Exploratory Data Analysis On AMEO data for the year 2015

About me

- My name is Manish Putnala, and I am currently pursuing the final year of B.Tech in Computer Science and Engineering.

- I'm inclined towards learning Data Science for the following reasons:
  1.The rapid growth and promising future of this domain.
  2.Numerous career opportunities across the field.
  3.I'm always curious to research and extract crucial insights, patterns, and relations from
  data to find potential solutions to tackle real-world organizational requirements.

- I am a student with no prior work experience seeking opportunities in the realm of Data Science.

- My Linkedin: https://www.linkedin.com/in/manishputnala

About AMEO Data

For every engineer, AMEO dataset provides anonymised bio data information along with their respective skill scores and employment outcome information. Specifically, the following information is available for every engineer:

1. Scores on Aspiring Minds' AMCAT - a standardized test of job skills. The test includes cognitive, domain and personality assessments.
2. Personal information like gender and date of birth.
3. Pre-university information like 10th and 12th grade marks, board of education and 12th grade graduation year.
4. University information like GPA, college major, college reputation proxy, graduation year and college location.
5. The following employment outcome information is available for every engineer:

- Job annual salary
- Job title
- Job location
- Date of joining and leaving of job

Full report and Implementation :

- https://drive.google.com/file/d/1w9FxTqvvG5iEk-Mujt3jV7ayvJDlir7v/view?usp=share_link

Objective

The main objective of the project is:

- To understand the business requirements
- To perform Exploratory Data Analysis
- To find patterns and relations between variables
- To extract the key findings and insights
- To find the solutions for the business needs

Summary of Data

Unnamed: 0	object
ID	int64
Salary	float64
DOJ	object
DOL	object
Designation	object
JobCity	object
Gender	object
DOB	object
10percentage	float64
10board	object
12graduation	int64
12percentage	float64
12board	object
CollegeID	int64
CollegeTier	int64
Degree	object
Specialization	object
collegeGPA	float64
CollegeCityID	int64
CollegeCityTier	int64
CollegeState	object
GraduationYear	int64
English	int64
Logical	int64
Quant	int64
Domain	float64
ComputerProgramming	int64
ElectronicsAndSemicon	int64
ComputerScience	int64
MechanicalEngg	int64
ElectricalEngg	int64
TelecomEngg	int64
CivilEngg	int64
conscientiousness	float64
agreeableness	float64
extraversion	float64
nueroticism	float64
openess_to_experience	float64

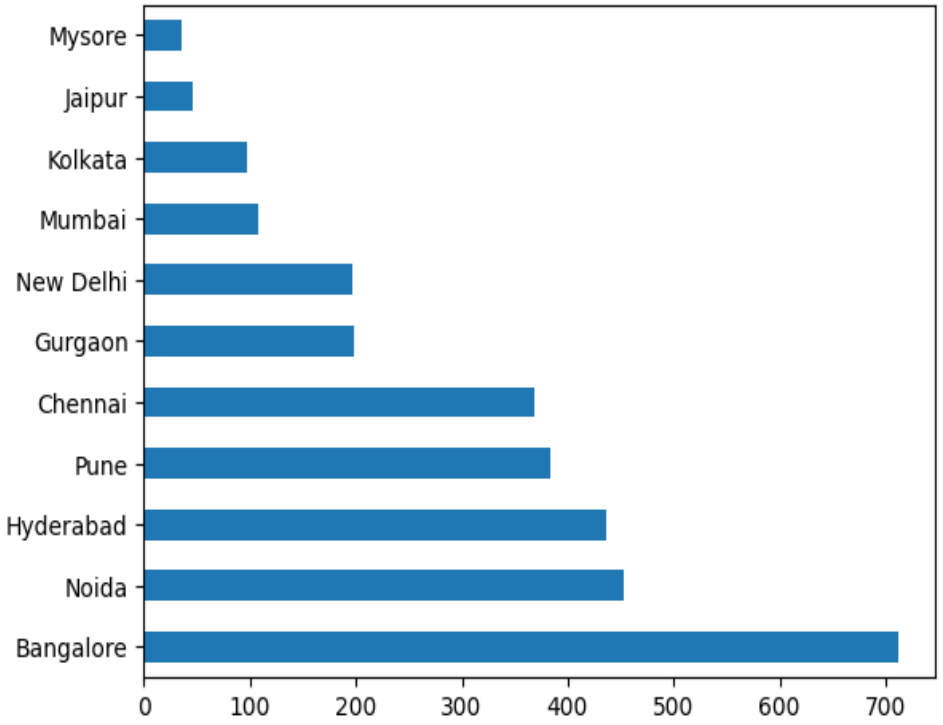
- The dataset contain 39 columns with Numerical and Categorical data types.
- It has 3998 rows of data .
- The dataset contains of engineering graduates.
- It contains the attributes such as graduates Salary, Date of joining, Date of leaving, Personal information such as Date of birth, Gender.
- It also contains data related to graduates academics such as 10th,12th and College academic performance, Graduation year and board.
- Contains the geographical data such as Job City, College State.
- It also contains standardized scored related to graduated Cognitive, Technical and Personality skills.

