IBM Data Analyst Capstone Project

Project Scenario

You have recently been hired as a Data Analyst by a global IT and business consulting services firm that is known for their expertise in IT solutions and their team of highly experienced IT consultants. In order to keep pace with changing technologies and remain competitive, your organization regularly analyzes data to help identify future skill requirements. As a Data Analyst, you will be assisting with this initiative and have been tasked with collecting data from various sources and identifying trends for this year's report on emerging skills. Your first task is to collect the top programming skills that are most in demand from various sources including job postings, training portals, and surveys. Once you have collected enough data, you will begin analyzing the data and identify insights and trends that may include the following:

 * What are the top programming languages in demand?
 * What are the top database skills in demand?
 * What are the popular IDEs?

You will begin by scraping internet web sites and accessing APIs to collect data in various formats like .csv files, excel sheets, and databases. Once this is completed, you will make that data ready for analysis using data wrangling techniques. When the data is ready you will then want to apply statistical techniques to analyze the data. Then bring all of your information together by using IBM Cognos Analytics to create your dashboard. And finally, show off your storytelling skills by sharing your findings in a presentation.

Technology Trends in Software Development

Prepared by: LJMcNeill

April 18, 2024

IBM Developer

SKILLS NETWORK

OUTLINE

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

IBM **Dev**_e_loper

SKILLS NETWORK

EXECUTIVE SUMMARY

Introduction:

Data sources were gathered and analyzed to identify trends in the use of certain languages, databases, and development environments in the software development industry as well as to consider the implications of the current demographics of the developer community.

The data analyzed included the 2019 Stack Overflow Developer Survey and data scraped from the GitHub job announcements. Analysis uncovered some notable technology trends and developer community demographic insights.

These insights provide value in informing strategic planning for quality of service readiness and maintaining a competitive edge in our global IT and business consulting services.

EXECUTIVE SUMMARY

Skills in Demand:

Some trends were found regarding demand of certain languages and databases.

- Old standards remaining in strong demand include:
  - **JavaScript, Python**, and **C** programming languages
  - **PostgreSQL** and **MongoDB** databases

- Rising stars and emerging contenders include:
  - **Go** and **Kotlin** programming languages
  - **DynamoDB, Elasticsearch,** and **Redis** databases.

IBM **Developer**

SKILLS NETWORK

EXECUTIVE SUMMARY

Demographic Insights:

Gender disparity is the primary demographic insight gleaned from the analysis.

- 14:1 Ratio Men to Women among developer survey respondents
- Average annual salary for Men is 32% higher than average annual salary for women

IBM **Dev**oper

SKILLS NETWORK

EXECUTIVE SUMMARY

Business Value: How this analysis can help business planning.

Technology Trends Analysis

- Conduct company-wide skills gap analysis to identify needed skills based on demand as identifies in the technology tend analysis.

- Higher new or upskill existing employees to close skill gaps.

- Plan to meet growing demand for emerging languages and databases poised for rapid demand growth in coming years.

IBM **Dev** loper

SKILLS NETWORK

EXECUTIVE SUMMARY

Business Value (cont'd)

Gender Gap

- Working toward gender equity is good for business. An International Labor Organization study cites the following benefits:
    - 63% increase profitability/productivity
    - 60% improved ability to attract/retain talent
    - 59% greater creativity innovation and openness
    - 58% better company reputation
    - 38% enhanced ability to gauge consumer interest/demand
- Actions to take:
    - Evaluate company compensation policies and practices to ensure no gender disparity.
    - Focus training and recruiting efforts to achieve more balanced workforce.
    - Foster company culture that embraces gender equality.

IBM Developer

SKILLS NETWORK

INTRODUCTION

GOALS:

1. To gain insights regarding Current Technology Usage and Future Technology Trends for:

- Languages
- Databases
- Platforms
- WebFrames

- 2. Understand the Demographics of the developer community:

- Gender
- Nationality
- Age
- Salary

IBM **Developer**

SKILLS NETWORK

METHODOLOGY

Data Sources:

- Stack Overflow Developer Survey 2019
- Job data gathered from GitHub Jobs API using Python Requests Library.
- Annual salary by programming language data scrapped from internet sources using Scrapy and BeautifulSoup Python libraries.

Data Cleaning: Cleaned and normalized data using Python.

Analysis: Statistically and computationally analyzed data with Python using Pandas, NumPy, Plotly, Matplotlib, Seaborn, and SciPy.

