Exploratory Data Analysis On AMEO Dataset

INNOMATICS
RESEARCH LABS

➡ About me

✓ Vijay sada
✓ B-tech in Electronics and Communications Engineering.
✓ Fresher

Data Description:

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.

Objective:

The main focus will be on providing insights based on the target column "salary" using the features like designation, gender, specialization, etc.
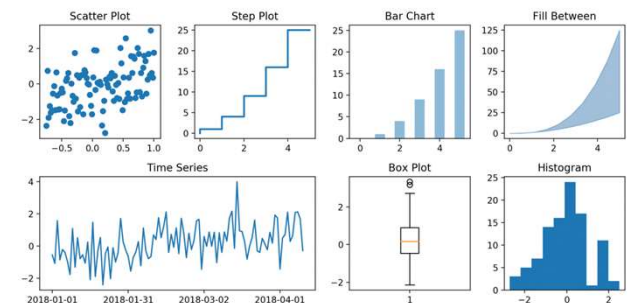
DATA CLEANING:

- Libraries Used : pandas, NumPy.

DATA ANALYSIS:

- Libraries Used : pandas, NumPy.

DATA VISUALIZATION:

- Libraries Used : pandas, matplotlib, seaborn.

RAW DATA:

	Unnamed: 0	ID	Salary	DOJ	DOL	Designation	JobCity	Gender	DOB	10percentage	...	ComputerScience	MechanicalEngg	ElectricalEngg
0	train	203097	420000.0	6/1/12 0:00	present	senior quality engineer	Bangalore	f	2/19/90 0:00	84.30	...	-1	-1	-1
1	train	579905	500000.0	9/1/13 0:00	present	assistant manager	Indore	m	10/4/89 0:00	85.40	...	-1	-1	-1
2	train	810601	325000.0	6/1/14 0:00	present	systems engineer	Chennai	f	8/3/92 0:00	85.00	...	-1	-1	-1
3	train	267447	1100000.0	7/1/11 0:00	present	senior software engineer	Gurgaon	m	12/5/89 0:00	85.60	...	-1	-1	-1
4	train	343523	200000.0	3/1/14 0:00	3/1/15 0:00	get	Manesar	m	2/27/91 0:00	78.00	...	-1	-1	-1
5	train	1027655	300000.0	6/1/14 0:00	present	system engineer	Hyderabad	m	7/2/92 0:00	89.92	...	407	-1	-1
6	train	947847	300000.0	8/1/14 0:00	5/1/15 0:00	java software engineer	Banglore	m	2/1/93 0:00	86.08	...	346	-1	-1
7	train	912934	400000.0	7/1/14 0:00	7/1/15 0:00	mechanical engineer	Bangalore	m	5/27/92 0:00	92.00	...	-1	469	-1
8	train	552574	600000.0	7/1/13 0:00	present	electrical engineer	Noida	m	9/17/91 0:00	90.00	...	-1	-1	-1
9	train	1203363	230000.0	7/1/14 0:00	present	project engineer	Kolkata	m	6/13/93 0:00	77.00	...	-1	-1	-1

CLEANED DATA:

|	ID	Salary	DOJ	DOL	Designation	JobCity	Gender	DOB	10percentage	12graduation	...	ElectricalEngg	TelecomEngg	CivilEngg	conscie
0	203097	420000.0	2012-06-01	2015-12-31	senior quality engineer	bangalore	F	1990-02-19	84.30	2007	...	NaN	NaN	NaN	
1	579905	500000.0	2013-09-01	2015-12-31	assistant manager	indore	M	1989-10-04	85.40	2007	...	NaN	NaN	NaN	
2	810601	325000.0	2014-06-01	2015-12-31	systems engineer	chennai	F	1992-08-03	85.00	2010	...	NaN	NaN	NaN	
3	267447	1100000.0	2011-07-01	2015-12-31	senior software engineer	gurgaon	M	1989-12-05	85.60	2007	...	NaN	NaN	NaN	
4	343523	200000.0	2014-03-01	2015-03-01	get	manesar	M	1991-02-27	78.00	2008	...	NaN	NaN	NaN	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
3953	47916	280000.0	2011-10-01	2012-10-01	software engineer	delhi	M	1987-04-15	52.09	2006	...	NaN	NaN	NaN	
3954	752781	100000.0	2013-07-01	2013-07-01	technical writer	hyderabad	F	1992-08-27	90.00	2009	...	NaN	NaN	NaN	
3955	355888	320000.0	2013-07-01	2015-12-31	associate software engineer	bangalore	M	1991-07-03	81.86	2008	...	NaN	NaN	NaN	
3956	947111	200000.0	2014-07-01	2015-01-01	software developer	bangalore	F	1992-03-20	78.72	2010	...	NaN	NaN	NaN	
3957	324966	400000.0	2013-02-01	2015-12-31	senior systems engineer	chennai	F	1991-02-26	70.60	2008	...	NaN	NaN	NaN	

