IBM Data Analyst Capstone Project: Adapting to the Emerging Technologies and Skills in the Global IT Sector

Abiyselassie M. Gashaw

October 22,2024

IBM Developer

SKILLS NETWORK

OUTLINE

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

EXECUTIVE SUMMARY

Summary of methodologies

- The data was collected from several datasets (e.g. Stack Overflow Survey, GitHub job postings) & web scraped.
- The data was then wrangled to remove duplicates & impute missing values from the survey dataset, and the cleaned data was normalized.
- The wrangled data was subjected to **EDA** (Exploratory Data Analysis) to determine the distribution of the normalized data, and to detect and remove outliers from the survey dataset.
- During Data Visualization, various charts were plotted to visualize the distribution, interesting relationships between features, composition and comparison of the data.
- **Cognos** was employed to create interactive dashboards to help analyze & present the data dynamically.

Summary of all results

- The findings from the dashboard showed that **JavaScript**, **HTML/CSS**, **SQL** & **Python** top the list of most used popular programming languages in demand by respondents for the current year and next year.
- **MySQL** & **Microsoft SQL Server** top the list of the most frequently used databases in demand by respondents for the current year, but **PostgreSQL** and **MongoDB** are expected to have more demand in the future.
- **jQuery** was identified as the most broadly used web frame in demand in the current year, but **React.js.** was expected to be the most desired web frame to learn next year by respondents.
- **Docker** was identified as the most desirable platform along with **Linux** & **Windows** platforms.
- Most of the respondents were males, had bachelor degree, located in the USA & were 28 yrs during the survey.

INTRODUCTION

- **Technical skills** required in the global IT sector & business consulting firms are constantly changing. Thus, Global IT and business consulting services firm should regularly analyze data to be competitive & keep pace with the ever-changing technologies.
- As part of a capstone project, I took on the role of a **Data Analyst** with this company to assist with collection, web scraping, exploration, wrangling, analysis and visualization of several datasets obtained from various sources to help identify trends for the current & next year's report on emerging technologies and skills.
- The cleaned dataset will be analyzed to identify insights & trends that answer following questions:
  - What are the **top programming languages** that are in demand?
  - What are the **top database** skills that are in demand?
  - What are the most popular IDEs (**platforms**, **web frames**)?
  - What do the **Demographic data** (gender, age, country, education level) look like ?
- **Cognos** will be employed to create interactive dashboards to help analyze & present the data dynamically.
- **Storytelling** skills will be used to provide a narrative & present the findings of the analysis to job recruiters, job seekers, tech innovators, employers, executives & professionals in the company.

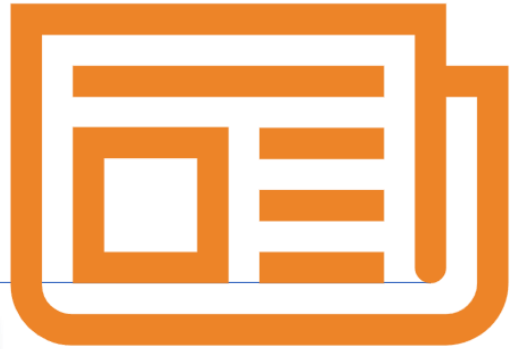
METHODOLOGY

- **Data Collection: API**
  - Job data were collected from **Jobs API** (job postings) for the technology skills that are in demand.
  - The collected data were stored into an Excel Spreadsheet named **'job-postings.xlsx**'

- **Data Collection: Webscraping**
  - **Language name** & **Average Annual salary** were scraped from a Programming_Languages.html website
  - The scraped data were written into a **popular-languages.csv** file.

- **Survey Dataset Exploration**
  - A **survey data** (a modified subset of **Stack Overflow** dataset) available at a dataset_url were loaded into a dataframe.
  - The dataset was explored for the technology skills that are in demand, and the data types of each column were identified.

METHODOLOGY…



- **Data Wrangling**
  - **Duplicate values** were identified & removed from the survey dataset.
  - **Missing values** were identified & imputed in the survey dataset.
  - The cleaned data were normalized in the survey dataset.

- **Exploratory Data Analysis**
  - Statistical techniques were applied to analyze the data and identify **patterns** & **trends**.
  - The distribution of the normalized data in the survey dataset was determined.
  - **Outliers** were detected and removed from the survey dataset.
  - Interesting **correlation** between different features was determined in the survey dataset.

- **Data Visualization**
  - Various **charts** were plotted to visualize the distribution, interesting relationships between features, composition and comparison of the data.
  - **Horizontal bar chart** was plotted to visualize comparison of Professional & non-professional Developers (**Appendix**).

METHODOLOGY…

- **Building an Interactive Dashboard with IBM Cognos**

2 CSV files, **m5_survey_data_demographics.csv** & **m5_survey_data_technologies_normalised.csv** were downloaded & then uploaded to the **IBM Cognos Dashboard**.

- A **Dashboard** renamed as **Current Technology Usage** was created where
  - Metrics such as Top 10 LanguageWorkedWith, Top 10 DatabaseWorkedWith, PlatformWorkedWith & Top 10 WebFrameWorkedWith were captured, and visualized as bar chart, column chart, word cloud chart and hierarchical bubble chart, respectively.

- A **Dashboard** renamed as **Future Technology Trend** was created where
  - Metrics such as Top 10 LanguageDesireNextYear, Top 10 DatabaseDesireNextYear, PlatformDesireNextYear & Top 10 WebFrameDesireNextYear were captured, and visualized as bar chart, column chart, tree map chart and hierarchical bubble chart, respectively.

- A **Dashboard** renamed as **Demographics** was created where
  - Metrics such as ‘Respondent classified by Gender’, ‘Respondent Count for Countries’, ‘Respondent Count by Age’, and ‘Respondent Count by Gender classified by Formal Education Level’ were captured, and visualized as pie chart, map chart, line chart and stacked bar chart, respectively.

