

EIDER ALEJANDRO PEÑA DAGUA

Prueba técnica DATAUP

1. Generar un dataset que agrupe por ítem y mes, desde enero 2020 a enero 2022, la cantidad total de ventas en kilos.

Desarrollo:

Para la solución del punto, se subió el archivo a Google Colaboratory en el archivo de Datos_de_DataUp.ipynb, al leerlo se muestra un archivo que debe ser depurado con fines prácticos para poder realizar un manejo más práctico de la información contenida en el documento. Para esto se eliminan algunas columnas y filas, también se renombran las columnas, al final se cambia el tipo de dato para las columnas de Venta_Neta_Kilos y Venta_Neta_ME de objeto a float, esto con fin de poder verificar que ninguna de las ventas sea negativa.

Por último se utiliza la sentencia de código:

```
Datosventas_Kilos =  
Datosventas.groupby(['Mes','Código_Material']).Venta_Neta_Kilos.sum().reset_index(name='Venta_Neta_Kilos')
```

Con la cual se agrupa los datos por Mes y Código_Material y se suma la columna Venta_Neta_Kilos

	Mes	Código_Material	Venta_Neta_Kilos
0	2020.01	1003831	960.0
1	2020.01	1003856	1705.0
2	2020.01	1003861	5510.0
3	2020.01	1003863	2900.0
4	2020.01	1003865	10790.0
...
2706	2022.02	3002507	7500.0
2707	2022.02	3014380	2400.0

2. Generar un dataset con el máximo, mínimo y promedio de venta en dinero por cada item desde enero 2021 a diciembre 2021.

Desarrollo:

Para generar este dataset primero se agrega una columna al dataframe llamada fecha que se extrae de la columna Mes esto con el fin de que esta columna luego se convierta en Index del dataframe, ya que con esta es más fácil extraer los datos correspondientes al año 2021, con esto listo se generan tres dataset que contienen el mínimo, el máximo y el promedio de las ventas.

```
#generamos 3 dataframe agrupados por mes y codigo del materia y a las ventas se les calcula el min max y mean
Datosventas_min = Datosventas2021.groupby(['Codigo_Material']).Venta_Neta_ME.min().reset_index(name = 'Venta_Neta_ME_Min')
Datosventas_max = Datosventas2021.groupby(['Codigo_Material']).Venta_Neta_ME.max().reset_index(name = 'Venta_Neta_ME_Max')
Datosventas_mean = Datosventas2021.groupby(['Codigo_Material']).Venta_Neta_ME.mean().reset_index(name = 'Venta_Neta_ME_Promedio')
```

Luego se hace una unión al para que todos los todos estén en un solo dataframe.

```
#creamos otra data frame donde hacemos un join para unir los data frame de min y max llamado Datosventas_min_max
Datosventas_min_max = pd.merge(Datosventas_min , Datosventas_max, on= ['Codigo_Material'], how='inner')
#se genera otro dataframe en donde a Datosventas_min_max se le añade el dataframe en donde se calculo el promedio
Datosventas_min_max_mean = pd.merge(Datosventas_min_max , Datosventas_mean, on= ['Codigo_Material'], how='inner')
Datosventas_min_max_mean
```

	Código_Material	Venta_Neta_ME_Min	Venta_Neta_ME_Max	Venta_Neta_ME_Promedio
0	1003831	0.0	8982669.0	1.215400e+06
1	1003856	0.0	6284320.0	6.399840e+05
2	1003861	0.0	22297684.0	5.177347e+06
3	1003863	66016.0	15985936.0	3.519141e+06
4	1003865	0.0	27709233.0	9.880328e+06
...
143	3024585	8813600.0	121091200.0	5.602088e+07
144	3025771	1830000.0	18300000.0	5.673000e+06

3. Generar un pronóstico de ventas en kilos por cada item para el mes en curso: febrero 2022. El modelo o técnica de pronóstico es de libre elección, puede ser tan simple o avanzado como se prefiera.

Desarrollo:

Para este punto se realizaron varios ensayos, utilizando diferentes técnicas como lo son las redes neuronales y regresiones lineales, con las redes neuronales se intento hacer uso de las redes neuronales recurrentes, pero no se obtuvo un resultado muy satisfactorio, al igual que con algunas redes densas no se logró una buena estimación, aunque siento que por este medio se puede lograr el tiempo no da para realizarlo. Por ende, el modelo

que más se ajusto fue una regresión lineal múltiple, que utiliza como valores de entrada el mes y las ventas, esto de cada código de material y de salida la venta neta en kilos, el archivo:

1. DataUp_modelo_lineal_Multiplescodigos.ipynb halla la predicción para todos lo códigos de material
2. DataUp_modelo_lineal.ipynb muestra el proceso que realiza DataUp_modelo_lineal_Multiplescodigos.ipynb pero solo para un código de material que este caso seria 1003861

Nota: ambos archivos se ejecutan con el archivo Datosventas_1.xlsx que se guarda al ejecutar el archivo Datos_de_DataUp.ipynb que este a su ves necesita del archivo ventas_nuevas.xlsx proporcionado por la empresa.

Nota 2: El archivo Datos_de_DataUp.ipynb también genera otros archivos como lo son:

Datosventas_kilos.xlsx correspondiente al dataset del punto 1

Datosventas_min_max_mean.xlsx correspondiente al punto 2

Datosventas.xlsx que es un archivo el cual contiene los datos depurados de Datos_nuevas.xlsx