Team Instinct Showcase

2022-06-25

AGENDA

1. Data Model
1. The Travelers
1. Forecasting and Retail Agents
1. Billboard Pitch
1. Seasonality

DATA MODEL

- Initial state of the dataset was a collection of *.csv files (some sharded) with many string fields (especially entity identifiers).

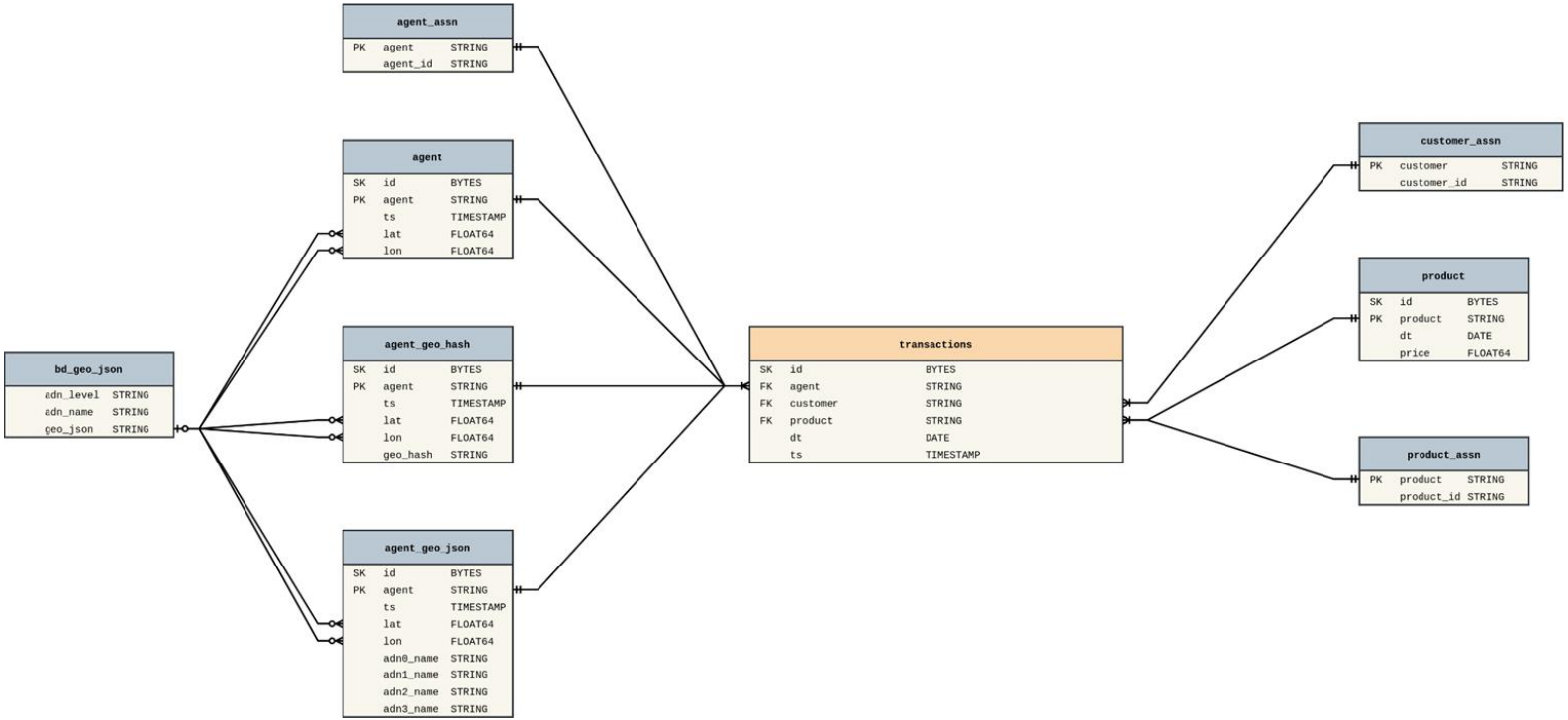
- Deterministic Ordinal Serialization encodes string identifier fields. Reduced storage complexity by a lossless compression ratio of ~10:4

- Engineered purpose designed data loaders to further reduce the memory complexity by a lossless compression ratio of ~10:2.3

DATA MODEL

- Need for a distributed cluster to perform some of our intended EDA still persisted
- Provisioned a Star Schema based data warehouse on Google BigQuery
- Provided ease of on-the-fly extendibility to the data warehouse
- The denormalized form allowed us to write small rough-and-ready queries for EDA
- Allowed us to offload the analytics computation to the cloud

