Tools: ER Studio/Navicat (Dimension Modeling), Databricks / Alteryx for Data Profiling, Databricks (bronze/silver zone pipelines in PySpark/SQL), Snowflake (If having challenges either with cost/credits/connection then use Databricks) as gold zone, Power BI or Tableau (visualization).

Objective: Implement an end-to-end solution and build analytics that answer the business questions.
Refer to the document attached for details: End-to-End-Solution-LA-Crime-Analysis.pdf Download End-to-End-Solution-LA-Crime-Analysis.pdf

Note:

1. Upload everything to GitHub
2. Follow best practices as discussed in class and implemented in previous projects

Deliverables

1. Alteryx file.
2. Revised ER model (ER Studio/Navicat) including professor-suggested improvements.
3. Databricks notebooks/scripts
4. Snowflake DDL & load scripts
5. Tableau or Power BI workbook + packaged extracts.
6. GitHub link
7. Document with all screenshots and insights

Business requirements:

Crime Rates Over Time:
• What is the overall trend in crime rates over the years?
• How have crime rates changed on a monthly basis?
• How have crime rates changed on a quarterly basis?

Day Time and Week Factors:
• Is there a correlation between the day of the week and the number of reported crimes?
• Do certain types of crimes tend to occur at specific times of the day?

Crime by Location:
• Where are the high-crime areas in Los Angeles?
We want to know (your inferences on):
   o By seeing the geo map, what are the crime hotspots?
   o Can you identify areas where Los Angeles performs better or worse in terms of crime rates?

Types of Crime:
• What are the most commonly used weapons in reported crimes?

Demographic Analysis:
• Show in the visualization the age patterns in crime.
• Show in the visualization the gender-related patterns in crime.

Arrests Ratio:
• What percentage of reported crimes result in Juvenile, Adult arrests?