رواد مصر الرقمية

وزارة الاتصالات
وتكنولوجيا المعلومات

Road Hazard Analytics

Prepared by :

Abdelrahman Mahmoud Ibrahim

Mohamed Ahmed Saad

Supervised by :

Eng. Marwan Mokhtar

Course :

Microsoft Power BI

Group code / Tech company :

CAI1 _ DAT2 _ S1e / AST

Content :

Introduction

This project is based on a countrywide classification of traffic incidents, covering 49 states in the USA. The dataset, collected between January 2018 and December 2019, records traffic incidents using multiple APIs that provide real-time data. These APIs gather and broadcast incident information from various sources, including federal and state departments of transportation, law enforcement agencies, traffic cameras, and road sensors. The dataset contains approximately 1,162,000 incidents.

Importance of the project

Traffic incident analysis is vital for understanding the factors and patterns that contribute to road accidents. By examining various elements such as time, location, weather conditions, and road features, this analysis helps identify high-risk areas and dangerous behaviors. This knowledge is essential for developing targeted safety measures, improving road designs, and creating data-driven policies to enhance overall road safety.

Furthermore, analyzing traffic incidents aids in predicting future accidents and evaluating the effectiveness of current safety interventions. Through detailed analysis, stakeholders can allocate resources effectively and take proactive steps to reduce accidents, injuries, and fatalities. Ultimately, this leads to safer communities and a better quality of life for all road users.

Dataset

✓ Background about the dataset:

	Unnamed: 0	ID	Source	Severity	Start_Time	End_Time	Start_Lat	Start_Lng	End_Lat	End_Lng	...	Station	Stop	Traffic_Calming	Traffic_
0	0	A-1	Source2	3	2016-02-08 05:46:00	2016-02-08 11:00:00	39.865147	-84.058723	NaN	NaN	...	False	False	False	
1	1	A-2	Source2	2	2016-02-08 06:07:59	2016-02-08 06:37:59	39.928059	-82.831184	NaN	NaN	...	False	False	False	
2	2	A-3	Source2	2	2016-02-08 06:49:27	2016-02-08 07:19:27	39.063148	-84.032608	NaN	NaN	...	False	False	False	
3	3	A-4	Source2	3	2016-02-08 07:23:34	2016-02-08 07:53:34	39.747753	-84.205582	NaN	NaN	...	False	False	False	
4	4	A-5	Source2	2	2016-02-08 07:39:07	2016-02-08 08:09:07	39.627781	-84.188354	NaN	NaN	...	False	False	False	

✓ Meta data:

- ID: A unique identification number for each incident.
- Source: Some source or the source by which the incident was reported.
- Severity: The degree of severity of the Road.

(Severity = [1,2,3,4])

1. The road you are about to cross is SAFE.
2. The road is safe but please mind your car speed.
3. The road has many risks please try to take another Road.
4. It's impossible to take that road you must look for another Road.

- Start_Time and End_Time: The time the incident started and the time it ended.
- Start_Lat and Start_Lng: Latitude and longitude of the accident start location.
- End_Lat and End_Lng: The latitude and longitude of the accident end location.
- Distance(mi): The distance affected by the accident in miles.
- Description: A brief description of the incident.
- Street: The name of the street or road.
- City, County, State : Location, city & state.
- Timezone: The time zone.
- Airport_Code : Code of the nearby airport.
- Temperature(F), Wind_Chill(F), Humidity(%), Pressure(in), Visibility(mi), etc.: Weather data related to the incident, such as temperature, humidity, barometric pressure, and visibility.
- Amenity, Bump, Crossing, Give_Way, etc. Various factors related to roads and infrastructure.
- Sunrise_Sunset, Civil_Twilight, Nautical_Twilight, Astronomical_Twilight: Sunrise, dusk, sand and astronomical twilight.

✓ Cleaning and features engineering:

➢ **Removing nulls:**

Data usually contain a nulls value that will affect on the accuracy of the analysis.

Unnamed: 0	0.00
ID	0.00
Source	0.00
Severity	0.00
Start_Time	0.00
End_Time	0.00
Start_Lat	0.00
Start_Lng	0.00
End_Lat	44.03
End_Lng	44.03
Distance(mi)	0.00
Description	0.00
Street	0.14
City	0.00
County	0.00
State	0.00
Zipcode	0.02
Country	0.00
Timezone	0.10
Airport_Code	0.29
Weather_Timestamp	1.56
Temperature(F)	2.12
Wind_Chill(F)	25.87
Humidity(%)	2.25
Pressure(in)	1.82
Visibility(mi)	2.29

To detect the average of null values in all features in the data set, the following code is used:

```
round(df.isna().sum()/ len(df) * 100, 2)
```

Findings:

1. The most null values in columns(End_Lat , End_Lng) with 44.03%.
2. Then precipitation with 28.51%.
3. Then wind_chill (F) with 25.87%.
4. Therefore, ……

This code `round(df.isna().sum()/ len(df) * 100, 2)` represents the percentage of null values in all features in the data set.

We addressed null values in the data by removing those with a null percentage exceeding 40%.

```
df.drop(columns=['End_Lat', 'End_Lng'], axis=1, inplace=True)
```

For instance, when there were null values in street names, we extracted city names and identified the most frequently occurring street within each city. This method helped in dealing with missing data effectively by filling in the street names using the most common street associated with each city.

For example:

We calculated the most frequent street in this city (“Salt Lake City”), then we put the name of this street in the missing values belonging to this city (“Salt Lake City”) in the data.

```
x=df[df["City"]=="Salt Lake City"]
y=x["Street"].mode()[0]
L=df[(df['Street'].isnull() ) & (df['City'] =='Salt Lake City')].index
df['Street'].iloc[L] = y
```

We repeated this process with most of the cities to address the missing values in them.

```
def fahrenheit_to_celsius(fahrenheit):
    celsius = (fahrenheit - 32) * 5.0/9.0
    return round(celsius)
df['Temperature(C)'] = df['Temperature(F)'].apply(fahrenheit_to_celsius)
```

After fixing the null values by removing some and substituting others by the mod:

```
round(df.isna().sum()/ len(df) * 100, 2)
```

Unnamed: 0	0.0
ID	0.0
Source	0.0
Severity	0.0
Start_Time	0.0
End_Time	0.0
Distance(mi)	0.0
Description	0.0
Street	0.0
City	0.0
County	0.0
State	0.0
Country	0.0
Timezone	0.0
Humidity(%)	0.0
Pressure(in)	0.0
Visibility(mi)	0.0
Wind_Direction	0.0
Wind_Speed(mph)	0.0
Weather_Condition	0.0
Amenity	0.0
Bump	0.0

Then we showing that the percentage of the null values become 0.

- **Feature Engineering:**

```
df['Start_Time'] = pd.to_datetime(df['Start_Time'], format='%Y-%m-%d %H:%M:%S',errors='coerce')
```

This line converts the Start_Time column in the dataframe df to a datetime format using the specified format %Y-%m-%d %H:%M:%S. The errors='coerce' argument handles errors during conversion by setting them to NaT (not a time).

```
df['Start_Date'] = df['Start_Time'].dt.date
```

This line creates a new column Start_Date in the dataframe df, which contains only the date portion (without the time) extracted from the Start_Time column.

```
df['Start_Time'] = df['Start_Time'].dt.time
```

This line modifies the Start_Time column in the dataframe df to contain only the time portion (without the date) extracted from the original Start_Time column.

Then: We repeated the same previous codes but in End_time column.

```
df["Start_Date"] = pd.to_datetime(df["Start_Date"])
```

The code converts the "Start_Date" column to a datetime format using pd.to_datetime, allowing for easier manipulation and analysis of the date data within the column.
The code changes the "Start_Date" column into a date format so you can work with it as a date.

```
df['Day_Name']=df['Start_Date'].dt.day_name()
df['Day']=df['Start_Date'].dt.day
df['Month']=df['Start_Date'].dt.month
df['Year']=df['Start_Date'].dt.year
```

```
df["Hour"]=df["Start_Time"].dt.hour
df['Minutes'] = df['Start_Time'].dt.minute
```

```
def map_months(x):
    if x in [12, 1, 2]:
        return 'Winter'
    elif x in [3, 4, 5]:
        return 'Spring'
    elif x in [6, 7, 8]:
        return 'Summer'
    elif x in [9, 10, 11]:
        return 'Autumn'

df['Season'] = df['Month'].apply(map_months)
```

The function map_months maps each month (1-12) to a season. It returns 'Winter' for months 12, 1, and 2; 'Spring' for months 3, 4, and 5; 'Summer' for months 6, 7, and 8; and 'Autumn' for months 9, 10, and 11. The code then uses df['Month']. apply(map_months) to apply this function to the 'Month' column and creates a new column 'Season' in the dataframe with the corresponding season for each month.

```
df["Month"]=df["Month"].astype(int)
df["Year"]=df["Year"].astype(int)
df["Hour"]=df["Hour"].astype(int)
df["Minutes"]=df["Minutes"].astype(int)
```

➢ The Dataset After cleaning:

	Source	Severity	Start_Lat	Start_Lng	Description	Street	City	County	State	Country	Timezone	Humidity(%)	Pressure(in)	Visil
2	Source2	2	39.063148	-84.032608	Accident on OH-32 State Route 32 Westbound at ...	State Route 32	Williamsburg	Clermont	OH	US	US/Eastern	100.0	29.67	
13	Source2	2	39.790760	-84.241547	Accident on Salem Ave at Hillcrest Ave / Kensi...	Salem Ave	Dayton	Montgomery	OH	US	US/Eastern	89.0	29.65	
17	Source2	2	39.752174	-84.239952	Accident on Delphos Ave at Brooklyn Ave. Expec...	Delphos Ave	Dayton	Montgomery	OH	US	US/Eastern	89.0	29.65	
18	Source2	2	39.740669	-84.184135	Accident on Stewart St near Rubicon St. Expect...	Rubicon St	Dayton	Montgomery	OH	US	US/Eastern	93.0	29.63	
19	Source2	2	39.790703	-84.244461	Accident on Hillcrest Ave at Piccadilly Ave. E...	W Hillcrest Ave	Dayton	Montgomery	OH	US	US/Eastern	89.0	29.65	

Business questions and Analytical SQL queries

1. Which times of the year see the most road incidents ?

Query: Count incidents per month (Seasonal Variation Analysis)

```
SELECT
    month,
    COUNT(*) AS incident_count
FROM
    road
GROUP BY
    month
ORDER BY
    incident_count DESC;
```

This query counts the number of incidents per month to identify which months experience the most accidents.

This query shows that:

	month smallint	incident_count bigint
1	10	130673
2	9	109389
3	5	103214
4	8	100968
5	4	97328
6	11	96912
7	7	94414
8	12	94357
9	3	91470
10	6	90217
11	1	78452
12	2	75062

2. What are the peak hours for road incidents, and how can traffic management be improved during these times?

Query: Count incidents by hour of the day (Rush Hour Analysis)

```
SELECT
    hour,
    COUNT(*) AS incident_count
FROM
    road
GROUP BY
    hour
ORDER BY
    incident_count DESC;
```

This query identifies the peak hours for incidents, providing insights into when road traffic management might need improvement.

This query shows that:

	hour smallint	incident_count bigint
1	8	107211
2	17	99199
3	7	97252
4	16	96259
5	15	78958
6	18	70809
7	9	64920
8	11	62334
9	14	62236
10	10	61019

3. Which geographic locations (cities, intersections, or roads) are most prone to accidents?

Query: Count incidents per city (Geographic Hotspot Analysis)

```
SELECT
    city,
    COUNT(*) AS incident_count
FROM
    road
GROUP BY
    city
ORDER BY
    incident_count DESC
LIMIT 15;
```

This query lists the top 15 cities with the most accidents, allowing for targeted road safety interventions in high-risk areas.

This query shows that:

	city text	incident_count bigint
1	Charlotte	30928
2	Houston	29630
3	Austin	23035
4	Raleigh	22713
5	Dallas	21207
6	Los Angeles	20129
7	Baton Rouge	16658
8	Nashville	16229
9	Miami	14552
10	Atlanta	14352
11	Oklahoma City	12773
12	Orlando	10804
13	San Antonio	10518
14	Portland	10405
15	Minneapolis	10158

4. How do weather conditions impact road safety?

Query: Count incidents by weather condition (Weather Impact Analysis)

```
SELECT
    weather_condition,
    COUNT(*) AS incident_count
FROM
    road
GROUP BY
    weather_condition
ORDER BY
    incident_count DESC;
```

This query examines the number of incidents based on different weather conditions to assess how weather impacts road safety.

This query shows that:

	weather_condition character varying (50)	incident_count bigint
1	Fair	264405
2	Clear	253983
3	Mostly Cloudy	194715
4	Partly Cloudy	139288
5	Overcast	121495
6	Cloudy	78762
7	Scattered Clouds	71580
8	Light Rain	26701
9	Thunderstorm	1759
10	Thunder in the Vicinity	1622

5. What types of road features (e.g., traffic signals, junctions) are linked to higher accident rates, and how can these be redesigned to improve safety?
Query: Incident count by road features like junctions, traffic signals (Road Feature Analysis)

```
SELECT
    junction,
    COUNT(*) AS incident_count
FROM
    road
GROUP BY
    junction
ORDER BY
    incident_count DESC;
```

This query counts incidents at road junctions to highlight intersections and areas where accidents are frequent, potentially needing redesigns or safety interventions.

This query shows that:

	junction character varying (50)	incident_count bigint
1	false	1074087
2	true	88369

6. How does the presence of traffic signals impact the number of traffic incidents?

Query: Count incidents based on the presence of traffic signals (Traffic Signal Impact Analysis)

```
SELECT
    traffic_signal,
    COUNT(*) AS incident_count
FROM
    road
GROUP BY
    traffic_signal
ORDER BY
    incident_count DESC;
```

This query counts the number of incidents based on whether a traffic signal was present at the location of the incident. It groups the data by the traffic_signal column, which categorizes the incidents by the presence (or absence) of traffic signals. The results, ordered by the count of incidents in descending order, help to determine if areas with or without traffic signals experience more accidents. This can inform decision-making about where to implement or enhance traffic signal systems to improve road safety.

This query shows that:

	traffic_signal character varying (50)	incident_count bigint
1	false	925673
2	true	236783

7. What roles do human factors (e.g., fatigue, distraction) play in causing accidents?

Query: Incident count by hour for late-night driving (Driver Fatigue Analysis)

```
SELECT
    hour,
    COUNT(*) AS incident_count
FROM
    road
WHERE
    hour >= 22 OR hour <= 5
GROUP BY
    hour
ORDER BY
    incident_count DESC;
```

This query looks at incidents that occur late at night or early in the morning to assess the role of driver fatigue in accidents.

This query shows that:

	hour smallint	incident_count bigint
1	5	28240
2	4	19604
3	22	17375
4	23	9489
5	0	7430
6	2	5436
7	1	5159
8	3	5043

8. How do traffic-calming measures (e.g., speed bumps, reduced speed zones) affect accident rates?

Query: Incident count by traffic-calming features (Traffic Calming Measures Analysis)

```
SELECT
    traffic_calming,
    COUNT(*) AS incident_count
FROM
    road
WHERE
    traffic_calming = 'True'
GROUP BY
    Traffic_calming;
```

This query counts incidents on roads with traffic-calming measures to evaluate their effectiveness in reducing accidents.

This query shows that:

	traffic_calming character varying (50)	incident_count bigint
1	True	1203

9. How do holidays or special events affect road safety?

Query: Incident count around holiday periods (Holiday Impact Analysis)

```
SELECT
    day_name,
    COUNT(*) AS incident_count
FROM
    road
WHERE
    month = 12
GROUP BY
    day_name
ORDER BY
    incident_count DESC;
```

This query looks at the number of incidents during holiday months like December to determine if special measures should be taken during these periods.

This query shows that:

	day_name character varying (50)	incident_count bigint
1	Monday	17758
2	Tuesday	16934
3	Thursday	15822
4	Friday	15763
5	Wednesday	15141
6	Saturday	6712
7	Sunday	6227

10. How do weekends (Saturday and Sunday) compare in terms of the number of traffic incidents?

Query: Count incidents occurring on weekends (Weekend Traffic Incident Analysis)

```
SELECT
    day_name,
    COUNT(*) AS incident_count
FROM
    road
WHERE
    day_name IN ('Saturday', 'Sunday')
GROUP BY
    day_name
ORDER BY
    incident_count DESC;
```

This query focuses on comparing the number of traffic incidents that occurred on weekends, specifically Saturdays and Sundays. It filters the data using the day_name column, counting incidents for each day. By grouping the data by day_name and ordering the results by the count of incidents, we can identify which weekend day has more incidents. This information can help in understanding traffic patterns over the weekend and implementing strategies to reduce accidents during these days.

This query shows that:

	day_name character varying (50)	incident_count bigint
1	Saturday	70402
2	Sunday	60633

11. What is the long-term trend in road safety across the years?

Query: Incident trend by year (Year-over-Year Trend Analysis)

```
SELECT
    year,
    COUNT(*) AS incident_count,
    AVG(severity) AS avg_severity
FROM
    road
GROUP BY
    year
ORDER BY
    year;
```

This query analyzes the long-term trend of incidents by year, including average severity, to assess whether safety measures have improved road conditions over time.

This query shows that:

	year smallint	incident_count bigint	avg_severity numeric
1	2018	545560	2.3780170833638830
2	2019	616896	2.3060321350762527

12. How do weather conditions, time of day, and location (state) contribute to severe traffic incidents (Severity 4)?

Query: Combined Factors Impact Analysis

```
SELECT
    weather_condition,
    hour,
    state,
   COUNT(*) AS incident_count
FROM
    road
WHERE
    severity = 4
GROUP BY
    weather_condition, hour, state
ORDER BY
    incident_count DESC;
```

This query looks at the number of severe incidents (severity 4) based on weather conditions, time of day, and state to identify patterns that could help in implementing targeted safety measures for high-risk conditions and locations.

This query shows that:

	weather_condition character varying (50)	hour smallint	state character varying (50)	incident_count bigint
1	Fair	0	FL	88
2	Fair	0	GA	86
3	Fair	9	FL	70
4	Mostly Cloudy	10	FL	66
5	Mostly Cloudy	17	FL	65
6	Partly Cloudy	11	FL	62
7	Fair	8	FL	59
8	Mostly Cloudy	15	FL	59
9	Mostly Cloudy	18	FL	58
10	Mostly Cloudy	12	FL	54
11	Fair	7	FL	53
12	Clear	16	CA	52
13	Mostly Cloudy	0	FL	52
14	Clear	15	GA	51
15	Fair	20	FL	50
16	Fair	23	FL	50

13. Do specific holidays (Halloween on October 31 and Thanksgiving on November 25) see an increase in traffic incidents, and on which days do they occur?

Query: Count incidents on Halloween and Thanksgiving (Holiday Incident Analysis)

```
SELECT
    day_name,
    month,
    COUNT(*) AS incident_count
FROM
    road
WHERE
    (month = 10 AND day = 31) OR (month = 11 AND day = 25)
GROUP BY
    day_name, month
ORDER BY
    incident_count DESC;
```

This query looks at the number of incidents on Halloween (October 31) and Thanksgiving (November 25) to assess if there is a spike in accidents during these specific holidays and identify the day of the week on which they occurred. This information could help in planning safety measures for high-risk holiday periods.

This query shows that:

	day_name character varying (50)	month smallint	incident_count bigint
1	Thursday	10	2609
2	Wednesday	10	2184
3	Monday	11	2052
4	Sunday	11	584

14. Which season experiences the highest number of traffic incidents?

Query: Count incidents by season (Seasonal Traffic Incident Analysis)

```
SELECT
    season,
    COUNT(*) AS incident_count
FROM
    road
GROUP BY
    season
ORDER BY
    incident_count DESC;
```

This query analyzes the number of traffic incidents by season to determine which season has the highest incident count. Understanding seasonal trends can help in allocating resources and implementing safety measures tailored to the specific risks associated with each season.

This query shows that:

	season character varying (50)	incident_count bigint
1	Autumn	336974
2	Spring	292012
3	Summer	285599
4	Winter	247871

15. How do different types of intersections and crossings (crossing, junction, railway) relate to the frequency of severe traffic incidents?

Query: Count severe incidents based on intersection types (Severe Incident Analysis at Crossings and Junctions)

```
SELECT
    crossing,
    junction,
    railway,
    COUNT(*) AS incident_count
FROM
    road
WHERE
    severity IN (3, 4)
GROUP BY
    crossing, junction, railway
ORDER BY
    incident_count DESC;
```

This query examines the number of severe traffic incidents (severity levels 3 and 4) based on the types of crossings, junctions, and railways to identify which intersection types are associated with higher frequencies of severe accidents. This information can guide infrastructure improvements and safety interventions at critical points in the road network.

This query shows that:

	crossing character varying (50)	junction character varying (50)	railway character varying (50)	incident_count bigint
1	false	false	false	309671
2	false	true	false	37434
3	true	false	false	12576
4	false	false	true	1135
5	true	false	true	651
6	true	true	false	506
7	false	true	true	232
8	true	true	true	78

16. Which streets in specific cities have a high number of traffic incidents?

Query: Count incidents per street and city (High Incident Analysis)

```
SELECT
    street,
    city,
    COUNT(*) AS incident_count
FROM
    road
GROUP BY
    street, city
HAVING
    COUNT(*) > 100
ORDER BY
    incident_count DESC;
```

This query identifies streets in various cities that have recorded more than 100 traffic incidents, allowing us to pinpoint high-risk areas. By focusing on streets with a significant number of incidents, this analysis can inform targeted interventions and safety measures to improve road safety in these locations.

This query shows that:

	street text	city text	incident_count bigint
1	I-95 S	Miami	2067
2	I-95 N	Miami	1633
3	I-45 N	Houston	1316
4	I-5 N	Seattle	1215
5	I-5 N	Portland	1158
6	I-405 N	Los Angeles	1141
7	I-10 E	Los Angeles	1111
8	Capital Blvd	Raleigh	1082
9	I-24 W	Nashville	1065
10	I-4 E	Orlando	1061

17. In which cities do traffic incidents frequently occur on hills?

Query: Count incidents related to hills by city (Hill-Related Incident Analysis)

```
SELECT
    city,
    description,
    COUNT(*) AS incident_count
FROM
    road
WHERE
    description LIKE '%hill%'
GROUP BY
    city, description
ORDER BY
    incident_count DESC;
```

This query analyzes traffic incidents that have a description related to "hills" to determine in which cities these incidents are most common. By grouping the data by city and description, the query counts incidents and orders the results to identify areas where hilly terrain may contribute to higher accident rates. This insight can guide road safety initiatives in cities with challenging topography.

This query shows that:

	city text	description text	incident_count bigint
1	Pleasanton	At Foothill Rd/San Ramon Rd - Accident.	52
2	Rancho Cucamonga	At CA-66/Foothill Blvd - Accident.	42
3	Baldwin Park	At I-210/Foothill Fwy - Accident.	39
4	Arcadia	At I-210/Foothill Fwy - Accident.	33
5	Claremont	At CA-210/Foothill Fwy - Accident.	27
6	Pomona	At Phillips Ranch Rd - Accident.	26

Analysis based on Street

1. What are the top 10 streets with the highest number of traffic incidents?

Query: Count incidents per street (Top Streets Incident Analysis)

```
SELECT
    street,
    COUNT(*) AS incident_count
FROM
    road
GROUP BY
    street
ORDER BY
    incident_count DESC
LIMIT 10;
```

This query identifies the top 10 streets with the highest number of traffic incidents by counting incidents for each street and ordering the results in descending order. By focusing on the streets with the most incidents, this analysis can help prioritize road safety measures and interventions in the most dangerous areas.

This query shows that:

	street text	incident_count bigint
1	I-5 N	13268
2	I-95 N	12375
3	I-95 S	11507
4	I-5 S	9017
5	I-10 E	8044
6	I-10 W	7824
7	I-405 N	6244
8	I-80 W	5455
9	I-75 N	5323
10	I-75 S	5069

2. Which streets experience the highest number of traffic incidents at specific hours?

Query: Count incidents per street and hour (Top Streets and Hours Incident Analysis)

```
SELECT
    street,
    hour,
    COUNT(*) AS incident_count
FROM
    road
GROUP BY
    street, hour
ORDER BY
    incident_count DESC
LIMIT 10;
```

This query identifies the top 10 combinations of streets and hours with the highest number of traffic incidents. By counting incidents for each street at each hour and ordering the results, the analysis highlights when and where incidents are most likely to occur. This information can assist in developing targeted traffic management strategies and safety initiatives during high-risk hours on specific streets.

This query shows that:

	street text	hour smallint	incident_count bigint
1	I-95 N	8	1235
2	I-5 N	17	1108
3	I-95 S	8	1084
4	I-5 N	15	1072
5	I-5 N	16	1067
6	I-5 N	14	1053
7	I-95 N	7	1014
8	I-95 S	7	1002
9	I-95 N	16	958
10	I-95 N	17	953

3. Which streets have the highest number of traffic incidents under specific weather conditions?

Query: Count incidents per street and weather condition (Top Streets Weather-Related Incident Analysis)

```
SELECT
    street,
    weather_condition,
    COUNT(*) AS incident_count
FROM
    road
GROUP BY
    street, weather_condition
ORDER BY
    incident_count DESC
LIMIT 10;
```

This query identifies the top 10 combinations of streets and weather conditions associated with the highest number of traffic incidents. By counting incidents for each street based on varying weather conditions, the analysis reveals which streets are particularly prone to accidents under specific weather scenarios. This information can guide targeted safety measures and improvements in road infrastructure, especially in areas most affected by adverse weather conditions.

This query shows that:

	street text	weather_condition character varying (50)	incident_count bigint
1	I-5 N	Fair	3338
2	I-95 N	Mostly Cloudy	2792
3	I-5 N	Clear	2717
4	I-5 S	Fair	2648
5	I-95 S	Mostly Cloudy	2625
6	I-10 W	Fair	2398
7	I-10 E	Fair	2368
8	I-95 S	Fair	2285
9	I-95 N	Fair	2244
10	I-95 N	Clear	2221

4. Which streets experience the highest number of traffic incidents during specific seasons?

Query: Count incidents per street and season (Top Streets Seasonal Incident Analysis)

```
SELECT
    street,
    season,
    COUNT(*) AS incident_count
FROM
    road
GROUP BY
    street, season
ORDER BY
    incident_count DESC
LIMIT 10;
```

This query identifies the top 10 combinations of streets and seasons with the highest number of traffic incidents. By counting incidents for each street during different seasons, the analysis helps to reveal patterns of accidents related to seasonal conditions. This information can inform traffic management strategies and safety initiatives tailored to the unique risks associated with each season on specific streets.

This query shows that:

	street text	season character varying (50)	incident_count bigint
1	I-5 N	Autumn	4303
2	I-95 N	Summer	3299
3	I-5 S	Autumn	3267
4	I-95 N	Spring	3253
5	I-5 N	Summer	3230
6	I-95 N	Autumn	3114
7	I-95 S	Summer	3055
8	I-95 S	Autumn	3016
9	I-5 N	Spring	2981
10	I-5 N	Winter	2754

5. How do traffic calming measures impact the number of incidents on specific streets?

Query: Count incidents per street and traffic calming measures (Top Streets Traffic Calming Incident Analysis)

```
SELECT
    street,
    traffic_calming,
    COUNT(*) AS incident_count
FROM
    road
GROUP BY
    street, traffic_calming
ORDER BY
    incident_count DESC
LIMIT 10;
```

This query analyzes the top 10 combinations of streets and their associated traffic calming measures in relation to the number of traffic incidents. By counting incidents for each street based on whether traffic calming measures are present, the analysis helps to assess the effectiveness of such measures in reducing accidents. This information can guide further traffic safety initiatives and the implementation of additional calming strategies on high-risk streets.

This query shows that:

	street text	traffic_calming character varying (50)	incident_count bigint
1	I-5 N	False	13263
2	I-95 N	False	12374
3	I-95 S	False	11504
4	I-5 S	False	9016
5	I-10 E	False	8044
6	I-10 W	False	7824
7	I-405 N	False	6244
8	I-80 W	False	5455
9	I-75 N	False	5323
10	I-75 S	False	5069

6. Which streets at specific junctions have the highest number of traffic incidents?

Query: Count incidents per street and junction (Top Streets Junction Incident Analysis)

```
SELECT
    street,
    junction,
    COUNT(*) AS incident_count
FROM
    road
WHERE
    junction IS NOT NULL
GROUP BY
    street, junction
ORDER BY
    incident_count DESC
LIMIT 10;
```

This query identifies the top 10 combinations of streets and their respective junctions that have recorded the highest number of traffic incidents. By counting incidents where junction information is available, the analysis reveals which streets at specific junctions are most prone to accidents. This information can inform traffic management decisions and safety measures at critical junctions to enhance road safety and reduce the likelihood of incidents.

This query shows that:

	street text	junction character varying (50)	incident_count bigint
1	I-5 N	false	11289
2	I-95 N	false	10399
3	I-95 S	false	9568
4	I-5 S	false	8021
5	I-10 E	false	7161
6	I-10 W	false	6872
7	I-405 N	false	5449
8	I-75 N	false	4707
9	I-75 S	false	4351
10	I-80 W	false	4199

7. Which streets in specific cities experience the highest number of severe traffic incidents?
   Query: Count severe incidents per street and city (Severe Incident Analysis by Location)

```
SELECT
    street,
    city,
    severity,
    COUNT(*) AS incident_count
FROM
    road
WHERE
    severity = 4
GROUP BY
    street, city, severity
ORDER BY
    incident_count DESC
LIMIT 100;
```

This query focuses on identifying the top 100 streets in various cities that have recorded the highest number of severe traffic incidents (severity level 4). By counting the incidents grouped by both street and city, the analysis highlights areas with critical safety concerns, allowing for targeted interventions and resource allocation to improve road safety in those high-risk locations.

This query shows that:

	street text	city text	severity smallint	incident_count bigint
1	I-95 S	Miami	4	253
2	I-95 N	Miami	4	164
3	GA-407 W	Atlanta	4	66
4	I-95 S	Hallandale	4	64
5	I-75 S	Atlanta	4	59
6	I-45 N	Houston	4	52
7	I-95 S	Hollywood	4	44
8	GA-407 S	Atlanta	4	43
9	I-95 S	Fort Lauderdale	4	41
10	GA-402 W	Atlanta	4	41

- We found that October was the most month have incident, The incidents increase from September to October and then decreased in November.

Analysis based on increase from September to October and decrease from October to November

1. How do weather conditions affect the number of incidents during the months of September, October, and November?

   Query: Count incidents by month and weather condition (Seasonal Weather Impact on Incidents Analysis)

```
SELECT
    month,
    weather_condition,
    COUNT(*) AS incident_count
FROM
    road
WHERE
    month IN (9, 10, 11)
GROUP BY
    month, weather_condition
ORDER BY
    month, incident_count DESC;
```

This query analyzes the number of traffic incidents during the fall months of September, October, and November, grouped by different weather conditions. By identifying how weather affects incident counts during these months, the analysis can help in understanding seasonal weather patterns that may contribute to increased road incidents. This insight can inform planning for safety measures tailored to specific weather conditions in the fall season.

This query shows that (Different parts of output):

	month smallint	weather_condition character varying (50)	incident_count bigint
1	9	Fair	39721
2	9	Mostly Cloudy	17073
3	9	Clear	15803
4	9	Partly Cloudy	15131
5	9	Overcast	6632
6	9	Cloudy	6571
7	9	Scattered Clouds	5609
8	9	Light Rain	1894

33	10	Overcast	9272
34	10	Scattered Clouds	6184
35	10	Light Rain	2909
36	10	Smoke	240
37	10	Light Drizzle	150
38	10	Rain	117
39	10	Patches of Fog	56
40	10	Thunder in the Vicinity	55

77	11	Drizzle	34
78	11	Thunder in the Vicinity	22
79	11	Wintry Mix	19
80	11	N/A Precipitation	17
81	11	Thunderstorm	10
82	11	Heavy Rain	10
83	11	T-Storm	10
84	11	Light Freezing Rain	9

2. What time of day sees the highest number of incidents during the months of September, October, and November?

Query: Count incidents by month and hour (Incident Timing Analysis for Fall Months)

```
SELECT
    month,
    hour,
    COUNT(*) AS incident_count
FROM
    road
WHERE
    month IN (9, 10, 11)
GROUP BY
    month, hour
ORDER BY
    month, incident_count DESC;
```

This query analyzes traffic incidents during the months of September, October, and November, grouped by the hour of the day. By identifying the times of day with the highest incident counts during these fall months, the analysis can help in adjusting traffic management, enforcement, and public safety campaigns to focus on peak hours when accidents are more likely to occur.

This query shows that (Different parts of output):

	month smallint	hour smallint	incident_count bigint
1	9	8	10024
2	9	17	9530
3	9	16	9425
4	9	7	8725
5	9	15	7482
6	9	18	6415
7	9	14	5994
8	9	9	5896
9	9	11	5680
10	9	13	5530

25	10	7	12027
26	10	8	11558
27	10	16	11327
28	10	17	11042
29	10	15	8843
30	10	6	7851
31	10	18	7451
32	10	14	6837
33	10	9	6524
34	10	13	6151

63	11	5	2569
64	11	20	2315
65	11	4	1935
66	11	21	1717
67	11	22	1278
68	11	23	762
69	11	0	565
70	11	3	435
71	11	2	435
72	11	1	413

3. Which states and cities experienced the largest increase in incidents from September to October, and the largest decrease from October to November?

Query: Incident increase and decrease by state and city (Fall Month Incident Change Analysis)

```
SELECT
    state,
    city,
    SUM(CASE WHEN month = 9 THEN 1 ELSE 0 END) AS
september_incidents,
    SUM(CASE WHEN month = 10 THEN 1 ELSE 0 END) AS
october_incidents,
    SUM(CASE WHEN month = 11 THEN 1 ELSE 0 END) AS
november_incidents,
    (SUM(CASE WHEN month = 10 THEN 1 ELSE 0 END) - SUM(CASE
WHEN month = 9 THEN 1 ELSE 0 END)) AS sept_to_oct_increase,
    (SUM(CASE WHEN month = 10 THEN 1 ELSE 0 END) - SUM(CASE
WHEN month = 11 THEN 1 ELSE 0 END)) AS oct_to_nov_decrease
FROM
    road
GROUP BY
    state, city
ORDER BY
    sept_to_oct_increase DESC, oct_to_nov_decrease DESC;
```

This query analyzes the changes in traffic incidents between September, October, and November across different states and cities. It calculates the increase in incidents from September to October and the decrease from October to November, identifying the areas that experienced the most significant shifts. This insight can help focus attention on areas that may require more resources or safety interventions during the fall months.

This query shows that:

	state character varying (50)	city text	september_incidents bigint	october_incidents bigint	november_incidents bigint	sept_to_oct_increase bigint	oct_to_nov_decrease bigint
1	NC	Charlotte	1689	2904	2519	1215	385
2	NC	Raleigh	1507	2547	1961	1040	586
3	CA	Los Angeles	2767	3470	1319	703	2151
4	MN	Minneapolis	798	1379	1100	581	279
5	VA	Richmond	643	1143	796	500	347
6	FL	Miami	1053	1517	1148	464	369
7	FL	Orlando	689	1081	970	392	111
8	MN	Saint Paul	620	981	725	361	256
9	TX	Houston	2537	2896	2311	359	585
10	TX	Austin	2138	2473	1926	335	547
11	NY	Rochester	593	900	434	307	466
12	GA	Atlanta	1045	1316	1071	271	245
13	IN	Indianapolis	971	1236	350	265	886

4. How do junctions and roundabouts contribute to traffic incidents during the fall months of September, October, and November?

Query: Count incidents by month, junction, and roundabout (Junction and Roundabout Incident Analysis for Fall Months)

```
SELECT
    month,
    junction,
    roundabout,
    COUNT(*) AS incident_count
FROM
    road
WHERE
    month IN (9, 10, 11)
GROUP BY
    month, junction, roundabout
ORDER BY
    month, incident_count DESC;
```

This query examines the number of traffic incidents during September, October, and November, focusing on how junctions and roundabouts contribute to accidents. By grouping the data by month, junction, and roundabout, this analysis reveals which combinations are most prone to incidents during the fall season. The results can guide infrastructure improvements and safety initiatives at specific traffic points to reduce accidents.

This query shows that:

	month smallint	junction character varying (50)	roundabout character varying (50)	incident_count bigint
1	9	false	False	100287
2	9	true	False	9100
3	9	false	True	2
4	10	false	False	120006
5	10	true	False	10665
6	10	false	True	1
7	10	true	True	1
8	11	false	False	90411
9	11	true	False	6496
10	11	false	True	3
11	11	true	True	2

5. How does the average severity of incidents change across the months of September, October, and November?

Query: Average incident severity per month (Severity Analysis for Fall Months)

```
SELECT
    month,
    AVG(severity) AS avg_severity
FROM
    road
WHERE
    month IN (9, 10, 11)
GROUP BY
    month
ORDER BY
    month;
```

This query calculates the average severity of traffic incidents during the months of September, October, and November. By grouping the data by month and calculating the average severity, the analysis helps to determine if certain months experience more severe incidents. The results can assist in understanding seasonal factors that may affect the seriousness of accidents, enabling better safety measures and planning during higher-risk periods.

This query shows that:

	month smallint	avg_severity numeric
1	9	2.3021693223267422
2	10	2.2789864776962341
3	11	2.3068350668647845

6. How do weekend incidents (Saturdays and Sundays) compare across the months of September, October, and November?

Query: Count incidents by month and weekend days (Weekend Incident Analysis for Fall Months)

```
SELECT
    month,
    day_name,
    COUNT(*) AS incident_count
FROM
    road
WHERE
    month IN (9, 10, 11)
    AND day_name IN ('Saturday', 'Sunday')
GROUP BY
    month, day_name
ORDER BY
    month, incident_count DESC;
```

This query examines the number of traffic incidents that occurred on weekends (Saturdays and Sundays) during the fall months of September, October, and November. By grouping the data by month and day of the week, the analysis helps identify whether certain weekends in the fall season have higher incident rates. This insight can guide safety measures and road interventions on weekends, particularly in months where accidents are more frequent.

This query shows that:

	month smallint	day_name character varying (50)	incident_count bigint
1	9	Saturday	8124
2	9	Sunday	7462
3	10	Saturday	7913
4	10	Sunday	6624
5	11	Saturday	6827
6	11	Sunday	5186

7. How do traffic incidents vary by day of the week during the months of September, October, and November?

Query: Count incidents by month and day of the week (Day-of-Week Incident Analysis for Fall Months)

```
SELECT
    month,
    day_name,
    COUNT(*) AS incident_count
FROM
    road
WHERE
    month IN (9, 10, 11)
GROUP BY
    month, day_name
ORDER BY
    month, incident_count DESC;
```

This query looks at the number of traffic incidents for each day of the week during the fall months of September, October, and November. By grouping incidents by both month and day of the week, this analysis identifies patterns in the occurrence of accidents on specific days. It helps in understanding which days have higher incident rates in each month, providing insights for targeted traffic management and safety measures during the fall season.

This query shows that:

	month smallint	day_name character varying (50)	incident_count bigint
1	9	Friday	19695
2	9	Tuesday	19302
3	9	Monday	18847
4	9	Wednesday	18594
5	9	Thursday	17365
6	9	Saturday	8124
7	9	Sunday	7462
8	10	Tuesday	25358
9	10	Wednesday	24650
10	10	Thursday	23155
11	10	Friday	21520
12	10	Monday	21453
13	10	Saturday	7913
14	10	Sunday	6624
15	11	Friday	21954
16	11	Wednesday	17329
17	11	Tuesday	16523
18	11	Monday	15155
19	11	Thursday	13938
20	11	Saturday	6827
21	11	Sunday	5186

8. Do more traffic incidents occur during the day or night in the fall months of September, October, and November?

Query: Count incidents by month and time period (Day vs. Night Incident Analysis for Fall Months)

```
SELECT
    month,
    CASE
        WHEN hour BETWEEN 6 AND 18 THEN 'Day'
        ELSE 'Night'
    END AS time_period,
    COUNT(*) AS incident_count
FROM
    road
WHERE
    month IN (9, 10, 11)
GROUP BY
    month, time_period
ORDER BY
    month, incident_count DESC;
```

This query analyzes whether more traffic incidents occur during the day (between 6 AM and 6 PM) or night (outside of those hours) for the months of September, October, and November. By grouping the incidents by month and classifying them into "Day" or "Night" periods, this analysis helps determine whether incidents are more frequent during the day or night in the fall months. This information can inform time-specific traffic safety measures and resource allocation.

This query shows that:

	month smallint	time_period text	incident_count bigint
1	9	Day	90604
2	9	Night	18785
3	10	Day	107318
4	10	Night	23355
5	11	Day	80347
6	11	Night	16565

Dashboard (Power BI)

- DAX measures:

1. Total incidents :

`Total Incidents = COUNTROWS(road)`
This measure calculates the total number of incidents

2. Severe incidents :

`Severe Incidents = CALCULATE([Total Incidents], road[Severity] = 4)`
This measure calculates the total number of incidents where the severity level is equal to 4

3. Incident by month :

`Incidents by Month = COUNTAX(road, road[Month])`
This measure calculates the total number of incidents of each month

- DAX for calculated col.s:

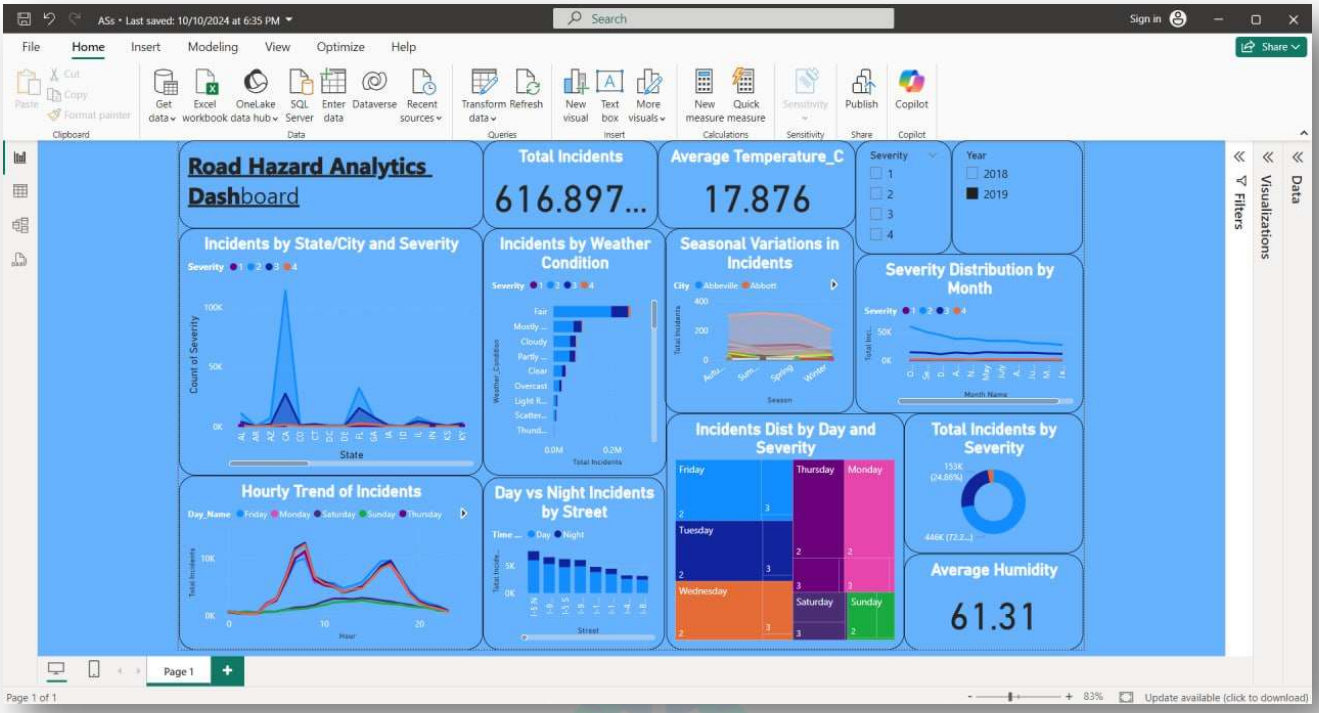
1. Calculated col. (Time period) to classify the hours as “Day” or “Night” :

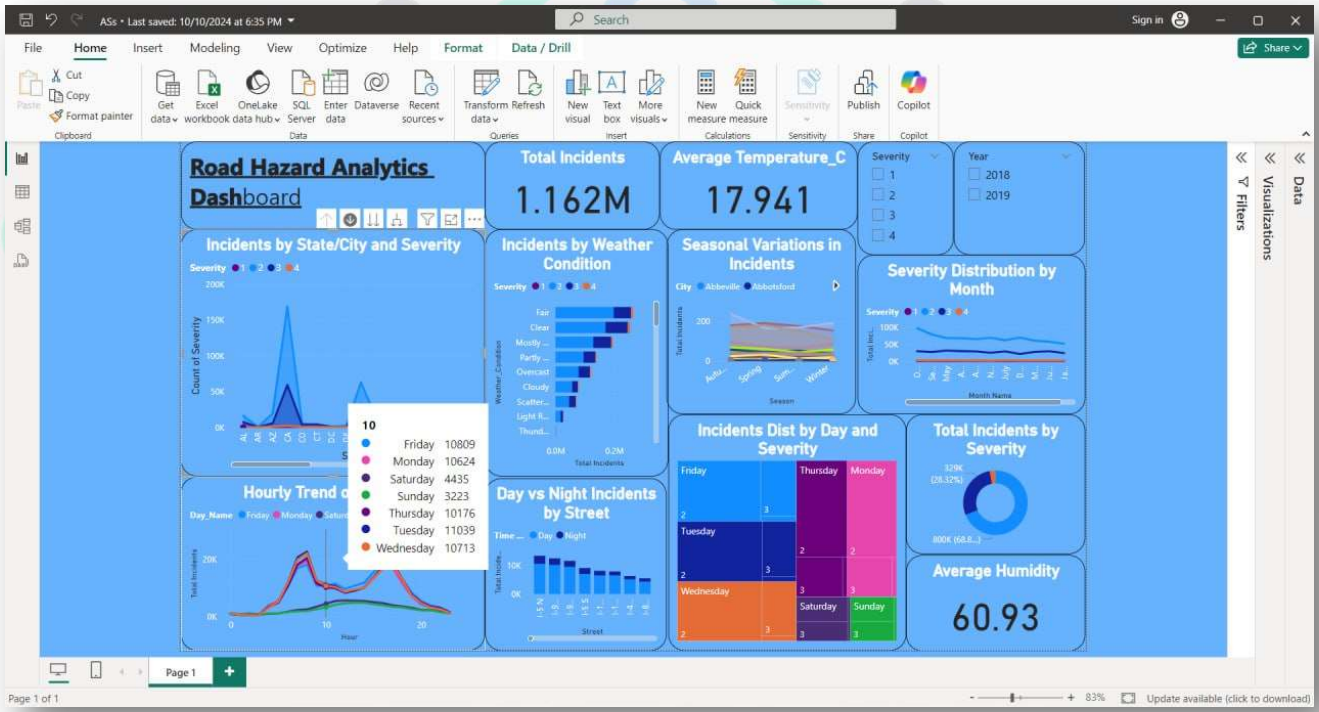
```
Time Period = IF(road[Hour] >= 6 &&
road[Hour] <= 18, "Day", "Night")
```

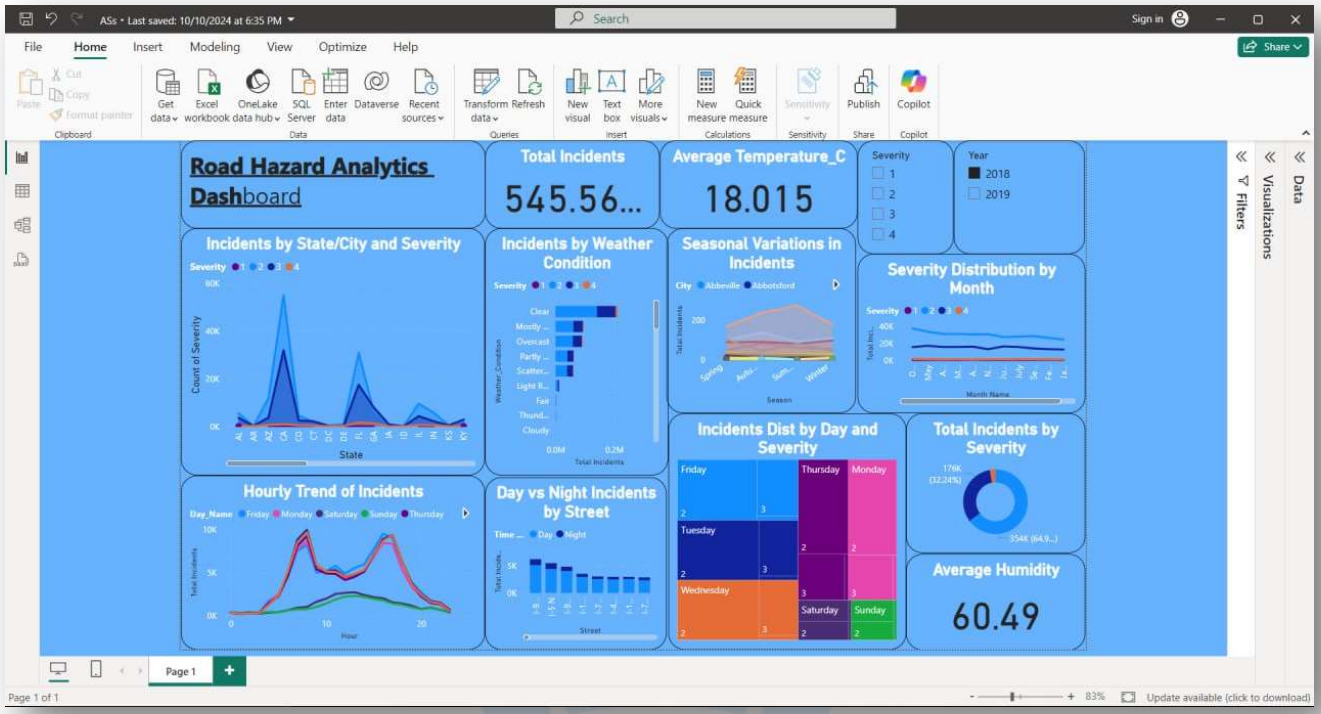
2. Calculated col. (Month name) to convert the numeric month value into the corresponding month name :

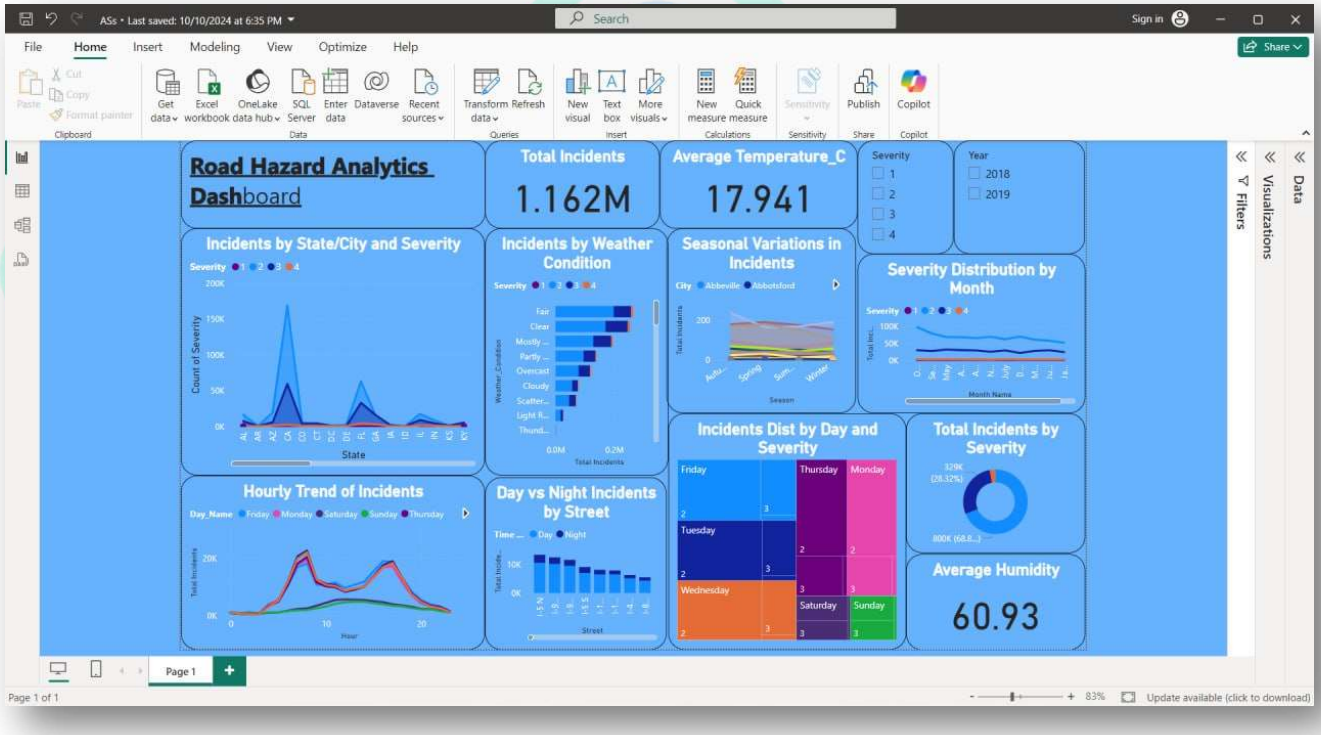
```
Month Name =
SWITCH(
    road[Month],
    1, "January",
    2, "February",
    3, "March",
    4, "April",
    5, "May",
    6, "June",
    7, "July",
    8, "August",
    9, "September",
    10, "October",
    11, "November",
    12, "December",
    "Invalid Month"  -- Optional: handle
invalid values
)
```

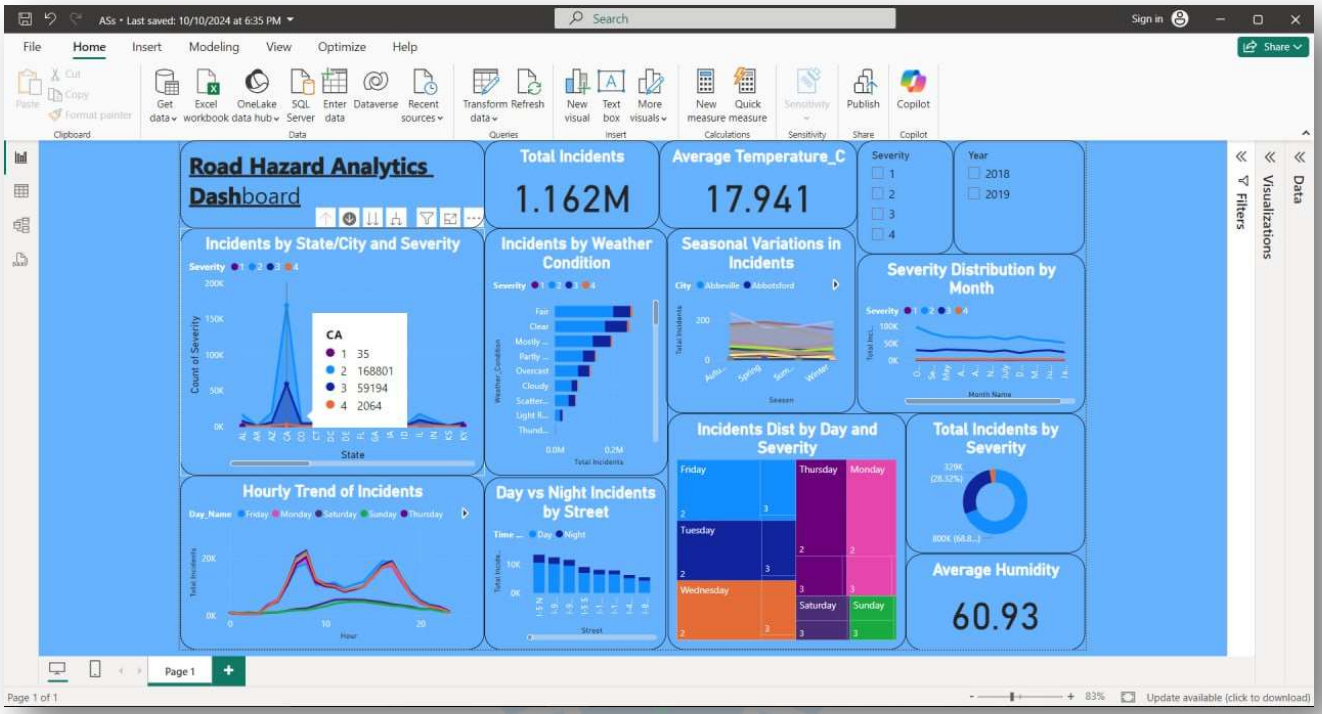
- Dashboard (Some screenshots for the visuals):

