SALES ANALYSIS AND CUSTOMER SEGMENTATION IN RETAIL

A VISUAL ANALYTICS PROJECT, 2020/2021

Alessandra Monaco (1706205)

TABLE OF CONTENTS

1 MOTIVATIONS

Customer Intelligence
How to increase customer satisfaction and loyalty?
How to do tailored advertising?

Resource Allocation
Which products to buy?

Manufacturers

Produce goods from raw material

$ Sell to

Retailers

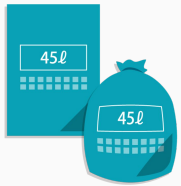
Sell goods to the public in relatively small quantities.

$ Sell to

Customers

Purchase according to their needs and interests

$ Sell to



Wholesalers

Sell goods in large quantities

$ Sell to

GOAL: Maximize profits

1 TASKS AND SUBTASKS

TASKS

- **Analyze the Market** (trending product categories, proficient time periods, seasonalities, …) for improved marketing strategy
- **Understand customer needs** to increase satisfaction
- **Treat customers differently** depending on their interests, loyalty, value (targeted advertising, targeted promotions and discounts, …)

SUBTASK

CUSTOMER SEGMENTATION : split customers into groups with similar characteristics:

- Purchase behavior (ex: *RFM Segmentation*) = how they buy
- Interests and needs (ex: *Unsupervised Segmentation*) = what they buy
- Demographic characteristics
- Custom Business rules
- …

2 RELATED WORKS

[1] Ricky Akbar, Meza Silvana, Mohammad Hafiz Hersyah, Miftahul Jannah. **Implementation of Business Intelligence for Sales Data Management Using Interactive Dashboard Visualization in XYZ Stores**. *International Conference on Information Technology Systems and Innovation (ICITSI)*, 2020

Inspirational dashboard to analyze sales and profit trends, sales by categories, apply drill-down operators, ...

[2] Ron Kohavi, Rajesh Parekh. **Visualizing RFM Segmentation**. *Proceedings of the 2004 SIAM international conference*, 2004

Explanation and application of RFM segmentation model, with segment visualization ideas.

[3] A. S. M. Shahadat Hossain. **Customer Segmentation using Centroid Based and Density Based Clustering Algorithms**. *3rd International Conference on Electrical Information and Communication Technology (EICT)*, 2017

Introduction to unsupervised segmentation comparing 2 algorithms

[4] Rahul Shirole, Laxmiputra Salokhe, Saraswati Jadhav. **Customer Segmentation using RFM Model and K-Means Clustering**. *International Journal of Scientific Research in Science and Technology (IJSRST)*, 2021

Application of unsupervised clustering on RFM data

[5] Julian Heinrich, Bertjan Broeksema. **Big Data Visual Analytics with Parallel Coordinates**. *Big Data Visual Analytics (BDVA)*, 2015

Use of parallel coordinates together with PCA and K-Means to discover structure in the data

3 DATASET

Customers

(5647,4)

Id	**DOB**	**Gender**	**City_Code**

Product Categories

(23,4)

Product_ category_ code	**Product_ category _ name**	Product_ subcategory_ code	**Product_ Subcategory_ name**

Transactions

(23053,10)

Id	**Cust _ Id**	**Date**	**Product_ subcategory _ code**	**Product_ category _ code**	**Qty**	**Rate**	**Tax**	**Total_ Amount**	**Store_ type ***

Merge join

Full Transactions Data

(23053,15)

Dataset available on Kaggle

* ***Multiple Retailing*** use case : products are offered to consumers through different channels (physical store, online store, tele-shop, …)

4 ANALYTICS - RFM Segmentation

- **R**ecency = n° days passed since last purchase
- **F**requency = n° purchases until most recent date in the dataset
- **M**onetary = total amount spent until most recent date in the dataset

