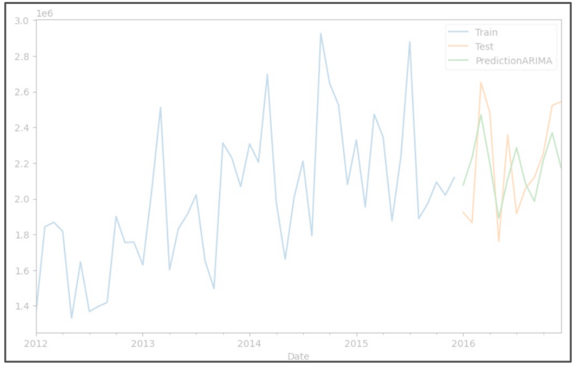
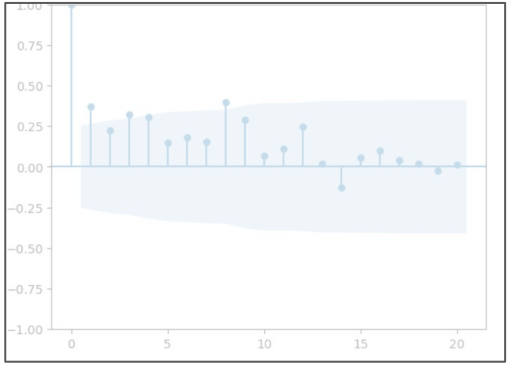
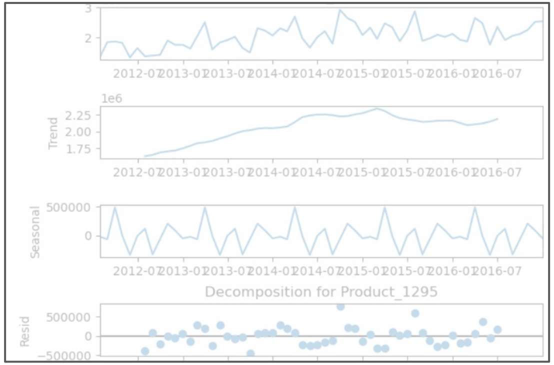
Demand Forecasting

Using ARIMA/SARIMAX Time Series Models





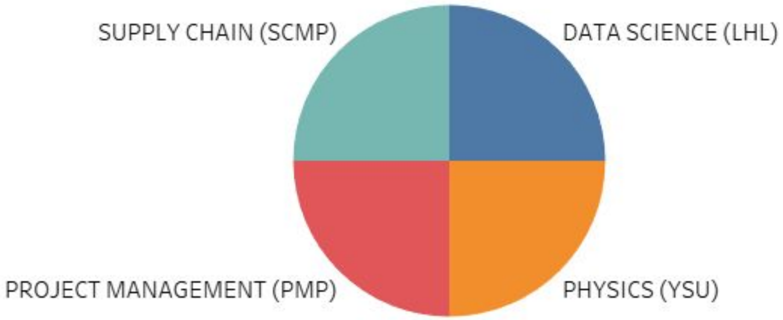


Garegin Manvelyan

October 10th, 2023

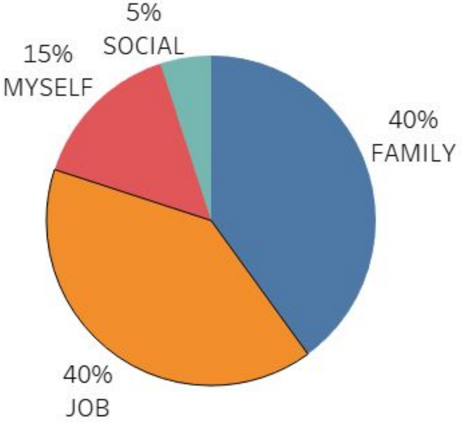
GAREGIN MANVELYAN

EDUCATION

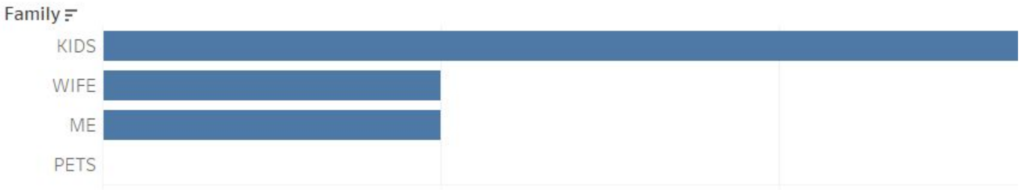


JOB

LIFE



FAMILY



PROJECT EXECUTION STEPS

- **Business Value** [Preventing Overstocking & Stockouts]
- **Data Gathering** [Kaggle online platform]
- **Preprocessing** [Cleaning, EDA, Reduction, Splitting]
- **Model Application** [ARIMA/SARIMAX Time Series Models]
- **Interpretation & Challenges** [Results]
- **Documentation** [Readme, Presentation]

