Olist E-Commerce: End-to-End Business Performance Review

SALES PERFORMANCE, DELIVERY PERFORMANCE, OPERATION EFFECIENCY

FRANCIS AUNG THU

Contents

1. Project Overview

This project analyses an e-commerce business using Python and an interactive Power BI dashboard. It focuses on three areas: **Sales Performance**, **Delivery Performance**, and **Operational Efficiency**.

Sales analysis highlights when customers buy, which categories drive revenue, and how behaviour changes over time. Cohort and RFM analysis show that most customers do not return after their first purchase, while a small group of high-value customers generates most revenue.

Delivery performance varies widely across regions. Metrics such as delivery gap, on-time rate, SLA success, and delivery cost show that late deliveries strongly affect customer satisfaction and retention.

Operational insights reveal peak-hour order spikes, high-cost cities, and fulfilment delays that impact delivery reliability. These findings point to opportunities to improve staffing, routing, and inventory placement.

The analysis was conducted in Python and visualised through an interactive Power BI dashboard. The Power BI dashboard brings all insights together, enabling stakeholders to explore sales trends, customer behaviour, delivery KPIs, and geographic performance interactively.

Overall, the project provides clear, data-driven recommendations to grow revenue, improve loyalty, and strengthen operational efficiency.

2. Dataset Summary

The Olist dataset contains multiple linked tables covering orders, customers, items, products, sellers, payments, and geolocation. Together, they represent the full customer journey—from order placement to final delivery—allowing a complete analysis of sales, customer behaviour, delivery reliability, and operations.

3. Methodology

The project followed a structured analytics workflow, using Python for data preparation, exploratory analysis, and feature engineering, and Power BI for interactive visualisation:

- **Business Understanding**: Defined key questions around sales performance, customer behaviour, delivery reliability, and operational efficiency.
- **Data Preparation (Python)**: Standardised data types, handled missing values, removed duplicates, and integrated multiple tables into a unified dataset.
- **Exploratory Data Analysis (Python)**: Identified trends, distributions, and anomalies across sales, delivery, and customer activity.
- **Feature Engineering (Python)**: Created delivery gap, fulfilment time, RFM scores, price segments, and profitability metrics.
- **Analytical Frameworks**: Built KPIs, segmentation models, and cohort structures to evaluate retention, spending behaviour, and operational performance.
- **Dashboard Development (Power BI)**: Designed an interactive dashboard with drill-downs, geographic views, and performance summaries.
- **Insight Synthesis**: Consolidated findings into clear business recommendations to support commercial and operational decision-making.

4. Tools & Technologies

- **Python** — used for data cleaning, exploratory analysis, feature engineering, and building analytical frameworks.
- **Power BI** — used for dashboard design, KPI visualisation, and geographic performance analysis.

5. Business Goals and Key Insights

1. Increase Sales Performance

Goal: Understand demand patterns, customer value, and product performance.

Key Insights:

- Sales peak during specific hours, days, and seasons.

- A few product categories generate most revenue (Pareto effect).

- Customer retention drops quickly; high-value customers drive most revenue.

- Revenue varies by day, time, and city.

- Customers show strong price sensitivity, especially in lower-priced segments.

2. Enhance Delivery Performance

Goal: Improve SLA compliance, delivery speed, and customer satisfaction.

Key Insights:

- Delivery cost and speed vary significantly by region.

- Many orders arrive late, reducing satisfaction and retention.

- On-time delivery rates differ widely across cities.

- Fulfilment delays occur at multiple stages, especially during peak periods.

- High-volume cities often have higher delivery costs due to routing inefficiencies.

3. Improve Operational Efficiency

Goal: Identify bottlenecks and reduce fulfilment and delivery delays.
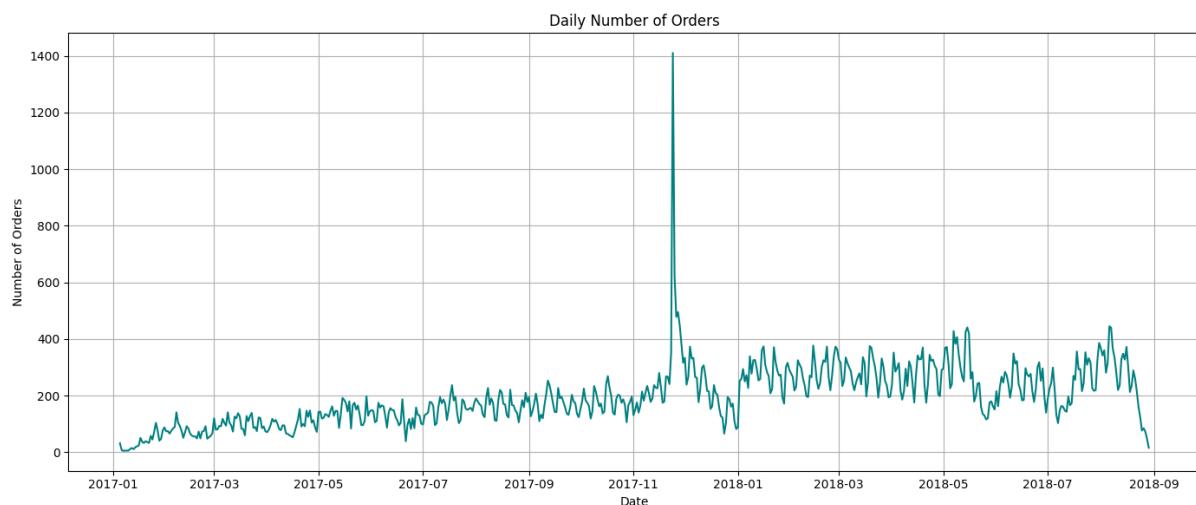
Key Insights:

- Strong hourly order spikes create pressure on fulfilment teams.

- Some cities consistently show higher delays and delivery costs.

- SLA performance varies across top cities.

- Delays occur in approval, carrier pickup, and last-mile delivery.

- Geographic demand is uneven, with some high-volume regions underserved operationally.

6. Key Insights
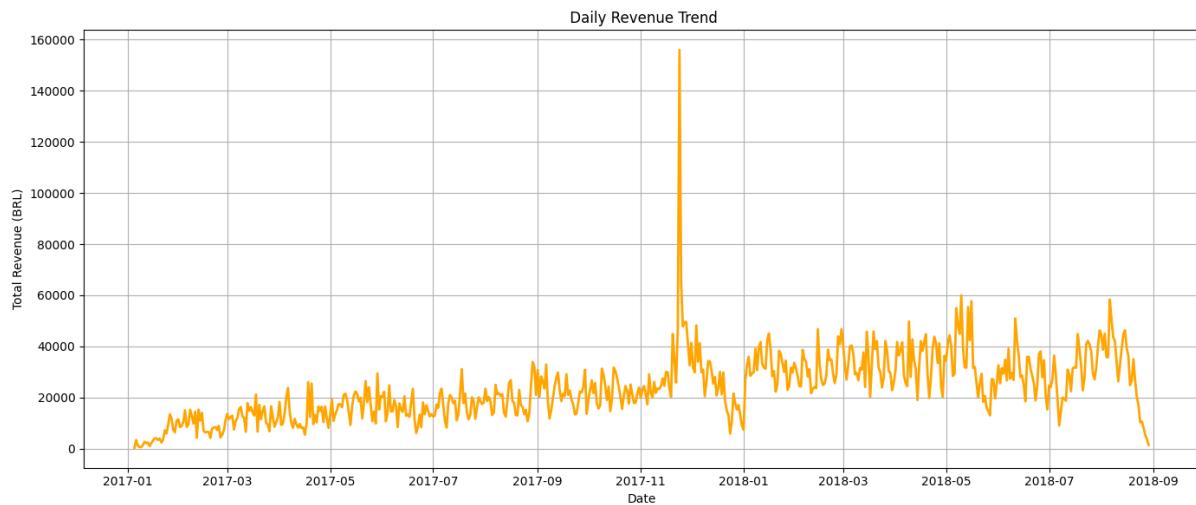
6.1 Sales Performance Analysis

Q1: The number of orders trend

- A significant spike in December 2017. Replicating the drivers behind this peak could inform future growth strategies.

