2019 STACK OVERFLOW DEVELOPER SURVEY

Yuvarekha Mahendran

January 12,2024

IBM Developer

SKILLS NETWORK

OUTLINE

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

EXECUTIVE SUMMARY

- Data analysing goal
- Description
  - Gathering data
  - Analysis of data
  - Visualization of data
- Results shown with the help of dashboard outputs as graphs
- Overall findings and implications with respect to the results shown in graph
- Overall conclusion

INTRODUCTION

- Stack overflow’s annual developer survey 2019 is the most comprehensive survey of people who code globally
- Currently,there are 90000 developers according to May 2023 data
- Data keeps changing every year as programming languages evolve
- Data is analysed by two ways
  - Technology trends
  - Demographics

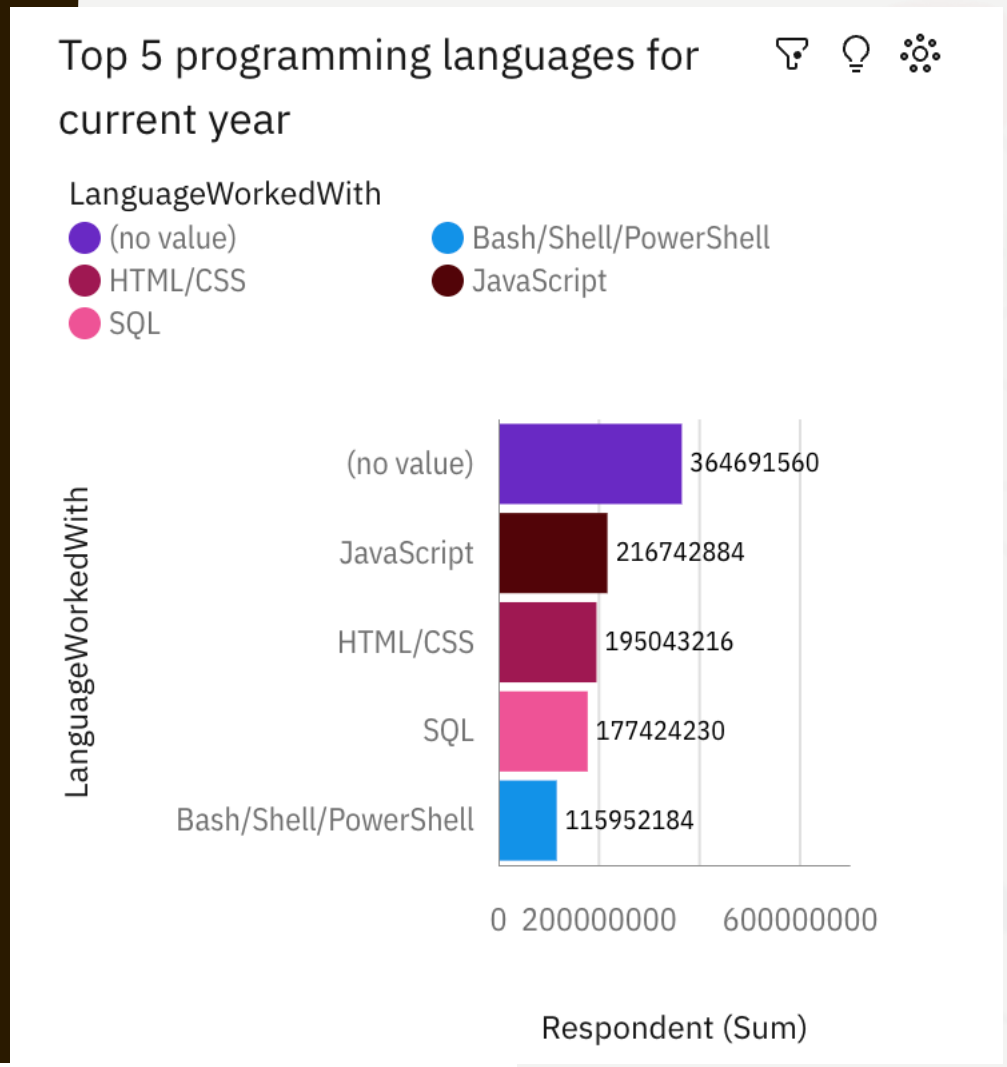
METHODOLOGY

- Data collection and glance through the dataset
  - 1)Scraping data from web
  - 2) API
  - 3)Importing requests library
- Data wrangling
- Exploratory data analysis
  - 1)Data analysis with data distribution
  - 2)Outlier detection
  - 3)Correlation of data
  - – Data visualization
  - – Output

