

LYFT DATA SCIENCE ASSIGNMENT

DEVELOP AND
RECOMMEND
LYFT CANCELLATION
POLICY

- PD



BUSINESS GOAL OF THIS ASSIGNMENT:

- When rider cancels a ride, it creates bad experience for both drivers and the riders on the Lyft platform. It also has a financial impact.
- The Rider Cancels team is working to develop a new cancellation policy for Lyft, that will help in determining how and when riders should be charged a fee for cancelling their ride requests.
- Business goal of this assignment is to explore and analyze the rider cancellation data to develop and recommend a Cancellation Fee policy for Lyft.

UNDERSTANDING RIDE DATA:

- **ride_id** - Unique identifier for the ride request.
- **rider_id** - Unique identifier for the rider who requested the ride.
- **driver_id** - Unique identifier for the driver.
- **ride_type** - Type of ride requested (shared, normal).
- **upfront_fare** - Final fare quote provided to the Rider before the request was made. This is surfaced to the rider after they enter both an origin and destination in the Lyft app.
- **rider_paid_amount** - Total amount of money the rider paid to Lyft.
- **eta_to_rider_pre_match** - ETA (estimated time to arrival) shown to the rider immediately before the ride request was made.
- **requested_at_local** - Time when ride was requested.
- **eta_to_rider_post_match** - ETA shown to the rider immediately after the ride request was matched to a specific driver driver.
- **accepted_at_local** - Time when the driver accepted the ride request.
- **arrived_at_local** - Time when the driver arrived at the pickup location.
- **picked_up_at_local** - Time when the rider was picked up from the pickup location.
- **dropped_off_at_local** - Time when the rider was dropped off.
- **actual_time_to_arrival** - Time (in seconds) for the driver to reach the designated pickup location after being matched with the ride request.
- **cancellation_flag** - Boolean flag for whether the ride was cancelled.
- **rider_request_number** - Sequential count of ride requests for each rider.

