CUSTOMER SEGMENTATION

MAXIMIZING REVENUE AND JUDGING CAMPAIGN EFFECTIVENESS

DATA SOURCE AND BRIEF DESCRIPTION

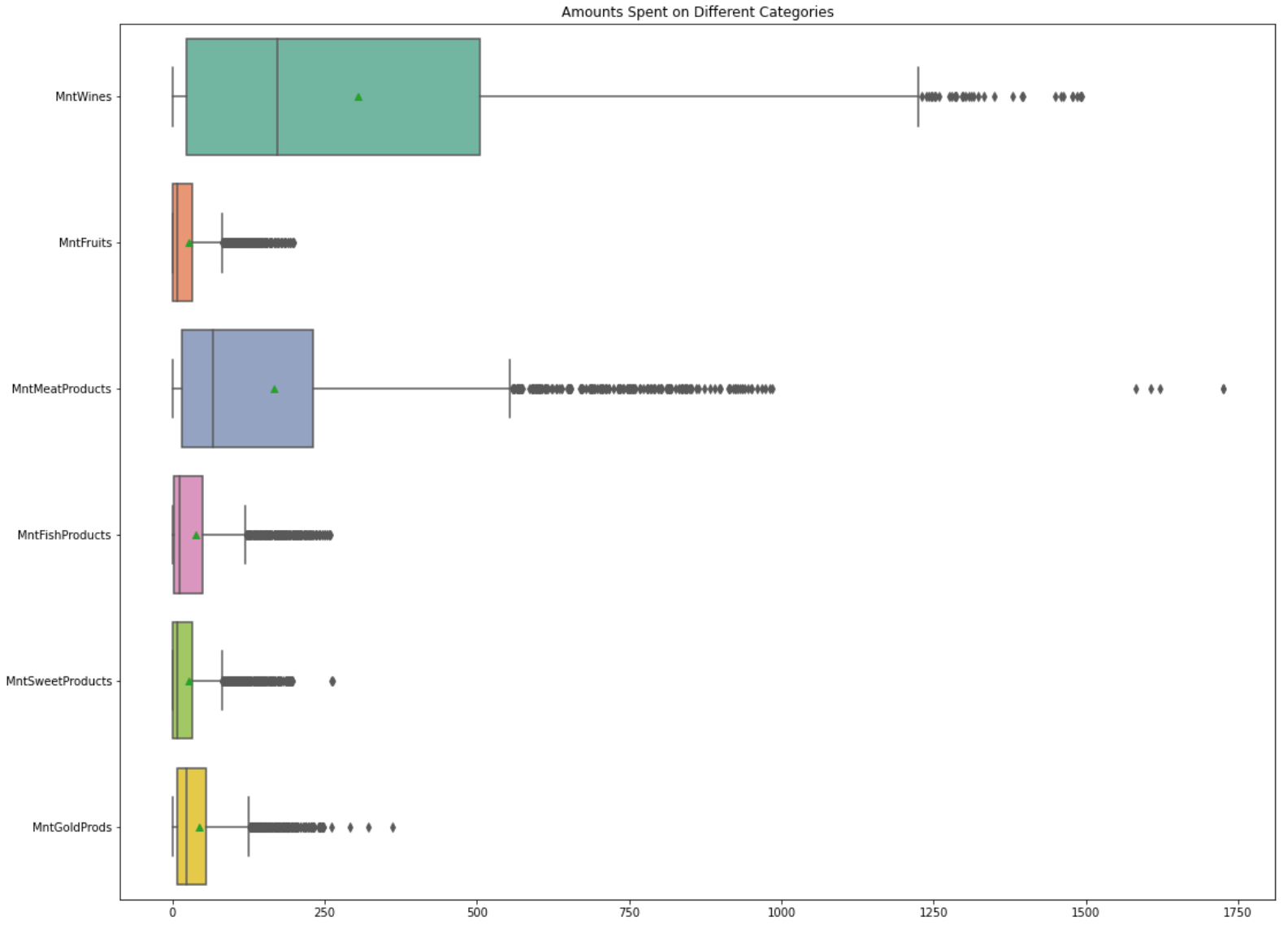
- Data source: https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis or https://www.kaggle.com/datasets/jackdaoud/marketing-data?select=ifood_df.csv. Provided by Dr. Omar Romero-Hernandez of UC Berkeley
- 2240 individual observations on
  - Customer characteristics (year of birth, education, income, number of children, etc.)
  - Products (aggregated amounts spent of wines, fruit, sweets, etc.)
  - Promotion acceptance for 5 campaigns
  - Number of purchases done on the web, using catalog, or in store, as well as the total number of web visits
  - Span 796 days (2012-2014)
  - Some missing income observations and extreme outliers in income and age

KEY QUESTIONS

- Who are the customers that bring the most revenue to the company?
- Who are the customers who could bring more revenue to the company?
- To which campaigns do these customers respond?
- Where do different groups of customer shop? (in store, online, over catalog)
- Any additional business insights about the customers?

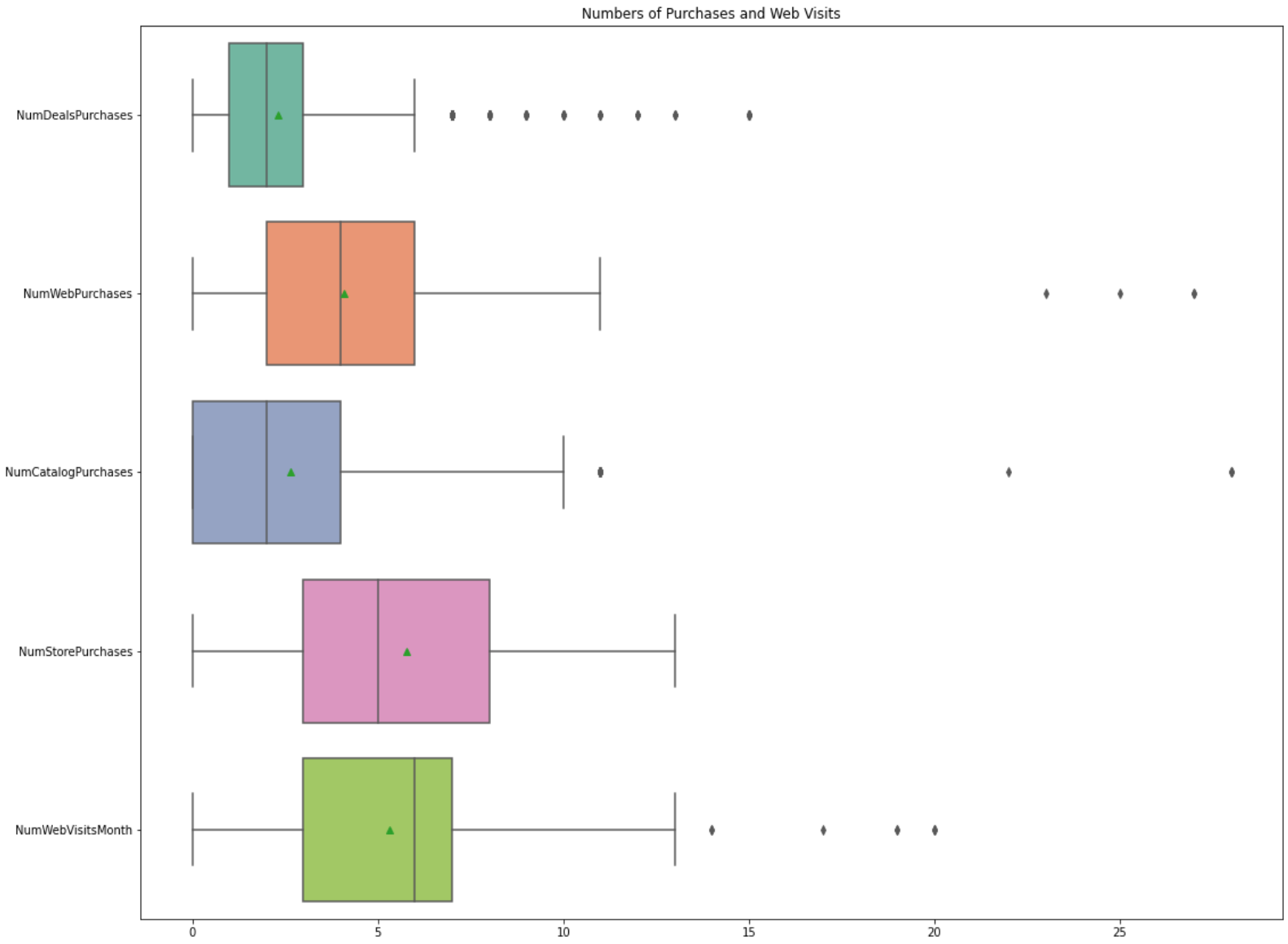
AMOUNTS SPENT BY CATEGORY

- High spending on wines and meat
- Spending amounts are skewed to the right for all categories
- Questions: What can we say about the spending patterns of different groups of customers?



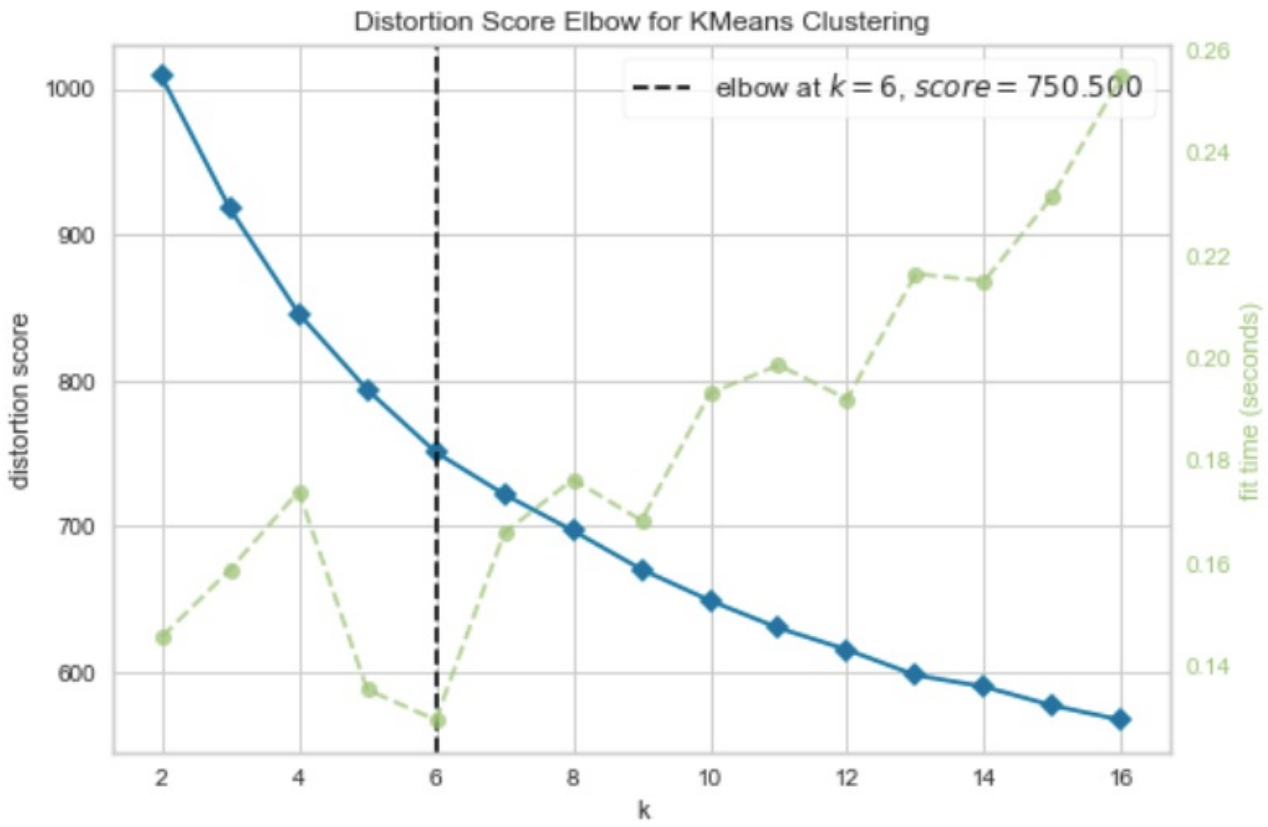
NUMBERS OF PURCHASES BY MEDIUM

- Customers shopping in store make the greatest median number of purchases, followed by web and catalog purchases
- High median number of monthly web visits
- Questions: Who shops in store? Who shops on the web? Who uses the catalog?

