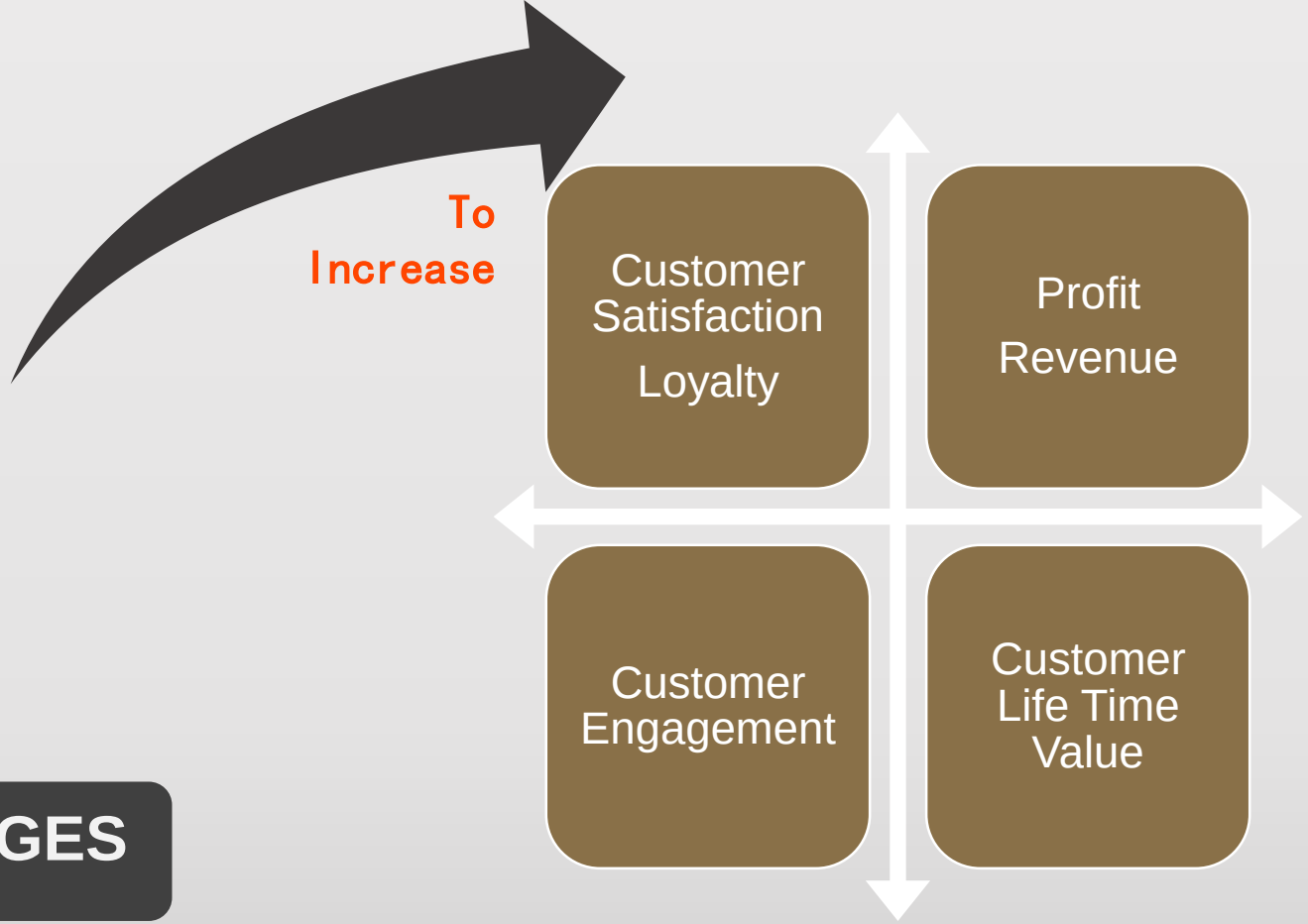
Customer Loyalty and Recommender System

Keyur Patel
Data Science Project
March 2019

Problem Statement

- For any company, customer loyalty and retention is critical to the business. With e-commerce boom, competitive advantage and more personalized experience for the stakeholders are needed.
- With increasing numbers of brands, growing number of users and changing environment, it is important to get insights into customers, product basis.
- Business decisions influenced by analytics can drive marketing efforts to increase customer retention, build loyal relationship with Users, and increase revenue and User engagement.
- Users have a huge choice of products to purchase but limited time. The real challenge is to provide recommendations of products that are relevant to the users, help users discover brands that they might never heard before or brands they might not know they would like.
- Filtering brand from entire catalog of brands which are relevant to the users is basically the key focus.

BUSINESS CASE



BUSINESS CHALLENGES

- Cost – of resources to maintain huge data
- Quality - services and tools
- Competition:

  - From Market Segment
  - Other Sources

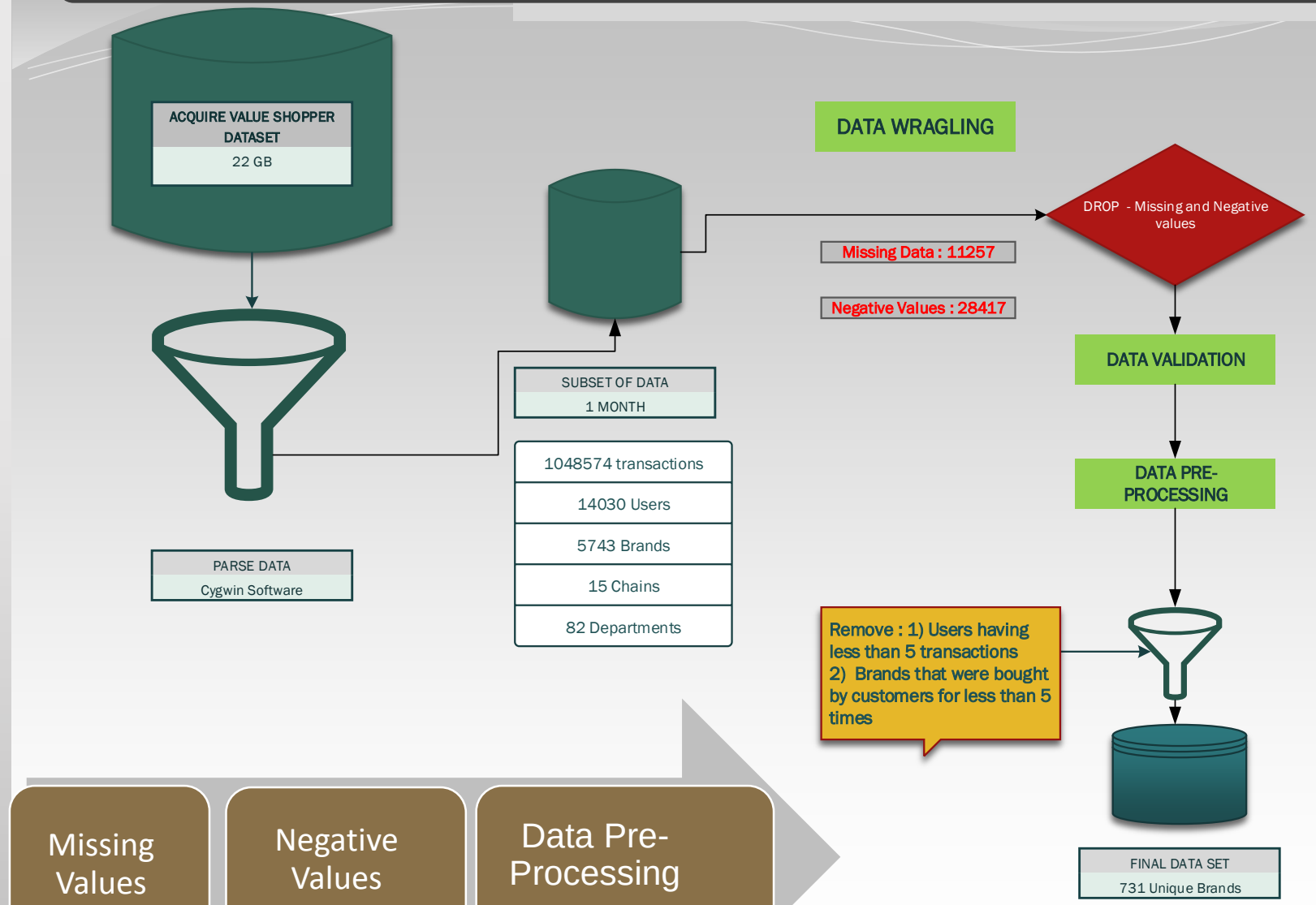
DATA SET

- The dataset from Kaggle’s Acquire Valued Shoppers Challenge.
- A year of shopping history of customers.
- There are 1048574 rows and 11 columns.

