

Parte II: proyecto de visualización

En esta segunda parte de la práctica, el estudiante deberá desarrollar una visualización de datos que demuestre un cierto conocimiento del campo de la visualización de datos, así como la elección apropiada de herramientas y técnicas. Se utilizará el conjunto o conjuntos de datos seleccionados y validados en la primera parte de la práctica, incluyendo en la medida de lo posible las mejoras sugeridas por el profesor.

Esta práctica está destinada a los siguientes objetivos:

- Escoger herramientas apropiadas para la creación de la/las visualizaciones presentadas
- Crear un proyecto de visualización de datos con una estructura, diseño y contenido profesional.
- Definir y responder razonadamente preguntas clave mediante la interacción con la visualización presentada.
- Diseñar los elementos de interactividad y de comunicación con el usuario de manera efectiva.

Además de la visualización, el estudiante tendrá que entregar un video que describa los puntos siguientes:

1. [20%] El proceso de creación seguido y las decisiones de diseño que se hayan tomado a lo largo del desarrollo.

Este dashboard nos permite decidir en base a la información obtenida, que proveedor de taxi tiene mayor apego con los clientes. En este caso, es el proveedor 2, el que tiene mayor demanda en los clientes. Asimismo, nos permite saber en que horarios la demanda de taxi es más solicitada, en la visualización podemos ver que hay mayor demanda entre las 3pm y 8pm. En relación a la duración del viaje, se tiene que el 95% de todos los viajes registrados son inferior a los 20 minutos.

2. [20%] La presentación in situ de la visualización, es decir, comentar las características de la visualización mientras se navega.

- Gráfico de barra: Nos permite representar la frecuencia de datos por cada categoría que se presenta en la variable, esto nos ayuda a comparar diferentes categorías o grupos.
- Histograma: Un histograma es una representación gráfica de la distribución de un conjunto de datos numéricos. Divide los datos en intervalos y muestra la frecuencia de los datos en cada intervalo.
- Dispersión de puntos: Nos permite medir la relación o correlación entre dos variables, detectar tendencias y patrones, identificar outliers, y analizar la distribución conjunta de dos variables.
- Gráfico de violín: Nos permite comparar distribuciones entre diferentes grupos o categorías, visualizar la densidad y la variabilidad de los datos.
- Diagrama de Árbol: Un diagrama de árbol representa datos jerárquicos usando rectángulos anidados. Cada rectángulo representa una categoría, y su tamaño es proporcional al valor que

representa. Las categorías pueden estar subdivididas en subcategorías.

3. [15%] Presentaremos brevemente las características más destacadas del conjunto de datos

Las características más relevantes que nos aporta en el “Dashboard de Viajes en Taxi: Análisis de desempeño, Distancia y Tarifas” tenemos a:

Variable	Tipo de datos	Descripción
VendorID	Cuantitativo	Identificador del proveedor de taxi.
tpep_pickup_datetime	Fecha	Fecha y hora en que el viaje en taxi comenzó.
tpep_dropoff_datetime	Fecha	Fecha y hora en que el viaje en taxi finalizó.
passenger_count	Cuantitativo	Número de pasajeros en el taxi durante el viaje.
trip_distance	Cuantitativo	Distancia recorrida durante el viaje en millas.
RatecodeID	Cuantitativo	Código tarifario que determina la tarifa del viaje (por ejemplo, tarifa estándar, tarifa nocturna, tarifa fija, etc.).
payment_type	Cuantitativo	Método de pago utilizado para el viaje (por ejemplo, efectivo, tarjeta de crédito, etc.).
total_amount	Cuantitativo	Monto total pagado por el viaje, incluyendo tarifa, impuestos, peajes, propina, etc. No incluye propinas en efectivo.

4. [20%] Las preguntas que responde la visualización.

- ¿En qué horarios hay mayor frecuencia de viajes?
- ¿Quién es el proveedor de taxi que tiene mayor demanda?
- ¿Hay alguna relación entre la distancia recorrida (millas) y el monto total?
- ¿Hay alguna relación entre el número de pasajeros y el monto total?
- ¿Qué tipo de pago es más usado? ¿Y cuál es el monto acumulado?
- ¿Cuál es la tarifa de viaje más recurrente? ¿Y cuál es el monto acumulado?
- Del monto acumulado:
 - ¿Cuál es el tipo de pago que abarca la mayor proporción del monto acumulado?

5. [15%] Los elementos interactivos disponibles como parte de la visualización.

- Filtros: Estos filtros nos permite seleccionar proveedores.
- Seleccionador de rango: Ente filtro de seleccionar rango lo usamos para delimitar la duración de viajes.
- Desplegables: Este filtro lo usamos para filtrar el tipo de pago y el tipo de tarifa.
- Leyendas interactivas: En algunas gráficas tenemos unas leyendas interactivas que nos permiten evaluar y seleccionar el tipo de proveedor.
- Botones: Entre los botones implementados tenemos el de cambiar de página y pasar de una página a la otra.

6. [10%] Reflexión: ¿Qué he aprendido del conjunto de datos? ¿Qué he aprendido de las técnicas utilizadas? ¿Qué me habría gustado hacer y no pude / supe?

A través del conjunto de datos, he podido aprender técnicas de extracción y limpieza de datos en Power BI. Esto ha sido posible mediante la integración con Python, lo que ha facilitado enormemente el proceso de limpieza y transformación de los tipos de datos en varios procesos. Entre las técnicas más utilizadas y aprendidas en Power BI, se destaca la extracción de datos. Esta herramienta me permite extraer datos de diversas fuentes, ya sea desde un motor de base de datos hasta vincular con servicios en la nube para desplegar visualizaciones. En mi caso, extraje los datos desde un archivo '.csv' y, con la extensión de Python, pude desarrollar la limpieza y transformación de datos con mayor facilidad, además de generar cálculos propios con DAX.

Sin embargo, me habría gustado analizar dos variables adicionales: PULocationID y DOLocationID, que representan las zonas de taxi donde se origina el recogido del usuario y la zona de destino del taxi. Este análisis no se pudo concretar porque no logré encontrar un enlace entre estos IDs y la ubicación geográfica. Mi intención era graficar en un mapa y situar los lugares con mayor demanda y los destinos más concurridos.

El video debe durar entre 4 y 6 minutos. Hay que respetar este rango de tiempo, se penalizará superar los márgenes porque también se evalúa la capacidad de síntesis y la calidad del guión.

La URL de la pantalla de visualización inicial tiene que ser pública y se tiene que poder explorar sin registrarse. Podéis usar, por ejemplo, las páginas GitHub (<https://pages.github.com/>), sorgeix.sh o colgar vuestra aplicación en un espacio web accesible. Si se usa código, el código se tiene que publicar en un repositorio público, como GitLab o GitHub, con una licencia de código abierto. Tenéis que incluir todos los ficheros necesarios para ejecutar la aplicación en un servidor web local. En el caso de Tableau, se puede publicar a Tableau Public, o también en otras herramientas en línea (Observable, Infogram, Flourish,...).