USER LIFE-TIME VALUE AND RECOMMENDATION ENGINE ON E-COMMERCE SALES DATA

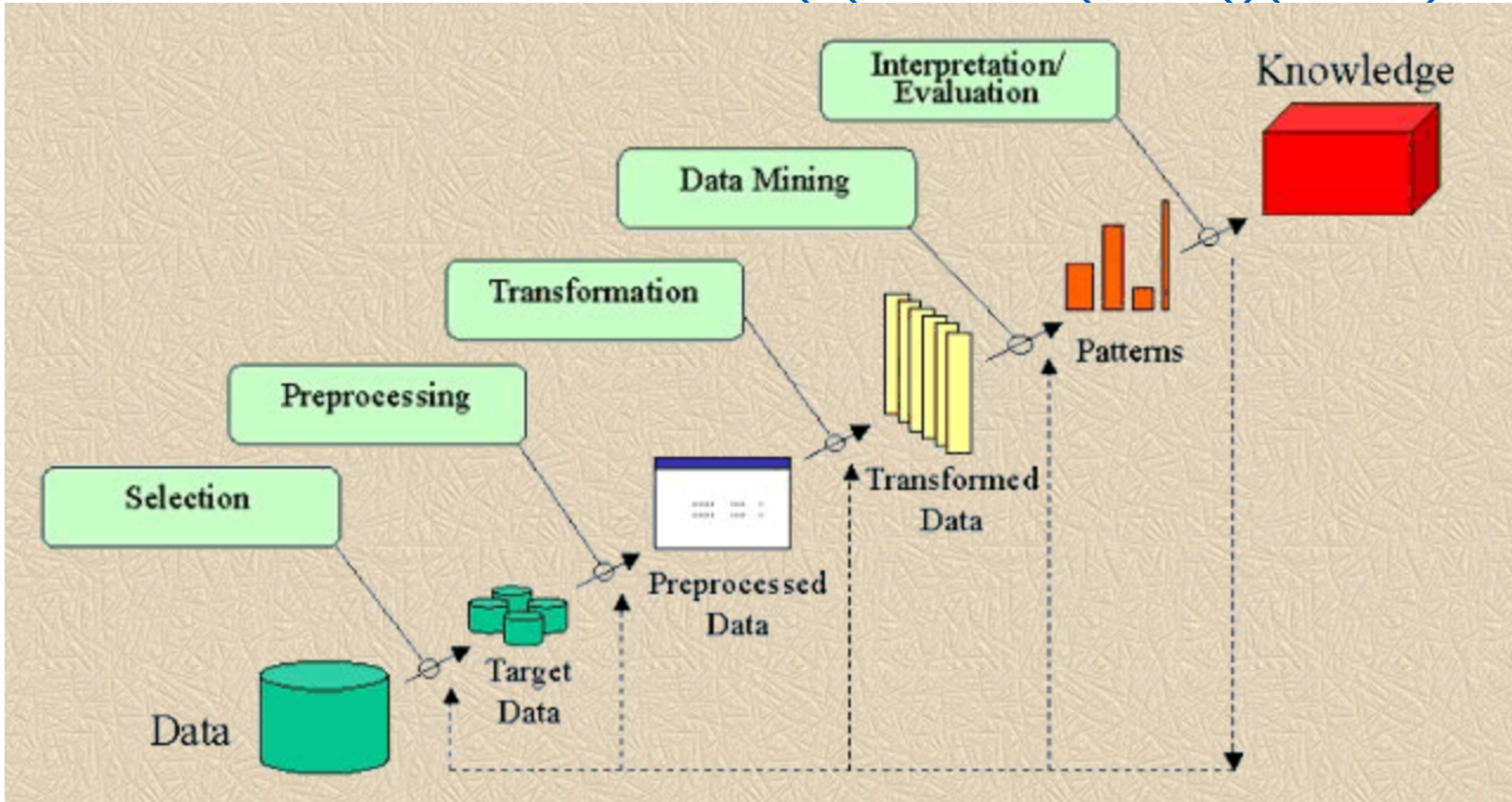
SHUBHAM SHARMA

ARCHIT SINGH

University at Buffalo
School of Engineering and Applied Sciences

CONTENTS

- About Superstore
- Problem Statement
- Data Description
- Data Preprocessing
- Data Analysis
- User Life-Time Value (LTV)
- Product Recommendation Engine



About Superstore

- Superstore is an online platform with various product listings and the sales data available to us is from the year 2009 – 2012.
- This data set is available publically and can be seen as a generic transaction level dataset
- This dataset helps us make a model that can be translated to any other transactional dataset
- The dataset has records corresponding to one transaction made by a user

Problem Statement

- To gauge how valuable a user is for the business based on his purchase history
- To create a recommendation engine for the store to recommend products to people who buy a particular product in future.

