Traffic Accident Data: Cleaning, Exploration and Visualization.

In [1]:
```
# importing libraries:
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
```

In [2]:
```
# importing dataset:
traffic=pd.read_csv("traffic.csv")
```

In [3]:
```
traffic
```

Out[3]:

	Time	Day_of_week	Age_band_of_driver	Sex_of_driver	Educational_level	Vehicle_driver_relation	Driving_experience	Ty
0	17:02:00	Monday	18-30	Male	Above high school	Employee	1-2yr	
1	17:02:00	Monday	31-50	Male	Junior high school	Employee	Above 10yr	
2	17:02:00	Monday	18-30	Male	Junior high school	Employee	1-2yr	L
3	1:06:00	Sunday	18-30	Male	Junior high school	Employee	5-10yr	
4	1:06:00	Sunday	18-30	Male	Junior high school	Employee	2-5yr	
...	...	...	...	...	...	...	...	...
12311	16:15:00	Wednesday	31-50	Male	NaN	Employee	2-5yr	
12312	18:00:00	Sunday	Unknown	Male	Elementary school	Employee	5-10yr	
12313	13:55:00	Sunday	Over 51	Male	Junior high school	Employee	5-10yr	
12314	13:55:00	Sunday	18-30	Female	Junior high school	Employee	Above 10yr	L
12315	13:55:00	Sunday	18-30	Male	Junior high school	Employee	5-10yr	

12316 rows × 32 columns

In [4]:
```
traffic.head()
```

Out[4]:

	Time	Day_of_week	Age_band_of_driver	Sex_of_driver	Educational_level	Vehicle_driver_relation	Driving_experience	Type_
0	17:02:00	Monday	18-30	Male	Above high school	Employee	1-2yr	A
1	17:02:00	Monday	31-50	Male	Junior high school	Employee	Above 10yr	P
2	17:02:00	Monday	18-30	Male	Junior high school	Employee	1-2yr	Lorry (
3	1:06:00	Sunday	18-30	Male	Junior high school	Employee	5-10yr	P
4	1:06:00	Sunday	18-30	Male	Junior high school	Employee	2-5yr	

5 rows × 32 columns

In [135…
```
traffic.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12316 entries, 0 to 12315
Data columns (total 32 columns):
 #   Column                       Non-Null Count  Dtype 
---  ------                       --------------  ----- 
 0   Time                         12316 non-null  object
 1   Day_of_week                  12316 non-null  object
 2   Age_band_of_driver           12316 non-null  object
 3   Sex_of_driver                12316 non-null  object
 4   Educational_level            11575 non-null  object
 5   Vehicle_driver_relation      11737 non-null  object
 6   Driving_experience           11487 non-null  object
 7   Type_of_vehicle              11366 non-null  object
 8   Owner_of_vehicle             11834 non-null  object
 9   Service_year_of_vehicle      8388 non-null   object
 10  Defect_of_vehicle            7889 non-null   object
 11  Area_accident_occured        12077 non-null  object
 12  Lanes_or_Medians             11931 non-null  object
 13  Road_allignment              12174 non-null  object
 14  Types_of_Junction            11429 non-null  object
 15  Road_surface_type            12144 non-null  object
 16  Road_surface_conditions      12316 non-null  object
 17  Light_conditions             12316 non-null  object
 18  Weather_conditions           12316 non-null  object
 19  Type_of_collision            12161 non-null  object
 20  Number_of_vehicles_involved  12316 non-null  int64 
 21  Number_of_casualties         12316 non-null  int64 
 22  Vehicle_movement             12008 non-null  object
 23  Casualty_class               12316 non-null  object
 24  Sex_of_casualty              12316 non-null  object
 25  Age_band_of_casualty         12316 non-null  object
 26  Casualty_severity            12316 non-null  object
 27  Work_of_casuality            9118 non-null   object
 28  Fitness_of_casuality         9681 non-null   object
 29  Pedestrian_movement          12316 non-null  object
 30  Cause_of_accident            12316 non-null  object
 31  Accident_severity            12316 non-null  object
dtypes: int64(2), object(30)
memory usage: 3.0+ MB
```

In [136…
```
traffic.describe()
```

Out[136…

	Number_of_vehicles_involved	Number_of_casualties
count	12316.000000	12316.000000
mean	2.040679	1.548149
std	0.688790	1.007179
min	1.000000	1.000000
25%	2.000000	1.000000
50%	2.000000	1.000000
75%	2.000000	2.000000
max	7.000000	8.000000

