Faculty of Science and Technology
Assignment Cover Sheet

Student ID number & Student Name	U3224132 Nelson Hall
Unit name	Software Technology 1
Unit number	(4483)
Unit Tutor	Mr. Pranav Gupta
Assignment name	ST1 Capstone Project – Semester 1 2023
Due date	Week 13 Friday 11.59 pm
Date submitted	11/05/2023

You must keep a photocopy or electronic copy of your assignment.

Student declaration
I certify that the attached assignment is my own work. Material drawn from other sources has been appropriately and fully acknowledged as to author/creator, source and other bibliographic details.

Signature of student: Nelson Hall
Date: 2/05/2023

Activity Log:

Weeks	Details
Week 8	• Chose Blood Transfusion Dataset to use for my Capstone Project. I selected this topic as I thought the attributes would be easy to understand and read for the project.
Week 9	• During Week 9 I started to develop Ideas for the project, specifically how I would use the dataset to create a Python GUI program
Week 10	• Review class templates in order to inform myself of the certain structure required to form the Assessment
Week 11	• Started working on my document inorder to develop a plan/algorithm for the task
Week 12	• Developed my Python GUI program, the code still has errors which need to be addressed before the final submission.
Week 13	• Went over code to make final changes • Submitted through github • Submitted on canvas

Introduction:

This report presents a comprehensive overview of the Python Capstone Project designed for the Software Technology 1 unit, which centres on the Blood Transfusion Service Center Data Set [1]. This dataset is obtainable from UCI [2], and the project requirements specified in the assignment handout were meticulously adhered to throughout the project development phase.

Blood transfusion is a medical procedure in which a patient receives blood or blood products from a donor [3]. This procedure is typically performed to replace lost blood due to injury, surgery, or a medical condition that results in anaemia. The process involves several steps, including donor screening, blood testing, and blood typing to ensure compatibility between the donor and recipient [4]. Once compatibility is confirmed, the blood is collected from the donor and processed to separate the various blood components. These components include red blood cells, platelets, plasma, and white blood cells, each with a unique role in the body. The necessary components are then transfused to the patient through a vein, where they are distributed throughout the body to perform their respective functions [5]. Blood transfusions are a vital aspect of modern medicine and have saved countless lives worldwide.

The objective of the project is to use the blood transfusion data set to create a model that predicts whether or not a donor will donate blood again. The data set includes several features, such as the number of times the donor has donated blood, the time since the donor's last donation, the total volume of blood donated, and the donor's age. The program will also present additional data visualisation in the hopes of answering questions 1 to 5 and provide extra analysis inorder to review the dataset.

Methodology:

1. Obtain the Blood Transfusion Service Center Data Set from UCI.

2. Familiarise yourself with the data set and its features, including the number of times the donor has donated blood, the time since the donor's last donation, the total volume of blood donated, and the donor's age.

3. Determine the project requirements specified in the assignment handout and ensure they are meticulously adhered to throughout the project development phase.

4. Develop a model that predicts whether or not a donor will donate blood again using the data set.

5. Create additional data visualisations to answer questions 1 to 5 and provide extra analysis to review the dataset.

6. Test and refine the model to improve its accuracy.

7. Present the project in a comprehensive report that outlines the methodology, findings, and recommendations.

8. Ensure that the report adheres to the requirements specified in the assignment handout and includes all necessary information.

9. Submit the report and any accompanying materials, such as code or visualisations, according to the assignment instructions.

10. Finally, review and evaluate the project outcome to identify areas for improvement and future research.

Algorithm Design Stage:

Pseudocode:

The pseudo code briefly explains the process I will take in order to develop the program while providing a guide to assist during its development and implementation.

```
1. Load the Blood Transfusion Service Center Data Set into the program.

2. Preprocess the data set to ensure it is clean and ready for analysis.

3. Split the data set into training and testing sets.

4. Select a suitable machine learning algorithm for the task, such as logistic regression or
decision tree.

5. Train the algorithm using the training set.

6. Test the trained algorithm using the testing set.

7. Evaluate the algorithm's performance using suitable metrics such as accuracy, precision,
and recall.

8. If the algorithm's performance is not satisfactory, refine the algorithm by adjusting its
parameters, selecting a different algorithm, or performing feature engineering.

9. Once the algorithm's performance is satisfactory, use it to predict whether or not a donor
will donate blood again.

10. Create additional data visualizations to provide extra analysis and insights into the
data set, such as histograms or scatter plots.
```

