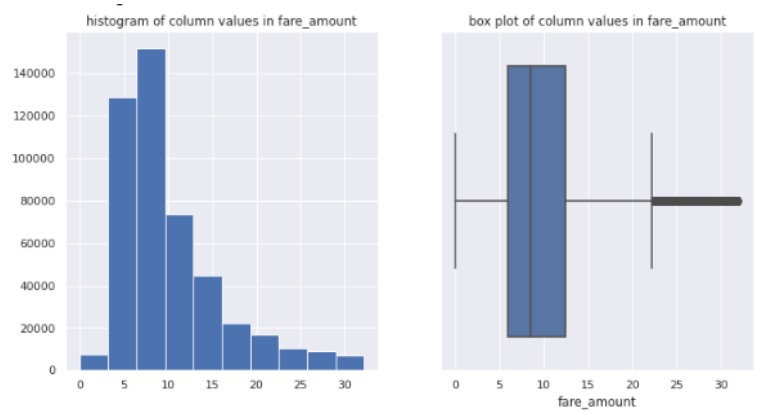
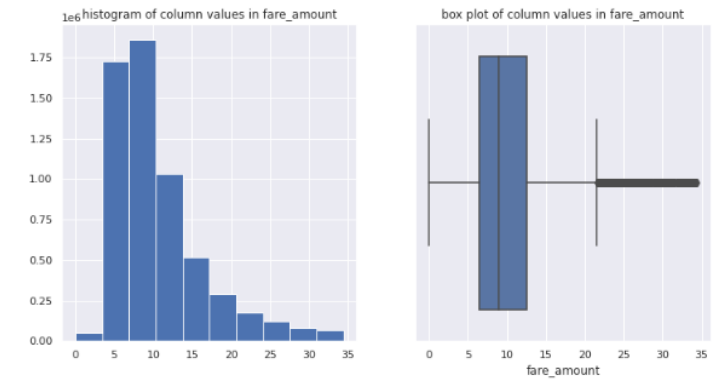
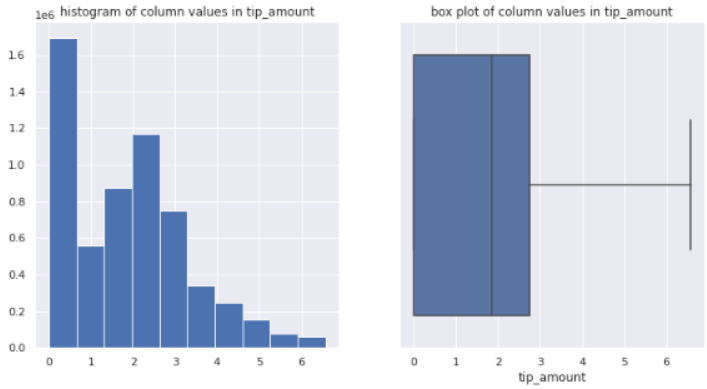
Feb