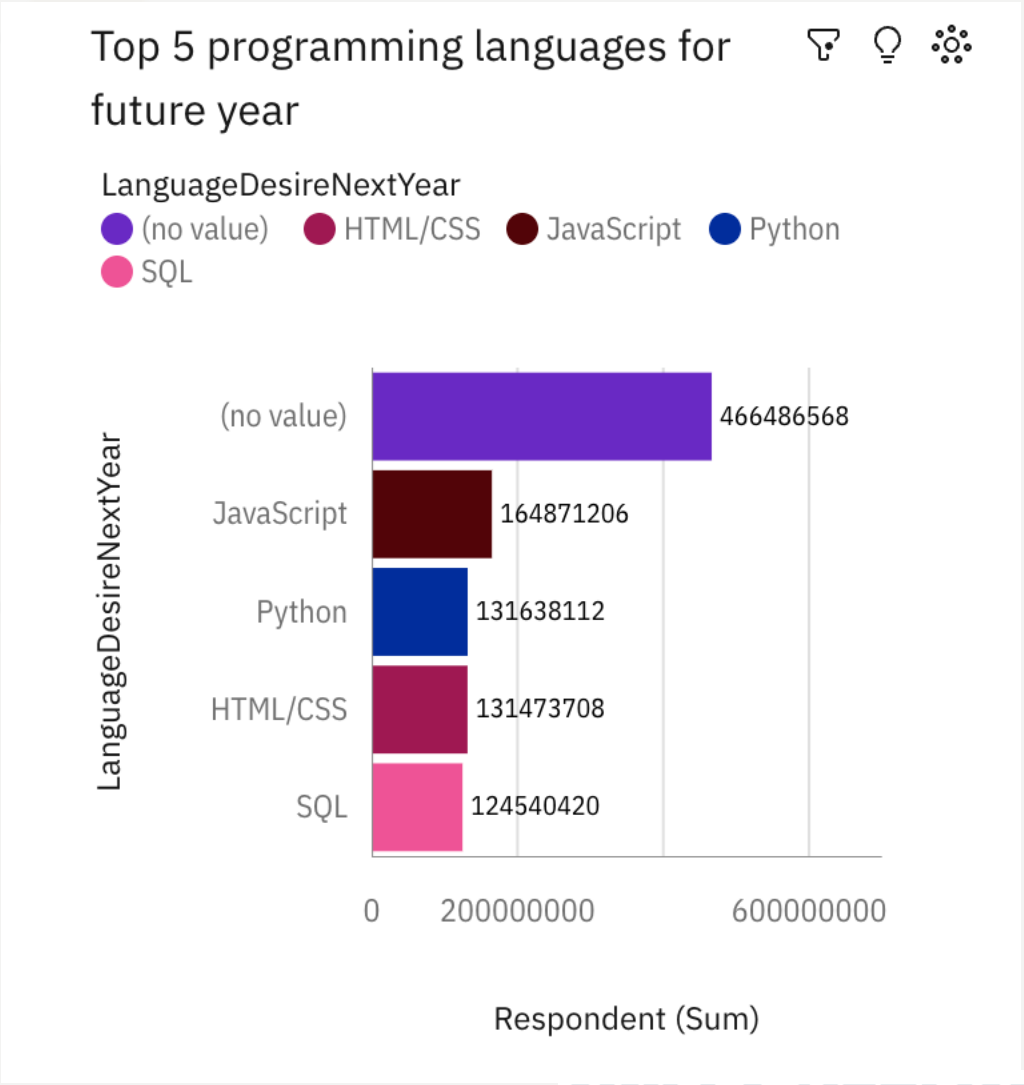
RESULTS

PROGRAMMING LANGUAGE TRENDS

Current Year



Next Year



IBM Developer

SKILLS NETWORK

PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

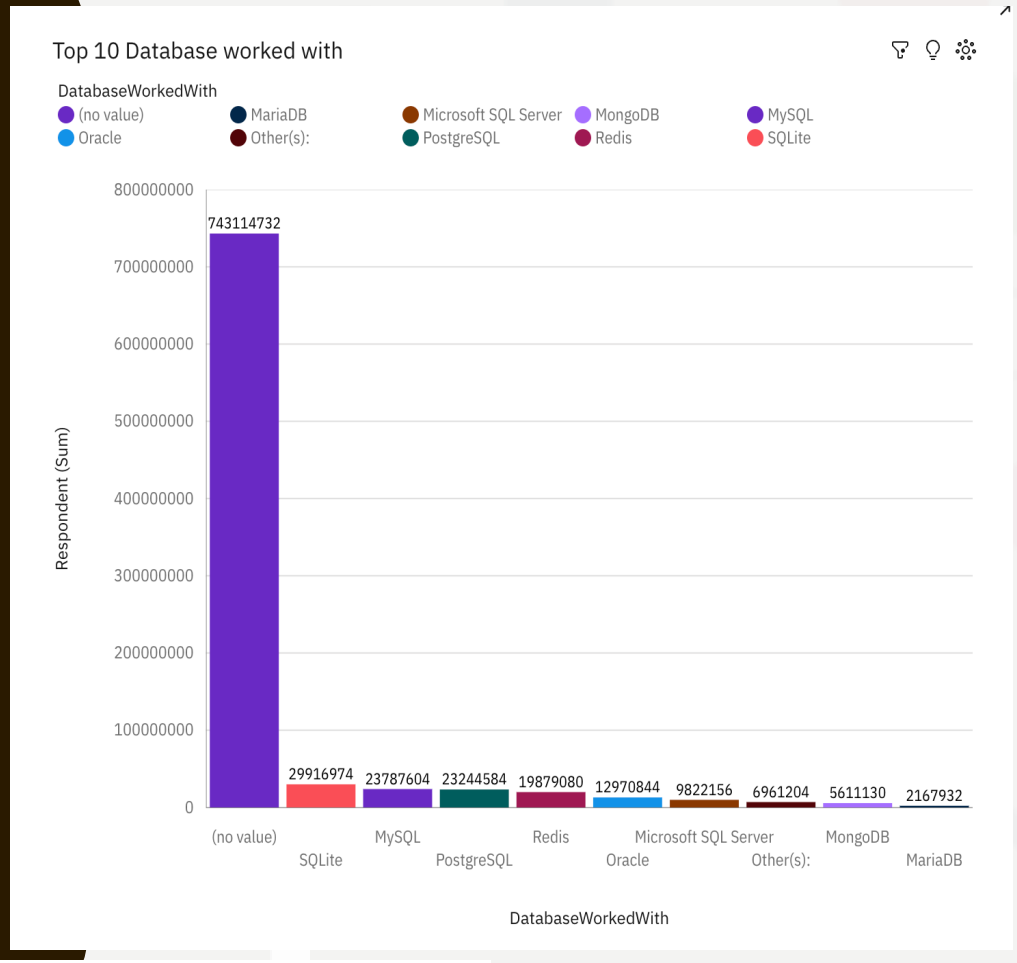
- Javascript tends to be second fastest growing language
- People's interest in scripting languages keeps increasing in current year
- Future trends displays python as one of the growing languages
- No interest towards Bash language in future trends

