Retail Forecasting

Virtual Internship

16-Sep-2023

Data Glacier

Your Deep Learning Partner

Data Exploration

- 12 Features
- Timeframe of the data: 2017-02-05 to 2020-12-27
- Total data points : 1218

Assumptions:

- Outliers are present in Sales feature but due to unavailability of sales details ,we are not treating this as outlier.

- Sales from 2020-11-22 to 2020-12-27 of SKU1 to SKU5 are 0 and the features of SKU6 in this timeframe are missing, so we dropped those data.

SKU1   SKU2   SKU3   SKU4   SKU5   SKU6
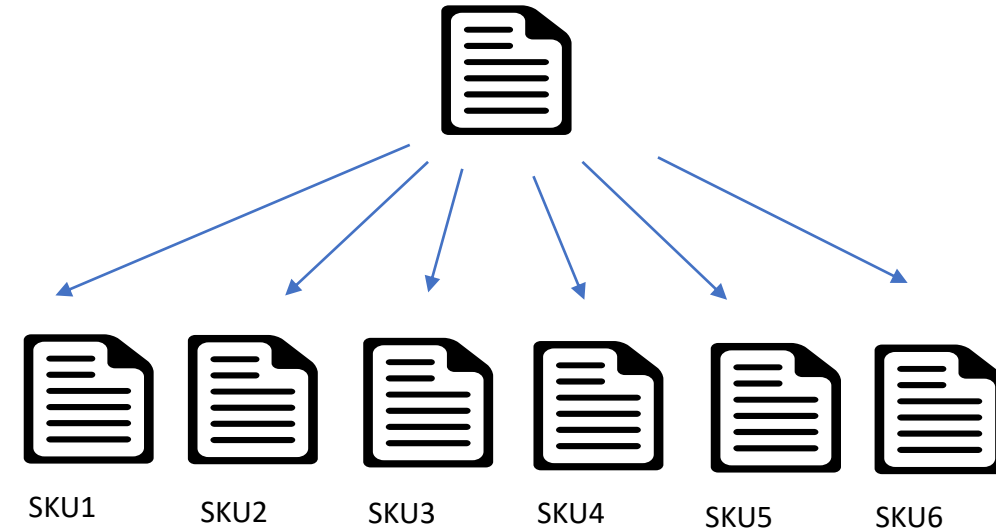
Area of investigation 1: Stationarity

