Aadhaar Student Enrolement Data Analysis Marach-July

BIG DATA ANALYTICS PROJECT REPORT

Name: A. Bhavani

Date: October 28 , 2025

Dataset Description

The dataset titled 'Enrolment_data_March–July.csv' contains student enrollment data across various departments for the months of March through July 2025. Each record provides information suchasstudentID,department,month,gender,andtotalenrollmentcount.Thisdatasetservesas the foundation for analyzing temporal and departmental enrollment patterns, gender participation, and monthly variations.

Domain Description

The project belongs to the Education Data Analytics domain. It focuses on processing and interpretingstudentenrollmentinformationtoidentifypatternsthataidinacademicandinstitutional decision-making. The analysis provides valuable insights for educational institutions to optimize admissions, resource allocation, and planning cycles using data-driven evidence.

Observed Insights & Hidden Facts

1. April 2025 Data Anomaly

Surface observation: A sudden drop in enrollments during April 2025 appears across multiple departments.Hiddenfact:Theuniformityofthedropindicatesalikelydataprocessingorcollection issue rather than a real event, suggesting a need for quality checks before future analysis.

2. Department Popularity and Concentration

Surfaceobservation:Certaindepartmentsconsistentlyshowhigherenrollmentcounts.Hiddenfact: Enrollment is concentrated in technical and management streams, highlighting strong student preferences and the need for balanced resource distribution.

3. Gender Participation Trends

Surfaceobservation:Bothmaleandfemaleenrollmentsincreaseovertime,withanoticeablerisein femaleparticipationbyJuly.Hiddenfact:Thistrendpointstogrowinginclusivity,possiblysupported by institutional initiatives and awareness programs.

4. Month-to-Month Growth Decoupling from Size

Surfaceobservation:Largerdepartmentsdonotnecessarilyshowfastermonth-over-monthgrowth. Hidden fact: Growth appears driven by specific campaigns or localized activities, emphasizing the importance of micro-level analysis.

Recommendations

1. ValidateandCleanAprilData:InspectAprilentriesforinconsistenciesandrectifyorexclude errors before future analyses.

2. ResourceOptimization:Aligndepartmentresourceswithenrollmentdemandtoprevent overloading popular programs.

3. EncourageFemaleParticipation:Continueinitiativesthatpromotebalancedgender representation across departments.

4. CreatePredictiveDashboards:Developliveanalyticsdashboardstomonitorandforecast department-wise and monthly enrollments.

Code Cells Description

1. Data Loading: The dataset was imported into a Spark DataFrame for distributed processing.

2. Data Cleaning: Missing values were replaced and columns standardized for uniformity.

3. Aggregation: Data was grouped by department and month to compute total enrollments.

4. Visualization: Pandas and Matplotlib were used to plot trends and departmental comparisons.

5. Export: Processed insights were summarized for reporting and visualization output.

Visualizations and Insights

Plot 1: Monthly Enrollment Trend (March–July 2025)

This plot highlights the overall monthly enrollment pattern. Enrollments increase steadily toward July,showingpeakadmissionactivity.Marchhasthelowestfigures,whileApril'sdipsuggestsdata inconsistencies. The consistent upward trend confirms academic-cycle-driven admissions.

Plot 2: Department-wise Enrollment Distribution

Thisvisualizationcomparesenrollmentsacrossdepartments.Technicalandmanagementcourses dominateduetohigherstudentdemand,whileartsandgeneralstudiesmaintainstableyetsmaller enrollments. This distribution underscores the need for balanced academic resource allocation.

Plot 3: Gender-wise Enrollment Comparison

This analysis compares male and female enrollment patterns. Both show growth, but female enrollmentsrisemoresharplyinlatermonths.Thenarrowinggendergapsuggestsincreased inclusivity and equal opportunity within institutional programs.

Plot 4: Heatmap - Enrollment Density by Month and Department

This analysis identifies the busiest months for each department. Bright cells indicate higher registrationintensity.Theheatmaphighlightsacademicpeaksandcanguidebetterschedulingand staffing decisions.

Plot 5: Average Enrollment per Month

Thissummaryshowsaverageenrollmentsacrossalldepartmentsforeachmonth.AprilandJuly record the highest averages, reflecting admission deadlines. These patterns help forecast institutional workload during academic cycles.

Conclusion

The Enrollment Data Analysis project successfully demonstrates how Python and Apache Spark can be integrated for efficient educational data analysis. The findings highlight temporal trends, departmentaldemand,andgenderparticipationimprovements.Byaddressingtheidentifieddata issuesandimplementingtherecommendations,institutionscanenhanceacademicplanningand operational efficiency.