Data Description

- The Data available is sales data for superstore from 2009 – 2012
- It has the following attributes: Row ID, Order ID, Order Date, Order Priority, Order Quantity, Sales, Discount, Ship Mode, Profit, Unit Price, Shipping Cost, Customer Name, Province, Region, Customer Segment, Product Category, Product Sub-Category, Product Name, Product Container, Product Base Margin, Ship Date

Data Preprocessing

- Rows with NULL values form close to 0.01% of the data therefore we can afford removing these records
- Label encoding customer name and product name in order to make computation faster for Recommendation system.

Data Analysis

- Trend of profit and sales over the time period (2009-2012)
- Shop's response to orders of different priority
- Profit share for different provinces
- Correlation between different variables
- Customer Behavior/ Performance

User Life-Time Value (LTV)

- Value attributed to the entire future relationship with a customer
- Measure of how important is a customer to the business
- Product recommendations can be improved by tailoring them based on the LTV of customers

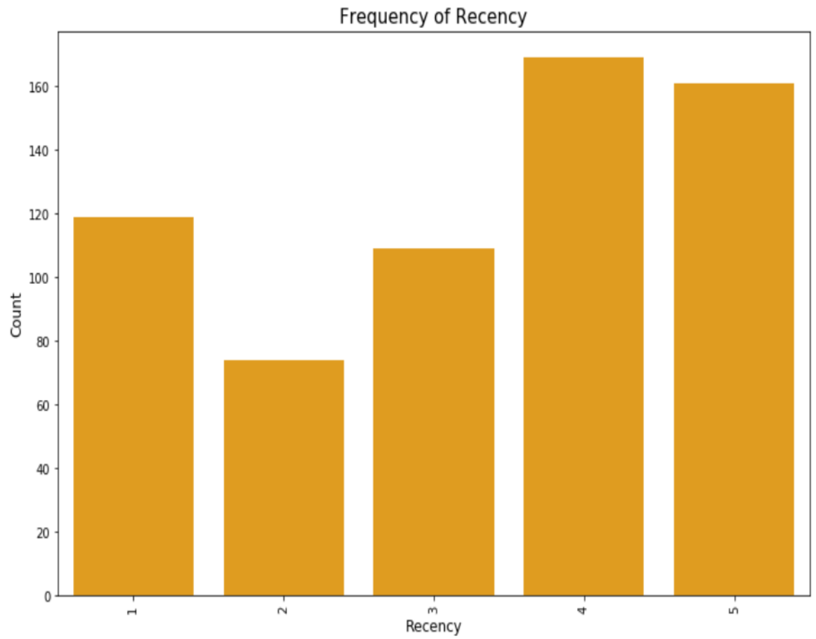
Defining LTV - RFM metrics

- Understanding the LTV of a user using RFM metrics
- RFM is a customer segmentation technique that uses past **purchase behavior** to divide customers into groups
- The components of RFM are:

  RECENCY (R): Time since last purchase
  FREQUENCY (F): Total number of purchases
  MONETARY VALUE (M): Total monetary Profit
- RFM results in an increased customer retention, response and profit

