IBM CAPSTONE PROJECT

Aruna Rakesh

09/16/2024

OUTLINE

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

IBM Developer

SKILLS NETWORK

EXECUTIVE SUMMARY

- About the Project: Data Science Project for a global IT and consulting services firm to identify on demand emerging skills(Technologies, Programming Languages, Databases) to remain competitive in the market.

Recommended Solution:

- Collecting Jobs Data from different sources
  - Find number of jobs for different locations for key skills across US
  - Identify Key skills and Pay range
  - Use different data formats to store the collected data

- Explore collected data and apply data wrangling techniques
  - Identify data columns, type
  - Identify duplicates, missing values in data
  - Normalize data

- Analyze data

  Identify frequency of different values in the collected data.
  - Identify outliers in the job data
  - Identify correlation between different data points.

- Visualize the findings of your data
  - Create different plots and charts to show distribution in the job data
  - Explain the relationship in the data
  - Compare different charts and arrive at your findings

- Create Dashboards
  - Categorize current technologies report, future technologies and demographics for the data.

IBM Developer

SKILLS NETWORK

INTRODUCTION

- Main purpose of the IBM Capstone project is about identifying demanding skills technologies, programming languages, databases, platforms and web-frames in the rapidly changing market.

- IBM Capstone project is for any business consulting firms who would like to stay up to date and keep pace with changing skills.

- What is the nature of the analysis and what information we gain from it? Nature of analysis includes analyzing and understanding demographics of survey data and handful of new technology options for developers.

- By reading this report, any business consulting firm will gain insights on the latest technologies, programming languages, database skills organizations require for their advancement and business needs.

- By looking into this report, skill gaps can be identified and required training programs can be developed by the organizations.

- By looking at the current and future trend data skills, organizations can identify the different places to recruit for their talents.

IBM Developer

SKILLS NETWORK

METHODOLOGY
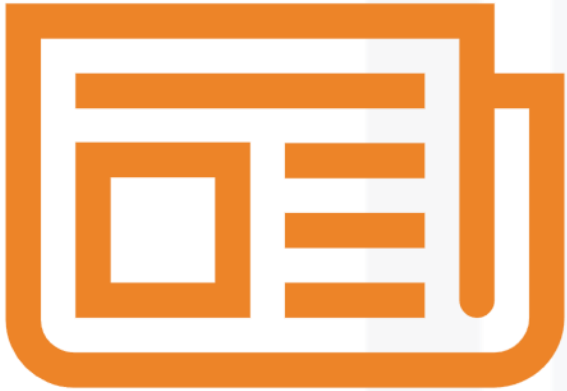
- IBM Capstone project collected data from online surveys called Kaggle Survey 2023 data. Survey data is about information collected from software professions across locations.

- Accessed latest jobs data using GitHub jobs API which uses JSON over HTTP to query online data.

- Explored data by extracting information from the websites using web scrapping by importing python Beautiful Soup library, pandas Dataframe.

- Using Data Wrangling techniques removed duplicates, missing values from the data.

- Conducted exploratory data analysis to plot the distribution of data, examining data to uncover patterns, relationships and insights. Implementing using python libraries such as matplotlib, seaborn.

- Implemented visualization charts, plots as Dashboards using IBM Cognos Analytics, Google Looker Studio to make data driven decisions.

- Calculate summary statistics (mean, median, mode) to understand central tendencies of data. Normalize data for data transformation and further analysis.

- Document the data findings, insights during analysis and present it to key stakeholders to reproduce results.

IBM Developer

SKILLS NETWORK

RESULTS

Data: Online survey of Respondents top choices of skills they are worked with in the current year and desiring to work with in the next year.

Category for Analysis: Top 10 Skills for (Languages, Database, Platform, Web-Frame) for Current year, Next year and Respondents Gender and Age.
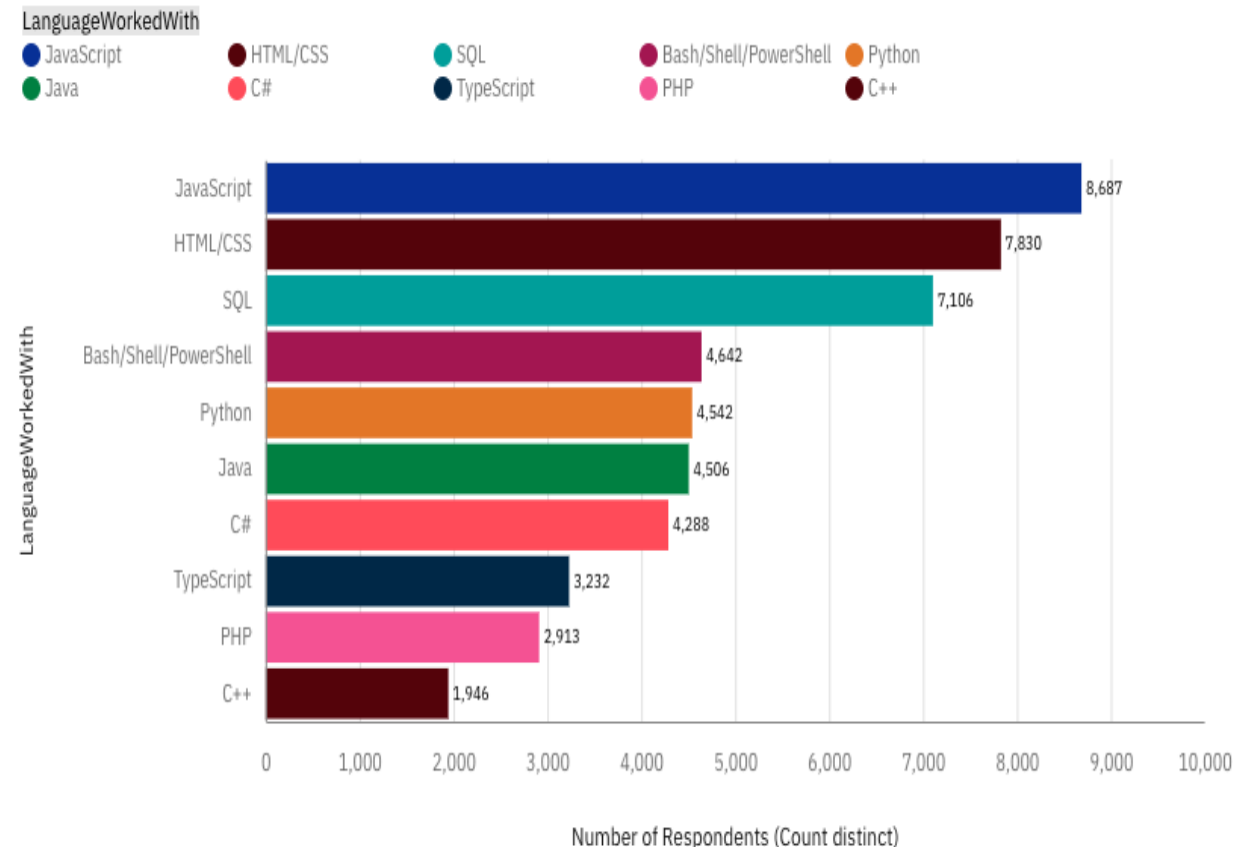
Insights and Trend observed from top 10 skills:

• Among the languages, JavaScript, HTML/CSS continues to be the top 2 popular programming languages. Python is one of the languages that has gained high importance and a strong desire among the respondents. SQL continues to retain its spot among the top 5 desired languages. GO and Kotlin are few other languages along with Typescript has increased importance from the current year and could be in high demand when compared to C# and Java.

• Among the databases, PostgreSQL, MongoDB, Redis, Elasticsearch and Firebase are the databases desired by the respondents over the current year, though the MySQL and Microsoft SQL server are the top 2 databases worked with in the current year. Approximately 5% of surveys increase for the PostgreSQL, 20% for MongoDB, 25% for Firebase, 32% increase for Redis and 42% increase for Elasticsearch.

• Linux, Windows, Docker and AWS are among the top choices for platforms worked with currently and desired for the future by the respondents from the survey. Followed by MacOS, Android, Google Cloud platform and Rasberry-Pi.

• Angular.js, React.js, ASP.net, Vue.js, JQuery were among the top 5 Web-frames of choice from the survey, with Vue.js gaining the top spot by more than 100% survey count.

• People contributing to the survey are spread across the globe with US and Canada being the highest count of 3058 respondents. Men are among the top contributors in the survey.
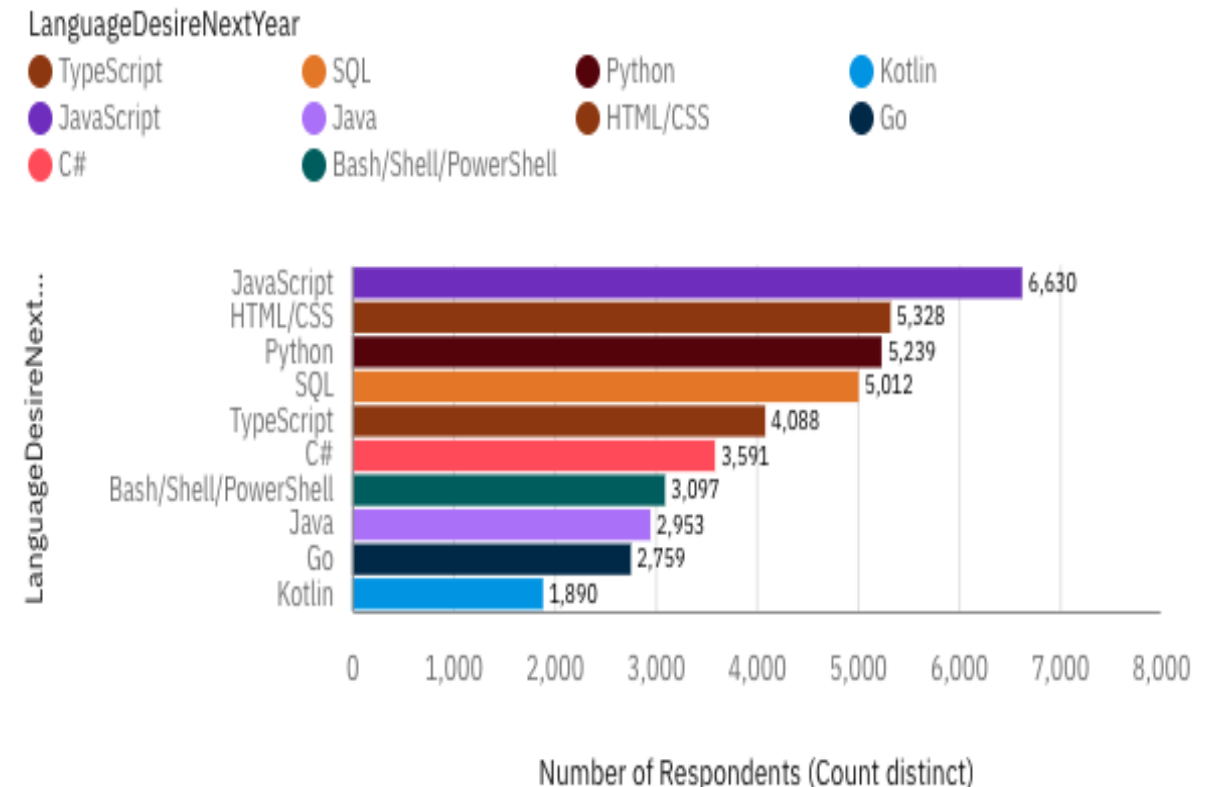
IBM Developer

SKILLS NETWORK

PROGRAMMING LANGUAGE TRENDS

Current Year

Next Year

PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

- JavaScript, HTML/CSS continues to be the top 2 programming languages, even though the count of respondents desired next year has dropped from the current year count of respondents worked with the languages.

- Python is one of the languages that gained importance among the respondents desiring next year in the future trend usage report.

- Count of respondents desiring to work with Typescript has increased from the current year when compared to C# and Java.

- Few languages which are worked with in the current year like PHP and C++ are not among the top 10 languages desired next year and taken over by languages like GO and Kotlin.
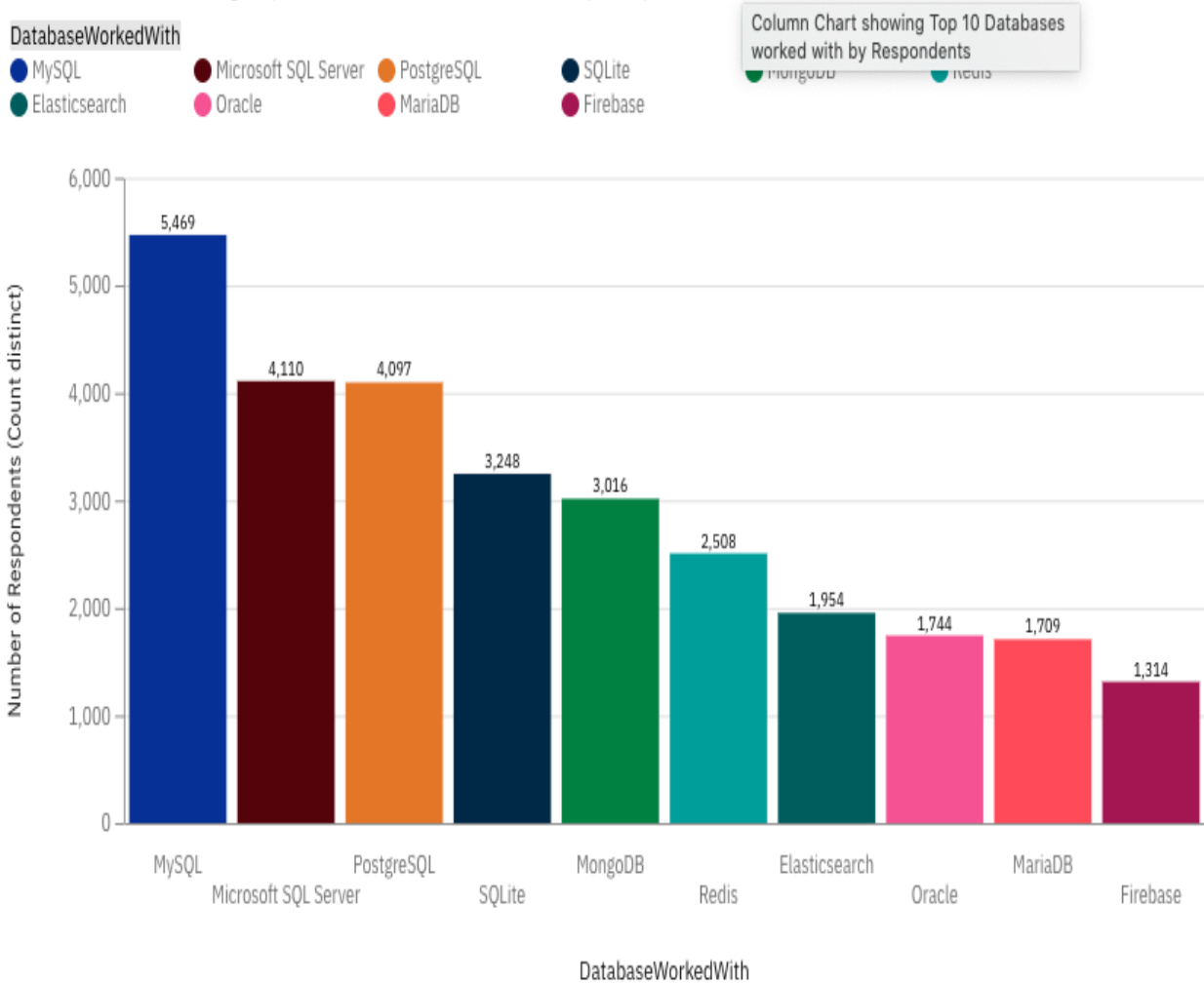
Implications

- Survey Respondents show their desire to the languages JavaScript, HTML/CSS as they retain their ranking in the top 2 spots.

- From the bar chart for top 10 languages desired next year, it clearly implies Python is that one language to reach the top spots with more than 5000 survey respondents desiring to work with it. Business should consider Python to be the valuable skill requirement to remain competitive in the market.

- Next language to keep a watch on from the findings is TypeScript which has gained a desire of 26% increase in surveys.

- C++ and PHP have lost their ranks to two other languages GO and Kotlin which can be analyzed further to gain more understanding about its scalability and Versatility.
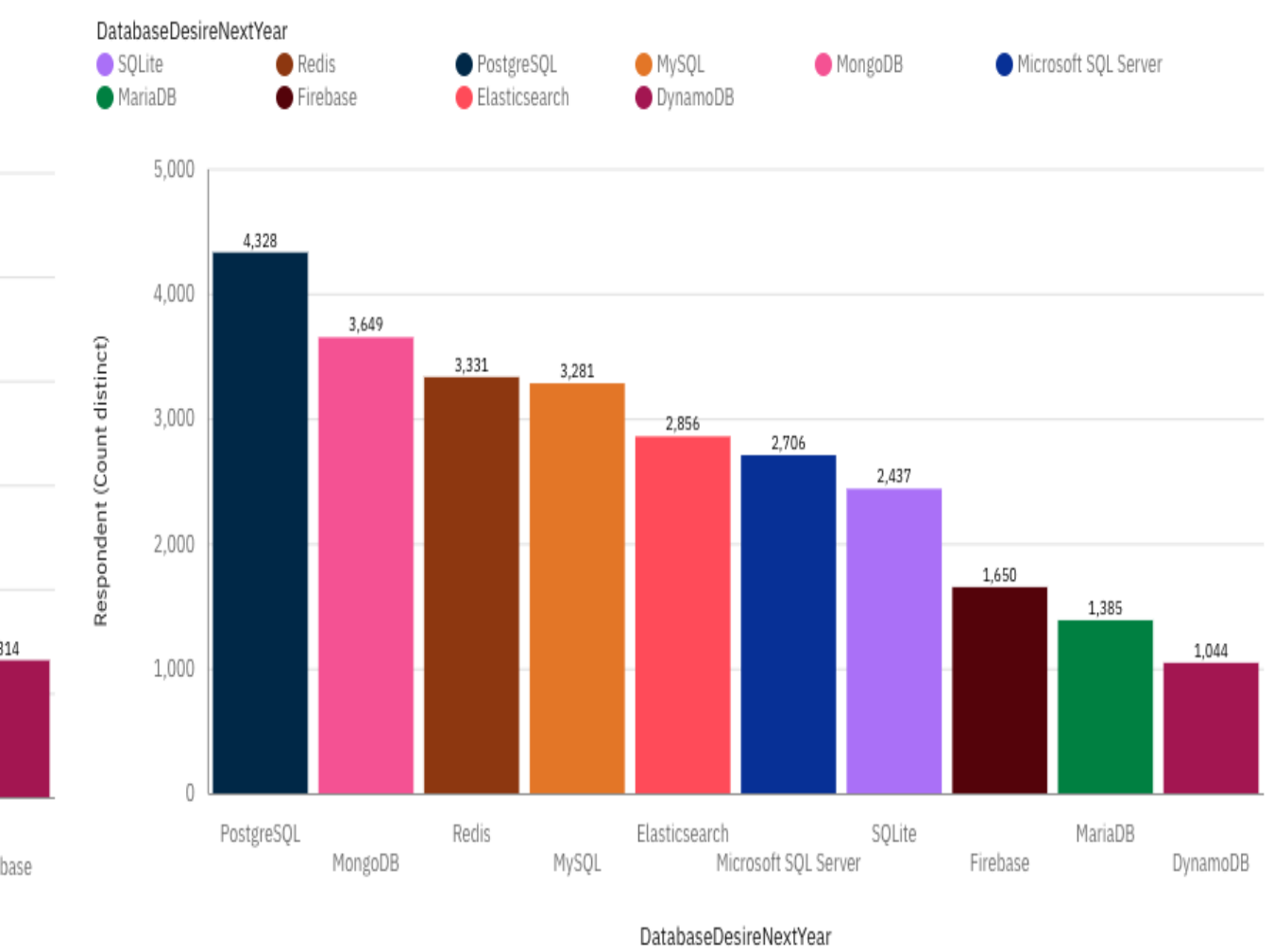
DATABASE TRENDS

Current Year

Next Year

Column Chart showing Top 10 Databases worked with by Respondents

DatabaseWorkedWith
● MySQL ● Microsoft SQL Server ● PostgreSQL ● SQLite ● MongoDB ● Redis
● Elasticsearch ● Oracle ● MariaDB ● Firebase

Number of Respondents (Count distinct)

Database	Count
MySQL	5,469
Microsoft SQL Server	4,110
PostgreSQL	4,097
SQLite	3,248
MongoDB	3,016
Redis	2,508
Elasticsearch	1,954
Oracle	1,744
MariaDB	1,709
Firebase	1,314

DatabaseWorkedWith

Column Chart showing Top 10 Database Desired Next Year By Respondents

DatabaseDesireNextYear
● SQLite ● Redis ● PostgreSQL ● MySQL ● MongoDB ● Microsoft SQL Server
● MariaDB ● Firebase ● Elasticsearch ● DynamoDB

Respondent (Count distinct)

Database	Count
PostgreSQL	4,328
MongoDB	3,649
Redis	3,331
MySQL	3,281
Elasticsearch	2,856
Microsoft SQL Server	2,706
SQLite	2,437
Firebase	1,650
MariaDB	1,385
DynamoDB	1,044

DatabaseDesireNextYear

IBM Developer

SKILLS NETWORK

DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- MySQL, Microsoft SQL Server being the top 2 Databases used by Respondents during the current year as per the survey, lost the count of respondents desiring to work with next year by 35% - 40%. This made the PostgreSQL and MongoDB to take the first two spots for the future desired language.

- PostgreSQL, MongoDB, Redis, Elasticsearch and Firebase are the databases desired by the respondents over the current year. Approximately 5% of surveys increase for the PostgreSQL, 20% for MongoDB, 25% for Firebase, 32% increase for Redis and 42% increase for Elasticsearch.

- Count of respondents desiring to work with Oracle has reduced and is not among the top 10 databases list for next year.

- Sqlite and MariaDB are also among the databases to have reduced desire count from survey respondents along with MYSQL and Microsoft SQL server.

Implications

- MYSQL, Microsoft SQL server and PostgreSQL are the top popular databases from the survey collected.

- From the bar chart for top 10 databases desired next year, PostgreSQL continues to be in the top with MongoDB, Elasticsearch, Redis and Firebase gaining its ranking to be among the desired databases.

- Apart from the survey collected, it is implied that though we have top 10 databases, it also depends on the type of application and data requirements upon which the businesses should decide on their databases.

- Scalability and performance should be considered when choosing between SQL or NoSQL databases. MongoDB, Redis are among the top 10 NoSQL databases desired by survey respondents.

IBM Developer

SKILLS NETWORK
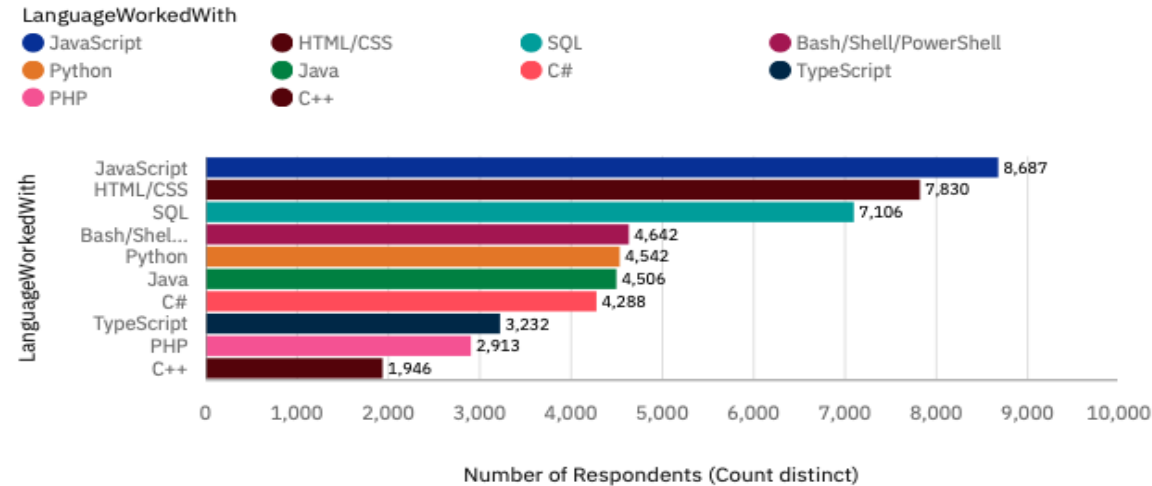
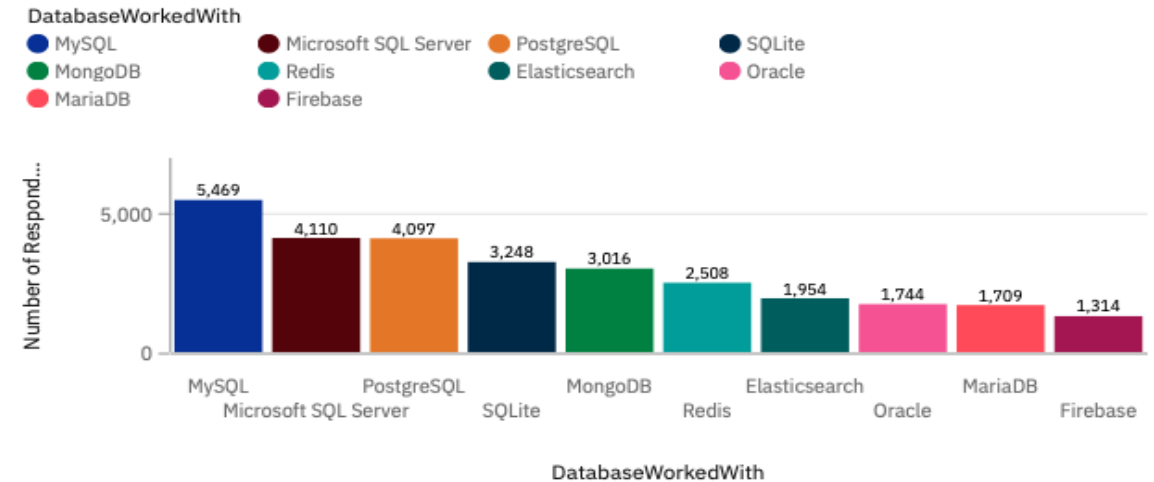# DASHBOARD

https://github.com/arunarakesh/DashboardRepo

DASHBOARD-CURRENT TECHNOLOGY USAGE

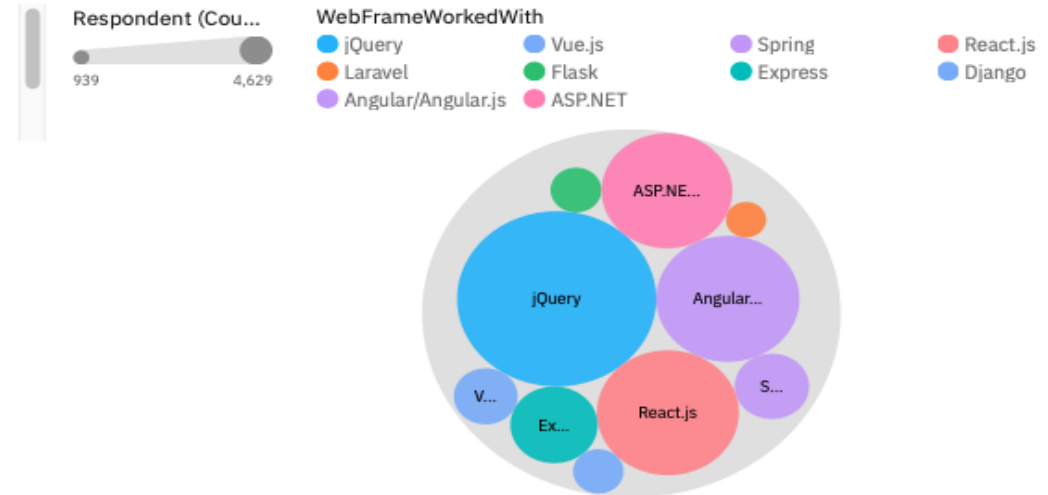Bar Chart Showing Top 10 Languages worked with by Respondents

Column Chart showing Top 10 Databases worked with by Respondents

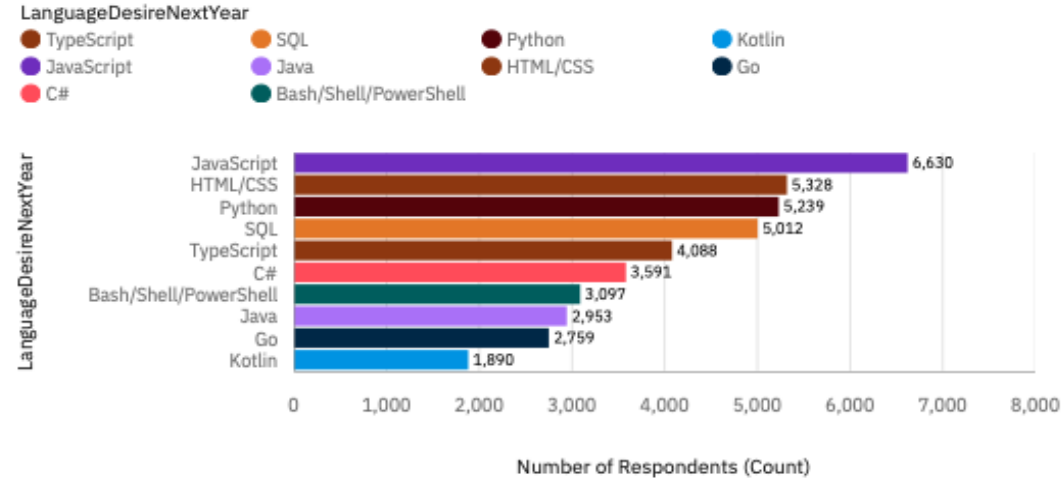WordCloud displaying different Platforms Worked with by Respondents

Hierarchy Bubble Chart displaying Top 10 WebFrame worked with by Respondents
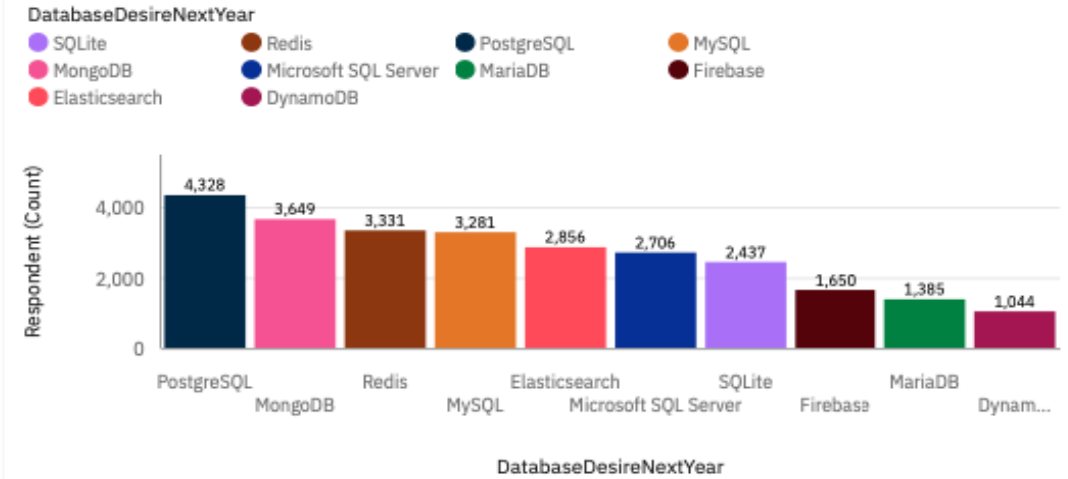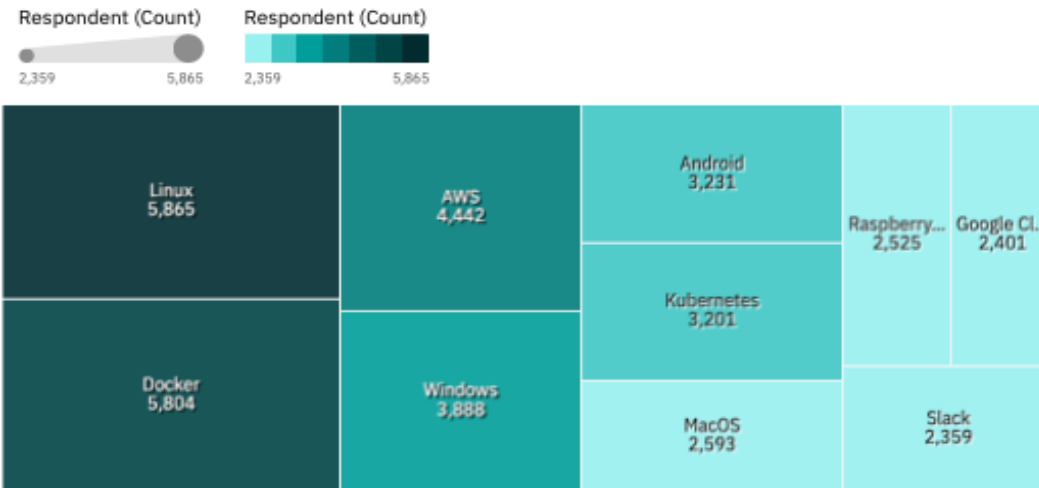
IBM Developer

SKILLS NETWORK

DASHBOARD-FUTURE TECHNOLOGY TREND

Bar Chart showing Top 10 Languages Desired Next Year By Respondents

LanguageDesireNextYear

- TypeScript
- JavaScript
- C#
- SQL
- Java
- Bash/Shell/PowerShell
- Python
- HTML/CSS
- Kotlin
- Go

LanguageDesireNextYear:

- JavaScript — 6,630
- HTML/CSS — 5,328
- Python — 5,239
- SQL — 5,012
- TypeScript — 4,088
- C# — 3,591
- Bash/Shell/PowerShell — 3,097
- Java — 2,953
- Go — 2,759
- Kotlin — 1,890

Number of Respondents (Count)

Column Chart showing Top 10 Database Desired Next Year By Respondents

DatabaseDesireNextYear

- SQLite
- MongoDB
- Elasticsearch
- Redis
- Microsoft SQL Server
- DynamoDB
- PostgreSQL
- MariaDB
- MySQL
- Firebase

Respondent (Count):

- PostgreSQL — 4,328
- MongoDB — 3,649
- Redis — 3,331
- MySQL — 3,281
- Elasticsearch — 2,856
- Microsoft SQL Server — 2,706
- SQLite — 2,437
- Firebase — 1,650
- MariaDB — 1,385
- Dynam... — 1,044

DatabaseDesireNextYear

Tree Map of Platforms Desire Next Year by Respondents

Respondent (Count): 2,359 — 5,865

- Linux 5,865
- Docker 5,804
- AWS 4,442
- Windows 3,888
- Android 3,231
- Kubernetes 3,201
- MacOS 2,593
- Raspberry... 2,525
- Google Cl... 2,401
- Slack 2,359

Hierarchy Bubble Chart showing WebFrame Desired Next Year by Respondents

Respondent (Count): 970 — 4,714

WebFrameDesireNextYear

- jQuery
- React.js
- Angular/Angular.js
- Vue.js
- Flask
- ASP.NET
- Spring
- Express
- Ruby on Rails
- Django

Bubbles: React.js, Vue.js, Angula..., ASP.N..., Ex..., jQu..., D..., S...
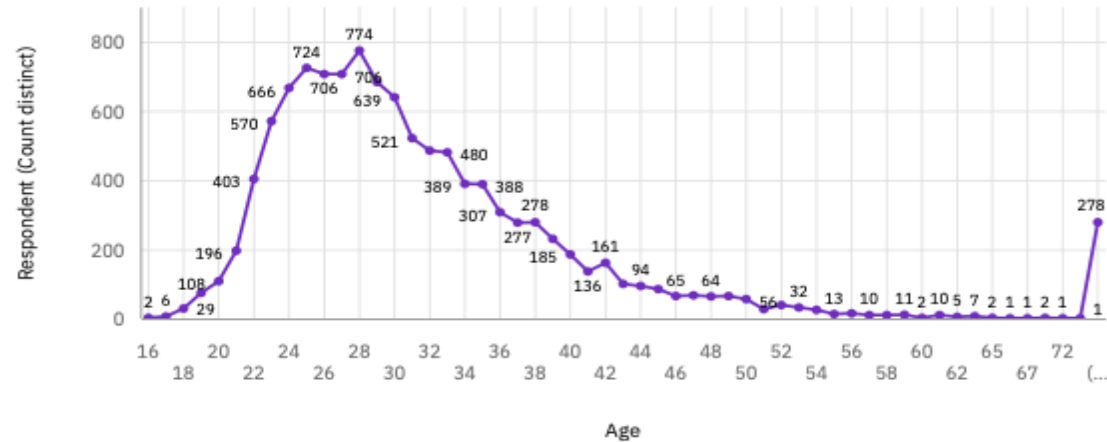
IBM Developer

SKILLS NETWORK
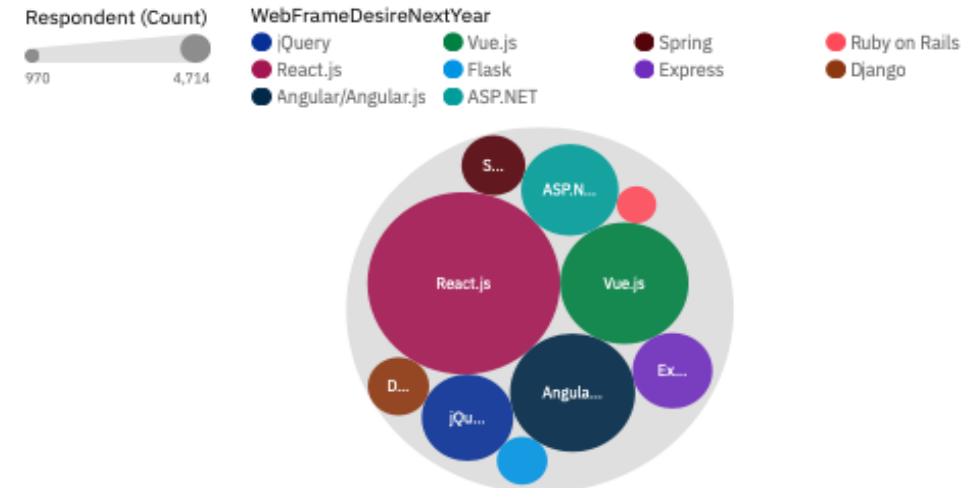
DASHBOARD-DEMOGRAPHICS

Pie Chart Capturing Respondent classified by Gender

Gender
- Man
- Woman

6.5%

93.5%
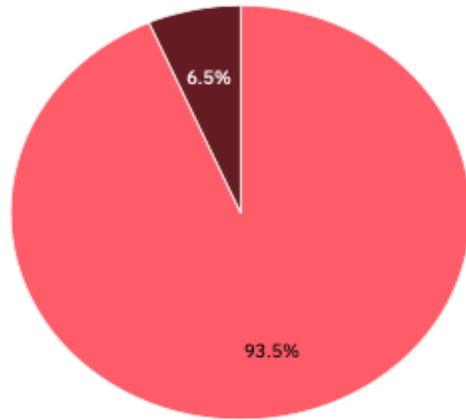
Line Chart showing Respondent Count by Age

Map Chart displaying Respondent Count for Countries

Respondent (Cou...

1   3,058

Norway   Russian Federation

Canada   Latvia
                Austria

United States   Algeria Israel   South Korea
Mexico   Nepal   Taiwan

        Côte d'Ivoire Somalia   Indonesia
Ecuador

Paraguay   South Africa   Australia
Argentina

© Mapbox © OpenStreetMap

Stacked Bar Chart to show Respondent Count by Gender, classified by Formal Education Level

EdLevel - Gender
- (no value) | Man
- Associate degree | Man
- Bachelor's degree (BA, BS, B.Eng....
- I never completed any formal edu...
- (no value) | Woman
- Associate degree | Woman
- Bachelor's degree (BA, BS, B.Eng....
- I never completed any formal edu...

Man: 5,341   2,484   482   1,280

Woman

0   1,000 2,000 3,000 4,000 5,000 6,000 7,000 8,000 9,000 10,000 11,000

Number of Respondents (Count distinct)
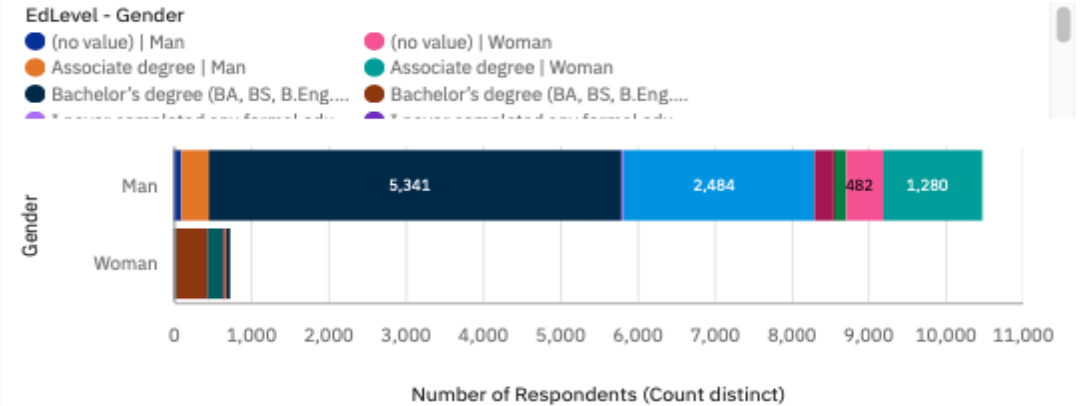
DISCUSSION

- Summary of the findings: JavaScript, PostgreSQL, React.js and Linux are the most desired skills for the next year from the surveys collected from respondents who hold either a bachelors or master's degree education level.
- Comparing the results from the current year to the next year survey, JavaScript is the most desired language with Python gaining popularity and PostgreSQL is the most desired database on top of MySql and Microsoft SQL server. Linux remains the top platform. React.js tops the web-frame desired list preferred over Jquery which is most used in the current year.

- Majority of Respondents who contributed to the survey in the age group of 25 and 35 hold an educational level of Bachelors or Masters degree.
- Removing duplicate records and eliminating Null values added more reliability to the survey data.
- Based on the findings, businesses can hire talents in the top skill areas.
- Constantly looking into and identifying the top skills will broaden the pool of available talent.
- Embracing hiring strategies and constantly training hiring managers about the skills required will benefit the IT firm in having smooth and efficient hiring process.

IBM **Developer**

SKILLS NETWORK

OVERALL FINDINGS & IMPLICATIONS

Findings

- JavaScript, HTML/CSS continues to be the top 2 programming languages, even though the count of respondents desired next year has dropped from the current year. Python is one of the languages that gained importance desiring next year. GO and Kotlin are also among the desired languages.

- MySQL, Microsoft SQL Server being the top 2 Databases used by Respondents during the current year lost the survey count desiring to work with next year by 35% - 40%. PostgreSQL, MongoDB, Redis, Elasticsearch and Firebase are the databases desired by the respondents over the current year.

- Linux, Windows, Docker and AWS are among the top choices for platforms worked with currently and desired for the future by the respondents from the survey. Followed by MacOS, Android, Google Cloud platform and Rasberry-Pi.

- Angular.js, React.js, ASP.net, Vue.js, JQuery were among the top 5 Webframes of choice from the survey, with Vue.js gaining the top spot by more than 100% survey count.

- People contributing to the survey are spread across the globe with US and Canada being the highest count of 3058 respondents. Men are among the top contributors in the survey.

Implications

- Programming languages like JavaScript, HTML/CSS and Python are among the top skills required and desiring to work with in the future. TypeScript has gained a desire of 26% increase in surveys. GO and Kotlin which can be analyzed further to gain more understanding about its scalability and Versatility.

- IT firm Business should consider hiring people with database skills like MYSQL, Microsoft SQL server and PostgreSQL to be their top popular databases along with a close watch on MongoDB, Elasticsearch, Redis and Firebase based on the projects data requirements and needs of their applications.

- Considering the top 10 platform skills, incorporating people with expertise in Linux, Windows, Docker and AWS will add more value to the talent resource of the IT business talent based recruitment.

- Business could benefit from identifying and hiring top Web-frame skills like Angular.Js, React.js, ASP.net, Vue.js and JQuery.

- Majority of Respondents who contributed to the survey in the age group of 25 and 35 hold a educational level of Bachelors or Master's degree.

IBM Developer

SKILLS NETWORK

CONCLUSION

- JavaScript, HTML/CSS and Python top the list of most used languages in both the years. Python is also the most desired language in the next year.
- PostGreSQL, MongoDB, Redis, ElasticSearch and Firebase top the list of most desired databases and Linux, Docker and AWS tops the list of desired platforms for the next year. React.js top the list of desired web frame.
- Hiring based on top skills will help organizations find the right talent and will open more opportunities to a large population of potential employees.
- Identifying top skills and hiring IT consultants with expertise will benefit the firm to remain competitive in the job market and benefit organizations looking to hire the right talent for their businesses.
- It can help businesses getting their job done easily more than considering professionals holding a degree.
- Knowing the top skills required also helps with properly planned training and development programs in an organization.
- Constantly looking into and identifying the top skills will broaden the pool of available talent.
- Embracing hiring strategies and constantly training hiring managers about the skills required will benefit the IT firm in having smooth and efficient hiring process.

APPENDIX

Pie chart of the top 5 databases that respondents wish to learn next year
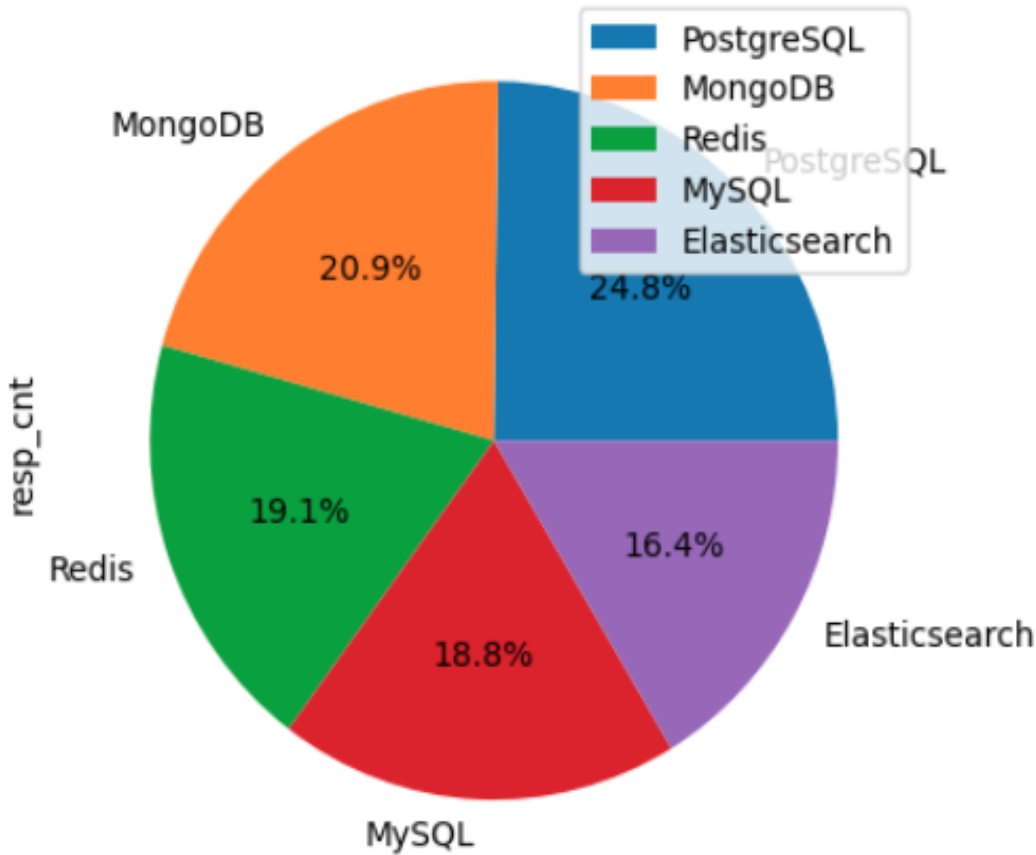
stacked chart of median WorkWeekHrs and CodeRevHrs for the age group 30 to 35
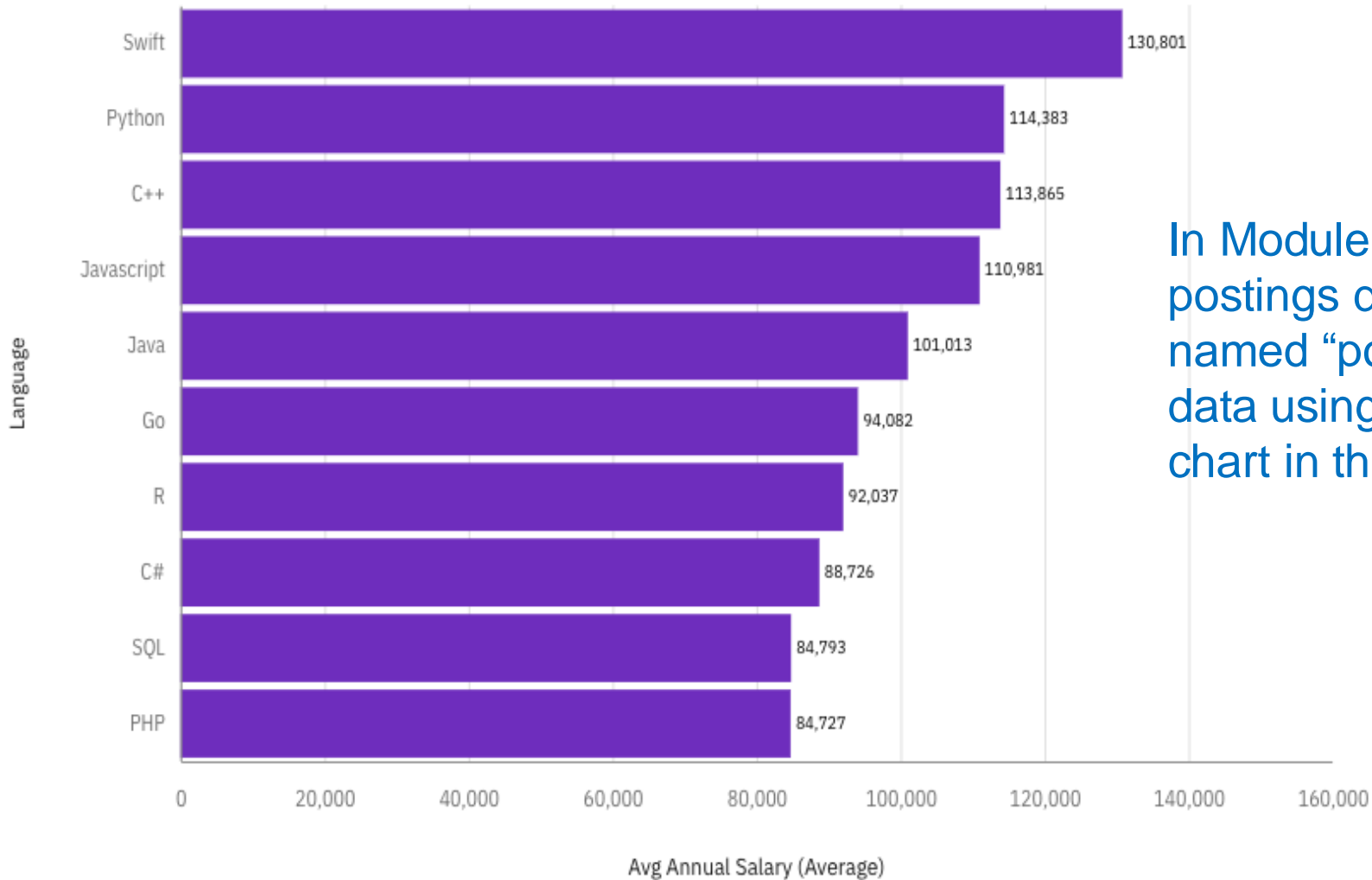

IBM Developer

SKILLS NETWORK
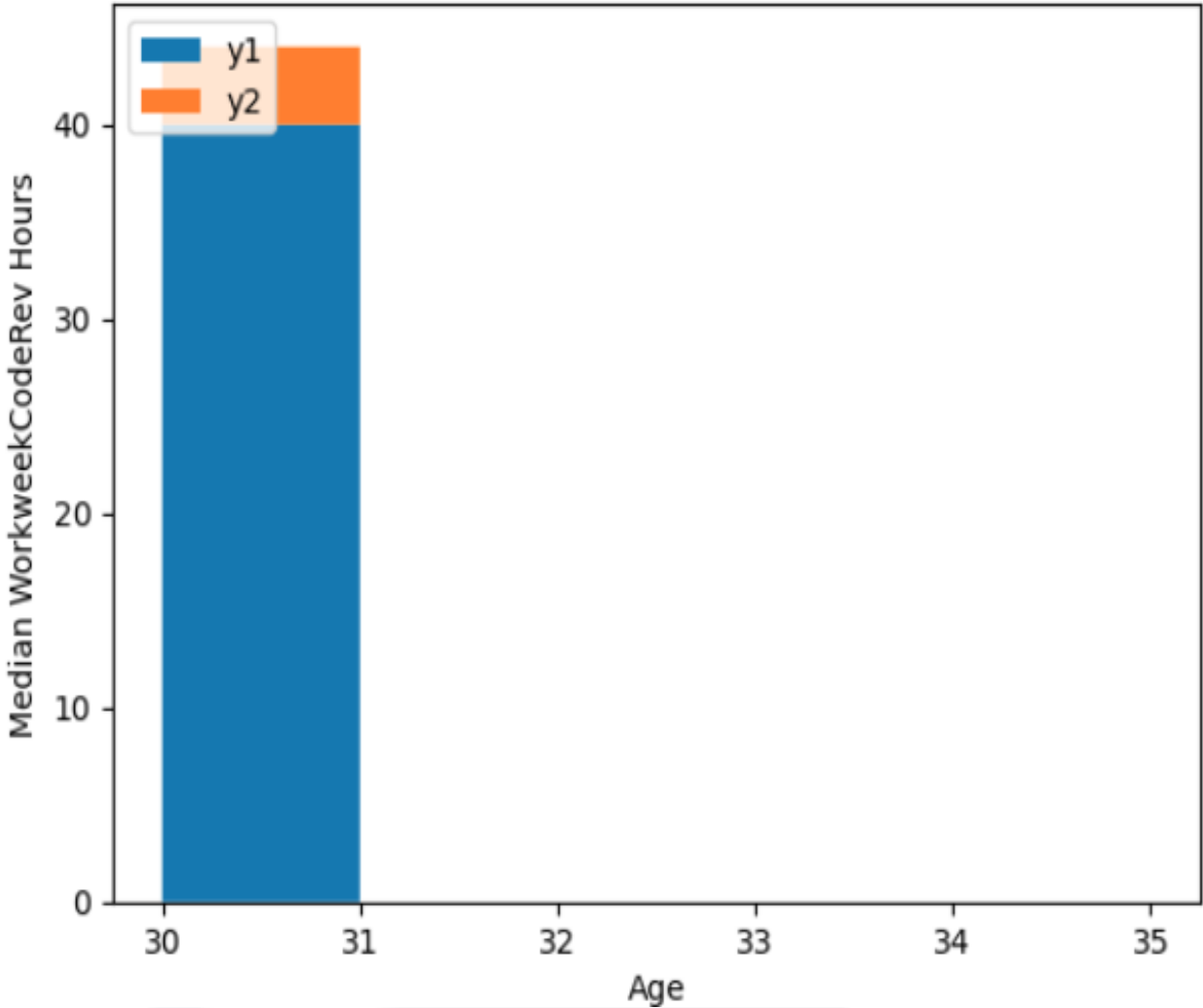
JOB POSTINGS

In Module 1 you have collected the job posting data using Job API in a file named "job-postings.xlsx". Present that data using a bar chart here. Order the bar chart in the descending order of the number of job postings.

Bar Chart showing Languages and Number of Job Postings
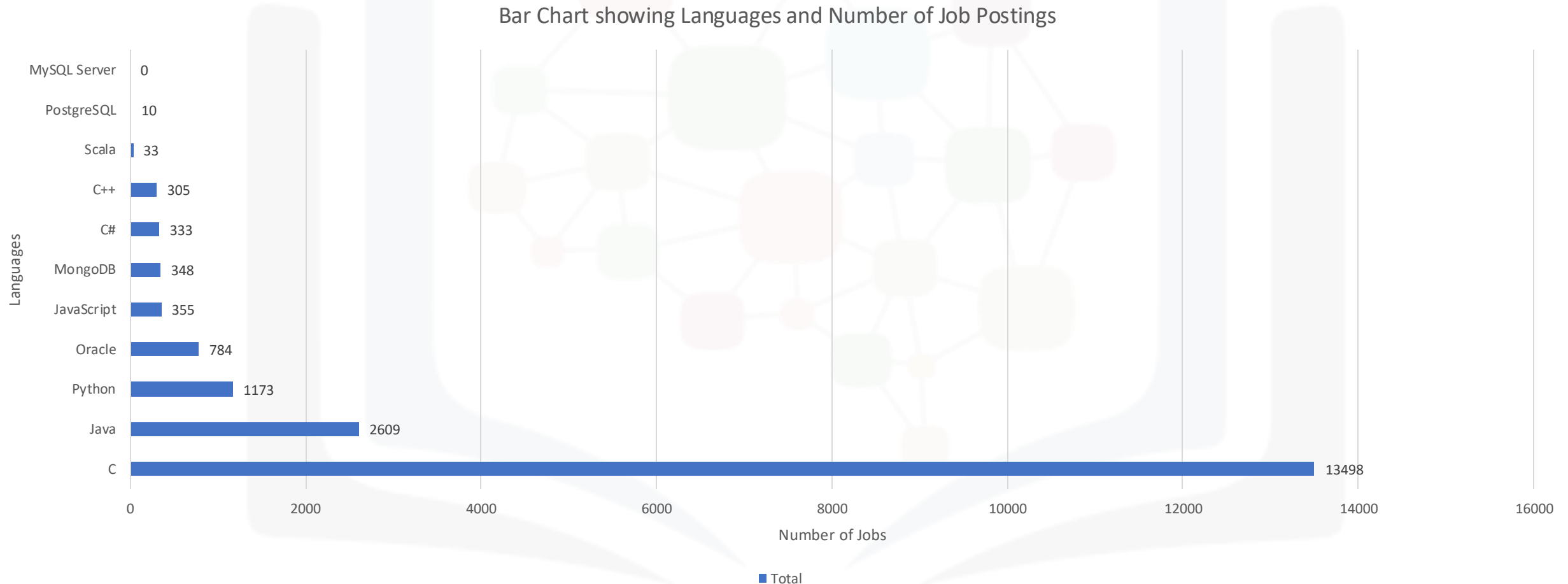
POPULAR LANGUAGES

Bar Chart showing Average Annual Salary by Programming 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.
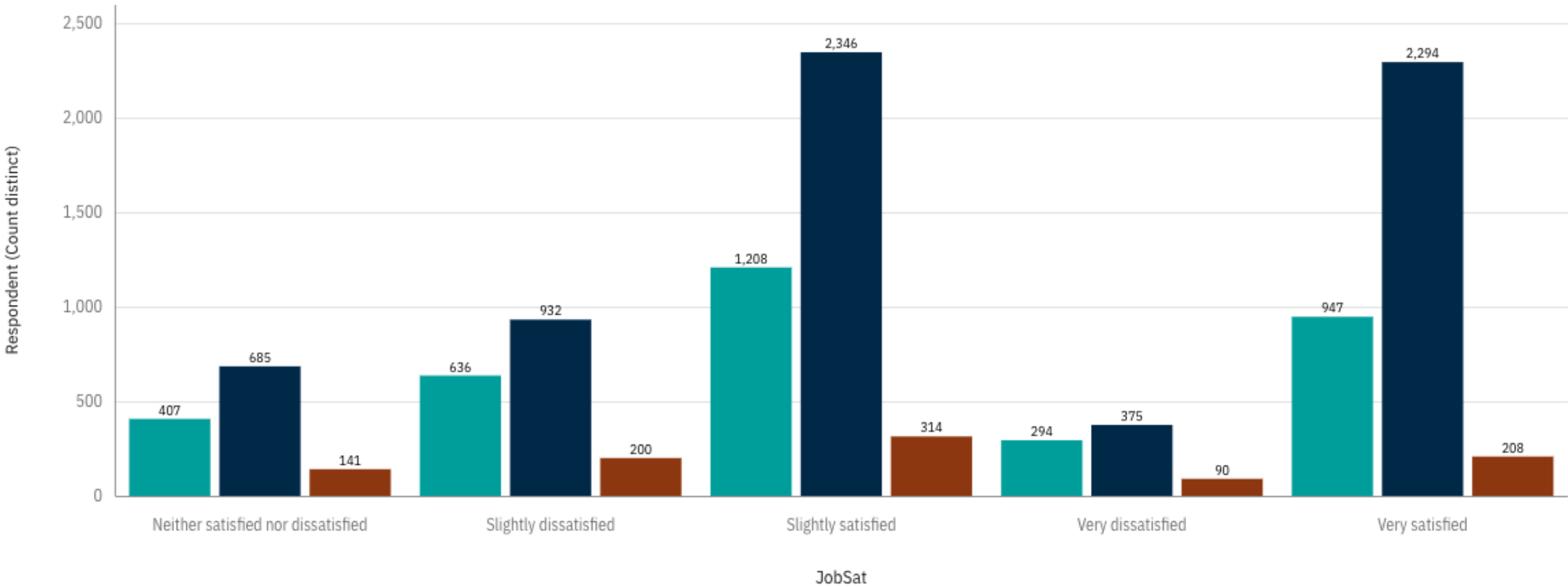
Further Value Added Analysis

Column Chart – Count of Respondent Job Satisfaction based on their work Location
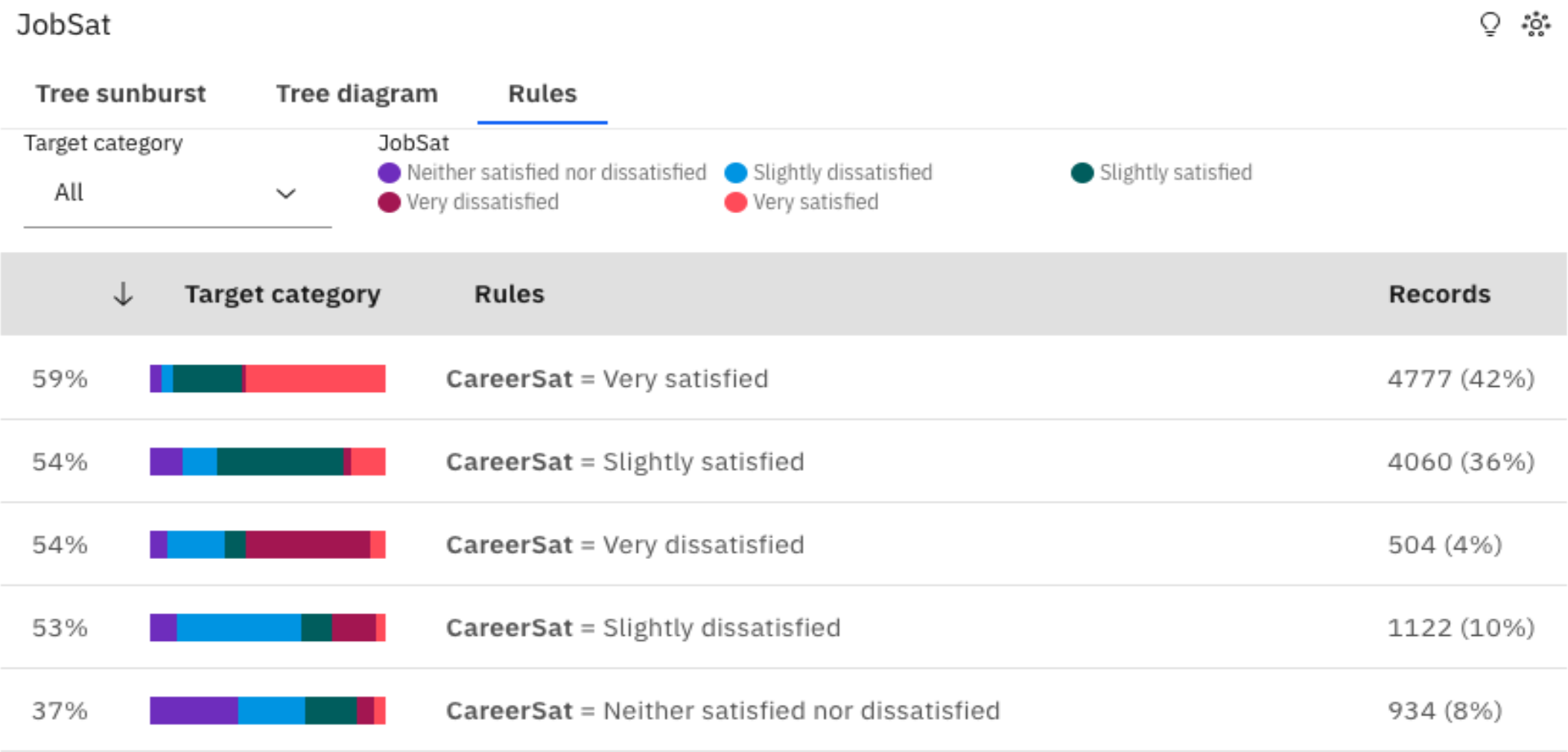

Respondent by JobSat colored by WorkLoc

WorkLoc
● Home   ● Office   ● Other place, such as a coworking ...

Respondent by JobSat colored by WorkLoc

Further Value Added Analysis

Tree Chart – Respondent Percentage based on Career Satisfaction from the survey dataset

JobSat

| Tree sunburst | Tree diagram | **Rules** |

Target category

JobSat
- Neither satisfied nor dissatisfied
- Slightly dissatisfied
- Slightly satisfied
- Very dissatisfied
- Very satisfied

All ⌄

| ↓	Target category	Rules	Records
59%		**CareerSat** = Very satisfied	4777 (42%)
54%		**CareerSat** = Slightly satisfied	4060 (36%)
54%		**CareerSat** = Very dissatisfied	504 (4%)
53%		**CareerSat** = Slightly dissatisfied	1122 (10%)
37%		**CareerSat** = Neither satisfied nor dissatisfied	934 (8%)

IBM Developer

SKILLS NETWORK

Further Value Added Analysis

Count of Respondents who are already Managers colored by EducationLevel

EdLevel
- Associate degree
- Bachelor's degree (BA, BS, B.Eng....
- Master's degree (MA, MS, M.Eng.,...
- Other doctoral degree (Ph.D, Ed.D...
- Professional degree (JD, MD, etc.)
- Secondary school (e.g. American ...
- Some college/university study wit...

Already a Manager

18
217
155
18
13
18
70

10   100   1,000

Respondent (Count distinct)
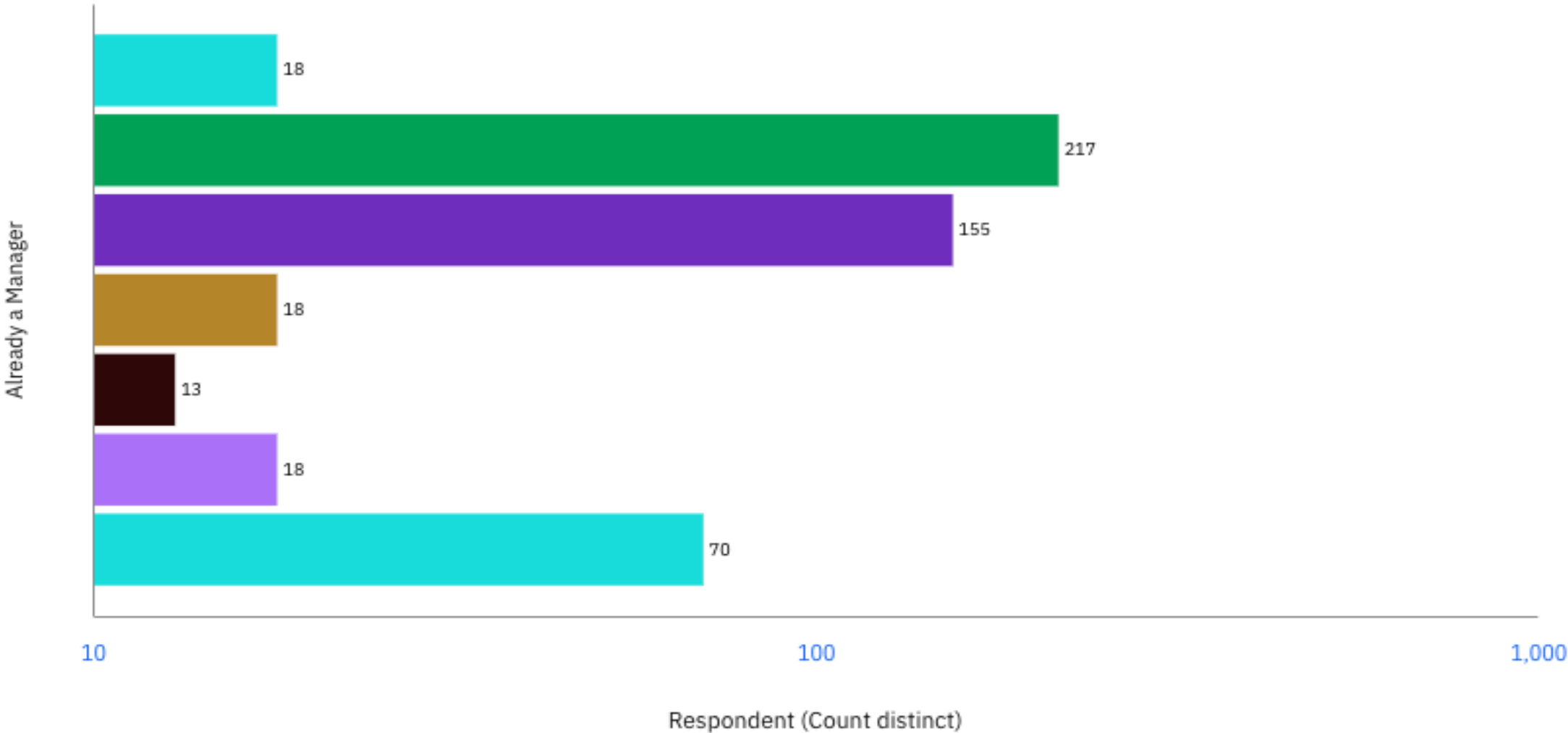
Further Value Added Analysis

Bar Chart showing People Salaries by their Age group

ConvertedComp by Age (Group)
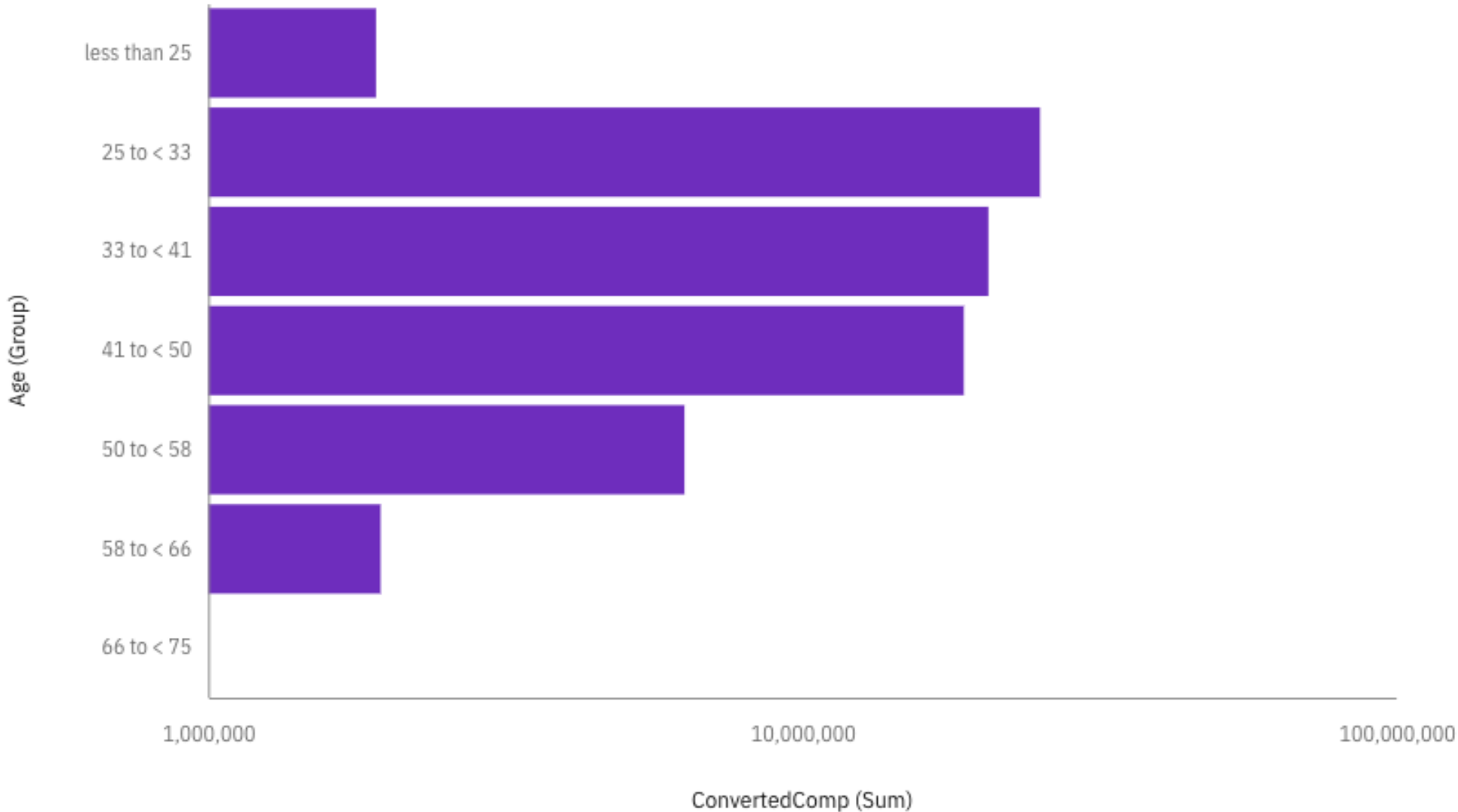
Findings from Further Value Added Analysis

- **From Further Analysis charts, Job Satisfaction of people working in office is a factor of 2 more than people working from home.**

- **From Tree Chart of overall analysis of Career Satisfaction, 42% of survey respondents are very satisfied with their careers.**

- **From analyzing education level of survey respondents and their role as managers, we can note that people holding Bachelor's and Master's degree are already managers in their career.**

- **Surprisingly, the counts of managers for people holding a PhD degree or secondary education seem to be roughly the same. The former could be because of the fact that people with a PhD degree are in academia rather than in the industry. On the other hand, the plot also shows that it is hard for people with just the secondary education to become managers.**

- **People in the age group between 25 and 33 earn higher salaries. With increase in age group, their earnings get reduced.**

Innovation Ideas:

- With a good grasp on concepts in data analytics, like data cleaning, exploratory data analysis, data visualization and Statistical analysis many real world data can be analyzed and products can be developed.
- By identifying trends and patterns in large volumes of dataset, example in the fields of product development and marketing, customer base can be segmented with respect to demographics, preferences and behavior.
- Considering Retail, Understanding customers, their buying patterns and choices, companies can market their product. By understanding individual customer preferences, companies can suggest their products, offer promotions and thereby increasing customer satisfaction and building loyalty.
- Collecting and analyzing sales data from different sources like sales, inventory, customer purchase history, retail industry can make better decisions about their pricing, promotions, hiring right customer service expert and marketing strategies. This can increase revenue and minimize costs.
- Proper Data driven decision making by maintaining data privacy and security can increase customer loyalty and satisfaction. Customer satisfaction can be measured from data collected from various sources using online surveys, purchase surveys, customer service calls, in-store feedback, focus groups and loyalty programs. It is a win-win for the customers and businesses.

IBM **Developer**

SKILLS NETWORK