



Identifying Emerging Technological Trends

Mhlongo M

21 August 2022

OUTLINE



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

EXECUTIVE SUMMARY



- Survey data of Software Professionals consisting of 11398 rows and 85 columns
- Data consists ~93% of Men
- Dominate Age Group Range: 20-35 years
- Top 5 programming Languages:
 - JavaScript, HTML/CSS, SQL, Bash/Powershell, and Python
- Most in demand programming Language: JavaScript
- Top 5 Databases:
 - MySQL, PostgreSQL, MS SQL Server, SQ Lite, and Mango DB
- Most Database Skill in demand: PostgreSQL

INTRODUCTION

- Data Analytics is a process of analyzing raw data in order to make conclusions about that information.
- Methods involved:
 - Qualitative Data Analysis
 - Quantitative Data Analysis
- There are typically 5 steps involved showed in the image on the right:
- **Aim:**
 - Analyze Several datasets to help identify trends for emerging technologies
 - **Objectives:**
 - Identify top programming languages in demand
 - Database skills in demand
 - Most popular IDE's
 - Demographic data for developers



METHODOLOGY

Data Collection

- API's and Web scrapping
- Data Used
 - Survey Data – Software Professionals
 - Rows: 11398
 - Columns: 85
 - Number of Jobs Opening:
 - Rows: 27005
 - Columns: 9

Data Cleaning

- Data Shape
 - # Columns
 - # Rows
- Data types
 - Objects
 - Integers
 - Floats
 - Datetime
- Duplicates
 - Drop duplicates
- Missing Values
 - Impute Missing Values
- Normalizing Data

Data Analysis

- Descriptive Statistics
 - Mean
 - Mode
 - Median
 - IQR
 - Outliers
- Data Visualization
 - Univariate Analysis
 - Distributions
 - Box Plots
 - Bi-Variate Analysis
 - Scatter Plots
 - Bubble Plots
 - Pie Charts, lines, and Bar Plots

Tools Used



RESULTS

■ Survey Data

- Shape (11398, 85)
- Three Datatypes:
- 79 Objects, 5 Float64, 1 int64

```
In [86]: 1 # data shape
2 df.shape

Out[86]: (11398, 85)

In [90]: 1 # datatypes
2 df.dtypes.value_counts()

Out[90]: object    79
float64    5
int64      1
dtype: int64
```

```
In [87]: 1 # view first 3 rows
2 df.head(3)
```

```
Out[87]:
```

	Respondent	MainBranch	Hobbyist	OpenSource	OpenSource	Employment	Country	Student	EdLevel	UndergradMajor	...	WelcomeChange	SOI
0	4	I am a developer by profession	No	Never	The quality of OSS and closed source software ...	Employed full-time	United States	No	Bachelor's degree (BA, BS, B.Eng., etc.)	Computer science, computer engineering, or sof...	...	Just as welcome now as I felt last year	wri devel
1	9	I am a developer by profession	Yes	Once a month or more often	The quality of OSS and closed source software ...	Employed full-time	New Zealand	No	Some college/university study without earning ...	Computer science, computer engineering, or sof...	...	Just as welcome now as I felt last year	
2	13	I am a developer by profession	Yes	Less than once a month but more than once per ...	OSS is, on average, of HIGHER quality than pro...	Employed full-time	United States	No	Master's degree (MA, MS, M.Eng., MBA, etc.)	Computer science, computer engineering, or sof...	...	Somewhat more welcome now than last year	wri devel

3 rows x 85 columns

■ Jobs Opening Data

- Shape (27005, 9)
- Two Datatypes:
 - 8 Object, 1 int64

```
In [35]: 1 # jobs shape
2 Jobs_df.shape
```

```
Out[35]: (27005, 9)
```

```
In [37]: 1 Jobs_df.dtypes.value_counts()
```

```
Out[37]: object    8
int64    1
dtype: int64
```

```
In [38]: 1 Jobs_df.head(3)
```

