Analysis of Cancellations at a Cab Portal Company

KYUNGJIN CHO
WISCONSIN SCHOOL OF BUSINESS
October, Fall 2018

Note: While your report should follow this rough format, you may be creative within the sections and present your analyses as you see best fit. Your aim is to build a stand-alone document that you can distribute as a writing sample as you apply for different programs and jobs. Do not write sentences like "As in Question 1,…", but rather respond using a self-contained description. We have placed a mapping of all the questions in Part III.

Executive Summary

In the given sample of data, IndoCabs drivers have a cancellation rate of 7.68 percent. Thoroughly, point to point travel had a highest cancellation rate followed by hourly rental. It is shown in the data sets that long distance travel had no cancellation. The online booking and mobile booking cancellation were not significantly distinct by the fact that cancellation rate of online booking was 1.91 percent higher than the rate of mobile bookings. Yet, it is more effective to further develop an online booking site because more than 6 times of customers use online bookings although their cancellation rate is similar. It can be inferred that there is a weak relationship between weekdays and percentage of bookings cancelled although cancellation occur remarkably on Sunday. It is inferred that the cancellation mostly occurred due to the fact that it is a weekend. Furthermore, there is also a weak correlation between an hour of booking and cancellation. We can infer that cancellation is not affected by the hour of booking. Indeed, the number of bookings by hour corresponded to the number of cancellations by hour. It is because number of bookings and number of cancellations is a cause and effect relationship; more booking leads to more cancellations. There is also a trend between the day of bins and booking window. Booking cancelled occurred noticeably in 1-day bin and especially, within 6 hours. Bins of one to seven showed a declining trend; however, more than seven-day bins booking cancelled were remarked as increasing.

Introduction

IndoCabs is an Indian cab company that is managed by the independent drivers and costumers. However, IndoCabs is facing a problem due to the drivers cancelling the ride. IndoCabs executives believe that it is due to a poor algorithm and try to find a trend between the data such as booking window, cancellation rate, customers booking trend, and weekday usage of cabs.

Analysis

A Look at Trip Durations

The average trip duration of customers was 4.06 minute, while the median of it

was 1.4 minute. It is important to include both average and median in this chart because the data, average trip duration, is significantly larger than the median, which means that the graph is skewed to the right. Since there is a large gap between average and median, it is also assumed that there will be some outliers affecting this result. Yet, although average booking window is larger than the median booking window, the difference between average and median of a booking window is not that large as it is for a trip duration. With the average and median, I also included interquartile range in order to examine the middle 50 percent of data. Moreover, interquartile range do not get affected by the outliers; therefore, I used average and interquartile range to see whether there are some existing outliers. The data of trip duration has large outliers, yet, booking window does not have a huge affecting outlier compared to a trip duration.

Summary Statistics of Trip Duration and Booking Window

	Trip Duration (min)	Booking Window (day)
Average	4.06	1.87
Median	1.4	0.42
Interquartile Range	1.73	0.79

The Magnitude of the Cancellation Problem at IndoCabs

The total number of bookings in the sample was 2083. Among those, cancellations occurred was 160; therefore, in percentage, 7.68% of cancellation was occurred due to a driver of IndoCabs. Although there is no specific cancellation rate that defines if 7.68% is a high number, it is inferred from IndoCabs that they are struggling to reduce the cancellation rate. This means that IndoCabs has to reduce their cancellation by analyzing the trend of booking cabs.

According to the data, long distance travel had the longest average trip duration following by hourly rental and point to point travel, respectively. Furthermore, the point to point travel had the highest percentage of cancellation following by hourly rental and long-distance travel. The sample data indicates that there is no cancellation for long distance travel. I expected this pattern of behavior since long distance travel is something that drivers plan more precisely than any other travel types. Moreover, the

highest percentage of cancellation was point to point travel which I was expecting it too, since there might be some unavoidable circumstances (ex. Distance of customers are far than expected, car got broken down, there is other customer who is closer to driver's location) where drivers have to cancel customers' request.

Cancellations by Travel Type

	Number of Bookings	Number of Cancellations	Percent Cancelled (%)
Long distance Travel	75	0	0%
Point to Point Travel	1623	142	8.75%
Hourly Rental Car	385	18	4.68%

795 online bookings were made in the given sample. Among those, 102 bookings got cancelled. Yet, only 119 bookings were made through a mobile site, and 13 were cancelled. By looking at the data given, there was no significant difference of percentage cancelled for online bookings and mobile site bookings because percentage of online bookings cancelled was 12.83% and percentage of mobile bookings that were cancelled was 10.92%. However, I would recommend IndoCabs to expand online site more because the booking number is significantly higher (693 bookings more than mobile booking) than it does on mobile site while the percentage of booking cancelled is similar to mobile booking (1.91% more than mobile booking). I would not recommend IndoCabs to shut down the mobile booking site, since it can be used for customers faced with unexpected use of cab.

When examining the percentage of cancelled trip by weekday, Sunday had a highest percentage of cancelled bookings. Sunday's booking cancelled was significantly high with the percentage of 10.95, given that the second highest booking cancelled was on Friday, with 8.33%. Moreover, Saturday had a lowest booking cancellation with the percentage of 5.08, which indicates that there is a wide range between Monday to Sunday's booking cancellation. Yet, those days that lie in middle had a similar range. I did not expect that Sunday will have the most car cancellation because I thought that drivers are most likely to cancel the cab on Friday since Friday is a last day of week and

drivers will face lot of impediments such as car traffic or accidents. However, Sunday is a reasonable day too because people usually do not prefer working on Sundays.

