

PRODIGY INFOTECH

TASK-5

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

```
In [32]: #Import required Libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from matplotlib.pyplot import figure
```

```
In [5]: #Load the data
Data=pd.read_csv("C:/Users/steph/Downloads/uk.csv")
Data
```

Out[5]:

	Unnamed: 0	Accident_Index	Location_Easting_OSGR	Location_Northing_OSGR	Longitude
0	0	200501BS00001	525680.0	178240	-0.191170
1	1	200501BS00002	524170.0	181650	-0.211708
2	2	200501BS00003	524520.0	182240	-0.206458
3	3	200501BS00004	526900.0	177530	-0.173862
4	4	200501BS00005	528060.0	179040	-0.156618
...
1048570	9117	201201LX50301	530140.0	170990	-0.129632
1048571	9118	201201LX50302	529820.0	170510	-0.134408
1048572	9119	201201LX50303	530650.0	176730	-0.120183
1048573	9120	201201LX50304	532220.0	172500	-0.099170
1048574	9121	201201LX50305	531180.0	177000	-0.112456

1048575 rows × 33 columns

In [6]:

```
#Create a dataframe  
df=pd.DataFrame(Data)  
df
```

Out[6]:

	Unnamed: 0	Accident_Index	Location_Easting_OSGR	Location_Northing_OSGR	Longitude
0	0	200501BS00001	525680.0	178240	-0.191170
1	1	200501BS00002	524170.0	181650	-0.211708
2	2	200501BS00003	524520.0	182240	-0.206458
3	3	200501BS00004	526900.0	177530	-0.173862
4	4	200501BS00005	528060.0	179040	-0.156618
...
1048570	9117	201201LX50301	530140.0	170990	-0.129632
1048571	9118	201201LX50302	529820.0	170510	-0.134408
1048572	9119	201201LX50303	530650.0	176730	-0.120183
1048573	9120	201201LX50304	532220.0	172500	-0.099170
1048574	9121	201201LX50305	531180.0	177000	-0.112456

1048575 rows × 33 columns

In [8]: #First 5 rows of the dataset
df.head()

Out[8]:

	Unnamed: 0	Accident_Index	Location_Easting_OSGR	Location_Northing_OSGR	Longitude	Latitude
0	0	200501BS00001	525680.0	178240	-0.191170	51.4890
1	1	200501BS00002	524170.0	181650	-0.211708	51.5200
2	2	200501BS00003	524520.0	182240	-0.206458	51.5253
3	3	200501BS00004	526900.0	177530	-0.173862	51.4824
4	4	200501BS00005	528060.0	179040	-0.156618	51.4957

5 rows × 33 columns

In [10]: #Last 5 rows of the dataset
df.tail()

Out[10]:

	Unnamed: 0	Accident_Index	Location_Easting_OSGR	Location_Northing_OSGR	Longitude
1048570	9117	201201LX50301	530140.0	170990	-0.129632
1048571	9118	201201LX50302	529820.0	170510	-0.134408
1048572	9119	201201LX50303	530650.0	176730	-0.120183
1048573	9120	201201LX50304	532220.0	172500	-0.099170
1048574	9121	201201LX50305	531180.0	177000	-0.112456

