DATA ANALYTICS PROJECT

ANJANI UPADHYAY

21-02-2023

IBM Developer

SKILLS NETWORK

OUTLINE

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

IBM Developer

SKILLS NETWORK

EXECUTIVE SUMMARY

- Analysis of future skill requirements with the aid of data analytics.
- Identification of insights and trends including :
  - Top Programming Languages in demand
  - Top Databases in demand
  - Most popular IDEs
- Comparison between current technology usage and future technology trends.
- Visualization of number of respondents according to various demographics.

INTRODUCTION

- Collection of data from various sources
- Cleaning the dataset
- Analysis of the dataset
- Creation of visualizations
- Building a dashboard to identify trends and patterns

METHODOLOGY

- **Data Collection** using:
  - APIs
  - Webscraping
- **Exploring** the dataset.
- **Data Wrangling** using NumPy and Python to :
  - Identify duplicate and missing values as well as replace/remove them
  - Normalization of data
- **Exploratory Data Analysis** using SQL and Python to :
  - Find the distribution of data in the dataset
  - Identify and remove outliers
  - Depict correlations between different features in the dataset.

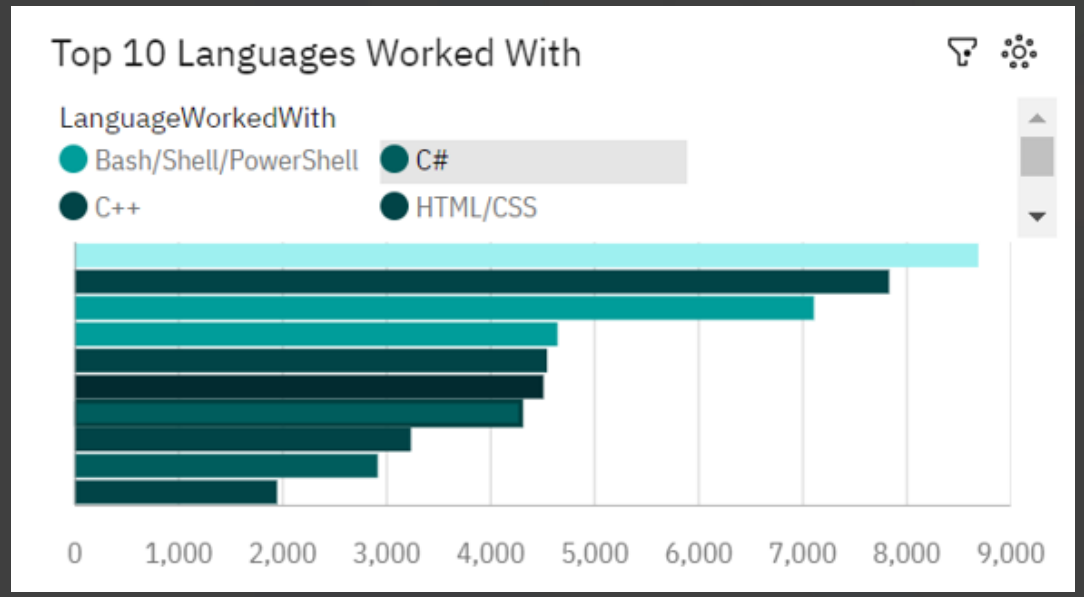
METHODOLOGY

- **Data Visualization** using Pandas and Matplotlib to:
  - Visualize distribution of data
  - Visualize relationship between two features
  - Visualize composition of data
  - Depict comparison of data using line and stacked bar charts.
- **Creation of a Dashboard** using IBM Cognos Analytics to:
  - Visualize current technology usage, future technology trends and deomgraphics
  - Portray KPIs (Key Performance Indicators) with the help of appropriate chart types.