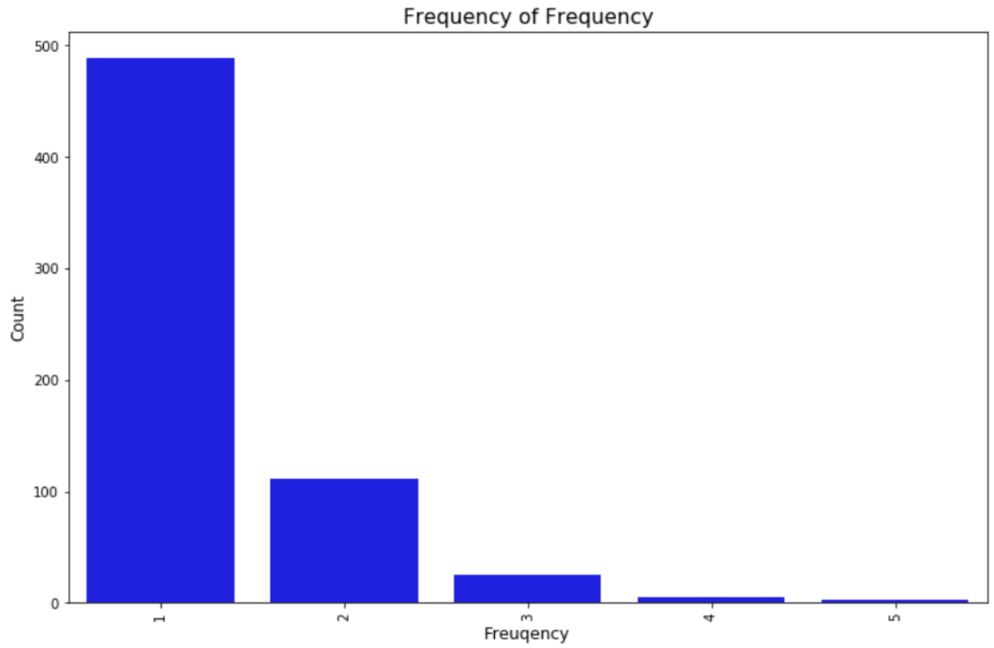
Recency

- Recency has been calculated w.r.t the last 2 months of 2012
- Customers are scored from 1-5 based on their most recent purchase
- Score is 5 if they bought in last 2 months, 4 if bought in last 4 months and so on



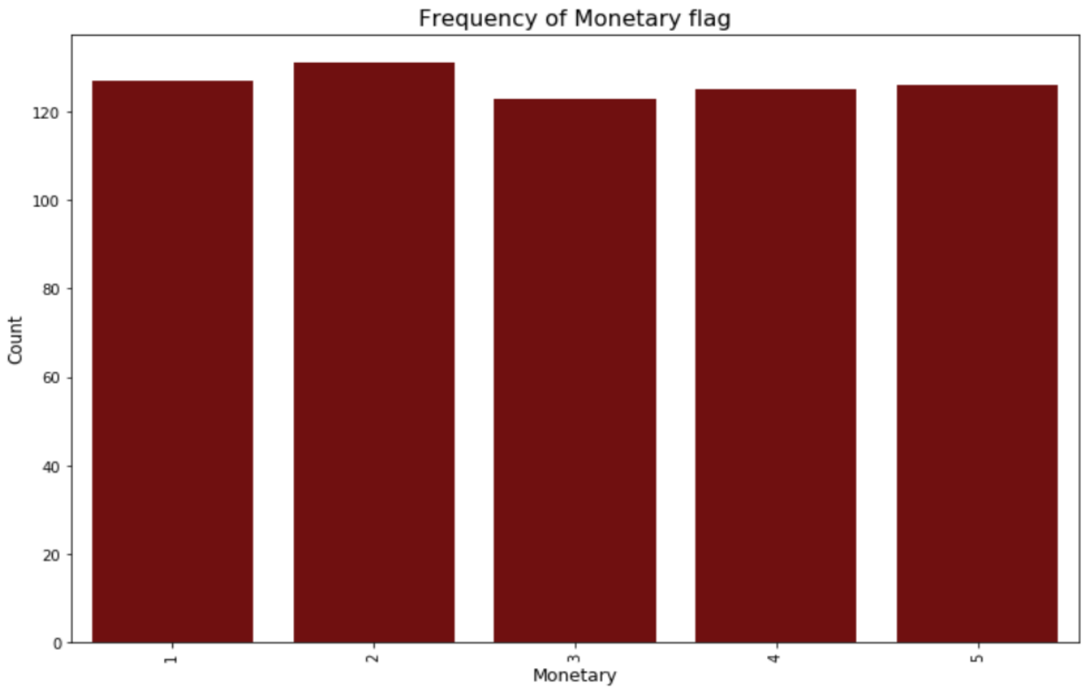
Frequency

- Frequency was divided into pentiles
- Pentiles were scored based on their value. For instance: pentile with highest value was given a score of 5 and the one with lowest value was scored 1