Full Transactions Data

Computing Recency, Frequency and Monetary value
for each customer

Quartile-based scoring
Equal-frequency binning (4 bins) for each metric

4x4x4=**64** segments

FILTERS
- Store type
- Customer gender
- Customer city

Analyze 64 segments

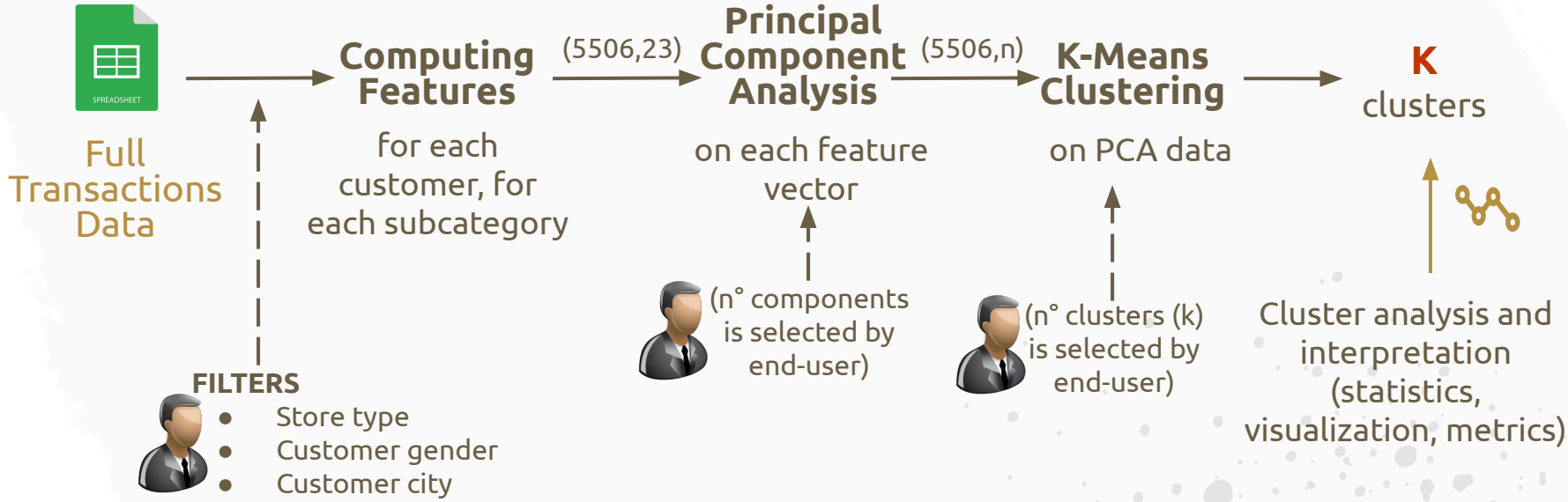
Analyze using a total score

Analyze using visualization

4 ANALYTICS - Unsupervised Segmentation

Learning-based segmentation

Feature$_i$ = total n° products belonging to subcategory i purchased by a customer



