IBM Data Analyst Capstone Project

Keerthi Katakam

12th August 2024

OUTLINE

- Executive Summary

- Introduction

- Methodology

- Results
  - Visualization – Charts
  - Dashboard

- Discussion
  - Findings & Implications

- Conclusion

- Appendix

IBM Developer

SKILLS NETWORK

EXECUTIVE SUMMARY

- Data contextualization and analysis goal.

- Methodology Description
  - Data gathering
  - Data analysis
  - Data visualizations

- Results presentation supported with graphs and trends

- Discussion of overall findings and implications regarding the results previously exposed

- Final conclusions of the carried out research

INTRODUCTION

- This presentation report uses data analytics to highlight current and projected trends in the need for skills related to programming languages, databases, platforms and web frames.

- The target audience for this research are IT professionals, HR managers, and anybody else with an interest in the IT sector who wants to learn about the top on-demand IT skills in their respective sectors that will also still be relevant in the future

- The following inquiries were investigated using the data:
  - Which programming languages are most in demand today?
  - What are the most in-demand database skills?
  - What popular IDEs or Web frames are there?

METHODOLOGY

- Collect survey data & explore its content by
  - Web Scraping
  - API
  - Request Library
- Data Wrangling
- Exploratory data analysis
  - Analyzing data distribution
  - Handling outliers
  - Correlations
- Data Visualization
  - Create visualizations using the developer survey data
  - Highlight the distribution of data, relationships between data, the composition of data, and comparison of data.
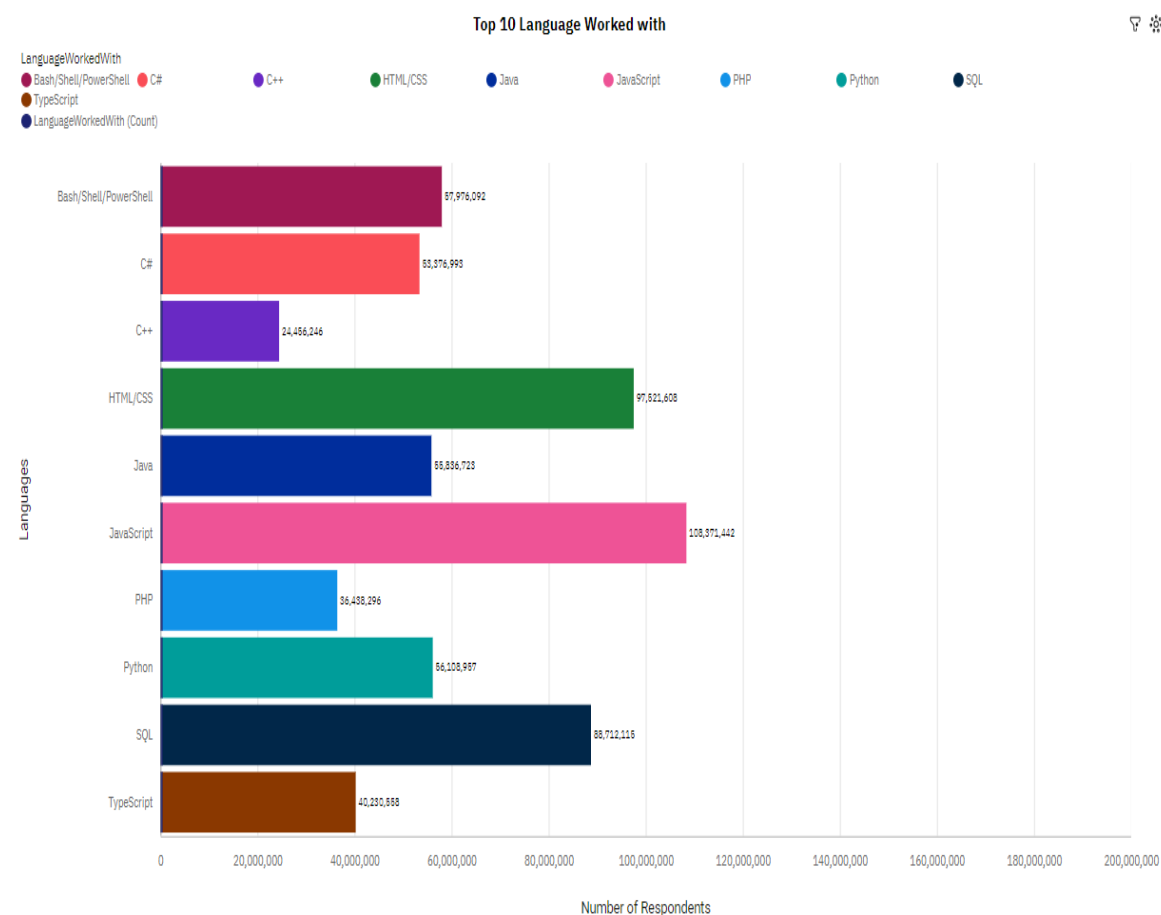
IBM Developer

