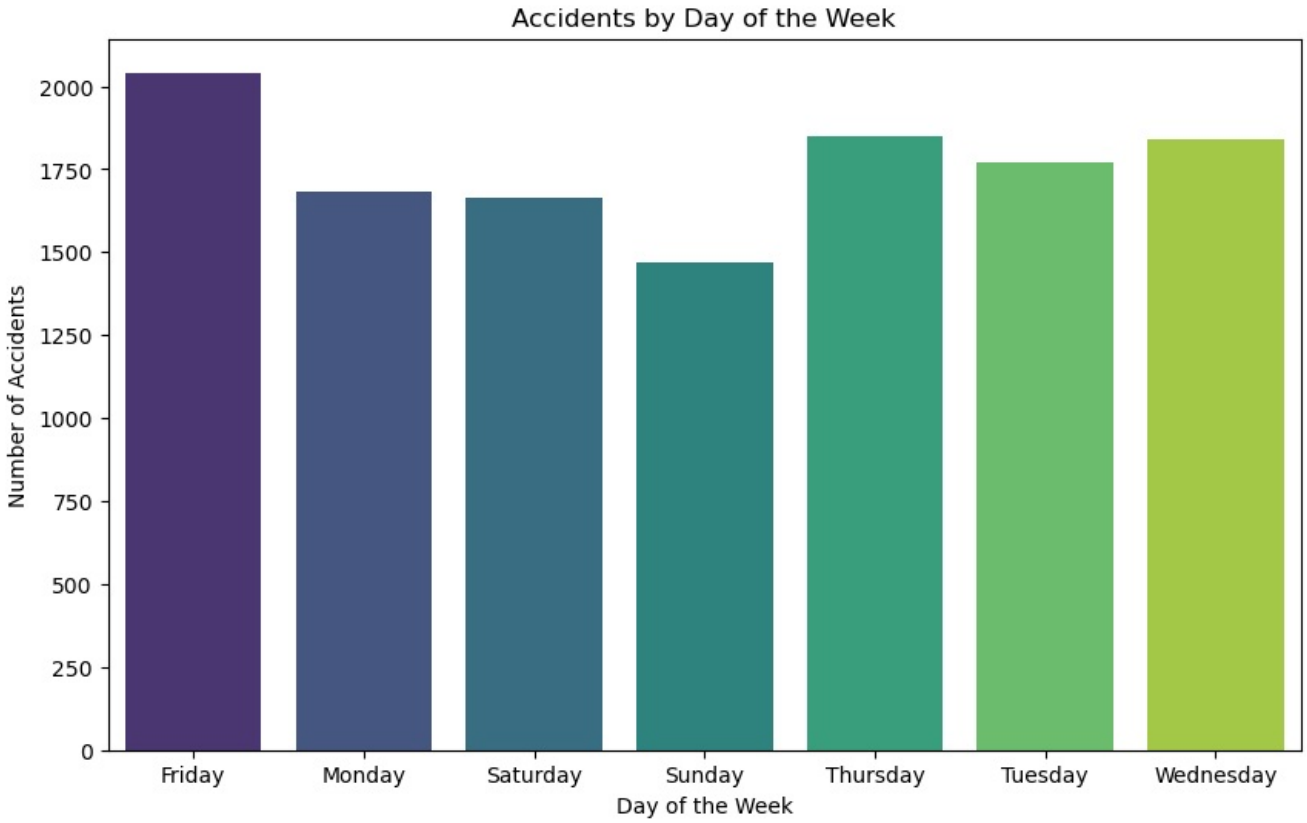
