



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

**INNOVATION. AUTOMATION. ANALYTICS**

## **PROJECT ON**

Exploratory Data Analysis (EDA)

### **Unveiling the Employment Landscape**

Analyzing Engineering Graduates' Outcomes through Cognitive,  
Technical, and Personality Skills

# About me

- I am currently pursuing a Master's degree in Business Analytics at University of Massachusetts at Amherst.  
I have 3 years of experience with Accenture as an Application Development Associate. I am eager to deepen my understanding of data-driven decision-making processes for big business.  
**LinkedIn** : <http://www.linkedin.com/in/kiran-bele>  
**GitHub** : <https://github.com/kiranbele11>

## Why This Project?

This project is a crucial step in my journey to becoming a data professional. It combines my passion for data analysis with my background in engineering, aiming to uncover valuable insights that can inform decisions in education and employment sectors.

# Agenda (This should be the PPT flow)

- Business Problem and Use case domain understanding(If Required)
- Objective of the Project
- Web Scraping – Details (Websites, Processor you followed)
- Summary of the Data
- Exploratory Data Analysis:
  - a. *Data Cleaning Steps*
  - b. *Data Manipulation Steps*
  - c. *Univariate Analysis Steps*
  - d. *Bivariate Analysis Steps*
- Key Business Question
- Conclusion (Key finding overall)
- Q&A Slide
- Your Experience/Challenges working on Web Scraping – Data Analysis Project.

# Objective:

- To conduct a thorough exploratory data analysis of the employment outcomes for engineering graduates, focusing on key factors like salary, specialization, and gender, and to address specific research questions related to these factors.

# Summary of the Data

- **Dataset Overview:**
  - The dataset consists of approximately 4,000 records of engineering graduates, with 40 independent variables including salary, job titles, academic scores, cognitive skills, technical skills, and personality traits.
- **Key Variables:**
  - **Salary:** The starting salary of graduates.
  - **Specialization:** The engineering discipline of the graduates.
  - **Gender:** Male or Female.
  - **Academic Scores:** 10th and 12th-grade scores, and college GPA.
  - **Skills:** Scores in areas like English, Logical, and Quantitative skills.
  - **Personality Traits:** Measures of traits like conscientiousness, extraversion, and openness to experience.

# Exploratory Data Analysis (EDA)

- **Data Cleaning Steps:** Removed duplicates and handled missing values. For example, any missing salary data was excluded from the analysis to ensure accurate results.
- **Data Manipulation Steps:** Grouped data by specialization and gender, calculated means and standard deviations of numerical columns, and created new columns for analysis (e.g., salary ranges).
- **Univariate Analysis:**
- **Salary Distribution:** Most graduates earn modest salaries, with a few outliers earning significantly more, indicating a right-skewed distribution.
- **Specializations:** A large concentration of students is found in specializations like Electronics and Communication Engineering and Computer Science & Engineering.
- **Personality Traits:** Traits like conscientiousness and agreeableness are normally distributed, showing a balanced mix among the graduates.
- **Bivariate Analysis:**
- **GPA vs. Salary:** There is a weak but positive correlation, suggesting that while academic performance influences salary, it is not the only determining factor. (Numerical vs numerical correlation analysis)
- **Gender vs. Specialization:** The analysis revealed that certain specializations are gender-dominated, with men dominating Mechanical and Civil Engineering, and women more represented in Biotechnology. (Categorical vs categorical crosstab analysis)
- **Salary by State:** The analysis showed significant variability in salaries across different states, with some states like Meghalaya offering higher salary potentials. (Categorical vs numerical analysis using box plots)

# Key Research Questions

- **Times of India Salary Claim Verification:**  
The claim that fresh graduates with a degree in Computer Science Engineering earn between 2.5-3 lakhs was tested. The analysis showed that the average salary for these roles is significantly higher, refuting the claim.
- **Gender and Specialization Relationship:**  
The analysis confirmed that gender does influence specialization choices. Men dominate in technical fields like Mechanical and Civil Engineering, while women are more prevalent in Biomedical Engineering and Biotechnology.
- **Gender and Salary:**  
In most specializations, there was no significant difference in starting salaries between male and female graduates. However, in fields like Electronics and Communication Engineering and Electrical Engineering, significant salary differences were observed, with men generally earning more.

# Conclusion

- **Key Insights:**  
The EDA revealed important trends in the employment outcomes of engineering graduates. **Specialization**, **gender**, and geographic **location** significantly impact starting salaries, though **academic** performance has a less pronounced effect. The analysis also highlighted **gender disparities in certain fields**, both in terms of representation and salary.
- **Implications:**  
These findings are valuable for educational institutions seeking to improve graduate employability, for employers aiming to address gender disparities, and for policymakers working to ensure equitable job opportunities.

# Learning Experience

Here's a summary of all the learning outcomes through this project :

- **Systematic Data Understanding:** Learned to approach large datasets methodically, focusing on understanding data types and patterns.
- **Data Cleaning & Manipulation:** Gained experience in handling missing data, correcting inconsistencies, and transforming variables, creating derived variables and grouping data to prepare it for effective analysis
- **Exploratory Data Analysis (EDA):** Used visualizations and statistical summaries to uncover patterns and trends across multiple factors.
- **Statistical Analysis:** Applied hypothesis testing, interpreting p-values, and selecting appropriate statistical tests.
- **Interdisciplinary Insights:** Integrated data from different domains to understand how various factors influence employment outcomes.
- **Critical Thinking:** Enhanced problem-solving skills by systematically addressing challenges like missing data and outliers.
- **Communication:** Improved ability to translate complex data findings into clear, actionable insights for a non-technical audience.

THANK  
YOU

