



Proyecto Fin de Semestre

Objetivo:

El propósito de este proyecto es evolucionar de la creación de reportes estáticos al desarrollo de herramientas interactivas de decisión. Los estudiantes deberán aplicar los principios de Storytelling with Data (STWD) aprendidos durante el semestre, implementándolos técnicamente mediante Python y Streamlit.

El reto consiste en procesar un conjunto de datos y transformarlo en un Dashboard Interactivo que permita a un usuario de negocio explorar la información, filtrar variables relevantes y obtener conclusiones estratégicas de manera autónoma.

Indicaciones:

- **Grupos:** el proyecto se realizará en equipos de 3 estudiantes (**obligatorio**).
- **Herramientas:** Python (VSCode), librerías de visualización (se recomienda Plotly Express o Altair) y Framework Web (Streamlit).
- **Entregables:**
 - **Código fuente y de respaldo:** script principal (app.py), el archivo de respaldo (Jupyter Notebook o Excel), y el dataset final a usar (.csv) en caso de que hayan hecho alguna limpieza de datos.
 - **Video:** un video entre 5 a 10 minutos donde el equipo (grupo) navegue por la aplicación, explicando el flujo narrativo y los hallazgos principales, simulando una presentación a la gerencia.
 - **URL de la App (puntos extra):** enlace a la aplicación desplegada en Streamlit Cloud (opcional).
- **Heteroevaluación:** se aplicará una encuesta de **heteroevaluación** al finalizar, donde evaluarán la contribución de cada integrante. Esta evaluación influirá en la nota individual del proyecto.

Requisitos Técnicos y de Diseño

Para que el proyecto sea evaluado, la aplicación debe cumplir con los siguientes estándares técnicos y de diseño:

1. Estructura de la aplicación:
 - Configuración de página completa (`st.set_page_config` con `layout="wide"`).
 - Uso de barra lateral (`st.sidebar`) para ubicar los filtros globales, manteniendo el área principal limpia para los gráficos.
 - Uso de métricas (`st.metric`) en la parte superior para destacar los indicadores clave (KPIs).
2. Visualización de datos:
 - Implementación de al menos 6 tipos de gráficos interactivos distintos.
 - Personalización de diseño: no utilizar los colores por defecto. Se debe definir y aplicar una paleta de colores corporativa o semántica en base a lo aprendido en STWD.
 - Limpieza visual: eliminación de elementos que no aporten valor.
3. Narrativa y contexto:
 - Los gráficos deben tener títulos conclusivos (que indiquen el hallazgo) en lugar de descriptivos.
 - Uso de componentes de texto (`st.markdown`, `st.info`) para proporcionar contexto, explicaciones y conclusiones escritas que guíen al usuario.



Selección del Caso de Negocio

Cada equipo deberá seleccionar **uno** de los siguientes tres casos para desarrollar su proyecto:

Opción A: Análisis de Rentabilidad y Operaciones (Caso Superstore)

Contexto:

Superstore Giant es un minorista con operaciones en todo el país. Aunque los volúmenes de venta son altos, la Dirección Financiera ha detectado que los márgenes de ganancia se están reduciendo en regiones específicas. Existe la preocupación de que la política de descuentos actual esté afectando negativamente la rentabilidad final, a pesar de generar mayores ingresos brutos.

Datos a tomar en cuenta:

- Sales y Profit: indicadores principales de desempeño.
- Discount: porcentaje de descuento aplicado a la venta.
- Category y Sub-Category: clasificación de productos.
- Region, State, City: ubicación geográfica de la transacción.

El reto del análisis

El dashboard debe responder: ¿en qué áreas y productos la estrategia de descuentos está comprometiendo la rentabilidad?

- Analizar la relación directa entre el nivel de descuento y la utilidad generada.
- Identificar qué categorías o productos específicos tienen un alto volumen de ventas pero generan pérdidas operativas.
- Evaluar la eficiencia de la estrategia de precios en las distintas categorías.

Fuente: Vivek468. (s.f.). Superstore dataset [Conjunto de datos]. Kaggle.

