Step 1: Data Profiling
- Check nulls, distinct counts, data types. Identify invalid/missing dates, times, lat/lon, or age.
- Document key observations and issues.

Step 2: Data Modeling
- Revise quiz dimension model with professor suggestions. The data model needs to be improved version of your submission and not from scratch.

Step 3: Bronze and Silver Layer (Databricks)
- Ingest raw data to DataBricks volume
- Perform cleaning and transformations – store in Silver layer.

Step 4: Gold Layer (Snowflake)
- Create dimension and fact tables in Snowflake
- Implement SCD as needed.

Step 5: Pipeline
- Use Databricks (PySpark or SQL) to load from silver → gold.

Step 6: Visualization (Power BI or Tableau)
- Build dashboard(s)/report(s) to answer below business questions.

Business Questions:

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?