- The uplift in daily order volume during 2018 suggests operational scalability and growing market traction.
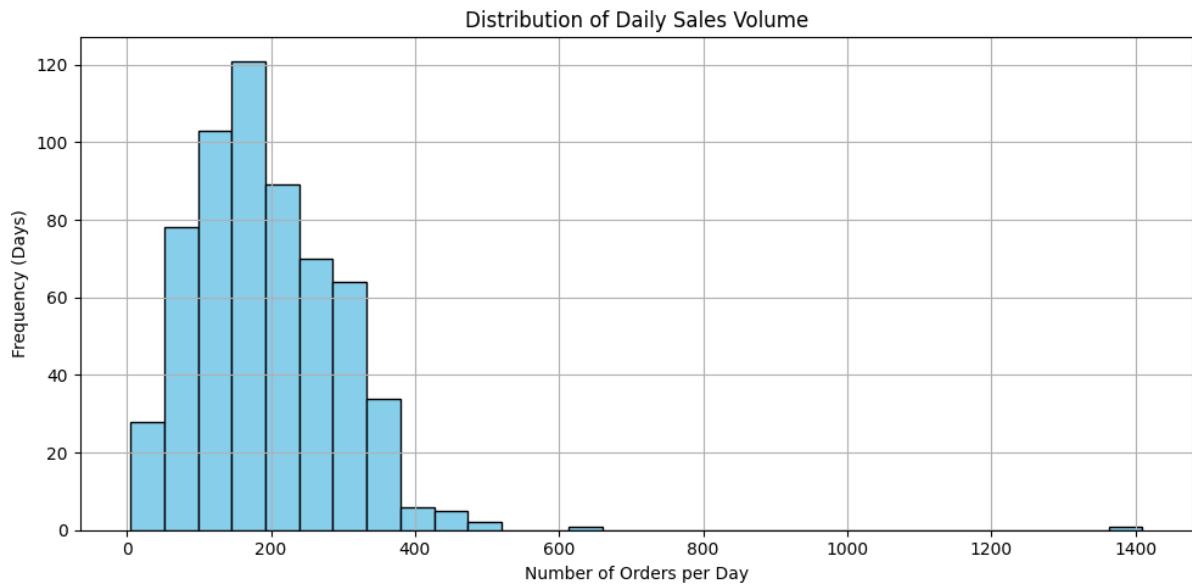
Q2: Daily Revenue Trend

Q3: Average Revenue per Order
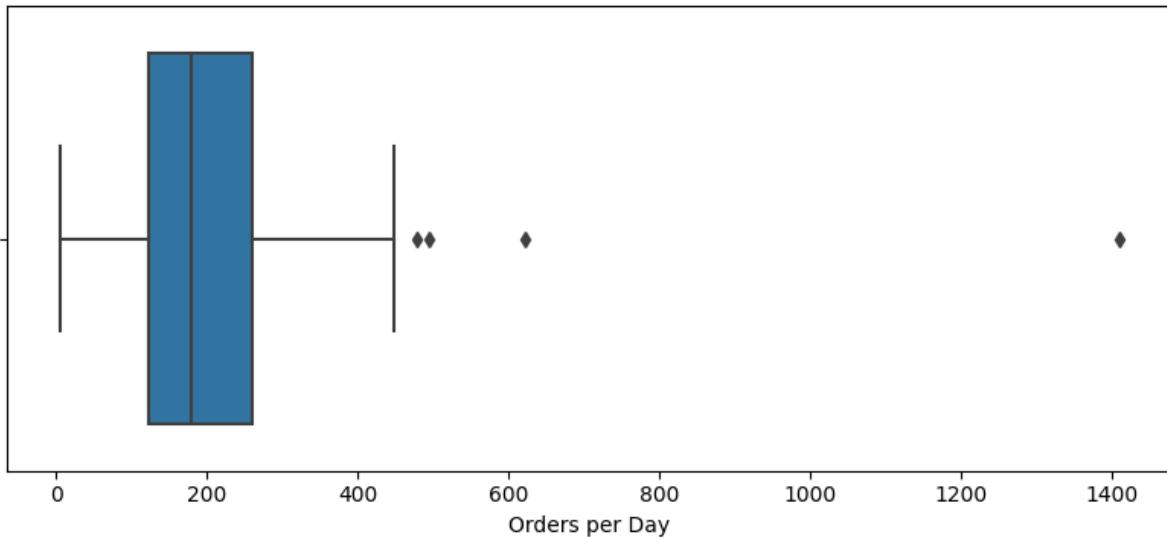
Average price per order: 143.95 BRL

Q4: Average Order per Day

Daily order volume is concentrated between 100–300 orders, with a strong peak around 200. The distribution is right-skewed, indicating occasional high-volume spikes that may reflect promotions, seasonal demand, or campaign effects.

Boxplot of Daily Sales Volume

Q5: Average Order per city
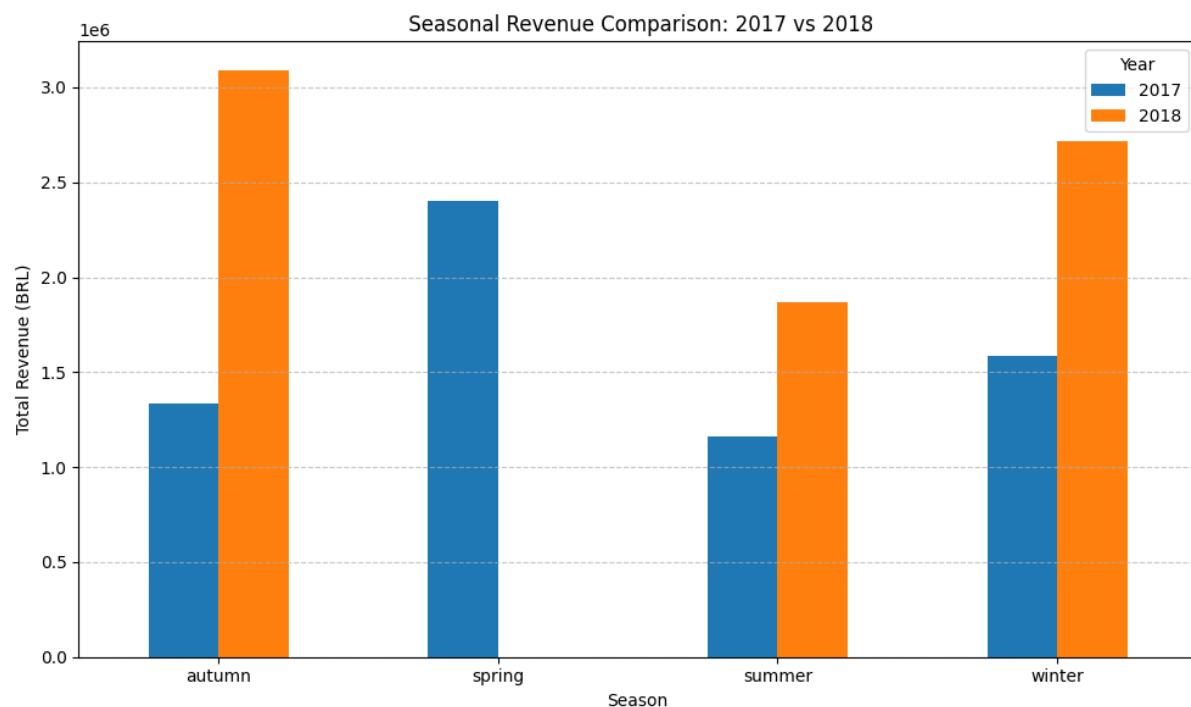
```
customer_city
agrestina                  3978.00
pirpirituba                2680.00
pianco                     2200.00
nova esperanca do piria    2199.00
engenheiro navarro         1997.00
mariental                  1799.00
loreto                     1599.99
amarante                   1520.00
periquito                  1481.50
ibitita                    1450.00
dtype: float64
```

