

The UT Dallas Morris Hite Center Marketing Analytics & Customer Insights Challenge

Rising Comets

Sneha Liza George

Vishakha Nangia

Seung Ah Shin

Linh Nhat Nguyen

Executive Summary

Problem statement:

- The client's sale has hardly experienced much growth in the past 12 months.

Outcome:

- The sales of the client's brand has increased by 20% as compared to last year

Insights:

- Define strategies to acquire more customers
- Formulate strategies to get a larger share of wallet and shift the spend from competitors to client

Findings:

- Potential areas of growth has been discovered
- Understanding the differences in customer profile of clients versus that of consumers shopping at their competitor

The transformation of dataset into insights leveraged R and Power Bi

Exploration of dataset

```
graph TD; A[Exploration of dataset] --> B[Formation of Hypothesis]; B --> C[Data Cleaning]; C --> D[Data Modeling]; D --> E[Data Visualization];
```

Formation of Hypothesis

Data Cleaning

Data Modeling

Data Visualization

Detailed description of the journey from data to information

Exploration of dataset

- Knowledge of data size
- Identification of dimensions and metrics by evaluating the dataset
- Categorization of variables in the dataset into customer demographics, social status and digital activity

Formation of hypothesis

- Definition of hypothesis with the selected dimensions and metrics

Data cleaning

- Import data into R
- Summarization of data
- Checking the presence null values and NAs and treating them wherever required
- Checking the presence of outliers and removing them wherever required

Contd...

Data Modelling in R

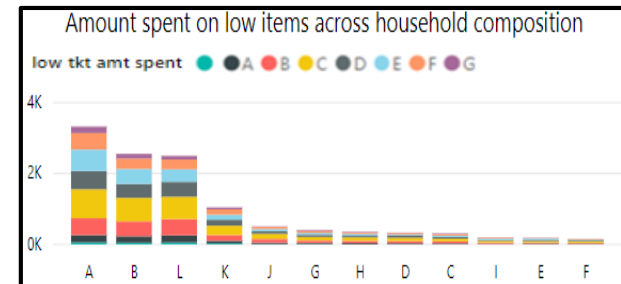
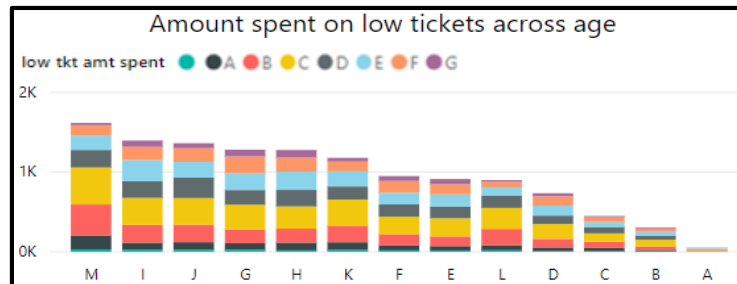
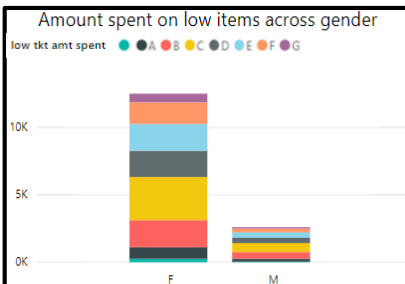
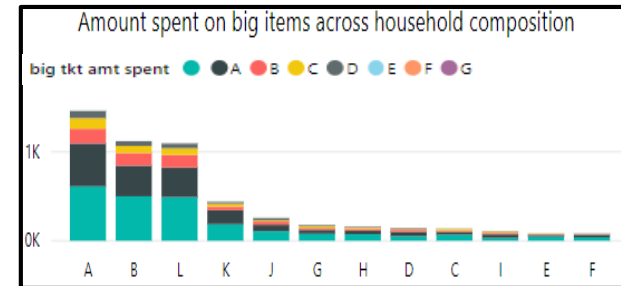
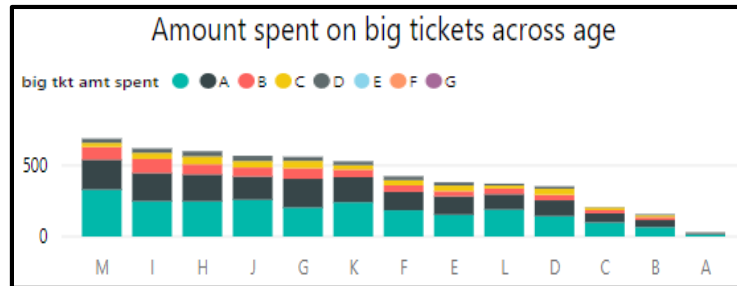
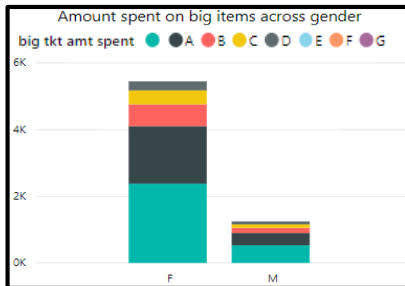
- Subset the cleaned data with the required dependent and independent variables
- Summarization of the subset data
- Determination of variables types(interval, nominal or ordinal)
- Imputation of missing data wherever required
- Choice of Chi Square test as the dependent and independent variables are nominal
- Leveraging p value to find significance value

Data visualization

- Import data into Power BI
- Usage of visuals to portray the relationship between dependent and independent variables
- Generation of findings and insights with the help of visuals

Client needs to focus on young female generation

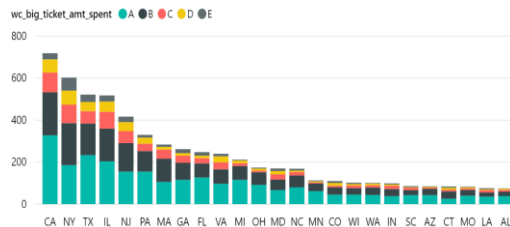
- 82% of the people who bought items are women out of which 80% bought low items where the majority spent between \$100 - \$199 and 35% bought big items where the majority spent less than \$100 yearly
- 13% of the people who bought tickets are 76+ years old out of which 96% bought low tickets where the majority spent between \$100 - \$199 and 43% bought big tickets where the majority spent less than \$100
- 65% of the people who bought items are above 45 years old
- 28% of the people who bought items are women married and having children out of which 27% bought low items where the majority spent between \$100 - \$199 and 12% bought big items where the majority spent less than \$100
- 80% of the people who bought items are married or householder women with and without children
- Competitors have generally younger generation (45-) shopping more whereas clients have older customers (76+) in the past 3 months