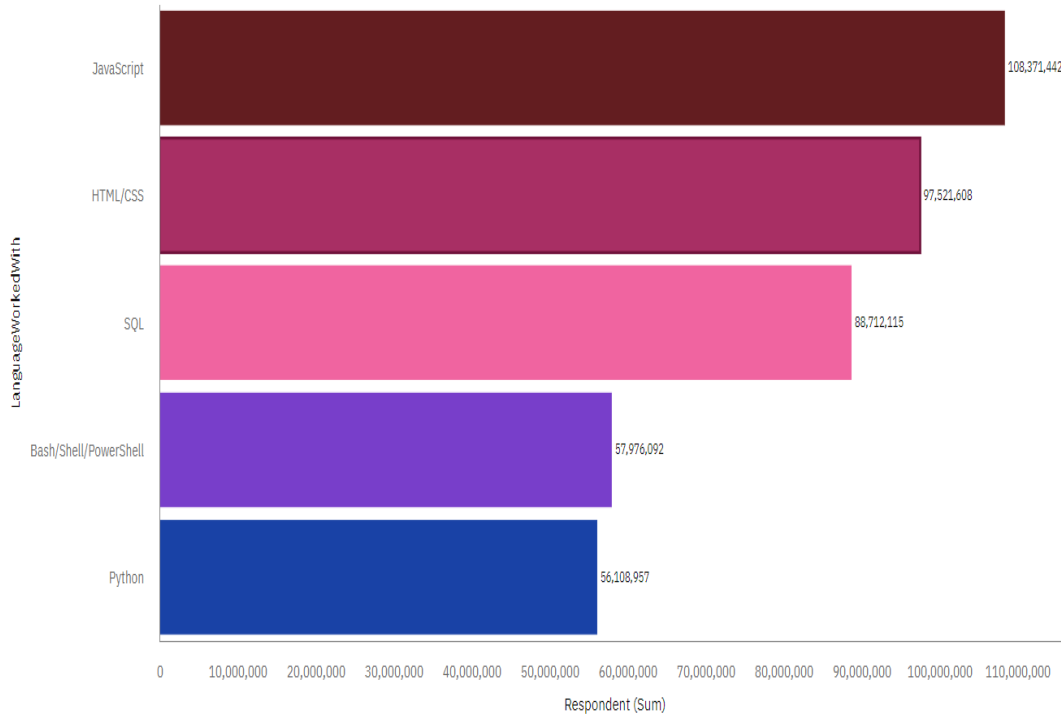
```
Out[38]:
```

	Id	Job Title	Job Experience Required	Key Skills	Role Category	Location	Functional Area	Industry	Role
0	0	Digital Media Planner	5 - 10 yrs	Media Planning Digital Media	Advertising	Los Angeles	Marketing , Advertising , MR , PR , Media Plan...	Advertising, PR, MR, Event Management	Media Planning Executive/Manager
1	1	Online Bidding Executive	2 - 5 yrs	pre sales closing software knowledge client...	Retail Sales	New York	Sales , Retail , Business Development	IT-Software, Software Services	Sales Executive/Officer
2	2	Trainee Research/ Research Executive- Hi- Tech...	0 - 1 yrs	Computer science Fabrication Quality check ...	R&D	San Francisco	Engineering Design , R&D	Recruitment, Staffing	R&D Executive