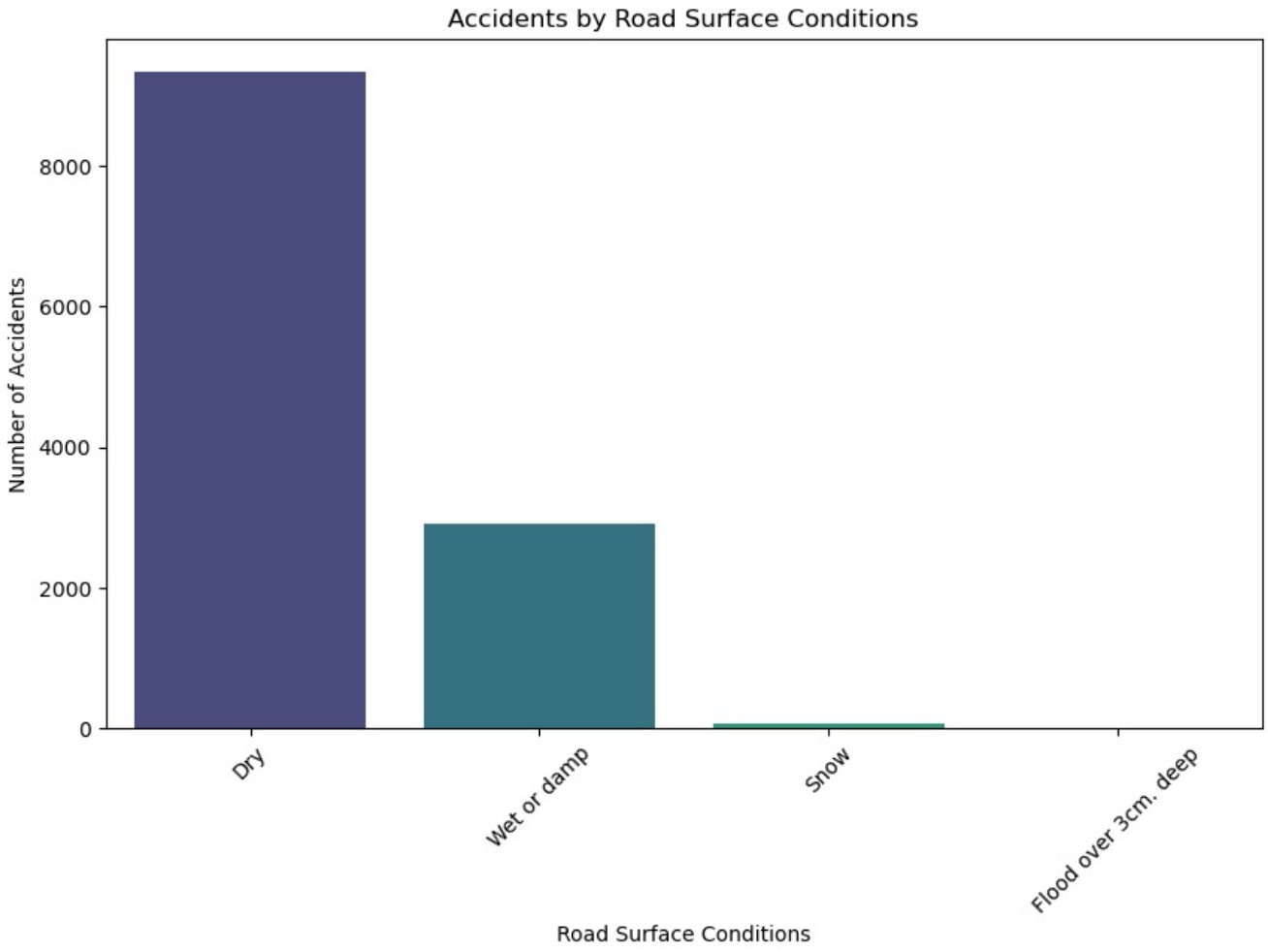
In [3]:

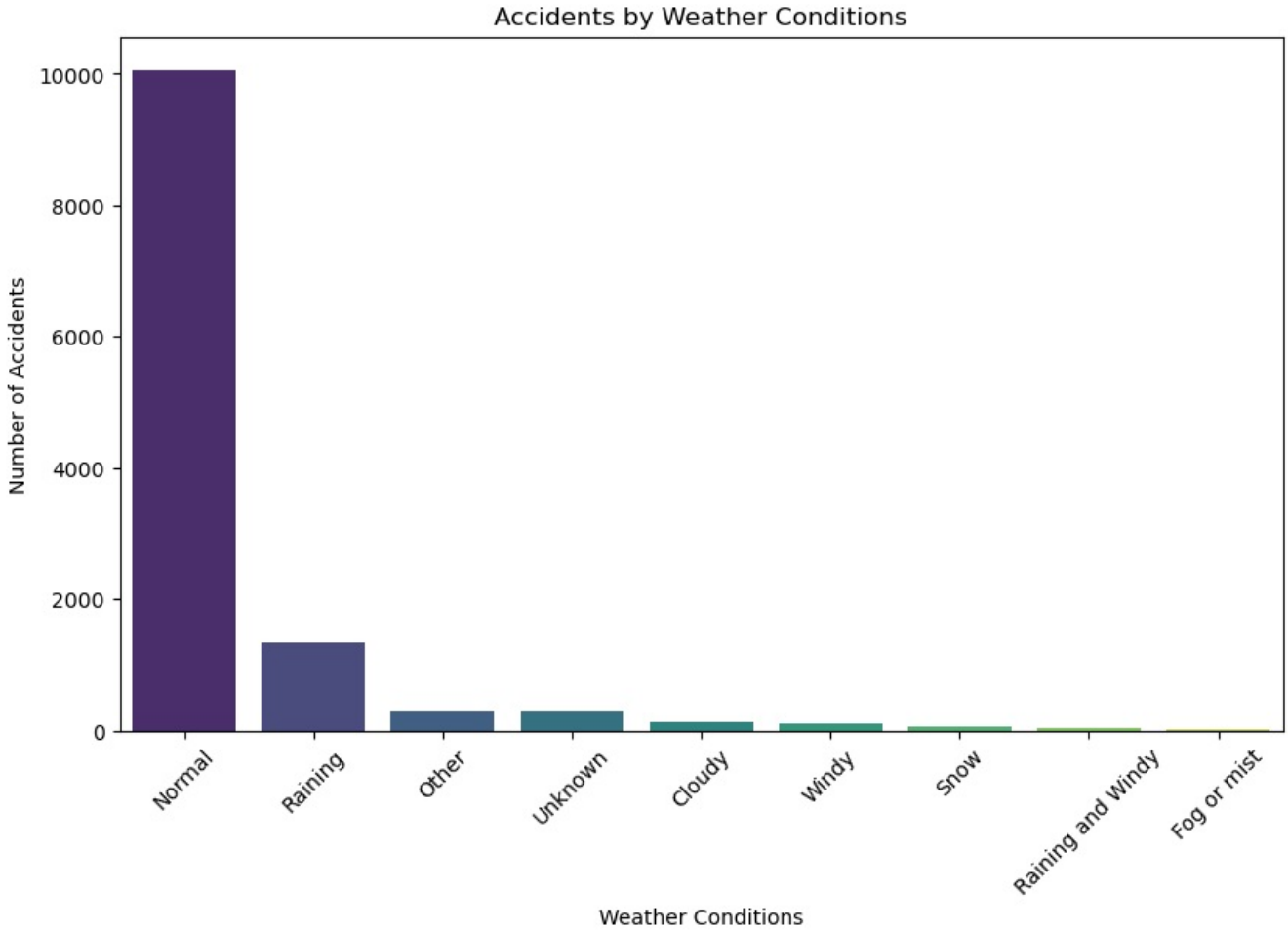
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import re