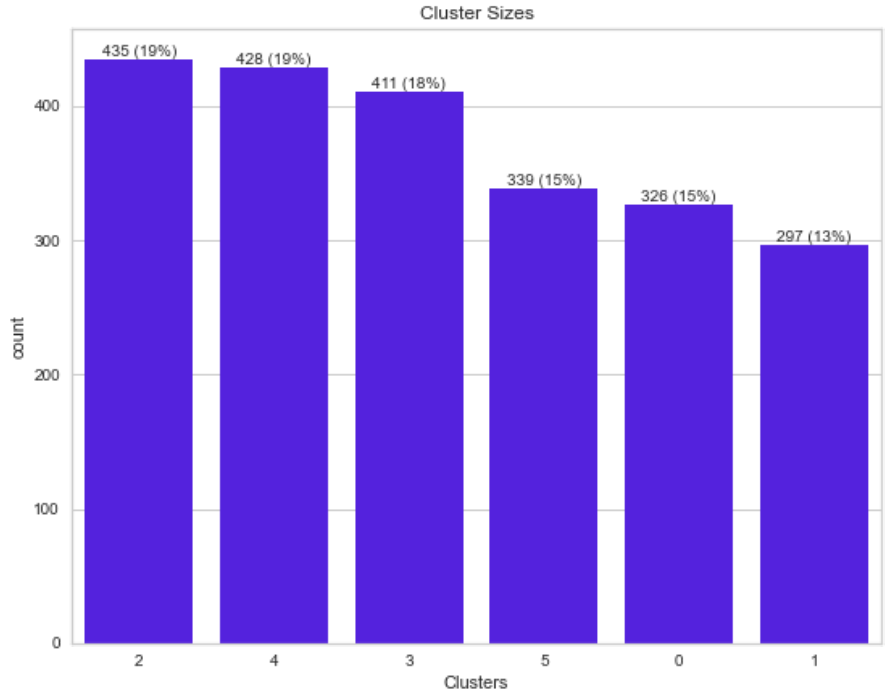


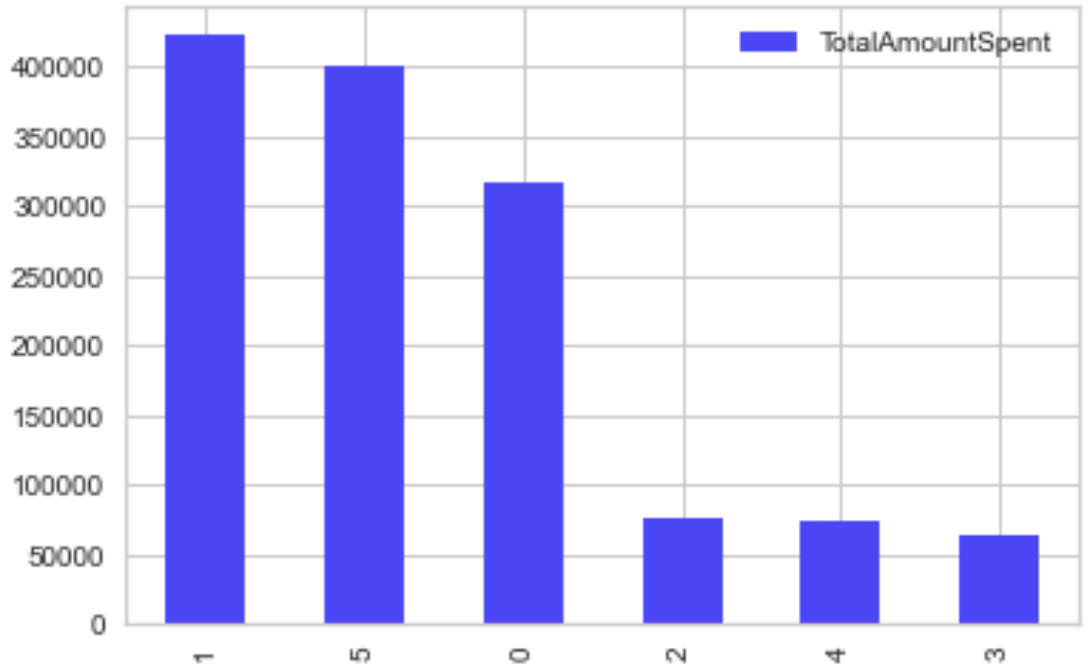
KMEANS CLUSTERING ALGORITHM

- Select k as the number of groups to cluster for.
- Randomly pick k points in the data as centroid points.
- Assign each non-centroid point to its closest centroid.
- Recalculate centroid points by taking the average of all surrounding points assigned to that cluster.
- Repeat steps 3-4 until calculated centroid points do not move anymore.



Good explanation at https://www.youtube.com/watch?v=hDmNF9JG3lo&ab_channel=ArtificialIntelligence-AllinOne



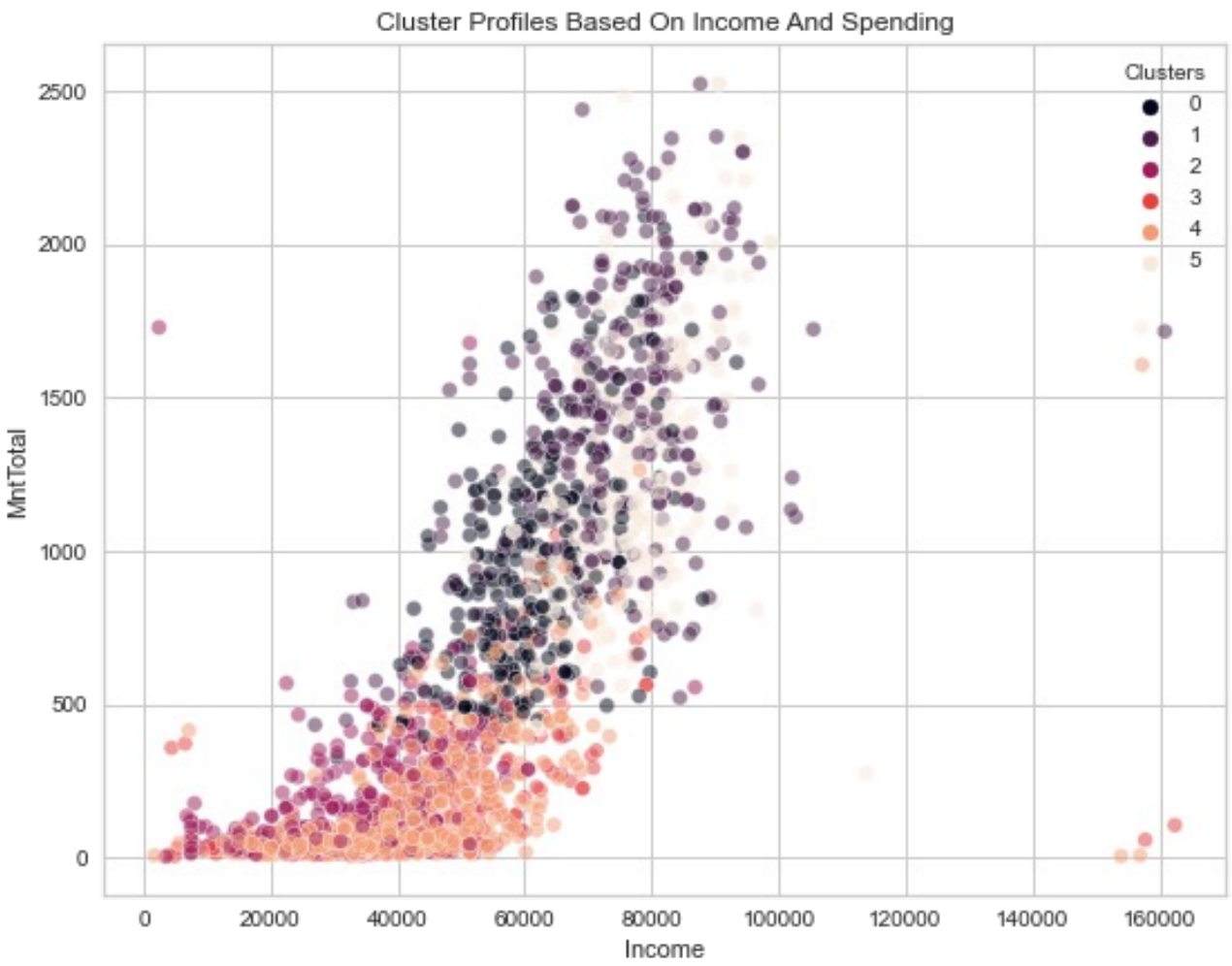


CLUSTER SIZES AND REVENUES

- 2 and 4 are the biggest clusters
- Clusters 1, 5, and 0 spend much more than the other clusters

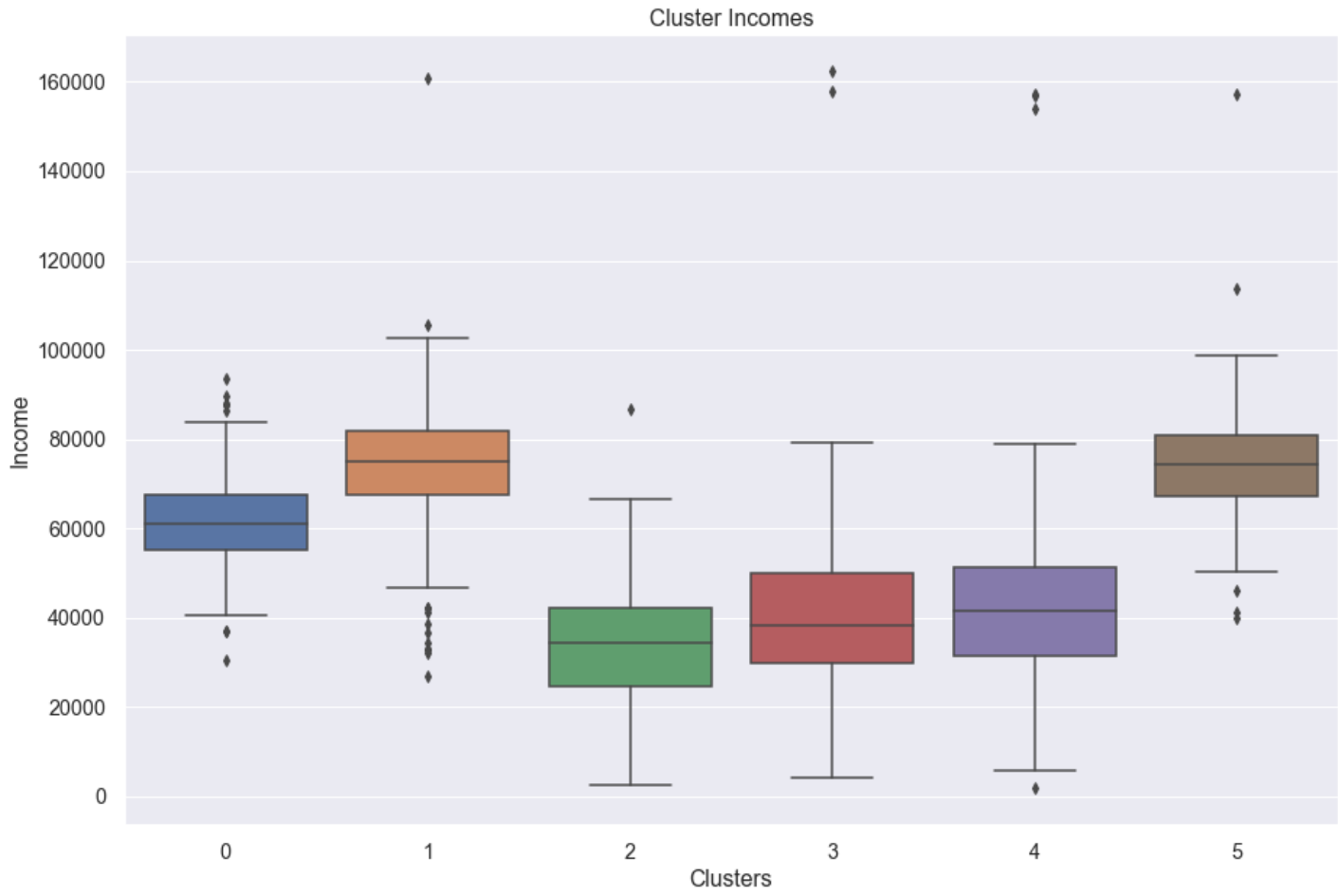
CLUSTER PROFILES BASED ON INCOME AND SPENDING

- A: Possibly not, as this look at income and spending for various clusters indicates



