INNOMATİCS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

Aspiring Minds from the Aspiring Mind Employment Outcome 2015 (AMEO) Data

About me

- My self C G KISORE
- I from Bachelor of Computer Science Background and present pursuing MCA
- Why I choose Data Science because from my teenage I have been very much interest in AI technologies. The evolution of AI in since 1950s if you want to develop AI we want lot of data to train. I want to pursue my career in AI. To get into AI first we have to learn how to handle data ,manage data that's why choose to pursue Data Science to learn.
- I didn't have any work experience .I am still a student
- Linkedin :- www.linkedin.com/in/cgkishore

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.**

INNOMATICS RESEARCH LABS

About the Data

The dataset was released by Aspiring Minds from the Aspiring Mind Employment Outcome 2015 (AMEO). The study is primarily limited only to students with engineering disciplines. The dataset contains the employment outcomes of engineering graduates as dependent variables (Salary, Job Titles, and Job Locations) along with the standardized scores from three different areas – cognitive skills, technical skills and personality skills. The dataset also contains demographic features. The dataset contains around 40 independent variables and 4000 data points. The independent variables are both continuous and categorical in nature. The dataset contains a unique identifier for each candidate. Below mentioned table contains the details for the original dataset.

Columns of the Data

'

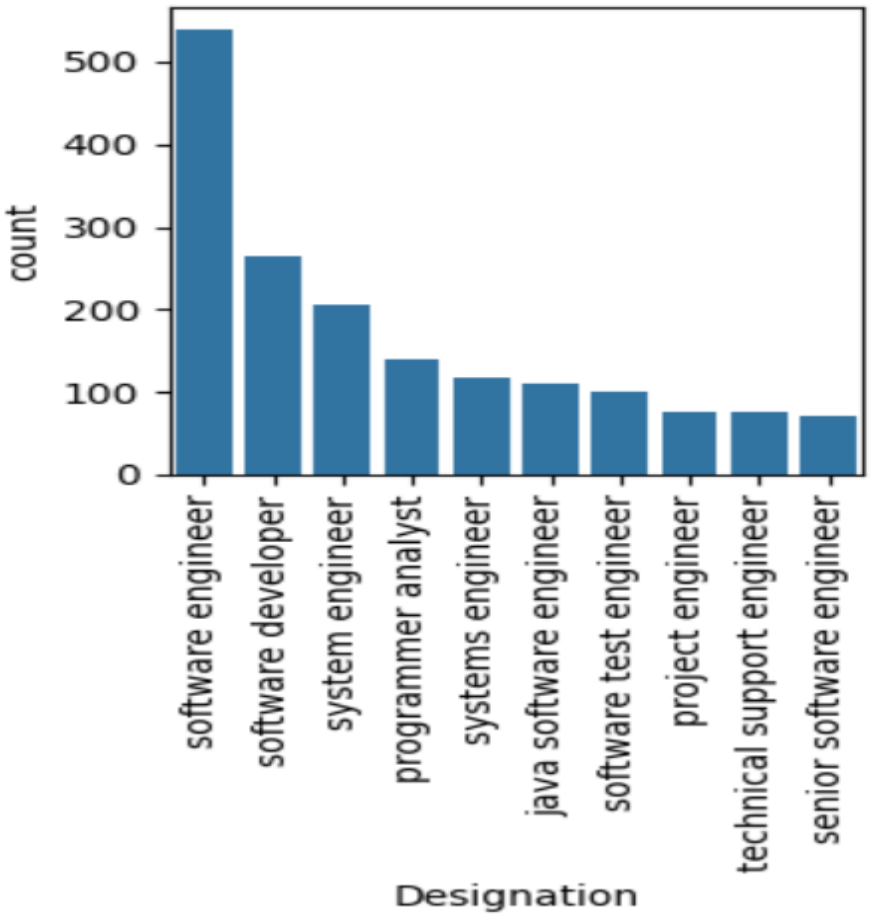
ID', 'Salary', 'DOJ', 'DOL', 'Designation', 'JobCity', 'Gender', 'DOB', '10percentage', '10board', '12graduation', '12percentage',
'12board', 'CollegeID', 'CollegeTier', 'Degree', 'Specialization', 'collegeGPA', 'CollegeCityID', 'CollegeCityTier', 'CollegeState',
'GraduationYear', 'English', 'Logical', 'Quant', 'Domain', 'ComputerProgramming', 'ElectronicsAndSemicon', 'ComputerScience', \
'MechanicalEngg', 'ElectricalEngg', 'TelecomEngg', 'CivilEngg', 'conscientiousness', 'agreeableness', 'extraversion', 'nueroticism',
'openess_to_experience', 'total'

Data Info and Data cleaning

- There are 39 columns and 3998 rows
- There are no null values and duplicated columns and rows
- Mostly all the features are in specific data types
- But some of the features like DOJ ,DOB are in object data type by using pandas to_datatime I change them into datetime datatype
- There is no outliers

