reading in files

March 26, 2024

```python
[163]: import pandas as pd
       import numpy as np
```

```python
[164]: df = pd.read_csv(r"D:\New folder\quantine\QVI_purchase_behaviour.csv")
       df.head()
```

```
[164]:    LYLTY_CARD_NBR                 LIFESTAGE PREMIUM_CUSTOMER
       0            1000   YOUNG SINGLES/COUPLES          Premium
       1            1002   YOUNG SINGLES/COUPLES       Mainstream
       2            1003           YOUNG FAMILIES           Budget
       3            1004   OLDER SINGLES/COUPLES       Mainstream
       4            1005  MIDAGE SINGLES/COUPLES       Mainstream
```

```python
[165]: import pandas as pd
```

```python
[166]: df2 = pd.read_excel(r'D:\New folder\quantine\QVI_transaction_data.xlsx')
```

```python
[167]: print(df2)
```

```
                DATE  STORE_NBR  LYLTY_CARD_NBR  TXN_ID  PROD_NBR  \
       0       43390          1            1000       1         5
       1       43599          1            1307     348        66
       2       43605          1            1343     383        61
       3       43329          2            2373     974        69
       4       43330          2            2426    1038       108
       ...       ...        ...             ...     ...       ...
       264831  43533        272          272319  270088        89
       264832  43325        272          272358  270154        74
       264833  43410        272          272379  270187        51
       264834  43461        272          272379  270188        42
       264835  43365        272          272380  270189        74

                                       PROD_NAME  PROD_QTY  TOT_SALES
       0          Natural Chip        Compny SeaSalt175g         2        6.0
       1                        CCs Nacho Cheese    175g         3        6.3
       2          Smiths Crinkle Cut  Chips Chicken 170g         2        2.9
       3          Smiths Chip Thinly  S/Cream&Onion 175g         5       15.0
       4      Kettle Tortilla ChpsHny&Jlpno Chili 150g         3       13.8
       ...                                    ...       ...        ...
```

```
264831        Kettle Sweet Chilli And Sour Cream 175g          2          10.8
264832                    Tostitos Splash Of  Lime 175g          1           4.4
264833                        Doritos Mexicana    170g          2           8.8
264834    Doritos Corn Chip Mexican Jalapeno 150g          2           7.8
264835                    Tostitos Splash Of  Lime 175g          2           8.8

[264836 rows x 8 columns]
```

[168]:
```python
#df2 = df2.drop(columns=['DATE', 'STORE_NBR', 'LYLTY_CARD_NBR', 'TXN_ID',
 ↪'PROD_NBR', '\'])
df2 = df2.drop(columns=['DATE', 'STORE_NBR', 'TXN_ID', 'PROD_NBR'])

print(df2)
```

```
        LYLTY_CARD_NBR                                PROD_NAME PROD_QTY  \
0                 1000    Natural Chip         Compny SeaSalt175g          2
1                 1307                CCs Nacho Cheese     175g          3
2                 1343    Smiths Crinkle Cut  Chips Chicken 170g          2
3                 2373    Smiths Chip Thinly  S/Cream&Onion 175g          5
4                 2426  Kettle Tortilla ChpsHny&Jlpno Chili 150g          3
...                  ...                                       ...       ...
264831          272319  Kettle Sweet Chilli And Sour Cream 175g          2
264832          272358               Tostitos Splash Of  Lime 175g          1
264833          272379                   Doritos Mexicana    170g          2
264834          272379  Doritos Corn Chip Mexican Jalapeno 150g          2
264835          272380               Tostitos Splash Of  Lime 175g          2

        TOT_SALES
0             6.0
1             6.3
2             2.9
3            15.0
4            13.8
...           ...
264831       10.8
264832        4.4
264833        8.8
264834        7.8
264835        8.8

[264836 rows x 4 columns]
```

