IndoCabs Cancellations Analysis

Michael Vitale
Wisconsin School of Business
March 2021

Executive Summary

IndoCabs is a cab portal company based in India that has temporarily suspended its operations. IndoCabs business model is unique in that the company connects customers to a variety of smaller cab companies rather than managing their own drivers. They have suspended operations due to a high proportion of cancellations by drivers, which threatens the success of their company. In the given sample, about 9% of all bookings were canceled. Within this proportion, 21% of mobile bookings and 13% of online bookings were canceled. As evident from the data, online and mobile bookings are being canceled at a relatively higher proportion than otherwise. Given the higher cancellation rate of mobile bookings, I would advise the company to remove mobile bookings from their business model and only service other travel types. Next, the vast majority of bookings and the highest proportion of cancellations occur within one day of an intended ride. I would advise the company to raise prices on bookings made less than one day in the future, to encourage customers to book in advance, and thus decrease cancellations. Third, of the services offered by IndoCabs, point to point urban travel has a significantly higher cancellation rate (9.8%), and I would advise that they marginally increase the price on such travel types to incentivize drivers not to cancel. Finally, there are a few metrics which outline periods of increased frequency of bookings, and I would encourage IndoCabs to recruit a larger number of drivers during this period to supplement increased cancellations. All considered, in order for IndoCabs to recommence operation, the company needs to tweak their business model to consider not only serving the customer, but also drivers. By recognizing and evaluating the tendencies of their customers and drivers, they will be more successful going forward.

Analysis of Trip Durations and Booking Window

Trip duration is a measure of the time period from requested trip start to requested trip end. There are a few valuable metrics to consider when looking at the entirety of data regarding trip duration. To start, the mean trip duration of all the data is 4.46 days, while the median value is 1.34 days. While the median is more telling of the most commonly occurring trip duration, it is important to mention both measures of center to display the disconnect between them. There exists a significant disconnect between the mean and the median largely because of the very wide range of the data, in which very high values pull up the mean. The range for trip duration is 207.40 days, meaning there exists very high values of trip duration. As I previously stated, this is important to mention because the median may be more representative of the true center, but it is important also to consider the wide range of possible trip duration values.

A very similar pattern exists when considering the variable booking window which measures the distance in time between when the booking was initially created and when it is supposed to begin. The mean value of the booking window is 2.00 Days, and the median value is .41 days. Once again, there exists a large difference between the two measures which is a result of the large upper values in the wide range, which is 32.23 days. Again, the median is likely a better measure of center in this instance but it is still important to mention the higher mean and large range.

Summary Statistics of Trip Duration and Booking Window

	Trip Duration	Booking Window
Mean	4.46	2.00
Median	1.34	.41
Range	207.40	32.23

When considering the magnitude of Cancellations:

Based on the bookings provided in the sample, there were 219 cancellations out of a total of 2486 bookings in the sample. This means that about 9% of total bookings were canceled. This is obviously an unreasonably high percentage– which Indocabs can not support going forward

Magnitude of Cancellations
Total number of cancellations: 219.00
Total number of bookings: 2486.00
Proportion of canceled bookings .09

Further Segmenting the Cancellation of Bookings

Based on the data, the vast majority of bookings are short distance trips from point to point in urban environments. 1955 out of 2486 bookings were of this sort. While this is the most common type of booking, it also has the highest proportion of cancellations of any of the travel types– about 9.82%. Hourly rentals has the second largest number of bookings from the sample data, with 419 bookings, though this is still more than 4 times smaller than the frequency of trips from point to point. Hourly rentals have a cancellation rate of 5.97% which is still large, though smaller than point to point travel. Finally, long distance trips have the smallest number of bookings, with 112 from the sample, and the smallest proportion of cancellations, with about 1.79% of bookings canceled.

