

```
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 30 items

In [6]: `df.head(30)`

Out[6]:

	Item Fat Content	Item Identifier	Item Type	Outlet Establishment Year	Outlet Identifier	Outlet Location Type	Outlet Size	Outlet Type	Visi
0	Regular	FDX32	Fruits and Vegetables	2012	OUT049	Tier 1	Medium	Supermarket Type1	0.10
1	Low Fat	NCB42	Health and Hygiene	2022	OUT018	Tier 3	Medium	Supermarket Type2	0.00
2	Regular	FDR28	Frozen Foods	2010	OUT046	Tier 1	Small	Supermarket Type1	0.02
3	Regular	FDL50	Canned	2000	OUT013	Tier 3	High	Supermarket Type1	0.02
4	Low Fat	DRI25	Soft Drinks	2015	OUT045	Tier 2	Small	Supermarket Type1	0.03
5	low fat	FDS52	Frozen Foods	2020	OUT017	Tier 2	Small	Supermarket Type1	0.00
6	Low Fat	NCU05	Health and Hygiene	2011	OUT010	Tier 3	Small	Grocery Store	0.09
7	Low Fat	NCD30	Household	2015	OUT045	Tier 2	Small	Supermarket Type1	0.02
8	Low Fat	FDW20	Fruits and Vegetables	2000	OUT013	Tier 3	High	Supermarket Type1	0.02
9	Low Fat	FDX25	Canned	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.10
10	LF	FDX21	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.08
11	Low Fat	NCU41	Health and Hygiene	2017	OUT035	Tier 2	Small	Supermarket Type1	0.09
12	Low Fat	FDL20	Fruits and Vegetables	2022	OUT018	Tier 3	Medium	Supermarket Type2	0.12
13	Low Fat	NCR54	Household	2000	OUT013	Tier 3	High	Supermarket Type1	0.09
14	Low Fat	FDH19	Meat	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.03
15	Regular	FDB57	Fruits and Vegetables	2017	OUT035	Tier 2	Small	Supermarket Type1	0.01
16	Low Fat	FDO23	Breads	2022	OUT018	Tier 3	Medium	Supermarket Type2	0.12
17	Low Fat	NCB07	Household	2012	OUT049	Tier 1	Medium	Supermarket Type1	0.07
18	Low Fat	FDJ56	Fruits and Vegetables	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.18
19	Low Fat	DRN47	Hard Drinks	2022	OUT018	Tier 3	Medium	Supermarket Type2	0.01
20	Regular	FDZ07	Fruits and Vegetables	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.00
21	Low Fat	NCK31	Others	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.02

	Item Fat Content	Item Identifier	Item Type	Establishment Year	Outlet Identifier	Outlet Location Type	Outlet Size	Outlet Type	Visi
22	Low Fat	FDJ41	Frozen Foods	2022	OUT018	Tier 3	Medium	Supermarket Type2	0.02
23	Low Fat	DRI51	Dairy	2022	OUT018	Tier 3	Medium	Supermarket Type2	0.04
24	Regular	FDC40	Dairy	2020	OUT017	Tier 2	Medium	Supermarket Type1	0.06
25	Low Fat	FDB53	Frozen Foods	2020	OUT017	Tier 2	Medium	Supermarket Type1	0.14
26	Low Fat	FDA16	Frozen Foods	2017	OUT035	Tier 2	Small	Supermarket Type1	0.03
27	Regular	FDO19	Fruits and Vegetables	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.01
28	Low Fat	FDO04	Frozen Foods	2010	OUT046	Tier 1	Small	Supermarket Type1	0.02
29	Regular	FDL25	Breakfast	2012	OUT049	Tier 1	Medium	Supermarket Type1	0.11

