Data Analysis and Power BI Dashboarding of Student Response Dataset

By

Sampik Kumar Gupta

PROBLEM STATEMENT AND OBJECTIVE OF THE TASK

Problem Statement:

The objective of this project is to analyze and visualize student performance data from a recent test to gain insights into individual and overall performance. The data includes student responses to a set of questions, which need to be compared against a coding sheet containing correct answers. This analysis aims to assess student scores, calculate test performance metrics, and visualize the data using business intelligence (BI) tools to provide a comprehensive overview of the results.

Objectives:

Data Analysis (Part 1):

- Calculate the "Student Score" by matching student responses with the correct answers provided in the coding sheet. Responses that match the correct answer will be scored as "1", while non-matching responses will be scored as "0".
- Determination of the Test Score for each student based on the number of correct answers.
- Computation of the percentage of correct answers for each student, representing the proportion of correct responses out of the total questions.
- Classification of the students into performance levels based on their percentage correct.

Data Visualization (Part 2) using a Business Intelligence (BI) Tool:

- Development of interactive slicers to filter data by Grade, Subject, District, and Block, ensuring synchronization across these filters.
- Construction of a map chart to visualize the distribution of participating students across different locations, with drill-down capabilities for Districts and Blocks.
- Creation of Key Performance Indicators (KPIs) in the form of cards to display total student count and average percentage correct.
- Presentation of a table showcasing the top 5 questions based on performance (highest correct responses) and highlighting the bottom 5 questions based on performance (lowest correct responses) along with their descriptions.

The study aims to provide educational stakeholders with actionable insights into student performance patterns, areas of improvement, and regional variations. The data visualization component will enable users to interactively explore the data, uncover trends, and make informed decisions to enhance educational outcomes.

METHODOLOGY

Data Analysis (Part 1):

Approach: To accomplish the task, I followed a structured approach using Excel functions and formulas. The steps involved were as follows:

Data Preparation:

To begin, I established a dedicated sheet within the provided Excel workbook to facilitate reference and analysis. This sheet was renamed "Final Result Sheet" for clarity. Subsequently, I extracted the necessary columns into this sheet, leaving the required columns blank for subsequent calculations, including all question columns (from 1 to 40), Test Score, Correct Answer Percent, and Levels.

Calculating Student Scores:

For assessing student scores, I employed the XLOOKUP function to verify if the student's response in columns [1….40] aligned with the correct answer from the "Coding Sheet." The formula utilized was:

```
=IFERROR(IF(XLOOKUP($K2&L$1,Codeframe[[Test Code]:[TestCode]]&Codeframe[[Question Number]:[Question Number]],Codeframe[[Answer]:[Answer]],,0)='Data Sheet'!I2,1,0),"NA")
```

This formula compared each student's response from the 'Data Sheet' to the 'Coding Sheet,' assigning a score of 1 for a correct answer and 0 for an incorrect answer. It also managed instances where XLOOKUP could not find a match, returning "NA." Here, "Codeframe" represents the table

name of the Coding Sheet, facilitating comparison of the lookup_array and return_array from the Coding Sheet.

Calculating Test Scores and Correct Answer Percentages:

Following this, the "Test Score" column was computed by summing the individual question scores using the formula:

```
=SUM(L2:AY2)
```

The "Correct Answer Percent" column was determined by dividing the "Test Score" by the total number of questions:

```
=H2/36
```

Calculating Student Levels:

To classify student performance levels based on their Correct Answer Percentages, the following formula was utilized:

```
=IF(I2 <= 0.3, "Beginner", IF(AND(0.3 < I2, I2 <= 0.5), "Basic", IF(AND(0.5 < I2, I2 <= 0.8), "Intermediate", "Advanced")))
```

This formula categorized students into levels:

- "Beginner" (Correct Answer Percentage ≤ 30%)
- "Basic" (31% ≤ Correct Answer Percentage ≤ 50%)
- "Intermediate" (51% ≤ Correct Answer Percentage ≤ 80%)
- "Advanced" (Correct Answer Percentage > 80%)

After following the outlined steps, the "Final Result Sheet" was prepared with all the necessary data for subsequent visualization and dashboard creation.

Data Visualization (Part 2) using a Business Intelligence (BI) Tool:

To begin the data visualization process, the coding sheet and result sheet were imported into Power BI, and necessary transformations were conducted using the Power Query Editor to prepare the data for visualization.

Step 1: Result Sheet Transformation

A copy of the result sheet was created to facilitate data manipulation. The following transformations were applied:

- The result sheet was utilized to calculate the "Question ID" by merging the "Test Code" and "Question Number."
- Columns 37, 38, 39, and 40 were deleted as they contained no valuable information.
- The "QuestionNumbers" columns [1, 2, 3, ... 36] were unpivoted, resulting in the generation of two additional columns: "Attribute" (containing the Question Number) and "Value" (indicating the student's marks as 1 or 0 for each UID).
- This transformation expanded each UID into 36 rows (corresponding to the 36 questions), where "Attribute" denoted the Question Number and "Value" represented the student's score (1 for correct, 0 for incorrect).

Step 2: Creating Question ID

A duplicate of the "Test Code" and "Attribute" columns after unpivoting was created. These columns were merged to generate an additional column named "Question ID," facilitating easier data referencing and analysis.

Step 3: Filtering for Subject and Loading Data

Since the provided dataset lacked data for Maths Subject in the coding sheet, the Result Sheet and its duplicate (Result Sheet 2) were filtered to include data only for the subject "Gujarati." The transformed and filtered data was then loaded into Power BI for visualization.

By conducting these data transformations and filtering, the dataset was prepared for insightful visualization, focusing specifically on the subject of "Gujarati." This refined dataset will serve as

the basis for creating interactive visualizations and insightful dashboards to analyze student performance effectively.

In the dashboard, the following features were utilized to present the analyzed data effectively:

A). Slicers for Grade, Subject, District, and Block:

Interactive slicers were implemented to filter data by Grade, Subject, District, and Block. These slicers were synchronized to ensure that filtering one criterion updates the others accordingly, providing a cohesive user experience.

B). Map Chart with Drill-Down Capability:

A map chart was created to visualize the distribution of students across different locations. The map chart allowed for drill-down functionality, enabling users to explore student participation at various levels such as Districts and Blocks.

C). Key Performance Indicators (KPIs):

KPIs were displayed using cards to present important metrics:

- Total Students: Showcased the total count of students participating in the test.
- Average Percentage Correct: Displayed the average percentage of correct answers across all students.

D). Top 5 Questions Table:

A table was generated to highlight the top 5 questions based on student performance. This table included the question descriptions along with corresponding performance metrics.

E). Bottom 5 Questions Table:

Similarly, a table was created to showcase the bottom 5 questions based on student performance. The table included question descriptions along with the respective performance metrics for these questions.

These components were integrated into the dashboard to provide a comprehensive and interactive overview of student performance data. Users could filter and explore the data dynamically using slicers, visualize geographic distribution through the map chart, and gain insights into specific

questions through the top and bottom question tables. The KPIs offered quick summaries of key metrics, enhancing the dashboard's utility for educational stakeholders seeking actionable insights from the test results.

I am including a screenshot of the dashboard to provide insights into student performance across different regions and to facilitate effective analysis of various metrics. To enhance insights using filters and maps, I have attached the Power BI(.pbix) file and the output result sheet. These resources will enable a comprehensive study of the data and its insights.

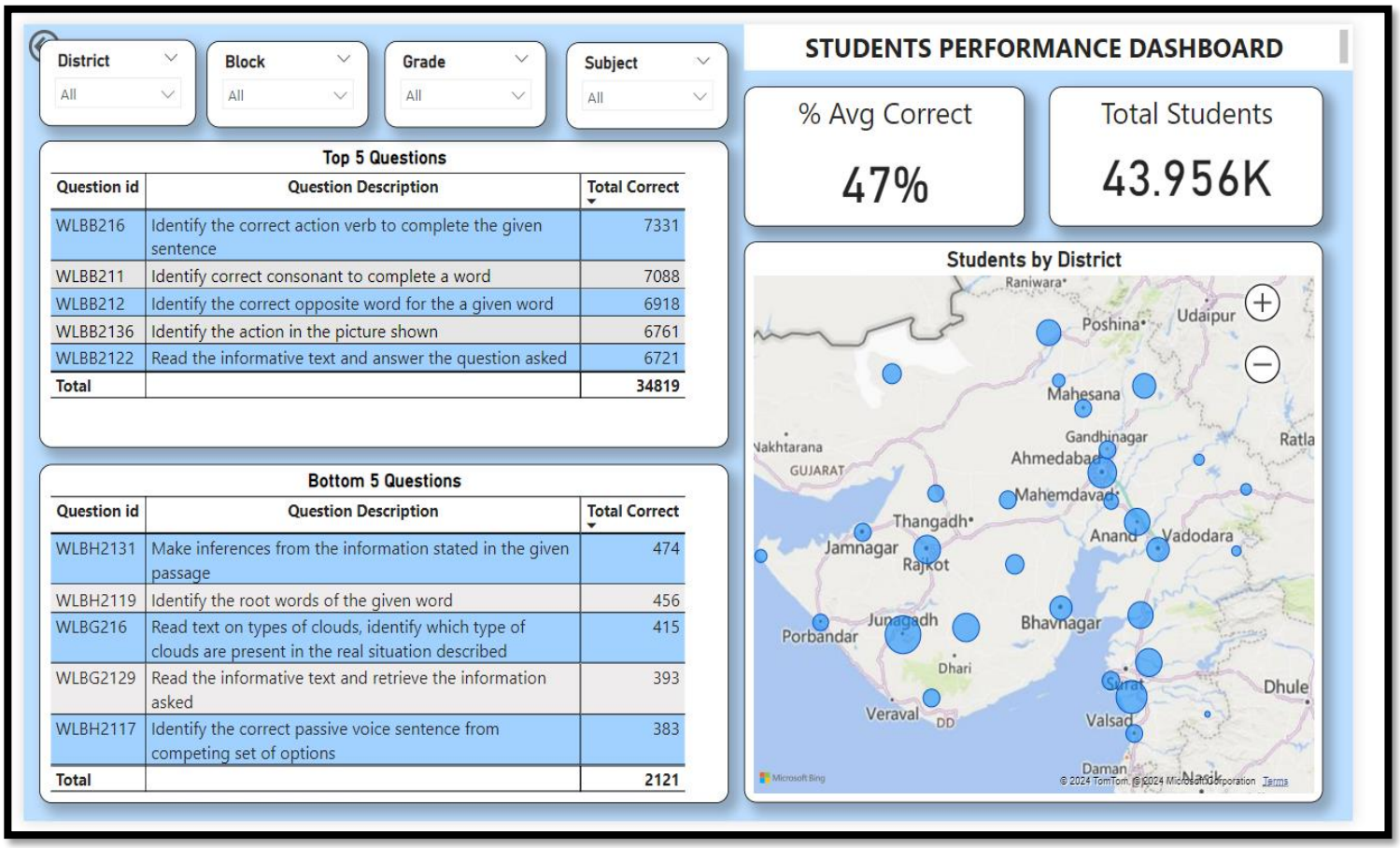

Figure: Students Performance Dashboard