

```
In [ ]: import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
import re
```

```
In [ ]: #  
pd.DataFrame
```

```
Out[ ]: pandas.core.frame.DataFrame
```

```
In [ ]: df= pd.read_csv(r'A:\Data Analytics\Forage)Quantum\QVI_transaction_data.csv')
```

```
In [ ]: df.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  object    
 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: float64(1), int64(5), object(2)  
memory usage: 16.2+ MB
```

```
In [ ]: df[ 'DATE' ] = pd.to_datetime(df[ 'DATE' ])
```

```
In [ ]: df[ 'PROD_NAME' ].unique()
```

```
Out[ ]: array(['Natural Chip      Comnpy SeaSalt175g',
   'CCs Nacho Cheese    175g',
   'Smiths Crinkle Cut Chips Chicken 170g',
   'Smiths Chip Thinly S/Cream&Onion 175g',
   'Kettle Tortilla ChpsHny&Jlpno Chili 150g',
   'Old El Paso Salsa   Dip Tomato Mild 300g',
   'Smiths Crinkle Chips Salt & Vinegar 330g',
   'Grain Waves         Sweet Chilli 210g',
   'Doritos Corn Chip Mexican Jalapeno 150g',
   'Grain Waves Sour   Cream&Chives 210G',
   'Kettle Sensations  Siracha Lime 150g',
   'Twisties Cheese     270g', 'WW Crinkle Cut     Chicken 175g',
   'Thins Chips Light& Tangy 175g', 'CCs Original 175g',
   'Burger Rings 220g', 'NCC Sour Cream & Garden Chives 175g',
   'Doritos Corn Chip Southern Chicken 150g',
   'Cheezels Cheese Box 125g', 'Smiths Crinkle     Original 330g',
   'Infzns Crn Crnchers Tangy Gcamole 110g',
   'Kettle Sea Salt     And Vinegar 175g',
   'Smiths Chip Thinly Cut Original 175g', 'Kettle Original 175g',
   'Red Rock Deli Thai Chilli&Lime 150g',
   'Pringles Sthrn FriedChicken 134g', 'Pringles Sweet&Spicy BBQ 134g',
   'Red Rock Deli SR    Salsa & Mzzrla 150g',
   'Thins Chips          Originl saltd 175g',
   'Red Rock Deli Sp    Salt & Truffle 150G',
   'Smiths Thinly        Swt Chli&S/Cream175G', 'Kettle Chilli 175g',
   'Doritos Mexicana   170g',
   'Smiths Crinkle Cut French OnionDip 150g',
   'Natural ChipCo      Hony Soy Chckn175g',
   'Dorito Corn Chp     Supreme 380g', 'Twisties Chicken270g',
   'Smiths Thinly Cut   Roast Chicken 175g',
   'Smiths Crinkle Cut Tomato Salsa 150g',
   'Kettle Mozzarella  Basil & Pesto 175g',
   'Infuzions Thai SweetChili PotatoMix 110g',
   'Kettle Sensations  Camembert & Fig 150g',
   'Smith Crinkle Cut   Mac N Cheese 150g',
   'Kettle Honey Soy    Chicken 175g',
   'Thins Chips Seasonedchicken 175g',
   'Smiths Crinkle Cut Salt & Vinegar 170g',
   'Infuzions BBQ Rib   Prawn Crackers 110g',
   'Grnwves Plus Btroot & Chilli Jam 180g',
   'Tyrrells Crisps     Lightly Salted 165g',
   'Kettle Sweet Chilli And Sour Cream 175g',
   'Doritos Salsa       Medium 300g', 'Kettle 135g Swt Pot Sea Salt',
   'Pringles SourCream  Onion 134g',
   'Doritos Corn Chips  Original 170g',
   'Twisties Cheese      Burger 250g',
   'Old El Paso Salsa   Dip Chnky Tom Ht300g',
   'Cobs Popd Swt/Chlli &Sr/Cream Chips 110g',
   'Woolworths Mild     Salsa 300g',
   'Natural Chip Co     Tmato Hrb&Spce 175g',
   'Smiths Crinkle Cut Chips Original 170g',
   'Cobs Popd Sea Salt  Chips 110g',
   'Smiths Crinkle Cut Chips Chs&Onion170g',
   'French Fries Potato Chips 175g',
   'Old El Paso Salsa   Dip Tomato Med 300g',
   'Doritos Corn Chips  Cheese Supreme 170g',
```

```
'Pringles Original    Crisps 134g',
'RRD Chilli&        Coconut 150g',
'WW Original Corn    Chips 200g',
'Thins Potato Chips  Hot & Spicy 175g',
'Cobs Popd Sour Crm &Chives Chips 110g',
'Smiths Crnkle Chip Orgnl Big Bag 380g',
'Doritos Corn Chips Nacho Cheese 170g',
'Kettle Sensations BBQ&Maple 150g',
'WW D/Style Chip     Sea Salt 200g',
'Pringles Chicken    Salt Crips 134g',
'WW Original Stacked Chips 160g',
'Smiths Chip Thinly CutSalt/Vinegr175g', 'Cheezels Cheese 330g',
'Tostitos Lightly   Salted 175g',
'Thins Chips Salt & Vinegar 175g',
'Smiths Crinkle Cut Chips Barbecue 170g', 'Cheetos Puffs 165g',
'RRD Sweet Chilli & Sour Cream 165g',
'WW Crinkle Cut     Original 175g',
'Tostitos Splash Of Lime 175g', 'Woolworths Medium Salsa 300g',
'Kettle Tortilla ChpsBtroot&Ricotta 150g',
'CCs Tasty Cheese   175g', 'Woolworths Cheese Rings 190g',
'Tostitos Smoked    Chipotle 175g', 'Pringles Barbeque 134g',
'WW Supreme Cheese  Corn Chips 200g',
'Pringles Mystery    Flavour 134g',
'Tyrrells Crisps    Ched & Chives 165g',
'Snbts Whlgrn Crisps Cheddr&Mstrd 90g',
'Cheetos Chs & Bacon Balls 190g', 'Pringles Slt Vingar 134g',
'Infuzions SourCream&Herbs Veg Strws 110g',
'Kettle Tortilla ChpsFeta&Garlic 150g',
'Infuzions Mango     Chutny Papadums 70g',
'RRD Steak &         Chimuchurri 150g',
'RRD Honey Soy       Chicken 165g',
'Sunbites Whlegrn    Crisps Frch/Onin 90g',
'RRD Salt & Vinegar 165g', 'Doritos Cheese Supreme 330g',
'Smiths Crinkle Cut Snag&Sauce 150g',
'WW Sour Cream &OnionStacked Chips 160g',
'RRD Lime & Pepper  165g',
'Natural ChipCo Sea Salt & Vinegr 175g',
'Red Rock Deli Chikn&Garlic Aioli 150g',
'RRD SR Slow Rst     Pork Belly 150g', 'RRD Pc Sea Salt 165g',
'Smith Crinkle Cut   Bolognese 150g', 'Doritos Salsa Mild 300g'],
dtype=object)
```

```
In [ ]: df['PDWT_grams'] = df['PROD_NAME'].replace('[a-zA-Z/&]', ' ', regex=True).astype('i')

In [ ]: df['PROD_NAME'] = df['PROD_NAME'].replace('[0-9]{2,3}[gG]$', ' ', regex=True).astype('str')

In [ ]: df['PROD_NAME'] = df['PROD_NAME'].replace('&', ' & ', regex=True).astype('str')

In [ ]: df['PROD_NAME'] = df['PROD_NAME'].replace(r'\s{2,4}', ' ', regex=True).astype('str')

In [ ]: df['PROD_NAME'] = df['PROD_NAME'].replace(r'[/*()_-]', '', regex=True).astype('str')

In [ ]: df['PROD_NAME'] = df['PROD_NAME'].str.strip()
```

```
In [ ]: df['PROD_NAME'].unique()
```

```
Out[ ]: array(['Natural Chip Comnpy SeaSalt', 'CCs Nacho Cheese',
   'Smiths Crinkle Cut Chips Chicken',
   'Smiths Chip Thinly SCream & Onion',
   'Kettle Tortilla ChpsHny & Jlpno Chili',
   'Old El Paso Salsa Dip Tomato Mild',
   'Smiths Crinkle Chips Salt & Vinegar',
   'Grain Waves Sweet Chilli', 'Doritos Corn Chip Mexican Jalapeno',
   'Grain Waves Sour Cream & Chives',
   'Kettle Sensations Siracha Lime', 'Twisties Cheese',
   'WW Crinkle Cut Chicken', 'Thins Chips Light & Tangy',
   'CCs Original', 'Burger Rings', 'NCC Sour Cream & Garden Chives',
   'Doritos Corn Chip Southern Chicken', 'Cheezels Cheese Box',
   'Smiths Crinkle Original', 'Infzns Crn Crnchers Tangy Gcamole',
   'Kettle Sea Salt And Vinegar', 'Smiths Chip Thinly Cut Original',
   'Kettle Original', 'Red Rock Deli Thai Chilli & Lime',
   'Pringles Sthrn FriedChicken', 'Pringles Sweet & Spcy BBQ',
   'Red Rock Deli SR Salsa & Mzzrla', 'Thins Chips Originl saltd',
   'Red Rock Deli Sp Salt & Truffle',
   'Smiths Thinly Swt Chli & SCream', 'Kettle Chilli',
   'Doritos Mexicana', 'Smiths Crinkle Cut French OnionDip',
   'Natural ChipCo Hony Soy Chckn', 'Dorito Corn Chp Supreme',
   'Twisties Chicken', 'Smiths Thinly Cut Roast Chicken',
   'Smiths Crinkle Cut Tomato Salsa',
   'Kettle Mozzarella Basil & Pesto',
   'Infuzions Thai SweetChili PotatoMix',
   'Kettle Sensations Camembert & Fig',
   'Smith Crinkle Cut Mac N Cheese', 'Kettle Honey Soy Chicken',
   'Thins Chips Seasonedchicken', 'Smiths Crinkle Cut Salt & Vinegar',
   'Infuzions BBQ Rib Prawn Crackers',
   'GnnWves Plus Btroot & Chilli Jam',
   'Tyrrells Crisps Lightly Salted',
   'Kettle Sweet Chilli And Sour Cream', 'Doritos Salsa Medium',
   'Kettle 135g Swt Pot Sea Salt', 'Pringles SourCream Onion',
   'Doritos Corn Chips Original', 'Twisties Cheese Burger',
   'Old El Paso Salsa Dip Chnky Tom Ht',
   'Cobs Popd SwtChlli & SrCream Chips', 'Woolworths Mild Salsa',
   'Natural Chip Co Tmato Hrb & Spce',
   'Smiths Crinkle Cut Chips Original', 'Cobs Popd Sea Salt Chips',
   'Smiths Crinkle Cut Chips Chs & Onion',
   'French Fries Potato Chips', 'Old El Paso Salsa Dip Tomato Med',
   'Doritos Corn Chips Cheese Supreme', 'Pringles Original Crisps',
   'RRD Chilli & Coconut', 'WW Original Corn Chips',
   'Thins Potato Chips Hot & Spicy',
   'Cobs Popd Sour Crm & Chives Chips',
   'Smiths Crnkle Chip Orgnl Big Bag',
   'Doritos Corn Chips Nacho Cheese', 'Kettle Sensations BBQ & Maple',
   'WW DStyle Chip Sea Salt', 'Pringles Chicken Salt Crips',
   'WW Original Stacked Chips', 'Smiths Chip Thinly CutSaltVinegr',
   'Cheezels Cheese', 'Tostitos Lightly Salted',
   'Thins Chips Salt & Vinegar', 'Smiths Crinkle Cut Chips Barbecue',
   'Cheetos Puffs', 'RRD Sweet Chilli & Sour Cream',
   'WW Crinkle Cut Original', 'Tostitos Splash Of Lime',
   'Woolworths Medium Salsa', 'Kettle Tortilla ChpsBtroot & Ricotta',
   'CCs Tasty Cheese', 'Woolworths Cheese Rings',
   'Tostitos Smoked Chipotle', 'Pringles Barbeque',
   'WW Supreme Cheese Corn Chips', 'Pringles Mystery Flavour'],
```

```
'Tyrrells Crisps Ched & Chives',
'Snbts Whlgrn Crisps Cheddr & Mstrd', 'Cheetos Chs & Bacon Balls',
'Pringles Slt Vingar', 'Infuzions SourCream & Herbs Veg Strws',
'Kettle Tortilla ChpsFeta & Garlic',
'Infuzions Mango Chutny Papadums', 'RRD Steak & Chimuchurri',
'RRD Honey Soy Chicken', 'Sunbites Whlegrn Crisps FrchOnin',
'RRD Salt & Vinegar', 'Doritos Cheese Supreme',
'Smiths Crinkle Cut Snag & Sauce',
'WW Sour Cream & OnionStacked Chips', 'RRD Lime & Pepper',
'Natural ChipCo Sea Salt & Vinegr',
'Red Rock Deli Chikn & Garlic Aioli',
'RRD SR Slow Rst Pork Belly', 'RRD Pc Sea Salt',
'Smith Crinkle Cut Bolognese', 'Doritos Salsa Mild'], dtype=object)
```

```
In [ ]: df.drop_duplicates(inplace=True)
```

```
In [ ]: df_sort = df['PROD_NAME'].value_counts().reset_index()
```

```
In [ ]: df = df.merge(df_sort, on='PROD_NAME').sort_values(by='count')
```

```
In [ ]: df = df.sort_values(by='count', ascending=False)
```

```
In [ ]: df.head()
```

| | | DATE | STORE_NBR | LYLTY_CARD_NBR | TXN_ID | PROD_NBR | PROD_NAME | PROD_C |
|---------------|--|------------|-----------|----------------|--------|----------|-----------|---------------------------------|
| 135054 | | 2019-06-09 | 271 | | 271234 | 269616 | 102 | Kettle Mozzarella Basil & Pesto |
| 240286 | | 2018-11-15 | 24 | | 24190 | 21376 | 102 | Kettle Mozzarella Basil & Pesto |
| 173501 | | 2018-10-20 | 26 | | 26318 | 23650 | 102 | Kettle Mozzarella Basil & Pesto |
| 93345 | | 2018-12-16 | 181 | | 181043 | 182743 | 102 | Kettle Mozzarella Basil & Pesto |
| 93307 | | 2019-01-21 | 180 | | 180226 | 182487 | 102 | Kettle Mozzarella Basil & Pesto |

```
In [ ]: df.drop(columns='count', inplace=True)
```

```
In [ ]: df[df['PROD_NAME'].str.contains('Salsa', regex=False)]
```

Out[]:

| | | DATE | STORE_NBR | LYLTY_CARD_NBR | TXN_ID | PROD_NBR | PROD_NAME | PROD_C |
|--------|------------|------|-----------|----------------|--------|----------|-----------|---|
| 71963 | 2019-02-20 | | 58 | | 58141 | 53435 | 65 | Old El Paso Salsa Dip Chnky Tom Ht |
| 249372 | 2018-07-06 | | 147 | | 147166 | 146720 | 65 | Old El Paso Salsa Dip Chnky Tom Ht |
| 2845 | 2019-06-11 | | 151 | | 151214 | 149924 | 65 | Old El Paso Salsa Dip Chnky Tom Ht |
| 187866 | 2019-04-24 | | 40 | | 40063 | 36416 | 65 | Old El Paso Salsa Dip Chnky Tom Ht |
| 225776 | 2018-07-03 | | 202 | | 202083 | 201749 | 65 | Old El Paso Salsa Dip Chnky Tom Ht |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 67075 | 2018-11-06 | | 222 | | 222198 | 222628 | 76 | Woolworths Medium Salsa |
| 168718 | 2018-12-17 | | 230 | | 230120 | 232728 | 76 | Woolworths Medium Salsa |
| 62194 | 2019-03-16 | | 122 | | 122066 | 125038 | 76 | Woolworths Medium Salsa |
| 24055 | 2019-06-13 | | 233 | | 233303 | 237074 | 76 | Woolworths Medium Salsa |
| 24059 | 2018-12-20 | | 233 | | 233444 | 237223 | 76 | Woolworths Medium Salsa |

18094 rows × 9 columns



In []: df.isna().value_counts()

```
Out[ ]: DATE STORE_NBR LYLTY_CARD_NBR TXN_ID PROD_NBR PROD_NAME PROD_QTY TOT_SALES
PDWT_grams
False False False False False False False False
False 264835
Name: count, dtype: int64
```

```
In [ ]: df[['PROD_QTY', 'TOT_SALES']].agg(['mean', 'max', 'min'])
```

```
Out[ ]: PROD_QTY TOT_SALES
mean 1.907308 7.304205
max 200.000000 650.000000
min 1.000000 1.500000
```

```
In [ ]: df[df['PROD_QTY'] == 200]
```

```
Out[ ]: DATE STORE_NBR LYLTY_CARD_NBR TXN_ID PROD_NBR PROD_NAME PROD_QTY
69763 2019-05-20 226 226000 226210 4 Dorito Corn Chp Supreme 20
69762 2018-08-19 226 226000 226201 4 Dorito Corn Chp Supreme 20
```

```
In [ ]: df[df['LYLTY_CARD_NBR'] == 226000]
```

```
Out[ ]: DATE STORE_NBR LYLTY_CARD_NBR TXN_ID PROD_NBR PROD_NAME PROD_QTY
69763 2019-05-20 226 226000 226210 4 Dorito Corn Chp Supreme 20
69762 2018-08-19 226 226000 226201 4 Dorito Corn Chp Supreme 20
```

```
In [ ]: df.drop(index=[69763, 69762], inplace=True)
```

```
In [ ]: df['DATE'].value_counts()
```

```
Out[ ]: DATE
2018-12-24    939
2018-12-23    917
2018-12-22    915
2018-12-19    906
2018-12-18    862
...
2018-12-12    664
2019-06-24    662
2019-06-13    659
2018-10-18    658
2018-11-25    648
Name: count, Length: 364, dtype: int64
```

```
In [ ]: dates = pd.date_range(start='2018-07-01', end='2019-06-30')
```