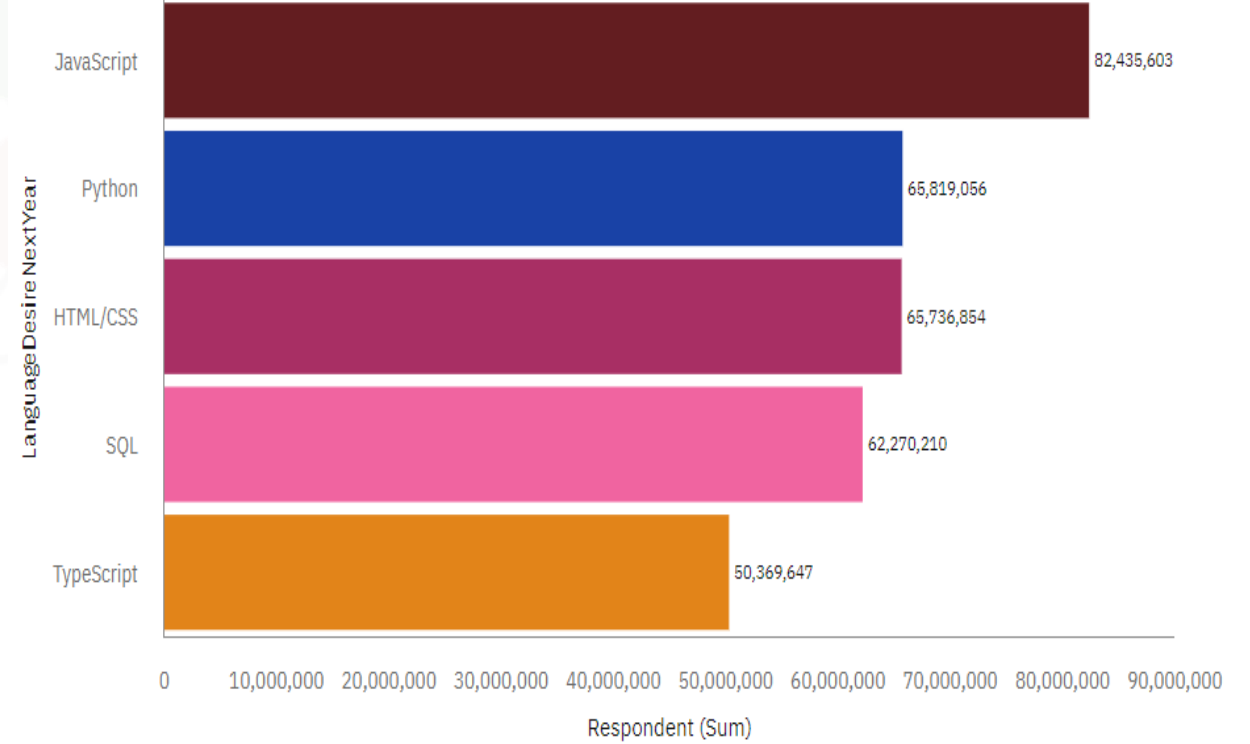
```
In [12]: 1 api_url = "http://127.0.0.1:5000/data"
2 def get_number_of_jobs_I(technology):
3     #your code goes here
```

PROGRAMMING LANGUAGE TRENDS

Top 5 Languages for Current Year



Top 5 Languages for Next Year



PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

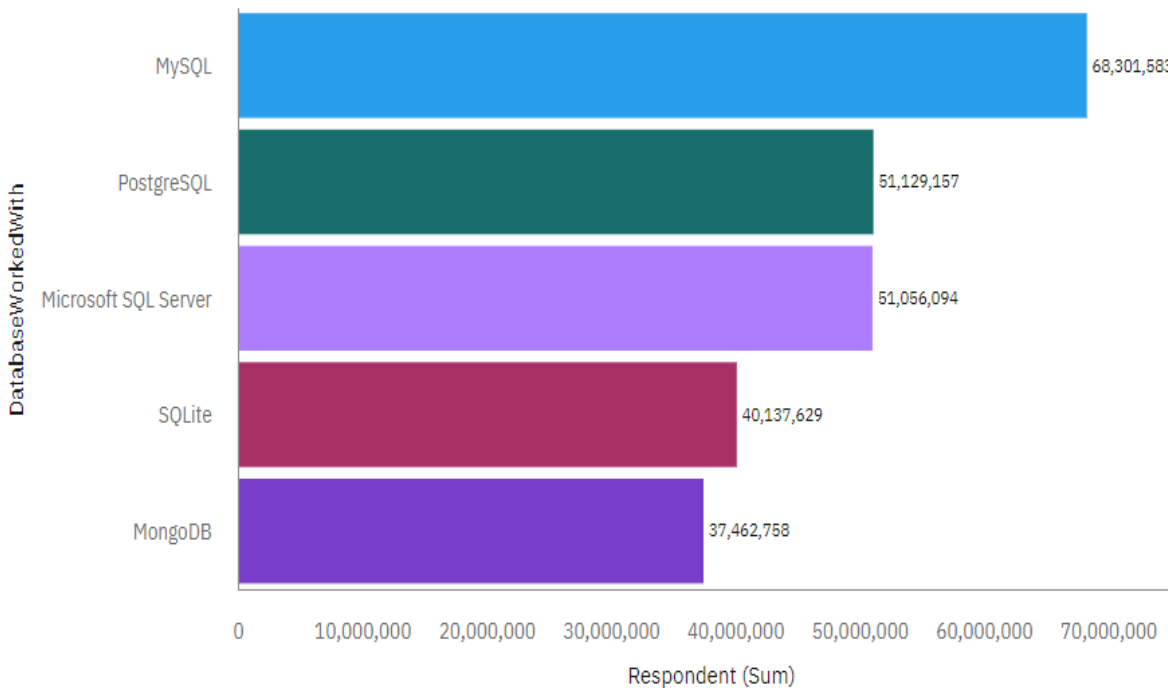
- Current Year Top 5 most popular Languages are:
 - JavaScript, HTML/CSS, SQL, Bash/Powershell, and Python
- JavaScript is the most popular Language, reported at about 108 Million Jobs.
- JavaScript is expected to still lead in the following year

