

Managing data using Python for a dataset from French retail companies

Importing the library

```
In [40]: import sqlite3
import pandas as pd
```

Setting Up Database Connection and Cursor

```
In [41]: connection = sqlite3.connect('kucher.db')
cursor=connection.cursor()
```

Loading the CSV Data into DataFrame

```
In [42]: df = pd.read_csv('c:\\Users\\NP\\Desktop\\werkstudent docu\\cover letter\\English\\kucher\\csv files\\combined_csv.csv')
```

C:\\Users\\NP\\anaconda3\\lib\\site-packages\\python_core\\interactiveshell.py:3644: DtypeWarning: Columns (1) have mixed types.Specify dtype option on import or set low_memory=False.

Committing Changes to the Database

```
In [43]: connection.commit()
```

Displaying the DataFrame

```
In [44]: df
Out[44]:
```

	Point_de_Vente	Nomero_TPV	Nomero_Transaction	Date_Transaction	Heure	Typologie_Magasin	Nomero_Fidelite	Type_de_Vente	Univers_Produit	Segment_Produit	...	Fedex_Numero	Fedex_Libelle	Cible_Genre_Age
0	PDV-00062	TPV_783	TID000002834579	31/01/2022	16:00:00	Typologie_Magasin_06	N_2563125	TV3	CL1_9	CL2_51	...	FedexNum2138	FedexLib0522	CGAF
1	PDV-00040	TPV_769	TID000002759437	31/01/2022	18:00:00	Typologie_Magasin_06	N_3420193	TV3	CL1_9	CL2_51	...	FedexNum2138	FedexLib0522	CGAF
2	PDV-000173	TPV_610	TID00000003371	31/01/2022	15:00:00	Typologie_Magasin_06	NAN	TV3	CL1_9	CL2_51	...	FedexNum2138	FedexLib0522	CGAF
3	PDV-000400	TPV_586	TID00000048672	31/01/2022	16:00:00	Typologie_Magasin_06	NAN	TV3	CL1_29	CL2_48	...	FedexNum1482	FedexLib0477	CGAF
4	PDV-000420	TPV_488	TID00000573995	31/01/2022	15:00:00	Typologie_Magasin_06	N_362364	TV3	CL1_29	CL2_48	...	FedexNum1482	FedexLib0477	CGAF
...
543956	PDV-000020	TPV_711	TID000001963968	17/03/2022	16:00:00	Typologie_Magasin_06	N_1731337	TV3	CL1_10	CL2_72	...	FedexNum500	FedexLib1337	CGAF1
543957	PDV-000192	TPV_411	TID000001797562	17/03/2022	10:00:00	Typologie_Magasin_06	NAN	TV3	CL1_10	CL2_72	...	FedexNum500	FedexLib1337	CGAF1
543958	PDV-000398	TPV_92	TID00000784967	18/03/2022	13:00:00	Typologie_Magasin_06	NAN	TV3	CL1_10	CL2_72	...	FedexNum500	FedexLib1337	CGAF1
543959	PDV-000446	TPV_793	TID000002967223	17/03/2022	12:00:00	Typologie_Magasin_06	NAN	TV3	CL1_10	CL2_72	...	FedexNum500	FedexLib1337	CGAF1
543960	PDV-000026	TPV_718	TID000005468352	14/03/2022	17:00:00	Typologie_Magasin_06	NAN	TV3	CL1_10	CL2_72	...	FedexNum500	FedexLib1337	CGAF1