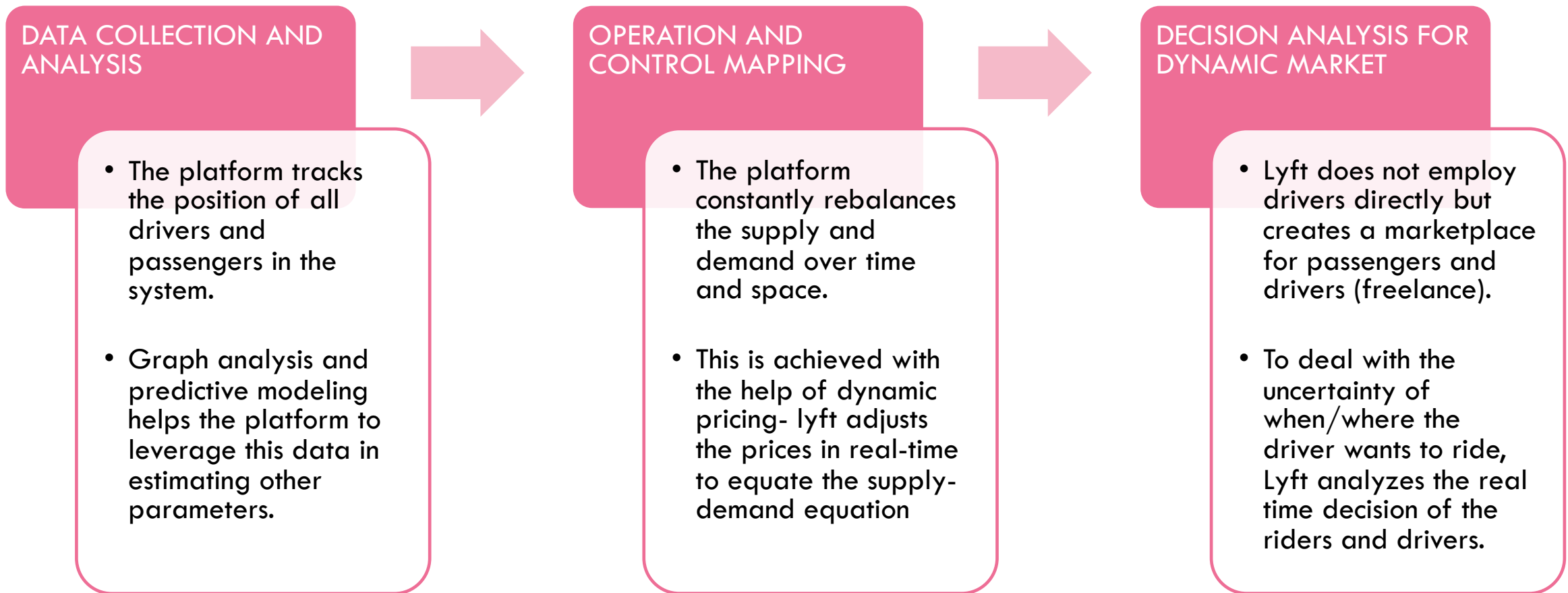


UNDERSTANDING EXPERIMENT DATA:

Lyft launched a randomized experiment to test the effect charging riders cancellation fees, of varying amounts, if they cancel a ride request. Riders were assigned to each variant and informed that the new cancel fee would apply to all future rides. This experiment was in effect for the entire duration of Ride Request Dataset.

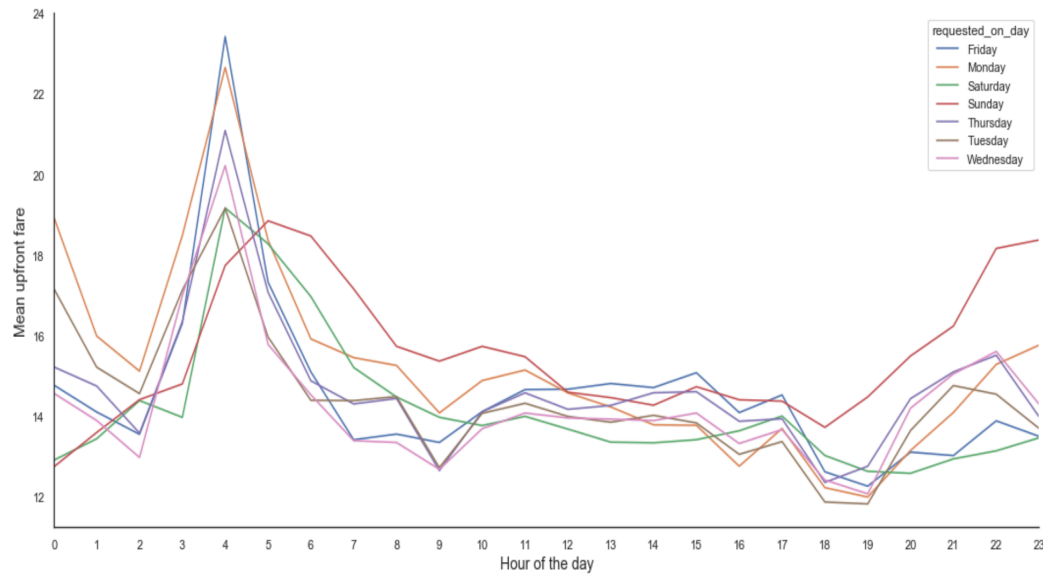
- **rider_id** - unique identifier for a Rider.
- **variant** - experiment group the Rider was in.
- **cancel_penalty** - cancellation penalty fee for the variant.

