

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: 🎵 • 📖 🔍 🍜 🌍 ✈️



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

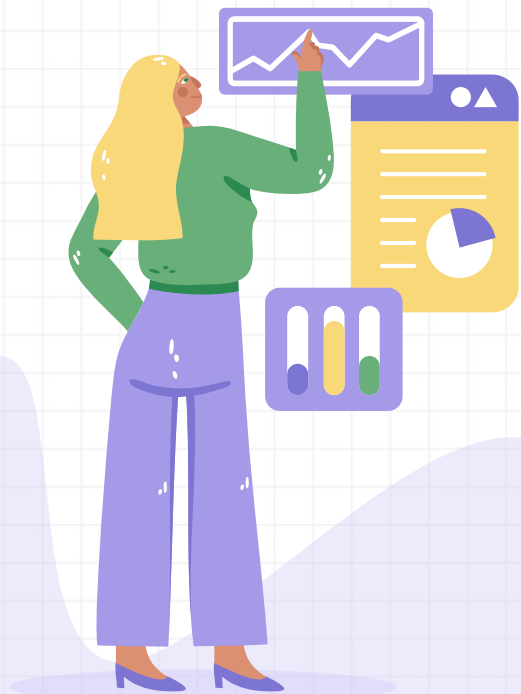


Model Evaluation

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



Optimization & Tuning using `auto.arima`



EDA Insights

Holidays & Events Calendar

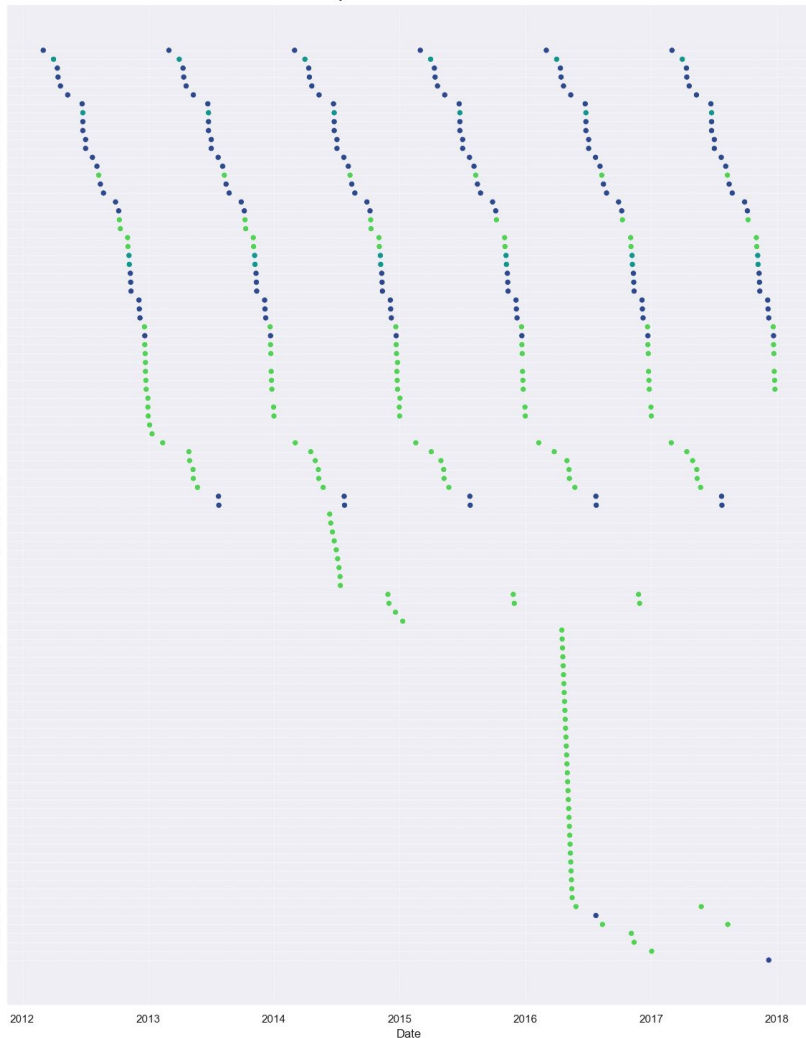


Holiday Description

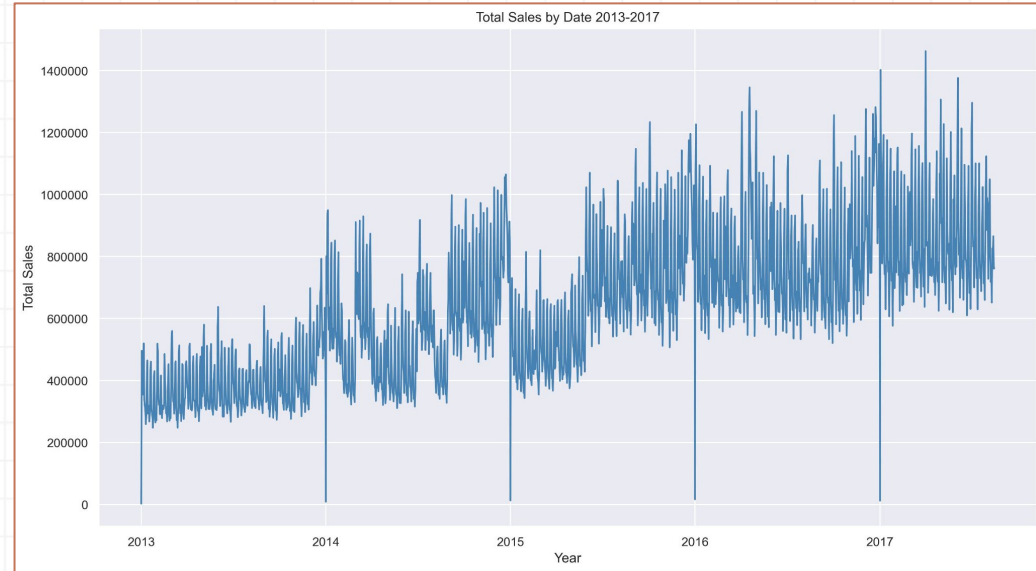
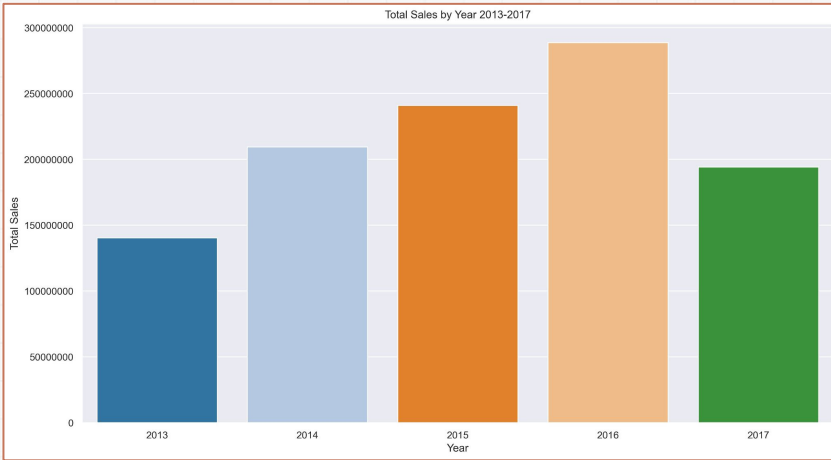
Fundacion de Manta
 Provincializacion de Cotopaxi
 Fundacion de Cuenca
 Cantonizacion de Libertad
 Cantonizacion de Riobamba
 Cantonizacion del Puyo
 Cantonizacion de Guaranda
 Provincializacion de Imbabura
 Cantonizacion de Latacunga
 Fundacion de Machala
 Fundacion de Santo Domingo
 Cantonizacion de El Carmen
 Cantonizacion de Cayambe
 Fundacion de Esmeraldas
 Primer Grito de Independencia
 Fundacion de Riobamba
 Fundacion de Ambato
 Fundacion de Ibarra
 Cantonizacion de Guevedo
 Independencia de Guayaquil
 Traslado Independencia de Guayaquil
 Dia de Difuntos
 Independencia de Cuenca
 Provincializacion de Santo Domingo
 Provincializacion Santa Elena
 Independencia de Guaranda
 Independencia de Latacunga
 Independencia de Ambato
 Fundacion de Quito-1
 Fundacion de Quito
 Fundacion de Loja
 Cantonizacion de Salinas
 Navidad-3
 Navidad-2
 Puente Navidad
 Navidad-1
 Puente Primer dia del ano
 Primer dia del ano-1
 Primer dia del ano
 Recupero puente Navidad
 Recupero puente primer dia del ano
 Carnaval
 Viernes Santo
 Dia del Trabajo
 Dia de la Madre-1
 Dia de la Madre
 Batalla de Pichincha
 Fundacion de Guayaquil-1
 Fundacion de Guayaquil
 Inauguracion Mundial de futbol Brasil
 Mundial de futbol Brasil- Ecuador-Suiza
 Mundial de futbol Brasil- Ecuador-Honduras
 Mundial de futbol Brasil- Ecuador-Francia
 Mundial de futbol Brasil- Octavos de Final
 Mundial de futbol Brasil- Cuartos de Final
 Mundial de futbol Brasil- Semifinales
 Mundial de futbol Brasil- Tercer y cuarto lugar
