Skills Network

IBM Data Analyst Capstone Project:

Stackoverflow Survey Analysis

BY JUNG-HYUN RYU

2025-01-01

OUTLINE

1. EXECUTIVE SUMMARY
2. INTRODUCTION
3. METHODODOLOGY
4. RESULTS
5. VISUALIZATION – LANGUAGE & DATABASE TRENDS
6. DASHBOARD
7. DISCUSSION & STRATEGIC IMPLICATIONS
8. CONCLUSION
9. APPENDIX
10. EXTRA CHARTS – JOB POSTING & POPULAR LANGUAGE

EXECUTIVE SUMMARY

Market Analysis Overview

- ### Key Technology Trends
  - JavaScript maintains market leadership while Python shows strongest growth potential
  - Traditional databases (MySQL, Oracle) declining as PostgreSQL and MongoDB gain momentum
  - Strong shift toward cloud platforms and containerization technologies
- ### Critical Metrics
  - Programming Languages: JavaScript leads with highest current usage and future demand
  - Databases: PostgreSQL shows 47% increase in future interest
  - Platforms: Cloud services (AWS, Azure) demonstrate significant growth trajectory
  - Demographics: 73.3% male vs 26.7% female representation in the industry
- ### Strategic Recommendations
  - Invest in cloud-native development capabilities
  - Prioritize full-stack development training, especially JavaScript ecosystem
  - Transition toward modern database solutions (PostgreSQL, MongoDB)
  - Implement targeted initiatives to improve gender diversity
  - Enhance continuous learning programs to address emerging technology demands
- ### Business Impact
  - These findings suggest a clear direction for technology investment and talent development strategies, emphasizing the need for modernization of technology stacks and diversity initiatives to maintain competitive advantage in the rapidly evolving IT landscape.

INTRODUCTION TO CAPSTONE PROJECT

- Background scenario
  - I am hired as a Data Analyst by a global leader in IT and business consulting, committed to staying at the forefront of technological advancements. As part of our annual initiative to identify future skill requirements, we have conducted a comprehensive analysis of emerging skills in the IT industry.
  - Business benefits: This data-driven approach will help us remain competitive and adapt our skillset to meet evolving industry needs.

- Key Focus Areas - Our analysis aims to identify trends in:
  - Top programming languages
  - In-demand database skills
  - Popular Integrated Development Environments (IDEs)
  - Other interesting information in IT

METHODOLOGY – Data Collection & Wrangling

- Data is from multiple sources to ensure a holistic view of the current IT landscape:
  1. Job Postings: Analyzed using Jobs API to extract in-demand skills
  2. Web Scraping: Utilized Python libraries (BeautifulSoup, requests) to collect data from relevant websites, including www.ibm.com
  3. Developer Survey by Stack Overflow: Examined a large-scale dataset from IBM Cloud to understand current trends

- Data link: 1. Job posting , 2. Developer Survey
- Github links: Lab1, Lab2, Lab3, Lab4

RESULTS – Exploratory Data Analysis Key Insights

Github link: Lab5

- Dataset Overview:
  - Large-scale developer survey with 85 columns of data
  - Diverse information including demographics, education, employment, and technical preferences

- Key Variables:
  - Demographics: Age, Gender, Country, Ethnicity
  - Education: Education Level, Undergraduate Major
  - Employment: Employment Status, Developer Role
  - Technical: Programming Languages, Open Source Contribution, IDE Preferences

- Initial Observations:
  - Respondents from various countries (e.g., United States, New Zealand, United Kingdom, Australia)
  - Predominantly male respondents in the sample shown
  - Most Common Age range of respondents in the sample: 22-29 years old
  - Most respondents have at least a Bachelor's degree, many with Computer Science backgrounds
  - Full-time employment is common among the respondents

