Data Glacier
Your Deep Learning Partner

Exploratory Data Analysis

Data Science - Retail Forecasting

February 9th, 2023

Hina Timothy, Junfei Liu, Siming Luo

Agenda

Executive Summary

EDA

Recommendations

Executive Summary

Beverage Industry Highlights
- Market volume, from 696.1 Liters to 994.1 (2009 to 2018)
- PepsiCo, 342 new non-alcoholic beverage (January 2016 and February 2021)
- Anheuser-Busch InBev, global learning company with over 54 million U.S. dollars sales

Problem Statement
- An Australian company which has beverage business has needs to forecast sales for each product. The company already has a in-house solution for predicting sales. However, they want to seek solutions based on AI/ML.
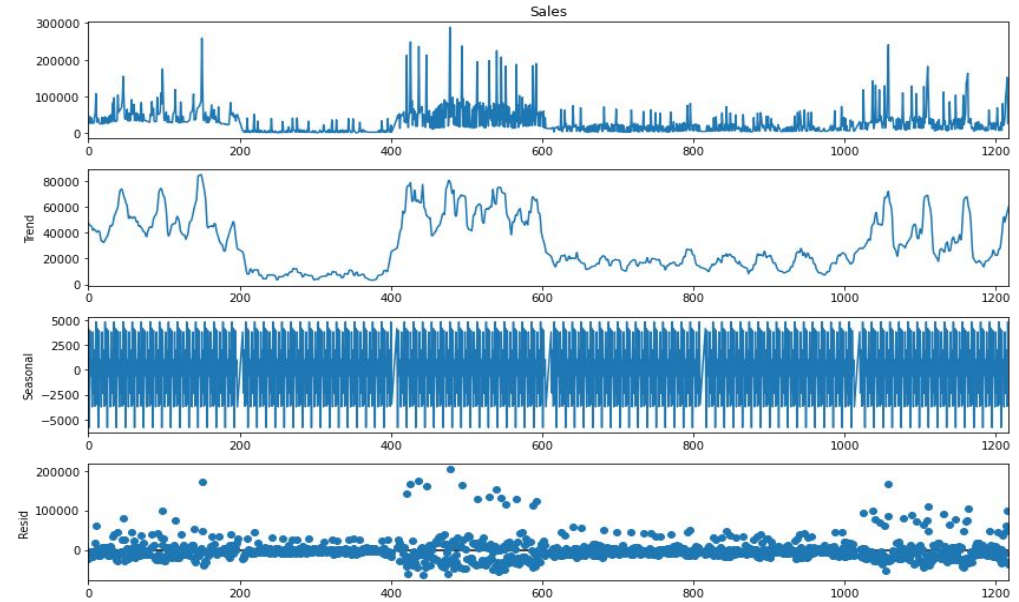
Exploratory Data Analysis Approach
- Seaborn, matplotlib.pyplot
- ydata_profiling

