Blinkit Grocery Analysis Using Pandas

DATASET COLUMN OVERVIEW

- 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

Item Fat C	Item Iden	Item Type	Outlet Est	Outlet Ide	Outlet Loc	Outlet Size	Outlet Typ	Item Visib	Item Weig	Sales	Rating
Regular	FDX32	Fruits and	2012	OUT049	Tier 1	Medium	Supermar	0.100014	15.1	145.4786	5
Low Fat	NCB42	Health an	2022	OUT018	Tier 3	Medium	Supermar	0.008596	11.8	115.3492	5
Regular	FDR28	Frozen Fo	2016	OUT046	Tier 1	Small	Supermar	0.025896	13.85	165.021	5
Regular	FDL50	Canned	2014	OUT013	Tier 3	High	Supermar	0.042278	12.15	126.5046	5
Low Fat	DRI25	Soft Drink	2015	OUT045	Tier 2	Small	Supermar	0.03397	19.6	55.1614	5
low fat	FDS52	Frozen Fo	2020	OUT017	Tier 2	Small	Supermar	0.005505	8.89	102.4016	5
Low Fat	NCU05	Health an	2011	OUT010	Tier 3	Small	Grocery St	0.098312	11.8	81.4618	5
Low Fat	NCD30	Household	2015	OUT045	Tier 2	Small	Supermar	0.026904	19.7	96.0726	5
Low Fat	FDW20	Fruits and	2014	OUT013	Tier 3	High	Supermar	0.024129	20.75	124.173	5
Low Fat	FDX25	Canned	2018	OUT027	Tier 3	Medium	Supermar	0.101562		181.9292	5
LF	FDX21	Snack Foo	2018	OUT027	Tier 3	Medium	Supermar	0.084555		109.8912	5
Low Fat	NCU41	Health an	2017	OUT035	Tier 2	Small	Supermar	0.052045	18.85	192.1846	5
Low Fat	FDL20	Fruits and	2022	OUT018	Tier 3	Medium	Supermar	0.128938	17.1	112.3886	5
Low Fat	NCR54	Household	2014	OUT013	Tier 3	High	Supermar	0.090487	16.35	195.211	5
Low Fat	FDH19	Meat	2018	OUT027	Tier 3	Medium	Supermar	0.032928		173.1738	5
Regular	FDB57	Fruits and	2017	OUT035	Tier 2	Small	Supermar	0.018802	20.25	222.1772	5
Low Fat	FDO23	Breads	2022	OUT018	Tier 3	Medium	Supermar	0.147024	17.85	93.7436	5
Low Fat	NCB07	Household	2012	OUT049	Tier 1	Medium	Supermar	0.077628	19.2	197.611	5
Low Fat	FDJ56	Fruits and	2018	OUT027	Tier 3	Medium	Supermar	0.182515		98.77	5
Low Fat	DRN47	Hard Drin	2022	OUT018	Tier 3	Medium	Supermar	0.016895	12.1	178.566	5
Regular	FDZ07	Fruits and	2018	OUT027	Tier 3	Medium	Supermar	0		60.2194	5

PROBLEM STATEMENT

1.OBJECTIVE:

Analyze grocery sales data to uncover trends in outlet demographics, item performance, and sales, focusing on visibility, fat content, and outlet characteristics to enhance retail efficiency and customer satisfaction.

2.GOAL:

- Examine outlet demographics by location, size, and establishment year.
- Identify top and underperforming item categories.
- Study the impact of Low Fat vs. Regular items on sales.
- Analyze the relationship between visibility and sales.
- Benchmark outlet performance to find growth opportunities.

3.OUTCOME:

- Tier 3 locations and medium-sized outlets are top performers; newer outlets show higher sales.
- Fruits & Vegetables and Snack Foods lead sales; Seafood and Starchy Foods need focus.
- Strong demand for Low Fat items across all outlet types.
- Higher visibility correlates with better sales performance.
- Target underperforming categories and expand in Tier 3 locations for growth.