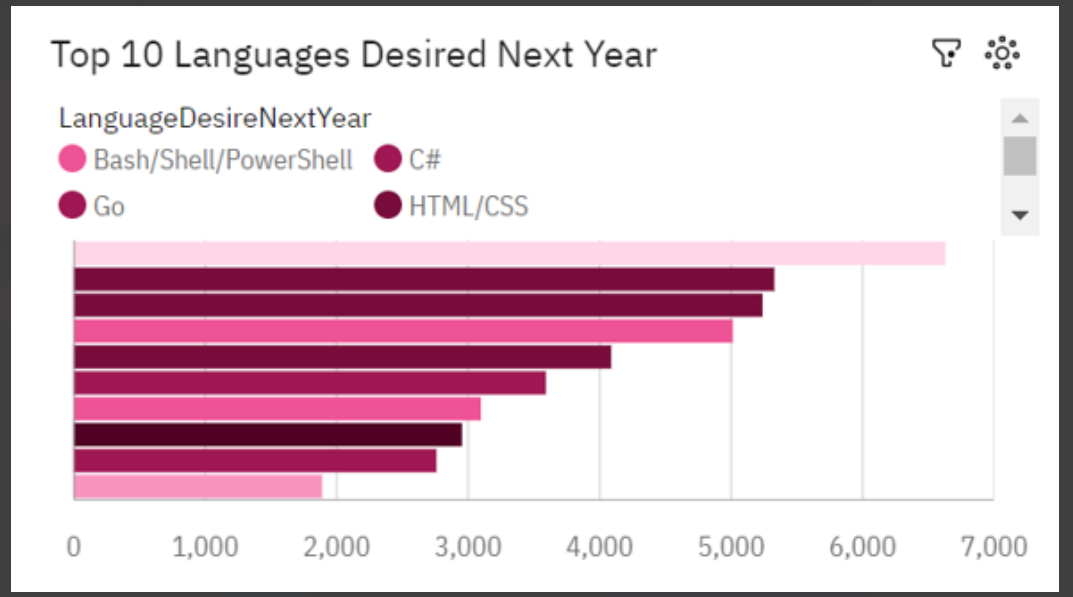
RESULTS

PROGRAMMING LANGUAGE TRENDS

CURRENT YEAR :



NEXT YEAR :



IBM Developer

SKILLS NETWORK

PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

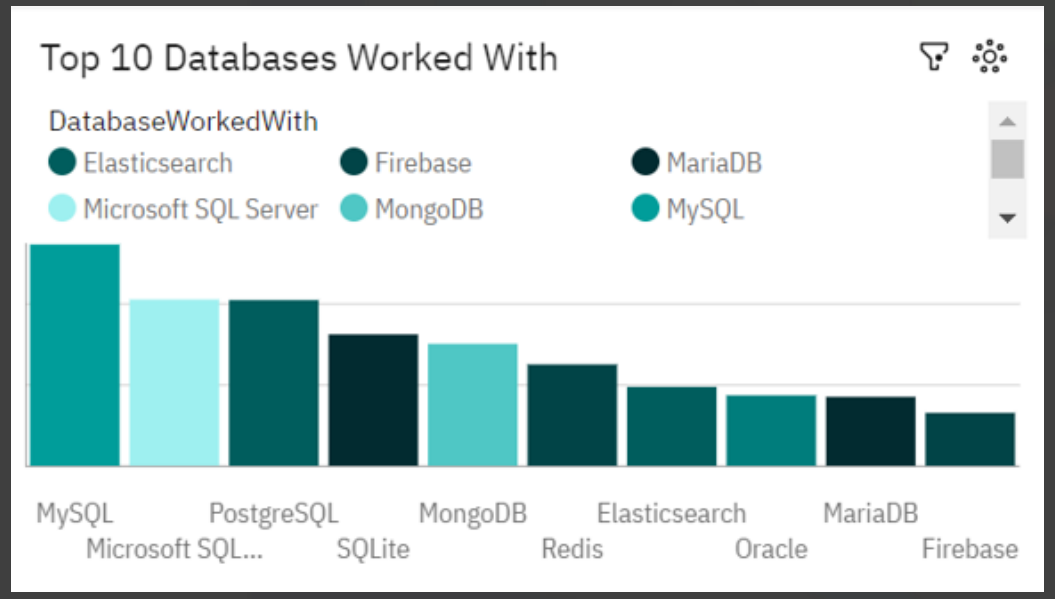
- JavaScript is the top programming language used currently with 8,687 respondents and C++ is at the bottom with 1,946 respondents.
- HTML/CSS, SQL, Python, Java, PHP are some of the other popular programming languages which are being used widely these days.
- JavaScript, HTML/CSS, SQLcontinue to be on top in the predictions for next year with a rise in demand for Python.