Implications

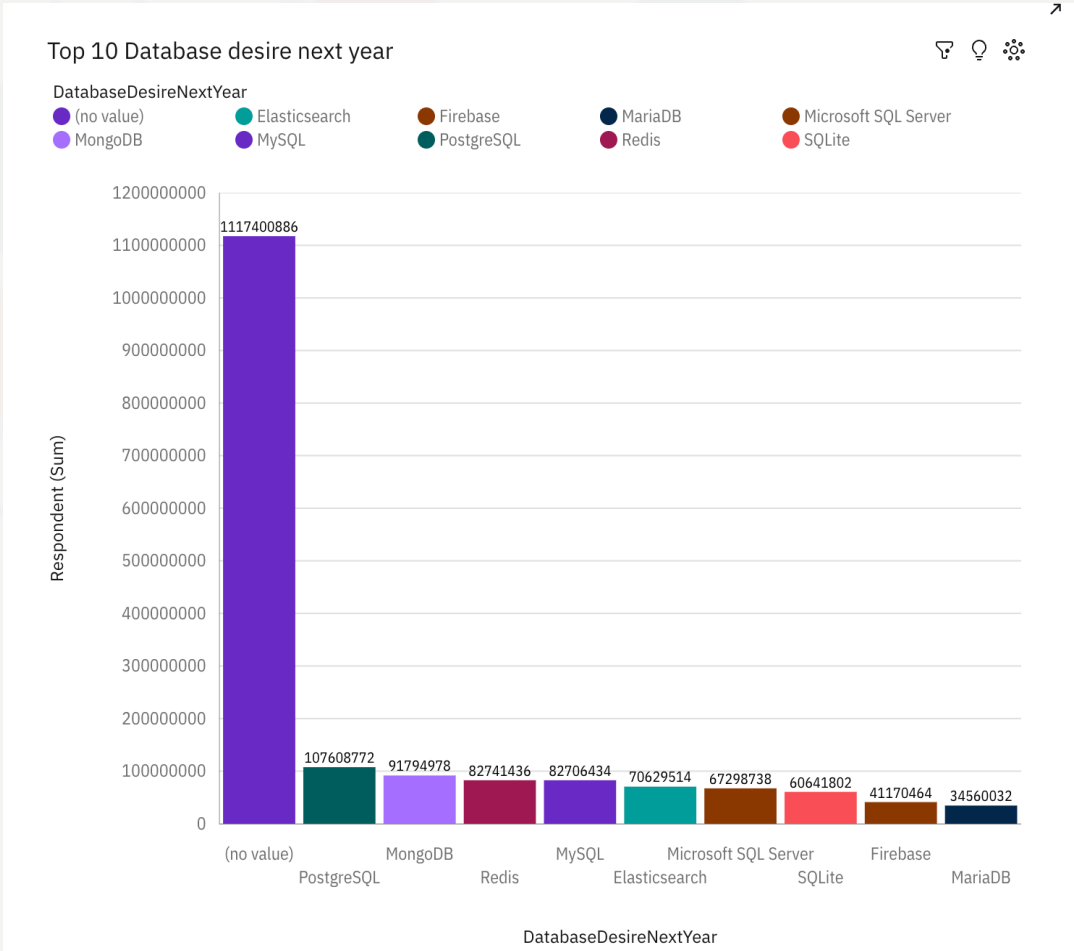
- Shift in first fastest growing language in future trends
- Fond for python programming by future developers
- Increased skillset with python programming language
- Less focus on SQL in future years

DATABASE TRENDS

Current Year



Next Year



IBM Developer

SKILLS NETWORK

DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- SQLite becomes the fastest evolving language
- Increased interest in MongoDB,MariaDB in future trends compared to current trends
- New interest in Firebase in future findings
- Decent trends in Oracle database currently

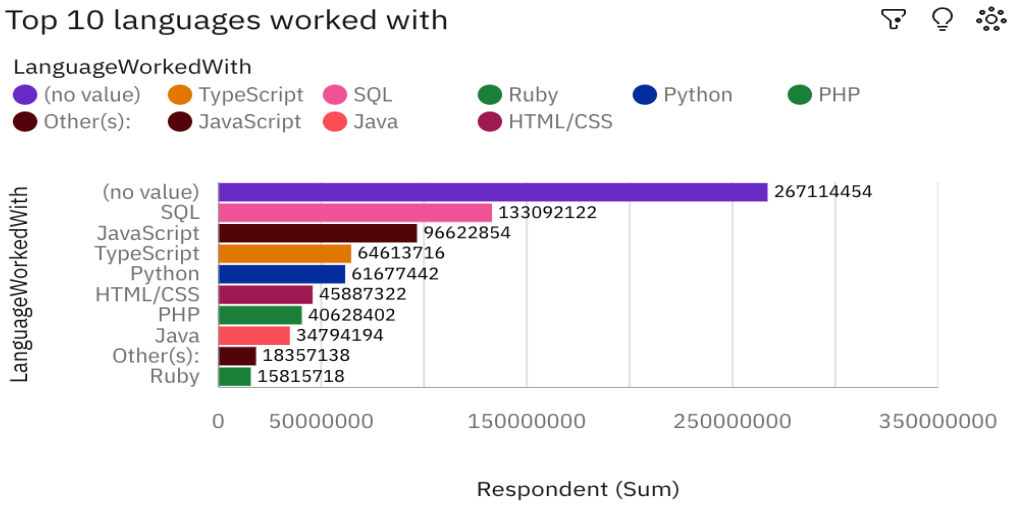
Implications

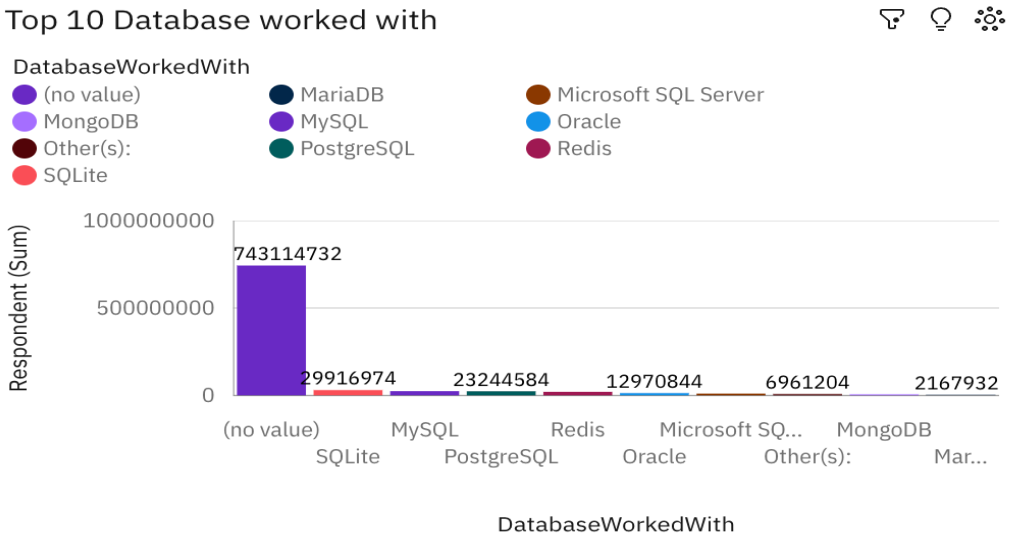
- Introduction to Firebase in evolving market
- Shift in interest towards MongoDB and MariaDB
- Microsoft SQL Server has increased interest
- Shift in database software according to the demand

