IBM Data Analyst Capstone Project

Name Surname

2024-05-23

IBM Developer

SKILLS NETWORK

OUTLINE

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

IBM **Developer**

SKILLS NETWORK

EXECUTIVE SUMMARY

❑ It is essential to keep up with the ever-changing technologies in the global IT sector. This reports uses current technics and tools common used in data analyses to highlight important findings and trends in datasets.

❑ Data was gathered from a stack overflow surveys, IBM site and Github job board postings. Datasets were collected, cleaned, subjected to exploratory analysis, and visualized on dashboards.

❑ The findings showed that Javascript is the most popular programming language and is anticipated to be so in future. MySQL has the highest database usage at the moment but Postgre SQL is protected to have more demand in the future

IBM Developer

SKILLS NETWORK

INTRODUCTION

❑ Presentation reports uses data analytics to highlight current trends in the need for skill related to programming languages, databases platforms and web frames

❑ These questions were investigated using the data:
  ✓ Which programming languages are most in demand?
  ✓ What are the most in-demand database skills?
  ✓ What popular IDEs or web frames are there?

❑ The audience for this research are IT professionals, HR managers, and anybody else with an interest in IT sector who wants to learn and upgrade their skills and would like to know what will also still be relevant in the future.

IBM Developer

SKILLS NETWORK

METHODOLOGY

❏ Data in several formats such as the number of jobs currently available for different technologies and for different places, were gathered using the Github Jobs ARI on Python

❏ To obtain the names of the programming languages and their yearly wages, the IBM website was scrapped. The dataset from a 2019 Stack Overflow developer Survey was downloaded and used.

❏ Python was used to clean and analyze the datasets. To access the distribution of data, the presence of outliers, and the correlation between various columns in the dataset, an exploratory data analyses was carried out.

❏ Charts, graphs, and dashboards were created using Python and Cognos analytics tools to visualize the data. All the Python analyses were carried out on Jupyter notebook through Visual Studio.
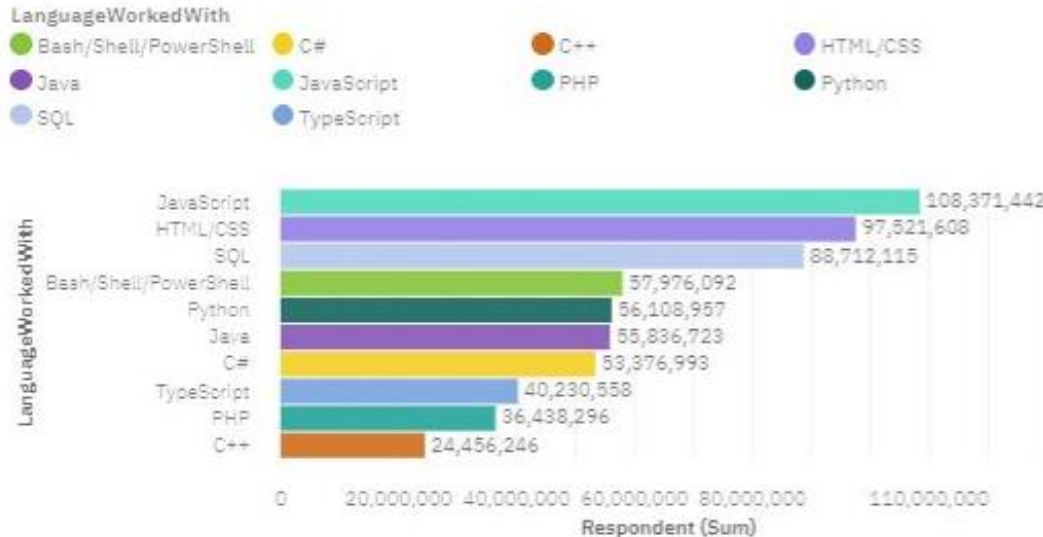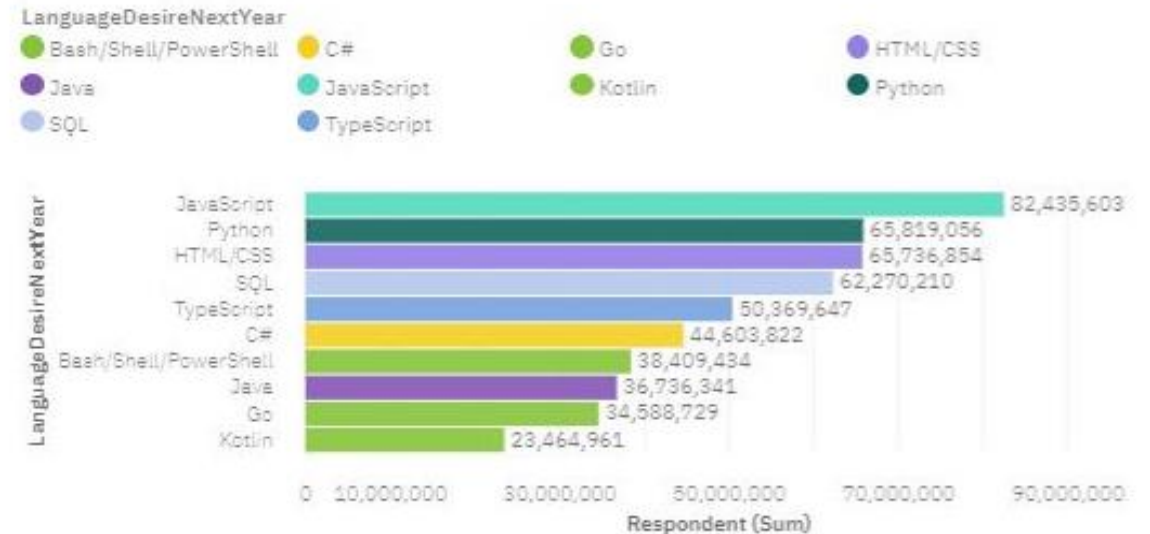
RESULTS

PROGRAMMING LANGUAGE TRENDS

Current Year

Top 10 LanguagesWorkedWith

LanguageWorkedWith
- Bash/Shell/PowerShell
- C#
- C++
- HTML/CSS
- Java
- JavaScript
- PHP
- Python
- SQL
- TypeScript

Language	Respondent (Sum)
JavaScript	108,371,442
HTML/CSS	97,521,608
SQL	88,712,115
Bash/Shell/PowerShell	57,976,092
Python	56,108,957
Java	55,836,723
C#	53,376,993
TypeScript	40,230,558
PHP	36,438,296
C++	24,456,246

Next Year

Top 10 Languages Desire Next Year

LanguageDesireNextYear
- Bash/Shell/PowerShell
- C#
- Go
- HTML/CSS
- Java
- JavaScript
- Kotlin
- Python
- SQL
- TypeScript

Language	Respondent (Sum)
JavaScript	82,435,603
Python	65,819,056
HTML/CSS	65,736,854
SQL	62,270,210
TypeScript	50,369,647
C#	44,603,822
Bash/Shell/PowerShell	38,409,434
Java	36,736,341
Go	34,588,729
Kotlin	23,464,961

PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

Javascript, HTML/CSS, SQL, Shell languages and Python are the most used languages currently.

Javascript, HTML/CSS, Python, SQL, and Typescript will be the most used languages next year and future years.

Python will have more demand than SQL next year.

Implications

Javascript and HTML are used for web development which means that web development as a tech skill has the highest demand, especially as Typescript is getting viral.

Python is gaining more and more traction due to the increase in demand for AI and ML skills.
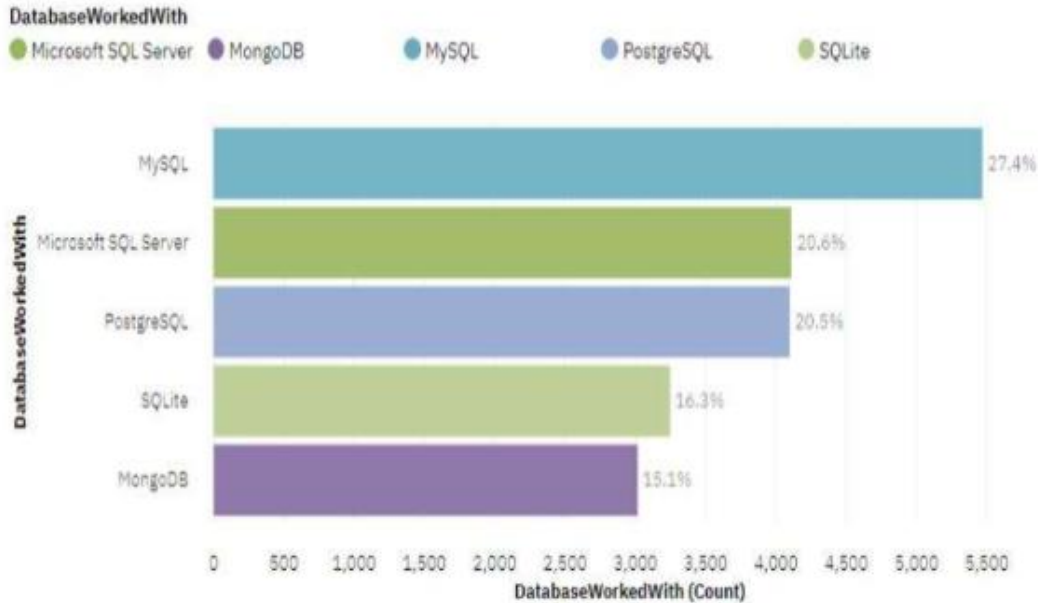
SQL is the still the most relevant language for data professionals. It is important for aspiring data analysts, scientists, business analysts etc to have SQL skills.
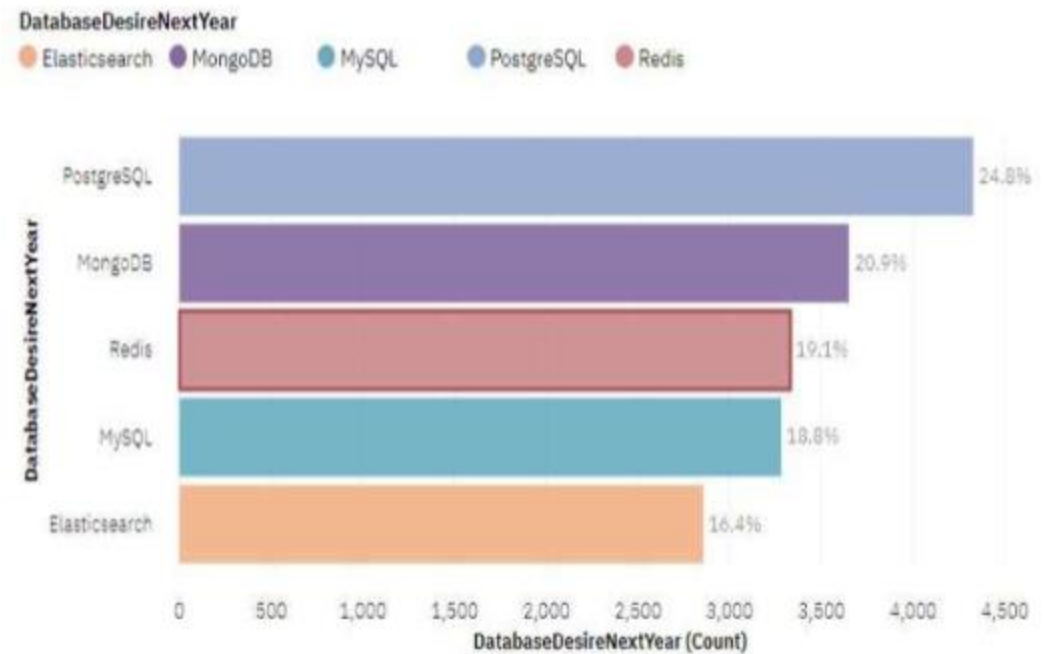
IBM Developer

SKILLS NETWORK

DATABASE TRENDS

Current Year

Next Year

Top 5 Databases in the Current year

DatabaseWorkedWith
● Microsoft SQL Server ● MongoDB ● MySQL ● PostgreSQL ● SQLite

MySQL — 27.4%
Microsoft SQL Server — 20.6%
PostgreSQL — 20.5%
SQLite — 16.3%
MongoDB — 15.1%

DatabaseWorkedWith (Count)

Top 5 Databases for the Next Year

DatabaseDesireNextYear
● Elasticsearch ● MongoDB ● MySQL ● PostgreSQL ● Redis

PostgreSQL — 24.8%
MongoDB — 20.9%
Redis — 19.1%
MySQL — 18.8%
Elasticsearch — 16.4%

DatabaseDesireNextYear (Count)

DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

MySQL, Microsoft SQL server, Postgre SQL, SQLite and MongoDB are the top 5 most used databases at the moment.

However, Postgre SQL, MongoDB, Redis, MySQL and Elasticsense are projected to become more popular in the future.

Redis and Elasticsense are relaatively new tools and are set to gain more traction in the IT space.

Implications

SQL is still a top tool to watch out for in data specialists.

Companies still prefer Open source databases.

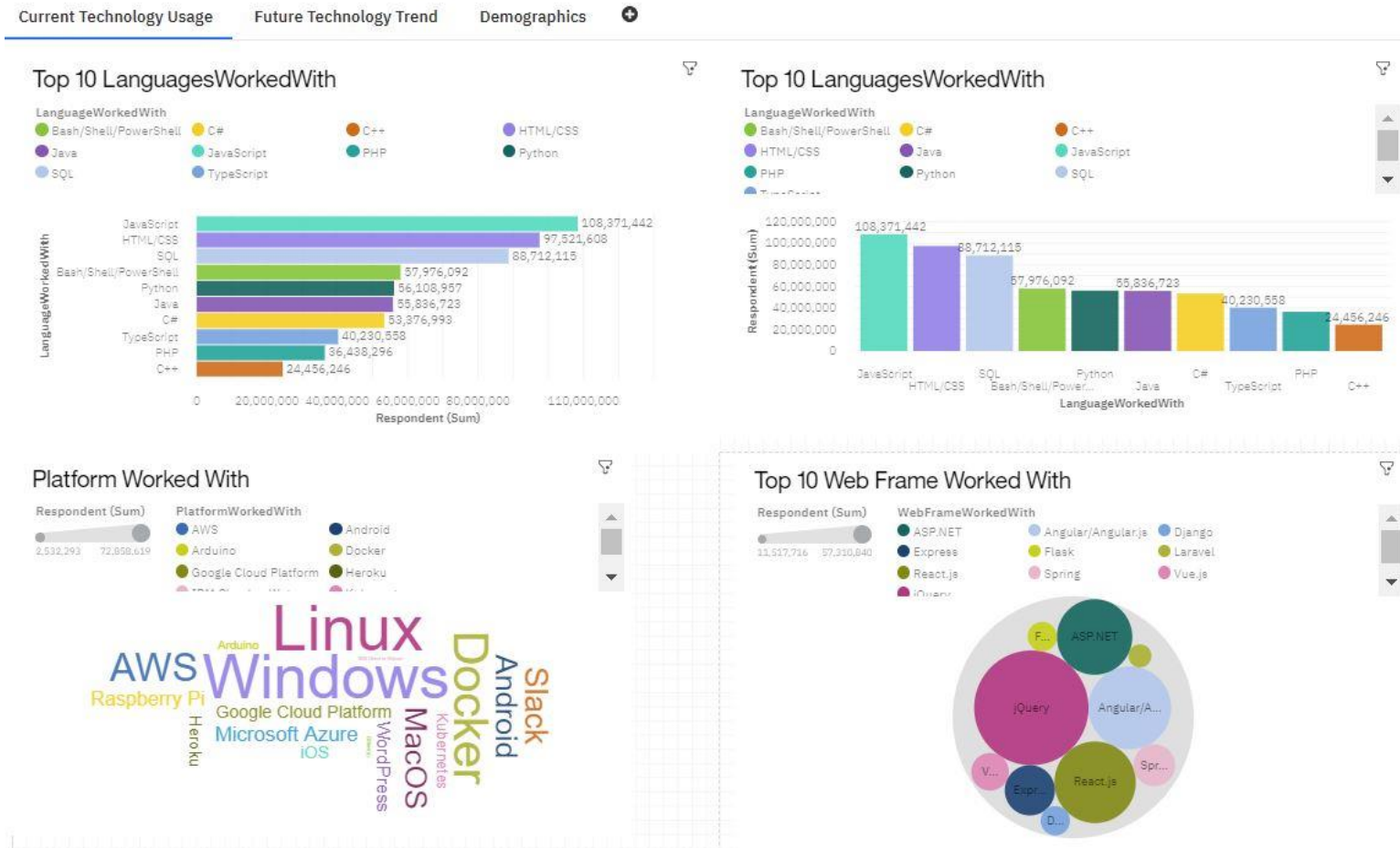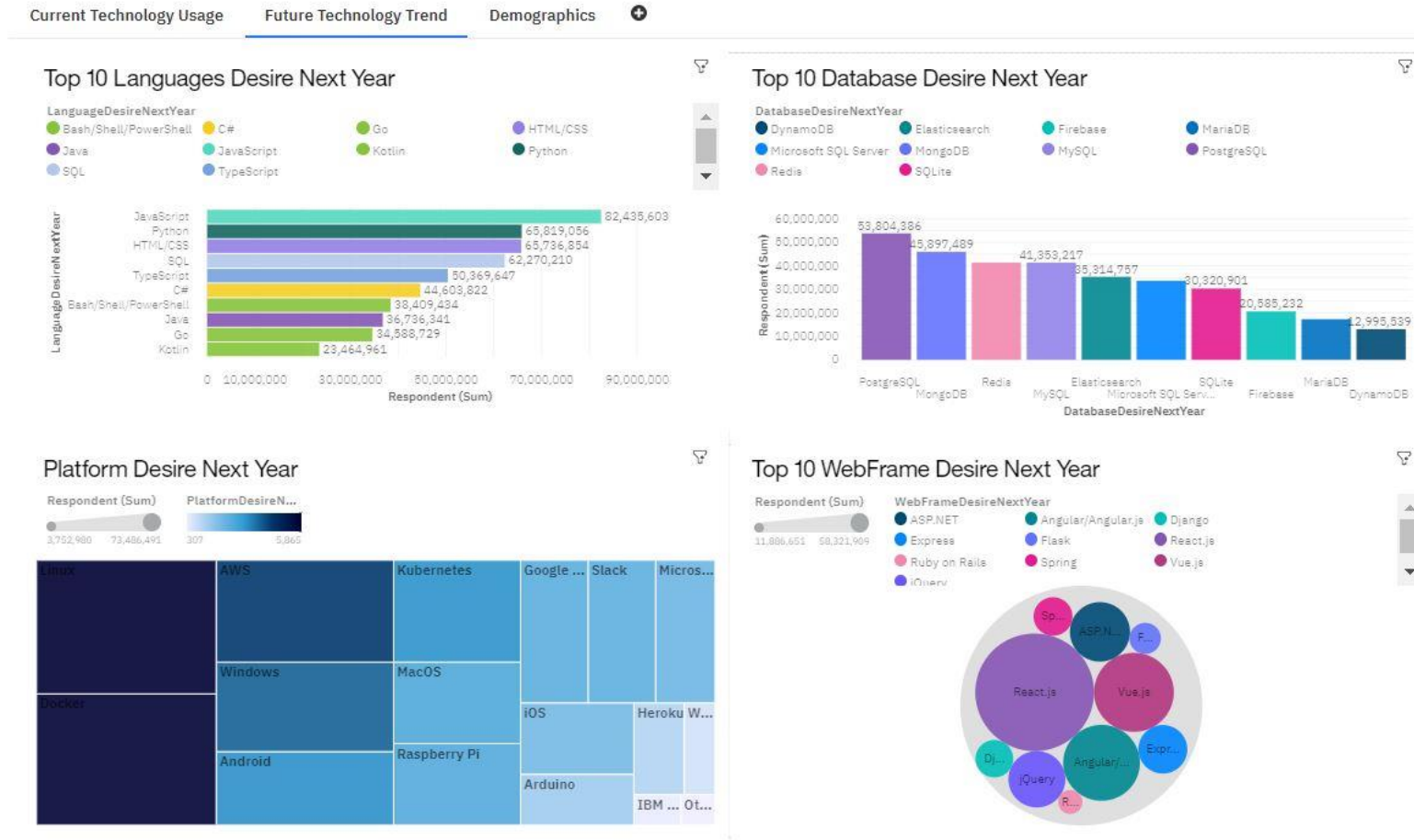Oracle SQL was not among the top 5. It is losing relevance as time passes.

IBM Developer

SKILLS NETWORK

DASHBOARD

Dashboard link

IBM Developer

SKILLS NETWORK
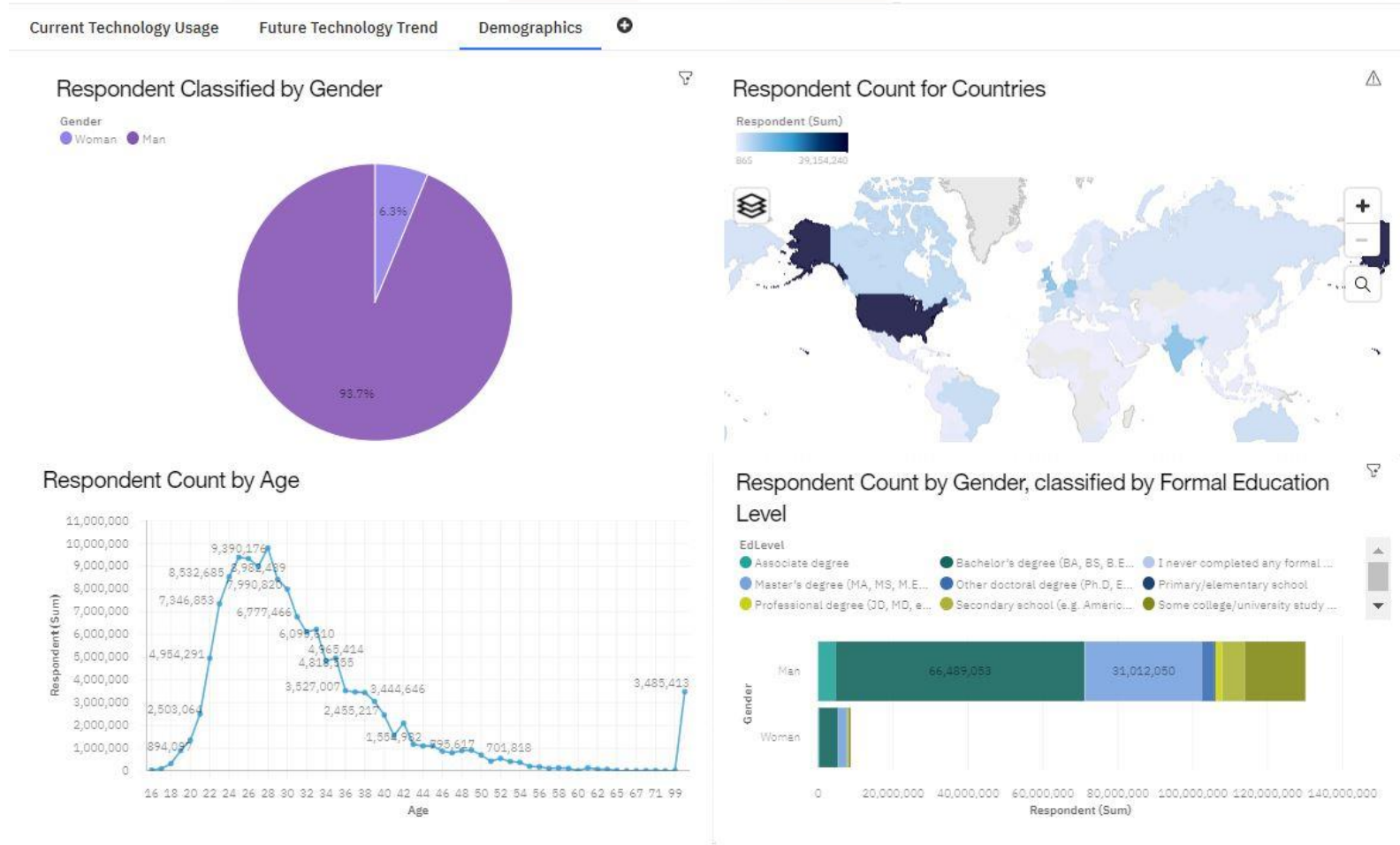
DASHBOARD TAB 1

DASHBOARD TAB 2

DASHBOARD TAB 3

DISCUSSION

❑Upskilling the Tech sector

❑How do we close the wide gender gap in the Tech sector?

❑Is completing a masters degree really a requirement?

❑More tech education, access and development in less developed regions in South East Asia, South America, Africa and parts of Europe

❑How relevant will Oracle SQL still be in the future?

IBM Developer

SKILLS NETWORK

OVERALL FINDINGS & IMPLICATIONS

Findings

❑ Most people in the Tech field have a Bachelors' degree

❑ Web development languages are the most popular and on-demand tools in the IT field

❑ Most responders want to learn Postgre SQL and React JS next year

Implications

❑ It is important for data professionals to develop their skills in NoSQL in addition to SQL databases

❑ Web development is still a very profitable skill

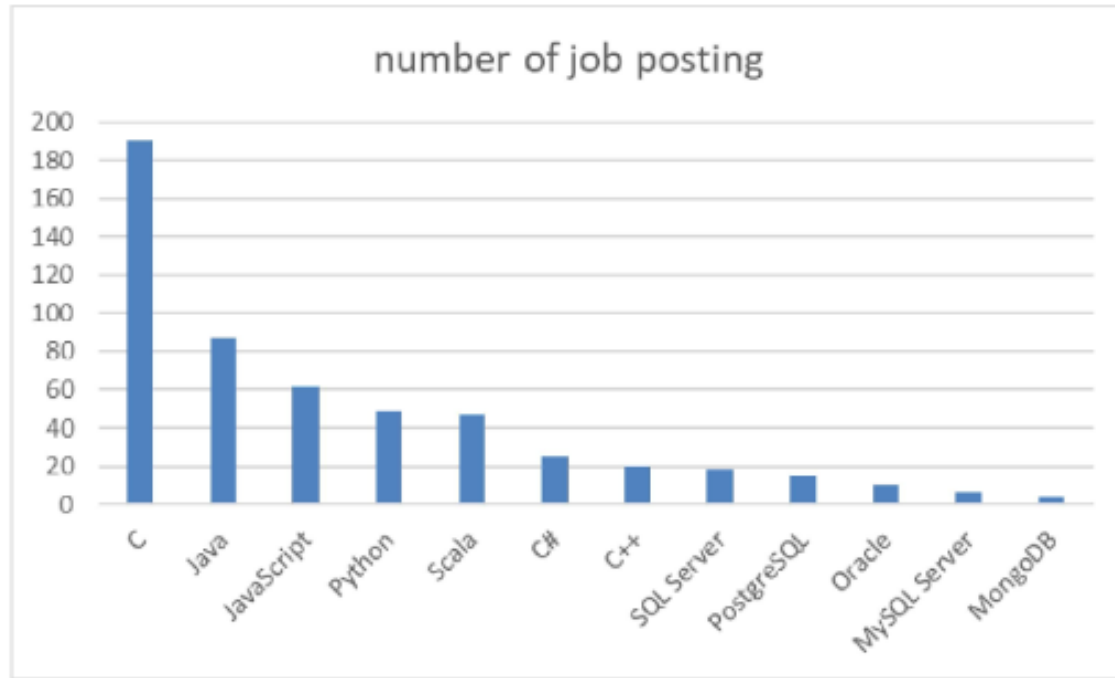❑ Less developed countries need more access to tech trainings and educations

CONCLUSION

- We wanted to analyze the dataset for the preferred current and future trends in technology while also considering the demographics of the respondents

- From the visualizations it is clear that the technology trends are preferred between the age groups between 21 and 43

- Among programming langauges, HTML and JavaScript are the most popular langauges to work with and will remain so in the coming years according to the future technology trends

- Among the databases to work with, we may see the popularity of MySQL decreasing and being replaced with PostgreSQL and MongoDB

- We may see an increase in the use of these technologies in other countries as Multinational Corporations diversify into this field to keep up with competition

JOB POSTINGS

number of job posting

POPULAR LANGUAGES

Popular languages and their average annual salary

Language	Average Annual Salary
Swift	130,801
Python	114,383
C++	113,865
Javascript	110,981
Java	101,013
Go	94,082
R	92,037
C#	88,726
SQL	84,793
PHP	84,727