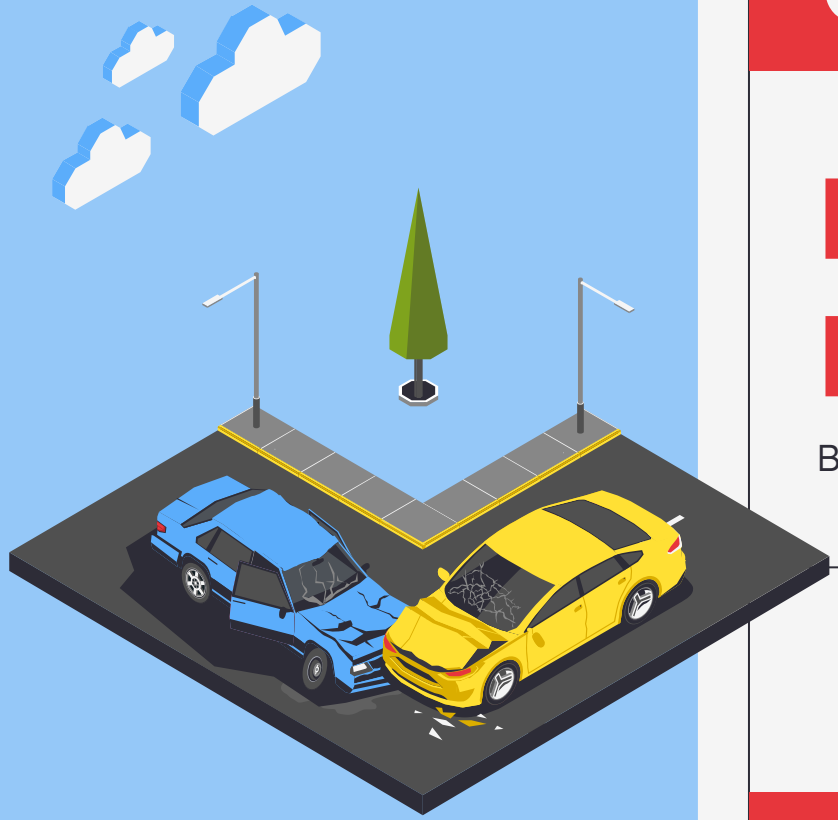
ROAD INJURIES IN CALIFORNIA

An Exploratory analysis using BI tools

shutterstock



01

BUSINESS PROBLEM

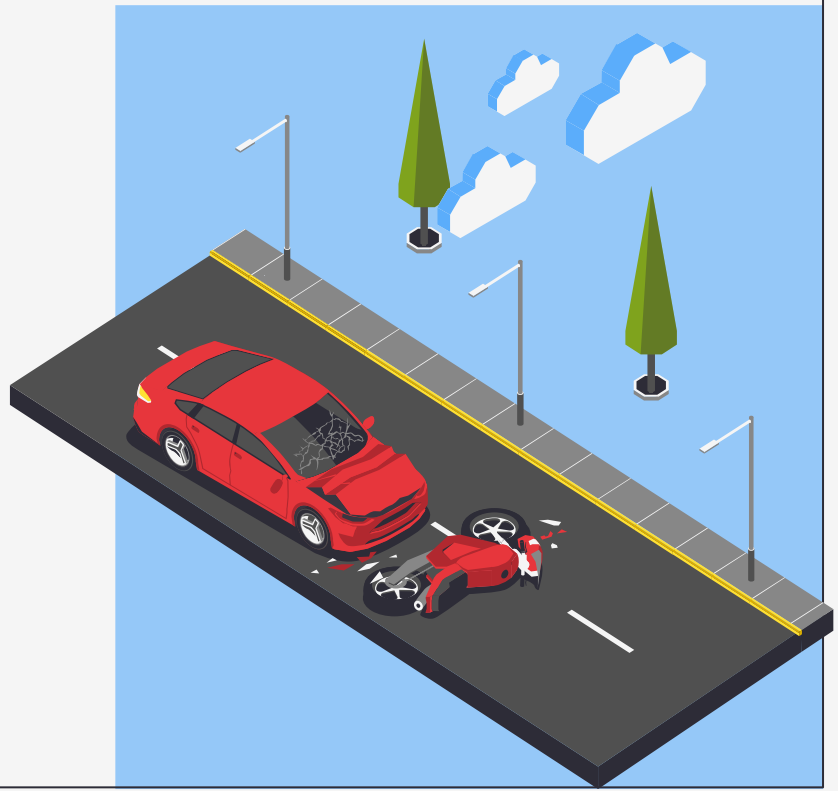
Basic Aim of the project

BUSINESS PROBLEM!

Every year, California is rocked by more road accidents compared to other states in the US. The main aim of the project is to :

- Reduce number of accidents
- Consequently have fewer casualities
- Fewer economic losses

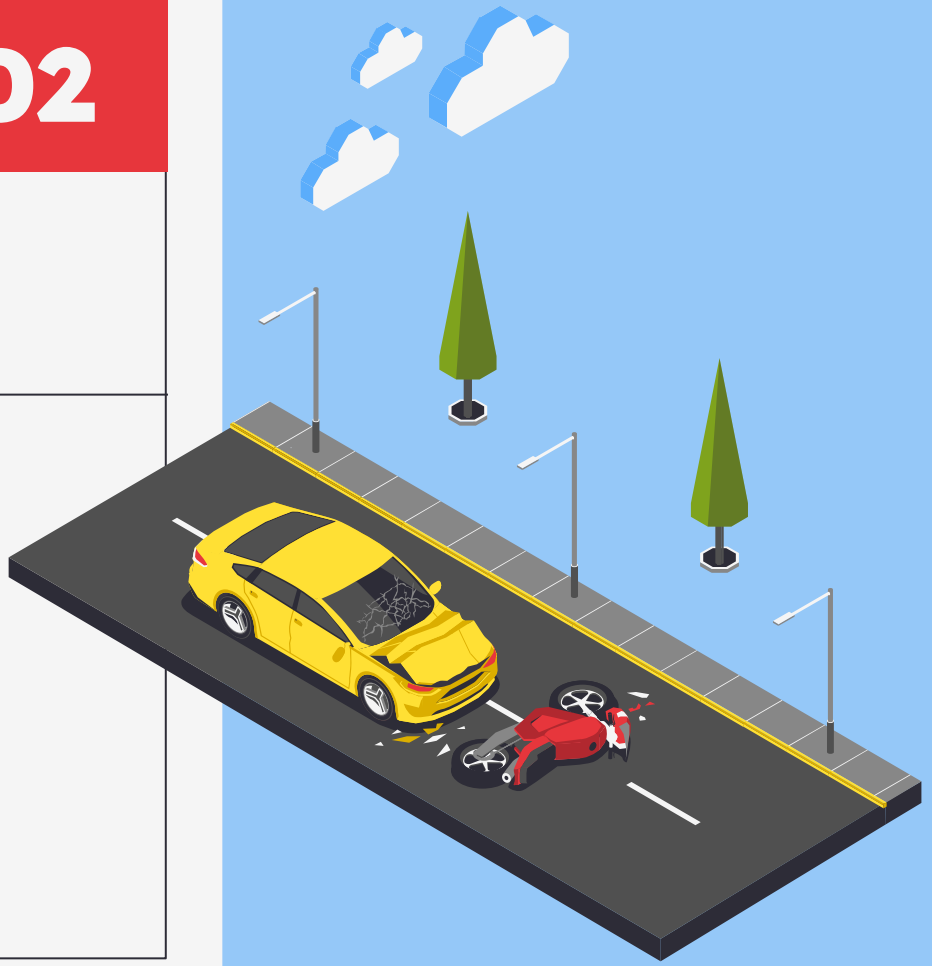
Exploratory analysis is performed on archivial data to identify trends, patterns and causes to plan further steps towards prevention.



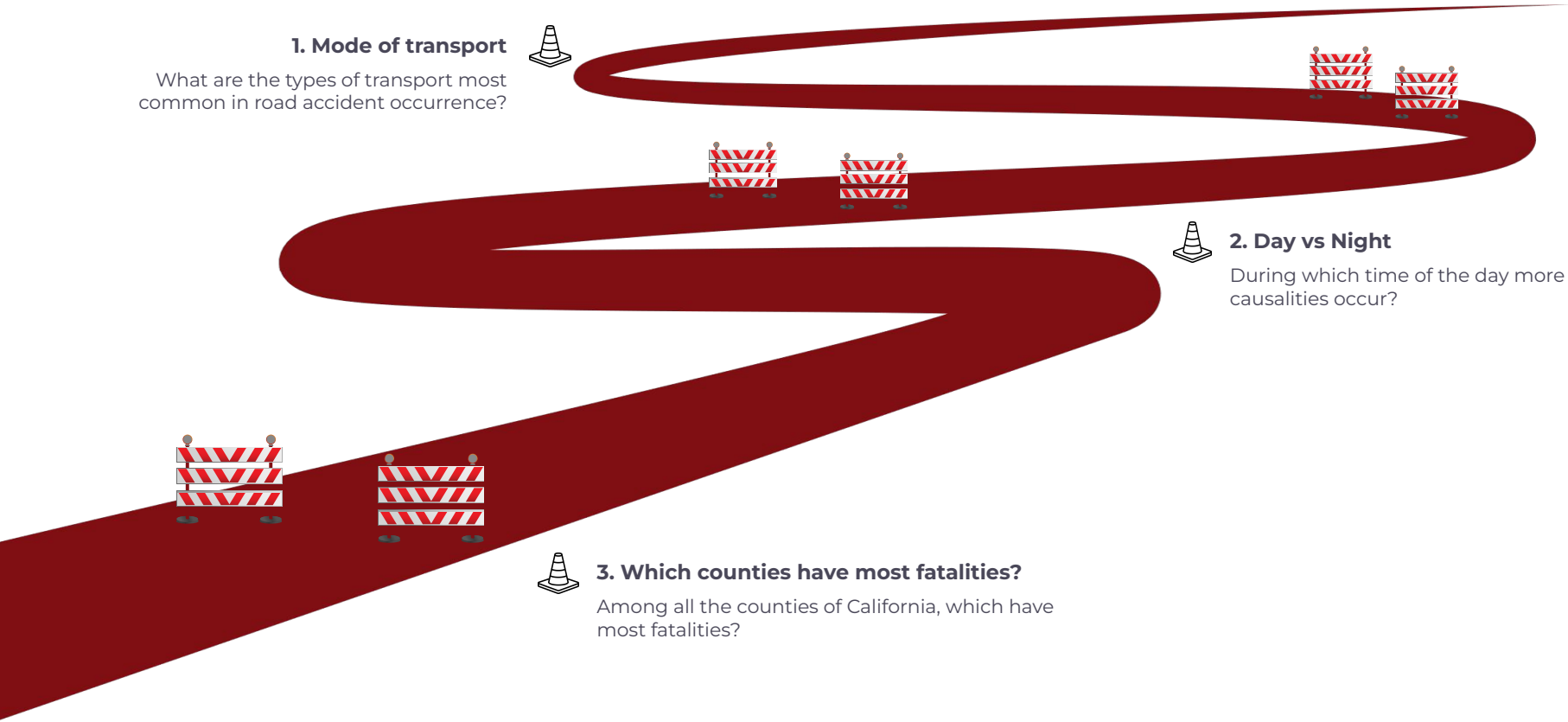
02

BUSINESS QUESTIONS

Focus on the decision-making aspect



BUSINESS QUESTIONS!



1. Mode of transport

What are the types of transport most common in road accident occurrence?

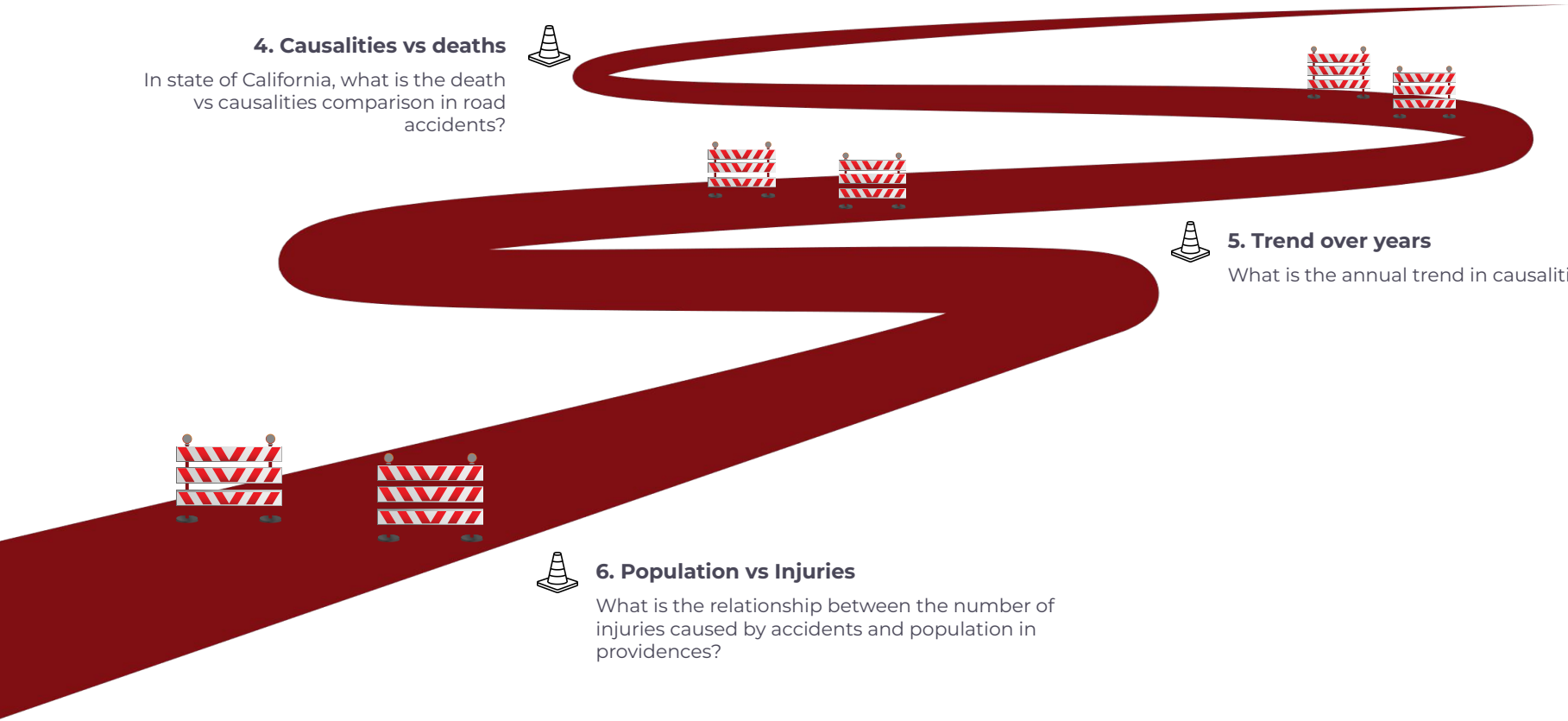
2. Day vs Night

During which time of the day more causalities occur?

3. Which counties have most fatalities?

Among all the counties of California, which have most fatalities?

BUSINESS QUESTIONS!



4. Causalities vs deaths

In state of California, what is the death vs causalities comparison in road accidents?

5. Trend over years

What is the annual trend in causalities?

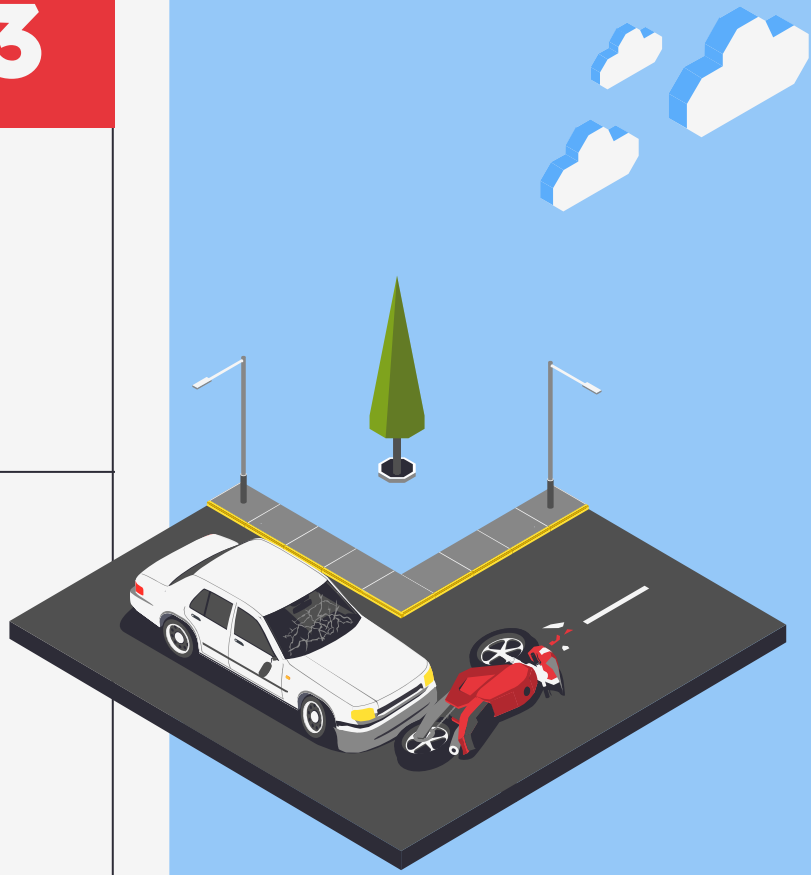
6. Population vs Injuries

What is the relationship between the number of injuries caused by accidents and population in providences?

03

DATA CLEANING & SQL CONNECTION

Birds eye view of road accidents dataset



Aerial view of dataset

- Road injuries in California dataset is a vast data giving information related to the injuries caused during road accidents in state of California.
- It includes details such as the year of the report, race/ethnicity, geography, county, region, mode of transportation, accident severity, number of injuries, and population data.

There are few key columns which helped in exploratory analysis such as,

i. **“reportyear” :** Year of the report
ii. **“county_name” :** Name of the county.
iii. **“region_name” :** Name of the region.
iv. **“mode” :** Mode of transportation involved in accident.
v. **“severity” :** Severity of accident.
vi. **“injuries” :** Number of injuries.
vii. **“totalpop” :** Total population.

Along with lot of information, dataset also contains noise. For noise mitigation cleaning is performed.

PowerBI Dashbord Link

Team7-Road Accidents Dashboard.pbix

Normalizing dataset!

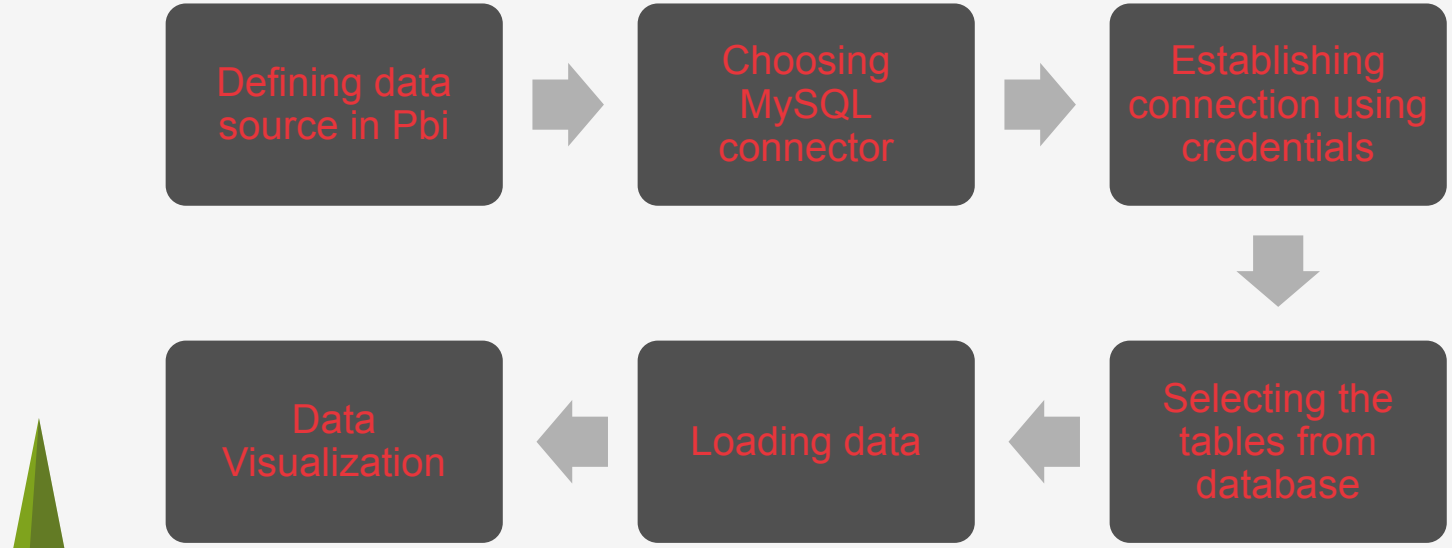
- Raw data is in denormalized form with many missing values, unwanted data.
- Efficient dashboard charts requires normalized data.
- Python libraries are used to clean the data.
- Missing values are replaced with mean and mode values.
- Unwanted rows are cleaned and dropped.

index	ind_id	ind_definit	reportyear	race_eth_c	race_eth_r	geotype	geotypeval	geoname	county_nar	county_fip	region_nar	region_coc	mode	severity	injuries	totalpop	poprate	poprate_s	poprate_rs	version
16065	753	Annual nur	2006-2010	9	Total	CD	6.07E+08	Idyllwild	Riverside	6065	Southern C	14	Vehicles	Severe Inju	13.6	12133	112.091	31.06263	27.71197	########
16066	753	Annual nur	2006-2010	9	Total	CD	6.03E+08	Imperial	Imperial	6025	Southern C	14	All modes	Killed	1.2	10230	11.73021	10.71956	91.38426	########
16067	753	Annual nur	2006-2010	9	Total	CD	6.03E+08	Imperial	Imperial	6025	Southern C	14	All modes	Severe Inju	6.8	10230	66.47116	25.64391	38.57899	########
16068	753	Annual nur	2006-2010	9	Total	CD	6.03E+08	Imperial	Imperial	6025	Southern C	14	Car/Pickup	Killed	1	10230	9.775171	9.783843	100.0887	########
16069	753	Annual nur	2006-2010	9	Total	CD	6.03E+08	Imperial	Imperial	6025	Southern C	14	Car/Pickup	Severe Inju	6.4	10230	62.56109	24.86951	39.75236	########
16070	753	Annual nur	2006-2010	9	Total	CD	6.03E+08	Imperial	Imperial	6025	Southern C	14	Motorcycle	Killed	0.2	10230	1.955034	4.372365	223.6465	########
16071	753	Annual nur	2006-2010	9	Total	CD	6.03E+08	Imperial	Imperial	6025	Southern C	14	Motorcycle	Severe Inju	0.2	10230	1.955034	4.372365	223.6465	########
16072	753	Annual nur	2006-2010	9	Total	CD	6.03E+08	Imperial	Imperial	6025	Southern C	14	Truck	Severe Inju	0.2	10230	1.955034	4.372365	223.6465	########
16073	753	Annual nur	2006-2010	9	Total	CD	6.03E+08	Imperial	Imperial	6025	Southern C	14	Vehicles	Killed	1.2	10230	11.73021	10.71956	91.38426	########
16074	753	Annual nur	2006-2010	9	Total	CD	6.03E+08	Imperial	Imperial	6025	Southern C	14	Vehicles	Severe Inju	6.8	10230	66.47116	25.64391	38.57899	########
16075	753	Annual nur	2006-2010	9	Total	CD	6.03E+08	Independe	Inyo	6027	Central/So	3	All modes	Killed	1.4	2316	60.44905	51.25891	84.79688	########
16076	753	Annual nur	2006-2010	9	Total	CD	6.03E+08	Independe	Inyo	6027	Central/So	3	All modes	Severe Inju	6.4	2316	276.3385	110.8856	40.12673	########
16077	753	Annual nur	2006-2010	9	Total	CD	6.03E+08	Independe	Inyo	6027	Central/So	3	Bicyclist	Severe Inju	0.4	2316	17.27116	27.33411	158.2645	########
16078	753	Annual nur	2006-2010	9	Total	CD	6.03E+08	Independe	Inyo	6027	Central/So	3	Car/Pickup	Killed	1.4	2316	60.44905	51.25891	84.79688	########
16079	753	Annual nur	2006-2010	9	Total	CD	6.03E+08	Independe	Inyo	6027	Central/So	3	Car/Pickup	Severe Inju	4.2	2316	181.3472	89.3695	49.2809	########
16080	753	Annual nur	2006-2010	9	Total	CD	6.03E+08	Independe	Inyo	6027	Central/So	3	Motorcycle	Severe Inju	1.4	2316	60.44905	51.25891	84.79688	########
24607	753	Annual nur	2006-2010	9	Total	CD	6.04E+08	Agoura Hill	Los Angele	6037	Southern C	14	All modes	Killed	8.8	62396	14.10347	4.764634	33.76342	########
24608	753	Annual nur	2006-2010	9	Total	CD	6.04E+08	Agoura Hill	Los Angele	6037	Southern C	14	All modes	Severe Inju	38.6	63257	61.02091	9.916118	16.25036	########
24609	753	Annual nur	2006-2010	9	Total	CD	6.04E+08	Agoura Hill	Los Angele	6037	Southern C	14	Bicyclist	Killed	0.4	8939	4.474773	7.079294	158.2045	########
24610	753	Annual nur	2006-2010	9	Total	CD	6.04E+08	Agoura Hill	Los Angele	6037	Southern C	14	Bicyclist	Severe Inju	3.6	45699	7.877634	4.156443	52.76258	########
24611	753	Annual nur	2006-2010	9	Total	CD	6.04E+08	Agoura Hill	Los Angele	6037	Southern C	14	Car/Pickup	Killed	6.2	57458	10.79049	4.340108	40.2216	########
24612	753	Annual nur	2006-2010	9	Total	CD	6.04E+08	Agoura Hill	Los Angele	6037	Southern C	14	Car/Pickup	Severe Inju	17.8	63257	28.13918	6.695446	23.79404	########
24613	753	Annual nur	2006-2010	9	Total	CD	6.04E+08	Agoura Hill	Los Angele	6037	Southern C	14	Motorcycle	Killed	1	20667	4.838632	4.841482	100.0589	########
24614	753	Annual nur	2006-2010	9	Total	CD	6.04E+08	Agoura Hill	Los Angele	6037	Southern C	14	Motorcycle	Severe Inju	14.6	45667	31.97057	8.41722	26.32802	########
24615	753	Annual nur	2006-2010	9	Total	CD	6.04E+08	Agoura Hill	Los Angele	6037	Southern C	14	Pedestrian	Killed	0.8	18395	4.349008	4.864663	111.8568	########
24616	753	Annual nur	2006-2010	9	Total	CD	6.04E+08	Agoura Hill	Los Angele	6037	Southern C	14	Pedestrian	Severe Inju	2.6	34197	7.603006	4.717656	62.04987	########
24617	753	Annual nur	2006-2010	9	Total	CD	6.04E+08	Agoura Hill	Los Angele	6037	Southern C	14	Truck	Killed	0.4	6818	5.866823	9.279423	158.1678	########
24618	753	Annual nur	2006-2010	9	Total	CD	6.04E+08	Agoura Hill	Los Angele	6037	Southern C	14	Vehicles	Killed	7.6	62396	12.18027	4.426403	36.34077	########
24619	753	Annual nur	2006-2010	9	Total	CD	6.04E+08	Agoura Hill	Los Angele	6037	Southern C	14	Vehicles	Severe Inju	32.4	63257	51.21963	9.080587	17.72872	########
24620	753	Annual nur	2006-2010	9	Total	CD	6E+08	Alameda	Alameda	6001	Bay Area	1	All modes	Killed	0.2	3426	5.837712	13.05987	223.7155	########
24621	753	Annual nur	2006-2010	9	Total	CD	6E+08	Alameda	Alameda	6001	Bay Area	1	All modes	Severe Inju	8.4	57939	14.49801	5.007492	34.53918	########

Table 1: Cleaned and normalized data

MySQL and Power Bi connector

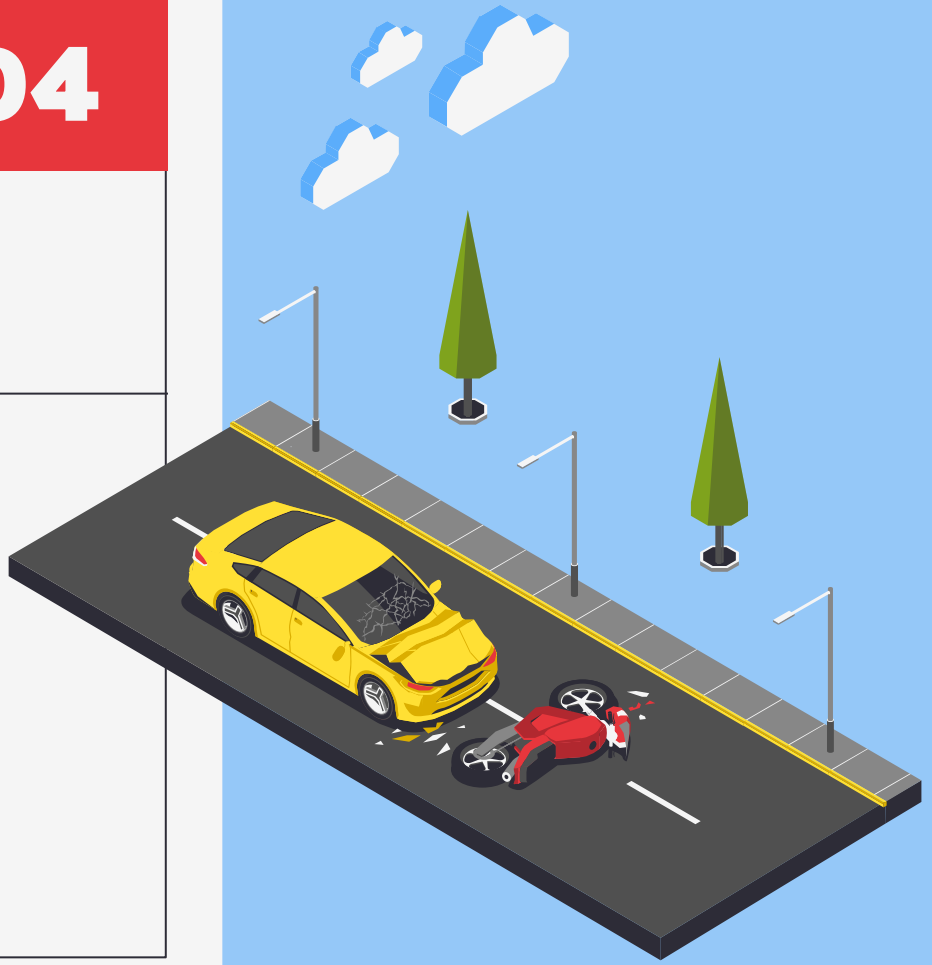
To connect database to Power Bi following steps are followed,



04

BUSINESS INSIGHTS

Transforming data into actionable Business insights.



INSIGHTS DRAWN FROM DASHBOARDS

- Injuries caused by road accidents are gradually decreasing every year.
- Among all providences in California, Los Angeles has more causalities on an average.
- Out of all causalities, 1/4th are dead.
- Car causes more causalities than other means of transport.
- During early nights and late nights, more accidents occur than mornings and afternoons.

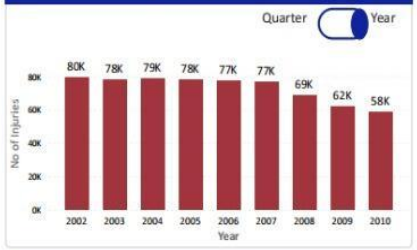
Road Traffic Injuries in California

Filter Menu

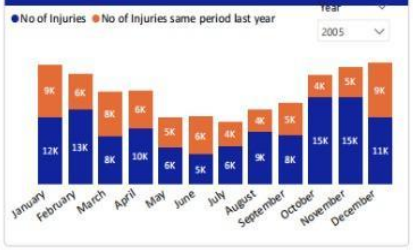
Region: All
County: All
Geo Name: All
Geo Type: All
Mode: All
Quarter: All
Year: All

Total Injuries by mode	=	Pedestrain	+	Bicyclist	+	Truck	+	Motorcycle	+	Car/Pickup	+	Bus	+	All Mode	+	Vehicles
988.36K		58.66K		21.76K		3.96K		47.45K		223.54K		886.67		356.26K		275.84K

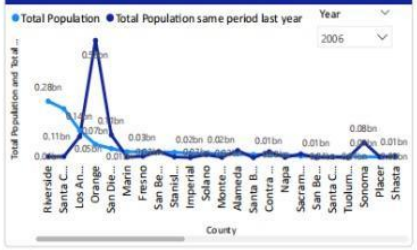
Injuries per Year



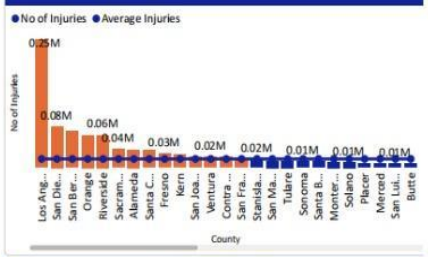
Injuries by Month



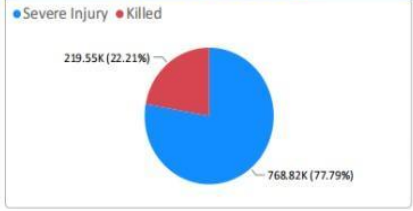
Population and Injuries of different county



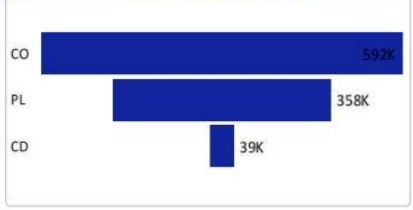
Injuries by County



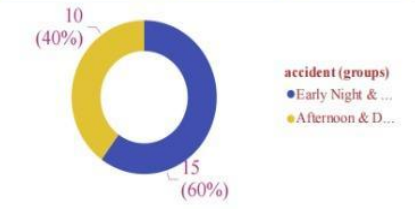
Injuries by Severity



Injuries by Geo Type



Severity based on Light



Injuries by Region

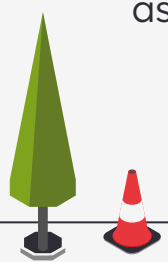
05

DATA DRIVEN DECISIONS

Recommendations for further analysis

Data Driven Decisions

- A thorough strategy to improving road safety begins with gathering extensive data on collisions, taking note of variables including location, time, weather, state of the road, and the cars involved.
- Statistical and visual aids are used to evaluate this data in order to find patterns and connections that help with the geospatial analysis process of identifying "accident hotspots".
- In order to identify contributing variables, root cause analysis explores underlying factors such as driving behavior and road design. By using previous data, predictive modeling forecasts the risk of accidents, allowing for preemptive safety measures.
- Intervention strategies are then put into practice, such as enhanced enforcement and improved roads, in accordance with the insights acquired.
- By directing continual changes and improving safety precautions, ongoing monitoring and assessment guarantee the effectiveness of these actions.

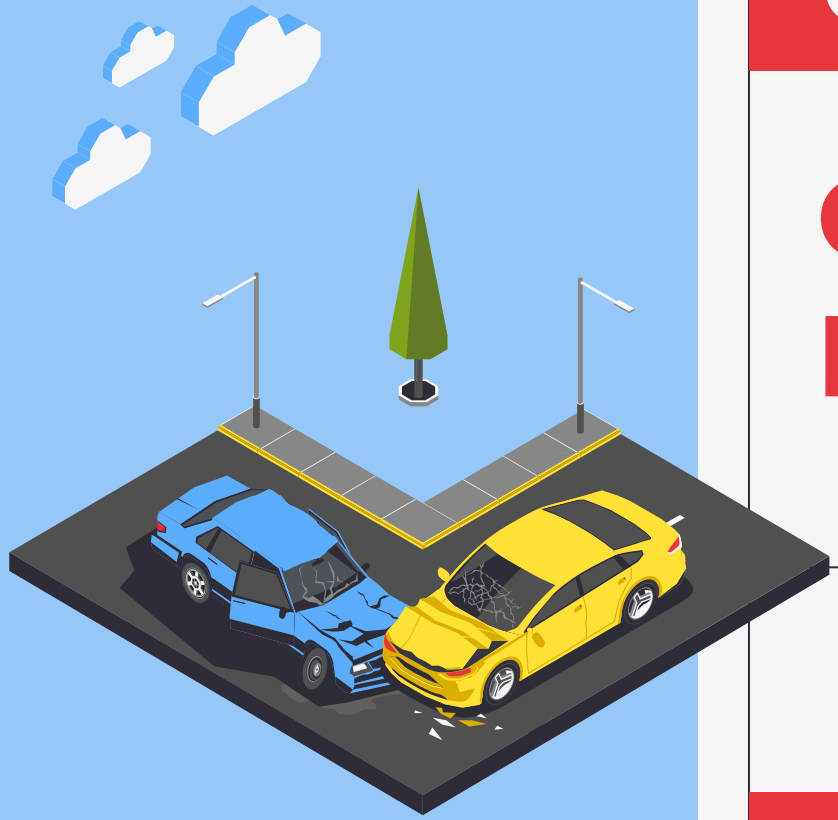


To reduce road accidents effectively:

- **Night Driving:** Limit non-essential nighttime driving when accidents are most likely to occur. When driving, employ caution, make sure you're well-rested, and utilize your headlights appropriately
- **Los Angeles Focus:** Give priority to safety measures such as increased law enforcement, road improvements, and awareness programs in high-risk locations like Los Angeles.

Recommendations from Data Analysis:

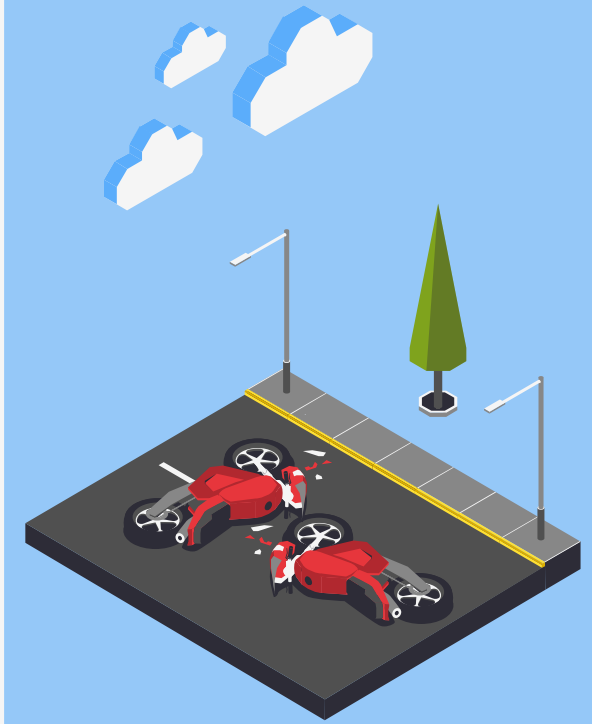
- **Factors Associated with Night Driving:** Examine certain causes of accidents at night, such as poor vision, tired drivers, and alcohol consumption.
- **Accident Severity:** To more effectively prioritize safety precautions, analyze the severity of incidents according to time and location.
- **Weather Conditions:** Use meteorological information to investigate relationships with accident frequencies.
- **Areas at High Risk:** For focused improvements, identify particular road sections or intersections that have a high frequency of accidents.
- **Demographic Analysis:** Look into the likelihood of accidents in various age groups or demographics for targeted education efforts, if the facts allow.



06

CONCLUSIO N

- In conclusion, a thorough strategy is required to lower traffic-related injuries in California.
- We can successfully adopt specialized safety measures by concentrating on high-accident regions like Los Angeles and risky hours like night driving.
- It is crucial to implement tactics like more law enforcement, improved roads, and awareness programs.
- To further improve these measurements, a detailed analysis of factors such as weather, road locations, and driver demographics would be necessary.
- By adopting these methods, we aim to make California's roads safer and reduce injuries from accidents.



THANK YOU