DATA ANALYSIS PYTHON PROJECT - BLINKIT ANALYSIS

Import Libraries

```
In [1]: import pandas as pd
   import numpy as np
   import matplotlib.pyplot as plt
   import seaborn as sns
```

Import Raw Data

```
In [4]: df = pd.read_csv(r"C:\Users\TEMP USER\Downloads\blinkit_data.csv")
```

Sample Data

```
In [6]: df.head(20)
```

Out[6]:

	Item Fat Content	Item Identifier	Item Type	Outlet Establishment Year	Outlet Identifier	Outlet Location Type	Outlet Size	Outlet Typ
0	Regular	FDX32	Fruits and Vegetables	2012	OUT049	Tier 1	Medium	Supermark Type
1	Low Fat	NCB42	Health and Hygiene	2022	OUT018	Tier 3	Medium	Supermark Typa
2	Regular	FDR28	Frozen Foods	2010	OUT046	Tier 1	Small	Supermark Typ:
3	Regular	FDL50	Canned	2000	OUT013	Tier 3	High	Supermark Type
4	Low Fat	DRI25	Soft Drinks	2015	OUT045	Tier 2	Small	Supermark Typ:
5	low fat	FDS52	Frozen Foods	2020	OUT017	Tier 2	Small	Supermark Type
6	Low Fat	NCU05	Health and Hygiene	2011	OUT010	Tier 3	Small	Groce Sto
7	Low Fat	NCD30	Household	2015	OUT045	Tier 2	Small	Supermark Type
8	Low Fat	FDW20	Fruits and Vegetables	2000	OUT013	Tier 3	High	Supermark Typ:
9	Low Fat	FDX25	Canned	1998	OUT027	Tier 3	Medium	Supermark Type
10	LF	FDX21	Snack Foods	1998	OUT027	Tier 3	Medium	Supermark Typ:
11	Low Fat	NCU41	Health and Hygiene	2017	OUT035	Tier 2	Small	Supermark Type
12	Low Fat	FDL20	Fruits and Vegetables	2022	OUT018	Tier 3	Medium	Supermark Typ:
13	Low Fat	NCR54	Household	2000	OUT013	Tier 3	High	Supermark Type
14	Low Fat	FDH19	Meat	1998	OUT027	Tier 3	Medium	Supermark Typ:
15	Regular	FDB57	Fruits and Vegetables	2017	OUT035	Tier 2	Small	Supermark Type
16	Low Fat	FDO23	Breads	2022	OUT018	Tier 3	Medium	Supermark Typ:

	Item Fat Content	ltem Identifier	Item Type	Outlet Establishment Year	Outlet Identifier	Outlet Location Type	Outlet Size	Outlet Typ
17	7 Low Fat	NCB07	Household	2012	OUT049	Tier 1	Medium	Supermark Type
18	B Low Fat	FDJ56	Fruits and Vegetables	1998	OUT027	Tier 3	Medium	Supermark Typ:
19	9 Low Fat	DRN47	Hard Drinks	2022	OUT018	Tier 3	Medium	Supermark Type

In [8]: df.tail(20)

