<u>Forecasting using Meta Prophet</u>

Prophet is an open-source library developed by Facebook and designed for automatic forecasting of univariate time series data.

- The Facebook Prophet model is a type of GAM (Generalized Additive Model) that specializes in solving business/econometric — time-series problems.
- Prophet helps to set up an additive model that takes trends, seasonality, holiday effects, etc. into consideration to achieve a good forecast
- It is designed to be easy and completely automatic, e.g. point it at a time series and get a forecast. As such, it is intended for internal company use, such as forecasting sales, capacity, etc.
- We use a decomposable time series model (Harvey & Peters 1990) with three main model components: trend, seasonality, and holidays.
- They are combined in the following equation:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t.$$

1. Here g(t) is the trend function which models non- periodic changes in the value of the time series,
2. s(t) represents periodic changes (e.g., weekly and yearly seasonality)
3. h(t) represents the effects of holidays which occur on potentially irregular schedules over one or more days.
4. The error term epsilon(t) represents any idiosyncratic changes which are not accommodated by the model; later we will make the parametric assumption that epsilon(t) is normally distributed.

When to use Prophet model?

The software is good for the business forecasts encountered at Facebook, which refers to hourly, daily or weekly observations with strong multiple seasonalities. Prophet is also designed to deal with holidays that are known in advance while missing observations and large outliers.

The challenge with Facebook Prophet is that it does not look for casual relationships between the past and the future. It simply finds the best curve to fit the data using a linear logistic curve component for the external regressor. Another major criticism against Prophet is that its underlying assumptions are simplistic and weak. Also, since Prophet does not directly consider the recent data points as compared to other models, this affects the performance in cases where prior assumptions do not fit.

Hence, Prophet is generally recommended only for time series where the only informative signals are trends, and the residuals are just noise.

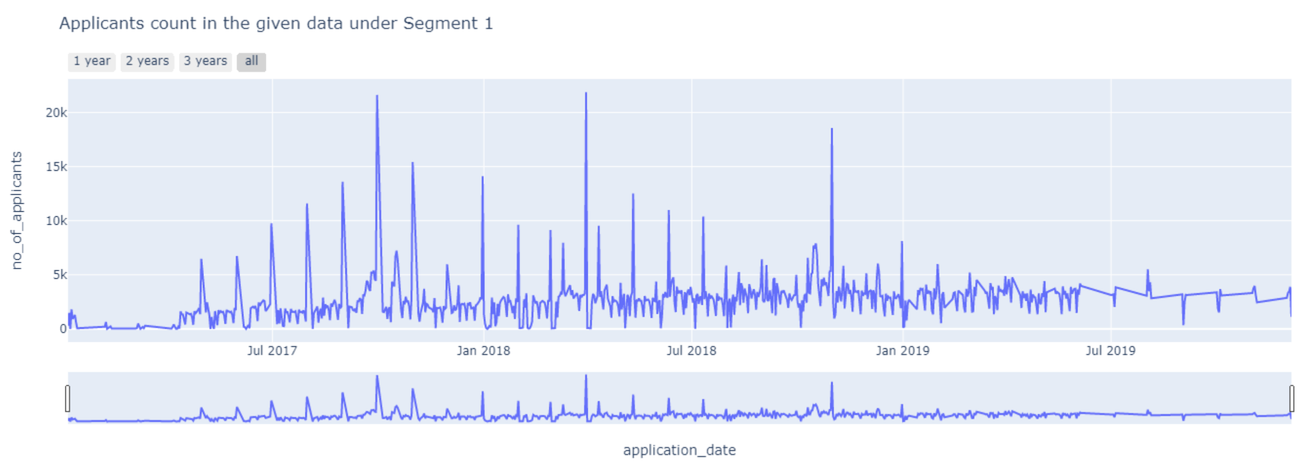
For the given problem statement the Prophet model was implemented on the individual segments using Python.

