**Team Members:** Timothy Chien, Meenakshi Kavaseri, Rohit Rao, Harita Vellalath

An Analysis of Higher Education

**Question → Private vs Public (Cost vs Early/Mid-Career Pay)**

I analyzed 2 CSV files (Payscale Bachelors.csv and Cost of Attendance 1 .csv) and merged both of them on school name. I cleaned up the data by removing all special characters and irrelevant data.

I then calculated the 4 year tuition( by adding 2014 and 2015 and multiplying by 2).

Then I proceeded with the following steps:

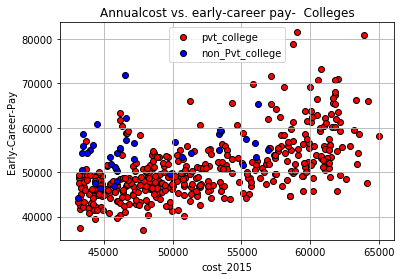
* Comparing top ten private vs. top ten non-private for Total cost.
* Comparing top ten Private vs. top ten non-private for Early-Career -Pay.
* Comparing top ten Private vs. top ten non-private for Mid-Career-Pay

After getting the information from above, I plotted the data using 2 scatter plots to find any correlation and trends.

1. Scatter plot based on cost\_2015 vs. early-career pay -all private college and all public colleges

Observations:

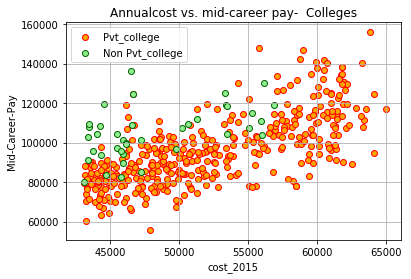
* Private colleges- There seems to be a correlation between what you pay and what you earn as early-career pay.
* The majority of the data points cluster around in the range of early $40,000s to mid $50,000s for both Cost and Pay.
* Cost of non- private schools are comparatively lesser than the private schools.
* For Public schools, there seems to be no real trend and data is less evenly distributed to draw conclusions.



(2) Scatter plot based on cost\_2015 vs. Mid-career pay (all private college and all public colleges)

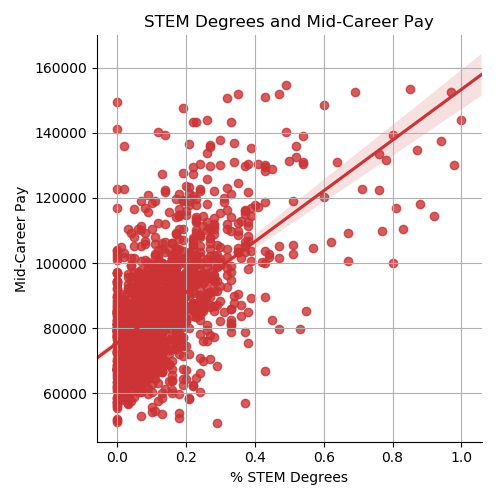
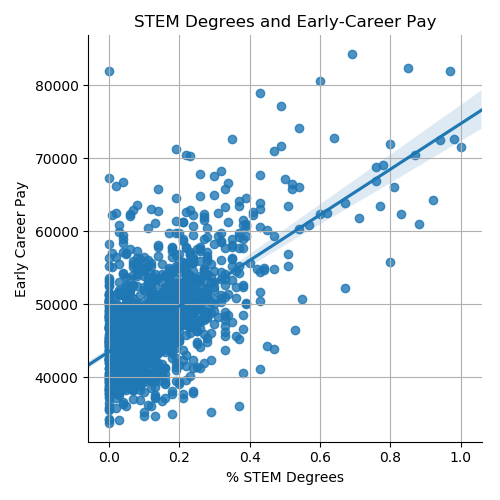
Observations:

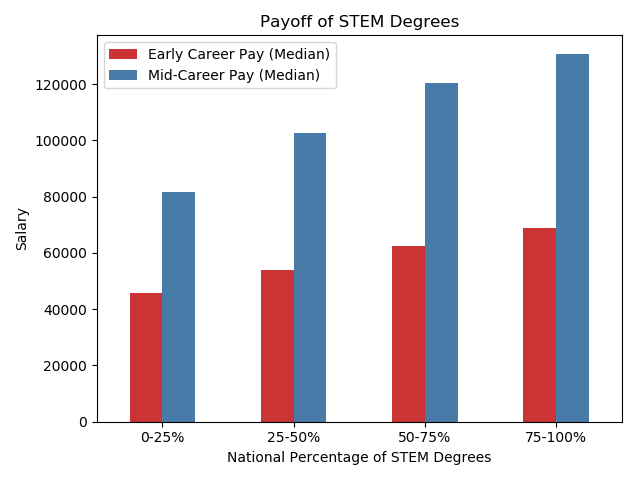
* Mid career pay is the same trend as the early career pay for private colleges. The scale seems to follow a linear trend.
* The trend for public colleges between early pay and mid career pay seems to be similar. No real trend.



# **Question → Does pursuing a STEM (Science, Technology, Engineering, and/or Mathematics) Major affect Early/Mid-Career Pay?**

After analyzing “payscaleAlumni.csv,” a dataset obtained from Payscale.com, I created the following plots to model the correlation between “% STEM,” a variable used to measure the percentage of students who graduated with degrees in STEM (Science, Technology, Engineering, and Mathematics) from colleges across the country, and both early/mid-career pay of graduates:





The two scatterplots were created using a combination of Seaborn and Matplotlib, while the bar chart was created using a combination of Matplotlib and the Pandas plotting function. In addition, I used the Pandas “cut” function (pd.cut()) to “bin” the data as shown above.

After analyzing all 3 plots, I observed a positive correlation between the percentage of students who graduated with STEM degrees and the salaries they received both in the early and mid stages of their careers.

# **Question → How do different types of schools affect Early or Mid-Career Pay?**

[How significant is an Ivy-League degree to your career pay? In a group whose cultures traditionally place strong emphasis on education this question popped out immediately, but what other classifications might be significant for a college applicant?](https://docs.google.com/document/d/1lbxq_9l7-W9tNMOsEB8Gy-HZHig9cuqi2md2XTrcaL8/edit?usp=sharing)

|  |  |
| --- | --- |
|  | * Graduates who went on to pursue further degrees and education in most cases only made ~2% more than the graduate who started working with just the bachelor’s degree * Attending an **Ivy League** (+43% / +60%), **Engineering** (+38% / +42%) schools seem to have significantly higher early and mid-career pay * Schools known for **Research** (+10% / +14%) or **Sports** (+7.6% / +11.6%) seem to have moderately higher career pay * **Business** school graduates do not seem to vary much from the public/private averages (within 2%) * **Religious** (-4.1% / -5.7%) school graduates tend to average below that of public/private schools |

# **Question → What are the college majors with the best / worst on investment?**

I worked with two sets of data. I used “early\_mid\_pay.csv” and “degrees-that-pay-back.csv”. The available data was very limited to answer my question. The only data set that was of use in order for me to answer my question was the early\_mid\_pay.csv. This data set gave me a list of more that 400 different and very specific college majors and their post - graduation early and mid - career pay. After cleaning up the list, and pulling only the columns that I needed, using jupyter notebook, I displayed the top 5 in a table and the bottom 5 in a table based on mid - career pay. With the table, I used matplotlib to create two bar graphs as visuals. From performing this analysis, I have gathered that, the top 5 majors with the greatest return on investment mostly have to do with science, with the exception of one math major, and the lowest return on investment mostly have to do with early childhood education with the exception of one veterinary technology. This analysis shows the significant difference between the top five mid-career pay (175,000) vs the bottom 5 (40,000).