Q6: Top 5 No. of customers per city

```
customer_city
sao paulo         14832
rio de janeiro     6545
belo horizonte     2643
brasilia           2050
curitiba           1450
Name: customer_unique_id, dtype: int64
```
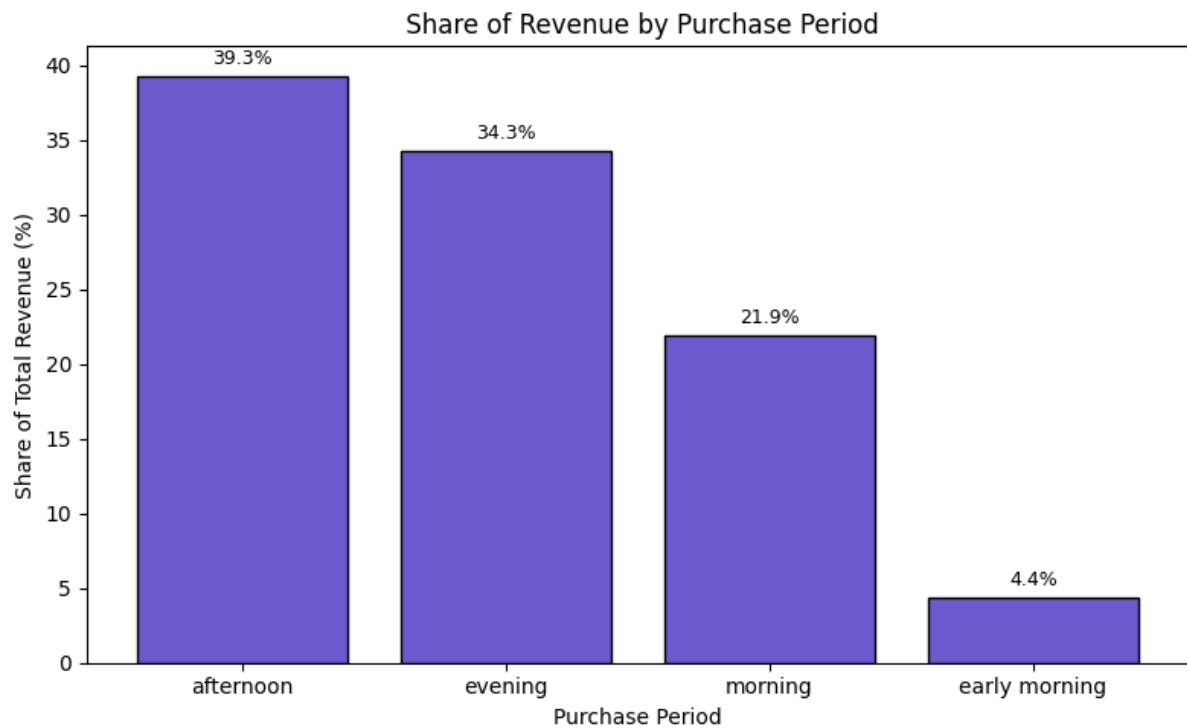
Customer distribution is heavily concentrated in São Paulo and Rio de Janeiro.

Q7: Seasonal Revenue Comparison: 2017 & 2018

Seasonal Revenue Comparison: 2017 vs 2018

Q8: Share of sales by Purchase Period

purchase_period	Total Revenue (BRL)	Revenue Share (%)
afternoon	5566260.70	39.320263
evening	4859135.60	34.325106
morning	3107138.36	21.948935
early morning	623679.73	4.405696

Share of Revenue by Purchase Period
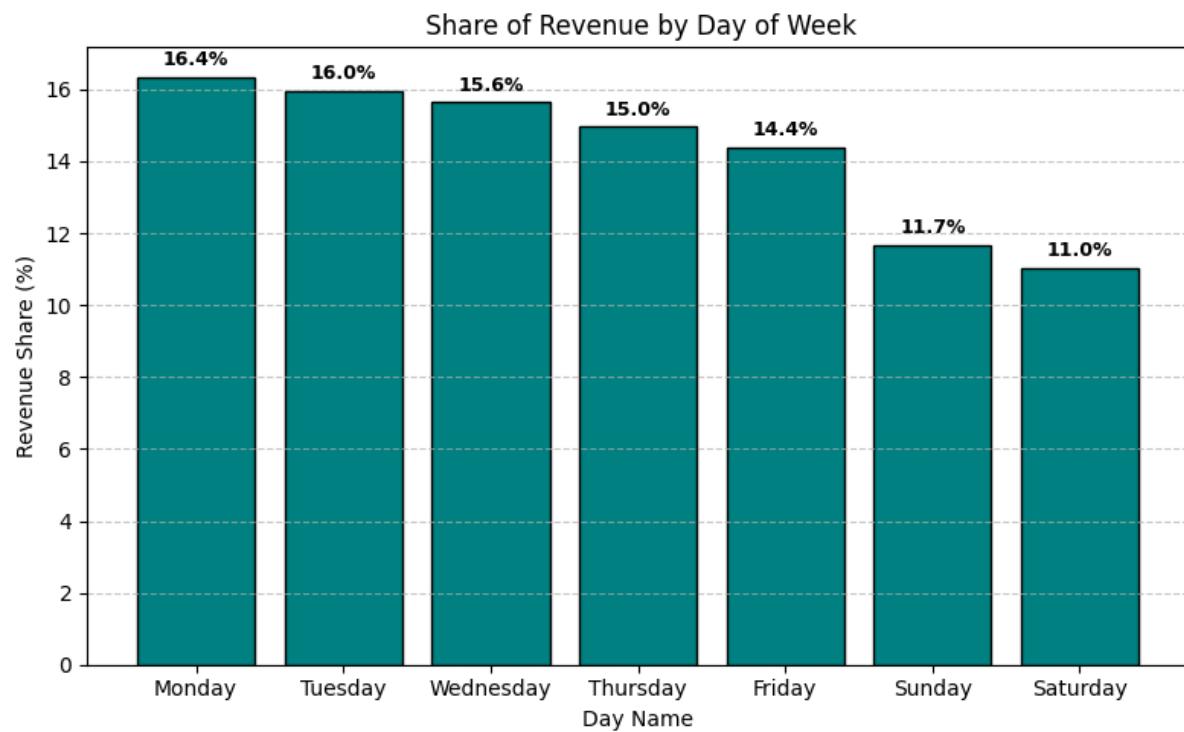
Afternoon and evening periods generate nearly three-quarters of total revenue, indicating peak purchasing hours and optimal windows for marketing, promotions, and inventory prioritisation.

Q9: Share of Revenue by Day Name
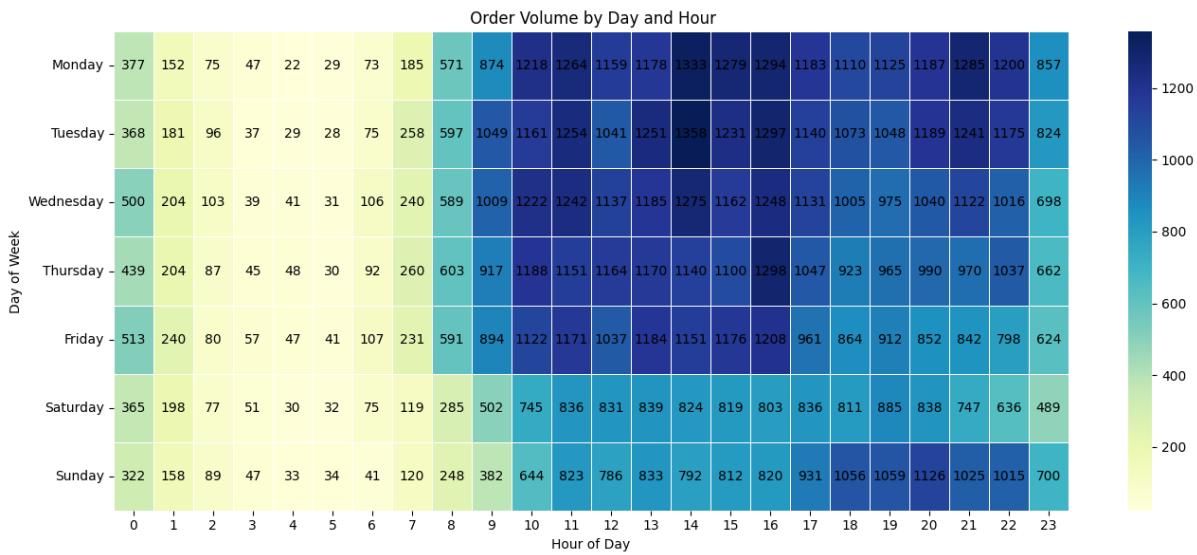
	Total Revenue (BRL)	Average Revenue per Day	Revenue Share (%)
day_name			
Monday	2314835.72	121.341706	16.352082
Tuesday	2258780.06	118.876904	15.956102
Wednesday	2213796.42	120.840416	15.638336
Thursday	2117275.45	120.780117	14.956509
Friday	2038528.40	122.045644	14.400237
Sunday	1650396.93	118.767770	11.658462
Saturday	1562601.41	123.301618	11.038272

Weekdays drive 88% of revenue, Monday leading in total share. Weekend sales lag despite strong per-day averages, suggesting lower traffic or fewer active selling hours.