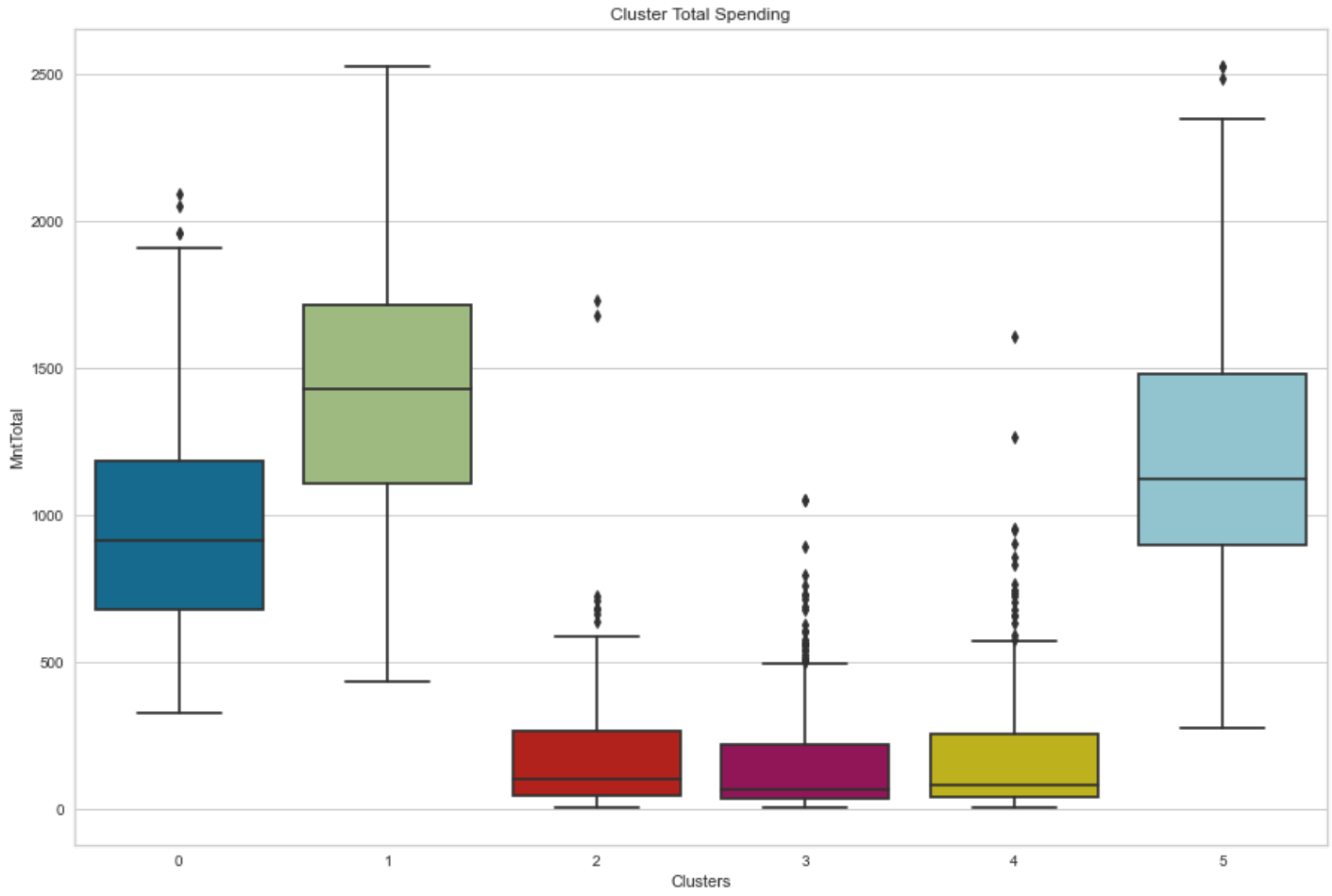
CLUSTERS BY INCOME

- Clusters 5 and 1 have the highest incomes, followed by Cluster 0.
- Clusters 2-4 have below median income for all customers.



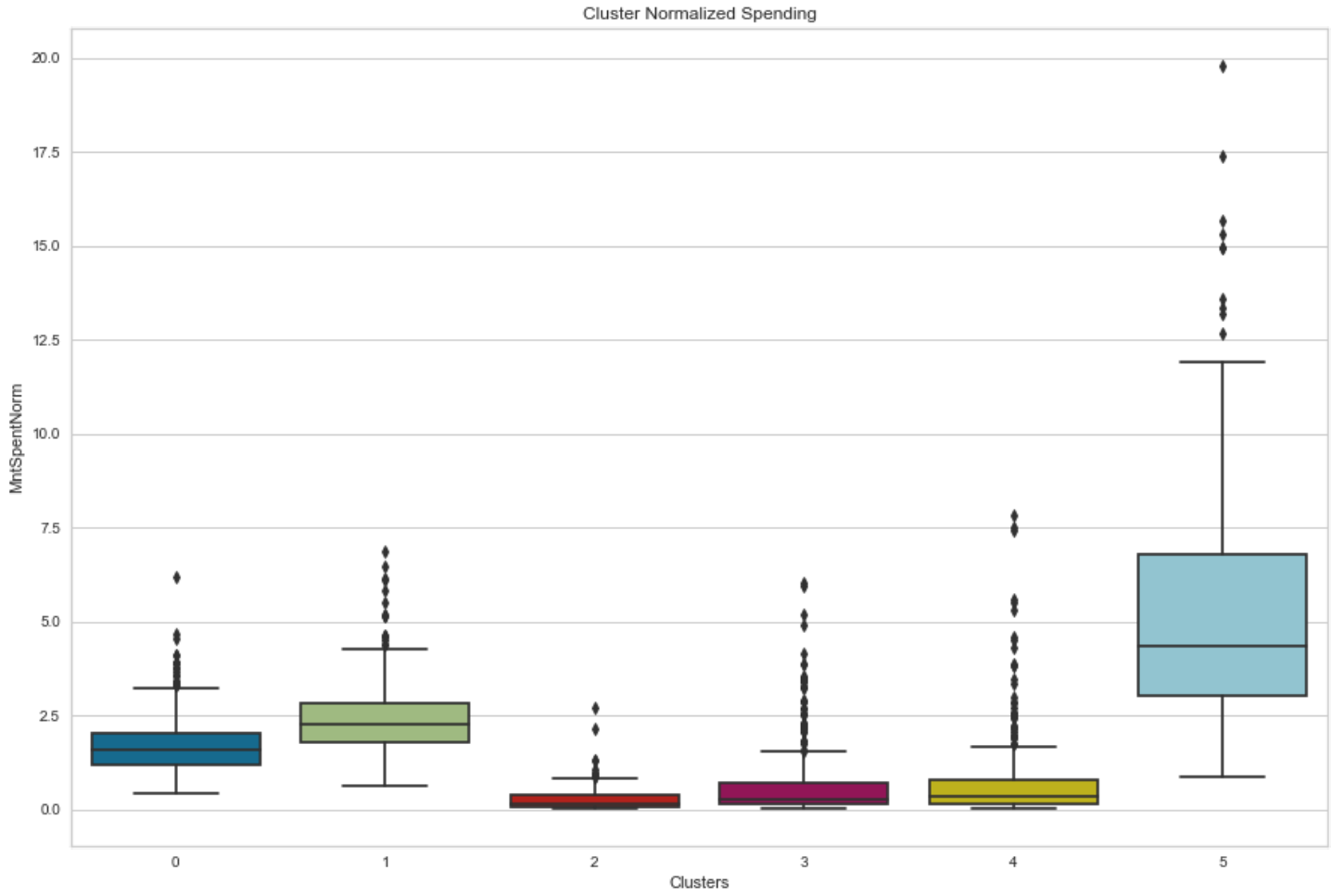
TOTAL SPENDING BY CLUSTER

- Cluster 1 customers spend the most, followed by Cluster 5 and 0 customers



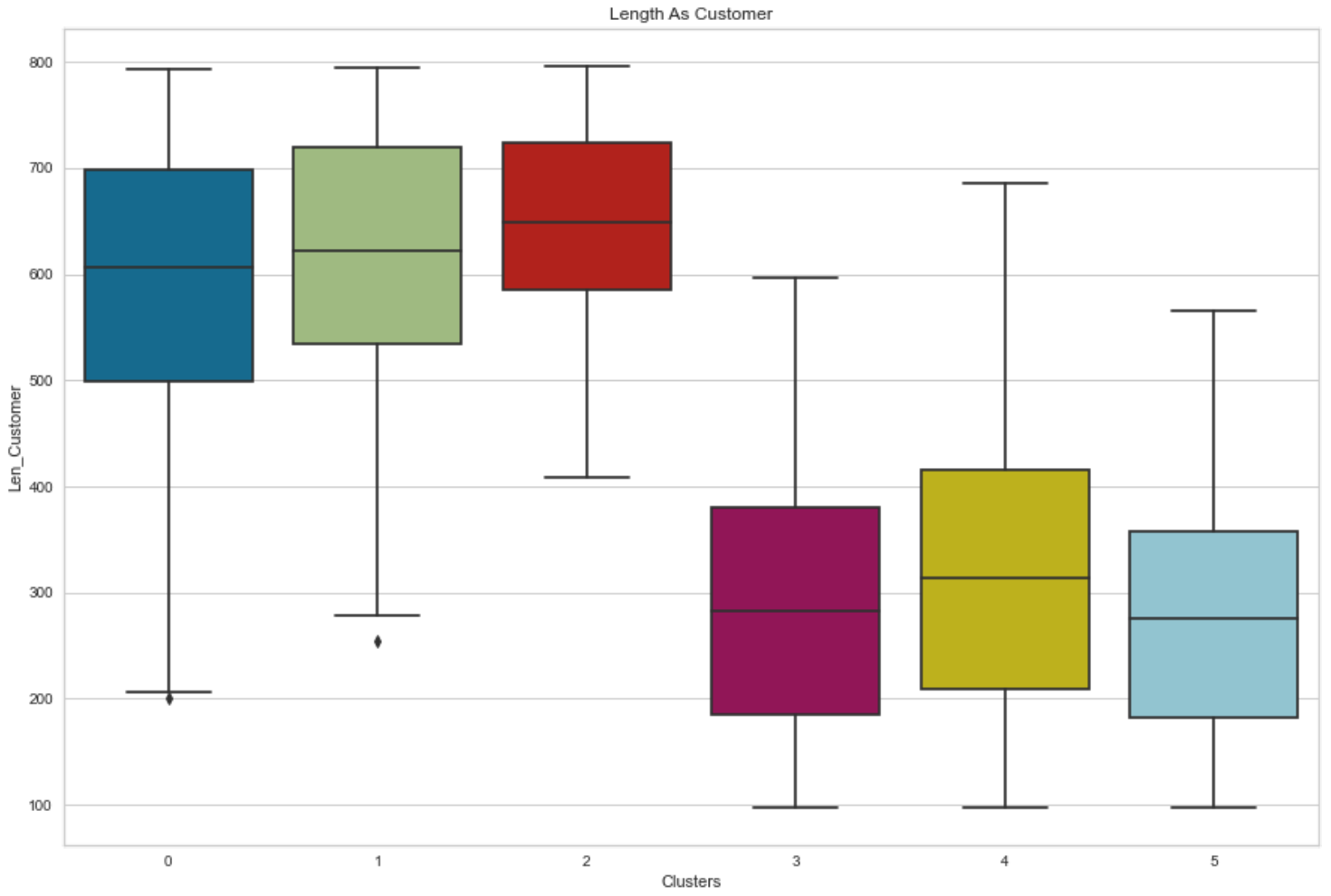
CLUSTER NORMALIZED SPENDING

- When normalized by length spent as customer, Cluster 5 customers spend much more than Cluster 1 customers!
- Cluster 0 customers follow
- These patterns hold by category as well (wine, meats, etc.)



