## Executive Summary

In this report, data visualization techniques are applied to visualize data present in 'recent-grads.csv' in order to analyze earning of students graduated from college depending upon their majors. Data is visualized by using line graph, histograms, and box plots to draw the interesting insights from it.

Here's a link of the file that I’ll be working with:

[recent-grads.csv](https://github.com/fivethirtyeight/data/blob/master/college-majors/recent-grads.csv) - employment data according to college major for recent college graduates

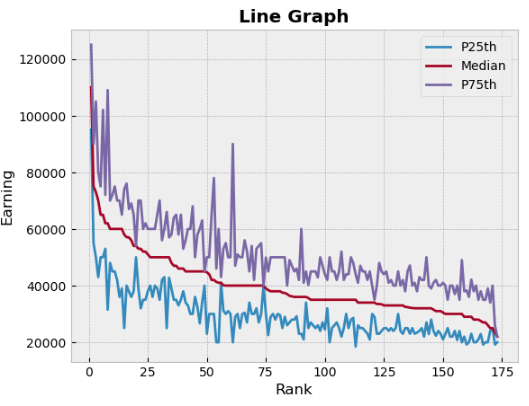
**Data Visualization**

My dataset contains some columns related to the earnings of students of different majors that are quite significant for data visualization. Here are the columns that I’ll use in this project:

* **"Rank"** is representing the major’s rank of each major of college.
* **"P25th"** is representing the 25th percentile of students’ earnings.
* **"P75th"** is representing the 75th percentile of students’ earnings.
* **"Median"** is representing median value of earnings of students.

**Creating Line Graph**

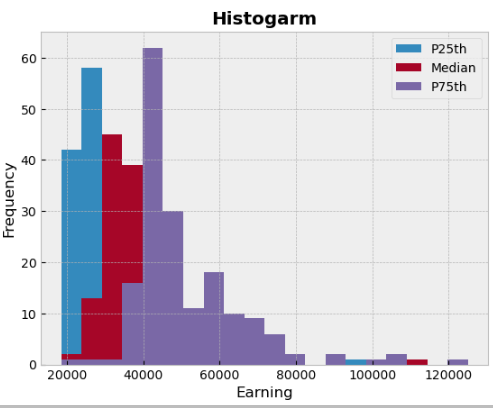
Line graphs also known as line charts are significant to demonstrate trends or changes over time. Let's begin exploring the data using line graph.



This graph is showing how earning of graduates changed over time by examining median, 25th and 75th percentile of it. It represents that students with major “PETROLEUM ENGINEERING” has highest earning whereas graduates with “LIBRARY SCIENCE” has lowest earning. It also shows that the majors having very small gaps in above line graph between the lines of 25th and 75th percentiles earn salaries close to the median salary.

**Creating Histogram**

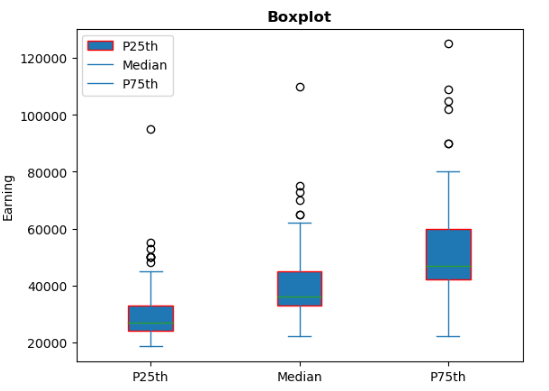
A histogram is used to represent data that is provided in a form of different groups with the help of bars.



The proposed histogram shows twenty bins ranging from 20,000 to $120,000. The most interesting fact that can be observed in this histogram is that the most common average salary range is 30,000 - 35,000 but only a few graduates are earning a median salary of over 80,000.

**Creating Box plot**

A **Box Plot** is used to exhibit summary of dataset values by highlighting its minimum, maximum, first quartile, third quartile and median.



The proposed boxplot is showing that maximum values for P25th, Median and P75th are 9,500, 11,000 and 12,000. Each box is showing there is a huge difference between them because median line of each box lies outside of other box.

**Github Link:** [**https://github.com/AaMNAHZaAFAR/AppliedDataScience**](https://github.com/AaMNAHZaAFAR/AppliedDataScience)