



Data Analyst Capstone

William Chaudoin

04/28/23

OUTLINE

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

EXECUTIVE SUMMARY

- ▶ Data Science and Information Technology are ever changing fields. With new applications and methods being developed constantly.
- ▶ To enter these career fields, requires a specific skillset and expertise in a variety of languages, programs, and statistical methods.
- ▶ To become the best candidate, an individual needs to know not only what the most desired current skills are but what will be the most desired skills in the future.
- ▶ This presentation will answer these questions by analyzing and visualizing current data associated with existing professionals within the field.



INTRODUCTION

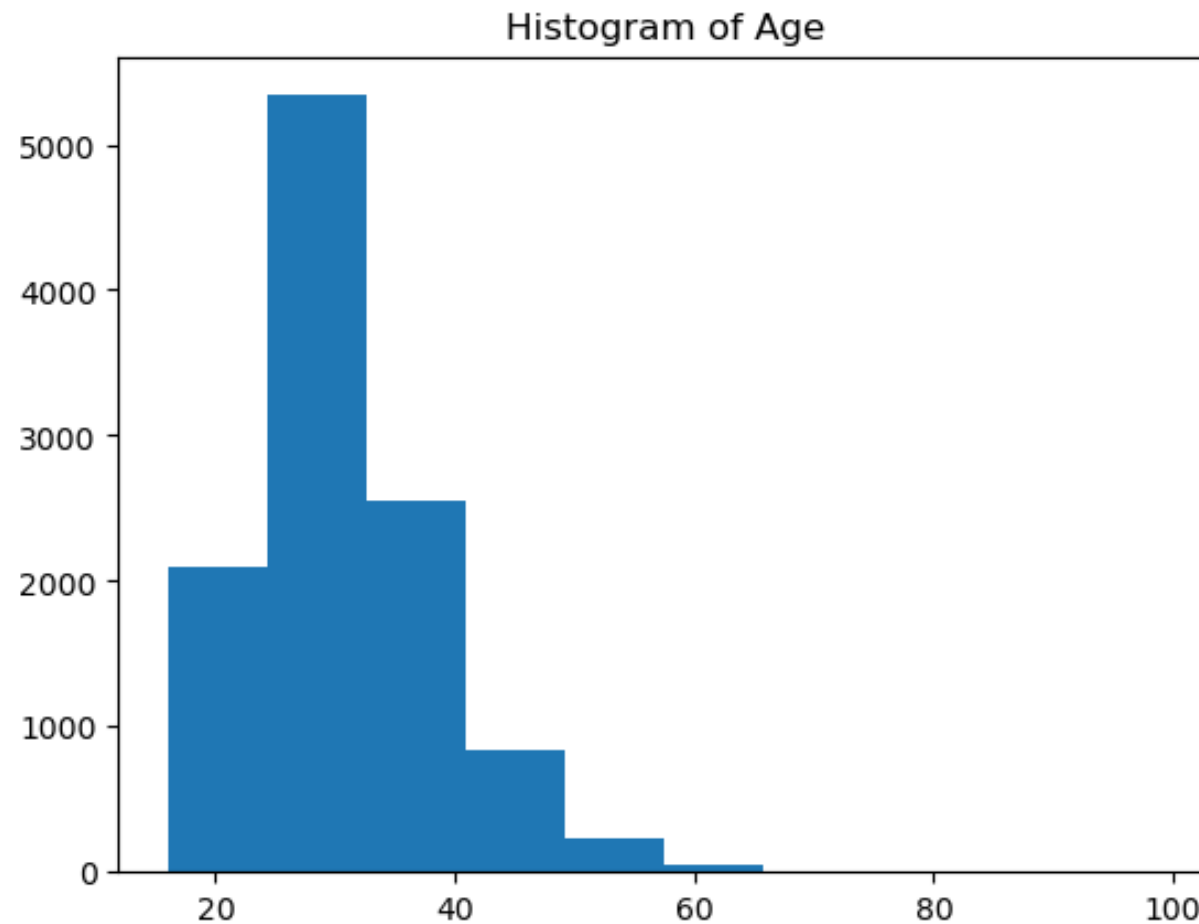
- This presentation is for individuals who wish to enter the Data Science or Information Technology fields along with existing professionals that are looking for ways to expand their skillset to become more marketable for other positions.
- All recommendations are based on an analysis and visualization of obtained data.
- This presentation seeks to not only answer what are the current in demand skills but what they will be in the future as well.

METHODOLOGY

- ▶ This project was conducted using a survey dataset, which prior to cleaning contained 11552 observations. This dataset asked respondents a variety of questions from whether they used open-source software to their country of origin.
- ▶ This dataset was cleaned in order to remove duplicates and to replace any missing values. In addition to compiling existing information into new variables for easier analysis.
- ▶ The data underwent an exploratory analysis to determine correlations, descriptive statistics, distribution, and to identify outliers.
- ▶ It was visualized in various ways from standalone plots to integrated enhanced dashboards using IBM Cognos Analytics.

RESULTS

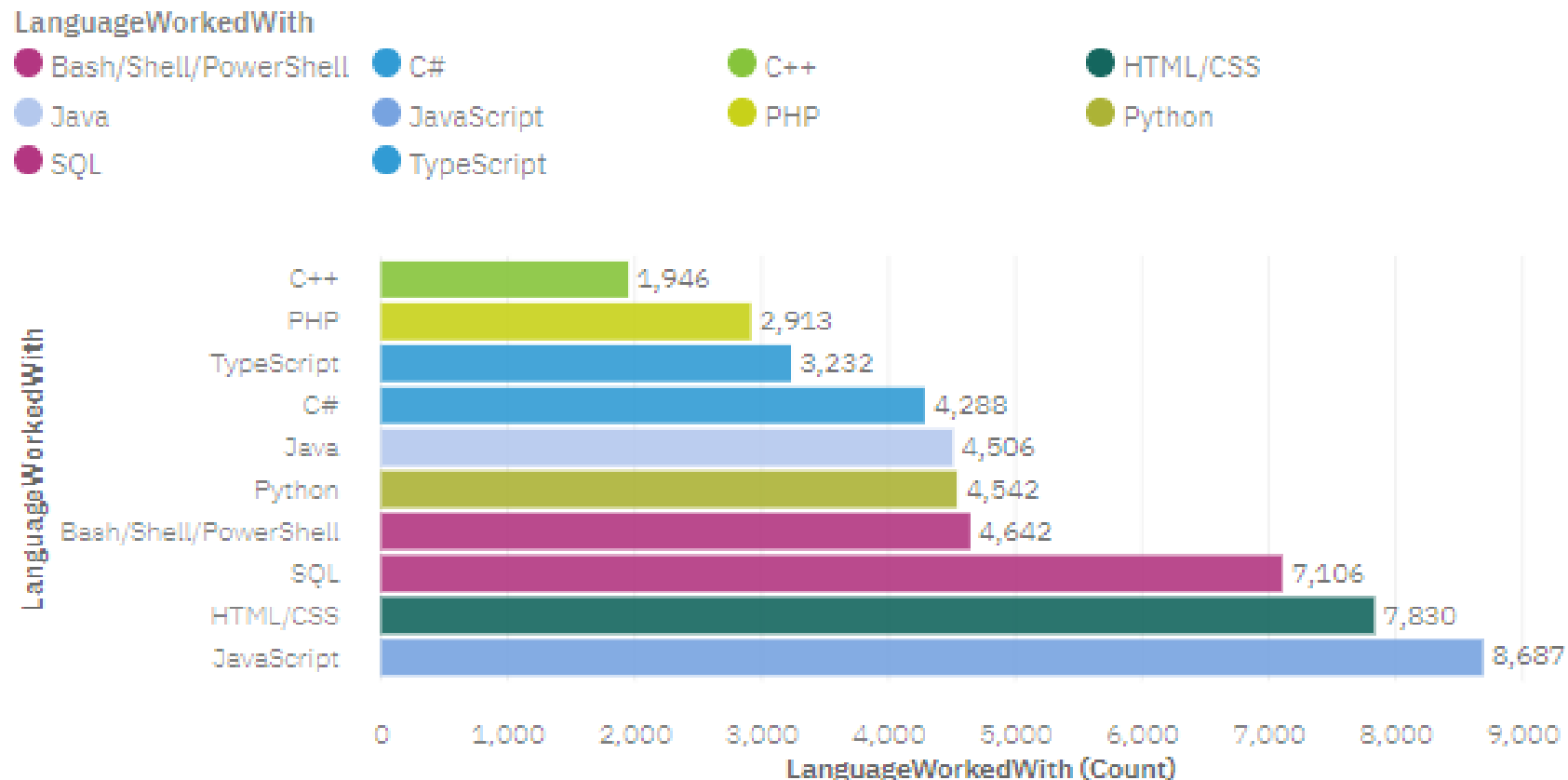
Distribution of Respondent Age



PROGRAMMING LANGUAGE TRENDS:

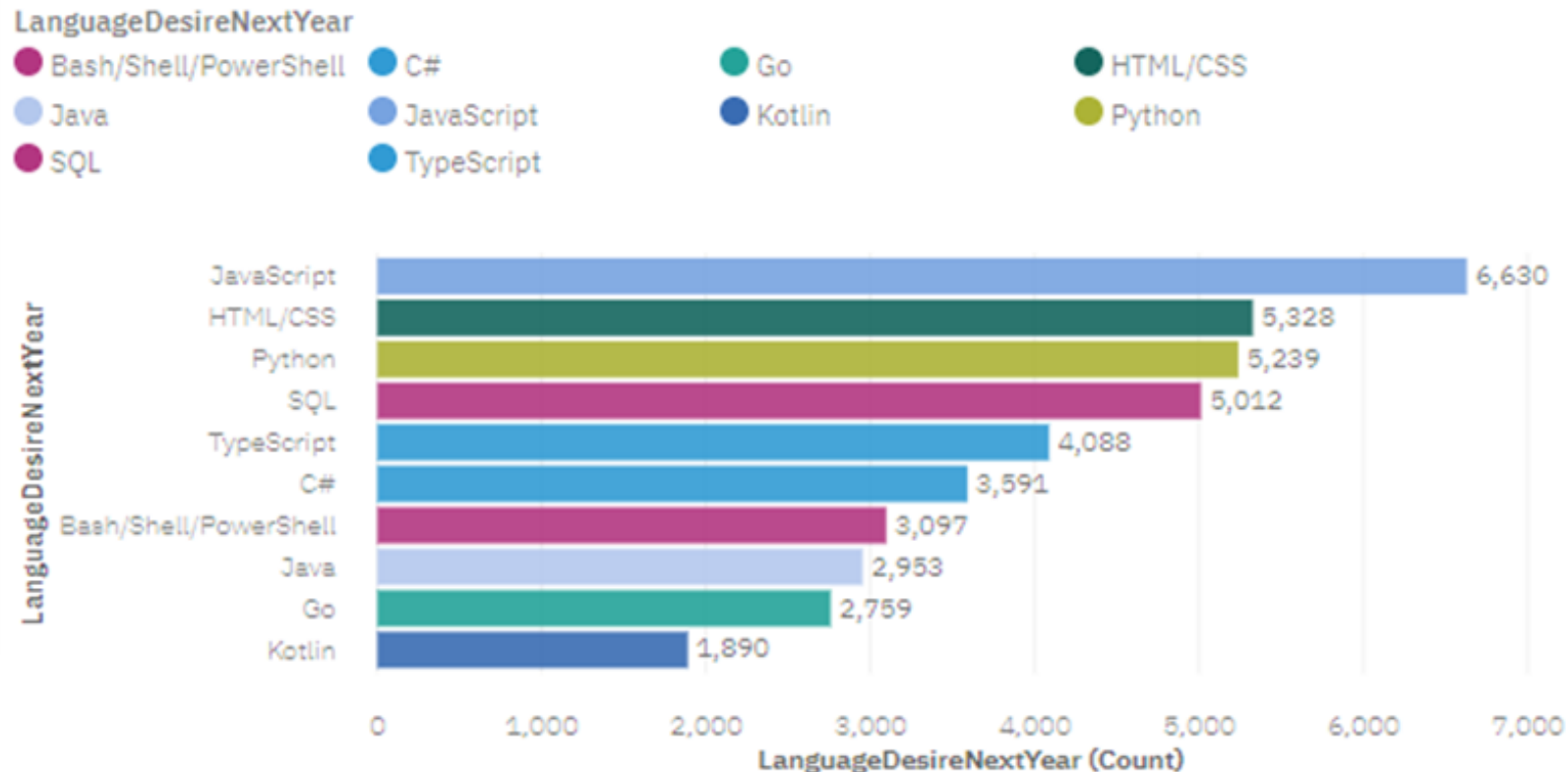
Current

Bar Chart of Top 10 Languages Worked With



PROGRAMMING LANGUAGE TRENDS: Next Year

Bar Chart of Top 10 Languages Desired Next Year



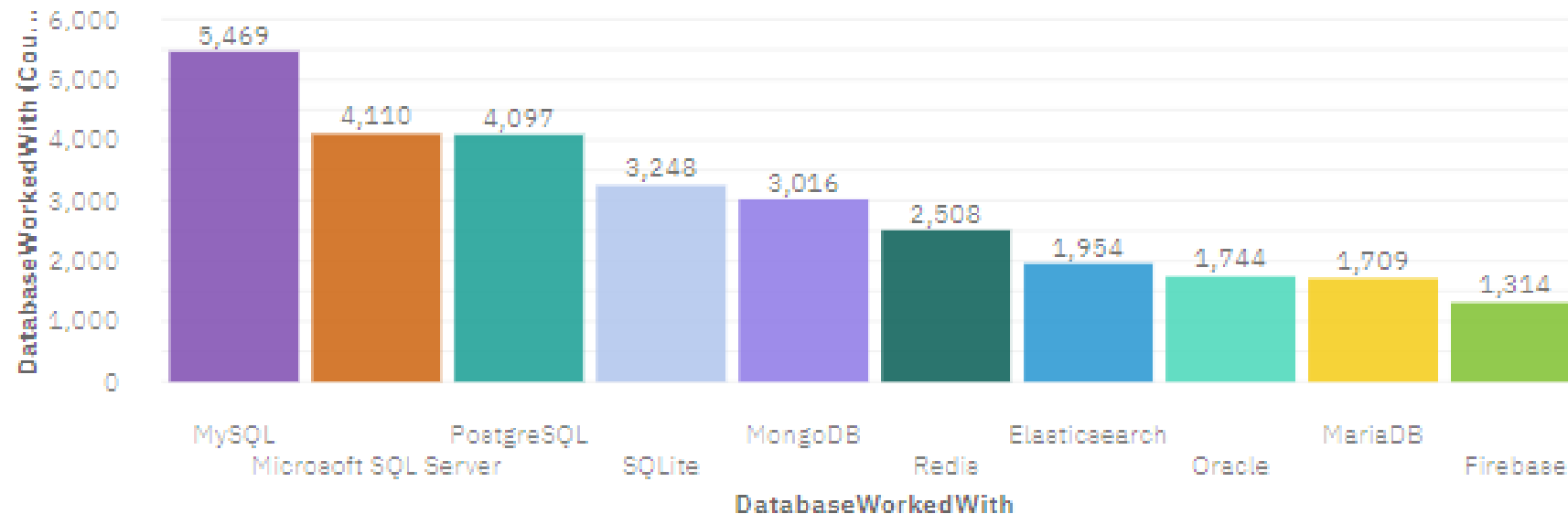
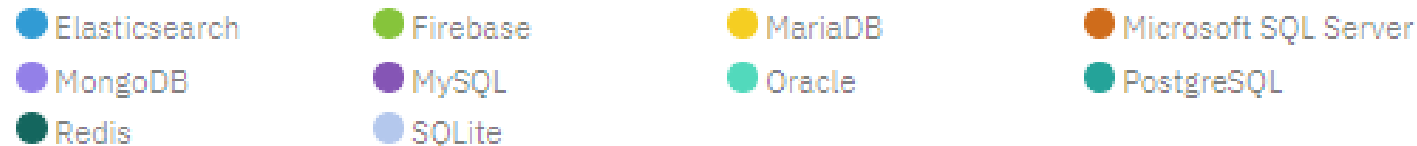
PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

- ▶ The data showed there is a strong overlap between what is desired now and what is desired next year.
- ▶ Javascript is the number one language desired for both periods with HTML being the runner.
- ▶ Meanwhile there has been change in interest in some languages such as Python, PowerShell, and SQL.
- ▶ Interest in Python grew, while PowerShell and SQL saw a decrease in interest.
- ▶ These trends imply that there is a demand for specific skills that do not fade with time.
- ▶ However, some of these skills will see slight increases and decreases overtime.
- ▶ The primary languages of interest seem to be JavaScript and HTML. Individuals with these skills are more likely to find employment, while Python looks to be more marketable in the future in comparison to SQL and PowerShell.

DATABASE TRENDS: Current

Column Chart of Top 10 Databases Worked With

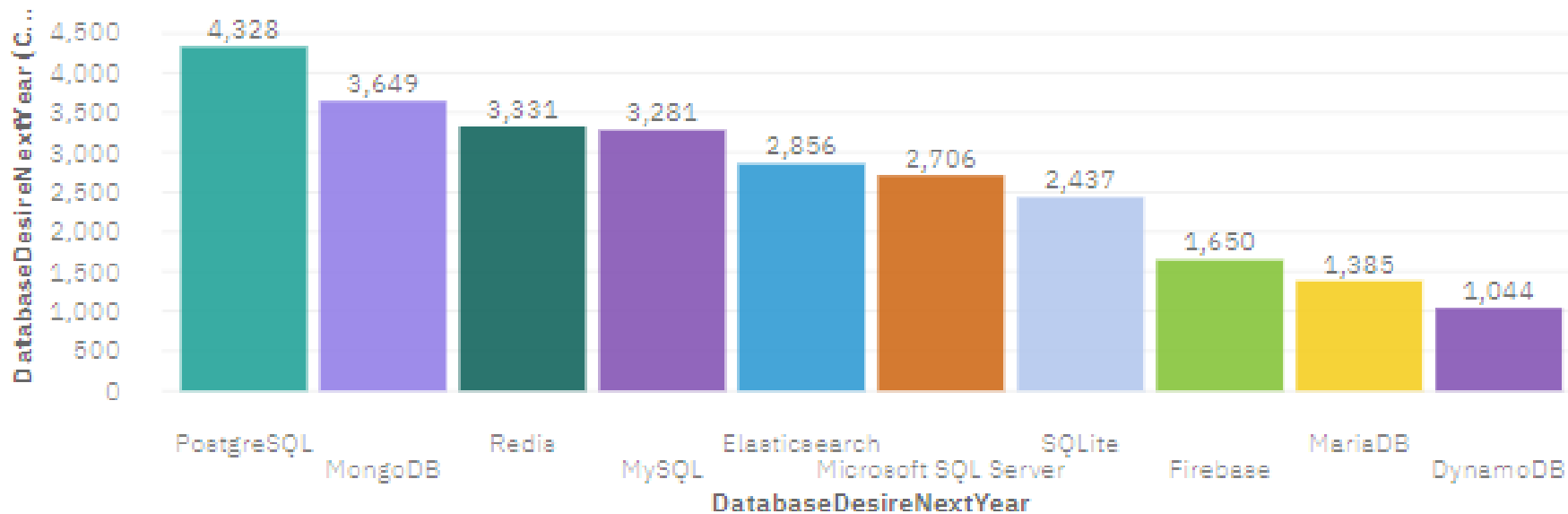
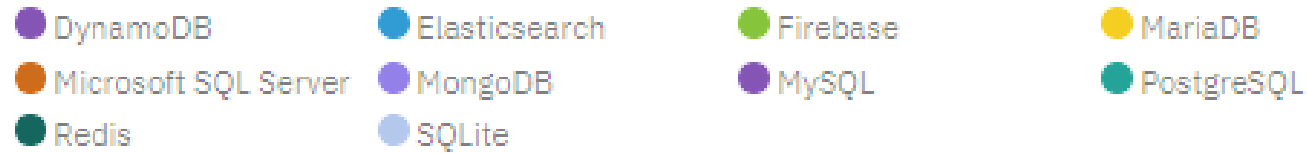
DatabaseWorkedWith



DATABASE TRENDS: Next Year

Column Chart of Top 10 Databases Desired Next Year

DatabaseDesireNextYear



DATABASE TRENDS - FINDINGS & IMPLICATIONS

- ▶ There are a lot of changes year to year when it comes to the preferred database.
- ▶ Currently MySQL and Microsoft SQL Server reign supreme, but next year the interest in these will fall to PostGres SQL and MongoDB.
- ▶ MongoDB appears to be growing in popularity as a more in-demand skill.
- ▶ Meanwhile, SQLite decreased significantly from being the 4th most desired to the 7th most desired.
- ▶ Individuals should evaluate whether their current knowledge of the preferred database is accurate.
- ▶ Moving forward, MongoDB appears to be playing a larger role in the Data Science and IT fields along with PostGres SQL.
- ▶ MySQL still has a role in the market but Microsoft SQL Server and SQLite are falling behind in desirability.



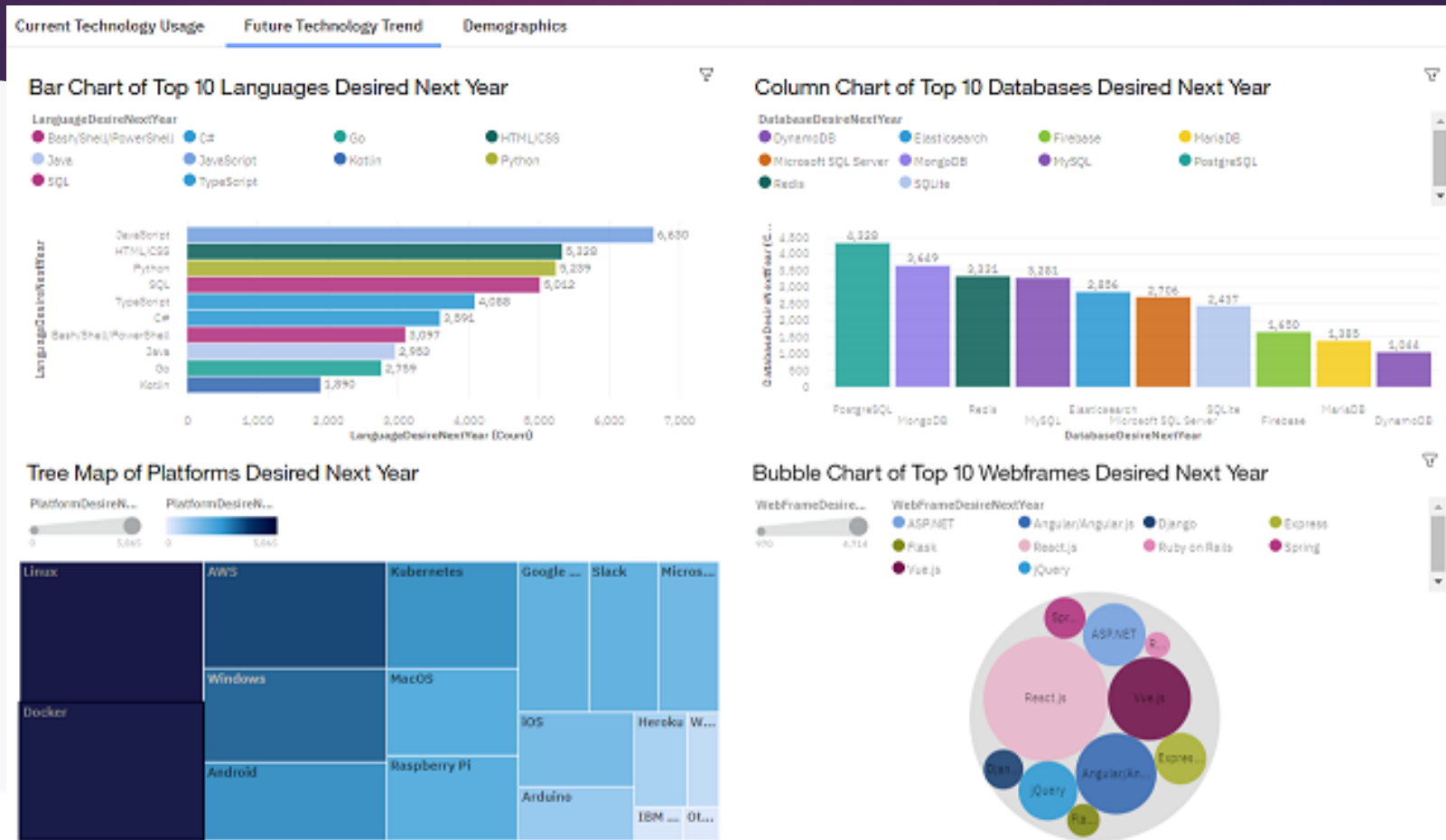
DASHBOARD

FOR FURTHER ANALYSIS CONSULT THE LINKED COGNOS ANALYICS
DASHBOARD: [LOCATED HERE](#)

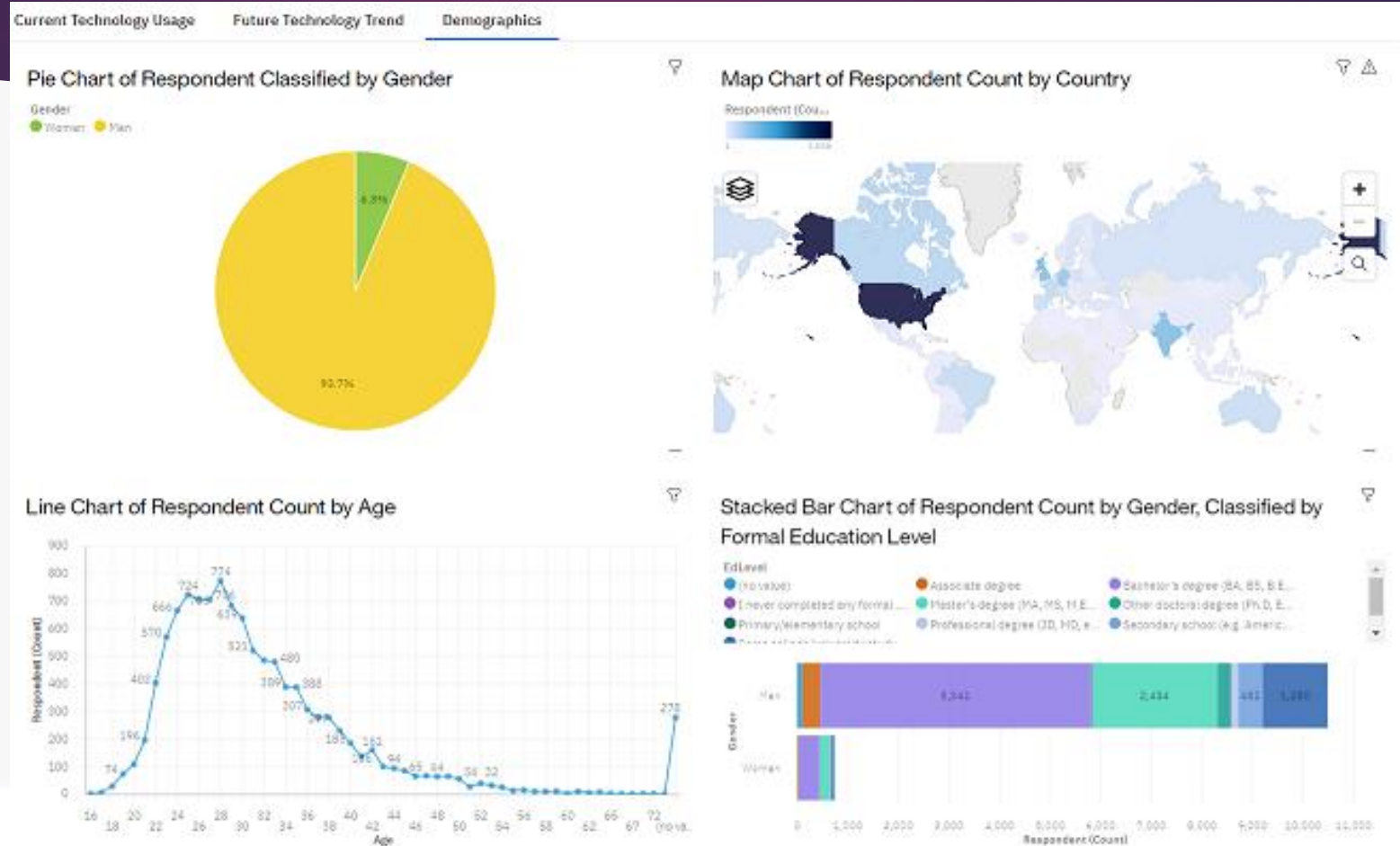
DASHBOARD TAB 1



DASHBOARD TAB 2



DASHBOARD TAB 3



DISCUSSION



OVERALL FINDINGS & IMPLICATIONS

- ▶ The top programming languages are JavaScript and HTML with Python, SQL, PowerShell, and TypeScript being other popular ones now or in the future.
- ▶ The desired databases are changing fairly quickly with MongoDB soon to become a top database while MySQL and Microsoft SQL Server are falling out of favor.
- ▶ Regarding Platforms, Linux is still a popular platform along with Windows, while the latter is falling out of favor for others such as Docker.
- ▶ jQuery is the current top Webframe but soon to be overtaken by React.js

A strong foundation in JavaScript, HTML, Python, and SQL will aid an individual currently but also prepare them moving forward. Learning TypeScript will prepare an individual for the future.

Understanding PostGres SQL, MongoDB, and MySQL are needed skills for tomorrow.

Knowledge on how to utilize Linux, Docker, and Windows are requirements for today and tomorrow.

jQuery is currently very popular, but React.js is growing.

CONCLUSION

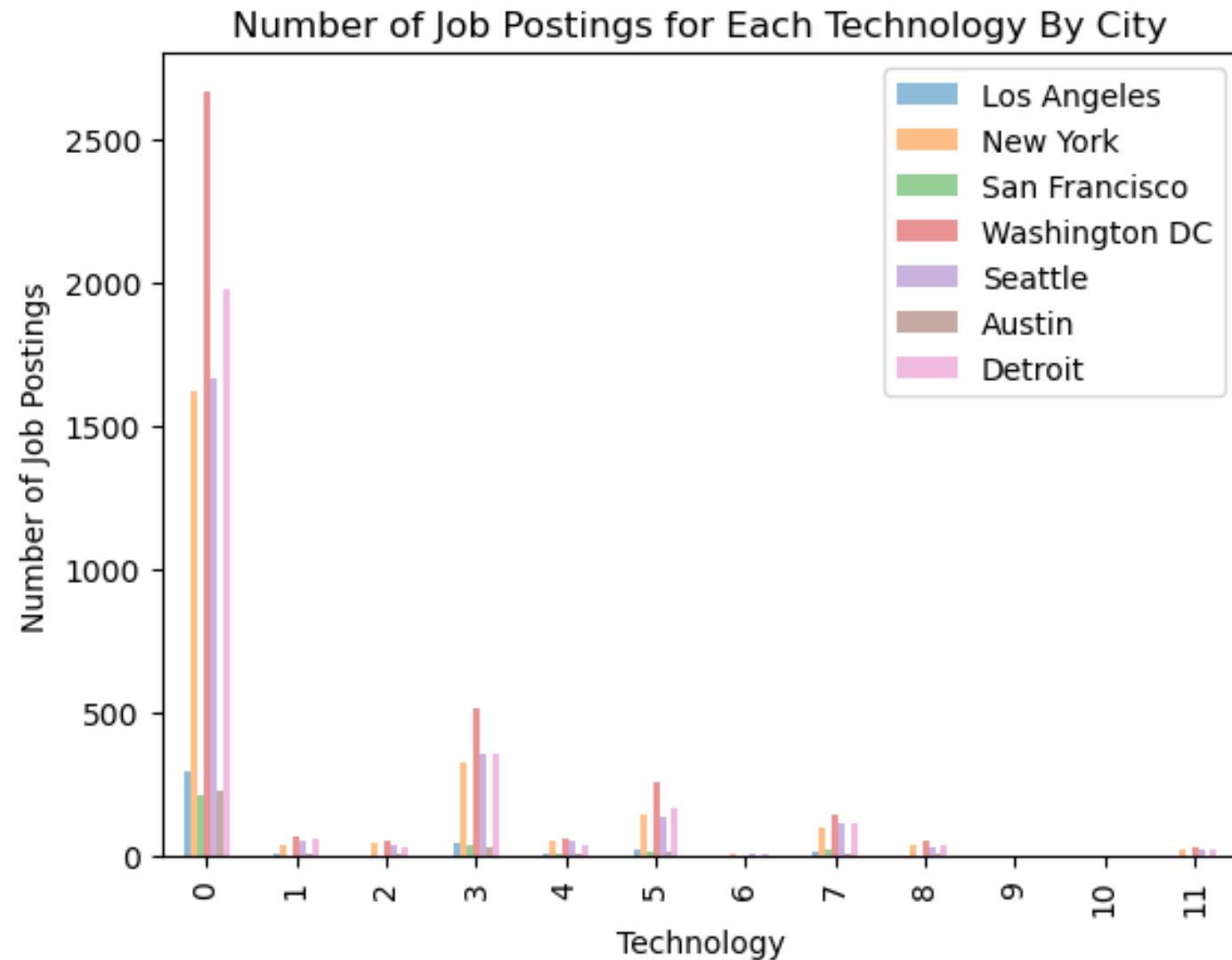


- ▶ Data Science and Information Technology are two very interesting but changing fields. The best outlook when entering these fields is to always be aware of changes and to find ways to always increase an individual's knowledge and skill level.
- ▶ Focus on gaining a strong foundation in current popular languages, databases, platforms, and webframes.
- ▶ Afterwards, continue to grow a skillset by tackling next year's most desired languages, databases, platforms, and webframes.

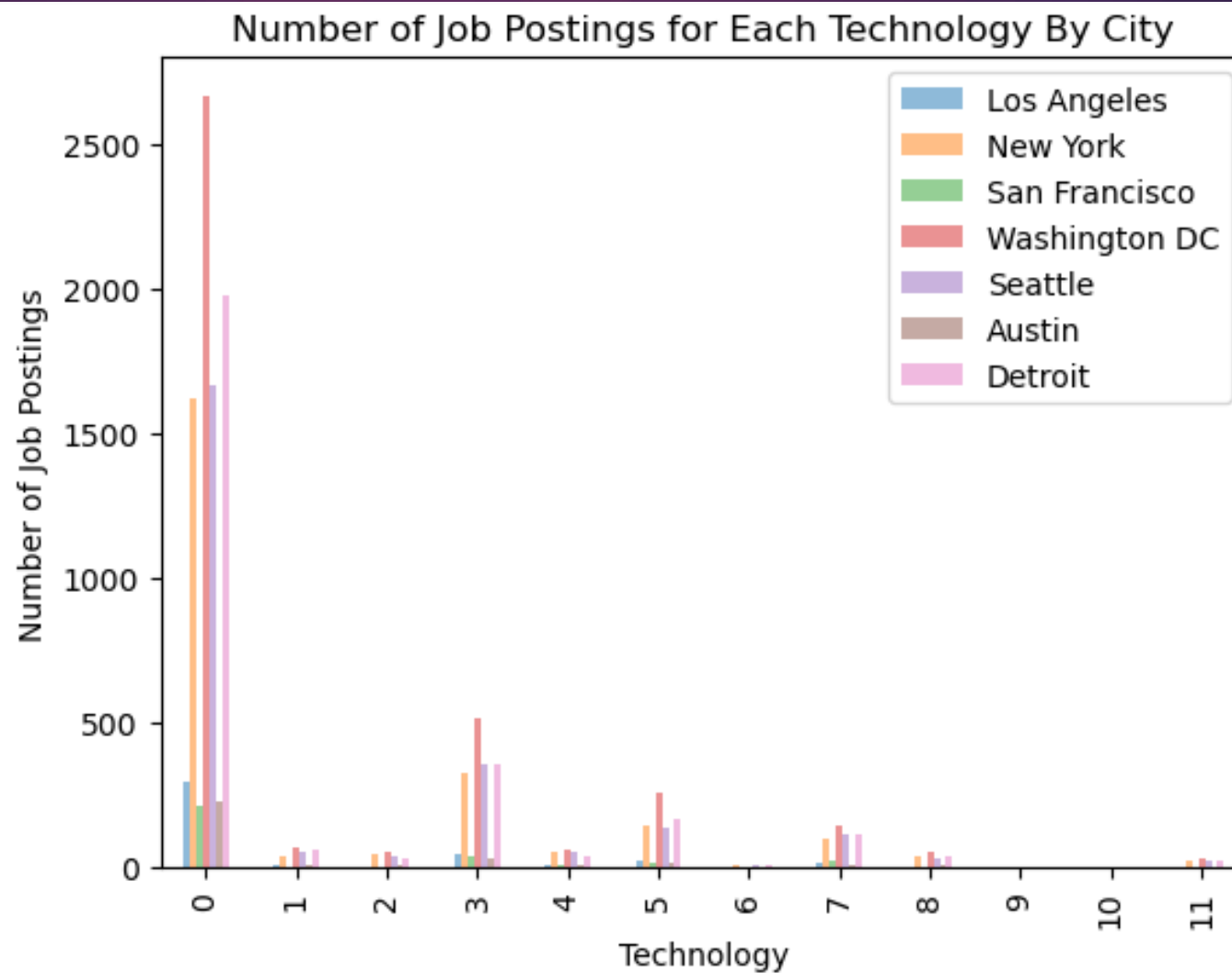
APPENDIX



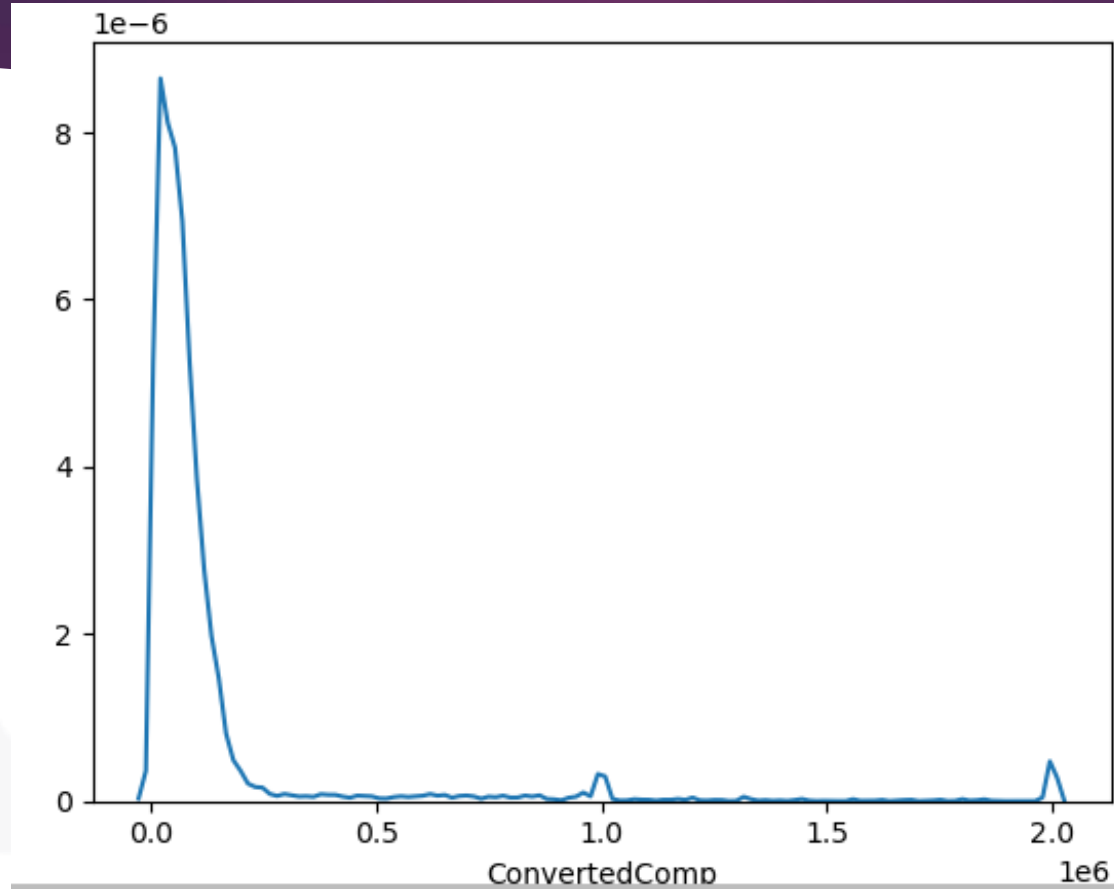
JOB POSTINGS



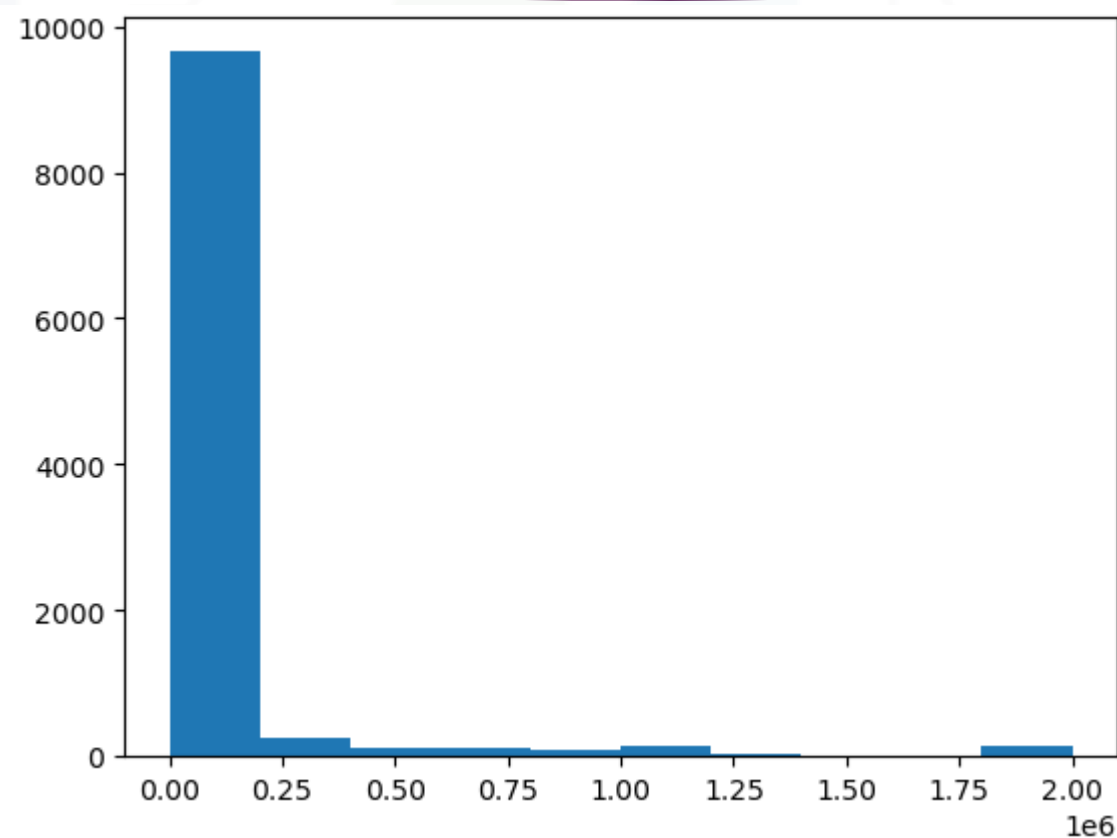
POPULAR LANGUAGES



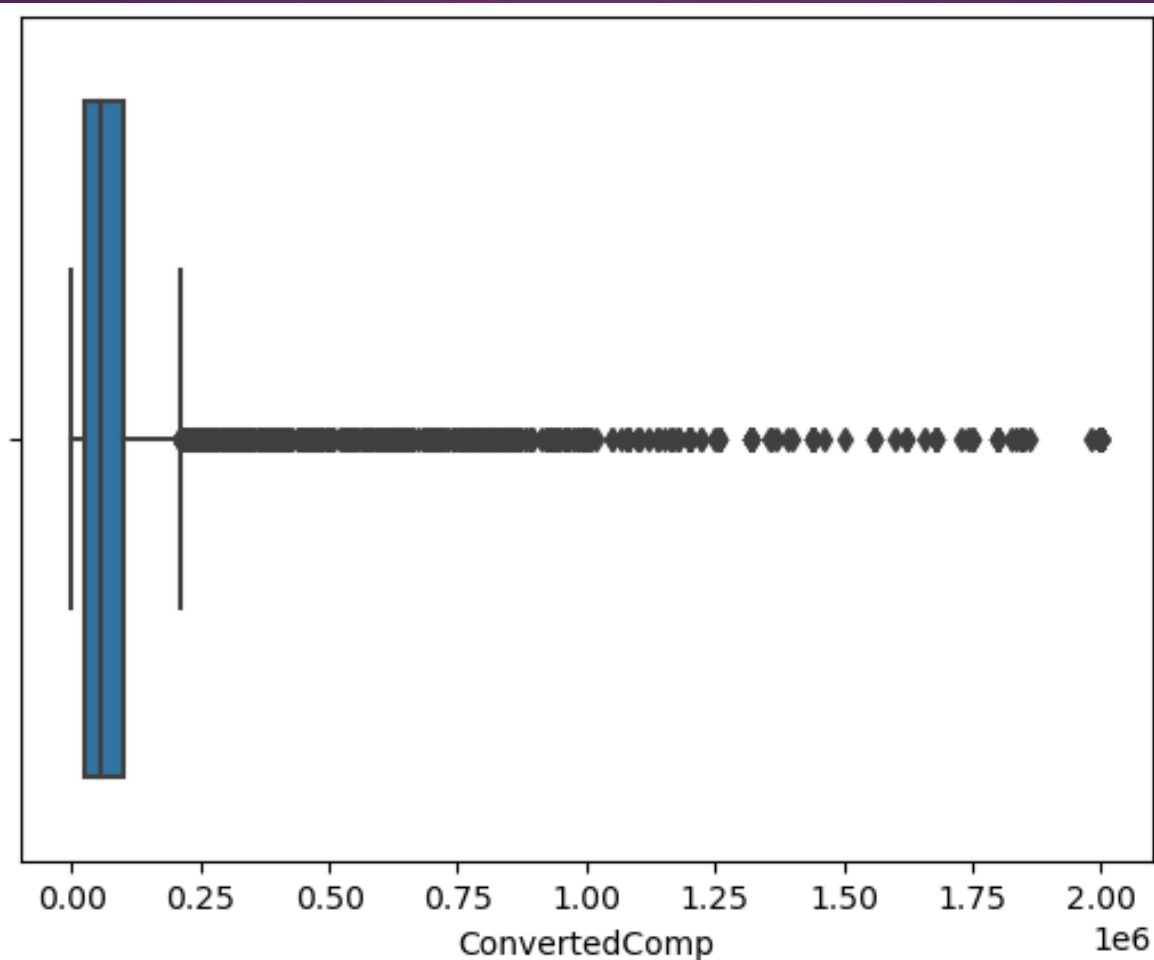
Line Graph of Converted Comp



Histogram of Converted Comp



Boxplot of Converted Comp to check for Outliers



Correlation Matrix for Age and other variables

	Respondent	CompTotal	ConvertedComp	WorkWeekHrs	CodeRevHrs	Age
Respondent	1.000000	-0.013490	0.002181	-0.015314	0.004621	0.004041
CompTotal	-0.013490	1.000000	0.001037	0.003510	0.007063	0.006970
ConvertedComp	0.002181	0.001037	1.000000	0.021143	-0.033865	0.105386
WorkWeekHrs	-0.015314	0.003510	0.021143	1.000000	0.026517	0.036518
CodeRevHrs	0.004621	0.007063	-0.033865	0.026517	1.000000	-0.020469
Age	0.004041	0.006970	0.105386	0.036518	-0.020469	1.000000

Relationship of Converted Comp and Age