In [137…
```
traffic.isnull().sum()
```

Out[137…
```
Time                           0
Day_of_week                    0
Age_band_of_driver             0
Sex_of_driver                  0
Educational_level            741
Vehicle_driver_relation      579
Driving_experience           829
Type_of_vehicle              950
Owner_of_vehicle             482
Service_year_of_vehicle     3928
Defect_of_vehicle           4427
Area_accident_occured        239
Lanes_or_Medians             385
Road_allignment              142
Types_of_Junction            887
Road_surface_type            172
Road_surface_conditions        0
Light_conditions               0
Weather_conditions             0
Type_of_collision            155
Number_of_vehicles_involved    0
Number_of_casualties           0
Vehicle_movement             308
Casualty_class                 0
Sex_of_casualty                0
Age_band_of_casualty           0
Casualty_severity              0
Work_of_casuality           3198
Fitness_of_casuality        2635
Pedestrian_movement            0
Cause_of_accident              0
Accident_severity              0
dtype: int64
```

In [138…
```
traffic.fillna(np.mean)
```

Out[138…

	Time	Day_of_week	Age_band_of_driver	Sex_of_driver	Educational_level	Vehicle_driver_relation	Driving_experienc
0	17:02:00	Monday	18-30	Male	Above high school	Employee	1-2y
1	17:02:00	Monday	31-50	Male	Junior high school	Employee	Above 10y
2	17:02:00	Monday	18-30	Male	Junior high school	Employee	1-2y
3	1:06:00	Sunday	18-30	Male	Junior high school	Employee	5-10y
4	1:06:00	Sunday	18-30	Male	Junior high school	Employee	2-5y
...	...	...	...	...	...	...	.
12311	16:15:00	Wednesday	31-50	Male	<function mean at 0x0000025B9AF585E0>	Employee	2-5y
12312	18:00:00	Sunday	Unknown	Male	Elementary school	Employee	5-10y
12313	13:55:00	Sunday	Over 51	Male	Junior high school	Employee	5-10y
12314	13:55:00	Sunday	18-30	Female	Junior high school	Employee	Above 10y
12315	13:55:00	Sunday	18-30	Male	Junior high school	Employee	5-10y

12316 rows × 32 columns

In [139…
```
traffic.shape
```

Out[139…
```
(12316, 32)
```

In [140…
```
traffic['Accident_severity'].value_counts()
```

Out[140…
```
Accident_severity
Slight Injury     10415
Serious Injury     1743
Fatal injury        158
Name: count, dtype: int64
```

In [141…
```
traffic['Age_band_of_driver'].value_counts()
```

Out[141…
```
Age_band_of_driver
18-30       4271
31-50       4087
Over 51     1585
Unknown     1548
Under 18     825
Name: count, dtype: int64
```

In [142…
```
traffic['Educational_level'].value_counts()
```

Out[142…
```
Educational_level
Junior high school    7619
Elementary school     2163
High school           1110
Above high school      362
Writing & reading      176
Unknown                100
Illiterate              45
Name: count, dtype: int64
```

In [143…
```
traffic['Types_of_Junction'].value_counts()
```

Out[143…
```
Types_of_Junction
Y Shape        4543
No junction    3837
Crossing       2177
Other           445
Unknown         191
O Shape         164
T Shape          60
X Shape          12
Name: count, dtype: int64
```

