ROAD TRAFFIC ACCIDENTS

About Dataset

This data set is collected from Addis Ababa Sub-city police departments for Masters research work. The data set has been prepared from manual records of road traffic accidents.

```
# import numpy
import numpy as np
# import pandas
import pandas as pd
# import matplotlib
import matplotlib as mpl
# import the dataset
df_road_data_set = r'C:\Users\HP\Desktop\Big_Data_Analytics\Future_Studies_Projects_Dataset\RTA Dataset.csv'
```

```
df_road_data_set
```

	Time	Day_of_week	Age_band_of_driver	Sex_of_driver	Educational_level	Vehicle_driver_relation	Driving_experience	Type_of_vehicle	Owner_of_vehicle	Service_year_of_vehicle	Defect_of_vehicle	Area_accident_occured	Lanes_or_Medians	Road_surface_type	Road_surface_conditions
0	17:02:00	Monday	18-30	Male	Junior high school	Employee	1-2y	Automobile	Owner	Above 5y				Asphalt roads	
1	17:02:00	Monday	31-50	Male	Junior high school	Employee	Above 10y	Public (> 45 seats)	Owner	5-10y				Asphalt roads	
2	17:02:00	Monday	18-30	Male	Junior high school	Employee	5-10y	Lorry (41-100Q)	Owner					Asphalt roads	
3	18:00:00	Sunday	18-30	Male	Junior high school	Employee	2-5y	Public (13-45 seats)	Governmental					Asphalt roads	
4	13:00:00	Sunday	18-30	Male	Junior high school	Employee	2-5y	Taxi	Owner					Asphalt roads	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
12315	16:30:00	Wednesday	51-58	Male	-	Taxi	-	Pickup upto 10Q	Owner	-		-	-	Asphalt roads	-
12316	23:00:00	Friday	18-30	Male	Junior high school	Employee	2-5y	Lorry (41-100Q)	Owner	-		-	-	Asphalt roads	-
12317	8:00:00	Monday	Over 58	Male	Elementary school	Employee	5-10y	Automobile	Owner	-		-	-	Asphalt roads	-
12318	14:00:00	Tuesday	Over 58	Male	Junior high school	Employee	5-10y	Bajaj	Owner	-		-	-	Asphalt roads	-
12319	17:00:00	Friday	18-30	Male	Junior high school	Employee	Above 10y	Lorry (41-100Q)	Owner	-		-	-	Asphalt roads	-
12320	12:00:00	Sunday	18-30	Male	Junior high school	Employee	5-10y	Other	Owner	-		-	-	Asphalt roads	-

12316 rows × 32 columns

```
# head function
df.head()
```

	Time	Day_of_week	Age_band_of_driver	Sex_of_driver	Educational_level	Vehicle_driver_relation	Driving_experience	Type_of_vehicle	Owner_of_vehicle	Service_year_of_vehicle	Defect_of_vehicle	Area_accident_occured	Lanes_or_Medians	Road_surface_type	Road_surface_conditions
0	17:02:00	Monday	18-30	Male	Junior high school	Employee	1-2y	Automobile	Owner	Above 5y				Asphalt roads	
1	17:02:00	Monday	31-50	Male	Junior high school	Employee	Above 10y	Public (> 45 seats)	Owner	5-10y				Asphalt roads	
2	17:02:00	Monday	18-30	Male	Junior high school	Employee	5-10y	Lorry (41-100Q)	Owner					Asphalt roads	
3	18:00:00	Sunday	18-30	Male	Junior high school	Employee	2-5y	Public (13-45 seats)	Governmental					Asphalt roads	
4	13:00:00	Sunday	18-30	Male	Junior high school	Employee	2-5y	Taxi	Owner					Asphalt roads	

5 rows × 32 columns

```
# tail function
df.tail()
```

	Time	Day_of_week	Age_band_of_driver	Sex_of_driver	Educational_level	Vehicle_driver_relation	Driving_experience	Type_of_vehicle	Owner_of_vehicle	Service_year_of_vehicle	Defect_of_vehicle	Area_accident_occured	Lanes_or_Medians	Road_surface_type	Road_surface_conditions
12316	16:30:00	Wednesday	51-58	Male	-	Taxi	-	Pickup upto 10Q	Owner	-		-	-	Asphalt roads	-
12317	23:00:00	Friday	18-30	Male	Junior high school	Employee	2-5y	Lorry (41-100Q)	Owner	-		-	-	Asphalt roads	-
12318	8:00:00	Monday	Over 58	Male	Elementary school	Employee	5-10y	Automobile	Owner	-		-	-	Asphalt roads	-
12319	14:00:00	Tuesday	Over 58	Male	Junior high school	Employee	5-10y	Bajaj	Owner	-		-	-	Asphalt roads	-
12320	17:00:00	Friday	18-30	Male	Junior high school	Employee	Above 10y	Lorry (41-100Q)	Owner	-		-	-	Asphalt roads	-
12321	12:00:00	Sunday	18-30	Male	Junior high school	Employee	5-10y	Other	Owner	-		-	-	Asphalt roads	-

