ECBS 5146: Term project 2

Team 1:

Kabdula Asset

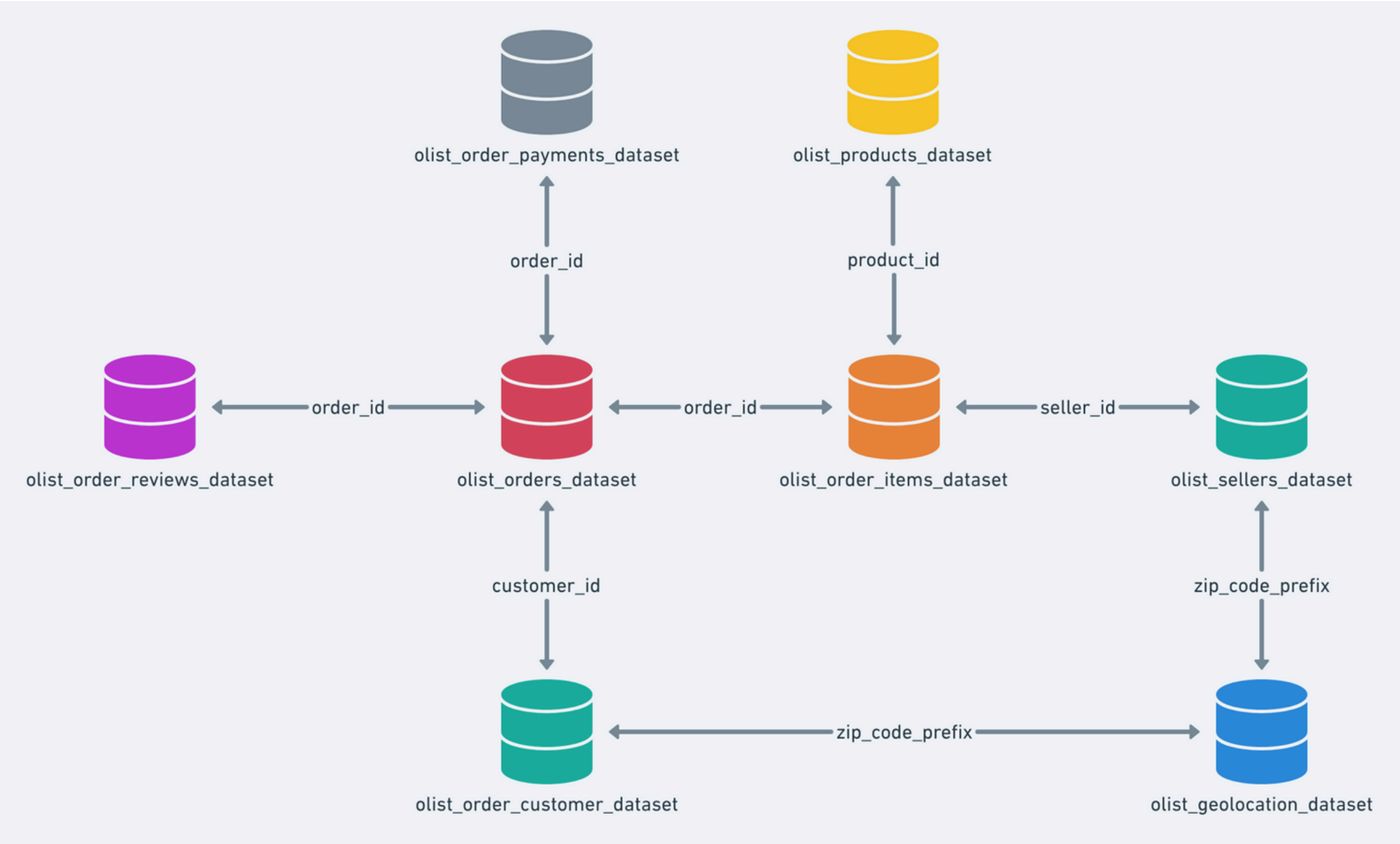
Sundu Mehmet

Toktargazy Ayazhan

Zheng Ying

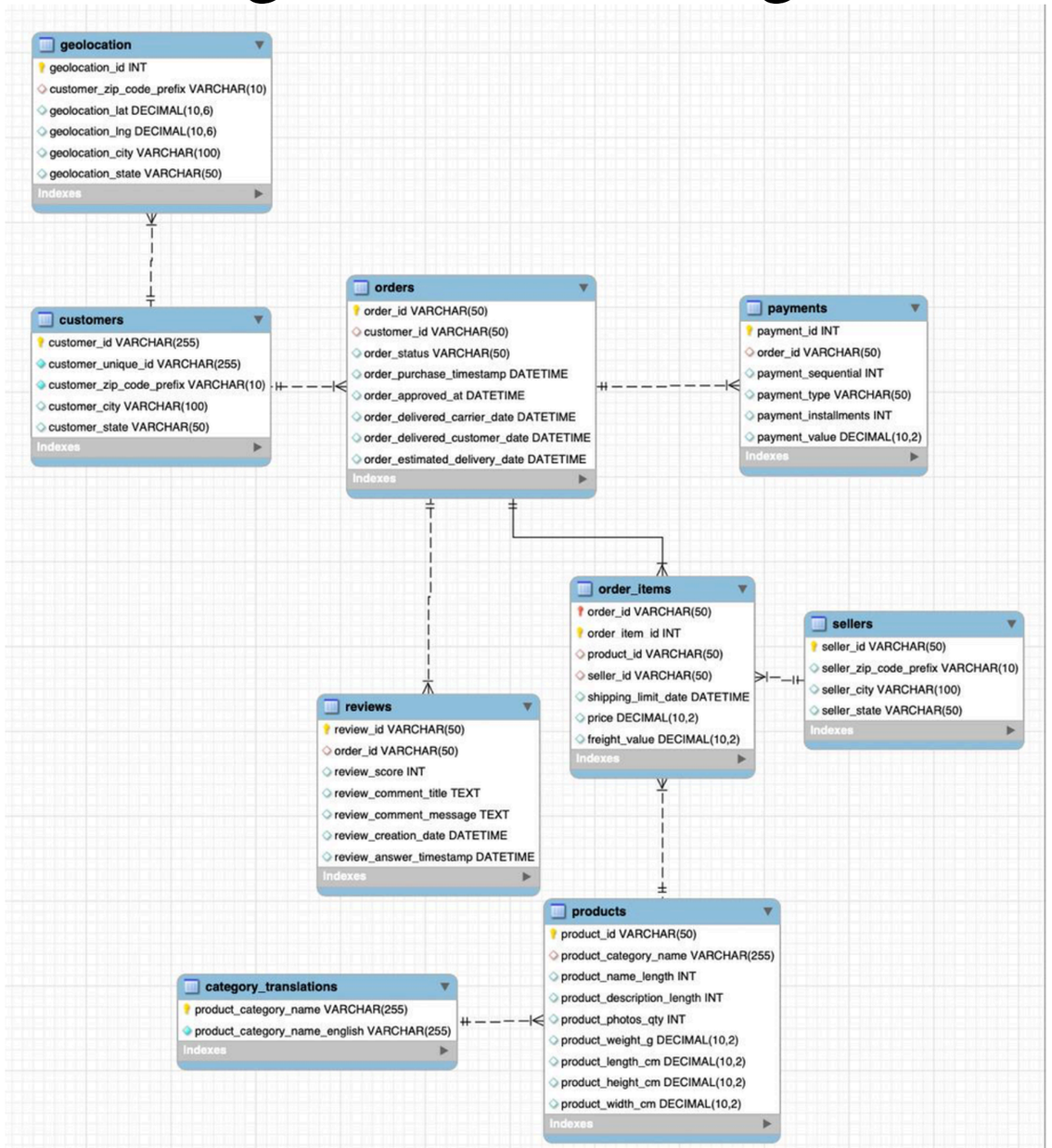
Contents

Image 1. Data Schema



