INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

Exploratory Data Analysis on Engineering Graduates Salary Dataset

About me

- **B-tech (2018-2022)**
- **Passion towards Data Science made me learn DS.**
- **No working experience**
- **Linkedin:**www.linkedin.com/in/uday-teja-dirsipamu
- **Github:**https://github.com/udaydirsipamu

INNOMATICS RESEARCH LABS

Index

- **Business Problem and Use case domain understanding.**
- **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.**

INNOMATICS RESEARCH LABS

Business Problem and Use Case Domain Understanding

- Understanding salary determinants for Engineering graduates
- Education and employment analysis in Engineering sector
- Understanding career decisions of a graduates on potential earnings

Objective of the project

- Explore and analyze factors influencing salaries of Engineering graduates
- To test claims about salary ranges for fresh graduates
- To provide actionable insights for Graduate students

Web Scraping

- Data set was provided.
- No processor was used as data set was available.
- The data set was available in excel file.
- The data was loaded using **pandas.**
- Preprocessing steps were used to clean data

Summary of the data

- **Data set description**:Key features are Salary,Designation,specialization and College Tier
- **Data size**:3998 rows and 39 columns
- **Data types**:Numerical (Salary,12 graduation etc)and Categorical columns(Designation,Specialization etc.)

- **Exploratory Data Analysis:**

  - **a.Data cleaning steps:**
  - Checked for missing values and dropped rows where **Salary** column is NAN.
  - Removed spaces and standardized text in **Specialization** and **Designation** columns.

b.Data manipulation steps:

- Filtered Computer science graduates with relevant jobs Programmer analyst,Software Engineering,Hardware Engineering and Associate Engineering.
- Calculated average salary based on Specialization,collegetier and experience.

c.Univariate Analysis steps:

- Plotted PDF with histograms for all numerical values like salary etc.
- Used boxplots to detect outliers in all columns.
- Created count plot for categorical values like Specialization and Designation.
- Salary distribution is right skewed.
- Outliers exist in higher salary ranges.
- Computer science and Information technology have highest average salary.

c.Bivariate Analysis:

- Analyzed relationship between numerical columns using Scatterplot and hexbin plot.
- Identified patterns between Categorical and numerical columns using box boxplot and barplot.
- Used stacked bar plots to explore relationships between Gender and Specialization.
- Graduates from Tier 1 colleges have high Salaries in average.
- Software Engineers and Programmer analyst have high salaries compare to Associate Engineers.

Key Business Question

- **Does the Times of India Claim Hold True?**: The article claimed that fresh Computer Science graduates earn between **2.5-3 lakhs** in certain roles
- My analysis showed that average salary for the given role is above 3 lakhs.

Conclusion

- The salary data reveals **variation across job roles and specializations**.
- Graduates from **Tier 1 colleges** and those in technical roles like **Software Engineer** earn higher salaries on average.
- The claim from the Times of India about salaries for fresh graduates was **contradicted by the dataset**, showing that actual salaries are higher than stated.

Q/A

- Do you have any clarifications on the salary trends across different job roles and specializations?
- What other factors, besides specialization and job title, might be influencing salary outcomes?
- Are there any noticeable salary differences based on gender for the same roles (e.g., Software Engineer)?

My Experience/Challenges working on Data Analysis Project.

- Handling missing values.
- Cleaning and standardizing text data for proper filtering.
- Testing the claim of research question.
- Gained hands-on experience in EDA
- Understanding the complexities of various numerical and categorical columns

THANK YOU

INNOMATICS RESEARCH LABS