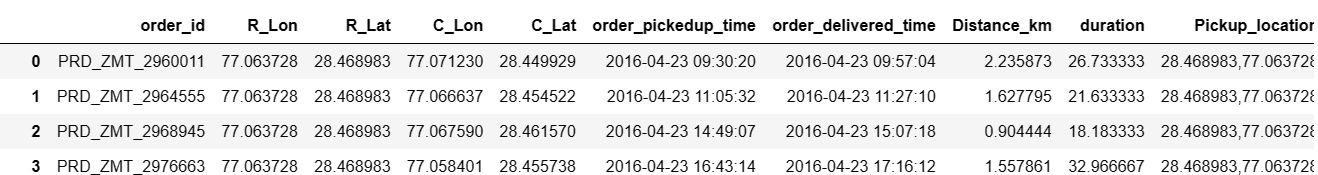
**Data Analysis of Delivery Data :**

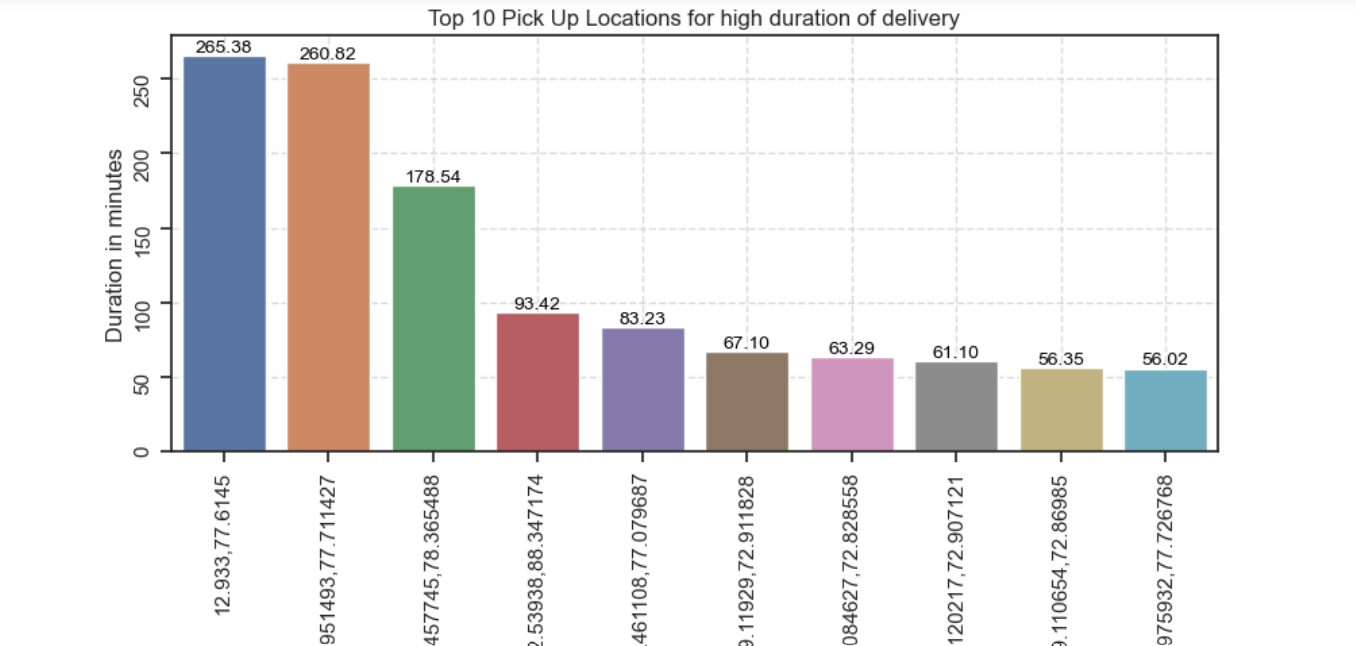
This datasets mostly comprises of Time Series Data tells about different product Id and their Delivery Information.

With the source and Destination’s latitude and longitude location it was possible for me to calculate the geo-distance which is the actual distance for each record. For time wise calculation , I found the duration of the delivery.

