SHAHI SAWARI

Report Presented By:
Sannia Nasir

Introduction of the Company

Shahi Sawari is a private limited that provides pick and drop service via auto-rickshaws for daily affordable commute.

The major features of this service include:

- **Safety and Reliability**
  The company verifies and recruits experienced drivers to ensure the reliability of the service and safety of the consumers.
- **Ride Tracking System**
  Furthermore, the Shahi Sawari app allows a customer to share his/her ride with a third party so that the ride can be tracked and hence, the security is assured.
- **Fixed Pricing**
  The service charges are fixed with respect to the base rate, time and distance. No surge applies.
- **Ride Pooling**
  The service allows customers to pool in a ride if they wish to split the charges.
- **Customer Support**
  The Customer Support is functional throughout the week from 7am to 11pm to assist the consumers in case of any problem or query.
- **Monthly Packages**
  In addition to day-to-day commute, Shahi Sawari also offers monthly packages to customers.

Competition

Being heavily dependent on the personal vehicles and mass transit systems, Pakistan does not have many players in the taxicab industry. Shahi Sawari shares the market with two multinational companies: **Uber and Careem** and also with local **independent auto rickshaws** in Pakistan.

- **Fares**

The fare breakdown of Shahi Sawari, Uber and Careem is as follows:

Company	Starting Fee (Rs)	Moving- per km (Rs)	Waiting-per min (Rs)	Minimum ride Fare (Rs)	Cancellation Fee (Rs)
Shahi Sawari	35	11	0	65	50
Careem: Rikshaw	15	12.56	3	70	60
Careem: Go Mini	65	10.99	4.9	95	100
Careem: Go	80	12	5.08	120	150
Uber: Auto	15.4	8.55	1.97	62	52
Uber: Go	80	7.12	3.18	120	120
Uber: Mini	64	5.69	2.55	96	96

- **Surge or Peak Factor Pricing**

Both companies, Uber and Careem are often criticized for surcharging prices when the supply is limited at a fare higher than the normal rate known as the “peak factor effect or surge”.

While Careem has a relatively more extensive network than Uber, having more captains and hence more cars makes the frequency of surge fares usually low however Uber faces the problem of less to no taxis available at times to a larger extent so higher fares are also common.

Moreover, Careem also offers monthly packages to the consumers which is relatively economical and surge pricing does not apply in that case.

On the other hand, although the fares are competitive among the regular auto rickshaws in the cities, yet there is no fixed rate and the price might fluctuate depending upon the negotiations between the driver and the customer.

- **Geographical Coverage**

While local auto rickshaws are available countrywide, Uber and Careem operate in only particular cities of the country.
Uber operates in Islamabad, Karachi, Lahore, Gujranwala, Faisalabad and Hyderabad while in addition to these services, Careem provides also services in Sialkot, Multan and Abbottabad.

- **Customer Service**

Both Careem and Uber have physical centers in the major cities to deal with queries and complaints of customers. Moreover, both companies offer a “help” option in their apps via which queries can be routed.

- **Safety**

To ensure the safety of customers, Uber offers a separate “Safety Toolkit” in the app via which consumers can contact the local police via emergency button, share their ride with the trusted contacts, call the driver anonymously and access information regarding the community guidelines, insurance and general safety tips.
On the other hand, while Careem too provides a “Safety” section to the consumers, it is not very extensive.

- **Marketing**

Both companies Uber and Careem are competing to partner with major events such as food festivals, concerts, literary festivals, plays etc. happening around the city.
Moreover, both services often also offer promos like “50% off (up to Rs.100)”, “flat150” etc.
On a comparative scale, Careem has been more active with respect to social media promotion and other campaigns such as Bakra-on-wheels, Rishta Aunty Service and Careem Air etc.

Literature Review

With the increasing population as well as the low-density urban sprawl in larger cities, the need for mobility in these urban centers is soaring alongside. The boosted demand for means of mobility necessitates a supply that not just meets the demand quantitatively but also manages the transport of the masses in the most optimal way. The supply side comprises of the Public Transport Systems and private vehicles. Another way of looking at the distinction between modes of travelling is via the number of people a single mode can accommodate at a time. While majority of the public transport systems such as bus, train or metro known as Mass Transport Systems (MTS) can cater to a greater number of travelers altogether yet the strict schedules, long routes, extended travel time and factors such as deteriorating infrastructure and safety concerns for women and children in developing block such as South or Southeast Asian countries have started to tilt the line of popularity a little away from the MTS.

On the contrary, taxis, cabs, rickshaws and carts have quite recently emerged as popular mode of conveyance due to the privacy and convenience in terms of travel time and experience. However, the downside of such private modes highlights the absence on an integrated management. In most cases, every cab or rickshaw driver acts as an independent organization setting his/her fares, routes, stations almost on his/her own in the case where the country does not regulate such services (Salanova et al.). Hence in recent years, capitalizing on the “no-central-policy” issue of taxis, cabs and rickshaws, companies have emerged to offer a solution; transportation networks to provide an indiscriminate, uniform and integrated private or shared pick and drop cab/rickshaw service for everyone.

These companies that provide uniform taxi services differ from one another in multiple ways such as revenue generation model, customer and driver relationship management, cost drivers, and most importantly taxi dispatch algorithms and objectives. Marc Kuo identifies different dispatching algorithms that might determine the type of business a company is running. The first one is to dispatch all the idle cars to receive the nearest ride. Second way is to receive multiple passengers from one location and then drop them one by one. The third algorithm enables the drives to pick up other passengers that intend to get dropped on the way of the destination of the ride already in progress just like carpooling, the technique which already is being used by services like Lyft Line and UberPool. Fourthly there is a way to delay departure until the maximum number of passengers are picked up ensuring that all of them are dropped off on time. Another smart way of boosting up the number of rides is to only accept the ones those are in close proximity, so the revenue is based upon the "quick short trips". Finally, the most common and efficient way of dispatching taxis and rickshaws is through route optimization where given the constraint of time, vehicles and other resources, shortest routes are calculated, and cars are dispatched accordingly (2016).

While the choice of a specific algorithm is affected by various objectives of each company that might be to minimize customer wait time or travel time, maximize the number of rides per driver etc. (Voytek 2014), all the operations and strategic management of all the companies is unanimously influenced by some common parameters. The basic factors that determine the quality and eventually the profitability of the integrated transport companies are dynamic passenger demand and dynamic taxi supply (Fei Miao et al.). The former spatiotemporal factor varies by the regular characteristics such as rushy areas or hours or irregular fluctuations including weather, events, holidays or traffic routines. On the other hand, the later factor is a bit more complex as the behavior of every driver differs according to his working hours, areas of cruising or personal preferences. While the former factor can be dealt with techniques of predictive modeling or forecasting, ironically it might be a challenge to manage the supply side because it involves the prediction and analysis of various human minds' thinking patterns and general psychology.

With every advancement in technology and techniques such as behavioral targeting, the cab services are coming up with mechanisms to identify and control the behaviors of drivers. For example, according to the international cab service company Uber conducted a research and found out that surge pricing plays an important role in influencing the cruising route or standing point of a driver. As their commission rate remains constant during the peak hours or in the peak areas, drivers are motivated to drive to areas of surging demand so that the demand is met and the total transactions are also maximized. Moreover, another behavior that this study revealed was that when a driver wanted a break whilst not willing to turn off the application in order to benefit from the hourly payment package, he/she would often park the cab in between the cabs that were already sharing rides so that they would not get any requests automatically (Kyung Lee et al.)

Moreover, it can be observed that for some drivers, their work behavior is a function of the extent of their knowledge on the assignment algorithm of the company's app (Simonite 2015). So the more a driver understands the algorithm, the easier it is for him/her to generate workarounds to escape the unwanted assignments while for the driver who lacks in such an understanding, "tackling" the undesirable or uneconomical rides might not be possible and he/she might reject those rides.

Consequently, to deal with such general behavioral trends, the companies must design policies and mechanisms so that quality is not compromised. For example, to assure minimum cancellation rate at driver's end, Uber checks a driver's performance in terms of his/her rating and the number of rides accepted and rejected by the driver per day (Rosenblat 2016).

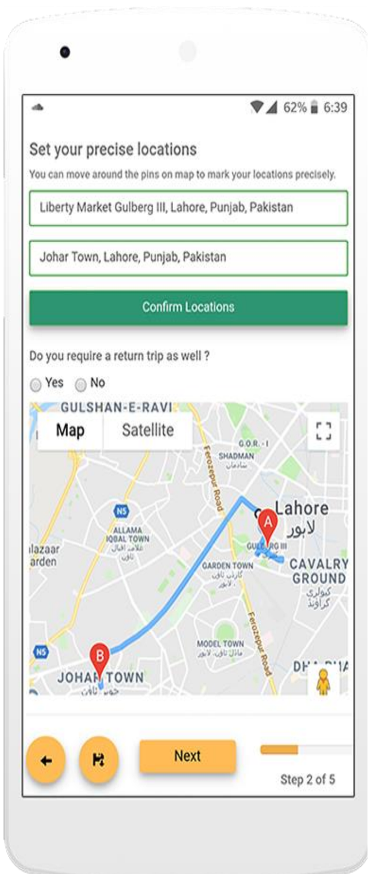
So, while worldwide companies have come up with general mechanisms to predict and control the actions of their employees; the drivers, yet their behavior changes depending upon the culture, policies, economy and various norms of the geographical region they are working in. Hence in order to get a deeper insight into this contractual relationship between the managers and the drivers, this study focuses upon one company from a South Asian country taken as a case study. The aim of this study is to focus on supply side of the local company "Shahi Sawari" that provides pick and drop service via rickshaws in the city of Lahore Pakistan. Therefore, through this study, a data-driven analysis of drivers' behaviors will be conducted via data exploration and unsupervised data analysis to identify any general trends or patterns and hence suggest a policy framework for Shahi Sawari to provide an optimal service.

Relationship with Customers

Shahi Sawari primarily targets students and working class that require an affordable conveyance to and from their colleges, universities and workplaces respectively.

Each customer can book a ride by downloading the "Shahi Sawari" app in his/her smart phone and then requesting the ride by specifying the pick-up and drop off location through maps via the app. Moreover, options of ride pooling and monthly package are also available to the customers.

The company requires a feedback from each customer at the end of every ride where the customer is required to give a rating out of 5 or any additional comments that correspond to the "average rating" for that particular driver. Furthermore, a customer can also contact the company at the helpline in case of any query or problem.



Relationship with Drivers

The company Shahi Sawari and the rickshaw drivers working for its share a **contractual relationship.**

Terms of the Contract

- Every driver is offered to choose one of the following three packages with minimum earning per day:

Package	**Minimum Guaranteed Earning per Day**
6 Hour	Rs. 1200
8 Hour	Rs. 1600
10 Hour	Rs. 2000

- According to the package chosen by a driver, he is respectively guaranteed a minimum amount of earning per day.
- If the driver earns below the minimum amount i.e. Rs. 1000 for a 6-Hour package, then company pays the driver on top of it i.e. Rs. 200 to make it up to the "guaranteed" earning.
- However, if the driver earns above or equal to the minimum earning decided, then there is no liability on the company to pay him any additional amount.
- Shahi Sawari draws its revenue from the 15% of the total fare or earning each driver makes per day (including the excess amount paid to the driver in case of earning below guaranteed amount).

Potential Risks:

The above-mentioned terms of the contract might invoke some erroneous behavior amongst some drivers in pursuit of working less but earning more which will deteriorate the service such as deliberately taking up less rides by either

- staying online yet taking up only private rides or
- choosing to park in a remote area where demand is the least etc. and then demanding the guaranteed amount from the company.

The drivers may also have some a "target earning per day" in their minds and they might start rejecting rides beyond that point.

In order to deal with such potential risks, Shahi Sawari imposes two additional conditions on the drivers:

- The average rating of each driver has to be above 4 stars
- Of the total rides allocated to each driver, he can only reject 15%

Focus of the Study

Given the terms and conditions of the contractual relationship between Shahi Sawari and the drivers and the dispatch algorithm used by the company, the supply (drivers) of the service is distinctively affected. The drivers exhibit various types of behavior for different motives including maximization of revenue, minimization of rides during duty hours and completion of only short trips as opposed to longer ones.

The thinking structure of every human mind is different; hence each driver might find a different way to "tackle" the contractual terms and achieve personal objectives. However, whatever the behavior might be, the impact on the service is huge. The goal is to employ unsupervised learning mechanisms in order to figure out patterns of distinctive driver behaviors and accordingly propose a policy suggestion to the company that optimizes their service. The study is aimed to analyze the following areas:

- Fairness of the Dispatch Algorithm System
- Relationship between the acceptance/rejection rate and fare earned by the drivers
- The variability of acceptance/rejection rate with respect to particular geographical area
- Relationship between the acceptance/rejection rate and the distance of the driver from the pickup location
- Relationship between the acceptance/rejection rate and the total trip distance (from pickup location to drop-off location)
- The fluctuation of demand with respect to time and day

Data Source and Description

The data used for the research has been extracted from the Shahi Sawari App.
It has details of 103600 trip requests over a period of 5 months elaborated in terms of the following variables:

Variables	Definitions
driver_id	Unique ID for each driver
trip_date	Dates extended over a period of five months
trip_time	The time at which a request for trip is processed
pickup_location_address	The address of pickup location
drop_location_address	The address of drop-off location
total_fare	The total fare per trip including discount, cancellation charges and wait charges
status	The trip status: Completed, Forcefully Completed, Rejected by Driver or Cancelled by customer
driver_rating	Rating (out of 5) per trip for the particular driver
distance	The distance (in meters) covered per trip

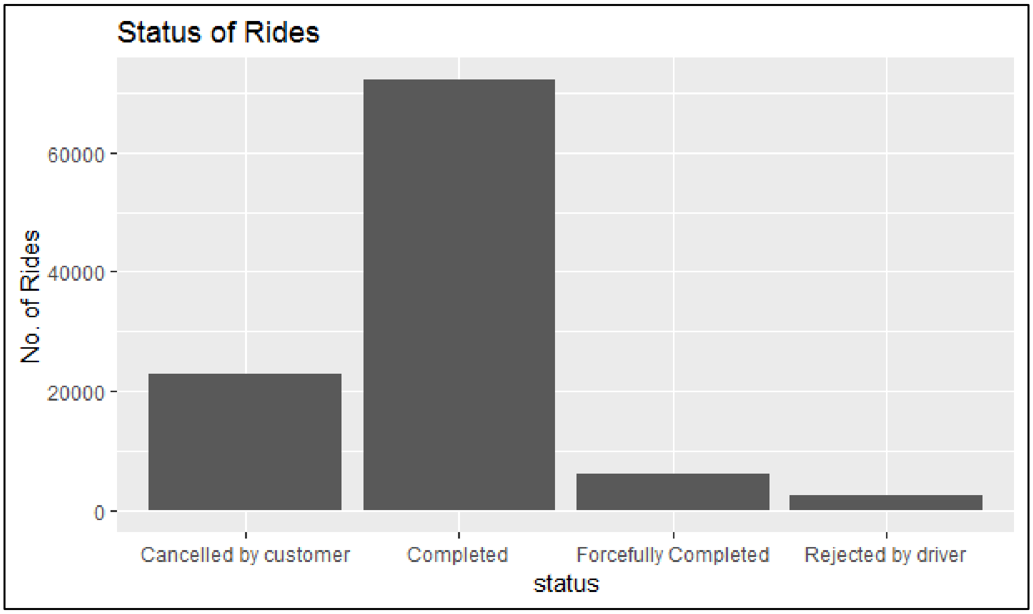
ride_time	The total time taken (in minutes) to reach drop-off location from pickup location
pickup_latitude, pickup_longitude, drop_latitude,drop_longitude,	The longitude and latitude of pickup and drop-off location of the trip
driver_accept_latitude, driver_accept_longitude	The longitude and latitude of the location where the driver is when he accepts the trip
ride_distance_from_google	The total distance (in meters) covered per trip
accept_distance	The distance of the “accept location” to the “pickup location”

Data Exploration

To understand the general nature and relationships of the data collected over five months, various visualizations were created and analyzed. The most significant results are as follows:

- **Visualization # 1 Proportion of Overall Ride Status**

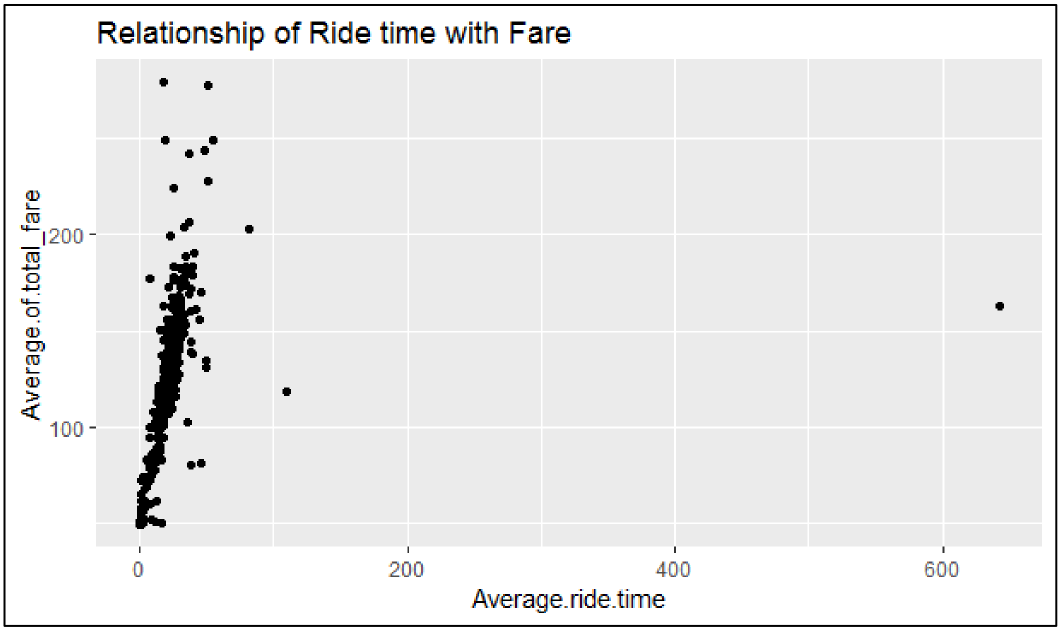
The analysis of ride status shows that during the period of five months, 69.6% of the total rides were completed. 5.9% if the rides were forcefully completed. Considering the rejection of rides, it can be seen that 21.9% of the rides were rejected by the customers themselves whereas 2.4% of the rides were rejected by the drivers themselves. The visual representation of the results can be seen below:



It can be observed that the overall rejection by driver rate is quite low. This can be explained by the fact that the terms of the contract restrict the proportion of rejection (15%) per driver.