Implications

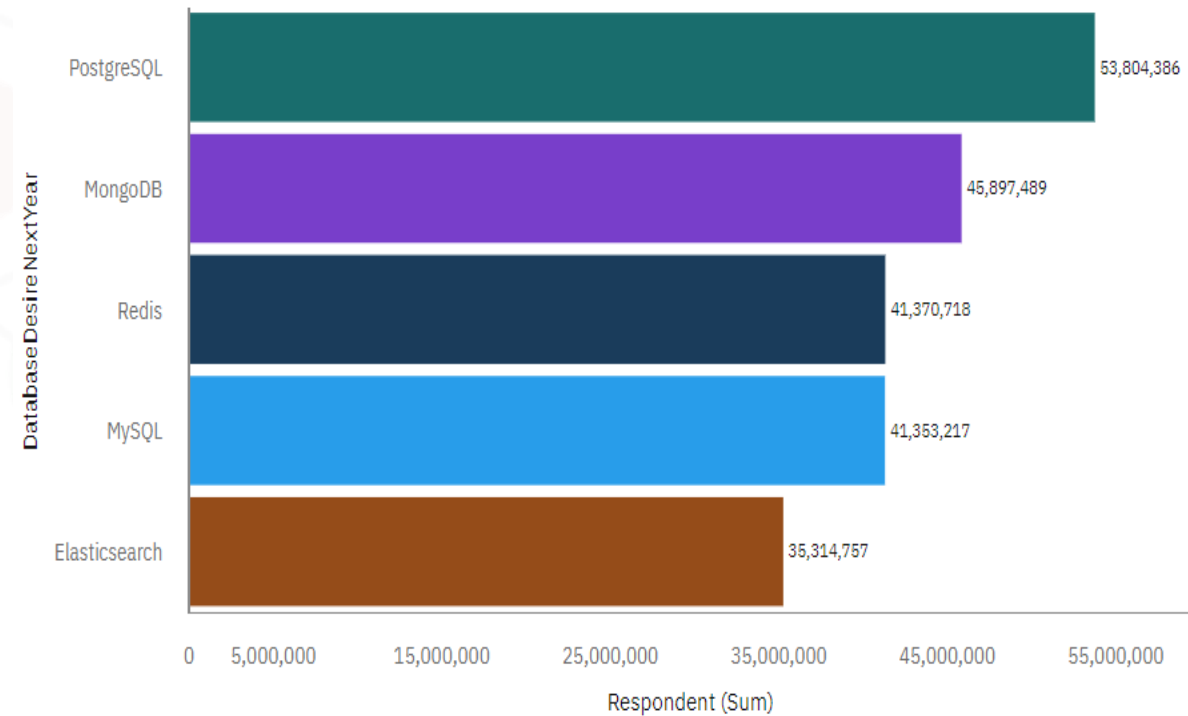
- Expected JavaScript popularity to drop by ~24% in the following year
- Python to Rank 2nd, a growth popularity by 17%
- TypeScript Popularity to rise and rank 5th, while Bash/Powershell popularity will drop out of the Top 5.

DATABASE TRENDS

Top 5 Database skills for Current Year



Top 5 Database Skills for Next Year



DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- Current Year Top 5 most popular Databases are:
 - MySQL, PostgreSQL, MS SQL Server, SQLite, and Mango DB
- MySQL is the most popular Database.
- PostgreSQL is expected to become the most popular Database in the following year

