DATA TEAM PRESENTS

Coffee Shop Sales Improvement

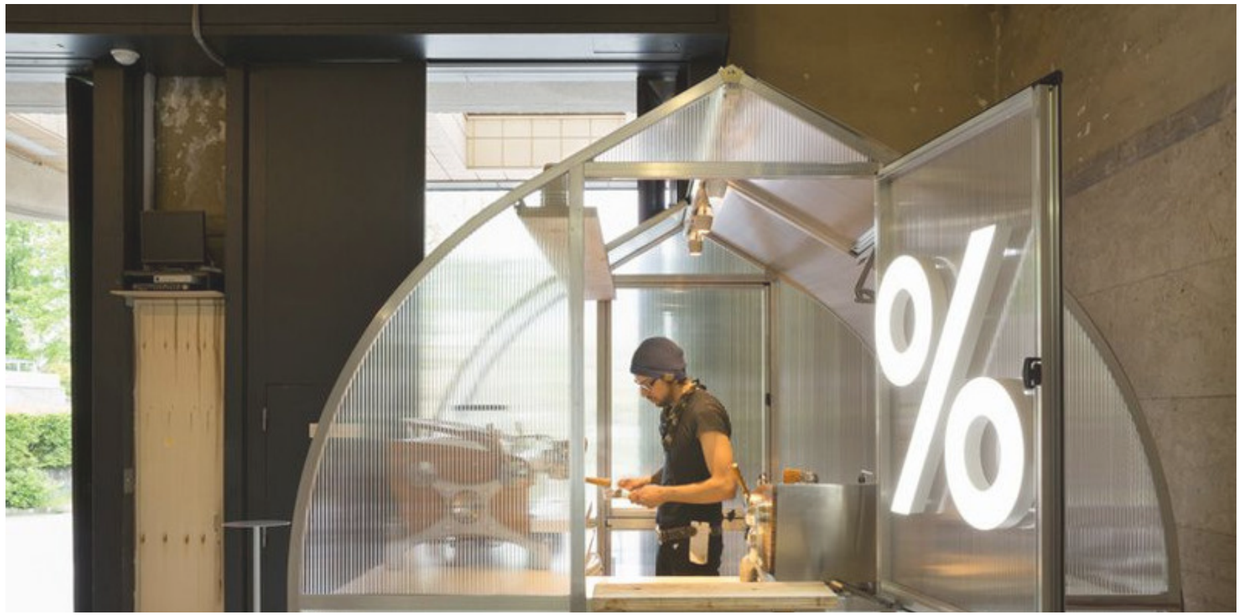
% ARABICA COFFEE %

Hi! I'm raissa!

A data-scientist in training

- Data Squad Member
- Coffee Lover
- Data Engineer & Analyst

What is our current problem?



Our current sales performances were just barely passing the target of the past month.

What else can we do to improve the **monthly** revenue sales?

Machine Learning Objective

Improve customer’s buying rate through **product recommendation** and **personalized advertisement/ promos** from customer segmentation

WHAT ARE THE DATA THAT WE USE?

CUSTOMER DATA

2.245 rows & 8 columns

APRIL SALES DATA

49.894 rows & 14 columns

PRODUCT AND INVENTORY DATA

two tables, 88 rows & 14 cols
307 rows & 7 cols

Source : Coffee Shop Dummy Data from IBM Cognos

Data Cleansing Process

STORE IN SQL WORKBENCH SCHEMA

Pull data with SQL queries to JupyterLab

REMOVE CERTAIN CHARACTERS & FIX DTYPE

Fixing certain data values and types

MERGE & ADD NEW COLUMNS

Binning, slicing, gaining further insights through existing columns

Dashboard step by step process

IMPORT DATA

using MongoDB compass

BUILD WEB-APP

using Flask and connect to MongoDB by pymongo

BUILD DASHBOARD VISUALIZATION

using MongoDB Atlas & EMBED TO WEB-APP

Cleaned and uploaded to MongoDB

10 DBS 28 COLLECTIONS

☆ FAVORITE

Filter your data

- admin
- coffeeshop
  - clustered
  - customer
  - inventory
  - product
  - sales
  - sales_with_day
- local
- sample_airbnb
- sample_analytics
- sample_geospatial
- sample_mflix

Documents

coffeeshop.clustered

Documents | Aggregations | Schema

FILTER

ADD DATA | VIEW


```
_id: ObjectId("60377695bde22baa1c2a47b4")
: "0"
customer_id: "1"
home_store: "3"
customer_name: "Kelly Key"
loyalty_card_number: "908-424-2890"
gender: "M"
birth_year: "1950"
generation: "Baby Boomers"
work: "edu"
member_duration: "2017"
member_month_joined: "2017-01"
office: "adipiscing"
total_spending: "29.199999809265137"
count(line_item_amount): "8"
avg_spending: "3.649999976158142"
member_for: "28"
label: "0"
```


Dashboards

coffee shop

Dashboard for %ARABICA% coffee shop

- Customer's Office by Gender
- Product Waste
- Sales Performance

+7 more charts — Modified an hour ago

coffeeshop.product

DOCUMENTS 88 | INDEXES 1

Documents | Aggregations | Schema | Explain Plan | Indexes | Validation

FILTER | OPTIONS | ANALYZE | RESET

Query returned 88 documents. This report is based on a sample of 88 documents (100.00%).