Source: Kaggle

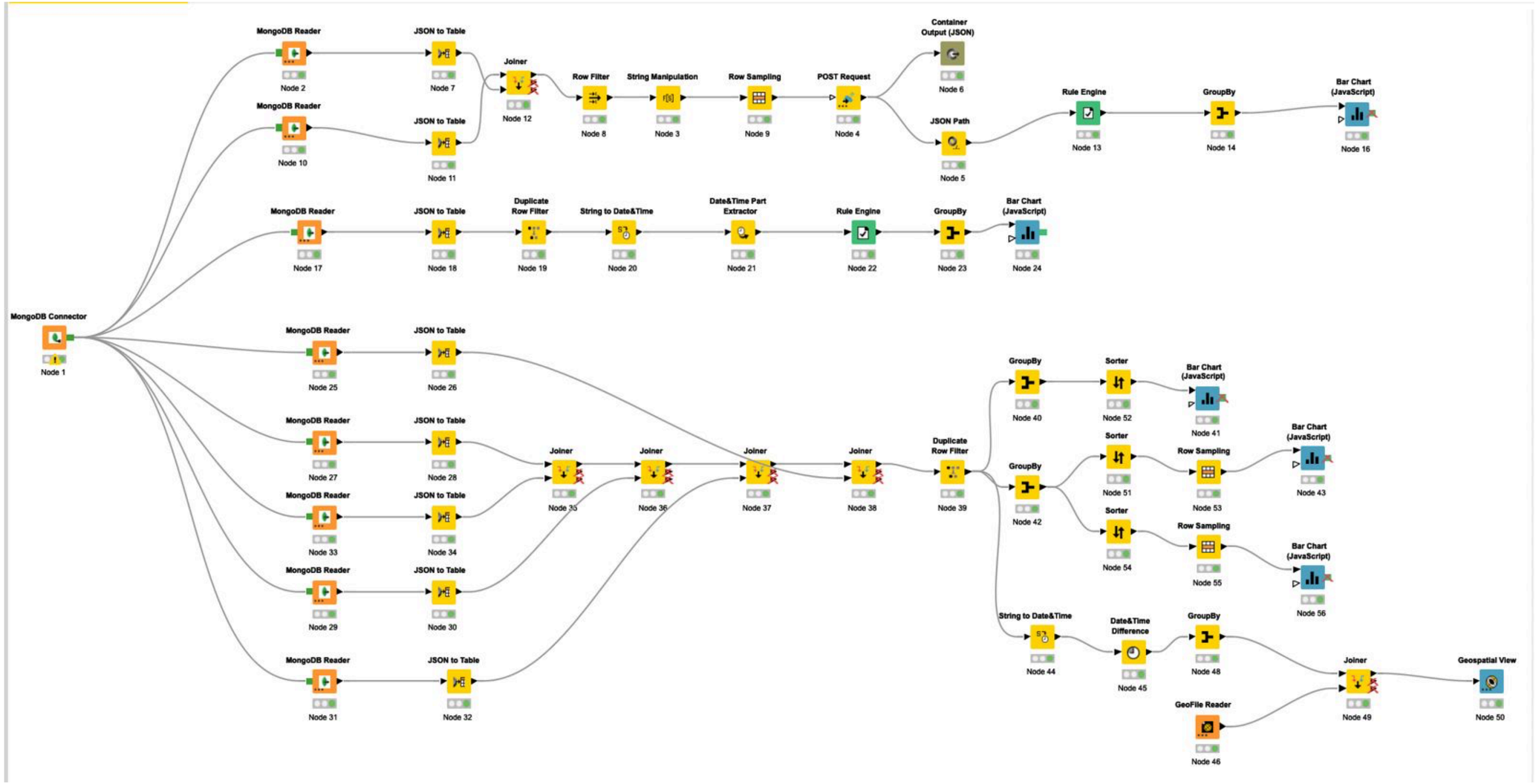
Image 2. EER diagram



Research Questions

- Q1. How does the delivery performance (on-time vs. delayed) affect customer sentiment in Brazilian e-commerce?
- Q2. Is there a clear seasonal trend in order volume? How to promote off-season sales?
- Q3. Which categories perform best and worst nationwide?
- Q4. Are there noticeable regional differences in the delivery periods?