<https://www.kaggle.com/datasets/vivek468/superstore-dataset-final>

Opción B: Análisis de Riesgo y Retención de Clientes (Caso de Banco)

Contexto:

Una institución financiera está experimentando una tasa elevada de cancelación de tarjetas de crédito (Churn). Existen dos hipótesis internas sobre la causa:

- **Hipótesis demográfica:** los clientes que se van son aquellos de menores ingresos o de ciertos rangos de edad.
- **Hipótesis de comportamiento:** los clientes que se van son aquellos que dejan de utilizar el producto gradualmente (inactividad). La gerencia necesita claridad para diseñar su próxima campaña de retención.

Datos a tomar en cuenta:

- Attrition_Flag: estado del cliente (Cliente Actual vs. Cliente que canceló).
- Total_Trans_Amt y Total_Trans_Ct: monto total y cantidad de transacciones realizadas.



Herramientas y Visualización

Proyecto Fin de Semestre

- **Credit_Limit**: límite de crédito asignado.
- **Income_Category**: rango de ingresos del cliente.
- **Months_Inactive**: meses de inactividad registrados.

El reto del análisis

El dashboard debe responder: ¿qué factores (demográficos o de comportamiento) predicen mejor la salida de un cliente?

- Validar si el nivel de ingresos influye realmente en la decisión de cancelar la tarjeta.
- Identificar patrones de uso (frecuencia y monto) que distinguen a los clientes leales de los que cancelan.
- Investigar sobre el comportamiento transaccional y la inactividad reciente como indicadores clave de riesgo, por encima del perfil demográfico.

Fuente: Sakshigoyal7. (s.f.). Credit card customers [Conjunto de datos]. Kaggle.

<https://www.kaggle.com/datasets/sakshigoyal7/credit-card-customers>

Opción C: Inteligencia de Marketing y Segmentación (Caso de Campañas)

Contexto:

El departamento de Marketing ha ejecutado 5 campañas publicitarias distintas con resultados mixtos. Para la planificación del próximo año, necesitan optimizar el presupuesto enfocándose únicamente en los segmentos de clientes más rentables. El objetivo es identificar el perfil ideal para promocionar productos de alto valor (Vinos y Productos exclusivos).

Datos a tomar en cuenta:

- **Response**: indicador de si el cliente aceptó la oferta en la última campaña.
- **MntWines, MntGoldProds**: gasto histórico en categorías específicas.
- **NumWebPurchases vs NumStorePurchases**: preferencia del canal de compra.
- **Education, Marital_Status**: perfil demográfico.
- **Kidhome, Teenhome**: composición familiar (número de niños/adolescentes).

El reto del análisis

El dashboard debe responder: ¿cuál es el perfil del cliente de alto valor y cuál es el canal más efectivo para contactarlo?

- Analizar cómo la composición familiar (hijos en el hogar) afecta el consumo de productos de alto valor.
- Determinar si los clientes de mayores ingresos prefieren comprar a través de la web o en tiendas físicas.
- Investigar sobre la influencia de la estructura familiar en el comportamiento de compra de productos de lujo y la efectividad de los canales.

Fuente: Rodsaldanha. (s.f.). Marketing campaign [Conjunto de datos]. Kaggle.

<https://www.kaggle.com/datasets/rodsaldanha/arketing-campaign>



Rúbrica de Calificación

Nota Total: 100 puntos (equivale al 25% de la nota final del curso)

Esta rúbrica está diseñada para evaluar su capacidad de aplicar los conceptos de análisis, visualización y narrativa para contar una historia con datos de manera efectiva y persuasiva.