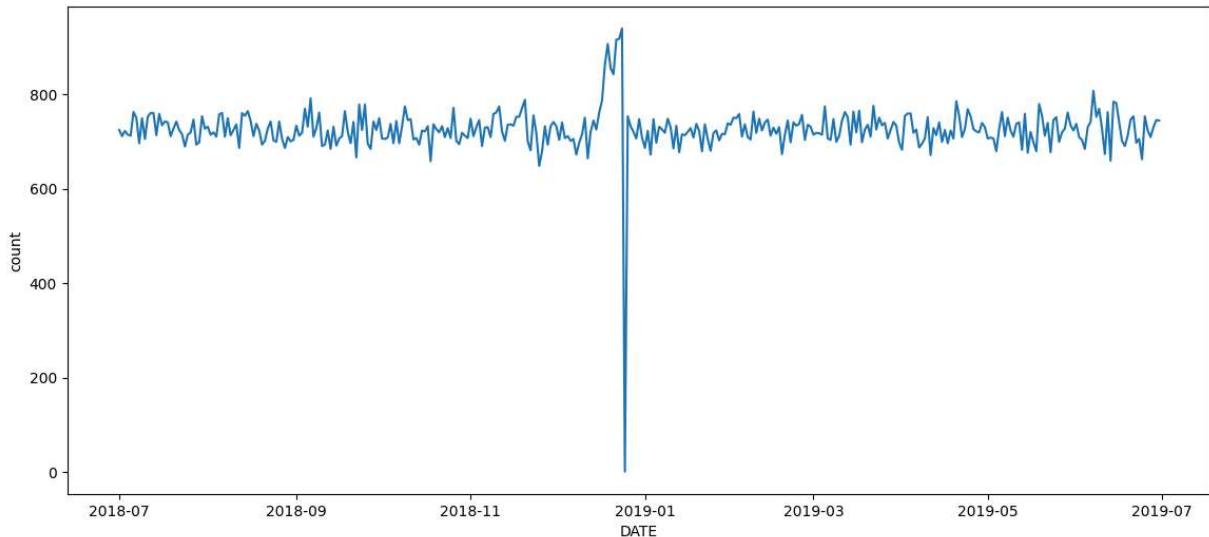
```
In [ ]: dates = pd.DataFrame(dates)
```

```
In [ ]: dates.columns = ['DATE']
```

```
In [ ]: df = df.merge(dates, on='DATE', how='outer')
```

```
In [ ]: plt.figure(figsize=(14,6))
sns.lineplot(data=df.value_counts(subset=df['DATE']).reset_index(), x='DATE', y='co
```

```
Out[ ]: <Axes: xlabel='DATE', ylabel='count'>
```

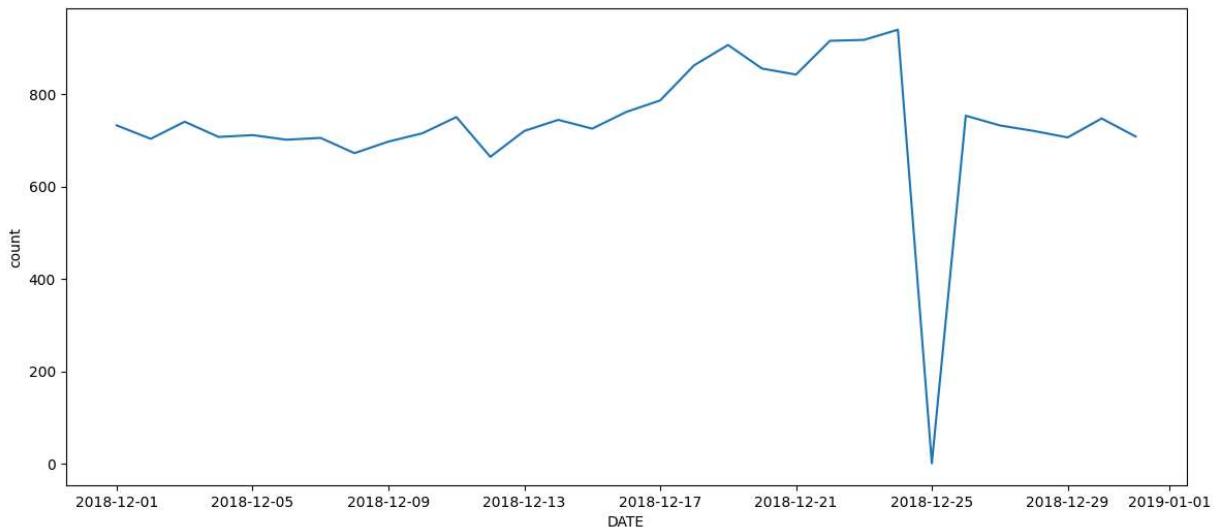


```
In [ ]: salecount = df.value_counts(subset=df['DATE']).reset_index()
```

```
In [ ]: sales_dec = salecount.query('"2018-12-01" <= DATE < "2019-01-01"').sort_values(by='
```

```
In [ ]: plt.figure(figsize=(14,6))
sns.lineplot(data=sales_dec, x='DATE', y='count')
```

```
Out[ ]: <Axes: xlabel='DATE', ylabel='count'>
```



```
In [ ]: df[['DATE', 'TOT_SALES']].groupby('DATE').sum()
```

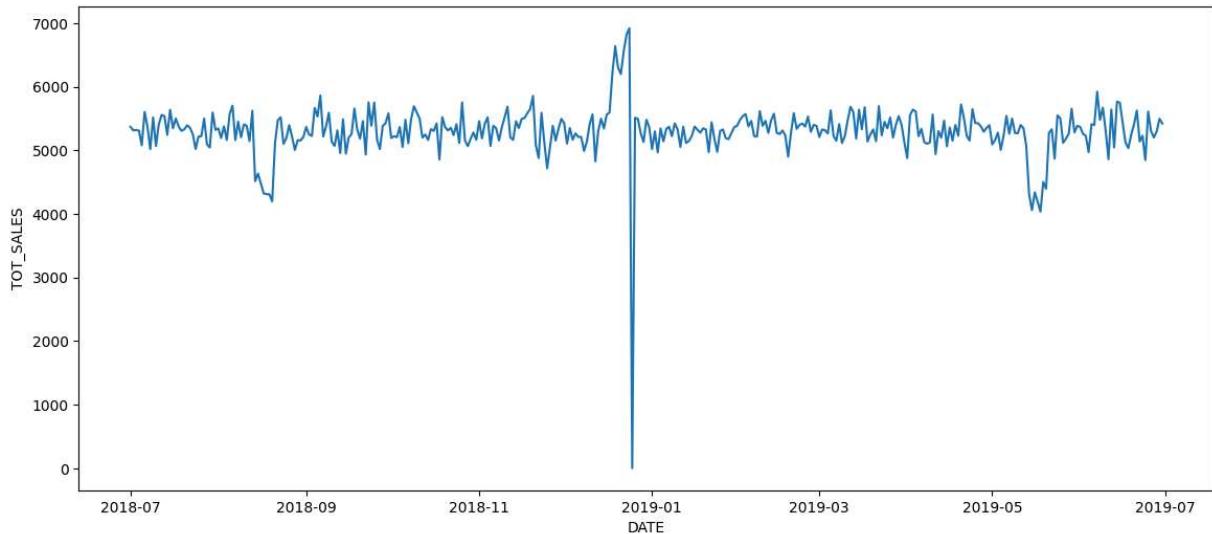
```
Out[ ]: TOT_SALES
```

| DATE | |
|------------|--------|
| 2018-07-01 | 5372.2 |
| 2018-07-02 | 5315.4 |
| 2018-07-03 | 5321.8 |
| 2018-07-04 | 5309.9 |
| 2018-07-05 | 5080.9 |
| ... | ... |
| 2019-06-26 | 5305.0 |
| 2019-06-27 | 5202.8 |
| 2019-06-28 | 5299.6 |
| 2019-06-29 | 5497.6 |
| 2019-06-30 | 5423.4 |

365 rows × 1 columns

```
In [ ]: plt.figure(figsize=(14,6))
sns.lineplot(data=df[['DATE', 'TOT_SALES']].groupby('DATE').sum(), x='DATE', y='TOT_
```

```
Out[ ]: <Axes: xlabel='DATE', ylabel='TOT_SALES'>
```

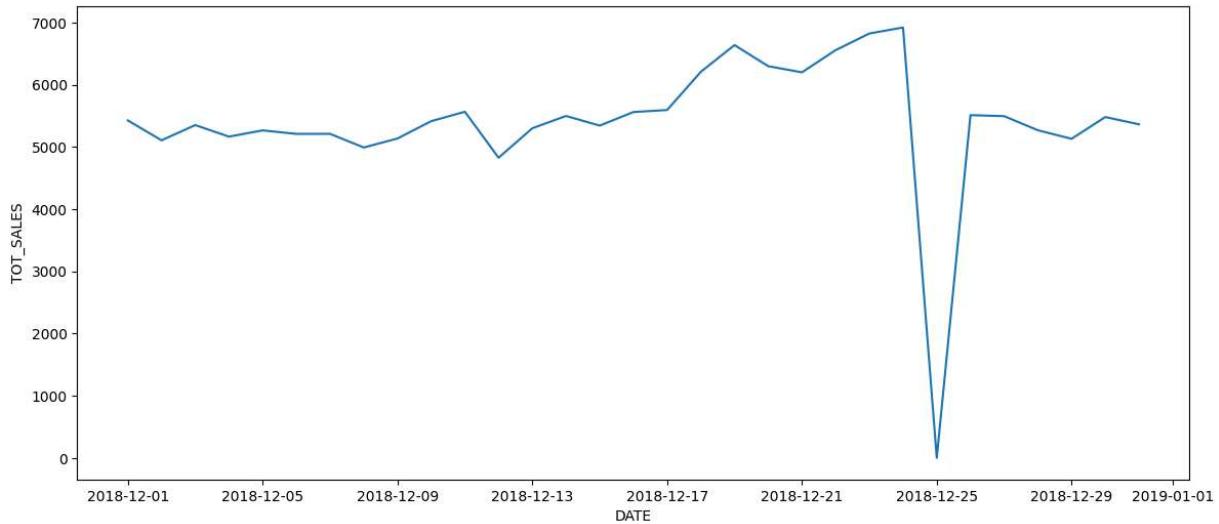


```
In [ ]: tot_sales = df[['DATE', 'TOT_SALES']].groupby('DATE').sum()
```

```
In [ ]: sales_dec = tot_sales.query('"2018-12-01" <= DATE < "2019-01-01"').sort_values(by='TOT_SALES')
```

```
In [ ]: plt.figure(figsize=(14,6))
sns.lineplot(data=sales_dec, x='DATE', y='TOT_SALES')
```

```
Out[ ]: <Axes: xlabel='DATE', ylabel='TOT_SALES'>
```



```
In [ ]: df.head()
```

Out[]: **DATE** **STORE_NBR** **LYLTY_CARD_NBR** **TXN_ID** **PROD_NBR** **PROD_NAME** **PROD_QTY**

| | | | | | | | |
|----------|------------|-------|----------|----------|-------|---------------------------------|-----|
| 0 | 2018-07-01 | 23.0 | 23124.0 | 19481.0 | 102.0 | Kettle Mozzarella Basil & Pesto | 2.0 |
| 1 | 2018-07-01 | 129.0 | 129170.0 | 133242.0 | 102.0 | Kettle Mozzarella Basil & Pesto | 2.0 |
| 2 | 2018-07-01 | 226.0 | 226290.0 | 227777.0 | 102.0 | Kettle Mozzarella Basil & Pesto | 2.0 |
| 3 | 2018-07-01 | 20.0 | 20196.0 | 17175.0 | 102.0 | Kettle Mozzarella Basil & Pesto | 1.0 |
| 4 | 2018-07-01 | 231.0 | 231257.0 | 234734.0 | 102.0 | Kettle Mozzarella Basil & Pesto | 2.0 |

◀ ▶

In []: `df.rename(columns={'PDWT_grams':'PACK_SIZE'}, inplace=True)`

In []: `df['PACK_SIZE'].unique()`

Out[]: `array([175., 150., 110., 165., 135., 380., 170., 134., 330., 270., 250., 210., 300., 90., 220., 190., 200., 70., 160., 180., 125., nan])`

In []: `df['PACK_SIZE'].agg(['mean', 'max', 'min'])`

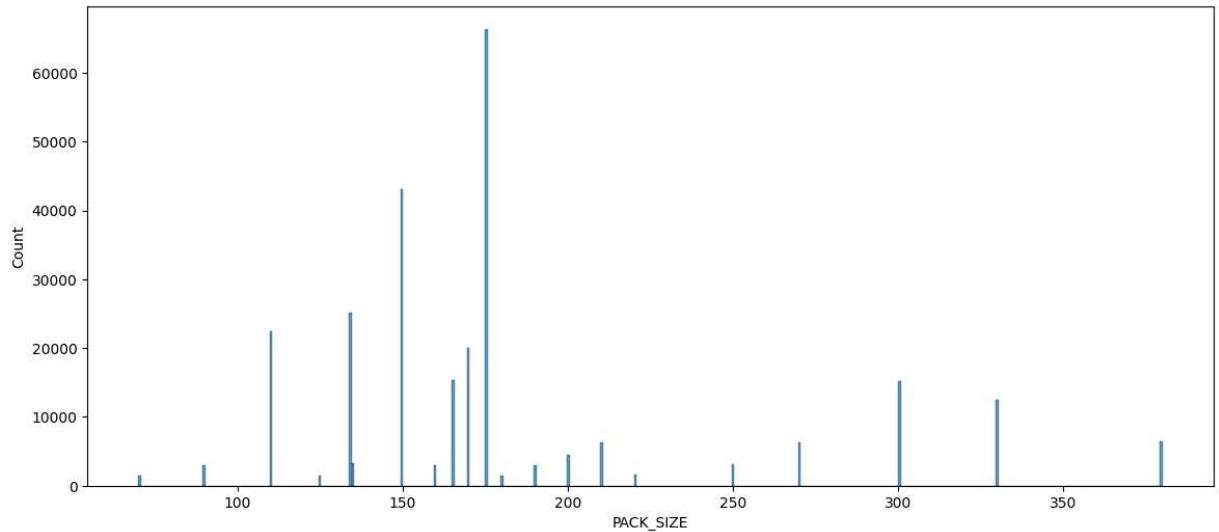
Out[]: `mean 182.42554
max 380.00000
min 70.00000
Name: PACK_SIZE, dtype: float64`

In []: `df.PACK_SIZE.value_counts()`

```
Out[ ]: PACK_SIZE  
175.0    66389  
150.0    43131  
134.0    25102  
110.0    22387  
170.0    19983  
165.0    15297  
300.0    15166  
330.0    12540  
380.0    6416  
270.0    6285  
210.0    6272  
200.0    4473  
135.0    3257  
250.0    3169  
90.0     3008  
190.0    2995  
160.0    2970  
220.0    1564  
70.0     1507  
180.0    1468  
125.0    1454  
Name: count, dtype: int64
```

```
In [ ]: plt.figure(figsize=(14,6))  
sns.histplot(df['PACK_SIZE'])
```

```
Out[ ]: <Axes: xlabel='PACK_SIZE', ylabel='Count'>
```



```
In [ ]: #BRAND column  
  
df['BRAND']= df['PROD_NAME'].str.split().str[0]
```

```
In [ ]: df.head()
```

Out[]:

| | DATE | STORE_NBR | LYLTY_CARD_NBR | TXN_ID | PROD_NBR | PROD_NAME | PROD_QTY |
|---|------------|-----------|----------------|----------|----------|-----------|-------------------------------------|
| 0 | 2018-07-01 | 23.0 | | 23124.0 | 19481.0 | 102.0 | Kettle Mozzarella Basil & Pesto 2.0 |
| 1 | 2018-07-01 | 129.0 | | 129170.0 | 133242.0 | 102.0 | Kettle Mozzarella Basil & Pesto 2.0 |
| 2 | 2018-07-01 | 226.0 | | 226290.0 | 227777.0 | 102.0 | Kettle Mozzarella Basil & Pesto 2.0 |
| 3 | 2018-07-01 | 20.0 | | 20196.0 | 17175.0 | 102.0 | Kettle Mozzarella Basil & Pesto 1.0 |
| 4 | 2018-07-01 | 231.0 | | 231257.0 | 234734.0 | 102.0 | Kettle Mozzarella Basil & Pesto 2.0 |

◀ ▶

In []: df['BRAND'].unique()

Out[]: array(['Kettle', 'Cobs', 'Tyrrells', 'Tostitos', 'Infuzions', 'Smiths', 'Thins', 'Doritos', 'Pringles', 'Dorito', 'Twisties', 'Grain', 'Cheezels', 'Infzns', 'Old', 'Snbts', 'Natural', 'Burger', 'CCs', 'RRD', 'Woolworths', 'Smith', 'WW', 'Red', 'Cheetos', 'GrnWves', 'Sunbites', 'NCC', 'French', nan], dtype=object)

In []: df['BRAND']=df['BRAND'].str.replace('Red', 'RRD', regex=False)

In []: df['BRAND']=df['BRAND'].str.replace('Infzns', 'Infuzions', regex=False)

In []: df['BRAND']=df['BRAND'].str.replace('Snbts', 'Sunbites', regex=False)

In []: df['BRAND']=df['BRAND'].str.replace('Natural', 'NCC', regex=False)

In []: df['BRAND']=df['BRAND'].str.replace('WW', 'Woolworths', regex=False)

In []: df['BRAND']=df['BRAND'].str.replace('Smiths', 'Smith', regex=False)

In []: df['BRAND'].unique()

Out[]: array(['Kettle', 'Cobs', 'Tyrrells', 'Tostitos', 'Infuzions', 'Smith', 'Thins', 'Doritos', 'Pringles', 'Dorito', 'Twisties', 'Grain', 'Cheezels', 'Old', 'Sunbites', 'NCC', 'Burger', 'CCs', 'RRD', 'Woolworths', 'Cheetos', 'GrnWves', 'French', nan], dtype=object)