Q10: Share of Revenue by Day of Week

Share of Revenue by Day of Week

Day Name	Revenue Share (%)
Monday	16.4%
Tuesday	16.0%
Wednesday	15.6%
Thursday	15.0%
Friday	14.4%
Sunday	11.7%
Saturday	11.0%

Q11: Order Traffic Heatmap

Order Volume by Day and Hour

Day of Week	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Monday	377	152	75	47	22	29	73	185	571	874	1218	1264	1159	1178	1333	1279	1294	1183	1110	1125	1187	1285	1200	857
Tuesday	368	181	96	37	29	28	75	258	597	1049	1161	1254	1041	1251	1358	1231	1297	1140	1073	1048	1189	1241	1175	824
Wednesday	500	204	103	39	41	31	106	240	589	1009	1222	1242	1137	1185	1275	1162	1248	1131	1005	975	1040	1122	1016	698
Thursday	439	204	87	45	48	30	92	260	603	917	1188	1151	1164	1170	1140	1100	1298	1047	923	965	990	970	1037	662
Friday	513	240	80	57	47	41	107	231	591	894	1122	1171	1037	1184	1151	1176	1208	961	864	912	852	842	798	624
Saturday	365	198	77	51	30	32	75	119	285	502	745	836	831	839	824	819	803	836	811	885	838	747	636	489
Sunday	322	158	89	47	33	34	41	120	248	382	644	823	786	833	792	812	820	931	1056	1059	1126	1025	1015	700

Q12: Share of Order by Price Segment

Half of all orders come from low-priced items, indicating strong demand elasticity and price sensitivity among customers. Medium and high-priced segments share the remaining volume equally, suggesting opportunities for bundling, upselling, or targeted promotions

Q13: Customer Retention Cohort Analysis (Based on First Purchase Month)
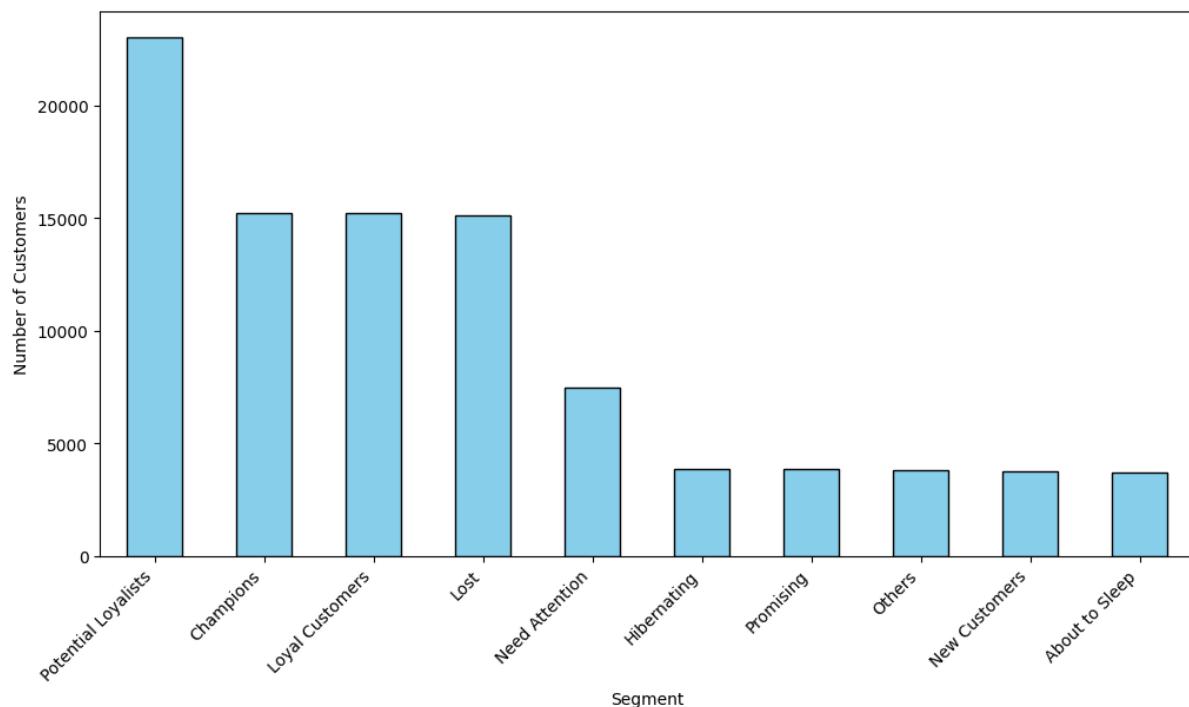
cohort_index cohort_month	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2017-01	753	3	2	1	3	1	3	1	1	0	3	1	5	3	1
2017-02	1693	4	5	2	7	2	4	3	1	3	2	5	2	3	2
2017-03	2595	13	9	10	9	4	4	8	9	2	10	3	6	3	4
2017-04	2340	14	5	4	8	6	8	7	7	4	6	2	2	1	2
2017-05	3560	17	17	14	11	12	15	5	9	10	9	12	9	1	7
2017-06	3114	15	11	12	8	12	12	7	4	7	10	11	5	4	6
2017-07	3844	20	14	9	11	8	12	4	7	10	8	11	5	10	0
2017-08	4149	28	14	11	15	22	12	11	6	6	10	8	4	0	0
2017-09	4091	28	22	12	18	9	9	10	12	8	11	3	0	0	0
2017-10	4413	31	11	4	10	9	10	16	12	9	9	0	0	0	0
2017-11	7218	40	28	13	14	13	8	14	10	4	0	0	0	0	0
2017-12	5442	12	15	19	14	11	9	1	12	0	0	0	0	0	0
2018-01	6984	24	26	20	20	11	12	16	0	0	0	0	0	0	0
2018-02	6422	25	25	19	17	14	13	0	0	0	0	0	0	0	0
2018-03	6948	31	22	20	9	8	0	0	0	0	0	0	0	0	0
2018-04	6709	39	21	16	9	0	0	0	0	0	0	0	0	0	0
2018-05	6606	35	18	14	0	0	0	0	0	0	0	0	0	0	0
2018-06	5935	25	16	0	0	0	0	0	0	0	0	0	0	0	0
2018-07	6054	31	0	0	0	0	0	0	0	0	0	0	0	0	0
2018-08	6238	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Q14: Customer Spending Cohort Analysis

cohort_index cohort_month	0	1	2	3	4	5
2017-01	128552.10	124.92	82.89	75.00	199.70	59.99
2017-02	259044.40	478.38	484.50	89.25	978.22	49.99
2017-03	395889.77	1493.86	1452.24	1873.01	989.33	1306.99
2017-04	389731.06	2638.53	800.38	670.78	2076.98	1046.38
2017-05	540960.93	1561.40	2777.79	1575.60	1088.59	1413.91
2017-06	447758.41	3716.20	1673.04	1238.65	1278.79	1421.68
2017-07	525039.80	4848.00	1878.25	1005.16	951.74	1653.48
2017-08	589265.37	4369.76	1905.19	804.74	1539.26	2541.32
2017-09	651800.29	3046.03	2845.74	1518.35	4171.39	1663.48
2017-10	682181.29	3996.34	1324.51	520.67	1199.00	1184.40
2017-11	1029924.98	8905.26	3852.36	1310.31	2161.32	1836.63
2017-12	752029.29	2167.08	1553.58	2490.15	2444.46	1861.43
2018-01	966934.08	3237.50	3205.84	2873.83	2284.46	1645.50
2018-02	866551.76	3162.25	3586.46	3245.69	2034.82	2556.34
2018-03	1007988.13	3886.14	2899.04	2899.34	1410.85	428.85
2018-04	1005289.49	4299.32	3253.66	1506.86	682.01	0.00
2018-05	1007100.30	7974.05	2203.76	2015.90	0.00	0.00

Q15: RFM Analysis

Customer Distribution by RFM Segment

The customer base is anchored by Potential Loyalists and Champions, indicating strong recent engagement and high-value behaviour. These segments should be prioritised for retention, loyalty rewards, and personalised marketing. Lost and low-engagement segments offer reactivation potential through targeted campaigns.
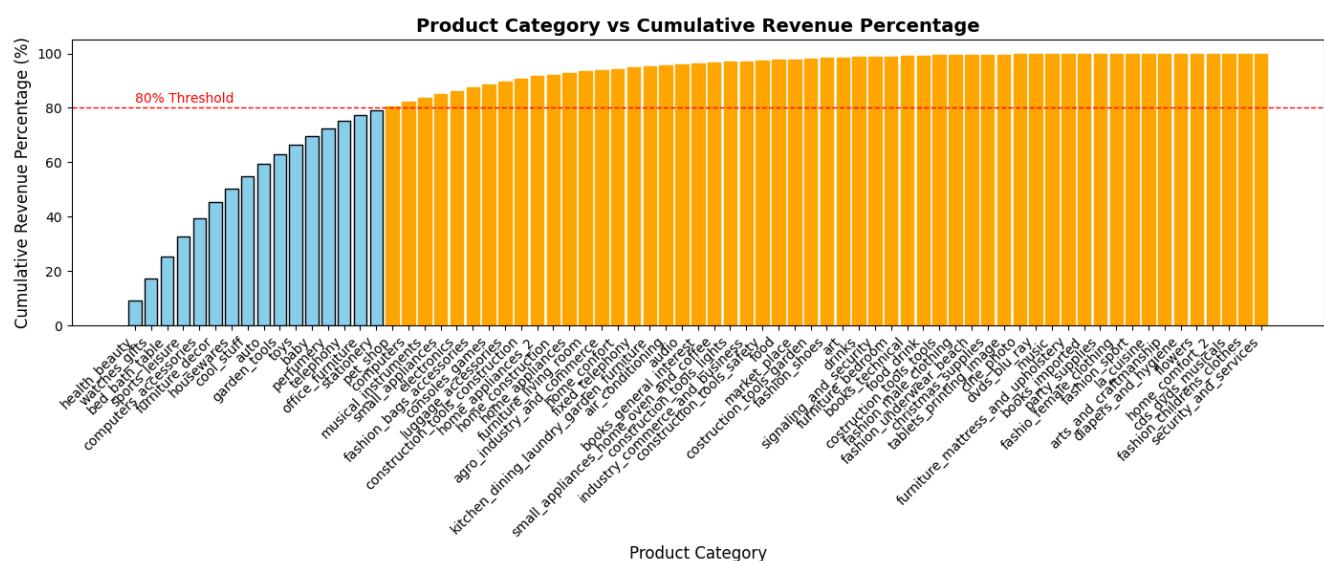
Q16: Pareto Analysis

Which categories/ number of products contribute to 80% of the revenue?