5 VISUALIZATIONS - Transaction Time Series

1 / 2

LINE CHART

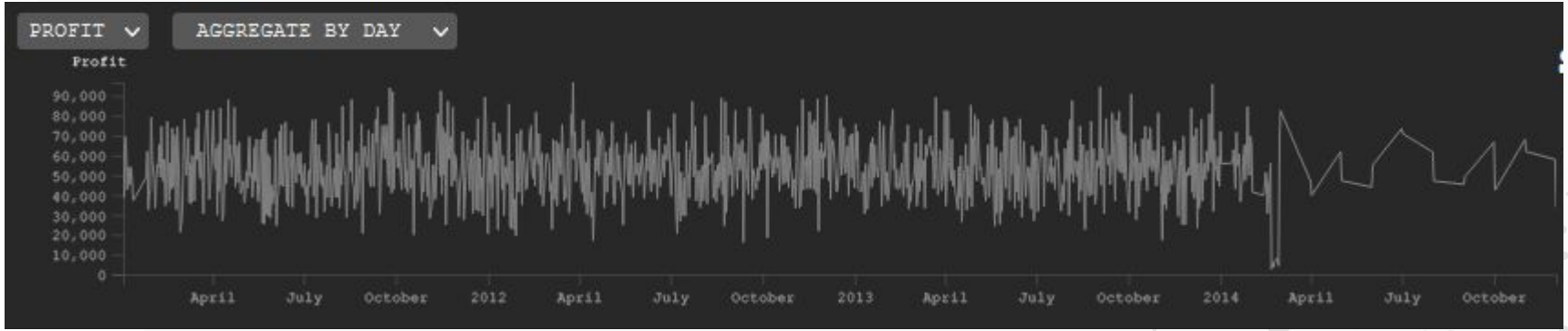
DATA : Transaction Time Series

VISUAL ENCODING

- Position : encodes x (date) and y (profit/sales) value

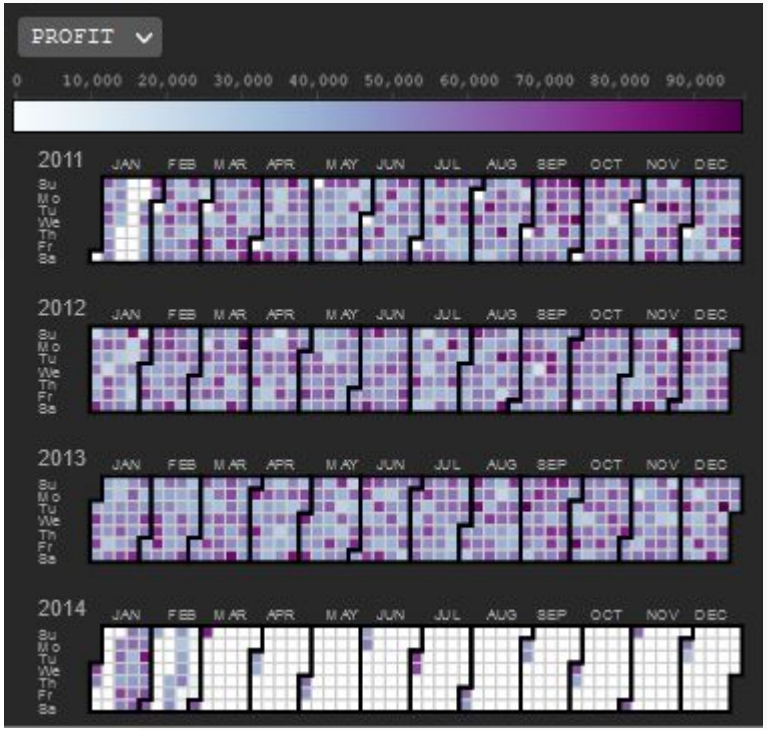
INSIGHTS

- Profit/sales trend (increasing/ decreasing/ stable / fluctuating, ..)
- Most proficient days (peaks of max)
- Most/Less proficient time periods
- Comparing monthly/yearly profits or sales



5 VISUALIZATIONS - Transaction Time Series

2 / 2



CALENDAR HEATMAP

DATA : Transaction Time Series

VISUAL ENCODING

- Color shades: encode total profit/sales for a cell (day)

INSIGHTS

- Seasonal patterns(week-day, month, season, year)
- Quickly identify most proficient days
- Most/Less proficient time periods

5 VISUALIZATIONS - Sales By Category

TREEMAP

DATA : Transaction Time Series

VISUAL ENCODING

- Color : encodes a category
- Rect size : encodes the total sales of (sub)category

INSIGHTS

- Most sold category/subcategory (in general or for a given segment/cluster)
- Hierarchical structure of product categories

5 VISUALIZATIONS - Clusters

1 / 2

2D PCA SCATTER PLOT

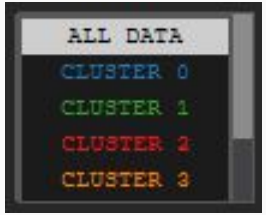
DATA : Clustered PCA data (2 components)

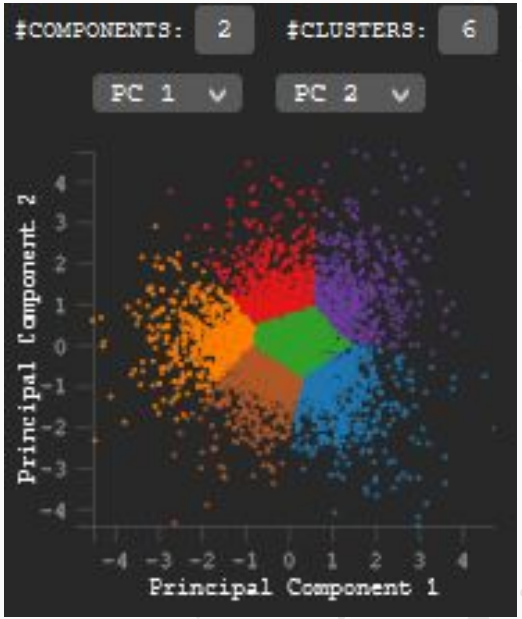
VISUAL ENCODING

- Position : encodes values of the chosen principal components
- Color : encodes correspondent cluster

INSIGHTS

- Clusters separation in the chosen components
- Presence of outliers in some clusters