DATA CHALLENGE

- Huge data
- Need sufficient hardware resources to process 22GB data
- A subset of transaction data selected for analysis (one month).

How data is acquired, cleaned and Pre-processed?



FEATURE ENGINEERING

FINAL DATA SET

731 Unique Brands

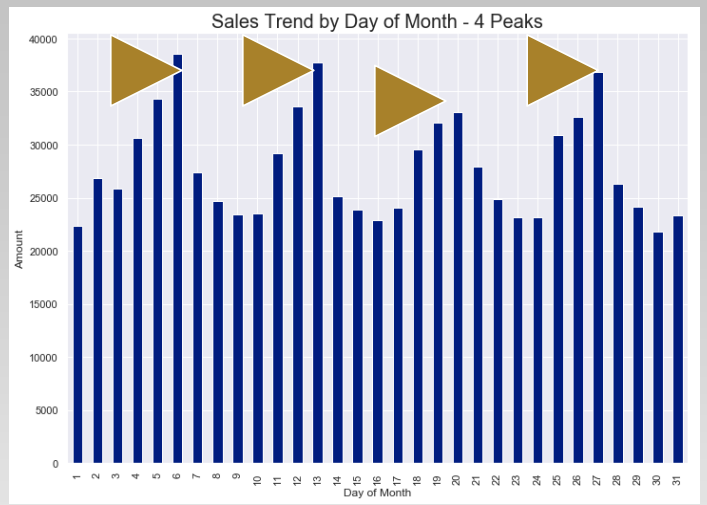
FEATURE ENGINEERING: Day, Weekday, Week

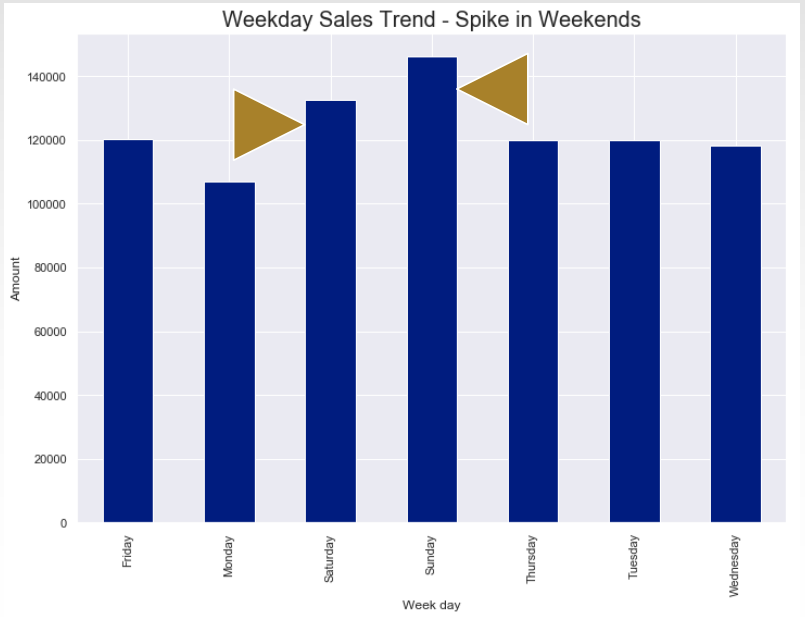
ANALYSIS

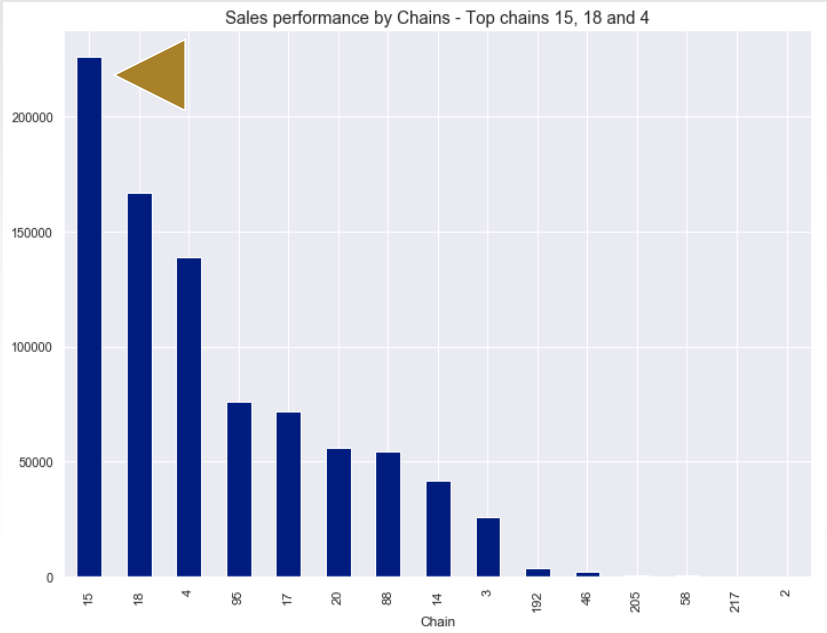
Day	6 TH January Maximum sales followed by 13th and 27th
Week	Maxiumum Sales on 4th Week and 3rd Week lowest Sales
Weekday	Maximum Sales during Weekends, and lowest sales on Monday
Chain	Chain # 15 provides highest Sales followed by 18 and 4. Chain# 2 provides least sales volume

Data Story


Weekly Sales Trend- 4th Week Maximum Sales
Amount
Week
0
50000
100000
150000
200000
250000
1
2
3
4