Percentage of Cancelled Trips by Weekday
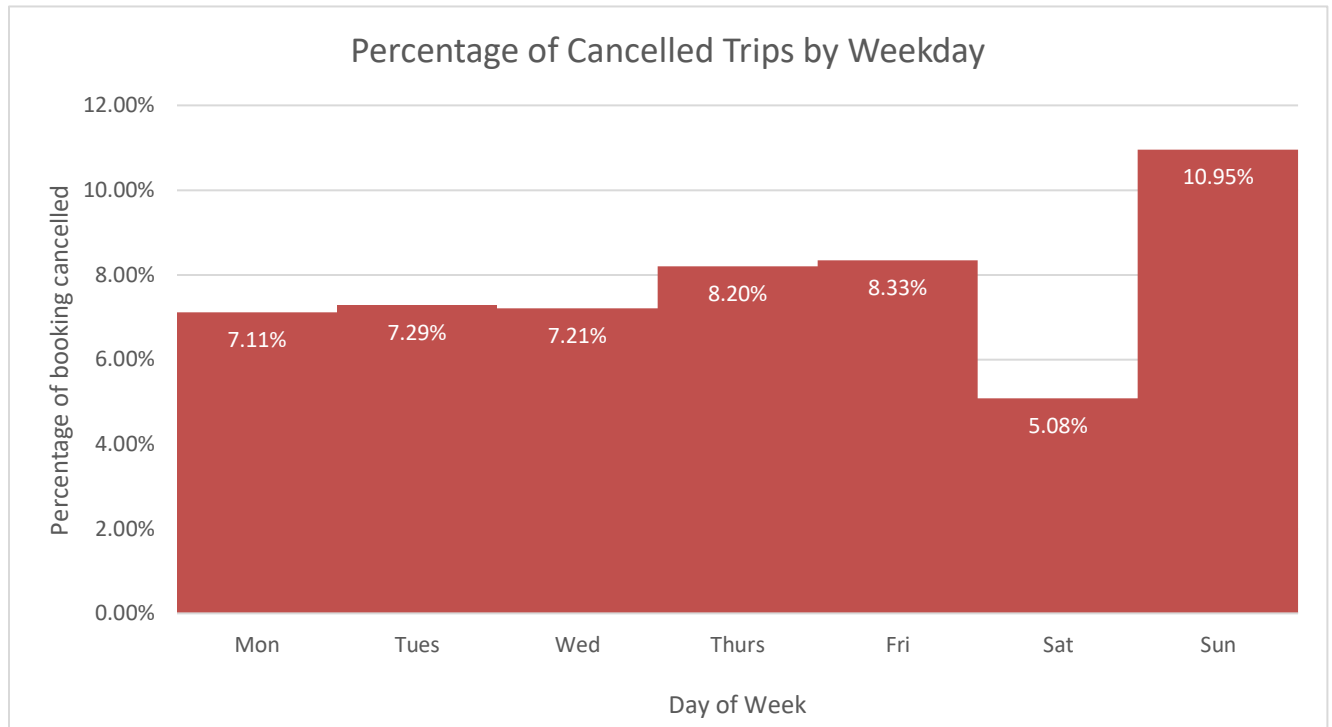
The Relationship between Booking Windows, Cancellations, and Trip Timing

According to the graph "number of booking per hour and sum of cancellations", there is almost no correlation between the two data sets. The correlation coefficient (r) is 0.34, which indicates weak linear positive relationship between hour of booking and cancellations.

Sum of Bookings and Cancellations by Hour
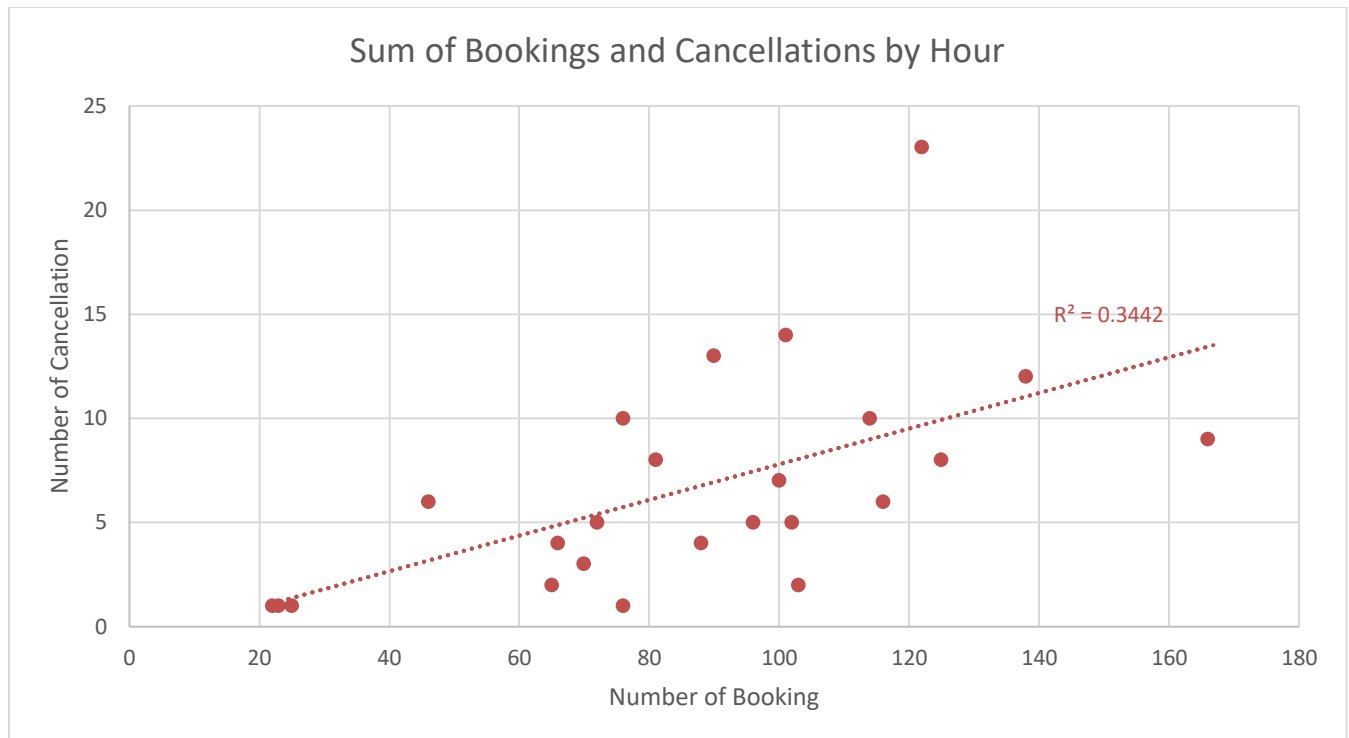
Among the three line charts below, the first graph – number of bookings by hour – indicates that there is a significant bookings at 9 A.M. The second graph – number of cancellations by hour – displays that cancellation happens the most at 7 P.M. Lastly, the third graph which shows percentage of booking cancelled by hour indicates that the cancellation happened mostly at 7 P.M with 19%. Including all the line charts, third one is the most informative one because it is easier to read the cancellation rate by hour as it is labeled as percentage. Furthermore, it is seen clearly that there is a strong relationship between the number of cancellation by hour and the percentage of cancellations by hour. When I analyzed two data by scatterplot, made a linear trendline, and got a correlation coefficient, they showed a strong positive linear relationship with r^2 equaling 0.77.