Area of investigation 1: Stationarity

Sales of SKU5

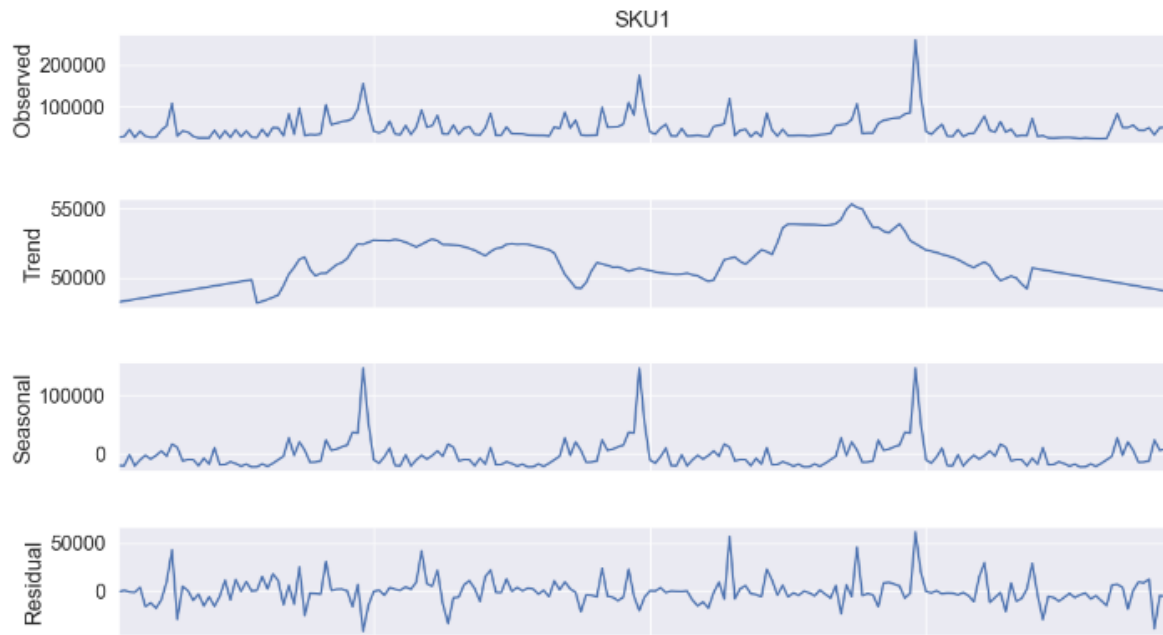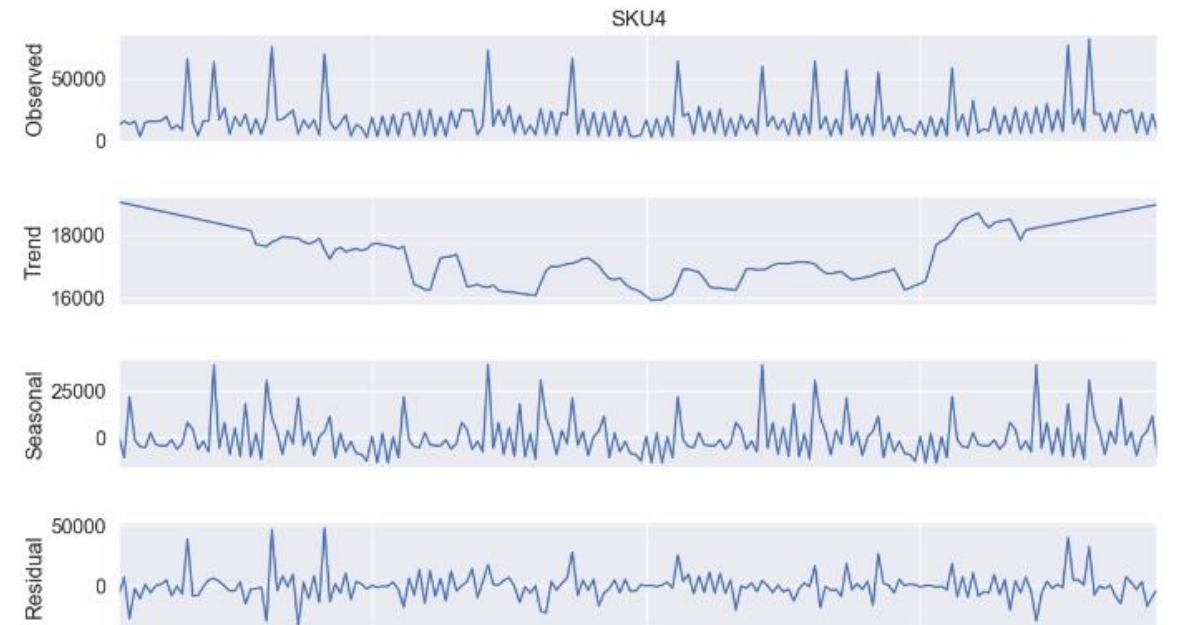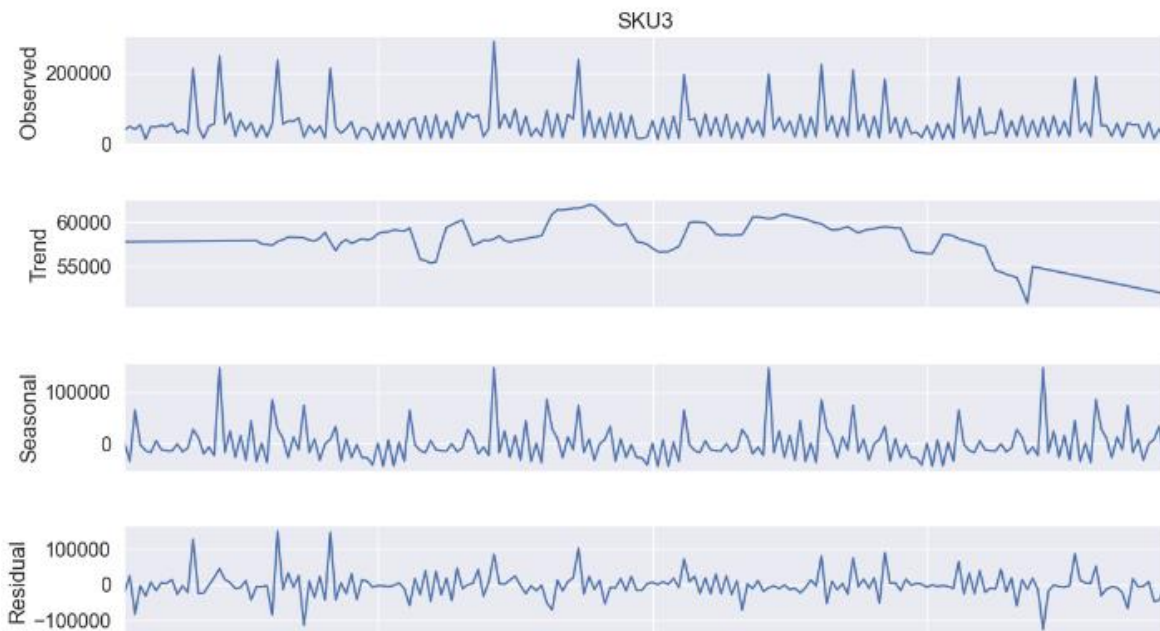Sales of SKU6