SKILLS NETWORK

RESULTS
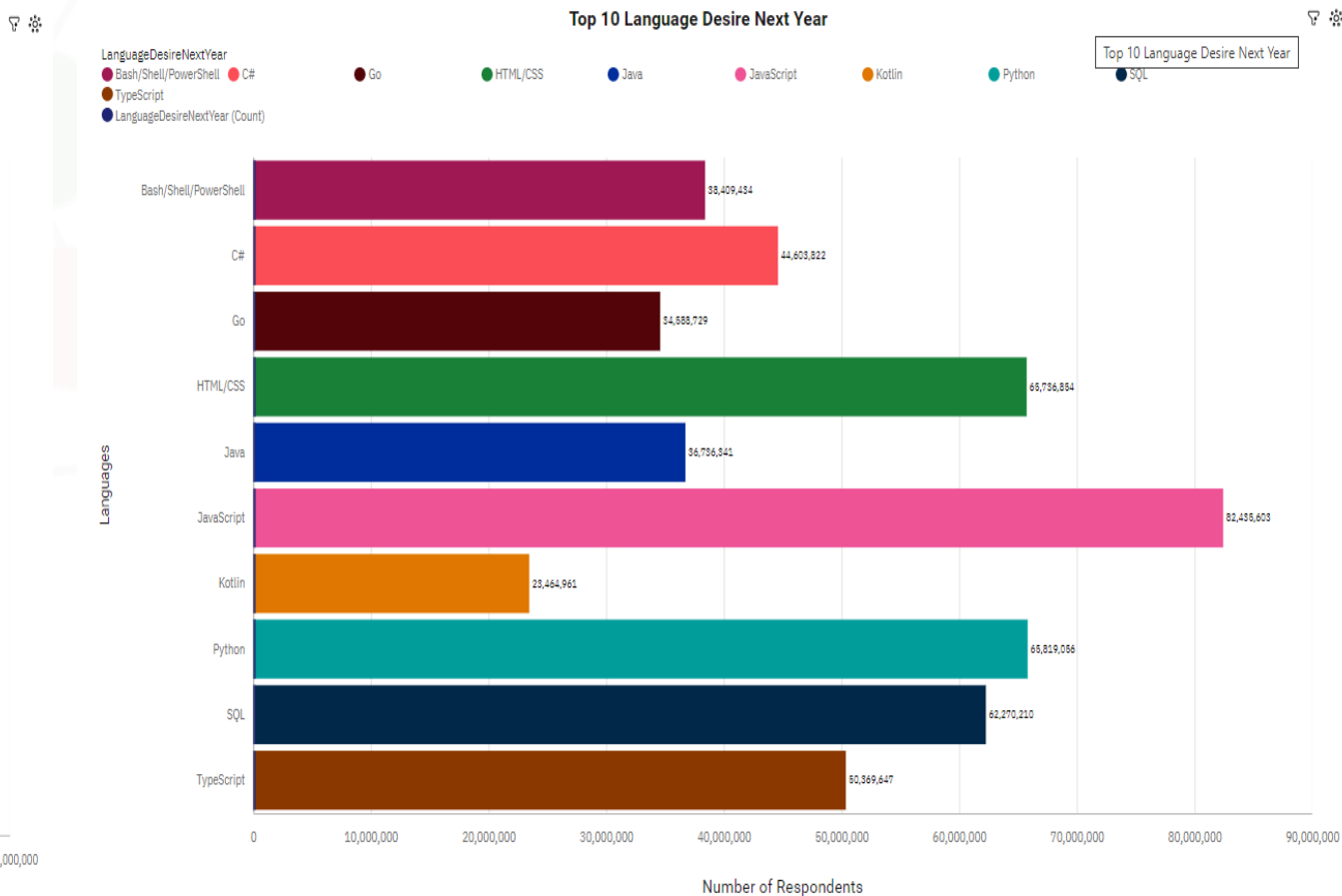
PROGRAMMING LANGUAGE TRENDS

Current Year

Next Year

PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

- Javascript,HTML/CSS,SQL,Shell languages and Python are the most used languages currently. JavaScript seems to keep as leading language.

- Javascript, HTML/CSS, Python, SQL, and Typescript will be the most used languages next year and future years. Great interest in Typescript

- Python will have more demand than SQL for next year.
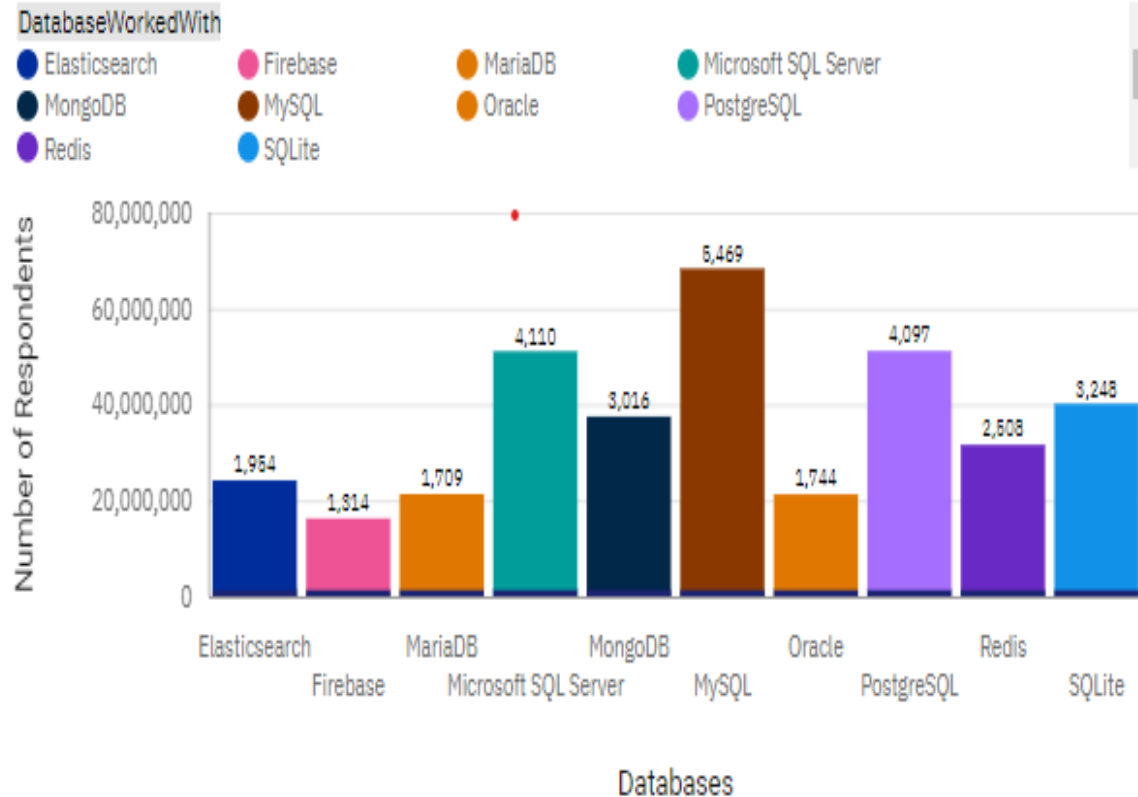
Implications

- Javascript and HTML are used for web development as a tech skill has the highest demand, especially as Typescript is getting viral. Possible developers migration from JavaScript to Typescript

- Python is gaining more and more traction due to the increase in demand for AI and ML skills.