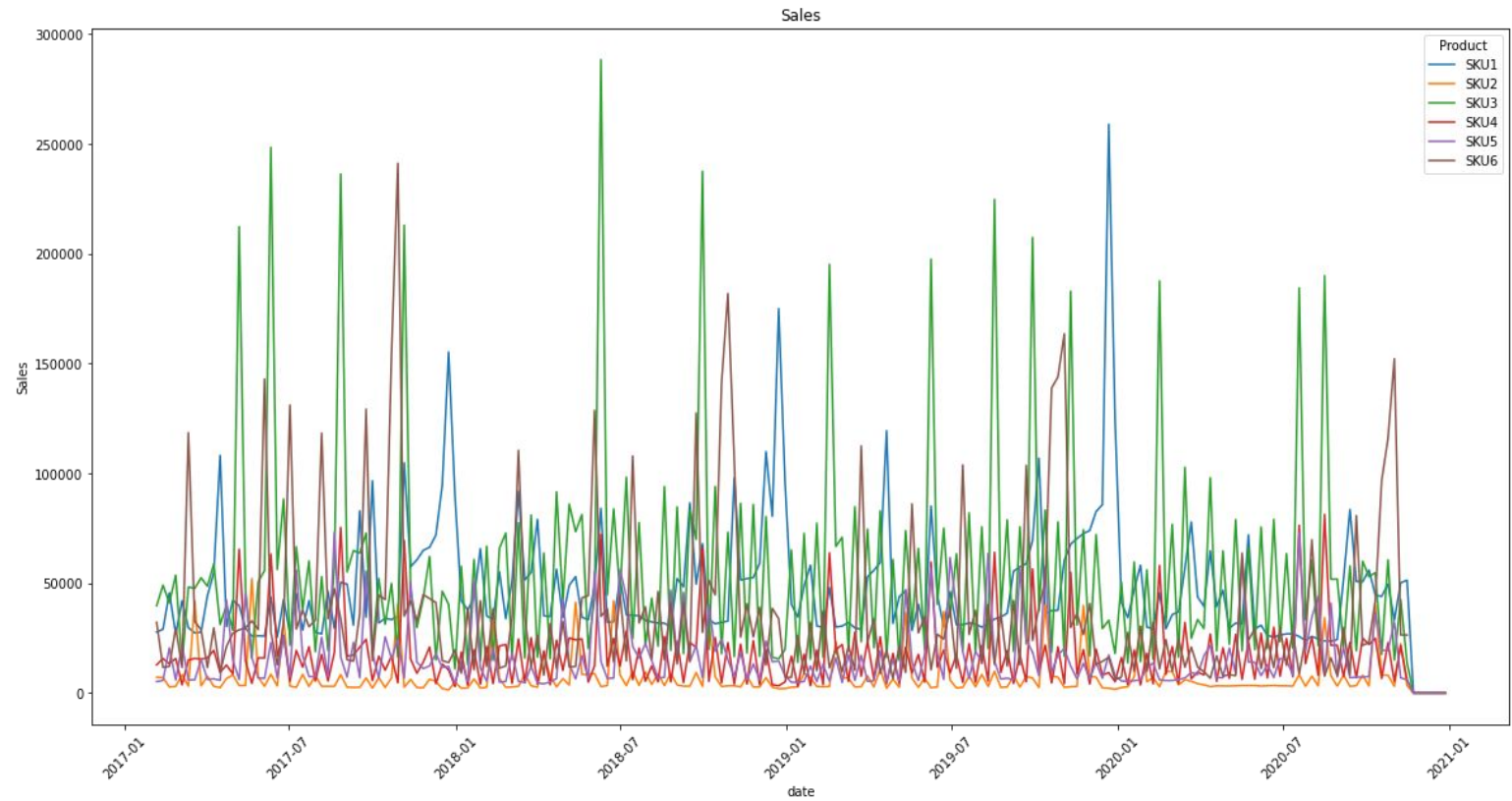# Exploratory Data Analysis
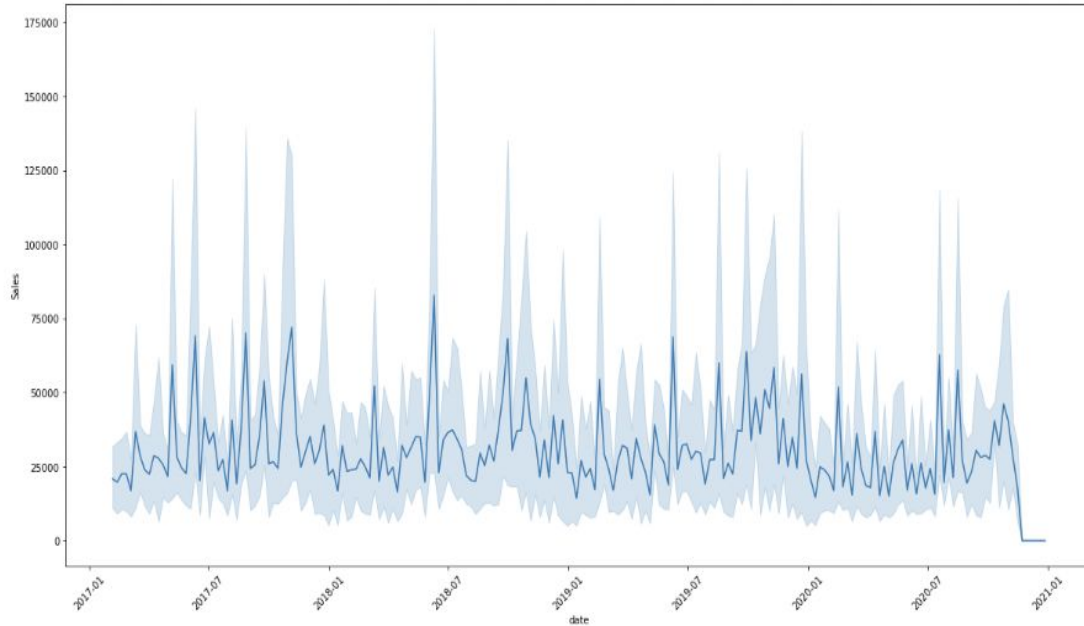
Sales overtime