# Load the dataset
df = pd.read_csv('C:/Users/sathi/Downloads/RTA.csv')

# Display the first few rows of the dataset
print(df.head())

# Analyze accident counts by day of the week
accidents_by_day = df['Day_of_week'].value_counts().sort_index()

# Plot accidents by day of the week
plt.figure(figsize=(10, 6))
sns.barplot(x=accidents_by_day.index, y=accidents_by_day.values, palette='viridis')
plt.title('Accidents by Day of the Week')
plt.xlabel('Day of the Week')
plt.ylabel('Number of Accidents')
plt.show()

# Analyze accident counts by road surface conditions
accidents_by_road_surface = df['Road_surface_conditions'].value_counts()

# Plot accidents by road surface conditions
plt.figure(figsize=(10, 6))
sns.barplot(x=accidents_by_road_surface.index, y=accidents_by_road_surface.values, palette='viridis')
plt.title('Accidents by Road Surface Conditions')
plt.xlabel('Road Surface Conditions')
plt.ylabel('Number of Accidents')
plt.xticks(rotation=45)
plt.show()

# Analyze accident counts by weather conditions
accidents_by_weather = df['Weather_conditions'].value_counts()

# Plot accidents by weather conditions
plt.figure(figsize=(10, 6))
sns.barplot(x=accidents_by_weather.index, y=accidents_by_weather.values, palette='viridis')
plt.title('Accidents by Weather Conditions')
plt.xlabel('Weather Conditions')
plt.ylabel('Number of Accidents')
plt.xticks(rotation=45)
plt.show()

# Clean the 'Time' column
def clean_time(time_str):
    match = re.search(r'(\d{1,2}):(\d{2})', str(time_str))
    if match:
        return f"{int(match.group(1)):02}:{match.group(2)}"
    return None

df['Cleaned_Time'] = df['Time'].apply(clean_time)

# Drop rows with invalid time values
df = df.dropna(subset=['Cleaned_Time'])

# Convert the cleaned time to datetime
df['Cleaned_Time'] = pd.to_datetime(df['Cleaned_Time'], format='%H:%M')
df['Hour'] = df['Cleaned_Time'].dt.hour

# Analyze accident counts by time of day
accidents_by_hour = df['Hour'].value_counts().sort_index()

# Plot accidents by time of day
plt.figure(figsize=(10, 6))
sns.barplot(x=accidents_by_hour.index, y=accidents_by_hour.values, palette='viridis')
plt.title('Accidents by Time of Day')
plt.xlabel('Hour of Day')
plt.ylabel('Number of Accidents')
plt.show()

# Identify and visualize accident hotspots (assuming 'Area_accident_occured' column exists)
accidents_by_area = df['Area_accident_occured'].value_counts().head(10)

