Seatwork 6.1 Exploratory Data Analysis on Your Own Dataset

Detchosa, Ralph Christian D.
CPE22S3

Computational Thinking with Python
Dr. Roman Richard

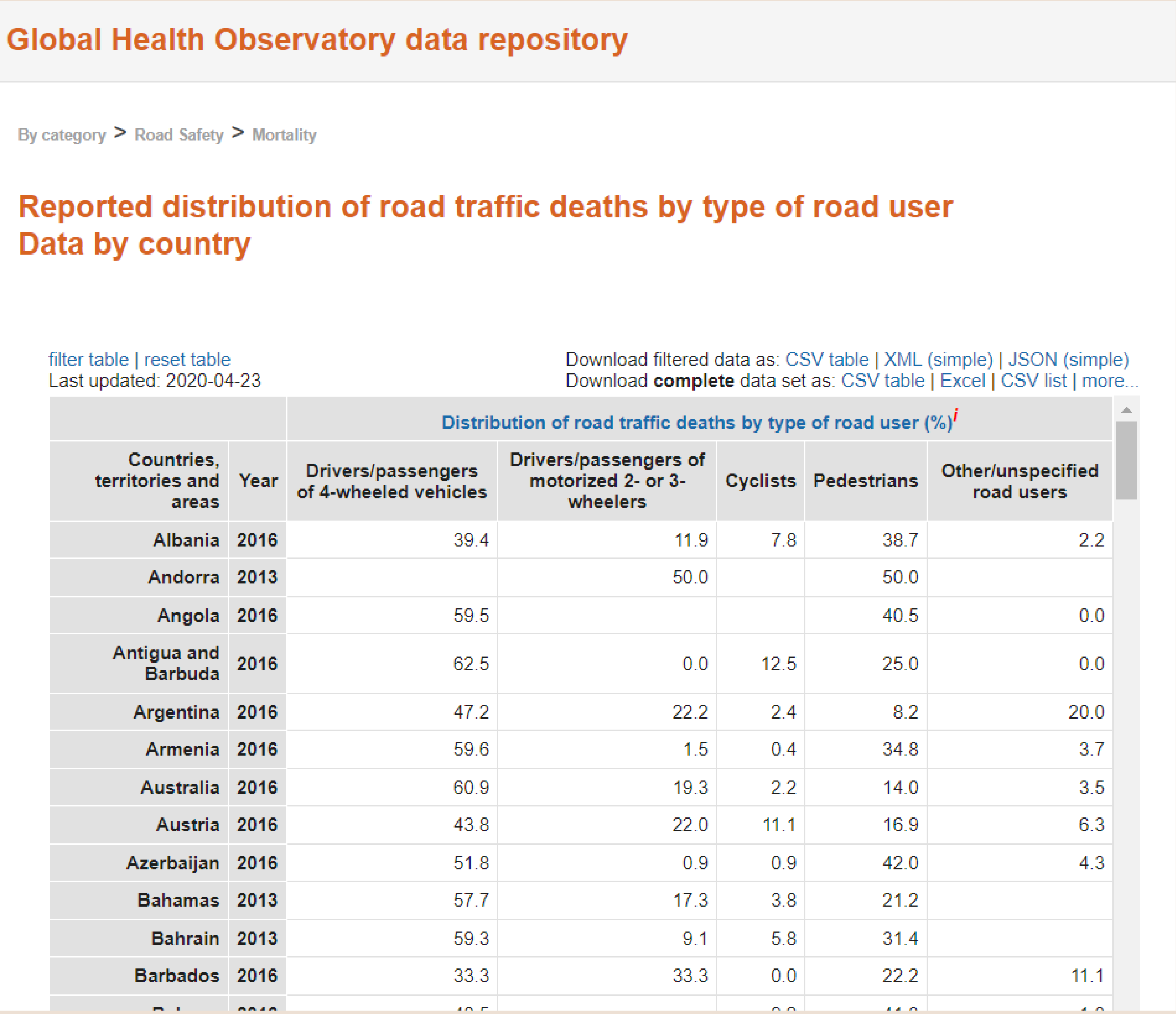
For all purpose

For Presenration

Data Set: Distribution of road traffic deaths by type of road user (%)

Country-collected data on road traffic deaths broken down by road user groups.

Data were collected from several different sectors and stakeholders in each country and were submitted to the World Health Organization.



Global Health Observatory data repository

By category > Road Safety > Mortality

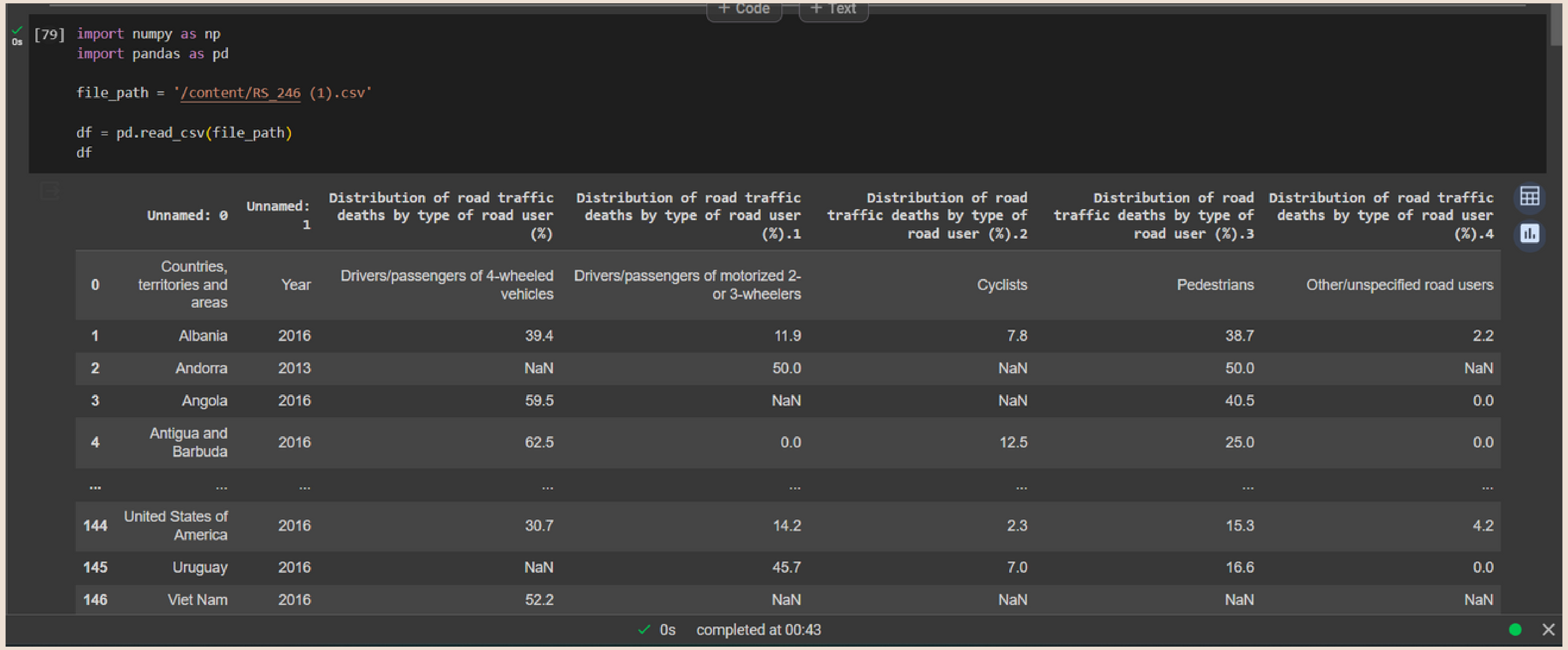
Reported distribution of road traffic deaths by type of road user
Data by country

