Task 5

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

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

```
df = pd.read_csv("/content/RTA Dataset.csv")
```

```
df.head()
```

	Time	Day_of_week	Age_band_of_driver	Sex_of_driver	Educational_level	Vehicle_driver_relation	Driving_ex
0	17:02:00	Monday	18-30	Male	Above high school	Employee	
1	17:02:00	Monday	31-50	Male	Junior high school	Employee	A
2	17:02:00	Monday	18-30	Male	Junior high school	Employee	
3	1:06:00	Sunday	18-30	Male	Junior high school	Employee	
4	1:06:00	Sunday	18-30	Male	Junior high school	Employee	

5 rows × 32 columns

```
df.tail()
```

	Time	Day_of_week	Age_band_of_driver	Sex_of_driver	Educational_level	Vehicle_driver_relation	Driving
8993	16:12:00	Wednesday	31-50	Male	Junior high school	Employee	
8994	16:12:00	Wednesday	Over 51	Male	Junior high school	Employee	
8995	16:35:00	Friday	18-30	Male	Elementary school	Owner	
8996	16:35:00	Friday	18-30	Male	Junior high school	Employee	
8997	16:35:00	Friday	18-30	Male	Junior high school	Employee	

5 rows × 32 columns

```
df.info()
```

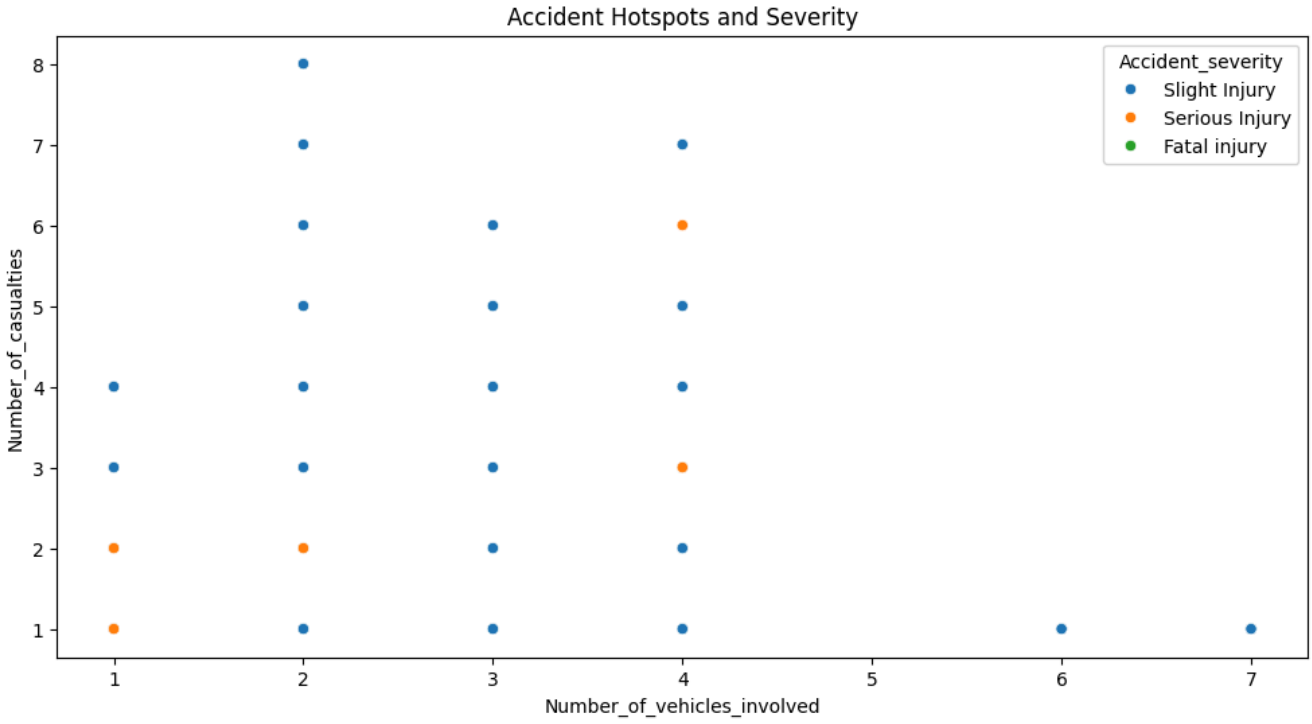
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8998 entries, 0 to 8997
Data columns (total 32 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Time                        8998 non-null   object
 1   Day_of_week                 8998 non-null   object
 2   Age_band_of_driver          8998 non-null   object
 3   Sex_of_driver               8998 non-null   object
 4   Educational_level           8477 non-null   object
 5   Vehicle_driver_relation     8621 non-null   object
 6   Driving_experience          8400 non-null   object
 7   Type_of_vehicle             8283 non-null   object
 8   Owner_of_vehicle            8647 non-null   object
 9   Service_year_of_vehicle     6041 non-null   object
 10  Defect_of_vehicle           5710 non-null   object
 11  Area_accident_occured       8822 non-null   object
 12  Lanes_or_Medians            8702 non-null   object
 13  Road_allignment             8884 non-null   object
```

