Analyzing stack overflow developer survey

Nimanthi Yaseema

September 02, 2023

OUTLINE

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

EXECUTIVE SUMMARY

- In this project, we analyze data to help identify future skill requirements in the IT field.
- We have two data sets based on technologies and demographics. Using the technology data, we analyze to identify present and future trends of technology. From demographics data, we identify technology insights based on gender, age, locations and education levels.
- We visualize our results using IBM Cognos Dashboard Embedded platform and created three different dashboard tabs, i.e., current data technologies, future data technologies and demographics of respondents.

INTRODUCTION

- To Visualize the data ,dashboards were created to emphasis on the present and future trends of data technologies.
- Through creating dashboards using IBM Cognos Dashboard, we would be able to visualize our data based on top 10 present and future data technology trends based on respondent gender, formal education ,age and countries. To Visualize the data ,dashboards were created to emphasis on the present and future trends of data technologies.
- Through creating dashboards using IBM Cognos Dashboard, we would be able to visualize our data based on top 10 present and future data technology trends based on respondent gender, formal education ,age and countries.

•Exploring Data

METHODOLOGY

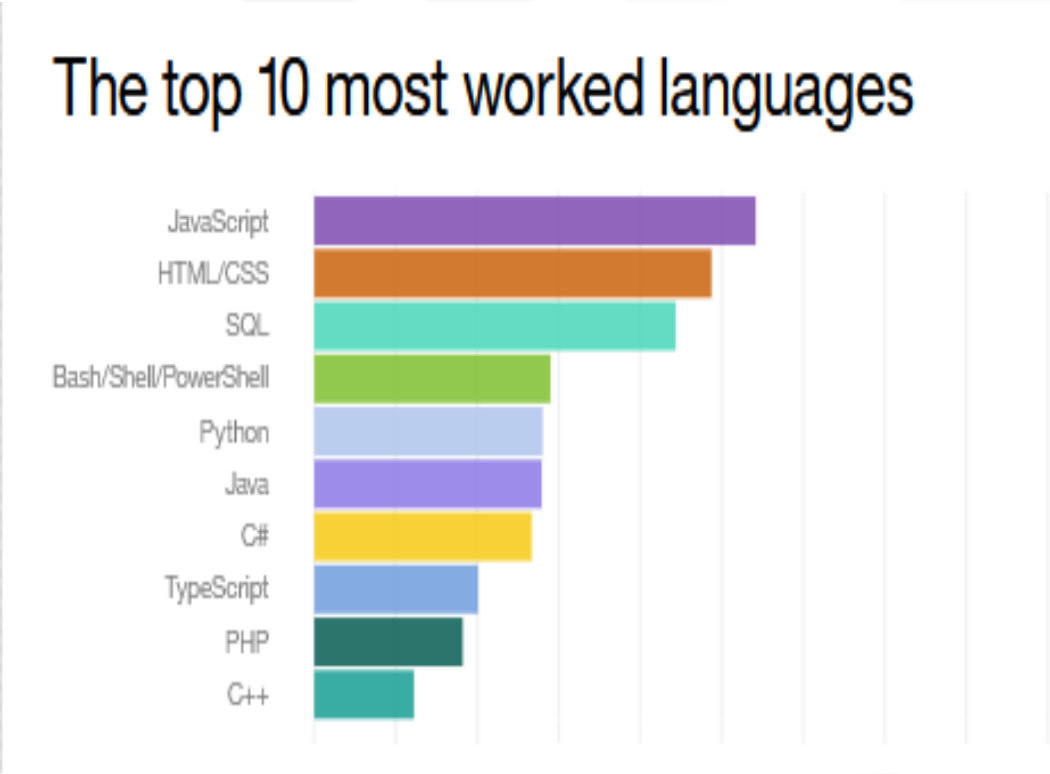
- We collected data using APIs and Webscraping methods. After data wrangling process, we analyze the data and determine distribution, outliers and correlation.
- A modified subset of Stack Overflow data consisting of two data sets based on the demographics and data technologies were used to visualize the present and future data trends based on their demographics.
- IBM Cognos Analytics was used to create different dashboards.

RESULTS

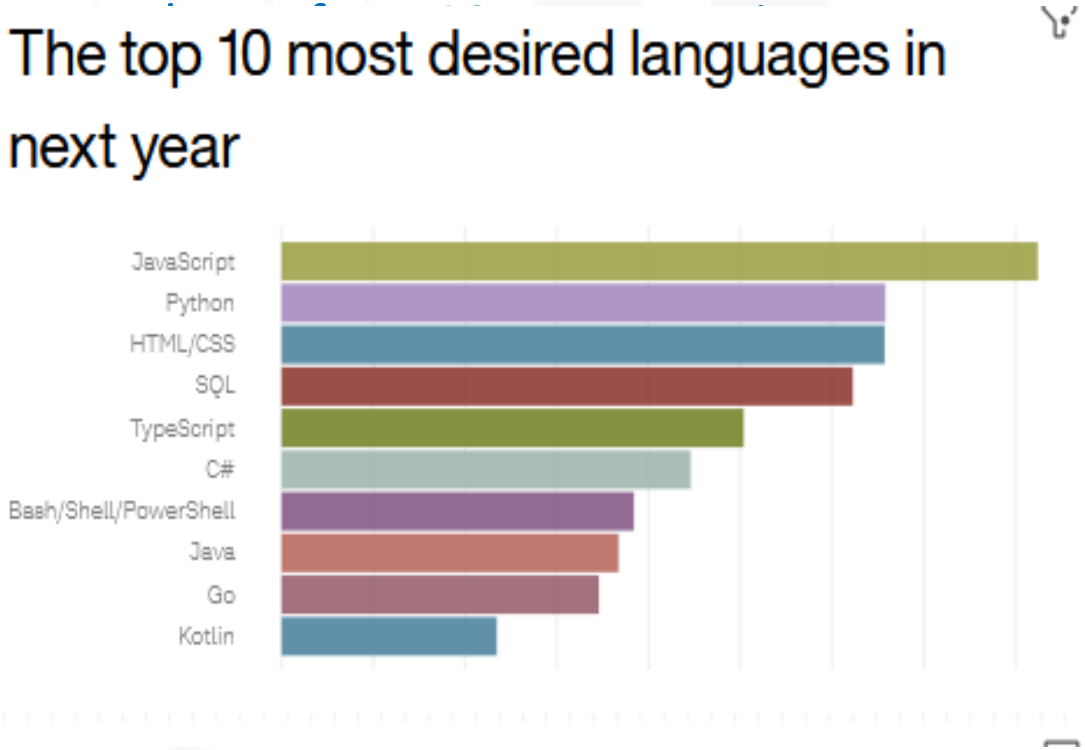
- The data was collected from stack overflow blog under Open Database License .
- It consisted of two data sets :
    - Survey data technology normalized consisted of approximately 75K respondents querying them about the present and future data technologies interest . Then using visualization tools such as bar, columns graphs along with tree maps and hierarchy trees were used to analysis the top trending present and future technologies trends.
    - Survey data demographics was more clustered data set a which involved around 11K respondents answering about various demographics questions such as country , age, education level, ethnicity, gender etc. The results of these finding were visualized in form of bar ,line charts along with maps.

PROGRAMMING LANGUAGE TRENDS

Current Year



Next Year



PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

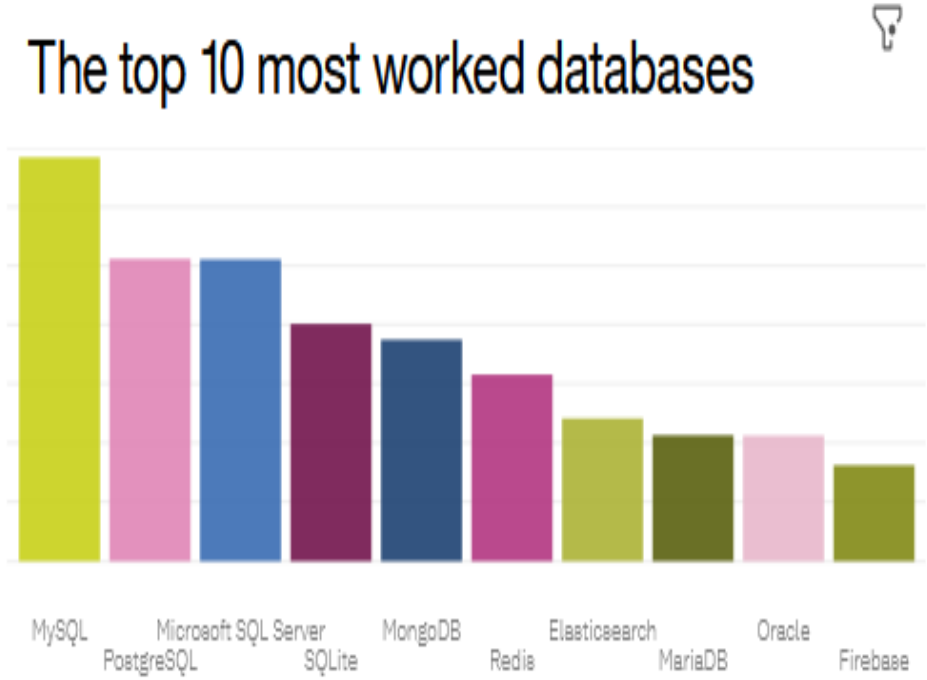
- JavaScript and HTML were the languages that were preferred and will continue to dominate in future too.
- They were further followed by SQL, Bash, and Python for current trends, but they may not be the next go to languages in future.

Implications

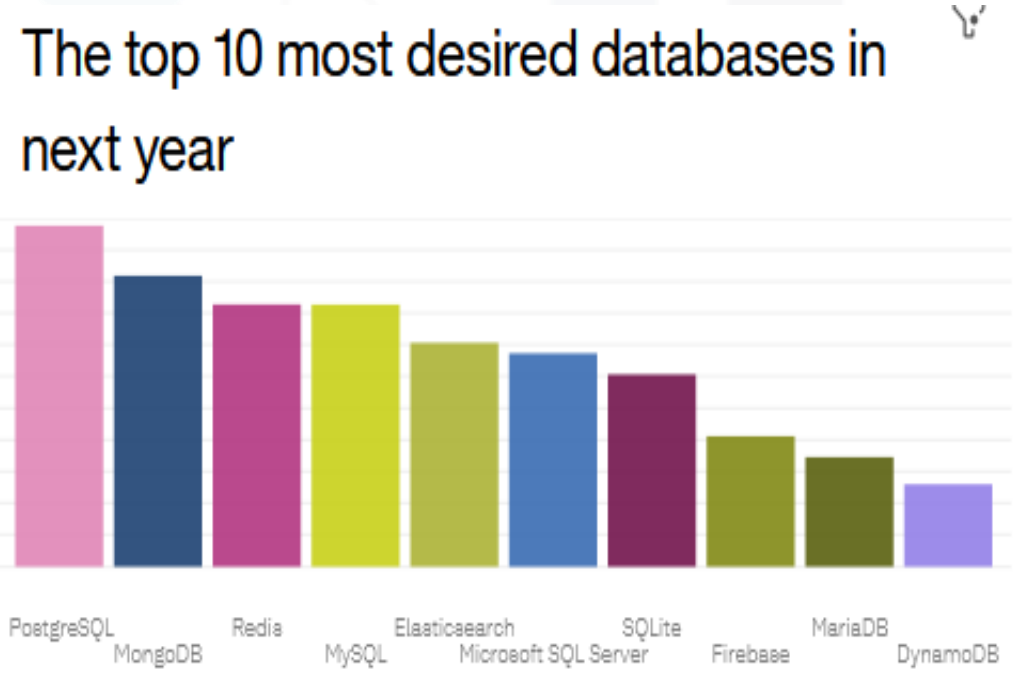
- The job market would see a higher influx of JavaScript and HTML engineer.
- This may decrease the present high demand of such developer and we may see a saturation point in coming years.

DATABASE TRENDS

Current Year



Next Year



DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- My SQL dominates the Database platform preferred by developers but in next few years we may see a steep decline in its preferences.
- MongoDB would be the database platform which would be most preferred Database platform to be preferred in future.

Implications

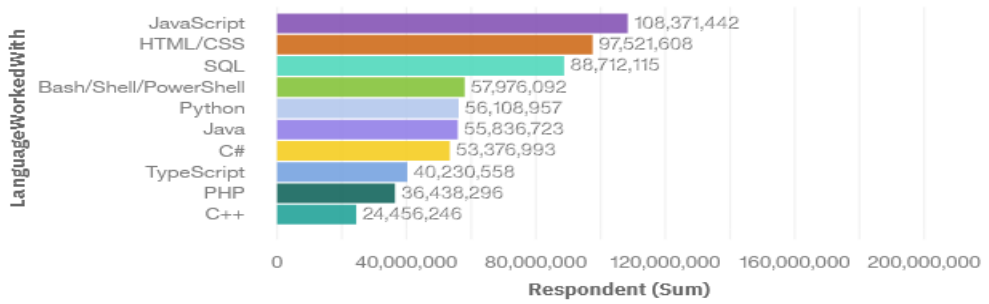
- We may see more organization preferring to work with MongoDB and PostgreSQL, as the popularity of them may increase in coming years.
- MySQL may reach a saturation point and wouldn't be sufficient to give you an edge over others in term of skills

DASHBOARD

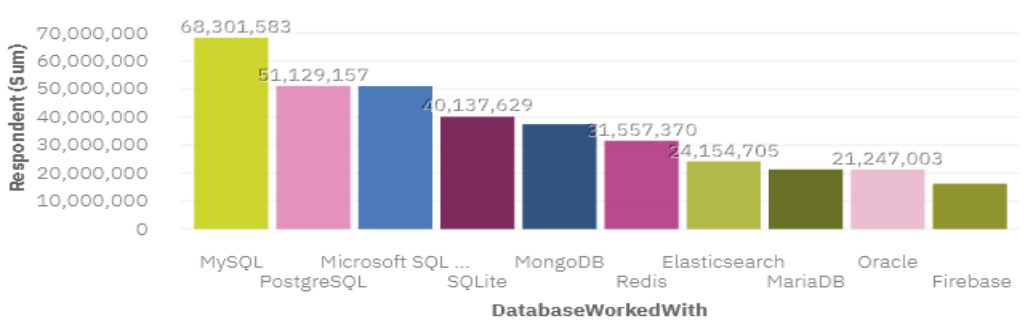
https://dataplatform.cloud.ibm.com/dashboards/a2988373-42c5-44b9-973b-8e6a81012f81/view/0023c1046e8a109f63cfc4e407902e5378662459b0bb8a57d3827b490d692297a8681a99c87c4b0bde400165a5ba100ccf

DASHBOARD TAB 1

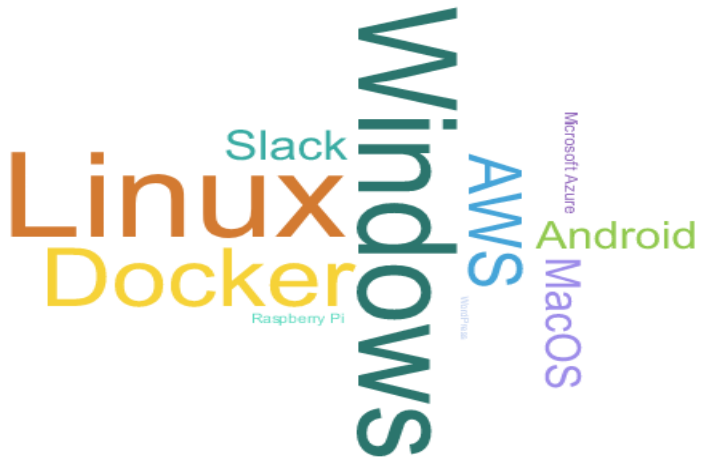
The top 10 most worked languages



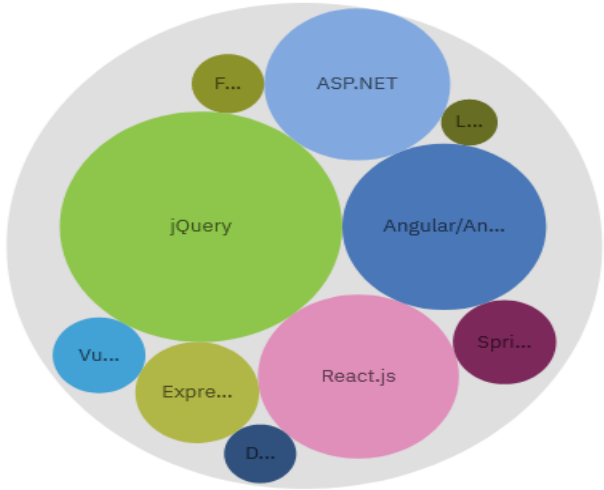
The top 10 most worked databases



The top 10 most worked platforms



The top 10 most worked web frames

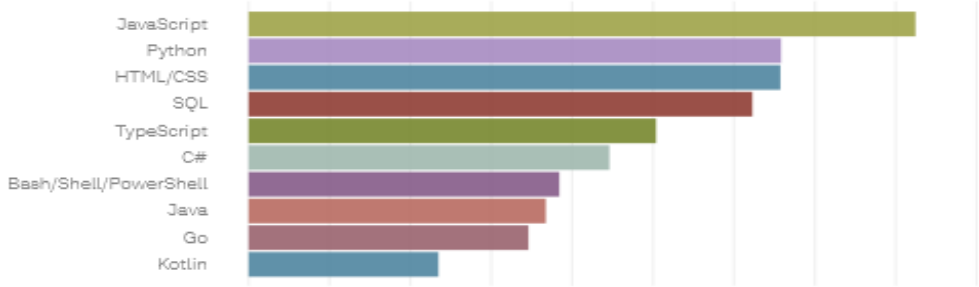


IBM Developer

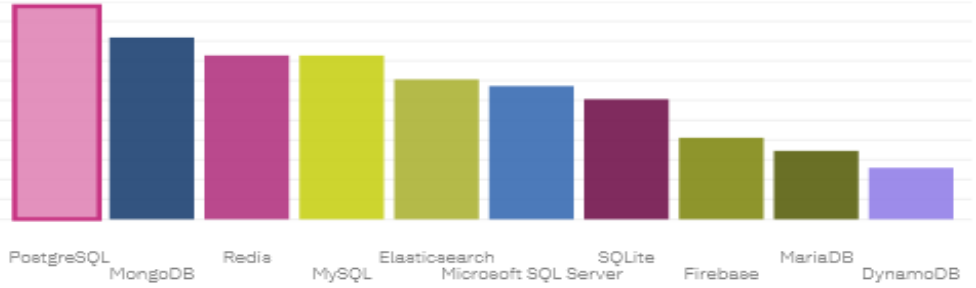
SKILLS NETWORK

DASHBOARD TAB 2

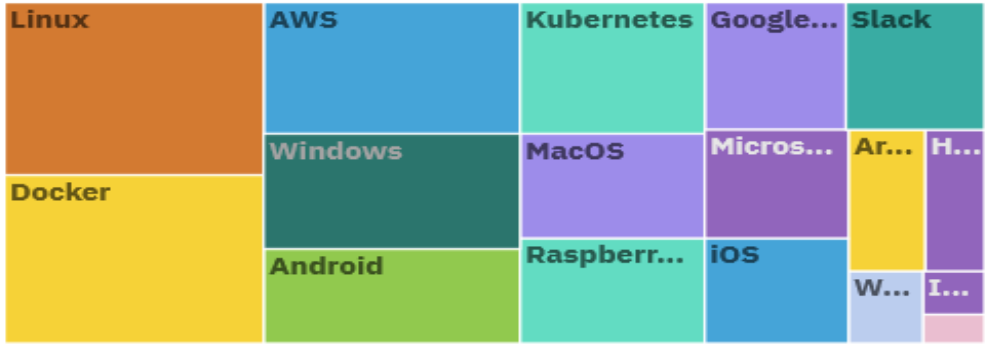
The top 10 most desired languages in next year



The top 10 most desired databases in next year

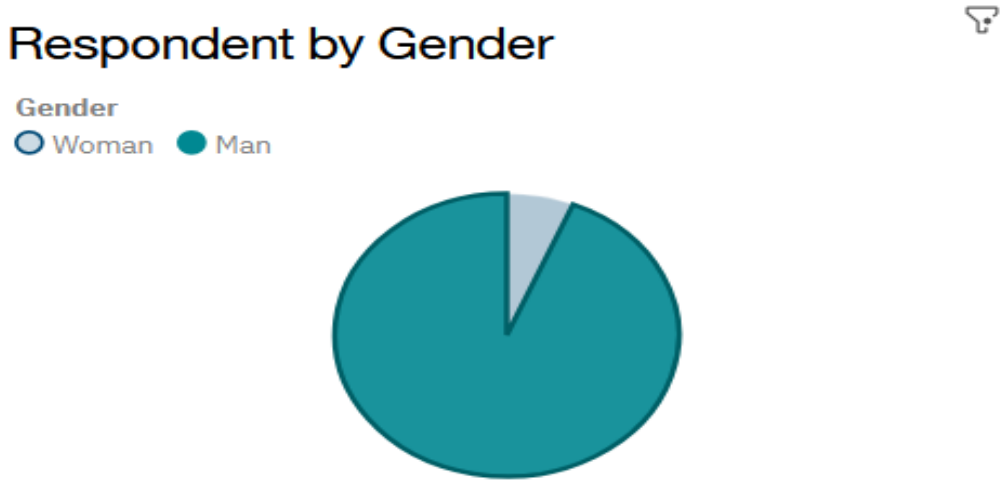


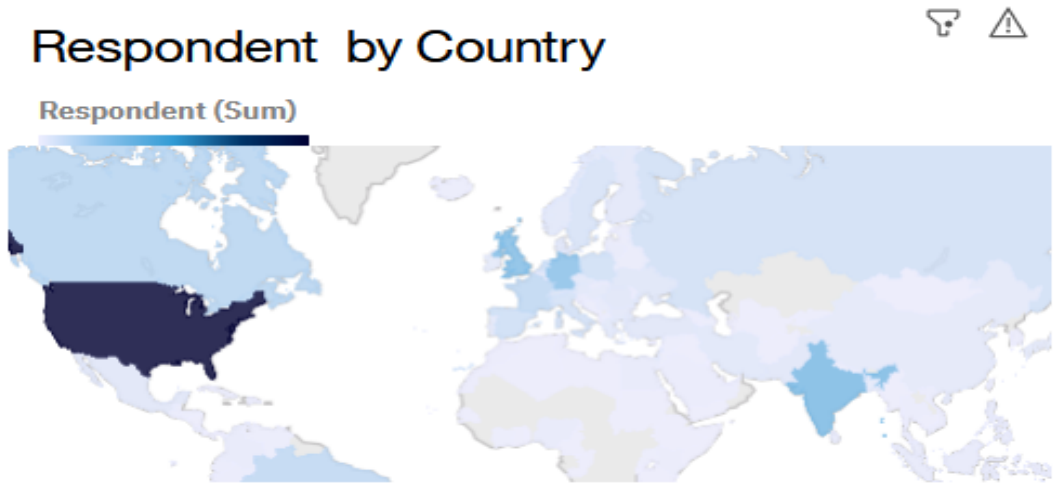
The most desired platforms in next year

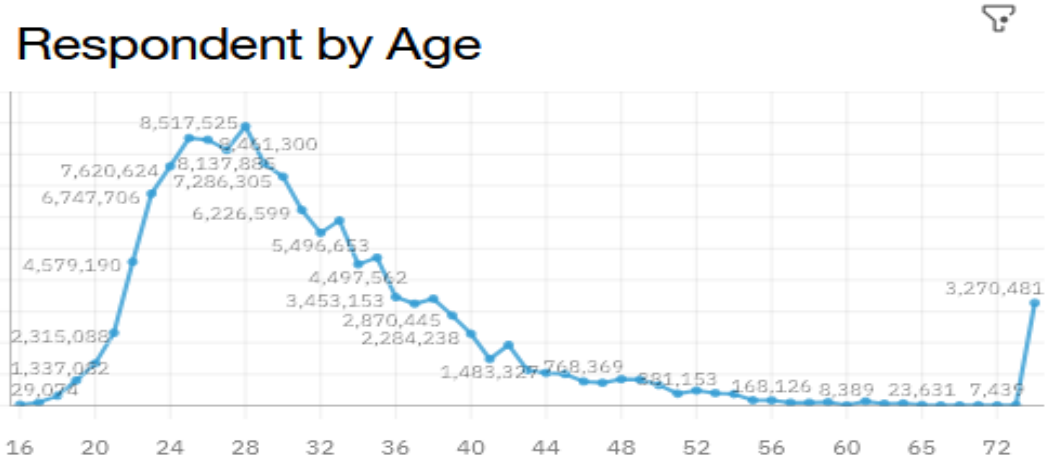


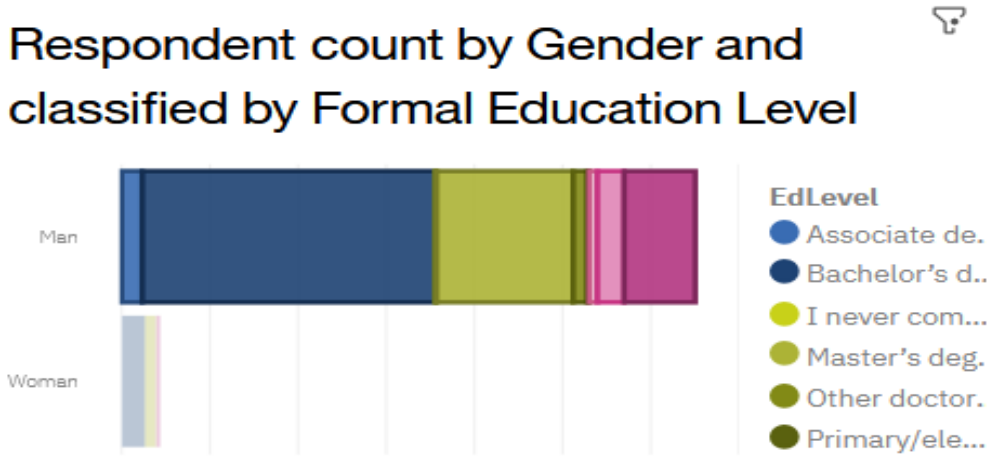
The top 10 most desired web frames in next year

DASHBOARD TAB 3









IBM Developer

SKILLS NETWORK

DISCUSSION

- Languages: Will SQL and Python tend to loose it dominance in upcoming years? Will other languages such as Go, C# would be the new emerging languages of future?

OVERALL FINDINGS & IMPLICATIONS

Findings

- The database technologies surveys results are dominated by male population, majority of them being USA respondents.
- Bachelor's and Master’s students highly prefer doing specialization in data technology.

Implications

- The database technologies surveys results are dominated by male population, majority of them being USA respondents.
- Bachelor's and Master’s students highly prefer doing specialization in data technology.

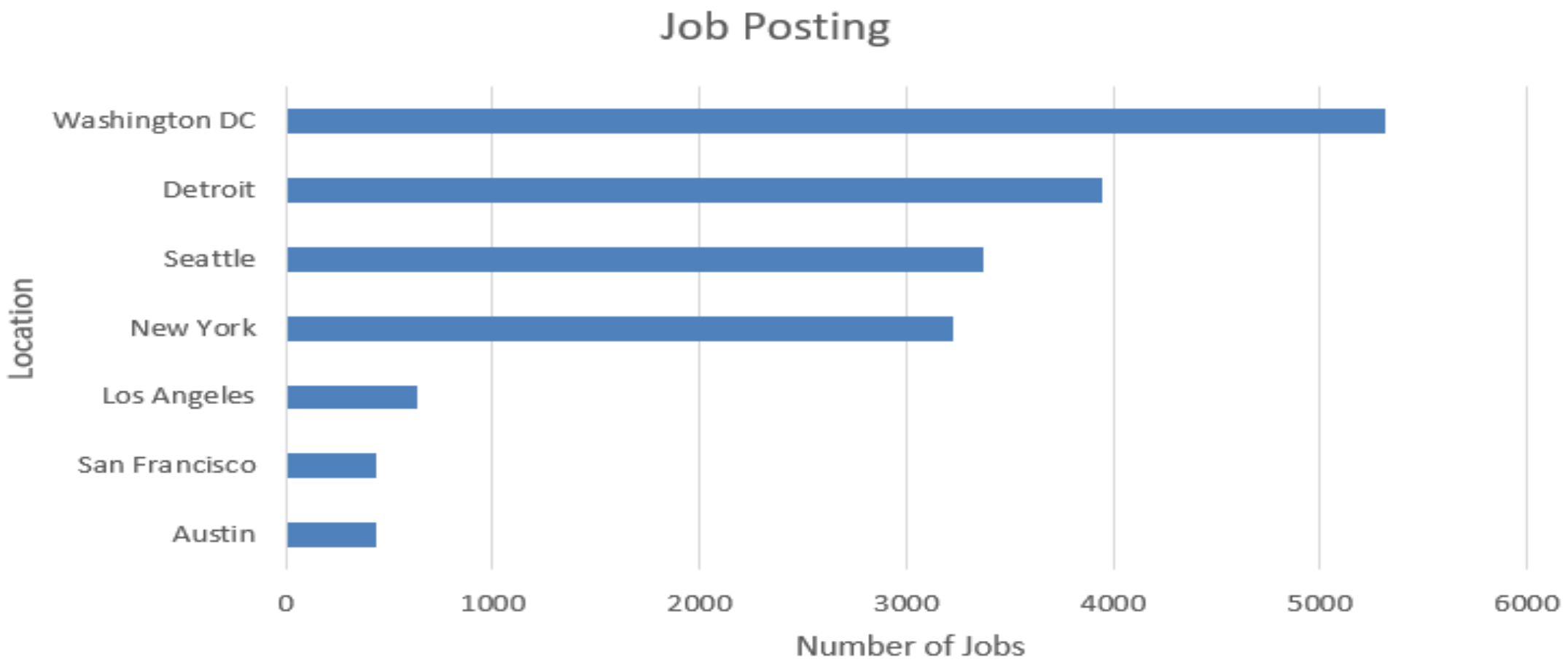
CONCLUSION

- We wanted to analyze the dataset for the preferred current and future trends in data technology while also considering the demographics of the respondent.
- From the data visualization it is clearly visible that the data technologies trends are preferred by population b/w 21 to 36 years
- It has a wider acceptance among male as compared to women.
- Among the languages , HTML and JavaScript will remain the most preferred languages to be worked with.
- Among the Database, we may see the popularity of SQL decreasing and which may be replaced by MongoDB and PostgreSQL

APPENDIX

- Resources used: Database taken from: Stack overflow blog under ODbl: Open Database License. IBM Cognos dashboard were used for creating Dashboards

JOB POSTINGS



POPULAR LANGUAGES