Univariate Analysis

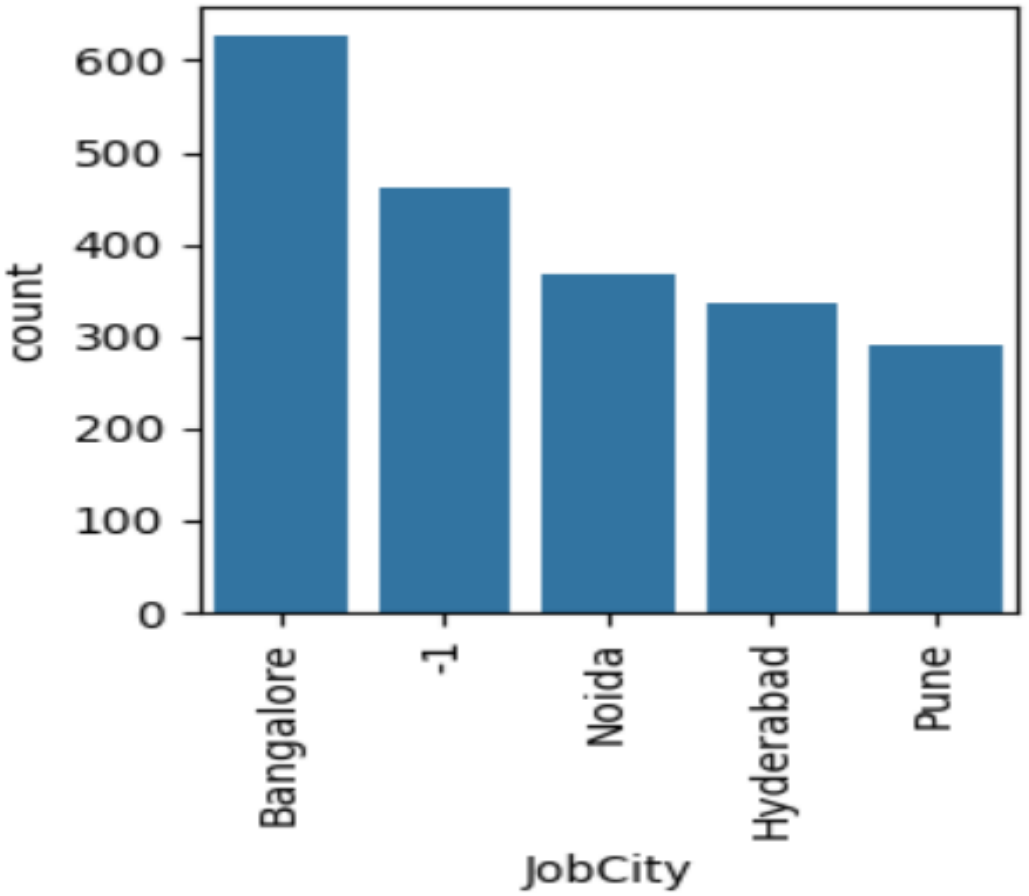
These plot shows the top Designations count



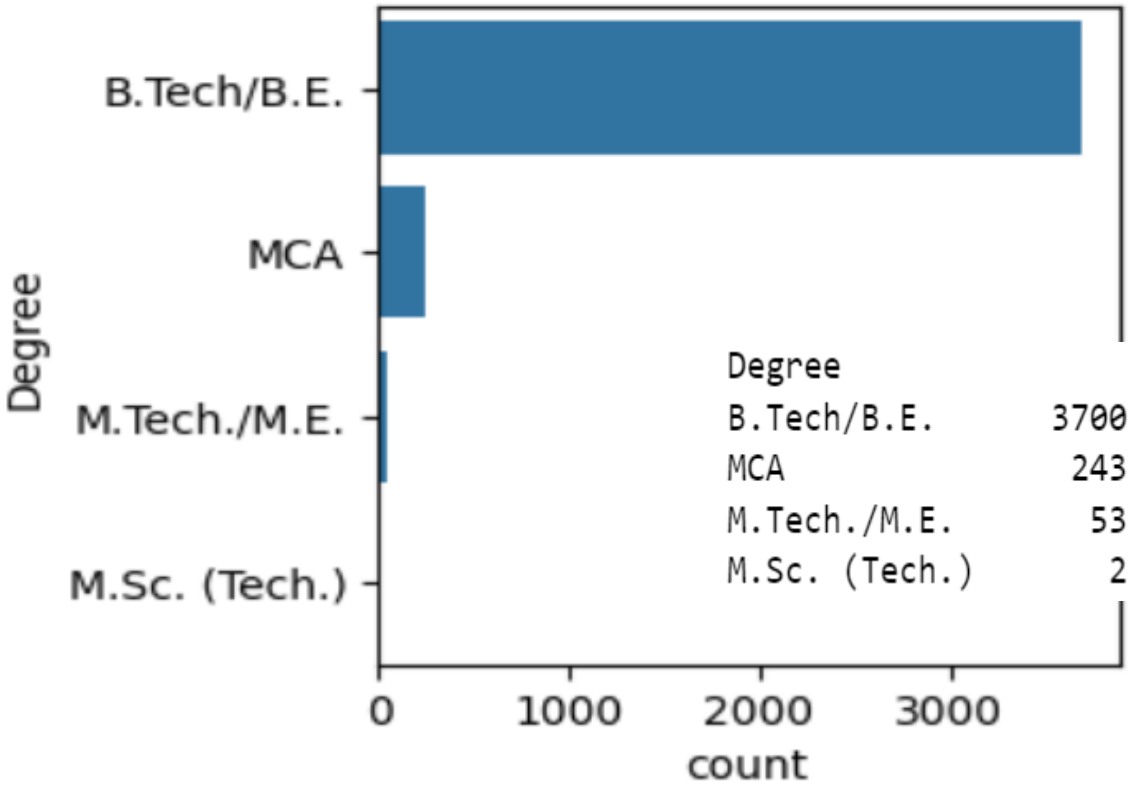
Designation	
software engineer	539
software developer	265
system engineer	205
programmer analyst	139
systems engineer	118
java software engineer	111
software test engineer	100
project engineer	77
technical support engineer	76
senior software engineer	72

