prodigy-ds-05

September 20, 2024

1 Task 5

Analyze traffic accident data to identify patterns related to road conditions, weather, and time of day. Visualize accident hotspots and contributing factors.

Loading Libraries and Data

[4]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
```

[5]:
```
df=pd.read_csv("RTA Dataset.csv")
df.head()
```

[5]:
```
       Time Day_of_week Age_band_of_driver Sex_of_driver   Educational_level  \
0  17:02:00      Monday              18-30          Male   Above high school
1  17:02:00      Monday              31-50          Male  Junior high school
2  17:02:00      Monday              18-30          Male  Junior high school
3   1:06:00      Sunday              18-30          Male  Junior high school
4   1:06:00      Sunday              18-30          Male  Junior high school

  Vehicle_driver_relation Driving_experience      Type_of_vehicle  \
0                Employee              1-2yr           Automobile
1                Employee         Above 10yr  Public (> 45 seats)
2                Employee              1-2yr      Lorry (41?100Q)
3                Employee             5-10yr  Public (> 45 seats)
4                Employee              2-5yr                  NaN

  Owner_of_vehicle Service_year_of_vehicle  … Vehicle_movement   \
0            Owner              Above 10yr  …  Going straight
1            Owner                 5-10yrs  …  Going straight
2            Owner                     NaN  …  Going straight
3     Governmental                     NaN  …  Going straight
4            Owner                 5-10yrs  …  Going straight
```

```
    Casualty_class Sex_of_casualty Age_band_of_casualty Casualty_severity  \
0               na              na                   na                na
1               na              na                   na                na
2  Driver or rider            Male                31-50                 3
3       Pedestrian          Female                18-30                 3
4               na              na                   na                na

  Work_of_casuality Fitness_of_casuality Pedestrian_movement  \
0               NaN                  NaN    Not a Pedestrian
1               NaN                  NaN    Not a Pedestrian
2            Driver                  NaN    Not a Pedestrian
3            Driver               Normal    Not a Pedestrian
4               NaN                  NaN    Not a Pedestrian

           Cause_of_accident Accident_severity
0            Moving Backward     Slight Injury
1                 Overtaking     Slight Injury
2   Changing lane to the left    Serious Injury
3  Changing lane to the right    Slight Injury
4                 Overtaking     Slight Injury

[5 rows x 32 columns]
```

[6]:
```
df.shape
```

[6]:
```
(12316, 32)
```

[7]:
```
df.describe()
```

[7]:
```
       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
```

[8]:
```
df.describe(include="all")
```

[8]:
```
            Time Day_of_week Age_band_of_driver Sex_of_driver  \
count      12316       12316              12316         12316
unique      1074           7                  5             3
top     15:30:00      Friday              18-30          Male
freq         120        2041               4271         11437
mean         NaN         NaN                NaN           NaN
```

```
std          NaN         NaN                NaN           NaN
min          NaN         NaN                NaN           NaN
25%          NaN         NaN                NaN           NaN
50%          NaN         NaN                NaN           NaN
75%          NaN         NaN                NaN           NaN
max          NaN         NaN                NaN           NaN

         Educational_level Vehicle_driver_relation Driving_experience  \
count                11575                   11737              11487
unique                   7                       4                  7
top     Junior high school                Employee             5-10yr
freq                  7619                    9627               3363
mean                   NaN                     NaN                NaN
std                    NaN                     NaN                NaN
min                    NaN                     NaN                NaN
25%                    NaN                     NaN                NaN
50%                    NaN                     NaN                NaN
75%                    NaN                     NaN                NaN
max                    NaN                     NaN                NaN

       Type_of_vehicle Owner_of_vehicle Service_year_of_vehicle  …  \
count            11366            11834                    8388  …
unique              17                4                       6  …
top         Automobile            Owner                 Unknown  …
freq              3205            10459                    2883  …
mean               NaN              NaN                     NaN  …
std                NaN              NaN                     NaN  …
min                NaN              NaN                     NaN  …
25%                NaN              NaN                     NaN  …
50%                NaN              NaN                     NaN  …
75%                NaN              NaN                     NaN  …
max                NaN              NaN                     NaN  …

       Vehicle_movement   Casualty_class Sex_of_casualty Age_band_of_casualty  \
count             12008            12316           12316                12316
unique               13                4               3                    6
top      Going straight  Driver or rider            Male                   na
freq               8158             4944            5253                 4443
mean                NaN              NaN             NaN                  NaN
std                 NaN              NaN             NaN                  NaN
min                 NaN              NaN             NaN                  NaN
25%                 NaN              NaN             NaN                  NaN
50%                 NaN              NaN             NaN                  NaN
75%                 NaN              NaN             NaN                  NaN
max                 NaN              NaN             NaN                  NaN

       Casualty_severity Work_of_casuality Fitness_of_casuality  \
count              12316              9118                 9681
unique                 4                 7                    5
top                    3            Driver               Normal
freq                7076              5903                 9608
mean                 NaN               NaN                  NaN
std                  NaN               NaN                  NaN
min                  NaN               NaN                  NaN
25%                  NaN               NaN                  NaN
50%                  NaN               NaN                  NaN
75%                  NaN               NaN                  NaN
max                  NaN               NaN                  NaN

       Pedestrian_movement Cause_of_accident Accident_severity
count                12316             12316             12316
unique                   9                20                 3
top       Not a Pedestrian     No distancing     Slight Injury
freq                 11390              2263             10415
mean                   NaN               NaN               NaN
std                    NaN               NaN               NaN
min                    NaN               NaN               NaN
25%                    NaN               NaN               NaN
50%                    NaN               NaN               NaN
75%                    NaN               NaN               NaN
max                    NaN               NaN               NaN

[11 rows x 32 columns]
```

[9]:

```
df.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
```

Exploratory Data Analysis

[11]:

```python
df.duplicated().sum()
```

[11]:

```
0
```

[12]:

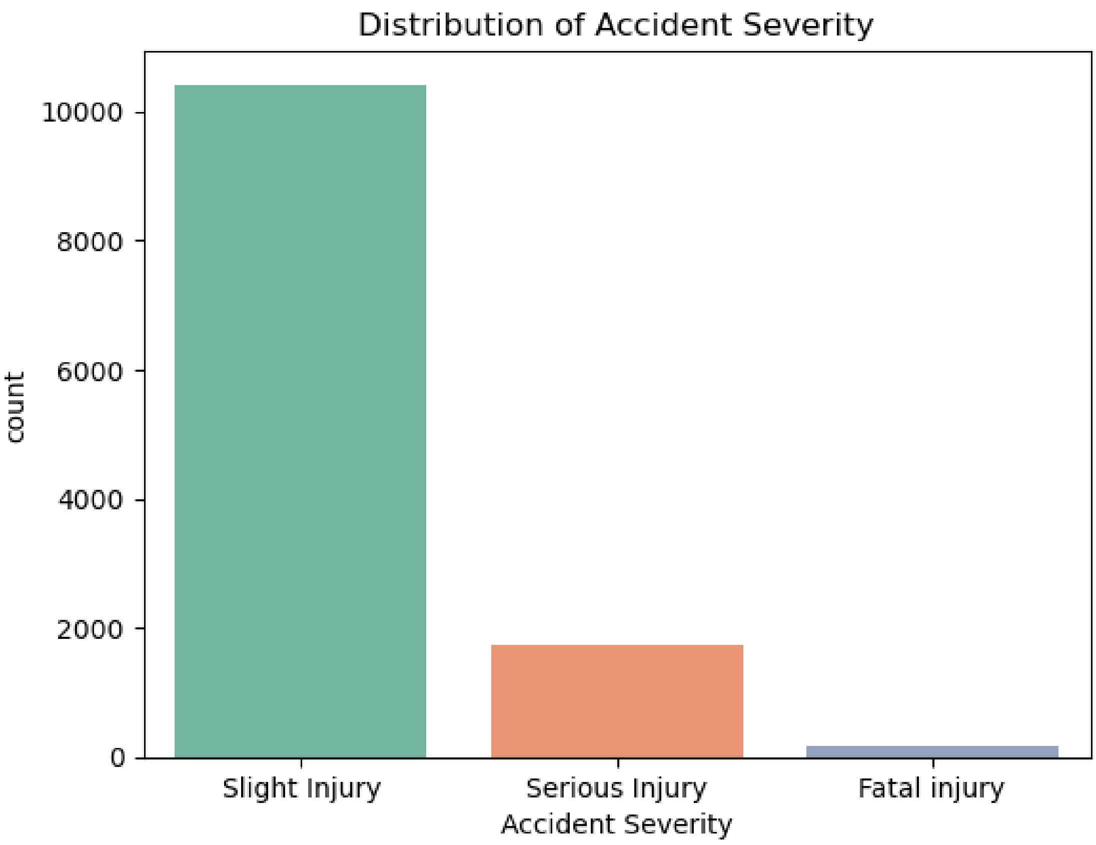
```python
df['Accident_severity'].value_counts()
```

[12]:

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

[13]:

```python
import seaborn as sns
import matplotlib.pyplot as plt
sns.countplot(x=df['Accident_severity'], palette='Set2')
plt.title('Distribution of Accident Severity')
plt.xlabel('Accident Severity')
plt.show()
```



Handling missing values

[15]:
```
df.isna().sum()
```

[15]:
```
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
```

[16]:

```
df.drop(['Service_year_of_vehicle','Defect_of_vehicle','Work_of_casuality',␣
 ↪'Fitness_of_casuality','Time'],
        axis = 1, inplace = True)
df.head()
```

[16]:

```
  Day_of_week Age_band_of_driver Sex_of_driver   Educational_level  \
0      Monday              18-30          Male   Above high school
1      Monday              31-50          Male  Junior high school
2      Monday              18-30          Male  Junior high school
3      Sunday              18-30          Male  Junior high school
4      Sunday              18-30          Male  Junior high school

  Vehicle_driver_relation Driving_experience      Type_of_vehicle  \
0                Employee              1-2yr           Automobile
1                Employee         Above 10yr  Public (> 45 seats)
2                Employee              1-2yr       Lorry (41?100Q)
3                Employee             5-10yr  Public (> 45 seats)
4                Employee              2-5yr                  NaN

  Owner_of_vehicle Area_accident_occured   Lanes_or_Medians  …  \
0            Owner     Residential areas                NaN  …
1            Owner          Office areas  Undivided Two way  …
2            Owner    Recreational areas              other  …
3     Governmental          Office areas              other  …
4            Owner      Industrial areas              other  …

  Number_of_vehicles_involved Number_of_casualties Vehicle_movement  \
0                           2                    2   Going straight
1                           2                    2   Going straight
```

```
2                           2                    2   Going straight
3                           2                    2   Going straight
4                           2                    2   Going straight

    Casualty_class Sex_of_casualty Age_band_of_casualty Casualty_severity  \
0               na              na                   na                na
1               na              na                   na                na
2  Driver or rider            Male                31-50                 3
3       Pedestrian          Female                18-30                 3
4               na              na                   na                na

   Pedestrian_movement           Cause_of_accident Accident_severity
0     Not a Pedestrian             Moving Backward     Slight Injury
1     Not a Pedestrian                  Overtaking     Slight Injury
2     Not a Pedestrian   Changing lane to the left    Serious Injury
3     Not a Pedestrian  Changing lane to the right     Slight Injury
4     Not a Pedestrian                  Overtaking     Slight Injury

[5 rows x 27 columns]
```

[17]:

```
categorical=[i for i in df.columns if df[i].dtype=='O']
print('The categorical variables are',categorical)
```

```
The categorical variables are ['Day_of_week', 'Age_band_of_driver',
'Sex_of_driver', 'Educational_level', 'Vehicle_driver_relation',
'Driving_experience', 'Type_of_vehicle', 'Owner_of_vehicle',
'Area_accident_occured', 'Lanes_or_Medians', 'Road_allignment',
'Types_of_Junction', 'Road_surface_type', 'Road_surface_conditions',
'Light_conditions', 'Weather_conditions', 'Type_of_collision',
'Vehicle_movement', 'Casualty_class', 'Sex_of_casualty', 'Age_band_of_casualty',
'Casualty_severity', 'Pedestrian_movement', 'Cause_of_accident',
'Accident_severity']
```

[18]:

```
for i in categorical:
    df[i].fillna(df[i].mode()[0],inplace=True)
```

[19]:

```
df.isna().sum()
```

[19]:

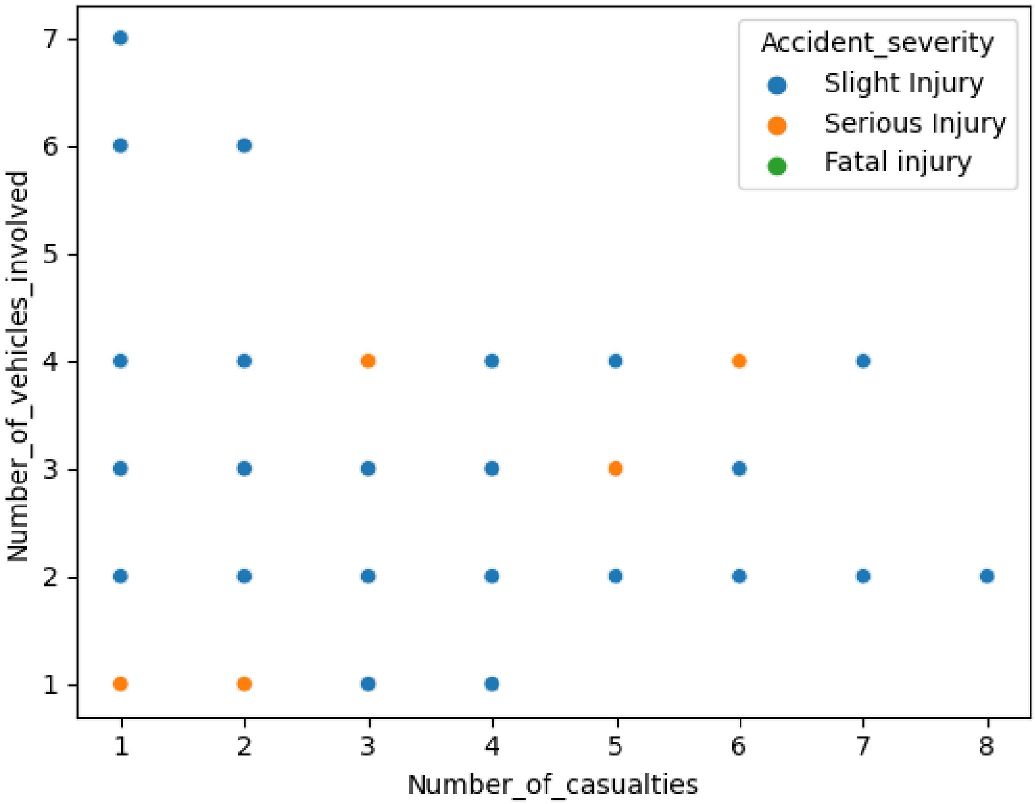
```
Day_of_week                     0
Age_band_of_driver              0
Sex_of_driver                   0
Educational_level               0
Vehicle_driver_relation         0
Driving_experience              0
Type_of_vehicle                 0
Owner_of_vehicle                0
Area_accident_occured           0
```

```
Lanes_or_Medians               0
Road_allignment                0
Types_of_Junction              0
Road_surface_type              0
Road_surface_conditions        0
Light_conditions               0
Weather_conditions             0
Type_of_collision              0
Number_of_vehicles_involved    0
Number_of_casualties           0
Vehicle_movement               0
Casualty_class                 0
Sex_of_casualty                0
Age_band_of_casualty           0
Casualty_severity              0
Pedestrian_movement            0
Cause_of_accident              0
Accident_severity              0
dtype: int64
```

Data Visualization