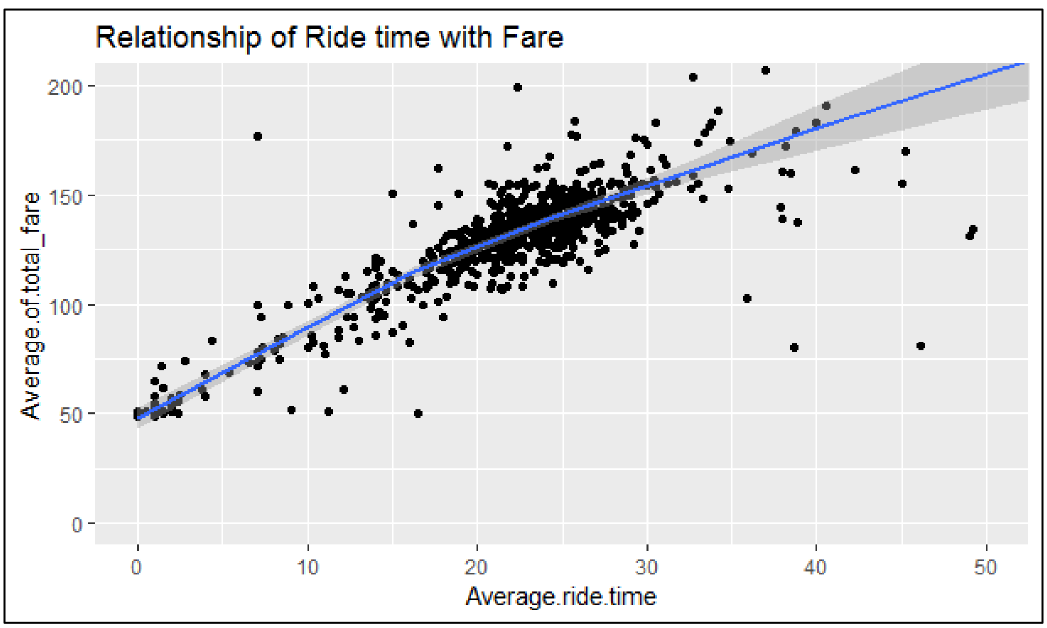
- **Visualization # 2 Relationship between Average Ride Time and Average Total Fare per Driver**

It can be generally assumed that if a driver on average takes trips that take a longer time to be completed, then he would earn more as the fare per minute would be higher in these trips. To verify the assumption, the averages for ride time and total fare were calculated for every unique driver ID and a scatter plot was made between the average ride time and average total fare:



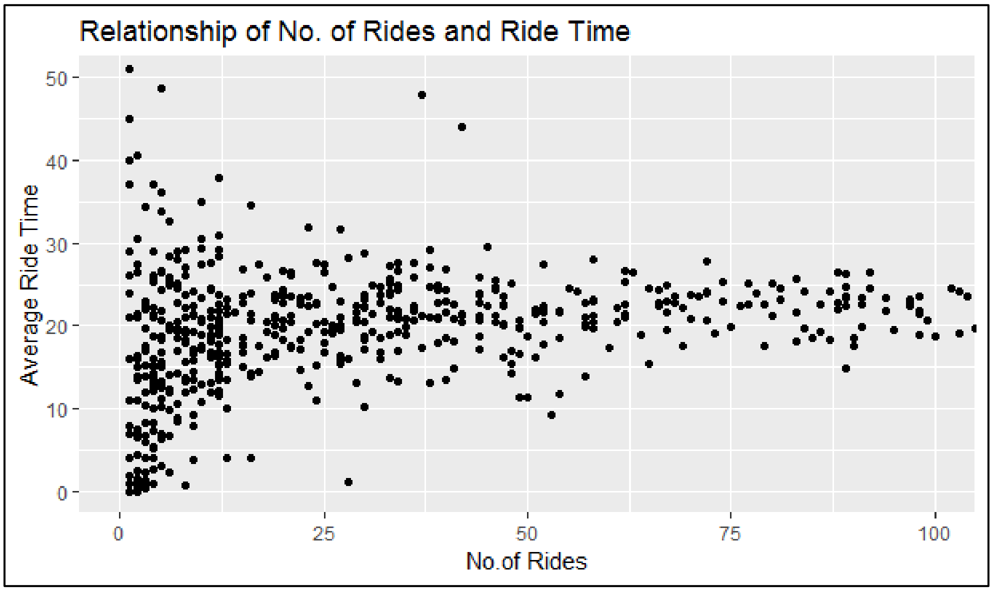
Apart from the outliers, it can be seen in the illustration below that there is a positive relationship between ride time and total fare which confirms the initial assumption.

The removal of outliers shows that most of the rides are clustered between an average ride time of 20-30 minutes and the average fare ranges between Rs.100-200.



- **Visualization # 3 Relationship between Average Ride Time and Average Number of Rides per Driver**

In order to inquire into the relationship between the number of rides and the average ride time per drive, exploration was initiated with a hypothesis that the drivers who complete long distance rides are likely to complete fewer rides on average per day and a scatter plot was made:

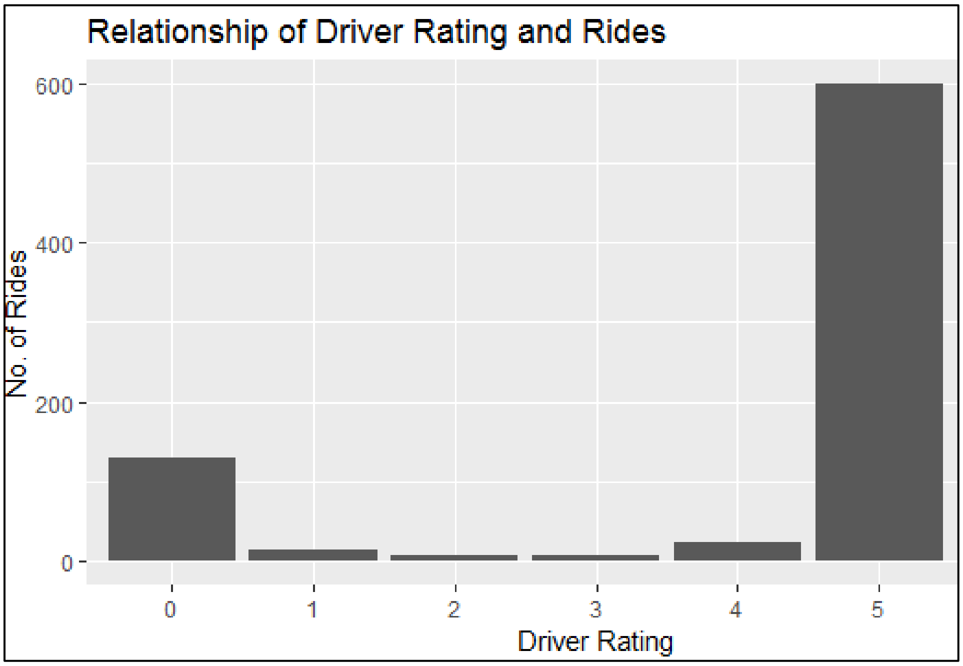


The analysis shows no specific trend. The rides are rather uniformly distributed. It can yet be seen from the figure that the drivers who completed rides with an average time of 40 minutes and above had lower rides compared to the drivers who completed rides of a shorter duration.

- **Visualization # 4 Relationship between Driver Rating and Total Number of Rides per Driver**

It is important to determine if the driver with higher ratings get greater number of rides as compared to drivers with lower earnings. According to the analysis, among all the rides carried out during the period of 5 months, most of them have been given high ratings. Out of the total rides, 71.56% of the ratings were 5 whereas a rating of 4 was received 2.89% times. The consumers didn't give any rating 6.57% of the time. The ratings of 1, 2, 3 and 4 were received 1.79%, 0.7%, 0.95% and 2.87% of the times respectively.

```
          0           1           2           3           4           5        N.A.
0.155316607 0.017921147 0.007168459 0.009557945 0.028673835 0.715651135 0.065710872
```



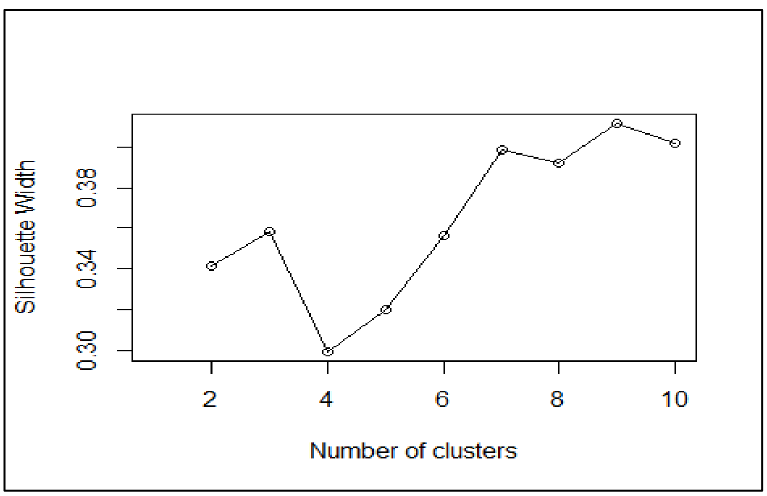
Unsupervised Learning Techniques

We carried out unsupervised learning techniques such as cluster analysis and association rules to discern the underlying patterns which rest of the analysis couldn't. Cluster Analysis was employed, and partitioning methods were used to determine that the rejection rates of drivers differ as a result of the categorization of ratings, working days, working hours and number of rides the driver receives. **Association Rules** is a measure of co-occurrence and correlation. It was used to identify the rules or combinations of driver's information that would be associated with a high or low rejection rate of a driver.

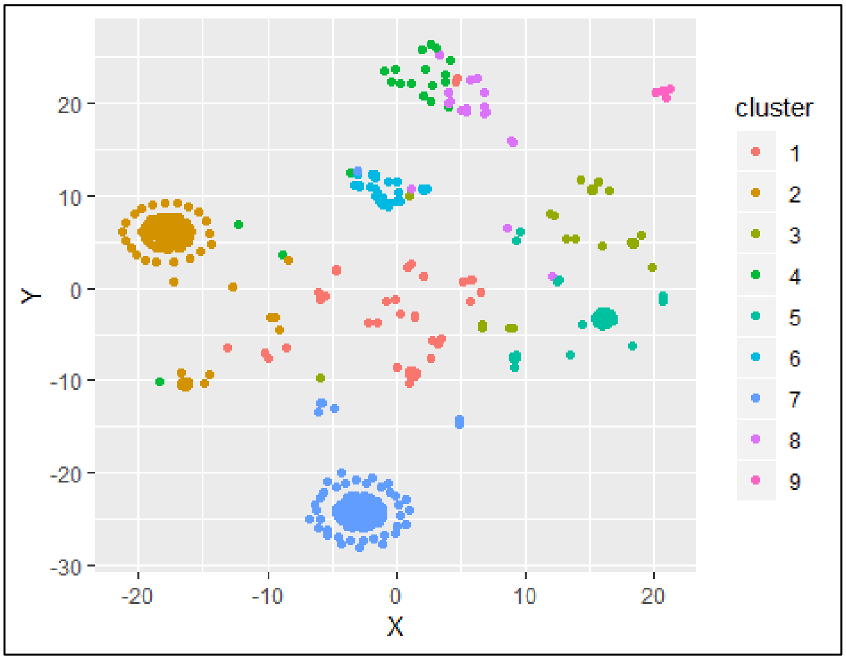
In order to employ both techniques and make the results easily interpretable, the variables were converted into categorical variables. From the driver's perspective, the factors affecting the rejection of rides were shortlisted as, working hours, working days, rating, the average number of rides' classification per day. All the variables were divided into 3 categories. In case of working hours, the drivers who worked **more than 7 hours** were classified as **"High",** those who worked **between 5-7 hours** were labelled as **"Medium**" and the remaining were referred as **"Low".** Similarly, in case of working days, the drivers who worked **more than 60 days** fell under the category of **"High**", the ones who worked **less than 30 days** were classified as **"Low"** and the remaining as **"Medium**". In case of the average ratings, the rating of **4 and above** was classified as High and the rating of **3 and below** was labelled as Low. While classifying the average number of rides completed per driver/day, the ride distribution was named as **"Above Average" (for rides greater than 5), "Average" (for rides between 3-5) and "Below Average" (for rides lower than 3/day).**

- **Cluster Analysis**

As the data was prepared based on categorical classification, clustering based upon Gower Distance was applied which converted k categories in k binary columns and then made the distance matrix on the basis of Dice coefficient. The Silhouette validation technique was employed to see how many clusters could be made given the data. It suggested that 9 clusters should be made as seen below.



In order to inspect how these clusters differed, their medoids were analyzed. To visualize the results, t-distributed stochastic neighborhood embedding (t-SNE) is used. The figure below shows the 9 well-separated clusters that partitioning around medoids (PAM) was able to detect.



The analysis reveals some interesting patterns which are in line with the previous analysis. The most interesting observation is that the drivers who completed above than average rides had lower ratings as compared to the drivers who performed below average on this dimension and had higher rating classification. The drivers who completed average rides had medium ratings i.e. a rating around 3 or 3.5.

While analyzing from the perspective of working hours, it can be seen that this variable is a clear differentiator among the clusters. It can be clearly seen that the drivers who performed below average were solely the ones who had low working hours/day and those who performed above average had a greater number of working hours/day. While looking at number of working days, it can be seen that above than average drivers worked for greater number of days than average or below than average drivers.

It is also interesting to note that the analysis doesn't show any relation between the rejection rate and ride classification i.e. it doesn't show that the number of rides that the driver completes in a day has anything to do with the number of rides that he rejects. This implies that there might be other factors such as pick up distance from driver location etc. which might be contributing to the rides' rejection pattern. The results are highlighted below for further referencing.

	Driver.ID	Ride.Classification	Rating.Classification	Working.Days	Working.Hours.Day	X..rejection
655	65048	Average	Medium	Low	Medium	Low
814	77049	Below Average	High	Low	Low	Low
697	66656	Above Average	Low	Medium	High	Low
11	20738	Below Average	High	Low	Low	Medium
817	77298	Above Average	Low	Low	High	Low
666	65444	Average	Medium	Low	Medium	High
823	77729	Below Average	Low	Low	Low	Low
713	68795	Above Average	Low	Low	High	Medium
846	NA					Low

- **Association Rules**

Association rules were used to identify the variables that are associated with the rejection rates of drivers. The R package, 'arules', was used for mining association rules within the dataset and to report the combinations of variables that contribute to the rejection rates of drivers. Since transaction matrix requires that variables are stored in the form of factors, the numeric variables were converted to factors with different labels assigned to different levels in the factors. The data was then stored in the transaction matrix in a sparse format. The Apriori function searched over different combinations of the variables with one item in the consequent and four in the antecedent.

High Rejection Rate:

A subset of rules was created to find the rules that would result in a high rejection rate of drivers. With a **minimum support of 0.01, minimum confidence of 0.1**. Support is an indication of how frequently the itemset appears in the dataset. The support of e.g. the number of working hours/day with respect to rejection rate is defined as the proportion of transactions in the dataset which contains the itemset working hours. In this case, a minimum support of 0.01 means that the variables should appear at least 1% of the time in all the entries. Confidence is an indication of how often the rule has been found to be true. A confidence of 0.1 suggests that there is at least 10% chance that if Rejection Rate for example is high then the driver rating must be low. In our dataset, there are about 1,177 rules that reflect the different possible combinations of variables that would result in high rejection rate of drivers. The minimum confidence was set to 0.1 in order to ensure that the rules that the algorithm returned would be strong predictors of rejection rate. The rules are sorted according to the lift whereby the rules with the highest lift ratio are listed first. The results display that the antecedents of the top eight rules follow more or less the same structure with some variables appearing in all eight rules.

```
    lhs                              rhs                    support confidence     lift count
[1] {Ride.Classification= Average,
     Working.Days= Medium}        => {X..rejection=High} 0.01063830  0.2250000 3.120492     9
[2] {Rating.Classification=Low,
     Working.Days= Medium}        => {X..rejection=High} 0.01182033  0.1538462 2.133670    10
[3] {Ride.Classification= Average,
     Rating.Classification=Low}   => {X..rejection=High} 0.01654846  0.1333333 1.849180    14
[4] {Working.Days= Medium}        => {X..rejection=High} 0.01536643  0.1214953 1.685001    13
[5] {Ride.Classification= Average} => {X..rejection=High} 0.02836879  0.1194030 1.655982    24
[6] {Ride.Classification= Average,
     Working.Hours.Day= Medium}   => {X..rejection=High} 0.01300236  0.1134021 1.572756    11
[7] {Working.Hours.Day= Medium}   => {X..rejection=High} 0.02482270  0.1060606 1.470939    21
[8] {Rating.Classification= Medium} => {X..rejection=High} 0.02482270  0.1024390 1.420712    21
```

Antecedents:

· Ride Classification = Average

· Working Days= Medium

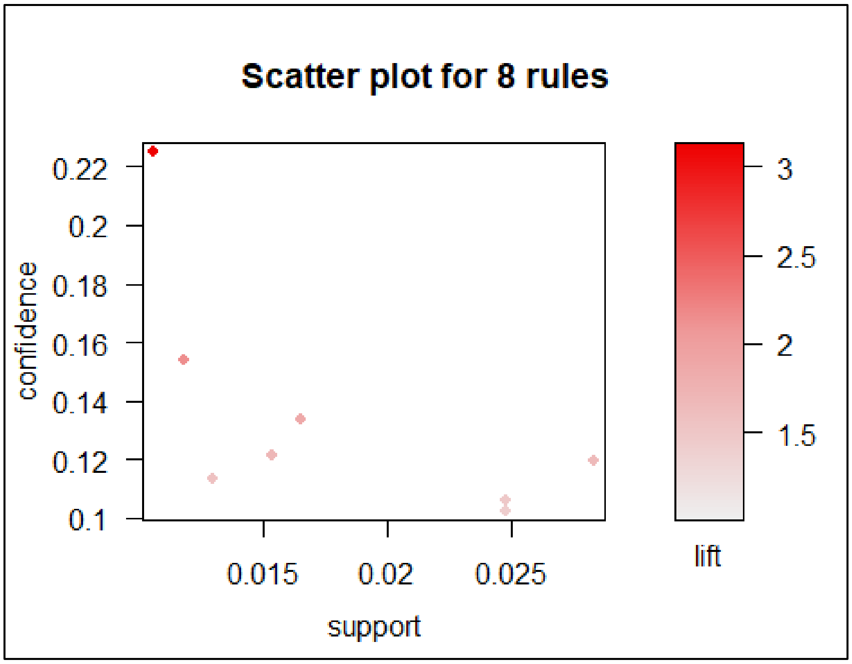
· Working Hours per day = Medium

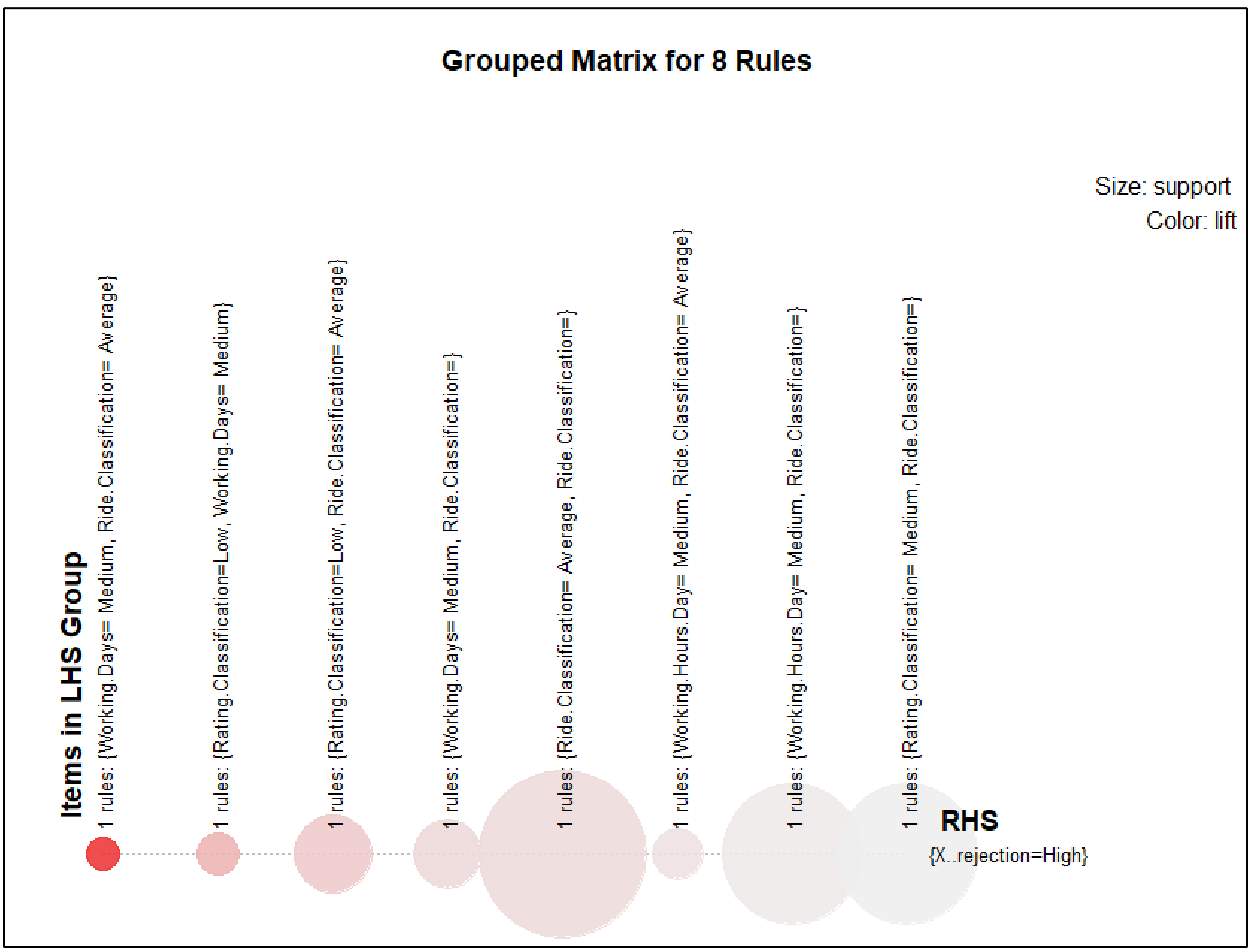
· Rating Classification = Low

Consequent:

Percentage Rejection Rate = High

The results show that the drivers who reject the highest number of rides are usually the ones who complete average number of rides/day i.e. around 3-4 rides/day. Moreover, their working hours are usually medium ranging from 5-7 hours and the working days are around 30. The most interesting observation is that their average rating is low. It makes intuitive sense that the drivers who reject rides often will have poor ratings on average. The lift ratio of the second rule, for example, indicates that the chances of a driver rejecting a ride is lifted by a factor of 2.13 if the driver has a lower rating and works for medium number of working days. The figure below can be seen for visual representation of top 8 rules.


