

Import Libraries

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

Import Raw Data
```

```
In [2]: df = pd.read_csv('blinkit_data.csv')
```

Sample Data

```
In [3]: df
```

| | 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 |
|------|------------------|-----------------|-----------------------|---------------------------|-------------------|----------------------|-------------|-------------------|-----------------|-------------|----------|--------|
| 0 | Regular | FDX32 | Fruits and Vegetables | 2012 | OUT049 | Tier 1 | Medium | Supermarket Type1 | 0.100014 | 15.10 | 145.4786 | 5.0 |
| 1 | Low Fat | NCB42 | Health and Hygiene | 2022 | OUT018 | Tier 3 | Medium | Supermarket Type2 | 0.008596 | 11.80 | 115.3492 | 5.0 |
| 2 | Regular | FDR28 | Frozen Foods | 2010 | OUT046 | Tier 1 | Small | Supermarket Type1 | 0.025896 | 13.85 | 165.0210 | 5.0 |
| 3 | Regular | FDL50 | Canned | 2000 | OUT013 | Tier 3 | High | Supermarket Type1 | 0.042278 | 12.15 | 126.5046 | 5.0 |
| 4 | Low Fat | DRD25 | Soft Drinks | 2015 | OUT045 | Tier 2 | Small | Supermarket Type1 | 0.033970 | 19.60 | 55.1614 | 5.0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 8518 | low fat | NCT53 | Health and Hygiene | 1998 | OUT027 | Tier 3 | Medium | Supermarket Type3 | 0.000000 | NaN | 164.5526 | 4.0 |
| 8519 | low fat | FDN09 | Snack Foods | 1998 | OUT027 | Tier 3 | Medium | Supermarket Type3 | 0.034706 | NaN | 241.6828 | 4.0 |
| 8520 | low fat | DRE13 | Soft Drinks | 1998 | OUT027 | Tier 3 | Medium | Supermarket Type3 | 0.027571 | NaN | 86.6198 | 4.0 |
| 8521 | reg | FDT50 | Dairy | 1998 | OUT027 | Tier 3 | Medium | Supermarket Type3 | 0.107715 | NaN | 97.8752 | 4.0 |
| 8522 | reg | FDM58 | Snack Foods | 1998 | OUT027 | Tier 3 | Medium | Supermarket Type3 | 0.000000 | NaN | 112.2544 | 4.0 |

8523 rows × 12 columns

```
In [4]: df.head(5)
```

| | 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 |
|---|------------------|-----------------|-----------------------|---------------------------|-------------------|----------------------|-------------|-------------------|-----------------|-------------|----------|--------|
| 0 | Regular | FDX32 | Fruits and Vegetables | 2012 | OUT049 | Tier 1 | Medium | Supermarket Type1 | 0.100014 | 15.10 | 145.4786 | 5.0 |
| 1 | Low Fat | NCB42 | Health and Hygiene | 2022 | OUT018 | Tier 3 | Medium | Supermarket Type2 | 0.008596 | 11.80 | 115.3492 | 5.0 |
| 2 | Regular | FDR28 | Frozen Foods | 2010 | OUT046 | Tier 1 | Small | Supermarket Type1 | 0.025896 | 13.85 | 165.0210 | 5.0 |
| 3 | Regular | FDL50 | Canned | 2000 | OUT013 | Tier 3 | High | Supermarket Type1 | 0.042278 | 12.15 | 126.5046 | 5.0 |
| 4 | Low Fat | DRD25 | Soft Drinks | 2015 | OUT045 | Tier 2 | Small | Supermarket Type1 | 0.033970 | 19.60 | 55.1614 | 5.0 |

```
In [5]: df.tail(5)
```

| | 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 |
|------|------------------|-----------------|--------------------|---------------------------|-------------------|----------------------|-------------|-------------------|-----------------|-------------|----------|--------|
| 8518 | low fat | NCT53 | Health and Hygiene | 1998 | OUT027 | Tier 3 | Medium | Supermarket Type3 | 0.000000 | NaN | 164.5526 | 4.0 |
| 8519 | low fat | FDN09 | Snack Foods | 1998 | OUT027 | Tier 3 | Medium | Supermarket Type3 | 0.034706 | NaN | 241.6828 | 4.0 |
| 8520 | low fat | DRE13 | Soft Drinks | 1998 | OUT027 | Tier 3 | Medium | Supermarket Type3 | 0.027571 | NaN | 86.6198 | 4.0 |
| 8521 | reg | FDT50 | Dairy | 1998 | OUT027 | Tier 3 | Medium | Supermarket Type3 | 0.107715 | NaN | 97.8752 | 4.0 |
| 8522 | reg | FDM58 | Snack Foods | 1998 | OUT027 | Tier 3 | Medium | Supermarket Type3 | 0.000000 | NaN | 112.2544 | 4.0 |

