Customer Segmentation
Lovelytics team 2

Chris Lee
Fulin Wang
Naijia Wu
Huan Sun
Kata Mezo
Mengting Wang
Yanlin Zhang

Agenda

1. Objectives
2. Strategies & Methods
3. Visualization Results
   - Customer Demography
   - Customer Segmentation
4. Modelling Results
5. Conclusions
6. Limitations & Deficiencies

Objectives


The objective of this project is to “provide our client with insights into their target customers.”

Strategy & Method

Strategy and Preprocessing

1.

Combine the two datasets

- Merging on Personal Sequence Number (PSN)
- Randomized chose of the data due to time pressure
- Total Sample Size (n=144,629)

2.

Data Cleaning & Preprocessing

- Drop "indicator" features
- Recode entries of "spending" variables by averages of their pre-defined intervals, and keep the highest categories only
- Group similar types of features up and impute missing values iteratively
- Sum "spending" variables up and drop their individual time-series components

3.

Modelling

- Unsupervised machine learning
  - K-Means clustering
  - Principal Component Analysis

K-Means clustering

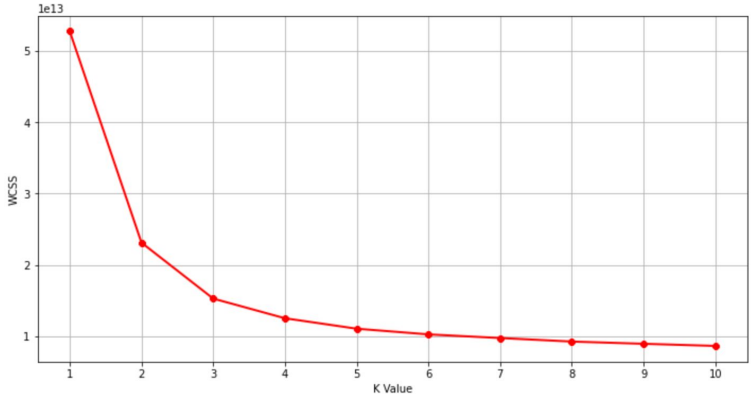
The logic of K-Means clustering

- Minimize the within-cluster variation
  - A good clustering is one for which the within-cluster variation(WCV) is as small as possible.

$$\underset{C_1,\ldots,C_K}{\text{minimize}} \left\{ \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2 \right\}$$

- Select the number of clusters
  - X label: number of clusters
  - Y label: reduction within cluster variation

- See the elbow chart and finalize with 5 clusters

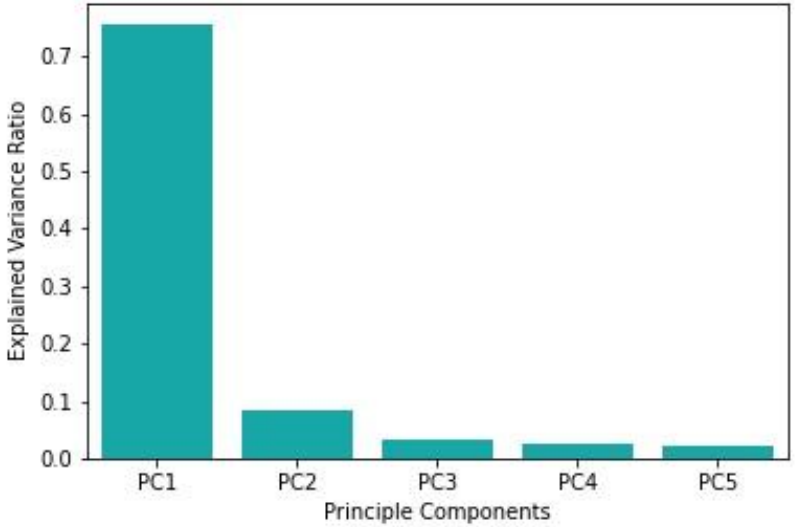


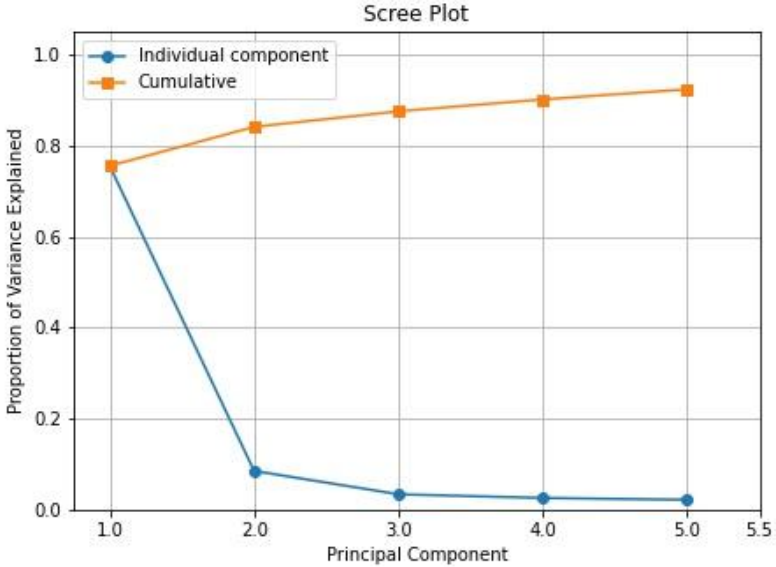
K-Means clustering

The logic of K-Means clustering

- After specifying the number of clusters:
- Select randomly 1-5 objects as the initial cluster centers
- Iterative until the cluster assignments stop changing:
    - For each of the 5 clusters, compute the cluster centroid
    - Assign each observation to the cluster whose centroid is closest (where closest is defined using Euclidean distance)

K-Means clustering and PCA





- PC1 explains nearly 80% of the total variance, while subsequent PC's each contribute less than 10%
- We then should primarily prioritize the subcomponent features of PC1

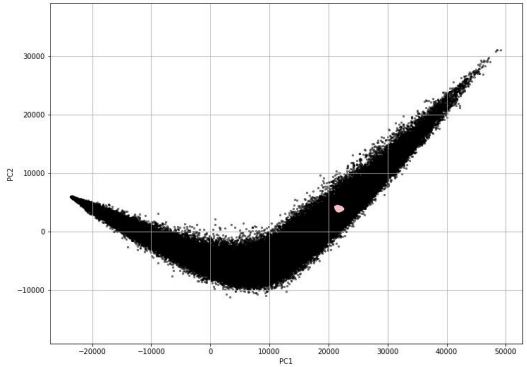
K-Means clustering and PCA

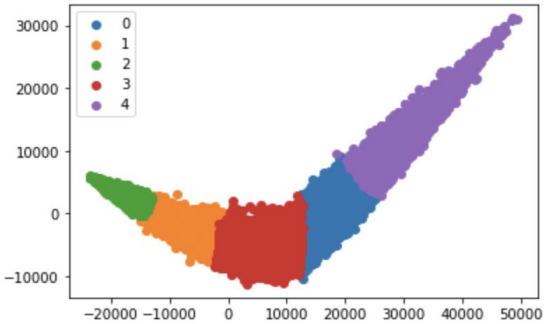
	PC1 subcomponent	Variable Names
13	0.636267	sum_overall
21	0.500260	sum_retail
19	0.373093	sum_other_retail
18	0.290034	sum_food_convenience_drug_store
23	0.241054	sum_travel

	PC2 subcomponent	Variable Names
13	0.701628	sum_overall
19	0.480100	sum_other_retail
23	0.384347	sum_travel
18	0.241398	sum_food_convenience_drug_store
20	0.133158	sum_restaurant

- The overall spending takes up 0.636267 of the PC1 loading, followed by spendings on retail and on other retail.
- The subcomponents of PC2 exhibit a fairly similar pattern, though the component itself accounts for far less total variability.

K-Means clustering and PCA





Original Data & PCA-Transformed Data Comparison

- The light points are the original data, while the dark points are the projected version.
- The information alongside the least important principal axes are removed, leaving only the components with highest variances.
- Despite reducing the dimension of the data by more than 50%, the overall relationship between the data points are mostly preserved.
- The five clusters are well separated along both PC1 and PC2 axes, implying their different consumption patterns

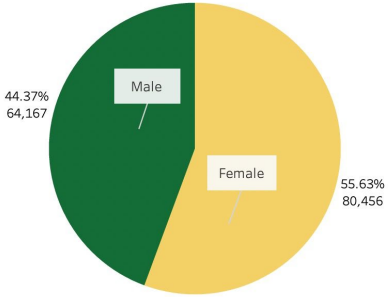
Visualization results

Customer Demography

Customer Segmentation

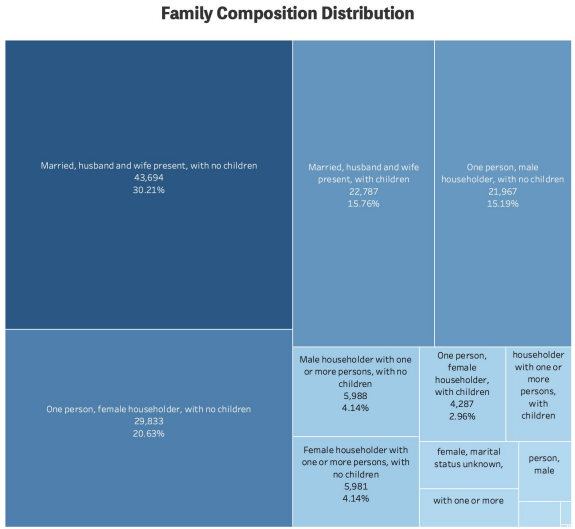
Visualization results - Customer Demography