Scatter plot for 8 rules
confidence
0.22
0.2
0.18
0.16
0.14
0.12
0.1
0.015
0.02
0.025
support
3
2.5
2
1.5
lift


Grouped Matrix for 8 Rules
Size: support
Color: lift
Items in LHS Group
1 rules: {Working.Days= Medium, Ride.Classification= Average}
1 rules: {Rating.Classification=Low, Working.Days= Medium}
1 rules: {Rating.Classification=Low, Ride.Classification= Average}
1 rules: {Working.Days= Medium, Ride.Classification=}
1 rules: {Ride.Classification= Average, Ride.Classification=}
1 rules: {Working.Hours.Day= Medium, Ride.Classification= Average}
1 rules: {Working.Hours.Day= Medium, Ride.Classification=}
1 rules: {Rating.Classification= Medium, Ride.Classification=}
RHS
{X..rejection=High}

Low Rejection Rate:

A subset of rules was created to find the rules that would result in a low rejection rate of drivers. With a minimum **support of 0.01, minimum confidence of 0.7**. This means that there are about 406 rules that reflect the different possible combinations of variables that would result in low rejection rate of drivers.

Antecedents:

· Ride Classification = Above Average

· Working Days= Low/Medium/High

· Working Hours per day = High

· Rating Classification = High

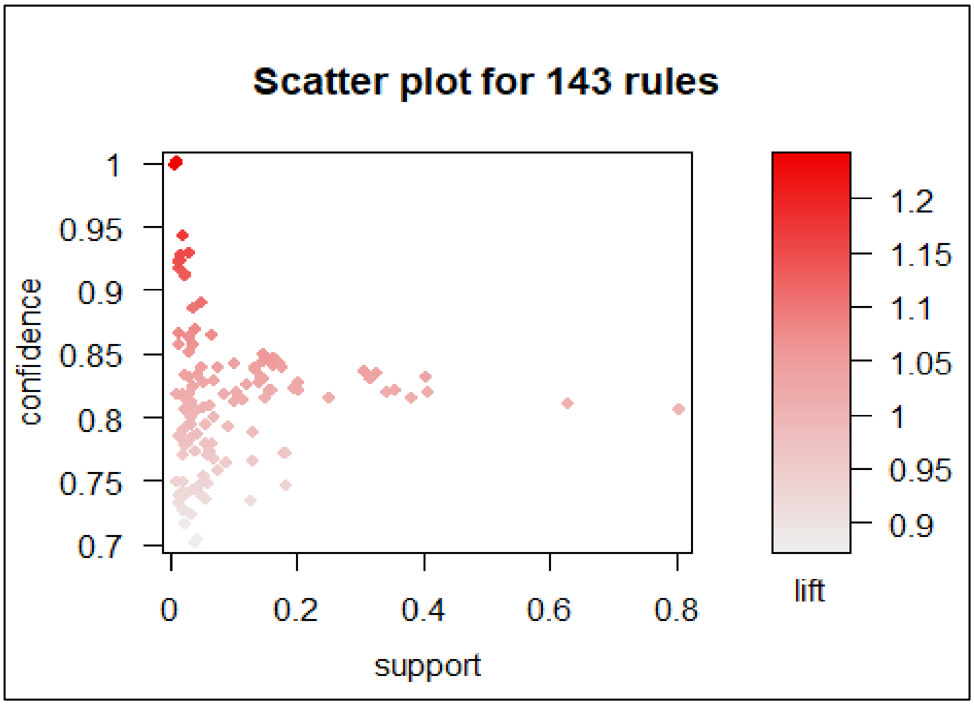
Consequent:

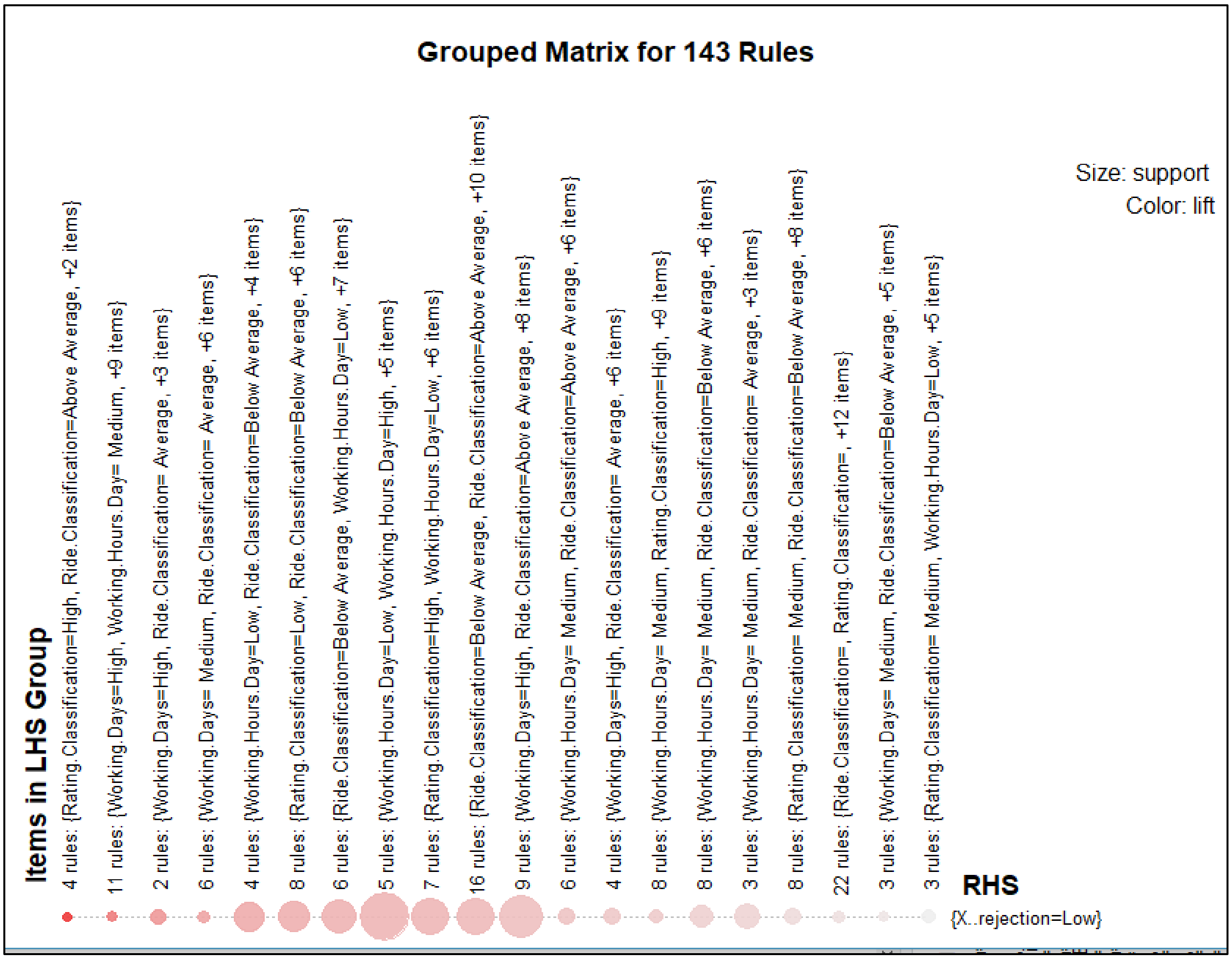
Percentage Rejection Rate = Low

The top 10 rules are as follows:

```
     lhs                                    rhs                   support confidence     lift count
[1]  {Ride.Classification=Above Average,
      Rating.Classification=High}        => {X..rejection=Low} 0.01418440  1.0000000 1.240469    12
[2]  {Ride.Classification=Above Average,
      Rating.Classification=High,
      Working.Hours.Day=High}            => {X..rejection=Low} 0.01300236  1.0000000 1.240469    11
[3]  {Ride.Classification=Above Average,
      Rating.Classification=High,
      Working.Days=Low}                  => {X..rejection=Low} 0.01182033  1.0000000 1.240469    10
[4]  {Ride.Classification=Above Average,
      Rating.Classification=High,
      Working.Days=Low,
      Working.Hours.Day=High}            => {X..rejection=Low} 0.01063830  1.0000000 1.240469     9
[5]  {Ride.Classification=Below Average,
      Working.Days= Medium}              => {X..rejection=Low} 0.01891253  0.9411765 1.167500    16
[6]  {Rating.Classification=Low,
      Working.Days=High,
      Working.Hours.Day=High}            => {X..rejection=Low} 0.03073286  0.9285714 1.151864    26
[7]  {Ride.Classification= Average,
      Rating.Classification= Medium,
      Working.Days=Low,
      Working.Hours.Day=High}            => {X..rejection=Low} 0.01536643  0.9285714 1.151864    13
[8]  {Ride.Classification= Average,
      Rating.Classification=High,
      Working.Hours.Day= Medium}         => {X..rejection=Low} 0.01418440  0.9230769 1.145048    12
[9]  {Ride.Classification=Above Average,
      Rating.Classification= Medium,
      Working.Hours.Day= Medium}         => {X..rejection=Low} 0.01418440  0.9230769 1.145048    12
[10] {Ride.Classification= Average,
      Rating.Classification=High,
      Working.Days=Low,
      Working.Hours.Day= Medium}         => {X..rejection=Low} 0.01418440  0.9230769 1.145048    12
>
```

The results show that the drivers who complete above than average rides, have high rating and their working hours/day are higher reject fewer rides. The lift ratio of the second rule, for example, indicates that the chances that a driver rejects fewer rides is lifted by a factor of 1.24 if the driver has a high rating, works for high number of working hours/day and completes more than average rides. The fourth rule also suggests that if the above mentioned 3 attributes are fulfilled and even if the driver has low number of working days, it would still lead the driver to rejecting fewer number of rides. In cases where the the working days are high along with the working hours, even if the rating of the driver is low, it would lead to fewer rejection of rides. The scatterplot of the visualization of all the rules is shown below along with the groped matrix which provides the summary of the rules along with their relative importance in terms of confidence and lift.





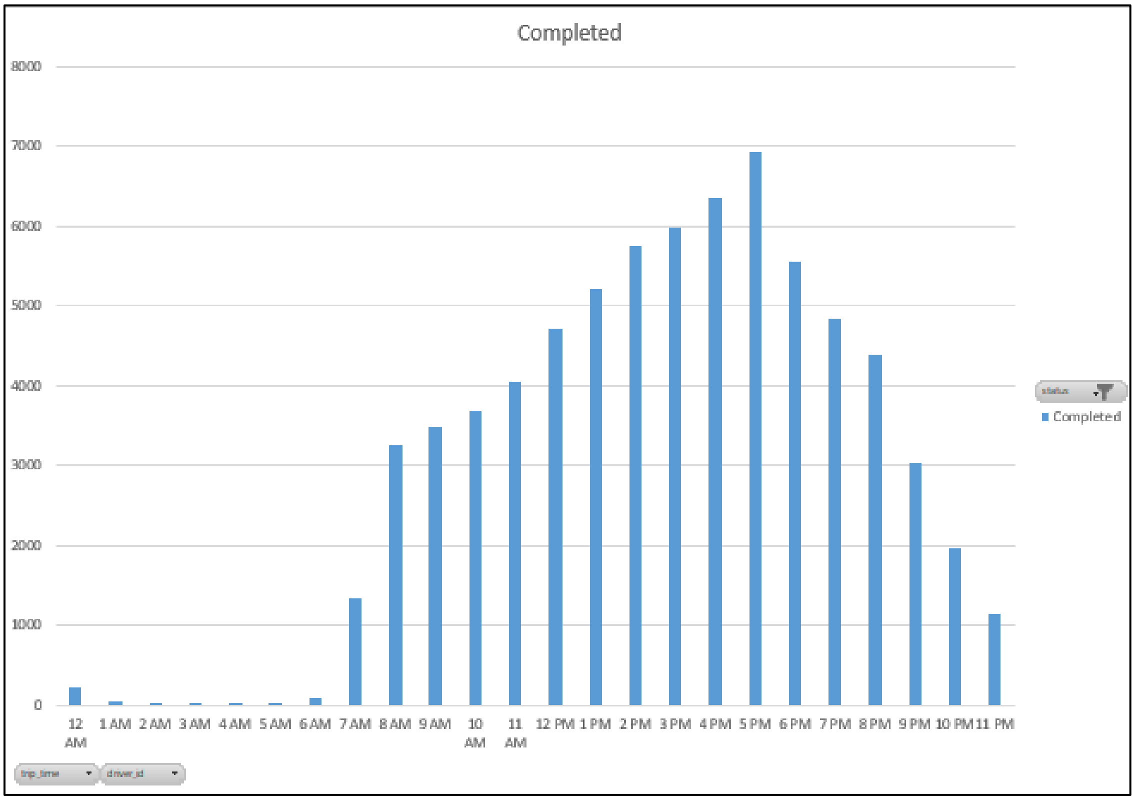
Analysis of Driver Behavior with respect to Day Hours

The demand for rickshaws (ride demand) varies throughout the day depending upon the time of the day. Naturally, the demand is highest during peak hours in the morning and in the evening. With the fluctuating demand, the drivers also develop coping mechanisms to boost up their profit and reduce work hours as much as they can. Hence, one aspect of this study was to analyze such behavior and determine any pattern that exists between it and hours of the day.

To conduct the analysis a visualization was created by plotting the status of the rides against the timings of the day:

A. Status: Complete

- **Graph:**

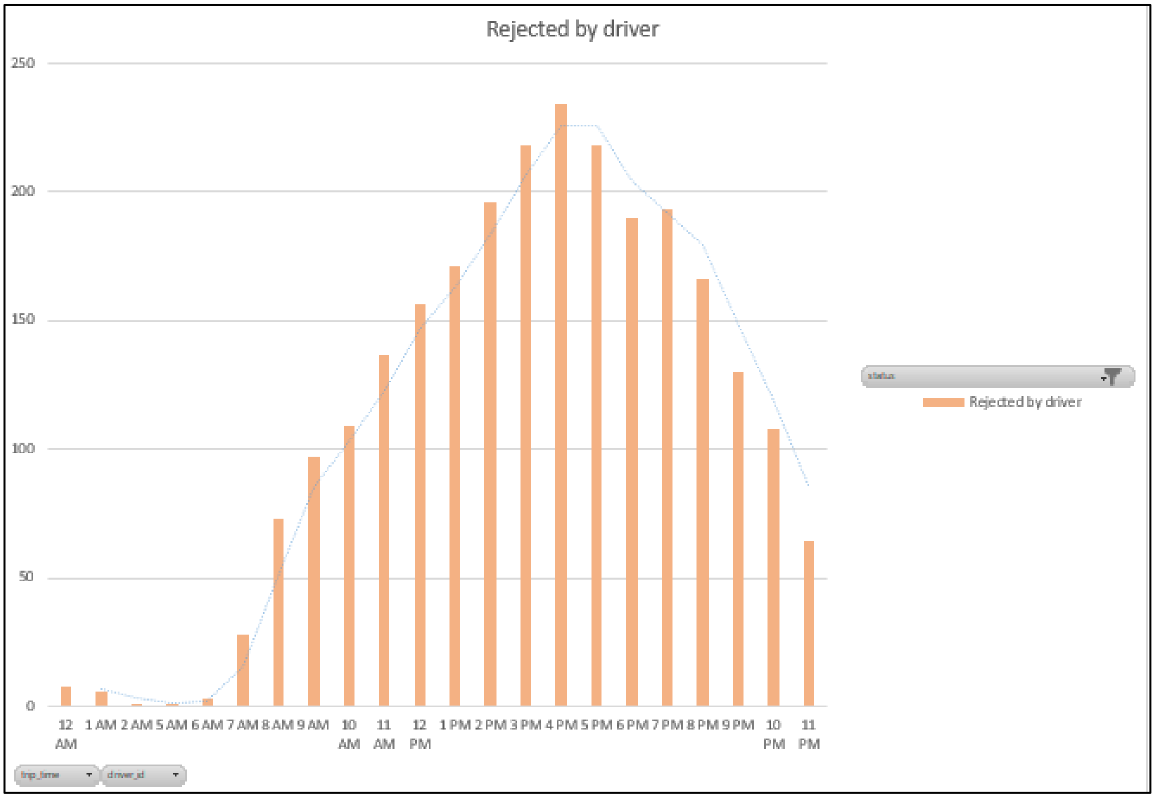


- **Interpretation**

  It was analysed that the maximum number of rides were completed during the evening from 4pm-5pm. The rides are normally distributed throughout the day. The patterns suggest that as expected, there is a dip in the number of rides at night after 11pm**.**

B. Status: Rejected

- **Graph B.1:**

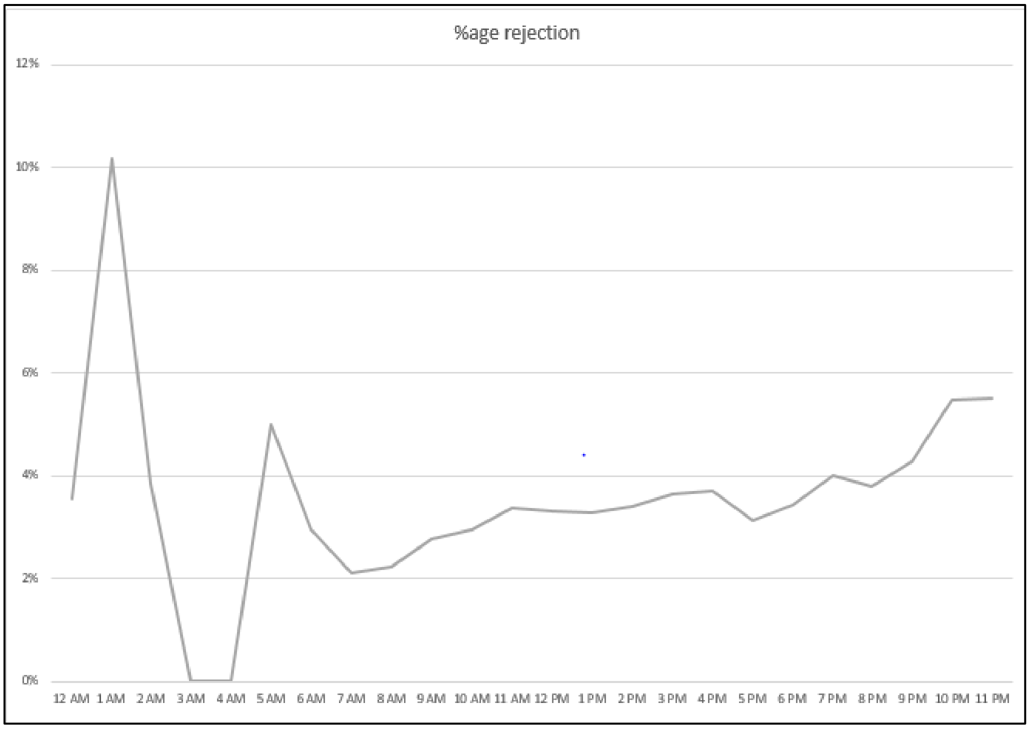


- **Interpretation**

  If the rejection of rides by driver is analyzed, it can be seen that it also follows the same trend. The maximum number of rides are rejected around 4pm. This can be primarily because the riders receive the maximum amount of requests at that time and there are only a limited amount of requests that they can cater to. Apart from this, it is also possible that by this time, the drivers achieve their daily earning target which is why they start cancelling rides.

