Talaria Forecasting - Financial Forecast

Sandesh Brand 4 - Discovery

By: Rishabh Singhal

Introduction:

My model is capable of predicting a wide range of values. I have submitted the **.ipynb** file that contains the code written in the language python 3.

For running this notebook, library requirements are - **Pandas**, **NumPy**, **Matplotlib**, **Sklearn**, **Statmodels**(for installation use **conda install -c conda-forge statsmodels**).

I have written separate codes for different variables starting from forecasting for -

' Leopard Volume Closing Base',
' Leopard Volume Leavers',
' Leopard Volume Gross Adds',
' Panther Volume Closing Base',
' Panther Volume Leavers',
' Panther Volume Gross Adds',
' Hyena Volume Closing Base',
' Hyena Volume Leavers',
' Hyena Volume Gross Adds',
' Panther - Leopard - Hyena Revenue Total Revenue'

in a serialized manner.

Let us label them as **variable1**, **variable2**, **variable3**, **variable4 upto variable10**.

So before deciding the model for training them I have followed some basic steps for each variable step :

- Created a new Dataframe d having columns time and sales
- For each variable, extract the horizontal row and store them in the sales column in the Dataframe d.
- Store time corresponding to each value in the time column in the dataset d.
- After doing the previous step, I have converted the above column into an index by using set_index, after which time values will be the index of the Dataframe d.
- After this, using the describe function I have analyzed the dataset.
- After this, I have plotted the Dataframe d.

For variable1 :

- After plotting, we observe that the curve is following a linear pattern, so I used a linear regression model to make the predictions.
- I have also plotted prediction-actual value which we found out is very less, which validates our approach.
- I also find that our model is stationary, which again validates our approach.
- I also predicted upto the next 12-time stamp forecasting values and labeled it as x1.

For Variable2, Variable3 upto Variable10:

- Each feature was analyzed using the Augment Dickey-Fuller test to determine if the data had a trend.
- For all these variables SARIMAX model is used for forecasting with the same prediction steps with a little variation in parameters.
- I have also plotted prediction-actual value which we found out is very less, which validates our approach.

- All of the prediction values are stored in the dataframe name **'Prediction'** in the format as asked in the problem statement and stored the values of submission in the **required format** and then downloading it as **'Prediction.xlsx'** file.

Robustness:

- I have submitted the required sliding windows file for robustness calculations in the required format.
- I have also built two robustness function one for linear regression and other for Sarimax which will save result in the sliding windows file.