 Mundial de futbol Brasil- Final
 Black Friday
 Cyber Monday
 Recupero Puente Navidad
 Recupero Puente Primer dia del ano
 Terremoto Manabi+1
 Terremoto Manabi+2
 Terremoto Manabi+3
 Terremoto Manabi+4
 Terremoto Manabi+5
 Terremoto Manabi+6
 Terremoto Manabi+7
 Terremoto Manabi+8
 Terremoto Manabi+9
 Terremoto Manabi+10
 Terremoto Manabi+11
 Terremoto Manabi+12
 Terremoto Manabi+13
 Terremoto Manabi+14
 Terremoto Manabi+15
 Terremoto Manabi+16
 Terremoto Manabi+17
 Terremoto Manabi+18
 Terremoto Manabi+19
 Terremoto Manabi+20
 Terremoto Manabi+21
 Terremoto Manabi+22
 Terremoto Manabi+23
 Terremoto Manabi+24
 Terremoto Manabi+25
 Terremoto Manabi+26
 Terremoto Manabi+27
 Terremoto Manabi+28
 Terremoto Manabi+29
 Terremoto Manabi+30
 Traslado Batalla de Pichincha
 Traslado Fundacion de Guayaquil
 Traslado Primer Grito de Independencia
 Puente Dia de Difuntos
 Recupero Puente Dia de Difuntos
 Traslado Primer dia del ano
 Traslado Fundacion de Quito

