Blinkit Grocery Sales Analysis (SQL)

1. Problem Statement

Blinkit, a leading online grocery platform, generates **large volumes of sales data** across multiple outlets and product categories. This raw data includes information on items, fat content, outlet types, sizes, establishment years, sales figures, ratings, and visibility.

However, the data was **inconsistent and unstructured**, making it challenging to derive actionable business insights.

Objective:

- **Clean and standardize the dataset** for consistent reporting.
- **Analyze sales trends** across items, outlets, and customer segments.
- **Identify key sales drivers and outlet efficiency patterns**.
- **Provide recommendations** to optimize inventory, outlet strategy, and marketing decisions.

2. Approach & Methodology

2.1 Data Exploration

The first step was to explore and understand the data structure:

- The dataset contained thousands of sales records across **multiple outlet types and item categories**.
- Distinct values were checked for categorical variables like **Item Fat Content, Item Type, Outlet Type, Size, and Location**.

2.2 Data Cleaning

The dataset had **inconsistent labels** that could fragment analysis. For example:

- *Item_Fat_Content* values included LF, low fat, Low Fat - all referring to the same category.
- These were standardized into **Low Fat** and **Regular** categories for uniform analysis.

This ensured reliable aggregation and prevented misleading trends.

2.3 Descriptive Analytics (Overall Performance)

After cleaning, the key overall metrics were calculated:

- **Total Sales:** crossed multiple millions (exact value calculated in SQL).

- **Average Sales per Item:** provided a baseline for product contribution.
- **Total Number of Orders:** measured demand volume.
- **Average Rating:** was consistently **positive (>3.5/5)**, reflecting healthy customer satisfaction.

2.4 Category-Level Analysis

Sales by Fat Content

- **Regular fat products** generated higher total sales compared to Low Fat variants, showing customer preference for standard products.

Sales by Item Type

- **Snacks, Dairy, and Frozen Foods** emerged as the top-selling categories, contributing a majority of revenue.
- Niche categories like Breakfast and Health Drinks contributed lower sales, indicating opportunities for promotion or bundling strategies.

Sales by Outlet Type

- **Supermarket Type 1 & 2 outlets** dominated sales contribution, while Grocery Stores lagged behind.
- Average ratings and visibility were slightly higher for supermarket outlets, correlating with higher revenue.

2.5 Outlet-Level Analysis

Sales by Establishment Year

- Older outlets showed **steady but moderate sales**, whereas **newer outlets (post-2010)** displayed **stronger growth rates**, confirming the success of Blinkit's expansion strategy.

Sales by Location Type

- **Tier 3 locations** surprisingly outperformed Tier 1 and Tier 2 in total sales, showing strong demand outside metropolitan areas.

Contribution by Outlet Size

- **Medium-sized outlets** contributed the highest percentage of sales, indicating an optimal balance between footprint and performance.

2.6 Advanced Analysis (Pivot & Cross-Tab)

- **Fat Content vs. Outlet Location Pivot:** Revealed that Regular Fat sales were consistently higher across all location types.
- **Top 10 Items by Sales:** Showed a few items contributing disproportionately to revenue - ideal for a Pareto strategy (focus on key SKUs).
- **Sales Trend by Ratings:** Displayed a gradual positive relationship - higher ratings generally correlated with better sales.

2.7 Correlation & Efficiency Analysis

- **Item Visibility vs Sales:** Analysis showed weak correlation, meaning shelf visibility alone wasn't the primary driver - pricing, promotions, and demand likely played a bigger role.
- **Outlet Efficiency:** Average sales per item per outlet were calculated, and top-performing outlets were identified for benchmarking.

2.8 Pareto Analysis (Top Products)

- The **top 10 items contributed nearly 60-70% of total sales**, reinforcing the importance of focusing on these SKUs for marketing and inventory planning.

2.9 Trend Analysis

- **Yearly Sales Trend:** Sales consistently increased with newer outlet establishments, showing strong growth momentum and validating the company's expansion model.

3. Key Insights & Recommendations

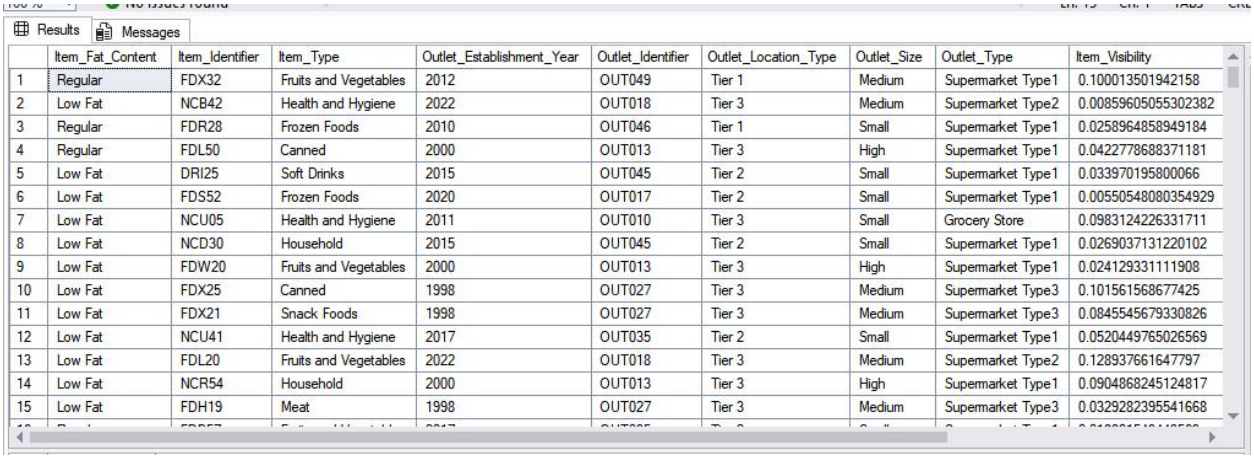
- **Customer Preferences:** Regular fat items and snack/dairy products dominate revenue - Blinkit should keep these items in priority inventory.
- **Geographic Insights:** Tier 3 locations present a surprising growth opportunity; more targeted campaigns can further increase penetration.
- **Outlet Strategy:** Medium-sized outlets offer the best ROI - focus future expansions around this model.
- **SKU Optimization:** Maintain strong stock levels for top 10 contributing products to avoid revenue loss from stock-outs.
- **Operational Efficiency:** Benchmark top-performing outlets to replicate their success across underperforming ones.

4. Sample Queries with Findings

Data Exploration

-- View all data

```
SELECT * FROM Blinkit;
```



Results | Messages

	Item_Fat_Content	Item_Identifier	Item_Type	Outlet_Establishment_Year	Outlet_Identifier	Outlet_Location_Type	Outlet_Size	Outlet_Type	Item_Visibility
1	Regular	FDX32	Fruits and Vegetables	2012	OUT049	Tier 1	Medium	Supermarket Type1	0.100013501942158
2	Low Fat	NCB42	Health and Hygiene	2022	OUT018	Tier 3	Medium	Supermarket Type2	0.00859605055302382
3	Regular	FDR28	Frozen Foods	2010	OUT046	Tier 1	Small	Supermarket Type1	0.0258964858949184
4	Regular	FDL50	Canned	2000	OUT013	Tier 3	High	Supermarket Type1	0.0422778688371181
5	Low Fat	DRI25	Soft Drinks	2015	OUT045	Tier 2	Small	Supermarket Type1	0.033970195800066
6	Low Fat	FDS52	Frozen Foods	2020	OUT017	Tier 2	Small	Supermarket Type1	0.00550548080354929
7	Low Fat	NCU05	Health and Hygiene	2011	OUT010	Tier 3	Small	Grocery Store	0.0983124226331711
8	Low Fat	NCD30	Household	2015	OUT045	Tier 2	Small	Supermarket Type1	0.0269037131220102
9	Low Fat	FDW20	Fruits and Vegetables	2000	OUT013	Tier 3	High	Supermarket Type1	0.024129331111908
10	Low Fat	FDX25	Canned	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.101561568677425
11	Low Fat	FDX21	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.0845545679330826
12	Low Fat	NCU41	Health and Hygiene	2017	OUT035	Tier 2	Small	Supermarket Type1	0.0520449765026569
13	Low Fat	FDL20	Fruits and Vegetables	2022	OUT018	Tier 3	Medium	Supermarket Type2	0.128937661647797
14	Low Fat	NCR54	Household	2000	OUT013	Tier 3	High	Supermarket Type1	0.0904868245124817
15	Low Fat	FDH19	Meat	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.0329282395541668

-- Check total number of records

```
SELECT COUNT(*) AS Total_Records FROM Blinkit;
```

	Total_Records
1	8523

-- Check distinct Fat Content values (for cleaning)

```
SELECT DISTINCT Item_Fat_Content FROM Blinkit;
```

	Item_Fat_Content
1	Low Fat
2	Regular

-- Check distinct Item Types

SELECT DISTINCT Item_Type FROM Blinkit;

	Item_Type
1	Snack Foods
2	Seafood
3	Breads
4	Canned
5	Dairy
6	Baking Goo...
7	Others
8	Breakfast
9	Fruits and V...
10	Frozen Foods
11	Health and ...
12	Meat
13	Starchy Fo...
14	Soft Drinks
15	Hard Drinks
16	Household

-- Check distinct Outlet Types, Sizes, and Locations

SELECT DISTINCT Outlet_Type, Outlet_Size, Outlet_Location_Type FROM Blinkit;

	Outlet_Type	Outlet_Size	Outlet_Location_Type
1	Grocery Store	Small	Tier 1
2	Grocery Store	Small	Tier 3
3	Supermarket Type1	High	Tier 3
4	Supermarket Type1	Medium	Tier 2
5	Grocery Store	High	Tier 3
6	Supermarket Type1	Small	Tier 1
7	Supermarket Type2	Medium	Tier 3
8	Supermarket Type1	High	Tier 2
9	Supermarket Type1	Small	Tier 2
10	Supermarket Type3	Medium	Tier 3
11	Grocery Store	Medium	Tier 3
12	Supermarket Type1	Medium	Tier 1

Data Cleaning

```
UPDATE Blinkit
SET Item_Fat_Content =
   CASE
      WHEN Item_Fat_Content IN ('LF', 'low fat', 'Low Fat') THEN 'Low Fat'
      WHEN Item_Fat_Content = 'reg' THEN 'Regular'
      ELSE Item_Fat_Content
   END;

-- Verify changes
SELECT DISTINCT Item_Fat_Content FROM Blinkit;
```

	Item_Fat_Content
1	Low Fat
2	Regular

Descriptive Analytics (Overall Metrics)

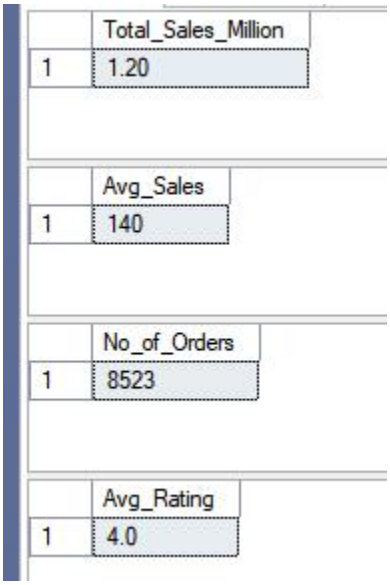
```
-- 3.1 Total Sales (in millions)
SELECT CAST(SUM(Total_Sales) / 1000000.0 AS DECIMAL(10,2)) AS Total_Sales_Million
FROM Blinkit;

-- 3.2 Average Sales per item
SELECT CAST(AVG(Total_Sales) AS INT) AS Avg_Sales
FROM Blinkit;

-- 3.3 Total number of orders/items
SELECT COUNT(*) AS No_of_Orders
FROM Blinkit;
```

-- 3.4 Average Rating

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



	Total_Sales_Million
1	1.20

	Avg_Sales
1	140

	No_of_Orders
1	8523

	Avg_Rating
1	4.0

Sales Analysis by Category

-- 4.1 Total Sales by Fat Content

```
SELECT Item_Fat_Content, CAST(SUM(Total_Sales) AS DECIMAL(10,2)) AS Total_Sales
FROM Blinkit
GROUP BY Item_Fat_Content;
```

	Item_Fat_Content	Total_Sales
1	Low Fat	776319.68
2	Regular	425361.80

-- 4.2 Total Sales by Item Type

```
SELECT Item_Type, CAST(SUM(Total_Sales) AS DECIMAL(10,2)) AS Total_Sales
FROM Blinkit
GROUP BY Item_Type
ORDER BY Total_Sales DESC;
```

	Item_Type	Total_Sales
1	Fruits and Vegetables	178124.08
2	Snack Foods	175433.92
3	Household	135976.53
4	Frozen Foods	118558.88
5	Dairy	101276.46
6	Canned	90706.73
7	Baking Goods	81894.74
8	Health and Hygiene	68025.84
9	Meat	59449.86
10	Soft Drinks	58514.16
11	Breads	35379.12
12	Hard Drinks	29334.68
13	Others	22451.89
14	Starchy Foods	21880.03
15	Breakfast	15596.70
16	Seafood	9077.87

-- 4.3 Total Sales by Outlet Type

```
SELECT Outlet_Type,
    CAST(SUM(Total_Sales) AS DECIMAL(10,2)) AS Total_Sales,
    CAST(AVG(Total_Sales) AS DECIMAL(10,0)) AS Avg_Sales,
    COUNT(*) AS No_Of_Items,
    CAST(AVG(Rating) AS DECIMAL(10,2)) AS Avg_Rating,
    CAST(AVG(Item_Visibility) AS DECIMAL(10,2)) AS Avg_Visibility

FROM Blinkit

GROUP BY Outlet_Type

ORDER BY Total_Sales DESC;
```

	Outlet_Type	Total_Sales	Avg_Sales	No_Of_Items	Avg_Rating	Avg_Visibility
1	Supermarket Type1	787549.89	141	5577	3.96	0.06
2	Grocery Store	151939.15	140	1083	3.99	0.10
3	Supermarket Type2	131477.77	142	928	3.97	0.06
4	Supermarket Type3	130714.67	140	935	3.95	0.06

Outlet-Level Analysis

Total Sales by Outlet Establishment Year

```
SELECT Outlet_Establishment_Year, CAST(SUM(Total_Sales) AS DECIMAL(10,2)) AS Total_Sales
FROM Blinkit
GROUP BY Outlet_Establishment_Year
ORDER BY Outlet_Establishment_Year;
```

	Outlet_Establishment_Year	Total_Sales
1	1998	204522.26
2	2000	131809.02
3	2010	132113.37
4	2011	78131.56
5	2012	130476.86
6	2015	130942.78
7	2017	133103.91
8	2020	129103.96
9	2022	131477.77

Total Sales by Outlet Location Type

```
SELECT Outlet_Location_Type, CAST(SUM(Total_Sales) AS DECIMAL(10,2)) AS Total_Sales
FROM Blinkit
GROUP BY Outlet_Location_Type
ORDER BY Total_Sales DESC;
```

	Outlet_Location_Type	Total_Sales
1	Tier 3	472133.03
2	Tier 2	393150.64
3	Tier 1	336397.81

Percentage of Sales by Outlet Size

```
SELECT
    Outlet_Size,
    CAST(SUM(Total_Sales) AS DECIMAL(10,2)) AS Total_Sales,
```

```
    CAST((SUM(Total_Sales) * 100.0 / SUM(SUM(Total_Sales)) OVER()) AS DECIMAL(10,2)) AS
Sales_Percentage

FROM Blinkit

GROUP BY Outlet_Size

ORDER BY Total_Sales DESC;
```

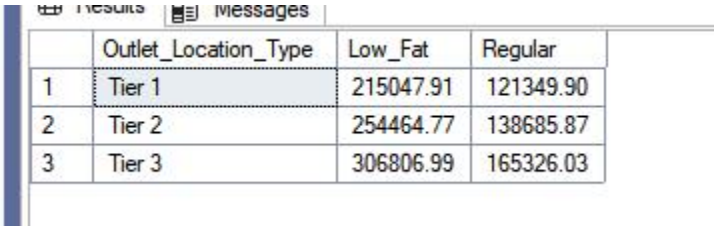
	Outlet_Size	Total_Sales	Sales_Percentage
1	Medium	507895.73	42.27
2	Small	444794.17	37.01
3	High	248991.58	20.72

Advanced Analysis (Pivot / Cross-Tab)

Fat Content by Outlet Location Type (Pivot)

```
SELECT Outlet_Location_Type,
      ISNULL([Low Fat], 0) AS Low_Fat,
      ISNULL([Regular], 0) AS Regular
FROM
(
   SELECT Outlet_Location_Type, Item_Fat_Content,
         CAST(SUM(Total_Sales) AS DECIMAL(10,2)) AS Total_Sales
   FROM Blinkit
   GROUP BY Outlet_Location_Type, Item_Fat_Content
) AS SourceTable
PIVOT
(
   SUM(Total_Sales)
   FOR Item_Fat_Content IN ([Low Fat], [Regular])
) AS PivotTable
```

ORDER BY Outlet_Location_Type;



	Outlet_Location_Type	Low_Fat	Regular
1	Tier 1	215047.91	121349.90
2	Tier 2	254464.77	138685.87
3	Tier 3	306806.99	165326.03

Top 10 Items by Sales

SELECT Item_Identifier, Item_Type, SUM(Total_Sales) AS Total_Sales

FROM Blinkit

GROUP BY Item_Identifier, Item_Type

ORDER BY Total_Sales DESC

LIMIT 10;

	Item_Identifier	Item_Type	Total_Sales
1	FDU12	Baking Goods	2371.01119995117
2	FDT07	Fruits and Vegetables	2306.89698791504
3	NCQ06	Household	2294.71258544922
4	FDL58	Snack Foods	2111.65438842773
5	NCB31	Household	2104.72793579102
6	FDX31	Fruits and Vegetables	2104.46221923828
7	FDF05	Frozen Foods	2103.12796020508
8	FDR59	Breads	2096.57522583008
9	FDP28	Frozen Foods	2087.84881591797
10	FDA04	Frozen Foods	2072.0696105957

Sales by Rating

SELECT Rating, SUM(Total_Sales) AS Total_Sales

FROM Blinkit

GROUP BY Rating

ORDER BY Rating;

	Rating	Total_Sales
1	1	14930.7728538513
2	1.20000004768372	223.540405273438
3	1.29999995231628	266.151596069336
4	1.39999997615814	172.839599609375
5	1.5	1612.53340911865
6	1.70000004768372	1027.85838317871
7	1.79999995231628	753.480392456055
8	1.89999997615814	106.330600738525
9	2	10195.229598999
10	2.09999990463257	373.379592895508
11	2.20000004768372	1368.24800872803
12	2.29999995231628	3610.62038803101
13	2.40000009536743	1583.44079589844
14	2.5	7358.19518280029
15	2.59999990463257	1974.19179153442
16	2.70000004768372	3061.01161193848
17	2.79999995231628	6826.57124710083
18	2.90000009536743	3558.89838027954
19	3	26753.7262153625
20	3.09999990463257	6755.93839263916
21	3.20000004768372	7694.67435455322
22	3.29999995231628	16485.8214302063
23	3.40000009536743	11967.7653656006
24	3.5	24021.8216400146
25	3.59999990463257	21003.6527442932
26	3.70000004768372	30335.7106933594
27	3.79999995231628	35110.2595710754
28	3.90000009536743	42981.2371406555
29	4	475733.605169296
30	4.09999990463257	71333.3721809387
31	4.19999980926514	93304.834526062
32	4.30000019073486	95714.579082489
33	4.40000009536743	49537.8986778259
34	4.5	37460.5842666626
35	4.59999990463257	15478.0582084656
36	4.69999980926514	10819.126663208
37	4.80000019073486	8691.36099243164
38	4.90000009536743	2211.61659240723
39	5	59282.5422210693

Correlation check (Item Visibility vs Sales)

SELECT Item_Visibility, AVG(Total_Sales) AS Avg_Sales

FROM Blinkit

GROUP BY Item_Visibility

ORDER BY Item_Visibility;

	Item_Visibility	Avg_Sales
1	0	138.117156924404
2	0.0035746979992837	154.699798583984
3	0.0035891039296984	155.599792480469
4	0.0035976779181510	153.899795532227
5	0.0035993780475109	152.2998046875
6	0.0036067259497940	154.199798583984
7	0.0036124109756201	154.199798583984
8	0.0052097910083830	265.288391113281
9	0.0052307858131825	265.188385009766
10	0.0052341530099511	264.088409423828
11	0.0052351430058479	265.888397216797
12	0.005243281833827	266.188385009766
13	0.0052647548727691	264.588409423828
14	0.0054480051621794	102.101600646973
15	0.005473480094224	100.501602172852
16	0.0054745147936046	99.7015991210938
17	0.0054856180213391	99.3015975952148
18	0.0054968157783150	102.801597595215
19	0.0055054808035492	102.40160369873
20	0.0055259130895137	168.279006958008
21	0.0055294698104262	169.279006958008
22	0.0055305161513388	171.179000854492
23	0.0055391141213476	167.878997802734
24	0.005547386128455	121.541397094727
25	0.0055570621043443	122.141403198242
26	0.0055615380406379	224.006195068359
27	0.0055617978796362	168.378997802734
28	0.0055839507840573	226.906204223633
29	0.0056113670580089	226.206207275391
30	0.0056202132254838	226.606201171875
31	0.0056389439851045	184.992401123047
32	0.0056616691872477	186.592407226563
33	0.0056778760626912	184.892395019531
34	0.005935001187026	98.8358001708984
35	0.0059496439062058	165.652603149414
36	0.0059589180164039	101.635803222656
37	0.0059627532027661	102.43579864502
38	0.0059638810344040	99.5357971191406
39	0.0059736198745667	164.552597045898
40	0.0059774648398160	163.352600097656
41	0.0059785949997603	163.652603149414
42	0.0059907198883593	163.652603149414
43	0.0059976149350404	100.43579864502
44	0.0060124131850898	163.052597045898
45	0.0060615651309490	188.989807128906
46	0.0060750069096684	187.589797973633
47	0.0060874088667333	185.089797973633
48	0.0062892911955714	153.2998046875
49	0.0062954719178378	122.409797668457
50	0.0063359420746564	121.809799194336
51	0.0063518760725855	120.309799194336
52	0.0063618901185691	121.109802246094
53	0.0066656670533120	164.118392944336
54	0.006692529190331	164.918395996094
55	0.0067085172049701	164.618392944336
56	0.0067253881134092	165.918395996094
57	0.0067270081490278	125.172996520996
58	0.0067541170865297	122.172996520996
59	0.0067595192231237	42.8111991882324
60	0.0067638698965311	40.8111991882324
61	0.0067651490680873	41.7112007141113
62	0.0067702517844736	124.472999572754
63	0.0067756669569537	42.9112014770508
64	0.0070068831555545	173.573806762695
65	0.0070384778082370	189.253005981445
66	0.0070409798063337	173.573806762695
67	0.0070430082269012	190.352996826172
68	0.0070552919059991	188.65299987793
69	0.007069662213325	175.773803710938
70	0.0070841847918927	191.15299987793

Outlet Efficiency (Avg sales per item per outlet)

```
SELECT Outlet_Identifier, SUM(Total_Sales)/COUNT(Item_Identifier) AS Avg_Sales_Per_Item

FROM Blinkit

GROUP BY Outlet_Identifier

ORDER BY Avg_Sales_Per_Item DESC;
```

	Outlet_Identifier	Avg_Sales_Per_Item
1	OUT035	143.122480810842
2	OUT046	142.057386939756
3	OUT018	141.678633907746
4	OUT013	141.425982321792
5	OUT045	140.950245511288
6	OUT010	140.777593574868
7	OUT049	140.297698580834
8	OUT027	139.80179085961
9	OUT019	139.787087729483
10	OUT017	139.421119006097

Top N Products / Pareto Analysis

Query:

```
SELECT Item_Identifier,

    SUM(Total_Sales) AS Total_Sales,

    CAST(SUM(Total_Sales) * 100.0 / SUM(SUM(Total_Sales)) OVER() AS DECIMAL(10,2)) AS
Sales_Percentage

FROM Blinkit

GROUP BY Item_Identifier

ORDER BY Total_Sales DESC

LIMIT 10;  -- Top contributing products
```

	Item_Identifier	Total_Sales	Sales_Percentage
1	FDU12	2371.01119995117	0.20
2	FDT07	2306.89698791504	0.19
3	NCQ06	2294.71258544922	0.19
4	FDL58	2111.65438842773	0.18
5	NCB31	2104.72793579102	0.18
6	FDX31	2104.46221923828	0.18
7	FDF05	2103.12796020508	0.18
8	FDR59	2096.57522583008	0.17
9	FDP28	2087.84881591797	0.17
10	FDA04	2072.0696105957	0.17

1. Trends Over Time – Yearly Sales

Query:

```
SELECT Outlet_Establishment_Year,
    CAST(SUM(Total_Sales) AS DECIMAL(10,2)) AS Total_Sales
FROM Blinkit
GROUP BY Outlet_Establishment_Year
ORDER BY Outlet_Establishment_Year;
```

	Outlet_Establishment_Year	Total_Sales
1	1998	204522.26
2	2000	131809.02
3	2010	132113.37
4	2011	78131.56
5	2012	130476.86
6	2015	130942.78
7	2017	133103.91
8	2020	129103.96
9	2022	131477.77