

police-dataset

January 30, 2024

0.1 Working on Real Project with Python

0.1.1 Police Dataset

Here, The data from Police Check Post is Given. This data is available as a CSV file. We are going to analyze this data set using the Pandas Data Frame.

```
[3]: import pandas as pd
```

```
[4]: data = pd.read_csv("D:/data analystics/Python for data analytics/Python_
↳Projects/Police DataSet/file.csv")
```

0.1.2 Question.1. Remove the column that only contains missing values

```
[5]: data.isnull().sum()
```

```
[5]: stop_date          0
stop_time             0
country_name        65535
driver_gender        4061
driver_age_raw       4054
driver_age           4307
driver_race          4060
violation_raw        4060
violation            4060
search_conducted      0
search_type          63056
stop_outcome         4060
is_arrested          4060
stop_duration        4060
drugs_related_stop    0
dtype: int64
```

```
[6]: data.drop( columns= 'country_name', inplace= True)
```

0.1.3 Question.2. For Speeding, were Men or Women stopped more Often?

```
[7]: data[data.violation == 'Speeding'].driver_gender.value_counts()
```

```
[7]: driver_gender
M    25517
F     11686
Name: count, dtype: int64
```

0.1.4 Question.3. Does gender affect who gets searched during stop?

```
[8]: data.groupby('driver_gender').search_conducted.sum()
```

```
[8]: driver_gender
F      366
M     2113
Name: search_conducted, dtype: int64
```

```
[10]: data.search_conducted.value_counts()
```

```
[10]: search_conducted
False    63056
True      2479
Name: count, dtype: int64
```

0.1.5 Question.4. What is mean stop_duration?

```
[12]: data.head()
```

```
[12]:   stop_date stop_time driver_gender  driver_age_raw  driver_age  driver_race \
0   1/2/2005    1:55             M         1985.0        20.0        White
1  1/18/2005    8:15             M         1965.0        40.0        White
2  1/23/2005   23:15             M         1972.0        33.0        White
3  2/20/2005   17:15             M         1986.0        19.0        White
4  3/14/2005   10:00             F         1984.0        21.0        White

      violation_raw violation  search_conducted  search_type  stop_outcome \
0           Speeding  Speeding             False         NaN      Citation
1           Speeding  Speeding             False         NaN      Citation
2           Speeding  Speeding             False         NaN      Citation
3  Call for Service    Other             False         NaN  Arrest Driver
4           Speeding  Speeding             False         NaN      Citation

      is_arrested stop_duration  drugs_related_stop
0           False    0-15 Min                False
1           False    0-15 Min                False
2           False    0-15 Min                False
```

3	True	16-30 Min	False
4	False	0-15 Min	False

```
[13]: data.stop_duration.value_counts()
```

```
[13]: stop_duration
0-15 Min      47379
16-30 Min     11448
30+ Min       2647
2              1
Name: count, dtype: int64
```

```
[14]: data['stop_duration'] = data['stop_duration'].map({'0-15 Min': 7.5, '16-30 Min': 24, '30+ Min': 45 })
```

```
[15]: data.head()
```

```
[15]:   stop_date stop_time driver_gender driver_age_raw driver_age driver_race \
0    1/2/2005      1:55             M          1985.0        20.0      White
1    1/18/2005      8:15             M          1965.0        40.0      White
2    1/23/2005     23:15             M          1972.0        33.0      White
3    2/20/2005     17:15             M          1986.0        19.0      White
4    3/14/2005     10:00             F          1984.0        21.0      White
```

	violation_raw	violation	search_conducted	search_type	stop_outcome	\
0	Speeding	Speeding	False	NaN	Citation	
1	Speeding	Speeding	False	NaN	Citation	
2	Speeding	Speeding	False	NaN	Citation	
3	Call for Service	Other	False	NaN	Arrest Driver	
4	Speeding	Speeding	False	NaN	Citation	

	is_arrested	stop_duration	drugs_related_stop
0	False	7.5	False
1	False	7.5	False
2	False	7.5	False
3	True	24.0	False
4	False	7.5	False

```
[16]: rounded_mean = round(data['stop_duration'].mean(), 2) # Round to 2 decimal places
print("mean of Stop Duration", rounded_mean)
```

mean of Stop Duration 12.19

0.1.6 Question.5. Compare the age distribution for each violation

```
[17]: data.groupby('violation').driver_age.describe()
```

```
[17]:
```

	count	mean	std	min	25%	50%	75%	\
violation								
Equipment	6507.0	31.682957	11.380671	16.0	23.0	28.0	39.0	
Moving violation	11876.0	36.736443	13.258350	15.0	25.0	35.0	47.0	
Other	3477.0	40.362381	12.754423	16.0	30.0	41.0	50.0	
Registration/plates	2240.0	32.656696	11.150780	16.0	24.0	30.0	40.0	
Seat belt	3.0	30.333333	10.214369	23.0	24.5	26.0	34.0	
Speeding	37120.0	33.262581	12.615781	15.0	23.0	30.0	42.0	

	max
violation	
Equipment	81.0
Moving violation	86.0
Other	86.0
Registration/plates	74.0
Seat belt	42.0
Speeding	88.0