Female Housing Power



- ***Female Housing Power Rocks!***
  - Female to Male Ratio: 5/4
  - Status of house owning, female to male : 4/3
  - Percentage gain: 6.67%

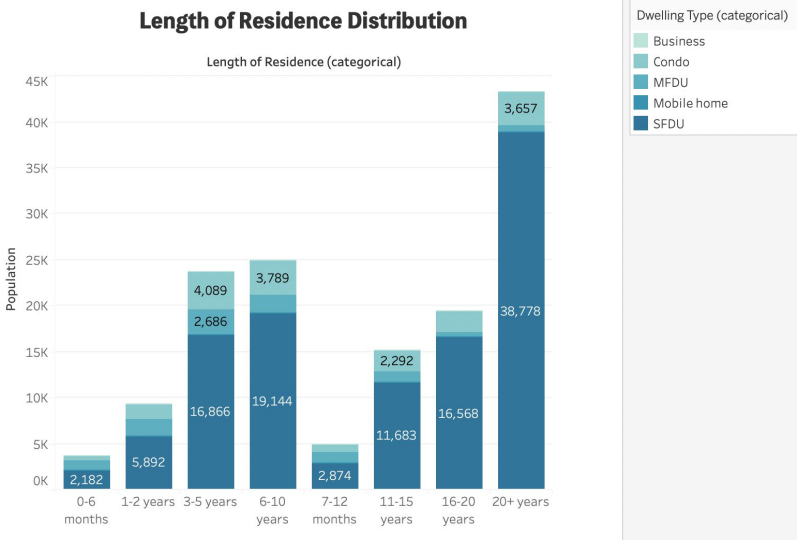
Family Composition Distribution

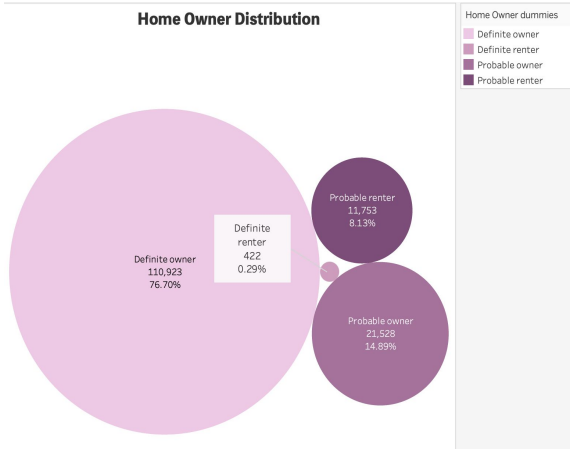


Visualization results - Customer Demography

Distribution of Dwelling Types

- SFDU prevails for the length of residence.
- The dominant subject of the dataset are owner instead of renter.





Visualization results - Customer Demography

The effect of marital status on the dwelling types

- Despite that different marital status have certain effect on the dwelling types, the most prevalent type of dwelling for both the married and single people is SFDU (Single-Family-Dwelling-Unit)

	Dwelling Type dummies					
Marital Status dummies	Business	Condo	MFDU	Mobile home	SFDU	Unknown
Married	147	8,939	3,370	295	80,051	1
Single	115	10,227	7,379	123	33,979	

	Dwelling Type dummies					
Marital Status dummies	Business	Condo	MFDU	Mobile home	SFDU	Unknown
Married	0.16%	9.63%	3.63%	0.32%	86.26%	0.00%
Single	0.22%	19.73%	14.24%	0.24%	65.57%	

Visualization results

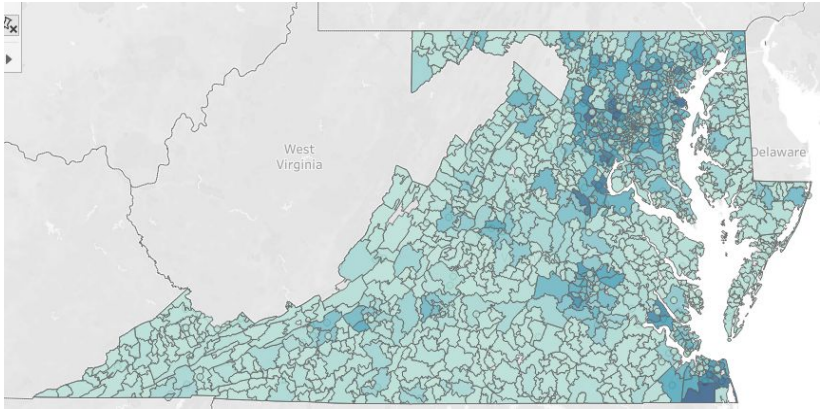
Customer Demography

Customer Spending

Visualization results - Customer Spending

Overall Spending by Region

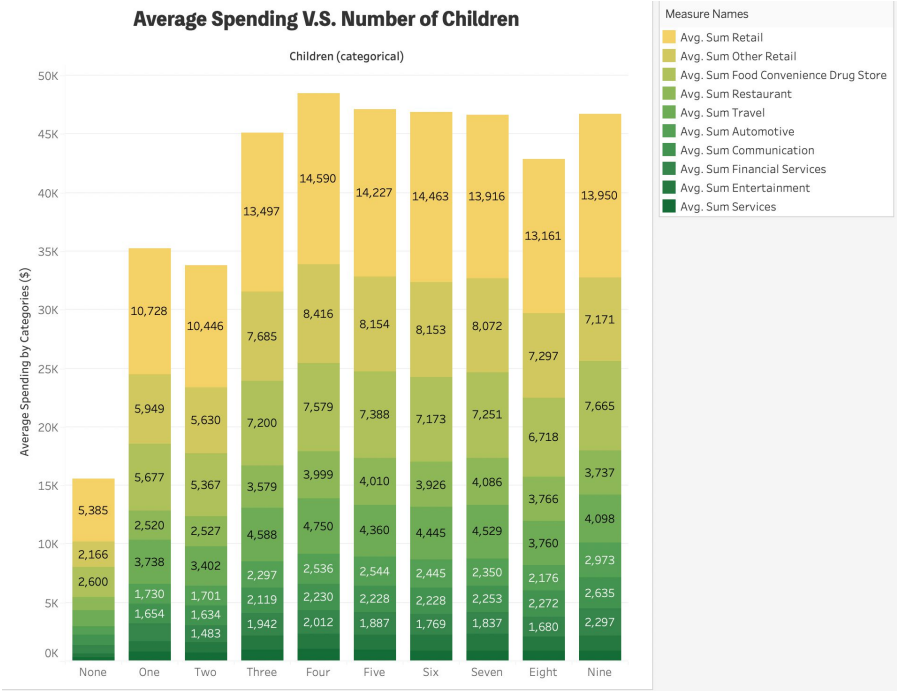
- Top 3 Overall Spending Areas: Chesapeake (VA), Potomac (MD), Pasadena (MD)



Category	No. 1	No.2	No. 3
Automotive	***Potomac, MD***	Ellicott City, MD	Chesapeake, VA
Communication	**Pasadena, MD**	Potomac, MD	Woodbridge, MD
Education	Yorktown, VA	Poquoson, VA	Virginia Beach, VA
Financial Service	***Potomac, MD***	Ashburn, VA	Pasadena, MD
Food Convenience	Ashburn, VA	Potomac, MD	Gaithersburg, MD
Retail	***Potomac, MD***	Ashburn, VA	Pasadena, MD
Other Retail	***Potomac, MD***	Ashburn, VA	Chesapeake, VA
Restaurant	**Chesapeake, VA**	Pasadena, MD	Ashburn, VA
Service	Stafford, VA	Virginia Beach, VA	Chesapeake, VA
Travel	***Potomac, MD***	Ashburn, VA	Chesapeake, VA