GOAL

Business Value [Preventing Overstocking & Stockouts]

The Aim is to Develop an Improved Demand Forecast Model for Inventory Optimization and 10% Cost Savings.

DATASET SELECTION

Data Gathering [Kaggle online platform]

The "Forecasts for Product Demand" dataset contains historical product demand for a global manufacturing company with thousands of products across various categories.

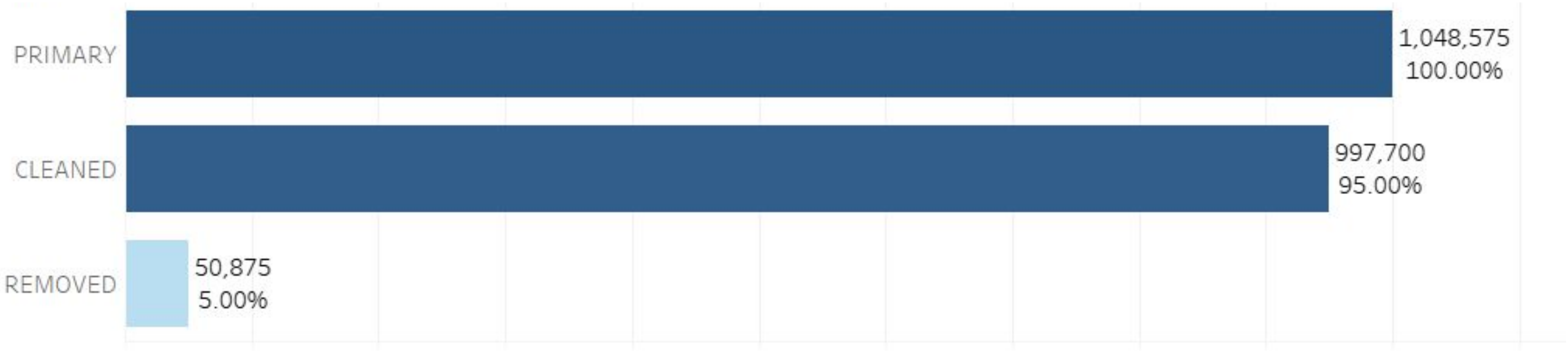
Records :	1 048 575
Time Period :	5 years / 60 months
Unique Products:	2 160
Warehouses:	4
Product Categories:	33

Dataset was downloaded from kaggle online platform

Data Preparation

Preprocessing

[Cleaning]



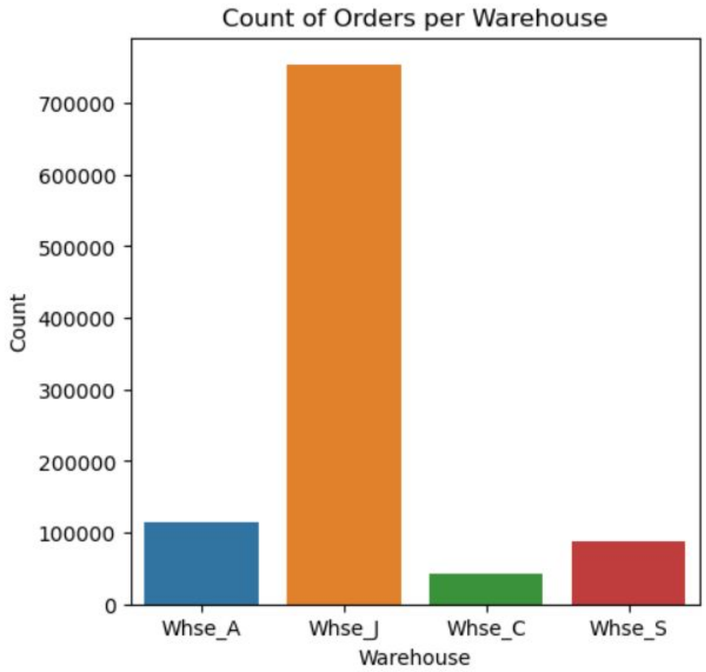
Null Values: **11 239** ....removed
Data Types : **2** .... converted
Single Negative Values: **26** ....removed
Negative/Positive Pairs:**11 746** ....removed

