Industry Overview

The New York City Taxi and Limousine Commission (TLC) is responsible for the regulation and licensing of the New York City's Medallion Taxi cabs. There used to be over 1,000,000 trips completed by the New York City yellow cabs, all done by around 200,000 TLC licenses. These taxi cabs are the only vehicles that are legally allowed to pick up passengers through prearranged services or street-hailing. The legal right to work is set through the purchase and licensing of the medallions, which must be present on the taxi. These medallions are auctioned by TLC and available for transfer through the open market by licensed brokers (Salam, 2021).

At one point, the Yellow Cab business value outpaced gold and the Dow Jones Industrial Index. However, this was all before 2014, the introduction of Uber and other companies such as Lyft. As shown in Figure 1, in 2013 Yellow Cab averaged more than 400,000 trips per day, however with the introduction of Uber in 2014, this number started decreasing steadily while the trips taken through Uber and Lyft increased rapidly. Within 4 years, the Yellow Cab taxi trips were around 250,000 while Uber's average trips skyrocketed to around 800,000. While there was a rapid downfall in trips during Covid-19, the return of taxis and ride-hailing apps is starting to rise again, however, the market share is mostly dominated by Uber and Lyft.

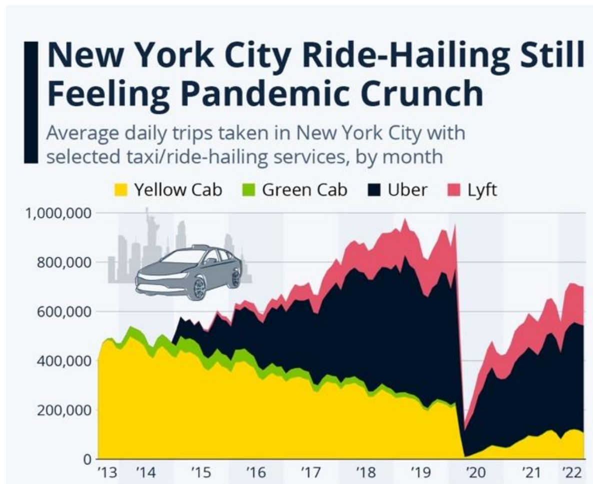


Figure 1 - Average Daily Trips (source: NYC TLC)

Current State

Decreasing Market Share

Since the emergence of ride-hailing apps such as Uber and Lyft in 2015, New York City yellow cabs have seen a steady decline in their market share. More specifically, yellow cabs have had their market share dwindle from nearly 80% in 2015 to 15% as of September 2022, while ride-hailing apps have gone on to claim 85% of the total NYC cab market.

NYC Taxi Market Share

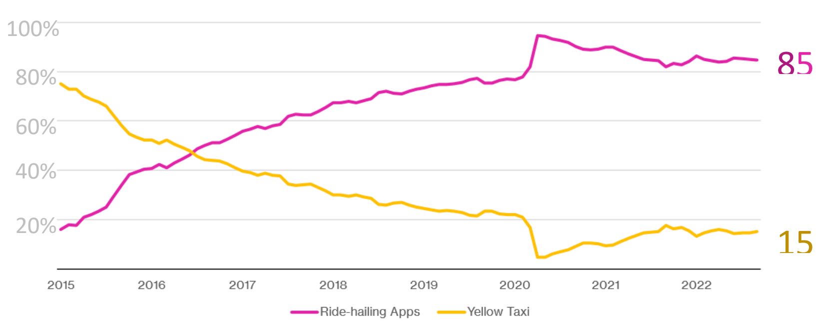


Figure 2 - NYC Taxi Market Share (Data from NYC TLC)

Pricing Structure

Currently, Yellow Cab follows a simple fixed pricing structure. A customer will be charged an initial cost of $2.50 plus an additional charge of $0.50 per 1/5 of a mile and two surcharges totaling $0.80 (Taxi Fare, 2022). The $0.80 in surcharges account for both the $0.30 improvement surcharge, and the $0.50 MTA state surcharge. In contrast, the ride-hailing apps are using a dynamic pricing model which has been proven successful as 85% of the market share has shifted from Yellow Taxi to Uber and Lyft. A single static pricing model does not consider the relationship between supply and demand. After our review, was determined that using a more efficient pricing model will help the company maximize the revenue.

CHARGES CONSIDERED	COST
Initial Charge	$2.50
Travel Charge	$0.50 per 1/5 of a mile
MTA State Surcharge	$0.50
Improvement Surcharge	$0.30

Prices to optimize

Table 1 - Yellow Cab Pricing Structure

PRO Opportunity

How can Yellow Cabs use pricing and revenue optimization to reclaim market share?

