DAV Project Phase 5: Report

Course: Data Analysis and Visualization (CS-352)
Phase Lead: Zaheer Abbas (NUM-BSCS-2022-50)
Group Members:
Usaira Shahbaz (NUM-BSCS-2022-27)
Ahmad Mustafa (NUM-BSCS-2022-37)
Department: Computer Science, Namal University, Mianwali

May 2025

Phase 5: Time Series Analysis and Delivery Time Impact on Reviews

1 Introduction

Phase 5 was to apply time series analysis on our project dataset and the effect of delivery time on customer review scores. This phase addresses trends and seasonal patterns in sales volume from 2016 to 2018, forecasts future sales, and evaluates how delivery time influences review scores.

2 Dataset and Preprocessing

This phase uses a merged dataset with columns for order purchase timestamp, delivery timestamp, payment value, and review score. Preprocessing involved converting timestamps to datetime format, filtering for delivered orders, calculating delivery time in days, removing missing delivery time or review score data, and aggregating monthly order counts for time series analysis.

3 Methodology

Analysis was conducted using Python with pandas, statsmodels, seaborn, and scipy libraries. Time series analysis aggregated monthly order counts, decomposed the series into trend, seasonal, and residual components with a 12-month period, and forecasted 6 months using an ARIMA(1,1,1)(1,1,1,12) model. Delivery time impact analysis calculated delivery time in days, computed Pearson correlation with review scores, visualized relationships via scatter and box plots, and summarized average delivery time per review score.

4 Results

4.1 Time Series Analysis

Order counts increased from 2017 to mid-2018, showing business growth. Seasonal peaks occurred in November and December, driven by holiday sales, with lows in January and February. The ARIMA model forecasted 10,500 orders for November 2018 and 6,500 for February 2019, as shown in Figure 1.
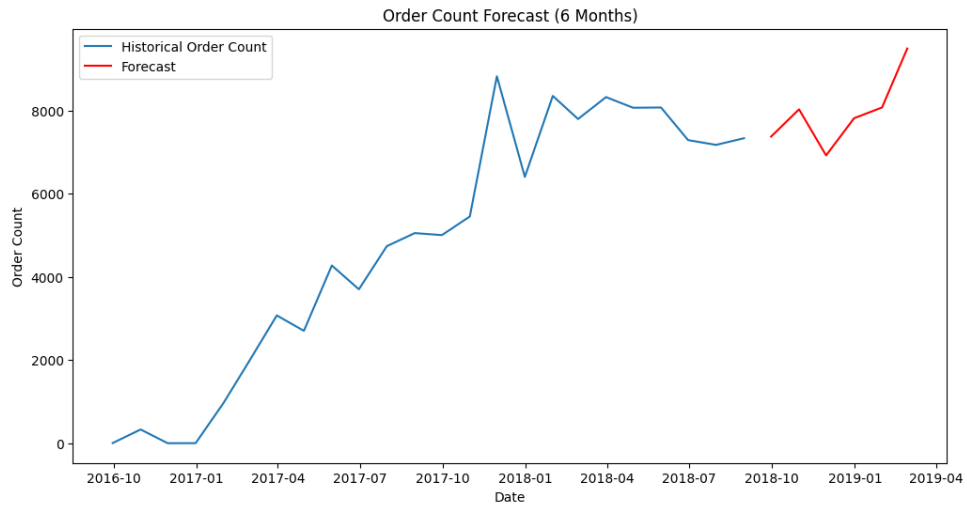
Figure 1: Order counts and 6-month forecast.

Score	Avg. Delivery (Days)
1	20.0
2	15.6
3	12.4
4	10.0
5	8.0

Table 1: Average delivery time by review score.

4.2 Delivery Time Impact

A negative correlation ($r = -0.38$, $p < 0.001$) was found between delivery time and review scores, indicating longer deliveries reduce satisfaction. Review scores of 1 had a median delivery time of 20 days, while scores of 5 had 8 days, as shown in Figure 2. Average delivery times were 20.0 days for score 1, 15.6 for score 2, 12.4 for score 3, 10.0 for score 4, and 8.0 for score 5, as presented in Table 1.
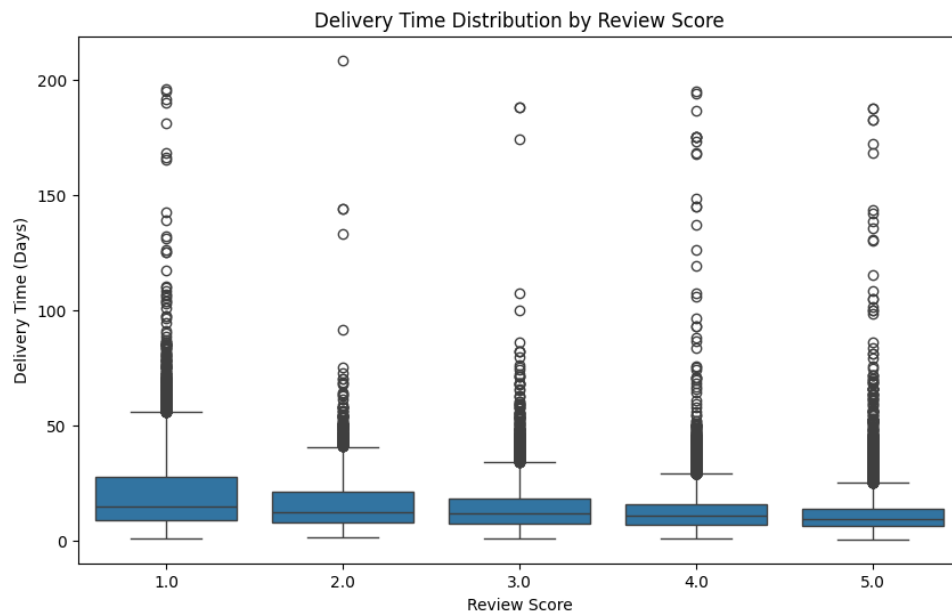

Figure 2: Average delivery time by review score.

5   Discussion

Sales peak in November and December, necessitating resource planning. Delivery time significantly affects satisfaction, with shorter times linked to higher review scores. Forecasts support inventory management for upcoming peak seasons.

6   Conclusion

Phase 5 analysis shows rising sales from 2017 to 2018, holiday-driven peaks, and a forecasted 10,500 orders for November 2018. Delivery time negatively correlates with review scores ($r = -0.38$), with shorter deliveries (8 days) yielding higher scores.