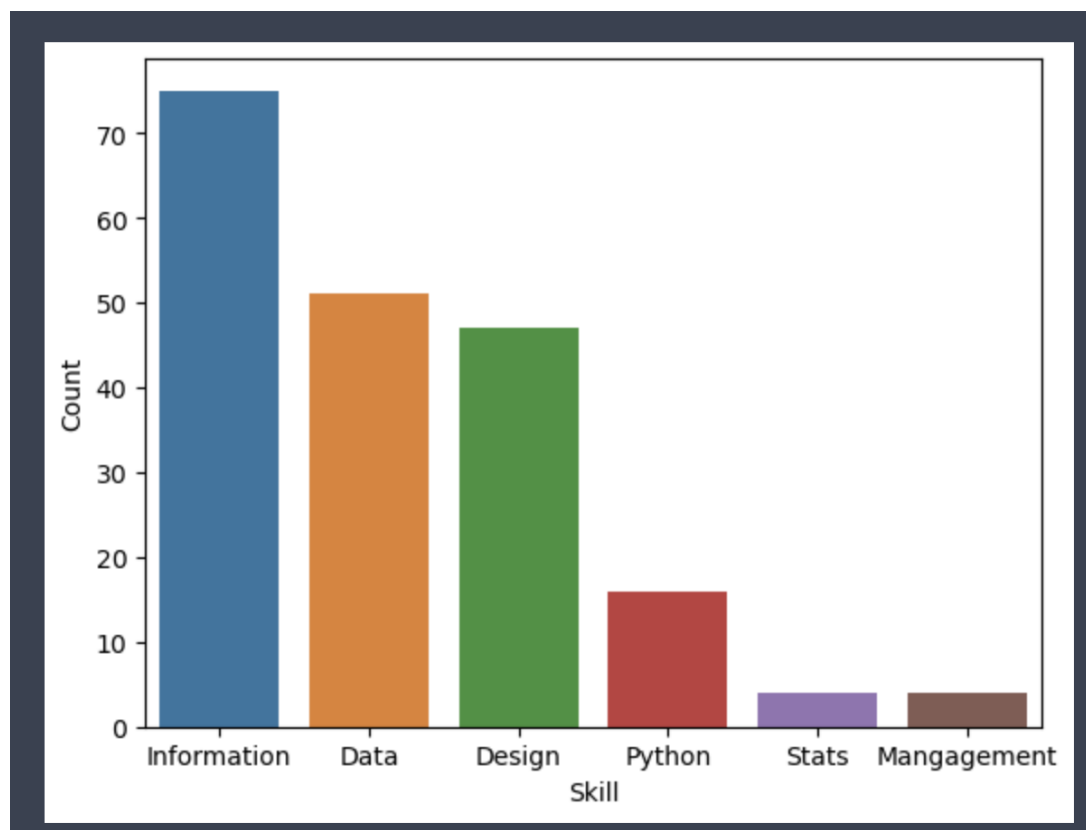


Research Question: What (data)jobs are available to students after they graduate with a bachelor's degree in Information Science, and what is the pay?

The original scope and purpose of this project is to identify hard skills needed to land a future job in Information Science, as well as to start to understand the associated salary with those jobs. I used CSV files that I found on the Kaggle data website as well as some rudimentary web data scraping skills with BeautifulSoup from the course curriculum for INFO classes. The purpose of taking data from the CU Boulder [website](#) was to visualize and get an idea of the skillset gained from the Bachelor's degree.

### 1.1 Information Skillset



From the course description it is worth noting that it is not a clear depiction of the skills gainable from the course due to vagueness in the writing. I turned the BeautifulSoup into a dictionary with

counts of mentions and then graphed it as shown above. Python was changed into python but initially, it was called ‘programming.’ However, most of the INFO courses teaching programming are specific to Python, so that is why I made that renaming choice. I think using the syllabus for the classes would have made for a more accurate ‘skillset’ gained. Perhaps this is an area for improvement and further research work for a later study.

