5303 Final Project

Food App Business

Regression and Classification

Dataset: Food App Business

Description:

This dataset records 27 variables about customers using a Food App to order different goods.

Demographic Statistics:

- Age, Income, Marital Status, # of Children, # of Teenagers

Target Variables:

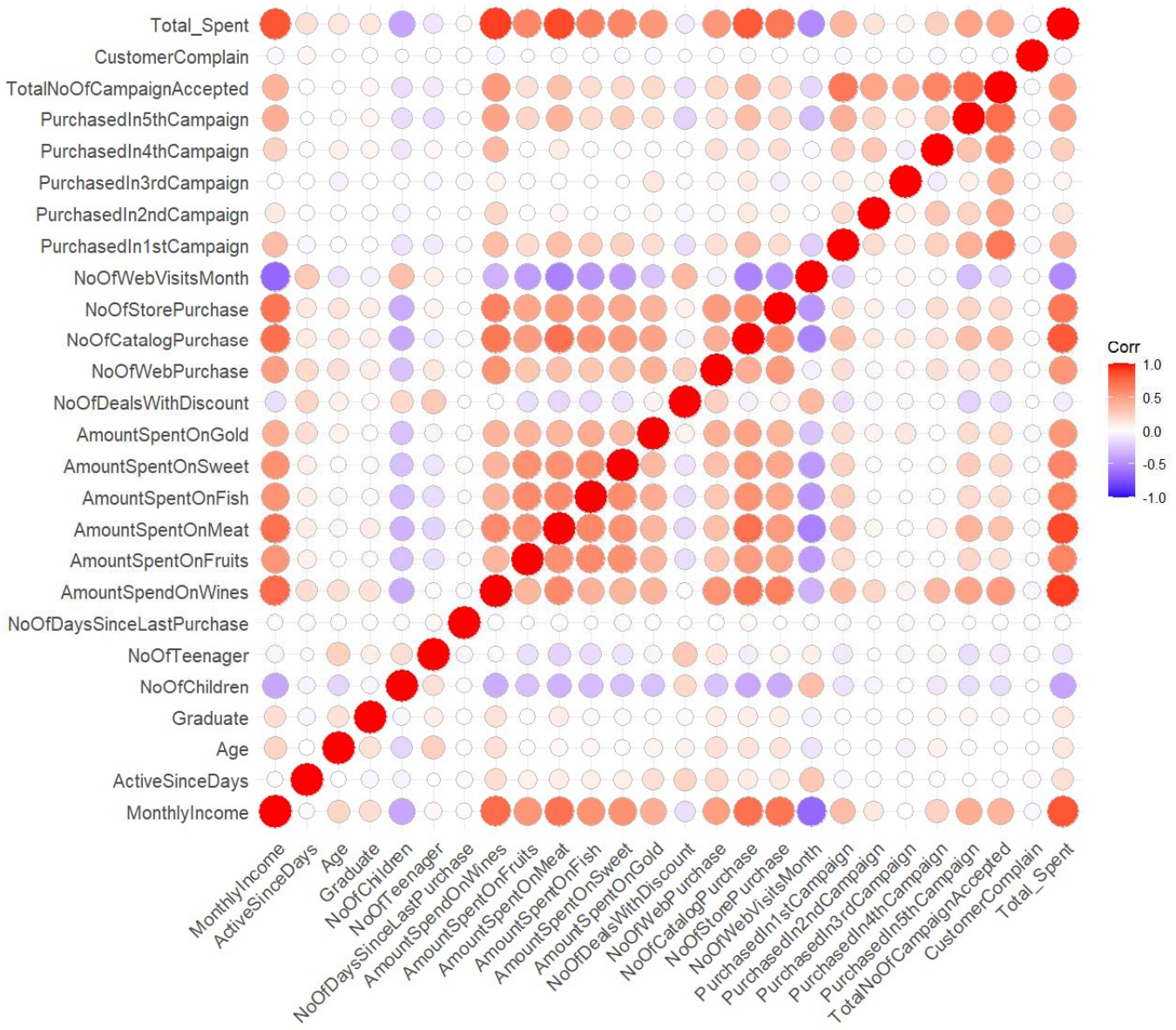
- Amount Spent on Each Category
- Amount Spent Total