- **Graph B.2**

It is important to see relative drivers' cancellation pattern to the total number of the rides. In order to calculate this, percentage cancellation of rides per hour was calculated and plotted as:



- **Interpretation**

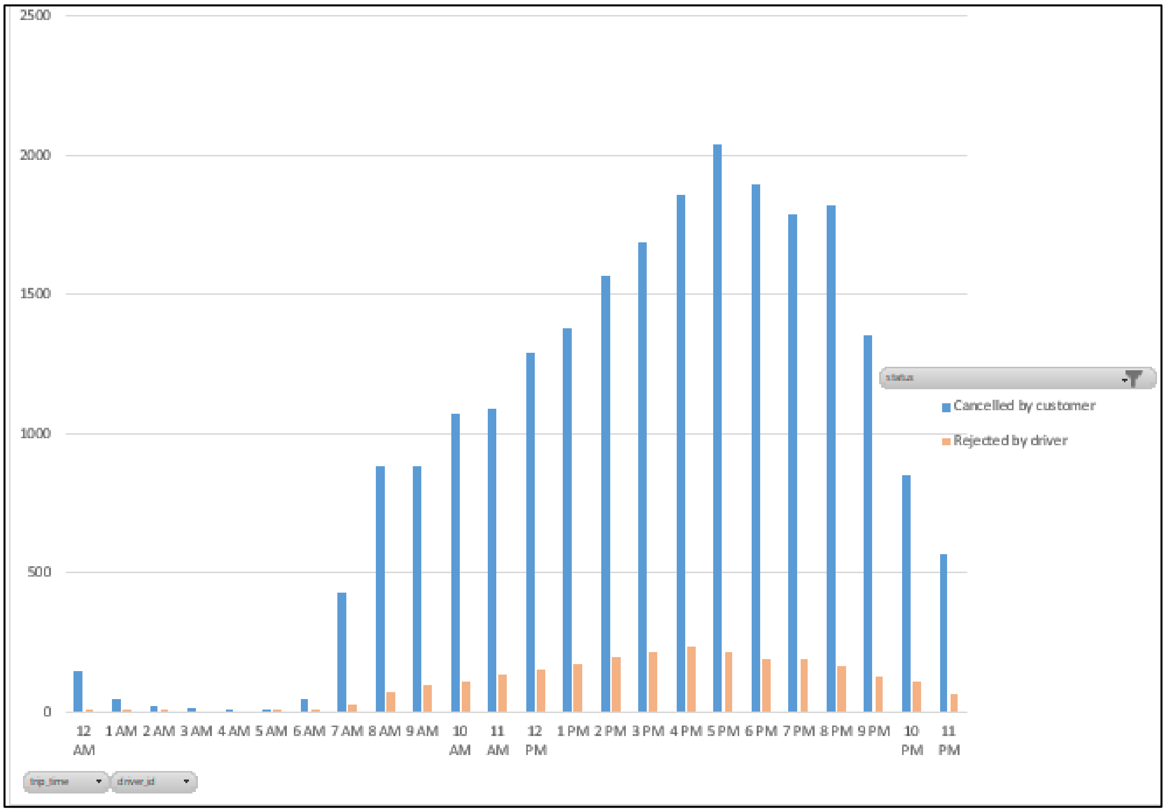
It can be seen that percentage of rejection by drivers keeps on increasing from 7 am to 11 am. The rejection rate then dips a little from 12 pm to 2 pm and decreases significantly from 4 pm to 5pm. This can be explained by the fact that most of the rides are allocated to the drivers during the peak hours and this is the time where a driver can expect to earn the maximum, therefore, he does not reject many rides then. The rejection percentage then keeps on increasing till 2 am and drops at 3am to 4 am probably due to the fact that no rides during these hours are actually requested.

C: Comparison of Status Cancelled by Customer and Status Rejected by Driver

The uncompleted rides are of two types majorly, the ones that are cancelled by customer himself or those that are rejected. The proportion of such rides against the time frames within a day can be more meaningful if plotted an analyzed in comparison.

Hence through such a comparison it can be seen that on average, the customers reject fewer rides as compared to the drivers. This is insightful because it explains the fact that since customers are penalized in the form of "cancellation charges" for every ride they cancel and while drivers have to keep the overall percentage of rejection under 15% as per the contract terms, drivers have a slightly more flexible cushion to reject rides as opposed to cancellation by the customers.

The comparison plot is given below:

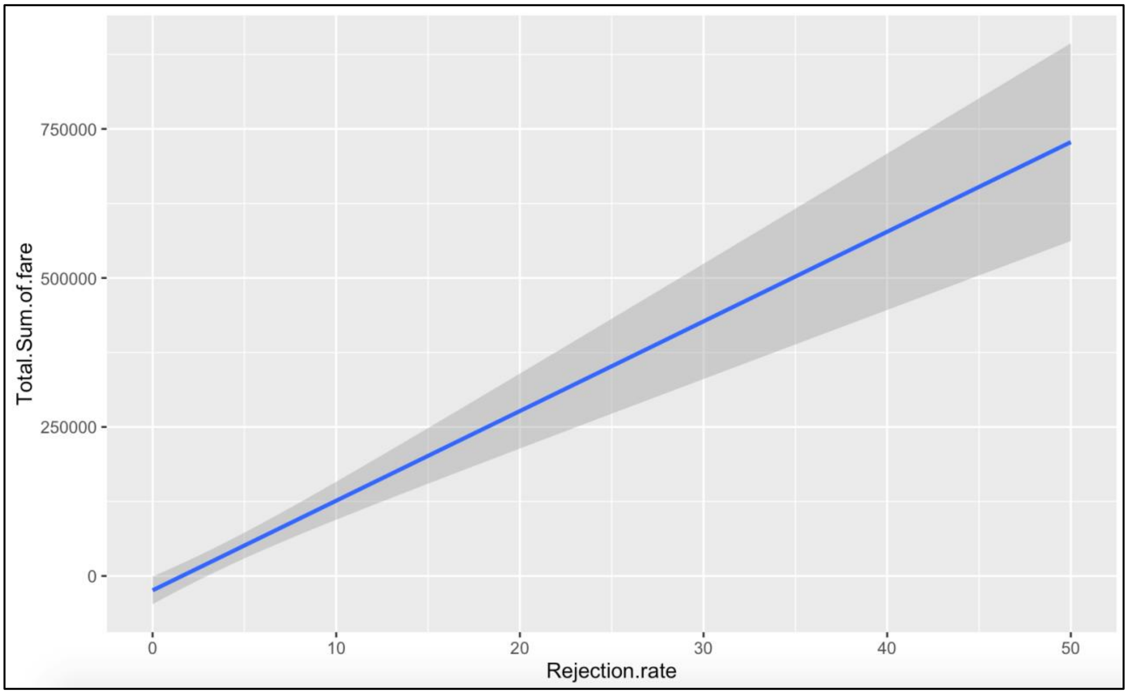


Inference:

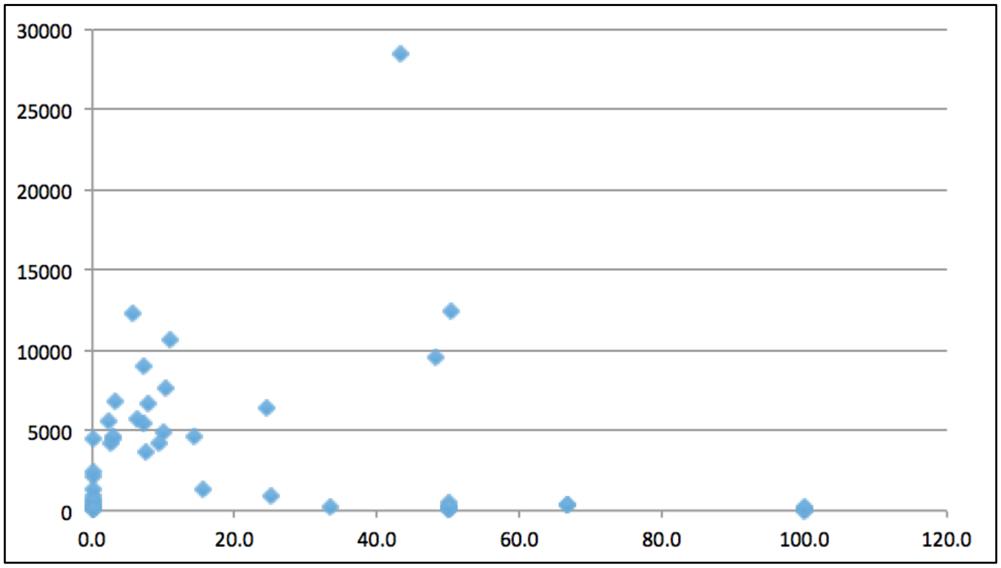
The overview of these visualizations reveals that the cancellation by customers and rejection by drivers roughly follow a normal distribution. However, if a percentage trend of rejection is observed with respect to total rides initiated, it is observed that drivers are more likely to reject rides in later hours of the night. This may be because of late hours, drivers are most likely to have earn the "benchmark" earning during the peak hours that he aims to make in a day. Hence policies can be revised to reduce the rejection rate in later hours.

Analysis of Drivers' Acceptance and Rejection Behavior with respect to Fare

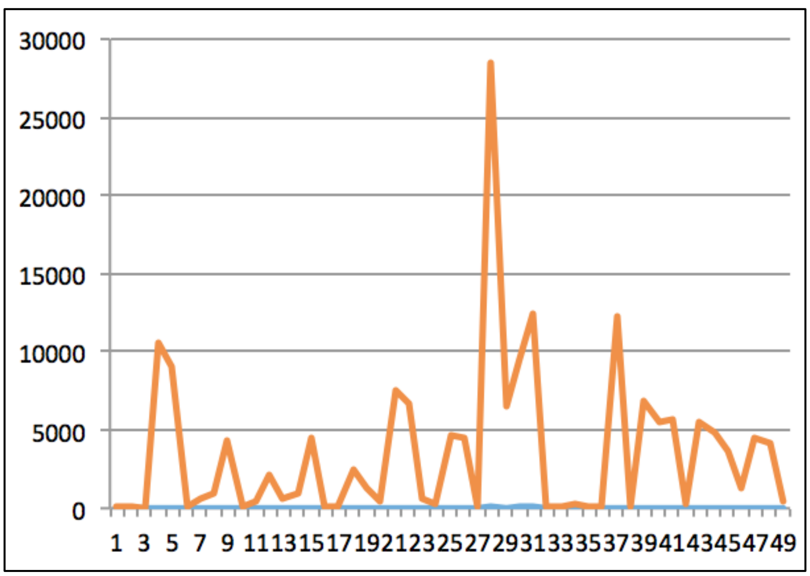
One method of checking driver behavior is by observing their ride acceptance versus cancellation behavior. The study on the basis of this method was begun by seeing how much money is earned per driver and comparing it to the amount of cancellations they have each made in total. This showed if there is any connection between these two factors.



The graph shows that as a driver's rejection rate increases, so does their total sum of fares. The standard errors also increase with it, showing that though the average fares are higher at higher rejection rate levels, the distribution is much more widely spread out, while there is more precision at lower rejection rates. This, however, may be caused by very low rejection rates by inactive drivers.

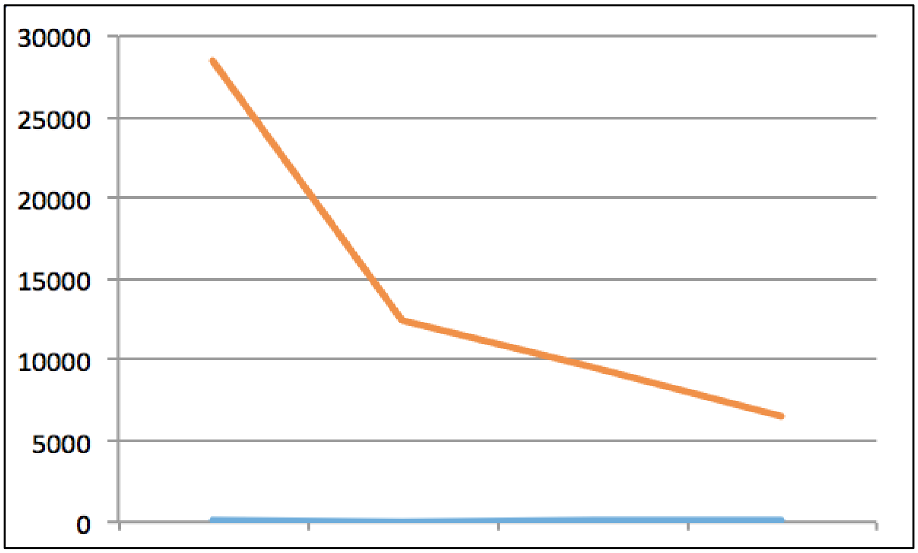


When looking at the rejection rates and fares for drivers by looking at their overall statistics, a more accurate analysis of the data can be done. What can be seen here is that a majority of active drivers have a rejection rate between 0 and 20%. The fare amounts for drivers who have a higher rejection rate is overall lower than those in the first 20%. This makes sense as those who are able to make the most money are rejecting less rides. What we want to see next is whether this trend can be seen by looking at drivers on a monthly basis

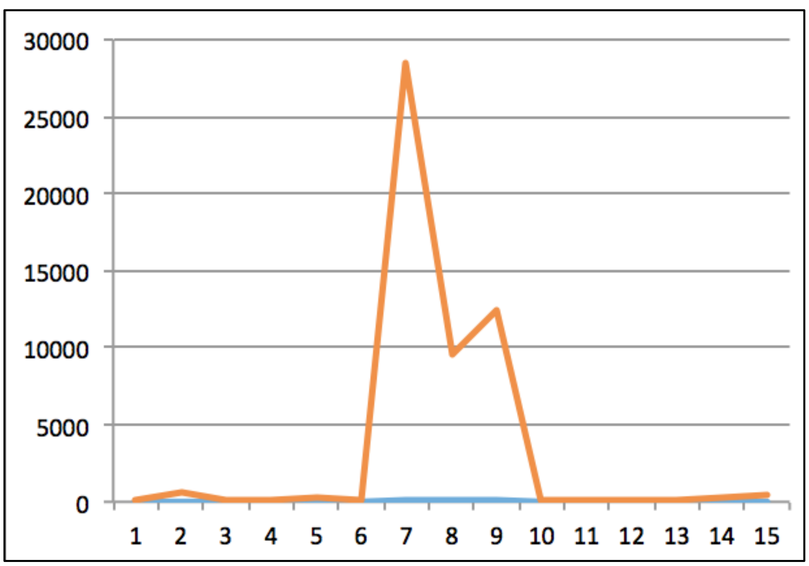


The monthly trend for driver fare in comparison to cancellations shows that those with very few cancellations make a lot of money while those with more cancellations on average make less.

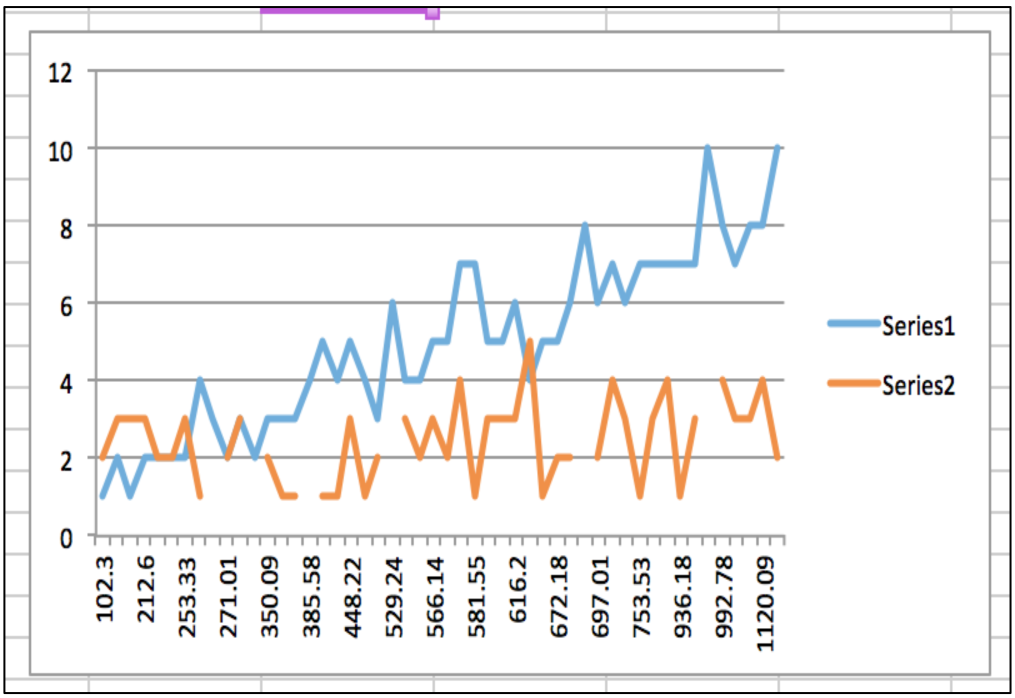
Confusion however is created from the large spike in the graph where those making the most amount of money have between 27 and 32 cancellations. This can be thought of being caused by more active drivers.



When looking at data of the 15 most active drivers, it was seen that those with lower rejections per month collected the highest amount of fare. The fare decreased almost linearly until a bend after which the graph became linear again. This shows that the most active members are making the most fare on average as well as rejecting the least number of rides. The question that still needs to be answered is why there is a sharp increase in the middle of the graph for all drivers.



The graph above is made for the most problematic drivers, those with a rejection rate of 40% and above. Some of these drivers are found to make high amounts of monthly income while rejecting large amounts of rides as well. The problem of drivers only accepting profitable rides can be found here. Though these drivers are few in number, they are the ones the company needs to work on taking preventative measures against the most. Though most drivers are not trying to only find rides in which they can make the most amount of money, those that are will quickly influence their peers and cause this habit to spread throughout the business.



The above graph shows the average daily acceptance and rejection pattern against fare. The blue line represents the number of accepted and completed rides while orange shows the number of rejected rides. The first interesting thing to note is the lack of pattern of rejected rides as fare goes up for a single day. There is less variance when the fare is low, but as it increases the slope of the lines increases quite visibly.

Analysis of Ride Allocation System

- **Data Preparation and Filtration**

In order to inquire into the proportion of total rides allocated to each driver, some data filtration was required:

1. The total number of drivers which were initially 846 was reduced to 823 after the removal of the outliers and the data points which had errors. There were multiple entries in the ride data column which had negative values. These were detected as errors as the average number of rides per driver can't take a value less than zero.

2. Total number of working days in the span of five months and average daily working hours were calculated for each driver. Drivers with total working days less than 15 and average daily working hours less than 7 were excluded from analysis.

3. Total number of rides (a) allocated to each filtered driver within the span of five months was calculated.

4. Total number of hours worked (b) by each filtered driver was also found.

5. Number of average rides per Hour allocated to each driver was found by dividing the (a) by (b).

6. Drivers were sorted into two categories "Drivers with More Rides" and "Drivers with Less Rides" with average rider per hour, 0.8 and 0.5 respectively with respect to the total average of 0.62 for separate analysis.

- **Distribution of Rides and Fairness of Allocation**

The next part of the analysis was to see if the rides are fairly allocated to the drivers and whether there are any differences in the distribution of rides. However, as the sample size was still very large, we progressed with creating several scenarios where the variables such as average no. of rides per day/ driver, number of working days of the drivers, driver rating etc. were varied and their impact was analyzed.