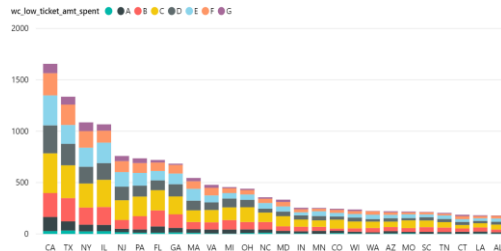
Client needs to establish its brand among minor ethnicities like Chinese and Vietnamese

- 35% of the people who bought items are from CA, TX, NY and IL out of which 80% bought low items where the majority spent between \$100 - \$199 and 45% bought big items where the majority spent less than \$100
- 29% of the people who bought tickets are English of which 98% bought low tickets where the majority spent between \$100 - \$199 and 40% bought big tickets where the majority spent less than \$100
- 90% of the people who bought items speak English out of which 88% bought low items where the majority spent between \$100 - \$199 and 38% bought big items where the majority spent less than \$100
- Apart from English, items are bought the most by people who are German, Hispanic, Irish, Scottish, Italian and French and by those who speak Spanish, Mandarin, French, Italian and Vietnamese
- Competitors tend to be popular among minor ethnicities like Chinese and Vietnamese

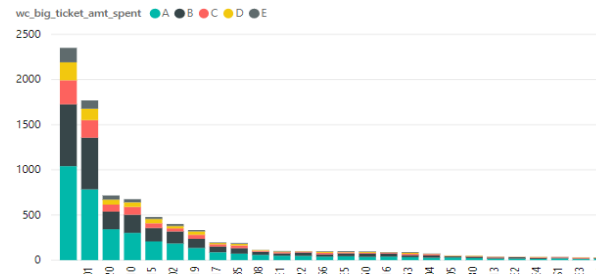
Amount spent on big items across top 25 states



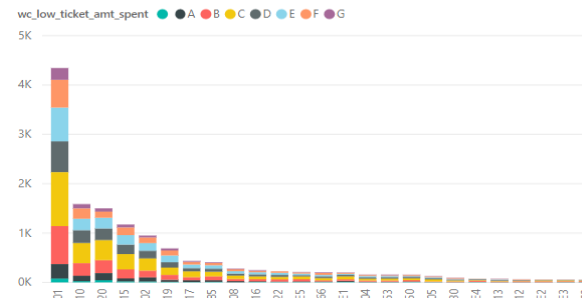
Amount spent on low items across top 25 states



Amount spent on big items across ethnicity

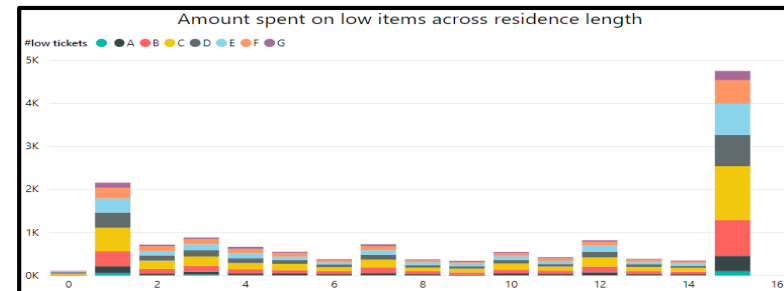
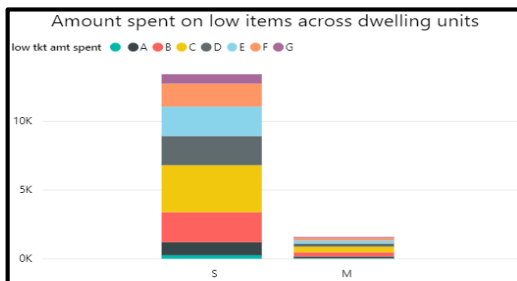
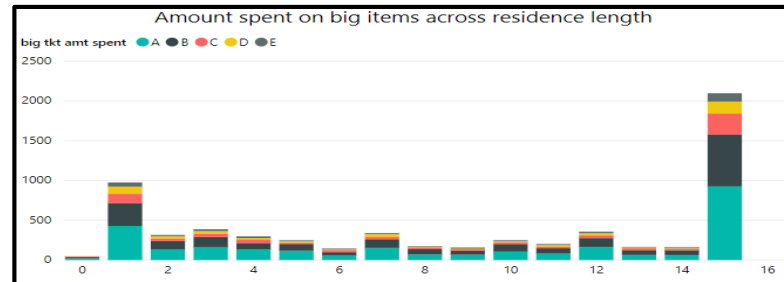
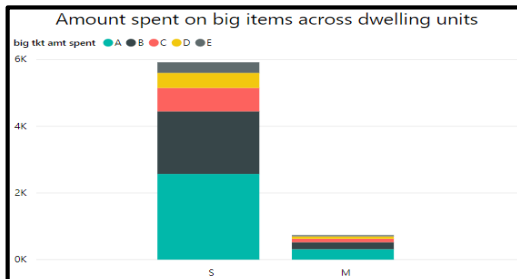


Amount spent on low items across ethnicity



Potential for customer acquisition is present among people living in single family dwelling units who are in their initial years of stay

