**EDA:**

We have done EDA for the data points in plotply so that we get an overview of the data trend.

Based on the EDA, we inferred that the month end and the day before the national holidays had a peak value of applications received, whereas the month start and the day of national days had a dip in the value of applications received. Below are the variables that were analysed and inferred from the EDA.

1. Holidays

2. Monsoon

3. Weekend

4. Month Start

5. Month End

**Feature Engineering:**

Since the insights from the EDA suggested, certain trends in the number of applications received in the holidays we have the external dataset of national holidays from the below links.

<https://www.calendarlabs.com/holidays/india/2017>

<https://www.calendarlabs.com/holidays/india/2018>

<https://www.calendarlabs.com/holidays/india/2019>

**Model Selection:**

We tried transformation of data points along with AUTO ARIMA as the initial approach, but as we had to capture the sudden increase in peaks especially during holidays, we had to remove transformation and hence change the model.

Hence we have converted a univariate time series analysis to multivariate time series analysis by considering the above mentioned variables.

Since we are using a multivariate data, the model selected was XGBoost. XGBoost was selected over other regression algorithms like Linear Regression because linear regression will capture the linearity between the data points whereas XGBoost captures the non-linearity of the data points due to its boosting nature.

**Inference:**

We have not done any hyper parameter tuning on the model, because we found out that the variables are more important for this particular model rather than the hyper parameter tuning.

**Final Submission:**

Though we had a better score of 16 in the public leader board, we have submitted 18.22 as the final submission so that the model generalizes more to the private data rather than overfitting the public data.