With dynamics changing in the transportation industry, Uber and Lyft fares are rising and wait times have become increasingly long. Uber and Lyft have acknowledged that their prices have increased amid rising ride demand and have experienced difficulties recruiting drivers. They have complete control over their dynamic pricing algorithm, which is known to be tweaked often.

On the other hand, yellow cabs adhere to city-regulated meters that take both distance and time into consideration and don't have surge pricing. Yellow Cab is far from a full recovery and this could be an ideal opportunity for them to come back into business by adopting a new pricing strategy for rides to optimize revenue and reduce the prevalence of ride-hailing alternatives.

Our Approach

To improve competitiveness and increase market share, it is necessary to identify the optimal pricing mechanism for both the **initial charge** and the **travel charge** that adopts the equilibrium of dynamic supply and demand in the taxi industry.

Specifically, the objective of introducing this new pricing system needs to satisfy:

- **Flexibility**: Prices are reflected and automatically adjusted from the non-stationary daily demand and supply in the market.
- **Competitiveness**: Ensure the cab vacancy rate is below an optimal level and demand is not slashed due to pricing changes.
- **Reliability**: Analysis is based on actual taxi data and effectively describes the relationships of demands influenced by all parameters.

PRO WORKFLOW & ARCHITECTURE

Historical Trip Data

Data on previous Yellow Cab trips must be collected to understand the market and the factors that may influence the model.

Aggregate/ Summarize

Collected data will then be summarized to identify demands, costs, and distances.

Dynamic Pricing Model

Optimize initial charge and travel charge prices using Excel Solver.

Implement Optimal Prices

Utilize the output of model and implement new prices to capture more revenue.

Data Collection & Processing

Data used to complete the optimization task was collected from the NYC Taxi & Limousine Commission (TLC) website. The TLC reports trip data for each cab type and month, making the data open-source and a valuable resource for growth opportunities such as the one presented in this paper.

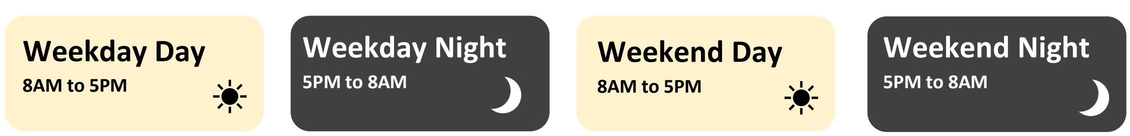
For the purposes of this project, yellow cab data was collected for June, July, and August. This was done with the intention of identifying how Yellow Cabs can improve their performance in the upcoming summer season. Key fields that were included in the dataset are displayed below in Table 2.

FIELD	DESCRIPTION	EXAMPLE
Pickup_datetime	Date and time meter was engaged.	2022-07-07 8:59:03 PM
Dropoff_datetime	Date and time meter was disengaged	2022-07-07 9:47:02 PM
Passenger_count	Number of passengers	3
PULocationID	TLC zone in which meter was engaged	6
DOLocationIID	TLC zone in which meter was disengaged	6
Payment_type	Numeric code for payment type (1=credit card, 2=cash, etc.)	1
Trip_distance	Distance travelled in miles	11.9
Fare_amount	The travel fare calculated by meter	7.50
MTA_tax	$0.50 MTA tax	0.50
Improvement_surcharge	$0.30 improvement surcharge	0.30

Table 2- Features in NYC TLC Provided Data

In preparation for the modeling phase, the pickup datetime field was used to segment the data into four categories based on the day and hour of the trip: weekday daytime, weekday evening, weekend daytime, and weekend evening. Identifying whether the trip took place on a weekday or weekend was achieved by using the weekday function in python’s datetime library. Identifying whether the trip was during the daytime or evening was achieved by simply labelling based on the timeframes set at the beginning of the project, in particular the daytime ranging from 8AM to 4:59PM and the evening ranging from 5PM to 7:59AM. Lastly, these two identifying features were combined to assign a unique segment to the trip such as “weekday_day”.

FOUR SEGMENTS



Minor cleaning was also carried out, as the dataset contained a small number of trips from both May and September due to technical errors. Trips with a distance equal to zero were also removed, as these were deemed as a quick stop-and-start event and would simply provide noise during the modeling process. Lastly, to narrow the scope of our approach it was decided that only trips within 10 miles would be modeled, meaning the results of this report provide insights on how to improve pricing and revenue for short trips only.

After creating segmentation fields and cleaning the data, it was then grouped by each segment and day to generate summary demand data. The output of this process was a dataset for each segment containing the daily number of trips (i.e., demand), the average distance of those trips, and the average number of passengers. It is with these summary datasets that the modelling process described in the following section utilized.

Modelling

Based on the data as described above, a model was developed to understand the demand behavior for multiple segments. Fares for Yellow Cab are made up of several fixed charges (e.g. surcharges) and variable charges (e.g. travel charge) which were determined based on factors such as jurisdiction and distance. Once the relationship between price and demand was determined through regression analysis, the intercept of the function and coefficient for price was used to optimize revenue in each segment.

The dependent variable, demand, was provided as the total number of trips completed by Yellow Cab daily for the selected time-period. The independent variable, or explanatory feature, was the total average fare of each trip that was taken. Since there were only two variables to consider, a linear regression model was selected as it was able to provide a good balance between simplicity and interpretability. Once the final dataset was available, Excel offered an excellent end-to-end model pipeline with its built-in regression data analysis tool pack.

The initial model was a linear regression using daily cost of fare and demand starting from June 2022 until August 2022. As seen in the below, the resulting R-square of 0.01 indicated that there was not an explainable correlation between the two variables, or it could not be represented as a linear relationship (see Figure 3 below). Without a reasonable level of certainty on the coefficients, it was not possible to obtain more insight from the data. Further analysis was required to get a better understanding of the data, specifically through segmentation.

Regression Statistics	
Multiple R	0.100216685
R Square	0.010043384
Adjusted R Square	0.004721037
Standard Error	8628.784326
Observations	188

Figure 3 - Statistical Summary of Regression for Unsegmented Data

The hypothesis was that consumer demand would be different based on whether a customer needed to make a trip on a weekday or weekend, and if the trip was during the day or evening. Thus, the dataset was split into four segments based on date and time of the trip. The R-square results can be seen below for each bivariate analysis. A significant improvement in model performance can be observed for segments during the day on weekends and weekdays (see Figure 4 below).

Regression Statistics (Weekday Day)	
Multiple R	0.731125678
R Square	0.534544758
Adjusted R Square	0.52727202
Standard Error	3529.072928
Observations	66

Regression Statistics(Weekend Day)	
Multiple R	0.694676723
R Square	0.48257575
Adjusted R Square	0.461016406
Standard Error	3426.119251
Observations	26

Regression Statistics (Weekday Evening)	
Multiple R	0.383640944
R Square	0.147180374
Adjusted R Square	0.133855067
Standard Error	6435.706029
Observations	66

Regression Statistic (Weekend Evening)	
Multiple R	0.12545969
R Square	0.015740134
Adjusted R Square	-0.020713935
Standard Error	11311.60164
Observations	29

Figure 4 - Statistical Summary of Regression for Segmented Data

Using the respective intercepts and coefficients of each demand function:

$$\boldsymbol{Demand = Intercept + Fare\ Coefficient * Fare}$$

The total revenue was optimized by maximizing the revenue function:

$$\boldsymbol{Revenue = \sum_i a + b * Fare_i}$$

And changing the decision variables, initial charge and travel charge, where the fare for each day of *i* equals:

$$\boldsymbol{Fare_i = Initial\ Charge + (Travel\ Charge * \ Distance_i) + Surcharge}$$

Subject to total demand less than or equal to the average trips per day per segment, which was estimated to be 51,026.

Results & Analysis

Compared to the original pricing structure, the model implements segmentation of pricing between weekday day, weekday evening, weekend day and weekend evening. As indicated by Figure 5 below, the original price structure has an initial charge of $2.50 and a travel charge of $0.50 per 1/5 of a mile. However, to capture more demand and in turn increase revenue, the weekday day initial price is increased by 10 cents to $2.60 while the travel charge is decreased by 25 cents to $0.25 per 1/5 of a mile. The initial price is further increased by 25 cents to $2.75 to take advantage of the demand during the weekday evening while travel charge no longer exists. The weekend day mirrors the weekday day pricing of $2.60 initial charge and 25 cents per 1/5 of a mile. Lastly, based on the optimal pricing results the weekend evening price should continue to use the original initial charge of $2.50 and implement a slight decrease to the travel charge bringing it down to $0.25 per 1/5 of a mile.

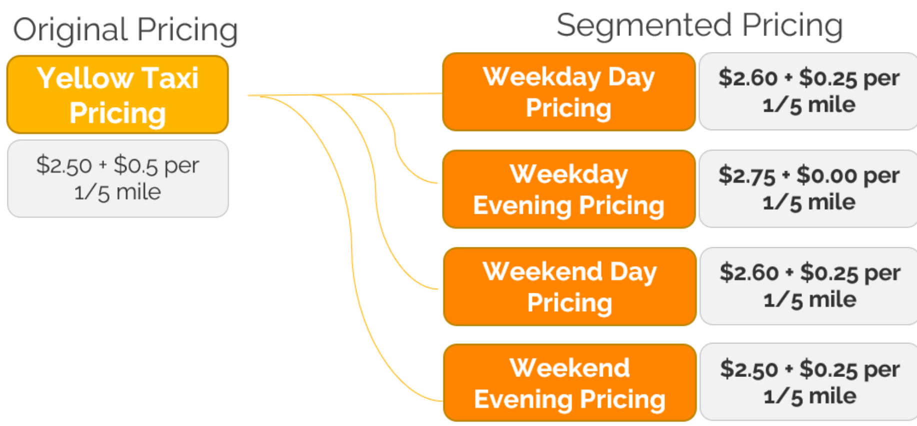


Figure 5 - Segmented Pricing Results

Through the adoption of the model, there is an overall increase in revenue of roughly $2 million over the three-month period analyzed. The weekday daytime revenue jumps from $10.95 million to $11.22 million, an increase of 2.4%. There is a more significant increase in revenue during the weekday evening where the revenue propels from $11.58 million to $11.94 million, an increment of 3.1%. This, however, does not compare to the change in revenue during the weekend daytime if the model is implemented; there is an increase in revenue of 30% where the pre-optimized revenue surges from $3.4 million to $4.4 million, indicating that this segment make benefit the most of pricing changes. Furthermore, the weekend evening post-optimized revenue also increases by 6.1%, increasing from $4.8 million to $5 million. From these results, it is evident that Yellow Cab truly can benefit from segmented dynamic pricing.

Segment	Revenue Pre-Optimization	Revenue Post-Optimization	Change
Weekday Daytime	$10.95M	$11.22M	+$0.26M (+2.4%)
Weekday Evening	$11.58M	$11.94M	+$0.36M (+3.1%)
Weekend Daytime	$3.4M	$4.4M	+$1M (+30%)
Weekend Evening	$4.8M	$5M	+0.29M (+6.1%)

Figure 6 - Revenue Changes as a Result of Pricing Decisions

The results above in conjunction with deeper segmentation and dynamic application of initial and travel charges per segment can capture an even greater demand and can maximize revenue beyond current thresholds, making yellow cab a major player in the transportation mobility industry.

The Way Forward

Moving forward, the model can be enhanced further by collecting ridership data through surveys to understand the willingness-to-pay across the four segments and even those that have not been explored in this analysis. With this survey data, it is expected that improvements will be seen both the demand predictions and revenue forecasts.

Furthermore, additional segmentation involving weather, destination zones, and seasonality can also be integrated to improve the model's accuracy and ensure it is properly accounting for influential pricing factors.

Finally, it is also worth exploring the effect of other pricing models such as choice pricing models to better consider competitors, discount pricing models to attract more riders, and loyalty pricing models to reward riders who have been supportive of Yellow Cab since the emergence of ride-hailing apps.

Over the past five years, Yellow Cab has struggled to compete let alone regain market share, and to reassert its hallmark presence in NYC it must be more dynamic – starting with their prices. With the results of this analysis, it is clear that even limited dynamic pricing can produce more demand and revenue for Yellow Cab, however in order to capture those increases Yellow Cab must do something they have haven't done in recent years – take action.

References

Taxi Fare. (2022). Retrieved from nyc: https://www.nyc.gov/site/tlc/passengers/taxi-fare.page#:~:text=%242.50%20initial%20charge.,Dutchess%2C%20Orange%20or%20Putnam%20Counties

TLC Trip Record Data. (2022). Retrieved from nyc: https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page

Harding, B. (2021, June 3rd). Where Have All of New York City's Yellow Cabs Gone? Taxi Numbers Still Low, Data Shows. Retrieved from NBC New York: City:https://www.nbcnewyork.com/news/local/where-have-all-of-new-york-citys-yellow-cabs-gone-taxi-numbers-still-low-data-shows/3087721/

Salam, E. (2021, October 2nd). 'They stole from us': the New York taxi drivers mired in debt over medallions. Retrieved from The Guardian: https://www.theguardian.com/us-news/2021/oct/02/new-york-city-taxi-medallion-drivers-debt