

# ORIE 4741

## Project Proposal

Team members: Aman jain (aj644@cornell.edu), Vineel Yellapantula (vy45@cornell.edu)

### Question Identified

“How should we price the X days subscription program for home-office-home commutes for a retail user”

### Why is it important

In the context of India, Daily commuting to the office and back home is a challenging task. Even though widespread presence of ondemand ride hailing services like Uber, commuting in peak hours still remains an unsolved problem. This is primarily owing to below par reliability of these services during office hours.

[Moveinsync](#) (mentioned as Company henceforth) has been working in the domain of employee transportation since 2009 and have recently started their offering to retail customers ([News](#)). This customer proposition provides lots of opportunities in terms of making data intelligent systems to run efficiently at scale.

This project tries to find a balance between service quality and cost of service - which is at the core of service.

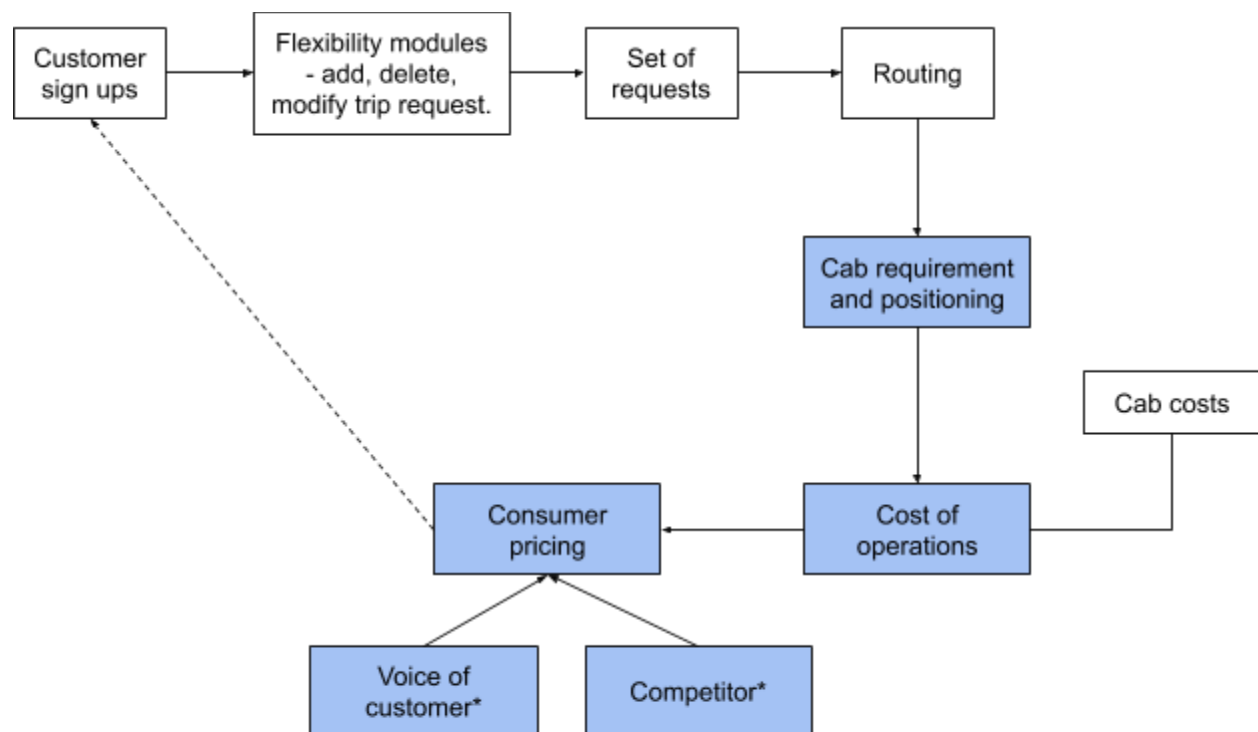
### Datasets

Details of these datasets is provided in the appendix.

No.	Name	Objective	Source
1	User attribute	To obtain user specific features.	Company Data
2	Trip attribute	Main input data for cab requirement algorithm.	Company Data

3	Cabs attributes	To understand constraints of available supply	Company Data
4	Cab Contract attributes	Main input data for operational costs.	Company Data
5	Foursquare APIs	To obtain any location specific feature.	Company Data
6	Uber pricing APIs	Competitive pricing - input to overall pricing.	Company Data

## How to use the dataset to answer the question



We will be focussing on the blue colored boxes of the process.

### Steps:

- Understanding data (EDA) (expected to be more than 100,000 rows)
  - Locations\*time densities.
  - Speeds - Google map API to draw out average speeds.
- Preprocessing data
  - Setting up required features (e.g. geo hashes etc).
- Consuming route data.

