



**INNOMATICS<sup>®</sup>**  
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

**PROJECT ON**

**# EDA Project: AMCAT Data  
Analysis**

# About me

<b>Name</b>	<b>Bankar Krushnakant</b>
<b>Background</b>	<b>BCA</b>
<b>Why you want to learn Data Science</b>	I want to learn Data Science because it's super cool to dig into data, find hidden patterns, and make sense of the numbers. Plus, it opens up a ton of opportunities to solve real-world problems in creative ways.
<b>Any work experience</b>	10 Months Internship Experience.
<b>Github</b>	<a href="https://github.com/krushna27">https://github.com/krushna27</a>

## Objective of the Project

- To explore the relationships between salary, educational background, and demographics.
- To test the claim that fresh Computer Science Engineering graduates earn between 2.5-3 lakhs.
- To investigate gender preferences for specializations.

# Summary of the Data

## Business Problem Understanding

- Objective: Analyze employee data to evaluate salary distributions and relationships among key factors.

## Data Collection

- Loaded employee dataset for analysis.

## Data Preprocessing & Cleaning

- Handled missing data and detected outliers.
- Converted categorical variables for analysis.

## Data Manipulation

- Normalized numerical data and aggregated by categorical variables.

## Univariate Analysis

- Explored distributions of numerical and categorical variables using PDFs, histograms, and count plots.
- Identified outliers in numerical columns.

## Bivariate Analysis

- Investigated relationships using scatter plots and pair plots for numerical variables.
- Used box plots and bar plots for categorical vs numerical analysis.
- Explored relationships between categorical variables with stacked bar plots.

# Process Flow

- **Hypothesis Testing**
- Tested salary claims for fresh graduates.
- Analyzed gender and specialization relationship using Chi-square test.
- **Conclusion & Key Findings**
- Summarized insights on salary distributions, education impact, and job market trends.
- **Insights & Recommendations**
- Suggested further analysis based on observed patterns and outliers.

# Summary of the Data

- **Total Records:** 3998
- **Columns:**
- **Numerical Variables:**
  - Salary, GPA, Exam Scores (English, Logical, Quant), Age-related attributes.
- **Categorical Variables:**
  - Designation, Job City, Gender, College Tier, Degree, Specialization, College State.
- **Key Observations:**
- **Salary:** Some outliers are there.
- **GPA:** Generally, higher GPA correlates with higher salaries.
- **Designation:** Most common designations are Senior Software Engineer and Assistant Manager.
- **Job City:** Major employment hubs include Bangalore, Gurgaon, and Chennai.
- **Gender Distribution:** Male employees outnumber females, indicating a gender imbalance.
- **College Tier:** Majority of employees graduated from Tier 2 colleges.

# Conclusion

- **Salary Trends:**
  - A positive correlation exists between college GPA and salary, indicating that higher academic performance can lead to better-paying jobs. However, some outliers reflect exceptions to this trend.
- **Designation Insights:**
  - Certain designations, particularly Senior Software Engineer and Assistant Manager, dominate the job market, highlighting the demand for these roles. Salary variations within designations suggest differences based on experience and skill levels.
- **Geographic Distribution:**
  - Major employment hubs such as Bangalore, Gurgaon, and Chennai offer more opportunities and higher salaries, indicating regional preferences in job placement.
- **Gender Disparities:**
  - The data reveals a notable gender imbalance in employment, with significantly more male employees than female. This suggests potential barriers or biases in hiring practices within the industry.
- **College Tier Influence:**
  - Graduates from Tier 1 colleges tend to earn higher salaries compared to those from Tier 2 colleges, indicating that college reputation may play a crucial role in initial salary offers.
- **Specialization and Gender Relationship:**
  - A significant relationship was found between gender and specialization preferences, rejecting the null hypothesis. This implies that gender influences the choice of specialization in engineering fields.

# Summary of the Data

- **Total Records:** 3998
- **Columns:**
- **Numerical Variables:**
  - Salary, GPA, Exam Scores (English, Logical, Quant), Age-related attributes.
- **Categorical Variables:**
  - Designation, Job City, Gender, College Tier, Degree, Specialization, College State.
- **Key Observations:**
- **Salary:** Some outliers are there.
- **GPA:** Generally, higher GPA correlates with higher salaries.
- **Designation:** Most common designations are Senior Software Engineer and Assistant Manager.
- **Job City:** Major employment hubs include Bangalore, Gurgaon, and Chennai.
- **Gender Distribution:** Male employees outnumber females, indicating a gender imbalance.
- **College Tier:** Majority of employees graduated from Tier 2 colleges.



THANK  
YOU