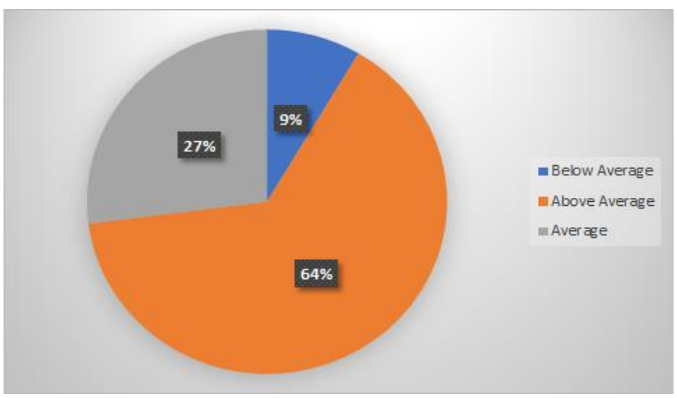
Before carrying out the main analysis, the performance of the drivers was divided into 3 categories.

1. Below Average
2. Average
3. Above Average

The drivers who complete 3 rides or less on average per day were classified as Below Average. The drivers who completed rides between 3-5/ day were labelled as Average while those who completed more than 5 rides per day were classified as Above Average. The aim was to see if the proportion of Above Average and Below Average drivers changes as other variables are changed.

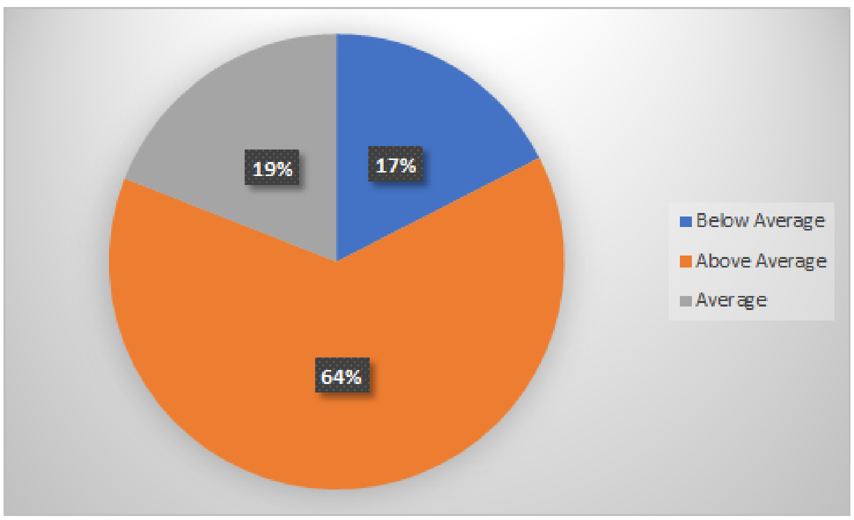
Scenario 1:

In this case, only drivers which had working hours of 7 and above were selected. This left us with a total of 259 observations. Among these observations, it could be clearly seen that 64% of the drivers fell in the category of above average i.e. they completed more than 5 rides per day. On the other hand, 27% of the drivers were labelled as average whereas 9% of the drivers were classified as below average. This shows a skewed distribution towards the left. It must be taken into account that this scenario considers all the drivers which were available within the 5-month time frame. These include drivers who were working consistently for all months as well as those who only worked for a period of few days.



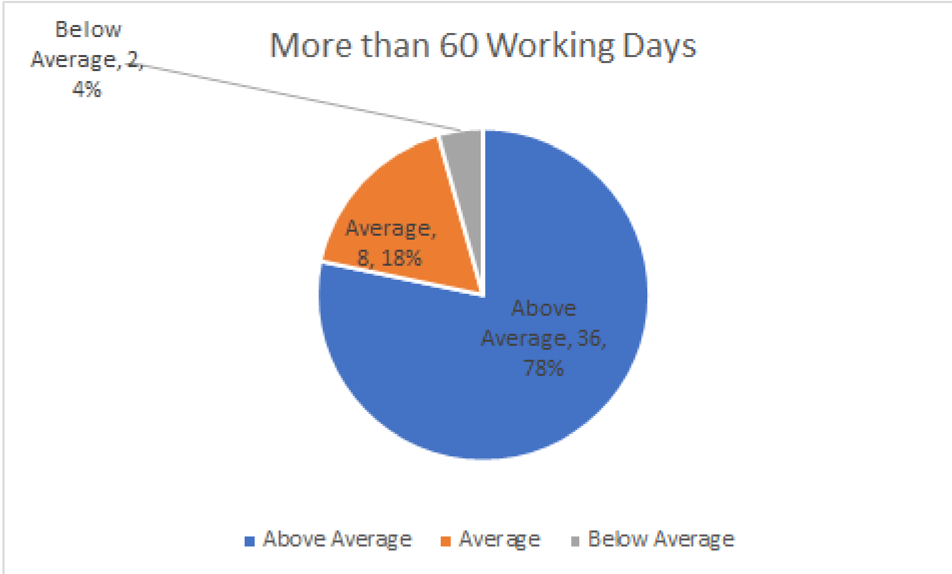
Scenario 2:

In this case, the drivers who had working hours of 7 and above and who worked for more than 30 days were included. It can be clearly seen that compared to the previous pie chart, the drivers who completed above than average rides remained 64%, however, the percentage of drivers who received average number of rides fell to 19% as the proportion of below than average drivers increased to 17%. This hints towards the fact that as the working days of the drivers increase, the distribution of rides tends to become more asymmetric. This shows the possibility that there might be some drivers i.e. a large proportion in this case, which are comparatively better off than the other drivers.



Scenario 3:

In order to see if the experience of drivers had to play a part in how many rides were completed by a driver/day, the working days were increased to 60 and above. The results showed that 78% of the drivers working for more than 2 months received above than average rides per day. The drivers who complete average number of rides almost remained the same i.e. 18% whereas the proportion of drivers receiving below than average rides reduced dramatically to 2%. This tends to show a positive relationship between the experience of drivers and average number of rides that they receive.



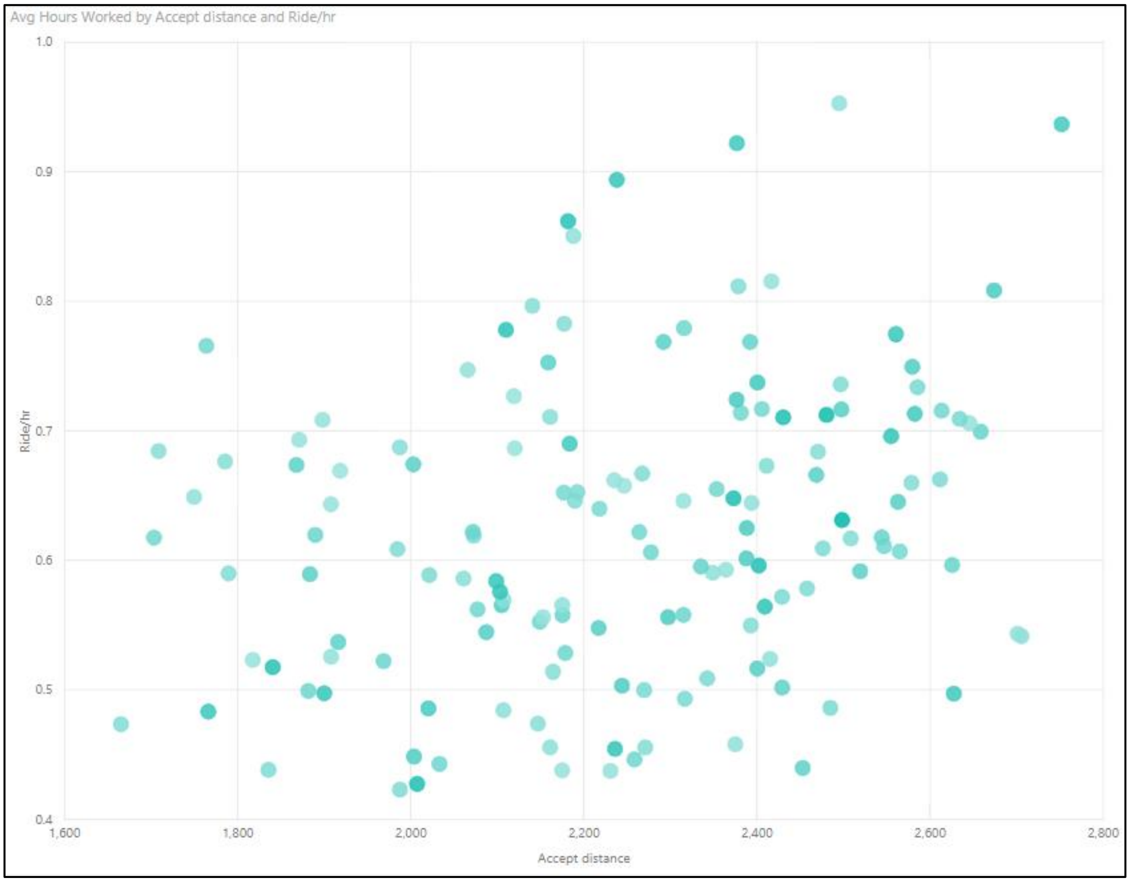
Analysis:

The visualizations show that the distribution of rides among drivers is skewed. The major proportion of the drivers don't fall in the within one standard deviation of the mean. A major chunk of the drivers is either two standard deviations above the average number of rides or are one standard deviation below the mean value. There can be several factors attributing to this pattern. The analysis suggests that as the working days of the drivers increase, and they tend to get more experienced, they tend to complete more rides on average per day than their counterparts who work for a shorter period.

In order to see, if the same drivers consistently performed better and received greater number of rides, the individual driver ID's were inspected. This showed that the drivers who performed below than average were usually the ones who worked for a very short duration. This confirmed the possibility that those with less experience tend to receive fewer rides on average per day. However, there were still drivers who worked for the same duration as those with above than average rides. In order to see, if there are factors other than experience contributing to this, we compared the ratings of above than average drivers as well as below than average drivers. It was very interesting to note that the former had an average rating of 3.6 whereas the latter had an average rating of 3.7. Although the difference is minor, this is further discussed later in the report.

- **Driver-wise Analysis**

  - **Relationship of Average Hours Worked, Accept Distance and Average Number of Rides per Hour**



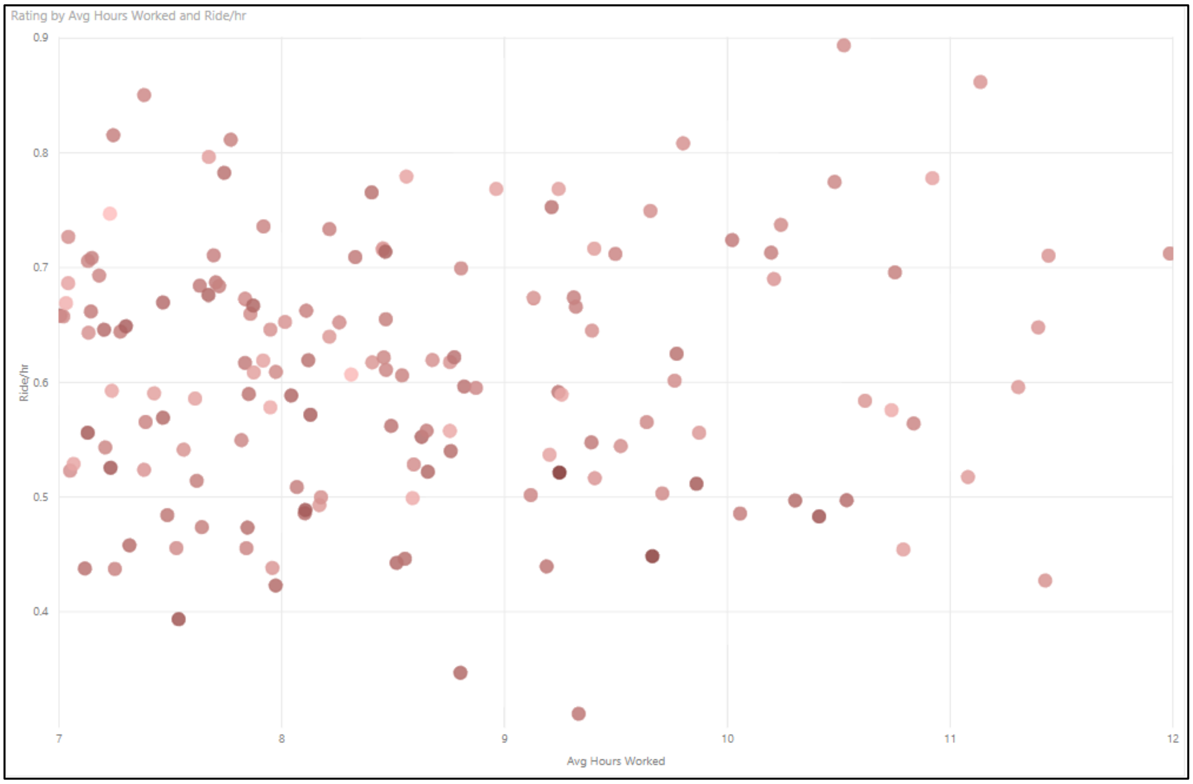
Accept Distance vs. Rides/Hour vs. Average Working Hours

Interpretation

Each point represents a driver with Accept Distance on x-axis and average number of rides allocated on y axis. The color of each point represents average number of hours per day worked by that driver, the darker the more.

It can be inferred from the above graph that drivers who are allocated more rides per hour are mostly allocated the pickup locations that are far off from their standing location. Also, that since darker points are mostly clustered at a large accept distance, this means that drivers that work for more hours on average are also allocated rides that are far off.

- **Relationship between Average Rating, Average Number of Rides per Hour and Average Hours Worked**



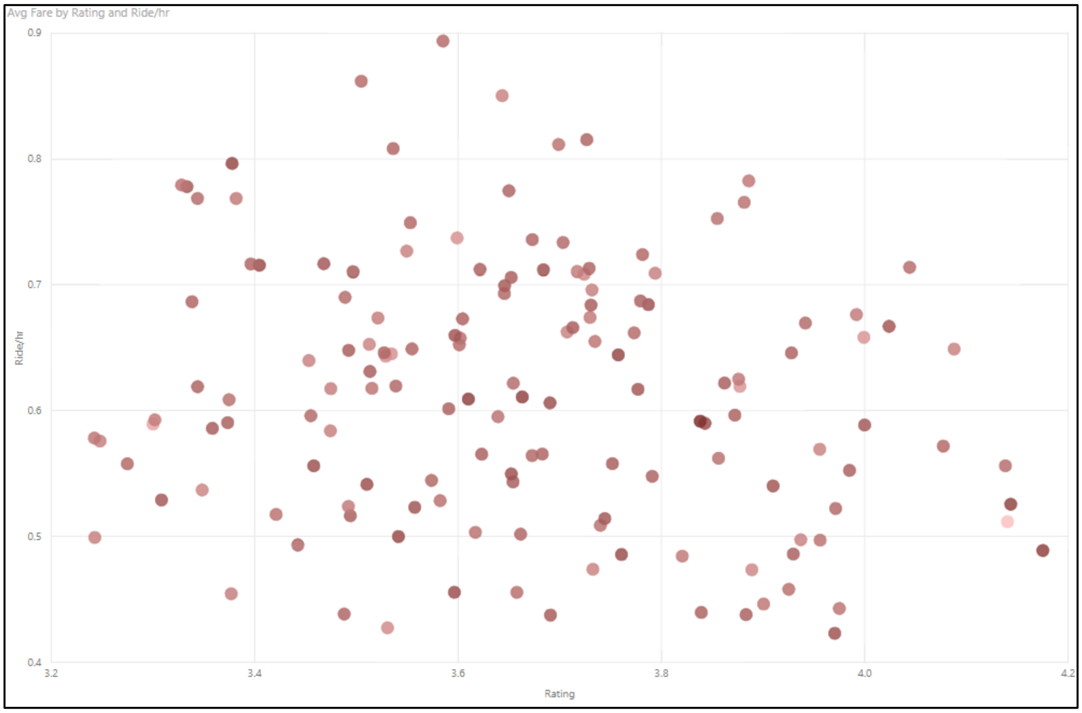
Average Rating vs. Average Number of Rides per Hour vs. Average Hours Worked

Interpretation

In this case the color of the points represents the average rating of drivers, the darker the higher. It can be concluded that drivers who who work for more hours on average have been allocated a slightly a greater number of rides per hour.

However, it has been observed that the darker points are clustered towards the lower average hours worked and slightly towards lower average rides per hour as well. This means that drivers who work for a smaller number of working hours on average, have a lower average number of rides allocated but their average rating is high. The average rating of drivers who get more than 0.80 rides per hour have an average rating of 3.82 while drivers who get less than 0.50 rides per hour have an average rating of 3.60.

- **Relationship between Average Rating, Average Fare earned and Average Number of Rides per Hour**

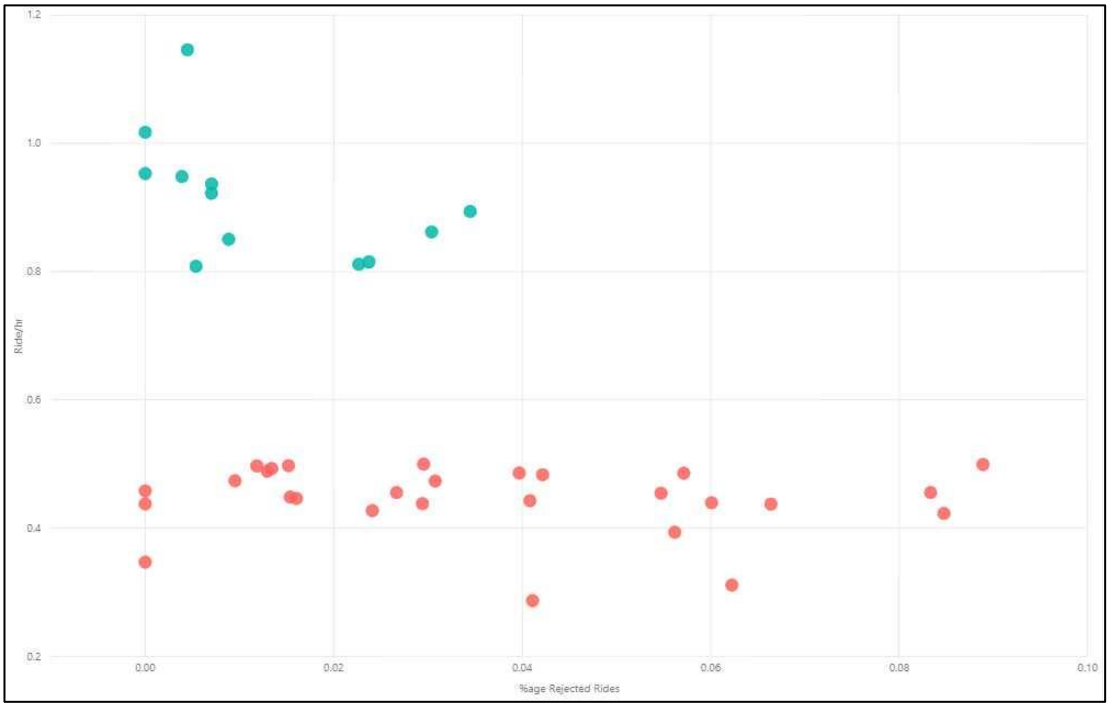


Average Rating vs. Average Fare earned vs. Average Number of Rides per Hour

Interpretation

The color of the points indicates the fare, the darker, the higher the average fare earned by the driver. This visualization affirms the interpretation of the previous graph as well that where average rating is higher, the driver has completed lesser number of rides on average. Moreover, the drivers who work for more hours on average have a rating between 3.5 to 4.5 roughly. Also, the drivers that no specific pattern can be associated with average fare earned based on this graph since points of every shade are scattered. The average fare is very similar with drivers with high and low number of rides per hour. Drivers with more rides than average earn PKR 138.79 in one ride on average while drivers who get less rides earn PKR 136.92 on average.

