Blinkit Ecommerce Data Analysis Project Report

PREPARED BY:- VINAY RAYKAR

INDEX

1. Introduction

This report presents a comprehensive analysis of the Blinkit ecommerce data aimed at understanding sales trends, customer behavior, delivery performance, inventory management, and marketing effectiveness. The goal is to generate actionable insights that help optimize business operations and enhance customer satisfaction. The analysis leveraged **Python (Pandas, Matplotlib, Seaborn)** for data cleaning, visualization, and advanced analytics.

2. Data Description

Multiple Blinkit datasets were utilized encompassing order details, customer profiles, delivery performance records, product inventories, marketing campaign metrics, and customer feedback. Key identifiers such as order ID, product ID, and customer ID were used to link datasets, enabling a holistic view of the Blinkit ecommerce ecosystem from product availability to customer experience.

3. Exploratory Analysis & Key Findings

1. **Customer Segmentation Distribution**

   This bar chart classifies customers based on their purchasing behavior using RFM scoring. It highlights the proportion of valuable customers, allowing targeted marketing approaches.

1. Who are Blinkit's most valuable customers?

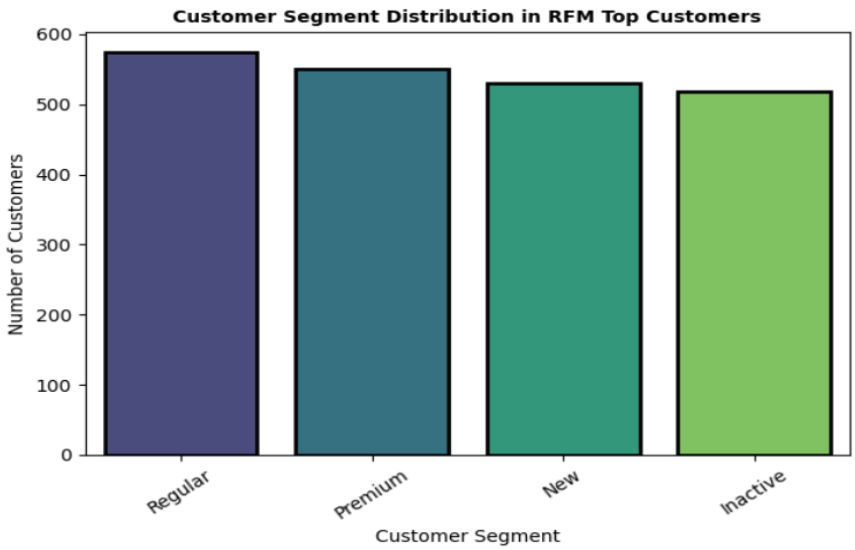
Segment customers based on frequency, monetary value, and recency (RFM analysis). Identify which city or segment is the most profitable.

[69]:
```
# Reference date = one day after latest order
reference_date = orders['order_date'].max() + pd.Timedelta(days=1)

# Group by customer and calculate RFM metrics
rfm = orders.groupby('customer_id').agg({
    'order_date': lambda x: (reference_date - x.max()).days,  # Recency
    'order_id': 'count',                                      # Frequency
    'order_total': 'sum'                                      # Monetary
}).reset_index()

rfm.columns = ['customer_id', 'Recency', 'Frequency', 'Monetary']

# Score each R, F, M from 1 (low) to 4 (high)
rfm['R_Score'] = pd.qcut(rfm['Recency'], 4, labels=[4, 3, 2, 1]).astype(int)
rfm['F_Score'] = pd.qcut(rfm['Frequency'].rank(method='first'), 4, labels=[1, 2, 3, 4]).astype(int)
rfm['M_Score'] = pd.qcut(rfm['Monetary'], 4, labels=[1, 2, 3, 4]).astype(int)

# Combine scores
rfm['RFM_Score'] = rfm[['R_Score', 'F_Score', 'M_Score']].sum(axis = 1)
rfm['RFM_Segment'] = rfm['R_Score'].astype(str) + rfm['F_Score'].astype(str) + rfm['M_Score'].astype(str)

# Merge with customer data
rfm = rfm.merge(customers[['customer_id', 'area', 'customer_segment']], on='customer_id', how='left')
```