Visualization results - Customer Spending

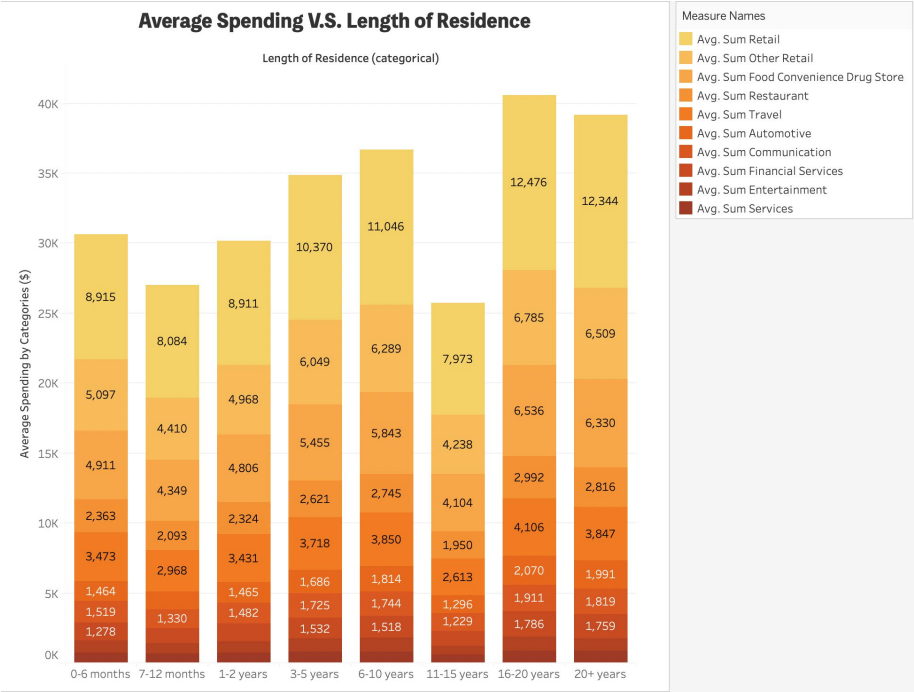
Average spending vs Number of Children



- a positive association between number of children and total spending per person
- the highest "slopes" : someone starts to have children and when the number of children increases from 2 to 3
- the average spending trend is relatively smooth and does not necessarily indicates an increasing pattern
- the rank of raw amount of spending regarding different categories is consistent, regardless of how many children one has

Visualization results - Customer Spending

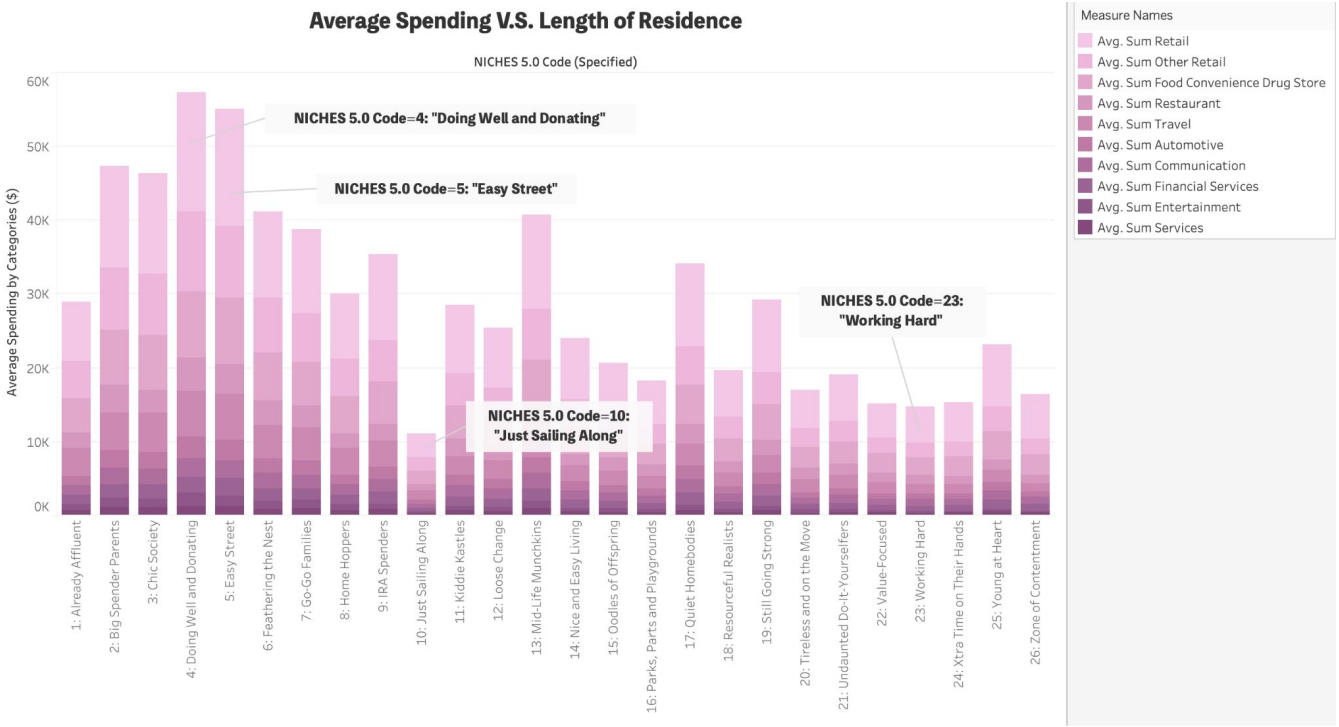
Average spending vs Length of Residence



- a positive correlation between residence length and average spending roughly
- the amount of spending actually drops as it transitions from 6 to 7 month
- customers with 11-15 years have the lowest spending per person
- the rank of raw amount of spending regarding different categories is consistent, regardless of how long one resides

Visualization results - Customer Spending

Average spending vs Niche 5.0 code



Visualization results - Customer Spending

NICHES 5.0 Code explained: "Doing Well and Donating"

- "These families are led by adults in their prime earning years, typically homeowners, who spend 3x more than the average population and donate to a wide variety of charitable causes. They are career-oriented, enjoy traveling, fitness and investing and own newer full-size and luxury vehicles."

Average age	46 years old
Average length of residence	13 years
Presence of children	Likely to have children
Marital status	Mostly married
Homeowners	Owners
Occupations	Finance, MIS/IT/Tech, Management, Marketing/Advertising, Service sector
Education	College degree

Spend behavior

3x average levels of discretionary spend, across all channels

Categories	Kids' apparel and merchandise, décor, fitness, furniture, gift cards, adult apparel and shoes, home improvement, hunting/fishing, active outdoor, sporting goods, business publications
Merchants	Dick's Sporting Goods, iTunes.com, Kohls.com, Target.com

Visualization results - Customer Spending

NICHES 5.0 Code explained: "Easy Street"

- "The households in this niche are typically older and educated, with grown children possibly still living at home. They are financially savvy and active investors, have the highest net worth of any niche and spend 2x the average across many categories."

Average age	63 years old
Average length of residence	15 years
Presence of children	Both with and without children
Marital status	Mostly married
Homeowners	Owners
Occupations	Owner, Management, Business/Financial Operations
Education	College degree

Spend behavior

2x average level of discretionary spend, across all channels

Categories	Male apparel, décor, dry cleaning, flowers, gift cards, men's publications, female apparel, modern décor, business publications
Merchants	Costco, DSW, Giant Food Stores, Macys.com, Nordstrom, Trader Joe's

Visualization results - Customer Spending

NICHES 5.0 Code explained: "Just Sailing Along"

- "These 30-somethings are either working on their degree or climbing the corporate ladder. As they work to establish themselves and build for the future, they are savvy spenders, opting for used vehicles and financial providers with rewards programs. They are often renters and enjoy fitness, travel and the arts."

Average age	35 years old
Average length of residence	5 years
Presence of children	Less likely to have children
Marital status	Mostly single
Homeowners	Renters
Occupations	College students, Banking, Management, Healthcare, Business/Financial Operations
Education	Some college or college degree

Spend behavior

Below-average levels of discretionary spend

Categories	Male apparel, furniture, modern décor, baby accessories, fitness, video games/systems
Merchants	BestBuy.com, Etsy.com, H&M, The Gap

Visualization results - Customer Spending

NICHES 5.0 Code explained: "Working Hard"

- "These hard-working households usually have children and rent their homes. They strive to achieve a high social status and enjoy changing brands for the sake of variety and novelty. They're also very receptive to coupons, offers and discounts."

Average age	40 years old
Average length of residence	8 years
Presence of children	Likely to have children
Marital status	Married and single
Homeowners	Renters
Occupations	Construction, Natural Resources, Sales, Temporarily unemployed
Education	Some college or high school

Spend behavior

Below-average levels of discretionary spend

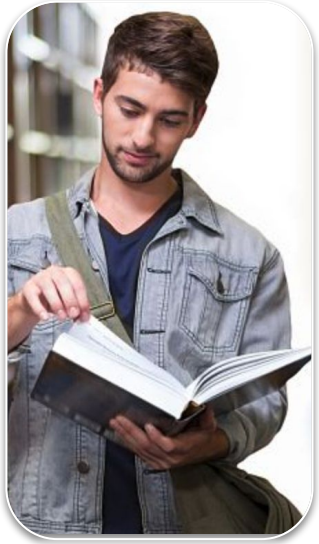
Categories	Laundromats, young-adult merchandise
Merchants	Burlington Coat Factory, Family Dollar, Payless, Save-A-Lot

Modelling results

K-Means Clustering Result

We get 5 distinctive customer segments based on KMeans

Rich and young white collar



Financially conservative single

Rich family-orientated married

Young enjoying life with spending

Married, good at budgeting

K-Means Clustering Result

Cluster 1: Rich and young white collar

- **Single or newly married without children, with relatively high income. Enjoy shopping, convenient food, have saving habit.**
  - Average number of children: 1.79
  - Family composition: **Mostly are married with no children (26.54%) and single female (24.6%)**
  - Discretionary Spending Income: **$55,000-$64,999**
  - Marital Status: 75% married
  - Gender: 43% Male
  - Average of overall spending: $29,949

K-Means Clustering Result

Cluster 2: Financially conservative single

- **Low income single, mostly female. With little need to spend on children's education, enjoy convenient food. With low spending incentive, prefer saving.**
  - Average number of children: 1.59
  - Family composition: Mostly are **single female (26.35%) and single male (19.9%)**
  - Discretionary Spending Income: $25,000-$34,999
  - Marital Status: 57% married
  - Gender: 38% male
  - Average of overall spending: **$3,532**