3958 rows × 35 columns
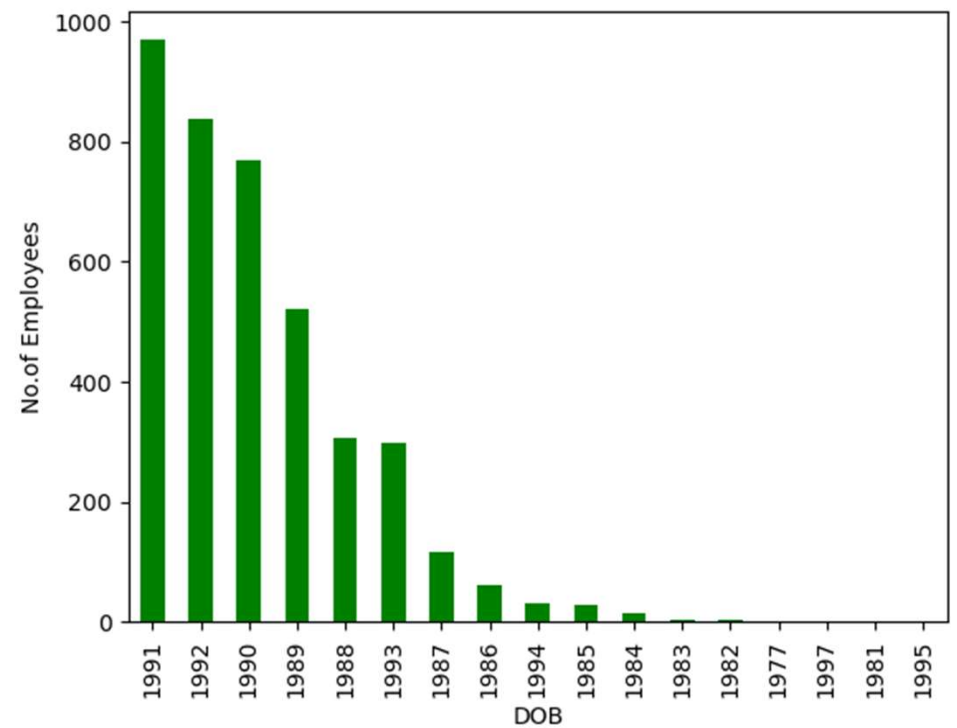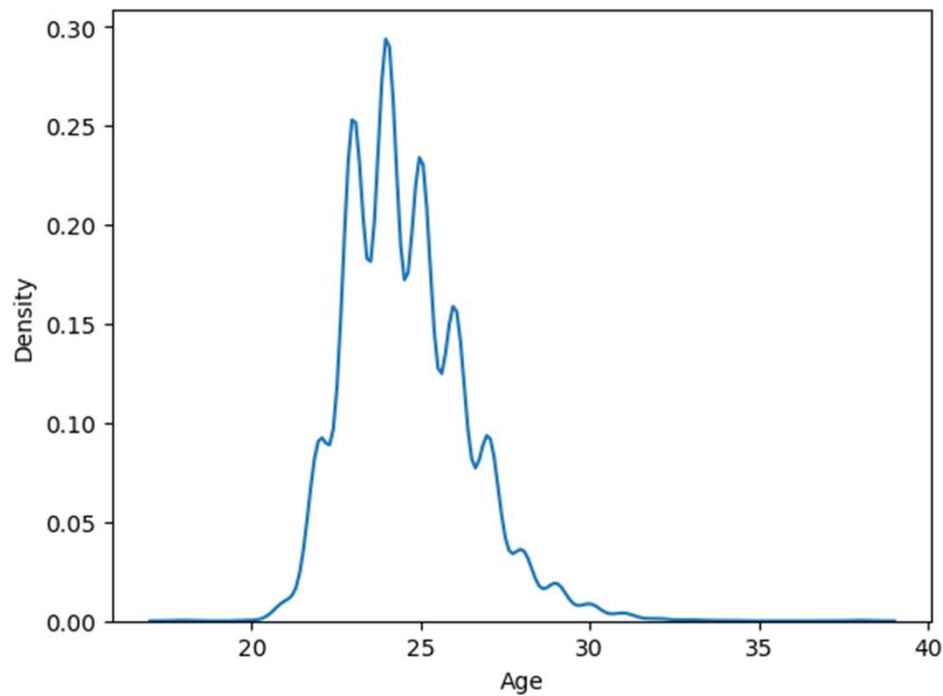
- From the above we can observe that
      1. Most of the employees are in an age range of 21 to 28
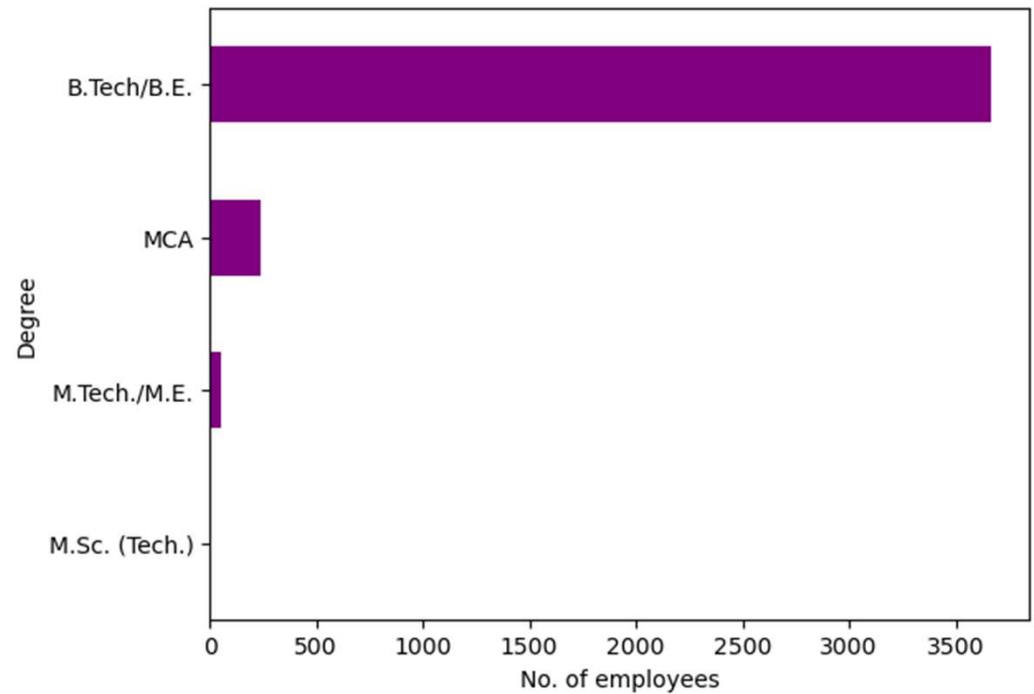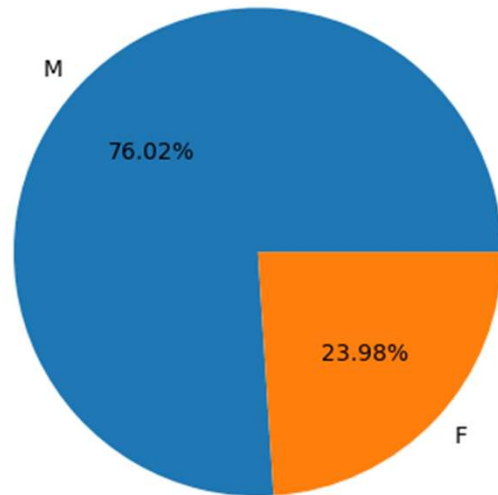   2. Most of the employees were born in between 1990 to 1992
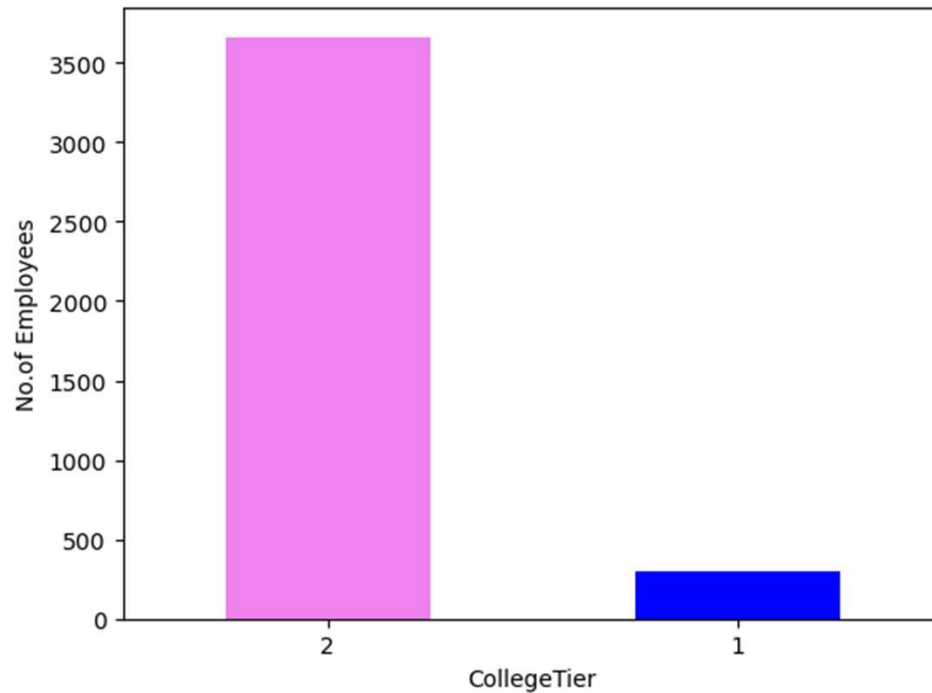
From above we can observe
- Most of the employees are form tier 2 colleges
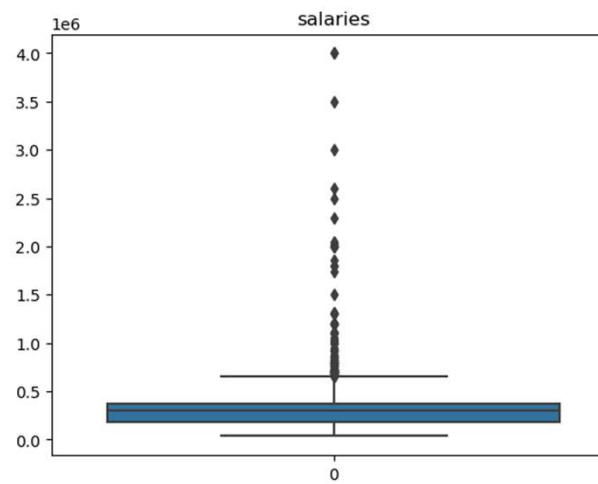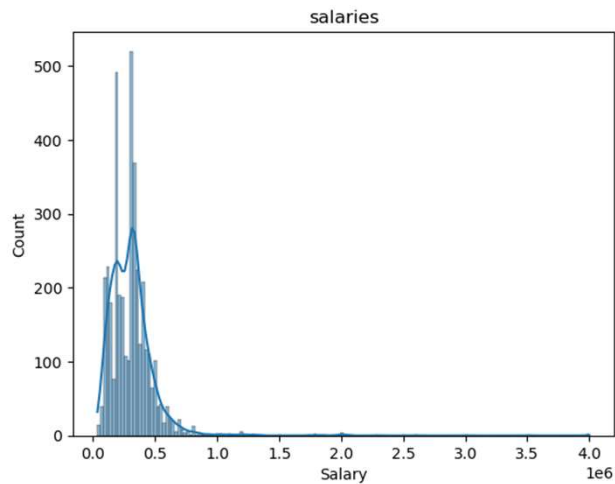- Most of the employees are from the state "Uttar Pradesh"

INNOMATICS
RESEARCH LABS
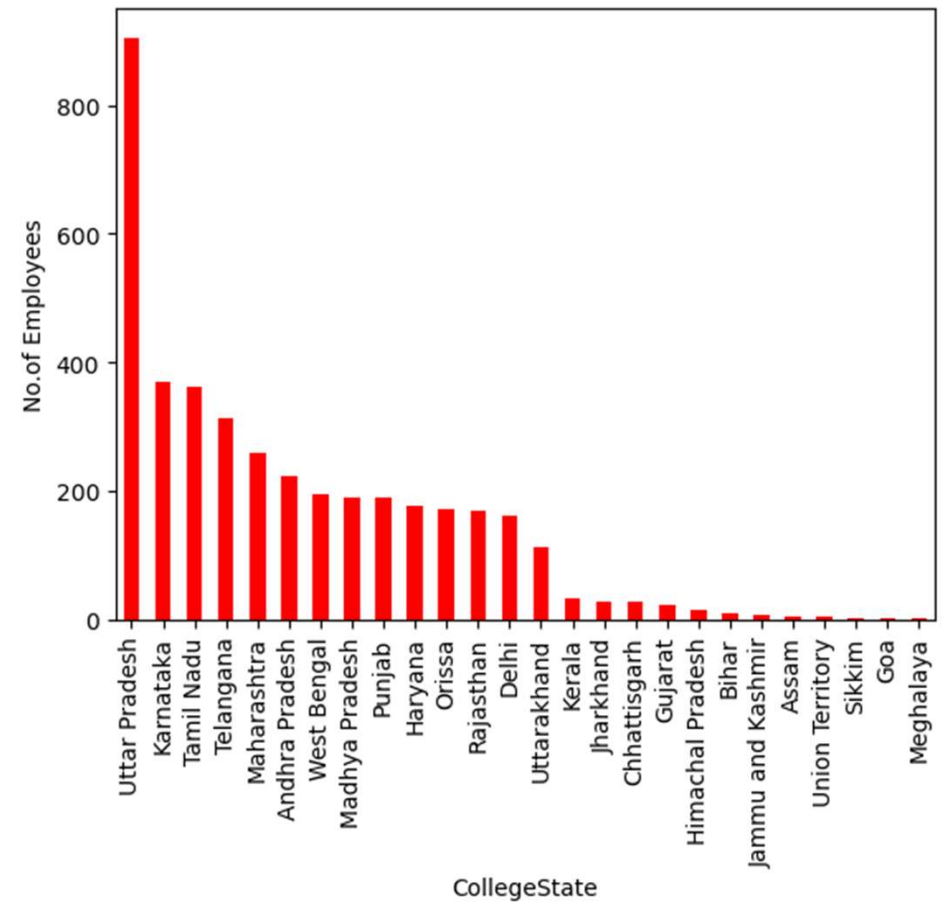
Percentage of Male vs Female employees
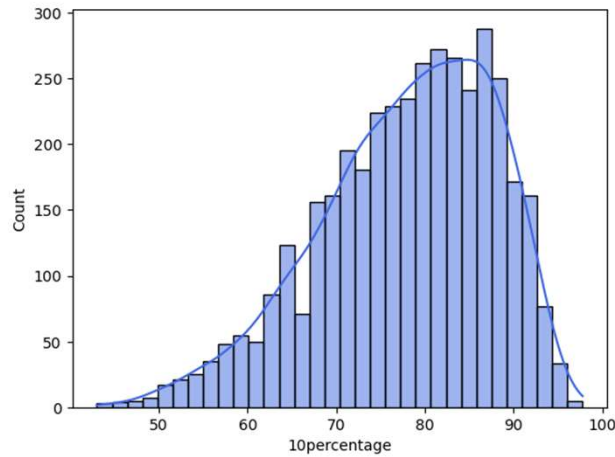
From above we can observe
 - Most of the employees are Male
 - Most of the employees are from the B.Tech/B.E. background

salaries

From above we can observe
 - Most of the employees are having an salary range of 35000 to 600000
 - There exist a lot of outliers in the salary because there might be employees having more experience, skills or higher position in the company.
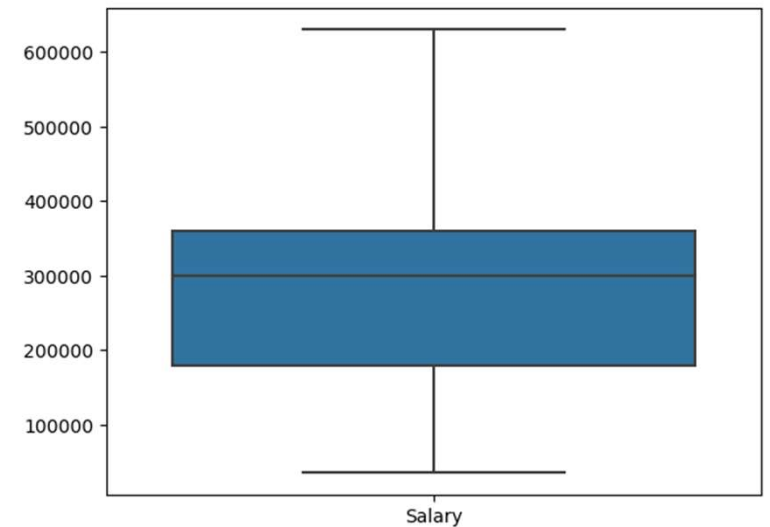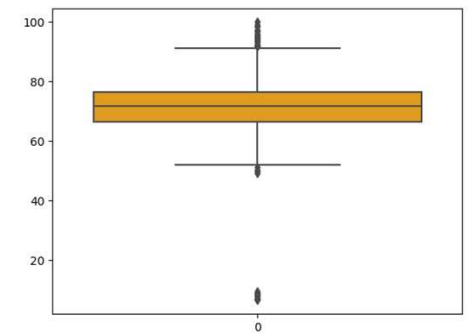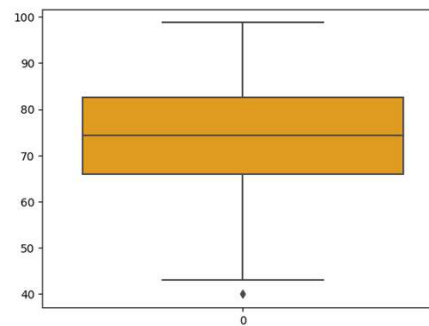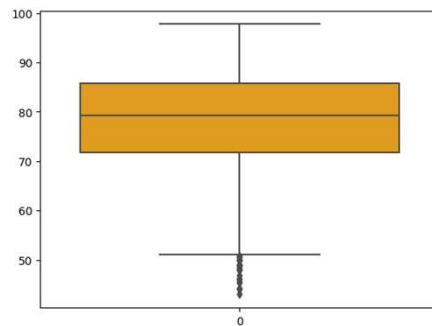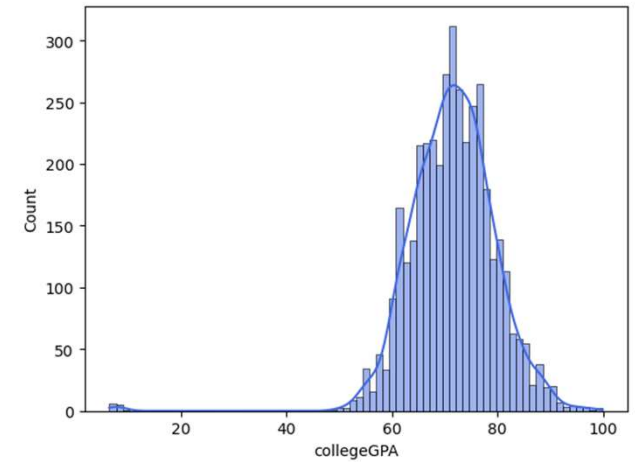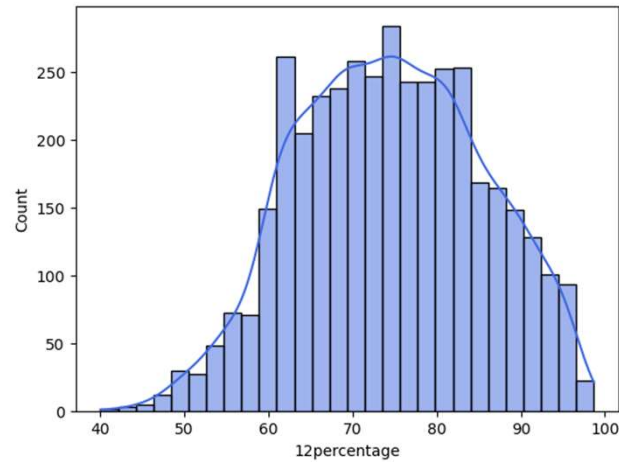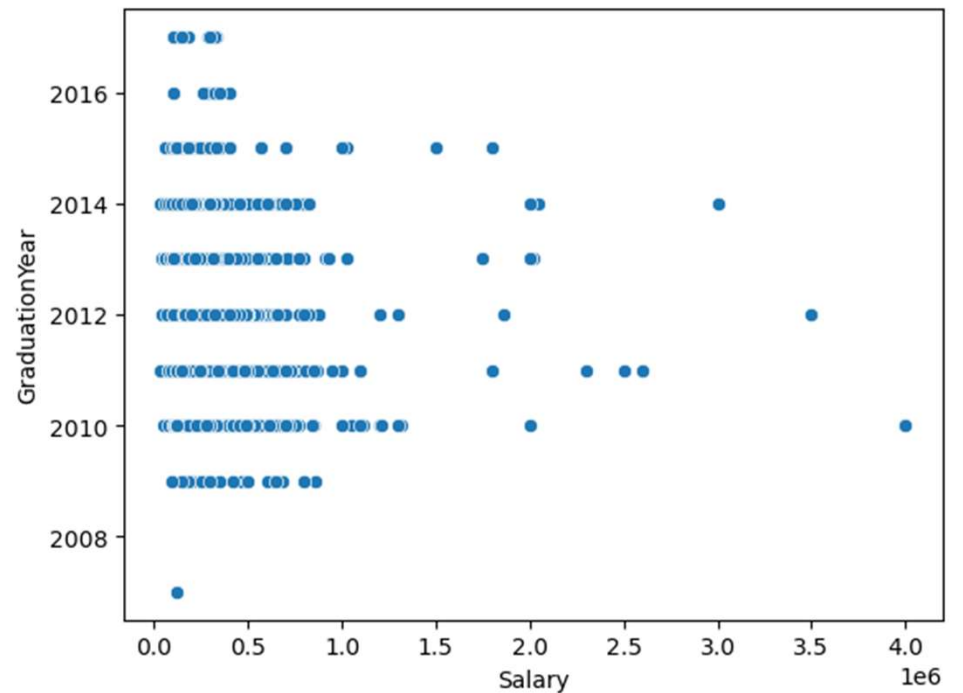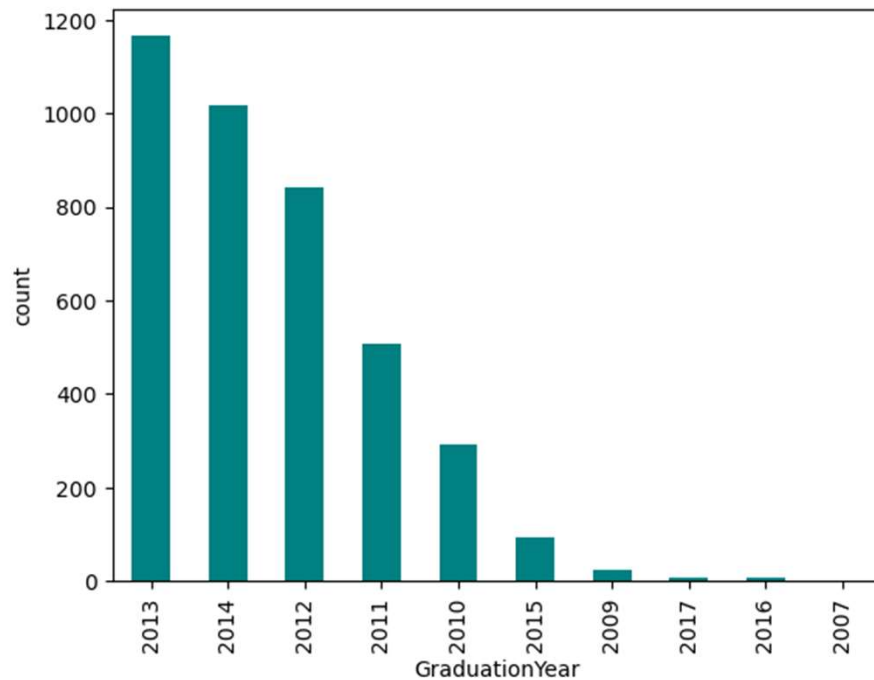
From above we can observe
 - Avg of '10percentage', '12percentage', 'collegeGPA' is around 80,75,70
 - There are some outliers in the data due to some low scorers.