```
[21]: sns.scatterplot(x=df['Number_of_casualties'],␣
       ↪y=df['Number_of_vehicles_involved'], hue=df['Accident_severity'])
```

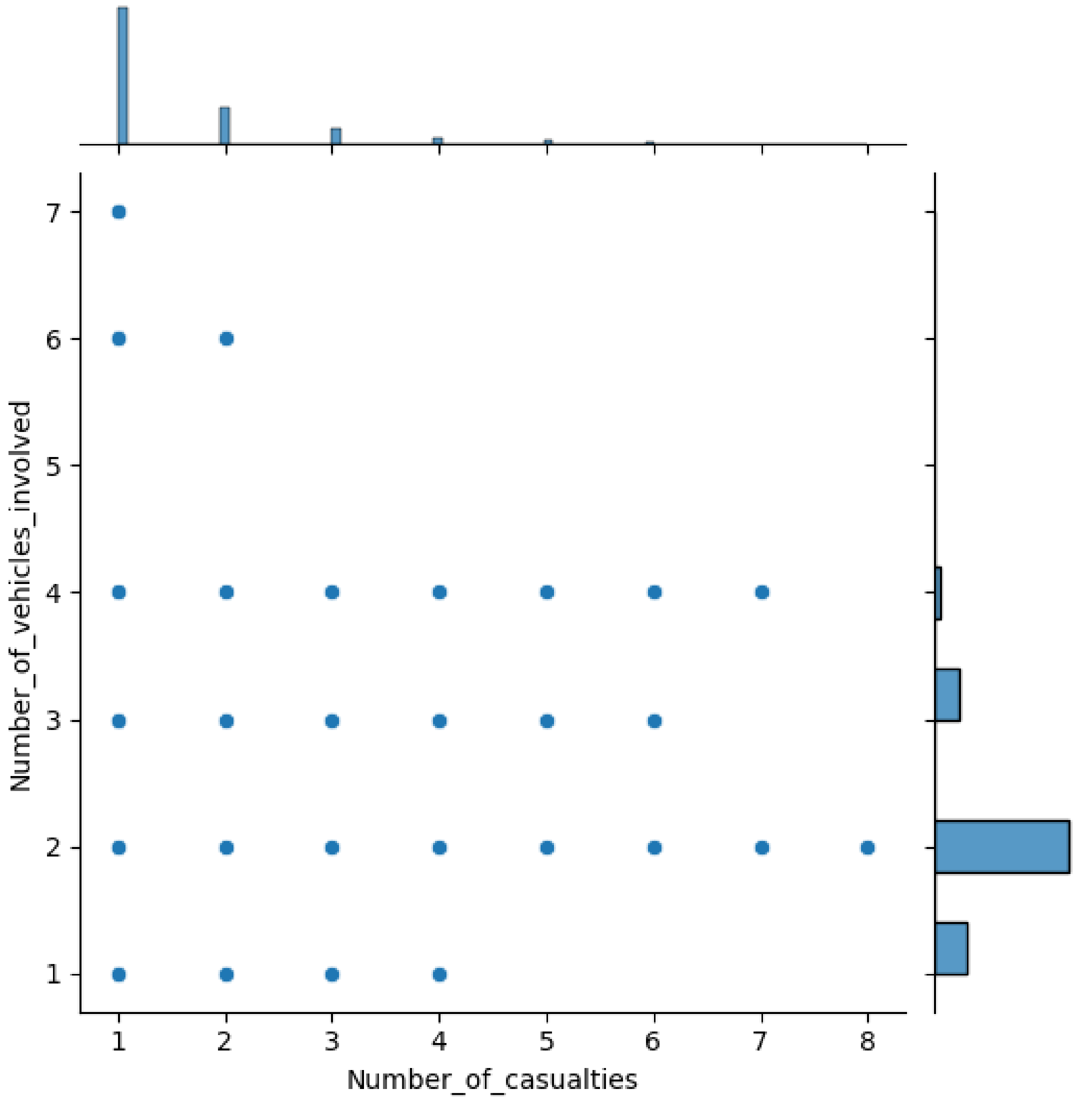
```
[21]: <Axes: xlabel='Number_of_casualties', ylabel='Number_of_vehicles_involved'>
```



There is no visible correlation between Number_of_casualties and Number_of_vehicles_involved columns

[23]:
```
sns.jointplot(x='Number_of_casualties',y='Number_of_vehicles_involved',data=df)
```

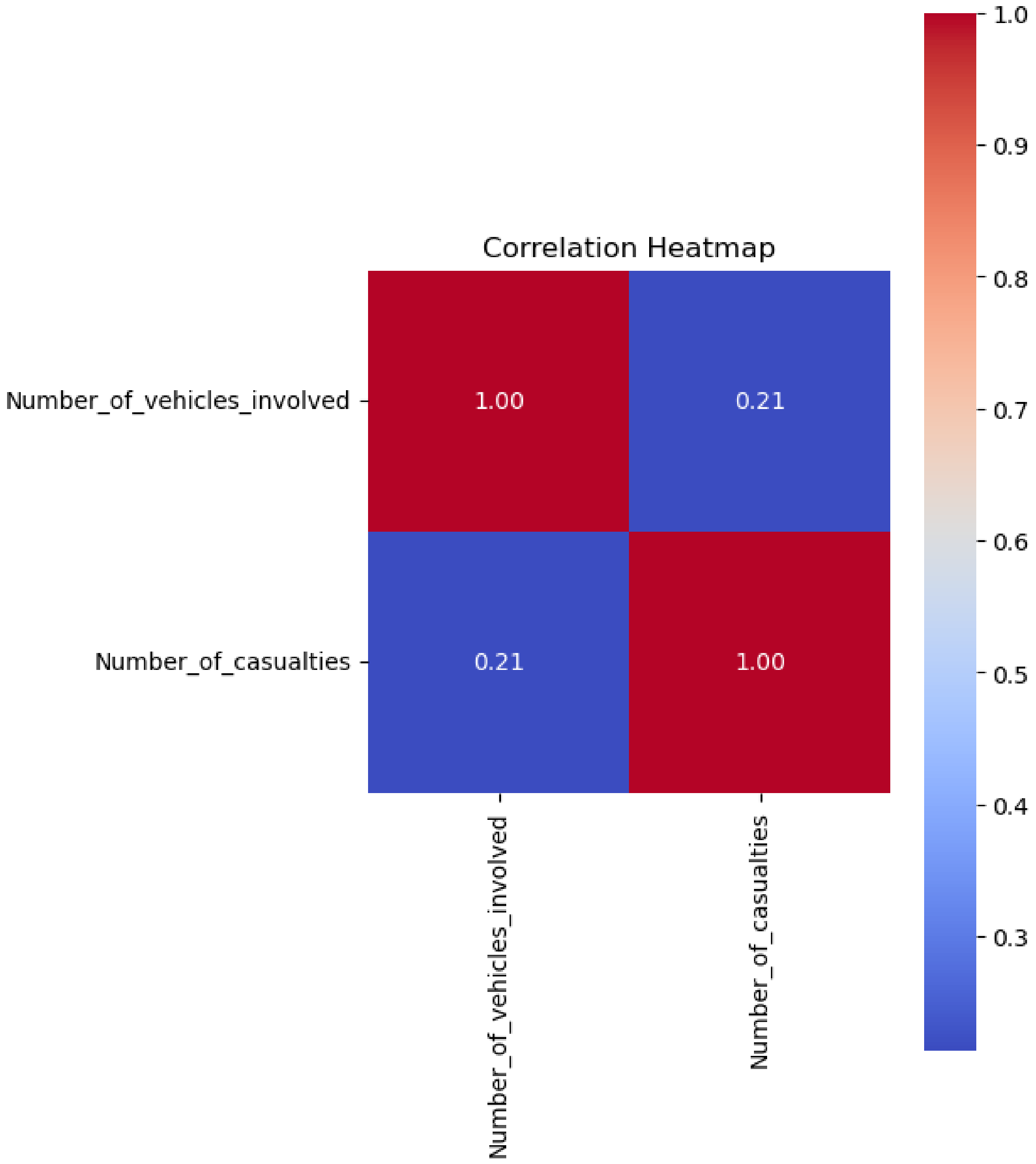
[23]:
```
<seaborn.axisgrid.JointGrid at 0x7fabfff47e10>
```



[24]:
```
print(df.info())
numeric_df = df.select_dtypes(include=['number'])
correlation_matrix = numeric_df.corr()
print(correlation_matrix)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12316 entries, 0 to 12315
Data columns (total 27 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   Day_of_week                  12316 non-null  object
 1   Age_band_of_driver           12316 non-null  object
 2   Sex_of_driver                12316 non-null  object
 3   Educational_level            12316 non-null  object
 4   Vehicle_driver_relation      12316 non-null  object
 5   Driving_experience           12316 non-null  object
 6   Type_of_vehicle              12316 non-null  object
 7   Owner_of_vehicle             12316 non-null  object
 8   Area_accident_occured        12316 non-null  object
 9   Lanes_or_Medians             12316 non-null  object
 10  Road_allignment              12316 non-null  object
 11  Types_of_Junction            12316 non-null  object
 12  Road_surface_type            12316 non-null  object
 13  Road_surface_conditions      12316 non-null  object
 14  Light_conditions             12316 non-null  object
 15  Weather_conditions           12316 non-null  object
 16  Type_of_collision            12316 non-null  object
 17  Number_of_vehicles_involved  12316 non-null  int64
 18  Number_of_casualties         12316 non-null  int64
 19  Vehicle_movement             12316 non-null  object
 20  Casualty_class               12316 non-null  object
 21  Sex_of_casualty              12316 non-null  object
 22  Age_band_of_casualty         12316 non-null  object
 23  Casualty_severity            12316 non-null  object
 24  Pedestrian_movement          12316 non-null  object
 25  Cause_of_accident            12316 non-null  object
 26  Accident_severity            12316 non-null  object
dtypes: int64(2), object(25)
memory usage: 2.5+ MB
None
                             Number_of_vehicles_involved  Number_of_casualties
Number_of_vehicles_involved                     1.000000              0.213427
Number_of_casualties                            0.213427              1.000000
```

[25]:
```
correlation_matrix = df.select_dtypes(include=['number']).corr()
plt.figure(figsize=(5, 8))
sns.heatmap(correlation_matrix, annot=True, fmt=".2f", cmap='coolwarm',␣
 ↪square=True)
plt.title('Correlation Heatmap')
plt.show()
```



[26]:

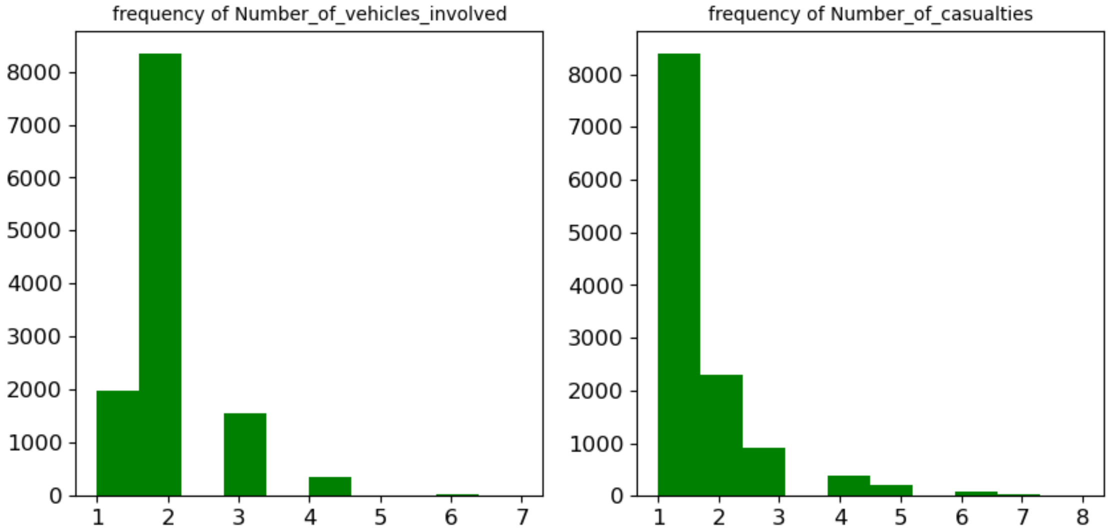
```
numerical=[i for i in df.columns if df[i].dtype!='O']
print('The numerica variables are',numerical)
```