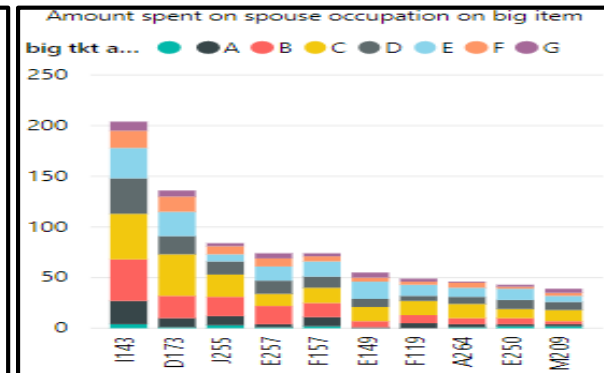
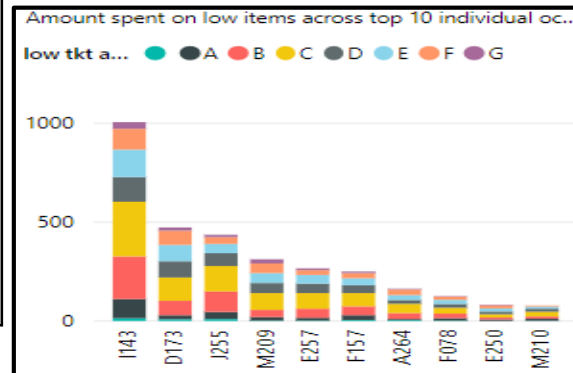
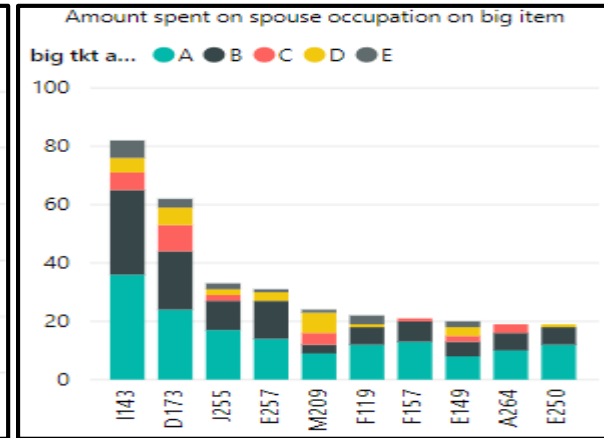
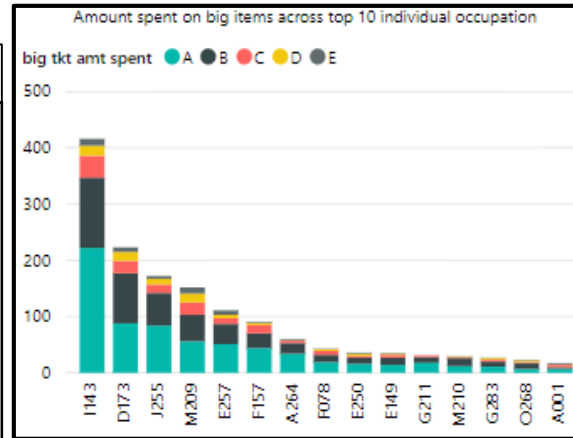
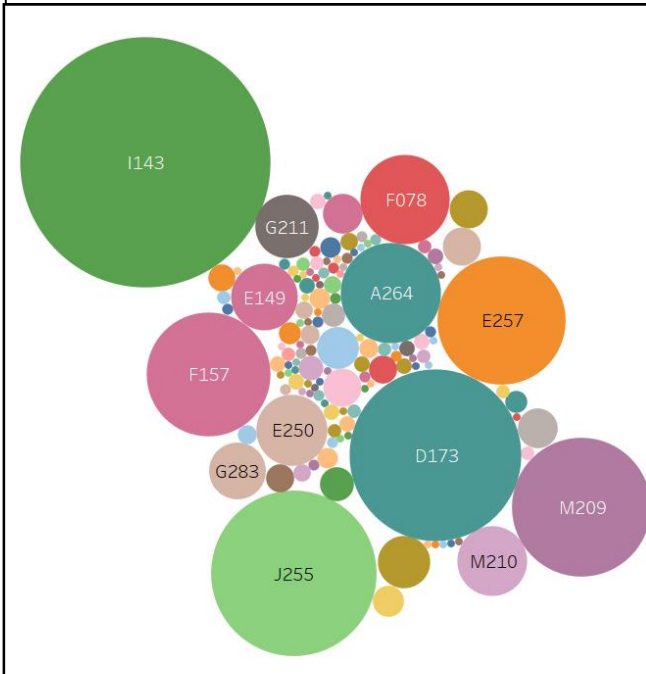
- Women clothing was bought the most by the people living in single family dwelling units for more than 14 years
- 89% of the people who bought tickets live in single-family dwelling units out of which 50% bought big tickets where the majority spent less than \$100 and 38% bought low tickets where majority spent between \$100 - \$199 for the past 12 months
- People living in the same address for more than 14 years bought 33% of the tickets out of which 19% bought big tickets where the majority spent less than \$100 and 14% bought low tickets where majority spent between \$100 - \$199 for the past 12 months
- The client is doing better than competitors for the last 3 months or in last 30 days



Client's brand needs to promoted across professionals

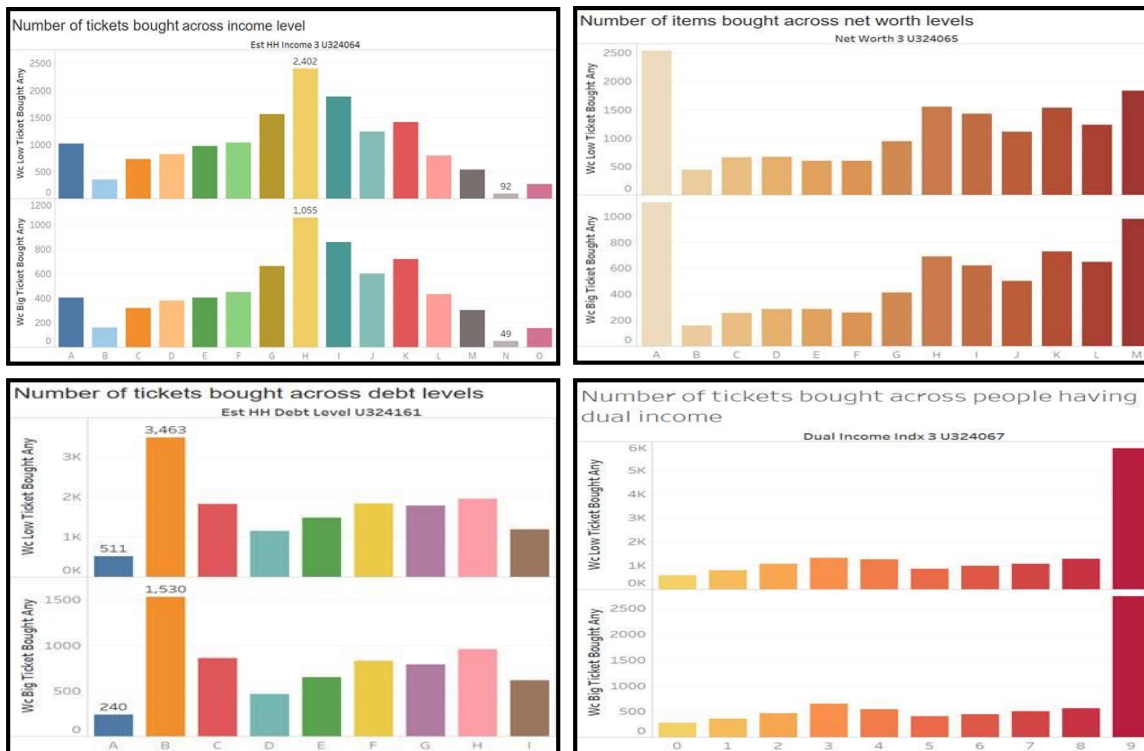
- Around 40% of the shoppers are managers and homemakers and 28% of the spouses follow similar trend
- The client is doing better than competitors for last 3 months or in last 30 days.