There exists a clear pattern in the data, which is in line with my expectations. First, I would assume long distance trips would have a larger proportion of cancellations and this is reinforced by the data. I believe long distance trips would have a larger threshold of commitment, as I would assume the driver would likely not take on the responsibility of a long trip lightly and would thus cancel with lower frequency. There is likely also a higher monetary payout for drivers for longer trips, which would be harder to rationalize forfeiting. Hourly rentals have a higher proportion of cancellations, likely because of the decreased time commitment and lessened net gain in terms of revenue. Point to point travel in urban areas has the highest cancellation proportion– possibly because it is seemingly the lowest commitment to the driver in terms of time and revenue. Overall, the pattern seems to be that drivers are more committed and less likely to cancel, when trips are longer in duration.

Cancellations by Travel Type

	Number of Bookings	Number of Cancellations	Proportion canceled
Long Distance Trips	112	2	0.0179
Trips From Point to Point in Cities	1955	192	0.0982
Hourly Rentals	419	25	0.0597

When Considering Booking Channels

It's important to consider two of the most common booking channels when analyzing the sample data: online bookings and mobile site bookings. Online bookings account for more than a third of total bookings, with 986 bookings in the sample data. Of these bookings, about 13% were canceled.Mobile bookings made up a much smaller amount of the total bookings in the sample data, with 141 of 2486 bookings being mobile bookings. Of these bookings, about 21% were canceled.

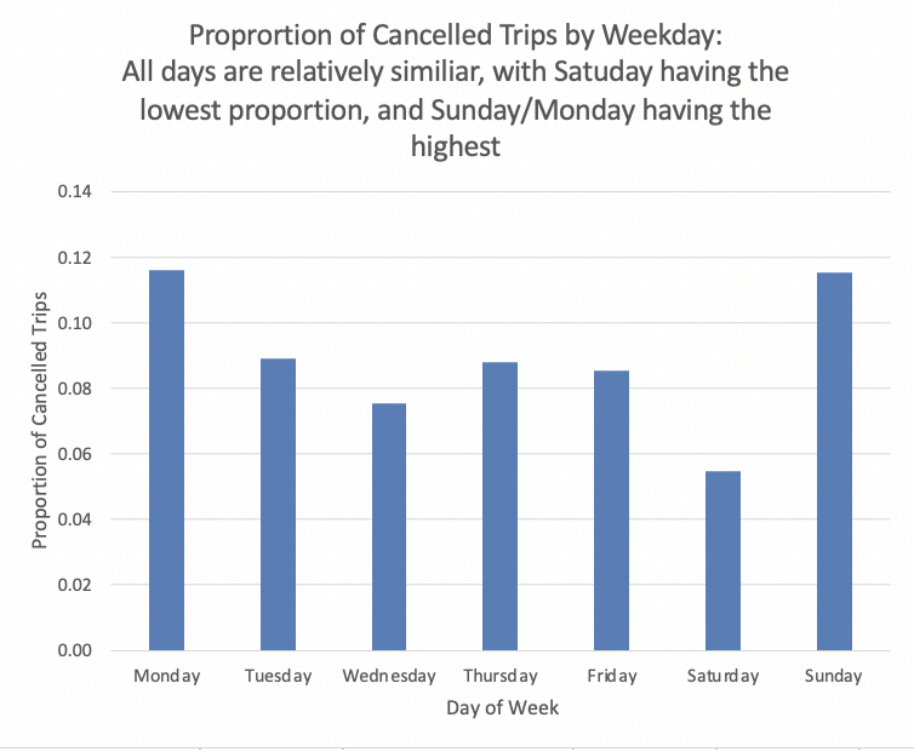
It is important to mention that the total proportion of canceled bookings among the entire sample set is 9%, meaning both mobile bookings and online bookings have a much higher cancellation rate than the total cancellation rate. Mobile bookings also have a much higher cancellation rate than online bookings. Overall, cancellations are more likely among both of these two booking methods when compared with the total cancellation rate and when comparing the two methods, mobile bookings have a higher cancellation rate. If I were IndoCabs, I would consider shutting down mobile bookings, or both mobile bookings and online bookings. That being said, there is a very high volume of bookings which are online bookings, so cutting it out of their business model would cause losses. It may make more sense to just cut mobile bookings.