# Plot accident hotspots
plt.figure(figsize=(10, 6))
sns.barplot(x=accidents_by_area.index, y=accidents_by_area.values, palette='viridis')
plt.title('Top 10 Accident Hotspots')
plt.xlabel('Area')
plt.ylabel('Number of Accidents')
```

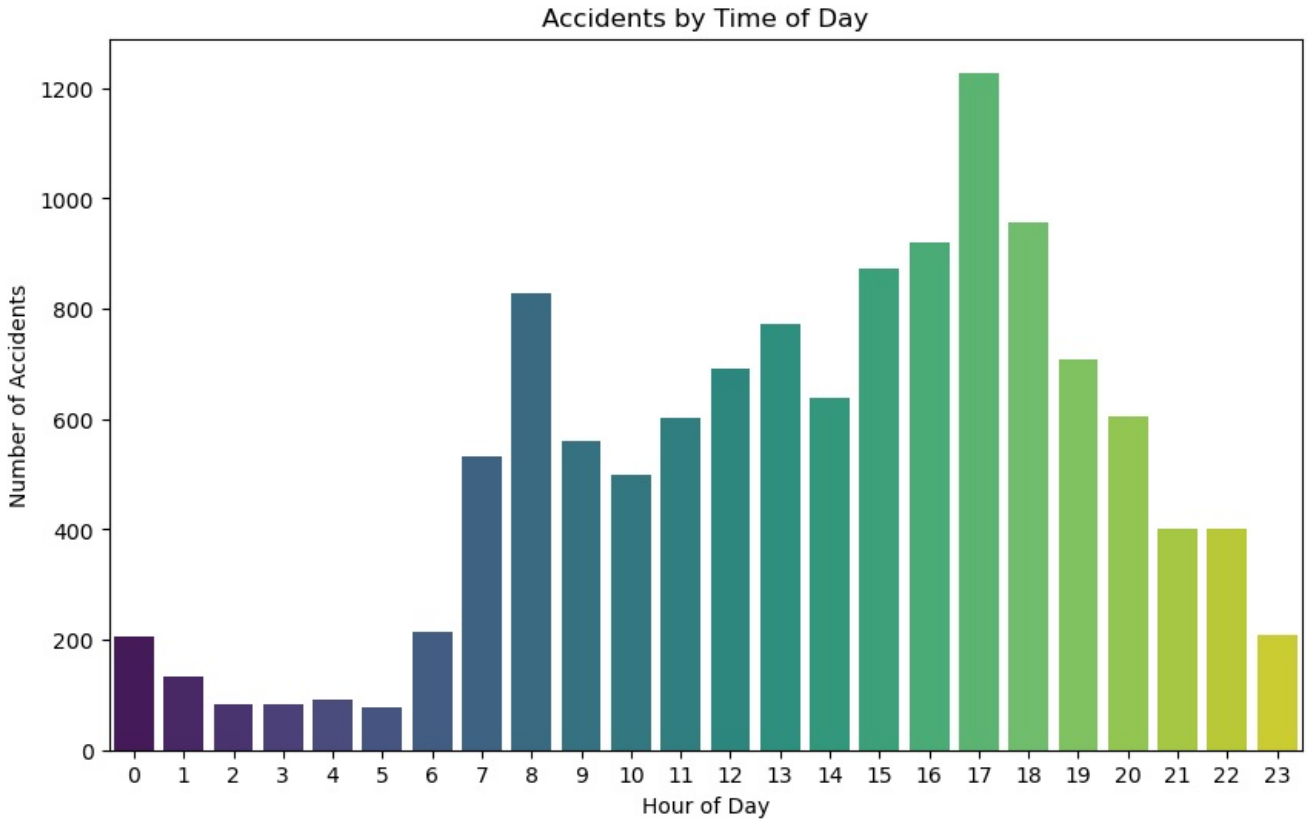
```
plt.xticks(rotation=45)
plt.show()

# Analyze contributing factors (assuming 'Cause_of_accident' column exists)
accidents_by_cause = df['Cause_of_accident'].value_counts().head(10)