5 rows × 33 columns

In [11]: #Check for all the columns of the dataset
df.columns

```
Out[11]: Index(['Unnamed: 0', 'Accident_Index', 'Location_Easting_OSGR',  
   'Location_Northing_OSGR', 'Longitude', 'Latitude', 'Police_Force',  
   'Accident_Severity', 'Number_of_Vehicles', 'Number_of_Casualties',  
   'Date', 'Day_of_Week', 'Time', 'Local_Authority_(District)',  
   'Local_Authority_(Highway)', '1st_Road_Class', '1st_Road_Number',  
   'Road_Type', 'Speed_limit', 'Junction_Control', '2nd_Road_Class',  
   '2nd_Road_Number', 'Pedestrian_Crossing-Human_Control',  
   'Pedestrian_Crossing-Physical_Facilities', 'Light_Conditions',  
   'Weather_Conditions', 'Road_Surface_Conditions',  
   'Special_Conditions_at_Site', 'Carriageway_Hazards',  
   'Urban_or_Rural_Area', 'Did_Police_Officer_Attend_Scene_of_Accident',  
   'LSOA_of_Accident_Location', 'Year'],  
  dtype='object')
```

```
In [12]: #Check for the number of rows and columns of the dataset  
df.shape
```

```
Out[12]: (1048575, 33)
```

```
In [13]: #Check for the information ,i.e, dtype and null value for each column  
df.info
```

```
Out[13]: <bound method DataFrame.info of
g_OSGR \
0          0  200501BS00001           525680.0
1          1  200501BS00002           524170.0
2          2  200501BS00003           524520.0
3          3  200501BS00004           526900.0
4          4  200501BS00005           528060.0
...
1048570    9117 201201LX50301           530140.0
1048571    9118 201201LX50302           529820.0
1048572    9119 201201LX50303           530650.0
1048573    9120 201201LX50304           532220.0
1048574    9121 201201LX50305           531180.0

      Location_Northing_OSGR  Longitude  Latitude  Police_Force \
0                  178240 -0.191170  51.489096           1
1                  181650 -0.211708  51.520075           1
2                  182240 -0.206458  51.525301           1
3                  177530 -0.173862  51.482442           1
4                  179040 -0.156618  51.495752           1
...
1048570    170990 -0.129632  51.422932           1
1048571    170510 -0.134408  51.418692           1
1048572    176730 -0.120183  51.474400           1
1048573    172500 -0.099170  51.436021           1
1048574    177000 -0.112456  51.476704           1

      Accident_Severity  Number_of_Vehicles  Number_of_Casualties  ... \
0                      2                      1                      1  ...
1                      3                      1                      1  ...
2                      3                      2                      1  ...
3                      3                      1                      1  ...
4                      3                      1                      1  ...
...
1048570    3                      2                      1  ...
1048571    3                      1                      1  ...
1048572    3                      1                      2  ...
1048573    3                      3                      1  ...
1048574    3                      2                      1  ...

      Pedestrian_Crossing-Physical_Facilities \
0                               Zebra crossing
1  Pedestrian phase at traffic signal junction
2  No physical crossing within 50 meters
3  No physical crossing within 50 meters
4  No physical crossing within 50 meters
...
1048570  No physical crossing within 50 meters
1048571  No physical crossing within 50 meters
1048572  Pedestrian phase at traffic signal junction
1048573  No physical crossing within 50 meters
1048574  No physical crossing within 50 meters

      Light_Conditions  Weather_Conditions \
0  Daylight: Street light present  Raining without high winds
1  Darkness: Street lights present and lit  Fine without high winds
2  Darkness: Street lights present and lit  Fine without high winds
3  Daylight: Street light present  Fine without high winds
4  Darkness: Street lighting unknown  Fine without high winds
```

...	
1048570	Daylight: Street light present	Fine without high winds	
1048571	Daylight: Street light present	Fine without high winds	
1048572	Daylight: Street light present	Raining with high winds	
1048573	Daylight: Street light present	Fine without high winds	
1048574	Daylight: Street light present	Raining without high winds	
	Road_Surface_Conditions	Special_Conditions_at_Site	\
0	Wet/Damp	None	
1	Dry	None	
2	Dry	None	
3	Dry	None	
4	Wet/Damp	None	
...	
1048570	Dry	None	
1048571	Dry	None	
1048572	Flood (Over 3cm of water)	None	
1048573	Dry	None	
1048574	Wet/Damp	None	
	Carriageway_Hazards	Urban_or_Rural_Area	\
0	None	1	
1	None	1	
2	None	1	
3	None	1	
4	None	1	
...	
1048570	None	1	
1048571	None	1	
1048572	None	1	
1048573	None	1	
1048574	None	1	
	Did_Police_Officer_Attend_Scene_of_Accident		\
0		Yes	
1		Yes	
2		Yes	
3		Yes	
4		Yes	
...		...	
1048570		Yes	
1048571		Yes	
1048572		Yes	
1048573		Yes	
1048574		Yes	
	LSOA_of_Accident_Location	Year	
0	E01002849	2005	
1	E01002909	2005	
2	E01002857	2005	
3	E01002840	2005	
4	E01002863	2005	
...	
1048570	E01003144	2012	
1048571	E01003115	2012	
1048572	E01003126	2012	
1048573	E01003167	2012	
1048574	E01003181	2012	

```
[1048575 rows x 33 columns]>
```

```
In [14]: #Check for Statistical Analysis  
df.describe
```

```
Out[14]: <bound method NDFrame.describe of
           ing_OSGR \>

          Unnamed: 0 Accident_Index Location_East
0            0 200501BS00001      525680.0
1            1 200501BS00002      524170.0
2            2 200501BS00003      524520.0
3            3 200501BS00004      526900.0
4            4 200501BS00005      528060.0
...
...       ...
1048570    9117 201201LX50301      530140.0
1048571    9118 201201LX50302      529820.0
1048572    9119 201201LX50303      530650.0
1048573    9120 201201LX50304      532220.0
1048574    9121 201201LX50305      531180.0

          Location_Northing_OSGR Longitude Latitude Police_Force \
0                  178240 -0.191170  51.489096      1
1                  181650 -0.211708  51.520075      1
2                  182240 -0.206458  51.525301      1
3                  177530 -0.173862  51.482442      1
4                  179040 -0.156618  51.495752      1
...
...       ...
1048570    170990 -0.129632  51.422932      1
1048571    170510 -0.134408  51.418692      1
1048572    176730 -0.120183  51.474400      1
1048573    172500 -0.099170  51.436021      1
1048574    177000 -0.112456  51.476704      1

          Accident_Severity Number_of_Vehicles Number_of_Casualties ... \
0                  2                      1                 1   ...
1                  3                      1                 1   ...
2                  3                      2                 1   ...
3                  3                      1                 1   ...
4                  3                      1                 1   ...
...
...       ...
1048570    3                      2                 1   ...
1048571    3                      1                 1   ...
1048572    3                      1                 2   ...
1048573    3                      3                 1   ...
1048574    3                      2                 1   ...

          Pedestrian_Crossing-Physical_Facilities \
0                  Zebra crossing
1  Pedestrian phase at traffic signal junction
2  No physical crossing within 50 meters
3  No physical crossing within 50 meters
4  No physical crossing within 50 meters
...
...
1048570  No physical crossing within 50 meters
1048571  No physical crossing within 50 meters
1048572  Pedestrian phase at traffic signal junction
1048573  No physical crossing within 50 meters
1048574  No physical crossing within 50 meters

          Light_Conditions Weather_Conditions \
0  Daylight: Street light present  Raining without high winds
1  Darkness: Street lights present and lit  Fine without high winds
2  Darkness: Street lights present and lit  Fine without high winds
3  Daylight: Street light present  Fine without high winds
4  Darkness: Street lighting unknown  Fine without high winds
```

