Analysis of Cancellations at a Cab Portal Company

ANNA GILHOOL
97
WISCONSIN SCHOOL OF BUSINESS
MARCH 2018

Executive Summary

IndoCabs is a cab company based in India, which acts as a go-between for customers and smaller cab companies. Through IndoCabs people can book long distance, point to point, hourly rental trips, and cabs rides within cities online, by mobile website, text, email, phone call, or web chat. Due to a numerous amount of cancellations by cab drivers on the customers, IndoCabs was forced to shut down its services. In this report, the magnitude of the problem has been analyzed and the relationship between booking windows, trip start time, and cancellations is evaluated.

From examining a sample of 2046 bookings, 8.80% of bookings were cancelled, showing the magnitude of the problem. Out of all the booking types, Point to Point bookings were cancelled the most at 10.01%. Point to Point is the most popular booking type, so it cannot be removed from IndoCabs services. However, to reduce this cancellation rate the price of Point to Point trips could be increased, so drivers would be more motivated to provide cab services. Furthermore, mobile and online bookings channels were analyzed. Mobile booking had a cancellation rate of 12.80% and online had a cancellation rate of 14.80%. Both of these rates are higher than the average rate of 8.80%, meaning that these channels are not efficient and should be diminished. It would be better for IndoCabs to utilize their other channels (text, email, phone call, and web chat).

Cancellations were also analyzed by day of the week, concluding Sunday at the highest cancellation rate of 11.43% and Saturday with the lowest rate at 4.92%. Also, Monday and Wednesday through Friday all had cancellation rates above 10%. This could be due to an increase need for cabs rides on these days. Since these days have a higher influx of requested rides, IndoCabs could implement higher prices to balance out cab requests and driver willingness to follow through with customers' requests.

Additionally, the correlation between number of bookings and number of cancellations by hour was investigated. A correlation of 62.30% was found, demonstrating that trip start time has a significant effect on the cancellation rate. The percentage of car cancellations by hour was examined, which is helpful in illustrating the correlation between the hour of the booking and the proportion of bookings cancelled in that hour. The results showed a higher number of cancelled trips in the morning, lunch, and late at night. This is expected as morning and lunch are rush hours and late at night has a lower number of drivers available. For example, hour 18 has the highest cancellation rate of 21.24%, which is probably correlated with the late trip time. Implementing a higher price at these hours could reduce the cancellation rate due to a higher fare being an incentive for the driver.

Finally, the distribution of bookings windows for cancelled trips in a .25-day interval was calculated. The results showed that the most cancellations (33%) occurred in 0-.25 days (0-6 hours).

Analysis

A Look at Trip Durations

Summary Statistics of Trip Duration and Booking Window (Table 1)

	Trip Duration (hours)	Booking Window (days)
Median	1.34	0.43
1st Quartile	0.95	0.17
3rd Quartile	2.66	0.94

Table 1 communicates the summary statics median, 1st Quartile, and 3rd Quartile for trip duration and booking window. The median is a measure of central tendency and corresponds to the middle values of the data. Using the median instead of the average allows for less distortion. If the average is used instead of the mean, then outliers can skew the data and the average would be misrepresented. The 1st quartile represents the 25th percentile, meaning 25% of the data is below this value, and the 3rd quartile represents the 75th percentile, the upper median. Analyzing the 1st and 3rd quartile with the median allows to see the range of data in the sample. The 1st quartile gives an idea about the lower end of the data set, the median represents the middle set of the data, while the 3rd quartile represents the upper data.

The Magnitude of the Cancellation Problem at IndoCabs

Examining a sample of 2046 bookings, 180 bookings were found to be cancelled, meaning 8.80% of all bookings were cancelled; this is a very significant complication within the company. Out of all bookings, the most cancelled bookings type is Point to Point, having a cancellation rate of 10.01% compared to a long distance booking at 1.41% or a hourly rental at 5.05% (illustrated in Table 2 below). Point to Point is the most common booking type, making up 1599 out of the 2046 bookings, which could be correlated to a higher cancellation rate. The cancellations could be caused by the lower fare rate the drivers earn, compared to Long Distance and Hourly Rental income. Since Point to Point is the most common booking type and has the highest cancellation rate, it would be most efficient to focus on lowering the Point to Point cancellations. This could be achieved by raising the fare of Point to Point travel, so drivers are less likely to cancel.