Number of items bought across occupation



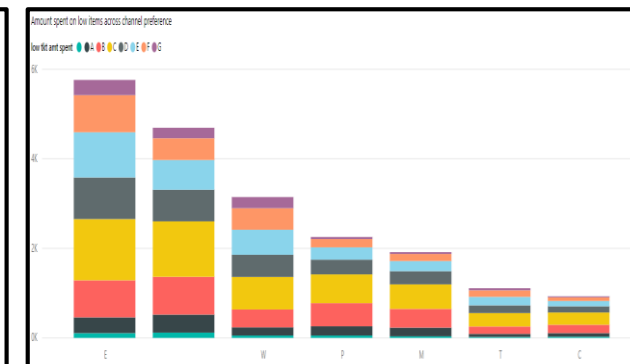
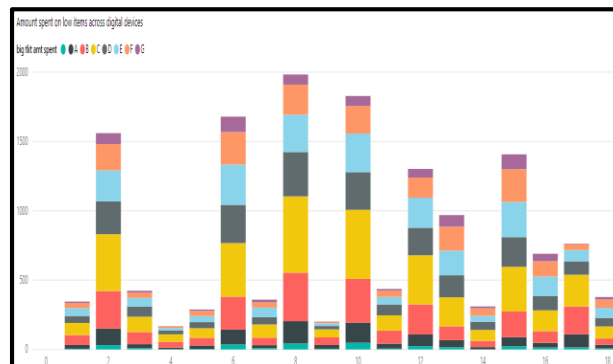
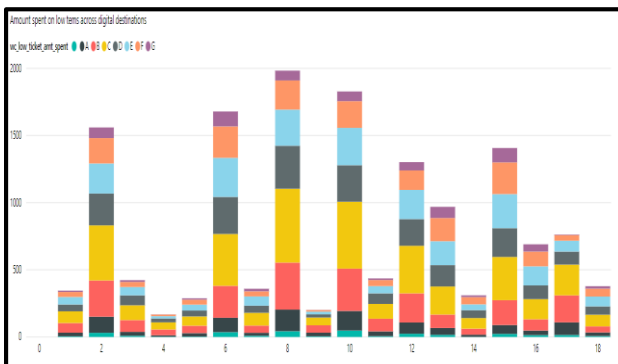
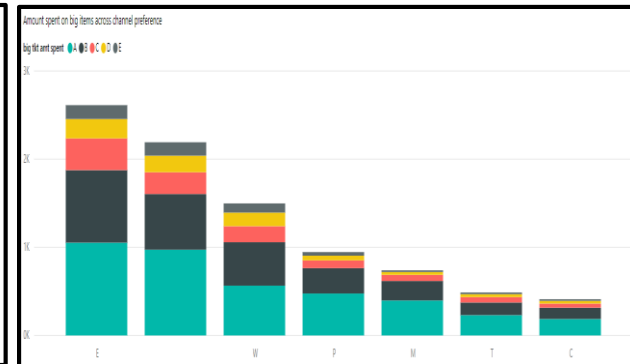
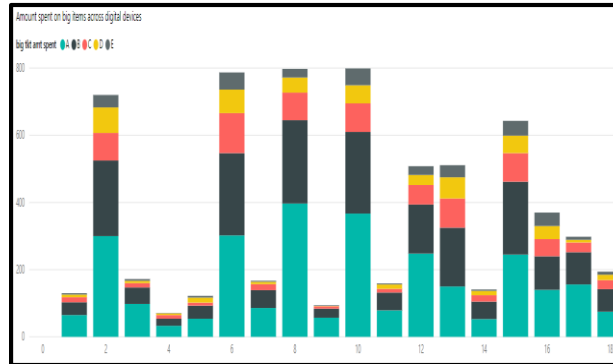
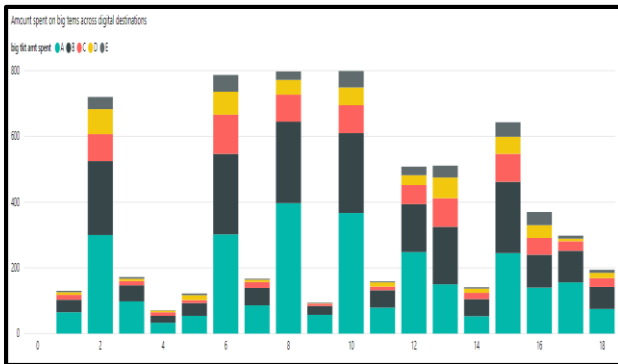
Promoting client's brand across affluent class may invite more sales of big items

- People earning \$75,000-\$99,999 buy the maximum number of tickets and to about 16% of the total big and low tickets bought.
- The number of tickets bought are maximum for the families having more than one source of income.
- The people having low annual financial debt (\$1 - \$1,499) buy the maximum number of tickets and contribute to 23% of the total tickets bought.
- The households with net worth of \$0 or less buy around 17% of the tickets.



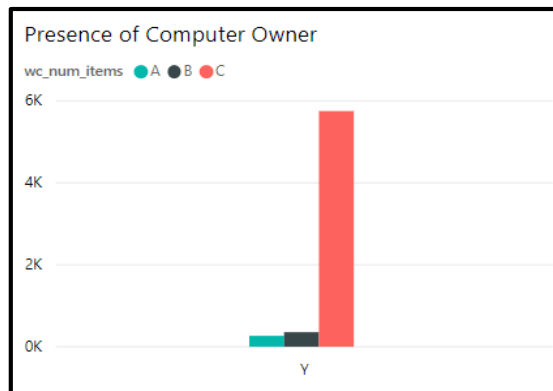
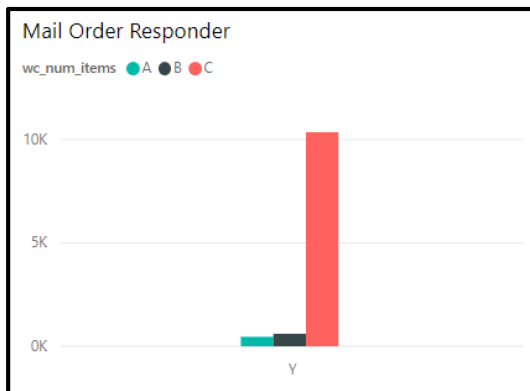
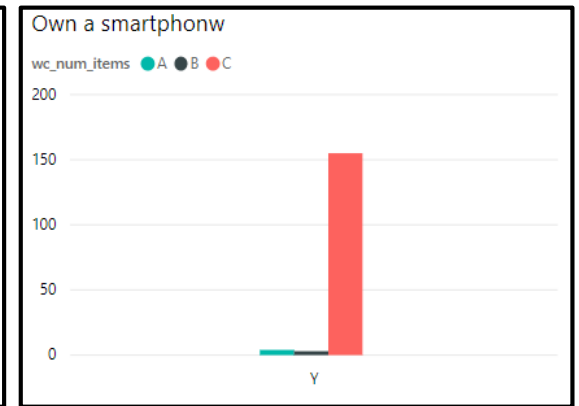
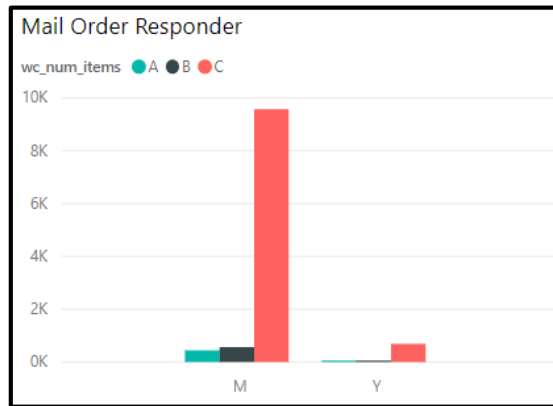
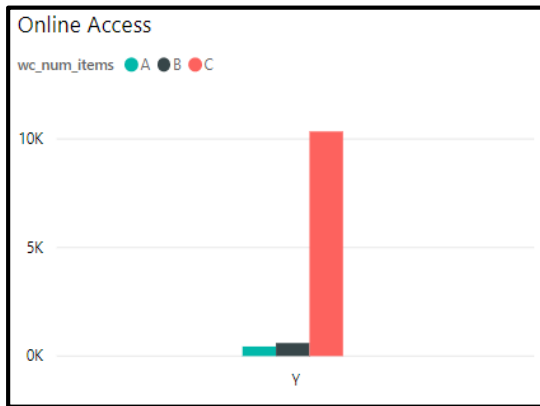
The client's brand needs to be promoted through digital advertising