...	
1048570	Daylight: Street light present	Fine without high winds	
1048571	Daylight: Street light present	Fine without high winds	
1048572	Daylight: Street light present	Raining with high winds	
1048573	Daylight: Street light present	Fine without high winds	
1048574	Daylight: Street light present	Raining without high winds	
	Road_Surface_Conditions	Special_Conditions_at_Site	\
0	Wet/Damp	None	
1	Dry	None	
2	Dry	None	
3	Dry	None	
4	Wet/Damp	None	
...	
1048570	Dry	None	
1048571	Dry	None	
1048572	Flood (Over 3cm of water)	None	
1048573	Dry	None	
1048574	Wet/Damp	None	
	Carriageway_Hazards	Urban_or_Rural_Area	\
0	None	1	
1	None	1	
2	None	1	
3	None	1	
4	None	1	
...	
1048570	None	1	
1048571	None	1	
1048572	None	1	
1048573	None	1	
1048574	None	1	
	Did_Police_Officer_Attend_Scene_of_Accident		\
0		Yes	
1		Yes	
2		Yes	
3		Yes	
4		Yes	
...		...	
1048570		Yes	
1048571		Yes	
1048572		Yes	
1048573		Yes	
1048574		Yes	
	LSOA_of_Accident_Location	Year	
0	E01002849	2005	
1	E01002909	2005	
2	E01002857	2005	
3	E01002840	2005	
4	E01002863	2005	
...	
1048570	E01003144	2012	
1048571	E01003115	2012	
1048572	E01003126	2012	
1048573	E01003167	2012	
1048574	E01003181	2012	

```
[1048575 rows x 33 columns]>
```

```
In [15]: #Datatype of each column  
df.dtypes
```

```
Out[15]: Unnamed: 0           int64  
Accident_Index          object  
Location_Easting_OSGR    float64  
Location_Northing_OSGR   int64  
Longitude                float64  
Latitude                 float64  
Police_Force              int64  
Accident_Severity         int64  
Number_of_Vehicles        int64  
Number_of_Casualties      int64  
Date                      object  
Day_of_Week               int64  
Time                      object  
Local_Authority_(District) int64  
Local_Authority_(Highway)  object  
1st_Road_Class            int64  
1st_Road_Number           int64  
Road_Type                 object  
Speed_limit               int64  
Junction_Control          object  
2nd_Road_Class            int64  
2nd_Road_Number           int64  
Pedestrian_Crossing-Human_Control object  
Pedestrian_Crossing-Physical_Facilities object  
Light_Conditions           object  
Weather_Conditions         object  
Road_Surface_Conditions    object  
Special_Conditions_at_Site object  
Carriageway_Hazards         object  
Urban_or_Rural_Area        int64  
Did_Police_Officer_Attend_Scene_of_Accident object  
LSOA_of_Accident_Location  object  
Year                      int64  
dtype: object
```

