Northeastern University
College of Professional Studies

Module 1 - Assignment

Technique Practice

ALY6040 80440 Data Mining Applications

Instructor: Prof. Justin Grosz

Date: January 15, 2023

By,

Jeseeka Shah
Shah.jes@northeastern.edu

Introduction:

In this assignment, I will perform a basic analysis of the Kickstarter startup company known as IceCubed using the dataset that has been supplied. This statistic focuses on the role that innovators and company owners have had in a variety of different initiatives, ideas, and endeavors through providing ice cream makers. Data are facts that may teach a charity more about its patrons and the actions they do. The amount of the gift, the donor's criteria, their gender, their income level, and several other facts are examples of data. To effectively complete this project, you will need to conduct an analysis during which you will evaluate the dataset in preliminary fashion.

I have gone over the information, which includes details regarding the donated ice cream maker and the answers to the numerous inquiries. Examining customer buying patterns and making suggestions for the product categories they are most likely to buy in the future are essential for boosting sales and determining the products that consumers prefer.

Data Clean-up:

Donate ID	0
Donate Date	0
Gender	0
Deposit Amount	40
Preferred Color of Device	0
Ice Cream Products Consumed Per Week	104
Favorite Flavor Of Ice Cream	0
Donated To Kick Starter Before	0
Household Income	4801
Do you own a Keurig	0
How many desserts do you eat a week	0

Table 1: Shows the Null Values in the dataset.

The data obtained from the Kickstarter campaign was used to create a CSV file with 10,000 observations and 11 variables. With this initial design of columns and data types, unique values, and missing values, the dataset's origins may be understood. Deposit Amount (0.4%), Ice Cream Products Consumed Weekly (1%), and Household Income (48%) are among the variables that are missing. Since Deposit Amount is the only part of the information that is missing and has a right-skewed histogram, the 40 rows may be removed while still retaining the data distribution. The mean of a whole column might be used to replace the 1% missing data from Ice cream products consumed on a weekly basis with a pretty normal distribution since only 104 out of 9960 values are missing. About half of the information is missing from Household Income. You may either entirely remove the column because it has no numerical data at all, or you can modify the mode of the column by adding new values to fill in the gaps left by the column's missing values. Instead of totally erasing the column data, it is preferable to impute the missing values using the mode.

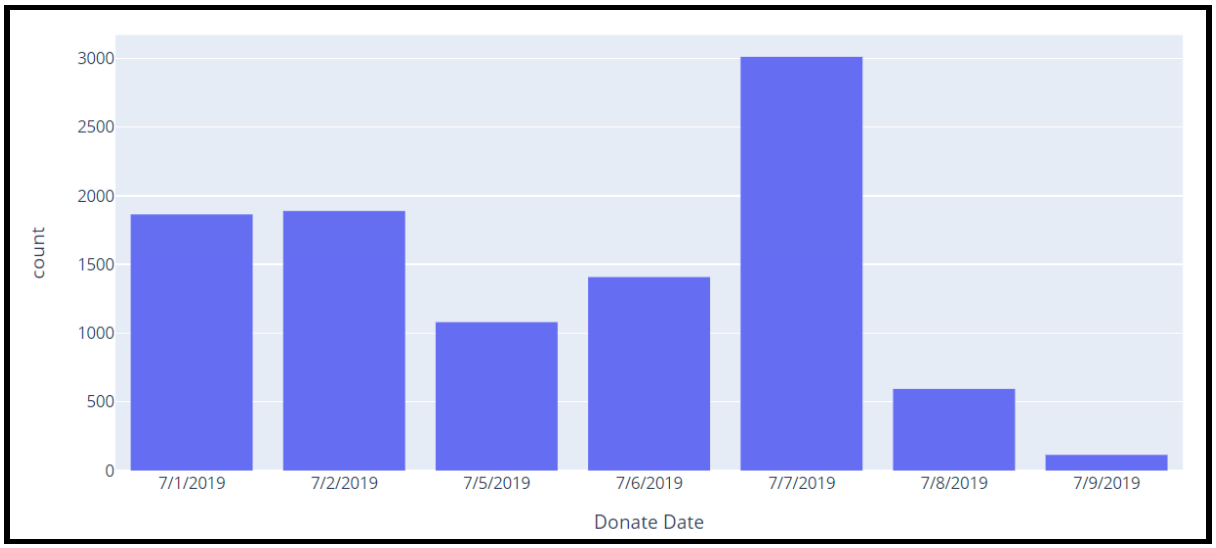
Data Analysis and Inference: -

The first step in identifying trends and creating business questions is to examine each column. Use the summary table to begin the basic examination of numerical data. Additionally, statistical data is included for each of the three number column in this table of data.

index	Donate ID	Deposit Amount	Ice Cream Products Consumed Per Week	How many desserts do you eat a week
count	10000.0000	9960.0000	9896.0000	10000.0000
mean	5000.5000	141.0700	4.9689	5.0607
std	2886.8957	127.2233	3.1753	2.8970
min	1.0000	100.0000	0.0000	0.0000
25%	2500.7500	100.0000	2.0000	3.0000
50%	5000.5000	100.0000	5.0000	5.0000
75%	7500.2500	120.0000	8.0000	7.0000
max	10000.0000	10000.0000	10.0000	100.0000

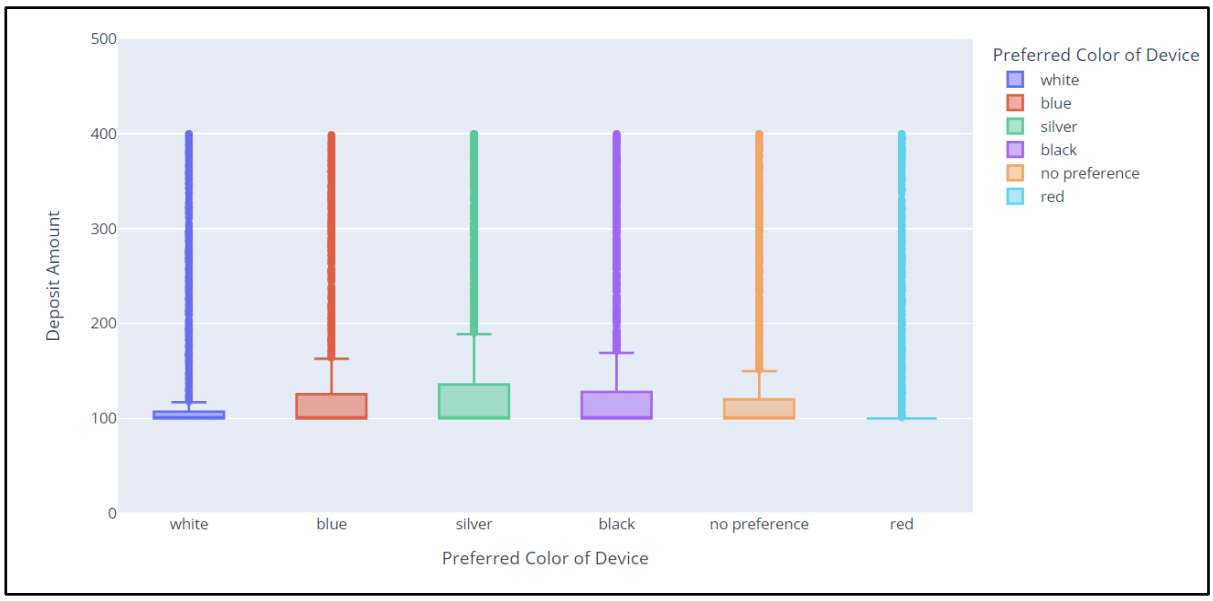
Table 1: Statistical summary.

This teaches us a lot, including the fact that the typical deposit is $141. The donated population eats five ice cream products and five sweets on average each week.



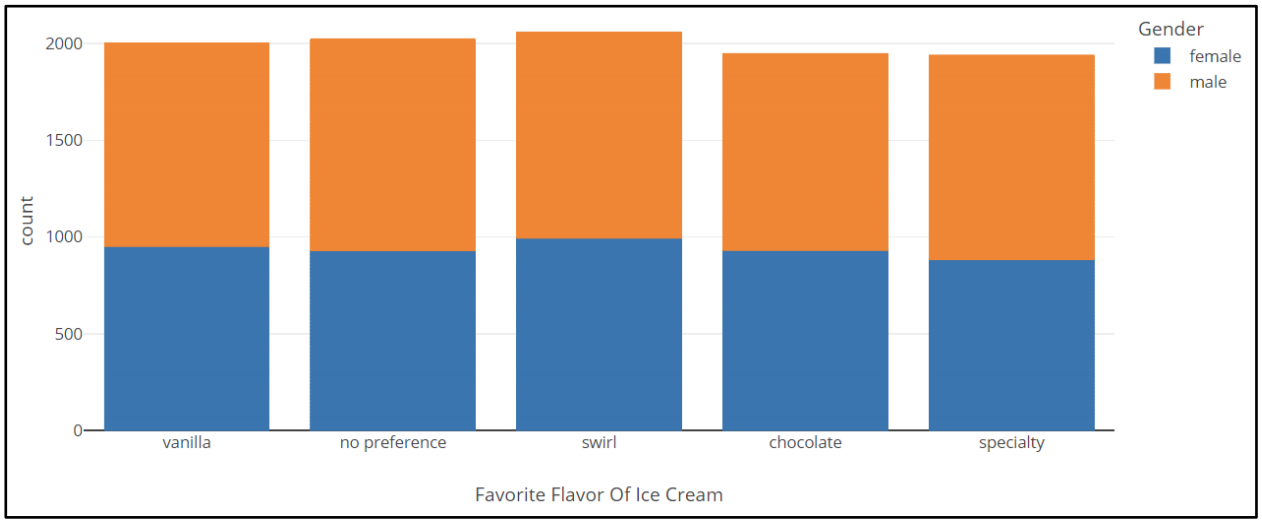
Graph 1: Donation date histogram.

On July 1 through July 9, 2019, a fundraising effort was planned. The histogram showed that the seventh and the first two days saw the bulk of the donations. A $10,000 donation was the largest ever received on July 7th, 2019. On July 3 and 4, however, no payments were made. Over the course of these nine days, $1,405,057.57 has been raised.



Graph 2 Amount Deposited based on device color preference

The above graph shows that people who have donated the most have a color preference of Silver, Blue and black whereas the people having color preference as red have donated the least amount.



Graph 4- Flavors of ice cream preference grouped by gender

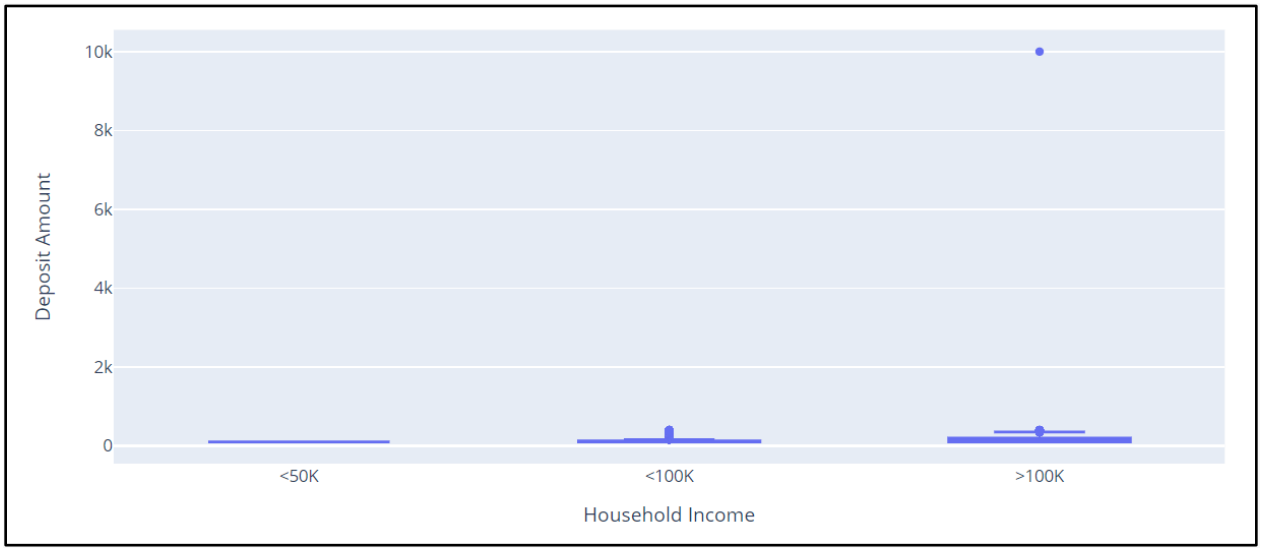
Additionally, ice cream flavor studies produced some striking results. I can see that swirl and vanilla are the most widely consumed flavors. It's also important to note that a significant percentage of voters said they had no preference at all. Most of the men preferred no option, but swirl was more well-liked by the female population.

Economics suggests that people are more likely to want to update a product if they are already familiar with it. Examples of this in action include these. Those who had previously owned a Keurig donated more than twice as frequently as those who had not. The primary target audience for marketing initiatives should be users who have previously owned a Keurig product. The fact that

3,235 of the 9,960 contributors were brand-new to the program is particularly noteworthy. This is a great concept to use as a key performance statistic.

Outlier Analysis:

Outliers in the data should be given more attention. Some data points may influence your thinking, leading you to draw the wrong conclusions.



Graph 5: Outlier Deposit Amount

In our data we see there is an outlier deposit amount of 10K which are made by the household earning of more then 100K.

Conclusion:

Exploratory data analysis using the Python programming language was completed successfully using Jupyter Notebook. Data cleansing is one of the initial jobs in every data analytics project, and I have tried to use those in this assignment.The visualization of the data helped to draw few inferences using Bamboolib. Based on my research, I believe it is ideal to make more silver and red colored items because customers are more likely to buy them. When we break it down by gender, we see that men prefer silver and women prefer red. Furthermore, persons who own a Keurig are more likely to donate than those who do not. This could be because they are increasingly accustomed to applications that provide instant gratification and demand more of it. One more thing that I found out is that in the month of July there was maximum donation recieved.This trend can be expected for coming years. Maybe they want this machine to be build so they can eat it at home or maybe they just want to eat fresh ice cream. Further analysis can be performed using gender distributions, past Keurig owners, device colors, and ice cream flavor choices.

References:

1. *Python Data Science for Everyone*. (n.d.). Retrieved November 6, 2022, from https://bamboolib.8080labs.com/

2. Plotly Python Graphing Library. (n.d.). Plotly: The front end for ML and data science models. https://plotly.com/python/