In []: df.query('BRAND == "CCs"')

Out[]:

| | DATE | STORE_NBR | LYLTY_CARD_NBR | TXN_ID | PROD_NBR | PROD_NAME | PROD_TYPE |
|--------|------------|-----------|----------------|----------|----------|------------------|-----------|
| 513 | 2018-07-01 | 33.0 | 33053.0 | 29755.0 | 91.0 | CCs Tasty Cheese | |
| 514 | 2018-07-01 | 24.0 | 24208.0 | 21476.0 | 91.0 | CCs Tasty Cheese | |
| 515 | 2018-07-01 | 152.0 | 152075.0 | 150697.0 | 91.0 | CCs Tasty Cheese | |
| 516 | 2018-07-01 | 118.0 | 118173.0 | 121672.0 | 91.0 | CCs Tasty Cheese | |
| 517 | 2018-07-01 | 78.0 | 78095.0 | 76010.0 | 91.0 | CCs Tasty Cheese | |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 264668 | 2019-06-30 | 106.0 | 106154.0 | 107674.0 | 66.0 | CCs Nacho Cheese | |
| 264669 | 2019-06-30 | 232.0 | 232129.0 | 235937.0 | 66.0 | CCs Nacho Cheese | |
| 264670 | 2019-06-30 | 121.0 | 121417.0 | 124536.0 | 66.0 | CCs Nacho Cheese | |
| 264671 | 2019-06-30 | 70.0 | 70078.0 | 67940.0 | 66.0 | CCs Nacho Cheese | |
| 264672 | 2019-06-30 | 138.0 | 138229.0 | 141975.0 | 66.0 | CCs Nacho Cheese | |

4551 rows × 10 columns



In []: df['BRAND']=df['BRAND'].str.replace('Grain', 'Grain Waves', regex=False)

In []: df['BRAND']=df['BRAND'].str.replace('GrnWves', 'Grain Waves', regex=False)

In []: df['BRAND']=df['BRAND'].str.replace('Dorito', 'Doritos', regex=False)

In []: df['BRAND']=df['BRAND'].str.replace('Doritoss', 'Doritos', regex=False)

In []: df['BRAND'].unique()

Out[]: array(['Kettle', 'Cobs', 'Tyrrells', 'Tostitos', 'Infuzions', 'Smith', 'Thins', 'Doritos', 'Pringles', 'Twisties', 'Grain Waves', 'Cheezels', 'Old', 'Sunbites', 'NCC', 'Burger', 'CCs', 'RRD', 'Woolworths', 'Cheetos', 'French', nan], dtype=object)

In []: df.head()

Out[]:

| | DATE | STORE_NBR | LYLTY_CARD_NBR | TXN_ID | PROD_NBR | PROD_NAME | PROD_QTY |
|---|------------|-----------|----------------|----------|----------|-----------|-------------------------------------|
| 0 | 2018-07-01 | 23.0 | | 23124.0 | 19481.0 | 102.0 | Kettle Mozzarella Basil & Pesto 2.0 |
| 1 | 2018-07-01 | 129.0 | | 129170.0 | 133242.0 | 102.0 | Kettle Mozzarella Basil & Pesto 2.0 |
| 2 | 2018-07-01 | 226.0 | | 226290.0 | 227777.0 | 102.0 | Kettle Mozzarella Basil & Pesto 2.0 |
| 3 | 2018-07-01 | 20.0 | | 20196.0 | 17175.0 | 102.0 | Kettle Mozzarella Basil & Pesto 1.0 |
| 4 | 2018-07-01 | 231.0 | | 231257.0 | 234734.0 | 102.0 | Kettle Mozzarella Basil & Pesto 2.0 |



In []: df.BRAND.value_counts().reset_index()

Out[]:

| | BRAND | count |
|----|-------------|-------|
| 0 | Kettle | 41288 |
| 1 | Smith | 31822 |
| 2 | Doritos | 28145 |
| 3 | Pringles | 25102 |
| 4 | RRD | 17779 |
| 5 | Woolworths | 14757 |
| 6 | Infuzions | 14201 |
| 7 | Thins | 14075 |
| 8 | Cobs | 9693 |
| 9 | Tostitos | 9471 |
| 10 | Twisties | 9454 |
| 11 | Old | 9324 |
| 12 | Grain Waves | 7740 |
| 13 | NCC | 7469 |
| 14 | Tyrrells | 6442 |
| 15 | Cheezels | 4603 |
| 16 | CCs | 4551 |
| 17 | Sunbites | 3008 |
| 18 | Cheetos | 2927 |
| 19 | Burger | 1564 |
| 20 | French | 1418 |

In []:

df.shape

Out[]:

(264834, 10)

In []:

df_cd = pd.read_csv(r'A:\Data Analytics\Forage\Quantum\QVI_purchase_behaviour.csv')

In []:

df_cd.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
```

In []: df_cd['LIFESTAGE'].value_counts(ascending=False).reset_index()

Out[]:

| LIFESTAGE | count |
|-----------|-----------------------------|
| 0 | RETIREES 14805 |
| 1 | OLDER SINGLES/COUPLES 14609 |
| 2 | YOUNG SINGLES/COUPLES 14441 |
| 3 | OLDER FAMILIES 9780 |
| 4 | YOUNG FAMILIES 9178 |
| 5 | MIDAGE SINGLES/COUPLES 7275 |
| 6 | NEW FAMILIES 2549 |

In []: df_cd_ls_c = df_cd['LIFESTAGE'].value_counts(ascending=False).reset_index()

In []: df_cd['PREMIUM_CUSTOMER'].value_counts(ascending=False).reset_index()

Out[]:

| PREMIUM_CUSTOMER | count |
|------------------|------------------|
| 0 | Mainstream 29245 |
| 1 | Budget 24470 |
| 2 | Premium 18922 |

In []: df_cd_pc_c = df_cd['PREMIUM_CUSTOMER'].value_counts(ascending=False).reset_index()

In []: df_cd[df_cd['LYLTY_CARD_NBR'] == 226000.0]

Out[]:

| LYLTY_CARD_NBR | LIFESTAGE | PREMIUM_CUSTOMER |
|----------------|-----------------------|------------------|
| 59694 | 226000 OLDER FAMILIES | Premium |

In []: df_cd.drop(index=59694, inplace=True)

In []: df.merge(df_cd, on='LYLTY_CARD_NBR', how='outer')

Out[]:

| | DATE | STORE_NBR | LYLTY_CARD_NBR | TXN_ID | PROD_NBR | PROD_NAME | PROD_CD |
|--------|------------|-----------|----------------|----------|----------|--|---------|
| 0 | 2018-10-17 | 1.0 | 1000.0 | 1.0 | 5.0 | Natural Chip Comnpy SeaSalt | |
| 1 | 2018-09-16 | 1.0 | 1002.0 | 2.0 | 58.0 | Red Rock Deli Chikn & Garlic Aioli | |
| 2 | 2019-03-07 | 1.0 | 1003.0 | 3.0 | 52.0 | Grain Waves Sour Cream & Chives | |
| 3 | 2019-03-08 | 1.0 | 1003.0 | 4.0 | 106.0 | Natural ChipCo Hony Soy Chckn | |
| 4 | 2018-11-02 | 1.0 | 1004.0 | 5.0 | 96.0 | WW Original Stacked Chips | |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 264829 | 2018-10-01 | 88.0 | 2370751.0 | 240394.0 | 60.0 | Kettle Tortilla ChpsFeta & Garlic | |
| 264830 | 2018-10-24 | 88.0 | 2370961.0 | 240480.0 | 70.0 | Tyrrells Crisps Lightly Salted | |
| 264831 | 2018-10-27 | 88.0 | 2370961.0 | 240481.0 | 65.0 | Old El Paso Salsa Dip Chnky Tom Ht | |
| 264832 | 2018-12-14 | 88.0 | 2373711.0 | 241815.0 | 16.0 | Smiths Crinkle Chips Salt & Vinegar | |
| 264833 | 2018-12-25 | NaN | NaN | NaN | NaN | NaN | |

264834 rows × 12 columns



In []: df_merged = df.merge(df_cd, on='LYLTY_CARD_NBR', how='outer')

In []: df_merged.isna().value_counts()

```
Out[ ]: DATE STORE_NBR LYLTY_CARD_NBR TXN_ID PROD_NBR PROD_NAME PROD_QTY TOT_SALES  
PACK_SIZE BRAND LIFESTAGE PREMIUM_CUSTOMER  
False False False False False False False False  
False False False False 264833  
True True True True True True True True  
True True True True 1  
Name: count, dtype: int64
```

```
In [ ]: df = df_merged
```

```
In [ ]: df
```

Out[]:

| | DATE | STORE_NBR | LYLTY_CARD_NBR | TXN_ID | PROD_NBR | PROD_NAME | PROD_ |
|--------|------------|-----------|----------------|----------|----------|--|-------|
| 0 | 2018-10-17 | 1.0 | 1000.0 | 1.0 | 5.0 | Natural Chip Comnpy SeaSalt | |
| 1 | 2018-09-16 | 1.0 | 1002.0 | 2.0 | 58.0 | Red Rock Deli Chikn & Garlic Aioli | |
| 2 | 2019-03-07 | 1.0 | 1003.0 | 3.0 | 52.0 | Grain Waves Sour Cream & Chives | |
| 3 | 2019-03-08 | 1.0 | 1003.0 | 4.0 | 106.0 | Natural ChipCo Hony Soy Chckn | |
| 4 | 2018-11-02 | 1.0 | 1004.0 | 5.0 | 96.0 | WW Original Stacked Chips | |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 264829 | 2018-10-01 | 88.0 | 2370751.0 | 240394.0 | 60.0 | Kettle Tortilla ChpsFeta & Garlic | |
| 264830 | 2018-10-24 | 88.0 | 2370961.0 | 240480.0 | 70.0 | Tyrrells Crisps Lightly Salted | |
| 264831 | 2018-10-27 | 88.0 | 2370961.0 | 240481.0 | 65.0 | Old El Paso Salsa Dip Chnky Tom Ht | |
| 264832 | 2018-12-14 | 88.0 | 2373711.0 | 241815.0 | 16.0 | Smiths Crinkle Chips Salt & Vinegar | |
| 264833 | 2018-12-25 | NaN | NaN | NaN | NaN | NaN | |

264834 rows × 12 columns



In []: df_tcd = df.groupby('LYLTY_CARD_NBR')['TOT_SALES'].sum().sort_values(ascending=False)

In []: df_tcd

Out[]: **LYLTY_CARD_NBR TOT_SALES**

| 0 | 230078.0 | 138.6 |
|--------------|----------|-------|
| 1 | 63197.0 | 132.8 |
| 2 | 259009.0 | 127.2 |
| 3 | 162039.0 | 126.8 |
| 4 | 58361.0 | 124.8 |
| ... | ... | ... |
| 72631 | 41089.0 | 1.5 |
| 72632 | 41042.0 | 1.5 |
| 72633 | 38262.0 | 1.5 |
| 72634 | 38391.0 | 1.5 |
| 72635 | 120121.0 | 1.5 |

72636 rows × 2 columns

In []: df['STORE_NBR']=df['STORE_NBR'].fillna(0).astype(int)

In []: df['LYLTY_CARD_NBR']=df['LYLTY_CARD_NBR'].fillna(0).astype(int)

In []: df['TXN_ID']=df['TXN_ID'].fillna(0).astype(int)

In []: df['PROD_NBR']=df['PROD_NBR'].fillna(0).astype(int)

In []: df['PROD_QTY']=df['PROD_QTY'].fillna(0).astype(int)

In []: df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 264834 entries, 0 to 264833
Data columns (total 12 columns):
 #   Column            Non-Null Count  Dtype  
--- 
 0   DATE              264834 non-null   datetime64[ns]
 1   STORE_NBR         264834 non-null   int64  
 2   LYLTY_CARD_NBR   264834 non-null   int64  
 3   TXN_ID            264834 non-null   int64  
 4   PROD_NBR          264834 non-null   int64  
 5   PROD_NAME         264833 non-null   object  
 6   PROD_QTY          264834 non-null   int64  
 7   TOT_SALES         264833 non-null   float64 
 8   PACK_SIZE          264833 non-null   float64 
 9   BRAND              264833 non-null   object  
 10  LIFESTAGE          264833 non-null   object  
 11  PREMIUM_CUSTOMER  264833 non-null   object  
dtypes: datetime64[ns](1), float64(2), int64(5), object(4)
memory usage: 24.2+ MB
```

```
In [ ]: df.query('LYLTY_CARD_NBR == 230078')
```

Out[]:

| | DATE | STORE_NBR | LYLTY_CARD_NBR | TXN_ID | PROD_NBR | PROD_NAME | PROD_C |
|--|------|-----------|----------------|--------|----------|-----------|--------|
|--|------|-----------|----------------|--------|----------|-----------|--------|

| | | | | | | | |
|---------------|------------|-----|--------|--------|----|-------------------------------------|--|
| 227711 | 2018-07-18 | 230 | 230078 | 232416 | 7 | Smiths Crinkle Original | |
| 227712 | 2018-07-31 | 230 | 230078 | 232417 | 85 | RRD Honey Soy Chicken | |
| 227713 | 2018-09-26 | 230 | 230078 | 232418 | 90 | Tostitos Smoked Chipotle | |
| 227714 | 2018-09-29 | 230 | 230078 | 232419 | 1 | Smiths Crinkle Cut Chips Barbecue | |
| 227715 | 2018-10-25 | 230 | 230078 | 232420 | 2 | Cobs Popd Sour Crm & Chives Chips | |
| 227716 | 2018-10-29 | 230 | 230078 | 232421 | 87 | Infuzions BBQ Rib Prawn Crackers | |
| 227717 | 2018-11-18 | 230 | 230078 | 232422 | 79 | Smiths Chip Thinly Cut Salt Vinegar | |
| 227718 | 2018-12-23 | 230 | 230078 | 232424 | 17 | Kettle Sensations BBQ & Maple | |
| 227719 | 2019-01-09 | 230 | 230078 | 232425 | 63 | Kettle 135g Swt Pot Sea Salt | |
| 227720 | 2019-02-18 | 230 | 230078 | 232426 | 88 | Kettle Honey Soy Chicken | |
| 227721 | 2019-03-21 | 230 | 230078 | 232427 | 82 | Smith Crinkle Cut Mac N Cheese | |
| 227722 | 2019-05-09 | 230 | 230078 | 232428 | 15 | Twisties Cheese | |
| 227723 | 2019-05-12 | 230 | 230078 | 232429 | 22 | Thins Chips Originl salted | |
| 227724 | 2019-05-17 | 230 | 230078 | 232430 | 68 | Pringles Chicken Salt Crips | |
| 227725 | 2019-05-27 | 230 | 230078 | 232431 | 42 | Doritos Corn Chip Mexican Jalapeno | |

| | DATE | STORE_NBR | LYLTY_CARD_NBR | TXN_ID | PROD_NBR | PROD_NAME | PROD_C |
|---------------|------------|-----------|----------------|--------|----------|-----------|-------------------------------|
| 227726 | 2019-06-01 | 230 | | 230078 | 232432 | 17 | Kettle Sensations BBQ & Maple |
| 227727 | 2019-06-20 | 230 | | 230078 | 232433 | 91 | CCs Tasty Cheese |

```
In [ ]: df.query('LYLTY_CARD_NBR == 259009')
```

Out[]:

| | | DATE | STORE_NBR | LYLTY_CARD_NBR | TXN_ID | PROD_NBR | PROD_NAME | PROD_C |
|--------|------------|------|-----------|----------------|--------|----------|-----------|--------------------------------------|
| 252263 | 2018-07-04 | | 259 | | 259009 | 257456 | 1 | Smiths Crinkle Cut Chips Barbecue |
| 252264 | 2018-07-19 | | 259 | | 259009 | 257457 | 15 | Twisties Cheese |
| 252265 | 2018-08-30 | | 259 | | 259009 | 257458 | 51 | Doritos Mexicana |
| 252266 | 2018-09-24 | | 259 | | 259009 | 257459 | 60 | Kettle Tortilla ChpsFeta & Garlic |
| 252267 | 2018-10-13 | | 259 | | 259009 | 257460 | 88 | Kettle Honey Soy Chicken |
| 252268 | 2018-10-13 | | 259 | | 259009 | 257460 | 41 | Doritos Salsa Mild |
| 252269 | 2018-12-26 | | 259 | | 259009 | 257461 | 50 | Tostitos Lightly Salted |
| 252270 | 2018-12-27 | | 259 | | 259009 | 257462 | 22 | Thins Chips Originl saltd |
| 252271 | 2019-02-21 | | 259 | | 259009 | 257463 | 90 | Tostitos Smoked Chipotle |
| 252272 | 2019-03-13 | | 259 | | 259009 | 257464 | 65 | Old El Paso Salsa Dip Chnky Tom Ht |
| 252273 | 2019-04-12 | | 259 | | 259009 | 257465 | 84 | GrnWves Plus Btroot & Chilli Jam |
| 252274 | 2019-04-25 | | 259 | | 259009 | 257466 | 4 | Dorito Corn Chp Supreme |
| 252275 | 2019-04-25 | | 259 | | 259009 | 257466 | 94 | Burger Rings |
| 252276 | 2019-04-26 | | 259 | | 259009 | 257467 | 88 | Kettle Honey Soy Chicken |
| 252277 | 2019-05-02 | | 259 | | 259009 | 257468 | 9 | Kettle Tortilla ChpsBtroot & Ricotta |