In [144…
```
traffic['Number_of_vehicles_involved'].value_counts()
```

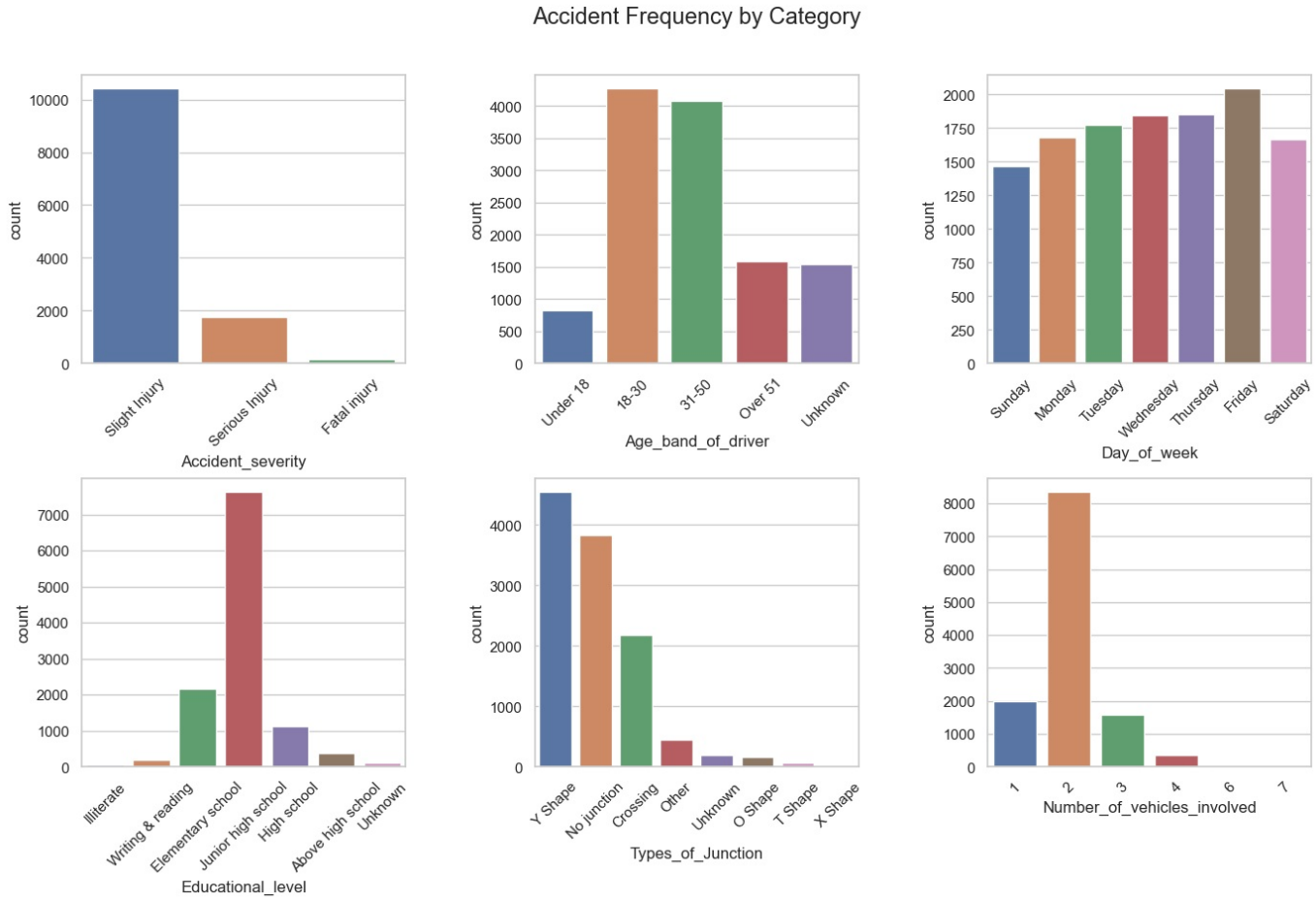
Out[144…
```
Number_of_vehicles_involved
2    8340
1    1996
3    1568
4     363
6      42
7       7
Name: count, dtype: int64
```

Exploratory Data Analysis

putting the data in a visualizing way and in bar charts to see the outcome

In [145…
```
def subplot(traffic, i, column, order=None, palette=None):
    plt.subplot(2, 3, i)
    sns.barplot(data=traffic[column].value_counts().reset_index(), x=column, y='count', palette=palette, order=
    plt.xticks(rotation=45)
```

In [146…
```
sns.set_style('whitegrid')
plt.figure(figsize=(16, 9))
plt.suptitle("Accident Frequency by Category", fontsize=17)

subplot(traffic, i=1, column = 'Accident_severity')

age_order = ['Under 18', '18-30', '31-50', 'Over 51', 'Unknown']
subplot(traffic, i=2, column = 'Age_band_of_driver', order= age_order)

day_order = ['Sunday','Monday','Tuesday','Wednesday','Thursday','Friday','Saturday']
subplot(traffic, i=3, column = 'Day_of_week', order=day_order)

education_order= ["Illiterate", "Writing & reading", "Elementary school", "Junior high school","High school","Al
subplot(traffic, i=4, column = 'Educational_level', order=education_order)

subplot(traffic, i=5, column = 'Types_of_Junction')
subplot(traffic, i=6, column = 'Number_of_vehicles_involved')

plt.subplots_adjust(left=0.1, right=0.9, bottom=0.1, top=0.9, wspace=0.4, hspace=0.4)
plt.show()
```



Results:

Graph 1 : The graph shows that most accident cases are in slight injury .

Graph 2 : The graph shows that the age band of drivers who involved in accidents are mostly in the categories of 18-50 i.e., higher in 18-30 and then in 31-50.

Graph 3 : This graph shows that the majority of accidents are held in the friday more than any other days. This could be due to the end of the work week and rush hours to go home.

Graph 4 : This graph shows that the drivers who are in elementary school education are mostly involved in accidents.Also reading and writing , junior high school drivers are involved.This could indicate a connection between low level of education and a tendency to encounter an accident.

Graph 5 : This graph shows that Y shapes and no junction as well as crossing road type led to the highest counts of Road accident.

Graph 6 : This graph shows that the number of vehicles involved in accidents are usually 1-2 or three.Mostly two.

CONCLUSION:

In conclusion, the analysis reveals various factors influencing road accidents, including day of the week, road and weather conditions, vehicle types, and driver characteristics. Understanding these patterns can inform targeted interventions and safety measures to reduce the frequency and severity of road accidents, such as improving road infrastructure, enhancing vehicle safety features, and promoting safe driving practices among all road users. Additionally, efforts should be made to debunk stereotypes and promote gender-neutral approaches to road safety.