MTA Targeted Marketing Applications

Problem Statement

- Leveraging Public Transportation Ridership Patterns to attract private investments and/or revenue from corporate sources.

- The main goal is for corporations to patron certain stations of the existing in exchange for running blanket ads.

Huh? What do you guys do again?

- We are a business facing project which provides a platform for businesses to run spatially targeted ads
- Provide rebates on ticket fare to the MTA user by leveraging in-coach ad rev
- Provide a spatially targeted billboard marketing solution

Who is it for?

- The primary persona of a user are for the Sales and Marketing division of medium to large corporations.

- The secondary consumer is the average Joe using the subway lines.

But, how?

Techie Jargon

- Mine rider data
- Develop robust live streaming solution
- Explore the idea of Local hub
- Make use of spatial nature of data

What we really mean

- Which are the most used stations?
- Make some Vibranium grade platform!
- Where do people get/switch rides?
- Which neighbours are really good at spreading the word?

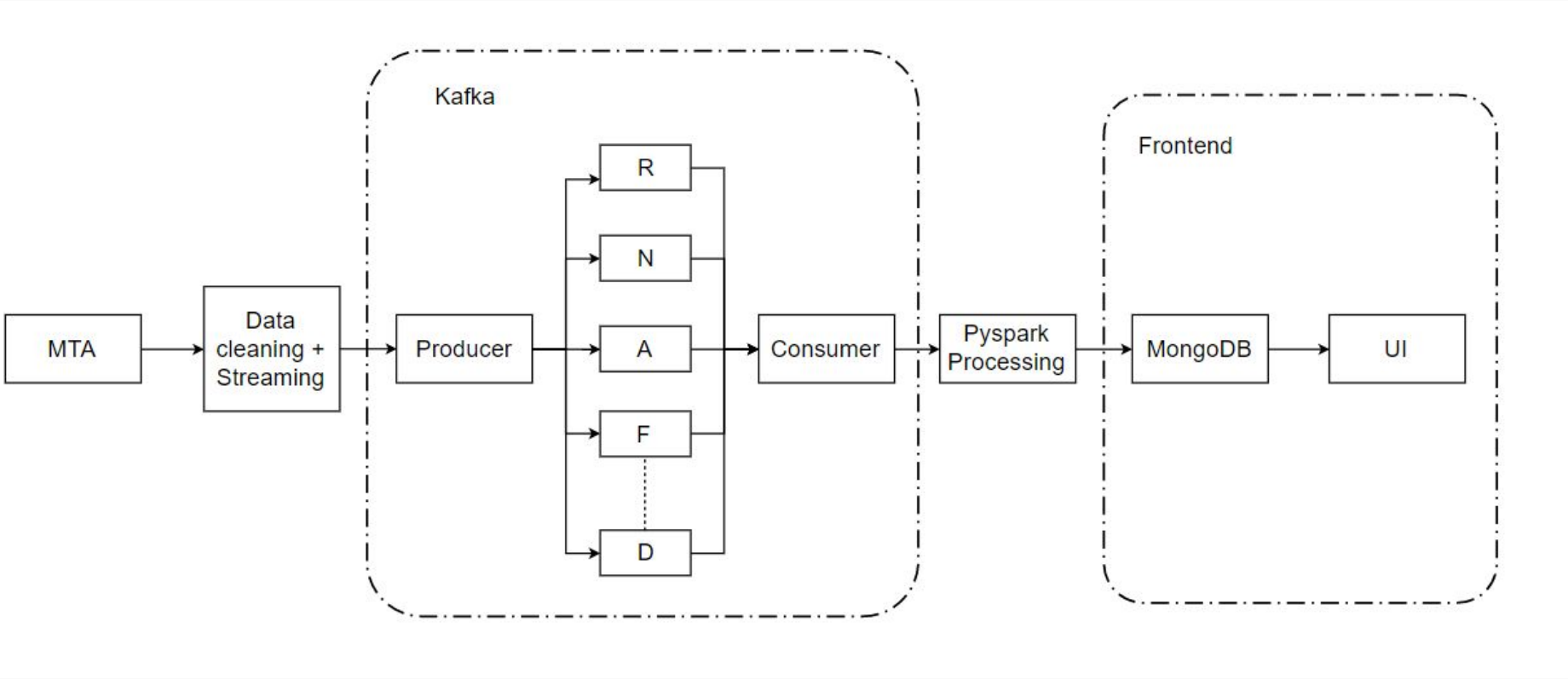
Can you be more specific?

