create database road\_accidents;

use road\_accidents;

#creating a table which has same columns as the dataset,

CREATE TABLE road\_accidents (

Accident\_Index VARCHAR(20),

Accident\_Date DATE,

Day\_of\_Week VARCHAR(20),

Junction\_Control VARCHAR(50),

Junction\_Detail VARCHAR(50),

Accident\_Severity VARCHAR(20),

Latitude FLOAT,

Light\_Conditions VARCHAR(50),

Local\_Authority\_District VARCHAR(100),

Carriageway\_Hazards VARCHAR(100),

Longitude FLOAT,

Number\_of\_Casualties INT,

Number\_of\_Vehicles INT,

Police\_Force VARCHAR(50),

Road\_Surface\_Conditions VARCHAR(50),

Road\_Type VARCHAR(50),

Speed\_limit INT,

Time TIME,

Urban\_or\_Rural\_Area VARCHAR(20),

Weather\_Conditions VARCHAR(100),

Vehicle\_Type VARCHAR(50)

);

SHOW VARIABLES LIKE 'secure\_file\_priv';

#importing Dataset

LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/road\_accident.csv'

INTO TABLE road\_accidents

FIELDS TERMINATED BY ','

ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 ROWS

(

Accident\_Index,

@Accident\_Date, -- use variable here

Day\_of\_Week,

Junction\_Control,

Junction\_Detail,

Accident\_Severity,

Latitude,

Light\_Conditions,

Local\_Authority\_District,

Carriageway\_Hazards,

Longitude,

Number\_of\_Casualties,

Number\_of\_Vehicles,

Police\_Force,

Road\_Surface\_Conditions,

Road\_Type,

Speed\_limit,

Time,

Urban\_or\_Rural\_Area,

Weather\_Conditions,

Vehicle\_Type

)

#changing the datatype of Accident\_Date

SET Accident\_Date = STR\_TO\_DATE(@Accident\_Date, '%d-%m-%Y');

#changing the datatype of Time

UPDATE road\_accidents SET Time = STR\_TO\_DATE(Time, '%H:%i') WHERE Time IS NOT NULL;

#displaying the first 10 rows from the dataset/table.

SELECT \* FROM road\_accidents LIMIT 10;

#displays column name and its datatype.

DESCRIBE road\_accidents;

### Univariate Analysis

#displays the unique values from Time Column

SELECT DISTINCT Time FROM road\_accidents ORDER BY Time LIMIT 10;

#displays the count of accidents (rows of the dataset)

SELECT COUNT(\*) FROM road\_accidents;

#count of accidents happened in each weather condition

SELECT Weather\_Conditions, COUNT(\*) AS total FROM road\_accidents GROUP BY Weather\_Conditions ORDER BY total DESC;

#count of accidents happened in each Speed Limit

SELECT Speed\_limit, COUNT(\*) AS total FROM road\_accidents GROUP BY Speed\_limit ORDER BY Speed\_limit;

#displays count in every Accident\_Severity column's data, can know frequency.

SELECT Accident\_Severity, COUNT(\*) AS total FROM road\_accidents GROUP BY Accident\_Severity;

#changing the anomaly. From 'Fetal' to 'Fatal'

UPDATE road\_accidents SET Accident\_Severity = 'Fatal' WHERE Accident\_Severity = 'Fetal';

#displays number of accidents happened in every location.

SELECT Local\_Authority\_District, COUNT(\*) AS total FROM road\_accidents GROUP BY Local\_Authority\_District ORDER BY total DESC LIMIT 10;

#displays accident count on each day.

SELECT Day\_of\_Week, COUNT(\*) AS total FROM road\_accidents GROUP BY Day\_of\_Week ORDER BY FIELD(Day\_of\_Week, 'Monday','Tuesday','Wednesday','Thursday','Friday','Saturday','Sunday');

#creates a new column Time\_of\_Day which hold in which part of time in the day accident was happened,

ALTER TABLE road\_accidents ADD COLUMN Time\_of\_Day VARCHAR(20);

#classifies the data into parts of the day based on Time Column.

