Can I be a Data Analyst or Data Scientist?

DATA 6505: Data Munging in Python
By: Caroline Donovan and Patrycja Ciemniewska
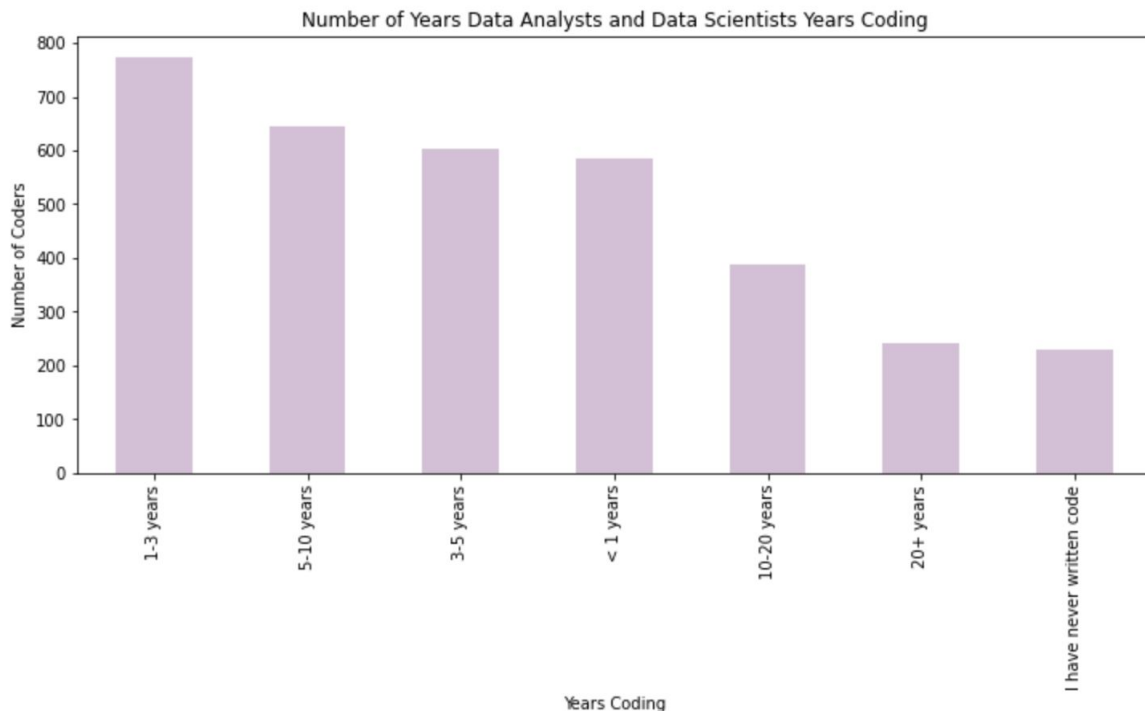
December 15th, 2022

Section 4:
Main Analysis- Analysis 1

Let's Take a Look at...

➔ **Data Analysts**
➔ **Data Scientists**

Q1: How Many Years Have Data Analysts and Data Scientists Been Coding?

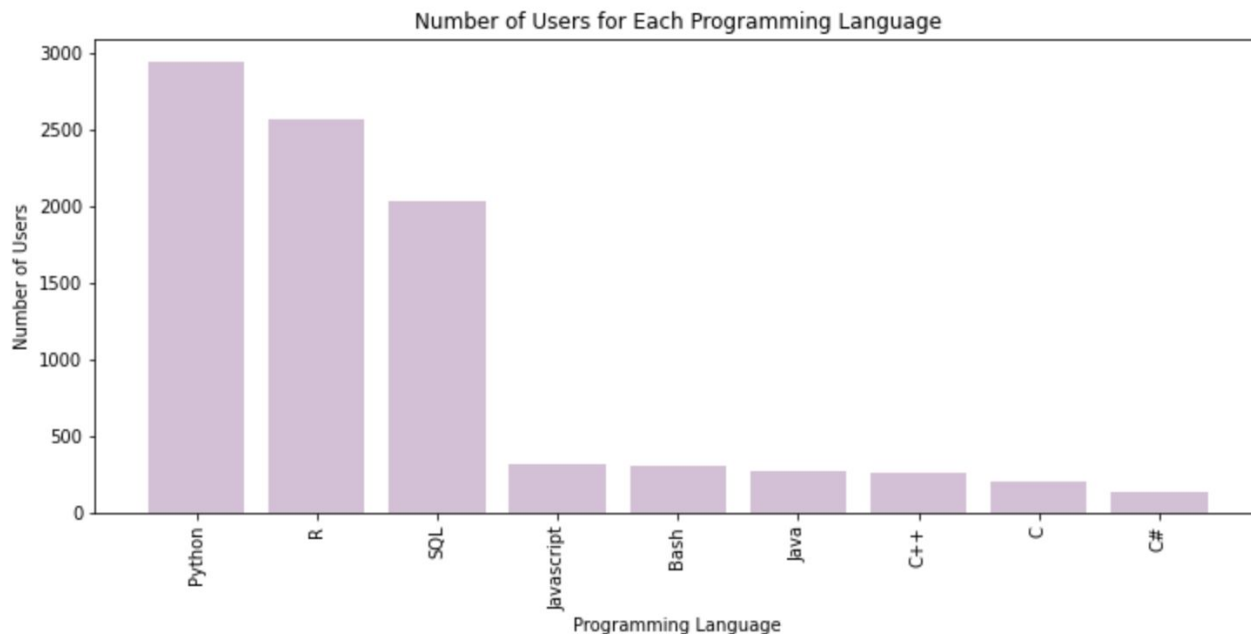Number of Years Data Analysts and Data Scientists Years Coding