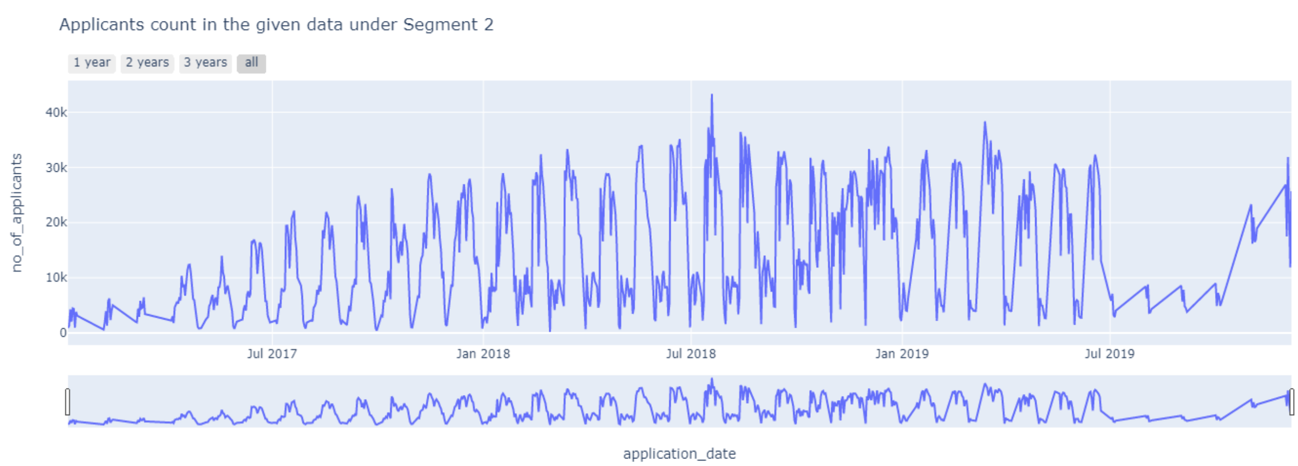
Approach

> Data

After the initial data type formatting the missing values in the data were not treated since by dropping them we induce bias. The data is grouped into two based on whether the record belongs to segment 1 or 2.

From the distributions of each segment :





We observe monthly seasonality between the periods April-2017 to July-2019 in Segment 2 wherein the number of applicants tend to increase mid-month whereas in Segment 1 the data tends to be stationary over the period between Mid-April 2017 to Mid-May 2019.

To account for the effect of holidays we explore two approaches- using the default list provided by Prophet and Adding customised list of holidays.

Furthur,each segment's records are trained separately using Prophet.

> Model

For each segment data is trained using all the available data in order to account for the inconsistency in the data during 2017 and after 2018.

For Segment 1 predictions we use 95% confidence interval.The default confidence interval of Prophet is 85%. The yhat(predicted values) and y(actual value) are as follows:

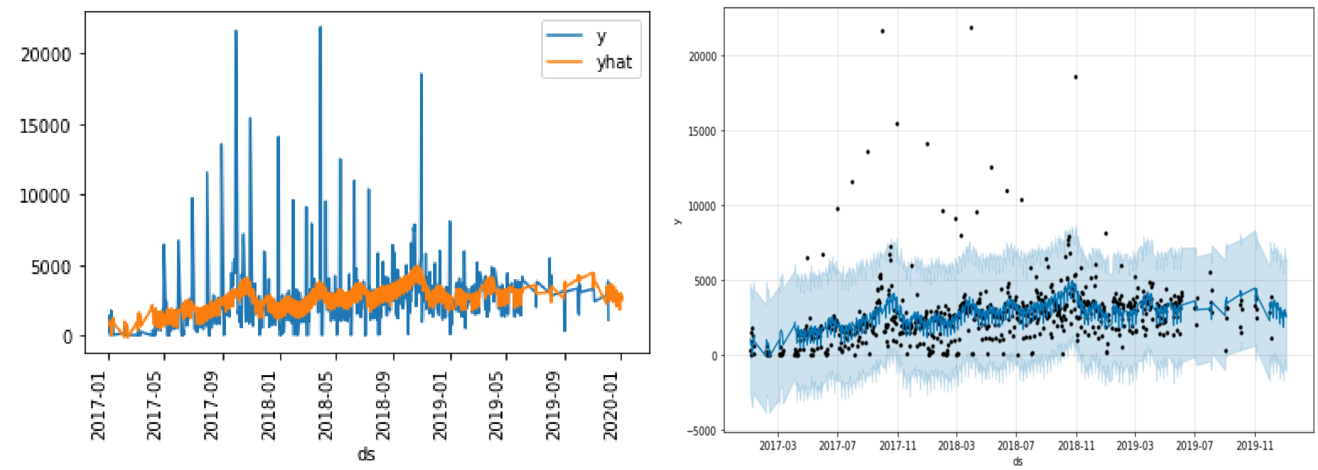


Fig1 (a) *Actual vs Predicted values for Segment 1* *(b) Forecasted values for the next 30 days Segment 1*