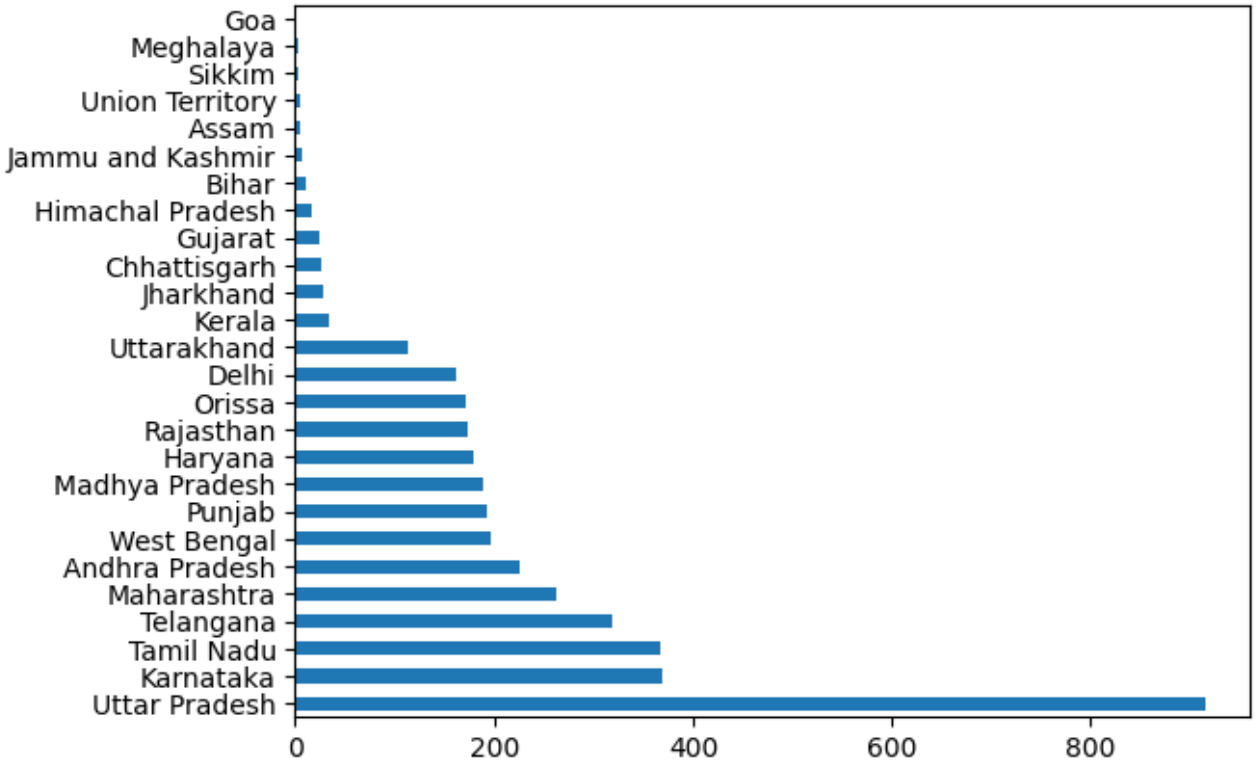
Data Cleaning and Transformation

```
RangeIndex: 3998 entries, 0 to 3997
Data columns (total 38 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   ID                     3998 non-null   int64
 1   Salary                 3998 non-null   float64
 2   DOJ                    3998 non-null   datetime64[ns]
 3   DOL                    3998 non-null   datetime64[ns]
 4   Designation            3998 non-null   object
 5   JobCity                3998 non-null   object
 6   Gender                 3998 non-null   object
 7   DOB                    3998 non-null   datetime64[ns]
 8   10percentage           3998 non-null   float64
 9   10board                3998 non-null   object
 10  12graduation           3998 non-null   int64
 11  12percentage           3998 non-null   float64
 12  12board                3998 non-null   object
 13  CollegeID              3998 non-null   int64
 14  CollegeTier            3998 non-null   int64
 15  Degree                 3998 non-null   object
 16  Specialization         3998 non-null   object
 17  collegeGPA             3998 non-null   float64
 18  CollegeCityID          3998 non-null   int64
 19  CollegeCityTier        3998 non-null   int64
 20  CollegeState           3998 non-null   object
 21  GraduationYear         3998 non-null   int64
 22  English                3998 non-null   int64
 23  Logical                3998 non-null   int64
 24  Quant                  3998 non-null   int64
 25  Domain                 3998 non-null   float64
 26  ComputerProgramming    3998 non-null   int64
 27  ElectronicsAndSemicon  3998 non-null   int64
 28  ComputerScience        3998 non-null   int64
 29  MechanicalEngg         3998 non-null   int64
 30  ElectricalEngg         3998 non-null   int64
 31  TelecomEngg            3998 non-null   int64
 32  CivilEngg              3998 non-null   int64
 33  conscientiousness      3998 non-null   float64
 34  agreeableness          3998 non-null   float64
 35  extraversion           3998 non-null   float64
 36  nueroticism            3998 non-null   float64
 37  openess_to_experience  3998 non-null   float64
```

- The dataset contains no missing values.
- Removed Unnamed column because it is not useful.
- Filtered columns into Numerical and Categorical columns.
- Changed datatype of Date columns from object type to datetime type.
- Handled Invalid values and replaced them with valid ones

