

Blinkit Grocery Data Analysis Project

Steps taken in the pre-processing of the data

1. Data Set :

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
Regular	FDX32	Fruits and Vegetables	2012	OUT049	Tier 1	Medium	Supermarket Type1	0.1000135	15.1	145.4786	5
Low Fat	NCB42	Health and Hygiene	2022	OUT018	Tier 3	Medium	Supermarket Type2	0.008596051	11.8	115.3492	5
Regular	FDR28	Frozen Foods	2016	OUT046	Tier 1	Small	Supermarket Type1	0.025896485	13.85	165.021	5
Regular	FDL50	Canned	2014	OUT013	Tier 3	High	Supermarket Type1	0.042277867	12.15	126.5046	5
Low Fat	DRI25	Soft Drinks	2015	OUT045	Tier 2	Small	Supermarket Type1	0.033970195	19.6	55.1614	5
low fat	FDS52	Frozen Foods	2020	OUT017	Tier 2	Small	Supermarket Type1	0.005505481	8.89	102.4016	5
Low Fat	NCU05	Health and Hygiene	2011	OUT010	Tier 3	Small	Grocery Store	0.098312421	11.8	81.4618	5
Low Fat	NCD30	Household	2015	OUT045	Tier 2	Small	Supermarket Type1	0.026903714	19.7	96.0726	5
Low Fat	FDW20	Fruits and Vegetables	2014	OUT013	Tier 3	High	Supermarket Type1	0.024129332	20.75	124.173	5
Low Fat	FDX25	Canned	2018	OUT027	Tier 3	Medium	Supermarket Type3	0.101561568		181.9292	5
LF	FDX21	Snack Foods	2018	OUT027	Tier 3	Medium	Supermarket Type3	0.084554569		109.8912	5

2. Reading the data:

Loaded the dataset, applied filters on columns, and checked each column along with its data types.

3. Handling missing values in Item Weight:

- Found more than 1400 missing values in the *Item Weight* column.
- Filled these missing values using the corresponding values mapped from the *Item Identifier* column.
- After this step, 4 missing values still remained in *Item Weight*.
- Filled those remaining values using the mean *Item Weight* of records grouped by the corresponding *Item Fat Content* and *Item Type* category.

4. Preprocessing code:

Documented the Python code used for the above preprocessing steps.

5. Data type adjustments:

Converted columns to appropriate data types and standardized the decimal places in numeric columns.

Data analysis with SQL

1. Creating database and table in MySQL : Created a database blinkit and blinkit_data table.

2. Data analysis according to problem statement and Query used :

KPIs :

- Total Sales: The overall revenue generated from all items sold.

Query : `Select concat(Round(Sum(Sales)/1000,2)," K") As `Overall Revenue in Thousands`
from blinkit_data`

	Overall Revenue in Thousands
▶	1201.68 K

- Average Sales: The average revenue per sale.

Query : `Select Round(Avg(Sales),1) as `Average Revenue per Sale` from blinkit_data`

	Average Revenue per Sale
▶	141

- The total count of items sold.

Query :

`SELECT COUNT(*) AS Total_No_of_Items_Sold
FROM blinkit_data;`

	Total_No_of_Items_Sold
▶	8523

- Average Rating: The average customer rating for items sold.

Query :

`SELECT CAST(AVG(Rating) AS DECIMAL(10,1)) AS Avg_Rating
FROM blinkit_data;`

	Overall_Avg_Rating
▶	4.0

1. Total Sales by Fat Content:

Objective: Analyse the impact of fat content on total sales.

Query :

`Select `Item Fat Content`, Round(sum(Sales),2) as Total_Sales
from blinkit_data
group by `Item Fat Content`
order by Total_sales Desc;`

	Item Fat Content	Total_Sales
▶	Low Fat	776319.69
	Regular	425361.8

2. Total Sales by Item Type:

Objective: Identify the performance of different item types in terms of total sales and no. of Items Sold from each category.

Query :

```
Select `Item Type`, Round(SUM(Sales),2) AS Total_Sales, Count(*) As `Total Different items sold` from blinkit_data
group by `Item Type`
order by Total_Sales Desc;
```

	Item Type	Total_Sales	Total Different items sold
►	Fruits and Vegetables	178124.08	1232
	Snack Foods	175433.92	1200
	Household	135976.53	910
	Frozen Foods	118558.88	856
	Dairy	101276.46	682
	Canned	90706.73	649
	Baking Goods	81894.74	648
	Health and Hygiene	68025.84	520
	Meat	59449.86	425
	Soft Drinks	58514.17	445
	Breads	35379.12	251
	Hard Drinks	29334.68	214
	Others	22451.89	169
	Starchy Foods	21880.03	148
	Breakfast	15596.7	110
	Seafood	9077.87	64

3. Fat Content by Outlet for Total Sales:

Objective: Compare total sales across different outlets segmented by fat content.

Query :

```
SELECT `Outlet Location Type`,
SUM(CASE WHEN `Item Fat Content` = 'Low Fat' THEN Sales ELSE 0 END) AS Low_Fat,
SUM(CASE WHEN `Item Fat Content` = 'Regular' THEN Sales ELSE 0 END) AS Regular
FROM blinkit_data
GROUP BY `Outlet Location Type`
ORDER BY `Outlet Location Type`;
```

	Outlet Location Type	Low_Fat	Regular
►	Tier 1	215047.9126000001	121349.89940000001
	Tier 2	254464.77940000014	138685.86819999994
	Tier 3	306806.9964000001	165326.0368

4. Total Sales by Outlet Establishment:

Objective: Evaluate how the age or type of outlet establishment influences total sales.

Query :

```
Select `Outlet Establishment Year`, Round(Sum(Sales)) as Total_sales from
blinkit_data
group by `Outlet Establishment Year`
order by `Outlet Establishment Year`;
```

	Outlet Establishment Year	Total_sales
▶	2011	78132
	2012	130477
	2014	131809
	2015	130943
	2016	132113
	2017	133104
	2018	204522
	2020	129104
	2022	131478

5. Percentage of Sales by Outlet Size:

Objective: Analyse the correlation between outlet size and total sales.

Query :

```
select `Outlet Size`, Round(sum(Sales)) As Total_sales,
Round(sum(Sales) * 100/sum(sum(Sales)) over(),1) As Sales_Percentage
from blinkit_data
group by `Outlet Size`
order by Total_Sales desc;
```

	Outlet Size	Total_sales	Sales_Percentage
▶	Medium	507896	42.3
	Small	444794	37
	High	248992	20.7

6. Sales by Outlet Location:

Objective: Assess the geographic distribution of sales across distinct locations.

Query :

```
SELECT `Outlet Location Type`, Round(SUM(Sales)) AS Total_Sales
FROM blinkit_data
GROUP BY `Outlet Location Type` ORDER BY Total_Sales DESC ;
```

	Outlet Location Type	Total_Sales
▶	Tier 3	472133
	Tier 2	393151
	Tier 1	336398

7. All Metrics by Outlet Type:

Objective: Provide a comprehensive view of all key metrics (Total Sales, Average Sales, Number of Items, Average Rating) broken down by different outlet types.

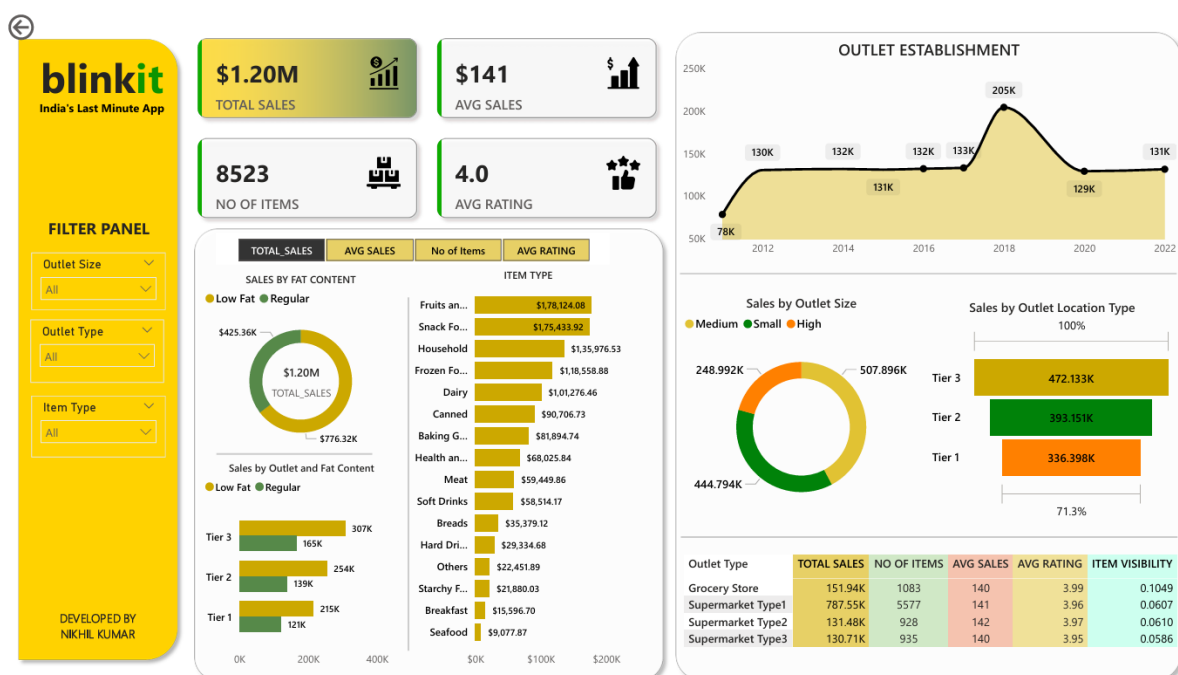
Query :

```
SELECT `Outlet Type`, Round(Sum(Sales)) As Total_Sales,
Round(AVG(Sales)) as Avg_Sales,
Count(*) As No_of_Items,
Round(Avg(Rating),2) as Avg_Rating
FROM blinkit_data
group by `Outlet Type`
order by Total_Sales;
```

