Capstone Project Report

NAME: Somdutt Acharya

DATE: 02-02-2021

OUTLINE

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

EXECUTIVE SUMMARY

- The purpose of the project was to determine the trends and distribution of the programming languages, databases and platform based on number of respondents.
- The trends of programming languages was analyzed based on 3 factors
  - Current Technology Usage
  - Future Technology Trend
  - Demographics
- Insights are generated for the programming language trends for the current year and next year, databases for current year and next year, platform for current year and next year and so on.
- Classification has been made on gender and no. Of respondents are recorded as per demographics
- Bar charts have been developed to know the trends of the top 5 and top 10 languages and databases and the rank.

INTRODUCTION

- **Problem Statement:** Collecting data from various sources and identifying trends for this years report on emerging skills.
- The task was to collect the top programming skills that are most in demand from various sources including:
    - Job Posstings
    - Training Portals
    - Surverys

- In this project I worked as a Data Analyst for a global IT and business solution company.
- The purpose was to perform data analysis and present the findings so that the company could keep pace with the changing technologies, remain competitive and identify trends for this years report on emerging skills.

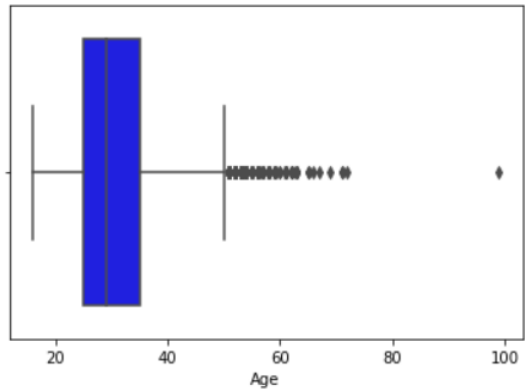
METHODOLOGY

- Various methods were used for collecting different types of data such as:
  - Job data through API's
  - Web data through Web Scraping
  - Survey data as CSV's
- Exploring the data,
  - At first basic exploratory data analysis was done on the CSV files.
  - The file was called using a dataframe object and then various methods were called to get the first 5 records of the data
  - Total number of rows and columns were retireived from the data
  - Name of the columns were retireived from the CSV dataset.
- Data Wrangling
  - Duplicates were detected and removed
  - Null values were detected and replaced
  - Finally, data was normalized
- Data Analysis and Visualizations
  - Distribution of data was checked
  - Outliers were identified and removed
  - Correlation between features were identified in the dataset
  - Finallly bar charts and other charts were plotted and dashboard was created

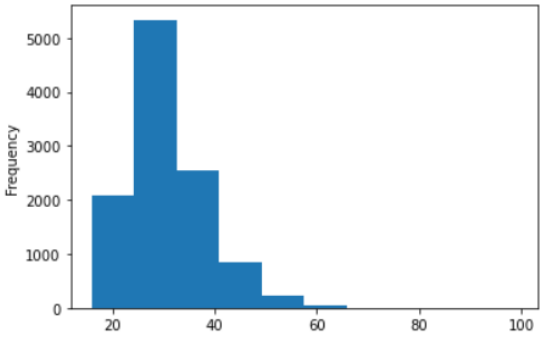
RESULTS

Initially the data was having duplicates and null values. Eventually data was cleaned, wrangled and normalized so that it could be used for analytics and visualisation. Various insights were captured which can be seen below

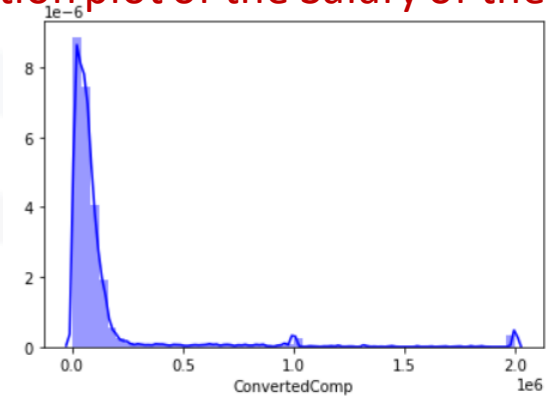
Boxplot of Age of respondents. We can see that there are outliers.These outliers were later removed.