Cancellation by Booking Type

	Count of Record	Sum of Car Cancellation	Average of Car Cancellation
Mobile Bookings	141	30	.21
Online Bookings	986	126	.13

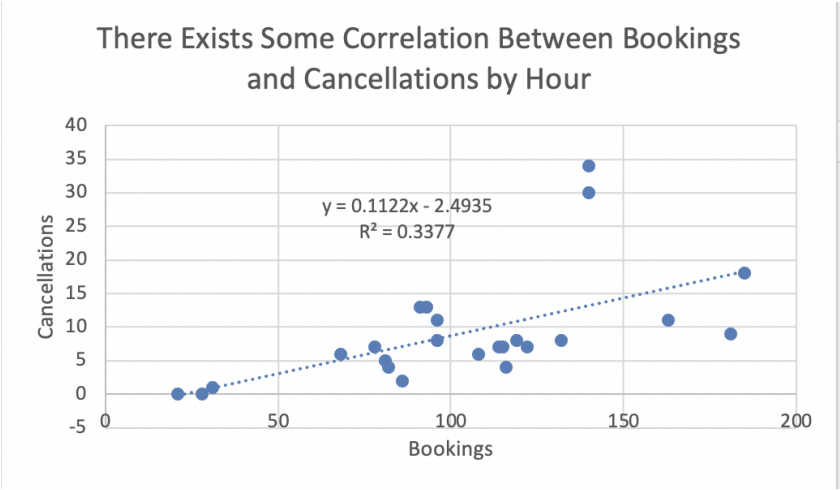
Significance of Weekday as an Indicator of Car Cancellation

The data is relatively consistent in terms of the amount of cancellations by weekday. There are no extreme differences as all proportions are under 12% of total bookings. The only foreseeable pattern is there is a higher proportion of cancellations on Sundays and Monday, and the smallest proportion of cancellations on Saturdays.



The Relationship Between Hourly Bookings and Hourly Cancellations

The R^2 value when considering the relationship between hourly bookings and hourly cancellations is about .34. This means that when looking at each individual hour of the day the number of bookings corresponds to the number of cancellations with a slight relationship. A R^2 value of .34 shows a correlation, but a minor correlation.

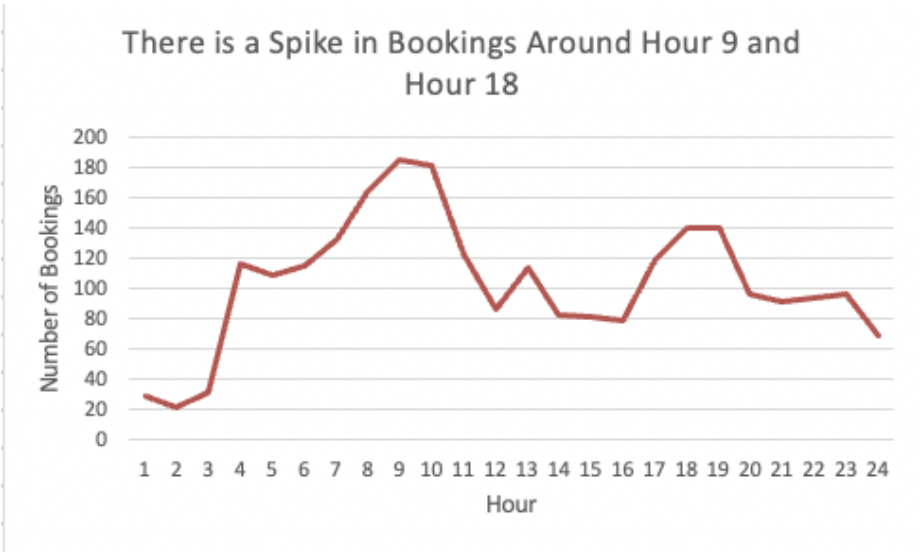


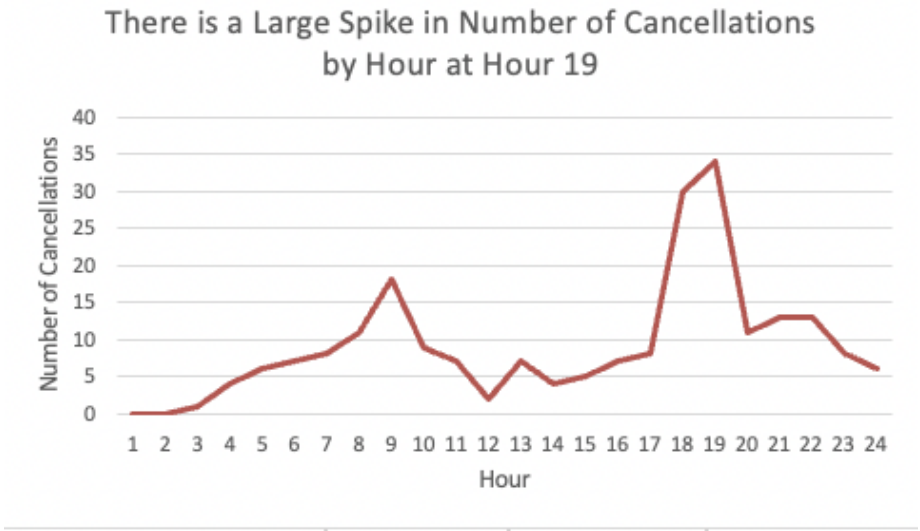
Analysis of the Number of Cancellations by Hour

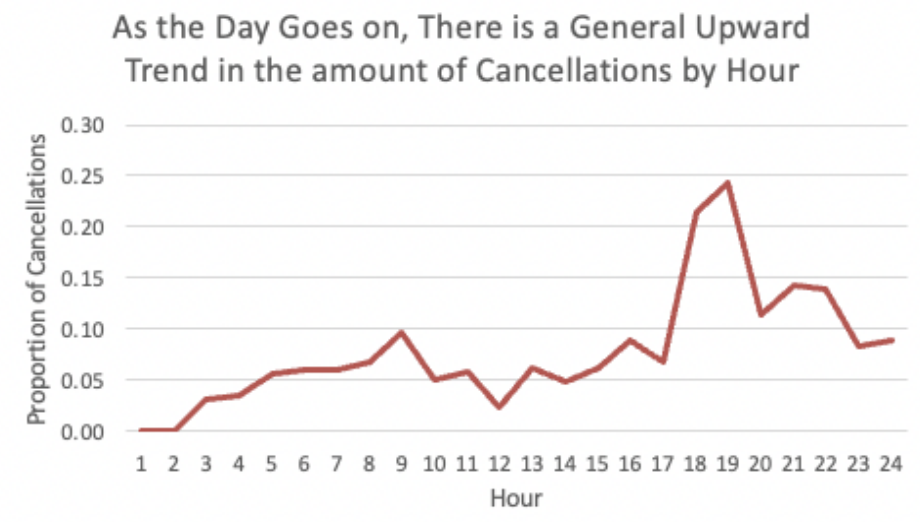
All three graphs pictured below seem to follow a general pattern. There are spikes around the hour 9 and the hour 18 mark. In a general sense, this seems predictable as when there are more bookings, there is a greater possibility for cancellations. The graph of the number of cancellations per hour and the proportion of cancellations per hour also seems particularly similar, which is understandable as they display the same variable, just in slightly different formats.

The most informative line chart would be the graph depicting the proportion of cancellations per hour, as the proportion is the most telling of how likely a cancellation is to happen. When considering solely the quantity of cancellations, there is no consideration of the total number of bookings within that hour.

When considering all that, it is most useful to compare the line charts depicting the total number of bookings per hour and the total proportion of cancellations per hour. There exists a somewhat subtle disconnect between these charts. It is clear that while there is a large spike in the number of bookings per hour between hours 7 and 11, there is only a small increase in the proportion of cancellations per hour. Between the hours of 17 and 19, there is a slightly smaller spike in the number of bookings, but there is a much larger spike in the proportion of total canceled bookings. This means that even when considering the higher amount of bookings between hour 17 and 19, there is an abnormally high proportion of cancellations.