Rows after cleaning : **997 700**

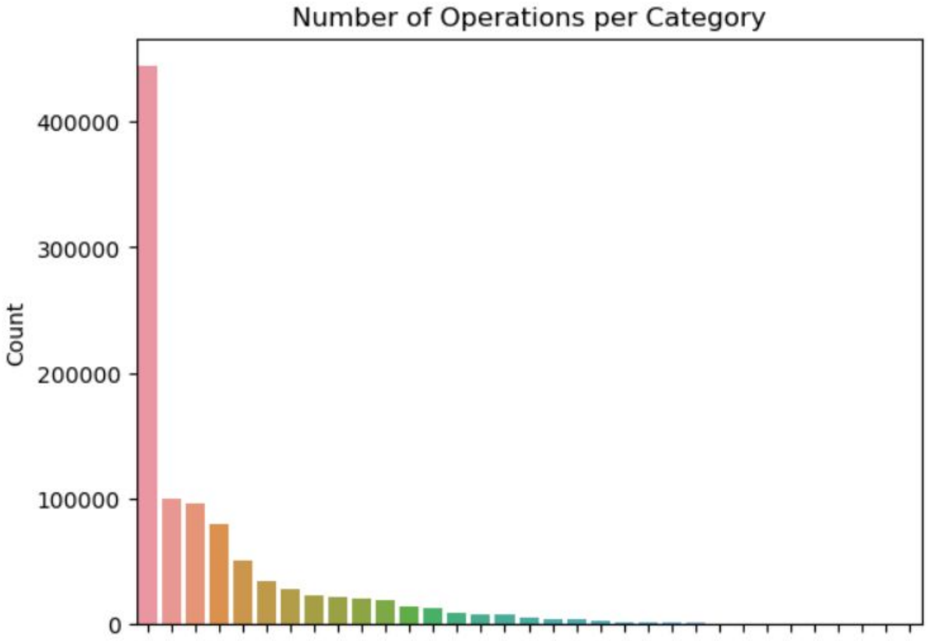
Data Preparation

Preprocessing [EDA]

Busiest warehouse : “J” (76%)



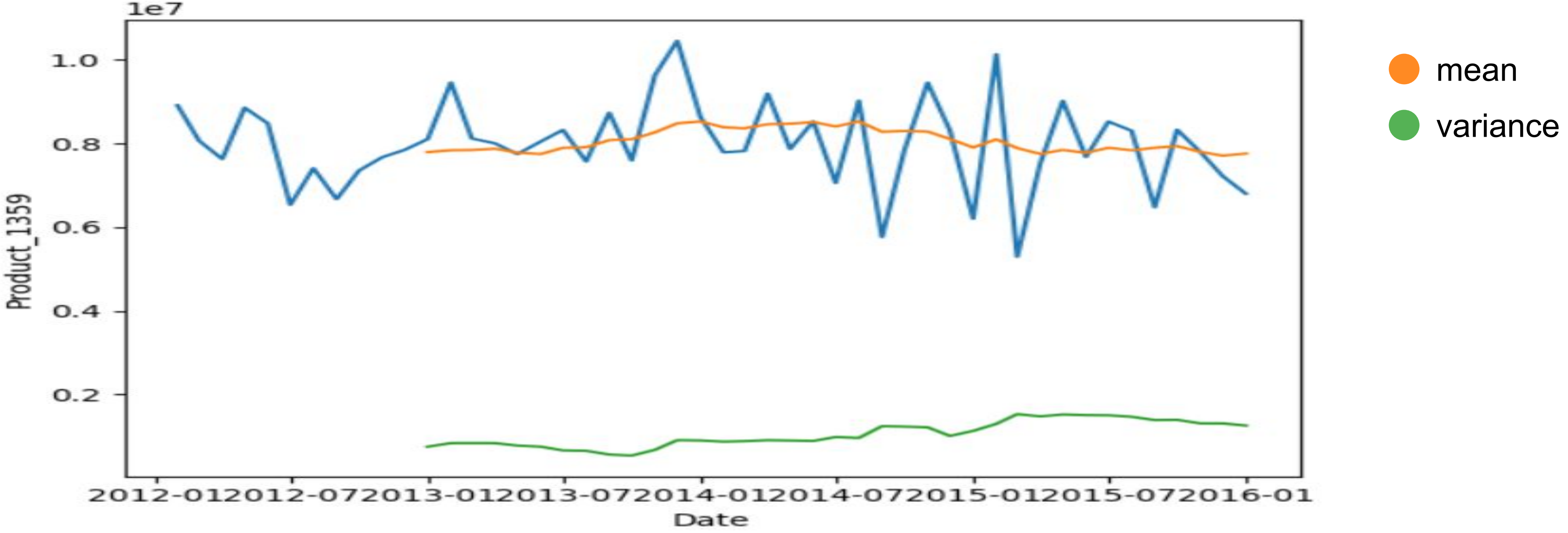
Most used product category : “#19” (43%)



