Project Report: SQL-Based Analysis of College Courses

Step-by-Step Task Completion

Step 1: Understanding the Dataset

- **Actions Taken:** Reviewed the columns and structure of `DA Batch 10 SQL & PowerBi Data set.csv`, including college details, course types, durations, and categories.
- **Analysis Conducted:** Mapped the dataset to SQL schema and identified key fields like `District`, `CourseName`, `CourseType`, `IsProfessional`, etc.
- **Results:** Dataset was clean and well-structured with 12 relevant columns for SQL analysis.

Step 2: Preparing the Data for SQL Queries

- **Actions Taken:** Loaded the CSV into Python using pandas, renamed columns for SQL compatibility, and created an SQLite in-memory database.
- **Analysis Conducted:** Verified data integrity and checked for missing values.
- **Results:** Table `college_courses` created successfully in SQLite for query execution.

Step 3: Writing and Testing SQL Queries

- **Actions Taken:** Formulated 20 SQL queries as per the internship task. Queries ranged from basic aggregations to CTEs and window functions.
- **Analysis Conducted:** Each query was tested individually for correctness and efficiency.
- **Results:** All queries executed successfully, producing meaningful insights.

Summary of Analyses Conducted

Q1: Top Districts Offering Professional Courses

- Pune, Mumbai Suburban, and Nagpur were among the top districts.

Q2: Average Course Duration by Course Type

- PG courses had the highest average duration.

Q3: Unique Colleges per Course Category

- Engineering and Science categories had the widest spread.

Q4: Colleges Offering Both UG & PG

- Identified institutions with diverse offerings.

Q5 to Q20:

- Included analyses such as aided vs unaided comparison, duration classifications, median filtering, specialization extraction, and professional course percentages.

Each query provided a new layer of understanding into the operational and academic structure of institutions.

Final Conclusion

This project enhanced my hands-on experience in structured query formulation and real-world data exploration. I gained practical insight into:

- Identifying institutional strengths and course trends.
- Classifying data for targeted decision-making.
- Leveraging SQL for business-like insights in the education sector.

The work lays the foundation for potential visualization and dashboarding in Power BI or Tableau, making it a complete data pipeline from raw data to decision-ready insights.