Taylor Imhof

Bellevue University | DSC 550

Final Project Milestone #1: Write-Up

2/02/2022

**CHOOSING A DATASET**

When searching for an interesting dataset to use for the final project, I happened across some marketing data on Kaggle.com. Based on the information from Kaggle’s description, it would seem that the data is “simulated” and was created by a professor for his students to use when studying machine learning. However, after doing a little bit of EDA, I felt that this dataset would still be useful for the purposes of this project.

Dataset Link: [Marketing Data](https://www.kaggle.com/jackdaoud/marketing-data) (Kaggle.com)

**IDENTIFYING A PROBLEM**

The marketing data set that I will be using for this project contains a ton of useful fields, such as the customer’s yearly income, education level, year of birth, marital status, and more! Apart from being able to see if certain categorical variables could predict numerical values, such as the number of children at home or the customer’s yearly income, I felt and interesting problem to try to solve would be to identify whether the deployed marketing campaigns were successful or not. There are five columns detailing whether a customer accepted an offer during one of the five campaigns. I felt this would be an interesting problem to put into practice the skills we have learned throughout the course.

**GRAPHICAL ANALYSIS**

To being to get a feel for how the data is distributed in the dataset, I began to create some basic visualizations of certain columns I felt would be useful for analysis. For more detail on how I got to the graphs, please see the associated Jupyter Notebook.

**Visualizations Produced in Notebook:**

**INCOME COLUMN**

**Income Histogram With Obvious Outliers**

**Chart, histogram

Description automatically generated**

**Income Boxplot**

Chart, box and whisker chart

Description automatically generated

**Income Histogram with KDE**

Chart, histogram

Description automatically generated

**YEAR\_BIRTH COLUMN**

**Year\_Birth Boxplot**

Chart, box and whisker chart

Description automatically generated

**Year\_Birth Histogram**

Chart, histogram

Description automatically generated

**Year\_Birth Histogram w/ Education Hue**

Chart, histogram

Description automatically generated

**EDUCATION COLUMN**

**Education Bar Chart (Membership Count)**

Chart, bar chart

Description automatically generated

**COUNTRY COLUMN**

**Country Piechart**

Chart, pie chart

Description automatically generated