	Item Fat Content	Item Identifier	Item Type	Outlet Establishment Year	Outlet Identifier	Outlet Location Type	Outlet Size	Outlet ⁻
8503	Regular	FDR22	Snack Foods	1998	OUT027	Tier 3	Medium	Superma T <u>'</u>
8504	Regular	FDS09	Snack Foods	1998	OUT027	Tier 3	Medium	Superma T <u>'</u>
8505	Regular	FDS34	Snack Foods	1998	OUT027	Tier 3	Medium	Superma T <u>'</u>
8506	Regular	FDU09	Snack Foods	1998	OUT027	Tier 3	Medium	Superma T <u>'</u>
8507	Regular	FDU33	Snack Foods	1998	OUT027	Tier 3	Medium	Superma T <u>'</u>
8508	Regular	FDU57	Snack Foods	1998	OUT027	Tier 3	Medium	Superma T <u>'</u>
8509	Regular	FDU58	Snack Foods	1998	OUT027	Tier 3	Medium	Superma T <u>'</u>
8510	Regular	FDX46	Snack Foods	1998	OUT027	Tier 3	Medium	Superma T <u>'</u>
8511	Regular	FDX57	Snack Foods	1998	OUT027	Tier 3	Medium	Superma T <u>'</u>
8512	Regular	FDY33	Snack Foods	1998	OUT027	Tier 3	Medium	Superma T <u>'</u>
8513	Regular	DRY23	Soft Drinks	1998	OUT027	Tier 3	Medium	Superma T <u>'</u>
8514	low fat	FDA11	Baking Goods	1998	OUT027	Tier 3	Medium	Superma T <u>'</u>
8515	low fat	FDK38	Canned	1998	OUT027	Tier 3	Medium	Superma T <u>'</u>
8516	low fat	FDO38	Canned	1998	OUT027	Tier 3	Medium	Superma T <u>'</u>
8517	low fat	FDG32	Fruits and Vegetables	1998	OUT027	Tier 3	Medium	Superma T <u>'</u>
8518	low fat	NCT53	Health and Hygiene	1998	OUT027	Tier 3	Medium	Superma T <u>'</u>
8519	low fat	FDN09	Snack Foods	1998	OUT027	Tier 3	Medium	Superma T <u>'</u>
8520	low fat	DRE13	Soft Drinks	1998	OUT027	Tier 3	Medium	Superma T <u>'</u>
	8504 8505 8506 8507 8508 8509 8510 8511 8512 8513 8514 8515 8516 8517	Fat Content 8503 Regular 8504 Regular 8505 Regular 8506 Regular 8508 Regular 8509 Regular 8511 Regular 8512 Regular 8513 Regular 8514 Iow fat 8515 Iow fat 8516 Iow fat 8517 Iow fat 8518 Iow fat 8519 Iow fat	Fat Content Item Identifier 8503 Regular FDR22 8504 Regular FDS09 8505 Regular FDU09 8506 Regular FDU33 8508 Regular FDU57 8509 Regular FDX46 8511 Regular FDX57 8512 Regular FDX57 8513 Regular DRY23 8514 Iow fat FDA11 8515 Iow fat FDK38 8516 Iow fat FDO38 8517 Iow fat FDG32 8518 Iow fat NCT53 8519 Iow fat FDN09	Fat ContentItem Type8503RegularFDR22Snack Foods8504RegularFDS09Snack Foods8505RegularFDS34Snack Foods8506RegularFDU09Snack Foods8507RegularFDU33Snack Foods8508RegularFDU57Snack Foods8510RegularFDU58Snack Foods8511RegularFDX57Snack Foods8512RegularFDX57Snack Foods8513RegularFDY33Snack Foods8514Iow fatFDA11Baking Goods8515Iow fatFDA38Canned8516Iow fatFDG32Fruits and Vegetables8517Iow fatFDG32Fruits and Hygiene8518Iow fatFDN03Snack Foods8519Iow fatFDN09Snack Foods8520Iow fatFDN09Snack Foods	Fat Content Item Type Establishment Year 8503 Regular FDR22 Snack Foods 1998 8504 Regular FDS09 Snack Foods 1998 8505 Regular FDS34 Snack Foods 1998 8506 Regular FDU09 Snack Foods 1998 8507 Regular FDU33 Snack Foods 1998 8508 Regular FDU57 Snack Foods 1998 8509 Regular FDU58 Snack Foods 1998 8510 Regular FDX57 Snack Foods 1998 8511 Regular FDX57 Snack Foods 1998 8512 Regular FDY33 Snack Foods 1998 8513 Regular FDY33 Snack Foods 1998 8514 Iow fat FDA11 Baking Goods 1998 8515 Iow fat FDX38 Canned 1998 8516 Iow fat FDG32 Fruits and Canned	Fat Content Item Type Item Type Item Type Establishment Year Counter Item Type 8503 Regular FDR22 Snack Foods 1998 OUT027 8504 Regular FDS09 Snack Foods 1998 OUT027 8505 Regular FDU34 Snack Foods 1998 OUT027 8506 Regular FDU09 Snack Foods 1998 OUT027 8507 Regular FDU57 Snack Foods 1998 OUT027 8508 Regular FDU58 Snack Foods 1998 OUT027 8509 Regular FDU58 Snack Foods 1998 OUT027 8510 Regular FDX57 Snack Foods 1998 OUT027 8511 Regular FDX57 Snack Foods 1998 OUT027 8512 Regular FDX57 Snack Foods 1998 OUT027 8513 Regular FDX13 Soft Foods 1998 OUT027 8514 Iow fat FDX38 </th <th>Fat Content Item Intentifier Content Item Type Establishment Year Outroit Location Type 8503 Regular FDR22 Snack Foods Foods 1998 OUT027 Tier 3 8504 Regular FDS09 Snack Foods Foods Foods 1998 OUT027 Tier 3 8505 Regular FDU09 Snack Foods Foods Foods 1998 OUT027 Tier 3 8506 Regular FDU33 Snack Foods Foods Foods Foods 1998 OUT027 Tier 3 8508 Regular FDU57 Snack Foods 1998 OUT027 Tier 3 8510 Regular FDX46 Snack Foods Foods</th> <th>Fat Content Item Type Establishment Year Outlet Gentifier Location Type Size 8503 Regular FDR22 Snack Foods 1998 OUT027 Tier 3 Medium 8504 Regular FDS09 Snack Foods 1998 OUT027 Tier 3 Medium 8505 Regular FDS34 Snack Foods 1998 OUT027 Tier 3 Medium 8506 Regular FDU09 Snack Foods 1998 OUT027 Tier 3 Medium 8507 Regular FDU33 Snack Foods 1998 OUT027 Tier 3 Medium 8508 Regular FDU57 Snack Foods 1998 OUT027 Tier 3 Medium 8510 Regular FDU58 Snack Foods 1998 OUT027 Tier 3 Medium 8511 Regular FDX57 Snack Foods 1998 OUT027 Tier 3 Medium 8512 Regular FDX57 Snack Foods 1998 OUT027<!--</th--></th>	Fat Content Item Intentifier Content Item Type Establishment Year Outroit Location Type 8503 Regular FDR22 Snack Foods Foods 1998 OUT027 Tier 3 8504 Regular FDS09 Snack Foods Foods Foods 1998 OUT027 Tier 3 8505 Regular FDU09 Snack Foods Foods Foods 1998 OUT027 Tier 3 8506 Regular FDU33 Snack Foods Foods Foods Foods 1998 OUT027 Tier 3 8508 Regular FDU57 Snack Foods 1998 OUT027 Tier 3 8510 Regular FDX46 Snack Foods	Fat Content Item Type Establishment Year Outlet Gentifier Location Type Size 8503 Regular FDR22 Snack Foods 1998 OUT027 Tier 3 Medium 8504 Regular FDS09 Snack Foods 1998 OUT027 Tier 3 Medium 8505 Regular FDS34 Snack Foods 1998 OUT027 Tier 3 Medium 8506 Regular FDU09 Snack Foods 1998 OUT027 Tier 3 Medium 8507 Regular FDU33 Snack Foods 1998 OUT027 Tier 3 Medium 8508 Regular FDU57 Snack Foods 1998 OUT027 Tier 3 Medium 8510 Regular FDU58 Snack Foods 1998 OUT027 Tier 3 Medium 8511 Regular FDX57 Snack Foods 1998 OUT027 Tier 3 Medium 8512 Regular FDX57 Snack Foods 1998 OUT027 </th