Decomposition of Sales (Trend, Seasonality, Residuals)

- By checking how sales perform overtime, it is evident that seasonality exists
- Further decomposition shows that there is a non-linear and additive trend with seasonality
- The series is stationary as no pattern appears in the residual plot
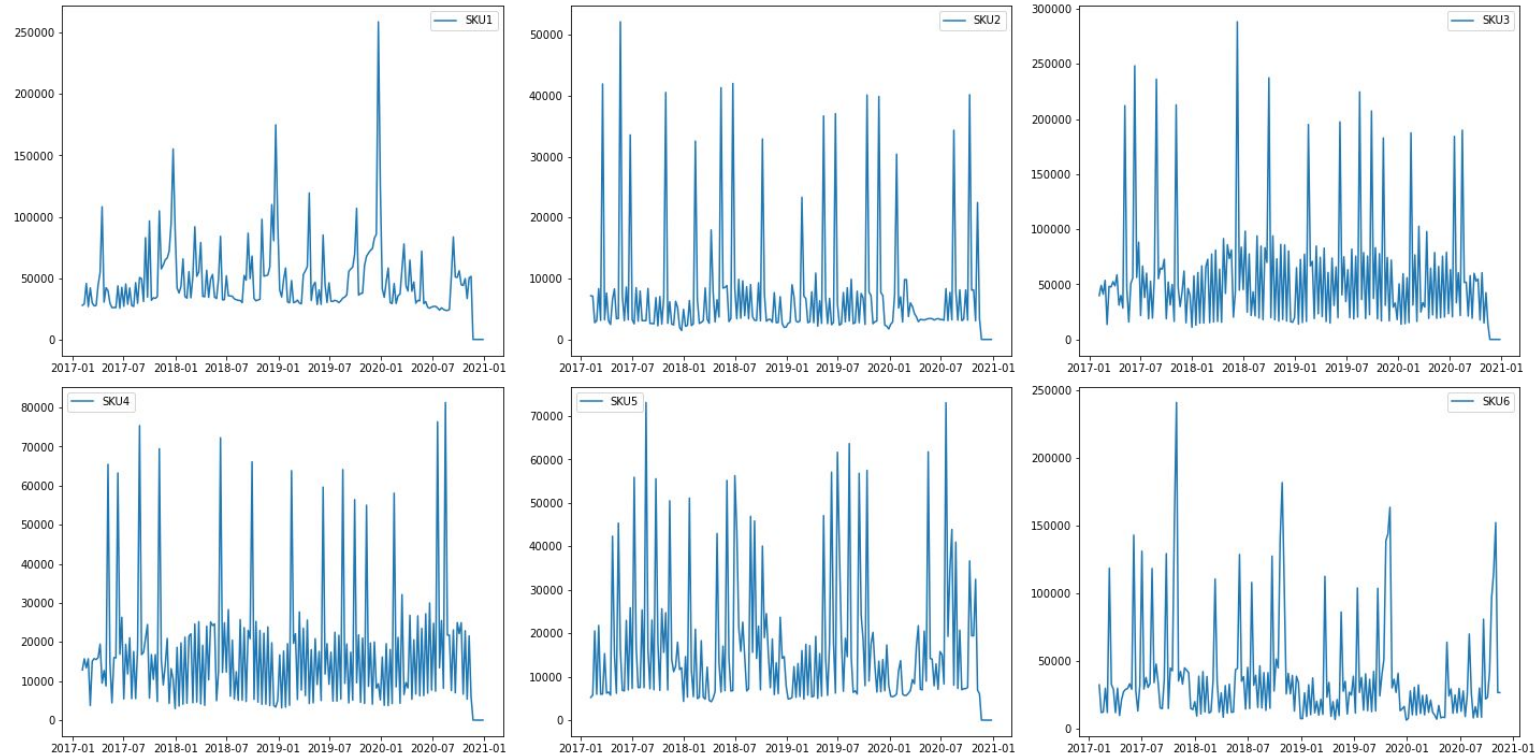
Exploratory Data Analysis