Number of Bookings by Hour

Number of Cancellations by Hour
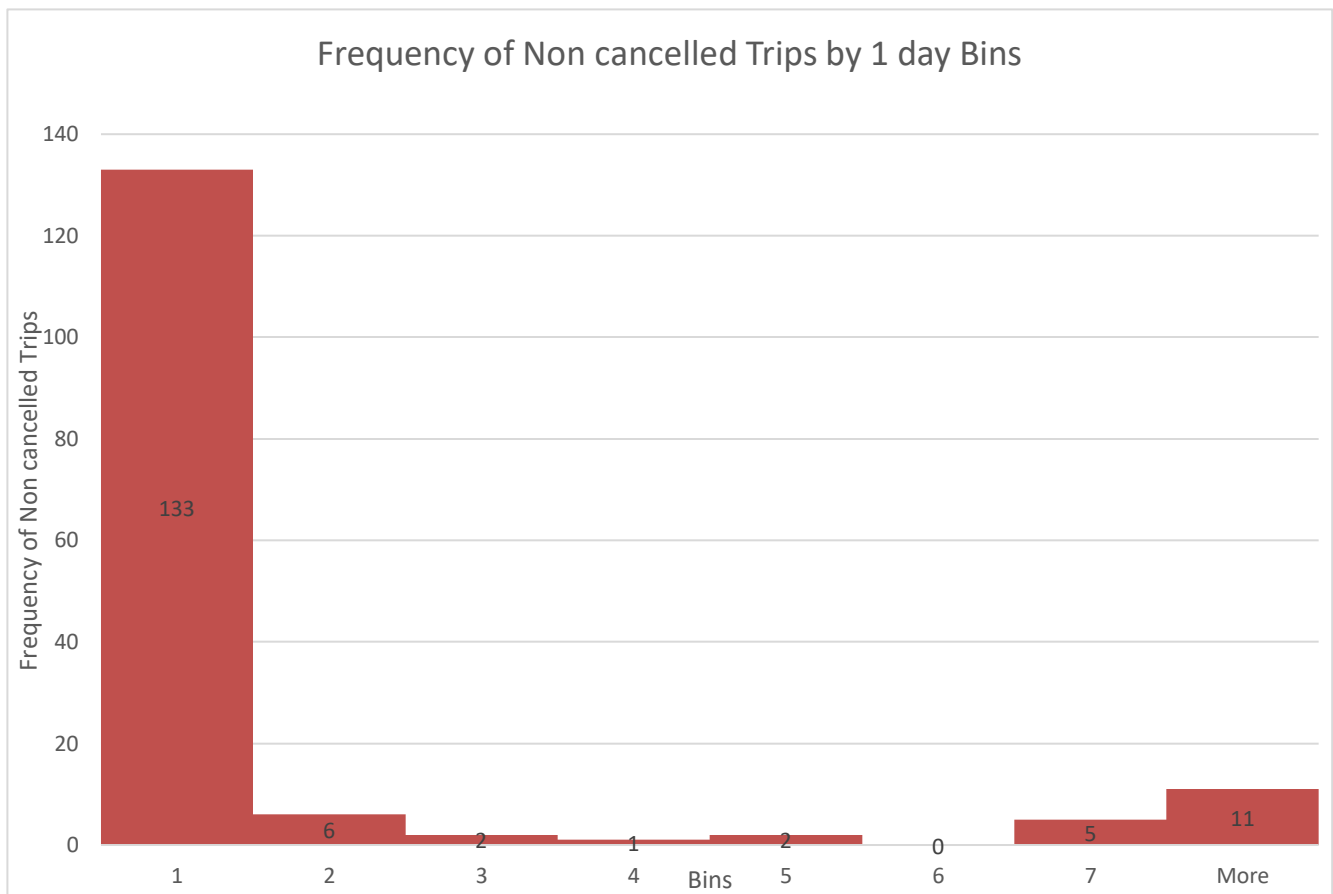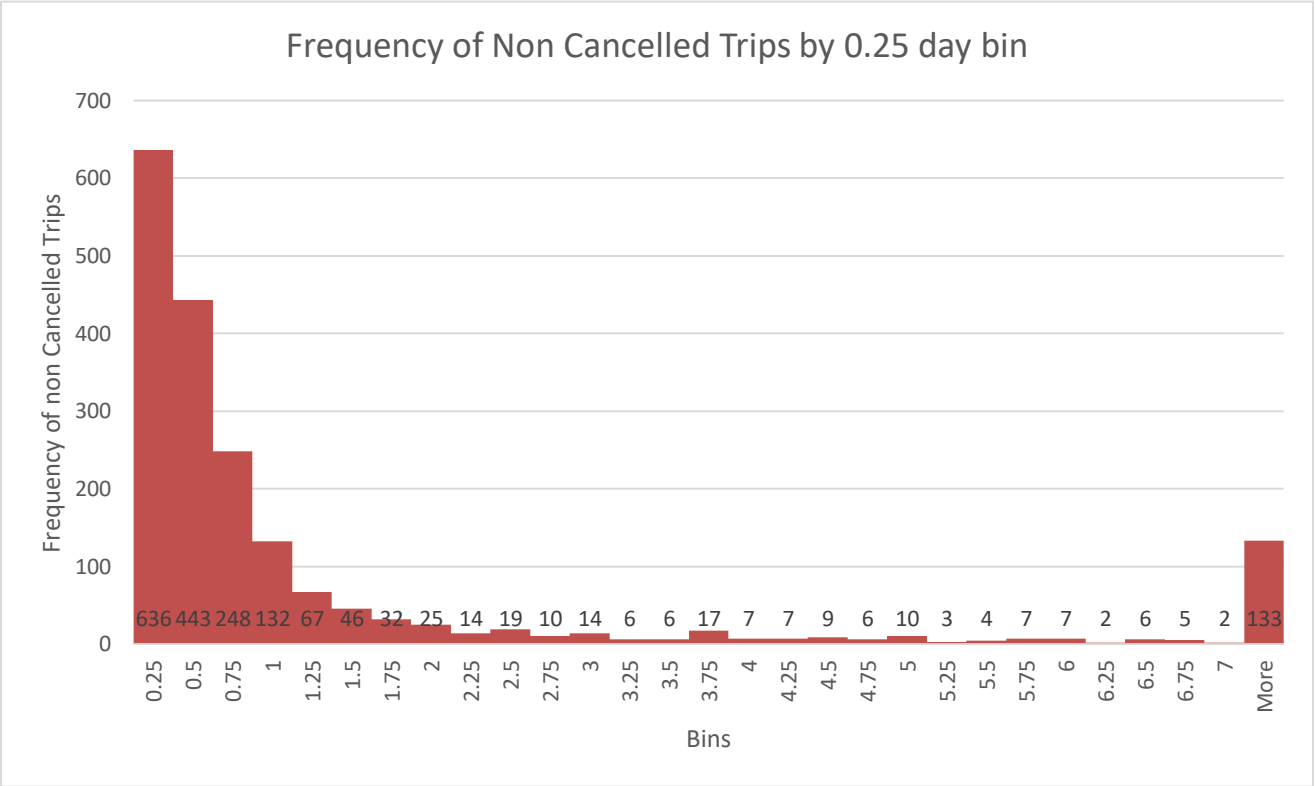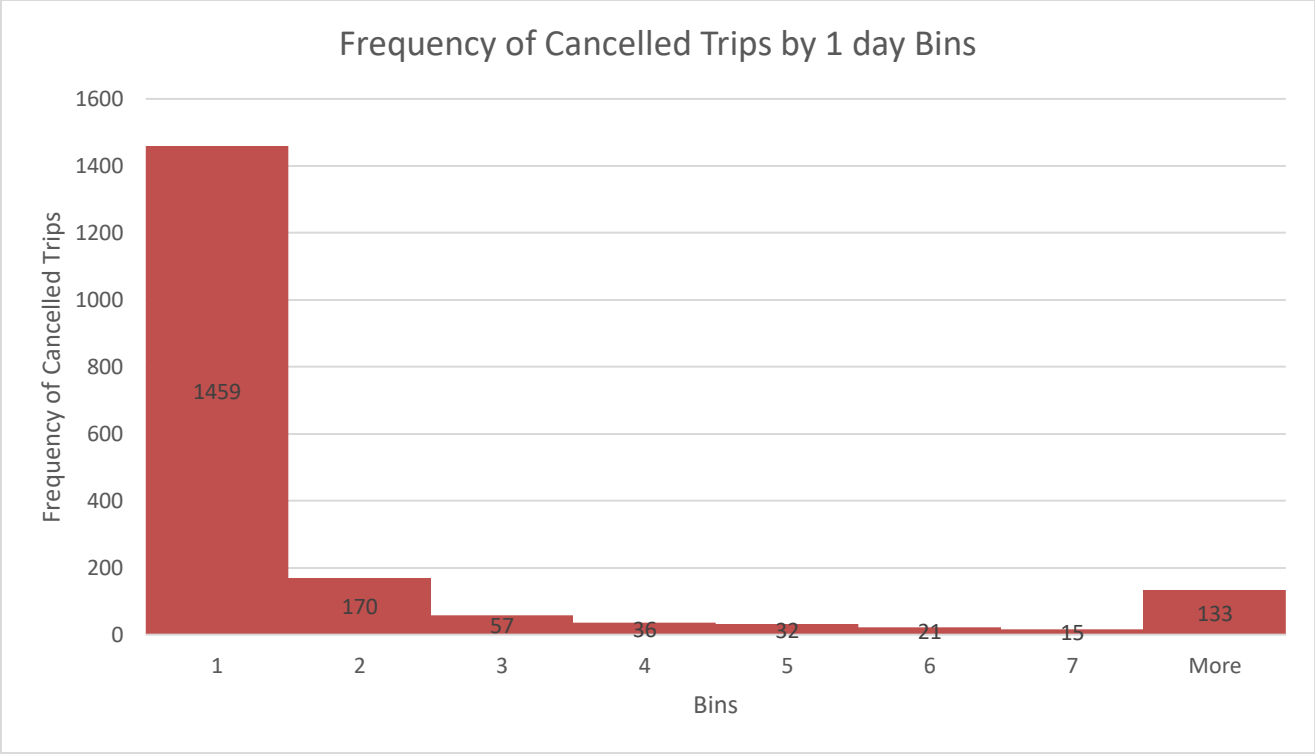
Percentage of Cancellations by Hour
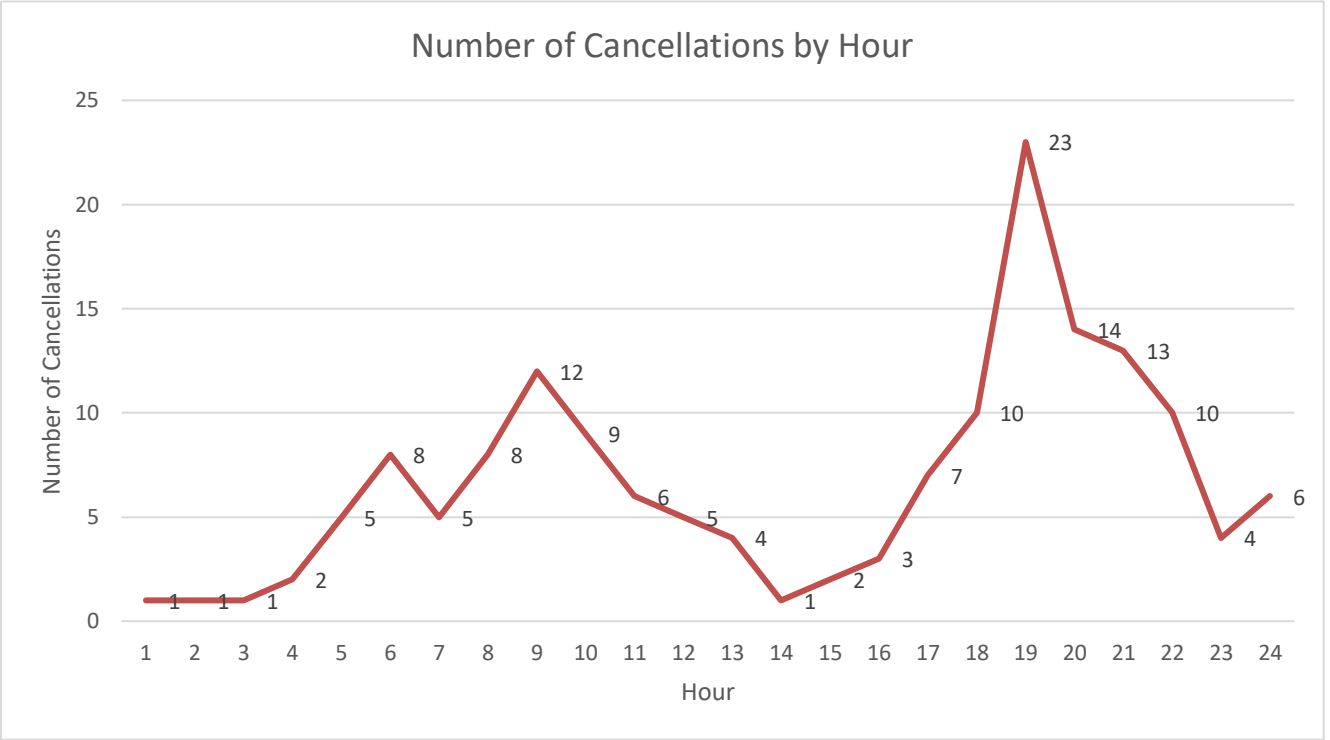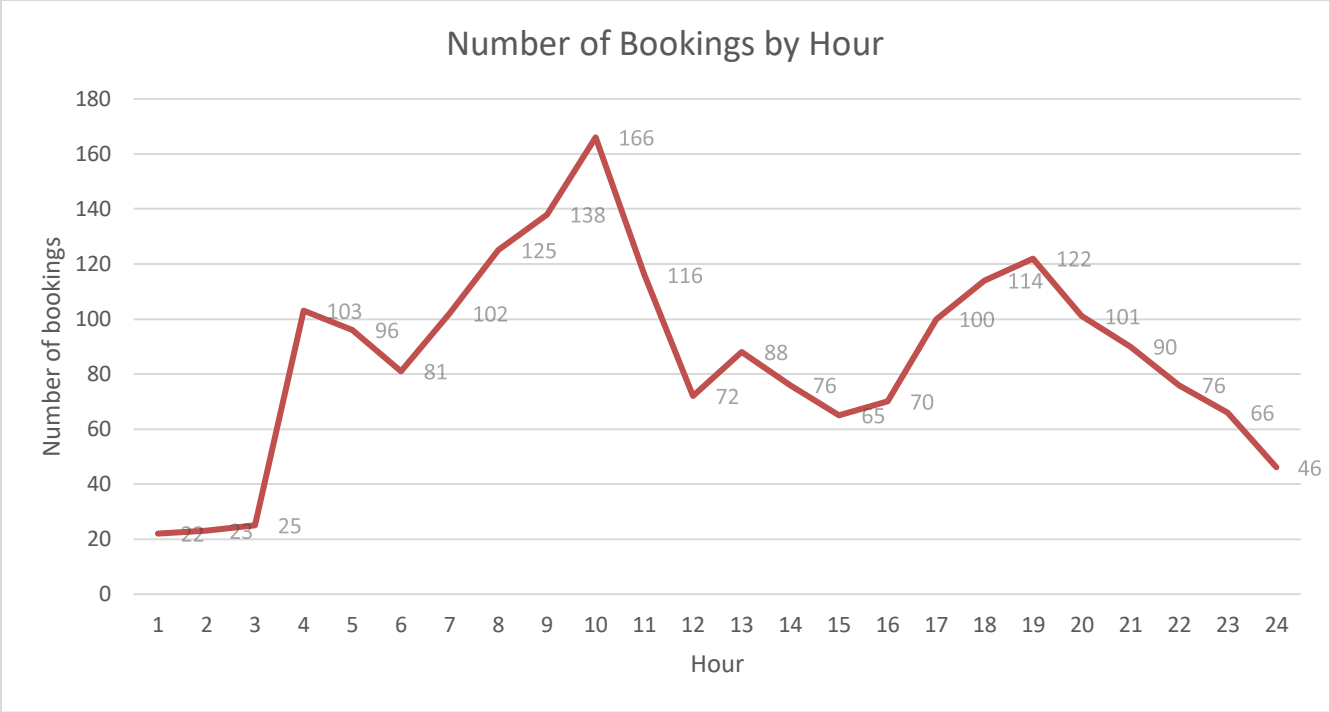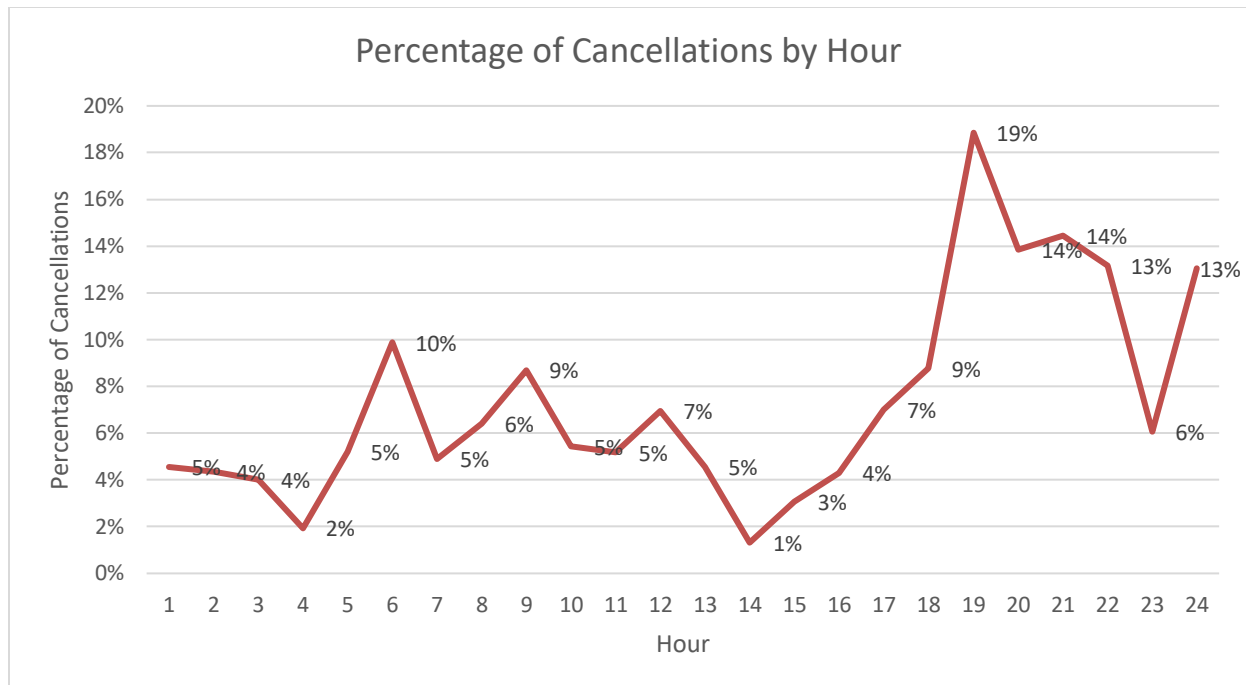
The two histograms analyze the data between booking windows and the frequency of bookings. By looking at the graph, first day bin has the most reservation significant to those of other bins. It is clearly shown that 76.43 percent of IndoCabs customers utilize one-day booking window. It is distinctly conveyed in second graph – 0.25 day bins- that customers travels within 0.25 days, which is 6 hours, to their destination. I would recommend IndoCabs about using 0.25 day of bins because it reveals specific amount of time in hour rather than depicting it as "less than one day". Specifically, 34.28 percent of costumers use IndoCabs for less than 6 hours travel and 76.43 percent of customers use cab to arrive their destination in one day bin.