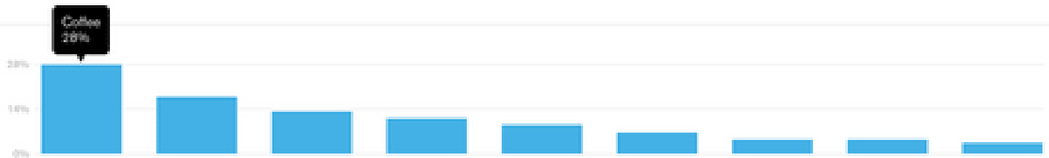
product

string

Almond Croissant, I Need My Bean! Diner mug, Brazilian - Organic, Latte Rg, Earl Grey Lg, Cranberry Scone, Traditional Blend Chai Lg, Peppermint Lg, Columbian Medium Roast Sm, Lemon Grass Rg, Our Old Time Diner Blend Lg, Our Old Time Diner Blend Rg, Brazilian Rg, Pumpkin Spice Latte, Ethiopia, Espresso shot, Jamaican Coffee River, Cappuccino, Oatmeal Scone, Peppermint

product_category

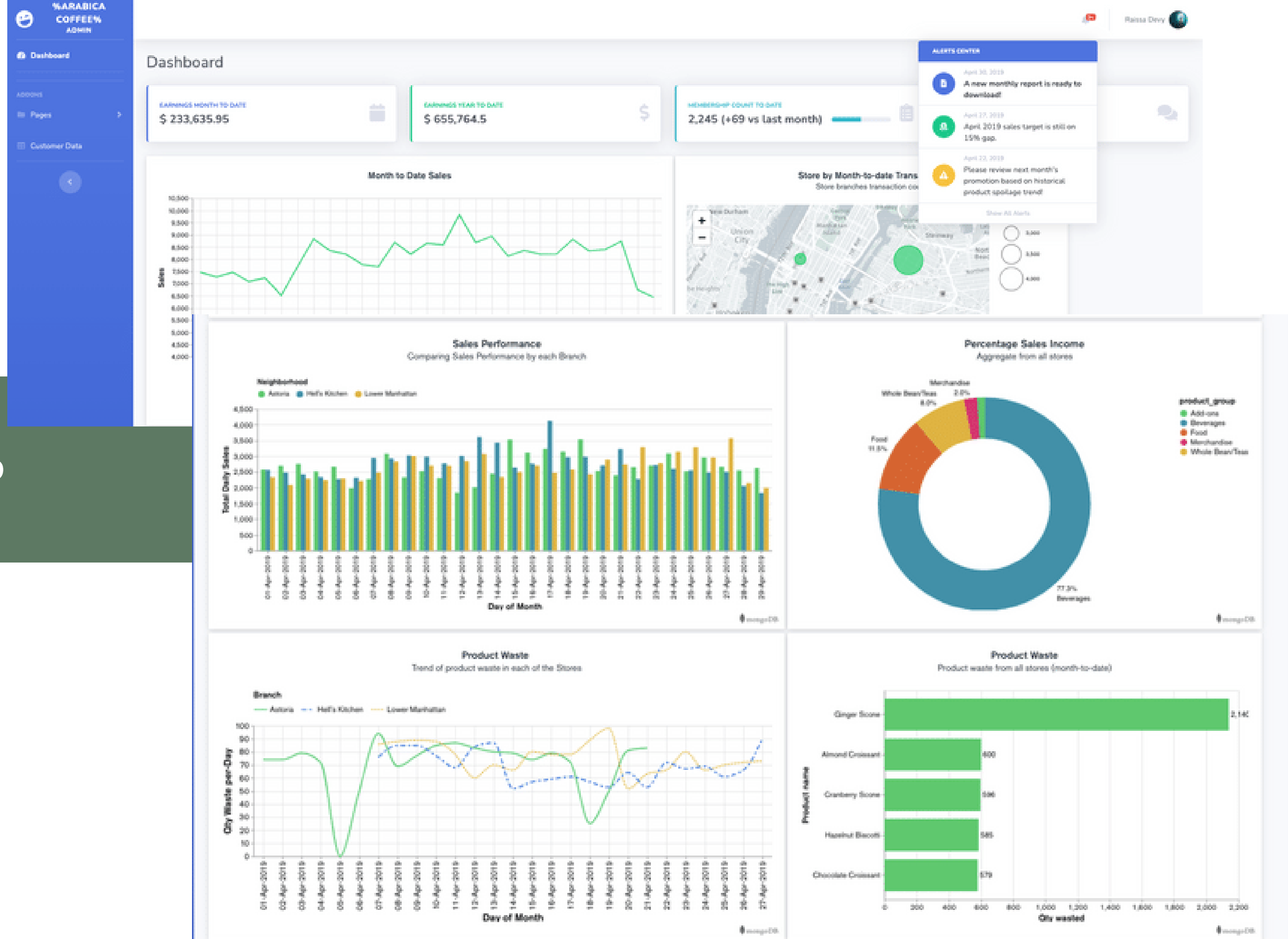
string



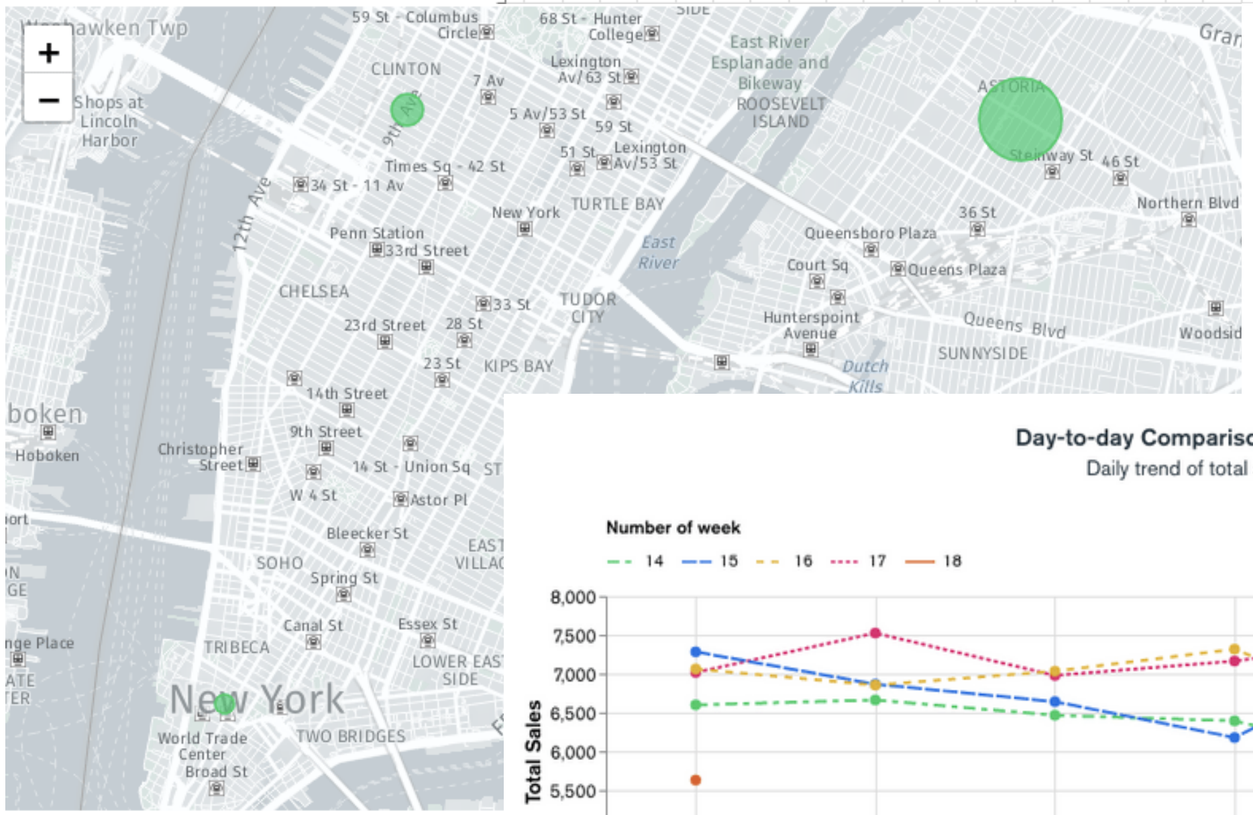
product_description

