

75.06/95.58 Organización de Datos

Primer Cuatrimestre de 2024

Trabajo Práctico 1: Enunciado

El trabajo práctico 1 de la materia se basa en el análisis de datos del registro histórico de consumos de la línea aérea Airways de en <https://www.kaggle.com/datasets/ghassenkhaled/airways-customer-data>. Los datos incluyen reservas, venta de pasajes y reviews. Datos

Customer_airways_data

- Num_passengers: Cantidad de pasajeros que viajaron
- Sales_channel: Canal de venta por el cual fue realizada la reserva.
- Trip_type: Puede ser Round Trip (ida y vuelta), One way (solo ida) o Circle trip (secuencia de tramos)
- Purchase_lead: Número de días entre el momento de la reserva y el momento del viaje.
- Length_of_stay: Número de días en el destino antes del regreso.
- Flight_hour: Horario de partida del vuelo
- Flight_day: Día de la semana de partida del vuelo
- Route: Destino
- Booking_origin: País desde donde se realizó la reserva.
- Wants_extra_bags: Si el cliente solicitó equipaje extra al hacer la reserva.

Cleaned-reviews

- Reviews: Texto conteniendo el review del usuario
- Rate: Puntaje (de 1 a 10) otorgado
- Date: Fecha
- Country: País de origen.
- Verified: Si el registro fue validado y se puede considerar un review legítimo
- Comments: Texto ingresado por el usuario.

Objetivo

El objetivo del primer TP es realizar un análisis exploratorio del set de datos. Queremos ver qué cosas podemos descubrir sobre los datos que puedan resultar interesantes.

Requisitos

Los requisitos de la primera entrega son los siguientes:

- El análisis debe estar hecho en Python Pandas.
- El análisis debe entregarse en formato pdf, conteniendo el análisis realizado, explicaciones, hipótesis, resultados, y visualizaciones que puedan resultar de interés.
- Debe informarse también el link a los notebooks / un repositorio Github en donde pueda bajarse el código completo para generar el análisis.
- El trabajo no debe limitarse a analizar cada archivo por separado, se espera que puedan buscar relaciones entre los datos de distintos archivos tratados en forma conjunta.
- El trabajo debe incluir al menos 10 visualizaciones, orientadas a entender mejor los datos y sus relaciones, y al menos una enfocada a mostrar alguna de las conclusiones a la que llegaron. Ésta no puede ser un plot simple, deben incluirse elementos de diseño y trabajarse la estética para que la misma sirva para ser presentada a un público no necesariamente experimentado en el mundo de la aviación.

Evaluación

La evaluación del TP se realizará en base al siguiente criterio:

- Originalidad del análisis exploratorio.
- Calidad del reporte. ¿Está bien escrito? ¿Es claro y preciso?
- Calidad del análisis exploratorio: qué tipo de preguntas se hacen y de qué forma se responden, ¿es la respuesta clara y concisa con respecto a la pregunta formulada?
- Calidad de las visualizaciones presentadas.
 - ¿Tienen todos los ejes su rótulo?
 - ¿Tiene cada visualización un título?
 - ¿Es entendible la visualización sin tener que leer la explicación?
 - ¿El tipo de plot elegido es adecuado para lo que se quiere visualizar?
 - ¿Es una visualización interesante?
 - ¿El uso del color es adecuado?
 - ¿Hay un exceso o falta de elementos visuales en la visualización elegida?
 - ¿La visualización es consistente con los datos?
- Conclusiones presentadas.
 - ¿Presenta el grupo un listado de "insights" aprendidos sobre los datos en base al análisis realizado? ¿Es interesante?
 - ¿Descubrieron cosas que pueden ser de interés a las empresas de aviación, o a futuros pasajeros en busca de pasajes?

Criterio de Reentrega

En caso de que la entrega no esté en condiciones de aprobar, se considerará la posibilidad de realizar una reentrega siempre y cuando el trabajo presentado cumpla con todas las condiciones antes descriptas para la entrega, y solamente se trate de mejorar algún punto que pueda requerir más desarrollo. No se considerarán para reentrega trabajos que hayan sido presentados en forma incompleta.

Asignación Ayudantes

Grupo	Apellido y nombre Integrante 1	Padrón Integrante 1	Apellido y nombre Integrante 2	Padrón Integrante 2	Ayudante
1	Flores Sosa Zoraida Yurico	87039	Choque Carmona Luis Alberto	93738	Nacho B
2	de San Vicente, Hernán Martín	108800	Converso, Lara Daniela	107632	Julieta
3	Jonathan Leonel Pepe	94692	Henseler Santiago	110732	Nacho B
4	Marianela Fernanda Gareca Janko	109606	Beltran Malbran	110036	Julieta
5	Tomás Ezequiel Apud Mora	107968	Juan Angel Gomez	87943	JuanMa/Alejo
6	Hsieh Cindy Teresa	108051	Dominguez Gonzalo Alejo	109759	Nacho A
7	Nemirovsky Federico	108560	Czerwiak Juan Joaquin	109640	Nico
8	Batallan David Leonardo	97529	Fontana Maria Agustina	108090	Joaco
9	Cristian Eduardo Roldan Montes	96713	Nicolas Rizzo Ehrenbock	109756	Martín
10	Pellicieri Agustin	108172	Dellisola Florencia	109897	Matías Rotondo
11	Simone Abril	110939	Fiorotto Camila	108239	Lucas Correa
12	Brandan, Esteban Ariel	109186	Alonso Schneider, Micaela Alonso	109551	Matías Rotondo
13	Martina Gualdi	110513	Cruz Marcela Jazmin	110066	Damu
14	Sol Guadalupe	109525	Federico Martin Forte	92389	Nati

	Urbano				
15	Leith Cristian	103494	Guttlein Gareis Alexis Daniel	104431	Juan Pablo
16	Franco Bricchi	104438	Matías Morales	106793	Nico
17	Pablo Ezequiel Ruiz	109407	Lucas Couttulenc	109726	Lucas Correa
18	Galarza Adrian	106194	Vargas Rodrigo	97076	Lucas Fiegl
19	Santellán Felipe Martin	109022	Brandan Ricardo Ezequiel	96970	JuanMa/Alejo
20	Lucas Incerti	106972	Martin Quilis	105005	Juan Pablo
21	Brizuela Lopez, Mariano Jesus	105079	Fernandez Moreno, Tomas Agustin	109951	Nati
22	Axel Aparicio Rotermund	96283	Ignacio Orona	108013	Lucas Fiegl
23	Aizenscher Tomas	105350	Facundo Lucio Brondo	97640	Damu
24	Graziosi German	104623	Condori Luis	98237	Martín
25	Idoia Larrayoz	183652	Cortiñas Sabrina	97052	Joaco
26	Emanuel Sagman	103439	Ariel Folgueira	109473	JuanMa/Alejo