Implications

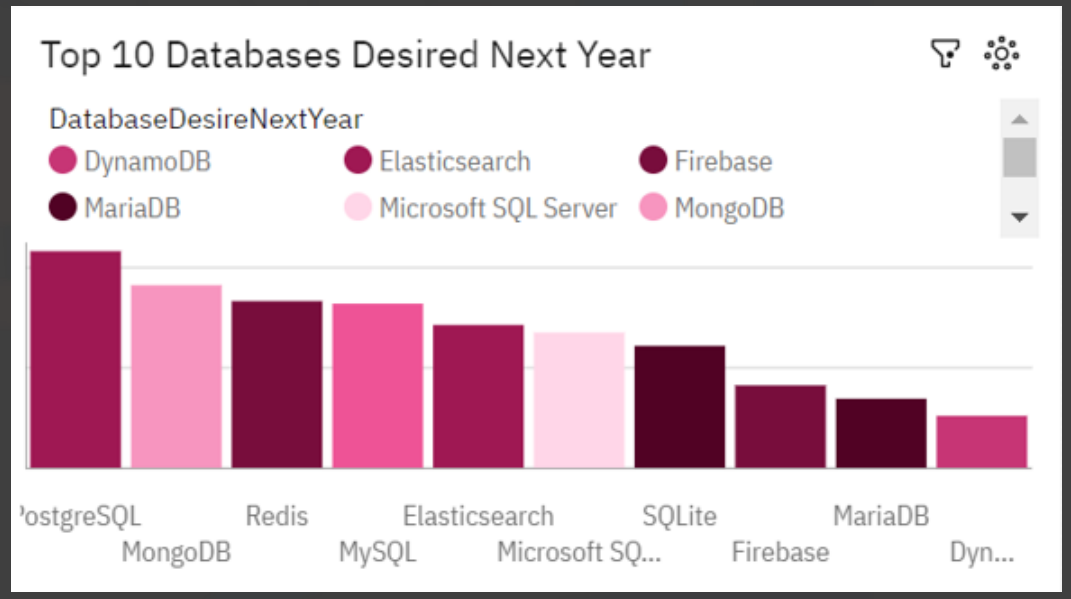
- The wide use of JavaScript as well as HTML/CSS languages worldwide implies huge demand for web developers in the present scenario.
- SQL and Python are also used by a large number of respondents indicating opportunities in the fields of Data Analytics, Data Science, Machine Learning and AI.
- C++ and PHP will likely be replaced by Go and Kotlin next year in the top 10 list.

DATABASE TRENDS

CURRENT YEAR :



NEXT YEAR :



IBM Developer

SKILLS NETWORK

DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- MySQL is the top database used currently with 5,469 respondents and Firebase is at the bottom with 1,314 respondents.
- Microsoft SQL Server, PostgreSQL, SQLite, MongoDB are some other popular databases.
- PostgreSQL, MongoDB and Redis move to the top of the list next year with a significant fall in the demand for MySQL database as well as Microsoft SQL Server.

Implications

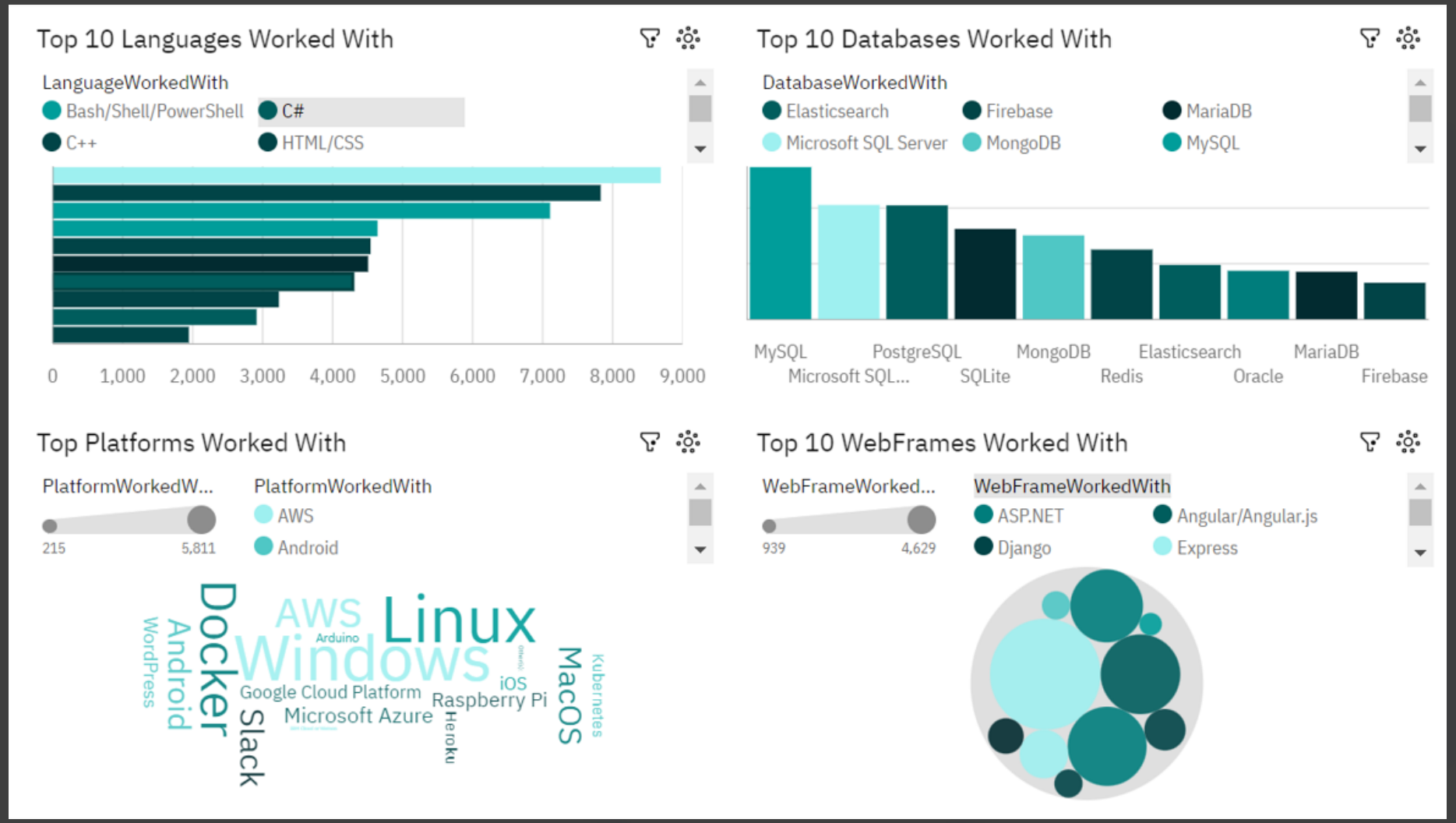
- The demand for currently less used databases will likely rise in the next year.
- Less scope for MariaDB and Firebase as per the current trends as well as future predictions.

DASHBOARD

Link to view the Dashboard –

https://eu2.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=.my_folders%2FData%2BAnalytics%2BCapstone%2BProject%2BDashboard&action=view&mode=dashboard&subView=model00000186747bfb95_00000000