Frequency of Booking Windows by 1-day Bins

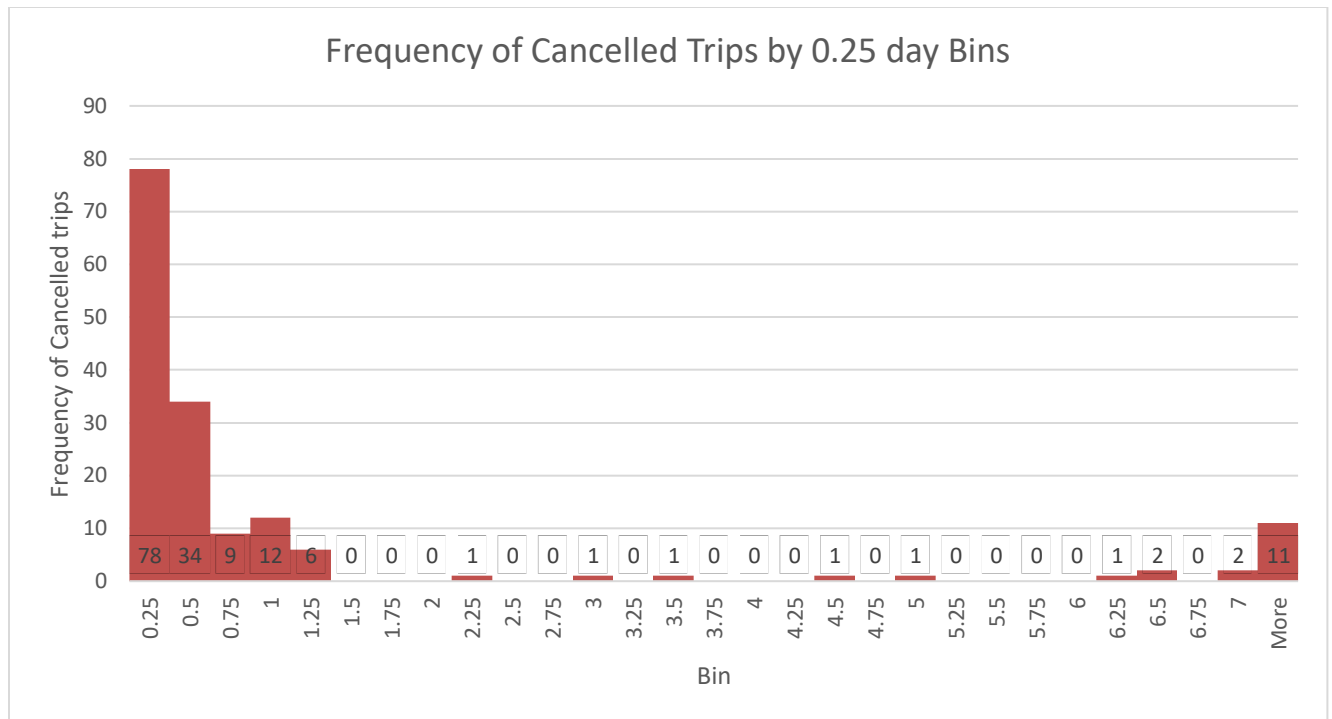Frequency of Bokking Windows by 0.25 day Bins
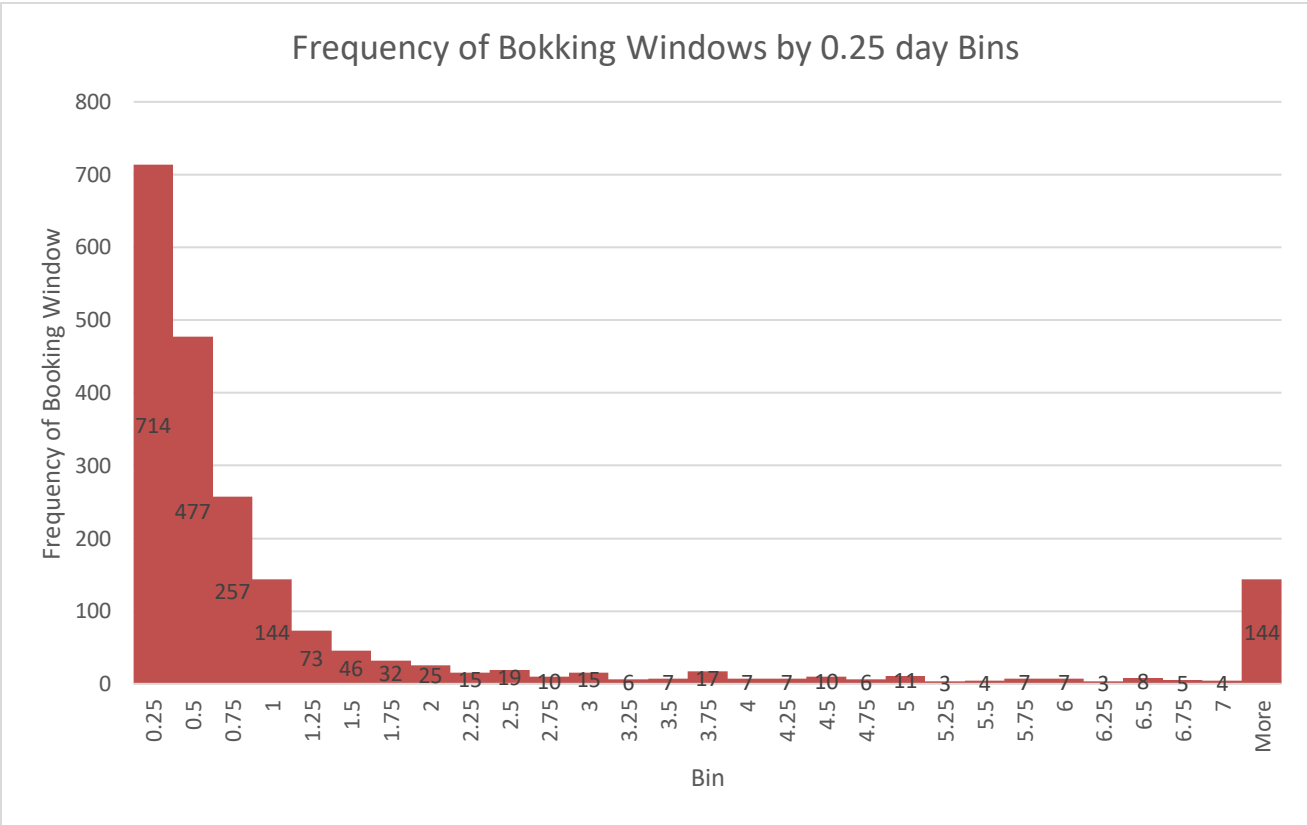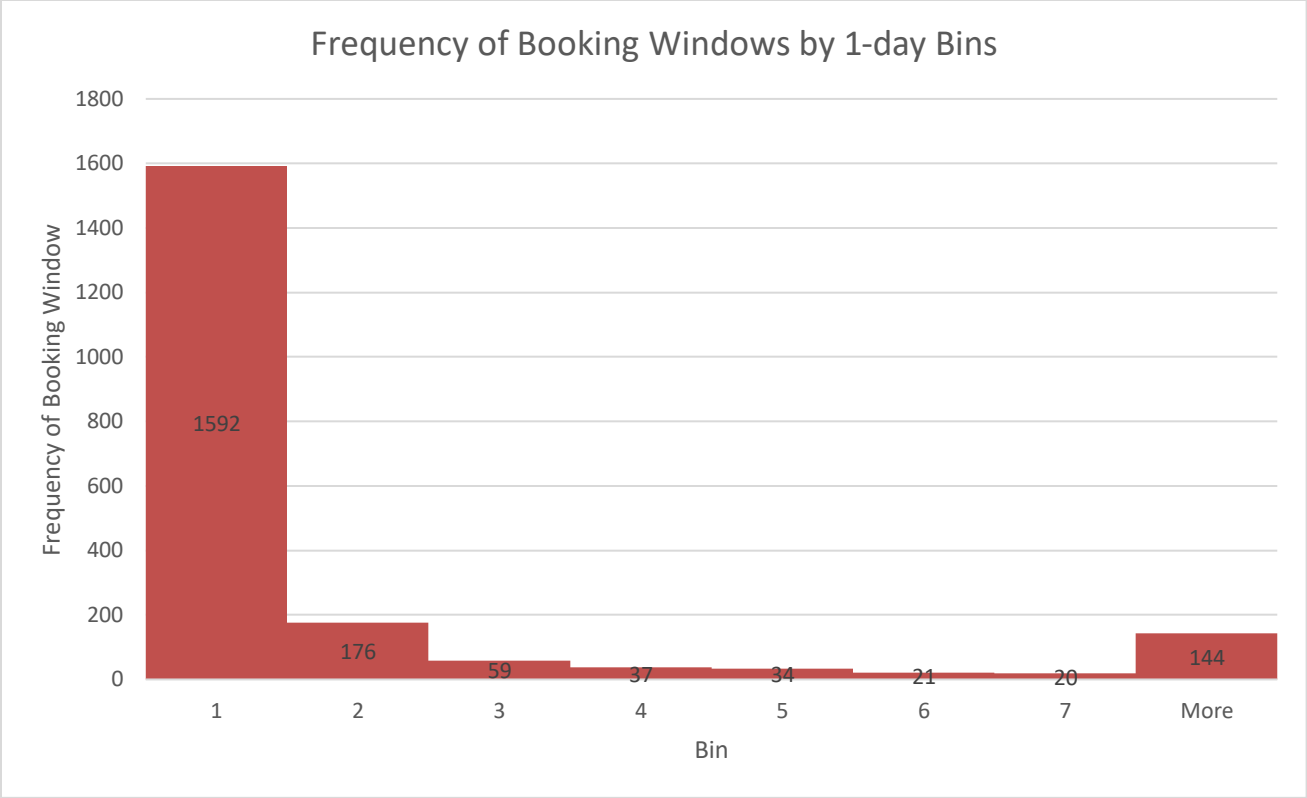
The four histograms indicate how booking windows affect cancellation rates. It is

