

Marketplace Health & Logistics Optimizer

Presented by: **Omer Metwally**

Last Updated: **10.th February 2026**

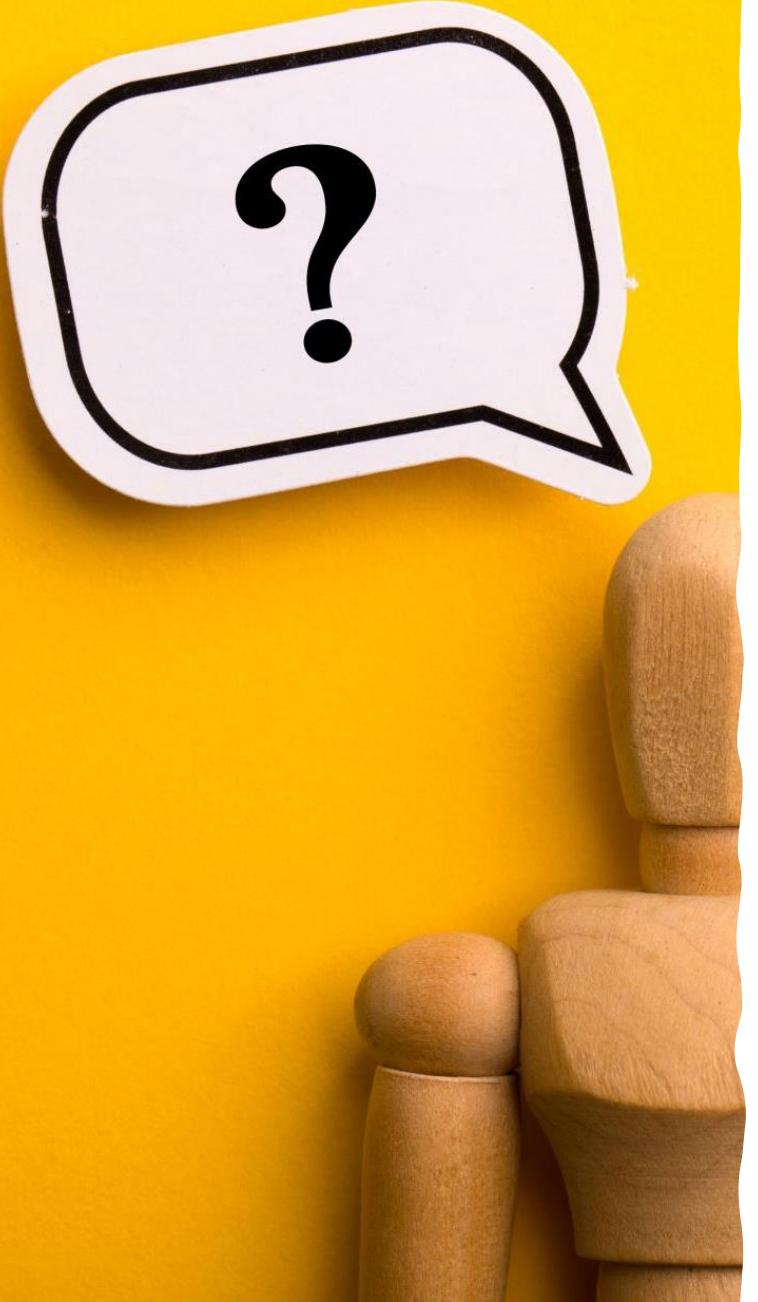


Table of Contents

Healthcare Operations

- Objective
- Business Task
- Data source
- Data cleaning summary
- The story with data.
- Conclusion
- Recommendations

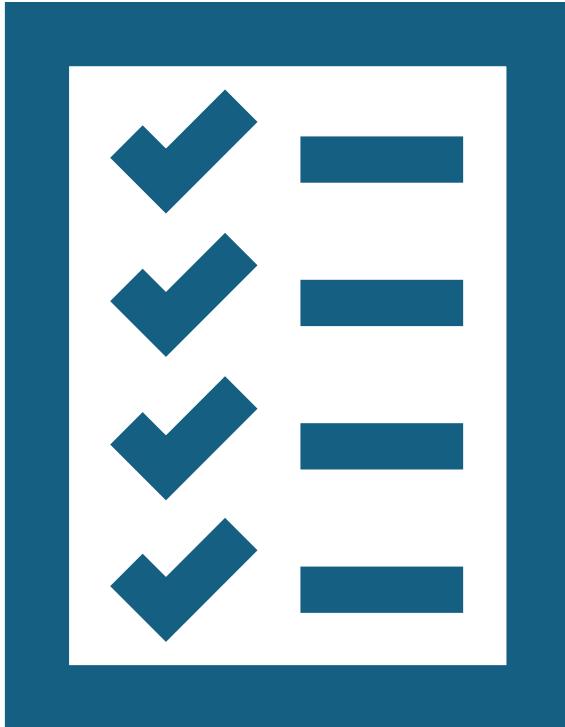
Objective



What are we talking about?

To analyze and visualize **Brazilian E-commerce marketplace data** to identify **key drivers of order cancellations, logistics fulfillment bottlenecks, and customer sentiment trends**, helping marketplace operators **reduce revenue leakage, optimize last-mile delivery efficiency, and enhance the overall customer experience**.

Business Task



The marketplace operator aims to analyze e-commerce transaction and logistics data to identify where and how order cancellations and revenue leakage are concentrated across different product categories, seller performance, and shipping routes.

The analysis focuses on answering key business questions to uncover:

1

Evaluate order lifecycles across diverse seller locations, product categories, and fulfillment stages to establish a baseline for operational performance.

2

Pinpoint underperforming sellers and high-friction categories (such as Office Furniture) that contribute most to the current **21.24% cancellation rate**.

3

Rank the impact of "Last-Mile" delivery times and shipping costs on order completion to prioritize resource allocation for shipping speed improvements.

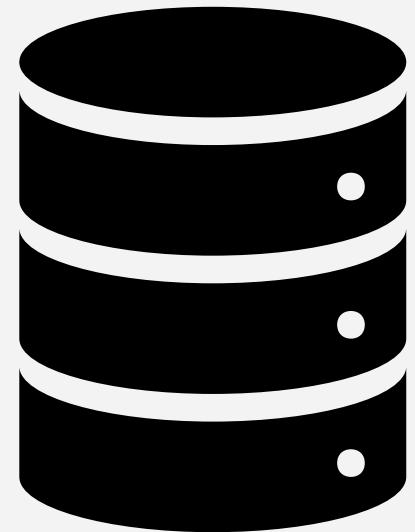
4

Detect regional logistical patterns to identify significant operational bottlenecks occurring within high-volume states like **São Paulo** and **Paraná**.

5

Deliver data-driven recommendations to support strategic scheduling, carrier SLA renegotiations, and pricing optimization to recover the **24% lost revenue**.

Data source



Data set

- The dataset used in this analysis comes from the [**Brazilian E-Commerce Public Dataset by Olist**](#) available on [Kaggle](#)

License: Open dataset provided for educational and analytical purposes.
The data has been used strictly for **learning and non-commercial analysis**.

- **Data License:**
© Motivate International Inc. | Public data used for educational purposes only.

Data cleaning summary

Data cleaning



✓ Handled missing values and orphans

- ✓ Created "Unknown" placeholder records during the SQL phase for orphaned Order, Customer, and Geolocation IDs to prevent data loss during joins.
- ✓ Mapped null or missing city and state entries to a standardized "Unmapped Region" category to maintain visual consistency in geographic reports.

✓ Optimized model relationships

- ✓ Identified and resolved circular dependencies by transitioning cross-filter directions from "Both" to "Single".
- ✓ Deactivated ambiguous paths between Geolocation and Customers to ensure the model correctly prioritizes Seller-to-Customer logistics.

✓ Standardized formats and data types

- ✓ Validate that all (Order Id, Customer Id, Seller Id) columns were formatted as Text to ensure accurate "Distinct Count" calculations across the Snowflake schema.
- ✓ Unified currency and measurement units, converting product dimensions into centimeters (cm) and weight into kilograms (kg) for volumetric analysis

Data cleaning



✓ Engineered logistical metrics

- ✓ Calculated **Waiting Days** by determining the delta between order purchase and delivery timestamps.
- ✓ Categorized **Shipping Cost Levels** (Low, Medium, High) and **Arrival Performance** (On Time vs. Late) to transform raw numbers into actionable business dimensions.

✓ Filtered irrelevant data

- ✓ Identified and excluded test records or orders with "0" dimensions that skewed the average shipping cost and volumetric weight results.
- ✓ Utilized the Filters Pane to hide placeholder "???" and "0" values from executive-level visuals without deleting the underlying transaction history.



Result:

Clean and structured Snowflake schema ready for identifying cancellation drivers and creating the **Logistics Optimization & Revenue Recovery** dashboard..



The story with Data

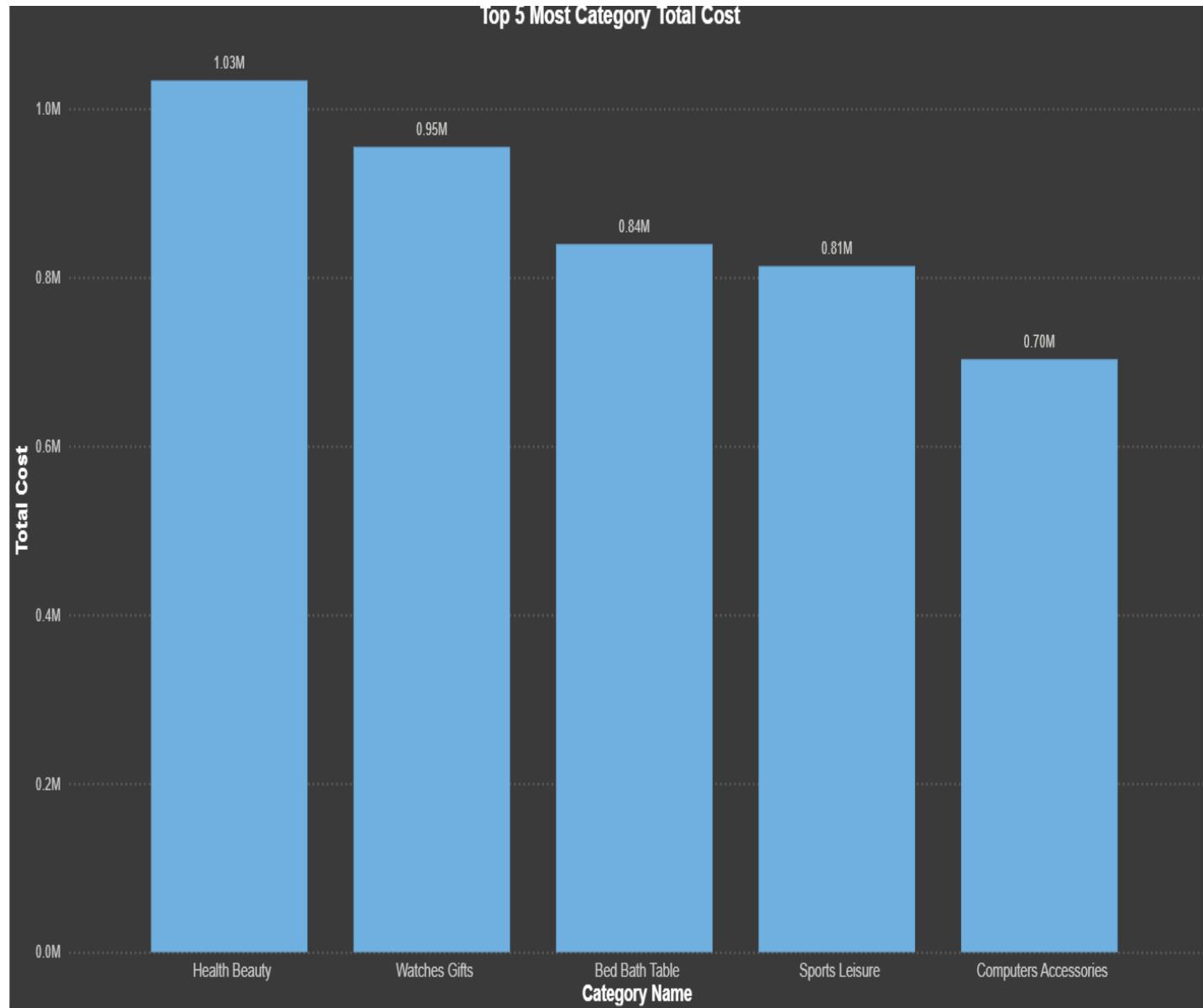


Executive Sales & Operations Summary.