In []: tp_cd = df_tcd['LYLTY_CARD_NBR'][:20]

```
In [ ]: tp_cd
```

```
Out[ ]: 0    230078.0
        1    63197.0
        2    259009.0
        3    162039.0
        4    58361.0
        5    230154.0
        6    94185.0
        7    130090.0
        8    179228.0
        9    172032.0
       10   199157.0
       11   116181.0
       12   114163.0
       13    3153.0
       14   59333.0
       15   28020.0
       16   164034.0
       17   156054.0
       18   116113.0
       19   199146.0
Name: LYLTY_CARD_NBR, dtype: float64
```

```
In [ ]: pd.set_option('display.max_rows',100)
```

```
In [ ]: df_cd['LIFESTAGE'].value_counts(ascending=False).reset_index()
```

```
Out[ ]:
```

| | LIFESTAGE | count |
|----------|------------------------|-------|
| 0 | RETIREES | 14805 |
| 1 | OLDER SINGLES/COUPLES | 14609 |
| 2 | YOUNG SINGLES/COUPLES | 14441 |
| 3 | OLDER FAMILIES | 9779 |
| 4 | YOUNG FAMILIES | 9178 |
| 5 | MIDAGE SINGLES/COUPLES | 7275 |
| 6 | NEW FAMILIES | 2549 |

```
In [ ]: df.query('LYLTY_CARD_NBR in @tp_cd').groupby('LYLTY_CARD_NBR').value_counts(ascending=False)
```

Out[]:

| | LYLTY_CARD_NBR | DATE | STORE_NBR | TXN_ID | PROD_NBR | PROD_NAME | PROD_QTY |
|-----|----------------|------------|-----------|--------|----------|-----------|--|
| 0 | 3153 | 2018-07-04 | | 3 | 1723 | 87 | Infuzions BBQ Rib Prawn Crackers |
| 1 | 3153 | 2018-07-13 | | 3 | 1724 | 30 | Doritos Corn Chips Cheese Supreme |
| 2 | 3153 | 2018-07-22 | | 3 | 1725 | 17 | Kettle Sensations BBQ & Maple |
| 3 | 3153 | 2018-08-21 | | 3 | 1726 | 46 | Kettle Original |
| 4 | 3153 | 2018-09-08 | | 3 | 1727 | 81 | Pringles Original Crisps |
| ... | ... | ... | | ... | ... | ... | ... |
| 295 | 259009 | 2019-04-12 | | 259 | 257465 | 84 | GrnWves Plus Btroot & Chilli Jam |
| 296 | 259009 | 2019-04-25 | | 259 | 257466 | 4 | Dorito Corn Chp Supreme |
| 297 | 259009 | 2019-04-25 | | 259 | 257466 | 94 | Burger Rings |
| 298 | 259009 | 2019-04-26 | | 259 | 257467 | 88 | Kettle Honey Soy Chicken |
| 299 | 259009 | 2019-05-02 | | 259 | 257468 | 9 | Kettle Tortilla ChpsBtroot & Ricotta |

300 rows × 13 columns



In []: df_tcd_plt = df.query('LYLTY_CARD_NBR in @tp_cd').groupby('LYLTY_CARD_NBR').value_c

In []: df_cd.query('LYLTY_CARD_NBR in @tp_cd')

Out[]:

| | LYLTY_CARD_NBR | LIFESTAGE | PREMIUM_CUSTOMER |
|--------------|----------------|------------------------|------------------|
| 810 | 3153 | MIDAGE SINGLES/COUPLES | Premium |
| 7846 | 28020 | OLDER FAMILIES | Premium |
| 16001 | 58361 | YOUNG FAMILIES | Budget |
| 16363 | 59333 | YOUNG FAMILIES | Premium |
| 17292 | 63197 | OLDER FAMILIES | Budget |
| 25958 | 94185 | YOUNG FAMILIES | Premium |
| 31050 | 114163 | OLDER FAMILIES | Budget |
| 31586 | 116113 | OLDER SINGLES/COUPLES | Premium |
| 31654 | 116181 | YOUNG FAMILIES | Premium |
| 35338 | 130090 | YOUNG FAMILIES | Budget |
| 41873 | 156054 | MIDAGE SINGLES/COUPLES | Premium |
| 42813 | 162039 | OLDER FAMILIES | Mainstream |
| 43370 | 164034 | RETIREEES | Budget |
| 45905 | 172032 | YOUNG SINGLES/COUPLES | Mainstream |
| 47959 | 179228 | YOUNG FAMILIES | Budget |
| 52864 | 199146 | YOUNG FAMILIES | Budget |
| 52875 | 199157 | YOUNG FAMILIES | Premium |
| 60925 | 230078 | OLDER FAMILIES | Budget |
| 61001 | 230154 | OLDER FAMILIES | Budget |
| 69017 | 259009 | OLDER SINGLES/COUPLES | Mainstream |

In []: df.groupby('LYLTY_CARD_NBR')[['TOT_SALES', 'PROD_QTY']].sum().reset_index()

Out[]:

| | LYLTY_CARD_NBR | TOT_SALES | PROD_QTY |
|--------------|----------------|-----------|----------|
| 0 | 0 | 0.0 | 0 |
| 1 | 1000 | 6.0 | 2 |
| 2 | 1002 | 2.7 | 1 |
| 3 | 1003 | 6.6 | 2 |
| 4 | 1004 | 1.9 | 1 |
| ... | ... | ... | ... |
| 72632 | 2370651 | 13.0 | 2 |
| 72633 | 2370701 | 7.2 | 2 |
| 72634 | 2370751 | 9.2 | 2 |
| 72635 | 2370961 | 18.6 | 4 |
| 72636 | 2373711 | 11.4 | 2 |

72637 rows × 3 columns

```
In [ ]: df_cd_agg = df.groupby('LYLTY_CARD_NBR')[['TOT_SALES','PROD_QTY']].sum().reset_index()
In [ ]: df_cd_merged = df_cd.merge(df_cd_agg, on='LYLTY_CARD_NBR', how='inner')
In [ ]: df_cd_merged.sort_values(by='TOT_SALES', ascending=False)
```

Out[]:

| | LYLTY_CARD_NBR | LIFESTAGE | PREMIUM_CUSTOMER | TOT_SALES | PROD_QTY |
|--------------|----------------|-------------------------|------------------|-----------|----------|
| 60924 | 230078 | OLDER FAMILIES | Budget | 138.6 | 36 |
| 17292 | 63197 | OLDER FAMILIES | Budget | 132.8 | 32 |
| 69016 | 259009 | OLDER SINGLES/COPLES | Mainstream | 127.2 | 30 |
| 42813 | 162039 | OLDER FAMILIES | Mainstream | 126.8 | 36 |
| 16001 | 58361 | YOUNG FAMILIES | Budget | 124.8 | 28 |
| ... | ... | ... | ... | ... | ... |
| 11310 | 41089 | YOUNG SINGLES/COPLES | Premium | 1.5 | 1 |
| 11282 | 41042 | RETIREES | Premium | 1.5 | 1 |
| 10464 | 38262 | YOUNG SINGLES/COPLES | Premium | 1.5 | 1 |
| 10546 | 38391 | OLDER SINGLES/COPLES | Budget | 1.5 | 1 |
| 32447 | 120121 | YOUNG SINGLES/COPLES | Budget | 1.5 | 1 |

72636 rows × 5 columns



In []: df_cd_merged.groupby('PREMIUM_CUSTOMER')[['TOT_SALES', 'PROD_QTY']].sum().sort_values

Out[]: TOT_SALES PROD_QTY

| PREMIUM_CUSTOMER | TOT_SALES | PROD_QTY |
|-------------------|-----------|----------|
| Mainstream | 750744.50 | 193965 |
| Budget | 676211.55 | 177898 |
| Premium | 506152.95 | 132859 |

In []: df_cd_pc = df_cd_merged.groupby('PREMIUM_CUSTOMER')[['TOT_SALES', 'PROD_QTY']].sum()

In []: df_cd_pc = df_cd_pc.merge(df_cd_pc_c, on='PREMIUM_CUSTOMER', how='inner')

In []: df_cd_pc['Chips_Per_Customer'] = df_cd_pc['PROD_QTY']/df_cd_pc['count']

In []: df_cd_pc

| | PREMIUM_CUSTOMER | TOT_SALES | PROD_QTY | count | Chips_Per_Customer |
|---|------------------|-----------|----------|-------|--------------------|
| 0 | Mainstream | 750744.50 | 193965 | 29245 | 6.632416 |
| 1 | Budget | 676211.55 | 177898 | 24470 | 7.270045 |
| 2 | Premium | 506152.95 | 132859 | 18922 | 7.021404 |

```
In [ ]: df_cd_pc['Sales_Per_Customer'] = df_cd_pc['TOT_SALES']/df_cd_pc['count']
```

```
In [ ]: df_cd_pc
```

| | PREMIUM_CUSTOMER | TOT_SALES | PROD_QTY | count | Chips_Per_Customer | Sales_Per_Cu |
|---|------------------|-----------|----------|-------|--------------------|--------------|
| 0 | Mainstream | 750744.50 | 193965 | 29245 | 6.632416 | 25. |
| 1 | Budget | 676211.55 | 177898 | 24470 | 7.270045 | 27. |
| 2 | Premium | 506152.95 | 132859 | 18922 | 7.021404 | 26. |

```
In [ ]: df_cd_merged.groupby('LIFESTAGE')[['TOT_SALES', 'PROD_QTY']].sum().sort_values(by='TOT_SALES')
```

```
Out[ ]:
```

TOT_SALES PROD_QTY

LIFESTAGE

| | | |
|-------------------------------|-----------|--------|
| OLDER SINGLES/COUPLES | 402420.75 | 104201 |
| RETIREES | 366470.90 | 94166 |
| OLDER FAMILIES | 352467.20 | 94592 |
| YOUNG FAMILIES | 316160.10 | 84561 |
| YOUNG SINGLES/COUPLES | 260405.30 | 66634 |
| MIDAGE SINGLES/COUPLES | 184751.30 | 47721 |
| NEW FAMILIES | 50433.45 | 12847 |

```
In [ ]: df_cd_ls = df_cd_merged.groupby('LIFESTAGE')[['TOT_SALES', 'PROD_QTY']].sum().sort_
```

```
In [ ]: df_cd_ls = df_cd_ls.merge(df_cd_ls_c, on='LIFESTAGE', how='inner')
```

```
In [ ]: df_cd_ls['Chips_Per_Customer'] = df_cd_ls['PROD_QTY']/df_cd_ls['count']
```

```
In [ ]: df_cd_ls
```

Out[]:

| | LIFESTAGE | TOT_SALES | PROD_QTY | count | Chips_Per_Customer |
|---|------------------------|-----------|----------|-------|--------------------|
| 0 | OLDER SINGLES/COUPLES | 402420.75 | 104201 | 14609 | 7.132658 |
| 1 | RETIREES | 366470.90 | 94166 | 14805 | 6.360419 |
| 2 | OLDER FAMILIES | 352467.20 | 94592 | 9780 | 9.671984 |
| 3 | YOUNG FAMILIES | 316160.10 | 84561 | 9178 | 9.213445 |
| 4 | YOUNG SINGLES/COUPLES | 260405.30 | 66634 | 14441 | 4.614223 |
| 5 | MIDAGE SINGLES/COUPLES | 184751.30 | 47721 | 7275 | 6.559588 |
| 6 | NEW FAMILIES | 50433.45 | 12847 | 2549 | 5.040016 |

In []: df_cd_ls['Sales_Per_Customer'] = df_cd_ls['TOT_SALES']/df_cd_ls['count']

In []: df_cd_ls

Out[]:

| | LIFESTAGE | TOT_SALES | PROD_QTY | count | Chips_Per_Customer | Sales_Per_Custom |
|---|------------------------|-----------|----------|-------|--------------------|------------------|
| 0 | OLDER SINGLES/COUPLES | 402420.75 | 104201 | 14609 | 7.132658 | 27.54608 |
| 1 | RETIREES | 366470.90 | 94166 | 14805 | 6.360419 | 24.75318 |
| 2 | OLDER FAMILIES | 352467.20 | 94592 | 9780 | 9.671984 | 36.03959 |
| 3 | YOUNG FAMILIES | 316160.10 | 84561 | 9178 | 9.213445 | 34.44760 |
| 4 | YOUNG SINGLES/COUPLES | 260405.30 | 66634 | 14441 | 4.614223 | 18.03231 |
| 5 | MIDAGE SINGLES/COUPLES | 184751.30 | 47721 | 7275 | 6.559588 | 25.39536 |
| 6 | NEW FAMILIES | 50433.45 | 12847 | 2549 | 5.040016 | 19.78558 |



In []: df_cd.groupby(['LIFESTAGE', 'PREMIUM_CUSTOMER']).count().rename(columns= {'LYLTY_CAR

Out[]:

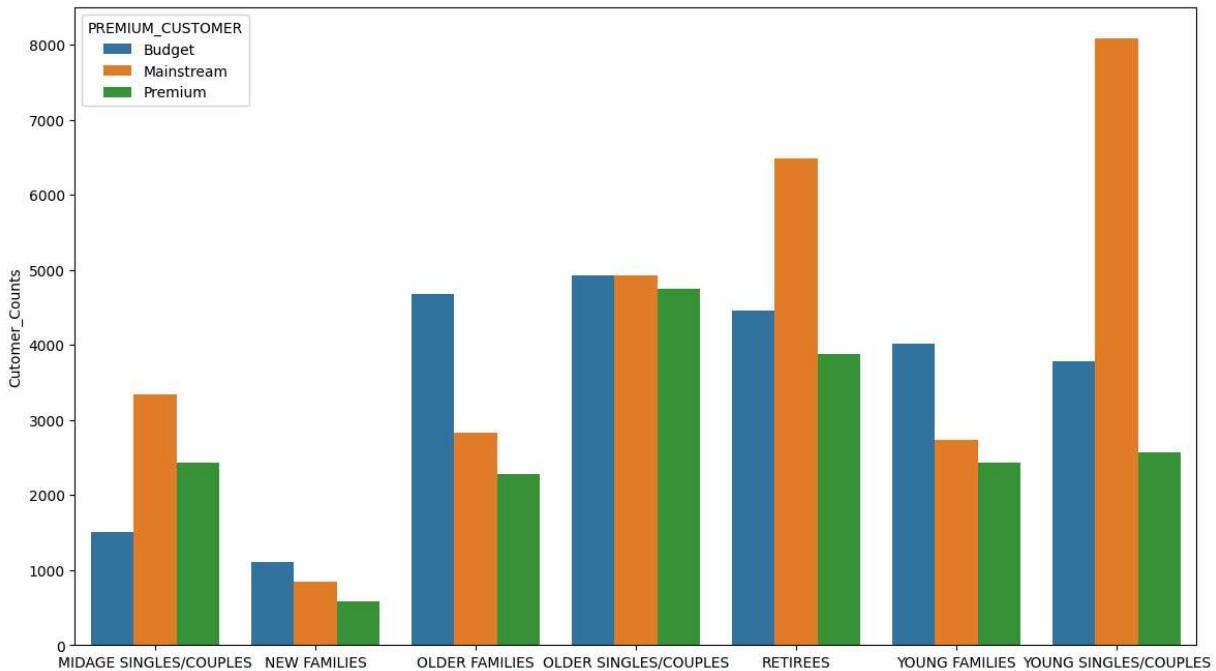
Cutomer_Counts

| LIFESTAGE PREMIUM_CUSTOMER | | |
|-----------------------------------|-------------------|------|
| MIDAGE SINGLES/COUPLES | Budget | 1504 |
| | Mainstream | 3340 |
| | Premium | 2431 |
| NEW FAMILIES | Budget | 1112 |
| | Mainstream | 849 |
| | Premium | 588 |
| OLDER FAMILIES | Budget | 4675 |
| | Mainstream | 2831 |
| | Premium | 2273 |
| OLDER SINGLES/COUPLES | Budget | 4929 |
| | Mainstream | 4930 |
| | Premium | 4750 |
| RETIREES | Budget | 4454 |
| | Mainstream | 6479 |
| | Premium | 3872 |
| YOUNG FAMILIES | Budget | 4017 |
| | Mainstream | 2728 |
| | Premium | 2433 |
| YOUNG SINGLES/COUPLES | Budget | 3779 |
| | Mainstream | 8088 |
| | Premium | 2574 |

In []: df_cd_seg_c = df_cd.groupby(['LIFESTAGE', 'PREMIUM_CUSTOMER']).count().rename(columns

In []: plt.figure(figsize=(14,8))
sns.barplot(data=df_cd_seg_c, x='LIFESTAGE', y='Cutomer_Counts', hue='PREMIUM_CUSTO
plt.xlabel(None)

Out[]: Text(0.5, 0, '')



```
In [ ]: df_cd.query('LIFESTAGE == "MIDAGE SINGLES/COUPLES"').count()
```

```
Out[ ]: LYLTY_CARD_NBR      7275
LIFESTAGE          7275
PREMIUM_CUSTOMER    7275
dtype: int64
```

```
In [ ]: df_cd_merged[['LIFESTAGE', 'PREMIUM_CUSTOMER', 'TOT_SALES', 'PROD_QTY']].groupby(['LIF
```

Out[]:

| LIFESTAGE | PREMIUM_CUSTOMER | TOT_SALES | PROD_QTY |
|------------------------|------------------|-----------|----------|
| OLDER FAMILIES | Budget | 168363.25 | 45065 |
| YOUNG SINGLES/COUPLES | Mainstream | 157621.60 | 38632 |
| RETIREES | Mainstream | 155677.05 | 40518 |
| YOUNG FAMILIES | Budget | 139345.85 | 37111 |
| OLDER SINGLES/COUPLES | Budget | 136769.80 | 35220 |
| | Mainstream | 133393.80 | 34997 |
| | Premium | 132257.15 | 33984 |
| RETIREES | Budget | 113147.80 | 28764 |
| OLDER FAMILIES | Mainstream | 103445.55 | 27756 |
| RETIREES | Premium | 97646.05 | 24884 |
| YOUNG FAMILIES | Mainstream | 92788.75 | 25044 |
| MIDAGE SINGLES/COUPLES | Mainstream | 90803.85 | 22699 |
| YOUNG FAMILIES | Premium | 84025.50 | 22406 |
| OLDER FAMILIES | Premium | 80658.40 | 21771 |
| YOUNG SINGLES/COUPLES | Budget | 61141.60 | 16671 |
| MIDAGE SINGLES/COUPLES | Premium | 58432.65 | 15526 |
| YOUNG SINGLES/COUPLES | Premium | 41642.10 | 11331 |
| MIDAGE SINGLES/COUPLES | Budget | 35514.80 | 9496 |
| NEW FAMILIES | Budget | 21928.45 | 5571 |
| | Mainstream | 17013.90 | 4319 |
| | Premium | 11491.10 | 2957 |