filter table | reset table
Last updated: 2020-04-23

Download filtered data as: CSV table | XML (simple) | JSON (simple)
Download **complete** data set as: CSV table | Excel | CSV list | more...

		Distribution of road traffic deaths by type of road user (%)[i]				
Countries, territories and areas	**Year**	**Drivers/passengers of 4-wheeled vehicles**	**Drivers/passengers of motorized 2- or 3-wheelers**	**Cyclists**	**Pedestrians**	**Other/unspecified road users**
Albania	**2016**	39.4	11.9	7.8	38.7	2.2
Andorra	**2013**		50.0		50.0	
Angola	**2016**	59.5			40.5	0.0
Antigua and Barbuda	**2016**	62.5	0.0	12.5	25.0	0.0
Argentina	**2016**	47.2	22.2	2.4	8.2	20.0
Armenia	**2016**	59.6	1.5	0.4	34.8	3.7
Australia	**2016**	60.9	19.3	2.2	14.0	3.5
Austria	**2016**	43.8	22.0	11.1	16.9	6.3
Azerbaijan	**2016**	51.8	0.9	0.9	42.0	4.3
Bahamas	**2013**	57.7	17.3	3.8	21.2	
Bahrain	**2013**	59.3	9.1	5.8	31.4	
Barbados	**2016**	33.3	33.3	0.0	22.2	11.1

Reformat



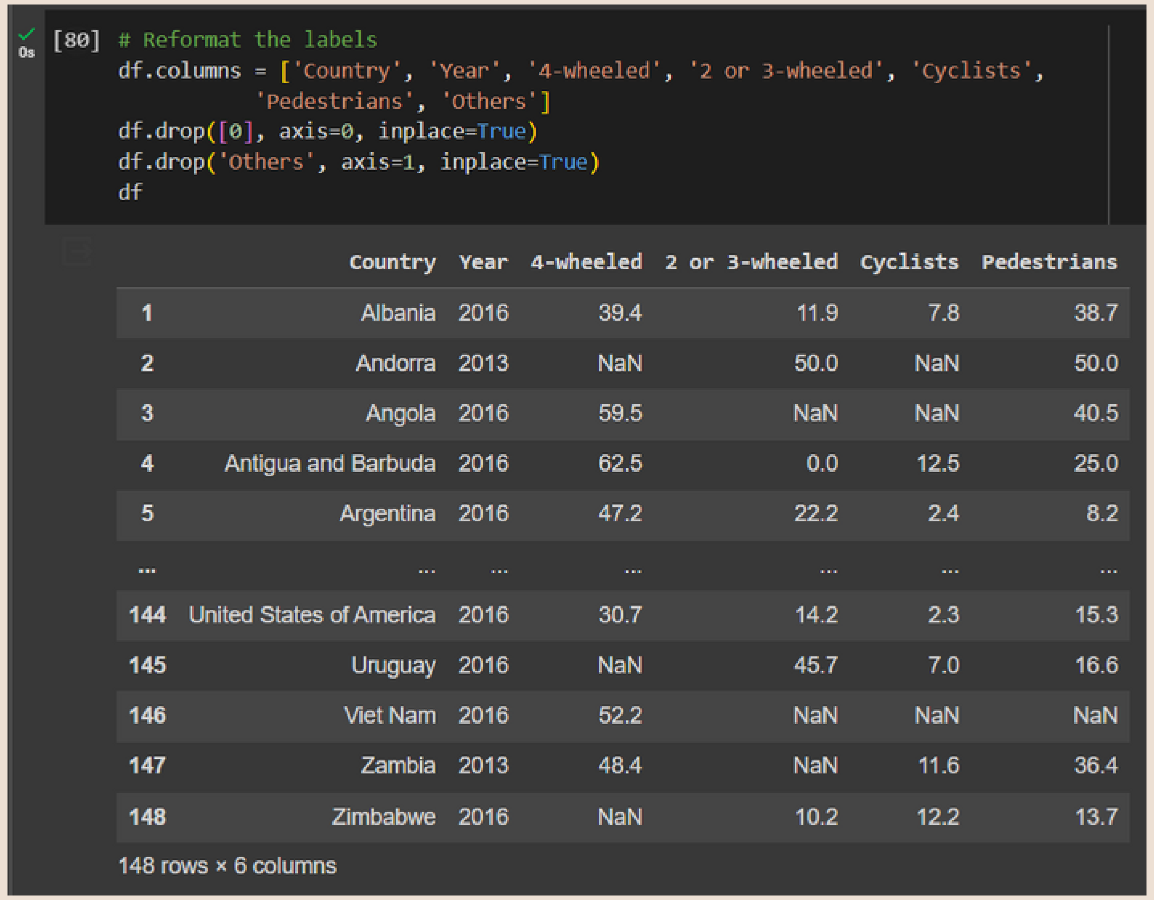
```
[79] import numpy as np
     import pandas as pd

     file_path = '/content/RS_246 (1).csv'

     df = pd.read_csv(file_path)
     df
```

	Unnamed: 0	Unnamed: 1	Distribution of road traffic deaths by type of road user (%)	Distribution of road traffic deaths by type of road user (%).1	Distribution of road traffic deaths by type of road user (%).2	Distribution of road traffic deaths by type of road user (%).3	Distribution of road traffic deaths by type of road user (%).4
0	Countries, territories and areas	Year	Drivers/passengers of 4-wheeled vehicles	Drivers/passengers of motorized 2- or 3-wheelers	Cyclists	Pedestrians	Other/unspecified road users
1	Albania	2016	39.4	11.9	7.8	38.7	2.2
2	Andorra	2013	NaN	50.0	NaN	50.0	NaN
3	Angola	2016	59.5	NaN	NaN	40.5	0.0
4	Antigua and Barbuda	2016	62.5	0.0	12.5	25.0	0.0
...	...	...	...	...	...	...	...
144	United States of America	2016	30.7	14.2	2.3	15.3	4.2
145	Uruguay	2016	NaN	45.7	7.0	16.6	0.0
146	Viet Nam	2016	52.2	NaN	NaN	NaN	NaN

