Cohort 4 Analysis - Insight Report

By Faith Maina

Introduction

I conducted a comprehensive analysis of the Cohort 4 dataset, examining participant demographics, learning patterns, experience levels, and performance outcomes. Through systematic data cleaning in Excel, SQL queries, and Power BI visualizations, I uncovered key insights about our cohort's composition and student success factors.

Data Cleaning and Methodology

I started by loading the raw dataset into Excel and found no duplicate records in the ID field, indicating good data collection practices. My cleaning process included:

- Standardizing text data using `=TRIM(PROPER(H2))` for proper formatting
- Converting experience ranges to numeric months using complex IF formulas (mapping "Less than six months" to 3 months, "More than 3 years" to 42 months)
- Transforming weekly learning hours into numeric values (3, 10, or 16 hours)
- Converting Yes/No responses to boolean values (1/0) for aptitude and graduation status

Key SQL Analysis Findings

Through SQLite queries, I discovered important patterns:

Data Quality: I confirmed the total_score field contained clean numeric data with no null values, validating my performance analysis reliability.

Track Performance: My analysis revealed significant performance variations across learning paths, with graduation rates calculated for each track to identify the most successful programs.

Geographic Distribution: I identified top countries represented, showing global diversity in our participant base.

Demographics: Age and gender analysis revealed participation patterns across different demographic groups.

completion rates and success factors.

Power BI Dashboard

I created an interactive dashboard featuring:

- Dynamic demographic breakdowns (age, gender, country)
- Track distribution visualizations showing program popularity
- Correlation charts between learning hours, experience, and performance scores
- Graduation trend analysis tracking success patterns

Conclusions and Recommendations

My analysis revealed that experience levels, learning time investment, and track selection significantly influence participant success. The systematic data cleaning approach, particularly converting categorical variables to numeric formats, was crucial for meaningful statistical analysis.

The Power BI dashboard provides stakeholders with actionable insights for program optimization. I recommend using these findings to guide student placement, adjust program structures, and identify at-risk participants early.