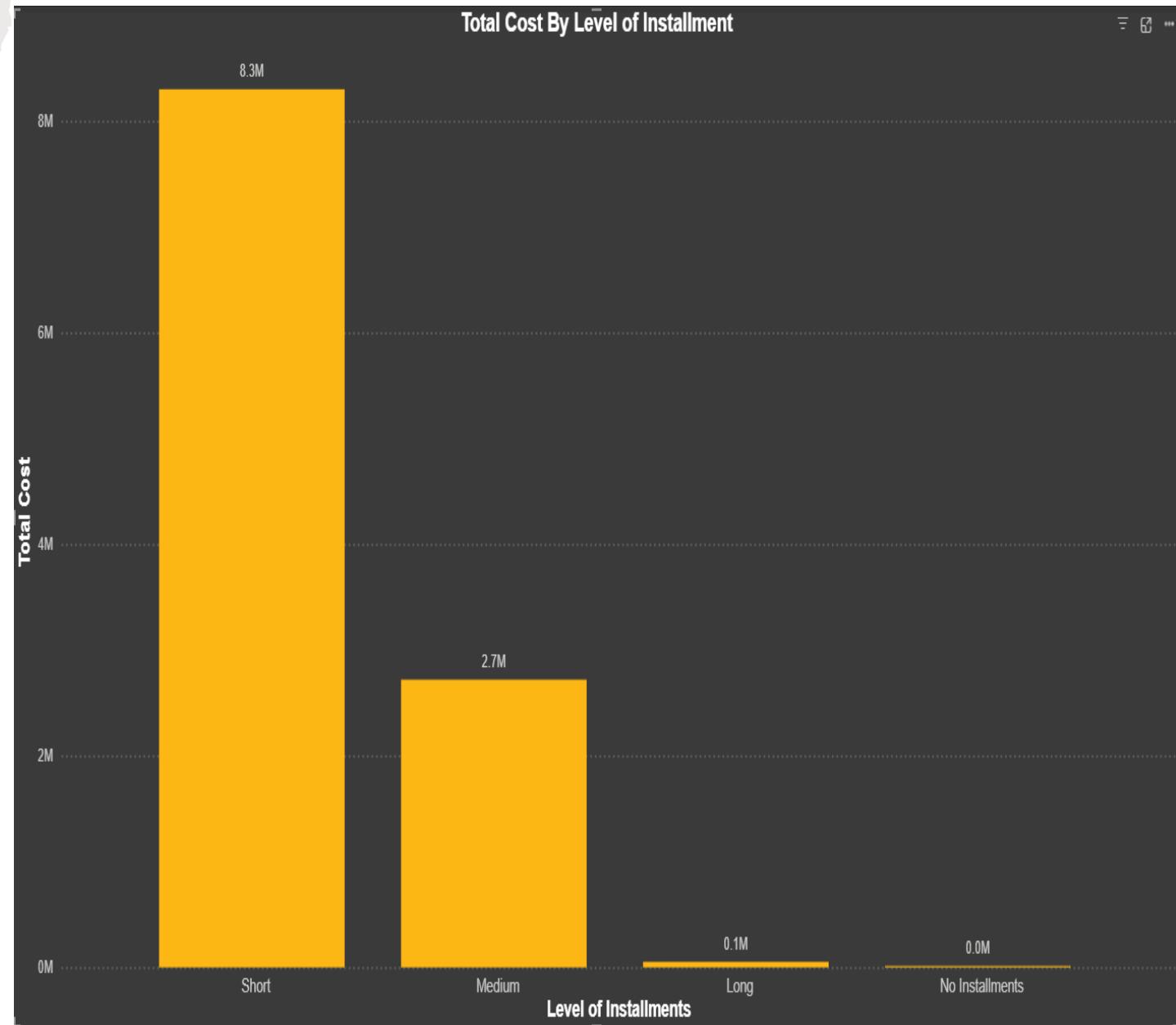
Total Customers	Total Cost	Total Orders	Avg Waiting Days	Avg Months of Installments
78.317K	14.42M	99.44K	9.14	2.93

- A significant **21%** of orders are cancelled, resulting in a **24%** loss of potential revenue.

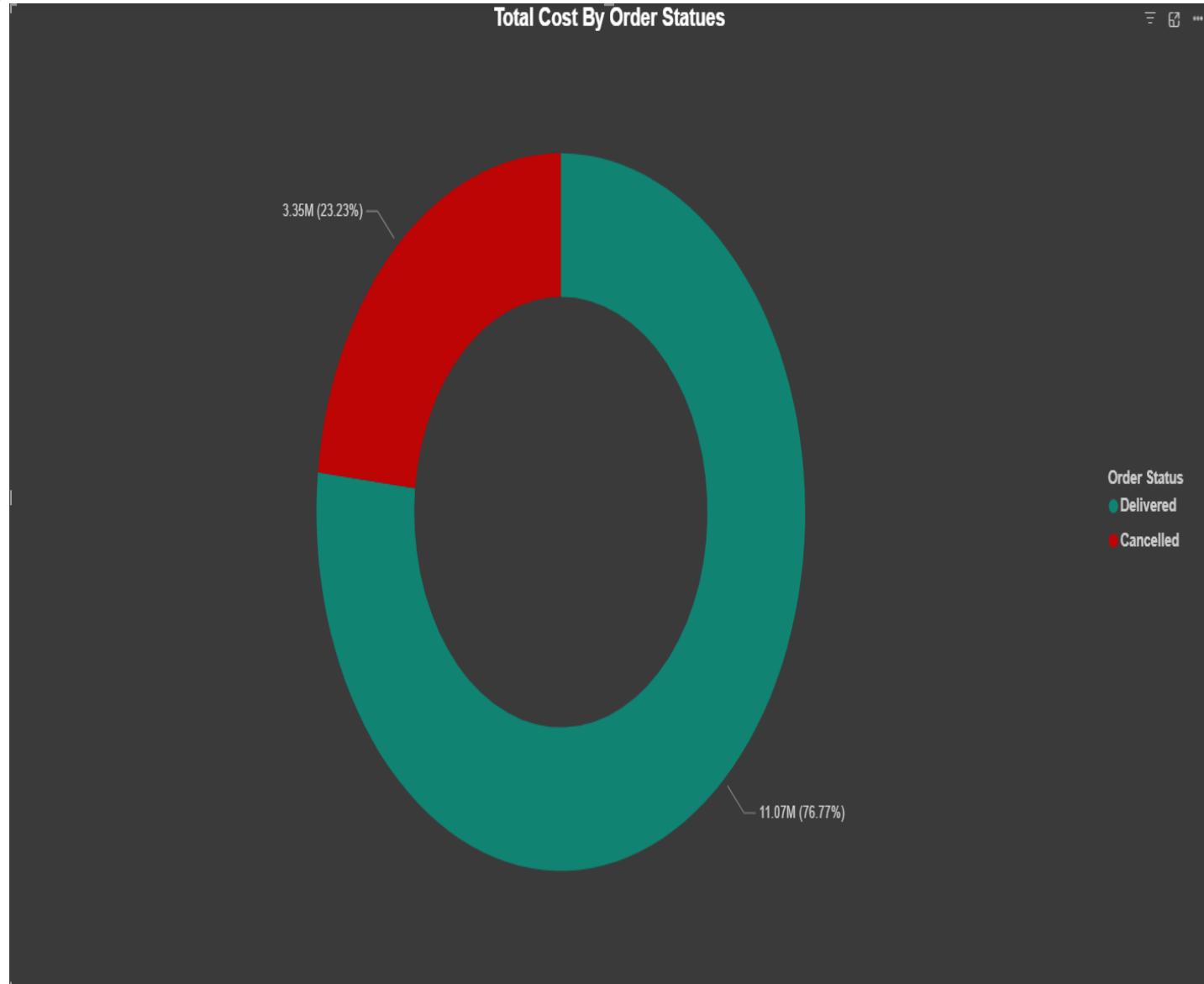
Health Beauty is Top Category by Total Cost With 1.03M\$



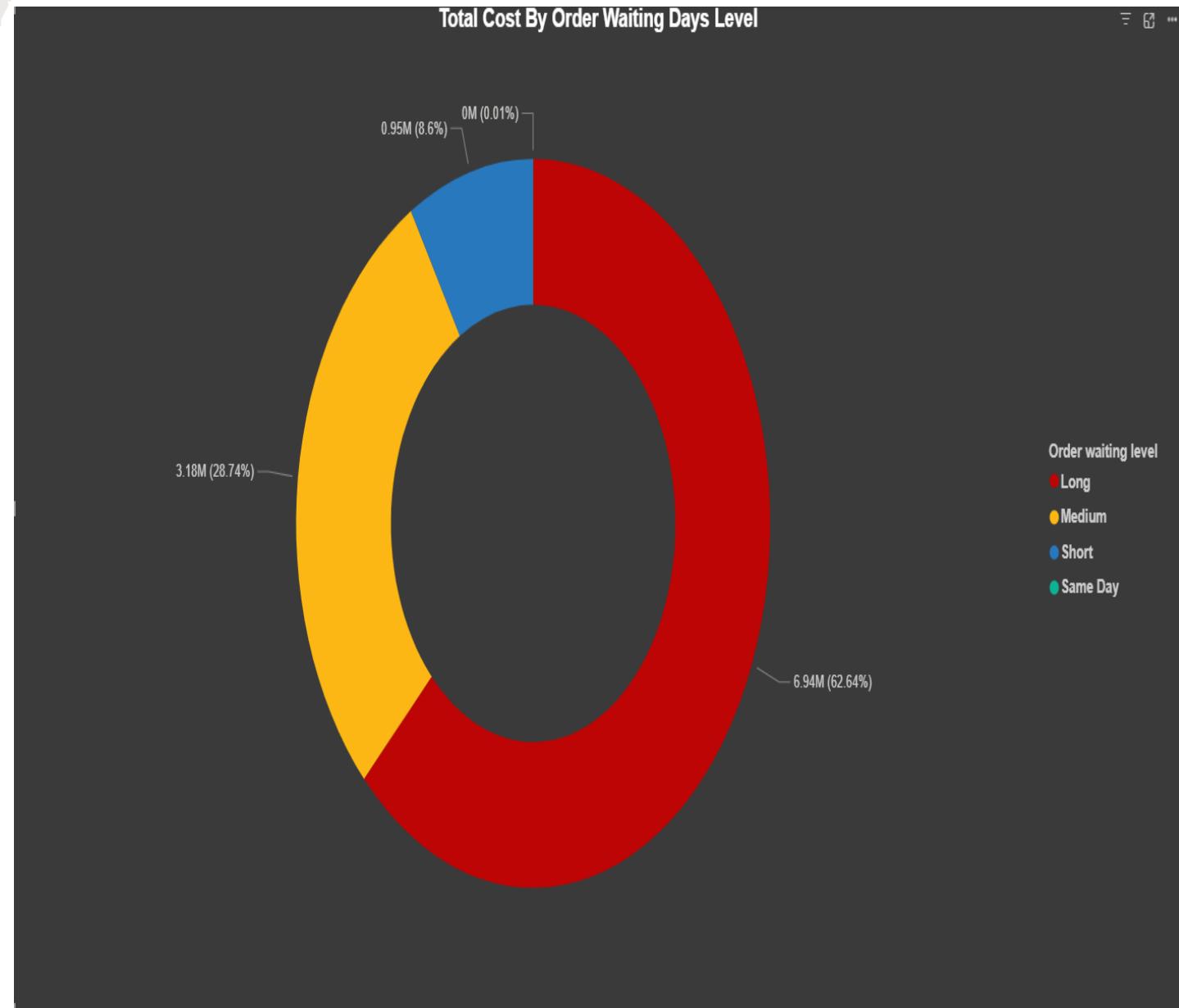
Most of Installments rely on short Period (1-3) Month With Total Value of 8.3M\$ And Least is Long Period

