Kaggle Competition Data Analysis and Accident Prediction Strategies for Road Accidents in Jordan

1.1 PROJECT SUMMARY:

This project focuses on analyzing accident data from Jordan through Exploratory Data Analysis (EDA) techniques. The aim is to identify key patterns related to accident severity, dates, time of day, road and weather conditions, alcohol involvement, and driver demographics. Based on the findings, strategic recommendations are proposed to reduce accidents, especially focusing on year-round conditions and peak construction zones.

```
import numpy as np
import pandas as pd
import matplotlib as mpl
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

```
df = pd.read_csv('/content/Accidents_Jordan.csv')
df.drop(columns=['Unnamed: 0'], inplace=True)
```

```
df.shape
```

```
(1000, 15)
```

	Date Name	City Name	Year	Month	Day of Week	Time of Day	Accident Severity	Vehicle Type Involved	Number of Casualties	Road Type	Road Condition	Lighting Conditions	Traffic Control Presence	Speed Limit (km/h)	Driver Age	Driver Gender	Driver Nationality	Alcohol involvement	Accident Location Details
0	Jeans and Kaftans	Irbid	2021	May	Monday	1:45	Serious	Car	1	Two-Way	National Highway	Dark	Signs	70	45	Male	Jordan	No	Intersection
1	Cake Festival	Ajloun	2016	January	Wednesday	17:30	Minor	1	Truck	1	Urban Road	City	Clear	Signs	50	30	Male	Jordan	Yes	Straight Road
2	Chicken Festival	Jerash	2020	May	Wednesday	8:07	Minor	1	Pedestrian	4	National Highway	Clear	Signs	120	18	Female	Jordan	No	Bridge
3	Cake Festival	Madaba	2018	June	Saturday	12:51	Minor	1	Bus	10	Two-Way	Dark	Signs	70	18	Female	Jordan	Yes	Straight Road
4	Dibbin	Zarqa	2021	August	Thursday	11:21	Minor	1	Cycle	2	Urban Road	Wet	Clear	Signs	110	30	Male	Jordan	No	Intersection
2000	Tent Rental	Ajloun	2021	January	Sunday	1:15	Minor	1	Truck	4	National Highway	Wet	Clear	Signs	70	40	Male	Expat	Yes	Intersection
2006	Ultrasound	Jerash	2018	July	Sunday	10:12	Fatal	1	Car	1	Urban Road	Dark	Clear	Signs	90	23	Female	Jordan	No	Intersection
2007	Weightlifting	Jerash	2021	January	Thursday	10:34	Minor	2	Two-Wheeler	4	Two-Way	City	Clear	Signs	47	19	Female	Jordan	No	Intersection
2008	Weightlifting	Jerash	2020	June	Monday	19:00	Minor	1	Car	1	Urban Road	Dark	Clear	Signs	50	20	Female	Expat	Yes	Bridge
2009	Accredited Financial	Irbid	2020	September	Monday	7:19	Fatal	1	Cycle	1	National Highway	Dark	Clear	Signs	40	36	Male	Jordan	No	Bridge

2000 rows × 15 columns

```
df.info()
```

```
<class> pandas.core.frame.DataFrame
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 15 columns):
 #   Date Name                  2000 non-null   object
 #   City Name                  2000 non-null   object
 #   Year                       2000 non-null   int64
 #   Month                      2000 non-null   object
 #   Day of Week                2000 non-null   object
 #   Time of Day                2000 non-null   object
 #   Accident Severity          2000 non-null   object
 #   Vehicle Type Involved      2000 non-null   object
 #   Number of Casualties       2000 non-null   int64
 #   Road Type                  2000 non-null   object
 #   Road Condition             2000 non-null   object
 #   Lighting Conditions        2000 non-null   object
 #   Traffic Control Presence   2000 non-null   object
 #   Speed Limit (km/h)         2000 non-null   int64
 #   Driver Age                 2000 non-null   int64
 #   Driver Gender              2000 non-null   object
 #   Driver Nationality         2000 non-null   object
 #   Alcohol involvement        2000 non-null   object
 #   Accident Location Details  2000 non-null   object
dtypes: int64(4), object(11)
memory usage: 234.4+ KB
```

