In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

In [2]:
```
df1 = pd.read_csv(r"C:\Users\Admin\Desktop\Python\Data\Quantium_Project\QVI_data_Pr
df1.head()
```

Out[2]:

	LYLTY_CARD_NBR	DATE	STORE_NBR	TXN_ID	PROD_NBR	PROD_NAME	PROD_QTY	T
0	1000	17-10-2018	1	1	5	Natural Chip Compny SeaSalt175g	2	
1	1002	16-09-2018	1	2	58	Red Rock Deli Chikn&Garlic Aioli 150g	1	
2	1003	07-03-2019	1	3	52	Grain Waves Sour Cream&Chives 210G	1	
3	1003	08-03-2019	1	4	106	Natural ChipCo Hony Soy Chckn175g	1	
4	1004	02-11-2018	1	5	96	WW Original Stacked Chips 160g	1	

In [3]:
```
df1.shape
```

Out[3]:
```
(264834, 12)
```

Checking null values

In [4]:
```
df1.isnull().sum()
```

Out[4]:
```
LYLTY_CARD_NBR      0
DATE                0
STORE_NBR           0
TXN_ID              0
PROD_NBR            0
PROD_NAME           0
PROD_QTY            0
TOT_SALES           0
PACK_SIZE           0
BRAND               0
LIFESTAGE           0
PREMIUM_CUSTOMER    0
dtype: int64
```

Checking data types

In [5]:
```
df1.dtypes
```

Out[5]:
```
LYLTY_CARD_NBR        int64
DATE                 object
STORE_NBR             int64
TXN_ID                int64
PROD_NBR              int64
PROD_NAME            object
PROD_QTY              int64
TOT_SALES           float64
PACK_SIZE             int64
BRAND                object
LIFESTAGE            object
PREMIUM_CUSTOMER     object
dtype: object
```

Date is in object type so, it will to be converted to be into datetime formate.

In [6]:
```
df1['DATE'] = pd.to_datetime(df1['DATE'])
df1.info()
```

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

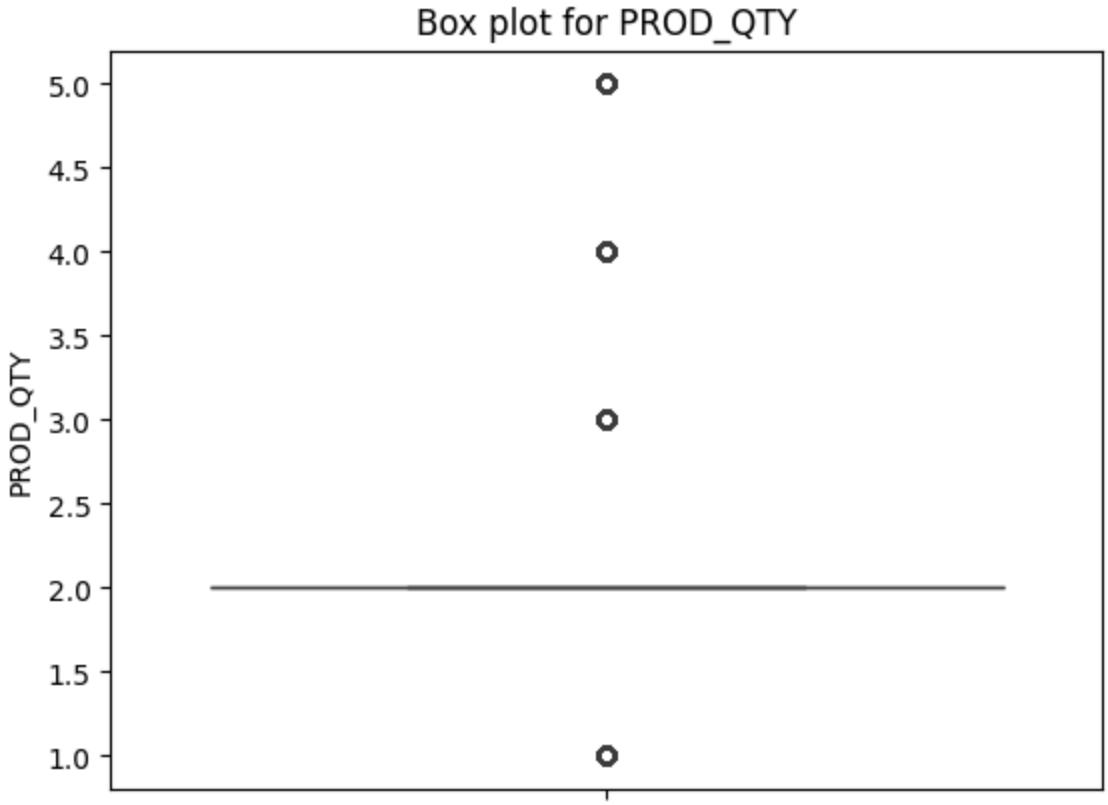
In [7]:
```
df1.describe()
```

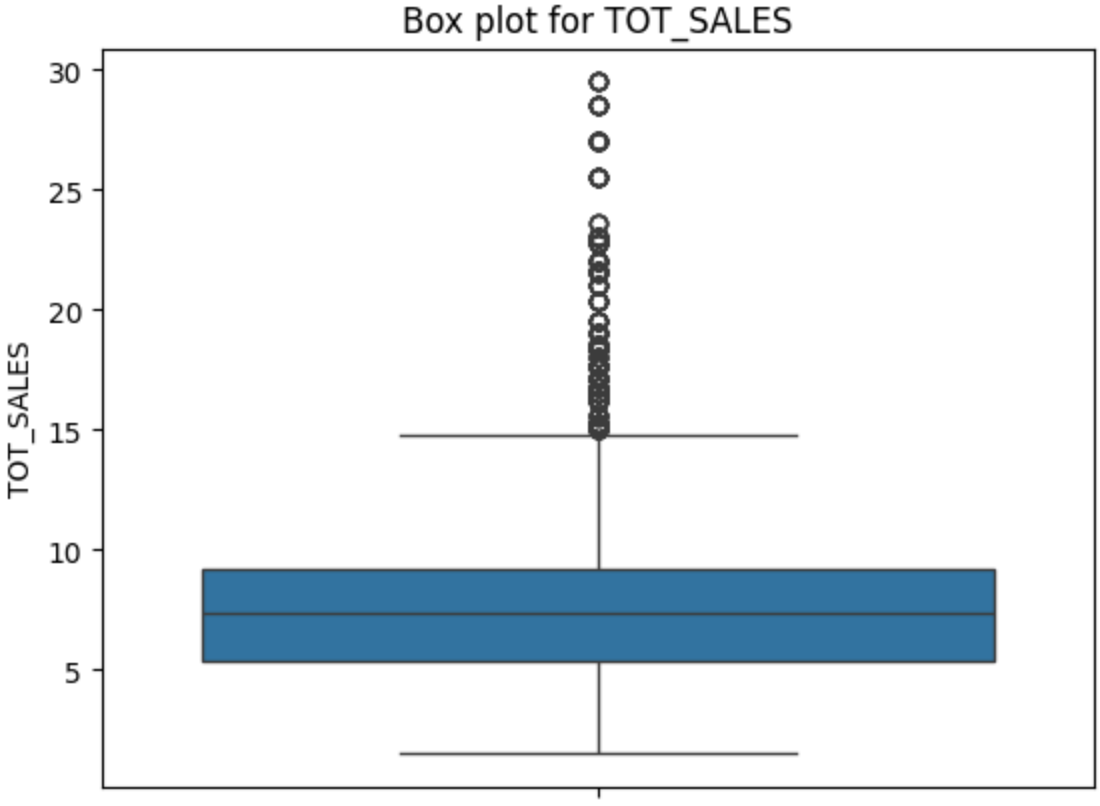
Out[7]:

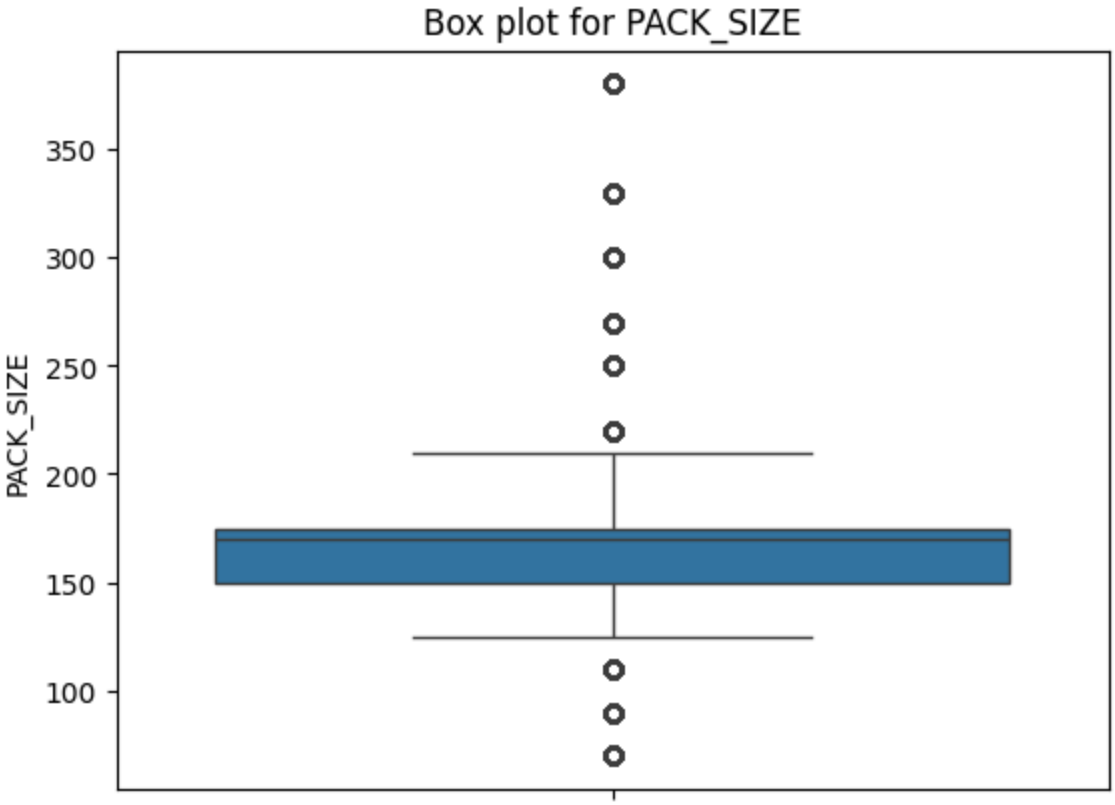
	LYLTY_CARD_NBR	DATE	STORE_NBR	TXN_ID	PROD_NBR
count	2.648340e+05	264834	264834.000000	2.648340e+05	264834.000000
mean	1.355488e+05	2018-12-30 00:52:10.292937984	135.079423	1.351576e+05	56.583554
min	1.000000e+03	2018-07-01 00:00:00	1.000000	1.000000e+00	1.000000
25%	7.002100e+04	2018-09-30 00:00:00	70.000000	6.760050e+04	28.000000
50%	1.303570e+05	2018-12-30 00:00:00	130.000000	1.351365e+05	56.000000
75%	2.030940e+05	2019-03-31 00:00:00	203.000000	2.026998e+05	85.000000
max	2.373711e+06	2019-06-30 00:00:00	272.000000	2.415841e+06	114.000000
std	8.057990e+04	NaN	76.784063	7.813292e+04	32.826444

Ploting box plot to check oulier

In [8]:
```
for i in ['PROD_QTY','TOT_SALES','PACK_SIZE']:
    plt.title(f'Box plot for {i}')
    sns.boxplot(df1[i])
    plt.show()
```


Box plot for PROD_QTY
PROD_QTY
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0


Box plot for TOT_SALES
TOT_SALES
30
25
20
15
10
5



In [9]:
```
df1.describe(include='object')
```

Out[9]:

	PROD_NAME	BRAND	LIFESTAGE	PREMIUM_CUSTOMER
count	264834	264834	264834	264834
unique	114	21	7	3
top	Kettle Mozzarella Basil & Pesto 175g	KETTLE	OLDER SINGLES/COUPLES	Mainstream
freq	3304	41288	54479	101988

In [10]:
```
for i in df1.describe(include='object').columns:
    print(i)
    print(df1[i].unique())
    print('---------------------------------------------')
```

```
PROD_NAME
['Natural Chip        Compny SeaSalt175g'
 'Red Rock Deli Chikn&Garlic Aioli 150g'
 'Grain Waves Sour    Cream&Chives 210G'
 'Natural ChipCo      Hony Soy Chckn175g' 'WW Original Stacked Chips 160g'
 'Cheetos Puffs 165g' 'Infuzions SourCream&Herbs Veg Strws 110g'
 'RRD SR Slow Rst     Pork Belly 150g' 'Doritos Cheese      Supreme 330g'
 'Doritos Mexicana    170g' 'Old El Paso Salsa   Dip Tomato Med 300g'
 'GrnWves Plus Btroot & Chilli Jam 180g'
 'Smiths Crinkle Cut  Chips Barbecue 170g'
 'Kettle Sensations   Camembert & Fig 150g'
 'Doritos Corn Chip Southern Chicken 150g' 'CCs Tasty Cheese    175g'
 'Tostitos Splash Of  Lime 175g' 'Kettle 135g Swt Pot Sea Salt'
 'RRD Salt & Vinegar  165g' 'Infuzions Mango     Chutny Papadums 70g'
 'Smiths Crinkle Cut  Snag&Sauce 150g' 'Smiths Crinkle      Original 330g'
 'RRD Sweet Chilli &  Sour Cream 165g'
 'Smiths Chip Thinly  S/Cream&Onion 175g'
 'Smiths Crinkle Chips Salt & Vinegar 330g'
 'Red Rock Deli SR    Salsa & Mzzrlla 150g'
 'Cobs Popd Sea Salt  Chips 110g' 'Natural ChipCo Sea  Salt & Vinegr 175g'
 'Natural Chip Co     Tmato Hrb&Spce 175g' 'Burger Rings 220g'
 'Woolworths Cheese   Rings 190g'
 'Smiths Thinly       Swt Chli&S/Cream175G'
 'Thins Chips Seasonedchicken 175g'
 'Smiths Thinly Cut   Roast Chicken 175g'
 'Tyrrells Crisps     Ched & Chives 165g'
 'Doritos Corn Chips  Cheese Supreme 170g'
 'Smiths Chip Thinly  Cut Original 175g'
 'Smiths Crinkle Cut  Chips Original 170g'
 'Thins Chips Light&  Tangy 175g' 'Doritos Corn Chips  Original 170g'
 'Kettle Sensations   Siracha Lime 150g'
 'Smiths Crinkle Cut  Salt & Vinegar 170g'
 'Smith Crinkle Cut   Bolognese 150g' 'Cheezels Cheese 330g'
 'Kettle Chilli 175g' 'Tyrrells Crisps     Lightly Salted 165g'
 'Twisties Cheese     270g' 'WW Crinkle Cut      Chicken 175g'
 'RRD Chilli&         Coconut 150g'
 'Infuzions BBQ Rib   Prawn Crackers 110g'
 'Sunbites Whlegrn    Crisps Frch/Onin 90g'
 'Doritos Salsa       Medium 300g' 'Kettle Tortilla ChpsFeta&Garlic 150g'
 'Smiths Crinkle Cut  French OnionDip 150g'
 'WW D/Style Chip     Sea Salt 200g'
 'Smiths Chip Thinly  CutSalt/Vinegr175g'
 'Kettle Sensations   BBQ&Maple 150g'
 'Old El Paso Salsa   Dip Tomato Mild 300g'
 'Tostitos Smoked     Chipotle 175g' 'RRD Lime & Pepper   165g'
 'CCs Nacho Cheese    175g' 'Snbts Whlgrn Crisps Cheddr&Mstrd 90g'
 'Kettle Tortilla ChpsBtroot&Ricotta 150g'
 'Pringles Sthrn FriedChicken 134g' 'Pringles Chicken    Salt Crips 134g'
 'French Fries Potato Chips 175g' 'Kettle Mozzarella   Basil & Pesto 175g'
 'CCs Original 175g' 'Tostitos Lightly    Salted 175g'
 'Smiths Crnkle Chip  Orgnl Big Bag 380g'
 'Smiths Crinkle Cut  Chips Chicken 170g'
 'Smiths Crinkle Cut  Chips Chs&Onion170g' 'Twisties Chicken270g'
 'Woolworths Medium   Salsa 300g'
 'Red Rock Deli Sp    Salt & Truffle 150G' 'RRD Pc Sea Salt     165g'
 'WW Supreme Cheese   Corn Chips 200g' 'WW Original Corn    Chips 200g'
```

```
 'Woolworths Mild     Salsa 300g' 'Cheezels Cheese Box 125g'
 'Doritos Salsa Mild  300g' 'Cobs Popd Swt/Chlli &Sr/Cream Chips 110g'
 'Infzns Crn Crnchers Tangy Gcamole 110g'
 'WW Sour Cream &OnionStacked Chips 160g'
 'Pringles Mystery    Flavour 134g' 'Pringles Barbeque   134g'
 'Grain Waves         Sweet Chilli 210g' 'Pringles Sweet&Spcy BBQ 134g'
 'Kettle Original 175g' 'Infuzions Thai SweetChili PotatoMix 110g'
 'Old El Paso Salsa   Dip Chnky Tom Ht300g'
 'Smiths Crinkle Cut  Tomato Salsa 150g' 'Cheetos Chs & Bacon Balls 190g'
 'Kettle Sweet Chilli And Sour Cream 175g'
 'Doritos Corn Chips  Nacho Cheese 170g'
 'Cobs Popd Sour Crm  &Chives Chips 110g'
 'Red Rock Deli Thai  Chilli&Lime 150g' 'Twisties Cheese     Burger 250g'
 'Kettle Sea Salt     And Vinegar 175g'
 'WW Crinkle Cut      Original 175g' 'Dorito Corn Chp     Supreme 380g'
 'Doritos Corn Chip Mexican Jalapeno 150g'
 'Pringles SourCream  Onion 134g'
 'Kettle Tortilla ChpsHny&Jlpno Chili 150g'
 'RRD Steak &         Chimuchurri 150g' 'Thins Chips Salt &  Vinegar 175g'
 'Thins Chips         Originl saltd 175g'
 'RRD Honey Soy       Chicken 165g' 'Kettle Honey Soy    Chicken 175g'
 'NCC Sour Cream &    Garden Chives 175g'
 'Pringles Original   Crisps 134g' 'Smith Crinkle Cut   Mac N Cheese 150g'
 'Thins Potato Chips  Hot & Spicy 175g' 'Pringles Slt Vingar 134g']
-----------------------------------------
BRAND
['NATURAL' 'RRD' 'GRNWVES' 'WOOLWORTHS' 'CHEETOS' 'INFUZIONS' 'DORITOS'
 'OLD' 'SMITHS' 'KETTLE' 'CCS' 'TOSTITOS' 'COBS' 'BURGER' 'THINS'
 'TYRRELLS' 'CHEEZELS' 'TWISTIES' 'SUNBITES' 'PRINGLES' 'FRENCH']
-----------------------------------------
LIFESTAGE
['YOUNG SINGLES/COUPLES' 'YOUNG FAMILIES' 'OLDER SINGLES/COUPLES'
 'MIDAGE SINGLES/COUPLES' 'NEW FAMILIES' 'OLDER FAMILIES' 'RETIREES']
-----------------------------------------
PREMIUM_CUSTOMER
['Premium' 'Mainstream' 'Budget']
-----------------------------------------
```

In [11]:
```
df1['MonthName'] = df1['DATE'].dt.month_name()
df1['Year'] = df1['DATE'].dt.year
df1['Quarter'] = df1['DATE'].dt.quarter
df1.head()
```

Out[11]:

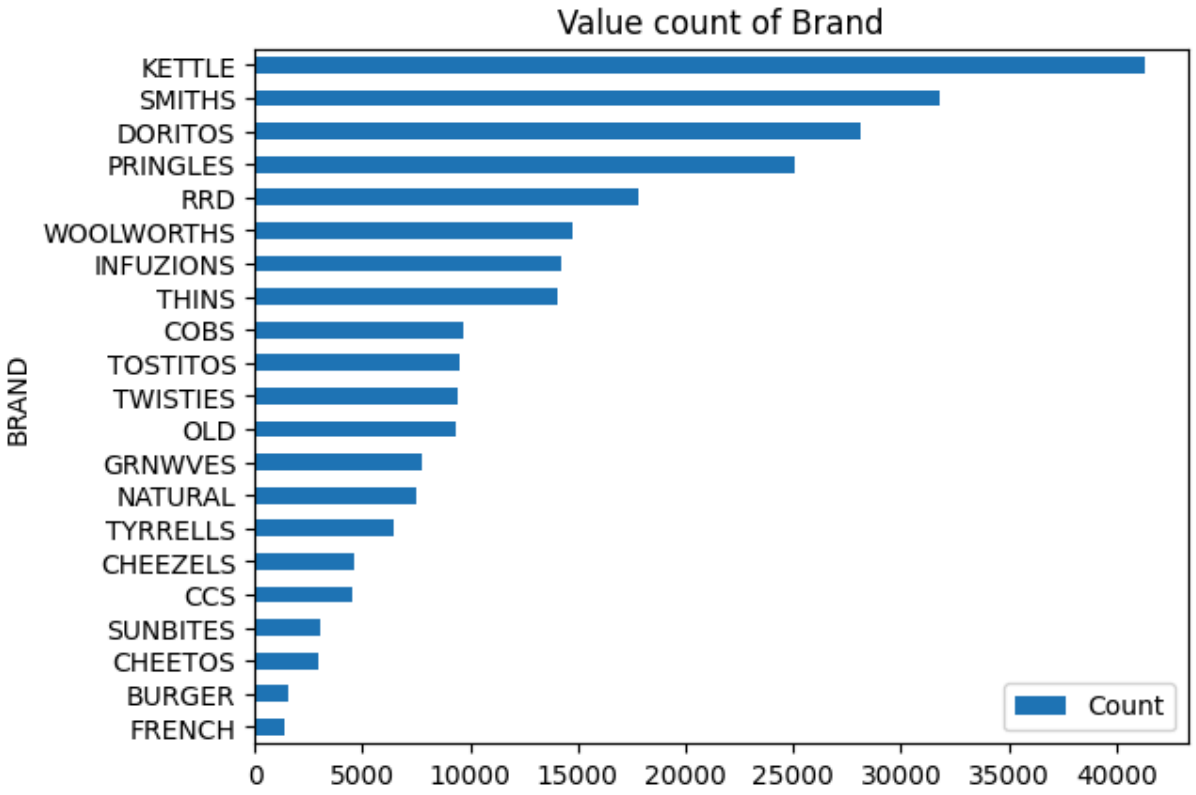
	LYLTY_CARD_NBR	DATE	STORE_NBR	TXN_ID	PROD_NBR	PROD_NAME	PROD_QTY
0	1000	2018-10-17	1	1	5	Natural Chip Compny SeaSalt175g	2
1	1002	2018-09-16	1	2	58	Red Rock Deli Chikn&Garlic Aioli 150g	1
2	1003	2019-03-07	1	3	52	Grain Waves Sour Cream&Chives 210G	1
3	1003	2019-03-08	1	4	106	Natural ChipCo Hony Soy Chckn175g	1
4	1004	2018-11-02	1	5	96	WW Original Stacked Chips 160g	1

Performing EDA

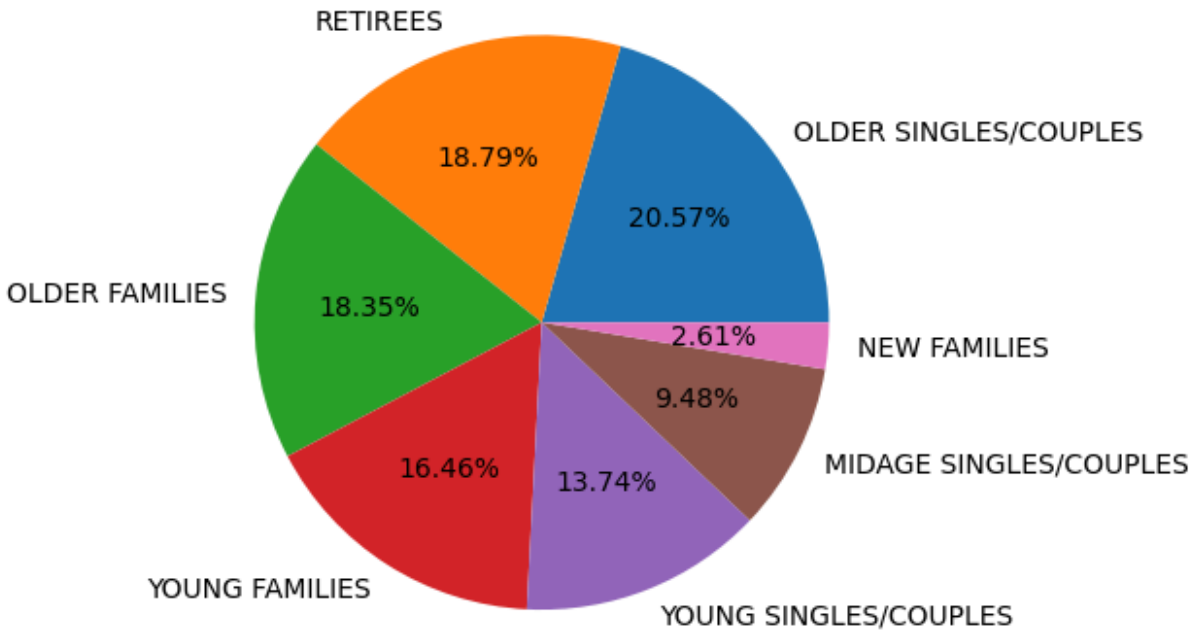
In [12]:
```
df1['PROD_NAME'].value_counts().nunique() #There are 101 total product.
```

Out[12]:
```
101
```

In [13]:
```
df1['BRAND'].value_counts().sort_values().plot(kind = 'barh', label = 'Count')#Most
plt.title('Value count of Brand')
plt.legend()
plt.show()
```



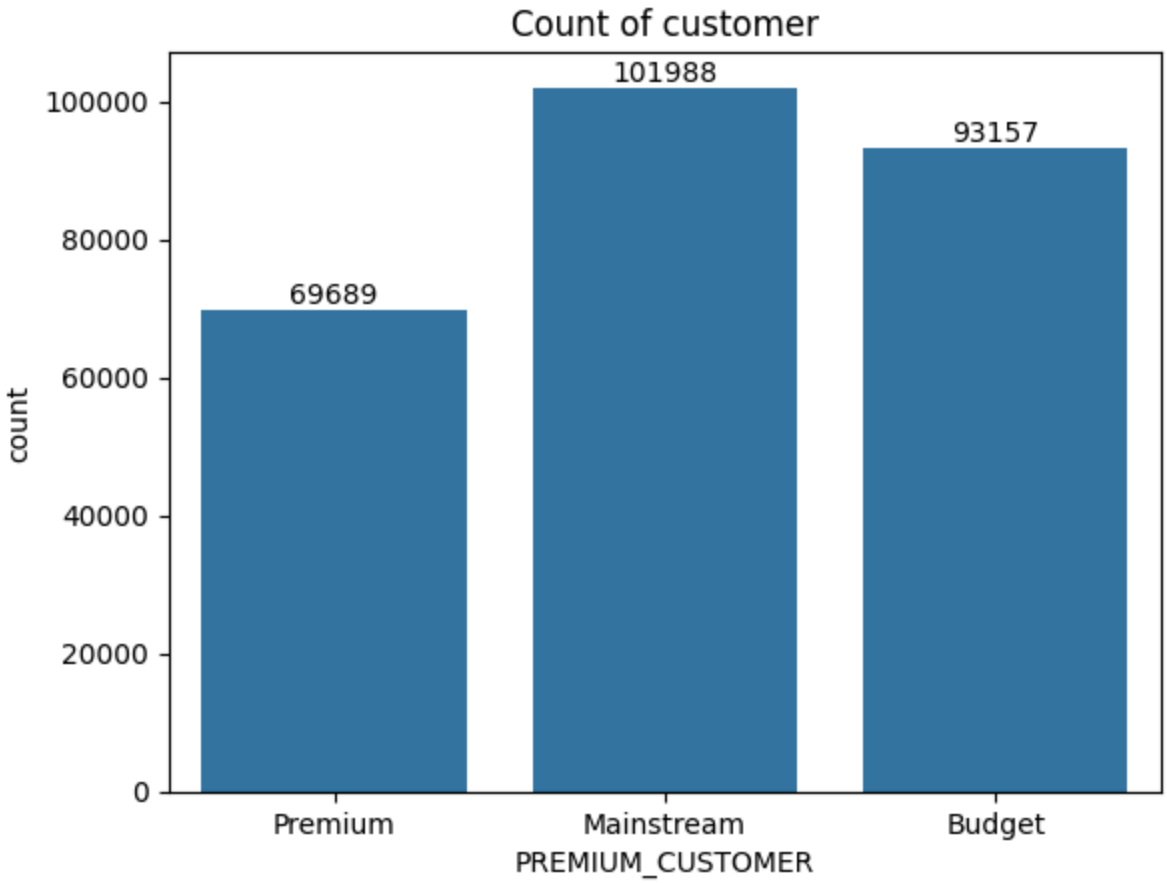
In [14]:
```
plt.pie(x = df1['LIFESTAGE'].value_counts().values,
        labels=df1['LIFESTAGE'].value_counts().index,
       autopct='%.2f%%')
plt.show()
```



After Kettle most opted brand is smiths.

In [15]:
```
plt.title('Count of customer')
ax = sns.countplot(x = df1['PREMIUM_CUSTOMER'])
ax.bar_label(ax.containers[0])
```

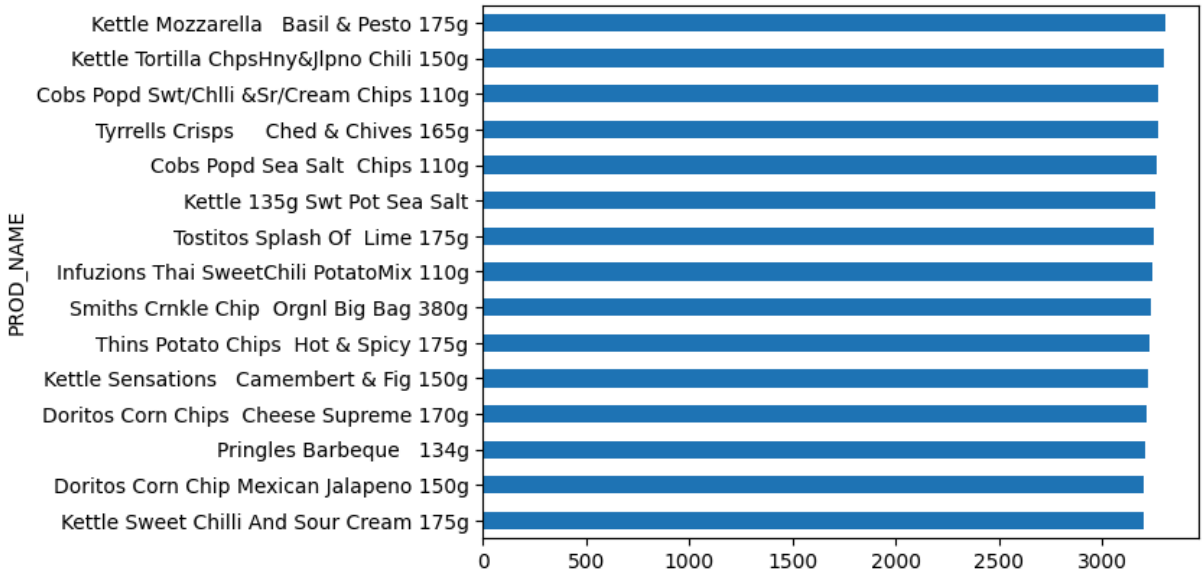
Out[15]:
```
[Text(0, 0, '69689'), Text(0, 0, '101988'), Text(0, 0, '93157')]
```



Most business given by mainstream and budget customers then premium customer giving business

In [16]:
```
r1 = df1['PROD_NAME'].value_counts().head(15).sort_values()
r1.plot(kind = 'barh')
```

Out[16]:
```
<Axes: ylabel='PROD_NAME'>
```



"Kettle Sweet Chilli And Sour Cream 175g" is most opted project by customer

In [17]:
```
df1.columns
```

Out[17]:
```
Index(['LYLTY_CARD_NBR', 'DATE', 'STORE_NBR', 'TXN_ID', 'PROD_NBR',
       'PROD_NAME', 'PROD_QTY', 'TOT_SALES', 'PACK_SIZE', 'BRAND', 'LIFESTAGE',
       'PREMIUM_CUSTOMER', 'MonthName', 'Year', 'Quarter'],
      dtype='object')
```

In [18]:
```
r2 = df1.groupby('PROD_NAME').agg(PROD_QTY_SUM=('PROD_QTY', 'sum'),sales_sum = ('TO
r2.head(15)
```

Out[18]:

	PROD_NAME	PROD_QTY_SUM	sales_sum
11	Dorito Corn Chp Supreme 380g	6109	39052.0
86	Smiths Crnkle Chip Orgnl Big Bag 380g	6164	36367.6
77	Smiths Crinkle Chips Salt & Vinegar 330g	6106	34804.2
33	Kettle Mozzarella Basil & Pesto 175g	6381	34457.4
76	Smiths Crinkle Original 330g	6018	34302.6
6	Cheezels Cheese 330g	6017	34296.9
12	Doritos Cheese Supreme 330g	5858	33390.6
39	Kettle Sweet Chilli And Sour Cream 175g	6120	33031.8
34	Kettle Original 175g	6064	32740.2
35	Kettle Sea Salt And Vinegar 175g	6035	32589.0
32	Kettle Honey Soy Chicken 175g	6033	32578.2
31	Kettle Chilli 175g	5792	31271.4
48	Old El Paso Salsa Dip Chnky Tom Ht300g	5986	30513.3
49	Old El Paso Salsa Dip Tomato Med 300g	5929	30237.9
50	Old El Paso Salsa Dip Tomato Mild 300g	5890	30033.9

"Dorito Corn Chp Supreme 380g" gives most sales

In [19]:
```
r2 = df1.groupby('PROD_NAME').agg(PROD_QTY_SUM=('PROD_QTY', 'sum'),
                                  sales_sum = ('TOT_SALES','sum')).reset_index().so
r2.head(15)
```

Out[19]:

	PROD_NAME	PROD_QTY_SUM	sales_sum
33	Kettle Mozzarella Basil & Pesto 175g	6381	34457.4
42	Kettle Tortilla ChpsHny&Jlpno Chili 150g	6309	29021.4
8	Cobs Popd Sea Salt Chips 110g	6277	23852.6
10	Cobs Popd Swt/Chlli &Sr/Cream Chips 110g	6256	23772.8
98	Tostitos Splash Of Lime 175g	6234	27429.6
102	Tyrrells Crisps Ched & Chives 165g	6227	26149.2
30	Kettle 135g Swt Pot Sea Salt	6212	26090.4
28	Infuzions Thai SweetChili PotatoMix 110g	6206	23582.8
95	Thins Potato Chips Hot & Spicy 175g	6185	20410.5
15	Doritos Corn Chips Cheese Supreme 170g	6180	27183.2
86	Smiths Crnkle Chip Orgnl Big Bag 380g	6164	36367.6
37	Kettle Sensations Camembert & Fig 150g	6157	28308.4
13	Doritos Corn Chip Mexican Jalapeno 150g	6125	23887.5
39	Kettle Sweet Chilli And Sour Cream 175g	6120	33031.8
51	Pringles Barbeque 134g	6112	22614.4

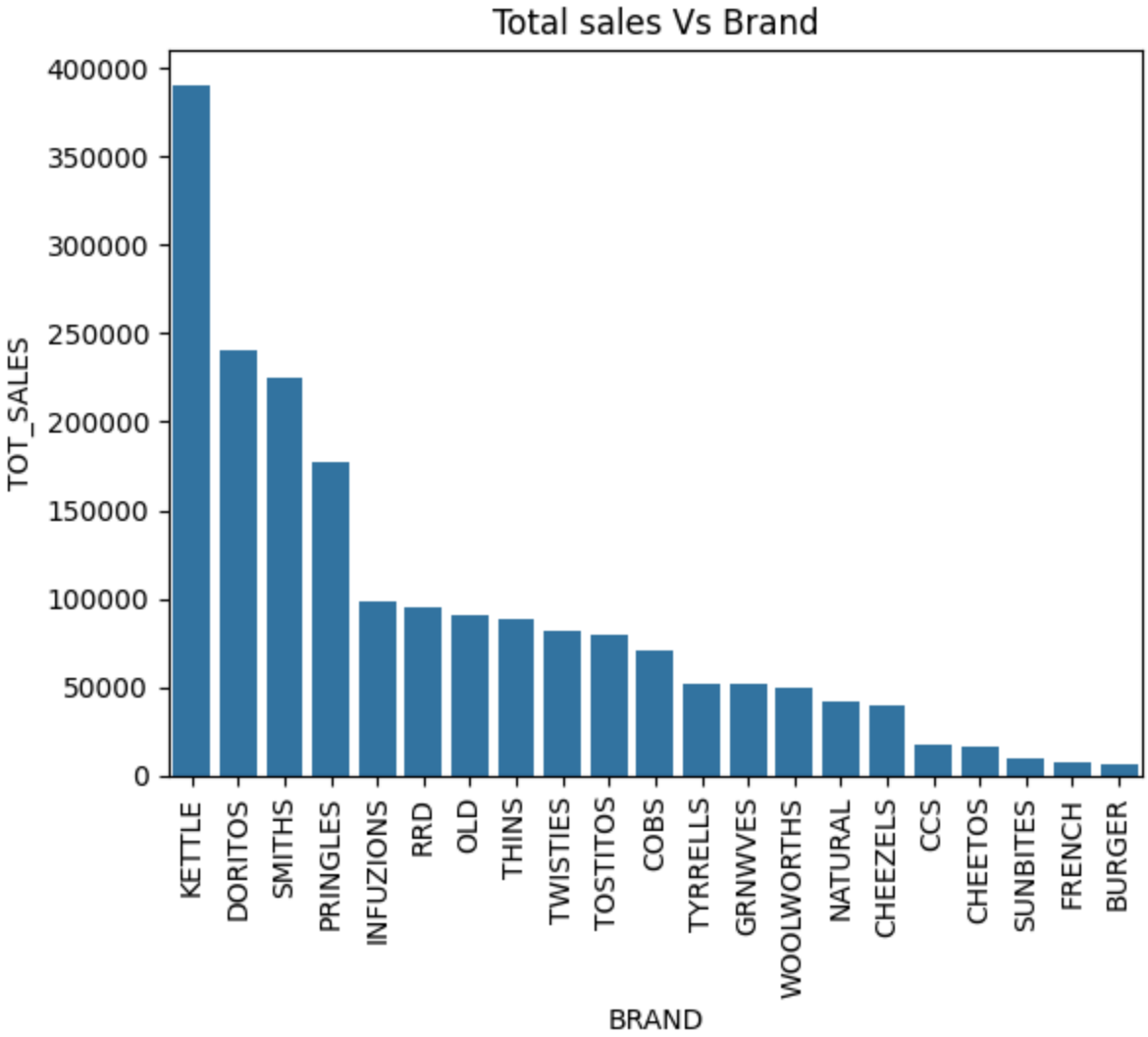
"Kettle Mozzarella Basil & Pesto 175g" is most qty ordered product.

In [20]:
```
r3 = df1.groupby('BRAND')[['TOT_SALES']].sum().reset_index().sort_values(by = 'TOT_
r3
```

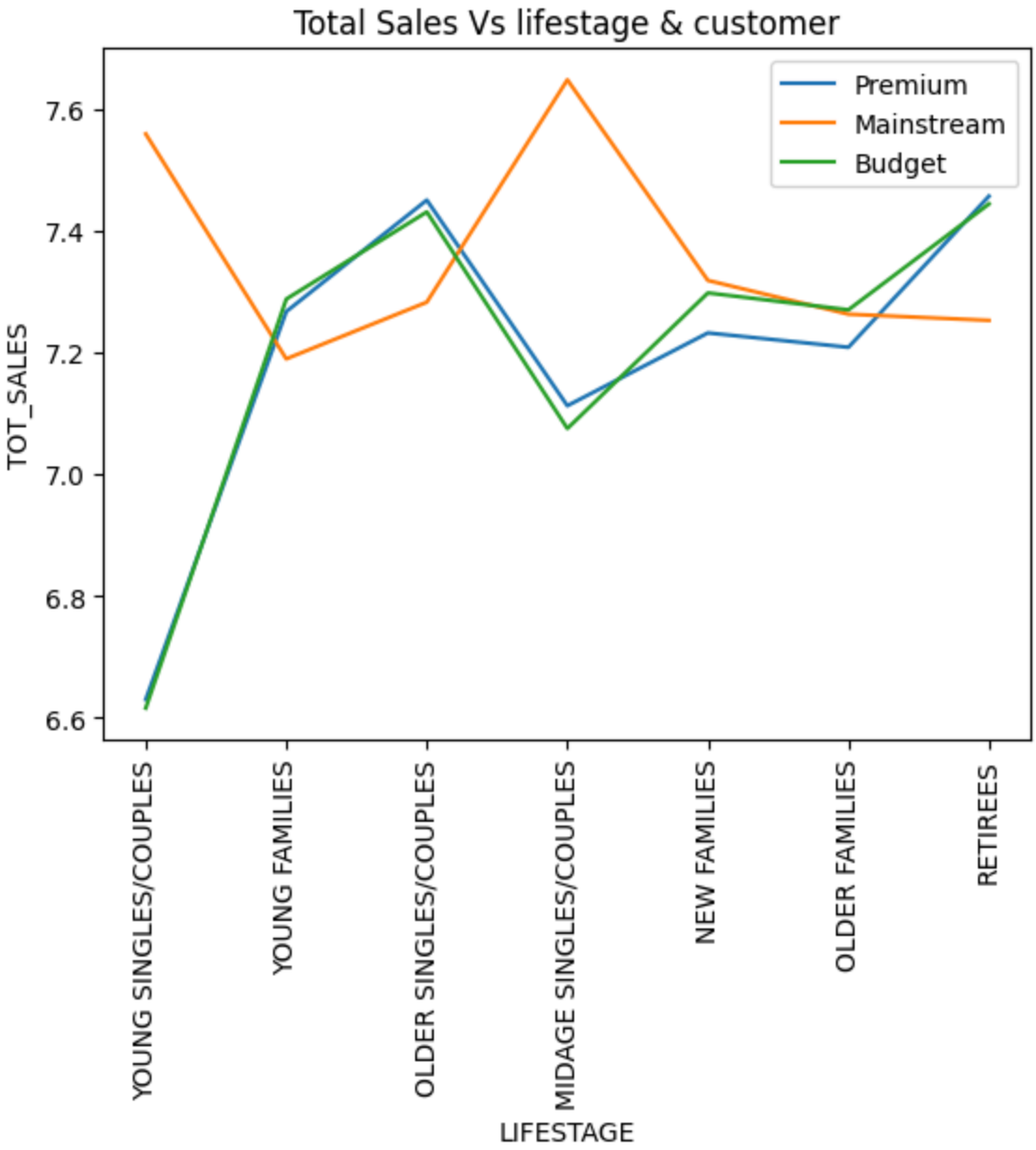
Out[20]:

	BRAND	TOT_SALES
9	KETTLE	390239.8
5	DORITOS	240590.9
14	SMITHS	224660.2
12	PRINGLES	177655.5
8	INFUZIONS	99047.6
13	RRD	95046.0
11	OLD	90785.1
16	THINS	88852.5
18	TWISTIES	81522.1
17	TOSTITOS	79789.6
4	COBS	70569.8
19	TYRRELLS	51647.4
7	GRNWVES	51617.2
20	WOOLWORTHS	49343.6
10	NATURAL	42318.0
3	CHEEZELS	40029.9
1	CCS	18078.9
2	CHEETOS	16884.5
15	SUNBITES	9676.4
6	FRENCH	7929.0
0	BURGER	6831.0

In [21]:
```
plt.title('Total sales Vs Brand')
sns.barplot(data = r3, x = 'BRAND', y = 'TOT_SALES')
plt.xticks(rotation = 90)
plt.show()
```



In [22]:
```
plt.title('Total Sales Vs lifestage & customer')
sns.lineplot(data = df1, y = 'TOT_SALES', x = 'LIFESTAGE', hue = 'PREMIUM_CUSTOMER'
plt.xticks(rotation = 90)
plt.legend()
plt.show()
```



"Customer mainstream" having life stage midage singles/couples giveing most sales

"Customer mainstream" giving most sales as compare to budget and premium customer

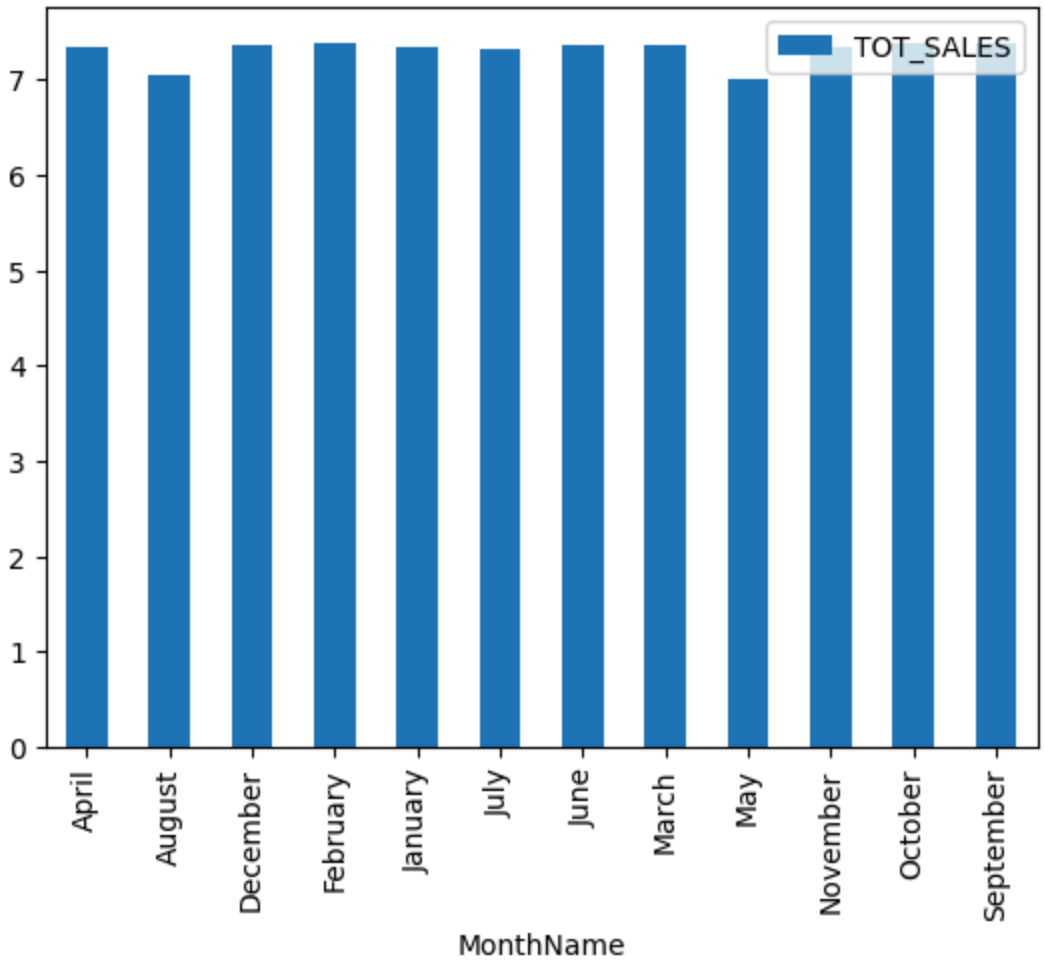
In [23]:
```
df1.groupby('MonthName')[['TOT_SALES']].mean().reset_index().sort_values(by = 'TOT_
```

Out[23]:

	MonthName	TOT_SALES
3	February	7.383729
11	September	7.382698
10	October	7.376871
7	March	7.359472
6	June	7.354373
2	December	7.353335
0	April	7.343798
4	January	7.339123
9	November	7.332679
5	July	7.325383
1	August	7.054041
8	May	6.999136

In [24]:
```
df1.groupby('MonthName')[['TOT_SALES']].mean().plot(kind = 'bar')
```

Out[24]:
```
<Axes: xlabel='MonthName'>
```



Fab month has most sales

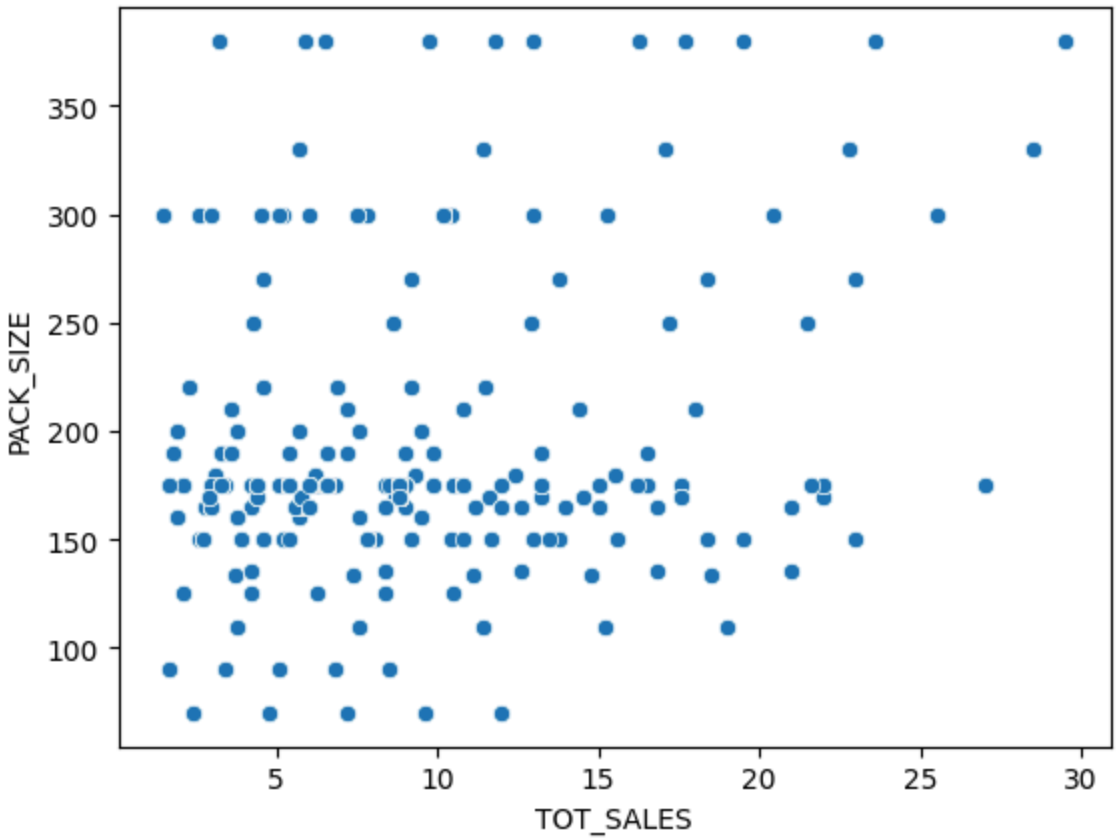
In [25]:
```
df1.groupby(['Year'])[['TOT_SALES']].mean()
```

Out[25]:

Year	TOT_SALES
2018	7.303771
2019	7.294835

2018 has most sales as campare to 2019

In [26]:
```
sns.scatterplot(data = df1, x = 'TOT_SALES', y = 'PACK_SIZE')
plt.show()
```


PACK_SIZE
350
300
250
200
150
100
5
10
15
20
25
30
TOT_SALES