Holidays & Events Calendar 2013-2017

Locale
 • Local
 • Regional
 • National

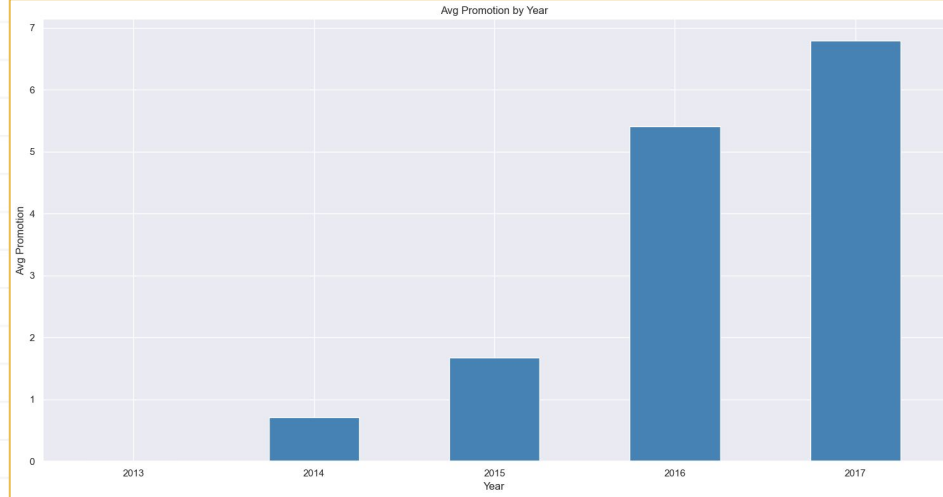


Daily and Yearly Total Sales

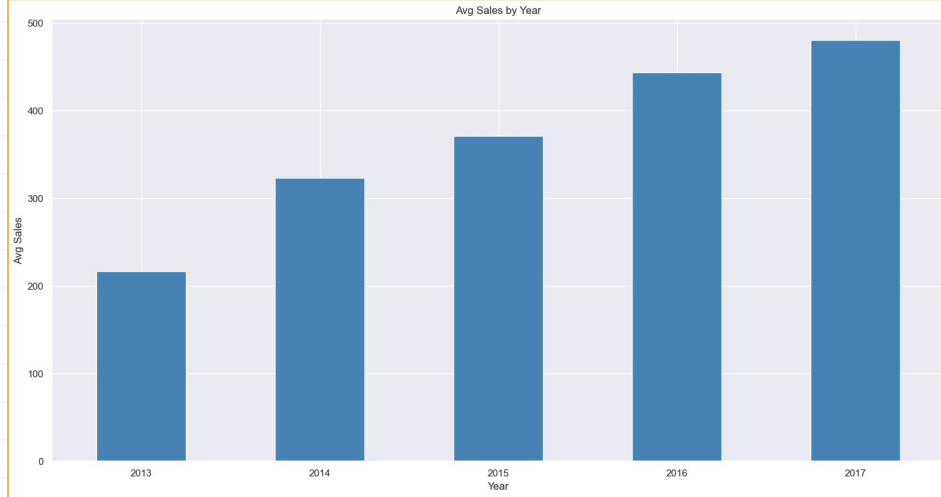


Avg Promotion vs Avg Sales Per Year

Avg Promotion by Year



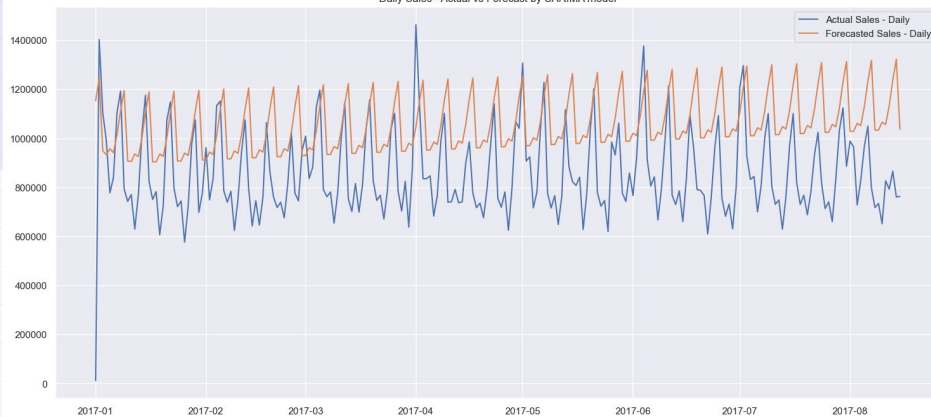
Avg Sales by Year



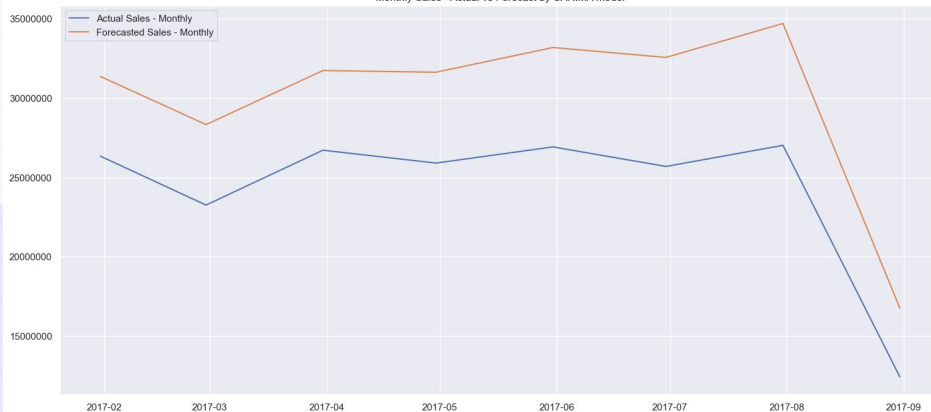


Forecast Results & Model Evaluation