INNOMATICS RESEARCH LABS

From this plot we can understand what are the top cities in which most of them do their jobs

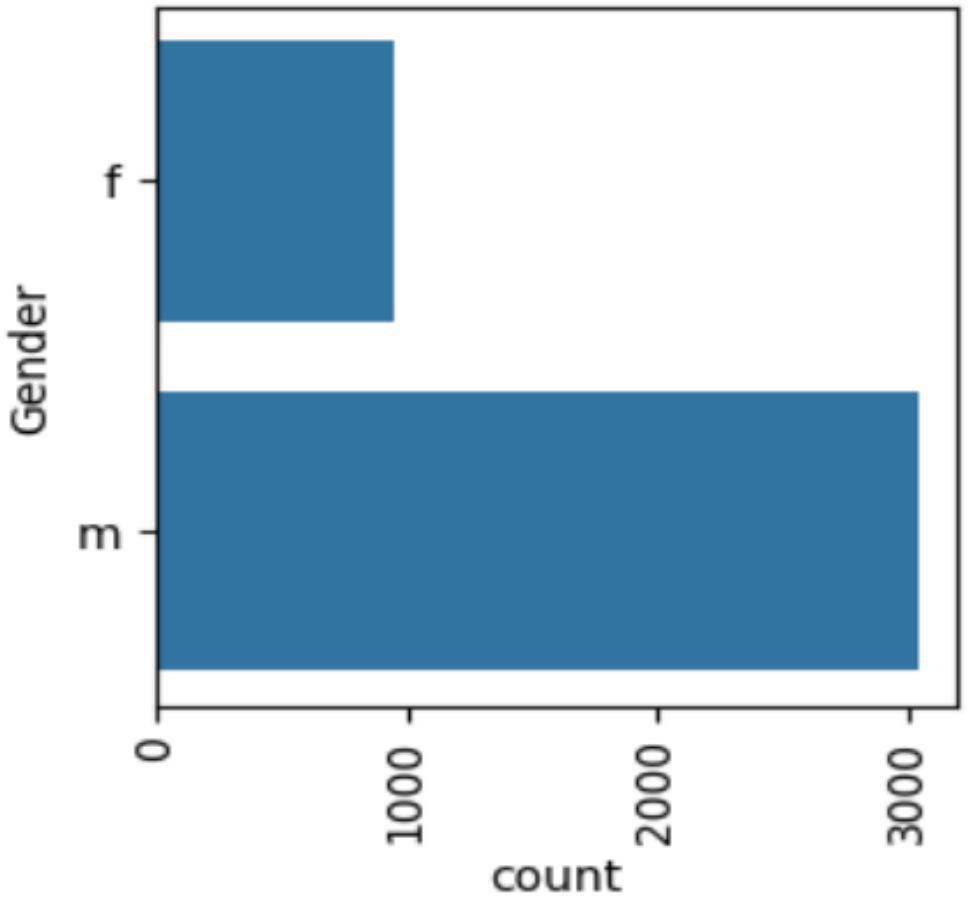


In this plot we can find most job persons from which background

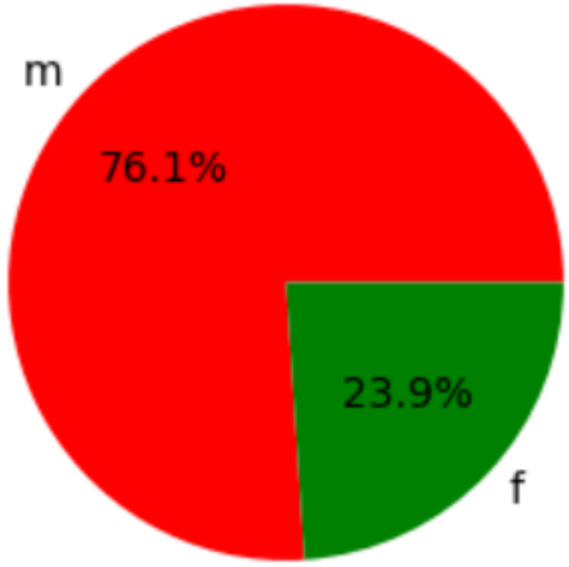


We can analyse the gender contribution

There are more Males than Females and we see in the below pie chart what is the percentage contribution of males and Females