- **Digital Devices:** Tickets are bought the most by Evolving - Aspiring Early Adopters and Highly Evolved – Aspiring Rural Early Adopters
- **Digital Destination:** Tickets are bought the most by retired or retiring executives, corporate homebodies and well-read executives
- 38% of people who bought items prefer e-mail whereas 21% people prefer web as channel of communication



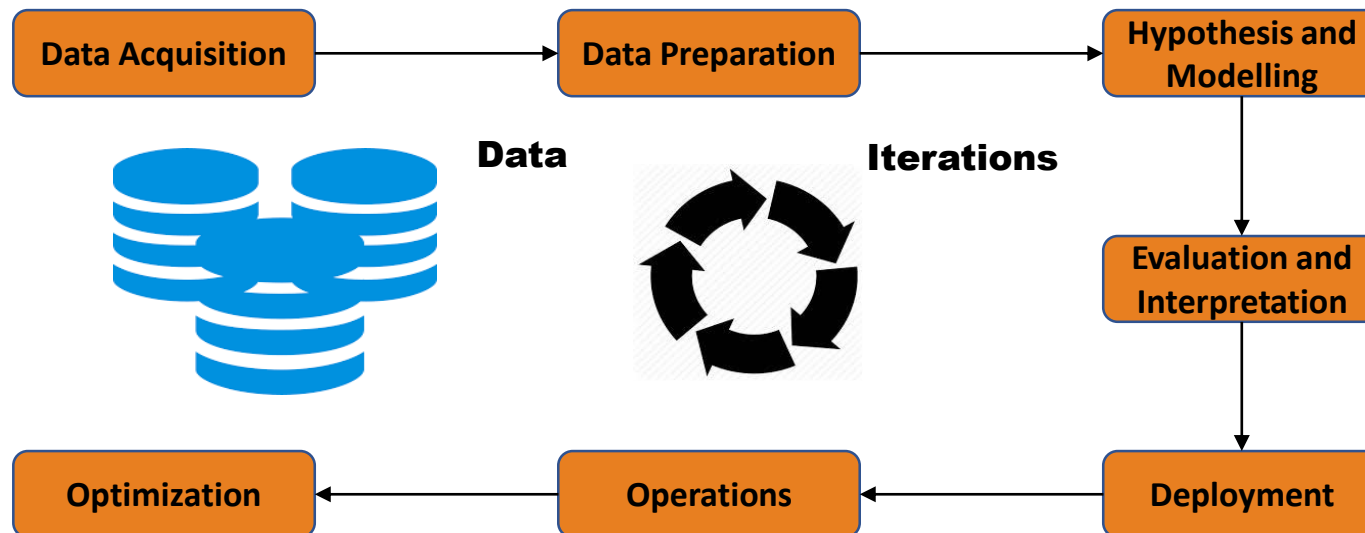
Digital advertising can be a source of potential growth

- 90% people who bought three or more items are digitally active
- 52% people who bought three or more items respond to mail order where the majority spent below \$100 on big ticket items and between \$100 - \$199 on low ticket items



Model Operationalization and Optimization

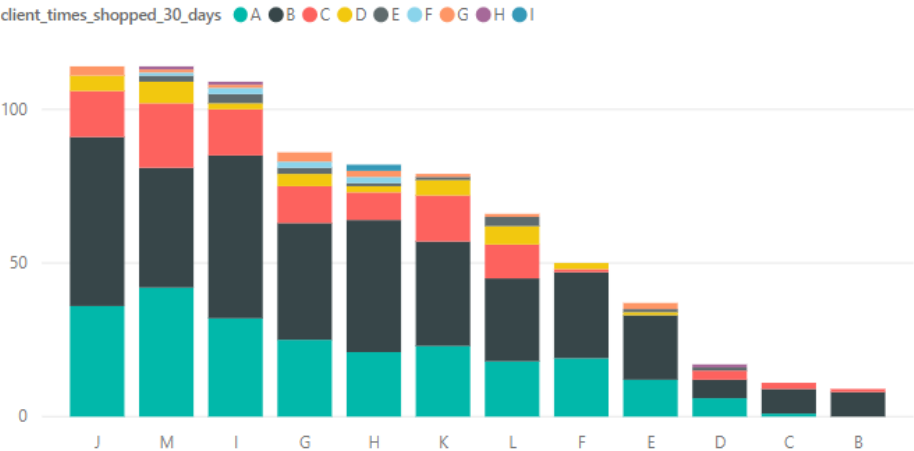
- Reduce the expected generalization error in the model to minimize the risk as it will be the topmost priority for the client in case the customer base size needs to be increased
- Fine-tune decisions, offers, or treatments to better align with the business priorities.
- Proper documentation of the decision process and domain expertise so that the decision logic can be shared, improved updated ,audited and simulated.
- Adaptability of the model to multiple decision scenarios.



