



Insights into Data Analyst Roles

S. Michael Stewart

May 30, 2023

OUTLINE



- Executive Summary
- Introduction
- Methodology
- Results
 - Visualization – Charts
 - Dashboard
- Discussion
 - Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY



- This study looked at demographics and job information for programming and data analysts
- Webscraping and APIs were used to gather data to answer questions like
 - What is the pay range and average pay of data analysts?
 - What is the demographic profile of data analysts?
 - What languages, platforms, and services are data analysts using and want to learn/use in the coming year?
- Data was processed and exploratory data analysis (EDA) was used to identify potential correlations
- Dashboards were built to see overall trends
- Programming language seemed to have the biggest impact on salary/compensation, and the most popular languages to use and learn were those closely associated with full-stack development

INTRODUCTION



- For this project, I looked at the current state of data analyst roles
- The questions to be answered were
 - What is the pay range and average pay of data analysts?
 - What is the demographic profile of data analysts?
 - What languages, platforms, and services are data analysts using and want to learn/use in the coming year?
- To answer these questions, I first compiled data from web scrapping and survey responses

METHODOLOGY



- Data was first acquired through
 - Webscraping of the IBM SkillsNetwork website using the BeautifulSoup package
 - Using the Kaggle API
- Data from both sources was then processed to remove duplicate entries, and blank salaries were replaced with the average salary
- Exploratory data analysis (EDA) was conducted by determining what correlations existed between the different demographic profiles, such as age
- SQL queries were then used to generate plots of the different correlations
- Finally, data was summarized and visualized using IBM's Congos dashboard application

Results – Part 1

- The table to the right shows the correlation matrix from the EDA
- The converted compensation represents the yearly salary for respondents
- Self correlation (e.g. respondent vs respondent) is omitted as we expect a variable to correlate with itself

	Converted Compensation	Average Hours Worked Per Week	Average Hours a Week Doing Code Revision	Age
Converted Compensation		0.021143	-0.033865	0.105386
Average Hours Worked Per Week	0.021143		0.026517	0.036518
Average Hours a Week Doing Code Revision	-0.033865	0.026517		-0.020469
Age	0.105386	0.036518	-0.020469	

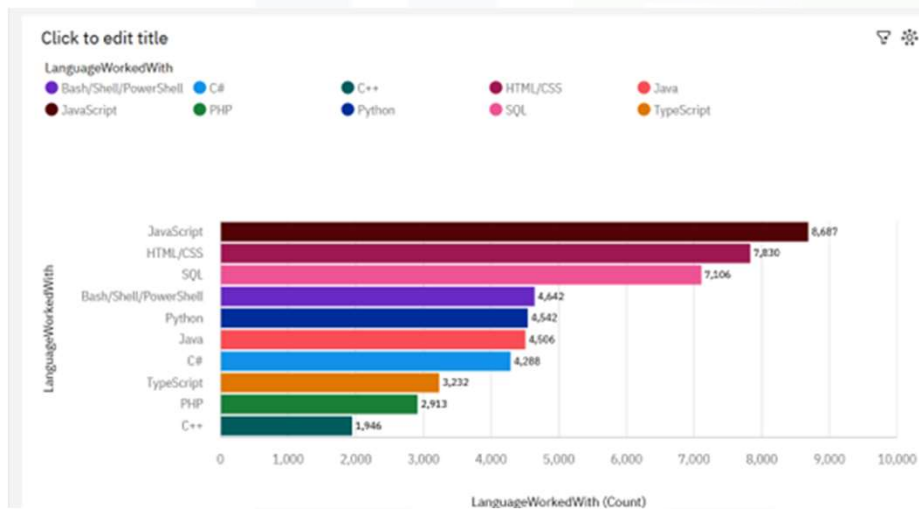
Results – Part 1

- From the table we see that compensation correlates most strongly with age
- While compensation correlates weakly with the average hours worked per week, it looks like compensation decreases with hours spent doing code revision

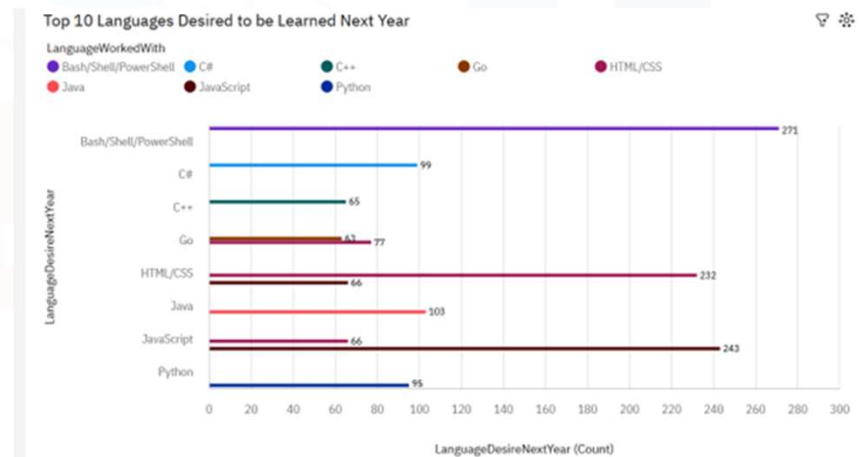
	Converted Compensation	Average Hours Worked Per Week	Average Hours a Week Doing Code Revision	Age
Converted Compensation		0.021143	-0.033865	0.105386
Average Hours Worked Per Week	0.021143		0.026517	0.036518
Average Hours a Week Doing Code Revision	-0.033865	0.026517		-0.020469
Age	0.105386	0.036518	-0.020469	

PROGRAMMING LANGUAGE TRENDS

Current Year



Next Year



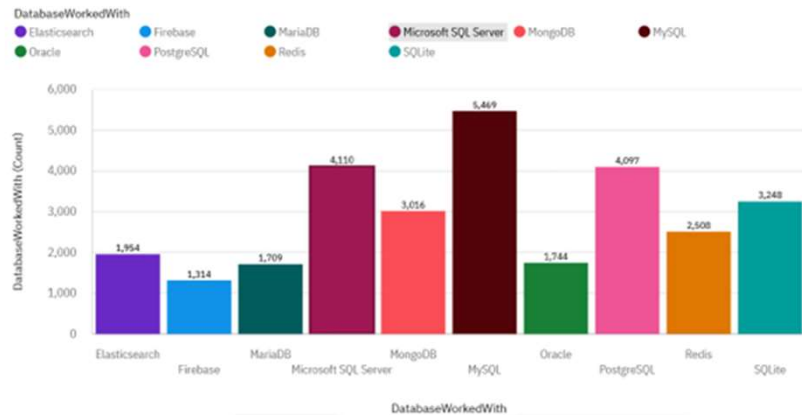
PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

- In the current year, the most popular language use was Javascript, followed by HTML/CSS then SQL
- Most respondents put BASH and Javascript as their top languages to learn in the coming year
- Of the popular languages currently being used and that programmers want to learn, the top three are front end languages like JavaScript, HTML, BASH/PowerShell
- The lower top-10 are all common backend languages, such as python
- In general, it looks like most programmer currently work or are looking to work in frontend/backend/full-stack

DATABASE TRENDS

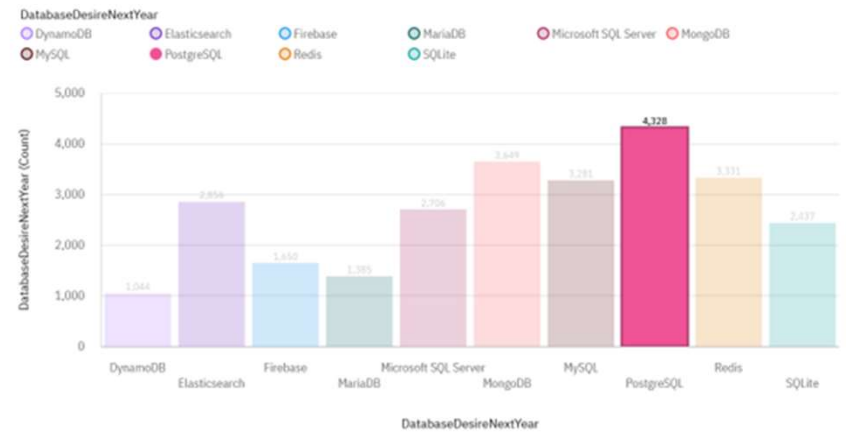
Current Year

Top 10 Databases Worked With



Next Year

Top 10 Databases Desired to be Learned Next Year



DATABASE TRENDS - FINDINGS & IMPLICATIONS

- The three most popular database languages are MySQL, MicrosoftSQL Server, and ProtegeSQL
- In the coming year, most programmers want to learn ProtégéSQL and MongoDB
- It is quite interesting that despite the growing number of non-categorical data sources, such as tweets and DSS streams, the most popular databases remain SQL servers
- There is some shift to NoSQL platforms, but in general SQL still reigns supreme

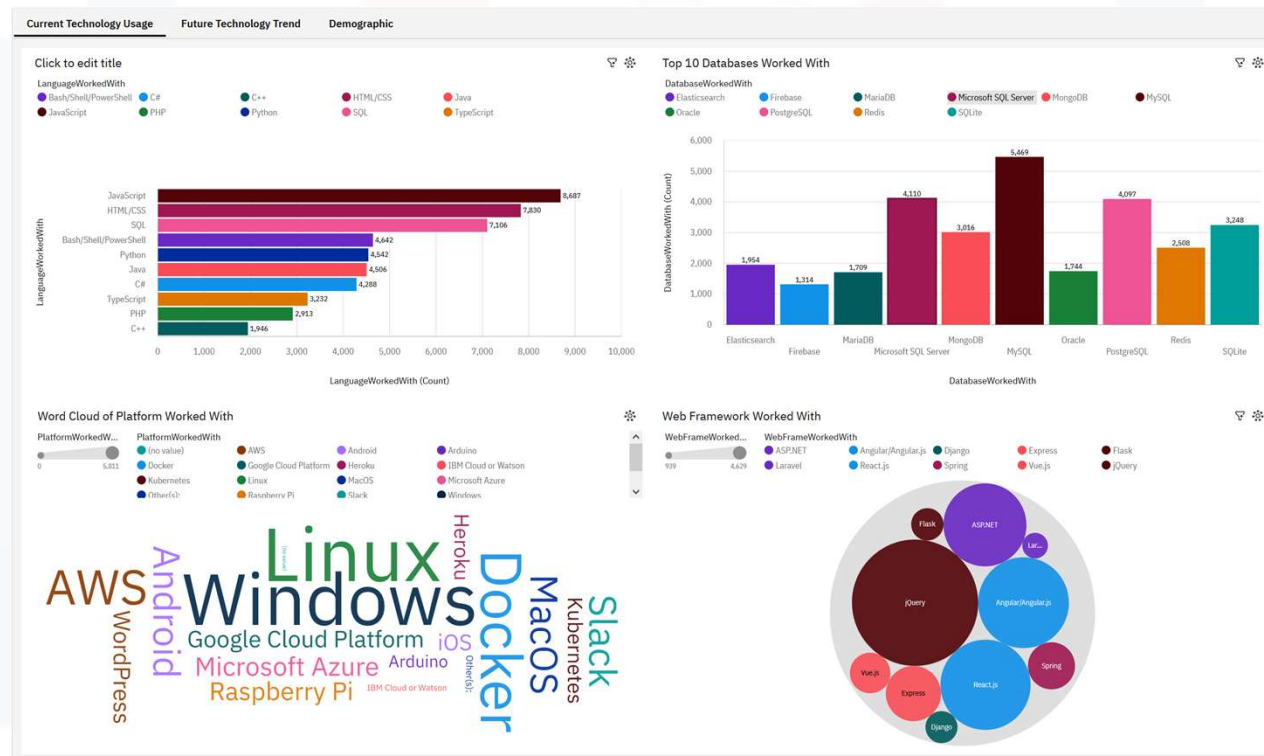
DASHBOARD



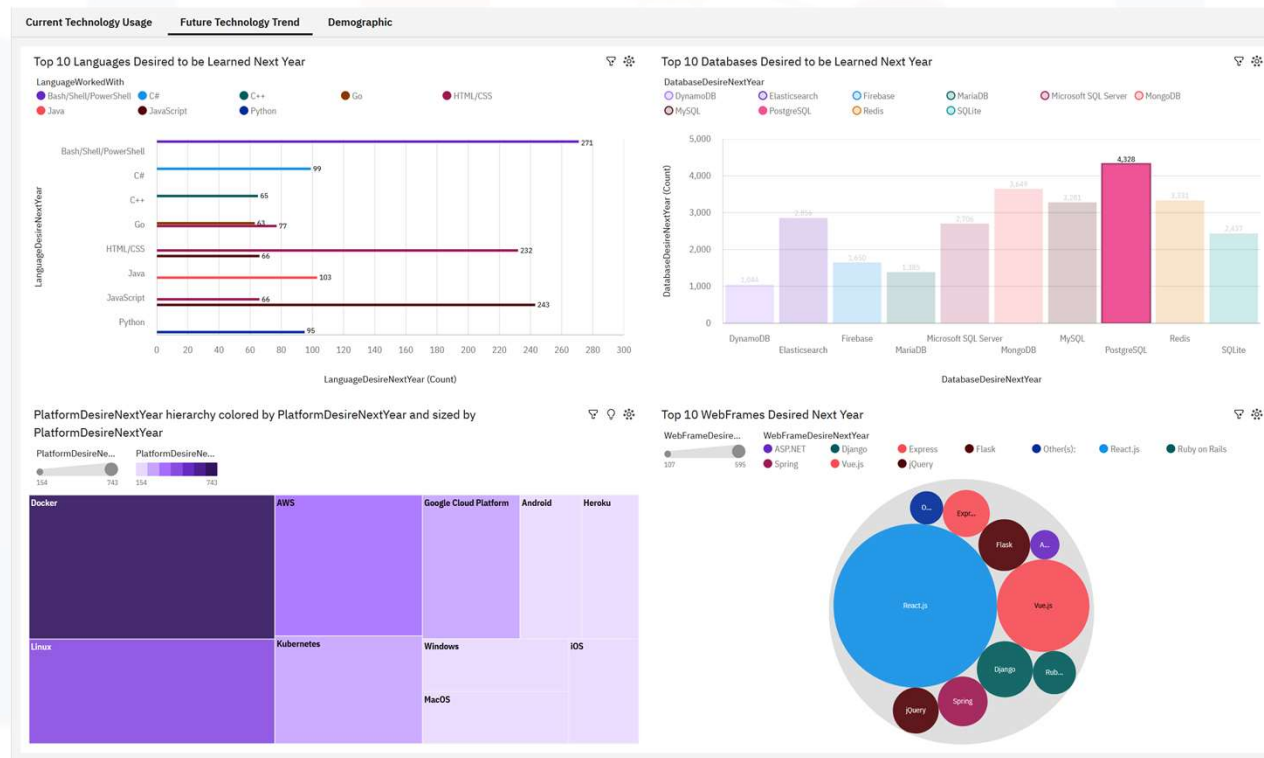
Link the dashboard is

https://us1.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=.my_folders%2FCapstoneDashboard&action=view&mode=dashboard&subView=model00000188626e5f76_00000000

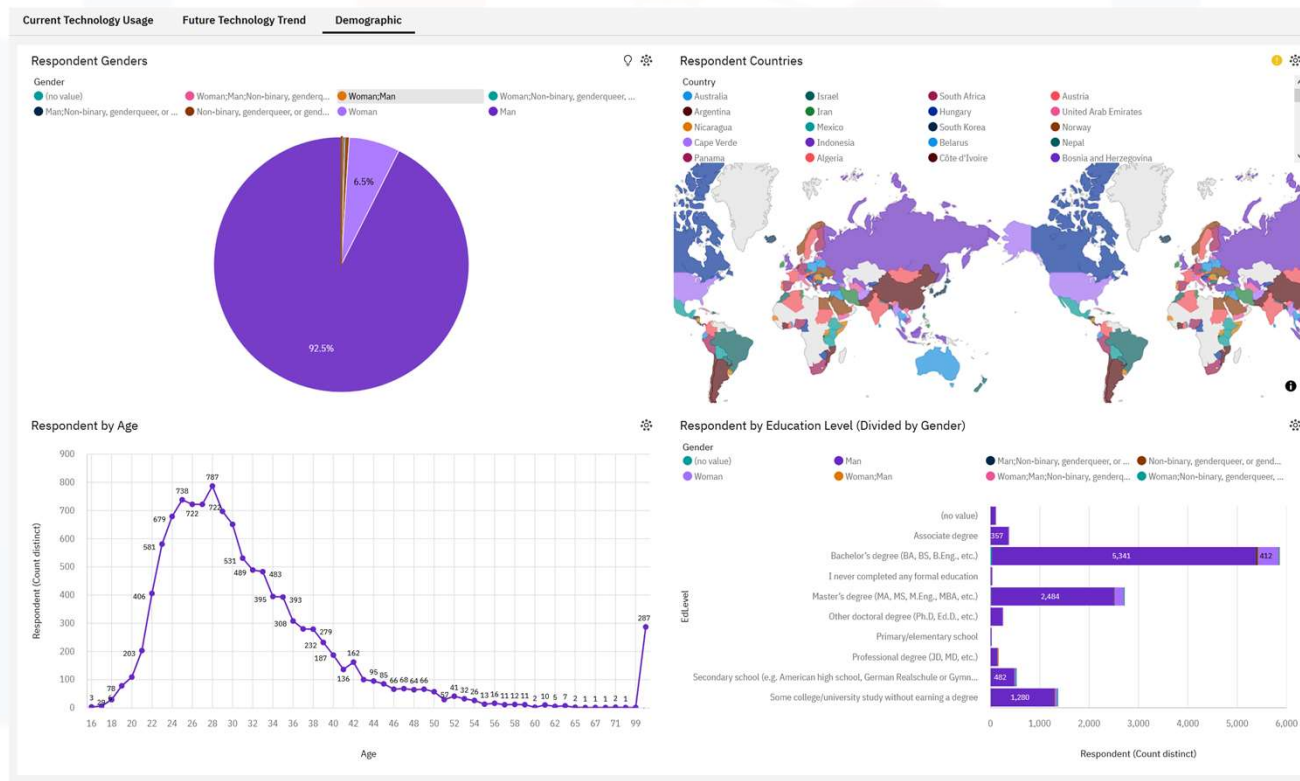
DASHBOARD TAB 1



DASHBOARD TAB 2



DASHBOARD TAB 3



DISCUSSION



- After looking into demographics and jobs information for programmer through both Webscraping and API requests several trends stood out
- The most popular and well-paying languages are those associated with front end development, such as Javascript and C++, as well as popular backend languages like Python and SQL
- While there is much overlap with these languages and jobs other than web development, it is telling how much overlap there is
- As for programmers themselves, they tend to be a homogeneous group consisting mostly of men between 20-40 with a bachelor's degree in computer science

OVERALL FINDINGS & IMPLICATIONS

- From this study, key findings include that most of the programming languages used today are strong associated with web development
- Most programmers are men in their 20s-40s with backgrounds in computer science
- Age seems to be the strongest correlator with pay beyond the language being used
- Due to the versatility of languages like Python and SQL, it makes sense why they would be so popular – they have applications to more than just web development but are still integral to web development
- If someone is looking to be successful in the data analyst field, they are best to learn Python and SQL before moving into more frontend languages like Dash and Swift

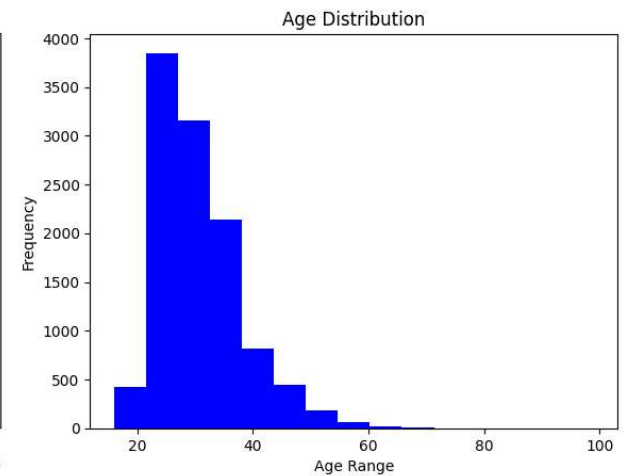
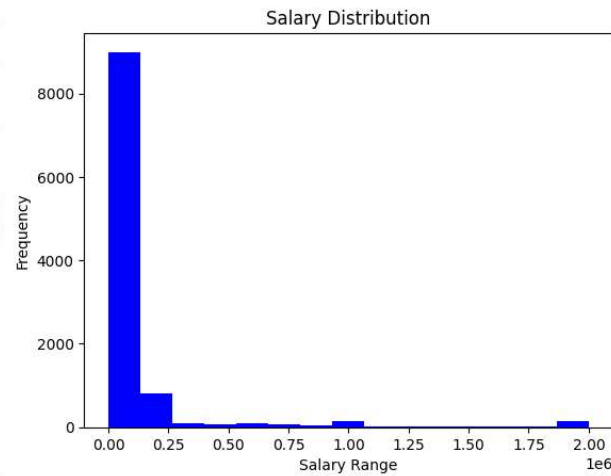
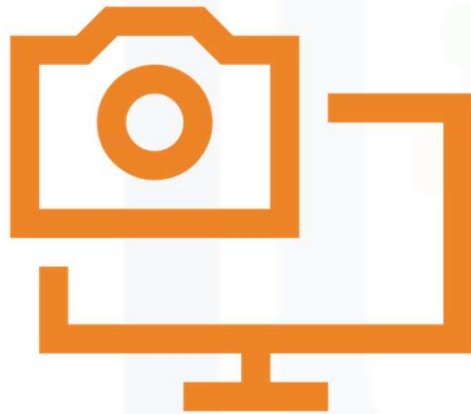
CONCLUSION



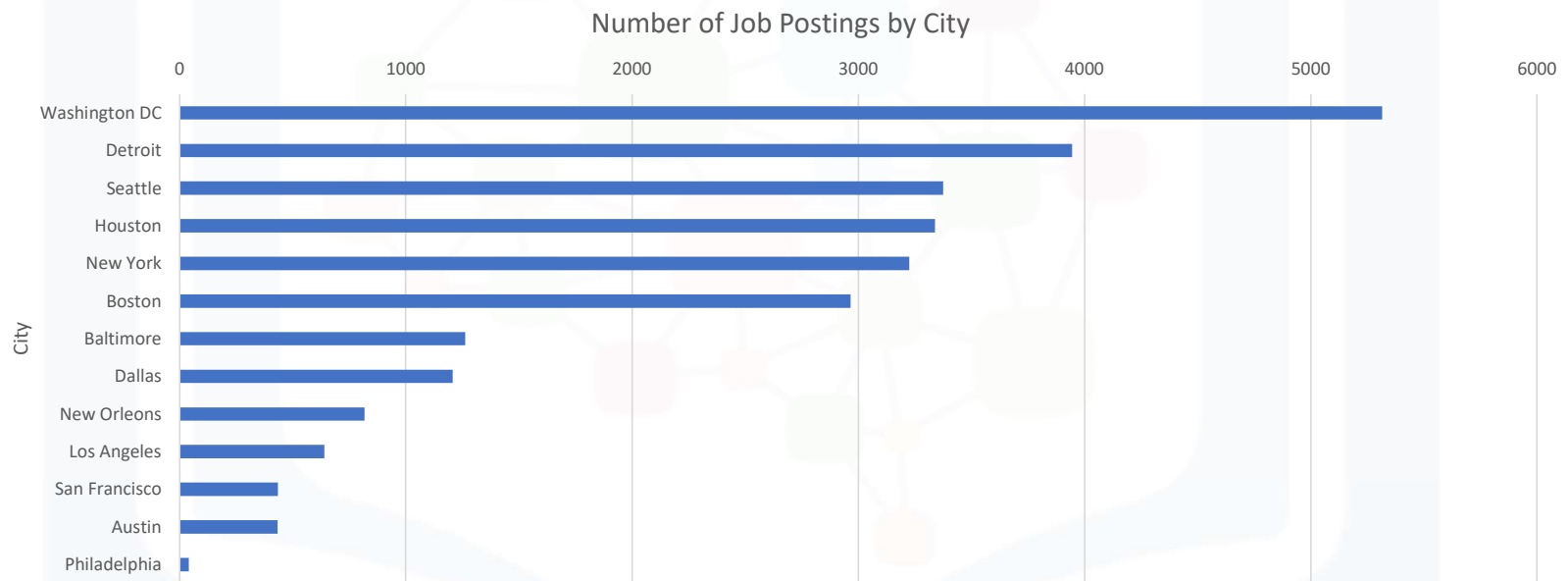
- The best paying languages are those that use Python, Swift, C++, and Javascript
- Beyond the languages learned, age seems to be the second largest factor in overall pay
- Most programmers are men in their 20s-40s
- Most programmers looking to grow their skills are looking for languages with applicability to full-stack, or frontend/backend development

APPENDIX

- Salary Distributions by Age



JOB POSTINGS



POPULAR LANGUAGES

