Favorita Sales Forecast

By: Hamza Al Bustanji
Mentor: Dipanjan Sarkar

Why are we forecasting sales?

Sales forecasting is an essential component of strategic planning for any retail company. It involves predicting future sales performance based on past data, market trends, and other relevant factors. The importance of forecasting sales lies in its ability to provide critical information that helps retail companies make informed decisions about their operations and future growth.

1. Data Collection

Our data were provided through:

➔ **Kaggle**

The company provided the data through a Kaggle competition.

➔ **The World Bank's Website**

We obtained Ecuadorian inflation data from the World Bank's website to analyze inflation-adjusted sales.

2. Data Wrangling

The tasks that we completed in this step were:

- **Displayed Our Data**
- **Data Cleaning**
  We cleaned the data from the World Bank's website so that it's in usable shape for our project.
- **Created Data Frames**
  We created different data frames for use in our EDA work.

3. EDA

The main aspects that we wished to investigate in our EDA were:

- ➔ **The Inherent Nature of Our Time-Series**
  Is there seasonality in our data? If so, what are the lengths of the seasons? Is there a trend? Is it an upward or a downward trend? etc..

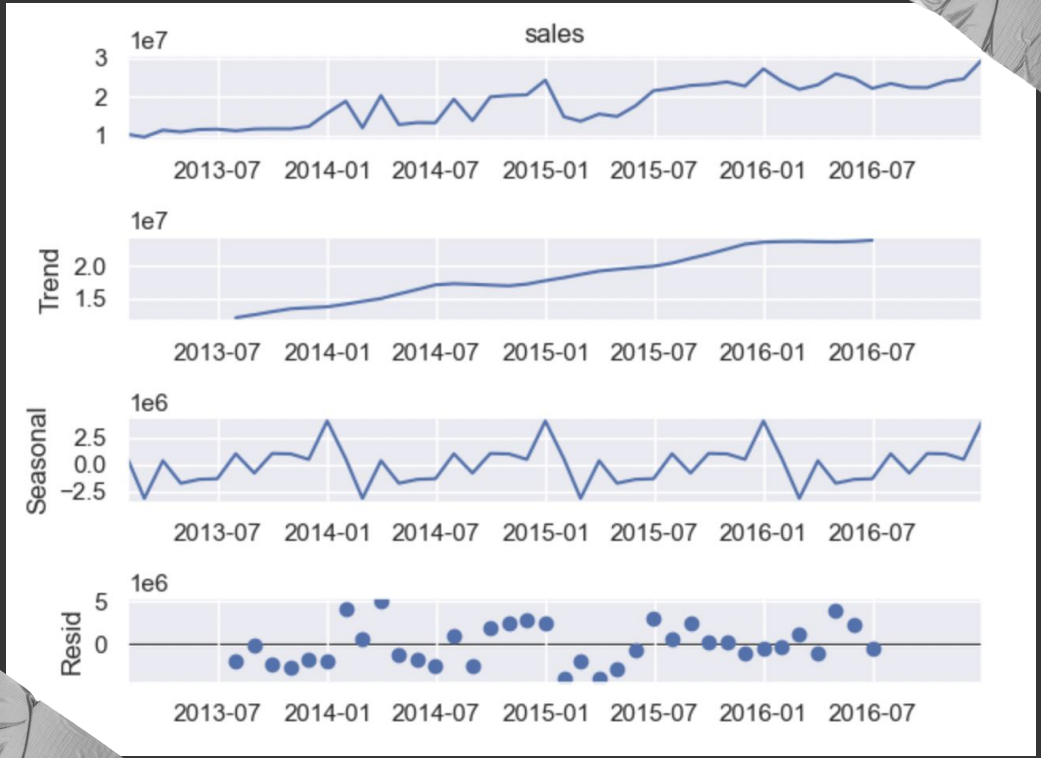
- ➔ **How Do sales relate to other data?**
  How do sales relate to gas prices? How do they relate to themselves? How does adjusting for inflation affect sales?

3. EDA - Seasonality and Trend

Seasonality and trend:

The plot on the left displays the annual trend. The plot on the right displays seasonality fluctuations each month.