[70]:
```
rfm.head()
```

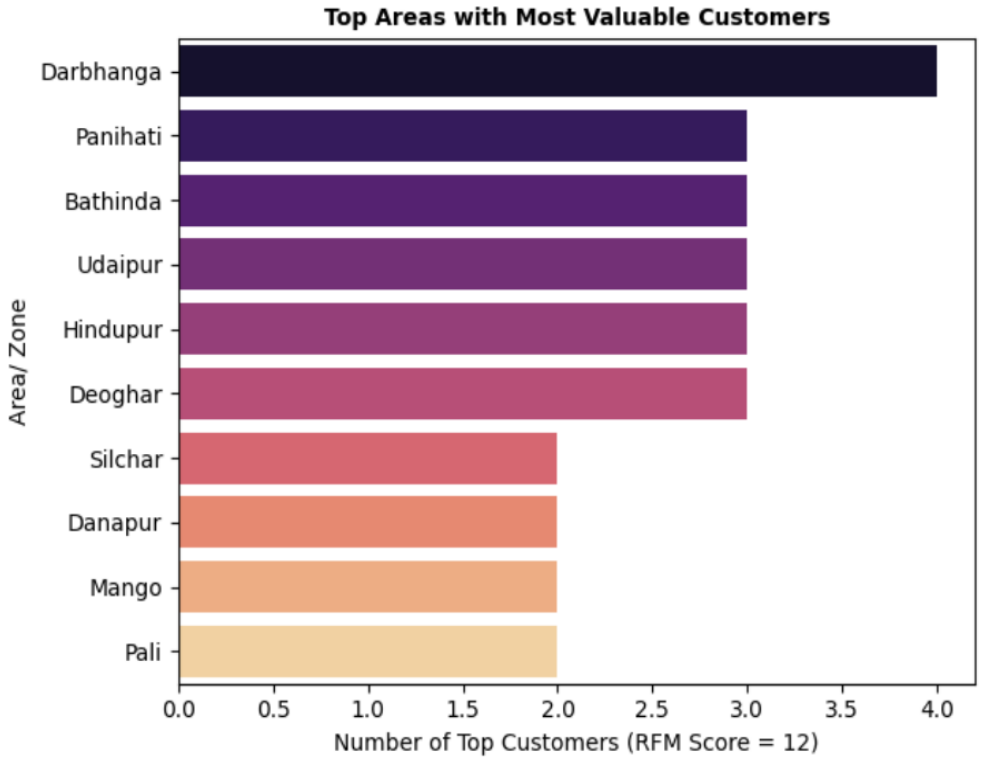
[70]:

	customer_id	Recency	Frequency	Monetary	R_Score	F_Score	M_Score	RFM_Score	RFM_Segment	area	customer_segment
0	31813	26	2	5726.04	4	2	3	9	423	Sultan Pur Majra	Regular
1	61020	66	3	7844.90	4	3	4	11	434	Kurnool	New
2	119099	165	4	14768.77	3	4	4	11	344	Kavali	Inactive
3	188838	260	2	5182.04	2	2	3	7	223	Madanapalle	Inactive
4	191616	165	2	4089.96	3	2	2	7	322	Pallavaram	Regular



2. **Top Geographic Areas with Valuable Customers**

   A horizontal bar chart visualizing the cities and areas with the highest concentration of loyal, high-value customers, helping prioritize regional strategies.



3. **Delivery Performance: On-Time vs Delayed Deliveries by City/Partner**

   This stacked bar chart compares delivery timeliness across cities and delivery partners, identifying areas for improving on-time delivery rates.

```
Total Percentage of deliveries are delayed by 49.39 %
```