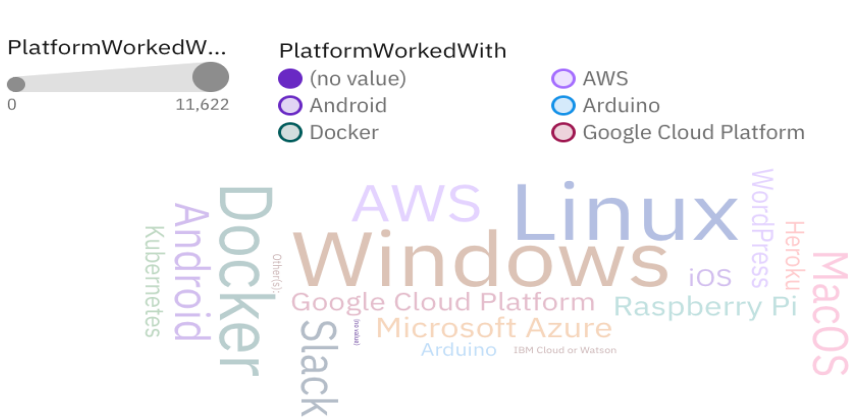
DASHBOARD

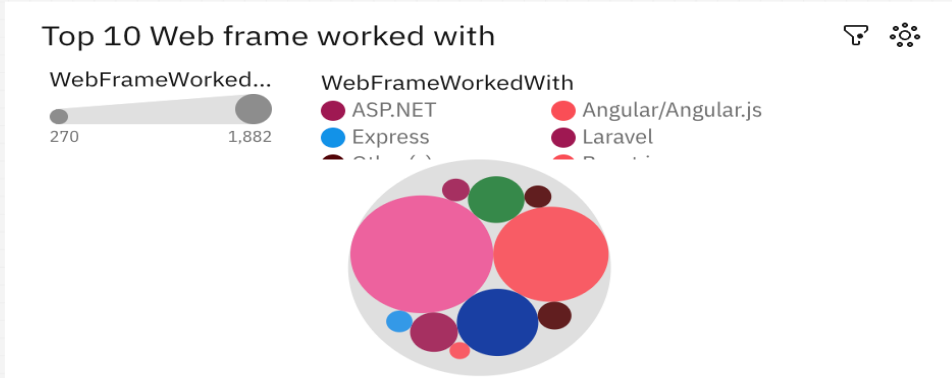
https://us3.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=.my_folders%2FCapstone&action=view&mode=dashboard&subView=model00000018cfb7900bc_00000000

DASHBOARD TAB 1





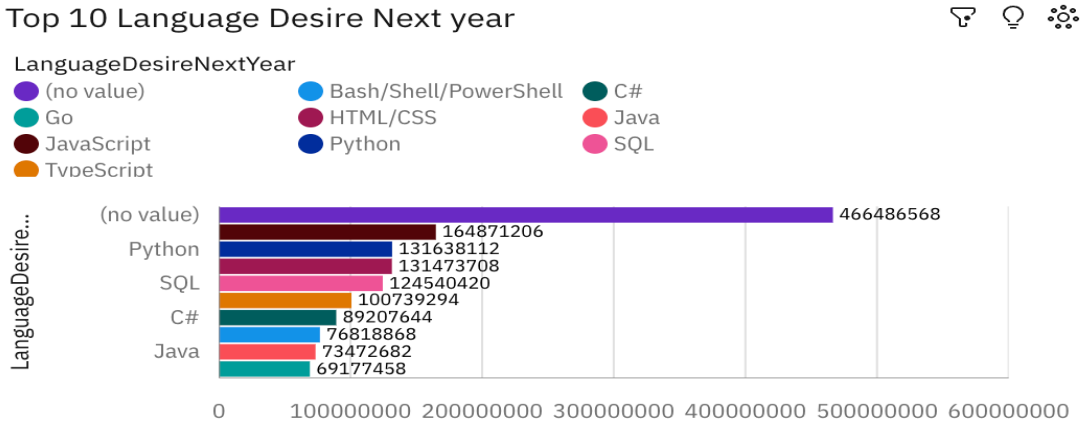


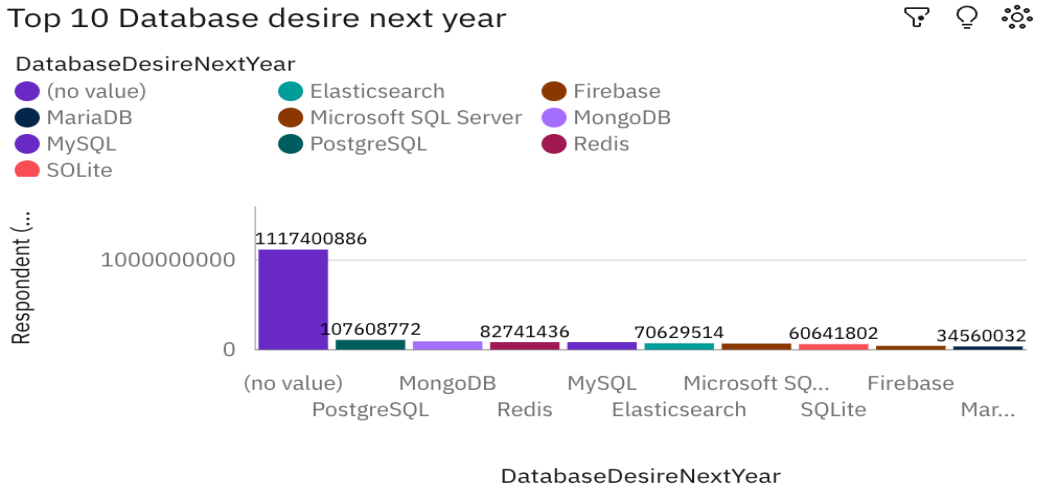


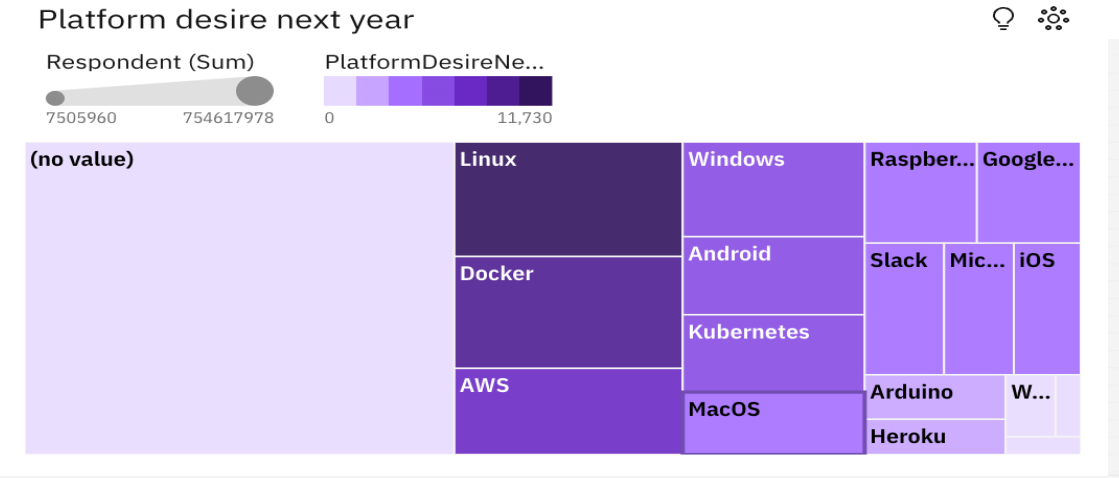
IBM Developer

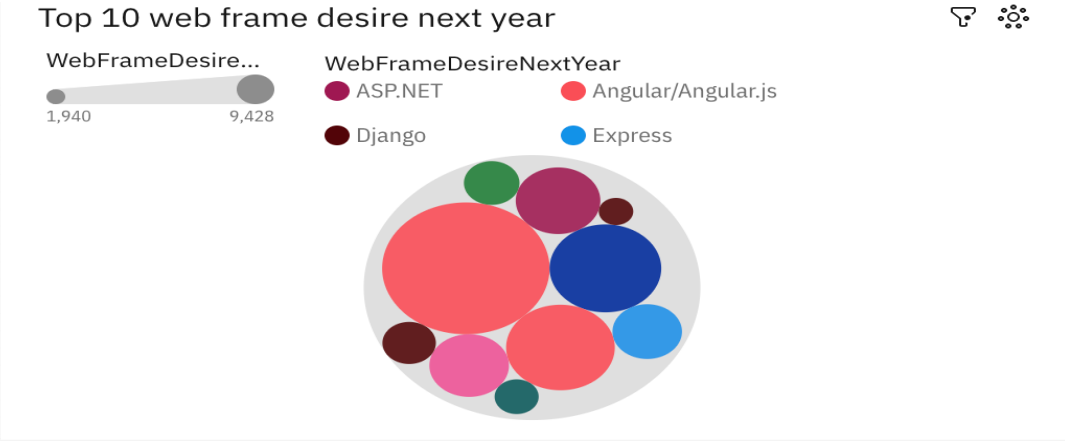
SKILLS NETWORK

DASHBOARD TAB 2





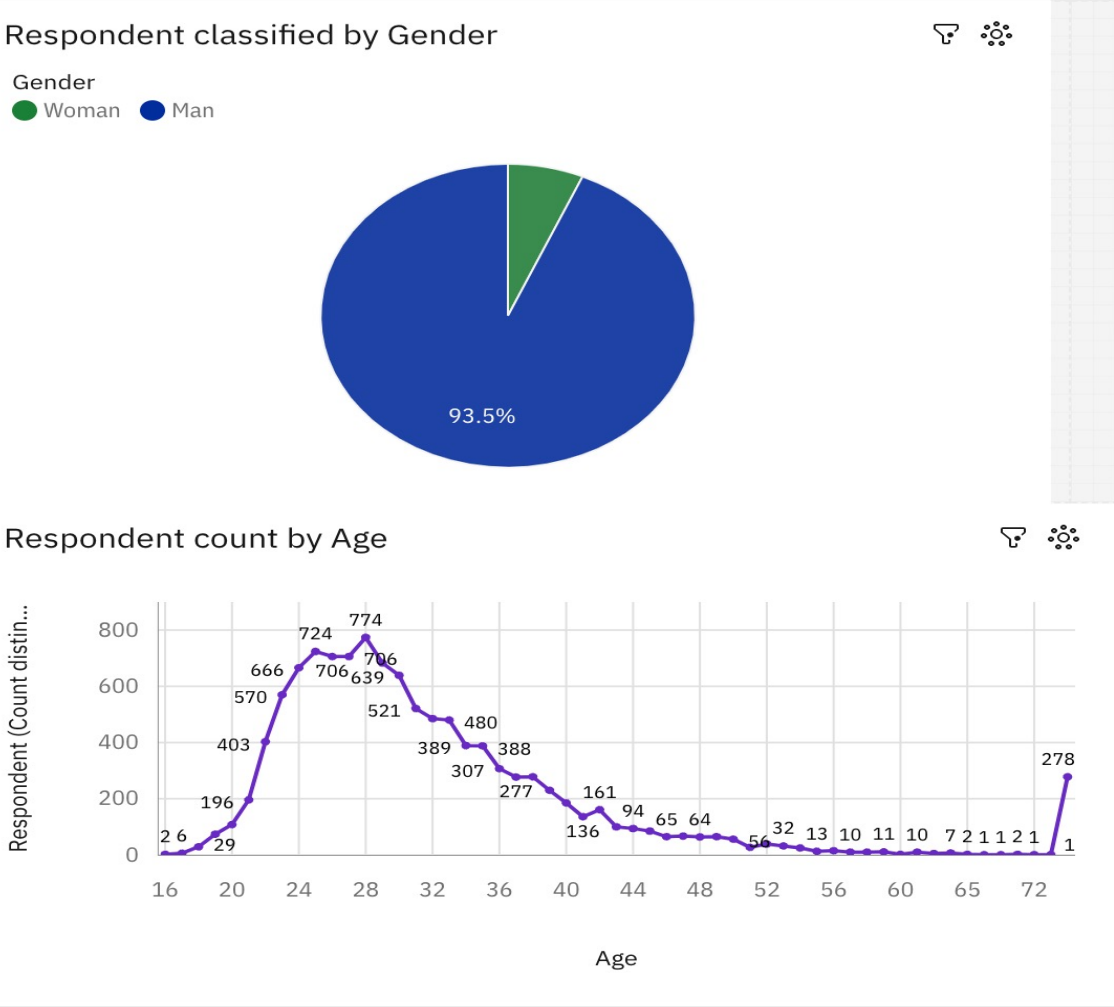


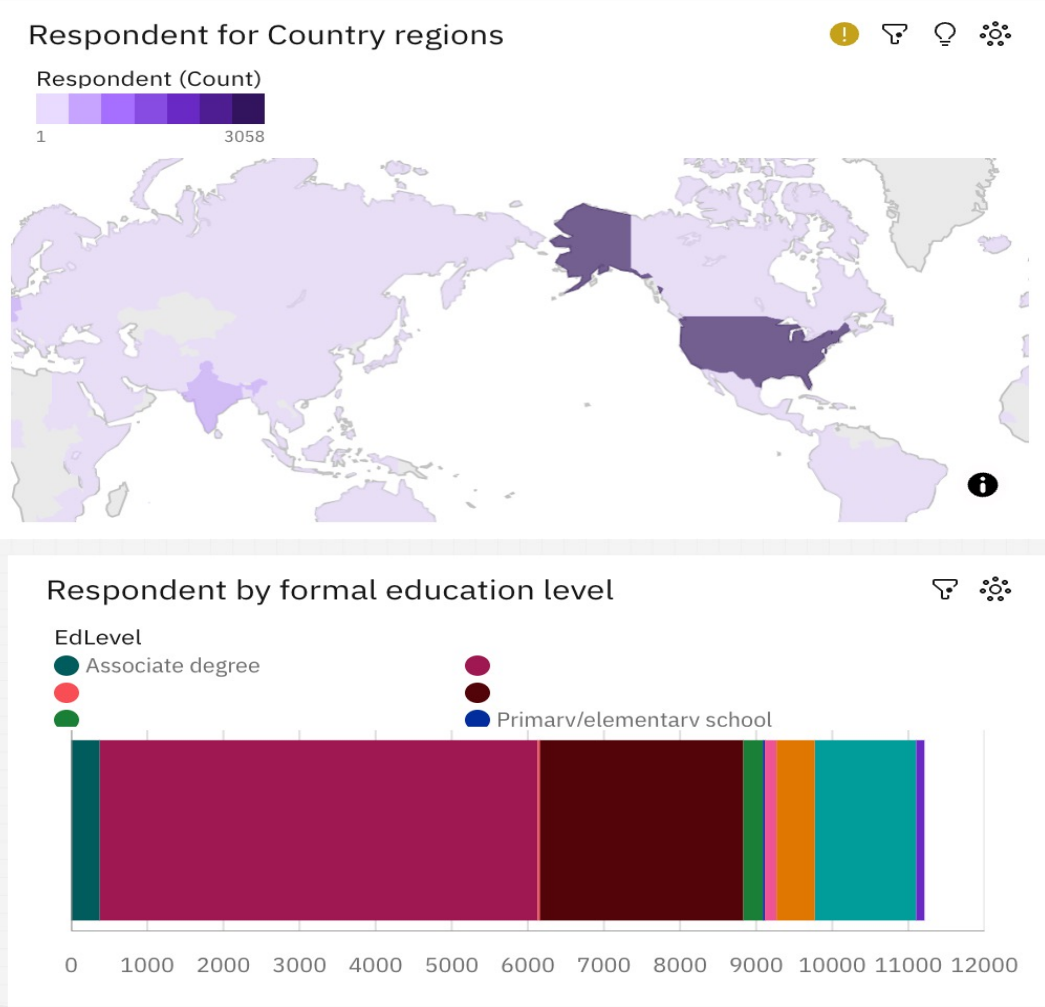


IBM Developer

SKILLS NETWORK

DASHBOARD TAB 3





IBM Developer

SKILLS NETWORK

DISCUSSION

- Increased interest towards database in future years
- Growth in technological sectors
- The increasing demand in database software like MongoDB,MariaDB in future years
- What about Swift and GoLang programming languages?

OVERALL FINDINGS & IMPLICATIONS

Findings

- Minimum qualification of all developers is BE in CSE
- Increased interest towards database softwares
- Python is fastest growing programming language
- Increased interest towards database languages

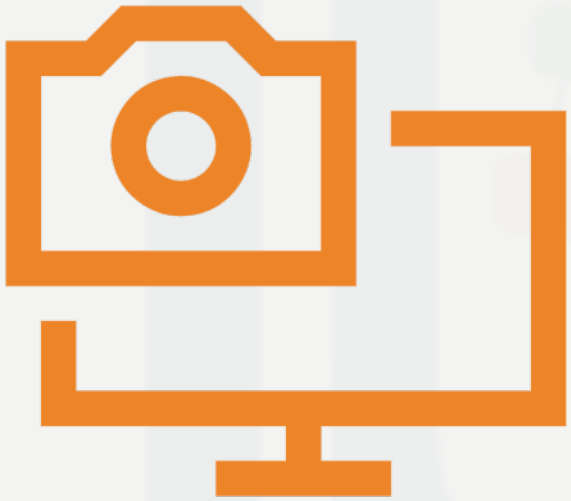
Implications

- Data professionals must be proficient in SQL,tableau and Excel with python
- Increased concentration in basic foundational topics in python
- Professional growth globally is essential
- Bash is not given importance in future years

CONCLUSION

- It is important for data professionals to be proficient in both scripting and programming languages for a successful career in upcoming technology improvements
- Importance to database is also equally important
- Data visualizations like Tableau should also be very helpful tool to view as dashboards
- Data professionals should be expert in calculations with Excel spreadsheet

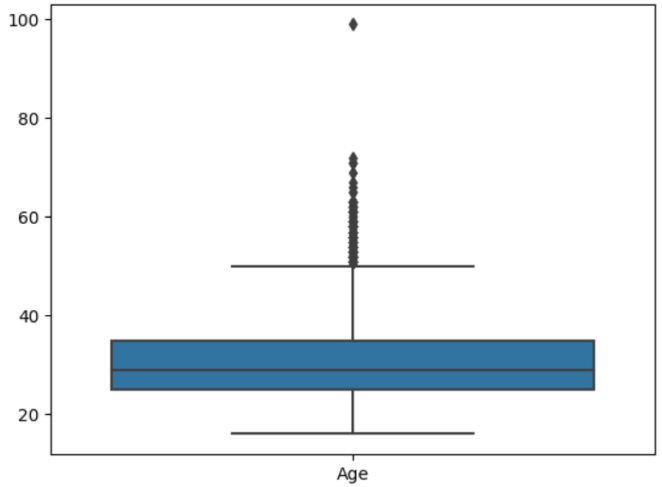
APPENDIX



Plot a box plot of `Age.`

```
# your code goes here
QUERY = """
SELECT Age
FROM master
"""
df = pd.read_sql_query(QUERY,conn)
df.head()
sns.boxplot(df,)

plt.show()
```



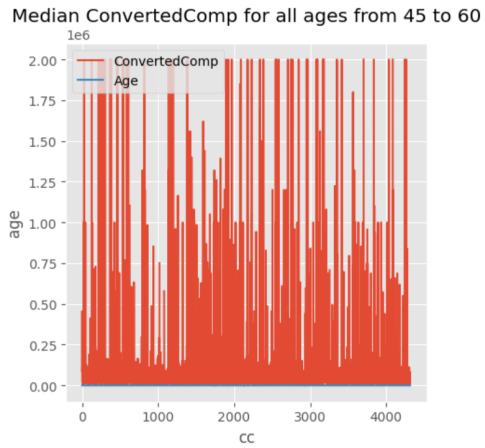
IBM Developer

SKILLS NETWORK

[ ]

	ConvertedComp	Age
count	4.012000e+03	4317.000000
mean	1.155598e+05	27.448923
std	2.786517e+05	1.683755
min	0.000000e+00	25.000000
25%	2.199600e+04	26.000000
50%	4.950000e+04	27.000000
75%	8.483875e+04	29.000000
max	2.000000e+06	30.000000

```
1 line_df.plot(kind = 'line', figsize = (5,5))
2 plt.xlabel('cc')
3 plt.ylabel('age')
4 plt.title('Median ConvertedComp for all ages from 45 to 60')
5 plt.show()
```



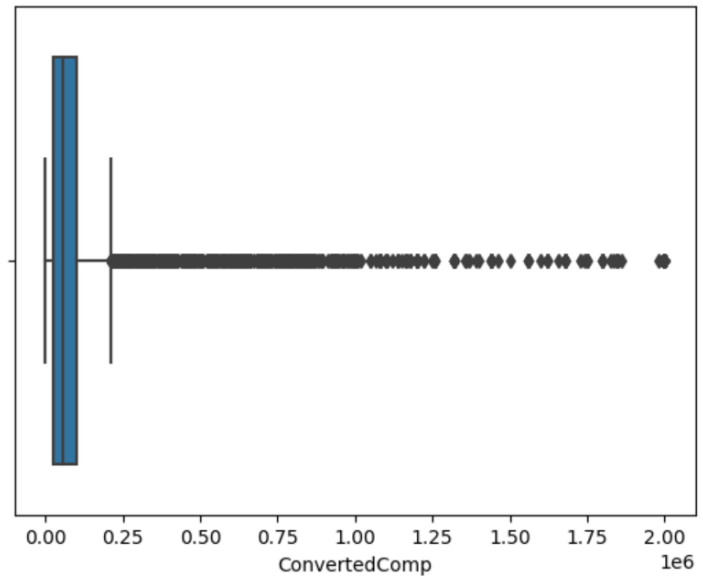
Finding outliers

Find out if outliers exist in the column `ConvertedComp` using a box plot?

[ ]

```
1 # your code goes here
2 sns.boxplot(x=df["ConvertedComp"])
```

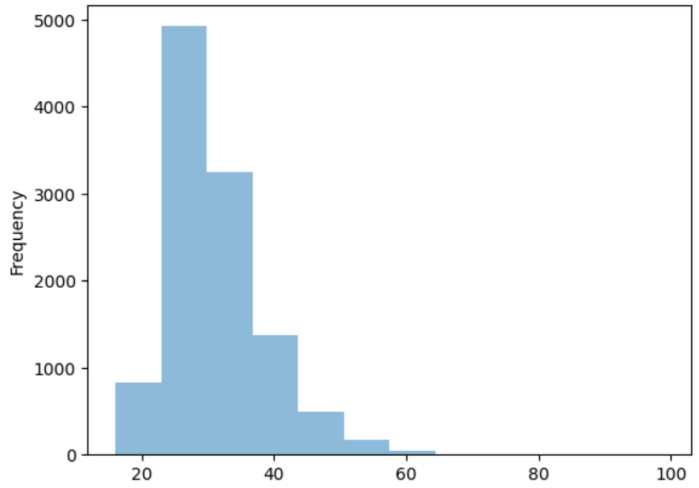
```
<Axes: xlabel='ConvertedComp'>
```

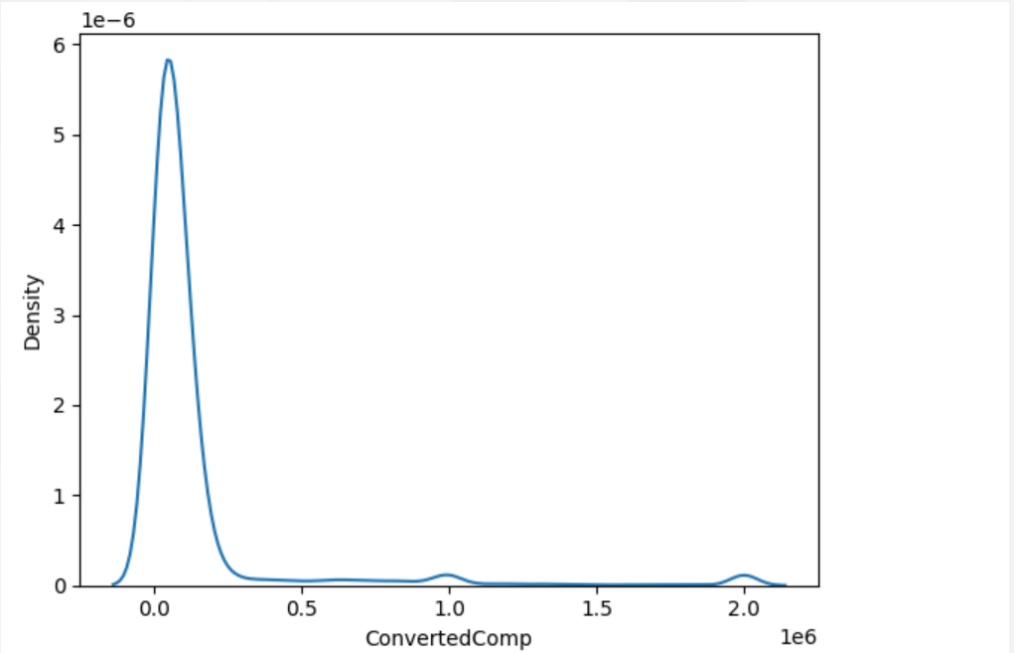


Plot a histogram of the column `Age`.