Daily Sales - Actual vs Forecast by SARIMA model



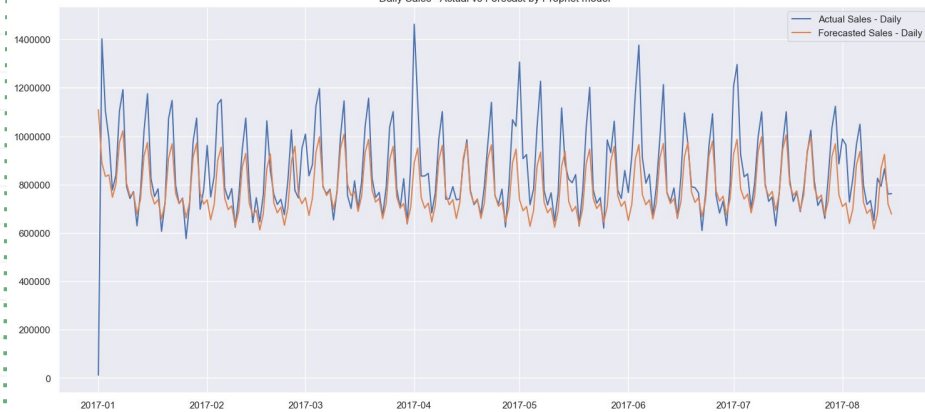
Monthly Sales - Actual vs Forecast by SARIMA model



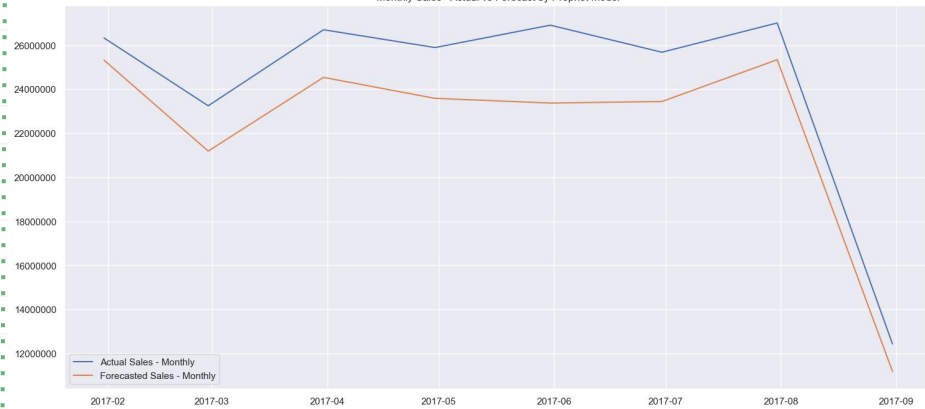
SARIMA

RMSE = 269919.81715

Daily Sales - Actual vs Forecast by Prophet model



Monthly Sales - Actual vs Forecast by Prophet model



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

CREDIT

Let's connect!