543961 rows × 22 columns

Previewing the dataframe

```
In [45]: print(df.head())
Out[45]:
```

	Point_de_Vente	Nomero_TPV	Nomero_Transaction	Date_Transaction	Heure	Typologie_Magasin	Nomero_Fidelite	Type_de_Vente	Univers_Produit	Segment_Produit	...	Fedex_Numero	Fedex_Libelle	Cible_Genre_Age						
0	PDV-00062	TPV_783	TID000002834579	31/01/2022	16:00:00	Typologie_Magasin_06	N_2563125	TV3	CL1_9	CL2_51	...	FedexNum2138	FedexLib0522	CGAF						
1	PDV-00040	TPV_769	TID000002759437	31/01/2022	18:00:00	Typologie_Magasin_06	N_3420193	TV3	CL1_9	CL2_51	...	FedexNum2138	FedexLib0522	CGAF						
2	PDV-000173	TPV_610	TID00000003371	31/01/2022	15:00:00	Typologie_Magasin_06	NAN	TV3	CL1_9	CL2_51	...	FedexNum2138	FedexLib0522	CGAF						
3	PDV-000400	TPV_586	TID00000048672	31/01/2022	16:00:00	Typologie_Magasin_06	NAN	TV3	CL1_29	CL2_48	...	FedexNum1482	FedexLib0477	CGAF						
4	PDV-000420	TPV_488	TID00000573995	31/01/2022	15:00:00	Typologie_Magasin_06	N_362364	TV3	CL1_29	CL2_48	...	FedexNum1482	FedexLib0477	CGAF						
...						
0	Typologie_Magasin_05	N_2563125	TV3	CL1_9	CL1_9	Typologie_Magasin_05	N_1420193	TV3	CL1_9	CL1_9	...	FedexNum500	FedexLib1337	CGAF1						
1	Typologie_Magasin_05	NAN	TV3	CL1_9	CL1_9	Typologie_Magasin_05	NAN	TV3	CL1_9	CL1_9	...	FedexNum500	FedexLib1337	CGAF1						
2	Typologie_Magasin_05	NAN	TV3	CL1_9	CL1_9	Typologie_Magasin_05	NAN	TV3	CL1_9	CL1_9	...	FedexNum500	FedexLib1337	CGAF1						
3	Typologie_Magasin_05	NAN	TV3	CL1_29	CL1_29	Typologie_Magasin_05	NAN	TV3	CL1_10	CL2_72	...	FedexNum500	FedexLib1337	CGAF1						
4	Typologie_Magasin_05	N_262364	TV3	CL1_29	CL1_29	Typologie_Magasin_05	NAN	TV3	CL1_10	CL2_72	...	FedexNum500	FedexLib1337	CGAF1						
...						
0	Segment_Produit	...	Fedex_Numero	Fedex_Libelle	Cible_Genre_Age	CL2_51	...	FedexNum2138	FedexLib0522	CGAF	CL2_51	...	FedexNum2138	FedexLib0522	CGAF					
1	CL2_51	...	FedexNum2138	FedexLib0522	CGAF	CL2_51	...	FedexNum2138	FedexLib0522	CGAF	CL2_51	...	FedexNum2138	FedexLib0522	CGAF					
2	CL2_51	...	FedexNum2138	FedexLib0522	CGAF	CL2_51	...	FedexNum2138	FedexLib0522	CGAF	CL2_51	...	FedexNum2138	FedexLib0522	CGAF					
3	CL2_48	...	FedexNum1482	FedexLib0477	CGAF	CL2_48	...	FedexNum1482	FedexLib0477	CGAF	CL2_48	...	FedexNum1482	FedexLib0477	CGAF					
4	CL2_48	...	FedexNum1482	FedexLib0477	CGAF	CL2_48	...	FedexNum1482	FedexLib0477	CGAF	CL2_48	...	FedexNum1482	FedexLib0477	CGAF					
...						