In []: df_cd_seg = df_cd_merged[['LIFESTAGE', 'PREMIUM_CUSTOMER', 'TOT_SALES', 'PROD_QTY']].g

In []: df_cd_seg

Out[]:

| LIFESTAGE | PREMIUM_CUSTOMER | TOT_SALES | PROD_QTY |
|------------------------|------------------|-----------|----------|
| OLDER FAMILIES | Budget | 168363.25 | 45065 |
| YOUNG SINGLES/COUPLES | Mainstream | 157621.60 | 38632 |
| RETIREES | Mainstream | 155677.05 | 40518 |
| YOUNG FAMILIES | Budget | 139345.85 | 37111 |
| OLDER SINGLES/COUPLES | Budget | 136769.80 | 35220 |
| | Mainstream | 133393.80 | 34997 |
| | Premium | 132257.15 | 33984 |
| RETIREES | Budget | 113147.80 | 28764 |
| OLDER FAMILIES | Mainstream | 103445.55 | 27756 |
| RETIREES | Premium | 97646.05 | 24884 |
| YOUNG FAMILIES | Mainstream | 92788.75 | 25044 |
| MIDAGE SINGLES/COUPLES | Mainstream | 90803.85 | 22699 |
| YOUNG FAMILIES | Premium | 84025.50 | 22406 |
| OLDER FAMILIES | Premium | 80658.40 | 21771 |
| YOUNG SINGLES/COUPLES | Budget | 61141.60 | 16671 |
| MIDAGE SINGLES/COUPLES | Premium | 58432.65 | 15526 |
| YOUNG SINGLES/COUPLES | Premium | 41642.10 | 11331 |
| MIDAGE SINGLES/COUPLES | Budget | 35514.80 | 9496 |
| NEW FAMILIES | Budget | 21928.45 | 5571 |
| | Mainstream | 17013.90 | 4319 |
| | Premium | 11491.10 | 2957 |

In []: df_cd_seg.merge(df_cd_seg_c, on=['LIFESTAGE', 'PREMIUM_CUSTOMER'], how='inner')

Out[]:

| | | | TOT_SALES | PROD_QTY | Cutomer_Counts |
|--|------------------------|------------------|-----------|----------|----------------|
| | LIFESTAGE | PREMIUM_CUSTOMER | | | |
| | OLDER FAMILIES | Budget | 168363.25 | 45065 | 4675 |
| | YOUNG SINGLES/COUPLES | Mainstream | 157621.60 | 38632 | 8088 |
| | RETIREES | Mainstream | 155677.05 | 40518 | 6479 |
| | YOUNG FAMILIES | Budget | 139345.85 | 37111 | 4017 |
| | OLDER SINGLES/COUPLES | Budget | 136769.80 | 35220 | 4929 |
| | | Mainstream | 133393.80 | 34997 | 4930 |
| | RETIREES | Premium | 132257.15 | 33984 | 4750 |
| | | Budget | 113147.80 | 28764 | 4454 |
| | | Mainstream | 103445.55 | 27756 | 2831 |
| | OLDER FAMILIES | Premium | 97646.05 | 24884 | 3872 |
| | RETIREES | Mainstream | 92788.75 | 25044 | 2728 |
| | MIDAGE SINGLES/COUPLES | Mainstream | 90803.85 | 22699 | 3340 |
| | YOUNG FAMILIES | Premium | 84025.50 | 22406 | 2433 |
| | OLDER FAMILIES | Premium | 80658.40 | 21771 | 2273 |
| | YOUNG SINGLES/COUPLES | Budget | 61141.60 | 16671 | 3779 |
| | | Premium | 58432.65 | 15526 | 2431 |
| | MIDAGE SINGLES/COUPLES | Premium | 41642.10 | 11331 | 2574 |
| | | Budget | 35514.80 | 9496 | 1504 |
| | NEW FAMILIES | Budget | 21928.45 | 5571 | 1112 |
| | | Mainstream | 17013.90 | 4319 | 849 |
| | | Premium | 11491.10 | 2957 | 588 |

In []: df_cd_seg = df_cd_seg.merge(df_cd_seg_c, on=['LIFESTAGE', 'PREMIUM_CUSTOMER'], how=

In []: df_cd_seg['Sales_Per_Seg'] = df_cd_seg['TOT_SALES']/df_cd_seg['Customer_Counts']

In []: df_cd_seg['Chips_Per_Seg'] = df_cd_seg['TOT_SALES']/df_cd_seg['Customer_Counts']

```
In [ ]: df_cd_seg['avg_price'] = df_cd_seg['TOT_SALES']/df_cd_seg['PROD_QTY']
```

```
In [ ]: df_cd_seg
```

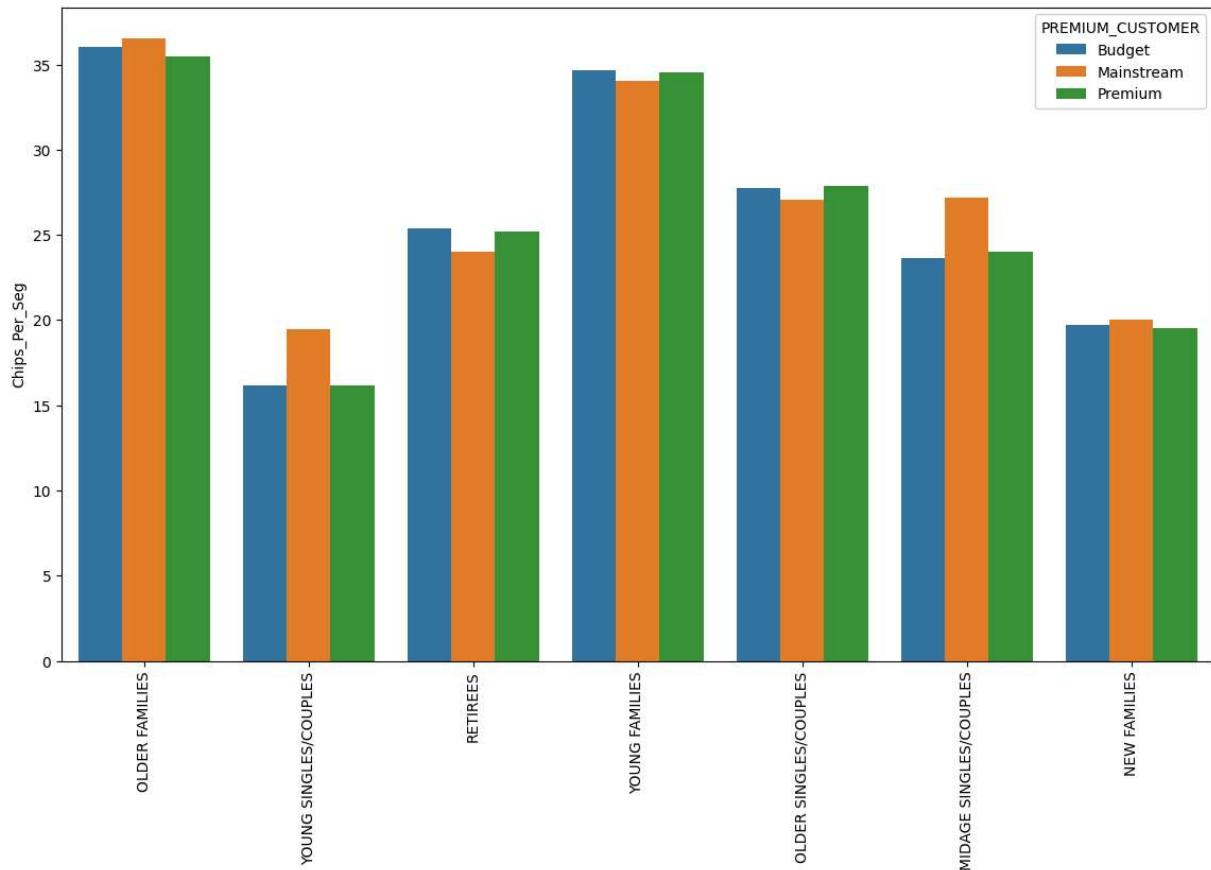
```
Out[ ]:
```

| | LIFESTAGE | PREMIUM_CUSTOMER | TOT_SALES | PROD_QTY | Cutomer_Counts | Sales_P |
|------------------------|------------|------------------|-----------|----------|----------------|---------|
| OLDER FAMILIES | Budget | 168363.25 | 45065 | 4675 | 9. | |
| YOUNG SINGLES/COUPLES | Mainstream | 157621.60 | 38632 | 8088 | 4. | |
| RETIREES | Mainstream | 155677.05 | 40518 | 6479 | 6. | |
| YOUNG FAMILIES | Budget | 139345.85 | 37111 | 4017 | 9. | |
| OLDER SINGLES/COUPLES | Budget | 136769.80 | 35220 | 4929 | 7. | |
| | Mainstream | 133393.80 | 34997 | 4930 | 7. | |
| | Premium | 132257.15 | 33984 | 4750 | 7. | |
| RETIREES | Budget | 113147.80 | 28764 | 4454 | 6. | |
| OLDER FAMILIES | Mainstream | 103445.55 | 27756 | 2831 | 9. | |
| RETIREES | Premium | 97646.05 | 24884 | 3872 | 6. | |
| YOUNG FAMILIES | Mainstream | 92788.75 | 25044 | 2728 | 9. | |
| MIDAGE SINGLES/COUPLES | Mainstream | 90803.85 | 22699 | 3340 | 6. | |
| YOUNG FAMILIES | Premium | 84025.50 | 22406 | 2433 | 9. | |
| OLDER FAMILIES | Premium | 80658.40 | 21771 | 2273 | 9. | |
| YOUNG SINGLES/COUPLES | Budget | 61141.60 | 16671 | 3779 | 4. | |
| MIDAGE SINGLES/COUPLES | Premium | 58432.65 | 15526 | 2431 | 6. | |
| YOUNG SINGLES/COUPLES | Premium | 41642.10 | 11331 | 2574 | 4. | |
| MIDAGE SINGLES/COUPLES | Budget | 35514.80 | 9496 | 1504 | 6. | |
| NEW FAMILIES | Budget | 21928.45 | 5571 | 1112 | 5. | |
| | Mainstream | 17013.90 | 4319 | 849 | 5. | |
| | Premium | 11491.10 | 2957 | 588 | 5. | |

```
In [ ]: plt.figure(figsize=(14,8))
sns.barplot(data=df_cd_seg, x='LIFESTAGE', y='Chips_Per_Seg', hue='PREMIUM_CUSTOMER')
```

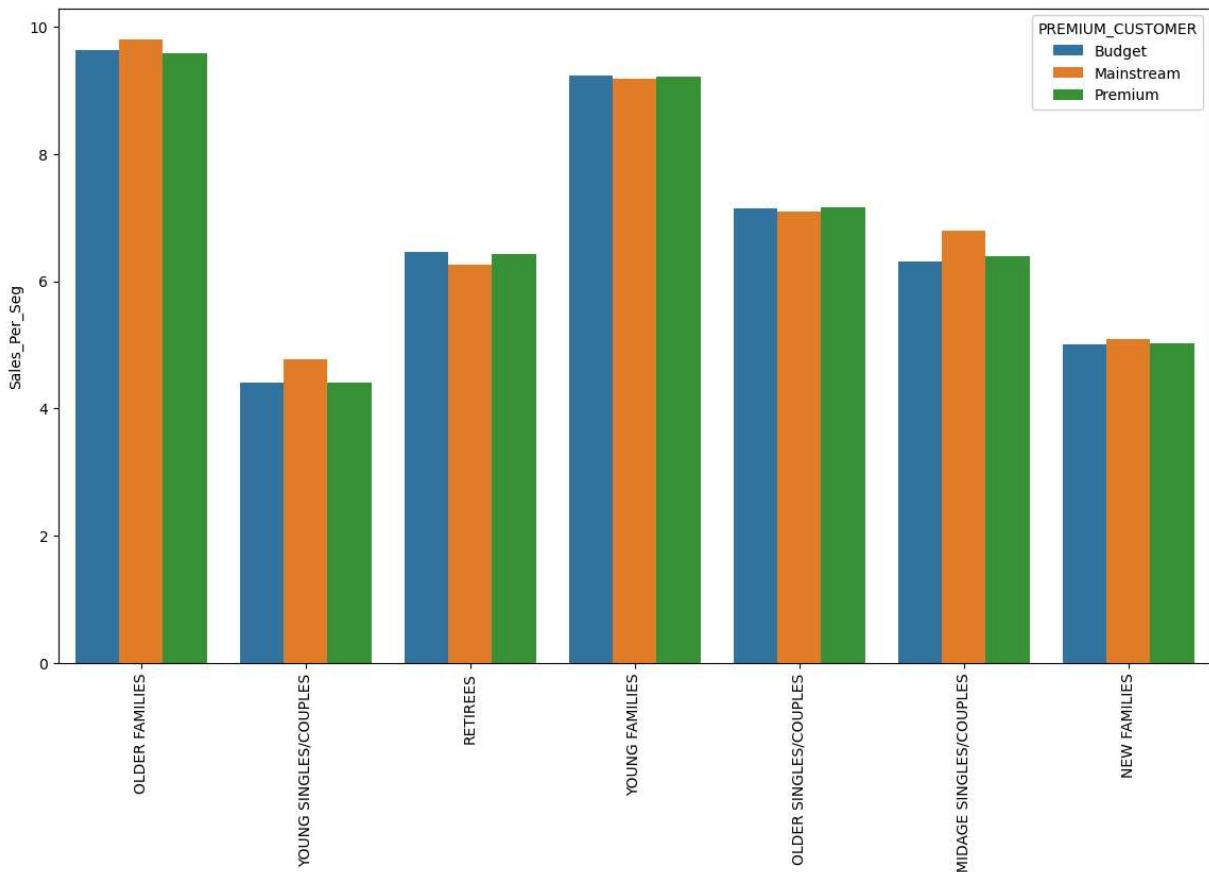
```
plt.xlabel(None)
plt.xticks(rotation=90)
```

Out[]: ([0, 1, 2, 3, 4, 5, 6],
[Text(0, 0, 'OLDER FAMILIES'),
Text(1, 0, 'YOUNG SINGLES/COUPLES'),
Text(2, 0, 'RETIREEES'),
Text(3, 0, 'YOUNG FAMILIES'),
Text(4, 0, 'OLDER SINGLES/COUPLES'),
Text(5, 0, 'MIDAGE SINGLES/COUPLES'),
Text(6, 0, 'NEW FAMILIES')])



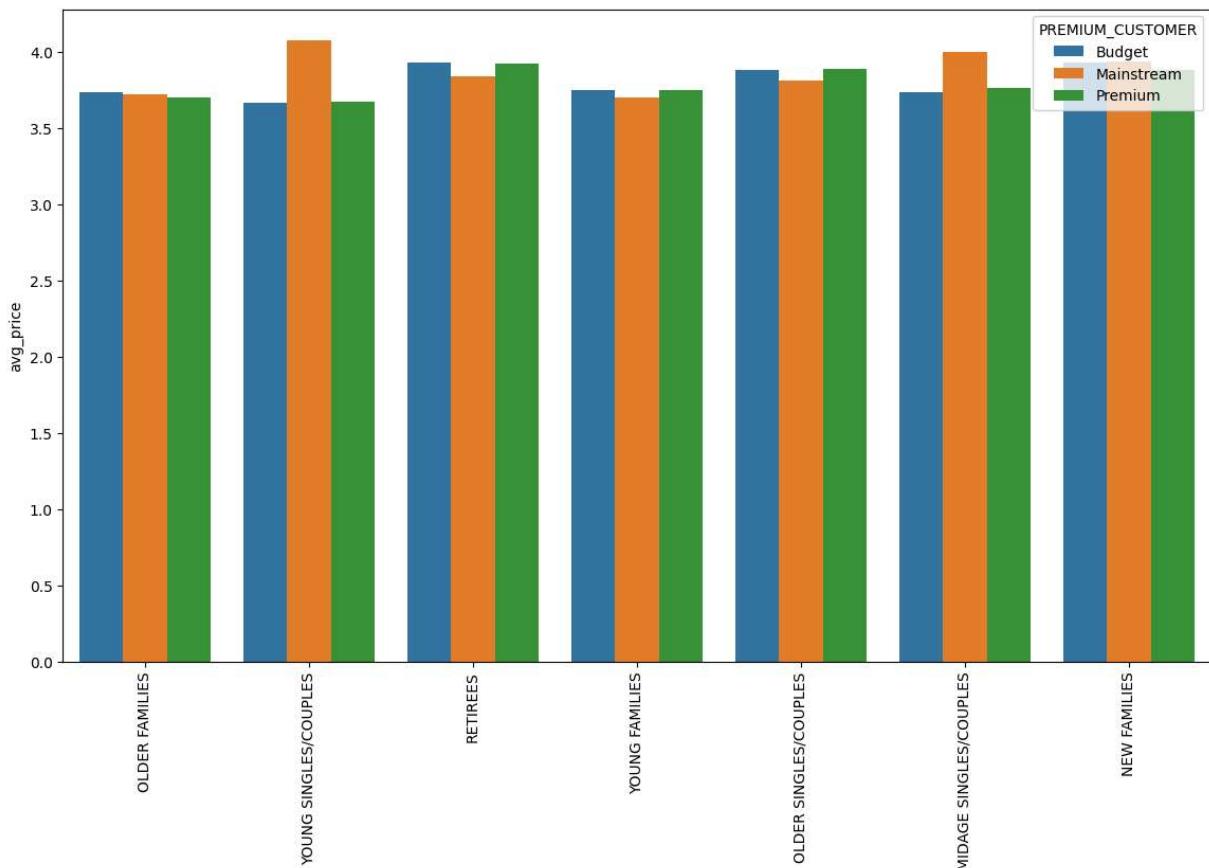
In []: plt.figure(figsize=(14,8))
sns.barplot(data=df_cd_seg, x='LIFESTAGE', y='Sales_Per_Seg', hue='PREMIUM_CUSTOMER')
plt.xlabel(None)
plt.xticks(rotation=90)

Out[]: ([0, 1, 2, 3, 4, 5, 6],
[Text(0, 0, 'OLDER FAMILIES'),
Text(1, 0, 'YOUNG SINGLES/COUPLES'),
Text(2, 0, 'RETIREEES'),
Text(3, 0, 'YOUNG FAMILIES'),
Text(4, 0, 'OLDER SINGLES/COUPLES'),
Text(5, 0, 'MIDAGE SINGLES/COUPLES'),
Text(6, 0, 'NEW FAMILIES')])



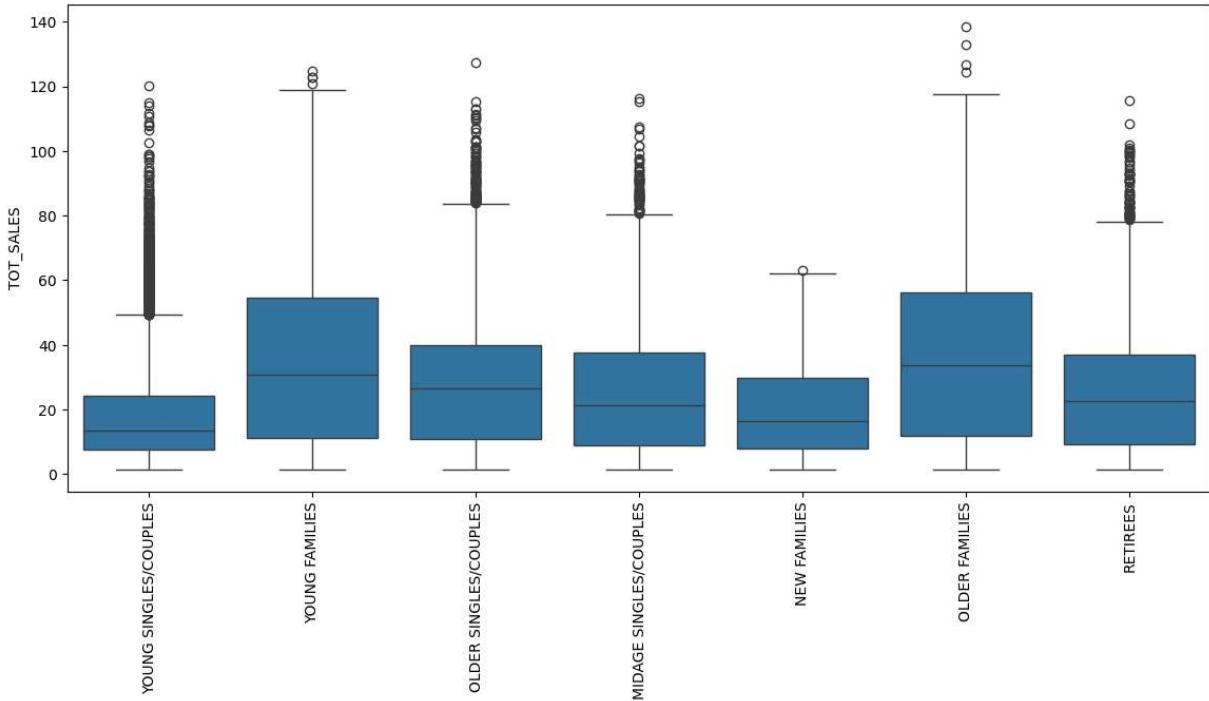
```
In [ ]: plt.figure(figsize=(14,8))
sns.barplot(data=df_cd_seg, x='LIFESTAGE', y='avg_price', hue='PREMIUM_CUSTOMER')
plt.xlabel(None)
plt.xticks(rotation=90)
```

```
Out[ ]: ([0, 1, 2, 3, 4, 5, 6],
 [Text(0, 0, 'OLDER FAMILIES'),
 Text(1, 0, 'YOUNG SINGLES/COUPLES'),
 Text(2, 0, 'RETIREES'),
 Text(3, 0, 'YOUNG FAMILIES'),
 Text(4, 0, 'OLDER SINGLES/COUPLES'),
 Text(5, 0, 'MIDAGE SINGLES/COUPLES'),
 Text(6, 0, 'NEW FAMILIES')])
```



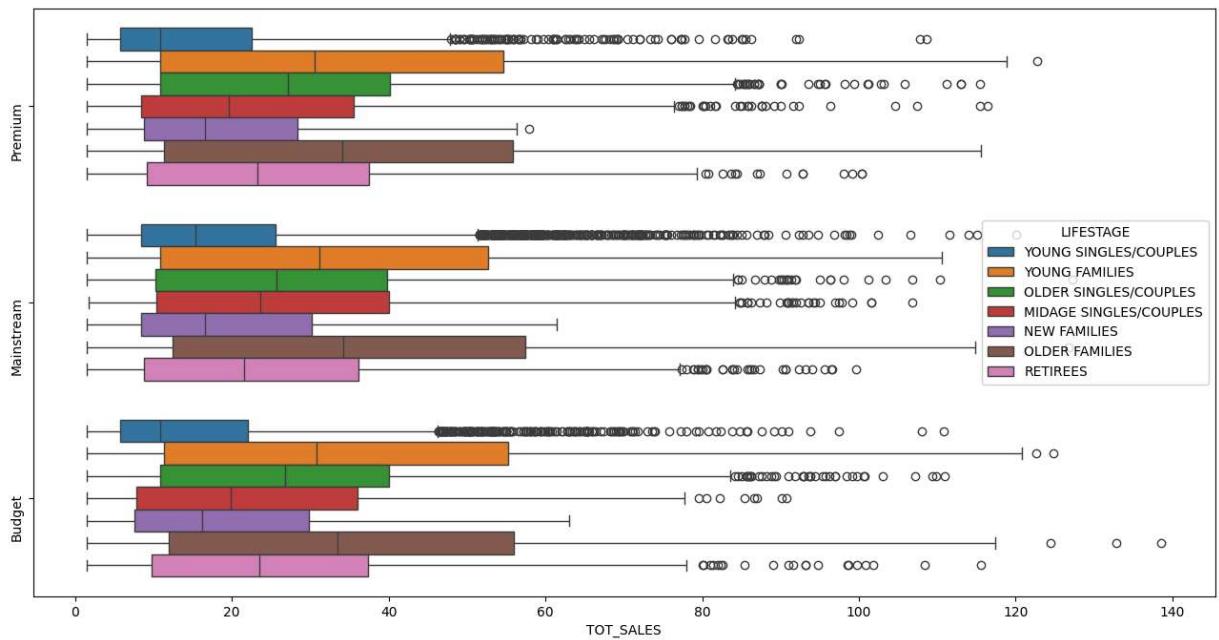
```
In [ ]: plt.figure(figsize=(14,6))
sns.boxplot(x=df_cd_merged['LIFESTAGE'], y=df_cd_merged['TOT_SALES'])
plt.xticks(rotation=90)
plt.xlabel(None)
```

Out[]: Text(0.5, 0, '')



```
In [ ]: plt.figure(figsize=(16,8))
sns.boxplot(x=df_cd_merged['TOT_SALES'], y=df_cd_merged['PREMIUM_CUSTOMER'], hue=df_cd_merged['LIFESTAGE'])
plt.yticks(rotation=90)
plt.ylabel(None)
```

Out[]: Text(0, 0.5, '')



In []: df_cd_seg

Out[]:

| LIFESTAGE | PREMIUM_CUSTOMER | TOT_SALES | PROD_QTY | CUSTOMER_COUNTS | Sales_P |
|-----------------------|------------------|-----------|----------|-----------------|---------|
| OLDER FAMILIES | Budget | 168363.25 | 45065 | 4675 | 9. |
| YOUNG SINGLES/COPLES | Mainstream | 157621.60 | 38632 | 8088 | 4. |
| RETIREES | Mainstream | 155677.05 | 40518 | 6479 | 6. |
| YOUNG FAMILIES | Budget | 139345.85 | 37111 | 4017 | 9. |
| OLDER SINGLES/COPLES | Budget | 136769.80 | 35220 | 4929 | 7. |
| | Mainstream | 133393.80 | 34997 | 4930 | 7. |
| RETIREES | Premium | 132257.15 | 33984 | 4750 | 7. |
| | Budget | 113147.80 | 28764 | 4454 | 6. |
| | Mainstream | 103445.55 | 27756 | 2831 | 9. |
| RETIREES | Premium | 97646.05 | 24884 | 3872 | 6. |
| YOUNG FAMILIES | Mainstream | 92788.75 | 25044 | 2728 | 9. |
| MIDAGE SINGLES/COPLES | Mainstream | 90803.85 | 22699 | 3340 | 6. |
| YOUNG FAMILIES | Premium | 84025.50 | 22406 | 2433 | 9. |
| OLDER FAMILIES | Premium | 80658.40 | 21771 | 2273 | 9. |
| YOUNG SINGLES/COPLES | Budget | 61141.60 | 16671 | 3779 | 4. |
| MIDAGE SINGLES/COPLES | Premium | 58432.65 | 15526 | 2431 | 6. |
| YOUNG SINGLES/COPLES | Premium | 41642.10 | 11331 | 2574 | 4. |
| MIDAGE SINGLES/COPLES | Budget | 35514.80 | 9496 | 1504 | 6. |
| NEW FAMILIES | Budget | 21928.45 | 5571 | 1112 | 5. |
| | Mainstream | 17013.90 | 4319 | 849 | 5. |
| | Premium | 11491.10 | 2957 | 588 | 5. |

In []: `from scipy.stats import ttest_ind`In []: `t_test_grp1 = df_cd_seg.query('LIFESTAGE == ["YOUNG SINGLES/COPLES", "MIDAGE SINGLE`In []: `t_test_grp2 = df_cd_seg.query('LIFESTAGE == ["YOUNG SINGLES/COPLES", "MIDAGE SINGLE`

```
In [ ]: ttest_ind(t_test_grp1, t_test_grp2)
```

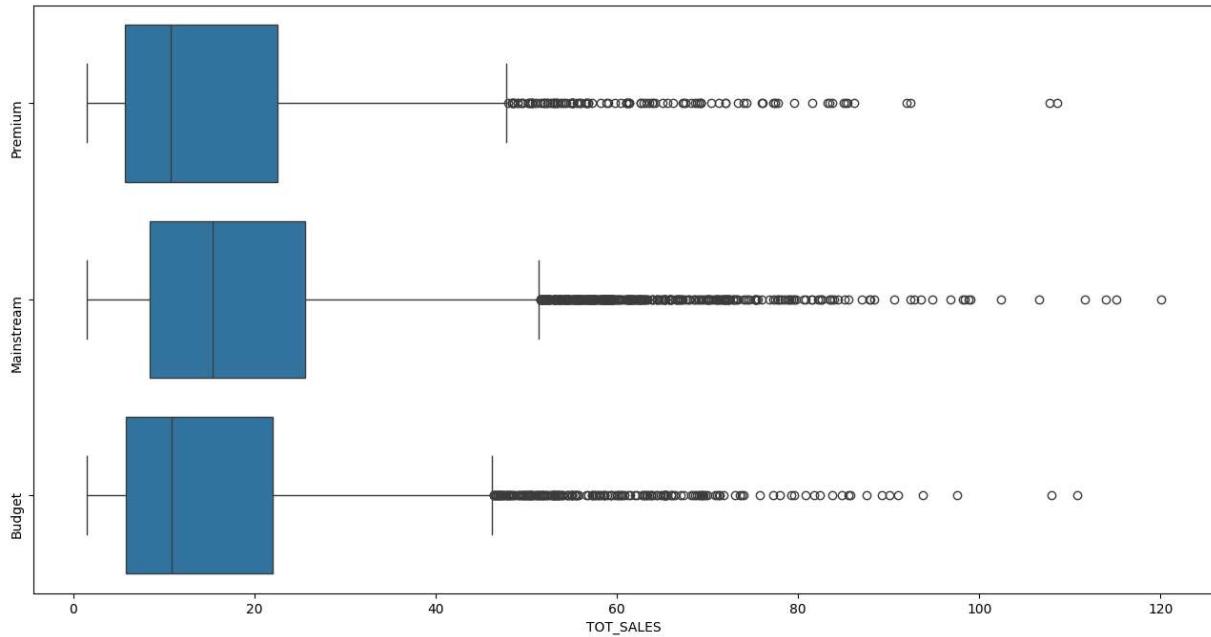
```
Out[ ]: TtestResult(statistic=array([7.60688692]), pvalue=array([0.00160278]), df=array([4.]))
```

P_value = 0.0016

Our Initial hypothesis of the difference between the 2 sample groups (1.YOUNG SINGLES/COUPLES","MIDAGE SINGLES/COUPLES : Mainstream, 2.YOUNG SINGLES/COUPLES","MIDAGE SINGLES/COUPLES:Budget & Premium) is significantly high is being REJECTED.

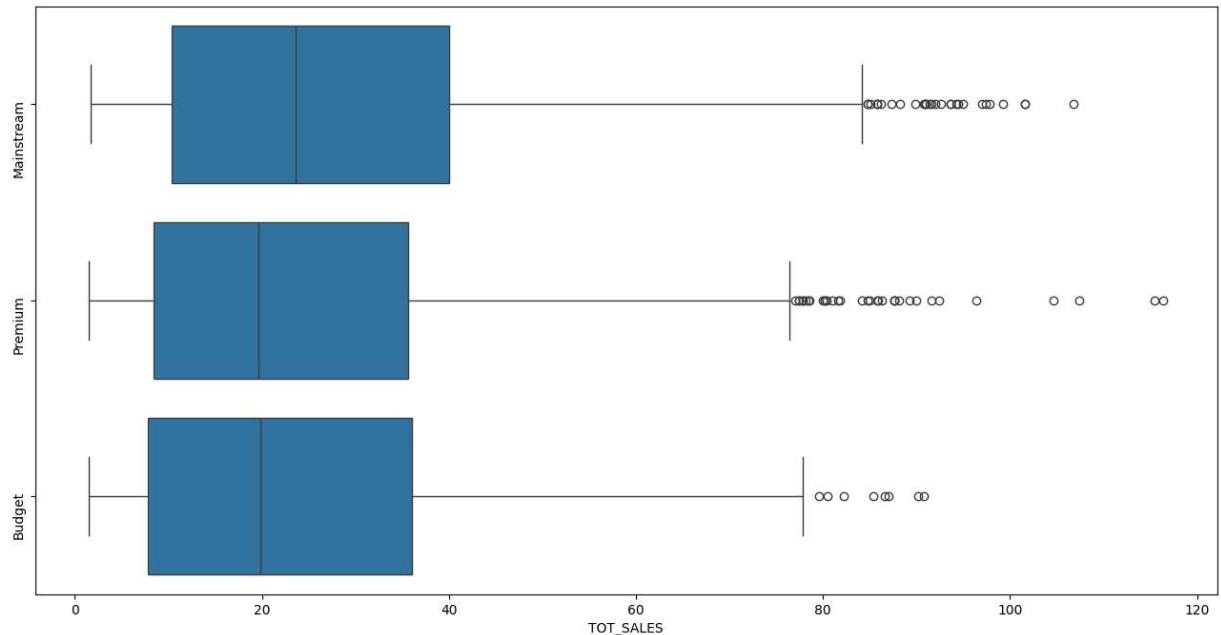
```
In [ ]: plt.figure(figsize=(16,8))
sns.boxplot(data=df_cd_merged.query('LIFESTAGE == "YOUNG SINGLES/COUPLES"'), x='TOT_SALES',
plt.yticks(rotation=90)
plt.ylabel(None)
```

```
Out[ ]: Text(0, 0.5, '')
```



```
In [ ]: plt.figure(figsize=(16,8))
sns.boxplot(data=df_cd_merged.query('LIFESTAGE == "MIDAGE SINGLES/COUPLES"'), x='TOT_SALES',
plt.yticks(rotation=90)
plt.ylabel(None)
```

```
Out[ ]: Text(0, 0.5, '')
```



```
In [ ]: df_merged
```

Out[]:

| | DATE | STORE_NBR | LYLTY_CARD_NBR | TXN_ID | PROD_NBR | PROD_NAME | PROD_C |
|--------|------------|-----------|----------------|---------|----------|-----------|-------------------------------------|
| 0 | 2018-10-17 | 1 | | 1000 | 1 | 5 | Natural Chip Comnpy SeaSalt |
| 1 | 2018-09-16 | 1 | | 1002 | 2 | 58 | Red Rock Deli Chikn & Garlic Aioli |
| 2 | 2019-03-07 | 1 | | 1003 | 3 | 52 | Grain Waves Sour Cream & Chives |
| 3 | 2019-03-08 | 1 | | 1003 | 4 | 106 | Natural ChipCo Hony Soy Chckn |
| 4 | 2018-11-02 | 1 | | 1004 | 5 | 96 | WW Original Stacked Chips |
| ... | ... | ... | | ... | ... | ... | ... |
| 264829 | 2018-10-01 | 88 | | 2370751 | 240394 | 60 | Kettle Tortilla ChpsFeta & Garlic |
| 264830 | 2018-10-24 | 88 | | 2370961 | 240480 | 70 | Tyrrells Crisps Lightly Salted |
| 264831 | 2018-10-27 | 88 | | 2370961 | 240481 | 65 | Old El Paso Salsa Dip Chnky Tom Ht |
| 264832 | 2018-12-14 | 88 | | 2373711 | 241815 | 16 | Smiths Crinkle Chips Salt & Vinegar |
| 264833 | 2018-12-25 | 0 | | 0 | 0 | 0 | NaN |

