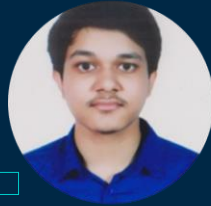


Not_So_Strong



Dept. of Computer Science & Engineering
Rajshahi University of Engineering & Technology

Problem

- i) Problem Statement
- ii) Analysis
- iii) Solution Breakdown

01

Problem Statement

We need to develop a machine learning model to accurately predict which new products (Pack_ID), customers (CUSTOMER_ID) from the provided list in 'Q1-customers.csv' are likely to purchase. If a customer is not predicted to purchase any products, we have to assign a value of -1.

Analysis

Exploratory Data Analysis (EDA)

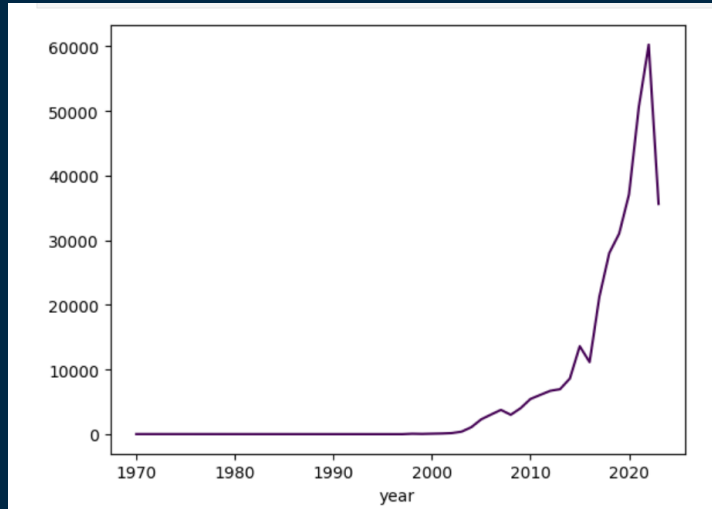


Fig. 01 First Connection vs. Year

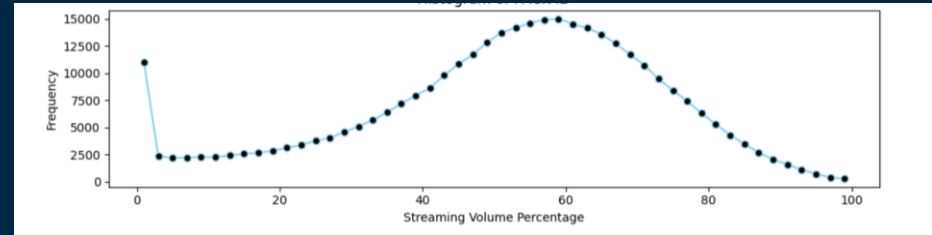


Fig. 02 Histogram of Streaming Volume Percentage.

Analysis

Exploratory Data Analysis (EDA)