```
df.isnull().sum()
```

```
Date Name                    0
City Name                    0
Year                         0
Month                        0
Day of Week                  0
Time of Day                  0
Accident Severity            0
Vehicle Type Involved        0
Number of Casualties         0
Road Type                    0
Road Condition               0
Lighting Conditions          0
Traffic Control Presence     0
Speed Limit (km/h)           0
Driver Age                   0
Driver Gender                0
Driver Nationality           0
Alcohol involvement          0
Accident Location Details    0
dtype: int64
```

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

```
0
```

```
df.describe(include='object')
```

	Date Name	City Name	Month	Day of Week	Time of Day	Accident Severity	Vehicle Type Involved	Road Type	Road Condition	Lighting Conditions	Traffic Control Presence	Driver Gender	Driver Nationality	Alcohol involvement	Accident Location Details
count	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000
unique	10	10	12	7	1440	3	10	4	4	4	4	3	3	2	3
top	Cake Festival	Irbid	July	Tuesday	17:30	Minor	1	Two-Way	Dark	Clear	Signs	Male	Jordan	No	Intersection
freq	208	208	180	294	3	1668	1668	1000	1000	1000	1000	1000	1000	1000	1000

```
df.describe(include='int64')
```

	Year	Number of Casualties	Road Condition	Lighting Conditions	Traffic Control Presence	Speed Limit (km/h)	Driver Age
count	2000.000000	2000.000000	2000.000000	2000.000000	2000.000000	2000.000000	2000.000000
mean	2020.000000	2.998000	0.998000	2.000000	1.500000	75.995000	34.177500
std	1.999500	1.420167	1.417597	1.417597	1.118034	19.750346	9.436388
min	2016.000000	1.000000	0.000000	0.000000	0.000000	30.000000	18.000000
25%	2018.000000	2.000000	0.000000	1.000000	0.750000	70.000000	26.000000
50%	2020.000000	3.000000	0.000000	2.000000	1.500000	70.000000	34.000000
75%	2022.000000	4.000000	2.000000	3.000000	2.250000	90.000000	43.000000
max	2024.000000	5.000000	3.000000	3.000000	3.000000	120.000000	70.000000

```
df.describe(include='float64')
```

```
KeyboardInterrupt: Traceback (most recent call last)
```

```
df.describe(include='all')
```

	Date Name	City Name	Year	Month	Day of Week	Time of Day	Accident Severity	Vehicle Type Involved	Number of Casualties	Road Type	Road Condition	Lighting Conditions	Traffic Control Presence	Speed Limit (km/h)	Driver Age	Driver Gender	Driver Nationality	Alcohol involvement	Accident Location Details
count	2000	2000	2000.000000	2000	2000	2000	2000	2000	2000.000000	2000	2000.000000	2000.000000	2000.000000	2000.000000	2000.000000	2000	2000	2000	2000
unique	10	10	NaN	12	7	1440	3	10	NaN	4	NaN	NaN	NaN	NaN	NaN	3	3	2	3
top	Cake Festival	Irbid	NaN	July	Tuesday	17:30	Minor	1	NaN	Two-Way	NaN	NaN	NaN	NaN	NaN	Male	Jordan	No	Intersection
freq	208	208	NaN	180	294	3	1668	1668	NaN	1000	NaN	NaN	NaN	NaN	NaN	1000	1000	1000	1000
mean	NaN	NaN	2020.000000	NaN	NaN	NaN	NaN	NaN	2.998000	NaN	0.998000	2.000000	1.500000	75.995000	34.177500	NaN	NaN	NaN	NaN
std	NaN	NaN	1.999500	NaN	NaN	NaN	NaN	NaN	1.420167	NaN	1.417597	1.417597	1.118034	19.750346	9.436388	NaN	NaN	NaN	NaN
min	NaN	NaN	2016.000000	NaN	NaN	NaN	NaN	NaN	1.000000	NaN	0.000000	0.000000	0.000000	30.000000	18.000000	NaN	NaN	NaN	NaN
25%	NaN	NaN	2018.000000	NaN	NaN	NaN	NaN	NaN	2.000000	NaN	0.000000	1.000000	0.750000	70.000000	26.000000	NaN	NaN	NaN	NaN
50%	NaN	NaN	2020.000000	NaN	NaN	NaN	NaN	NaN	3.000000	NaN	0.000000	2.000000	1.500000	70.000000	34.000000	NaN	NaN	NaN	NaN
75%	NaN	NaN	2022.000000	NaN	NaN	NaN	NaN	NaN	4.000000	NaN	2.000000	3.000000	2.250000	90.000000	43.000000	NaN	NaN	NaN	NaN
max	NaN	NaN	2024.000000	NaN	NaN	NaN	NaN	NaN	5.000000	NaN	3.000000	3.000000	3.000000	120.000000	70.000000	NaN	NaN	NaN	NaN