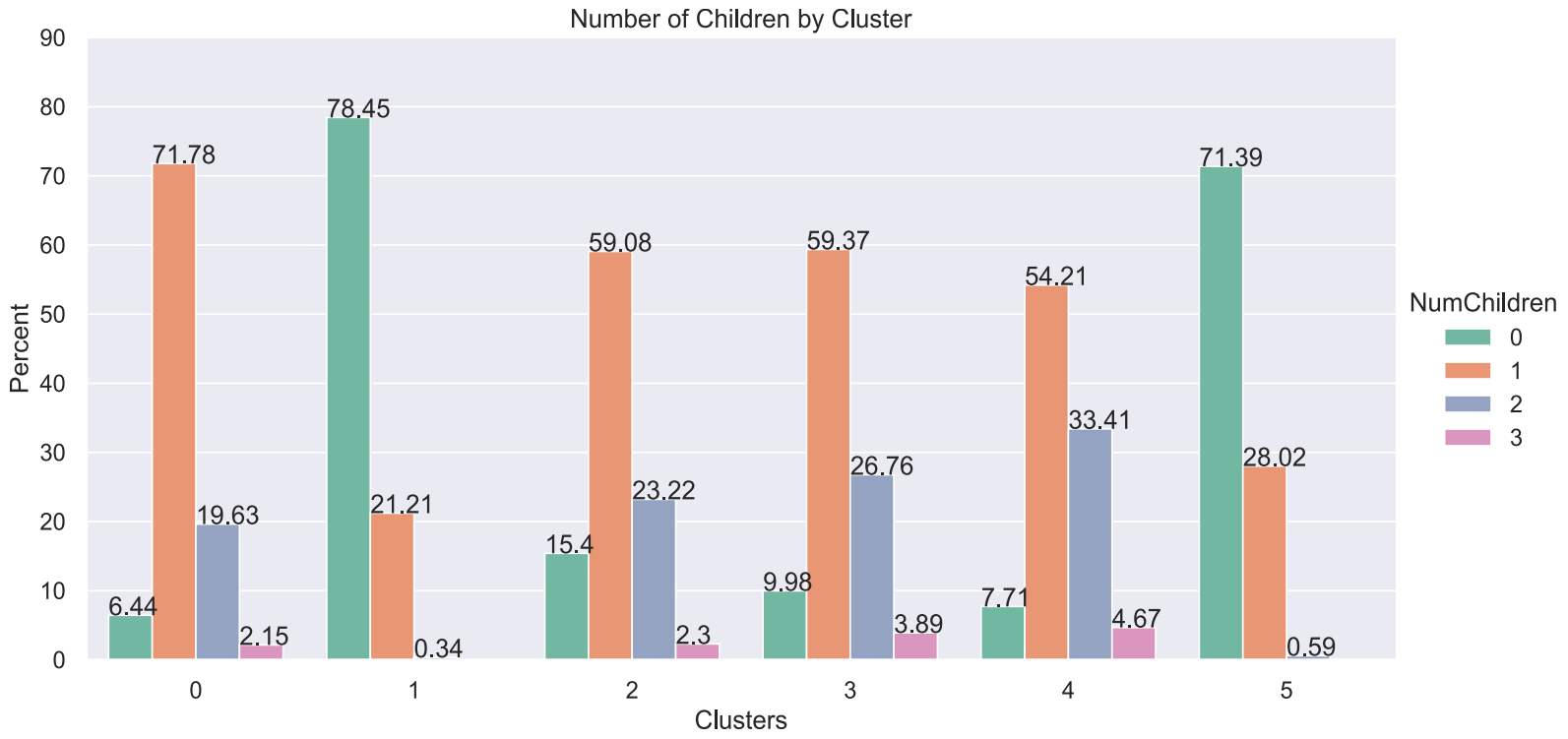
LENGTH AS CUSTOMER

- Cluster 5 customers are more recent, along with those in clusters 3 and 4
- Cluster 1 customers are loyalty customers, along with those in clusters 0 and 3



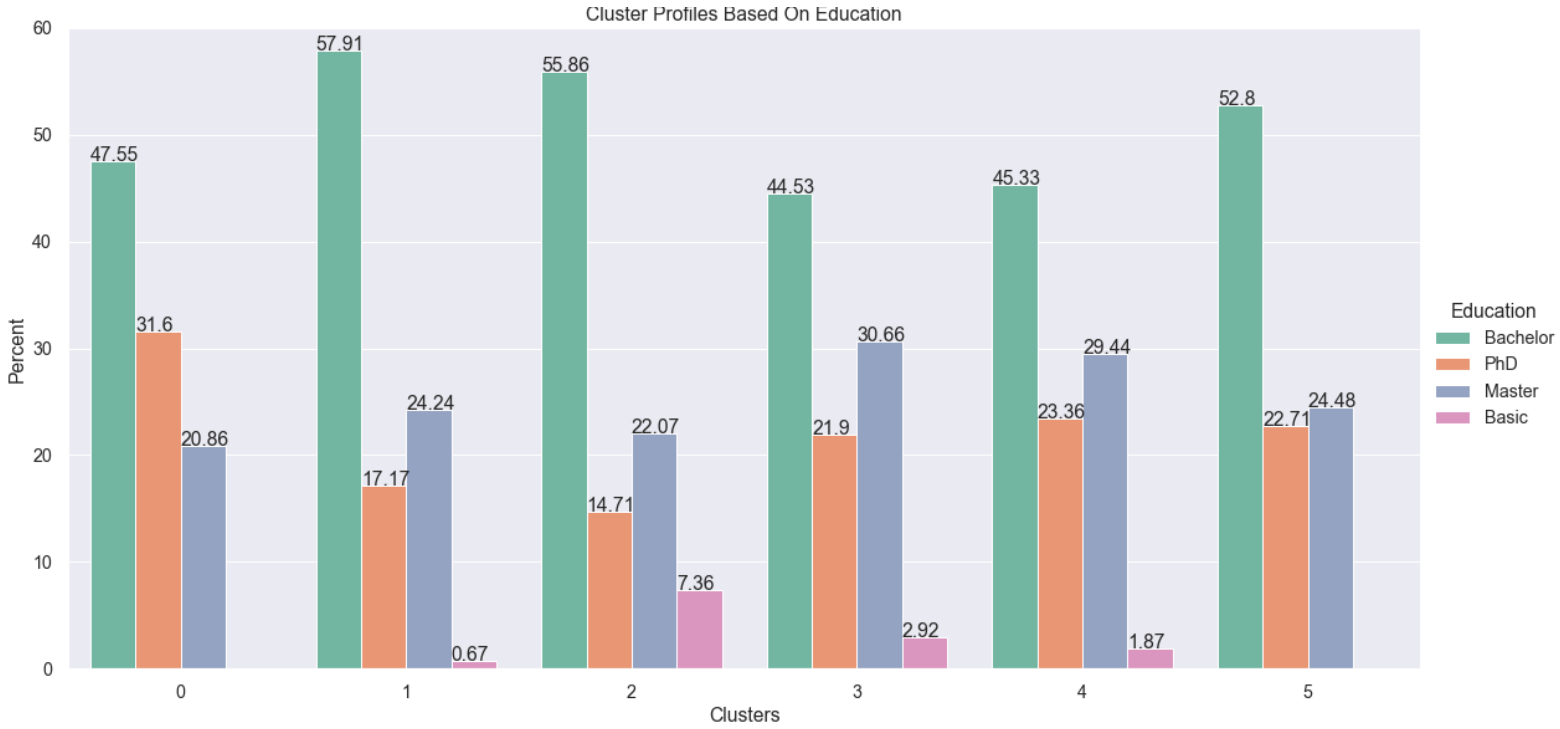
CLUSTERS BY NUMBER OF CHILDREN

- Cluster 5 customers are slightly more likely to have at least once child than Cluster 1
- Over 90% of Cluster 1 customers have at least 1 child, which could explain their lower spending



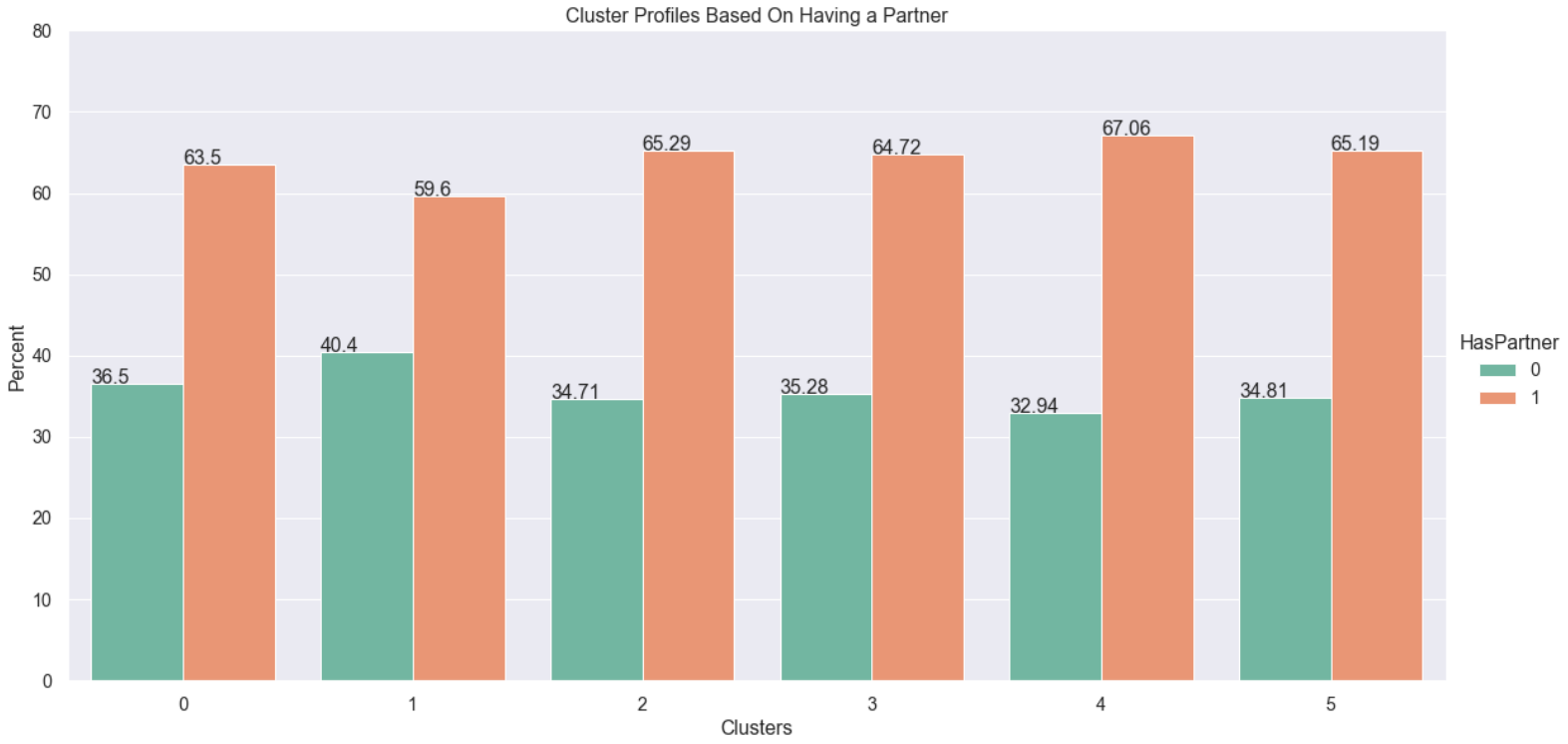
CLUSTERS BY EDUCATION

- Cluster 5 have more advanced degrees than Cluster 1
- Cluster 1 customers have the most PhDs, but probably have less disposable income due to having more children



CLUSTERS BASED ON HAVING A PARTNER

- Clusters are similar in that respect, with Cluster 4 customers most likely to have a partner and Cluster 1 least likely