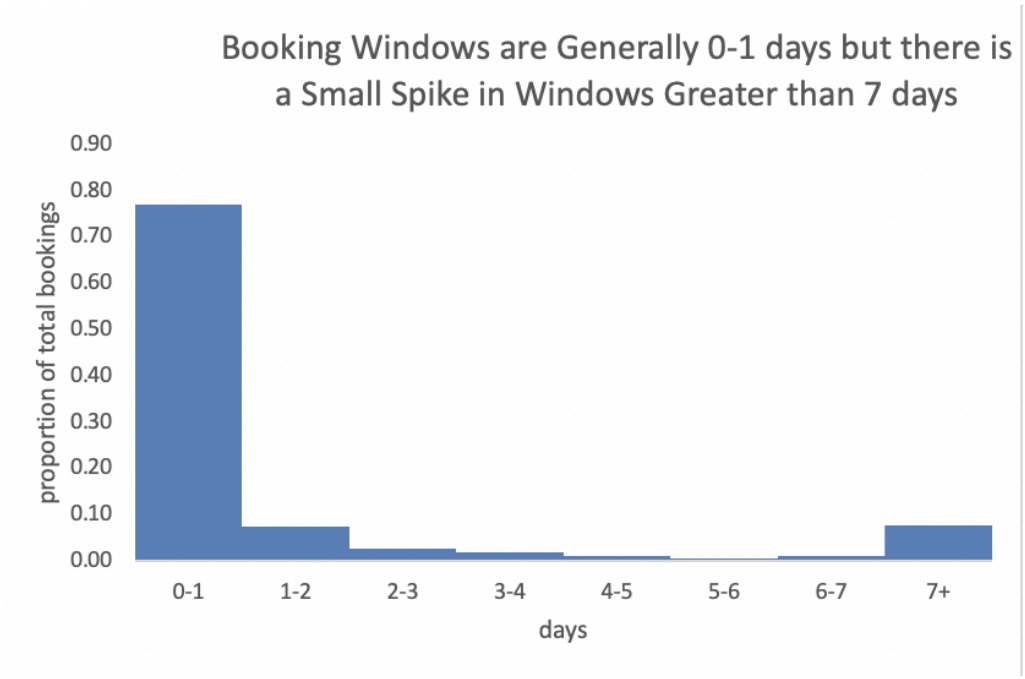

There is a Spike in Bookings Around Hour 9 and Hour 18
Number of Bookings
200
180
160
140
120
100
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Hour

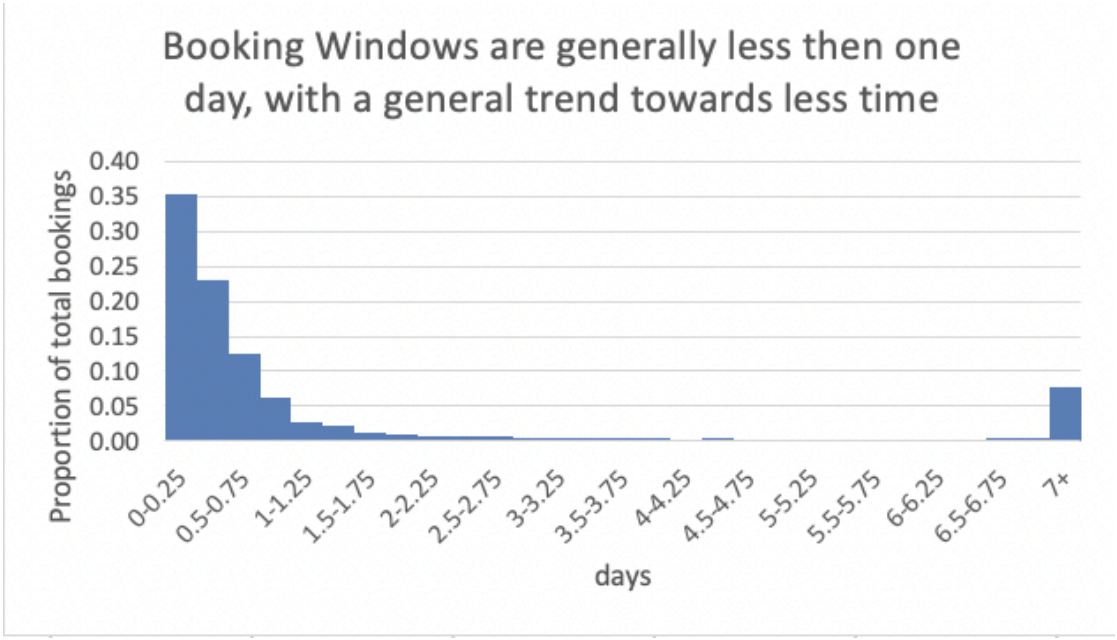

There is a Large Spike in Number of Cancellations by Hour at Hour 19
Number of Cancellations
40
35
30
25
20
15
10
5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Hour