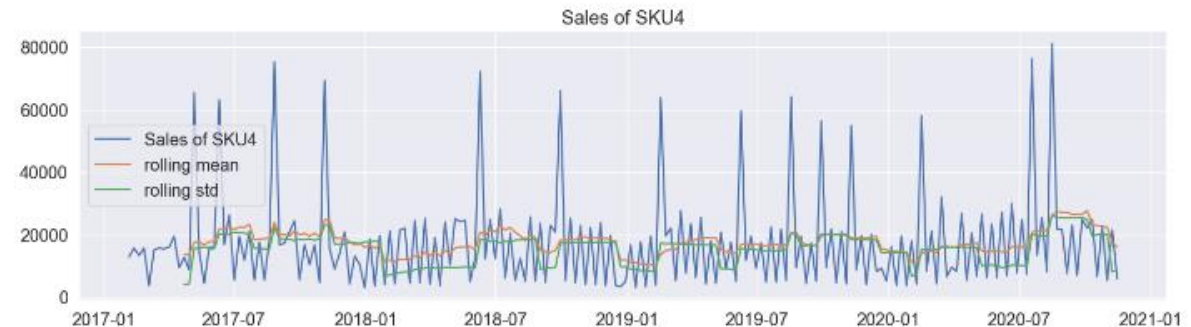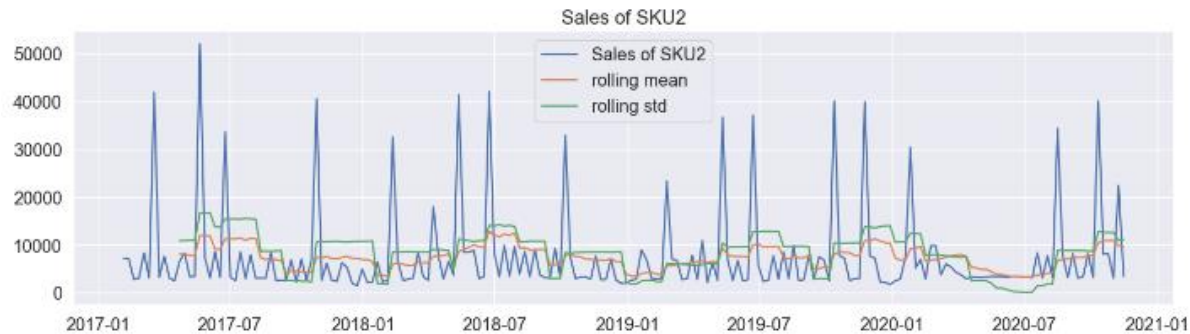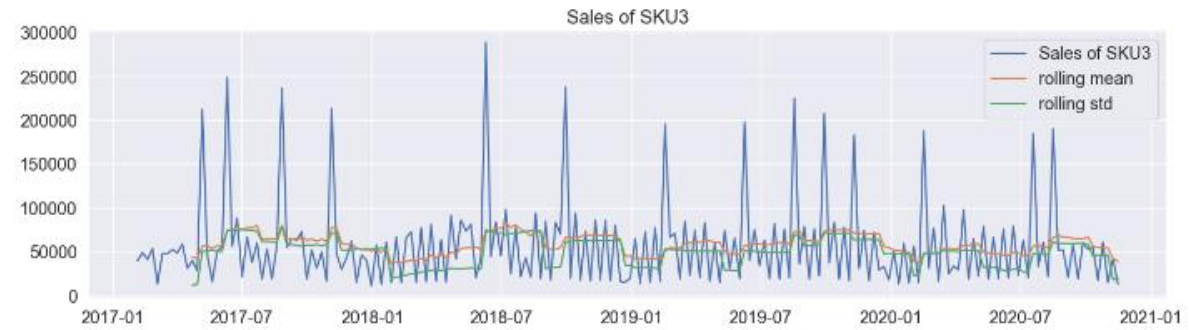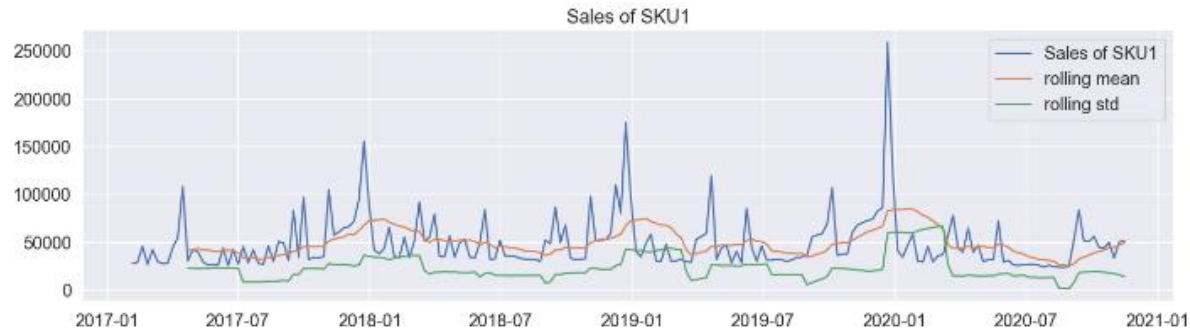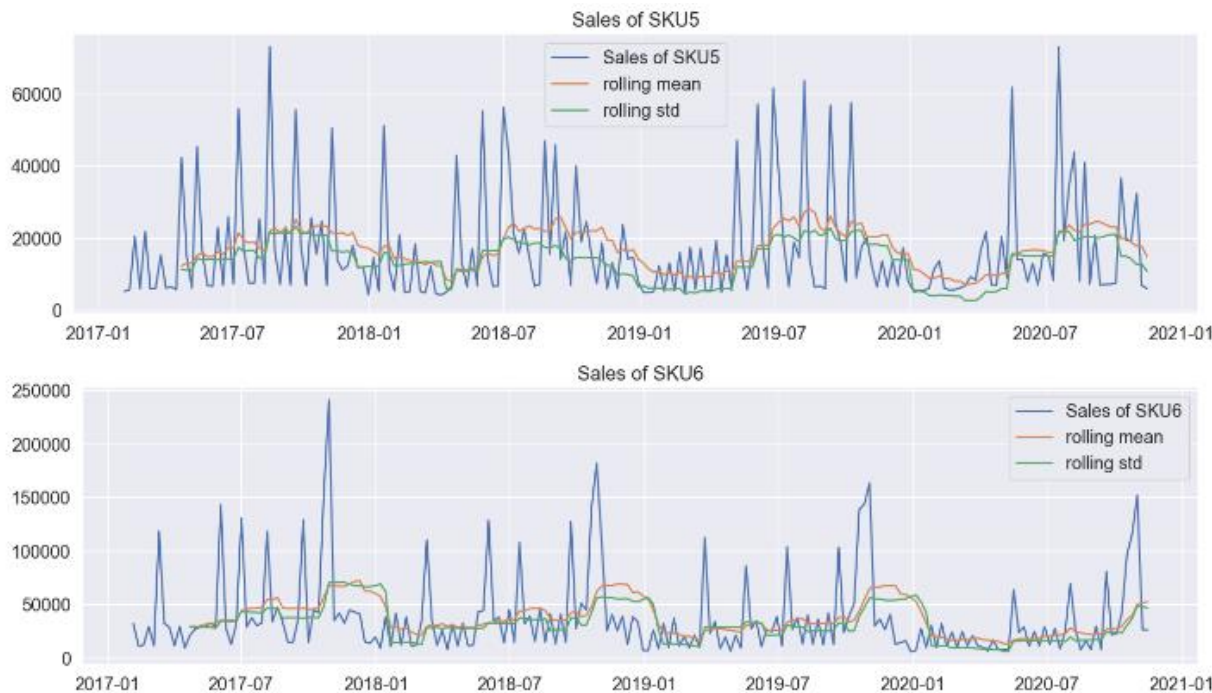
From the graph, mean and standard