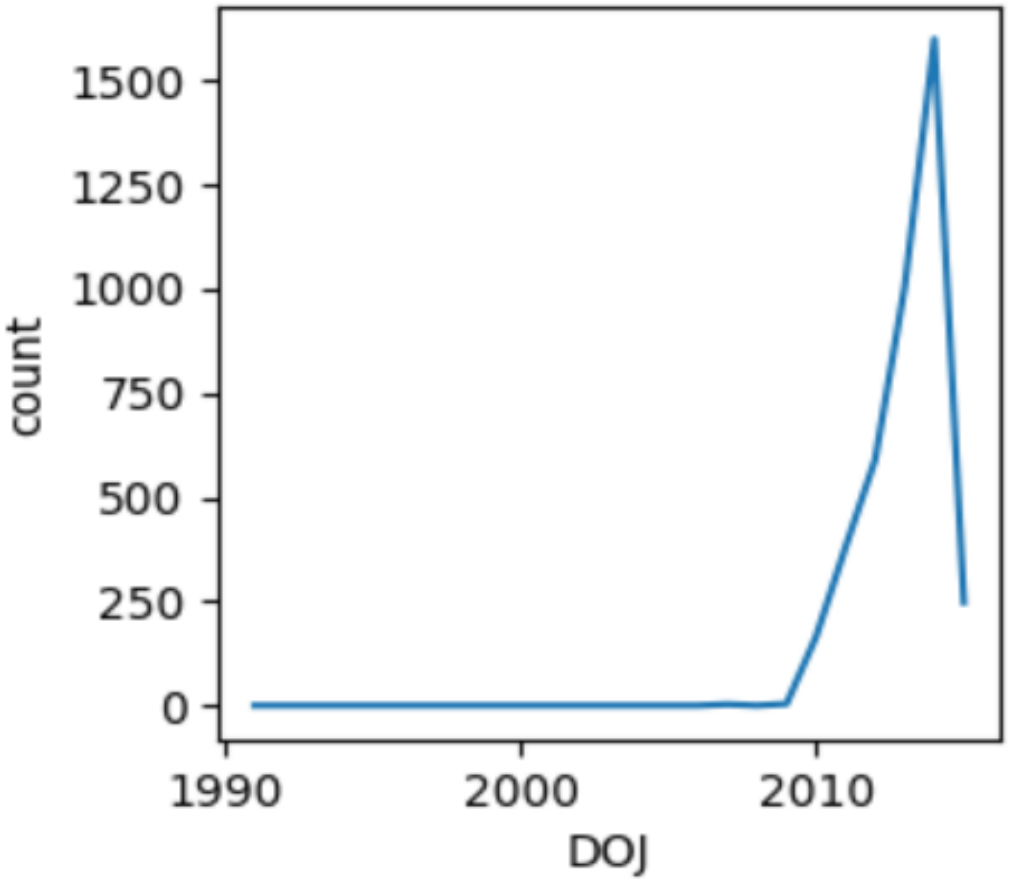



```
Gender
m     3041
f      957
```



INNOMATICS RESEARCH LABS

Count of Joining in various years



Year of joining - count

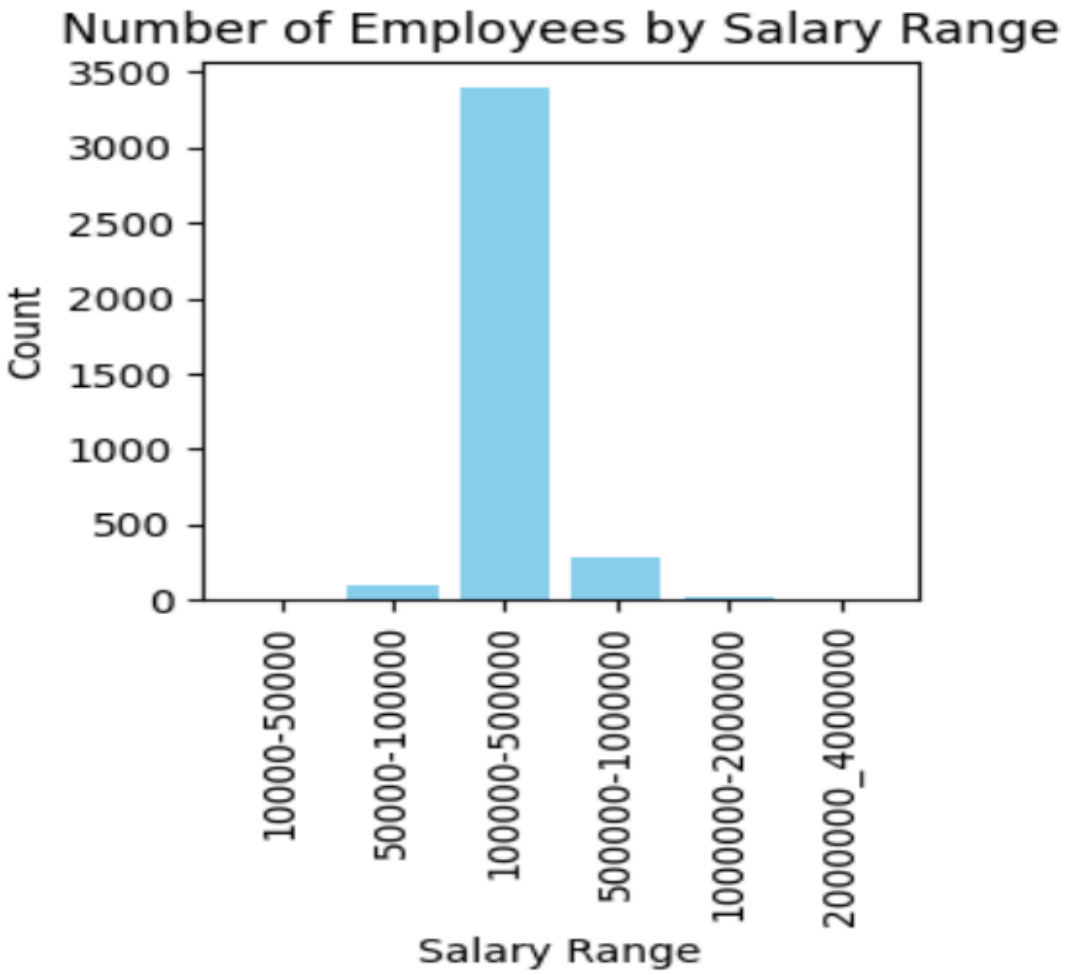
2014	1596
2013	1004
2012	590
2011	381
2015	248