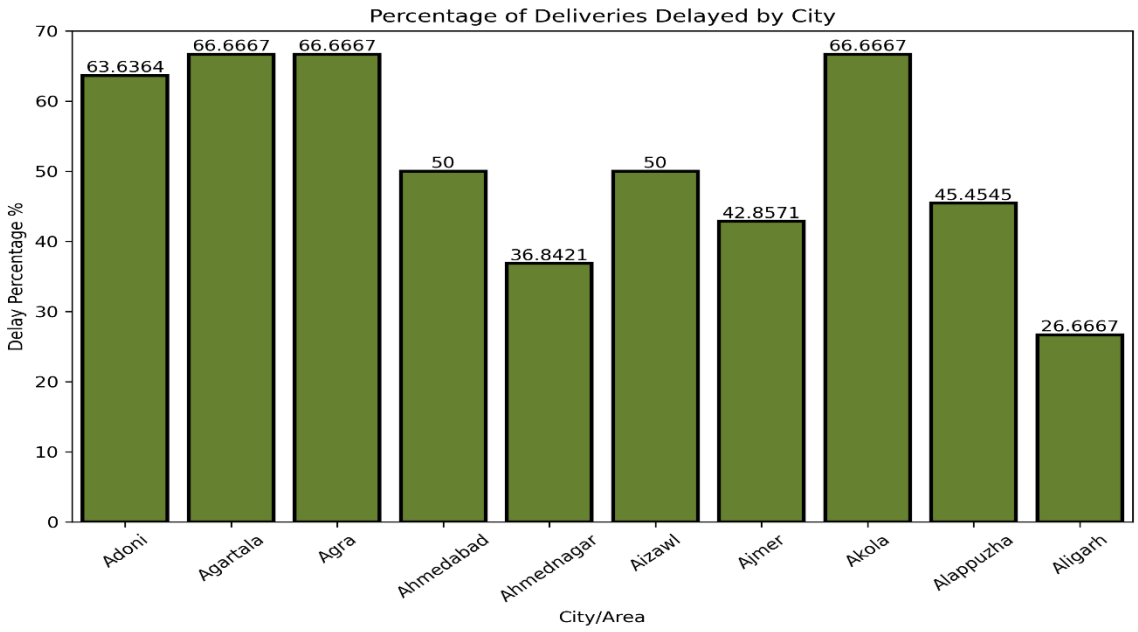
```
[77]: print("Total Percentage of deliveries delayed of city is")
      round(pct_city,2).head()
```

```
Total Percentage of deliveries delayed of city is
```

[77]:

	area	Delay%
0	Adoni	63.64
1	Agartala	66.67
2	Agra	66.67
3	Ahmedabad	50.00
4	Ahmednagar	36.84

```
Total Percentage of deliveries delayed by Delivery Man:
   delivery_partner_id  Delay%
3                  102   100.0
0                   43     0.0
1                   66     0.0
2                   70     0.0
4                  158     0.0
```



4. **Sentiment Analysis of Customer Feedback**

   This chart categorizes customer feedback sentiments, highlighting negative sentiments linked to delivery delays and product issues, guiding service enhancements.

[101]:

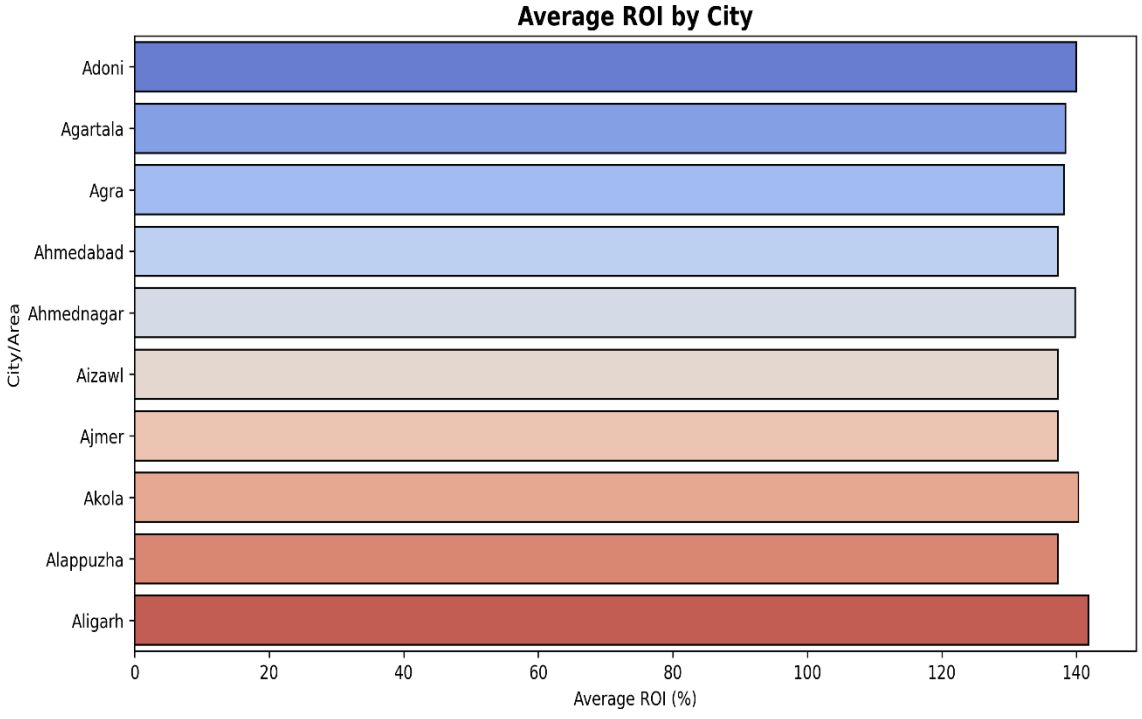
	sentiment	product_name	delivery_time_minutes	feedback_text
7	Negative	Toothpaste	13	I had a bad experience.
5	Negative	Popcorn	13	Items were missing from my order.
6	Negative	Vitamins	12	Customer service was not helpful.
8	Negative	Toilet Cleaner	12	The order was incorrect.
4	Negative	Cough Syrup	8	Not worth the price I paid.
3	Negative	Lotion	3	Product was damaged during delivery.
2	Negative	Bread	2	I had a bad experience.
0	Negative	Orange Juice	2	The order was incorrect.
9	Negative	Baby Wipes	2	Product was damaged during delivery.
1	Negative	Nuts	1	Product was damaged during delivery.