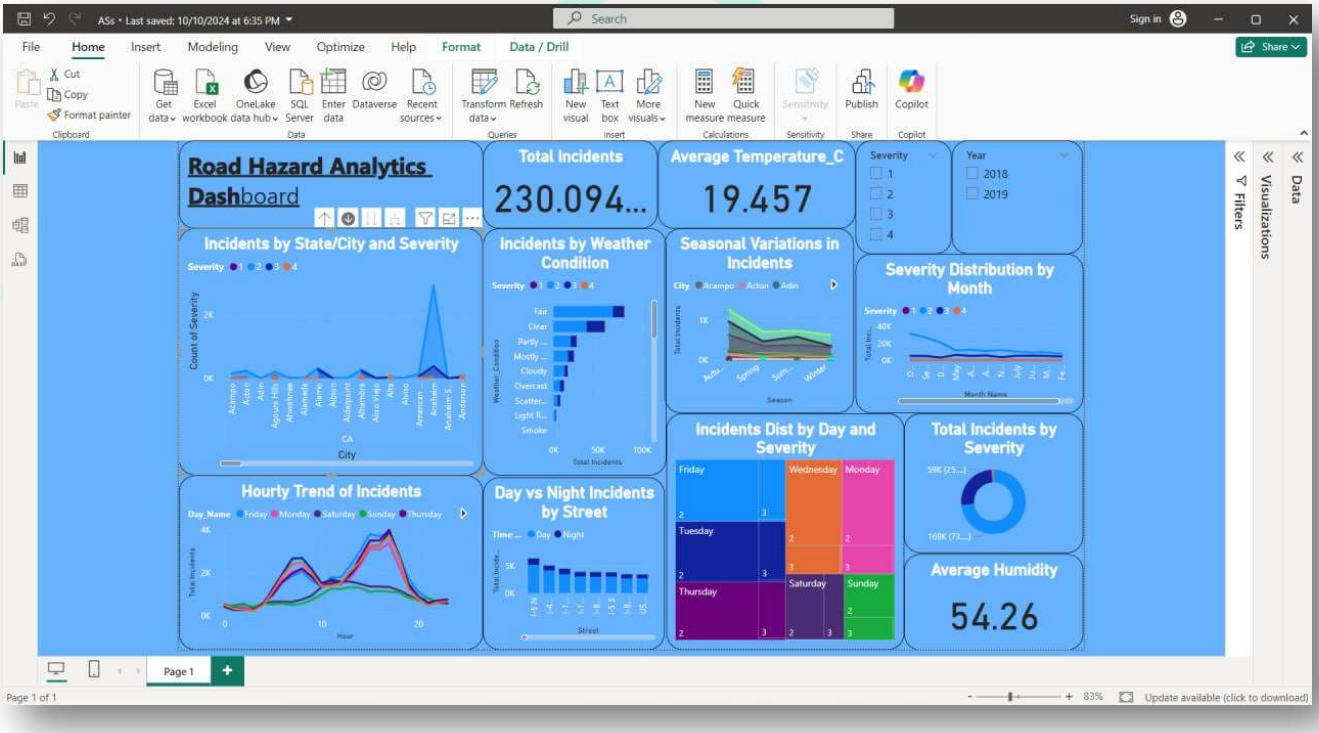





Road Hazard Analytics Dashboard
Total Incidents
545.56...
Average Temperature_C
18.015
Incidents by State/City and Severity
Incidents by Weather Condition
Seasonal Variations in Incidents
Severity Distribution by Month
Hourly Trend of Incidents
Day vs Night Incidents by Street
Incidents Dist by Day and Severity
Total Incidents by Severity
Average Humidity
60.49


Road Hazard Analytics Dashboard
Total Incidents
1.162M
Average Temperature_C
17.941
Incidents by State/City and Severity
Incidents by Weather Condition
Seasonal Variations in Incidents
Severity Distribution by Month
Hourly Trend of Incidents
Day vs Night Incidents by Street
Incidents Dist by Day and Severity
Total Incidents by Severity
Average Humidity
60.93


Road Hazard Analytics Dashboard
Total Incidents
1.162M
Average Temperature_C
17.941
Incidents by State/City and Severity
Incidents by Weather Condition
Seasonal Variations in Incidents
Severity Distribution by Month
Incidents Dist by Day and Severity
Total Incidents by Severity
Hourly Trend of Incidents
Day vs Night Incidents by Street
Average Humidity
60.93


Road Hazard Analytics Dashboard
Total Incidents
230.094...
Average Temperature_C
19.457
Incidents by State/City and Severity
Incidents by Weather Condition
Seasonal Variations in Incidents
Severity Distribution by Month
Incidents Dist by Day and Severity
Total Incidents by Severity
Hourly Trend of Incidents
Day vs Night Incidents by Street
Average Humidity
54.26

Recommendations

- ✓ Implement seasonal safety campaigns during the fall, especially in October, when incidents are highest. This could include increased road safety advertising, public awareness campaigns about driving conditions during autumn, and proactive road maintenance.

- ✓ Issue weather-specific driving advisories to encourage safer driving behaviors during adverse conditions. Authorities could also consider deploying more traffic control personnel and resources during bad weather and ensuring roads are cleared and treated (e.g., salting icy roads).

- ✓ Optimize traffic signal timings, install roundabouts where feasible, and implement additional safety features like better signage, rumble strips, or traffic calming measures around junctions.

- ✓ Enhance law enforcement and road safety monitoring on weekends. Consider more stringent DUI checks and speed monitoring, as well as targeted road safety campaigns encouraging cautious driving during weekends.

- ✓ Focus safety improvements on identified high-risk streets. Install advanced safety measures such as better street lighting, increased police presence, or even physical road modifications (e.g., speed bumps) in these areas.

- ✓ Improve street lighting, enhance visibility of road markings, and consider night driving safety programs, such as offering free vision checks or distributing anti-glare glasses to drivers.

- ✓ Develop a real-time incident detection and response system that uses live data to alert authorities about potential hazards. This could help in deploying emergency services faster, setting up roadblocks, or sending alerts to drivers.

- ✓ Allocate more resources, such as traffic patrols and emergency services, during these peak months. Conduct infrastructure inspections beforehand to ensure roads are well-maintained.

Conclusion

The analysis of road hazard data revealed important insights regarding the patterns and factors associated with traffic incidents. A notable finding was the increase in incidents from September to October, followed by a decrease in November. This seasonal pattern suggests a need for targeted safety measures during high-risk periods, especially in October.

Weather conditions also played a significant role in incident rates, with poor visibility, rain, and fog contributing to higher numbers. Roads with traffic signals, junctions, and specific street locations were frequently linked to more severe accidents, indicating potential areas for safety improvements.

Incidents tended to be more frequent during weekends, particularly on Saturdays and Sundays, which may be linked to increased travel and social activities. Additionally, nighttime driving posed a higher risk, possibly due to reduced visibility and driver fatigue.

By addressing these factors—such as focusing on high-risk months, improving road safety measures during adverse weather, optimizing traffic management at intersections, and enhancing nighttime visibility—it is possible to make roads safer and reduce the number of traffic incidents. The analysis also highlights the need for continuous monitoring and proactive interventions to adapt to changing conditions and patterns.