We can see that age group of respondents are between 24 to 32 from the histogram

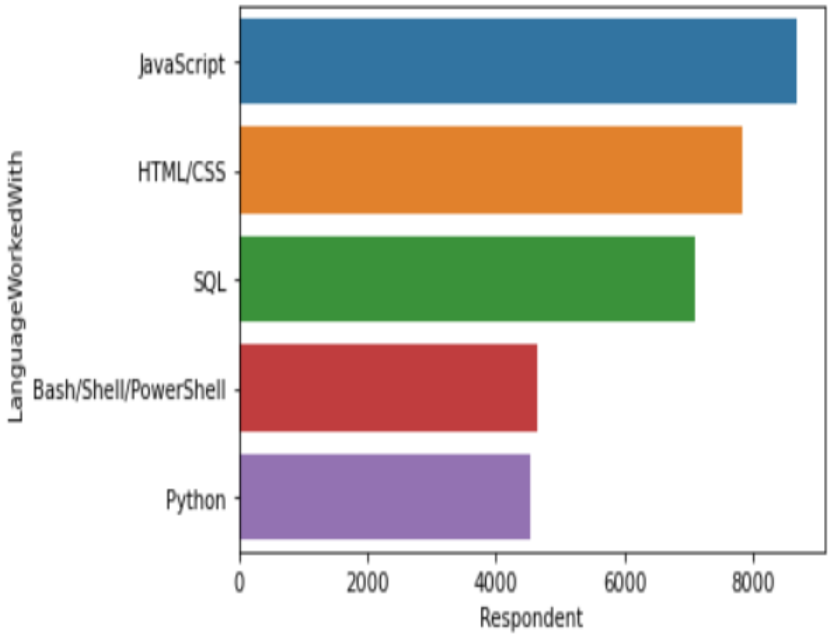


Distribution plot of the Salary of the respondents

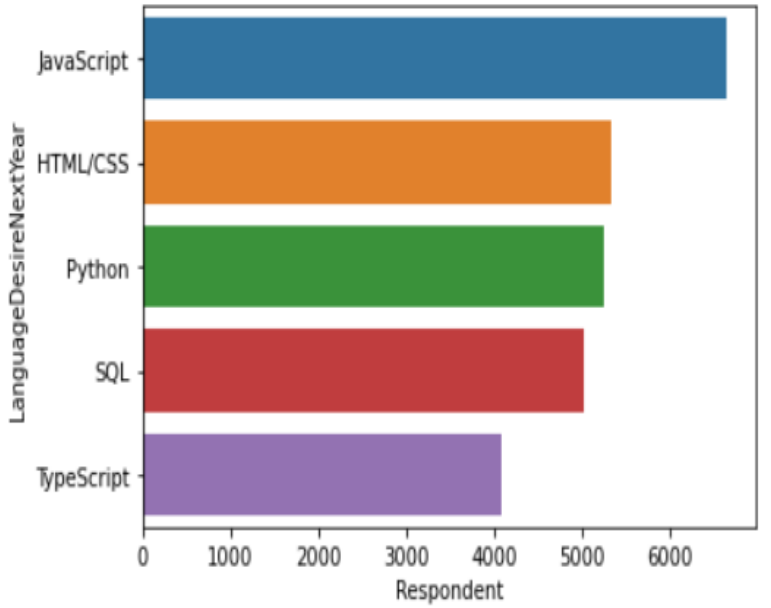


PROGRAMMING LANGUAGE TRENDS

Current Year



Next Year



PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

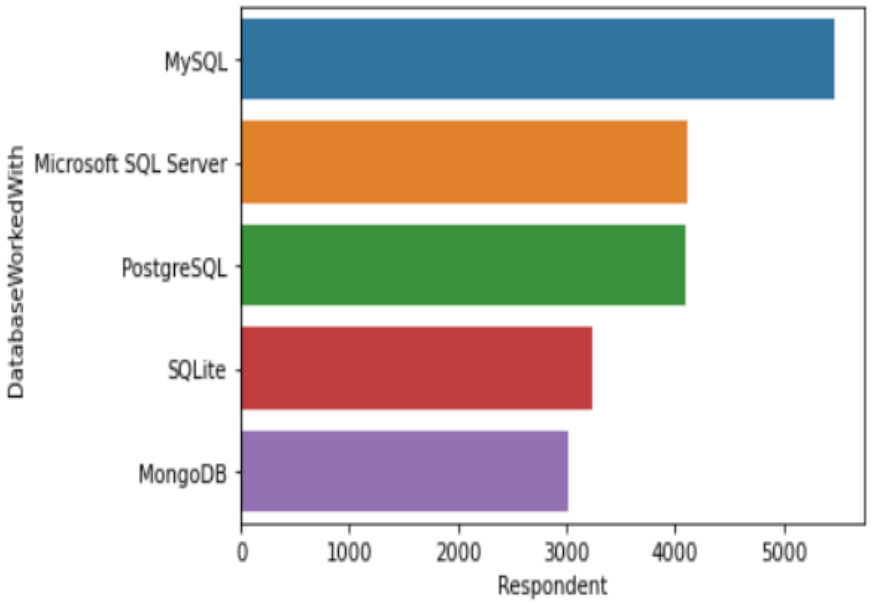
- **Finding 1**: Javascript Remains the top trending language
- **Finding 2:** HTML/CSS remains 2nd rank language in the current year and for next year
- **Finding 3:** Increase in the demand for Python in the coming year as compared to present year.

Implications

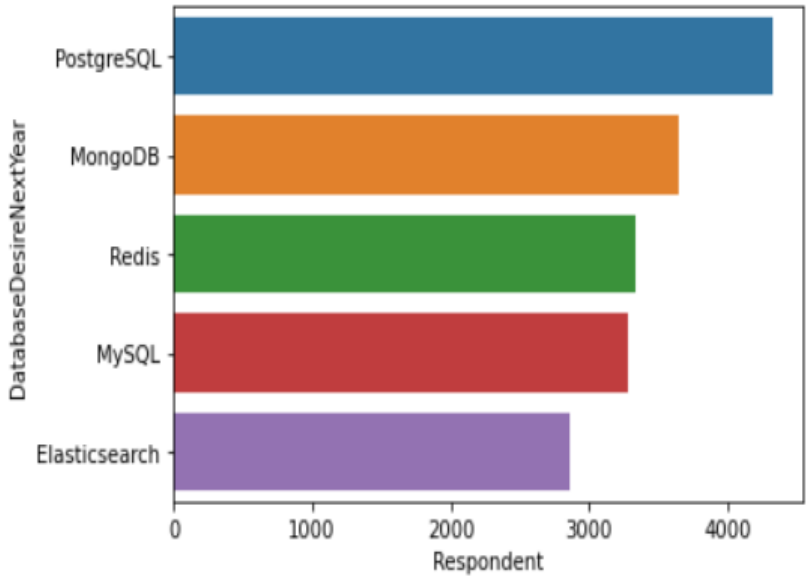
- **Implication 1:** Javascript shows a good sign for both current year and next year and thus professionals can learn that language as it will remain top in demand.
- **Implication 2:** The demand for HTML/CSS will reduce in the coming year but it will remain the 2nd most popular language
- **Implication 3:** The demand of python language is going to increase in the coming year and thus it is a good option to learn python.

DATABASE TRENDS

Current Year



Next Year



DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- **Finding 1:** There will be a significant increase in the demand of PostgreSQL(becoming no.1 rank databse in next year) and decrease in the demand for MySQL(no.3 rank in next year)
- **Finding 2:** Current year rank of Mongo DB-5, Next year rank of MongoDB-1.
- **Finding 3:** Elasticsearch is in the list of top 5 database desired next year.

