Taxi Trips in Chicago

Agastya Jha
Ashwin Singh
Pravek Sharma
Saatvik Dutt
Sameep Dahal

Goal

Enhance cab booking experiences in the city of Chicago.

How?

Dataset: Taxi trips dataset collected by Business Affairs and Consumer Protection, released by Department of Affairs and Consumer Protection.

Model: Poisson process, with taxi trips as arrivals.

About the Dataset

- Date of the trip
- Start Time and End Time of each trip
- Distance travelled during the trip
- Total Fare of the trip
- Pickup location
- Drop Off location

Preprocessing the Data

- Weekdays
- 7AM to 10AM
- 5PM to 8PM
- No vacation months (June, July, August, December)

Uniformity of data helps establish identical nature of random variables, thus easing the analysis.

Modelling as a Poisson Process

“Start Trip” and “End Trip” are modelled as a Poisson process.

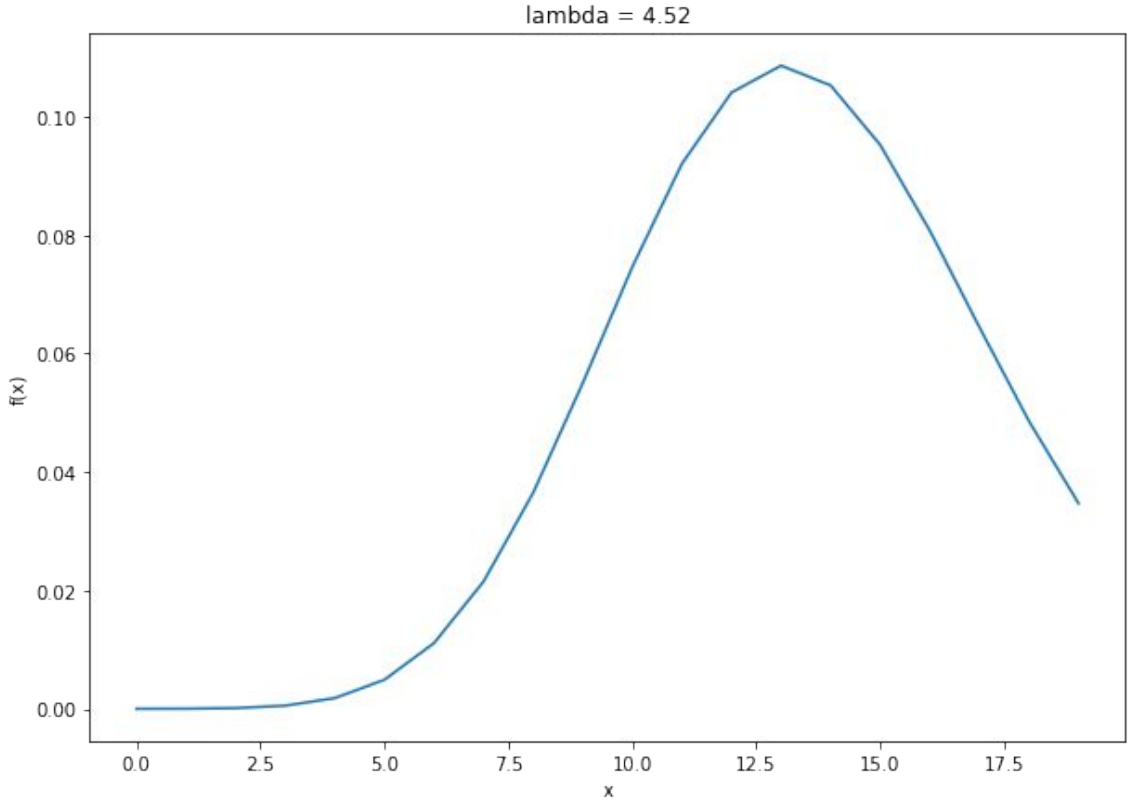
Counting process:

- $N(0) = 0$
- $N(t) \in \{0,1,2,...\}$, for all $t \in [0,\infty)$
- If $s \leq t$ then $N(s) \leq N(t)$

Independent Increments: The number of arrivals in disjoint slots, are independent of each other.

Distribution of # of taxis

Poisson(λt)



How do we plan to enhance the cab booking experience?

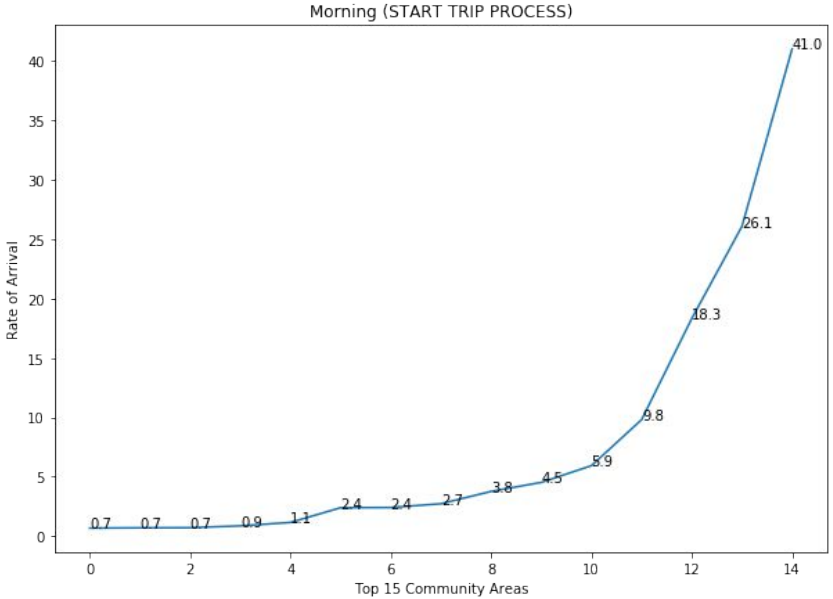
Users

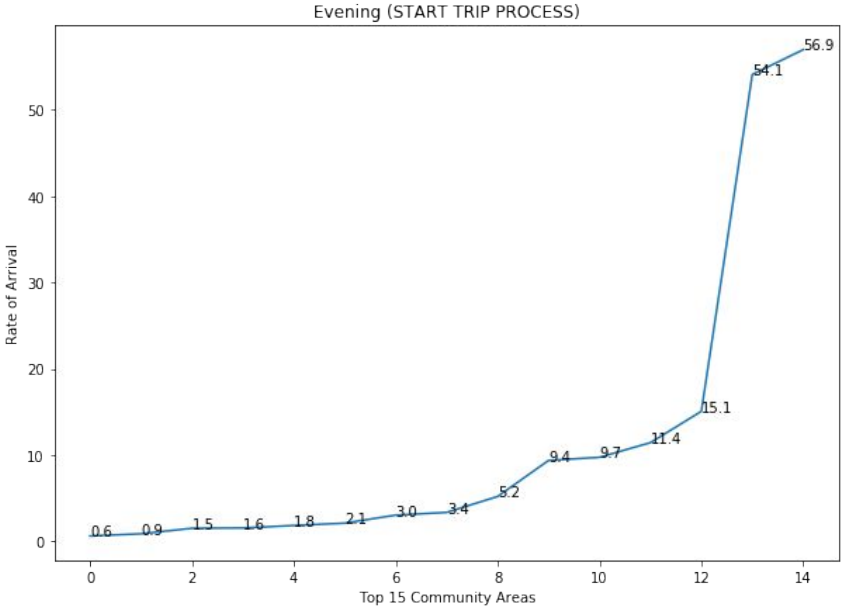
- Providing them an expected wait time according to the community area they’re in while booking a cab.
- Suggesting them when to pool depending upon their start location and destination.