deviation of SKU 1-6 do not vary much in time. We could also use the Augmented Dickey Fuller test to determine the stationarity of the time series data. Since the data are stationary, we could use Holt-Winters' Seasonal Method and SARIMA to model the data
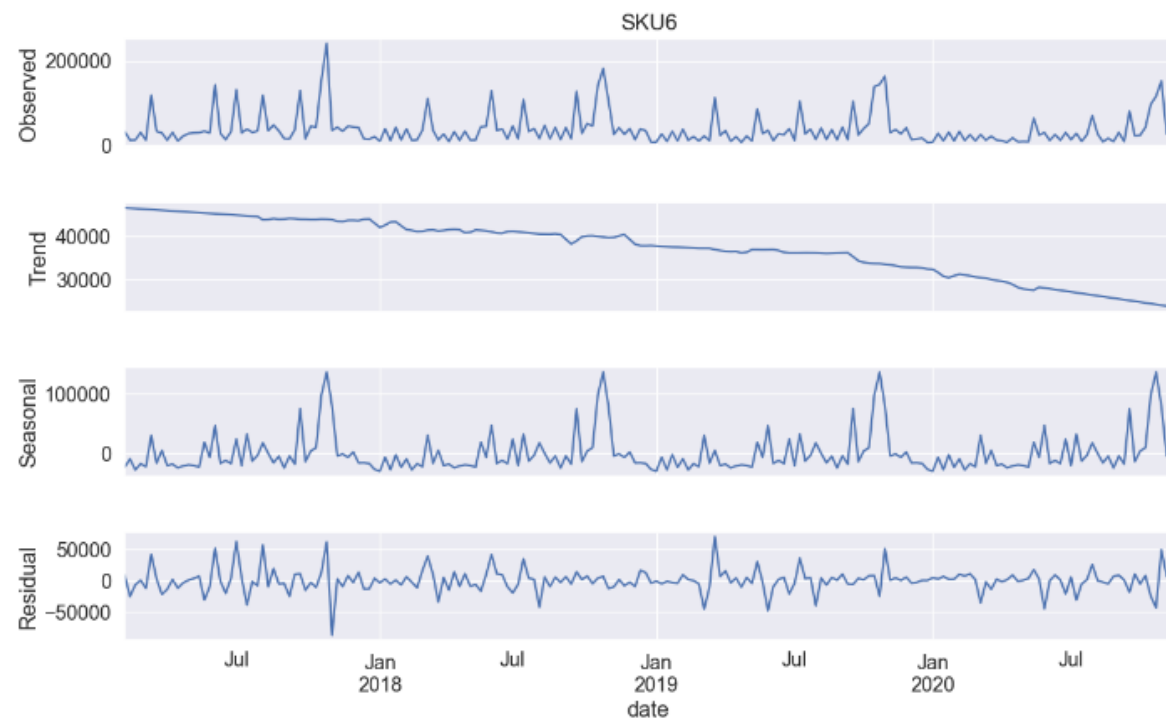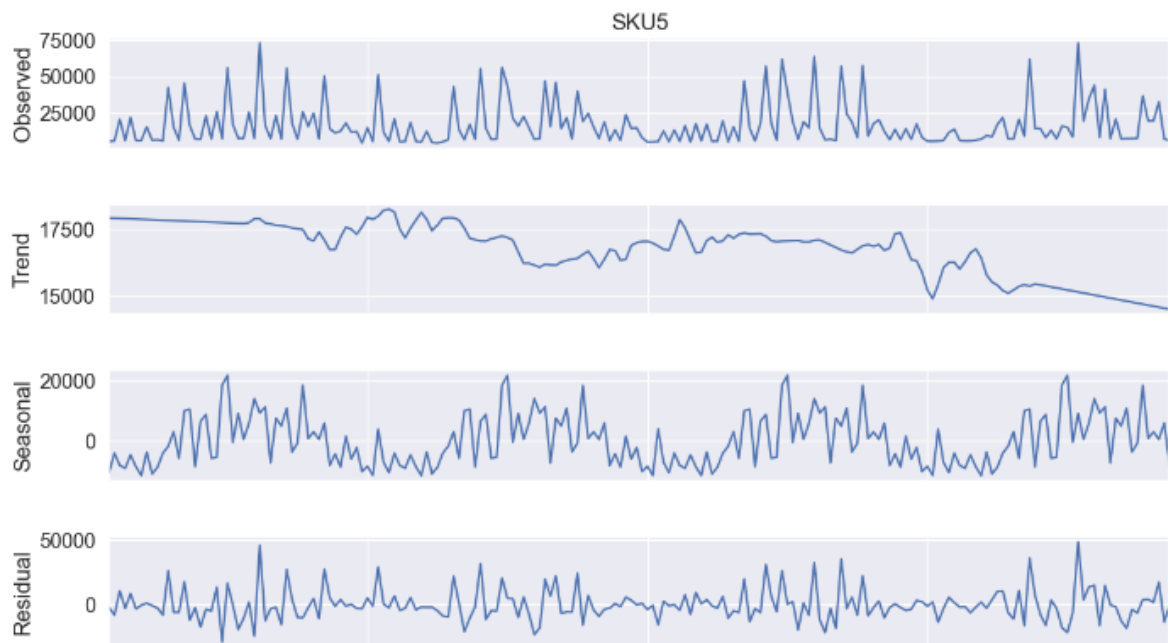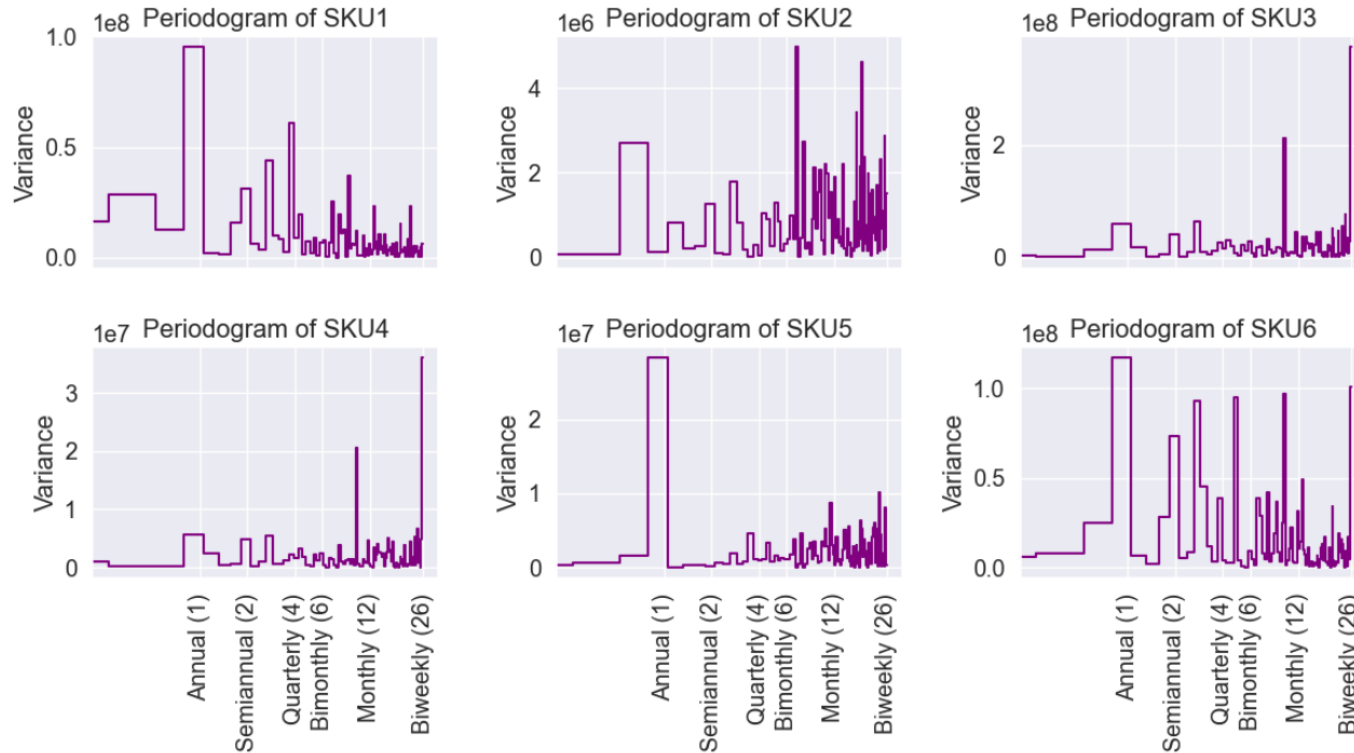
Area of investigation 2: Trend and Seasonality

