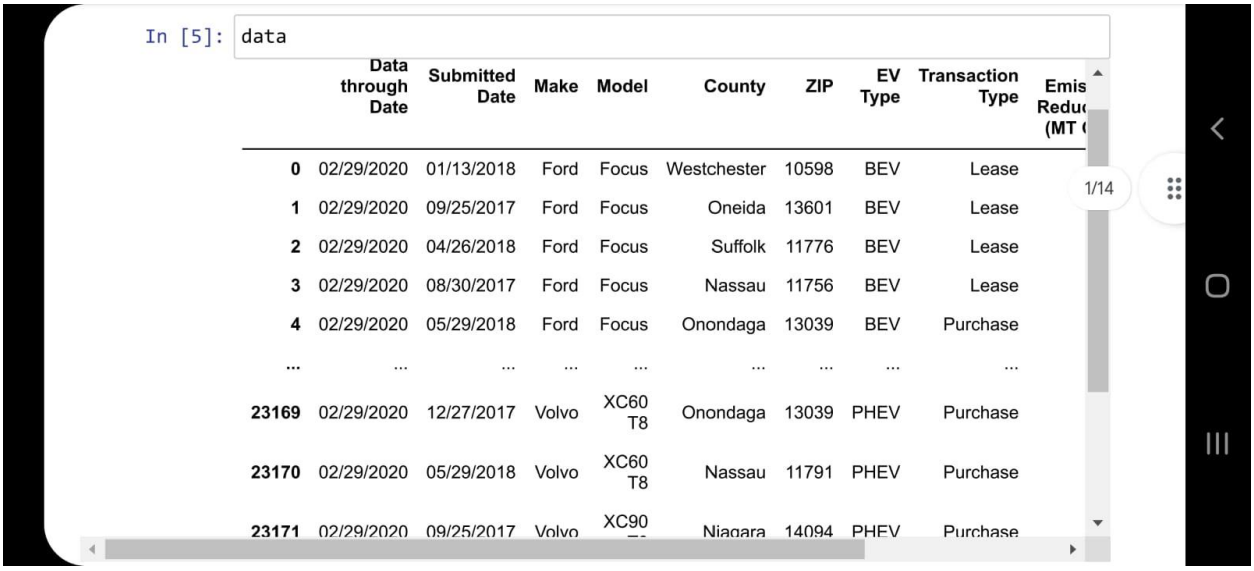
EV Report

Anusha Narayanan



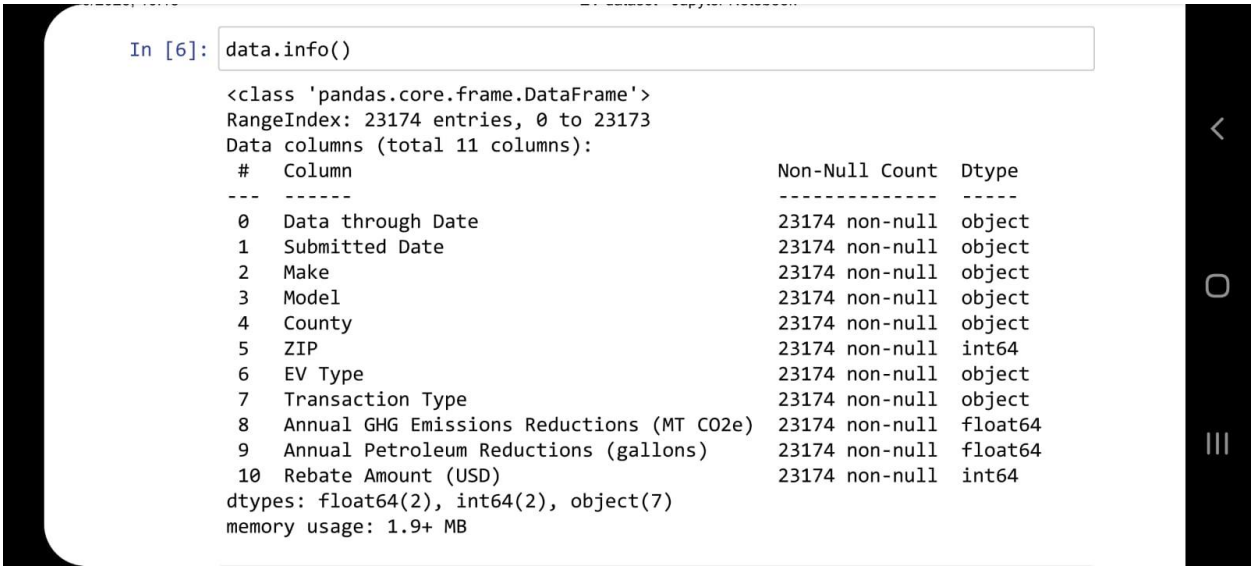
In [5]: data

	Data through Date	Submitted Date	Make	Model	County	ZIP	EV Type	Transaction Type	Emis Reduc (MT
0	02/29/2020	01/13/2018	Ford	Focus	Westchester	10598	BEV	Lease	
1	02/29/2020	09/25/2017	Ford	Focus	Oneida	13601	BEV	Lease	
2	02/29/2020	04/26/2018	Ford	Focus	Suffolk	11776	BEV	Lease	
3	02/29/2020	08/30/2017	Ford	Focus	Nassau	11756	BEV	Lease	
4	02/29/2020	05/29/2018	Ford	Focus	Onondaga	13039	BEV	Purchase	
...	...	...	...	...	...	...	...	...	
23169	02/29/2020	12/27/2017	Volvo	XC60 T8	Onondaga	13039	PHEV	Purchase	
23170	02/29/2020	05/29/2018	Volvo	XC60 T8	Nassau	11791	PHEV	Purchase	
23171	02/29/2020	09/25/2017	Volvo	XC90	Niagara	14094	PHEV	Purchase	

1/14

Observations

1. Date should be converted.
2. No need of County Column
3. Finding difference between ‘Data through Date’ and ‘Submitted Date ’ to know how much time is taken to identify the details.



In [6]: data.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 23174 entries, 0 to 23173
Data columns (total 11 columns):
 #   Column                                  Non-Null Count  Dtype
---  ------                                  --------------  -----