Top 7 Job roles

Top 7 Specializations
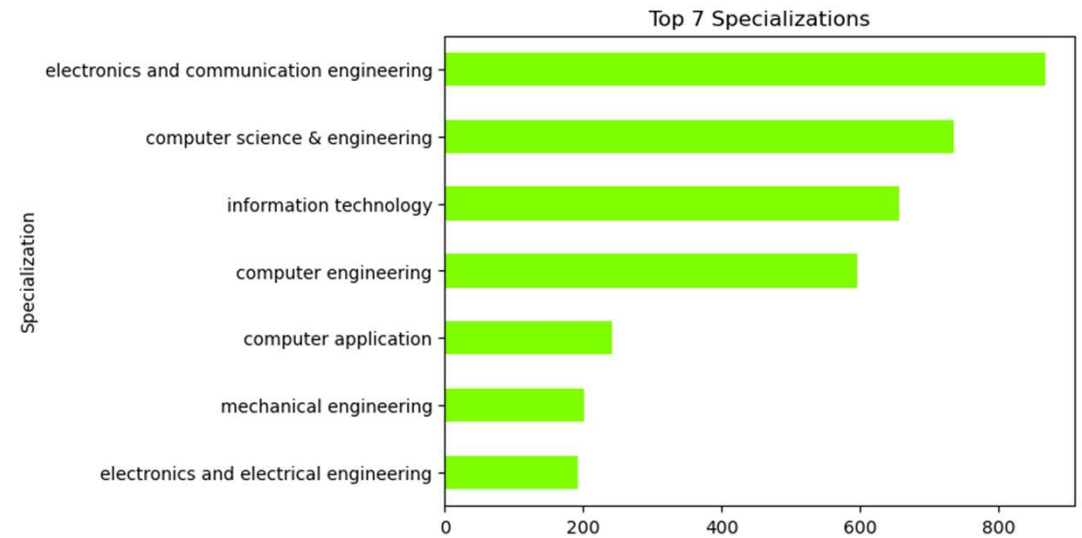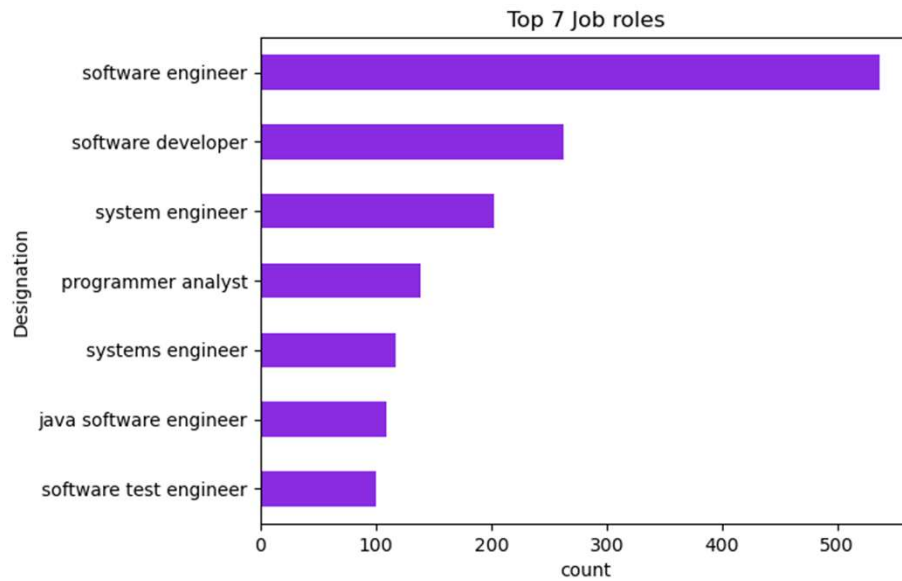
From above we can observe
 - Most of the employees are working under the role software engineer
 - Most of the employees have specialized their graduation in Electronics and Communication Engg.
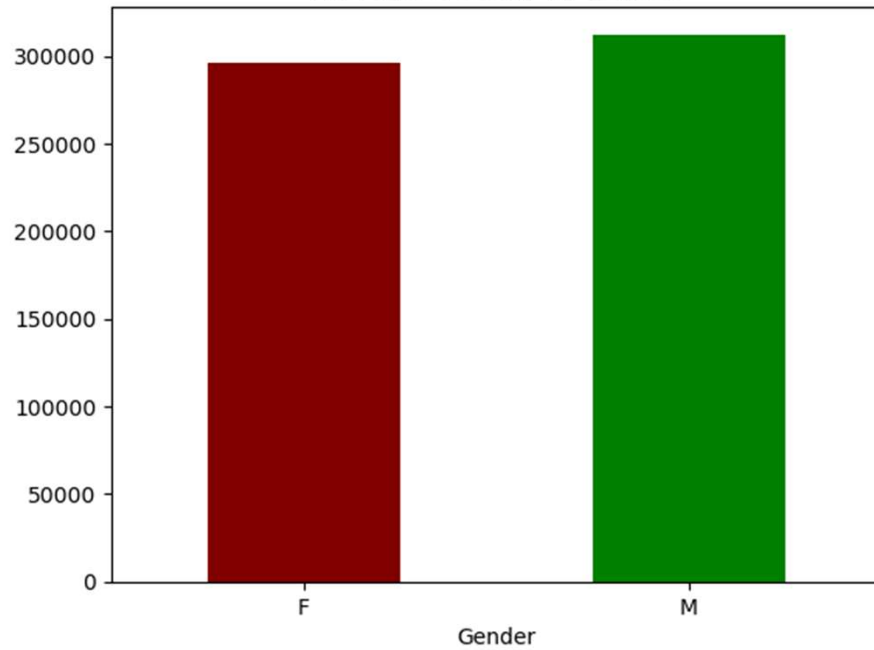
From above we can observe
- Most of the employees have completed their graduation between 2012 to 2014
