

Robi Datathon 3.0

ACI_ServerDown - Team ID: 322

Problem 1

Statement

Predict out of the given customer list, who will purchase which of the new products.

Data

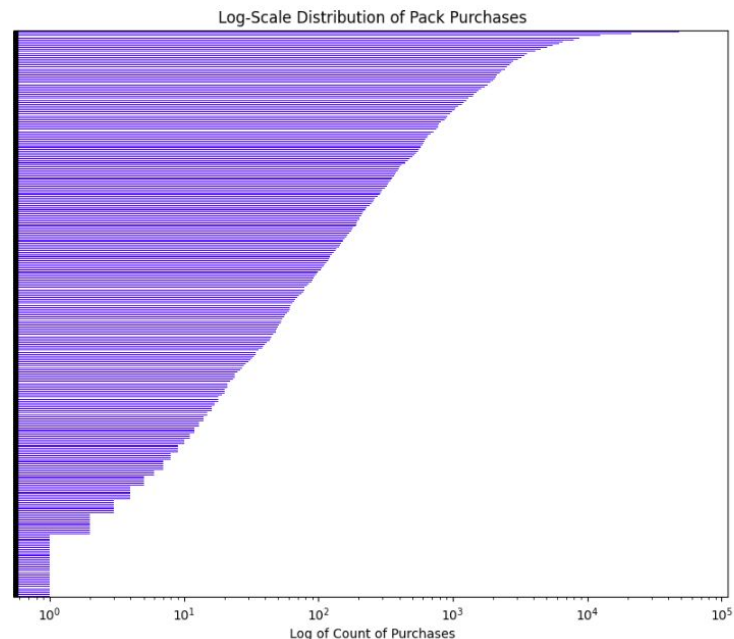
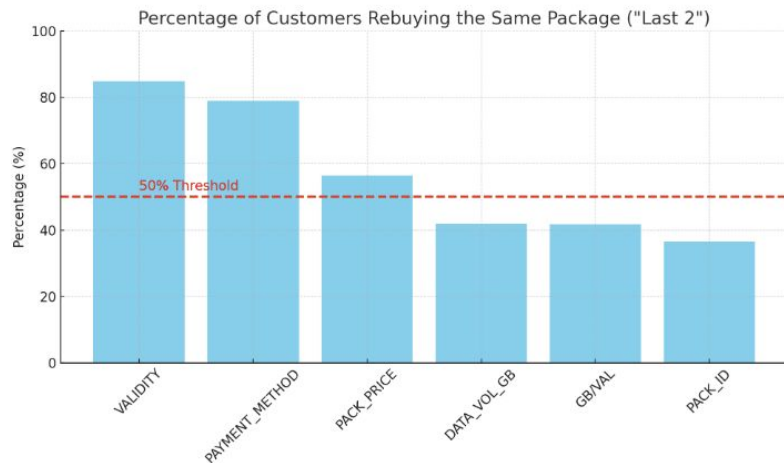
- Customer Profiles
- Customer Purchase History
- Customer Daily Internet Usage History
- New and old product catalog

Focus

Identifying the new packs that are the most similar to the packs customers are most used to buying, making them the most probable choice.