✓ 0s completed at 00:43

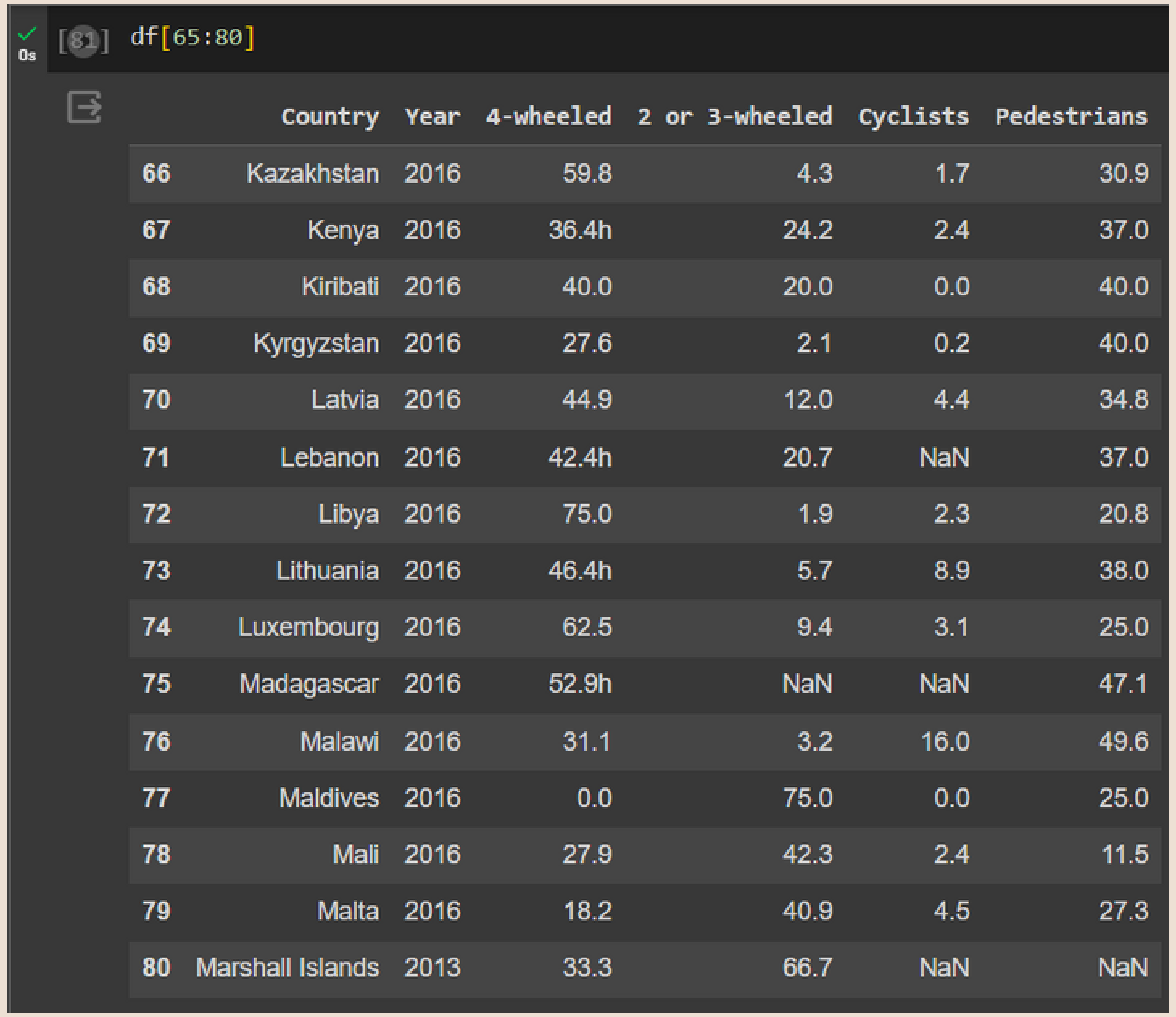


```
[80] # Reformat the labels
     df.columns = ['Country', 'Year', '4-wheeled', '2 or 3-wheeled', 'Cyclists',
                   'Pedestrians', 'Others']
     df.drop([0], axis=0, inplace=True)
     df.drop('Others', axis=1, inplace=True)
     df
```

	Country	Year	4-wheeled	2 or 3-wheeled	Cyclists	Pedestrians
1	Albania	2016	39.4	11.9	7.8	38.7
2	Andorra	2013	NaN	50.0	NaN	50.0
3	Angola	2016	59.5	NaN	NaN	40.5
4	Antigua and Barbuda	2016	62.5	0.0	12.5	25.0
5	Argentina	2016	47.2	22.2	2.4	8.2
...	...	...	...	...	...	...
144	United States of America	2016	30.7	14.2	2.3	15.3
145	Uruguay	2016	NaN	45.7	7.0	16.6
146	Viet Nam	2016	52.2	NaN	NaN	NaN
147	Zambia	2013	48.4	NaN	11.6	36.4
148	Zimbabwe	2016	NaN	10.2	12.2	13.7