Data Preparation

Preprocessing [EDA]

Demand monthly plot over 5 years & Stationarity Check



Data Preparation

Preprocessing

[reshaping]

Pivoting Data for Product-Level Analysis*

Product_Code	Warehouse	Category	Date	Demand
Product_0965	Whse_A	Category_006	2011-11-18	1
Product_0504	Whse_J	Category_015	2011-12-05	1
Product_2165	Whse_C	Category_024	2011-12-06	1
Product_1699	Whse_J	Category_026	2011-12-07	1
Product_1680	Whse_S	Category_021	2011-12-09	1
Product_0965	Whse_A	Category_006	2011-12-16	1
Product_1757	Whse_J	Category_001	2011-12-20	1
Product_0609	Whse_J	Category_001	2011-12-20	1
Product_0620	Whse_J	Category_001	2011-12-20	1
Product_0620	Whse_J	Category_001	2011-12-21	1
Product_0258	Whse_J	Category_001	2011-12-21	1
Product_0260	Whse_J	Category_001	2011-12-21	1

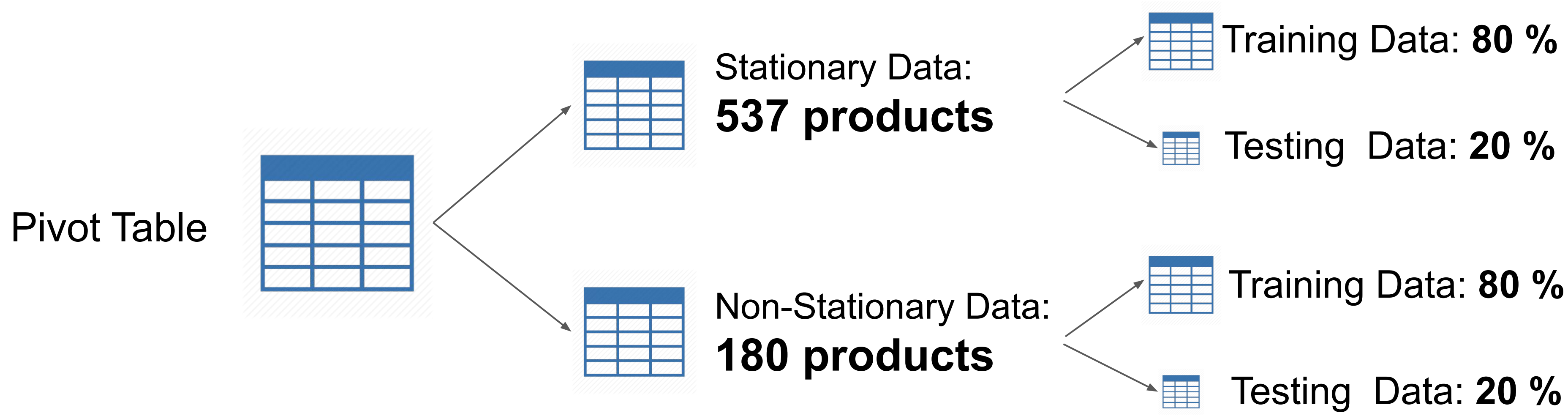
Date	Product_1359 Whse_J	Product_1360 Whse_J	Product_1367 Whse_J	Product_1368 Whse_J
2012-01-31	8910000	1178000	11300	3300
2012-02-29	8061000	1162000	7300	5500
2012-03-31	7625000	1222000	13100	6100
2012-04-30	8850000	864000	12200	5800
2012-05-31	8475000	937000	6700	2100
2012-06-30	6531000	1113000	11500	4500
2012-07-31	7406000	1021000	11400	4800
2012-08-31	6667000	827000	8200	5500
2012-09-30	7349000	957000	9300	10200
2012-10-31	7668000	780000	11500	5600
2012-11-30	7843000	1319000	13900	1800
2012-12-31	8102000	920000	12800	2200
2013-01-31	9455000	885000	13800	1500