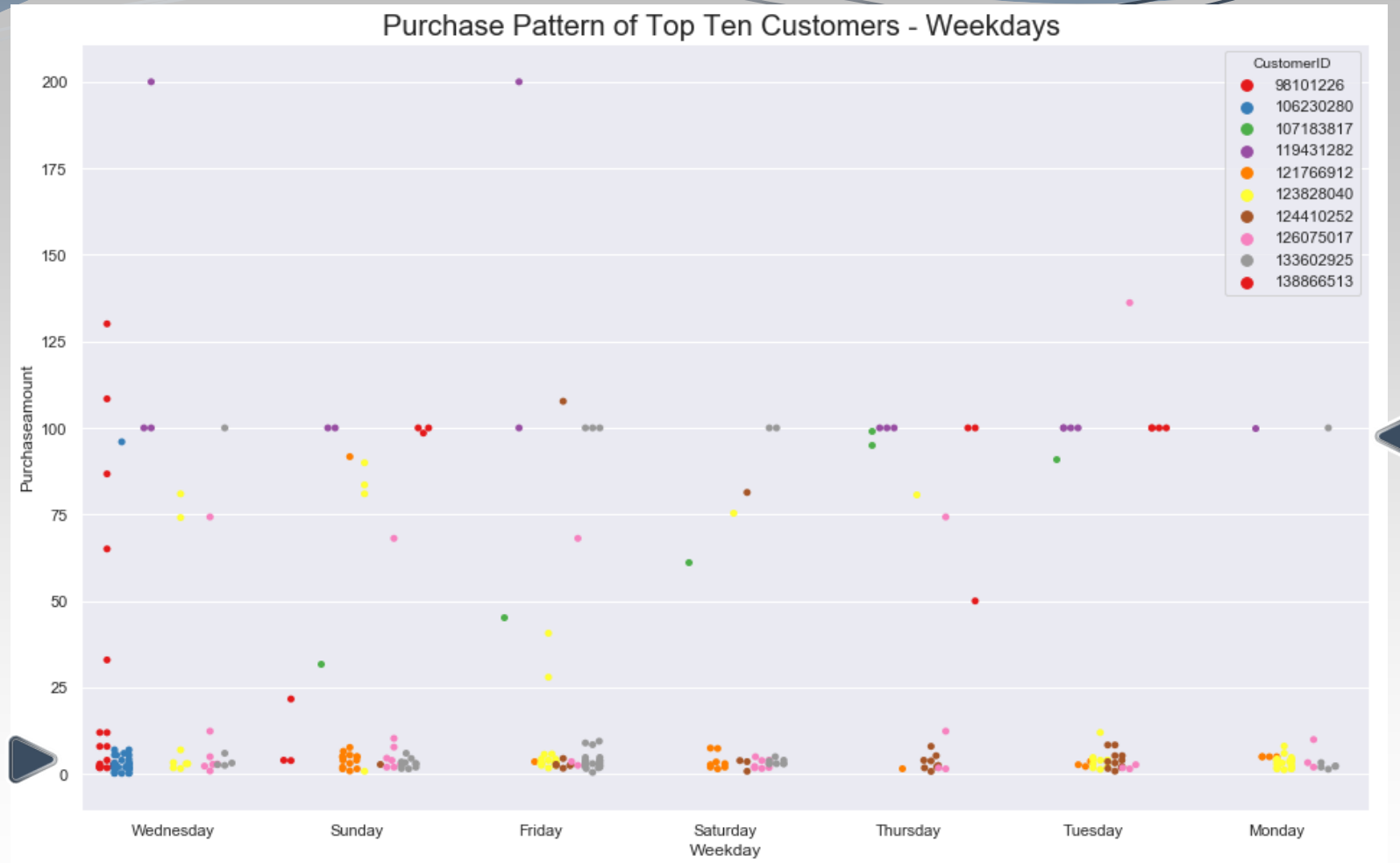

Sales Trend by Day of Month - 4 Peaks
Amount
Day of Month
0
5000
10000
15000
20000
25000
30000
35000
40000


Weekday Sales Trend - Spike in Weekends
Amount
Week day
0
20000
40000
60000
80000
100000
120000
140000
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday


Sales performance by Chains - Top chains 15, 18 and 4
Chain
0
50000
100000
150000
200000
15
18
4
96
17
20
88
14
3
192
46
205
58
217
2

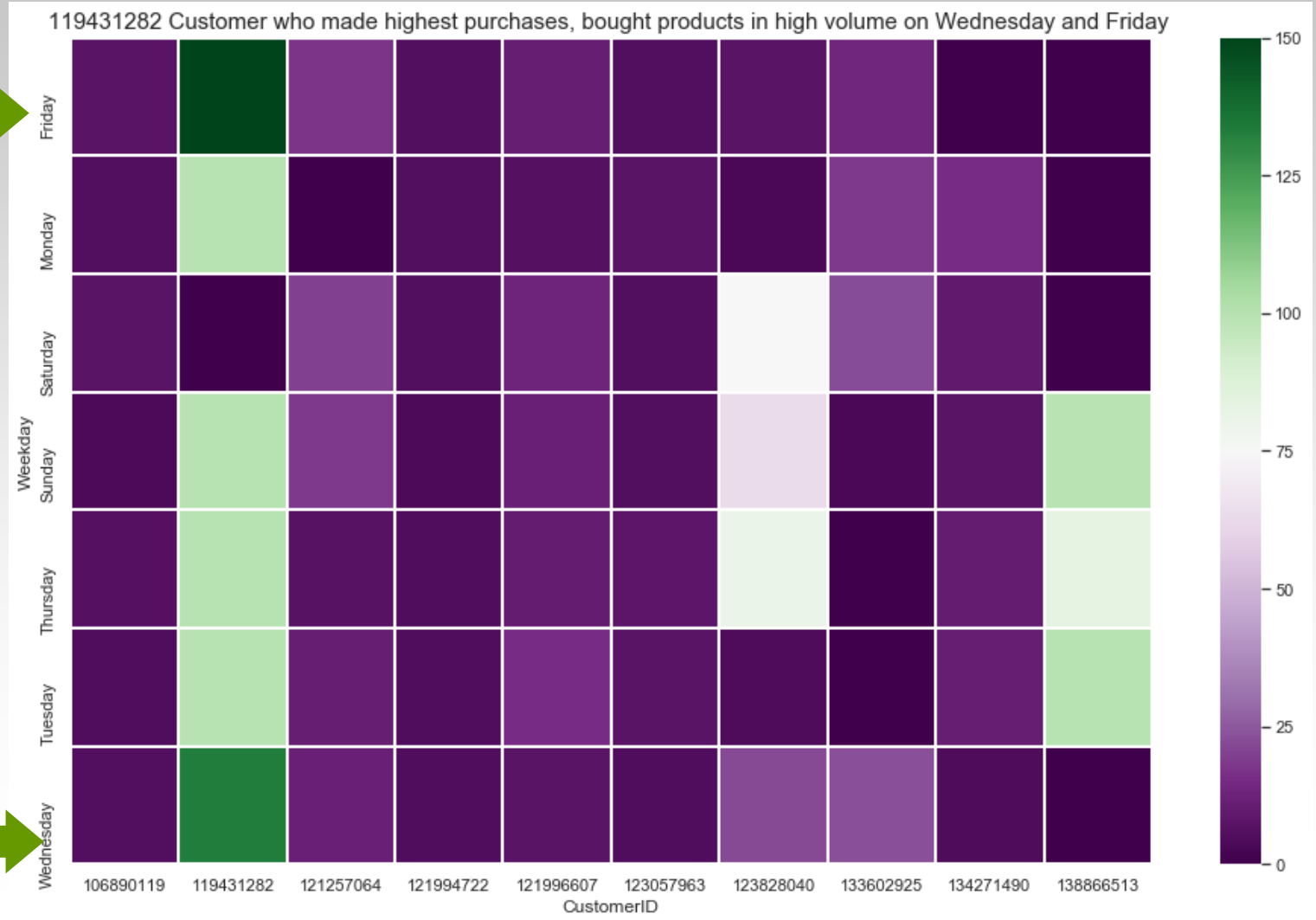
Exploratory Data Analysis

- **Top 10 customers** with high purchase amount/transaction
- **Total 322** transactions
- Duration – **1 month**