Dataset Description:

The Blood Transfusion Service Center Data Set [1] used in this project was collected from the Blood Transfusion Service Center in Hsin-Chu City, Taiwan. The dataset includes data from 748 donors, with each donor having a record of their R (Recency - months since last donation), F (Frequency - total number of donation), M (Monetary - total blood donated in c.c.), T (Time - months since first donation), and a binary variable indicating whether the donor donated blood in March 2007 (1 stands for donating blood; 0 stands for not donating blood). The goal of this project is to create a program that predicts whether or not a donor will donate blood again using this dataset.

Exploratory Data Analysis:

Imported libraries:

```
import tkinter as tk
import pandas as pd
import plotly.graph_objs as go
import plotly.express as px
import numpy as np
```

Understanding the data and providing the program with libraries such as tkinter, pandas, and plotly is a necessary step in any data analysis or visualisation project. These libraries provide powerful tools for manipulating and presenting data in a user-friendly manner.

Tkinter is a built-in Python library used for creating graphical user interfaces (GUI) and is commonly used for creating applications that require a simple interface. Pandas is a popular library used for data manipulation and analysis. It provides an easy way to work with data and perform tasks such as reading and writing data to and from files, manipulating dataframes, and filtering data.

Plotly is a library used for creating interactive and dynamic visualisations. It offers a wide range of charts and graphs that can be customised to suit various needs. This library provides high-quality and interactive visualisations that can be used to communicate complex data in a clear and concise manner.

Numpy is another important library that is commonly used in data analysis and scientific computing. It provides powerful tools for working with arrays and matrices, performing mathematical operations, and generating random numbers.

Overall, the use of these libraries is essential in any data analysis project as they provide the necessary tools for data manipulation, analysis, and visualisation. They make it easier to perform complex tasks and present data in a clear and concise manner, making it more accessible to users.

Loading the dataset:

```
# Load the data
data = pd.read_csv('transfusion.data')
```

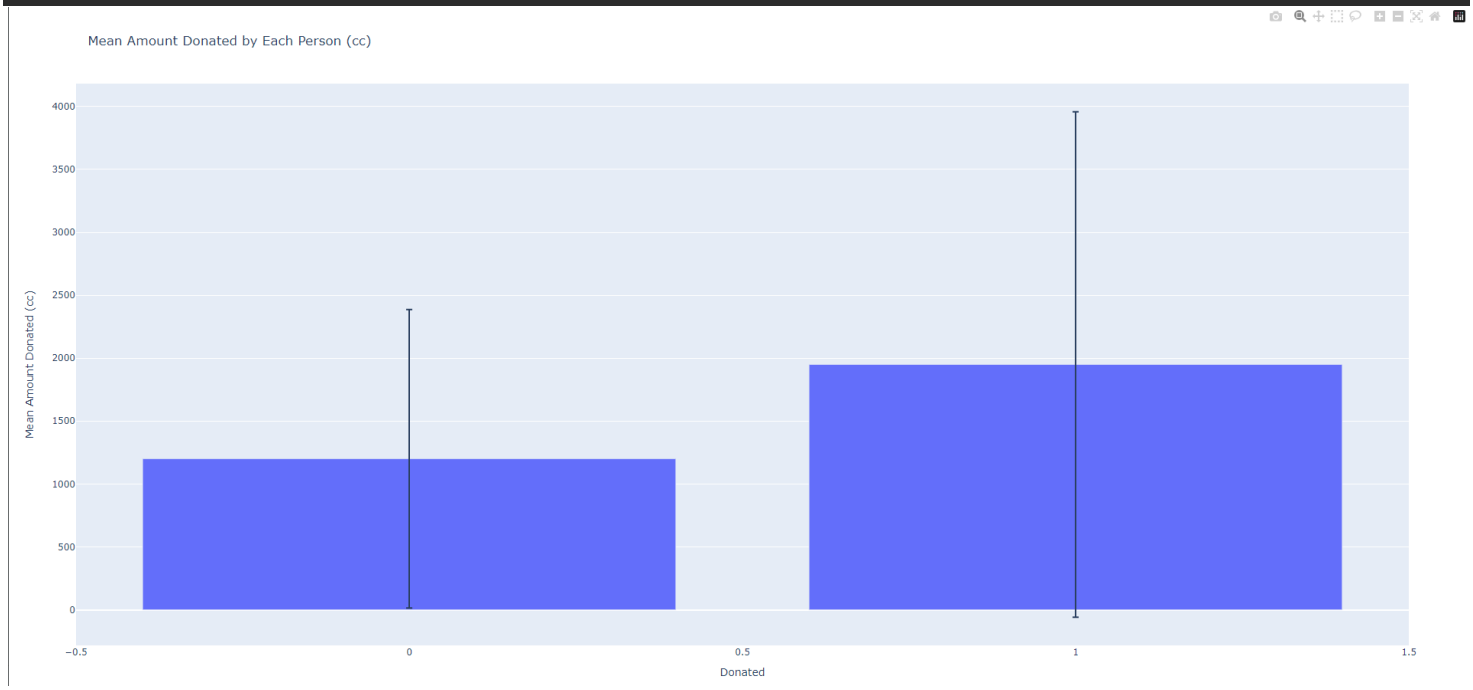
The next step of uncoding the data is that it must load the data from a CSV file named 'transfusion.data' into a pandas dataframe called 'data' and be able to interpret the following data set. Without loading the data, we would not be able to perform any analysis on the data or build any models to predict blood donation patterns.

List Attributes:

R (Recency - months since last donation)
F (Frequency - total number of donation)
M (Monetary - total blood donated in c.c.)
T (Time - months since first donation), and a binary variable representing whether he/she donated blood in March 2007 (1 stands for donating blood; 0 stands for not donating blood).

Visualising Data Distribution In Detail:

```
fig2 = go.Figure(data=[go.Bar(x=donation_stats['Donated_Mar_2007'],
y=donation_stats['mean'],
                              error_y=dict(type='data', array=donation_stats['std']))])
fig2.update_layout(title='Mean Amount Donated by Each Person (cc)',
                   xaxis_title='Donated',
                   yaxis_title='Mean Amount Donated (cc)')
fig2.show(renderer='browser', auto_open=False)
```



Algorithm Implementation Stage:

The following algorithm developed upon through this project has proven to be successful and will be the algorithm expressed within my code development, the algorithm can be found within the below github repository:
https://github.com/NelsonHall0/CapstoneProjectST1u3224132.py

Software Deployment Stage:

This code will be a Python program that uses the Tkinter library to create a GUI window for a blood donation predictor. The user will be able to input their donation history, and the program will predict the likelihood of them donating blood in the future. Additionally, there will be a button to display visualisations of the donation data. The program will read in a CSV file containing the donation data and renames the columns for clarity. I will then create the Tkinter window and define the input fields for the user to input their donation history. There will also be an error message label in case the user inputs invalid data. This can all be found in the github repository below.
https://github.com/NelsonHall0/CapstoneProjectST1u3224132.py

References:

[1] Blood Transfusion Service Center Data Set
https://archive.ics.uci.edu/ml/datasets/Blood+Transfusion+Service+Center

[2] UCI Machine Learning Repository
https://archive.ics.uci.edu/ml/datasets.php

[3] National Heart, Lung, and Blood Institute. What is a blood transfusion?
https://www.nhlbi.nih.gov/health-topics/blood-transfusion

[4] World Health Organization. Blood safety and availability.
https://www.who.int/news-room/fact-sheets/detail/blood-safety-and-availability

[5] American Red Cross. Blood transfusion.
https://www.redcrossblood.org/donate-blood/blood-donation-process/what-happens-to-donated-blood/blood-transfusion.html