264834 rows × 12 columns



In []: df_merged.query('LIFESTAGE == "YOUNG SINGLES/COUPLES" and PREMIUM_CUSTOMER == "Main

Out[]:

| | BRAND | count |
|----|-------------|-------|
| 0 | Kettle | 3844 |
| 1 | Doritos | 2541 |
| 2 | Pringles | 2315 |
| 3 | Smith | 1988 |
| 4 | Infuzions | 1250 |
| 5 | Thins | 1166 |
| 6 | RRD | 969 |
| 7 | Twisties | 900 |
| 8 | Tostitos | 890 |
| 9 | Old | 867 |
| 10 | Cobs | 864 |
| 11 | Grain Waves | 646 |
| 12 | Tyrrells | 619 |
| 13 | Woolworths | 599 |
| 14 | NCC | 394 |
| 15 | Cheezels | 346 |
| 16 | CCs | 222 |
| 17 | Cheetos | 166 |
| 18 | Sunbites | 128 |
| 19 | French | 78 |
| 20 | Burger | 62 |

In []: df_merged.groupby('PACK_SIZE')['BRAND'].value_counts()

```
Out[ ]: PACK_SIZE BRAND
70.0      Infuzions    1507
90.0      Sunbites     3008
110.0     Infuzions   12694
          Cobs        9693
125.0     Cheezels     1454
134.0     Pringles    25102
135.0     Kettle       3257
150.0     Kettle      19009
          RRD         10372
          Smith       7374
          Doritos     6376
160.0     Woolworths   2970
165.0     RRD         7407
          Tyrrells    6442
          Cheetos     1448
170.0     Doritos     12613
          Smith       7370
175.0     Kettle      19022
          Thins        14075
          Tostitos    9471
          Smith       7506
          NCC          7469
          CCs          4551
          Woolworths   2877
          French      1418
180.0     Grain Waves  1468
190.0     Woolworths   1516
          Cheetos     1479
200.0     Woolworths   4473
210.0     Grain Waves  6272
220.0     Burger       1564
250.0     Twisties     3169
270.0     Twisties     6285
300.0     Old          9324
          Doritos     2921
          Woolworths   2921
330.0     Smith       6339
          Cheezels    3149
          Doritos     3052
380.0     Smith       3233
          Doritos     3183
```

Name: count, dtype: int64

```
In [ ]: df_merged.query('LIFESTAGE == "YOUNG SINGLES/COUPLES" and PREMIUM_CUSTOMER == "Main
```

Out[]: **BRAND count**

| 0 | Smith | 928 |
|----------|------------|-----|
| 1 | Old | 867 |
| 2 | Doritos | 768 |
| 3 | Cheezels | 287 |
| 4 | Woolworths | 120 |

In []: df_merged.query('LIFESTAGE == "YOUNG SINGLES/COUPLES" and PREMIUM_CUSTOMER == "Budget"')

Out[]: **BRAND count**

| 0 | Doritos | 315 |
|----------|------------|-----|
| 1 | Old | 279 |
| 2 | Smith | 261 |
| 3 | Woolworths | 159 |
| 4 | Cheezels | 85 |

In []: df_merged.query('LIFESTAGE == "YOUNG SINGLES/COUPLES" and PREMIUM_CUSTOMER == "Premium"')

Out[]: **BRAND count**

| 0 | Doritos | 206 |
|----------|------------|-----|
| 1 | Smith | 203 |
| 2 | Old | 178 |
| 3 | Woolworths | 105 |
| 4 | Cheezels | 65 |

In []: df_merged.query('LIFESTAGE == "OLDER FAMILIES" and PREMIUM_CUSTOMER == "Mainstream"')

Out[]: BRAND

| | |
|------------|-----|
| Doritos | 512 |
| Smith | 486 |
| Old | 474 |
| Woolworths | 173 |
| Cheezels | 133 |

Name: count, dtype: int64

In []: df_merged.query('LIFESTAGE == "OLDER FAMILIES" and PREMIUM_CUSTOMER == "Budget" and

```
Out[ ]: BRAND
Doritos      826
Smith        820
Old          758
Cheezels     275
Woolworths   275
Name: count, dtype: int64
```

```
In [ ]: df_merged.query('LIFESTAGE == "OLDER FAMILIES" and PREMIUM_CUSTOMER == "Premium" an
```

```
Out[ ]: BRAND
Smith        388
Doritos     376
Old          357
Woolworths   145
Cheezels     108
Name: count, dtype: int64
```

```
In [ ]: df_merged.groupby(['PACK_SIZE', 'BRAND'])[['LIFESTAGE', 'PREMIUM_CUSTOMER']].value_c
```

| Out[]: | PACK_SIZE | BRAND | LIFESTAGE | PREMIUM_CUSTOMER |
|---------|-----------|-------|------------------------|------------------|
| 330.0 | Cheezels | | OLDER SINGLES/COUPLES | Budget 223 |
| | | | RETIREEES | Budget 201 |
| | | | OLDER SINGLES/COUPLES | Mainstream 190 |
| | | | MIDAGE SINGLES/COUPLES | Mainstream 186 |
| | | | RETIREEES | Premium 149 |
| | | | YOUNG FAMILIES | Mainstream 145 |
| | | | | Premium 142 |
| | | | OLDER FAMILIES | Mainstream 133 |
| | | | | Premium 108 |
| | | | MIDAGE SINGLES/COUPLES | Premium 102 |
| | | | YOUNG SINGLES/COUPLES | Budget 85 |
| | | | | Premium 65 |
| | | | MIDAGE SINGLES/COUPLES | Budget 47 |
| | | | NEW FAMILIES | Mainstream 28 |
| | | | | Budget 26 |
| | | | | Premium 22 |
| Doritos | | | YOUNG SINGLES/COUPLES | Mainstream 303 |
| | | | OLDER FAMILIES | Budget 259 |
| | | | RETIREEES | Mainstream 240 |
| | | | OLDER SINGLES/COUPLES | Budget 224 |
| | | | YOUNG FAMILIES | Budget 217 |
| | | | OLDER SINGLES/COUPLES | Mainstream 211 |
| | | | RETIREEES | Budget 204 |
| | | | OLDER SINGLES/COUPLES | Premium 197 |
| | | | OLDER FAMILIES | Mainstream 164 |
| | | | RETIREEES | Premium 162 |
| | | | MIDAGE SINGLES/COUPLES | Mainstream 137 |
| | | | YOUNG FAMILIES | Mainstream 130 |
| | | | | Premium 115 |
| | | | OLDER FAMILIES | Premium 111 |
| | | | YOUNG SINGLES/COUPLES | Budget 100 |
| | | | MIDAGE SINGLES/COUPLES | Premium 83 |
| Smith | | | | Budget 55 |
| | | | YOUNG SINGLES/COUPLES | Premium 54 |
| | | | NEW FAMILIES | Budget 42 |
| | | | | Mainstream 29 |
| | | | | Premium 15 |
| | | | YOUNG SINGLES/COUPLES | Mainstream 605 |
| | | | OLDER FAMILIES | Budget 558 |
| | | | RETIREEES | Mainstream 509 |
| | | | OLDER SINGLES/COUPLES | Premium 458 |
| | | | YOUNG FAMILIES | Budget 448 |
| | | | OLDER SINGLES/COUPLES | Budget 445 |
| | | | | Mainstream 404 |
| | | | RETIREEES | Budget 389 |
| | | | MIDAGE SINGLES/COUPLES | Mainstream 345 |
| | | | RETIREEES | Premium 324 |
| | | | OLDER FAMILIES | Mainstream 322 |
| | | | YOUNG FAMILIES | Mainstream 285 |
| | | | OLDER FAMILIES | Premium 275 |
| | | | YOUNG FAMILIES | Premium 246 |
| | | | YOUNG SINGLES/COUPLES | Budget 176 |
| | | | MIDAGE SINGLES/COUPLES | Premium 160 |
| | | | YOUNG SINGLES/COUPLES | Premium 128 |
| | | | MIDAGE SINGLES/COUPLES | Budget 105 |

| | | | | |
|-------|---------|------------------------|------------|-----|
| | | NEW FAMILIES | Budget | 70 |
| | | | Mainstream | 46 |
| | | | Premium | 41 |
| 380.0 | Doritos | YOUNG SINGLES/COUPLES | Mainstream | 303 |
| | | RETIREES | Mainstream | 251 |
| | | OLDER FAMILIES | Budget | 248 |
| | | OLDER SINGLES/COUPLES | Premium | 236 |
| | | | Mainstream | 232 |
| | | | Budget | 216 |
| | | YOUNG FAMILIES | Budget | 216 |
| | | RETIREES | Premium | 191 |
| | | | Budget | 177 |
| | | OLDER FAMILIES | Mainstream | 162 |
| | | MIDAGE SINGLES/COUPLES | Mainstream | 138 |
| | | YOUNG FAMILIES | Premium | 136 |
| | | | Mainstream | 132 |
| | | OLDER FAMILIES | Premium | 123 |
| | | MIDAGE SINGLES/COUPLES | Premium | 102 |
| | | YOUNG SINGLES/COUPLES | Budget | 98 |
| | | | Premium | 71 |
| | | MIDAGE SINGLES/COUPLES | Budget | 60 |
| | | NEW FAMILIES | Budget | 41 |
| | | | Mainstream | 35 |
| | | | Premium | 15 |
| | Smith | YOUNG SINGLES/COUPLES | Mainstream | 323 |
| | | OLDER FAMILIES | Budget | 262 |
| | | RETIREES | Mainstream | 246 |
| | | OLDER SINGLES/COUPLES | Budget | 240 |
| | | | Mainstream | 223 |
| | | YOUNG FAMILIES | Budget | 215 |
| | | OLDER SINGLES/COUPLES | Premium | 209 |
| | | RETIREES | Budget | 198 |
| | | | Premium | 178 |
| | | MIDAGE SINGLES/COUPLES | Mainstream | 172 |
| | | OLDER FAMILIES | Mainstream | 164 |
| | | YOUNG FAMILIES | Mainstream | 151 |
| | | | Premium | 129 |
| | | OLDER FAMILIES | Premium | 113 |
| | | MIDAGE SINGLES/COUPLES | Premium | 102 |
| | | YOUNG SINGLES/COUPLES | Budget | 85 |
| | | | Premium | 75 |
| | | MIDAGE SINGLES/COUPLES | Budget | 54 |
| | | NEW FAMILIES | Budget | 41 |
| | | | Mainstream | 30 |
| | | | Premium | 23 |

Name: count, dtype: int64

```
In [ ]: df_merged.groupby('BRAND')['BRAND'].value_counts().sort_values(ascending=False)
```

```
Out[ ]: BRAND
Kettle      41288
Smith       31822
Doritos    28145
Pringles   25102
RRD        17779
Woolworths 14757
Infuzions  14201
Thins       14075
Cobs        9693
Tostitos   9471
Twisties    9454
Old         9324
Grain Waves 7740
NCC         7469
Tyrrells   6442
Cheezels   4603
CCs         4551
Sunbites   3008
Cheetos    2927
Burger     1564
French     1418
Name: count, dtype: int64
```

```
In [ ]: df_merged.groupby('PACK_SIZE')['PACK_SIZE'].value_counts().sort_values(ascending=False)
```

```
Out[ ]: PACK_SIZE
175.0    66389
150.0    43131
134.0    25102
110.0    22387
170.0    19983
165.0    15297
300.0    15166
330.0    12540
380.0    6416
270.0    6285
210.0    6272
200.0    4473
135.0    3257
250.0    3169
90.0     3008
190.0    2995
160.0    2970
220.0    1564
70.0     1507
180.0    1468
125.0    1454
Name: count, dtype: int64
```

Task 1 Finished with

Young Families & Old Families with Highest Spending segment

Mainstream: YOUNG SINGLES/COUPLES & MIDAGE SINGLES/COUPLES with High Avg_Price_Per_Chips

Kettle,Smith & Doritos as the Most sold Brands

175g Most sold Pack Size

Smith & Doritos Highest sold larger packs chips

```
In [ ]: pd.set_option('display.max_rows', None)
```

```
In [ ]: df.query('STORE_NBR == 77 | STORE_NBR == 86 | STORE_NBR == 88').count()
```

```
Out[ ]: DATE            3974  
STORE_NBR        3974  
LYLTY_CARD_NBR  3974  
TXN_ID           3974  
PROD_NBR         3974  
PROD_NAME        3974  
PROD_QTY          3974  
TOT_SALES        3974  
PACK_SIZE         3974  
BRAND            3974  
LIFESTAGE        3974  
PREMIUM_CUSTOMER 3974  
dtype: int64
```

```
In [ ]: df.query('STORE_NBR == 77 | STORE_NBR == 86 | STORE_NBR == 88').groupby(['BRAND'])[[
```

Out[]:

PROD_QTY TOT_SALES

| BRAND | | |
|--------------------|------|--------|
| Kettle | 1337 | 6601.0 |
| Pringles | 773 | 2860.1 |
| Doritos | 768 | 3452.7 |
| Smith | 765 | 3011.2 |
| Infuzions | 475 | 1732.2 |
| Thins | 455 | 1501.5 |
| RRD | 423 | 1194.3 |
| Cobs | 357 | 1356.6 |
| Woolworths | 333 | 595.5 |
| Tostitos | 307 | 1350.8 |
| Twisties | 297 | 1338.9 |
| Old | 257 | 1310.7 |
| Grain Waves | 215 | 753.5 |
| Tyrrells | 210 | 882.0 |
| NCC | 196 | 588.0 |
| Cheezels | 134 | 652.2 |
| CCs | 120 | 252.0 |
| Sunbites | 77 | 130.9 |
| Cheetos | 65 | 203.5 |
| Burger | 50 | 115.0 |
| French | 42 | 126.0 |

In []: df.query('STORE_NBR == 77').groupby(['BRAND'])[['PROD_QTY', 'TOT_SALES']].sum().sort

Out[]:

PROD_QTY TOT_SALES

| BRAND | | |
|--------------------|-----|-------|
| Smith | 134 | 446.7 |
| Kettle | 108 | 537.2 |
| RRD | 81 | 225.6 |
| Woolworths | 79 | 143.2 |
| Pringles | 66 | 244.2 |
| Doritos | 62 | 271.5 |
| Thins | 44 | 145.2 |
| NCC | 37 | 111.0 |
| Infuzions | 33 | 118.4 |
| CCs | 32 | 67.2 |
| Grain Waves | 31 | 111.1 |
| Tostitos | 28 | 123.2 |
| Old | 23 | 117.3 |
| Cobs | 19 | 72.2 |
| Twisties | 18 | 79.5 |
| Tyrrells | 16 | 67.2 |
| Sunbites | 16 | 27.2 |
| Burger | 15 | 34.5 |
| Cheezels | 15 | 53.1 |
| French | 10 | 30.0 |
| Cheetos | 5 | 14.5 |

In []: df.query('STORE_NBR == 77').groupby(['LIFESTAGE'])[['PROD_QTY', 'TOT_SALES']].sum().

Out[]:

PROD_QTY TOT_SALES**LIFESTAGE**

| YOUNG SINGLES/COUPLES | 235 | 799.5 |
|-------------------------------|-----|-------|
| RETIREES | 209 | 714.8 |
| OLDER SINGLES/COUPLES | 155 | 561.1 |
| OLDER FAMILIES | 99 | 348.2 |
| YOUNG FAMILIES | 74 | 265.4 |
| MIDAGE SINGLES/COUPLES | 57 | 205.0 |
| NEW FAMILIES | 43 | 146.0 |

In []: df.query('STORE_NBR == 77').groupby(['LIFESTAGE', 'BRAND'])[['PROD_QTY', 'TOT_SALES']]

Out[]:

| LIFESTAGE | BRAND | PROD_QTY | TOT_SALES |
|-----------------------|------------|----------|-----------|
| RETIREES | Smith | 38 | 121.9 |
| YOUNG SINGLES/COUPLES | Smith | 35 | 105.7 |
| OLDER SINGLES/COUPLES | Smith | 29 | 107.0 |
| YOUNG SINGLES/COUPLES | RRD | 27 | 75.0 |
| | Woolworths | 25 | 45.0 |
| | Doritos | 23 | 100.1 |
| RETIREES | Kettle | 23 | 113.0 |
| YOUNG SINGLES/COUPLES | Kettle | 22 | 112.8 |
| OLDER SINGLES/COUPLES | Kettle | 21 | 105.4 |
| RETIREES | NCC | 20 | 60.0 |
| | RRD | 20 | 56.7 |
| YOUNG SINGLES/COUPLES | Pringles | 19 | 70.3 |
| OLDER FAMILIES | Kettle | 17 | 83.8 |
| OLDER SINGLES/COUPLES | RRD | 16 | 44.4 |
| YOUNG SINGLES/COUPLES | Thins | 14 | 46.2 |
| OLDER FAMILIES | Woolworths | 14 | 25.2 |
| | Pringles | 13 | 48.1 |
| YOUNG FAMILIES | Pringles | 12 | 44.4 |
| YOUNG SINGLES/COUPLES | Infuzions | 12 | 40.0 |
| RETIREES | Doritos | 12 | 51.1 |

In []: df.query('STORE_NBR == 86').groupby(['BRAND'])[['PROD_QTY', 'TOT_SALES']].sum().sort

Out[]:

PROD_QTY TOT_SALES

| BRAND | | |
|--------------------|-----|---------|
| Smith | 445 | 1493.50 |
| Kettle | 381 | 1875.40 |
| RRD | 342 | 968.70 |
| Woolworths | 254 | 452.30 |
| Doritos | 240 | 1006.15 |
| Pringles | 187 | 691.90 |
| NCC | 159 | 477.00 |
| Thins | 147 | 485.10 |
| Infuzions | 140 | 466.20 |
| CCs | 88 | 184.80 |
| Cobs | 88 | 334.40 |
| Tostitos | 80 | 352.00 |
| Twisties | 79 | 355.60 |
| Grain Waves | 78 | 260.80 |
| Sunbites | 61 | 103.70 |
| Tyrrells | 61 | 256.20 |
| Cheetos | 60 | 189.00 |
| Old | 60 | 306.00 |
| Cheezels | 49 | 200.10 |
| Burger | 35 | 80.50 |
| French | 32 | 96.00 |

In []: df.query('STORE_NBR == 86').groupby(['LIFESTAGE'])[['PROD_QTY', 'TOT_SALES']].sum().