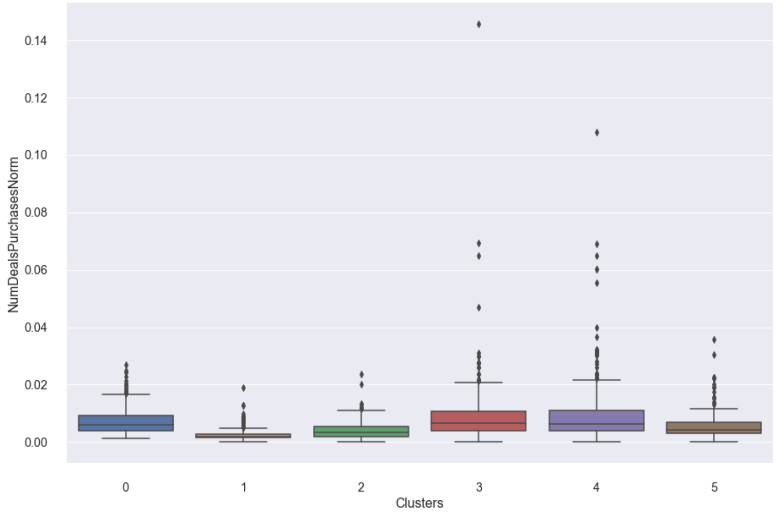


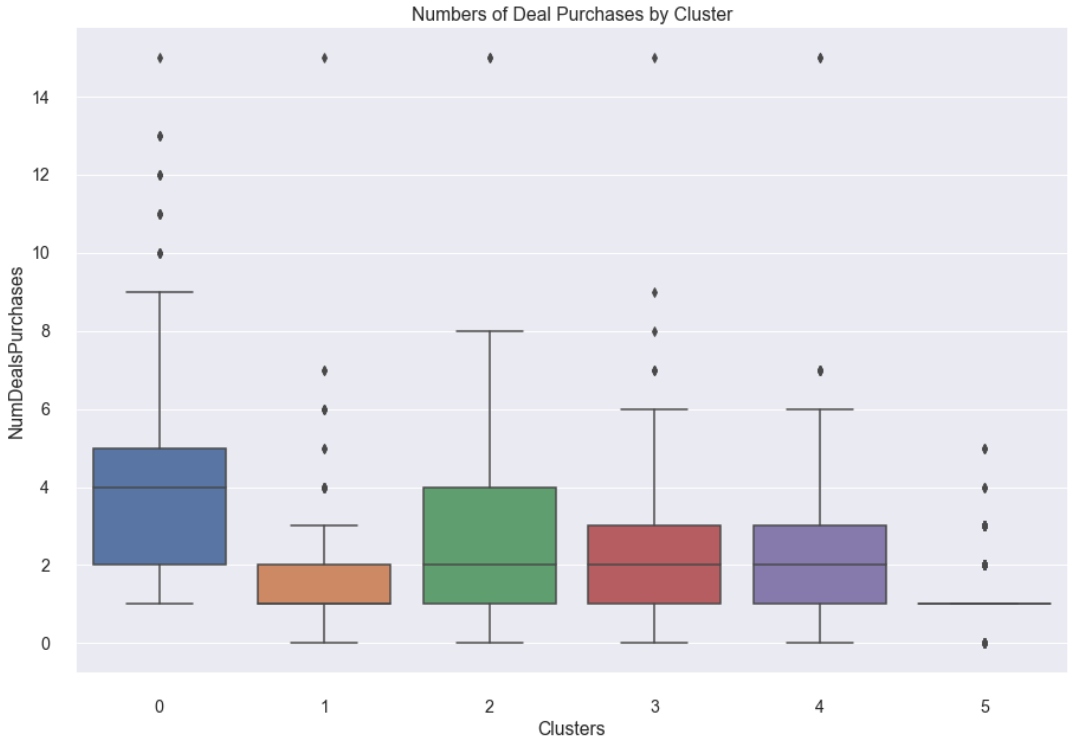
CLUSTER CHARACTERISTICS SUMMARY

- Clusters 1 and 5 are the high spenders with high income. Cluster 1 consists of loyalty customers but Cluster 5 customers spend the most per day. Apart from time spent as customers, they are similar in most key dimensions, with Cluster 1 being slightly better educated, having more children, and having slightly higher likelihood of having a partner.
- Clusters 0-2 are the loyalty customers and Clusters 3-5 are the newer customers.
- Cluster 1 customers also have higher than average income, but probably have less disposable income due to having children
- Clusters 2-4 have lower income, are more likely to have children, and spend less

DEALS

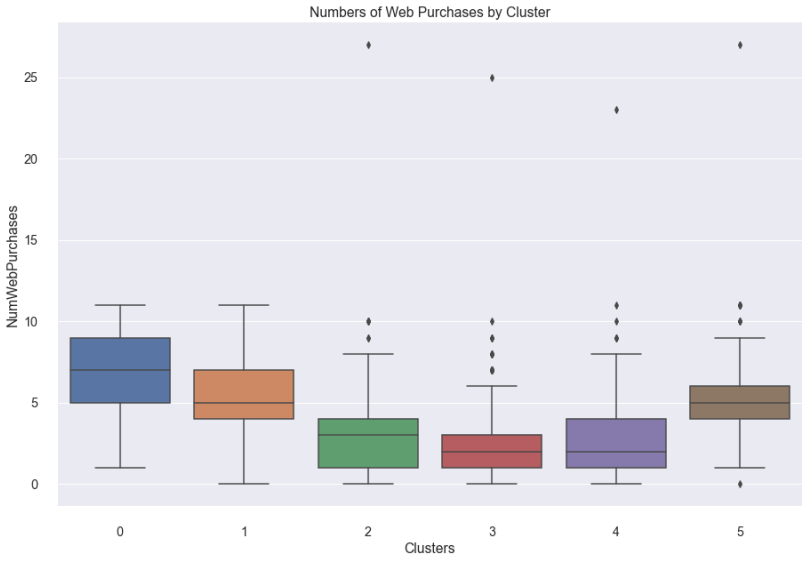
- Takeaway: If we want to reach Cluster 0, target deals at that cluster
- Clusters 1 and 5 are not very responsive to deals
- Upon normalization, Clusters 2-4 are also responsive to deals. *Target deals at these customers.*

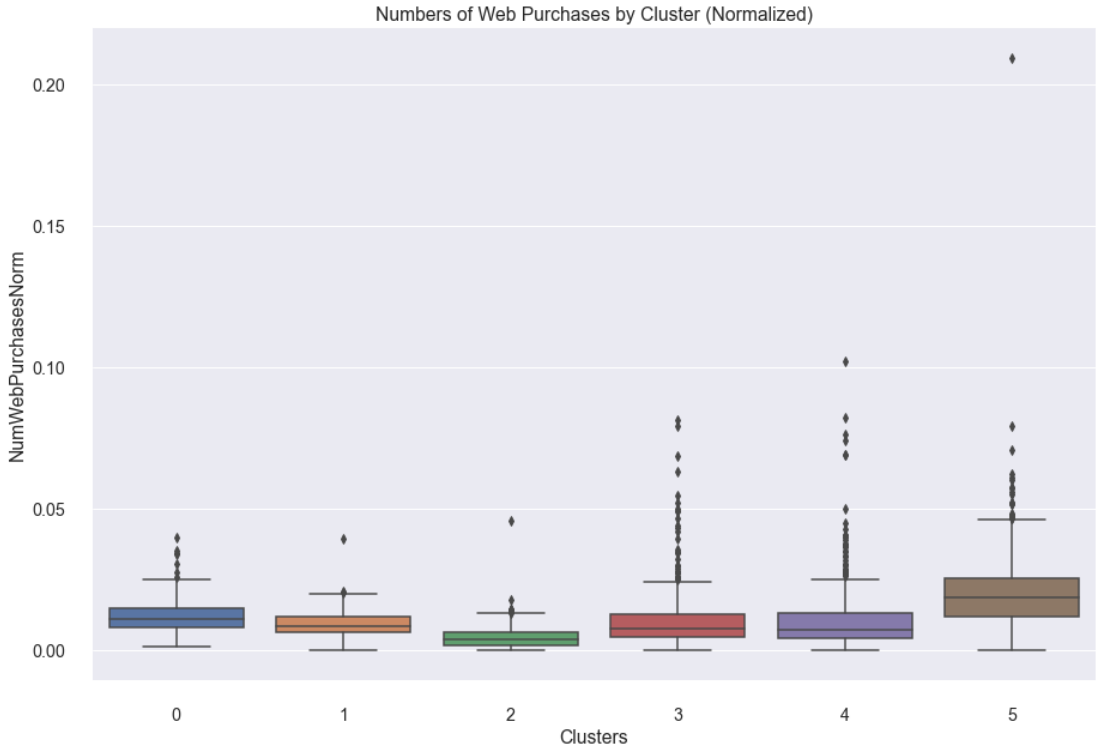




PURCHASE CHANNEL (WEB)

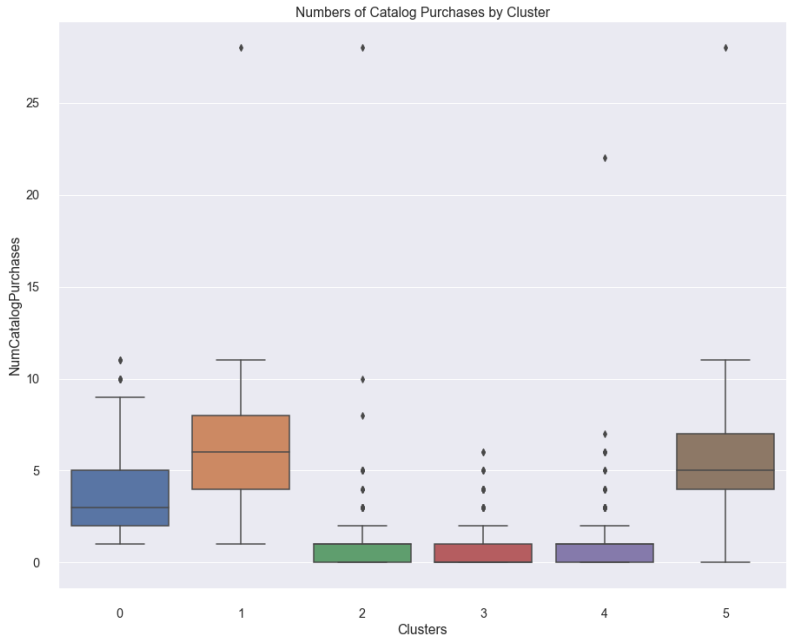
- When normalized, Cluster 5 shops on the web more than the other clusters. *Concentrate on making/keeping the web page appealing.*
- Clusters 0 and 1 follow

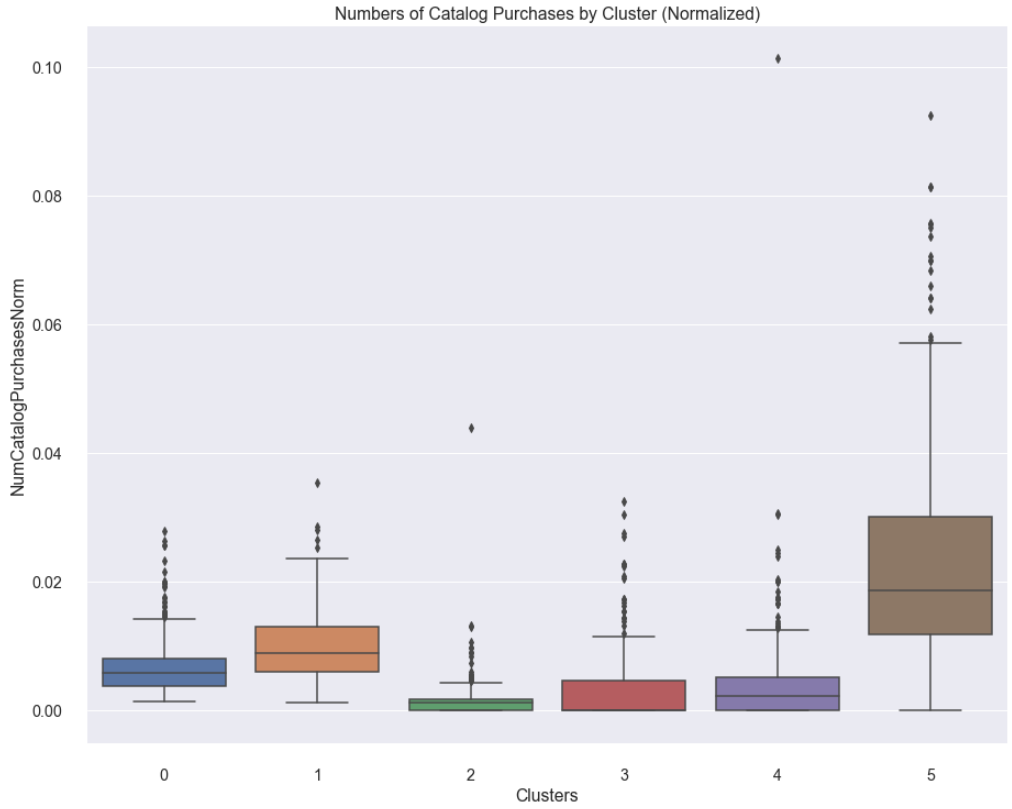




PURCHASE CHANNEL (CATALOG)

