Customer Segmentation for Grocery Stores

ML Masters

Project Importance & Research Questions

01

Customer Characteristics

What types of customers are present in the dataset?

02

Customer Clusters

How many customer clusters are present in the dataset?

03

Product Categories

Are certain product categories associated with higher grocery spend / customer characteristics?

Project Architecture

Database for grocery customer data

Deploy HTML server + DB endpoint

Webpage format + styling

Data manipulation + graphing

ETL

01

Extract

Kaggle .CSV file containing grocery store customer data read into a Pandas data frame

02

Transform

Feature engineering, data cleansing, and initial analysis

03

Load

SQLAlchemy "to_sql" function used to insert grocery store customer data from Pandas data frame to SQLite database

What types of customers are present in the dataset?

Age Group

68% between 40-65

Parental Status

71% have kids

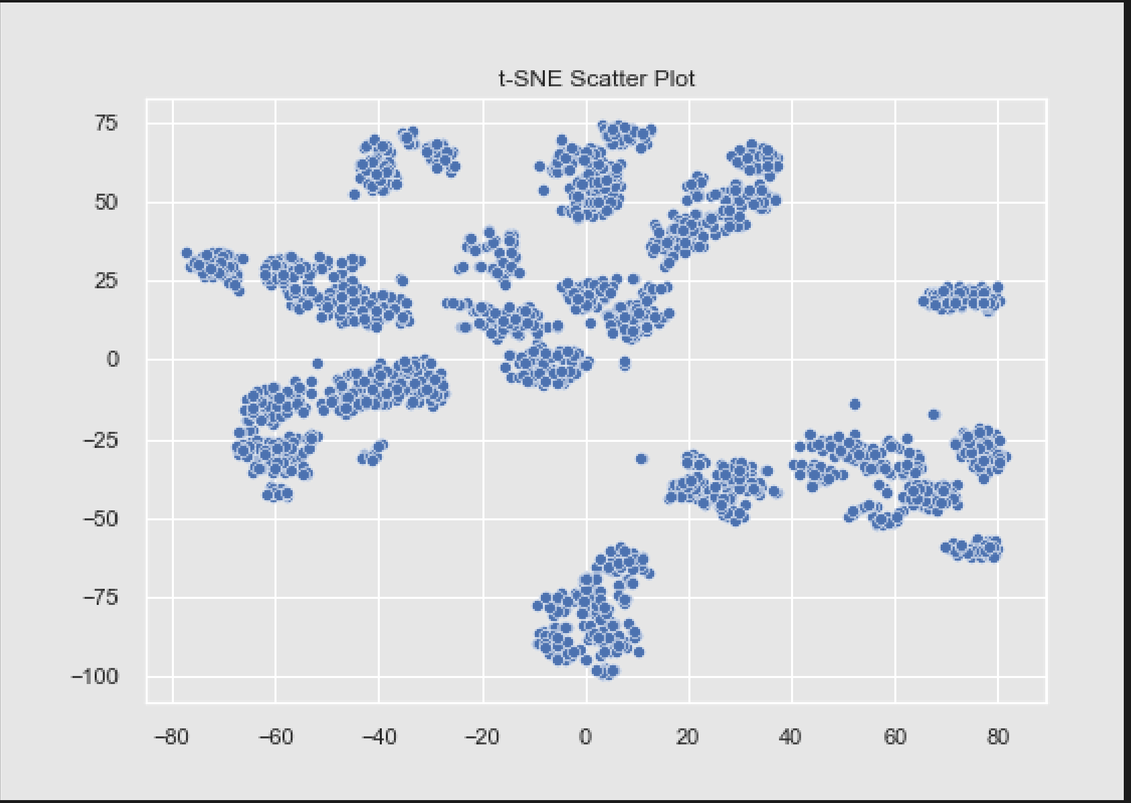
Education Level

50% high school education
47% college education

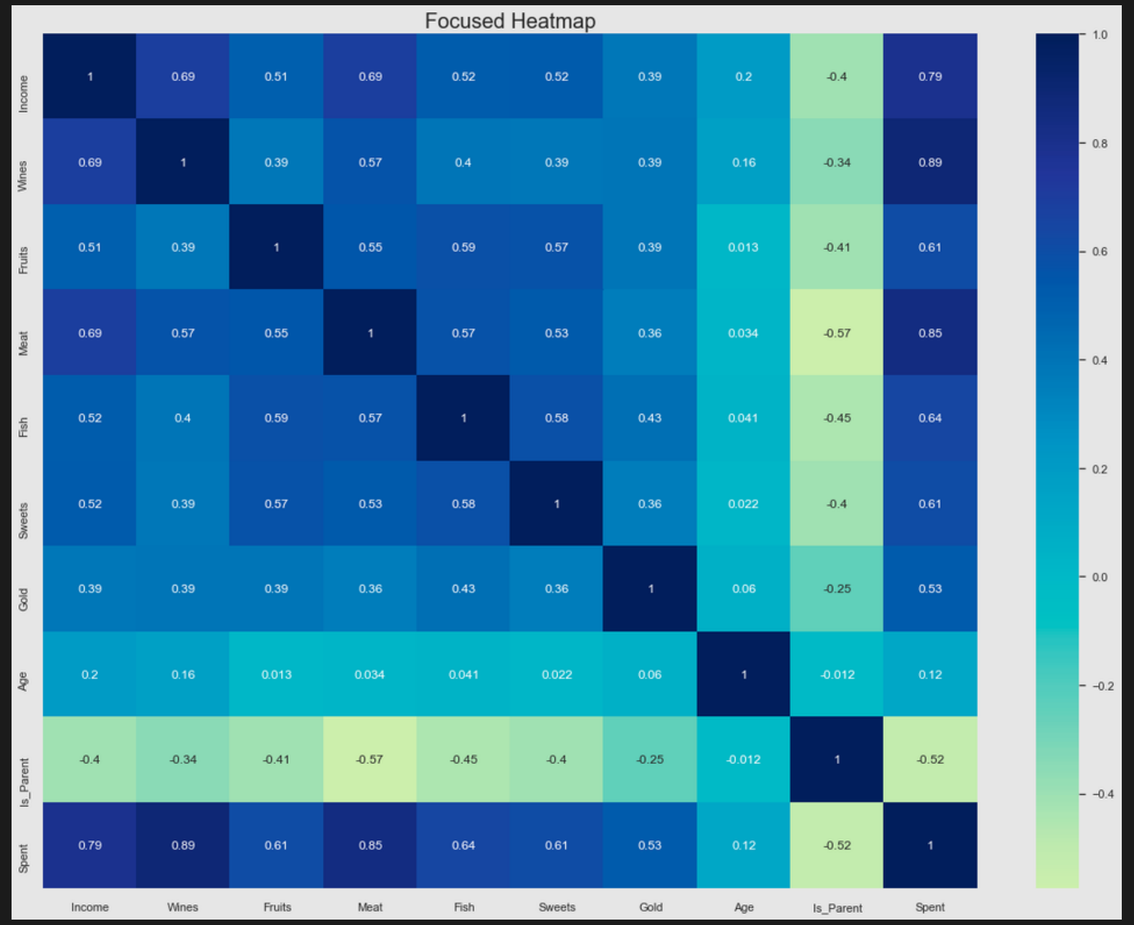
Marriage Status

64% married

How many customer clusters are present in the dataset?



Are certain product categories correlated with higher grocery spend?



Final Thoughts & Potential Uses

Businesses

Businesses with customer data could perform similar analyses in order to better understand and target their customers.

Non-profits

Universities and other non-profits could perform similar analyses in order to target donors via the most effective channel for a given certain demographic.

Government

IRS could perform similar analyses in order to determine tax payer characteristics are most likely to result in a failed audit / additional tax revenues.

Questions

References

Dataset

Grocery Store Customer Data (Kaggle)