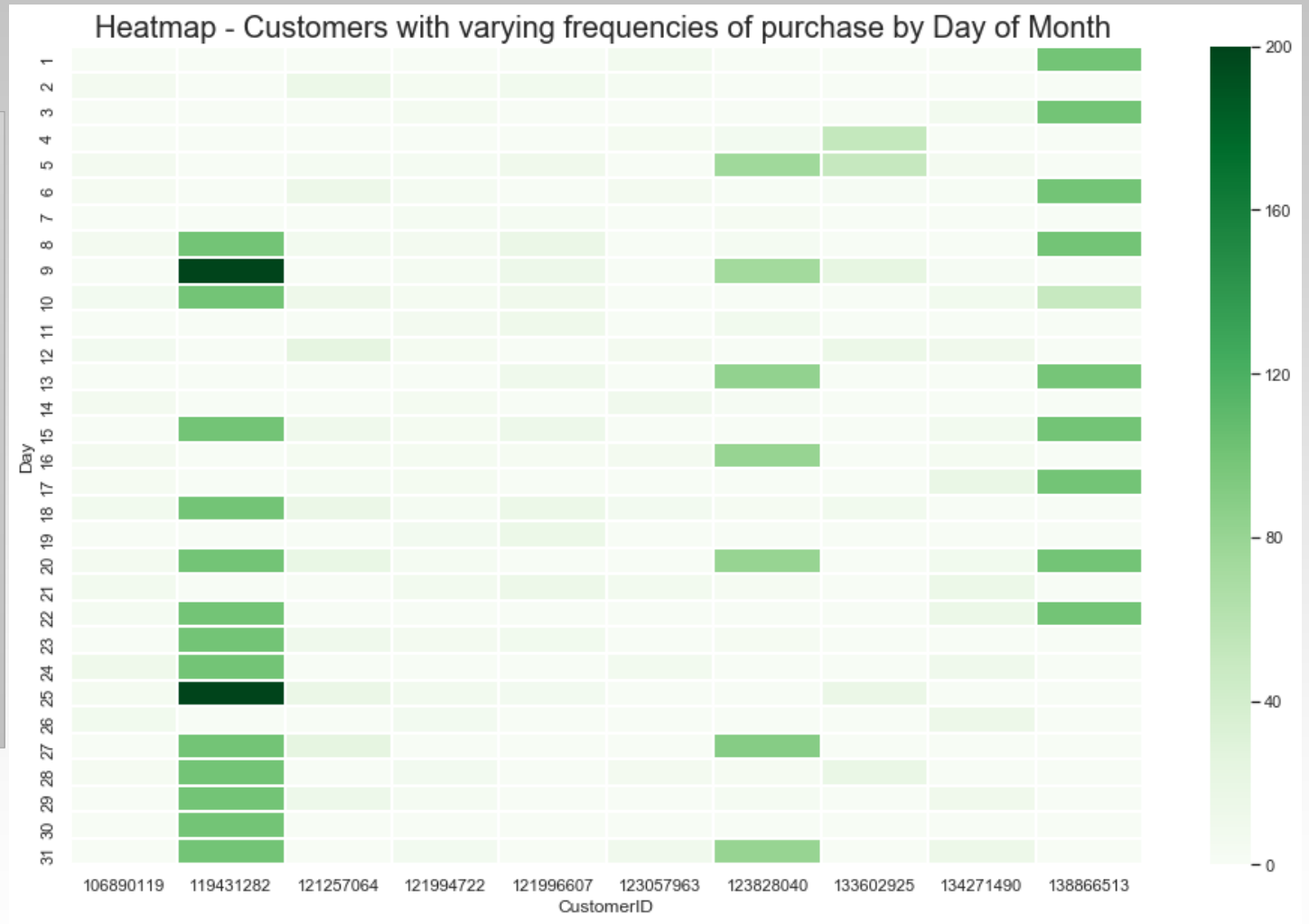
Exploratory Data Analysis

- **Purchase Pattern** by day of the Week
- High sales on Wednesday and Friday



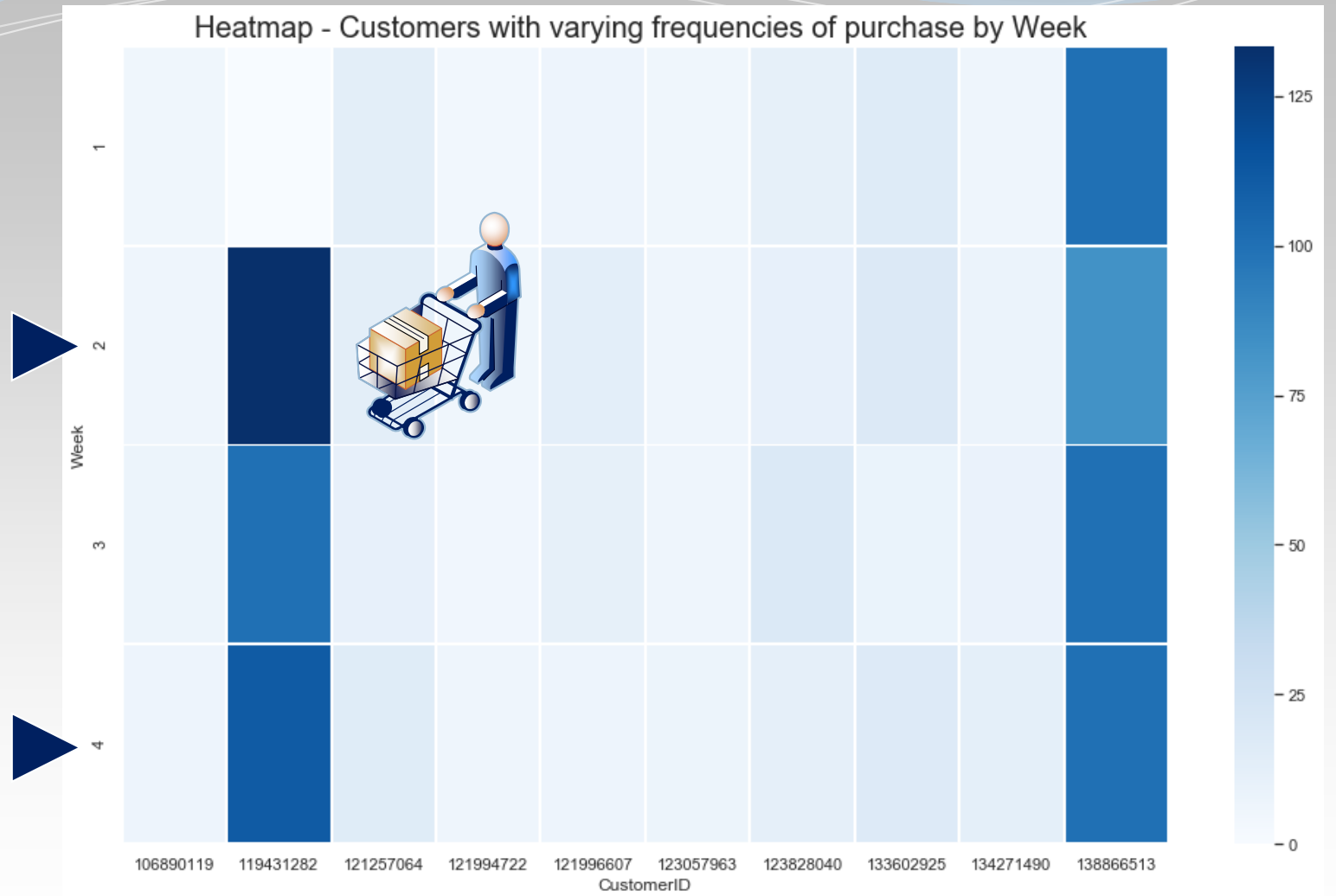
Exploratory Data Analysis

- **Top 10 customers** purchase trend
- **Low, medium and High** frequencies of purchase
- High Sales on 9th and 25th