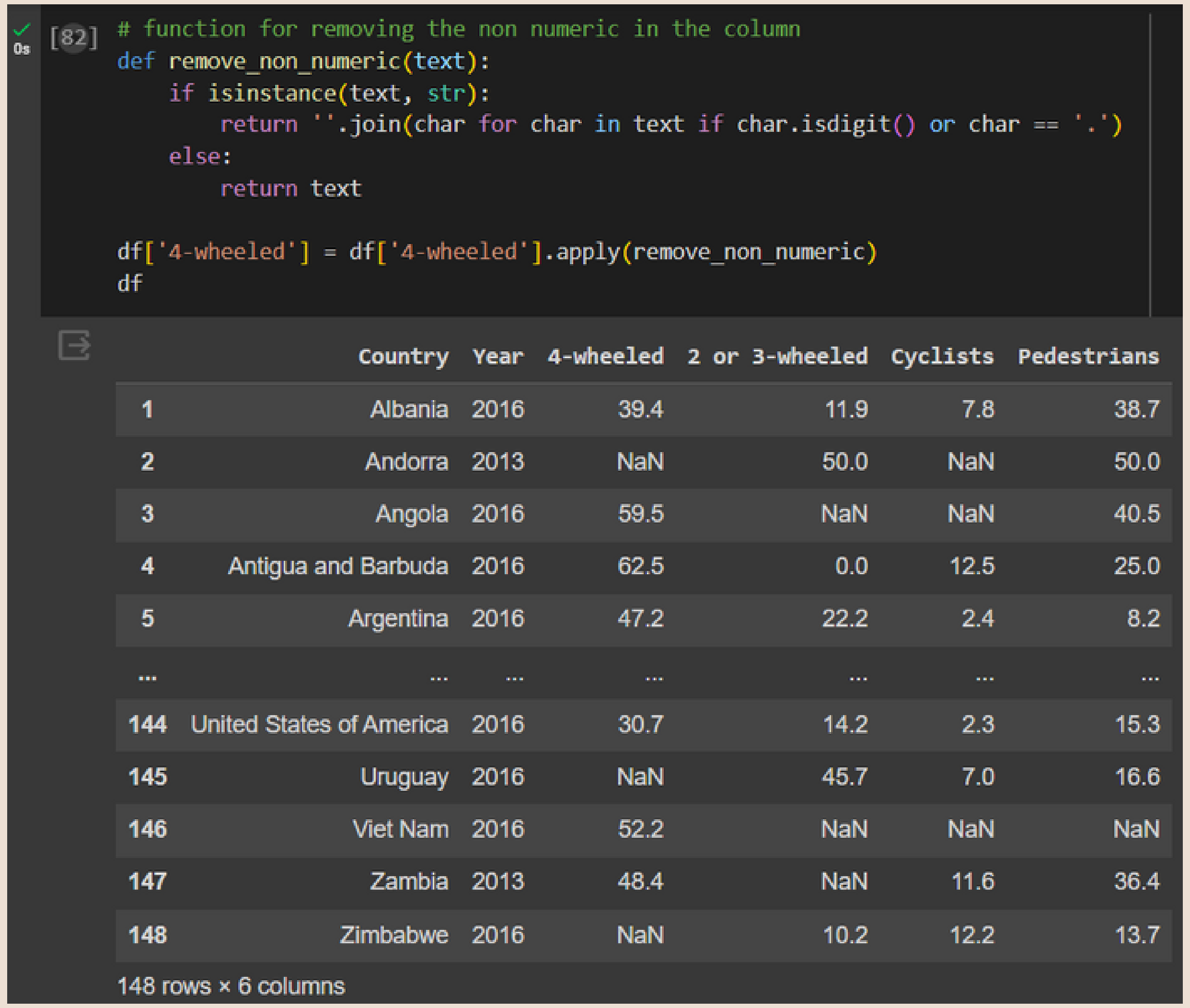
148 rows × 6 columns

Mean



[81] df[65:80]

	Country	Year	4-wheeled	2 or 3-wheeled	Cyclists	Pedestrians
66	Kazakhstan	2016	59.8	4.3	1.7	30.9
67	Kenya	2016	36.4h	24.2	2.4	37.0
68	Kiribati	2016	40.0	20.0	0.0	40.0
69	Kyrgyzstan	2016	27.6	2.1	0.2	40.0
70	Latvia	2016	44.9	12.0	4.4	34.8
71	Lebanon	2016	42.4h	20.7	NaN	37.0
72	Libya	2016	75.0	1.9	2.3	20.8
73	Lithuania	2016	46.4h	5.7	8.9	38.0
74	Luxembourg	2016	62.5	9.4	3.1	25.0
75	Madagascar	2016	52.9h	NaN	NaN	47.1
76	Malawi	2016	31.1	3.2	16.0	49.6
77	Maldives	2016	0.0	75.0	0.0	25.0
78	Mali	2016	27.9	42.3	2.4	11.5
79	Malta	2016	18.2	40.9	4.5	27.3
80	Marshall Islands	2013	33.3	66.7	NaN	NaN



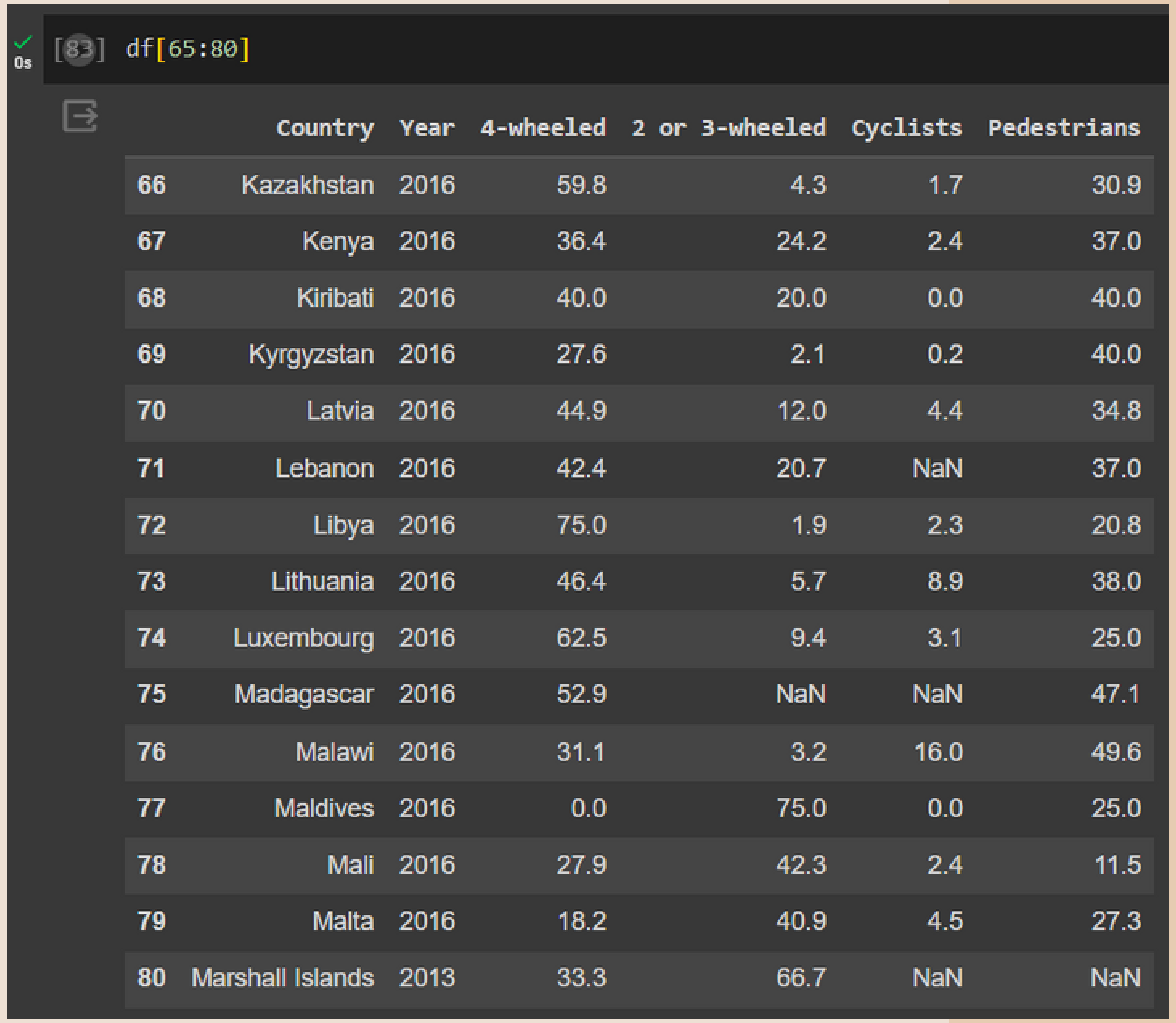
[82]

```
# function for removing the non numeric in the column
def remove_non_numeric(text):
    if isinstance(text, str):
        return ''.join(char for char in text if char.isdigit() or char == '.')
    else:
        return text

df['4-wheeled'] = df['4-wheeled'].apply(remove_non_numeric)
df
```