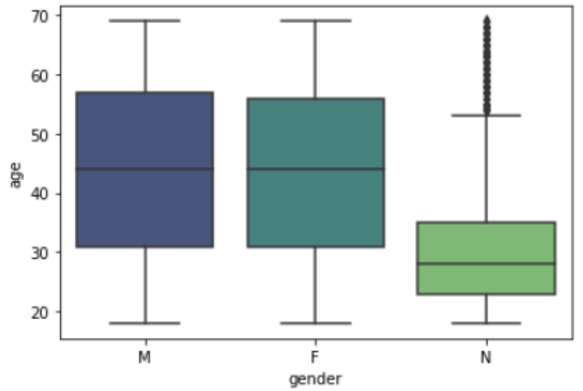
string

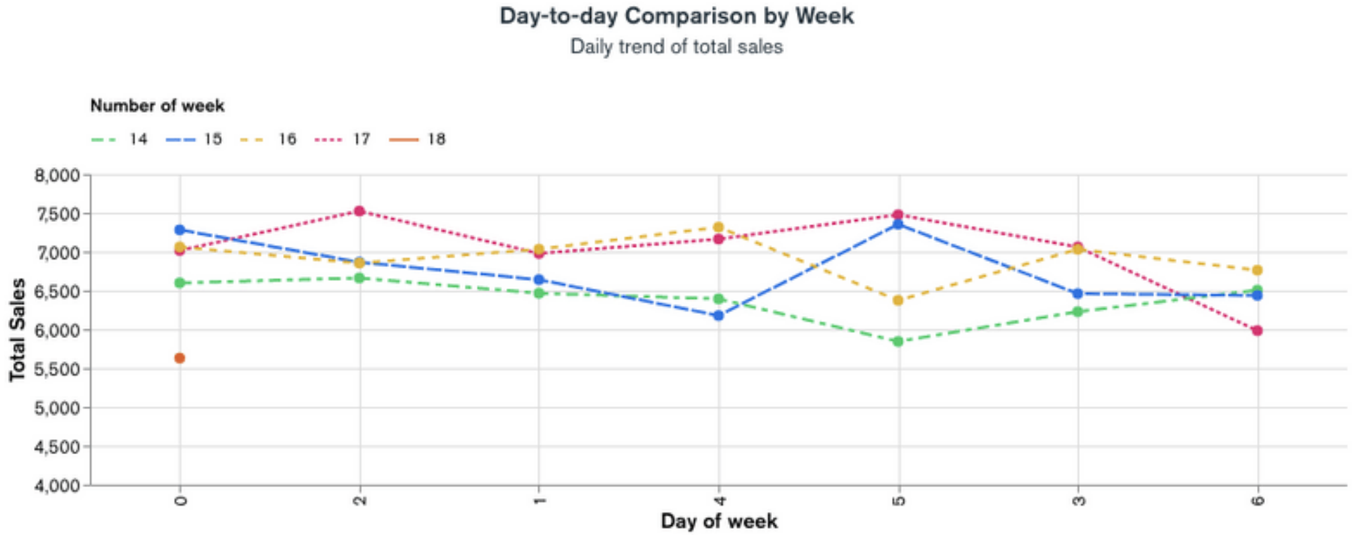
DASHBOARD


%ARABICA COFFEE%
ADMIN
Dashboard
Pages
Customer Data
Dashboard
EARNINGS MONTH TO DATE
$ 233,635.95
EARNINGS YEAR TO DATE
$ 655,764.5
MEMBERSHIP COUNT TO DATE
2,245 (+69 vs last month)
ALERTS CENTER
A new monthly report is ready to download!
April 2019 sales target is still on 15% gap.
Please review next month's promotion based on historical product spoilage trend
Month to Date Sales
Sales
Store by Month-to-date Trans
Store branches transaction co
Sales Performance
Comparing Sales Performance by each Branch
Neighborhood
Astoria
Hell's Kitchen
Lower Manhattan
Total Daily Sales
Day of Month
Percentage Sales Income
Aggregate from all stores
Merchandise
Whole Bean/Teas
8.0%
2.0%
Food
11.5%
77.3%
Beverages
product_group
Add-ons
Beverages
Food
Merchandise
Whole Bean/Teas
Product Waste
Trend of product waste in each of the Stores
Branch
Astoria
Hell's Kitchen
Lower Manhattan
Qty Waste per-Day
Day of Month
Product Waste
Product waste from all stores (month-to-date)
Ginger Scone
Almond Croissant
Cranberry Scone
Hazelnut Biscotti
Chocolate Croissant
Product name
Qty wasted