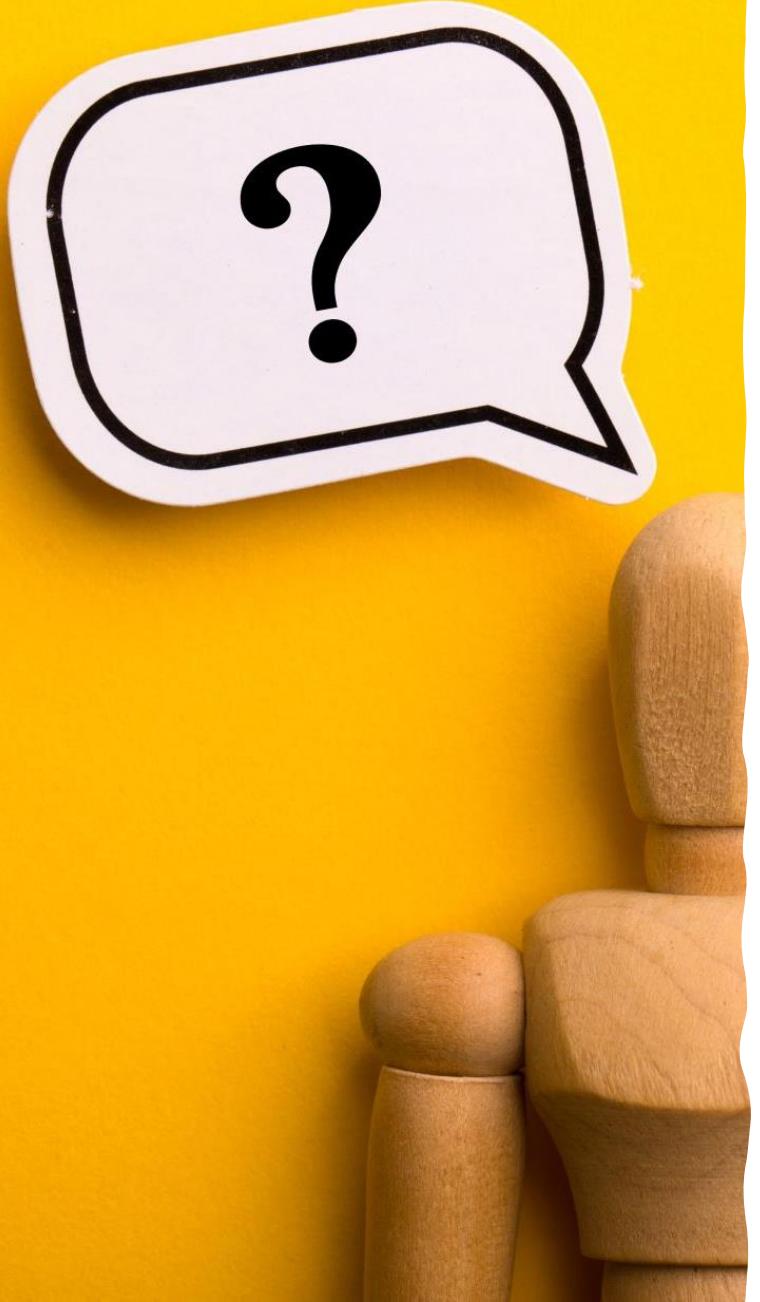


Nearly 23% of Orders Have
been Cancelled With Loss
Nearly 3.35M\$



Nearly 63% Total Cost 6.9M\$
Depends on Long Waiting
Level (15-45) Days





Which States have Long Waiting
Time for Orders and Cancelled
Orders ?

Total States	Avg Long Waiting Days	Total Cancelled Orders
27	12.20	21.124K

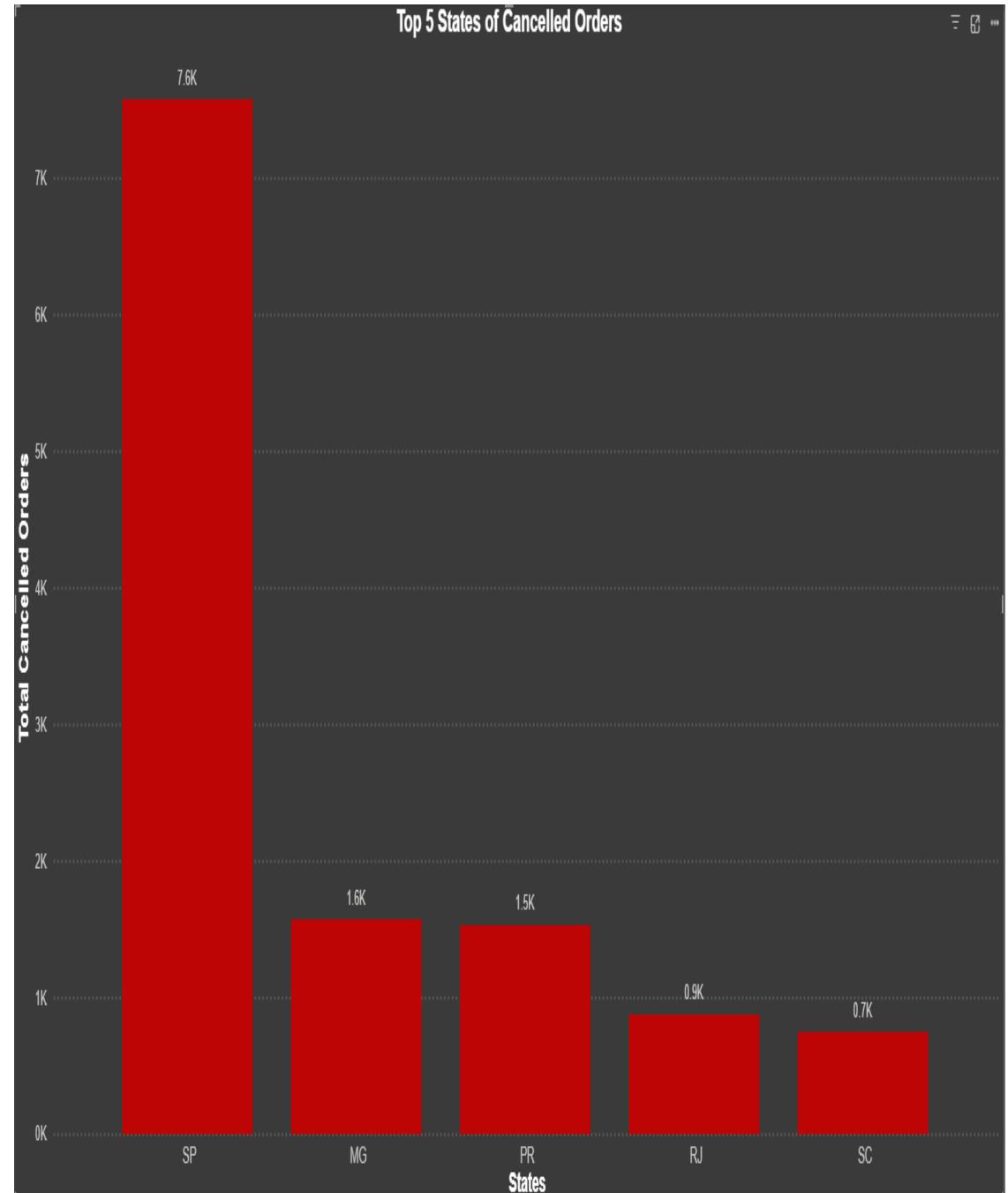
With 27 States There are nearly 21k Orders have been Cancelled and average of waiting long days is 12.2 Days

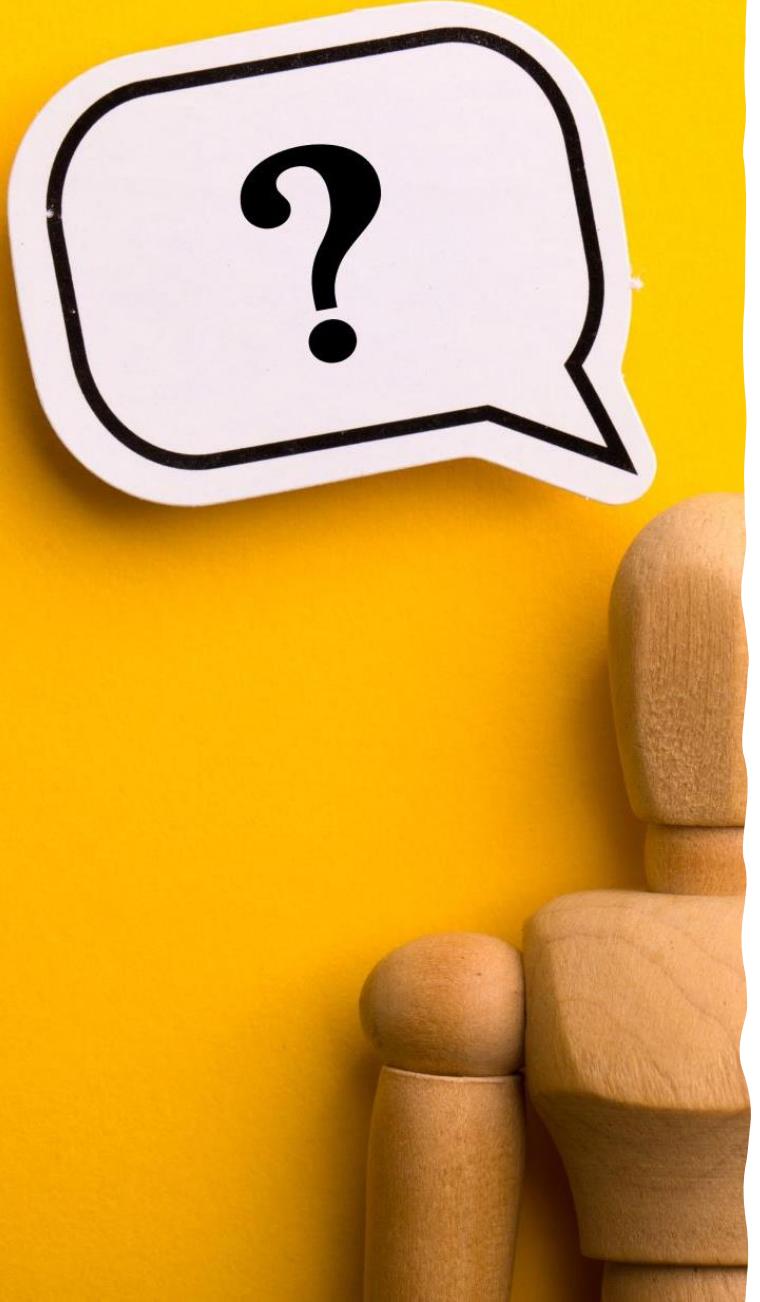
Top 5 States with average
waiting long time at the top