Cancellations by Travel Type (Table 2)

	Number of Bookings	Number of Cancellations	Percent Cancelled
Long Distance	71	1	1.41%
Point to Point	1599	160	10.01%
Hourly Rental	376	19	5.05%

Comparing the online and mobile booking channels, there was no significant difference. 117 out of 881 online bookings were cancelled, meaning 14.43% of bookings made online were cancelled, and 16 out of 125 mobile bookings were cancelled, meaning 12.80% of bookings made through mobile booking were cancelled. The 2% difference between these bookings methods is very small, but both have a significant difference from the average of cancellations at 8.80%. It would be recommended to shutdown both mobile and online services and expand to booking by the other channels offered at IndoCabs (text, email, phone call, and web chat).

Chart 1

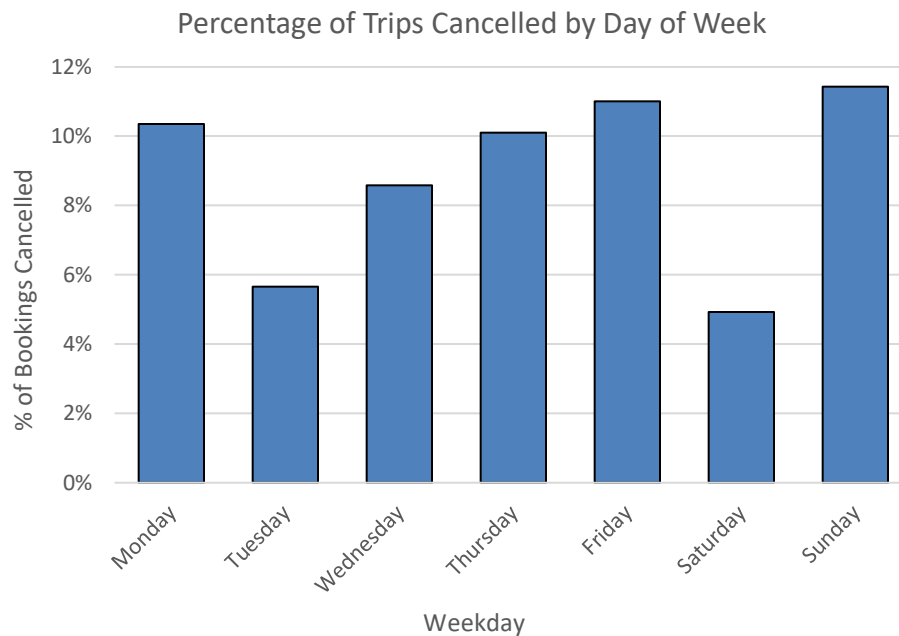**Percentage of Trips Cancelled by Day of Week**

Chart 1 displays the percentage of cancellations by weekday, conveying Sunday with the highest cancellation rate at 11.43% and Saturday with the lowest rate at 4.92%. Most of the cancellations have a percentage around 10%, but only Tuesday and Saturday stand out as having a cancellation rate of lower than 6%. The pattern is somewhat surprising as cancellations would be expected to be the highest on the weekend due to less drivers wanting to work. Sunday follows this expected pattern, but Saturday has the lowest cancellation rate, which is unexpected. This could be due to workers having more time available on Saturday solely for IndoCabs bookings. Again, implementing a higher fare on high cancellations days could reduce cancellations.