```
In [6]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8523 entries, 0 to 8522
Data columns (total 12 columns):
 #   Column              Non-Null Count  Dtype
---  -
 0   Item Fat Content    8523 non-null   object
 1   Item Identifier     8523 non-null   object
 2   Item Type           8523 non-null   object
 3   Outlet Establishment Year  8523 non-null   int64
 4   Outlet Identifier   8523 non-null   object
 5   Outlet Location Type  8523 non-null   object
 6   Outlet Size         8523 non-null   object
 7   Outlet Type         8523 non-null   object
 8   Item Visibility     8523 non-null   float64
 9   Item Weight         7060 non-null   float64
10   Sales               8523 non-null   float64
11   Rating              8523 non-null   object (?)
dtypes: float64(4), int64(1), object(7)
memory usage: 799.2+ KB
```

```
In [8]: df.describe()
```

| | Outlet Establishment Year | Item Visibility | Item Weight | Sales | Rating |
|-------|---------------------------|-----------------|-------------|-------------|-------------|
| count | 8523.000000 | 8523.000000 | 7060.000000 | 8523.000000 | 8523.000000 |
| mean | 2010.831867 | 0.066132 | 12.857645 | 140.992782 | 3.965857 |
| std | 8.371760 | 0.051598 | 4.643456 | 62.275067 | 0.605651 |
| min | 1998.000000 | 0.000000 | 4.555000 | 31.290000 | 1.000000 |
| 25% | 2000.000000 | 0.026989 | 8.773750 | 93.826500 | 4.000000 |
| 50% | 2012.000000 | 0.053931 | 12.600000 | 143.012800 | 4.000000 |
| 75% | 2017.000000 | 0.094585 | 16.850000 | 185.643700 | 4.200000 |
| max | 2022.000000 | 0.328391 | 21.350000 | 266.884000 | 5.000000 |

```
In [9]: df.isnull()
```

| | 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 |
|------|------------------|-----------------|-----------|---------------------------|-------------------|----------------------|-------------|-------------|-----------------|-------------|-------|--------|
| 0 | False | False | False | False | False | False | False | False | False | False | False | False |
| 1 | False | False | False | False | False | False | False | False | False | False | False | False |
| 2 | False | False | False | False | False | False | False | False | False | False | False | False |
| 3 | False | False | False | False | False | False | False | False | False | False | False | False |
| 4 | False | False | False | False | False | False | False | False | False | False | False | False |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 8518 | False | False | False | False | False | False | False | False | False | True | False | False |
| 8519 | False | False | False | False | False | False | False | False | False | True | False | False |
| 8520 | False | False | False | False | False | False | False | False | False | True | False | False |
| 8521 | False | False | False | False | False | False | False | False | False | True | False | False |
| 8522 | False | False | False | False | False | False | False | False | False | True | False | False |

8523 rows × 12 columns

```
In [10]: df.isnull().sum()
```

| | |
|---------------------------|------|
| Item Fat Content | 0 |
| Item Identifier | 0 |
| Item Type | 0 |
| Outlet Establishment Year | 0 |
| Outlet Identifier | 0 |
| Outlet Location Type | 0 |
| Outlet Size | 0 |
| Outlet Type | 0 |
| Item Visibility | 0 |
| Item Weight | 1463 |
| Sales | 0 |
| Rating | 0 |
| dtype: int64 | |

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

(8523, 12)

