Store Sales

Forecasting

(Tee) Thanh Nguyen - Nov 2023

>>> print("(Tee) Thanh Nguyen")

>>> print("About Me:")

+3 years in: Marketing / Digital/ Insights

Industry: Tech/ Marketing / Finance

Hobby: ๑^•ﻌ•^๑📖🍳🍜🌍✈️

Agenda

01
The Problem

02
Project Goal

03
Workflow

04
Results

05
Conclusion

The Context

- Forecasting is a crucial component of any successful grocery retail business.
- Accurate sales forecasting and prediction of future trends are more critical for grocery retailers than ever, against the complexity from the environments.

The Company

- Corporación Favorita is a large Ecuador-based grocery retailer. They operate hundreds of supermarkets, with over 200,000 different products on their shelves.
- The dataset is adapted from a large Kaggle time series forecasting dataset.

The Problem

The task:

It's all about "**Accurate Predictions**" of sales!

Accurate foresights of the future

=

Better operational, tactical, and strategic decisions

The solution:

Time-stamped data & **time series forecasting**

Project Goal

- Utilize statistical modeling & machine learning to build a practical time series model to accurate forecasts the future daily sales, based on historical data
- Discover patterns in the historical sales data, such as trend, seasonality, cyclical or irregular variations

Workflow

01 Data Preprocessing

Load, investigate, and transform data, feature engineering, train test split

02 Exploratory Data Analysis

Visualize, analyze data to discover trends, seasonality, patterns and check assumptions

03 Model Development

Develop 2 time series models to solve the task: **SARIMA** & **Facebook Prophet**

04 Model Evaluation

Using the RMSE to compare the 2 models. Best model: **Facebook Prophet**

Optimization & Tuning using `auto.arima`
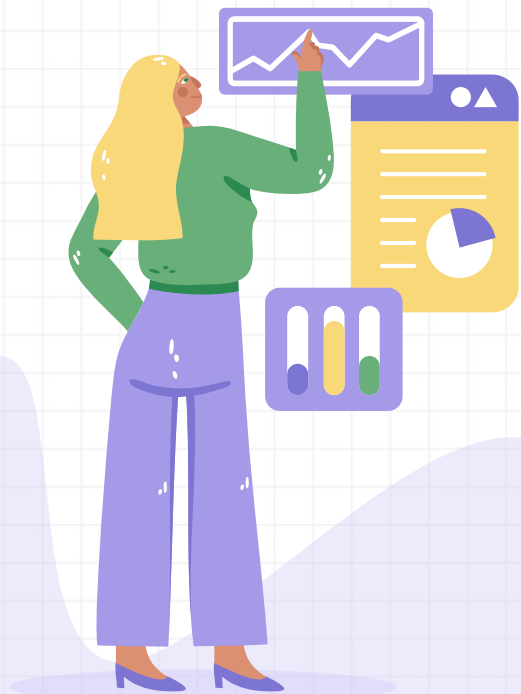
EDA Insights