Visualizations: Created visualizations with IBM Cognos Analytics, Excel, Seaborn, and Matplotlib.
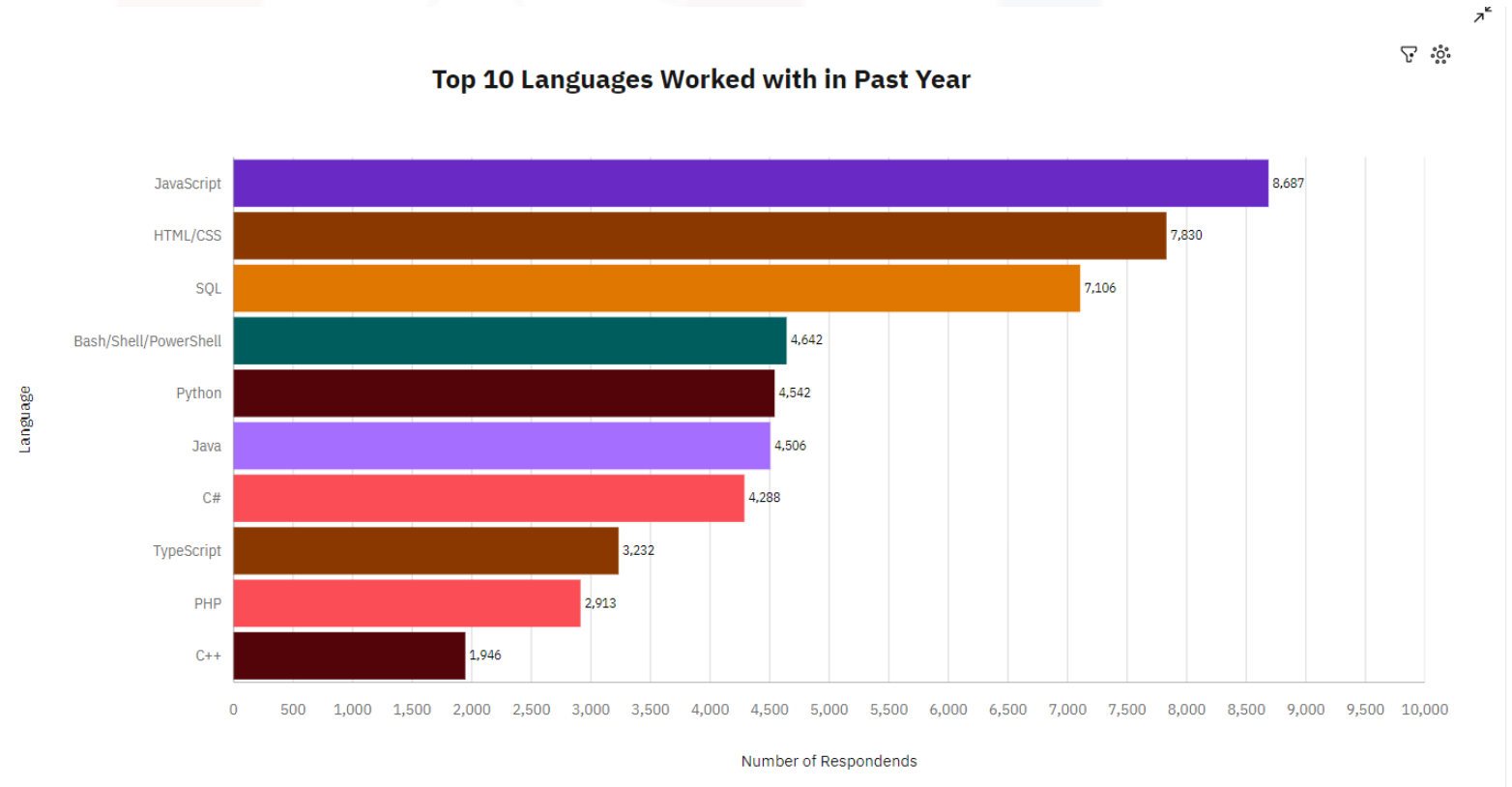
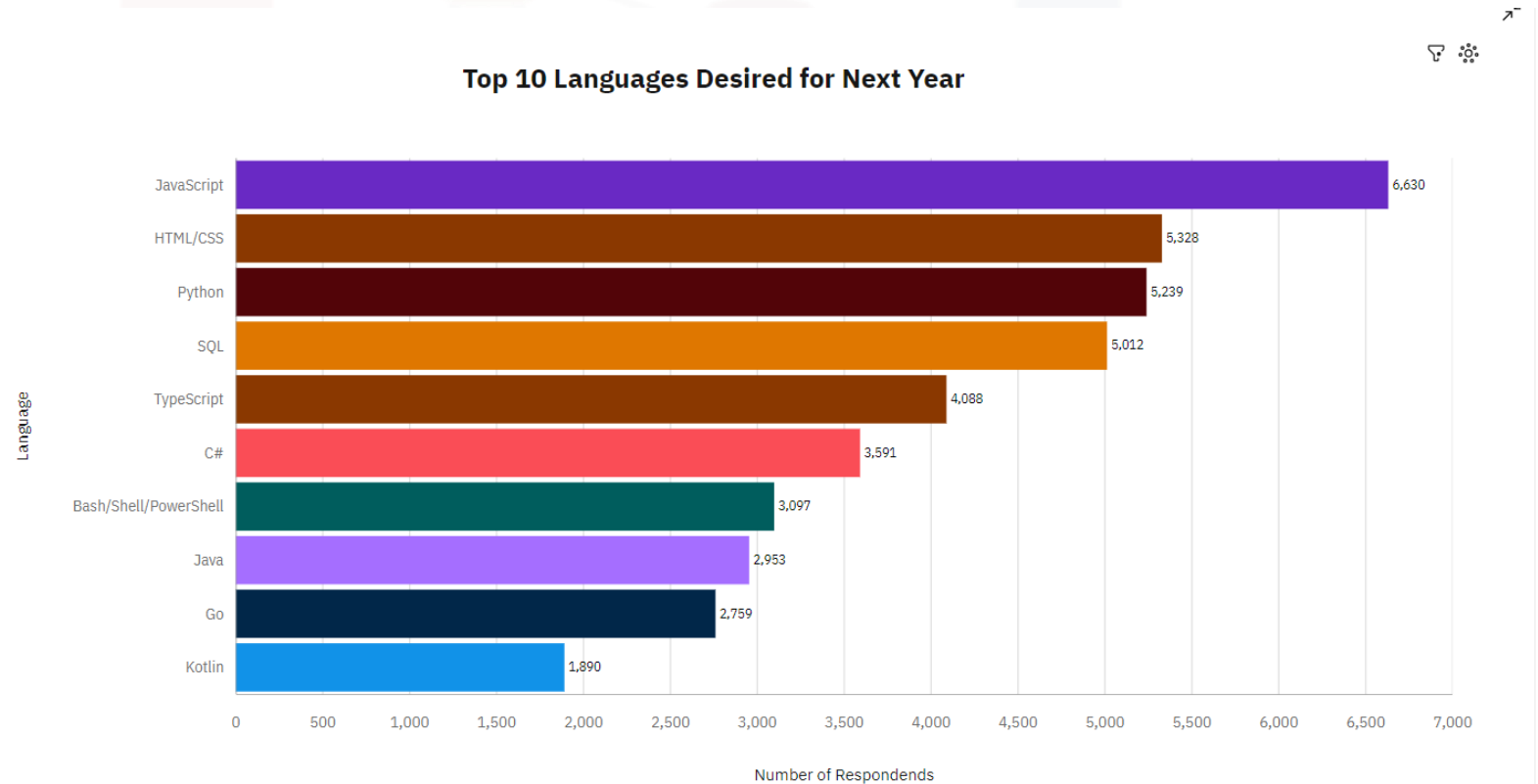# RESULTS

