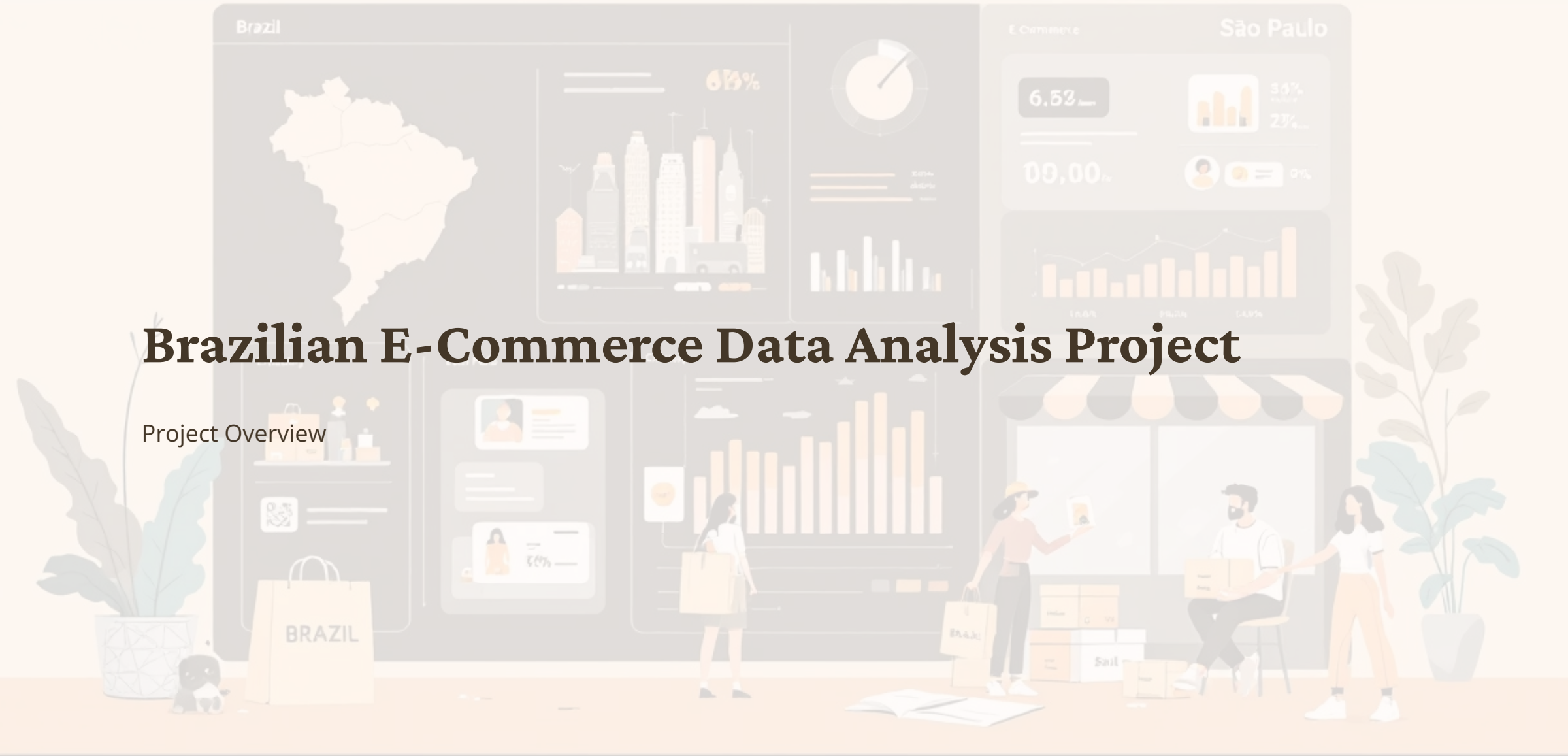


Brazilian E-Commerce Data Analysis Project

Project Overview

Project Overview

This project presents a comprehensive Exploratory Data Analysis (EDA) of the Brazilian E-Commerce Public Dataset by Olist, covering marketplace transactions from 2016 to 2018. The analysis explores customer behaviour, seller performance, product dynamics, and delivery logistics across Brazil's multi-vendor e-commerce platform.