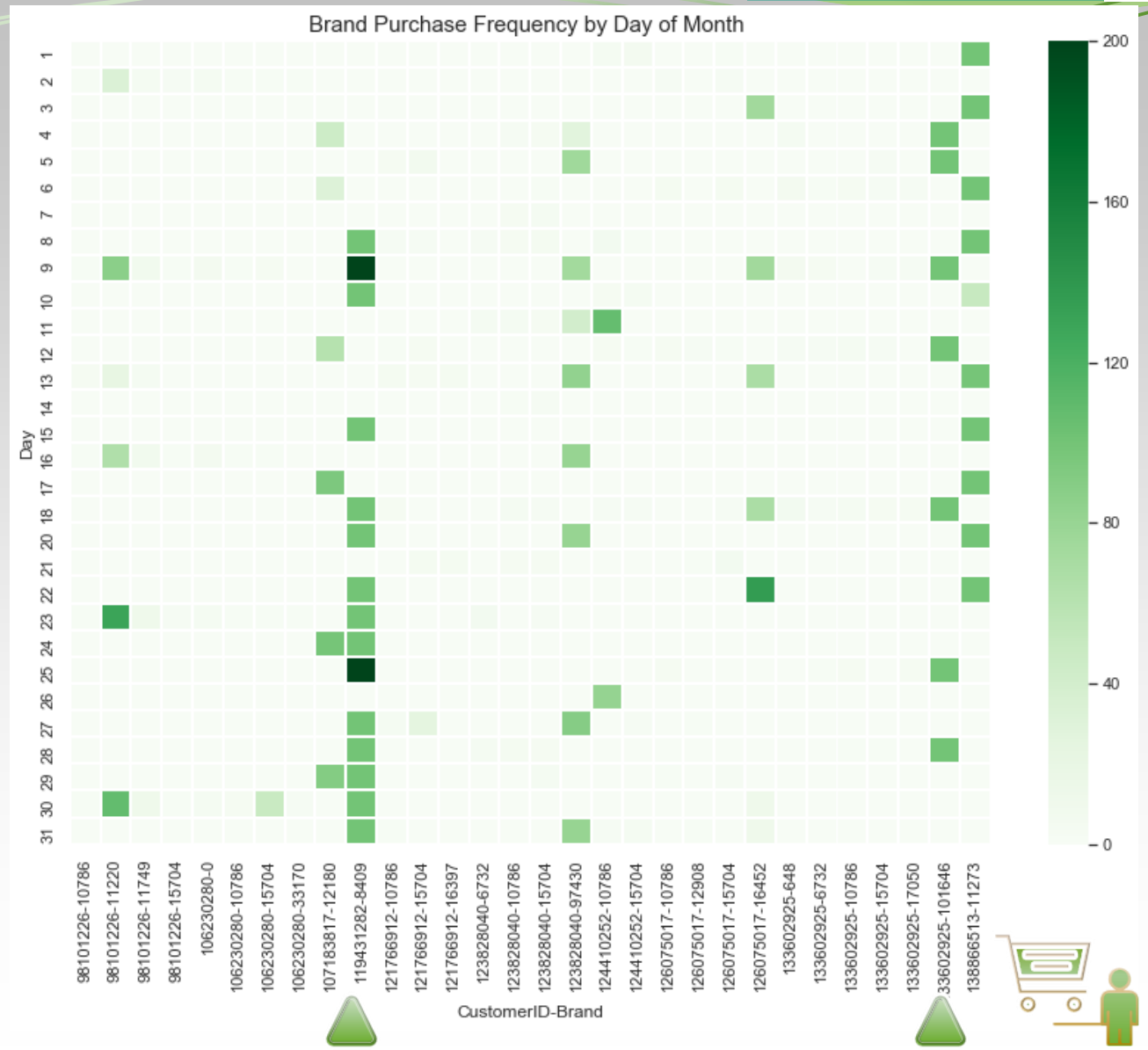


Exploratory Data Analysis

- **Purchase Pattern** by Week
- High Sales during 2nd and 4th Week

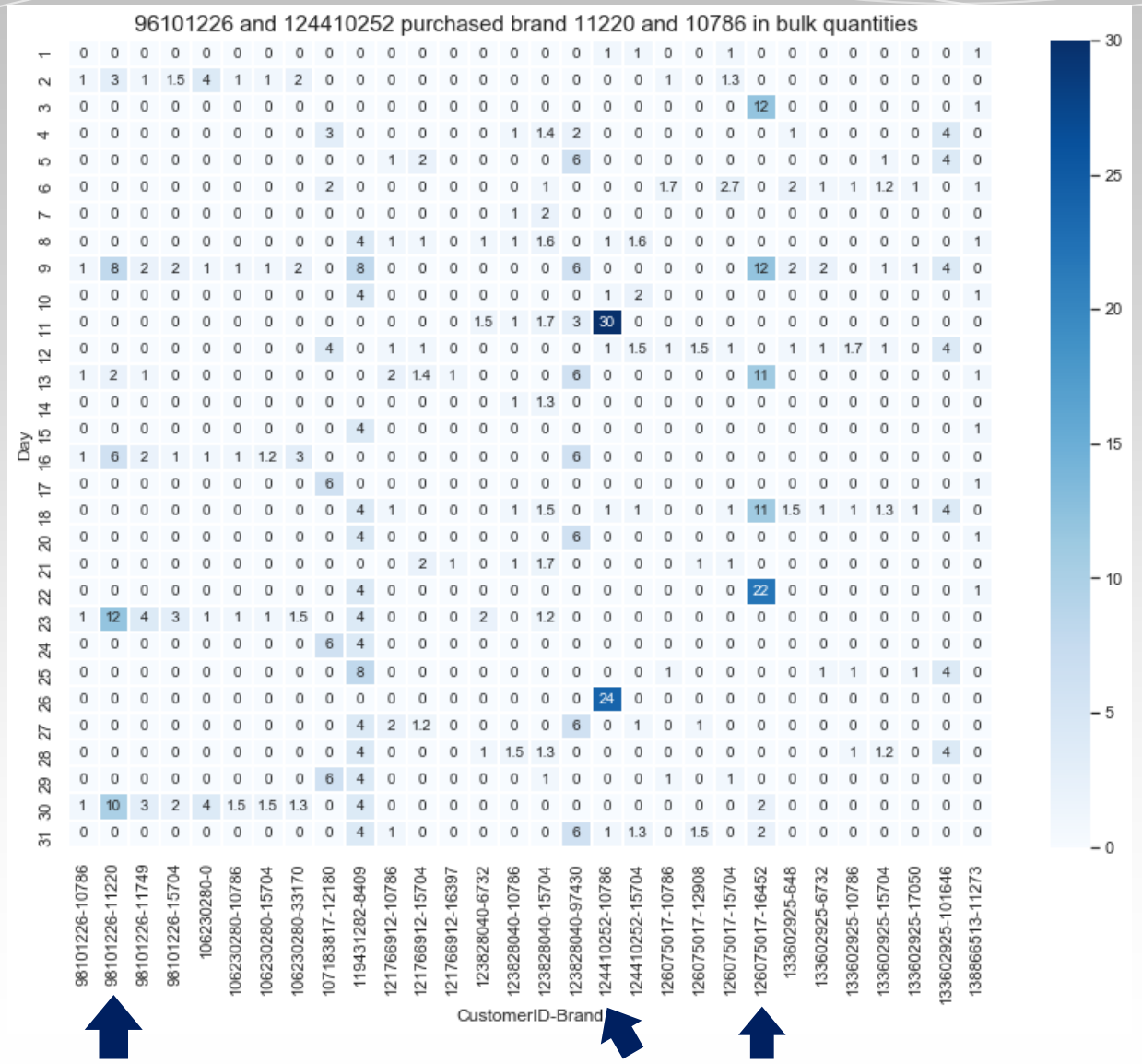


- **Top 10 customers** purchase trend
- **Customer preferences for brands**



Exploratory Data Analysis

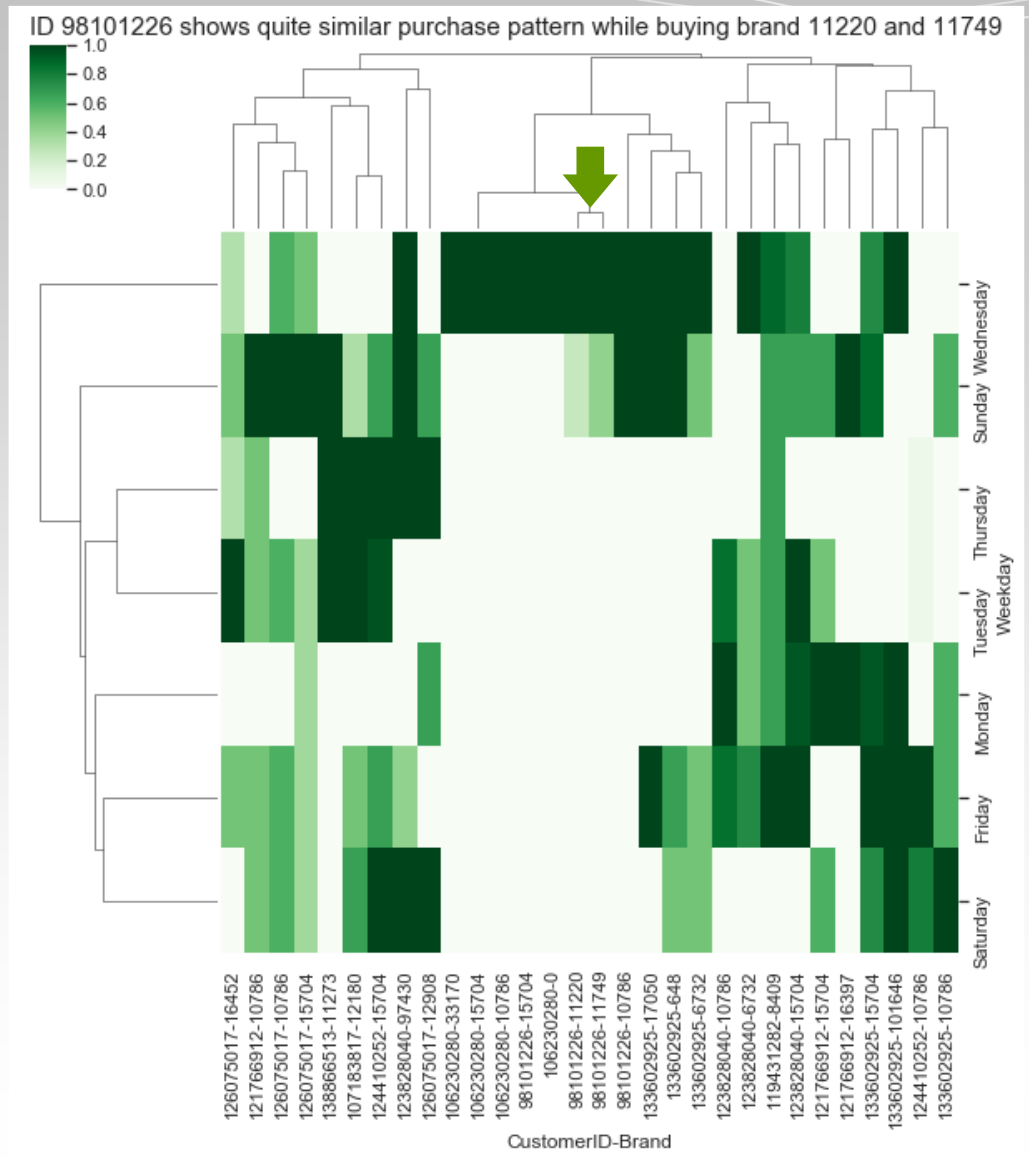
- **Top 10 customers**
- Customer-Brand Purchase behavior by volume