# Plot contributing factors
plt.figure(figsize=(10, 6))
sns.barplot(x=accidents_by_cause.index, y=accidents_by_cause.values, palette='viridis')
plt.title('Top 10 Contributing Factors to Accidents')
plt.xlabel('Cause of Accident')
plt.ylabel('Number of Accidents')
plt.xticks(rotation=45)
plt.show()
```

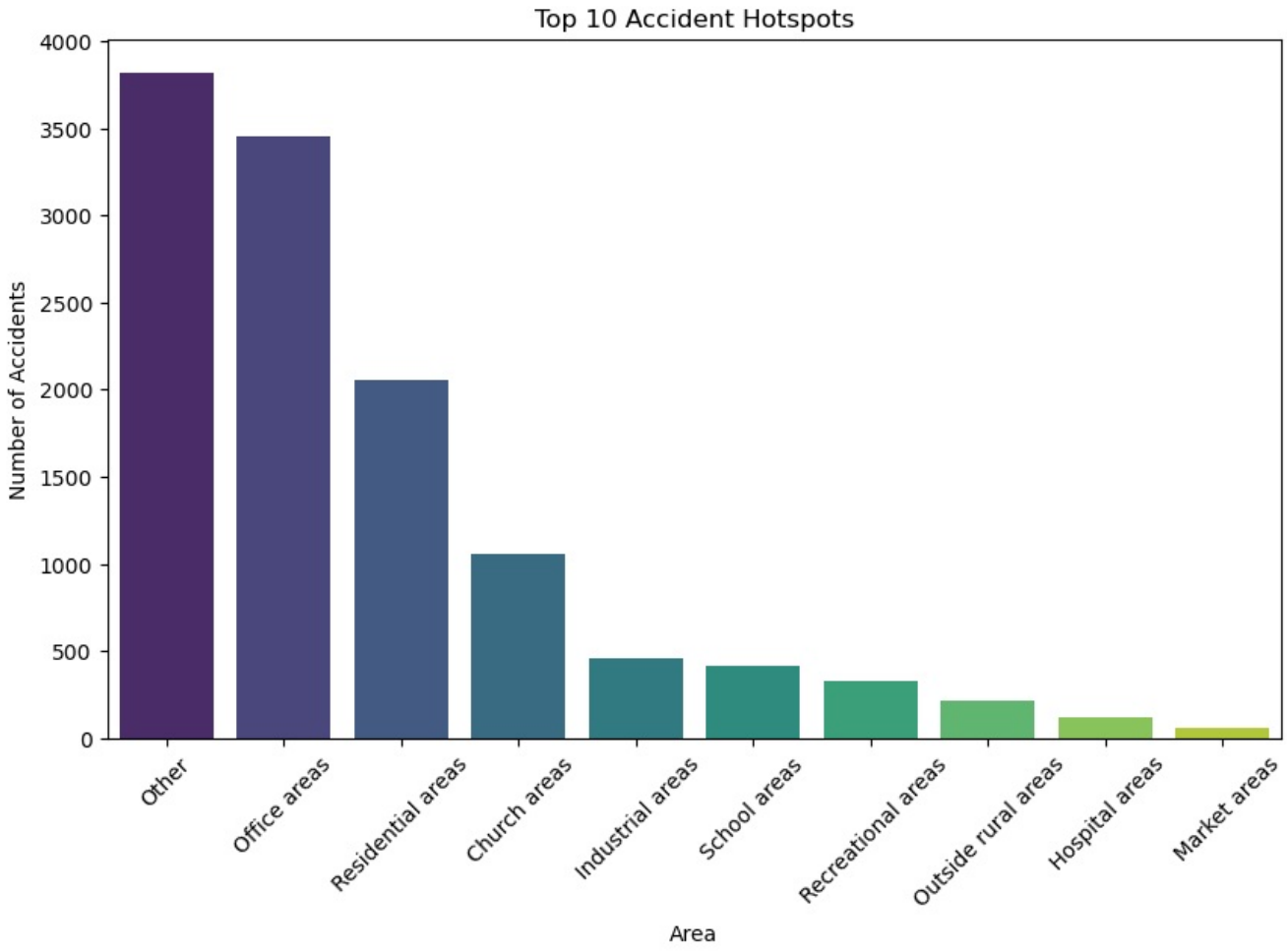
```
       Time Day_of_week Age_band_of_driver Sex_of_driver   Educational_level  \
0  17:02:00      Monday              18-30          Male   Above high school
1  17:02:00      Monday              31-50          Male  Junior high school
2  17:02:00      Monday              18-30          Male  Junior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

