Analysis of 2019 Stack Overflow Developer Survey

Phone Sett Paing

April 2022

IBM **Developer**

SKILLS NETWORK

OUTLINE

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

IBM Developer

SKILLS NETWORK

EXECUTIVE SUMMARY

- This presentation summarize key findings from analyzing **2019 Stack Overflow Developer Survey Results** dataset.

- Analyzing the dataset give insights to the following:
    - **Current Technology Usage** (e.g., popular language, database, platform or web frame among developers)
    - **Future Technology Trend** (e.g., most language, database, platform or web frame that developers want to work in over the next year)
    - **Demographics** (e.g., gender gap among developers, age groups of developers)

- These findings are relevant in order to keep pace with changing technologies, remain competitive and identify future skill requirements for the organization.

INTRODUCTION

- With ever changing technologies in the field of IT, new problems will emerge that require new solutions.

- As a global IT and Business consulting services frim, we must be able to provide expertise in IT solutions to these emerging new problems to remain competitive.

- To solve these problems, we need teams of highly experienced IT consultants in these changing technologies and skills.

- This analysis will identify future skill requirements and give insights and trends in the following:
  - What are the top programming languages in demand?
  - What are the top database skills in demand?

METHODOLOGY

- **Data Source:** The subset (provided by IBM) of original data set (2019 stack overflow developer survey results) from Stack Overflow, a popular website for developers. This randomized subset contains around 1/10th of original data set.

- **Data Exploration:** Find out number of rows, columns and identify the data types of each column to get familiar with the data.

- **Data Wrangling:** Identify and remove duplicate values, identify and impute missing values and normalize data.

- **Exploratory Data Analysis:** Identify the distribution of data in the dataset, identify and remove outliers from the dataset and identify correlation between features in the dataset.

- **Data Visualization:** Visualize the distribution of data, relationships in data, composition of data and comparison of data using various Python libraries and using Cognos Dashboard Embedded (CDE) through IBM Cloud to create the following dashboard:
  - Current Technology Usage
  - Future Technology Trend
 - Demographics

IBM **Developer**

SKILLS NETWORK

RESULTS

- The initial subset data contain **11552 rows** and **85 columns**.

- After removing duplicates, imputing missing values and removing outliers, the data contain **10519 rows** and **85 columns**.

- The survey was collected from developers in **135 countries** world wide

- Survey Respondents had a **median age of 29** and **majority** of respondents were **male**.