Holidays & Events Calendar
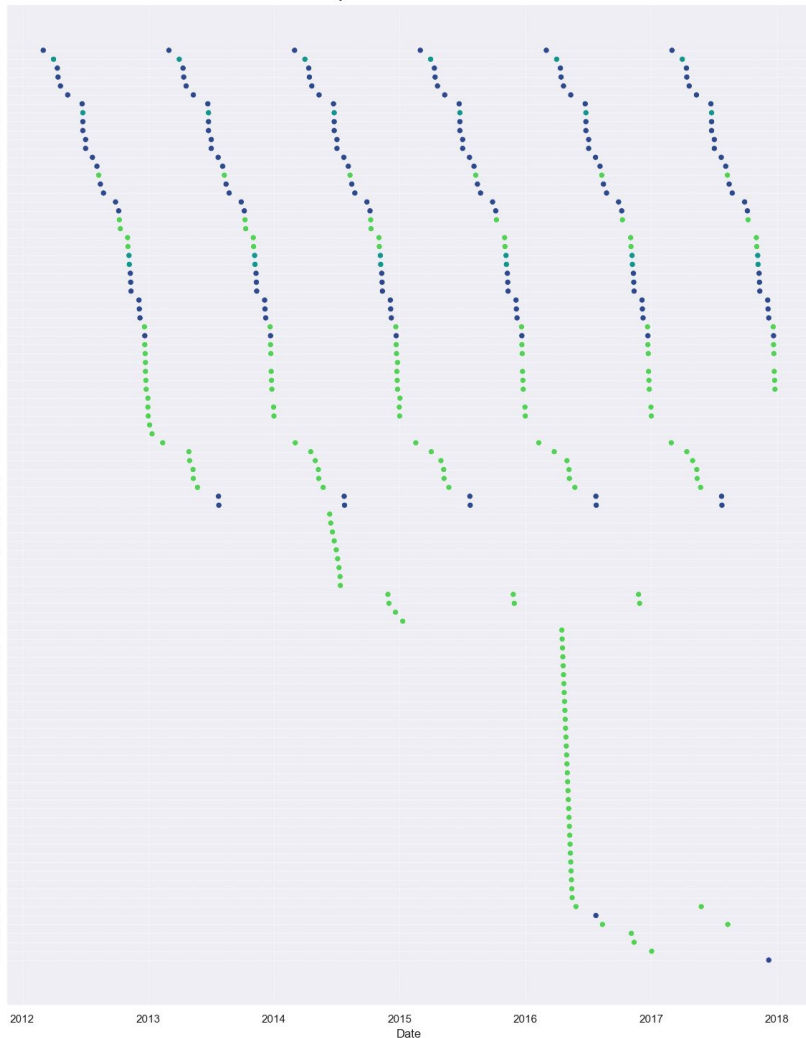
Daily and Yearly Total Sales
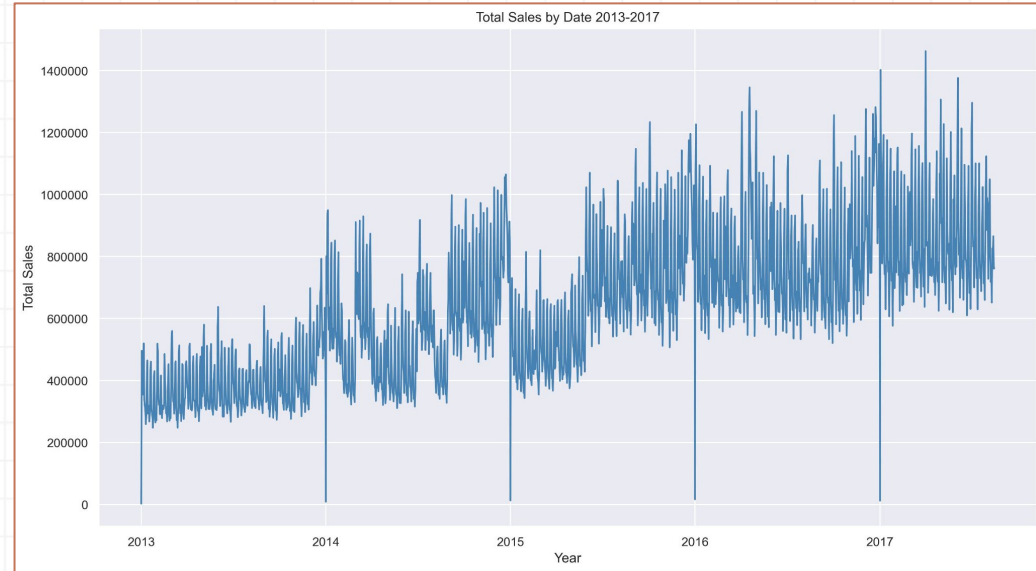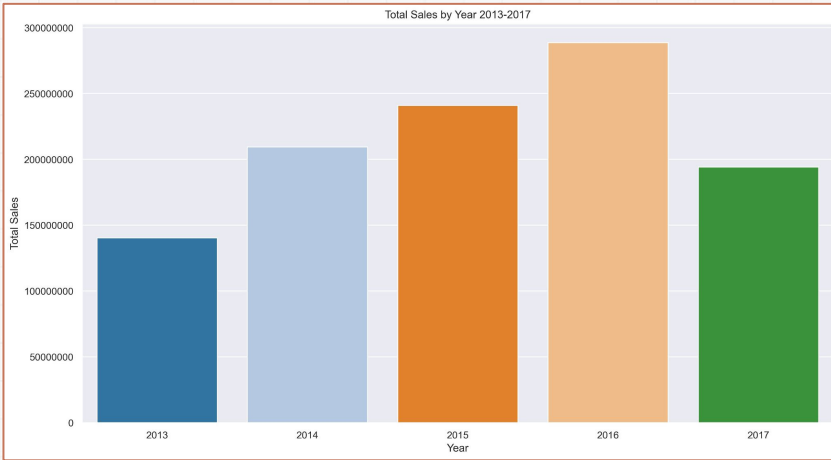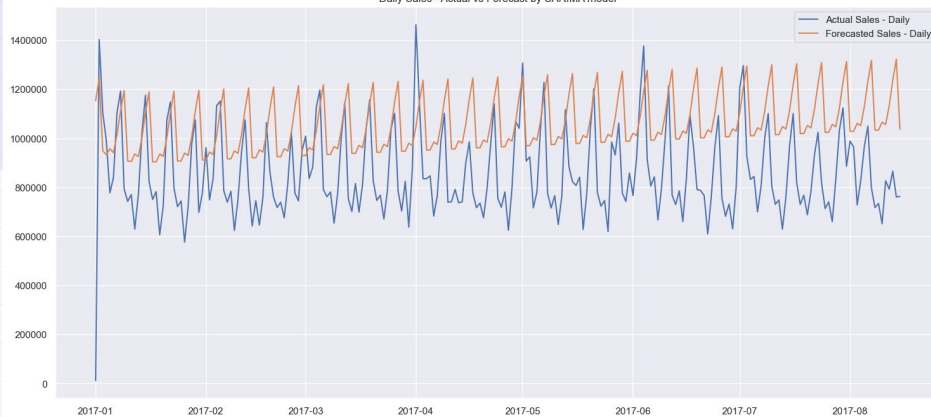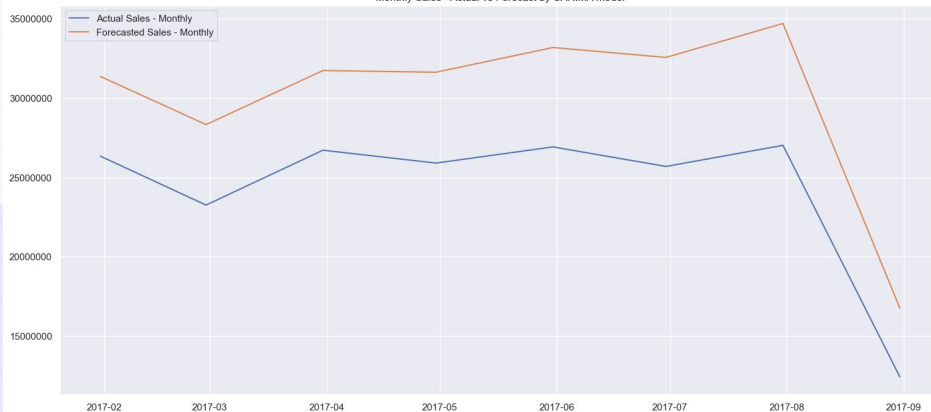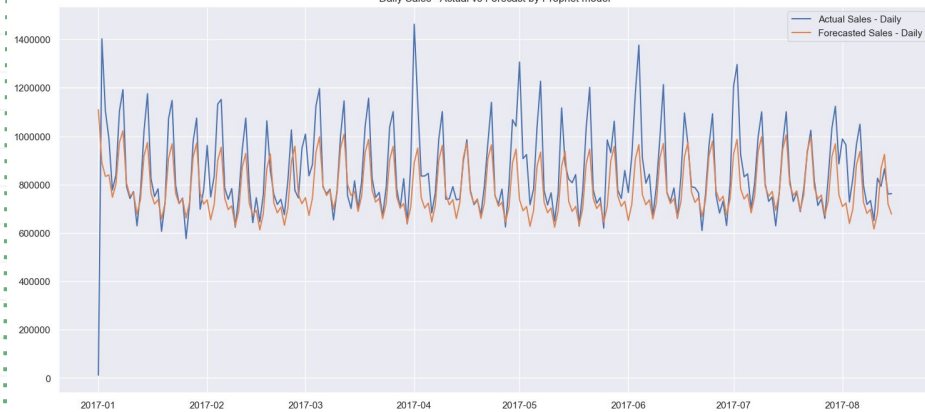
Avg Promotion vs Avg Sales Per Year
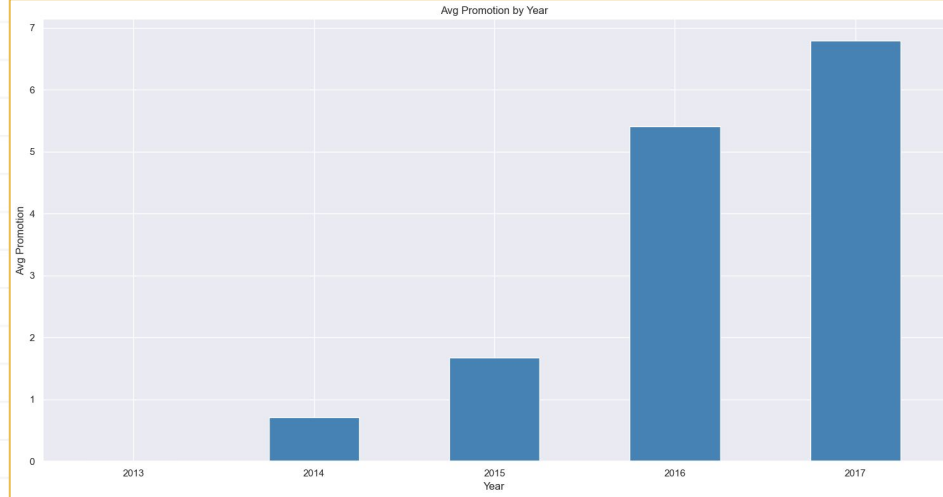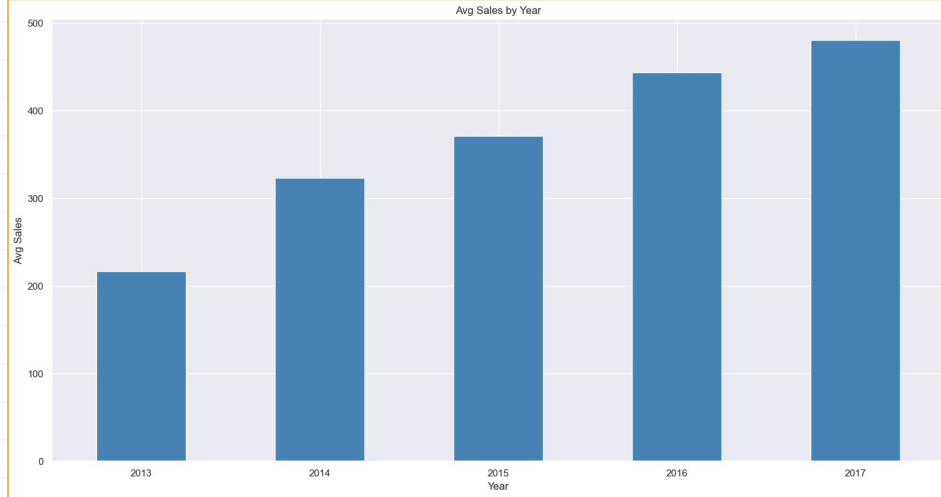
Forecast Results &
Model Evaluation