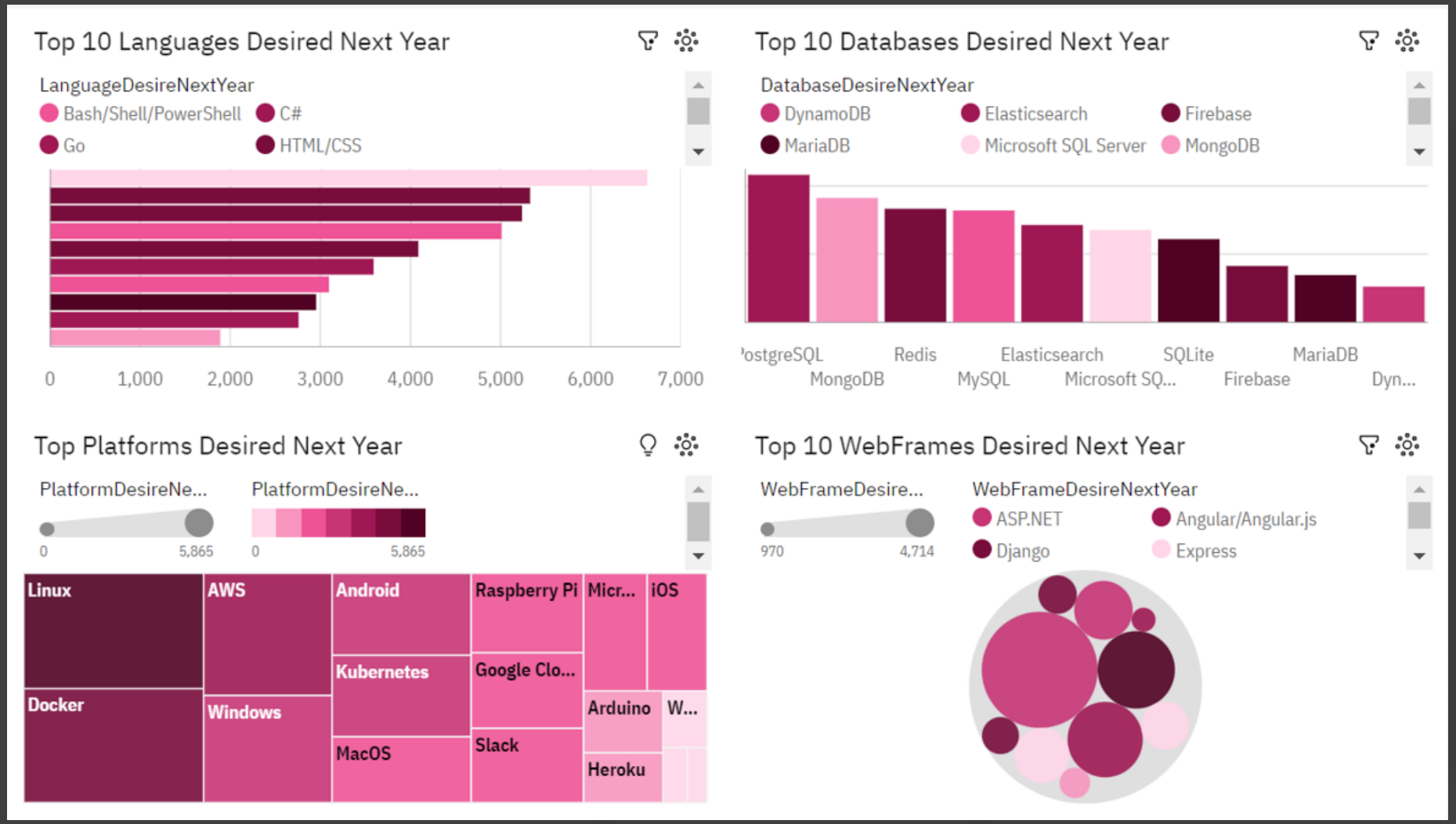
CURRENT TECHNOLOGY USAGE



IBM Developer

SKILLS NETWORK

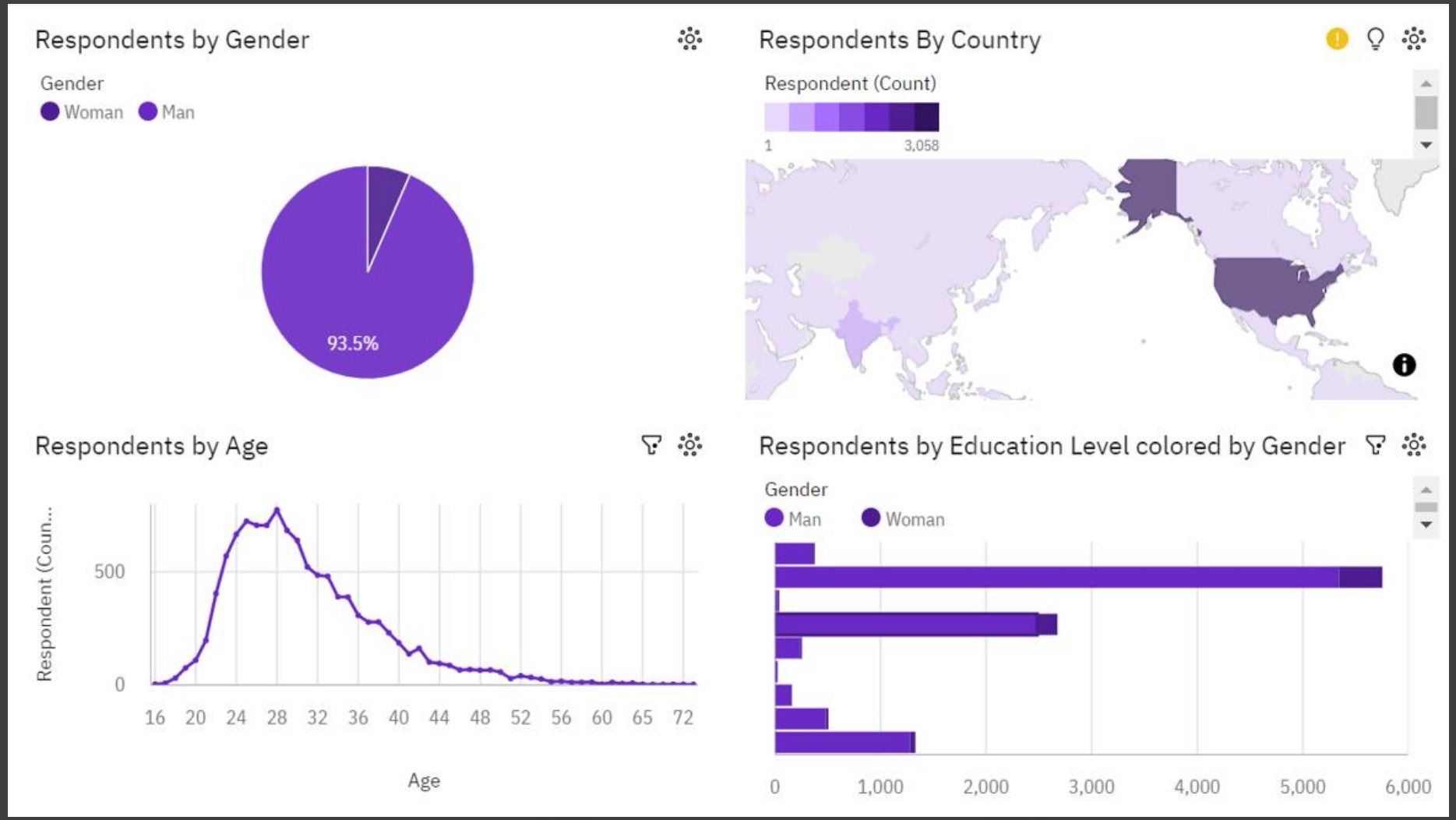
FUTURE TECHNOLOGY TREND



IBM Developer

SKILLS NETWORK

DEMOGRAPHICS



IBM Developer

SKILLS NETWORK

OVERALL FINDINGS & IMPLICATIONS

Findings

- JavaScript is the top language worked with currently as well as according to the next year trends.
- MySQL is the most popular database being used currently.
- Linux is the top platform worked with and continues to remain at top in the future.
- jQuery is the top WebFrame worked with currently.
- Majority of the respondents in this survey are men of age 28 years with a Bachelor's degree and maximum respondents are from the United States with a count of 3,058.

Implications

- MySQL will likely be replaced by PostgreSQL at the top position next year.
- jQuery will likely observe a huge fall in demand next year and be replaced by React.js as the top desired WebFrame.
- Only 6.5% of the respondents are women indicating a drastic decrease in the amount of women working in the fields of Technology, Computer Science etc. thus making them male-dominated fields.

CONCLUSION

- JavaScript and Linux are in demand currently as well as in the near future.
- PostgreSQL and React.js are some of the technologies that will experience a rise in demand next year and should be used currently so as to keep up.
- A significant shortage of women is evident in the Technology sector and is something which needs to be laid emphasis upon.

APPENDIX

- **Data Collection using API (Source)** - https://www.kaggle.com/promptcloud/jobs-on-naukricom
- **Data Collection using Webscraping (Source)** - https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DA0321EN-SkillsNetwork/labs/datasets/Programming_Languages.html
- **Exploratory Data Analysis** - https://jupyterlab-1-labs-prod-jupyterlab-us-east-0.labs.cognitiveclass.ai/hub/user-redirect/lab/tree/labs/module%203/M3ExploratoryDataAnalysis-lab.ipynb
- **Data Visualization** - https://jupyterlab-1-labs-prod-jupyterlab-us-east-0.labs.cognitiveclass.ai/hub/user-redirect/lab/tree/labs/module%204/Data%20Visualisation.ipynb

IBM Developer

SKILLS NETWORK