VISUALIZATION: CHARTS

GITHUB LINK

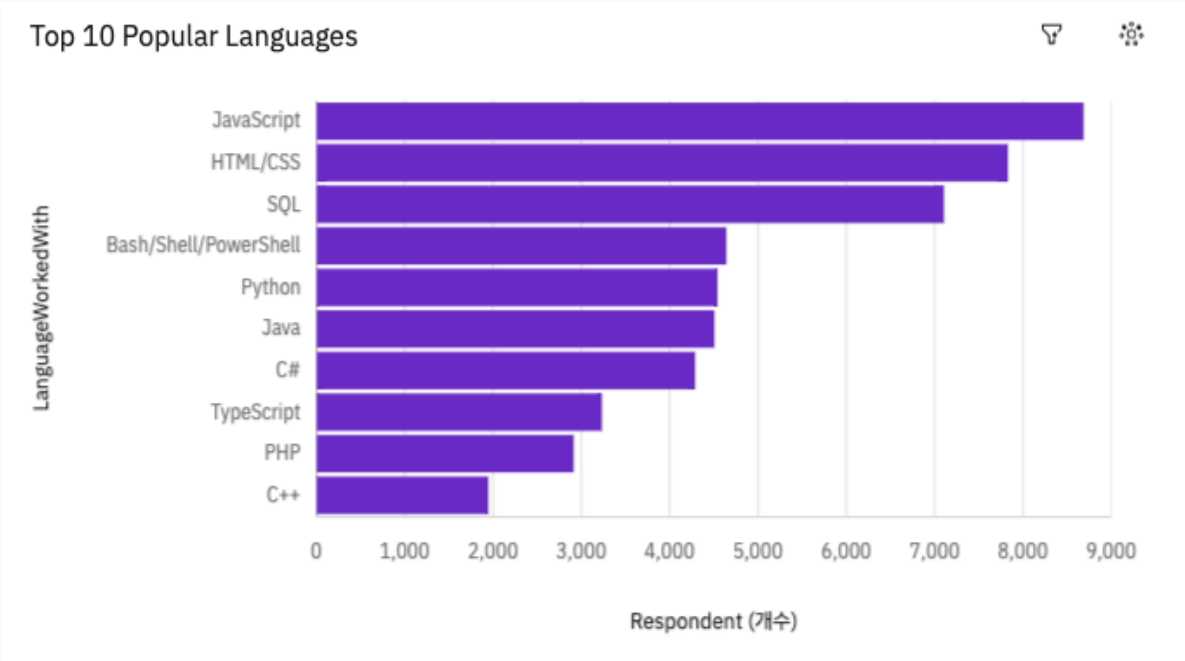
IBM COGNOS LINK

PROGRAMMING LANGUAGE TRENDS

Github link: Dashboards

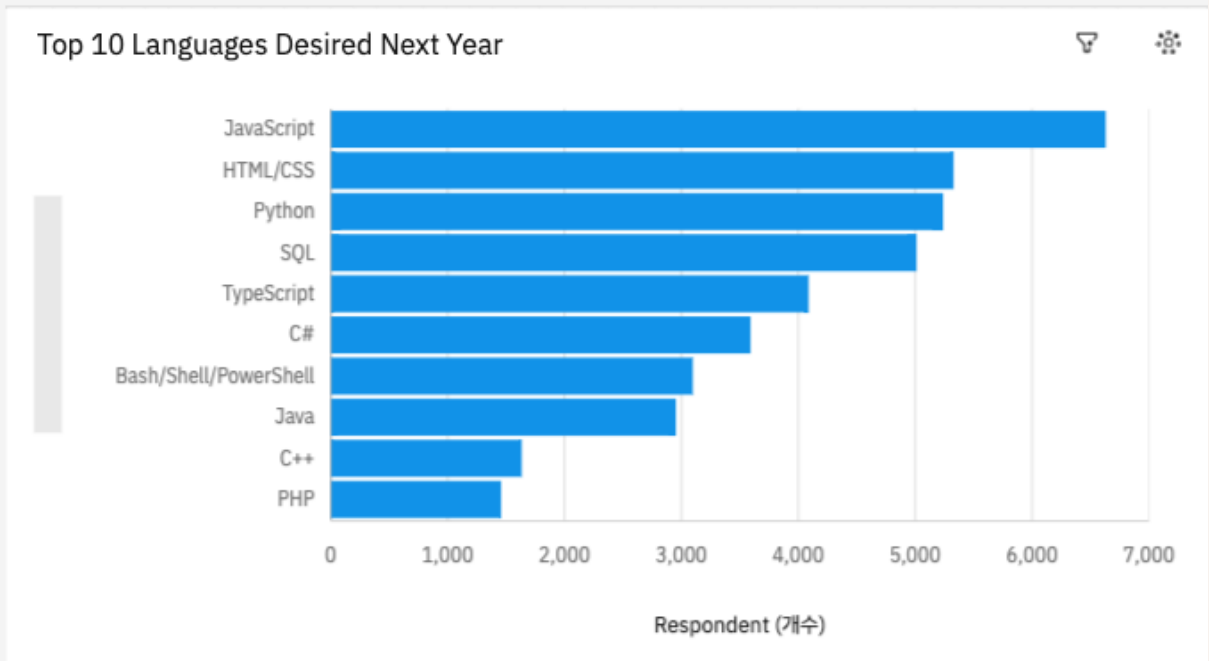
CURRENT YEAR

Current Technology Usage | Future Technology Usage | +



NEXT YEAR

Current Technology Usage | Future Technology Usage | +



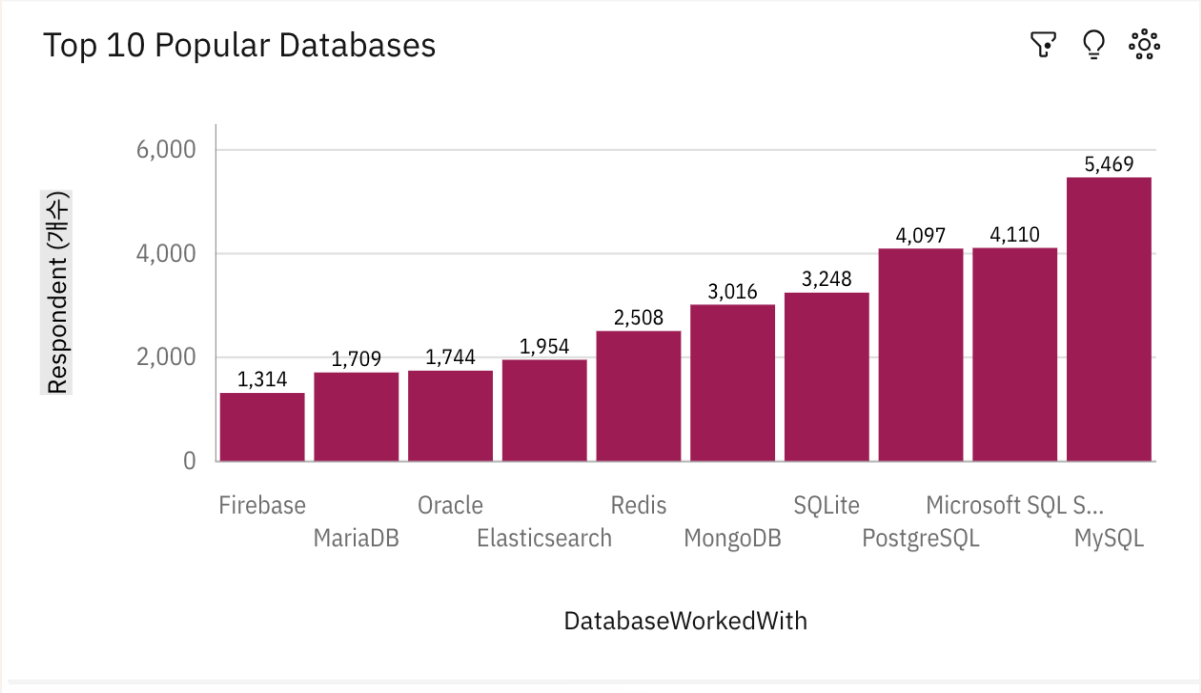
LANGUAGE TRENDS – FINDINGS & IMPLICATIONS

1. JavaScript's Continued Dominance: JavaScript maintains its top position in both current and desired languages, indicating its ongoing importance in web development and its adaptability to new trends.
2. Rising Popularity of Python: Python's jump from 5th place in current usage to 3rd in desired languages suggests a growing interest in data science, machine learning, and backend development. This trend may lead to increased demand for Python-related skills and tools.
3. Shift Towards Statically-Typed Languages: The rise of TypeScript from 8th to 5th place in desired languages implies a growing preference for statically-typed languages in large-scale applications, potentially due to improved maintainability and developer productivity.