 0   Data through Date                       23174 non-null  object
 1   Submitted Date                          23174 non-null  object
 2   Make                                    23174 non-null  object
 3   Model                                   23174 non-null  object
 4   County                                  23174 non-null  object
 5   ZIP                                     23174 non-null  int64
 6   EV Type                                 23174 non-null  object
 7   Transaction Type                        23174 non-null  object
 8   Annual GHG Emissions Reductions (MT CO2e)  23174 non-null  float64
 9   Annual Petroleum Reductions (gallons)   23174 non-null  float64
 10  Rebate Amount (USD)                     23174 non-null  int64
dtypes: float64(2), int64(2), object(7)
memory usage: 1.9+ MB
```

Observations

1. As per info() there is no any Nan values in any rows

```
In [18]: data_1
```

Out[18]:

	Annual GHG Emissions Reductions (MT CO2e)	Annual Petroleum Reductions (gallons)	Rebate Amount (USD)	Difference	Make_BMW	Make_Chevrolet	Make_Chry
0	2.76	592.89	1700	777	0	0	
1	2.76	592.89	1700	887	0	0	
2	2.76	592.89	1700	674	0	0	
3	2.76	592.89	1700	913	0	0	
4	2.76	592.89	1700	641	0	0	
...	...	...	...	...	...	...	
23169	0.40	200.58	500	794	0	0	
23170	0.40	200.58	500	641	0	0	
23171	0.68	231.77	500	887	0	0	

Observations

1. Since computer understands no.s .Every Categorical values are converted through dummies

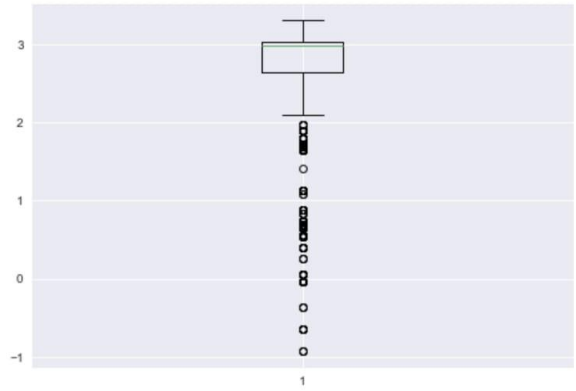
```
In [19]: data_1.describe()
```

Out[19]:

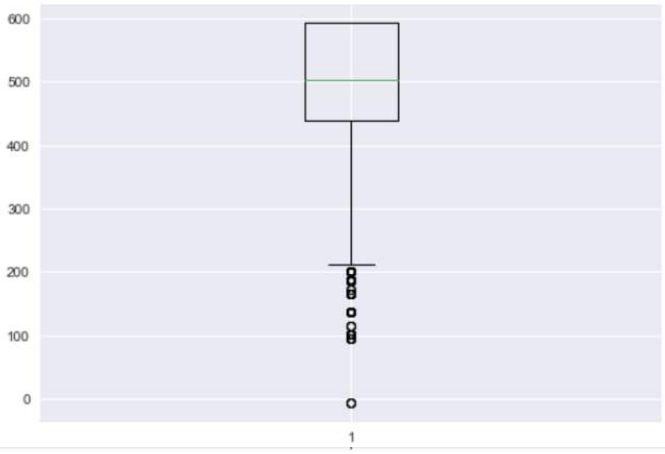
	Annual GHG Emissions Reductions (MT CO2e)	Annual Petroleum Reductions (gallons)	Rebate Amount (USD)	Difference	Make_BMW	Make_Chevrole
count	23174.000000	23174.000000	23174.000000	23174.000000	23174.000000	23174.00000
mean	2.692061	488.940697	1408.893588	477.686243	0.047510	0.13398
std	0.641441	108.241707	494.254611	303.124878	0.212732	0.34064
min	-0.930000	-7.110000	500.000000	0.000000	0.000000	0.00000
25%	2.650000	440.110000	1100.000000	204.000000	0.000000	0.00000
50%	2.990000	503.600000	1100.000000	460.000000	0.000000	0.00000
75%	3.030000	592.890000	2000.000000	722.000000	0.000000	0.00000
max	3.310000	592.890000	2000.000000	1072.000000	1.000000	1.00000

8 rows × 130 columns

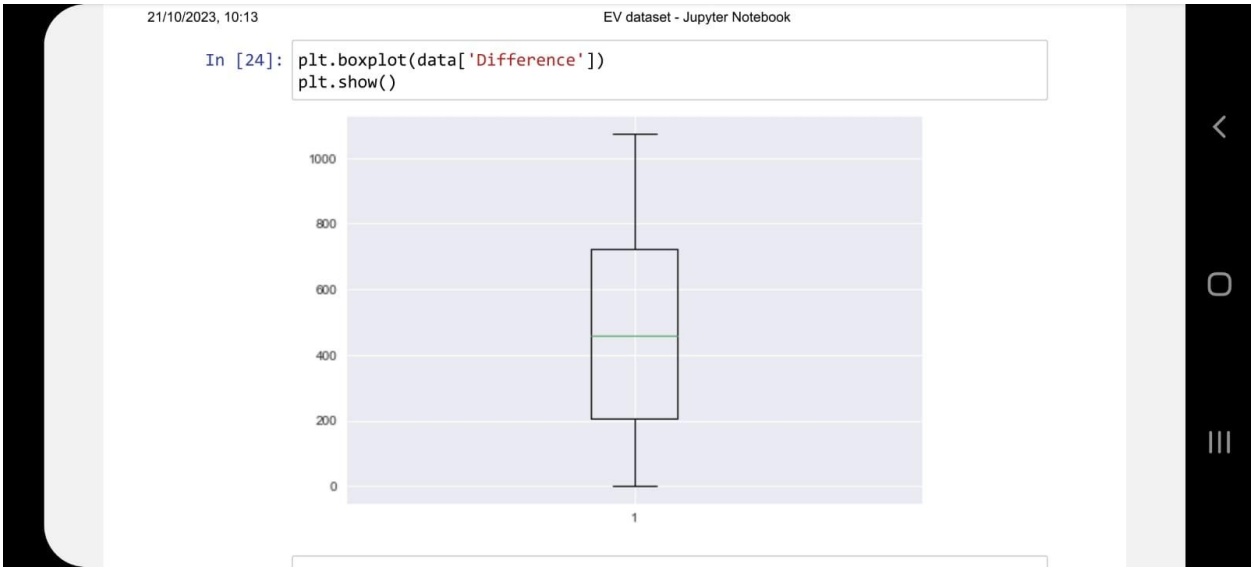
In [21]:
```
plt.boxplot(data['Annual GHG Emissions Reductions (MT CO2e)'])
plt.show()
```



In [22]:
```
plt.boxplot(data['Annual Petroleum Reductions (gallons)'])
plt.show()
```



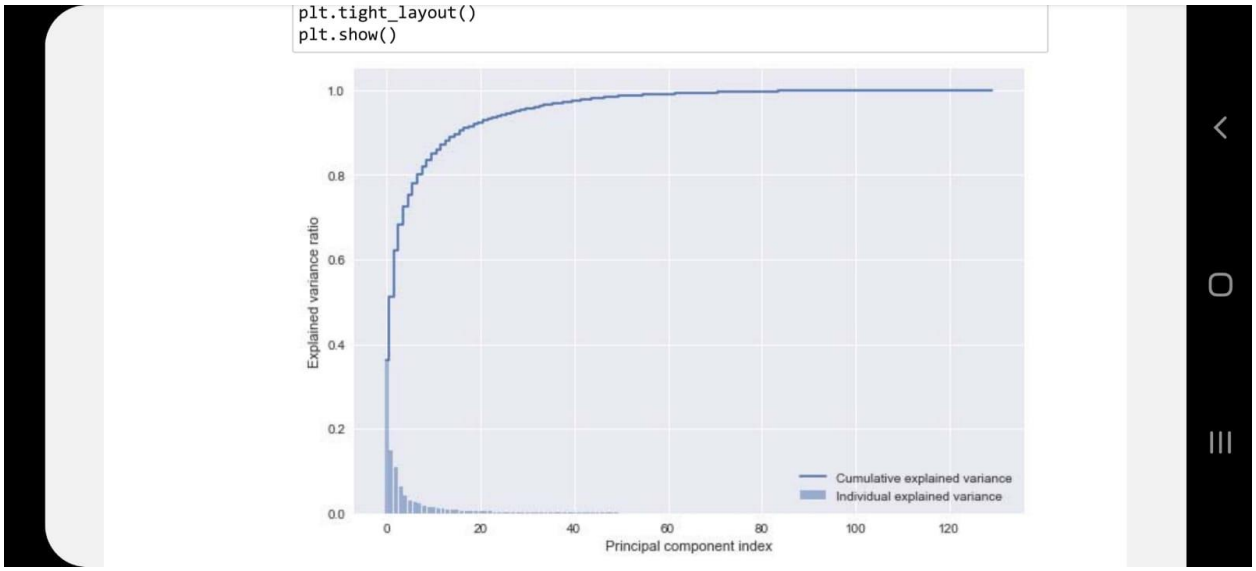
In [23]:
```
plt.boxplot(data['Rebate Amount (USD)'])
plt.show()
```



Observations

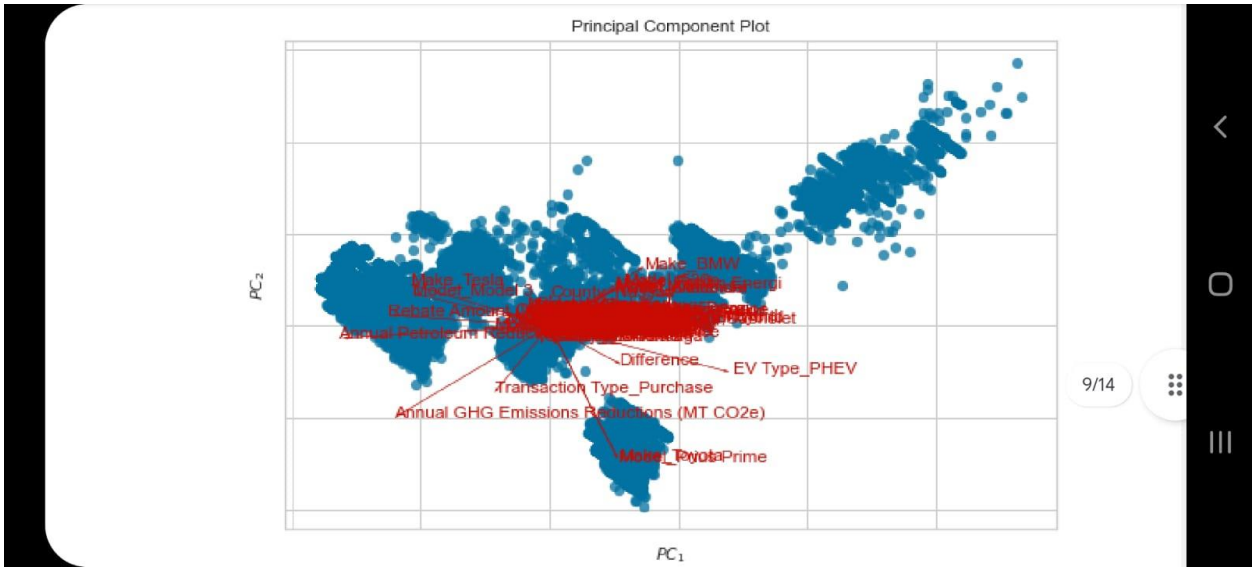
1. As per 'Annual GHG Emissions Reductions (MT CO2e)' boxplot there are many outliers in which it should be removed from the data set and it is not symmetric
2. As it can be seen from the output generated by describe command, One quarter of EV are less than 2.65 ; half of the EV are lessthan 2.99; and three quarters of EV are less than 3.03 . The oldest EVis 3.31 .It is astonishing EV less than 2, or has made a mistake when completing the data.
3. As per 'Annual Petroleum Reductions (gallons)' boxplot there are many outliers in which it should be removed from the data set and it is not symmetric
4. As it can be seen from the output generated by describe command, One quarter of EV are less than 440.11 ; half of the EV are less than 503.6 ; and three quarters of EV are less than 592.89 . The oldest EV is 592.89 .It is astonishing EV less than 210, or has made a mistake when completing the data.
5. As per 'Rebate Amount (USD)' boxplot ,there are no outliers. As it can be seen from the output generated by describe command,

One quarter of EV are less than 1100 ; half of the EV are less than 1100 ; and three quarters of EV are less than 2000 . The oldest EV is 2000 .



Observations

1. The latter two are the most important pieces of information. Principal component 1 explains about one fifth (18%) of the variance of the original data; principal component 2 about one tenth (16%). Together, they explain 24% of the variation in the original data. Principal components 3 to 130 explain of the original variation.


Principal Component Plot
Make_BMW
Difference
EV Type_PHEV
Transaction Type_Purchase
Annual GHG Emissions Reductions (MT CO2e)
Model_Prime
PC_2
PC_1