As the Day Goes on, There is a General Upward Trend in the amount of Cancellations by Hour
Proportion of Cancellations
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Hour

Proportion of Total Bookings Organized by Booking Window

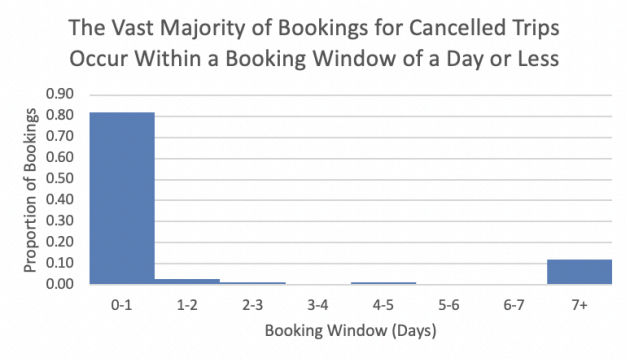
Below are two histograms comparing the proportion of total bookings to the length of the booking window, but the first has bins of 1 day, and the second has bins of .25 days (6 hours). Of these, the histogram with bins of .25 days is much more telling of the overall pattern, as we can clearly see that within the larger bin of 0-1 days, there are significant differences. As a general rule, as the number of days (or hours) of booking window decreases, there are a greater quantity of bookings. The opposite is also true, as when the booking window increases in time, the number of bookings generally decreases. This is important information for IndoCabs because they can alter their business model or pricing to better match their customers behavior.