Implications

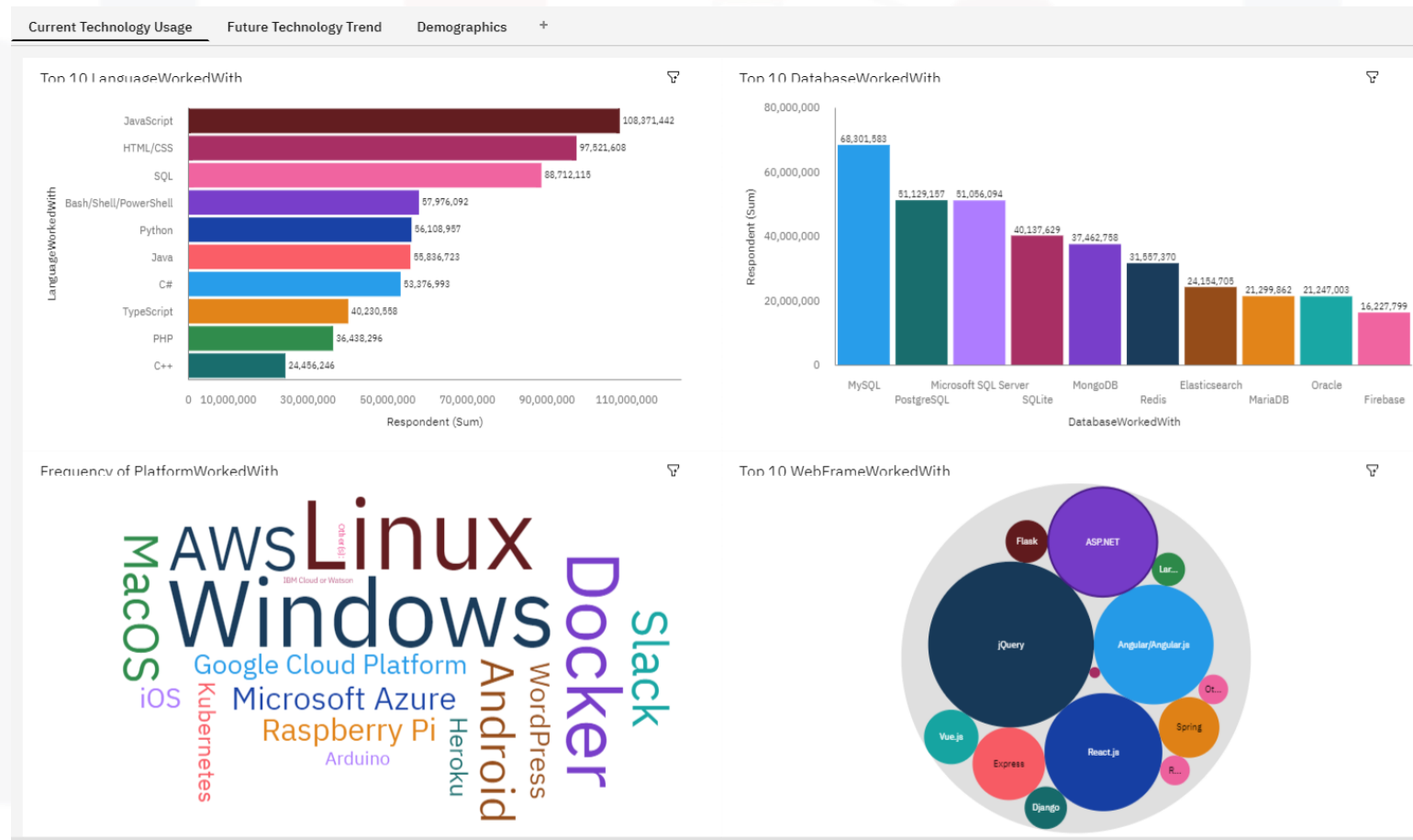
- Expected MySQL popularity to drop by ~40% in the following year
- **Elasticsearch** will be the **fastest growing Database**, with a growth rate of ~46% !
- Redis will be the 2nd fastest growing Databases, with a growth rate of 31%.