[169]:
```python
print(df)
```

```
        LYLTY_CARD_NBR              LIFESTAGE PREMIUM_CUSTOMER
0                 1000  YOUNG SINGLES/COUPLES          Premium
1                 1002  YOUNG SINGLES/COUPLES       Mainstream
2                 1003          YOUNG FAMILIES           Budget
3                 1004  OLDER SINGLES/COUPLES       Mainstream
```

```
4              1005  MIDAGE SINGLES/COUPLES       Mainstream
...                ...                      ...              ...
72632        2370651  MIDAGE SINGLES/COUPLES       Mainstream
72633        2370701          YOUNG FAMILIES       Mainstream
72634        2370751          YOUNG FAMILIES          Premium
72635        2370961          OLDER FAMILIES           Budget
72636        2373711   YOUNG SINGLES/COUPLES       Mainstream

[72637 rows x 3 columns]
```

[170]:
```python
df = df.merge(df2, on="LYLTY_CARD_NBR")
df.head()
```

[170]:
```
   LYLTY_CARD_NBR               LIFESTAGE PREMIUM_CUSTOMER  \
0            1000   YOUNG SINGLES/COUPLES         Premium
1            1002   YOUNG SINGLES/COUPLES      Mainstream
2            1003          YOUNG FAMILIES          Budget
3            1003          YOUNG FAMILIES          Budget
4            1004  OLDER SINGLES/COUPLES      Mainstream

                          PROD_NAME  PROD_QTY  TOT_SALES
0  Natural Chip        Compny SeaSalt175g         2        6.0
1   Red Rock Deli Chikn&Garlic Aioli 150g         1        2.7
2   Grain Waves Sour    Cream&Chives 210G         1        3.6
3  Natural ChipCo      Hony Soy Chckn175g         1        3.0
4             WW Original Stacked Chips 160g         1        1.9
```

[172]:
```python
df["LIFESTAGE"].unique()
```

[172]:
```
array(['YOUNG SINGLES/COUPLES', 'YOUNG FAMILIES', 'OLDER SINGLES/COUPLES',
       'MIDAGE SINGLES/COUPLES', 'NEW FAMILIES', 'OLDER FAMILIES',
       'RETIREES'], dtype=object)
```

[173]:
```python
df["LIFESTAGE"] = df["LIFESTAGE"].str.replace("/", " ")
df["LIFESTAGE"].unique()
```

[173]:
```
array(['YOUNG SINGLES COUPLES', 'YOUNG FAMILIES', 'OLDER SINGLES COUPLES',
       'MIDAGE SINGLES COUPLES', 'NEW FAMILIES', 'OLDER FAMILIES',
       'RETIREES'], dtype=object)
```

[174]:
```python
df["PREMIUM_CUSTOMER"].unique()
```

[174]:
```
array(['Premium', 'Mainstream', 'Budget'], dtype=object)
```

[175]:
```python
df.head()
```

[175]:
```
   LYLTY_CARD_NBR               LIFESTAGE PREMIUM_CUSTOMER  \
0            1000   YOUNG SINGLES COUPLES         Premium
```

```
1          1002   YOUNG SINGLES COUPLES       Mainstream
2          1003           YOUNG FAMILIES          Budget
3          1003           YOUNG FAMILIES          Budget
4          1004   OLDER SINGLES COUPLES       Mainstream

                             PROD_NAME  PROD_QTY  TOT_SALES
0  Natural Chip        Compny SeaSalt175g         2        6.0
1   Red Rock Deli Chikn&Garlic Aioli 150g         1        2.7
2   Grain Waves Sour    Cream&Chives 210G         1        3.6
3  Natural ChipCo      Hony Soy Chckn175g         1        3.0
4          WW Original Stacked Chips 160g         1        1.9
```

[192]:
```python
#grouping function
#df.groupby('PREMIUM_CUSTOMER')['TOT_SALES'].sum()
df.groupby('PREMIUM_CUSTOMER')['TOT_SALES'].agg(['sum', 'mean'])
```

[192]:
```
                        sum       mean
PREMIUM_CUSTOMER
Budget            676211.55   7.258838
Mainstream        750744.50   7.361106
Premium           507458.95   7.281556
```

