

## Short description for Problem 1

In this problem, I used a time series analysis using the Autoregressive Integrated Moving Average (ARIMA) modeling technique in Python which is one of the most general class of models for forecasting a time series.

The ARIMA technique seeks to model these trends in data over time and then extrapolate into the future to obtain forecasts. Time series modeling is a powerful method to derive insights from time-based data and make informed decisions about future trends.

In my solution, the notebook has completed with some main steps:

- 1/ Get and Transform Data
- 2/ Visualize time-series data
- 3/ Stationarize the time series, do some testing
- 4/ Find optimal parameter and build ARIMA model
- 5/ Perform and visualize forecasting
- 6/ Predict and write the result to the CSV file.

And In revenue of week, Sum sale value was instead by Mean() to keep the stability of the model.

## Short description for Problem 2

In this problem, we need to classify new users into high-value or not. After trying with some ML Model, I decided to use Logistic Regression for this case. Before start, we need to generate some features for each user. And create label data for the model.

We need to create data with the label. Because in this problem, we need to the classification of some users, which are created and logged in the first time in month X and classification base on their transaction in next 3 months (X+1, X+2, X+3). Follow the question, we take into account in total transaction amount of user in the next 3 months, we put them into high-value (value 1) and not (value 0).

Furthermore, we find some characteristics of the user from the transaction, action data. About transaction, I figure out the Period (Average time between 2 transactions), Average transaction amount, Number of transactions...And about actions, I find for each user, how long from create account to first logged in. Moreover, I count number of each campaign by month and use them as features of user for my model.

After creating feature for the user, I keep the user, which are have first logged in time before 2016-04-01 to satisfy the time range (3 months) – Detail in y code file for Exercise 2 for my label data. I try with SMOTE for imbalanced data, however, it doesn't improve my accuracy, therefore, I haven't used it in the final. (I will try to find the reason for it later)

I predict for users satisfied with the condition of problem and with others, I grant for 0.