Dataset Description

The dataset comprises 9 interconnected tables containing:

- Orders: 99,441 records with order lifecycle information
- Customers: Customer demographics and location data
- Sellers: Seller information and geographic distribution
- Products: Product catalog with 32,951 items across 73 categories
- Order Items: Line-item details for each order
- Payments: Payment methods and transaction values
- Reviews: Customer ratings and feedback
- Geolocation: Geographic coordinates for Brazilian ZIP codes

Methodology — Data Cleaning

Converted timestamp columns to datetime format

- Removed invalid orders with illogical date sequences
- Filtered out zero-price items and undefined payment types
- Removed duplicate reviews and handled missing product categories
- Standardised city names and removed accents for consistency
- Imputed missing product dimensions using median values

Methodology — Data Wrangling

Merged 9 datasets into a unified commerce dataframe

- Joined customer and geolocation data for spatial analysis
- Created temporal features (day, month, day of week)
- Engineered delivery delay metrics

1

Unified dataset

Single dataframe to power analysis and visualizations

2

Spatial features

Enable mapping and regional insights

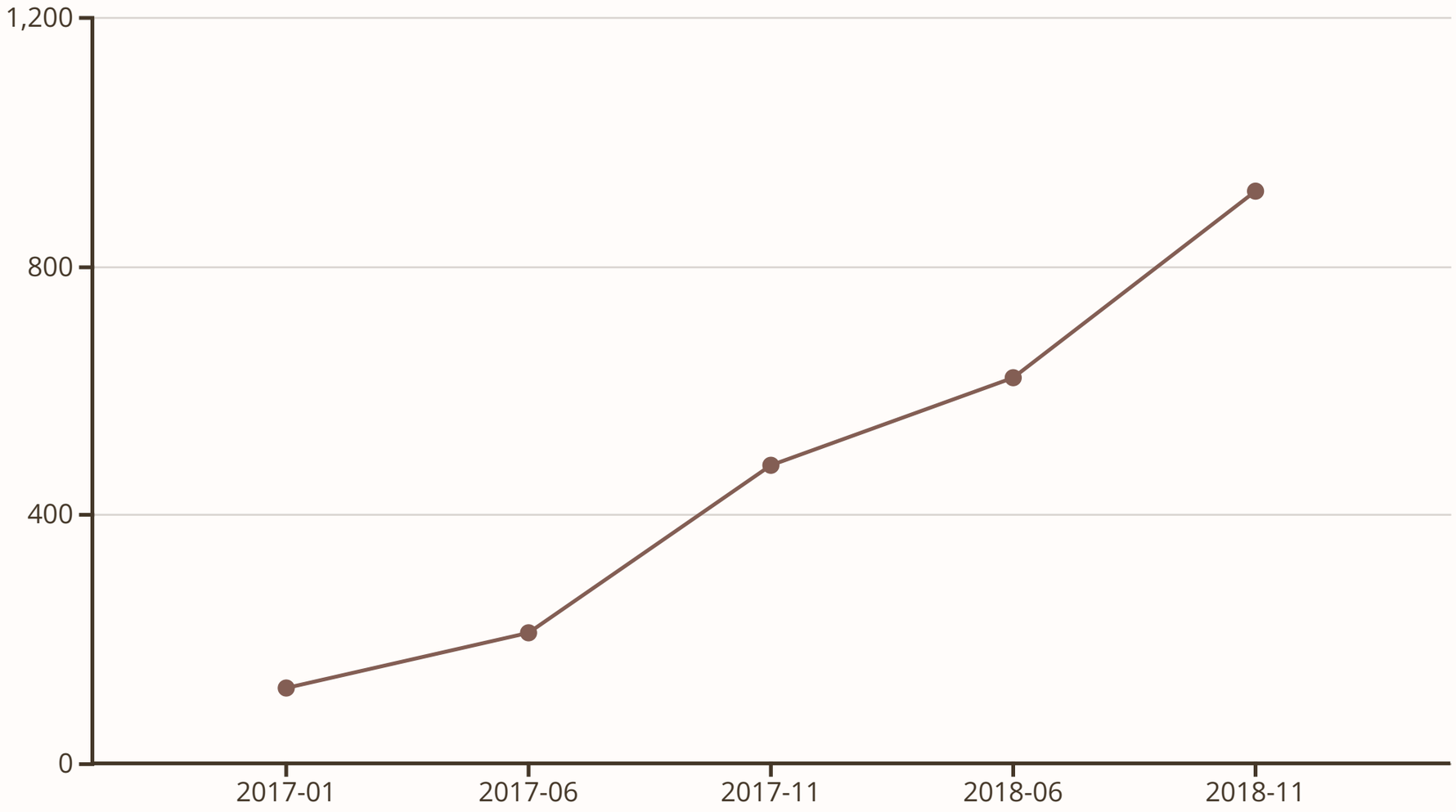
3

Temporal features

Support time-series and seasonal analysis

Key Findings — Order Performance

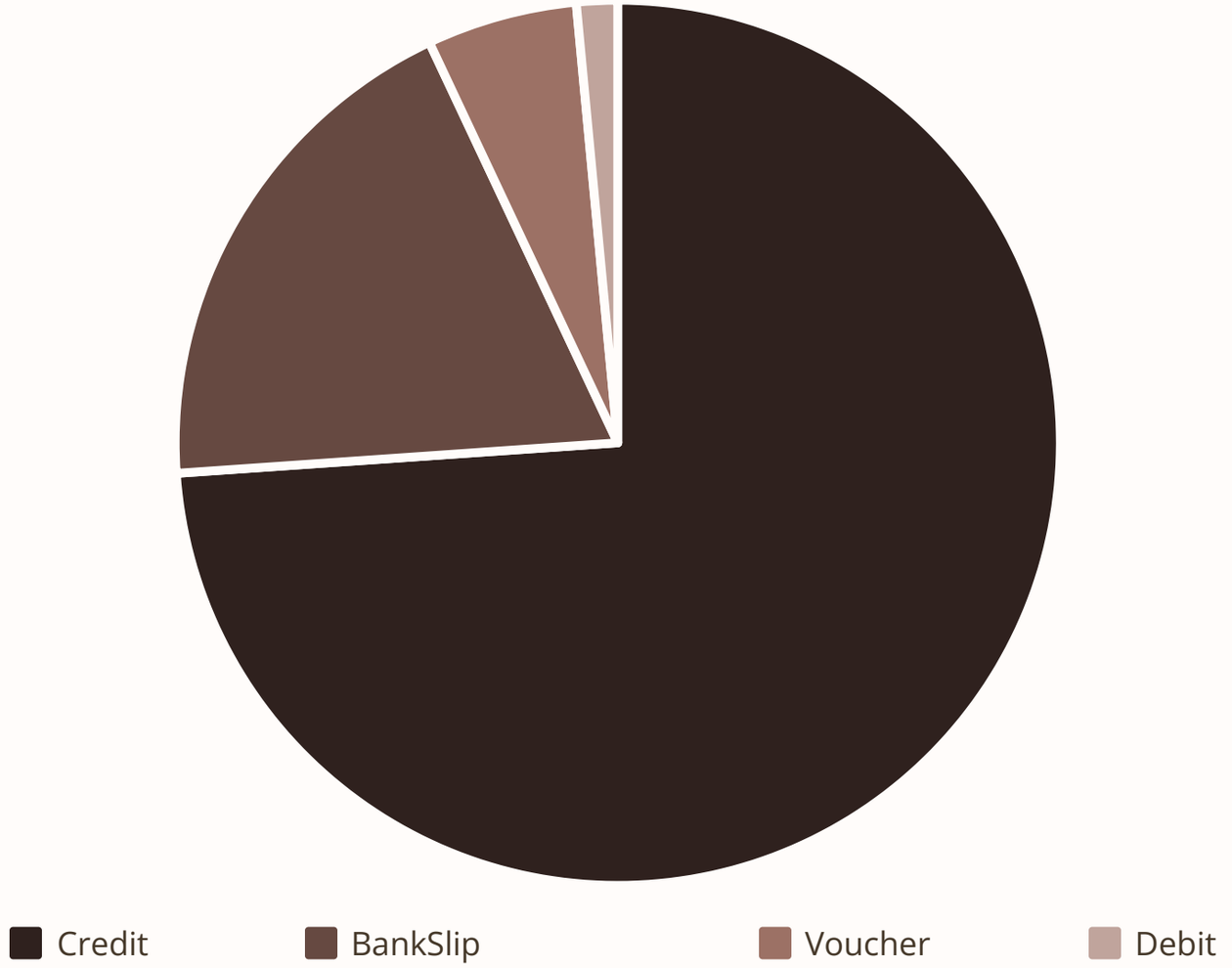
- 96.5% of orders were successfully delivered
- Peak ordering days: Monday and Tuesday
- Highest order volume: November (holiday season impact)
- Daily order trends show consistent growth throughout 2017-2018