PR



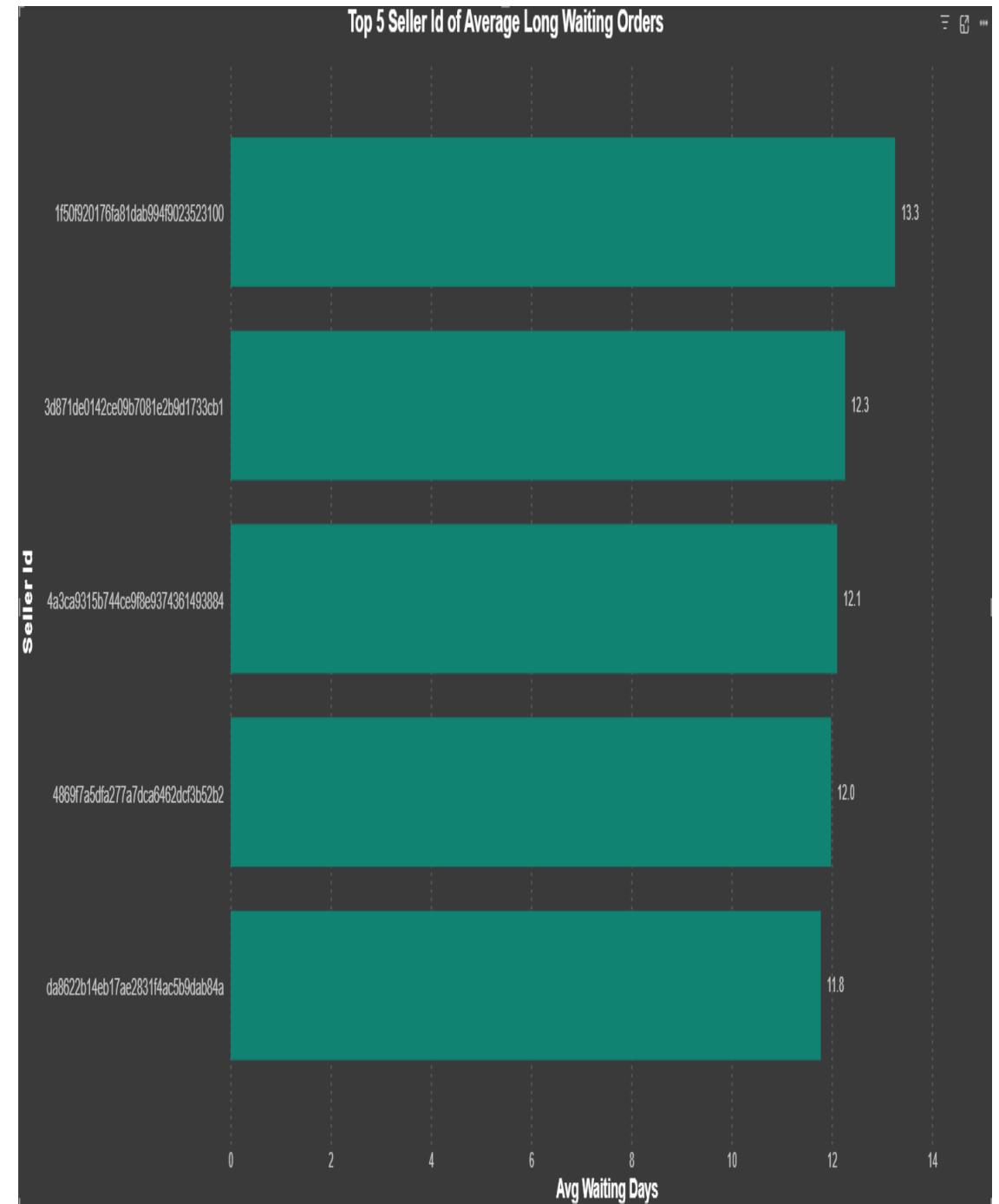
Top 5 States with Cancelled
Orders at the top SP with
Total 7.6k Orders



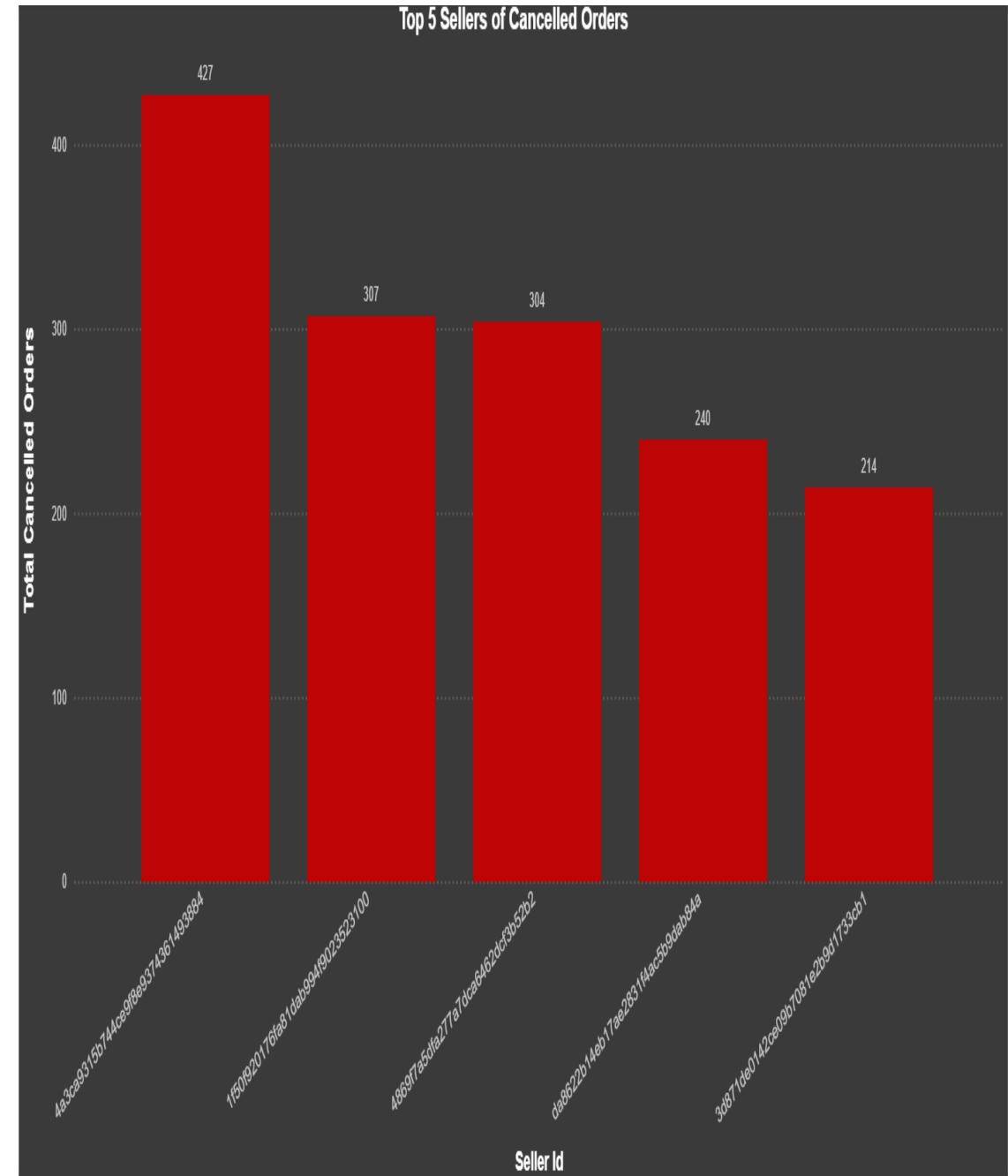


Which Sellers have Long Waiting Time for Orders and Cancelled Orders ?

Top 5 Sellers ID with average waiting long time with nearly 12.5 Days per Seller Id



**Top 5 Sellers ID with
Cancelled Orders Total 1.5k
Orders**



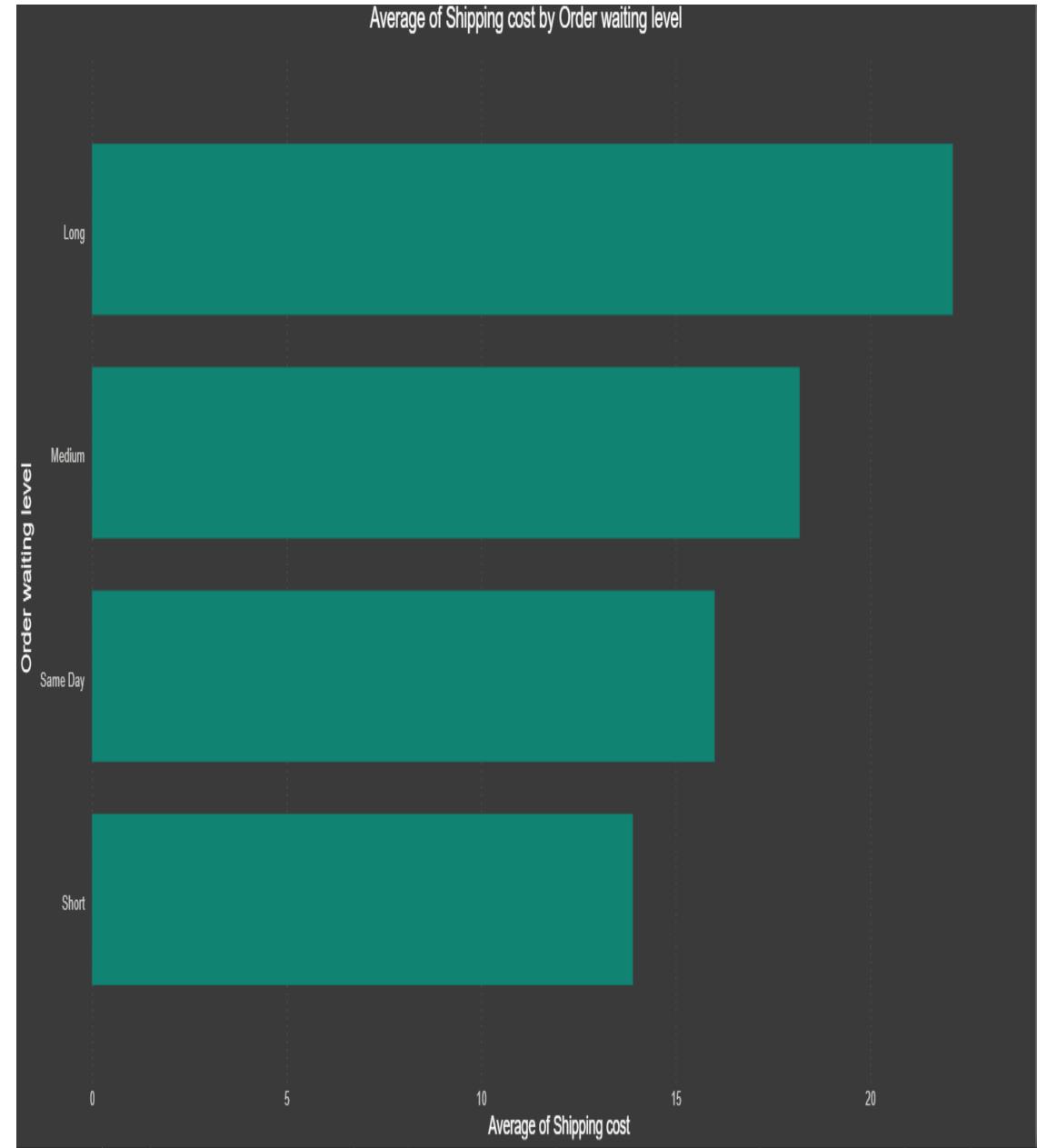


Does a High Shipping Cost directly
Correlate with a Low Review Score ?

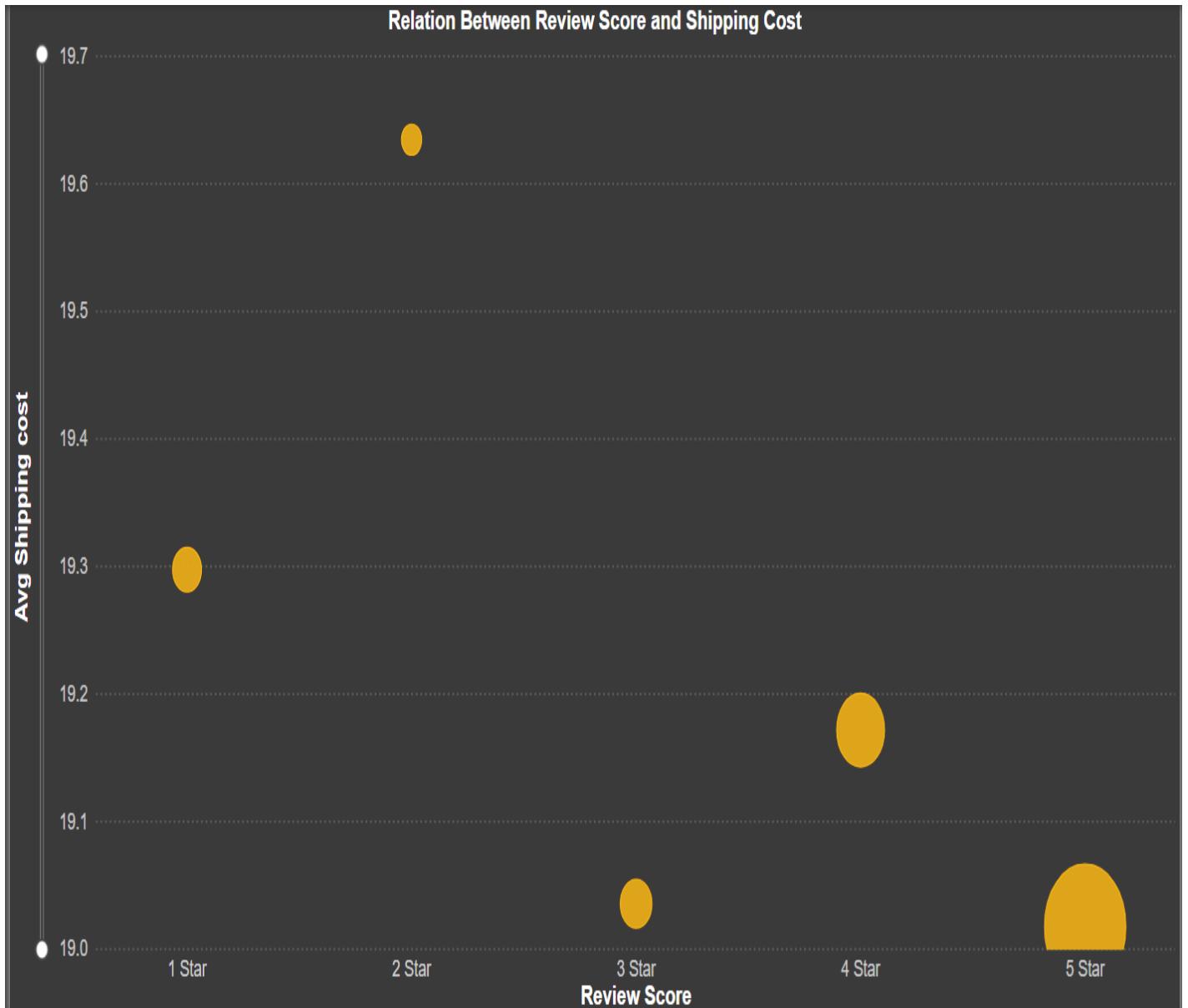
Total Shipping Cost	Avg Shipping Cost	% Shipping Cost	Avg Review Score
1.50M	19.13	0.18	4.33