The Relationship between Booking Windows, Cancellations, and Trip Timing

Chart 2

Number of Bookings and Cancellations by Hour

$R^2 = 0.3881$
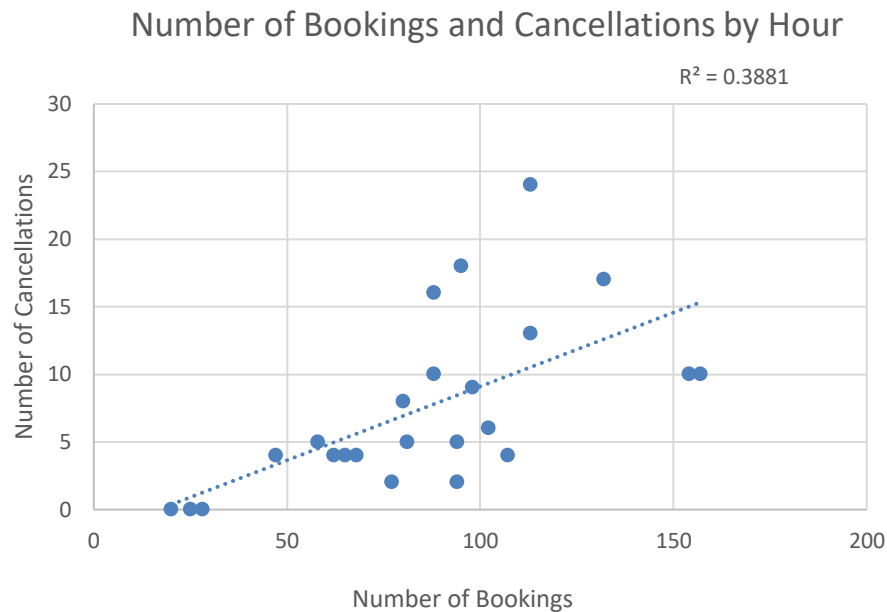
The scatterplot above illustrates the relationship between the number of bookings and the number of cancellations by hour. The R-square illustrated on the chart represents how close the data is to the regression line. The square root of R-squared is the correlation between the two variables. Therefore, the R-squared value of .3381 on Chart 2 conveys the strong positive correlation of 62.30% between hourly number of bookings and the hourly number of cancellations. This demonstrates that trip start time has a significant effect on the cancellation rate.
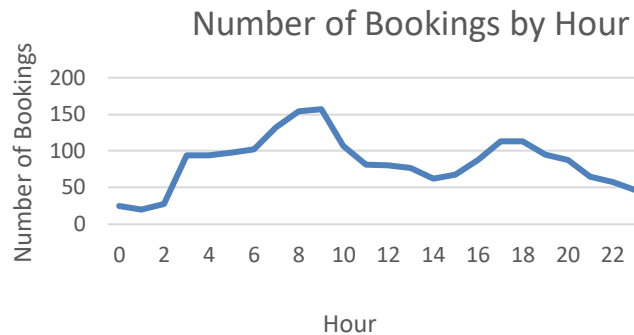
Chart 3

Number of Bookings by Hour

Chart 4

Car Cancellations by Hour
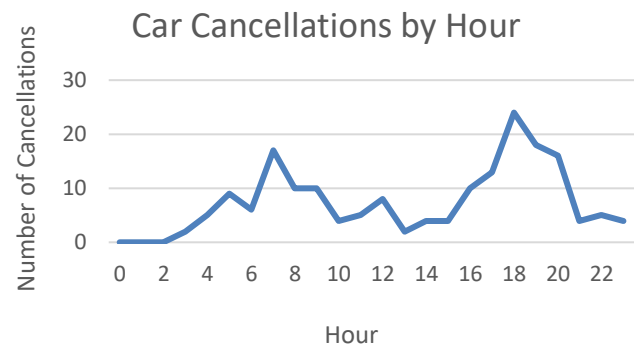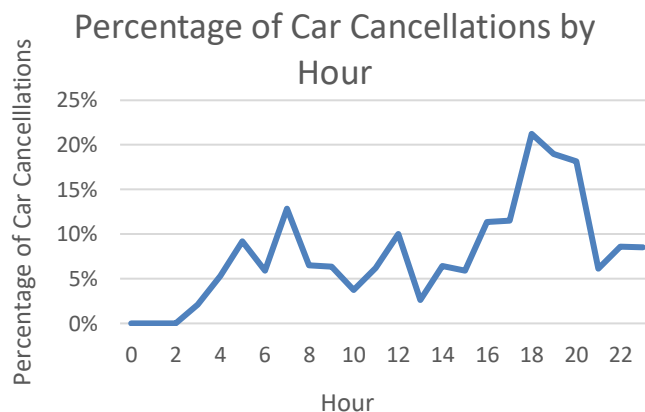