```
In [7]: df.tail(30)
```

Out[7]:

	Item Fat Content	Item Identifier	Item Type	Establishment Year	Outlet Identifier	Outlet Location Type	Outlet Size	Outlet Type	v
8493	Regular	FDU27	Meat	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8494	Regular	FDW27	Meat	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8495	Regular	FDX03	Meat	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8496	Regular	FDY03	Meat	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8497	Regular	FDG22	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8498	Regular	FDG46	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8499	Regular	FDJ21	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8500	Regular	FDK58	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8501	Regular	FDN34	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8502	Regular	FDP21	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8503	Regular	FDR22	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8504	Regular	FDS09	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8505	Regular	FDS34	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8506	Regular	FDU09	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8507	Regular	FDU33	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8508	Regular	FDU57	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8509	Regular	FDU58	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8510	Regular	FDX46	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8511	Regular	FDX57	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8512	Regular	FDY33	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
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

	Item Fat Content	Item Identifier	Item Type	Establishment Year	Outlet Identifier	Outlet Location Type	Outlet Size	Outlet Type	v
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	reg	FDT50	Dairy	1998	OUT027	Tier 3	Medium	Supermarket Type3	0
8522	reg	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}')
```

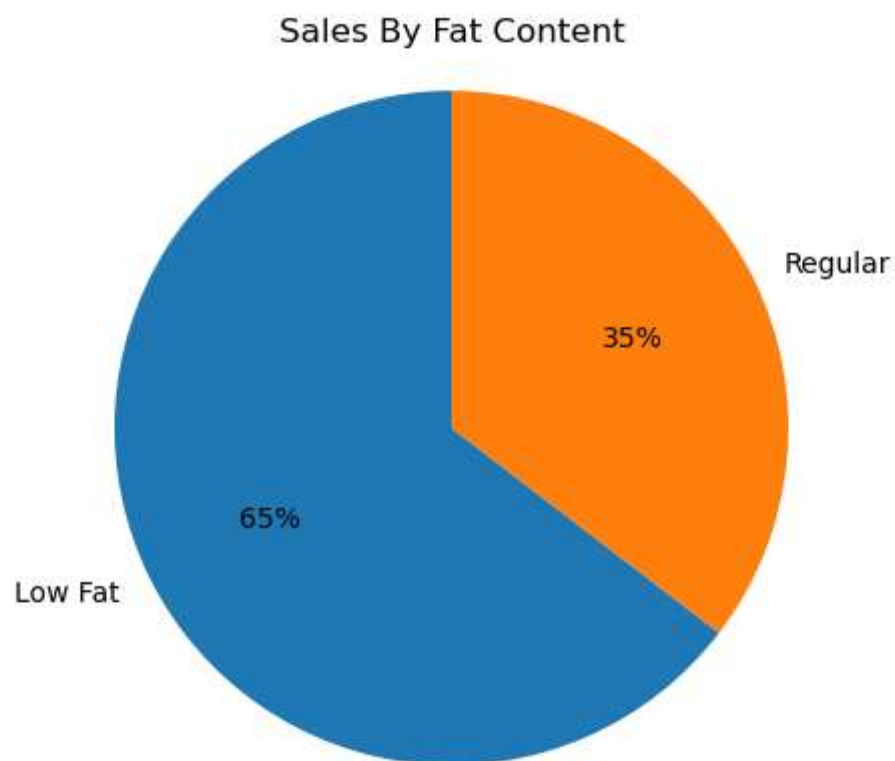
```
print(f'Average Ratings:{avg_ratings: ,.1f}')
```

```
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()
```