Data Cleaning and Preparation

Standardized 'Item Fat Content' Categories

- Replaced inconsistent labels (e.g., 'low fat', 'LF') with uniform categories like 'Low Fat' and 'Regular'.

Removed Null Values & Duplicates

- Dropped irrelevant columns (e.g., 'Item Weight' due to missing data).
- Checked and removed duplicate entries for data consistency.

Feature Engineering

- Created new feature, **Year_count** (difference between 2024 and outlet establishment year).
- **sales_by_outlet** to sum sales for each outlet.
- **Avg Rating by Outlet**: Calculates average rating per outlet, retaining original data structure.
- **Avg Sales by Item Type**: Calculates average sales per item type, preserving original data.

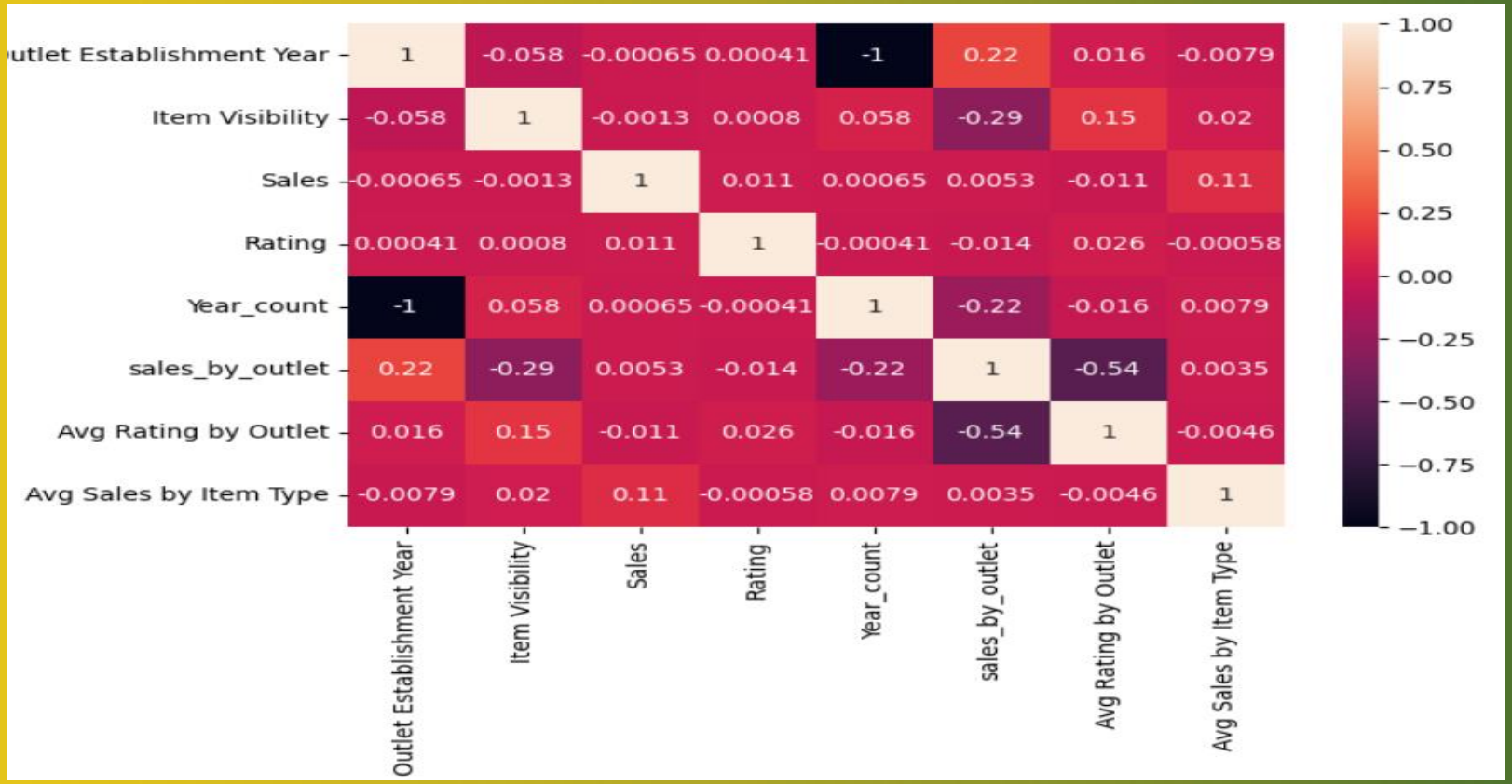
Outlier Detection

- Used **Interquartile Range (IQR)** to identify and flag outliers in 'Sales'.

Analysis

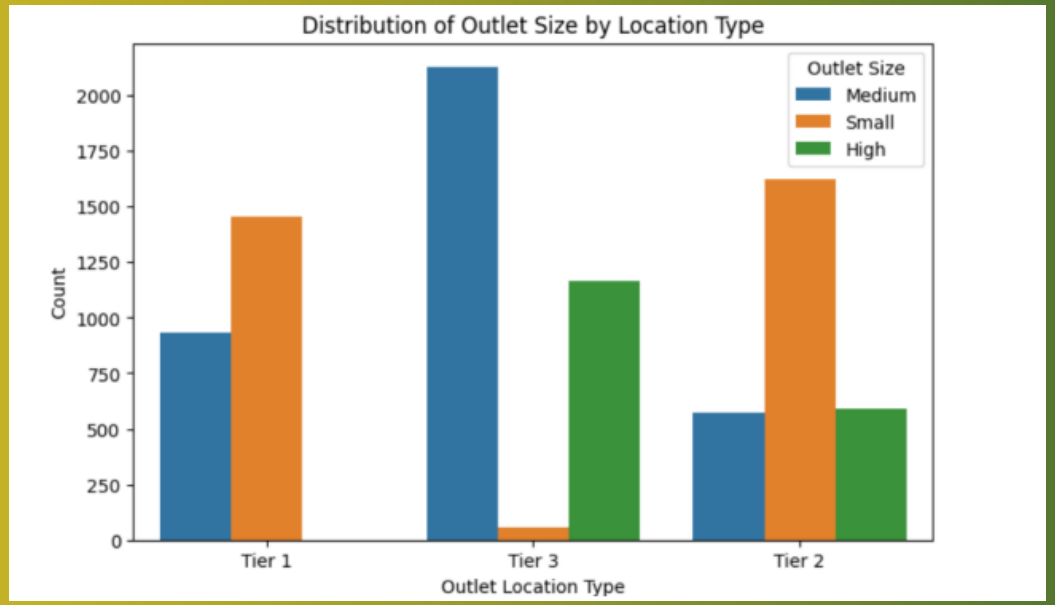
1.Correlation Heatmap

- *Year Count & Outlet Establishment Year(-1): This is because "Year Count" is derived from "Outlet Establishment Year".*
- *Item Visibility shows weak correlations with most variables, suggesting it has minimal direct linear relationships with other numerical features.*
- *As the outlet gets older (Year Count increases), sales tend to slightly decrease, but the weak correlation suggests other factors likely have a stronger impact on sales.*