RESULTS

Programming Language Trends

- Findings & Implications

Database Trends

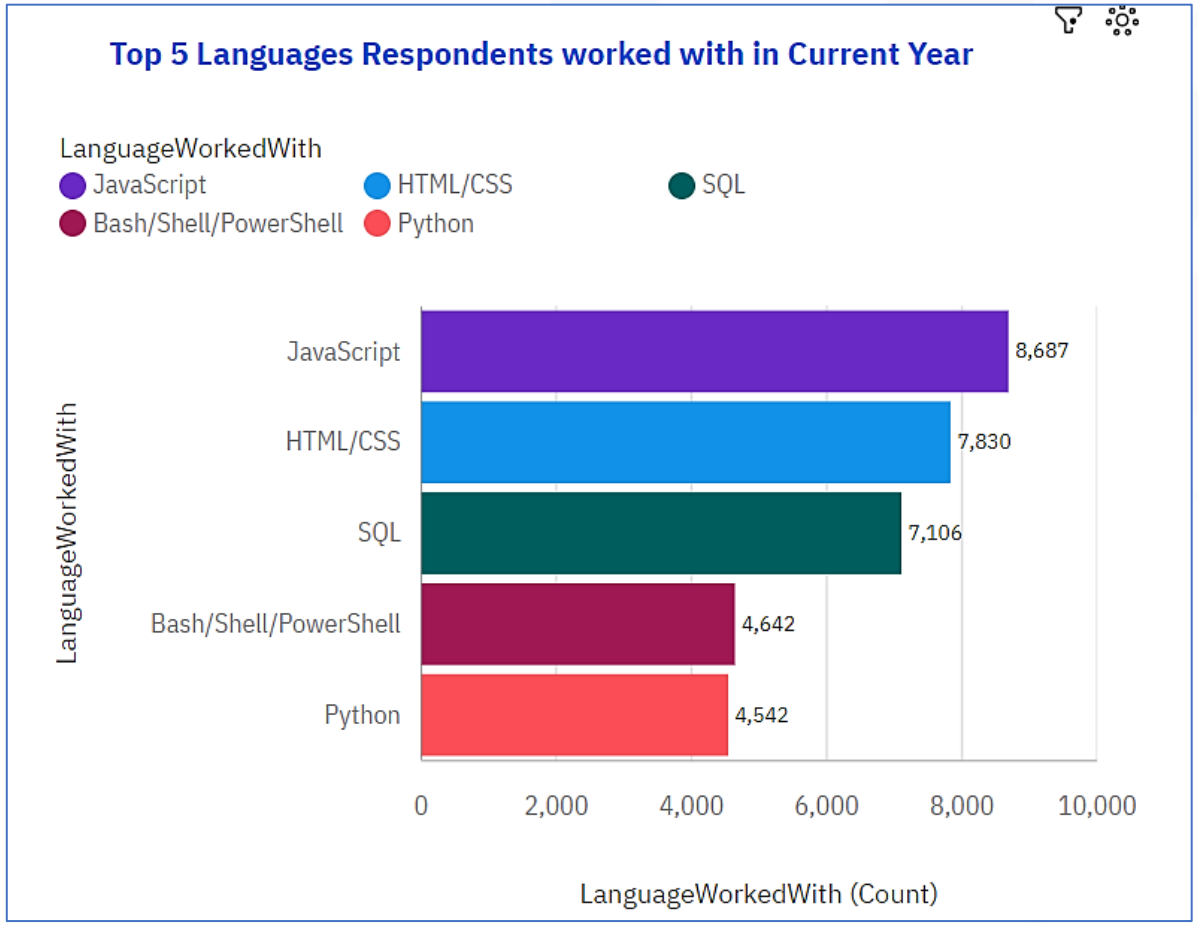
- Findings & Implications

Dashboard

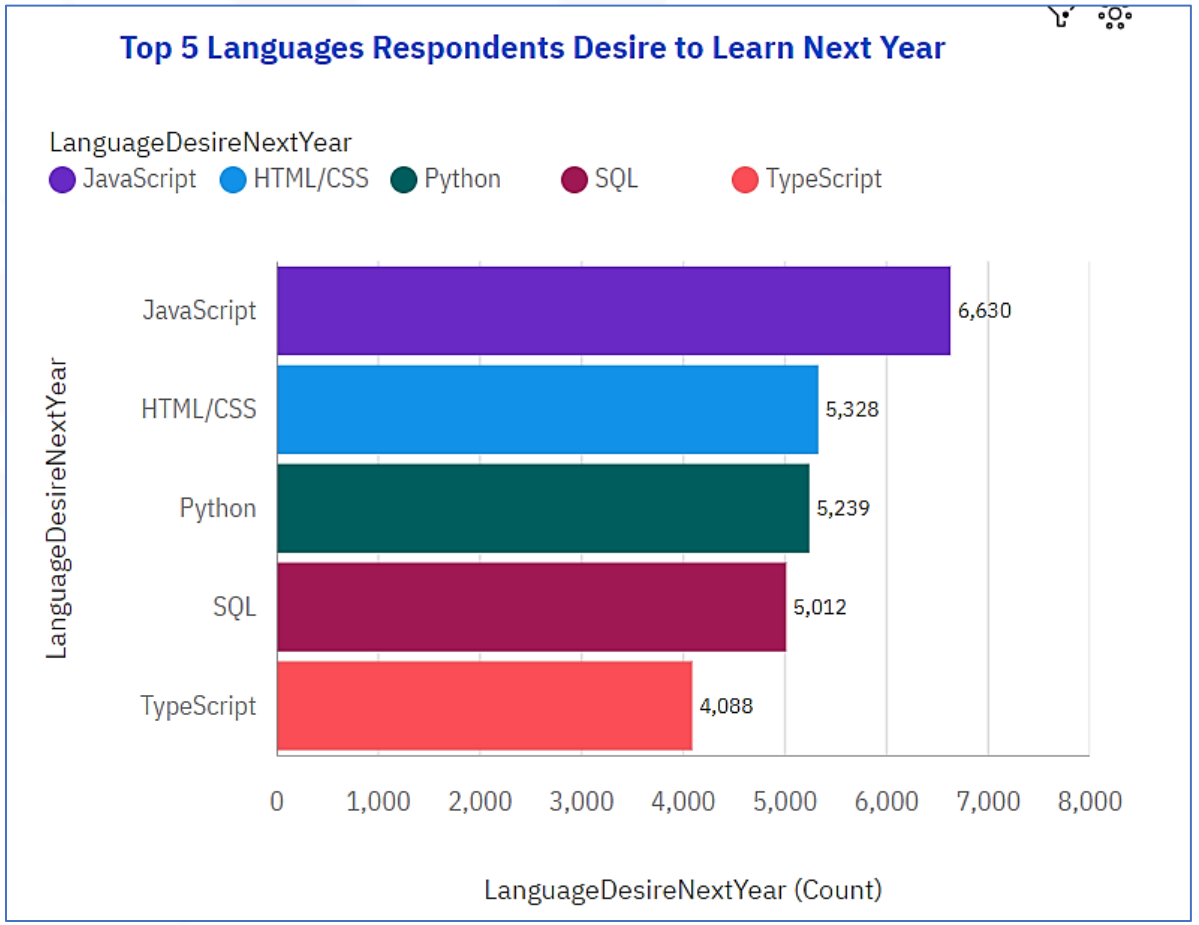
- Dashboard Tab 1
- Dashboard Tab 2
- Dashboard Tab 3