```
df.describe(include='datetime')
```

```
ValueError: Cannot select dtypes: no columns matched
```

```
df.describe(include=['object', 'int64', 'datetime64'])
```

	Date Name	City Name	Year	Month	Day of Week	Time of Day	Accident Severity	Vehicle Type Involved	Number of Casualties	Road Type	Road Condition	Lighting Conditions	Traffic Control Presence	Speed Limit (km/h)	Driver Age	Driver Gender	Driver Nationality	Alcohol involvement	Accident Location Details
count	2000	2000	2000.000000	2000	2000	2000	2000	2000	2000.000000	2000	2000.000000	2000.000000	2000.000000	2000.000000	2000.000000	2000	2000	2000	2000
unique	10	10	NaN	12	7	1440	3	10	NaN	4	NaN	NaN	NaN	NaN	NaN	3	3	2	3
top	Cake Festival	Irbid	NaN	July	Tuesday	17:30	Minor	1	NaN	Two-Way	NaN	NaN	NaN	NaN	NaN	Male	Jordan	No	Intersection
freq	208	208	NaN	180	294	3	1668	1668	NaN	1000	NaN	NaN	NaN	NaN	NaN	1000	1000	1000	1000
mean	NaN	NaN	2020.000000	NaN	NaN	NaN	NaN	NaN	2.998000	NaN	0.998000	2.000000	1.500000	75.995000	34.177500	NaN	NaN	NaN	NaN
std	NaN	NaN	1.999500	NaN	NaN	NaN	NaN	NaN	1.420167	NaN	1.417597	1.417597	1.118034	19.750346	9.436388	NaN	NaN	NaN	NaN
min	NaN	NaN	2016.000000	NaN	NaN	NaN	NaN	NaN	1.000000	NaN	0.000000	0.000000	0.000000	30.000000	18.000000	NaN	NaN	NaN	NaN
25%	NaN	NaN	2018.000000	NaN	NaN	NaN	NaN	NaN	2.000000	NaN	0.000000	1.000000	0.750000	70.000000	26.000000	NaN	NaN	NaN	NaN
50%	NaN	NaN	2020.000000	NaN	NaN	NaN	NaN	NaN	3.000000	NaN	0.000000	2.000000	1.500000	70.000000	34.000000	NaN	NaN	NaN	NaN
75%	NaN	NaN	2022.000000	NaN	NaN	NaN	NaN	NaN	4.000000	NaN	2.000000	3.000000	2.250000	90.000000	43.000000	NaN	NaN	NaN	NaN
max	NaN	NaN	2024.000000	NaN	NaN	NaN	NaN	NaN	5.000000	NaN	3.000000	3.000000	3.000000	120.000000	70.000000	NaN	NaN	NaN	NaN