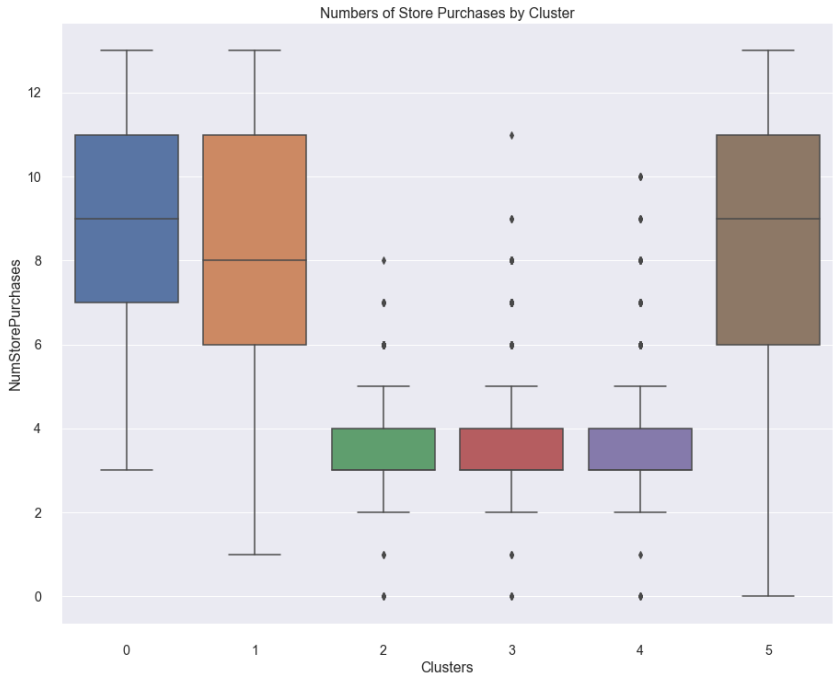
- When normalized, Cluster 5 shops via catalog more than the other clusters. *Make sure to have a polished catalog.*
- Clusters 1 and 0 follow

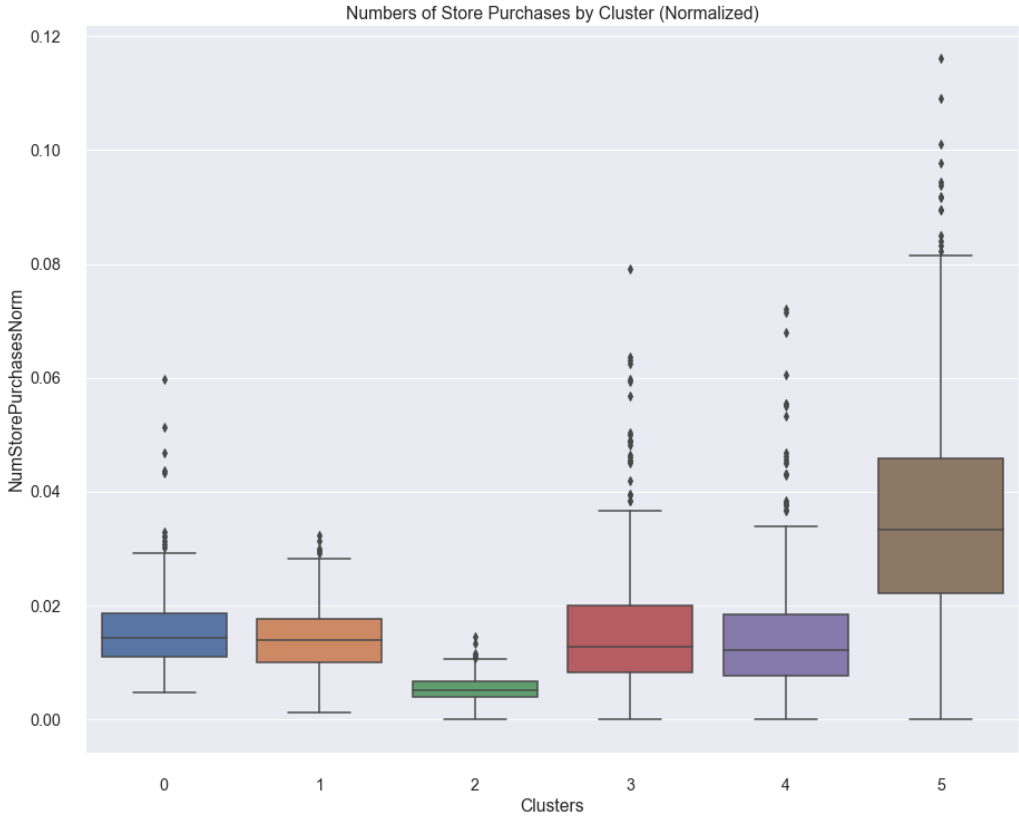




PURCHASE CHANNEL (STORE)

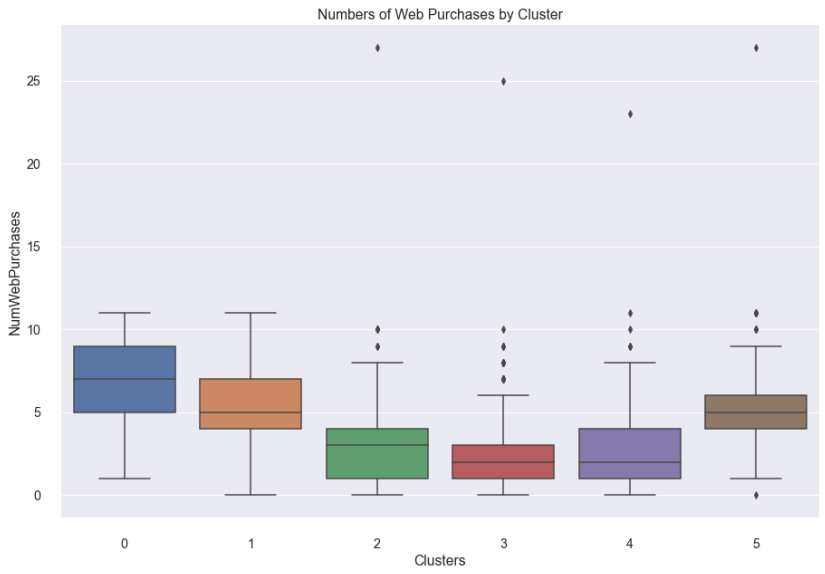
- Cluster 5 has a strong propensity to shopping in store, buying more than any other cluster and about as much as they buy using the catalog and web combined. *Make these customers' in-store experience customers pleasant.*