Goal:

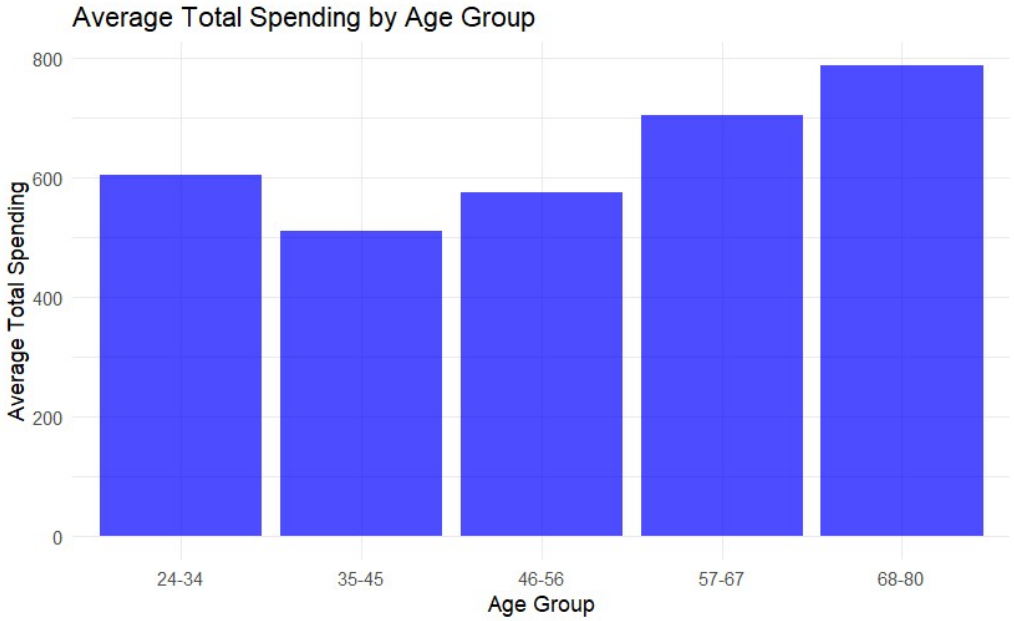
- Create actionable business insights and data driven proposals to add value to the company

