

# Story 4 – How much do we get paid?

By: Shoshana Farber

# Objective:

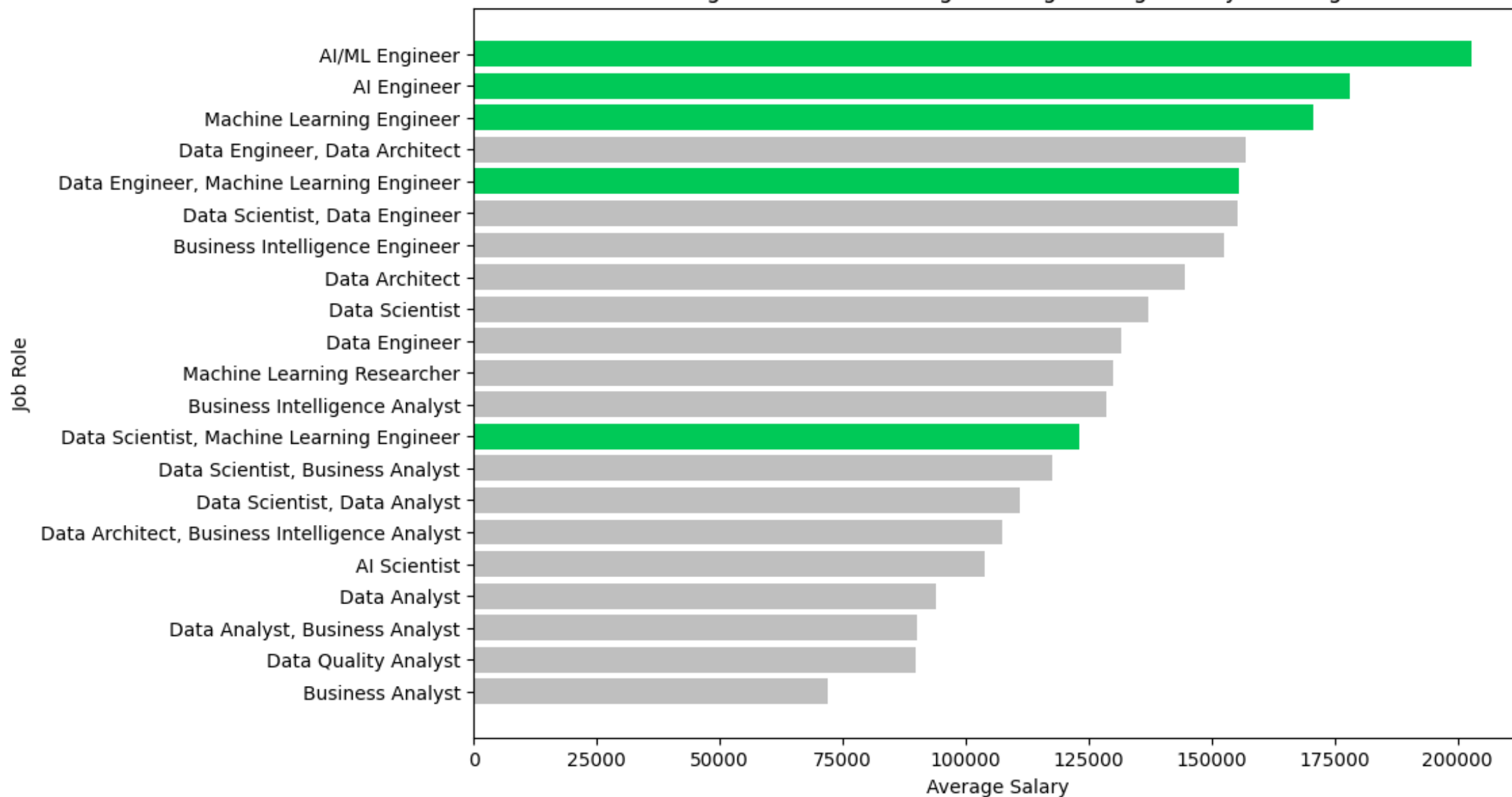
- ▶ I have introduced the term "Data Practitioner" as a generic job descriptor because we have so many different job role titles for individuals whose work activities overlap including Data Scientist, Data Engineer, Data Analyst, Business Analyst, Data Architect, etc.
- ▶ For this story we will answer the question, "**How much do we get paid?**" Your analysis and data visualizations must address the variation in average salary based on role descriptor and state.

# Sources:

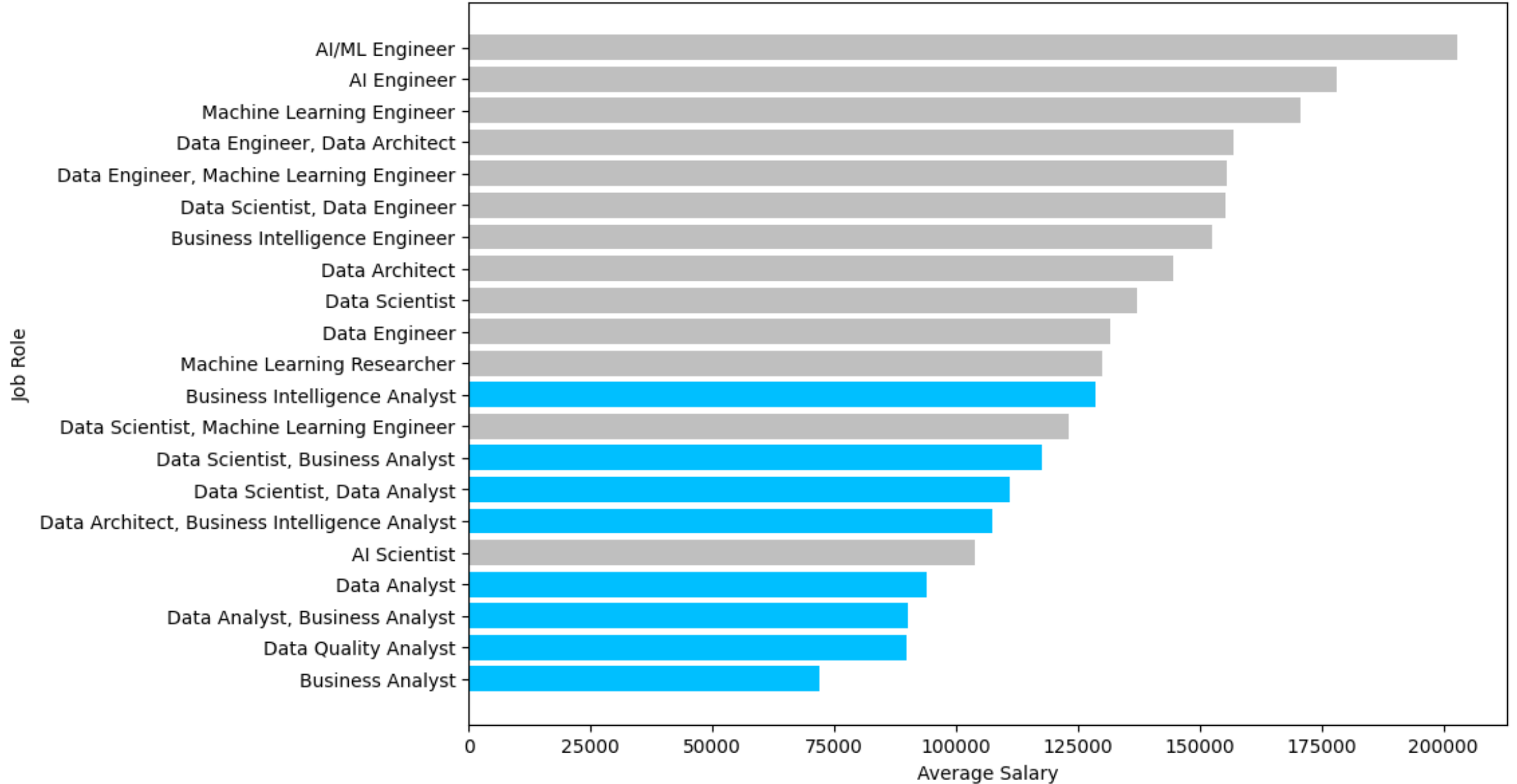
- ▶ Data was initially acquired through scraping from [ai-jobs.net](https://ai-jobs.net).
- ▶ This data was supplemented using a dataframe of already scraped jobs from Kaggle
  - ▶ [https://www.kaggle.com/datasets/thedevastator/jobs-dataset-from-glassdoor?select=salary\\_data\\_cleaned.csv](https://www.kaggle.com/datasets/thedevastator/jobs-dataset-from-glassdoor?select=salary_data_cleaned.csv)

# Salary vs. Job Role

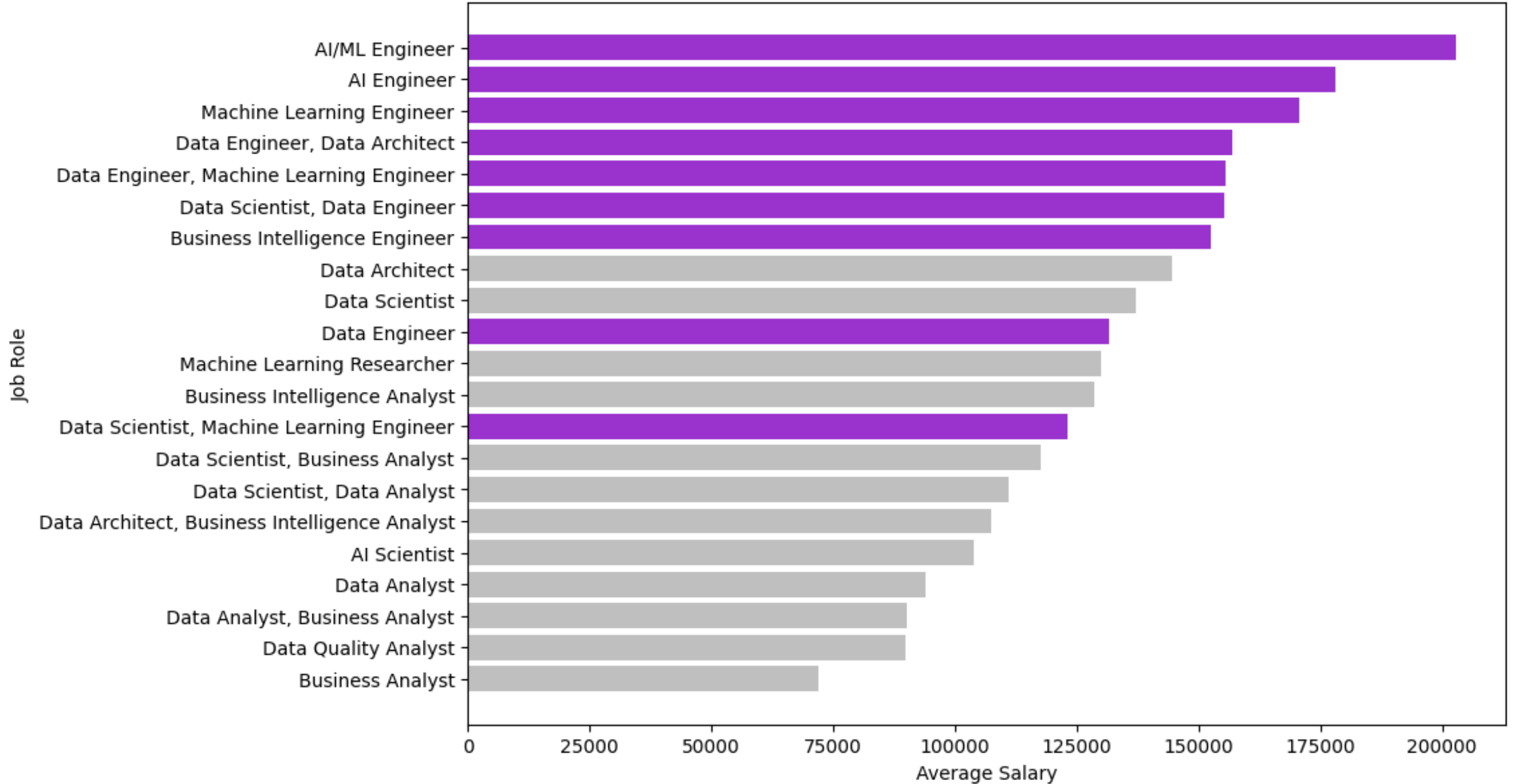
Machine Learning and Artificial Intelligence Engineers generally have higher salaries



## "Analysts" are generally paid the least



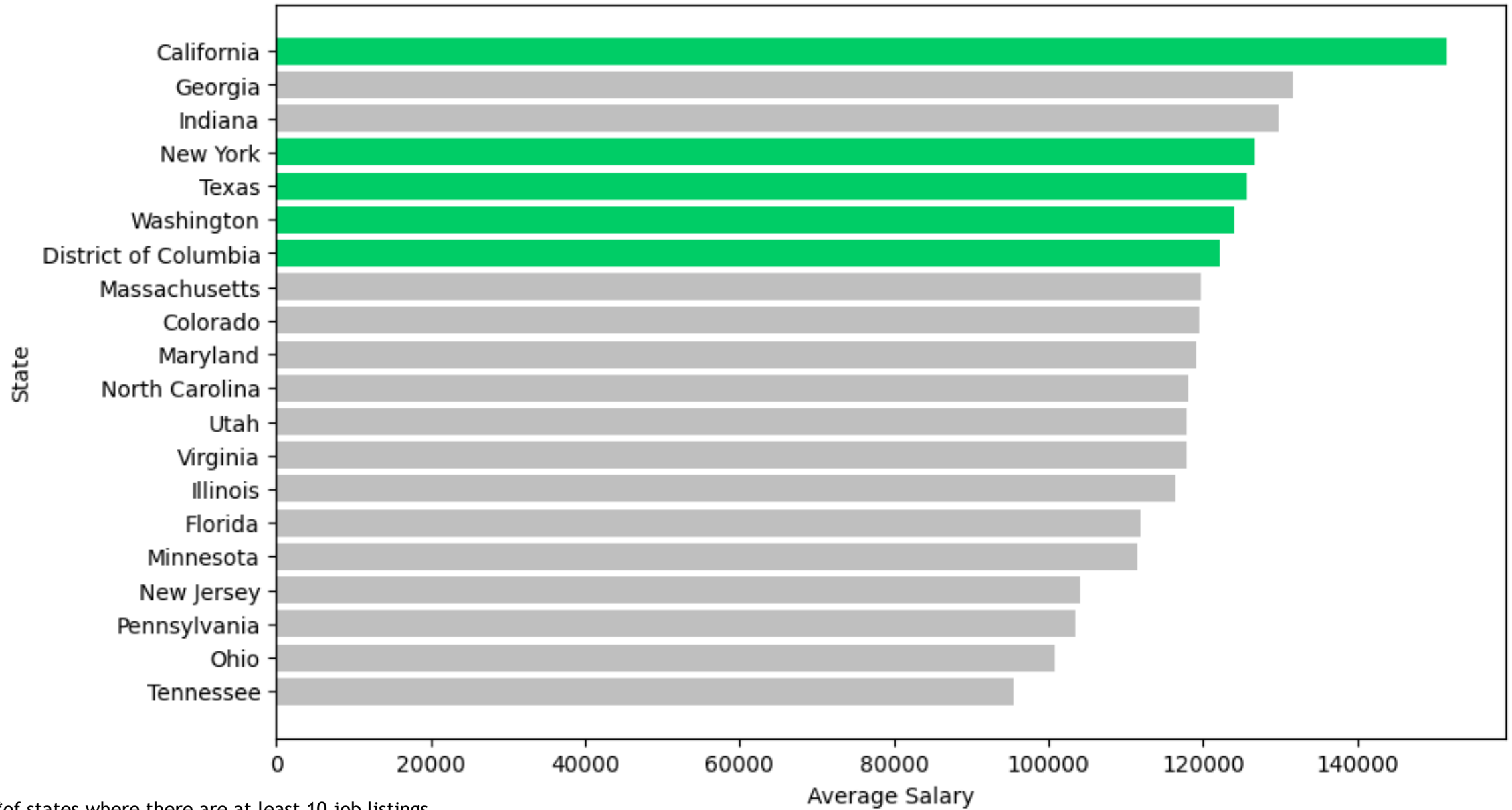
"Engineers" tend to have higher salaries



# Salary vs. State



# California, New York, DC, Texas, and Washington are amongst the highest paying states\*



\*of states where there are at least 10 job listings

# Conclusions:

- ▶ Machine Learning Engineers tend to have the highest salaries.
- ▶ “Analysts” tend to be paid the least while “engineers” are paid more.
- ▶ California, New York, Texas, and DC are among the highest paying states, which makes sense as cities within these states tend to be seen as technology hubs (such as Los Angeles CA, San Francisco CA, New York City NY, Seattle WA, Austin TX, Dallas TX, and Washington DC).

# Notes:

- ▶ This assignment was done using job listing scraped from a data related job site.
- ▶ The initial scrape consisted of almost 4,000 job listings. The Glassdoor job listings which were used to supplement consisted of about 1,000 additional job listings.
- ▶ Not all of the listings were based in the United States. As such, the data had to be filtered down for the salary vs. state analysis. The results may not truly reflect the average salaries for each state, as some states had very or only one listing which skewed the data towards the single salary for these states. For this reason, the data was filtered for only states with more than 10 listings.
- ▶ Ideally I would have visualized salary vs. state with a choropleth, but I did not pull enough data for this visualization to be meaningful.