DATA CLEANING

```
In [16]: #Check for the null values  
print(df.isnull().sum())
```

Unnamed: 0	0
Accident_Index	0
Location_Easting_OSGR	101
Location_Northing_OSGR	0
Longitude	101
Latitude	0
Police_Force	0
Accident_Severity	0
Number_of_Vehicles	0
Number_of_Casualties	0
Date	0
Day_of_Week	0
Time	104
Local_Authority_(District)	0
Local_Authority_(Highway)	0
1st_Road_Class	0
1st_Road_Number	0
Road_Type	0
Speed_limit	0
Junction_Control	0
2nd_Road_Class	0
2nd_Road_Number	0
Pedestrian_Crossing-Human_Control	17
Pedestrian_Crossing-Physical_Facilities	34
Light_Conditions	0
Weather_Conditions	0
Road_Surface_Conditions	0
Special_Conditions_at_Site	0
Carriageway_Hazards	0
Urban_or_Rural_Area	0
Did_Police_Officer_Attend_Scene_of_Accident	0
LSOA_of_Accident_Location	79520
Year	0
dtype: int64	

```
In [17]: #Remove unnecesary columns
df1=df.drop(['Location_Easting_OSGR','Longitude',
             'LSOA_of_Accident_Location','Pedestrian_Crossing-Human_Control',
             'Pedestrian_Crossing-Physical_Facilities'],axis=1)
df1
```

Out[17]:

	Unnamed: 0	Accident_Index	Location_Northing_OSGR	Latitude	Police_Force	Accident_S
0	0	200501BS00001		178240	51.489096	1
1	1	200501BS00002		181650	51.520075	1
2	2	200501BS00003		182240	51.525301	1
3	3	200501BS00004		177530	51.482442	1
4	4	200501BS00005		179040	51.495752	1
...
1048570	9117	201201LX50301		170990	51.422932	1
1048571	9118	201201LX50302		170510	51.418692	1
1048572	9119	201201LX50303		176730	51.474400	1
1048573	9120	201201LX50304		172500	51.436021	1
1048574	9121	201201LX50305		177000	51.476704	1

1048575 rows × 28 columns

In [18]: #New dataframe without unnecesary columns
print(df1.isnull().sum())

```
Unnamed: 0          0
Accident_Index      0
Location_Northing_OSGR 0
Latitude           0
Police_Force       0
Accident_Severity   0
Number_of_Vehicles  0
Number_of_Casualties 0
Date               0
Day_of_Week         0
Time               104
Local_Authority_(District) 0
Local_Authority_(Highway) 0
1st_Road_Class     0
1st_Road_Number    0
Road_Type          0
Speed_limit        0
Junction_Control  0
2nd_Road_Class    0
2nd_Road_Number   0
Light_Conditions   0
Weather_Conditions 0
Road_Surface_Conditions 0
Special_Conditions_at_Site 0
Carriageway_Hazards 0
Urban_or_Rural_Area 0
Did_Police_Officer_Attend_Scene_of_Accident 0
Year               0
dtype: int64
```