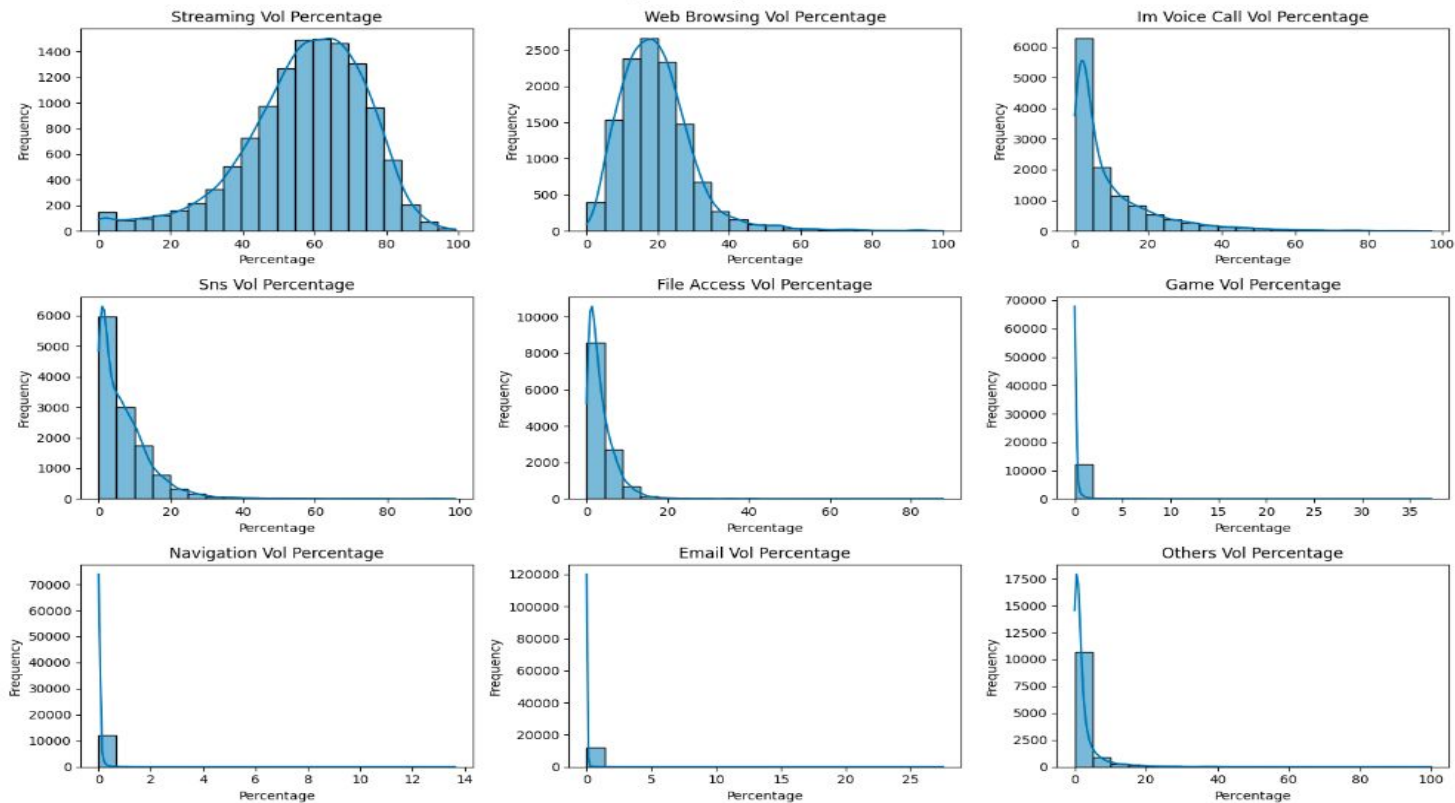
Data Exploration

- Customers have a strong preference to keep buying the same amount
- Goal is to find the ideal amount the customer is likely to buy and predict the next new pack purchase using that.



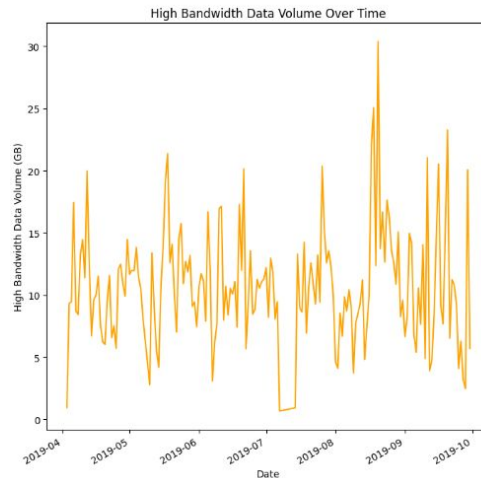
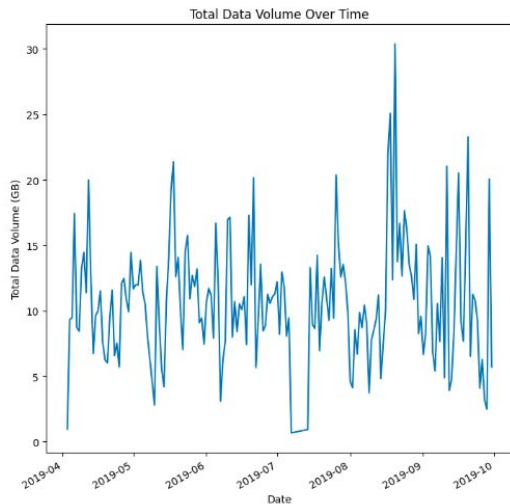
EDA Cont.