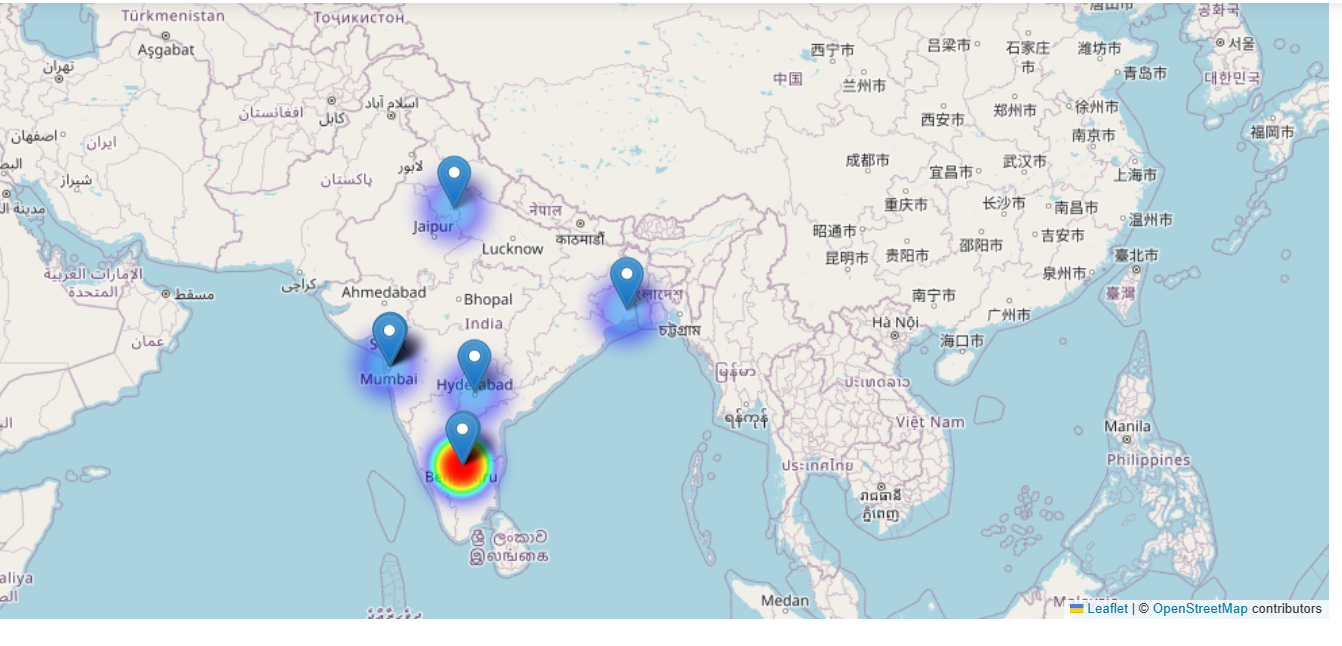
Now the Data Looks this below.