💡 **Insights:**
When looking at years Data Analysts and Data Scientists have been coding, the most popular is between 1-3 years- we are in luck!

This makes sense since the roles are prominently filled by more entry level to mid level positions. People would normally not be in these roles for an extended amount of time such as 20+ years coding.

Fairfield
DOLAN
Fairfield University
Dolan School of Business

Q2: What Programming Language's are the Most Popular Amongst Data Analysts and Data Scientists?
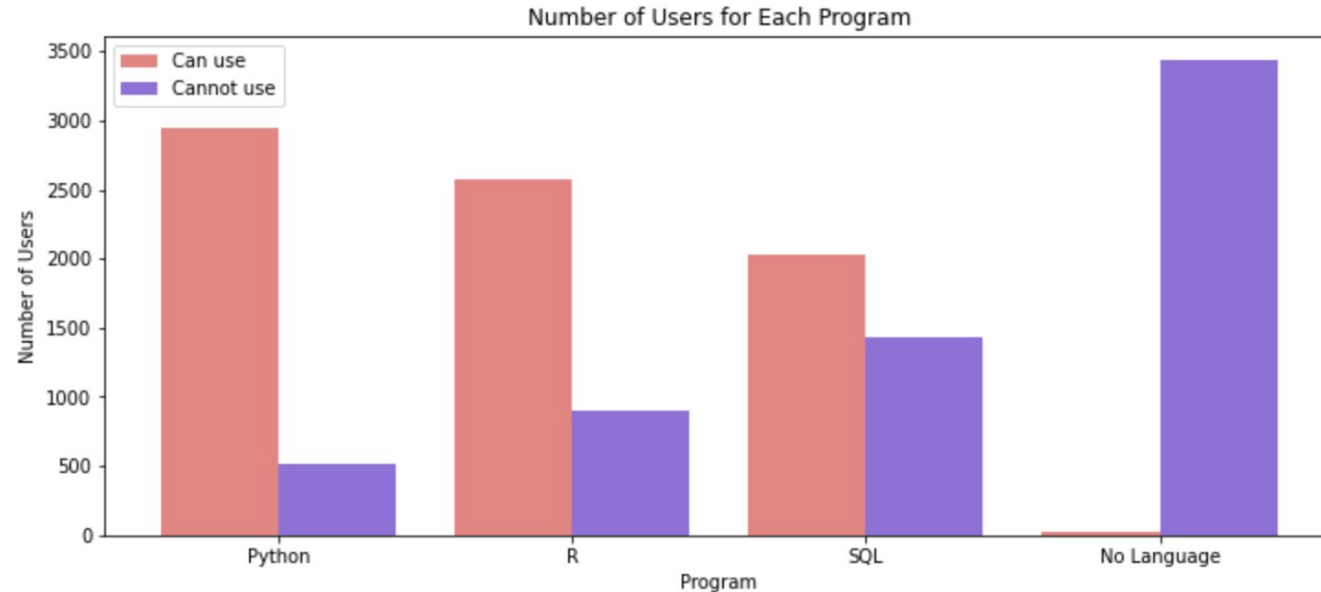
💡**Insights:**
When looking at programming languages used Data Analysts and Data Scientists mostly use Python, R, and SQL. So they took DATA 6510, DATA 6505, and DATA 6540?

If you have ever asked yourself if there are better languages out there- ask no further!

Number of Users for Each Programming Language

Q3: How Do Our Coding Abilities Compare Against Data Analysts and Data Scientists?