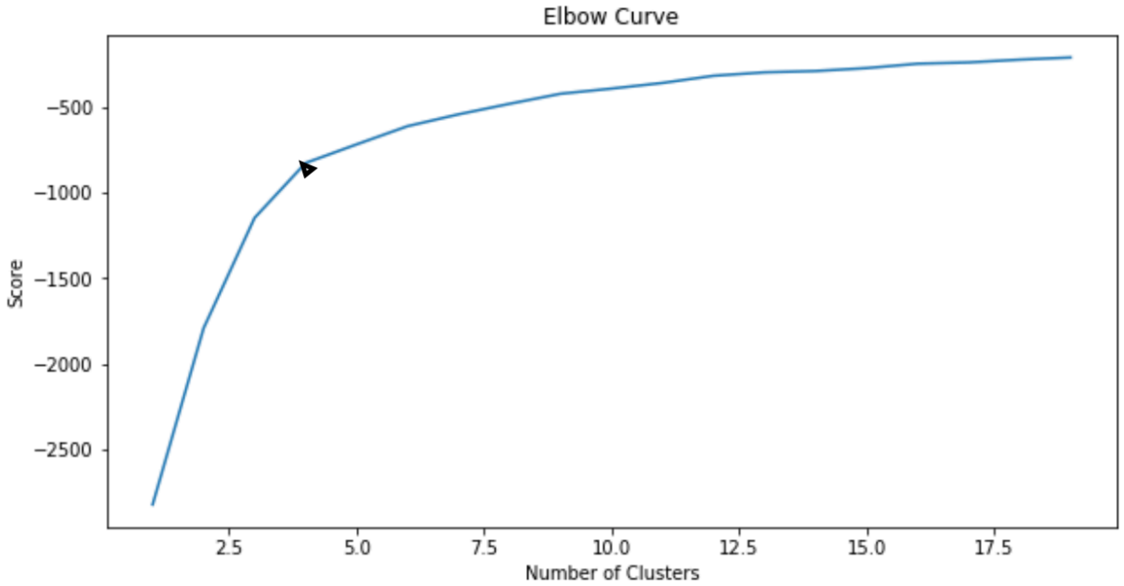
Monetary

- Similar to Frequency Monetary values were also divided into pentiles
- Pentile with highest value was given a score of 5 and the one with lowest value was scored 1



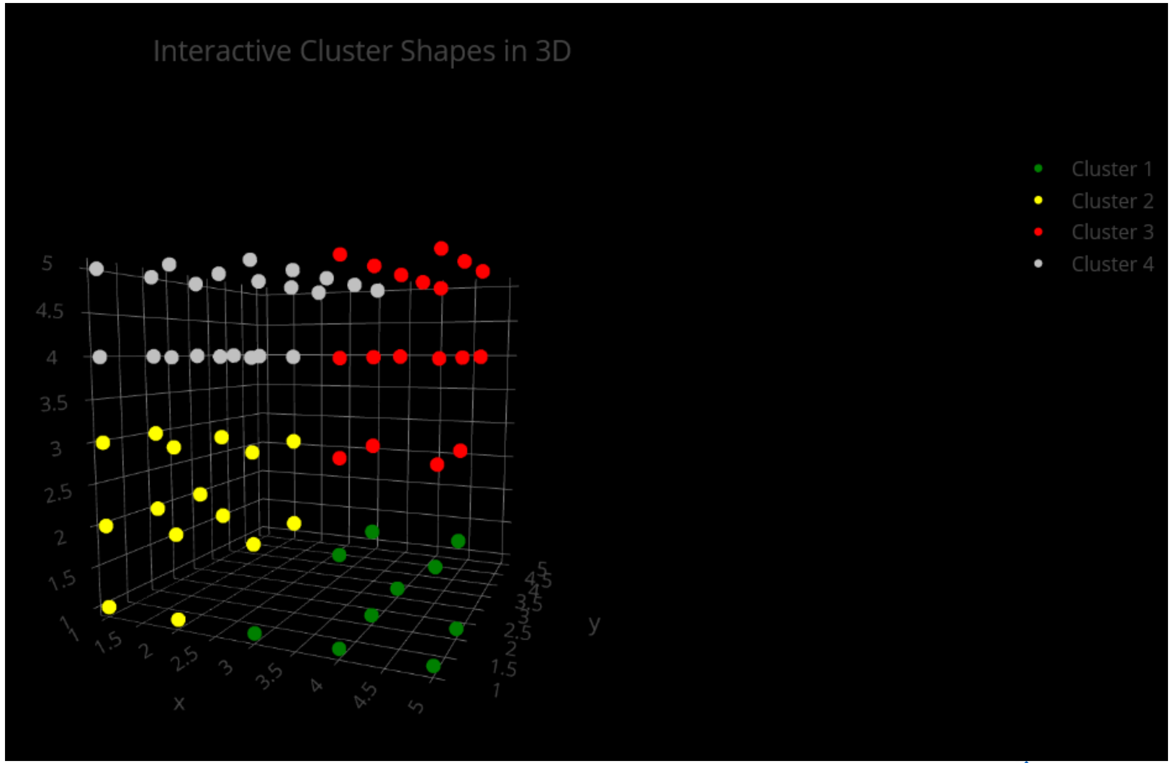
Clustering – Deciding the number of clusters

- Generally RFM values are used to define a fixed number of segments
- Since we are keeping the learning unsupervised the number of clusters are decided on the go
- We used elbow method to determine the number of clusters
- The elbow was observed at k = 4 and hence those many clusters were decided to be taken