- Employees who have passed out between 2010 to 2014 seem to have a higher salary as the salary increases along with the expertise and experience.
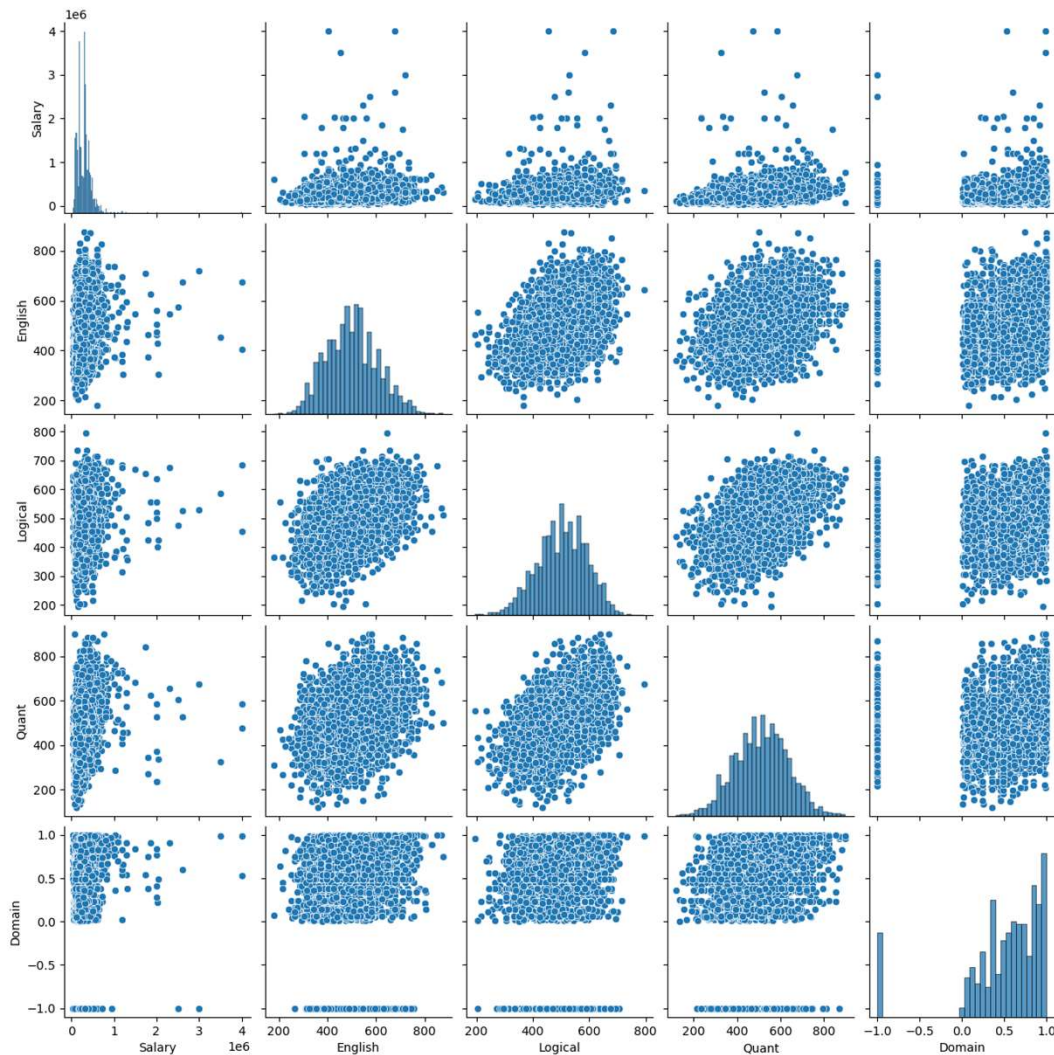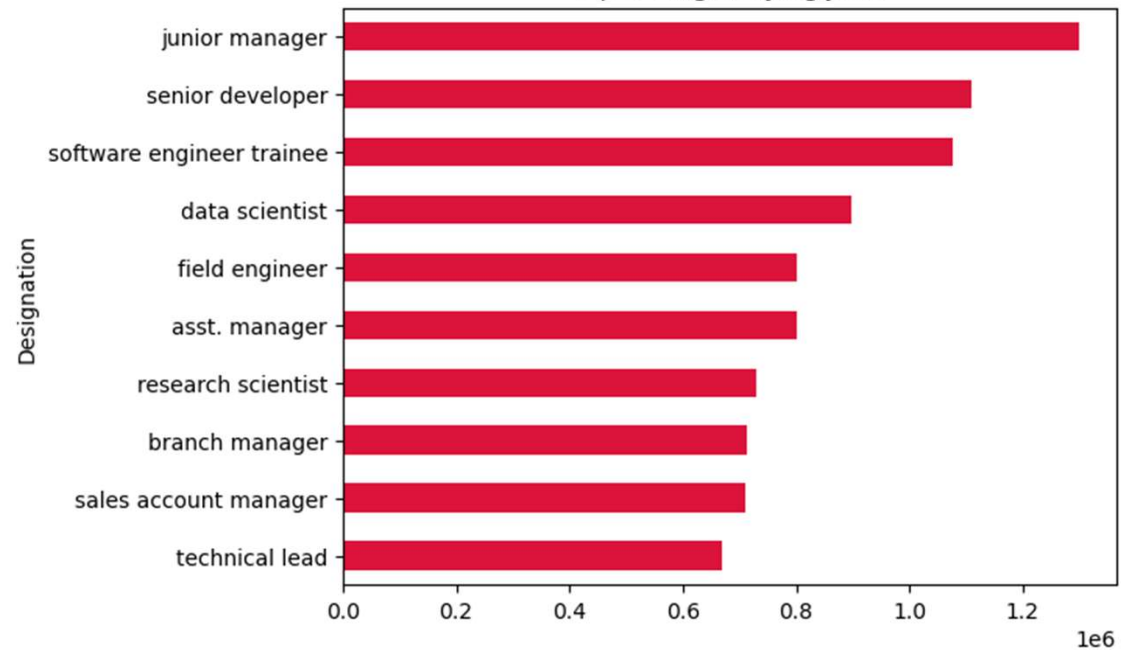
Pair Plot of Salary vs marks
- We can observe that the scores like Logical and quant ability have a positive co-relation .
- the English and logical skills have a positive co-relation.
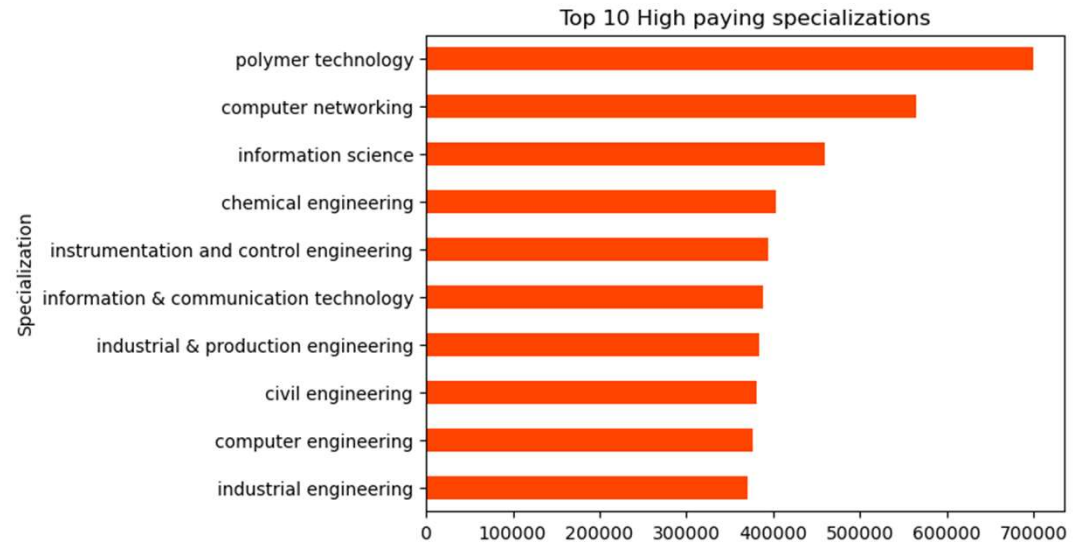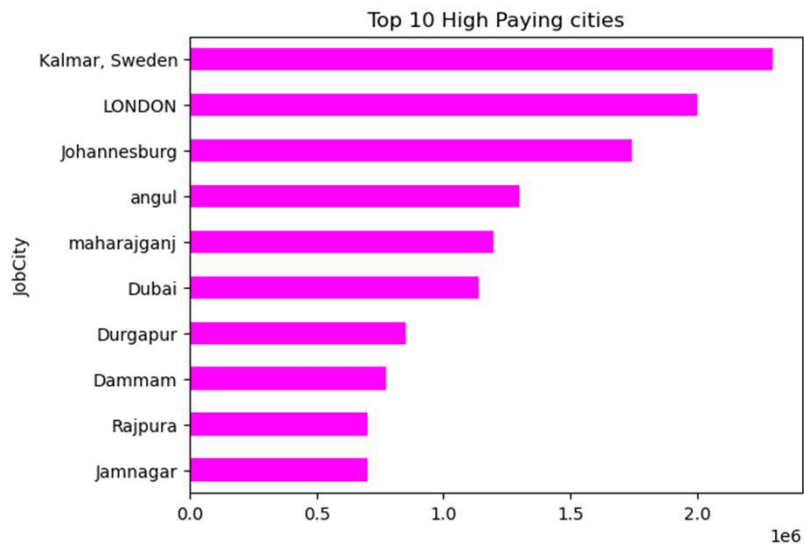
Average of salary by gender

Top 10 High Paying JOBs

From above we can observe
- The average salaries of male and female differ by around 2k.
- On right we have a graph showing the highest paying roles in the market.

INNOMATICS
RESEARCH LABS

Top 10 High Paying cities

Top 10 High paying specializations

From above we can observe
- Most of the highest paying cities are outside India.
- The specialization fields like polymer technology and computer networking offers a higher salary package.
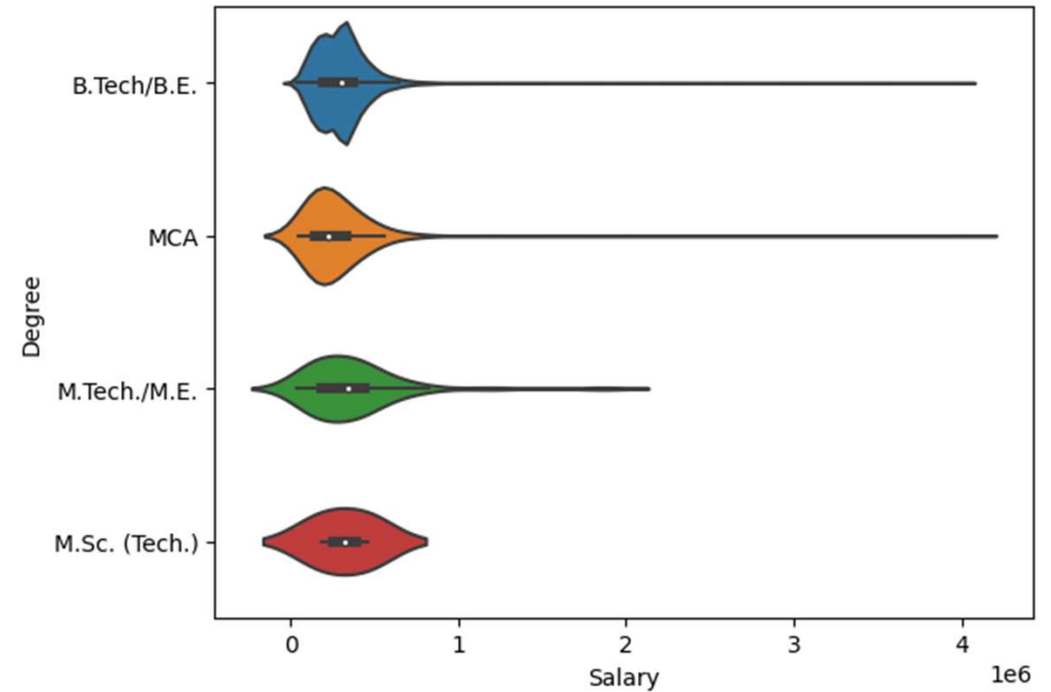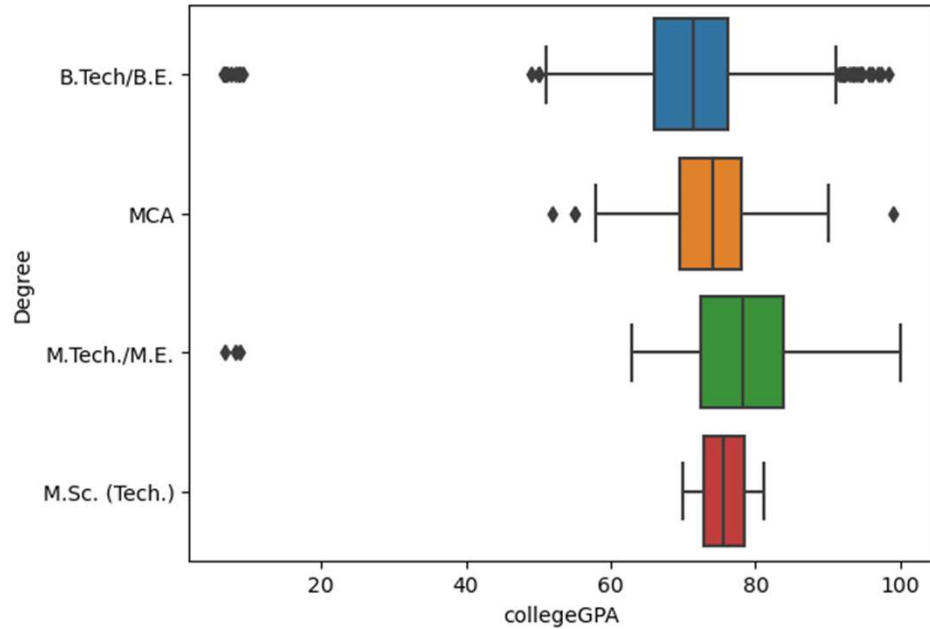
From above we can observe
- From left most of the outliers are in the B.tech degree as there might be a input error or because of different grading system maybe.
- The people from B.tech and MCA have a higher max salary compared to the rest degrees.
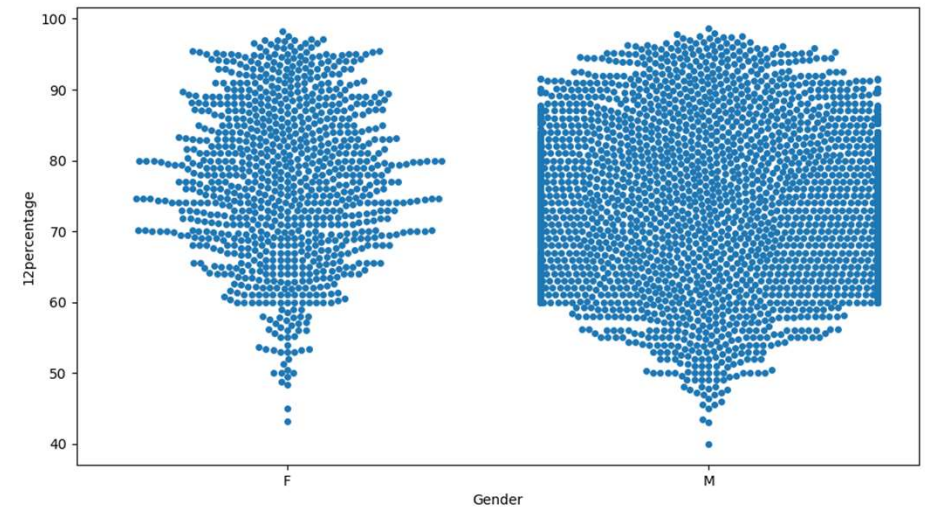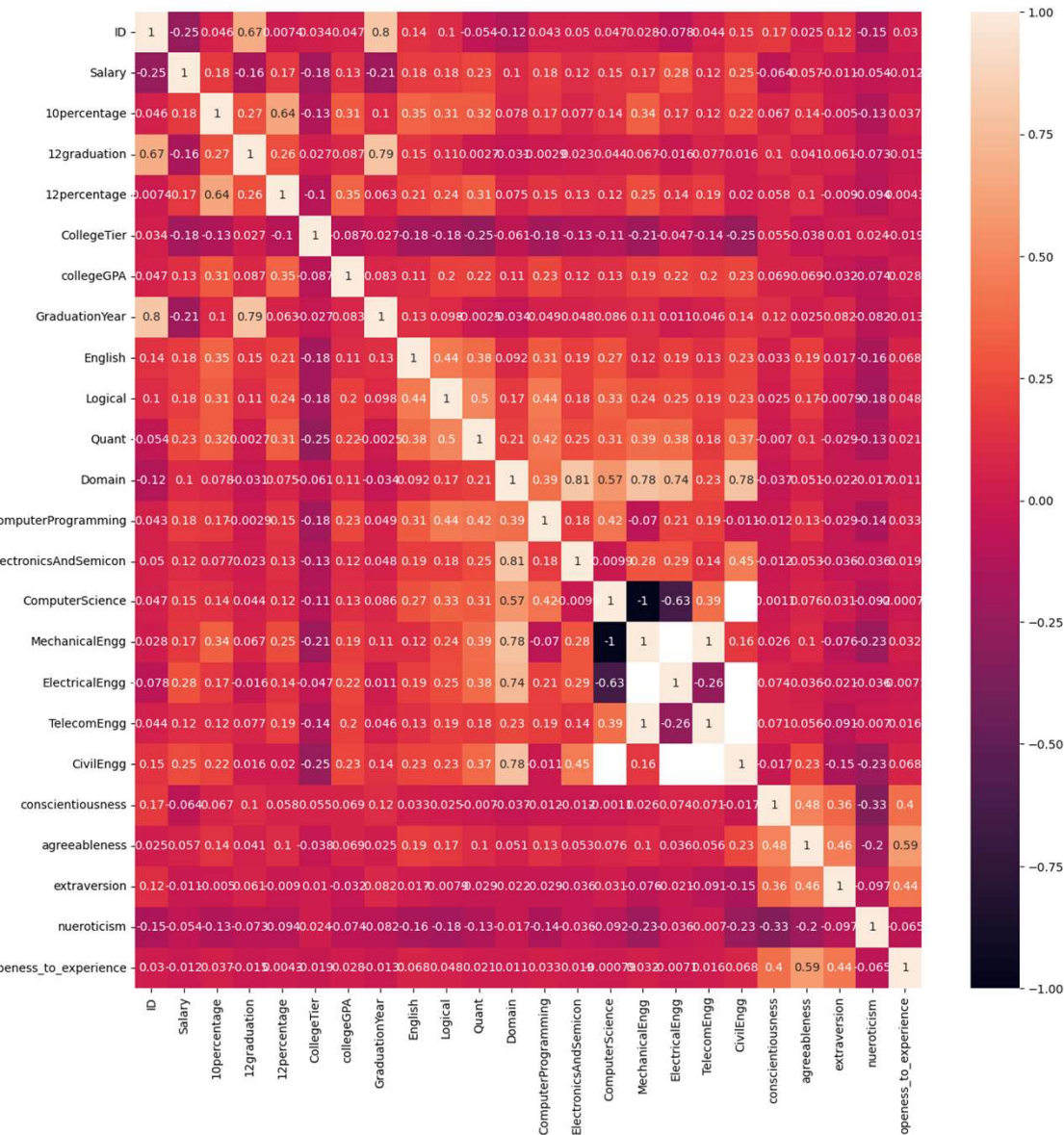
From above we can observe
 - Most of the columns are slightly correlated.
 - some columns are totally unrelated.
 - domain column has strong relation with their exam specialization.
 - Most of the marks in the 12percentage are between 60 to 90 range.
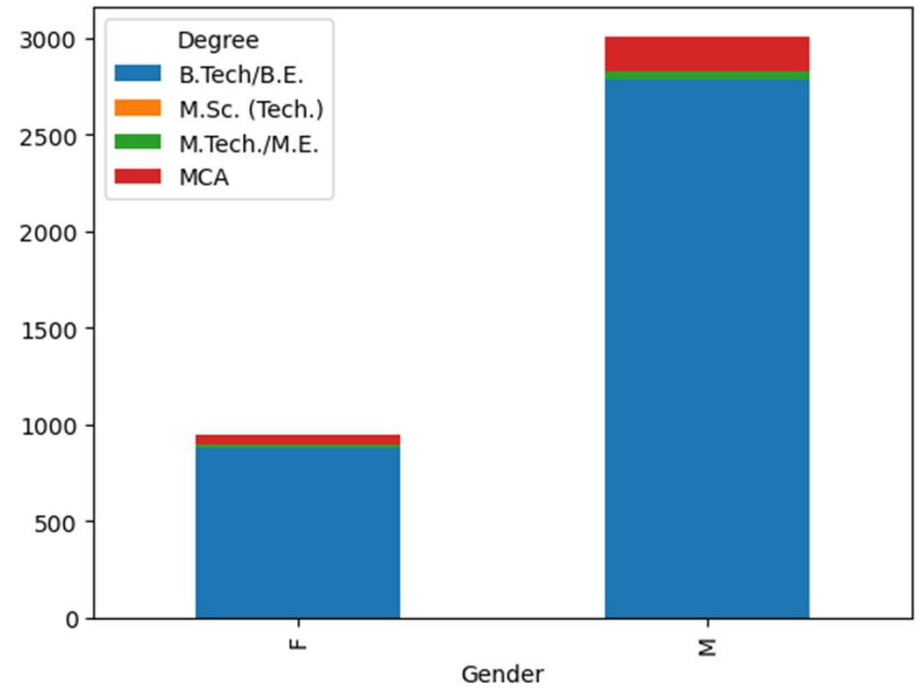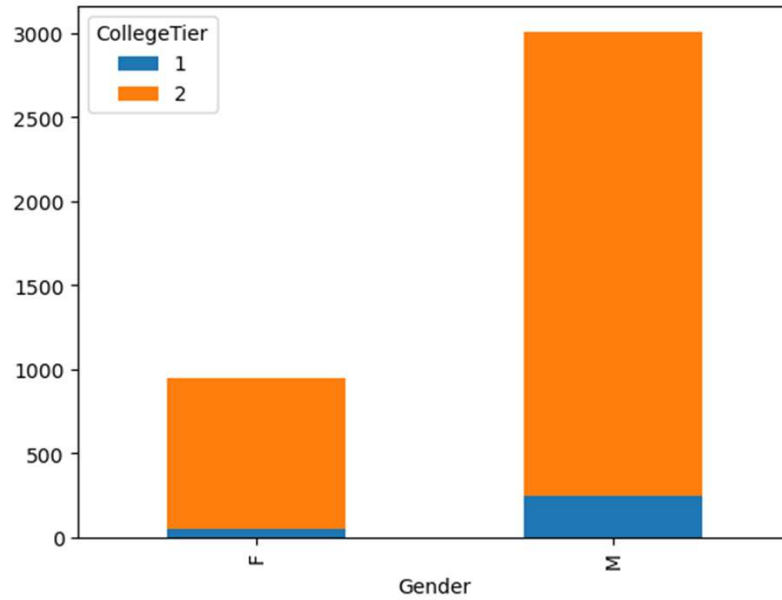
From above we can observe
- Most of the male and female are from tier 2 colleges and most of them are from B.tech background.
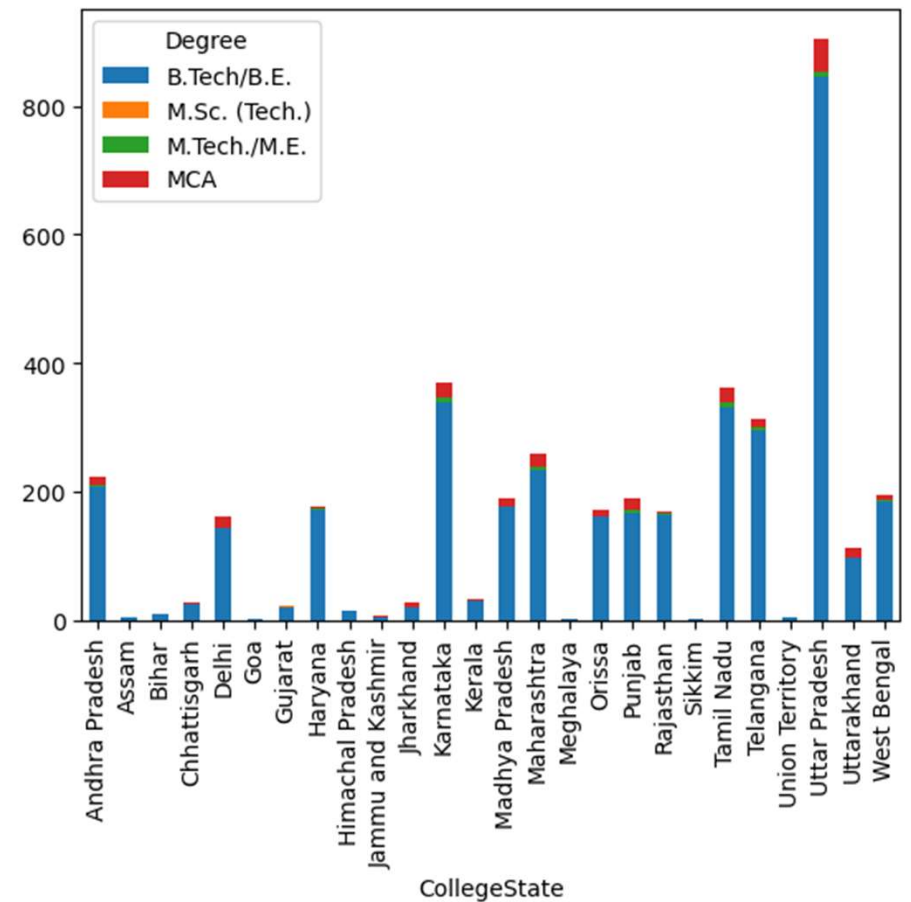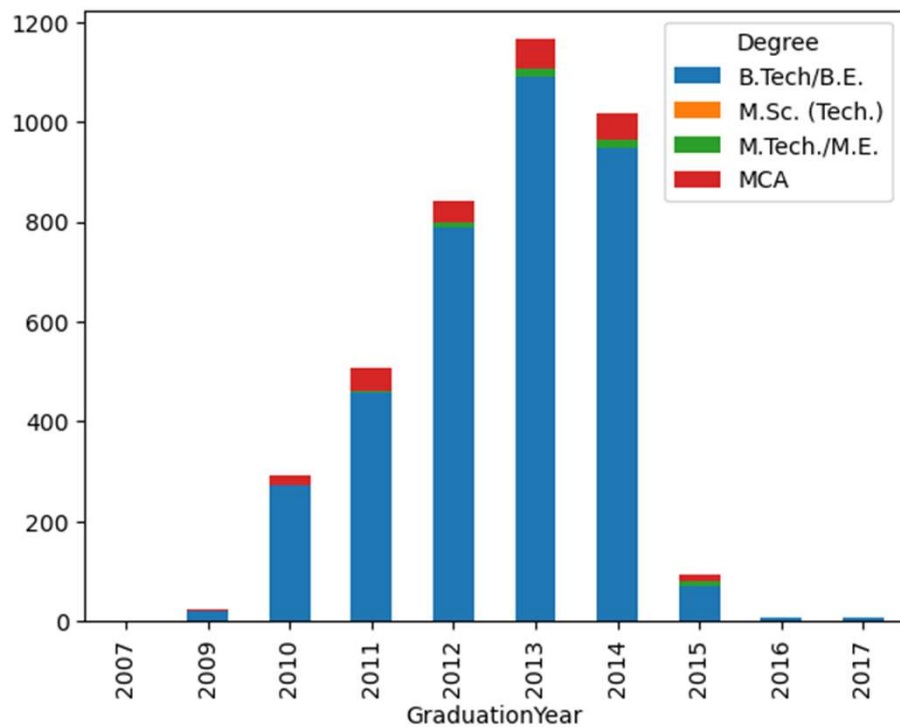
From above we can observe
 - Most of the B.tech graduates are from uttar Pradesh and many have graduated in 2013

From above we can observe
- Most of the bars show that there are twice the men working in the same field as women.

Research Questions

•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."

```
Designation
associate engineer    290000.000000
hardware engineer     281250.000000
programmer analyst    339784.172662
software engineer     342020.484171
Name: Salary, dtype: float64
```
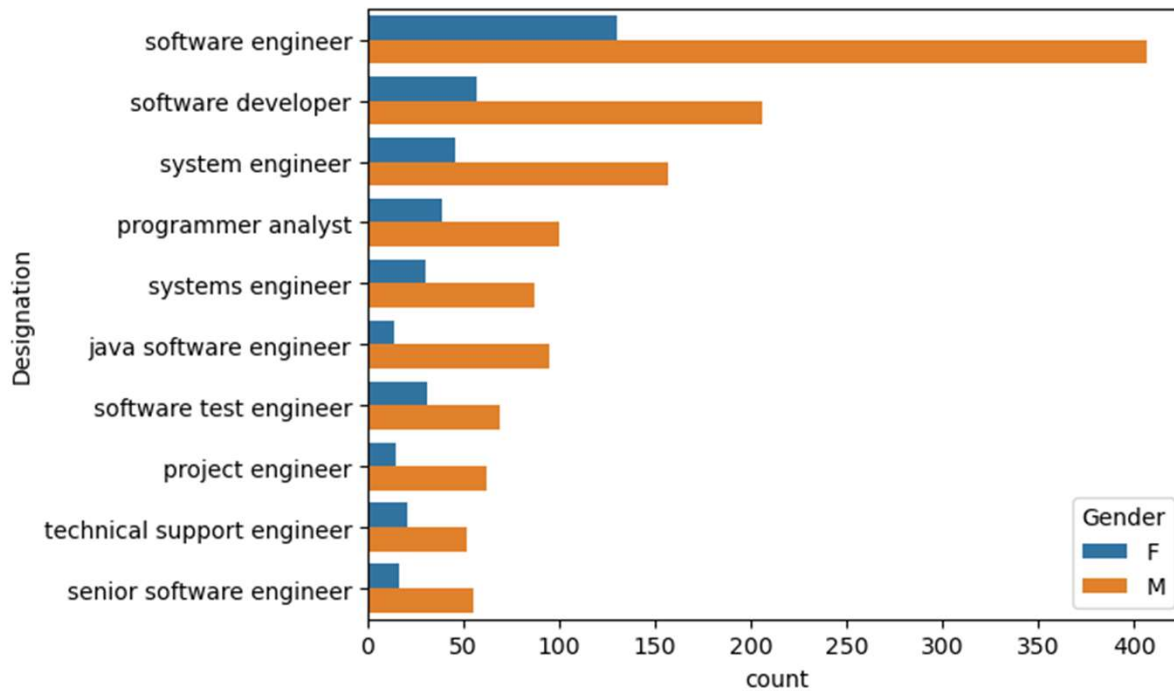
```
Designation           Gender
associate engineer    F        240000.000000
                      M        315000.000000
hardware engineer     F        290000.000000
                      M        278333.333333
programmer analyst    F        329230.769231
                      M        343900.000000
software engineer     F        337653.846154
                      M        343415.233415
Name: Salary, dtype: float64
```

From above we can observe that there the given insight was correct.

- Is there a relationship between gender and specialization? (i.e. Does the preference of Specialisation depend on the Gender?

```
Gender  Specialization
M       electronics and communication engineering       657
        computer science & engineering                  554
        information technology                          484
        computer engineering                            422
F       electronics and communication engineering       210
M       mechanical engineering                          191
        computer application                            182
F       computer science & engineering                  181
        computer engineering                            174
        information technology                          173
M       electronics and electrical engineering          159
        electronics & telecommunications                93
        electrical engineering                          63
F       computer application                            59
        electronics and electrical engineering          33
Name: Salary, dtype: int64
```

From above we can observe
 - Most of the men and women choose ECE as first priority and CSE as second priority.

Conclusion:

- Most of the people have done B.Tech
- Most of the employees are male
- We found varying salary trends across various designations.
- We found that most of the high paying jobs are related to computers.

THANK
YOU

INNOMATICS
RESEARCH LABS