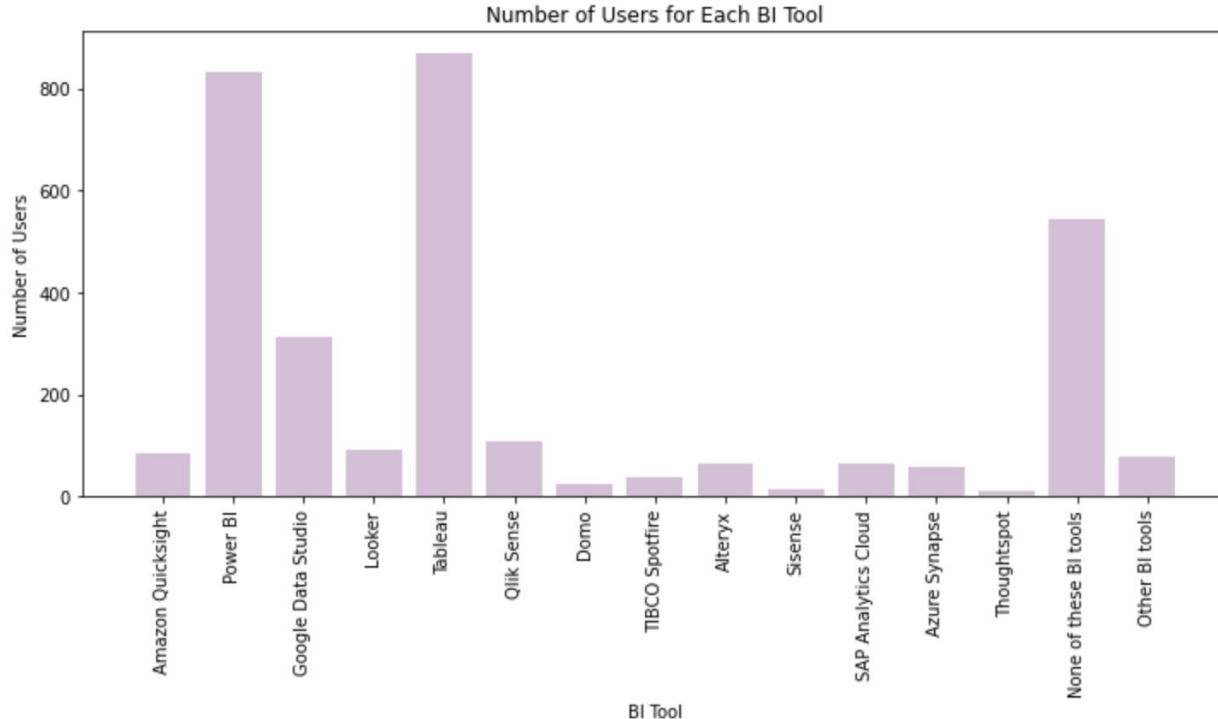Number of Users for Each Program

💡**Insights:**
Fairfield University MSBA program offers courses teaching Python, R, and SQL.

This allows us students to be prepared to enter these roles.

When looking at SQL, we can conclude this will help differentiate us from other students.

Fairfield
DOLAN
Fairfield University
Dolan School of Business

Q5: What BI Tools do Data Scientists and Data Analysts Use?

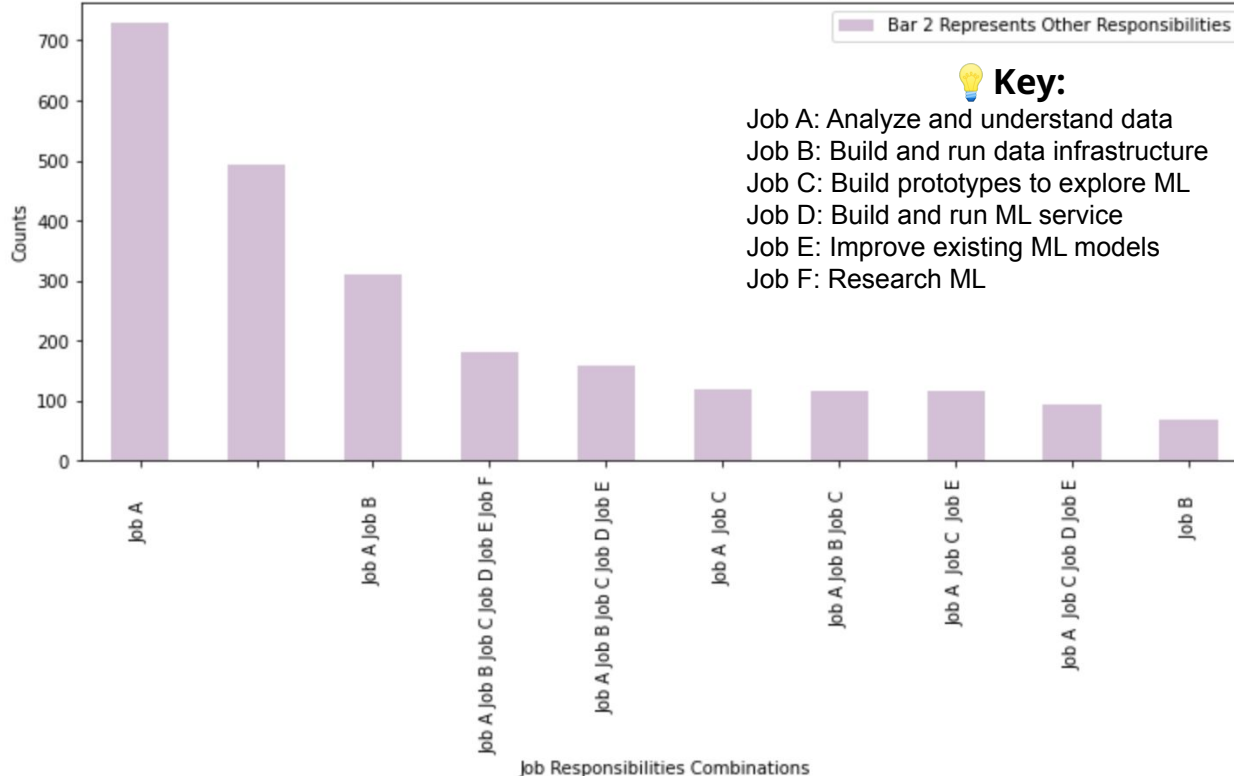Number of Users for Each BI Tool

💡 **Insights:**
It is interesting that Tableau is the most commonly used BI tool followed by Power BI.
Did someone say 6540 already?

Additionally it is interesting that the BI tool coming in 3rd place is not using BI tools at all. So if not DATA 6540, maybe 6520? Or 6545? Not all Analysts and Scientists do BI!

Fairfield
DOLAN
Fairfield University
Dolan School of Business

Q7: What are the Most Popular Combinations of Job Responsibilities Amongst Data Analysts and Data Scientists?

Top 10 Data Scientist and Data Analyst Responsibilities

Bar 2 Represents Other Responsibilities

💡**Key:**
Job A: Analyze and understand data
Job B: Build and run data infrastructure
Job C: Build prototypes to explore ML
Job D: Build and run ML service
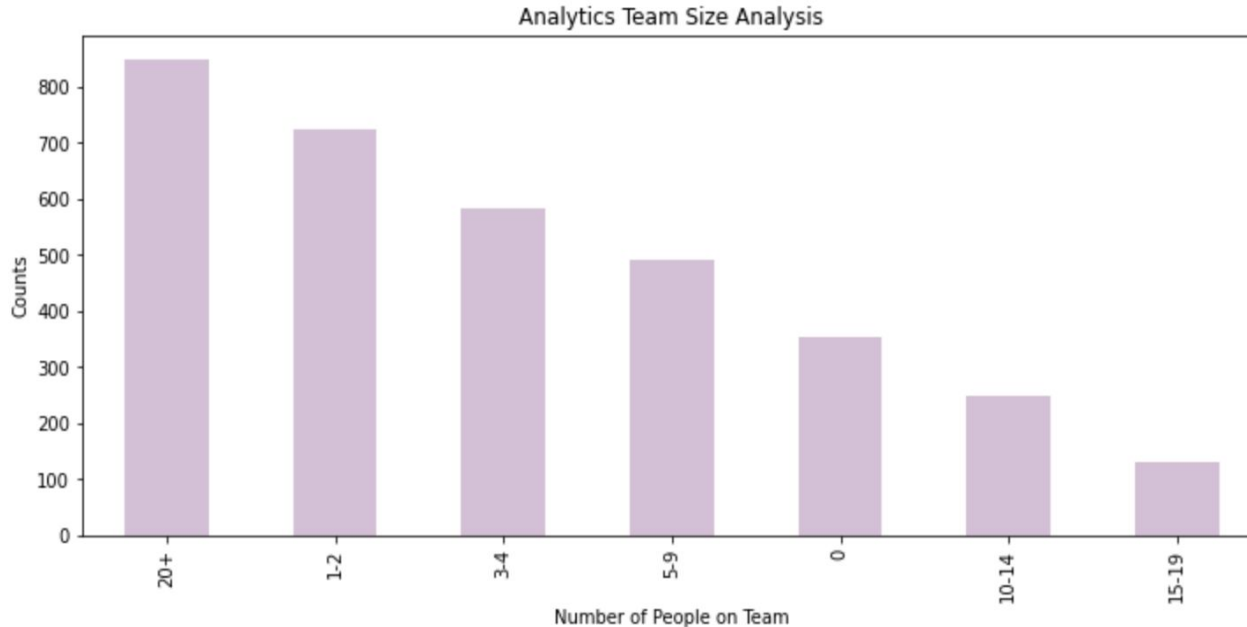Job E: Improve existing ML models
Job F: Research ML

💡**Insights:**
Since the roles we selected to look at are Data Analysts and Data Scientists it is expected Job 1 is the most popular. What is interesting that Job 1 and Job 2 together are performed together more frequently then Job B alone. This could be because we are not looking at Data Engineers in this data.

Q8: How Big are the Analytics Teams?

Analytics Team Size Analysis

💡 **Insights:**

What stands out most is that the top 2 responses for how many people are performing the same role as you. Either people are working by themselves or with only one other person or the polar opposite where they are working on large teams.

Why might this be?

Q9: How Does the Amount of People Differ for a Data Analyst vs. a Data Scientist?

```
Job Title                                                      Q26
Data Analyst (Business, Marketing, Financial, Quantitative, etc)  1-2   352
                                                               20+   326
                                                               3-4   249
                                                               0     231
                                                               5-9   189
                                                               10-14  94
                                                               15-19  51
Data Scientist                                                 20+   524
                                                               1-2   371
                                                               3-4   335
                                                               5-9   302
                                                               10-14 153
                                                               0     124
                                                               15-19  78
Name: Q26, dtype: int64
```
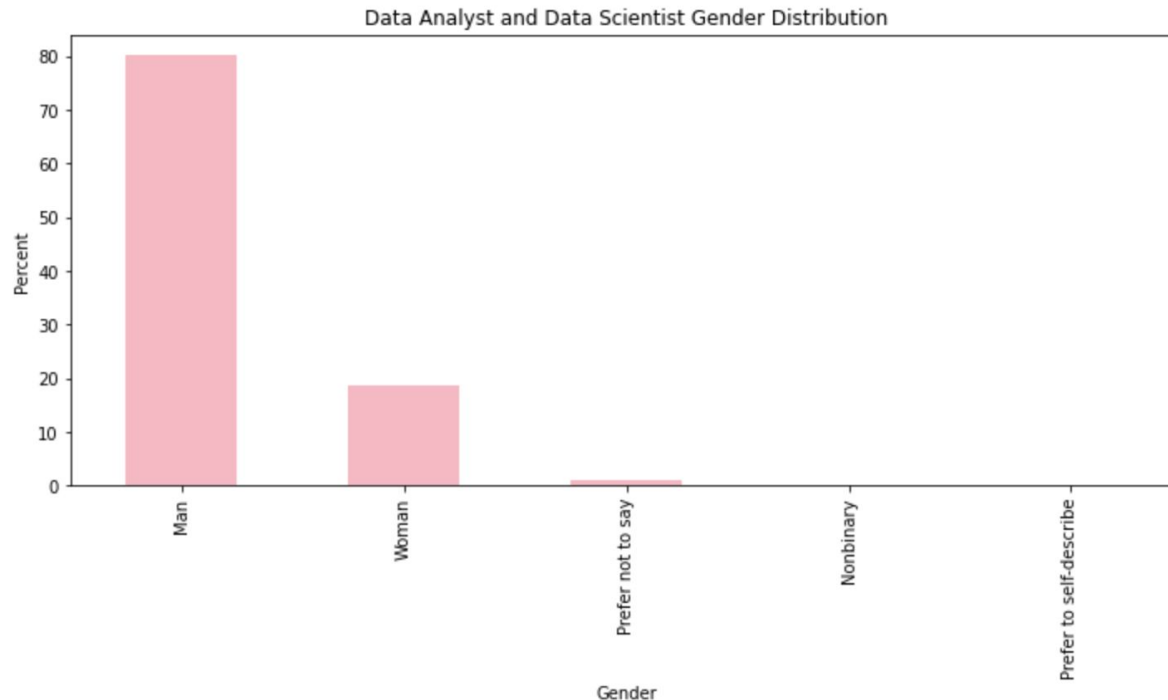
💡 **Insights:**
Let's not forget that Data Analysts and Data Scientists are not the same…. They are grouped for this presentation because we could do both.

Scientists are more likely to be surrounded by a group of more than 20 people, whereas analysts are more likely to be lone wolves.

What would you prefer to be?

Fairfield
DOLAN
Fairfield University
Dolan School of Business

Q10a: What are the Distributions for Gender for Data Analysts and Scientists?

Data Analyst and Data Scientist Gender Distribution
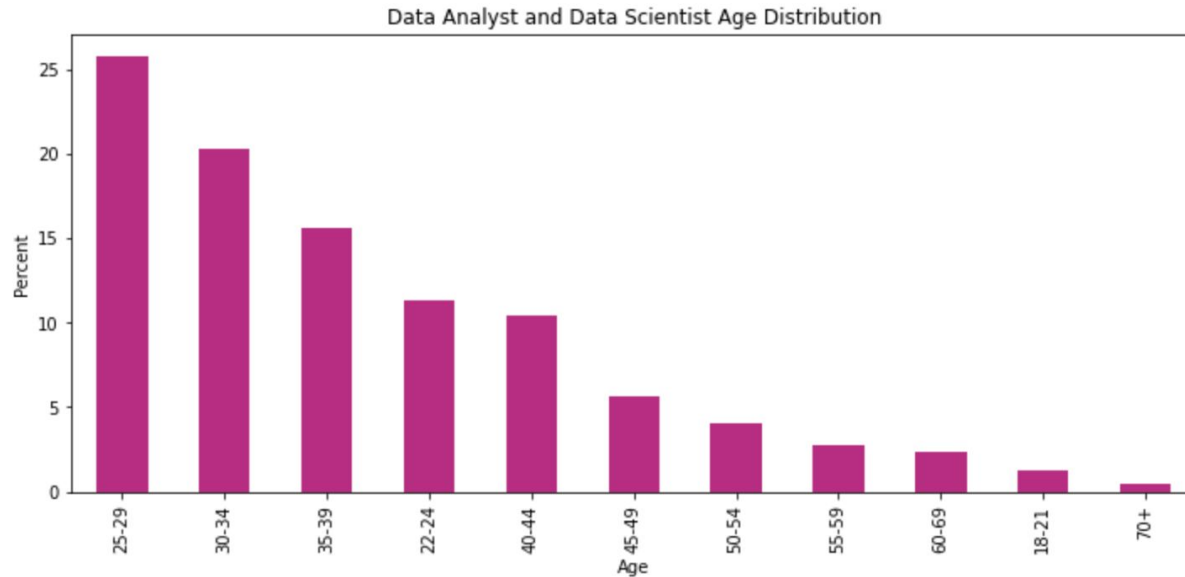
💡**Insights:**
Unfortunately, the job titles of Data Analyst and Data Scientist are still dominated mostly by males.

However, there is a trend here..... To be continued!

Fairfield
DOLAN
Fairfield University
Dolan School of Business

Q10b: What are the Distributions for Age for Data Analysts and Scientists?

Data Analyst and Data Scientist Age Distribution

💡 **Insights:** The most popular age group for a Data Analyst or Data Scientist is 25-29. This makes sense considering the amount of time it takes to develop the necessary skills!

Q11: What is the Average Salary for a Data Analyst and Data Scientist?

```
[101]   1 DA_DS_df['Salary_averages'].mean().round(2)
```
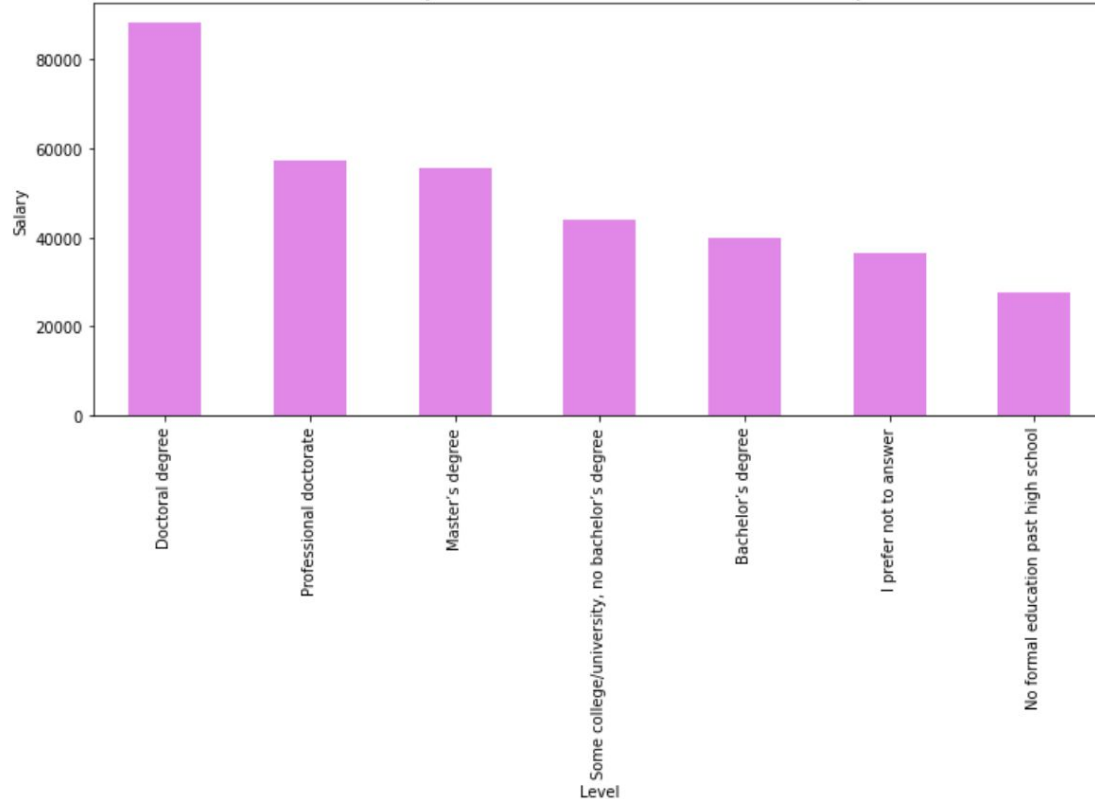
53272.44

💡**Insights:**
For reference, if you are thinking this looks a little low- this average is taking into account all of the co-ops, graduate assistantships, research jobs, and internships completed where people are called 'Analysts'. Synchrony and The Hartford people- you! The average represents several different incomes

Q12: How Does Level of Education Affect Salary?

Data Analyst and Data Scientist Education Level vs Salary

💡 **Insights:**

If you have ever asked yourself, 'Did I need to do the Master's?' - here you go.

The answer is- good choice! The Master's degree moves your salary average up nearly $20,000 compared to the Bachelor's.

Thinking of continuing? We recommend going for your Doctorate considering a Professional Doctorate is fairly similar to Master's!
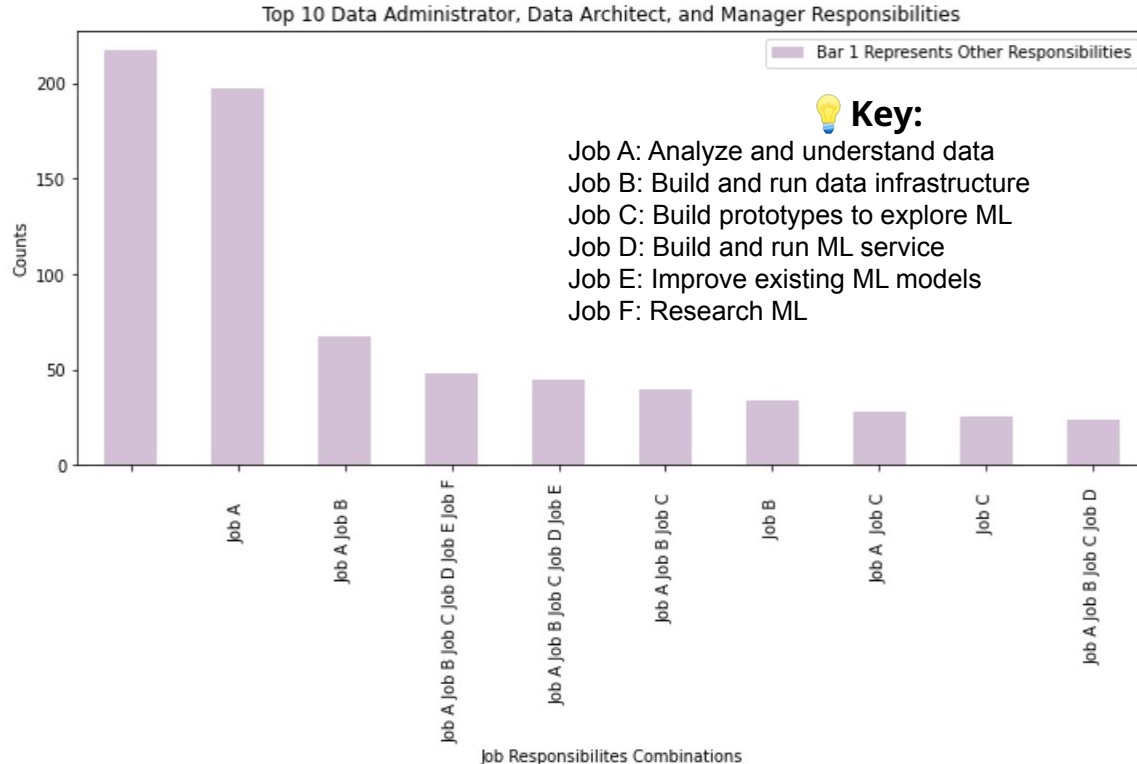
Fairfield
DOLAN
Fairfield University
Dolan School of Business

Section 5:
Main Analysis- Analysis 2

Let's Take a Look at... **Upper level positions!**

→ Manager (Program, Project, Operations, Executive-level, etc)
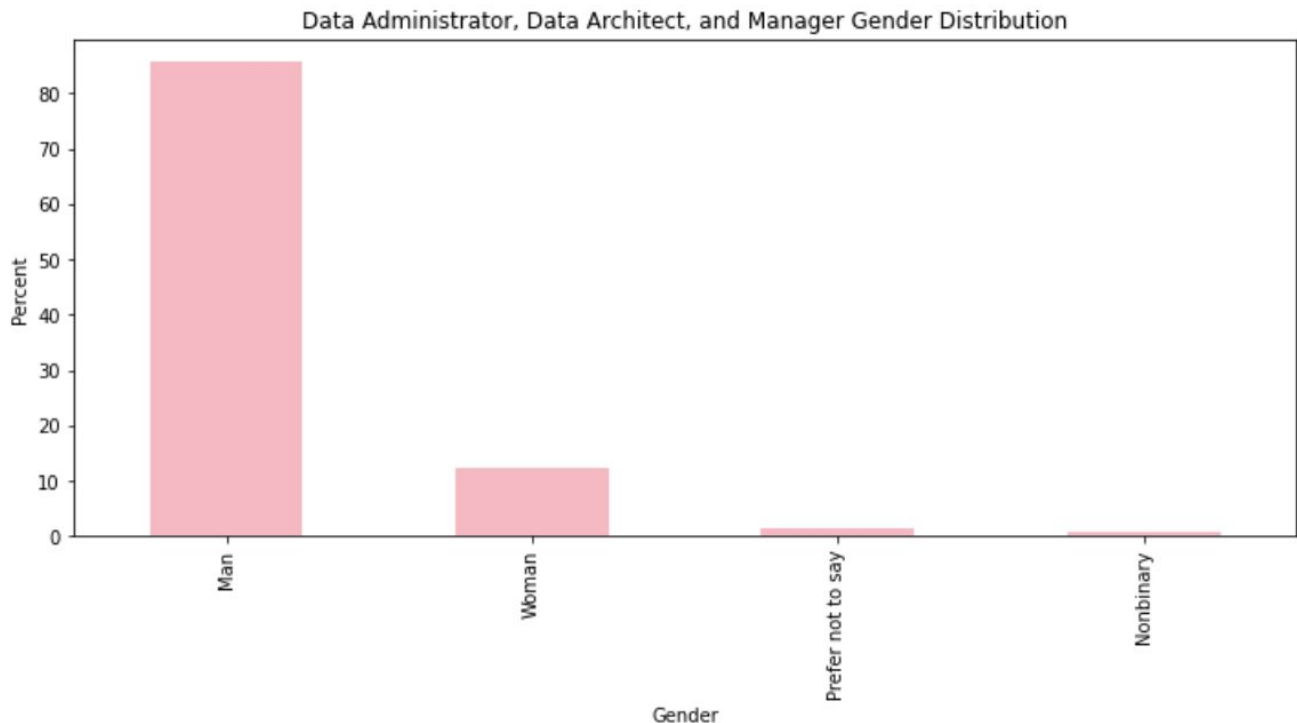→ Data Architect
→ Data Administrator

Q1: What do Job Responsibilities look like for Upper Level Positions in Analytics?

Top 10 Data Administrator, Data Architect, and Manager Responsibilities

💡**Key:**
Job A: Analyze and understand data
Job B: Build and run data infrastructure
Job C: Build prototypes to explore ML
Job D: Build and run ML service
Job E: Improve existing ML models
Job F: Research ML

💡**Insights:**
We can conclude that this dataset doesn't tell us much about upper level roles. We can see that the majority of these job titles do not align with any of the given responsibilities. That being said, we can still conclude that job A is important for any role in data analytics.

Q2a: How does Gender Change for Upper Level Positions Compared to Analysts and Scientists?

Data Administrator, Data Architect, and Manager Gender Distribution

💡 **Insights:**
When Remember that trend we said we would come back to? It looks like in the higher up roles there are even less women - meaning a change is happening as we speak!
We would hope if we take a look at this in 2032, that women are making up more and more of this field!

Q2b: How does Age Change for Upper Level Positions Compared to Analysts and Scientists?

Data Administrator, Data Architect, and Manager Age Distribution

💡**Insights:**
As expected, it is going to take a few years to reach the upper level positions.

That being said, it does not mean that you won't make a pivot in between your first role and reaching an upper level!

Q3: How Does Salary Change Moving up to an Upper Level Position?

```
[102]  1 Upper_level_df['Salary_averages'].mean().round(2)

       93403.71
```

```
[100]  1 Upper_level_df['Salary_averages'].value_counts().head(10).plot(kind='bar', color='deeppink')
       2 plt.title('Top 10 Salary Responses')
       3 plt.xlabel("Salary")
       4 plt.ylabel("Counts");
       5
```


Top 10 Salary Responses

💡**Insights:**

Finally, the big sell- the more you move up, the more money you will make! But remember- years coding and education help with this too!

So what's the **story**?

Can we be a Data Analyst or a Data Scientist?

The story is…….

If you have only been coding for 1-3 years…. You **CAN** do it!

If you are a man OR a woman in Analytics…… You **CAN** do it!

If you know Python, SQL, and R…….. You **CAN** do it!

If you want to do BI…… You **CAN** do it!

If you are still unsure what exactly you want to do just know…… You **CAN** do it!

If you are willing to keep learning and growing…… You **CAN** do it!

Thank you for listening!

Questions?