

# Case problem

**Copyright MBAG; Do Not Distribute and Share**

# Problem

- You have been provided with a time series in “train.csv” file which you have to forecast for next 12 time steps.
- Train data has 74 data points - from 01.01.2015 to 01.02.2021; Test data has to be submitted in the same format as that of “train.csv” but for dates 01.03.2021 till 01.02.2022.
- We recommend using open source package and it has to be coded in python 3.7 or 3.8 or 3.9 versions; preferred 3.7
- Use established open-source packages, like statsmodels and etc.; preferred sktime; Please don't use Prophet as we want to assess your model know-how.

# What you will be assessed upon

Topic	score (in %)	What we will be looking for?
Accuracy	30	Against test set that is not shared.
Model pipeline	50	How easy it is to run pipeline with only one command; what are the components of pipeline; what pre-processing and post-processing steps have you included in pipeline?
Explainability of models	20	Can you explain how the models and methods you have used work? This is focused mainly on Time Series knowledge.
Uniqueness	10	Whether you simply copy-pasted or not.

# Some advices for time constraints

- If time is limited, then focus on working upon higher scoring topics as shown in previous slide.
- We will only review code as per your github link. We will not be accepting code submission, but only the “test.csv”. But this output also can be generated and written in advance to the git repo. We will download the file from repo and assess the accuracy locally against test set.
- Coding style is important; if you are running short on time, then minimise your work on doc-strings, comments but not on class creations for example.
- You can stick to specific package based pipeline creation, like SKTIME or DATS or whatever you find best, as long as it is not Prophet package.