- **Relationship between Average Rating, Rejection Rate and Average Number of Rides per Hour**



Average Rating vs. Average Rejection Rate vs. Type of Driver (with more or less rides)

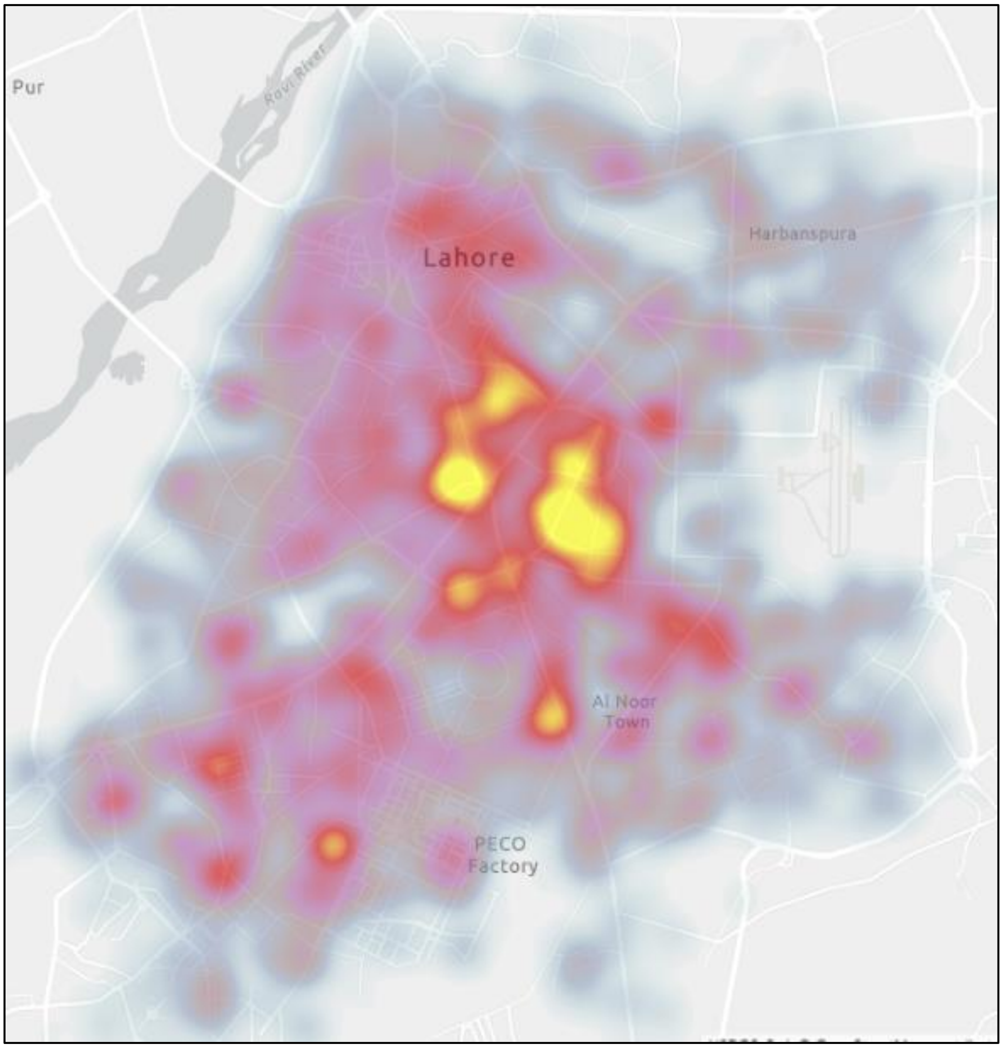
Interpretation

The x axis and y axis represent the average rejection rate and average number of rides respectively. The color however represents "Drivers with more Rides" in Blue color and "Drivers with less Rides" in Red color as compared to the average rides per hour. Drivers that have completed less rides per hour have higher rejection rates than the others.

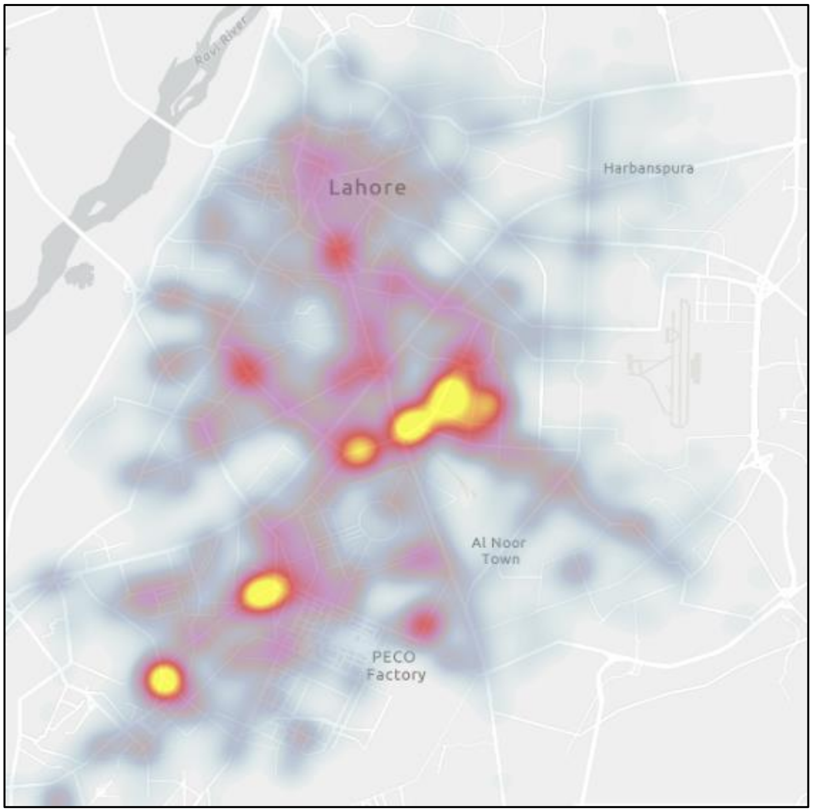
The rejection rate for drivers with less rides is 1.42% which is considerably lower than the overall average of 2.42%. The drivers who get less than 0.50 rides per hour reject 3.78% of the rides which is a significant difference from the drivers with higher allocation rate.

- **Comparison of Pickup Location with Driver Location**

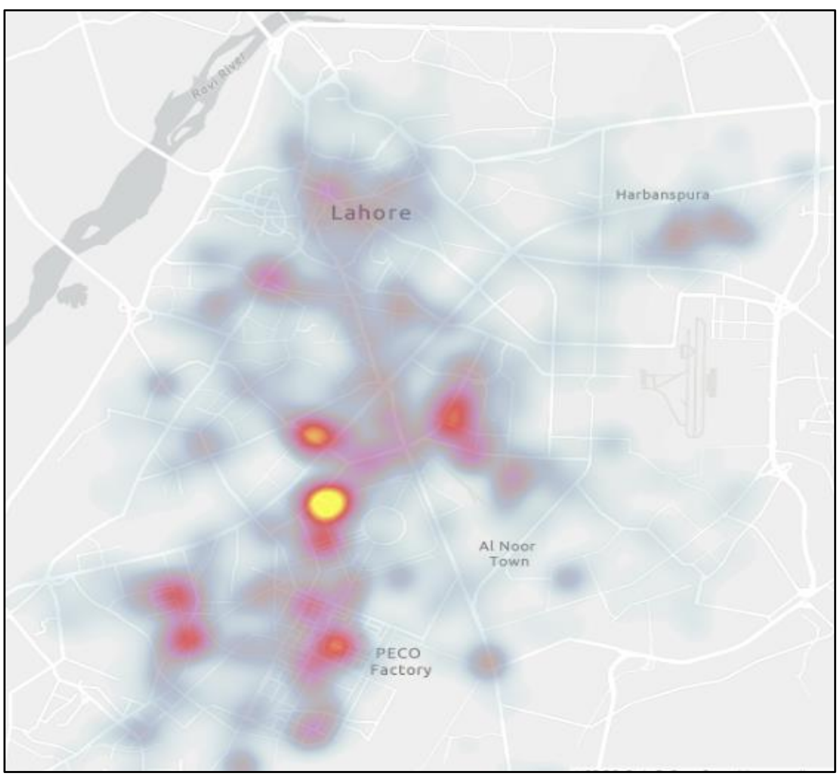
The following heat maps indicate the pickup/customer location for rides that were requested and the location of drivers belonging to each category:



A. Pick-up Location Map



B. Location of "Drivers with More Rides" Map



C. Location of "Drivers with Less Rides" Map

Comparison of A and B

A comparison of A and B map reveals that the drivers who get more rides are usually clustered around the locations where customer demand is high as the two maps show concentration " in yellow" at almost the same areas.

Comparison of A and C

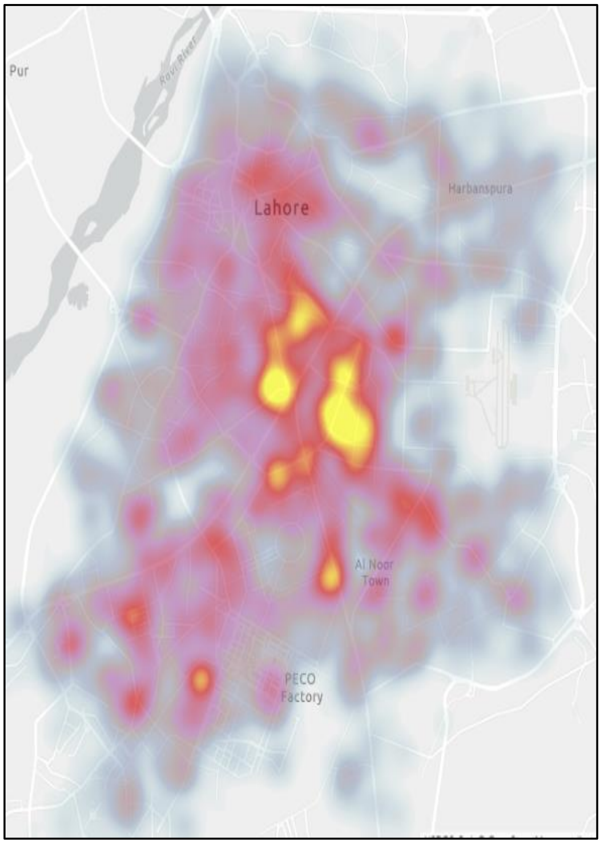
A comparison of A and C map reveals that the drivers who get less rides are usually clustered around the locations where customer demand is not much as the two maps show concentration "in yellow" at different areas.

These comparisons indicate that the reason that distinguishes the two categories of the drivers is the fact that the one with higher number of rides choose to stand at the location where simply the demand is high.

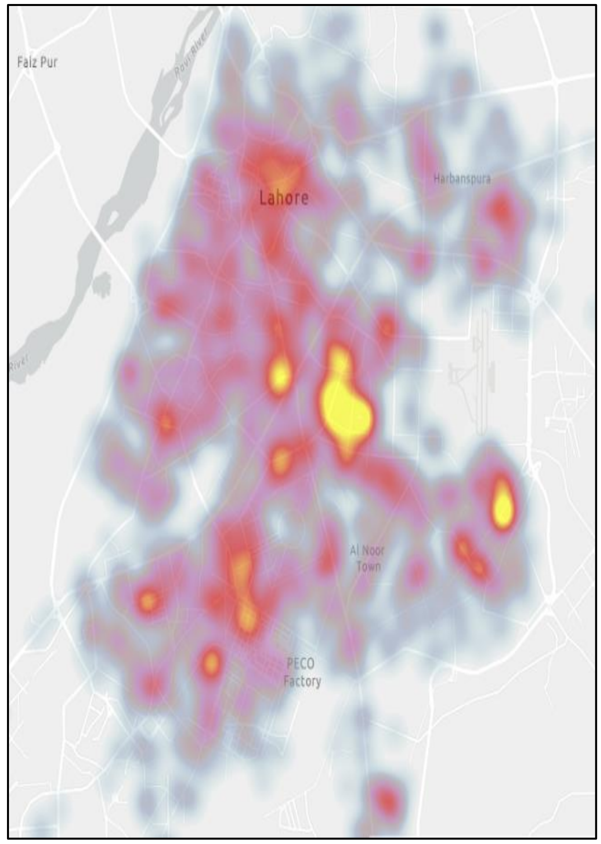
Comparison of Allocated Rides Location and Rejected Rides Location

- **Overall Comparison**

For over the span of five months, location of all the rides initiated and location of all the rides rejected was plotted on the map to compare:



Location of Total Rides Initiated



Location of Total Rides Rejected

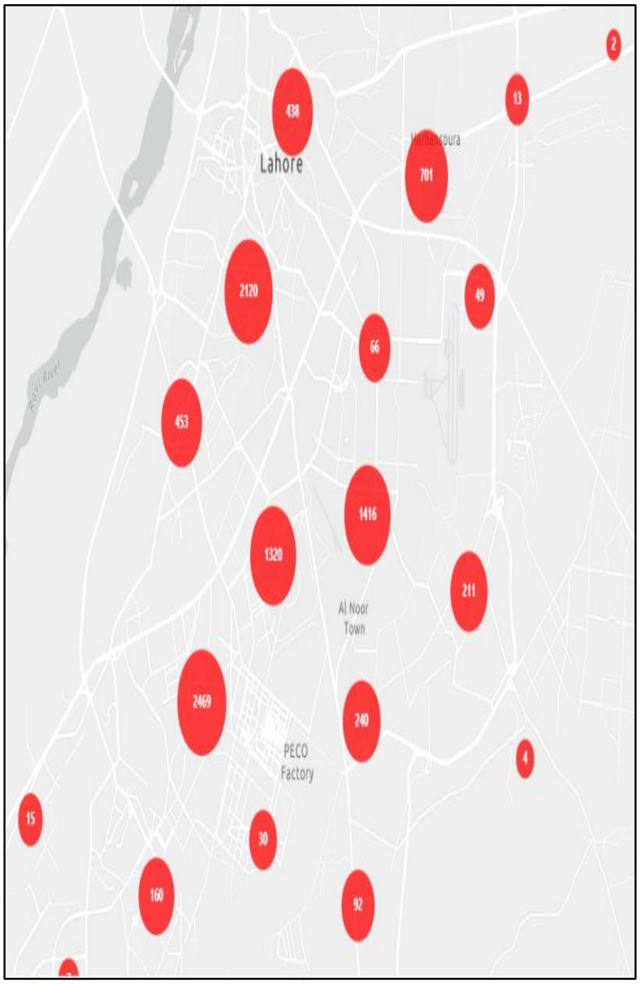
Analysis

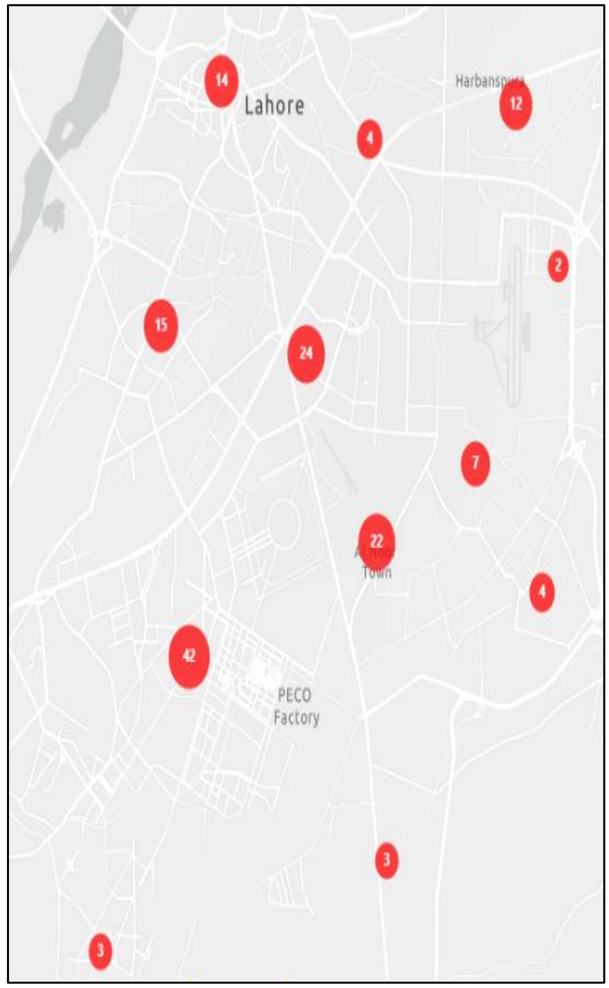
It can be seen that the two map plots do not differ significantly which can be interpreted as the fact that pick-up location of customers does not significantly contribute to the reason for which drivers reject the rides. It can be inferred that rejection pattern is not linked with the placement of customer.

- **Comparison with respect to Day Time**

Now in order to analyze and compare the relationship on a deeper scale, the location of total rides initiated, and location of total rides rejected has been mapped with respect to different time slots within a day.

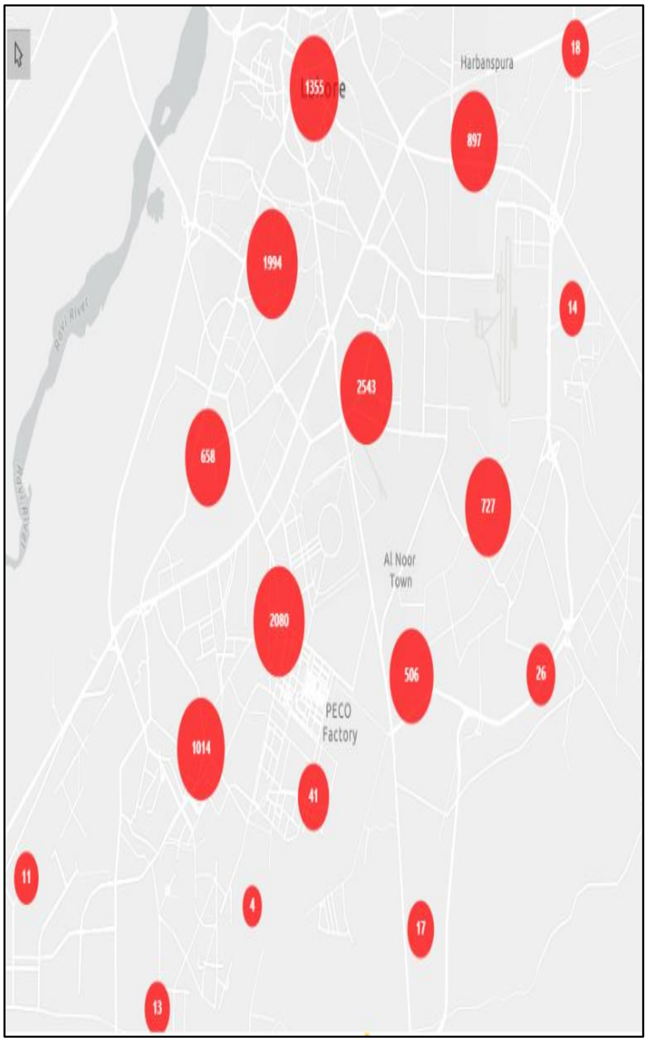
Slot # 1 (8 am to 10 am)



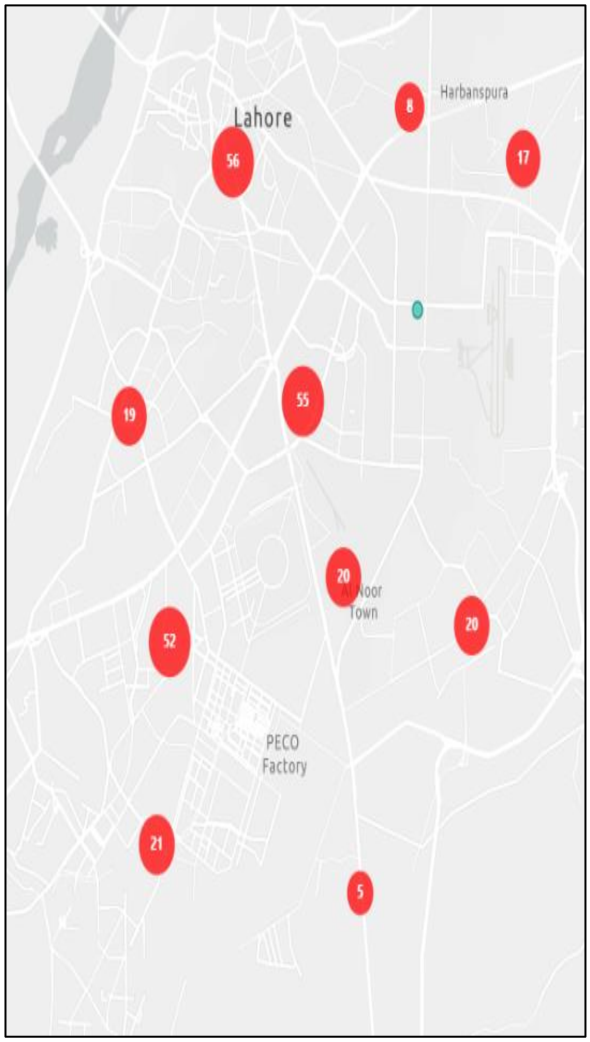


Total Rides Initiated **Total Rides Rejected**

Slot # 2 (10 am to 12 pm)