```
The numerica variables are ['Number_of_vehicles_involved',
'Number_of_casualties']
```

[27]:

```
#distribution for numerical columns
plt.figure(figsize=(10,10))
plotnumber = 1
```

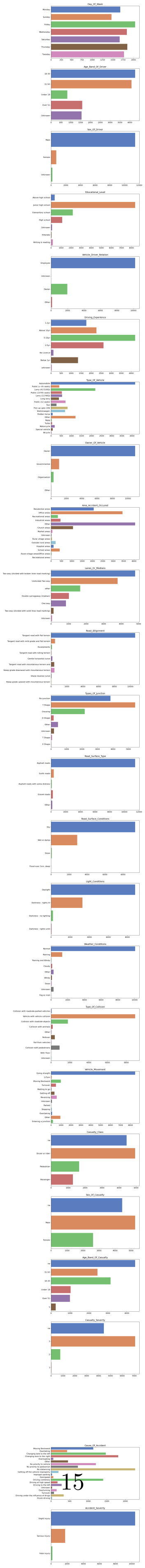
```
for i in numerical:
    if plotnumber <= df.shape[1]:
        ax1 = plt.subplot(2,2,plotnumber)
        plt.hist(df[i],color='green')
        plt.xticks(fontsize=12)
        plt.yticks(fontsize=12)
        plt.title('frequency of '+i, fontsize=10)
    plotnumber +=1
```



Most accidents are occured when 2 vehicles are involved and 1 casuality is happend mostly in the accidents.

[29]:

```
#count plot for categorical values
plt.figure(figsize=(10,200))
plotnumber = 1
for col in categorical:
    if plotnumber <= df.shape[1] and col!='Pedestrian_movement':
        ax1 = plt.subplot(28,1,plotnumber)
        sns.countplot(data=df, y=col, palette='muted')
        plt.xticks(fontsize=12)
        plt.yticks(fontsize=12)
        plt.title(col.title(), fontsize=14)
        plt.xlabel('')
        plt.ylabel('')
    plotnumber +=1
```


Day_Of_Week
Monday
Sunday
Friday
Wednesday
Saturday
Thursday
Tuesday
0 250 500 750 1000 1250 1500 1750 2000
Age_Band_Of_Driver
18-30
31-50
Under 18
Over 51
Unknown
0 500 1000 1500 2000 2500 3000 3500 4000
Sex_Of_Driver
Male
Female
Unknown
0 2000 4000 6000 8000 10000 12000
Educational_Level
Above high school
Junior high school
Elementary school
High school
Unknown
Illiterate
Writing & reading
0 1000 2000 3000 4000 5000 6000 7000 8000
Vehicle_Driver_Relation
Employee
Unknown
Owner
Other
0 2000 4000 6000 8000 10000
Driving_Experience
1-2yr
Above 10yr
5-10yr
2-5yr
No Licence
Below 1yr
unknown
0 500 1000 1500 2000 2500 3000 3500 4000
Type_Of_Vehicle
Automobile
Public (> 45 seats)
Lorry (41?100Q)
Public (13?45 seats)
Lorry (11?40Q)
Long lorry
Public (12 seats)
Taxi
Pick up upto 10Q
Stationwagen
Ridden horse
Other
Bajaj
Turbo
Motorcycle
Special vehicle
Bicycle
0 500 1000 1500 2000 2500 3000 3500 4000
Owner_Of_Vehicle
Owner
Governmental
Organization
Other
0 2000 4000 6000 8000 10000
Area_Accident_Occured
Residential areas
Office areas
Recreational areas
Industrial areas
Other
Church areas
Market areas
Unknown
Rural village areas
Outside rural areas
Hospital areas
School areas
Rural village areasOffice areas
Recreational areas
0 500 1000 1500 2000 2500 3000 3500 4000
Lanes_Or_Medians
Two-way (divided with broken lines road marking)
Undivided Two way
other
Double carriageway (median)
One way
Two-way (divided with solid lines road marking)
Unknown
0 1000 2000 3000 4000 5000
Road_Allignment
Tangent road with flat terrain
Tangent road with mild grade and flat terrain
Escarpments
Tangent road with rolling terrain
Gentle horizontal curve
Tangent road with mountainous terrain and
Steep grade downward with mountainous terrain
Sharp reverse curve
Steep grade upward with mountainous terrain
0 2000 4000 6000 8000 10000
Types_Of_Junction
No junction
Y Shape
Crossing
O Shape
Other
Unknown
T Shape
X Shape
0 1000 2000 3000 4000 5000
Road_Surface_Type
Asphalt roads
Earth roads
Asphalt roads with some distress
Gravel roads
Other
0 2000 4000 6000 8000 10000 12000
Road_Surface_Conditions
Dry
Wet or damp
Snow
Flood over 3cm. deep
0 2000 4000 6000 8000
Light_Conditions
Daylight
Darkness - lights lit
Darkness - no lighting
Darkness - lights unlit
0 2000 4000 6000 8000
Weather_Conditions
Normal
Raining
Raining and Windy
Cloudy
Other
Windy
Snow
Unknown
Fog or mist
0 2000 4000 6000 8000 10000
Type_Of_Collision
Collision with roadside-parked vehicles
Vehicle with vehicle collision
Collision with roadside objects
Collision with animals
Other
Rollover
Fall from vehicles
Collision with pedestrians
With Train
Unknown
0 2000 4000 6000 8000
Vehicle_Movement
Going straight
U-Turn
Moving Backward
Turnover
Waiting to go
Getting off
Reversing
Unknown
Parked
Stopping
Overtaking
Other
Entering a junction
0 1000 2000 3000 4000 5000 6000 7000 8000
Casualty_Class
na
Driver or rider
Pedestrian
Passenger
0 1000 2000 3000 4000 5000
Sex_Of_Casualty
na
Male
Female
0 1000 2000 3000 4000 5000
Age_Band_Of_Casualty
na
31-50
18-30
Under 18
Over 51
5
0 1000 2000 3000 4000
Casualty_Severity
na
3
2
1
0 1000 2000 3000 4000 5000 6000 7000
Cause_Of_Accident
Moving Backward
Overtaking
Changing lane to the left
Changing lane to the right
Overloading
Other
No priority to vehicle
No priority to pedestrian
No distancing
Getting off the vehicle improperly
Improper parking
Overspeed
Driving carelessly
Driving at high speed
Driving to the left
Unknown
Overturning
Turnover
Driving under the influence of drugs
Drunk driving
0 500 1000 1500 2000
Accident_Severity
Slight Injury
Serious Injury
Fatal injury
0 2000 4000 6000 8000 10000

Handling Categorical Values

[31]:
```
df.dtypes
```

[31]:
```
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
Area_accident_occured          object
Lanes_or_Medians               object
Road_allignment                object
Types_of_Junction              object
Road_surface_type              object
Road_surface_conditions        object
Light_conditions               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
Pedestrian_movement            object
Cause_of_accident              object
Accident_severity              object
dtype: object
```

Since there are so many categorical values, we need to use feature selection We need to perform label encoding before applying chi square analysis

[33]:
```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
df1=pd.DataFrame()
#adding all the categorical columns except the output to new data frame
for i in categorical:
    if i!= 'Accident_severity':
        df1[i]=le.fit_transform(df[i])
```

[34]:

```
#confirming the data type
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12316 entries, 0 to 12315
Data columns (total 24 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   Day_of_week              12316 non-null  int64
 1   Age_band_of_driver       12316 non-null  int64
 2   Sex_of_driver            12316 non-null  int64
 3   Educational_level        12316 non-null  int64
 4   Vehicle_driver_relation  12316 non-null  int64
 5   Driving_experience       12316 non-null  int64
 6   Type_of_vehicle          12316 non-null  int64
 7   Owner_of_vehicle         12316 non-null  int64
 8   Area_accident_occured    12316 non-null  int64
 9   Lanes_or_Medians         12316 non-null  int64
 10  Road_allignment          12316 non-null  int64
 11  Types_of_Junction        12316 non-null  int64
 12  Road_surface_type        12316 non-null  int64
 13  Road_surface_conditions  12316 non-null  int64
 14  Light_conditions         12316 non-null  int64
 15  Weather_conditions       12316 non-null  int64
 16  Type_of_collision        12316 non-null  int64
 17  Vehicle_movement         12316 non-null  int64
 18  Casualty_class           12316 non-null  int64
 19  Sex_of_casualty          12316 non-null  int64
 20  Age_band_of_casualty     12316 non-null  int64
 21  Casualty_severity        12316 non-null  int64
 22  Pedestrian_movement      12316 non-null  int64
 23  Cause_of_accident        12316 non-null  int64
dtypes: int64(24)
memory usage: 2.3 MB
```

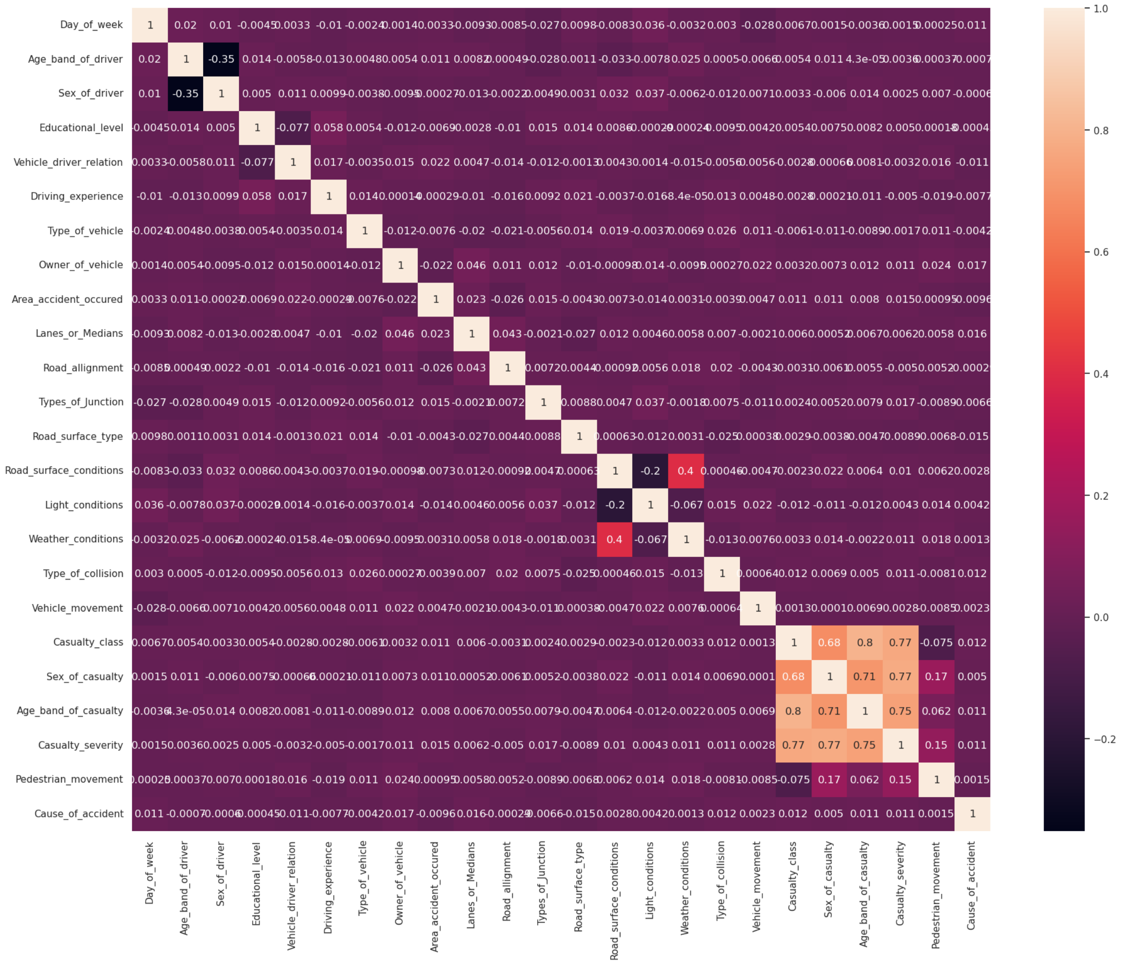
Correlation

[36]:

```
plt.figure(figsize=(22,17))
sns.set(font_scale=1)
sns.heatmap(df1.corr(), annot=True)
```

[36]:

```
<Axes: >
```



```
[37]: df1.head()
```

```
[37]:    Day_of_week  Age_band_of_driver  Sex_of_driver  Educational_level  \
      0            1                   0              1                  0
      1            1                   1              1                  4
      2            1                   0              1                  4
      3            3                   0              1                  4
      4            3                   0              1                  4

         Vehicle_driver_relation  Driving_experience  Type_of_vehicle  \
      0                        0                   0                0
      1                        0                   3               11
      2                        0                   0                5
      3                        0                   2               11
      4                        0                   1                0

         Owner_of_vehicle  Area_accident_occured  Lanes_or_Medians  …  \
      0                 3                      9                 2  …
      1                 3                      6                 4  …
      2                 3                      1                 6  …
      3                 0                      6                 6  …
      4                 3                      4                 6  …

         Light_conditions  Weather_conditions  Type_of_collision  Vehicle_movement  \
      0                 3                   2                  3                 2
      1                 3                   2                  8                 2
      2                 3                   2                  2                 2
      3                 0                   2                  8                 2
      4                 0                   2                  8                 2

         Casualty_class  Sex_of_casualty  Age_band_of_casualty  Casualty_severity  \
      0               3                2                     5                  3
      1               3                2                     5                  3
      2               0                1                     1                  2
      3               2                0                     0                  2
      4               3                2                     5                  3

         Pedestrian_movement  Cause_of_accident
      0                    5                  9
      1                    5                 16
      2                    5                  0
      3                    5                  1
      4                    5                 16

      [5 rows x 24 columns]
```

[38]:

```
from sklearn.feature_selection import chi2
f_p_values=chi2(df1,df['Accident_severity'])
```

[39]:

```
#f_p_values will return Fscore and pvalues
f_p_values
```

[39]:

```
(array([ 0.15822071,  8.91539214,  0.1431894 ,  0.17458477,  5.34534549,
         4.49967858,  1.07767124,  1.10426215,  3.61654037,  3.28161464,
         0.1319306 ,  3.08648691,  6.99480557,  0.61510308, 16.08282359,
         1.14934538, 10.09632283,  2.20071197,  3.2168602 ,  0.12594479,
        13.77841337,  0.20273788,  0.39747982,  3.19366551]),
 array([9.23937958e-01, 1.15890328e-02, 9.30908116e-01, 9.16409114e-01,
        6.90673790e-02, 1.05416165e-01, 5.83427189e-01, 5.75721597e-01,
        1.63937473e-01, 1.93823502e-01, 9.36163348e-01, 2.13686893e-01,
        3.02759144e-02, 7.35244973e-01, 3.21854237e-04, 5.62889079e-01,
        6.42112839e-03, 3.32752607e-01, 2.00201664e-01, 9.38969394e-01,
        1.01872169e-03, 9.03599597e-01, 8.19763078e-01, 2.02536988e-01]))
```

[40]:

```
#for better understanding and ease of access adding them to a new dataframe
f_p_values1=pd.DataFrame({'features':df1.columns, 'Fscore': f_p_values[0],␣
 ↪'Pvalues':f_p_values[1]})
f_p_values1
```

[40]:

```
                   features     Fscore   Pvalues
0               Day_of_week   0.158221  0.923938
1        Age_band_of_driver   8.915392  0.011589
2             Sex_of_driver   0.143189  0.930908
3         Educational_level   0.174585  0.916409
4   Vehicle_driver_relation   5.345345  0.069067
5        Driving_experience   4.499679  0.105416
6           Type_of_vehicle   1.077671  0.583427
7          Owner_of_vehicle   1.104262  0.575722
8     Area_accident_occured   3.616540  0.163937
9          Lanes_or_Medians   3.281615  0.193824
10          Road_allignment   0.131931  0.936163
11        Types_of_Junction   3.086487  0.213687
12        Road_surface_type   6.994806  0.030276
13  Road_surface_conditions   0.615103  0.735245
14         Light_conditions  16.082824  0.000322
15       Weather_conditions   1.149345  0.562889
16        Type_of_collision  10.096323  0.006421
17         Vehicle_movement   2.200712  0.332753
18           Casualty_class   3.216860  0.200202
19          Sex_of_casualty   0.125945  0.938969
20      Age_band_of_casualty  13.778413  0.001019
21        Casualty_severity   0.202738  0.903600
22      Pedestrian_movement   0.397480  0.819763
23        Cause_of_accident   3.193666  0.202537
```

[41]:

```
#since we want lower Pvalues we are sorting the features
f_p_values1.sort_values(by='Pvalues',ascending=True)
```

[41]:

```
                   features     Fscore   Pvalues
14         Light_conditions  16.082824  0.000322
20     Age_band_of_casualty  13.778413  0.001019
16        Type_of_collision  10.096323  0.006421
1        Age_band_of_driver   8.915392  0.011589
12        Road_surface_type   6.994806  0.030276
4   Vehicle_driver_relation   5.345345  0.069067
5        Driving_experience   4.499679  0.105416
8     Area_accident_occured   3.616540  0.163937
9          Lanes_or_Medians   3.281615  0.193824
18           Casualty_class   3.216860  0.200202
23        Cause_of_accident   3.193666  0.202537
11        Types_of_Junction   3.086487  0.213687
```

```
17         Vehicle_movement   2.200712  0.332753
15       Weather_conditions   1.149345  0.562889
7          Owner_of_vehicle   1.104262  0.575722
6           Type_of_vehicle   1.077671  0.583427
13  Road_surface_conditions   0.615103  0.735245
22      Pedestrian_movement   0.397480  0.819763
21        Casualty_severity   0.202738  0.903600
3         Educational_level   0.174585  0.916409
0               Day_of_week   0.158221  0.923938
2             Sex_of_driver   0.143189  0.930908
10          Road_allignment   0.131931  0.936163
19          Sex_of_casualty   0.125945  0.938969
```

we need higher Fscore and lower the Pvalues, so by evaluating, we can remove Owner_of_vehicle, Type_of_vehicle, Road_surface_conditions, Pedestrian_movement,Casualty_severity,Educational_level,Day_of_week,Sex_of_driver,Road_allignment, Sex_of_casualty

[43]:
```
#after evaluating we are removing lesser important columns and storing to a new␣
 ↪data frame
df2=df.drop(['Owner_of_vehicle', 'Type_of_vehicle', 'Road_surface_conditions',␣
 ↪'Pedestrian_movement',
                                                                               ␣
 ↪'Casualty_severity','Educational_level','Day_of_week','Sex_of_driver','Road_allignment',
          'Sex_of_casualty'],axis=1)
df2.head()
```

[43]:
```
  Age_band_of_driver Vehicle_driver_relation Driving_experience  \
0              18-30                Employee              1-2yr
1              31-50                Employee          Above 10yr
2              18-30                Employee              1-2yr
3              18-30                Employee             5-10yr
4              18-30                Employee              2-5yr

  Area_accident_occured                                  Lanes_or_Medians  \
0     Residential areas  Two-way (divided with broken lines road marking)
1          Office areas                                  Undivided Two way
2    Recreational areas                                              other
3          Office areas                                              other
4      Industrial areas                                              other

  Types_of_Junction Road_surface_type       Light_conditions  \
0       No junction     Asphalt roads               Daylight
1       No junction     Asphalt roads               Daylight
2       No junction     Asphalt roads               Daylight
3           Y Shape       Earth roads  Darkness - lights lit
4           Y Shape     Asphalt roads  Darkness - lights lit
```

```
  Weather_conditions                        Type_of_collision  \
0             Normal  Collision with roadside-parked vehicles
1             Normal           Vehicle with vehicle collision
2             Normal          Collision with roadside objects
3             Normal           Vehicle with vehicle collision
4             Normal           Vehicle with vehicle collision

   Number_of_vehicles_involved  Number_of_casualties Vehicle_movement  \
0                            2                     2   Going straight
1                            2                     2   Going straight
2                            2                     2   Going straight
3                            2                     2   Going straight
4                            2                     2   Going straight

    Casualty_class Age_band_of_casualty           Cause_of_accident  \
0               na                   na             Moving Backward
1               na                   na                  Overtaking
2  Driver or rider                31-50   Changing lane to the left
3       Pedestrian                18-30  Changing lane to the right
4               na                   na                  Overtaking

  Accident_severity
0     Slight Injury
1     Slight Injury
2    Serious Injury
3     Slight Injury
4     Slight Injury
```

[44]:

```python
df2.shape
```

[44]:

```
(12316, 17)
```

[45]:

```python
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12316 entries, 0 to 12315
Data columns (total 17 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   Age_band_of_driver       12316 non-null  object
 1   Vehicle_driver_relation  12316 non-null  object
 2   Driving_experience       12316 non-null  object
 3   Area_accident_occured    12316 non-null  object
 4   Lanes_or_Medians         12316 non-null  object
 5   Types_of_Junction        12316 non-null  object
 6   Road_surface_type        12316 non-null  object
 7   Light_conditions             12316 non-null  object
 8   Weather_conditions           12316 non-null  object
 9   Type_of_collision            12316 non-null  object
 10  Number_of_vehicles_involved  12316 non-null  int64
 11  Number_of_casualties         12316 non-null  int64
 12  Vehicle_movement             12316 non-null  object
 13  Casualty_class               12316 non-null  object
 14  Age_band_of_casualty         12316 non-null  object
 15  Cause_of_accident            12316 non-null  object
 16  Accident_severity            12316 non-null  object
dtypes: int64(2), object(15)
memory usage: 1.6+ MB
```

[46]:

```
#to check distinct values in each categorical columns we are storing them to a␣
 ↪new variable
categorical_new=[i for i in df2.columns if df2[i].dtype=='O']
print(categorical_new)
```

```
['Age_band_of_driver', 'Vehicle_driver_relation', 'Driving_experience',
'Area_accident_occured', 'Lanes_or_Medians', 'Types_of_Junction',
'Road_surface_type', 'Light_conditions', 'Weather_conditions',
'Type_of_collision', 'Vehicle_movement', 'Casualty_class',
'Age_band_of_casualty', 'Cause_of_accident', 'Accident_severity']
```

[47]:

```
for i in categorical_new:
    print(df2[i].value_counts())
```

```
Age_band_of_driver
18-30       4271
31-50       4087
Over 51     1585
Unknown     1548
Under 18     825
Name: count, dtype: int64
Vehicle_driver_relation
Employee    10206
Owner        1973
Other         123
Unknown        14
Name: count, dtype: int64
Driving_experience
5-10yr        4192
2-5yr         2613
Above 10yr    2262
1-2yr         1756
Below 1yr     1342
No Licence     118
unknown         33
```

```
Name: count, dtype: int64
Area_accident_occured
Other                              4058
Office areas                       3451
Residential areas                  2060
 Church areas                      1060
 Industrial areas                   456
School areas                        415
  Recreational areas                327
 Outside rural areas                218
 Hospital areas                     121
  Market areas                       63
Rural village areas                  44
Unknown                              22
Rural village areasOffice areas      20
Recreational areas                    1
Name: count, dtype: int64
Lanes_or_Medians
Two-way (divided with broken lines road marking)    4796
Undivided Two way                                   3796
other                                               1660
Double carriageway (median)                         1020
One way                                              845
Two-way (divided with solid lines road marking)      142
Unknown                                               57
Name: count, dtype: int64
Types_of_Junction
Y Shape        5430
No junction    3837
Crossing       2177
Other           445
Unknown         191
O Shape         164
T Shape          60
X Shape          12
Name: count, dtype: int64
Road_surface_type
Asphalt roads                       11468
Earth roads                           358
Gravel roads                          242
Other                                 167
Asphalt roads with some distress       81
Name: count, dtype: int64
Light_conditions
Daylight                   8798
Darkness - lights lit      3286
Darkness - no lighting      192
Darkness - lights unlit      40
```

```
Name: count, dtype: int64
Weather_conditions
Normal               10063
Raining               1331
Other                  296
Unknown                292
Cloudy                 125
Windy                   98
Snow                    61
Raining and Windy       40
Fog or mist             10
Name: count, dtype: int64
Type_of_collision
Vehicle with vehicle collision             8929
Collision with roadside objects            1786
Collision with pedestrians                  896
Rollover                                    397
Collision with animals                      171
Collision with roadside-parked vehicles      54
Fall from vehicles                           34
Other                                        26
Unknown                                      14
With Train                                    9
Name: count, dtype: int64
Vehicle_movement
Going straight         8466
Moving Backward         985
Other                   937
Reversing               563
Turnover                489
Getting off             339
Entering a junction     193
Overtaking               96
Unknown                  88
Stopping                 61
U-Turn                   50
Waiting to go            39
Parked                   10
Name: count, dtype: int64
Casualty_class
Driver or rider    4944
na                 4443
Pedestrian         1649
Passenger          1280
Name: count, dtype: int64
Age_band_of_casualty
na          4443
18-30       3145
```

```
31-50       2455
Under 18    1035
Over 51      994
5            244
Name: count, dtype: int64
Cause_of_accident
No distancing                           2263
Changing lane to the right              1808
Changing lane to the left               1473
Driving carelessly                      1402
No priority to vehicle                  1207
Moving Backward                         1137
No priority to pedestrian                721
Other                                    456
Overtaking                               430
Driving under the influence of drugs     340
Driving to the left                      284
Getting off the vehicle improperly       197
Driving at high speed                    174
Overturning                              149
Turnover                                  78
Overspeed                                 61
Overloading                               59
Drunk driving                             27
Unknown                                   25
Improper parking                          25
Name: count, dtype: int64
Accident_severity
Slight Injury     10415
Serious Injury     1743
Fatal injury        158
Name: count, dtype: int64
```

[48]:
```
dummy=pd.get_dummies(df2[['Age_band_of_driver', 'Vehicle_driver_relation',␣
 ↪'Driving_experience',
                          'Area_accident_occured', 'Lanes_or_Medians',␣
 ↪'Types_of_Junction', 'Road_surface_type',
                          'Light_conditions', 'Weather_conditions',␣
 ↪'Type_of_collision', 'Vehicle_movement',
                          'Casualty_class', 'Age_band_of_casualty',␣
 ↪'Cause_of_accident']],drop_first=True)
dummy.head()
```

[48]:
```
   Age_band_of_driver_31-50  Age_band_of_driver_Over 51  \
0                     False                       False
1                      True                       False
2                     False                       False
```

```
3                     False                        False
4                     False                        False

   Age_band_of_driver_Under 18  Age_band_of_driver_Unknown  \
0                        False                       False
1                        False                       False
2                        False                       False
3                        False                       False
4                        False                       False

   Vehicle_driver_relation_Other  Vehicle_driver_relation_Owner  \
0                          False                          False
1                          False                          False
2                          False                          False
3                          False                          False
4                          False                          False

   Vehicle_driver_relation_Unknown  Driving_experience_2-5yr  \
0                            False                     False
1                            False                     False
2                            False                     False
3                            False                     False
4                            False                      True

   Driving_experience_5-10yr  Driving_experience_Above 10yr  …  \
0                      False                          False  …
1                      False                           True  …
2                      False                          False  …
3                       True                          False  …
4                      False                          False  …

   Cause_of_accident_No distancing  \
0                            False
1                            False
2                            False
3                            False
4                            False

   Cause_of_accident_No priority to pedestrian  \
0                                        False
1                                        False
2                                        False
3                                        False
4                                        False

   Cause_of_accident_No priority to vehicle  Cause_of_accident_Other  \
0                                     False                    False
```

```
1                                     False                    False
2                                     False                    False
3                                     False                    False
4                                     False                    False

   Cause_of_accident_Overloading  Cause_of_accident_Overspeed  \
0                          False                        False
1                          False                        False
2                          False                        False
3                          False                        False
4                          False                        False

   Cause_of_accident_Overtaking  Cause_of_accident_Overturning  \
0                         False                          False
1                          True                          False
2                         False                          False
3                         False                          False
4                          True                          False

   Cause_of_accident_Turnover  Cause_of_accident_Unknown
0                       False                      False
1                       False                      False
2                       False                      False
3                       False                      False
4                       False                      False

[5 rows x 102 columns]
```

[49]:
```
df3=pd.concat([df2,dummy],axis=1)
df3.head()
```

[49]:
```
  Age_band_of_driver Vehicle_driver_relation Driving_experience  \
0              18-30                Employee              1-2yr
1              31-50                Employee         Above 10yr
2              18-30                Employee              1-2yr
3              18-30                Employee             5-10yr
4              18-30                Employee              2-5yr

  Area_accident_occured                                  Lanes_or_Medians  \
0     Residential areas  Two-way (divided with broken lines road marking)
1          Office areas                                 Undivided Two way
2    Recreational areas                                             other
3          Office areas                                             other
4      Industrial areas                                             other