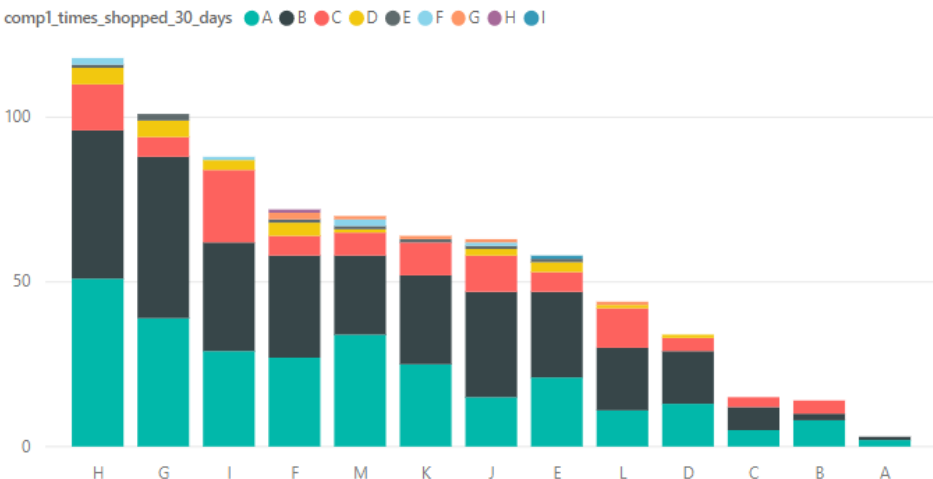
Appendix:

Number of items bought by client vs competitors over the past 30 days across age

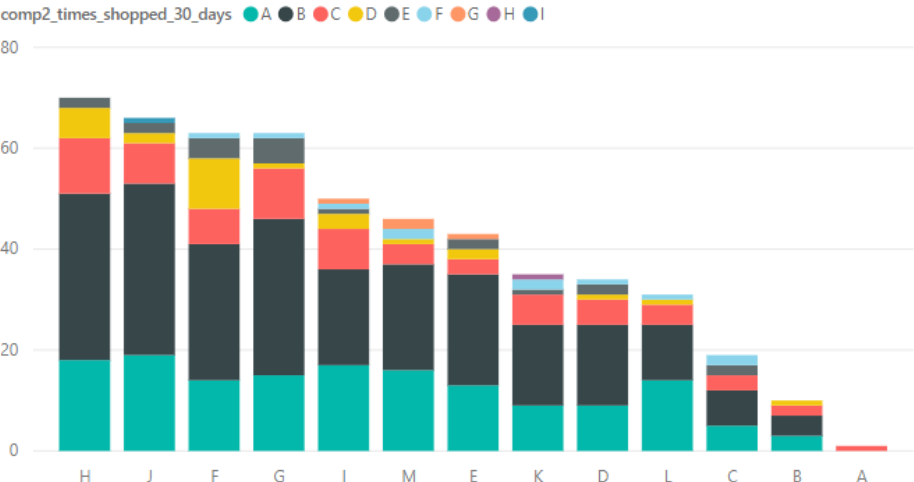
Number of times client bought in the last 30 days across age groups



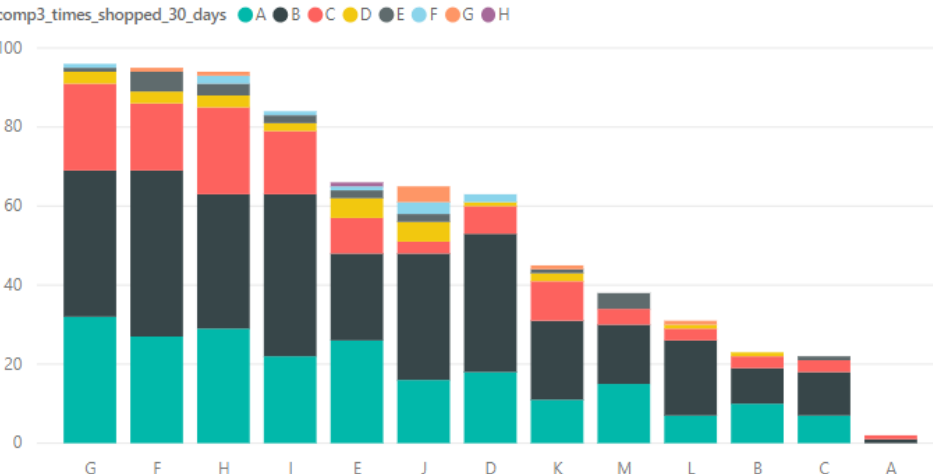
Number of times competitor 1 bought in the last 30 days across age groups



Number of times competitor 2 bought in the last 30 days across age groups

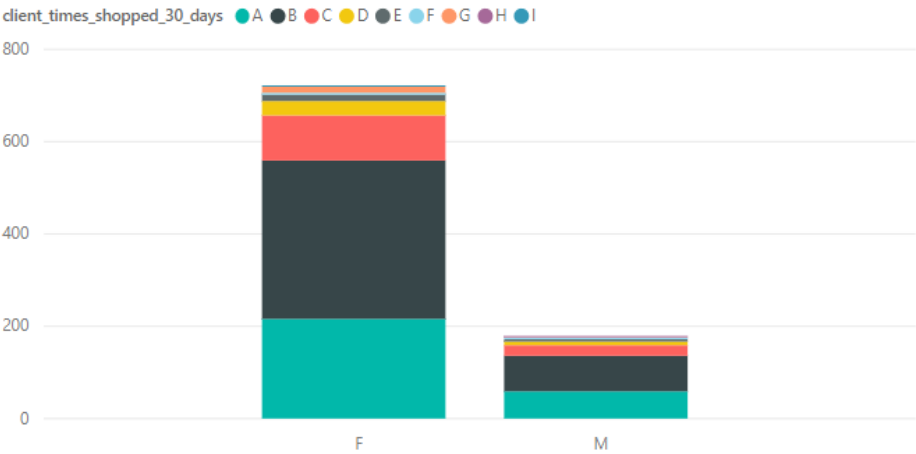


Number of times competitor 3 bought in the last 30 days across age groups

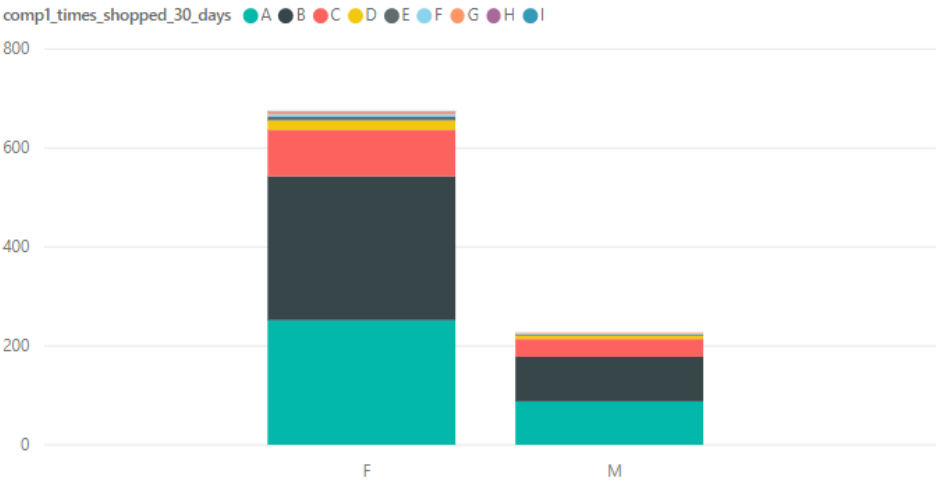


Number of items bought by client vs competitors over the past 30 days across gender