Distribution of Data Usage Percentages for Premium Service Customers



Feature Engineering & Modeling

- First, segregate customers by SERVICE_TYPE, as these are exclusive.
- Customers preferred the same payment methods (84% Last Two, 73.5% Last Three, 63% Last Four)
- For 7 & 30 days, we stick to that as our target feature space
- No 3 Days and 15 Days package in new packets. So, for 3 days we consider that the customer is unlikely to buy. For 15 days we perform further feature engineering.
- We consider the lowest price of the same volume in the new packages, and afterwards consider the volume customers get for that same price, and pick the most ideal.
- The remaining features are PRICE and GB_VOL.
- We calculate a Moving Average (MA) over the GB to consider customer's previous history and to put more weights on more recent pr
- We find the most likely pack by putting these features in a normalized feature space, and finding the closest neighbor



Problem 2: Have to predict probable uptake or purchase count of the products given products

Datasets:

PACK_PURCHASE.csv: Contains history of pack purchase for each customer and products

PRODUCT_CATALOGUE.csv : Contains product packages available for customer to purchase.

Basic Analysis:

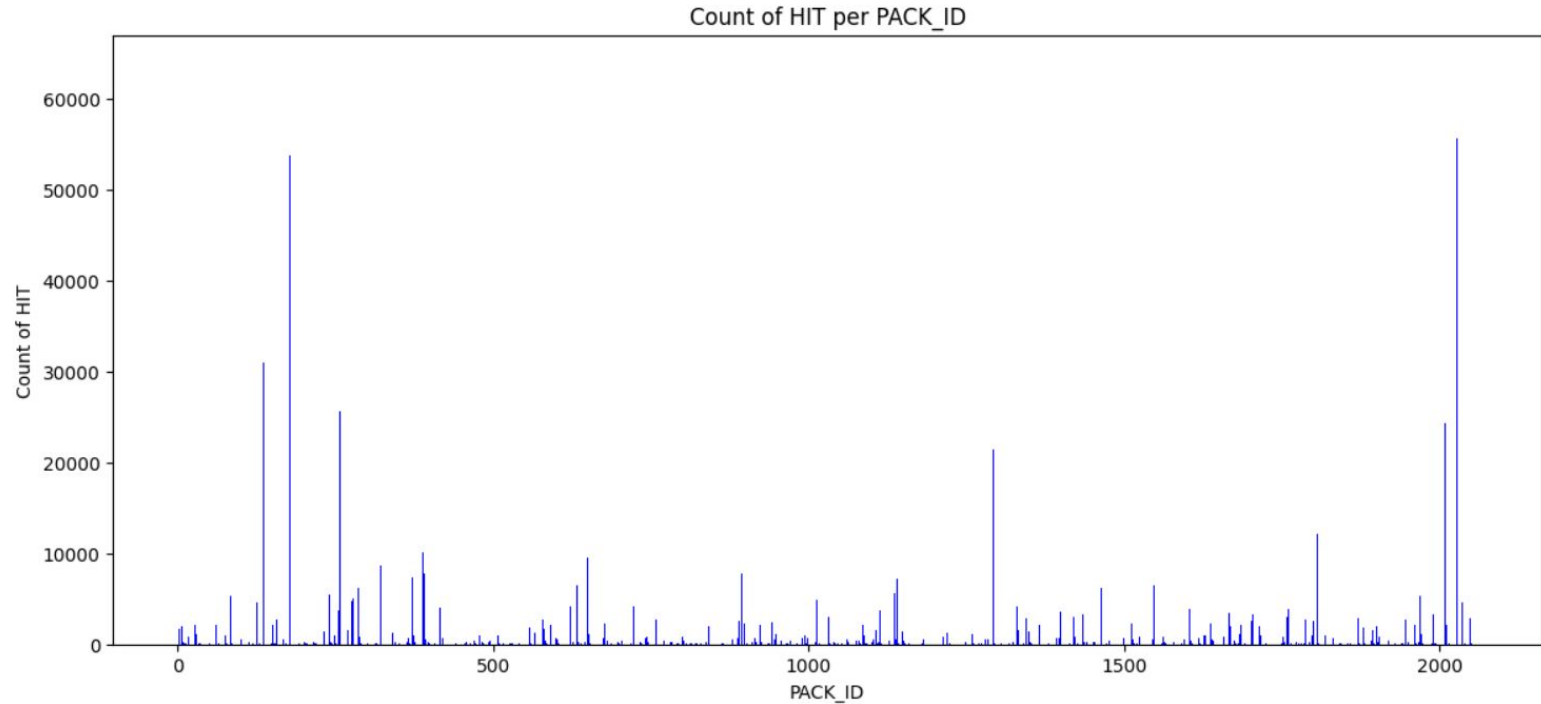
Total Product: 2049

Total Purchase history: 1208710

Have to predict for 20 product.

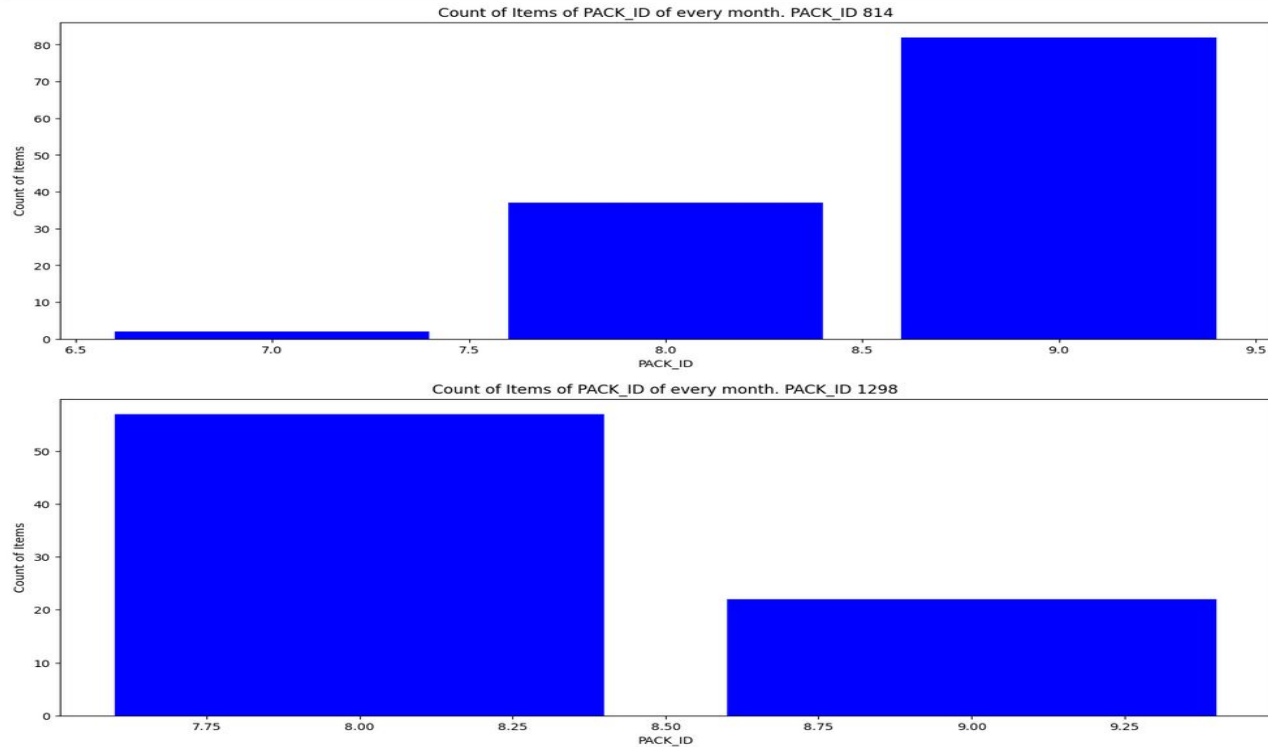
Data Exploration and Feature Engineering:

Hit counts exploration for each package for the whole time



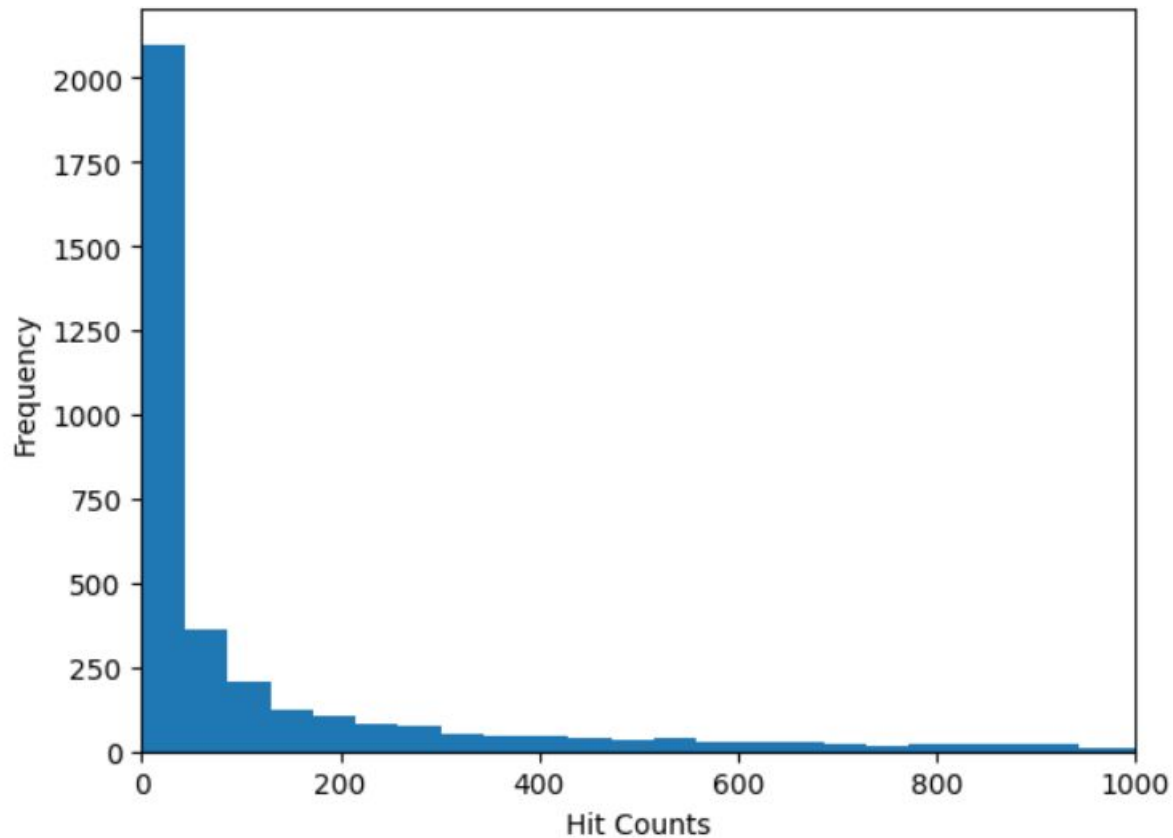
Data Exploration and Feature Engineering

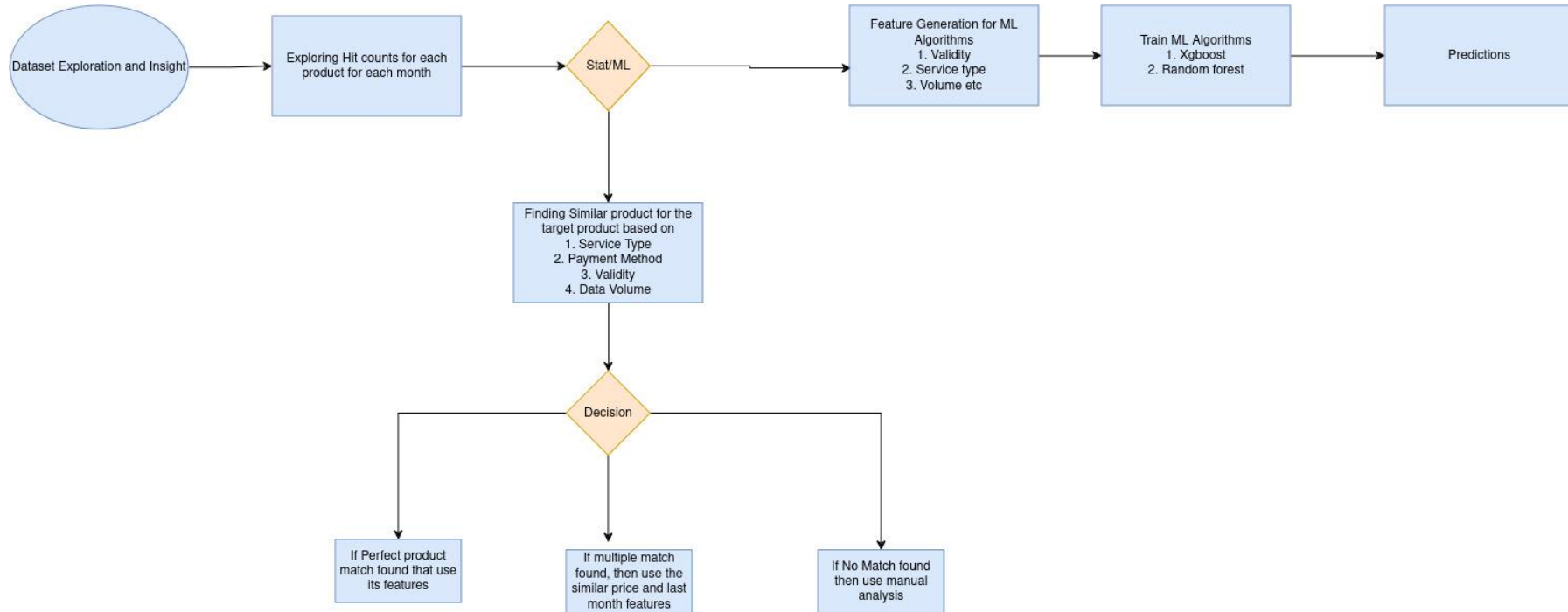
Hit counts exploration for package purchase for different particular month



Data Exploration and Feature Engineering

Hit counts
exploration for
package
purchase for
different
particular month





Problem 3: Positioning the billboards

Data used:

[Road Network Data of Bangladesh](#) consists of open access data collected from Open Street Map.

[Population Density in 400 m H3 Hexagon](#) data collected from Fixed up fusion of GHSL, Facebook, Microsoft Buildings, Copernicus Global Land Service Land Cover, Land Information New Zealand, and OpenStreetMap data. These are Vector H3 hexagons with population counts at 400m resolution.

Division polygons (Administrative level 1) collected from government resources

Why these data were chosen?

The reasoning behind the selection of population data and road network is: billboards are good for two reasons, how many people can see them and how long will they see it.

Road network data enables us to select busy intersections of road where traffic congestion is bound to happen, population data ensures that there are enough people in that congestion to make the economy of billboard work.