  Types_of_Junction Road_surface_type       Light_conditions  \
0       No junction     Asphalt roads               Daylight
```

```
1       No junction     Asphalt roads             Daylight
2       No junction     Asphalt roads             Daylight
3           Y Shape       Earth roads  Darkness - lights lit
4           Y Shape     Asphalt roads  Darkness - lights lit

  Weather_conditions                        Type_of_collision  …  \
0             Normal  Collision with roadside-parked vehicles  …
1             Normal           Vehicle with vehicle collision  …
2             Normal          Collision with roadside objects  …
3             Normal           Vehicle with vehicle collision  …
4             Normal           Vehicle with vehicle collision  …

   Cause_of_accident_No distancing  \
0                            False
1                            False
2                            False
3                            False
4                            False

   Cause_of_accident_No priority to pedestrian  \
0                                        False
1                                        False
2                                        False
3                                        False
4                                        False

  Cause_of_accident_No priority to vehicle Cause_of_accident_Other  \
0                                    False                   False
1                                    False                   False
2                                    False                   False
3                                    False                   False
4                                    False                   False

  Cause_of_accident_Overloading Cause_of_accident_Overspeed  \
0                         False                       False
1                         False                       False
2                         False                       False
3                         False                       False
4                         False                       False

  Cause_of_accident_Overtaking  Cause_of_accident_Overturning  \
0                        False                          False
1                         True                          False
2                        False                          False
3                        False                          False
4                         True                          False
```

```
   Cause_of_accident_Turnover  Cause_of_accident_Unknown
0                        False                      False
1                        False                      False
2                        False                      False
3                        False                      False
4                        False                      False

[5 rows x 119 columns]
```

[50]:
```
#dropping dummied columns
df3.drop(['Age_band_of_driver', 'Vehicle_driver_relation',␣
 ↪'Driving_experience', 'Area_accident_occured', 'Lanes_or_Medians',
          'Types_of_Junction', 'Road_surface_type', 'Light_conditions',␣
 ↪'Weather_conditions', 'Type_of_collision',
          'Vehicle_movement','Casualty_class', 'Age_band_of_casualty',␣
 ↪'Cause_of_accident'],axis=1,inplace=True)
df3.head()
```

[50]:
```
   Number_of_vehicles_involved  Number_of_casualties Accident_severity  \
0                            2                     2     Slight Injury
1                            2                     2     Slight Injury
2                            2                     2    Serious Injury
3                            2                     2     Slight Injury
4                            2                     2     Slight Injury

   Age_band_of_driver_31-50  Age_band_of_driver_Over 51  \
0                     False                       False
1                      True                       False
2                     False                       False
3                     False                       False
4                     False                       False

   Age_band_of_driver_Under 18  Age_band_of_driver_Unknown  \
0                        False                       False
1                        False                       False
2                        False                       False
3                        False                       False
4                        False                       False

   Vehicle_driver_relation_Other  Vehicle_driver_relation_Owner  \
0                          False                          False
1                          False                          False
2                          False                          False
3                          False                          False
4                          False                          False

   Vehicle_driver_relation_Unknown  …  Cause_of_accident_No distancing  \
0                            False  …                            False
1                            False  …                            False
2                            False  …                            False
3                            False  …                            False
4                            False  …                            False

   Cause_of_accident_No priority to pedestrian  \
0                                        False
1                                        False
2                                        False
3                                        False
4                                        False

   Cause_of_accident_No priority to vehicle  Cause_of_accident_Other  \
0                                     False                    False
1                                     False                    False
2                                     False                    False
3                                     False                    False
4                                     False                    False

   Cause_of_accident_Overloading  Cause_of_accident_Overspeed  \
0                          False                        False
1                          False                        False
2                          False                        False
3                          False                        False
4                          False                        False

   Cause_of_accident_Overtaking  Cause_of_accident_Overturning  \
0                         False                          False
1                          True                          False
2                         False                          False
3                         False                          False
4                          True                          False

   Cause_of_accident_Turnover  Cause_of_accident_Unknown
0                       False                      False
1                       False                      False
2                       False                      False
3                       False                      False
4                       False                      False

[5 rows x 105 columns]
```

Seperating Independent and Dependent

[52]:
```
x=df3.drop(['Accident_severity'],axis=1)
x.shape
```

```
[52]: (12316, 104)
```

```
[53]: x.head()
```

```
[53]:    Number_of_vehicles_involved  Number_of_casualties  \
      0                            2                     2
      1                            2                     2
      2                            2                     2
      3                            2                     2
      4                            2                     2

         Age_band_of_driver_31-50  Age_band_of_driver_Over 51  \
      0                     False                       False
      1                      True                       False
      2                     False                       False
      3                     False                       False
      4                     False                       False

         Age_band_of_driver_Under 18  Age_band_of_driver_Unknown  \
      0                        False                       False
      1                        False                       False
      2                        False                       False
      3                        False                       False
      4                        False                       False

         Vehicle_driver_relation_Other  Vehicle_driver_relation_Owner  \
      0                          False                          False
      1                          False                          False
      2                          False                          False
      3                          False                          False
      4                          False                          False

         Vehicle_driver_relation_Unknown  Driving_experience_2-5yr  …  \
      0                            False                     False  …
      1                            False                     False  …
      2                            False                     False  …
      3                            False                     False  …
      4                            False                      True  …

         Cause_of_accident_No distancing  \
      0                            False
      1                            False
      2                            False
      3                            False
      4                            False

         Cause_of_accident_No priority to pedestrian  \
      0                                        False
      1                                        False
      2                                        False
      3                                        False
      4                                        False

         Cause_of_accident_No priority to vehicle  Cause_of_accident_Other  \
      0                                     False                    False
      1                                     False                    False
      2                                     False                    False
      3                                     False                    False
      4                                     False                    False

         Cause_of_accident_Overloading  Cause_of_accident_Overspeed  \
      0                          False                        False
      1                          False                        False
      2                          False                        False
      3                          False                        False
      4                          False                        False

         Cause_of_accident_Overtaking  Cause_of_accident_Overturning  \
      0                         False                          False
      1                          True                          False
      2                         False                          False
      3                         False                          False
      4                          True                          False

         Cause_of_accident_Turnover  Cause_of_accident_Unknown
      0                       False                      False
      1                       False                      False
      2                       False                      False
      3                       False                      False
      4                       False                      False

      [5 rows x 104 columns]
```

[54]:

```
y=df3.iloc[:,2]
y.head()
```

[54]:

```
0     Slight Injury
1     Slight Injury
2    Serious Injury
3     Slight Injury
4     Slight Injury
Name: Accident_severity, dtype: object
```

[55]:

```
y.value_counts()
```

```
[55]: Accident_severity
      Slight Injury     10415
      Serious Injury     1743
      Fatal injury        158
      Name: count, dtype: int64
```

Splitting the data

KNN Model Creation

Prediction

```
[64]: from sklearn.model_selection import train_test_split

      x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2,␣
       ↪random_state=42)
      from sklearn.neighbors import KNeighborsClassifier

      model_KNN = KNeighborsClassifier(n_neighbors=5)
      model_KNN.fit(x_train, y_train)
```

```
[64]: KNeighborsClassifier()
```

```
[66]: y_pred=model_KNN.predict(x_test)
```

```
[67]: y_pred
```

```
[67]: array(['Slight Injury', 'Slight Injury', 'Slight Injury', …,
             'Slight Injury', 'Slight Injury', 'Slight Injury'], dtype=object)
```

Checking Accuracy, Classification Report, Confusion Matrix

```
[ ]: from sklearn.metrics import␣
      ↪confusion_matrix,classification_report,,accuracy_score,ConfusionMatrixDisplay
```

Classification Report

```
[ ]: report_KNN=classification_report(y_test,y_pred)
     print(report_KNN)
```

Accuracy Score

```
[ ]: accuracy_KNN=accuracy_score(y_test,y_pred)
     print(accuracy_KNN)
```

Confusion Matrix

[ ]:

```
matrix_KNN=confusion_matrix(y_test,y_pred)
print(matrix_KNN,'\n')
print(ConfusionMatrixDisplay.from_predictions(y_test,y_pred))
```