Univariate Analysis

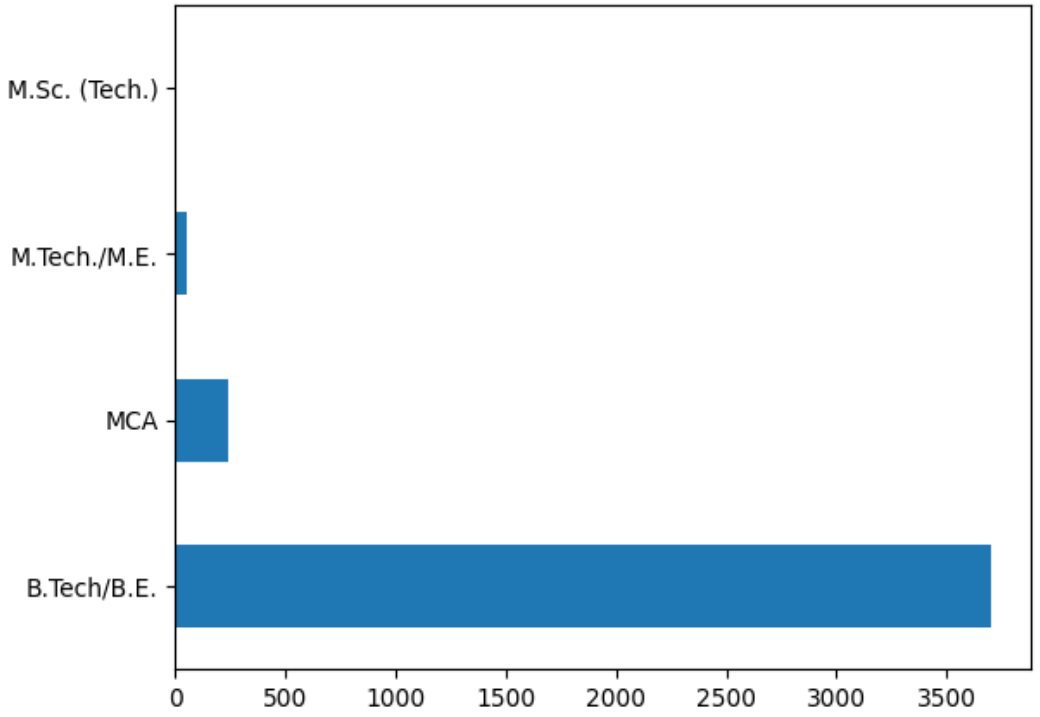


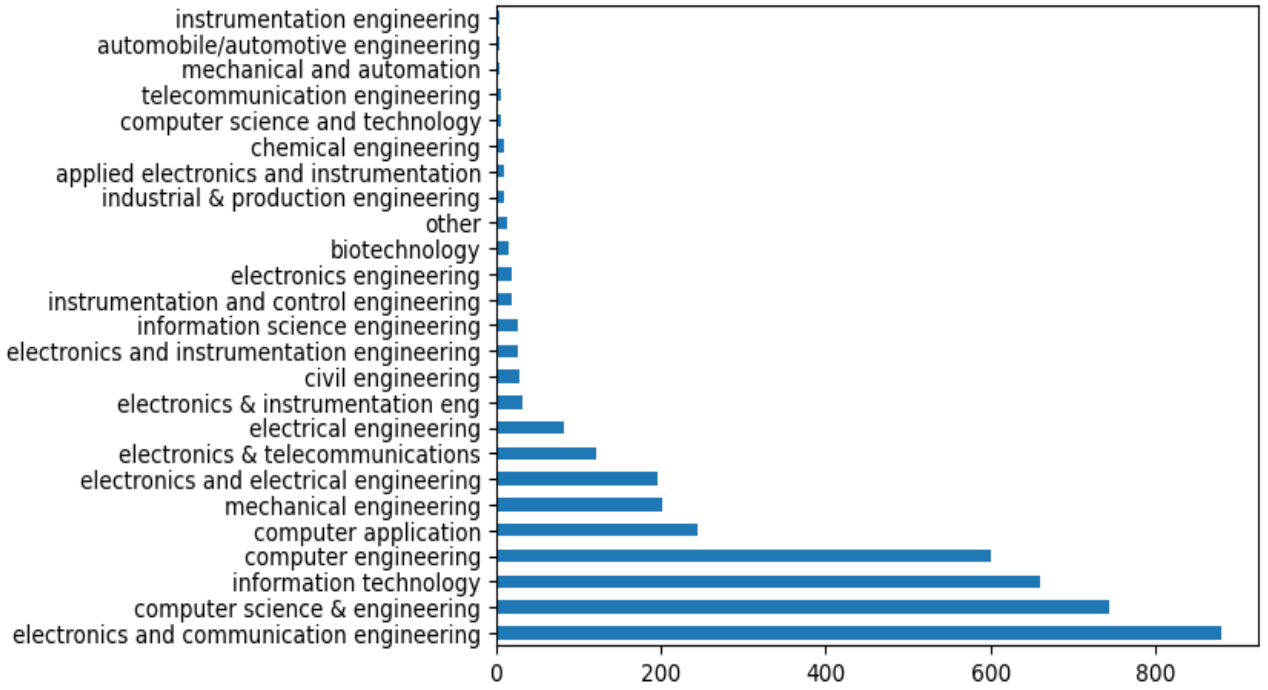


- Most of the graduates are working in cities like Bangalore, Noida, Hyderabad and Pune.

- Most of the graduates pursued their graduation in Uttar Pradesh followed by Karnataka, Tamil Nadu, Telangana and Maharashtra.

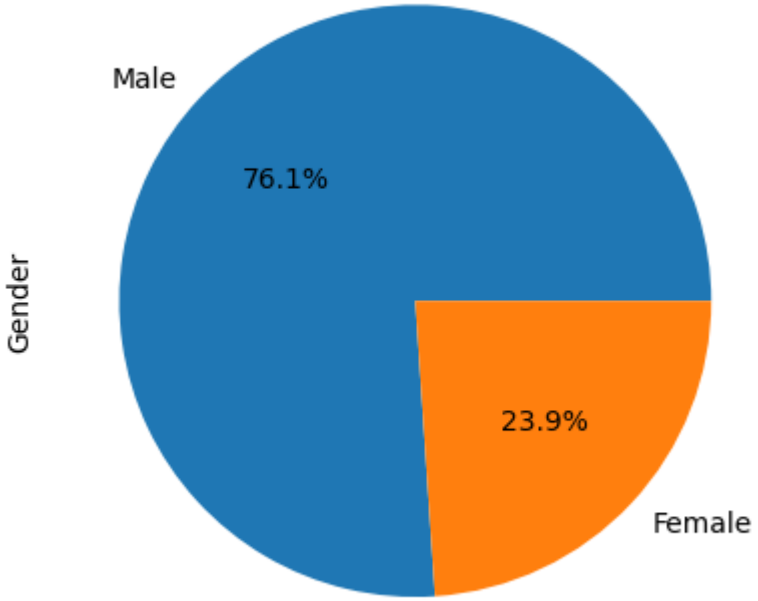
Univariate Analysis

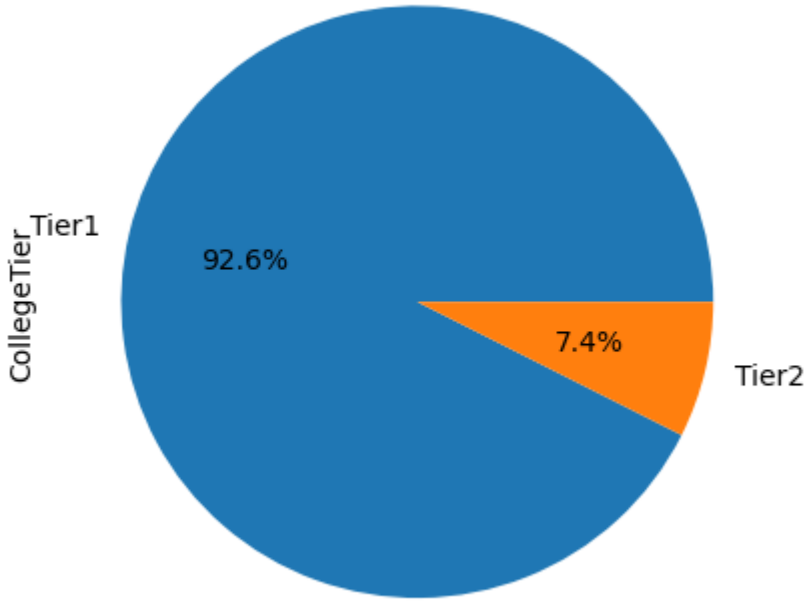




- Most of the graduates are working in cities like Bangalore, Noida, Hyderabad and Pune.
- Most of the graduates pursued their graduation in Uttar Pradesh followed by Karnataka, Tamil Nadu, Telangana and Maharashtra.