Area of investigation 2: Trend and Seasonality

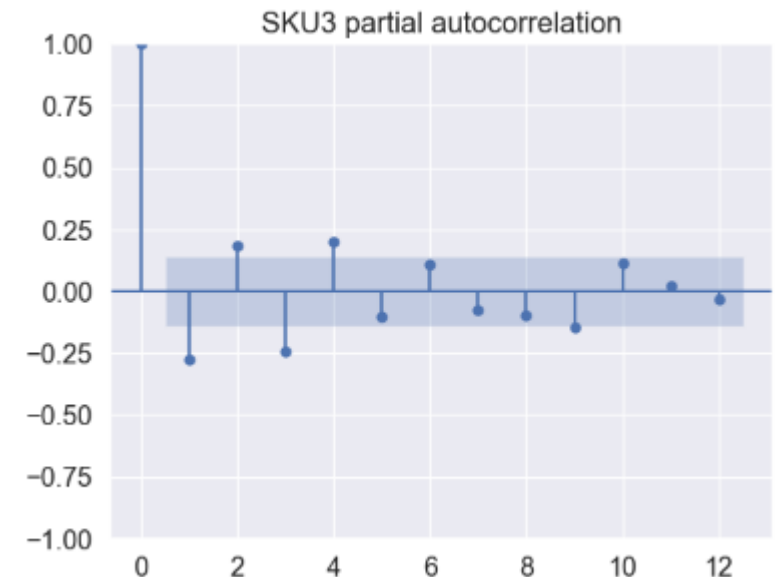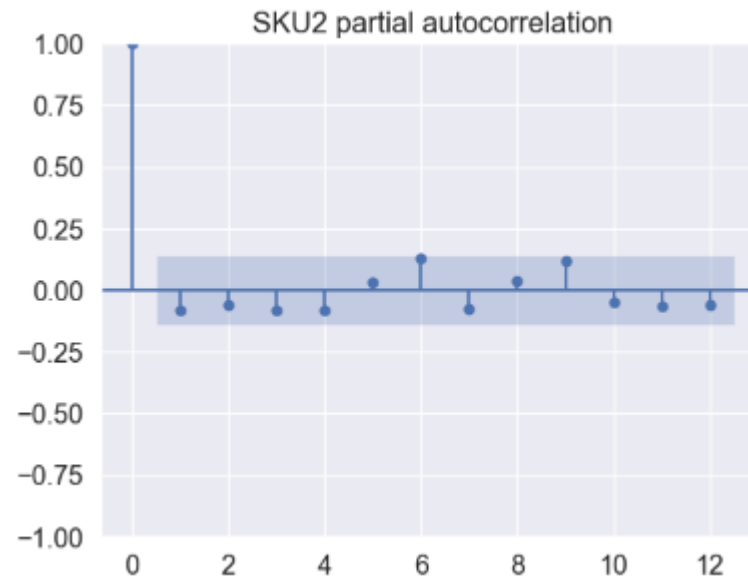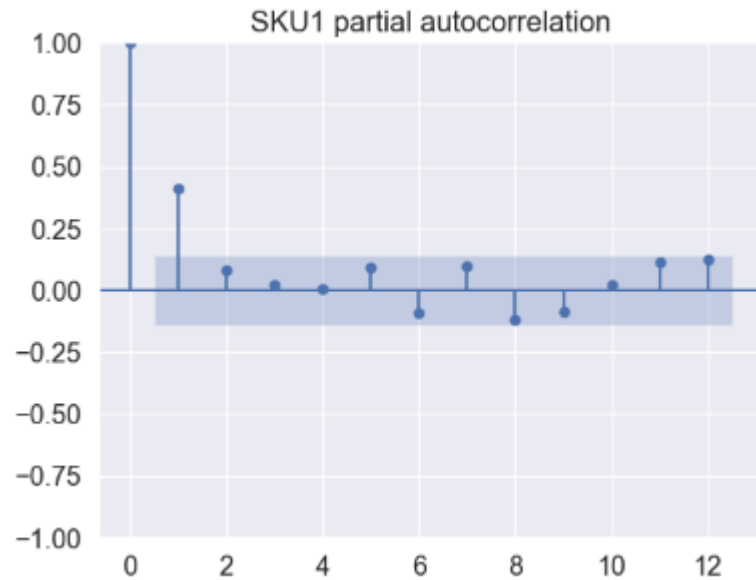# Area of investigation 2: Trend and Seasonality