2010	166
2009	5
2007	4
2004	1
2008	1

Pay Scale Range and their count

Salary

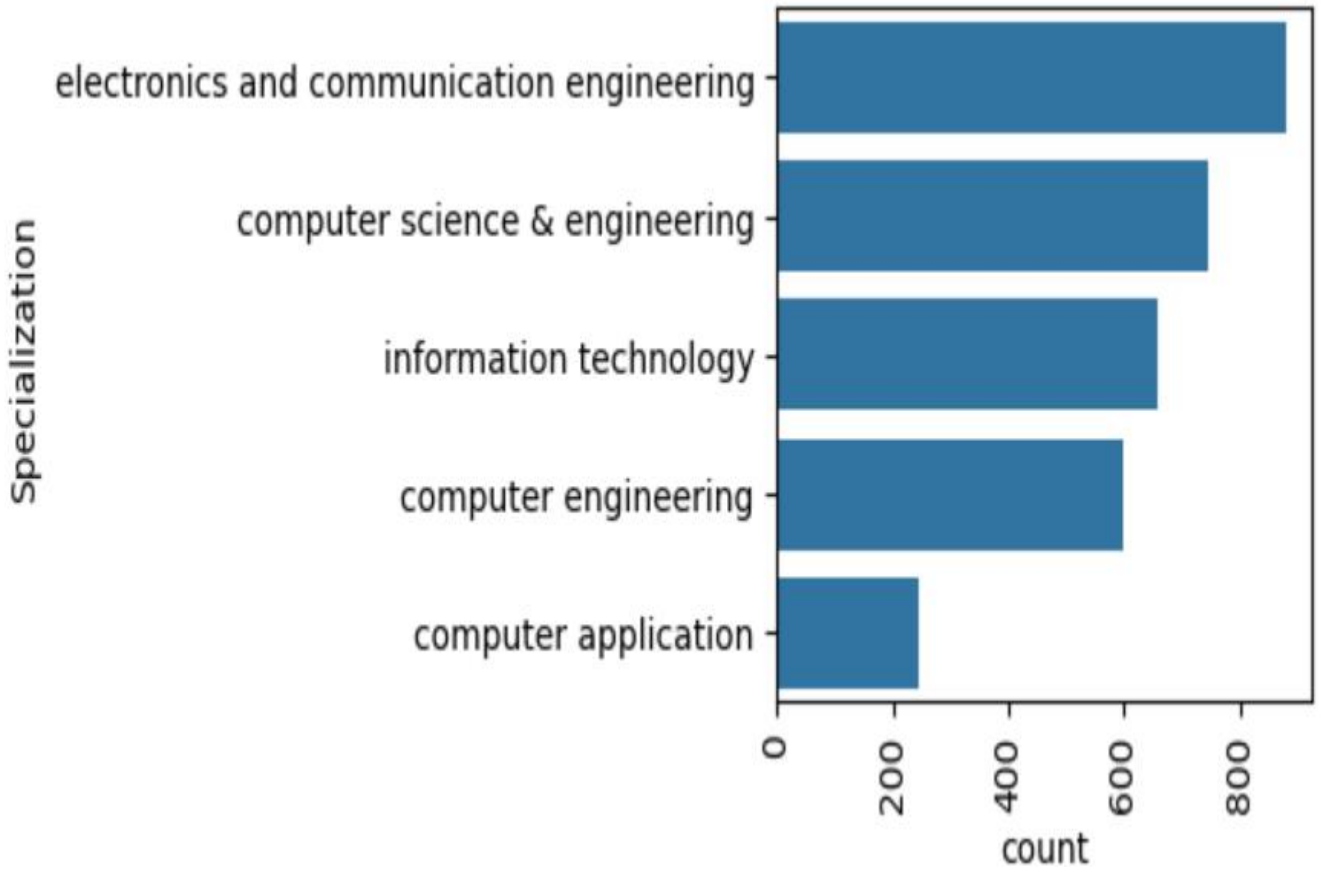
- Min 35000.0
- Max 4000000
- Mean 307699.84
- Senior Software engineer and automation engineer have salary of 4000000 this is the highest salary

Range	Count
10000-50000	9
50000-100000	101
100000-500000	3394
500000-1000000	276
1000000-2000000	23
2000000_4000000	7



INNOMATICS RESEARCH LABS

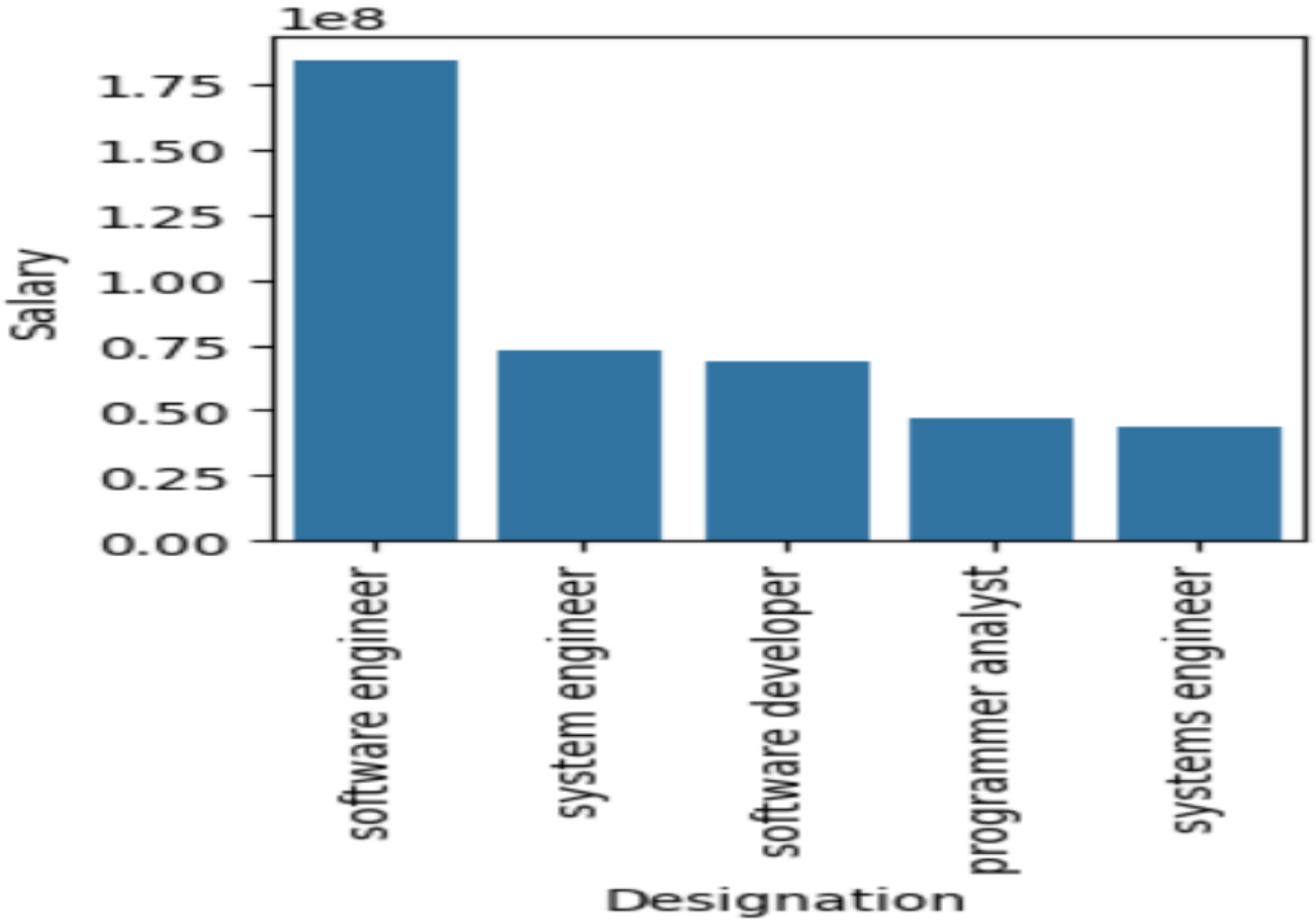
Top 5 specializations doing more jobs



Specialization	count
electronics and communication engineering	880
computer science & engineering	744
information technology	660
computer engineering	600
computer application	244
mechanical engineering	201