- SQL is the still the most relevant language for data professionals. It is important for aspiring data analysts, scientists, business analysts etc to have SQL skills.

IBM Developer

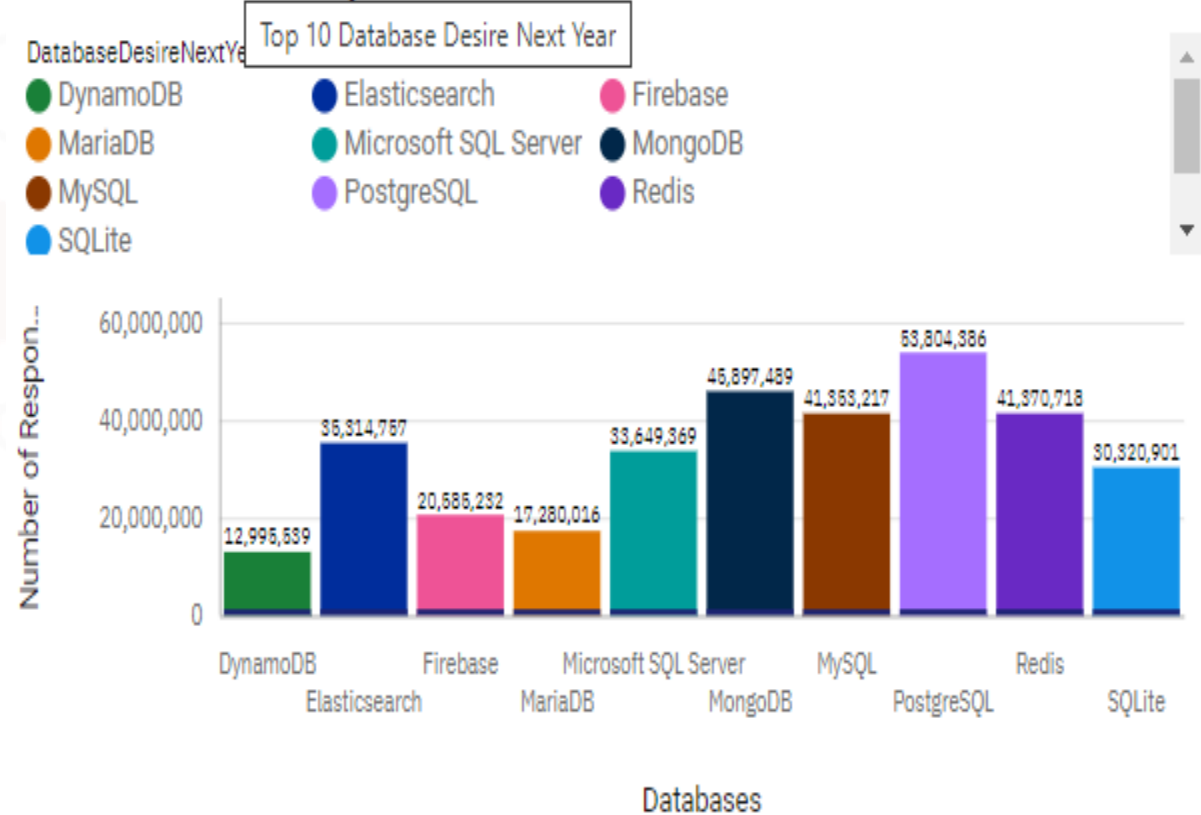SKILLS NETWORK

DATABASE TRENDS

Current Year

Next Year

Top 10 Database Worked With

Top 10 Database Desire Next Year

DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- MySQL, Microsoft SQL server, PostgreSQL SQL, SQLite and MongoDB are the top 5 most used databases at the moment. MySQL as most used database

- However, PostgreSQL SQL, MongoDB, Redis, MySQL and Elastic sense are projected to become more popular in the future.

- Redis and Elastic sense are relatively new tools and are set to gain more traction in the IT space.

Implications

- SQL is still a top tool to watch out for in data specialists.

- Companies still prefer Open source databases.

- Oracle SQL was not among the top 5. It is losing relevance as time passes.

- PostgreSQL and MongoDB establishment in the market
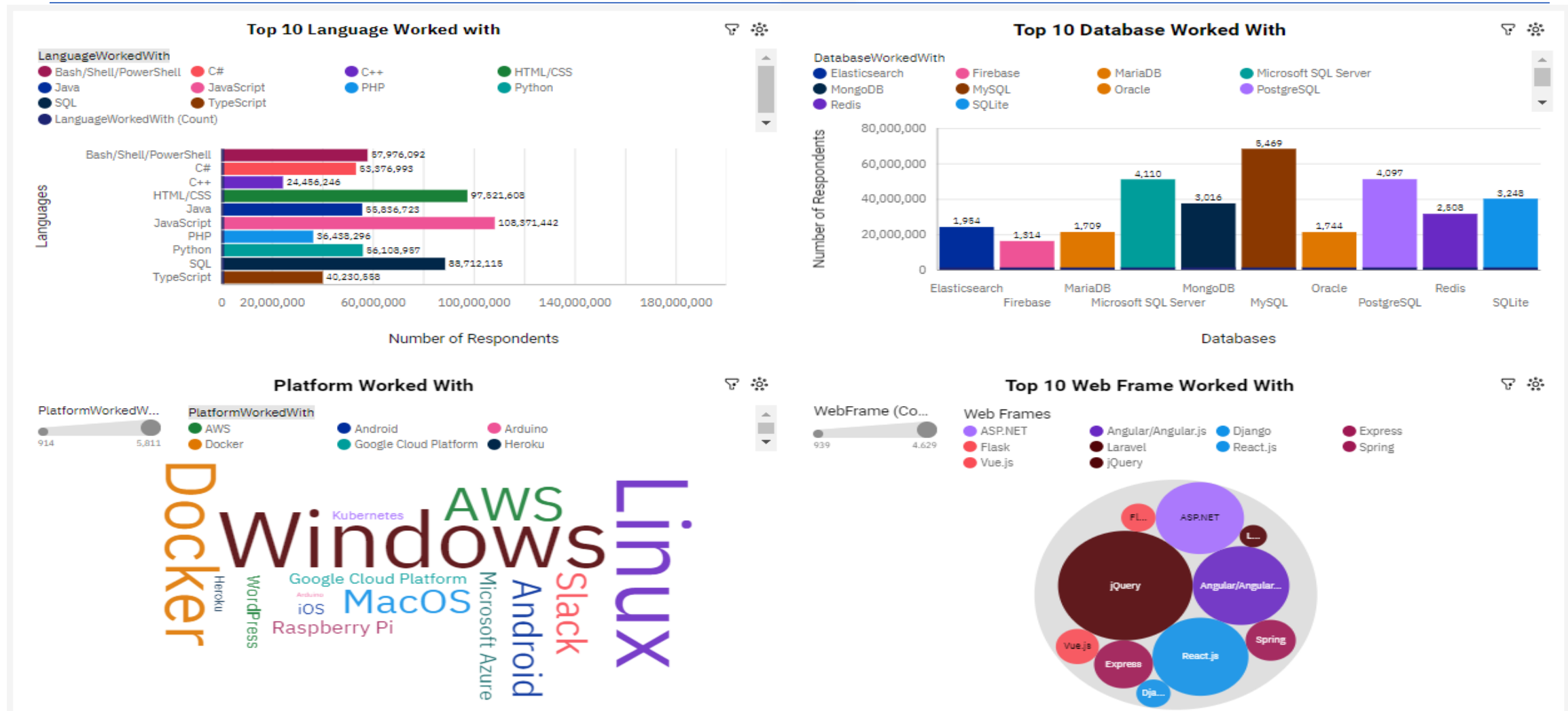
IBM Developer

SKILLS NETWORK

DASHBOARD

link IBM Cognos Analytics Dashboard Module:

https://us3.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=.my_folders%2Fm5_survey_tech%2Bdashboard&action=view&mode=dashboard&subView=model0000019133ec2664_00000002
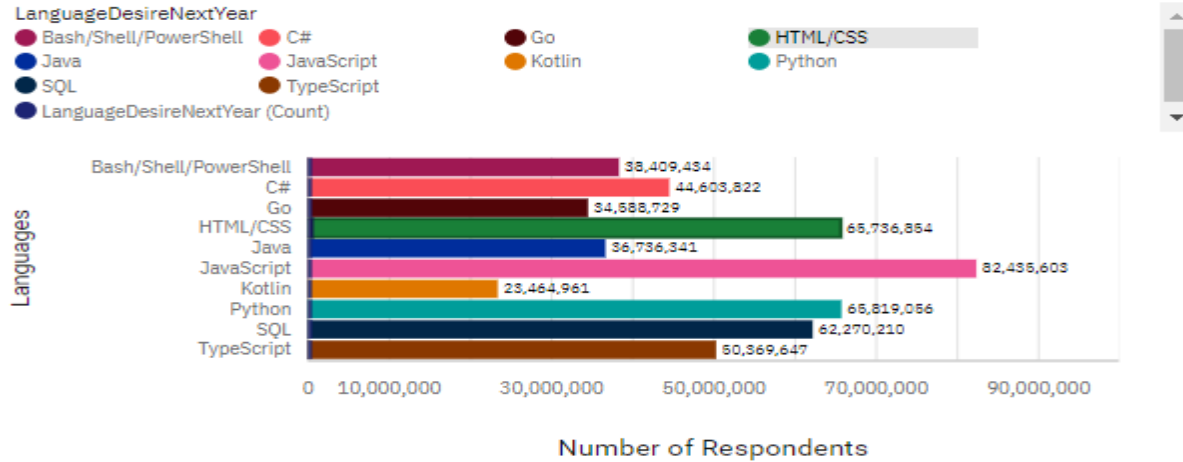
Git Hub link for cognos analytics dashboard pdf graphs

https://github.com/Keerthi-KataKam/IBM-Data-Analyst-Capstone-Project/blob/main/m5_survey_tech%20dashboard.pdf

IBM Developer

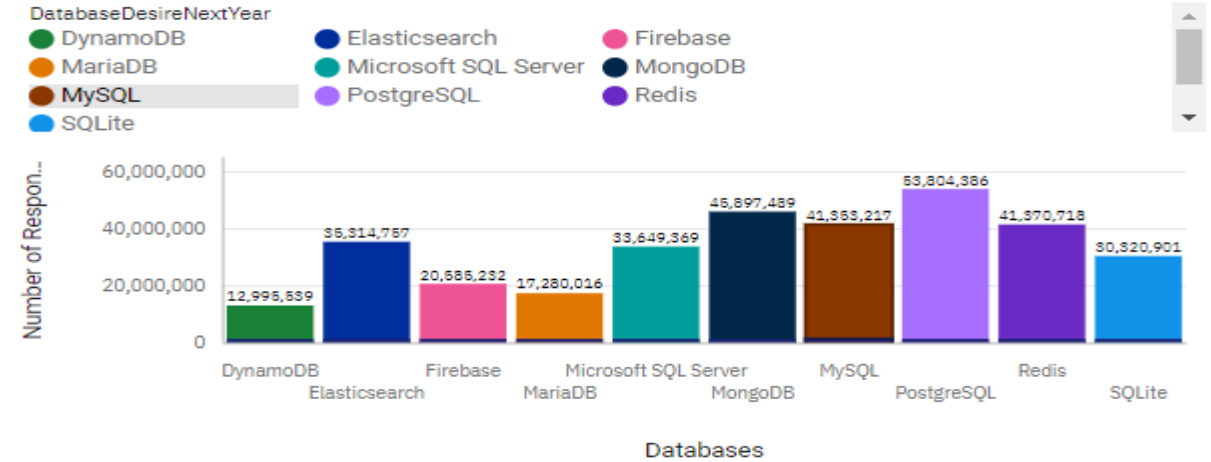SKILLS NETWORK

DASHBOARD TAB 1

DASHBOARD TAB 2

Top 10 Language Desire Next Year

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

Languages:
- Bash/Shell/PowerShell — 38,409,434
- C# — 44,603,822
- Go — 34,588,729
- HTML/CSS — 65,736,854
- Java — 36,736,341
- JavaScript — 82,435,603
- Kotlin — 23,464,961
- Python — 65,819,056
- SQL — 62,270,210
- TypeScript — 50,969,647