5 VISUALIZATIONS - Clusters

2 / 2

PARALLEL COORDINATES

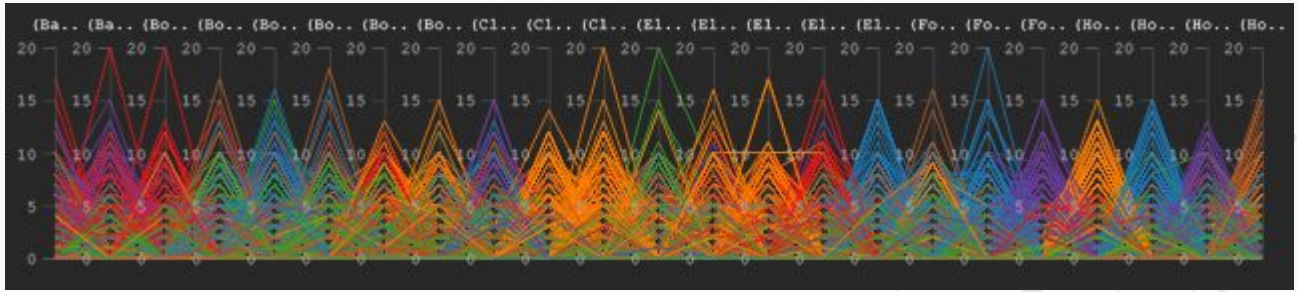
DATA : Clustered data (original features)

VISUAL ENCODING

- Position on dimensional axis : encodes the value of the dimension
- Color : encodes correspondent cluster

INSIGHTS

- Cluster interpretation : which features mostly characterize the cluster and how? Are the clustering results meaningful?
- Possible correlations between purchased sub categories (in the full dataset or in a clusters)



5 VISUALIZATIONS - Rfm Segments

HEATMAP

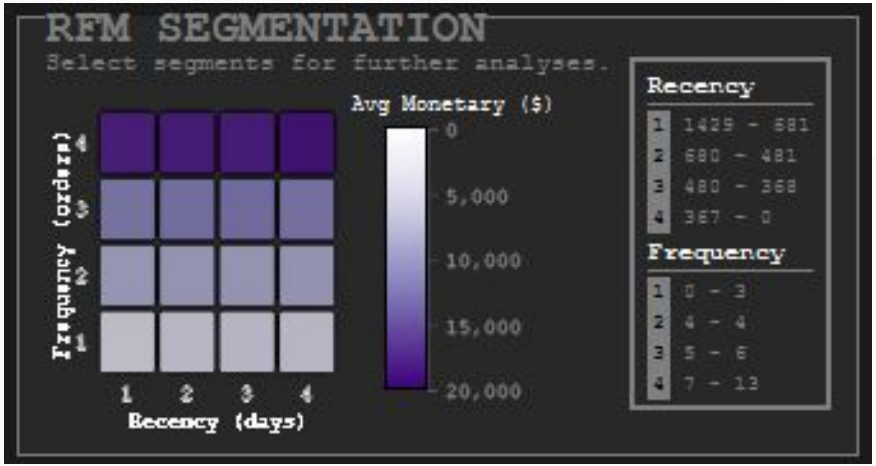
DATA : RFM data with Avg Monetary

VISUAL ENCODING

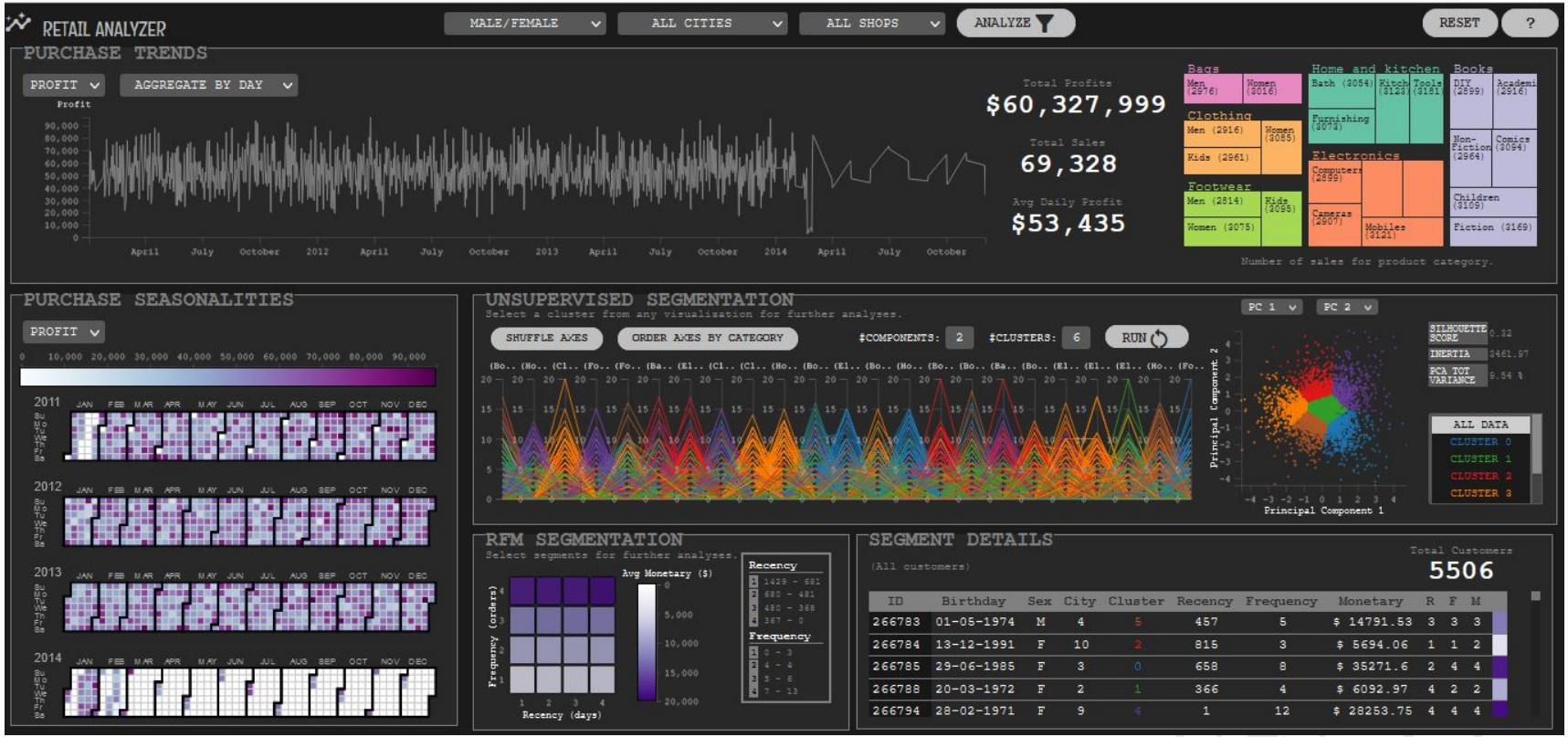
- Color shades : encode the average monetary value of the cell

INSIGHTS

- **Quickly identify important customer segments**
- How much does the monetary value change between different segments?
- What is the behavior of Big Spenders? (are they frequent/recent buyers also?)
- How much have frequent buyers spent on avg?



6 LIVE DEMO


RETAIL ANALYZER
MALE/FEMALE
ALL CITIES
ALL SHOPS
ANALYZE
RESET
?
PURCHASE TRENDS
PROFIT
AGGREGATE BY DAY
Total Profits
$60,327,999
Total Sales
69,328
Avg Daily Profit
$53,435
Number of sales for product category.
PURCHASE SEASONALITIES
UNSUPERVISED SEGMENTATION
Select a cluster from any visualization for further analyses.
SHUFFLE AXES
ORDER AXES BY CATEGORY
#COMPONENTS: 2
#CLUSTERS: 6
RUN
SILHOUETTE SCORE 0.32
INERTIA 3461.97
PCA TOT VARIANCE 9.54 %
ALL DATA
CLUSTER 0
CLUSTER 1
CLUSTER 2
CLUSTER 3
RFM SEGMENTATION
Select segments for further analyses.
Recency (days)
Frequency (orders)
Avg Monetary ($)
SEGMENT DETAILS
(All customers)
Total Customers
5506
ID Birthday Sex City Cluster Recency Frequency Monetary R F M
266783 01-05-1974 M 4 5 457 5 $ 14791.53 3 3 3
266784 13-12-1991 F 10 2 815 3 $ 5694.06 1 1 2
266785 29-06-1985 F 3 0 658 8 $ 35271.6 2 4 4
266788 20-03-1972 F 2 1 366 4 $ 6092.97 4 2 2
266794 28-02-1971 F 9 4 1 12 $ 28253.75 4 4 4

CONCLUSIONS

+ Better Resource Allocation : retailers buy from wholesalers/manufacturers only what is likely to be purchased by customers;

+ More effective Targeted Advertising, focused on groups of customers with similar characteristics;

+ More proficient Marketing, treating customers according to their value, loyalty, proficiency, …

+ Improved Customer Satisfaction : the market is aligned with customers' needs and interests.

THANK YOU FOR YOUR ATTENTION

Full code available on
https://github.com/AlessandraMonaco/Visual-Analytics