```
In [25]: 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 [26]: df.sample(5)
```

| | 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 |
|------|------------------|-----------------|--------------|---------------------------|-------------------|----------------------|-------------|-------------------|-----------------|-------------|----------|--------|
| 160 | Low Fat | FDK41 | Frozen Foods | 2015 | OUT045 | Tier 2 | Medium | Supermarket Type1 | 0.127800 | 14.30 | 83.6224 | 5.0 |
| 3012 | Regular | FDX12 | Baking Goods | 2011 | OUT010 | Tier 3 | High | Grocery Store | 0.043627 | 18.20 | 241.4196 | 4.1 |
| 643 | Regular | FDU51 | Meat | 2017 | OUT035 | Tier 2 | Small | Supermarket Type1 | 0.096495 | 20.20 | 175.6028 | 4.6 |
| 3769 | Low Fat | FDD36 | Baking Goods | 1998 | OUT027 | Tier 3 | Medium | Supermarket Type3 | 0.026175 | NaN | 127.1020 | 4.0 |
| 5971 | Low Fat | FDW28 | Frozen Foods | 2015 | OUT045 | Tier 2 | Small | Supermarket Type1 | 0.089004 | 18.25 | 196.7452 | 4.0 |

```
In [27]: df.index
```

RangeIndex(start=0, stop=8523, step=1)

```
In [28]: len(df)
```

8523

```
In [31]: df.nlargest(5, 'Rating')
```

| | 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 |
|---|------------------|-----------------|-----------------------|---------------------------|-------------------|----------------------|-------------|-------------------|-----------------|-------------|----------|--------|
| 0 | Regular | FDX32 | Fruits and Vegetables | 2012 | OUT049 | Tier 1 | Medium | Supermarket Type1 | 0.100014 | 15.10 | 145.4786 | 5.0 |
| 1 | Low Fat | NCB42 | Health and Hygiene | 2022 | OUT018 | Tier 3 | Medium | Supermarket Type2 | 0.008596 | 11.80 | 115.3492 | 5.0 |
| 2 | Regular | FDR28 | Frozen Foods | 2010 | OUT046 | Tier 1 | Small | Supermarket Type1 | 0.025896 | 13.85 | 165.0210 | 5.0 |
| 3 | Regular | FDL50 | Canned | 2000 | OUT013 | Tier 3 | High | Supermarket Type1 | 0.042278 | 12.15 | 126.5046 | 5.0 |
| 4 | Low Fat | DRD25 | Soft Drinks | 2015 | OUT045 | Tier 2 | Small | Supermarket Type1 | 0.033970 | 19.60 | 55.1614 | 5.0 |

```
In [33]: df.nsmallest(2, 'Sales')
```

| | 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 |
|------|------------------|-----------------|-------------|---------------------------|-------------------|----------------------|-------------|-------------------|-----------------|-------------|-------|--------|
| 5872 | Low Fat | DRK12 | Soft Drinks | 1998 | OUT027 | Tier 3 | Medium | Supermarket Type3 | 0.041683 | NaN | 31.29 | 4.0 |
| 7880 | Low Fat | DRK12 | Soft Drinks | 2000 | OUT013 | Tier 3 | High | Supermarket Type1 | 0.041651 | 9.5 | 31.49 | 4.0 |

```
In [35]: df.empty
```

False

```
In [40]: df.count()
```

| | |
|---------------------------|------|
| Item Fat Content | 8523 |
| Item Identifier | 8523 |
| Item Type | 8523 |
| Outlet Establishment Year | 8523 |
| Outlet Identifier | 8523 |
| Outlet Location Type | 8523 |
| Outlet Size | 8523 |
| Outlet Type | 8523 |
| Item Visibility | 8523 |
| Item Weight | 7060 |
| Sales | 8523 |
| Rating | 8523 |
| dtype: int64 | |

```
In [45]: df['Sales'].median()
```

143.0128

```
In [60]: df.dtypes
```

| | |
|---------------------------|---------|
| 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 | object |
| dtype: object | |

Data Cleaning

```
In [47]: df['Item Fat Content'].unique()
```

Array(['Regular', 'Low fat', 'low fat', 'LF', 'reg'], dtype=object)

```
In [48]: df['Item Fat Content'] = df['Item Fat Content'].replace({'Low fat': "Low Fat",
                                                                'LF': "Low Fat",
                                                                'reg': "Regular"
                                                                })
```