DATA MODEL

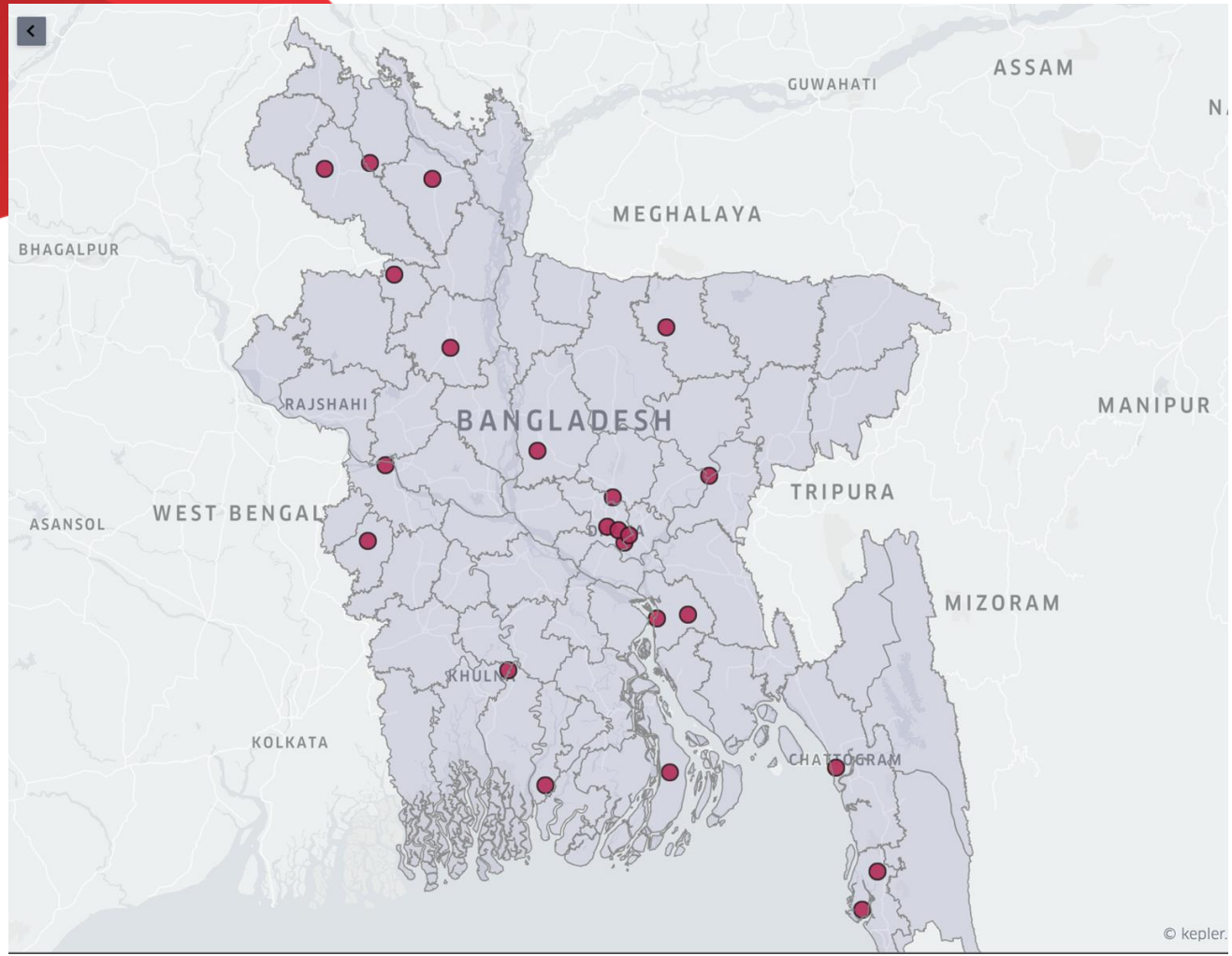

agent_assn
PK agent STRING
agent_id STRING
agent
SK id BYTES
PK agent STRING
ts TIMESTAMP
lat FLOAT64
lon FLOAT64
bd_geo_json
adn_level STRING
adn_name STRING
geo_json STRING
agent_geo_hash
SK id BYTES
PK agent STRING
ts TIMESTAMP
lat FLOAT64
lon FLOAT64
geo_hash STRING
agent_geo_json
SK id BYTES
PK agent STRING
ts TIMESTAMP
lat FLOAT64
lon FLOAT64
adn0_name STRING
adn1_name STRING
adn2_name STRING
adn3_name STRING
transactions
SK id BYTES
FK agent STRING
FK customer STRING
FK product STRING
dt DATE
ts TIMESTAMP
customer_assn
PK customer STRING
customer_id STRING
product
SK id BYTES
PK product STRING
dt DATE
price FLOAT64
product_assn
PK product STRING
product_id STRING

THE TRAVELERS

- Our definition of a top ‘Traveler’ is based on 3 key metrics and 1 supporting metric:

  - Average Distance traveled per transaction
  - Unique Agents Transacted
  - Number of transactions
  - Agent Uniqueness

- Only customers who were in the 95th percentile for the key metrics were considered, which accounted for around 2400 people.