We have Identified two main use cases:

- Active Marketing (in-coach): Make the tickets cheaper while sharing who helped with the fare.
- Passive Marketing (Bill board): The good old poster/billboard marketing space

Let’s have a look under the hood

Architecture Diagram



*Currently Processing only R F and D lines

Data Streaming/ Kafka

- MTA provides number of entry/exits for each turnstile of each station.
- File for each day contains the above data for every four hours and is dumped every night on the MTA website.
- We use this data to create a Kafka stream of station data creating a topic for each line.

STATION	DATE	TIME	ENTRIES	EXITS
JAY ST-METROTEC	04/13/2019	4	514.0	473.0
DEKALB AV	04/13/2019	4	4082.0	1850.0
ATL AV-BARCLAY	04/13/2019	4	7139.0	8115.0

Analysis: Pyspark Streaming

Weight of each station for subsidy -

$$stationWeight_i = \left(\sum_{i=1}^{n} \#entries \right) / n$$

Weight of each station for ads -

$$stationWeight_i = \left(\sum_{i=1}^{n} (\#entries) + (\#exits) \right) / n$$

Price of each station for subsidy -

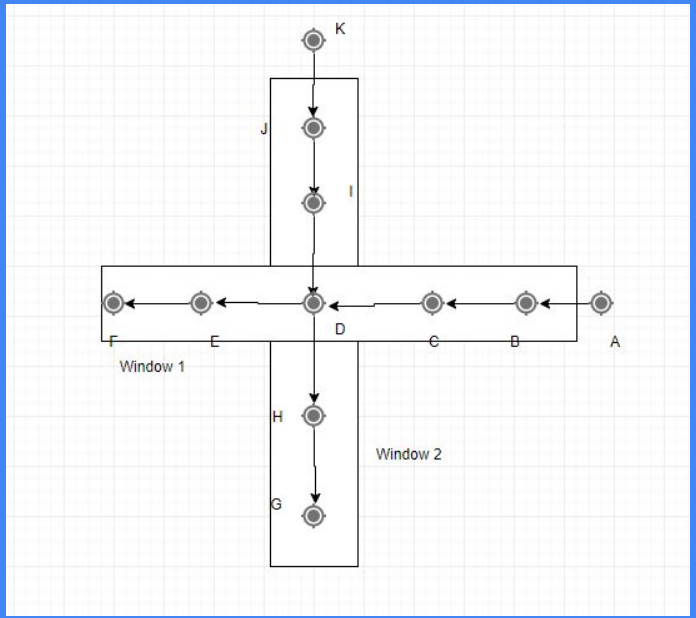
$$stationPrice_i = stationWeight_i * (\#entries)$$

Price of each station for ads -

$$stationPrice_i = stationWeight_i * (\#entries + \#exits)$$

Why a sliding window?

- To understand the local hubs
- Example: Traffic st Atlantic street is more or less independent of traffic at Lexington Ave



MongoDB and NodeJS

- After processing the MTA data, this data is dumped into a MongoDB collection.
- Each row of the collection contains the Station name, incoach price, advertisement price, number of entries and exits
- This data is then used by the front end to provide UI for a bidding system

Place Your Bid

#	Station Name	In Coach Price ($)	Bill Board ($)	Entries	Exits	Place Bid On In-Coach Price	Place Bid on Bill Board Price
1	FOREST HILLS 71	2632.41	3736.97	4922	3551	Place Bid	Place Bid
2	67 AV	172.56	207.12	1509	787	Place Bid	Place Bid
3	63 DR-REGO PARK	495.87	582.32	2772	1370	Place Bid	Place Bid
4	WOODHAVEN BLVD	1344.13	2034.05	3993	3253	Place Bid	Place Bid
5	GRAND-NEWTOWN	270.51	391.79	2300	1894	Place Bid	Place Bid
6	ELMHURST AV	95.17	99.91	1288	365	Place Bid	Place Bid

Feasibility and Scalability

- The producer-consumer system can scale as much as required owing to data being funneled into topics for each MTA Line
- Feasibility is guaranteed as the system is independent of MTA operations and no extra data is required

Demo - Show me the money!

Future work ?

- Stress testing the subsidy metric
- Building modular and appealing frontend for various user personas
- Providing cheaper options for local ads

The Team: Open for questions!

Utkarsh Prakash

Masters in Computer Science Student at NYU, Tandon

Tanay Varshney

Masters in Computer Science Student at NYU, Tandon

Ashim Aggarwal

Masters in Computer Science Student at NYU, Tandon

Jay Shah

Masters in Computer Science Student at NYU, Tandon