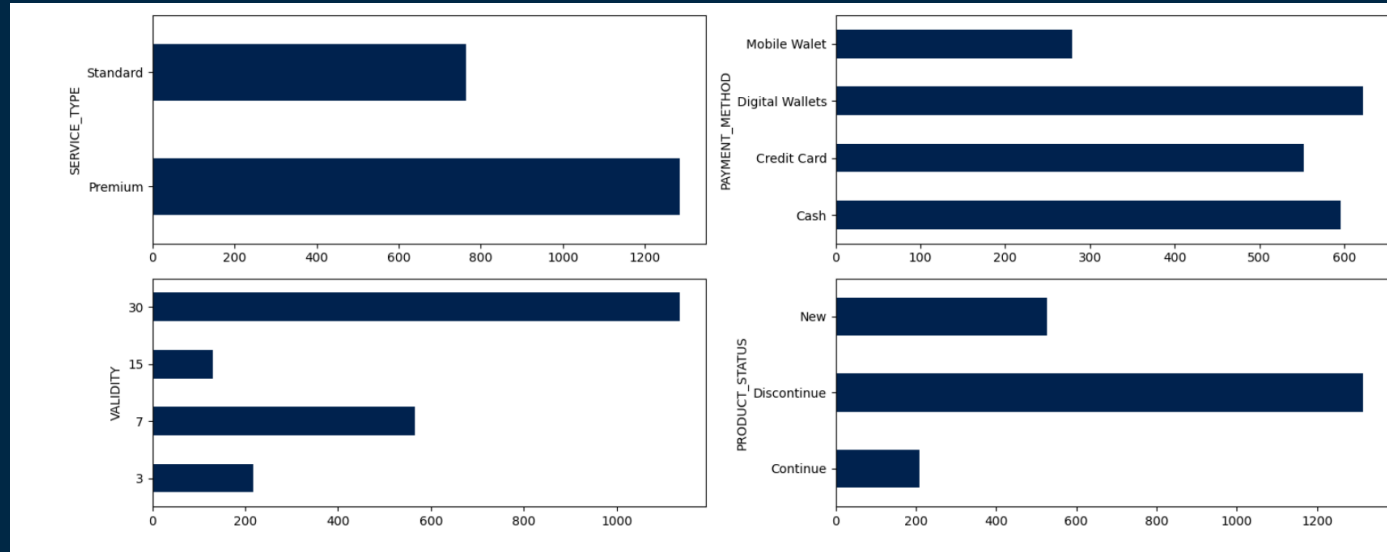


Fig. 03 Categorical Description of Product Catalogue

Analysis

Exploratory Data Analysis (EDA)

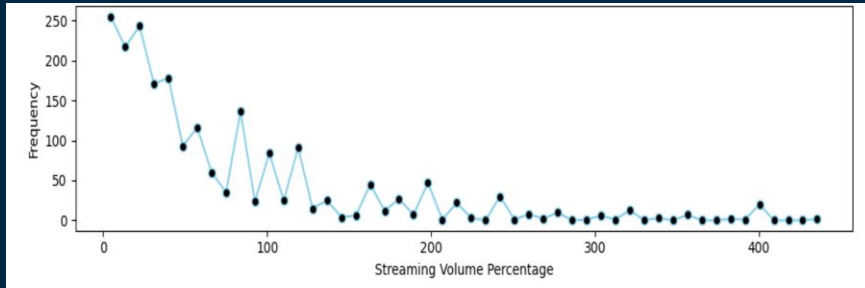


Fig. 04 Histogram of Data Volume (GB)

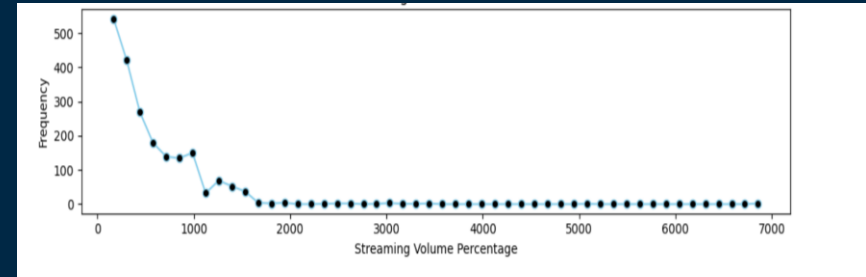


Fig. 05 Histogram of Pack Price.

Analysis

Exploratory Data Analysis (EDA)

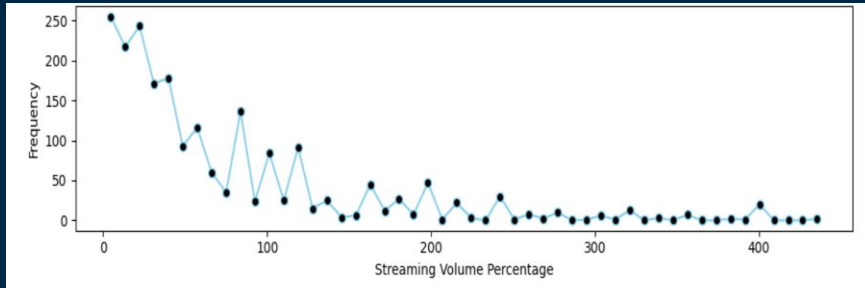


Fig. 04 Histogram of Data Volume (GB)

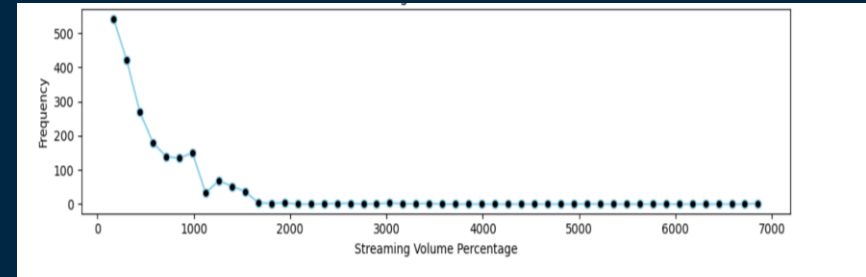


Fig. 05 Histogram of Pack Price.

