IBM DATA ANALYST CAPSTONE PROJECT: PROVIDING A BETTER IT SOLUTIONS IN THE ORGANIZATION.

PRESENTED BY

EZINNE IMO KALU

13th AUGUST, 2023

IBM Developer

SKILLS NETWORK

OUTLINE

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

EXECUTIVE SUMMARY

- In summary, for the IT Sector to meet up with the present and be competitive in the future, there is need to meet up with the ever change in Technology.
- JavaScript, HTML/CSS, Python, Go and Bash/Shell/PowerShell are the most desirable programming languages for next year.
- PostgreSQL, MongoDB, Redis, MySQL, and ElasticSearch are the most desirable databases for next year.
- There is a wide gap between both genders in the IT sector/company. The male gender are more in the industry, than the female gender.

INTRODUCTION

- This presentation uses data analytics to highlight the present and future trends technology, relating to programming languages, database, platforms and web frame.
- The target audience for this report is for professional, companies and recruiter who wants to know the current trend in their field of work.
- Able to gain insight on demographics of respondent
  - Gender in the workforce.
  - Education level.

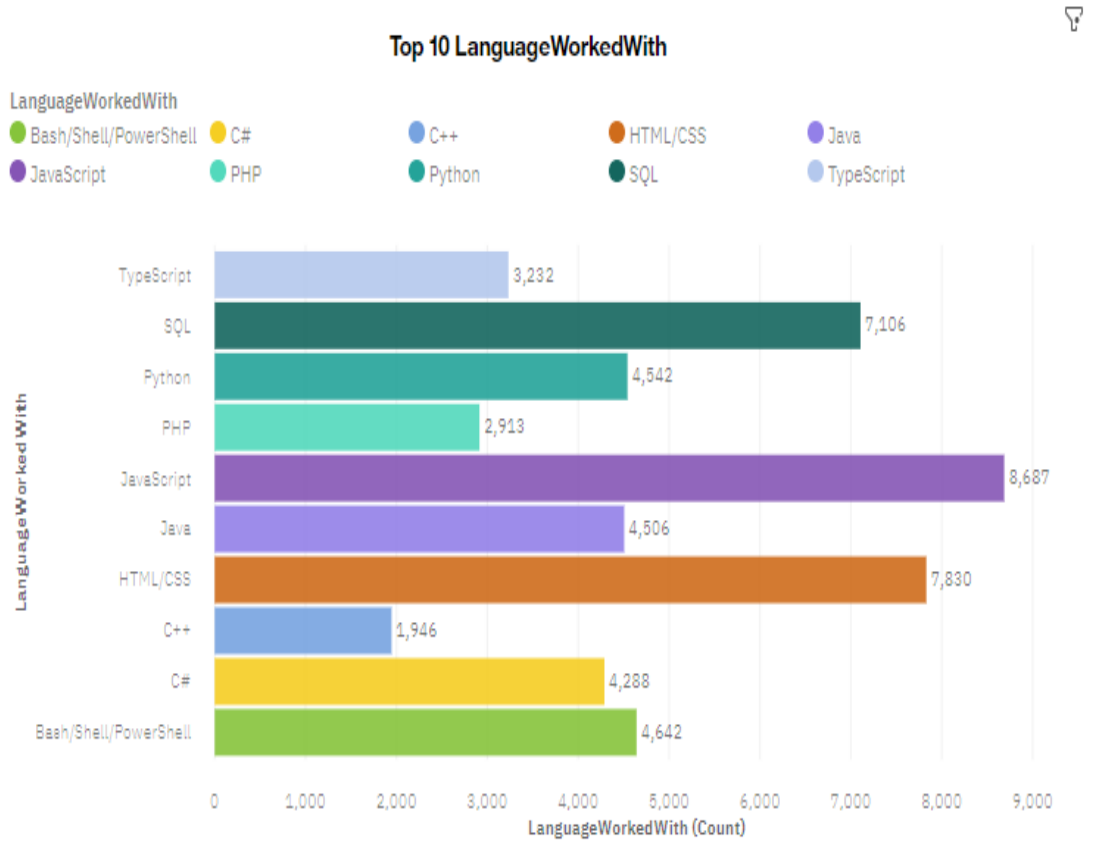
METHODOLOGY

- The data source for the online survey is from Stack Overflow Developer Survey 2019 data where the responses are compiled into csv files.
- The data source for job posting is from Jobs API where API get request call was used to obtain the result
- Python Programming Language on Skill Network
- Data Visualization
  - IBM Cloud Dashboard
  - Visualization on Jupiter Skill Network