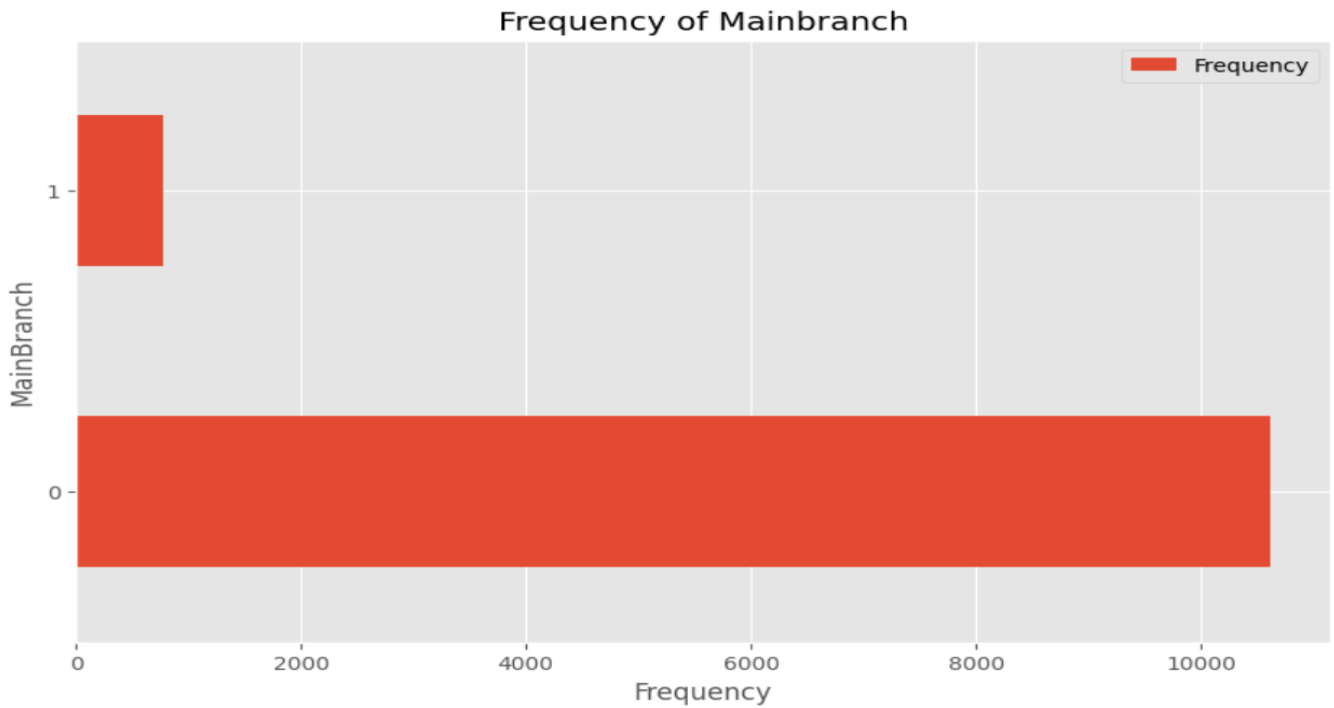
```
[ ]  1 # your code goes here
     2 df['Age'].plot.hist(bins=12, alpha=0.5)
```

```
<Axes: ylabel='Frequency'>
```

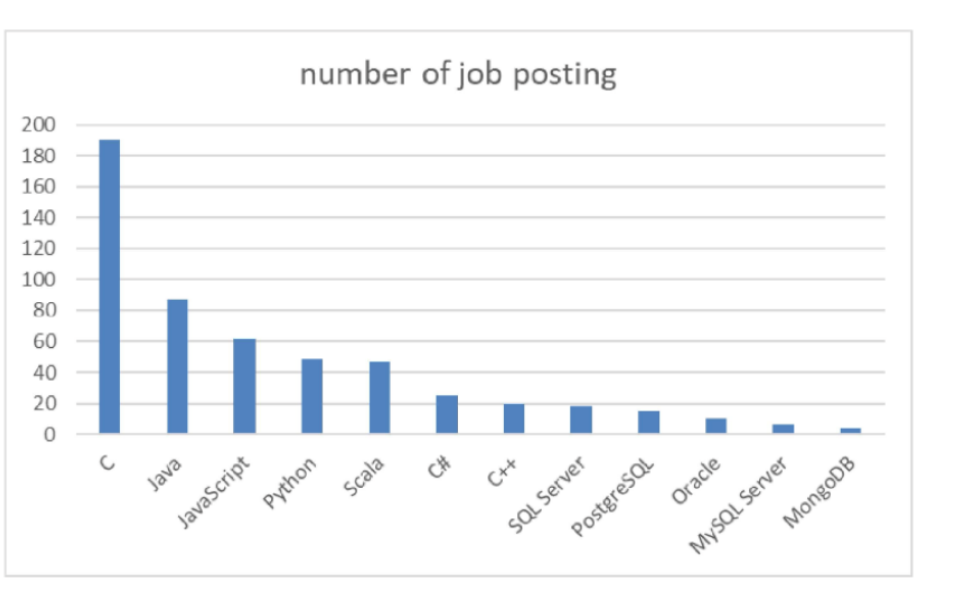




```
# your code goes here
bar_df.plot(kind = 'barh', figsize=(10,6))
plt.xlabel('Frequency')
plt.ylabel('MainBranch')
plt.title('Frequency of Mainbranch')
plt.show()
```



JOB POSTINGS



POPULAR LANGUAGES