

IBM Data Analyst Capstone Project: Analyzing Emerging Technology Trends



CHERONO MERCY

September 3rd, 2024

OUTLINE



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

EXECUTIVE SUMMARY



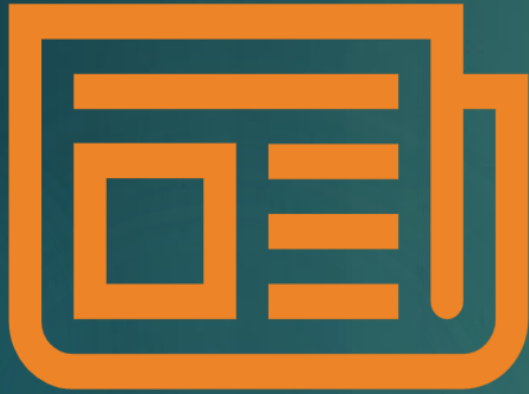
- ▶ To be competitive in the global IT sector, keeping up with the fast-paced technologies is essential. This report uses data analytics and visualisations to highlight current and projected trends in the need for skills related to programming languages, databases and other technologies. It also studies the demographics of professionals in the technology sector.
- ▶ Data was collected from a Stack Overflow survey and Github job postings. It was collected, cleaned, explored and visualised on dashboards.
- ▶ The findings indicated that JavaScript is the most popular programming language currently and is expected to be so in future. MySQL has the highest database usage currently but PostgreSQL is projected to have more demand in future.
- ▶ Moreover, the majority of the survey respondents were males aged 28 years from the USA.

INTRODUCTION



- ❖ This presentation report uses data analytics and visualisations to highlight current and expected trends in the demand for programming languages, databases, platforms and web frames.
- ❖ The following questions were investigated using the data:
 1. Which programming languages are most in demand today?
 2. What are the most in demand database skills?
 3. What popular IDEs or Web frames are there?
- ❖ The target audience for this research is IT professionals, IT students, HR managers and anyone interested in the IT sector, who yearns to learn about the top-demanded technology skills that will still be relevant in future.

METHODOLOGY



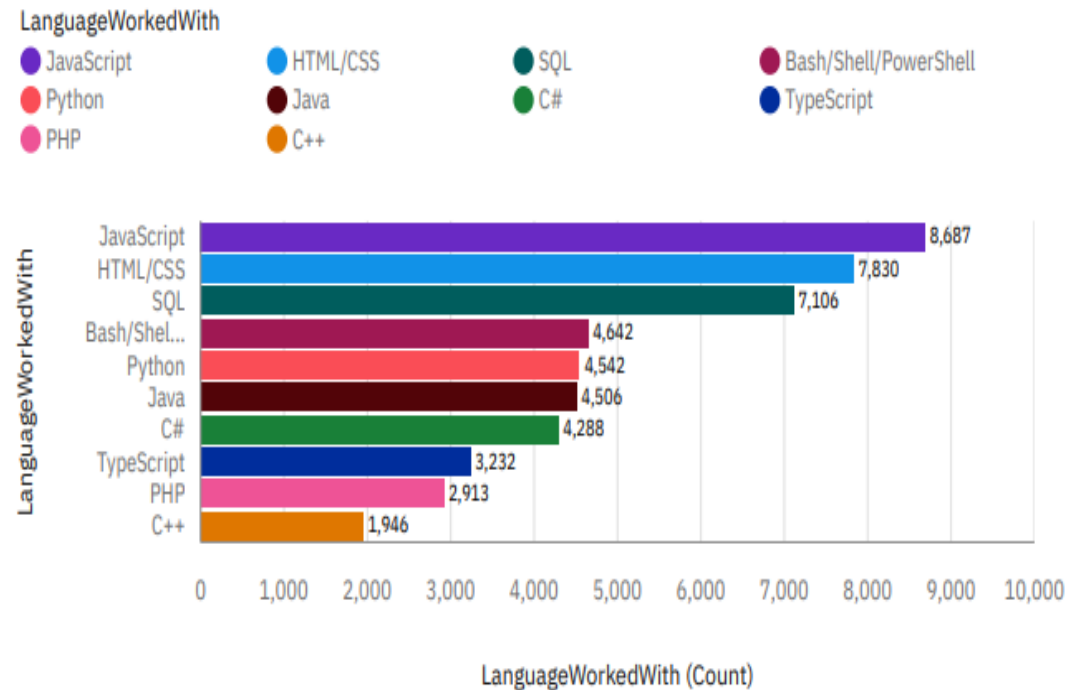
1. Data in several different formats such as, the number of jobs currently available for different technologies, were gathered using the GitHub jobs API in Python.
2. The IBM website was scraped to obtain the names of the programming languages and the annual salaries. The 2019 Stack Overflow developer survey dataset was downloaded and saved.
3. Python programming language was used to clean and analyse the data. An exploratory data analysis was done to assess the distribution of data, the presence of outliers and the correlation between various columns in the dataset.
4. Python libraries were used to create charts and graphs. IBM Cognos analytics was used to create dashboards. All the python analyses were performed on Jupyter notebook through a virtual environment.