*drop NaN and 0 after transformation

Data Preparation

Preprocessing [splitting]

Split Data to Stationary and Non-Stationary Datasets*



Time Series Model Application

ARIMA
time series model

[pdq -hyperparameters]

p,d,q hyperparameters (explanation)

p (AutoRegressive order): quantifies how many past time steps are used to predict the current value

Description	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Product_1359 Actual	150	100	25	15	20	54	220	260	300

p=2

q (Moving Average order): quantifies how many past forecast errors are used to predict the current value

Description	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Product_1359 Actual	150	100	25	15	20	54	220	260	300
Product_1359 Predicted	125	110	30	14	21	52	240	280	370
Forecast Error	25	-10	-5	1	-1	2	-20	-20	-70

q=5

d (Integration order): represents the number of differences needed to make the series stationary

Description	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	
Product_1359 Actual	150	100	25	15	20	54	220	260	300	
Product_1359 Differ by 1 Month	100	25	15	20	54	220	260	300	<==	**1 step left**
Differenced Data	50	75	10	-5	-34	-166	-40	-40		

d=1

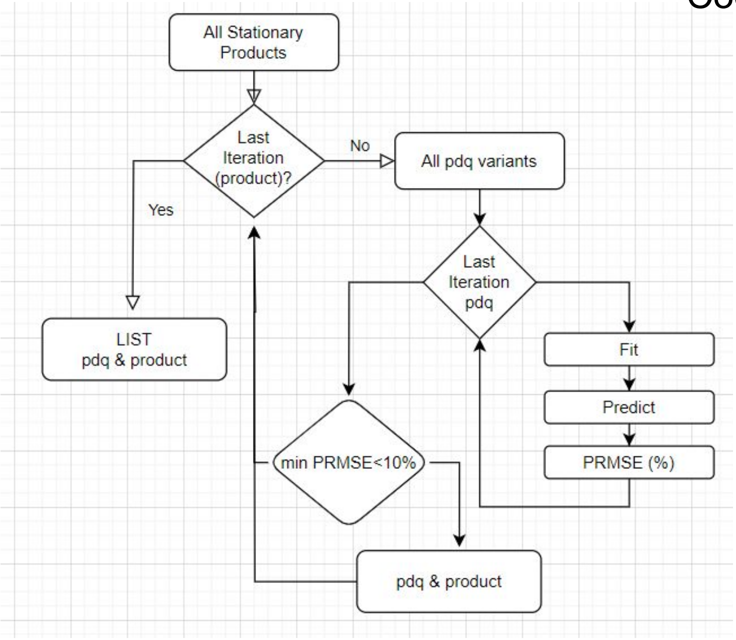
Time Series Model Application

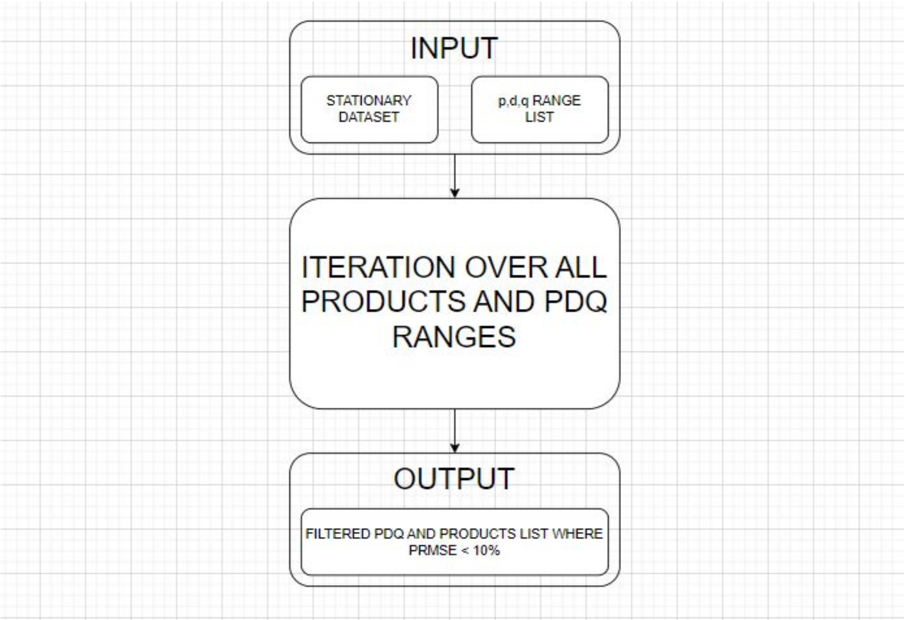
ARIMA time series model

[pdq -hyperparameters]

Get optimal p,d,q hyperparameters for all products in stationary data

Code Flowcharts



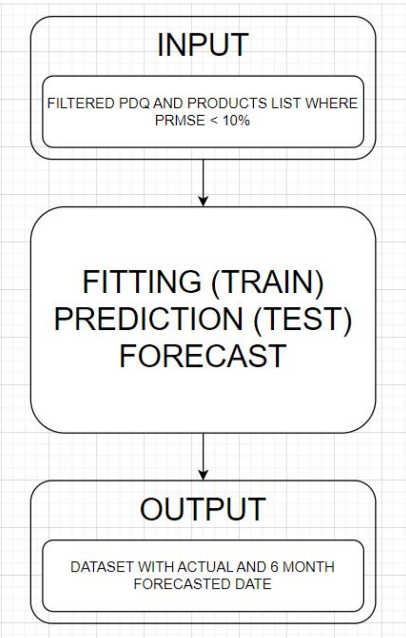


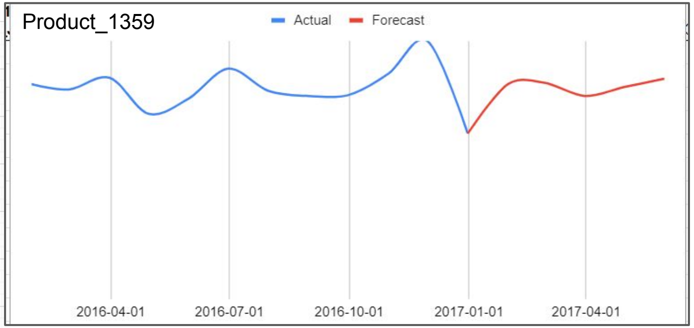
Time Series Model Application

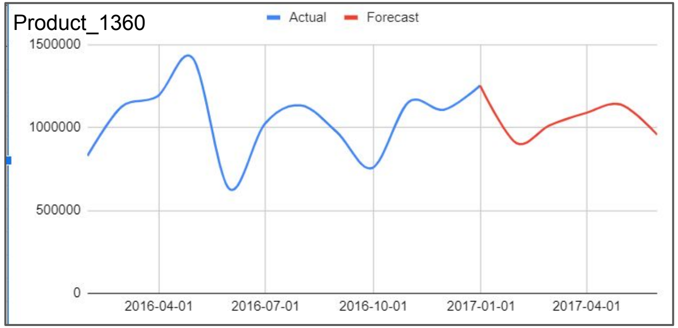
ARIMA
time series model

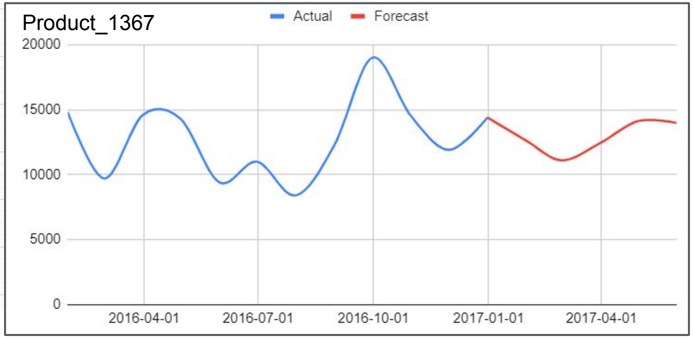
[pdq -hyperparameters]