PROGRAMMING LANGUAGES PAST YEAR

Top 10:

1. JavaScript
2. HTML/CSS
3. SQL
4. Bash/PowerShell
5. Python
6. Java
7. C#
8. TypeScript
9. PHP
10. C++

Top 10 Languages Worked with in Past Year

Language	Number of Respondends
JavaScript	8,687
HTML/CSS	7,830
SQL	7,106
Bash/Shell/PowerShell	4,642
Python	4,542
Java	4,506
C#	4,288
TypeScript	3,232
PHP	2,913
C++	1,946

PROGRAMMING LANGUAGES DESIRED NEXT YEAR

Top 10:

1. JavaScript
2. HTML/CSS
3. Python
4. SQL
5. TypeScript
6. C#
7. Bash/Shell/PowerShell
8. Java
9. Go
10. Kotlin

Top 10 Languages Desired for Next Year

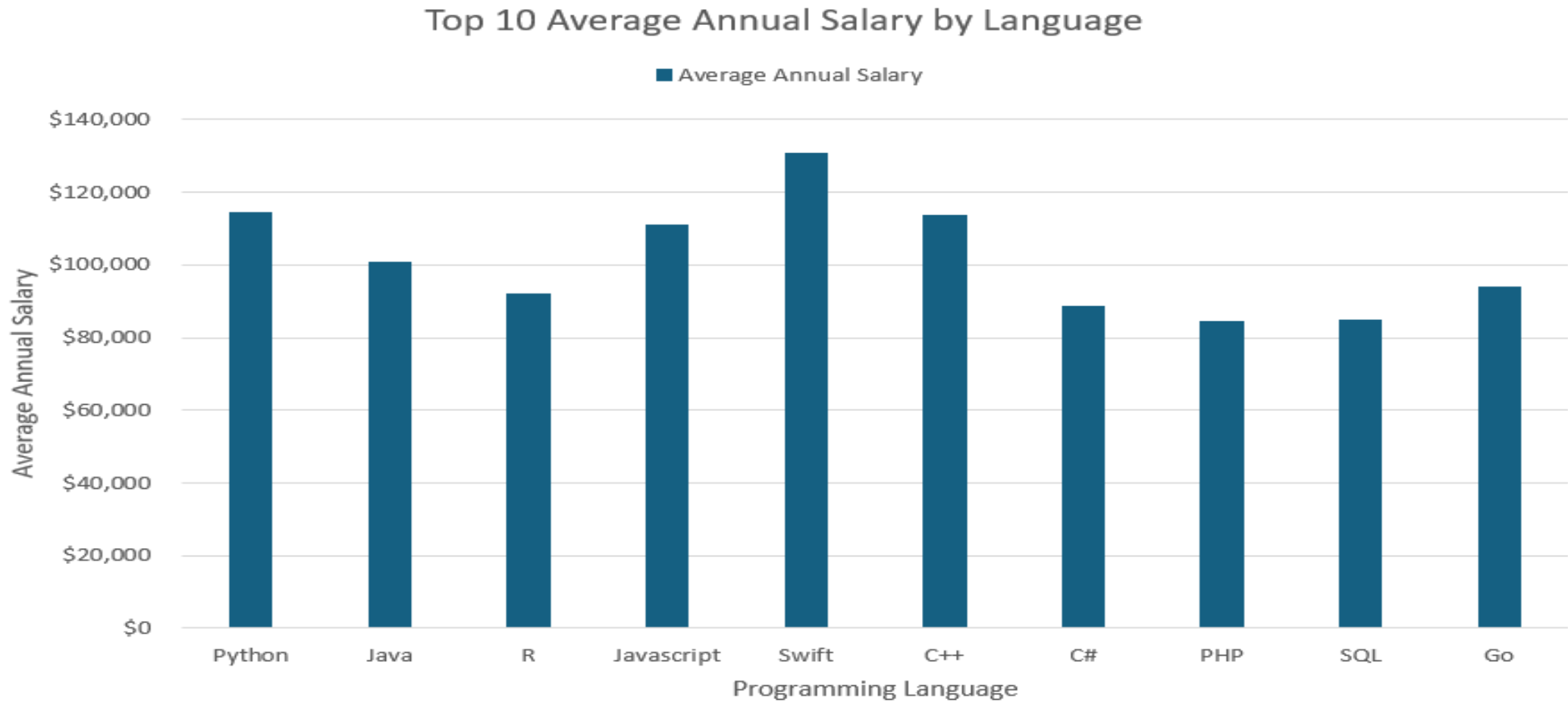| Language	Number of Respondends
JavaScript	6,630
HTML/CSS	5,328
Python	5,239
SQL	5,012
TypeScript	4,088
C#	3,591
Bash/Shell/PowerShell	3,097
Java	2,953
Go	2,759
Kotlin	1,890

IBM Developer

SKILLS NETWORK

Top Paying Languages

Top 10 Average Annual Salary by Language

■ Average Annual Salary

PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

- Python, TypeScript, and C# were higher for desired next year than for used in past year.

- TypeScript saw the biggest change.

- Go and Kotlin appeared in top 10 desired next year but not in used in past year.

- SQL, Bash/PowerShell, and Java were lower for desired next year than for used in past year.

- PHP and C++ appeared in top 10 for used in past year, but not in desired next year.

- Swift is highest paying language even though it does not appear in either top ten list for used or desired.
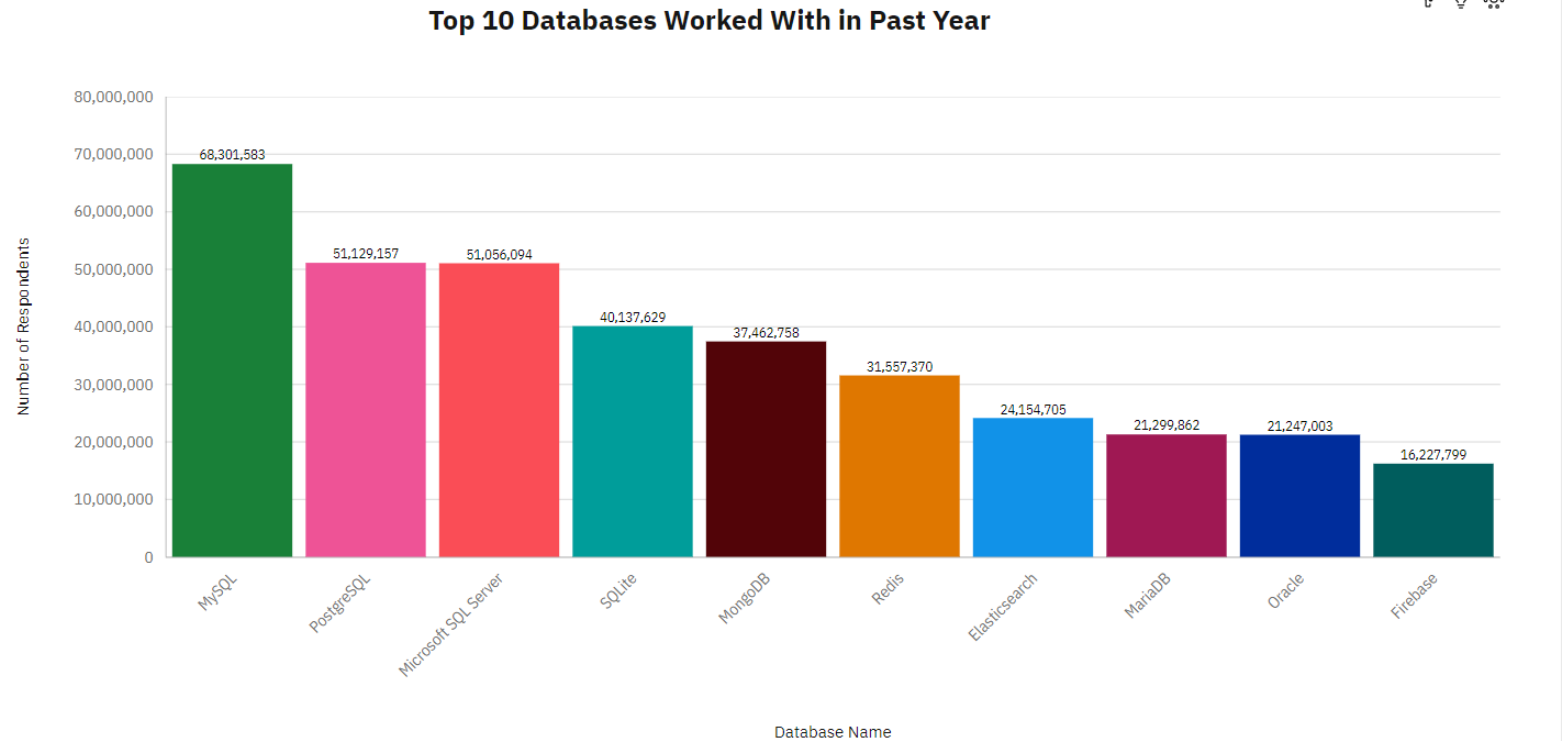
Implications

- Demand for JavaScript, Python, and TypeScript programming remains high and shows higher average salaries.

- TypeScript demand is growing at a higher rate than others.

- Go and Kotlin are newer languages that are rising in demand.

- Demand for SQL, Bash/PowerShell, Java, PHP, and C++ programmers is decreasing

- Swift language is specific to Apple which may be why it commands a higher pay yet is not as widely used or desired by developers.

IBM Developer

SKILLS NETWORK

DATABASES USED IN PAST YEAR

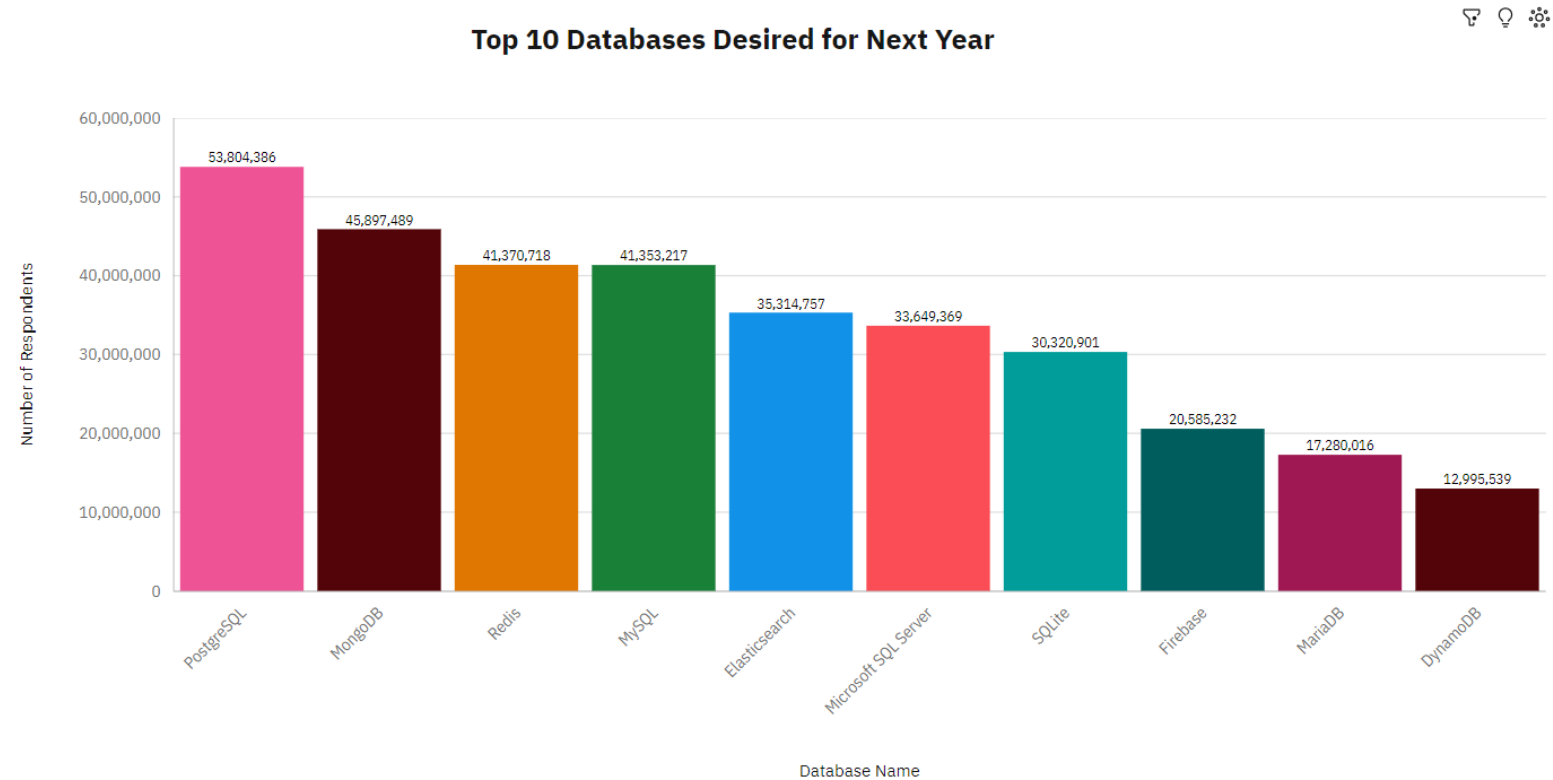Top 10 Databases Worked with in Past Year:

1. My SQL
2. PostgreSQL
3. Microsoft SQL Server
4. SQLite
5. MongoDB
6. Redis
7. Elasticsearch
8. MariaDB
9. Oracle
10. Firebase

Top 10 Databases Worked With in Past Year

IBM **Developer**

SKILLS NETWORK

DATABASES DESIRED FOR NEXT YEAR

Top 10 Databases Desired for Next Year:

1. PostgreSQL
2. MongoDB
3. Redis
4. MySQL
5. Elasticsearch
6. Microsoft SQL Server
7. SQLite
8. Firebase
9. MariaDB
10. DynamoDB

Top 10 Databases Desired for Next Year

Database	Number of Respondents
PostgreSQL	53,804,386
MongoDB	45,897,489
Redis	41,370,718
MySQL	41,353,217
Elasticsearch	35,314,757
Microsoft SQL Server	33,649,369
SQLite	30,320,901
Firebase	20,585,232
MariaDB	17,280,016
DynamoDB	12,995,539

IBM Developer

SKILLS NETWORK

Top 10 Databases
Used in Past Year Compared to Desired Next Year

Top 10 Worked with in Past Year:

1. My SQL
2. PostgreSQL
3. Microsoft SQL Server
4. SQLite
5. MongoDB
6. Redis
7. Elasticsearch
8. MariaDB
9. Oracle
10. Firebase

Top 10 Desired for Next Year:

1. PostgreSQL   +1
2. MongoDB   +2
3. Redis   +3
4. MySQL   - 3
5. Elasticsearch   +2
6. Microsoft SQL Server   -3
7. SQLite   -3
8. Firebase   +2
9. MariaDB   -1
10. DynamoDB

IBM **Developer**

SKILLS NETWORK

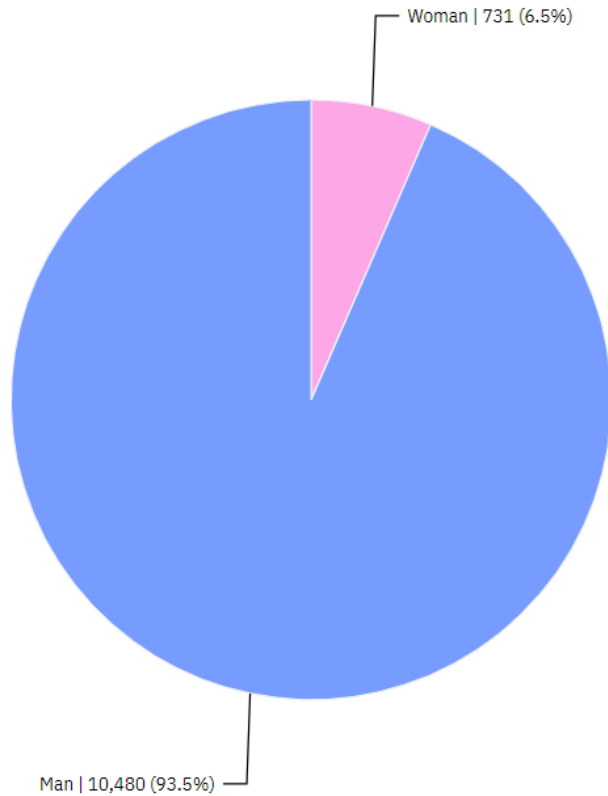# DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- PostgreSQL, MongoDB, Redis, Elasticsearch, and Firebase were higher for desired next year than for used in past year.

- Redis saw the biggest change.

- DynamoDB appeared in top 10 desired next year but not in used in past year.

- MySQL, Microsoft SQL Server, SQLite, MariaDB, and Oracle were lower for desired next year than for used in past year.

- Oracle appeared in top 10 for used in past year, but not in desired next year.
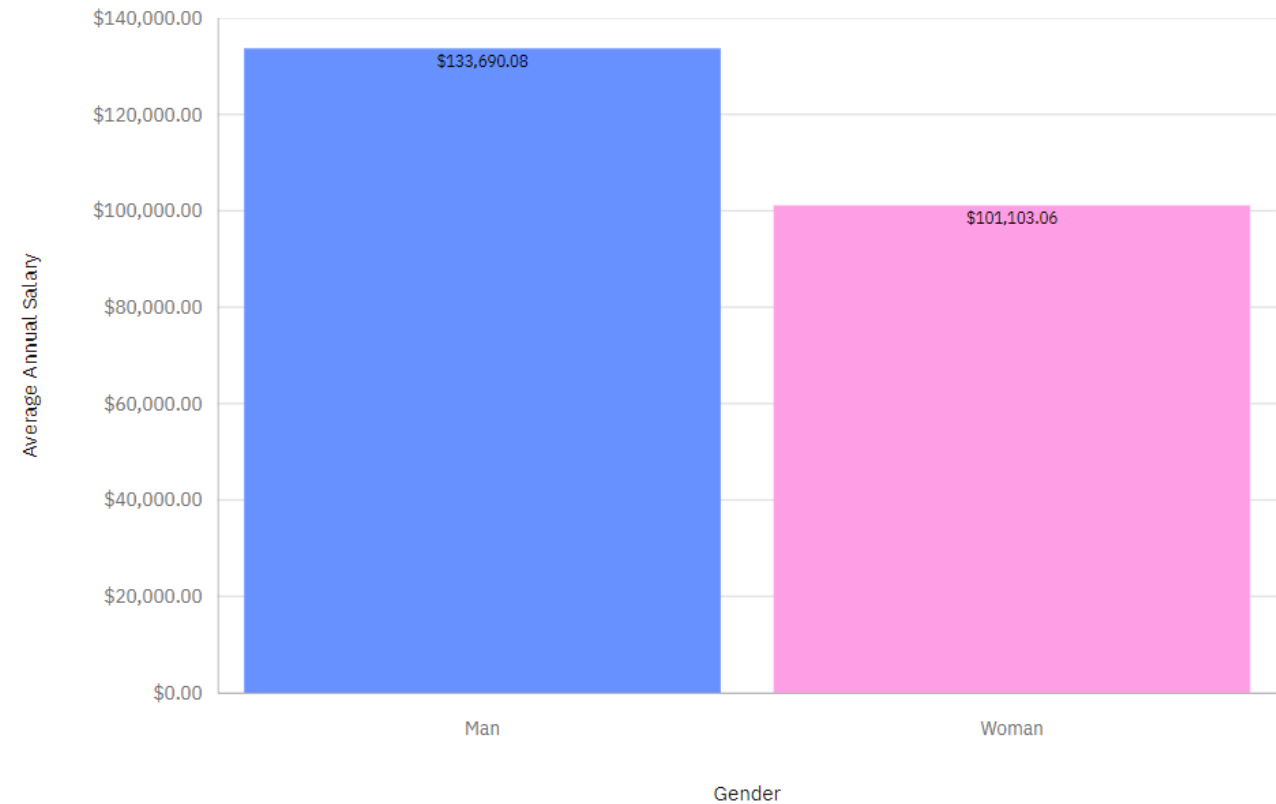
Implications

- Demand for PostgreSQL, MongoDB, Redis, Elasticsearch, and Firebase is growing.

- Redis demand is growing at a higher rate than others.

- DynamoDB (Amazon's fully managed NoSQL database) is rising in use and demand.

- MySQL, Microsoft SQL Server, SQLite, MariaDB, and Oracle demand is decreasing.

GENDER GAP AMONG SURVEY RESPONDENTS

Gender of Survey Respondents

Woman | 731 (6.5%)

Man | 10,480 (93.5%)

Average Annual Salary by Gender

$140,000.00

$120,000.00

$100,000.00

$80,000.00

$60,000.00

$40,000.00

$20,000.00

$0.00

Average Annual Salary

$133,690.08

$101,103.06

Man

Woman

Gender

DISCUSSION

Languages:

- Demand for JavaScript, Python, and TypeScript programming remains high and shows higher average salaries.
- Go and Kotlin are emerging languages that are likely to experience steep growth in demand in coming years.
- Demand for SQL, Bash/PowerShell, Java, PHP, and C++ programmers is on the decline.
- Swift language demand appears to be a niche market commanding high salaries but is not as widely used or desired among developers.

Databases:

- Demand for PostgreSQL, MongoDB, Redis, Elasticsearch, and Firebase is growing.
- Redis, DynamoDB appear poised for steep growth in coming years.
- MySQL, Microsoft SQL Server, SQLite, MariaDB, and Oracle demand is decreasing.

Demographics:

- Gender disparity is evidenced by the difference both in number and average annual salary of respondents to the developer survey.
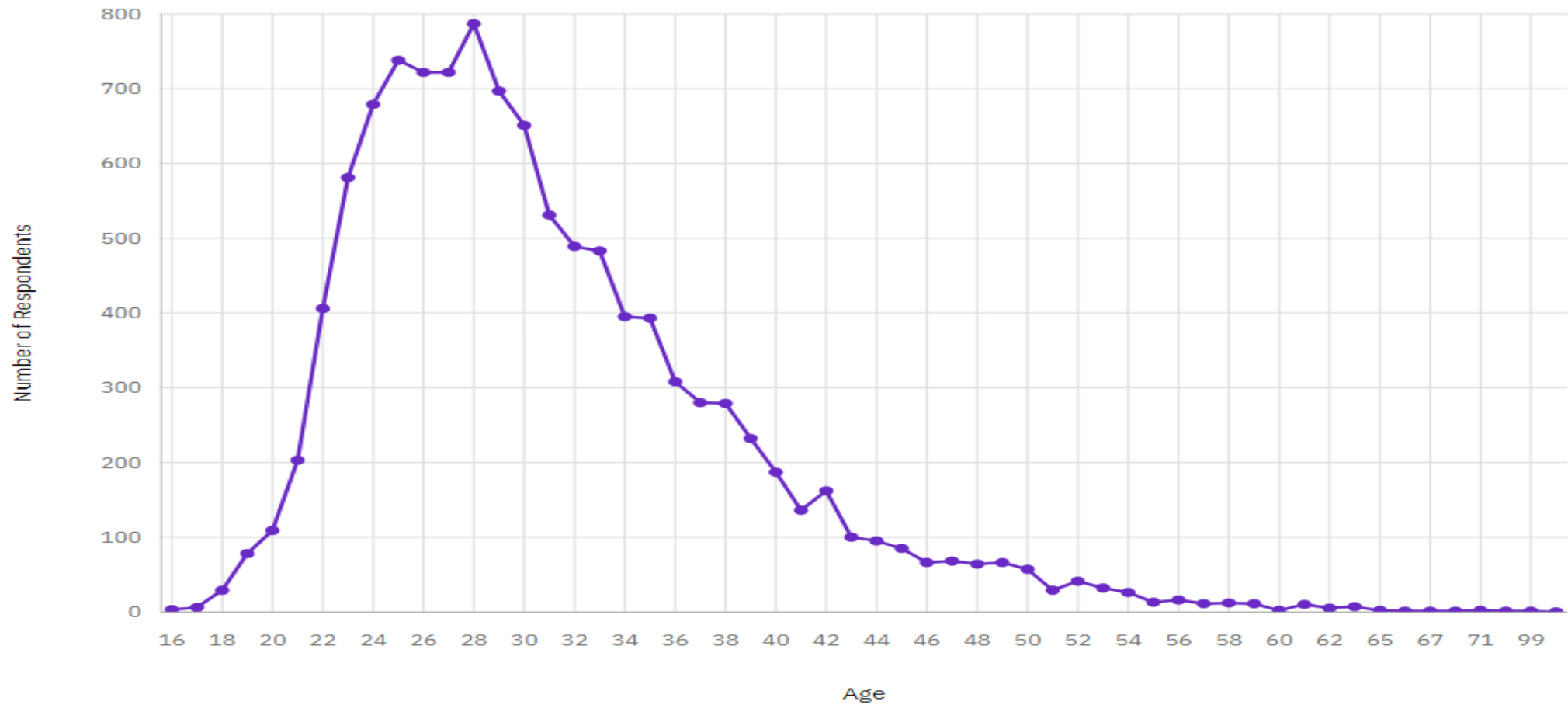
IBM Developer

SKILLS NETWORK

CONCLUSION

- The company should take an inventory of employee skills to ensure it is well equipped to meet trends in languages and databases.

- While not new languages, JavaScript and Python remain in strong demand and should remain a high priority.

- Go and Kotlin are new languages quickly growing in demand. This new need may require hiring new or upskilling current employees.

- PostgreSQL, MongoDB, Redis, Elasticsearch, Firebase, and DynamoDB are all database skills growing in demand. These skills should be the focus of hiring and internal training efforts for the company.

- Overcoming the gender gap in the developer community should be an area of focus in recruiting and training. The company should evaluate it's own compensation practices and policies to ensure there is no gender disparity in pay for equal work.

IBM Developer

SKILLS NETWORK

APPENDIX

Some additional data that was considered, but did not significantly contribute to the analysis or conclusions is presented in the following slides.
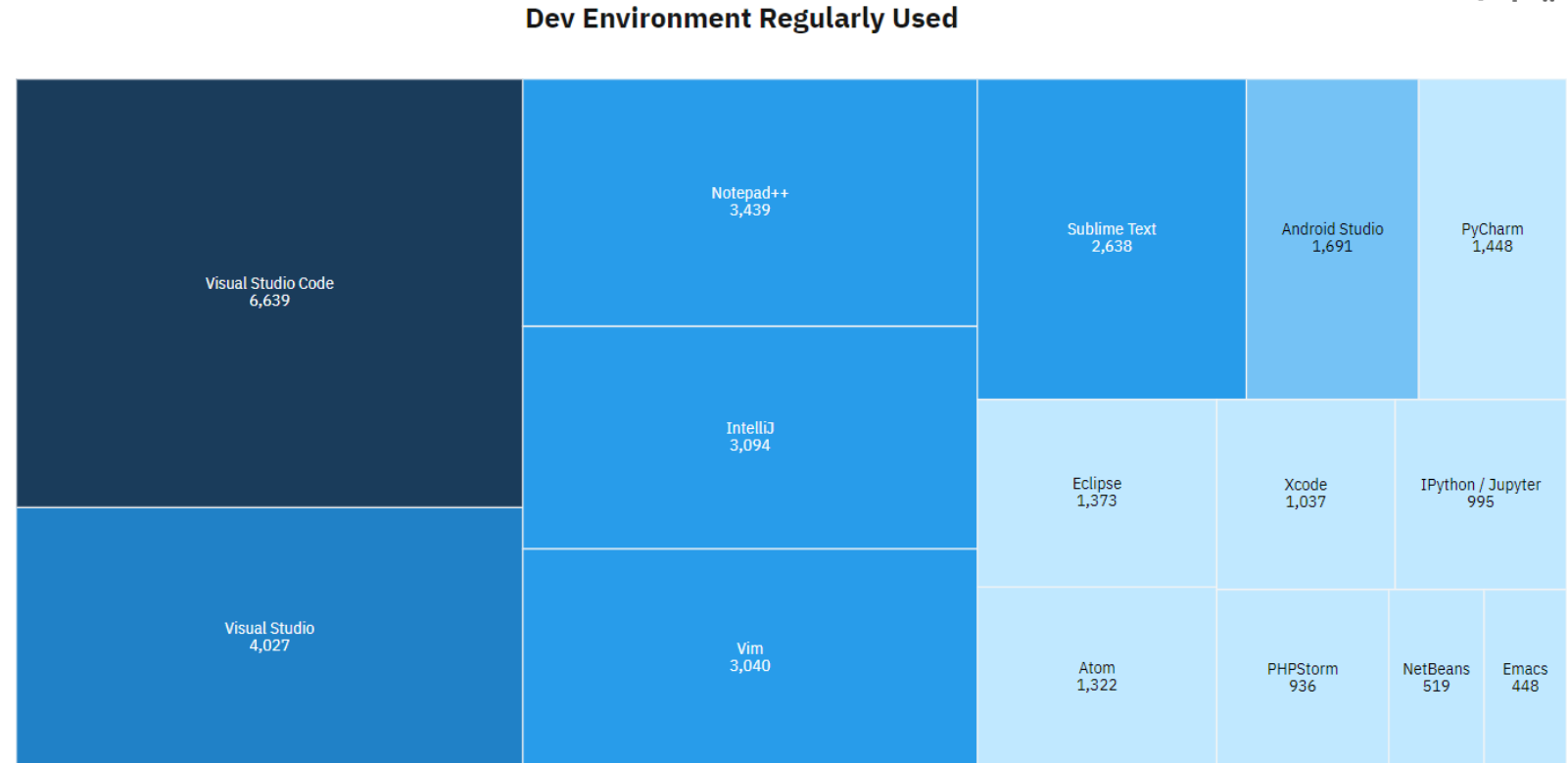
IBM **Developer**

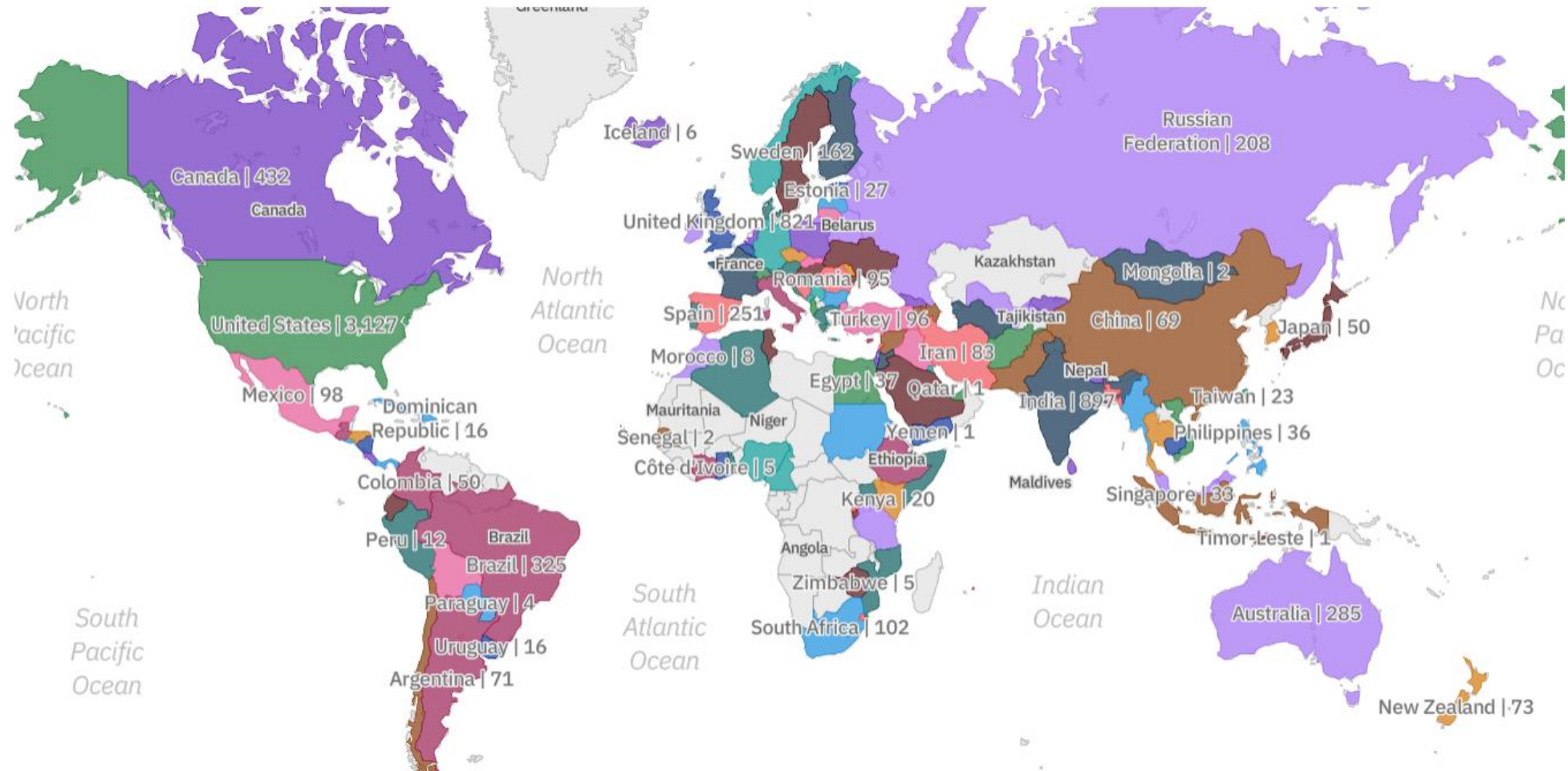SKILLS NETWORK

AGE of SURVEY RESPONDENTS

DEV ENVIRONMENTS REGULARLY USED

Top 10:
1.   Visual Studio Code
2.   Visual Studio
3.   Notepad ++
4.   IntelliJ
5.   Vim
6.   Sublime Text
7.   Android Studio
8.   PyCharm
9.   Eclipse
10. Atom

Dev Environment Regularly Used

Visual Studio Code
6,639

Visual Studio
4,027

Notepad++
3,439

IntelliJ
3,094

Vim
3,040

Sublime Text
2,638

Android Studio
1,691

PyCharm
1,448

Eclipse
1,373

Xcode
1,037

IPython / Jupyter
995

Atom
1,322

PHPStorm
936

NetBeans
519

Emacs
448

IBM **Dev**eloper

SKILLS NETWORK

SURVEY RESPONDENTS BY COUNTRY

JOB POSTINGS

Number of Job Postings from GitHub Jobs API by Skill Required

(Y-axis: Skill Required — MySQL Server, PostgreSQL, Scala, MongoDB, SQL Server, C++, C#, JavaScript, Oracle, Python, Java, C)

(X-axis: Number of Job Postings — 0, 2000, 4000, 6000, 8000, 10000, 12000, 14000, 16000)