```
 14  Types_of_Junction            8997 non-null   object
 15  Road_surface_type            8871 non-null   object
 16  Road_surface_conditions      8997 non-null   object
 17  Light_conditions             8997 non-null   object
 18  Weather_conditions           8997 non-null   object
 19  Type_of_collision            8892 non-null   object
 20  Number_of_vehicles_involved  8997 non-null   float64
 21  Number_of_casualties         8997 non-null   float64
 22  Vehicle_movement             8787 non-null   object
 23  Casualty_class               8997 non-null   object
 24  Sex_of_casualty              8997 non-null   object
 25  Age_band_of_casualty         8997 non-null   object
 26  Casualty_severity            8997 non-null   object
 27  Work_of_casuality            6646 non-null   object
 28  Fitness_of_casuality         7055 non-null   object
 29  Pedestrian_movement          8997 non-null   object
 30  Cause_of_accident            8997 non-null   object
 31  Accident_severity            8997 non-null   object
dtypes: float64(2), object(30)
memory usage: 2.2+ MB
```

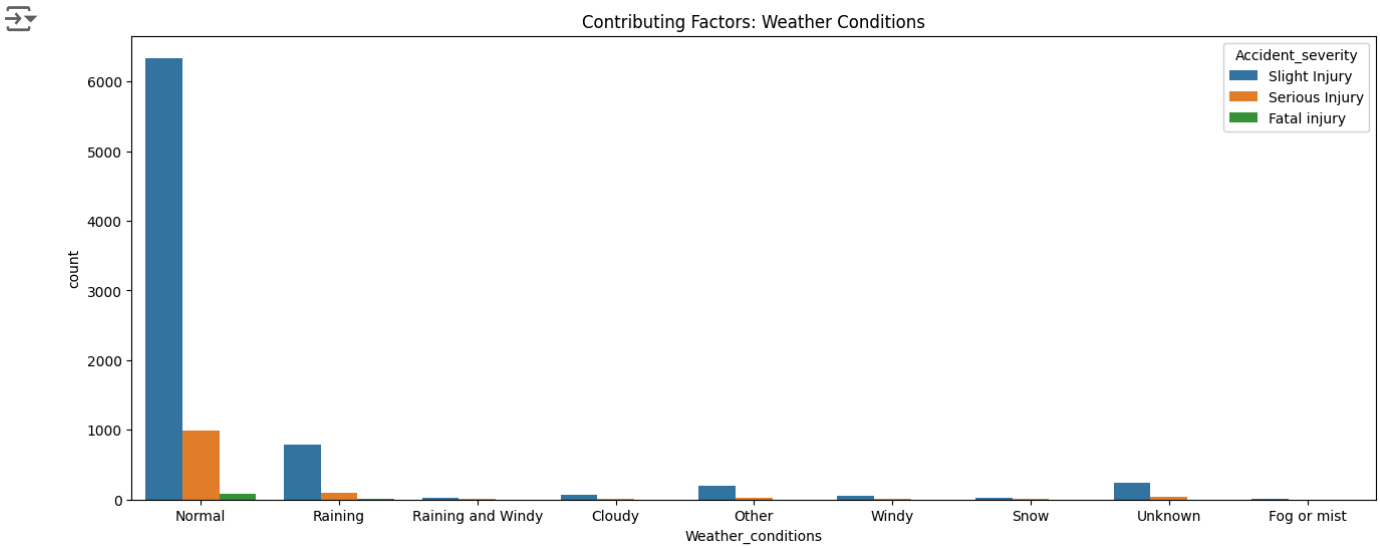
```
df.describe()
```

	Number_of_vehicles_involved	Number_of_casualties
count	8997.000000	8997.000000
mean	2.018673	1.522952
std	0.641743	1.004022
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

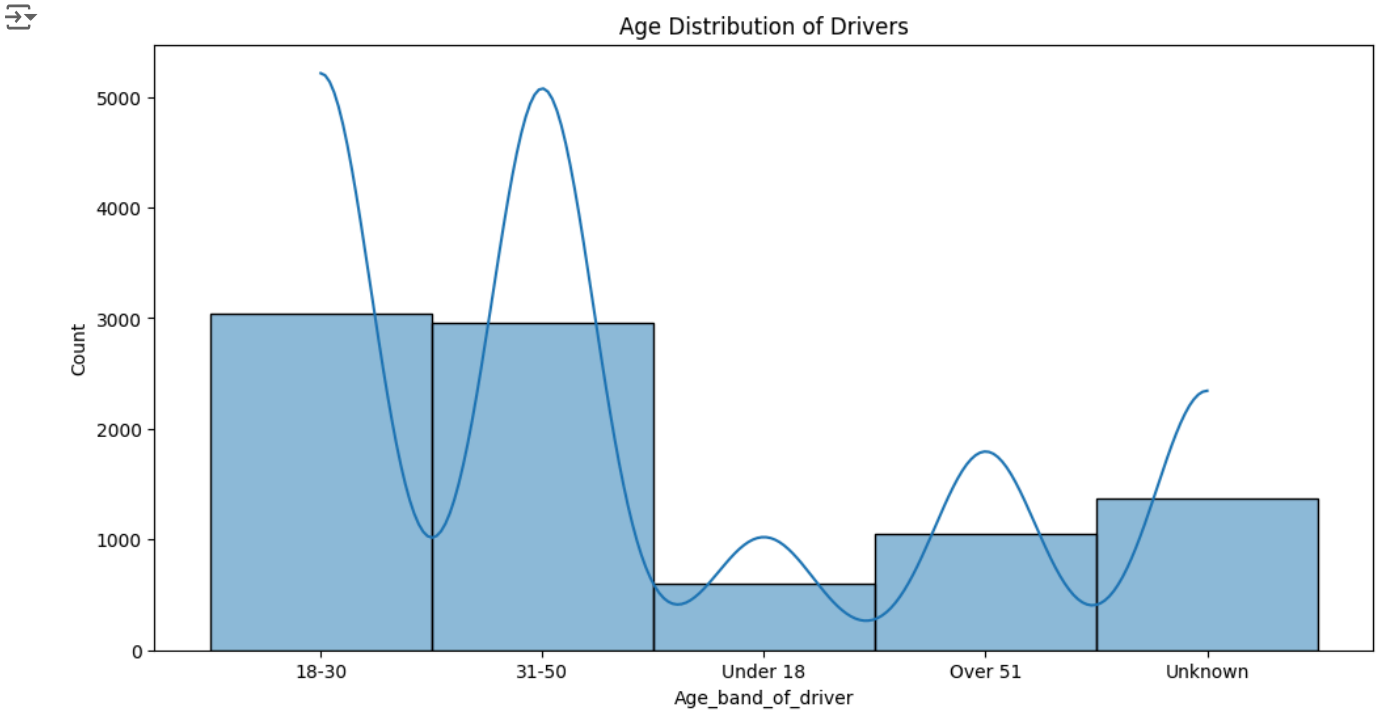
```
# Visualize accident hotspots
plt.figure(figsize=(12, 6))