- **Top 10 customers**
  Dendrogram

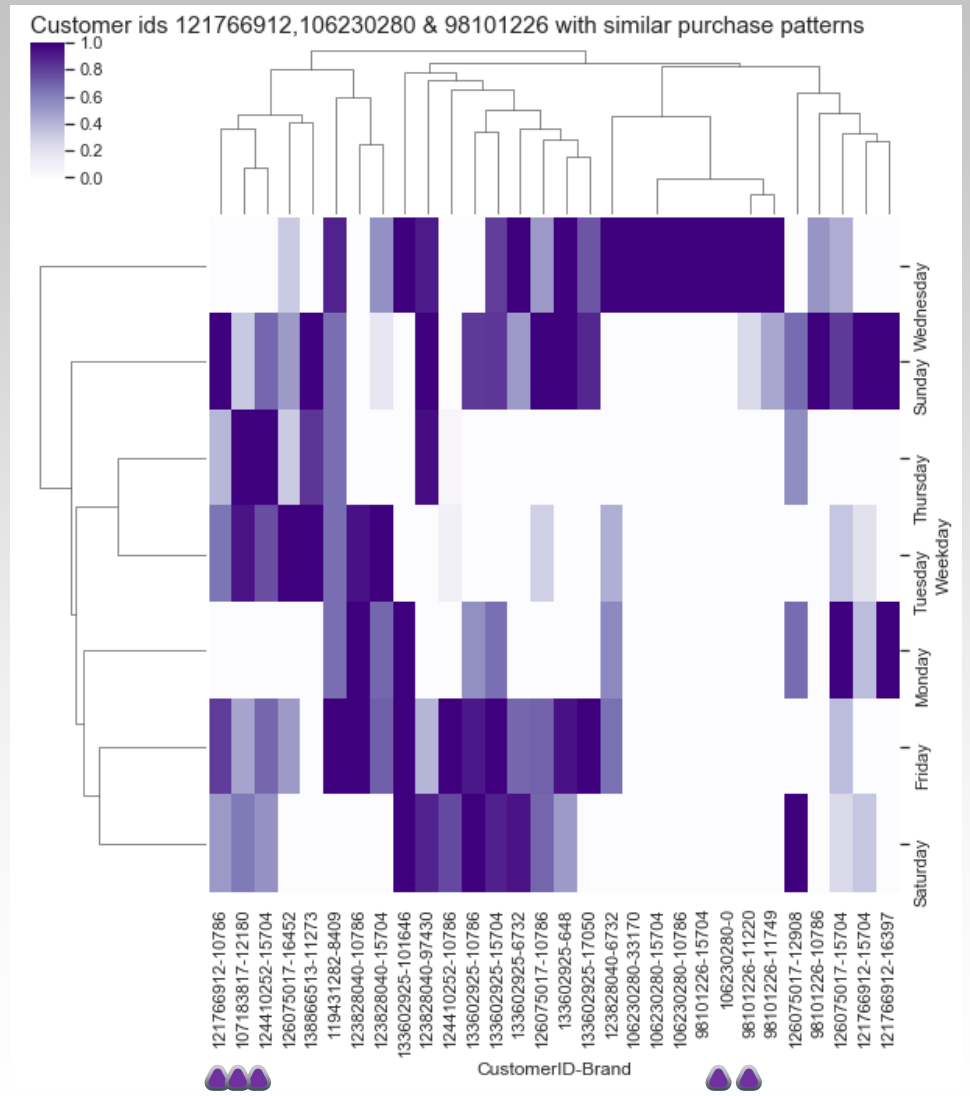
- **Similarity in purchase behavior**
  by Day of Week

A dendrogram represents hierarchical clustering, it illustrates the arrangement of the clusters produced by the corresponding analyses



- **Top 10 customers**
  Dendrogram

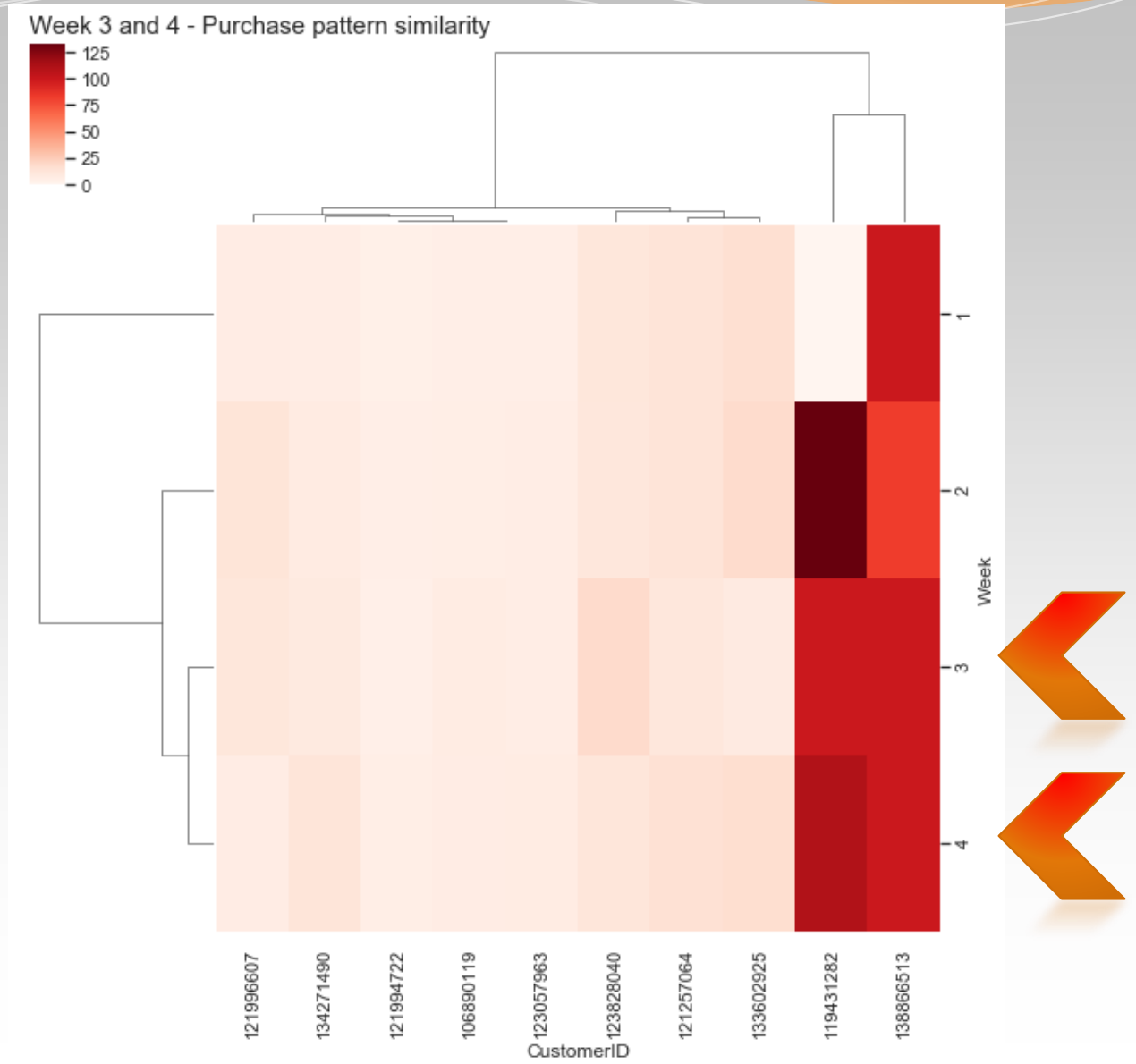
- **Similarity in purchase behavior**
  by Day of Week