Chart 5

Percentage of Car Cancellations by Hour

The patterns in Charts 3, 4, and 5 are all very similar. However, Charts 4 and 5 are more informative; they give data on the number of cancellations, instead of just reporting the number of trips at a certain hour as Chart 3 does. Charts 4 and 5 illustrate the cancellation problem at each hour and provide information about at what hour cancellation rates are high or low. Overall, Chart 5 is the most informative because it uses the percentage of cancellations, which includes the number of bookings and the number of cancellations, providing the most accurate data. Chart 5 concludes that the most cancellations occur in the morning, around lunch, and late at night. During these hours it would be effective to raise the pricing.
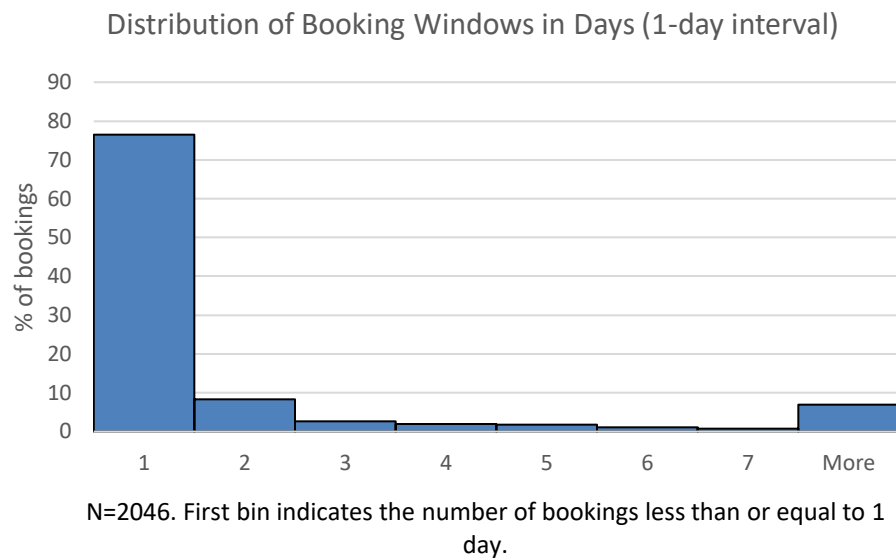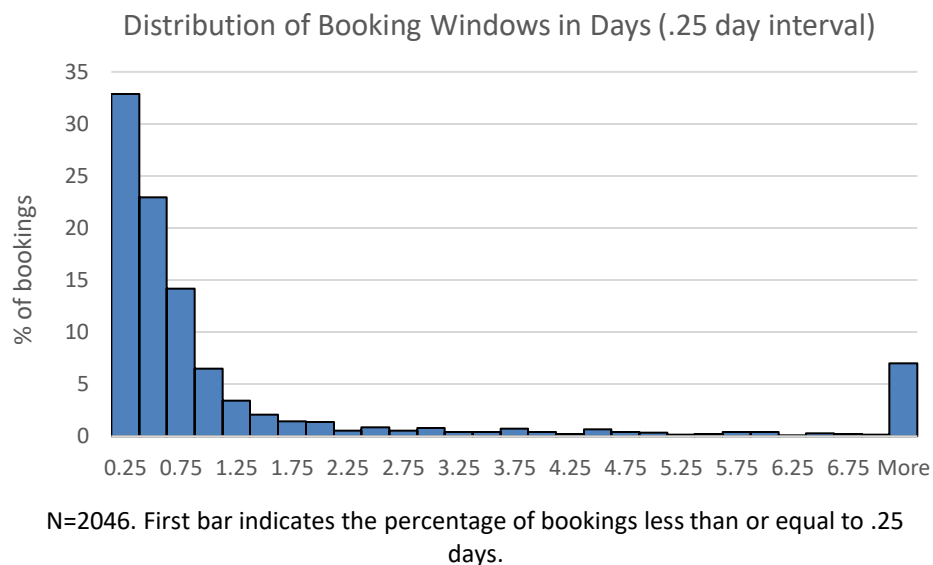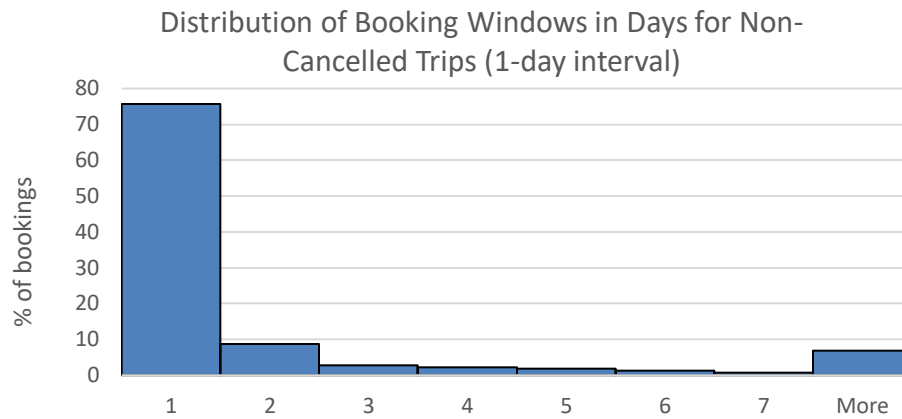
Chart 6

