Ride Cancellation: Analyzing Patterns and Strategies for Improvement - Summary

Introduction

The dataset contains a total of 163851 rides requested through the application between 11th January and 17th January 2020. However, (39500) 24.1% of requested rides were cancelled.

	% distribution of ride status	Count
Dropped off	75.9%	124351
Passenger cancelled	24.1%	39500

Calculation of the repeat requests after cancellation

Of the cancelled rides, (23) 0.06% immediately placed requests for another ride. To achieve this result, comparison into the immediate re-appearance of a customer's Id in Column C after cancellation would need to be established:

Step 1: Insert a new column (column O) to be used as a "helper column" to store the previous values from column C.
Step 2: In the first row of values (in this case C3, enter the formula =OFFSET(C3,1,0). This formula copies the value from the next row in column C.
Step 3: Drag the formula in cell O2 down to fill the entire column with the corresponding values from the next row in column C.
Step 4: In another tab, use the formula "=QUERY('Trip Data'!A:N, "SELECT * WHERE C = O")".

This will compare column C with the helper column O to check for matching values.

Possible reasons for cancellations

Day of the week:

Although Sunday has the lowest cancellation rate, compared to the other days of the week, it also has the lowest overall request rate.

	Saturday	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday
Dropped off	88.5%	92.0%	84.0%	84.6%	83.8%	84.6%	82.6%

Passenger cancelled	11.5%	8.0%	16.0%	15.4%	16.2%	15.4%	17.4%
Total requests	**20620**	**12792**	**25321**	**25925**	**26662**	**25414**	**27117**

The highest cancellation rate, when placed on a 100% scale, is seen on Friday with 17% of requested rides resulting in cancellations.

Time of the day:

Cancellation of rides was most prominent during the hours of 4pm to 6pm.
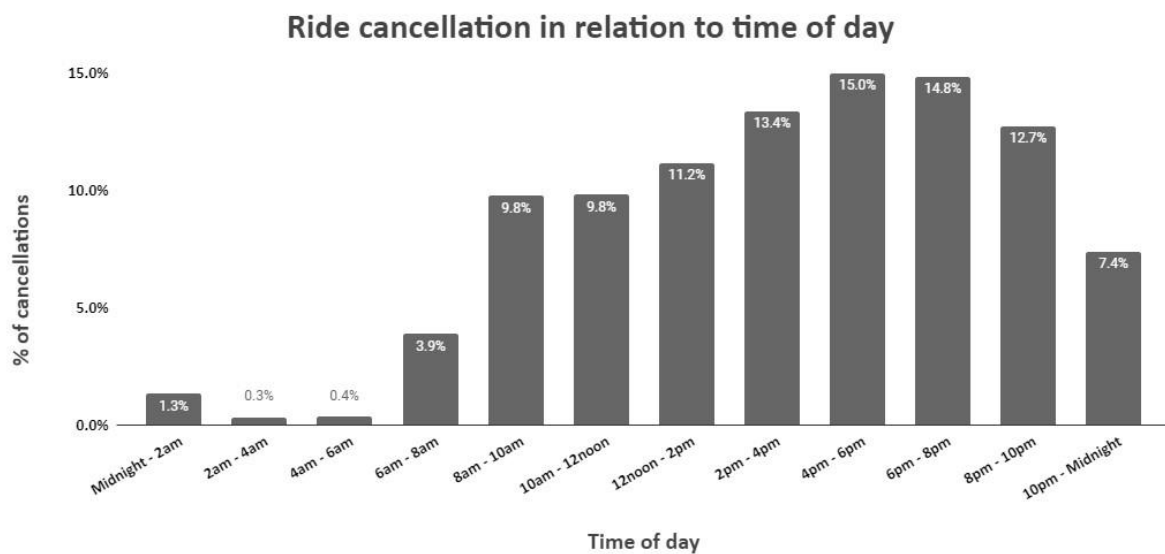
However, when looked at relative to the total number of ride requests placed at specific times of the day, the percentage of rides cancelled at different hours of the day does not fluctuate as much. This could be because the number of rides being cancelled at any given time increased as the number of requests for rides overall increased as well.

Time of day	No of cancelled rides	Total number of rides requested	% of cancelled rides
Midnight - 2am	523	1796	29.1%
2am - 4am	131	465	28.2%
4am - 6am	144	451	31.9%
6am - 8am	1536	6111	25.1%
8am - 10am	3859	16420	23.5%
10am - 12noon	3876	18193	21.3%
12 noon - 2pm	4407	19734	22.3%
2pm - 4pm	5288	20958	25.2%
4pm - 6pm	5921	24113	24.6%

6pm - 8pm	5865	23998	24.4%
8pm - 10pm	5036	20015	25.2%
10pm - Midnight	2914	11597	25.1%

Therefore, this graph more accurately shows the percentage of rides being cancelled in relation to the overall number of rides being requested during the different hours of the day.