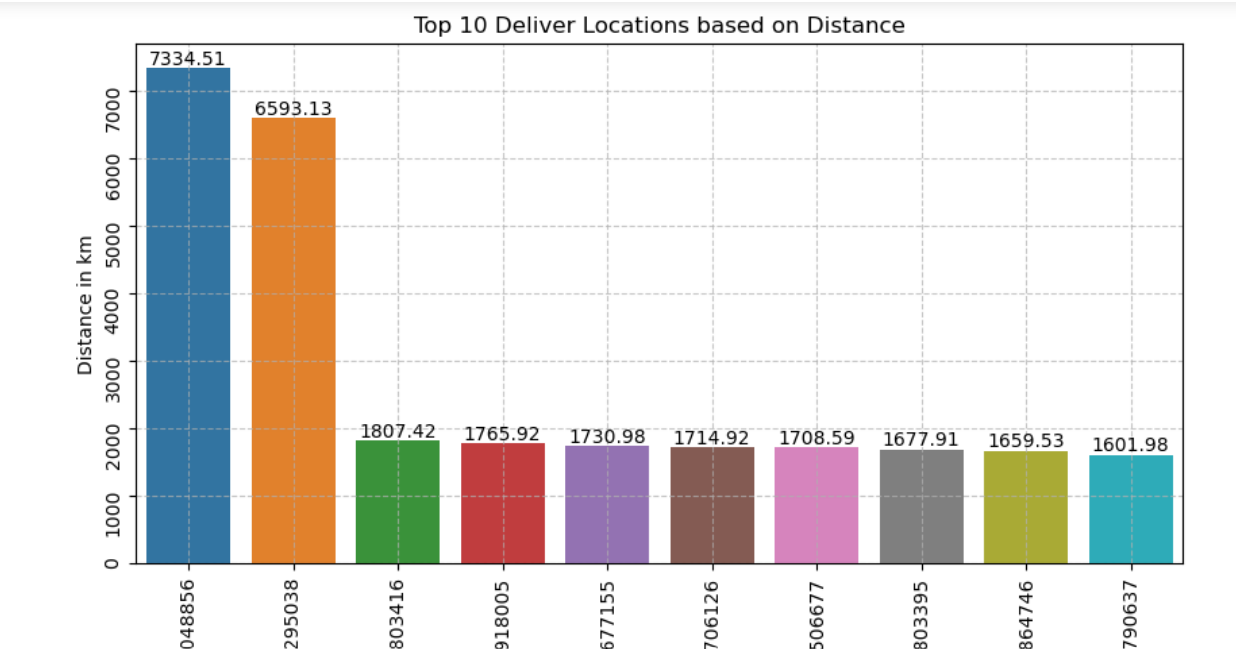
Now we will do and EDA to find out which pickup location has the highest mean of time. That means which customer location has to face the highest waiting time. I am choosing top 10 values here in descending order.



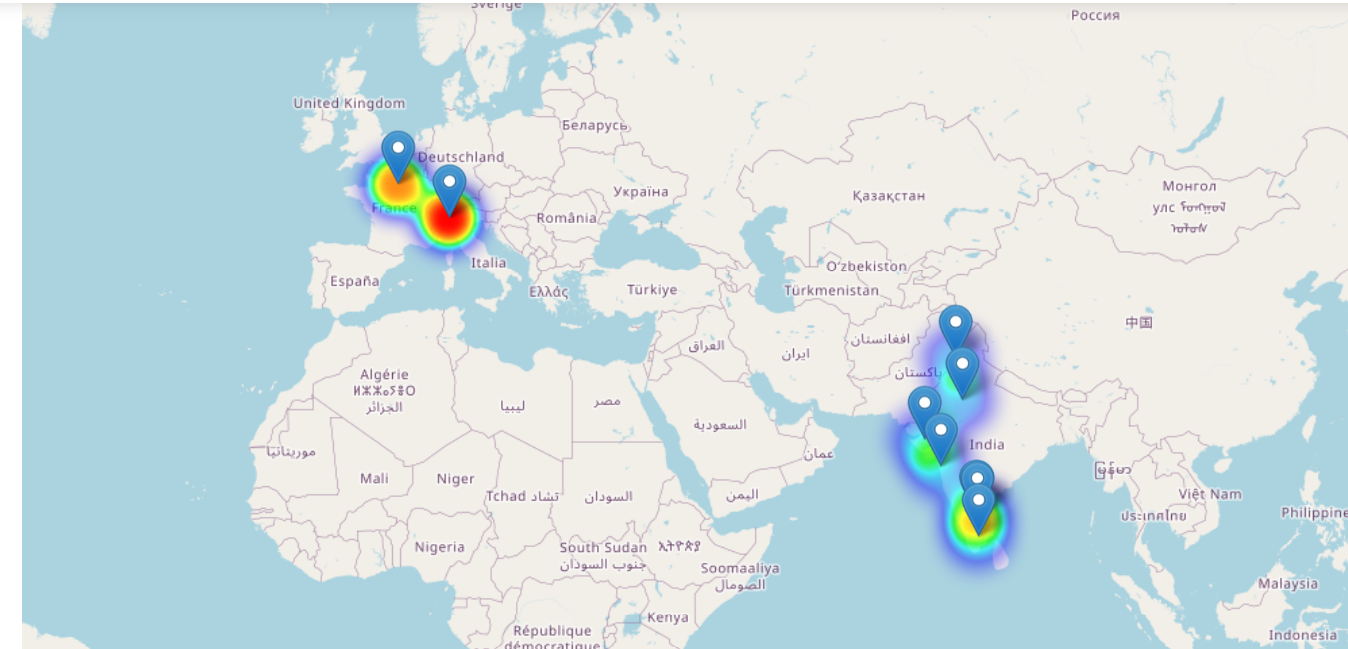
Now I have plot the geographic heatmap based on the duration intensity of the delivery with package folium.