- **Medium** Annual Salary: **$52,704 USD**

- **Average** Annual Salary: **$59,883 USD**

- Annual Salary and Age were positively correlated: **r = .40**

PROGRAMMING LANGUAGE TRENDS

Top 5 Popular Languages Current Year
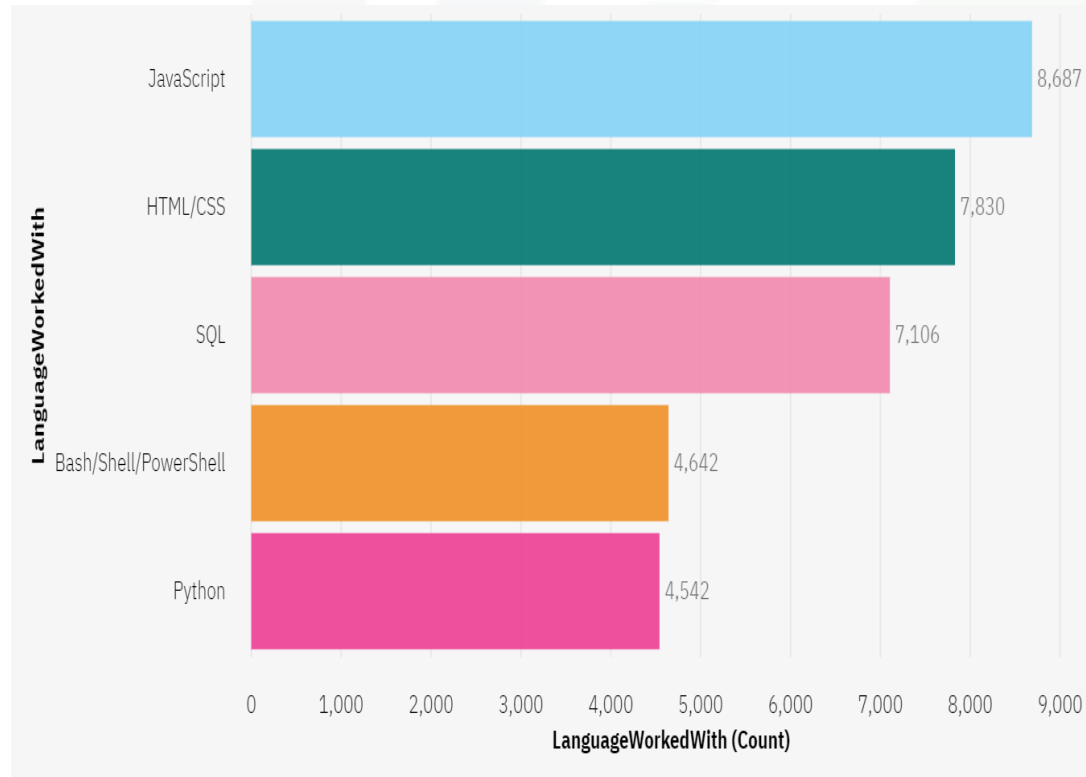
Top 5 Languages Desired Next Year
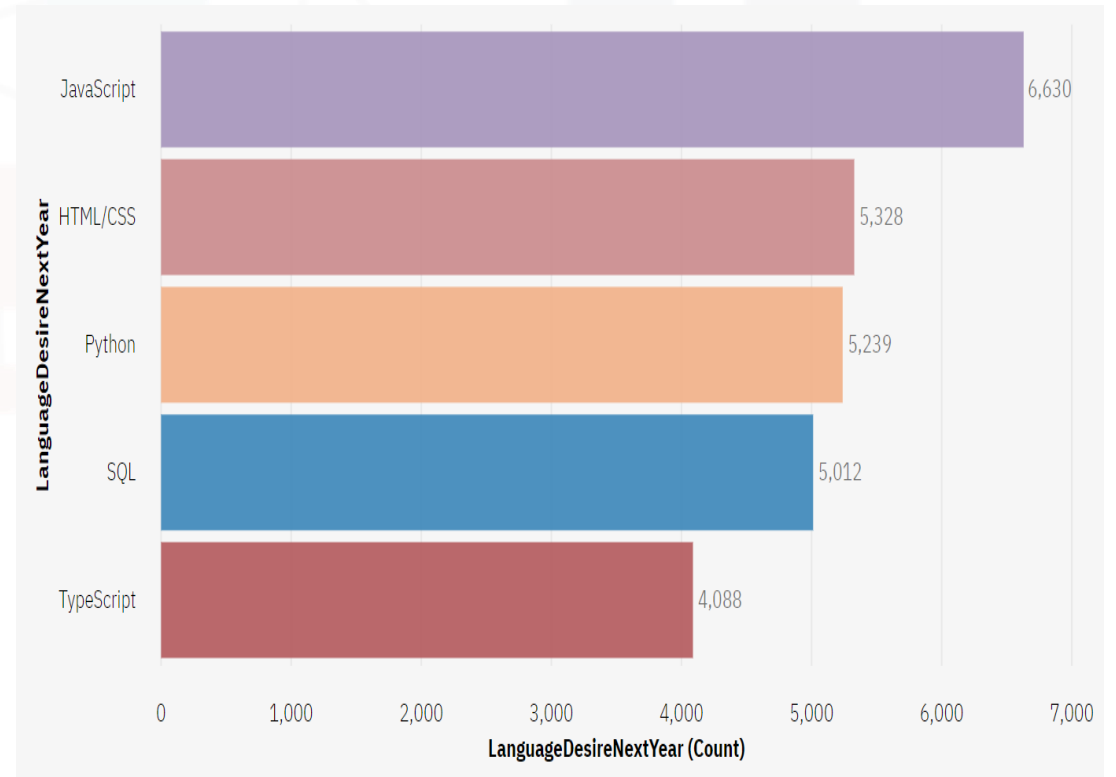
PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

- JavaScript and HTML/CSS are the top 2 popular languages for current year and next year.