Number of Respondents

Top 10 Database Desire Next Year

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

Number of Respon...
- DynamoDB — 12,995,639
- Elasticsearch — 36,314,757
- Firebase — 20,686,232
- MariaDB — 17,280,016
- Microsoft SQL Server — 33,649,369
- MongoDB — 46,897,489
- MySQL — 41,353,217
- PostgreSQL — 53,804,386
- Redis — 41,370,718
- SQLite — 30,320,901

Databases

Platform Desire Next Year
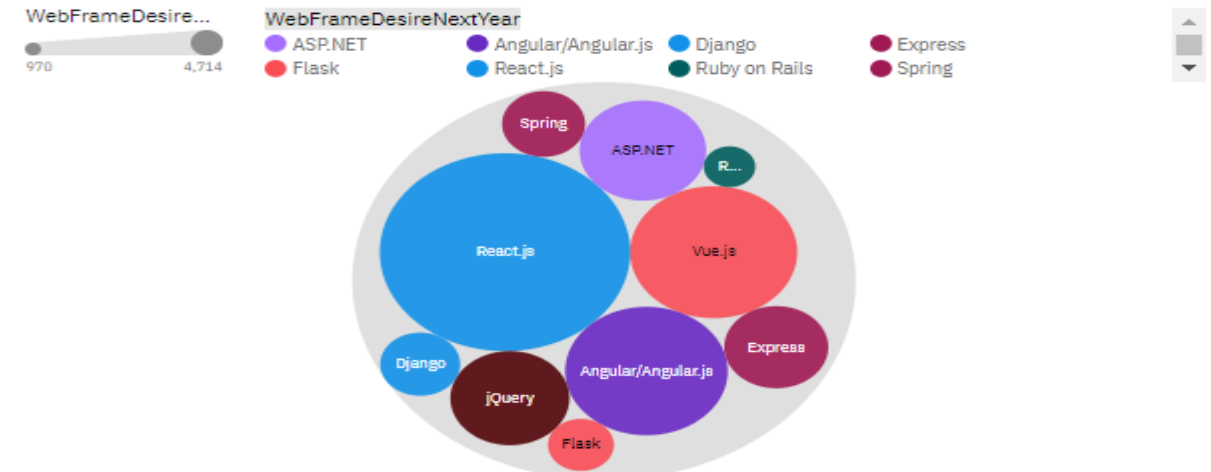
PlatformDesireNe... 0 — 5,865
PlatformDesireNe... 0 — 5,865

- Linux
- Docker
- AWS
- Windows
- Android
- Kubernetes
- MacOS
- Raspberry Pi
- Google Cloud Pl...
- Microsoft Azure
- iOS
- Slack
- Arduino
- Heroku
- WordP...
- IB...

Top 10 Web Frame Desire Next Year

WebFrameDesire... 970 — 4,714

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

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

IBM Developer

SKILLS NETWORK

DASHBOARD TAB 3
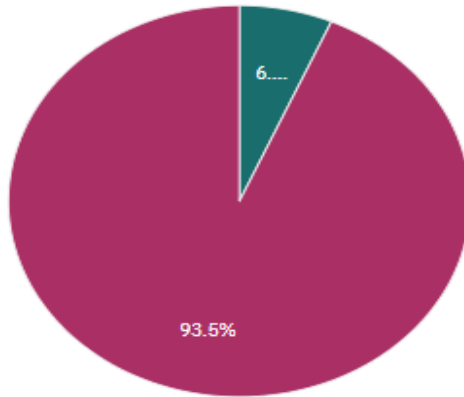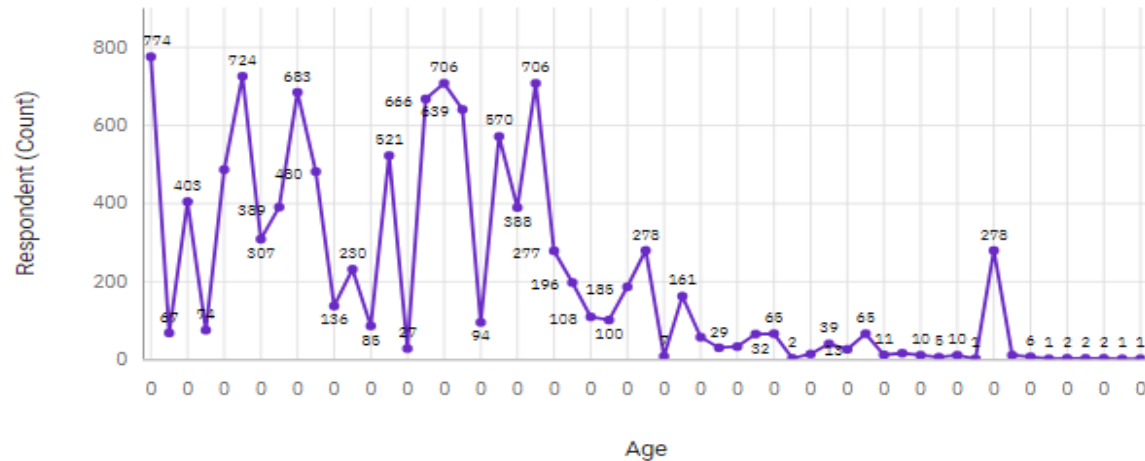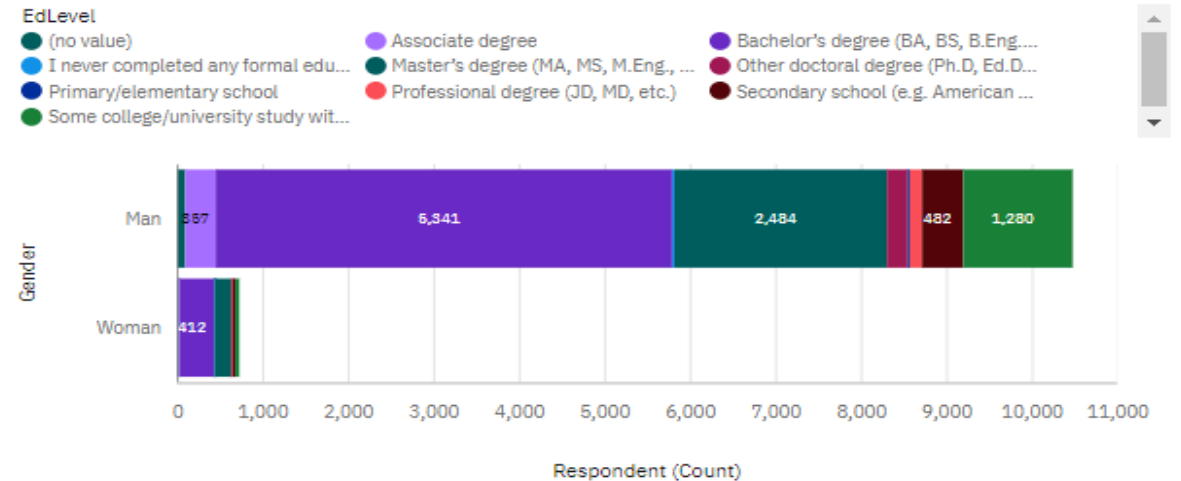
DISCUSSION