```
In [49]: df['Item Fat Content'].unique()
```

Array(['Regular', 'Low Fat'], dtype=object)

Business Requirements

KPI's Requirement

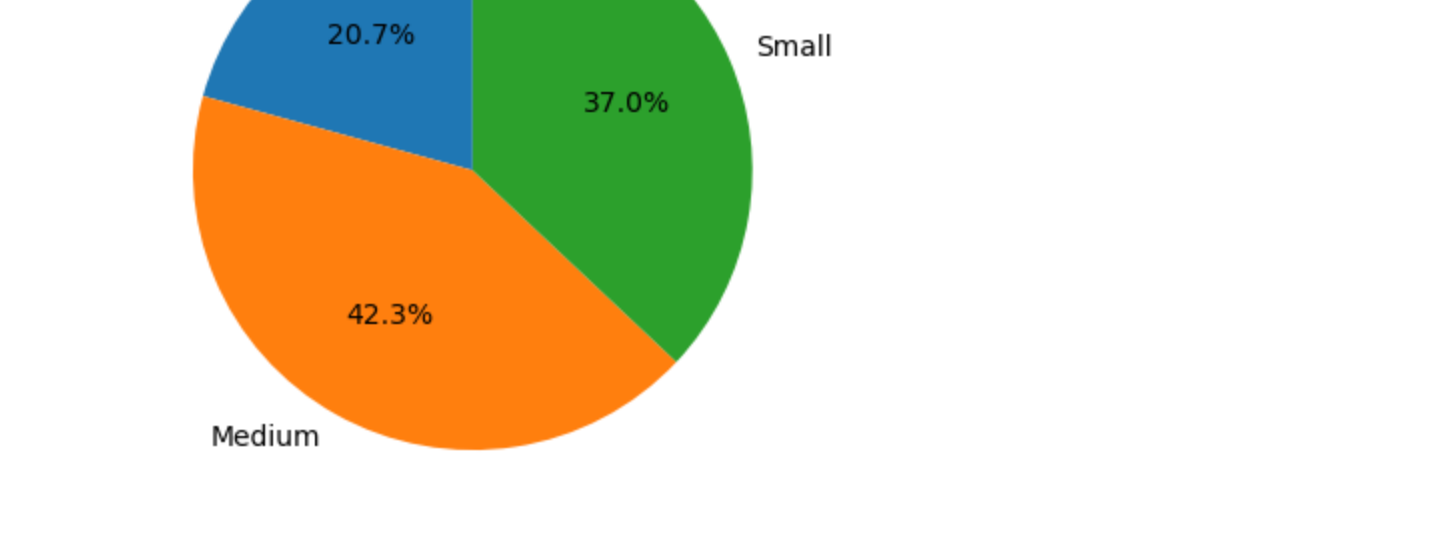
```
In [53]: #Total Sales
total_sales = df['Sales'].sum()
#Average Sales
avg_sales = df['Sales'].mean()
#No of Items Sold
no_of_items_sold = df['Sales'].count()
#Average Ratings
avg_ratings = df['Rating'].mean()

#Display
print(f"Total Sales: {total_sales:,0f}")
print(f"Average Sales: {avg_sales:,.1f}")
print(f"No of Items Sold: {no_of_items_sold:,0f}")
print(f"Average Ratings: {avg_ratings:,.1f}")