noted that booking windows impact cancellations because there is a noticeable trend that first day bin has a significant cancellation rate. Moreover, the fourth graph – 0.25 bins of cancelled trips – indicates that the trip is most likely to be cancelled within 0.25 days, which means that the maximum of 6 hours trip is vulnerable to the cancellation. Due to the same reason with the paragraph above, I would recommend IndoCabs business executives to utilize 0.25-day bins because it categorize which specific time of day the cancellation occurs frequently. It is noticeable that there is a significant cancellation gap also in between the booking window of one day.

Frequency of Non cancelled Trips by 1 day Bins

Frequency of Cancelled Trips by 1 day Bins

Frequency of Non Cancelled Trips by 0.25 day bin

Frequency of Cancelled Trips by 0.25 day Bins

Histogram with y-axis "Frequency of Cancelled trips" (0 to 90) and x-axis "Bin". Bar values: 0.25 = 78, 0.5 = 34, 0.75 = 9, 1 = 12, 1.25 = 6, 1.5 = 0, 1.75 = 0, 2 = 0, 2.25 = 1, 2.5 = 0, 2.75 = 0, 3 = 1, 3.25 = 0, 3.5 = 1, 3.75 = 0, 4 = 0, 4.25 = 0, 4.5 = 1, 4.75 = 0, 5 = 1, 5.25 = 0, 5.5 = 0, 5.75 = 0, 6 = 0, 6.25 = 1, 6.5 = 2, 6.75 = 0, 7 = 2, More = 11

Recommendations and Conclusion

As IndoCabs suffers from a driver's high cancellation rate, the purpose of this business analyzation was to find a best way to reduce the percent of cancellation. However, it is not the best way to comply with what data has given to IndoCabs executives because as I mentioned in Data Preparation section, this is only a sample data, not a population data. Yet, IndoCabs will reach the positive outcome if they agree to follow these recommendations: support more long-distance travel, give drivers a responsibility and provide incentives, pay closer attention to 0.25-day bin.

Since there was no cancellation rate of long distance travel, IndoCabs might disregard the long-distance travel; however, IndoCabs can make a profit out of it by supporting more long-distance travel. Due to the fact that both drivers and customers plan their trips more carefully when they are traveling long distance, promoting more long-distance travel will benefit customers and it will eventually make them to appreciate IndoCabs service.

Giving drivers a responsibility and provide incentives is also my recommendation towards IndoCabs. It is realistically inevitable to shut down the online booking system;

11

although it has a highest rate of cancellation rate, the online booking significantly has the highest rate of booking. Therefore, when drivers do not show up more than three times, IndoCabs can make drivers to pay a certain commission. Another recommendation would be to delay their matching time with the customers. IndoCabs can rate driver's personal cancellation rate, and if it exceeds more than 5%, IndoCabs will match other drivers with customers first, followed by those drivers. While giving drivers a responsibility, I recommend IndoCabs to provide incentives too. Incentives will allow drivers to motivate not to cancel the ride.

Due to the fact that booking cancelled occurred mostly within six hours, I would advise IndoCabs to pay closer attention to 0.25 day bins. In 0.25 day bins, IndoCabs executives can give stricter penalty if drivers cancel the bookings. However, before that, IndoCabs should enforce the system so that the drivers take unfair penalty due to the traffic, accident or any impediments.

Elevator Charts

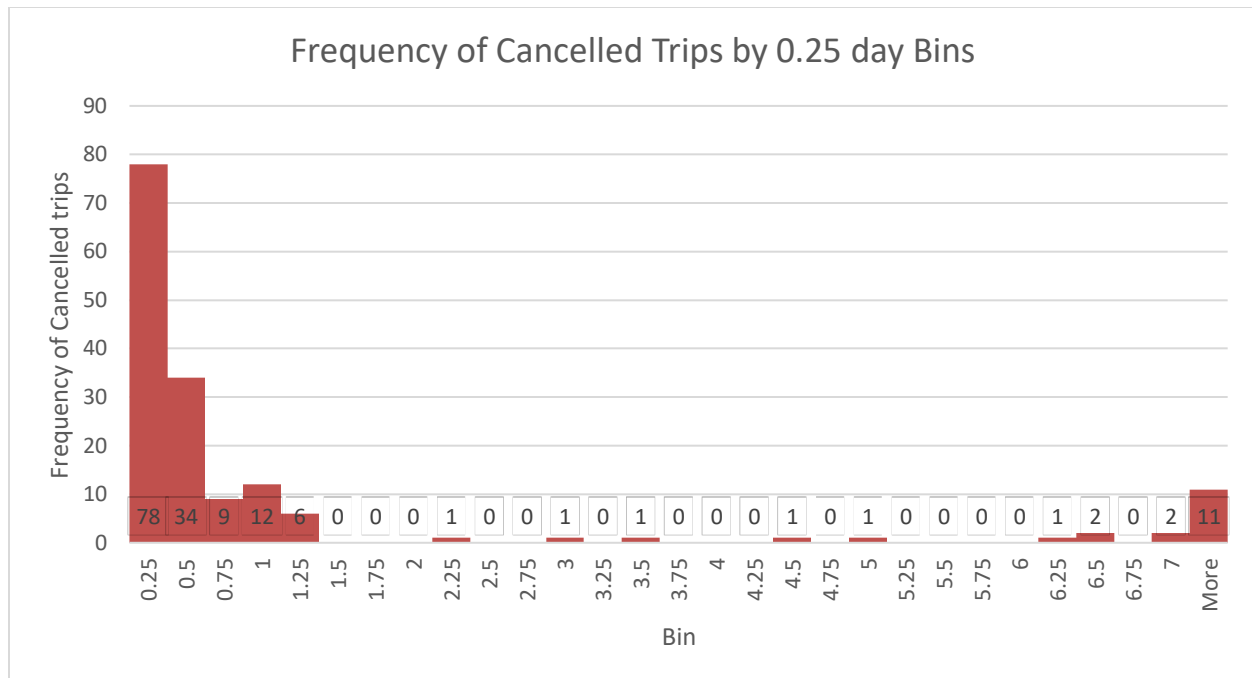Sum of Bookings and Cancellations by Hour