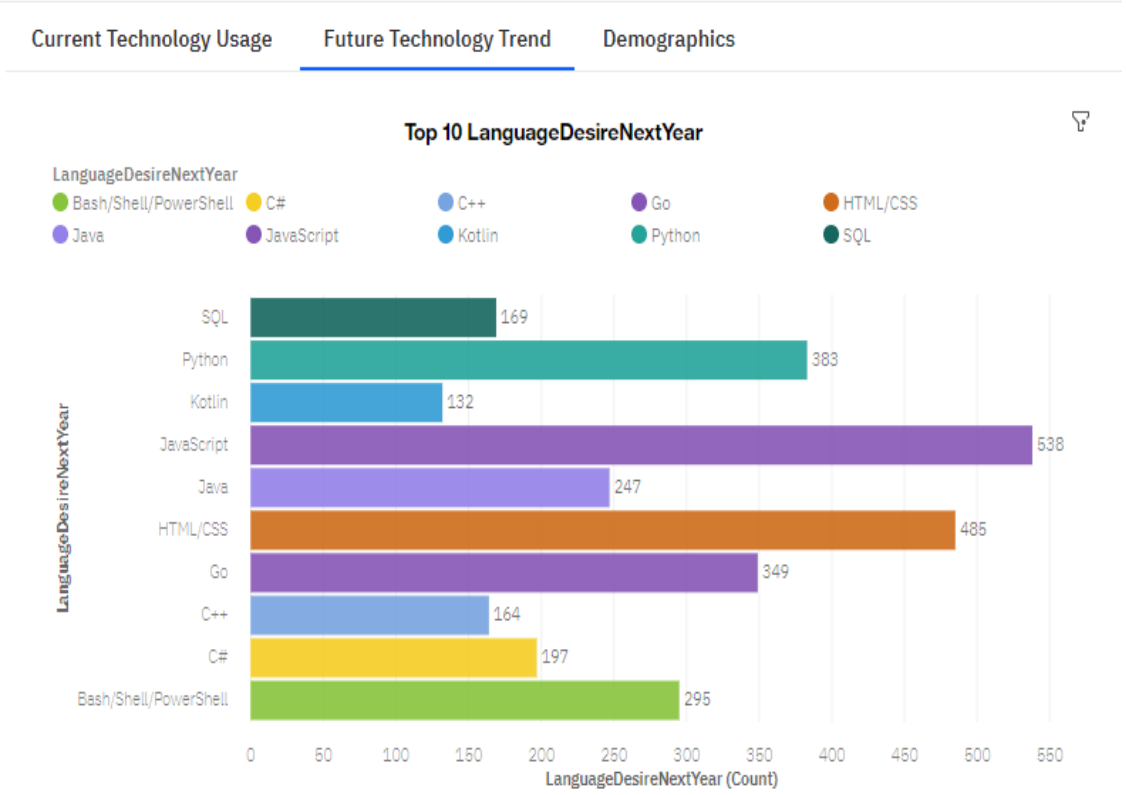
RESULTS

PROGRAMMING LANGUAGE TRENDS

Current Year



Next Year



IBM Developer

SKILLS NETWORK

PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

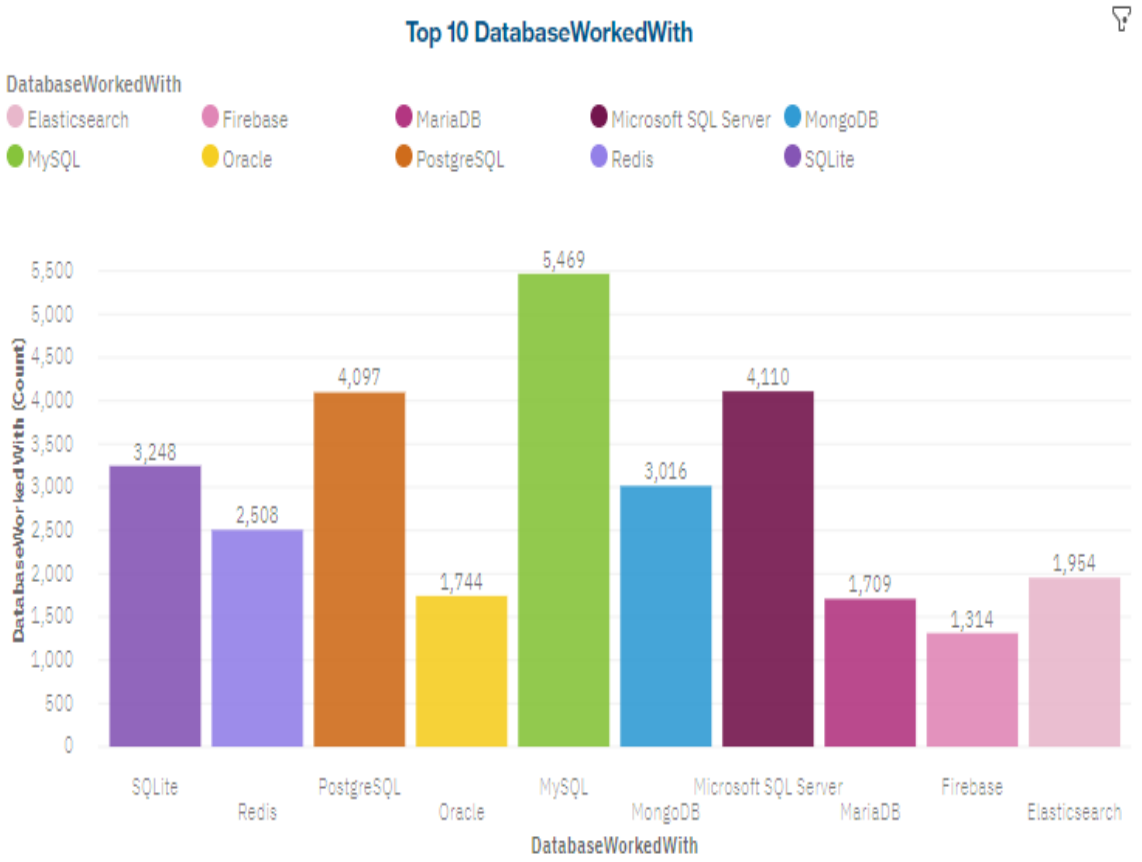
- From the graph above, we find out that JavaScript(8,687), HTML/CSS(7,830), SQL(7,106), Bash/Shell/PowerShell(4642) and Python(4542) are the most commonly use programming languages
- From the graph above, we find out that JavaScript (538), HTML/CSS(485), Python (383), Go(349) and Bash/Shell/PowerShell(295) are the most commonly use programming languages
- From the graph above we also find out that Python which is presently on the 5th position will rise to the 3rd position and Go which was not among the 10 top programming languages seems to be the 4th programming languages in the future.

Implications

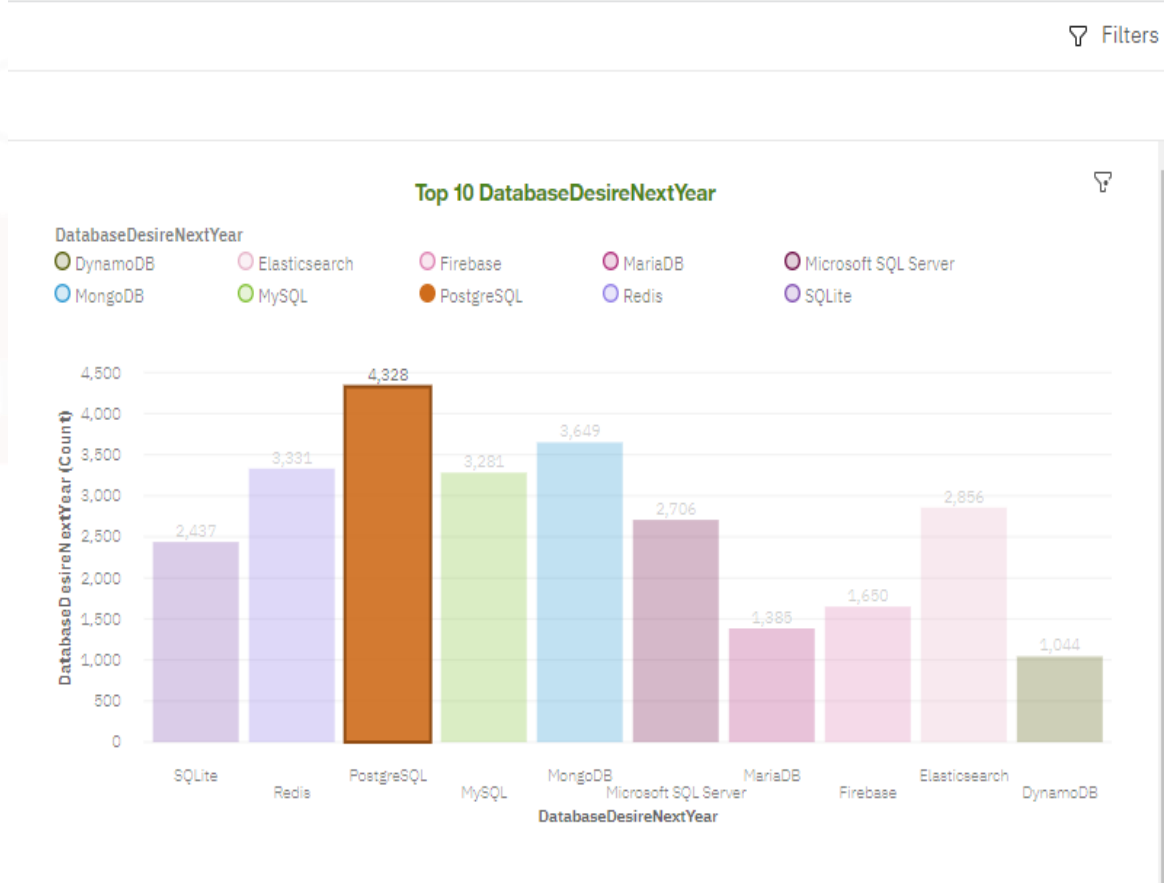
- This implies that JavaScript is presently the highest programming language in use.
- In the future JavaScript will still maintain its position as the highest programming language in use
- Finally, there is going to be some changes in the programming languages use. Example, Python and Go programming language will be more in use.

DATABASE TRENDS

Current Year



Next Year



DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- From the present Database graph we find out the 5 top database in use are: MySQL(5469), Microsoft SQL Server(4110), PostgreSQL(4097), and SQLite(3248).
- From the future Database graph we find out that the 5 top database that will be in use are: PostgreSQL(4328), MongoDB(3649), Redis(3331), MySQL(3281) and Elasticsearch(2856).
- From both present and future Database graph, PostgreSQL move from 4th position to 1st position in the future and MySQL from 1st position to 4th position in the future.

Implications

- It implies that MySQL is presently the highest Database trend.
- PostgreSQL Database is going to be the highest Database in trend.
- From graph, MySQL which is presently the highest Database in use would be any more but will be over taken by PostgreSQL.

DASHBOARD

THIS THE DASHBOARD PERMANENT LINK:

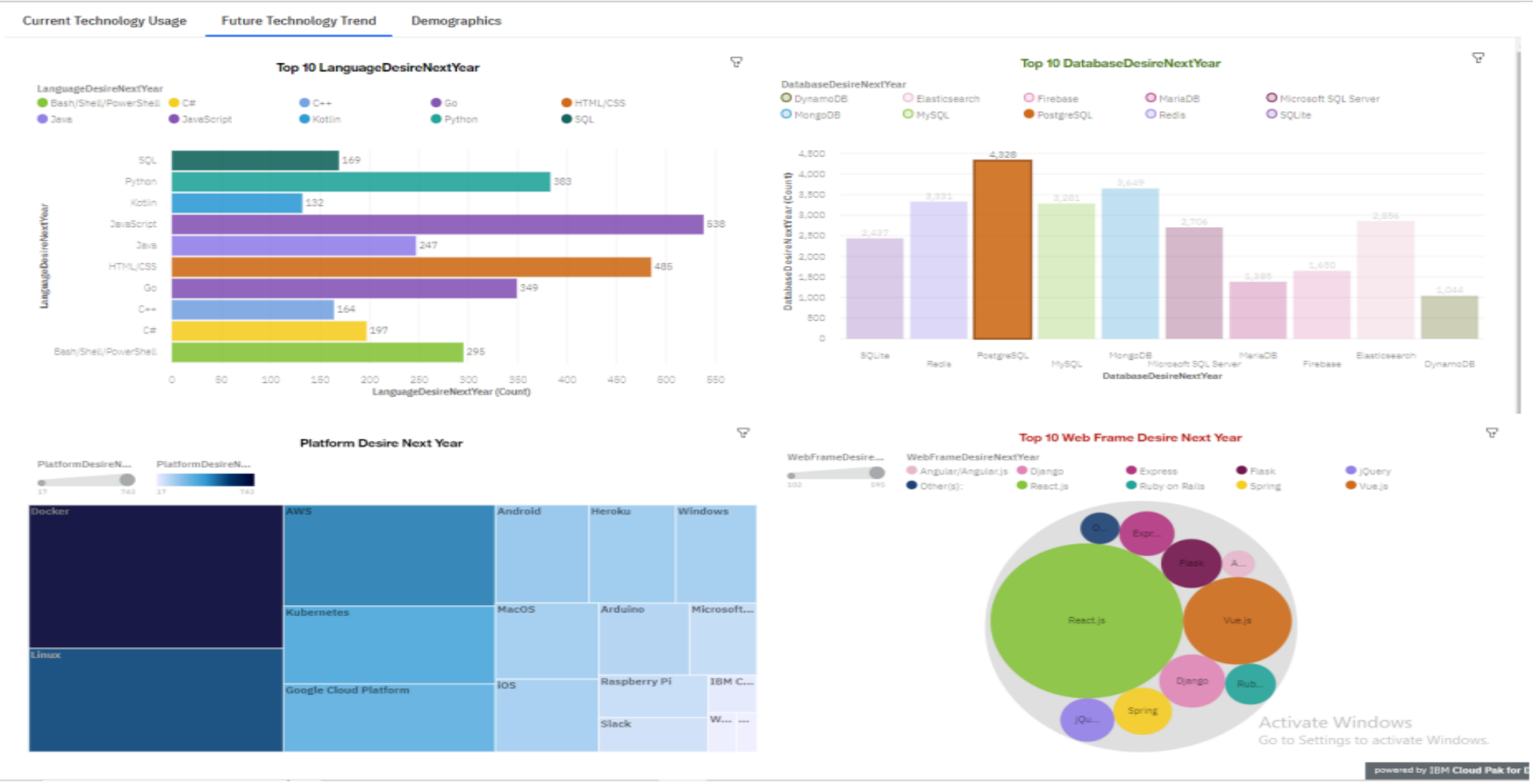
https://dataplatform.cloud.ibm.com/dashboards/32863e89-7d86-4be6-9cde-6025bcdc4aaa/view/5233a23f2e8d11ce40fceee407912c557d372455e7bbd50b81807b495e322597f06d10c2c827495d8f110532a2eb160cc0

DASHBOARD TAB 1

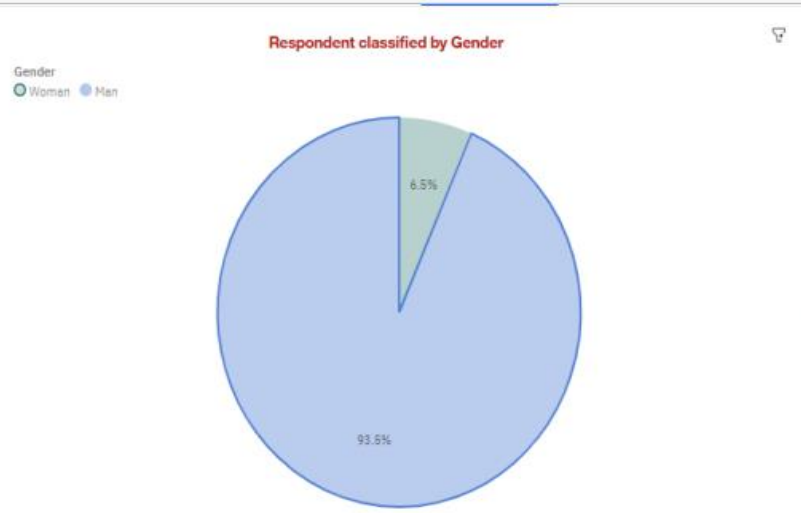
IBM Developer

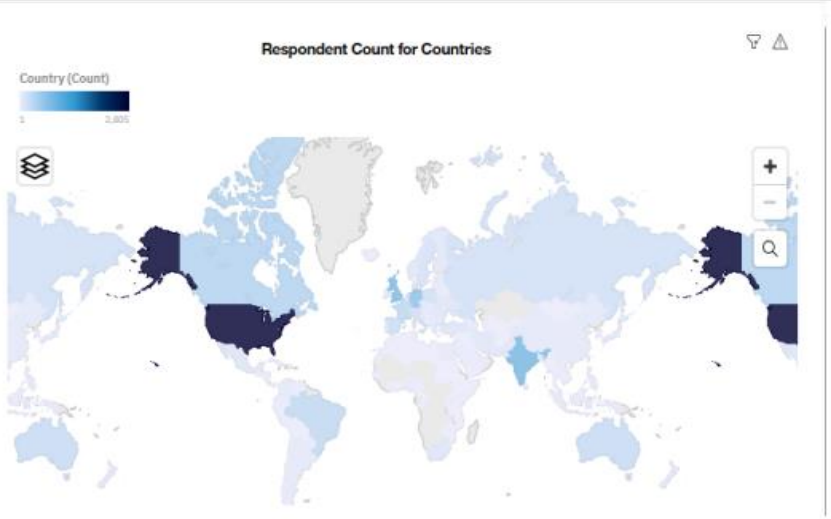
SKILLS NETWORK

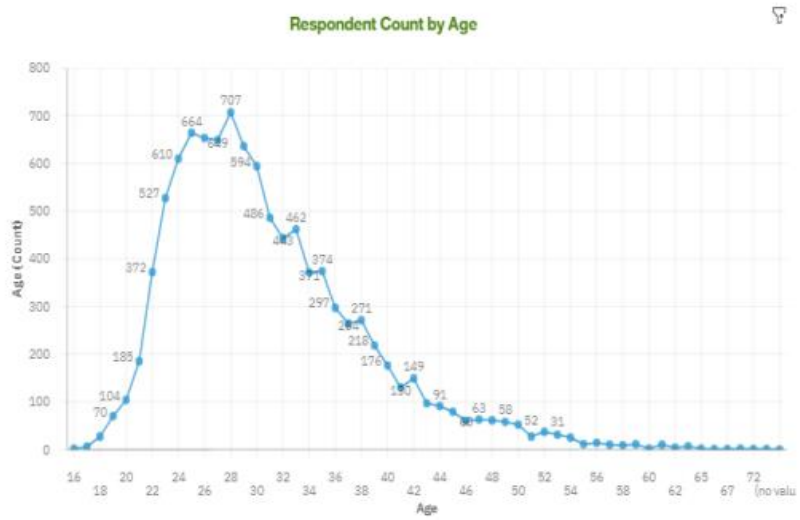
DASHBOARD TAB 2

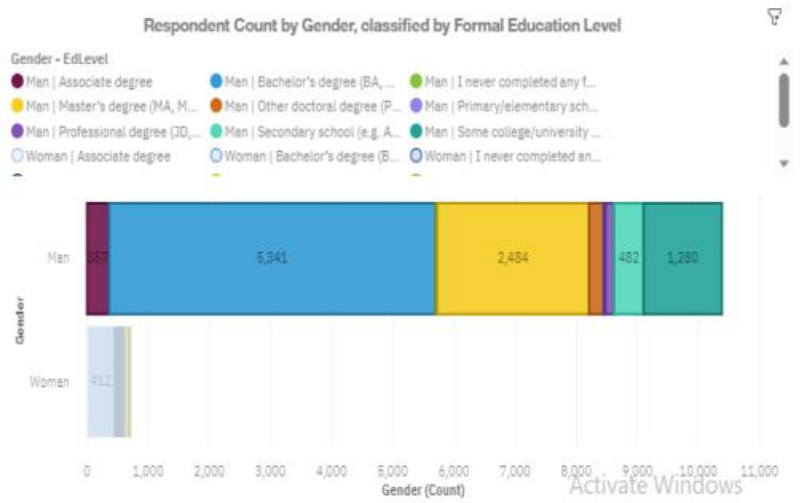


DASHBOARD TAB 3









DISCUSSION

- The majority of respondents are developer by professional.
- The mean age of respondents is 30 years old and the range of most respondents is between 20 – 40 years old.
- A total of 11398 survey respondents were recorded.
- This field of work is dominated by male over female with a ratio of 93.7% to 6.3% respectively.
- Most professional have at least a degree for educational level.
- For the age range of 25 - 30 year old, the compensation has a positive correlation with age.

OVERALL FINDINGS & IMPLICATIONS

Findings

- Python programming language is a fast growing languages
- PostgreSQL is a Database that many company are working with.
- React.js is the future trend for web frame
- Presently, we have less women in the IT sector

Implications

- In future is possible that many IT company will trend to use Python programming language.
- It implies that PostgreSQL will be more in use and many IT companies will like to work with it.
- Professional are migrating from Window platform and jQuery web frame.
- Women have less interest in the IT sector.

CONCLUSION

- IT companies should constantly stay updated in the trends in the organization, in order not to be outdated.
- Women should be more involved in the IT sectors

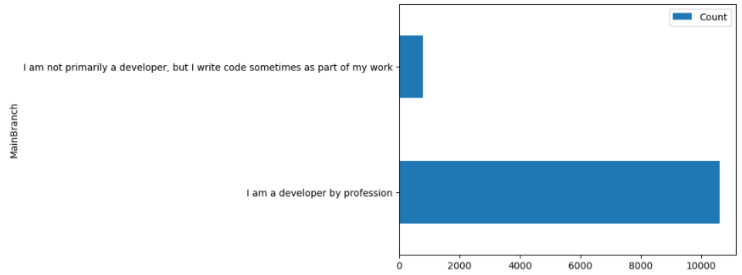
APPENDIX

Bar Chart

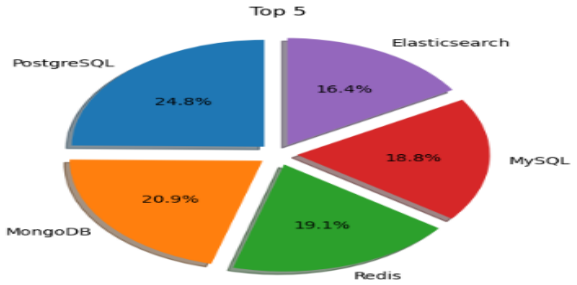
Create a horizontal bar chart using column `MainBranch.`

```
[16]: # your code goes here
QUERY = """
SELECT MainBranch,count(*) as Count
FROM master group by MainBranch
"""

# the read_sql_query runs the sql query and returns the data as a dataframe
df = pd.read_sql_query(QUERY,conn)
df.plot.barh(x='MainBranch', y='Count')
```

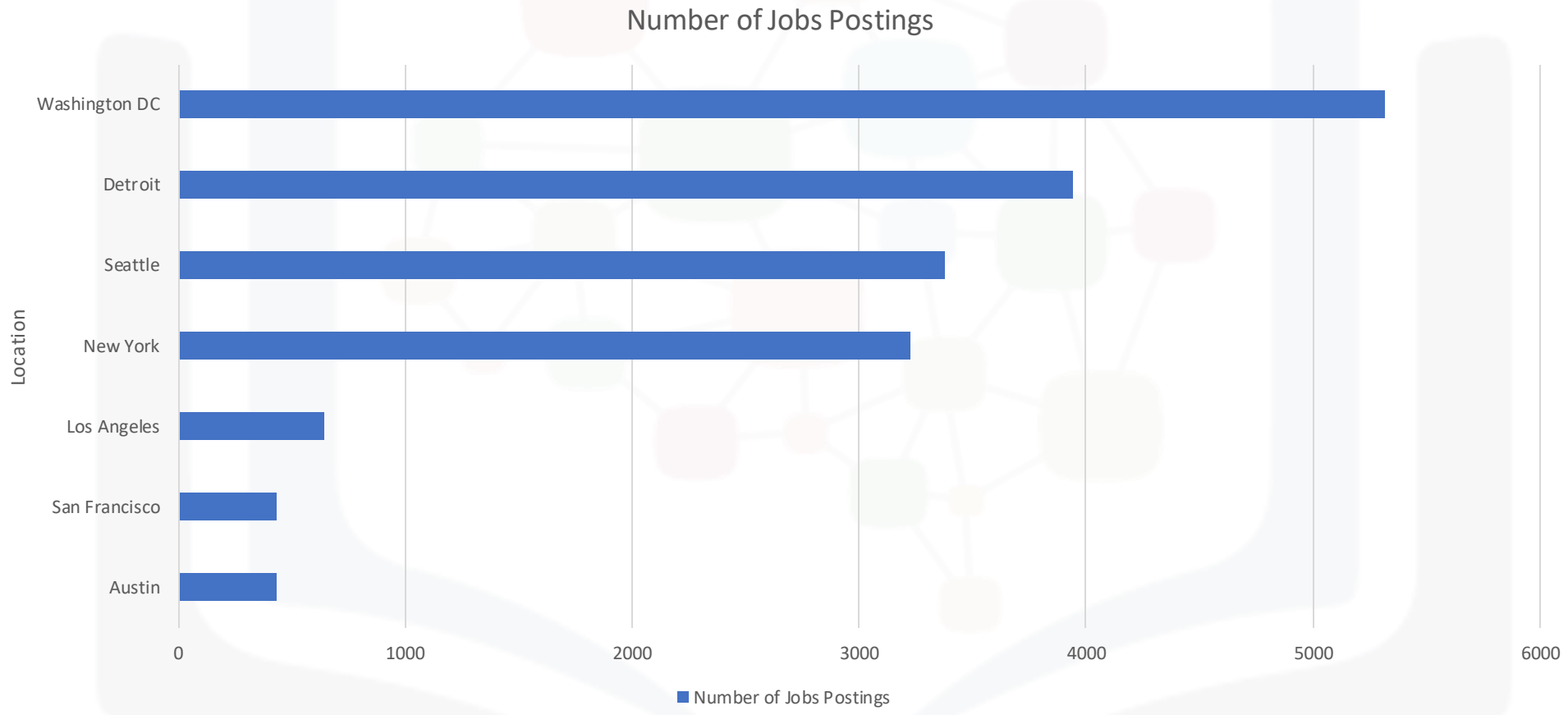
```
[16]: <AxesSubplot:ylabel='MainBranch'>
```




```
[13]: QUERY = """
select count(DatabaseDesireNextYear)_
as Count,DatabaseDesireNextYear_
from DatabaseDesireNextYear_
group by DatabaseDesireNextYear_
order by count(DatabaseDesireNextYear) DESC LIMIT 5

"""
df4 = pd.read_sql_query(QUERY,conn)
df4.set_index('DatabaseDesireNextYear', inplace=True)
lab = df4.index
explode_list = [0.1, 0.1, 0.1, 0.1, 0.1]
sizes = df4.iloc[:,0]
plt.pie(sizes, labels = lab , startangle=90,shadow=True,autopct='%1.1f%%', explode=explode_list)
plt.title('Top 5')
plt.show()
```



JOB POSTINGS



POPULAR LANGUAGES