4. Declining Interest in Some Established Languages: C++ and PHP, while currently popular, have lower popularity in the top 10 desired languages. This suggests a potential shift away from these languages, which could impact legacy systems and traditional web development practices.
5. Continued Importance of Web Technologies: HTML/CSS and SQL remain consistently popular, highlighting the ongoing significance of web development and database management skills in the industry.

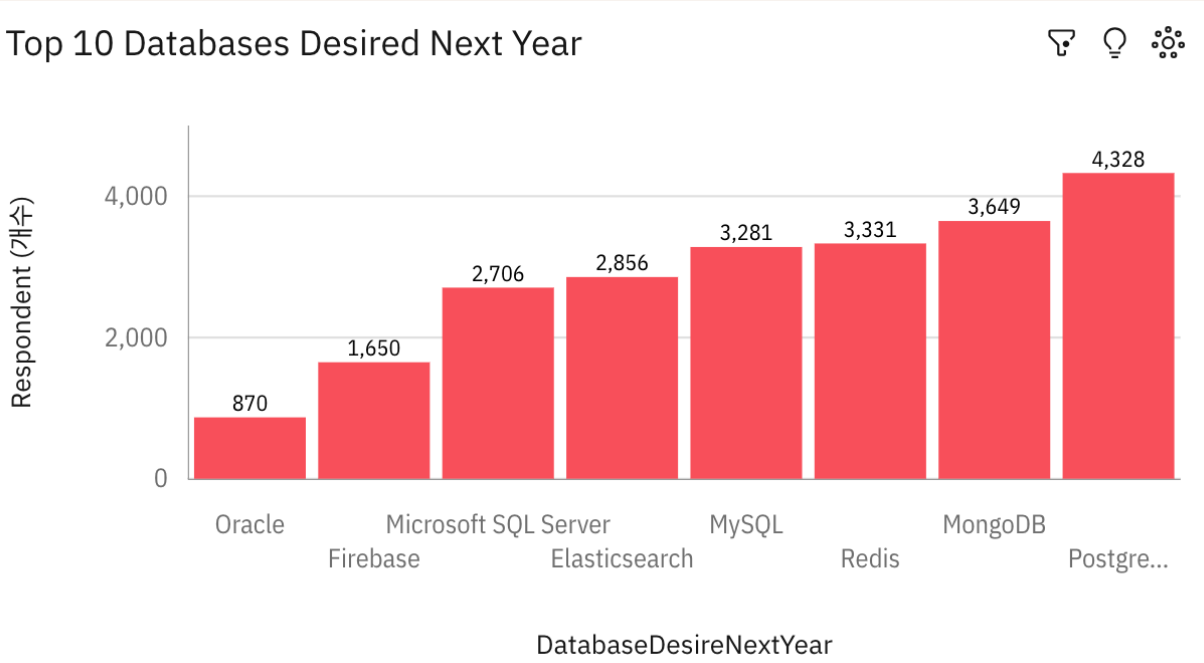
DATABASE TRENDS

Github link: Dashboards

CURRENT YEAR



NEXT YEAR



DATABASE TRENDS – FINDINGS & IMPLICATIONS

1. Shift in Traditional Databases

- PostgreSQL's rise to the top position indicates a growing preference for open-source, feature-rich relational databases
- MySQL's significant drop from 1st to 4th suggests a potential shift away from traditional market leaders
- Oracle's disappearance from the top 10 implies decreasing interest in proprietary database solutions

2. NoSQL Growth

- MongoDB's climb from 5th to 2nd place reflects the increasing importance of document-based NoSQL databases
- Redis's upward movement shows growing demand for in-memory database solutions
- The emergence of DynamoDB in the future top 10 indicates rising interest in cloud-native database services

3. Industry Trends

- The overall pattern suggests a move toward cloud-friendly and scalable database solutions (e.g., AWS, Docker, Kubernetis)
- Open-source databases are gaining more traction over proprietary solutions
- There's a clear trend toward databases that support microservices and distributed architectures

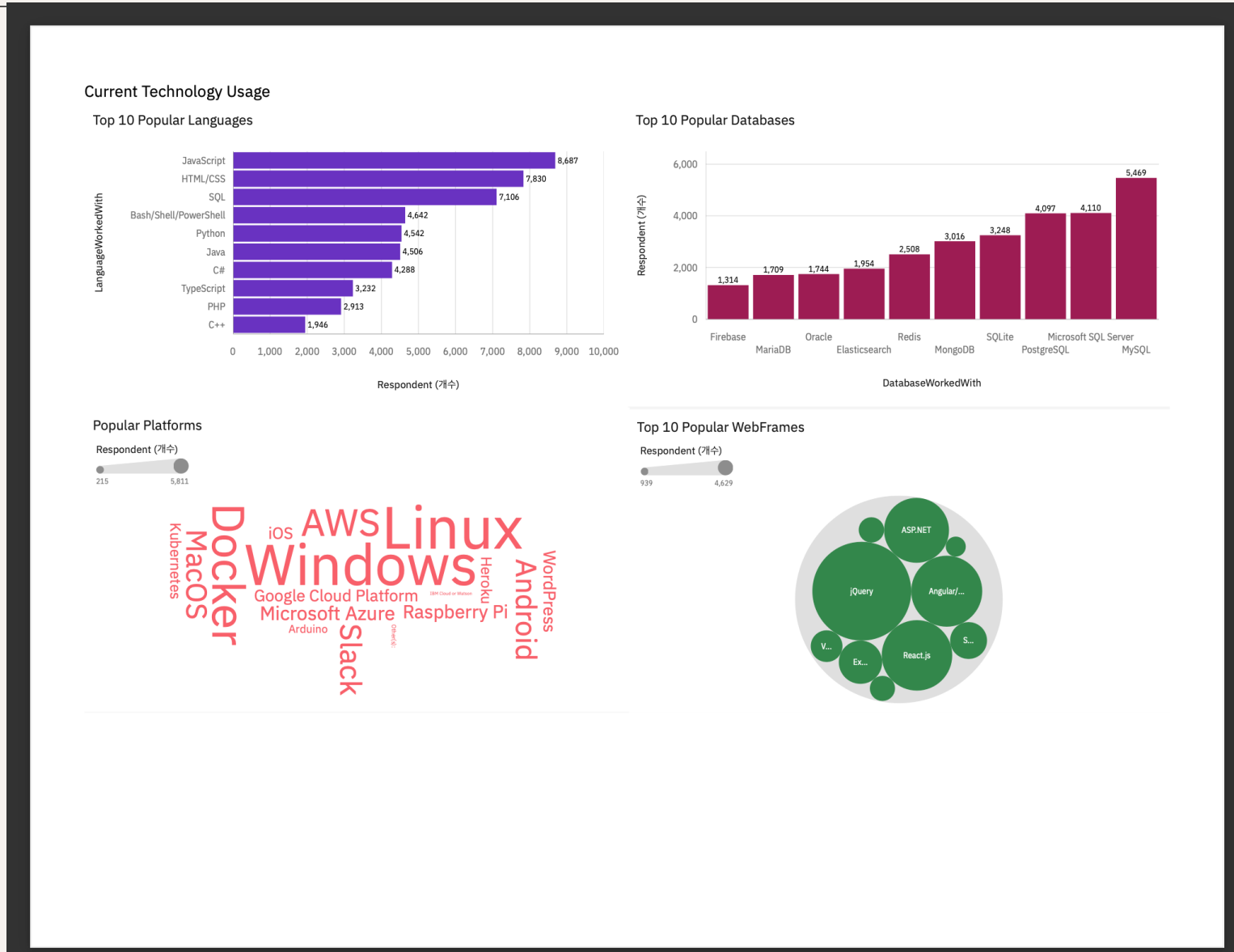
DASHBOARDS

GITHUB LINK

IBM COGNOS LINK

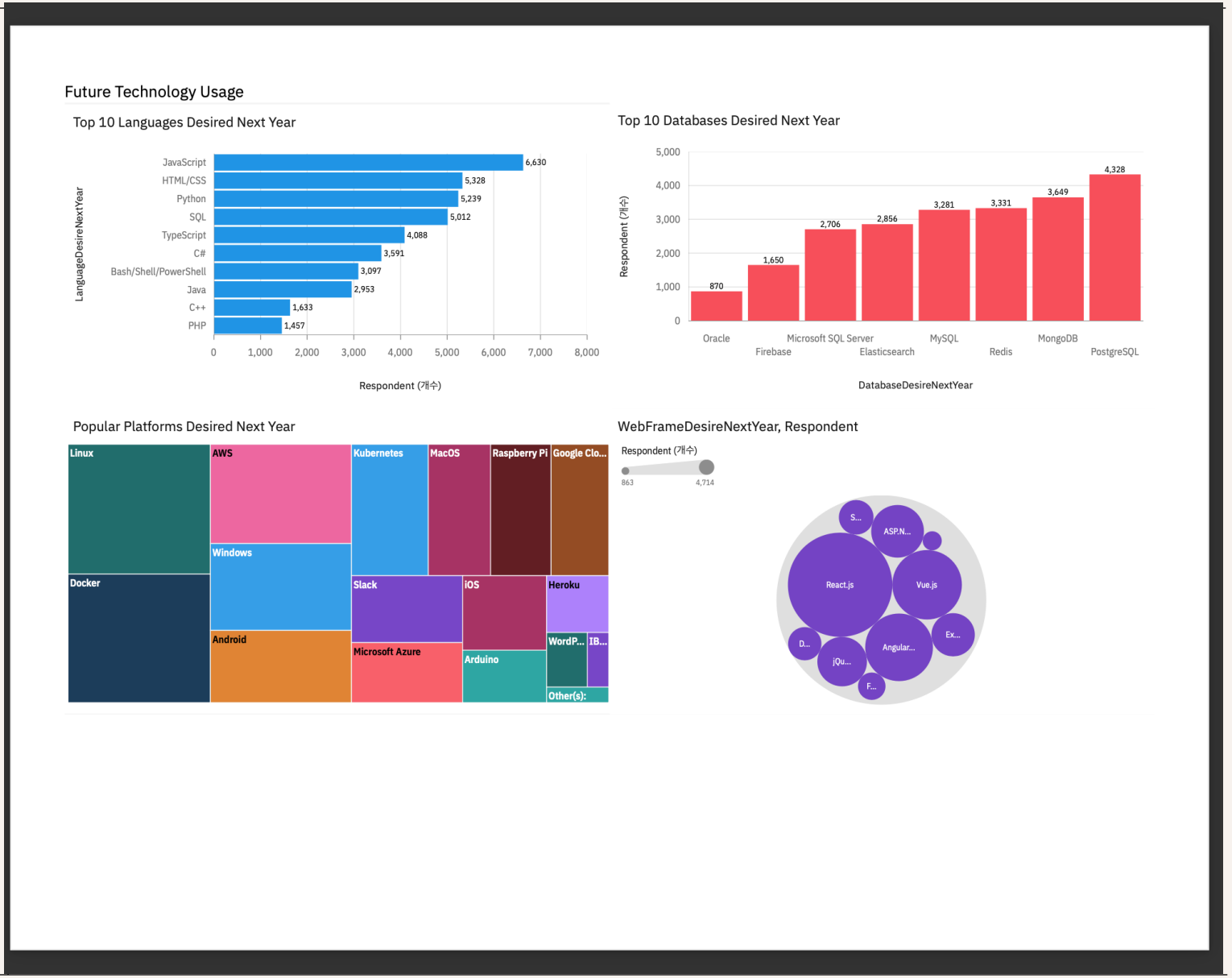
DASHBOARD TAB 1:

CURRENT TECH TREND



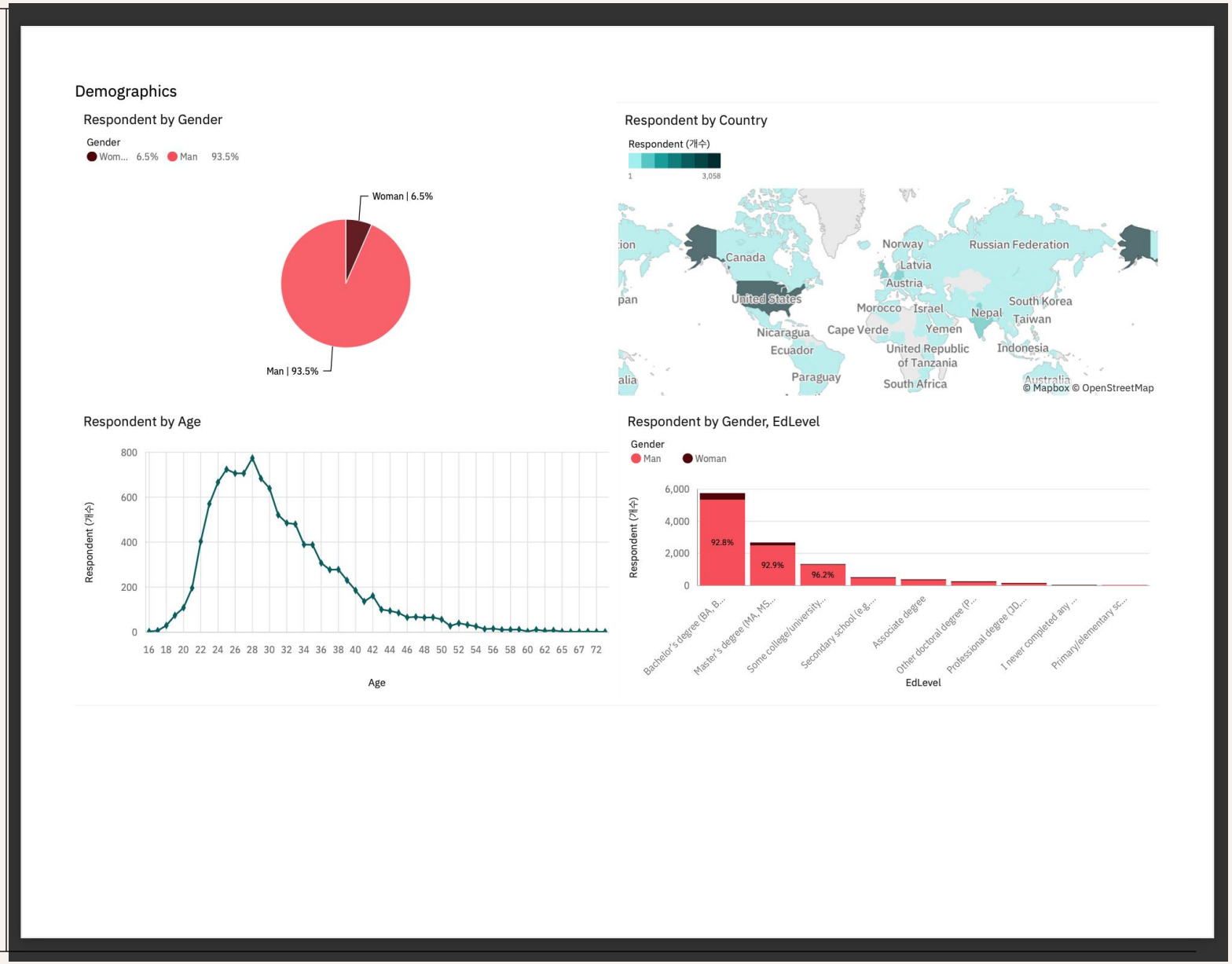
DASHBOARD TAB 2:

FUTURE TECH TREND



DASHBOARD TAB 3:

DEMOGRAPHIC TREND



DISCUSSION (1)

- Demographics and Gender Distribution - Male-Dominated Industry Reality
  - The survey shows a significant gender disparity with 73.3% male and 26.7% female representation, indicating a continuing need for initiatives to increase gender diversity in tech.
- Educational Background Patterns - Bachelor's Degree Dominance
  - The majority of respondents hold bachelor's degrees, with notably higher representation among both men and women, suggesting this remains the standard educational requirement in the tech industry.
- Geographic Distribution - US-Centric Technology Hub
  - The darker shading on the US in the map indicates a higher concentration of respondents, suggesting continued dominance of the US in the global tech landscape.

DISCUSSION (2)

- Platform Evolution - Cloud Platform Acceleration

- The shift from current to desired platforms shows increased interest in cloud services (AWS, Azure) and containers(Docker, Kubernetes), indicating a strong move toward cloud-native development.

- Web Framework Transitions - Modern Framework Adoption

- The bubble charts show React.js and Angular maintaining strong positions, with Vue.js gaining significant interest for the future, suggesting a trend toward modern, component-based frameworks.

- Development Environment Changes - DevOps Integration

- The increased presence of tools like Docker and Kubernetes in desired platforms indicates a stronger emphasis on DevOps practices and containerized development environments.These trends suggest a technology landscape that's becoming more cloud-centric, with stronger emphasis on modern web frameworks and containerized deployments, while still grappling with demographic imbalances in the workforce.

CONCLUSION (1)

