Analyzing and predicting Technological skills

Saurabh Paliwal

05/09/2021

IBM Developer

SKILLS NETWORK

Index

EXECUTIVE SUMMARY

Scope and Objective of this project:-

As a data analyst from global IT and consulting firm. I am presenting here the project on Analyzing and predicting Technological skills which are in demand.

Our project Includes following Steps:-

- Data collection by using Web Scrapping and APIs
- Data Wrangling
- Exploratory Data Analysis
- Data Visualization
- Dashboard of various trends
- Final presentation and outcomes of analysis

INTRODUCTION

- This project shows the analysis and prediction of changing Technologies to identify and keep pace with future skill requirement.
- The report is presented to a global IT & business consulting services firm to provide the insights from the data
- This will help the company to make suitable business decisions on the basis of technology growth.
- As a Data Analyst, I 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.
- This will also enhance the employee development, continues growth and higher satisfaction index.
- This will help the company to make suitable business decisions on the basis of technology growth, which will increase company’s Revenue

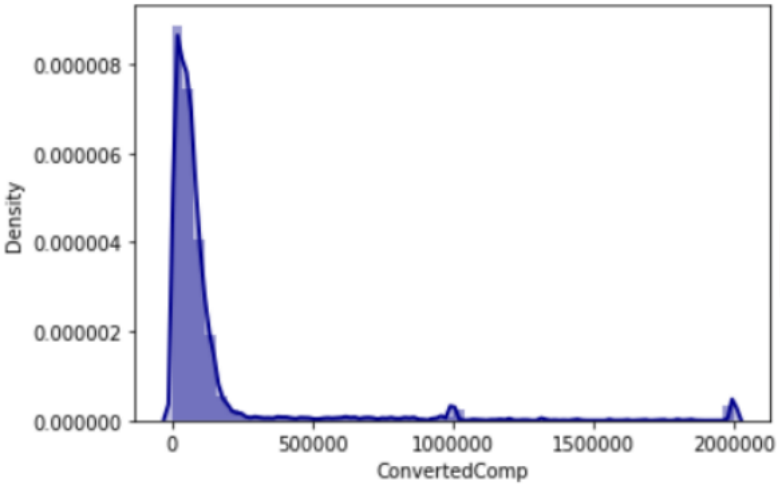
METHODOLOGY

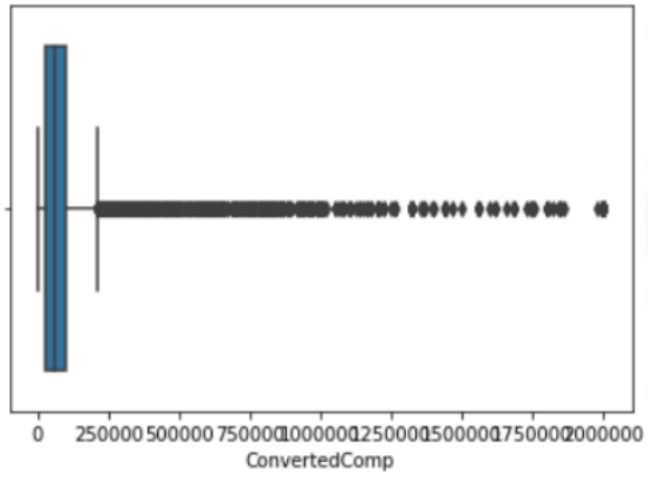
Data Collection

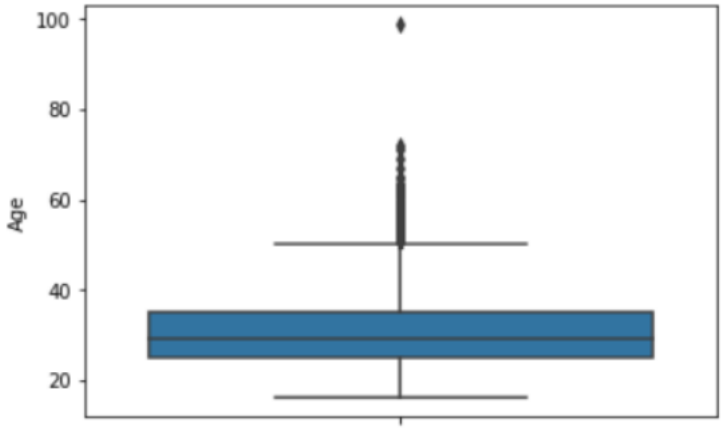
1. The number of jobs are determined currently open for various technologies (languages)
2. We called the API for given technologies
3. The results are written in excel spreadsheet
4. Data saved in form of excel file which listed languages
5. By Web scraping, we extracted information from given website using beautiful soup library
6. Then variables are scrapped from website