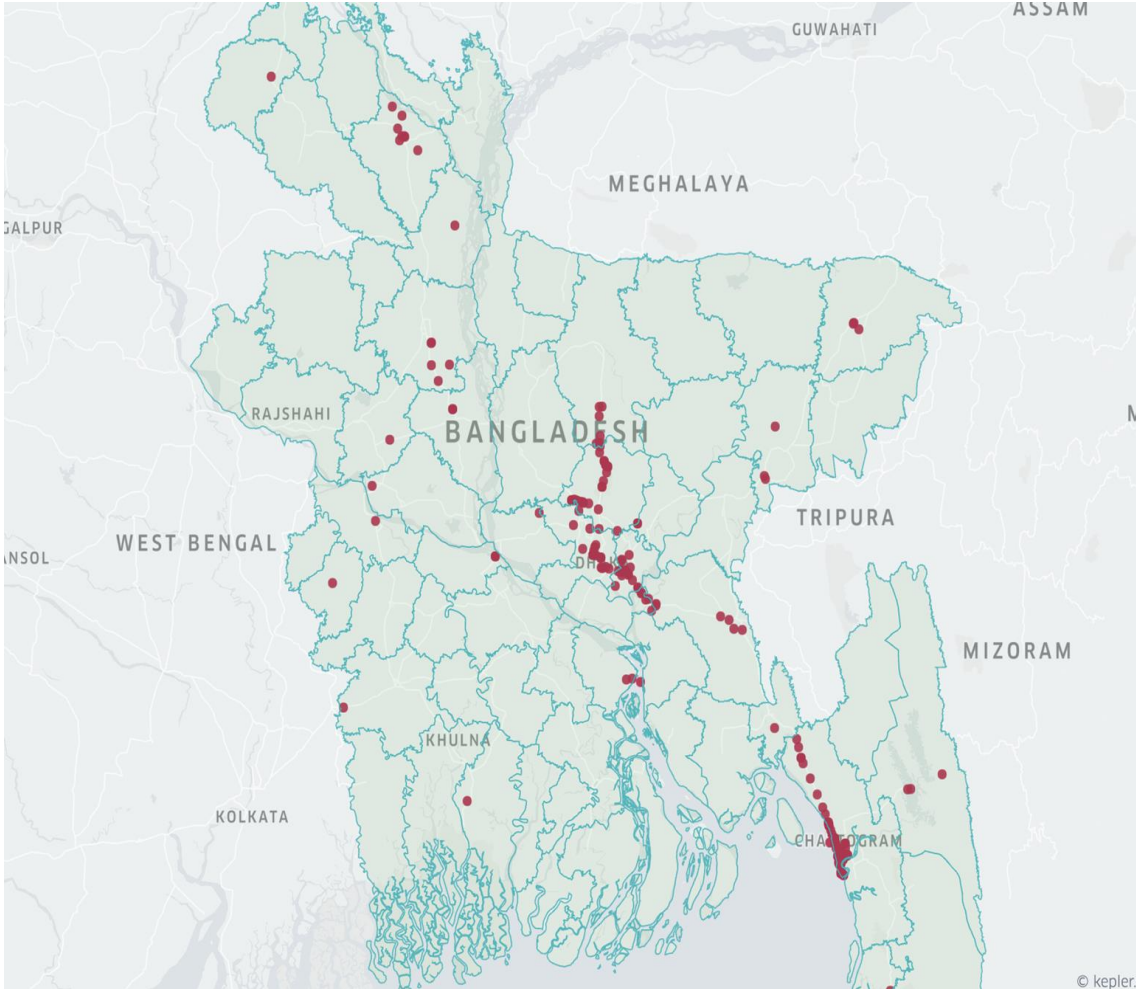

GUWAHATI
ASSAM
MEGHALAYA
BHAGALPUR
RAJSHAHI
BANGLADESH
MANIPUR
TRIPURA
WEST BENGAL
ASANSOL
MIZORAM
KOLKATA
CHATTOGRAM
© kepler.

THE TRAVELERS

- Retails serving highest number of top travelers tend to be along major highways near Dhaka and approaching Chittagong. This indicates that many travelers transact while on break during transit.

- New product for travelers.
    - Bundle voice and data package for short duration
    - Will be tailored to make the offer lucrative to travelers
    - It can only be availed via chosen retailers