UPDATE road\_accidents SET Time\_of\_Day = CASE

WHEN HOUR(Time) BETWEEN 0 AND 5 THEN 'Midnight'

WHEN HOUR(Time) BETWEEN 6 AND 11 THEN 'Morning'

WHEN HOUR(Time) BETWEEN 12 AND 16 THEN 'Afternoon'

WHEN HOUR(Time) BETWEEN 17 AND 20 THEN 'Evening'

WHEN HOUR(Time) BETWEEN 21 AND 23 THEN 'Night'

ELSE 'Unknown'

END;

#displays accident count in each part of the day.

SELECT Time\_of\_Day, COUNT(\*) AS total FROM road\_accidents GROUP BY Time\_of\_Day ORDER BY FIELD(Time\_of\_Day, 'Midnight','Morning','Afternoon','Evening','Night','Unknown');

###Bivariate Analysis

#displays the number of accident in every Accident\_Severity and Time\_of\_Day combination.

#we can know that which time in a day is dangerous like night or evening

SELECT Time\_of\_Day, Accident\_Severity, COUNT(\*) AS total FROM road\_accidents GROUP BY Time\_of\_Day, Accident\_Severity ORDER BY FIELD(Time\_of\_Day, 'Midnight','Morning','Afternoon','Evening','Night','Unknown'), total DESC;

#displays the number of accident in every Accident\_Severity and Weather\_Condition combination.

#we can know that which weather is dangerous

SELECT Weather\_Conditions, Accident\_Severity, COUNT(\*) AS total FROM road\_accidents GROUP BY Weather\_Conditions, Accident\_Severity ORDER BY Weather\_Conditions;

#displays the number of accident in every Accident\_Severity and Road\_Type.

#we can know that which type of road is dangerous

SELECT Road\_Type, Accident\_Severity, COUNT(\*) AS total FROM road\_accidents GROUP BY Road\_Type, Accident\_Severity ORDER BY Road\_Type;

#displays the number of accident in every Accident\_Severity and Vehicle\_Type combination.

#we can know that which type of vehicle types gets dangerous accidents

SELECT Vehicle\_Type, Accident\_Severity, COUNT(\*) AS total FROM road\_accidents GROUP BY Vehicle\_Type, Accident\_Severity ORDER BY Vehicle\_Type;

#displays the number of accident in every Accident\_Severity and Speed\_Limit combination.

#we can know that which speed limit zones are dangerous

SELECT Speed\_limit, Accident\_Severity, COUNT(\*) AS total FROM road\_accidents GROUP BY Speed\_limit, Accident\_Severity ORDER BY Speed\_limit;

###Multivariate Analysis

#displays the number of accident in every Accident\_Severity, Time\_of\_Day and Weather\_Condition combination.

#we can know that in which type of weather and what time in a day causes dangerous accident.

SELECT Time\_of\_Day, Weather\_Conditions, Accident\_Severity, COUNT(\*) AS total FROM road\_accidents GROUP BY Time\_of\_Day, Weather\_Conditions, Accident\_Severity ORDER BY total DESC;

#displays the number of accident in every Accident\_Severity, Road\_Type and Speed\_Limit combination.

#we can know that in which type of road at what speed zone causes dangerous accidents.

SELECT Speed\_limit,Road\_Type, Accident\_Severity, COUNT(\*) AS total FROM road\_accidents GROUP BY Speed\_limit, Road\_Type, Accident\_Severity ORDER BY Speed\_limit, total DESC;

#displays the number of accident in every Accident\_Severity, Urban\_or\_Rural\_Area and Time\_of\_Day combination

#we can know that in which type of area during which Time\_of\_Day causes dangerous accidents

SELECT Urban\_or\_Rural\_Area, Time\_of\_Day, Accident\_Severity,COUNT(\*) AS total FROM road\_accidents GROUP BY Urban\_or\_Rural\_Area, Time\_of\_Day, Accident\_Severity ORDER BY Urban\_or\_Rural\_Area, Time\_of\_Day;