RESULTS

PROGRAMMING LANGUAGE TRENDS

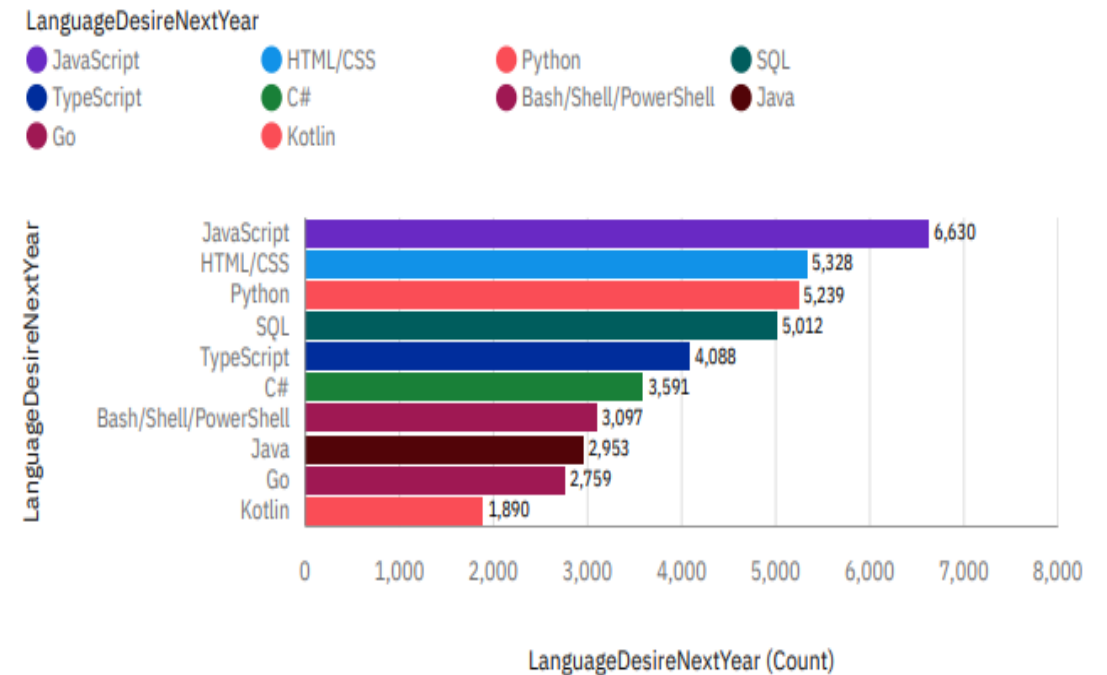
Current Year

Bar Chart of Top 10 LanguageWorkedWith



Next Year

Bar Chart of Top 10 LanguageDesireNextYear



PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

- ▶ Javascript, HTML/CSS, SQL, Shell languages and Python are the most currently used languages.
- ▶ Javascript, HTML/CSS, Python, SQL and Typescript are expected to be utilised the most next year and in future.
- ▶ Python will have more demand than SQL next year.