MY UNDERSTANDING OF LYFT'S MATCHING & DISPATCHING RIDES (DRIVERS AND RIDERS)



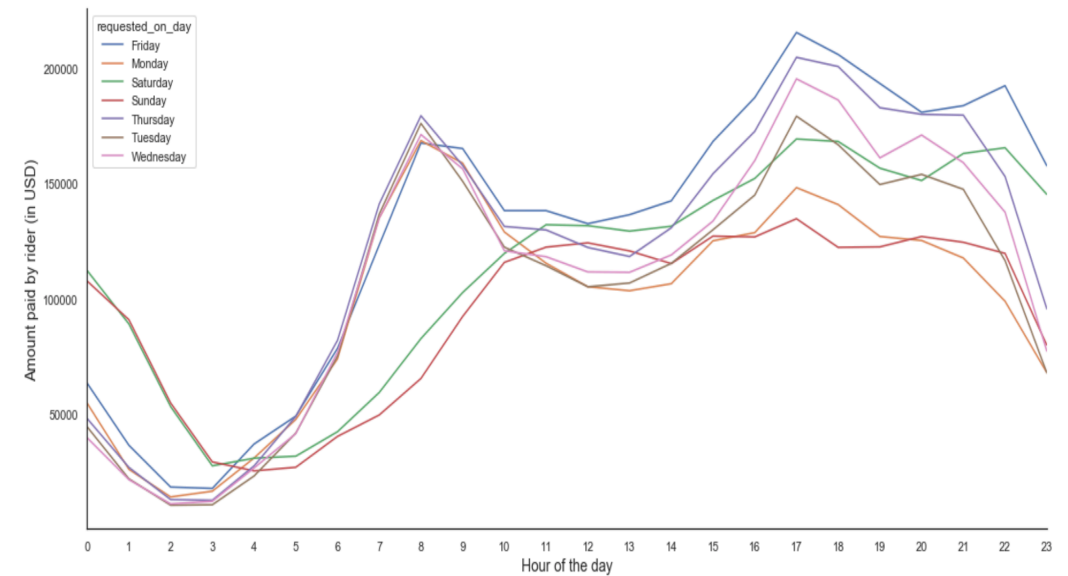
EXPLORATORY ANALYSIS ON THE GIVEN DATASET

Distribution of mean upfront fare shown to the rider per hour of the day



- Upfront fare shown to the rider is more in the early hours of the day.
- This makes sense as less riders are present in that timeframe. To meet the supply and demand equality, upfront fare during the early hours of the day is generally high.

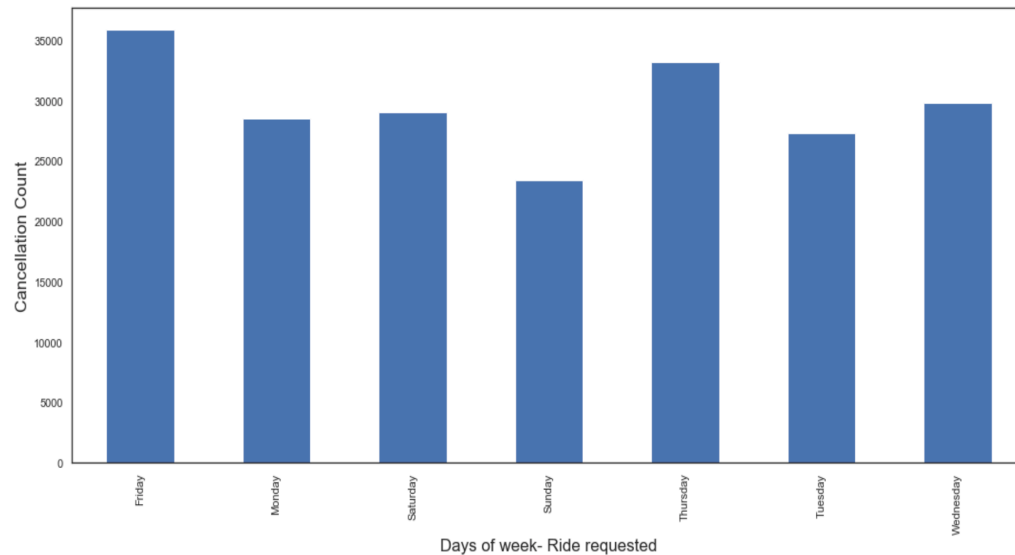
Distribution of total rider amount paid per hour of the day



- Revenue generation from both completed and cancelled rides is more for all the weekdays from 7:00 am to 9:00 am America/Los Angeles.
- The revenue is also high from 16th to 19th hour. We also see a sudden increase in the revenue from 11:00 am on weekends.

