DATA ANALYSIS PYTHON PROJECT -BLINKIT ANALYSIS

import libraries

```
In [29]:  import pandas as pd
          import numpy as np
          import matplotlib.pyplot as plt
          import seaborn as sns
```

```
In [30]:  df =pd.read_csv(r"C:\Users\HP\Desktop\Portfolio Data Analyst\Blinkit Portfolio P
```

```
In [31]:  df
```

Out[31]:

	Item Fat Content	Item Identifier	Item Type	Outlet Establishment Year	Outlet Identifier	Outlet Location Type	Outlet Size	Outlet Type
0	Regular	FDX32	Fruits and Vegetables	2012	OUT049	Tier 1	Medium	Supermarket Type1
1	Low Fat	NCB42	Health and Hygiene	2022	OUT018	Tier 3	Medium	Supermarket Type2
2	Regular	FDR28	Frozen Foods	2016	OUT046	Tier 1	Small	Supermarket Type1
3	Regular	FDL50	Canned	2014	OUT013	Tier 3	High	Supermarket Type1
4	Low Fat	DRI25	Soft Drinks	2015	OUT045	Tier 2	Small	Supermarket Type1
...	...	...	...	...	...	...	...	...
8518	low fat	NCT53	Health and Hygiene	2018	OUT027	Tier 3	Medium	Supermarket Type3
8519	low fat	FDN09	Snack Foods	2018	OUT027	Tier 3	Medium	Supermarket Type3
8520	low fat	DRE13	Soft Drinks	2018	OUT027	Tier 3	Medium	Supermarket Type3
8521	reg	FDT50	Dairy	2018	OUT027	Tier 3	Medium	Supermarket Type3
8522	reg	FDM58	Snack Foods	2018	OUT027	Tier 3	Medium	Supermarket Type3

8523 rows × 12 columns

```
In [32]:  df.head(10)
```

Out[32]:

	Item Fat Content	Item Identifier	Item Type	Outlet Establishment Year	Outlet Identifier	Outlet Location Type	Outlet Size	Outlet Type	
0	Regular	FDX32	Fruits and Vegetables	2012	OUT049	Tier 1	Medium	Supermarket Type1	0
1	Low Fat	NCB42	Health and Hygiene	2022	OUT018	Tier 3	Medium	Supermarket Type2	0
2	Regular	FDR28	Frozen Foods	2016	OUT046	Tier 1	Small	Supermarket Type1	0
3	Regular	FDL50	Canned	2014	OUT013	Tier 3	High	Supermarket Type1	0
4	Low Fat	DRI25	Soft Drinks	2015	OUT045	Tier 2	Small	Supermarket Type1	0
5	low fat	FDS52	Frozen Foods	2020	OUT017	Tier 2	Small	Supermarket Type1	0
6	Low Fat	NCU05	Health and Hygiene	2011	OUT010	Tier 3	Small	Grocery Store	0
7	Low Fat	NCD30	Household	2015	OUT045	Tier 2	Small	Supermarket Type1	0
8	Low Fat	FDW20	Fruits and Vegetables	2014	OUT013	Tier 3	High	Supermarket Type1	0
9	Low Fat	FDX25	Canned	2018	OUT027	Tier 3	Medium	Supermarket Type3	0

In [33]: `df.tail(25)`

	Item Fat Content	Item Identifier	Item Type	Outlet Establishment Year	Outlet Identifier	Outlet Location Type	Outlet Size	Outlet Type
8498	Regular	FDG46	Snack Foods	2018	OUT027	Tier 3	Medium	Supermarket Type3
8499	Regular	FDJ21	Snack Foods	2018	OUT027	Tier 3	Medium	Supermarket Type3
8500	Regular	FDK58	Snack Foods	2018	OUT027	Tier 3	Medium	Supermarket Type3
8501	Regular	FDN34	Snack Foods	2018	OUT027	Tier 3	Medium	Supermarket Type3
8502	Regular	FDP21	Snack Foods	2018	OUT027	Tier 3	Medium	Supermarket Type3
8503	Regular	FDR22	Snack Foods	2018	OUT027	Tier 3	Medium	Supermarket Type3
8504	Regular	FDS09	Snack Foods	2018	OUT027	Tier 3	Medium	Supermarket Type3
8505	Regular	FDS34	Snack Foods	2018	OUT027	Tier 3	Medium	Supermarket Type3
8506	Regular	FDU09	Snack Foods	2018	OUT027	Tier 3	Medium	Supermarket Type3
8507	Regular	FDU33	Snack Foods	2018	OUT027	Tier 3	Medium	Supermarket Type3
8508	Regular	FDU57	Snack Foods	2018	OUT027	Tier 3	Medium	Supermarket Type3
8509	Regular	FDU58	Snack Foods	2018	OUT027	Tier 3	Medium	Supermarket Type3
8510	Regular	FDX46	Snack Foods	2018	OUT027	Tier 3	Medium	Supermarket Type3
8511	Regular	FDX57	Snack Foods	2018	OUT027	Tier 3	Medium	Supermarket Type3
8512	Regular	FDY33	Snack Foods	2018	OUT027	Tier 3	Medium	Supermarket Type3
8513	Regular	DRY23	Soft Drinks	2018	OUT027	Tier 3	Medium	Supermarket Type3
8514	low fat	FDA11	Baking Goods	2018	OUT027	Tier 3	Medium	Supermarket Type3
8515	low fat	FDK38	Canned	2018	OUT027	Tier 3	Medium	Supermarket Type3
8516	low fat	FDO38	Canned	2018	OUT027	Tier 3	Medium	Supermarket Type3
8517	low fat	FDG32	Fruits and Vegetables	2018	OUT027	Tier 3	Medium	Supermarket Type3
8518	low fat	NCT53	Health and Hygiene	2018	OUT027	Tier 3	Medium	Supermarket Type3
8519	low fat	FDN09	Snack Foods	2018	OUT027	Tier 3	Medium	Supermarket Type3
8520	low fat	DRE13	Soft Drinks	2018	OUT027	Tier 3	Medium	Supermarket Type3
8521	reg	FDT50	Dairy	2018	OUT027	Tier 3	Medium	Supermarket Type3
8522	reg	FDM58	Snack Foods	2018	OUT027	Tier 3	Medium	Supermarket Type3

```
In [34]: print("Size of dataset :",df.shape)

         Size of dataset : (8523, 12)
```

```
In [35]: print(df.columns)

         Index(['Item Fat Content', 'Item Identifier', 'Item Type',
                'Outlet Establishment Year', 'Outlet Identifier',
                'Outlet Location Type', 'Outlet Size', 'Outlet Type', 'Item Visibility',
                'Item Weight', 'Sales', 'Rating'],
               dtype='object')
```

```
In [36]: df.dtypes
```

```
Out[36]: Item Fat Content             object
         Item Identifier              object
         Item Type                    object
         Outlet Establishment Year     int64
         Outlet Identifier            object
         Outlet Location Type         object
         Outlet Size                  object
         Outlet Type                  object
         Item Visibility             float64
         Item Weight                 float64
         Sales                       float64
         Rating                      float64
         dtype: object
```

Data Cleaning

```
In [38]: print(df['Item Fat Content'].unique())

         ['Regular' 'Low Fat' 'low fat' 'LF' 'reg']
```

```
In [39]: df['Item Fat Content'] =df['Item Fat Content'].replace({'LF':'Low Fat',
                                                                  'low fat': 'Low Fat',
                                                                  'reg':'Regular'})
```

```
In [40]: print(df['Item Fat Content'].unique())

         ['Regular' 'Low Fat']
```

Business Requirements

KPI's REUIREMENTS

```
In [43]: #total Sales

         total_sales =df['Sales'].sum()

         #Average sales
         avg_sales =df['Sales'].mean()

         #No of Items Sold
         No_of_items_Sold =df['Sales'].count()

         #Average Ratings
         avg_ratings =df['Rating'].mean()

         #Display
         print(f"Total Sales : ${total_sales:.1f}")
```

```
print(f"Average Sales : ${avg_sales:.1f}")

print(f"Number of Items Sold : ${No_of_items_Sold:,.0f}")

print(F"Average Ratings :{avg_ratings:,.1f}")
```

```
Total Sales : $1201681.5
Average Sales : $141.0
Number of Items Sold : $8,523
Average Ratings :4.0
```
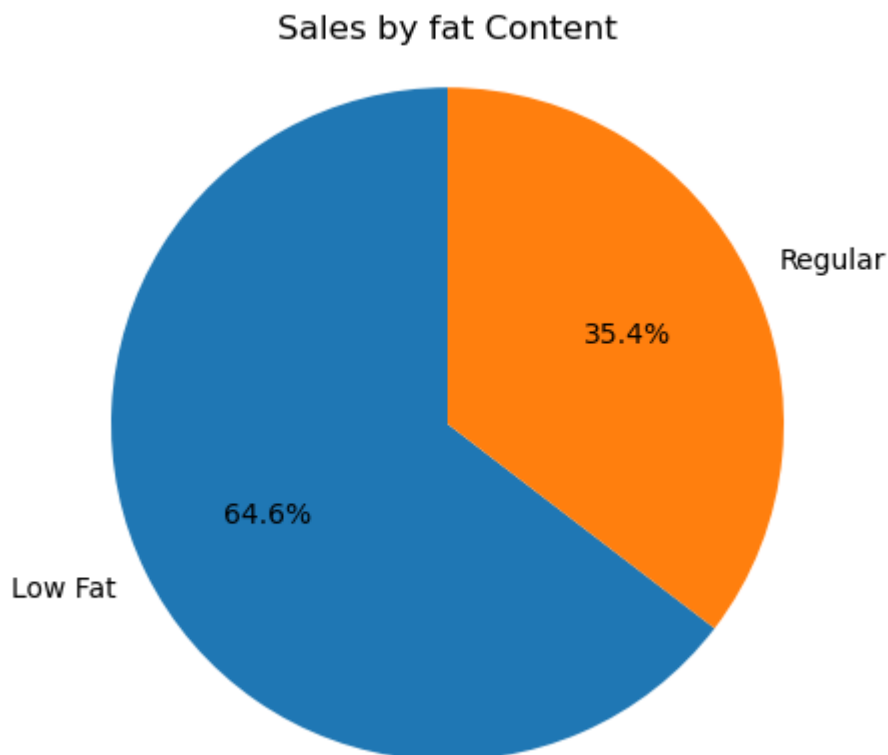
CHART REQUIREMENTS

Total Sales by Fat Content

In [46]:
```python
sales_by_Fat =df.groupby('Item Fat Content')['Sales'].sum()

plt.pie(sales_by_Fat,labels =sales_by_Fat.index,
        autopct ='%.1f%%',
        startangle =90)
plt.title('Sales by fat Content')
plt.axis('equal')
plt.show()
```


Total Sales by Item Type

In [74]:
```python
Sales_by_type =df.groupby('Item Type')['Sales'].sum().sort_values(ascending=Fals
plt.figure(figsize=(10,6))
bars =plt.bar(Sales_by_type.index,Sales_by_type.values)

plt.xticks(rotation=-90)
plt.xlabel('Item Type')
plt.ylabel('Total_Sales')
plt.title('Total Sales by Item type')
```

```
for bar in bars:
    plt.text(bar.get_x() +bar.get_width()/2,bar.get_height(),
             f'{bar.get_height():,.0f}',ha ='center',va='bottom',fontsize =8)
plt.tight_layout()
plt.show()
```


Total Sales by Item type
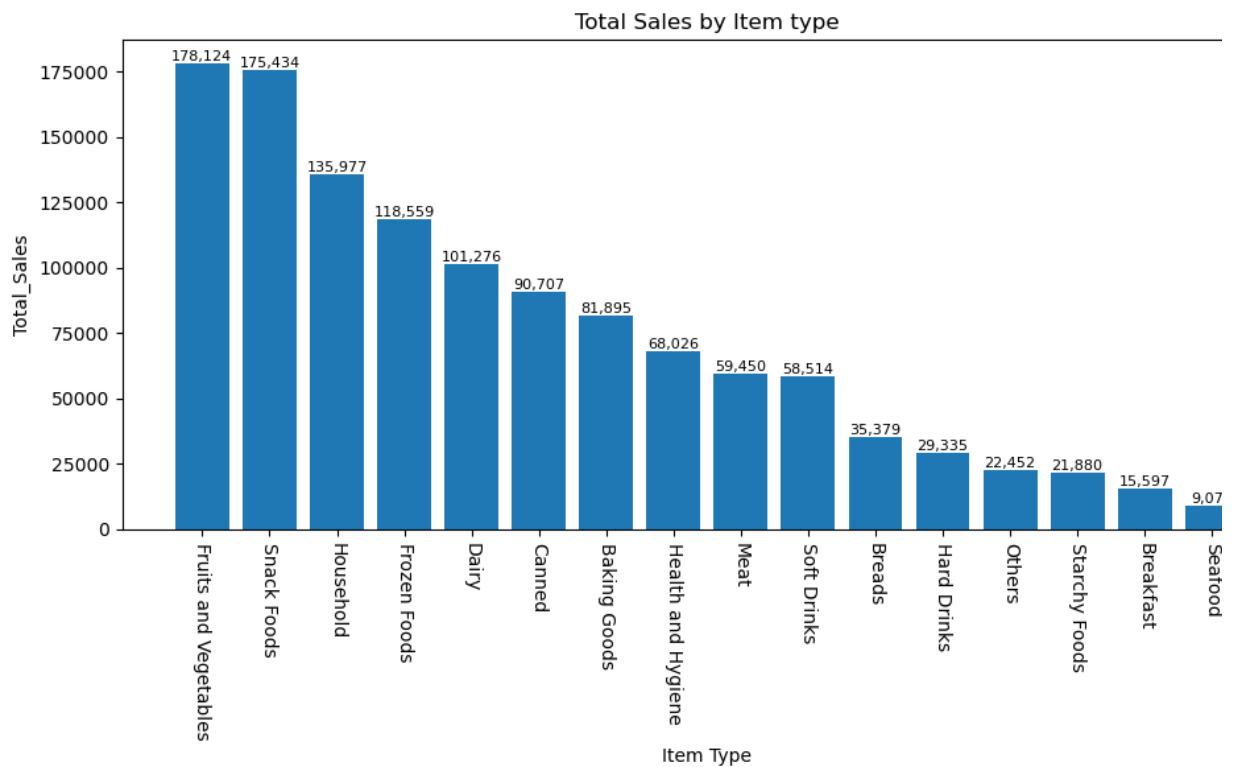
Fat Conent by Outlet For Total Sales

In [52]:
```
grouped = df.groupby(['Outlet Location Type','Item Fat Content'])['Sales'].sum()
grouped =grouped[['Regular', 'Low Fat']]
ax =grouped.plot(kind ='bar',figsize=(8,5), title='Outlet Tier by Item Fat Conte
plt.xlabel('Outlet Location Tier')
plt.ylabel('Total Sales')
plt.legend(title ='Item Fat Content')
plt.tight_layout()
plt.show()
```

## Outlet Tier by Item Fat Content
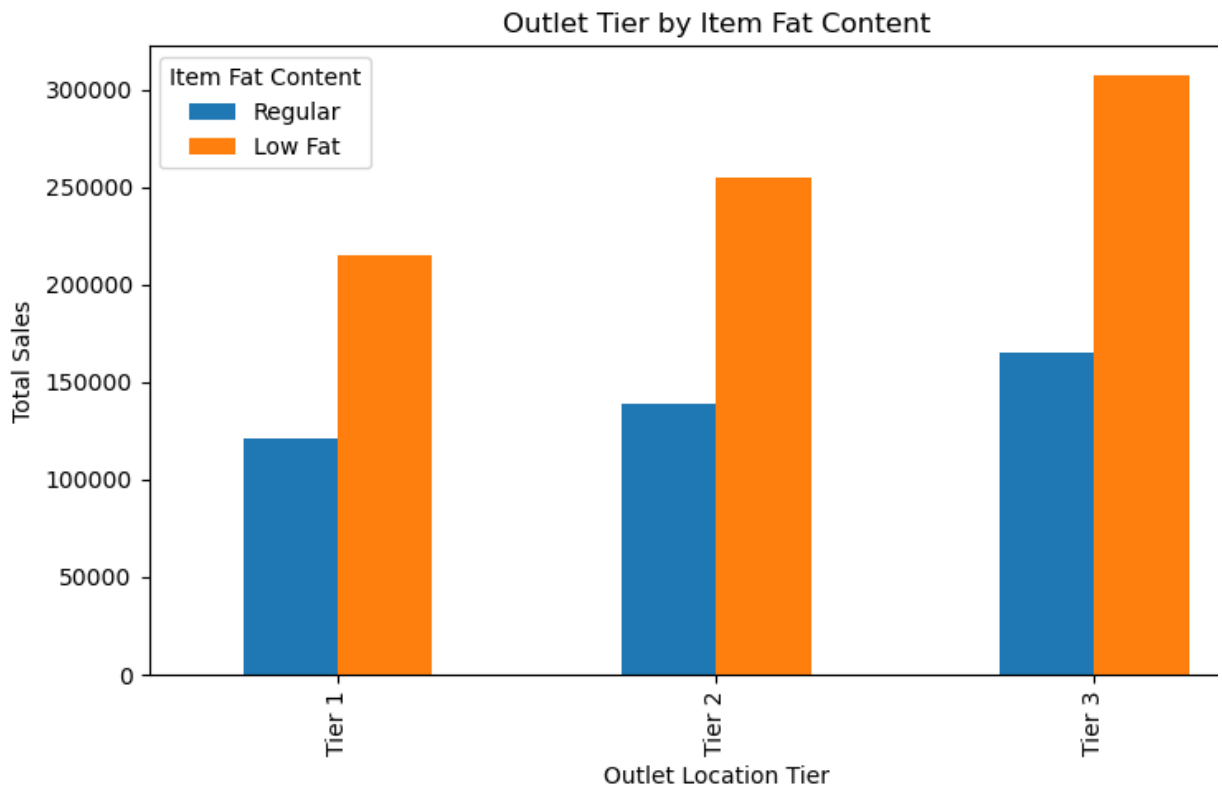

```
In [ ]: df.columns
```

Total Sales by Outlet Establishment

```
In [54]: sales_by_year =df.groupby('Outlet Establishment Year')['Sales'].sum().sort_index

plt.figure(figsize=(9,5))
plt.plot(sales_by_year.index,sales_by_year.values,marker ='o',linestyle='-')

plt.xlabel('Outlet Establishment Year')
plt.ylabel('Total Sales')
plt.title('Outlet Establishment')

for x,y in zip(sales_by_year.index,sales_by_year.values):
    plt.text(x,y,f'{y:,.0f}',ha ='center', va='bottom',fontsize=8)
plt.tight_layout()
plt.show()
```
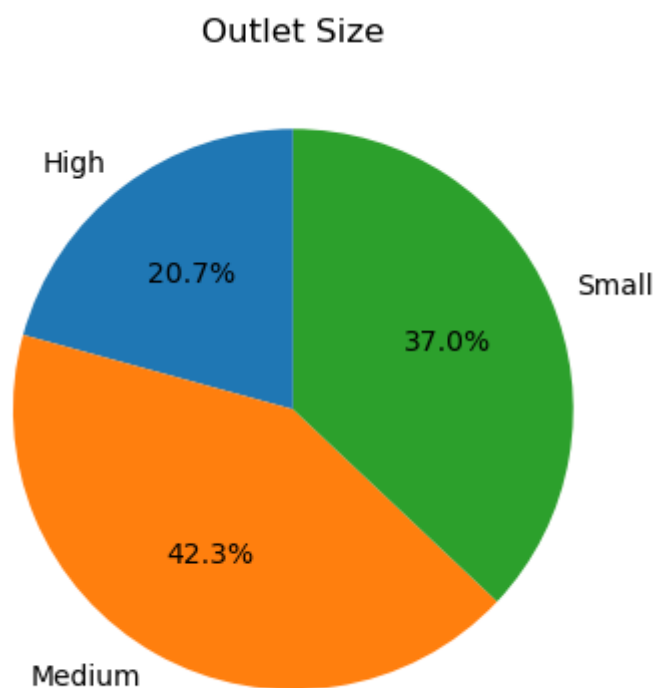
Outlet Establishment
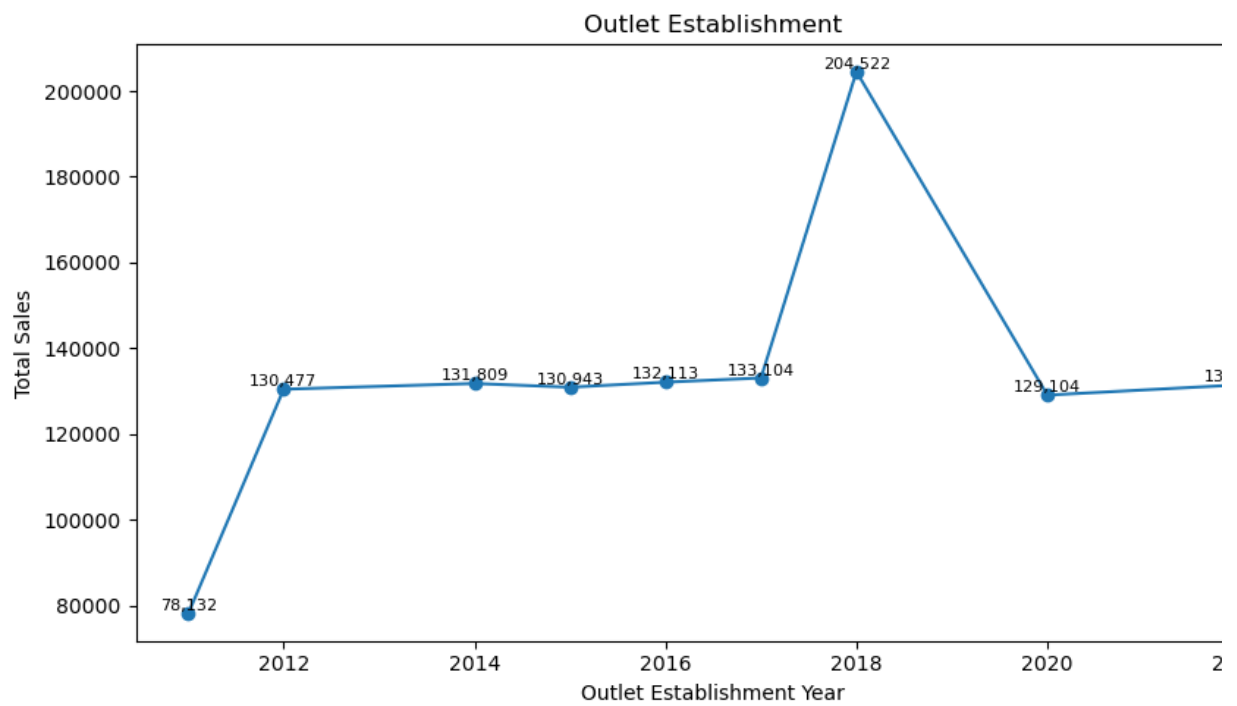
Sales by Outlet Size

```
In [61]:  sales_by_size =df.groupby('Outlet Size')['Sales'].sum()

          plt.figure(figsize=(4,4))
          plt.pie(sales_by_size,labels =sales_by_size.index,autopct='%1.1f%%',startangle =
          plt.title('Outlet Size')
          plt.tight_layout()
          plt.show()
```


Outlet Size

```
In [57]:  df.columns
```

```
Out[57]:  Index(['Item Fat Content', 'Item Identifier', 'Item Type',
                 'Outlet Establishment Year', 'Outlet Identifier',
                 'Outlet Location Type', 'Outlet Size', 'Outlet Type', 'Item Visibility',
                 'Item Weight', 'Sales', 'Rating'],
                dtype='object')
```
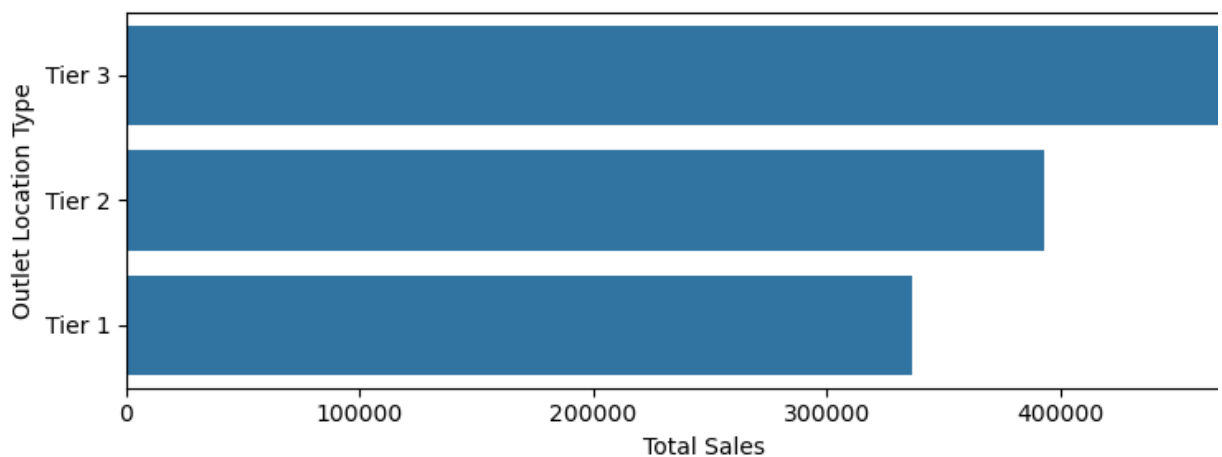
Sales by Outlet Location

```python
In [72]: sales_by_location =df.groupby('Outlet Location Type')['Sales'].sum().reset_index
         sales_by_location =sales_by_location.sort_values('Sales',ascending =False)

         plt.figure(figsize=(8,3))
         ax =sns.barplot(x='Sales',y ='Outlet Location Type', data=sales_by_location)
         plt.xlabel("Total Sales")
         plt.ylabel('Outlet Location Type')

         plt.tight_layout()
         plt.show
```

```
Out[72]:  <function matplotlib.pyplot.show(close=None, block=None)>
```



```
In [ ]:
```