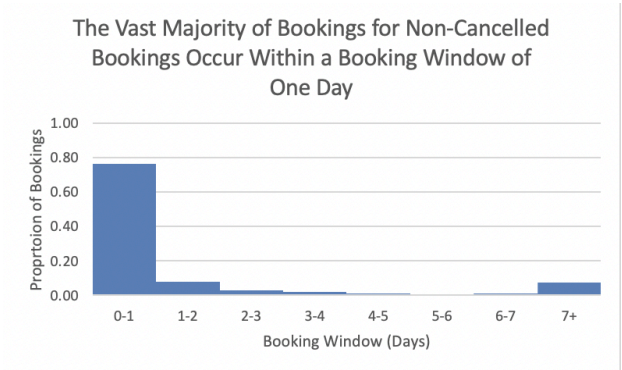


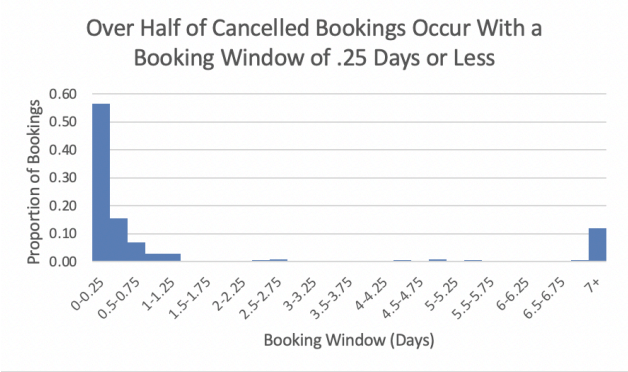


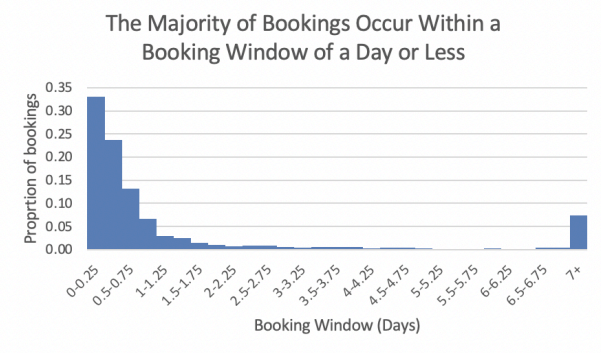
When considering these Histograms, there is a recurring pattern, which was mirrored in my previous analysis– that being that all the histograms have data heavily biased towards shorter booking windows, specifically within a booking window of one day. Trips are more likely to be booked, and canceled within a day of their actuation. Being that the graphs depicting bookings and the graphs depicting cancellations have this universal trend, these graphs are not very telling on how booking window affects cancellations.

When comparing the histograms with bins of 1 days, to the histograms with bins of .25 days, the histograms with bins of .25 days are much more useful. Being that there is such a high volume of data within a booking window of 0-1 days, it's useful to break up the data into smaller intervals. A pattern becomes apparent as well, that within day 0-1 less bookings occur as booking window increases, and the vast majority of bookings and cancellations occur within 0-.25 days.









General Recommendations and Conclusions

IndoCabs primary operational difficulties lie in the large sum of cancellations that occur by drivers. There are a variety of factors which increase the likelihood of a car cancellation, which can be supplemented by changes to the company.

There are a few recommendations I would make to the executives of Indocabs:

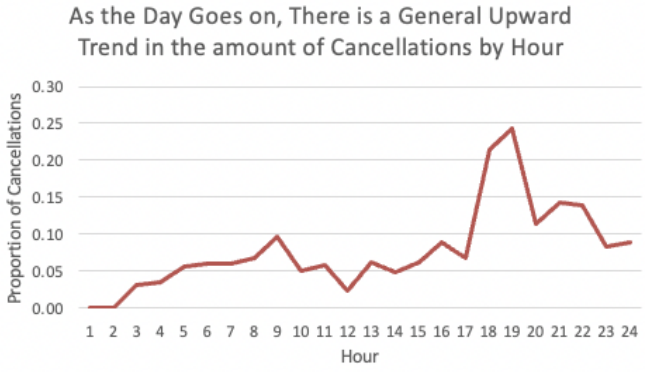
1. The first and most obvious trend within the data is that there is a much higher proportion of canceled rides when rides are booked on short notice. In order to supplement this problem, I would recommend that IndoCabs raise the prices for trips scheduled within one day of their respective start. This would incentivize customers to order rides in advance, meaning drivers will be less likely to cancel. Additionally, this strategy will generate greater revenue for the company as increasing prices is justified and will yield greater profit margins.
2. The second suggestion I would provide for IndoCabs is to completely eliminate the mobile bookings option. While it may be convenient for customers to have more venues to book rides, only 141 of 2486 total bookings were made via mobile booking. That's only about 6% of total bookings, and mobile bookings have a massive cancellation rate of 21%-- 12% higher than the total total proportion of cancellations (9%). By eliminating this option, total cancellations will decrease and with little damage to revenue for the company.
3. Third, IndoCabs would benefit from recruiting a larger quantity of drivers during periods of increased bookings. A larger work force would be beneficial to the company, as when drivers cancel, there is a larger quantity of alternative drivers to pick up the slack. According to the data, there is a higher quantity of bookings on Sundays and Mondays, and there is a lower quantity of bookings on Saturdays. Additionally, there is a higher quantity of bookings at hour 9 and hour 19. Both of these metrics are useful to the company in offering a supplement for high cancellation rates. While I understand the company does not directly manage employees, the company can encourage smaller cab companies to hire more employees– potentially through the institution of financial incentives for smaller companies. IndoCabs could also incentivize smaller companies to hire a greater quantity of workers at the aforementioned times through an equity investment.

4. Finally, I would suggest IndoCabs raise the prices on point to point urban travel. This makes up a massive percentage of overall trips (1955 of 2486 total bookings), so I would not recommend a dramatic change in pricing or customers may be deterred from using the service. That being said, 9.8% of point to point trips were canceled, which is the largest of all the travel types. A marginal increase in pricing may incentivize drivers not to cancel on rides and will simultaneously increase net revenue. If IndoCabs feels an increase to the overall price will be detrimental to their total sales, the company could offer a higher proportion of the income generated from rides to drivers and take a cut to their own profit, in an effort to cure the problem.

