

Final Project EDA

Victor Huang, Parker Johnson, Elijah Jones

5/24/2022

Data Description

Our data consists of the total number of daily transactions at 54 Corporación Favoritas grocery stores in Ecuador, recorded by the stores from 2013 to 2017. The dataset includes the date (our predictor variable) and the total number of products purchased in the store on that day (our response variable).

Each of the 54 stores are located in 1 of 22 cities in Ecuador. For our analysis, we have chose to analyze Store 1, which is located in Quito, and Store 28, which is located in Guayaquil. We will be looking at the total number of products sold daily, as well as the total number of products sold by week at these stores.

Our dataset, 'transactions.csv', was found on kaggle.com under the title "Store Sales - Time Series Forecasting". The link is included here: <https://www.kaggle.com/competitions/store-sales-time-series-forecasting/overview>

Research Description

Guayaquil and Quito are Ecuador's largest and second largest cities. We will address the following questions:

Do holidays have an effect on sales? In other words, are sales best explained using a non-seasonal or seasonal time series model?

Can sales of stores in Guayaquil and Quito be explained using the same time series model? How do sales differ for stores in Guayaquil versus Quito?

Initial Graphical Summaries: Stores 1 and 28

Figure 1 shows the number of daily transactions from 2013 to 2017 in Store 1.

To make it easier to observe trends in the data, we grouped sales by week. Figure 2 shows the number of weekly transactions from 2013 to 2017 in Store 1. Sales seem to follow a yearly cycle, peaking in late December. The mean and variance of the time series seem to be fairly constant. It appears that a seasonal model may be suitable for these data.

Figure 3 shows the number of daily transactions from 2013 to 2017 in Store 28.

To make it easier to observe trends in the data, we grouped sales by week. Figure 4 shows the number of weekly transactions from 2013 to 2017 in Store 28. Sales seem to follow a yearly cycle, peaking in late December. However, we see additional spikes scattered at other weeks throughout the year. Additionally, the mean of the series is increasing. We may need to remove the mean trend before further analysis of this series.

Analysis Plans

First, we will remove the mean trends in both time series. Then, we will use the ACF, PACF, and periodogram to fit the best model to each time series. Then, we will compare and contrast the models. We will use the models to explain trends in sales, and conduct future research to provide rationale for differences in the trends.