ABC_Class	Count	Percentage (%)
Class C	12038	36.719131
Class B	11659	35.563080
Class A	9087	27.717789

Class A products represent only 27% of the assortment yet contribute approximately 80% of total revenue. This concentration highlights a highly productive core range. The company should prioritise these Class A items in inventory planning, marketing, and operational focus to maximise revenue efficiency and ensure consistent availability

| ABC_Class	Count	Percentage (%)
Class C	39	54.929577
Class A	16	22.535211
Class B	16	22.535211
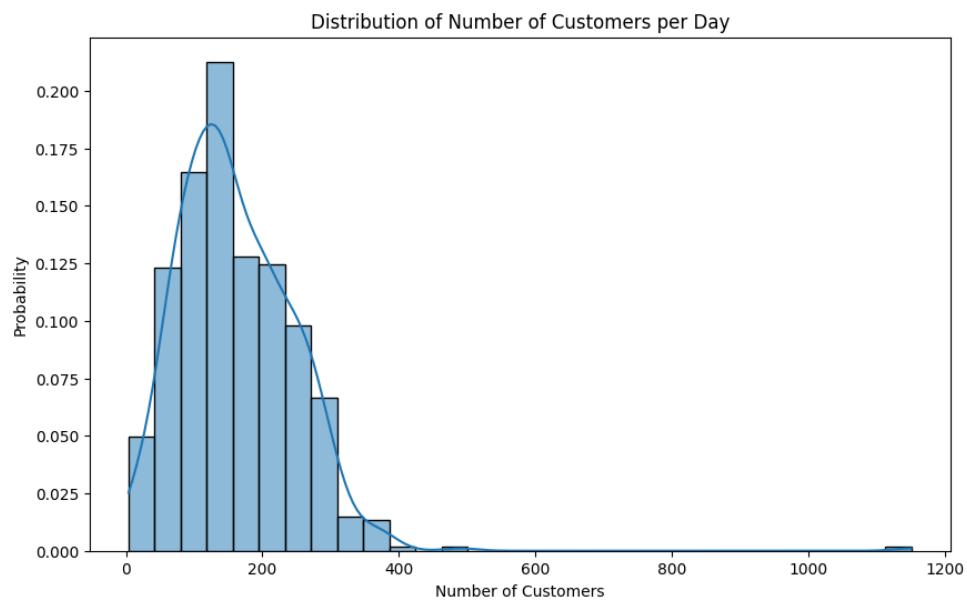
22% of all categories generate 80% of total revenue, indicating a classic Pareto distribution. Within this high-impact group, **Health & Beauty emerges as the single largest revenue contributor.**

6.2. Customer Analysis

Q1: Unique number of customers

95108

## Q2: Distribution of Unique number of customers per day

Distribution of Number of Customers per Day

## 6.3. Delivery Analysis
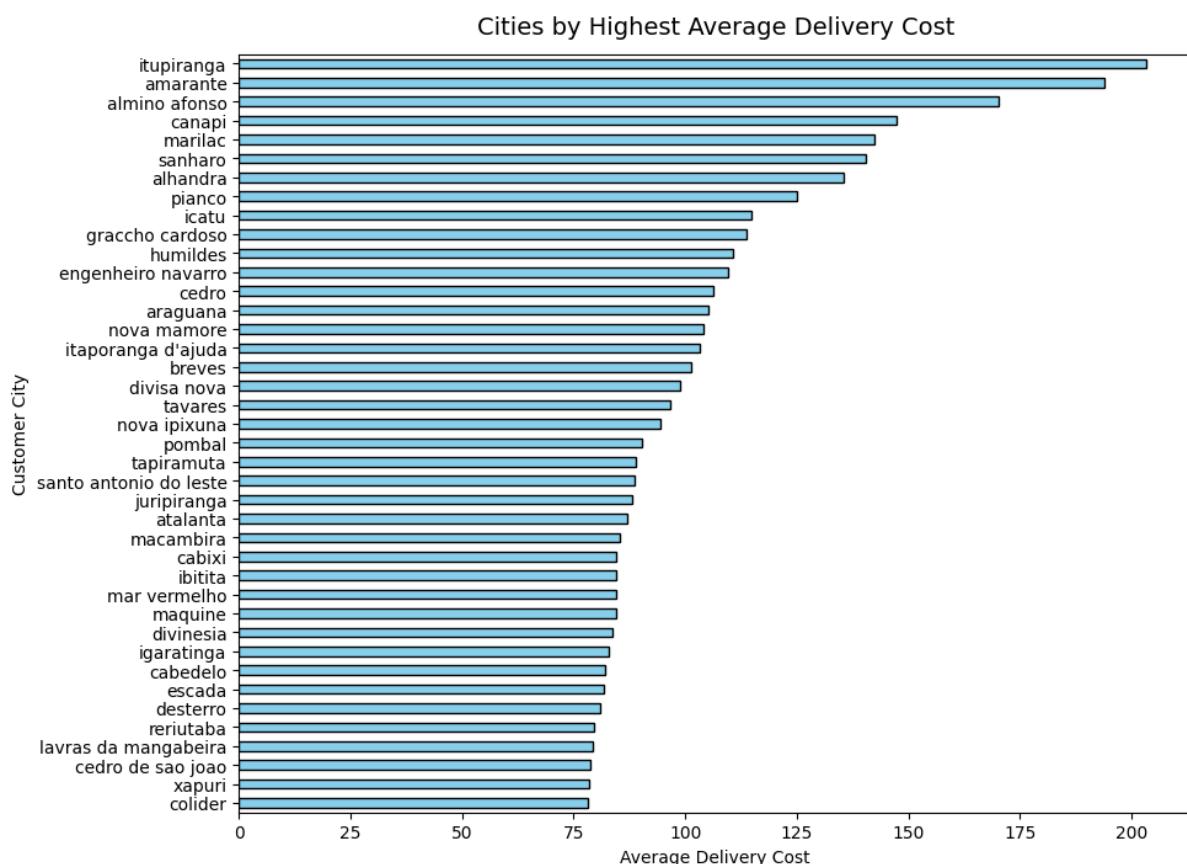
Q1: AVG Delivery Cost per order

23.89
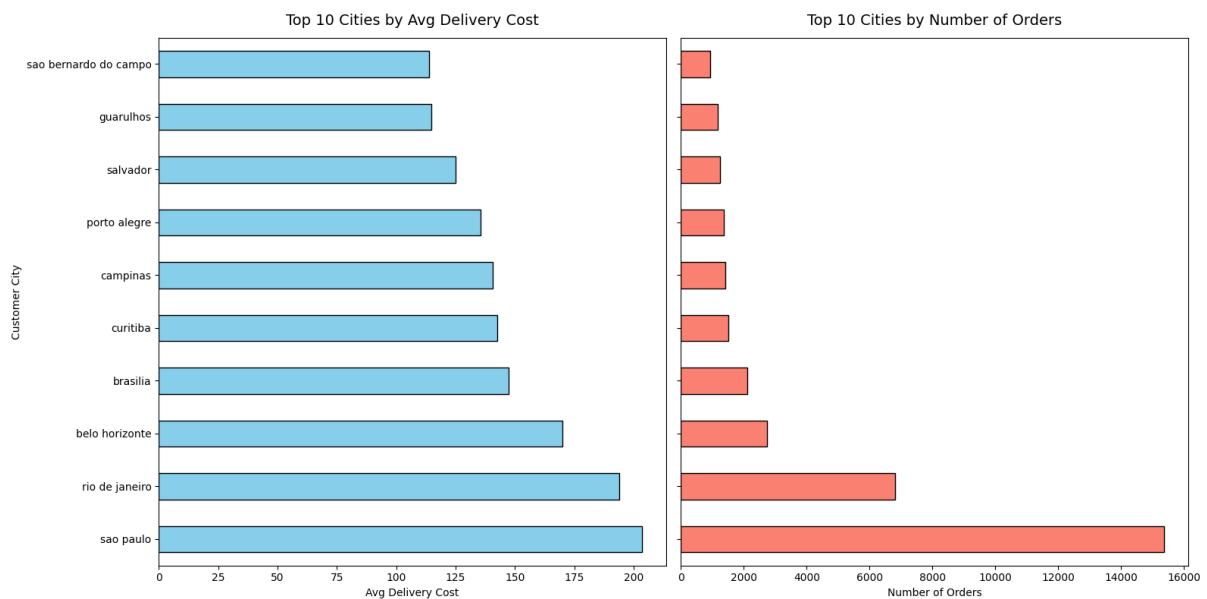
Q2: AVG Delivery Cost per city

	customer_city	avg_delivery_cost	order_count
0	sao paulo	14.380280	15364
1	rio de janeiro	20.810253	6801
2	belo horizonte	19.567772	2738
3	brasilia	21.104226	2110
4	curitiba	18.877635	1504
5	campinas	14.997459	1422
6	porto alegre	20.769787	1366
7	salvador	25.456699	1238
8	guarulhos	14.723764	1177
9	sao bernardo do campo	13.594301	925