Out[]:

PROD_QTY TOT_SALES

| LIFESTAGE | PROD_QTY | TOT_SALES |
|-------------------------------|----------|-----------|
| OLDER FAMILIES | 744 | 2534.90 |
| OLDER SINGLES/COUPLES | 613 | 2146.70 |
| YOUNG FAMILIES | 554 | 1913.35 |
| RETIREES | 510 | 1778.10 |
| YOUNG SINGLES/COUPLES | 297 | 1036.90 |
| MIDAGE SINGLES/COUPLES | 249 | 871.10 |
| NEW FAMILIES | 99 | 354.30 |

In []: df.query('STORE_NBR == 86').groupby(['LIFESTAGE', 'BRAND'])[['PROD_QTY', 'TOT_SALES']]

Out[]:

| LIFESTAGE | BRAND | PROD_QTY | TOT_SALES |
|------------------------|------------|----------|-----------|
| OLDER FAMILIES | Smith | 113 | 378.90 |
| OLDER SINGLES/COUPLES | Smith | 100 | 318.40 |
| YOUNG FAMILIES | Smith | 91 | 317.00 |
| OLDER FAMILIES | RRD | 87 | 246.60 |
| RETIREES | Kettle | 79 | 389.00 |
| YOUNG FAMILIES | RRD | 76 | 213.60 |
| OLDER FAMILIES | Woolworths | 74 | 129.80 |
| | Kettle | 72 | 355.20 |
| OLDER SINGLES/COUPLES | Kettle | 72 | 354.00 |
| | Doritos | 65 | 277.70 |
| OLDER FAMILIES | Doritos | 61 | 257.40 |
| RETIREES | RRD | 60 | 171.60 |
| YOUNG FAMILIES | Kettle | 59 | 286.60 |
| RETIREES | Smith | 54 | 188.20 |
| | Woolworths | 53 | 96.20 |
| OLDER SINGLES/COUPLES | RRD | 47 | 132.90 |
| YOUNG SINGLES/COUPLES | Kettle | 42 | 210.00 |
| OLDER SINGLES/COUPLES | Thins | 42 | 138.60 |
| MIDAGE SINGLES/COUPLES | Smith | 41 | 144.70 |
| YOUNG FAMILIES | Doritos | 41 | 170.85 |

In []: df.query('STORE_NBR == 88').groupby(['BRAND'])[['PROD_QTY', 'TOT_SALES']].sum().sort

Out[]:

PROD_QTY TOT_SALES

| BRAND | PROD_QTY | TOT_SALES |
|--------------------|----------|-----------|
| Kettle | 848 | 4188.40 |
| Pringles | 520 | 1924.00 |
| Doritos | 466 | 2175.05 |
| Infuzions | 302 | 1147.60 |
| Thins | 264 | 871.20 |
| Cobs | 250 | 950.00 |
| Twisties | 200 | 903.80 |
| Tostitos | 199 | 875.60 |
| Smith | 186 | 1071.00 |
| Old | 174 | 887.40 |
| Tyrrells | 133 | 558.60 |
| Grain Waves | 106 | 381.60 |
| Cheezels | 70 | 399.00 |

In []: df.query('STORE_NBR == 88').groupby(['LIFESTAGE'])[['PROD_QTY', 'TOT_SALES']].sum()

Out[]:

PROD_QTY TOT_SALES

| LIFESTAGE | PROD_QTY | TOT_SALES |
|-------------------------------|----------|-----------|
| OLDER SINGLES/COUPLES | 952 | 4187.65 |
| OLDER FAMILIES | 735 | 3232.80 |
| RETIREES | 681 | 2993.90 |
| YOUNG FAMILIES | 580 | 2538.30 |
| MIDAGE SINGLES/COUPLES | 409 | 1800.20 |
| YOUNG SINGLES/COUPLES | 311 | 1357.40 |
| NEW FAMILIES | 50 | 223.00 |

In []: df.query('STORE_NBR == 88').groupby(['LIFESTAGE', 'BRAND'])[['PROD_QTY', 'TOT_SALES']]

Out[]:

| LIFESTAGE | BRAND | PROD_QTY | TOT_SALES |
|------------------------|-----------|----------|-----------|
| OLDER SINGLES/COUPLES | Kettle | 211 | 1037.80 |
| OLDER FAMILIES | Kettle | 175 | 870.60 |
| RETIREES | Kettle | 146 | 720.40 |
| OLDER SINGLES/COUPLES | Pringles | 138 | 510.60 |
| YOUNG FAMILIES | Kettle | 125 | 607.40 |
| OLDER SINGLES/COUPLES | Doritos | 125 | 585.05 |
| OLDER FAMILIES | Doritos | 100 | 471.20 |
| MIDAGE SINGLES/COUPLES | Kettle | 96 | 486.40 |
| OLDER SINGLES/COUPLES | Infuzions | 94 | 357.20 |
| RETIREES | Pringles | 92 | 340.40 |
| OLDER FAMILIES | Pringles | 89 | 329.30 |
| RETIREES | Doritos | 80 | 353.40 |
| YOUNG FAMILIES | Pringles | 80 | 296.00 |
| YOUNG SINGLES/COUPLES | Kettle | 79 | 387.40 |
| YOUNG FAMILIES | Doritos | 74 | 358.30 |
| MIDAGE SINGLES/COUPLES | Pringles | 69 | 255.30 |
| OLDER FAMILIES | Thins | 67 | 221.10 |
| OLDER SINGLES/COUPLES | Thins | 64 | 211.20 |
| RETIREES | Infuzions | 58 | 220.40 |
| OLDER SINGLES/COUPLES | Tostitos | 56 | 246.40 |

In []: df.head()

Out[]:

| | DATE | STORE_NBR | LYLTY_CARD_NBR | TXN_ID | PROD_NBR | PROD_NAME | PROD_QTY | T |
|---|------------|-----------|----------------|--------|----------|-----------|------------------------------------|---|
| 0 | 2018-10-17 | 1 | | 1000 | 1 | 5 | Natural Chip Comnpy SeaSalt | 2 |
| 1 | 2018-09-16 | 1 | | 1002 | 2 | 58 | Red Rock Deli Chikn & Garlic Aioli | 1 |
| 2 | 2019-03-07 | 1 | | 1003 | 3 | 52 | Grain Waves Sour Cream & Chives | 1 |
| 3 | 2019-03-08 | 1 | | 1003 | 4 | 106 | Natural ChipCo Honey Soy Chckn | 1 |
| 4 | 2018-11-02 | 1 | | 1004 | 5 | 96 | WW Original Stacked Chips | 1 |

In []: `df['DATE'].agg(['min', 'max'])`

Out[]: min 2018-07-01
max 2019-06-30
Name: DATE, dtype: datetime64[ns]

In []: `df.groupby('STORE_NBR')['TOT_SALES'].sum().sort_values(ascending=False).reset_index`

Out[]:

| | STORE_NBR | TOT_SALES |
|----|-----------|-----------|
| 0 | 226 | 17605.45 |
| 1 | 88 | 16333.25 |
| 2 | 165 | 15973.75 |
| 3 | 40 | 15559.50 |
| 4 | 237 | 15539.50 |
| 5 | 58 | 15251.45 |
| 6 | 199 | 14797.00 |
| 7 | 4 | 14647.65 |
| 8 | 203 | 14551.60 |
| 9 | 26 | 14469.30 |
| 10 | 81 | 14361.95 |
| 11 | 201 | 14298.70 |
| 12 | 130 | 14289.65 |
| 13 | 72 | 14204.40 |
| 14 | 181 | 14108.45 |
| 15 | 217 | 13993.60 |
| 16 | 95 | 13915.50 |
| 17 | 261 | 13859.75 |
| 18 | 238 | 13708.40 |
| 19 | 210 | 13689.25 |
| 20 | 123 | 13468.40 |
| 21 | 125 | 13352.85 |
| 22 | 49 | 13287.70 |
| 23 | 75 | 13089.60 |
| 24 | 153 | 13042.05 |
| 25 | 194 | 13035.95 |
| 26 | 65 | 13015.05 |
| 27 | 231 | 12996.00 |
| 28 | 216 | 12993.10 |
| 29 | 59 | 12952.35 |

| | STORE_NBR | TOT_SALES |
|----|-----------|-----------|
| 30 | 60 | 12927.70 |
| 31 | 250 | 12924.20 |
| 32 | 93 | 12813.80 |
| 33 | 3 | 12802.45 |
| 34 | 154 | 12699.25 |
| 35 | 7 | 12486.90 |
| 36 | 43 | 12384.55 |
| 37 | 213 | 12364.60 |
| 38 | 230 | 12287.00 |
| 39 | 36 | 12098.25 |
| 40 | 63 | 12085.00 |
| 41 | 156 | 12044.30 |
| 42 | 166 | 12029.85 |
| 43 | 179 | 11864.70 |
| 44 | 152 | 11855.80 |
| 45 | 79 | 11831.20 |
| 46 | 100 | 11830.70 |
| 47 | 128 | 11806.20 |
| 48 | 168 | 11797.60 |
| 49 | 175 | 11768.30 |
| 50 | 80 | 11756.90 |
| 51 | 133 | 11573.95 |
| 52 | 71 | 11550.80 |
| 53 | 259 | 11540.90 |
| 54 | 94 | 11474.65 |
| 55 | 112 | 11392.40 |
| 56 | 119 | 11338.50 |
| 57 | 178 | 11302.50 |
| 58 | 270 | 11293.95 |
| 59 | 69 | 11288.85 |

| STORE_NBR | TOT_SALES |
|-----------|-----------|
| 60 | 223 |
| 61 | 157 |
| 62 | 257 |
| 63 | 269 |
| 64 | 33 |
| 65 | 114 |
| 66 | 91 |
| 67 | 137 |
| 68 | 184 |
| 69 | 160 |
| 70 | 180 |
| 71 | 138 |
| 72 | 55 |
| 73 | 106 |
| 74 | 164 |
| 75 | 13 |
| 76 | 247 |
| 77 | 86 |
| 78 | 155 |
| 79 | 227 |
| 80 | 236 |
| 81 | 62 |
| 82 | 225 |
| 83 | 113 |
| 84 | 172 |
| 85 | 57 |
| 86 | 232 |
| 87 | 105 |
| 88 | 102 |
| 89 | 97 |

| | STORE_NBR | TOT_SALES |
|-----|-----------|-----------|
| 90 | 229 | 10417.90 |
| 91 | 196 | 10408.20 |
| 92 | 191 | 10404.70 |
| 93 | 109 | 10399.10 |
| 94 | 30 | 10359.50 |
| 95 | 116 | 10354.50 |
| 96 | 104 | 10344.40 |
| 97 | 221 | 10330.70 |
| 98 | 67 | 10326.00 |
| 99 | 207 | 10317.20 |
| 100 | 23 | 10305.80 |
| 101 | 10 | 10302.30 |
| 102 | 48 | 10213.40 |
| 103 | 39 | 10182.05 |
| 104 | 122 | 10101.60 |
| 105 | 101 | 10077.70 |
| 106 | 32 | 9983.50 |
| 107 | 183 | 9977.60 |
| 108 | 118 | 9938.55 |
| 109 | 83 | 9924.90 |
| 110 | 190 | 9884.20 |
| 111 | 147 | 9869.90 |
| 112 | 222 | 9788.60 |
| 113 | 162 | 9786.10 |
| 114 | 219 | 9771.60 |
| 115 | 129 | 9771.10 |
| 116 | 271 | 9721.80 |
| 117 | 110 | 9688.70 |
| 118 | 209 | 9666.70 |
| 119 | 45 | 9508.50 |

| STORE_NBR | TOT_SALES |
|-----------|-------------|
| 120 | 5 9500.80 |
| 121 | 15 9486.05 |
| 122 | 70 9483.00 |
| 123 | 241 9399.60 |
| 124 | 78 9381.25 |
| 125 | 28 9374.35 |
| 126 | 107 9296.70 |
| 127 | 208 9294.75 |
| 128 | 262 9261.10 |
| 129 | 24 9230.00 |
| 130 | 19 8972.90 |
| 131 | 212 8873.90 |
| 132 | 148 8819.00 |
| 133 | 144 8707.75 |
| 134 | 56 8393.05 |
| 135 | 202 5991.05 |
| 136 | 246 5954.20 |
| 137 | 242 5816.80 |
| 138 | 235 5728.15 |
| 139 | 173 5702.75 |
| 140 | 200 5544.60 |
| 141 | 260 5432.25 |
| 142 | 84 5396.30 |
| 143 | 142 5388.00 |
| 144 | 54 5370.85 |
| 145 | 25 5297.65 |
| 146 | 134 5246.35 |
| 147 | 21 5154.90 |
| 148 | 66 5142.00 |
| 149 | 27 5120.70 |

| | STORE_NBR | TOT_SALES |
|-----|-----------|-----------|
| 150 | 12 | 5009.90 |
| 151 | 150 | 4970.00 |
| 152 | 253 | 4918.70 |
| 153 | 64 | 4894.60 |
| 154 | 149 | 4865.00 |
| 155 | 126 | 4791.70 |
| 156 | 37 | 4728.45 |
| 157 | 171 | 4725.90 |
| 158 | 234 | 4712.30 |
| 159 | 29 | 4697.50 |
| 160 | 243 | 4677.60 |
| 161 | 251 | 4666.45 |
| 162 | 240 | 4655.50 |
| 163 | 272 | 4653.95 |
| 164 | 17 | 4636.60 |
| 165 | 174 | 4577.80 |
| 166 | 143 | 4554.70 |
| 167 | 47 | 4507.95 |
| 168 | 8 | 4499.20 |
| 169 | 182 | 4489.90 |
| 170 | 96 | 4386.70 |
| 171 | 215 | 4354.50 |
| 172 | 245 | 4305.60 |
| 173 | 34 | 4266.20 |
| 174 | 170 | 4248.30 |
| 175 | 228 | 4236.30 |
| 176 | 248 | 4186.10 |
| 177 | 82 | 4103.50 |
| 178 | 239 | 4086.85 |
| 179 | 9 | 4012.40 |

| STORE_NBR | TOT_SALES |
|-----------|-----------|
| 180 | 3991.60 |
| 181 | 3990.35 |
| 182 | 3927.50 |
| 183 | 3904.20 |
| 184 | 3882.70 |
| 185 | 3769.60 |
| 186 | 3733.10 |
| 187 | 3689.40 |
| 188 | 3534.70 |
| 189 | 3529.00 |
| 190 | 3491.40 |
| 191 | 3466.90 |
| 192 | 3388.20 |
| 193 | 3269.70 |
| 194 | 3163.10 |
| 195 | 3150.40 |
| 196 | 3086.00 |
| 197 | 3040.00 |
| 198 | 3025.40 |
| 199 | 3023.45 |
| 200 | 3009.80 |
| 201 | 3008.20 |
| 202 | 2966.80 |
| 203 | 2909.70 |
| 204 | 2868.60 |
| 205 | 2835.30 |
| 206 | 2826.90 |
| 207 | 2752.90 |
| 208 | 2720.40 |
| 209 | 2715.05 |

| STORE_NBR | TOT_SALES |
|-----------|-------------|
| 210 | 6 2684.90 |
| 211 | 163 2635.70 |
| 212 | 195 2608.25 |
| 213 | 268 2601.05 |
| 214 | 41 2570.20 |
| 215 | 264 2519.80 |
| 216 | 20 2519.25 |
| 217 | 131 2470.40 |
| 218 | 265 2433.40 |
| 219 | 111 2407.70 |
| 220 | 1 2393.60 |
| 221 | 89 2390.10 |
| 222 | 167 2297.60 |
| 223 | 74 2244.90 |
| 224 | 124 2118.40 |
| 225 | 136 2022.30 |
| 226 | 2 2005.80 |
| 227 | 51 2005.50 |
| 228 | 186 1906.70 |
| 229 | 151 1843.20 |
| 230 | 120 1810.15 |
| 231 | 266 1742.70 |
| 232 | 16 1632.10 |
| 233 | 254 1621.30 |
| 234 | 35 1608.90 |
| 235 | 98 1589.20 |
| 236 | 189 1386.90 |
| 237 | 61 562.90 |
| 238 | 132 495.40 |
| 239 | 52 465.30 |

| | STORE_NBR | TOT_SALES |
|-----|-----------|-----------|
| 240 | 204 | 456.00 |
| 241 | 218 | 419.00 |
| 242 | 14 | 410.70 |
| 243 | 224 | 392.90 |
| 244 | 135 | 380.10 |
| 245 | 127 | 376.90 |
| 246 | 192 | 361.30 |
| 247 | 158 | 350.90 |
| 248 | 139 | 340.30 |
| 249 | 159 | 338.90 |
| 250 | 44 | 336.20 |
| 251 | 258 | 335.90 |
| 252 | 161 | 335.20 |
| 253 | 244 | 331.75 |
| 254 | 263 | 300.50 |
| 255 | 267 | 275.40 |
| 256 | 146 | 275.10 |
| 257 | 42 | 257.80 |
| 258 | 198 | 252.70 |
| 259 | 140 | 244.90 |
| 260 | 99 | 221.90 |
| 261 | 177 | 211.20 |
| 262 | 117 | 161.80 |
| 263 | 31 | 14.80 |
| 264 | 85 | 13.90 |
| 265 | 193 | 13.10 |
| 266 | 92 | 9.20 |
| 267 | 206 | 7.60 |
| 268 | 252 | 7.40 |
| 269 | 11 | 6.70 |

| STORE_NBR | TOT_SALES |
|-----------|-----------|
| 270 | 76 |
| 271 | 211 |
| 272 | 0 |

In []: