Voiceover Script for SQL Project Presentation

Slide 1: Title Slide "Hello everyone, I'm [Your Name], and in this video, I will walk you through my SQL project titled 'Higher Education Course Analysis', done as part of the Data Analyst Internship at Soulvibe.Tech."

Slide 2: Introduction "The main goal of this task was to use SQL to analyze a real-world dataset about higher education courses across Maharashtra. I wrote 20 queries that explore trends in course types, durations, institutions, and funding. Now let's look at each analysis step by step."

Query 1: Top 5 Districts with Most Professional Colleges "I filtered for rows where the course type is labeled 'Professional Course', grouped them by district, and counted distinct college names. After sorting in descending order of count, I picked the top 5 using the LIMIT clause. This showed which districts lead in professional education infrastructure."

Query 2: Average Duration by Course Type "Using the AVG function, I calculated the average course duration for each Course Type. Grouping by Course Type and sorting in descending order revealed which types of programs generally require more time. Postgraduate and Diploma courses stood out for their length."

Query 3: Unique College Count per Course Category "I grouped the data by Course Category and counted how many unique colleges were associated with each using COUNT(DISTINCT CollegeName). This showed how widely offered each course category is."

Query 4: Colleges Offering Both UG & PG "I used a self-join on the CollegeName column to find colleges offering both 'Under Graduate' and 'Post Graduate' programs. This identified versatile institutions with broader course offerings."

Query 5: Universities with >10 Unaided, Non-Professional Courses "I filtered rows for 'Unaided' and 'Non-Professional Course', grouped them by University, and used HAVING COUNT(*) > 10. This highlighted universities that support broader non-professional education without aid."

Query 6: Engineering Colleges with Above-Average Duration "I first computed the average duration for Engineering courses and then filtered courses exceeding this value using a subquery. These colleges offer longer-than-usual programs."

Query 7: Rank Courses by Duration within College "Using RANK() OVER PARTITION BY CollegeName ORDER BY Duration DESC, I assigned a rank to each course within a college. This ranked them from longest to shortest."

Query 8: Colleges with Duration Gap > 24 Months "By calculating MAX and MIN duration per college and checking the difference, I filtered colleges where that gap was greater than 24 months. These have a broad duration range in their course offerings."

Query 9: Cumulative Professional Courses by University "I filtered for 'Professional Course', grouped by University, counted the courses, and sorted alphabetically. This gave a complete view of professional course presence by university."

Query 10: Colleges with More than One Course Category "Using GROUP BY and COUNT(DISTINCT CourseCategory), I selected colleges with more than one category, highlighting multi-disciplinary institutions."

Query 11: Talukas Above District Avg Course Duration "With a CTE, I calculated district-wise average durations and compared each taluka's average to its district's. Talukas above the average were selected to find high-intensity education zones."

Query 12: Classify Duration into Short, Medium, Long "Using a CASE statement, I classified each course as Short (<12), Medium (12-36), or Long (>36) months, then counted how many of each existed per Course Category."

Query 13: Extract Specialization from Course Name "Using string functions, I split the Course Name to extract only the specialization part, such as 'Electrical' from 'Bachelor of Engineering - Electrical'."

Query 14: Courses with 'Engineering' in Name "A basic LIKE '%Engineering%' condition in the WHERE clause was used to count course names that include the word 'Engineering'."

Query 15: Unique Combinations of Course Name, Type, Category "Using SELECT DISTINCT, I extracted all unique combinations of Course Name, Course Type, and Course Category to evaluate diversity in course structuring."

Query 16: Courses Not Offered by Government Colleges "I used a NOT IN or EXCEPT clause to remove courses offered by government colleges. The remaining list shows privately managed courses."

Query 17: University with 2nd Highest Aided Courses "Grouped by University, I filtered for aided courses and used DENSE_RANK() to get the second-highest based on count."

Query 18: Courses Above Median Duration "I found the median course duration and filtered all courses with a duration above that value. These are the most time-intensive programs."

Query 19: % of Unaided Courses That Are Professional "Grouped by University, I used conditional aggregation to find the percentage of unaided courses marked as 'Professional'."

Query 20: Top 3 Course Categories by Avg Duration "I grouped by Course Category, calculated average duration, sorted descending, and picked the top 3. This identified the most time-consuming academic fields."

Conclusion "This project allowed me to use SQL to clean, filter, and analyze complex data. It improved my ability to use joins, aggregations, subqueries, and CTEs, and gave insights that are valuable for policy and planning in education."

Thank You "Thank you for your time. I hope this presentation helped you understand how SQL can be used to analyze real-world educational data."