THE TRAVELERS

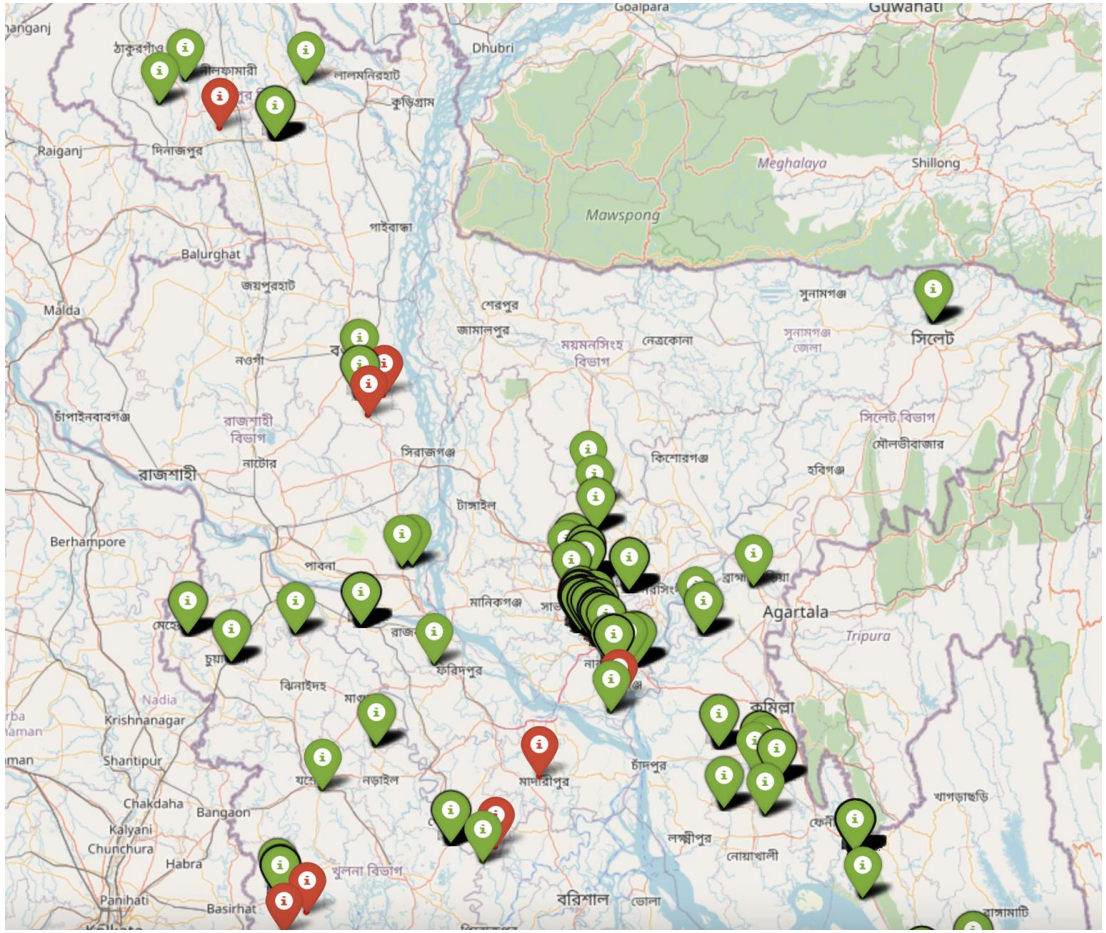


TOP 100 AGENTS

- Amongst top selling 100 retail, 8 of the retails have no other retail point within 500 meters radius.

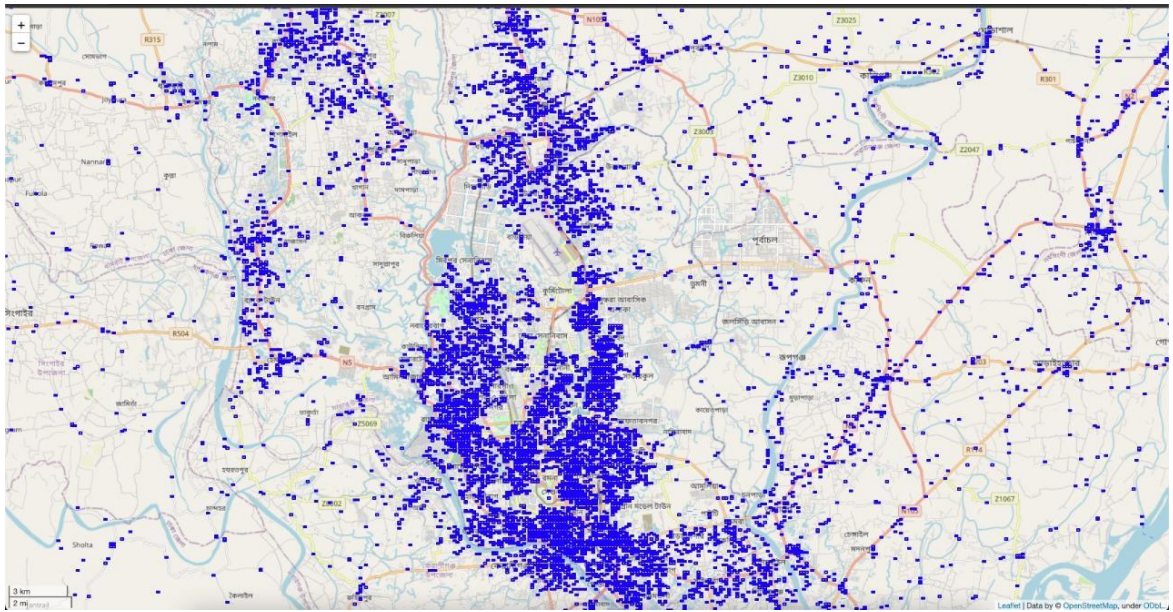
- Recommendation for these retails are:
  - These retails deal with a lot of customers, hence, good medium to market new/other products/services.
  - New retailers around these retails can be introduced.

AGENT


Goalpara
Guwahati
Dhubri
Meghalaya
Shillong
Mawspong
Raiganj
Balurghat
Malda
Berhampore
Nadia
Krishnanagar
Shantipur
Chakdaha
Kalyani
Chunchura
Habra
Panihati
Basirhat
Bangaon
Agartala
Tripura
বরিশাল
সিলেট
রাজশাহী

AGENT

- There are 54,099 geohash (1.2 Km x 0.6 Km) with 89,610 retailer.