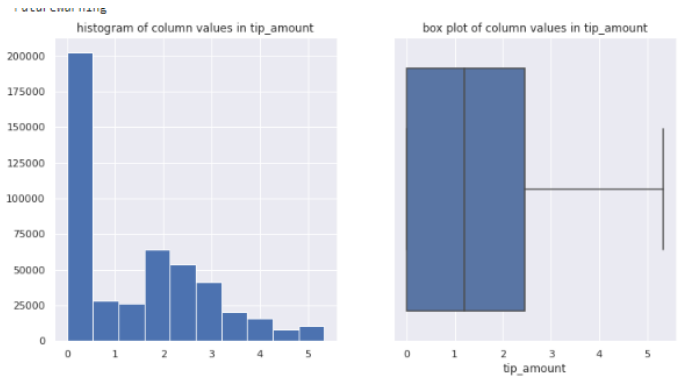
June 2020



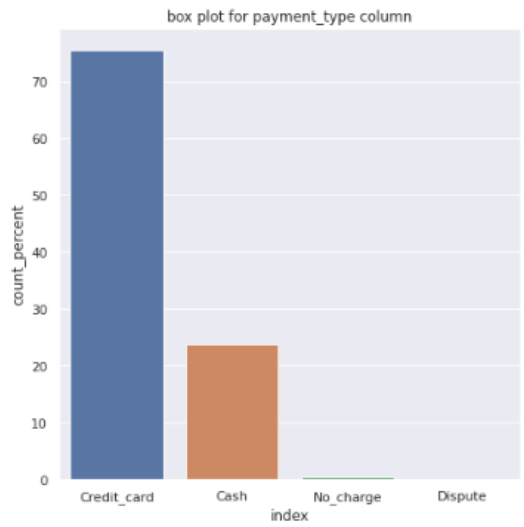
Feb



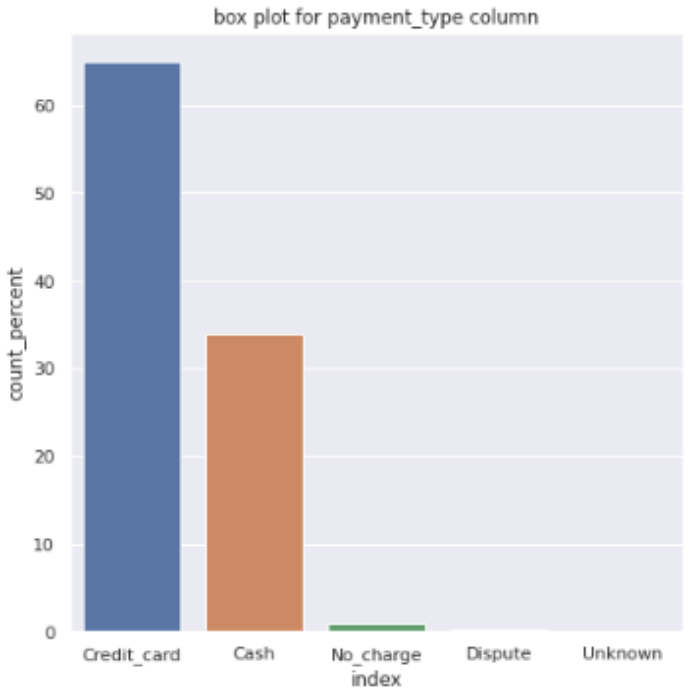
June 2020



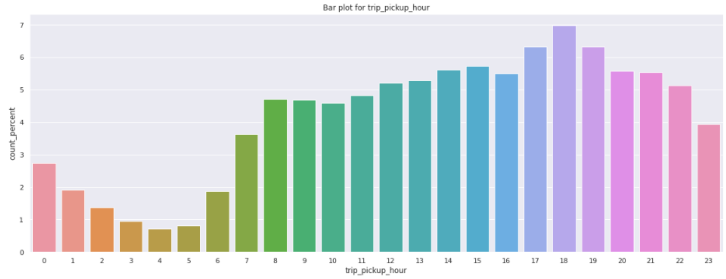
Feb

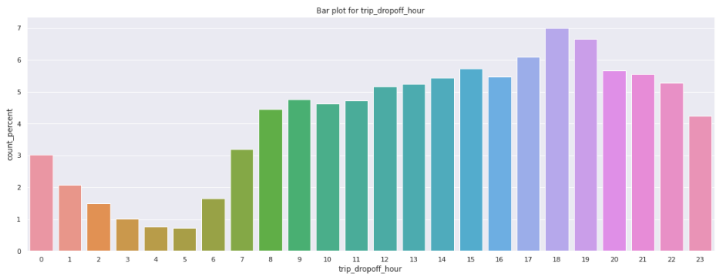


June 2020

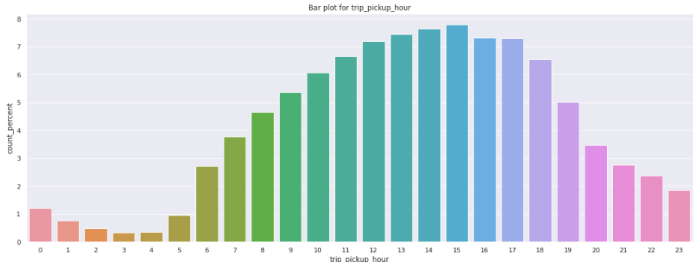