sns.scatterplot(x='Number_of_vehicles_involved', y='Number_of_casualties', hue='Accident_severity', data=df)
plt.title('Accident Hotspots and Severity')
plt.show()
```



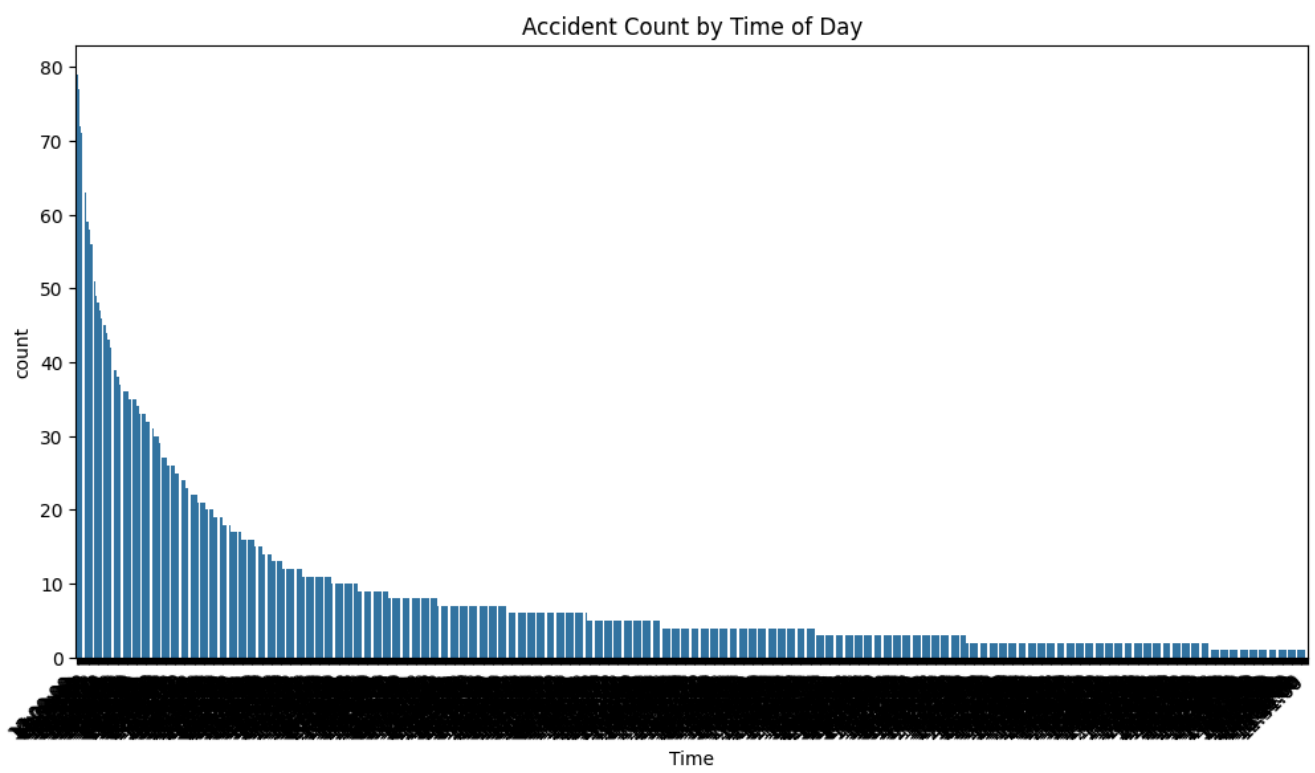
```
# Visualize contributing factors
plt.figure(figsize=(16, 6))
sns.countplot(x='Weather_conditions', hue='Accident_severity', data=df)
plt.title('Contributing Factors: Weather Conditions')
plt.show()
```



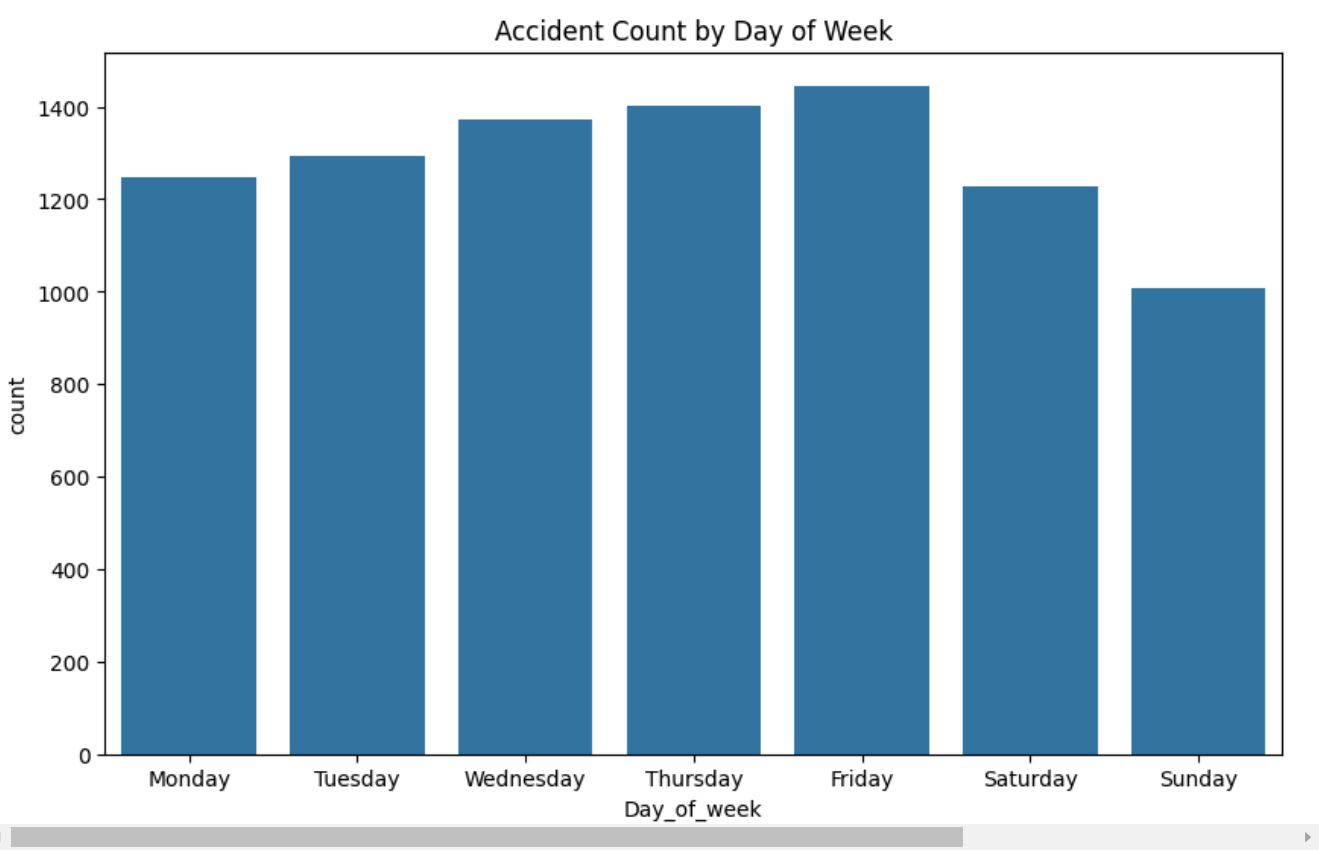

```
# Visualize Age distribution of drivers
plt.figure(figsize=(12, 6))
sns.histplot(x='Age_band_of_driver', data=df, bins=20, kde=True)
plt.title('Age Distribution of Drivers')
plt.show()
```



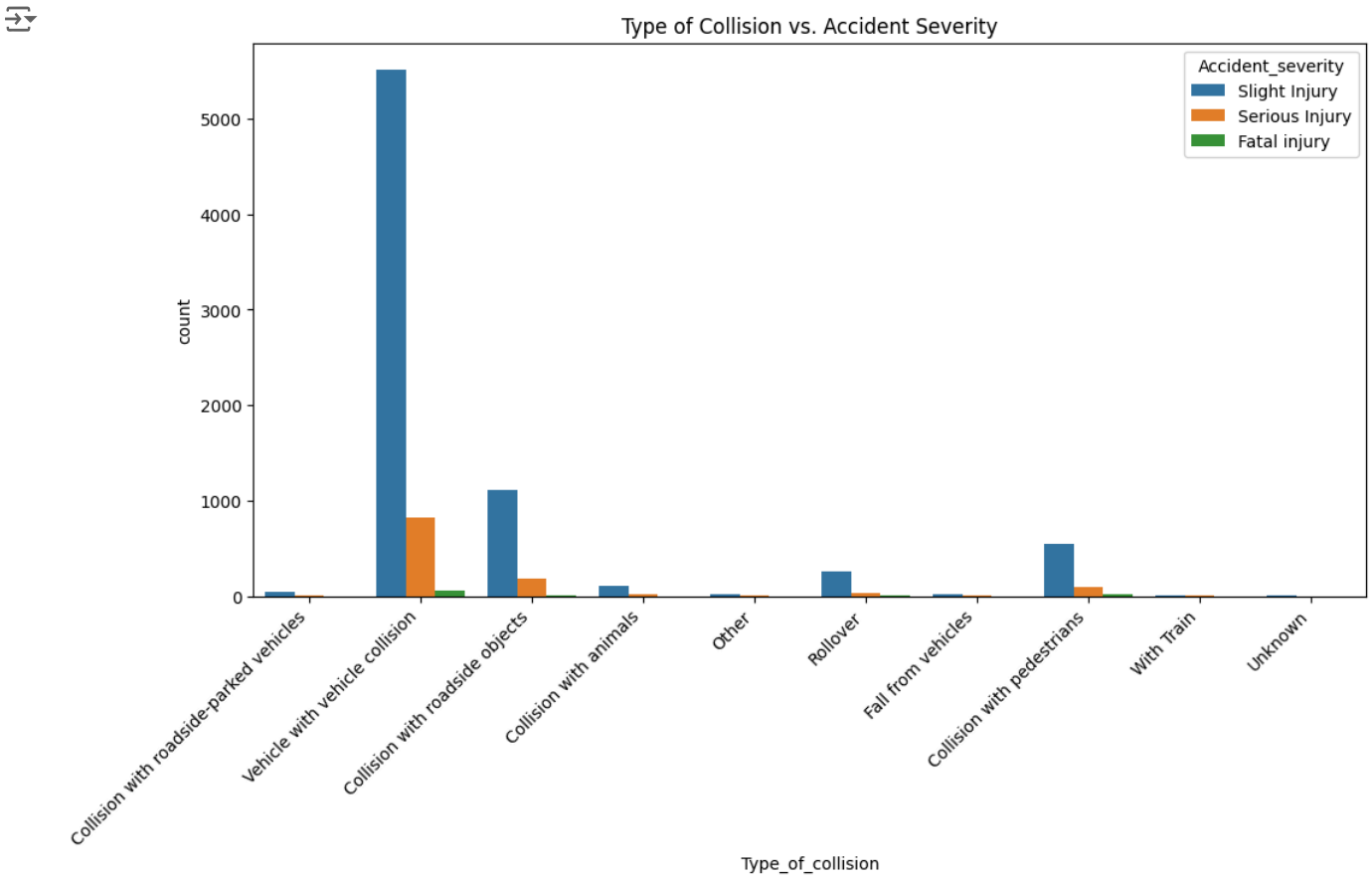

