

Examen 1

Python para datos


Datasets

1- Recibos Débitos Créditos y Facturas:

<https://drive.google.com/file/d/1cX0WnmqxbD8BMebgmvggDA9Gnr0wlnh5/view?usp=sharing>

g

2- dolar blue:

 dolar_blue

Consignas para desarrollo

No podrá usar la computadora personal. Saber cómo utilizar en cualquier computadora jupyter notebook desde cero es parte del examen también.

Todas las preguntas requieren código + interpretación.

Entrega

Imprimir en la jupyter lo requerido en la pregunta. Pasar a word la respuesta.

2 archivos de entrega:

1- word con las respuestas teóricas

2- Jupyter notebook resuelto con las soluciones impresas en la jupyter.

Contexto: Tú Crédito.

Este dataset corresponde a una empresa que cobra préstamos y a ese dinero lo vuelve a colocar en distintas estrategias financieras. Ya sea prestar más o bonos/acciones.

Entienda las siguientes preguntas en ese contexto.

1. Cuáles son las características principales del DF?
 - a. Cantidad de rows y columnas.
 - b. Clases de datos
2. El dataset cuenta con datos faltantes?
 - a. Muestre la cantidad de faltantes por columnas
 - b. Que porcentaje de faltantes hay
 - c. Muestre las filas con datos faltantes
 - d. Sí hay faltantes, ¿Cómo procederán a hacer el tratamiento de los mismos?
 - e. Haga el tratamiento de los valores faltantes en caso de creerlo necesario.
3. Imprima las estadísticas básicas del dataset. Hay un método para eso tanto en python como en R (count, mean, 25%,etc)
4. ¿Qué es un dato outlier? ¿Cómo procedería al análisis del mismo?
5. Cambie el nombre de las columnas a "nro_comprobante", "fecha", "monto", "id_cliente"
6. ¿Cuántos clientes únicos hay en el dataset?
7. Cree una nueva columna que llamada "tipo" que consista en agarrar la columna "nro_comprobante" y separar el primer string dejando sólo si es recibo, factura, nota de débito o nota de crédito.

Hints!

- a. `.apply(lambda x:x.split(" ")[0])`
8. Tipos de Comprobante.Responda.
 - a. Cuántos registros se tienen de cada uno?
 - b. monto total de los recibos
 - c. monto total de las notas de débito
 - d. monto total de las notas de crédito
 - e. monto total de lo facturado
9. ¿Dé qué tipo de datos es la columna fecha? Pasarla a datetime
10. Crear la columna año- mes.
11. Sumar sólo lo cobrado (REC y N/D) y mostrar una tabla que tenga como indice fecha(año-mes) y columna monto_cobrado, cantidad_recibos. (group by)
12. ¿Cuántos valores hay mayores a 60000 de la columna monto en el dataset general?
13. ¿Cual es el valor máximo de la columna monto? ¿Debería eliminarlo? ¿Por qué? (en el dataset general)
14. No tenga en cuenta el valor máximo para los siguientes análisis.
15. En el total del dataset.
 - a. Cuanto representa lo facturado sobre lo cobrado.

- b. Si sabe que lo facturado es el IVA ¿Tiene sentido? ¿Cómo lo interpretas?

16. Dataset dolar:

- a. Repita puntos 1,2,3 ahora para este dataset
 - b. Pasar a datetime la columna Fecha
 - c. Crear año-mes
17. Cree un dataset del valor promedio de la venta del dólar por año-mes (*groupby*).
(fijarse ejercicio siguiente en caso de que no de nada de output)
18. En python pase la columna Dolar Venta a Float.
(**Hint** en python dolar_df.Venta.apply(lambda x:x.replace(",",".")))
19. Cree un dataset que sea el resultado de hacer un merge/join entre el dataset agrupado de recibos_facts_debitos con el agrupado de dolar del punto anterior.
(Hint en python pd.merge(rec_mensual, dolar_df_mensual,how="left", left_index=True,right_index=True))
20. Hacer un lineplot del total cobrado por año mes
21. lineplot con clientes únicos por año mes en otro gráfico
22. Poner ambas líneas en dos gráficos alineados en la misma columna
23. Haga otro gráfico que le parezca interesante. ¿Qué idea transmite con ese gráfico?
24. Imprima la ubicación de este notebook (el path)

Teoría (en menos de una carilla)

25. Teniendo en cuenta este dataset ¿Qué problema podrías resolver con Machine Learning?
26. En caso de embarcarse en una tarea de Machine Learning. Describa el proceso CRISPDM y explíquelo para este caso.
27. ¿Qué decisión del negocio o tarea se podría automatizar o escalar? ¿Por qué?