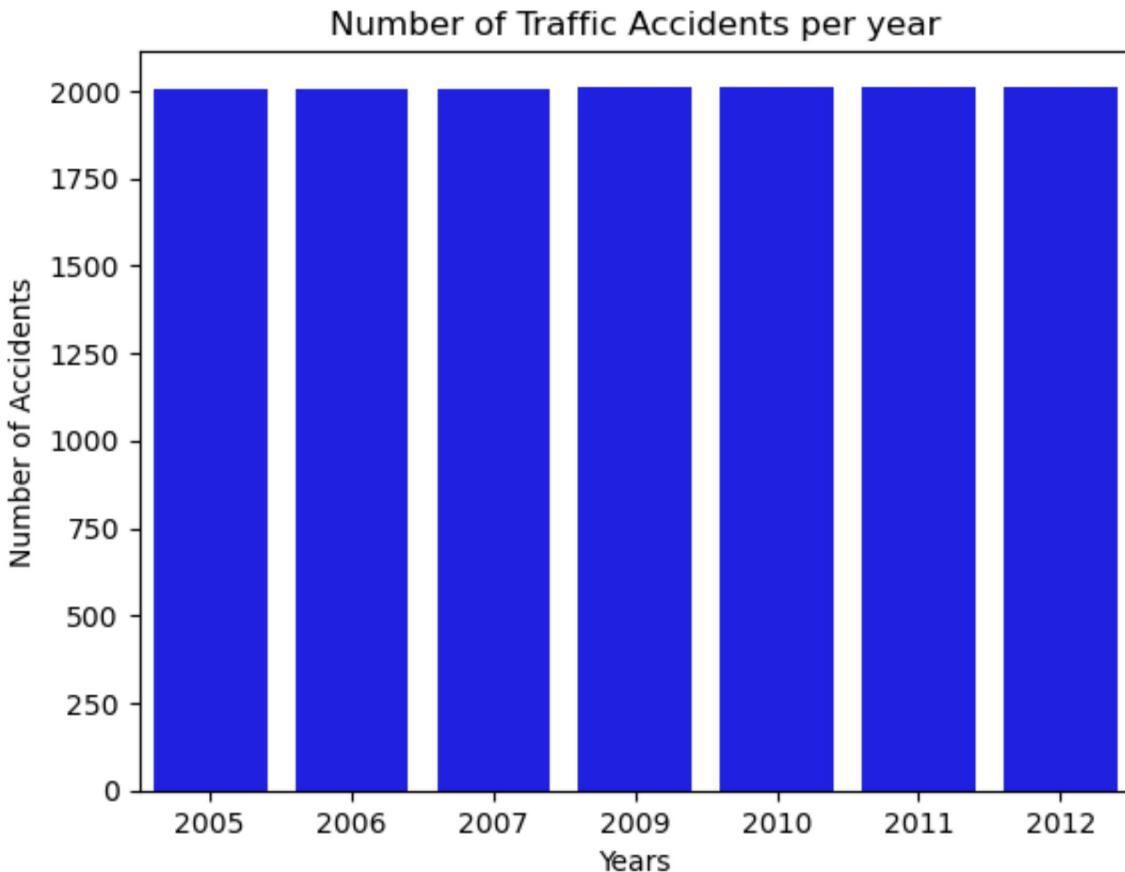
DATA VISUALIZATION

```
In [22]: #Accident rates by year
```

```
year=df1["Year"]
years=df1["Year"].unique()
num=year.values

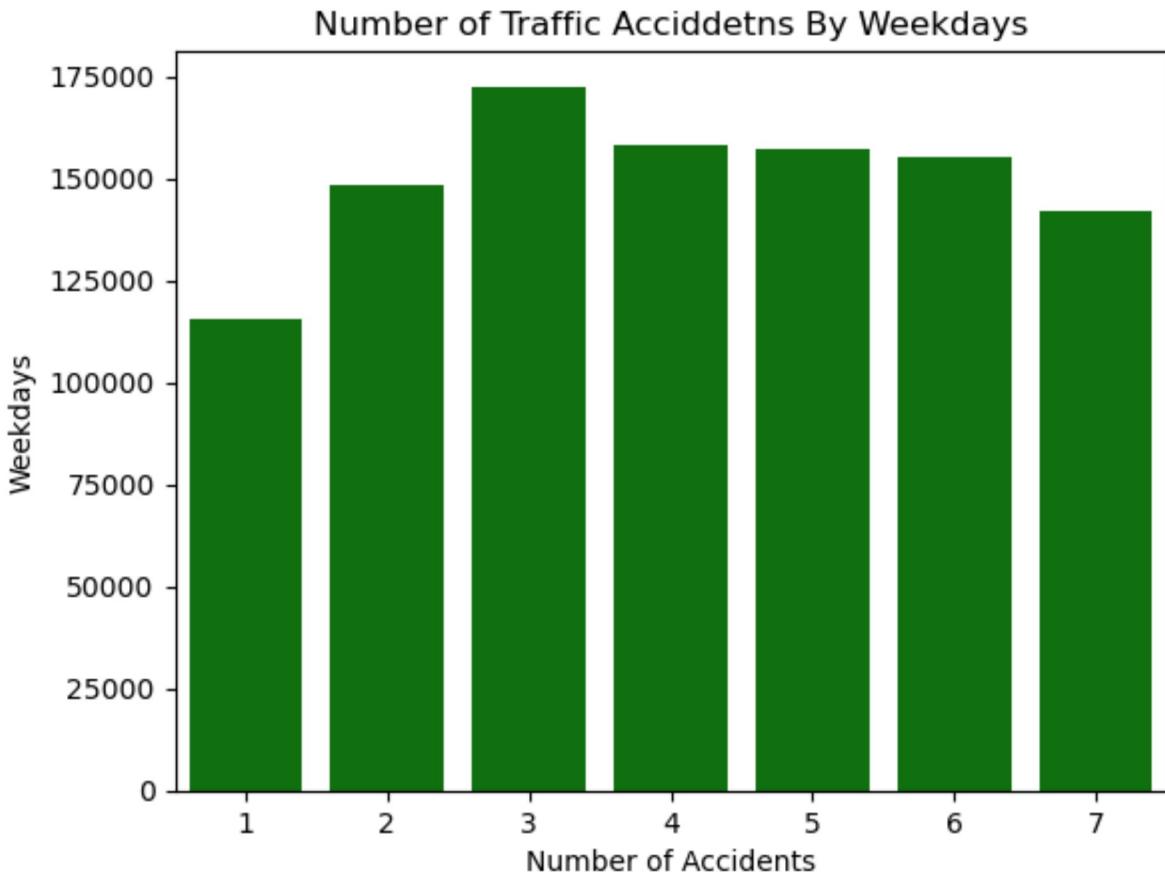
sns.barplot(x=year, y= num,color='blue')
plt.title("Number of Traffic Accidents per year")
plt.xlabel("Years")
plt.ylabel("Number of Accidents")
```

```
Out[22]: Text(0, 0.5, 'Number of Accidents')
```