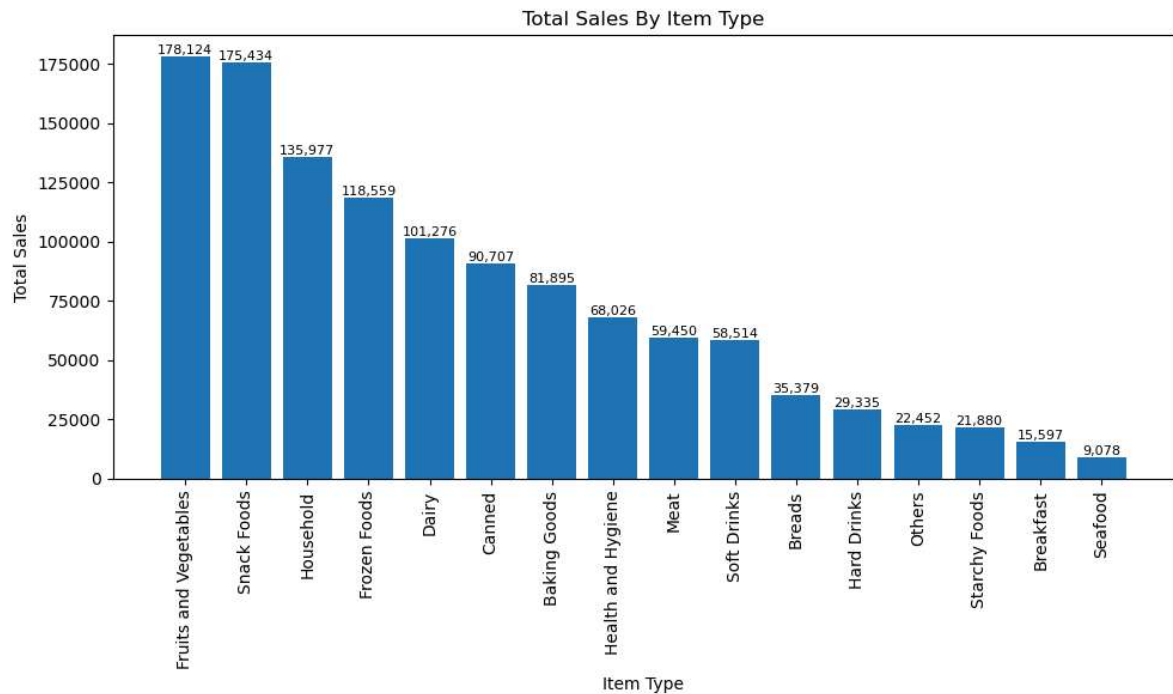


Total Sales By Item Type

```
In [18]: sales_by_type=df.groupby('Item Type')['Sales'].sum().sort_values(ascending=False)
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()
```



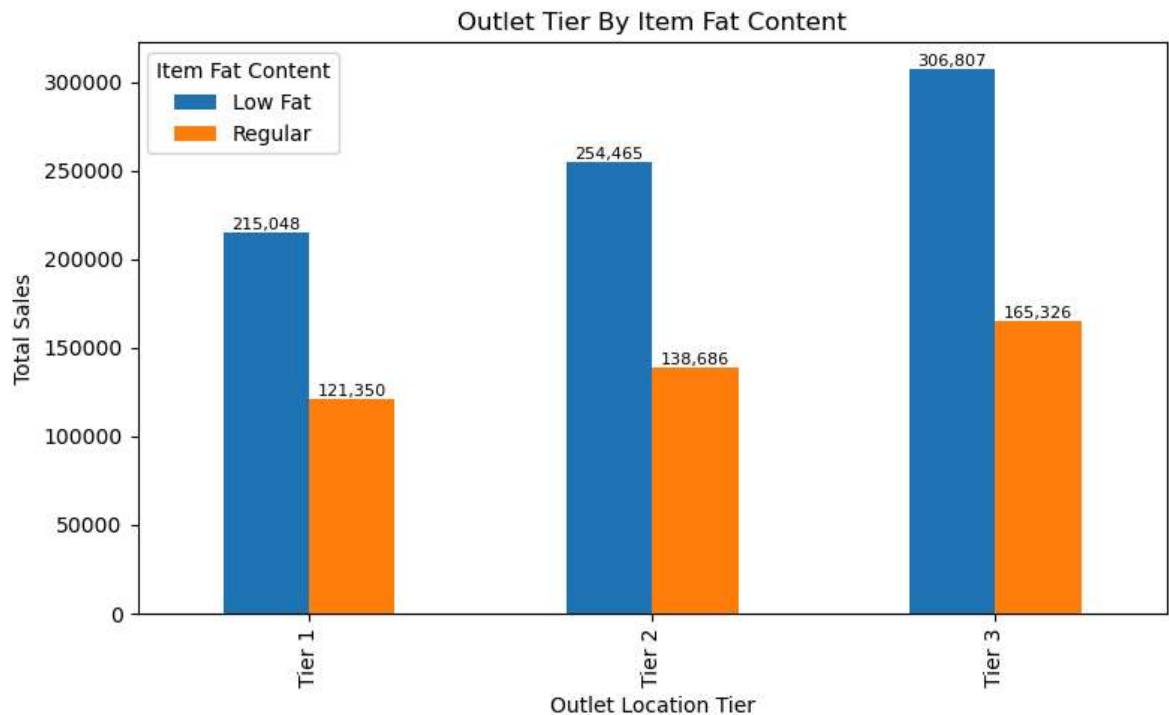
Fat Content By Outlet Type For Total Sales

```
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()
```