Distribution of Booking Windows in Days (1-day interval)

N=2046. First bin indicates the number of bookings less than or equal to 1 day.

Chart 7

Distribution of Booking Windows in Days (.25 day interval)

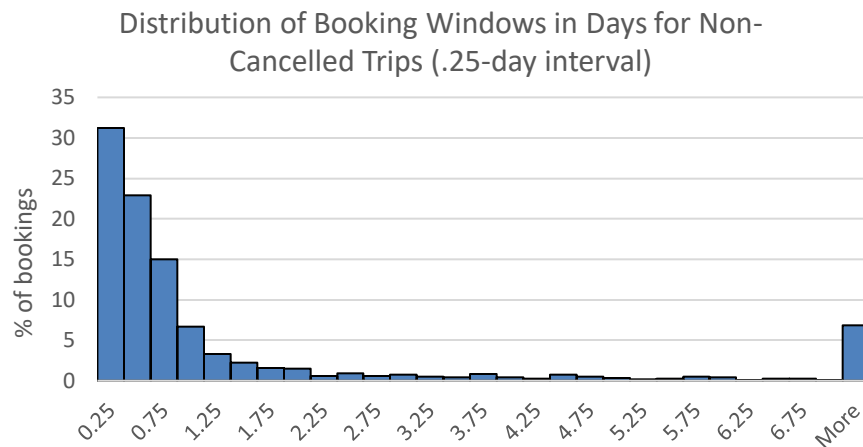N=2046. First bar indicates the percentage of bookings less than or equal to .25 days.

Charts 6 and 7 demonstrate the distribution of booking windows in 1-day intervals and .25-day intervals. Chart 7, which has bin sizes of .25 days, is the most revealing of the underlying patterns of the distribution of booking windows. When the booking windows are categorized into bins of only 7 days, this distorts the amount of booking windows, since some were less than one day. The 1 bin day in Chart 6 has a frequency of 70%, but the data does not illustrate that about 33% of the bookings were .25 days. Therefore, the .25 day bins are more specific and more accurate to draw conclusions from.

Chart 8

Distribution of Booking Windows in Days for Non-Cancelled Trips (1-day interval)

N=1866. First bin indicates the number of non-cancelled bookings less than or equal to 1 day.

Chart 9

Distribution of Booking Windows in Days for Non-Cancelled Trips (.25-day interval)
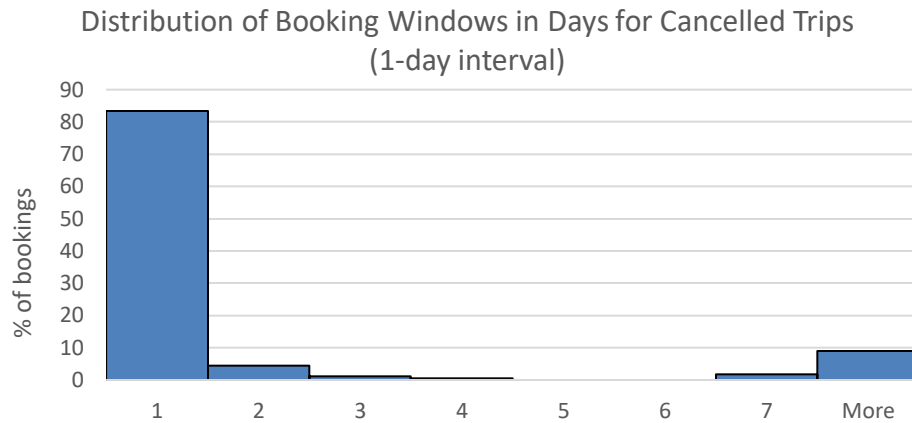

