Data Analysis & Visualization Report

1. Objective

The objective of this analysis is to clean, combine, and analyze student data collected from multiple raw Excel sheets and transform it into a single, structured dataset. Using this dataset, meaningful insights were derived through interactive visualizations to understand student distribution across branches, courses, age groups, and institutions.

2. Data Preparation & Processing

The raw data was provided across multiple Excel sheets with inconsistent column structures and missing values. The following steps were performed using Power Query:

- Combined multiple sheets into a single consolidated table using Append Queries
- Standardized column names (Name, Email, Mobile, Age, Branch, Course, Institution, Enrollment Number)
- Handled null values and removed empty rows
- Merged duplicate columns (such as multiple Email and Mobile fields) into single final columns
- Removed duplicate records based on E-mail to ensure data uniqueness
- Verified data consistency and integrity after merging

After processing, a clean dataset containing 761 unique student records was prepared for visualization.

3. Tools Used

- Microsoft Excel (Power Query) – Data cleaning and transformation
- Power BI Desktop – Data modeling and visualization

4. Key Insights from Analysis

Overall Statistics

- Total Students: 761
- Total Branches: 18
- Average Age: 22.48 years

Branch & Course Analysis

- Diploma courses account for the majority of enrollments
- Information Technology, Civil Engineering, Electronics, and Mechanical-related branches have the highest student counts
- Certain specialized branches show lower enrollment, indicating niche interest areas

Age Distribution

- Most students fall within the 18–25 age group
- The age distribution is concentrated around early 20s, indicating a typical diploma-level student demographic

Institution Analysis

- A small number of institutions contribute a large proportion of total enrollments
- Government Polytechnic Ghaziabad appears as a major contributor

5. Visualizations Created

The following visualizations were designed in Power BI to represent insights clearly:

- KPI cards for Total Students, Total Branches, and Average Age
- Bar charts showing student count by Branch and Institution
- Donut chart for Course distribution
- Line chart for student count across Age Groups
- Bar chart for Average Age by Branch
- Interactive slicers for Branch and Course to enable dynamic filtering

6. Conclusion

This analysis successfully transformed raw, unstructured data into a meaningful and interactive dashboard. The insights help in understanding student demographics, popular branches, age trends, and institutional contributions. The final dashboard is user-friendly, visually intuitive, and supports data-driven decision-making.

7. Future Improvements

- Include time-based data (admission year) for trend analysis
- Add geographic visualization using institution location
- Perform predictive analysis on enrollment trends