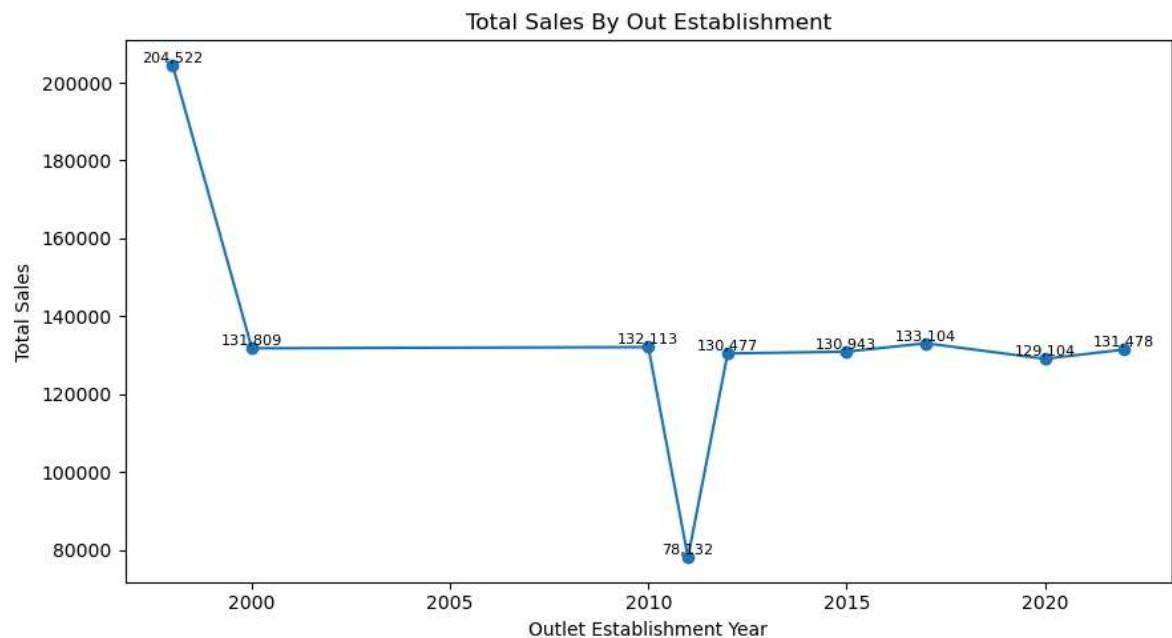


Total Sales By Outlet Establishment Year

```
In [42]: sales_by_year=df.groupby('Outlet Establishment Year')['Sales'].sum().sort_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('Total Sales By Out Establishment')

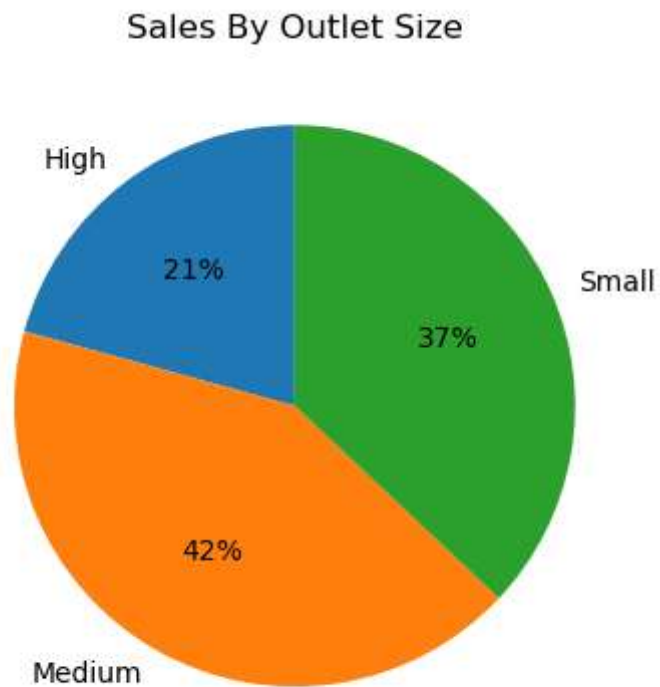
for x,y in zip(sales_by_year.index,sales_by_year.values):
    plt.text(x,y,f'{y:,.0f}',ha='center',va='bottom',fontsize=8)

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_location)

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