Analysis

Exploratory Data Analysis (EDA)

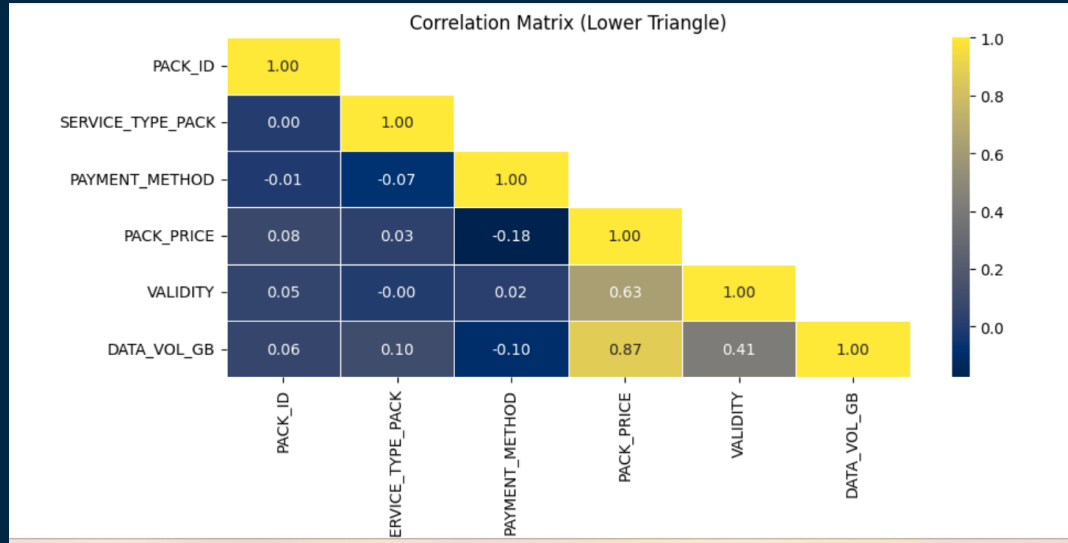


Fig. 06 Correlation between Product Features

Analysis

Feature Engineering

1. Preprocessing

- Dropping Null Values
- Data Filtering
- Categorical to Numerical Conversion

2. Data Merge

- Merged all table
- Grouped by 'CUSTOMER_ID'
- Sorted by 'EVENT_DATE' (Recency)
- Added New Feature ('PER_GB')

3. Experimentation

- Features considered
'SERVICE_TYPE_PACK',
'PACK_PRICE',
'PAYMENT_METHOD',
'VALIDITY',
'DATA_VOL_GB',
'PER_GB'

Solution Breakdown

Statistical Approach

We created a new data frame which only contains 'New' of PRODUCT_STATUS from PRODUCT_CATALOGUE.

Based on customer's full purchase history, we have filtered the new data frame (i.e., mean, mode, median), and also done some optimizations, and finally got our expected result using 'Mean' on the features.

Problem

- i) Problem Statement
- ii) Analysis
- iii) Solution Breakdown

02

Problem Statement

We need to predict probable uptake or purchase count of the products given in 'Q2-new-products.csv'.

Analysis

Exploratory Data Analysis (EDA)