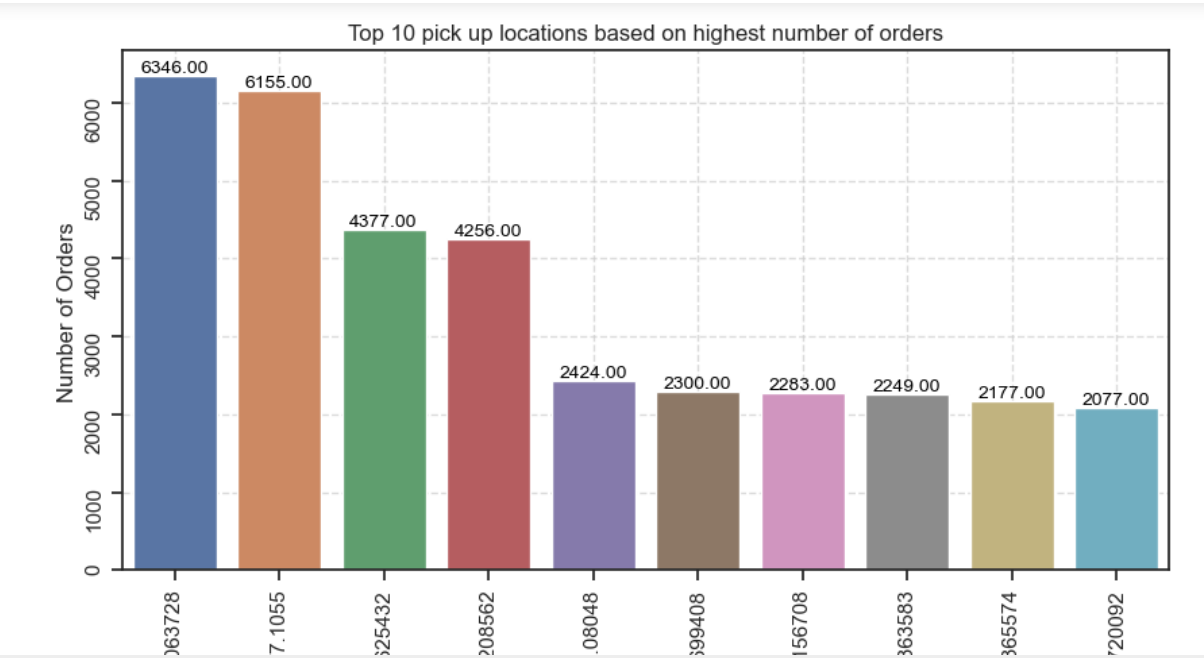
We will try to see which delivery location has the highest mean of distance of delievery. That means customer has to ship furthest. We will also choose here top 10 locations. **Highest waiting customers are from Bengaluru**