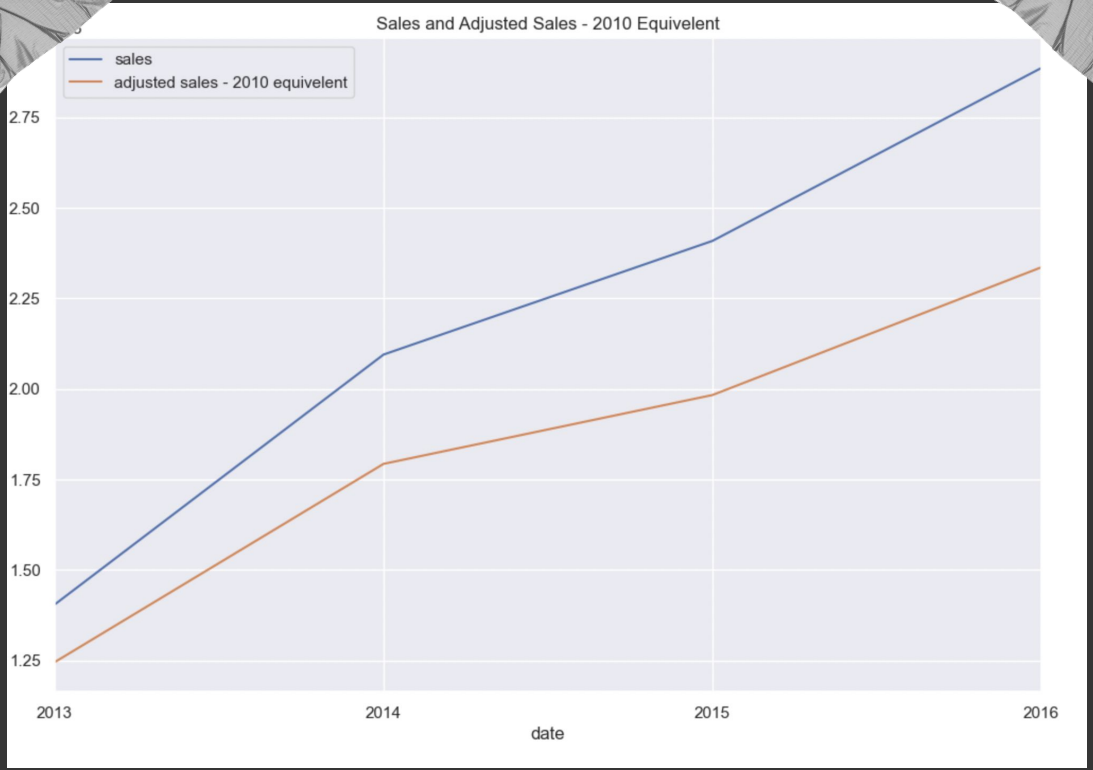
3. EDA - Seasonality and Trend



ETS decomposition:

First row shows our actual data. Second row shows the isolated trend. Third row shows the seasonal component. Last row shows the residual component.

3. EDA - Inflation Adjusted Sales



Inflation Adjusted Sales:

The Important aspect of this plot is observing the rate of change of each line (the slope) to determine to which extent higher sales are driven by inflation.

4. Pre-processing and Modeling

In this notebook we did the following:

➔ **Train/Test split**
We performed a train/test split with the last available 6 months being the testing set.

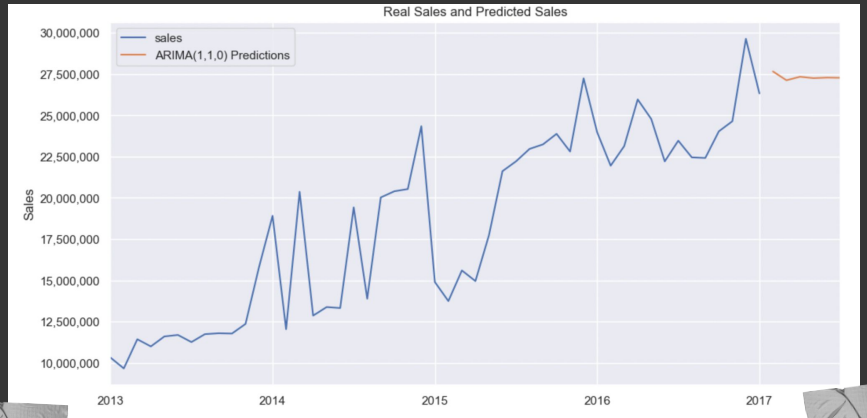
➔ **Grid-search**
We performed a grid-search to find the best parameters for the ARIMA and SARIMA models