Drivers

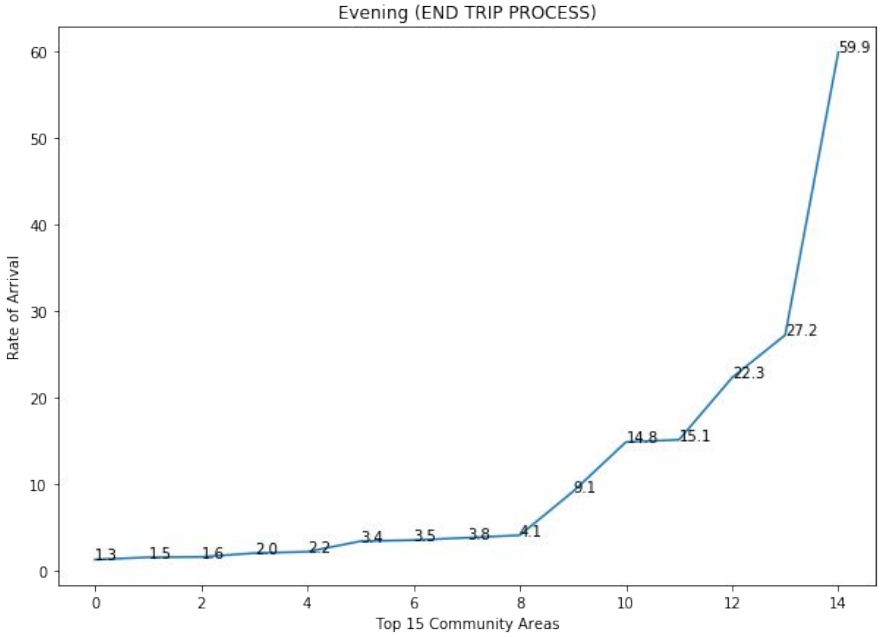
- Showing them areas where surge or demand is more during rush hours.
- Suggesting them pickup areas based on their preferences in terms of trip distance.

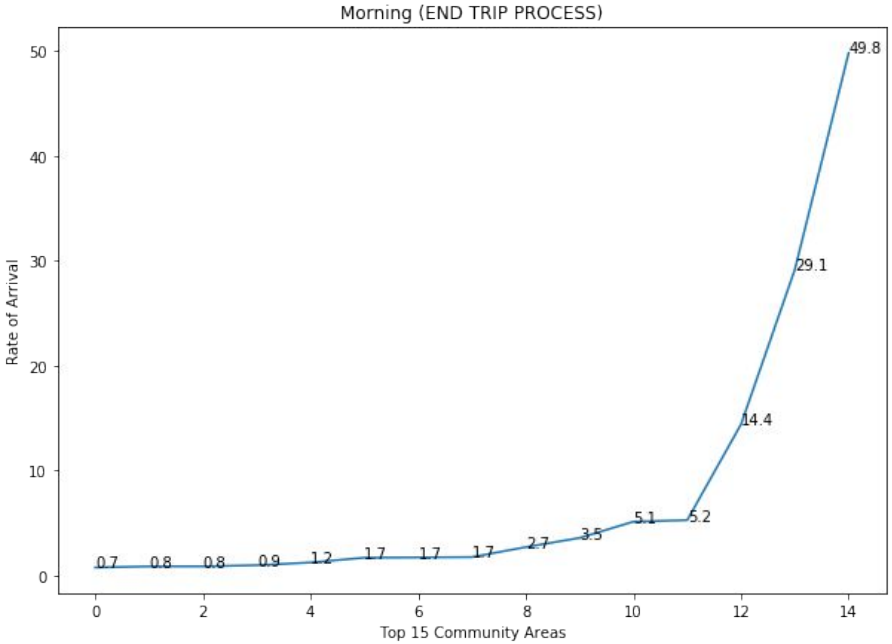
What the modelled data looks like? **(Start Trip)**