Q3: Cities by Highest Average Delivery Cost

Cities by Highest Average Delivery Cost

Q4: Top 10 cities by avg delivery cost per city and no. of orders
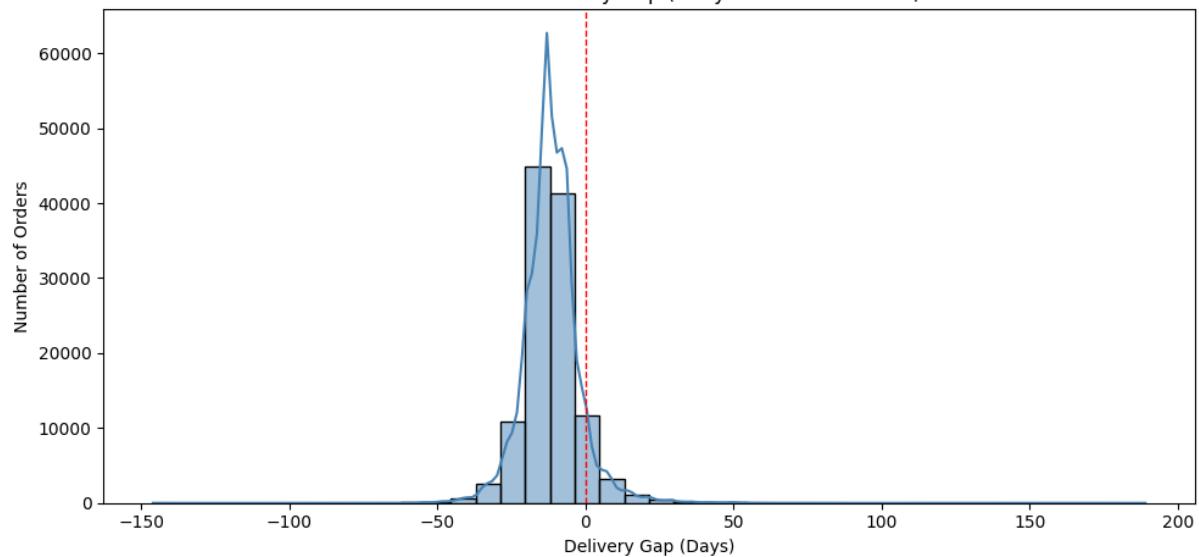
Q5: On-Time Delivery Rate