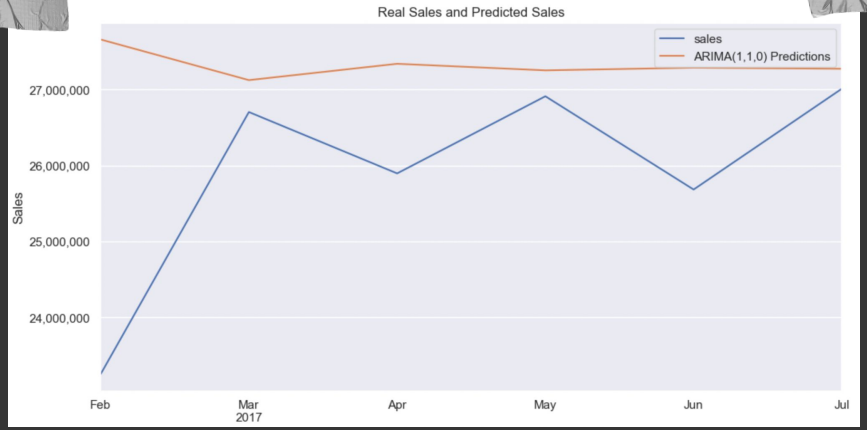
➔ **Plotted each model's predictions**
We plotted the values that were predicted by each model.

➔ **Created a Prophet model**
We went through the necessary pre-processing to be able to fit a Prophet model

➔ **Model Evaluation**
We evaluated the results of each model. We used the RMSE as our metric.

4. Pre-processing... - ARIMA and SARIMA

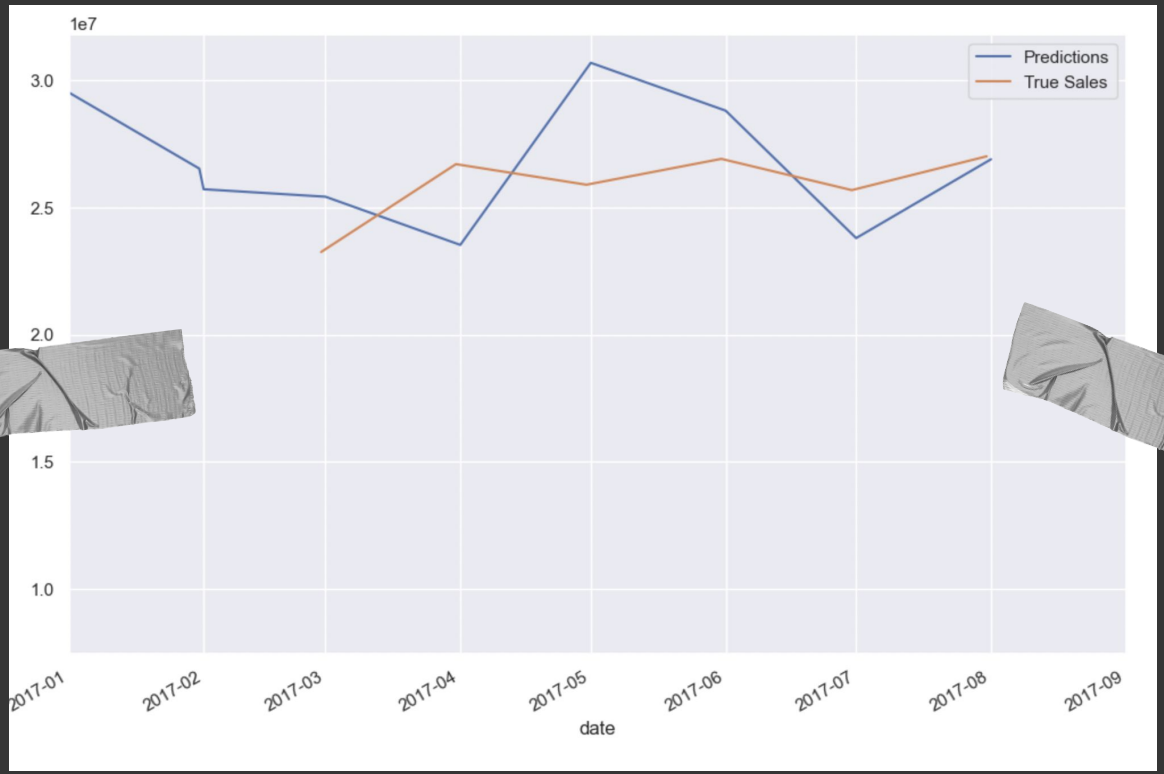




ARIMA and SARIMA:

Both the ARIMA and SARIMA models performed the same, which was due to Auto-ARIMA not picking up on any seasonality in the data..

4. Pre-processing... - Prophet Model



Prophet:

The Prophet model performed slightly worse than the ARIMA model, although it seems to have captured more of the seasonality of the data.

5.Evaluation Results

Model	RMSE
ARIMA	$ 2019662
SARIMA	$ 2019662
Prophet	$ 2736002