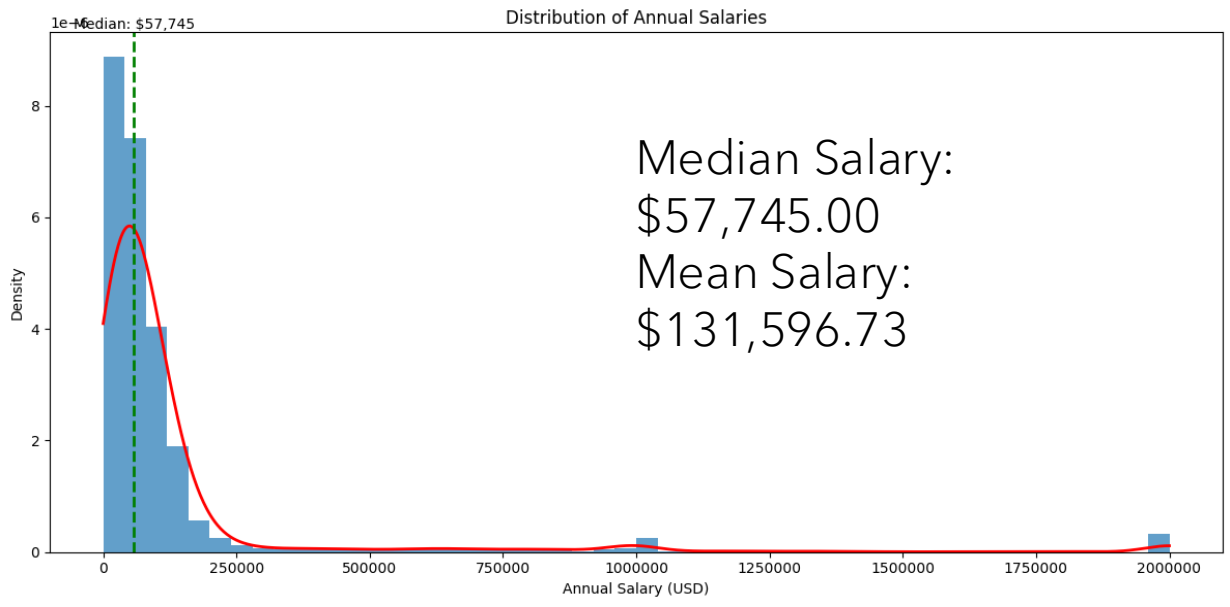
- Technology Stack Evolution
  - JavaScript and HTML/CSS maintain dominance in programming languages, reflecting the continued importance of web development
  - MySQL leads current database usage (5,469 users) but shows declining future interest, suggesting a shift in database preferences
  - Cloud platforms (AWS, Azure, Google Cloud) and containerization tools (Docker) show strong presence, indicating industry direction
- Framework and Platform Trends
  - jQuery, Angular, and React.js dominate the web framework landscape
  - Linux and Windows remain the primary operating platforms, with growing cloud platform adoption
  - Increasing interest in containerization and cloud-native technologies suggests a shift toward modern development practices
- Demographic Insights
  - Significant gender disparity exists with 73.3% male and 26.7% female representation
  - Most respondents are between ages 25-35, indicating a relatively young workforce
  - Bachelor's degree holders form the largest educational demographic across both genders
  - Strong geographic concentration in the United States suggests continued North American tech industry dominance

CONCLUSION (2) – Strategic Implications

- Organizations need to prioritize:
  - Full-stack development capabilities, especially JavaScript ecosystem
  - Cloud and containerization expertise
  - Open-source database solutions
  - Diversity and inclusion initiatives to address gender imbalance
  - Continuous learning programs to keep pace with evolving technology landscape
- Future Outlook:
  - Transition toward cloud-native and containerized development environments
  - Growing importance of full-stack development skills
  - Increasing emphasis on modern web frameworks and tools
  - Need for stronger initiatives to improve gender diversity in tech
  - Continued importance of formal education in technology careers

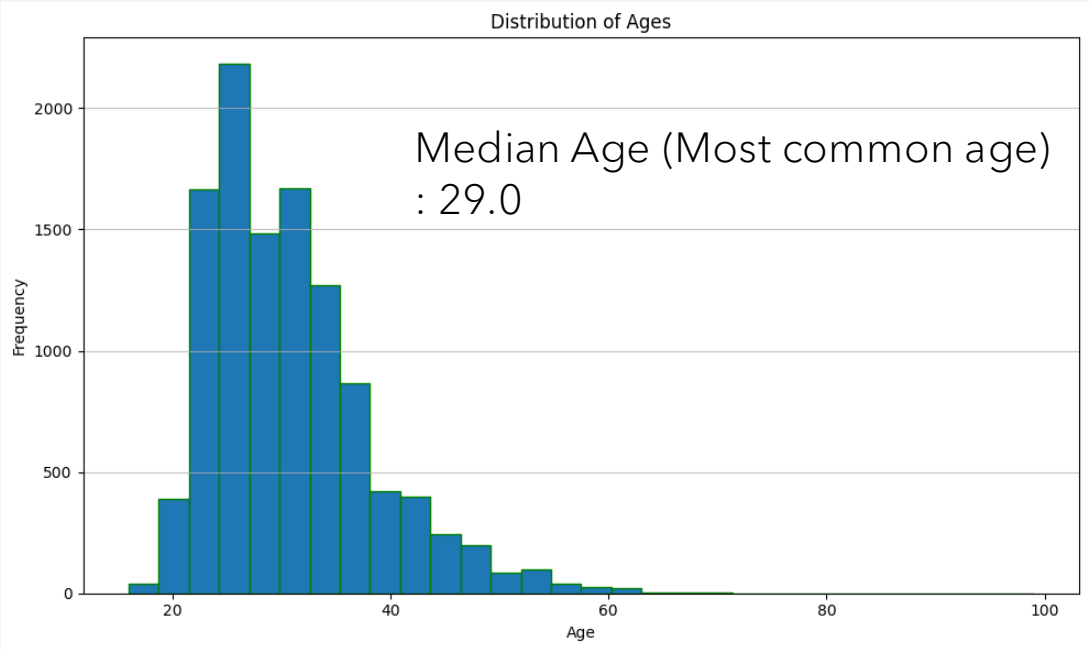
APPENDIX – any relevant additional charts, or tables created during the analysis phase.

DISTRIBUTION OF SALARY



Median Salary: $57,745.00
Mean Salary: $131,596.73

DISTRIBUTION OF AGE



Median Age (Most common age) : 29.0

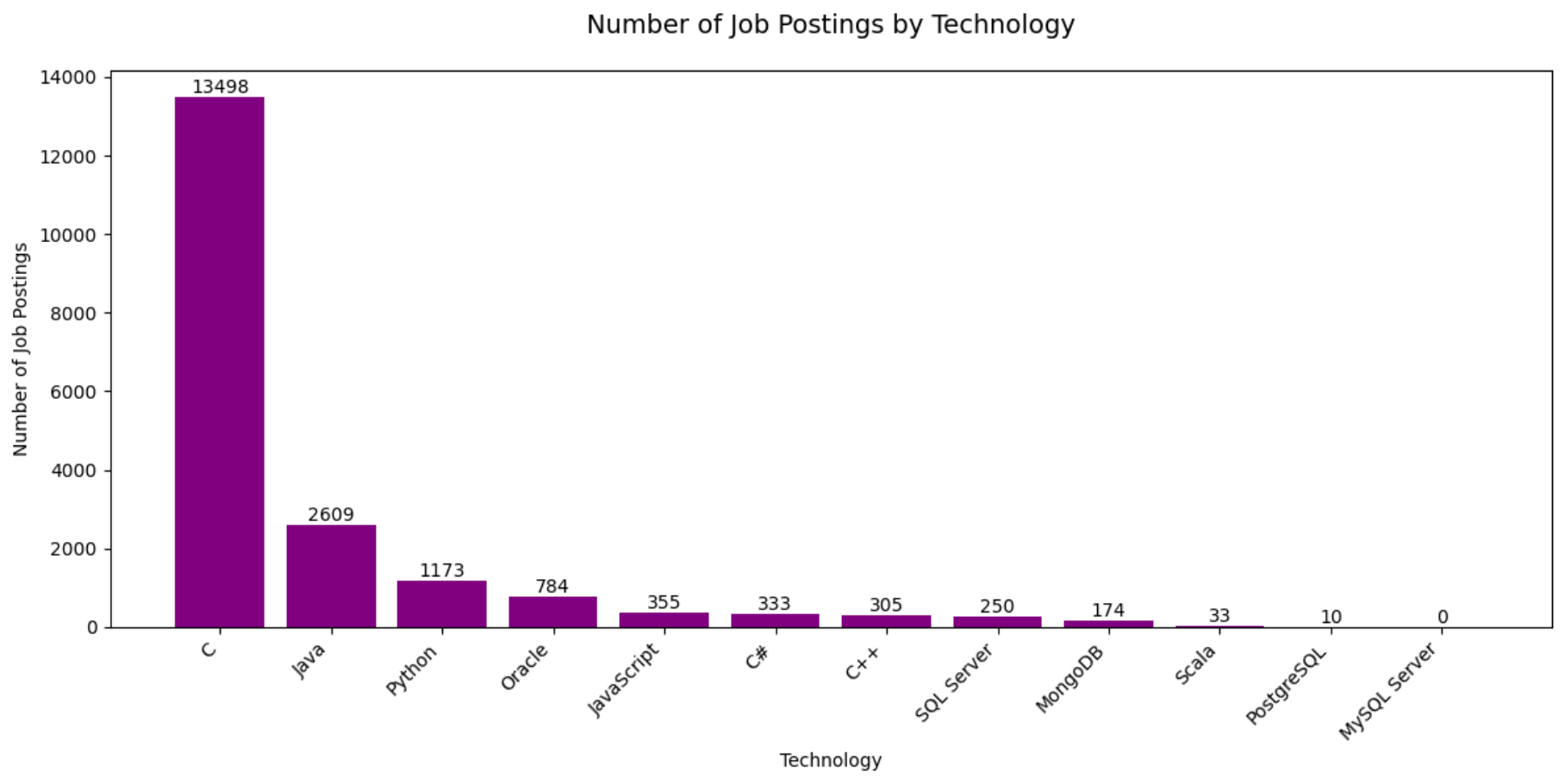
THANK YOU!

From Jung Hyun RYU

SPECIAL THANKS TO IBM TEAM

JOB POSTINGS

In Module 1 you have collected the job posting data using Job API in a file named “job-postings.xlsx”. Present that data using a bar chart here. Order the bar chart in the descending order of the number of job postings.



POPULAR LANGUAGES

In Module 1 you have collected the job postings data using web scraping in a file named “popular-languages.csv”. Present that data using a bar chart here. Order the bar chart in the descending order of salary.