PROGRAMMING LANGUAGE TRENDS

Current Year



Next Year



IBM Developer

SKILLS NETWORK

PROGRAMMING LANGUAGE TRENDS-FINDINGS & IMPLICATIONS

Findings

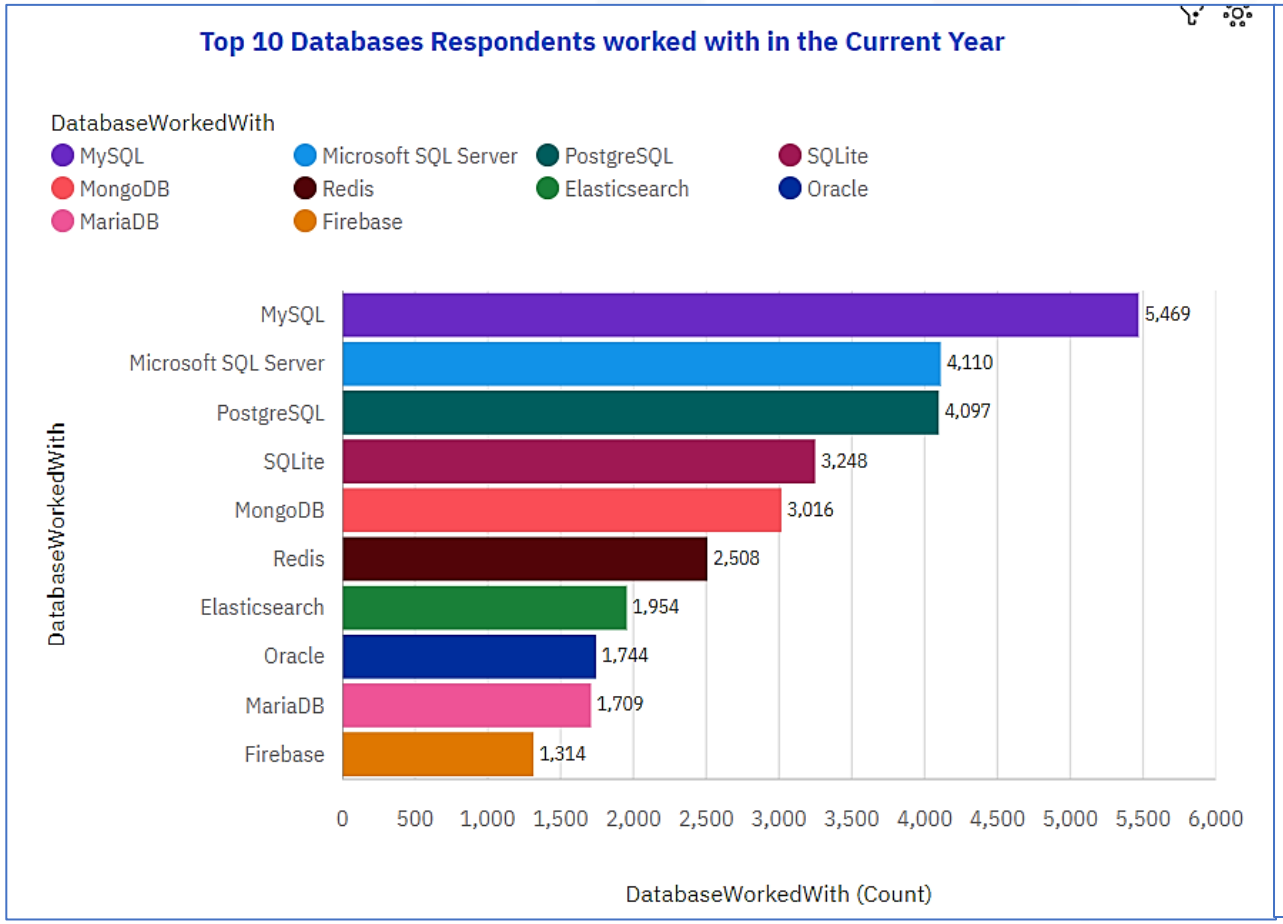
- The plotted bar charts for the **Top 5** Programming language show that **JavaScript** & **HTML/CSS** remain to be the 1st and 2nd choices that respondents have currently worked with and desire to learn in the future.
- **Python** has risen in the ranks of programming language from 5th in the current year to 3rd next year.
- **TypeScript** has joined the **Top 5** category of programming languages that respondents desire to learn next year, edging out **Bash/Shell/PowerShell**.

Implications

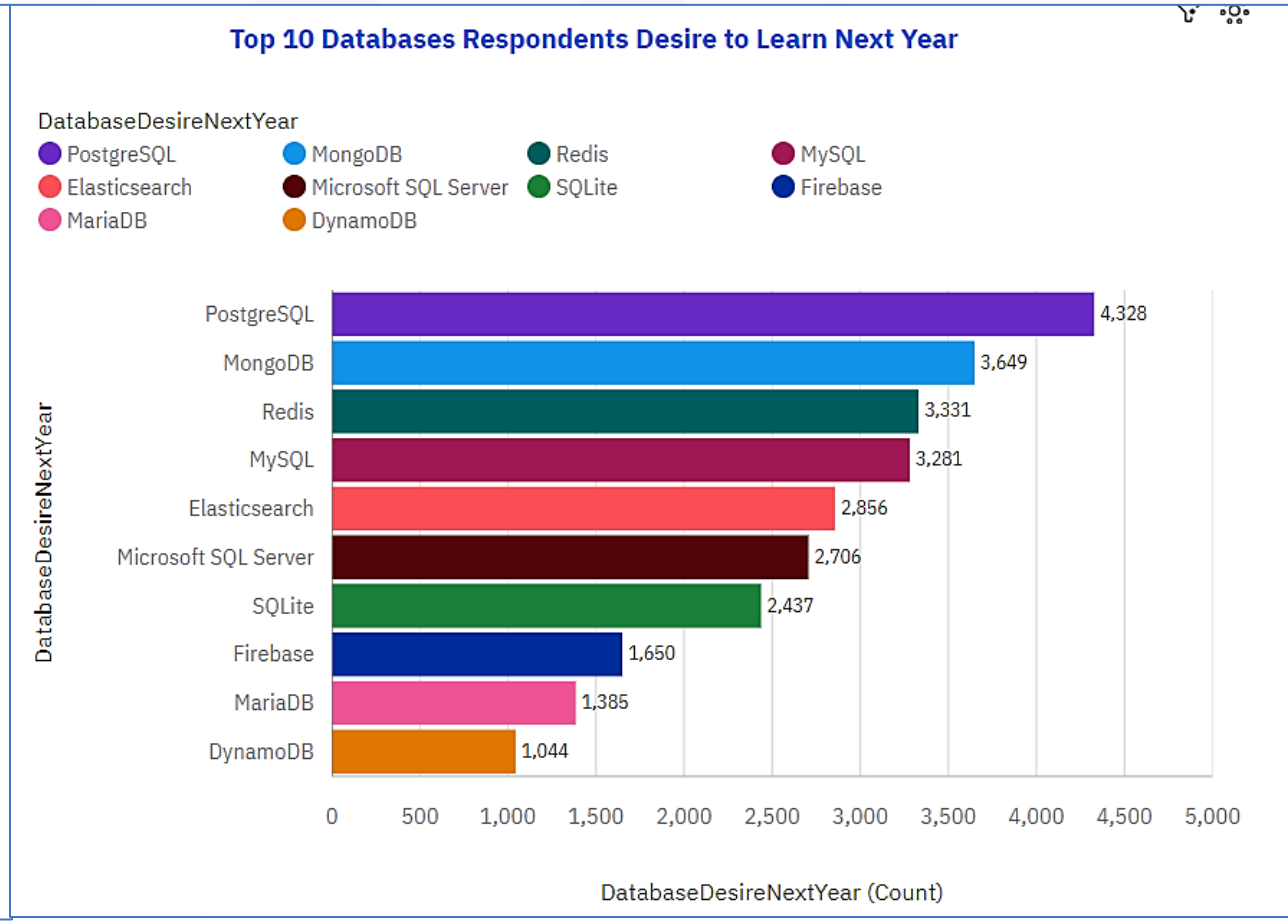
- **JavaScript's** & **HTML/CSS's** remaining top of all programming languages in both current and future situations implies that they are widely used & required by respondents & professional developers.
- **Python's** dramatic rise in demand in the top list of all programming languages in the next year implies that it has become the fastest growing major programming language & will become the 3rd most desired language in demand by respondents for next year.
- The impressive rise of **TypeScript** into the **top 5** category implies that there is an upward trend for using this language for industry & professional development.

DATABASE TRENDS

Current Year



Next Year



DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- The bar charts plotted for the **Top 10** Databases show that **MySQL** has become the most frequently used database by respondents currently, but **PostgreSQL** will become the 1st choice that respondents desire to learn in the future, edging ahead of **Microsoft SQL Server**.
- **MongoDB** has risen in the ranks of databases from 5th in the current year to 2nd in the next year.
- **Redis** & **Elasticsearch** have joined the **Top 5** category of databases that respondents desire to learn next year, kicking out **Microsoft SQL Server** & **SQLite**.

Implications

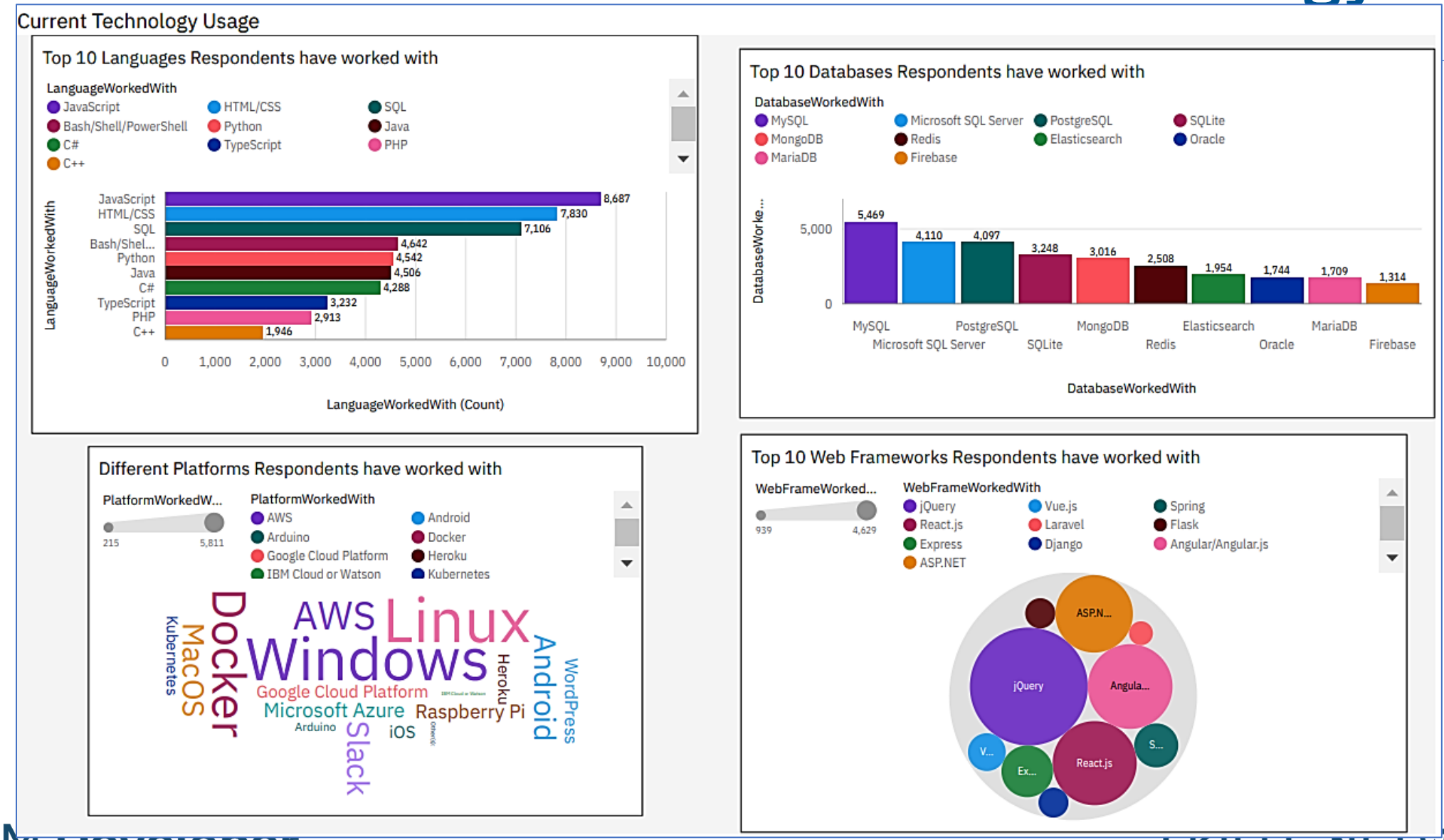
- **MySQL's** reign over all other databases in the current year and **PostgreSQL's** dramatic rise to the top of all databases in the future implies that both are widely used and required by respondents & professional developers.
- **MongoDB's** exciting rise in demand in the next year implies that it will be the 2nd most desired database in demand by respondents.
- The remarkable rise of **Redis & Elasticsearch** into the **top 5** category implies that there is an upward trend for using these databases for industry & professional development in the coming years.

DASHBOARD

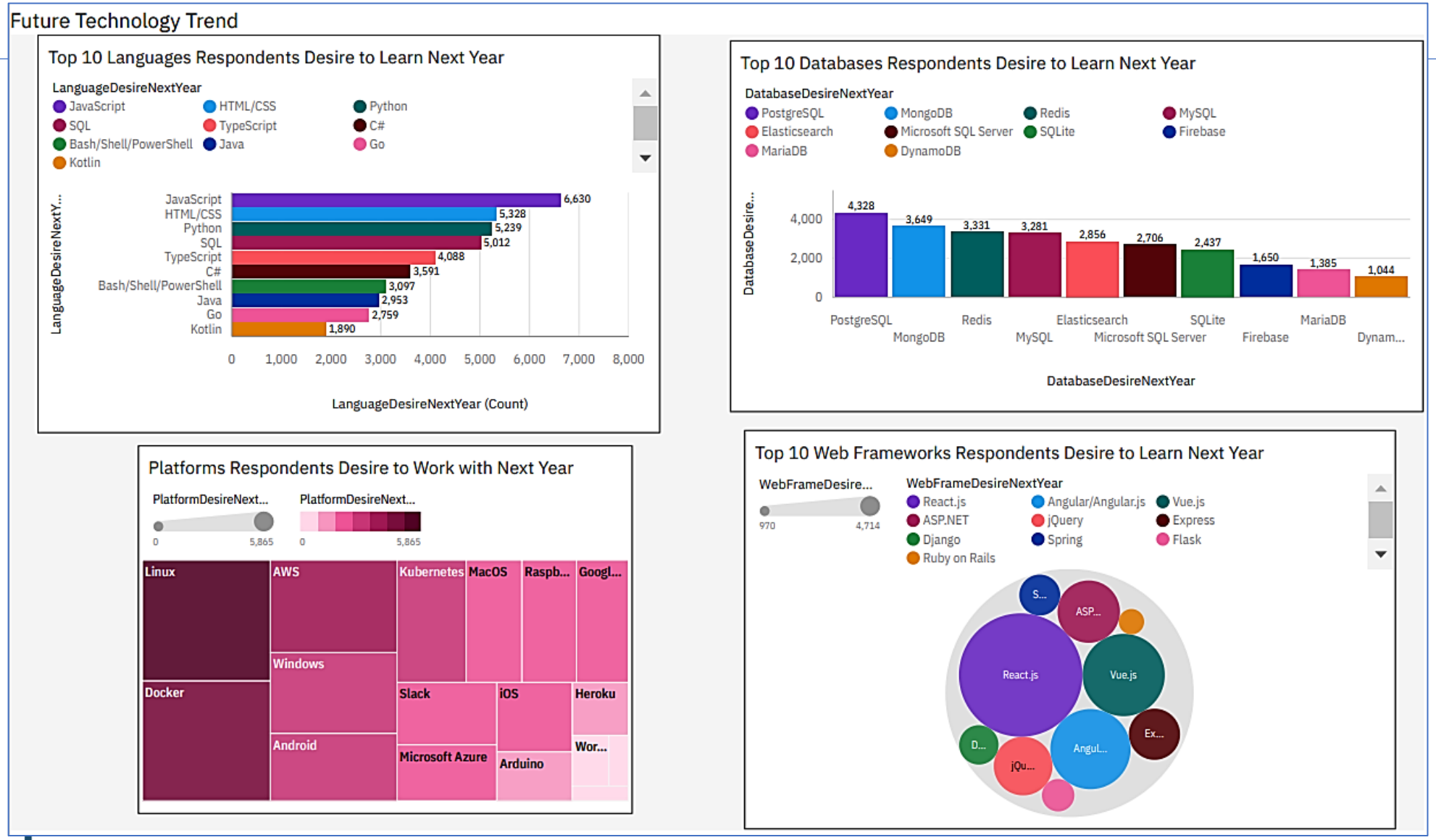
The GitHub link of the Cognos,

https://github.com/abiyselassie22/IBM-Data-Analyst-Capstone/blob/master/7.%20Building%20Dashboard%20IBM%20Cognos%20Analytics.pdf

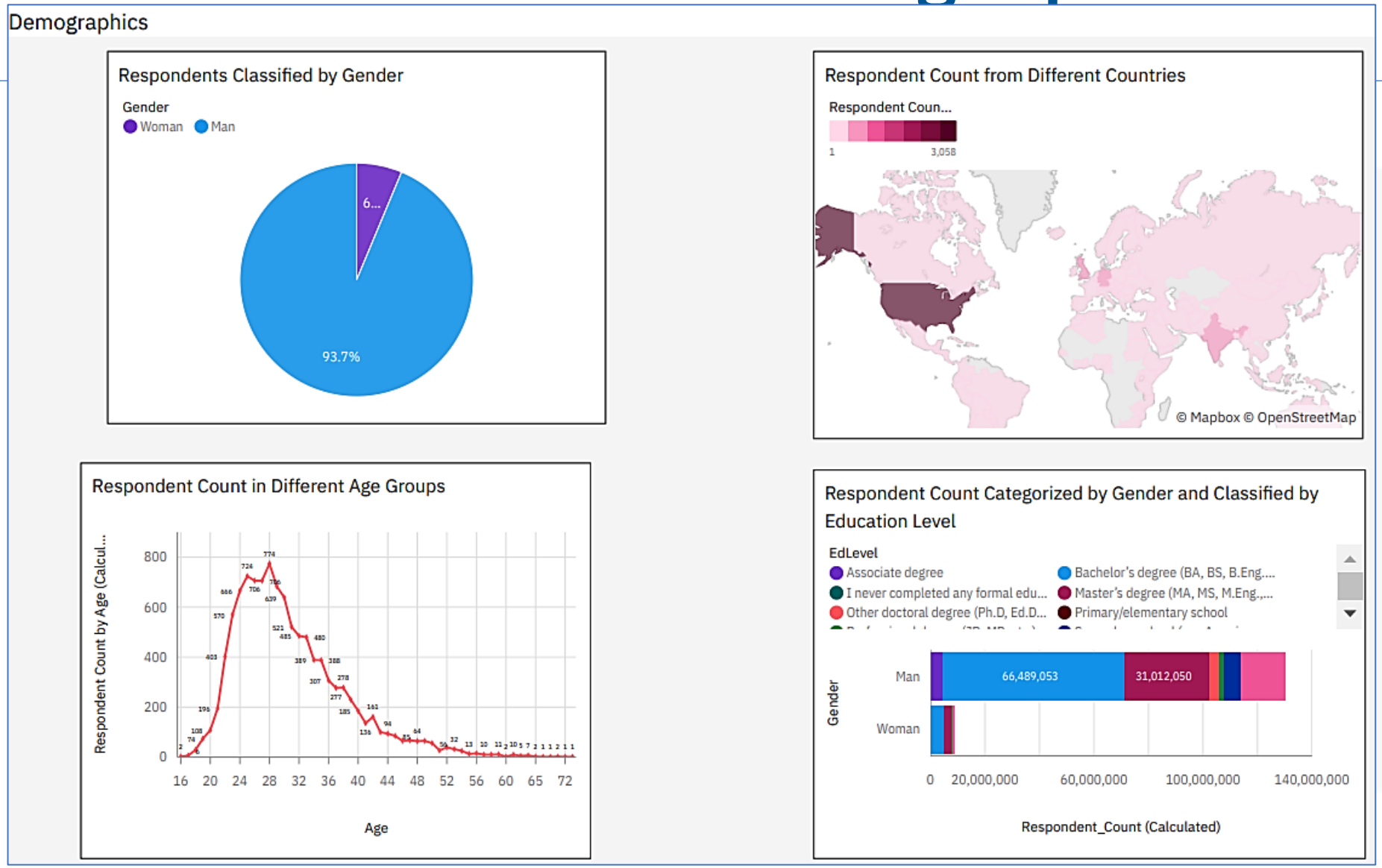
DASHBOARD TAB 1 – Current Technology Usage



DASHBOARD TAB 2 – Future Technology Usage



DASHBOARD TAB 3 - Demographics



DISCUSSION

OVERALL FINDINGS & IMPLICATIONS

- The overall findings and implications would help the Firm make more informed data-driven decisions.

OVERALL FINDINGS & IMPLICATIONS

Findings

- **JavaScript**, **HTML/CSS** & **SQL** top the list of most used popular programming languages in demand by respondents for the current year. And **Python** will be the most desired language to learn by respondents for the next year.

- **MySQL**, **Microsoft SQL Server** & **PostgreSQL** top the list of the most frequently used databases in demand by respondents currently. But **PostgreSQL** & **MongoDB** will be the most desired databases to learn by respondents in the future, edging ahead of **MySQL** & **Microsoft SQL Server**.

Implications

- It is imperative for job recruiters & technological innovators to adapt to the increasing trend of the most desired Python in addition to **JavaScript**, **HTML/CSS** & **SQL**.

- It is essential for job recruiters & tech innovators to give prominence to **MongoDB** in addition to **MySQL**, **Microsoft SQL** Server & **PostgreSQL** in the future.

OVERALL FINDINGS & IMPLICATIONS…

Findings

- **jQuery** was identified as the most broadly used web frame in the current year. But **React.js**. will be the most desired web frame to learn next year by respondents.
- **Linux**, **Windows** & **Docker** were the most common platforms by respondents in the current year. **Linux** & **Docker** will be the most used platforms that respondents desire to learn next year.
- **Most** of the respondents were males (**93.7%**), located in the USA & were at age of **28** during the survey (**Demographic Dashboard**).
- **Most** Respondents were developers by their profession (**Appendix**) & had Bachelor degrees (**Demographic Dashboard**).

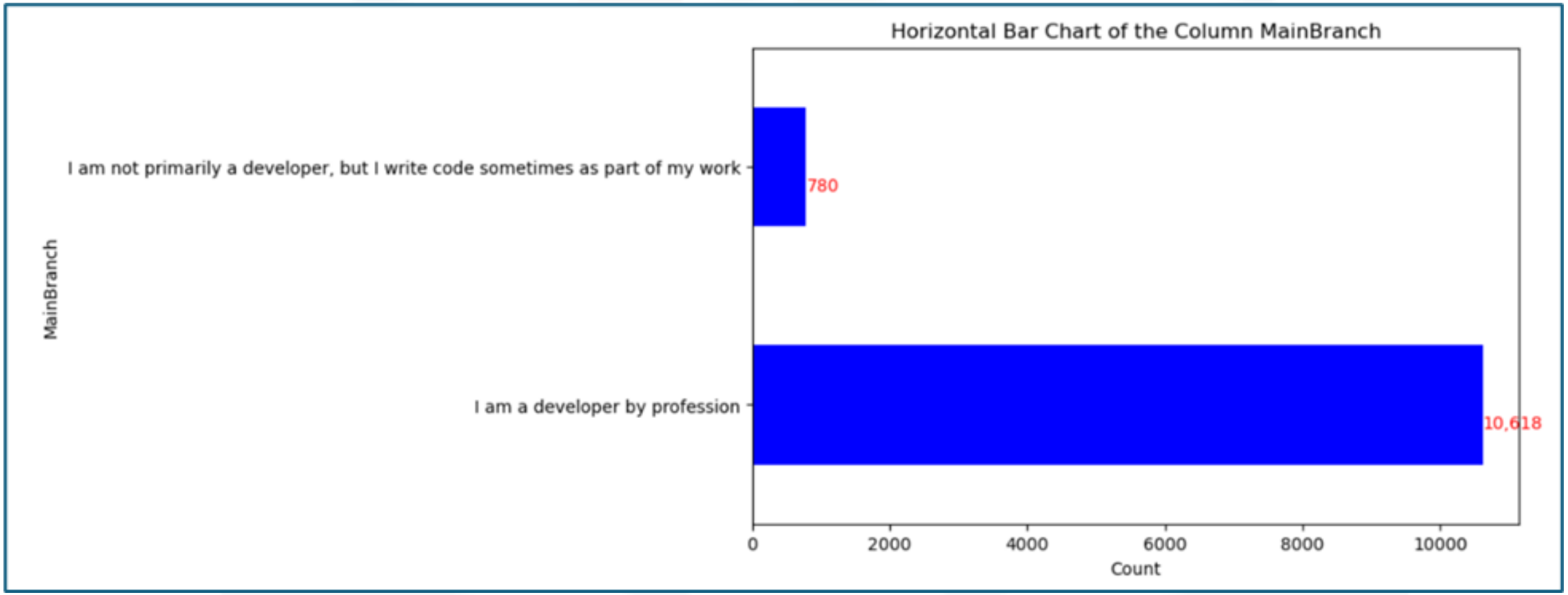
Implications

- It is relevant for job marketers to make a shift from the currently worked with web frames to the most desired **React.js** web frame in addition to **jQuery** in the next year.
- It is appropriate for employers to increasingly look for developers skilled in the **Docker** platform in addition to **Linux** & **Windows**.
- **Fewer** female respondents (**6.3%**) & increasing trend in count of younger respondents (**< 30yrs**) in the tech skills implies that females & older respondents are not welcomed in the tech sector.
- **Decreased** Respondent counts from less developed countries suggest that the tech skills are not accessible to them, or respondents cannot afford to take training in tech skills.

CONCLUSION

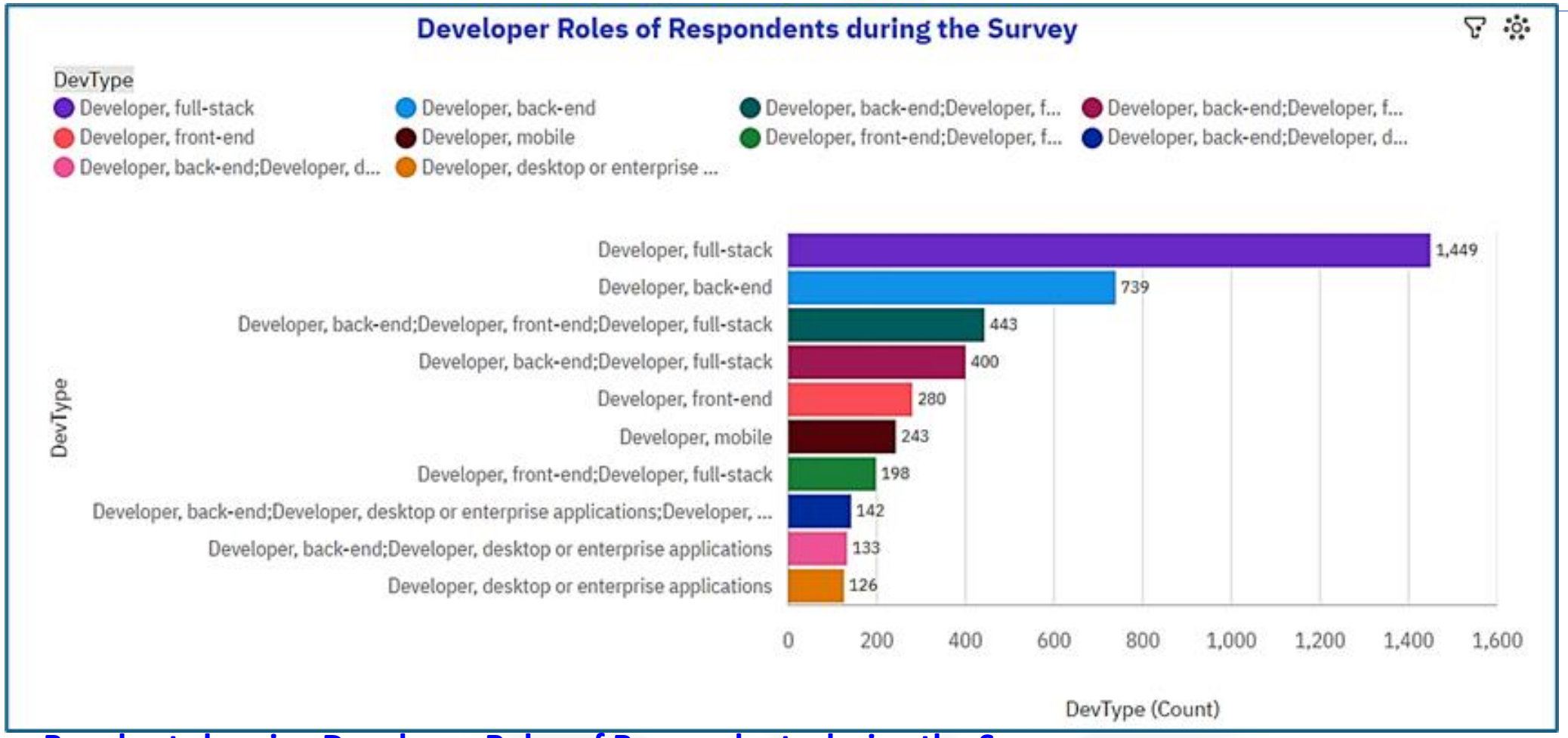
- **Python**, along with **JavaScript**, **HTML/CSS** & **SQL**, must be included in the Firm's list of the next year's job postings as this rapidly growing language shows a dramatic rise in demand from respondents in the next year.
- **MongoDB** should also be added to the list of top databases as it shows an impressive rise in demand from respondents & professional developers next year.
- **React.js** must be added in next year Job postings as the most desired & highly demanded web frame by respondents, along with **jQuery**, **JavaScript** competitively.
- **Docker** should be added to the list as the most desirable platform along with **Linux** & **Windows** platforms.
- The wider gender gap in tech must be closed by encouraging more females to enter to the tech field. Older respondents should not be discouraged but given more opportunities using internship and apprenticeship programs.
- Respondents in less developed countries should be given more access to high-tech skills through sponsored or funded training.
- So, the Data Analyst found it indispensable to adapt to trends in the emerging technologies & skills in the Global IT sector as the trends change from year to year worldwide.

APPENDIX



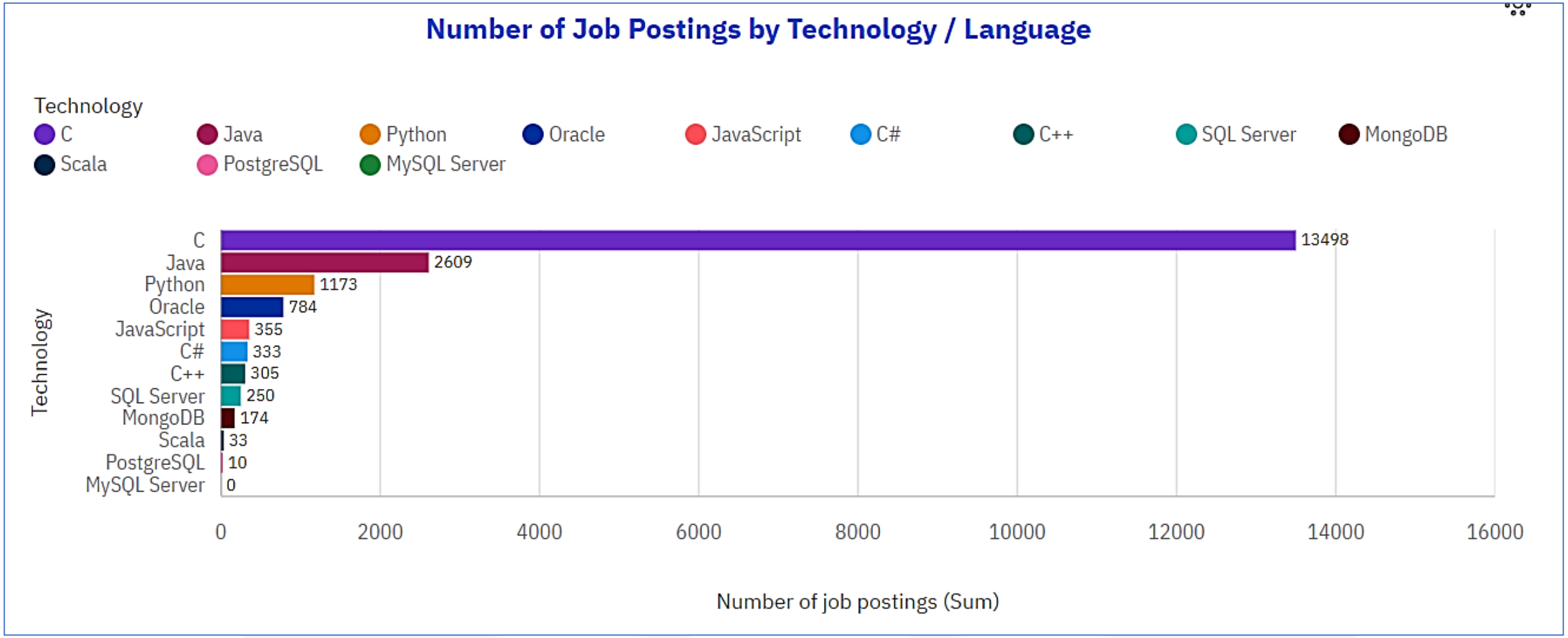
Horizontal Bar Chart showing Comparison of Professional & Non-professional Developers during the Survey.

APPENDIX…



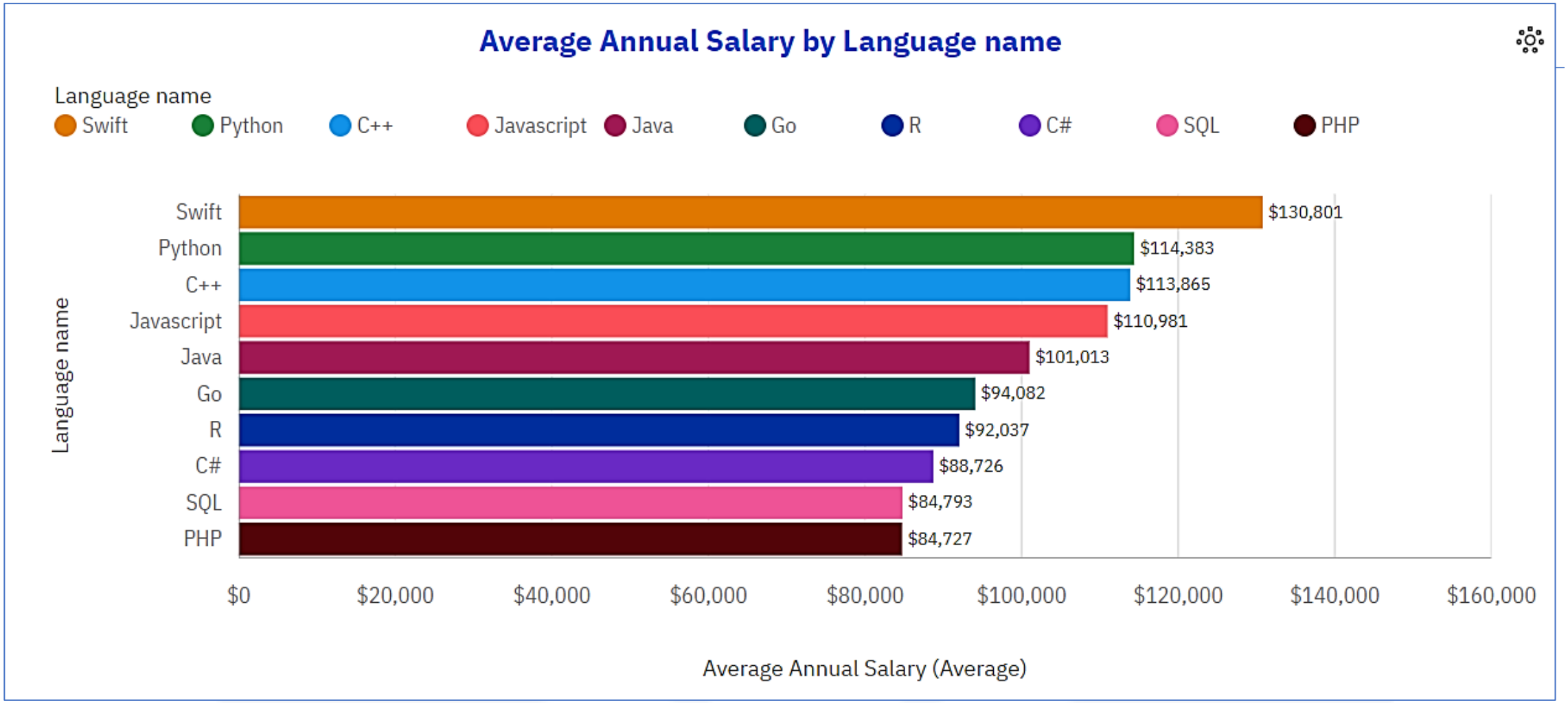
Bar chart showing Developer Roles of Respondents during the Survey.

JOB POSTINGS



Bar Chart showing the Job Posting by Technology / Language (Result for GitHub Job Posting API).

POPULAR LANGUAGES



Bar Chart showing Popular languages & their Average Annual Salary (Result for Survey Dataset Web scraping).

Source: Medium