Homework 1

Author: DeAndre Thomas

Professor: Dr. Debbie Landowski

Course: IST 652

Date: 05/08/2022

**Data**

The data used is a recording of historical donor demographic information. It was compiled into the ‘donors\_data.csv’. This file contains 3,120 rows and 24 columns. This data was also provided by our course instructor, Dr. Debbie Landowski.

**Questions**

1. Which gender per home-ownership group received the most promotions?
2. Which gender per wealth group has received the most promotions?
3. Which gender per wealth group provides the highest donations?
4. Which gender per home-ownership groups provides the highest donations?

**Analysis**

Data Cleaning

The following columns were removed from the original dataset for this analysis due to them not being used for further analysis:

* Row Id
* Row Id.
* zipconvert\_2
* zipconvert\_3
* zipconvert\_4
* zipconvert\_5
* IC15
* TARGET\_B
* TARGET\_D

The following column names were changed to more appropriate labels:

* homeowner dummy: Homeowner
* NUMCHLD: Number\_of\_Children
* INCOME: Income
* gender dummy: Gender
* WEALTH: Wealth
* HV: Home\_Value
* Icmed: Median\_Income
* NUMPROM: Promotions
* RAMNTALL: Total\_Donation
* MAXRMNT': Max\_Donation
* LASTGIFT: Last\_Donation
* TIMELAG: Time\_Lag
* AVGGIFT: Average\_Gift
* MAXRAMNT: Max\_Donation
* totalmonths: Total\_Months

**Program Description**

The pandas, numpy, and matplotlib libraries were imported into python to perform the analysis for this assignment. The pandas library is used for data manipulation and analysis. The numpy library enable mathematical operations on data. Last, but not least, the matplotlib library enables useful visualizations to be created in Python.

Before answering question 1 (which gender per home-ownership status group received the most promotions?), a pivot chart was developed that was indexed by Gender and Home-Ownership Status, with the mean (or average amount) of Promotions being the sought value.

Graphical user interface, text

Description automatically generated

With this data from the above-described pivot table, a stacked bar chart was created to visualize the data.

Chart, bar chart

Description automatically generated

To answer question 2, (which gender per wealth group has received the most promotions?), a grouping function was used to produce an answer. In effort to find an average amount of promotions received, individuals were groups by their wealth groups, genders, and home-ownership status.

Graphical user interface

Description automatically generated with low confidence

To answer question 3 (Which gender per wealth group provides the highest donations?), a grouping function was used to produce an answer. In effort to find an average amount of donations received, individuals were groups by their wealth groups, and genders,

Table

Description automatically generated

Finally, in order to answer question 4 (Which gender per home-ownership groups provides the highest donations?), a grouping function was used to produce an answer. In effort to find an average amount of donations received, individuals were groups by their genders and home-ownership status.

**Results**

(Q1) From the pivot table, we were able to see that, assuming that Gender 1 = Male, there were more males that received promotions than females. Additionally, amongst male recipients, it appeared to be an even number of homeowners and non-homeowners that received promotions.

(Q2) Assuming that Gender 1 = Male, Non-homeowning males, in wealth group 6, on average, receive the most promotions.

(Q3) Assuming that Gender 1 = Male, on average, males in wealth group 2 give the highest donation.

(Q4) Assuming that Gender 1 = Male, on average, males that do not own their home give the highest donations.

**Conclusion**

This was a great dataset to work with. It was relatively ‘clean’, and very easy to work with. The datatypes were ready to use for analysis. I believe that this would be a great dataset to do machine learning techniques on, such as K-Means Clustering and Logistic Regression. The biggest take-aways for this assignment was gaining a better understanding of the grouping function and its syntax. It was also great being able to create a stacked bar chart!