[196]:
```python
df.sort_values(by='TOT_SALES', ascending=False).head(20)
```

[196]:
```
        LYLTY_CARD_NBR                 LIFESTAGE PREMIUM_CUSTOMER  \
221625          226000           OLDER FAMILIES          Premium
221626          226000           OLDER FAMILIES          Premium
39548            44350            YOUNG FAMILIES          Budget
187012          190113           OLDER FAMILIES       Mainstream
118381          118021                 RETIREES       Mainstream
20349            24095                 RETIREES          Premium
190554          194308    OLDER SINGLES COUPLES          Premium
91451            94148   MIDAGE SINGLES COUPLES       Mainstream
44849            49303            OLDER FAMILIES          Premium
217026          222005            YOUNG FAMILIES       Mainstream
216848          221351    OLDER SINGLES COUPLES          Premium
151151          154199                 RETIREES          Budget
179406          181129    OLDER SINGLES COUPLES          Budget
170708          172239   MIDAGE SINGLES COUPLES          Premium
242078          243345    OLDER SINGLES COUPLES          Budget
98610           101110                 RETIREES          Budget
178850          180233            YOUNG FAMILIES          Budget
61934            66014            OLDER FAMILIES          Budget
236667          237317    OLDER SINGLES COUPLES       Mainstream
94040            96203   MIDAGE SINGLES COUPLES          Budget

                        PROD_NAME  PROD_QTY  TOT_SALES
```

```
221625         Dorito Corn Chp     Supreme 380g      200      650.0
221626         Dorito Corn Chp     Supreme 380g      200      650.0
39548    Smiths Crnkle Chip  Orgnl Big Bag 380g        5       29.5
187012   Smiths Crnkle Chip  Orgnl Big Bag 380g        5       29.5
118381   Smiths Crnkle Chip  Orgnl Big Bag 380g        5       29.5
20349    Smiths Crnkle Chip  Orgnl Big Bag 380g        5       29.5
190554   Smiths Crnkle Chip  Orgnl Big Bag 380g        5       29.5
91451    Smiths Crnkle Chip  Orgnl Big Bag 380g        5       29.5
44849    Smiths Crnkle Chip  Orgnl Big Bag 380g        5       29.5
217026       Smiths Crinkle      Original 330g         5       28.5
216848                     Cheezels Cheese 330g        5       28.5
151151       Doritos Cheese      Supreme 330g          5       28.5
179406                     Cheezels Cheese 330g        5       28.5
170708       Smiths Crinkle      Original 330g         5       28.5
242078  Smiths Crinkle Chips Salt & Vinegar 330g       5       28.5
98610   Smiths Crinkle Chips Salt & Vinegar 330g       5       28.5
178850                     Cheezels Cheese 330g        5       28.5
61934                      Cheezels Cheese 330g        5       28.5
236667       Doritos Cheese      Supreme 330g          5       28.5
94040        Smiths Crinkle      Original 330g         5       28.5
```

[197]: `df["PROD_NAME"].unique()`

```
[197]: array(['Natural Chip        Compny SeaSalt175g',
       'Red Rock Deli Chikn&Garlic Aioli 150g',
       'Grain Waves Sour    Cream&Chives 210G',
       'Natural ChipCo      Hony Soy Chckn175g',
       'WW Original Stacked Chips 160g', 'Cheetos Puffs 165g',
       'Infuzions SourCream&Herbs Veg Strws 110g',
       'RRD SR Slow Rst     Pork Belly 150g',
       'Doritos Cheese      Supreme 330g', 'Doritos Mexicana    170g',
       'Old El Paso Salsa   Dip Tomato Med 300g',
       'GrnWves Plus Btroot & Chilli Jam 180g',
       'Smiths Crinkle Cut  Chips Barbecue 170g',
       'Kettle Sensations   Camembert & Fig 150g',
       'Doritos Corn Chip Southern Chicken 150g',
       'CCs Tasty Cheese    175g', 'Tostitos Splash Of  Lime 175g',
       'Kettle 135g Swt Pot Sea Salt', 'RRD Salt & Vinegar  165g',
       'Infuzions Mango     Chutny Papadums 70g',
       'Smiths Crinkle Cut  Snag&Sauce 150g',
       'Smiths Crinkle      Original 330g',
       'RRD Sweet Chilli &  Sour Cream 165g',
       'Smiths Chip Thinly  S/Cream&Onion 175g',
       'Smiths Crinkle Chips Salt & Vinegar 330g',
       'Red Rock Deli SR    Salsa & Mzzrlla 150g',
       'Cobs Popd Sea Salt  Chips 110g',
       'Natural ChipCo Sea  Salt & Vinegr 175g',
```