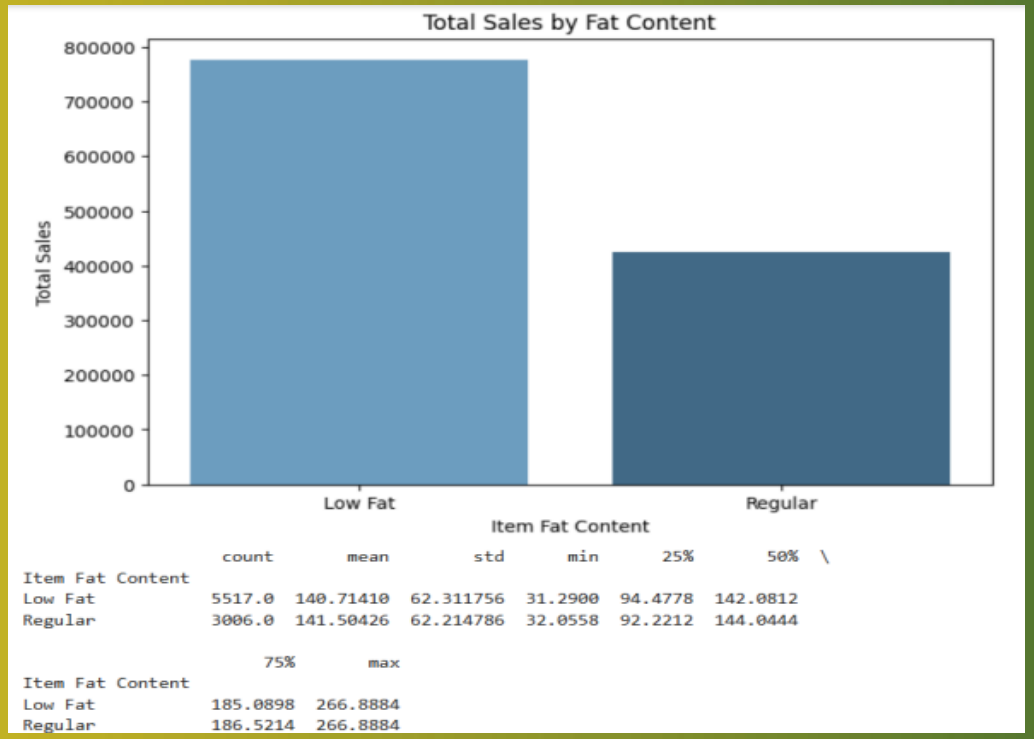
2.Relationship between Outlet Location Type and Outlet Size(Count Plot)

- *Tier 1 has more Small outlets than Medium outlets.*
- *Tier 2 has a more balanced distribution across High, Medium, and Small outlets.*
- *Tier 3 has a higher number of Medium and High outlets, but very few Small outlets.*



3.Impact of Fat Content on Total Sales(Barplot)

- *Total Sales: Low Fat items have higher total sales, indicating greater consumer demand.*
- *Mean Sales: Average sales are similar for both categories (140.71 vs 141.50).*
- *Sales Distribution: Both categories show similar variability with a standard deviation around 62.*
- *Range and Median: Sales distributions are comparable, with slight differences in the interquartile range and median.*