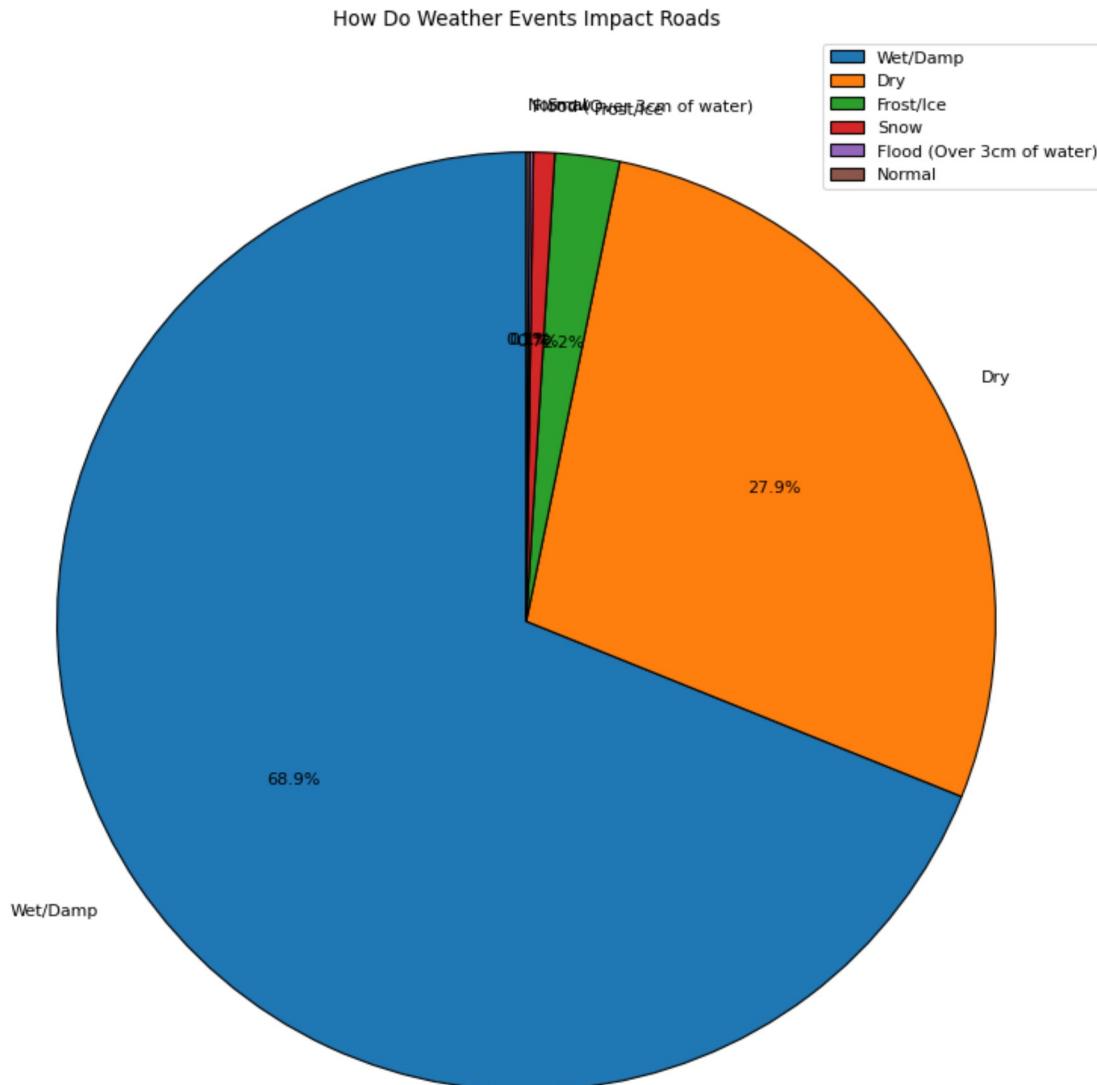
```
In [23]: #Accident rates by week
week = df1["Day_of_Week"].value_counts()
week
weeks = df1["Day_of_Week"].unique()
num_weeks = week.values
sns.barplot(x=weeks,y=num_weeks,color='green')
plt.title("Number of Traffic Accidetns By Weekdays")
plt.xlabel("Number of Accidents")
plt.ylabel("Weekdays")
```

```
Out[23]: Text(0, 0.5, 'Weekdays')
```



```
In [28]: #Accident rates due to road conditions
road_conditions = df1["Road_Surface_Conditions"].value_counts()
road_conditions_values = df1["Road_Surface_Conditions"].unique()

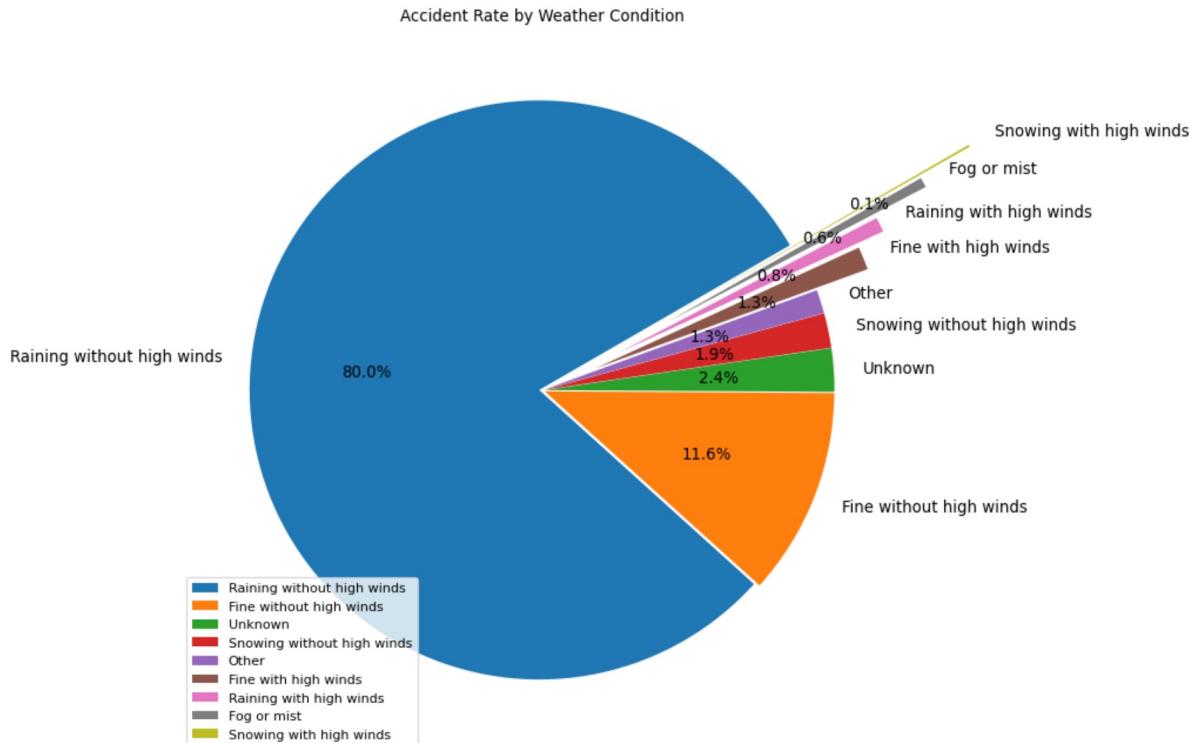
figure(figsize=(10, 10), dpi=80)
plt.pie(road_conditions, labels = road_conditions_values, autopct="%1.1f%%", wedgeprops
        startangle=90)
plt.tight_layout()
plt.legend()
plt.title("How Do Weather Events Impact Roads")
plt.show()
```



```
In [30]: #Accident rates due to weather conditions
weather_cond = df1["Weather_Conditions"].value_counts()
weather_cond_values= df1["Weather_Conditions"].unique()
weather_num_acc_arr = weather_cond.values

figure(figsize=(10, 10), dpi=80)
plt.pie(weather_cond, labels = weather_cond_values,startangle = 30, textprops={'size': 10})
plt.legend(loc ="lower left")
plt.title("Accident Rate by Weather Condition")
```