Implications

- **Implication 1:** It is a good option to learn PostgreSQL as it is going to be top ranking database to be desired for. There will be decrease in the demand for SQL but it can still be preferred to be learned as it will remain rank 3 in the desired database next year.
- **Implication 2:** It is a wise decision to learn Mongo DB as the rank of Mongo DB will jump to 2nd as the most desried database in the next year.
- **Implication 3:** The demand for elastic search is increasing for next year and will be in the top 5 languages desired next year.

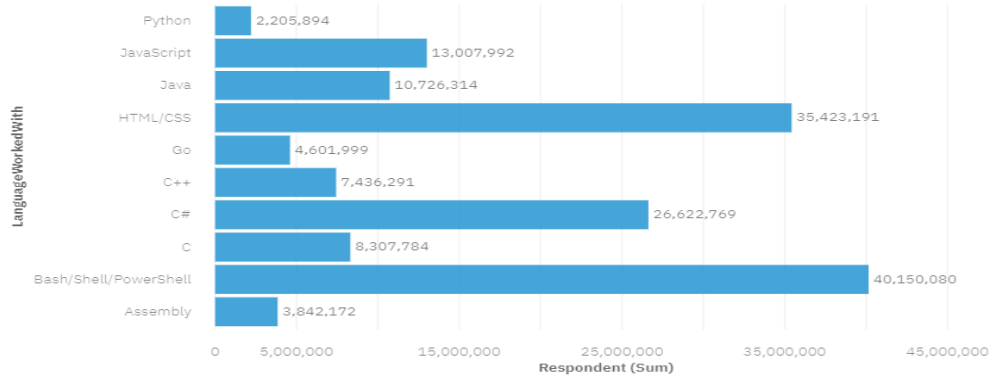
DASHBOARD

The permalink to Cognos Dashboard:

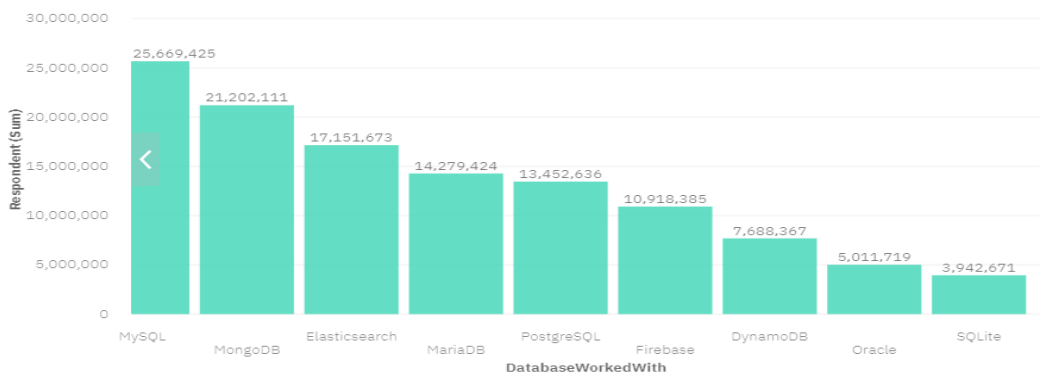
https://eu-gb.dataplatform.cloud.ibm.com/dashboards/054b8101-8ffb-4c34-bddc-bd0536af9211/view/5512e33a1b9634f609e8d4e407cd78072e327608babbd20384827b4906642397a96f1692c87d1f09da450661f5ef4459cd

DASHBOARD TAB 1

Top 10 languages worked with

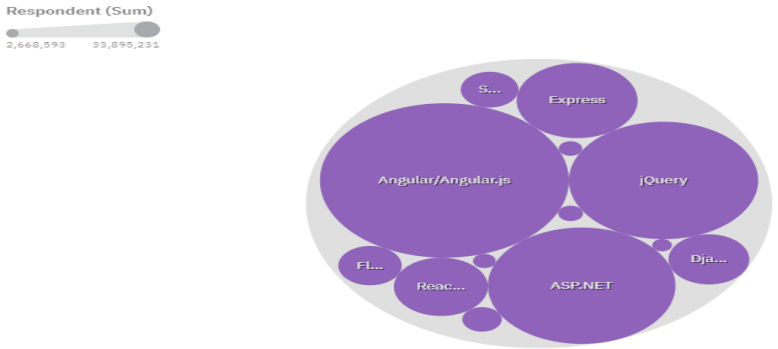


Top 10 Databases Worked With



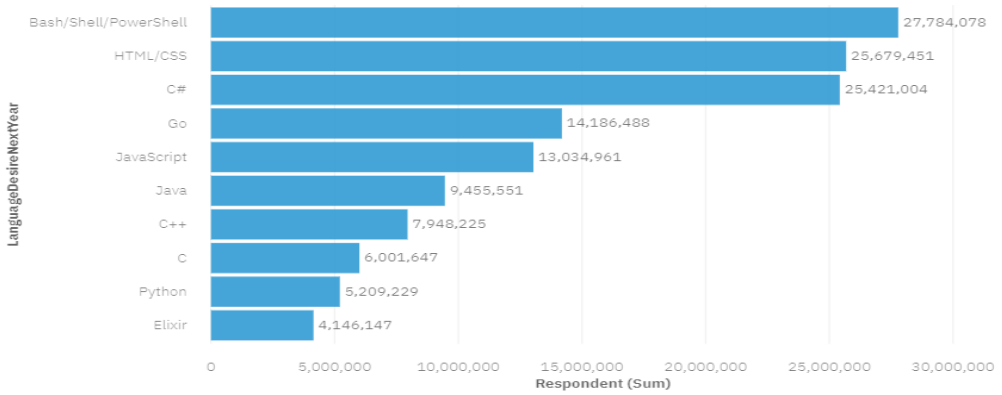
Platform Worked With in Words

Top 10 web frame worked with as bubble chart

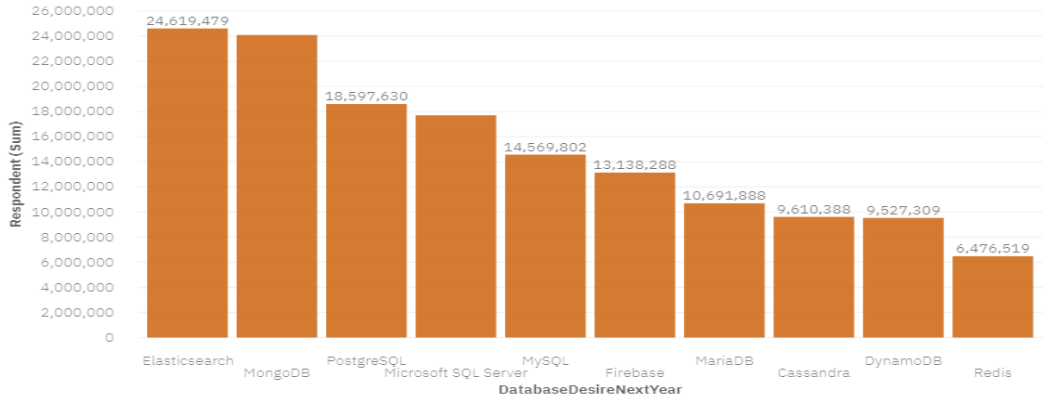


DASHBOARD TAB 2

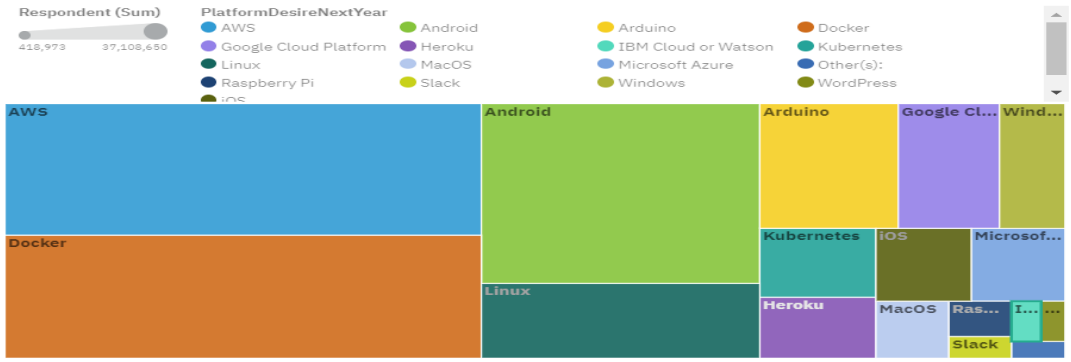
Top 10 language desired next year



Top 10 database desired next year



Treemap for Platform desired next year

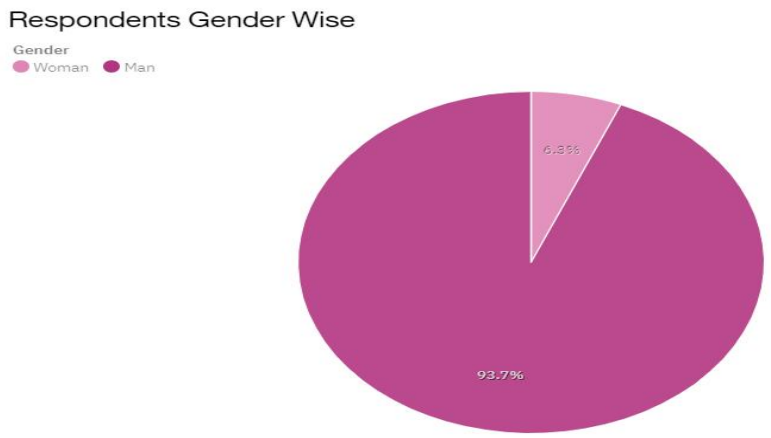


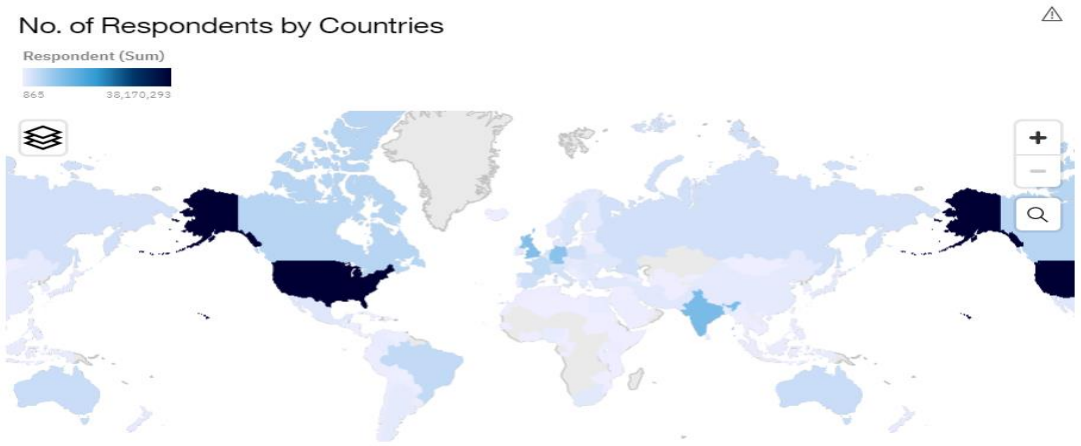
Top 10 webframe desire next year

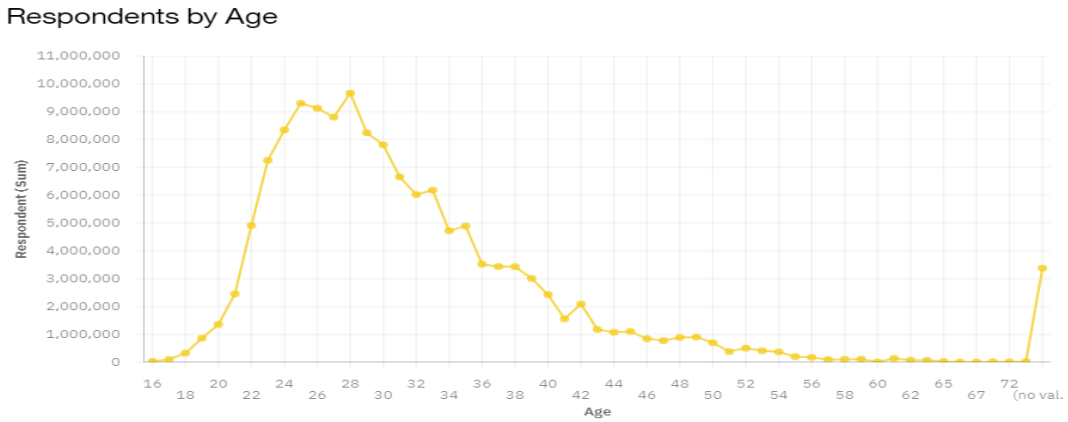
IBM Developer

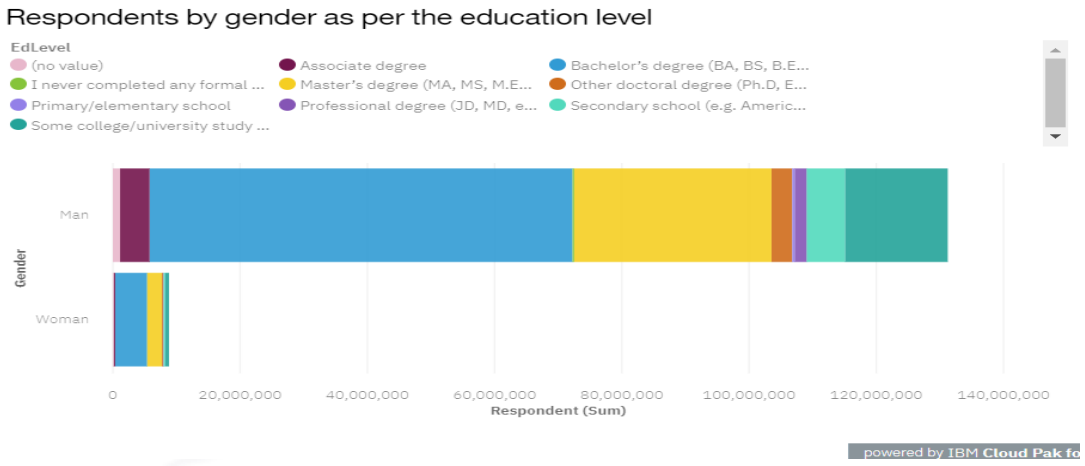
SKILLS NETWORK

DASHBOARD TAB 3









DISCUSSION

1. Powershell is the top language worked with in the current year.
2. MySQL is the top databse worked with in the current year.
3. Android and Docker are the top platforms worked with in the current year.
4. Powershell is the top language desired next year.
5. Elastic search is the top database desired next year.
6. AWS and Docker are the most desired platform for next year.
7. The number of men are much higher in the programmming language field.
8. The highest number of coders are from USA and Alaska

OVERALL FINDINGS & IMPLICATIONS

Findings

- **Finding 1:** The Powershell, Javascript, HTML and Pyhton are top in demand.
- **Finding 2:** PostgreSQL, MySQL, MongoDB and Elasticsearch are high in demand
- **Finding 3:** The number of respondents are in the age group of 24 to 32 and most of them are men.

Implications

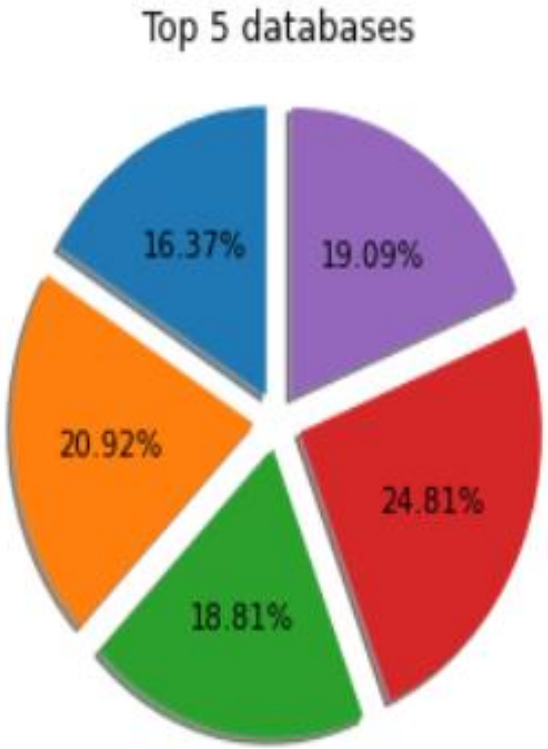
- **Implication 1:** This means that it is better for the professionals in the programming language field to learn Powershell, Javascript, HTML and Python.
- **Implication 2:** These databases will be in demand in the coming year as well as in the present year. It would be a good option to learn these databases.
- **Implication 3:** If you are a man and is in the age range of 24 to 32, then there are chances of you getting hired as per the results.

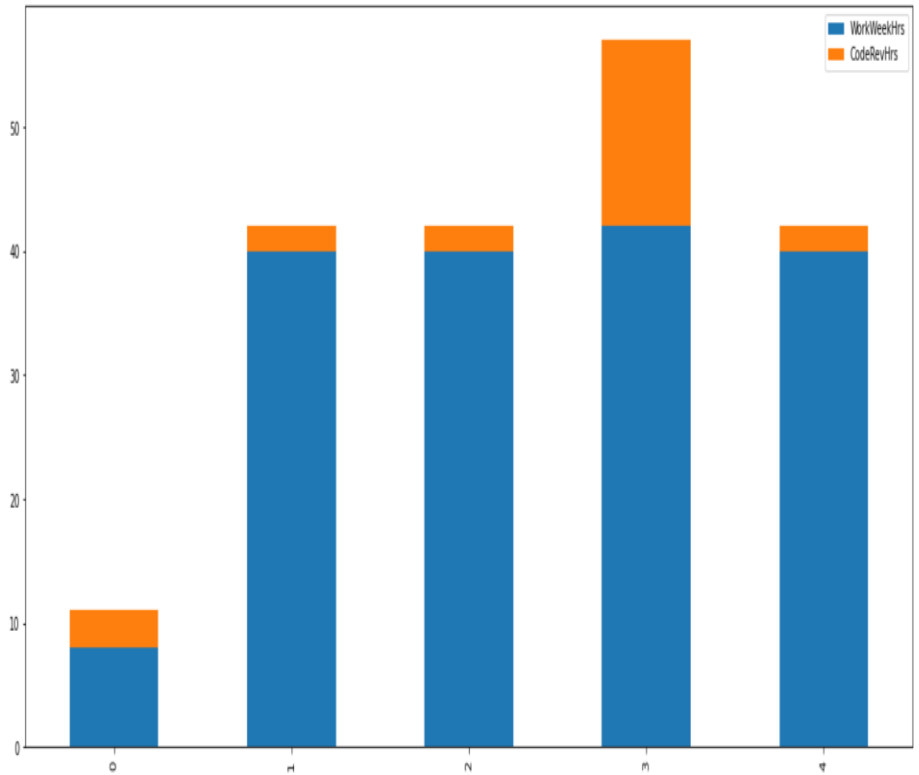
CONCLUSION

- The purpose was to find the trend of the demand of popular programming lanuages, databases and platforms.
- It can be inferred from the findings that Powershell, Javascript, HTML and Python languages are high in demand.
- Databases and platforms such as PostgreSQL,MySQL, Elasticsearch are high in demand.
- It is advised to the software professionals of age group 24 to 32 to learn these databases and languages, especially to the men because a lot of men are in this industry between the age group of 24 to 32.

APPENDIX

- Percent pie chart of top 5 databases desired next year and coding hours and coding revision hours stacked bar chart





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.