N=1866. First bin indicates the number of non-cancelled bookings less than or equal to .25 days.
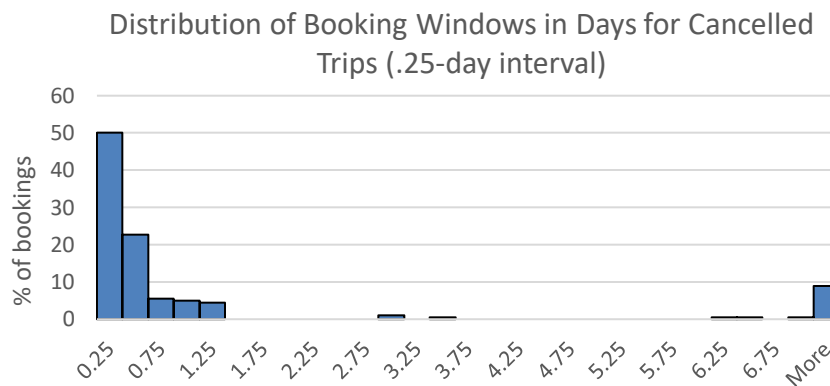
Charts 8 and 9 show the distribution of booking windows for non-cancelled trips. Chart 8 illustrates the booking window through 1 day intervals and Chart 9 illustrates the booking windows through .25 day intervals. Chart 9 is preferred because there is more detail. From Chart 8 it is noticeable that most booking windows are in bin 1. However, with Chart 9 the distribution is more specific, so the company can see that the distribution is really the highest at .25 days, not 1 day.

Chart 10

Distribution of Booking Windows in Days for Cancelled Trips (1-day interval)

N=180. First bin indicates the number of cancelled bookings less than or equal to 1 day.

Chart 11

Distribution of Booking Windows in Days for Cancelled Trips (.25-day interval)

N=180. The first bin indicates the number of of cancelled bookings less than or equal to .25 days.

Charts 10 and 11 illustrate the distribution of bookings windows for cancelled trips in 1 day and .25 day intervals. Chart 11 is preferred due to the same reasons between Charts 8 and 9 (Chart 11 is more precise). Therefore, .25-day intervals are more helpful in communicating insights to bookings and cancellations. From all of the histograms, it is noticeable that booking windows affect cancellations. Chart 7 shows that the largest distribution is .25 days (0-6 hours), and Chart 11 shows the distribution of cancelled bookings is highest at the .25 days. Both the number of bookings and number of cancellations are high at this point, proving the booking window and cancellation rate are correlated.

Elevator Charts

Chart 1

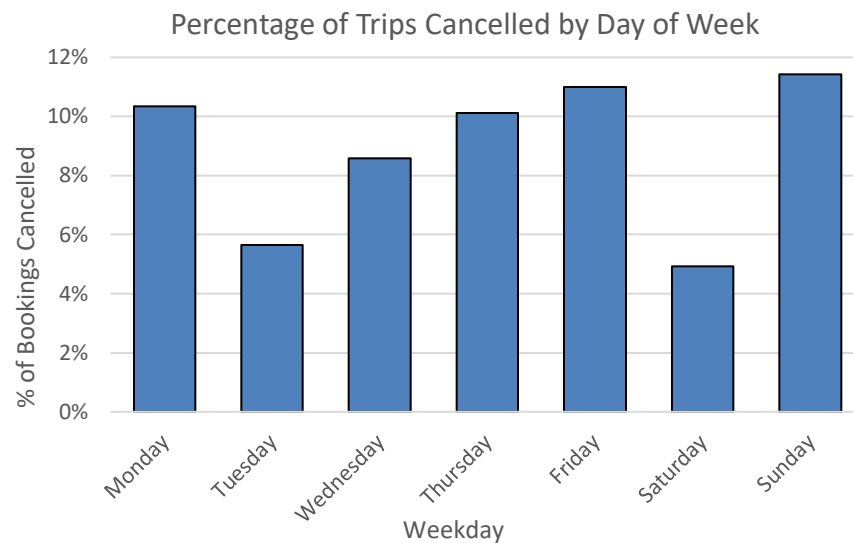### Percentage of Trips Cancelled by Day of Week

Chart 1 highlights key days in which IndoCabs needs to minimize cancellations, which is very easy to analyze by this graph. The bar chart is simple to read and effectively gets across the message that most days of the week, 10% or more trips are cancelled – telling the company something needs to be changed, such as number of employers working or higher pricing on days with more cancellations.

