

Faculty of Science and Technology

Assignment Coversheet

Student ID number & Student Name	u3243692 Burton Vanderstok
Unit name	Software Technology 1
Unit number	4483
Unit Tutor	Mr. Pranav Gupta
Assignment name	ST1 Capstone Project – Semester 1 2023
Due date	12/05/2023
Date submitted	12/05/2023

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: ____burton____

Date: _____

Table of Contents

Algorithm Design Stage	3
Dataset Description	3
Exploratory Data Analysis	3
Logbook.....	6

Algorithm Design Stage

Starting with cleaning the dataset to allow for the predictive analysis then present the new data with different visuals to show the information in a clear view

Dataset Description

For my data I have selected a traffic camera offence where the data set has recorded; Offence_Month, Rego_State, Clt_Catg, Camera_Type, Location_Code, Location_Desc, Offence_Desc, Sum_Pen_Amt, Sum_Inf_Count, Sum_With_Amt and Sum_With_Count, which describes offences in different states and when and how they were caught.

Exploratory Data Analysis

```
import pandas as pd
```

```
df= pd.read_csv('Traffic_camera_offences_and_fines.csv')  
df.head(5)
```

C:\Users\burto\AppData\Local\Temp\ipykernel_29604\2715242604.py:3: DtypeWarning: Columns (4) have mixed types. Specify dtype option on import or set low_memory=False.
df= pd.read_csv('Traffic_camera_offences_and_fines.csv')

	Offence_Month	Rego_State	Clf_Catg	Camera_Type	Location_Code	Location_Desc	Offence_Desc	Sum_Pen_Amt	Sum_Inf_Count	Sum_With_Amt	Sum_With_Count
0	Mar 2022	ACT	Diplomatic	FIXED ONLY SPEED CAMERA	1035	TUGGERANONG PARKWAY NEAR COTTER ROAD OVERPASS ...	20 Non-school zone exceed speed limit by <= 15...	301.0	1	0.0	0
1	Mar 2022	ACT	Diplomatic	FIXED ONLY SPEED CAMERA	1027	BARTON HIGHWAY BETWEEN GUNGAHLIN DRIVE AND ELL...	20 Non-school zone exceed speed limit by <= 15...	602.0	2	0.0	0
2	Mar 2022	ACT	Diplomatic	MOBILE SPEED CAMERA	2040	BLAMEY CRESCENT - CANBERRA	20 School Zone Exceed Speed Limit	1305.0	1	0.0	0

```
df.tail(5)
```

	Offence_Month	Rego_State	Clf_Catg	Camera_Type	Location_Code	Location_Desc	Offence_Desc	Sum_Pen_Amt	Sum_Inf_Count	Sum_With_Amt	Sum_With_Count
120997	Apr 2023	WA	OTHER PRSN	RED LIGHT AND SPEED CAMERA	1024	CANBERRA AVENUE/CAPTAIN COOK CRESCENT/MANUKA C...	59 (1) Enter Intersection or Marked Foot Cross...	502.0	1	0.0	0
120998	Apr 2023	WA	OTHER PRSN	RED LIGHT AND SPEED CAMERA	1002	NORTHBOURNE AVENUE/BARRY DRIVE/COOYONG STREET	20 Non-School Zone Exceed Speed Limit > 15 But...	908.0	2	0.0	0
120999	Apr 2023	WA	OTHER PRSN	RED LIGHT AND SPEED CAMERA	1020	BARRY DRIVE/MARCUS CLARKE STREET	20 Non-school zone exceed speed limit by <= 15...	1535.0	5	0.0	0
121000	Apr 2023	WA	OTHER PRSN	RED LIGHT AND SPEED CAMERA	1002	NORTHBOURNE AVENUE/BARRY DRIVE/COOYONG STREET	20 Non-school zone exceed speed limit by <= 15...	3377.0	11	0.0	0

```
df.info(5)
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 121002 entries, 0 to 121001
```

```
Data columns (total 11 columns):
```

#	Column	Non-Null Count	Dtype
0	Offence_Month	121002 non-null	object
1	Rego_State	121000 non-null	object
2	Clf_Catg	112187 non-null	object
3	Camera_Type	121002 non-null	object
4	Location_Code	121002 non-null	object
5	Location_Desc	121002 non-null	object
6	Offence_Desc	121002 non-null	object
7	Sum_Pen_Amt	120110 non-null	float64
8	Sum_Inf_Count	121002 non-null	int64
9	Sum_With_Amt	120227 non-null	float64
10	Sum_With_Count	121002 non-null	int64

```
dtypes: float64(2), int64(2), object(7)
```

```
memory usage: 10.2+ MB
```

```
df.describe()
```

	Sum_Pen_Amt	Sum_Inf_Count	Sum_With_Amt	Sum_With_Count
count	1.201100e+05	121002.000000	120227.000000	121002.000000
mean	2.782476e+03	7.492529	206.178812	0.509537
std	1.516218e+04	45.292916	1261.494230	3.582583
min	0.000000e+00	0.000000	0.000000	0.000000
25%	2.970000e+02	1.000000	0.000000	0.000000
50%	7.600000e+02	1.000000	0.000000	0.000000
75%	1.950000e+03	4.000000	0.000000	0.000000
max	1.766569e+06	5869.000000	134246.000000	446.000000

```
df2 = df
for x in df['Offence_Month'].unique().tolist():
    y = x.split()
    if y[0] == "Jan":
        y[0] = "1"
    if y[0] == "Feb":
        y[0] = "2"
    if y[0] == "Mar":
        y[0] = "3"
    if y[0] == "Apr":
        y[0] = "4"
    if y[0] == "May":
        y[0] = "5"
    if y[0] == "Jun":
        y[0] = "6"
    if y[0] == "Jul":
        y[0] = "7"
    if y[0] == "Aug":
        y[0] = "8"
    if y[0] == "Sep":
        y[0] = "9"
    if y[0] == "Oct":
        y[0] = "10"
    if y[0] == "Nov":
        y[0] = "11"
    if y[0] == "Dec":
        y[0] = "12"
    df2 = df2.replace(to_replace=x,value=y[0]+y[1])
df2.tail(5)
```

```
df2.corr(numeric_only=True)
```

	Sum_Pen_Amt	Sum_Inf_Count	Sum_With_Amt	Sum_With_Count
Sum_Pen_Amt	1.000000	0.907804	0.250299	0.298889
Sum_Inf_Count	0.907804	1.000000	0.277102	0.358364
Sum_With_Amt	0.250299	0.277102	1.000000	0.850923
Sum_With_Count	0.298889	0.358364	0.850923	1.000000

```
from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import chi2
```

```
X= df2.iloc[:,1:14]
Y= df2.iloc[:,1]
```

```
best_features= SelectKBest(score_func=chi2, k=3,)
fit= best_features.fit(X,Y)
```

```
df_scores= pd.DataFrame(fit.scores_)
df_columns= pd.DataFrame(X.columns)
```

```
features_scores= pd.concat([df_columns, df_scores], axis=1)
features_scores.columns= ['Features', 'Score']
features_scores.sort_values(by = 'Score')
```

```
X= df2[['SEP', 'JUN', 'JUL']] the top 3 features
Y= df2[['Sem_Pen_Amt']] the target output
```

```
X_train,X_test,y_train,y_test=train_test_split(X,Y,test_size=0.4,random_state=100)
```

```
logreg= LogisticRegression()
logreg.fit(X_train,y_train)
```

```
y_pred=logreg.predict(X_test)
```

```
y_pred_proba= logreg.predict_proba(X_test)[:,1]
```

```
false_positive_rate, true_positive_rate, _ = metrics.roc_curve(y_test, y_pred_proba)
```

```
auc= metrics.roc_auc_score(y_test, y_pred_proba)
```

Logbook

Week 10:

Researching the libraries to use for this assignment while working on other projects

Week 11:

Little work towards this project

Week 12:

Starting to program and finding dataset

Week 13:

Most of all work done programming database and starting predictive algorithm.