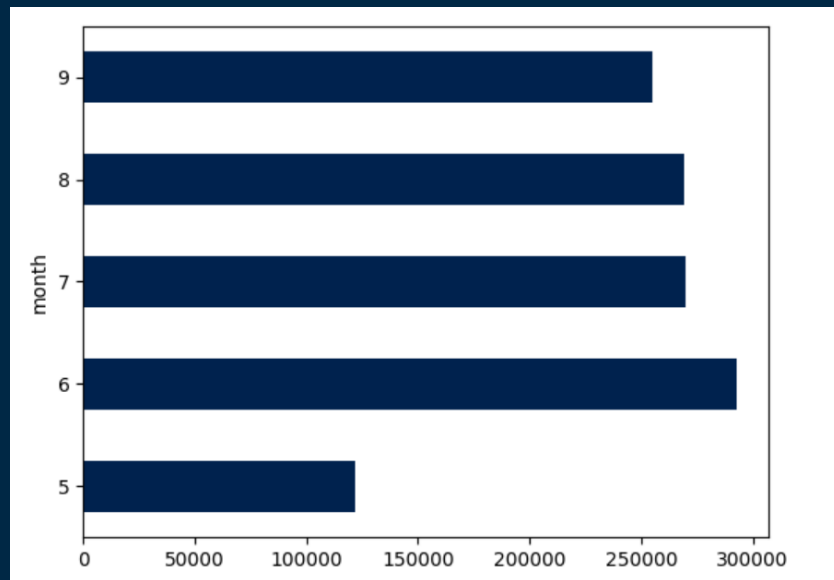


Fig. 07 Histogram of Per Month Sells

Analysis

Feature Engineering

1. Preprocessing

- Dropping Null Values
- Data Filtering
- Categorical to Numerical Conversion

2. Data Merge

- Merged 'PRODUCT_PURCHASE' & 'PRODUCT_CATALOGUE'
- created a new data frame which only contains 'New' of PRODUCT_STATUS from PRODUCT_CATALOGUE

Solution Breakdown

Statistical Function

- Unique pack count
- Mean count of important features
- Calculated frequency and recency ratio
- Calculated score based on previous calculated value
- Sorted BOX_ID
- Mapping new packs with BOX_ID and verifying compatibility of new boxes
- If compatible, then assigning frequency of BOX_ID to purchasing new pack
- Optimizing output using statistical values

Problem

- i) Problem Statement
- ii) Public Dataset Description
- iii) Analysis
- iv) Solution Breakdown

03

Problem Statement

We need to propose top 10 places where to place billboards for promoting these new products across Bangladesh using any public data source to support our recommendation.

Public Dataset Description

The dataset comprises information on districts in Bangladesh, including their names, respective areas, and populations (Census 2022-06-14). This data has been scrapped from the website

<https://www.citypopulation.de/en/bangladesh/cities/> using **BeautifulSoup**.

The extracted data contains:

1. District Name
2. Area (km²).
3. Population

Analysis

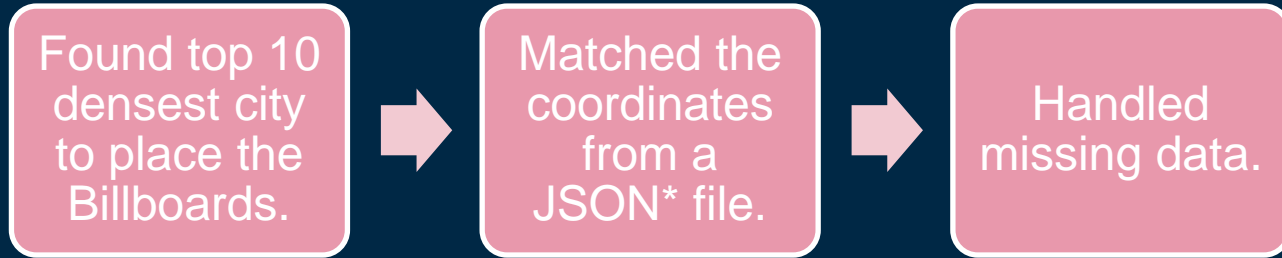
Insight



Where there is people, there should be the
billboards!