delivery_status
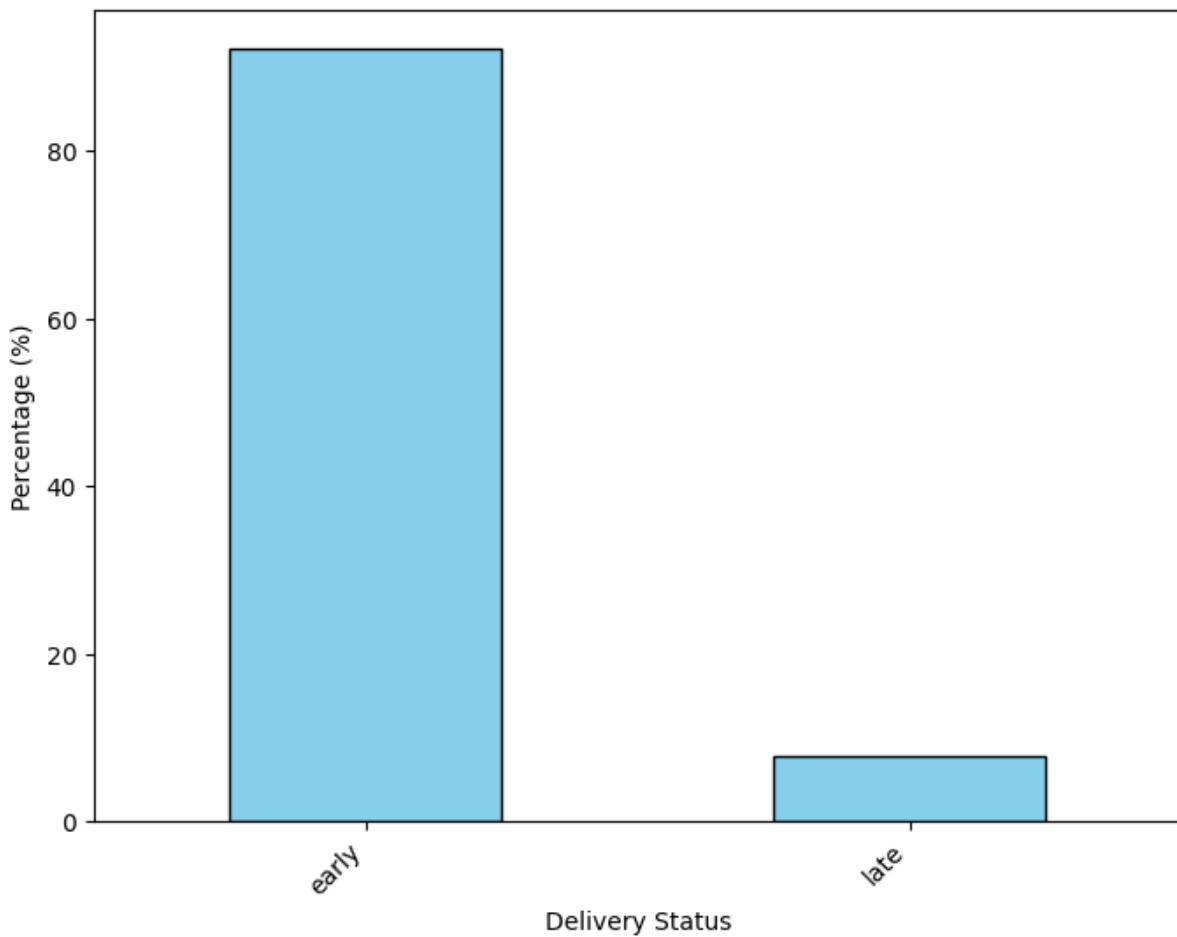
early   92.202218

late   7.797782
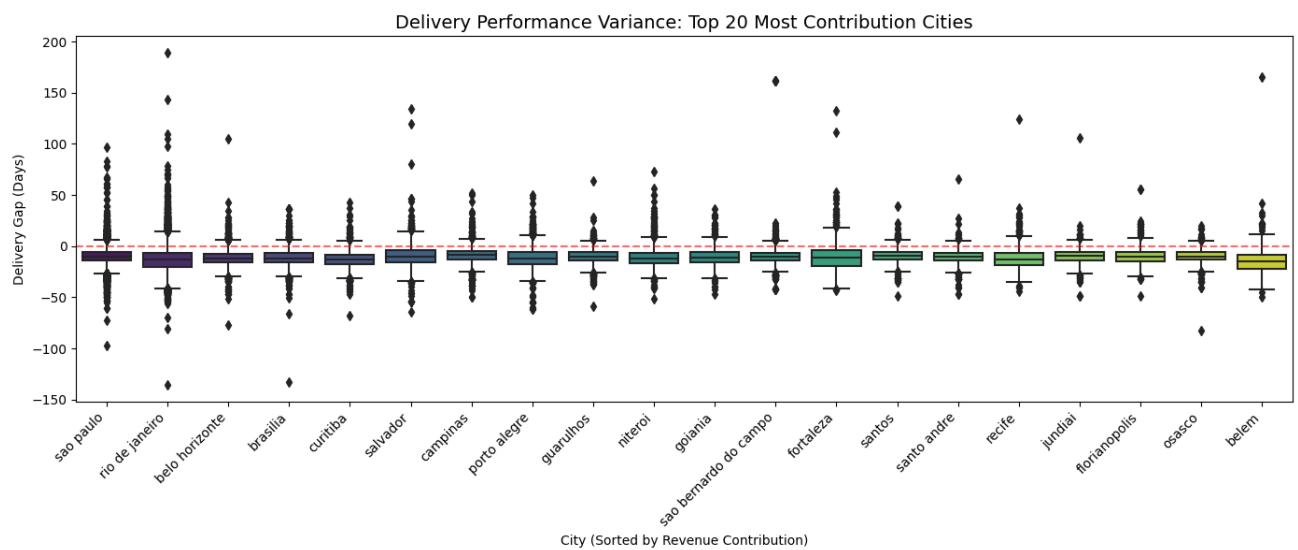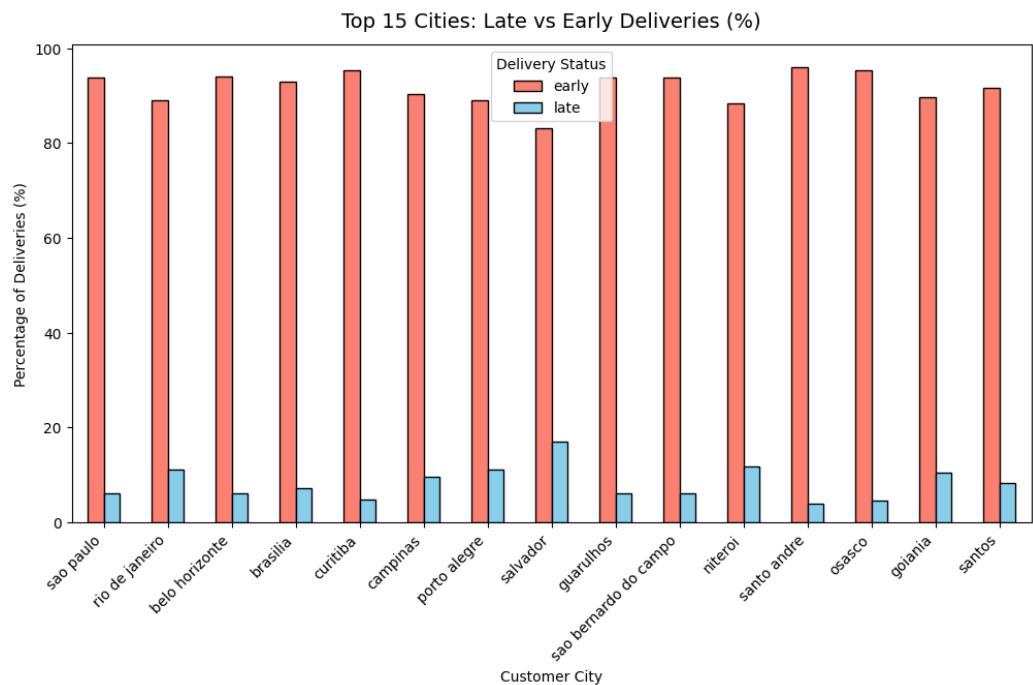
Distribution of Delivery Gap (Early vs Late Deliveries)

Delivery Status Distribution

Q6: Logistics Efficiency: SLA Success Rate Across Top 15 Cities

Top 15 Cities: Late vs Early Deliveries (%)

Delivery Performance Variance: Top 20 Most Contribution Cities

Q7: Fulfilment Efficiency Metrics

7.1 Top 10 Sellers with the Worst Purchase Approval Rate
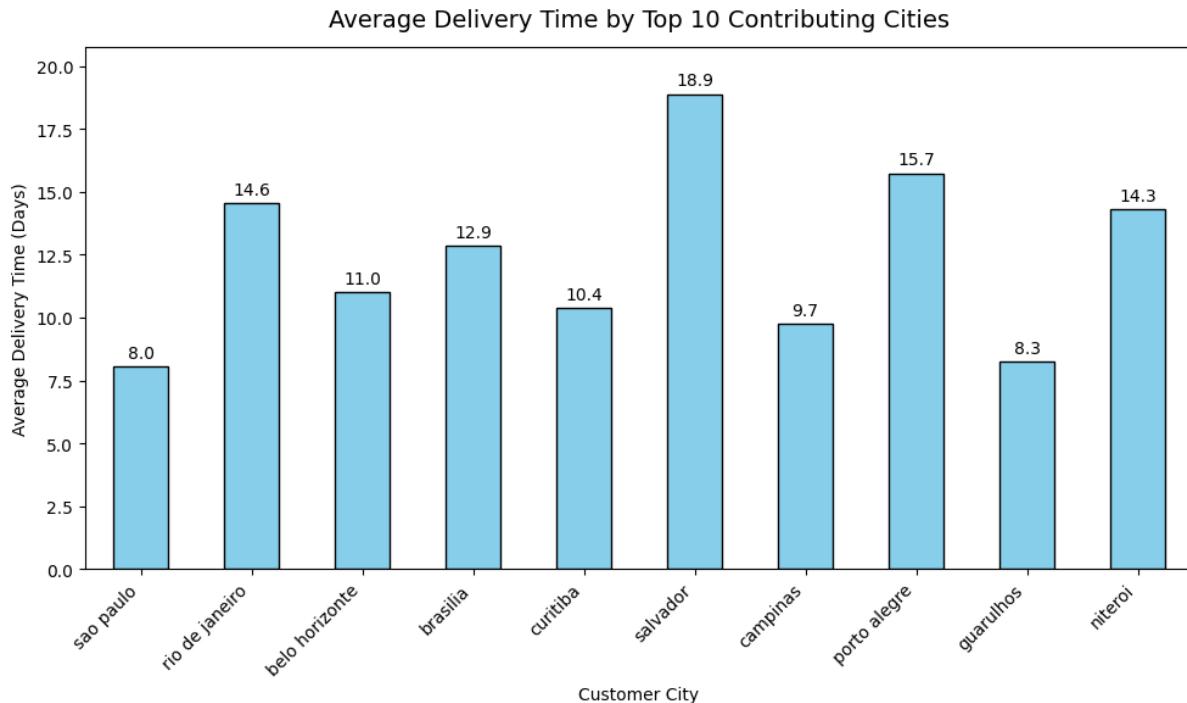
AVG time from purchase to approval minutes

747.0203174732303

7.2 Top 10 Slowest Sellers with AVG time from purchase to approval

seller_id	from_purchase_to_approved_minutes	seller_city	seller_state	order_count	pct_contribution
6560211a19b47992c3666cc44a7e94c0	583.183789	sao paulo	SP	1854	0.907657
4a3ca9315b744ce9f8e9374361493884	597.403188	ibitinga	SP	1806	1.506756
cc419e0650a3c5ba77189a1882b7556a	678.824580	santo andre	SP	1706	0.766022
1f50f920176fa81dab994f9023523100	783.118526	sao jose do rio preto	SP	1404	0.787187
da8622b14eb17ae2831f4ac5b9dab84a	662.908481	piracicaba	SP	1314	1.195925
955fee9216a65b617aa5c0531780ce60	784.043848	sao paulo	SP	1287	0.970139
7a67c85e85bb2ce8582c35f2203ad736	606.944220	sao paulo	SP	1157	1.057485
ea8482cd71df3c1969d7b9473ff13abc	715.687651	sao paulo	SP	1146	0.270043
4869f7a5dfa277a7dca6462dcf3b52b2	578.508432	guariba	SP	1132	1.680303
3d871de0142ce09b7081e2b9d1733cb1	510.281561	campo limpo paulista	SP	1080	0.702078

