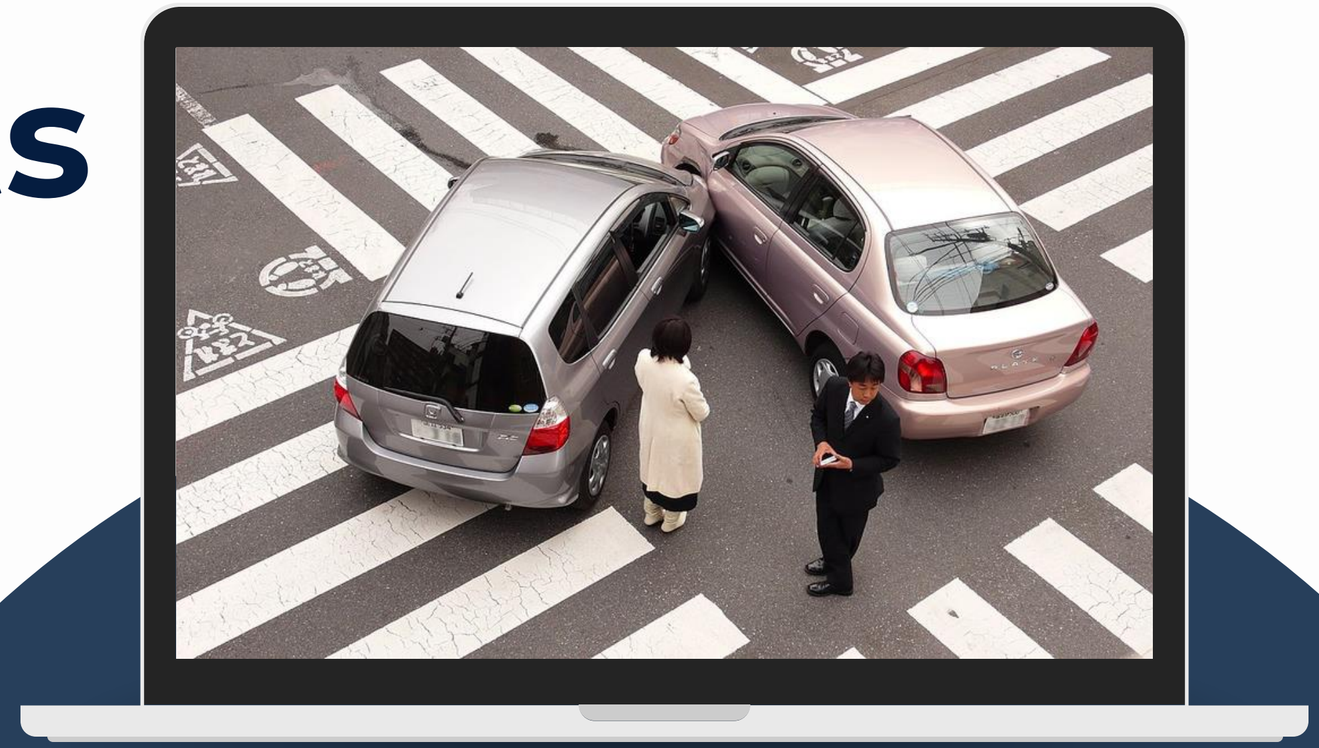
رواد مصر الرقمية

Car Accidents Analysis



وزارة الاتصـــــالات
وتكنولوجيا المعلومات

Our Team

Arwa Nabil

Mohamed Ibrahim

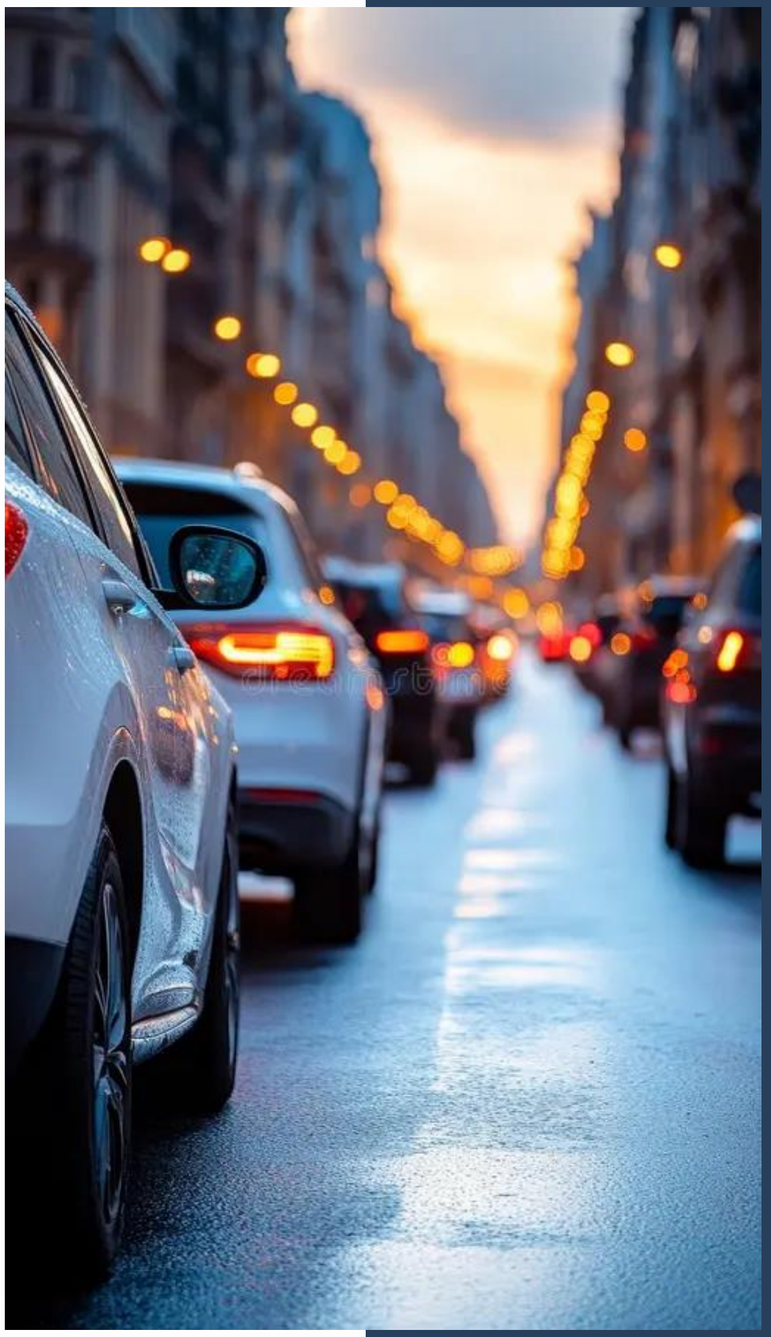
Kareem Farahat

Mahmoud Ahmed

Omar Mohammed

Muhammed Megahed

Agenda



01 Problem Statement

According to the Central Agency for Public Mobilization and Statistics (CAPMAS) in Egypt, road accident fatalities reached 5,260, while road traffic injuries increased from 71,016 in 2023 to 76,362 in 2024 (an increase of 7.5%).

In this project, we analyze car accident data from the United States as a model to uncover the main causes of accidents and explore ways to reduce them.

02 Project Objectives

The aim of this dashboard is to identify the main reasons behind car accidents, and to analyze how, when, and under what conditions they most frequently occur.

Question 01

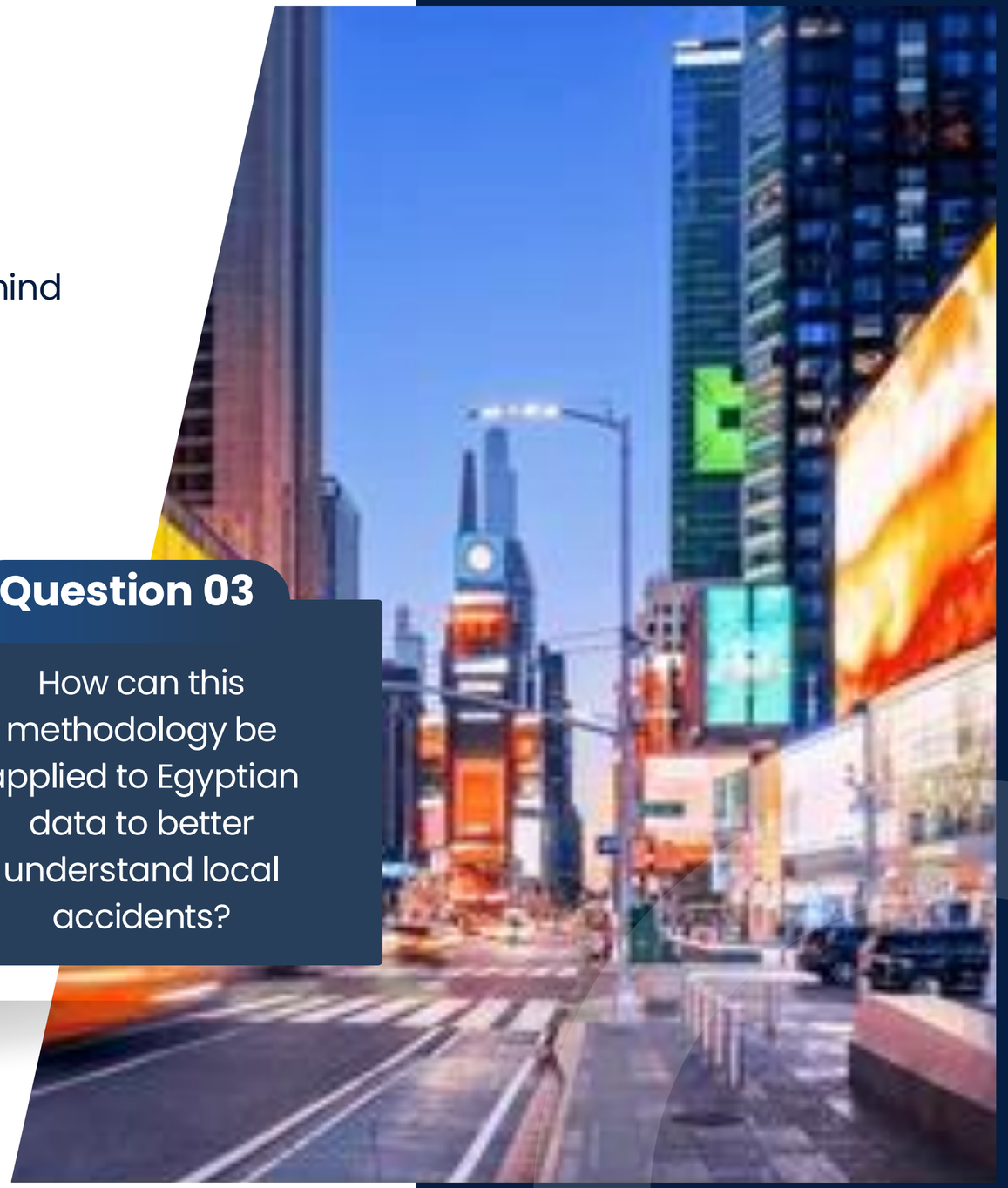
How has car accident data changed over the past few years?

Question 02

Which factors, such as traffic, weather, lighting, and road conditions, contribute most to accidents?

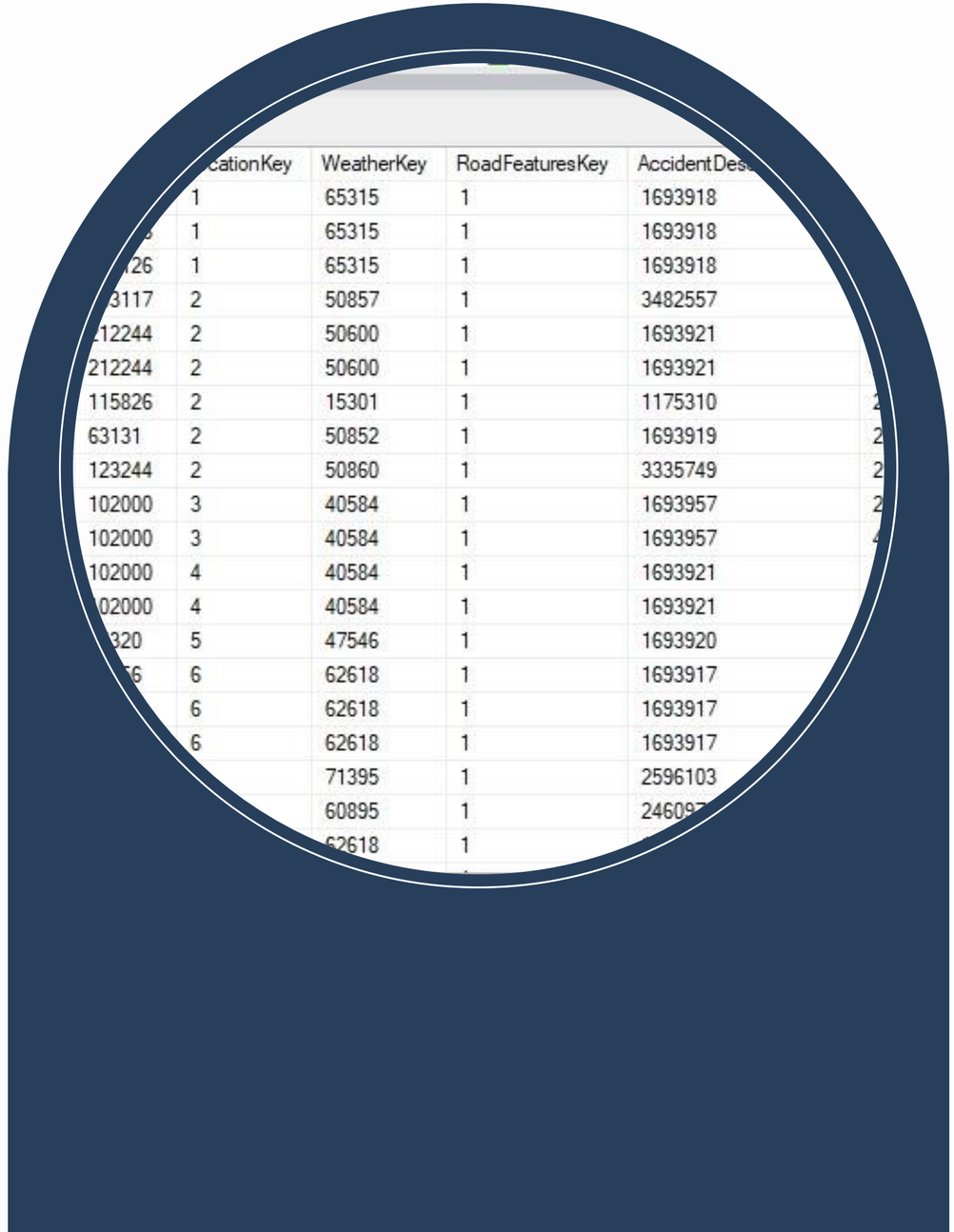
Question 03

How can this methodology be applied to Egyptian data to better understand local accidents?



03 Dataset Overview

- US Car Accidents (CSV)
- From Kaggle
- Time Range: from February 2016 to March 2023
- The data covers 49 States of the USA
- It contains approximately 7.7 million accident records



cationKey	WeatherKey	RoadFeaturesKey	AccidentDes
1	65315	1	1693918
1	65315	1	1693918
1	65315	1	1693918
2	50857	1	3482557
2	50600	1	1693921
2	50600	1	1693921
2	15301	1	1175310
2	50852	1	1693919
2	50860	1	3335749
3	40584	1	1693957
3	40584	1	1693957
4	40584	1	1693921
4	40584	1	1693921
5	47546	1	1693920
6	62618	1	1693917
6	62618	1	1693917
6	62618	1	1693917
	71395	1	2596103
	60895	1	246097
	62618	1	

04 Data Cleaning

Using Power Query

01 Loaded Data

02 Removed Duplicates & Irrelevant Records

03 Handled Missing Values

04 Created New Columns

05 Built Dimensional Tables

04 Data Cleaning

Using Power Query

01 Loaded Data

02 Removed Duplicates & Irrelevant Records

03 Handled Missing Values

04 Created New Columns

05 Built Dimensional Tables

04 Data Cleaning

Using Power Query

01 Loaded Data

02 Removed Duplicates & Irrelevant Records

03 Handled Missing Values

04 Created New Columns

05 Built Dimensional Tables

04 Data Cleaning

Using Power Query

01 Loaded Data

02 Removed Duplicates & Irrelevant Records

03 Handled Missing Values

04 Created New Columns

05 Built Dimensional Tables

04 Data Cleaning

Using Power Query

01 Loaded Data

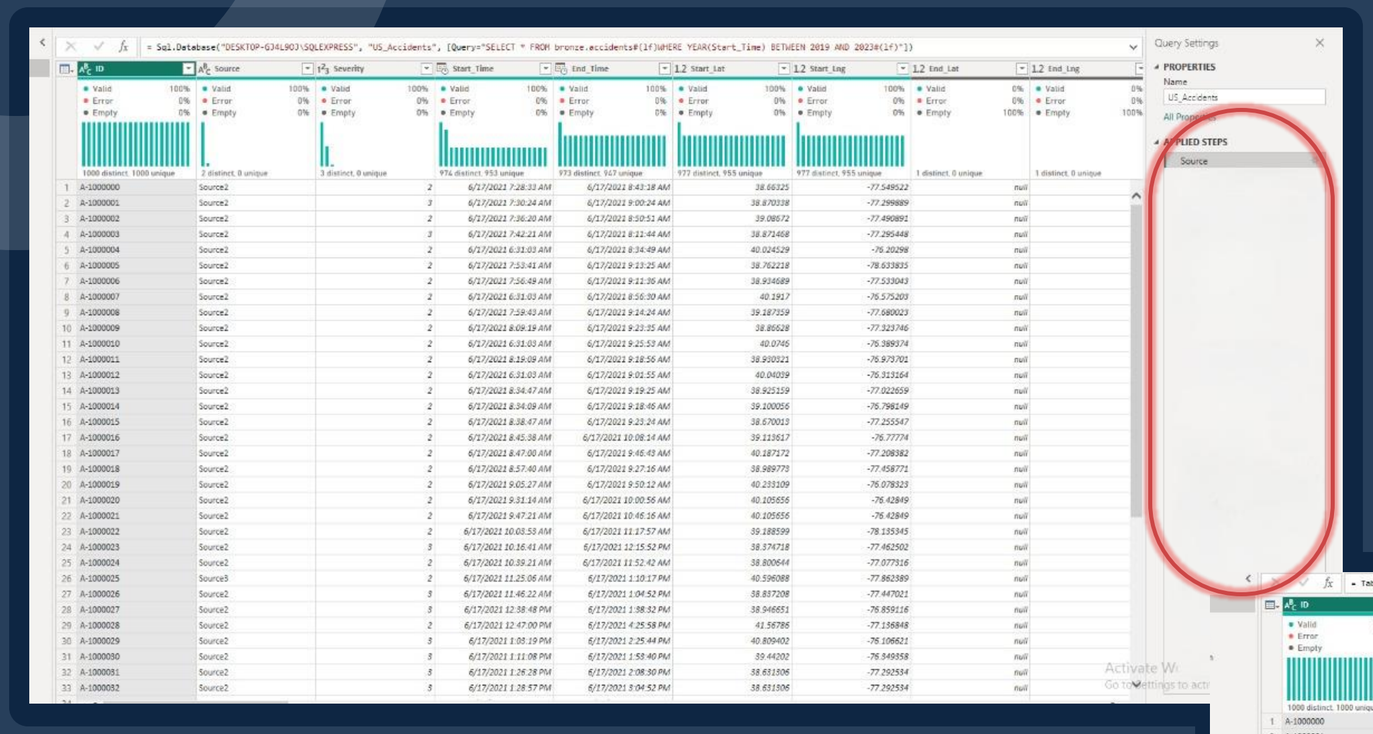
02 Removed Duplicates & Irrelevant Records

03 Handled Missing Values

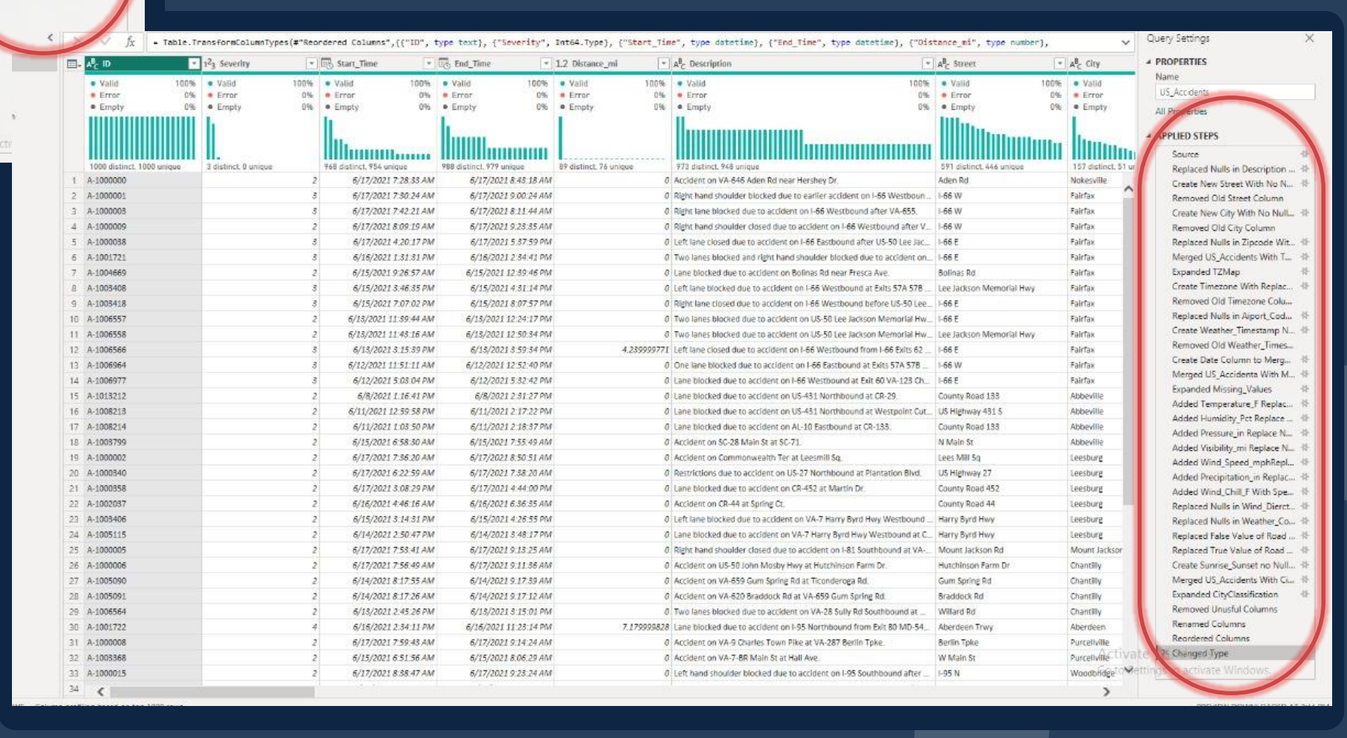
04 Created New Columns

05 Built Dimensional Tables

Before

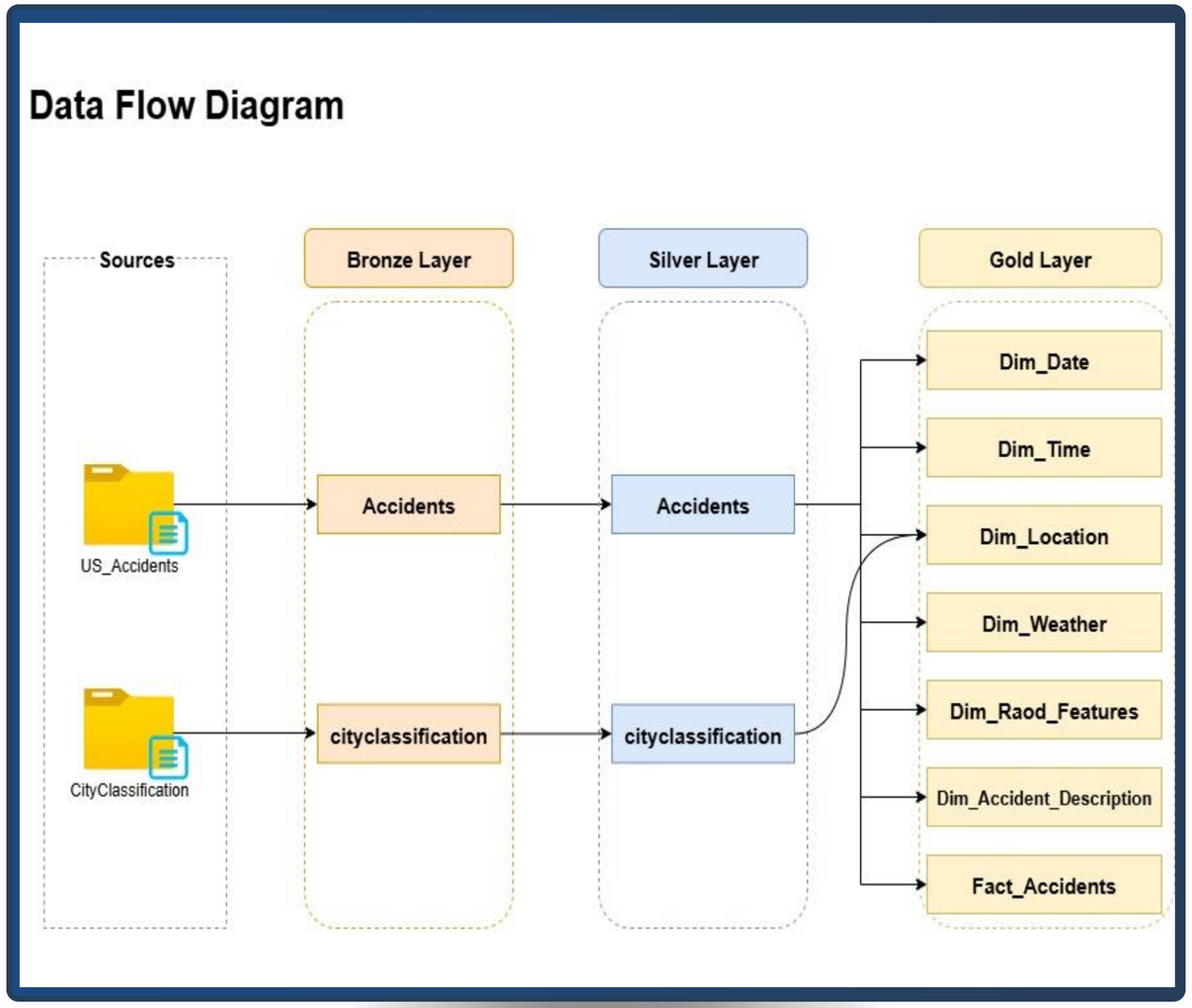


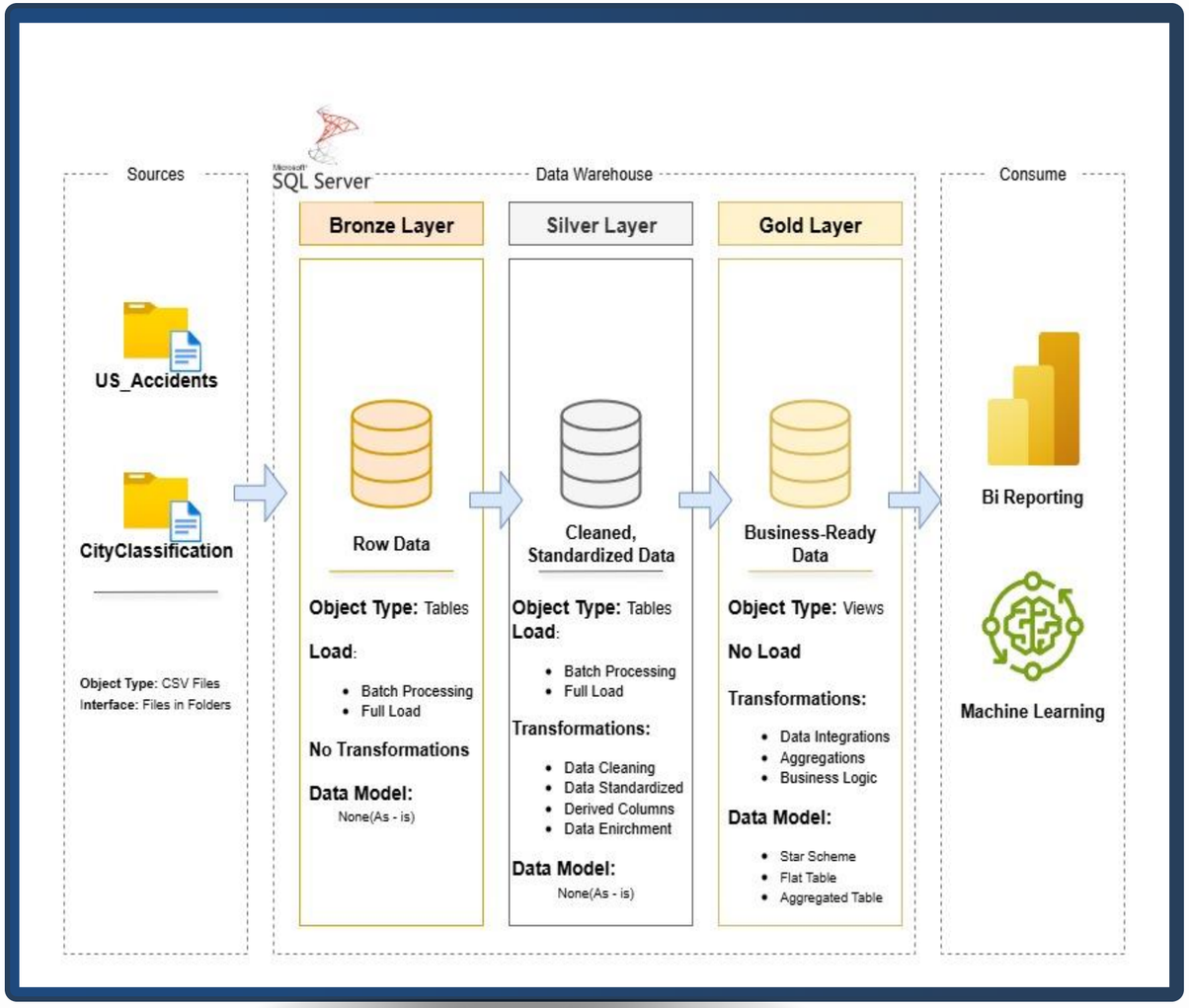
After



Data Cleaning

Using SQL





Data Cleaning Using SQL

ad_bronze_l...P-GJ4L9OJ\pc (59))* | DDL_Bronze_Layer.s...P-GJ4L9OJ\pc (58))*

```
EATE OR ALTER PROCEDURE bronze.load_bronze AS
GIN
  DECLARE @batch_start_time DATETIME, @batch_end_time DATETIME
  BEGIN TRY
      SET @batch_start_time = GETDATE();
      PRINT '==========================================';
      PRINT 'Loading Bronze Layer';
      PRINT '==========================================';


      PRINT '------------------------------------------';
      PRINT 'Loading US Accidents File'
      PRINT '------------------------------------------';

      PRINT '>> Truncating Table: bronze.accidents';
      TRUNCATE TABLE bronze.accidents;
      PRINT '>> Loading Date Into: bronze.accidents';
      BULK INSERT bronze.accidents
      FROM 'C:\DEPI Project\Database\US_Accidents_March23_UTF8.csv'
      WITH(
          FIRSTROW = 2,
          FIELDTERMINATOR = ',',
          ROWTERMINATOR = '\n',
          KEEPNULLS,
          TABLOCK
      )

      PRINT '------------------------------------------';
      PRINT 'Loading US CityClassification File'
      PRINT '------------------------------------------';

      PRINT '>> Truncating Table: bronze.cityclassification';
      TRUNCATE TABLE bronze.cityclassification;
      PRINT '>> Loading Date Into: bronze.cityclassification';
      BULK INSERT bronze.cityclassification
      FROM 'C:\DEPI Project\Database\CityClassification.csv'
      WITH(
          FIRSTROW = 2,
          FIELDTERMINATOR = ',',
          ROWTERMINATOR = '\n',
          KEEPNULLS,
          TABLOCK
      )
      SET @batch_end_time = GETDATE();
      PRINT '==============================================';
      PRINT 'Loading Bronze Layer is Completed';
      PRINT '>> Total Load Duration: ' + CAST(DATEDIFF(SECOND, @batch_start_time, @batch_end_time) AS NVARCHAR) + ' second';
      PRINT '==============================================';

  END TRY
  BEGIN CATCH
      PRINT '==============================================';
      PRINT 'ERROR OCCURED DURING LOADING BRONZE LAYER';
      PRINT 'Error Message: ' + ERROR_MESSAGE();
      PRINT 'Error Number: ' + CAST(ERROR_NUMBER() AS VARCHAR);
      PRINT 'Error Statue: ' + CAST(ERROR_STATE() AS VARCHAR);
      PRINT '==============================================';
  END CATCH;
D;

EC bronze.load_bronze
```

DDL_Gold_Layer.sql...P-GJ4L9OJ\pc (58))

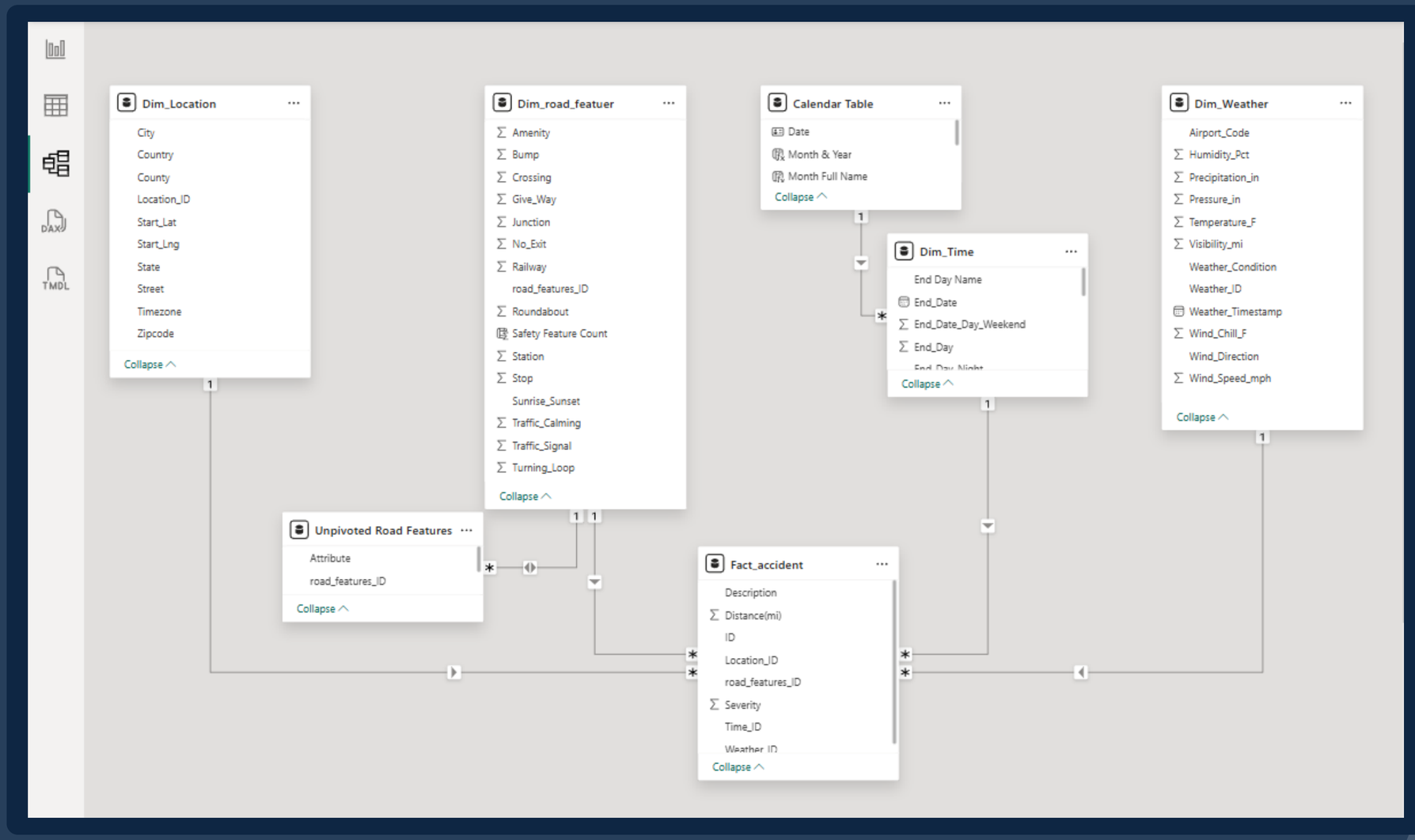
```
      Visibility_mi,
      Wind_Speed_mph,
      Precipitation_in
  FROM (
      SELECT DISTINCT
          State,
          CAST(Weather_Timestamp AS DATE) AS WeatherDate,
          DATEPART(HOUR, Weather_Timestamp) AS WeatherHour,
          Weather_Condition,
          Wind_Direction,
          Temperature_F,
          Humidity_Pct,
          Pressure_in,
          Visibility_mi,
          Wind_Speed_mph,
          Precipitation_in
      FROM silver.accidents
      WHERE Temperature_F IS NOT NULL AND Humidity_Pct IS NOT NULL AND Pressure_in IS NOT NULL AND Visibility_mi IS NOT NULL AND Wind_Speed
        AND Precipitation_in IS NOT NULL AND Weather_Condition IS NOT NULL
  ) t
  GO


CREATE OR ALTER VIEW gold.Dim_Raod_Features AS
  SELECT
      DENSE_RANK() OVER (
          ORDER BY Amenity, Bump, Crossing, Give_Way, Junction, No_Exit,
                   Railway, Roundabout, Station, Stop,
                   Traffic_Calming, Traffic_Signal, Turning_Loop
      ) AS RoadFeaturesKey,
      Amenity,
      Bump,
      Crossing,
      Give_Way,
      Junction,
      No_Exit,
      Railway,
      Roundabout,
      Station,
      Stop,
      Traffic_Calming,
      Traffic_Signal,
      Turning_Loop
  FROM (
      SELECT DISTINCT
          Amenity,
          Bump,
          Crossing,
          Give_Way,
          Junction,
          No_Exit,
          Railway,
          Roundabout,
          Station,
          Stop,
          Traffic_Calming,
          Traffic_Signal,
          Turning_Loop
      FROM silver.accidents
  ) t
  GO

CREATE OR ALTER VIEW gold.Dim_Accident_Description AS
  SELECT
      DENSE_RANK() OVER (ORDER BY Description) AS AccidentDescriptionKey,
      Description
  FROM (
      SELECT DISTINCT
          Description
      FROM silver.accidents
  ) t
  GO

CREATE OR ALTER VIEW gold.Fact_Accidents AS
  SELECT
```

Data Cleaning Using Python

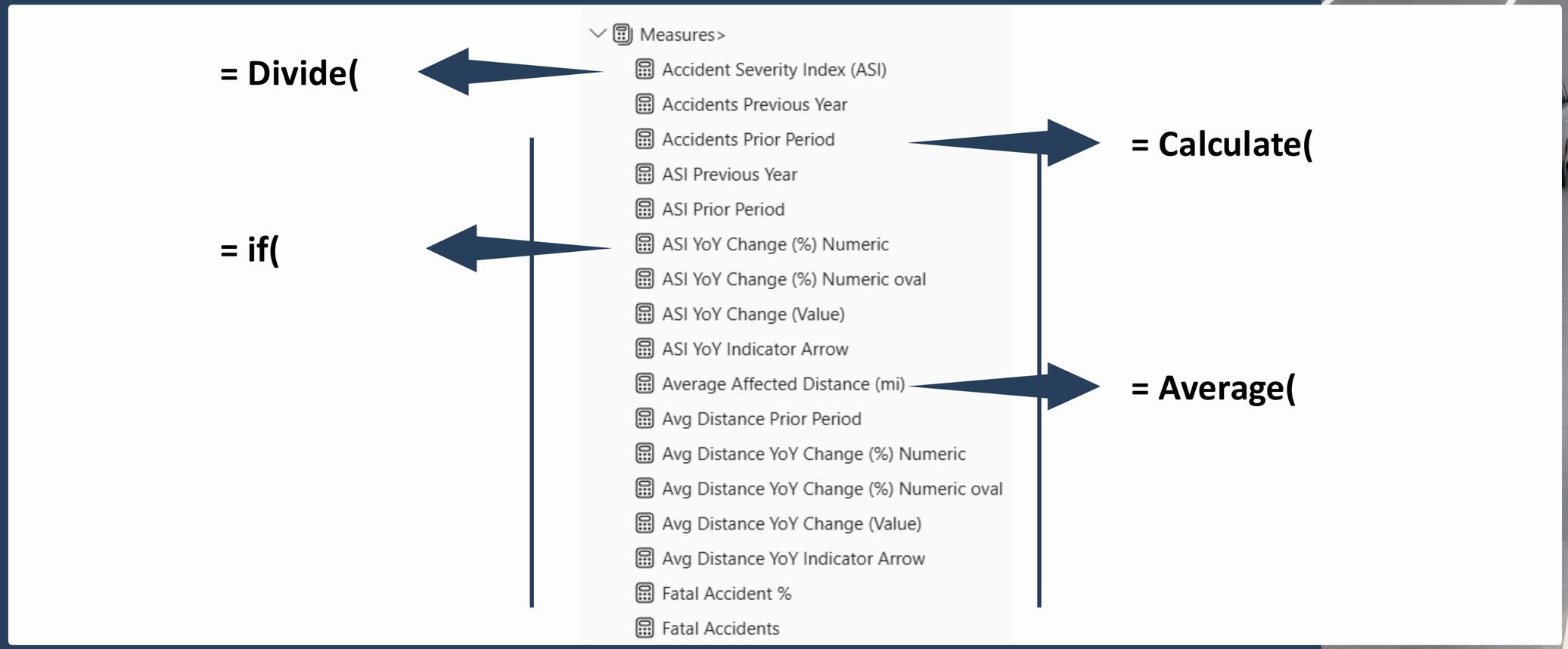
```
road_features = [
    "Amenity", "Bump", "Crossing", "Give_Way", "Junction", "No_Exit", "Railway",
    "Roundabout", "Station", "Stop", "Traffic_Calming", "Traffic_Signal",
    "Turning_Loop","Sunrise_Sunset", "Civil_Twilight", "Nautical_Twilight",
     "Astronomical_Twilight"
]

dim_road_features = Fact_Accident_with_WeatherID_LocationID[road_features].drop_duplicates().reset_index(drop=True)

# 2️⃣ Add surrogate key (road_features_ID starts at 3001 like in Power Query)
dim_road_features.insert(0, "road_features_ID", range(3001, 3001 + len(dim_road_features)))

# 3️⃣ Merge fact table with DimLocation to replace details with road_features_ID
Fact_Accident_with_WeatherID_LocationID_road_features_ID = Fact_Accident_with_WeatherID_LocationID.merge(dim_road_features, on=road_features, how="left")

# 4️⃣ Drop location details from fact table (keep only Location_ID as FK)
# fact_with_location_ID = fact_with_location_ID.drop(columns=location_cols)
Fact_Accident_with_WeatherID_LocationID_road_features_ID.head()
```

```
weather_cols = [
    "Airport_Code", "Weather_Timestamp", "Temperature_F", "Wind_Chill_F",
    "Humidity_Pct", "Pressure_in", "Visibility_mi", "Wind_Direction",
    "Wind_Speed_mph", "Precipitation_in", "Weather_Condition"
]

dimweather = df[weather_cols].drop_duplicates().reset_index(drop=True)

# 2️⃣ Add surrogate key (Weather_ID starts at 1001 like in Power Query)
dimweather.insert(0, "Weather_ID", range(1001, 1001 + len(dimweather)))

# 3️⃣ Merge fact table with DimWeather to replace details with Weather_ID
fact_with_weather_ID = df.merge(dimweather, on=weather_cols, how="left")
```

05 Data Modeling



06 Dax


= Divide(
= if(
Measures>
Accident Severity Index (ASI)
Accidents Previous Year
Accidents Prior Period
ASI Previous Year
ASI Prior Period
ASI YoY Change (%) Numeric
ASI YoY Change (%) Numeric oval
ASI YoY Change (Value)
ASI YoY Indicator Arrow
Average Affected Distance (mi)
Avg Distance Prior Period
Avg Distance YoY Change (%) Numeric
Avg Distance YoY Change (%) Numeric oval
Avg Distance YoY Change (Value)
Avg Distance YoY Indicator Arrow
Fatal Accident %
Fatal Accidents
= Calculate(
= Average(

07 Sample Visuals

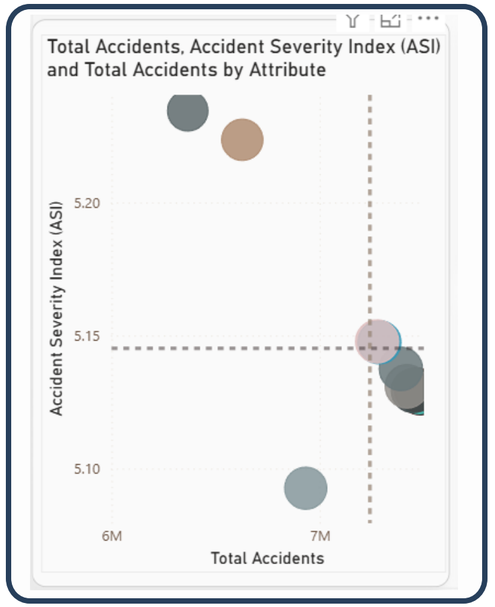
1) Table:

Showing the effect of different road features and distance on the number of accidents, and Accident Severity Index (ASI)

Feature	Accidents	ASI	Distance
Amenity	7,390,397.00	5.14	0.5
Bump	7,480,534.00	5.13	0.5
Crossing	6,628,206.00	5.22	0.5
Give_Way	7,448,797.00	5.13	0.5
Junction	6,933,051.00	5.09	0.5
No_Exit	7,464,873.00	5.13	0.5
Railway	7,419,423.00	5.13	0.5
Roundabout	7,483,716.00	5.13	0.5
Total	**7,483,959.00**	**5.13**	**0.5**

2) Bubble Chart:

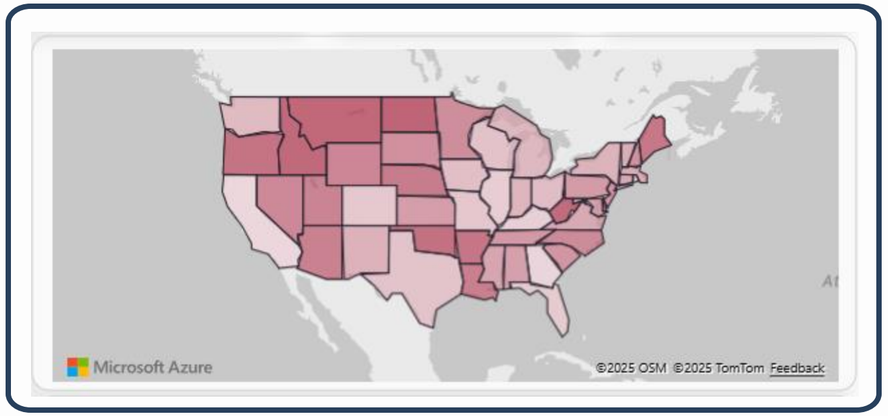
Showing Total accidents, Accident Severity Index (ASI), and total accidents by road feature



07 Sample Visuals

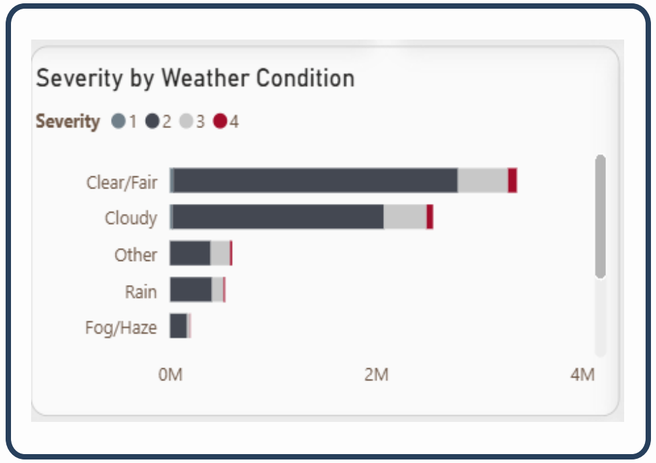
3) Map:

Showing number of accidents in each country

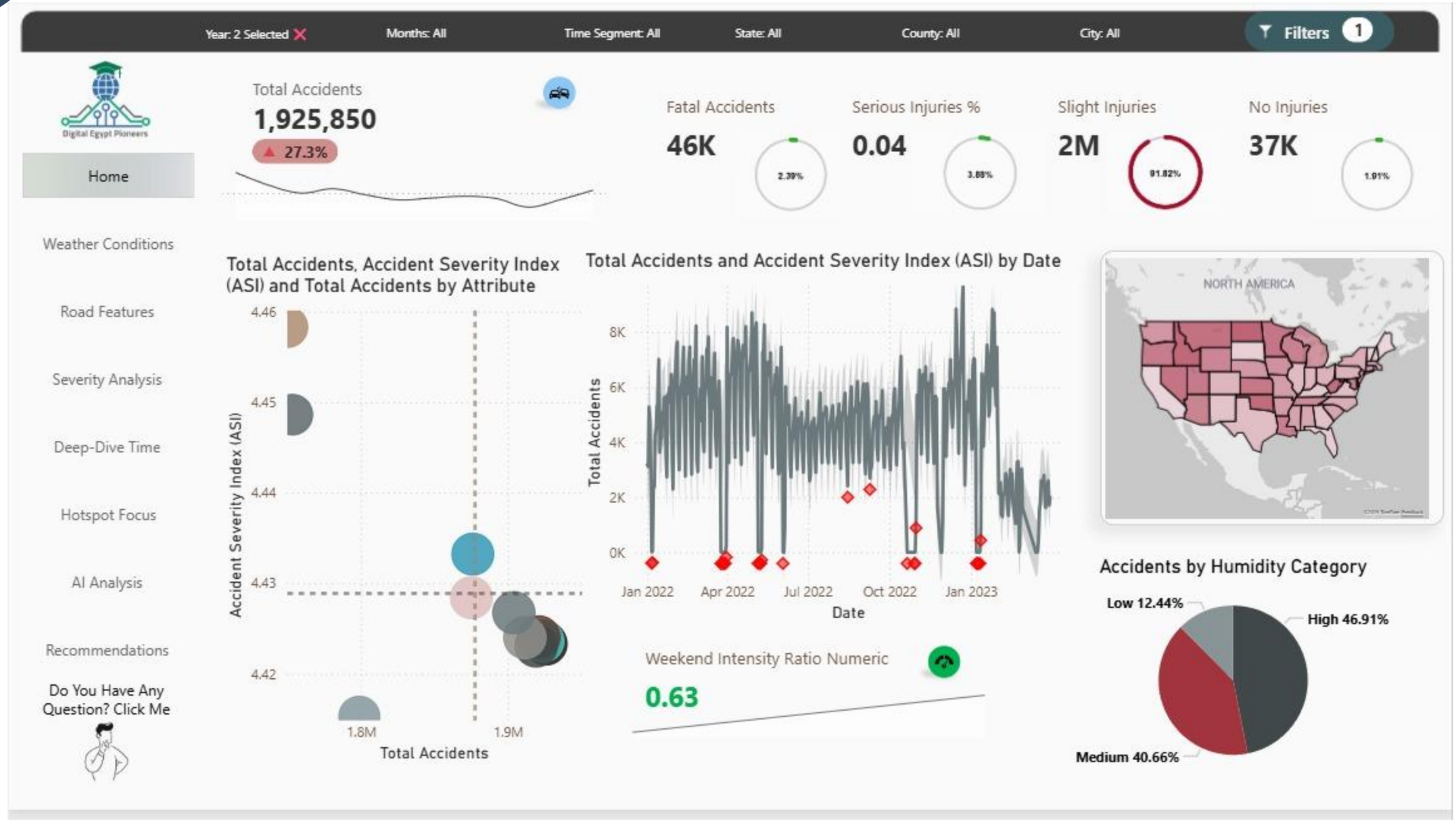


4) Bar Chart:

Showing Accidents Severity by different weather Conditions



9. Dashboard & Reports


Year: 2 Selected
Months: All
Time Segment: All
State: All
County: All
City: All
Filters 1
Digital Egypt Pioneers
Home
Weather Conditions
Road Features
Severity Analysis
Deep-Dive Time
Hotspot Focus
AI Analysis
Recommendations
Do You Have Any Question? Click Me
Total Accidents
1,925,850
27.3%
Fatal Accidents
46K
2.39%
Serious Injuries %
0.04
3.88%
Slight Injuries
2M
91.82%
No Injuries
37K
1.91%
Total Accidents, Accident Severity Index (ASI) and Total Accidents by Attribute
Accident Severity Index (ASI)
4.46
4.45
4.44
4.43
4.42
1.8M
1.9M
Total Accidents
Total Accidents and Accident Severity Index (ASI) by Date
Total Accidents
8K
6K
4K
2K
0K
Jan 2022
Apr 2022
Jul 2022
Oct 2022
Jan 2023
Date
Weekend Intensity Ratio Numeric
0.63
NORTH AMERICA
Accidents by Humidity Category
Low 12.44%
High 46.91%
Medium 40.66%

Year: 2 Selected | Months: All | Time Segment: All | State: All | County: All | City: All | Filters 1

Digital Egypt Pioneers

Home

Weather Conditions

Road Features

Severity Analysis

Deep-Dive Time

Hotspot Focus

AI Analysis

Recommendations

Do You Have Any Question? Click Me

Accidents in Adverse Weather

969K

accident rate at adverse weather

50.32%

Accident Count High Pressure

404K

accident rate at high pressure

20.96%

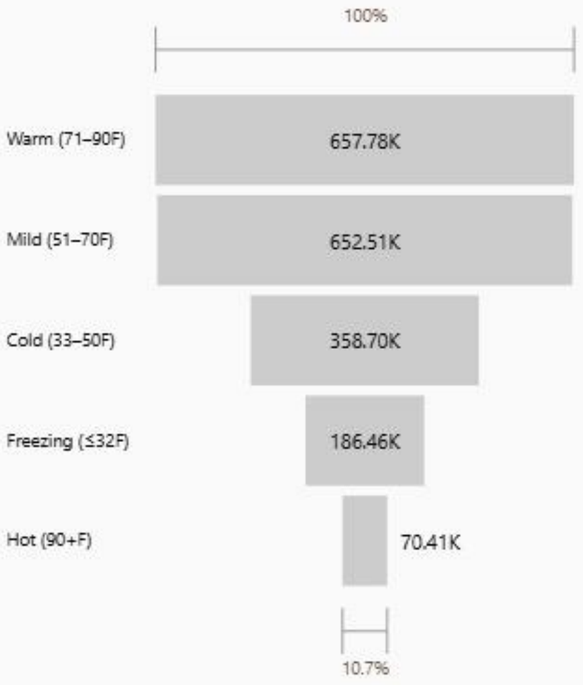
Accident At High Wind Speed

603K

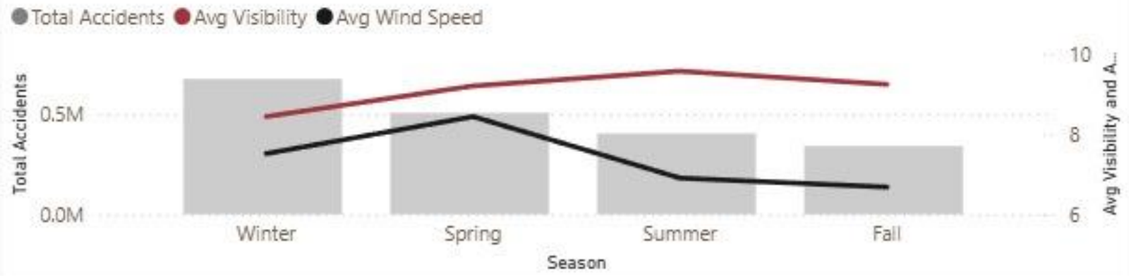
accident rate at high wind speed

31.33%

Accidents by Temperature Category



Total Accidents by Season and Avg visibility

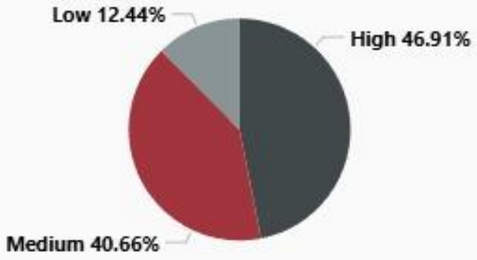


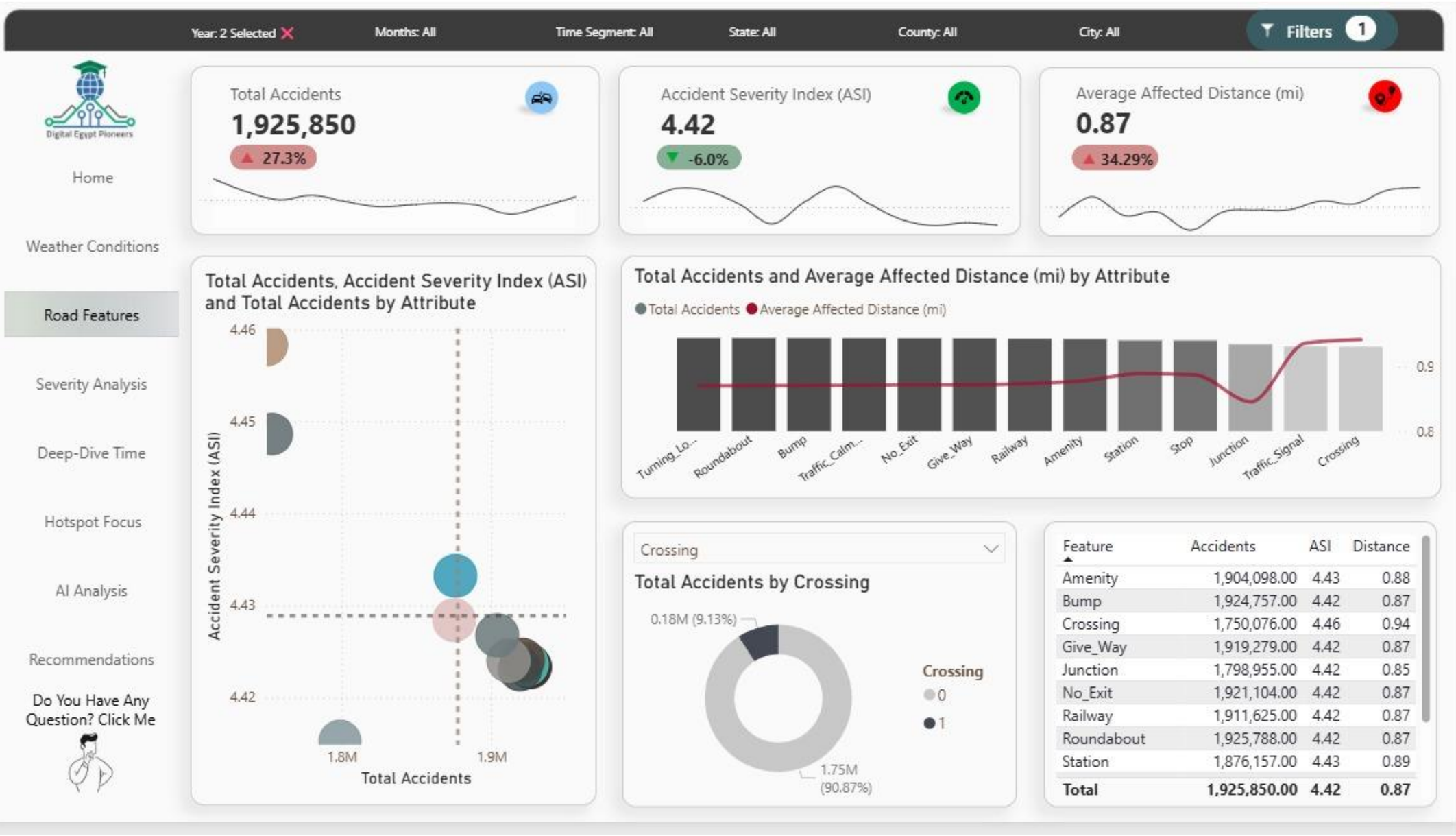
Fair
942,411.00
Total Accidents

Cloudy
267,075.00
Total Accidents

Mostly Cloudy

Accidents by Humidity Category




Year: 2 Selected
Months: All
Time Segment: All
State: All
County: All
City: All
Filters 1
Digital Egypt Pioneers
Home
Weather Conditions
Road Features
Severity Analysis
Deep-Dive Time
Hotspot Focus
AI Analysis
Recommendations
Do You Have Any Question? Click Me
Total Accidents
1,925,850
27.3%
Accident Severity Index (ASI)
4.42
-6.0%
Average Affected Distance (mi)
0.87
34.29%
Total Accidents, Accident Severity Index (ASI) and Total Accidents by Attribute
4.46
4.45
4.44
4.43
4.42
Accident Severity Index (ASI)
1.8M
1.9M
Total Accidents
Total Accidents and Average Affected Distance (mi) by Attribute
Total Accidents
Average Affected Distance (mi)
0.9
0.8
Turning_Lo...
Roundabout
Bump
Traffic_Calm...
No_Exit
Give_Way
Railway
Amenity
Station
Stop
Junction
Traffic_Signal
Crossing
Crossing
Total Accidents by Crossing
0.18M (9.13%)
1.75M (90.87%)
Crossing
0
1
Feature
Accidents
ASI
Distance
Amenity 1,904,098.00 4.43 0.88
Bump 1,924,757.00 4.42 0.87
Crossing 1,750,076.00 4.46 0.94
Give_Way 1,919,279.00 4.42 0.87
Junction 1,798,955.00 4.42 0.85
No_Exit 1,921,104.00 4.42 0.87
Railway 1,911,625.00 4.42 0.87
Roundabout 1,925,788.00 4.42 0.87
Station 1,876,157.00 4.43 0.89
Total 1,925,850.00 4.42 0.87

Year: 2 Selected | Months: All | Time Segment: All | State: All | County: All | City: All | Filters 1

Digital Egypt Pioneers

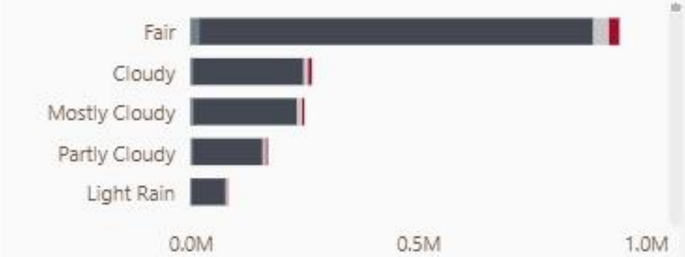
- Home
- Weather Conditions
- Road Features
- Severity Analysis
- Deep-Dive Time
- Hotspot Focus
- AI Analysis
- Recommendations

Do You Have Any Question? Click Me

Fatal Accidents	Serious Injuries %	Slight Injuries	No Injuries
46K	0.04	2M	37K
2.39%	3.88%	91.82%	1.91%

Severity by Weather Condition

Severity 1 2 3 4



Severity Over Years, Months

Severity 1 2 3 4

Top 5 Most repeated Accident Description

- A crash has occurred causing no to minimum delays. Use ...
- A crash has occurred with minimal delay to traffic. Prepar...
- A crash has occurred. Prepare to slow or move over for w...
- Accident
- An unconfirmed report of a crash has been received. Use ...

Severity by Wind Direction

Severity by Light Condition

Year: 2 Selected | Months: All | Time Segment: All | State: All | County: All | City: All | Filters 1

Digital Egypt Pioneers

- Home
- Weather Conditions
- Road Features
- Severity Analysis
- Deep-Dive Time
- Hotspot Focus
- AI Analysis
- Recommendations

Do You Have Any Question? Click Me

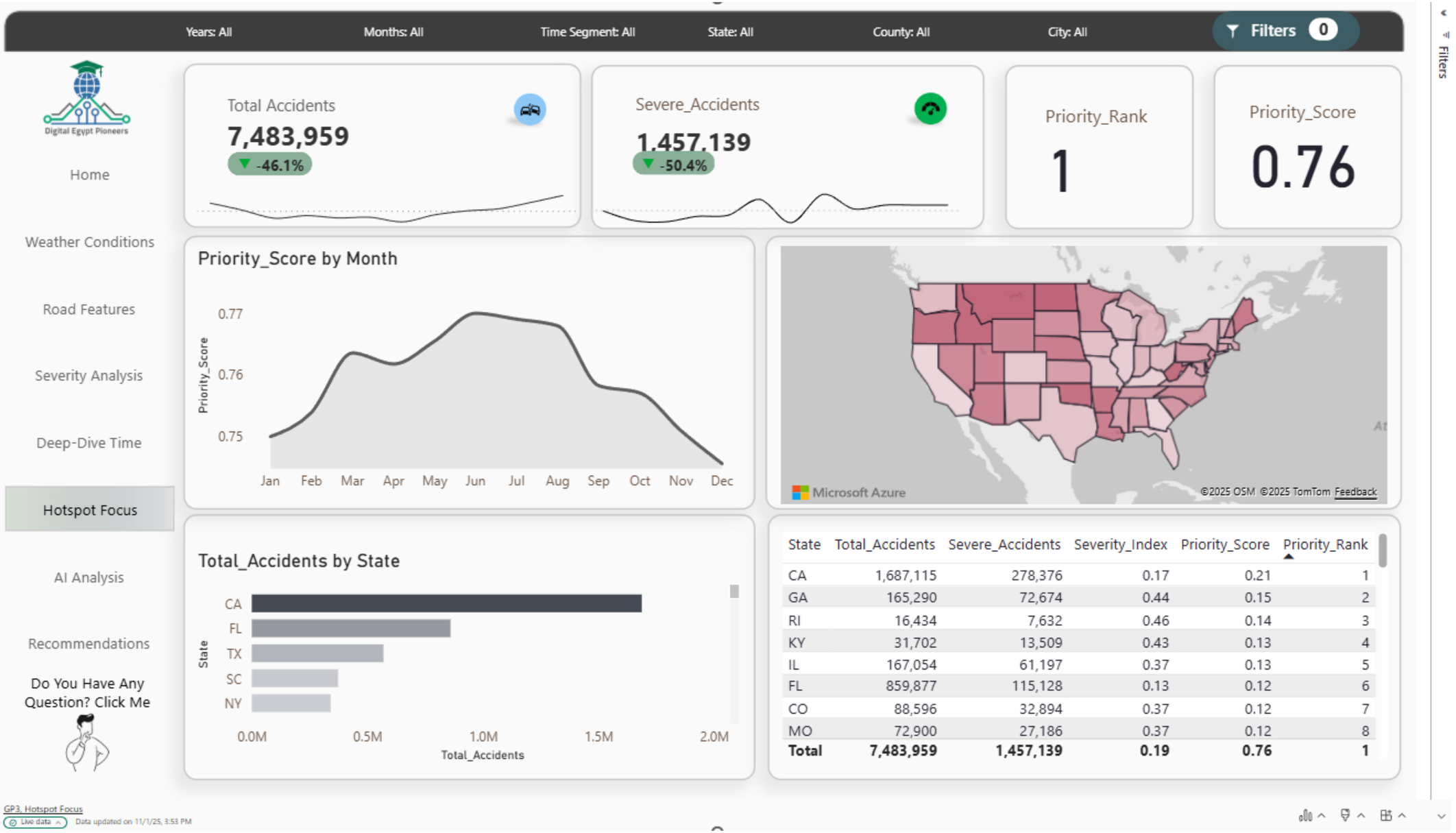
Total Accidents
1,925,850
▲ 27.3%

Weekend Intensity Ratio Numeric
0.63
▲ 0.03

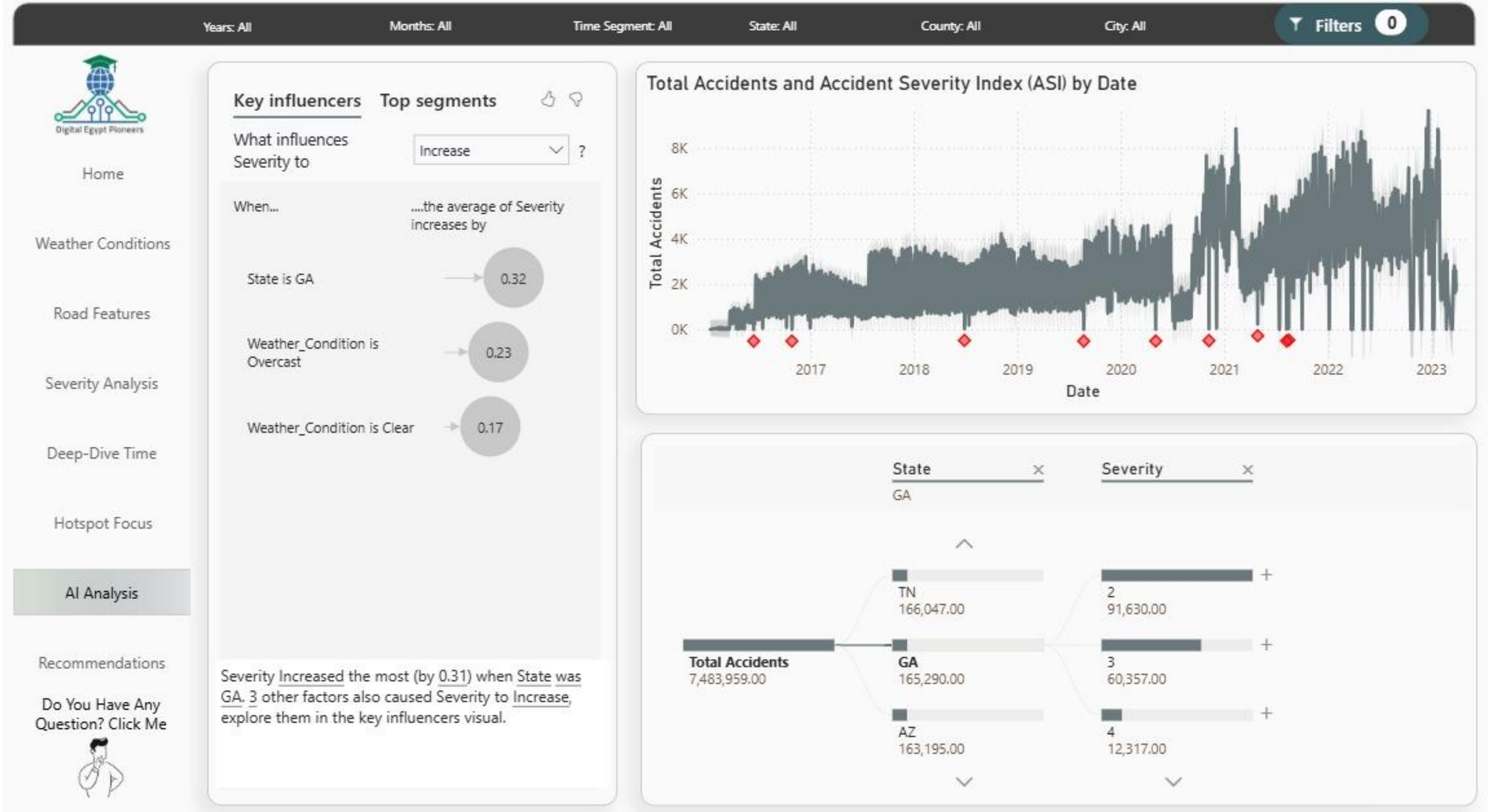
Peak Hour Ratio %
0.35
▲ 0.00

Total Accidents and Seasonal Trend Analysis by Month Name

Month Name	Friday	Monday	Saturday	Sunday	Thursday	Tuesday	Wednesday
April	61,558.00	41,383.00	36,739.00	21,965.00	44,757.00	43,399.00	44,046.00
August	36,068.00	33,334.00	23,349.00	20,228.00	34,098.00	39,346.00	39,698.00
December	56,750.00	31,791.00	38,691.00	22,922.00	53,231.00	37,340.00	38,528.00
February	58,890.00	46,340.00	29,466.00	25,263.00	59,669.00	52,516.00	55,242.00
January	69,617.00	74,007.00	50,577.00	51,764.00	58,654.00	70,244.00	64,520.00
July	41,135.00	25,743.00	27,934.00	24,085.00	31,416.00	29,426.00	29,977.00
June	25,064.00	28,959.00	16,148.00	14,848.00	39,117.00	28,415.00	40,802.00
March	42,027.00	32,734.00	25,445.00	21,447.00	43,194.00	43,205.00	46,762.00
May	34,272.00	35,659.00	20,226.00	21,478.00	40,529.00	42,247.00	44,459.00
November	30,397.00	25,047.00	19,700.00	15,941.00	31,157.00	36,817.00	35,611.00



State	Total_Accidents	Severe_Accidents	Severity_Index	Priority_Score	Priority_Rank
CA	1,687,115	278,376	0.17	0.21	1
GA	165,290	72,674	0.44	0.15	2
RI	16,434	7,632	0.46	0.14	3
KY	31,702	13,509	0.43	0.13	4
IL	167,054	61,197	0.37	0.13	5
FL	859,877	115,128	0.13	0.12	6
CO	88,596	32,894	0.37	0.12	7
MO	72,900	27,186	0.37	0.12	8
Total	**7,483,959**	**1,457,139**	**0.19**	**0.76**	**1**


Years: All
Months: All
Time Segment: All
State: All
County: All
City: All
Filters 0
Digital Egypt Pioneers
Home
Weather Conditions
Road Features
Severity Analysis
Deep-Dive Time
Hotspot Focus
AI Analysis
Recommendations
Do You Have Any Question? Click Me
Key influencers
Top segments
What influences Severity to
Increase
?
When...
....the average of Severity increases by
State is GA
0.32
Weather_Condition is Overcast
0.23
Weather_Condition is Clear
0.17
Severity Increased the most (by 0.31) when State was GA. 3 other factors also caused Severity to Increase, explore them in the key influencers visual.
Total Accidents and Accident Severity Index (ASI) by Date
Total Accidents
8K
6K
4K
2K
0K
2017
2018
2019
2020
2021
2022
2023
Date
State
GA
Severity
Total Accidents
7,483,959.00
GA
165,290.00
TN
166,047.00
AZ
163,195.00
2
91,630.00
3
60,357.00
4
12,317.00

Years: All | Months: All | Time Segment: All | State: All | County: All | City: All | Filters 0

Digital Egypt Pioneers

- Home
- Weather Conditions
- Road Features
- Severity Analysis
- Deep-Dive Time
- Hotspot Focus
- AI Analysis
- Recommendations

Do You Have Any Question? Click Me

Wind Speed	Visibility	Temperature
25	2	90

Accident_Risk_Score

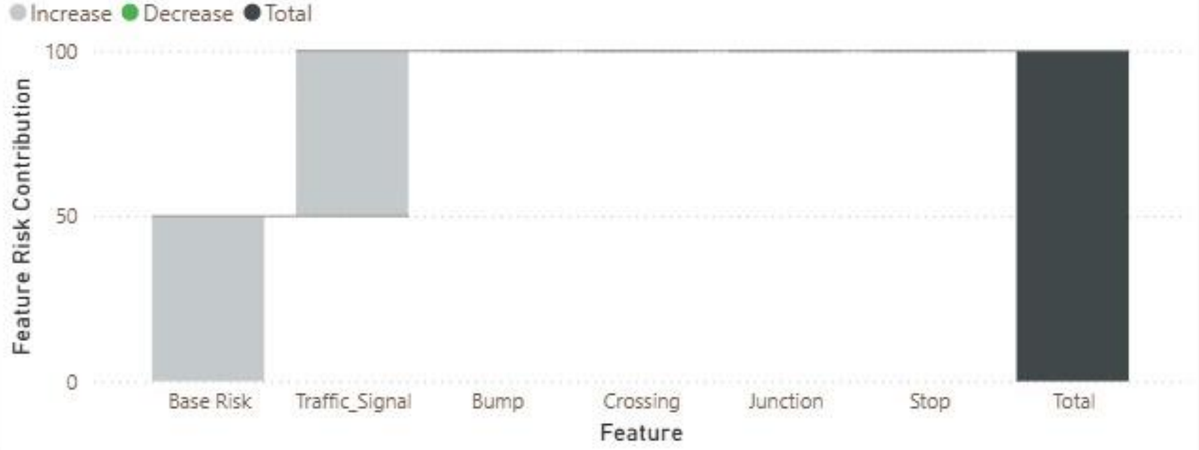
High Risk: Unsafe to travel under current conditions.

Junction	Crossing	Traffic Signal	Bump	Stop
0	0	1	0	0

Simulated Road Feature Risk

Low Risk: Road is safe.

Feature Risk Contribution by Road Feature



09 Key Results

California (CA) has the highest number of Accidents

1.6M

Accidents

followed by **Florida (FL)** and **Georgia (GA).**

Accidents Severity:

97.67%

Slight Injuries

2.58%

Fatal Injuries

Fair is the most common weather

1.3M

Accidents

- **Rain, Cloudy, and Clear** conditions show higher severity.

Road Features:

94.57%

No Crossing

7.3M

Accidents at Junctions

09 Key Results

- Accidents peak in **December, January, and November**
- **Tuesday** has the highest number of accidents, especially in **December**.

3M

Accidents during day

2M

Accidents At Night

10 Recommendation

Focus on California

Allocate resources, awareness campaigns, and infrastructure reviews.

Prioritize Seasonal & Weekly Safety

Increase alerts in December, January, November, and focus on Tuesdays.

Review Road Infrastructure

Assess high-accident junctions and no-crossing areas to improve safety.

THANK YOU!