K-Means Clustering Result

Cluster 3: Rich and successful family-orientated married people

- **High-income and married people. Focus on children's education spending, have enough money to spend much on luxury goods like entertainment and travel. Have a balance in saving and spending.**
  - Average number of children: 1.88
  - Family composition: Mostly are **married with children (35.18%) and married with no children (46.9%)**
  - Discretionary Spending Income: **$65,000 - $74,999**
  - Marital Status: **83% is married**
  - Gender: **47% male**
  - Average of overall spending: $30,000

K-Means Clustering Result

Cluster 4: Young enjoying life with spending

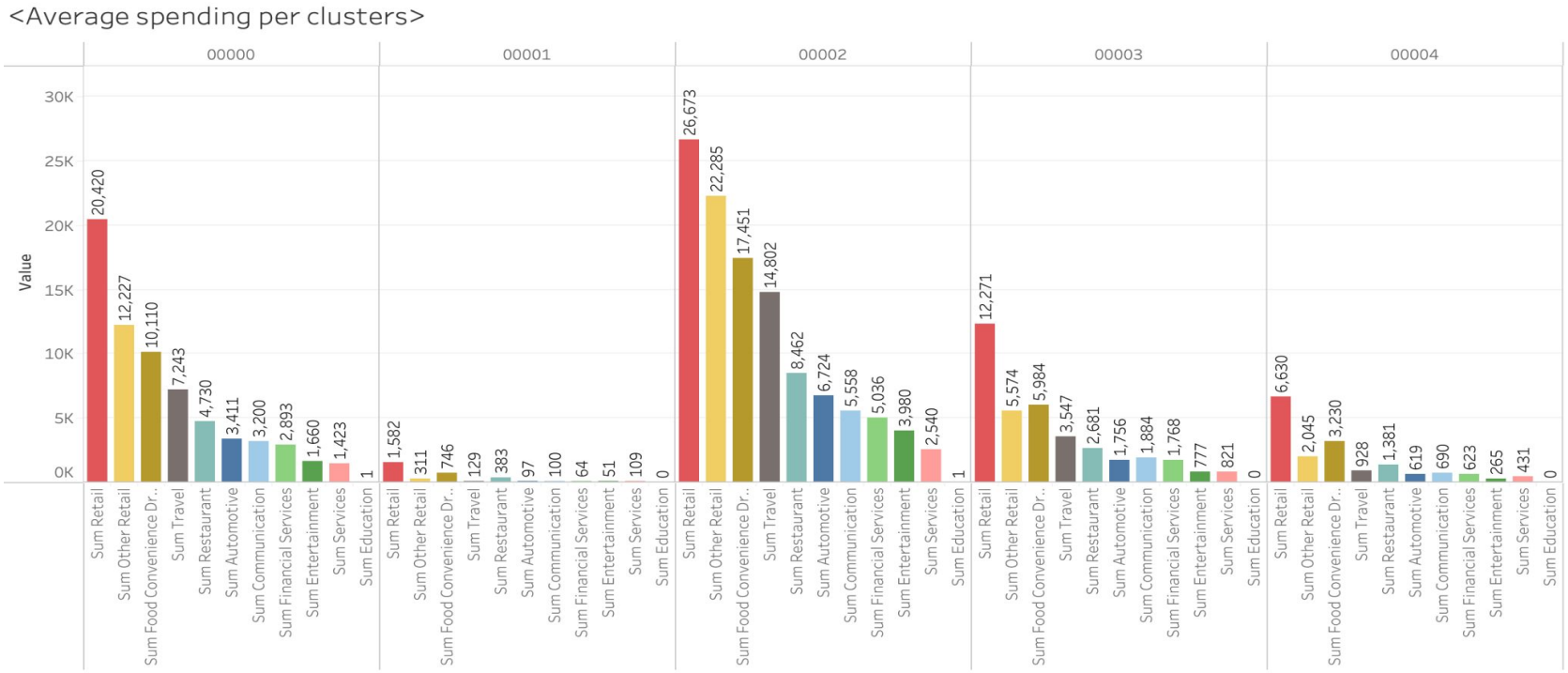
- **Single or newly married without children, with relatively low income. Like shopping, enjoy convenient food and have interest in financial services, save less and spend more.**
  - Average number of children: 1.67
  - Family composition: Mostly are married with no children (30.9%) and single female (20.2%)
  - Discretionary Spending Income: **$35,000 - $44,999**
  - Marital Status: 64% is married
  - Gender: 44% male
  - Average of overall spending: **$27,307**

K-Means Clustering Result

Cluster 5: Married, good at budgeting

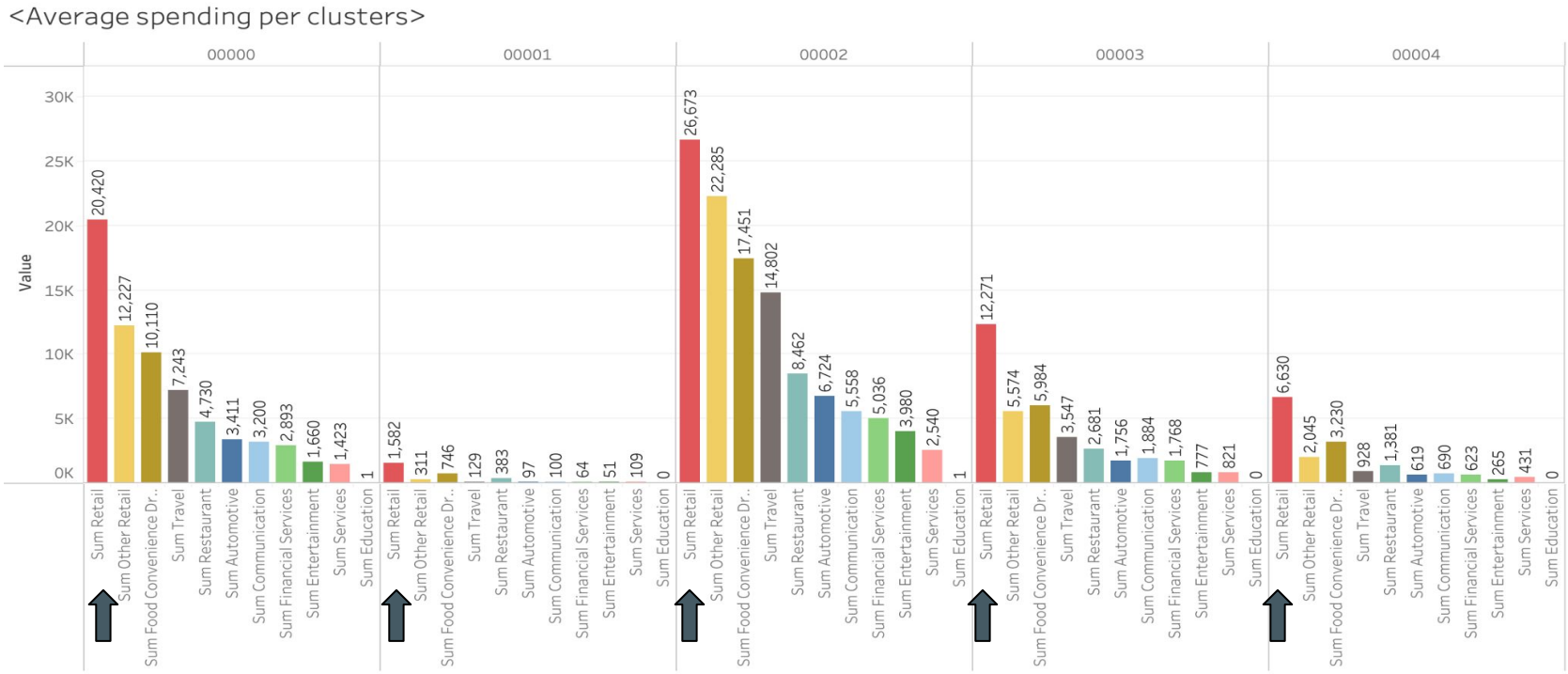
- **Low income married, like shopping and convenient food. Have a good balance in saving and spending.**
  - Average number of children: 1.62
  - Mostly are **married with children (40.4%) and married with no children (23.4%)**
  - Discretionary Spending Income: $25,000 - $34,999
  - Marital Status: 61% is married
  - Gender: 46% male
  - Average of overall spending: **$15,960**