Total 1.5M\$ Total of Shipping Cost With 19.3 Per
each order with total (18%) of Total Cost
With Average 4.3 Score

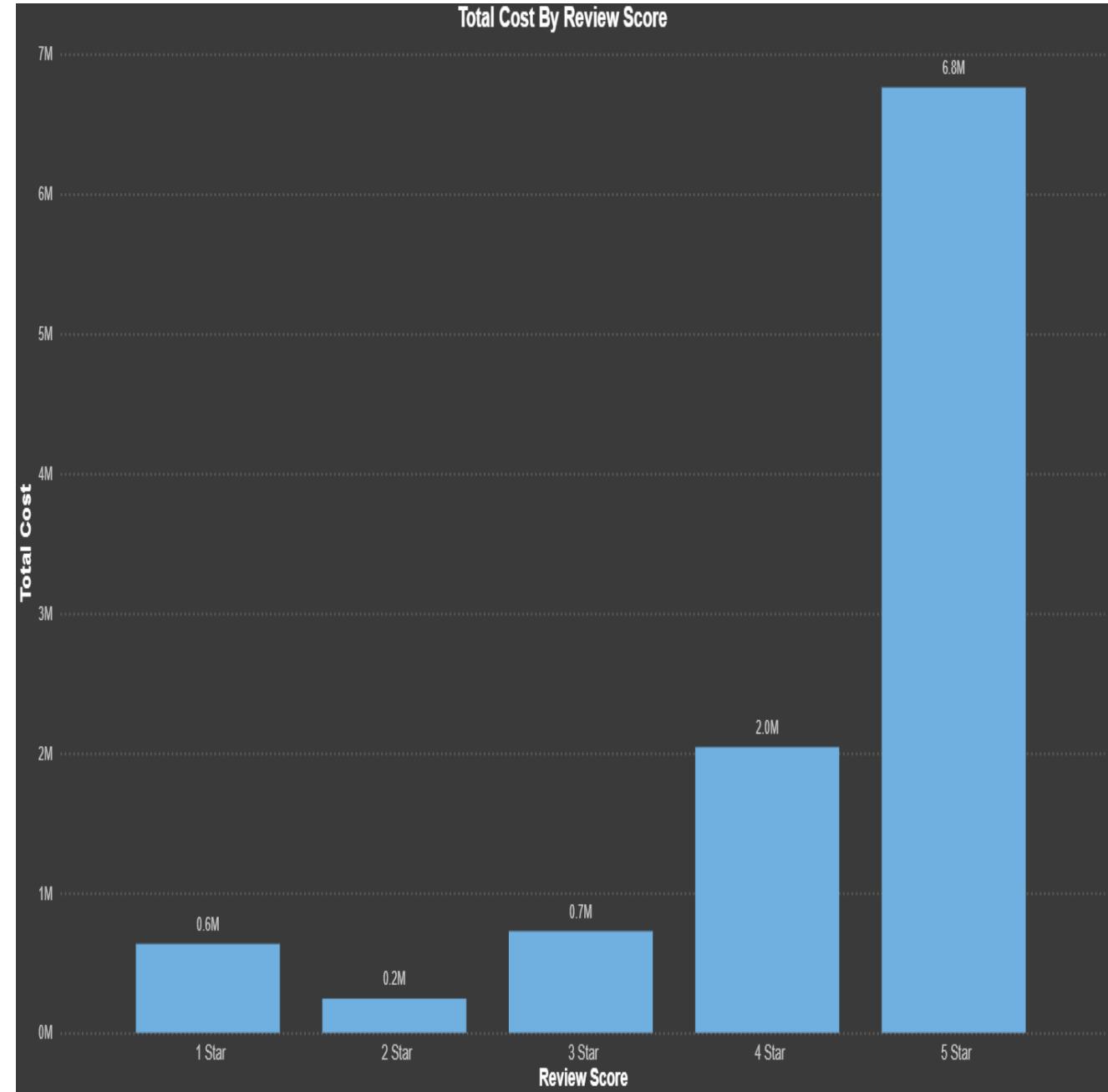
Long Level Waiting Days has highest Shipping Cost with avg 22.12\$ Per each Order with 31.4K Order

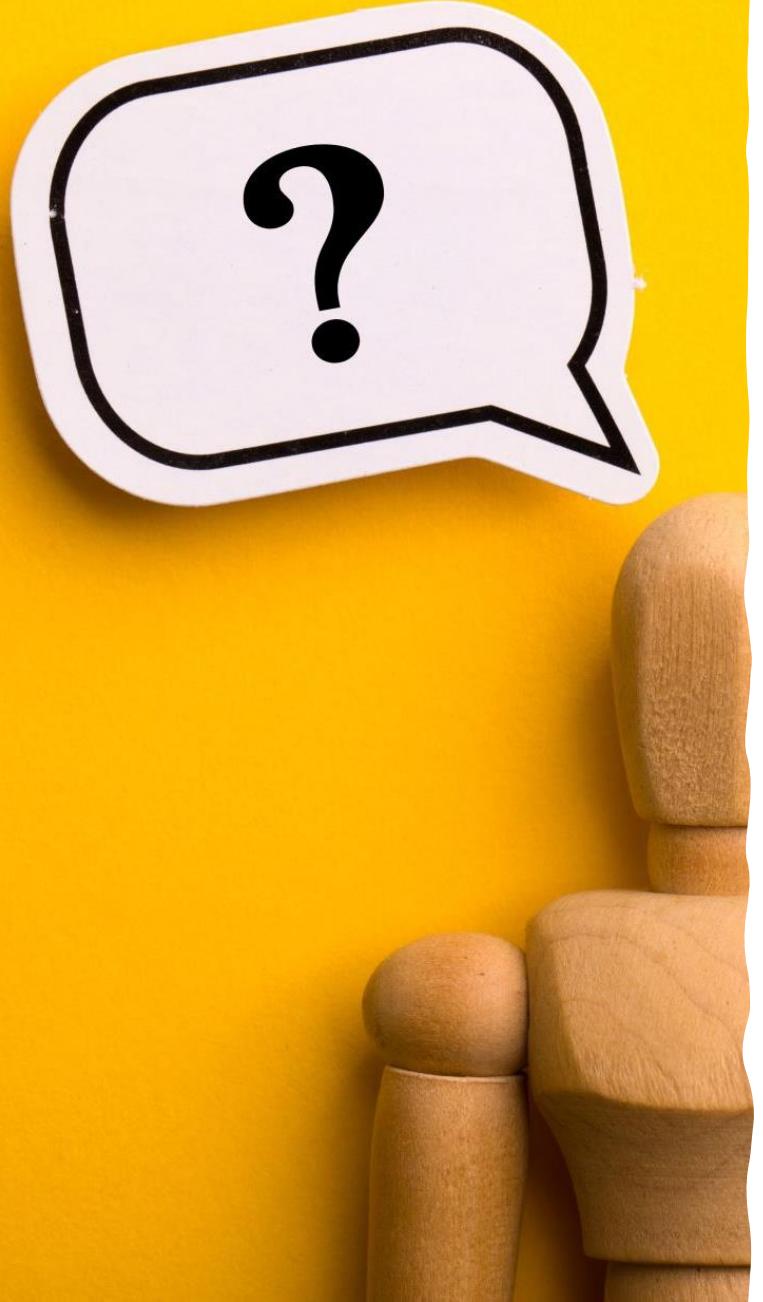


With Increasing Shipping Cost The Score Decrease



Nearly 47K Orders Got 5 Stars with total 6.8M\$ Total Cost with average 8.7 Days Waiting



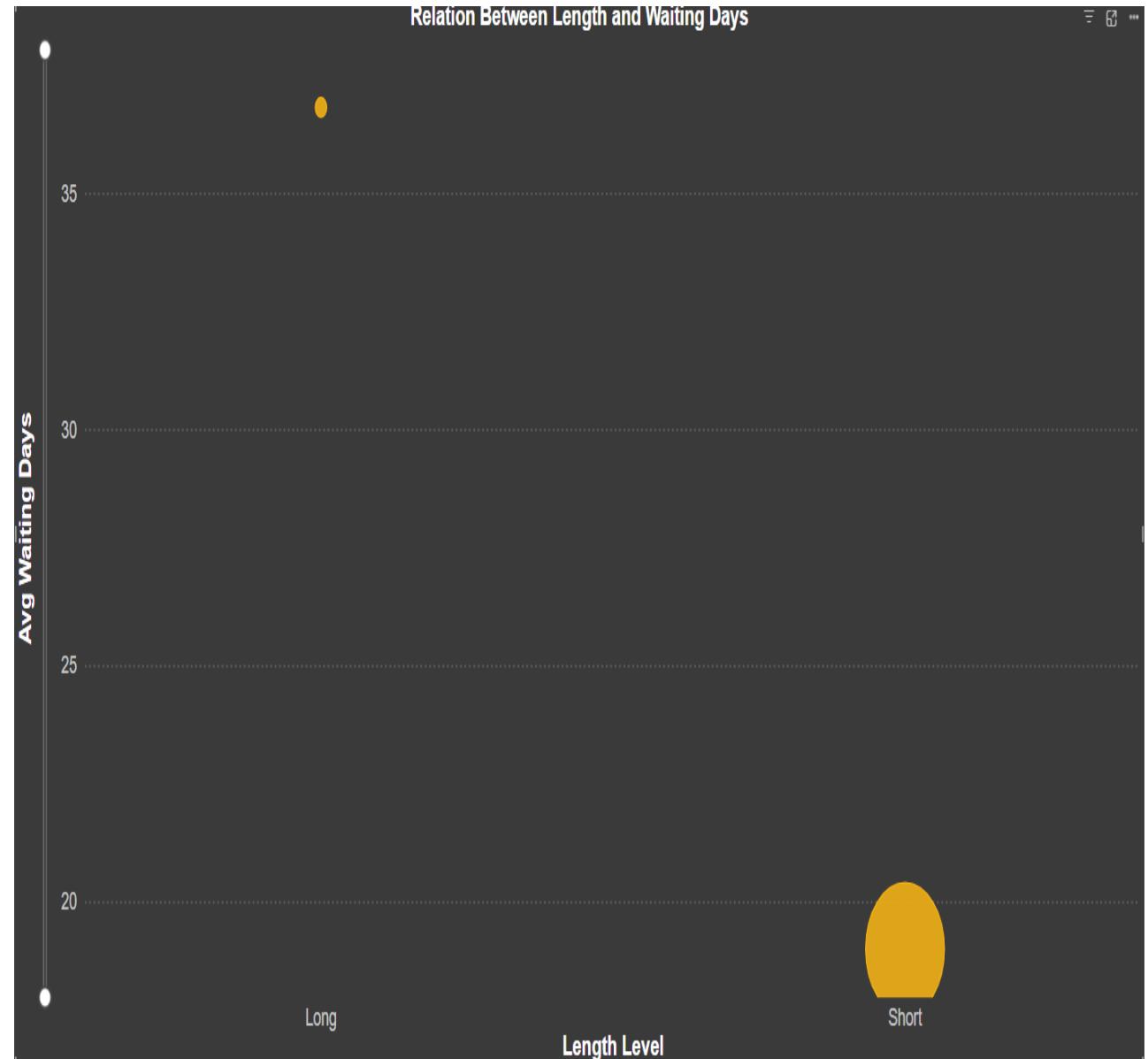


Does a Size of the Product directly
Correlate with a High Shipping Cost ?