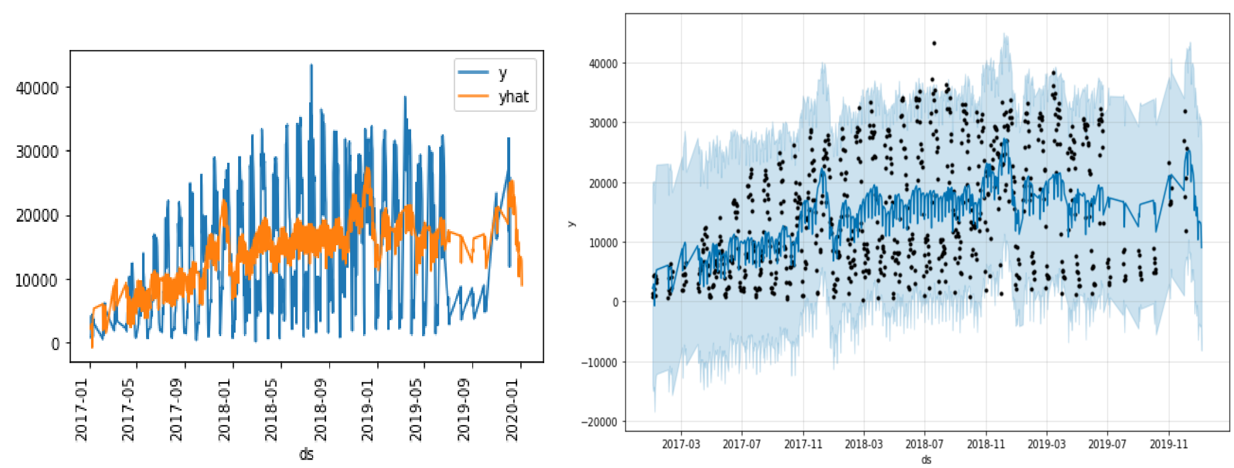


Fig2 (a) *Actual vs Predicted values for Segment 2* *(b) Forecasted values for the next 30 days Segment 2*

In fig1(b) and 2(b) the dark blue line shows the predicted values, the light blue region is the confidence interval and the black points show the actual values.

In case of both the segments we observe that the model has captured the seasonality and the trend but has failed to capture the outlier points. In fig2(a) we observe that the predicted values after 2019-05 are inconsistent with the actual trend. Thus we observe that the model gives more weightage to the historical data than the immediately preceding time-period.

On calculating the metrics for each of the models we obtain:

METRIC	SEGMENT 1	SEGMENT 2
SMAPE **(Symmetric Mean Absolute Percentage Error)**	0.2755	0.7110
MAPE **(Mean Absolute Percentage Error)**	0.2389	0.4921

Notebook link :

https://colab.research.google.com/drive/1x8UpZkJWzXyN0Nz8WKrWxjHdhYIyZi3E?usp=sharing

References:

1. https://towardsdatascience.com/time-series-analysis-with-facebook-prophet-how-it-works-and-how-to-use-it-f15ecf2c0e3a
2. https://www.youtube.com/watch?v=VtItg-J6-CI&t=97s
3. https://peerj.com/preprints/3190/
4. https://www.stat.tamu.edu/~suhasini/teaching673/linear.pdf
5. https://www.youtube.com/playlist?list=PLUgZaFoyJafhfcggaNzmZt_OdJq32-iFW
6. https://www.youtube.com/playlist?list=PLUgZaFoyJafhB73-1JUTRT0y5u_5fjFCR
7. https://www.youtube.com/playlist?list=PL_onPhFCkVQi4tqlo5beRDoQINTZazoP