Transactions for Store 1 in Quito vs. Date

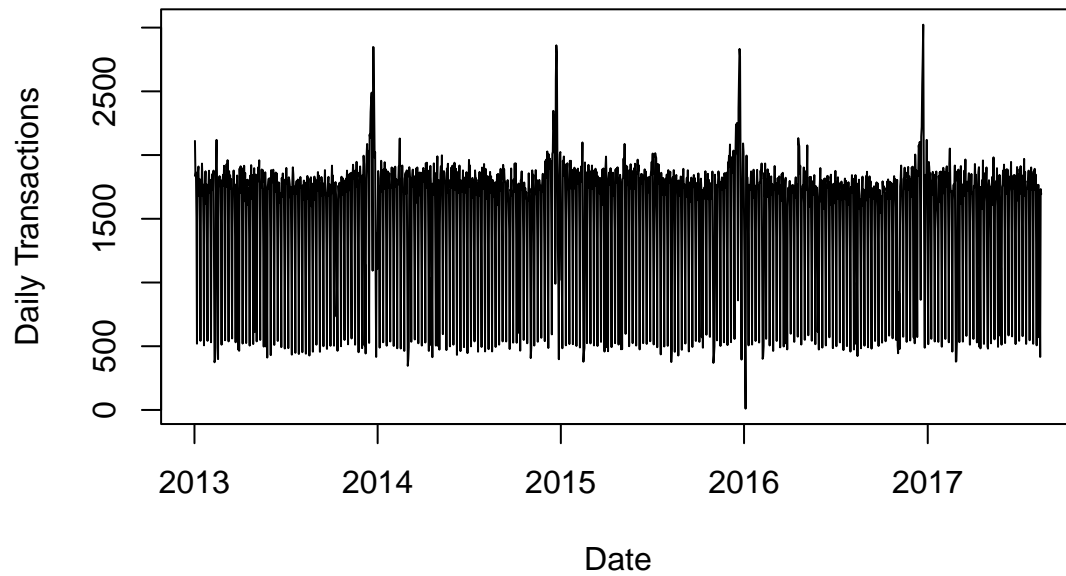


Figure 1: Daily transactions for store 1

Transactions for Store 1 in Quito vs. Week Number

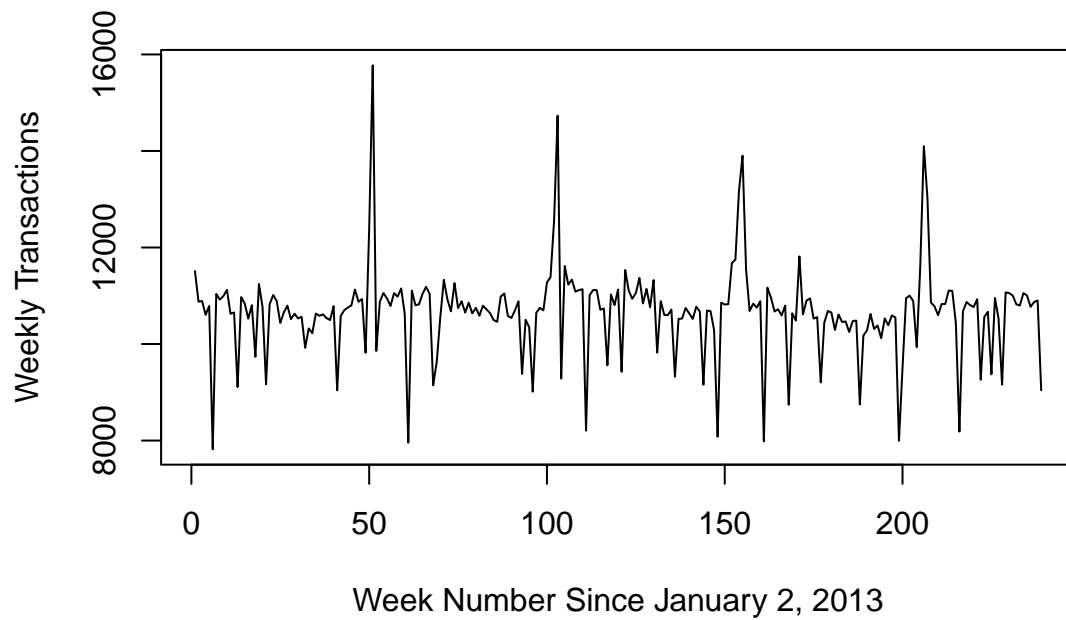


Figure 2: Weekly transactions for store 1

Date Versus Transactions for Store 28 in Guayaquil

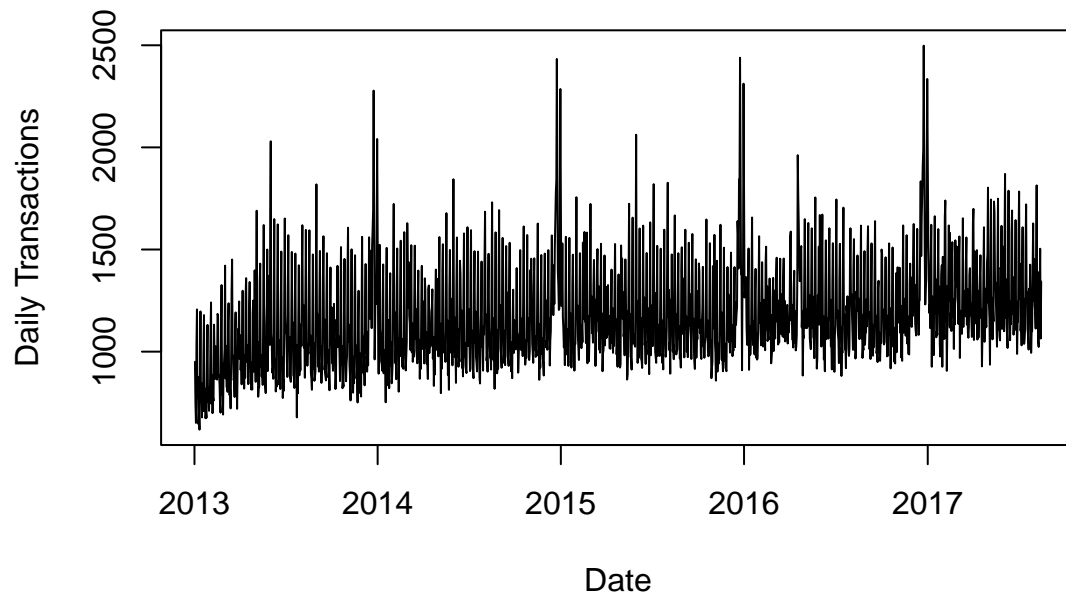


Figure 3: Daily transactions for store 28

Week Number Versus Transactions for Store 28 in Guayaquil

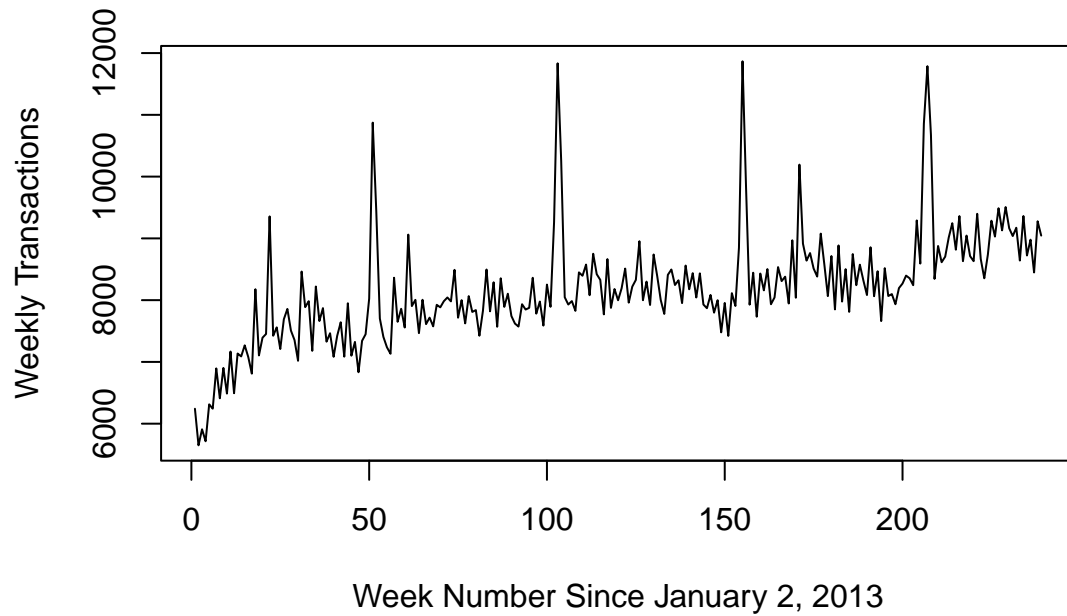


Figure 4: Weekly transactions for store 28