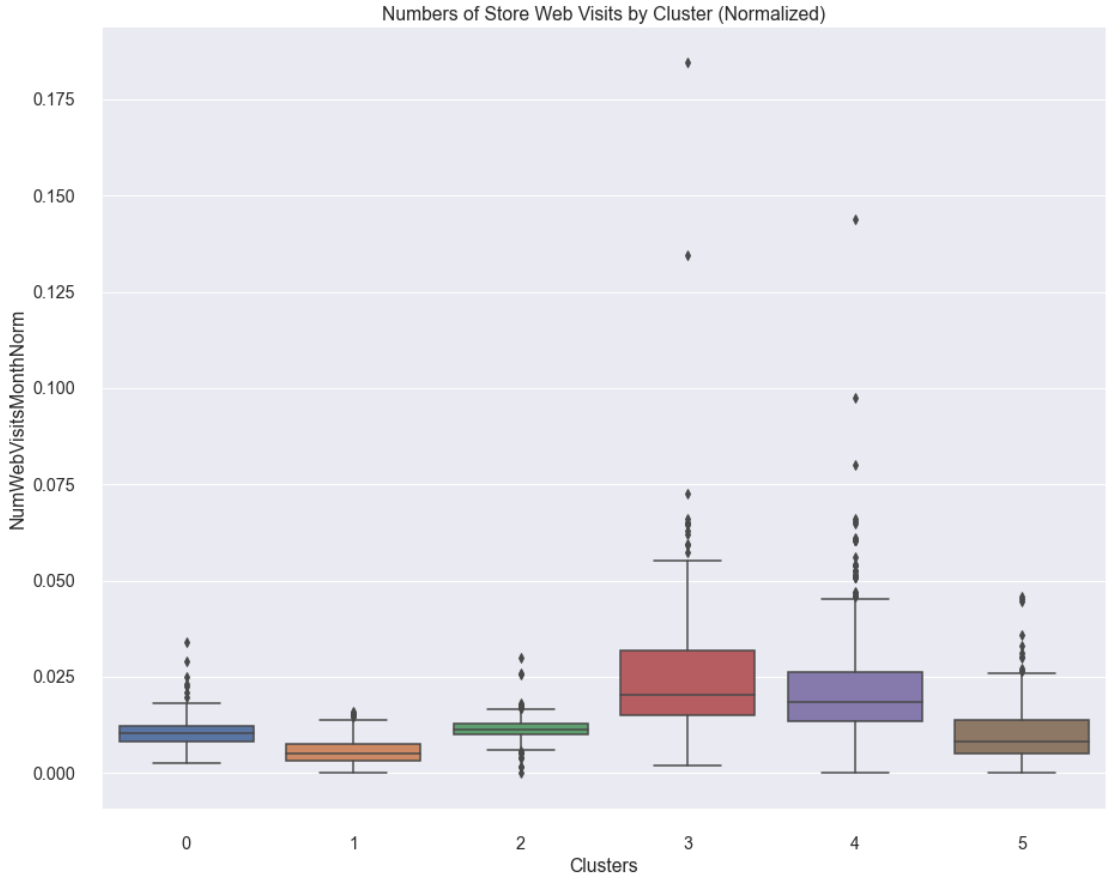




WEB VISITS

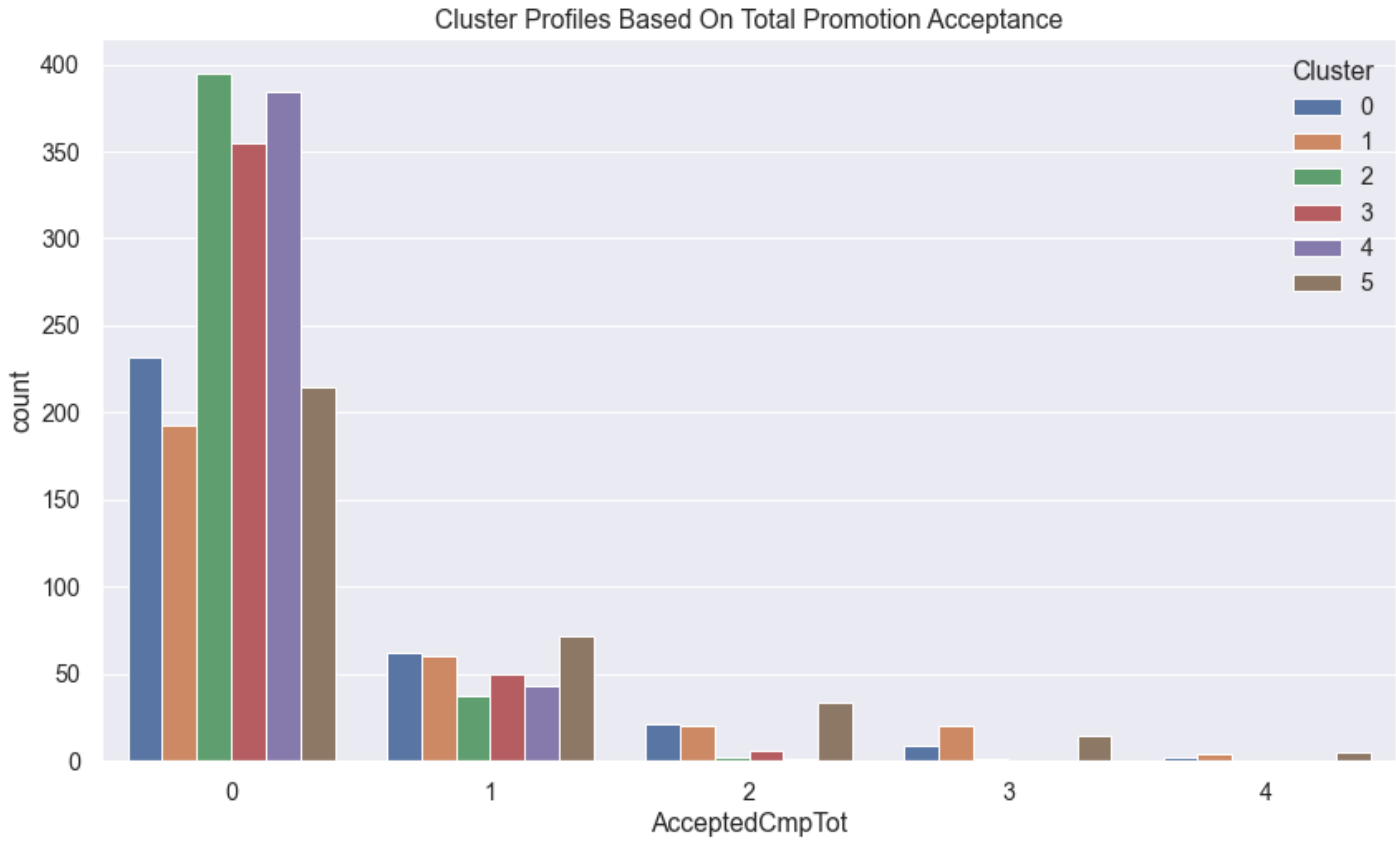
- Clusters 2-4 are the most active on the web when normalized by length as customers. Historically, Cluster 1 has also been active. *Cluster 5 probably use the web with the intention of making a purchase rather than looking for deals. Do not worry about targeting online deals at them.*





CAMPAIGN ACCEPTANCE

- Campaigns were moderately successful
- No customer accepted all 5 campaigns

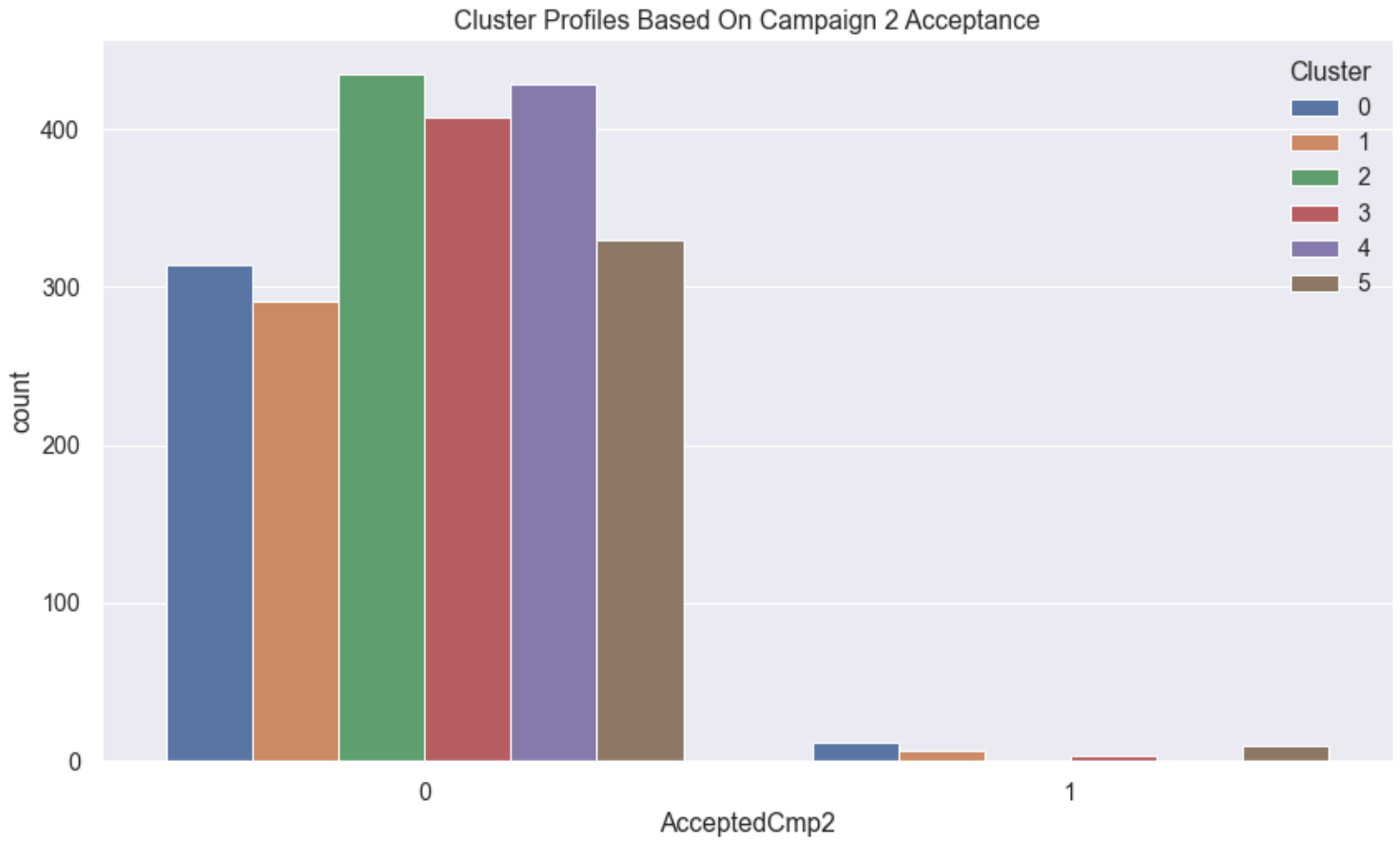


CAMPAIGN 1 ACCEPTANCE

- Campaign 1 was most successful with Clusters 1 and 5
- *Try to replicate this campaign if want to attract those customers*

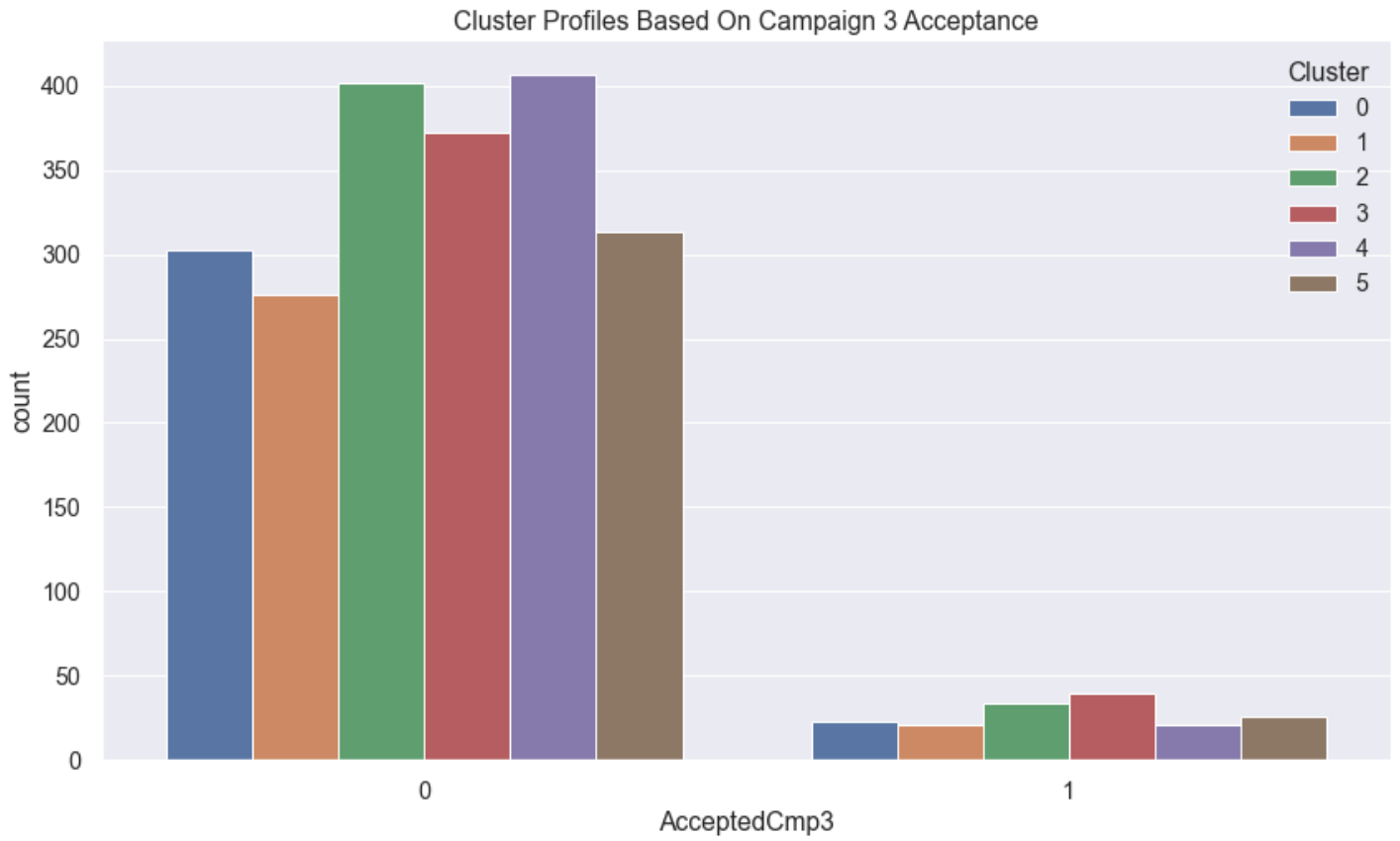
CAMPAIGN 2 ACCEPTANCE

- Campaign 2 had a very low response rate
- *Learn from this campaign in order not to repeat it*