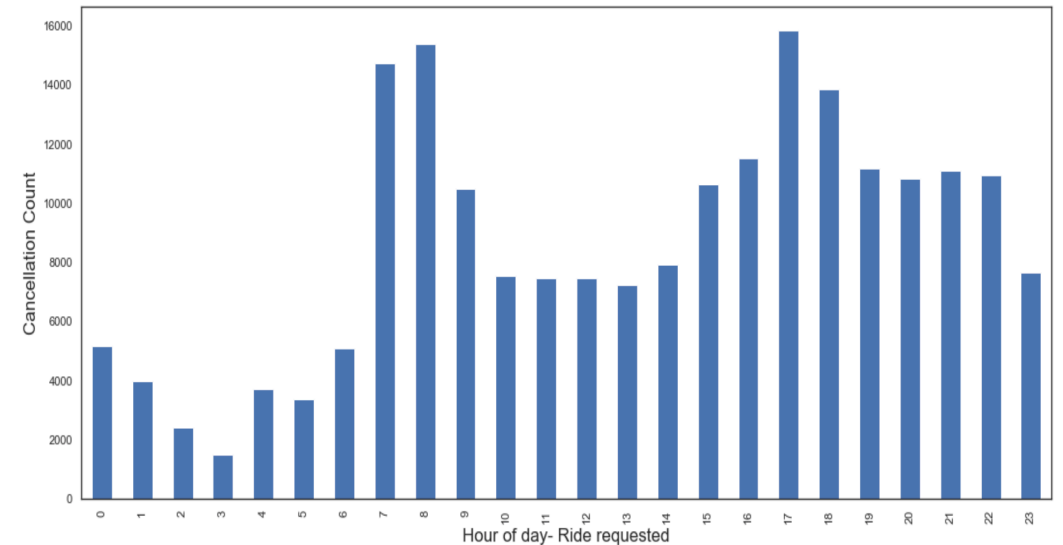
EXPLORATORY ANALYSIS ON THE GIVEN DATASET

Distribution of number of rides cancelled after the request per day of the week



- Friday sees the highest number of cancellations 35,869 (around 17.32% of the total cancellations), over the period from April 14, 2019 to May 26, 2019.

Distribution of number of rides cancelled after the request per hour of the day



- Over the given timeframe, hours that see significant cancellations are 7, 8, 17 and 18.



KEY INSIGHTS FROM THE GIVEN DATASET

- **May 3, 2019, the 123rd** day of the year 2019 saw the highest number of ride cancellations (count = 6,548).
- The cancellation rate at 9th hour of the day experiences a drop down of 31.79% from the 8th hour of the day. It keeps decreasing till 13th hour but sees an 34.20% surge from 14th to 15th hour, and then keeps increasing till 22nd hour.

For SHARED rides:

- Total Revenue from CANCELLED rides : **\$ 2,21,365.00**
- Total Revenue from COMPLETED rides : **\$ 3,829,143.02**

For STANDARD rides:

- Total Revenue from CANCELLED rides : **\$ 3,15,211.00**
- Total Revenue from COMPLETED rides : **\$ 14,154,846.75**

KEY INSIGHTS FROM THE GIVEN DATASET

Cancellation count:

- No. of shared rides cancelled : 83,355
- No. of standard rides cancelled : 123,695

Total Revenue:

- Cancelled rides : \$ **5,36,576**
- Completed rides : \$ **1,79,83,989**
- Of the total revenue over from April 14, 2019 to May 26, 2019, **cancelled rides constituted nearly 3 % (~2.98%)** of it.

- From the perspective of company's revenue growth, **having a cancellation policy, will be a good move.**
- For instance, cancellation policy during the given timeframe resulted in half-million dollars revenue alone.



LYFT'S CURRENT CANCELLATION POLICY

Lyft charges the rider, a fee for cancelling a ride when both of the following occur:

- The cancellation window has passed after a driver accepted the ride request.
- The driver is on time to arrive within 5 minutes of the original estimated arrival time.

For Shared rides, Lyft may charge a cancellation fee should the rider cancel any time after a driver accepts the ride request.

Repeat cancellations:

- Rider will be charged a \$2 fee if the rider cancels three or more rides of any type in a 15-minute timeframe, even if it's within the cancellation window.

In most cities, the amount charged is \$5 for Shared and Standard rides.

FACTORS TO CONSIDER IN DETERMINING THE CANCELLATION POLICY

An optimized Logistic Regression model with an accuracy score of 0.98 and precision of 0.95 resulted in the following features to be the best predictor of the ride cancellation indicator.

- **upfront_fare**
- **eta_to_rider_post_match**

The rider may cancel the ride after seeing the fare that the platform is showing, before the request.

After requesting, the platform will match a driver to the rider and then, if the estimated time of arrival after the match is higher, the rider may cancel the ride because of time constraint.

MY RECOMMENDATIONS FOR LYFT'S CANCELLATION POLICY

After taking the results from exploratory analysis and that of the modeling into consideration, I have the following recommendations:

- Lyft can have a higher cancellation fee (\$7) during the peak hours on weekdays (Especially Friday). From the analysis, the peak hours are from 7:00 am to 9:00 am and 4:00 pm to 8:00 pm. For the rest of the time range, cancellation penalty can be minimized to \$3 (Treatment-1 group).
- Lyft can impose a low cancellation penalty (\$3), if the rider cancels the request before the match, however ETA post match plays an important role in cancellation. Once the platform matches the rider with the driver, Lyft can impose a higher cancellation fee (\$5 during normal hours) or (\$7 during peak hours)
- Upfront fare is also one of the reason the rider cancels the ride. For the given dataset, the highest difference (for completed rides) between the upfront fare and amount paid by the rider was \$2.24. Lyft can devise a strategy to identify the most accurate fare quote before the request.



BUSINESS IMPACT OF MY RECOMMENDED CANCELLATION POLICY

- The revenue from cancellation resulted in nearly 3% of the total revenue from the time period of April 14, 2019 to May 26, 2019.
- After increasing the cancellation penalty to \$7 in peak hours, there will be a definite increase in Lyft's revenue.
- Ride cancellation creates a negative experience for both drivers and riders. Lyft acts as a marketplace for freelance drivers and having a bad experience with rides may cause high driver churn.
- During the non-peak hours, having a low cancellation fee (\$3), will not have a major impact on the platform's cancelled ride share of income.



LIMITATIONS

- Given dataset is only for the time period of April 14, 2019 to May 26, 2019. Findings like higher ride requests and cancellation on Friday might be only valid for this time period.
- Taking only cancellation into consideration without understanding driver's behavior, driver's cancellation and factors like distance to destination, time taken by the platform to match the rider with the driver etc., will only provide a narrow perspective.
- Working cross-functionally with different teams that handle driver's data, team that estimates the quote for upfront fare and the team that manages ETA pre and post match will lead to a more accurate estimation and modeling.
- Having both rider's and driver's data will help better understand the factors that lead to ride cancellation.



FUTURE WORK

Moving forward with the data analysis, it would be interesting to understand:

- The cohort analysis of the driver, to gain a better pattern of the driver's behavioral pattern and what factors results in driver's ride cancellation decision.
- Get to look around different features in the driver's data set to try and optimize the model that identifies driver's churn rate and different factors for predicting the driver lifetime value.
- Lastly, with so much happening around, and world coming on standstill with the economy shutting down, It would be interesting to see the situation analysis of driver and rider's current behavior with the ride sharing platform, and different factors that are currently contributing to the rate of ride cancellation.

THANK YOU