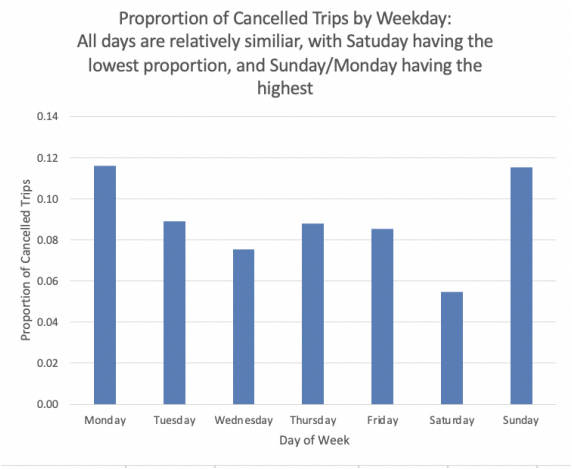
Hopefully these suggestions can be useful to IndoCabs, and they consider implementing them into their business model.

Elevator Charts

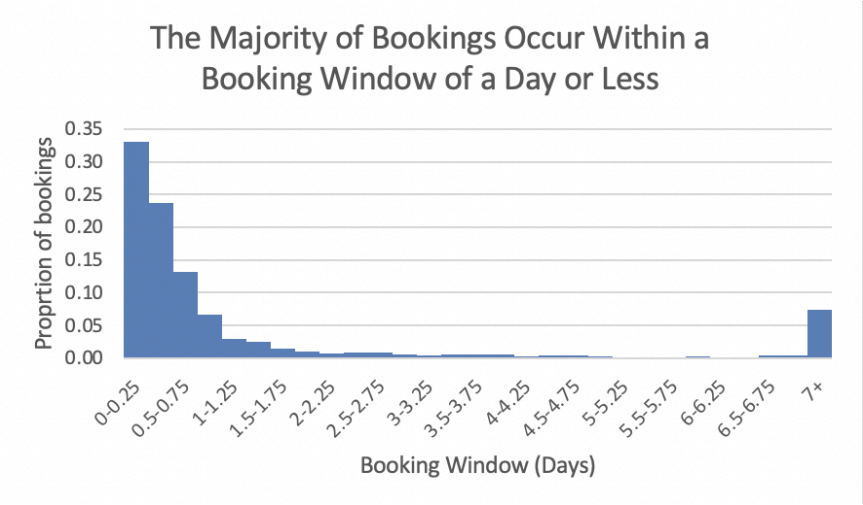
The following chart depicts the proportion of cancellations by hour for IndoCabs. This is useful to the company because during hours of increased cancellations, additional drivers can be recruited to offer a solution and alternative for customers who's drivers cancel.



The following chart depicts the proportion of canceled trips by weekday. In a similar fashion to the aforementioned chart depicting cancellations by hour, IndoCabs can clearly see which days of the week present the greatest difficulty in terms of cancellations. IndoCabs can then use this graph as a means to recruit more employees during periods of greater cancellations.



The following chart depicts the proportion of bookings by booking window for IndoCabs. This chart provides a clear overview of a very apparent pattern in IndoCabs customer base. The vast majority of bookings occur within very short notice of a future intended ride. This is useful to the company when considering pricing for rides.



Notes on Data Preparation

The data used in this analysis was provided by IndoCabs. The sample consisted of 2486 total bookings after removing irrelevant and erroneous data. Several types of data cleansing took place prior to my analysis.First, all duplicate observations were removed. Next some of the data far predated the other entries, so it was deemed irrelevant and deleted. Finally, some of the booking dates occurred after the trip start date so those entries were removed. The data set I was originally provided with had 2593 entries, so over 100 were removed– which I believe speaks to the extent of the data cleansing which occurred. I have no concerns with the data I used in this analysis and was efficiently able to organize and analyze it. Nothing ever seemed concerning or out of the ordinary through my continued investigation.