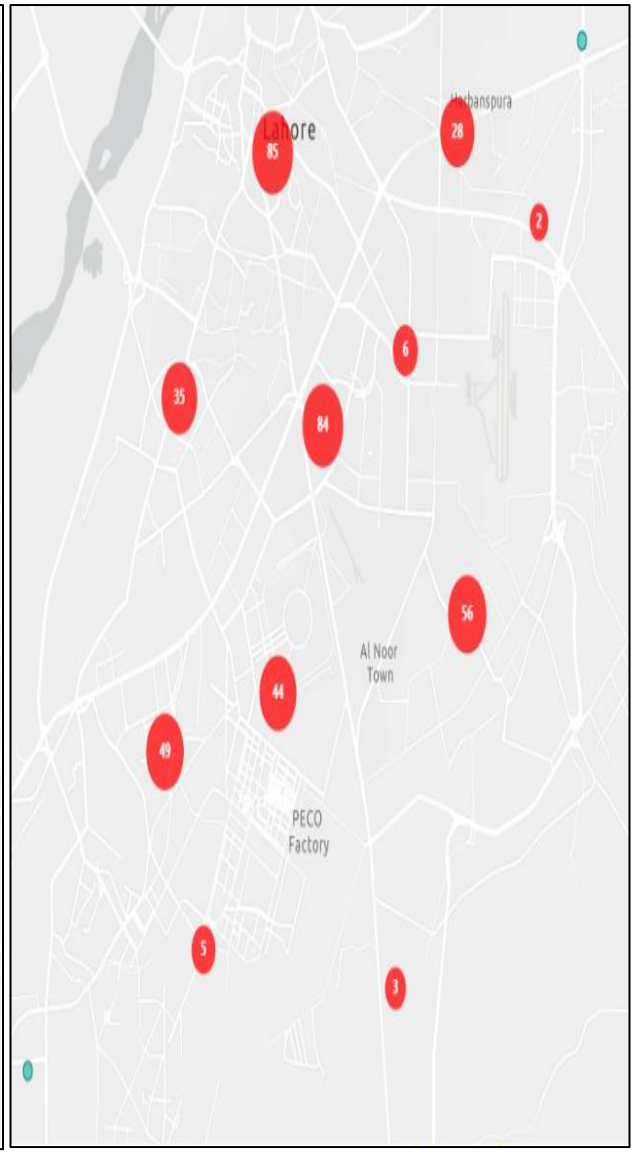
Total Rides Initiated



Total Rides Rejected

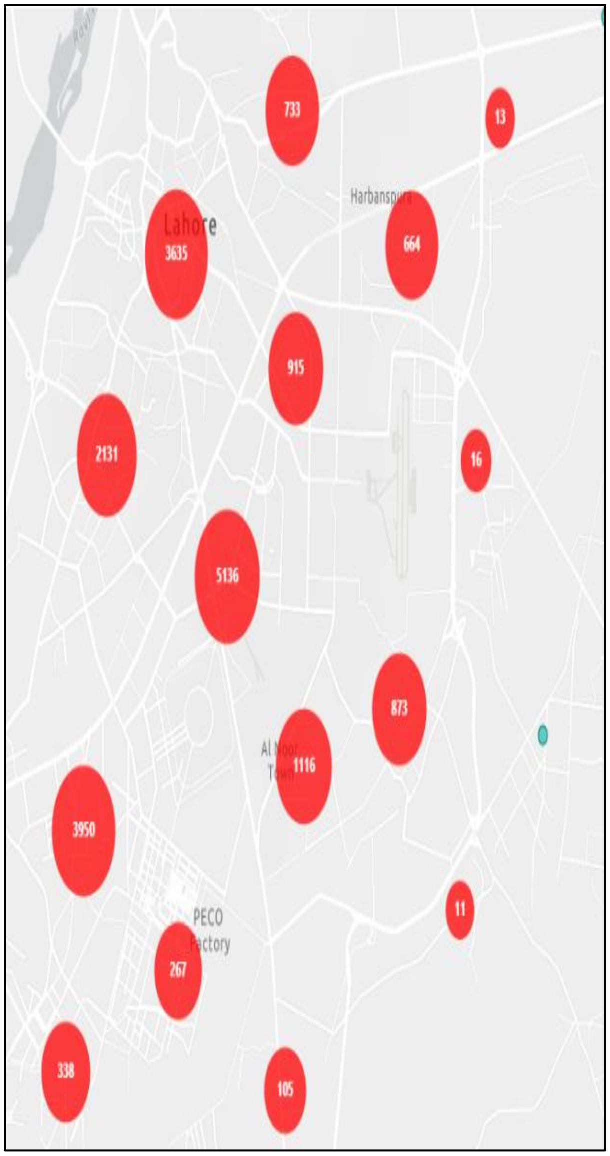
Slot # 3 (12 pm to 2 pm)

Total Rides Initiated

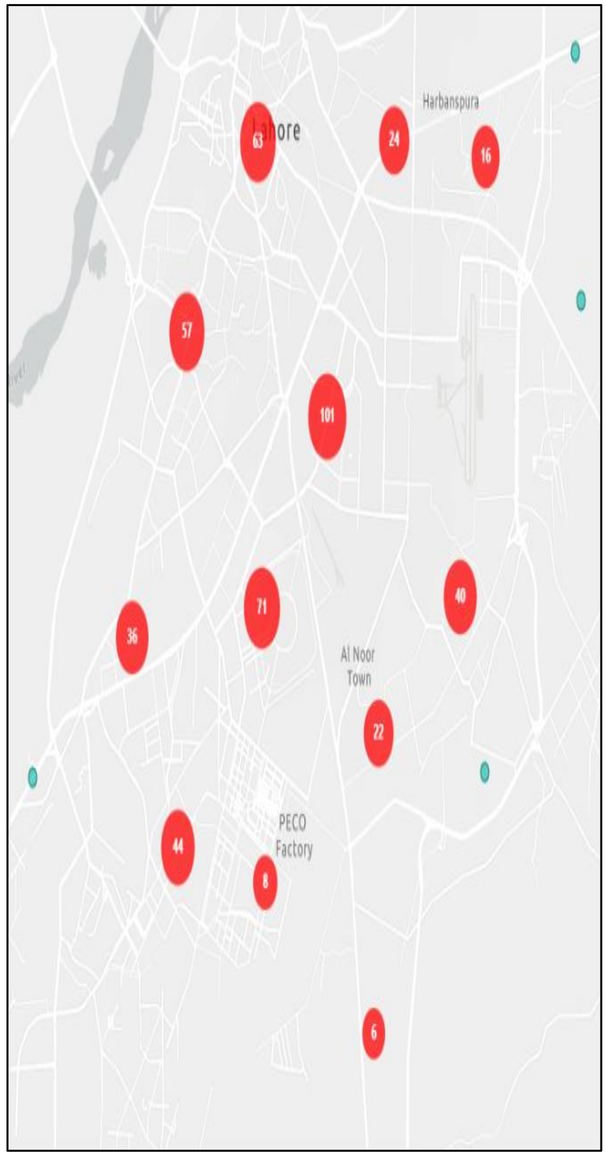


Total Rides Rejected

Slot # 4 (2 pm to 4 pm)

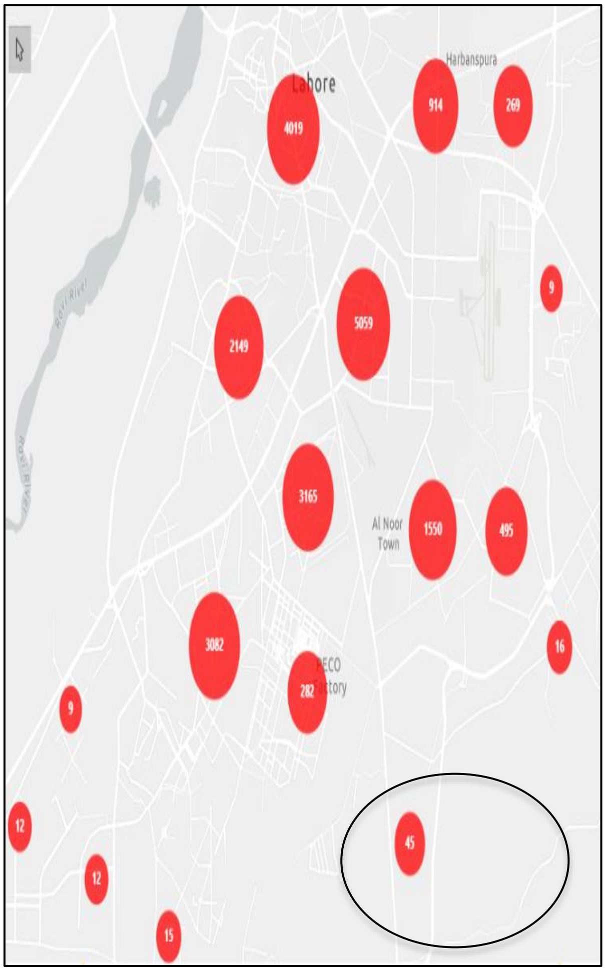


Total Rides Initiated

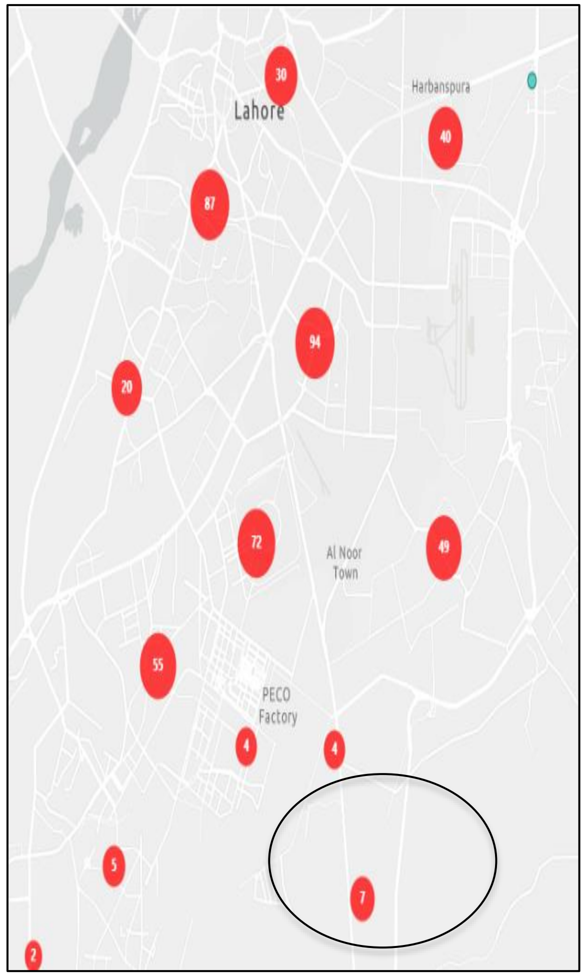


Total Rides Rejected

Slot # 5 (4 pm to 6 pm)

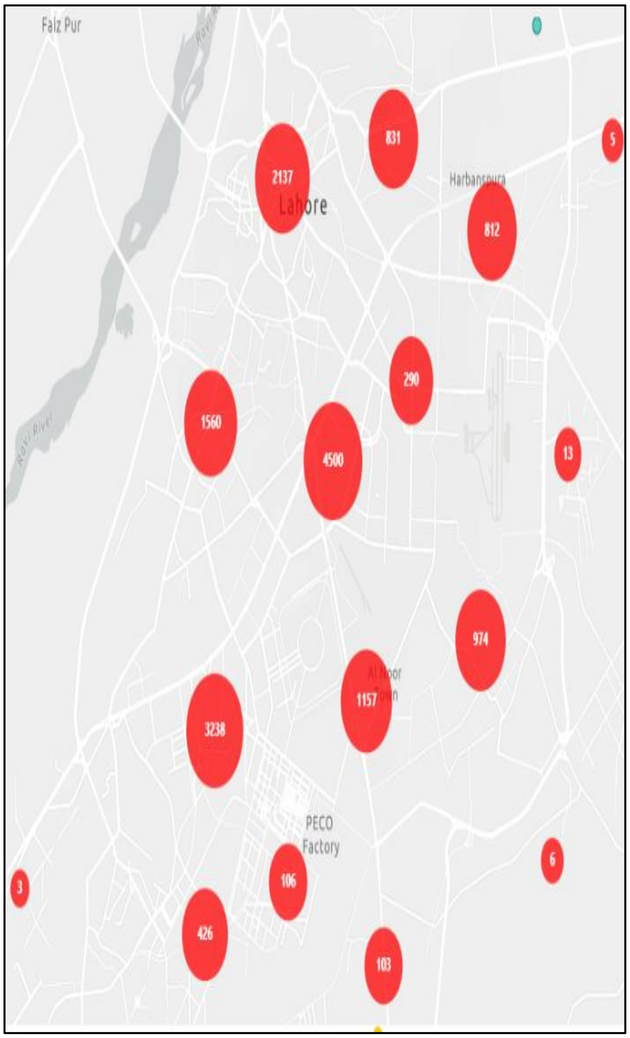


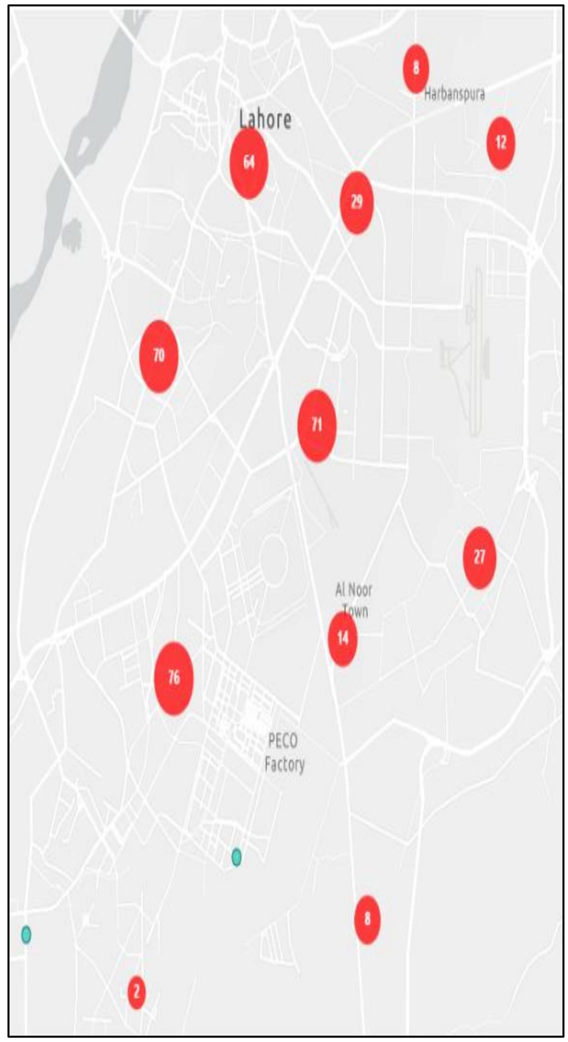
Total Rides Initiated



Total Rides Rejected

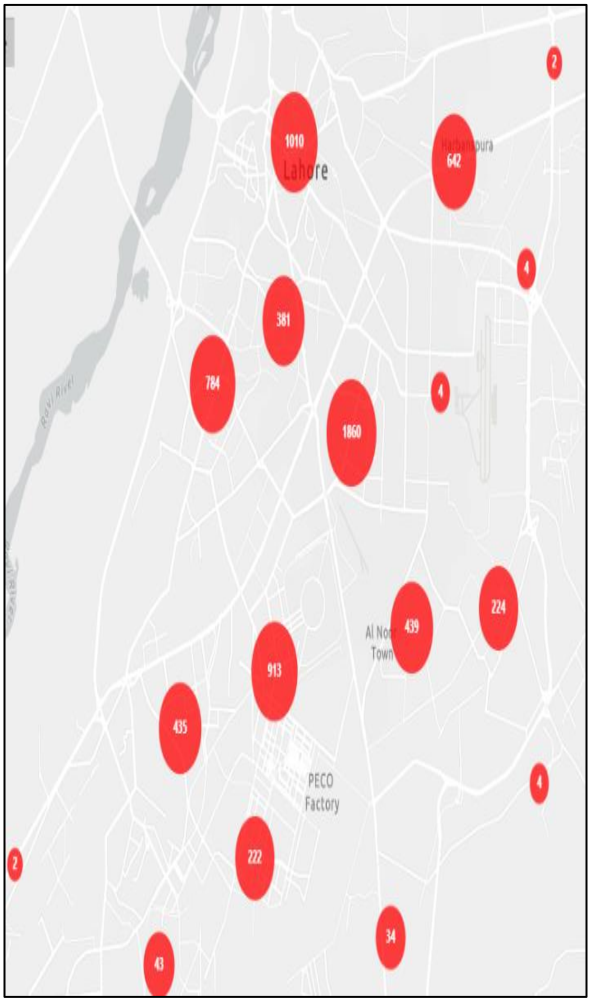
Slot # 6 (6 pm to 8 pm)





Total Rides Initiated **Total Rides Rejected**

Slot # 7 (8 pm to 10 pm)



Total Rides Initiated

Total Rides Rejected

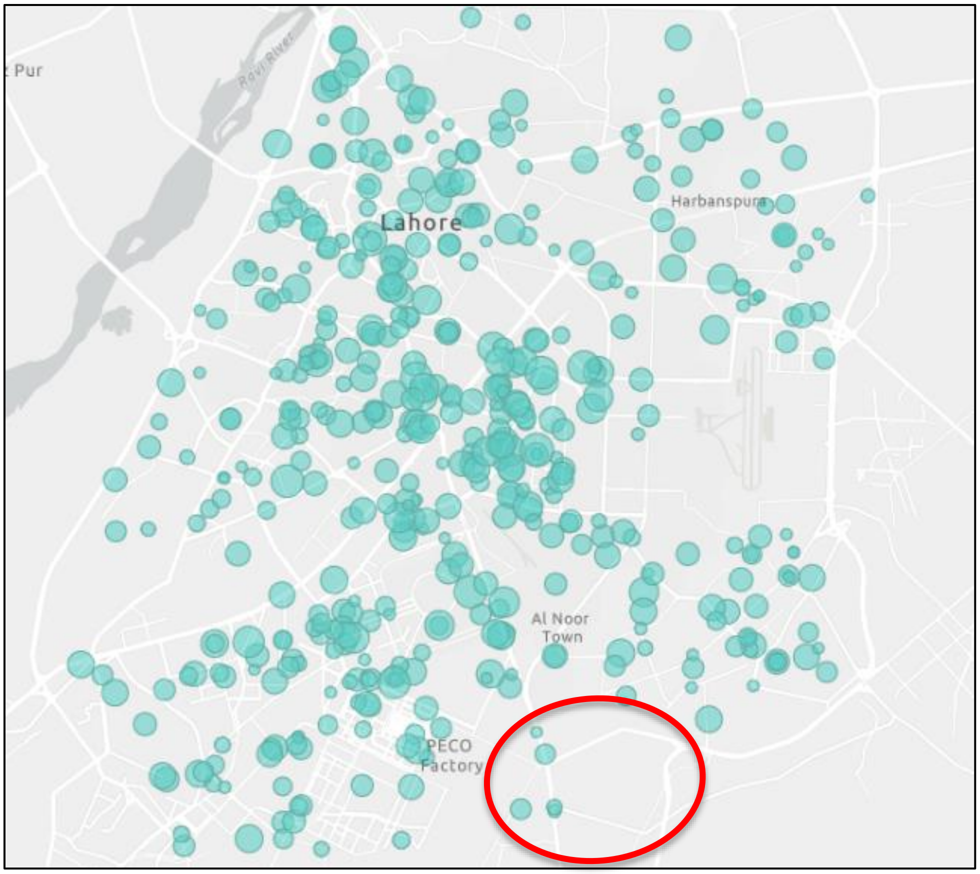
Analysis:

Slots (1:4, 6:7):

Upon zooming the time slots maps within a day except slot 5, it was noticed that the number of rides rejected relative to the total rides initiated (both indicated by the size of red circles) within a particular area was not very significant meaning that the proportion of rejected rides out of all rides requested was observed to be small throughout.