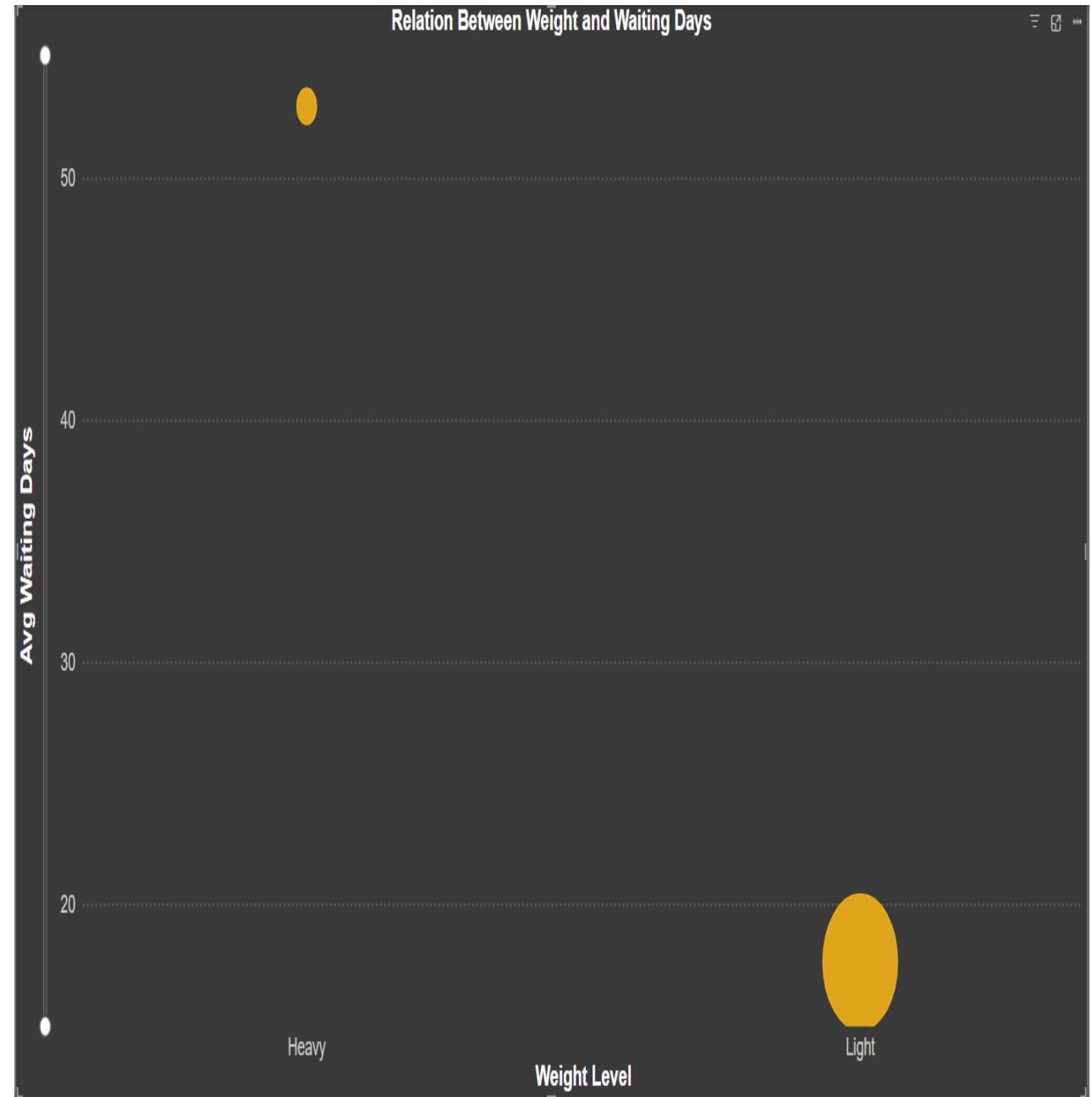
Total Orders	Avg Product Length (cm)	Avg Product Weight
78.316K	31.12	2.20K

Average Length of items is (31.12)cm with (2.2)kg Weight

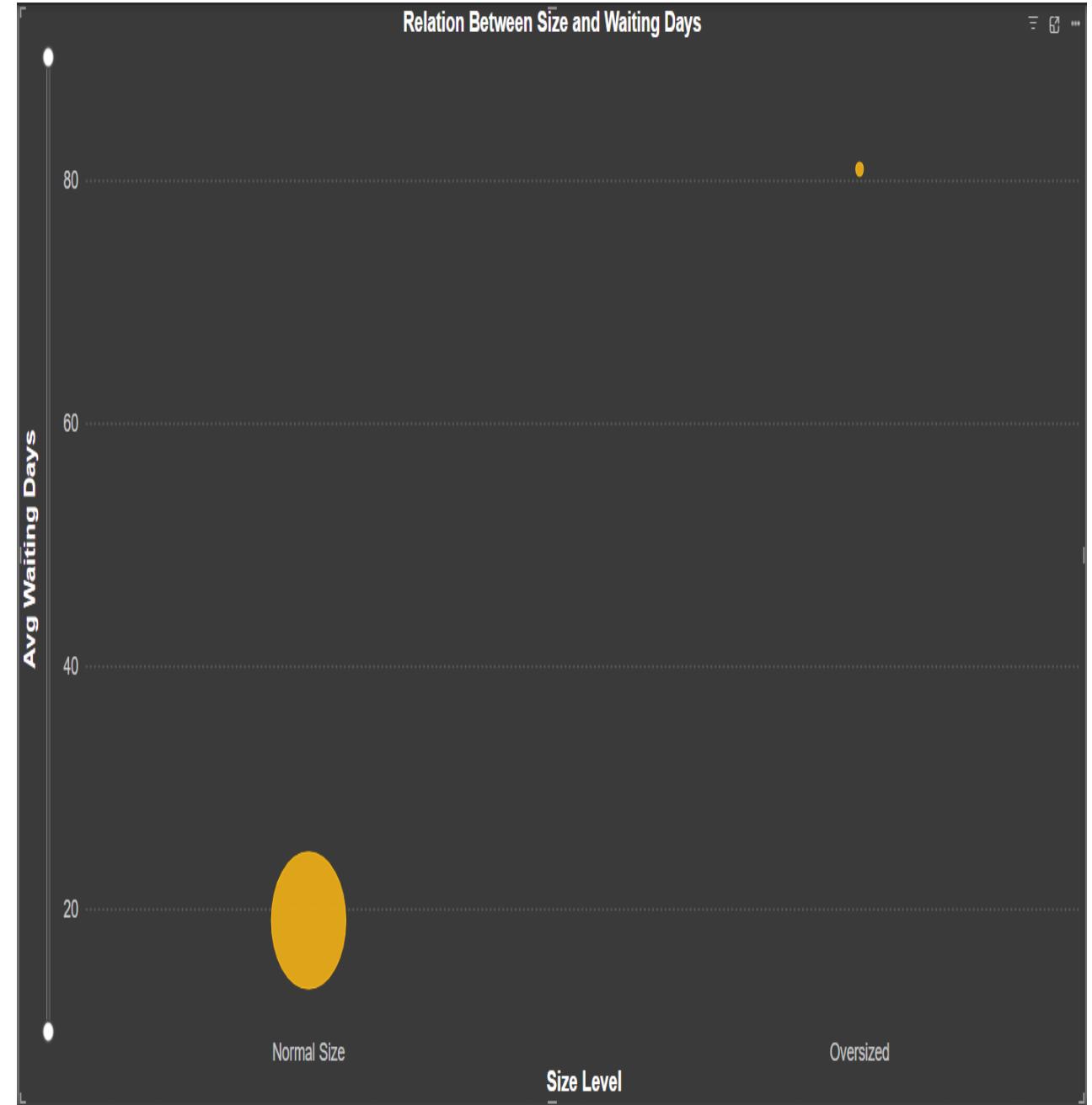
With increasing length of an Item the Shipping Cost increases and vice versa



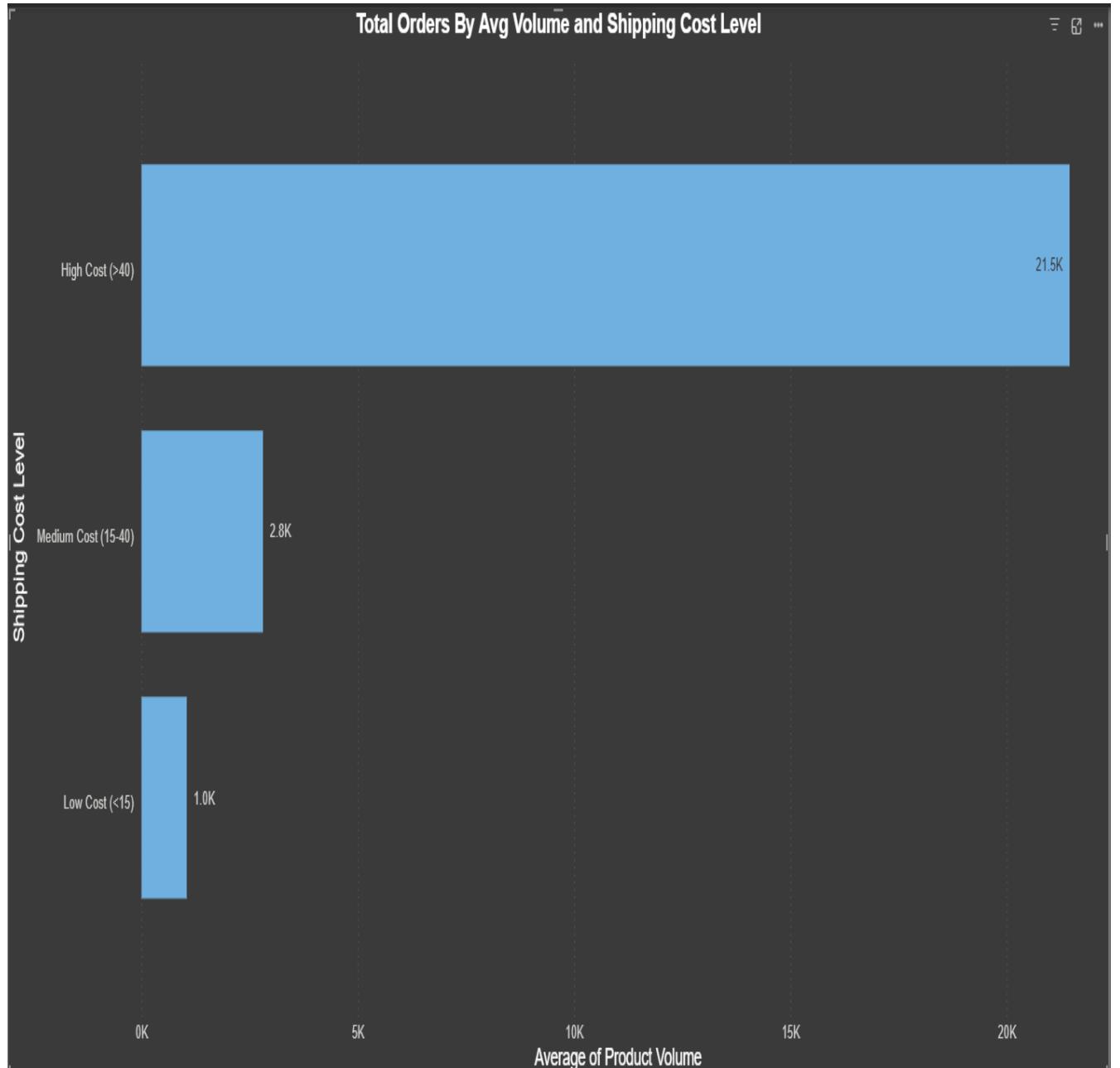
With increasing Weight of an Item the Shipping Cost increases and vice versa

