IBM Data Analyst Capstone Project

Reporter: Sheldon

Date: 2023-4-11

OUTLINE

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

EXECUTIVE SUMMARY

- This project aims to learn about the rapidly changing IT world from a tool perspective, in order to stay competitive.
- This project uses a subset of the 2019 StackOverflow online survey dataset.
- Analysis from several aspects:
  - Exploratory Data Analysis
  - Data Visualisation
  - Building A Dashboard
- Discuss both the trends and outliers.
- Conclusion provides a deeper view of the current IT market.

INTRODUCTION

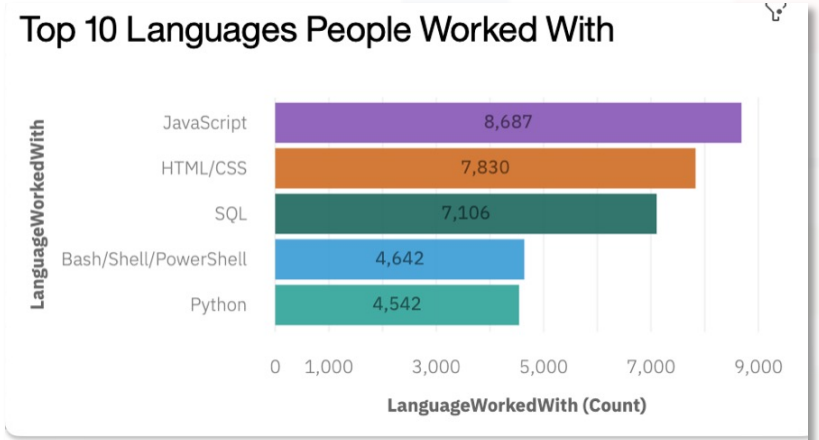
- Nature of the analysis:
  The goal of this project is getting insights in programmer perference by data analysis and data visualisation.
- Problem Statement:
  The IT industry is growing rapidly, so it is significant need to become familiar with the current trends.
- Questions for analysis:
  - Top 5 programming languages for current year and future year.
  - Top 10 databases for current year and future year.
  - Recognize the overall trends.

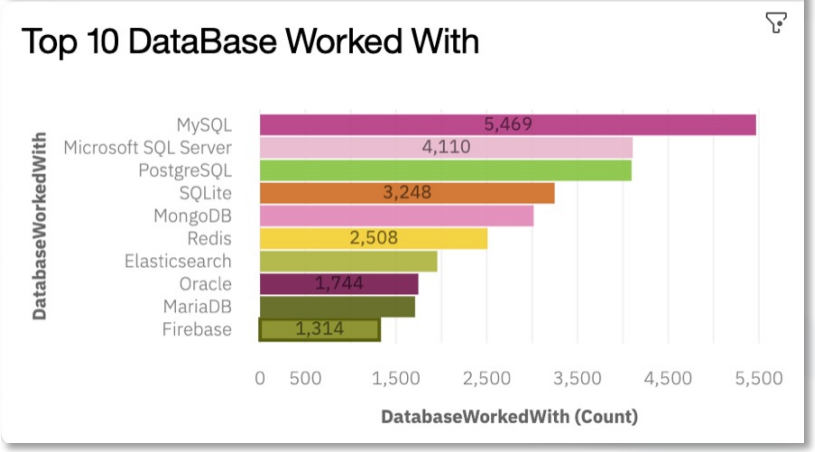
METHODOLOGY

- The dataset includes commonly used programming languages, databases, and platforms among programmers, as well as their desired languages for future learning..
- The data was collected through the website's API.
- Various charts will be generated to display trends.
  Bar chart
  Pie chart
  Line chart etc.

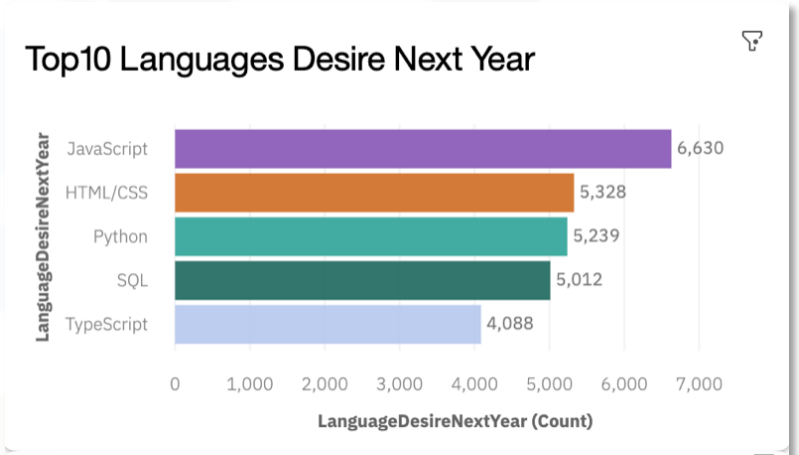
RESULTS

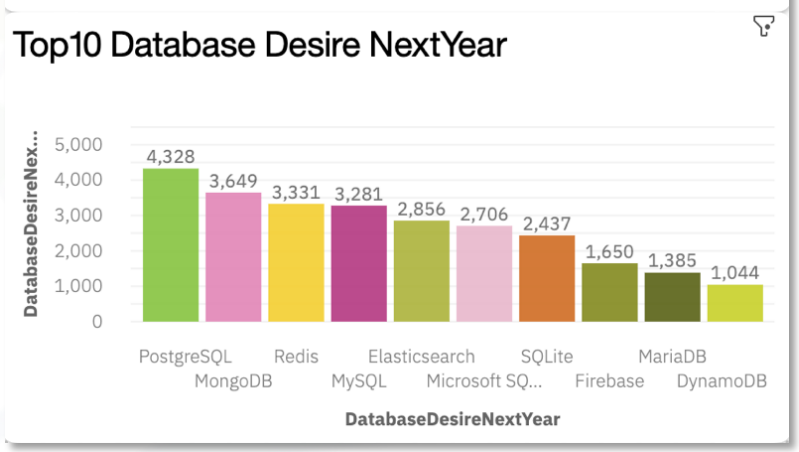
Top 5 programming languages and top 10 databases for current year.





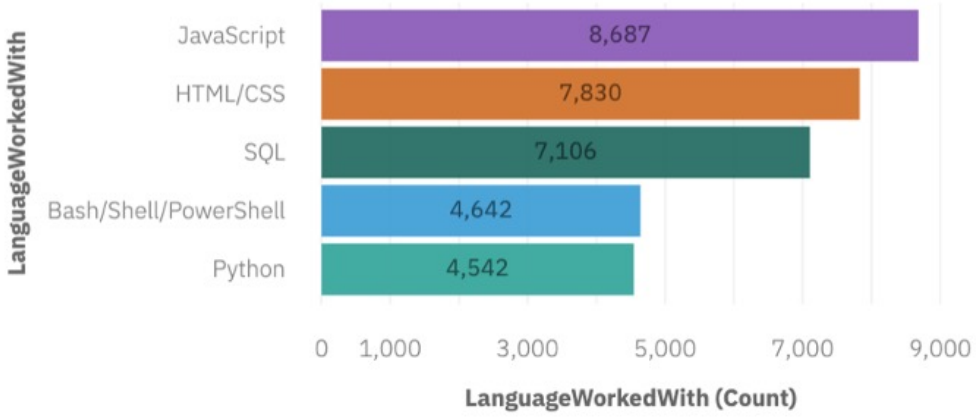
Top 5 programming languages top 10 databases for current year and future year



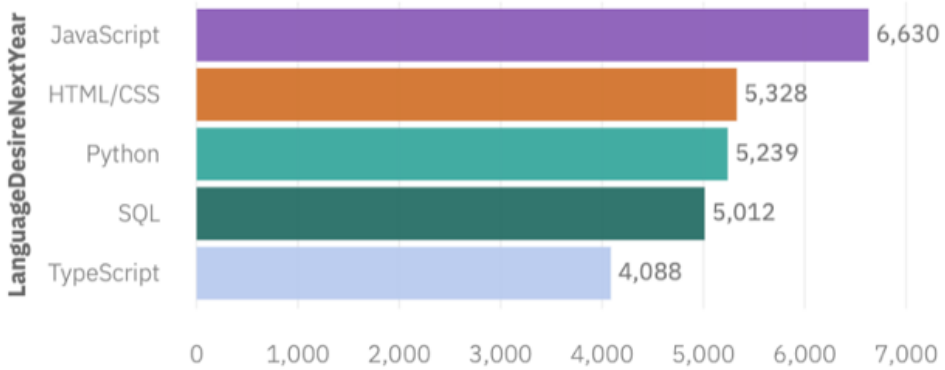


PROGRAMMING LANGUAGE TRENDS

Current Year



Next Year



PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

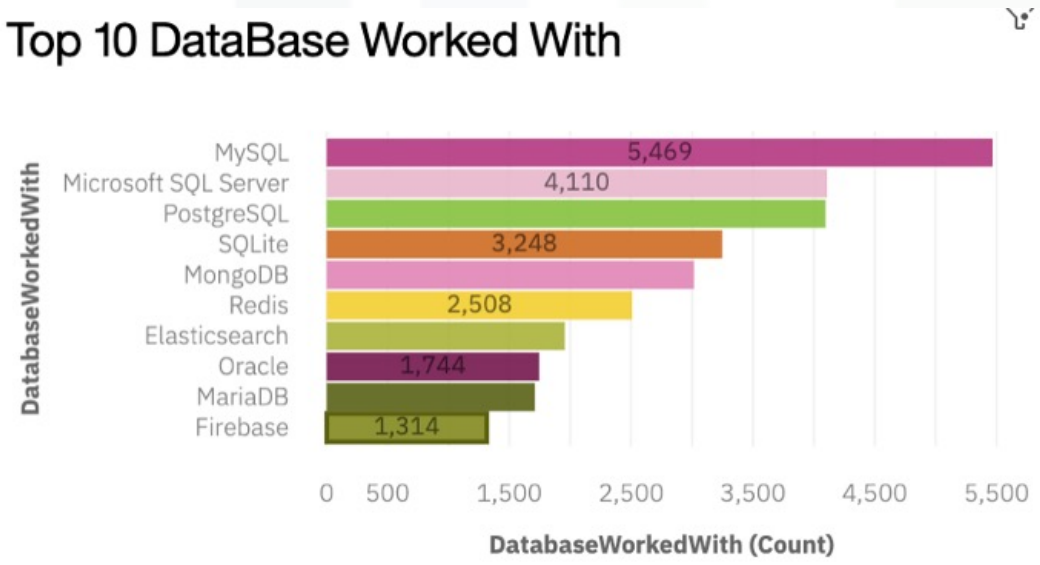
- The top five programming languages for current and future years are mostly the same, with Typescript replacing Shell in the future.
- Python will rise from the bottom to the third position.

Implications

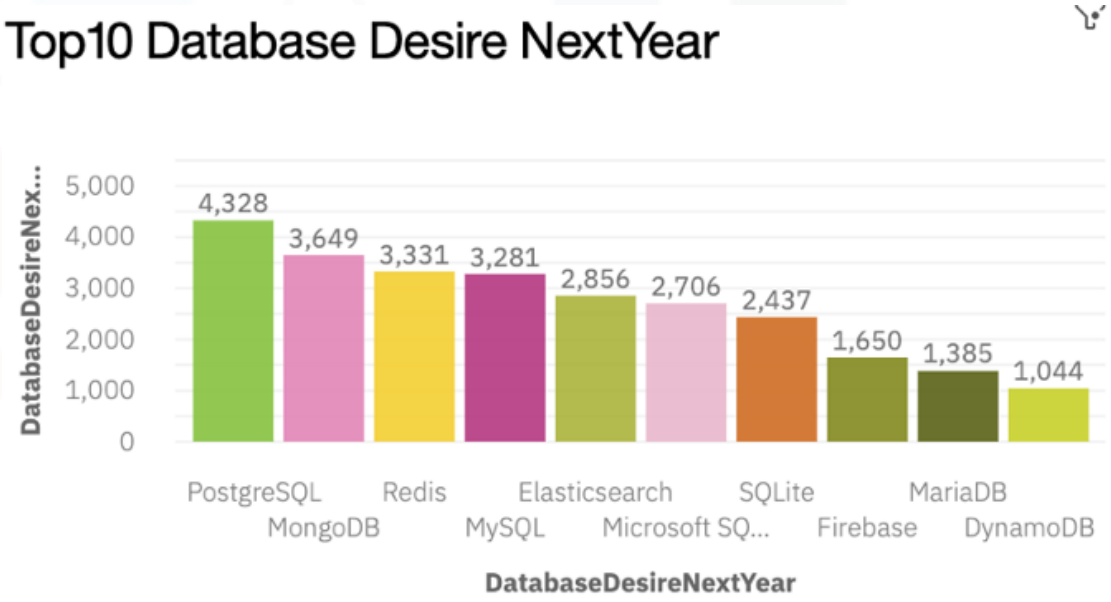
- The company should focus on maintaining its business in JavaScript, HTML/CSS, Python, and SQL.
- The company should consider expanding its business in Python and Typescript while reducing its reliance on Shell.

DATABASE TRENDS

Current Year



Next Year



DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- Oracle will drop out of the top 10 most popular databases next year, and DynamoDB will take its place.
- The top two databases have changed from MySQL and Microsoft SQL Server to PostgreSQL and MongoDB.

Implications

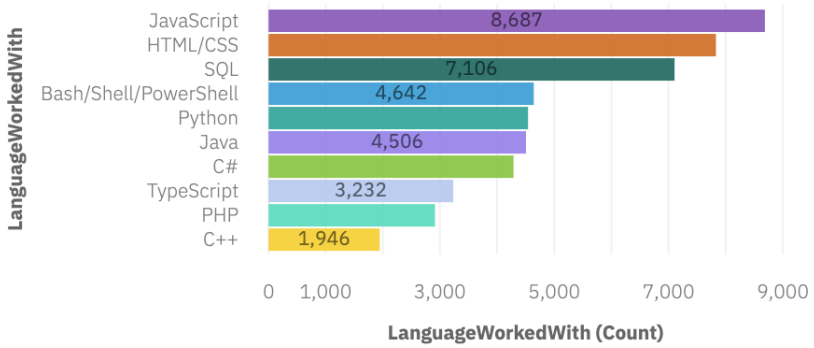
- The company should gradually reduce its reliance on MySQL, Microsoft SQL Server, and Oracle.
- The company should consider increasing its use of PostgreSQL, MongoDB, Redis, and SQLite.

DASHBOARD

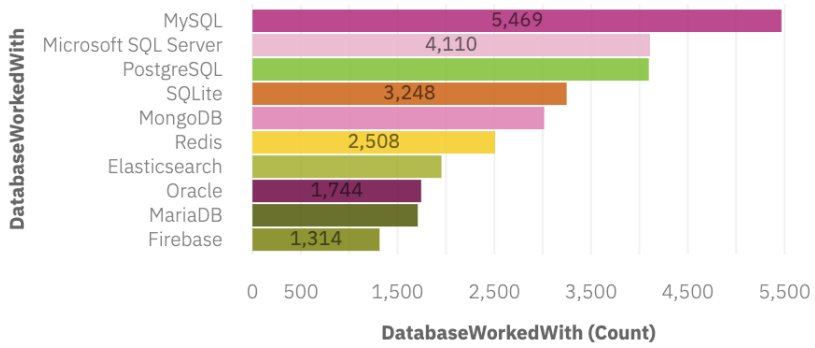
https://eu-gb.dataplatform.cloud.ibm.com/dashboards/65d5735e-f3f8-464d-af0b-882746caa14e/view/0115d70312ea75fe09b5eee4079d2b037465700bb6bbd052d5877b490e352497a96d1092c8284a59dc130663fbeb115bca

DASHBOARD TAB 1: Current Technology Usage

Top 10 Languages People Worked With

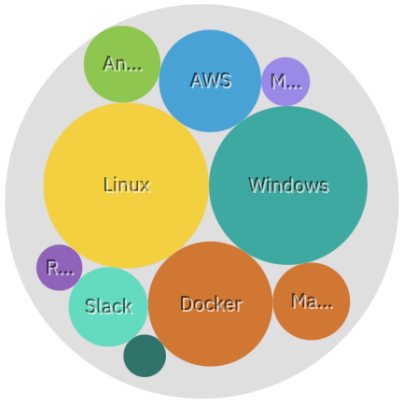


Top 10 DataBase Worked With



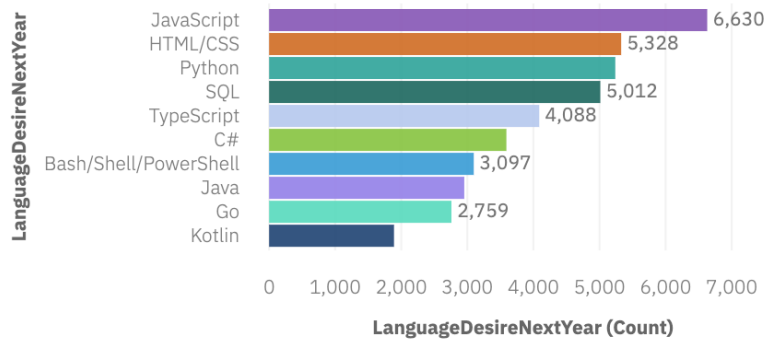
Platform Worked With

Top10 WebFrame Worked With

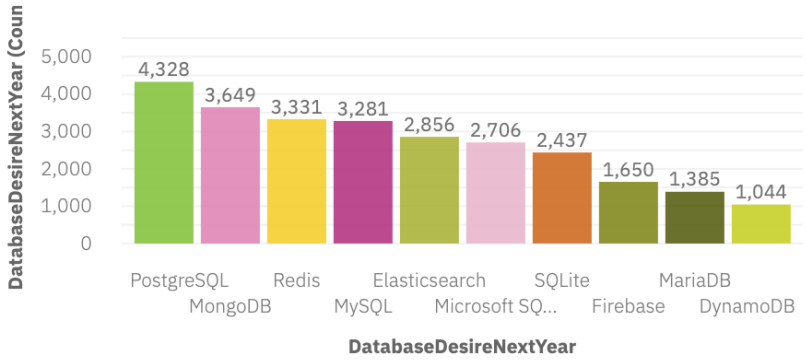


DASHBOARD TAB 2: Future Technology Usage

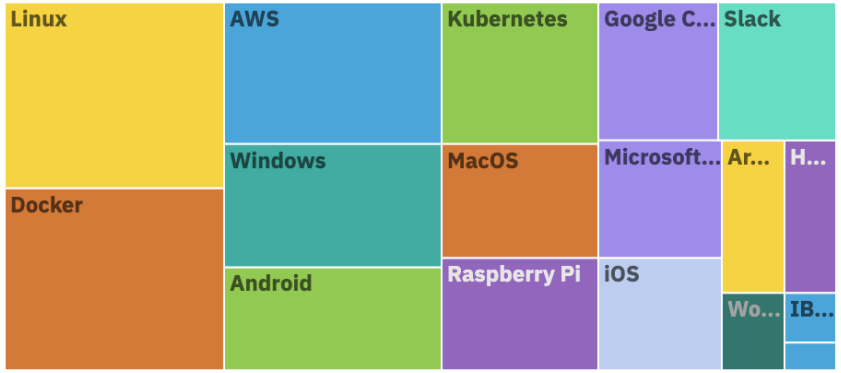
Top10 Languages Desire Next Year



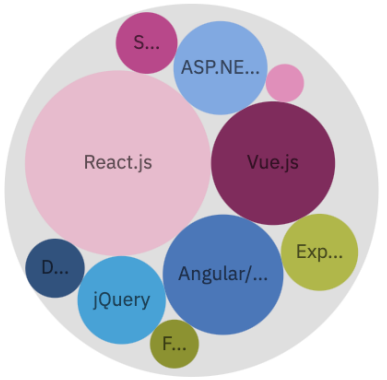
Top10 Database Desire NextYear



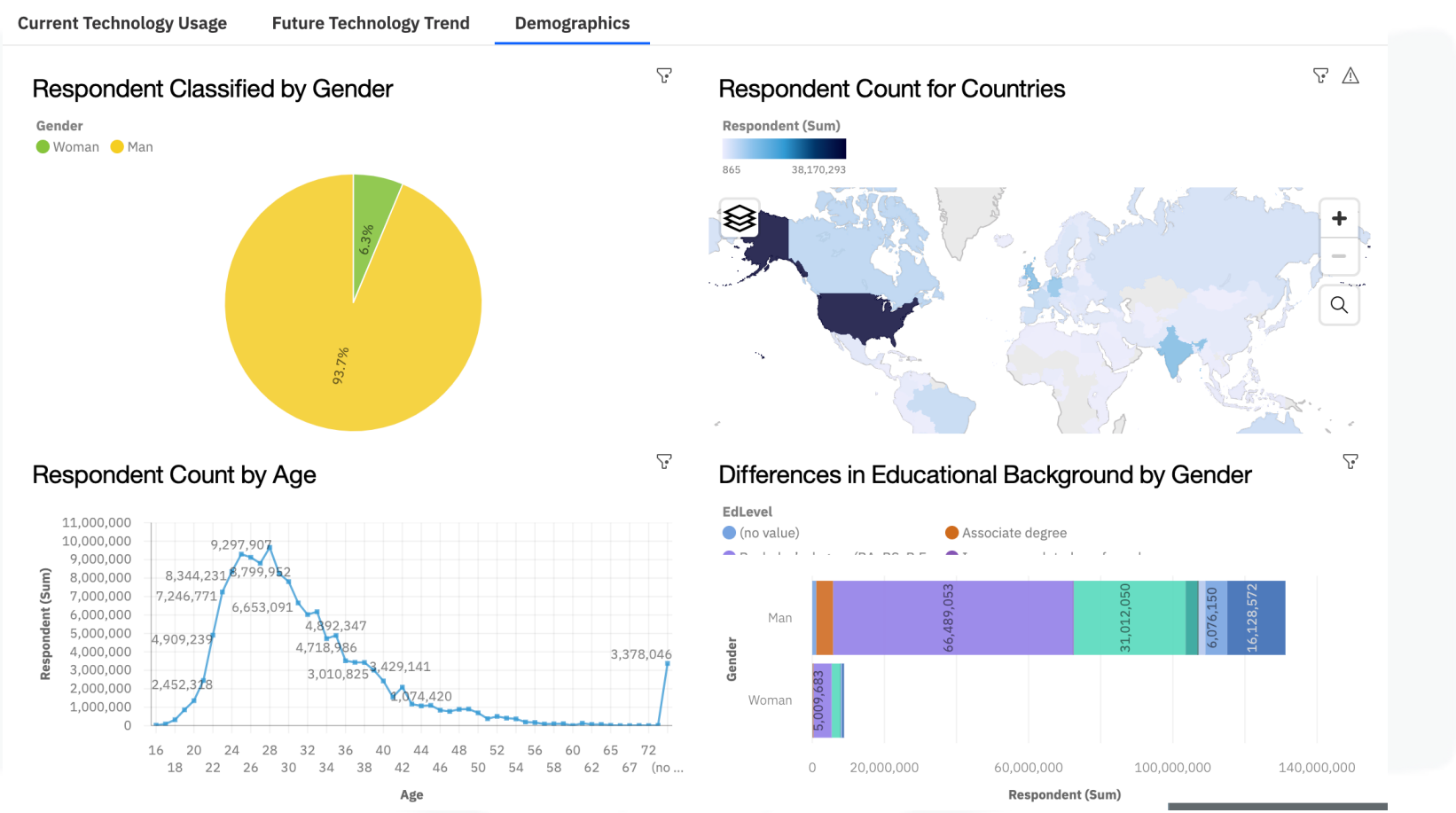
Platform Desire NextYear



Top 10 Frame Desire NextYear



DASHBOARD TAB 3: Demographics



DISCUSSION

- Aside from programming languages and databases, patterns were also observed in platforms, web frameworks, gender, age, countries, and educational background.
- Docker usage is increasing, while React.js, Vue.js, and Angular.js are the top three most popular web frameworks.
- The majority of programmers are aged between 22 to 32, and most hold either a Master's degree or no formal education.
- The United States has the largest number of programmers, and there are more male programmers than female programmers.

OVERALL FINDINGS & IMPLICATIONS

Findings

1. Popular programming languages include JavaScript, HTML/CSS, and Python, with emerging interest in Typescript. PostgreSQL and MongoDB are the most commonly used databases, with decreasing reliance on MySQL, Microsoft SQL Server, and Oracle.
2. Docker and React.js, Vue.js, and Angular.js are popular platform and web framework choices. The majority of programmers are aged 22-32, hold a Master's degree or no formal education, and are predominantly male.
3. The United States has the most programmers globally.

Implications

1. Companies should focus on maintaining business in popular programming languages while exploring Typescript, and increase use of PostgreSQL and MongoDB while reducing reliance on MySQL, Microsoft SQL Server, and Oracle.
2. Companies should adopt Docker and popular web frameworks, aim to attract younger programmers with a Master's degree or no formal education, and strive for a more balanced gender ratio in their programmer workforce.
3. Companies should consider the potential market in the United States.

CONCLUSION

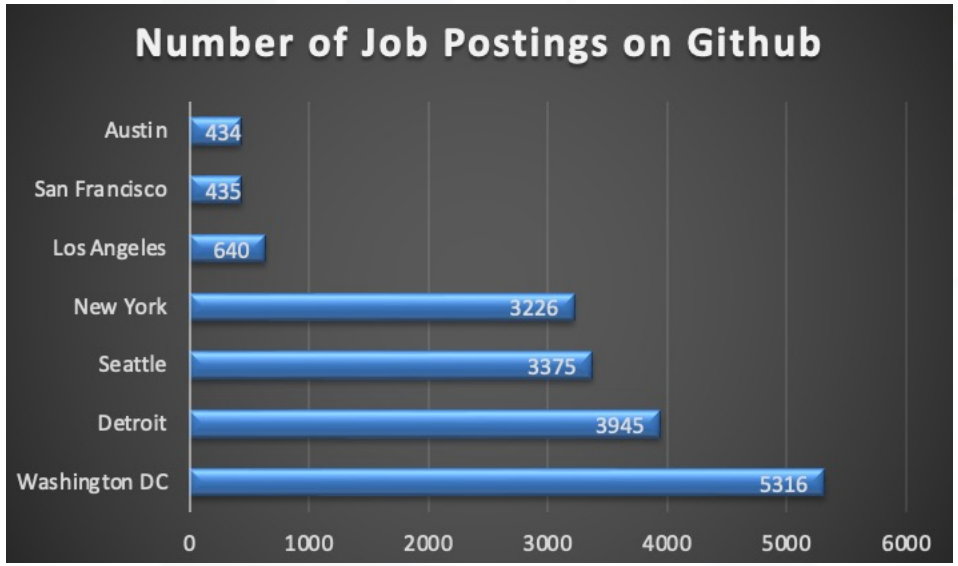
The data analysis highlights clear trends and patterns among programmers. Companies should focus on popular languages, explore emerging languages, increase use of PostgreSQL and MongoDB, adopt Docker and popular web frameworks, strive for a balanced gender ratio, and consider the potential market in the United States to adapt to the changing programming landscape and position themselves for success.

APPENDIX

Thanks for the dataset given from Stack overflow.

JOB POSTINGS

Job posting data using Github Job API :



POPULAR LANGUAGES

Collected the job postings data using web :