**Python Project**

Data Analysis

Link: [**https://drive.google.com/file/d/1M0SOPww1w2qIebfJbPHwuVuxIra9SIHO/view?usp=sharing**](https://drive.google.com/file/d/1M0SOPww1w2qIebfJbPHwuVuxIra9SIHO/view?usp=sharing)

**Data Information:**

**1.sl\_no : Serial Number**

**2.ssc\_p : Secondary Education percentage- 10th Grade**

**3.ssc\_b : Board of Education- Central/ Others**

**4.hsc\_p : Higher Secondary Education percentage- 12th Grade**

**5.hsc\_b : Board of Education- Central/ Others**

**6.hsc\_s : Specialization in Higher Secondary Education**

**7.degree\_p : Degree Percentage**

**8.degree\_t : Under Graduation(Degree type)- Field of degree education**

**9.workex : Work Experience**

**10.etest\_p : Employability test percentage ( conducted by college)**

**11.specialisation : Post Graduation(MBA)- Specialization**

**12.mba\_p : MBA percentage**

**13.status : Status of placement- Placed/Not placed**

**14.salary : Salary if the student is placed**

* **Check the first four rows of the dataframe**
* **Get all feature names**
* **Find the number of records and columns**
* **Use the .info() method to find the number of Non Null entries and Data Type of each feature**
* **What is the average Secondary Education percentage - 10th Grade**
* **What is the max Secondary Education percentage - 10th Grade**
* **How many toppers where there in 10th Grade?**
* **Is the student who got highest Secondary Eduaction percentage, placed or not?**
* **How many students are placed or unplaced?**
* **How many unique degrees are there in the dataset?**
* **Is there a correlation between 10th and 12th percentage**
* **Find the correlation matrix?**
* **Data Pre-processing**
* **Identify the column which can be removed?(only 1)**
* **Remove the unnecessary column**
* **Check number of null values in each column**
* **Fill the missing values with appropriate values and check number of null values in each column again**
* **Data Visualization**
* **Import matplotlib and seaborn**
* **Draw a scatter plot between 10th and 12th percentage with labels and title**
* **Draw the scatter plot between 10th and 12th class percentage of students grouped based on placement data**
* **Draw a boxplot for 10th percentage of the students**
* **Draw a boxplot for 12th percentage of the students**
* **Draw a boxplot for 12th percentage of the students for placed and unplaced students**
* **Draw lineplot for 10th, 12th, degree and MBA percentage**
* **Find correlation between continous columns**
* **Draw heatmap of correlation**