



**INNOVATION. AUTOMATION. ANALYTICS**

**PROJECT ON**

# **Expolartory\_Data\_Analysis on Aspiring Mind Employment Outcome 2015 (AMEO)**



# About me

## Radhika Aggarwal

Data Science // Statistics // Mathematics

Innovative and detail-oriented professional with a passion for leveraging data to drive strategic decision-making. Adept at utilizing advanced analytical and programming skills to extract meaningful insights from complex datasets. Proven ability to communicate technical findings to both technical and non-technical stakeholders. Seeking a challenging role as a Data Scientist where I can contribute my skills in statistical analysis, machine learning, and data visualization to solve real-world problems.

## Education History

- **Bachelor of Mathematics Hons 2019-2022 (81.6 %)**  
Vedic mathematics certification  
Water Warriors, advocating water conservation and sustainability
- **Master of Data Science and Analytics 2022-2024 (9.25 SGPA)**  
Innov8 Chief Editor  
Performed in Anveshna and Onam

## Embracing the Data Science Journey: From Curiosity to Creative Problem-Solving"

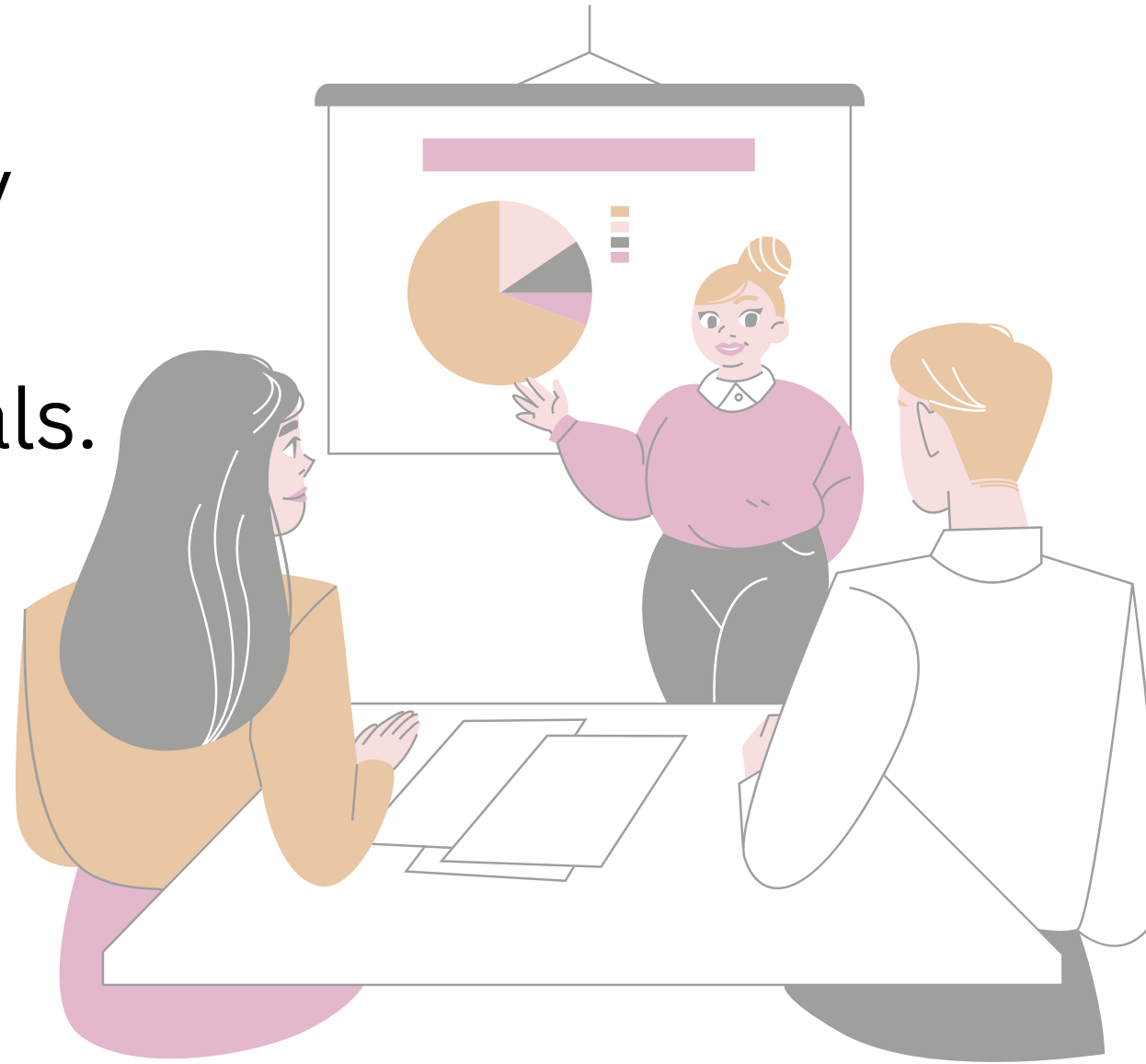
My passion for data science is deeply rooted in my lifelong curiosity about mathematics and its practical applications, constantly seeking connections in the world around me. It goes beyond a mere academic pursuit; it's about living my dream, especially when integrated with the marvels of artificial intelligence. I see data science as a means to address real-world challenges, utilizing analytical thinking to make optimal decisions. In a world filled with problem creators, I am determined to be a problem solver, finding satisfaction in contributing meaningful solutions. The creative aspect of transforming seemingly useless information into innovations that can amaze the world is what truly excites me.



