
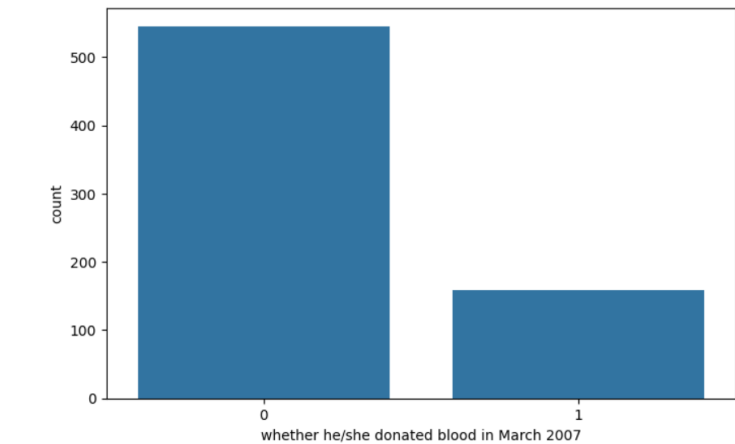


Data Collection and Preprocessing Phase

Date	14 July 2024
Team ID	739949
Project Title	Blood Donation Prediction
Maximum Marks	6 Marks

Data Exploration and Preprocessing Template

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	Description
Data Overview	
Univariate Analysis	<p>Code + Text</p> <p>UNIVARIATE ANALYSIS</p> <pre>sns.countplot(x='whether he/she donated blood in March 2007',data=df)</pre> <p><Axes: xlabel='whether he/she donated blood in March 2007', ylabel='count'></p> 

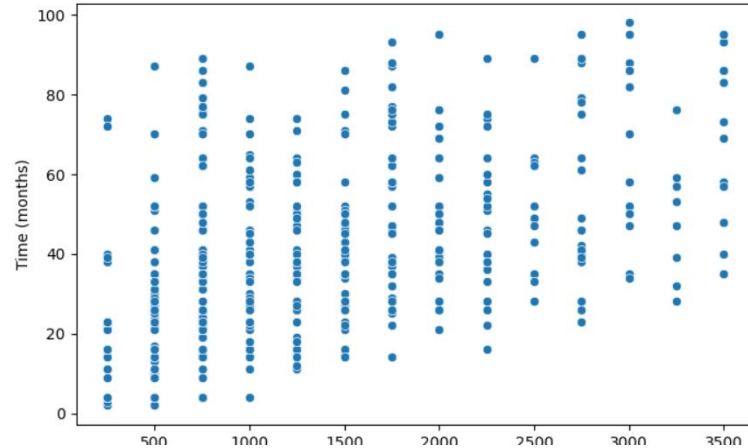
Bivariate Analysis

Code + Text

BIVARIATE ANALYSIS

```
sns.scatterplot(x=df['Monetary (c.c. blood)'],y=df['Time (months)'])
```

<Axes: xlabel='Monetary (c.c. blood)', ylabel='Time (months)'

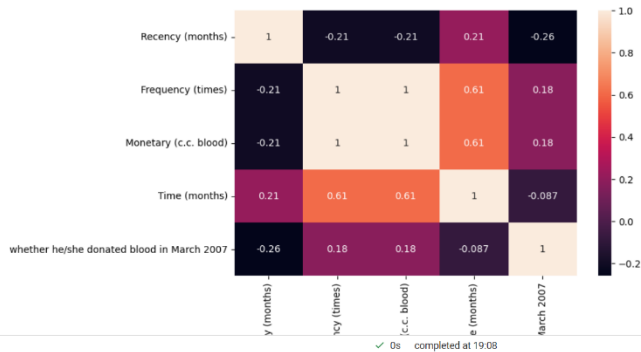


Multivariate Analysis

MULTIVARIATE ANALYSIS

```
[92] sns.heatmap(df.corr(),annot=True)
```

<Axes: >

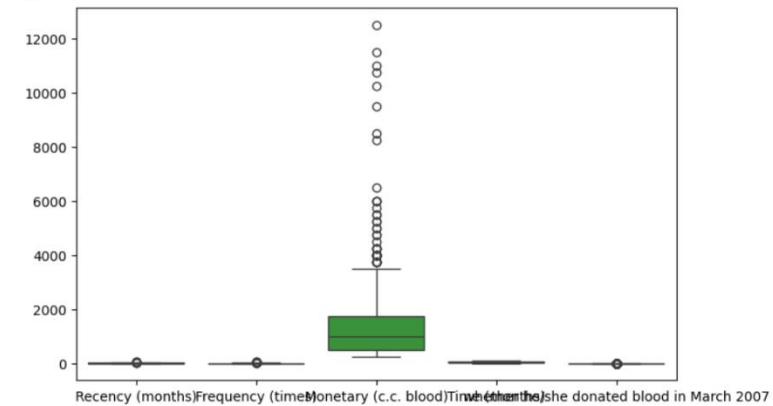


Outliers and Anomalies

Code + Text

```
sns.boxplot(df)
```

<Axes: >



Data Preprocessing Code Screenshots

Loading Data

```
[62]: df = pd.read_csv("/content/transfusion (2).csv")
df.head()
```

	Recency (months)	Frequency (times)	Monetary (c.c. blood)	Time (months)	whether he/she donated blood in March 2007
0	2	50	12500	98	1
1	0	13	3250	28	1
2	1	16	4000	35	1
3	2	20	5000	45	1
4	1	24	6000	77	0

Handling Missing Data

```
df.isnull().sum()
```

```
Recency      0
Frequency    0
Monetary     0
Time         0
Donated      0
dtype: int64
```

Data Transformation

```
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
scaled_features = scaler.fit_transform(df.drop(columns=["whether he/she donated blood in March 2007"]))
scaled_df = pd.DataFrame(scaled_features, columns=df.columns[:-1])

scaled_df["whether he/she donated blood in March 2007"] = df["whether he/she donated blood in March 2007"].values

# Display the transformed DataFrame
print("\nTransformed DataFrame:")
print(scaled_df.head())
```

```
Transformed DataFrame:
   Recency (months)  Frequency (times)  Monetary (c.c. blood)  Time (months) \
0      0.000000      0.923077      0.923077      0.270833
1      0.054054      0.230769      0.230769      0.020833
2      0.027027      0.461538      0.461538      0.125000
3      0.013514      0.846154      0.846154      0.343750
4      0.027027      0.615385      0.615385      0.208333

   whether he/she donated blood in March 2007
0              1
1              0
2              1
3              0
4              1
```

Feature Engineering

Attached the codes in final submission.

Save Processed Data

-