## 1.2 CSV Skillsets

After this graphic, I transitioned into using CSV data sources and pandas, the majority of this project. This data was collected from Kaggle, a link to this is [here](#). I then made a similar graphic that would show the most sought-after skills for a data professional. Most job postings

```
jobsDF[['title']].query('title == "Data Scientist"')  
# 384 of these positions are equal to data scientist
```

	title
194	Data Scientist
246	Data Scientist
351	Data Scientist
379	Data Scientist
380	Data Scientist
...	...
31111	Data Scientist
31112	Data Scientist
31310	Data Scientist
31342	Data Scientist
31469	Data Scientist

384 rows × 1 columns

```
jobsDF[['title']].query('title == "Data Analyst"')  
# as opposed to the 4119 Data Analysts
```

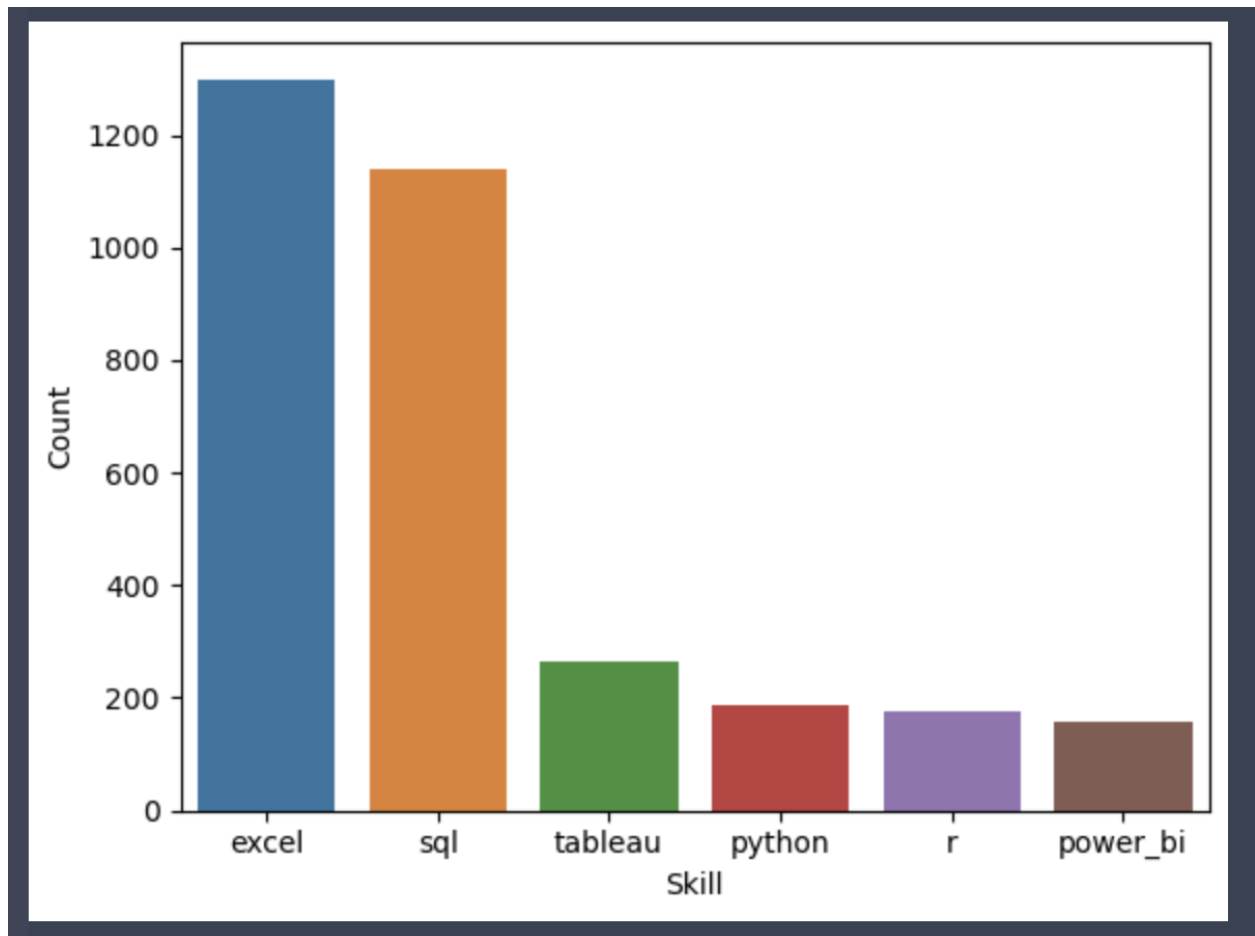
	title
0	Data Analyst
2	Data Analyst
3	Data Analyst
4	Data Analyst
6	Data Analyst
...	...
35135	Data Analyst
35143	Data Analyst
35182	Data Analyst
35202	Data Analyst
35203	Data Analyst

4119 rows × 1 columns

were for analysts of some sort, but there was variation within the data job family.

To provide context, the majority of the job postings were “Data Analyst” postings, compared to the 384 postings under “Data Scientist,” both of these occupations being plausible after having graduated from CU with an INFO degree. While this query shows a general trend/makeup of our first dataset, it doesn’t include the many different types of analysts, scientists, and data engineers (of which there are 194). This query is a rough estimate of the makeup of the data and only returns a number if the position title is exactly

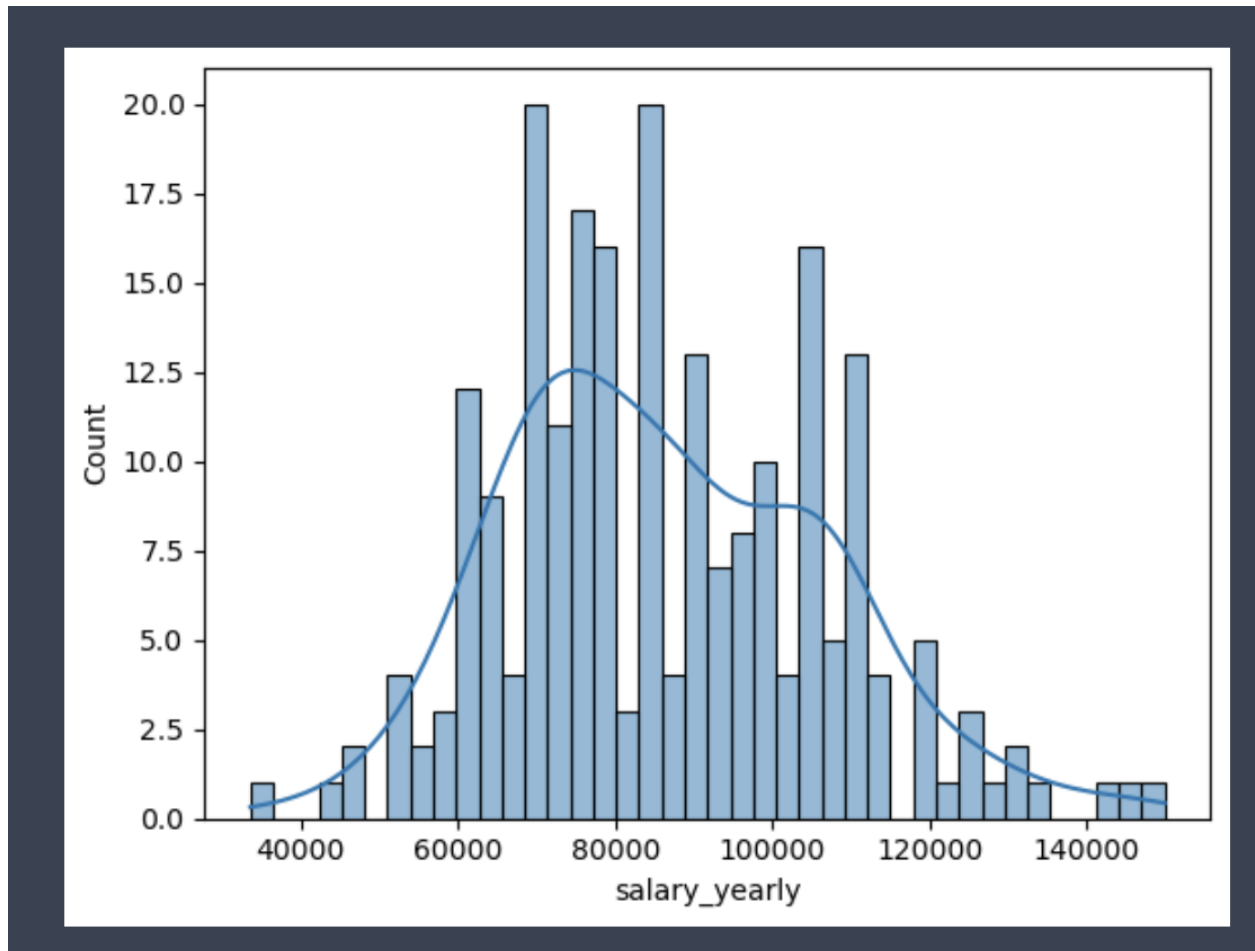
equivalent to ‘Data Scientist’ or ‘Data Analyst.’ After describing the makeup of the dataset, I went headfirst into exploring the skills associated with these data jobs. I didn’t differentiate between the data jobs as potentially these jobs could be earned by an INFO graduate.



These are the top sought-after skills for data professionals. This gap between sought-after skills and ‘taught’ skills does exist. For example, an SQL course is very necessary in today’s job market. Excel competence is required in most jobs throughout most industries, so I feel that is the ‘communication’ of soft skills. It is worth noting that Excel is ‘taught’ in the basic statistics course taught in INFO, so it is not left unsatisfied. There is one ‘R’ class taught as an upper-division elective in INFO. Tableau was touched on briefly in INFO 1111. However, I think

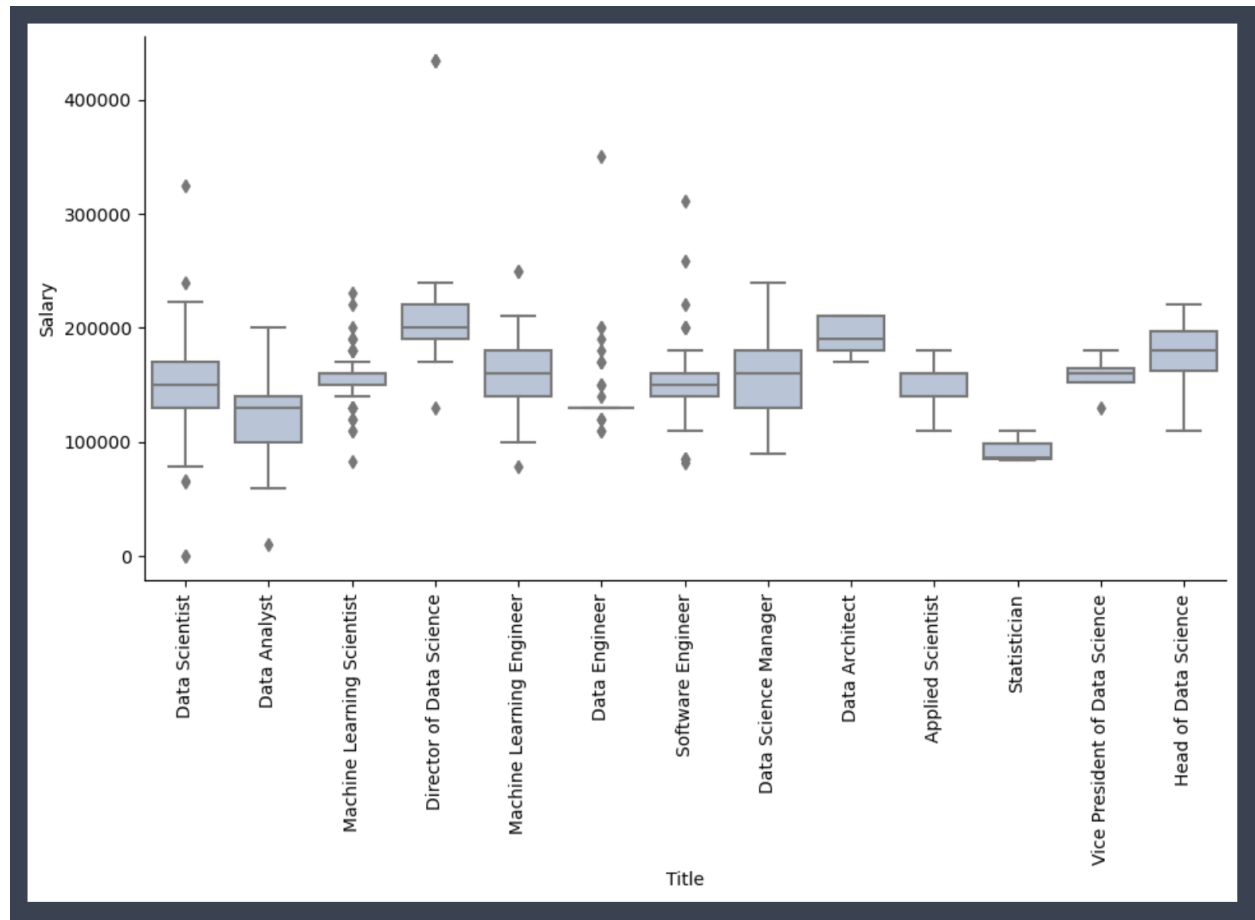
having an INFO class that reinforces the less technical but still sought-after skills like Excel, Tableau, and maybe PowerBI would be useful. Power BI and Tableau are very similar products, and I think that if only one of them were taught adequately, then this requirement would be satisfied.