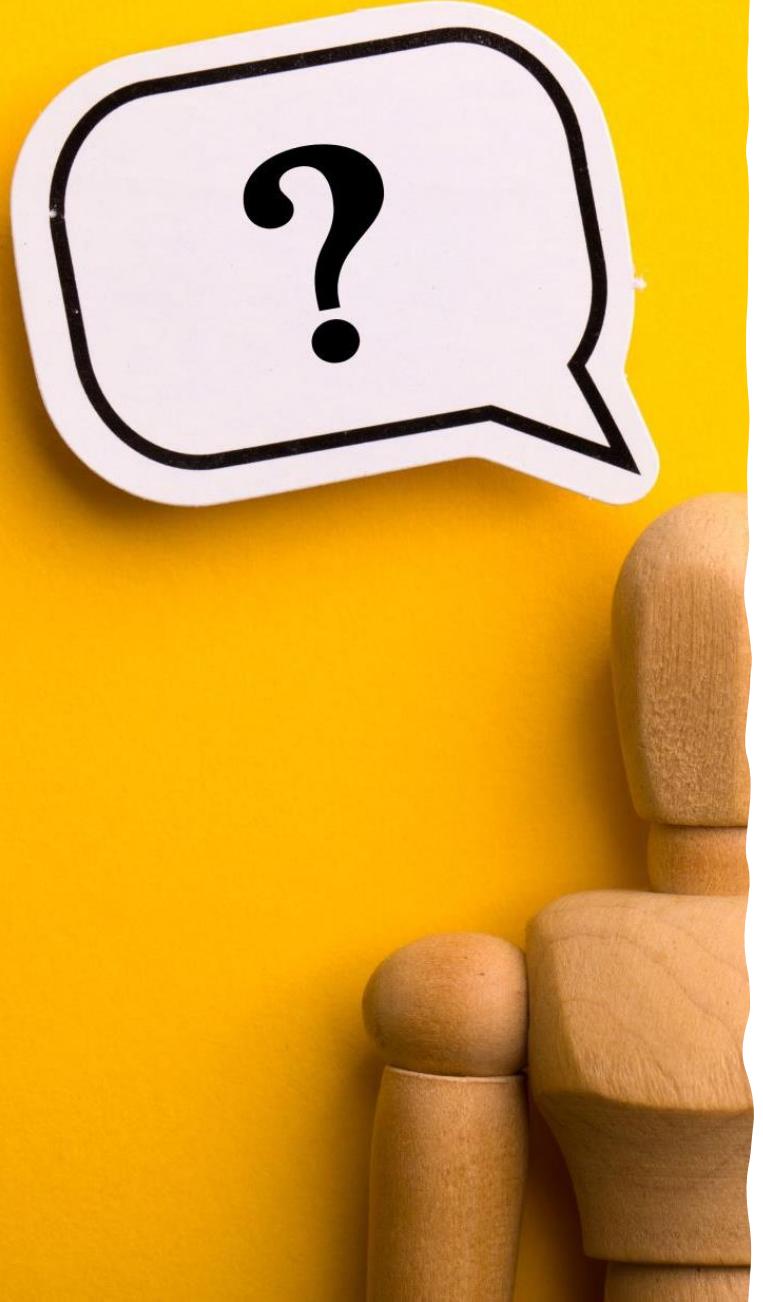


With increasing Size of an Item the Shipping Cost increases and vice versa



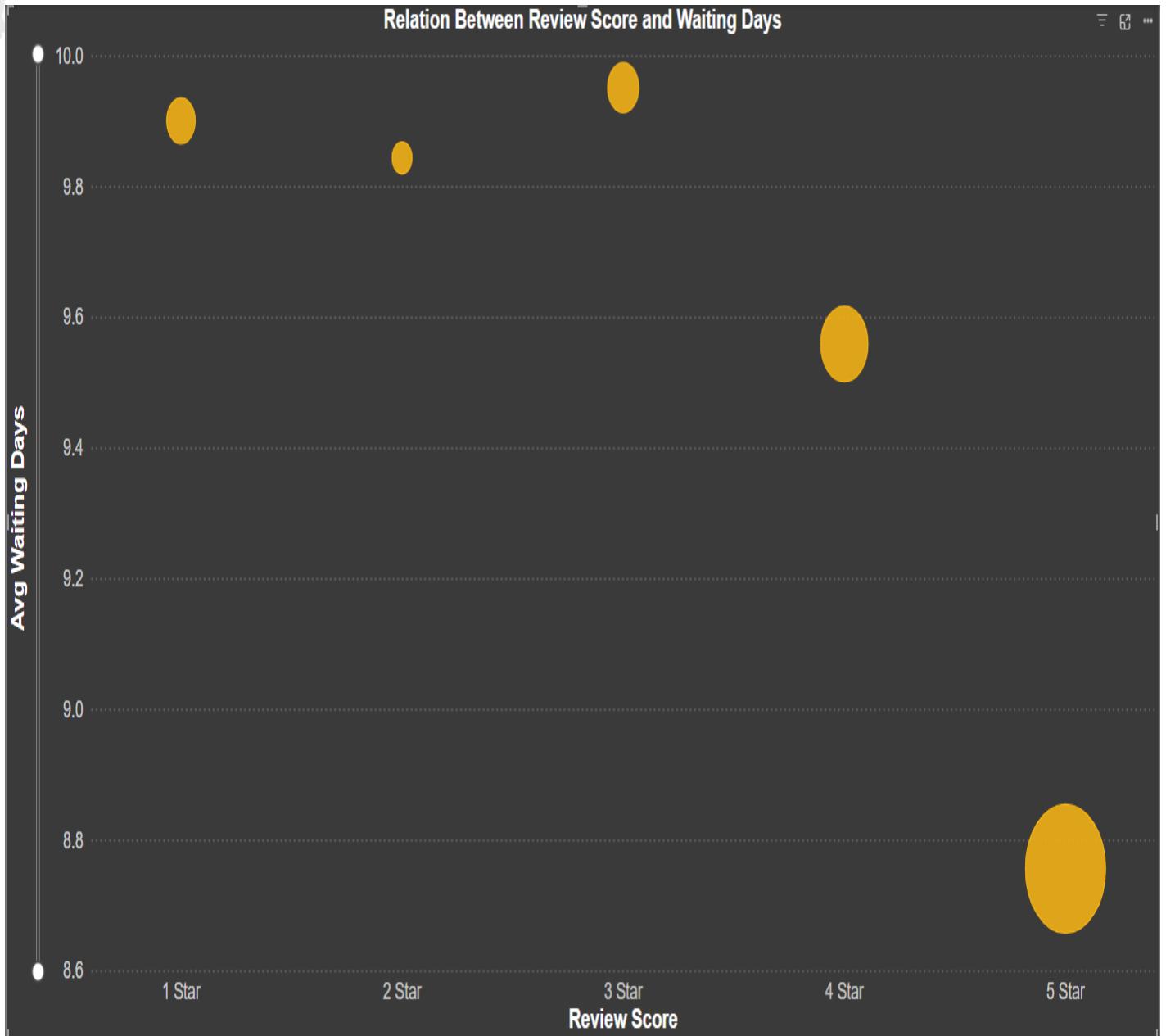
With increasing Volume of
an Item the Shipping Cost
increases and vice versa



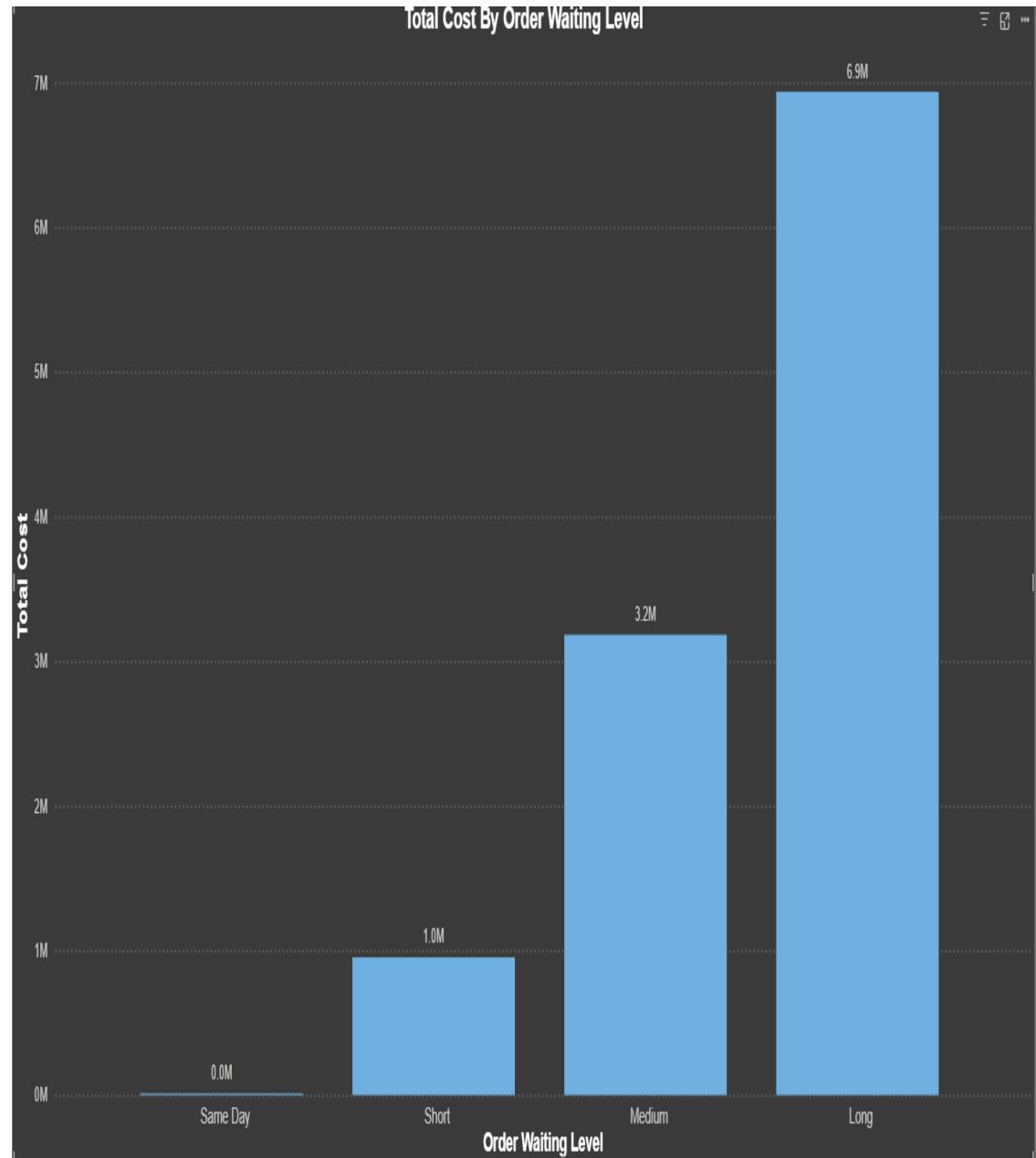


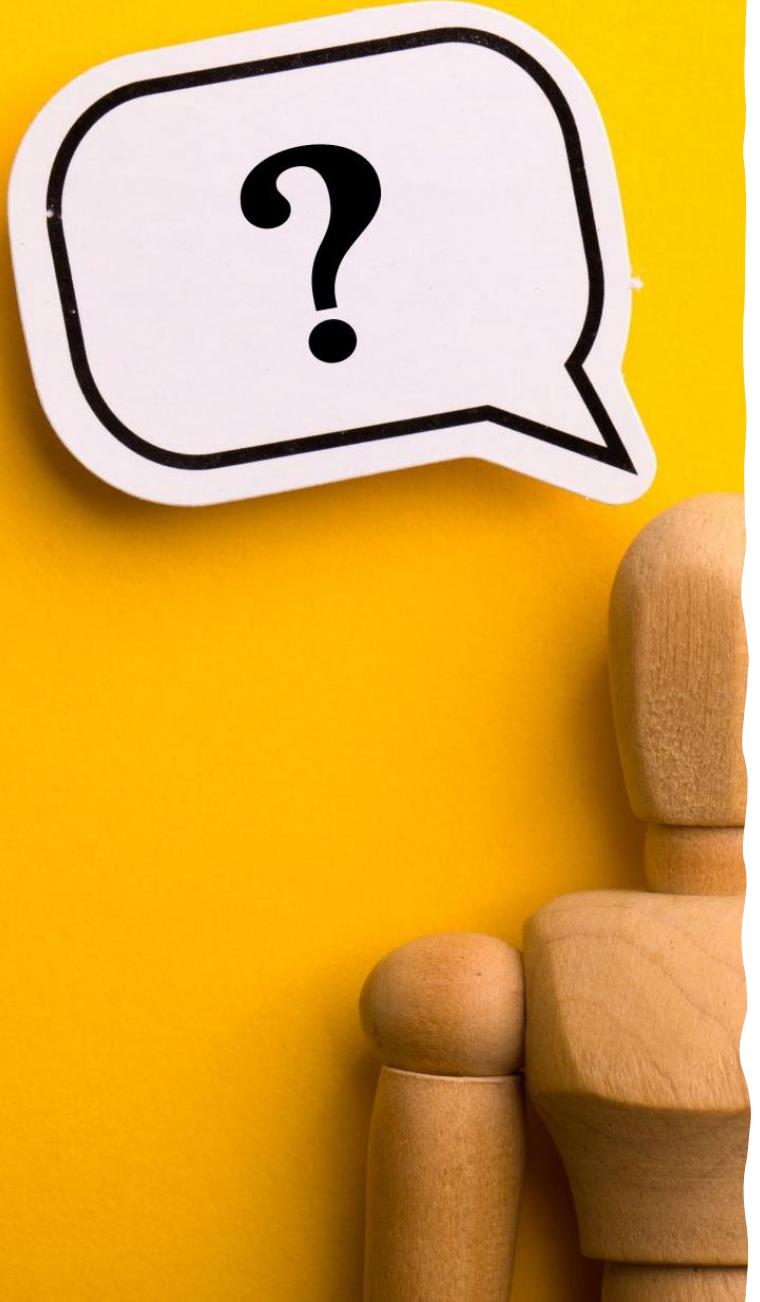
Does a long Waiting Days directly
Correlate with a Low Review Score?

