* Capstone Project 1 (1 week) :
  + Capstone Project: Real-time Crime Rate Prediction and Analysis
* Project Overview
  + Participants will build a real-time analytics pipeline to predict and analyze crime rates in major cities using historical crime data. The project will leverage Databricks for data processing, Spark for real-time streaming, Databricks SQL for analytics, and MLflow for machine learning model management.
* Data Source
  + Dataset: Use the "US Crime Rates" dataset available from the City of Chicago's
  + Data Portal. This dataset includes details such as the date, type of crime, location,

and arrest records.

* + Additional Data: Weather data from the NOAA which can be used to analyze the

impact of weather conditions on crime rates.

* Key Phases of the Project
  + Data Ingestion and Storage
  + Ingest real-time and historical crime data using Databricks Auto-loader.
  + Store data in Delta Lake to leverage ACID transactions and efficient data querying.

2. Data Processing and Transformation

Cleanse and preprocess data using PySpark.

Develop a Bronze/Silver/Gold layer architecture for processed data.

Implement streaming data pipelines to ingest real-time crime data.

3. Analytics and Business Intelligence

Use Databricks SQL to create queries and dashboards for analyzing crime trends and patterns.

Integrate external tools for enhanced visualization (e.g., Tableau or PowerBI

embedded in Databricks notebooks).

4. Machine Learning Model Development

Develop a predictive model to forecast crime rates based on historical data

and other factors like weather conditions.

● Use MLflow to track experiments, manage machine learning models, and

serve the final model.

5. Model Deployment and Serving

● Deploy the model using Databricks’ MLflow Model Serving.

● Set up a real-time prediction endpoint to provide on-demand crime rate

predictions.

6. Project Presentation

● Document the entire pipeline from data ingestion to model deployment.

● Prepare a final presentation outlining methodologies, insights, challenges,

and performance metrics.

● Demonstrate a live dashboard and model prediction results.

Expected Outcomes

● Participants will gain hands-on experience with building end-to-end solutions on the

Databricks platform.

● The project helps understand the integration of various Databricks components and

external tools.

● Participants will learn to handle real-time data streams, perform advanced analytics,

and deploy machine learning models in a production-like environment.