Bi-variate Analysis

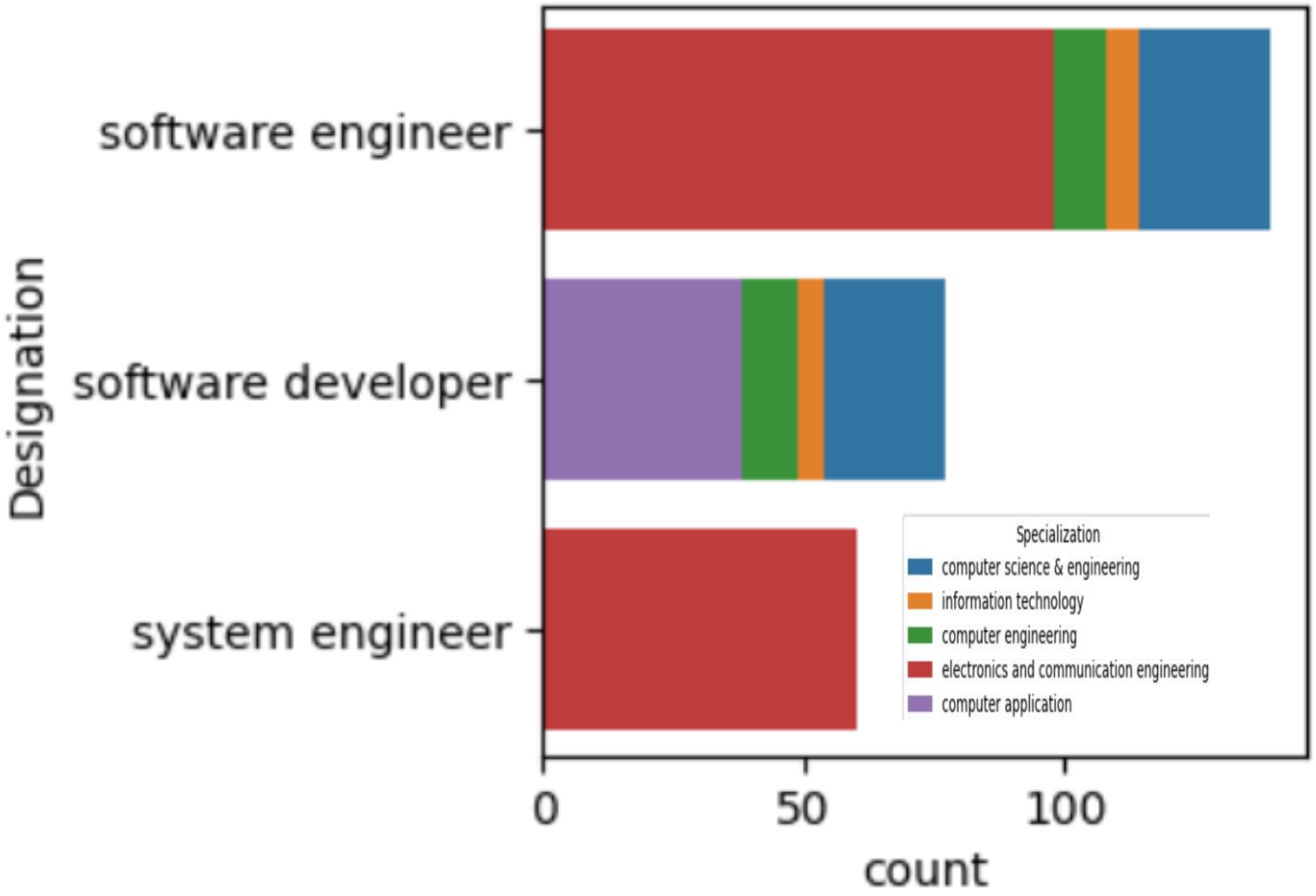
Top Designations that having high paying scale



Designation	Salary
software engineer	183915000.0
system engineer	72580000.0
software developer	68470000.0
programmer analyst	47230000.0
systems engineer	43585000.0