Image 3. ETL pipeline in Knime



Solution & Technical Choices

Q1: Delivery Performance vs. Customer Sentiment

- Data Prep: Cleaned ~41,000 observations, sampled 3,134 (98% confidence, 2% margin).
- Sentiment Analysis: Used Google Cloud NL API for sentiment scores.
- Analysis: Categorized delivery performance and analyzed its relation to sentiment.

Visualization of the Results: Q1

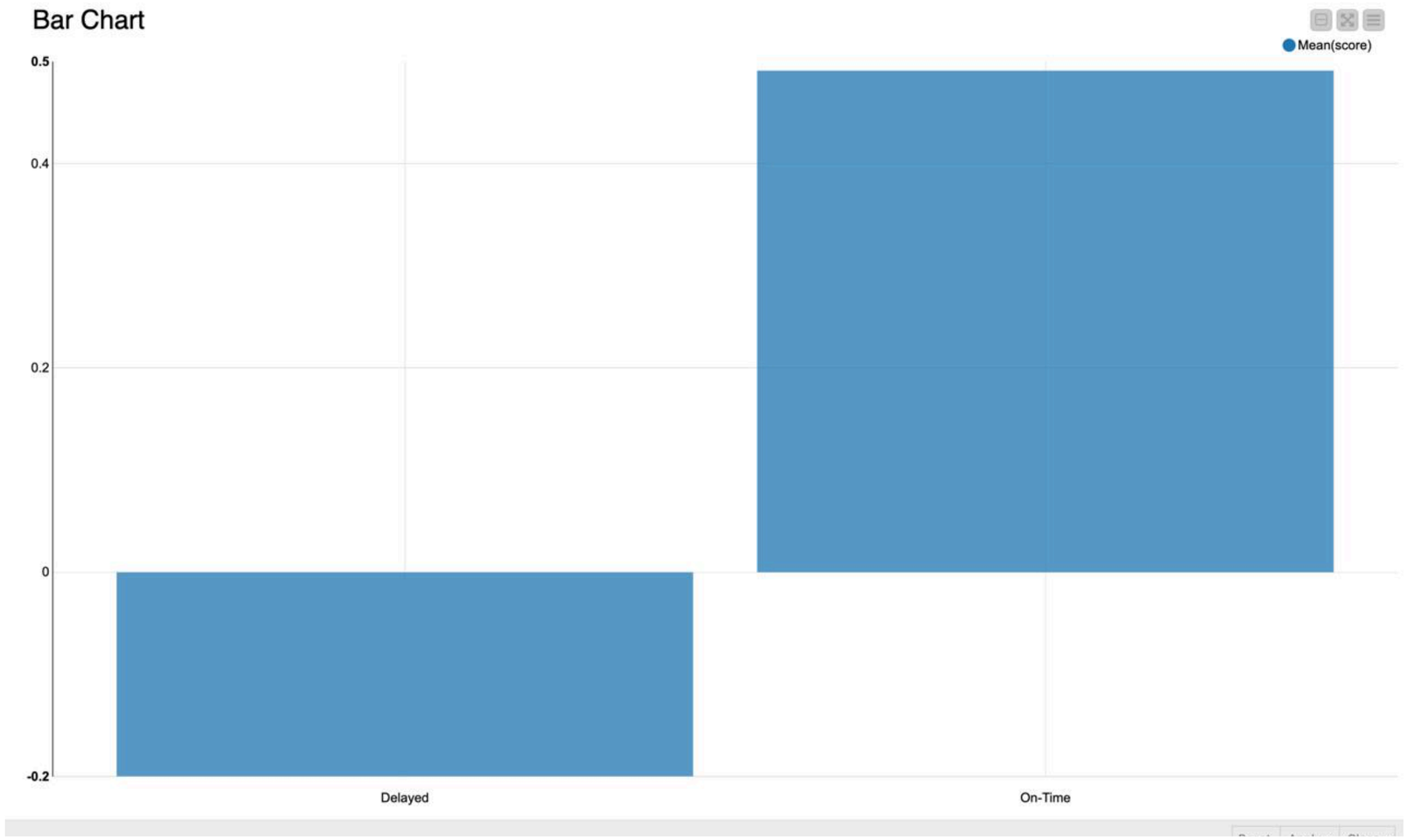


Chart 1. Correlation between sentiment and delivery categorization

Solution & Technical Choices

Q2: Seasonal Order Volume

- Data Prep: Removed duplicates, formatted dates, derived seasons.
- Analysis: Grouped order volumes by season.

Visualization of the Results: Q2

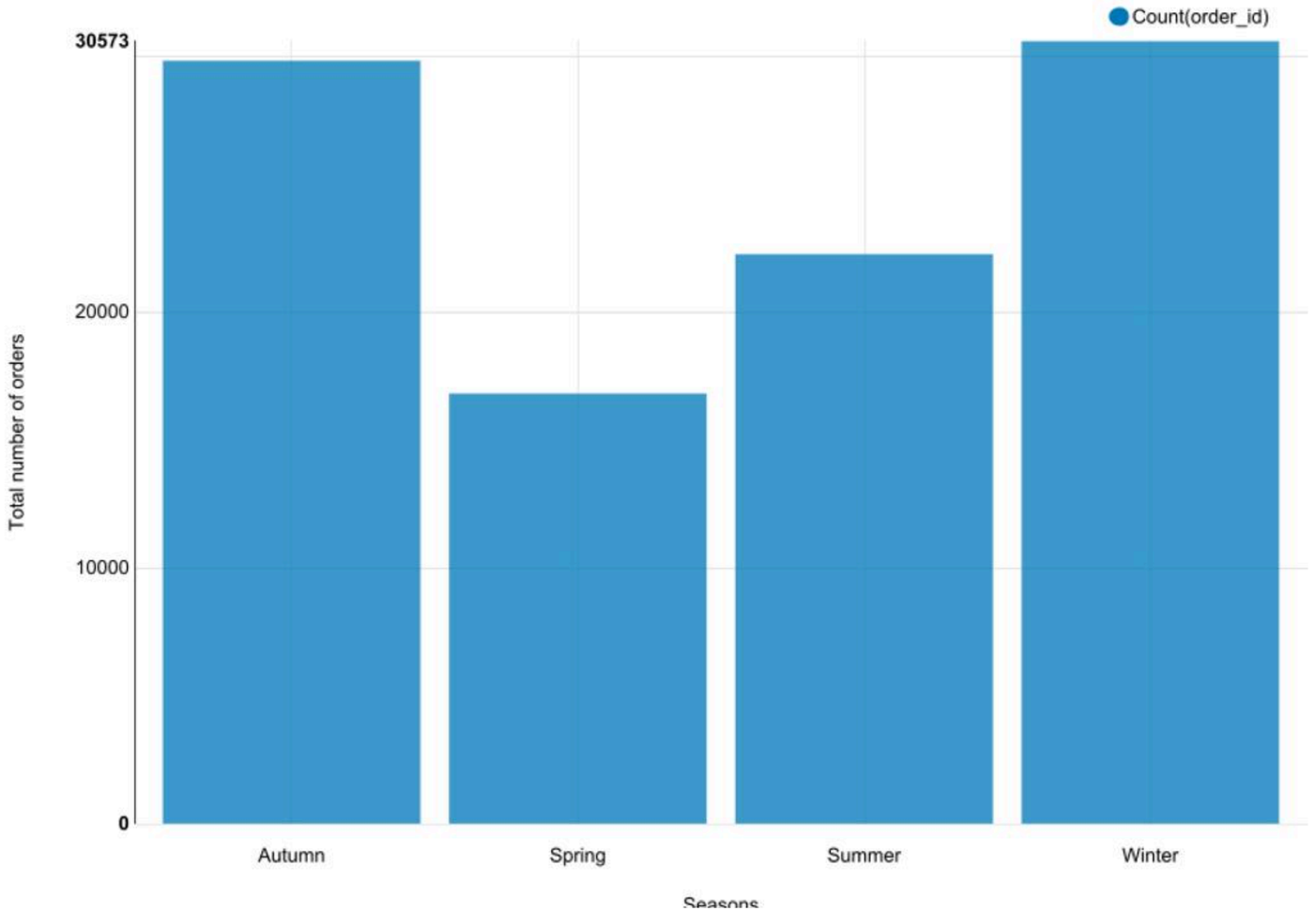


Chart 2. Seasonality character of orders

Solution & Technical Choices

Q3: Product Category Popularity

- Data Prep: Cleaned data.
- Analysis: Identified 10 most and least popular product categories.