'Natural Chip Co   Tmato Hrb&Spce 175g', 'Burger Rings 220g',
'Woolworths Cheese   Rings 190g',
'Smiths Thinly   Swt Chli&S/Cream175G',
'Thins Chips Seasonedchicken 175g',
'Smiths Thinly Cut   Roast Chicken 175g',
'Tyrrells Crisps   Ched & Chives 165g',
'Doritos Corn Chips   Cheese Supreme 170g',
'Smiths Chip Thinly   Cut Original 175g',
'Smiths Crinkle Cut   Chips Original 170g',
'Thins Chips Light&   Tangy 175g',
'Doritos Corn Chips   Original 170g',
'Kettle Sensations   Siracha Lime 150g',
'Smiths Crinkle Cut   Salt & Vinegar 170g',
'Smith Crinkle Cut   Bolognese 150g', 'Cheezels Cheese 330g',
'Kettle Chilli 175g', 'Tyrrells Crisps   Lightly Salted 165g',
'Twisties Cheese   270g', 'WW Crinkle Cut   Chicken 175g',
'RRD Chilli&   Coconut 150g',
'Infuzions BBQ Rib   Prawn Crackers 110g',
'Sunbites Whlegrn   Crisps Frch/Onin 90g',
'Doritos Salsa   Medium 300g',
'Kettle Tortilla ChpsFeta&Garlic 150g',
'Smiths Crinkle Cut   French OnionDip 150g',
'WW D/Style Chip   Sea Salt 200g',
'Smiths Chip Thinly   CutSalt/Vinegr175g',
'Kettle Sensations   BBQ&Maple 150g',
'Old El Paso Salsa   Dip Tomato Mild 300g',
'Tostitos Smoked   Chipotle 175g', 'RRD Lime & Pepper   165g',
'CCs Nacho Cheese   175g', 'Snbts Whlgrn Crisps Cheddr&Mstrd 90g',
'Kettle Tortilla ChpsBtroot&Ricotta 150g',
'Pringles Sthrn FriedChicken 134g',
'Pringles Chicken   Salt Crips 134g',
'French Fries Potato Chips 175g',
'Kettle Mozzarella   Basil & Pesto 175g', 'CCs Original 175g',
'Tostitos Lightly   Salted 175g',
'Smiths Crnkle Chip   Orgnl Big Bag 380g',
'Smiths Crinkle Cut   Chips Chicken 170g',
'Smiths Crinkle Cut   Chips Chs&Onion170g', 'Twisties Chicken270g',
'Woolworths Medium   Salsa 300g',
'Red Rock Deli Sp   Salt & Truffle 150G',
'RRD Pc Sea Salt   165g', 'WW Supreme Cheese   Corn Chips 200g',
'WW Original Corn   Chips 200g', 'Woolworths Mild   Salsa 300g',
'Cheezels Cheese Box 125g', 'Doritos Salsa Mild 300g',
'Cobs Popd Swt/Chlli &Sr/Cream Chips 110g',
'Infzns Crn Crnchers Tangy Gcamole 110g',
'WW Sour Cream &OnionStacked Chips 160g',
'Pringles Mystery   Flavour 134g', 'Pringles Barbeque   134g',
'Grain Waves   Sweet Chilli 210g',

```