- Understanding costs
  - Design algorithm to understand cab requirements.
    - Variation of clustering and sequencing algorithms.
- Building cost models (Focus only on following two costs)
  - Vehicle utilization etc.
    - Based on utilization in the above part.
    - Suitable pricing model with vehicle.
  - Cost of flexibility to customers.
- Understanding feedback of current customers on pricing. (expected: 500 survey responses)
- Understanding and selecting pricing philosophies.
  - Premium
  - Penetration
  - Economic
  - Behavioural economics application.
- Suggesting a pricing strategy for retail users.

## Appendix

<b>Question Identified</b>	<b>1</b>
<b>Why is it important</b>	<b>1</b>
<b>How to use the dataset to answer the question</b>	<b>2</b>
<b>Appendix</b>	<b>3</b>
Objective	3
Introduction	4
Motivation	4
Scope & Assumptions	4
Data description	4
Foursquare APIs	7
Uber pricing APIs	7

## Objective

This exercise/project is being done to fulfill the course requirement of ORIE 4741: Learning with big messy data, at Cornell. In this view the key objective of the project is to apply concepts and tools learnt during this course. This project will be a mix of Data analysis and algorithm development - Formulating important questions, and show how to use big messy data to answer (or try to answer) the questions. We will be working with real life data of an organization working in the transportation domain in India. The final deliverables will be a white paper suitable to submission to NeurIPS, ICML, KDD and present to subject matter experts.

## Introduction

### Motivation

Apart from fulfilling course requirements, we are highly motivated by working on a real life problem and hopefully see some of our work put in action.

### Scope & Assumptions

- We will not be focusing on routing.
- We will be limiting analysis to one market only.
- Voice of customers and competitive pricing are subjective to time availability.

### Data description

#### User attributes

No.	Column name	Datatype	Description	Mandatory?
1	User key	int	To identify users. Can be masked	Yes
2	Pick up lat long	float		Yes
3	Drop lat long	float		Yes
4	Pick up time	datetime		Yes
5	Drop time			Yes
6	Gender			Yes

7	Plan sign up time			Yes
8	is_change_request			Yes
9	% of no show			Yes

\*masked

### Trips attributes

No.	Column name	Datatype	Description	Mandatory?
1	Route id	int	To identify routes.	Yes
2	Route start time	datetime		Yes
3	Route end time	datetime		Yes
4	Trip type	Class	Pick, Drop or mix	Yes
5	Sequenced Set of stop lat long	List of lat long pairs		Yes
6	Sequenced Set of stop type	List of class	Pick, Drop	Yes
7	Sequenced Set of employee key	List of ids		Yes
8	Route Priority	Class	Low, medium, high	Optional
9	Start_Buffer	int	Buffer to add to start time	Optional
10	End_Buffer	int	Buffer to add to end time	Optional
11	Travel time variability	Class	Low, medium, high	Optional
12	Pricing zone of the trip			Optional

### Cabs attributes

No.	Column name	Datatype	Description	Mandatory?
1	Cab id	int	Can be masked.	Yes
2	Seats	int	4 seats	Yes
3	Possible contract ids	List of contract id		Yes

### Cab Contract attributes

No.	Column name	Datatype	Description	Mandatory?
1	Contract id	int	Can be masked.	Yes
2	Type of contract	Class		Yes
3	Min kms a day	int		Optional
4	Min kms a week	int		Optional
5	Min kms a month	int		Optional
6	Min daily guarantee	int		Optional
7	Min weekly guarantee	int		Optional
8	Min monthly guarantee	int		Optional
9	Per km rate over min	int		Optional
10	Per trip cost	int		Optional
11	Flat payment	int		Optional

## Foursquare APIs

API name	Doc link	Fields of interest	Objective*
Venue details	<a href="https://developer.foursquare.com/docs/api/venues/details">https://developer.foursquare.com/docs/api/venues/details</a>	Popular, rating, created_At, Like, dislikes.	5
Indicate venue incorrect	<a href="https://developer.foursquare.com/docs/api/venues/flag">https://developer.foursquare.com/docs/api/venues/flag</a>	Venue ID and problem	5
Get venue recommendation	<a href="https://developer.foursquare.com/docs/api/venues/explore">https://developer.foursquare.com/docs/api/venues/explore</a>	Venue ID, location, name and categories	3,4

\*objectives mentioned in “Data source” table above.

## Uber pricing APIs

<https://developer.uber.com/docs/riders/references/api/v1.2/estimates-price-get>