IBM Data Analyst Capstone Project

Shril Patel

January 23, 2025

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results: (Visualization-Charts/Dashboard)
- Overall Findings and Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY

- This comprehensive report explores the dynamic landscape of the IT industry, providing valuable insights into the most sought-after programming languages, databases, platforms, and web frameworks
- Notably, HTML/CSS, JavaScript, PowerShell, Java, and SQL emerge as the top programming languages in demand, with MySQL, Microsoft SQL Server, PostgreSQL, MongoDB, Redis, and SQLite leading the database landscape. Additionally, Google Cloud Platform, AWS, IBM Cloud, Windows, and Microsoft Azure are identified as the top platforms sought after in the industry. React.js, Angular, JQuery, ASP.NET, and Express stand out as the most utilized Web Frameworks.
- As we delve into the present demand landscape, it's crucial to anticipate future technology trends for comprehensive insights. By considering emerging technologies, this report offers invaluable guidance to IT professionals and organizations, empowering them to make informed strategic decisions and proactively address skill development initiatives.

INTRODUCTION

This comprehensive report delves into the evolving landscape of the IT industry, offering valuable insights into emerging skills and technologies.

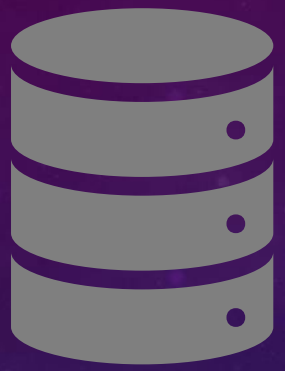
- The following inquiries were investigated using the data:
  1) Which Programming languages are most in demand today?
  2) Which database technologies are currently most sought after?
  3) What popular IDEs or WEB frames are there?

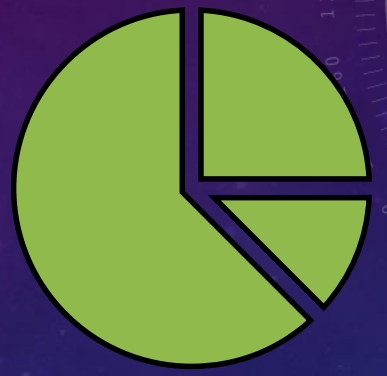
METHODOLOGY

Two Datasets used were scraped from a modified subset of data available on Stack Overflow consisting of present and future technology trends and the demographics of the respondents in the survey

Python was used to clean and analyze the data; the presence of outliers, and correlation between various columns in the dataset, an exploratory data analysis was carried out.

Charts, graph and dashboards were created using Python and Cognos analytics to Visualize the data. All the Python analyses were carried out on Juypter notebook through Visual Studio Code

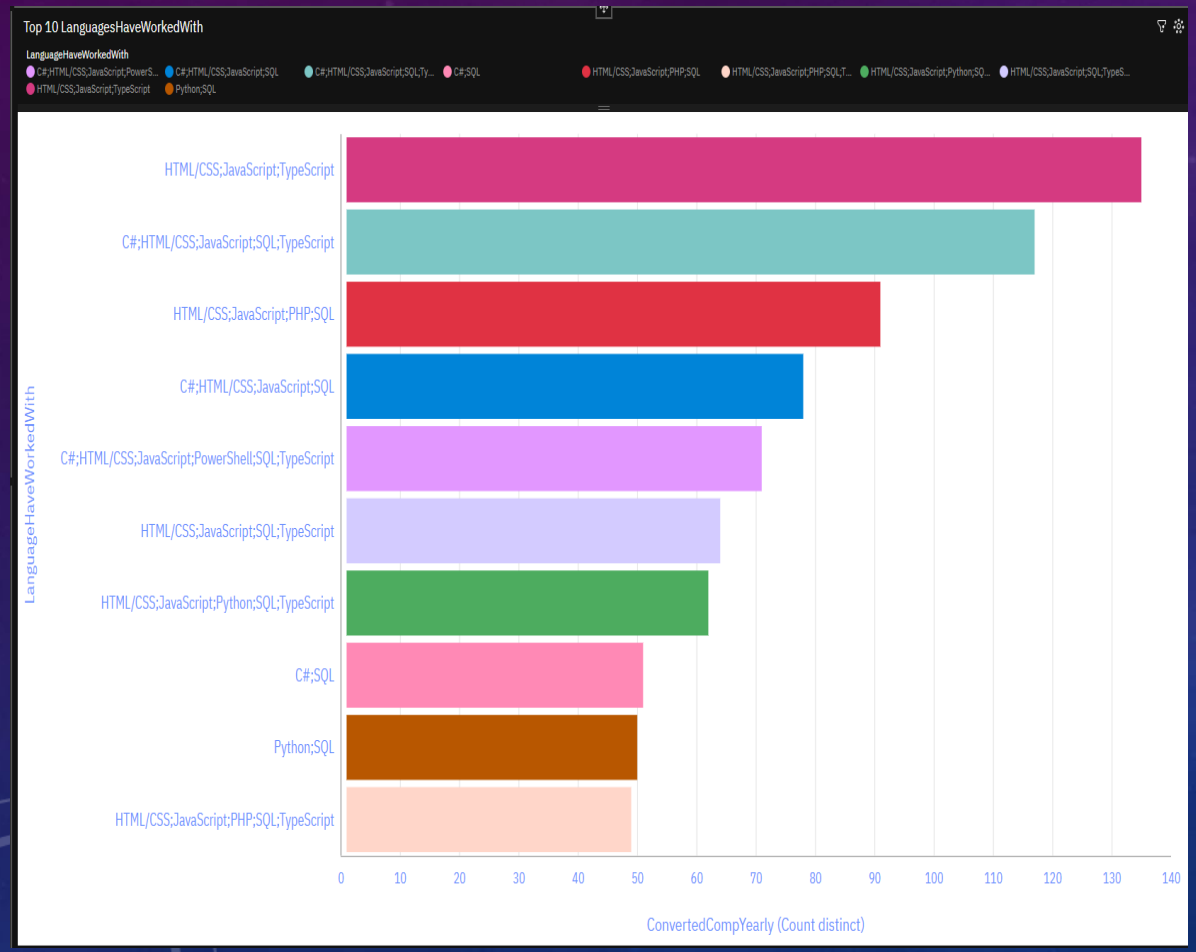




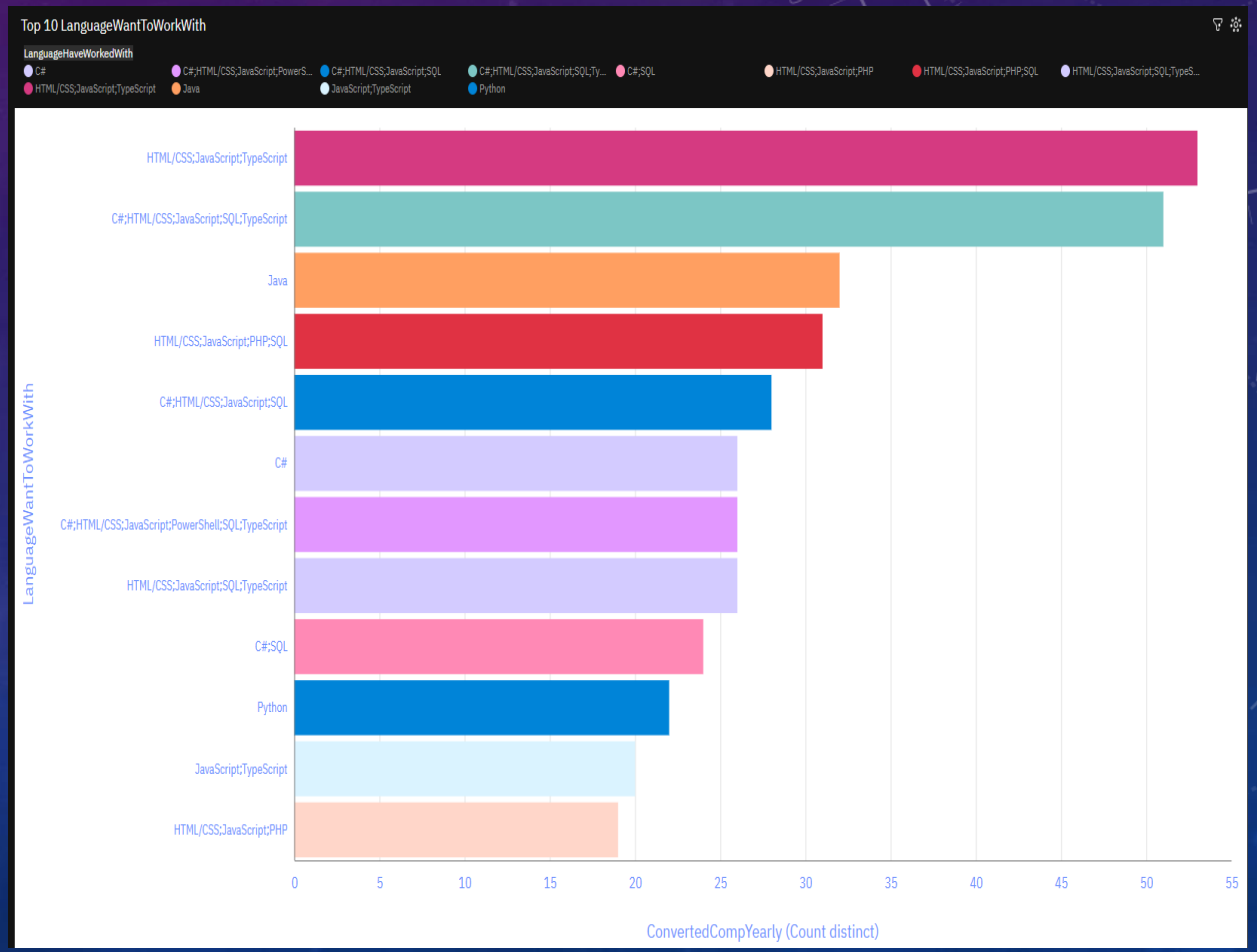
Results:

Programming Language Trends

Current Year



Next Year



Programming Language Trends – Findings & Implications

Finding:

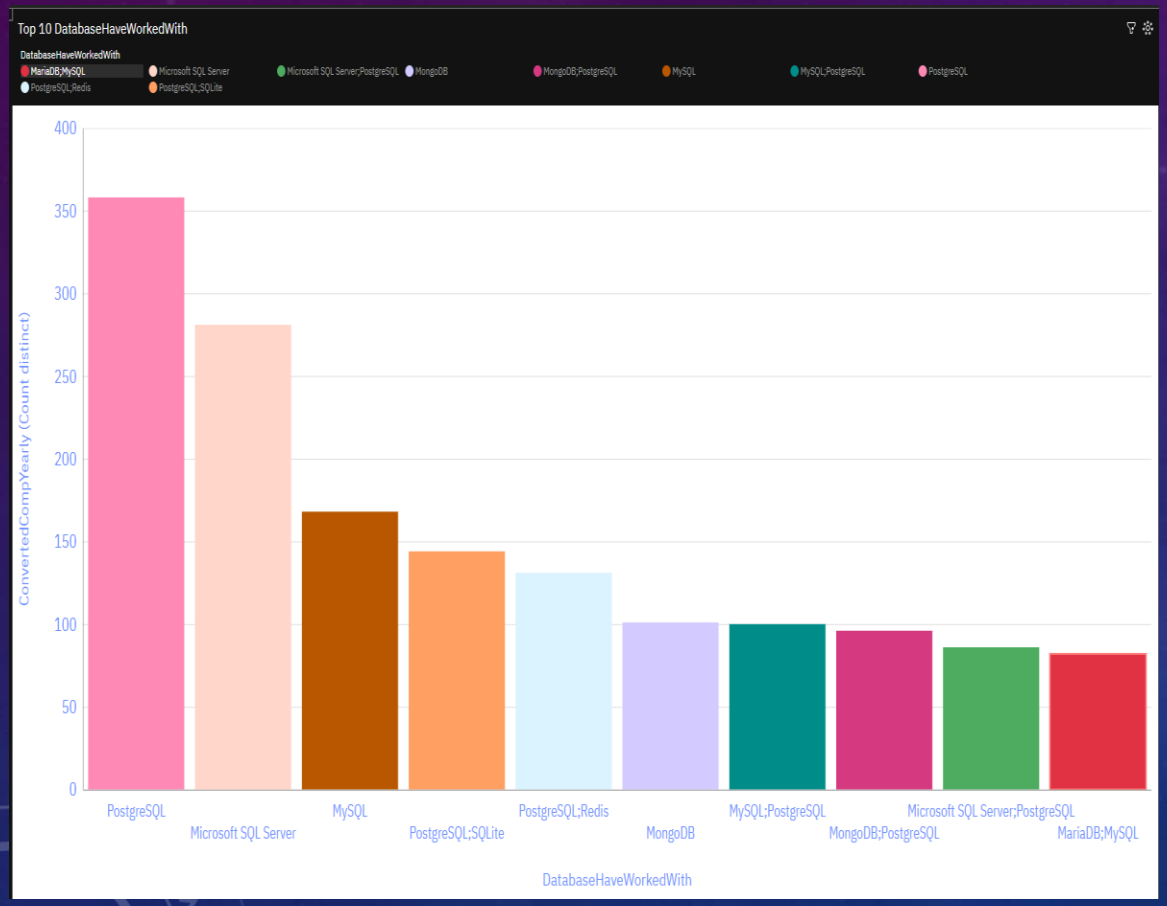
- HTML/CSS, the standard markup language for web pages, remains pivotal for web development, signaling its enduring popularity.
- JavaScript, a dynamic scripting language, empowers interactive content creation and multimedia control, securing its position as the second most prominent language, set to maintain its status in the future.
- Java, widely used for web application development, faces potential challenges from emerging contenders like Python and Go in future market dynamics
- C#, a versatile high-level language, witnesses increasing demand and is poised for further utilization in the upcoming year

Implications:

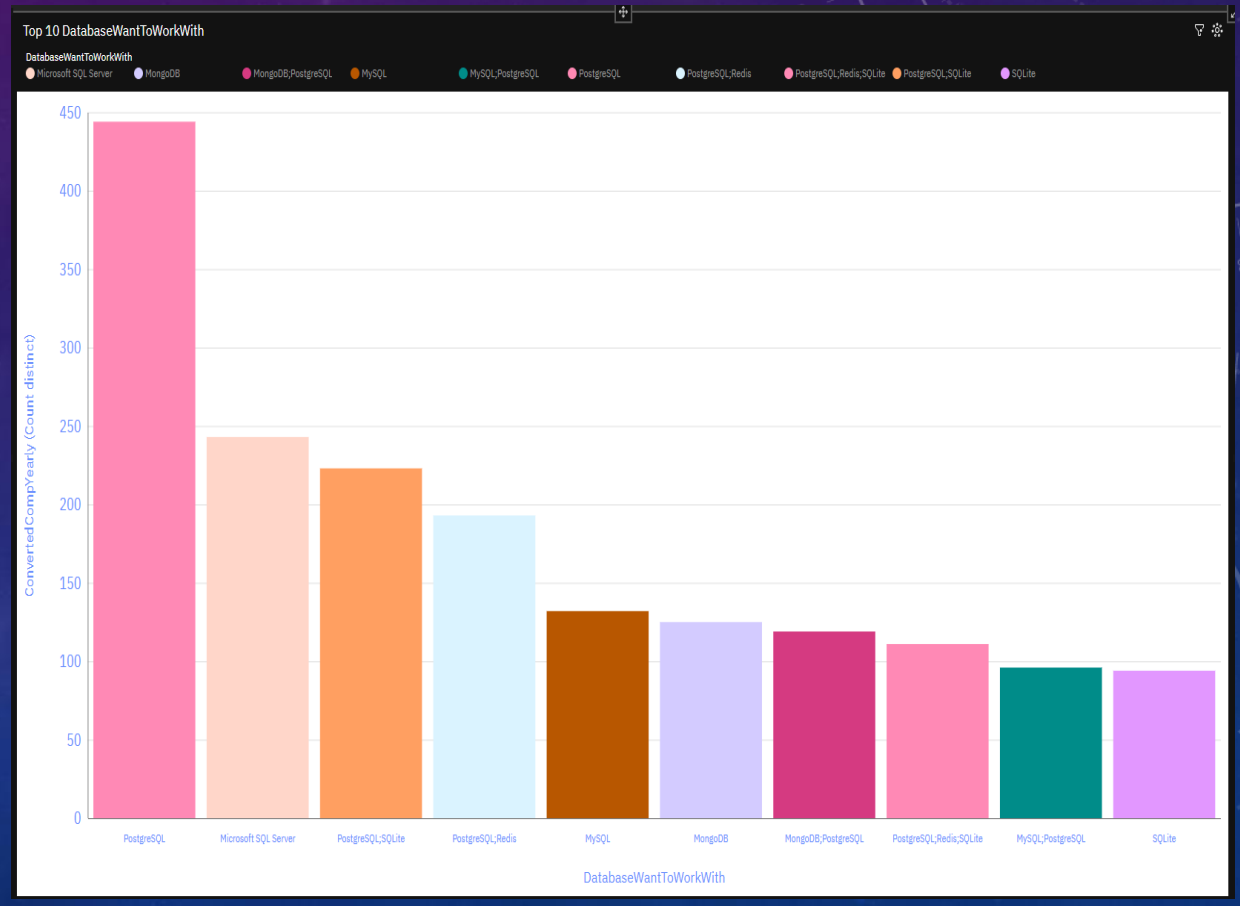
- JavaScript's sustained prominence in enabling dynamic web content highlights its critical role in delivering interactive user experiences, necessitating ongoing investment in JavaScript based technologies.
- Python is gaining more traction due to the increase in demand for AI and ML skills.
- SQL is more relevant language for Data Professions.

Database Trends

Current Year



Next Year



Database Trends – Findings & Implications

Findings:

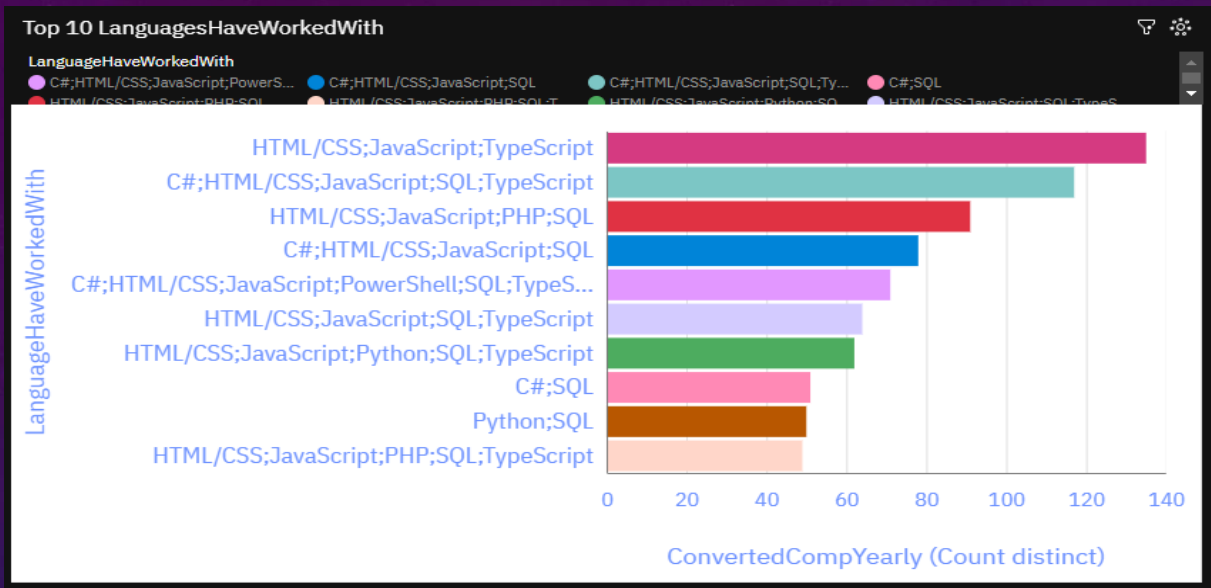
- PostgreSQL, a leading object-relational database, offers advanced capabilities.
- Microsoft SQL Server, renowned for its robustness, is currently second in demand but will likely fall to fourth place next year.

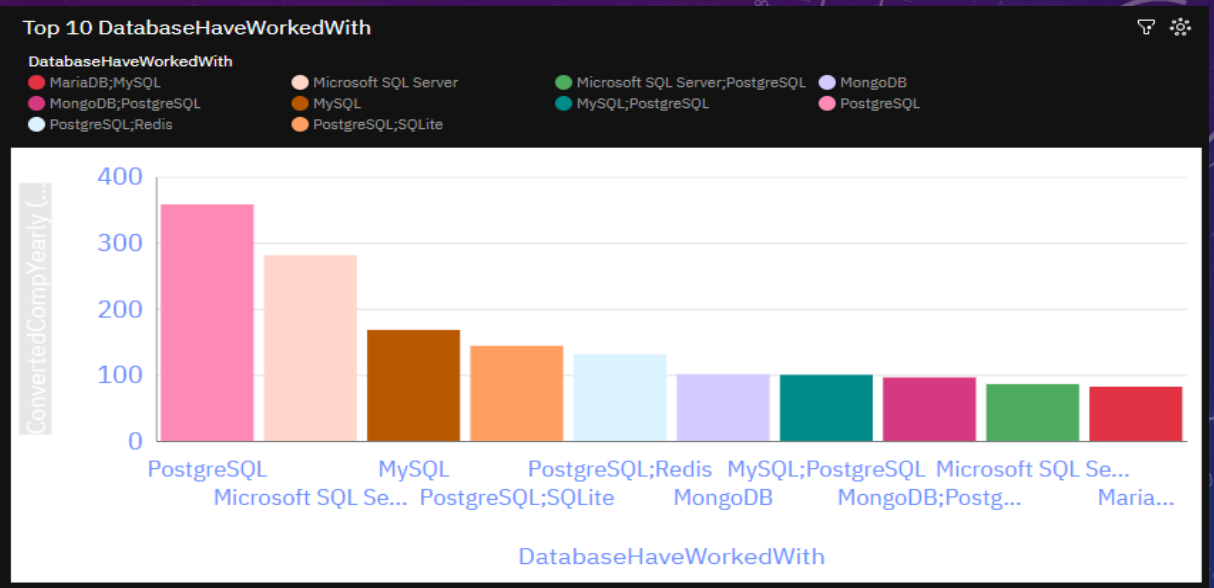
Implications:

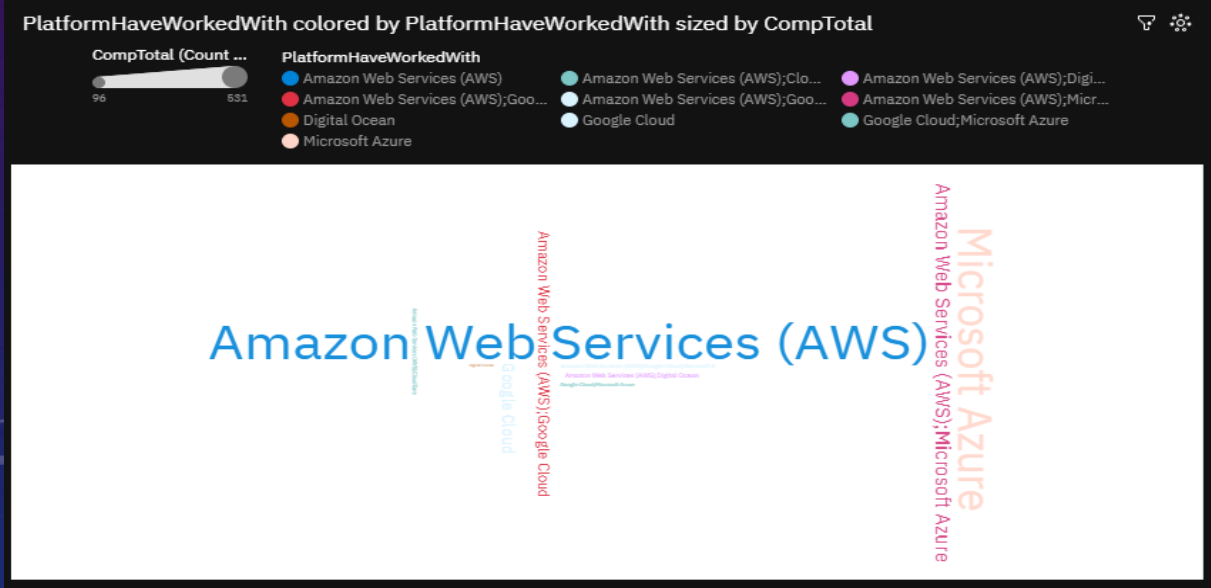
- PostgreSQL for the upcoming year emerges as the leading database solution, offering advanced capabilities and an object-oriented design that effectively addresses complex data management requirements, remaining leading and potentially picking up users to the decline of Microsoft SQL and My SQL.
- Microsoft SQL Server, currently second in demand, anticipates a drop off of users, signaling the need for organizations to reassess their database strategies and consider alternative solutions to maintain competitiveness in a changing landscape.

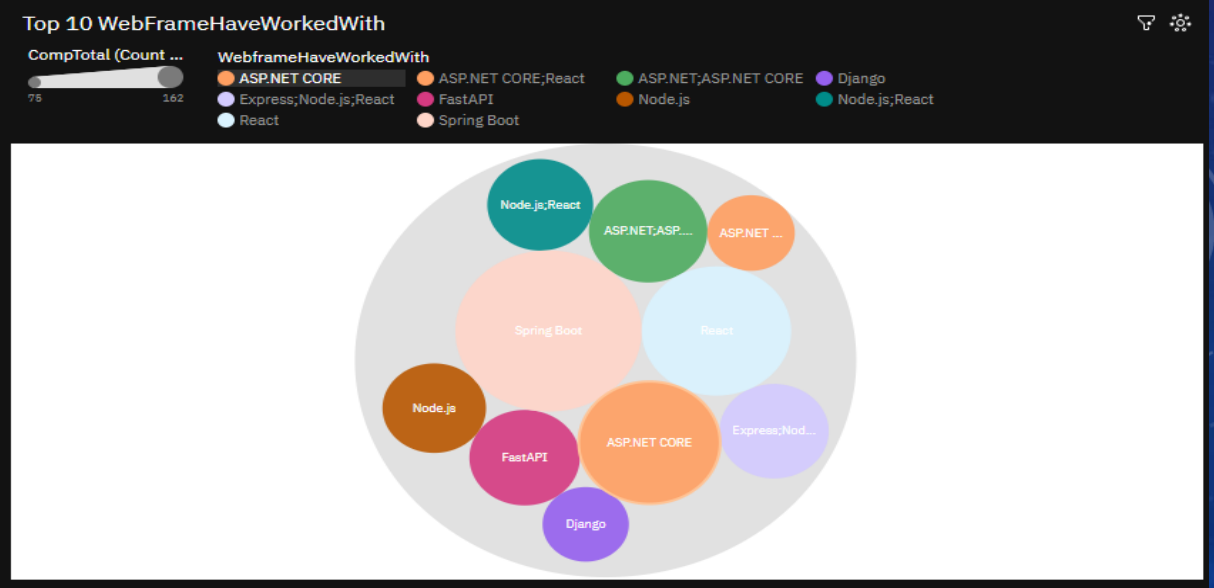
DASHBOARD

Dashboard Tab 1





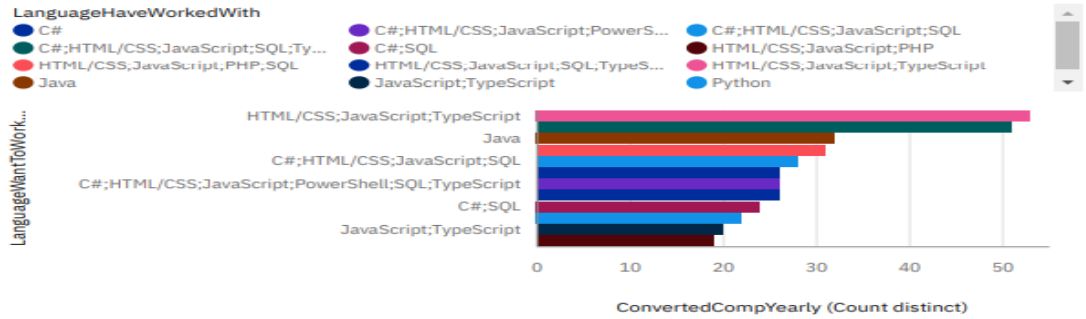




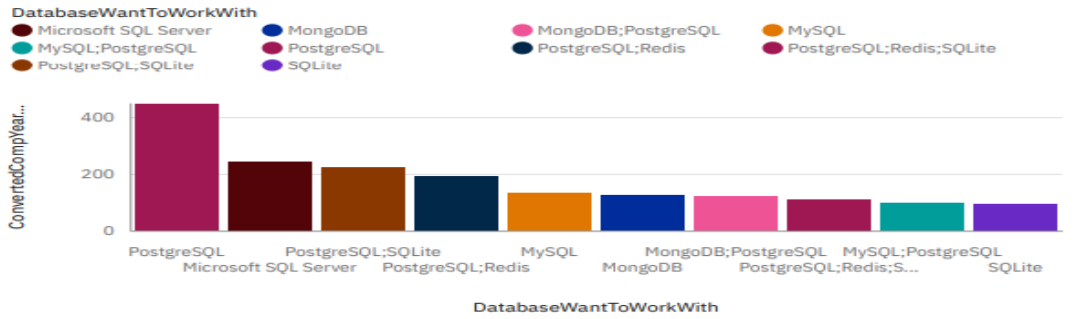
Dashboard Tab 2

Future Technology Trend

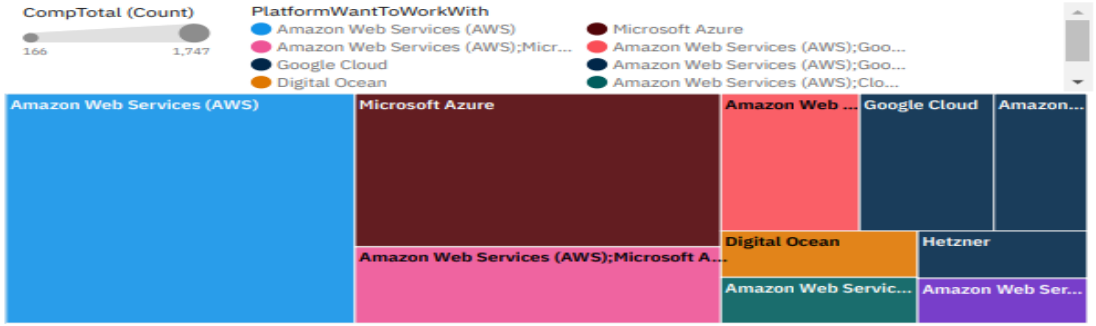
Top 10 LanguageWantToWorkWith



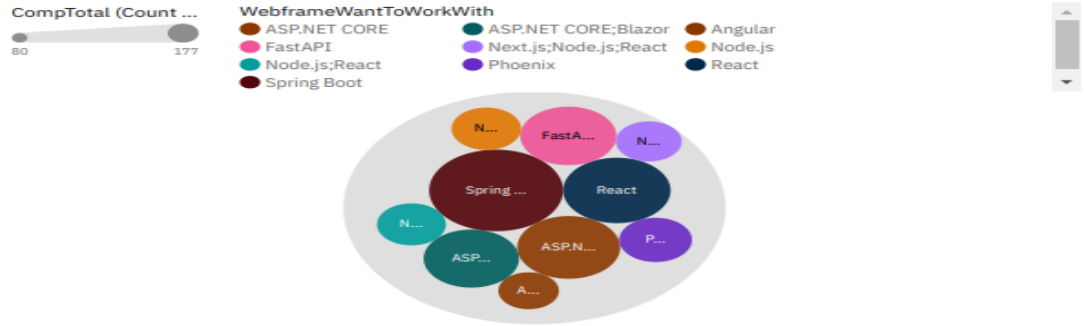
Top 10 DatabaseWantToWorkWith



Top 10 PlatformWantToWorkWith



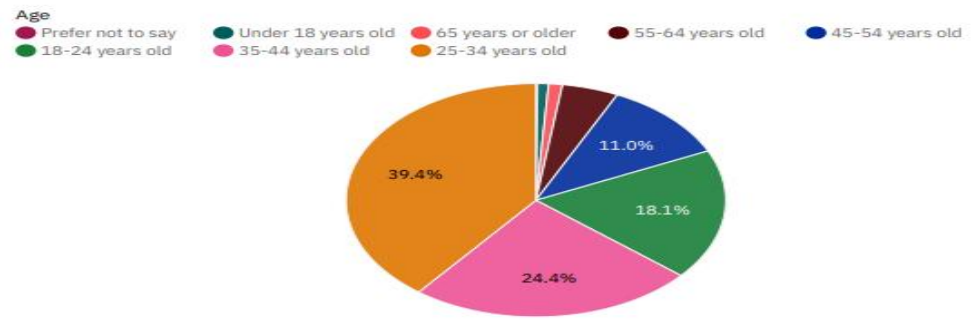
Top 10 WebframeWantToWorkWith



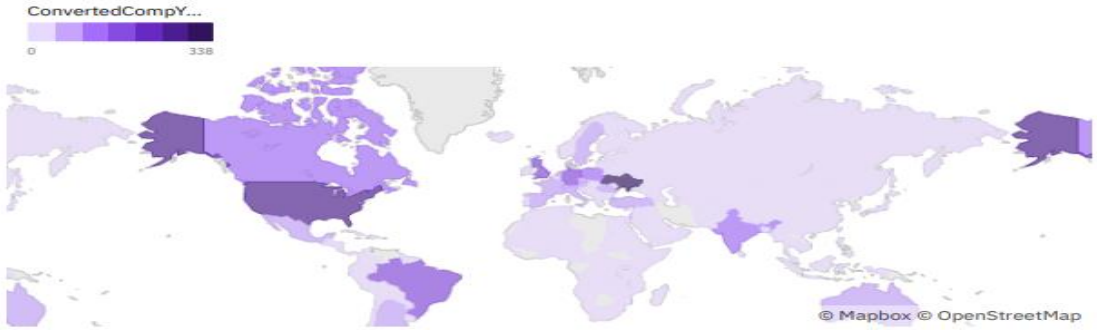
Dashboard Tab 3

Demographics

Respondent distribution by Age

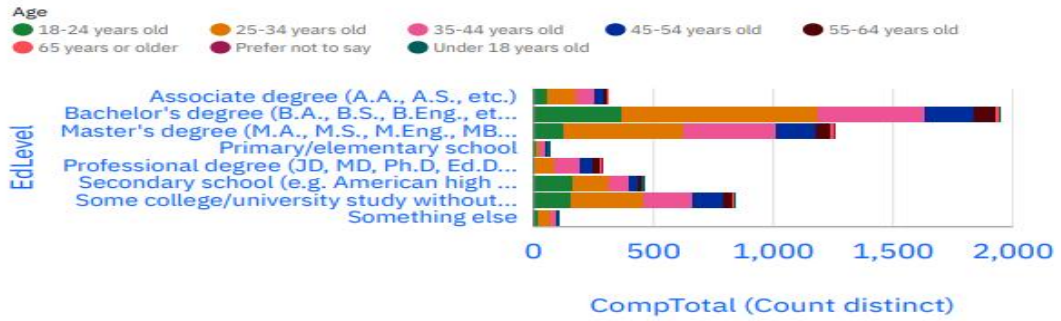


Capture Respondent Count by Country



Capture Respondent distribution by Formal Education Level

Capture Respondent Count by Age, classified by Education Level



Insights from Dashboards

- **Programming Languages**: Strong dominance of Python and JavaScript across demographics.
- **Databases**: PostgreSQL remains the top choice, with increasing interest in cloud-native databases.
- **Demographics**: Younger developers favor emerging technologies, while experienced developers prefer stability.

Overall Findings & Implications

Findings:

- Web development languages emerge as the predominant and sought-after tools within the IT sector presently, highlighting their critical role in driving technological advancements.
- The IT landscape primarily comprises young professionals below the age of 40, suggesting a youthful workforce driving innovation and progress within the industry.
- The desire to acquire skills in Postgre SQL and React JS among most respondents underscores the importance of staying abreast of emerging technologies to remain competitive in the field.

Implications:

- The continued importance of web development as a lucrative skill underscores the significance of investing in and advancing expertise in this domain.
- Outside of well developed countries, emerging developing countries like Brazil and India are growing in tech training and educations; Less developed countries need more access to tech trainings and educations.

Conclusion

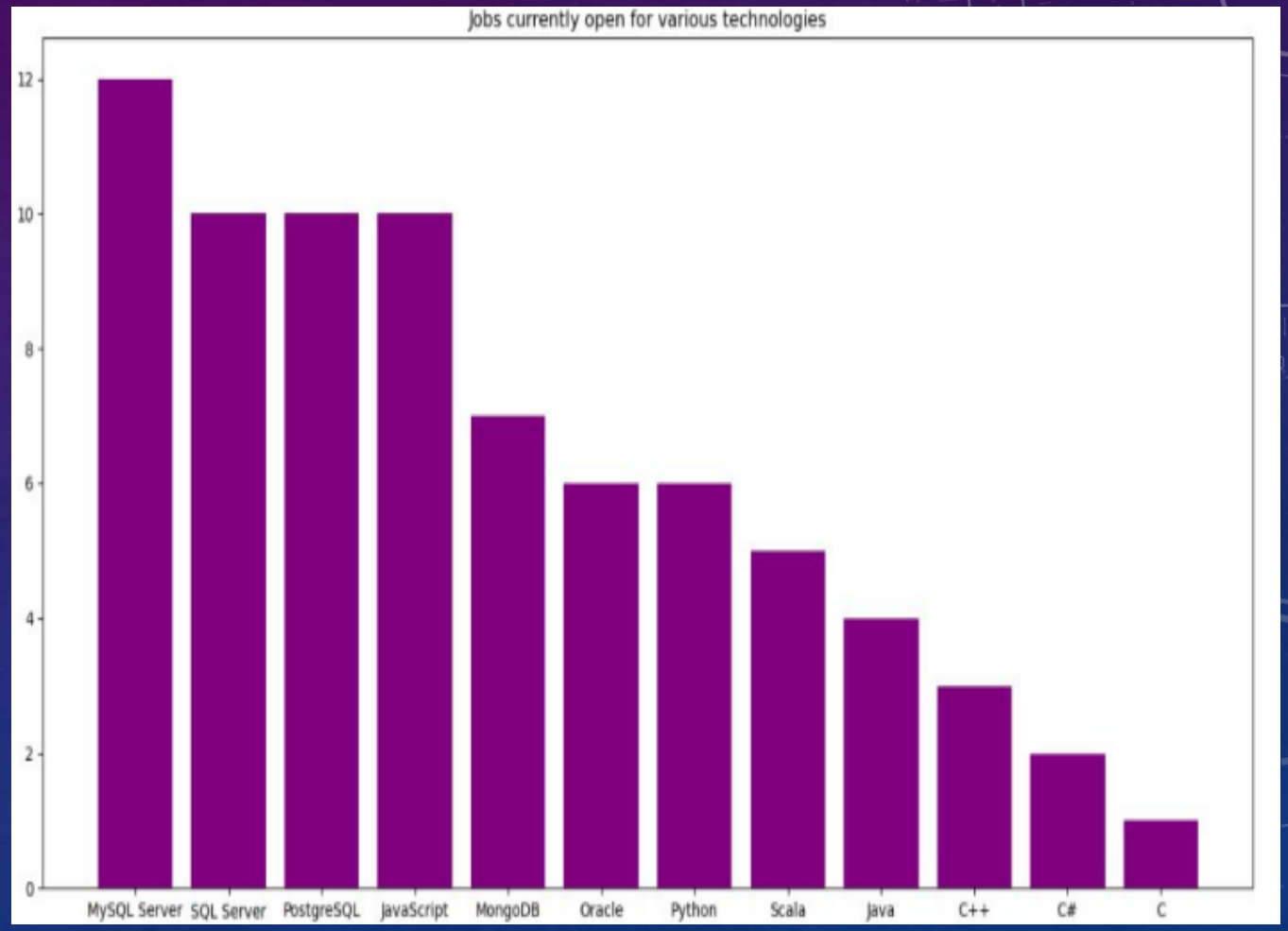
- **Programming Language Trend**: The findings reveal that JavaScript, HTML/CSS, PowerShell, Java, and C# emerge as the top five programming languages in demand for the current year, with Go replacing Java's position. This underscores the industry's robust demand for both front-end and back-end development expertise, as well as proficient data analysis and management skills.
- **Database Trend**:
- **Platform Trend**: Google Cloud Platform, AWS, IBM Cloud, Windows, and Microsoft Azure as the top platforms for the current year, while Docker, AWS, Android, Linux, Google Cloud Platform, and Kubernetes are forecasted as high-demand platforms for the upcoming year. This underscores developers' versatility and adaptability in navigating diverse platforms and technologies.
- **Web Framework Trend**: Results indicate that React.js, Angular, JQuery, ASP.NET, and Express remain the top five desired web frameworks for both the current and upcoming year. This highlights the increasing importance of front-end development skills and underscores the necessity for developers to excel in constructing dynamic and user-friendly websites.
- **Demographic Trend**: North America emerges with the highest number of respondents with a significant proportion holding a degree, following India and Brazil coming in second and third in respondents.

Appendix

- Resources:
  - Data was scraped from a Stack Overflow survey under an OBDl: Open Database License
  - IBM Cognos was used for creating the Dashboards and various visualizations
  - BeatifulSoup library in Python was used to scrape GitHub job postings along with their average salaries
- Acknowledgements:
  - Coursera for allowing access to this course
  - Stack Overflow for making survey data available for use in analysis

GITHUB JOB POSTINGS

Job posting data meticulously gathered through the utilization of the GitHub Job API.



Popular Languages

Prominent programming languages alongside their corresponding average annual salaries. This data set is the result of a comprehensive web scraping endeavor, meticulously extracting insights from GitHub job postings, subsequently saved in a CSV format for analysis