Daily Sales - Actual vs Forecast by SARIMA model
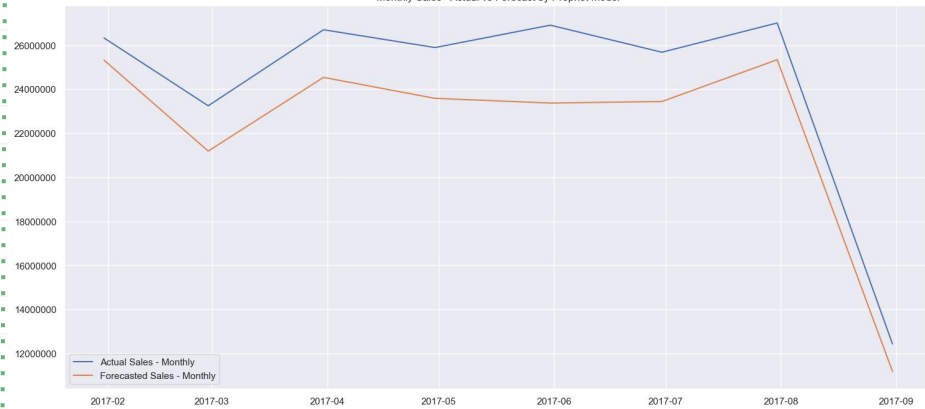
Daily Sales - Actual vs Forecast by Prophet model

Monthly Sales - Actual vs Forecast by SARIMA model

Monthly Sales - Actual vs Forecast by Prophet model

SARIMA

RMSE = 269919.81715

Prophet

RMSE = 147375.03819

Conclusion

Skills Gained

Solve with a common real-world business problem with time series sales & demand forecasting.

Challenges

Handle a large grocery dataset with over 3 million rows. Research how to perform auto.arima and git push for huge data.

Grateful For

The opportunity to learn about the new "time series" concept and apply into a full data science project.

Future Goals

Improve the Time Series Forecasting models with LLM, and deploy in Streamlit.

Thank you!

thanhnguyen.ulaval@gmail.com
flowcv.me/thanhnguyen

github.com/TeeNguyenDA

linkedin.com/in/thanhnguyentn

Let's connect!