0	Modele_Couleur_Ref	Modele_Couleur_Libelle	Type_de_Vente	NPS	Quantite_Vendue	MCR37031	MCL63842	NPS3	1	MCR37031	MCL63842	NPS3	1	MCR37031	MCL63842	NPS3	1			
1	MCR173774	MCL183848	NPS3	1	MCR173774	MCL183848	NPS3	1	MCR173774	MCL183848	NPS3	1	MCR173774	MCL183848	NPS3	1	MCR173774	MCL183848	NPS3	1
2	MCR173774	MCL183848	NPS3	1	MCR173774	MCL183848	NPS3	1	MCR173774	MCL183848	NPS3	1	MCR173774	MCL183848	NPS3	1	MCR173774	MCL183848	NPS3	1
3	MCR69542	MCL138327	NPS3	1	MCR69542	MCL138327	NPS3	1	MCR69542	MCL138327	NPS3	1	MCR69542	MCL138327	NPS3	1	MCR69542	MCL138327	NPS3	1
4	MCR69542	MCL138327	NPS3	1	MCR69542	MCL138327	NPS3	1	MCR69542	MCL138327	NPS3	1	MCR69542	MCL138327	NPS3	1	MCR69542	MCL138327	NPS3	1
...
0	CA_Net_TTC	CA_Net_MT	Marge_Nette_Magasin	0	23,835,092,548,385,289	18,714,557,919,757,469	68,169,272,144,355,468	0	23,835,092,548,385,289	18,714,557,919,757,469	68,169,272,144,355,468	0	23,835,092,548,385,289	18,714,557,919,757,469	68,169,272,144,355,468	0	23,835,092,548,385,289	18,714,557,919,757,469	68,169,272,144,355,468	0
1	36,846,934,649,269,188	29,836,105,116,967,768	66,983,961,239,713,089	0	36,846,934,649,269,188	29,836,105,116,967,768	66,983,961,239,713,089	0	36,846,934,649,269,188	29,836,105,116,967,768	66,983,961,239,713,089	0	36,846,934,649,269,188	29,836,105,116,967,768	66,983,961,239,713,089	0	36,846,934,649,269,188	29,836,105,116,967,768	66,983,961,239,713,089	0
2	29,484,916,461,858,788	23,961,462,089,083,058	5,116,114	0	29,484,916,461,858,788	23,961,462,089,083,058	5,116,114	0	29,484,916,461,858,788	23,961,462,089,083,058	5,116,114	0	29,484,916,461,858,788	23,961,462,089,083,058	5,116,114	0	29,484,916,461,858,788	23,961,462,089,083,058	5,116,114	0
3	69,185,277,644,915,888	56,152,658,157,577,688	12,664,219,748,875,768	0	69,185,277,644,915,888	56,152,658,157,577,688	12,664,219,748,875,768	0	69,185,277,644,915,888	56,152,658,157,577,688	12,664,219,748,875,768	0	69,185,277,644,915,888	56,152,658,157,577,688	12,664,219,748,875,768	0	69,185,277,644,915,888	56,152,658,157,577,688	12,664,219,748,875,768	0
4	11,847,639,386,187,288	89,754,488,789,072,108	86,747,878,489,135,588	0	11,847,639,386,187,288	89,754,488,789,072,108	86,747,878,489,135,588	0	11,847,639,386,187,288	89,754,488,789,072,108	86,747,878,489,135,588	0	11,847,639,386,187,288	89,754,488,789,072,108	86,747,878,489,135,588	0	11,847,639,386,187,288	89,754,488,789,072,108	86,747,878,489,135,588	0
...
5	rows × 22 columns