$R^2 = 0.3442$
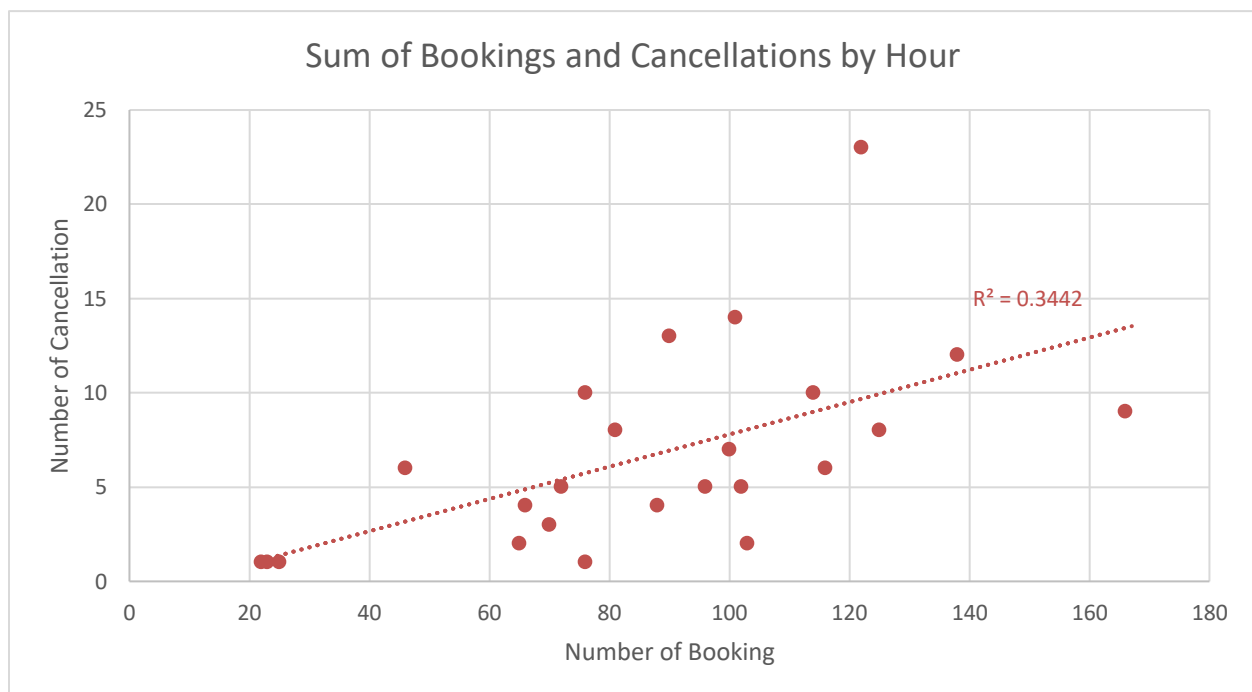
It was surprising to find out that there is no strong correlation between the number of bookings by hour and number of cancellation by hour. We expected them to correlate each other; however, we were not able to find the correlation between those two data

given.
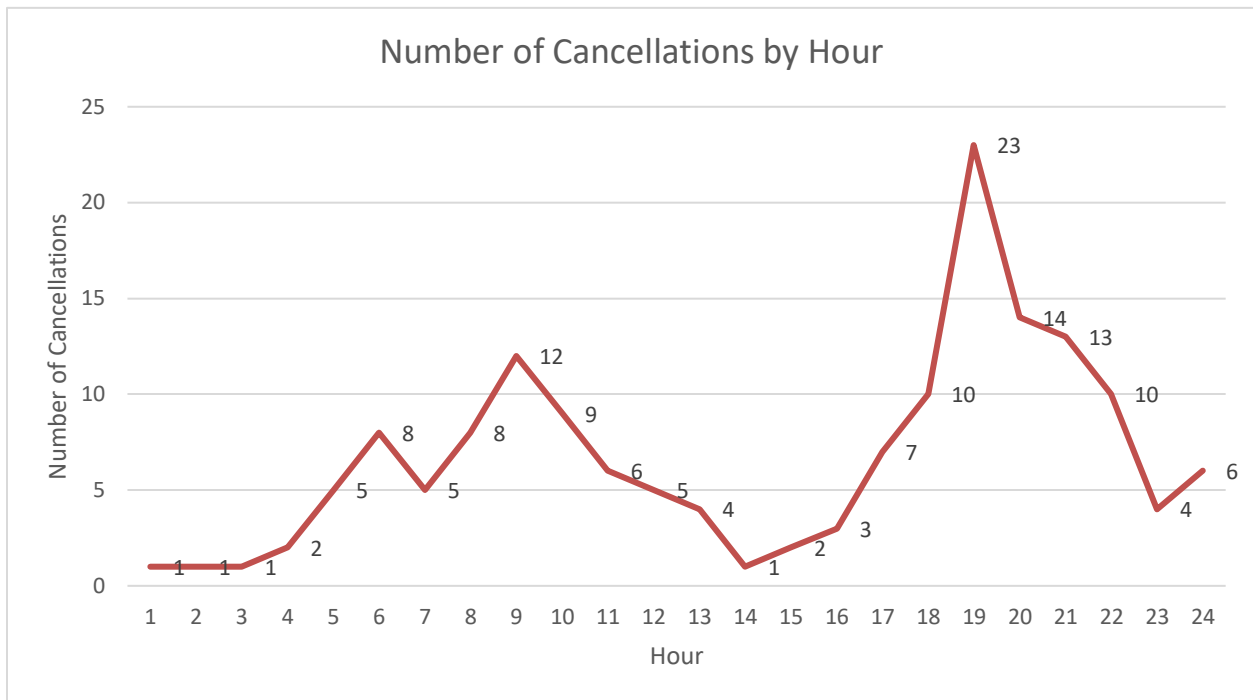
This graph shows the number of cancellations by hour, and it was interesting in a way that the cancellation happened mostly at 7 P.M. In my opinion, I think that this graph can be inferred that IndoCabs system has more errors at seven than it does in any other times because the number of customers dramatically arise at this point of the day due to the time of leaving work.

Frequency of Cancelled Trips by 0.25 day Bins

Bin	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.25	3.5	3.75	4	4.25	4.5	4.75	5	5.25	5.5	5.75	6	6.25	6.5	6.75	7	More
Frequency	78	34	9	12	6	0	0	0	1	0	0	1	0	1	0	0	0	1	0	1	0	0	0	0	1	2	0	2	11

This graph indicates that more than half of the car cancellation occurs within six hours of booking window. Yet, it is already revealed that most of the reservation is made within one day of booking window. I found it interesting because there should be more data explaining "how to reduce the cancellation rate within one day bin.

Notes on Data Preparation

When analyzing the data given by IndoCabs, there was a significant number of erroneous data in my sample given. Although my data had no duplicates, there were some travel date that fall before January 1, 2013. I found some dates that were written as 2070 instead of 2013. Since I did not edit 2070 as 2013 but rather erased all the data containing wrong dates as the instruction given to me, I think those data containing erroneous "to_date" might have changed my analysis. Moreover, since I used sample data instead of a population data, the analysis might have been different to those who used other sample than mine or population data analysis.