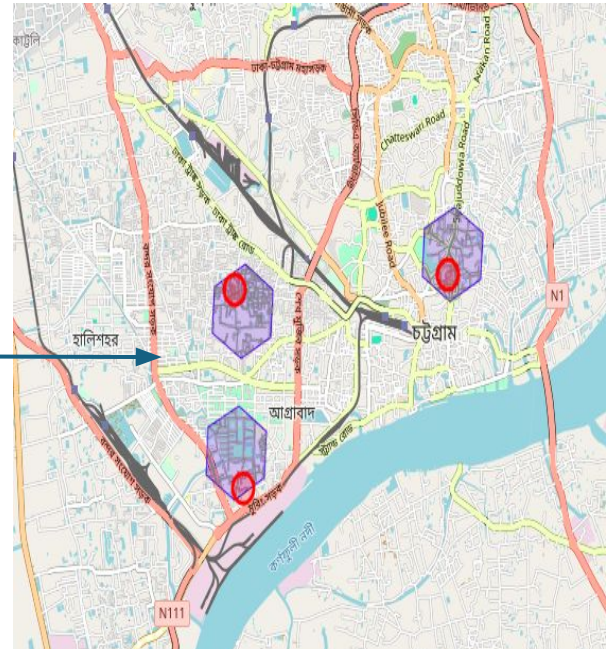
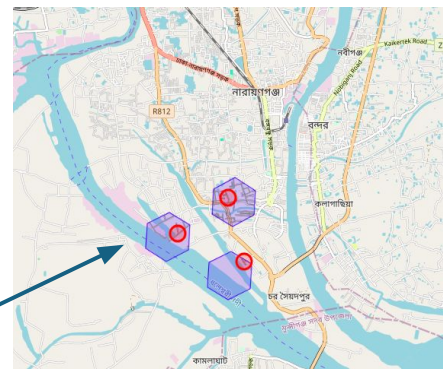
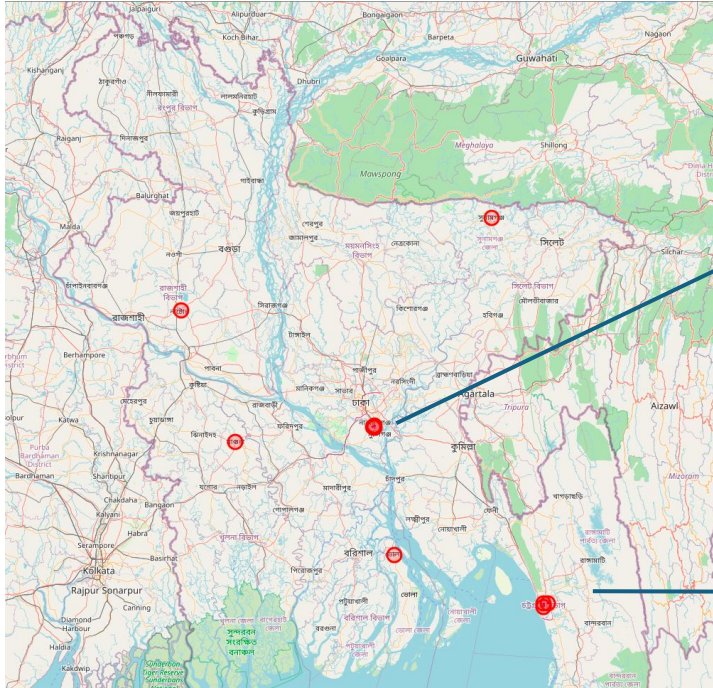
Methodology

1. Getting division polygons: Dissolve all Upazillas of a same division together to merge a large polygons for the division.
2. Getting population centers: Finding out the H3 polygon with highest population density in each division.
3. Determining how many from which division: Based on the Access and use of ICT by the Households and Individuals 2022 data delivered by [government statistics](#) we chose 3 points in both Dhaka and Chittagong (the two leading market) and one each in every other district.

Methodology (Continued)

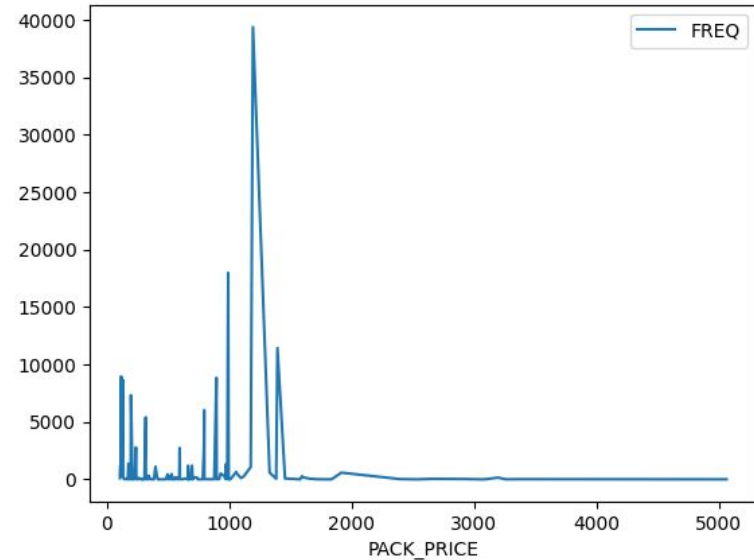
4. Then we overlay the road connection geojson files on each of the selected 10 H3 Hexagons. Using these road data, we find out intersections of major roads within that hexagon. We then identify the largest intersections based on road covered within that intersection which will cater to thje highest number of commuters.
5. We identify these 10 intersections as the 10 chosen points for placing our billboards.

Prediction



Problem 4: New Product Recommendation (Approaches)

- Regarding comparable items that have already been released, determining clusters based on the quantity of the package purchased (HIT COUNT)



Problem 4: New Product Recommendation (Approaches)

- Observing the range within which a product's price can rise without negatively impacting sales
- Analyzing customer's package switching pattern
- Analyzing prospects of promotion on particular "Data Consumption" category