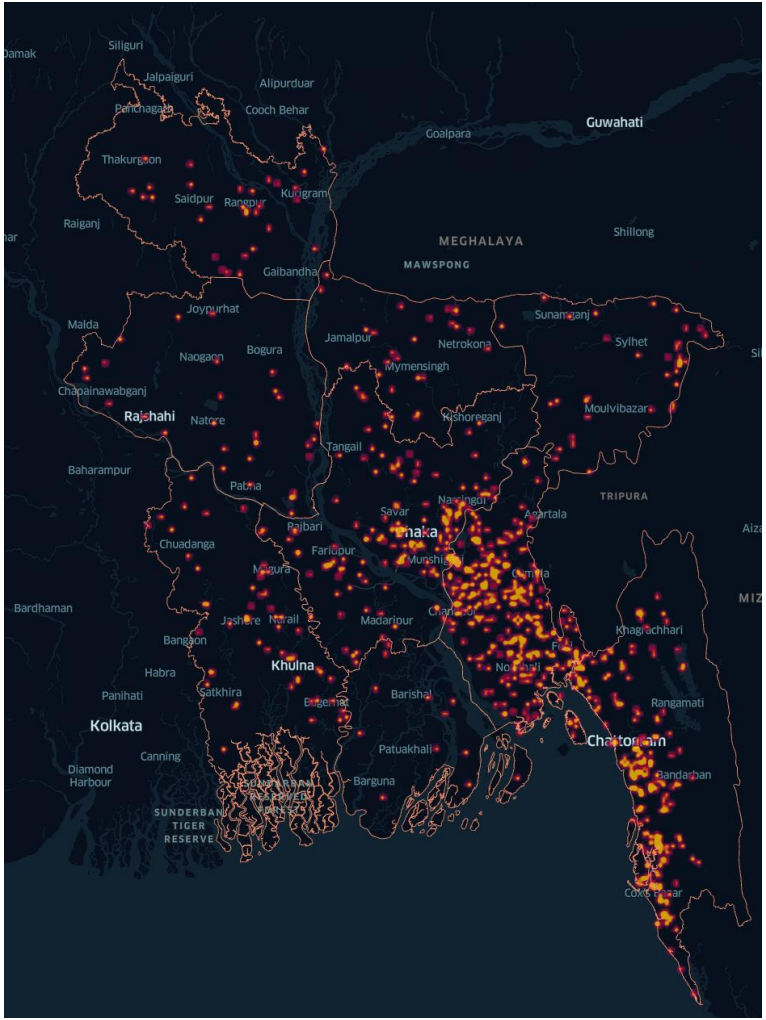


MONOPOLY AGENTS

- For the next 10 retailer deployment, the zones having good sales but low agents should be considered.

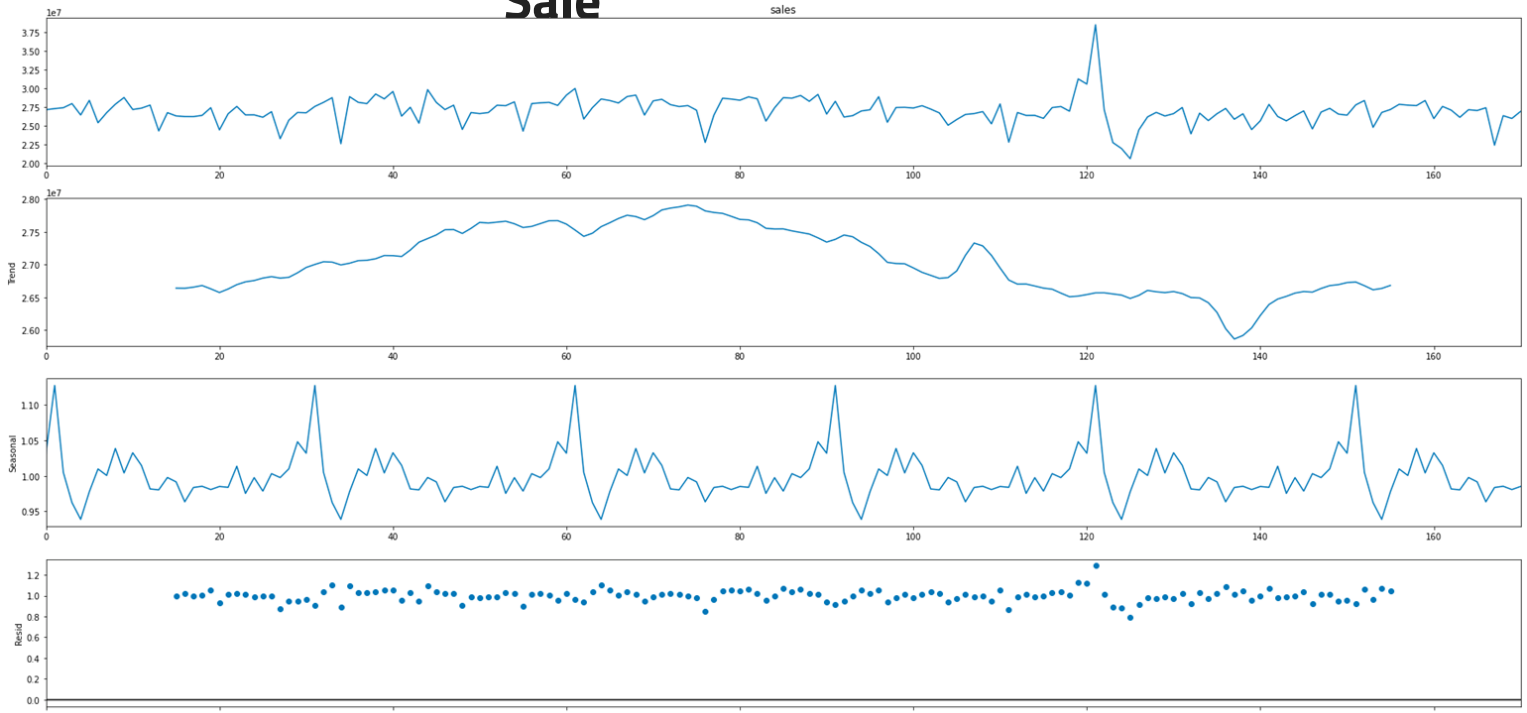
- The top districts with highest number of monopoly agents:
  - Chittagong
  - Comilla
  - Cox's Bazar
  - Noakhali
  - Chandpur