K-Means clustering results



K-Means clustering results

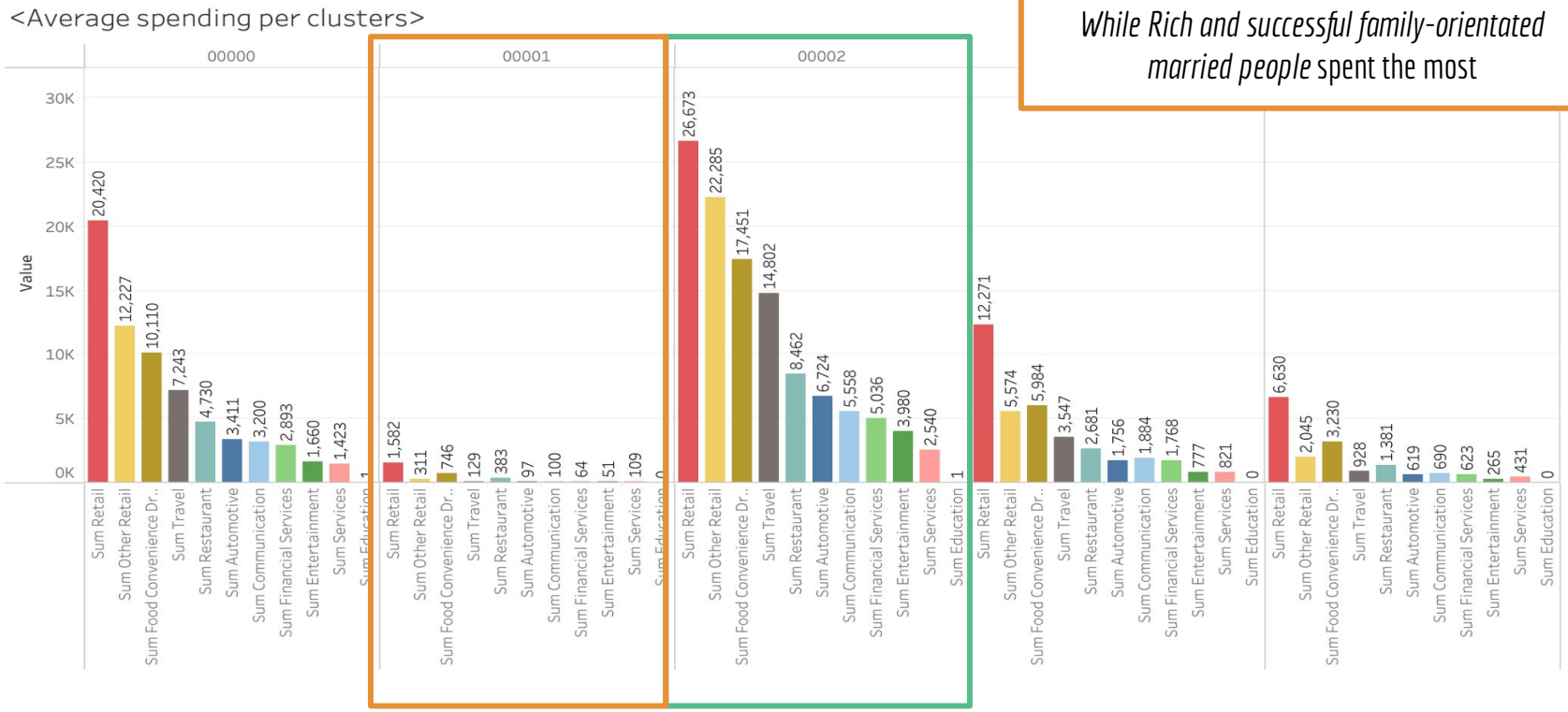
In each cluster, everyone spent the most on retail



K-Means clustering results

Financially conservative singles spent the least amount of money in every category