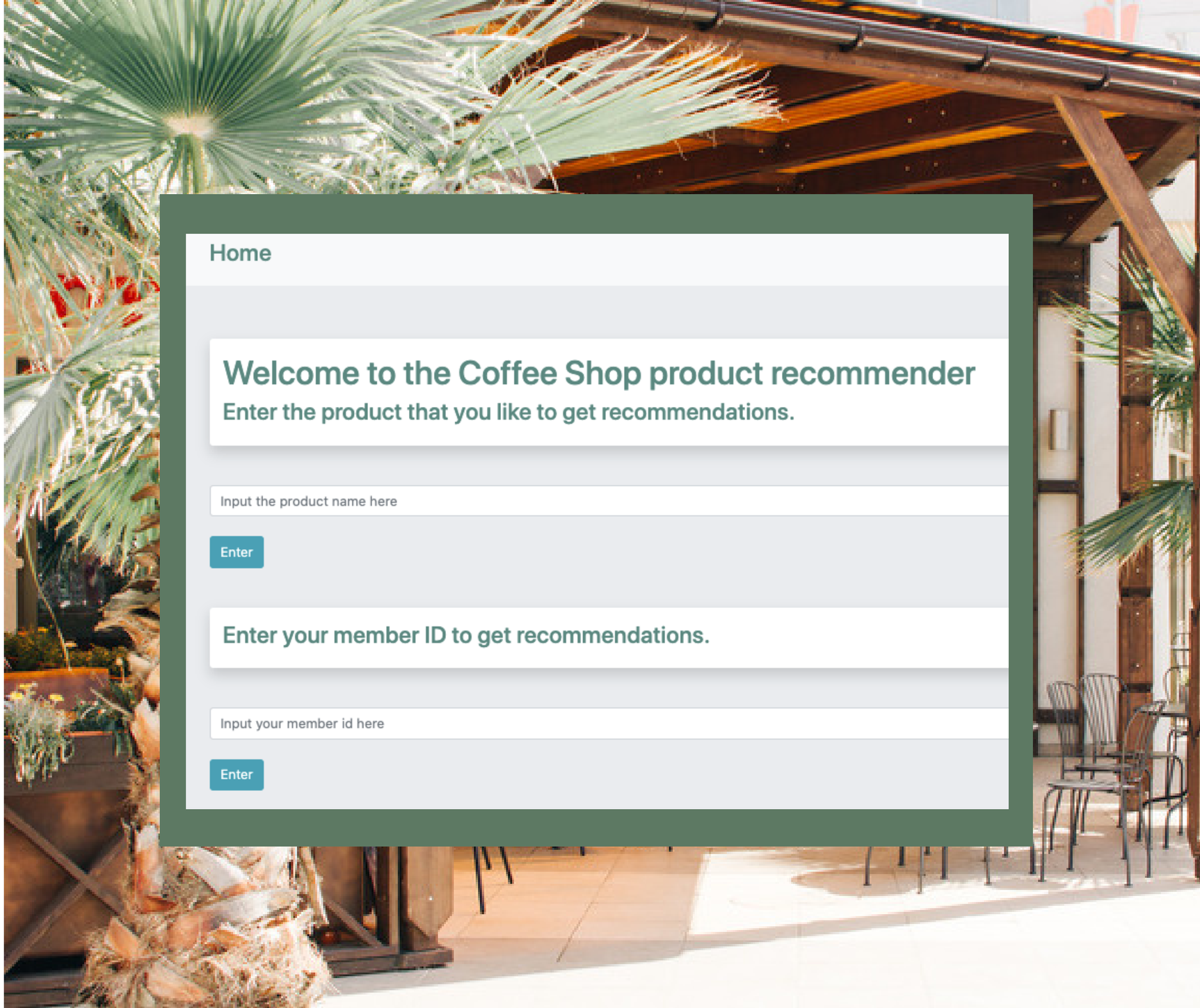
Data Insights





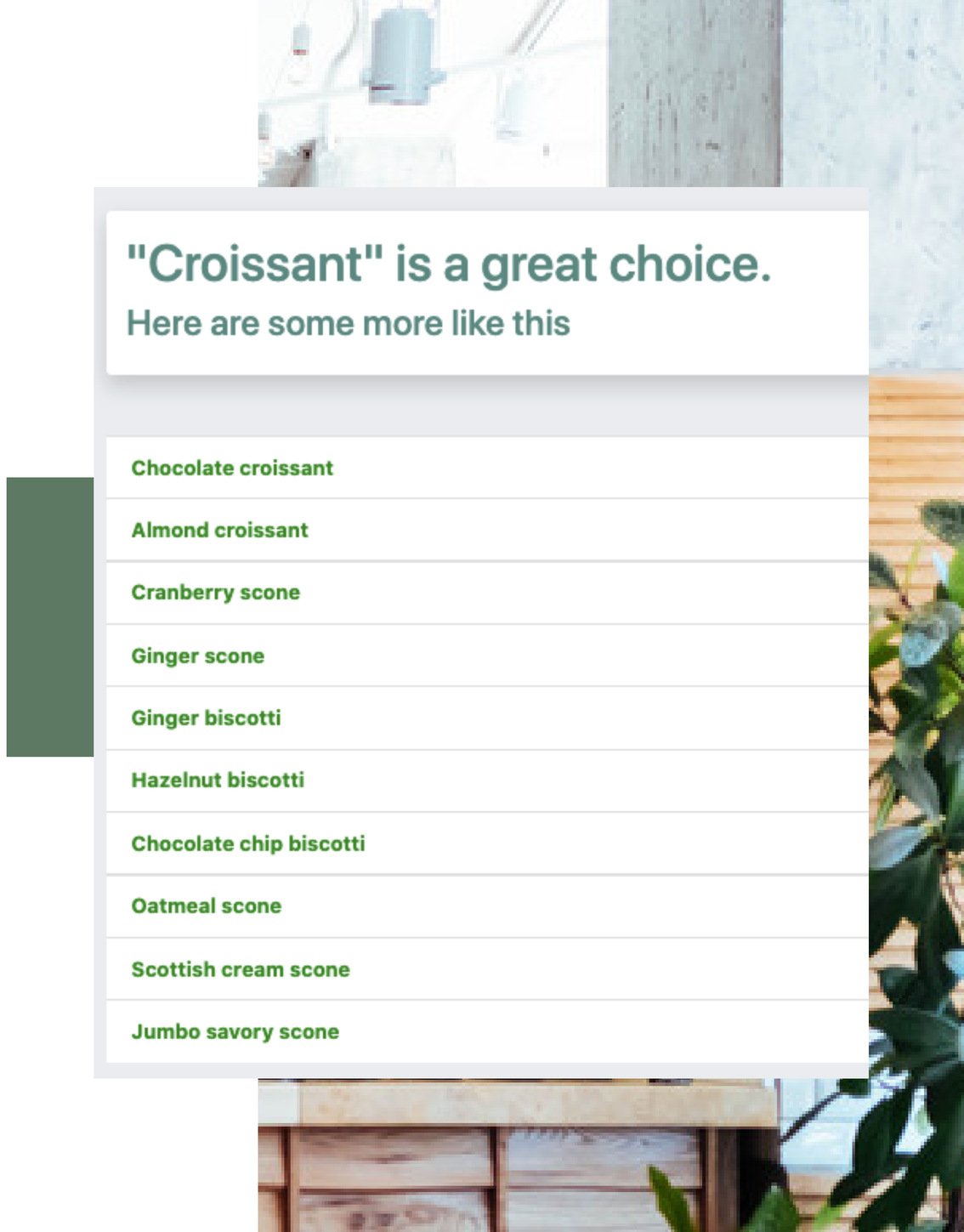


Product Recommender Systems



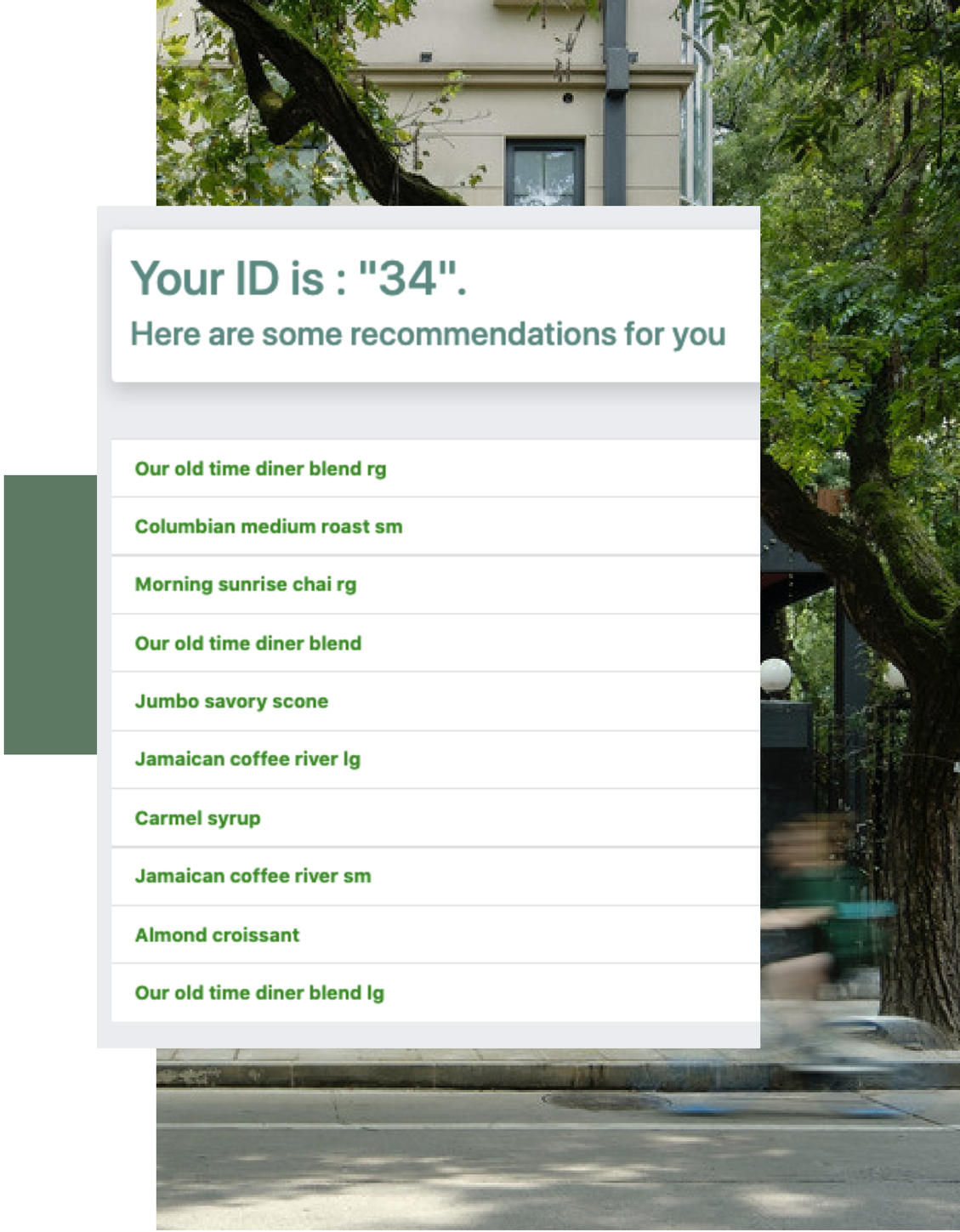