With Increasing Waiting
Days The Score Decrease



Nearly 46K Orders from long waiting days with total 6.9M\$
Total Cost with average 12.2 Days Waiting



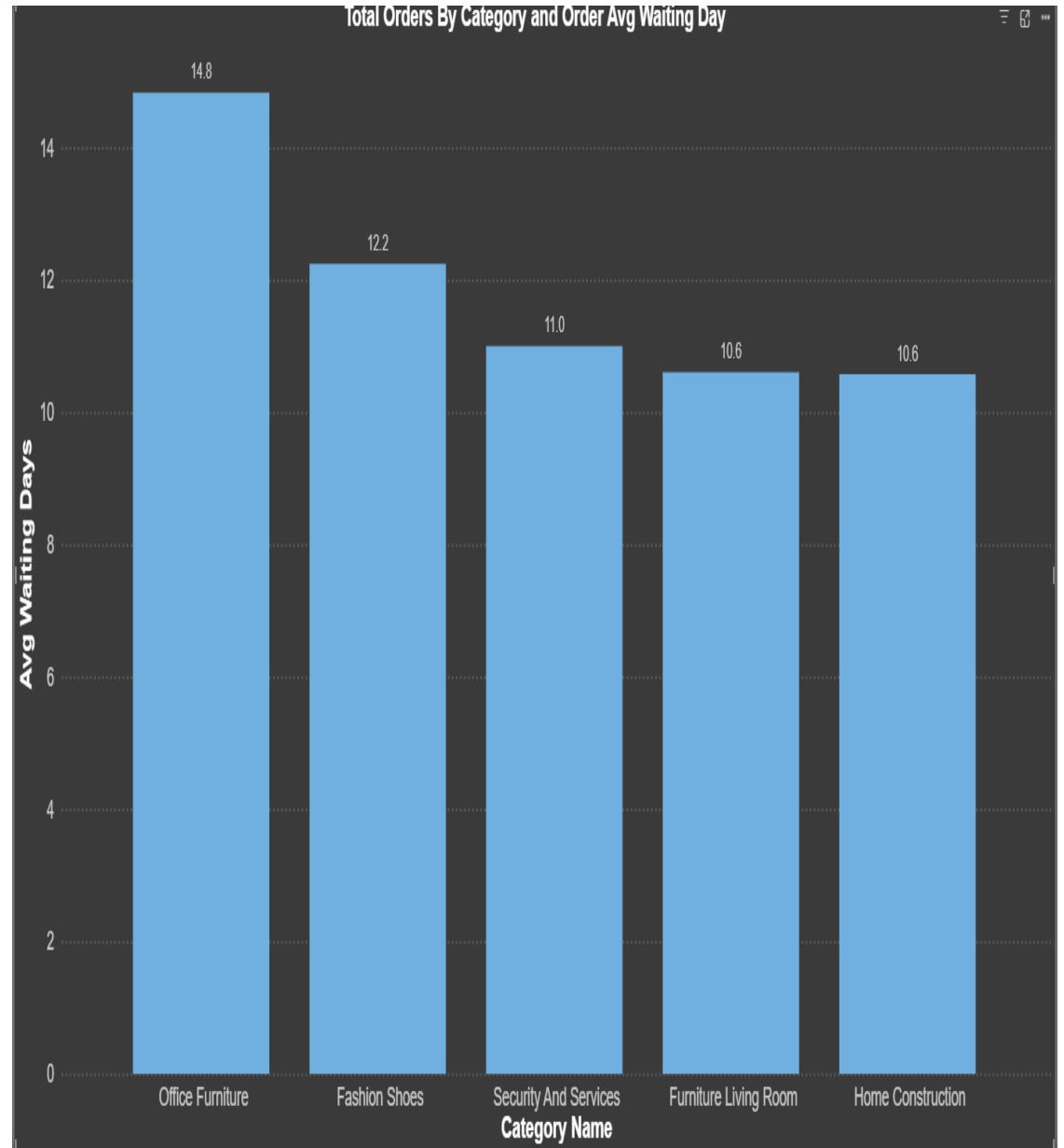


What Are The Reasons of Long Waiting Days ?

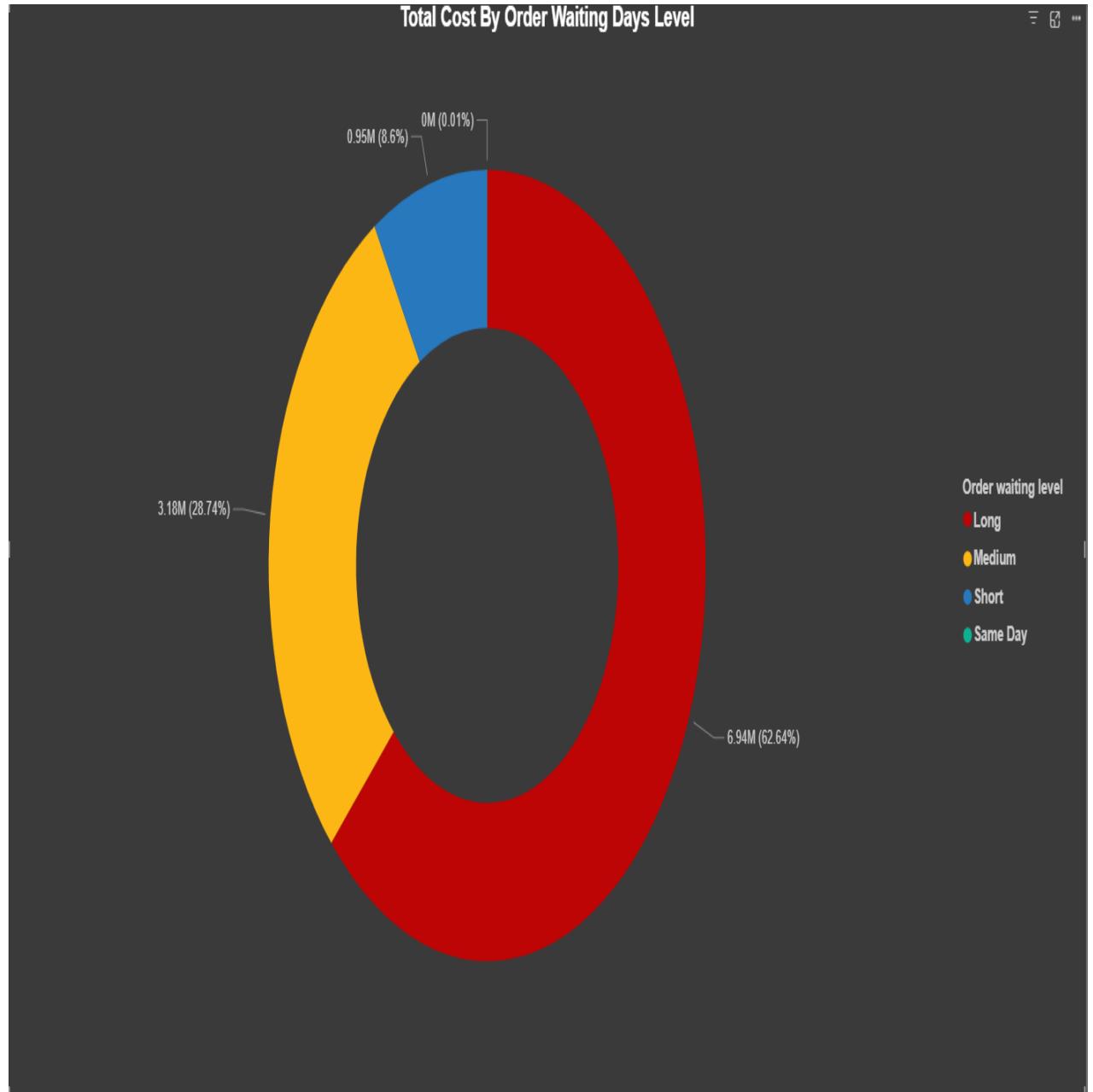
Avg Waiting Days	Approved Avg Days	Carrier Avg Days	Delivered Avg Days
9.14	0.29	1.99	8.58

while order approval takes only **0.3 days** and carrier hand-off takes **2 days**, the "Last-Mile" delivery to the customer averages **8.5 days**

The **Office Furniture** category is the top category with waiting long days with nearly (15) Days



With Increasing Item Volume
The Waiting Days Increase



Conclusion

Strategic Insights.

1

A significant **21%** of orders are cancelled, resulting in a **24%** loss of potential revenue. Data suggests the majority of these cancellations are triggered by high shipping costs and extended lead times.

2

São Paulo (SP) leads the nation in order volume but also in cancellations, with **7.6K** failed orders. Meanwhile, **Paraná (PR)** experiences the most severe delays, averaging **12.4 days** for delivery.

3

The Top 5 sellers contribute to a combined **1,492** cancellations, with an average delivery delay of **12.5 days**—significantly higher than the platform average.

Strategic Insights.

4

Shipping costs represent roughly **18%** of the total transaction value (Avg. **\$19.13** per order). There is a direct inverse correlation between shipping costs and customer sentiment; as logistics fees rise, review scores plummet.

5

Product volume is the primary driver of shipping expense. Approximately **4,000** "Heavy/Bulky" items are currently suffering from prohibitively high shipping costs.

6

With an average wait time of **9 days**, delivery speed is the primary factor in negative reviews. Extended delays directly erode brand trust and discourage recurring revenue.

Recommendations

Recommendations Summary

1

Accelerate fulfillment workflows and optimize shipping price structures to prevent cart abandonment and post-purchase cancellations.

2

Prioritize logistics infrastructure and carrier partnerships in **SP** and **PR** to reduce regional shipping friction and enhance service reliability.

3

Implement an "Efficiency Audit" for the top 5 underperforming sellers to streamline their internal dispatch processes.

Recommendations Summary

4

Implement localized shipping subsidies or flat-rate options to boost review scores, as high-rated orders correlate with the highest customer lifetime value (LTV).

5

Introduce targeted "Free Shipping" incentives or tiered discounts for heavy items for high-frequency customers to move bulky inventory more efficiently.

6

Set aggressive "Last-Mile" delivery targets to reduce waiting days, ensuring higher review scores and improved customer retention.

Thank You!



- **Omer Metwally | Aspiring Data Analyst**
- Tools Used:
 - Excel
 - Python
 - Exploratory Data Analysis (EDA)
 - Advanced Data Wrangling
 - Pandas & NumPy
 - SQL
 - Data Cleaning & Joins
 - "Unknown" Placeholder Handling
 - Power Bi
 - Power Query (ETL)
 - DAX
 - AI-Powered Visuals, Interactive Reporting Tools
- Last Updated: **10.th February 2026**