5 rows × 32 columns

```
# shape function
df.shape
```

```
(12316, 32)
```

```
# import the pyplot
import matplotlib.pyplot as plt
# import seaborn
import seaborn as sns
```

```
# describe function
df.describe()
```

	Time	Day_of_week	Age_band_of_driver	Sex_of_driver	Educational_level	Vehicle_driver_relation	Driving_experience	Type_of_vehicle	Owner_of_vehicle	Service_year_of_vehicle	Defect_of_vehicle	Area_accident_occured	Lanes_or_Medians	Road_surface_type	Road_surface_conditions
count	12316	12316	12316	12316	11575	11087	11048	11733	11621	10732	10161	12077	12030	12081	12064
unique	1141	7	5	2	7	5	8	17	4	6	2	14	8	5	5
top	16:30:00	Friday	18-30	Male	Junior high school	Employee	5-10y	Automobile	Owner	Unknown	Unknown	Other	Unknown	Asphalt roads	Dry
freq	187	1840	4907	11732	9486	9342	3299	3705	9783	3391	9591	3819	4878	9987	8897

```
# describe function
df.describe(include=['object'])
```

	Time	Day_of_week	Age_band_of_driver	Sex_of_driver	Educational_level	Vehicle_driver_relation	Driving_experience	Type_of_vehicle	Owner_of_vehicle	Service_year_of_vehicle	Defect_of_vehicle	Area_accident_occured	Lanes_or_Medians	Road_surface_type	Road_surface_conditions
count	12316	12316	12316	12316	11575	11087	11048	11733	11621	10732	10161	12077	12030	12081	12064
unique	1141	7	5	2	7	5	8	17	4	6	2	14	8	5	5
top	16:30:00	Friday	18-30	Male	Junior high school	Employee	5-10y	Automobile	Owner	Unknown	Unknown	Other	Unknown	Asphalt roads	Dry
freq	187	1840	4907	11732	9486	9342	3299	3705	9783	3391	9591	3819	4878	9987	8897

5 rows × 28 columns

```
# dtypes function
df.dtypes
```

```
Time                           object
Day_of_week                    object
Age_band_of_driver             object
Sex_of_driver                  object
Educational_level              object
Vehicle_driver_relation        object
Driving_experience             object
Type_of_vehicle                object
Owner_of_vehicle               object
Service_year_of_vehicle        object
Defect_of_vehicle              object
Area_accident_occured          object
Lanes_or_Medians               object
Road_surface_type              object
Road_surface_conditions        object
Light_conditions               object
Work_of_casualty               object
Weather_conditions             object
Type_of_collision              object
Number_of_vehicles_involved     int64
Number_of_casualties            int64
Vehicle_movement               object
Casualty_class                 object
Sex_of_casualty                object
Age_band_of_casualty           object
Casualty_severity              object
Work_of_casualty               object
Fitness_of_casualty            object
Pedestrian_movement            object
Cause_of_accident              object
accident_severity              object
dtype: object
```

```
# count function
df.count()
```

```
Time                           12316
Day_of_week                    12316
Age_band_of_driver             12316
Sex_of_driver                  12316
Educational_level              11575
Vehicle_driver_relation        11087
Driving_experience             11048
Type_of_vehicle                11733
Owner_of_vehicle               11621
Service_year_of_vehicle        10732
Defect_of_vehicle              10161
Area_accident_occured          12077
Lanes_or_Medians               12030
Road_surface_type              12081
Road_surface_conditions        12064
Light_conditions               12316
Work_of_casualty               10672
Weather_conditions             11914
Type_of_collision              11811
Number_of_vehicles_involved    12316
Number_of_casualties           12316
Vehicle_movement               11834
Casualty_class                 12316
Sex_of_casualty                12316
Age_band_of_casualty           12316
Casualty_severity              12316
Work_of_casualty               10677
Fitness_of_casualty            10681
Pedestrian_movement            12316
Cause_of_accident              12316
accident_severity              12316
dtype: object
```

```
# count function
df.groupby('Time')
```

```
<pandas.core.frame.DataFrame object at 0x1c1c0ed0>
```

```
# count function
df.groupby('Time').count
```

```
<bound method DataFrame.count of                      Time  Day_of_week  Age_band_of_driver  Sex_of_driver  \
Time                                                
```