ML Hackathon- Problem statement

Objective: Given is the dataset of various loan application counts spread over different parts of the country. In order to accurately scale their services this company wants to predict the upcoming count of the number of applications over the next 3 months. It should be done -
w.r.t **GROUP, BRANCH_ID, STATE,** predict the incoming number of applications for this hierarchy.

The given dataset is a daily level record of application counts. The train data has data until 2019 April, candidates should predict for the next 3 months which is for ~90 days.

Super Task: Improve the MAPE of the next three months forecast.
Metric: MAPE = mean(abs(Actual - Forecast)/Actual)
Where MAPE > 1; then set MAPE = 1.

1. **Predict the null values** in the zone which belongs to Group 2 using any appropriate ML algorithm.

2. Create a **good set of features** which would help in model building.
The created features **should explain the incoming applications behavior.**

3. **Apply clustering** and add the features you get from it as features.

4. Make valid **Encodings** and **Standardization** as required.

5. Analyze the improvements you get over **clustering** and make a valid judgement.

6. Analyze the improvements you get over various features you've added through **feature engineering.**

7. Build 2 models one which is a **LinearRegression** and the other which is a **RandomForestRegressor** and make the predictions for the upcoming next 3 months.

8. Make valid judgements and write them down about **moving variance** while applying your ML models.

9. Create a valid **validation dataset** which would give confidence about your next 3 months forecasts.

9. Take the weighted average of both LinearRegression and RandomForestRegressor predictions and make a judgement about their improvements. Explain how you decided the weights.

10. Use some features which explain the seasonality. Features such as national holidays, festivals etc... You should find a way to get this info, for example through any library or any datasets which have info about holidays.

Submission Files: Predictions for test data, Explanation Document.