Total Sales: 61,201,461
Average Sales: 6141.0
No of Items Sold: 8,523
Average Ratings: 4.0
```

Charts Requirements

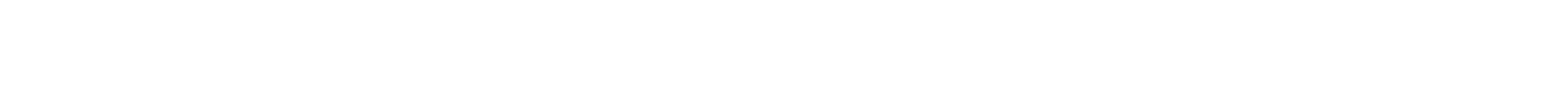
Total Sales by Fat Content



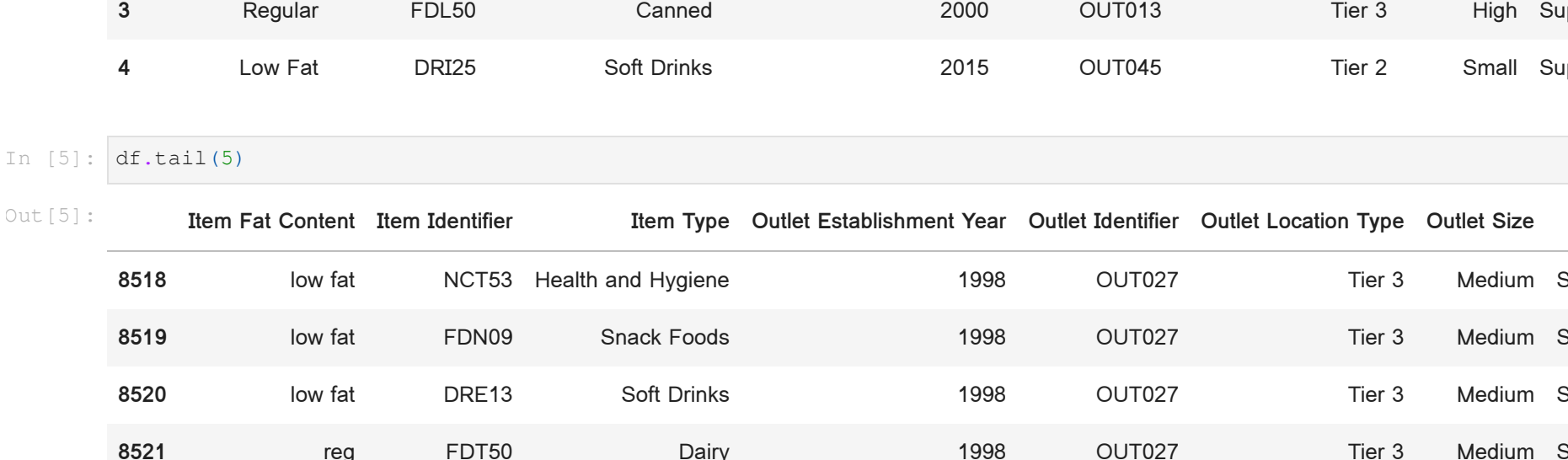
Total Sales by Item Type



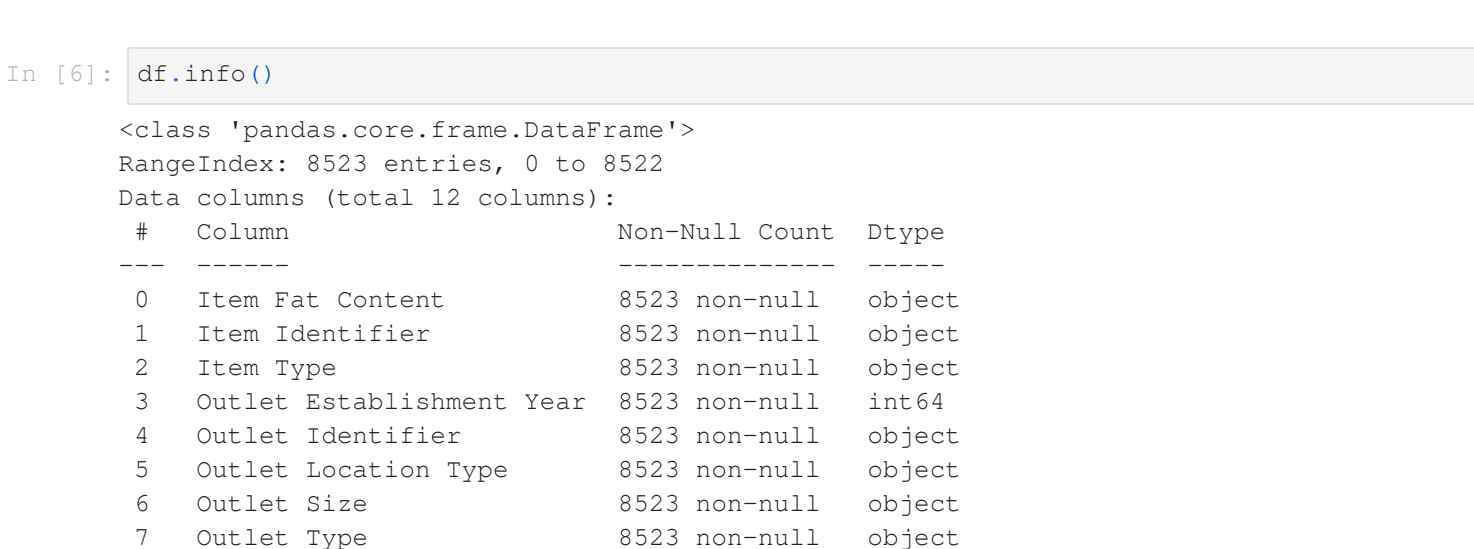
Fat Content by Outlet for Total Sales



Total Sales by Outlet Establishment



Sales by Outlet Size



Sales by Outlet Location