What the modelled data looks like? **(End Trip)**





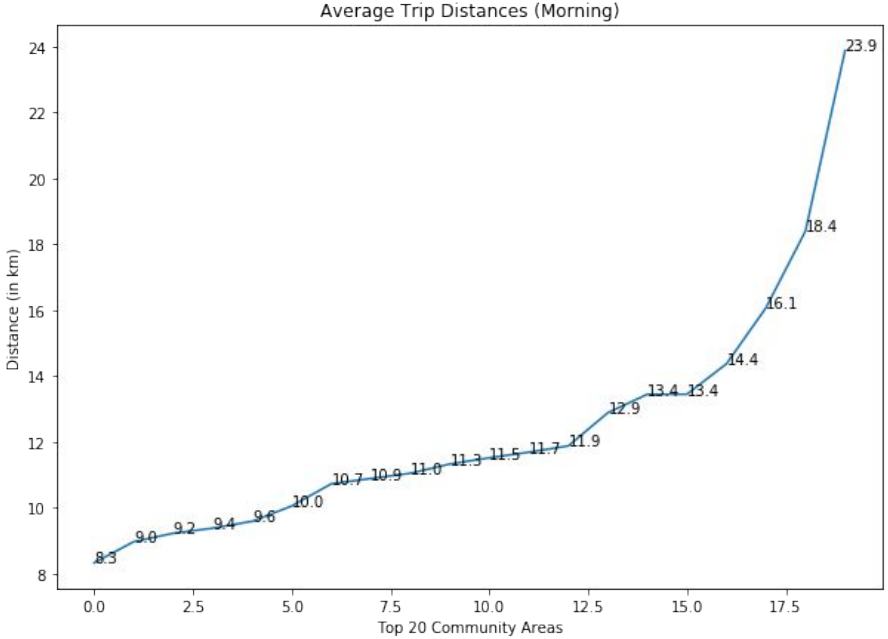
Expected Wait Time for Users

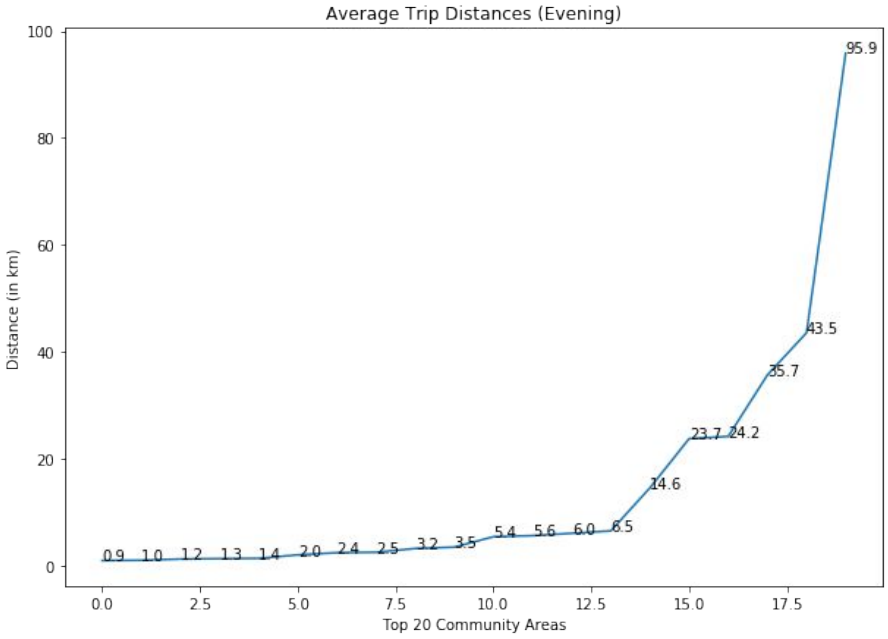
A user waiting for a taxi is the case of a *random incident* within the poisson process with an expected waiting time equal to the expected value of the interarrival time.

$$E[X_1] = 1/\lambda$$

It does not depend upon arrivals that occured in the past.