DASHBOARD

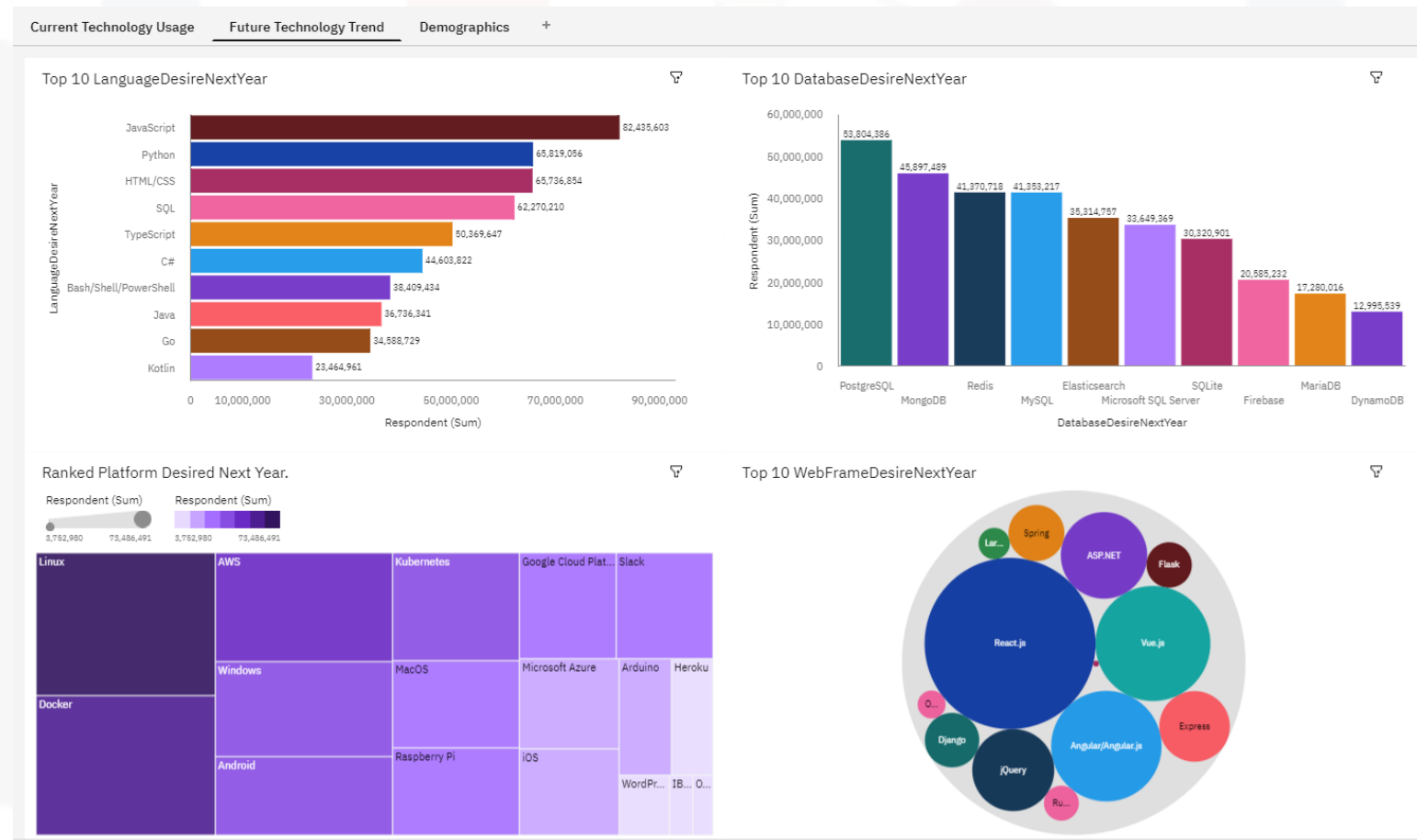


https://us1.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=.my_folders%2FBuilding%2Ba%2Bdashboard%2Bwith%2BCognos%2BDashboard%2BEmbedded%2B%2528CDE%2529&action=view&mode=dashboard&subView=model00000182bb1cd26d_00000000