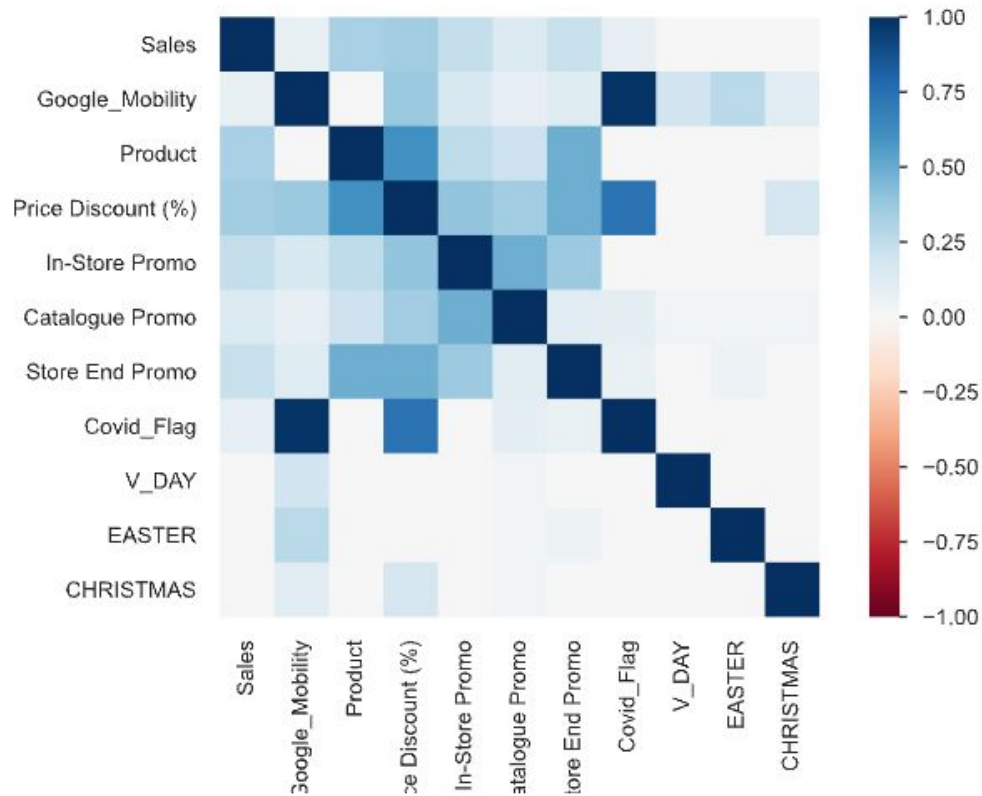● Here we first demonstrate the overall distribution of sales over time.

Exploratory Data Analysis

- Here we explore the overall distribution of sales by product over time.

Exploratory Data Analysis

Discover the correlation:

- Besides the last three features, Sales has positive correlation with all other features to different extents.

- The noticeable correlations:
  - Sales with Price Discount (%) - 0.347978
  - Product with Price Discount (%) - 0.603060
  - Covid_Flag with Price Discount (%) - 0.734857
  - Covid_Flag with Google Mobility - 0.982582