Checking for Duplicate Rows in the DataFrame

```
In [46]: num_duplicates = df.duplicated().sum()
print(f'Number of duplicate rows: {num_duplicates}')
```

Number of duplicate rows: 1

Printing duplicate rows in a DataFrame

```
In [47]: duplicate_rows = df[df.duplicated()]
print(duplicate_rows)
Out[47]:
```

	Point_de_Vente	Nomero_TPV	Nomero_Transaction	Date_Transaction	Heure	Typologie_Magasin	Nomero_Fidelite	Type_de_Vente	Univers_Produit	Segment_Produit	...	Fedex_Numero	Fedex_Libelle	Cible_Genre_Age								
424228	PDV-00180311	NAN	TID00000535838	16/02/2022	16:02:28	Typologie_Magasin_05	N_2843578	TV3	CL1_9	CL2_51	...	FedexNum748	FedexLib18868	CGAF1								
424178	18:00:00	Typologie_Magasin_05	N_2843578	TV3	CL1_9	CL2_51	...	FedexNum748	FedexLib18868	CGAF1	...	FedexNum748	FedexLib18868	CGAF1								
424228	18:00:00	Typologie_Magasin_05	N_2843578	TV3	CL1_9	CL2_51	...	FedexNum748	FedexLib18868	CGAF1	...	FedexNum748	FedexLib18868	CGAF1								
424228	18:00:00	Typologie_Magasin_05	N_2843578	TV3	CL1_9	CL2_51	...	FedexNum748	FedexLib18868	CGAF1	...	FedexNum748	FedexLib18868	CGAF1								
...								
0	Univers_Produit	Segment_Produit	...	Fedex_Numero	Fedex_Libelle	CL1_21	CL2_4	FedexNum748	FedexLib18868	CGAF1	...	FedexNum748	FedexLib18868	CGAF1								
424178	CL1_21	CL2_4	...	FedexNum748	FedexLib18868	CGAF1	...	FedexNum748	FedexLib18868	CGAF1	...	FedexNum748	FedexLib18868	CGAF1								
424228	CL1_21	CL2_4	...	FedexNum748	FedexLib18868	CGAF1	...	FedexNum748	FedexLib18868	CGAF1	...	FedexNum748	FedexLib18868	CGAF1								
...								
0	Cible_Genre_Age	Modele_Couleur_Ref	Modele_Couleur_Libelle	CA_Net_TTC	CA_Net_MT	CGAF1	MCR95737	MCL11988	CGAF1	MCR95737	MCL11988	CGAF1	MCR95737	MCL11988	CGAF1	MCR95737	MCL11988	CGAF1	MCR95737	MCL11988	CGAF1	
424178	Type_de_Vente	NPS	Quantite_Vendue	CA_Net_TTC	CA_Net_MT	NPS2	2	0	0	NPS2	2	0	0	NPS2	2	0	0	NPS2	2	0	0	
424228	Marge_Nette_Magasin	0	23,835,092,548,385,289	18,714,557,919,757,469	68,169,272,144,355,468	0	23,835,092,548,385,289	18,714,557,919,757,469	68,169,272,144,355,468	0	23,835,092,548,385,289	18,714,557,919,757,469	68,169,272,144,355,468	0	23,835,092,548,385,289	18,714,557,919,757,469	68,169,272,144,355,468	0	23,835,092,548,385,289	18,714,557,919,757,469	68,169,272,144,355,468	0
...
5	rows × 22 columns

To display the duplicates

```
In [48]: if num_duplicates > 0:
    print(f'Number of duplicate rows: {num_duplicates}')
    df[df.duplicated(keep=False)]
    print(df[df.duplicated(keep=False)])
Out[48]:
```

	Point_de_Vente	Nomero_TPV	Nomero_Transaction	Date_Transaction	Heure	Typologie_Magasin	Nomero_Fidelite	Type_de_Vente	Univers_Produit	Segment_Produit	...	Fedex_Numero	Fedex_Libelle	Cible_Genre_Age
424178	PDV-00180311	NAN	TID00000535838	16/02/2022	16:02:28	Typologie_Magasin_05	N_2843578	TV3	CL1_9	CL2_51	...	FedexNum748	FedexLib18868	CGAF1
424228	PDV-00180311	NAN	TID00000535838	16/02/2022	16:02:28	Typologie_Magasin_05	N_2843578	TV3	CL1_9	CL2_51	...	FedexNum748	FedexLib18868	CGAF1
...
0	Univers_Produit	Segment_Produit	...	Fedex_Numero	Fedex_Libelle	CL1_21	CL2_4	FedexNum748	FedexLib18868	CGAF1	...	FedexNum748	FedexLib18868	CGAF1
424178	CL1_21	CL2_4	...	FedexNum748	FedexLib18868	CGAF1	...	FedexNum748	FedexLib18868	CGAF1	...	FedexNum748	FedexLib18868	CGAF1
424228	CL1_21	CL2_4	...	FedexNum748	FedexLib18868	CGAF1	...	FedexNum748	FedexLib18868	CGAF1	...	FedexNum748	FedexLib18868	CGAF1