Visualization of the Results: Q3

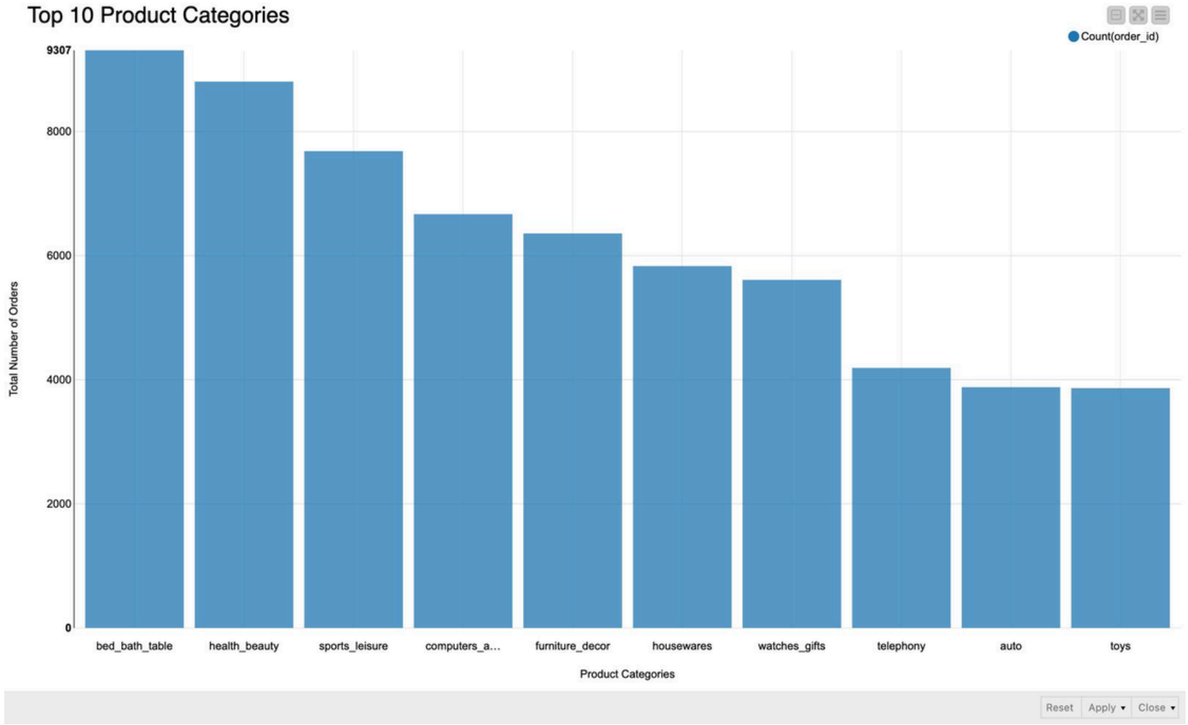


Chart 3a. Top-10 popular product categories

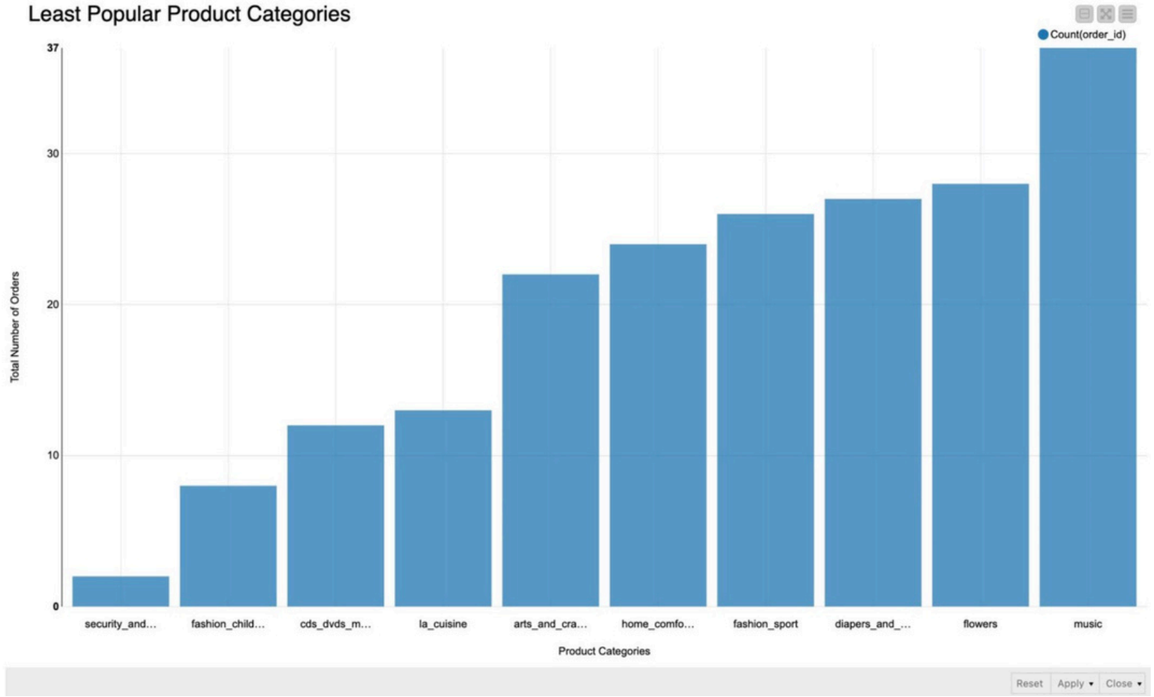


Chart 3b. The least popular 10 product categories

Solution & Technical Choices

Q4: Delivery Time & Geospatial Trends

- Data Prep: Calculated delivery times, joined geospatial data.
- Geospatial Analysis: Analyzed trends using GeoJSON files.
- Analysis: Grouped orders by state to identify high-volume regions.