```
df.describe(include=['object', 'int64', 'datetime64', 'bool', 'category'])
```

	Date Name	City Name	Year	Month	Day of Week	Time of Day	Accident Severity	Vehicle Type Involved	Number of Casualties	Road Type	Road Condition	Lighting Conditions	Traffic Control Presence	Speed Limit (km/h)	Driver Age	Driver Gender	Driver Nationality	Alcohol involvement	Accident Location Details
count	2000	2000	2000.000000	2000	2000	2000	2000	2000	2000.000000	2000	2000.000000	2000.000000	2000.000000	2000.000000	2000.000000	2000	2000	2000	2000
unique	10	10	NaN	12	7	1440	3	10	NaN	4	NaN	NaN	NaN	NaN	NaN	3	3	2	3
top	Cake Festival	Irbid	NaN	July	Tuesday	17:30	Minor	1	NaN	Two-Way	NaN	NaN	NaN	NaN	NaN	Male	Jordan	No	Intersection
freq	208	208	NaN	180	294	3	1668	1668	NaN	1000	NaN	NaN	NaN	NaN	NaN	1000	1000	1000	1000
mean	NaN	NaN	2020.000000	NaN	NaN	NaN	NaN	NaN	2.998000	NaN	0.998000	2.000000	1.500000	75.995000	34.177500	NaN	NaN	NaN	NaN
std	NaN	NaN	1.999500	NaN	NaN	NaN	NaN	NaN	1.420167	NaN	1.417597	1.417597	1.118034	19.750346	9.436388	NaN	NaN	NaN	NaN
min	NaN	NaN	2016.000000	NaN	NaN	NaN	NaN	NaN	1.000000	NaN	0.000000	0.000000	0.000000	30.000000	18.000000	NaN	NaN	NaN	NaN
25%	NaN	NaN	2018.000000	NaN	NaN	NaN	NaN	NaN	2.000000	NaN	0.000000	1.000000	0.750000	70.000000	26.000000	NaN	NaN	NaN	NaN
50%	NaN	NaN	2020.000000	NaN	NaN	NaN	NaN	NaN	3.000000	NaN	0.000000	2.000000	1.500000	70.000000	34.000000	NaN	NaN	NaN	NaN
75%	NaN	NaN	2022.000000	NaN	NaN	NaN	NaN	NaN	4.000000	NaN	2.000000	3.000000	2.250000	90.000000	43.000000	NaN	NaN	NaN	NaN
max	NaN	NaN	2024.000000	NaN	NaN	NaN	NaN	NaN	5.000000	NaN	3.000000	3.000000	3.000000	120.000000	70.000000	NaN	NaN	NaN	NaN

```
df.nunique()
```

```
Date Name                    10
City Name                    10
Year                          5
Month                        12
Day of Week                   7
Time of Day                1440
Accident Severity             3
Vehicle Type Involved        10
Number of Casualties          5
Road Type                     4
Road Condition                4
Lighting Conditions           4
Traffic Control Presence      4
Speed Limit (km/h)            6
Driver Age                   53
Driver Gender                 3
Driver Nationality            3
Alcohol involvement           2
Accident Location Details     3
dtype: int64
```

Time of Day is a high number of unique values (exact accident time) ; maybe better to group by hour later.

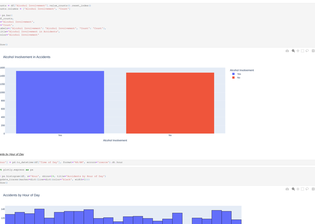
2 Visual Exploratory Data Analysis (EDA)

Accident Severity Distribution:

```
fig = px.pie(df, names='Accident Severity', title='Accident Severity Distribution')
fig.show()
```

Accidents by Date

```
df_count = df.groupby('Date Name').size().reset_index()
fig = px.bar(df_count, x='Date', y='Count', title='Accidents by Date')
fig.show()
```



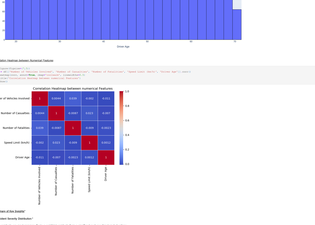
Accidents by Weather Condition

```
import plotly.express as px

df_count = df.groupby('Weather Condition').size().reset_index()
df_count.columns = ['Weather Condition', 'Count']

fig = px.bar(df_count, x='Weather', y='Count',
             color='Weather Condition',
             title='Weather Condition in Accidents',
             labels={'Weather': 'Weather', 'Count': 'Number of Accidents'},
             color_discrete=px.colors.qualitative.Pastel)

fig.show()
```



Accidents by Road Condition

```
df_count = df.groupby('Road Condition').size().reset_index()
df_count.columns = ['Road Condition', 'Count']

px.bar(df_count,
       x='Road Condition',
       y='Count',
       color='Road Condition',
       title='Road Condition During Accidents',
       labels={'Road Condition': 'Road Condition', 'Count': 'Accidents'},
       color_discrete=px.colors.qualitative.Pastel,
       text_auto=True)

fig.show()
```

Alcohol Involvement in Accidents

```
df_count = df.groupby('Alcohol involvement').size().reset_index()
df_count.columns = ['Alcohol Involvement', 'Count']

px.bar(df_count,
       x='Alcohol Involvement',
       y='Count',
       color='Alcohol Involvement',
       title='Alcohol Involvement in Accidents',
       labels={'Alcohol Involvement': 'Alcohol Involvement', 'Count': 'Count'},
       color_discrete=px.colors.qualitative.Pastel,
       text_auto=True)

fig.show()
```


Accidents By Hour of Day

```
df['Hour'] = pd.to_datetime(df['Time of Day'], format='%H:%M', errors='coerce').dt.hour

import plotly.express as px

fig = px.histogram(df, x='Hour', nbins=24, title='Accidents by Hour of Day')
fig.update_xaxis(dtick=1, tickprefix='', tickformat='d')
fig.show()
```


Driver's Age Distribution

```
fig = px.histogram(df, x='Driver Age', nbins=20, title='Driver's Age Distribution')
fig.update_xaxis(dtick=5, tickprefix='', tickformat='d')
fig.show()
```


Correlation Heatmap between Numerical Features

```
df.select_dtypes(include=['int64', 'float64', 'datetime64', 'bool', 'category'].
corr = df[['Number of Casualties', 'Road Condition', 'Lighting Conditions', 'Traffic Control Presence', 'Speed Limit (km/h)', 'Driver Age']].corr()
fig = px.imshow(corr, text_auto=True, aspect="auto", color_continuous_scale='RdBu_r', title='Correlation Heatmap between Numerical Features')
fig.show()
```


"Summary of Key Insights"

1 Accident Severity Distribution:

Minor accidents are most common. Serious and fatal accidents form a significant part, meaning targeted actions.

2 High Accident Dates:

Cake Festival and Eid days show the highest accident counts.

3 Weather Conditions:

Most accidents happen in fairly weather. Snowy and rainy conditions also significantly contribute.

4 Road Conditions:

Majority of accidents occur on under-construction road, but and dry roads and damaged roads are very low.

5 Alcohol Involvement:

A considerable number of accidents involve alcohol. Strong need for anti-drunk-driving measures.

6 Time of Day:

More accidents happen around 17:00, 8:00 and 18:00. Peak hours are critical for accident prevention.

7 Driver Demographics:

Most drivers involved are aged between 18 to 70, 30 to 45 years. Targeted awareness campaigns needed for this group.

8 Correlations:

Number of hospitalized is positively correlated with Number of Casualties. No strong correlation with speed limit or driver age directly.

Recommendations to Reduce Road Accidents in Jordan

1. Strike Law Enforcement in High-Accident Dates:

Increased deployment of traffic police and automated surveillance (e.g., speed cameras, red light cameras) in dates with the highest accident dates.

Increase fines and implement stricter penalties for violations such as overspeeding, reckless driving, and failure to wear seatbelts/helmets.

Conduct regular road safety audits and enforce corrective measures immediately.

2 Nightime Lighting Improvements and Monitoring:

Install solar street lighting on highways, rural roads, and accident-prone urban intersections to improve night visibility.

Use smart lighting systems that adjust brightness based on weather and traffic conditions.

Increase night-time patrolling and set up camera checkpoints to deter drunk driving.

3. Anti-Alcohol Driving Campaigns:

Launch nationwide awareness campaigns highlighting the dangers of drinking and driving, especially targeting festival seasons and weekends.

Implement stricter blood alcohol concentration (BAC) limits and conduct frequent random breathalyzer tests.

Collaborate with bars, restaurants, and event organizers to provide designated driver programs and incentives for ride-sharing.

4. Weather-Related Driving Alerts:

Develop real-time weather advisory systems integrated into GPS apps and highway signboards to warn drivers about fog, rain, or poor road conditions.

Enforce speed limit reductions during adverse weather and provide designated safe parking zones during extreme weather events.

Educate drivers on safe driving techniques in different weather conditions through training programs and public service announcements.

5.Clear Advance Warning:

Install large, highly visible warning signs several hundred meters before the construction site.

Use flashing lights, reflective signs, and electronic boards (especially at night or in low-visibility conditions).

6. Training Programs for Younger and Male Aged Drivers:

Eliminate mandatory defensive driving courses for young drivers (under 25) and risk-aged drivers (30-45), who statistically show higher accident involvement.

Require refresher courses for commercial drivers and fleet operators.

Incentivize participation through discounts on insurance premiums or tax benefits for individuals completing certified safety training.

Conclusion:

Through data-driven insights, the analysis identifies major factors contributing to road accidents in Jordan. The project recommends a combination of infrastructure improvements, stricter law enforcement, driver education, and real-time hazard management to significantly reduce accident rates and save lives.