K Means clustering to segment the Customers

- The data(R, F and M) were segmented into 4 clusters
- The Cluster 2 stands out to be of the highest value considering R, F and M metrics thus any user belonging to this cluster can be viewed as a high valued customer
- The Cluster 1 users can be seen as the users who have churned since their monetary and frequency values are fairly high but recency value is low.
- The Cluster 4 users doesn't seem to be significant with low M value, medium F and high R value



Product Recommendation Engine

- Recommender system using item-item similarity to help the shop recommend products to new users.
- User-User similarity measure is difficult to get here. Therefore final approach is item-item similarity. The similarity measure is number of purchases made for a product

	Customer Name	Product Name	Order Quantity
0	Aaron Bergman	Acme? Preferred Stainless Steel Scissors	23
1	Aaron Bergman	Avery 49	5
2	Aaron Bergman	Canon S750 Color Inkjet Printer	35
3	Aaron Bergman	SANFORD Liquid Accent? Tank-Style Highlighters	13
4	Aaron Bergman	V70	26

Product Recommendation Engine

- User based similarity measure is difficult to get here since small changes in user behaviors can make big changes in predictions. Therefore final approach is item based similarity since it has larger support and users are added more frequently than items thus have implications in model upkeep.
- The similarity measure is number of purchases made for a product

	Customer Name	Product Name	Order Quantity
0	Aaron Bergman	Acme? Preferred Stainless Steel Scissors	23
1	Aaron Bergman	Avery 49	5
2	Aaron Bergman	Canon S750 Color Inkjet Printer	35
3	Aaron Bergman	SANFORD Liquid Accent? Tank-Style Highlighters	13
4	Aaron Bergman	V70	26

Utility Matrix

- The matrix components show the number of purchases and NaN depicts that there was no purchase corresponding to that user and product.

customerId	0	1	2	3	4	5	6	7	8	9	...	785	786	787	788	789	790	791	792	793	794
productId																					
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	23.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	41.0	NaN	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	86.0	NaN	NaN	NaN	NaN	NaN	NaN	7.0	NaN	NaN

Normalized Utility Matrix

- Normalization is carried by projecting the range of purchase quantity for each item between 0 and 1.
- Better cosine similarity measure is obtained between two vectors

customerId	0	1	2	3	4	5	6	7	8	9	...	785	786	787	788	789	790	791	792	793	794
productId																					
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.341463	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000	0.0	0.888889	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	1.0	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0

5 rows × 795 columns

Prediction

- Top 10 recommendations for user who buy “Acme? Preferred Stainless Steel Scissors”

	0
0	Barricks 18" x 48" Non-Folding Utility Table w...
1	Vinyl Sectional Post Binders
2	Avery 520
3	Xerox 213
4	Xerox 1880
5	Staples Wirebound Steno Books, 6" x 9", 12/Pack
6	Hon 4070 Series Pagoda? Armless Upholstered St...
7	SANFORD Liquid Accent? Tank-Style Highlighters
8	GBC DocuBind P100 Manual Binding Machine
9	Accessory39

Conclusion

- Users were segmented into different clusters based on the metric chosen
- Priority can be assigned to different clusters when it comes to marketing and campaigning
- The recommendations can be tailored based on the RFM information to target the users more efficiently
- Item based similarity was found to be better than user based similarity for collaborative filtering in this case.

Links

- Data: https://raw.githubusercontent.com/curran/data/gh-pages/superstoreSales/superstoreSales.csv
- Tableau Dashboard: https://tabsoft.co/2UbgwBw
- Github: https://github.com/shub91/Superstore_Sales_Analysis


Diolch
Kiitos
Sheun
umesc
Kasih
Todah
Mamnoon
Shnorhakalutiun
Dekuju/Dekujeme
Te ekkür
Hvala
Shokriya
Dziękuję
Shokrun
Spaas
Mul
Ači
Gamsahapnida
Ngiyabonga
Xie
Dank
Takk
Cam
Gra
or
Dhanyavaad
al
Kop
Salamat
Merci
Dankie
Go
Grazie
Faleminderit
Dakujem
Waad
Dhanyavaadaalu
Daw
Dhanyavad
Khopjai
Kruthagnathalu
ThankYou
kun
Arigatou
krap
Tack
Shukriya
Or
Dhonnobaad
ederim
Asante
Gracias
Nandree
Blagodariya
Gomapsupnida
Euxaristo
Kun
Hain
Dhan
Grazzi
raibh
Fyrir
Terima
Enkosi
Danke
dank
daa