Content based recommender

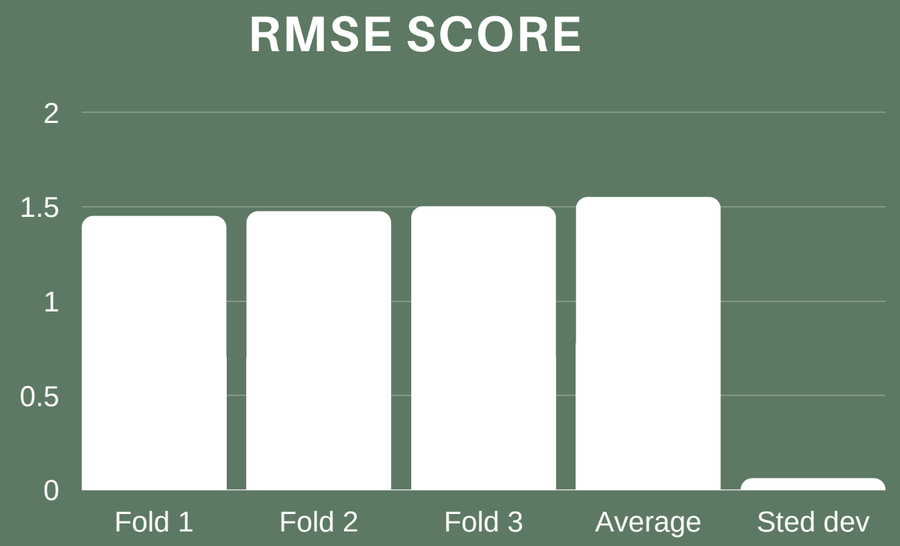
RECOMMEND CUSTOMER BASED ON PRODUCT SIMILARITIES



Collaborative filtering recommender

RECOMMEND BASED ON CUSTOMER'S HISTORICAL PURCHASE SIMILARITIES





Using RMSE & MAE as Validation Metrics

Best model : KNN-basic()