        'Pringles Sweet&Spcy BBQ 134g', 'Kettle Original 175g',
        'Infuzions Thai SweetChili PotatoMix 110g',
        'Old El Paso Salsa   Dip Chnky Tom Ht300g',
        'Smiths Crinkle Cut  Tomato Salsa 150g',
        'Cheetos Chs & Bacon Balls 190g',
        'Kettle Sweet Chilli And Sour Cream 175g',
        'Doritos Corn Chips  Nacho Cheese 170g',
        'Cobs Popd Sour Crm  &Chives Chips 110g',
        'Red Rock Deli Thai  Chilli&Lime 150g',
        'Twisties Cheese     Burger 250g',
        'Kettle Sea Salt     And Vinegar 175g',
        'WW Crinkle Cut      Original 175g',
        'Dorito Corn Chp     Supreme 380g',
        'Doritos Corn Chip Mexican Jalapeno 150g',
        'Pringles SourCream  Onion 134g',
        'Kettle Tortilla ChpsHny&Jlpno Chili 150g',
        'RRD Steak &         Chimuchurri 150g',
        'Thins Chips Salt &  Vinegar 175g',
        'Thins Chips         Originl saltd 175g',
        'RRD Honey Soy       Chicken 165g',
        'Kettle Honey Soy    Chicken 175g',
        'NCC Sour Cream &    Garden Chives 175g',
        'Pringles Original   Crisps 134g',
        'Smith Crinkle Cut   Mac N Cheese 150g',
        'Thins Potato Chips  Hot & Spicy 175g', 'Pringles Slt Vingar 134g'],
      dtype=object)
```

```
[198]: most_sold_product = df['PROD_NAME'].value_counts().index[0]
       print(most_sold_product)
```

```
Kettle Mozzarella    Basil & Pesto 175g
```

```
[201]: count_product_Kettle_Mozzarella_Basil_Pesto_175g_customers = df[df['PROD_NAME']␣
       ↪== 'Kettle Mozzarella    Basil & Pesto 175g']['PREMIUM_CUSTOMER'].
       ↪value_counts().sum()
       print(count_product_Kettle_Mozzarella_Basil_Pesto_175g_customers)
```

```
3304
```

```
[205]: min_qty_product = df.groupby('PROD_NAME')['PROD_QTY'].min().idxmin()
       # Find the minimum quantity sold
       min_qty = df.groupby('PROD_NAME')['PROD_QTY'].min().min()
       # Find the number of times that quantity was sold
       count_min_qty = df[df['PROD_QTY'] == min_qty]['PROD_NAME'].
       ↪value_counts()[min_qty_product]
```

```
[206]: print(f'The minimum quantity of the product {min_qty_product} sold is {min_qty}␣
       ↪and it was sold {count_min_qty} times.')
```