7.3 Fulfilment Cycle Time by the Top 10 Most Contribution City

Average Delivery Time by Top 10 Contributing Cities

## 7.4   Delivery Performance Summary Sheet by City
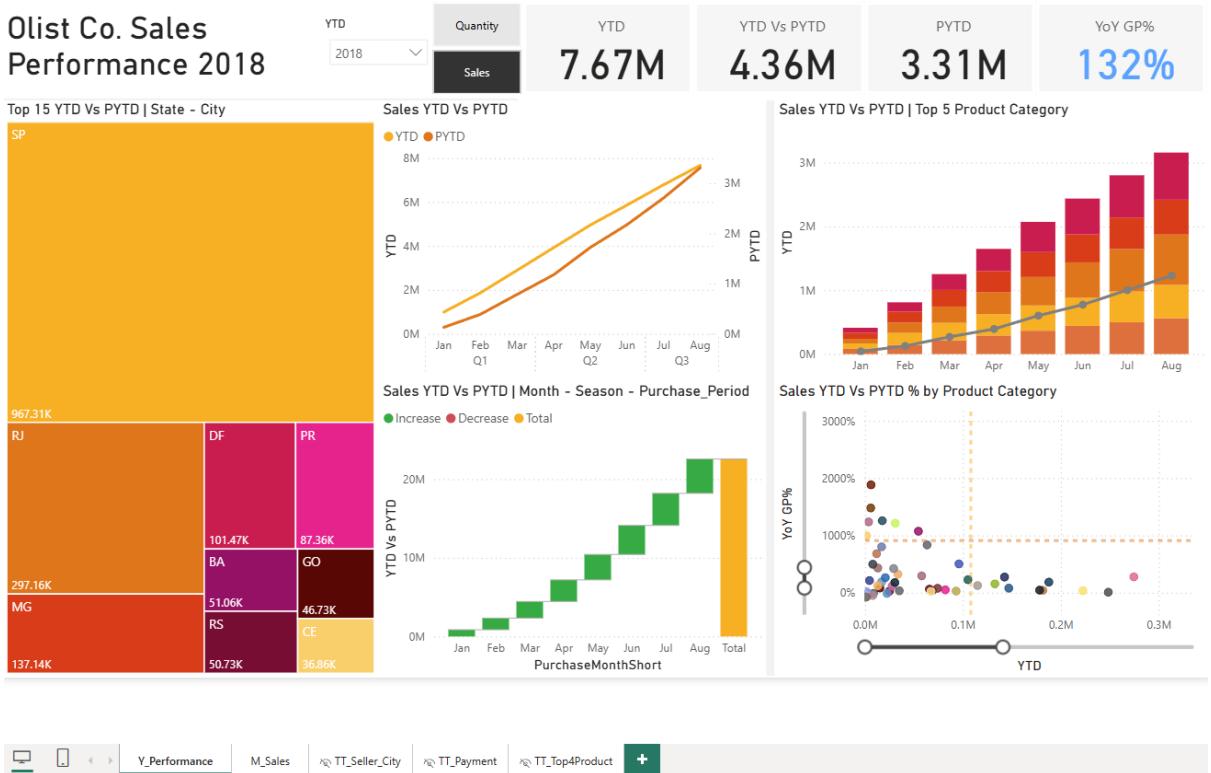
|	customer_city	customer_state	total_sales	contribution_percentage	from_purchase_to_customer_days	early	late	avg_delivery_cost	order_count
0	sao paulo	SP	2003911.01	14.155698	8.039054	93.777634	6.222366	14.380280	15364
1	rio de janeiro	RJ	1022850.06	7.225449	14.566980	88.887526	11.112474	20.810253	6801
2	belo horizonte	MG	365476.09	2.581736	11.026547	93.904703	6.095297	19.567772	2738
3	brasilia	DF	311230.80	2.198545	12.861731	92.813393	7.186607	21.104226	2110
4	curitiba	PR	218114.79	1.540771	10.386689	95.269894	4.730106	18.877635	1504
5	salvador	BA	196675.55	1.389323	18.884086	83.038638	16.961362	25.456699	1238
6	campinas	SP	195045.01	1.377805	9.741502	90.371230	9.628770	14.997459	1422
7	porto alegre	RS	194570.52	1.374453	15.738427	88.888889	11.111111	20.769787	1366
8	guarulhos	SP	155136.95	1.095893	8.264014	93.857143	6.142857	14.723764	1177
9	niteroi	RJ	132316.48	0.934688	14.313903	88.263359	11.736641	20.938063	844

7. PowerBI Visualization
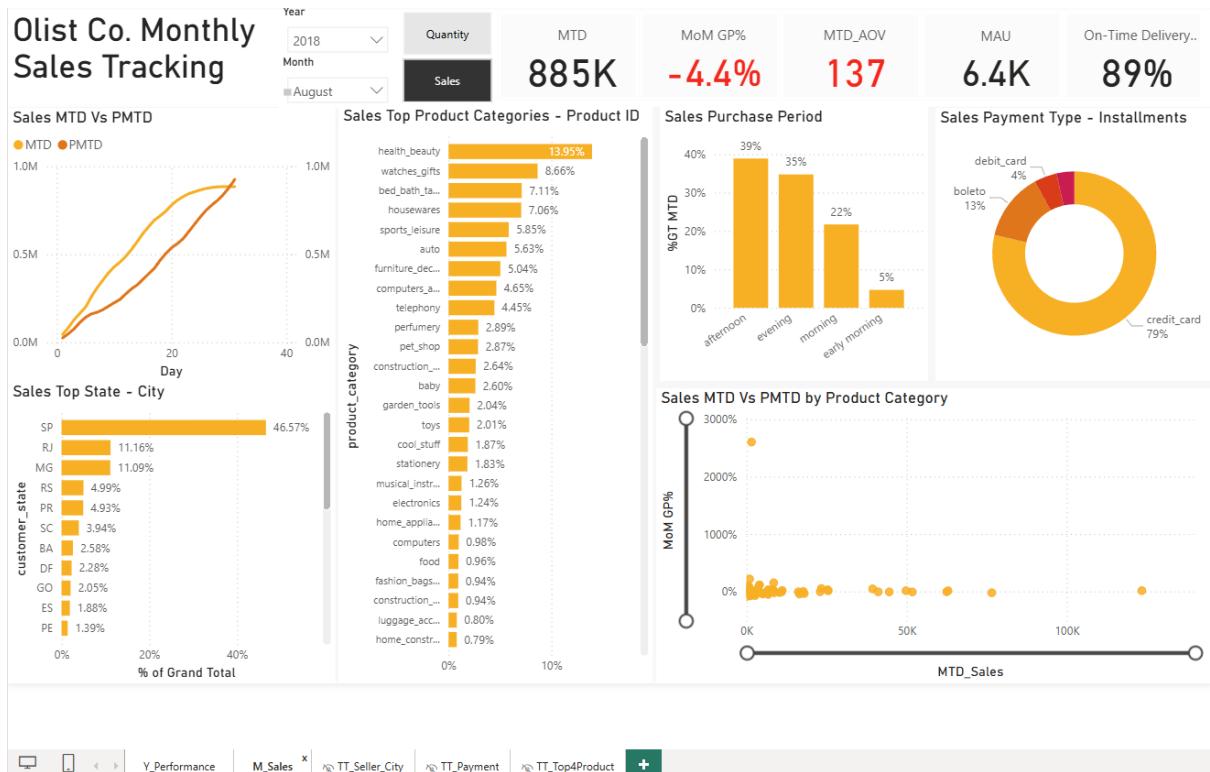
Dashboards were developed to visualise key business metrics:

- Yearly Performance
- Monthly Sales Performance

Power BI serves as the presentation layer, enabling stakeholders to monitor these insights on a daily basis.

Yearly Sales Performance Dashboard

Monthly Sales Performance Dashboard

8. Recommended Actions

1. Sales Performance

- **Prioritise High-Impact Product Categories**
  Focus commercial efforts on the top-performing Class A "Hero Products," which represent 27% of SKUs but generate 80% of revenue. Allocate marketing, stock, and content resources toward these categories, while reviewing low-performing SKUs for delisting, bundling, or targeted promotions to reduce operational waste.

- **Align Promotions With Demand Peaks**
  Run campaigns during the hours, days, and seasons where customer activity is highest. With 73% of revenue occurring in the afternoon and evening, shifting promotional timing to match these peaks can significantly increase conversion.

- **Strengthen Retention and Win-Back Strategies**
  Develop personalised retention journeys for high-value RFM segments (Champions, Loyal) and targeted win-back campaigns for At-Risk customers to reduce churn and protect long-term revenue.

- **Optimise Pricing and Bundling**
  Leverage strong demand in lower-priced segments by offering bundles, incentives, or entry-level offers. Low-priced items account for ~50% of orders, indicating clear price sensitivity and an opportunity to drive volume efficiently.

- **Improve Regional Sales Balance**
  Increase marketing presence and product availability in underperforming cities.

Use regional promotions to unlock growth in emerging markets and reduce over-reliance on top-performing regions.

2. Delivery Performance

- **Improve SLA compliance** by reviewing underperforming cities and addressing carrier or routing issues. Improve delivery SLAs in underserved regions. **Reduce delivery delays** by monitoring delivery gap trends and prioritising improvements in late-delivery regions. Partner with local logistics providers to reduce shipping times.

- **Optimise delivery cost** by renegotiating carrier contracts in high-cost cities and improving route planning.

- **Enhance fulfilment speed** by reducing delays between purchase, approval, and carrier pickup. São Paulo and Rio de Janeiro dominate customer counts.

- **Use delivery performance data** to set more accurate estimated delivery dates and improve customer expectations.

3. Operational Efficiency

- **Adjust warehouse staffing** during peak order hours identified in the order traffic heatmap to reduce fulfilment delays.

- **Improve inventory placement** by aligning stock levels with high-demand cities and regions.

- **Streamline fulfilment processes** to reduce approval time, dispatch delays, and carrier handover issues.