Chart 5

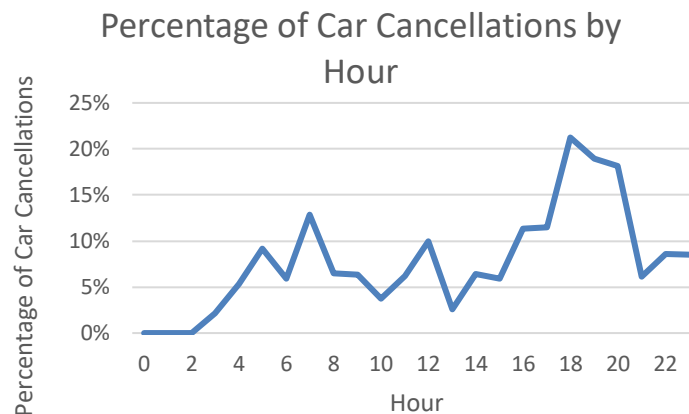### Percentage of Car Cancellations by Hour

Chart 5 provides significant information about the correlation between hour and the percentage of car cancellations, which allows the company to see what hour is most problematic (what hour has the most cancellations relative to the number of bookings in that hour). This chart is also easy to read; the maximum points clearly convey hours at which there are many cancellations and minimums convey low cancellation.

Chart 11

Distribution of Booking Windows in Days for Cancelled
Trips (.25-day interval)

N=180. The first bin indicates the number of of cancelled bookings less than
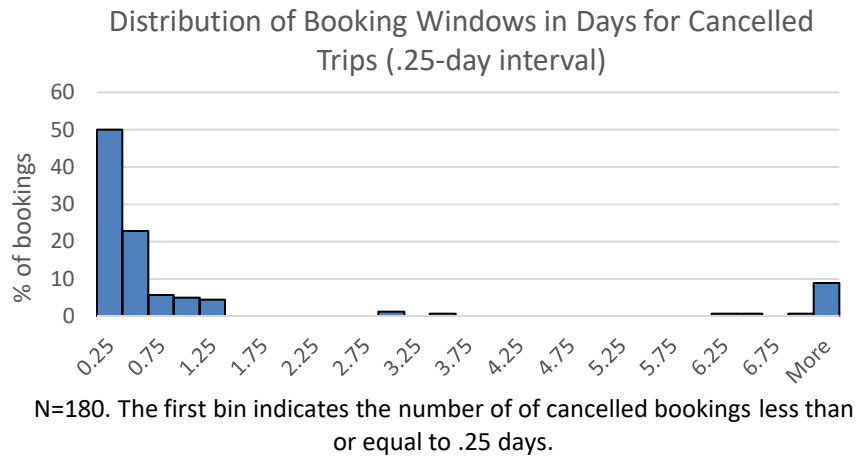or equal to .25 days.

Chart 11 gives significant information on the distribution of the booking window for cancelled trips. Using the .25 day intervals allows for the most precision, while using the booking window for cancelled trips allows IndoCabs to see the when cancellations are most likely to occur.

Notes on Data Preparation

To prepare the data, three error types were removed; duplicate observations, date entries before January 1, 2013, and any data entry errors in the booking date occurring after the trip start date were removed. The data had 100 duplicate values out of 25,114 records, which is a very low number for the number of records in this data set. After the data was cleaned, a sample of 2046 records was taken from a total of 25,144 records. From the sample data there was only 1 duplicate found, but there were dates before January 1, 2013 and negative values found in the booking window. Since only a small sample of data was used this could have resulted in a misrepresentation of the data. Also, additional variables were created including weekday, hour, duration, duration (hour), and booking window. These new variables were also filtered and cleaned by removing any negative values for the booking window.