# Bussiness Problem

## Salary Analysis Initiative

- Company HR faces challenges in accurately determining salaries for engineering professionals.
- Current approach lacks precision and neglects crucial factors: market trends, individual performance, and specialized skills.



# Objectives

- **Enhance Salary Precision:**  
Develop a data-driven model for accurate and consistent salary determination.
- **Recognize Specialized Skills:**  
Implement a system acknowledging and rewarding engineering expertise.

# Exploratory Data Analysis

## Data Summary

1. The dataset, capturing the employment outcomes of engineering graduates, comprises around 40 variables and 4000 data points. **Categorical variables** like 'Designation,' 'JobCity,' 'Gender,' and 'Specialization' provide insights into non-numeric attributes, reflecting job roles, locations, gender distribution, and academic specialties.
2. **Numerical variables**, including 'Salary,' '10percentage,' '12percentage,' and AMCAT assessment scores, offer quantifiable metrics for salary distributions, academic performance, and competency assessments. This dual representation of categorical and numerical data provides a comprehensive perspective on the dynamics of engineering graduates' employment outcomes.



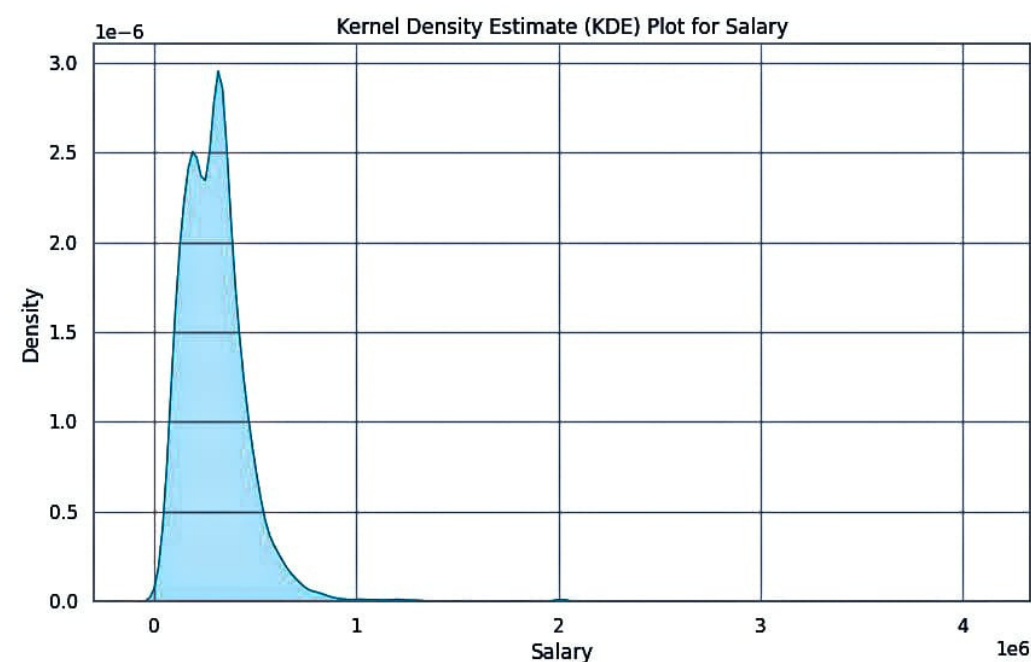
## Data Cleaning Steps

1. **Column Names Standardization:**
  - Enhancing Consistency and Readability
2. **Correct Data Formats:**
  - Ensuring Accuracy and Minimizing Errors
3. **Conversion of Data Types:**
  - Promoting Uniformity and Compatibility
4. **Titlecase Application:**
  - Aesthetic Standardization for Consistent Presentation
5. **Handling Missing Data:**
  - Comprehensive Examination and Strategies for Completeness
6. **Identification and Treatment of Outliers:**
  - Ensuring Robust Analysis Through Proper Handling