- SQL popularity will drop in next year but will still remain in top 5 popular languages.

- Decreasing interest in Bash/Shell/PowerShell while interest in TypeScript increases.

- Massive increase of interest in Python.

Implications

- JavaScript and HTML/CSS remain as the most popular language for Web Development.

- SQL will remain as most used language for querying and storing of data.

- Increase of interest in Python, due to the increase of popularity in Data Science, Machine Learning and Deep Learning AI.
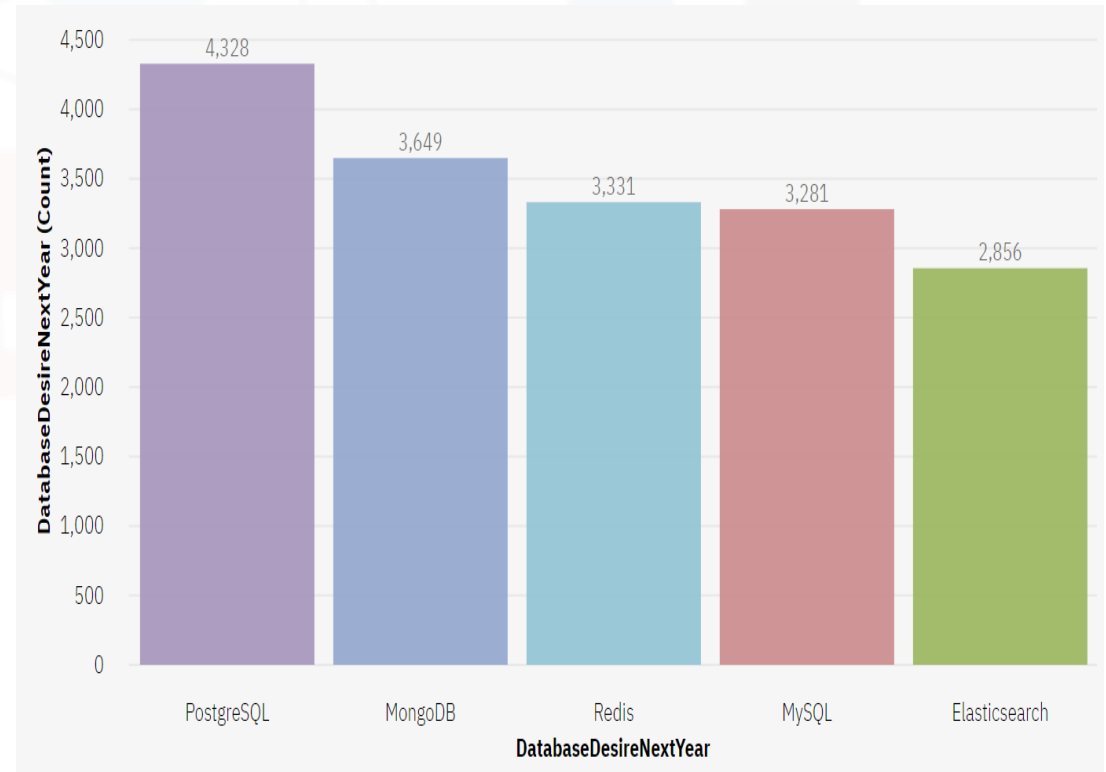
IBM Developer