Slot 5:

During the time slot 4 pm to 6 pm, an anomaly was observed in the area towards the bottom right corner of Ferozepur road. On zooming, it was found that the 7 out of 45 rides were rejected in that area giving a high rejection rate of almost 15% as opposed to an overall rejection rate of 2.5% in this time slot. So, in order to inquire further into this time slot, rejected rides were plotted on a map with respect to driver's distance from the pickup location:



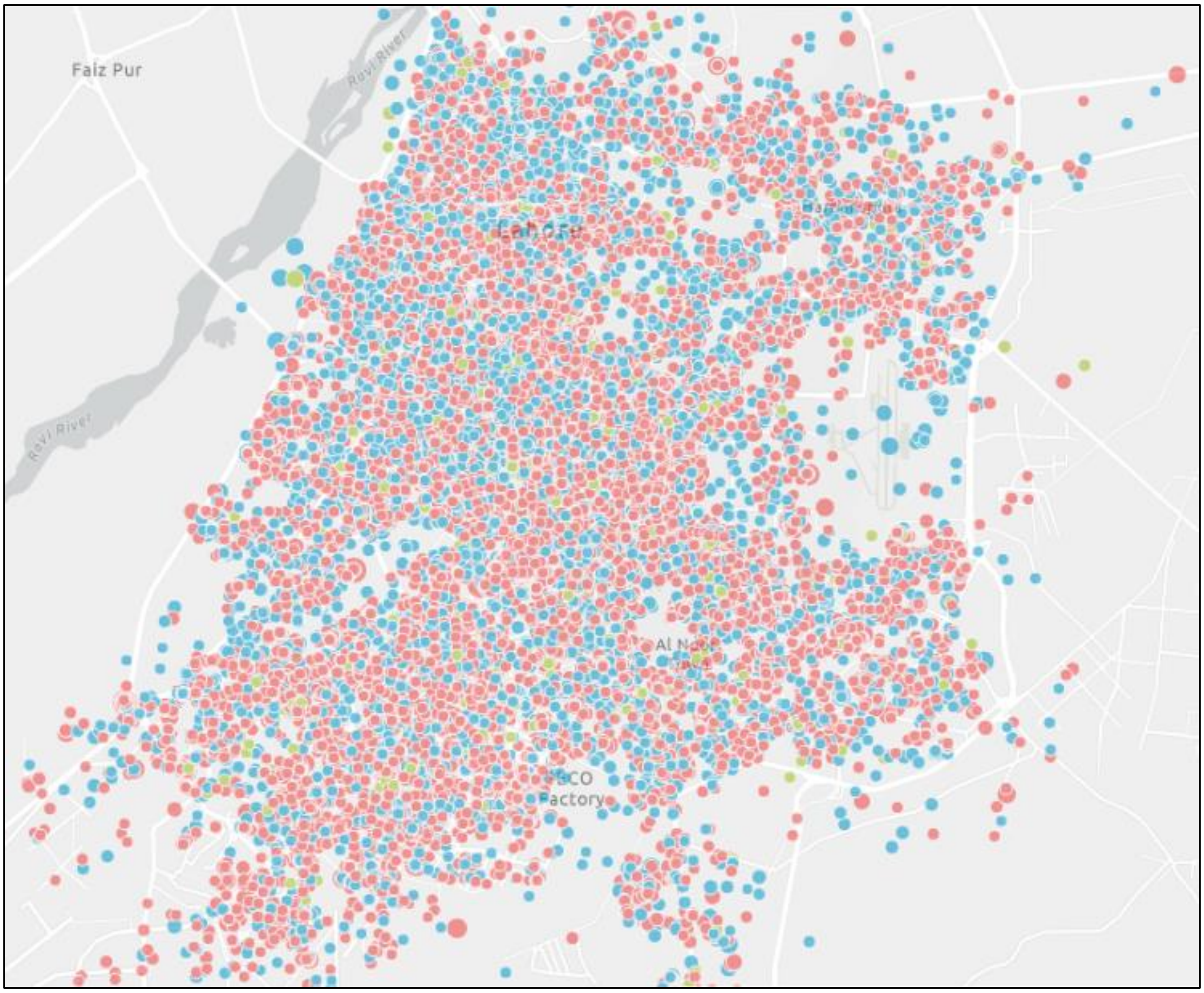
Rejected Rides with respect to Accept Distance in Slot 4 pm to 6 pm

As the size of the circles represent the accept distance, it can be said that in the same area (bottom right Ferozepur road), the size of the circles ranges from big to small indicating that the accept distance of the rides does not have a specific relationship with the likelihood of the rides being cancelled. This inference is contradictory to the assumption that drivers tend to cancel rides, the pickup location for which are far off.

Also, it is important to remember that the total number of rejected rides in this area was merely 7 so not a lot can be asserted and generalized based on this small number.

Relationship of Ride Status with Accept Distance

Following the above analysis, an overall visualization was created to inquire into the relationship of ride status with the accept distance:



Ride Accept Distance vs. Ride Status

Interpretation:

- Each circle represents a ride.
- The size of the circle represents the "accept distance". The bigger the circle, the more the accept distance.
- The color of the circle represents the ride status:
    - ➢ Red for Completed

- Green for Rejected by Driver
- Blue for Cancelled by Customer

Analysis:

Upon careful review of the visualization it was inferred that on average the more circles in blue color were big in size as compared to those in green which meant that the for the rides in which the pickup location of the customers was far off from the that of drivers, cancellation by customers was more than the rejection by drivers. It can also be verified by table below that states that the average accept distance of the rides rejected is even lower than the rides completed and is highest for the ones cancelled by the customers.

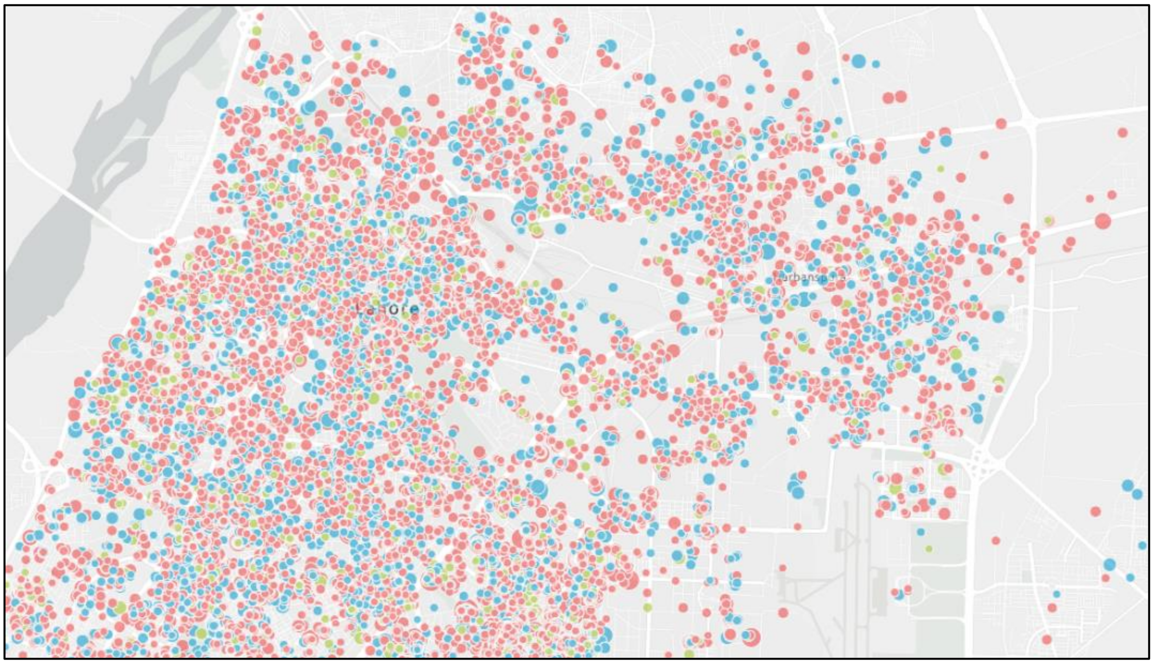
Average accept distance by status	
Cancelled by customer	3,049
Forcefully Completed	2,866
Completed	2,314
Rejected by driver	2,246

This can be due to a natural fact that customers generally do not want to wait for the drivers to pick them up for very long or maybe the drivers get in contact with customers and ask them to cancel the rides instead of rejecting them on their own.

Driver Rating

This research was also focused to analyze the reasons for which a customer might rate the driver poorly so that a recommendation can be given to improve the service experience.

- **Relationship with Ride Time**



Driver Rating vs. Ride Time

Interpretation:

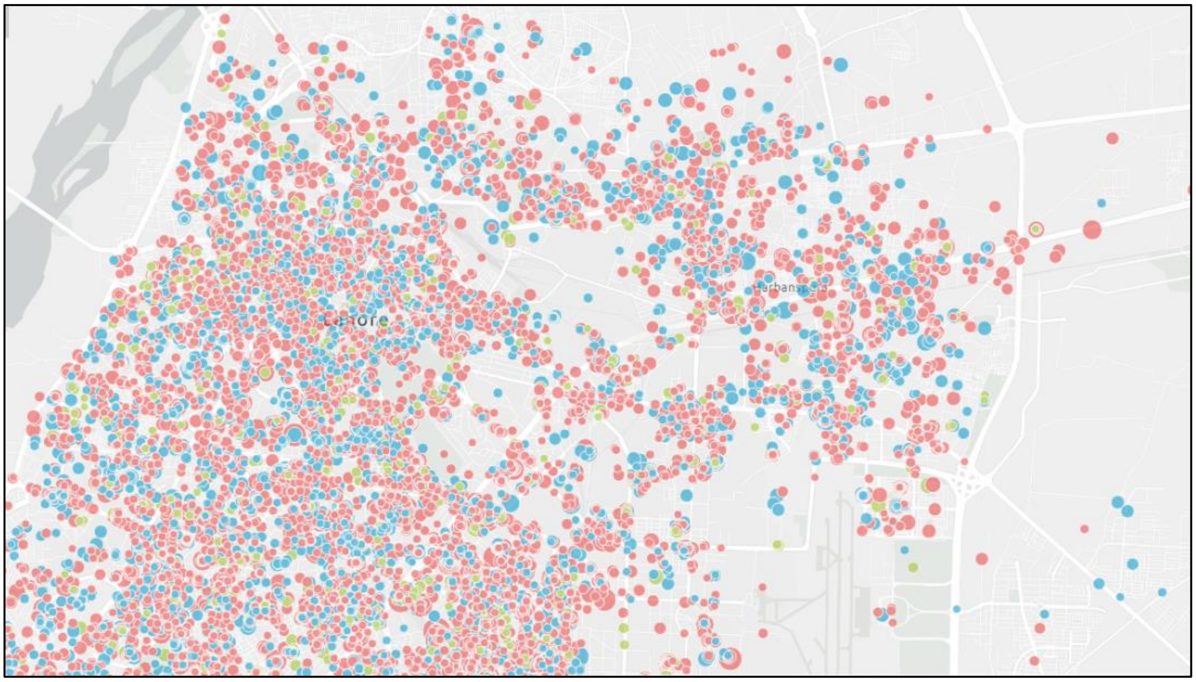
- Each circle represents a ride.
- The size of the circle represents the “ride time”. The bigger the circle, the more ride time.
- The color of the circle represents the driver rating for each ride:
  - Red for 5
  - Green for 4
  - Blue for 0

The map has been zoomed in for just the area where drivers’ ratings were relatively low; more blue circles.

Analysis:

It was found that the drivers that completed the rides, for which the ride time was long, were given a rating of 0 more frequently than the ones completing the shorter timed rides; bigger blue circles as opposed to the frequency of bigger green or red circles.

- **Relationship with Trip Fare**



Driver Rating vs. Ride Fare

Interpretation:

- Each circle represents a ride.
- The size of the circle represents the "Fare". The bigger the circle, the higher the fare.
- The color of the circle represents the driver rating for each ride:
  - ➢ Red for 5
  - ➢ Green for 4
  - ➢ Blue for 0

The map has been zoomed in for just the area where drivers' ratings were relatively low; more blue circles.

Analysis:

No specific pattern was identified in this visualization as the drivers that were rated poorly appeared to have completed rides for a range of fares (size of blue circles varying among big, average and small).

Hence, overall it can be inferred that while the amount of fare charged does not get bad reviews for the drivers however if the ride time is long, the customer might feel uncomfortable enough to rate the driver poorly.

Recommendations

- **Feature to set Desired Destinations**

It might be useful to allow drivers to set a desired location that they want to drop off in or towards which they are heading. Rides can be allocated to them that have the same or a drop off location within 1 km proximity of the desired location by driver. This feature would help to maximize time and acceptance at the driver's end particularly during the times when he has completed his working hours and heading back to his house.

However, there is a drawback to this recommendation, allowing drivers to set the desired location as many times as they wish can create a supply deficit at particular routes and hence affect company's market place negatively. This could also be discouraging for other drivers since they would be allocated rides at the routes where there is a supply deficit and they would have to first cover a long distance to reach the pickup location hence a major chunk of time would be spent in driving without earning.

In order to avoid this potential issue, it might help to set the limit of specifying the desired location once or twice per day per driver. So that this option could only be availed at the start or end of driver's working hours. Also, another restriction of non-availability of the option of specifying a destination during the peak hours can also be put upon the drivers in that case.

- **Balance between Retaining and Recruitment**

While expansion and growth of Shahi Sawari is significant and can be achieved by competing with other companies to recruit maximum number of drivers however the operations of the company might suffer. The resources to manage a large human capital and respond to its needs would be expended in merely increasing the load even more.

It is advisable to invest in attractive recruitment programs and increasing the supply to eventually cater to more demand, but the company must also balance out such resources with the ones needed to train and personally incentivize the drivers.

As the switching cost of a driver is not so large, it should be the utmost priority of the company to retain the current drivers in addition to recruiting more.

- **Specific Ride Allocation to New Recruits and Slacking Drivers**

The ride allocation should be designed so to channel more rides to the drivers that are newly recruited in order to ensure them a profitable contract and to the driver that are lagging behind so that they are kept motivated to continue the contract with Shahi Sawari.

Moreover, the pool requests could also not be channeled to the new drivers for some time just to make sure that they are first well acquainted with process and terms of the contract. So, the ease in provision of service by drivers should be prioritized first

- **Specific Ride Allocation to Highly Rated Drivers**

The analysis in this research revealed that there might exist a relationship between the ride allocation and the rating of the driver. It appeared that the lower the rating, the more likely is the driver to get a ride first. So, if Shahi Sawari can recode the algorithm as such that by default while allocating ride, it should prefer drivers with highest rating over the others in the nearest proximity of the pickup location, this would act as an added bonus incentive for good drivers. This would create a double win for Shahi Sawari, overall the company would enjoy a better image due to better service by good drivers and since the drivers will be competing for a better standing and a high rating, the service would automatically improve.

Yet, there is a drawback to this as well. Selective allocation can marginalize some drivers altogether as the ones with already low ratings would be worse off. Also, it would make driving for Shahi Sawari unrealistic for some highly rated drivers too who are not seeking to work for the company as a full-time job. So maybe another requirement of being a "full-time occupant" driver in addition to being "highly rated" within the nearest distance can be set for such a discriminated yet practical allocation of rides.

- **A Suggestive App**

The Shahi Sawari application can be redesigned to be more directive. Based on the real time monitoring and data driven analytics, it can offer live suggestions or directions to drivers regarding which route to take and most importantly in which area to work depending upon the demand dynamics.

Also, at the start of the day or working shift, the application can also be programmed to show the highlights of the day such as some important events or gatherings and the general forecasted demand with respect to geographical area and time to the drivers so that they can plan ahead of time.

- **Training**

While it might be more advisable for to have a standardized training course for drivers nationwide to be implemented by all the cab companies however even in its absence, Shahi Sawari should make sure that drivers are not set out to work in the field without being briefed and trained about the requirements and expectations from their job position.

- **Bonus**

In order to encourage the drivers to take up rides the pickup locations for which are out of the way for drivers or really far off, additional fixed bonus can be awarded to the drivers for completing such rides. In this way, the usual behavior of going for "short quick trips" observed in the analysis can be curbed to some extent.

- **Surge Pricing**

Unlike the competitors, Shahi Sawai has not yet employed surge pricing mechanism. This can be one of the reasons for which drivers might not be interested to cater to a boosted demand in certain geographical areas at certain times. Therefore, maintaining a surge price for such demand will incentivize the drivers to not reject such rides and hence the supply would be ensured.

However, this might be a drawback for the company on the customer end which can be pronounced into a negative effect on drivers. Since high price could lead to low demand which would decrease the opportunity to earn for drivers. Therefore, while setting the surge price, these factors should also be kept into consideration. A balanced fare should be charged which is neither too high for customers but is still encouraging enough for drivers.

- **Focus on particular area**

Since an anomaly was identified near the Ferozepur road in terms of high rejection rate, further investigation and focus should be laid to find out and solve the reason for which drivers tend to reject the rides in this area.

Limitations

- **Missing Drop-off Location for Rides Cancelled or Rejected**

If a ride is cancelled by customer or rejected by driver, the information about the drop-off location in this case is missing from the data, hence this research does not inquire into the rejections due to longer rides which otherwise can be a major factor.

- **Subjective Factors in Drivers' Ratings**

The rating that a driver gets for each ride might be dependent on a lot of subjective factors, such as the interaction between driver and the consumer hence, such factors are not gauged in this analysis.

- **Less points of Rejected Rides for Analysis**

When the analysis was conducted with respect to different time frames within a day, very few data points for rejected rides were found. Hence generalizing any inference based on those might create an issue of external validity.

- **Arbitration in Supervised Techniques**

In data preparation for supervised techniques, the numeric variables were divided into categories by setting arbitrary cut offs of values. Hence due to these subjective choices, the results may vary if cut off values are further changed.

Also, the choice of variables was kept limited. Only 4 variables were used, more can be added for more insightful results.

Works Cited

Kuo, Marc. "Taxi Dispatch Algorithms: Why Route Optimization Reigns." *Routific*, April 26, 2016. https://blog.routific.com.

Lee, Min Kyung, Daniel Kusbit, Evan Metsky, and Laura Dabbish. "Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers." *Human-Computer Interaction Institute,*.

Miao Fie, Shan Lin, Sirajum Munir, John A. Stankovic, Hua Huang, Desheng Zhang, Tian He, and George J. Pappas. "Taxi Dispatch with Real-Time Sensing Data in Metropolitan Areas — a Receding Horizon Control Approach." 2015, 100-09. doi:10.1145/2735960.2735961.

Rosenbalt, Alex. "The Truth About How Uber's App Manages Drivers." Review. *Harvard Business Review*, April 6, 2016.

Salanova, Joseph Maria, et al. "A Review of the Modeling of Taxi Services ." *Www.sciencedirect.com*, 2011, core.ac.uk/download/pdf/82647278.pdf.

Simonite, Tom. "When Your Boss Is an Uber Algorithm." Review. *MIT Technology Review*, December 1, 2015. https://www.technologyreview.com/s/543946/when-your-boss-is-an-Uber-algorithm/.

Voytek. "Optimizing a Dispatch System Using an AI Simulation Framework." Www.Uber.com. August 11, 2014. https://www.Uber.com/newsroom/semi-automated-science-using-an-ai-simulation-framework.