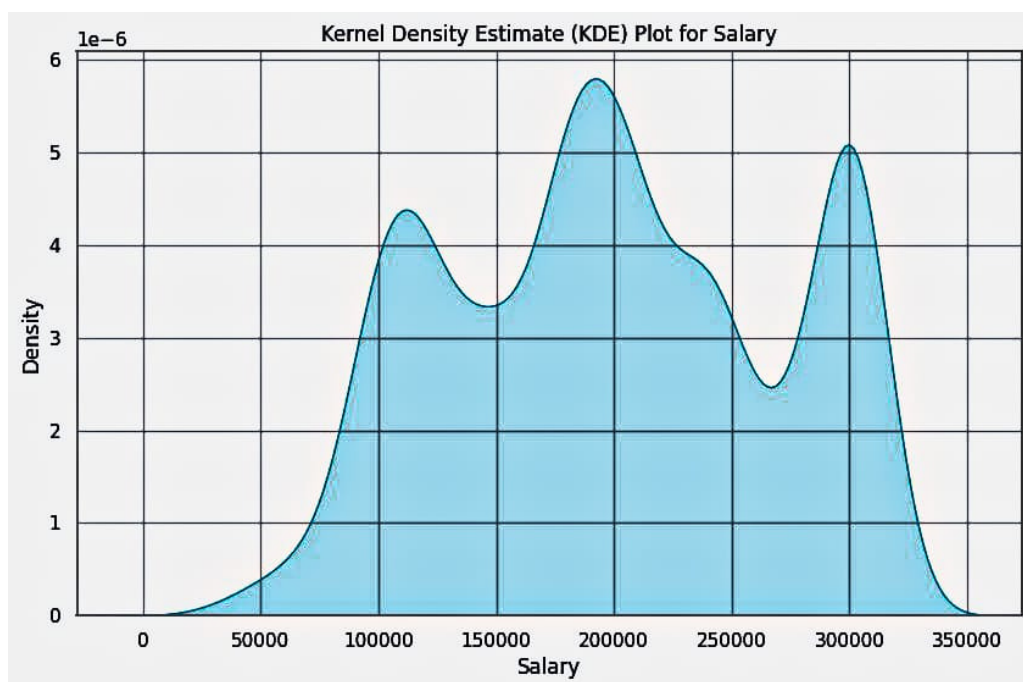


# Analysis of Data

Before Outlier Treatment



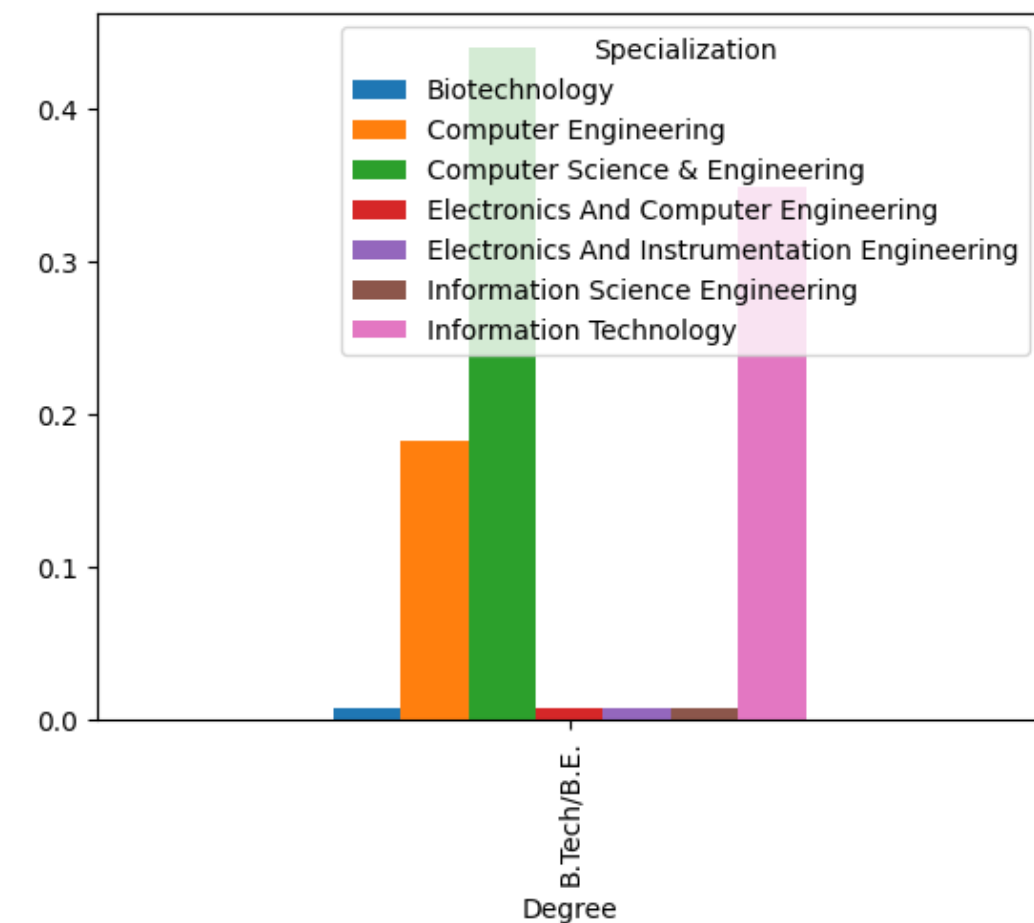
After Outlier Treatment



← **Univariate**

Following the outlier treatment applied to the 'Salary' column, the salary range has been adjusted, resulting in a more condensed distribution that appears to align more closely with a normal distribution. By mitigating the influence of extreme values, the data is now characterized by a narrower spread of salary values, contributing to a more representative and refined depiction of the salary distribution within the dataset."

## Bivariate



Based on the visible bars, it appears that Computer Science & Engineering has the highest percentage of graduates, followed by Electronics And Computer Engineering and Information Technology. The other specializations seem to have lower percentages.

# Research Question

## Testing the Times of India Claim

**Null Hypothesis (H0):** The average salary for fresh graduates with designations 'Programming Analyst,' 'Software Engineer,' 'Hardware Engineer,' and 'Associate Engineer' is equal to or less than 2.5-3 lakhs.

**Alternative Hypothesis (H1):** The average salary for fresh graduates with these designations is significantly greater than 3 lakhs.

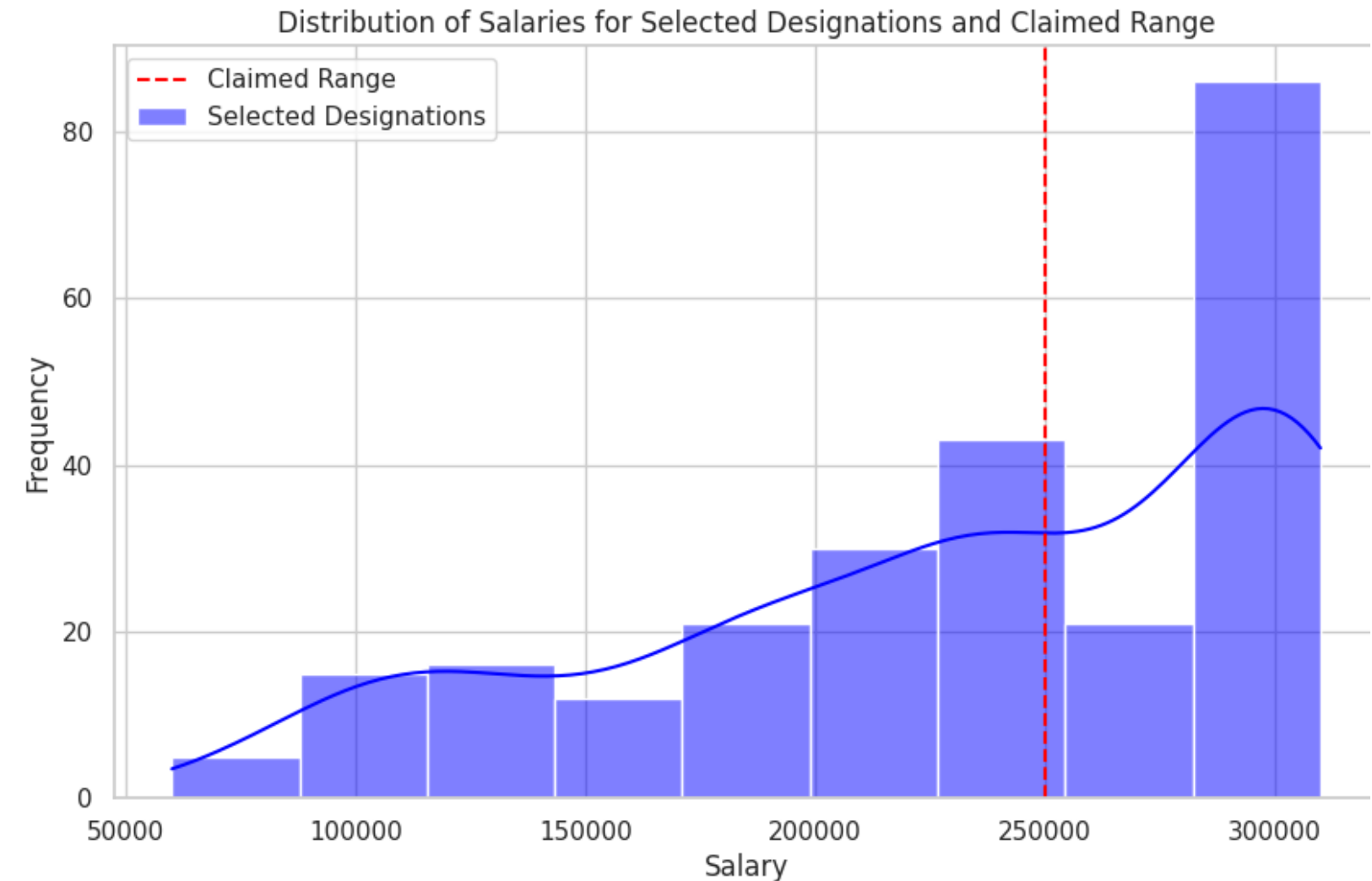
### Statistical Test Result:

t-Statistic: [-3.9194240092136106]

P-Value: [0.00010123189655940822]

### Conclusion:

- **Result:** Reject the null hypothesis.
- **Interpretation:** The average salary is significantly greater than the claimed range (\$250,000).



Set significance level **alpha** = 0.05

# Hypothesis Based on Gender and Specialization

## Null Hypothesis (H0):

There is no significant relationship between gender and specialization preference among engineering graduates.

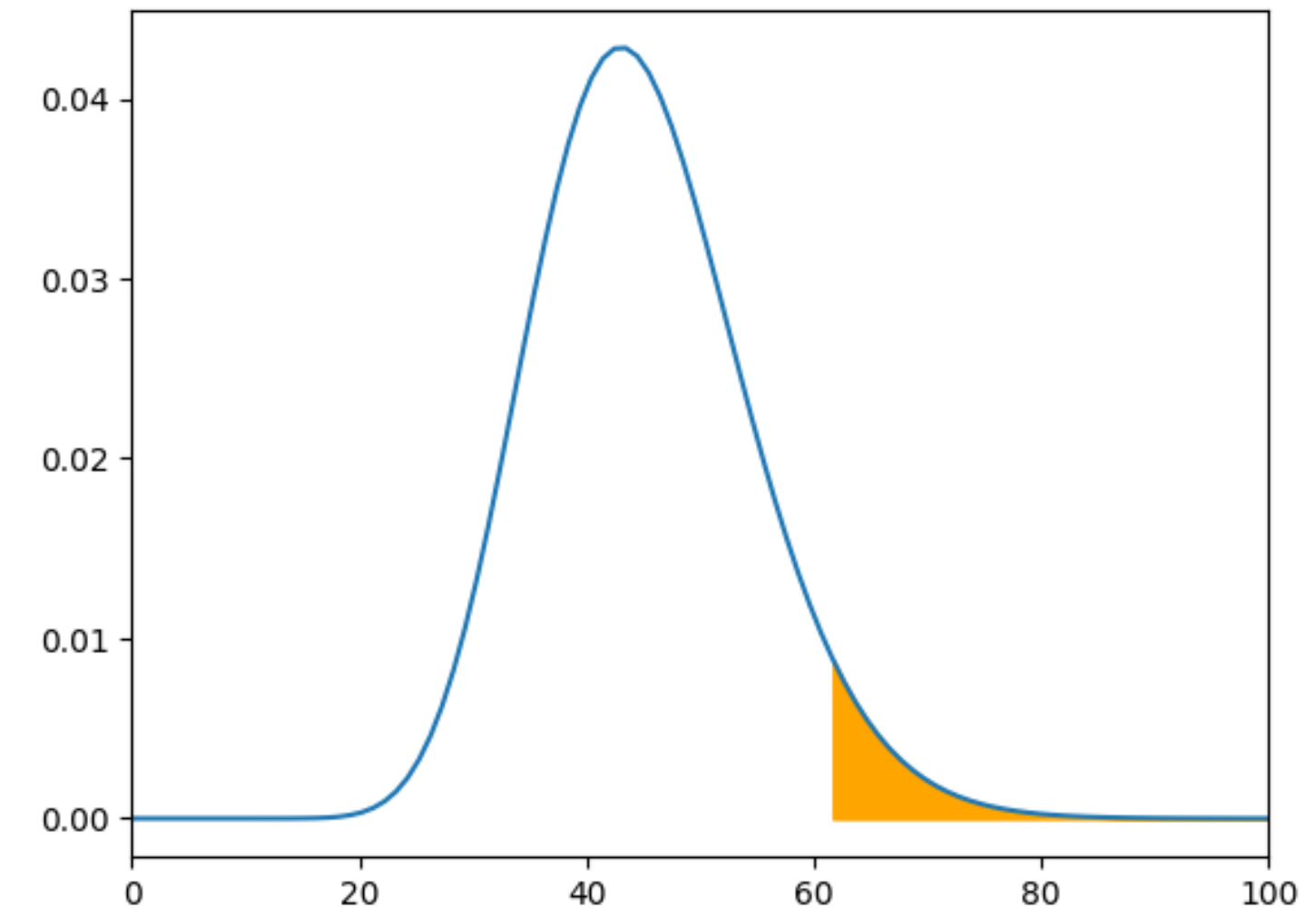
## Alternative Hypothesis (H1):

There is a significant relationship between gender and specialization preference among engineering graduates.

Critical Value at 90% Confidence Level: 61.65623337627955

Calculated Chi-Squared Test Statistic: 104.46891913608455

Confidence Level: 90% Degrees of Freedom: 45



Set significance level  $\alpha = 0.05$

**Calculated chi-squared test statistic > Critical Value at 90% Confidence Level**

- **Interpretation:** Reject the null hypothesis.
- **Conclusion:** Strong evidence suggests that the Gender and Specialization are not independent.

# CONCLUSION

After thoroughly analyzing the data in Jupyter notebook, here are the key takeaways:

1. **Top-Paying Field:** Computer Science & Engineering specialization stands out, offering the highest median salary.
2. **Gender Enrollment Trends:** Males tend to enroll more in specialization courses compared to females.
3. **Lucrative Domain:** The Software Engineer domain leads with the highest average salary at ₹209,166.67 per year.
4. **Salary Gender Gap:** On average, females earn around ₹203,648.65, slightly surpassing the mean salary of ₹194,105.26 for males.
5. **Workforce Dominance:** Software Engineer Domain has the largest number of employees, indicating its significance in the job market.
6. **Educational Alignment:** Graduates in Computer Science & Engineering show a strong alignment between their degree and specialization choices.
7. **Gender-Based Preferences:** Females favor Information Technology, while males lean towards Computer Science and Engineering in their chosen specializations.



# Thank You

For more better understanding of my project  
please refer to my LinkedIn and GitHub