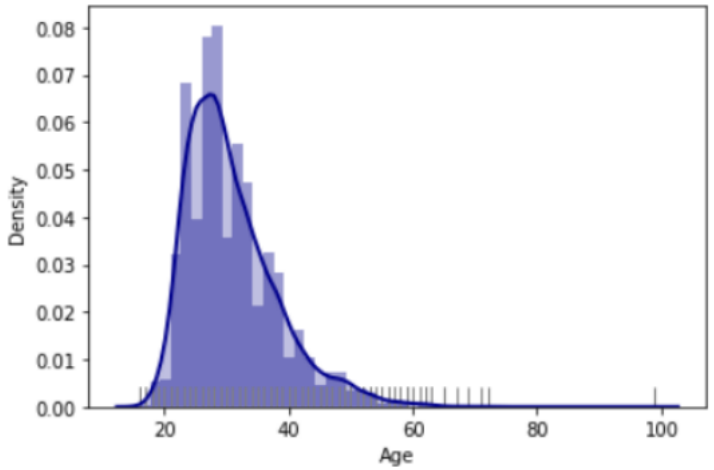
Methodology- Process Flow

RESULTS









- The density of change in salary compensation from 0 to 25000 is at its peak but it drastically falls down from 250000 to 2000000
- In the box plot from 250000 and age over 50 are outliers which are about 879. This outliers were removed to get the accurate result of analysis.
- The density of change in compensation from age 10 to 55 is high and it reduces after 60 age of person

Exploratory Analysis- Correlation

	Respondent	CompTotal	ConvertedComp	WorkWeekHrs	CodeRevHrs	Age
Respondent	1.000000	-0.019364	0.010878	-0.015275	0.002980	0.003950
CompTotal	-0.019364	1.000000	-0.063561	0.004975	0.017536	0.006371
ConvertedComp	0.010878	-0.063561	1.000000	0.034351	-0.088934	0.401821
WorkWeekHrs	-0.015275	0.004975	0.034351	1.000000	0.031963	0.037452
CodeRevHrs	0.002980	0.017536	-0.088934	0.031963	1.000000	-0.017961
Age	0.003950	0.006371	0.401821	0.037452	-0.017961	1.000000

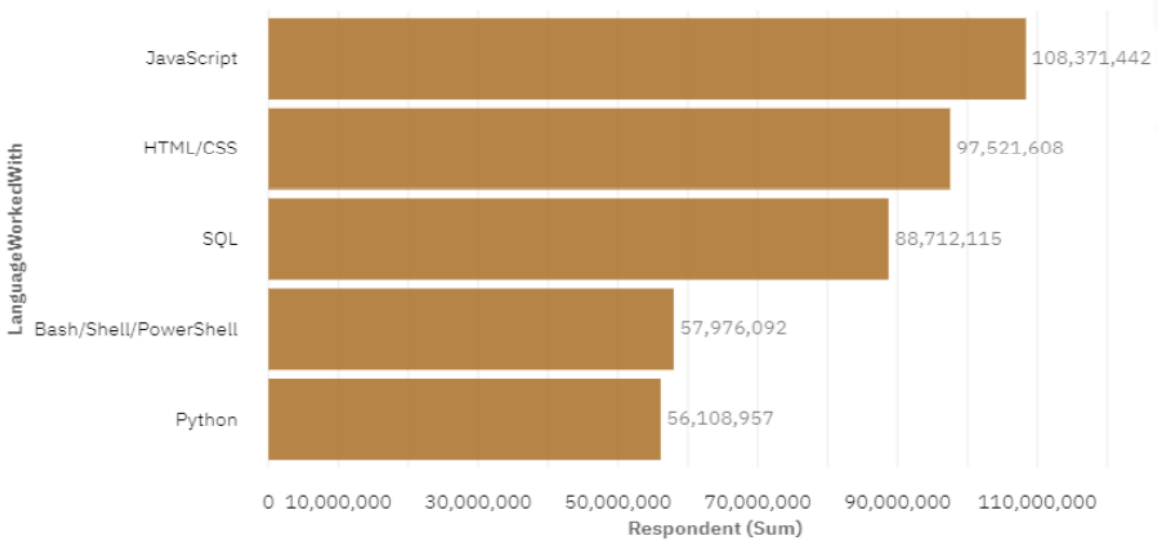
Insights:

- • There is a correlation between age of the person and Converted compensation.
- · On all other parameters, there is a weak correlation

PROGRAMMING LANGUAGE TRENDS

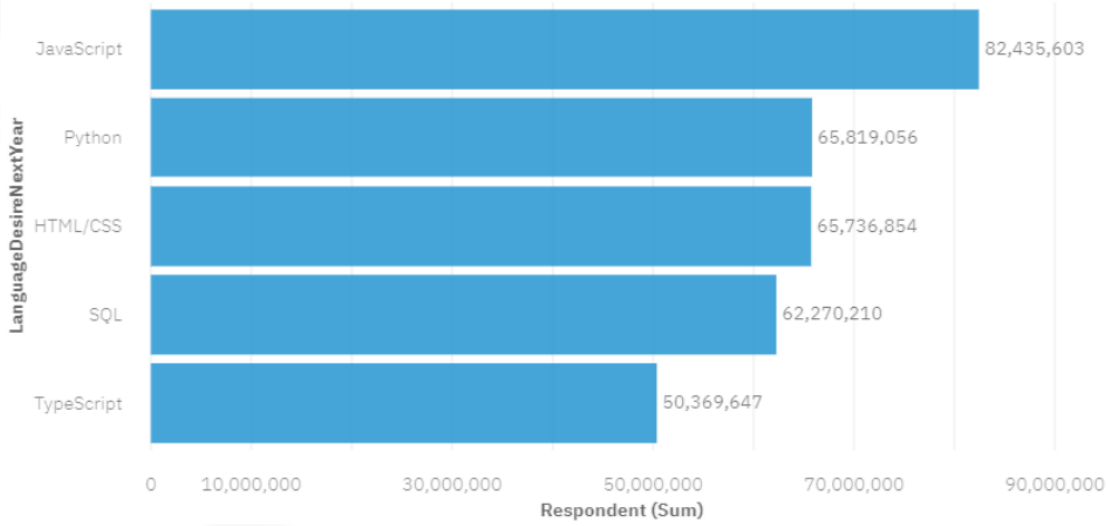
Current Year

Bar chart of top 5 programming languages for the current year



Next Year

Bar chart of top 5 programming languages for the next year



PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

- JavaScript is at top list in most popular languages with 108371442 users in the current year. Although reduction in users with 24% it remains popular next year.
- HTML/CSS and SQL are also most popular languages which holds 2nd & 3rd Rank
- Python and Shell are least popular languages with ranking of 4 & 5. But Due to high level language and easy to understand its Ranking increases from next year with 65819056 Respondents

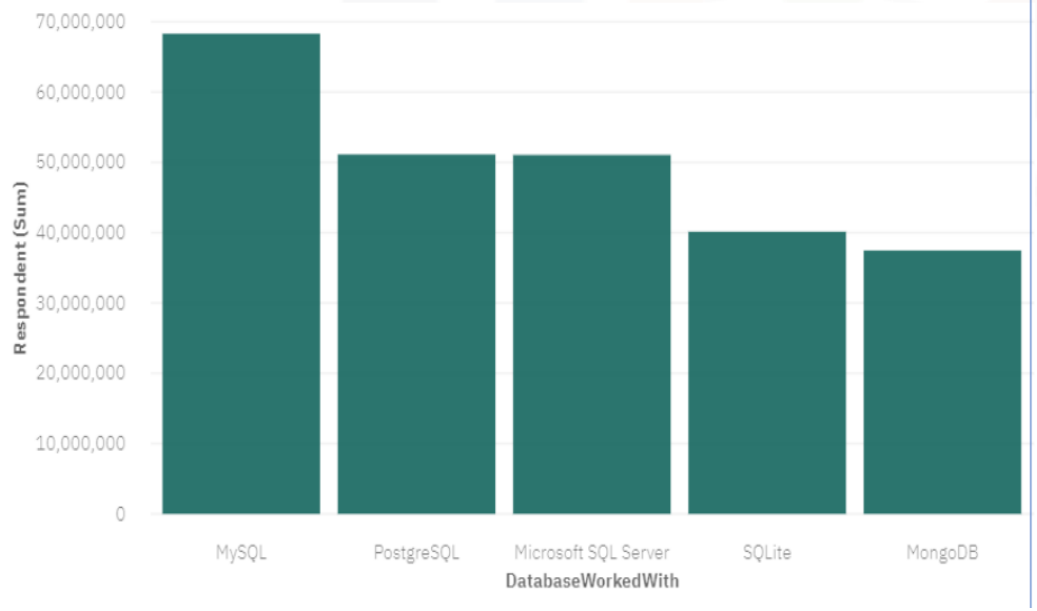
Implications

- JavaScript is high-level, often just-in-time compiled, and multi-paradigm. It has curly-bracket syntax, dynamic typing, prototype-based object-orientation, and first-class functions.
- Due to some advantages of JavaScript over CSS. CSS will become less popular in next year
- Due to increase in demand of Data Science & ML python becomes most popular language.
- Typescript is a new language which replaces the PowerShell.

DATABASE TRENDS

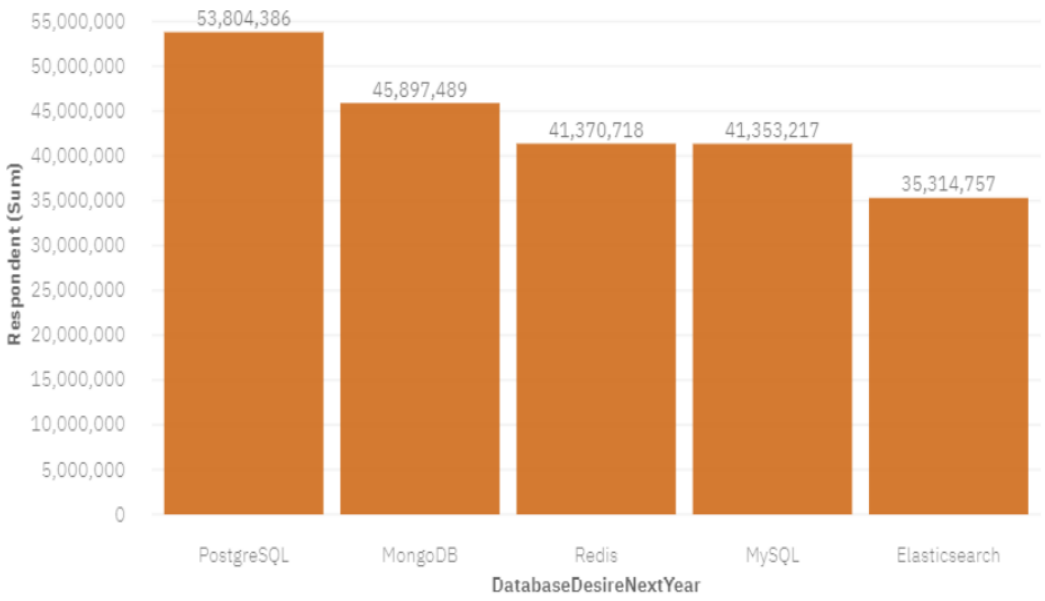
Current Year

Bar chart of top 5 databases for the current year



Next Year

Bar chart of top 5 databases for the next year



DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- MySQL is most popular programming language in current year with 1st rank highest respondents 68000000 but its uses will expected to drop by 34%
- The number of respondents uses PostgreSQL and Microsoft SQL server are same
- SQLite and Mongo DB have less respondents as compared to other Databases
- In the next year the use of PostgreSQL and MongoDB increases
- Highest growth is expected by ElasticSearch by 46% Followed by Redis(31%) MongoDB(23%)

Implications

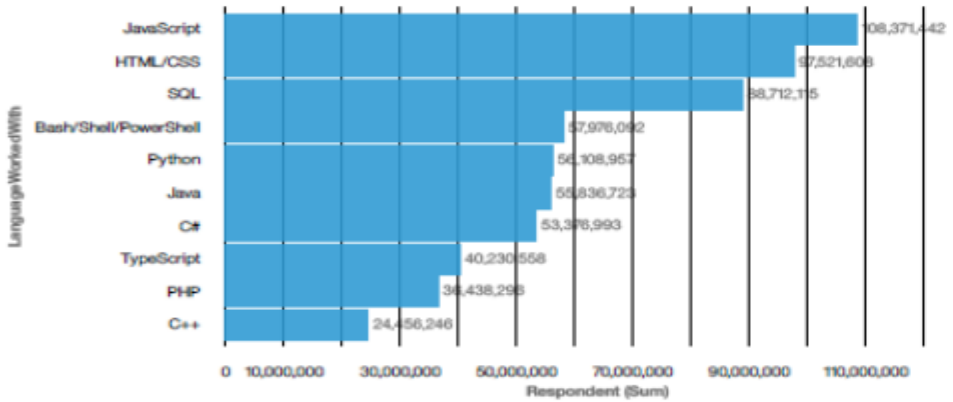
- Elastic Search, MongoDB and Redis have the highest growth potential
- IT companies may need to train their employees in above Databases
- PostgreSQL demand increses in next year becauses of because it support modern application and Materialized views
- In next year Redis and MySQL will have same number of respondents

DASHBOARD

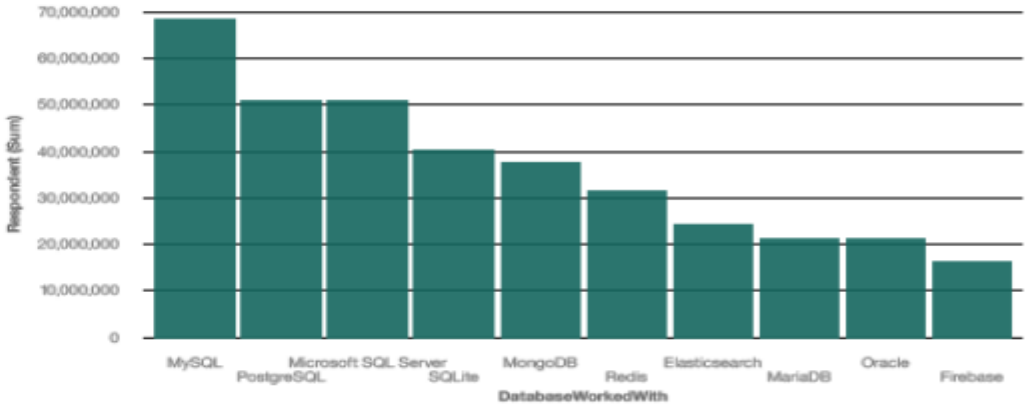
https://jp-tok.dataplatform.cloud.ibm.com/dashboards/5daf9620-6aaf-4dbf-88f3-ee057106cdfc/view/4167a202798a3fc715b4dce407cc7a0f7f362c59e7bb8b5789d47b490a692497f06f4793c82c420e8e170630f5eb440d9a

DASHBOARD TAB 1 – Current Technologies Uses

Top 10 Language Worked With



Top 10 Database Worked With



Platform Worked With

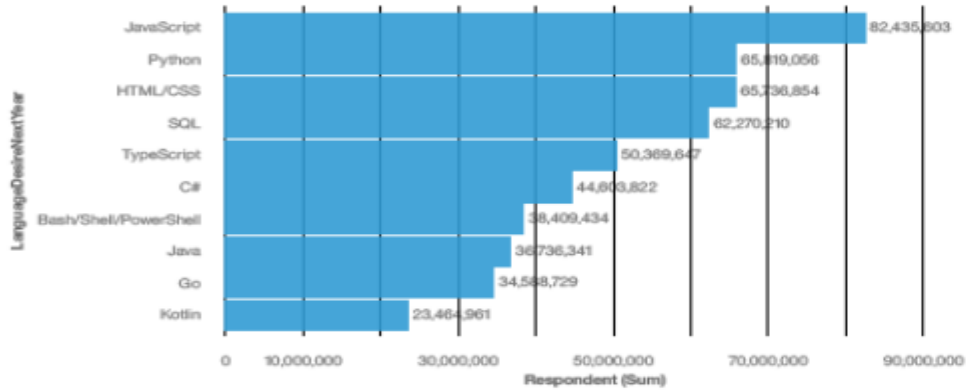
Top 10 WebFrame Worked With

IBM Developer

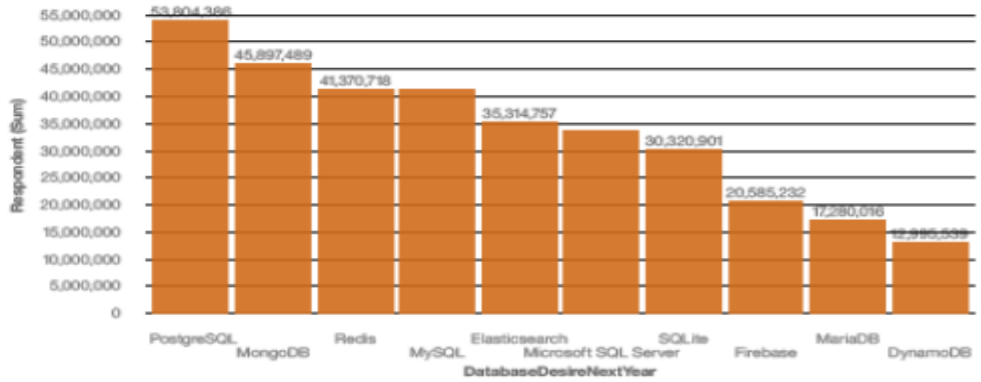
SKILLS NETWORK

DASHBOARD TAB 2:- Future Technology Trends

Top 10 Language Desire Next Year



Top 10 Database Desire Next Year



Platform Desire Next Year

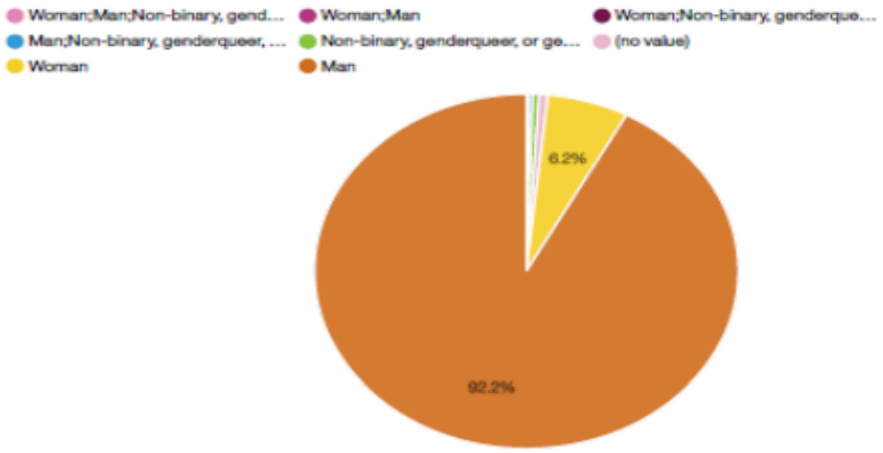
Top 10 Web Frame Desire Next Year

IBM Developer

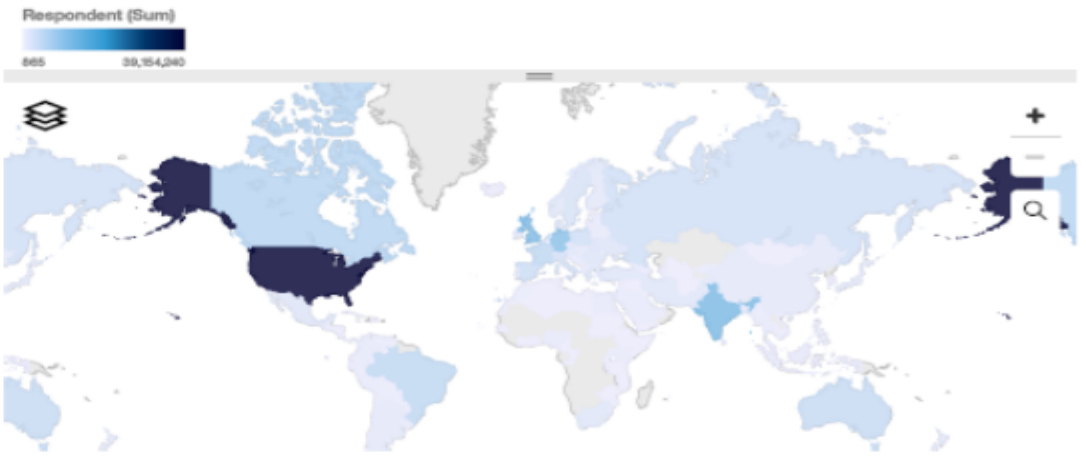
SKILLS NETWORK

DASHBOARD TAB 3:-Demographics

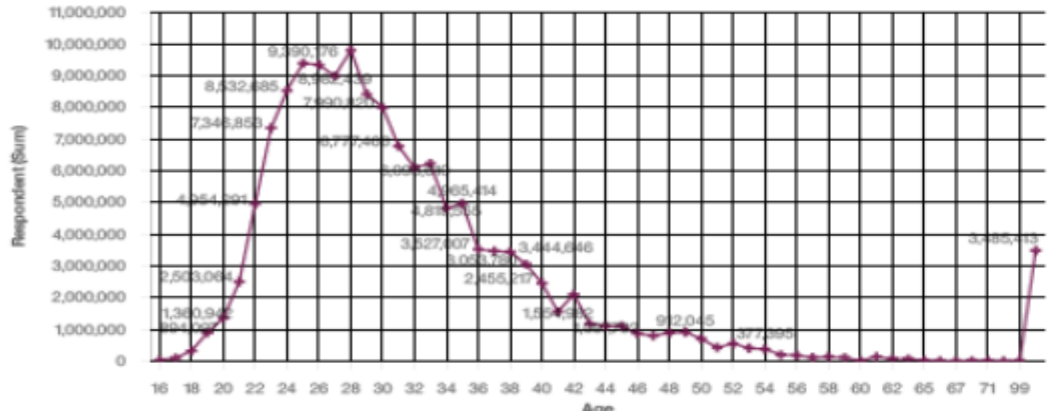
Respondent classified by Gender



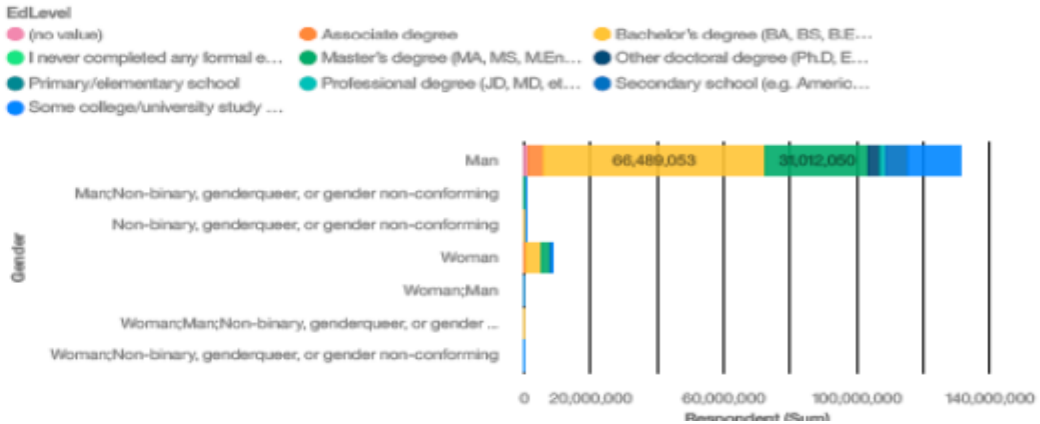
Respondent Count for Countries



Respondent Count by Age



Respondent Count by Gender, classified by Formal Education Level



IBM Developer

SKILLS NETWORK

DISCUSSION

- In the current Technologies Trends and Future Technologies trends Windows, Linux and Dockers are widely used
- On the other hand, the uses of windows will drop down next year
- JQuery is widely used in current trends as it has highest ranking in Top webframes
- But in Next year ReactJs will get more popularity, Because **React** Is Faster Than **jQuery**.
- One of the biggest things that **React** has going for it is the use of the Virtual DOM (Document Object Model) instead of the traditional DOM.

DISCUSSION

- Majority of the respondents are from USA, followed by India and Canada

- Male respondents are more than the Female respondents

- Majority of them lie between the age 22 and 35 years

- Maximum respondents have bachelor's degree followed by Master’s degree

OVERALL FINDINGS & IMPLICATIONS

- Findings:
- JavaScript is the top language worked with. It will be at the top although with reduction by 24%
- Desire for increase in usage for Python & Typescript by developers
- HTML/CSS & C# language users are expected to reduce
- In future there is reduction in desire for usage is expected in Microsoft SQL Server (-34%) and SQLite by -24%
- Firebase is a new technology trend expected in databases

- Implications:
- Female respondents has to be increased as it is 0.2% as compared to Male respondents(92%)
- TypeScript & Python languages are expected to growth in usage
- Elasticsearch, Redis & MongoDB databases are expected to growth
- On training & development, the companies should focus on the “TypeScript” & “Python” languages and the above stated databases

CONCLUSION

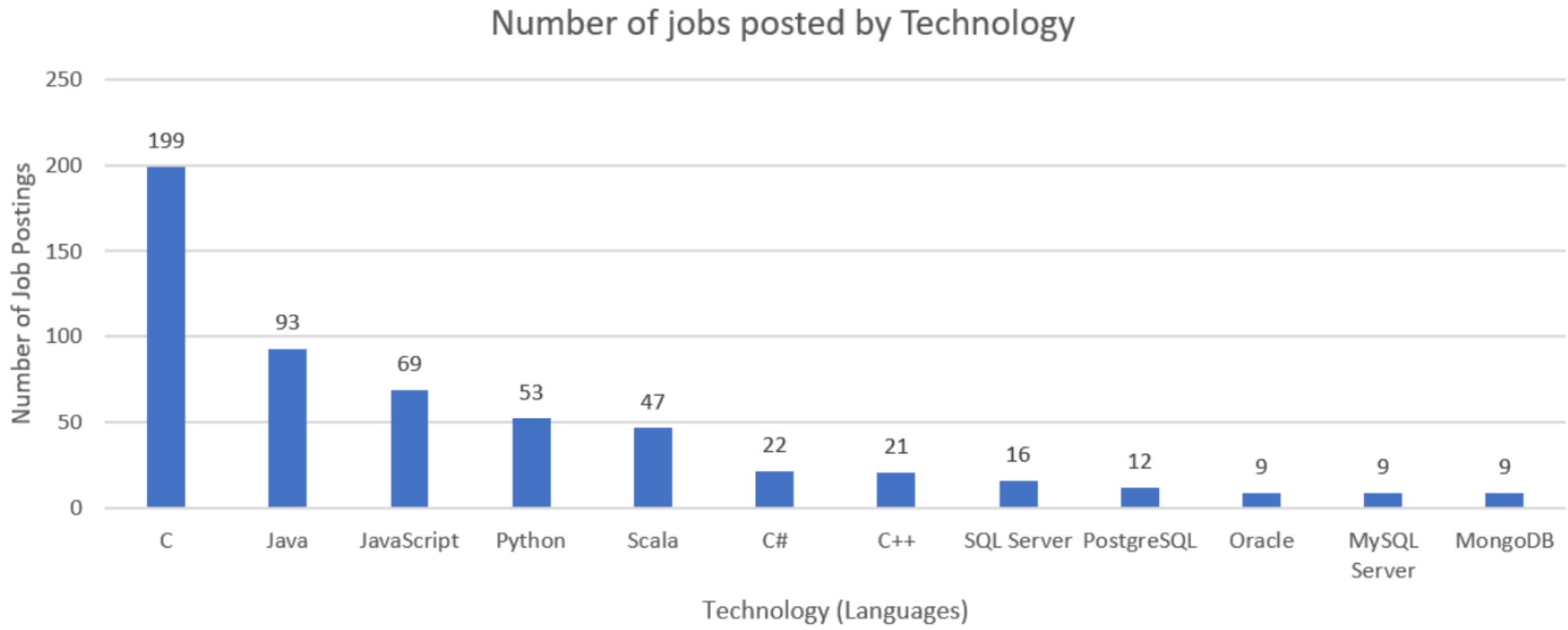
- The study provide good insights on the technology current usage and its future trend.
- Based on the trend presented in this project, the IT companies can plan resource capability development for upcoming technologies through training and development (ex: “TypeScript” & “Python” languages and Elasticsearch, Redis & MongoDB databases)
- The IT infrastructure companies should plan to develop the infrastructure for the next generation languages and databases
- At the same time, there are technologies on which continued focus is required (SQL & JavaScript)
- Also, the companies can ramp-down the services related to technologies which have lesser desire from developers in future (For example: HTML/CSS and Microsoft SQL Server)

APPENDIX

- GitHub job posting - Number of jobs posted by Technology
- Programming languages - Average Annual salary ($) by Technology

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.