Data Science Capstone Project Ideas

The following are the three ideas I have for the Capstone project.

1. Dallas Crime data
2. Chicago Taxi Trips
3. Traffic collision data
4. Crime Data from 2010 to Present
5. **Dallas Crime data**

The dataset from [https://www.dallasopendata.com](https://www.dallasopendata.com/Public-Safety/Police-Incidents/qv6i-rri7) contains all the crime incidents happened in Dallas from June 2014 till August 2019. Each row in the dataset is an incident reported to the Dallas Police Department and it has the following information

* Incident Number
* Type of Incident
* Date, Month and Year of Occurrence
* Offense status
* Modus Operandi (MO codes)
* Weapon Used
* Reporting Area
* Type Location
* Victim Type, Gender, Age, Race, Ethnicity

Problem Statement: The objective is to analyze the trend in crimes from June 2014 till August 2019 as follows

* Analyze the trend of different categories of crime
* Find the most used weapon in the crime scene
* Identify the areas with less crimes
* Predict the crime rate (for each category of crime) for the next year

1. **Chicago Taxi trips**

This dataset from [https://catalog.data.gov](https://catalog.data.gov/dataset/taxi-trips) contains the various details about the millions of taxi rides in the year 2015, traveling through the streets of Chicago. Each row in the dataset is a distinct taxi ride and has the following information

* Which taxi
* Trip start and end time
* Length of the trip in time and distance
* Pick up point and destination of the trip
* Cost of the trip
* Type of payment
* Taxi company

Problem Statement: Perform trend analysis on taxi trips based on the metrics like locality, cost, duration like

* Total number of trips taken per day, Total number of trips taken per month
* What are the peak hours for taking a taxi?
* Where does the most pickup and drop off happen?
* Estimate the cost of the trip based on given pick up and drop off location and time
* Predict the trip time based on the historical data
* Determine the average revenue per hour

1. **Traffic collision data**

This dataset is from [https://catalog.data.gov](https://catalog.data.gov/dataset/traffic-collision-data-from-2010-to-present), containing the traffic collision incidents happened in Los Angeles from 2010 till Present. It contains the following data

* The DR numbers
* Date and time when the incident happened and reported
* Area it occurred and reported
* Crime codes (Modus Operandi code)
* Victim details
* Location details etc.

Problem Statement:

* Review the MO codes to identify various reasons for collisions
* To identify the locations where most collisions happened based on the type of victim and the time of the day
* Collisions happened late nights, the reason why and where it happened the most
* Collisions causing major injuries and the type of victim

1. **Crime data from 2010 to Present**

The dataset from [https://data.lacity.org](https://data.lacity.org/A-Safe-City/Crime-Data-from-2010-to-Present/y8tr-7khq) contains original crime reports in the city of Los Angeles dating back to 2010. Each row is an incident which contains the following data.

* Date/Time of incident
* Area Name
* Crime code/description
* Victim details
* Premise details
* Weapon details, etc.

Problem Statement:

* Predict the date when next crime will happen in different areas of city and its probability