5. **Marketing Campaign ROI by Campaign**

   A bar chart presenting the return on ad spend for various marketing campaigns, helping to identify the most effective channels.

[34]:

	area	ROI (%)
0	Adoni	139.975
1	Agartala	138.379
2	Agra	138.151
3	Ahmedabad	137.243
4	Ahmednagar	139.844
5	Aizawl	137.243
6	Ajmer	137.243
7	Akola	140.279
8	Alappuzha	137.243
9	Aligarh	141.807

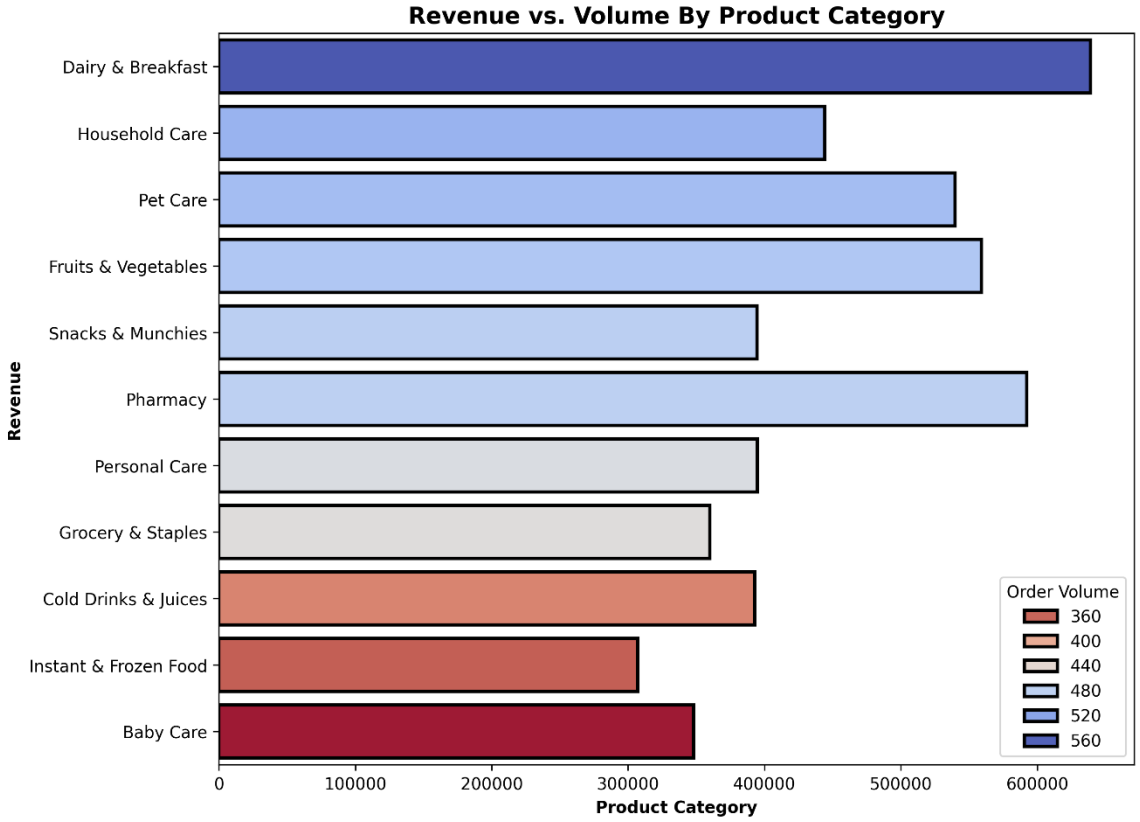


6. **Top Product Categories by Sales Volume and Revenue**

   Bar chart summarizing sales volume and revenue across product categories to ascertain high-performing segment

[48]:

	category	Volume	Revenue
0	Dairy & Breakfast	566	639222.19
1	Household Care	509	444244.25
2	Pet Care	501	539888.75
3	Fruits & Vegetables	492	559053.08
4	Snacks & Munchies	483	394648.71
5	Pharmacy	481	592368.57
6	Personal Care	454	394894.61
7	Grocery & Staples	449	359937.82
8	Cold Drinks & Juices	375	392717.62
9	Instant & Frozen Food	356	307212.65
10	Baby Care	334	348227.18

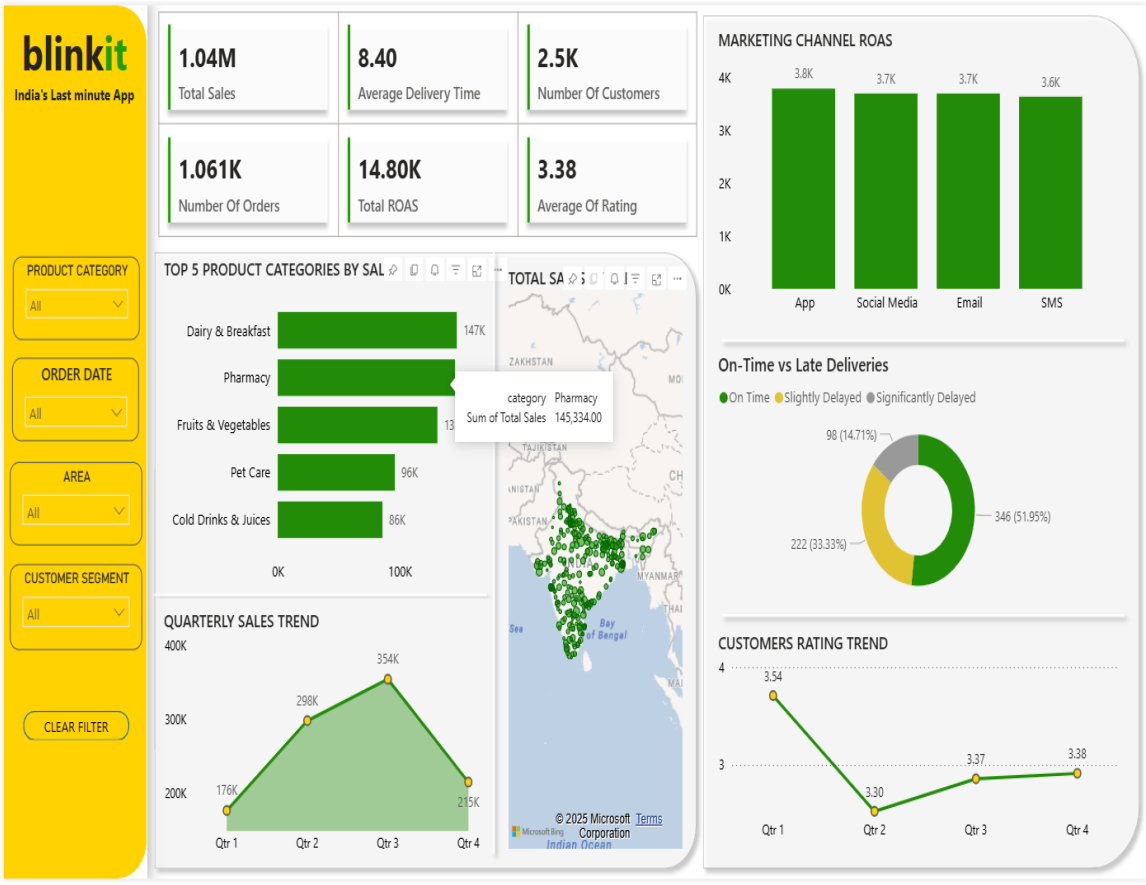


4. Methodology

Data preparation involved extensive cleaning, type conversions, and integration of multiple datasets. Analytical techniques included RFM analysis, correlation studies, and visualization techniques using Python libraries to uncover patterns and trends.

5. Dashboard Overview

This visualization tool supports real-time business decision making.



6. Conclusion

The Blinkit ecommerce data analysis established that optimizing delivery logistics and inventory management, coupled with focused marketing strategies tailored to high-value customers, can substantially improve profitability and customer experience. Continued monitoring of feedback and performance metrics is recommended for sustained growth.