Visualization of the Results: Q4

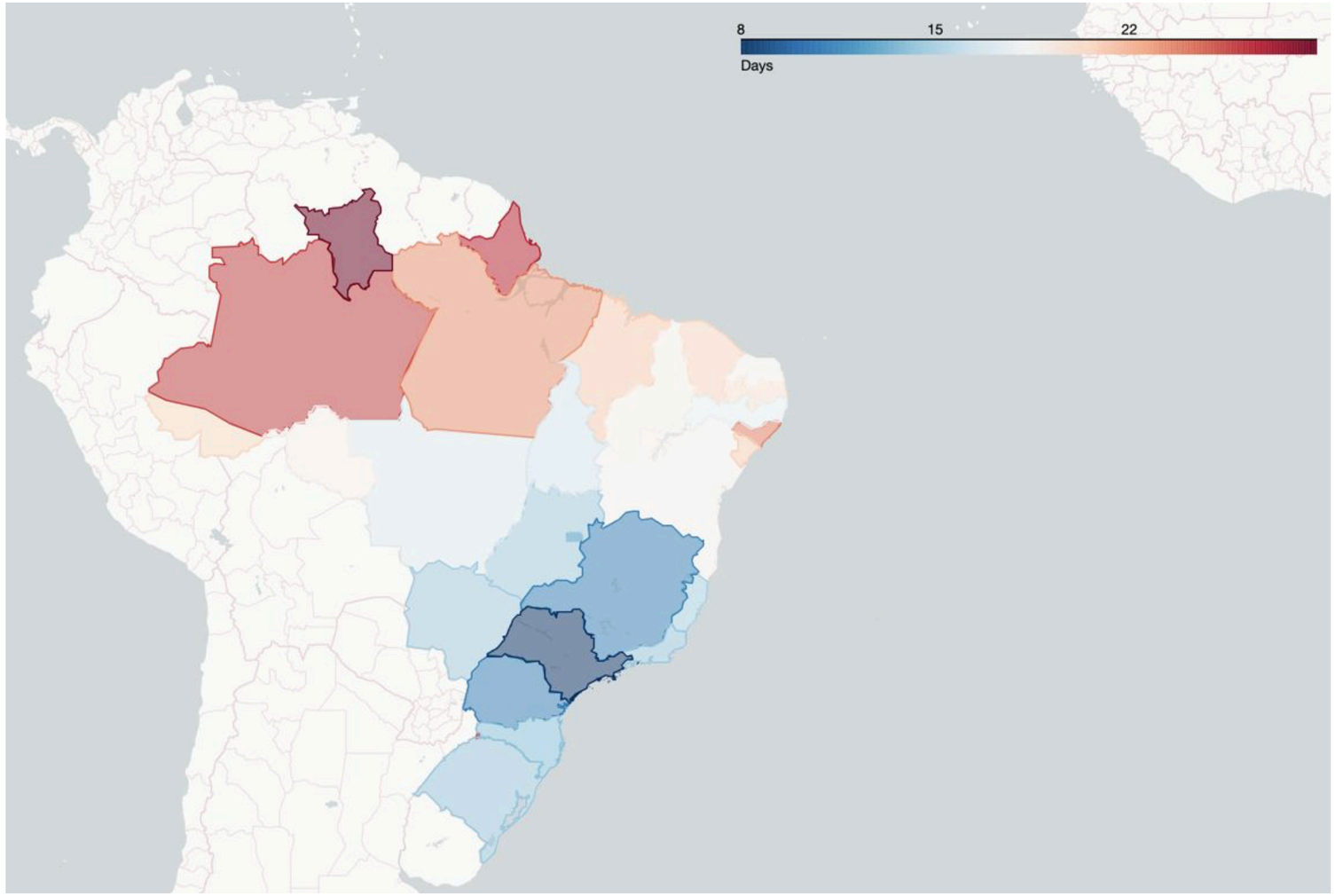


Chart 4. Heatmap of delivery periods

Conclusion and Future Work

Research Question	Aspect	Key Insight	Actionable Takeaway
Q1	**Timely Deliveries**	• on-time deliveries boost positive sentiment; • delays hurt satisfaction.	Focus on improving logistics to ensure timely deliveries.
Q2	**Seasonal Sales Trends**	• Winter/Autumn drive sales; • Spring needs creative strategies.	Develop targeted promotions for Spring to boost sales.
Q3	**Product Performance**	• Least popular categories need further analysis: drop or optimize.	Conduct deeper analysis to decide on elimination or improvement.
Q4	**Regional Delivery Insights**	• Northern regions face logistical delays, requiring improvement.	Address logistical challenges to enhance regional performance.

- Expand geospatial analytics for better logistics.
- Analyze time trends (2016–2018 vs. recent data).
- Integrate external data (weather, social media) for deeper insights.