Importing Libraries

In [4]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Import Data

In [5]:
```
df = pd.read_csv("C:/Lekshmi/DATA ANALYSIS PROJECTS/Blinkit - Python Project/b
```

View the first and last 10 items

In [88]:
```
df.head(10)
```

Out[88]:

	Item Fat Content	Item Identifier	Item Type	Outlet Establishment Year	Outlet Identifier	Outlet Location Type	Outlet Size	Outlet Type	Visib
0	Regular	FDX32	Fruits and Vegetables	2012	OUT049	Tier 1	Medium	Supermarket Type1	0.100
1	Low Fat	NCB42	Health and Hygiene	2022	OUT018	Tier 3	Medium	Supermarket Type2	0.008
2	Regular	FDR28	Frozen Foods	2010	OUT046	Tier 1	Small	Supermarket Type1	0.025
3	Regular	FDL50	Canned	2000	OUT013	Tier 3	High	Supermarket Type1	0.042
4	Low Fat	DRI25	Soft Drinks	2015	OUT045	Tier 2	Small	Supermarket Type1	0.033
5	Low Fat	FDS52	Frozen Foods	2020	OUT017	Tier 2	Small	Supermarket Type1	0.005
6	Low Fat	NCU05	Health and Hygiene	2011	OUT010	Tier 3	Small	Grocery Store	0.098
7	Low Fat	NCD30	Household	2015	OUT045	Tier 2	Small	Supermarket Type1	0.026
8	Low Fat	FDW20	Fruits and Vegetables	2000	OUT013	Tier 3	High	Supermarket Type1	0.024
9	Low Fat	FDX25	Canned	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.101

In [89]:

```
df.tail(10)
```

Out[89]:

	Item Fat Content	Item Identifier	Item Type	Outlet Establishment Year	Outlet Identifier	Outlet Location Type	Outlet Size	Outlet Type	V
8513	Regular	DRY23	Soft Drinks	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8514	Low Fat	FDA11	Baking Goods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8515	Low Fat	FDK38	Canned	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8516	Low Fat	FDO38	Canned	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8517	Low Fat	FDG32	Fruits and Vegetables	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8518	Low Fat	NCT53	Health and Hygiene	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8519	Low Fat	FDN09	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8520	Low Fat	DRE13	Soft Drinks	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8521	Regular	FDT50	Dairy	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8522	Regular	FDM58	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0

In [84]:

```
df.dtypes
```

Out[84]:

```
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
```

No of rows and Columns

```
In [86]: df.shape
```

```
Out[86]: (8523, 12)
```

```
In [87]: print('size of data is ',df.shape)
```

```
size of data is  (8523, 12)
```

Headers

```
In [11]: df.columns
```

```
Out[11]: Index(['Item Fat Content', 'Item Identifier', 'Item Type',
                'Outlet Establishment Year', 'Outlet Identifier',
                'Outlet Location Type', 'Outlet Size', 'Outlet Type', 'Item Visibilit
         y',
                'Item Weight', 'Sales', 'Rating'],
               dtype='object')
```

Data Cleaning

```
In [12]: print(df['Item Fat Content'].unique())
```

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

Normalizing - Converting to Standard Value

```
In [13]: df['Item Fat Content']=df['Item Fat Content'].replace({'LF':'Low Fat','low fat
```

```
In [14]: print(df['Item Fat Content'].unique())
```

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

KPI Requirement

In [15]:

```
total_sales=df['Sales'].sum()
avg_sales=df['Sales'].mean()
no_of_items_sold=df['Sales'].count()
avg_ratings=df['Rating'].mean()
```

In [16]:

```
print(f'Total Sales: ${total_sales: ,.1f}')
print(f'Average Sales: ${avg_sales: ,.0f}')
print(f'Total no of items sold:{no_of_items_sold: ,.0f}')
print(f'Average Ratings:{avg_ratings: ,.1f}')
```

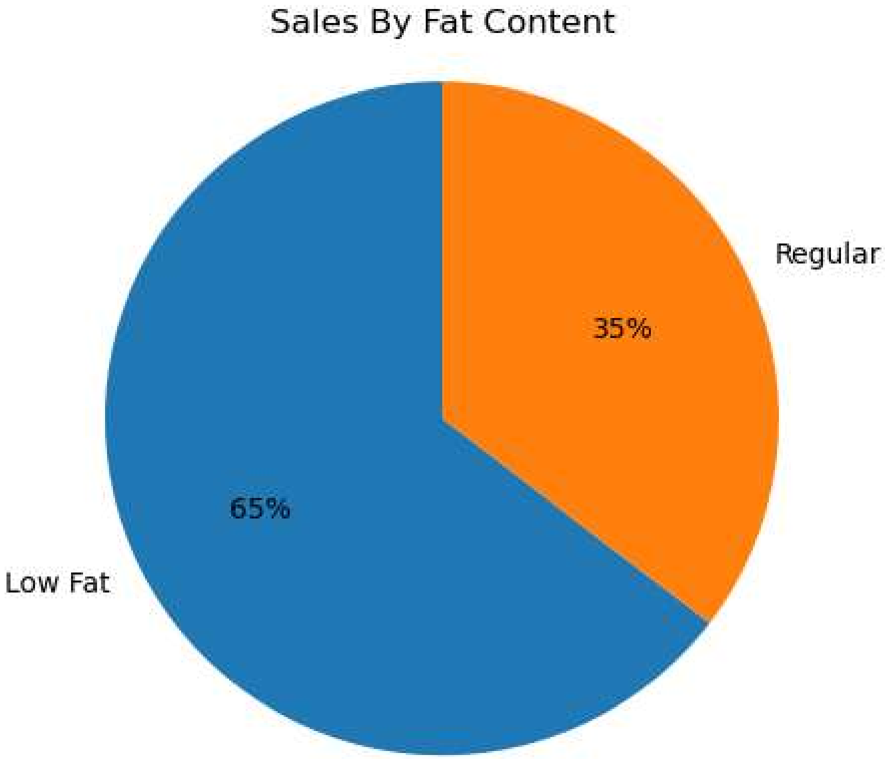
```
Total Sales: $ 1,201,681.5
Average Sales: $ 141
Total no of items sold: 8,523
Average Ratings: 4.0
```

Chart Requirements

Total Sales By Fat Content

In [17]:

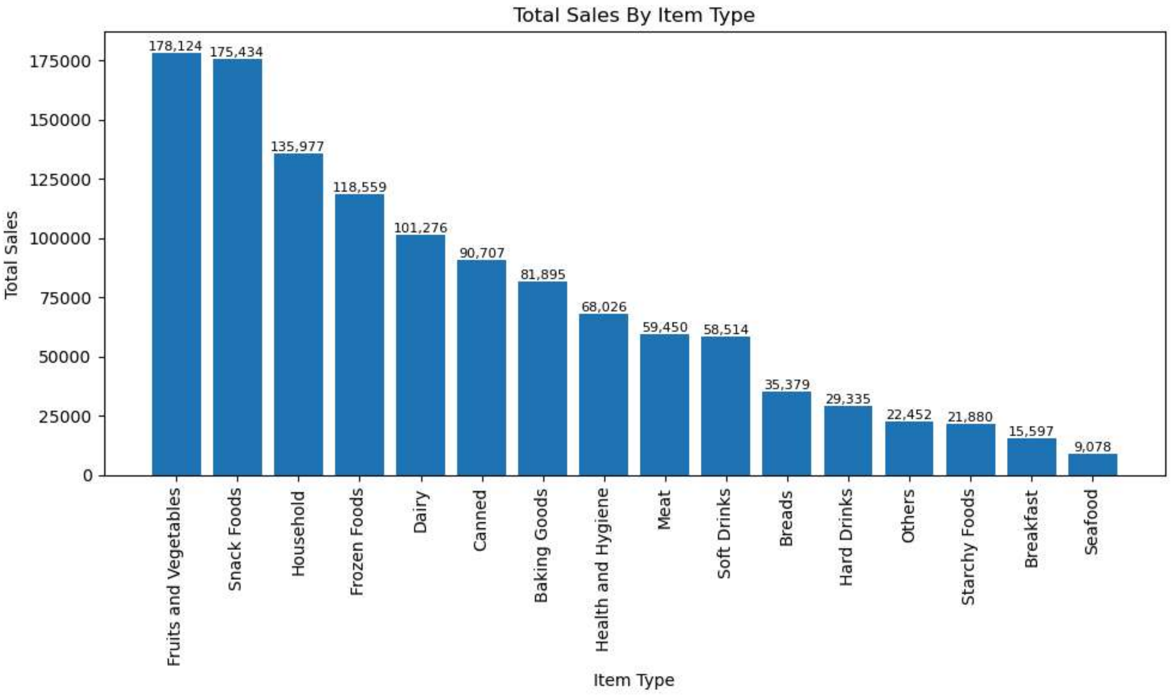
```
sales_by_fat_content=df.groupby('Item Fat Content')['Sales'].sum()
plt.pie(sales_by_fat_content,labels=sales_by_fat_content.index,autopct='%.0f%
plt.title('Sales By Fat Content')
plt.axis('equal')
plt.show()
```



Total Sales By Item Type

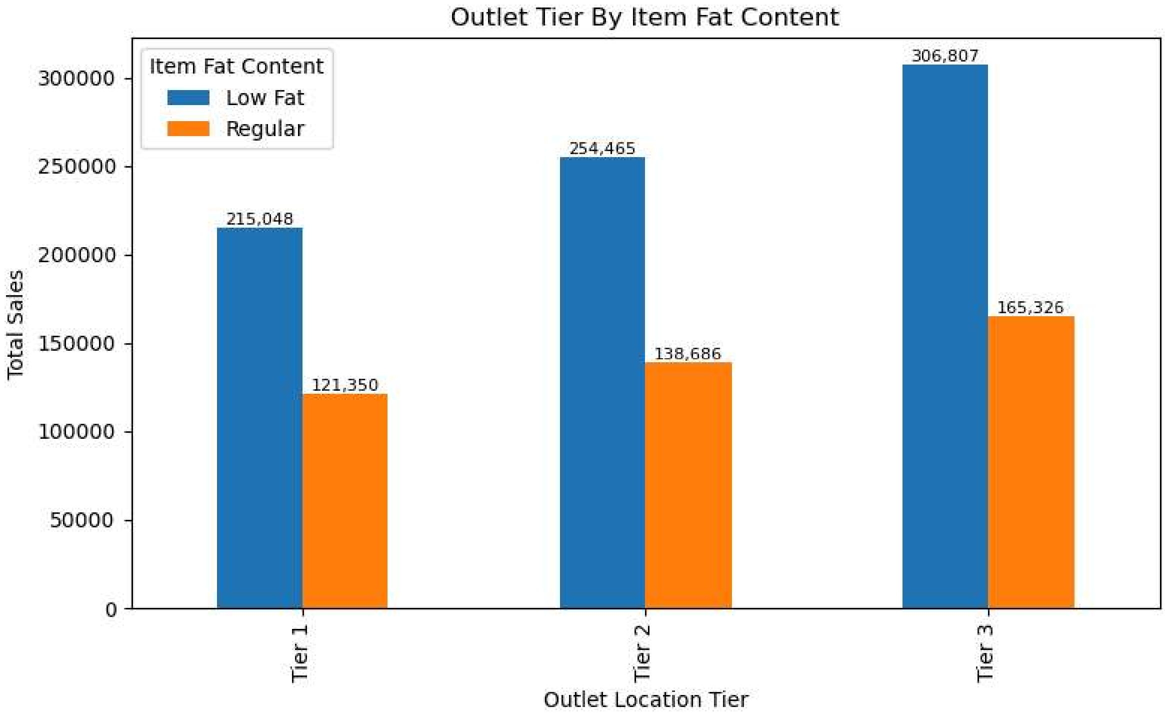
In [18]:

```
sales_by_type=df.groupby('Item Type')['Sales'].sum().sort_values(ascending=Fal
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:
        x=bar.get_x()+bar.get_width()/2
        y=bar.get_height()
        plt.text(
        x,
        y,
        f'{y:,.0f}',
        ha='center',
        va='bottom',
        fontsize=8)

plt.tight_layout()
plt.show()
```



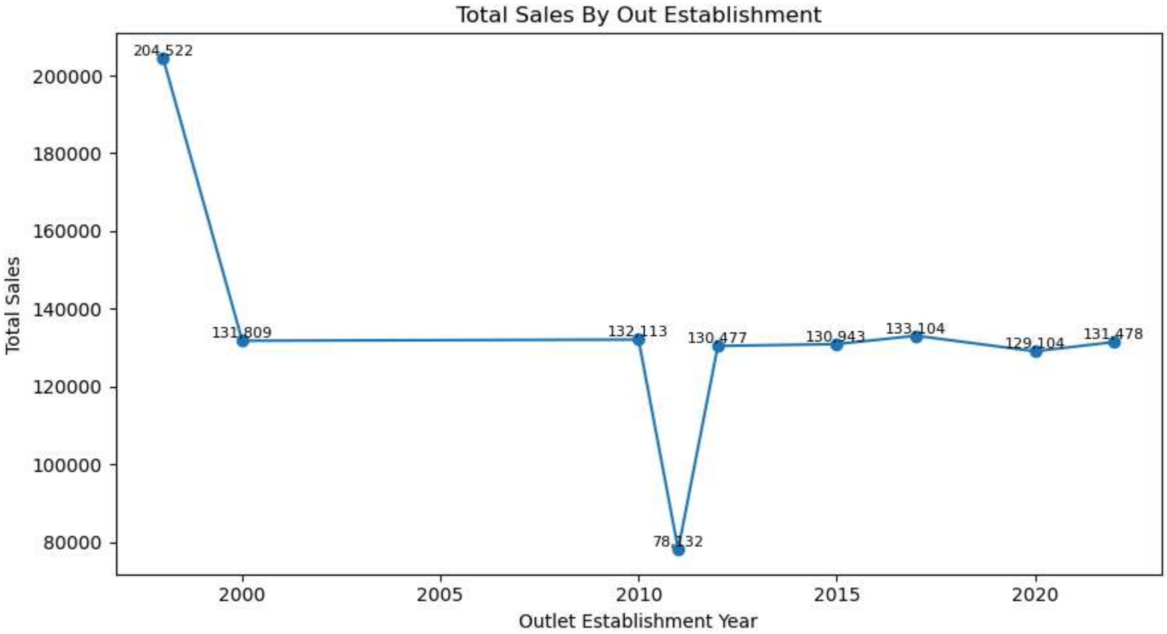
Fat Content By Outlet Type For Total Sales

In [21]:
```
grouped=df.groupby(['Outlet Location Type','Item Fat Content'])['Sales'].sum()
ax=grouped.plot(
    kind='bar',
    figsize=(8,5),
    title='Outlet Tier By Item Fat Content')

plt.xlabel('Outlet Location Tier')
plt.ylabel('Total Sales')
plt.legend(title='Item Fat Content')

for container in ax.containers:
    ax.bar_label(container, labels=[f'{v:,.0f}' for v in container.datavalues]
                 label_type='edge',fontsize=8)


plt.tight_layout()
plt.show()
```



Total Sales By Outlet Establishment Year

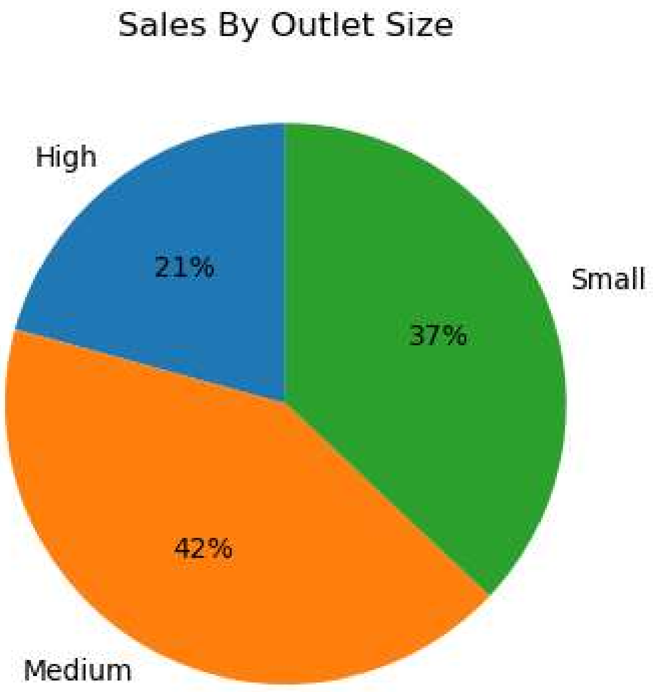
In [42]:
```
sales_by_year=df.groupby('Outlet Establishment Year')['Sales'].sum().sort_inde
plt.figure(figsize=(9,5))
plt.plot(sales_by_year.index,sales_by_year.values,marker='o',linestyle='-')
plt.xlabel('Outlet Establishment Year')
plt.ylabel('Total Sales')
plt.title('Total Sales By Out Establishment')

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()
```



Sales By Outlet Size

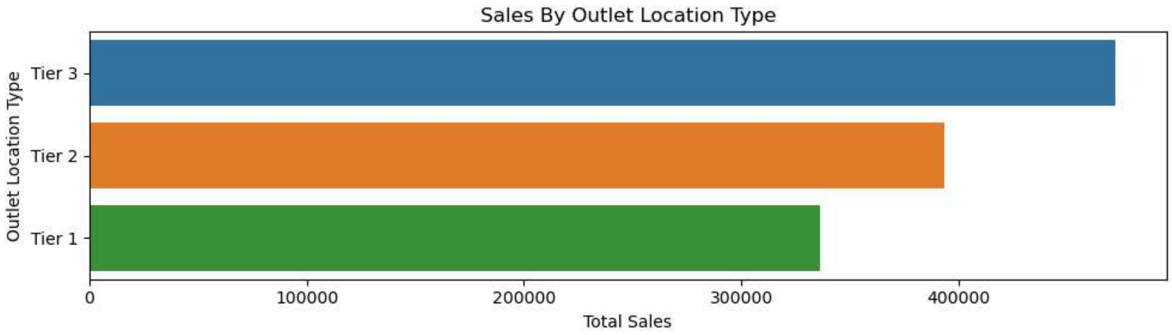
In [64]:

```
sales_by_outlet_size=df.groupby('Outlet Size')['Sales'].sum()
plt.figure(figsize=(4,4))
plt.pie(sales_by_outlet_size,labels=sales_by_outlet_size.index,autopct='%.0f%
plt.title('Sales By Outlet Size')
plt.tight_layout()
plt.show()
```



Sales By Outlet Location

In [83]:
```
sales_by_outlet_location=df.groupby('Outlet Location Type')['Sales'].sum()
sales_by_outlet_location=sales_by_outlet_location.sort_values(ascending=False
plt.figure(figsize=(10,3))
ax=sns.barplot(x='Sales',y='Outlet Location Type',data=sales_by_outlet_locatio

plt.title('Sales By Outlet Location Type')
plt.xlabel('Total Sales')
plt.ylabel('Outlet Location Type')
plt.tight_layout()
plt.show()
```


Sales By Outlet Location Type
Outlet Location Type
Tier 3
Tier 2
Tier 1
Total Sales