Data Analysis of Top Technologies Trends

By Ioana Bucuri

08.02.2021

OUTLINE

EXECUTIVE SUMMARY Part 1/3

- This data analysis was commissioned to examine the top programming languages and databases trends which are important for any IT and business consulting services firm
- Data from the following sources has been used
  - Job postings
  - Training portals
  - Surveys
- Key questions to be asked are:
  - What are the top programming languages in demand?
  - What are the top database skills in demand?
  - What are the popular IDEs?

EXECUTIVE SUMMARY Part 2/3

- Data collection, data exploration, data wrangling and data visualization have been performed for answering these questions
- Tools used for this project include Python programming language, Jupyter Notebooks, IBM Watson Studio, Microsoft Excel and IBM Cognos
- After the above steps have been performed using the listed tools, considering the quality of the remained data, the results can be considered accurate and trustful

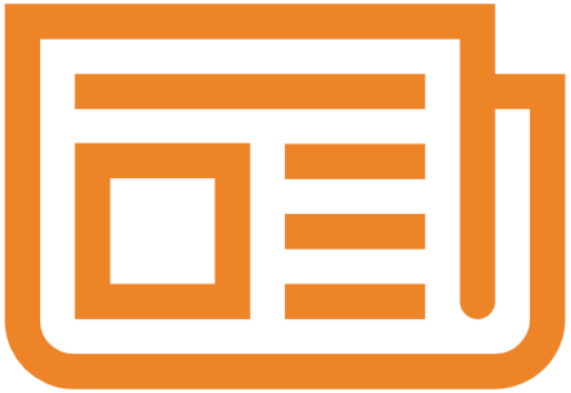
EXECUTIVE SUMMARY Part 3/3

- Using analytics and graphics, the questions being asked can be answered to, as follows:
  - On top programming languages for both the current year and the next year we can find JavaScript, HTML/CSS, SQL, Python, with the last one having an increased trend
  - MySQL, PostgreSQL and MongoDB are found in top 5 database trends for current and next year, while technologies not being on top in the current year seem to have an increased trend, like Redis and Elasticsearch
  - Top platforms for current year and next year are Windows, AWS, Linux, and Docker, with the 2 latter's popularity increasing.
  - The preferred Web Frames are React.js, Vue.js, Angular/ Angular.js and ASP.NET
- In conclusion, Company shall develop or continue developing the skills required for the above languages, databases, platforms and web frames.

INTRODUCTION

- With the goal of remaining competitive while technologies are evolving at a fast pace, identifying future skills requirements is an essential task for a global IT and business consulting services company.
- The data was collected from various sources, including job postings, training portals and surveys.
- Key questions we want to answer to in this analysis include:
  - What are the top programming languages in demand?
  - What are the top database skills in demand?
  - What are the popular IDEs?

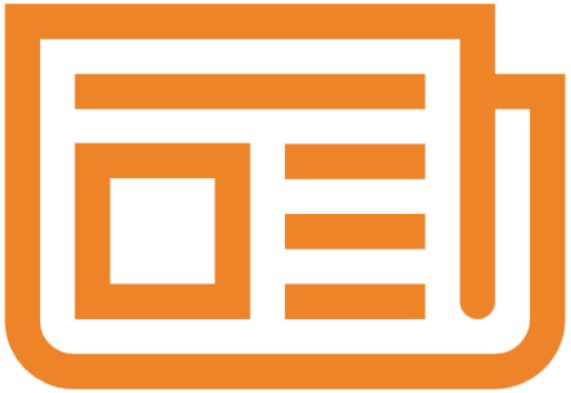
METHODOLOGY Part 1/4



1. Data Collection

- The data was collected from various sources with different techniques using Python in Jupyter Notebooks:
  - Jobs API on Github was used to download the job postings for the top programming languages
  - Scraping a website containing salary surveys to find out the programming languages that are most in demand, based on the average annual salary
  - Downloading the worldwide survey dataset available on Stack Overflow, a popular website for developers. The respondents of this survey provided information regarding languages and databases worked with, as well as their desire for the next year.

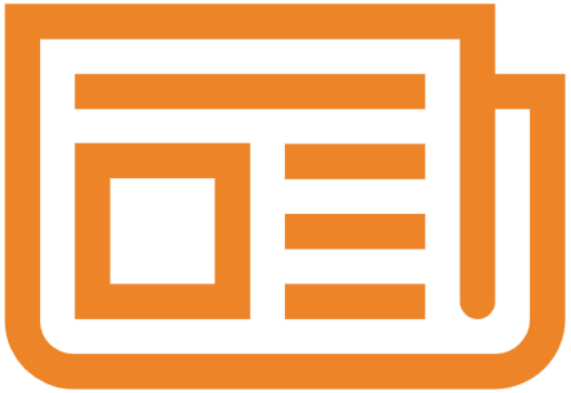
METHODOLOGY Part 2/4



2. Data Wrangling

- Data Wrangling was performed to clean up the dataset and make it ready for data analysis:
  - A new data frame containing only required columns has been created
  - Duplicate rows have been identified and deleted
  - The missing values were identified and deleted or replaced by the most appropriate values (like mean, median, frequency)
  - The data types and format were checked and corrected
  - Finally, data has been normalized so that the features are expressed using the same parameters (for example, yearly and monthly salaries were converted to yearly salaries)

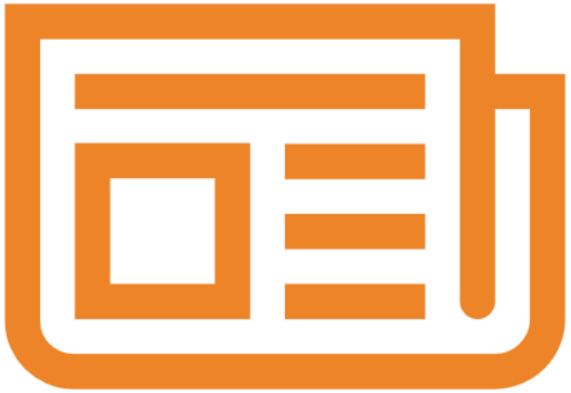
METHODOLOGY Part 3/4



3. Data Exploration

- Data Exploration has been performed to understand the collected data and to gain initial insights:
  - Data was loaded in a pandas dataframe and was explored in terms of number and content of rows and columns and data types
  - Descriptive statistics was used to understand the data in hand, like the mean of the respondents' ages, the countries where most of the respondents live, the data distribution and correlation between columns.
  - Outliers have been detected and dealt with
  - Considering there is a big amount of data which includes information regarding the current and past technologies, this step confirms the collected data is representative for the questions to be answered.

METHODOLOGY Part 4/4



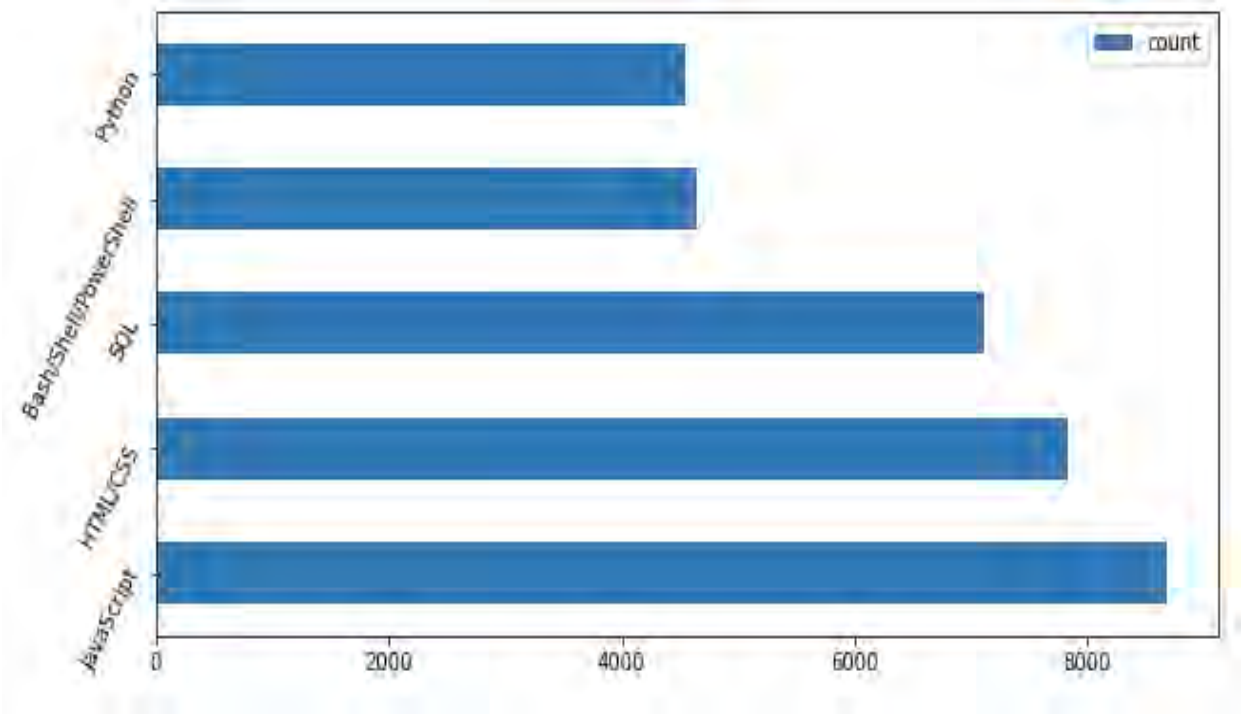
3. Data Visualization

- Data Visualization has been performed using Python in Jupyter Notebooks and IBM Cognos:
  - SQL queries and Pandas library were used to extract necessary data from RDBMS
  - Matplotlib library in Python was used to visualize the distribution and composition of data, the relationship between features and the data comparison
  - Dashboards were used in IBM Cognos to show the usage of top programming languages and databases, as well as the future trends, using different charts like bars and columns charts, word clouds and bubble charts

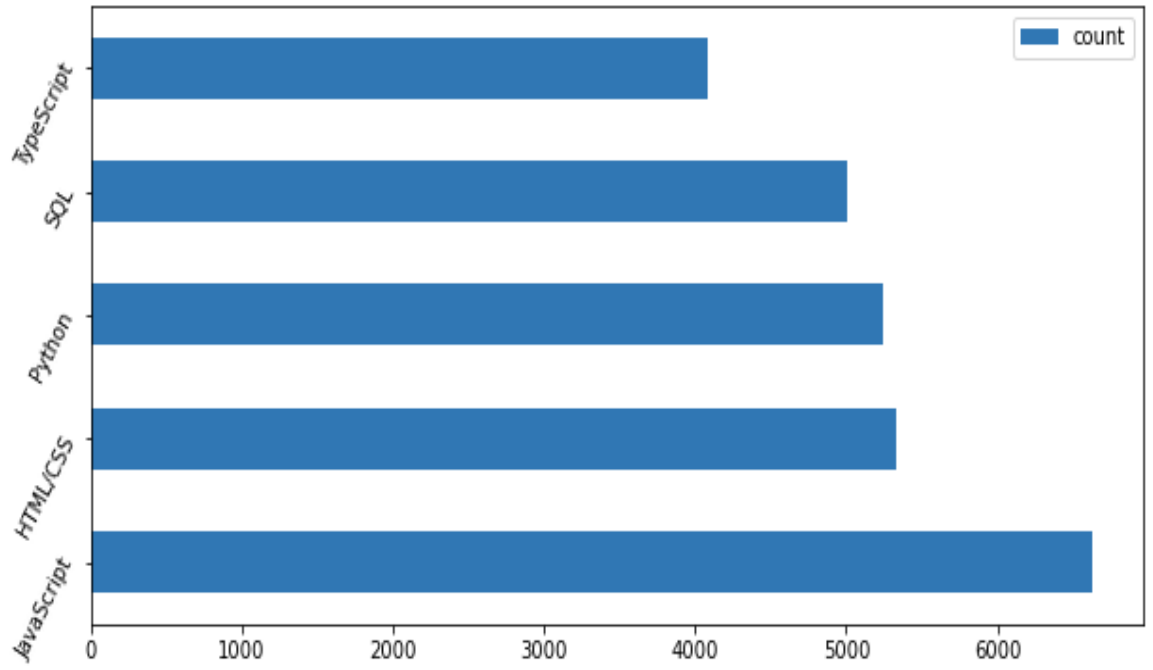
RESULTS

- Considering the collected data was consistent and representative for the questions being asked, we can count on it to reflect an accurate result.
- The figures on the next slides show the programming languages and database trends for the current and next year, the popular IDEs, the age and gender range of the surveyed respondents, as well as their distribution across countries.

PROGRAMMING LANGUAGE TRENDS



Current Year



Next Year

PROGRAMMING LANGUAGE TRENDS

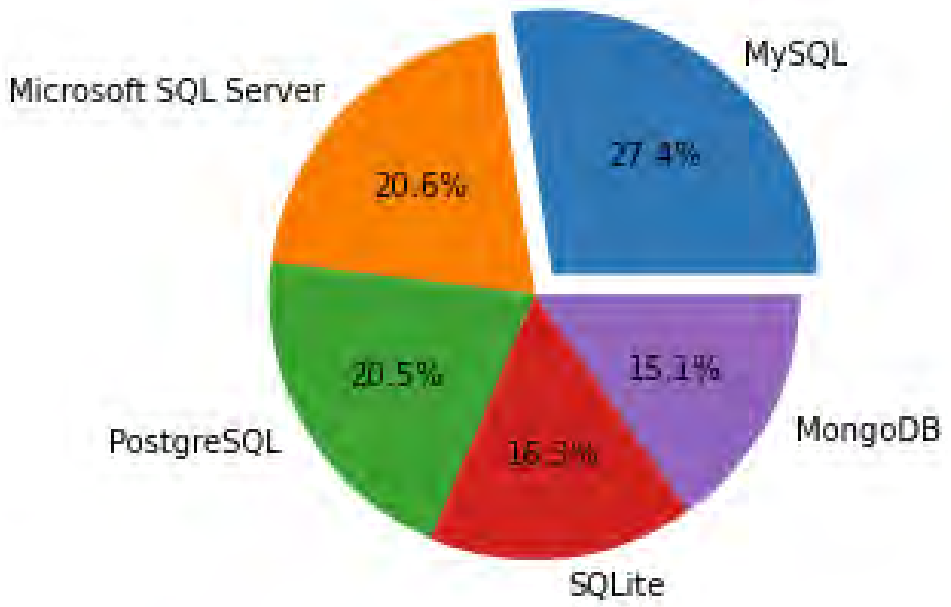
Findings

- JavaScript and HTML are the most popular languages in the current year and next year
- The popularity of Python and Typescript seems to be increasing in the next year
- SQL remains in top 5, while Bash/Shell/Powershell seems to decrease in popularity next year

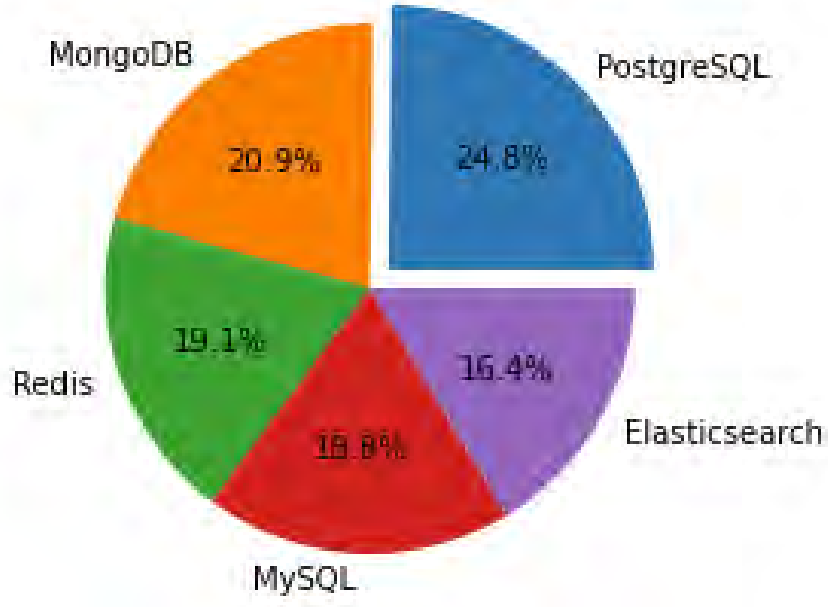
Implications

- Investment in developing Javascript and HTML skills is necessary
- Considering the evolving popularity of Python and Typescript, the same amount of attention shall be given to them too
- Investment in Bash /Shell/ Powershell skills shall not be as important as investment in SQL and in evolving technologies

DATABASE TRENDS



Current Year



Next Year

DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- Top 5 Database trends this year include MySQL followed by Ms SQL Server and PostgreSQL close to each other and SQLite and MongoDB
- In top 5 for next year, PostgreSQL replaces MySQL on top and MongoDB's popularity also increases, while SQL Server and SQLite are not on the top anymore, being replaced by Redis and Elasticsearch

Implications

- Attention shall be given to skills related to MySQL, PostgreSQL and MongoDB, as they are in top 4 trends for the current and next year
- Redis and Elasticsearch seem to be evolving fast, therefore, the company shall make sure to include them as must have skills

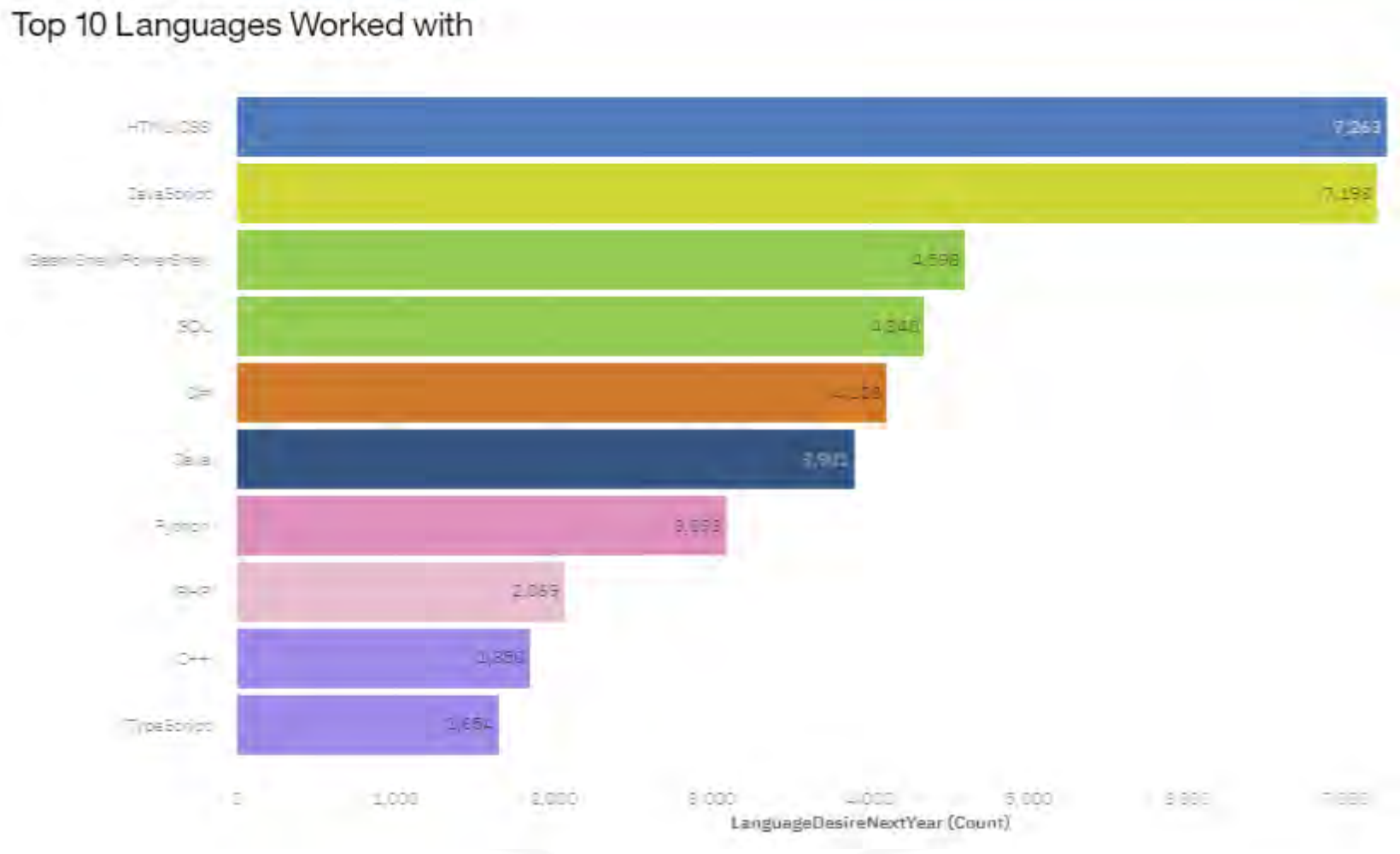
DASHBOARD

The following slides present the content of the dashboards which were created to show top 10 Programming Languages and Databases trends, Top 10 Platforms and Trends and brief information regarding Respondents' Ages, Countries, Gender and Studies.

You may use the link below to see the complete Dashboard.

https://eu-de.dataplatform.cloud.ibm.com/dashboards/3b86913e-d15c-4f3d-ac63-4aaf804535a7/view/0267e97f05ae2dc167d5f2e407cf7d062961745ee6bb860b85d37b495b617597f06c1090c87c435cdd450462faee130b9c

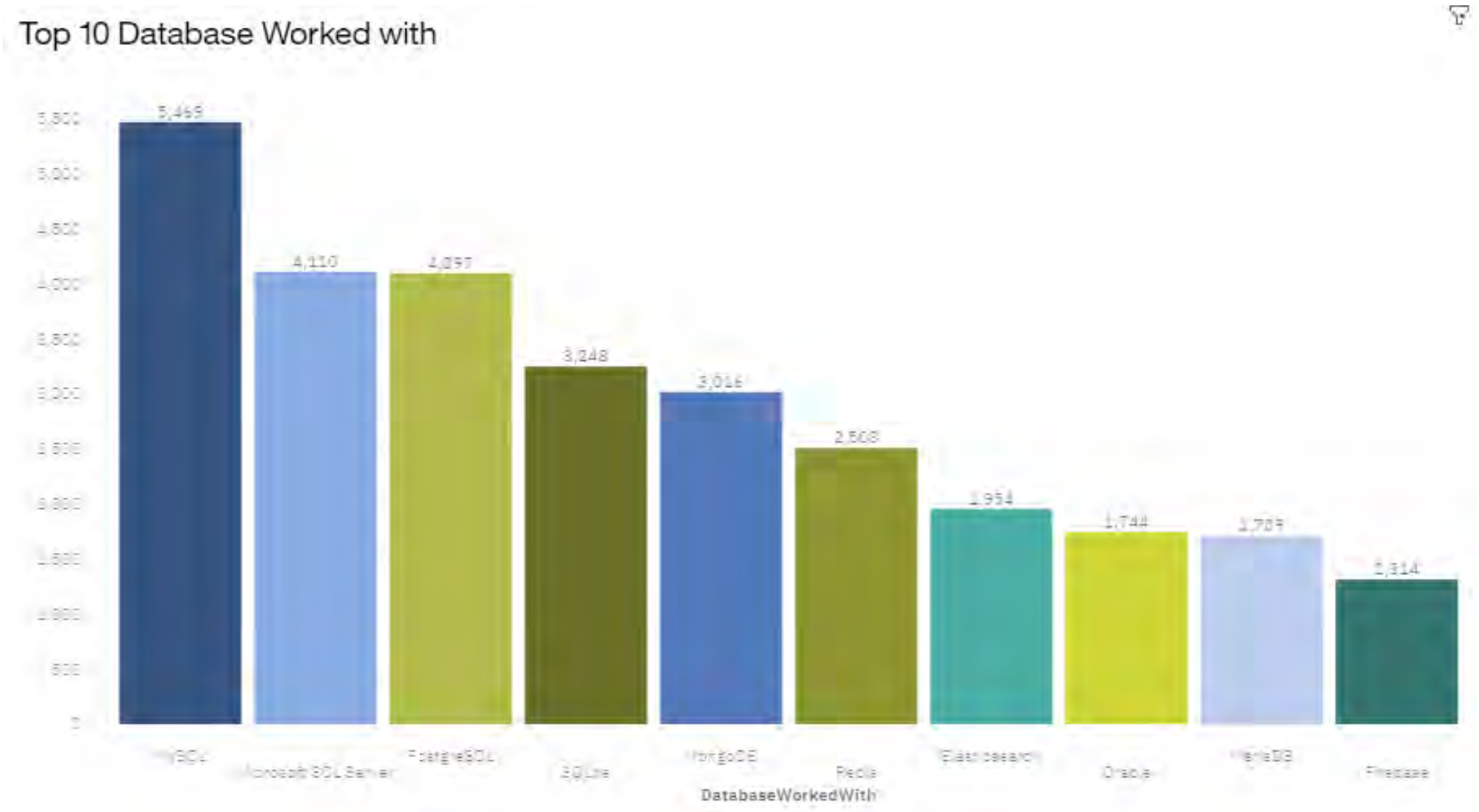
DASHBOARD TAB 1 Part 1/4



IBM Developer

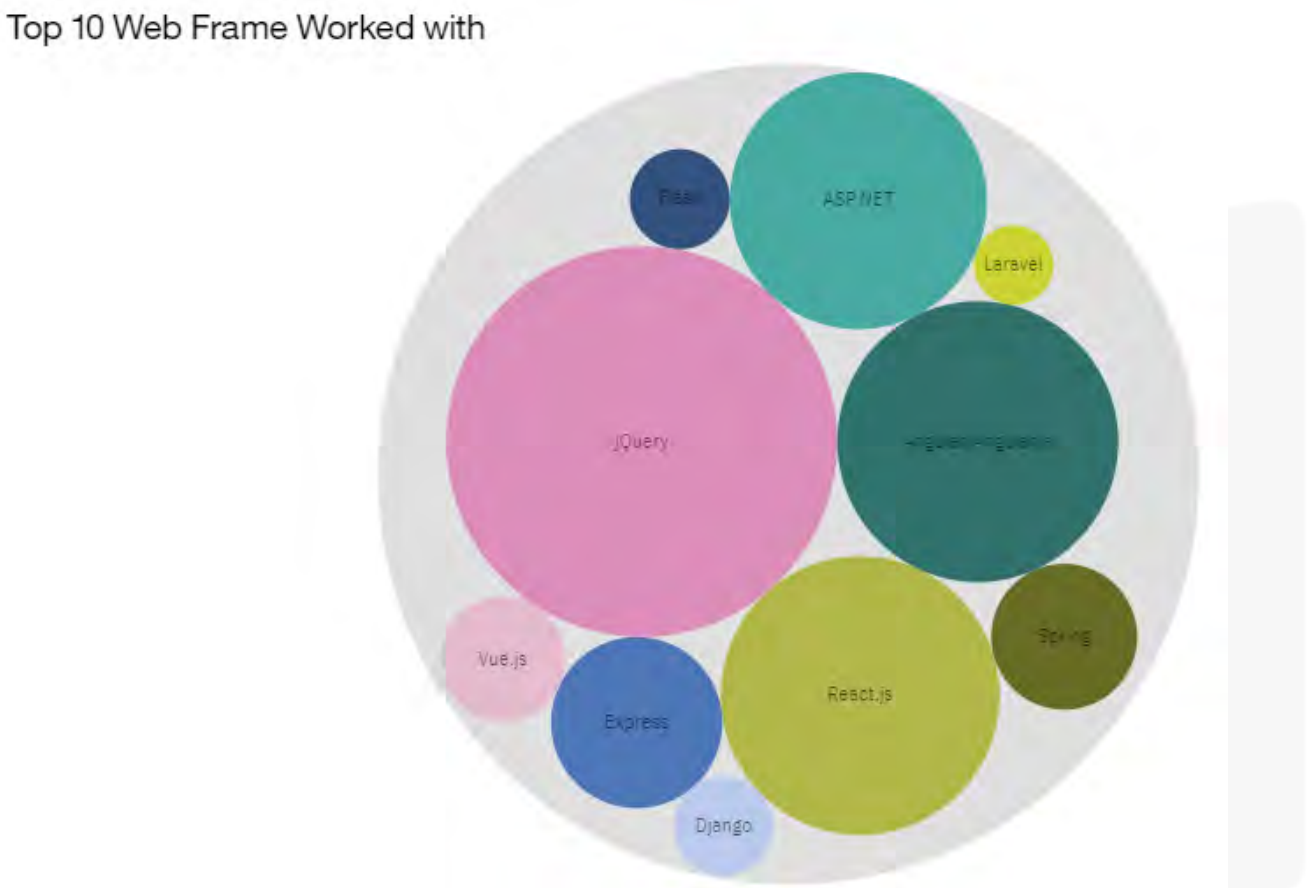
SKILLS NETWORK

DASHBOARD TAB 1 Part 2/4

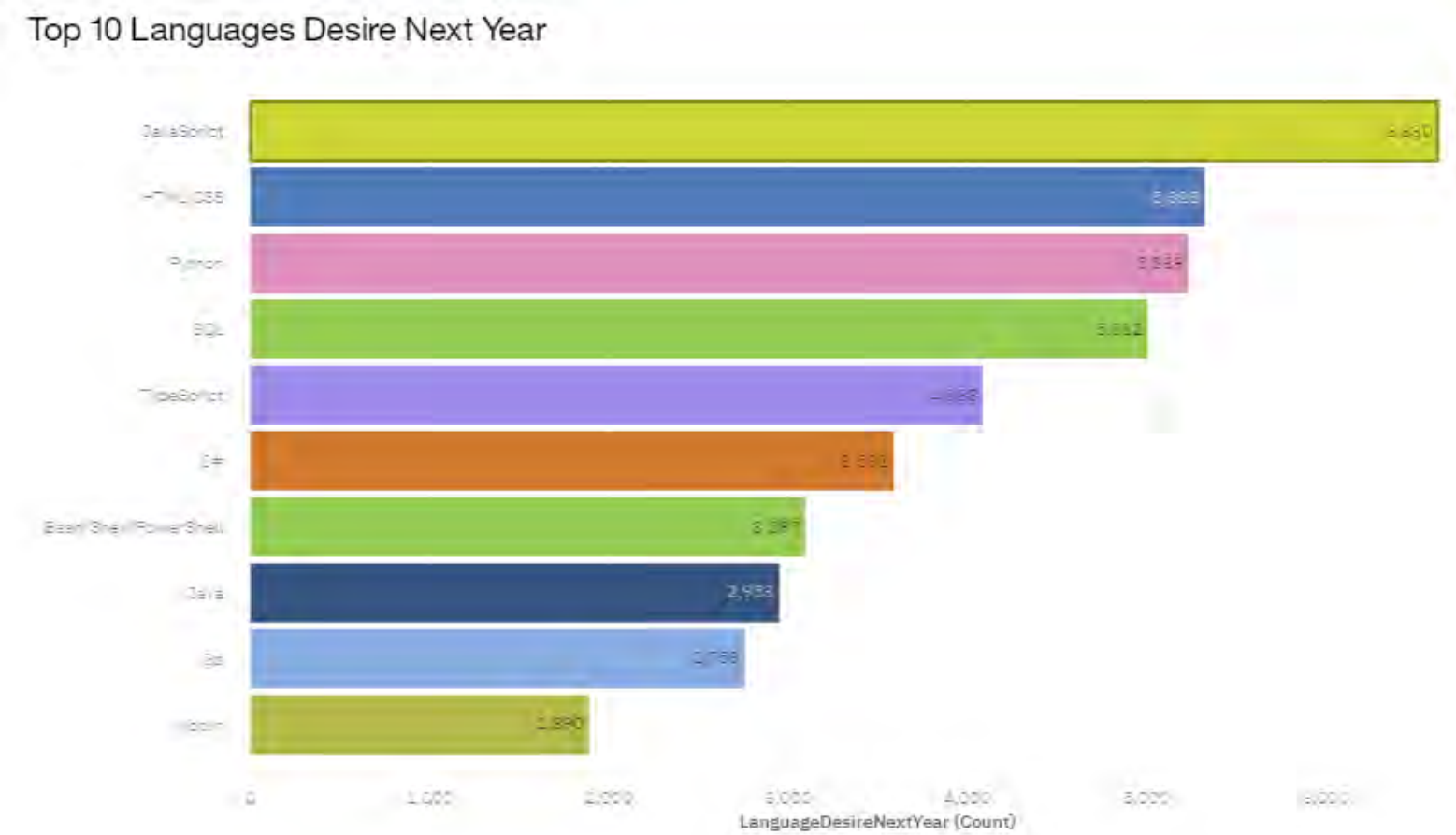


DASHBOARD TAB 1 Part 3/4

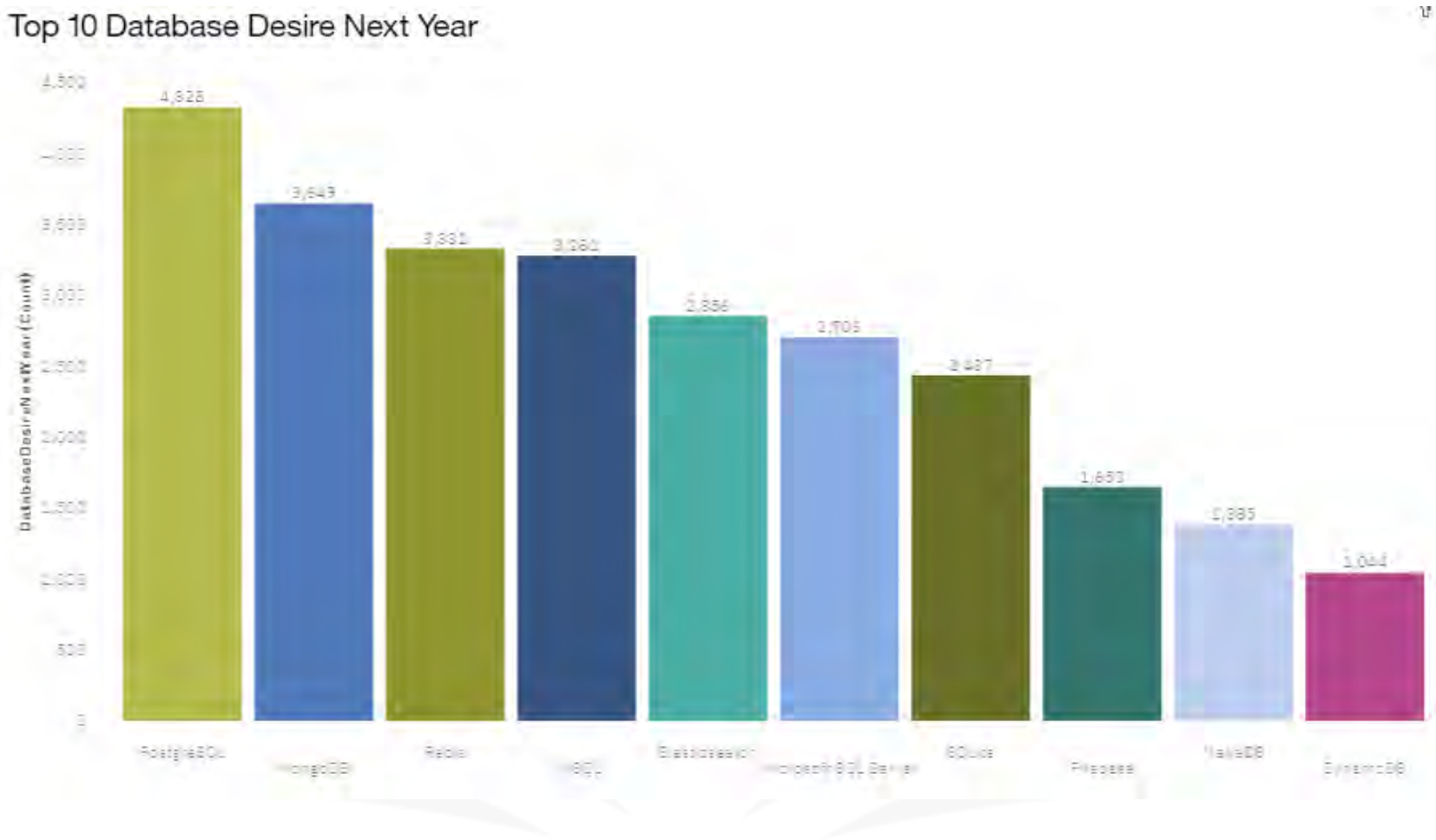
DASHBOARD TAB 1 Part 4/4



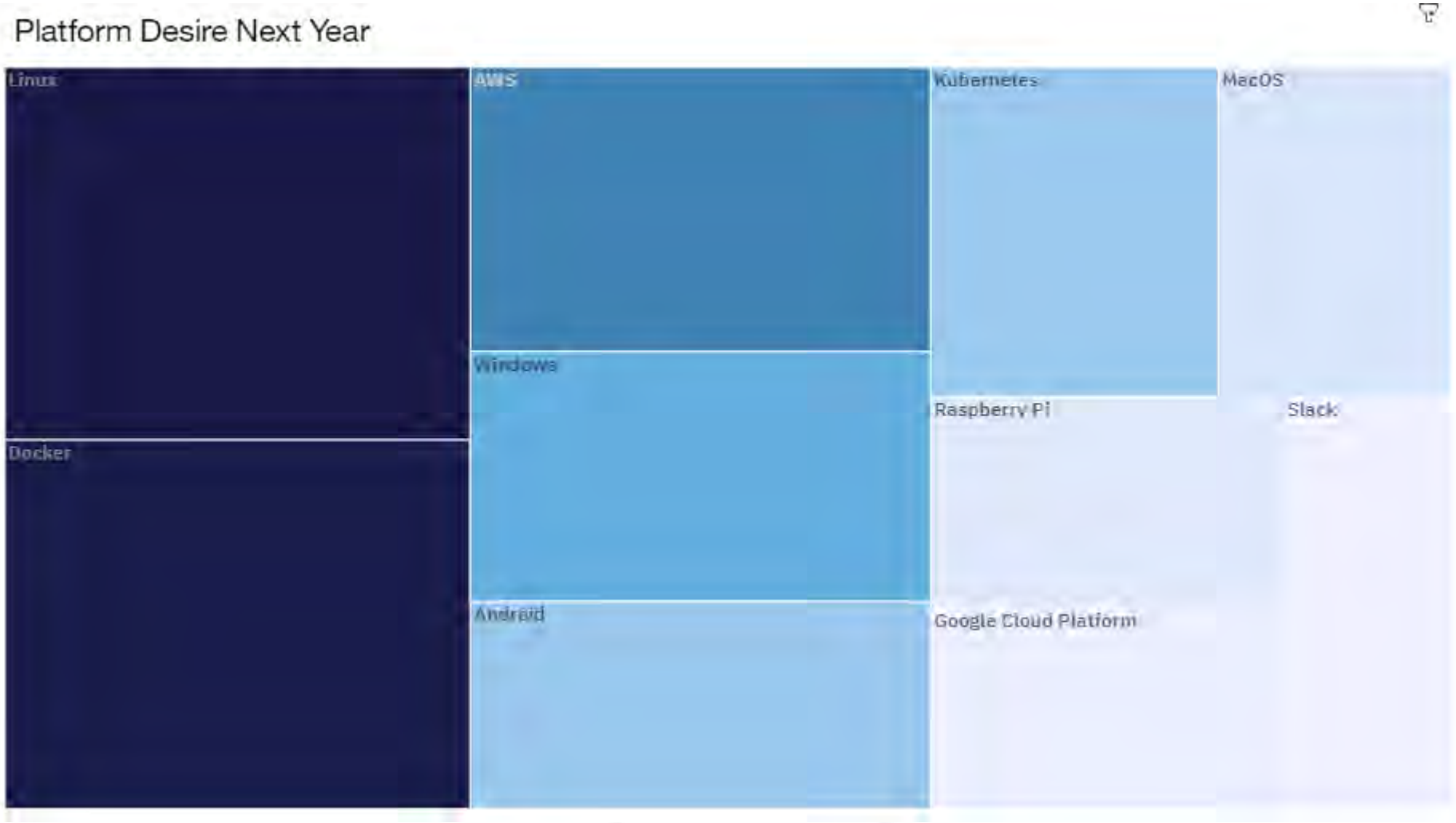
DASHBOARD TAB 2 Part 1/4



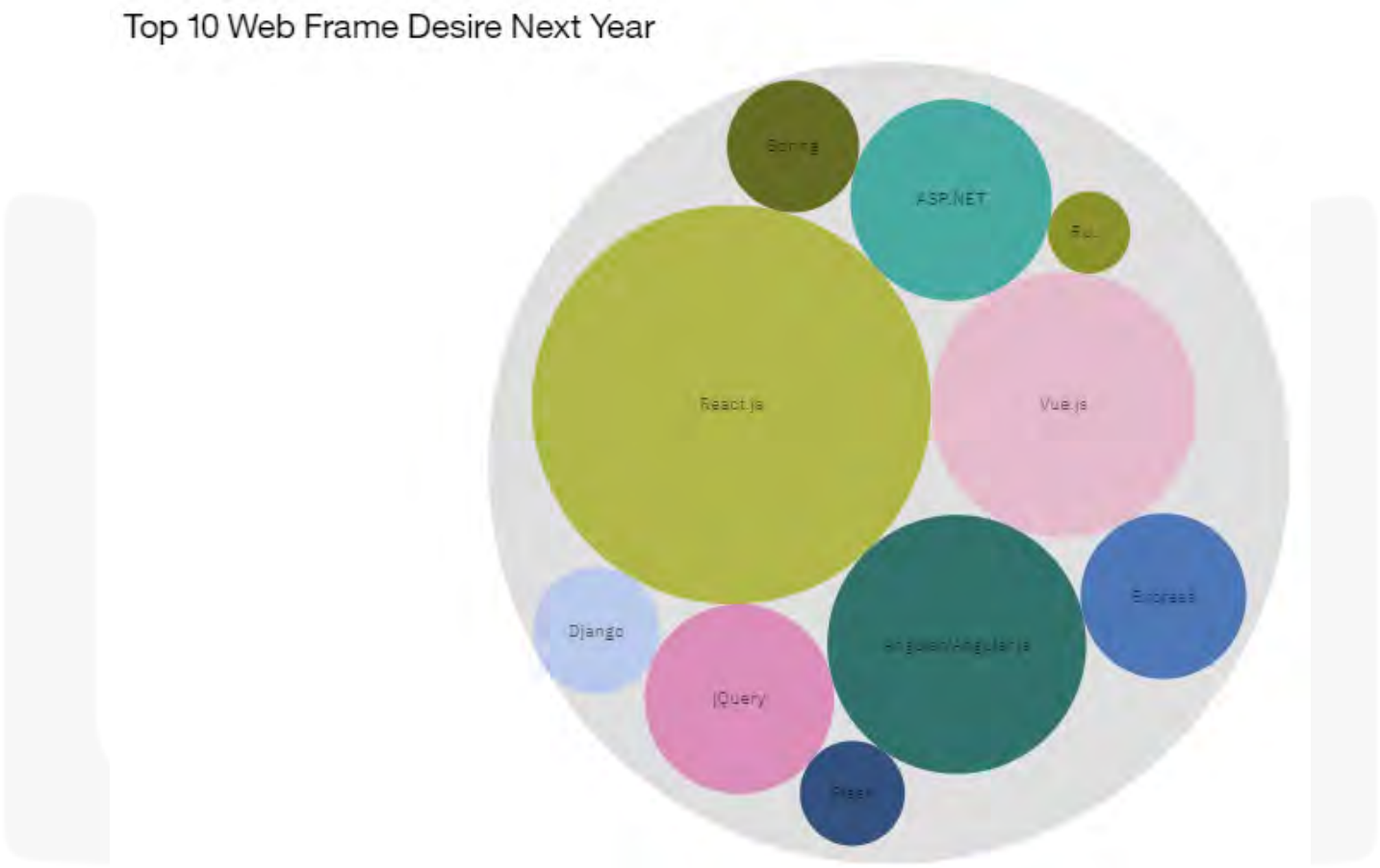
DASHBOARD TAB 2 Part 2/4



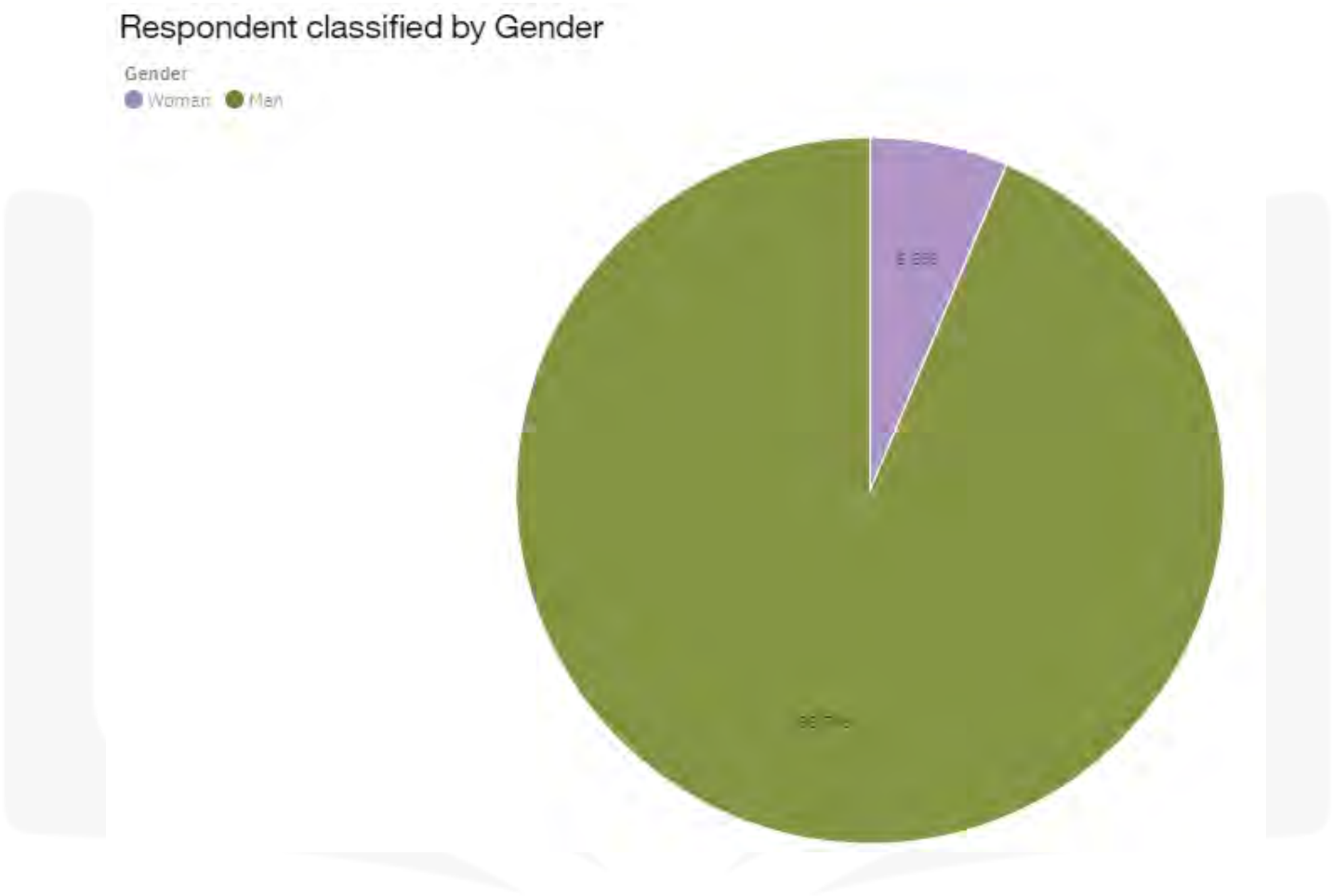
DASHBOARD TAB 2 Part 3/4



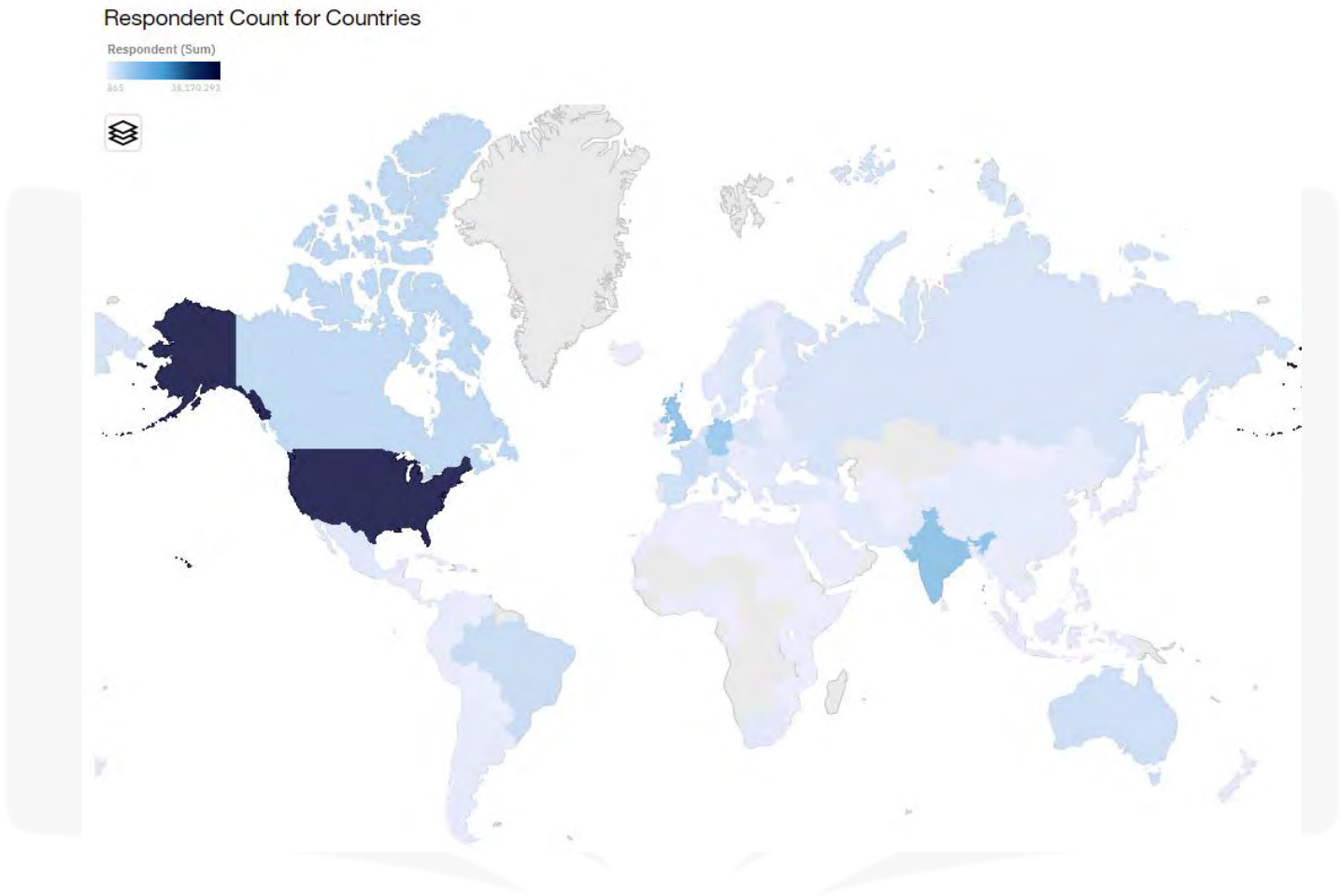
DASHBOARD TAB 2 Part 4/4



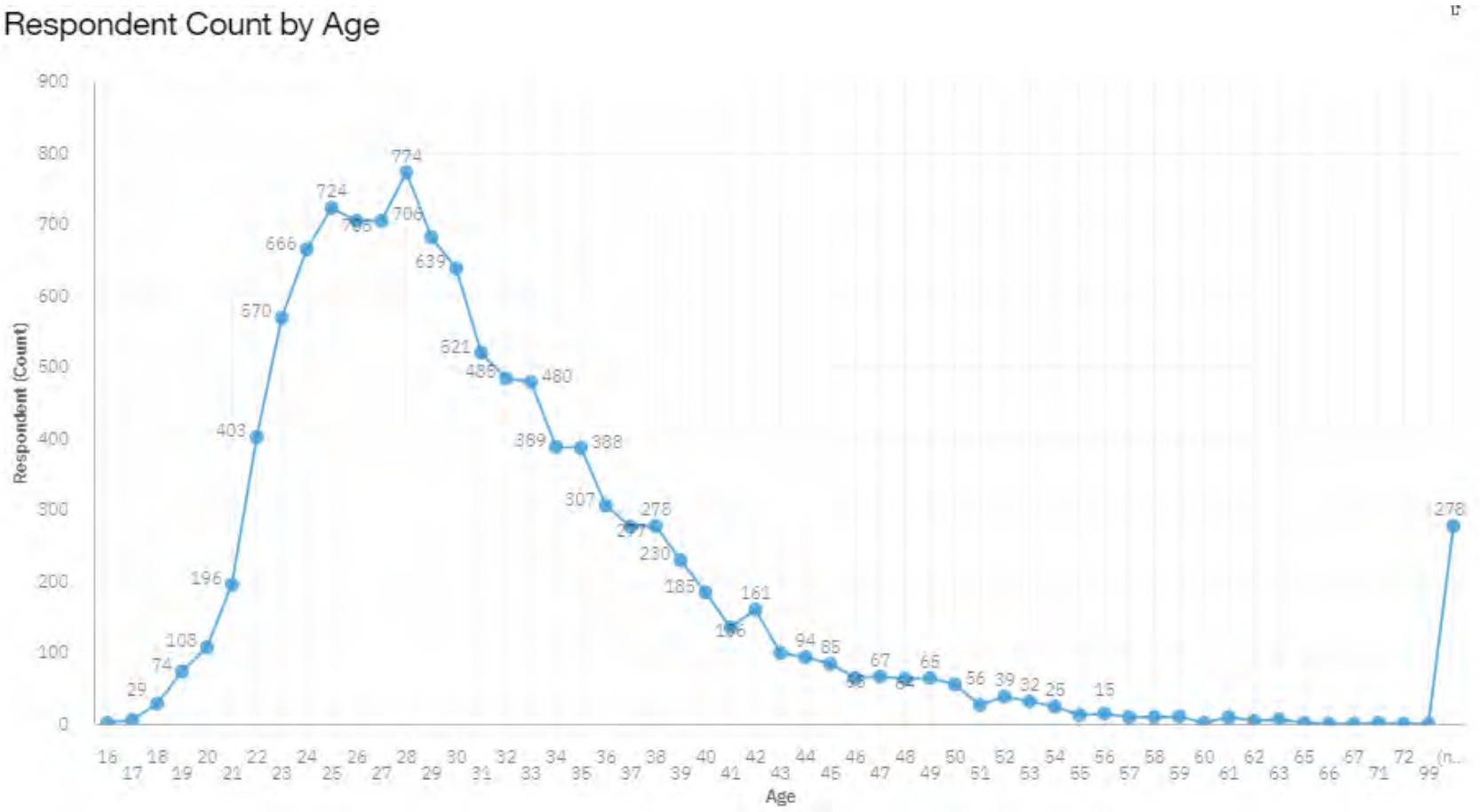
DASHBOARD TAB 3 Part 1/ 4



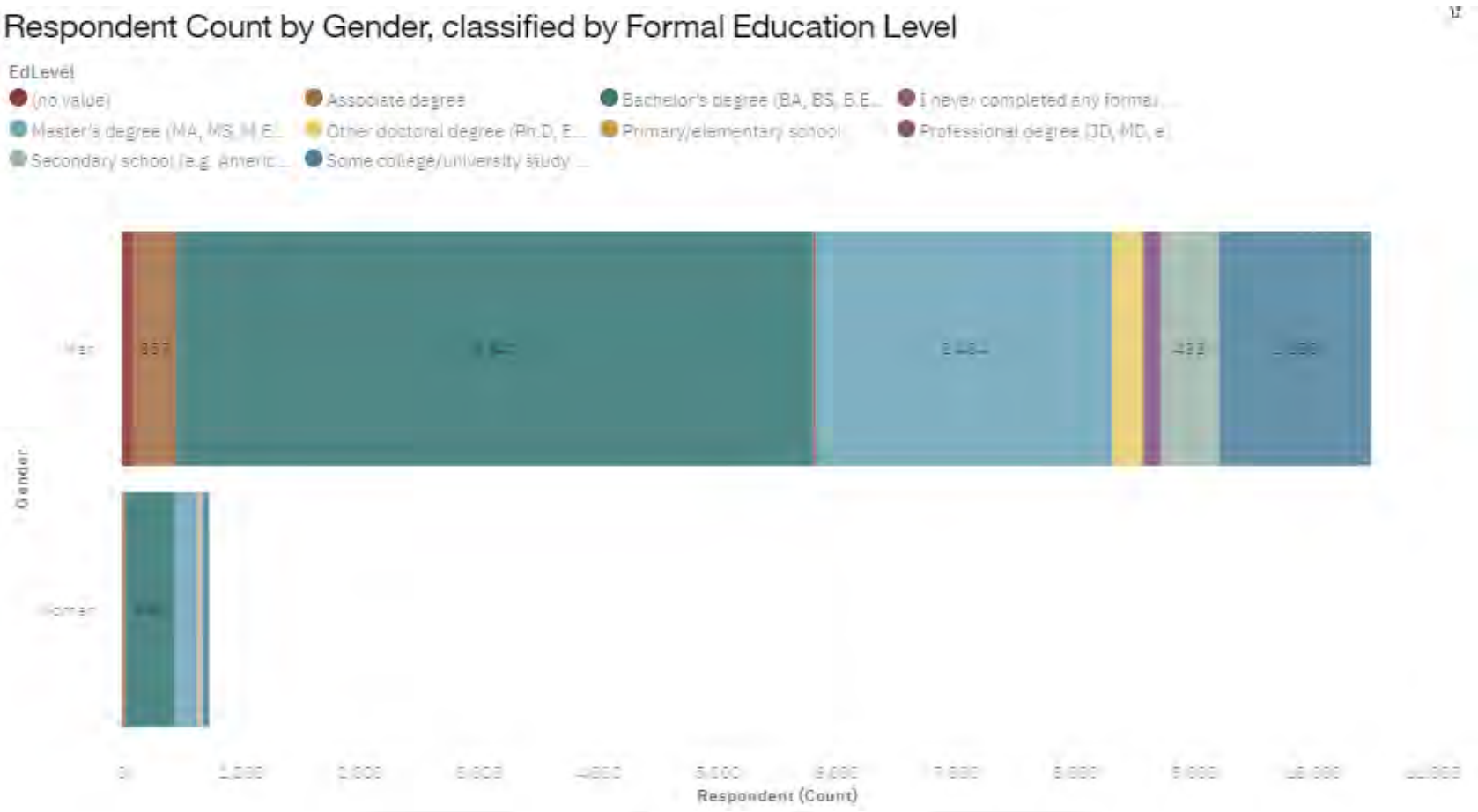
DASHBOARD TAB 3 Part 2/ 4



DASHBOARD TAB 3 Part 3/ 4



DASHBOARD TAB 3 Part 4/ 4



DISCUSSION

- The findings are mostly based on Respondents' answers in surveys.
- The current trend can be trusted, as the results are based on the technologies they work on.
- But the popular technologies for next year are based on their desire which won't necessarily reflect the real trend
- One suggestion is to consider the technologies which are searched for in job postings and the annual average salaries as well

OVERALL FINDINGS & IMPLICATIONS Part 1/2

Findings

- On top programming languages for both the current year and the next year we can find JavaScript, HTML/CSS, SQL, Python, with the last one having an increased trend
- MySQL, PostgreSQL and MongoDB are found in top 5 database trends for current and next year, while technologies not being on top in the current year seem to have an increased trend, like Redis and Elasticsearch

Implications

- Company shall concentrate the developing of JavaScript, HTML/CSS, SQL and Python skill, with special focus on the latter one, which has an increased popularity
- The top databases on which company shall develop its skills are MySQL, PostgreSQL and MongoDB, followed by technologies which increase in popularity at a fast pace, like Redis and Elasticsearch

OVERALL FINDINGS & IMPLICATIONS Part 2/2

Findings

- Top platforms for current year and next year are Windows, AWS, Linux, and Docker, with the 2 latter's popularity increasing.
- The preferred Web Frames are React.js, Vue.js, Angular/ Angular.js and ASP.NET

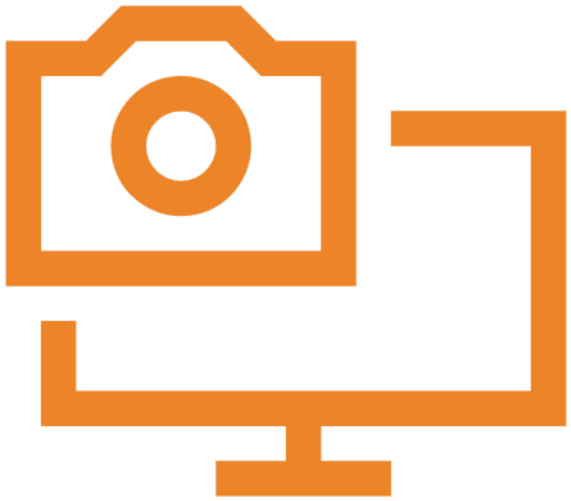
Implications

- Windows and Linux are the most popular platforms in the current year, but Company shall also focus on the platforms with increased popularity trends for next year, like AWS and Docker
- Company's web frames skills shall focus on the first 4 popular web frames which are by far the most popular ones.

CONCLUSION

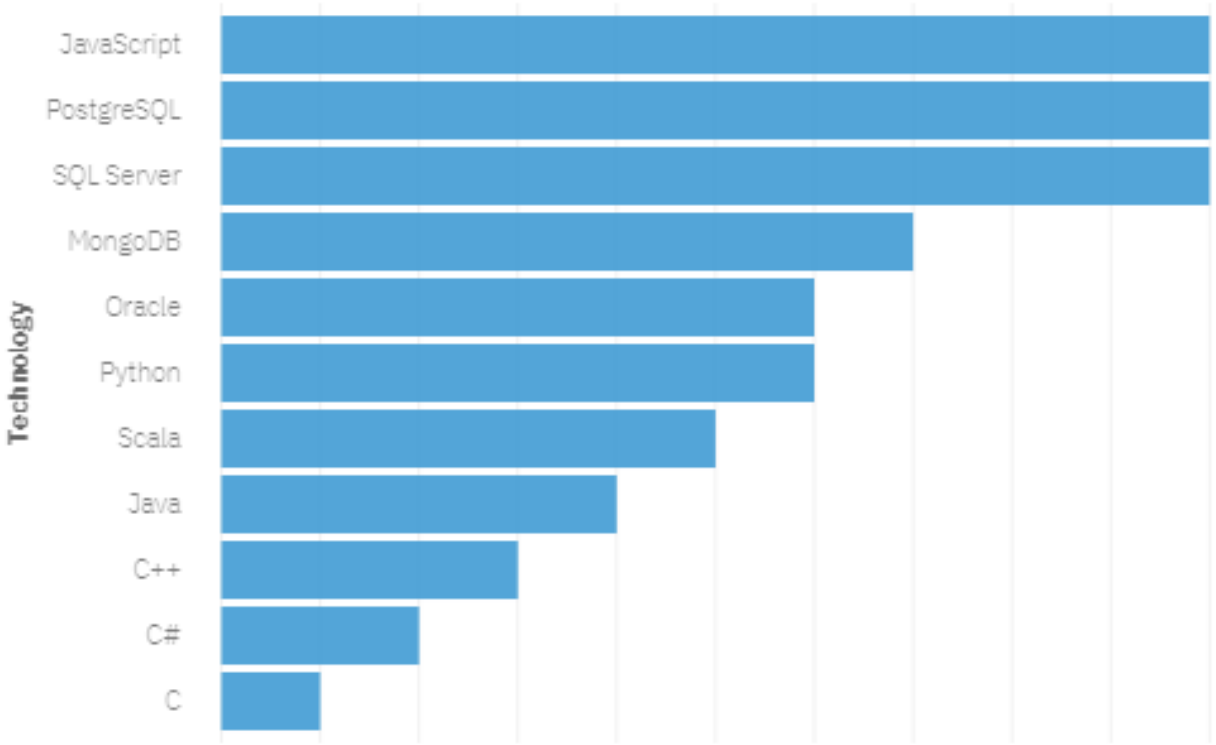
- Datasets used in this project are consistent and relevant for this study.
- Top programming languages in demand are: JavaScript, HTML/CSS, SQL and Python
- Top database skills in demand currently and in future are MySQL, PostgreSQL and MongoDB, with Redis and Elasticsearch being foreseen to evolve in future
- Preferred platforms are Windows, AWS, Linux, and Docker, with the former one having a decreasing trend for next year

APPENDIX



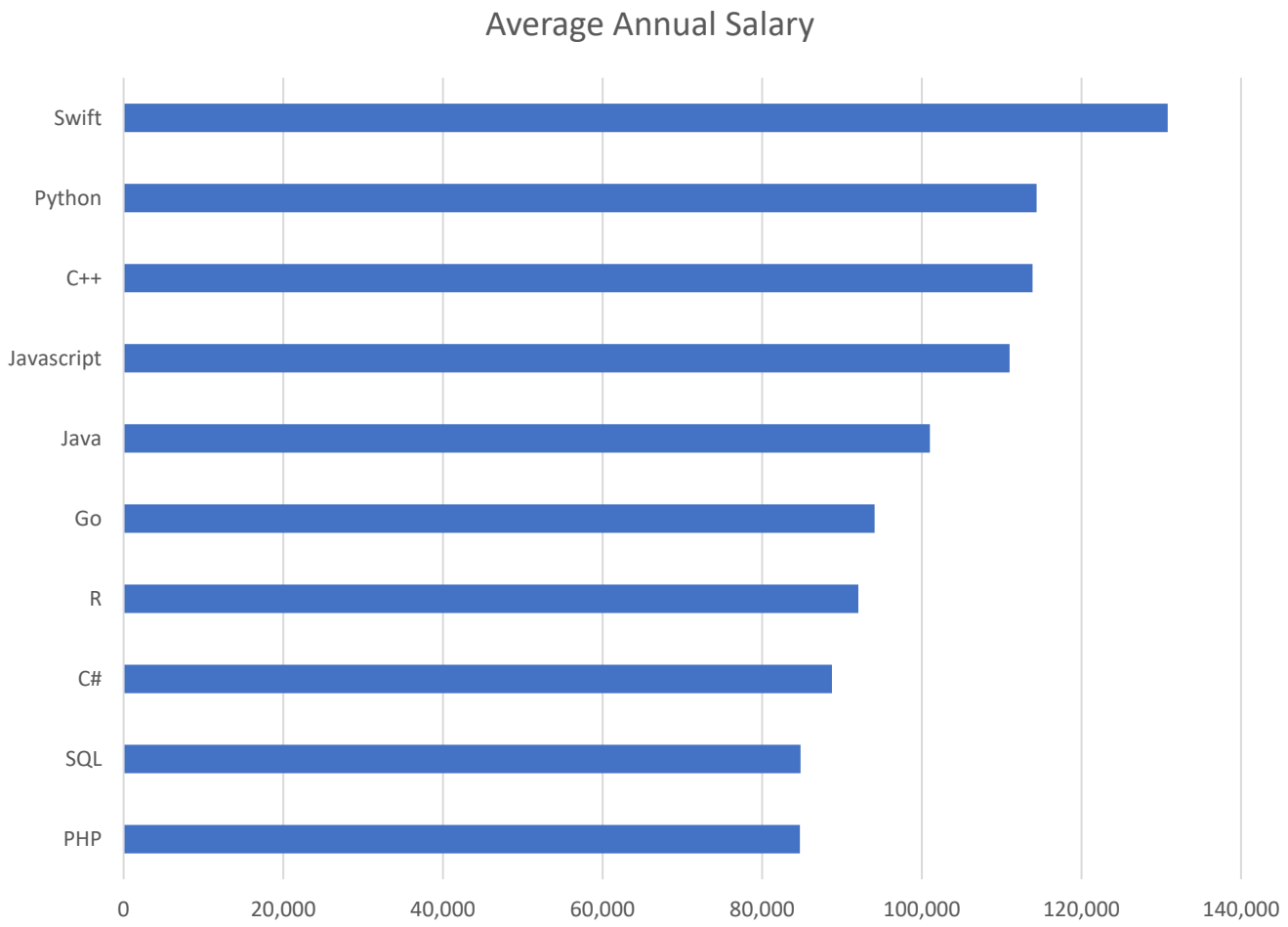
- The popularity of software has been presented in terms of Respondent's preference mostly, but let's see on the next slides 2 graphics showing:
  - The top programming languages and databases from job postings
  - The jobs with the highest average annual salaries.

GITHUB JOB POSTINGS



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