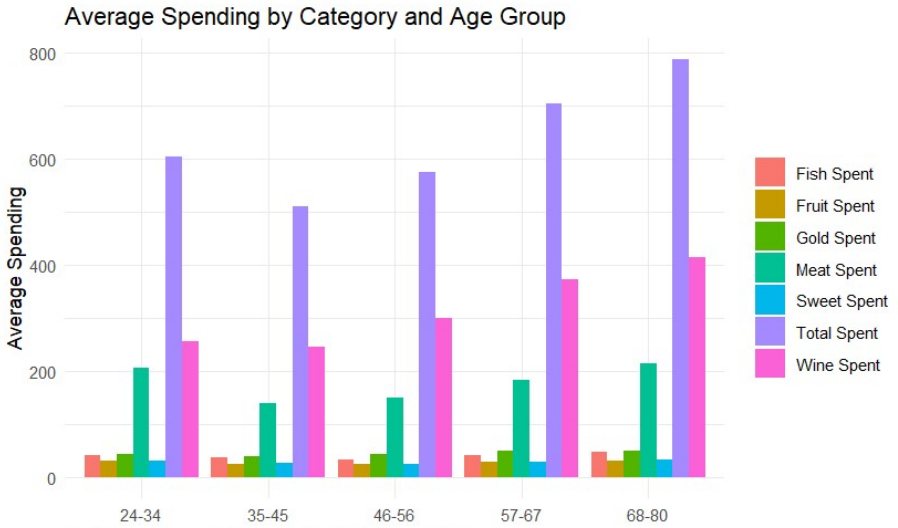
Data Preprocessing

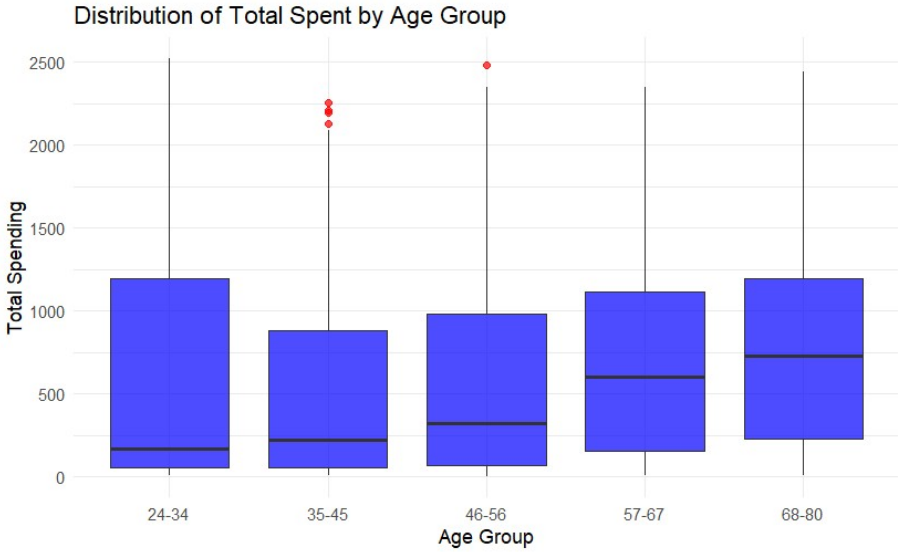
- There are 2205 entries, no missing values.
- All entries are integers, numerical.
- Combined columns, added Total_spent and Age_group columns


Total_Spent
CustomerComplain
TotalNoOfCampaignAccepted
PurchasedIn5thCampaign
PurchasedIn4thCampaign
PurchasedIn3rdCampaign
PurchasedIn2ndCampaign
PurchasedIn1stCampaign
NoOfWebVisitsMonth
NoOfStorePurchase
NoOfCatalogPurchase
NoOfWebPurchase
NoOfDealsWithDiscount
AmountSpentOnGold
AmountSpentOnSweet
AmountSpentOnFish
AmountSpentOnMeat
AmountSpentOnFruits
AmountSpendOnWines
NoOfDaysSinceLastPurchase
NoOfTeenager
NoOfChildren
Graduate
Age
ActiveSinceDays
MonthlyIncome
Corr
1.0
0.5
0.0
-0.5
-1.0

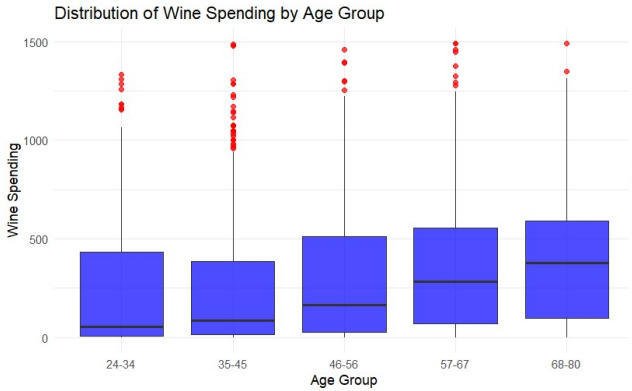
Initial Data Exploration

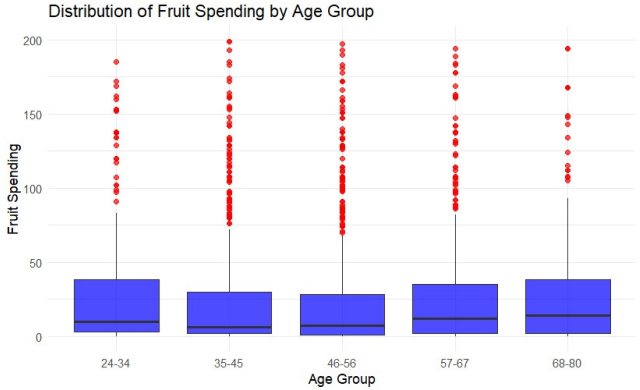


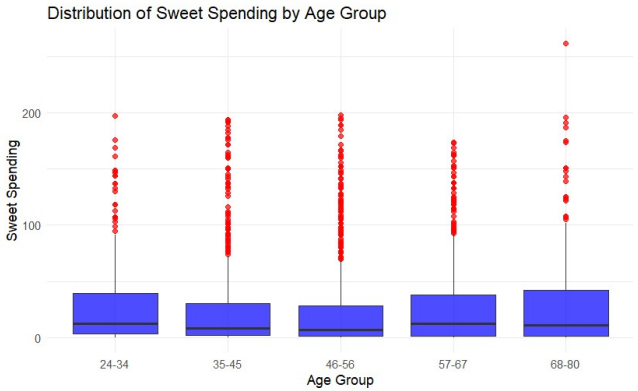


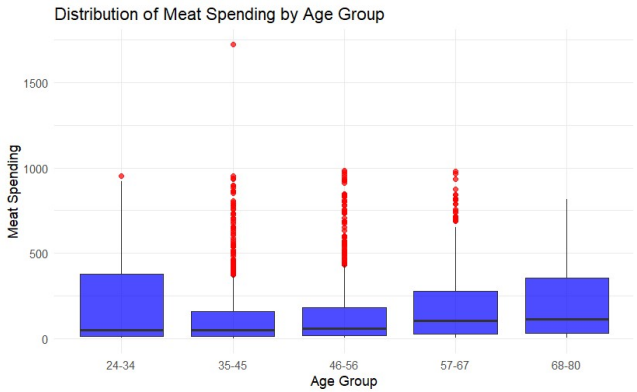


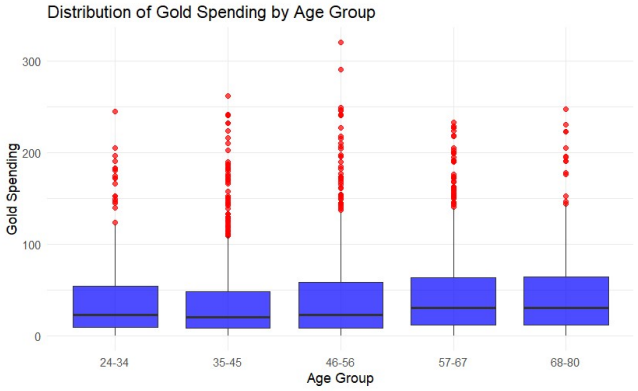
Distributions of Spending by Age Group

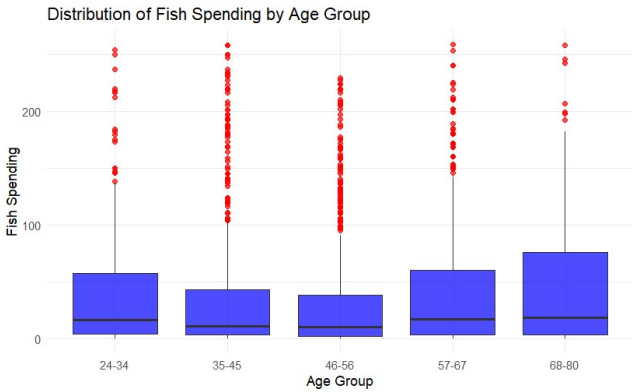






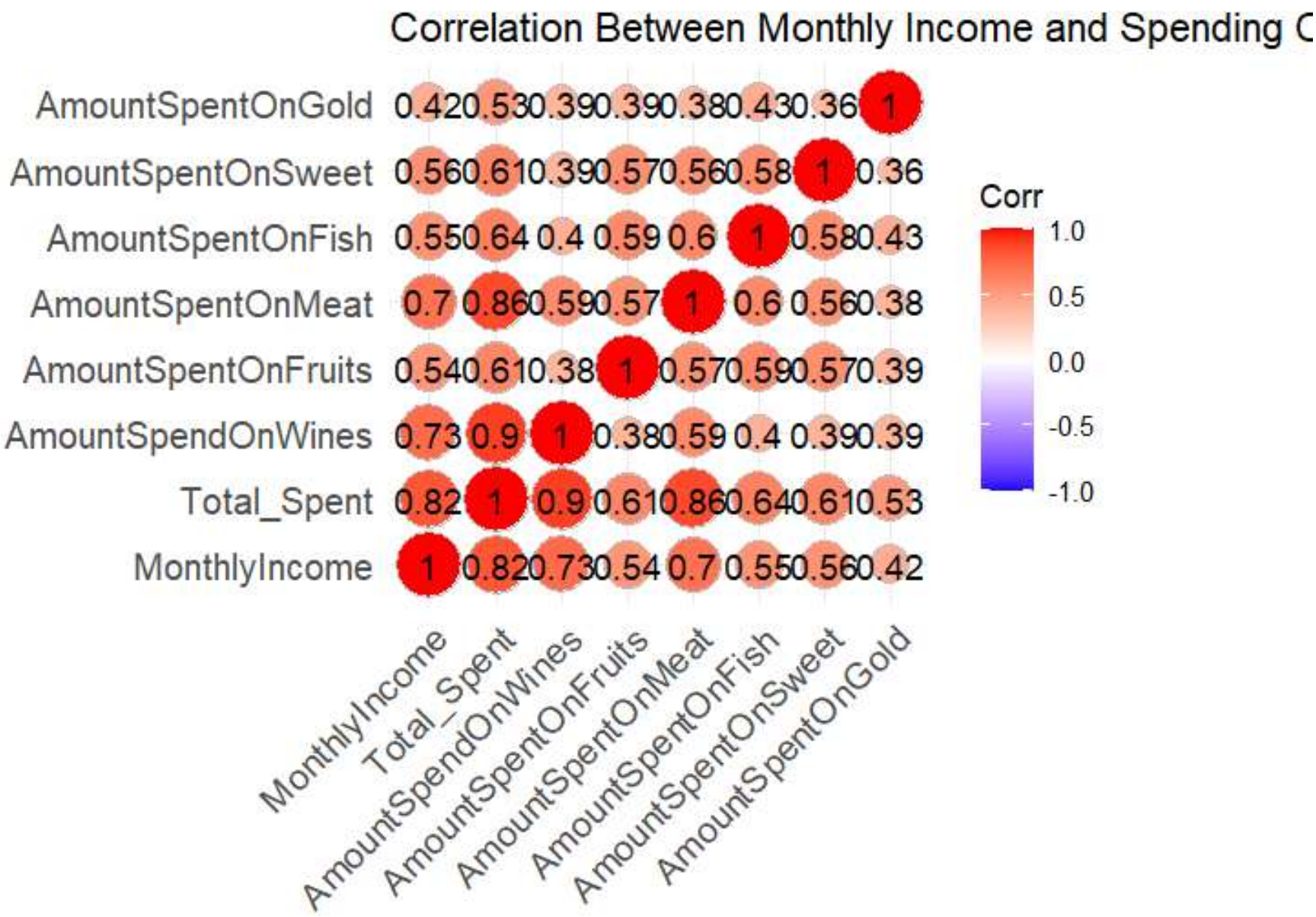






ANOVA / TURKEY Test?

- Results:
- Total Spent:
- 68-80 age group spent more than 24-34 group by $184.34
- 68-80 age group spent more than 35-45 group by $277.51
- 68-80 age group spent more than 45-56 group by $214.10
- 57-67 age group spent more than 35-45 group by $192.40
- 57-67 age group spent more than 46-56 group by $128.99
- Wine:
- 68-80 age group spent more than 24-34 group by $158.10
- 68-80 age group spent more than 35-45 group by $168.24
- 68-80 age group spent more than 45-56 group by $114.69
- 57-67 age group spent more than 24-34 group by $116.99
- 57-67 age group spent more than 35-45 group by $127.13
- 46-56 age group spent more than 35-45 group by $53.56
- Meat:
- 68-80 age group spent more than 35-45 group by $74.34
- 68-80 age group spent more than 46-56 group by $63.56
- 57-67 age group spent more than 35-45 group by $44.64
- 24-34 age group spent less than 35-45 group by $66.22
- 24-34 age group spent more than 46-56 group by $55.44
- Fish:
- 68-80 age group spent more than 46-56 group by $15.17
- 57-67 age group spent more than 46-56 group by $8.80
- Gold:
- 57-67 age group spent more than 35-45 group by $8.97
- Income and Spending Habit

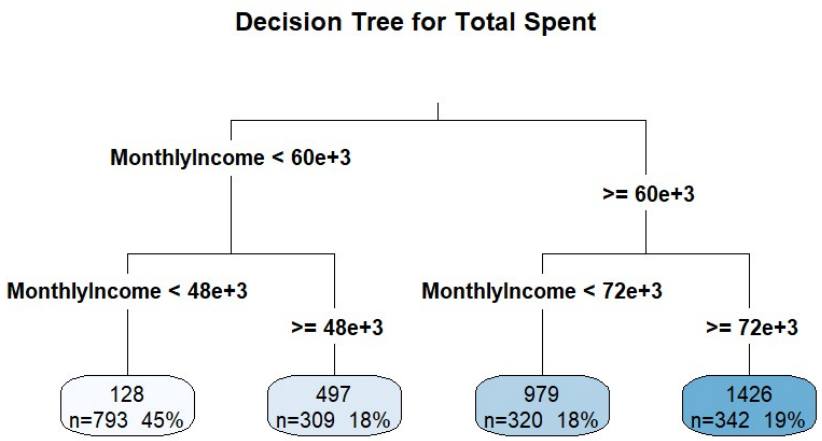

Correlation Between Monthly Income and Spending C
AmountSpentOnGold 0.42 0.53 0.39 0.39 0.38 0.43 0.36 1
AmountSpentOnSweet 0.56 0.61 0.39 0.57 0.56 0.58 1 0.36
AmountSpentOnFish 0.55 0.64 0.4 0.59 0.6 1 0.58 0.43
AmountSpentOnMeat 0.7 0.86 0.59 0.57 1 0.6 0.56 0.38
AmountSpentOnFruits 0.54 0.61 0.38 1 0.57 0.59 0.57 0.39
AmountSpendOnWines 0.73 0.9 1 0.38 0.59 0.4 0.39 0.39
Total_Spent 0.82 1 0.9 0.61 0.86 0.64 0.61 0.53
MonthlyIncome 1 0.82 0.73 0.54 0.7 0.55 0.56 0.42
MonthlyIncome
Total_Spent
AmountSpendOnWines
AmountSpentOnFruits
AmountSpentOnMeat
AmountSpentOnFish
AmountSpentOnSweet
AmountSpentOnGold
Corr
1.0
0.5
0.0
-0.5
-1.0

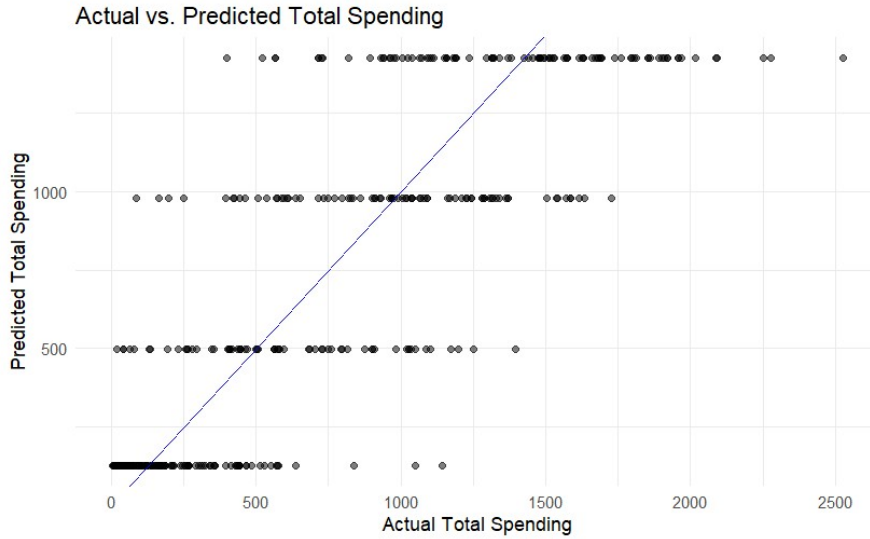
Linear Regression Models

Total Spending	y = 628.4 + 0.0239x
Wine	y = -308.3 + 0.0119x
Fruit	y = -26.9 0.001x
Meat	y = -216 + 0.0074x
Fish	y = -37.63 + 0.0015x
Sweets	y = -29.82 + 0.0011x
Gold	y = -9.795 + 0.001x

Decision Tree

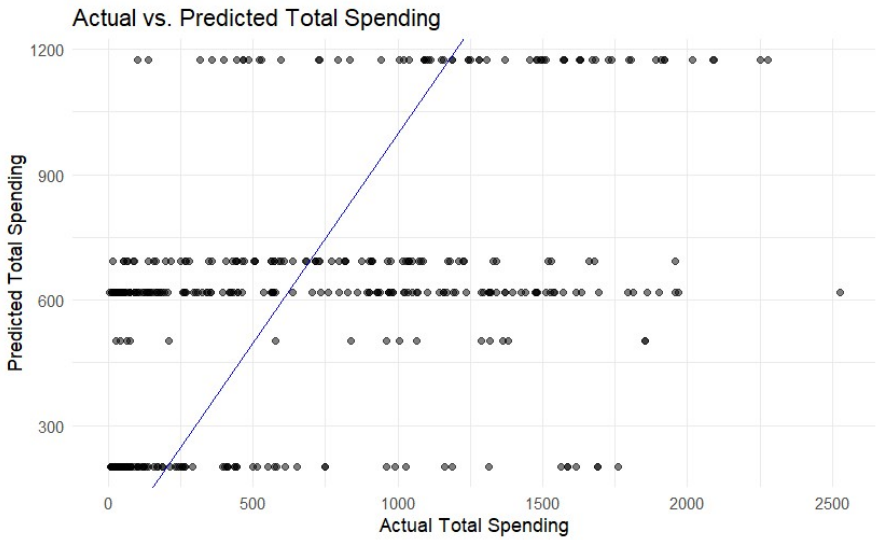
Mean Squared Error:
97140.515410515

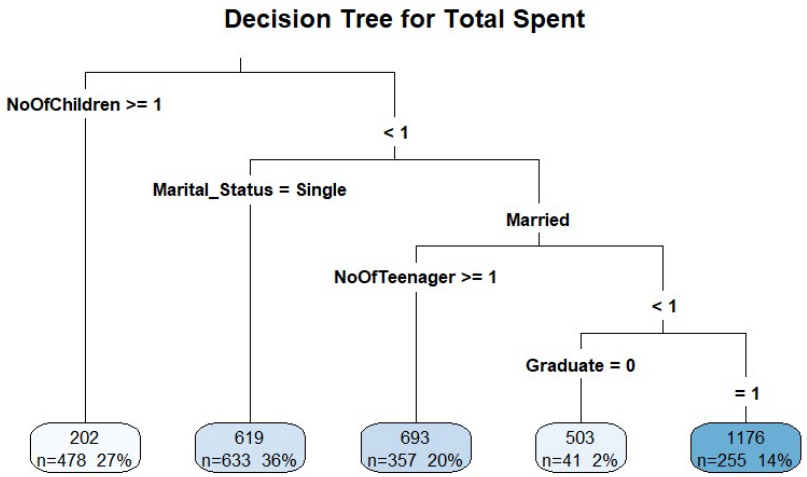




Model

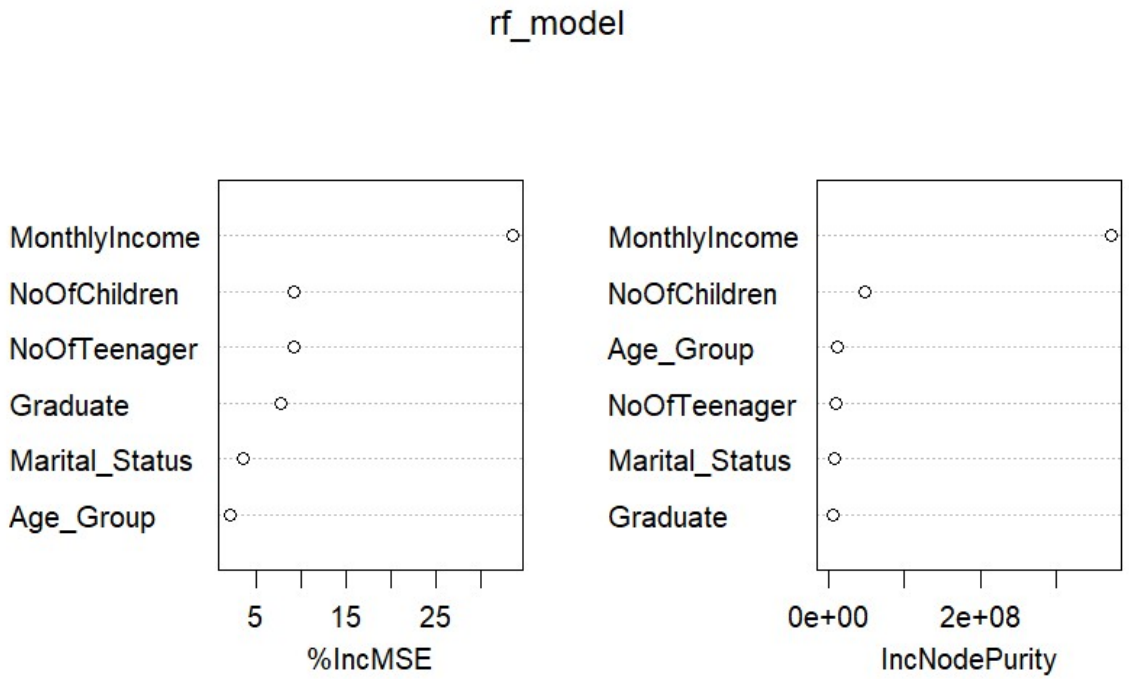
Mean Squared Error:
269454.116955018





Findings

- Monthly Income was the most important factor in Amount Spent. Number of Children is second most important.



Extra Modeling

- Linear Regression Models:
- Web:

  y = 0.6678 + 0.0000665x
- Catalog:

  y = -2.307 + 0.00009594x
- Store:

  y = 0.2714 + 0.0001076x