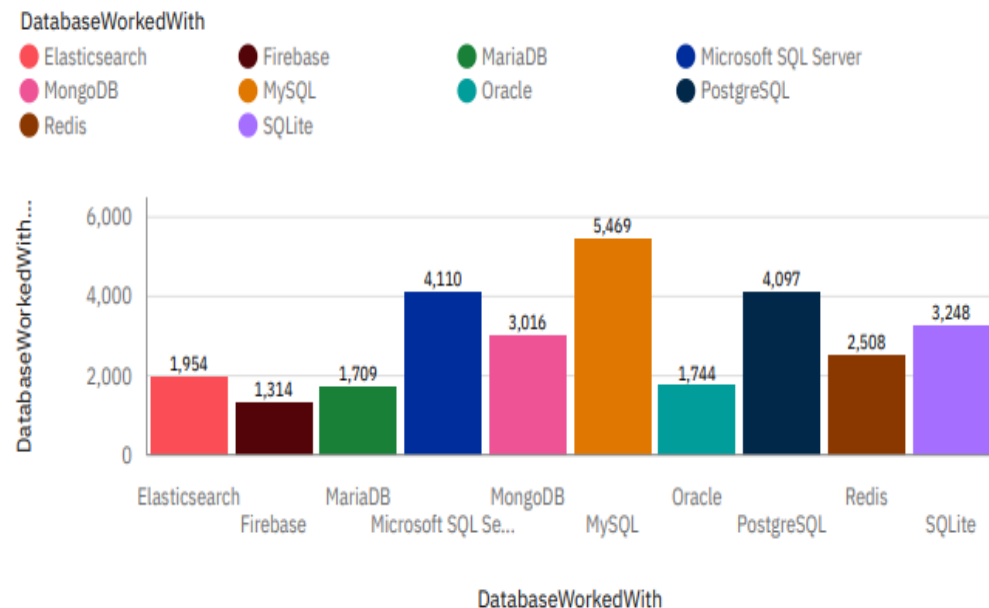
Implications

- ▶ Javascript and HTML are in high demand since they are used for web development and especially as Typescript is getting popular.
- ▶ Python is gaining traction due to the increase in demand for AI and ML skills.
- ▶ SQL is still the most relevant language for data professionals, thus it is important to have SQL skills.

DATABASE TRENDS

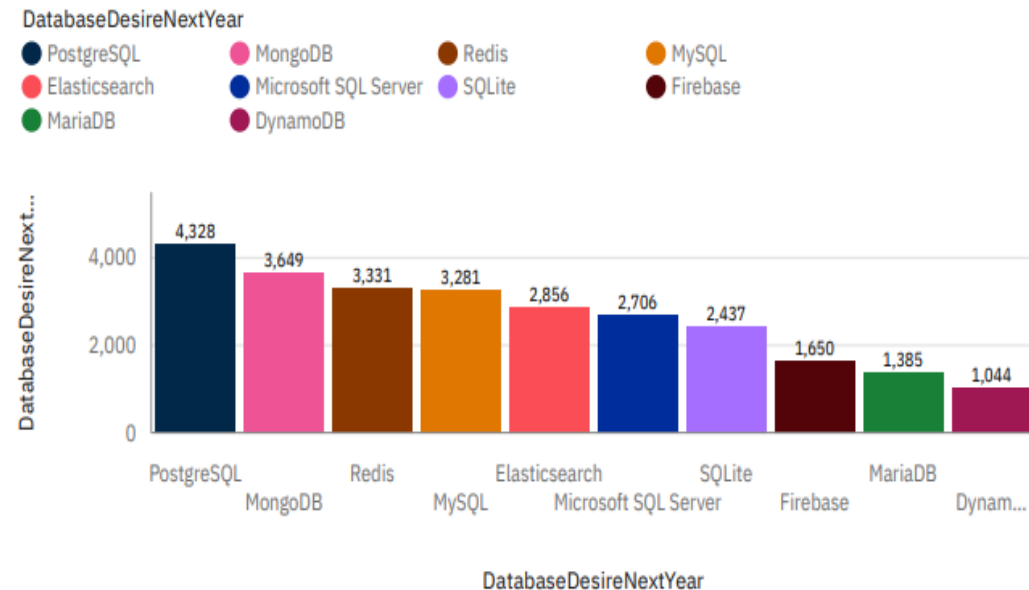
Current Year

Column Chart of Top 10 DatabaseWorkedWith



Next Year

Column Chart of Top 10 DatabaseDesireNextYear



DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- ▶ MySQL, Microsoft SQL Server, PostgreSQL, SQLite and MongoDB are the top 5 most used databases currently.
- ▶ PostgreSQL, MongoDB, Redis, MySQL and Elasticsearch are projected to become more popular next year.
- ▶ Redis and Elasticsearch are relatively new technological tools and are expected to gain popularity in the IT sector.

Implications

- ▶ SQL is still a top tool in the IT profession.
- ▶ Companies prefer Open Source databases such as MySQL.
- ▶ Technologies such as Oracle SQL are losing relevance as time passes.

DASHBOARD



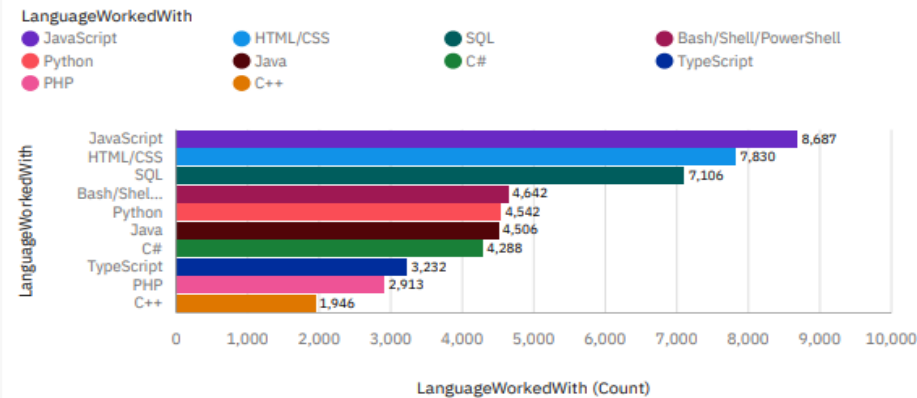
The permanent link to the IBM Cognos dashboard:

https://github.com/MCHERONO137/assignmentrepo/blob/main/_%20New%20Cognos%20Dashboard%202.pdf

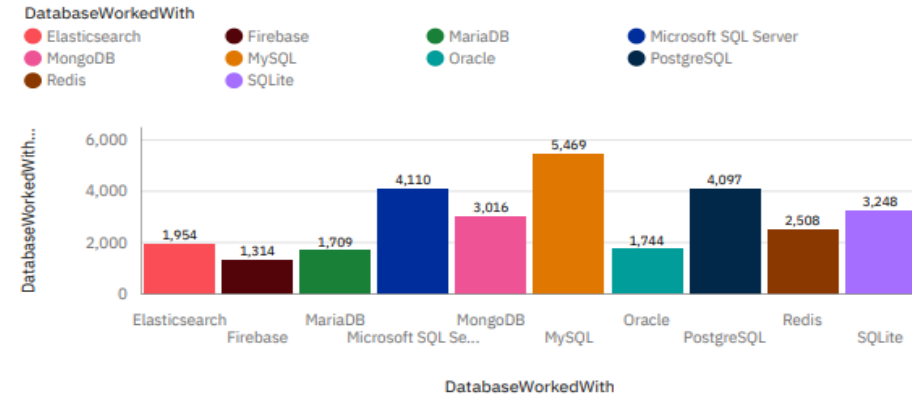
DASHBOARD TAB 1

Current Technology Usage

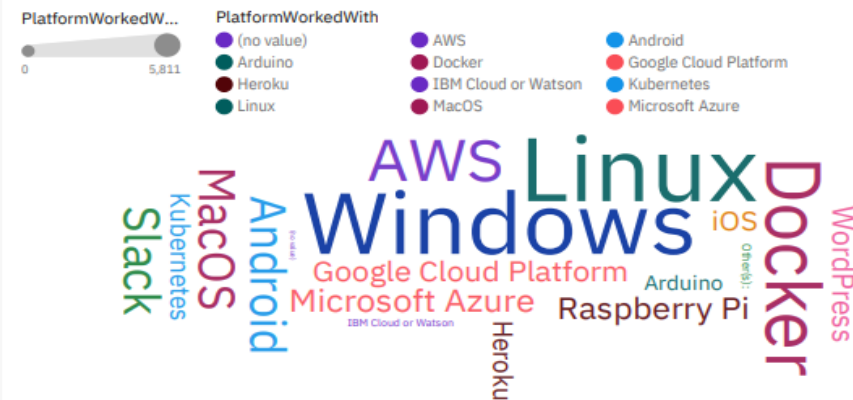
Bar Chart of Top 10 LanguageWorkedWith



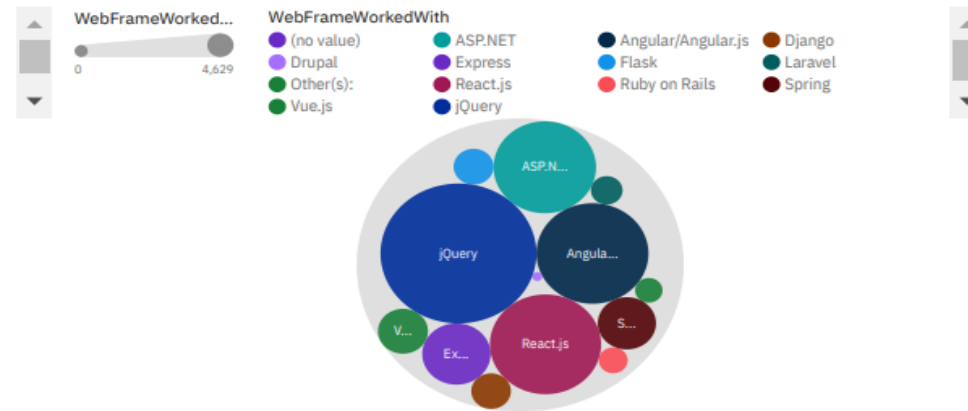
Column Chart of Top 10 DatabaseWorkedWith



Word Cloud Chart of PlatformWorkedWith



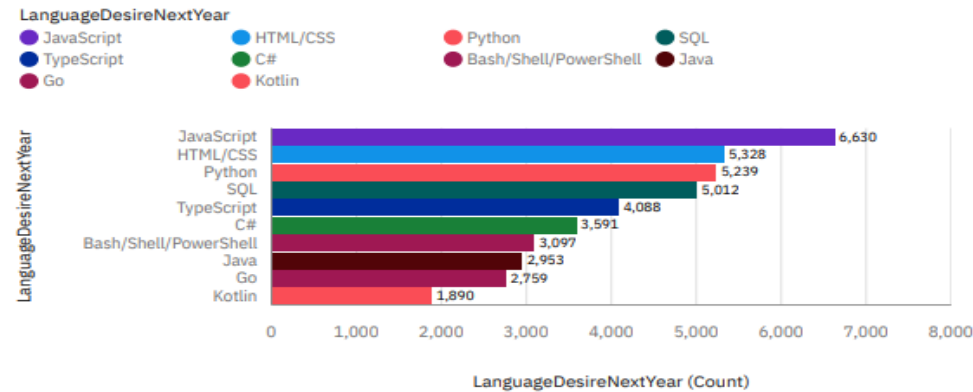
Hierarchy Bubble Chart of Top 10 WebFrameWorkedWith



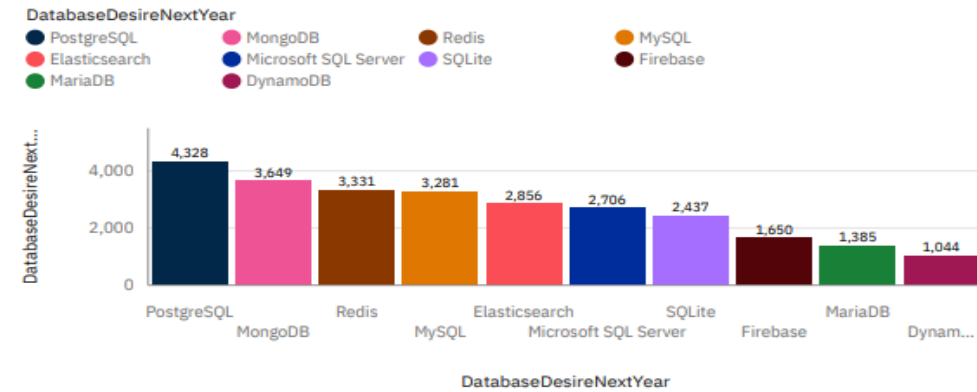
DASHBOARD TAB 2

Future Technology Trend

Bar Chart of Top 10 LanguageDesireNextYear



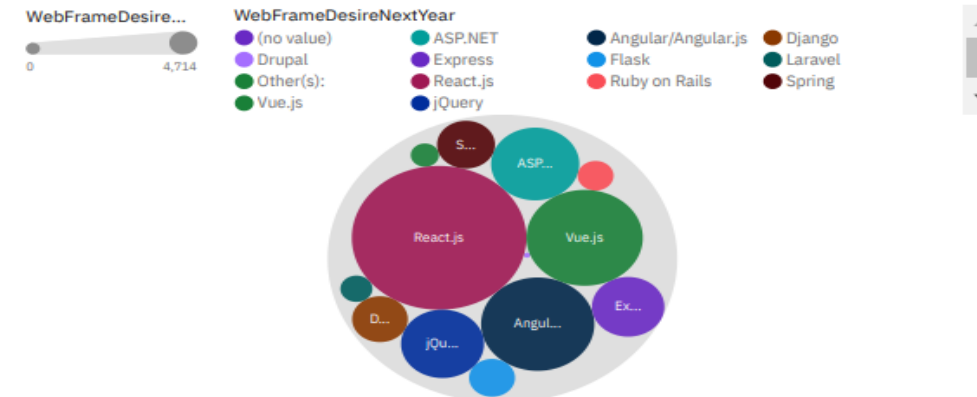
Column Chart of Top 10 DatabaseDesireNextYear



Tree Map Chart of PlatformDesireNextYear



Hierarchy Bubble Chart of Top 10 WebFrameDesireNextYear

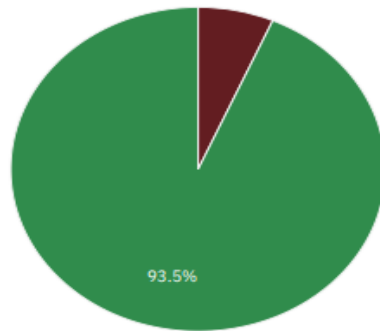


DASHBOARD TAB 3

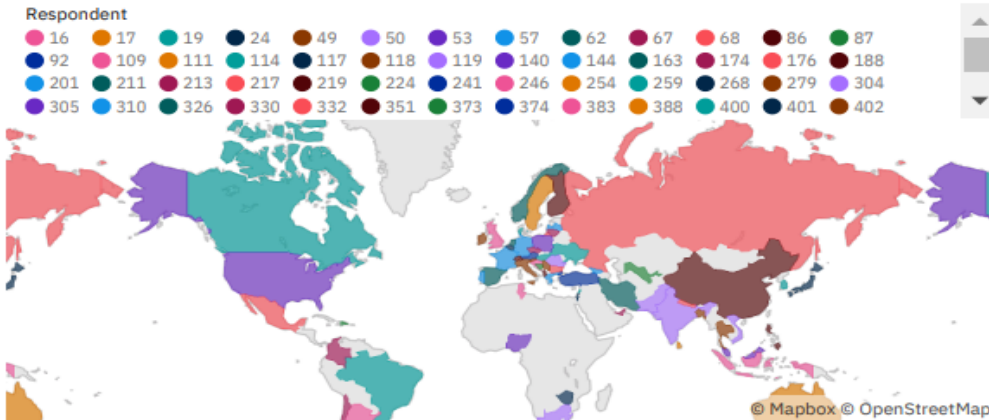
Demographics

Pie Chart of Respondent Count by Gender

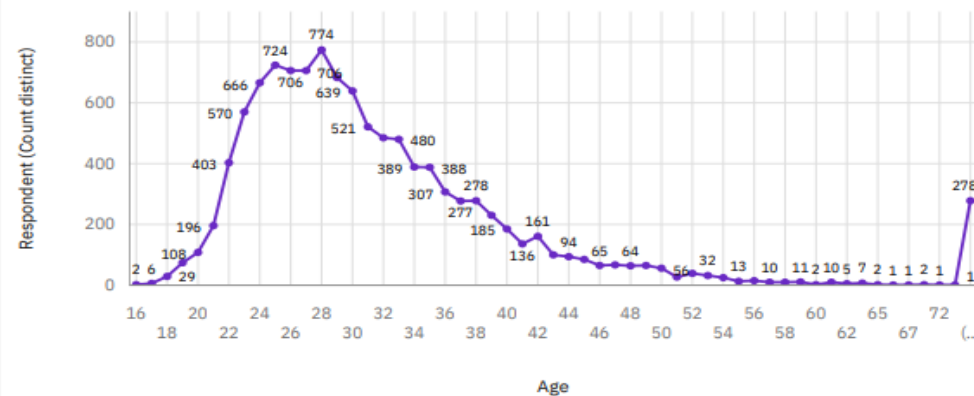
Gender
● Woman ● Man



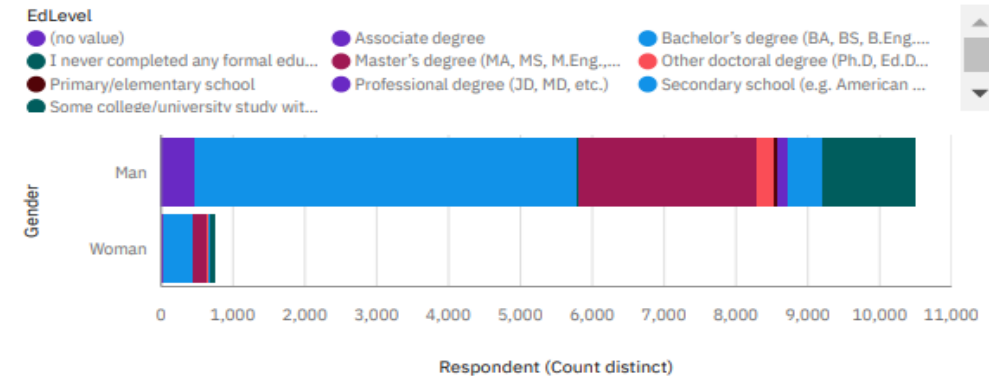
Map Chart of Respondent Count for Countries



Line Chart of Respondent Count by Age



Stacked Bar Chart of Respondent Count by Gender and classified by Formal Education Level



DISCUSSION

- ▶ It is quite clear that IT professionals should upskill to remain relevant in the IT sector.
- ▶ How do we close the wide gender gap in the technology sector?
- ▶ Is completing a masters or doctorate mandatory?
- ▶ More technology education and access are required in less developed regions, like in South East Asia, South America, Africa and some parts of Europe.
- ▶ Will OracleSQL be relevant in the future?

OVERALL FINDINGS & IMPLICATIONS

Findings

- ▶ Web development languages are the most popular and in-demand skills in the IT field currently.
- ▶ Most respondents want to learn PostgreSQL and React.js next year.
- ▶ The IT sector has young people of age under 40 years.
- ▶ Most people in the IT field have a Bachelor's degree.

Implications

- ▶ Web development is still a very lucrative skill.
- ▶ It is important for data professionals to develop proficiency in NoSQL on top of SQL databases.
- ▶ Less developed countries need more access to technological training and education.

CONCLUSION

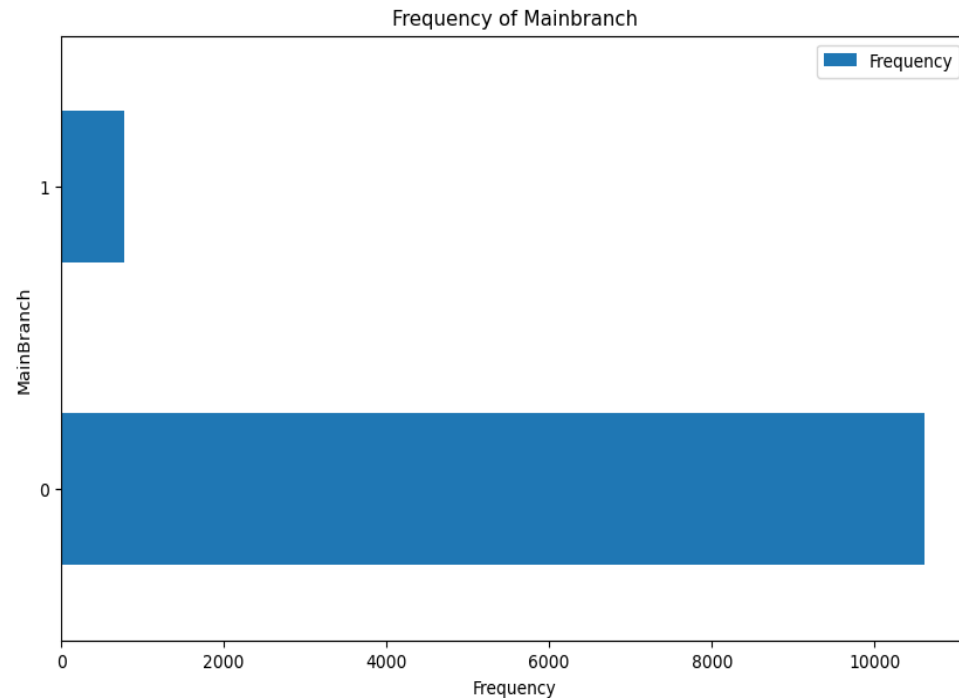


- ▶ It is important to stay updated with the current technologies in this fast-paced world.
- ▶ There is a need to solve the gender gap.
- ▶ Less developed countries need more access and education on technological skills.

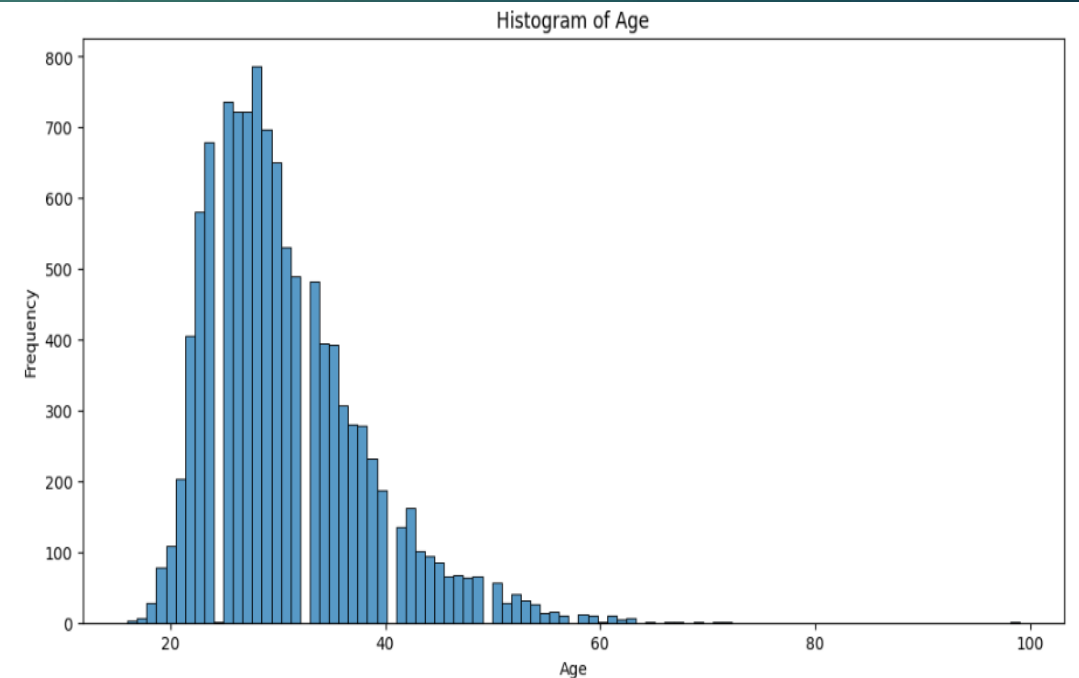
APPENDIX

```
In [40]: # your code goes here
Query8 = """SELECT MainBranch, count(*) as Frequency FROM master GROUP BY MainBranch"""
bar_df = pd.read_sql_query(Query8, conn)
bar_df

bar_df.plot(kind = 'barh', figsize=(10,6))
plt.xlabel('Frequency')
plt.ylabel('MainBranch')
plt.title('Frequency of Mainbranch')
plt.show()
```

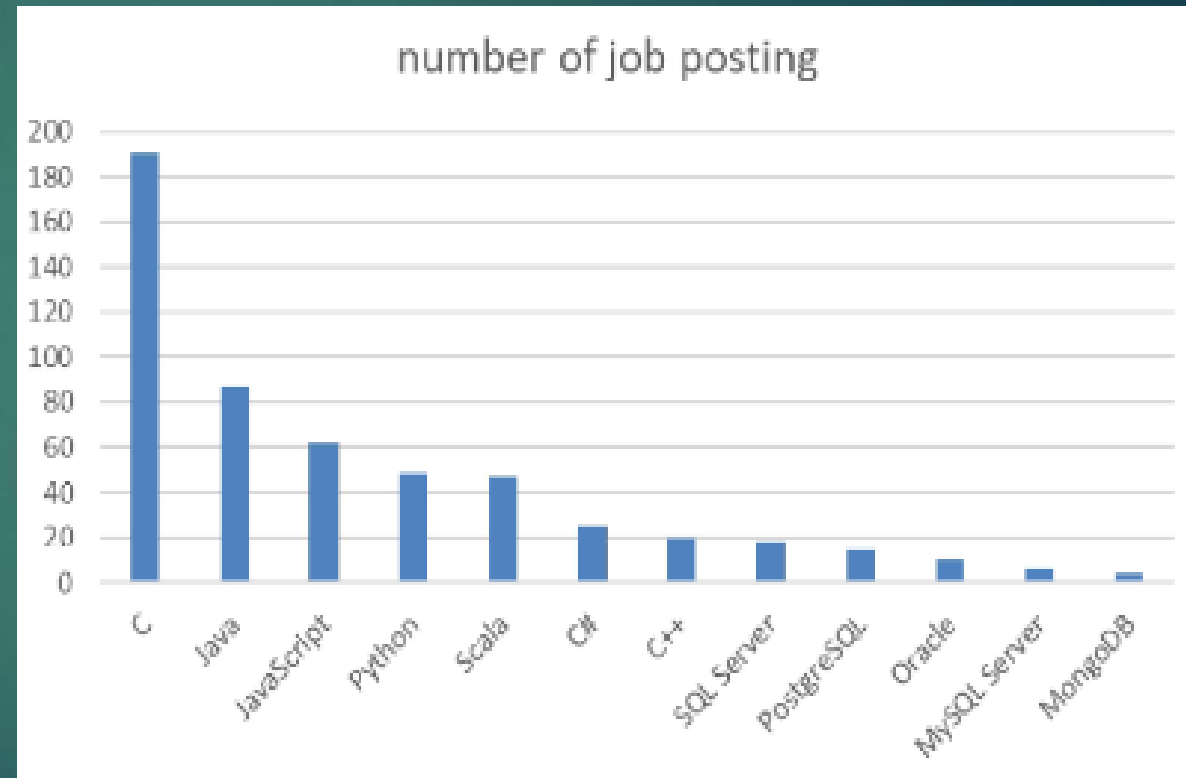


| | |
|---------------------------|-----------|
| Respondent | 0.004041 |
| CompTotal | 0.006970 |
| ConvertedComp | 0.105386 |
| WorkWeekHrs | 0.036518 |
| CodeRevHrs | -0.020469 |
| Age | 1.000000 |
| Name: Age, dtype: float64 | |



JOB POSTINGS

Bar chart presenting the job posting data collected using GitHub Job API.



POPULAR LANGUAGES



Bar chart displaying popular languages and their average annual salaries. The data was collected through webscraping the GitHub jobs data and saved in a csv file.