MONOPOLY AGENTS



FORECASTING

Next Week's Daily Total Sale



FORECASTING

Next Week's Daily Total Sale

- The seasonal decomposition shows a multiplicative nature.
- Has trend component, seasonality component.
- The model trained is an ARIMA(0, 1, 0) - a random walk model.

FORECASTING

Forecasting Next Day's Sales for Each Retail

- Fitting an individual model for each of the retails is not a scalable solution.
- One single time series model won’t be able to generalize well.
- we thought of attacking the problem by looking at the data in a tabular form.

FORECASTING

Features

- **Day of Week** - 0,1,2,3,4,5,6,7
- **No of Neighbor Retails** - Retail points within 500m radius.
- **Distinct Products** - Distinct products/services the retailer has.
- **Distinct Product Value** - Monetary value of distinct products.
- **Previous 5 Weekdays Sales** - E.g., for predicting this Tuesday's sales for a retailer we consider previous 5 Tuesday's data.
- **District**

FORECASTING

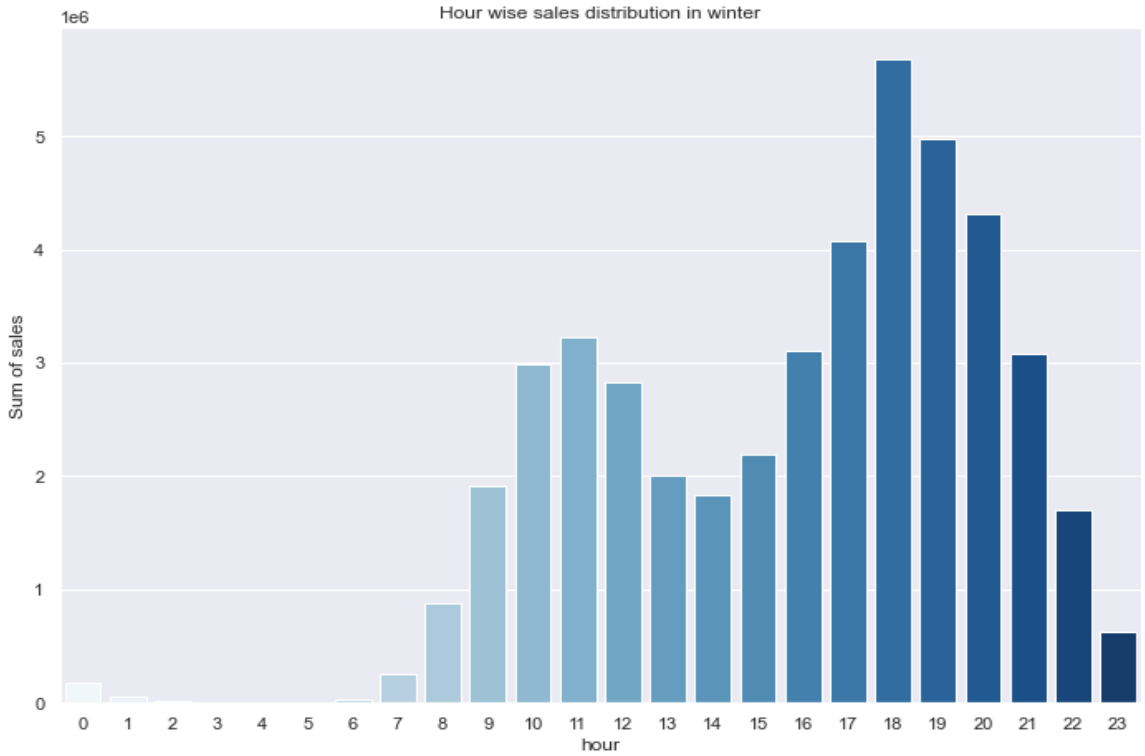
Modelling & Evaluation

- Initially, we used rolling average of the previous 5 weekdays sales values to establish a baseline of the model's accuracy. We got an MAE of ~230.

- With the newly engineered featured, feature preprocessing, a hypertuned XGBoost regressor trained on 80% of the training set, we obtained an MAE of around ~110.

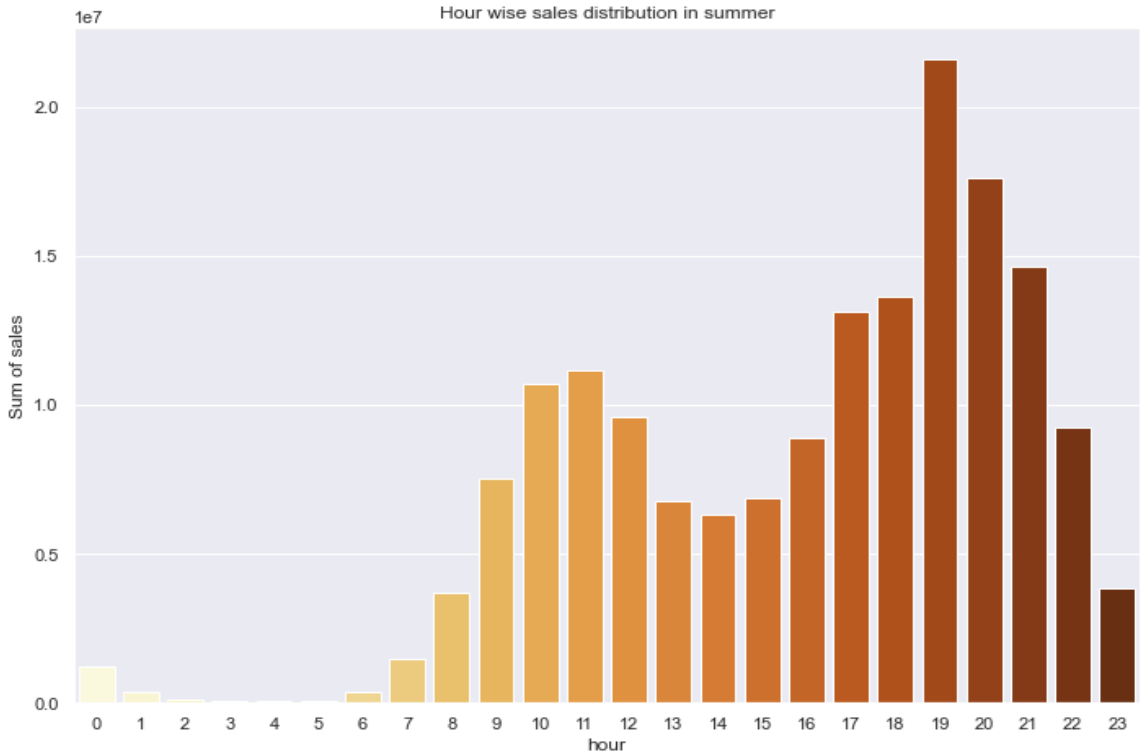
SEASONALITY

- During winter season the peak hour is between 6PM and 7PM
- People w/ purchasing power are in office during morning hours. Hence, morning peak is lower than evening peak



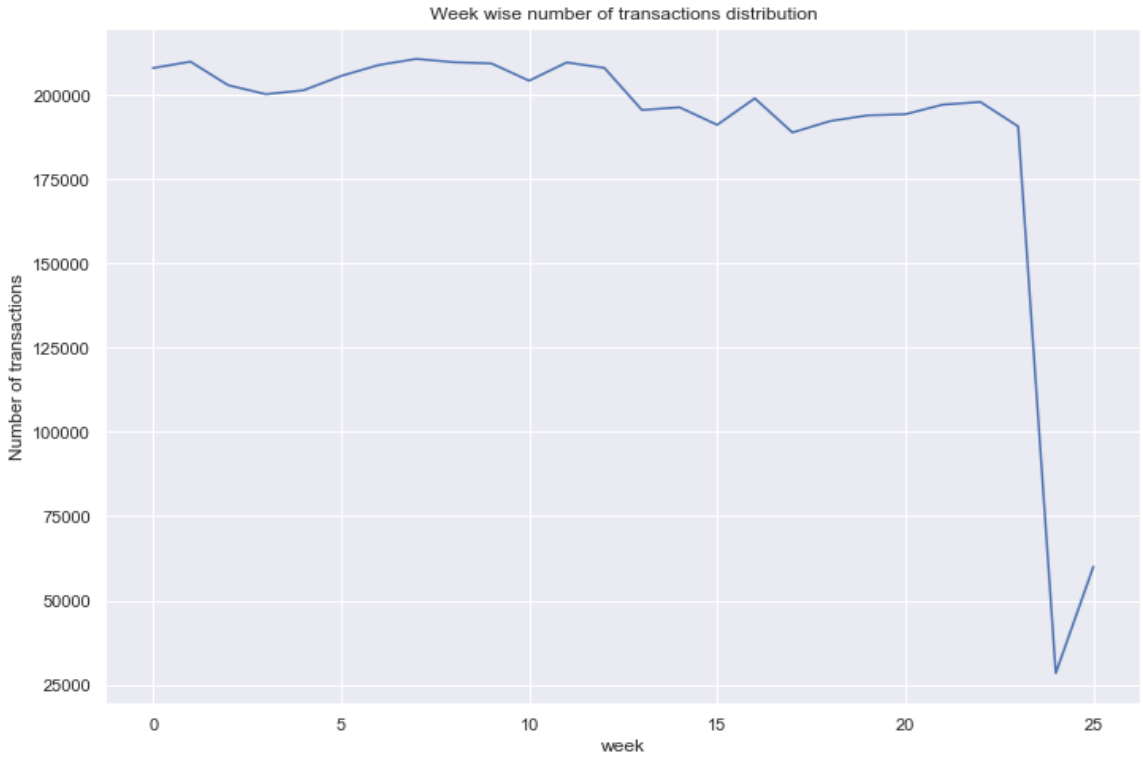
SEASONALITY

- During summer the peak hour shifts to 7PM upto 8PM
- This peak hour shift gives us an idea about natural daylight saving phenomenon



SEASONALITY

- Transactions fell during Ramadan
- Has stayed the same since. This could be due to inflation
- Lockdown and Covid restrictions impact the transaction volume negatively



SEASONALITY

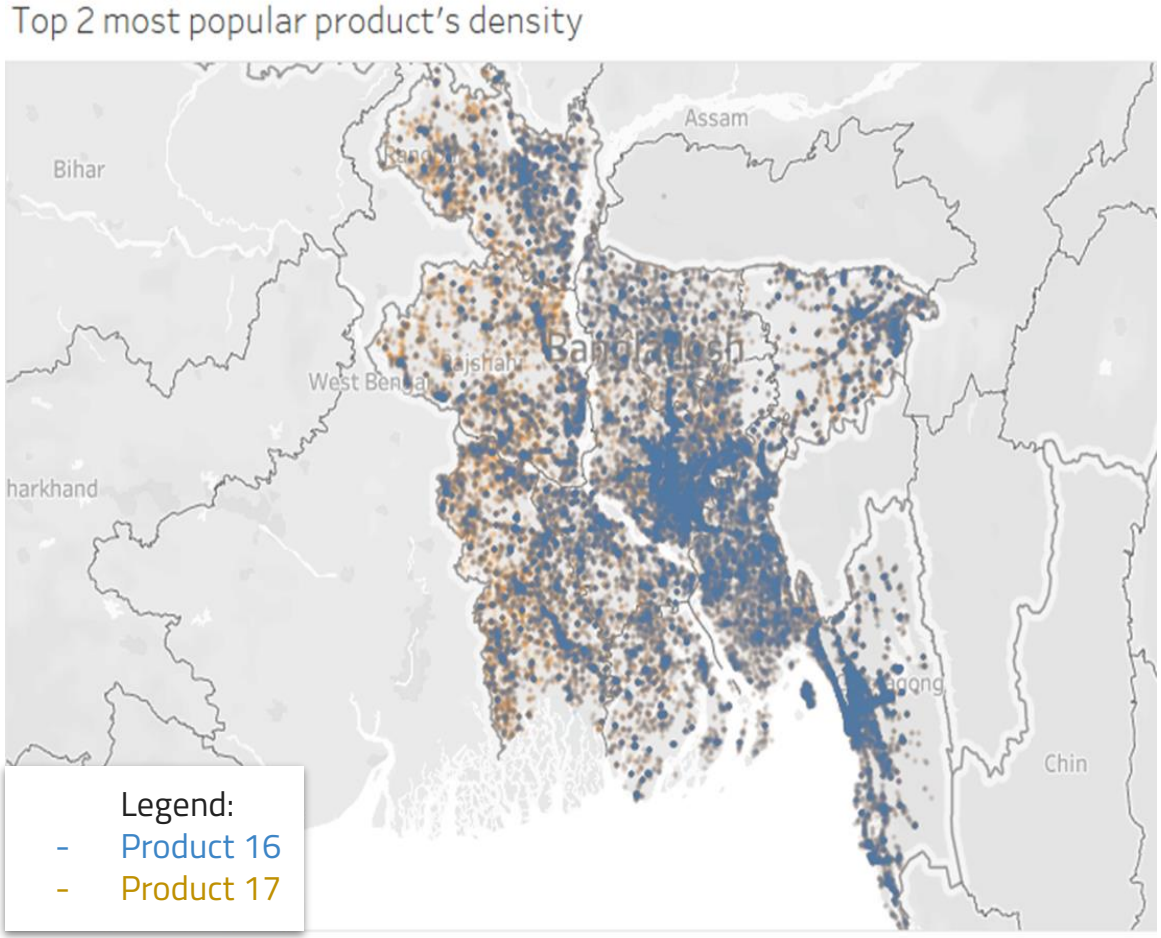
- During the month of Ramadan we can notice a sharp decline in transactions during Iftar hour 6PM to 6:59PM
- Transactions fell during Ramadan
- Has stayed the same since then. This could be due to inflation
- We can release a new products targeted for the price sensitive users to mitigate the fall in transaction volume

BILLBOARDING

- Top 2 most popular products are Product 16 and Product 17 with ~8.8 million and ~8.6 million transactions respectively
- Product 16 has high density in Dhaka and Chittagong
- Product 17 is more prevalent across the country

BILLBOARDING

- Product 16 is a send cash or digital payment feature
- Product 17 might be a cash out feature



BILLBOARDING

- Dhaka/Chittagong are already saturated

- Recommend to place the billboard in emerging markets (e.g. Khulna)
  - Product 16 has a low barrier to entry (high volume, low price)
  - Serves to effectively acquire new customers
  - Establishes the MFS' presence through visibility/engagement

- Goal is to capture the emerging market and gain the first-mover's advantage

THANK YOU

All our supplementary resources are available here