Exploratory Data Analysis
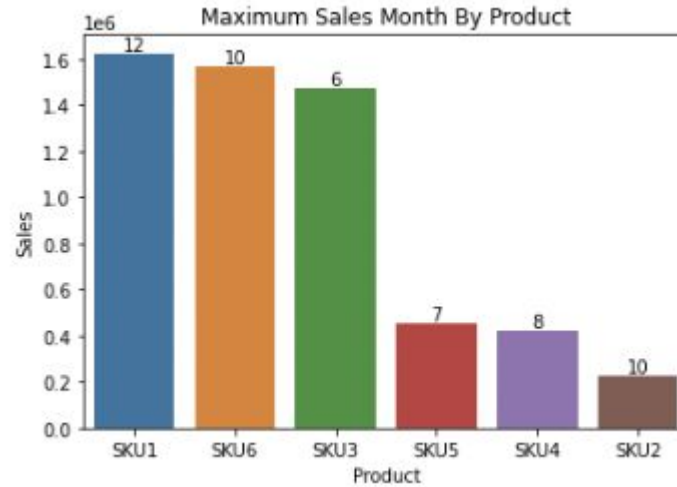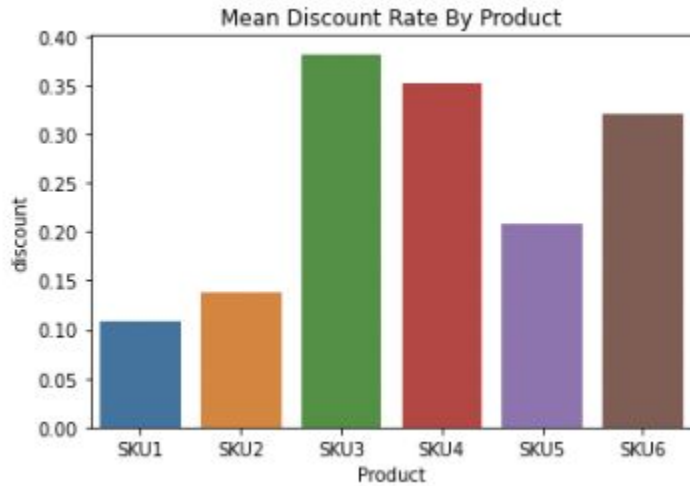
- From sales by product boxplot, SKU3 is the product that performs relatively better
- SKU6 and SKU3 are products which have higher sales because of heavy discount
- SKU1 is the product that rarely has discount but still incur high sales
- SKU4 has a wide range of discount but sales amount is bad
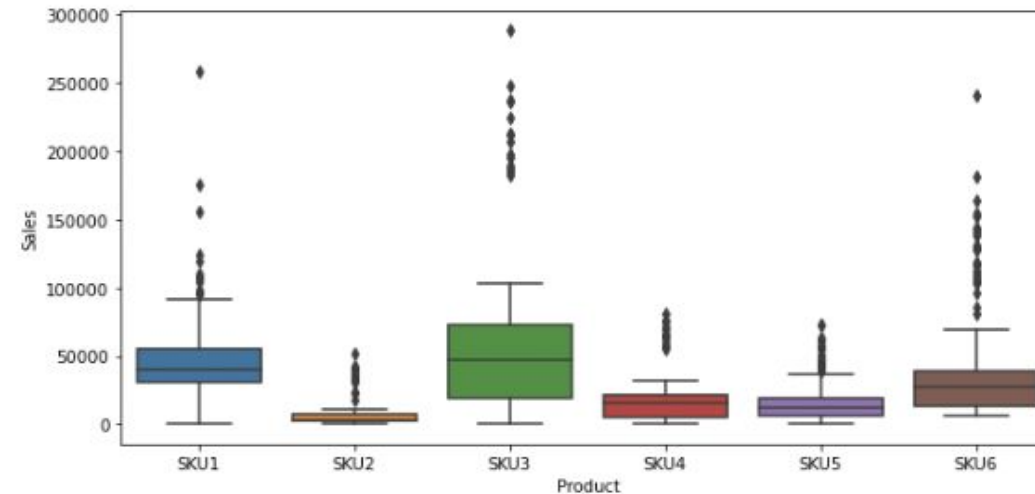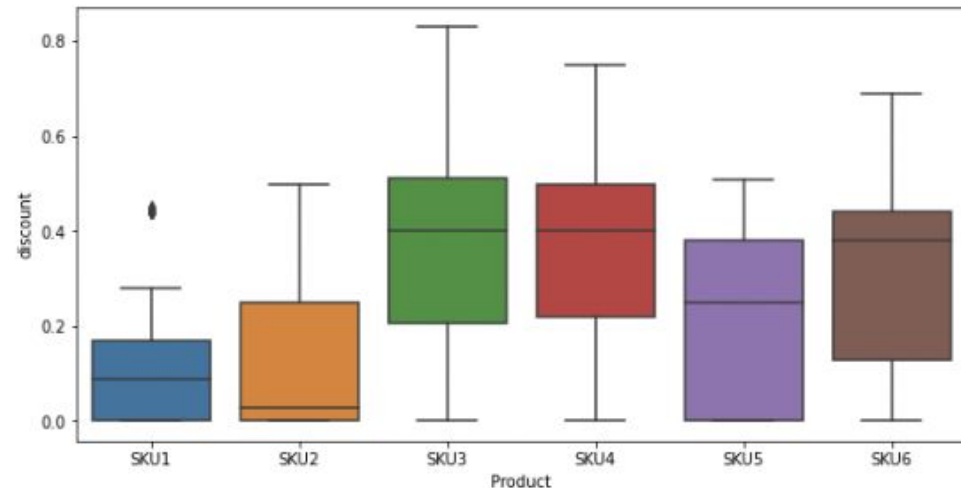
Recommendations

To sum up, we recommend to try the following models based on our EDA:
- ARIMA
- SARIMA
- Prophet (Cope with nonlinear trend with seasonality)
- DeepAR (Recurrent Neural Network)
- LSTM
- Linear Regression with external factors

Thank You