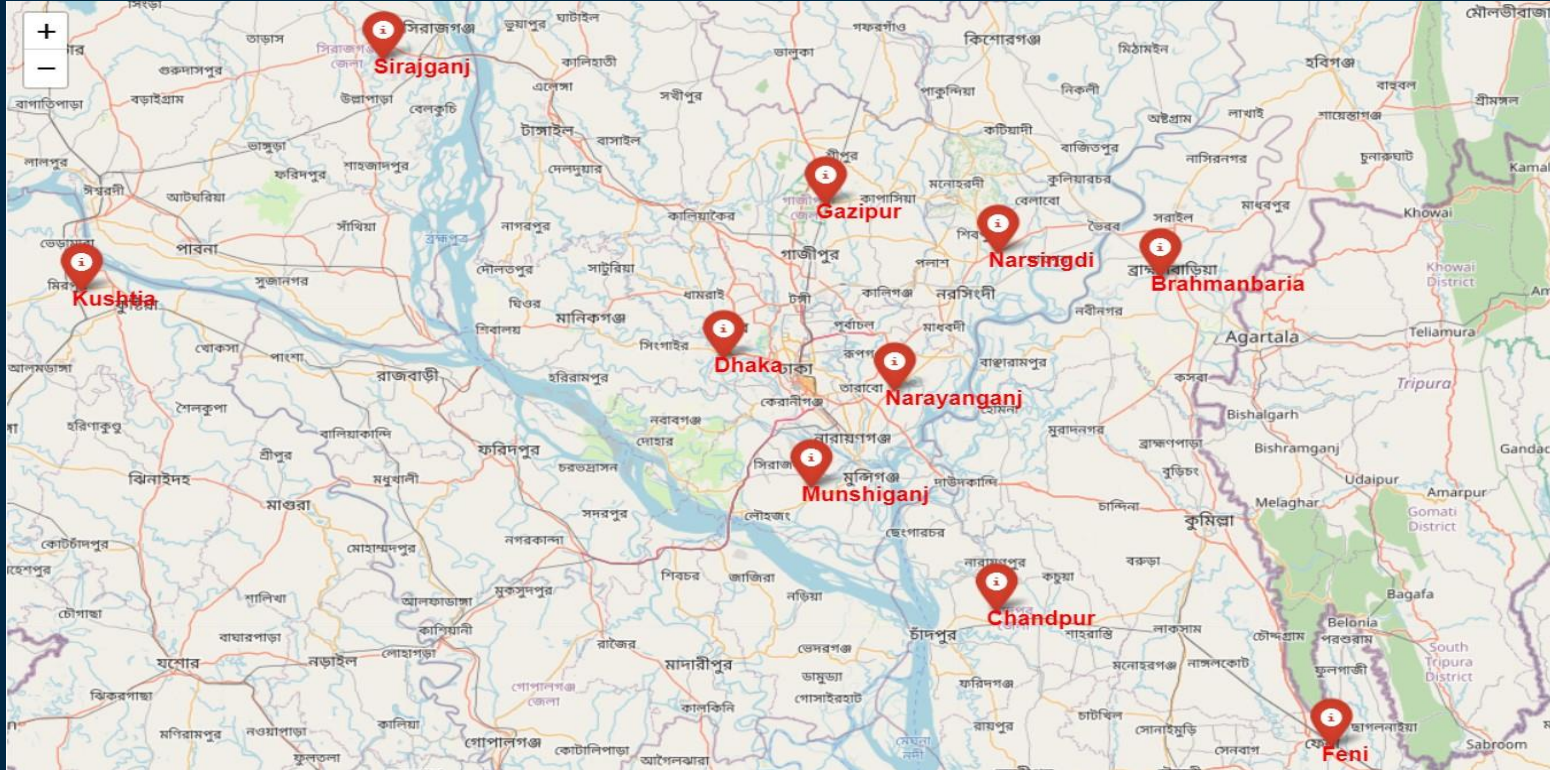
Density Matters!

Solution Breakdown



*https://github.com/ahnaf-tahmid-chowdhury/Choropleth-Bangladesh/blob/master/bangladesh_geojson_adm2_64_districts_zillas.json

Solution Breakdown



Problem

04

- i) Problem Statement
- ii) Analysis
- iii) Solution Breakdown

Problem Statement

We need to share our recommendation on what an ideal product catalog should be for “Premium” customer with “Cash” payment and for 30 days validity.

- a) Recommending 10 products that needs to be added to the product list
- b) Recommending 10 products that can be dropped from new product.

Analysis

Feature Engineering

1. Preprocessing

- Data filtering according to statement

2. Data Merge

- Merged
'CUSTOMER_PROFILE' &
'PRODUCT_CATALOGUE'

Solution Breakdown

Statistical Function

- Calculated frequency of purchased box, sorted it by maximum count
- Selected important features such as 'PACK_PRICE', 'DATA_VOL_GB' & run some queries on merged dataset
- Calculated per GB price near to maximum counted box from the outcome
- Predicted best package related to maximum counted box

We are open to questions now!

Thank You