### 1.3 Salary EDA

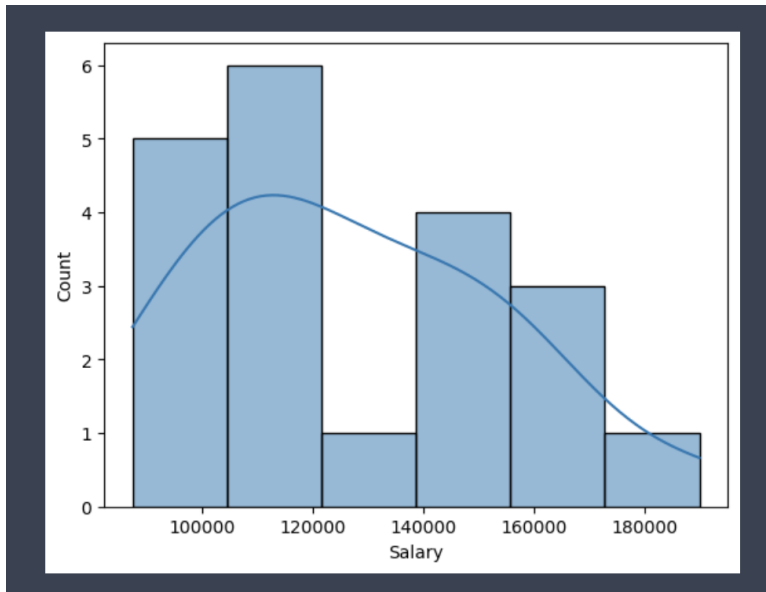


This is a generalized distribution of salary across all job families in the dataset, all data professional data jobs. From 60k to 120k is a nice distribution that shows me that there is plenty

of room for improvement and growth within jobs in the industry.



Looking at this awesome graphic, all job families have a pretty good spread of money. The most common ones, “Data Scientist and Analyst,” are both relatively high, and there is room for growth in both job families. From this graphic, I learned that not all Data Analysts will switch to data science because their salary can be just as high if they have enough experience. The diamonds outside the IQR plots represent outliers. Outliers are not the best thing to base the generalized information on. The slash in the middle of the box represents the median salary.



This chart shows the ‘entry-level’ salaries for the amounts of careers in Data Science.

I think most data science jobs, even entry-level positions, typically require some experience as a data professional.

That is why the salary is already in 6 figure areas. The average salary of Data Scientists is \$126,736 a year.

	Title	Salary	Levels
0	Data Scientist	110000.0	Jr.
4	Data Scientist	130000.0	Jr.
5	Data Scientist	100000.0	Jr.
33	Data Scientist	87360.0	Jr.
44	Data Scientist	150000.0	Jr.
48	Data Scientist	120000.0	Jr.
126	Data Scientist	160000.0	Jr.
142	Data Scientist	140000.0	Jr.
170	Data Scientist	120000.0	Jr.
194	Data Scientist	110000.0	Jr.
556	Data Scientist	NaN	Jr.
576	Data Scientist	100000.0	Jr.
582	Data Scientist	160000.0	Jr.
598	Data Scientist	160000.0	Jr.
699	Data Scientist	100000.0	Jr.
712	Data Scientist	140000.0	Jr.
805	Data Scientist	120000.0	Jr.
839	Data Scientist	140000.0	Jr.
1082	Data Scientist	110000.0	Jr.
1151	Data Scientist	190000.0	Jr.
1203	Data Scientist	87360.0	Jr.
1250	Data Scientist	NaN	Jr.

```
print("Mean Entry-Level Data Scientist Salary: ")
Mean Entry-Level Data Scientist Salary: 126736.0
```

To synthesize my findings, there is a multitude of pathways that students can follow after earning their bachelor’s degree, including grad school. Many different jobs are available in the data industry as well, and their salaries are relatively high. It is also evident that the INFO curriculum could be tweaked to accommodate the changing job field, to incorporate teachings of SQL as well as more software like Tableau and Power\_BI. I think that to satisfy more on the data science side of things, more statistics should also be incorporated into the curriculum.