```
# Visualize Time of day vs. Accident count
plt.figure(figsize=(12, 6))
sns.countplot(x='Time', data=df, order=df['Time'].value_counts().index)
plt.title('Accident Count by Time of Day')
plt.xticks(rotation=45, ha='right')
plt.show()
```



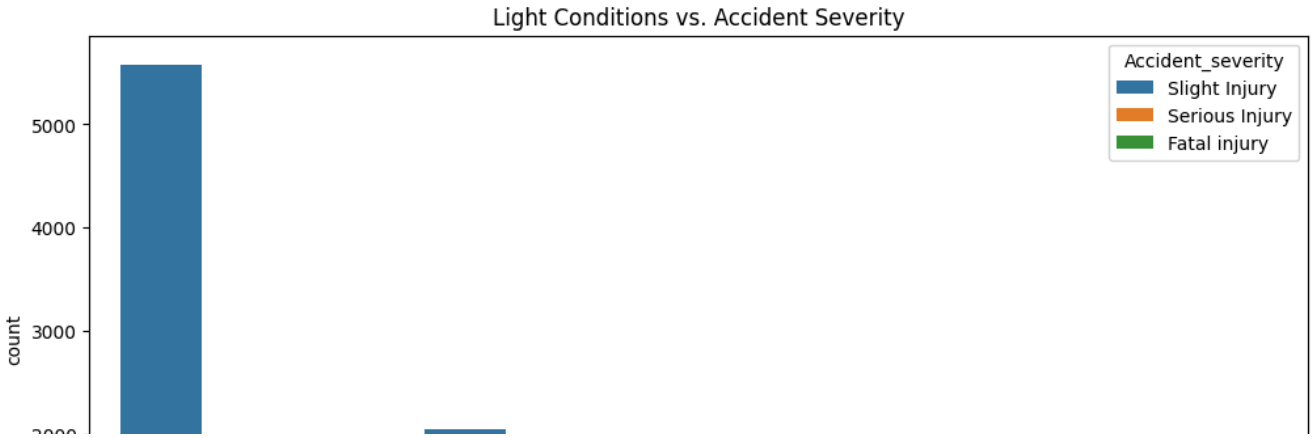

```
# Visualize Day of week vs. Accident count
plt.figure(figsize=(10, 6))
sns.countplot(x='Day_of_week', data=df, order=['Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday', 'Su
plt.title('Accident Count by Day of Week')
plt.show()
```



```
# Visualize Type of Collision vs. Accident severity
plt.figure(figsize=(12, 6))
sns.countplot(x='Type_of_collision', hue='Accident_severity', data=df)
plt.title('Type of Collision vs. Accident Severity')
plt.xticks(rotation=45, ha='right')
plt.show()
```




```
# Visualize Light conditions vs. Accident severity
plt.figure(figsize=(12, 6))
sns.countplot(x='Light_conditions', hue='Accident_severity', data=df)
plt.title('Light Conditions vs. Accident Severity')
plt.show()
```


Light Conditions vs. Accident Severity
Accident_severity
Slight Injury
Serious Injury
Fatal injury
count
5000
4000
3000