	Country	Year	4-wheeled	2 or 3-wheeled	Cyclists	Pedestrians
1	Albania	2016	39.4	11.9	7.8	38.7
2	Andorra	2013	NaN	50.0	NaN	50.0
3	Angola	2016	59.5	NaN	NaN	40.5
4	Antigua and Barbuda	2016	62.5	0.0	12.5	25.0
5	Argentina	2016	47.2	22.2	2.4	8.2
...	...	...	...	...	...	...
144	United States of America	2016	30.7	14.2	2.3	15.3
145	Uruguay	2016	NaN	45.7	7.0	16.6
146	Viet Nam	2016	52.2	NaN	NaN	NaN
147	Zambia	2013	48.4	NaN	11.6	36.4
148	Zimbabwe	2016	NaN	10.2	12.2	13.7

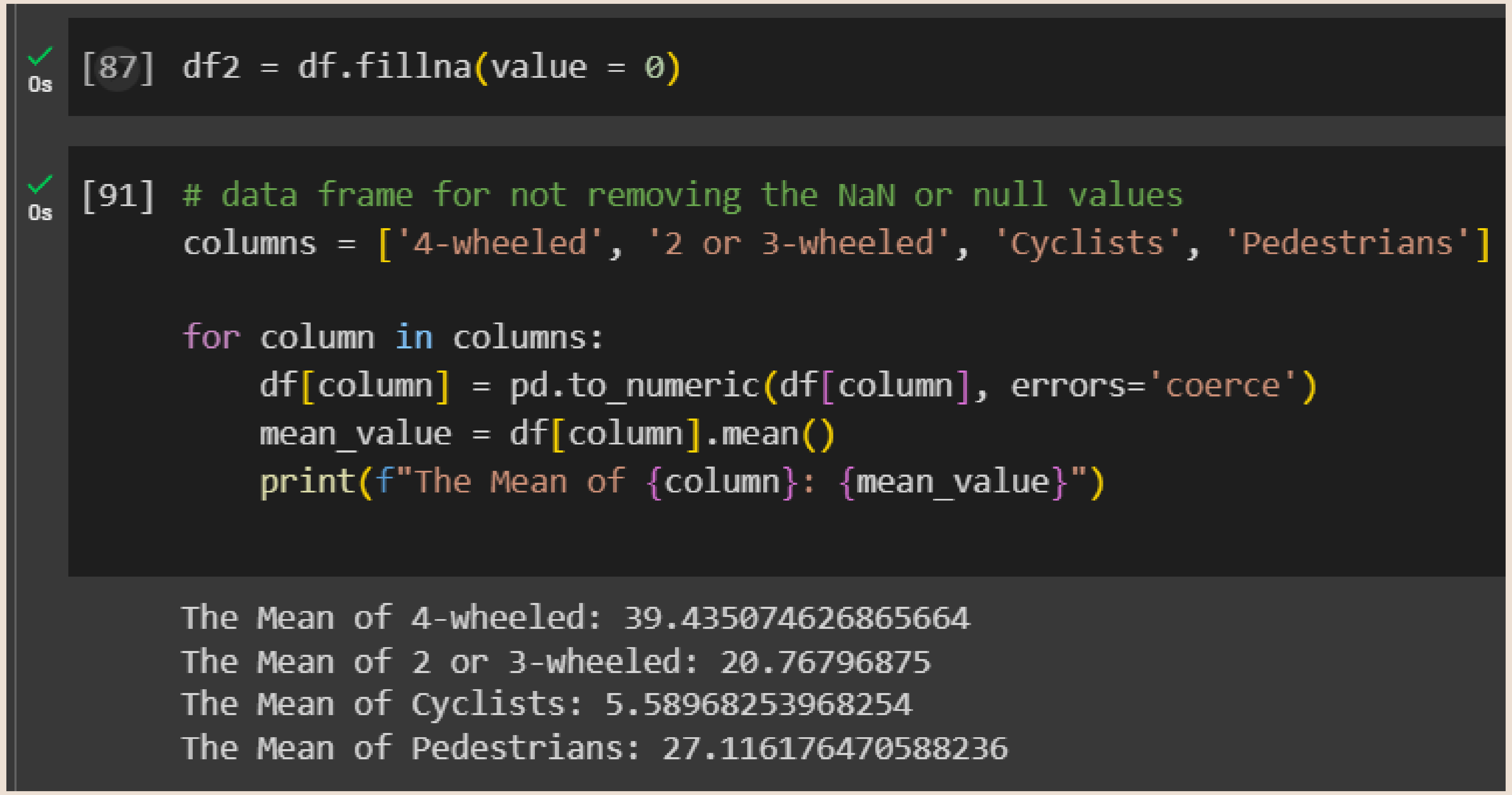
148 rows × 6 columns



[83] df[65:80]

	Country	Year	4-wheeled	2 or 3-wheeled	Cyclists	Pedestrians
66	Kazakhstan	2016	59.8	4.3	1.7	30.9
67	Kenya	2016	36.4	24.2	2.4	37.0
68	Kiribati	2016	40.0	20.0	0.0	40.0
69	Kyrgyzstan	2016	27.6	2.1	0.2	40.0
70	Latvia	2016	44.9	12.0	4.4	34.8
71	Lebanon	2016	42.4	20.7	NaN	37.0
72	Libya	2016	75.0	1.9	2.3	20.8
73	Lithuania	2016	46.4	5.7	8.9	38.0
74	Luxembourg	2016	62.5	9.4	3.1	25.0
75	Madagascar	2016	52.9	NaN	NaN	47.1
76	Malawi	2016	31.1	3.2	16.0	49.6
77	Maldives	2016	0.0	75.0	0.0	25.0
78	Mali	2016	27.9	42.3	2.4	11.5
79	Malta	2016	18.2	40.9	4.5	27.3
80	Marshall Islands	2013	33.3	66.7	NaN	NaN

Median, Variance, Standard Deviation

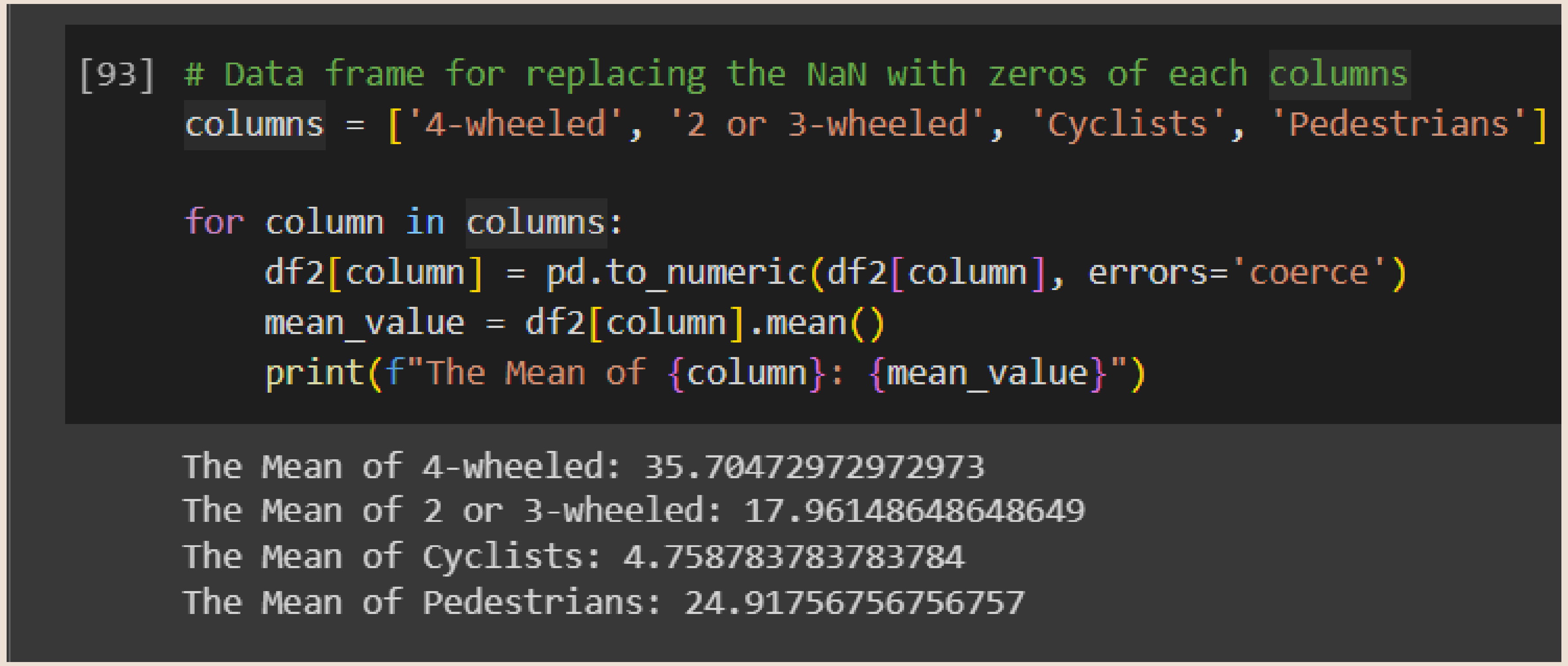


```
[87] df2 = df.fillna(value = 0)
```

```
[91] # data frame for not removing the NaN or null values
     columns = ['4-wheeled', '2 or 3-wheeled', 'Cyclists', 'Pedestrians']

     for column in columns:
         df[column] = pd.to_numeric(df[column], errors='coerce')
         mean_value = df[column].mean()
         print(f"The Mean of {column}: {mean_value}")
```

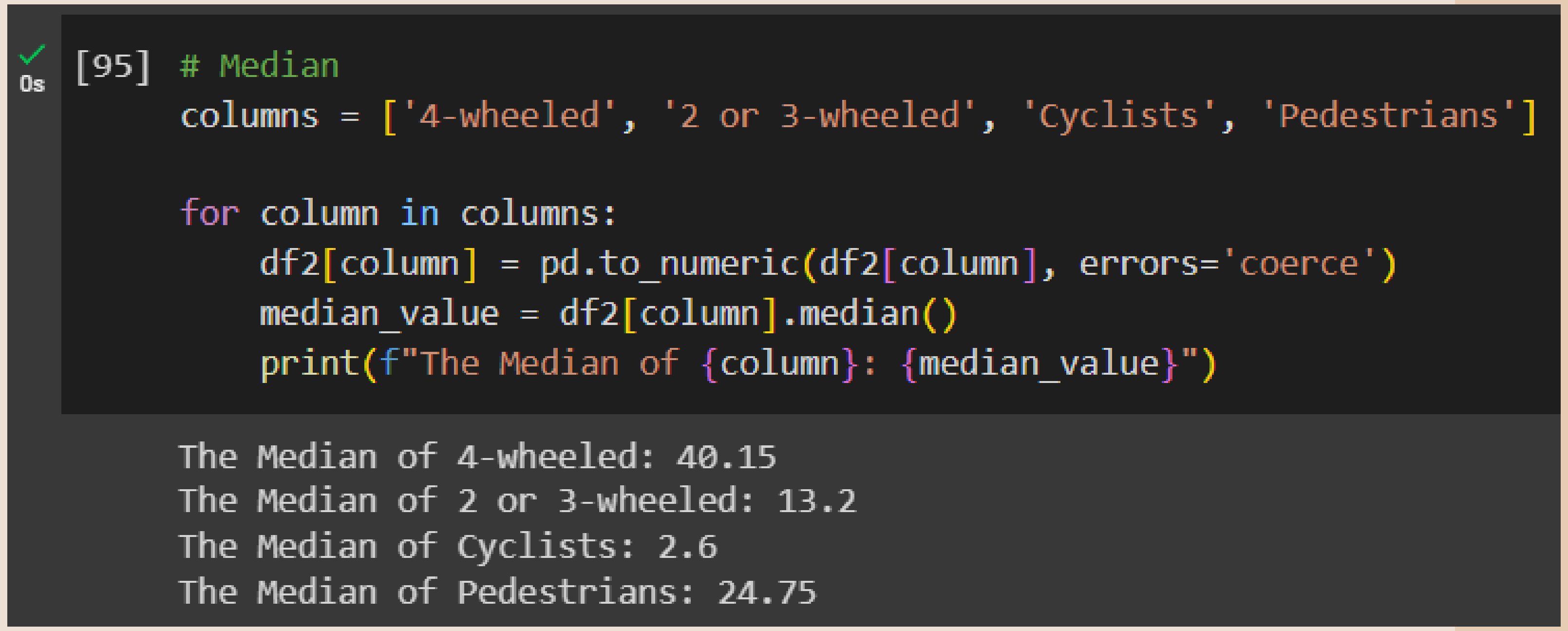
```
The Mean of 4-wheeled: 39.435074626865664
The Mean of 2 or 3-wheeled: 20.76796875
The Mean of Cyclists: 5.58968253968254
The Mean of Pedestrians: 27.116176470588236
```



```
[93] # Data frame for replacing the NaN with zeros of each columns
     columns = ['4-wheeled', '2 or 3-wheeled', 'Cyclists', 'Pedestrians']

     for column in columns:
         df2[column] = pd.to_numeric(df2[column], errors='coerce')
         mean_value = df2[column].mean()
         print(f"The Mean of {column}: {mean_value}")
```

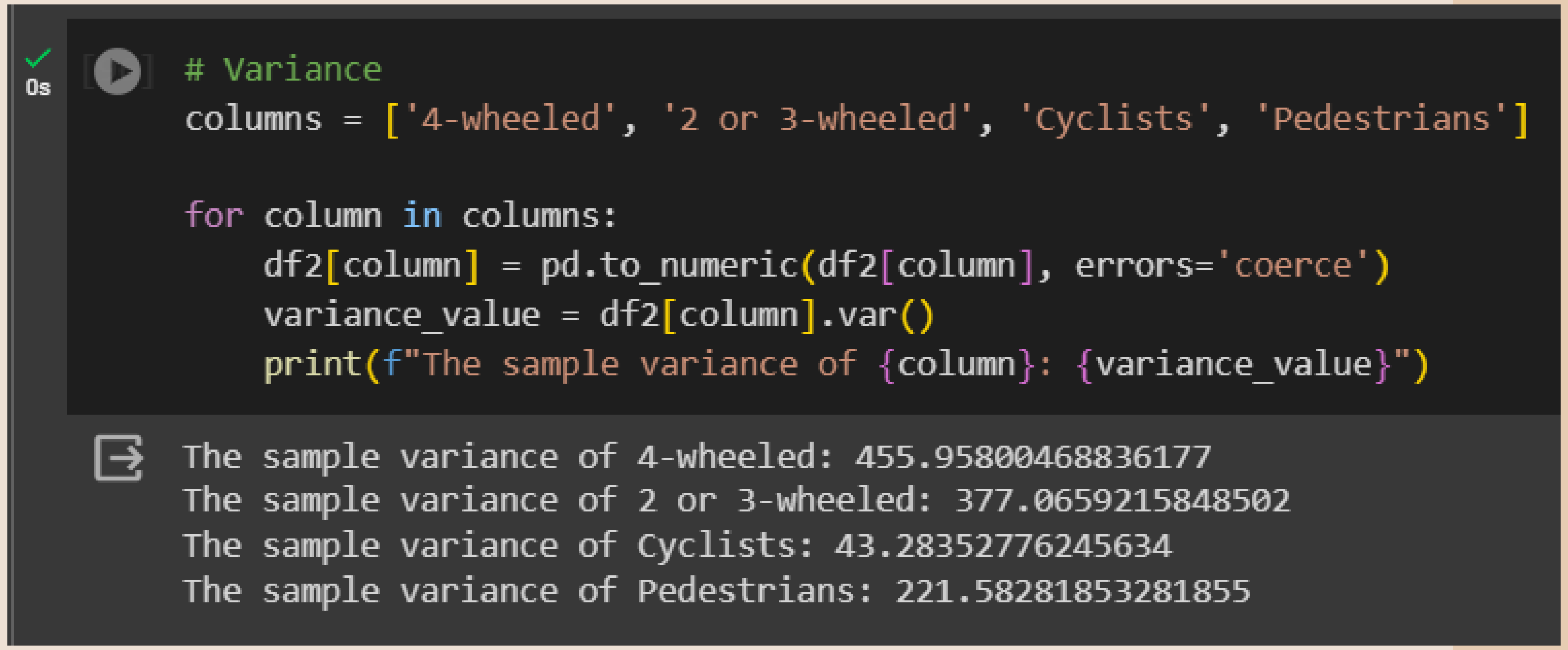
```
The Mean of 4-wheeled: 35.70472972972973
The Mean of 2 or 3-wheeled: 17.96148648648649
The Mean of Cyclists: 4.758783783783784
The Mean of Pedestrians: 24.91756756756757
```



```
[95] # Median
     columns = ['4-wheeled', '2 or 3-wheeled', 'Cyclists', 'Pedestrians']

     for column in columns:
         df2[column] = pd.to_numeric(df2[column], errors='coerce')
         median_value = df2[column].median()
         print(f"The Median of {column}: {median_value}")
```

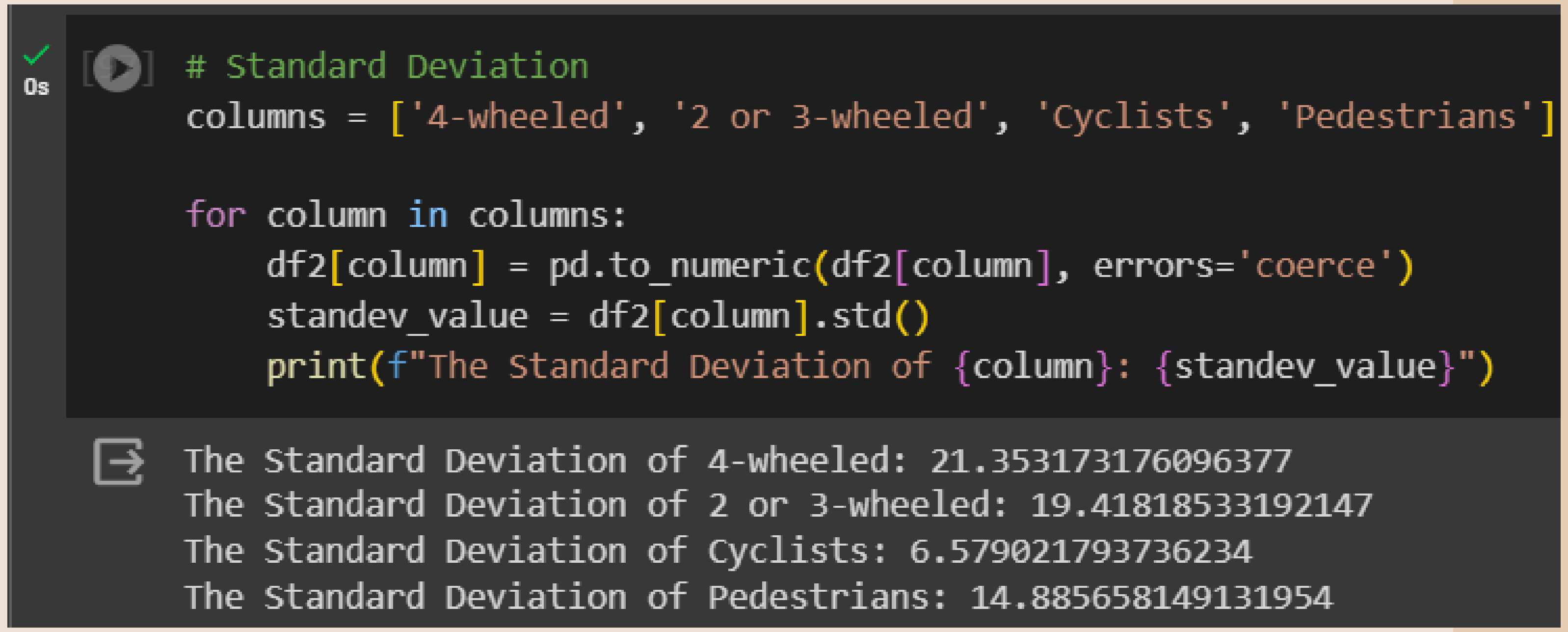
```
The Median of 4-wheeled: 40.15
The Median of 2 or 3-wheeled: 13.2
The Median of Cyclists: 2.6
The Median of Pedestrians: 24.75
```



```
# Variance
columns = ['4-wheeled', '2 or 3-wheeled', 'Cyclists', 'Pedestrians']

for column in columns:
    df2[column] = pd.to_numeric(df2[column], errors='coerce')
    variance_value = df2[column].var()
    print(f"The sample variance of {column}: {variance_value}")
```

```
The sample variance of 4-wheeled: 455.95800468836177
The sample variance of 2 or 3-wheeled: 377.0659215848502
The sample variance of Cyclists: 43.28352776245634
The sample variance of Pedestrians: 221.58281853281855
```

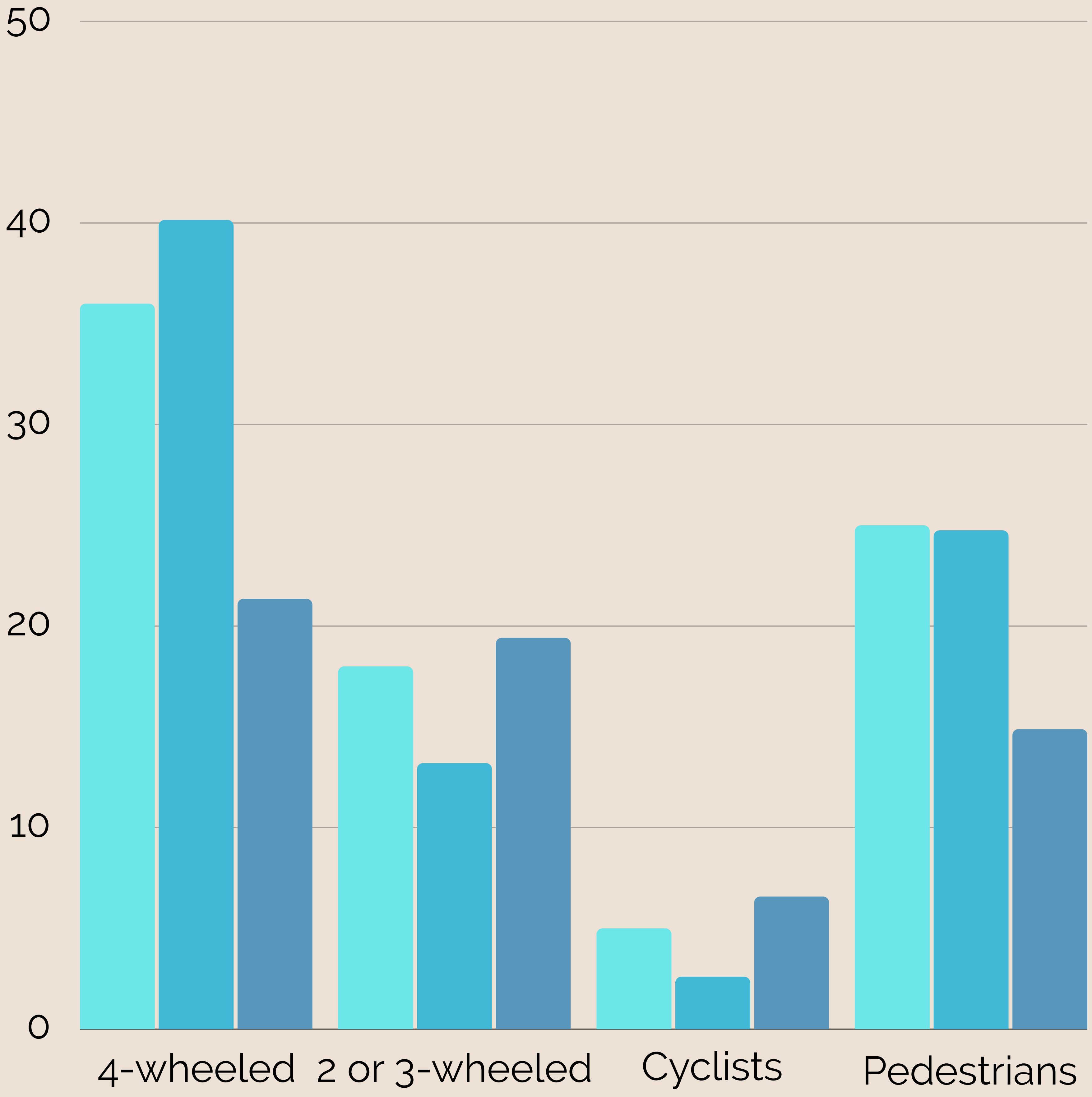


```
# Standard Deviation
columns = ['4-wheeled', '2 or 3-wheeled', 'Cyclists', 'Pedestrians']

for column in columns:
    df2[column] = pd.to_numeric(df2[column], errors='coerce')
    standev_value = df2[column].std()
    print(f"The Standard Deviation of {column}: {standev_value}")
```

```
The Standard Deviation of 4-wheeled: 21.353173176096377
The Standard Deviation of 2 or 3-wheeled: 19.41818533192147
The Standard Deviation of Cyclists: 6.579021793736234
The Standard Deviation of Pedestrians: 14.885658149131954
```

Data Analysis



- Variation in Road User Types
- Central Tendency
- Conclusion

Thank you for listening!

For all purpose

For Presenration