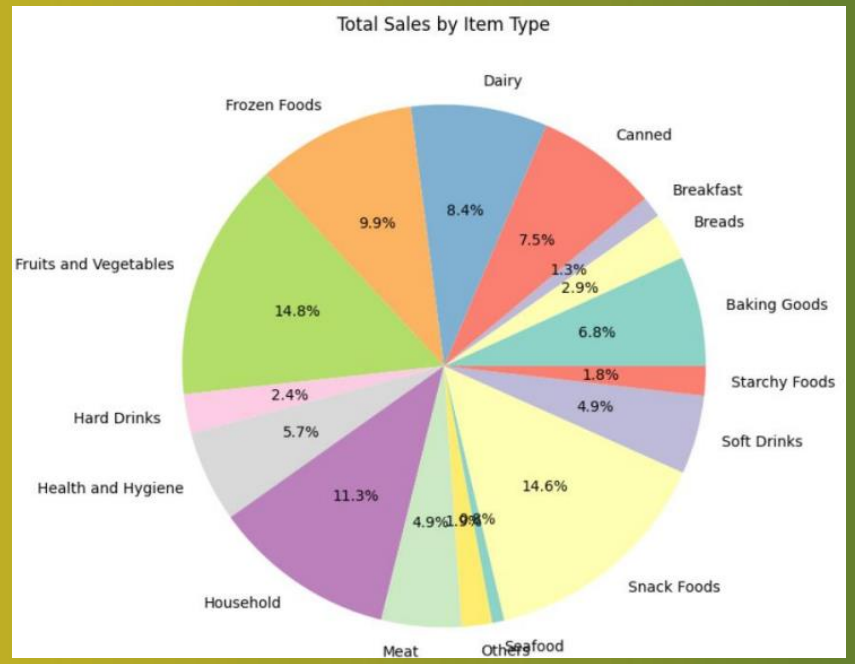


Item Fat Content	count	mean	std	min	25%	50%	\
Low Fat	5517.0	140.71410	62.311756	31.2900	94.4778	142.0812	
Regular	3006.0	141.50426	62.214786	32.0558	92.2212	144.0444	

Item Fat Content	75%	max
Low Fat	185.0898	266.8884
Regular	186.5214	266.8884

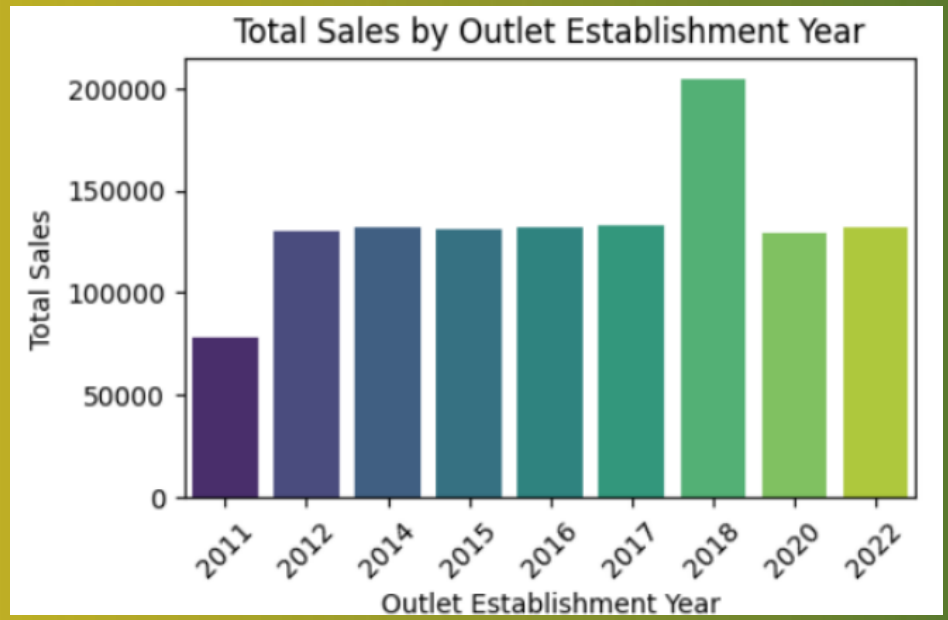
4.Total sales by item type

- *Top Performers: Fruits and Vegetables lead with 178,124.08 in sales, followed by Frozen Foods (118,558.88) and Dairy (101,276.46).*
- *Lowest Sales: Seafood has the least sales at 9,077.87, with Starchy Foods and Others also underperforming.*
- *Popular Categories: Snack Foods (175,433.92), Household (135,976.53), and Baking Goods (81,894.74) show strong demand.*
- *Moderate Performance: Canned, Health and Hygiene, Meat, and Breads have steady but average sales.*
- *Growth Potential:Seafood and Starchy Foods offer opportunities for growth.*



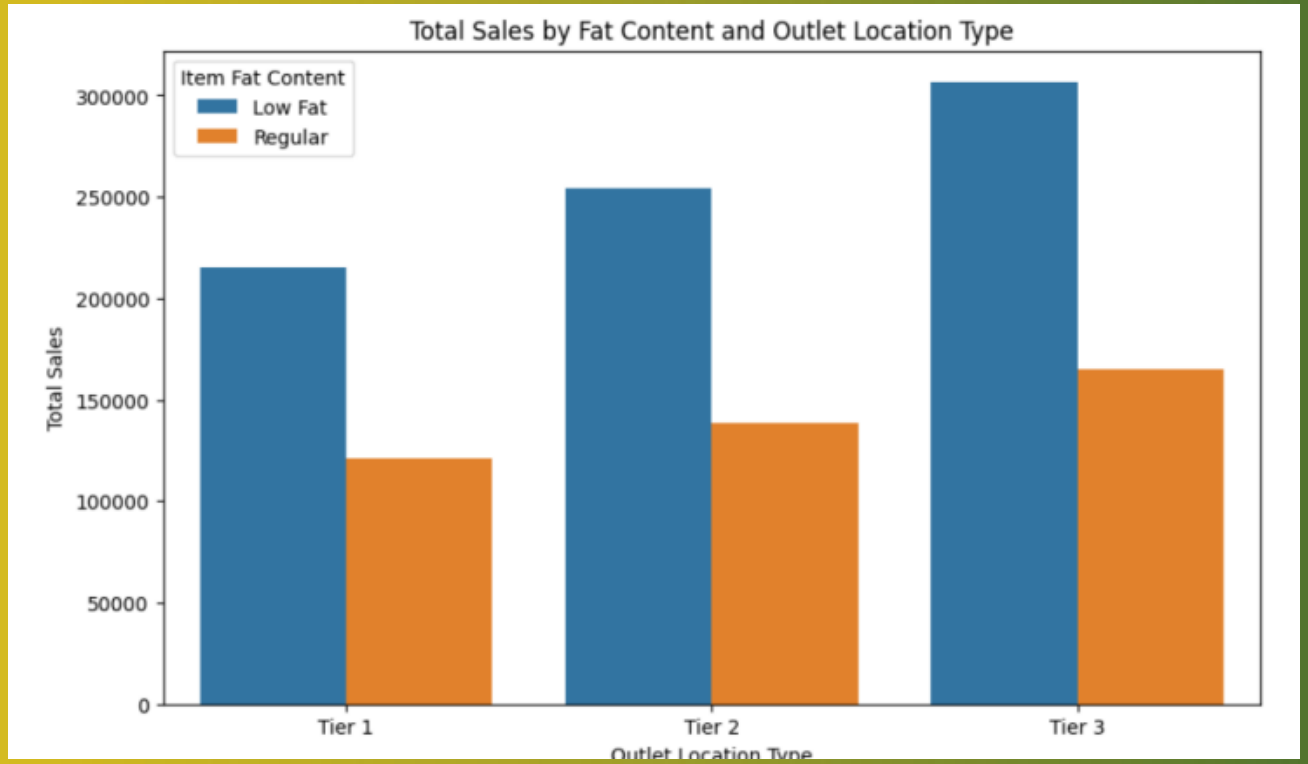
5.Total Sales by Outlet Establishment Year

- *Increasing Sales: Newer outlets, especially from 2017 onward, show higher total sales, with 2018 having the peak sales.*
- *Stable Early Years: Sales for outlets established between 2011-2016 remain stable around 130,000.*
- *Lower Sales in Older Outlets: Outlets from 2011 and 2012 show comparatively lower sales.*
- *Overall, newer outlets tend to have higher sales, with the 2018 outlet leading.*



6.Fat content by outlet for total sales

- *Low Fat items have a significant advantage in terms of total sales across all outlet types.*
- *Tier 3 outlets have the highest sales, likely contributing to better overall performance in the sales data.*



Conclusion

The Blinkit Grocery Data Analysis provides key insights:

1.Newer Tier 3 outlets show higher sales.

2. Fruits, Snacks, and Frozen Foods lead in sales; Seafood and Starchy Foods offer growth opportunities.

3.Low Fat items outperform Regular ones.

4.Supermarket Type 1 outlets in Tier 3 have the highest sales.

5.Need to give more focus on outlet size, location, and item visibility to boost sales.

6.Need to target underperforming categories and expand in high-performing areas.