IBM Data Analyst Capstone Project

Van Dzung

April 2024

IBM Developer

SKILLS NETWORK

OUTLINE

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

IBM **Dev**oper

SKILLS NETWORK

EXECUTIVE SUMMARY

- Current Technology Usage Trend
- Future Technology Trend
  - Language
  - Database
  - Platform
  - Web frame
- Demographics Survey
- Country & Gender Difference

INTRODUCTION

- Analyze the Technology and programming data to help :
  - Identify the most well-known programming languages, databases, platforms, and web frames in demand
  - Identify future skill requirements in the industry
  - Determine the gap in human resources

- Data must be collected from various resources then analyzed.

- Audience for this Presentation :
  - IT developers and investors
  - Students who study DA and programming.

IBM Developer

SKILLS NETWORK

METHODOLOGY

- Data Collection (Sources)
  - Stack overflow developer 2019 survey
  - API
- Data Wrangling
- Data Exploration
- Data Cleaning
- Data Visualization
  - Python matplotlib & turtle
  - Creating dashboard with IBM Cognos
- Presentation

IBM **Dev** loper

SKILLS NETWORK

RESULTS – Data collection Using API

- Washington DC has highest Job Posting.

- 1173 jobs posting in Python

```
api_url="http://127.0.0.1:5000/data"
def get_number_of_jobs_T(technology):
    payload={"Key Skills": technology}
    response=requests.get(api_url, params=payload)
    if response.ok:
        data=response.json()
#       print(data)
        number_of_jobs = len(data)

    return technology, number_of_jobs
```

Calling the function for Python and checking if it works.

```
get_number_of_jobs_T("Python")
```

```
('Python', 1173)
```

	Technology	Los Angeles	New York	San Francisco	Washington DC	Seattle	Austin	Detroit
0	C	296	1622	214	2664	1668	224	1973
1	C#	5	41	3	68	49	5	60
2	C++	3	43	3	55	41	4	32
3	Java	43	326	38	516	354	32	353
4	JavaScript	7	51	7	61	52	5	41
5	Python	24	143	17	258	133	15	170
6	Scala	0	8	0	3	4	1	5
7	Oracle	17	95	19	143	110	11	115
8	SQL Server	3	36	2	53	31	5	34
9	MySQL Server	0	0	0	0	0	0	0
10	PostgreSQL	0	1	0	3	1	0	2
11	MongoDB	2	25	2	32	21	1	25

RESULTS – Exploratory Data survey

- Data shape: 11552 rows, 85 columns

- Average Age of Surveys Takers: 30.77

- 135 Countries took the survey

```
[9]:  # your code goes here
      df['Age'].mean(axis=0)

[9]:  30.77239449133718
```

The dataset is the result of a world wide survey. Print how many unique countries are there in the Country column.

```
[10]:  # your code goes here
       len(df['Country'].unique())

[10]:  135
```

RESULTS - Data Wrangling

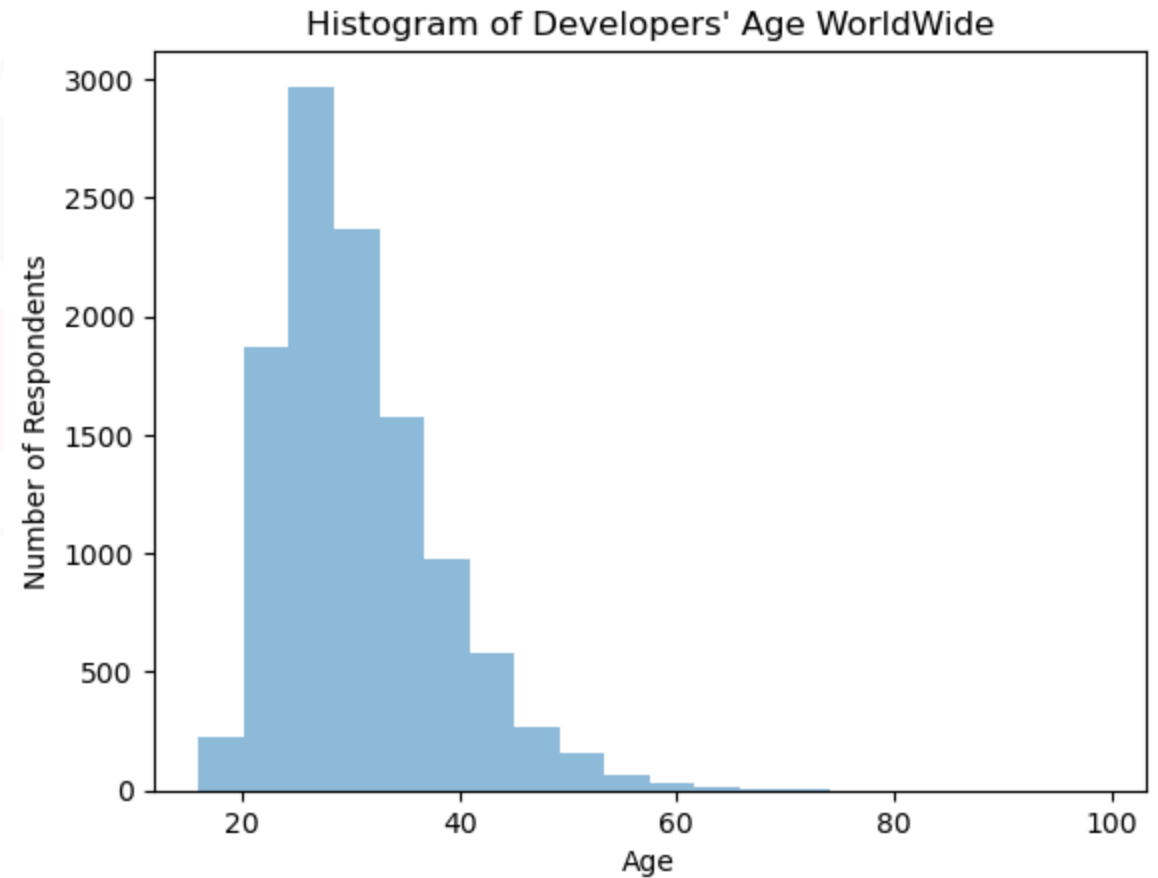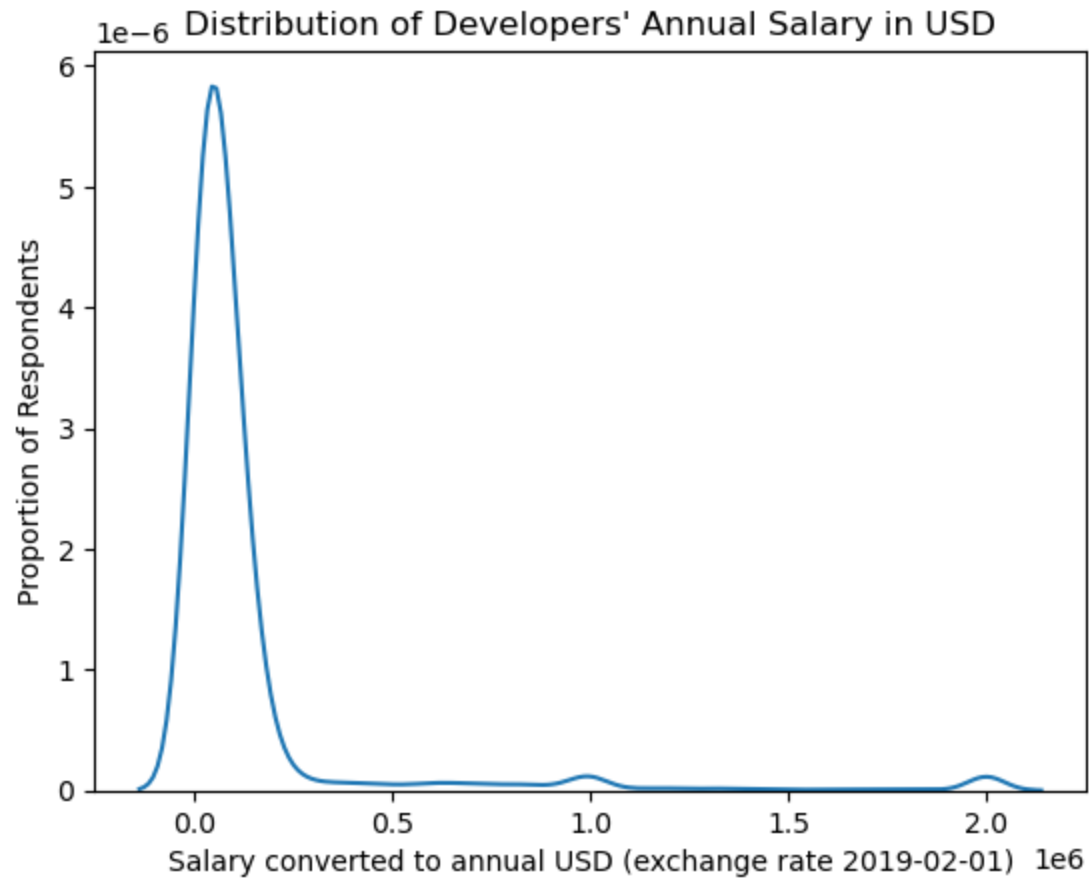After finding, then removing duplicate values:

(11552, 85) => (11398, 85)

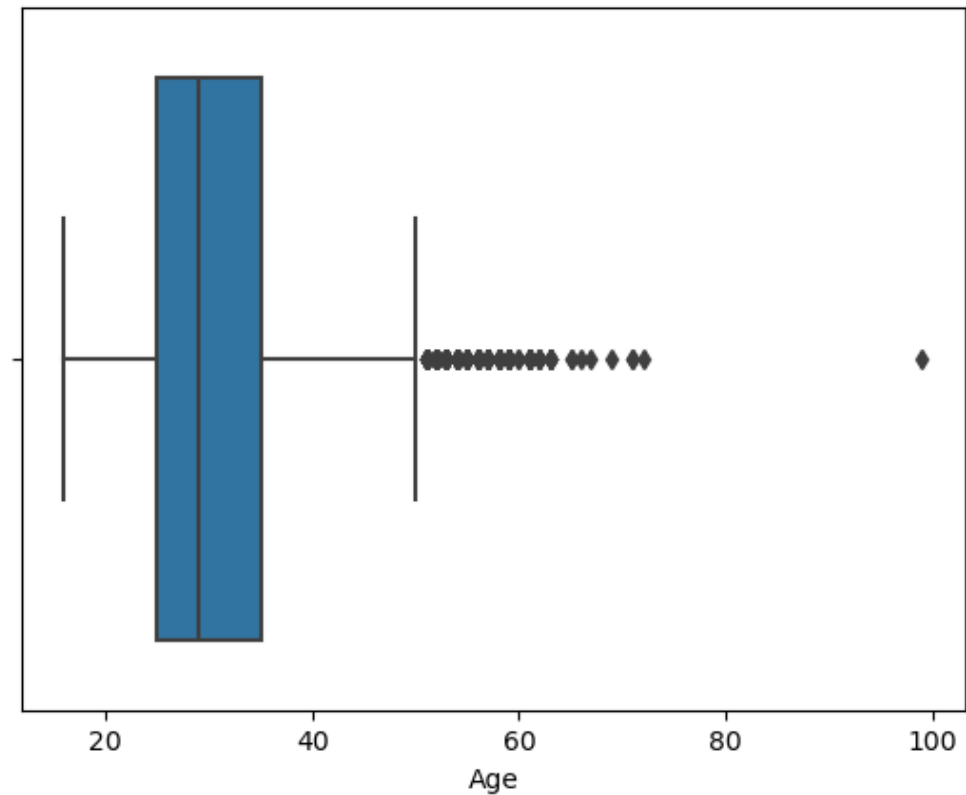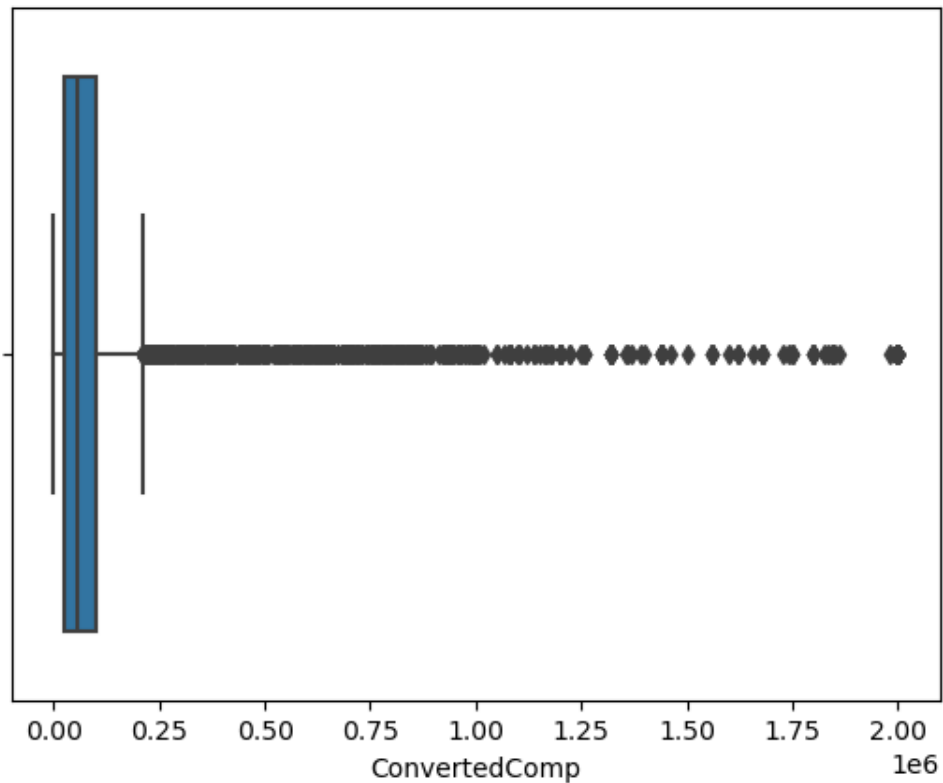Data Normalized Annual Compensation:

100000.0

RESULTS – EDA - Distribution

Distribution of Developers' Annual Salary in USD

Histogram of Developers' Age WorldWide

RESULTS – EDA - Outliers
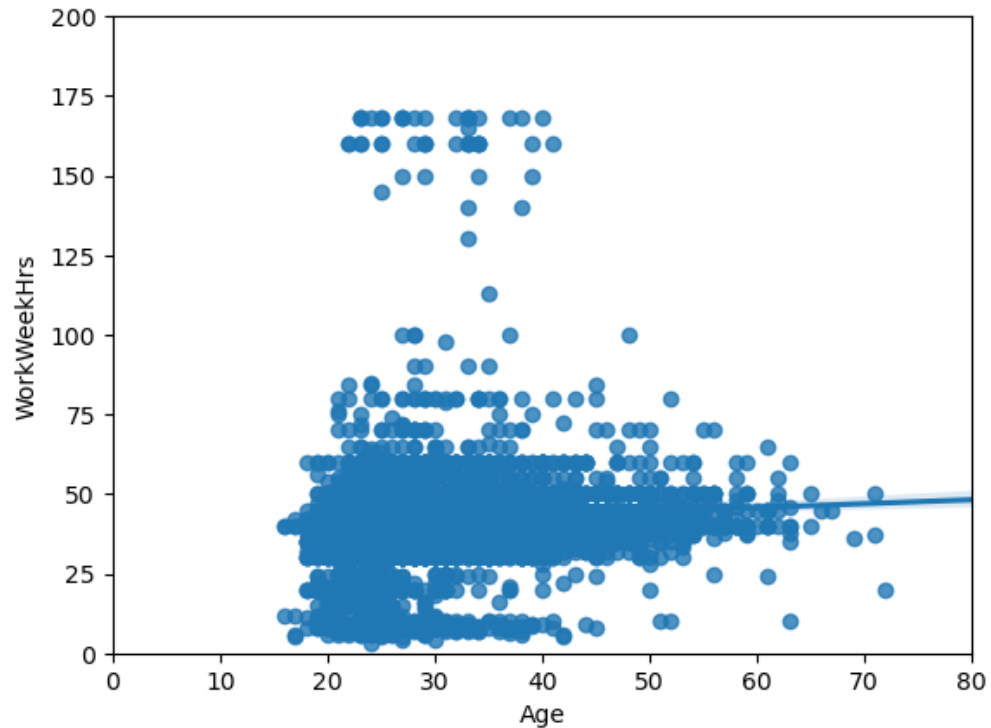
Most Outliers are found after Q3

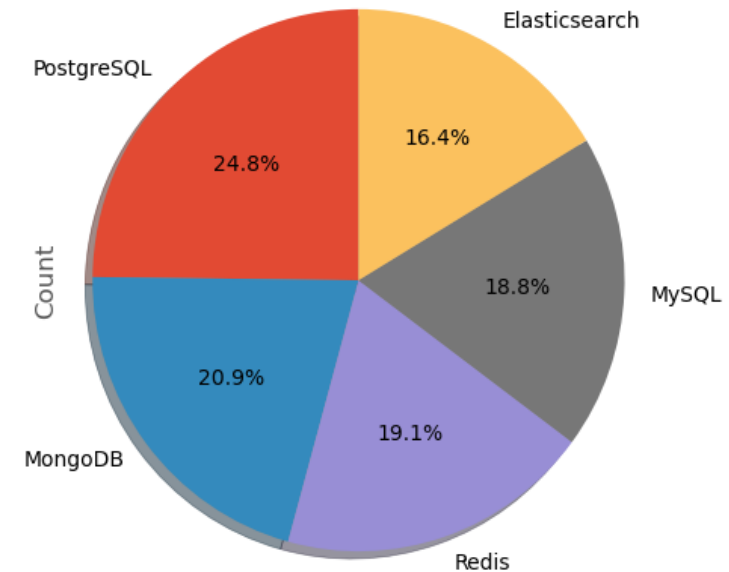# RESULTS – EDA - Correlation

```
Respondent           0.004041
CompTotal            0.006970
ConvertedComp        0.105386
WorkWeekHrs          0.036518
CodeRevHrs          -0.020469
Age                  1.000000
Name: Age, dtype: float64
```

RESULTS – Data Visualization using SQL

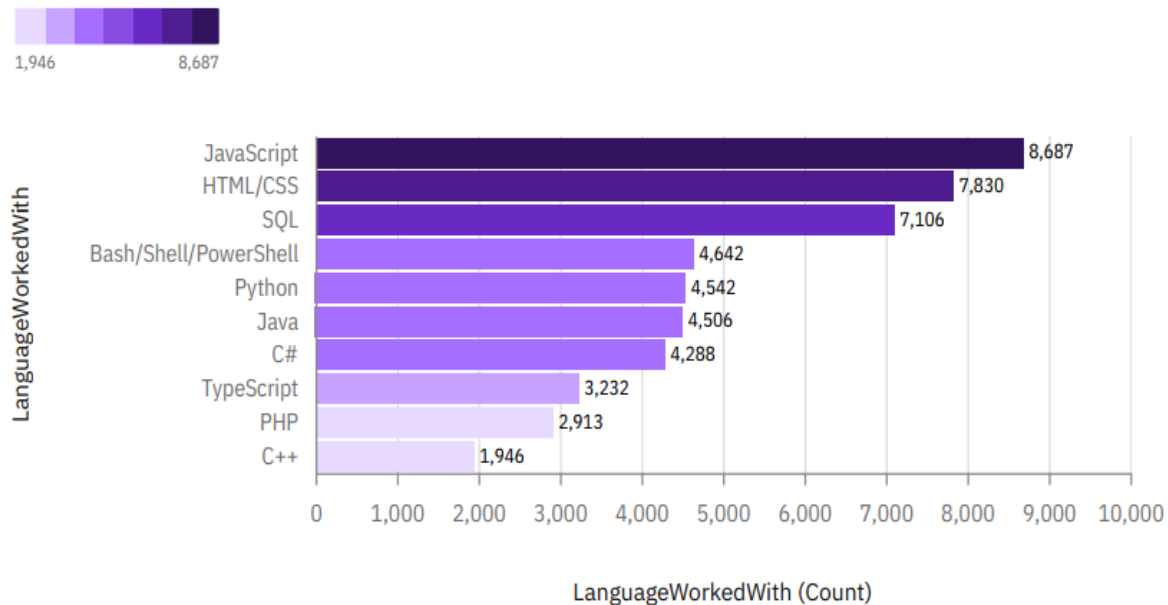Top 5 databases that respondents wish to learn next year

PROGRAMMING LANGUAGE TRENDS

Current Year

Next Year

Top 10 Language Worked With

1,946   8,687

LanguageWorkedWith	Count
JavaScript	8,687
HTML/CSS	7,830
SQL	7,106
Bash/Shell/PowerShell	4,642
Python	4,542
Java	4,506
C#	4,288
TypeScript	3,232
PHP	2,913
C++	1,946

LanguageWorkedWith (Count)

Top 10 Language Desire Next Year

LanguageDesireNextYear	Count
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

LanguageDesireNextYear (Count)

IBM Developer

SKILLS NETWORK
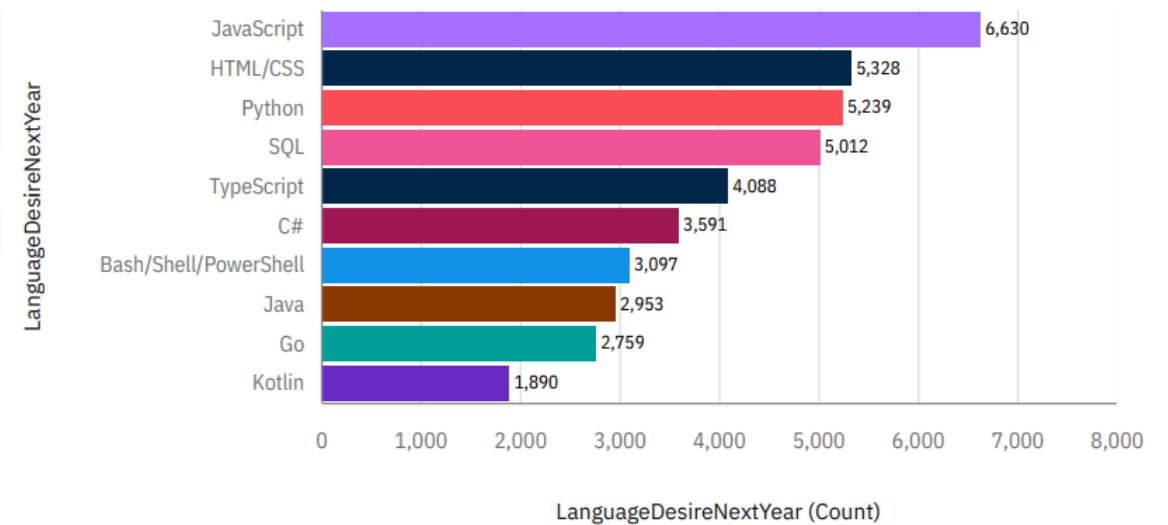
PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

- JavaScript is the most commonly used programming language,
but Python has risen in the ranks again.

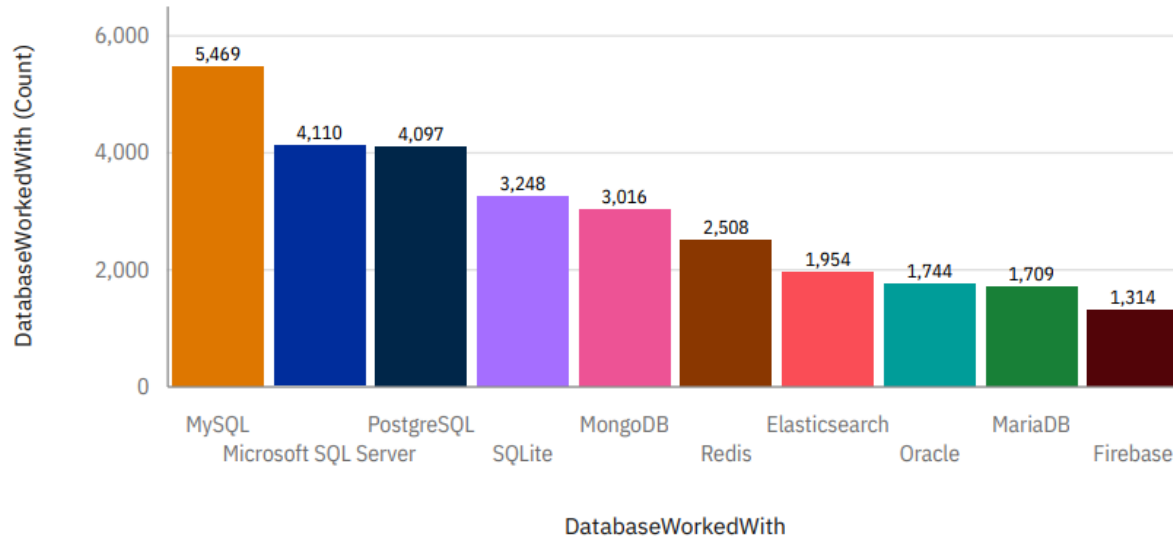- HTML, SQL are still top 4 popular languages now and future

Implications

- Python is the fastest-growing major programming language today.

- Web and application development are still demanded job.

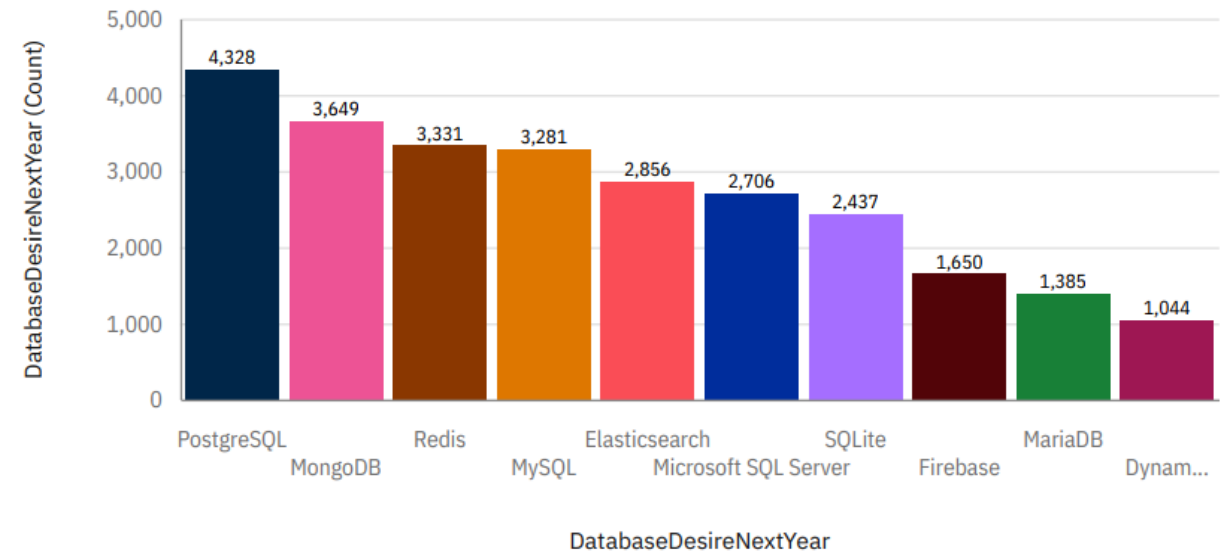- Database manipulation is also an important skill demanded.

DATABASE TRENDS

Current Year

Next Year

Top10 Database Worked With

Top 10 Database Desire Next Year

IBM Developer

SKILLS NETWORK

DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- MySQL is the most commonly used database

- MySQL, MongoDB, and PostgreSQL are still popular databases now and in the future.

- Microsoft SQL and SQL lite will be replaced by Elasticsearch and Redis.

Implications

- Private Company owned databases competitions are increasing.

- Open-Source database skills are still in high demand.
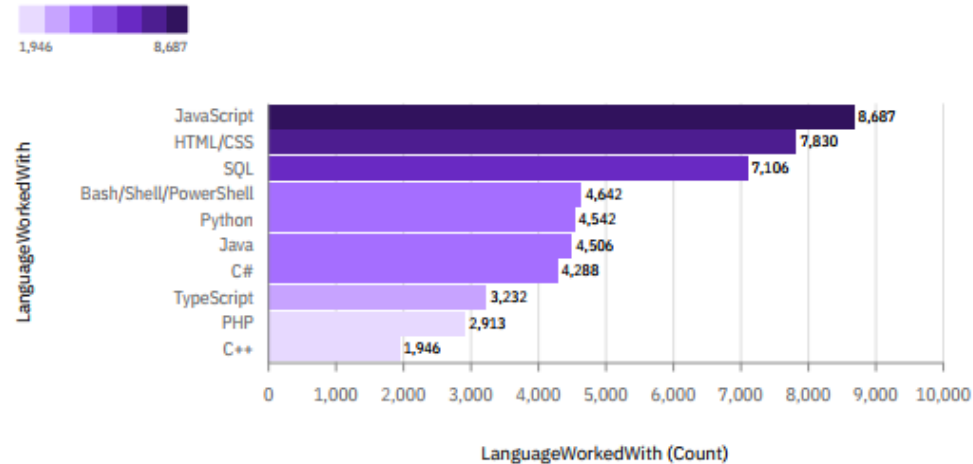
- Non relational database are also on the rise.

IBM Developer

SKILLS NETWORK

DASHBOARD

<The GitHub link of the Cognos dashboard goes here.>

Dvan4/IBM-Data-Analyst-Capstone- (github.com)

IBM Developer

SKILLS NETWORK

DASHBOARD TAB 1

Current Technology Usage

Top 10 Language Worked With

1,946   8,687

Language	Count
JavaScript	8,687
HTML/CSS	7,830
SQL	7,106
Bash/Shell/PowerShell	4,642
Python	4,542
Java	4,506
C#	4,288
TypeScript	3,232
PHP	2,913
C++	1,946

LanguageWorkedWith (Count)

Top10 Database Worked With

DatabaseWorkedWith (Count)

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

Platform Worked With

0   5,811

- (no value)
- AWS
- Android
- Arduino
- Docker
- Google Cloud ...
- Heroku
- IBM Cloud or ...
- Kubernetes
- Linux
- MacOS
- Microsoft Azure
- Other(s):
- Raspberry Pi
- Slack
- Windows
- WordPress
- iOS

AWS Docker Raspberry Pi Linux Windows Android MacOS WordPress Heroku Kubernetes Google Cloud Platform Microsoft Azure Arduino Slack IBM Cloud or Watson iOS Other(s):
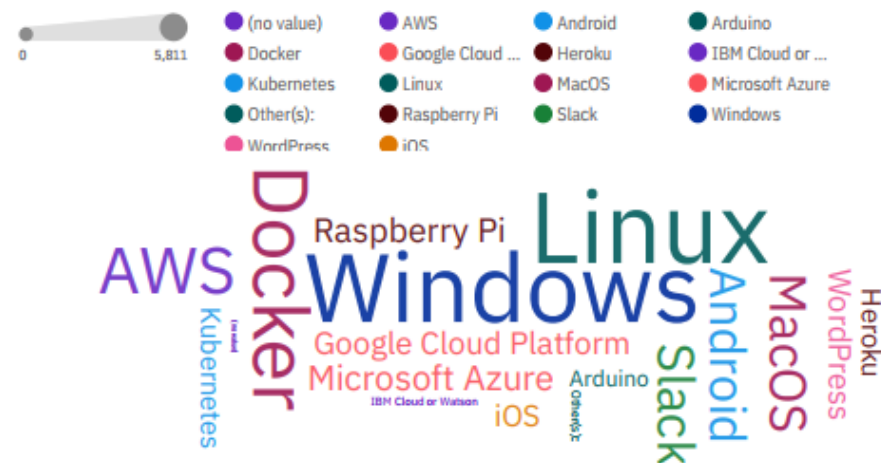
Top 10 WebFrame Worked With

939   4,629

- ASP.NET
- Angular/Angular.js
- Django
- Express
- Flask
- Laravel
- React.js
- Spring
- Vue.js
- jQuery

ASP.NET · jQuery · Angular... · React.js · V... · Ex... · S...
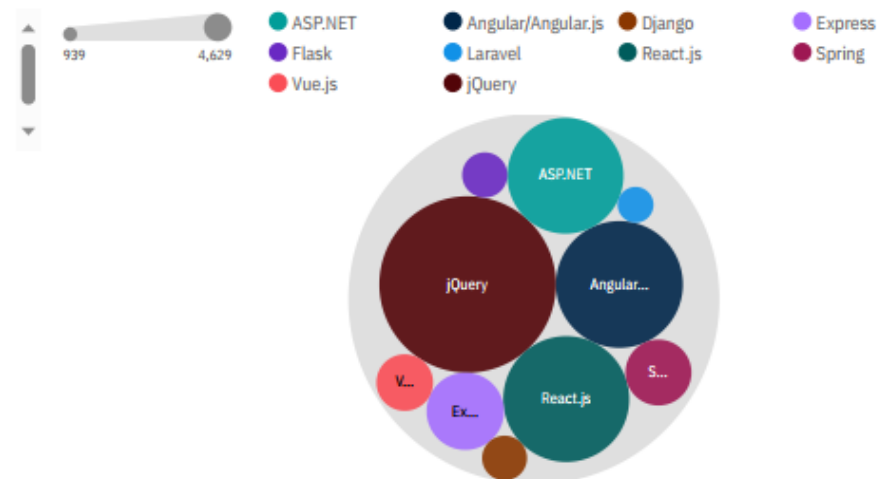
DASHBOARD TAB 2

Future Technology Trend

Top 10 Language Desire Next Year

Language	Count
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

LanguageDesireNextYear (Count)

Top 10 Database Desire Next Year

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
Dynam...	1,044
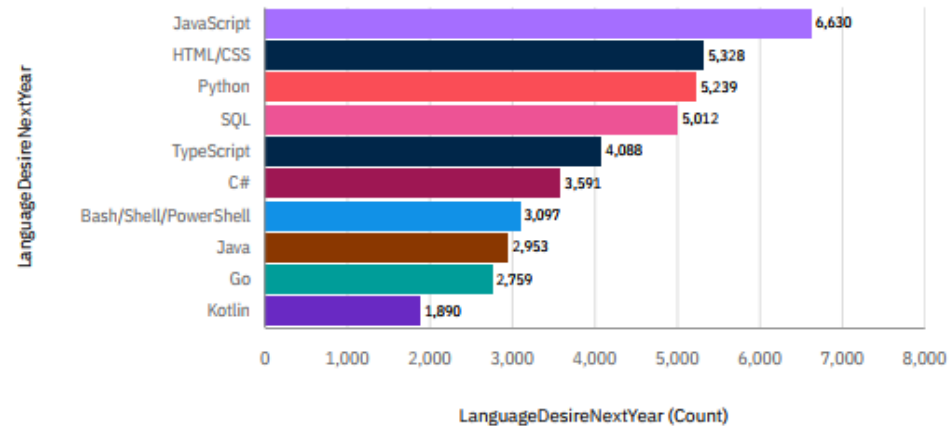
DatabaseDesireNextYear

Platform Desire Next Year
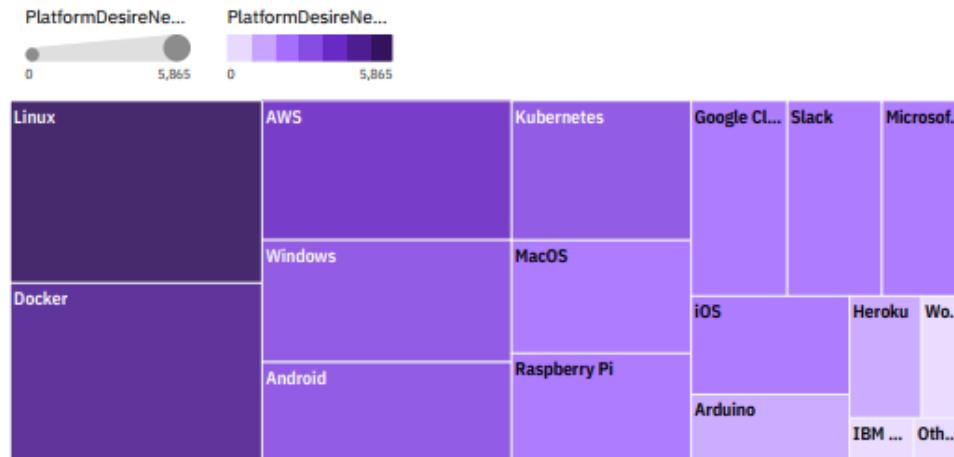
Top 10 WebFrame Desire Next Year

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

IBM Developer

SKILLS NETWORK
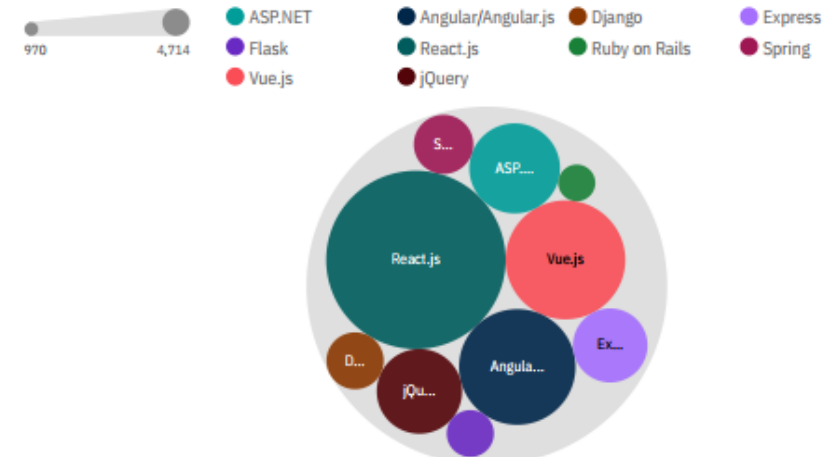
DASHBOARD TAB 3

Demographics

Respondent classified by Gender

93.5%

● Woman ● Man
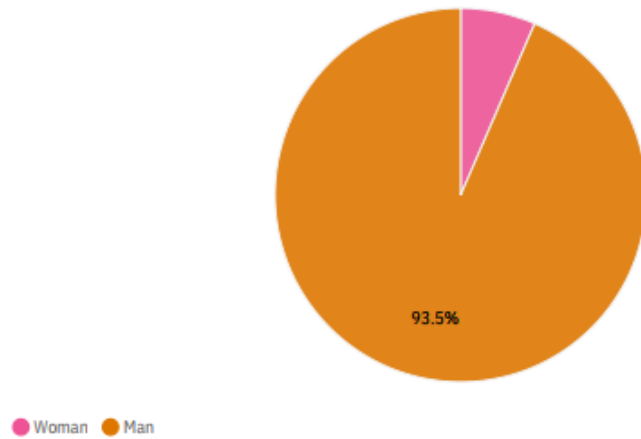
Respondent Count for Countries

Respondent (Count)

1 ─────── 3,127

© Mapbox © OpenStreetMap

Respondent Count by Age

647
452
1,062
887
1,209
1,171
791
617
422
247
124
111 85 63 49 31 19 26 18 6 7 6 8 2 3 3 1 2 1 2 1
57 107
342
163
38
13

10   14   18   22   26   30   34   38   43   7   9   (no...)
  12   16   20   24   28   32   36   40   5   9
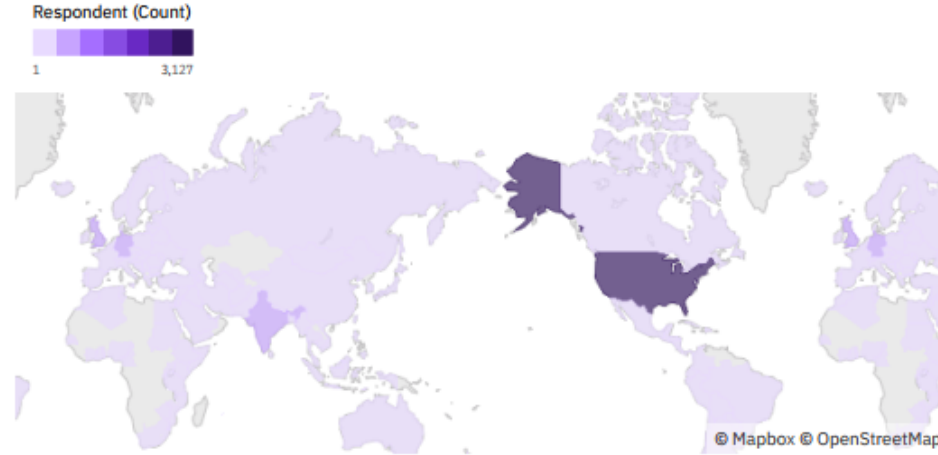
Age1stCode
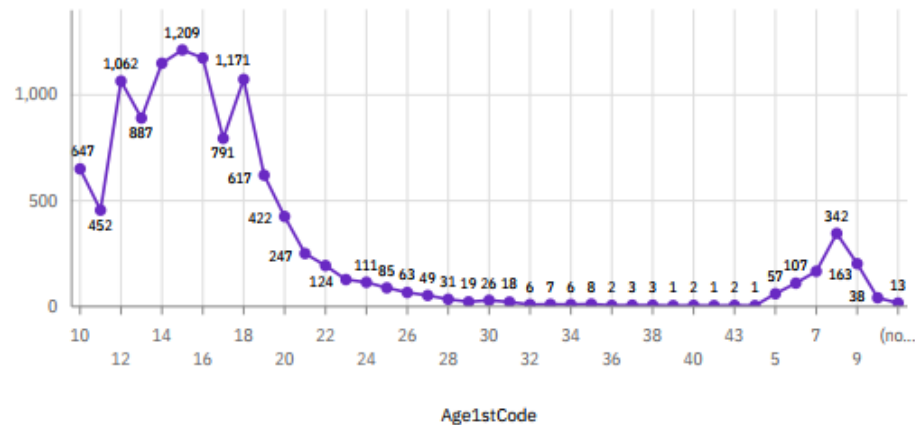
Respondent Count by Gender and Formal Education Level

● Man ● Woman

(no value)
Associate degree
Bachelor's degree (BA, BS, B.Eng., etc.)   5,341   412
I never completed any formal education
Master's degree (MA, MS, M.Eng., MBA, etc.)   2,484
Other doctoral degree (Ph.D, Ed.D., etc.)
Primary/elementary school
Professional degree (JD, MD, etc.)
Secondary school (e.g. American high school, German...) 482
Some college/university study without earning a degree 1,280

EdLevel

0   1,000   2,000   3,000   4,000   5,000   6,000

Respondent (Count)

IBM Developer

SKILLS NETWORK

DISCUSSION

- In this survey, Python has climbed to the second place outperforming Java in the fastest growth of major programming languages (behind Rust)

OVERALL FINDINGS & IMPLICATIONS

Findings

- The highly used languages and databases till stay in high demand in the future
- ThePlatform used right are very similar to the predicted platforms that will be used in the future
- jQuery will take the Lead in terms of web Frame work with in the future

Implications

- Man are still the leading gender in technology
- Other continent are still lagging behind America and Europe
- 24-34 is the highest age group for IT

CONCLUSION

- Technology is still hot field to be interested now and, in the future

- Due to high demand in the field, competition is getting high and there is a need to quick adaptation to change.

- Companies need to look to new Horizon like Central America and Africa.

- Artificial Intelligence and the like are new area impacting the technology field and are probably skills ones must seek to acquire.

APPENDIX

- Include any relevant additional charts, or tables that you may have created during the analysis phase.

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.

Number of Job Posting

Legend: Los Angeles, New York, San Francisco, Washington DC, Seattle, Austin, Detroit
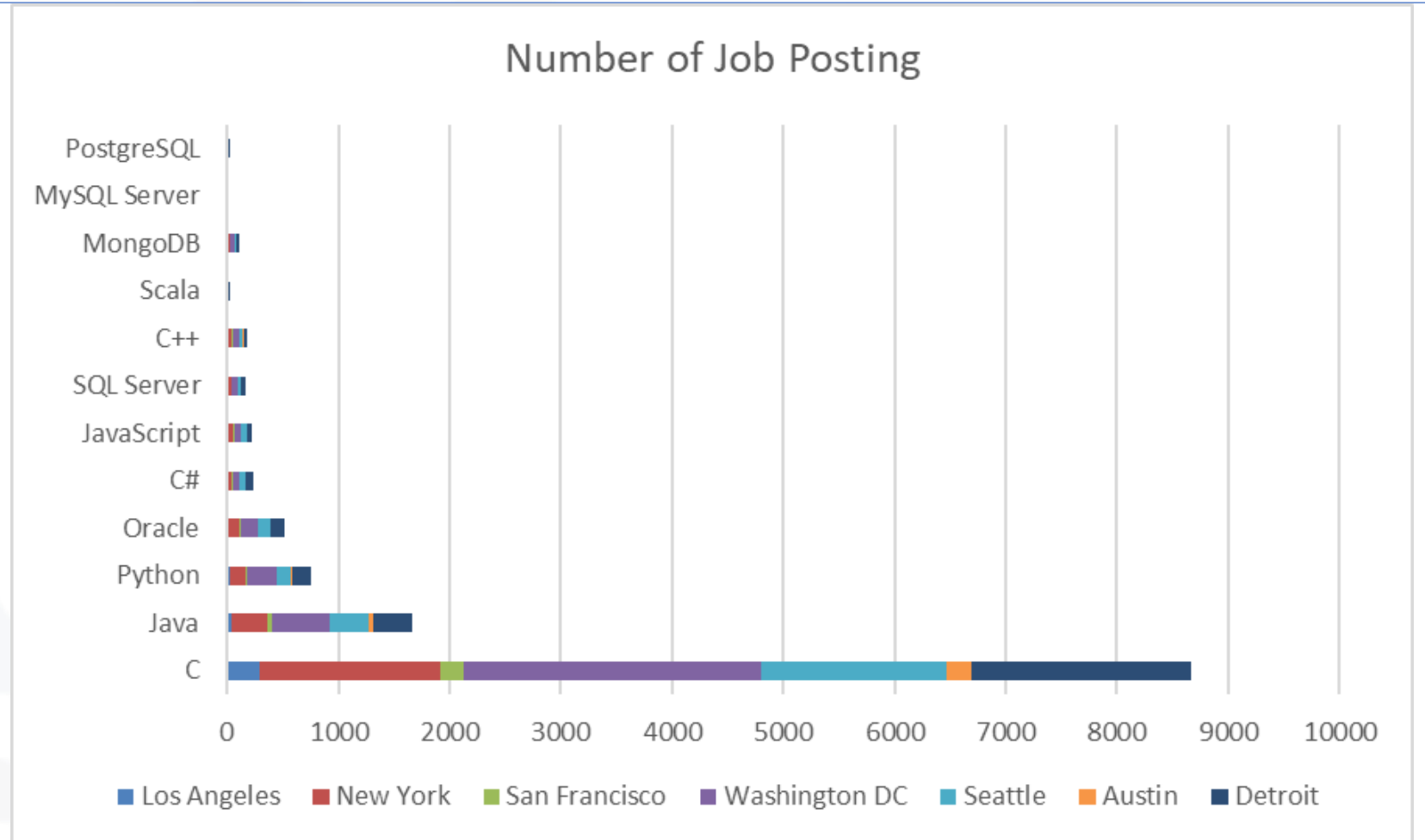
POPULAR LANGUAGES

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.

IBM **Developer**

SKILLS NETWORK