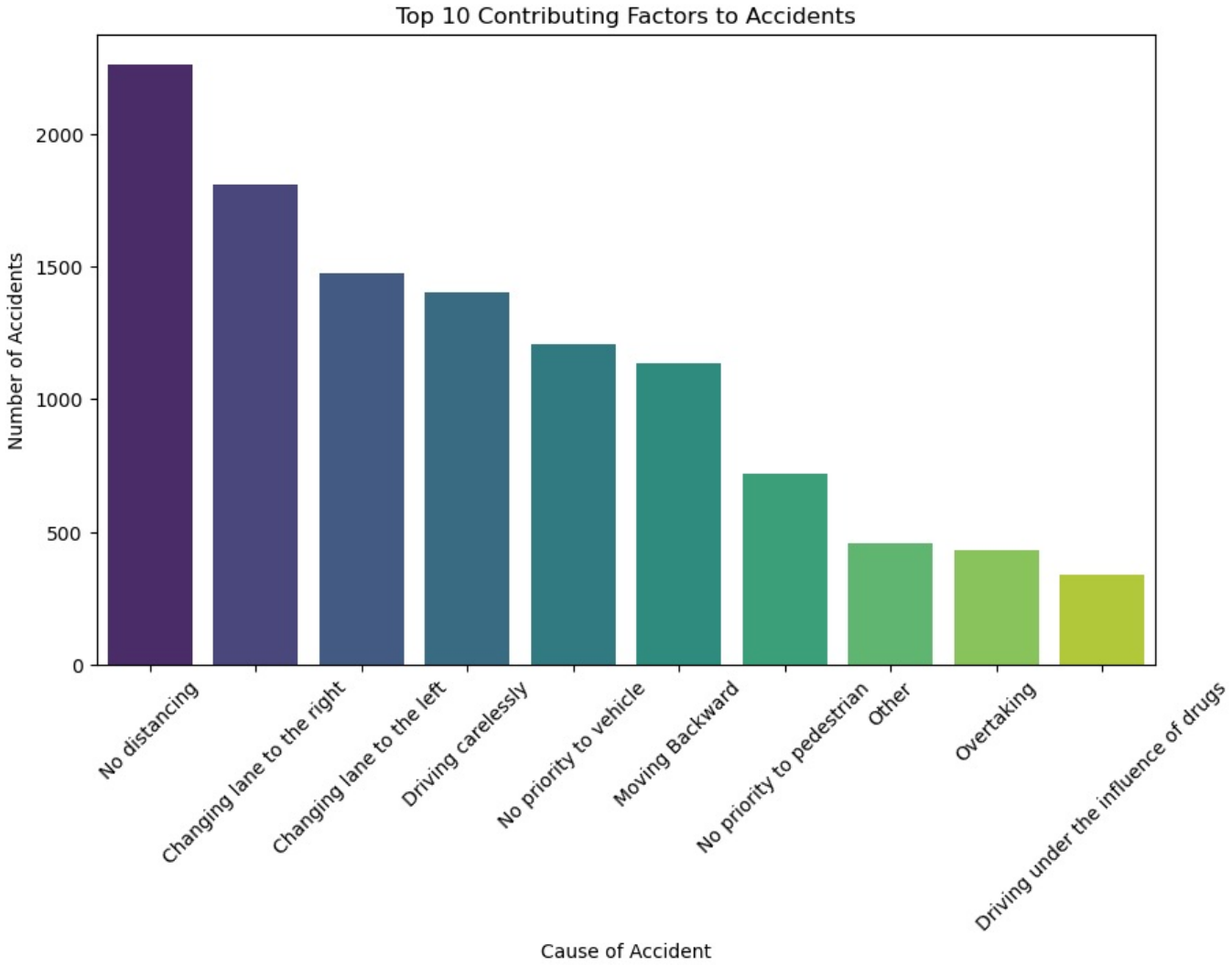

Accidents by Day of the Week
Number of Accidents
0
250
500
750
1000
1250
1500
1750
2000
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Day of the Week


Accidents by Road Surface Conditions
Number of Accidents
8000
6000
4000
2000
0
Dry
Wet or damp
Snow
Flood over 3cm. deep
Road Surface Conditions


Accidents by Weather Conditions
Number of Accidents
10000
8000
6000
4000
2000
0
Normal
Raining
Other
Unknown
Cloudy
Windy
Snow
Raining and Windy
Fog or mist
Weather Conditions


Accidents by Time of Day
Number of Accidents
0
200
400
600
800
1000
1200
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hour of Day


Top 10 Accident Hotspots
Number of Accidents
4000
3500
3000
2500
2000
1500
1000
500
0
Other
Office areas
Residential areas
Church areas
Industrial areas
School areas
Recreational areas
Outside rural areas
Hospital areas
Market areas
Area



In [ ]: