

Anexo B: Dashboard interactivo en Looker Studio

Se recoge el dashboard interactivo desarrollado en Looker Studio, estructurado en cuatro páginas: Visión general, Rendimiento por campaña, Análisis de creatividades y Análisis de conversiones y ventas, cuyo diseño y desarrollo se describe en la sección 5.4 del presente trabajo.

El dashboard puede consultarse en el siguiente enlace:

https://lookerstudio.google.com/reporting/15ebed93-556a-48ab-9f1c-10c25111f182

Anexo C: Dashboard en Power BI

Se recoge el archivo .pbix del dashboard desarrollado en Power BI Desktop, estructurado en tres páginas: Visión general, Rendimiento por campaña y Análisis de creatividades, cuyo diseño y desarrollo se describe en la sección 5.5 del presente trabajo.

El archivo puede descargarse directamente desde el repositorio de GitHub del proyecto, el enlace corresponde al del Anexo A.

Anexo D: Formulario de evaluación SUS y resultados

Se recoge el formulario de evaluación de usabilidad aplicado a los 8 participantes del estudio, compuesto por el cuestionario SUS, tareas de interpretación del dashboard y preguntas de valoración general, cuyos resultados se describen en la sección 5.7 del presente trabajo.

El formulario y sus resultados pueden consultarse en el siguiente enlace:

https://docs.google.com/forms/d/e/1FAIpQLSe57y-K480YJ5jPXl7-OUnY6hGIXLJNeMtws480toLjq_WNqQ/viewform?usp=dialog

Anexo E: Declaración de uso de Inteligencia Artificial:

En la elaboración del presente TFG se han utilizado herramientas de

Inteligencia Artificial (IA) como apoyo en distintas fases del trabajo académico.

- Redacción académica:

Herramienta: Claude (Anthropic), plan Pro

Funciones: mejora de la redacción académica y adaptación del texto a un

registro formal

Prompts utilizados: “Revisa el siguiente texto y mejora su redacción para

adaptarlo a un registro académico formal, manteniendo el contenido y las ideas

originales: [texto]”

- Gestión de referencias:

Herramienta: Claude (Anthropic), plan Pro

Funciones: formateo y redacción de referencias bibliográficas

Prompts utilizados: “Formatea la siguiente referencia según las normas APA 7ª

edición: [referencia]”

- Desarrollo de código:

Herramienta: Claude (Anthropic), plan Pro

Funciones: apoyo puntual en la depuración de errores, adaptación de fragmentos de código específicos y orientación sobre librerías y métodos de Python

Prompts utilizados: "El siguiente script da este error: [error]. Identifica la causa y propón una solución: [código]"

DECLARO QUE:

- El contenido final presentado es resultado de mi trabajo personal y académico.
- He revisado y validado críticamente toda la información obtenida mediante herramientas de IA.

- El uso de IA no ha sustituido el análisis, la reflexión ni las conclusions propias exigidas en el TFG.
- El trabajo cumple las normas de integridad académica y propiedad intellectual establecidas por la universidad.

Nombre del estudiante: Luis Ballester Tamarit

Firma:

BALLESTER TAMARIT LUIS - 23324956N

Firmado digitalmente por BALLESTER TAMARIT LUIS - 23324956N
Fecha: 2026.05.18 09:04:48 +02'00'