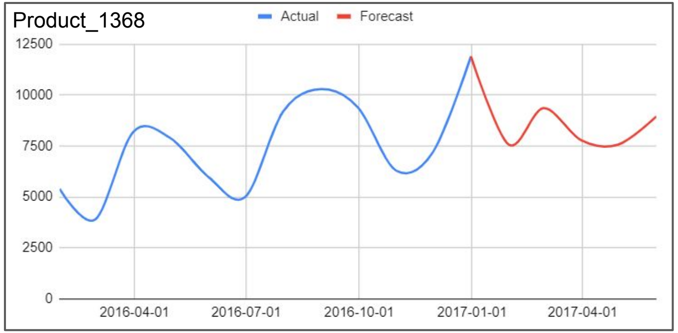
Initializing ARIMA model to Forecast Demand for Specified PDQ and Error<10%











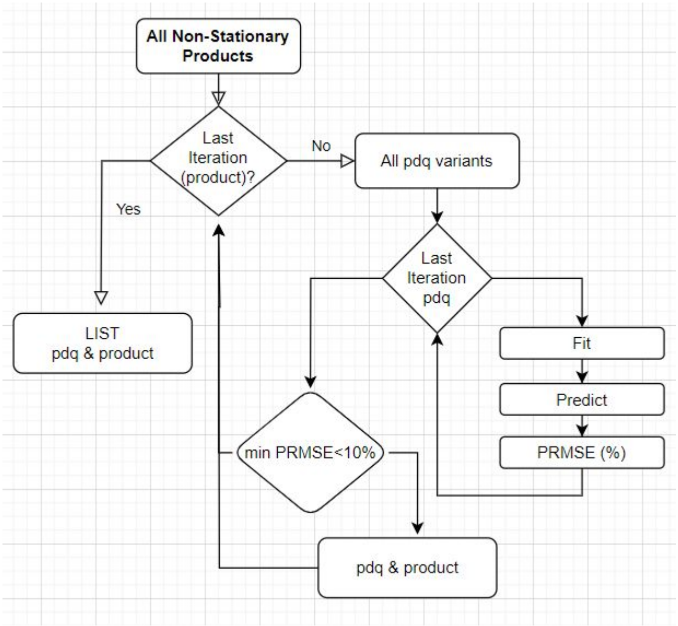
Time Series Model Application

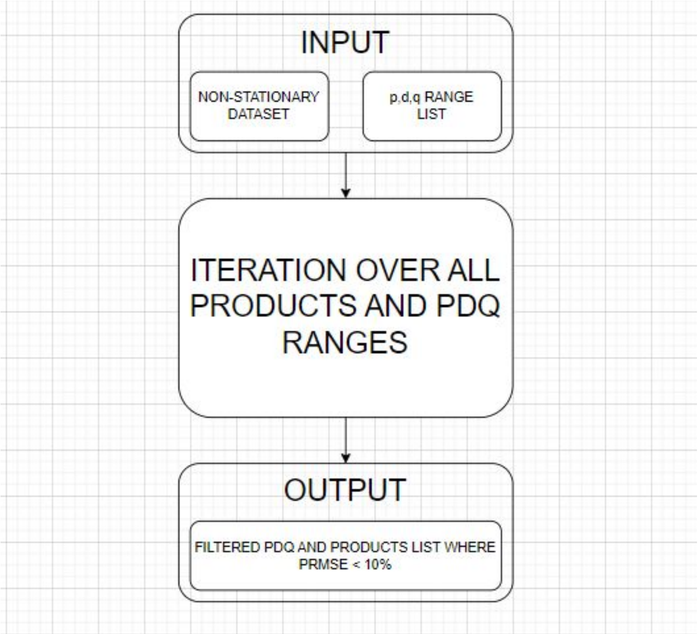
SARIMAX
time series model

[pdq -hyperparameters]