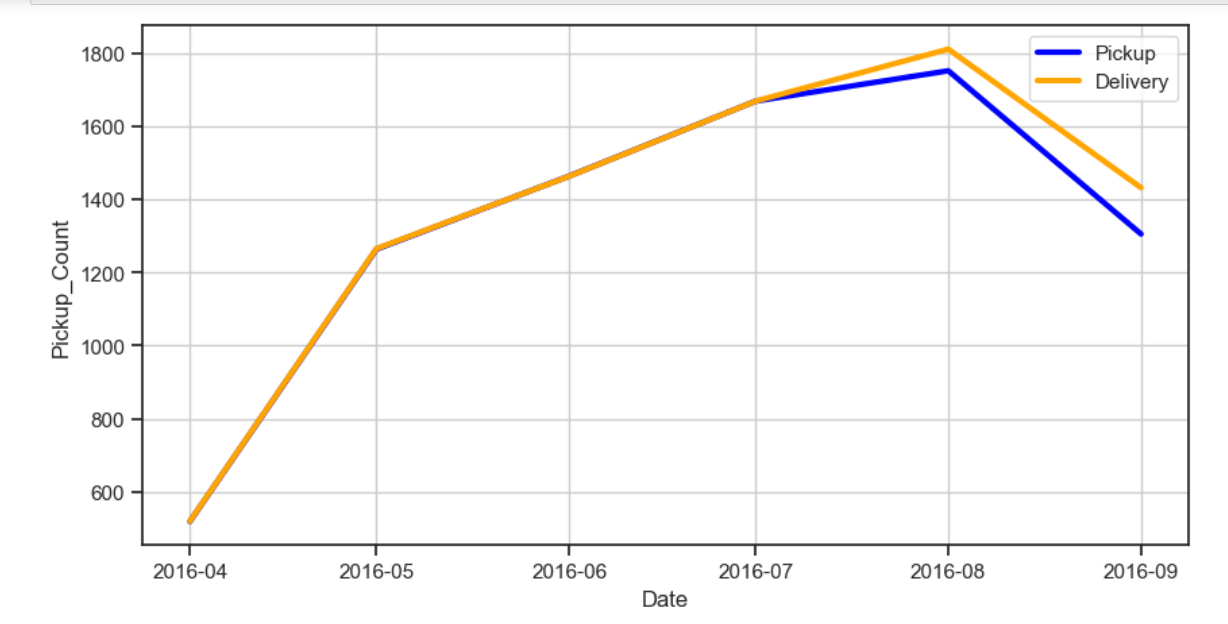
**Furthest location is Germany and Italy for delivery**



Trying to find the top pickup location with highest number of orders.



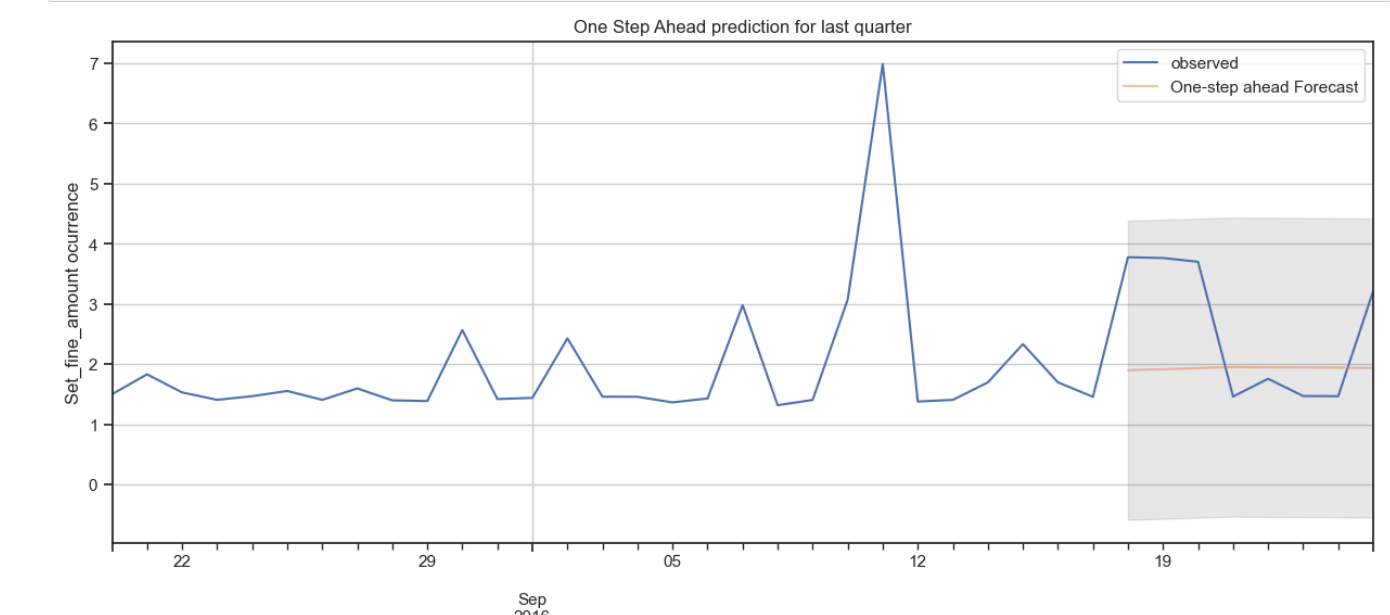
Plotting to see the trend of pick up and delivery over the time. **Seems like in last two months less number of orders has been picked up and more number of orders has been delivered. that means some packages which were ordered previously they delivered later before they were either on hold or they travelled long distance. Took more time.**