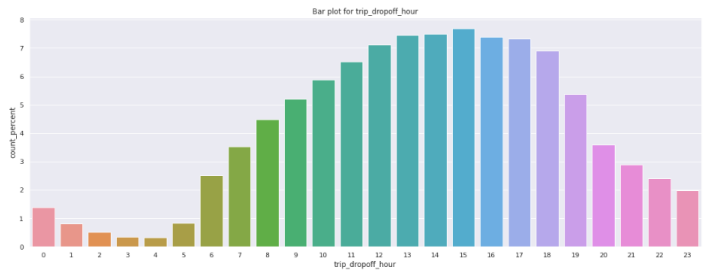
Feb 2020 pick-up – drop



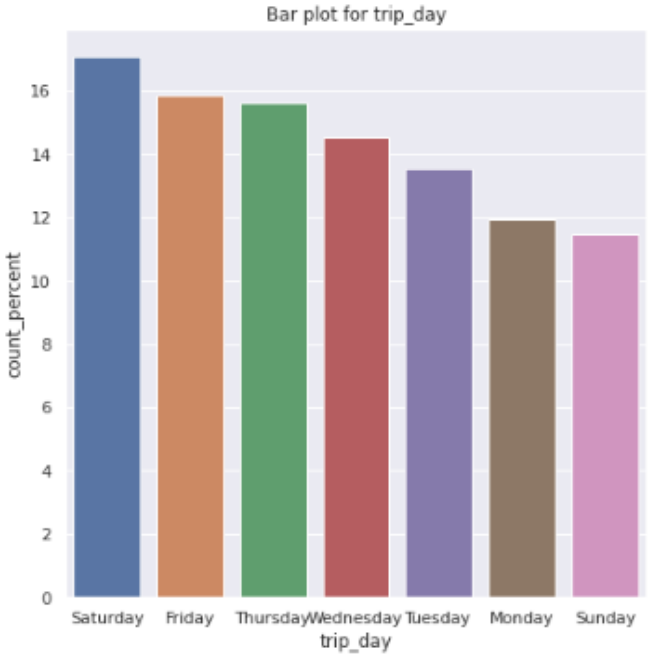


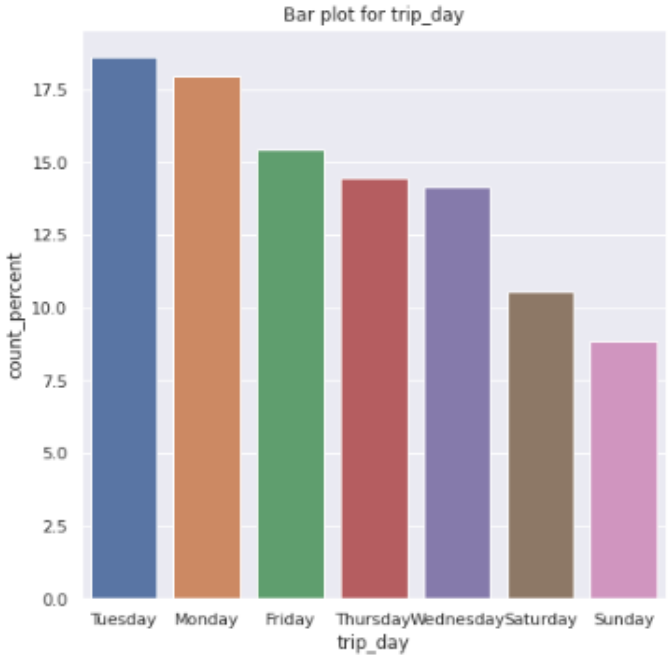
June 2020 pick-up – drop





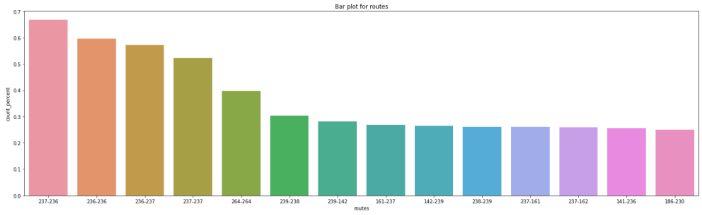
Weekly



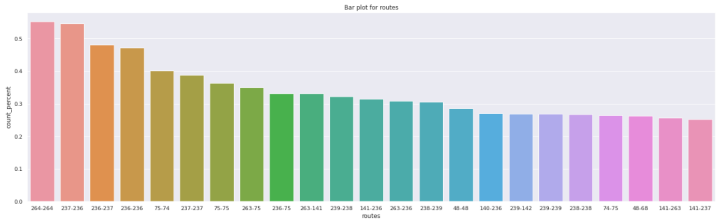


Most booked routes

Feb



June



Feb 2020 Observations:

Following insights would be useful for our company's product launch in New York

- fare_amount - most of the fare amount is within 9 dollar value as is shown by the median value. Though there are some significant outliers, the maximum of which is 6052 dollars.
- tip_amount - most of the tip amount is within 2 dollar as is shown by the median value. Though again here too we have outliers, the maximum of which is around 549 dollars.
- tolls_amount - most of the tolls_amount value is 0 so it seems most of the trips do not have to pay for tolls.
- total_taxes - most of the total_taxes values is within 1.3 dollars as is shown by the median value. Though we have outliers in this case but it is not as signiificant as the case for tip and fare.
- total_amount - most of the total_amount values is within 11 dollars as is shown by the median value. Again the outliers in this case seems mostly because of outliers in fare_amount.
- duration - most of the values in duration is within 10.76 minutes range as is shown by the median value. We do have some outliers which are beyond the range of 3147.6 minutes.
- trip_distance - most of the trip_distance is within 1.6 miles value as is shown by the median. The outlier in this case is 369.9.
- Trip Hour
    - the dropoff and pick up hour distribution looks almost same, it is because the trip duration in most of the cases is less than an hour with the median duration value as 10 min.
    - Peak hour for the pick up and drop off is around evening from 6 to 18. The busiest time is 6PM.
    - there is less traffic during night times and only after 8AM in morning does the pickup and drop off starts picking up pace.
- Trip day
    - Sunday has the lowest taxi uses.
    - Weekdays except Monday have heavy taxi uses.
    - Among weekends Staturday has taxi highest taxi uses as compare to all days.
- From the above plot we can observe that 5 busiest route are following:
  237-236, 236-236, 236-237, 237-237, 264-264
- From the hour 8PM to 5AM the median taxes seem to be a bit higher than other hours, it may be due to some overnight surcharges.
- Evening from 4PM to 7PM have quite variable taxes and is a bit higher than other times, it may be due to higher traffic charges.
- We discovered from the dataset that even for the busiest pickup location the median fare_amount is a bit lower than other busier pickup locations. So just choosing busy pickup locations for higher revenue won't work, we may have to choose locations taking into consideration both busy traffic and higher median fare_amount.
- Early morning hours of 5AM to 6AM have shorter duration trips

June 2020 Observations:

This is the result of some of the most important insights after doing univariate analysis:

- fare_amount - most of the fare amount is within 8-9 dollar value as is shown by the median value 8.5 dollar. Though there are some significant outliers, the maximum of which is 941 dollars.
- tip_amount - most of the tip amount is within 1-2 dollar as is shown by the median value. Though again here too we have outliers, the maximum of which is around 422 dollars.
- tolls_amount - most of the tolls_amount value is 0 so it seems most of the trips do not have to pay for tolls.
- total_taxes - most of the total_taxes values is within 1.3 dollars as is shown by the median value. Though we have outliers in this case but it is not as signiificant as the case for tip and fare. max value is 88
- total_amount - most of the total_amount values is within 13-14 dollars as is shown by the median value. Again the outliers in this case seems mostly because of outliers in fare_amount.
- duration - most of the values in duration is within 8.5 minutes range as is shown by the median value. We do have some outliers which are beyond the range of 4497 minutes.
- trip_distance - most of the trip_distance is within 1.71 miles value as is shown by the median. The outlier in this case is very large value 22543
- Credit card is the most preferred mode of payment followed by cash.
- Trip Hour
    - the dropoff and pick up hour distribution looks almost same, it is because the trip duration in most of the cases is less than an hour with the median duration value as 10 min.
    - Peak hour for the pick up and drop off is around evening from 6AM to 12PM . The busiest time is 3:00 PM.
    - there is less traffic during night times and only after 6AM in morning does the pickup and drop off starts picking up pace.
- Trip day
    - Sunday has the lowest taxi uses.
    - Tuesday & Moday has nearly same heavy taxi uses.
    - Friday, Thurday & Wednesday has nearly similar taxi uses
    - Saturday has bit more taxi use than Sunday
- The busiest location in terms of pickup and dropoff are 75, 236 and 237.
- Four of the busiest routes are - 264-264 237-236 236-237 236-236 75-74
- Mostly 1 or 2 passenger avail the cab. Group rides are less common.
- From the hour 8PM to 5AM the median taxes seem to be a bit higher than other hours, it may be due to some overnight surcharges.
- Evening from 4PM to 7PM have quite variable taxes and is a bit higher than other times, it may be due to higher traffic charges.
- We discovered from the dataset that even for the busiest pickup location the median fare_amount is a bit lower than other busier pickup locations. So just choosing busy pickup locations for higher revenue won't work, we may have to choose locations

taking into consideration both busy traffic and higher median fare_amount.

- Early morning hours of 5AM to 6AM have shorter duration trips