Project Journal : X23350458 Shahida Shaik

Week 1: Project Selection and Dataset Research

- **Tasks:** Explored potential topics and objectives for the project. Conducted research on public datasets related to crime, drug overdose rates, and life expectancy. Finalized the datasets to be used: Los Angeles crime data, CDC drug overdose data, and life expectancy data from national health databases.
- **Time Spent:** 10 hours
- **Challenges:** Finding datasets that were recent, detailed, and publicly available. Addressed this by using government databases and filtering data for the required years.

Week 2: Data Cleaning and Preprocessing

- **Tasks:** Cleaned and preprocessed datasets to ensure consistency. Removed null values, handled missing data, and standardized formats across datasets.

  Time Spent: 12 hours

- **Challenges:** Handling missing or incomplete data in life expectancy datasets since there were certain columns with almost half missing values. Used imputation techniques to address gaps. Some dataset I ignored the blanks since I was not sure if they had meaning.

Week 3: Initial Data Analysis and Coding

- **Tasks:** Conducted exploratory data analysis (EDA) on each dataset to understand key trends and variables. Used Python libraries such as Pandas, Matplotlib, for visualizations. Started writing code for data wrangling and integration.
- **Time Spent:** 15 hours
- **Challenges:** Interpreting trends in life expectancy data due to its large volume. Conducted te analysis in chunks.

Week 4: Database Integration

- **Tasks:** Set up SQLite database to store cleaned and processed datasets for efficient querying. Wrote SQL queries to extract relevant data subsets.

- **Time Spent:** 10 hours
- **Challenges:** Ensuring compatibility between Python and the database. Used the SQLite3 library in Python to resolve integration issues.

Week 5: Advanced Analysis and Visualization

- **Tasks:** Developed advanced visualizations to highlight key insights, such as the correlation between crime rates and drug overdose deaths, and the variation in life expectancy by geography and demographics.
- **Time Spent:** 14 hours
- **Challenges:** Designing clear, meaningful visualizations. Iterated through several plot types to find the most effective representation.

Week 6: Report Writing and Results Evaluation

- **Tasks:** Contributed to writing the results, evaluation, and discussion sections of the report. Interpreted findings in relation to project objectives and research questions. Reviewed the statistical validity of results.
- **Time Spent:** 12 hours
- **Challenges:** Balancing between detail and clarity in reporting results. Peer-reviewed with the team to refine.

Week 7: Finalization and Presentation Preparation

- **Tasks:** Participated in preparing the final report . Tested the code and visualizations to ensure reproducibility. Practiced delivering the presentation.
- **Time Spent:** 10 hours
- **Challenges:** Time management during final reviews. Addressed by dividing tasks among team members and setting deadlines.