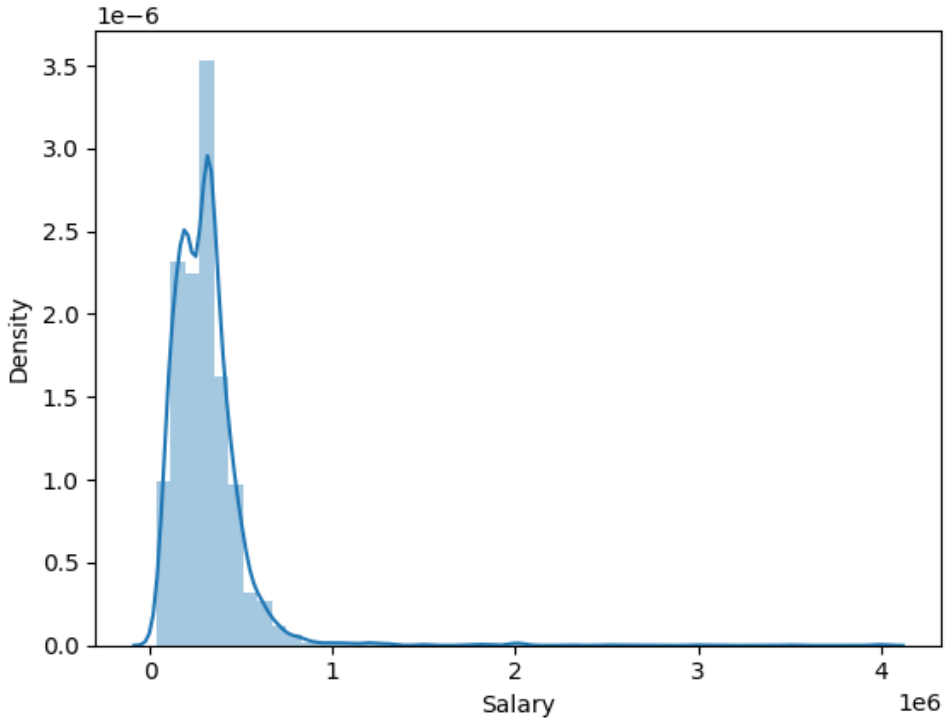
Univariate Analysis

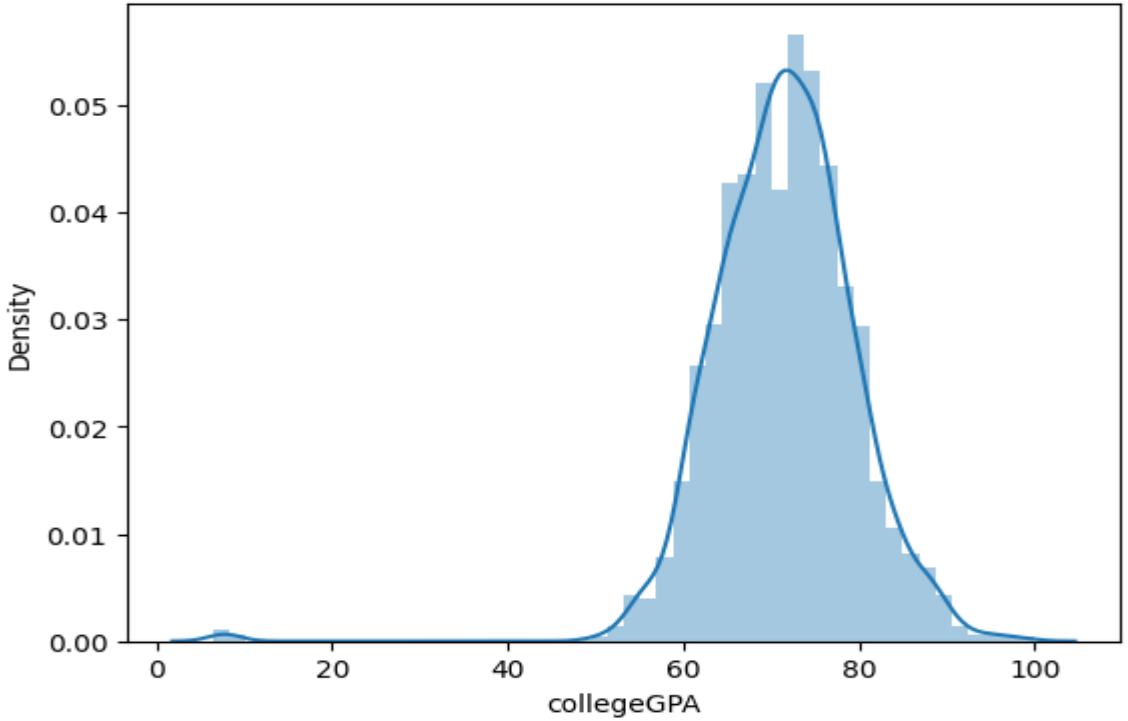




- 76.1% of the graduates are men, The male to female ratio is approximately 3:1.
- Most of the graduates belong to Tier 2 college there are only 7.4% of graduates who graduated from a Tier 1 College.

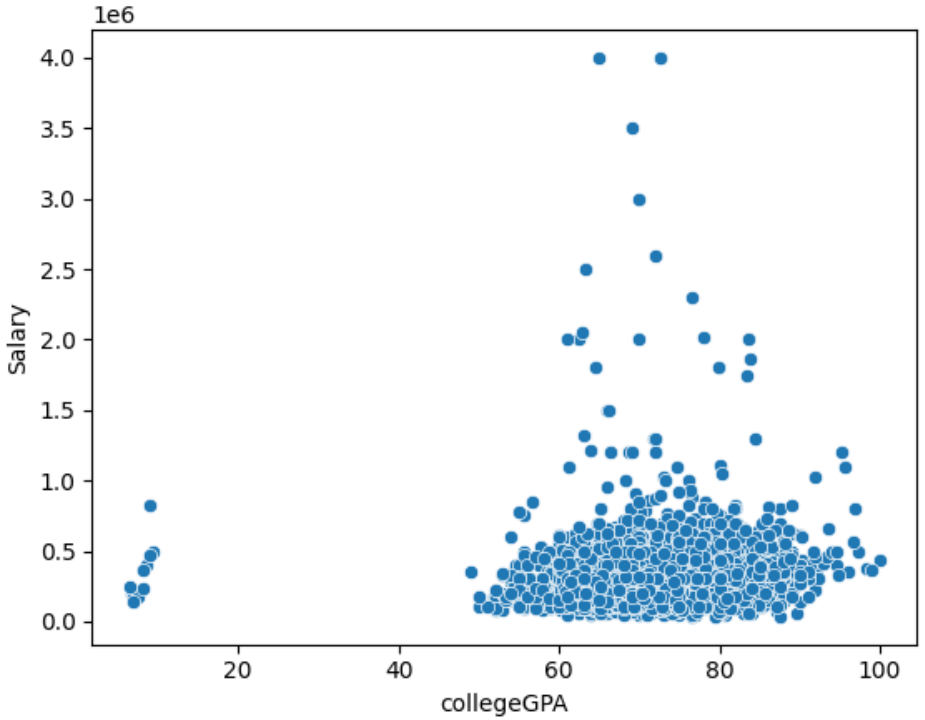
Univariate Analysis

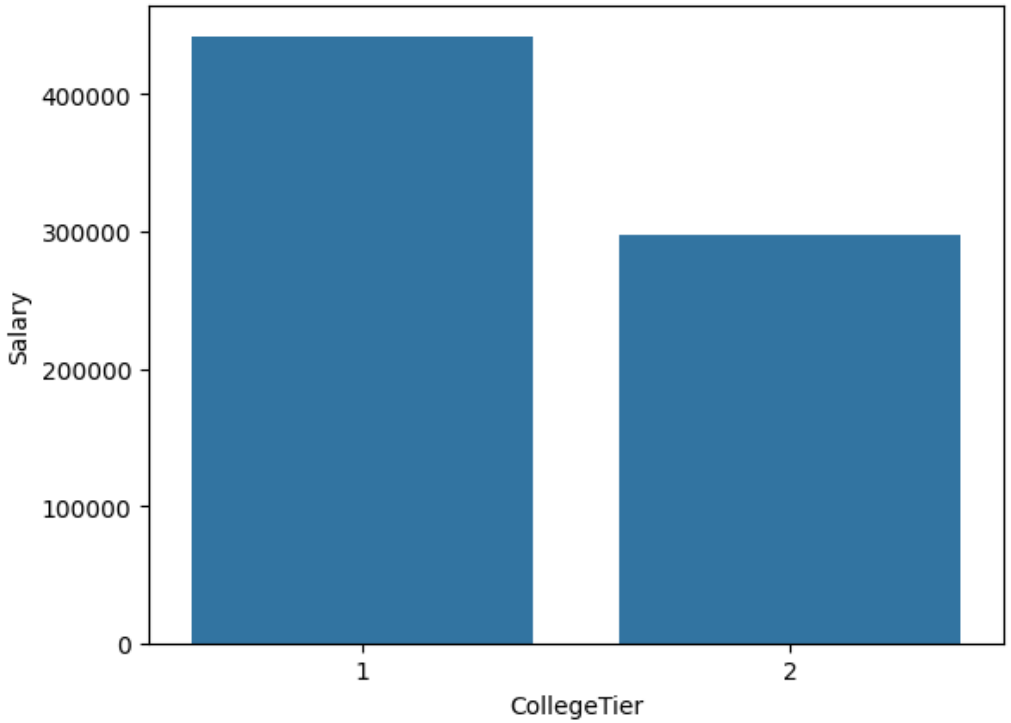




- The graphs display the distribution of Salary on left and Distribution of College GPA on Right
- Most of the graduates earn approximately 2.5 lakhs - 3 lakhs
- Most of the graduates have a College GPA between 6.0 - 8.0