The minimum quantity of the product Burger Rings 220g sold is 1 and it was sold 186 times.

```
[208]: avg_product_sales = df['PROD_QTY'].mean()
       #the average Q
       print(avg_product_sales)
```

1.907308674047335

```
[211]: #top_product = df.nlargest(n=1, columns='TOT_SALES')['PROD_NAME'].iloc[0]
       #num_customers_bought_top_product = df[df['PROD_NAME'] ==␣
        ↪top_product]['PREMIUM_CUSTOMER'].value_counts().sum()
       #print(num_customers_bought_top_product)
       top_10_products = df.nlargest(n=10, columns='TOT_SALES')
       print(top_10_products)
```

```
        LYLTY_CARD_NBR              LIFESTAGE PREMIUM_CUSTOMER  \
221625          226000          OLDER FAMILIES          Premium
221626          226000          OLDER FAMILIES          Premium
20349            24095                RETIREES          Premium
39548            44350          YOUNG FAMILIES           Budget
44849            49303          OLDER FAMILIES          Premium
91451            94148  MIDAGE SINGLES COUPLES       Mainstream
118381          118021                RETIREES       Mainstream
187012          190113          OLDER FAMILIES       Mainstream
190554          194308  OLDER SINGLES COUPLES          Premium
61934            66014          OLDER FAMILIES           Budget

                                    PROD_NAME  PROD_QTY  TOT_SALES
221625          Dorito Corn Chp     Supreme 380g       200      650.0
221626          Dorito Corn Chp     Supreme 380g       200      650.0
20349    Smiths Crnkle Chip  Orgnl Big Bag 380g         5       29.5
39548    Smiths Crnkle Chip  Orgnl Big Bag 380g         5       29.5
44849    Smiths Crnkle Chip  Orgnl Big Bag 380g         5       29.5
91451    Smiths Crnkle Chip  Orgnl Big Bag 380g         5       29.5
118381   Smiths Crnkle Chip  Orgnl Big Bag 380g         5       29.5
187012   Smiths Crnkle Chip  Orgnl Big Bag 380g         5       29.5
190554   Smiths Crnkle Chip  Orgnl Big Bag 380g         5       29.5
61934                   Cheezels Cheese 330g           5       28.5
```

```
[212]: count_last_10_products = df['PROD_NAME'].value_counts().tail(10).sum()
       print(count_last_10_products)
```

14293

```
[213]: last_10_products = df['PROD_NAME'].value_counts().tail(10)
       print(last_10_products)
```

```
PROD_NAME
Thins Chips        Originl saltd 175g        1441
```

```
Smiths Chip Thinly   CutSalt/Vinegr175g         1440
Smiths Crinkle Cut   French OnionDip 150g        1438
Red Rock Deli Chikn&Garlic Aioli 150g           1434
Sunbites Whlegrn     Crisps Frch/Onin 90g        1432
RRD Pc Sea Salt      165g                        1431
Woolworths Medium    Salsa 300g                  1430
NCC Sour Cream &     Garden Chives 175g          1419
French Fries Potato  Chips 175g                  1418
WW Crinkle Cut       Original 175g               1410
Name: count, dtype: int64
```

[218]: `df`

[218]:

	LYLTY_CARD_NBR	LIFESTAGE	PREMIUM_CUSTOMER	\
0	1000	YOUNG SINGLES COUPLES	Premium	
1	1002	YOUNG SINGLES COUPLES	Mainstream	
2	1003	YOUNG FAMILIES	Budget	
3	1003	YOUNG FAMILIES	Budget	
4	1004	OLDER SINGLES COUPLES	Mainstream	
...	...	...	...	
264831	2370701	YOUNG FAMILIES	Mainstream	
264832	2370751	YOUNG FAMILIES	Premium	
264833	2370961	OLDER FAMILIES	Budget	
264834	2370961	OLDER FAMILIES	Budget	
264835	2373711	YOUNG SINGLES COUPLES	Mainstream	

	PROD_NAME	PROD_QTY	TOT_SALES	\
0	Natural Chip Compny SeaSalt175g	2	6.0	
1	Red Rock Deli Chikn&Garlic Aioli 150g	1	2.7	
2	Grain Waves Sour Cream&Chives 210G	1	3.6	
3	Natural ChipCo Hony Soy Chckn175g	1	3.0	
4	WW Original Stacked Chips 160g	1	1.9	
...	...	...	...	
264831	Grain Waves Sweet Chilli 210g	2	7.2	
264832	Kettle Tortilla ChpsFeta&Garlic 150g	2	9.2	
264833	Tyrrells Crisps Lightly Salted 165g	2	8.4	
264834	Old El Paso Salsa Dip Chnky Tom Ht300g	2	10.2	
264835	Smiths Crinkle Chips Salt & Vinegar 330g	2	11.4	

	duplicated_PROD_NAME
0	False
1	False
2	False
3	False
4	False
...	...
264831	True

```
264832                    True
264833                    True
264834                    True
264835                    True

[264836 rows x 7 columns]
```

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]: