# Quantium Virtual Internship - Retail Strategy and Analytics

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

data_customer = pd.read_csv("QVI_purchase_behaviour.csv")
data_customer
```

```
       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]
```

In QVI_purchase_behaviour file we have 72637 rows and 3 columns

```
data_purchase = pd.read_excel("QVI_transaction_data.xlsx")
data_purchase
```

```
              DATE  STORE_NBR  LYLTY_CARD_NBR  TXN_ID  PROD_NBR  \
0      2018-10-17          1            1000       1         5
1      2019-05-14          1            1307     348        66
2      2019-05-20          1            1343     383        61
3      2018-08-17          2            2373     974        69
4      2018-08-18          2            2426    1038       108
...           ...        ...             ...     ...       ...
264831 2019-03-09        272          272319  270088        89
264832 2018-08-13        272          272358  270154        74
264833 2018-11-06        272          272379  270187        51
264834 2018-12-27        272          272379  270188        42
264835 2018-09-22        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]
```

In QVI_transaction_data file we have 264836 rows and 8 columns

```
data_customer.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 72637 entries, 0 to 72636
Data columns (total 3 columns):
 #   Column            Non-Null Count  Dtype 
---  ------            --------------  ----- 
 0   LYLTY_CARD_NBR    72637 non-null  int64 
 1   LIFESTAGE         72637 non-null  object
 2   PREMIUM_CUSTOMER  72637 non-null  object
dtypes: int64(1), object(2)
memory usage: 1.7+ MB

data_purchase.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 264836 entries, 0 to 264835
Data columns (total 8 columns):
 #   Column          Non-Null Count   Dtype         
---  ------          --------------   -----         
 0   DATE            264836 non-null  datetime64[ns]
 1   STORE_NBR       264836 non-null  int64         
 2   LYLTY_CARD_NBR  264836 non-null  int64         
 3   TXN_ID          264836 non-null  int64         
 4   PROD_NBR        264836 non-null  int64         
 5   PROD_NAME       264836 non-null  object        
 6   PROD_QTY        264836 non-null  int64         
 7   TOT_SALES       264836 non-null  float64
```

```
dtypes: datetime64[ns](1), float64(1), int64(5), object(1)
memory usage: 16.2+ MB
```

```
data_purchase['PROD_QTY'].value_counts()
```

```
PROD_QTY
2      236039
1       27518
5         450
3         430
4         397
200         2
Name: count, dtype: int64
```

```
data_purchase['PROD_NAME'] =
data_purchase['PROD_NAME'].str.replace('Chp', 'Chip', case=False)
data_chip =
data_purchase[data_purchase['PROD_NAME'].str.contains('Chip')]
data_chip
```

| | DATE | STORE_NBR | LYLTY_CARD_NBR | TXN_ID | PROD_NBR | \ |
|---|---|---|---|---|---|---|
| 0 | 2018-10-17 | 1 | 1000 | 1 | 5 | |
| 2 | 2019-05-20 | 1 | 1343 | 383 | 61 | |
| 3 | 2018-08-17 | 2 | 2373 | 974 | 69 | |
| 4 | 2018-08-18 | 2 | 2426 | 1038 | 108 | |
| 6 | 2019-05-16 | 4 | 4149 | 3333 | 16 | |
| ... | ... | ... | ... | ... | ... | |
| 264816 | 2019-05-29 | 271 | 271193 | 269366 | 75 | |
| 264824 | 2019-03-13 | 272 | 272193 | 269906 | 9 | |
| 264826 | 2019-03-25 | 272 | 272194 | 269908 | 75 | |
| 264830 | 2018-11-12 | 272 | 272319 | 270087 | 44 | |
| 264834 | 2018-12-27 | 272 | 272379 | 270188 | 42 | |

| | PROD_NAME | PROD_QTY | TOT_SALES |
|---|---|---|---|
| 0 | Natural Chip Compny SeaSalt175g | 2 | 6.0 |
| 2 | Smiths Crinkle Cut Chips Chicken 170g | 2 | 2.9 |
| 3 | Smiths Chip Thinly S/Cream&Onion 175g | 5 | 15.0 |
| 4 | Kettle Tortilla ChipsHny&Jlpno Chili 150g | 3 | 13.8 |
| 6 | Smiths Crinkle Chips Salt & Vinegar 330g | 1 | 5.7 |
| ... | ... | ... | ... |
| 264816 | Cobs Popd Sea Salt Chips 110g | 2 | 7.6 |
| 264824 | Kettle Tortilla ChipsBtroot&Ricotta 150g | 1 | 4.6 |

```
264826             Cobs Popd Sea Salt  Chips 110g          2         7.6

264830             Thins Chips Light&  Tangy 175g          2         6.6

264834    Doritos Corn Chip Mexican Jalapeno 150g          2         7.8


[87335 rows x 8 columns]
```

```
data_chip['PROD_QTY'].value_counts()
```

```
PROD_QTY
2      77921
1       9011
5        147
3        140
4        114
200        2
Name: count, dtype: int64
```

```
data_chip
```

```
             DATE  STORE_NBR  LYLTY_CARD_NBR  TXN_ID  PROD_NBR  \
0      2018-10-17          1            1000       1         5
2      2019-05-20          1            1343     383        61
3      2018-08-17          2            2373     974        69
4      2018-08-18          2            2426    1038       108
6      2019-05-16          4            4149    3333        16
...           ...        ...             ...     ...       ...
264816 2019-05-29        271          271193  269366        75
264824 2019-03-13        272          272193  269906         9
264826 2019-03-25        272          272194  269908        75
264830 2018-11-12        272          272319  270087        44
264834 2018-12-27        272          272379  270188        42

                                        PROD_NAME  PROD_QTY  TOT_SALES
0          Natural Chip        Compny SeaSalt175g         2        6.0

2          Smiths Crinkle Cut  Chips Chicken 170g         2        2.9

3          Smiths Chip Thinly  S/Cream&Onion 175g         5       15.0

4       Kettle Tortilla ChipsHny&Jlpno Chili 150g         3       13.8

6        Smiths Crinkle Chips Salt & Vinegar 330g         1        5.7

...                                           ...       ...        ...

264816             Cobs Popd Sea Salt  Chips 110g         2        7.6
```

```
264824   Kettle Tortilla ChipsBtroot&Ricotta 150g          1        4.6

264826              Cobs Popd Sea Salt  Chips 110g          2        7.6

264830              Thins Chips Light&  Tangy 175g          2        6.6

264834    Doritos Corn Chip Mexican Jalapeno 150g          2        7.8


[87335 rows x 8 columns]
```

```
data_chip['PROD_NAME'].value_counts()
```

```
PROD_NAME
Kettle Tortilla ChipsHny&Jlpno Chili 150g    3296
Cobs Popd Swt/Chlli &Sr/Cream Chips 110g     3269
Cobs Popd Sea Salt  Chips 110g               3265
Smiths Crnkle Chip  Orgnl Big Bag 380g       3233
Thins Potato Chips  Hot & Spicy 175g         3229
Doritos Corn Chips  Cheese Supreme 170g      3217
Doritos Corn Chip Mexican Jalapeno 150g      3204
Smiths Crinkle Chips Salt & Vinegar 330g     3197
Thins Chips Light&  Tangy 175g               3188
Dorito Corn Chip     Supreme 380g            3185
Doritos Corn Chip Southern Chicken 150g      3172
Doritos Corn Chips  Nacho Cheese 170g        3160
Cobs Popd Sour Crm  &Chives Chips 110g       3159
Kettle Tortilla ChipsBtroot&Ricotta 150g     3146
Tostitos Smoked     Chipotle 175g            3145
Kettle Tortilla ChipsFeta&Garlic 150g        3138
Doritos Corn Chips  Original 170g            3121
Thins Chips Seasonedchicken 175g             3114
Thins Chips Salt &  Vinegar 175g             3103
Smiths Chip Thinly  Cut Original 175g        1614
Natural Chip Co     Tmato Hrb&Spce 175g      1572
Natural ChipCo Sea  Salt & Vinegr 175g       1550
WW Supreme Cheese   Corn Chips 200g          1509
WW Original Corn    Chips 200g               1495
Smiths Crinkle Cut  Chips Barbecue 170g      1489
WW Original Stacked Chips 160g               1487
Smiths Crinkle Cut  Chips Chicken 170g       1484
WW Sour Cream &OnionStacked Chips 160g       1483
Smiths Crinkle Cut  Chips Chs&Onion170g      1481
Smiths Chip Thinly  S/Cream&Onion 175g       1473
WW D/Style Chip     Sea Salt 200g            1469
Natural Chip        Compny SeaSalt175g       1468
Smiths Crinkle Cut  Chips Original 170g      1461
Natural ChipCo      Hony Soy Chckn175g       1460
Thins Chips         Originl saltd 175g       1441
Smiths Chip Thinly  CutSalt/Vinegr175g       1440
```

```
French Fries Potato Chips 175g               1418
Name: count, dtype: int64
```

```
data_chip.describe()
```

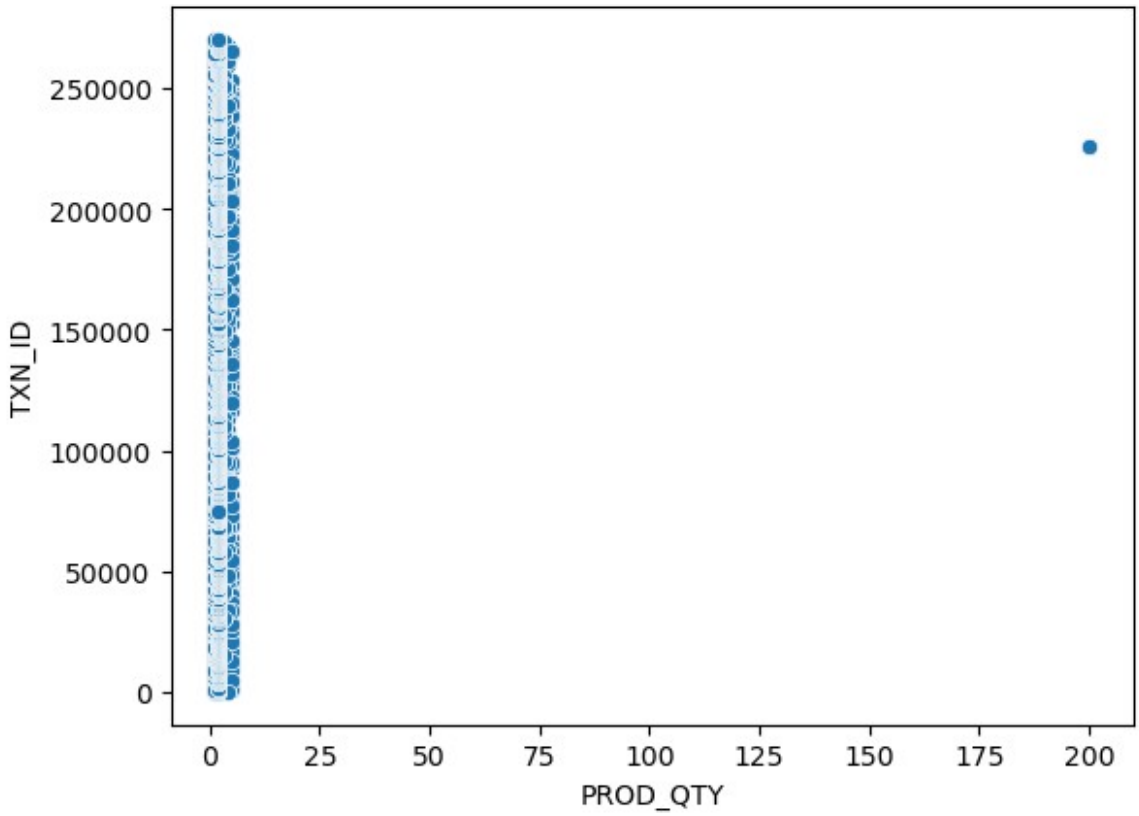
```
                                DATE     STORE_NBR  LYLTY_CARD_NBR  \
count                          87335  87335.000000    8.733500e+04
mean   2018-12-29 21:06:07.893742336    135.525185    1.360239e+05
min              2018-07-01 00:00:00      1.000000    1.000000e+03
25%              2018-09-30 00:00:00     70.000000    7.010000e+04
50%              2018-12-29 00:00:00    131.000000    1.314830e+05
75%              2019-03-31 00:00:00    203.000000    2.032830e+05
max              2019-06-30 00:00:00    272.000000    2.373711e+06
std                              NaN     76.853472    8.095134e+04

              TXN_ID      PROD_NBR      PROD_QTY     TOT_SALES
count   87335.000000  87335.000000  87335.000000  87335.000000
mean   135604.567127     49.949619      1.910620      7.286420
min         1.000000      1.000000      1.000000      1.900000
25%     68056.500000     16.000000      2.000000      6.000000
50%    135822.000000     42.000000      2.000000      6.600000
75%    203488.000000     78.000000      2.000000      8.800000
max    270209.000000    111.000000    200.000000    650.000000
std     78071.398801     34.450103      1.007451      3.943887
```

```
data_chip['PROD_QTY'].value_counts()
```

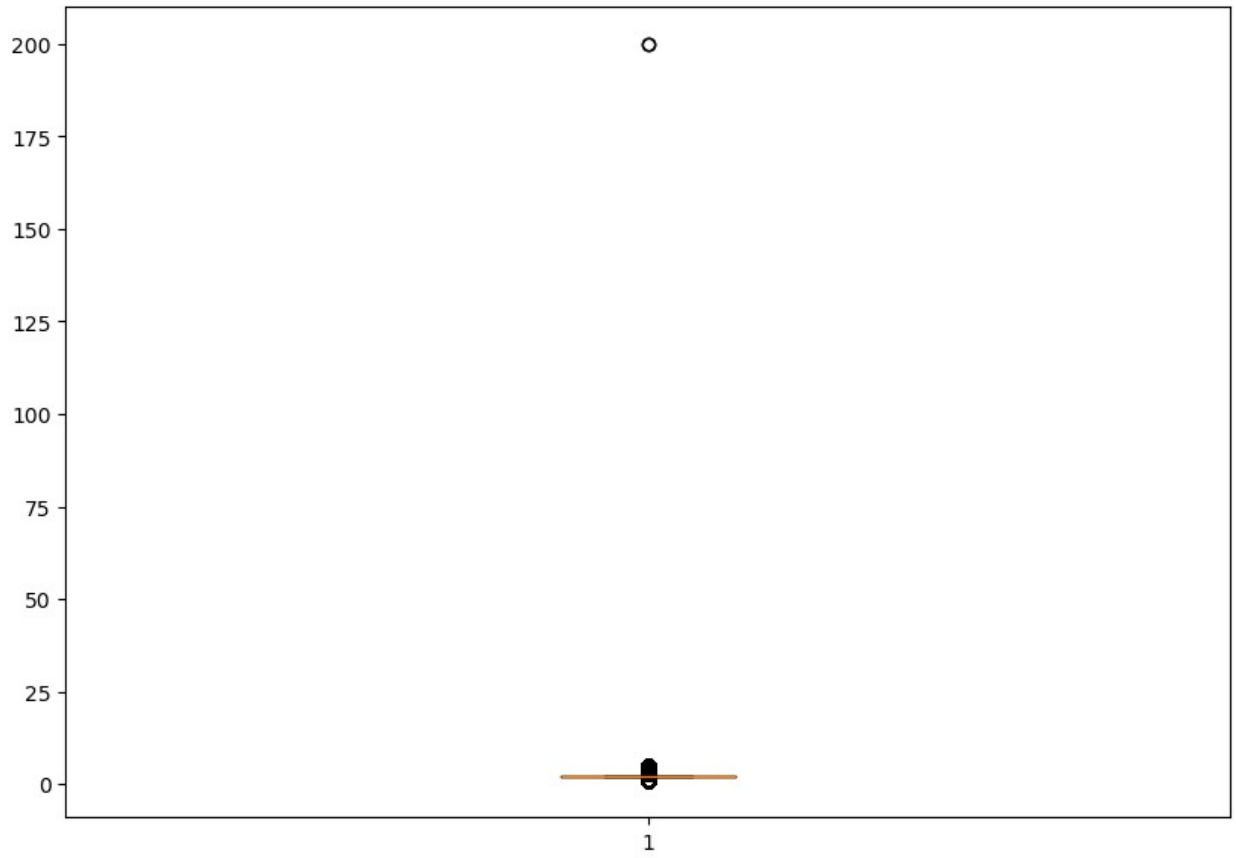
```
PROD_QTY
2      77921
1       9011
5        147
3        140
4        114
200        2
Name: count, dtype: int64
```

```
sns.scatterplot(x=data_chip['PROD_QTY'], y=data_chip['TXN_ID'])
```

```
<Axes: xlabel='PROD_QTY', ylabel='TXN_ID'>
```




```
fig = plt.figure(figsize =(10, 7))
plt.boxplot(data_chip['PROD_QTY'])
plt.show()
```




Using boxplot and scatterplot we can see the outlier in PROD_QTY

```
data_chip[data_chip['PROD_QTY']==200]
```

```
            DATE  STORE_NBR  LYLTY_CARD_NBR  TXN_ID  PROD_NBR  \
69762 2018-08-19        226          226000  226201         4
69763 2019-05-20        226          226000  226210         4

                                PROD_NAME  PROD_QTY  TOT_SALES
69762  Dorito Corn Chip     Supreme 380g       200      650.0
69763  Dorito Corn Chip     Supreme 380g       200      650.0
```

Two transactions made for qty 200

```
data_chip.drop([69762, 69763], inplace=True)
data_chip
```

```
C:\Users\suskr\AppData\Local\Temp\ipykernel_16592\911433495.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
```

```
returning-a-view-versus-a-copy
  data_chip.drop([69762, 69763], inplace=True)
```

```
             DATE  STORE_NBR  LYLTY_CARD_NBR  TXN_ID  PROD_NBR  \
0      2018-10-17          1            1000       1         5
2      2019-05-20          1            1343     383        61
3      2018-08-17          2            2373     974        69
4      2018-08-18          2            2426    1038       108
6      2019-05-16          4            4149    3333        16
...           ...        ...             ...     ...       ...
264816 2019-05-29        271          271193  269366        75
264824 2019-03-13        272          272193  269906         9
264826 2019-03-25        272          272194  269908        75
264830 2018-11-12        272          272319  270087        44
264834 2018-12-27        272          272379  270188        42

                                        PROD_NAME  PROD_QTY  TOT_SALES
0          Natural Chip        Compny SeaSalt175g         2        6.0
2          Smiths Crinkle Cut  Chips Chicken 170g         2        2.9
3          Smiths Chip Thinly  S/Cream&Onion 175g         5       15.0
4       Kettle Tortilla ChipsHny&Jlpno Chili 150g         3       13.8
6        Smiths Crinkle Chips Salt & Vinegar 330g         1        5.7
...                                           ...       ...        ...
264816             Cobs Popd Sea Salt  Chips 110g         2        7.6
264824   Kettle Tortilla ChipsBtroot&Ricotta 150g         1        4.6
264826             Cobs Popd Sea Salt  Chips 110g         2        7.6
264830             Thins Chips Light&  Tangy 175g         2        6.6
264834    Doritos Corn Chip Mexican Jalapeno 150g         2        7.8

[87333 rows x 8 columns]
```

```
data_chip[data_chip['LYLTY_CARD_NBR']==226000]
```

```
Empty DataFrame
Columns: [DATE, STORE_NBR, LYLTY_CARD_NBR, TXN_ID, PROD_NBR,
PROD_NAME, PROD_QTY, TOT_SALES]
Index: []
```

Removed the loyalty card number for further analysis

```
data_chip[data_chip['PROD_NAME']=='salsa']
```

```
Empty DataFrame
Columns: [DATE, STORE_NBR, LYLTY_CARD_NBR, TXN_ID, PROD_NBR,
PROD_NAME, PROD_QTY, TOT_SALES]
Index: []
```

```
data_grouped = data_chip.groupby('DATE').count().reset_index()
data_grouped[['DATE', 'TXN_ID']]
```

```
          DATE  TXN_ID
0   2018-07-01     198
1   2018-07-02     224
2   2018-07-03     238
3   2018-07-04     225
4   2018-07-05     243
..         ...     ...
359 2019-06-26     223
360 2019-06-27     212
361 2019-06-28     272
362 2019-06-29     247
363 2019-06-30     252

[364 rows x 2 columns]
```

```
data_grouped['month'] = data_grouped['DATE'].dt.year
data_grouped.drop(columns='month', inplace=True)
```

```
data_grouped
```

```
          DATE  STORE_NBR  LYLTY_CARD_NBR  TXN_ID  PROD_NBR  PROD_NAME
\
0   2018-07-01        198             198     198       198        198

1   2018-07-02        224             224     224       224        224

2   2018-07-03        238             238     238       238        238

3   2018-07-04        225             225     225       225        225

4   2018-07-05        243             243     243       243        243

..         ...        ...             ...     ...       ...        ...

359 2019-06-26        223             223     223       223        223

360 2019-06-27        212             212     212       212        212

361 2019-06-28        272             272     272       272        272
```

```
362 2019-06-29        247             247     247        247         247

363 2019-06-30        252             252     252        252         252


     PROD_QTY  TOT_SALES
0         198        198
1         224        224
2         238        238
3         225        225
4         243        243
..        ...        ...
359       223        223
360       212        212
361       272        272
362       247        247
363       252        252

[364 rows x 8 columns]
```

```
data_grouped['MONTH'] = data_grouped['DATE'].dt.month
data_grouped
```

```
           DATE  STORE_NBR  LYLTY_CARD_NBR  TXN_ID  PROD_NBR  PROD_NAME
\
0    2018-07-01        198             198     198        198         198

1    2018-07-02        224             224     224        224         224

2    2018-07-03        238             238     238        238         238

3    2018-07-04        225             225     225        225         225

4    2018-07-05        243             243     243        243         243

..          ...        ...             ...     ...        ...         ...

359 2019-06-26        223             223     223        223         223

360 2019-06-27        212             212     212        212         212

361 2019-06-28        272             272     272        272         272

362 2019-06-29        247             247     247        247         247

363 2019-06-30        252             252     252        252         252


     PROD_QTY  TOT_SALES  MONTH
0         198        198      7
1         224        224      7
```

```
2         238         238      7
3         225         225      7
4         243         243      7
..        ...         ...    ...
359       223         223      6
360       212         212      6
361       272         272      6
362       247         247      6
363       252         252      6

[364 rows x 9 columns]
```

```
data_grouped['YEAR'] = data_grouped['DATE'].dt.year
data_grouped
```

```
           DATE  STORE_NBR  LYLTY_CARD_NBR  TXN_ID  PROD_NBR  PROD_NAME
\
0    2018-07-01        198             198     198       198        198

1    2018-07-02        224             224     224       224        224

2    2018-07-03        238             238     238       238        238

3    2018-07-04        225             225     225       225        225

4    2018-07-05        243             243     243       243        243

..          ...        ...             ...     ...       ...        ...

359 2019-06-26        223             223     223       223        223

360 2019-06-27        212             212     212       212        212

361 2019-06-28        272             272     272       272        272

362 2019-06-29        247             247     247       247        247

363 2019-06-30        252             252     252       252        252


     PROD_QTY  TOT_SALES  MONTH  YEAR
0         198        198      7  2018
1         224        224      7  2018
2         238        238      7  2018
3         225        225      7  2018
4         243        243      7  2018
..        ...        ...    ...   ...
359       223        223      6  2019
360       212        212      6  2019
361       272        272      6  2019
362       247        247      6  2019
```

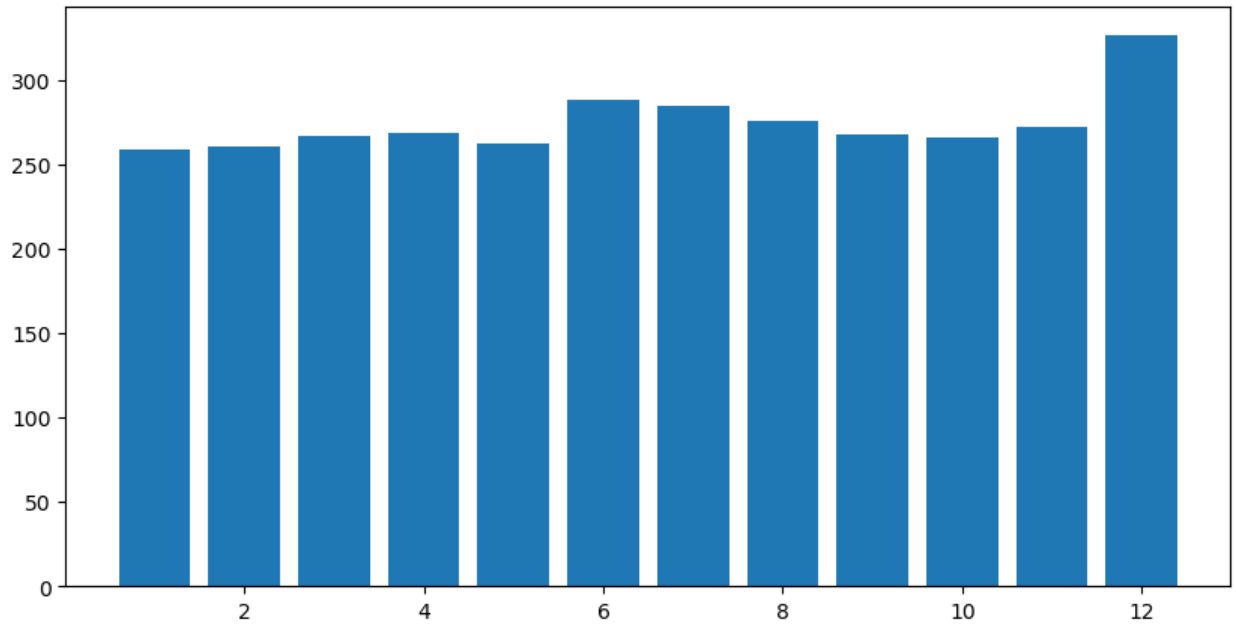
```
363       252        252      6  2019

[364 rows x 10 columns]
```

```
data_grouped['DAY'] = data_grouped['DATE'].dt.day
data_grouped
```

```
          DATE  STORE_NBR  LYLTY_CARD_NBR  TXN_ID  PROD_NBR  PROD_NAME
\
0   2018-07-01        198             198     198       198        198

1   2018-07-02        224             224     224       224        224

2   2018-07-03        238             238     238       238        238

3   2018-07-04        225             225     225       225        225

4   2018-07-05        243             243     243       243        243

..         ...        ...             ...     ...       ...        ...

359 2019-06-26        223             223     223       223        223

360 2019-06-27        212             212     212       212        212

361 2019-06-28        272             272     272       272        272

362 2019-06-29        247             247     247       247        247

363 2019-06-30        252             252     252       252        252


     PROD_QTY  TOT_SALES  MONTH  YEAR  DAY
0         198        198      7  2018    1
1         224        224      7  2018    2
2         238        238      7  2018    3
3         225        225      7  2018    4
4         243        243      7  2018    5
..        ...        ...    ...   ...  ...
359       223        223      6  2019   26
360       212        212      6  2019   27
361       272        272      6  2019   28
362       247        247      6  2019   29
363       252        252      6  2019   30

[364 rows x 11 columns]
```

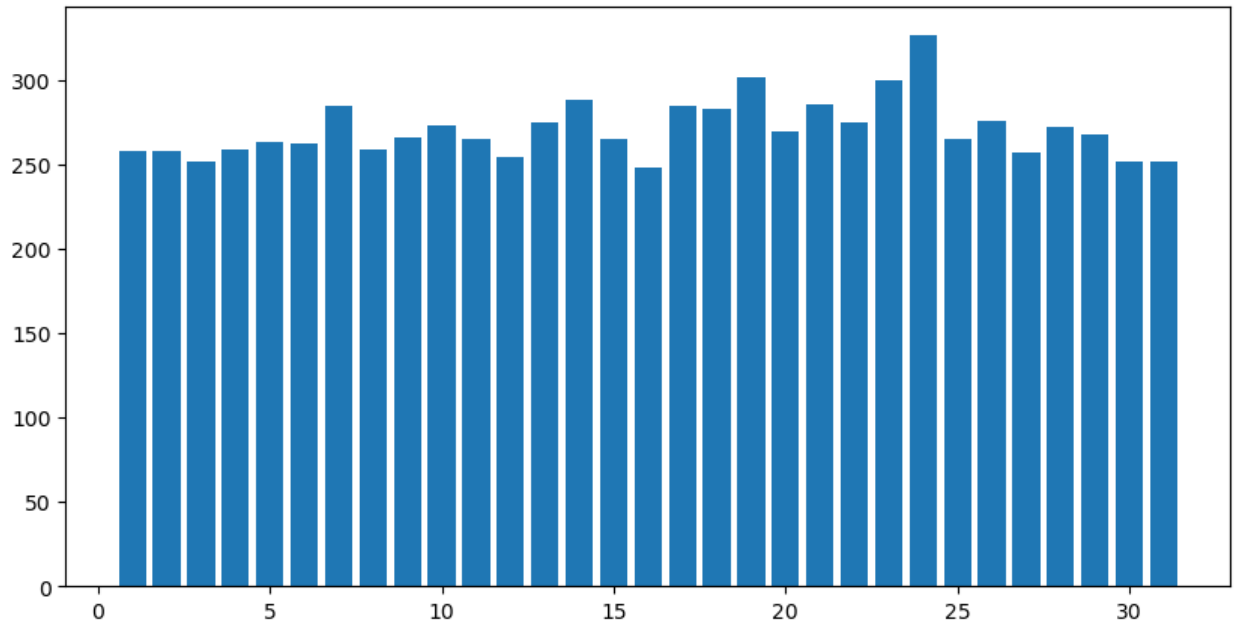
```
fig = plt.figure(figsize = (10, 5))
plt.bar(data_grouped['MONTH'], data_grouped['TXN_ID'])
```

```
<BarContainer object of 364 artists>
```




```
fig = plt.figure(figsize = (10, 5))
plt.bar(data_grouped['DAY'], data_grouped['TXN_ID'])
```

```
<BarContainer object of 364 artists>
```




We can see in december month around christmas we have max sales

```
data_chip
```

```
             DATE  STORE_NBR  LYLTY_CARD_NBR  TXN_ID  PROD_NBR  \
0      2018-10-17          1            1000       1         5
```

```
2      2019-05-20          1            1343     383        61
3      2018-08-17          2            2373     974        69
4      2018-08-18          2            2426    1038       108
6      2019-05-16          4            4149    3333        16
...           ...        ...             ...     ...       ...
264816 2019-05-29        271          271193  269366        75
264824 2019-03-13        272          272193  269906         9
264826 2019-03-25        272          272194  269908        75
264830 2018-11-12        272          272319  270087        44
264834 2018-12-27        272          272379  270188        42

                                       PROD_NAME  PROD_QTY  TOT_SALES

0          Natural Chip        Compny SeaSalt175g         2        6.0

2          Smiths Crinkle Cut  Chips Chicken 170g         2        2.9

3          Smiths Chip Thinly  S/Cream&Onion 175g         5       15.0

4       Kettle Tortilla ChipsHny&Jlpno Chili 150g         3       13.8

6        Smiths Crinkle Chips Salt & Vinegar 330g         1        5.7

...                                        ...       ...        ...

264816             Cobs Popd Sea Salt  Chips 110g         2        7.6

264824   Kettle Tortilla ChipsBtroot&Ricotta 150g         1        4.6

264826             Cobs Popd Sea Salt  Chips 110g         2        7.6

264830             Thins Chips Light&  Tangy 175g         2        6.6

264834    Doritos Corn Chip Mexican Jalapeno 150g         2        7.8


[87333 rows x 8 columns]
```

```
data_chip['SIZE'] = data_chip['PROD_NAME'].str[-4:]
data_chip
```

```
C:\Users\suskr\AppData\Local\Temp\ipykernel_16592\3072499435.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  data_chip['SIZE'] = data_chip['PROD_NAME'].str[-4:]
```

```
             DATE  STORE_NBR  LYLTY_CARD_NBR  TXN_ID  PROD_NBR  \
0      2018-10-17          1            1000       1         5
2      2019-05-20          1            1343     383        61
3      2018-08-17          2            2373     974        69
4      2018-08-18          2            2426    1038       108
6      2019-05-16          4            4149    3333        16
...           ...        ...             ...     ...       ...
264816 2019-05-29        271          271193  269366        75
264824 2019-03-13        272          272193  269906         9
264826 2019-03-25        272          272194  269908        75
264830 2018-11-12        272          272319  270087        44
264834 2018-12-27        272          272379  270188        42

                                        PROD_NAME  PROD_QTY  TOT_SALES
SIZE
0          Natural Chip        Compny SeaSalt175g         2        6.0
175g
2          Smiths Crinkle Cut  Chips Chicken 170g         2        2.9
170g
3          Smiths Chip Thinly  S/Cream&Onion 175g         5       15.0
175g
4       Kettle Tortilla ChipsHny&Jlpno Chili 150g         3       13.8
150g
6        Smiths Crinkle Chips Salt & Vinegar 330g         1        5.7
330g
...                                           ...       ...        ...
...
264816             Cobs Popd Sea Salt  Chips 110g         2        7.6
110g
264824   Kettle Tortilla ChipsBtroot&Ricotta 150g         1        4.6
150g
264826             Cobs Popd Sea Salt  Chips 110g         2        7.6
110g
264830             Thins Chips Light&  Tangy 175g         2        6.6
175g
264834    Doritos Corn Chip Mexican Jalapeno 150g         2        7.8
150g

[87333 rows x 9 columns]
```

```
data_chip['PROD_NAME'] = data_chip['PROD_NAME'].str[:-4]
data_chip
```

```
C:\Users\suskr\AppData\Local\Temp\ipykernel_16592\3038674784.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
```

```
returning-a-view-versus-a-copy
  data_chip['PROD_NAME'] = data_chip['PROD_NAME'].str[:-4]
```

```
             DATE  STORE_NBR  LYLTY_CARD_NBR  TXN_ID  PROD_NBR  \
0      2018-10-17          1            1000       1         5
2      2019-05-20          1            1343     383        61
3      2018-08-17          2            2373     974        69
4      2018-08-18          2            2426    1038       108
6      2019-05-16          4            4149    3333        16
...           ...        ...             ...     ...       ...
264816 2019-05-29        271          271193  269366        75
264824 2019-03-13        272          272193  269906         9
264826 2019-03-25        272          272194  269908        75
264830 2018-11-12        272          272319  270087        44
264834 2018-12-27        272          272379  270188        42

                                    PROD_NAME  PROD_QTY  TOT_SALES
SIZE
0          Natural Chip        Compny SeaSalt         2        6.0
175g
2          Smiths Crinkle Cut  Chips Chicken          2        2.9
170g
3          Smiths Chip Thinly  S/Cream&Onion          5       15.0
175g
4       Kettle Tortilla ChipsHny&Jlpno Chili          3       13.8
150g
6        Smiths Crinkle Chips Salt & Vinegar          1        5.7
330g
...                                       ...       ...        ...   .
..
264816             Cobs Popd Sea Salt  Chips          2        7.6
110g
264824   Kettle Tortilla ChipsBtroot&Ricotta          1        4.6
150g
264826             Cobs Popd Sea Salt  Chips          2        7.6
110g
264830             Thins Chips Light&  Tangy          2        6.6
175g
264834    Doritos Corn Chip Mexican Jalapeno          2        7.8
150g

[87333 rows x 9 columns]
```

```
data_chip['SIZE'].value_counts()
```

```
SIZE
175    29215
150    15956
170    15413
110     9693
```

```
380     6416
200     4473
330     3197
160     2970
Name: count, dtype: int64
```

```
data_chip['SIZE'] = data_chip['SIZE'].str[:-1]
data_chip
```

```
C:\Users\suskr\AppData\Local\Temp\ipykernel_16592\1766603709.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  data_chip['SIZE'] = data_chip['SIZE'].str[:-1]
```

| | DATE | STORE_NBR | LYLTY_CARD_NBR | TXN_ID | PROD_NBR | PROD_NAME | PROD_QTY | TOT_SALES | SIZE |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 2018-10-17 | 1 | 1000 | 1 | 5 | Natural Chip Compny SeaSalt | 2 | 6.0 | 175 |
| 2 | 2019-05-20 | 1 | 1343 | 383 | 61 | Smiths Crinkle Cut Chips Chicken | 2 | 2.9 | 170 |
| 3 | 2018-08-17 | 2 | 2373 | 974 | 69 | Smiths Chip Thinly S/Cream&Onion | 5 | 15.0 | 175 |
| 4 | 2018-08-18 | 2 | 2426 | 1038 | 108 | Kettle Tortilla ChipsHny&Jlpno Chili | 3 | 13.8 | 150 |
| 6 | 2019-05-16 | 4 | 4149 | 3333 | 16 | Smiths Crinkle Chips Salt & Vinegar | 1 | 5.7 | 330 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 264816 | 2019-05-29 | 271 | 271193 | 269366 | 75 | Cobs Popd Sea Salt Chips | 2 | 7.6 | 110 |
| 264824 | 2019-03-13 | 272 | 272193 | 269906 | 9 | Kettle Tortilla ChipsBtroot&Ricotta | 1 | 4.6 | 150 |
| 264826 | 2019-03-25 | 272 | 272194 | 269908 | 75 | | | | |
| 264830 | 2018-11-12 | 272 | 272319 | 270087 | 44 | | | | |
| 264834 | 2018-12-27 | 272 | 272379 | 270188 | 42 | | | | |

```
264826             Cobs Popd Sea Salt  Chips          2        7.6
110
264830             Thins Chips Light&  Tangy          2        6.6
175
264834    Doritos Corn Chip Mexican Jalapeno          2        7.8
150

[87333 rows x 9 columns]
```

```
data_chip['SIZE'].value_counts()
```

```
SIZE
175    29215
150    15956
170    15413
110     9693
380     6416
200     4473
330     3197
160     2970
Name: count, dtype: int64
```

```
data_chip['SIZE'] = data_chip['SIZE'].apply(pd.to_numeric)
data_chip.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 87333 entries, 0 to 264834
Data columns (total 9 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   DATE            87333 non-null  datetime64[ns]
 1   STORE_NBR       87333 non-null  int64
 2   LYLTY_CARD_NBR  87333 non-null  int64
 3   TXN_ID          87333 non-null  int64
 4   PROD_NBR        87333 non-null  int64
 5   PROD_NAME       87333 non-null  object
 6   PROD_QTY        87333 non-null  int64
 7   TOT_SALES       87333 non-null  float64
 8   SIZE            87333 non-null  int64
dtypes: datetime64[ns](1), float64(1), int64(6), object(1)
memory usage: 6.7+ MB

C:\Users\suskr\AppData\Local\Temp\ipykernel_16592\3428170374.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  data_chip['SIZE'] = data_chip['SIZE'].apply(pd.to_numeric)
```

## Changed SIZE column to numeric

```
data_chip['SIZE'].describe()
```

```
count    87333.000000
mean       183.840644
std         67.201016
min        110.000000
25%        150.000000
50%        170.000000
75%        175.000000
max        380.000000
Name: SIZE, dtype: float64
```

## MIN - 110 and MAX - 380

```
x = data_chip['SIZE']
plt.hist(x)
```

```
(array([ 9693., 18926., 44628.,  4473.,     0.,     0.,     0.,
0.,
         3197.,  6416.]),
 array([110., 137., 164., 191., 218., 245., 272., 299., 326., 353.,
380.]),
 <BarContainer object of 10 artists>)
```