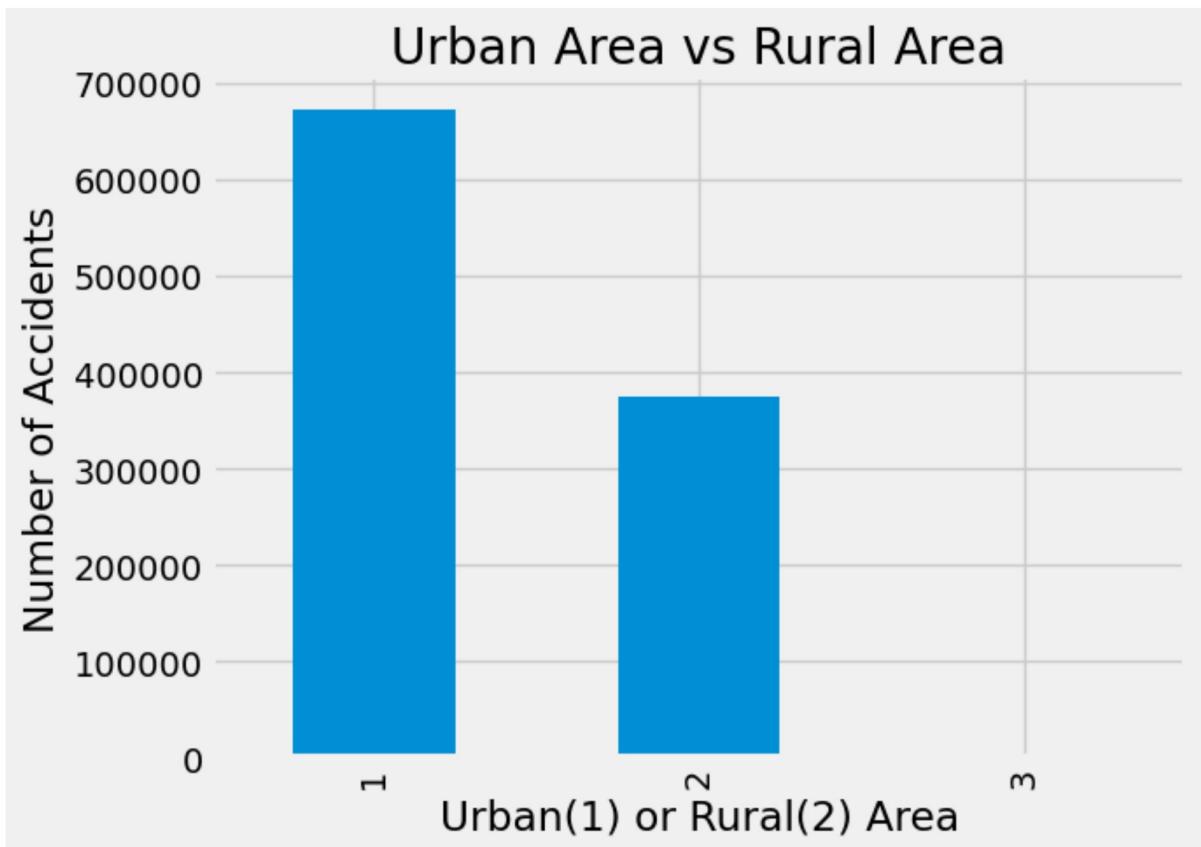
Out[30]: Text(0.5, 1.0, 'Accident Rate by Weather Condition')



In [31]: *#Accidents rates in urban and rural areas*

```
plt.style.use('fivethirtyeight')
df1["Urban_or_Rural_Area"].value_counts().plot(kind='bar')
plt.ylabel("Number of Accidents")
plt.xlabel("Urban(1) or Rural(2) Area")
plt.title("Urban Area vs Rural Area")
```

Out[31]: *Text(0.5, 1.0, 'Urban Area vs Rural Area')*



```
In [9]: #Number of injured in accidents and the severity of accidents
sns.barplot(x="Year",y="Number_of_Casualties",data=df1,hue="Accident_Severity")
plt.title("Accident_Severity")
plt.ylabel("Number of Casualties")
plt.legend(loc="upper right")
plt.show()
```