SKILLS NETWORK

DATABASE TRENDS

Top 5 Popular Database Current Year
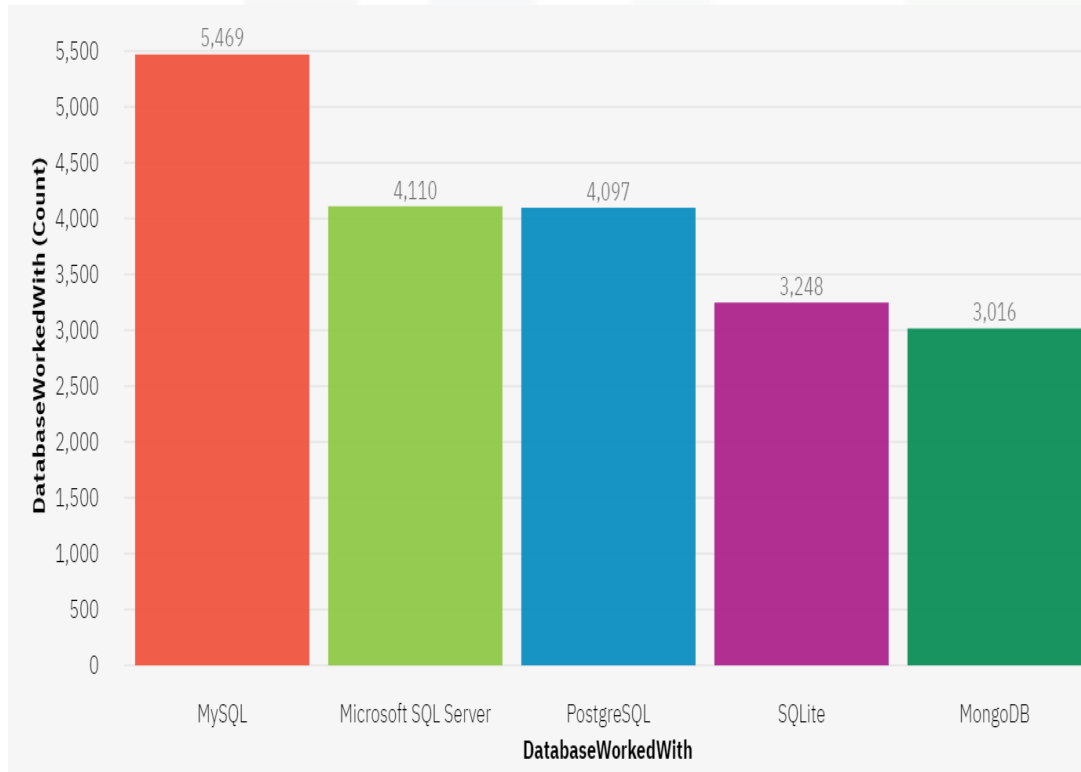
Top 5 Database Desired Next Year

DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- Increasing interest in PostgreSQL and MongoDB.

- Decreasing interest in MySQL.

- Lost of interest in Microsoft SQL Server and SQLite while gain of interest in Redis and Elasticsearch.

Implications

- Developers prefer to use open-source databases.

- Increasing interest in NoSQL databases, due to an increase need in handling structure and non-structure data for social media and web applications.

- Increase of interest in real-time databases, due to an increase popularity in Internet of Things (IoT) devices.
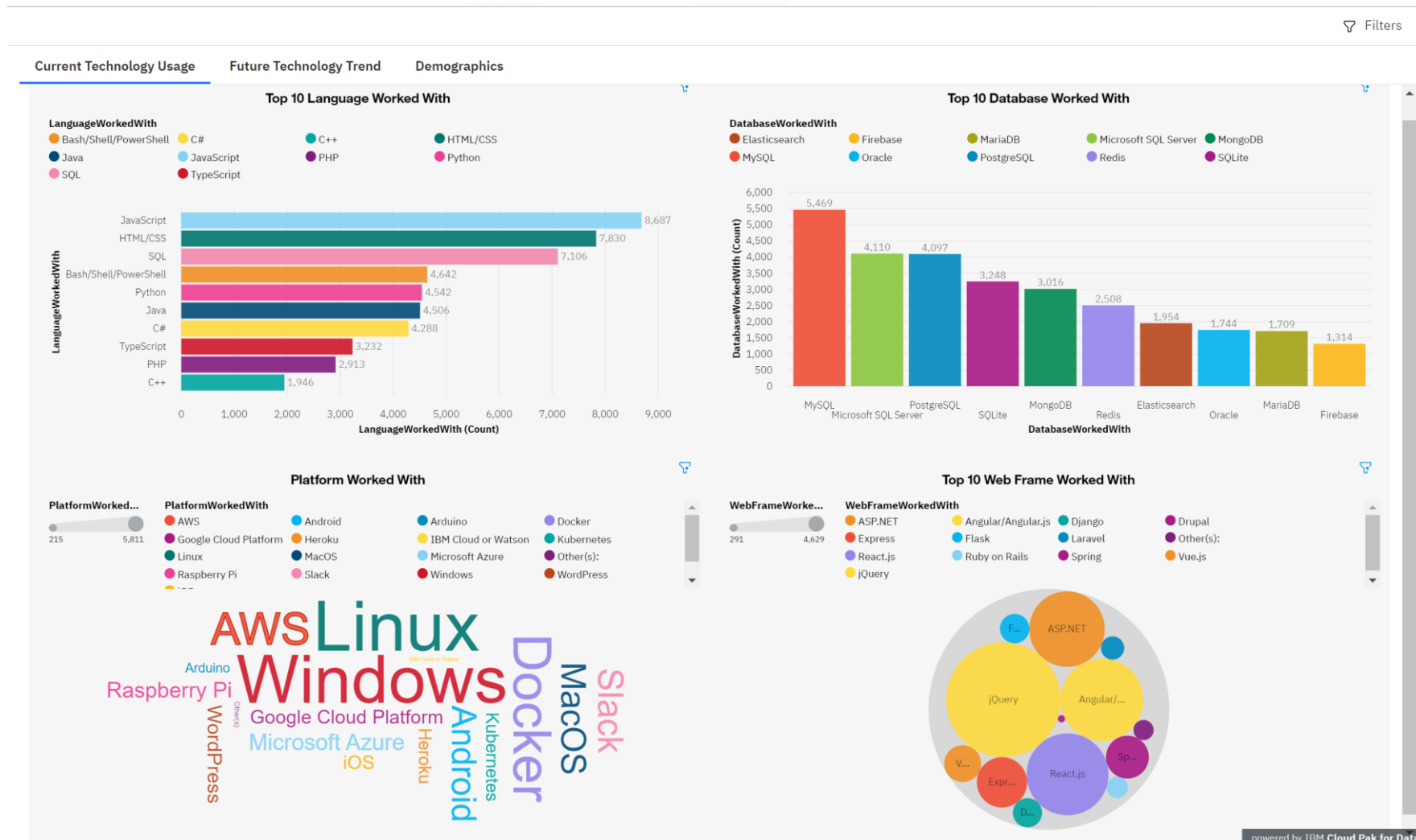
IBM Developer

SKILLS NETWORK

DASHBOARD

The following link contains the full, interactive Cognos dashboard:

Click here to open the dashboard **(Ctrl+Click)**

Screenshots of the dashboard are shown in the next three slides.
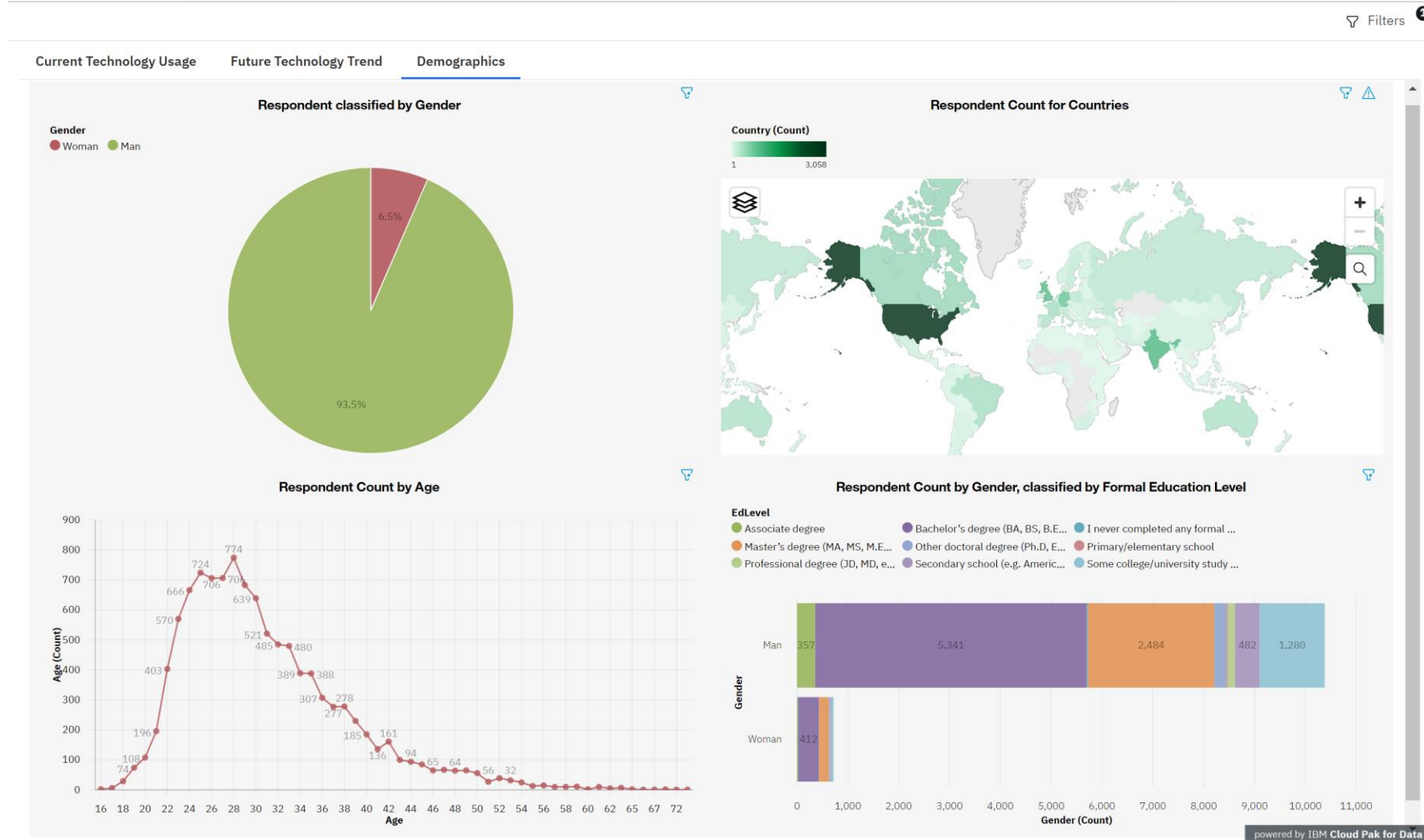
IBM **Developer**

SKILLS NETWORK

DASHBOARD A: CURRENT TECHNOLOGY USAGE

DASHBOARD B: FUTURE TECHNOLOGY TREND
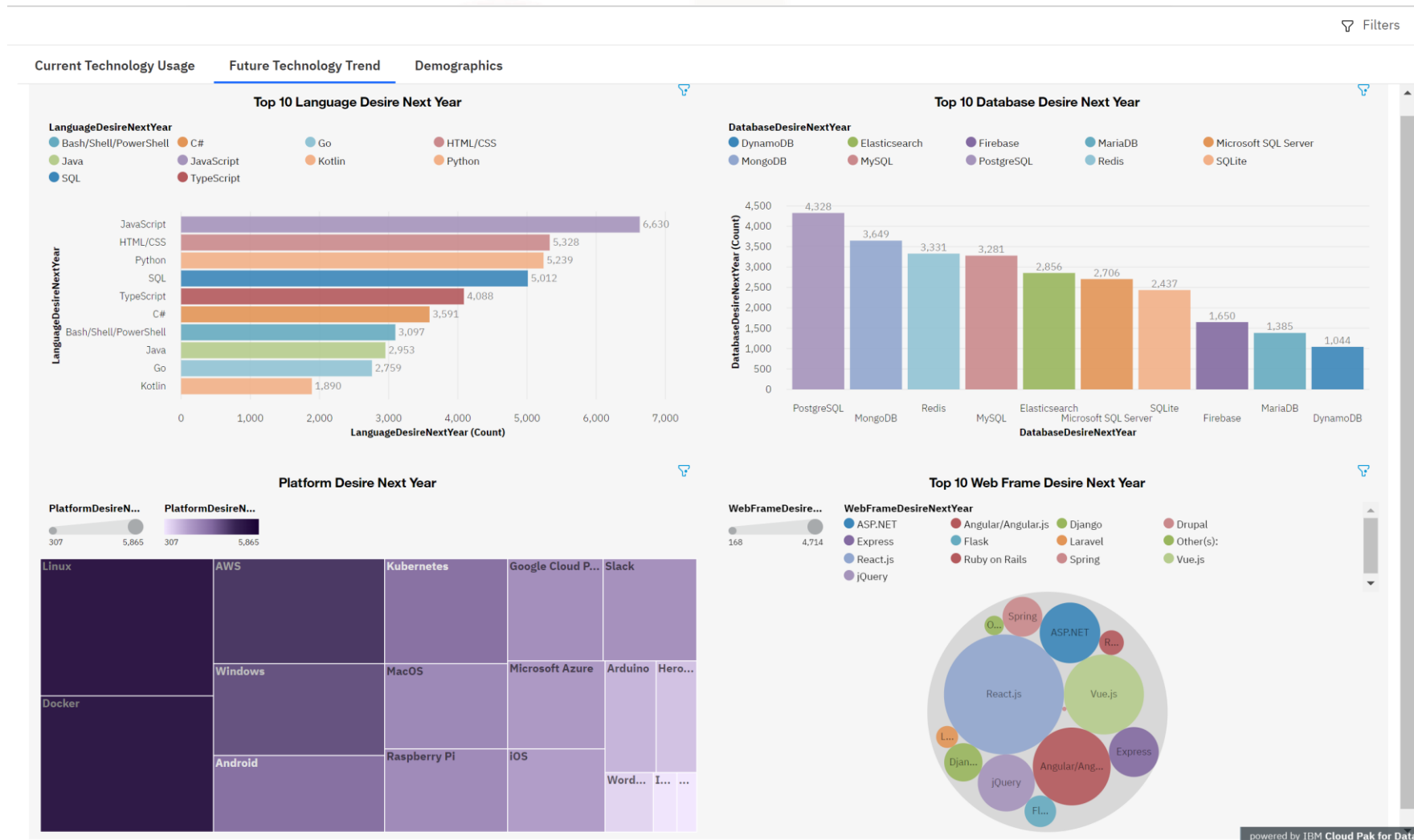
DASHBOARD C: DEMOGRAPHICS

DISCUSSION

The findings yield insights into the following questions:

- What are the top programming languages in demand and their future trend?

- What are the top database skills in demand and their future trend?

- Team of consultant with which skill will be needed?

- Which trainings should you provide to your employees?

- Should you train your existing team of consultants or hire new team of consultants?

OVERALL FINDINGS & IMPLICATIONS

Findings

- JavaScript and HTML/CSS are the most used languages for this year and the next.
- Increase of interest in Python and TypeScript.
- Increase of interest in PostgreSQL, MongoDB, Redis and Elasticsearch.

Implications

- Since Web Development languages are the most used programming languages among developer, Web development is still in high demand.
- Increase usage of Python might be due to increasing demands in Data Science, Deep Learning AI and Machine Learning.
- Increase usage of NoSQL databases as it is more scalable, provide superior performance and are easy for developers to work with. Apart form these it might also be due to growing needs to handle big unstructured data.

IBM **Developer**

SKILLS NETWORK

CONCLUSION

- Since Web Development is high in demands, team of consultants who are proficient in web development languages such as JavaScript and HTML/CSS are needed.

- If we already have enough teams, we need to provide Typescript (superset of JavaScript) training to them as it is gaining popularity in following year.

- If we do not have a team of consultants who are proficient in Python and are Data Science/Deep Learning AI/Machine Learning Professional, we will need to hire them as it is increasing in demand in coming year.

- Team of consultants who are proficient in PostgreSQL will need to be hire as it will increase in demands in next year.

- Team of consultants who are proficient in NoSQL databases such as MongoDB and team of consultants proficient in real-time databases such as Redis and Elasticsearch will need to be hire.
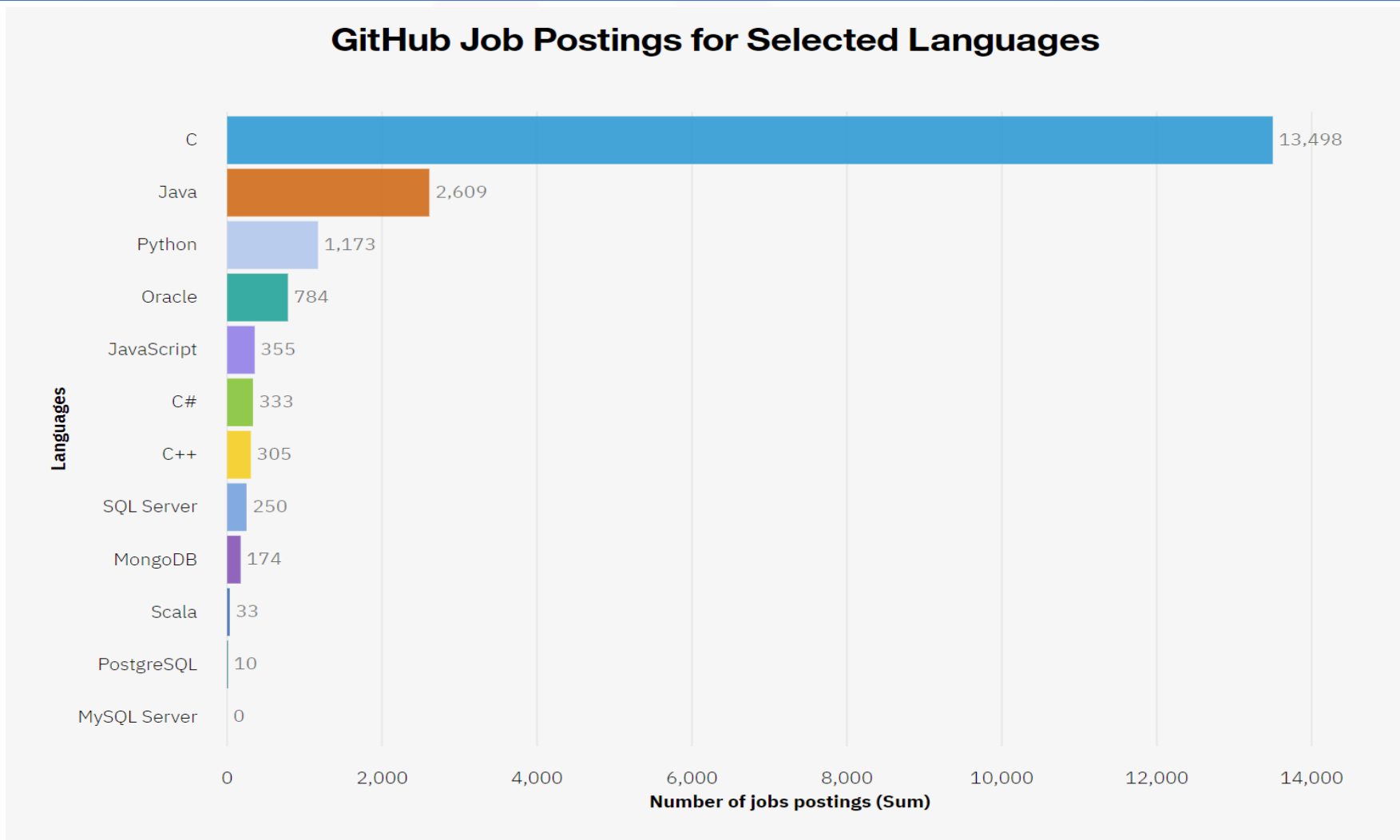
APPENDIX: RESOURCE

1. The subset of 2019 stack overflow developer survey results dataset provided by IBM. Link: **m1_survey_data**

2. Two modified subset of original dataset provided by IBM for building a dashboard with Cognos Dashboard Embedded (CDE). Link: **m5_survey_data_demographics.csv and m5_survey_data_technologies_normalised.csv**

3. Json file contents of GitHub Jobs API. Link: **Json file**

4. Survey data regarding Popular Programming Languages provided by IBM. Link: **Popular Programming Languages**

IBM **Developer**

SKILLS NETWORK

APPENDIX A: GITHUB JOB POSTINGS

GitHub Job Postings for Selected Languages

Languages	Number of jobs postings (Sum)
C	13,498
Java	2,609
Python	1,173
Oracle	784
JavaScript	355
C#	333
C++	305
SQL Server	250
MongoDB	174
Scala	33
PostgreSQL	10
MySQL Server	0

APPENDIX B: POPULAR LANGUAGES BY SALARY

Popular Programming Languages by 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