	Item Fat Content	Item Identifier	Item Type	Outlet Establishment Year	Outlet Identifier	Outlet Location Type	Outlet Size	Outlet ⁻
8521	reg	FDT50	Dairy	1998	OUT027	Tier 3	Medium	Superma T <u>'</u>
8522	reg	FDM58	Snack Foods	1998	OUT027	Tier 3	Medium	Superma T <u>'</u>

Size of Data

```
In [11]: print("size of Data:", df.shape)
    size of Data: (8523, 12)
```

Field info

Data Types

```
In [13]:
         df.dtypes
Out[13]: 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
                                       float64
         Rating
         dtype: object
```

Data Cleaning

BUSINESS REQUIREMENTS

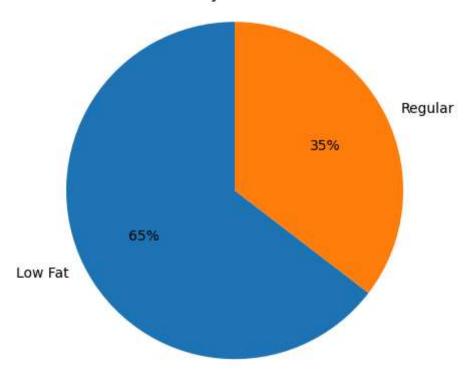
KPI's REQUIREMENTS

```
In [28]: #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_rating = df['Rating'].mean()
         #Display
         print(f"Total sales: ${total_sales:,.0f}")
         print(f"Average sales: ${avg sales:,.1f}")
         print(f"No_of_Items_sold: {no_of_items_sold:,.0f}")
         print(f"Average Ratings: {avg_rating:,.1f}")
        Total sales: $1,201,681
        Average sales: $141.0
        No_of_Items_sold: 8,523
        Average Ratings: 4.0
```

CHARTS REQUIREMENTS

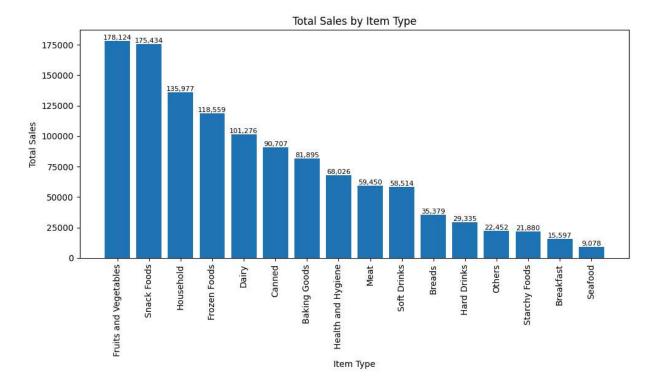
Total Sales by Fat Content:

Sales by fat Content



Total Sales by Item Type

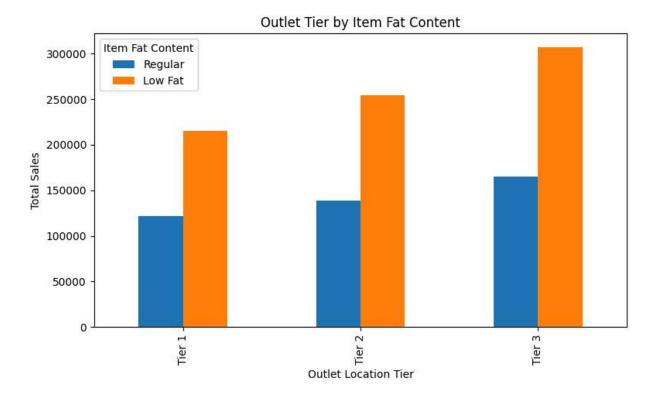
```
In [36]: sales_by_type = df.groupby('Item Type')['Sales'].sum().sort_values(ascending=False)
         plt.figure(figsize=(10, 6)) # Ensure reasonable size
         bars = plt.bar(sales_by_type.index, sales_by_type.values)
         plt.xticks(rotation=90)
         plt.xlabel('Item Type')
         plt.ylabel('Total Sales')
         plt.title('Total Sales by Item Type')
         # Annotate bars
         for bar in bars:
             plt.text(
                  bar.get_x() + bar.get_width() / 2,
                 bar.get_height(),
                  f'{bar.get_height():,.0f}', # CORRECTED f-string
                 ha='center',
                 va='bottom',
                 fontsize=8
             )
         plt.tight_layout()
         plt.show()
```



Fat Content by Outlet for Total Sales

```
In [38]: grouped = df.groupby(['Outlet Location Type', 'Item Fat Content'])['Sales'].sum().u
grouped = grouped[['Regular', 'Low Fat']]

ax = grouped.plot(kind='bar', figsize=(8, 5), title='Outlet Tier by Item Fat Conten
plt.xlabel('Outlet Location Tier')
plt.ylabel('Total Sales')
plt.legend(title='Item Fat Content')
plt.tight_layout()
plt.show()
```

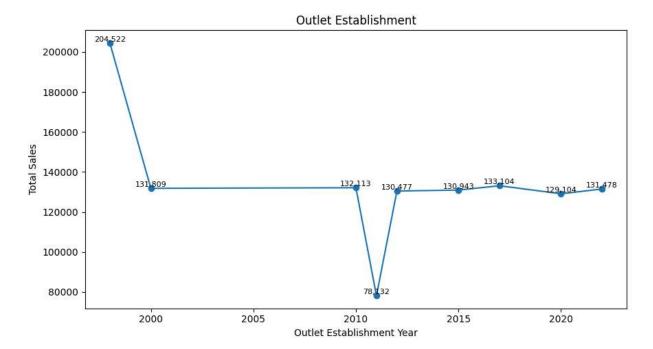


Total Sales by Outlet Establishment

```
In [39]: 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()
```

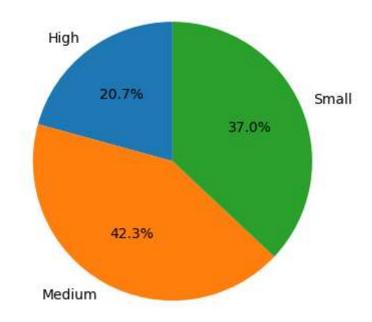


Sales by Outlet Size

```
In [40]: 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=90
plt.title('Outlet Size')
plt.tight_layout()
plt.show()
```

Outlet Size



Sales by Outlet Location

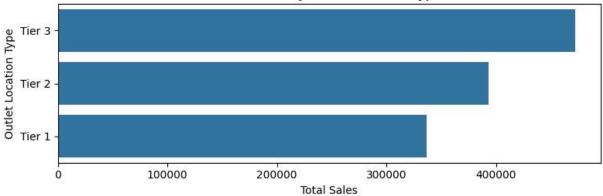
```
In [41]:
    sales_by_location = df.groupby('Outlet Location Type')['Sales'].sum().reset_index()
    sales_by_location = sales_by_location.sort_values('Sales', ascending=False)

plt.figure(figsize=(8, 3)) # Smaller height enough width
    ax = sns.barplot(x='Sales', y='Outlet Location Type', data=sales_by_location)

plt.title('Total Sales by Outlet Location Type')
    plt.xlabel('Total Sales')
    plt.ylabel('Outlet Location Type')

plt.tight_layout() # Ensure Layout fits without scroll
    plt.show()
```

Total Sales by Outlet Location Type



In []: