Analysis of College Majors & Graduates

Authors: Pablo Ilabaca, John Martinez, Andrew Mejia, Uriel Guijarro, Michael Miranda
Department of Information Systems, California State University Los Angeles
CIS-4200, Data Warehousing & Business Intelligence

1. Introduction

Thousands of people graduate from universities across the United States. All these graduates have different degrees, interests, and goals that they plan to have within their careers. However, what is the difference in pay for these graduates? We know that all degrees get different salaries, but with this project, we used some data to look at how vastly different these graduates get paid and what other factors can also change their salaries.

The main objective of this project is to look at the highest and lowest paying majors if the school that someone attends can vastly affect their pay if people who attend universities on the West and east coast have higher paying jobs, and also what are majors that have the highest difference in pay from the starting salary to a mid-career.

This is important because when people are applying to college, primarily high school students, they should know what they are getting into when both are applying to a college and applying with a major in mind. The type of college can impact whether that student can see if it is worth the investment as "higher end" universities such as engineering and the Ivy League can be more expensive and competitive to get into. The region can determine if to move regions to attend a college that is likely to pay a better salary. For example, if a student lives in the western region and, on average, he has lower salaries than in an eastern region. It may be worth it for the student to move and try to find a college that can end up paying later in that region. The salary increase in the major is also significant because, in the dataset, there could be instances in which the first few years of a job may not pay so well. However, after some time, the job may end up paying a salary well worth the time put in.

2. Dataset

The dataset we used came from Kaggle, but it originally came from *"The Wall Street Journal"* website, and is an accumulation of three separate csv files: (1) ***Degrees that Pay you Back,*** which shows the different types of college majors and their average annual income in different stages of their career paths (2) ***Salaries by Colleges Type***: shows the average salary at different stages of a career path by the type of college, and (3) ***Salaries for Colleges By Region,*** which showcases the average salary at different stages of college major specific careers by region type. Each csv file has a column dedicated to the starting and mid-career median salary at the 10th, 25th, 75th, and 90th percentile. This dataset was last updated 6 years ago, and each csv file contains similar values and columns, with little to no difference between the three.

The tools we used to analyze these datasets were just Tableau for almost all of the visualization and Excel for data preprocessing.

2.1 Data Preprocessing

For our dataset, out of the three csv files, two of the files had missing values. Within the ***Salaries by College Type*** file, there are missing values within the columns of Mid Career 10th Percentile Salary and within the Mid Career 90th Percentile Salary. On the ***Salaries by Region Type*** csv file, there are also missing values within the columns of the Mid Career 10th Percentile Salary and the Mid Career 90th Percentile Salary:

Mid-Career 10th Percentile Salary	Mid-Career 90th Percentile Salary
$68,400.00	$257,000.00
N/A	
N/A	N/A
$59,500.00	N/A
N/A	$201,000.00
$55,000.00	
$51,300.00	N/A
$51,700.00	$178,000.00
N/A	$193,000.00
$49,700.00	
$52,000.00	$177,000.00
$47,800.00	N/A
$50,700.00	$201,000.00

Figure 2.1.1- Missing values within the both the Salaries by College Type and Salaries By Region Type csv file.

We decided to fill in the values rather than leave them as N/A because it would disrupt the results during the visualization and analysis process. We did so by using the average function (=AVERAGE[:]) in Excel to find out the average annual salaries for both the 10th and 90th Percentile columns in those incomplete datasets. In the ***Salaries by Region Type*** file, within the Mid Career 10th Percentile Salary column, we got an average of **$45,253.11**, and for the Mid Career 90th Percentile Salary column, the average was **$160,442.12**. For the ***Salaries by College Type*** file, applying the average function resulted in **$44,250.65** for mid-career 10th percentile salary, and **$157,705.63** for the mid-career 90th percentile salary column. With all columns now filled, we could prepare for visualization and analysis.

3. Analysis Questions

Based on this dataset, we set to find out the results from the following questions:

1. **What are the highest and lowest paying majors per annual average salary?**
   We will look at the minimum and maximum average salaries for different majors. We can see if there is correlation between majors such as STEM, that pay around the same or majors in the liberal arts.

2. **Do people who attend "better schools" (such as private or Ivy Leagues) actually have higher paying jobs or are able to obtain raises more than those who attended other schools?**
   For this question, we would like to find out if Ivy Leagues have higher paying jobs than other schools. As well as an overall comparison to see how states, party, liberal arts, and engineering schools fare against one another.

3. **Do people who attend universities on the east and west coast tend to have higher paying jobs?**
   This question is more straightforward, it seems as if there is common knowledge that east and west parts of the US have better jobs, so likewise, the universities may have a similar attribute. We can also look at if the type of university has an impact on this, such as states, parties, etc.

4. **Are there majors that typically pay a lower entry-level job and then pay better after years of experience?**
   A part of our data set involves the starting median salary and the mid-career median salary. We can compare which undergraduate majors pay back the best as well as which ones pay back the worst amongst all applicable majors. In order to find this out, we utilized the dataset of "degrees-that-pay-back". We wanted our focus to be on how much of a change there was from the "starting median salary" to "mid-career salary". Using Tableau, we were able to find this change by making the measure "Percentage Difference from Starting to Mid-Career Salary" by subtracting "mid-career salary" from "starting career salary" and turning that result into a percent.

4. Group Members and Responsibilities

Below are the responsibilities that were given to each team member for this project in order to provide clean and sound analysis of this dataset:

Pablo: Assist in researching topics related to this dataset, assist in writing the report, and also aiding in any excel related tasks that are needed

Andrew: Will assist in handling data visualizations and inputting any relevant data into visualization tools (Excel, Tableau, etc.)

John: Will clean up the data and fill out any missing values, remove duplicates, etc. Will also assist in data visualizations and research into as to why we get the results we have

Michael: Data visualizations and will assist in researching and analyzing the data processing

Uriel: Will help in researching the topic and look at common trends and relationships. Will also assist in analyzing data, creating visuals, and turn it into writing

5. Analysis and Visualization

We used a combination of Excel and Tableau to help visualize our data for us to better understand and help answer our questions.

Starting with the first question, we decided to use bar graphs to demonstrate the highest and lowest-paying majors per annual average salary. The data had both starting median salary and mid-career median salary. This way, we could also see if there are degrees that continue to pay well or continue not to pay so well.

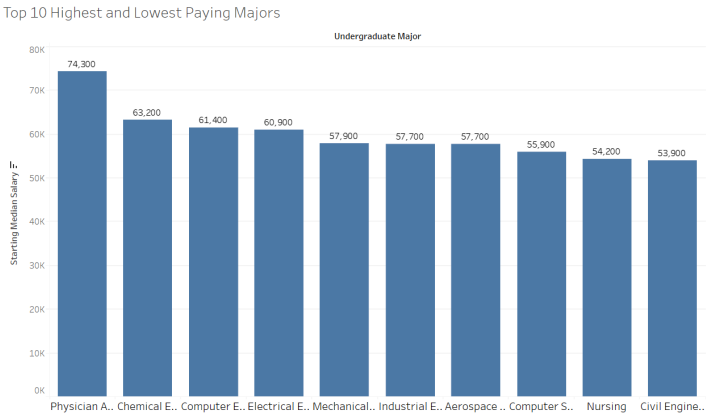

Figure 5.1.1 - Bar Chart for Top 10 Highest Paying Majors, Starting Salary

***Dimensions**: Undergraduate Major; Filtered by: Highest to Lowest, Top 10*
***Measures**: (AVG) Starting Median Salary*

This chart demonstrates the top 10 highest paying majors by starting median salary. It's worth noting that when looking at our graphs, we included all 50 majors, however, due to visibility issues, we just included the top 10. In the visualization with all 50, we color-grouped the degrees into 1 of 2 fields, STEM or Liberal Arts. Meaning degrees that were most closely related to STEM were blue, and degrees that were most similar to Liberal Arts were orange. The highest paying degree in this graph was Physician Assistant.

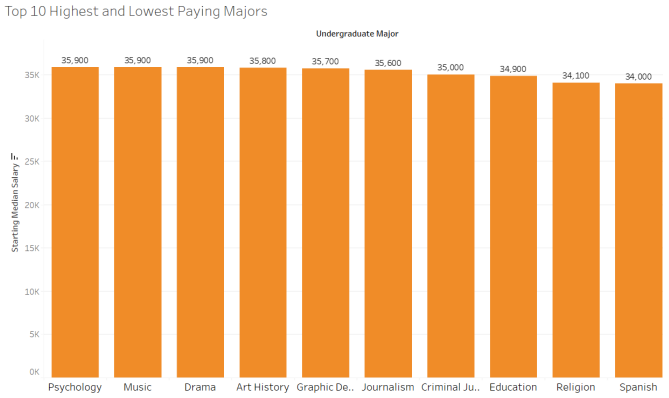

Figure 5.1.2 - Bar Chart for Bottom 10 Lowest Paying Majors, Starting salary

***Dimensions**: Undergraduate Major; Filtered by: Highest to Lowest, Bottom 10*
***Measures**: (AVG) Starting Median Salary*

This chart uses the same dimensions and measures, only that it shows the lowest 10 paying degrees based on starting median salary. The chart shows that the lowest-paying degrees for starting median salary all fall under the Liberal Arts category. The lowest-paying degree was Spanish.

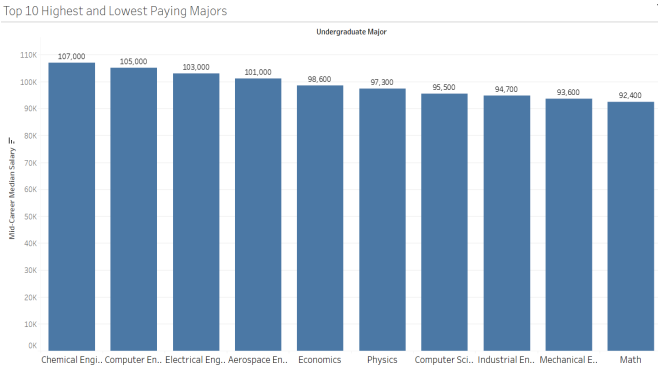

Figure 5.1.3 - Bar Chart for Top 10 Highest Paying Majors, Mid-Career Salary

***Dimensions**: Undergraduate Major; Filtered by: Highest to Lowest, Bottom 10*
***Measures**: (AVG) Mid-Career Median Salary*

For this graph, we looked at the mid-career median salary for the top 10 degrees. Again, we found that the top 10 highest-paying salaries were in a STEM-related field. With the highest paying degree being Chemical Engineering at $107,000.

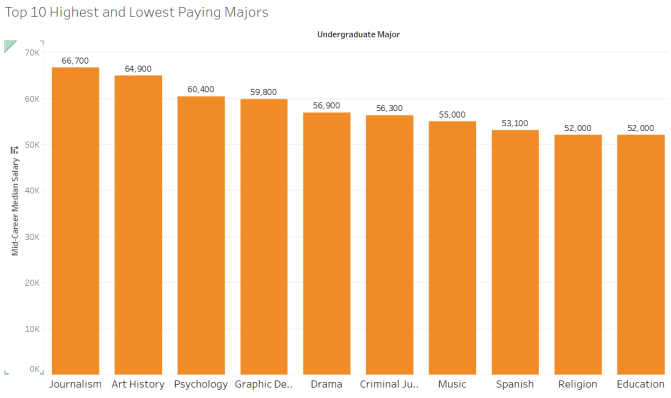

Figure 5.1.4 - Bar Chart for Bottom 10 Lowest Paying Majors, Mid-Career Salary

***Dimensions**: Undergraduate Major; Filtered by: Highest to Lowest, Bottom 10*
***Measures**: (AVG) Mid-Career Median Salary*

This last graph, pertaining to our first question, we looked at the bottom 10 mid-career median salaries and once more, the degrees were under the umbrella of Liberal Arts. This time, the lowest paying degree is Education at $52,000. Overall, it was apparent that degrees relating in some way to STEM were the highest paying degrees, both in starting and mid-career salary. This demonstrates our technology-forward world is continuing to expand its workforce to be technologically positive in some way. STEM jobs are highly sought out so an individual can think for themselves that a STEM-related job may be worth it for the short and long term.

For the second question we used bar graphs to separate each graduate's average starting salary and then we could better compare each one to each other.

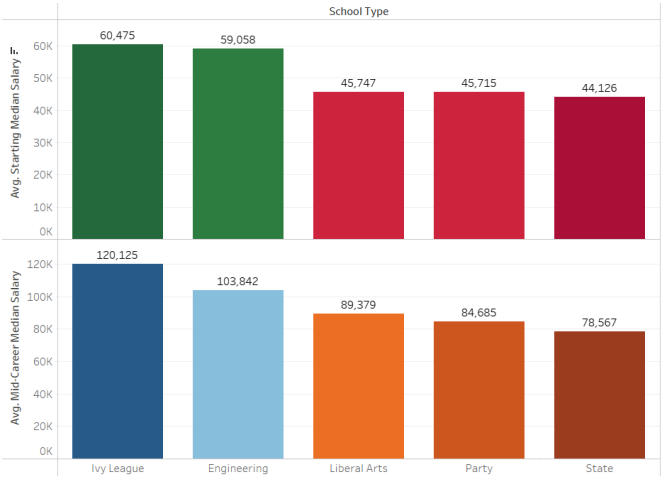

Figure 5.2.1 Bar chart of each starting salary by school type

***Dimensions**: School Type*
***Measures**: (AVG) Starting Median Salary and (AVG) Mid-Career Median Salary*

From the bar chart above, we can see that those who graduated from Ivy schools do tend to have a higher average starting salary than other types of colleges, with an annual starting salary of 60k, engineering schools were a close second with 59k being the average starting salary; just short of 1k difference between the two. State schools had the lowest starting salaries with 44k.

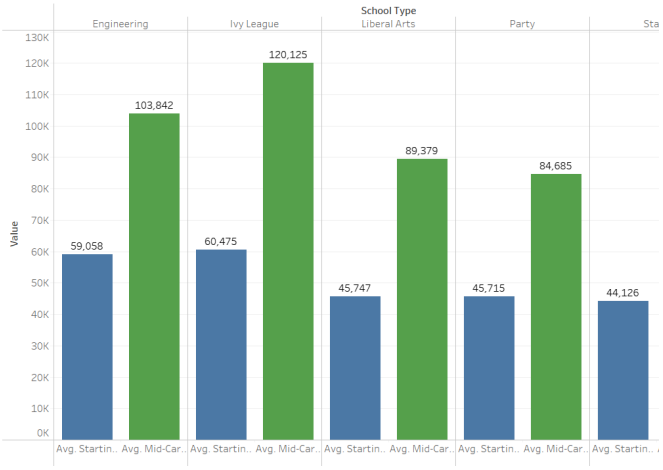

Figure 5.2.2 Bar chart of comparison of start and mid-career salaries

***Dimensions**: School Type*
***Measures**: (AVG) Starting Median Salary and (AVG) Mid-Career Median Salary*

People who attend universities on the east coast and west coast do tend to have higher paying jobs than those in the midwest and the southern regions. We combined the Californian and Western regions into one called West Coast to fit our question. Keep in mind that the count of schools in each region is not equal. We will also show you examples of schools that top the specific regions and the salary differences.

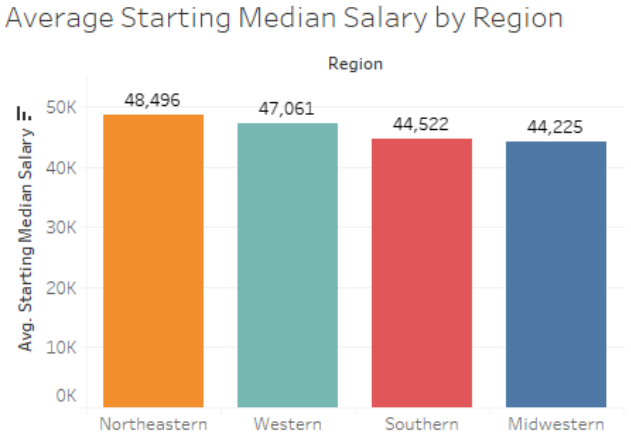

Figure 5.3.1 - Average Starting Median Salary by Region

***Dimensions**: Region (Western,Southern,etc.)*
***Measure**: Starting Median Salary (48,496) (Average)*
Descending Highest to Lowest

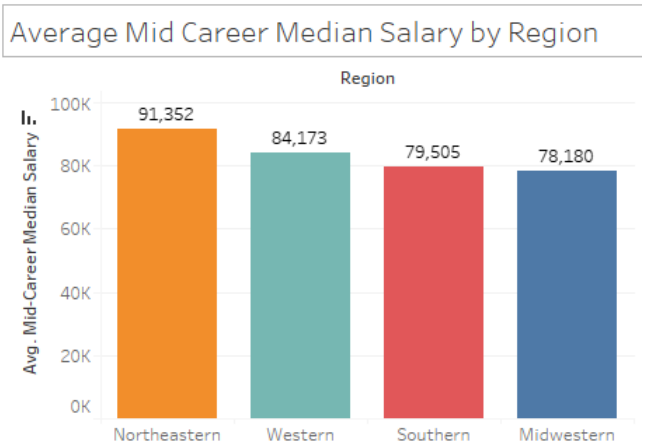

Figure 5.3.2 - Average Mid-Career Median Salary by Region

***Dimensions**: Region (Western,Southern,etc.)*
***Measure**: Mid Career Salary (91,352) (Average)*
Descending Highest to Lowest

Row Labels	Average of Starting Median Salary	Average of Mid-Career Median Salary
Northeastern	$48,496.00	$91,352.00
Western	$47,061.43	$84,172.86
Southern	$44,521.52	$79,505.06
Midwestern	$44,225.35	$78,180.28
Grand Total	**$46,253.44**	**$83,934.38**

Figure 5.3.3 - Pivot table view of Average Starting and Mid-Career Median Salary by Region

***Dimensions:** Region (Western)*
***Measures:** Average Starting and Mid Career Salary*
Filtered by Highest to Lowest

There is a $4k difference in salaries for the starting median salary so it is not a big difference, but it is still higher. The Mid-Career Median Salary is higher for the West and East coast compared to other regions, but we do see a bigger difference compared to the starting median salaries. The East Coast has a $12-13k difference; the West has a $6k difference from the Midwest and South.

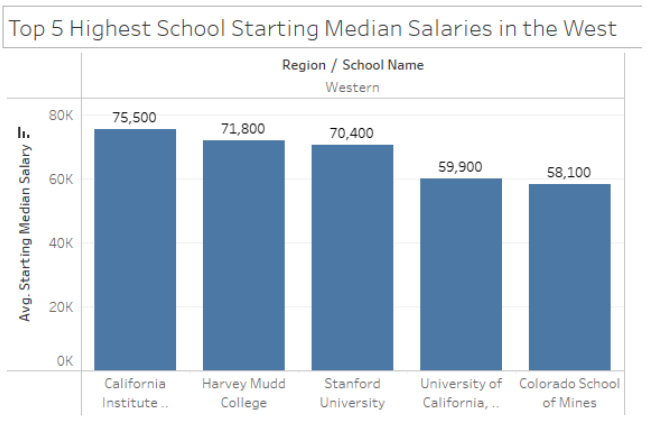

Figure 5.3.4 - Top 5 Highest Schools Starting Median Salary in the West

***Dimensions**:Region-Western, School Name (Stanford)*
***Measures**:Average of Starting Median Salary*
Filtered by Top 5 of Avg. Starting Median Salary by School Name Descending Highest to Lowest

You will see CIT top the list for starting salary in the west while Stanford and Berkeley also make the list, with some notable household name universities.

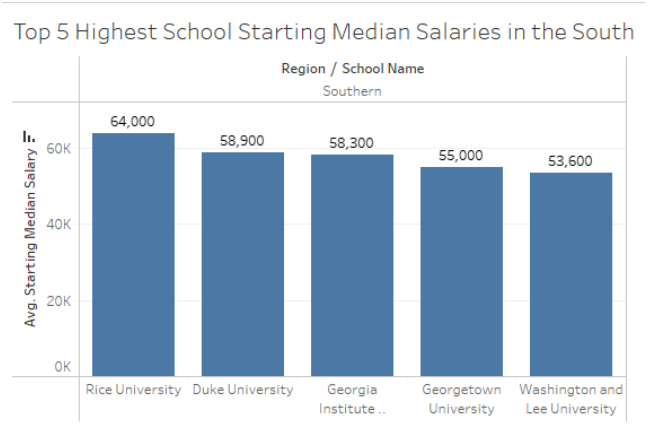

Figure 5.3.5 - Top 10 Highest Schools Starting Median Salary in the South

***Dimensions**:Region-Southern, School Name (Rice)*
***Measures**:Average of Starting Median Salary*
Filtered by Top 5 of Avg. Starting Median Salary by School Name Descending Highest to Lowest

This chart shows the top 10 schools in the south by starting median salary. The highest university, Rice University in Texas, has a $64k starting median salary, lower than the highest school in the west, CIT, by around $10k.

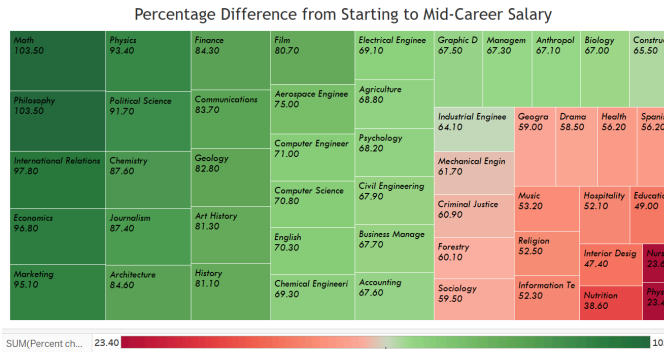

Figure 5.4.1 - Percentage Difference from Starting to Mid-Career Salary

***Dimensions**: Undergraduate Major; Filtered by: Lowest to Highest*
***Measures**: (SUM)Percent change from Starting to Mid-Career Salary*

The tree map above displays the highest percentage difference from starting salary to mid-career salary is Math, and Philosophy respectively with 103.5%. We can also defer that the lowest percentage difference from starting salary to mid-career salary is Physician Assistant with 23.4%. A factor to note is potentially already high paying starting careers may not increase salary over time compared to other majors. For example, Nursing may not provide much of a high percentage difference due to already paying well, unlike when we denote an Economics major. Economics may pay poorly in the entry-level position, but over time with a mid-career salary could be higher paying.

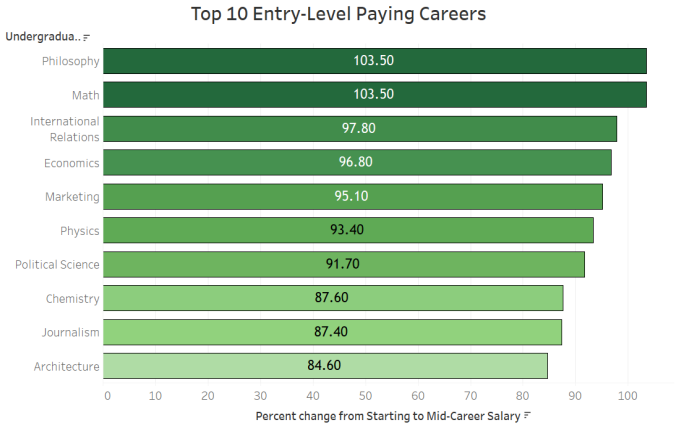

Figure 5.4.2 - Bar Chart of top 10 starting salary careers

Dimensions: Undergraduate Major
Measures: (SUM)Percent change from Starting to Mid-Career Salary; Filtered by: (SUM)Percent change from Starting to Mid-Career Salary>84.6 & Highest to Lowest

Here by filtering out the dataset with presenting the top ten starting salary paying careers, can we make an educated assumption that there are careers that'll pay more than others. Out of the 50 majors given to us within the dataset, Philosophy, Economics, Physics, to Architecture these are the most promising, and profitable majors. Not only is Philosophy and Math the most paying salaries, but just beneath them are International Relations and Economics, which is surprising. Some individuals might infer Physics, and/or Chemistry would be among the top ten. In this instance, they are not.

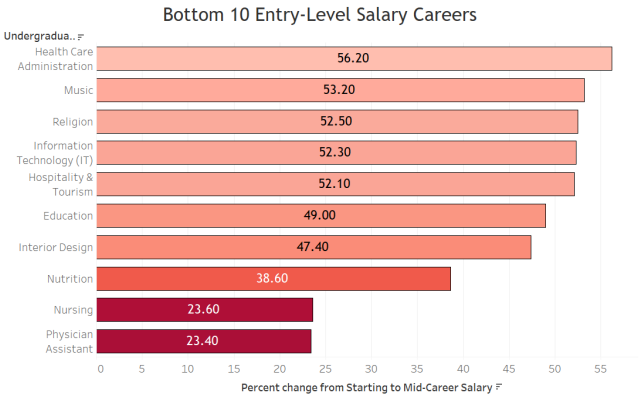

Figure 5.4.3- Bar Chart of bottom ten starting salary careers

***Dimensions**: Undergraduate Majors; Filtered by: Highest to Lowest*
***Measures**: (SUM)Percent change from Starting to Mid-Career Salary; Filtered by: (SUM)Percent change from Starting to Mid-Career Salary<56.2 & Highest to Lowest*

By filtering out the data with only presenting the lowest ten percent difference from starting to mid-career salary, we can see which fields are worse than others and by what margins exactly. From the 50 majors given to us, the top of the bottom ten majors is Health Care Administration, with Music right underneath. The lowest of the group being Physician Assistant, and Nursing being not too far away. In between these are Nutrition, Information Technology, and Religion. It's interesting to see Nursing ranked second-to-last from the bottom of this bar chart because that would mean being in the Nursing field there is not much growth for increased salary over the years compared to other majors. The same for Physician Assistant as well.

6. Limitations and Further Analysis

It was difficult to do further analysis and connect our 3 sheets when analyzing the region data sheet specifically. The school type data sheet had inconsistencies and had major school count differences by school type, so it would not provide an accurate view of the money you can make from some colleges. For example, there would only be one "party" school in California while having 20 "state" schools in that region. If we did want to read into the analysis, Ivy Leagues and Engineering schools have higher average salaries than those in other school types but are held back by the number of schools that each school type has.

7. Conclusion

From our visualizations and findings, we can conclude that most of what we hypothesized for each question was correct.

When questioning the top 10 highest-paying majors and the bottom 10 lowest-paying majors, we assumed that most, if not all of the top 10 highest-paying majors would be in STEM. This is because with rapid growth, high pay, job security in the technology field, engineering, and medical field, STEM majors would be highly looked upon by students. Based on our analysis, we can confirm this to be almost true, as nine out of the ten highest-paying majors tend to be in STEM, with the one exception being Economics placing in the top 5. The bottom 10 lowest paying majors were more towards liberal arts.

Before analyzing, we assumed that people who attended Ivy League schools tend to have both a higher entry-level and mid-career salary than those who attended other college types. We can confirm this to be true, as those who attended Ivy League schools tend to start out with higher pay than the other schools in both entry-level and mid-career. What was surprising, however, is that when analyzing the entry-level average salaries for all college types, engineering colleges made almost the same as Ivy League colleges, with a difference of only one thousand dollars. However, when analyzing mid-career salaries, Ivy League schools had a huge increase in comparison to engineering colleges, with a difference of seventeen thousand dollars. This is probably because Ivy Leagues tend to have grandeur alumni networks and have prestigious faculty in comparison to the other schools. State schools tend to make the least in both early and mid-career average salary.

On the third question, we also guessed that the West and east coast schools would be getting paid higher than those schools in the middle part of the United States. We show that our initial guess was true as this was also the cause but the same trend continues that the starting salaries had a small difference while the mid-career salaries had a significantly larger difference. This also made sense as east and west coast areas tend to have higher costs of living which means that they also need to have higher paying jobs as well.

For our last question, we were actually unsure of what the correct answer would be. We were sure how much the change of pay of a job would be from starting to mid-career. We were surprised to discover that it was philosophy and math that were the ones who saw the highest increases. Philosophy does make sense as it does seem like a career that would start off paying low but then could definitely see the increase as a professor or someone who decides to pursue higher education. Mathematics also makes sense as professors from any field tend to get paid lower but math could be one that sees a higher increase due to the ceiling that mathematics has, as well as being able to teach simple subjects such as algebra to more complex subjects such as multivariable calculus. We were also surprised to learn that jobs in the medical field such as physician's assistants and nursing had the lowest increase in pay over time. This also did make sense as those jobs tend to have higher paying salaries at first compared to most jobs but then they stagnate over time.