Get optimal p,d,q hyperparameters for all products in NON-stationary data

Code Flowcharts



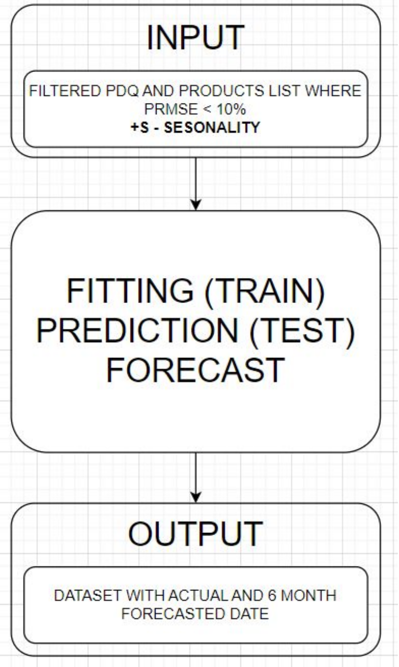


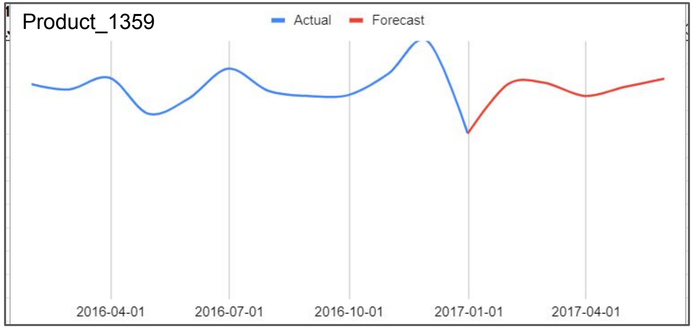
Time Series Model Application

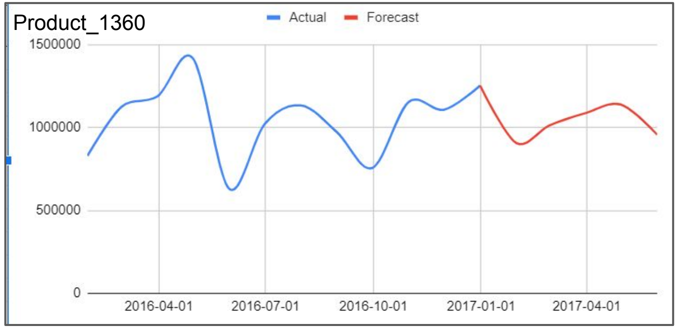
SARIMAX
time series model

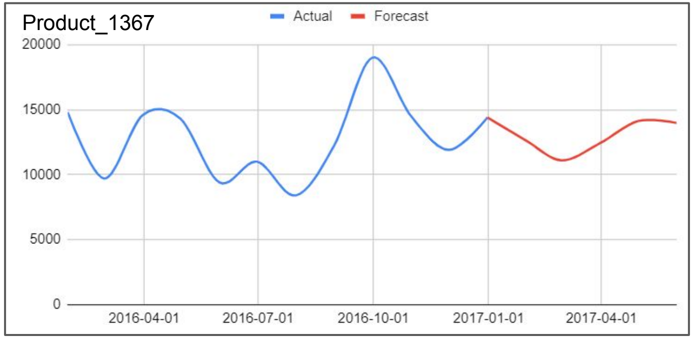
[pdq -hyperparameters]

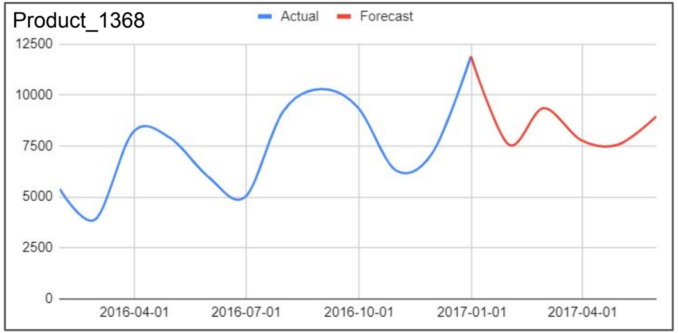
Initializing SARIMAX model to Forecast Demand for Specified PDQSand Error<10%











Results / Challenges

1) **Arima and Sarimax models demonstrate positive performance and can provide a 10% improvement in forecast accuracy compared to a simple naive forecast.**

2) **Despite positive performance, the models gave accurate results for only 30% of the products.**

3) **Hyperparameters tuning required big amount of time.**

ARIMA [580 items] : Elapsed time: 5 hours

SARIMAX [180 items] : Elapsed time: 57 hours

Future Goals

Implementation of Models on the Provided Time Series Dataset

1) **Random Forest Model**
2) **XGBoost Model**
3) **Multivariate Time Series Analysis**

Thank you!

Garegin Manvelyan

gareginm@gmail.com | Github | LinkedIn | 639-994-5999