What the modelled data looks like? **(Distances)**

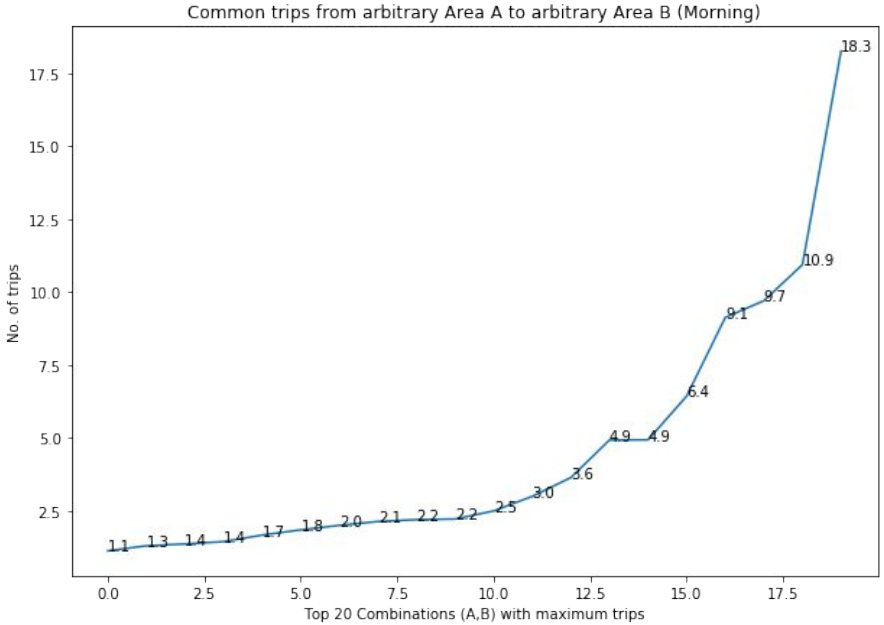


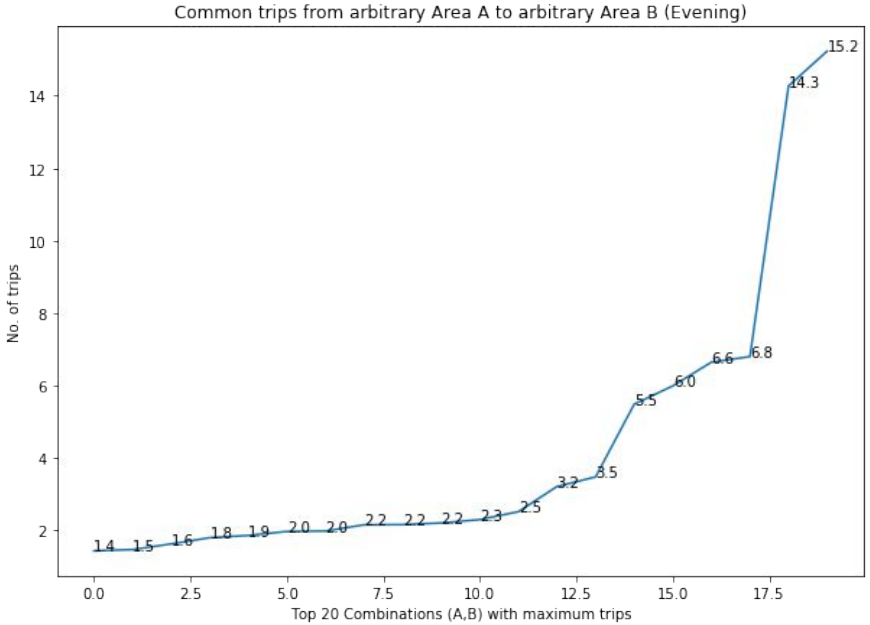


Distance Preferences for Drivers

Estimating the distances travelled by taxi trips while starting at different pickup areas gives an idea of whether the trip from a random area is more likely to be a longer trip or a shorter trip in terms of distance.

What the modelled data looks like? **(Trips from A → B)**





Suggesting the users to pool taxi trips

By estimating the number of trips across all pairs of community areas in Chicago, we can comment upon the likelihood of the expected destination and suggest the user to pool his ride in case the likelihood is good enough.

Thank You