Cancellation of rides was most prominent during the hours of 4am and 6am when 31.9% of the rides requested were cancelled by their passengers.

Ride cancellation by time of day
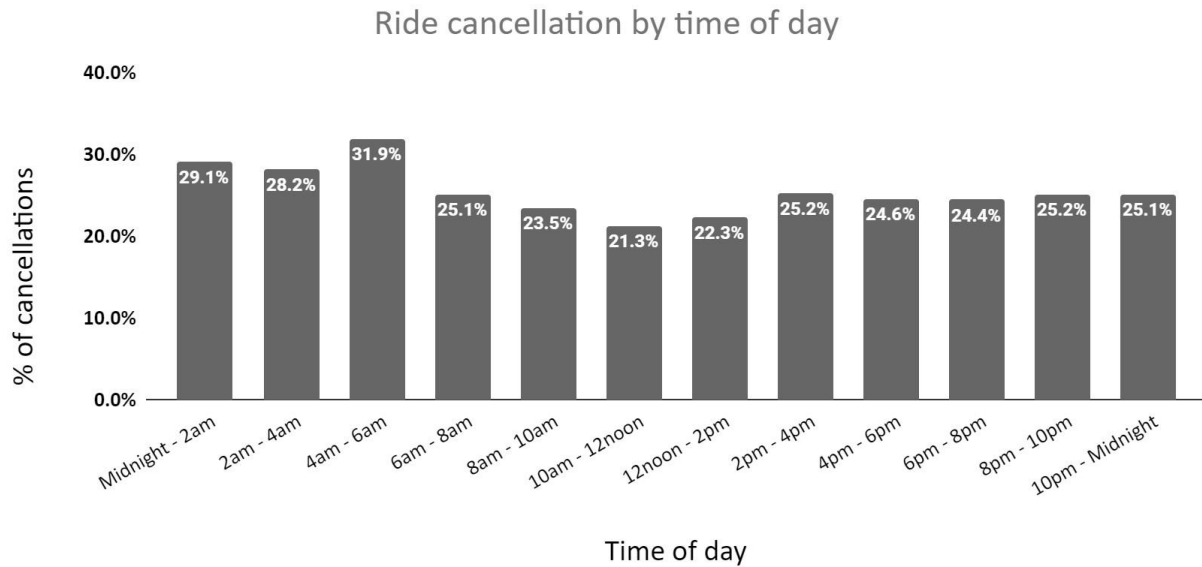
Duration of wait:

Most cancellations occurred with less than 5 minutes of passenger waiting time indicating that this might not be the reason for most cancellations. It is also unlikely that passengers would wait 1 hour or more for transportation via the app. This would potentially mark the 167 instances where this happened as an error that would require further assessment.

Waiting time	% of cancellations	Count of cancellations
Less than 5 minutes	83.11%	32830
Between 5 and 10 minutes	11.69%	4617
Between 10 and 20 minutes	4.04%	1596
Between 20 and 30 minutes	0.47%	187
Between 30 minutes and 1 hour	0.26%	103

Between 1 hour and 10 hours	0.21%	84
Between 10 hours and 24 hours	0.12%	47
More than 24 hours	0.09%	36
	100.00%	39500

Distance between customer and driver:

According to the data, the greater the distance between the passenger and the rider at the time of the request, the higher the rate of cancellation. Although there were instances of passengers cancelling with a distance of less than 200 metres between themselves and the riders, there was a steady increase in the cancellation rate with an increase in the distance between rider and passenger. There were also 6 instances of passengers not cancelling even with a distance of more than 1800 metres between themselves and the riders. However, these would be considered outliers. Almost half of the passengers (46%) who were assigned riders between 1600 and 1800 metres away cancelled their assigned rides.

Further calculations

Given more time, I would be interested in looking at these questions towards understanding ride cancellation:

1. Cancellations based on both the day of the week and the time of day. This could potentially provide more insight into the mindsets of the passengers.
2. Cancellations based on location of where the request was made vis-a-vis the distance to the destination and distance from rider. "Are passengers anticipating longer rides more prone to cancelling if they perceive that the designated rider is far from their pick up spot?" "Are passengers in particular areas more likely to cancel rides?"
3. Passenger waiting time in relation to individual riders. "Do particular riders have higher passenger waiting times than others? How does this affect the passenger cancellation rate?"
4. Cancellation based on passenger gender and time. "How do we market the app to the respective genders based on their timely needs?"
5. Cancellation and immediate pairing request. "Are passengers cancelling rides in preference of a rider currently present?"

Additional Data needs
1. Gender of the passengers
2. Pick up location of passengers (geo-location)
3. Passenger destination (geo-location)