While Rich and successful family-orientated married people spent the most

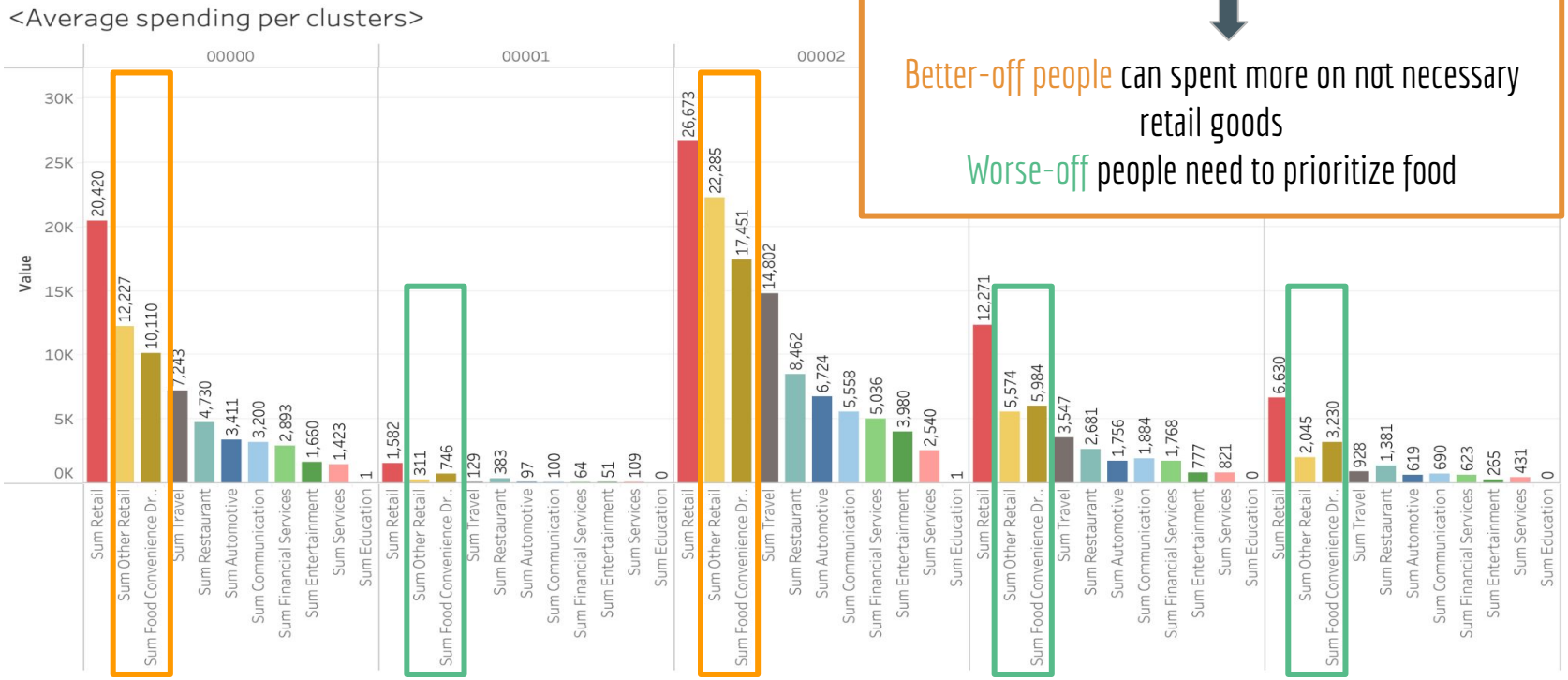


K-Means clustering results

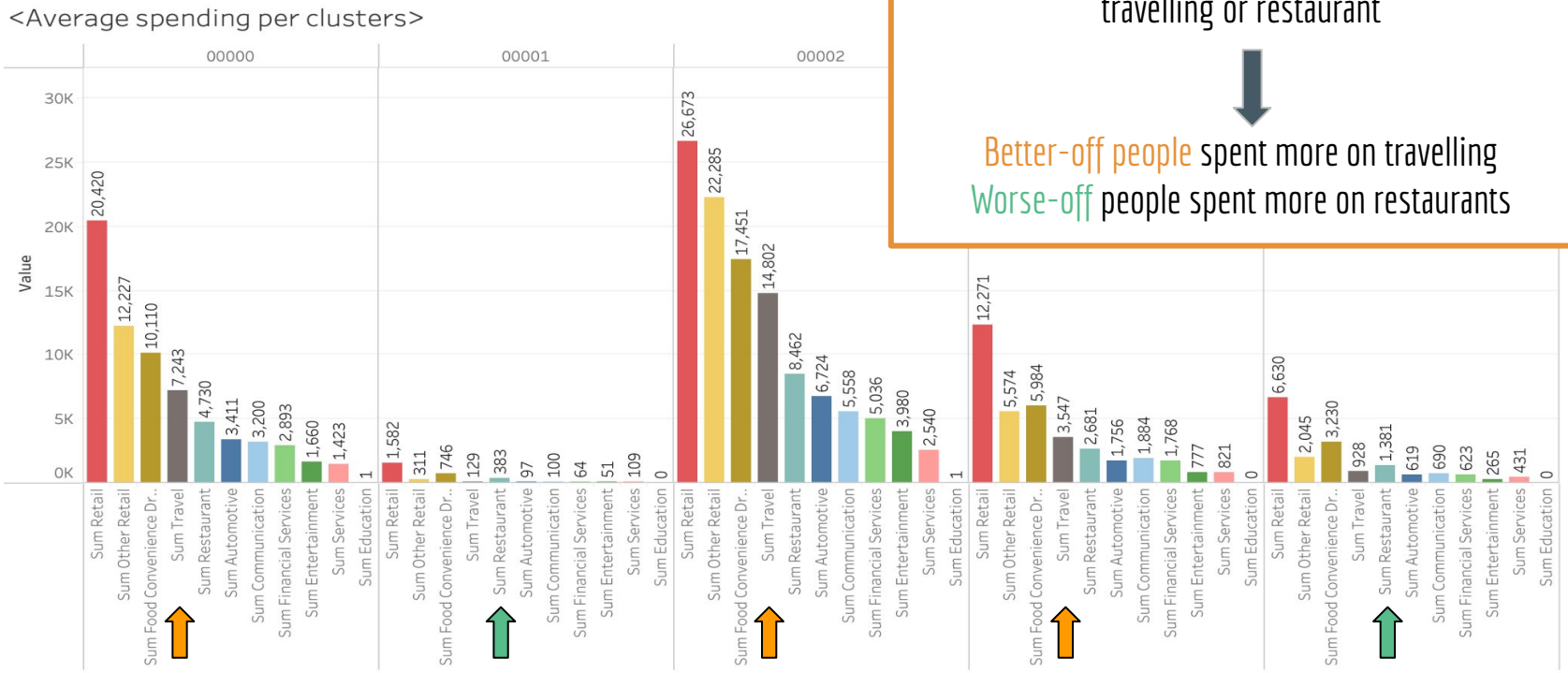
Second highest spending category varies

Better-off people can spent more on not necessary retail goods

Worse-off people need to prioritize food



K-Means clustering results



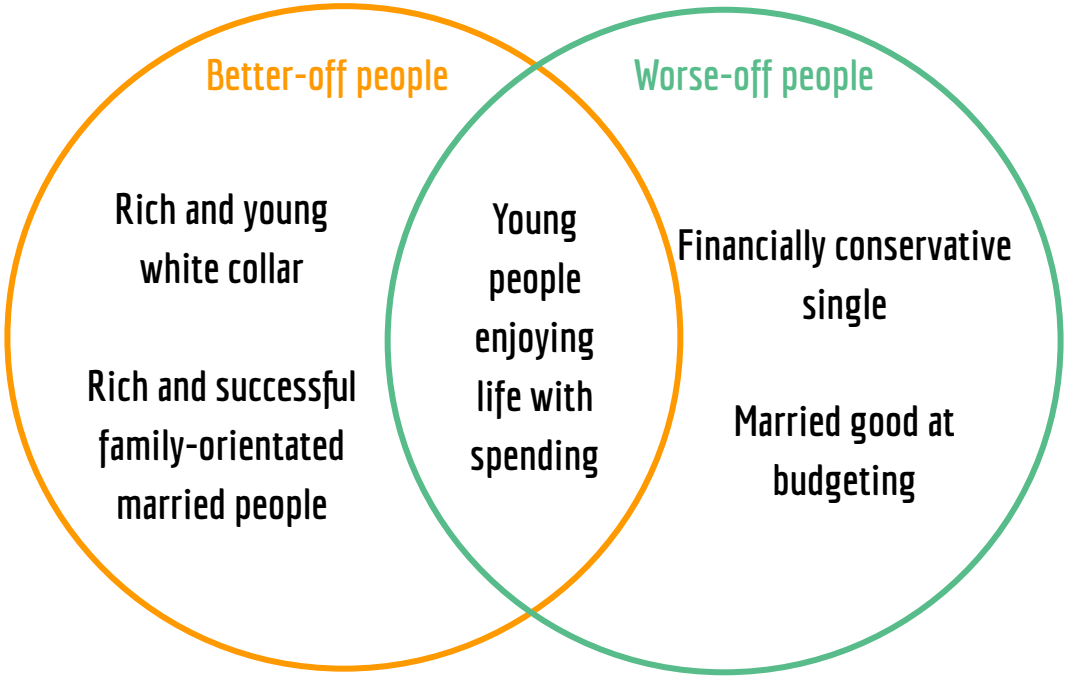
Income level determines whether they spent more on travelling or restaurant

Better-off people spent more on travelling
Worse-off people spent more on restaurants

Conclusions

K-Means clustering results

Key takeaways



K-Means clustering results

Key takeaways

Based on the clustering result, further marketing strategies should be conducted based on the saving behavior and the spending preference in different categories.

Focus on those prefer spending than saving

Cluster	Overall spending / Discretionary Spending Income
1	50%
2	12%
3	42%
4	68%
5	53%

Provide category specific recommendations

cluster	Highest spending category
1	retail, financial service, education
2	retail, convenient food
3	entertainment, education, travel
4	retail, convenient food
5	retail, convenient food

Limitations & Deficiencies

Limitations & Deficiencies

Limitations

- There are so many missing values (i.e. EDU) in original datasets
- Due to limited time, we didn't cover hierarchical clustering

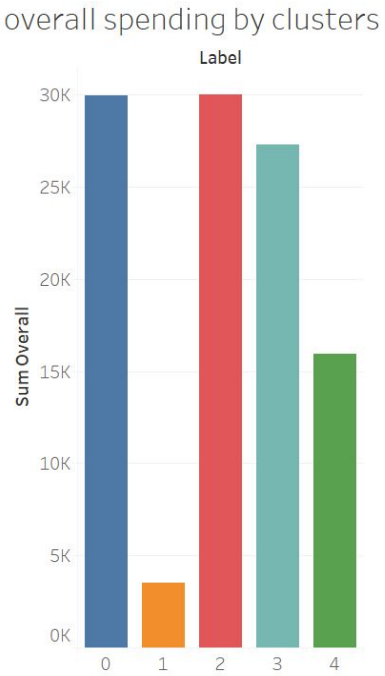
Deficiencies

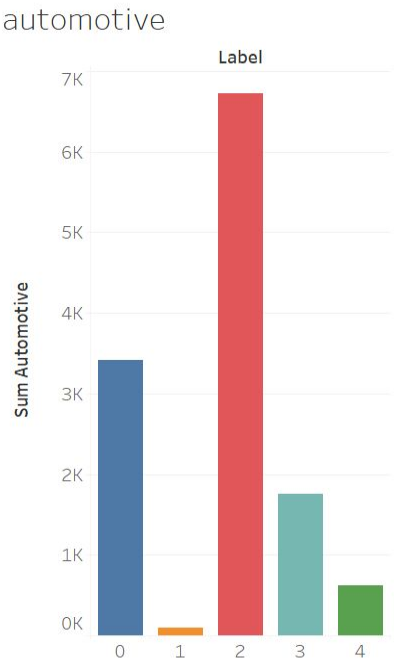
- We did post selection randomly from dataset 2, and didn't include all the data
- We sum all the spending among different categories from 2019Q3 to 2021Q2 to eliminate time series
- We didn't use any regression models to predict the potential spending with customer features

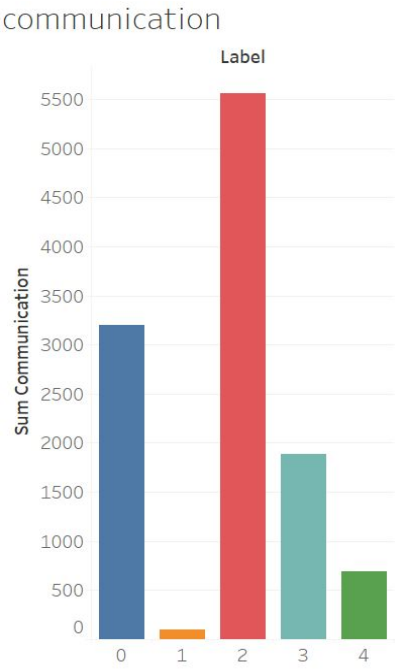
Thanks for your attention!