Criterio de Evaluación	Excelente (100%)	Bueno (80%)	Regular (60%)	Deficiente (<50%)	Ponderación
1. Funcionalidad e Interactividad	La aplicación funciona fluidamente. Los filtros afectan lógicamente a todos los gráficos. El código es estructurado y eficiente.	La aplicación funciona, aunque algún filtro presenta inconsistencias menores o lentitud leve. Código funcional.	La aplicación presenta errores frecuentes, lentitud notable o filtros que no responden.	La aplicación no ejecuta o tiene errores críticos que impiden su uso.	15 pts
2. Calidad de Visualización	Gráficos avanzados y adecuados para el dato. Diseño limpio, sin elementos distractores. Personalización de colores efectiva.	Gráficos correctos. Buen intento de limpieza visual y personalización, aunque con detalles menores.	Gráficos básicos o con configuración por defecto (ruido visual). Elección de gráficos mejorable.	Gráficos ilegibles, sin etiquetas claras o que no comunican la información.	15 pts
3. Storytelling y Diseño	Títulos de gráficos conclusivos ("Action titles"). Narrativa clara apoyada en texto (st.markdown). Diseño visual coherente.	Buenos títulos y uso de texto. La narrativa visual es coherente, aunque podría ser más enfática.	Títulos puramente descriptivos sin hallazgos. Diseño visual desordenado o colores aleatorios.	Ausencia total de hilo narrativo o contexto escrito en la aplicación.	15 pts
4. Profundidad del Análisis	La aplicación responde a la pregunta de negocio con hallazgos profundos y accionables para la gerencia.	Se responden las preguntas con hallazgos claros, aunque con menor profundidad estratégica.	Análisis superficial. Se presentan datos, pero no se deriva información útil para decidir.	No se identifica la problemática de negocio ni se ofrecen conclusiones.	15 pts
5. Comunicación y Presentación (Video)	Todos los integrantes participan y explican una parte. El video es claro, profesional y respeta el tiempo límite. Se simula un tono ejecutivo.	La mayoría participa. La explicación es clara, aunque la calidad del audio/video o el tono podría mejorar.	Solo habla una persona o la participación es desbalanceada. El video excede el tiempo o es confuso.	Video no entregado, problemas graves de audio o falta de profesionalismo.	20 pts
5. Heteroevaluación (Evaluación de Pares)	La nota en esta categoría se basará en el promedio de las calificaciones otorgadas por los demás miembros del grupo.				20 pts
PUNTAJE TOTAL					100 pts

3 Puntos Extra (Opcional): Publicación del Dashboard en Streamlit Cloud

Objetivo: El valor real de una herramienta de visualización radica en su accesibilidad. Para obtener estos puntos extra, el equipo deberá desplegar su aplicación en un servidor web, entregando una URL pública funcional (ej. <https://mi-proyecto.streamlit.app>) que permita acceder al dashboard desde cualquier dispositivo sin necesidad de ejecutar código.

Consideraciones Técnicas para el Despliegue: Streamlit Cloud es un servicio gratuito que permite alojar aplicaciones. Para que el despliegue sea exitoso y la aplicación funcione correctamente en la nube, deben cumplir con los siguientes tres requisitos técnicos:

- Alojamiento del código en GitHub:** Streamlit Cloud no carga archivos directamente desde su computadora personal; se sincroniza con un repositorio en la nube.
 - Requisito: deben crear una cuenta en GitHub y subir sus tres archivos esenciales (`app.py`, `requirements.txt`, `dataset.csv`) a un repositorio público.
- Gestión de dependencias (requirements.txt):** el servidor en la nube es un entorno nuevo que no tiene instaladas las librerías que ustedes utilizaron (como `pandas` o `plotly`). Es necesario indicarle explícitamente qué instalar.
 - Instrucción:** deben incluir en su repositorio un archivo de texto plano llamado `requirements.txt`. Este archivo debe contener una lista simple de las librerías requeridas
 - Contenido del archivo:**
`streamlit`
`pandas`
`plotly`



3. **Uso de rutas relativas:** un error común es referenciar archivos usando la ubicación de su disco duro local (ej: C:/Usuarios/Nombre/Proyecto/datos.csv). Esta ruta no existe en el servidor de la nube, lo que provocará un error inmediato.
 - Pista: utilicen rutas relativas. Si el archivo CSV está en la misma carpeta que su script principal en GitHub, Python solo necesita el nombre del archivo para encontrarlo.
 - Solución correcta: `pd.read_csv("ventas_2024.csv")`
 - Solución incorrecta: `pd.read_csv("C:/Usuarios/Documentos/ventas_2024.csv")`

Pasos para el despliegue:

1. Carguen su código final y archivos al repositorio de GitHub.
2. Ingresen a **share.streamlit.io** y vinculen su cuenta de GitHub.
3. Seleccionen su repositorio y el archivo principal (`app.py`).
4. Hagan clic en "Deploy" para publicar la aplicación y obtener su enlace web.

Para **validar el puntaje adicional**, deben copiar el enlace público generado (URL) y pegarlo en la sección de comentarios de la carpeta de entrega del proyecto en D2L, o incluirlo dentro de un archivo de texto junto con los entregables principales.