INNOMATICS RESEARCH LABS

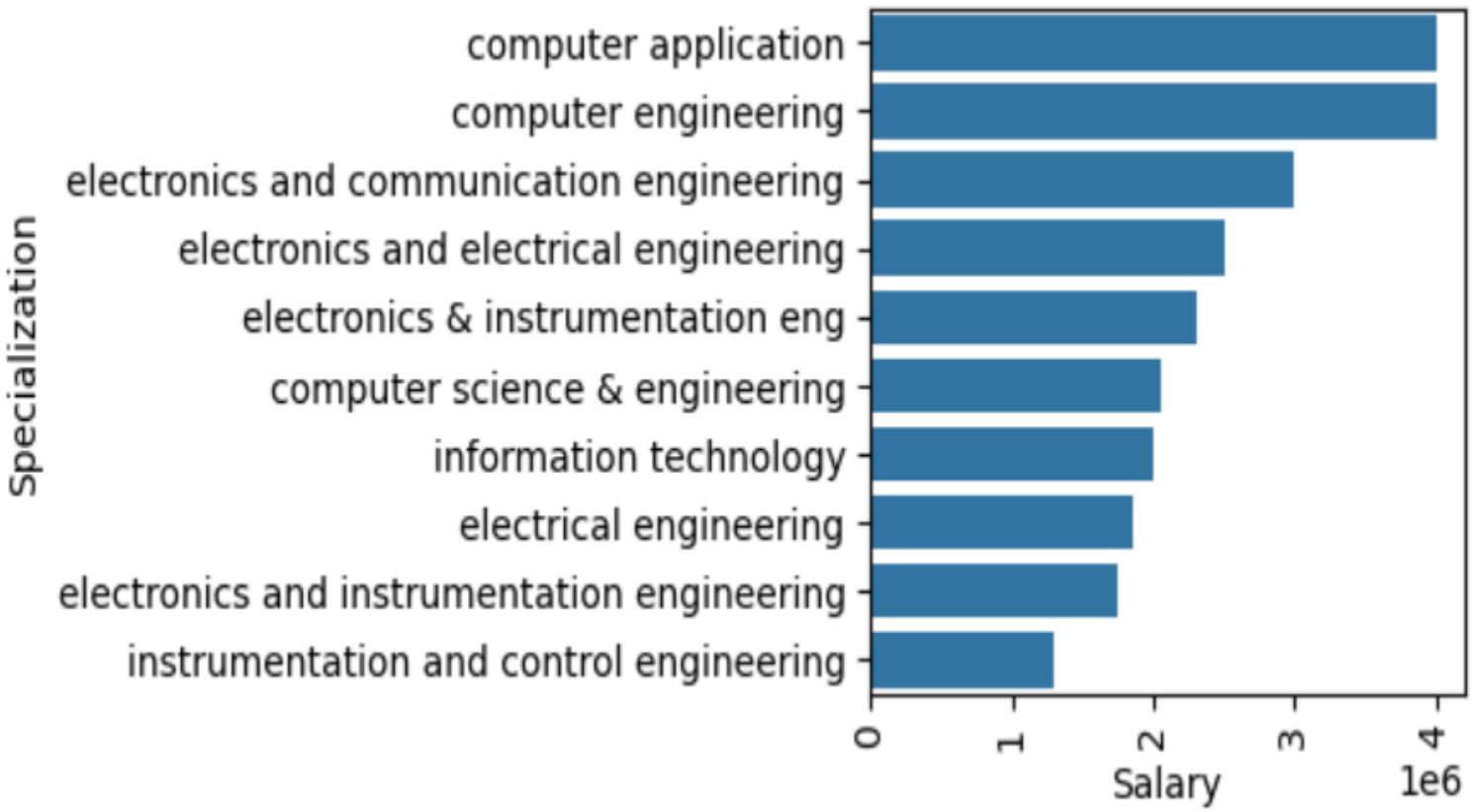
Top pic of different specializations from different Designations



Designation	Specialization	count
software engineer	computer science & engineering	139
	information technology	114
	computer engineering	108
	electronics and communication engineering	98
software developer	computer science & engineering	77
system engineer	electronics and communication engineering	60
software developer	information technology	54
	computer engineering	49
system engineer	computer engineering	41
software developer	computer application	38

INNOMATICS RESEARCH LABS

These are the different Specializations with high paying scale range



Specialization	Salary
computer application	4000000.0
computer engineering	4000000.0
electronics and communication engineering	3000000.0
electronics and electrical engineering	2500000.0
electronics & instrumentation eng	2300000.0
computer science & engineering	2050000.0
information technology	2000000.0
electrical engineering	1860000.0
electronics and instrumentation engineering	1745000.0
instrumentation and control engineering	1300000.0

INNOMATICS RESEARCH LABS

Summary and Conclusion

- Software engineer and soft developer designation are the top two having most members working
- Most of them are working in Bangalore city
- Most of them are from B-tech/B.E background
- From 2004-2014 the jobs are increasing rapidly
- Most of them are male compared to female
- Electronics and communication engineering and Computer science specializations are the highest among other specialization
- Mostly the pay scale range between 100000 to 500000
- The contribution of soft engineers having high salary among other designation
- Computer application and computer engineering are the top two high paying specialization.

THANK YOU

INNOMATICS RESEARCH LABS