Edits and Cancellations prediction model (in progress)

* M. Kaleeswaran, Analytics Intern, Swiggy NCR under Mr. Vinit Kumar.

Objective:

To make a fair prediction of at what time in future would a restaurant initiated item edit or cancellation would occur.

The idea:

* Create an hour-wise demand forecast on restaurants generating higher number of orders (it is not possible to predict the same for restaurants generating low orders with high significance).
* Forecast at which hour of the day and at what day of the week, the most frequently edited item is getting ordered.
* Forecast once after what frequency of orders in an hour of the day the item gets edited.
* Generalize that it would go out of stock, once when the item gets ordered after the number of orders during the hour has crossed the forecasted threshold value.
* This way with the previous forecasted hour-wise demands we can take action pro-actively.

The workflow:

Conducting an hour-wise demand forecast on a zone:

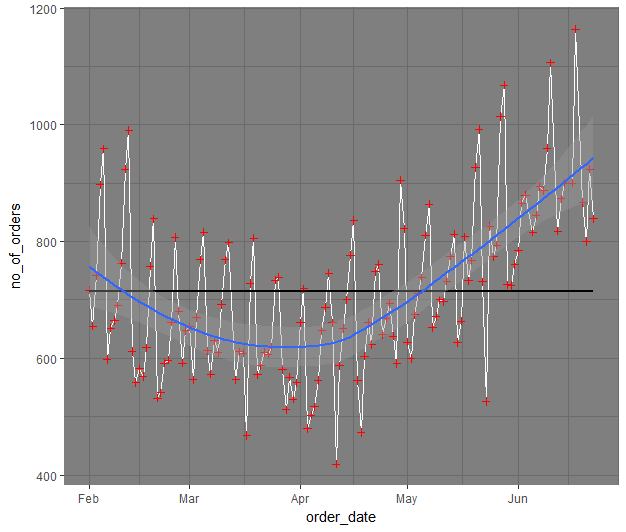
The purpose of doing this would be is that, if this could be done successfully on a zone-wise, it could be done on a restaurant with high order frequency.

The whole model has been implemented in R programming.

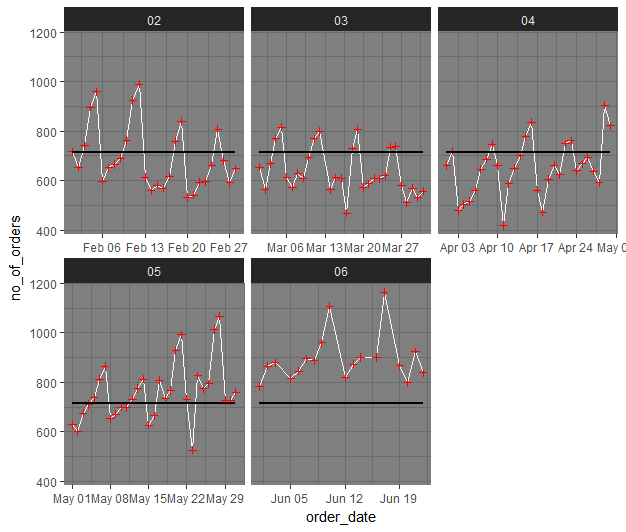
I have used ARIMA modelling to predict the demand. ARIMA – Auto-Regressive Integrated Moving Averages is a forecasting technique that would account for the Trend, seasonality and residuals in our Time-series data.

The data worked upon is the orders-items data for Janakpuri area restaurants during the month of February to June.

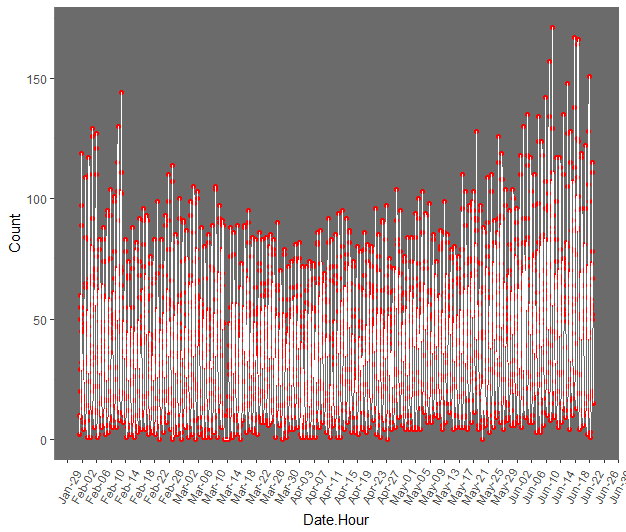
As a first step, created a time-series data of orders between each 1 hour time period in Janakpuri area to identify the general trend and seasonality in the data.



The overall trend in the orders in Janakpuri area between the months of February to June. This helps us identify the seasonality in our data. There is a seasonality at 7 period interval.

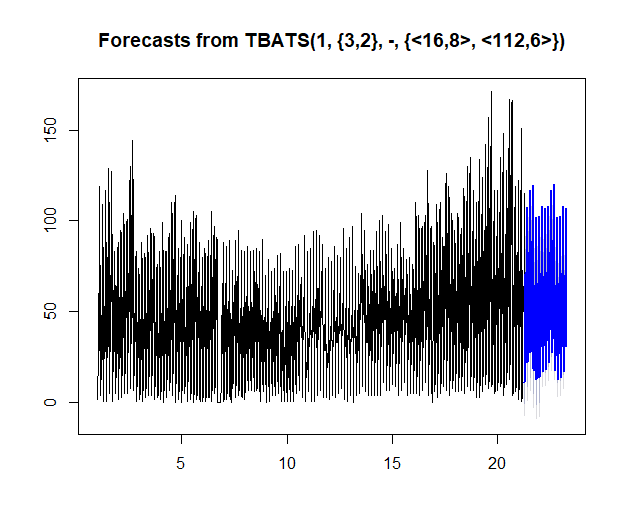


Similar break-up on a month-wise basis for better understanding the trend. The black-line represent the average orders over the timeframe.



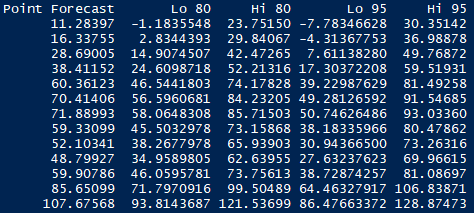
Hour-wise break of orders in Janakpuri area

Closer look at the hour-wise data reveals two kinds of seasonality, one within the days of the week and other within the hours of the day. Since our data is considered for only hours 8 – 23, our seasonality would be 16 and 16 \* 7.



Black part shows the order frequency from our data and Blue part is the prediction over next few time periods. It may visually seem that the prediction is undermining the order frequency, but it is not so, since most of the data points during June were outliers. Thanks to the Champions Trophy.

In order to deal with these outliers, we need to identify outliers and replace it with a prediction made using past data before the outliers and conduct further research, which would make the model more efficient.

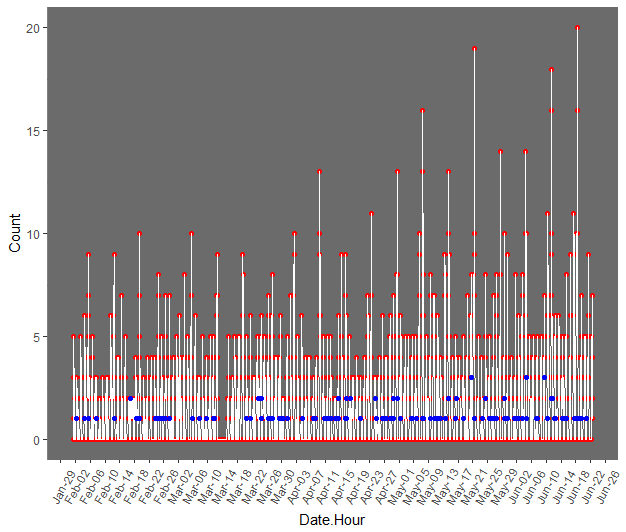


Predictions over first few hours.

Now, given that we have predicted on the whole Janakpuri area, we would have to extend this at restaurant level. It could be done by two ways:

* The hard way: Create ARIMA model on the restaurant to make prediction on hourly order frequency, just the way we did for overall zone.
* The easy way: Identify the fraction of total orders on an average the restaurant generates to the total orders in the zone at each hour. Make a prediction on the restaurants order as this fraction of orders predicted for the zone.

Method 1 would make a better estimates on the order frequency for obvious reasons.



An analysis on “Rama Chole Bhature”. The red points denote the count of orders during each hour between February to June and the blue point would represent any of the edits that has occurred.

This is what I have done up till now on this project, the way forward would be the exact same been described in the “The idea” section of the doc.

Rephrasing on the way to proceed:

* Make a prediction on hour-wise demand for each high hourly order frequency restaurants.
* Forecast at which hour of the day and at what day of the week, the most frequently edited item is getting ordered.
* Forecast once after what frequency of orders in an hour of the day the item gets edited.
* Generalize that it would go out of stock, once when the item gets ordered after the number of orders during the hour has crossed the forecasted threshold value.
* This way with the previous forecasted hour-wise demands we can take action pro-actively.

Issues:

There are so many discrepancies with the data. The figures don’t match to RHI data at certain fields. If the data is much more refined, we could have a much better significant estimates.