Time series demonstrates growth and a clear November peak tied to seasonal demand.

Key Findings — Payment Insights

- Credit card is the dominant payment method (73.9% of transactions)
- Other methods: Bank slip (19.1%), Voucher (5.5%), Debit card (1.5%)
- Average transaction values vary significantly by payment type



Payment mix highlights customer preference for credit options, suggesting installment and credit-driven strategies.

Key Findings — Product & Geographic Analysis

- Top-selling categories: Health & Beauty, Watches & Gifts, Bed/Bath/Table
- Strong correlation between product weight and freight costs
- 73 distinct product categories analysed
- São Paulo leads with the highest order volume
- Rio de Janeiro and Belo Horizonte follow as major markets
- Geographic analysis reveals concentration in Brazil's southeastern region

Visualizations & Tools

The project includes:

- Time series analysis of daily/monthly order trends
- Seasonal patterns by day of week and month
- Payment method distribution (pie charts and bar plots)
- Product category sales comparison
- Geographic distribution of orders
- Review score distribution
- Correlation heatmap of numerical variables
- Scatter plot analysis of weight vs. freight costs

Libraries and tools used:

- pandas, numpy - Data manipulation and analysis
- matplotlib, seaborn, plotly - Data visualization
- kagglehub - Dataset acquisition
- geopandas - Geospatial analysis

Data Quality Improvements, Business Insights & Conclusion

Data Quality Improvements

- Removed 610 invalid orders with timestamp inconsistencies
- Eliminated 1,244 zero-price items
- Dropped 610 rows with missing product categories
- Removed duplicate reviews (keeping first occurrence)
- Standardized 32,951 product records

Business Insights

- Logistics Optimization: Strong correlation between product weight and freight value suggests opportunities for shipping cost optimization
- Payment Strategy: Credit card dominance indicates customer preference for installment options
- Seasonal Planning: November peak demands increased inventory and logistics capacity
- Geographic Focus: São Paulo market represents significant revenue opportunity
- Customer Experience: Bimodal review pattern suggests need for improved quality consistency

Conclusion

This comprehensive EDA reveals critical insights into Brazilian e-commerce operations, highlighting opportunities for operational improvements, customer experience enhancement, and strategic market expansion. The analysis provides a solid foundation for data-driven decision-making in the e-commerce domain.

> Author: Sandeep Batra
>
> Date: October 2025
>
> Dataset Source: Olist Brazilian E-Commerce Dataset (Kaggle)