|  |  |  |
| --- | --- | --- |
| **University of Petra** | ***شعار جامعة البترا5 -*** | **جامعة البترا** |
| **Faculty of Information Technology** | **كلية تكنولوجيا المعلومات** |

**Data Exploration & Visualization**

**Analyzing Traffic Accident Data to Improve Road Safety**

**Prepared By:**

**Kinda Dardese (202110811)**

**Nagham Alotaibi (202110560)**

**Eng.Ghayda AlHyasat**

***1. Project Abstract:***

The project aims to analyze Pennsylvania crash data, focusing on Washington County's 2015 crash records, to identify trends and patterns in road accidents. By exploring factors such as crash locations, contributing causes, and time-specific trends, the analysis seeks to provide actionable insights for improving road safety strategies. Through data visualization, the project will support policymakers in making informed decisions to enhance safety measures and reduce traffic incidents.

***2. Project Requirements:***

***Data Source:***

The dataset used for this project is titled 2015washington.csv, a subset of the Washington County crash data available on Kaggle. This dataset provides detailed records of accidents that occurred in 2015, formatted as CSV (Comma-Separated Values) files. It includes 188 columns and 945 rows, encompassing comprehensive crash-related information. Additionally, the project utilizes supplementary crash data provided in MS Excel, which contains related tables such as CRASH, PERSON, ROADWAY, FLAG, and others, all stored in CSV format. These datasets collectively provide a robust foundation for analyzing and visualizing traffic accident trends.

***Power BI Tools:***

**Import Process**:

Connect Power BI and get data type csv.

Use Power Query to load required tables (CRASH, FLAG, ROADWAY, etc.).

**Transformation**:

Clean and standardize column names.

Remove duplicates and irrelevant fields.

Convert text fields like dates (CRASH\_DATE) to datetime formats.

Handle null values by substituting with default placeholders or filtering out incomplete rows.

**Data Model:**

Create a relational data model linking tables based on common fields like Accident ID or Location

Develop a star schema if additional data tables are integrated, with a fact table containing accident details and dimension tables for dates, locations, and weather

Establish relationships between tables:

CRASH (primary table) linked with FLAG, PERSON, ROADWAY, and VEHICLE using the CRN field.

***DAX Calculations:***

**Calculated Columns:**

Season (ex: Summer/Winter) derived from CRASH\_DATE.

Crash Severity categorized into Minor, Major, and Fatal based on FATAL\_COUNT.

Accident Severity Category to group accidents into severity levels.

Accident Hour to extract the hour from the timestamp for time-based analysis.

**Measures**:

Total Fatal Crashes = COUNTX(FILTER(CRASH, CRASH[FATAL\_COUNT] > 0), CRASH[CRN])

Injury Rate by Crash = AVERAGEX(CRASH, CRASH[TOT\_INJ\_COUNT])

Total Accidents = COUNT(CRASH[CRN])

Accidents per Location = COUNTX(VALUES(CRASH[LOCATION]), CRASH[CRN])

Accidents by Weather = COUNTX(FILTER(CRASH, NOT(ISBLANK(CRASH[WEATHER\_CONDITION]))), CRASH[CRN]

Accidents per Day = COUNTX(VALUES(CRASH[DAY\_OF\_WEEK]), CRASH[CRN])

**Calculated Tables**:

Filtered table for crashes involving alcohol:

FILTER(CRASH, FLAG[ALCOHOL\_RELATED] = 1)

A custom date table to enable time intelligence features for year, month, and day-level analyses:

DateTable =

(ADDCOLUMNS

CALENDAR(DATE(2015, 1, 1), DATE(2015, 12, 31)),

"Year", YEAR([Date]),

"Month", FORMAT([Date], "MMMM"),

"MonthNumber", MONTH([Date]),

"Quarter", "Q" & QUARTER([Date]),

"Weekday", FORMAT([Date], "dddd"),

Day", DAY([Date]),

WeekNumber", WEEKNUM([Date], 1) ",

YearMonth", FORMAT([Date], "YYYY-MM")) "

***Visualizations:***

**Charts and Graphs**:

**Bar Chart**:

Crash count by day of the week.

Accidents by severity

**Line Chart**:

Monthly crash trends over the past five years.

Trends in accidents over months

**Pie Chart**:

Proportion of crash types (ex: Alcohol-related, Aggressive driving).

Accidents distribution by weather conditions

**Map**:

Geographical distribution of crashes by severity.

Geographical distribution of accidents

**KPIs**:

Total Fatal Crashes.

Average injuries per crash.

**Cards:**

display key metrics like total accidents

**Slicers and Filters**:

Filters for year, location (county/municipality), crash type, severity, and weather conditions to enable interactivity.

Slicers for dynamic interactivity:

Select crash severity.

Time period range.