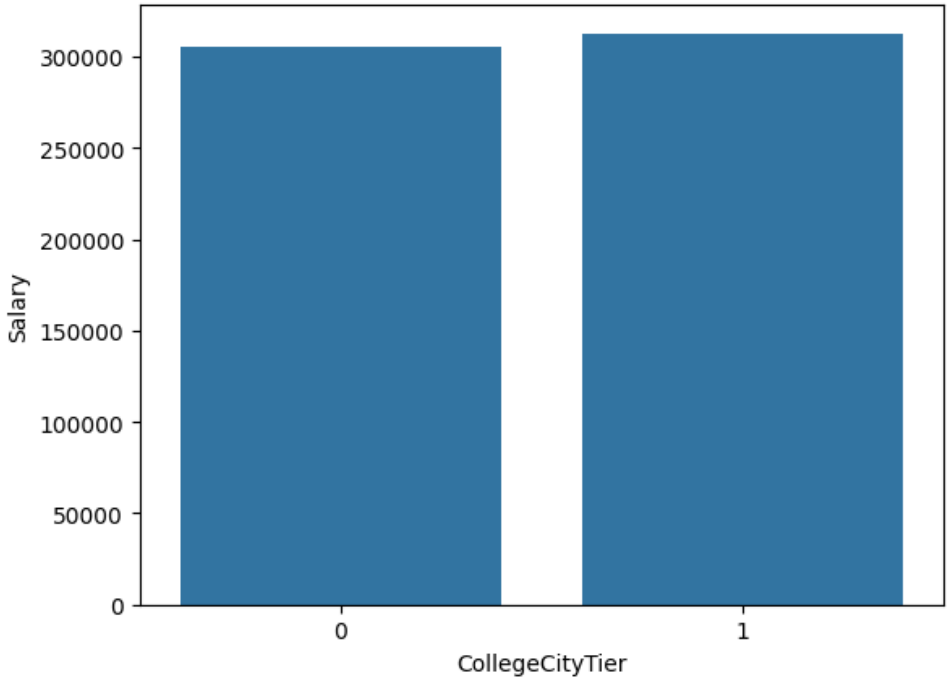
Bivariate Analysis

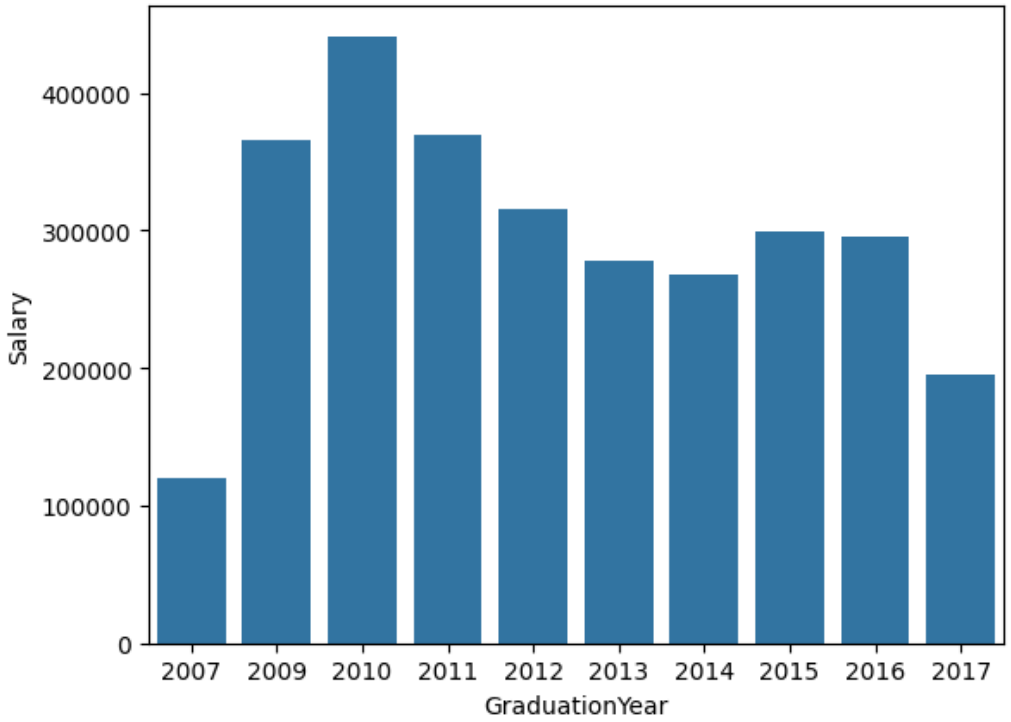




- The graphs show the relationship between the Salary and College GPA on the left and Salary and College Tier.
- The Salary and College GPA has a positive correlation as most of the graduates with higher College GPA secured higher salary package.
- The graduates from Tier 1 colleges have higher salary packages than Tier 2 college graduates.

Bivariate Analysis

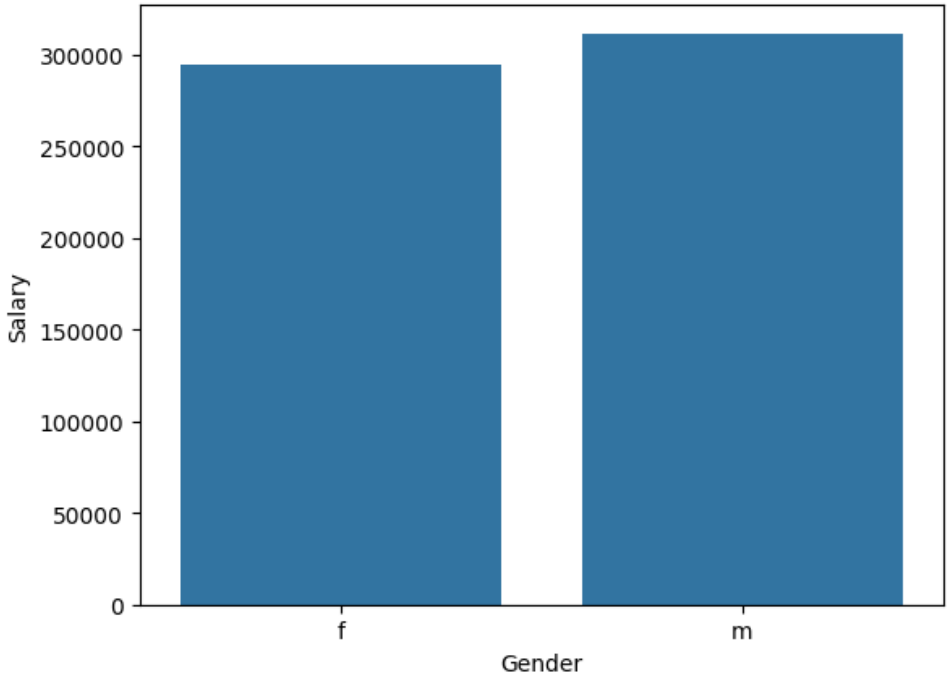


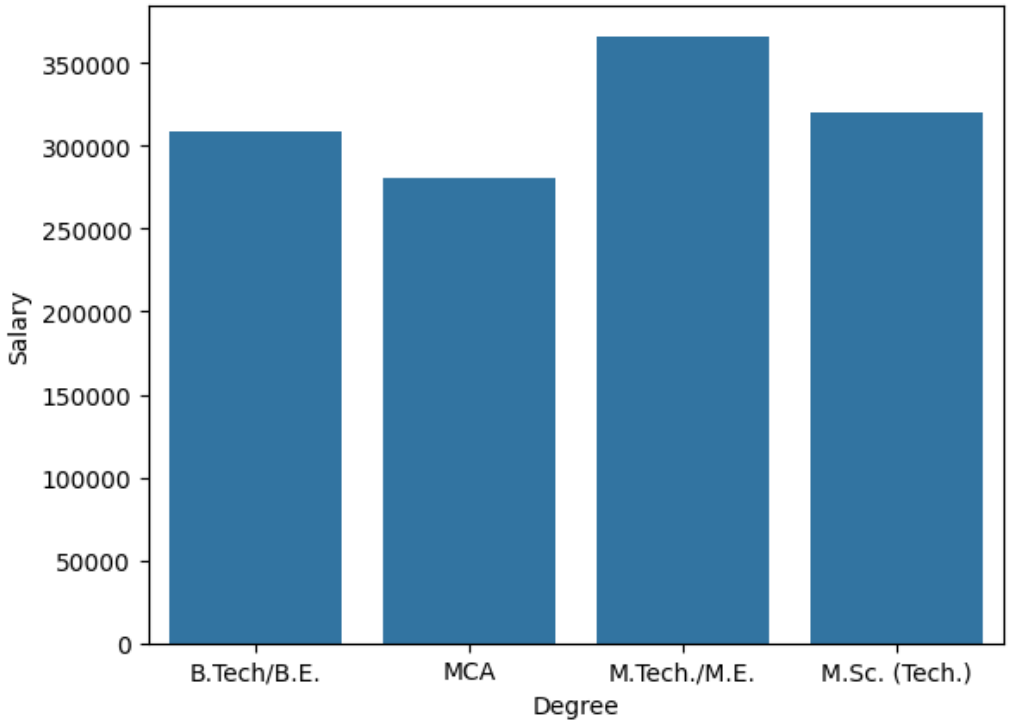


- The graduates from colleges located in Tier 1 cities earn slightly more than graduates from colleges located in Tier 2 cities.

- The graduates who graduated in the year 2010 have secured higher salary package compared to others.

Bivariate Analysis





- Males salary is slightly higher than the females.
- M.Tech/M.E graduates earn higher salary compared to others.

Research Question Analysis

- Question -1 Times of India article dated Jan 18, 2019 states that "*After doing your Computer Science Engineering if you take up jobs as a Programming Analyst, Software Engineer, Hardware Engineer and Associate Engineer you can earn up to 2.5-3 lakhs as a fresh graduate.*" Test this claim with the data given to you.

```
[77] ameo['Specialization'] = ameo['Specialization'].str.lower()
     ameo['Designation'] = ameo['Designation'].str.lower()
     hypo_1 = ameo[(ameo['Specialization'] == 'computer science & engineering') &
                   (ameo['Designation'].isin(['programmer analyst', 'software engineer', 'hardware engineer', 'associate engineer']))]
     print(hypo_1.iloc[0:,[4,16]].value_counts())
     print(f'Total Favourable Count: {len(hypo_1)}')
```

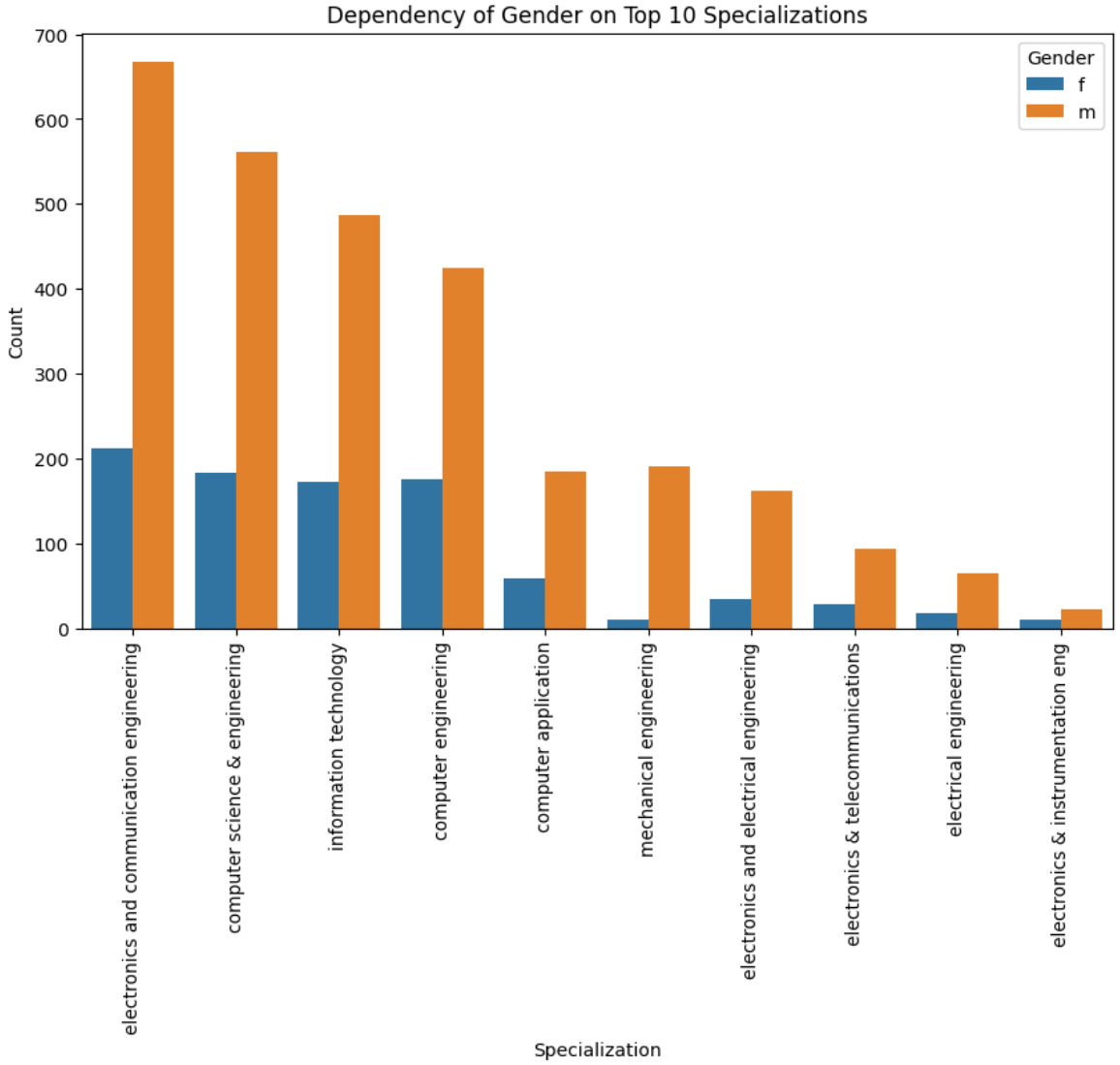
```
Designation         Specialization
software engineer   computer science & engineering    139
programmer analyst  computer science & engineering     26
associate engineer  computer science & engineering      2
dtype: int64
Total Favourable Count: 167
```

```
[91] hypo_1['Salary'].mean().round(2)
```

```
331526.95
```

- The dataset only has 167 records that meet the criteria hence the data is inadequate to test the claim
- However if we find the average of the available data it is approximately 331526.95
- With respect to the outcome of the data we can say that a person with computer science engineering specialization can earn up more than 3 LPA if they take up jobs as Programming Analyst, Software Engineer, Hardware Engineer or as an Associate Engineer.

Question –2 Is there a relationship between gender and specialization? (i.e. Does the preference of Specialization depend on the Gender?



- This count plot show us the relationship between genders and specialization.
- By looking at the graph we can say that there is a relationship between gender and specialization.
- The graph indicates that every specialization is male dominant.
- The male to female ratio is significantly high in specializations such as Mechanical Engineering and Electronics and Electrical Engineering.
- Most of the females prefer specializations such as Electronics and Communication Engineering, Computer Science Engineering and Information Technology.
- From the available data we can conclude that there is a dependency between Gender and Specialization.

Conclusion

- The dataset contains data of engineering graduates which was released by AMEO in the year 2015.
- The dataset contains data of engineering graduates such as graduates personal, academic and professional information.
- It also contains information related to their Technical, Cognitive and Personality skills.
- Dataset has no missing values but few columns contain invalid values.
- Around 76% of the data is of male graduates.
- Average salary of these engineering graduates is approximately 3 lakhs per annum.
- The Technical, Cognitive and Personality skills all are positively correlated with salary of the graduates.

Conclusion

- Graduates from Tier 1 colleges and Tier 1 cities tend to earn more compared to those from Tier 2 colleges and Tier 2 cities.
- Most of them work in cities like Bangalore, Noida, Hyderabad, Pune, and Chennai.
- A majority of these graduates hail from states such as Uttar Pradesh, Karnataka, Tamil Nadu, Telangana, and Maharashtra.
- The majority of graduates hold a B.Tech/B.E degree with specializations in branches such as Electronics and Communication Engineering, Computer Science Engineering, and Information Technology.
- Graduates with good academic performance tend to have higher salaries compared to those who performed poorly during their academic years.
- Graduates with an M.Tech/M.E degree earn the highest salaries, followed by those with an M.Sc and B.Tech/B.E degree.

Conclusion

- The data is inadequate to test and answer the question, namely whether a graduate with a specialization in Computer Science Engineering, taking up a job as a Programmer Analyst, Software Engineer, Hardware Engineer, or Associate Engineer, can earn between 2.5 lakhs and 3 lakhs per annum.
- However, with the limited data available, the average salary meeting the above criteria is approximately 3.3 lakhs per annum. So, if a Computer Science Engineering graduate takes up a job as a Programmer Analyst, Software Engineer, Hardware Engineer, or Associate Engineer, the expected salary range would be between 2.5 and 3 lakhs per annum.
- Furthermore, the data suggests that Specialization and Gender are dependent on each other. Women predominantly consider specializations such as Electronics and Communication Engineering, Computer Science Engineering, and Information Technology, whereas men are diversely spread across different specializations according to the data.

THANK YOU