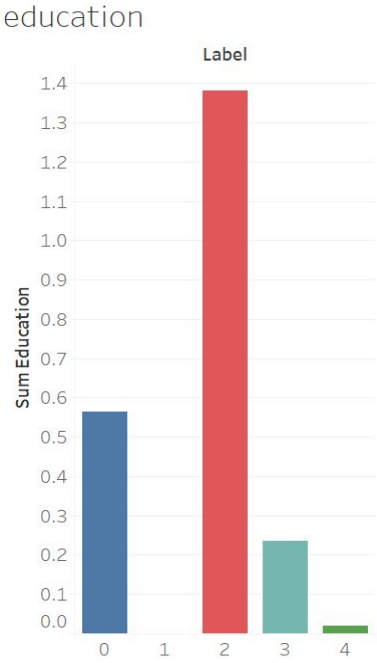
Appendix

Clustering results





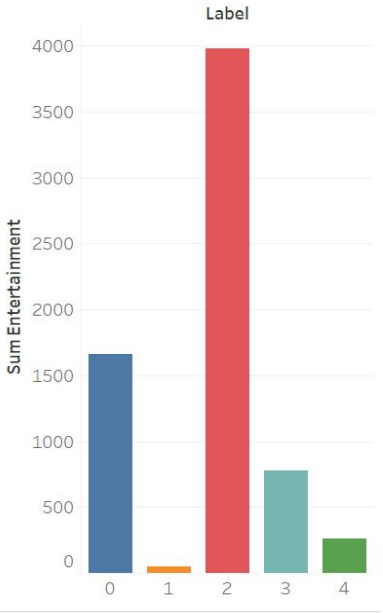




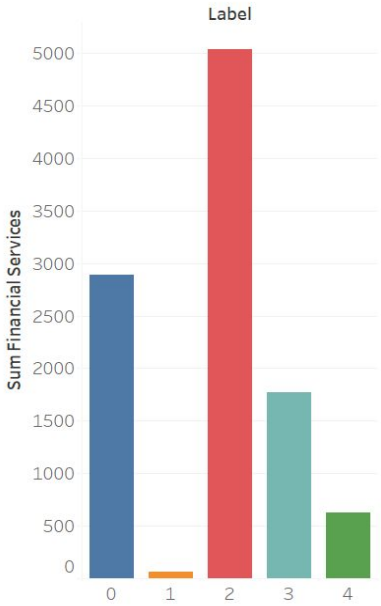
Appendix

Clustering results

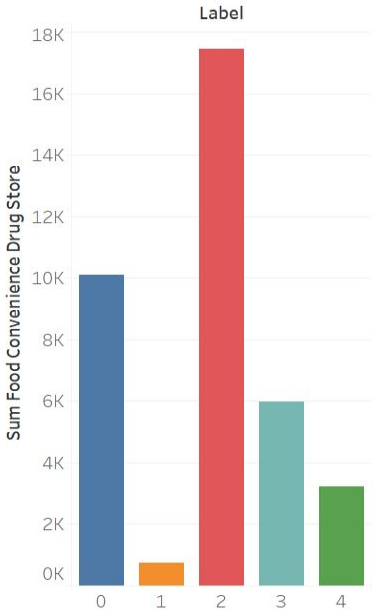
entertain



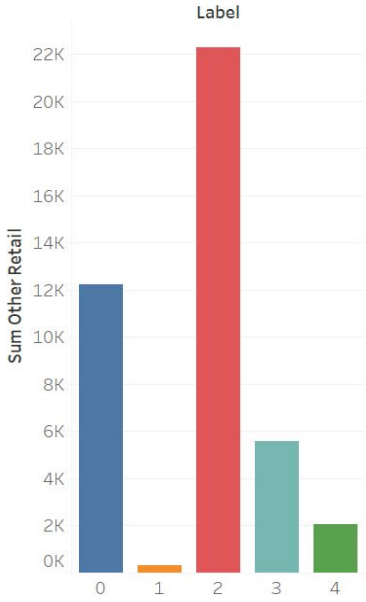
financial services



food convenience



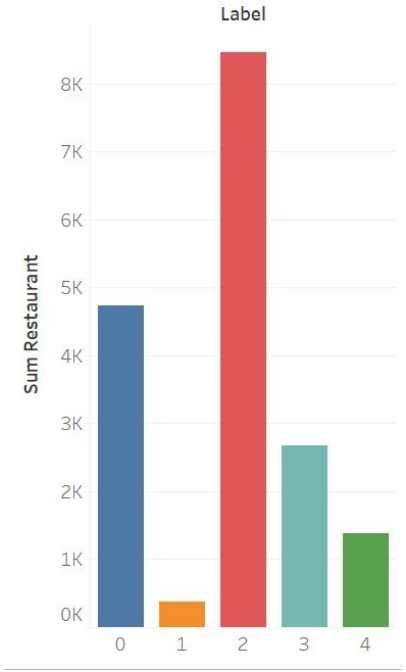
other retail



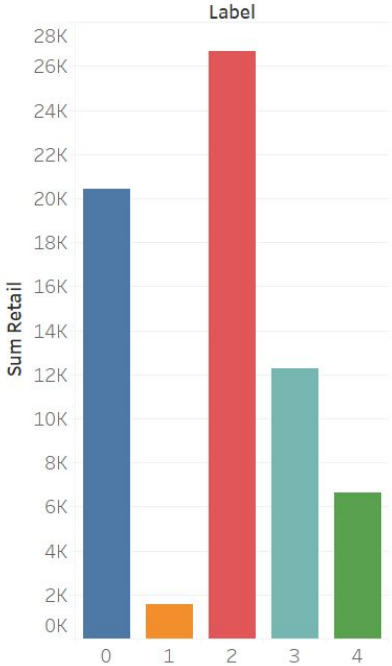
Appendix

Clustering results

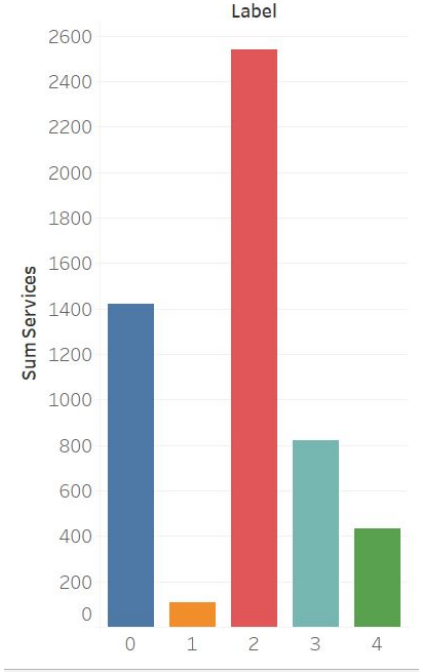
restaurant



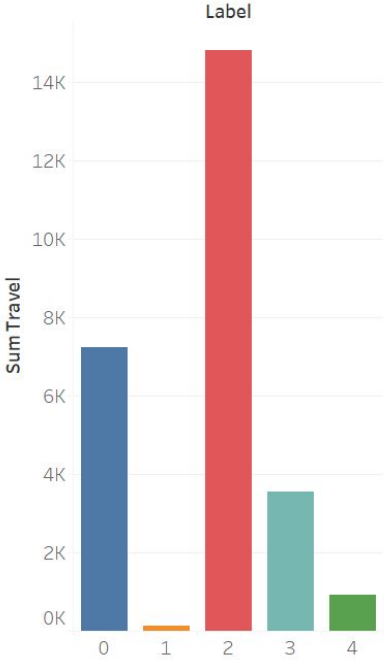
retail



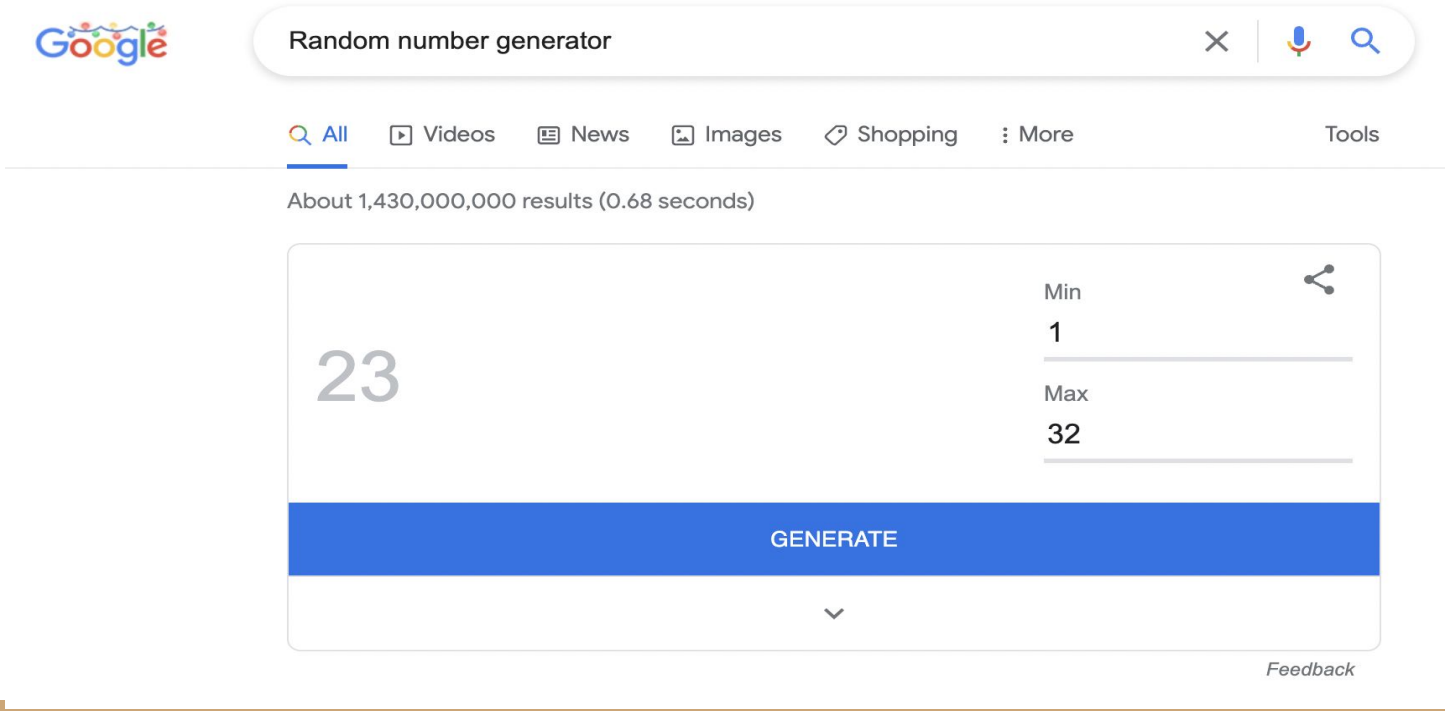
services



travel



1. Randomly generate 1 dataset from Epsolin Datasets 2



2. Combine Data Sets 1 & Data Set 2_23

- Combine the two datasets using Personal Sequence Number (PSN)
- Total Sample Size (n=144,629)