Time Series Seasonality inspection:

**Now we will see the data orders by date. Seeing the time series we will try to see the pattern in the time series data.**

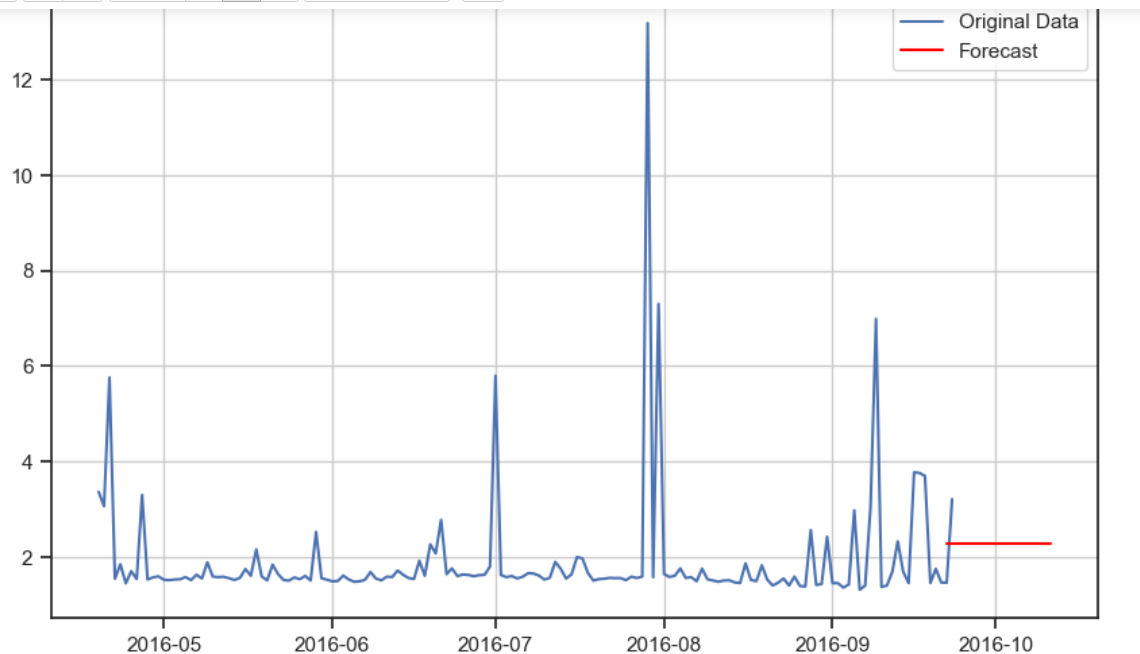
**Now for Time series forecasting with SARIMAX modelI have seen some abnormality here. I have seen while applying the interval for leadt AIC score Even though the data is normalized but the p value, skewness and heteroskeasdictiy gets worse. It happens to be due various reasons.**

****

**1. The Data is not very large.**

**2. There is a abrupt change in the data. There is one single row in the dataset which has no pickup information but has high delievry information. Also, Some geographical location has very high delivery/ pick up information. OR let's say in some day theres very high number of delivery.**

**So applied exponential smoothening**

****