Tahiyat Mahmood

Professor Xie

CSC 1302

29 April 2023

<div align="center">Dataset Project: Blood Transfusion Service Center Data Set</div>

In this paper, I analyze the Blood Transfusion Service Center Data Set, which contains information about the donation patterns of a blood donor service center in Hsin-Chu City, Taiwan. The data set published in 2008 includes information about the number of times each donor has donated blood, the recency of their last donation, and other factors that might have influenced their decisions to donate blood. My goal is to see whether there is a strong correlation between some factors and the decision to donate blood.

The data set contains the information for 748 donors chosen at random from the donor database. The donors' information includes 5 features: Recency (how many months since the last donation), Frequency (total number of donations), Monetary (total blood donated in c.c.), Time (months since the first donation), and a binary variable representing whether the donor donated blood in March 2007 with 1 for "yes" and 0 for "no". The binary variable's column name changed to "Donated March" to simplify to tables' results.
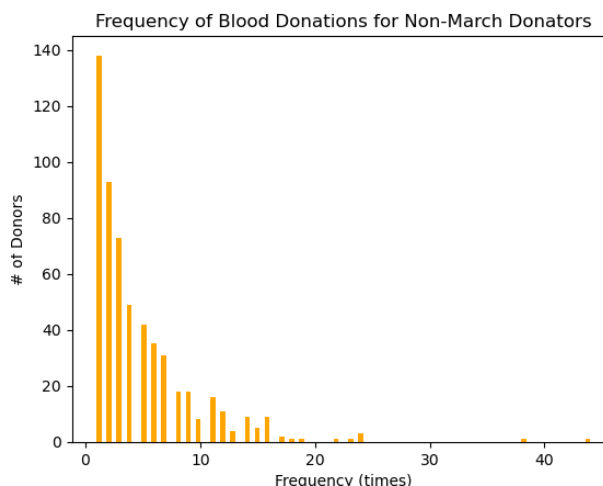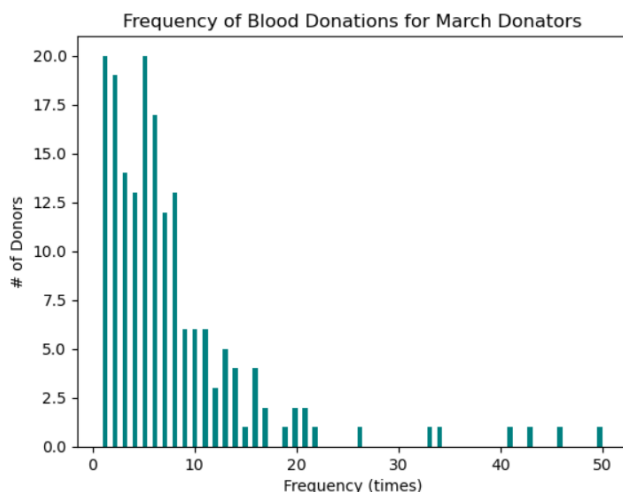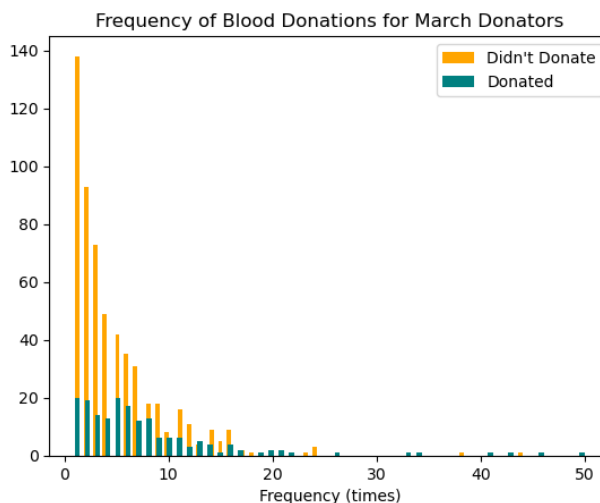

Donations in March 2007
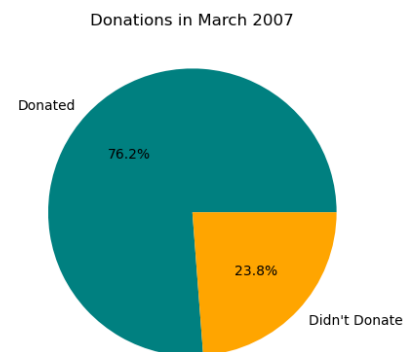
First, the data set presents an imbalance. One feature of the data is whether the donor donated in March. 23.8% of instances of donors donated, while 76.2% of instances did not donate. The data set is skewed towards those that didn't donate in March 2007. For the total data set, Time has a mean of 34.28 months, Recency with a mean of 9.5 months, and Frequency of 5.5 times. These statistics show, on average, donors would typically donate blood every 34 months, as recently as 9 months since they first donated, and about 5.8 times in total.

|	Time (months)	Recency (months)	Frequency (times)
count	748.000000	748.000000	748.000000
mean	34.282086	9.506684	5.514706
std	24.376714	8.095396	5.839307
min	2.000000	0.000000	1.000000
25%	16.000000	2.750000	2.000000
50%	28.000000	7.000000	4.000000

However, when comparing the statistics between those that donated in March 2007 versus those that hadn't, the averages change somewhat. For those that donated in March, the averages decreased to every 33 months, decreased to as recently as 5 months since they first donated, and increased to as frequently as 7.8 times in total. For those that didn't donate in March, the averages increased to every 35 months, increased to as recently as 11 months since they first donated, and decreased to as frequently as 4.8 times in total.

Donated in March

	Time (months)	Recency (months)	Frequency (times)
count	178.000000	178.000000	178.000000
mean	32.719101	5.455056	7.797753
std	23.629431	5.170615	8.036892

Did Not Donate in March

	Time (months)	Recency (months)	Frequency (times)
count	570.000000	570.000000	570.000000
mean	34.770175	10.771930	4.801754
std	24.605388	8.425039	4.746768

I graphed the Frequency of those who donated in March versus those who hadn't. The graphs are histograms to show the distribution of the Frequency as Frequency ranged from 1 to 50 and the number of donors could range from 0 to almost 140. Histograms are useful because they show the shape of the distributions. Based on the graphs, donors with 1st, 2nd, and 5th Frequency times were those more likely to donate during March. For both groups, there's a much less likely chance for donors to continue donating after 20 donations, with a noticeable decrease in chance as their frequency approached 10 donations. For March, those who chose not to donate were mostly those who had already donated less than 10 times, with around half of them those who had already donated once or twice.

Frequency of Blood Donations for March Donators

Frequency of Blood Donations for March Donators

Frequency of Blood Donations for Non-March Donators

Recency was plotted against Frequency in a scatter plot to visualize the relationship between the Frequency of blood donation and the time since the last donation. By plotting Recency against Frequency, it's difficult to determine a relationship as there's considerable overlap between the two. It is apparent that those with a higher Frequency of donations and have a shorter Recency period of less than 5 months were much more likely to donate in March. Those who chose not to donate in March were more clustered together around less than 25 months of Recencey and less than 25 times of frequency.

Donated vs Not Donated Frequency
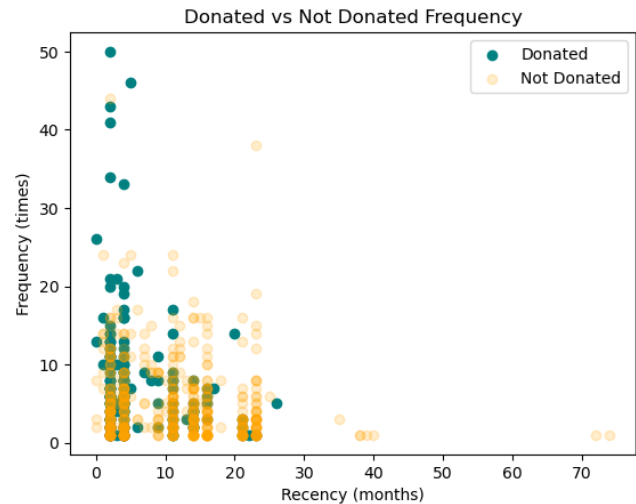
The analysis of the Blood Transfusion Service Center data set revealed a few interesting insights into the blood donation patterns of the center's donors. There is still a pattern that those with a recency of less than 5 months were the most likely to donate blood. Blood donation Recency and Frequency of donations alone don't seem to determine whether one would donate blood. This is apparent as the Frequency differences between those that donated in March versus those that didn't aren't vastly different with the dataset imbalance is taken into account. Determining whether a person would donate blood based on frequency is difficult due to the imbalance in the dataset.