Blood Transfusion Service Centre

Description and Preprocessing of Data

The tranfusion.csv is made up of 748 rows and 5 columns. The columns are as follows:

- **Recency (months)** - Number of months since the donor's most recent donation
- **Frequency (times)** - Total number of donations that the donor has made.
- **Monetary (c.c. blood)** - Total amount of blood that the donor has donated (cubic centimeters). I decided to drop this column because the Frequency already captures the data that Monetary does, since the Monetary value will be dictated by the Frequency variable.
- **Time (months)** - Number of months since the donor's first donation. I also decided to drop this variable because my initial thought was that Recency is a better way to predict whether a person will donate blood than Time is in this instance. However, I will explore an alternative where I include this variable in the model.
- **whether he/she donated blood in March 2007** - This is a binary variable which represents whether the donor donated blood in March 2007

Project Description

The aim for this project was to predict whether a person would donate blood in March 2007 using a Logistic Regression model which was coded using Python. The model created finished up with an accuracy of approximately 78% and a guessed correctly approximately 75% of data it had not seen before. Therefore using this model we would have predicted whether or not a person would donate blood with approximately 75% accuracy.