Perform **Exploratory Data Analysis(EDA)** on the data-set given below.

Download the dataset from this link. [**CLICK HERE TO DOWNLOAD DATASET**](https://drive.google.com/file/d/1mV_ifvzzrlPWgdSzrwW5ZK-aF3S9tydZ/view?usp=sharing)

The detailed description about the dataset can be found here. [**Dataset Description**](https://docs.google.com/document/d/14d4n90nryAUgA-IdNSXJAFzmynn6WTT4XfS7N4tum7s/edit?usp=sharing)

**If you are facing any difficulty in performing EDA, follow the steps mentioned below:**

**Step - 1 -** Introduction -> Give a detailed data description and objective - done

**Step - 2 -** Import the data and display the head, shape and description of the data. - done

**Step - 3 -** Univariate Analysis -> PDF, Histograms, Boxplots, Countplots, etc..

* Find the outliers in each numerical column
* Understand the probability and frequency distribution of each numerical column
* Understand the frequency distribution of each categorical Variable/Column
* Mention **observations** after each plot.

**Step - 4 -** Bivariate Analysis

* Discover the relationships between numerical columns using Scatter plots, hexbin plots, pair plots, etc..
* Identify the patterns between categorical and numerical columns using swarmplot, boxplot, barplot, etc..
* Mention **observations** after each plot.

**Step - 5 - 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.*” Test this claim with the data given to you.
* Is there a relationship between gender and specialisation? (i.e. Does the preference of Specialisation depend on the Gender?)

**Step - 6 -** Conclusion

**NOTE:** Mention **observations** after each plot.

**For the below mentioned step do your own research (use Google). Hints are given below.**

**Step - 7 -** Perform feature transformation:

* For Numerical Features -> Do Column Standardization
* For Categorical -> if more than 2 categories, use dummy variables. Otherwise convert the feature to Binary.

**--------------------------------------------------- END OF TASK---------------------------------------------------**