3. Process missing data

-

4. Modeling

- Machine Learning Models

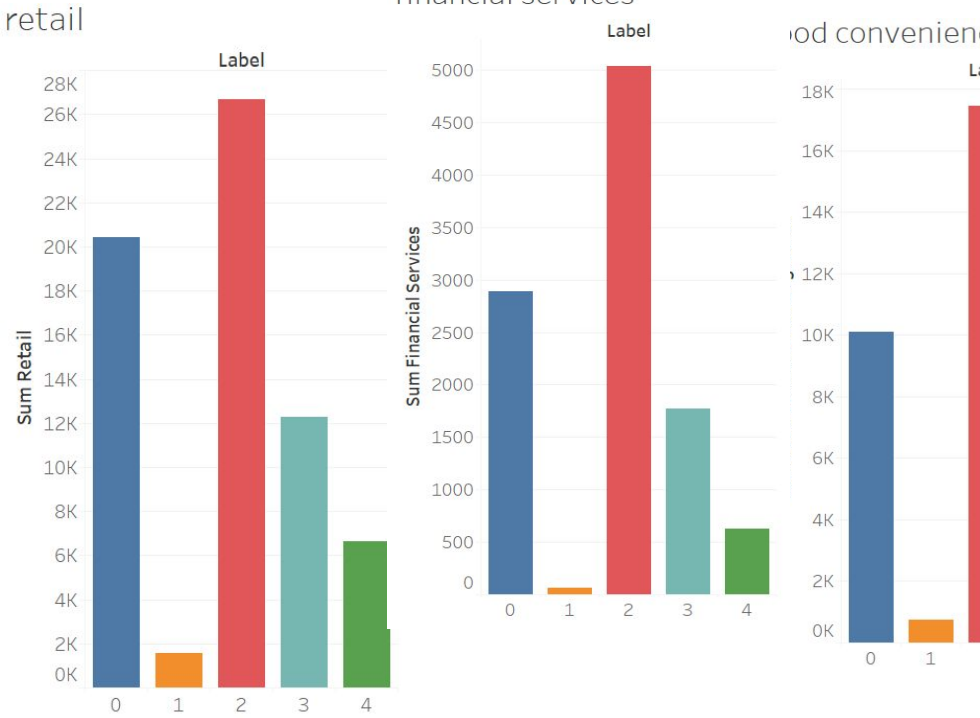
  KMeans clustering

  PCA with KMeans clustering

KMeans clustering results

Rich and young white collar

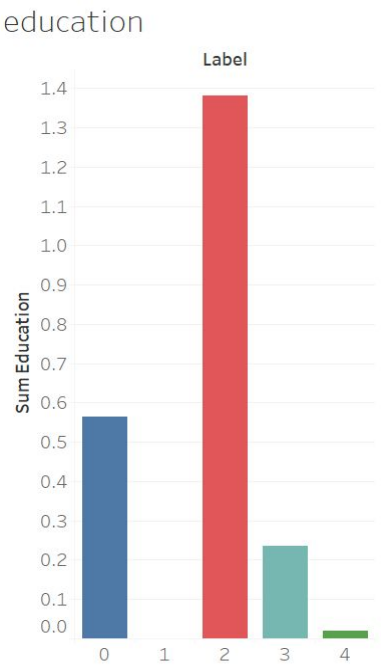
- cluster 1: Single or newly married without children, with relatively high income. Enjoy shopping, convenient food, have saving habit.
- Average number of children: 1.79
- Family composition: Mostly are married with no children (26.54%) and single female (24.6%)
- Discretionary Spending Income: $55,000-$64,999
- Marital Status: 75% married
- Gender: 43% Male
- Average of overall spending: $29,949

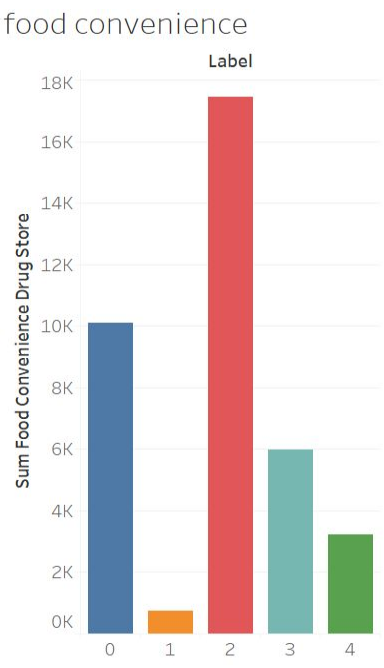


KMeans clustering results

Financially conservative single

th little need to spend on children's education, enjoy convenient food. With low spending power, prefer saving.
Average number of children: 1.59
Family composition: Mostly are single female (26.35%) and single male (19.9%)
Discretionary Spending Income: $25,000-$34,999
Marital Status: 57% married
Gender: 38% male
Average of overall spending: $3,532

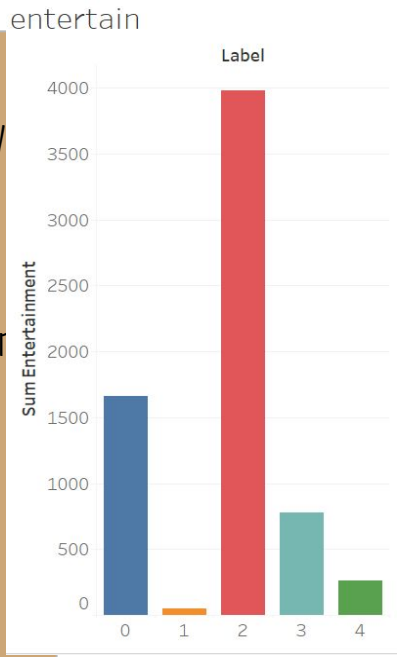


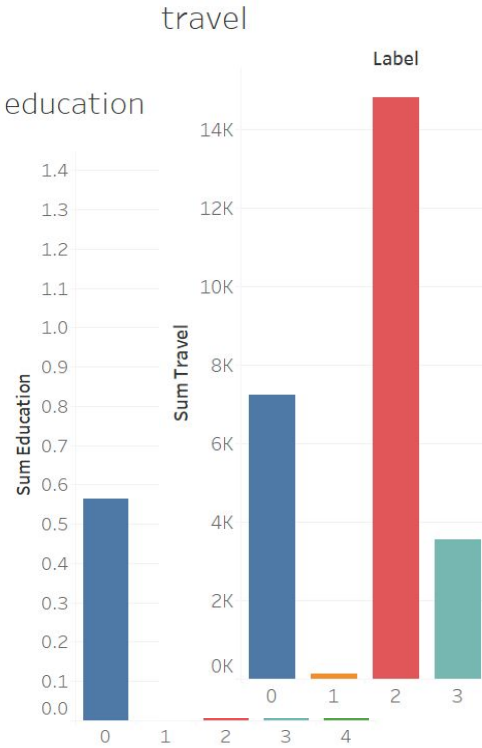


KMeans clustering results

Rich and successful family-orientated married people

en's
education spending, have enough money to spend much on luxury goods like entertainment and travel. Have a balance in saving and spending.
Average number of children: 1.88
Family composition: Mostly are married with children (35.18%) ar
married with no children (46.9%)
Discretionary Spending Income: $65,000 - $74,999
Marital Status: 83% is married
Gender: 47% male
Average of overall spending: $30,000

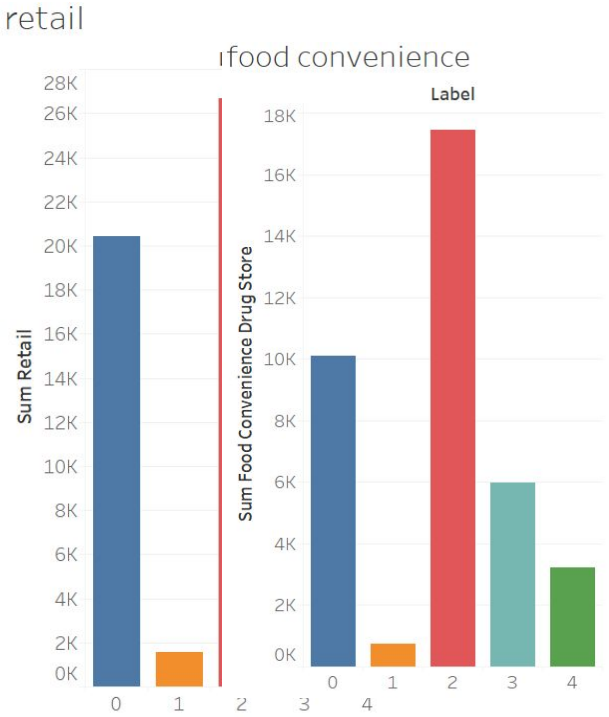




KMeans clustering results

Young people enjoying life with spending

ren, with relatively low income. Like shopping, enjoy convenient food and have interest in financial services, save less and spend more.
Average number of children: 1.67
Family composition: Mostly are married with no children (30.9%) and single female (20.2%)
Discretionary Spending Income: $35,000 - $44,999
Marital Status: 64% is married
Gender: 44% male
Average of overall spending: $27,307



KMeans clustering results

Married good at budgeting

nd convenient food.
Have a good balance in saving and spending.
Average number of children: 1.62
Mostly are married with children (40.4%) and married with no children (23.4%)
Discretionary Spending Income: $25,000 - $34,999
Marital Status: 61% is married
Gender: 46% male
Average of overall spending: $15,960