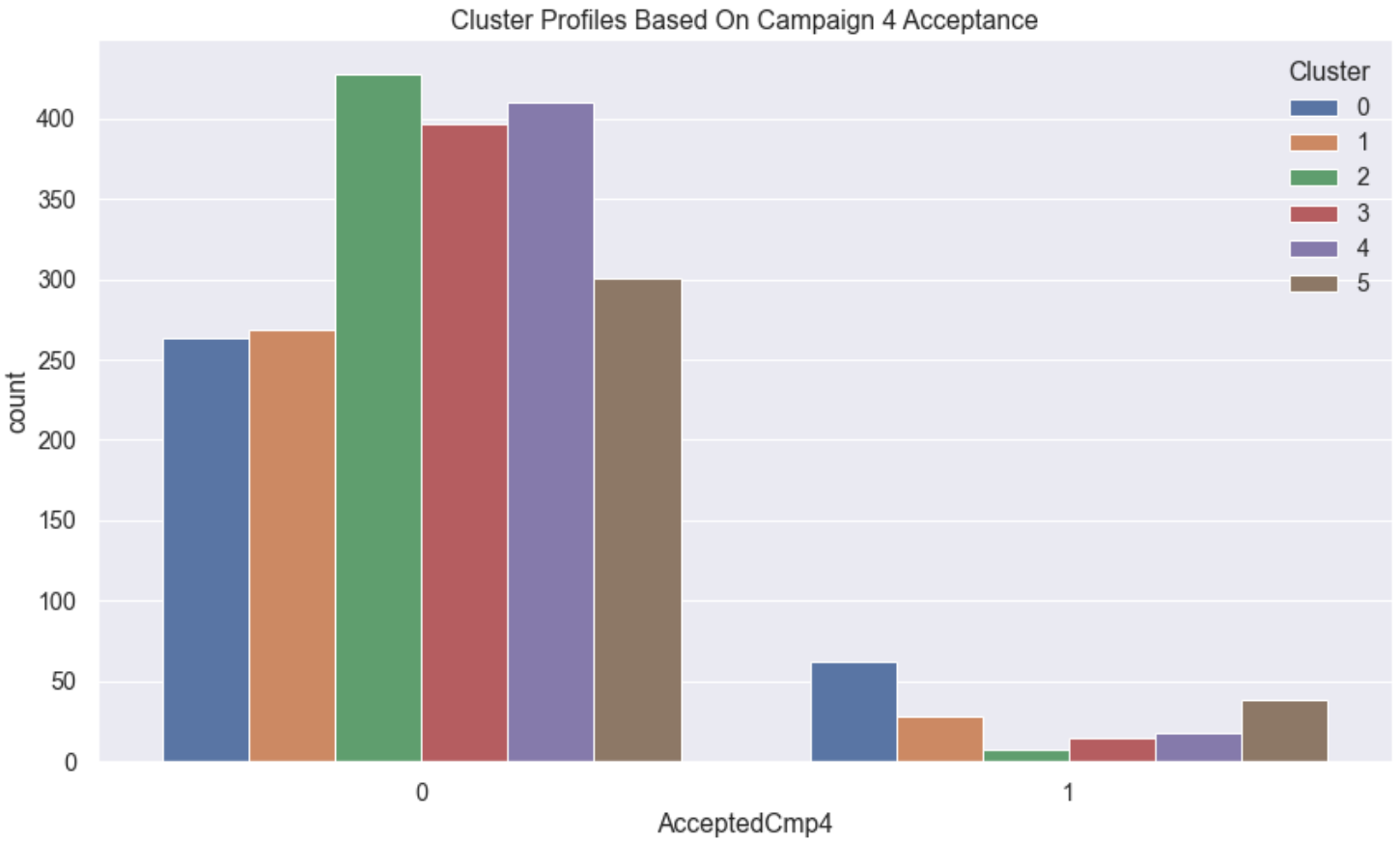
CAMPAIGN 3 ACCEPTANCE

- Campaign 3 was particularly successful with Clusters 2 and 3
- *Target campaigns like this at these customers*



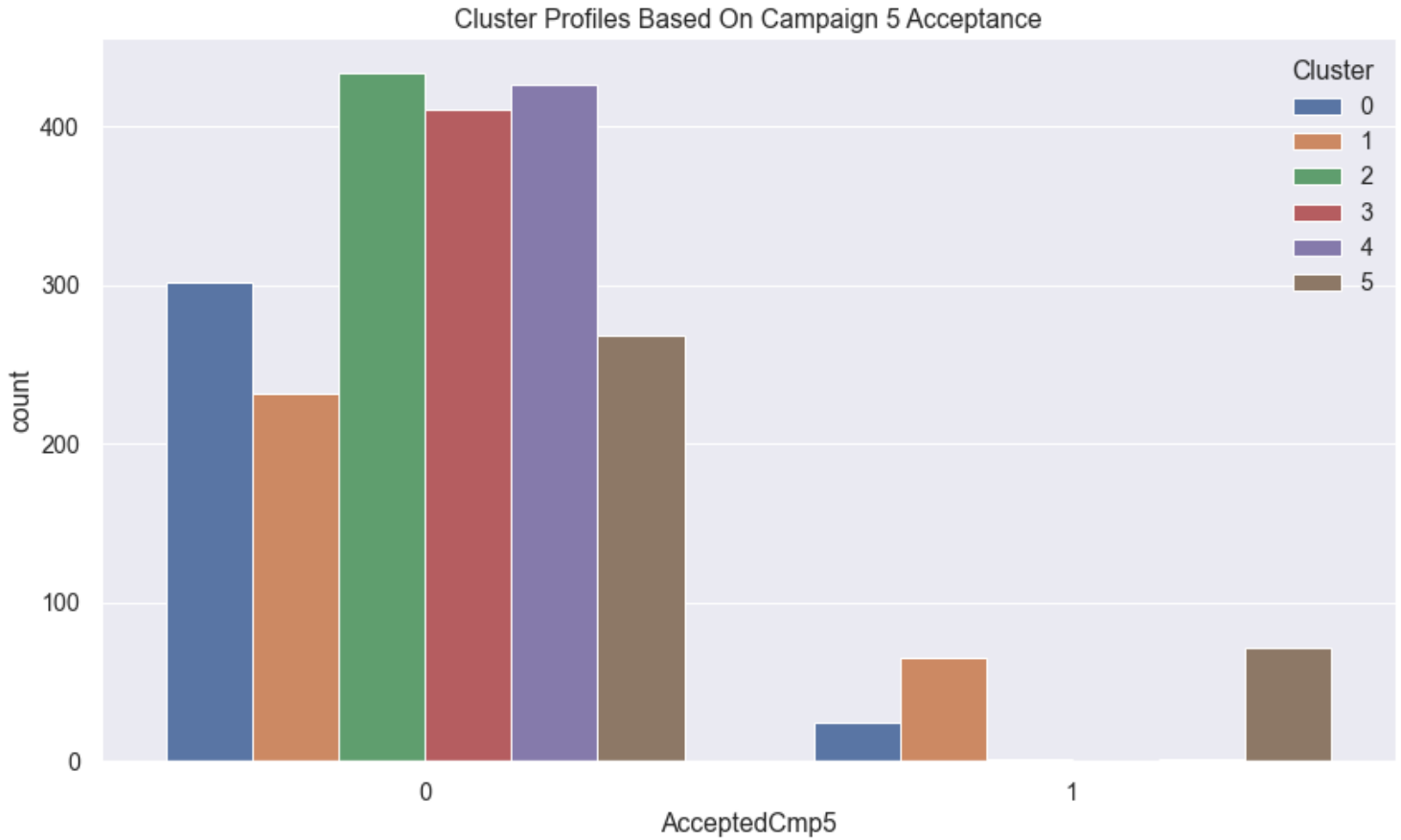
CAMPAIGN 4 ACCEPTANCE

- Campaign 4 was particularly successful with Clusters 0, which is a high-potential cluster
- *Replicate the best parts of this campaign to attract those clusters*



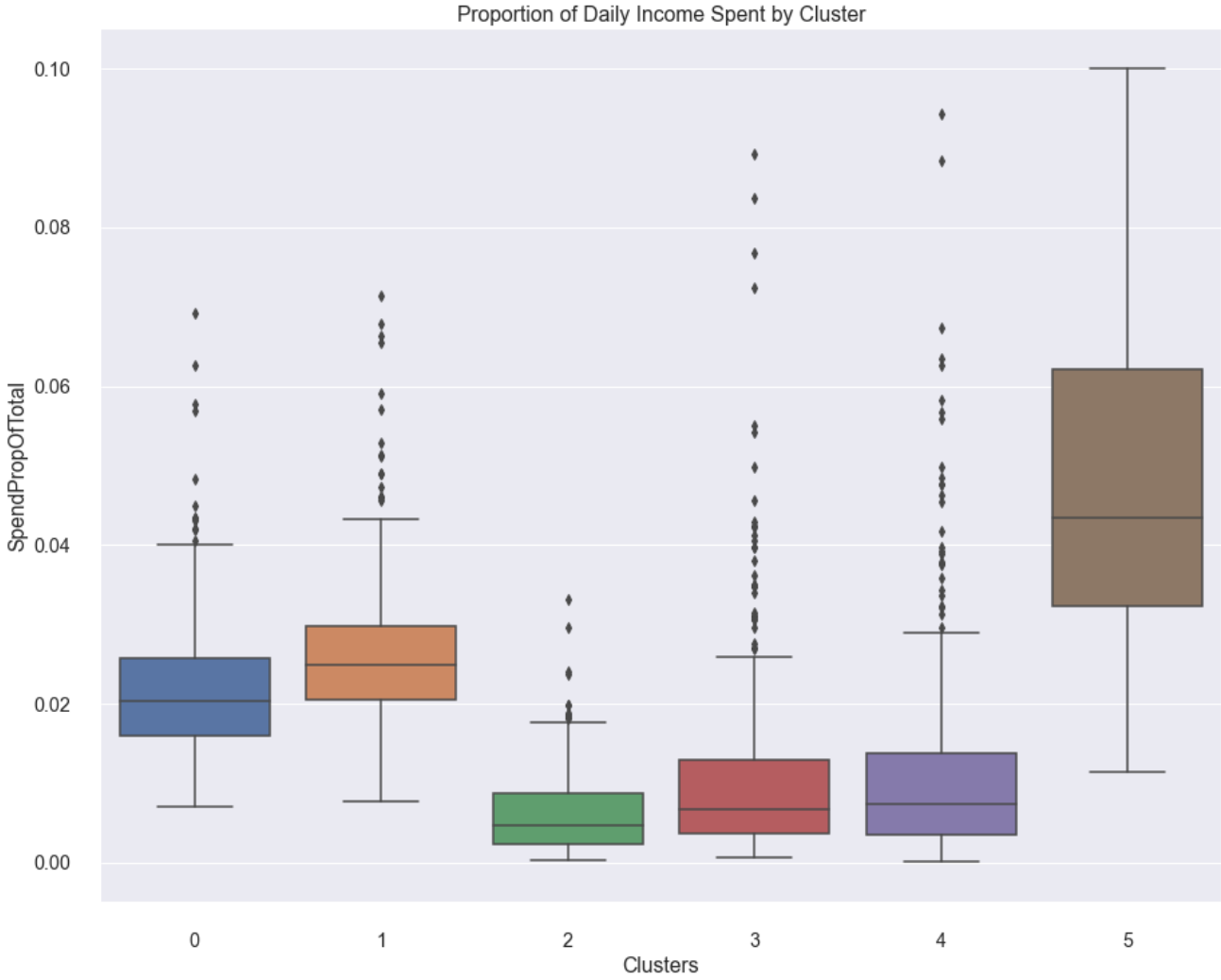
CAMPAIGN 5 ACCEPTANCE

- Like Campaign 1, Campaign 5 was particularly successful with Clusters 1 and 5
- The store had a good idea with these campaigns
- *Improve campaigns like this one to attract even more such customers*



SPENDING AS PROPORTION OF DAILY INCOME

- Cluster 5 is still in the lead followed by Clusters 1 and 0. Cluster 5 spends over 4% of their income at the store while Clusters 0 and 1 spend just over 2%.
- *Look for ways to attract customers away from the alternatives*



SUMMARY OF FINDINGS

- If run deals, target these at Clusters 2-4
- Concentrate on making/keeping the web page appealing for Clusters 0, 1, and 5
- Make sure to have a polished catalog, especially for Cluster 5 customers, who could bring in the most revenue in the future
- Make customers' in-store experience customers pleasant as high-spending clusters seem to value it
- Learn from Campaign 2 in order to not repeat it
- Concentrate on running and improving campaigns like Campaign1 and 5, as these attract the high-spending customers
- Concentrate on identifying and attracting customers like Cluster 5 for highest future expected revenue
- Look for ways to attract customers away from the alternatives as customers spend a moderate proportion of their daily income at the store

FUTURE WORK (PART 1)

- It would be helpful to have more information about the data set and the store in order to answer the following questions:
  - What do we know about the rationale behind each campaign? What distinguishes the campaigns?
  - What more can we learn about our customers? Specifically, are there factors that differentiate Cluster 5 that are not in the data? (Surveys)
  - What else can be learned about the way the store is making customers in-store and online shopping experience pleasant? Are there way to improve?
  - Can the store learn to do profitable business with Cluster 2-4?

FUTURE WORK (PART II)

- Get more granular data on store purchases
- Get data on campaign costs
- Get more data on profitability rather than just revenue, as profitability is the key objective
- Do A/B testing to judge the effectiveness of recommendations
- Are Cluster 5 customers just a more recent group of Cluster 1 customers? If so, can we reengage Cluster 1 and find ways to keep Cluster 5 continually engaged?

NOTES

- Please see the full repo at https://github.com/dapopov-st/Customer-Segmentation for further details
- Results of preliminary experiments are in the Preliminary_work_and_experiments directory
- Special thanks to Zeyu Zhang, David Kressley, Kyle Gallatin, and Philippe Heitzmann for crucial feedback and support