Periodogram

To use machine learning, we add features that simulate the trend and seasonality of the time series data. We use fourier feature to simulate the feature of seasonality.
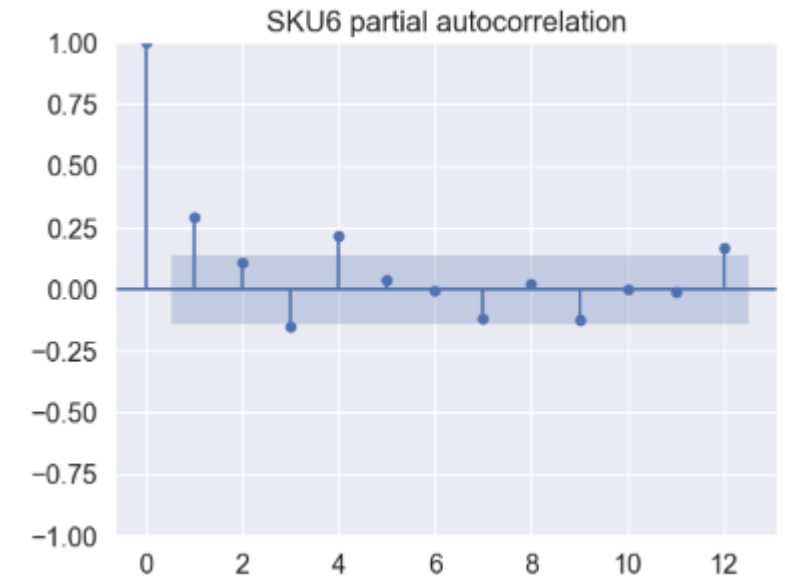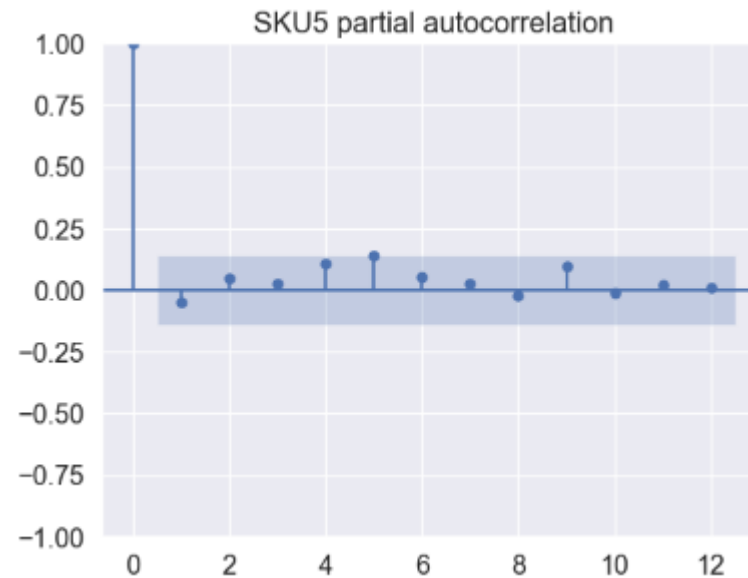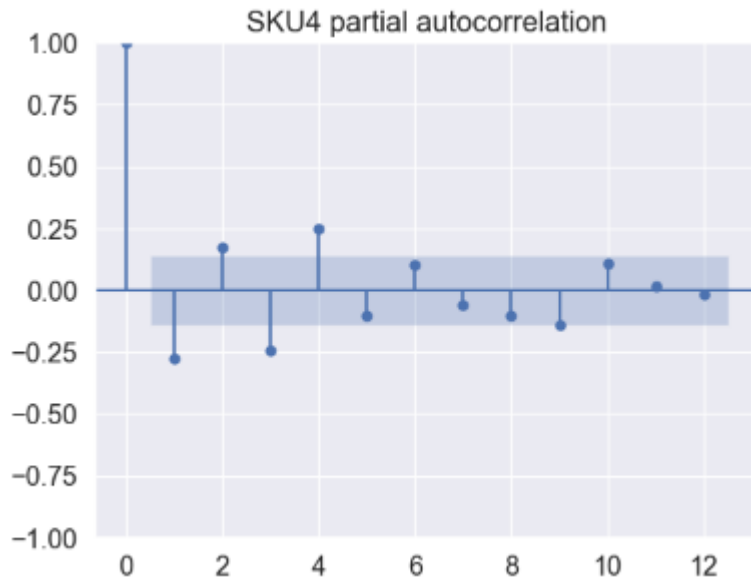
Laggings

Laggings
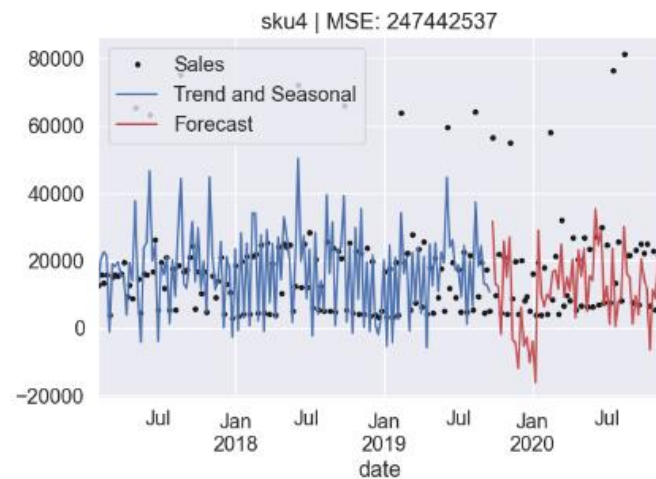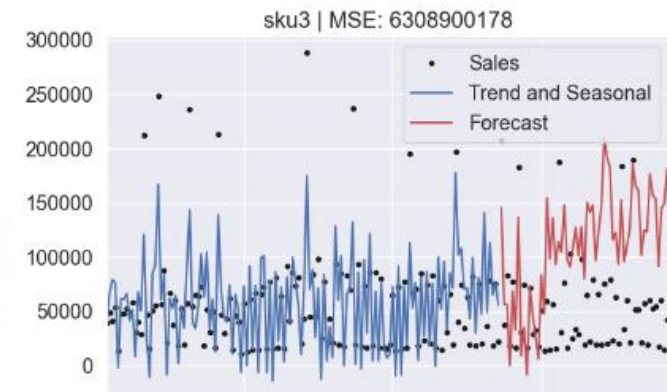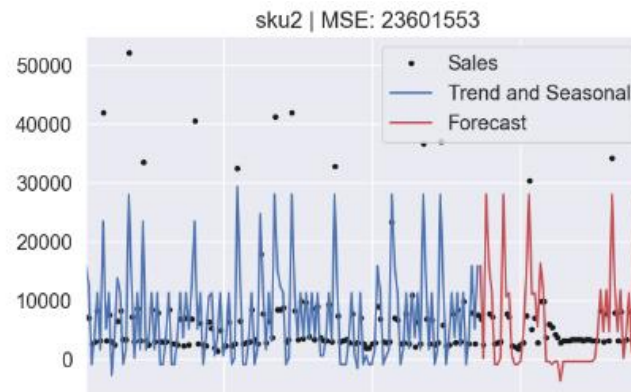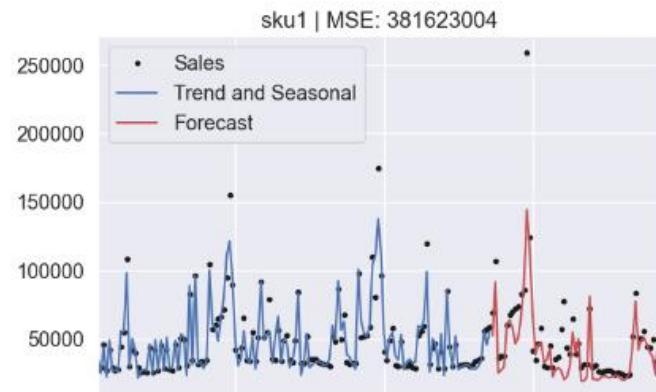

Laggings time series data could be useful for modelling. From the partial autocorrelation plot, lag1 is significant to SKU1 and SKU6, and lag1-4 are significant to SKU3 and SKU4

Linear Regression
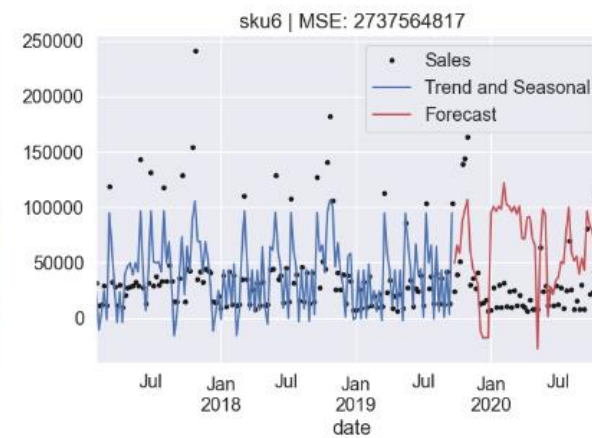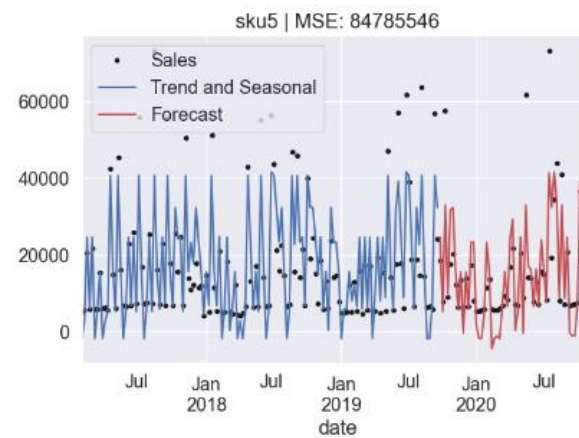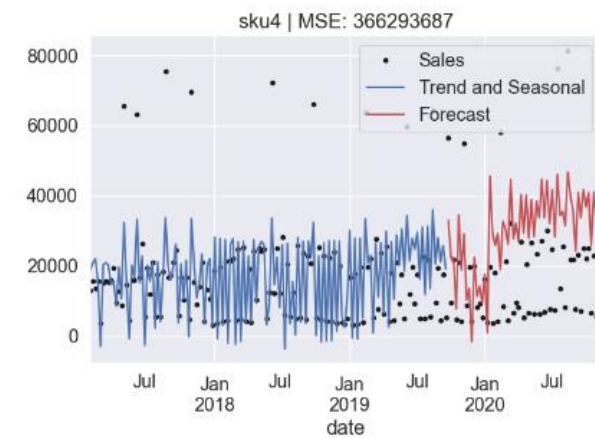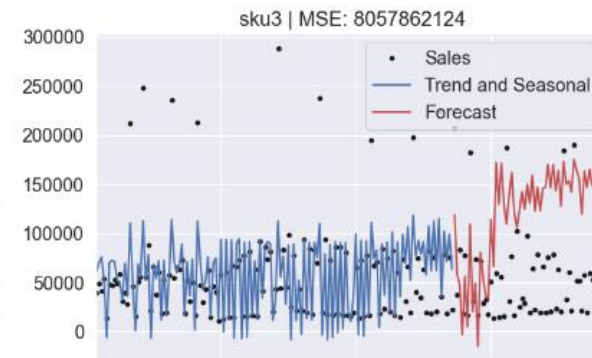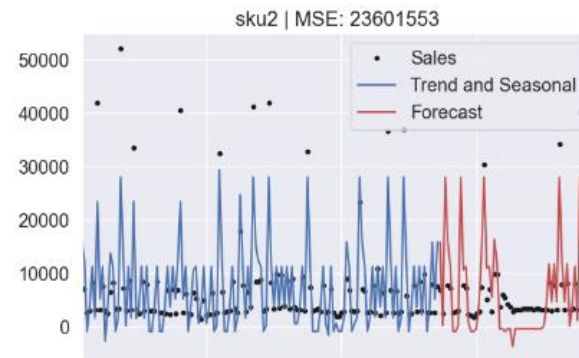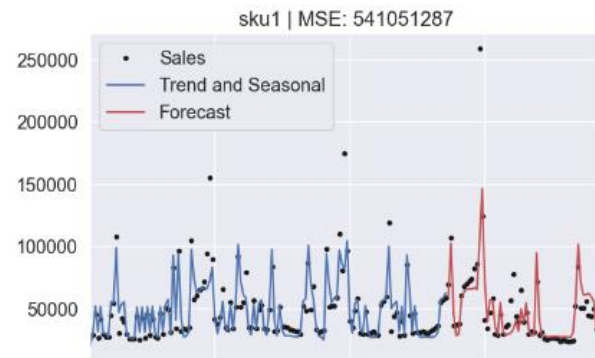
Linear Regression

Hyperparameters tuning

Thank You