Exploratory Data Analysis

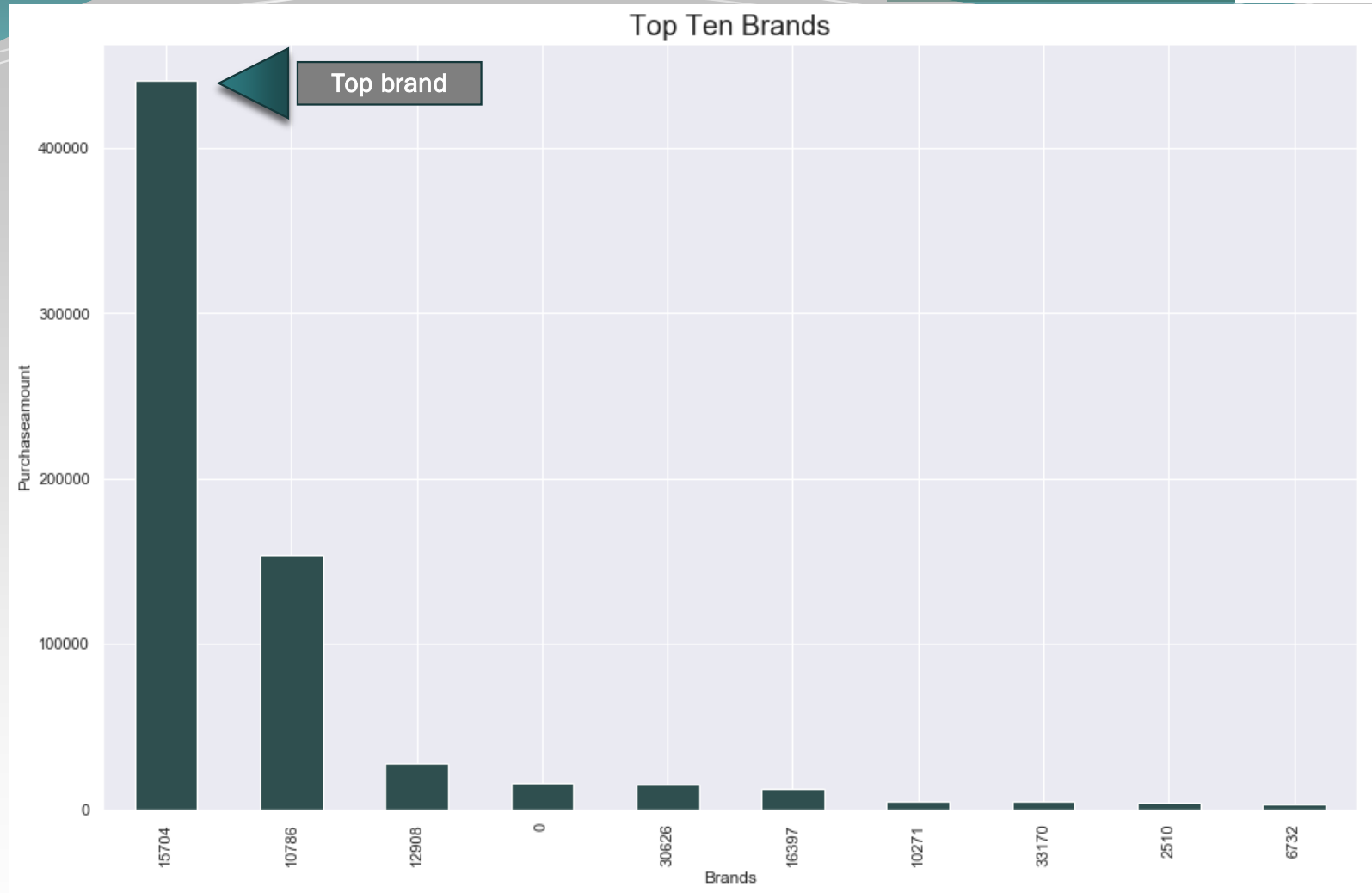
- **Top 10 customers**
  Dendrogram

- **Similarity** in purchase behavior by Week

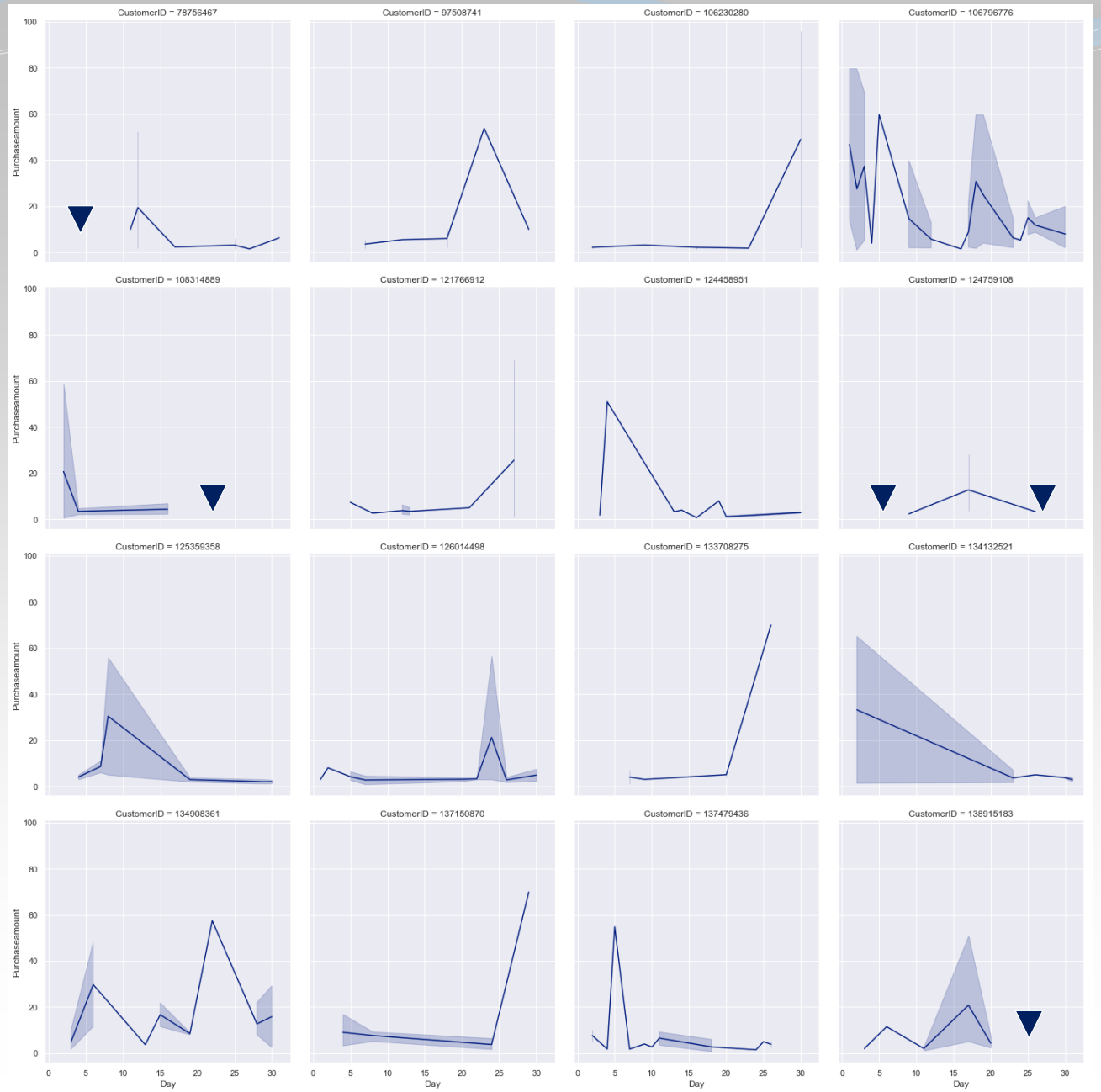


- **Sales by Top Brands**

- Top Brand # 15704



Analysis of Purchase Pattern of Customers of Top Brand # 15704



Improvement Opportunities: Which customer's of brand # 15704 had low sales ?

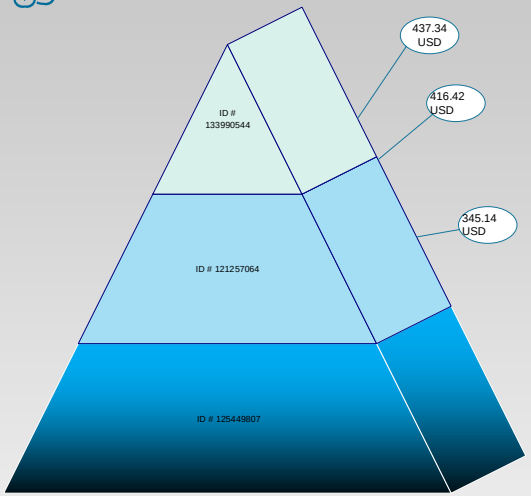
SALES INSIGHTS

81.7 %
Customers Prefer to buy during Weekends

Customers who exclusively buy in Weekends
353

Customers who exclusively buy in the Weekdays
1693

Customer Id 119431282 spent highest average amount per transaction - 113.33 USD



Weekend – Top 3 Transactions

What are the assumptions?

- Products purchased by a customer on a given day are a part of single transaction. That means, user visited the shop once in a single day and purchased the products once.
- Each Brand of a category from a specific department and developed by a company is a unique product.

How can we find a list of products bought in a single transaction?

Unique ID’s

- Create Unique transaction ID’s by concatenating 'CustomerID','Chain' and 'Day' together.

- Create Unique Product ID’s by concatenating 'Dept', 'Category', 'Company' and 'Brand' together separated by dash.

What is Clustering?

- Unsupervised machine learning approach
- Objective - minimize the distance between points in a cluster and maximize the distance between clusters
- Distance between points in a cluster is Within-cluster sum of squares (WCSS)
- The perfect clustering solution – minimum WCSS
- Use elbow method to select optimum number of clusters
- Each Brand of a category from a specific department and developed by a company is a unique product.

How to achieve ?

- Cluster products based on Chain, Dept., Category and Brand
- Goal – to find similar products in a cluster

How clustering is tied with business objective ?

- Design better upselling, cross-selling offers

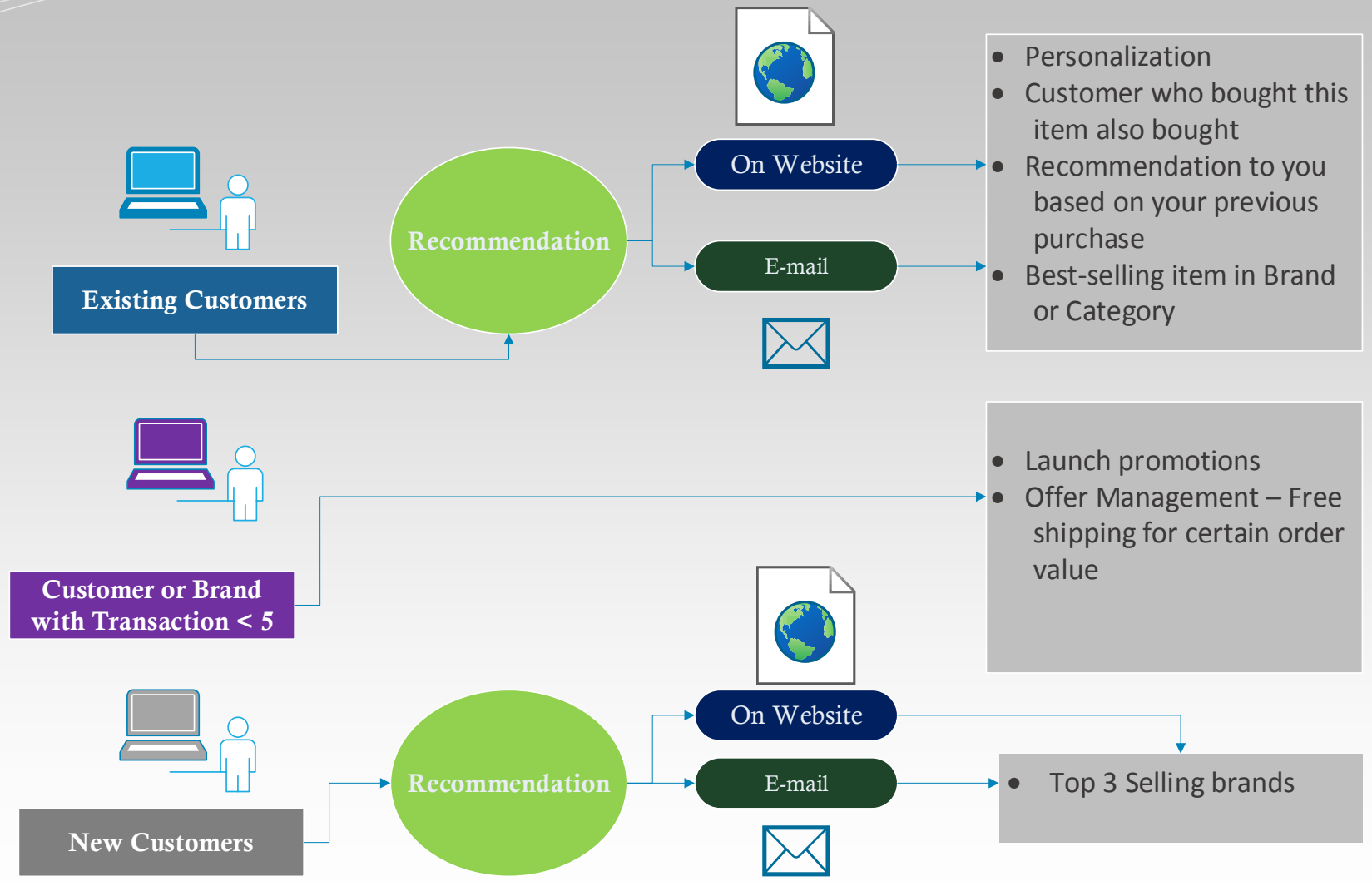
Which clustering technique to use? Why?

- Execute K-Modes Clustering
- K-Modes clustering is suitable for categorical data types
- Optimum clusters = 5 (K = 5)

How does the Recommender System works?

- If a customer id is given, top three product recommendation can be found from the clusters.
- First the clusters are found which the customer id belongs to. The cluster where maximum number of times this customer id appears is the ideal cluster.
- Then top three brand from that cluster for that customer id are found out for recommendation.

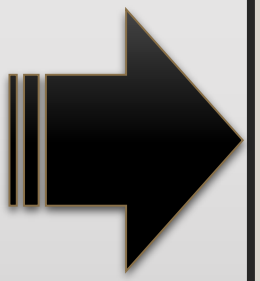
Project Deliverables : Strategies to Increase Average Order Value



INSIGHTS FROM EDA

- Most Frequently Bought Items
- Customer Preference For Specific Brands
- Purchase Patterns By Quantity
- Similarity In Purchase Patterns By Customers
- Similarity In Purchase Pattern By Brands

- **Top 10 Brands**
- **Customers who exclusively purchase products during weekdays and weekends**
- **Top 3 Weekend Customers – by Total purchase amount**
- **Average $ spent per transaction**



BUSINESS VALUE / IMPACT

- **Help demand management, forecast sales**
- **Improve planning process to reduce Inventory levels- by cross-selling, upselling and substitute selling**
- **Adopt Customer Centric approach – delight customers**
- **Apply Machine Learning algorithm and new technology to gain competitive advantage**
- **Ramp up Order-fulfillment process to meet sales demand**