CROSSVAL 3 FOLDS

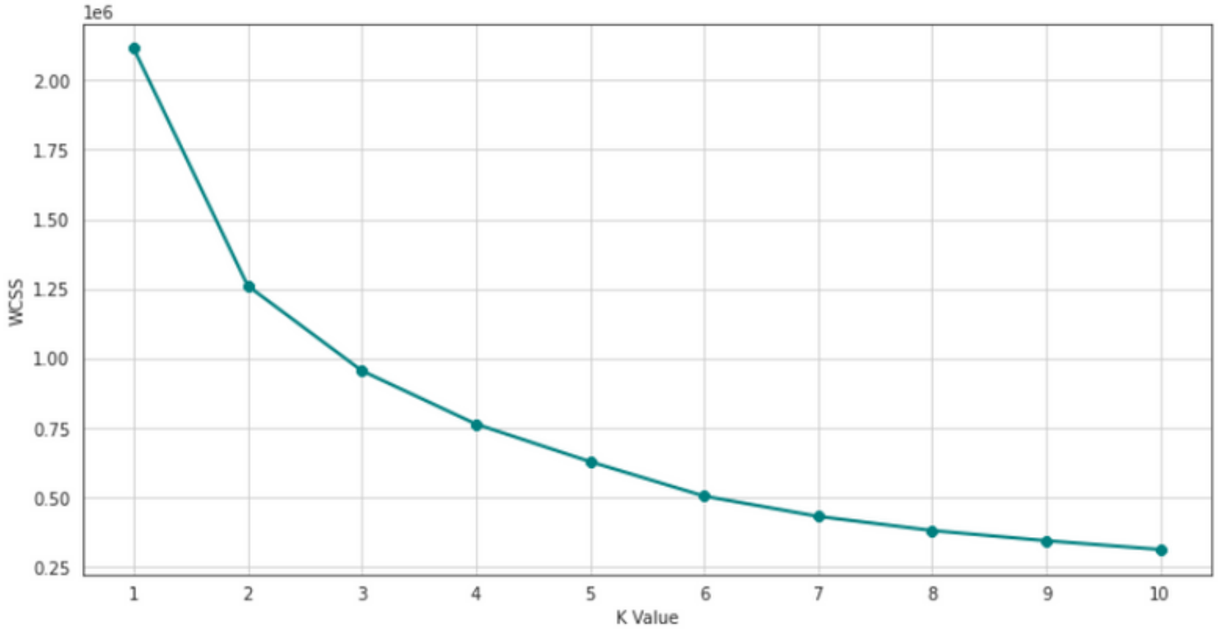
The 4 Clusters

Customer Segmentation using K-means

FOR PERSONALIZED MARKETING & BUNDLING PROMOS

Elbow method

BEST **K** = 4




```
for k in range(2, 7):
    kmeans = KMeans(n_clusters=k)
    kmeans.fit(X)
    label=kmeans.predict(X)
    print(f'Silhouette Score(n={k}): {silhouette_score(X, label)}')
```

```
Silhouette Score(n=2): 0.3899231849963964
Silhouette Score(n=3): 0.342641289693967
Silhouette Score(n=4): 0.33982712460825987
Silhouette Score(n=5): 0.3412836098154714
Silhouette Score(n=6): 0.34220128383774834
```

FINDING K - VALUE

CLUSTER 0

Old people. Average spending.
New members.

CLUSTER 1

Low-spenders.
Mixed age. Mixed membership.

CLUSTER 2

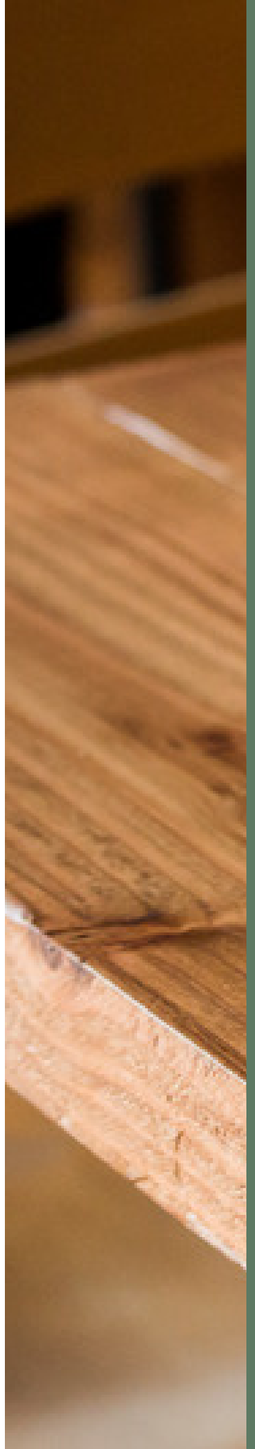
Young people. Average spenders.
New members.

CLUSTER 3

The high-spenders.
Mixed age. Mixed membership.

CUSTOMER CLUSTERS

Business Impact Calculation



15% PROMOTIONAL EMAIL CONVERTION RATE

As of historical campaigns, customers will respond to the promotion as high as 15% of all the total members

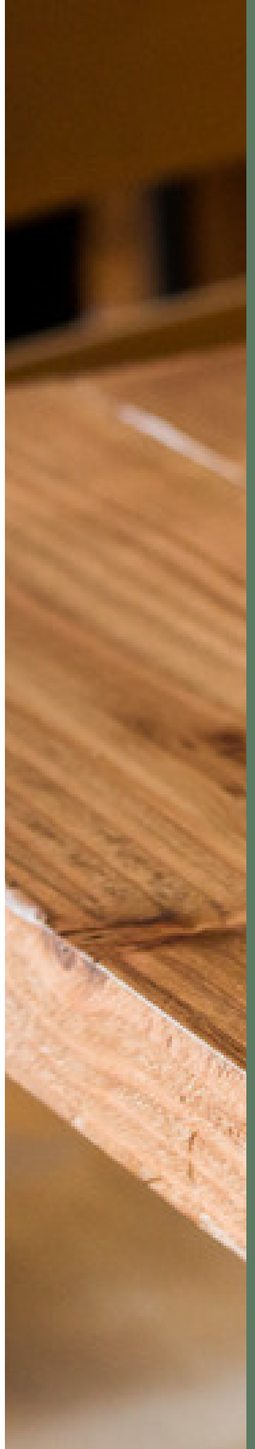
Total members 2.245 account.
15/100 * 2245 = 336 customers will respond to the promotion

```
[172]: avg_transaction = df_avg.groupby(['new_trans_id']).sum()
[173]: avg_transaction['line_item_amount'].mean()
[173]: 5.8277631646021355
```

Average value for each transaction $ 5.8

So the opportunity is to gain :
5.8 * 336 = **$ 1948,8/month**

Business Impact Calculation



Personalized content increases the efficiency of marketing spending by 10 to 30 percent.

MCKINSEY & COMPANY - 2019

Thus we are able to cut marketing budget up to $ 300/month

MARKETING BUDGET :
$ 1000 / MONTH

Future Improvements

IMPROVE UI/UX DESIGN

For recommender systems

UTILIZE MARKET BASKET ANALYSIS & ADDING NON-NUMERICAL INSIGHTS

For bundling promos & deepen customer segmentation using categorical insights

ADDING MORE DATAS

From the previous months to capture more insights

ANY QUESTIONS?

Thank you!