- Up skilling in the Technology sector.

- How do we close the wide gender gap in the Technology sector?

- Is completing a masters or doctorate degree really a requirement?

- The increasing demand for mobile development as Kotlin is getting popular.

- More tech education, access and development in less developed regions in South east Asia, South America, Africa and some parts of Europe.

- How relevant will Oracle SQL still be in the future?

OVERALL FINDINGS & IMPLICATIONS

Findings

- Web development languages are the most popular and on-demand tools in the IT field currently

- Over 90% young male Developers. The Tech sector is filled with majorly young people under 40 years of age.

- Developers mostly located in developed countries

- Most people in the IT field have a Bachelors' degree.

Implications

- It is important for data professionals to develop proficiency in NoSQL in addition to SQL databases.

- Web development is still a very lucrative skill.

- Young developers without postgrad studies are there

- Less developed countries need more access to tech training and education

**IBM Developer**

SKILLS NETWORK

CONCLUSION

- Developers are people with very marked characteristics.

- A good idea of popularity trends of different tools, platforms and languages can be obtained.

- It is expedient to stay updated in the Tech sector as the trends keep changing over time.

- There is a job to be done to spread accessibility of this labor market to countries in development

IBM **Developer**

SKILLS NETWORK

APPENDIX

Plot the distribution curve for the column `ConvertedComp`.

```python
# your code goes here
import seaborn as sns
sns.displot(df['ConvertedComp'], kde = True)
```
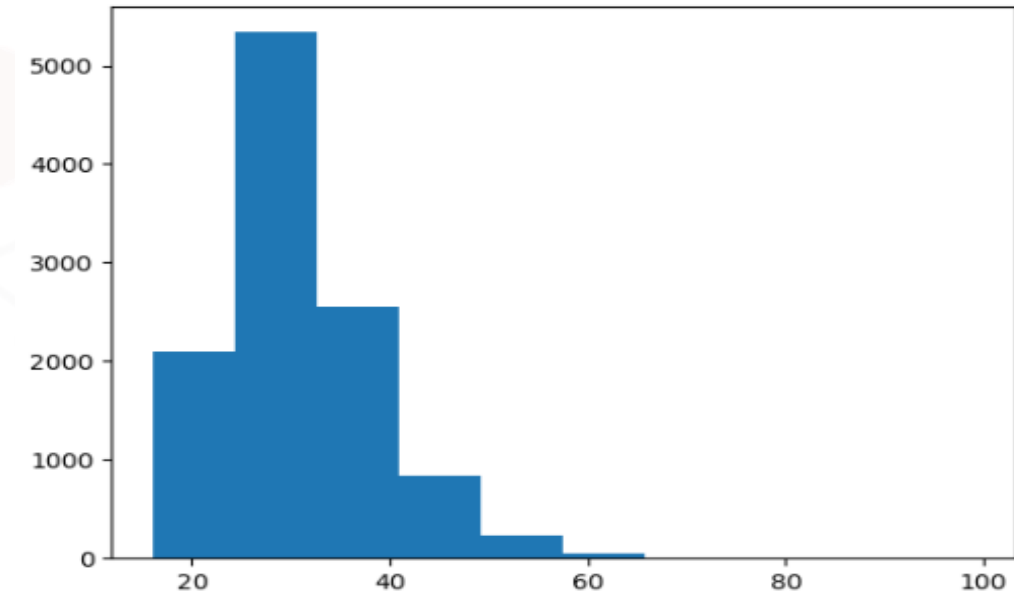
```
<seaborn.axisgrid.FacetGrid at 0x20e010c32c0>
```
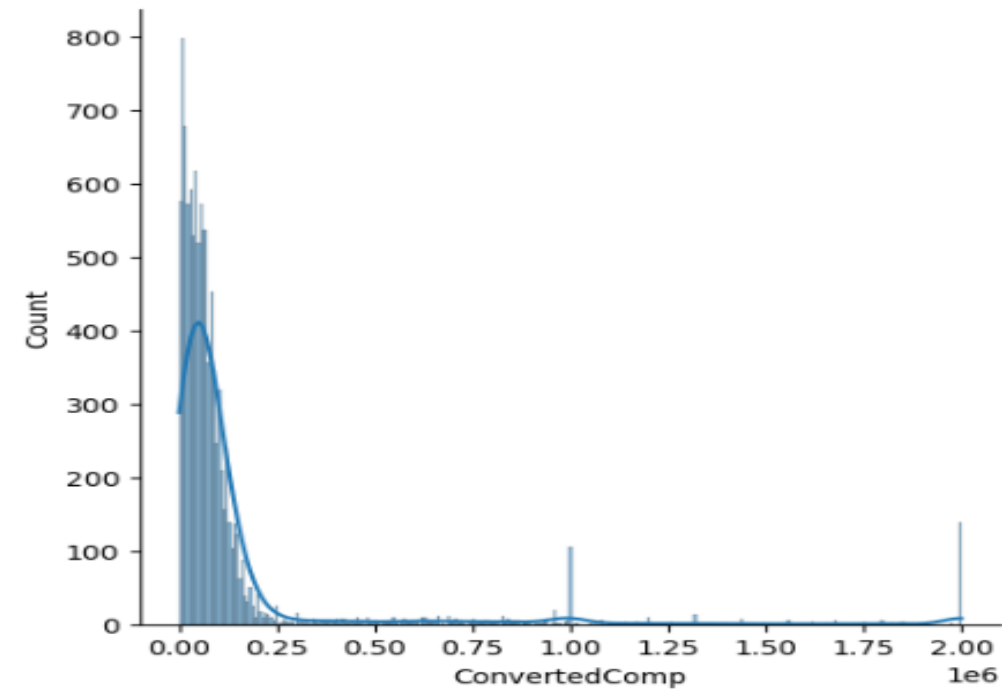