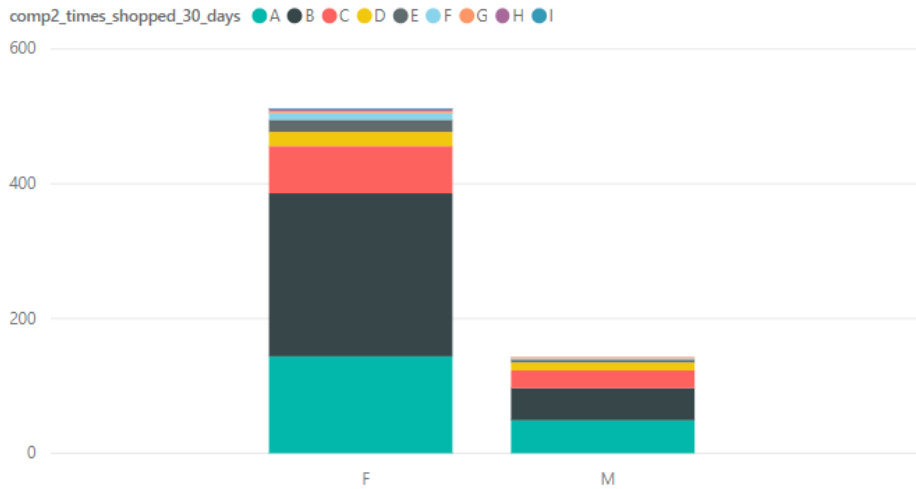
Number of times client bought in the last 30 days across gender groups



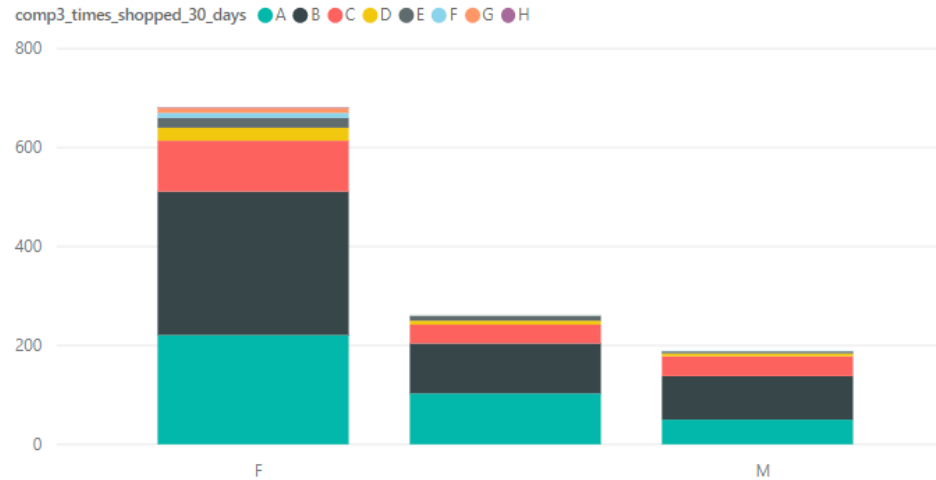
Number of times competitor 1 bought in the last 30 days across gender groups



Number of times competitor 2 bought in the last 30 days across gender groups

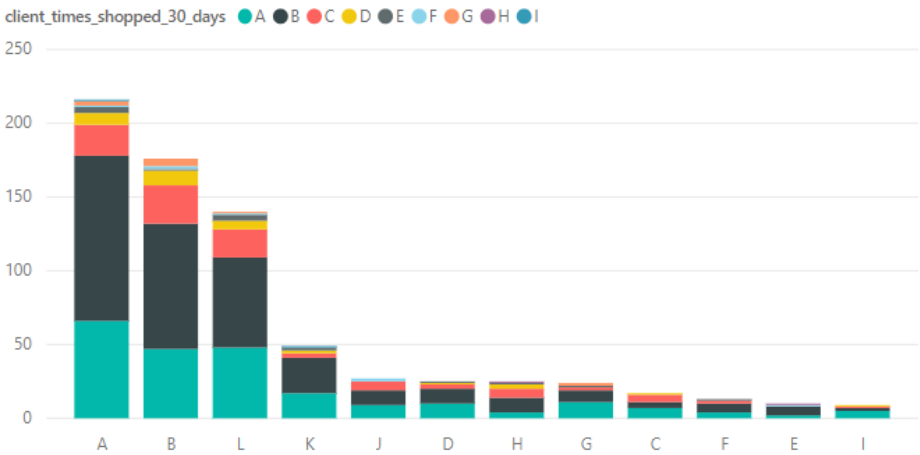


Number of times competitor 3 bought in the last 30 days across gender groups

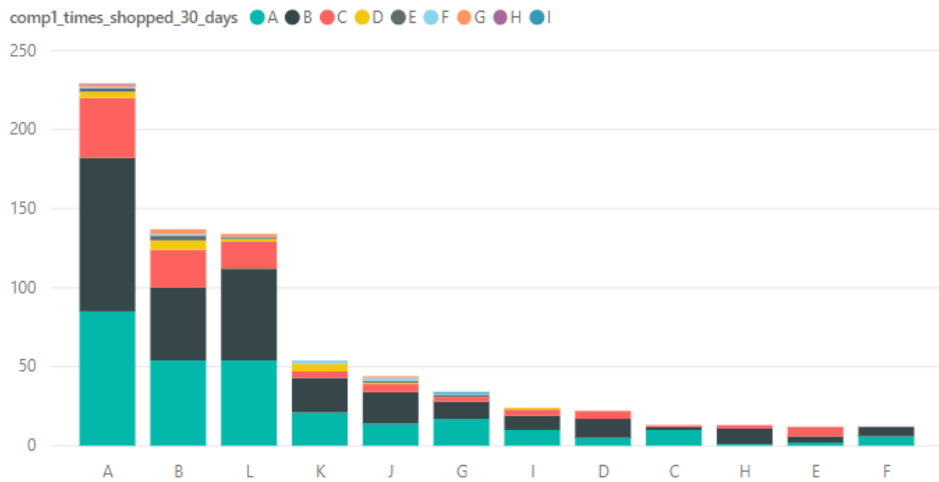


Number of items bought by client vs competitors over the past 30 days across marital status

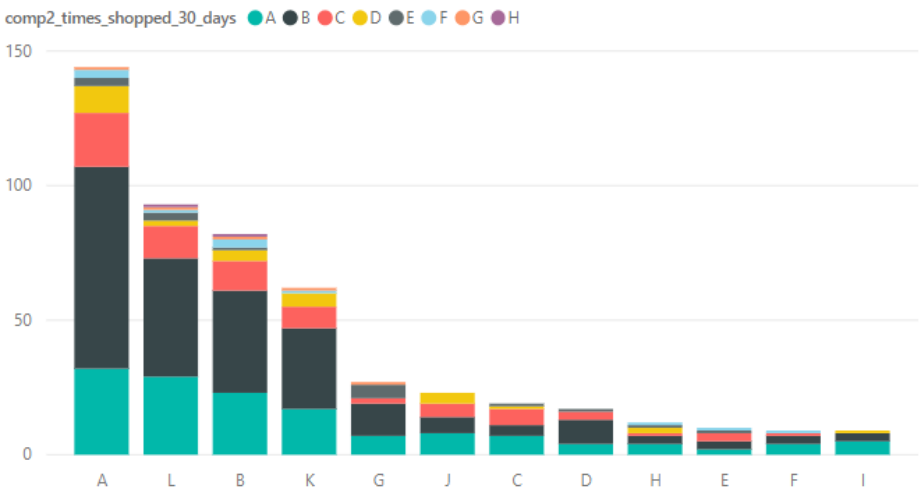
Number of times client bought in the last 30 days across marital status



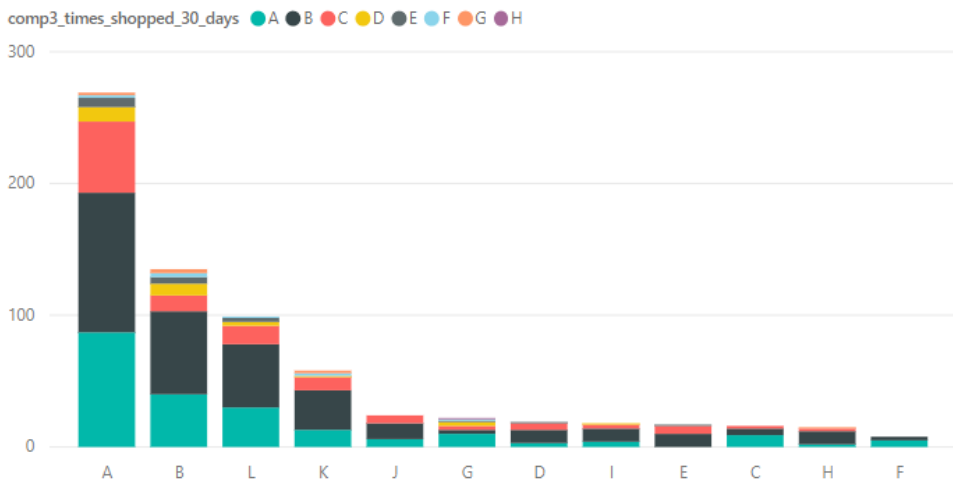
Number of times competitor 1 bought in the last 30 days across marital status



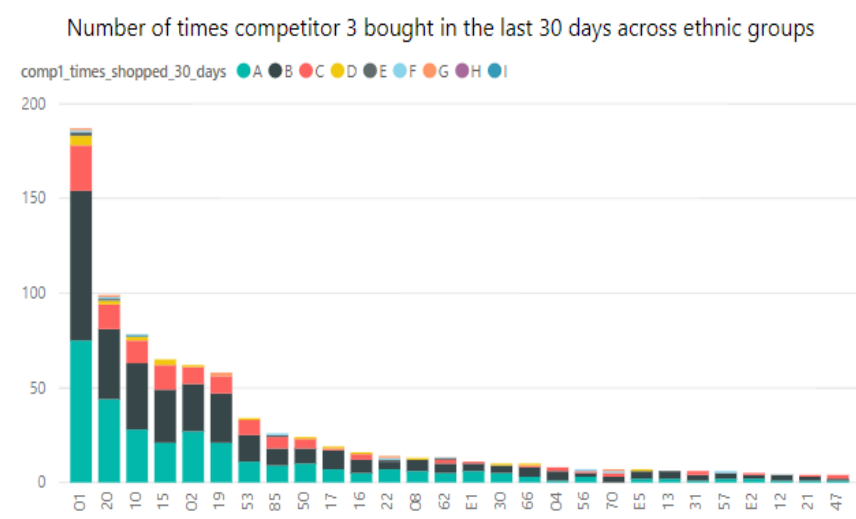
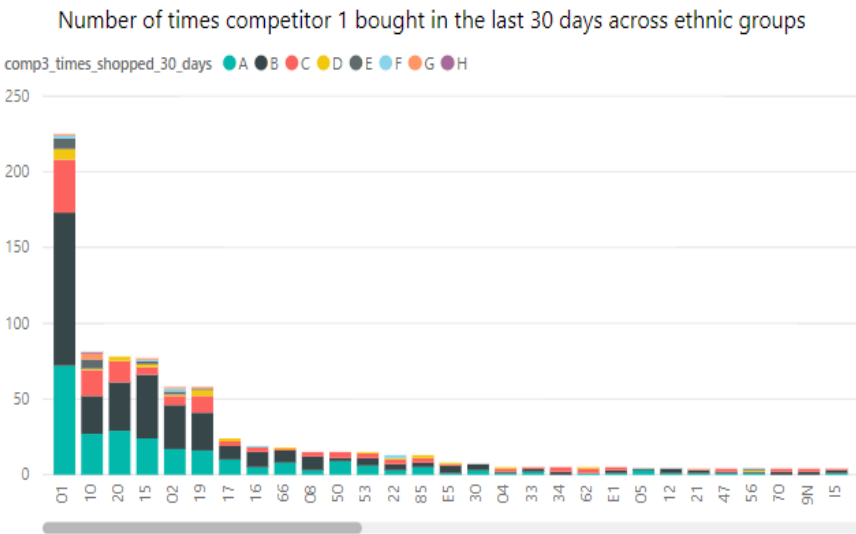
Number of times competitor 2 bought in the last 30 days across marital status



Number of times competitor 3 bought in the last 30 days across marital status



Number of items bought by client vs competitors for past 30 days across the ethnicities



Number of items bought by client vs competitors for past 30 days across the language