	Outlet Type	Total_Sales	Avg_Sales	No_of_Items	Avg_Rating
▶	Supermarket Type3	130715	140	935	3.95
	Supermarket Type2	131478	142	928	3.97
	Grocery Store	151939	140	1083	3.99
	Supermarket Type1	787550	141	5577	3.96

Created Dashboard using Power BI

1. Imported data from MySQL database and then transformed the datatype of numeric columns.
2. Created some measurements columns and Metrics.



Analysis & Insights from Power BI Dashboard :

Executive Summary :

The dashboard reveals a total sales revenue of **\$1.20 million** from 8,523 items, with an average sale value of \$141 and a strong average customer rating of 4.0. The business shows significant recent growth, a clear customer preference for low-fat products, and a heavy reliance on its "Supermarket Type1" outlets. The top-performing product categories are Fruits & Vegetables and Snack Foods.

1. Sales Performance Over Time

- **Recent Growth Surge:** There was a dramatic increase in sales in **2022, reaching \$205K**. This is a significant jump from the relatively stable period between 2014 and 2021, where annual sales hovered around \$130K.
- **Historical Performance:** The business saw a dip in its second year of operation (2013) but quickly recovered and maintained consistent performance for eight years before the recent surge.

Insight : The key takeaway is to investigate the drivers behind the **massive sales growth in 2022**. This could be due to successful marketing campaigns, expansion, or changes in consumer behaviour that the company should identify and replicate.

2. Top Product & Customer Insights

- **Key Categories: Fruits & Vegetables (\$178K)** and **Snack Foods (\$175K)** are the highest-grossing categories, followed by Household items (\$136K). These three categories are the primary revenue drivers.
- **Customer Preference for Low-Fat:** Low-fat products account for **\$776K (65%)** of sales compared to \$425K from regular-fat products. This strong preference is consistent across all store locations and types.

Insight: Marketing and inventory should be heavily focused on the top three categories. The clear preference for **low-fat options** indicates a health-conscious customer base, which can be leveraged in product placement and promotions.

3. Outlet & Location Performance

- **Dominant Outlet Type: Supermarket Type1** is the star performer, generating **\$788K in sales**, which is more than five times the revenue of the next closest outlet type (Grocery Store at \$152K).
- **Best Performing Size: Medium (\$508K)** and **Small (\$472K)** sized outlets generate most of the revenue. "High" size outlets are the lowest performers (\$249K).
- **Top Location Tier:** Outlets in **Tier 3 locations (\$445K)** generate the most sales, followed by Tier 2 (\$393K) and Tier 1 (\$336K).

Insight: The business should analyse the success factors of **Supermarket Type1**—be it location, product mix, or management—and apply these lessons to other outlet types. The data suggests that focusing expansion on **Medium and Small outlets in Tier 3 locations** would likely yield the best return on investment.

X

NIKHIL KUMAR