Plot a histogram of the column `Age`.

```python
# your code goes here
plt.hist(df['Age'])
```

```
(array([2.094e+03, 5.337e+03, 2.557e+03, 8.420e+02, 2.250e+02, 4.900e+01,
        6.000e+00, 0.000e+00, 0.000e+00, 1.000e+00]),
 array([16. , 24.3, 32.6, 40.9, 49.2, 57.5, 65.8, 74.1, 82.4, 90.7, 99. ]),
 <BarContainer object of 10 artists>)
```

APPENDIX

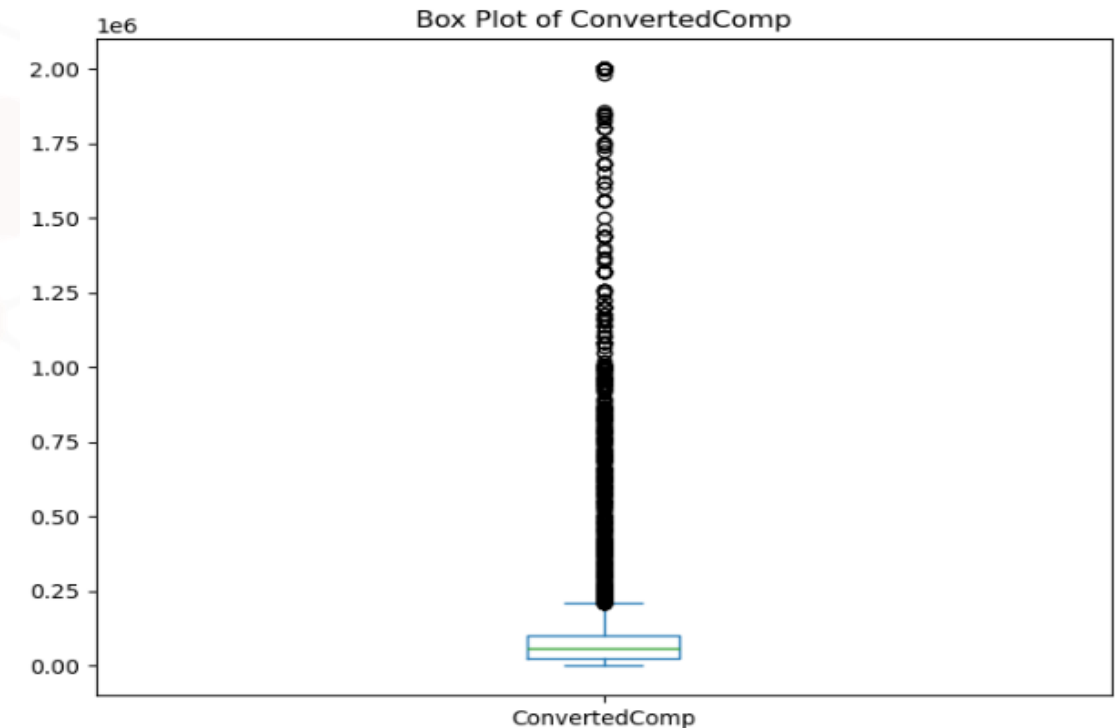Plot the histogram for the column `ConvertedComp` .

```python
# your code goes here
import matplotlib as mpl
%matplotlib inline
import matplotlib.pyplot as plt
plt.hist(df['ConvertedComp'])
plt.title('Histogram of ConvertedComp')
plt.xlabel('ConvertedComp')
plt.ylabel('Number of Responders')
plt.show()
```
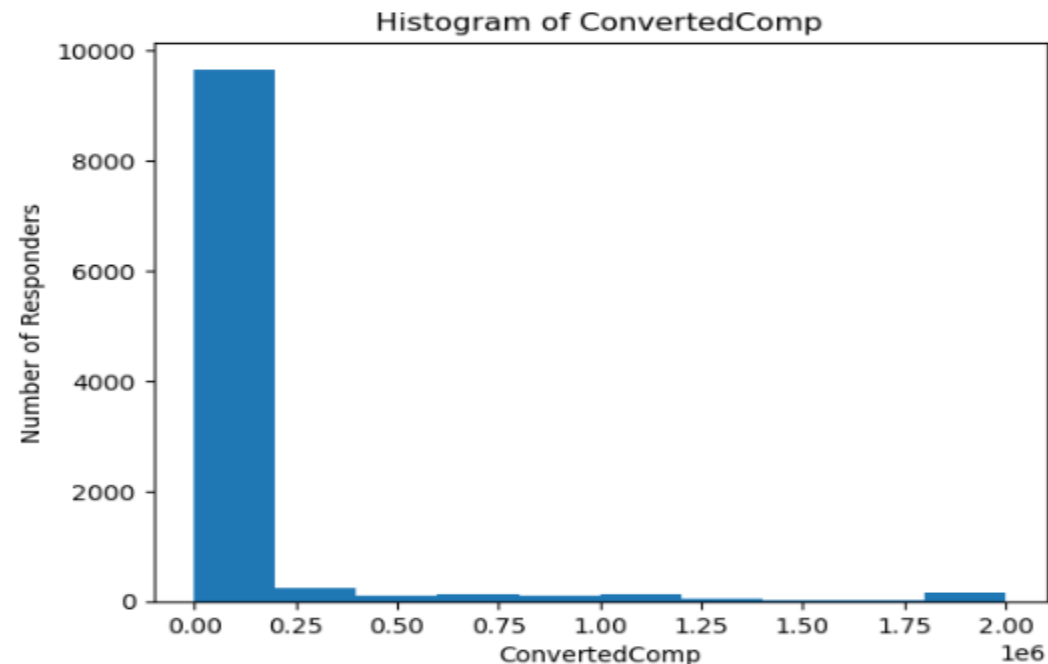
Finding outliers
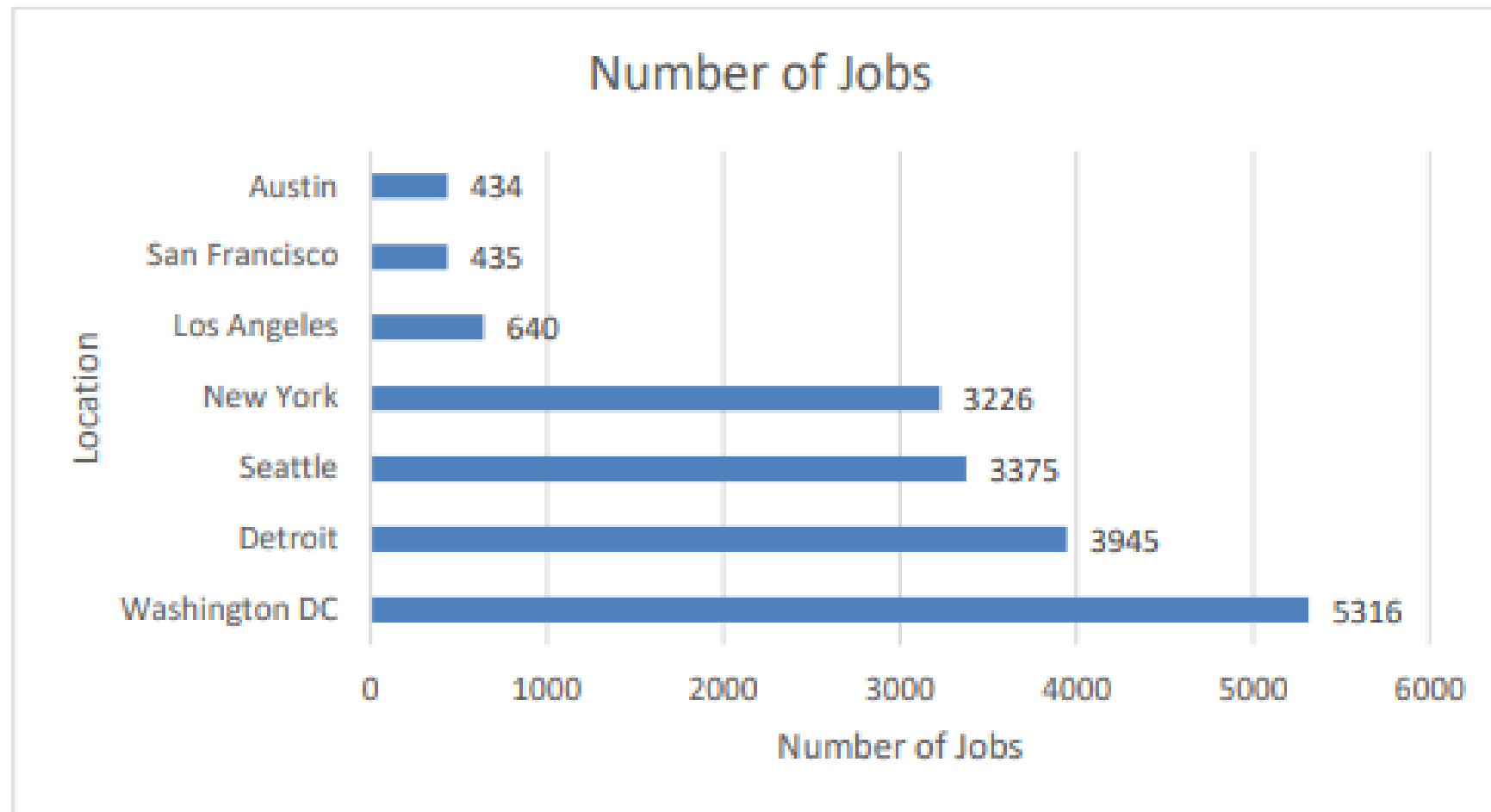
Find out if outliers exist in the column `ConvertedComp` using a box plot?

```python
# your code goes here
df_cv = pd.DataFrame(data=df['ConvertedComp'])
df_cv.plot(kind='box', figsize=(8,6), title = 'Box Plot of ConvertedComp')
plt.show()
```

JOB POSTINGS

Number of Jobs

Location	Number of Jobs
Austin	434
San Francisco	435
Los Angeles	640
New York	3226
Seattle	3375
Detroit	3945
Washington DC	5316

IBM Developer

SKILLS NETWORK

POPULAR LANGUAGES

Average Annual Salary

Language	Salary
Go	$94,082
SQL	$84,793
PHP	$84,727
C#	$88,726
C++	$113,865
Swift	$130,801
Javascript	$110,981
R	$92,037
Java	$101,013
Python	114,383

Average Annual Salary