Willow Cash Flow Prediction Challenge - 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 find the best parameters for different variables for forecasting using MAPE (Mean Absolute Percentage Error) Analysis.
These parameters are stored in the CSV file "Parameters_complete.csv". I have submitted this CSV file in the submissions.

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

- Find all the unique variables of Market-mapping.
- Separated all variables then accordingly and then analyzed them by plotting their curve.
- If the size of a given variable is greater than seven then -

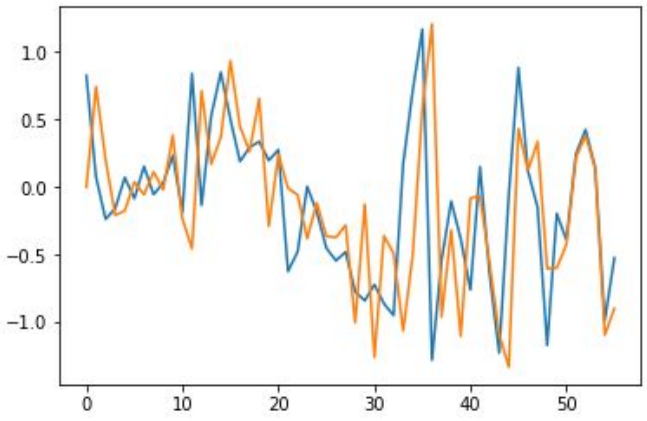
- Find the best parameters for which MAPE error of the last five values is minimum. (While training the model initially these last five values were not included but during final prediction, the complete dataset for training/fitting was used)

Otherwise -

- All values were used to train and MAPE error of complete data is calculated.

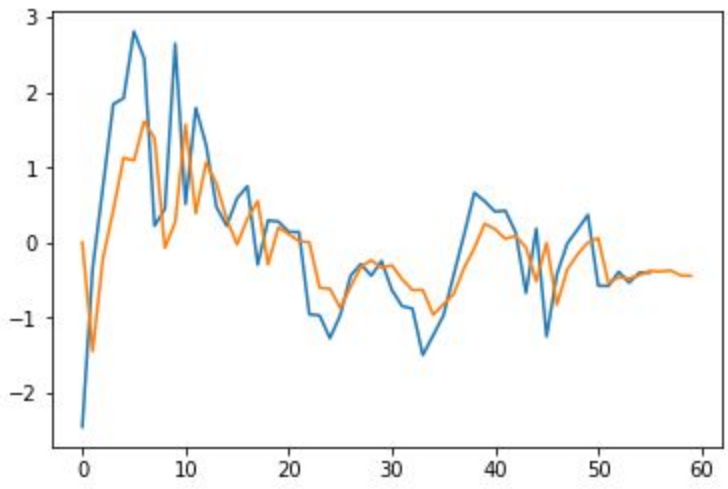
- After this, I have plotted the initial training set forecast, which was found to be very close to the actual value which validates our approach.

```
56 E16 Payables and Accruals
for these parameters -  (1, 1, 0) n (1, 1, 1, 9)  this -  21.878447186516148 is MAPE ERROR
forecasting values
```



- In the next column, I have first calculated the next 4 forecasting values as in the empty_test_sheet Year was only 6 and the Month was ranging from 3 to 6 only, so, I Calculated the next 4 values and If the month is 3 then the first value will be the answer or if the month is 5 then the third value will be the right answer which I saved them into test_sheet accordingly.

```
56 A01 Inventory
[-0.38671084 -0.37070456 -0.43805104 -0.44294185] $
```



- For all these variables SARIMAX model is used for forecasting with the same prediction steps with a little variation in parameters.

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