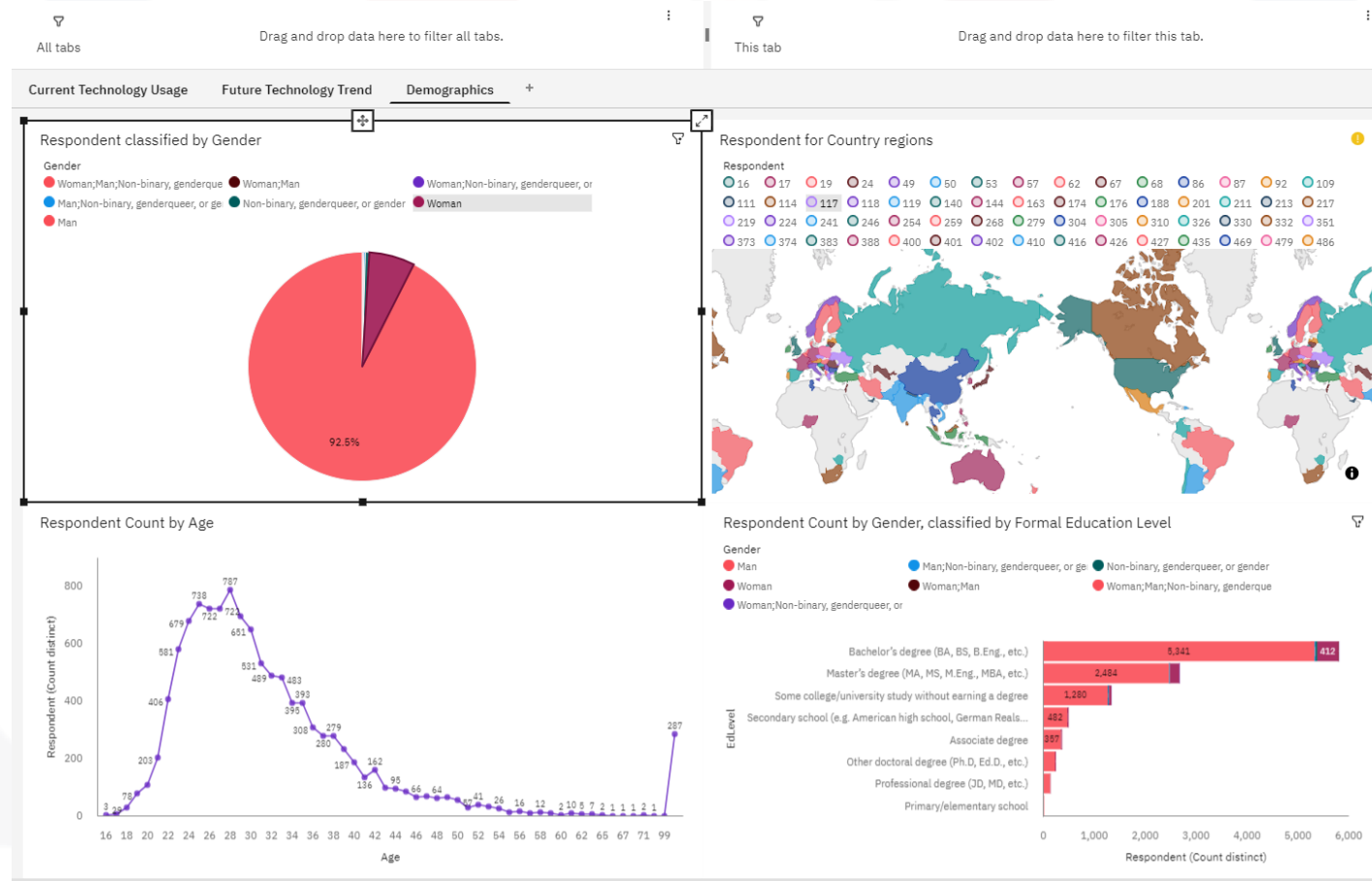
DASHBOARD TAB 1



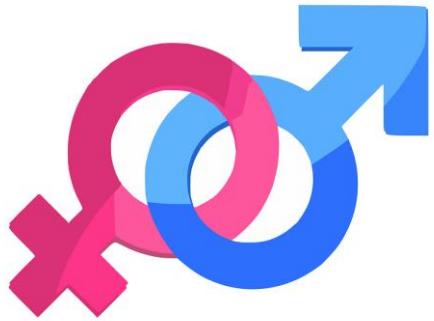
DASHBOARD TAB 2



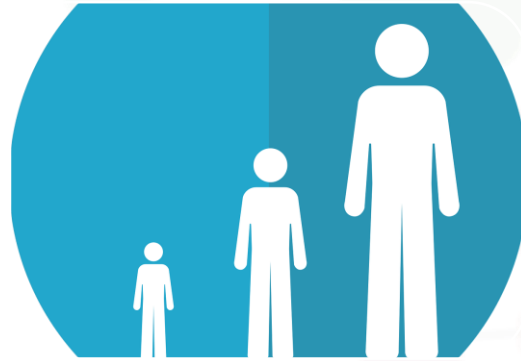
DASHBOARD TAB 3



DISCUSSION



92.5% Men
6.5% Women



Most Age Group
20-35

- Most Used Platform: Linux
- Most Used Webframe: jQuery
- Next most Desired Platforms:
 - Linux
 - Docker
- Next most desired Webframe: React.js

OVERALL FINDINGS & IMPLICATIONS



Most Popular
Language



Fastest Growing
Language



Most Popular
Database



Fastest Growing
Database

CONCLUSION



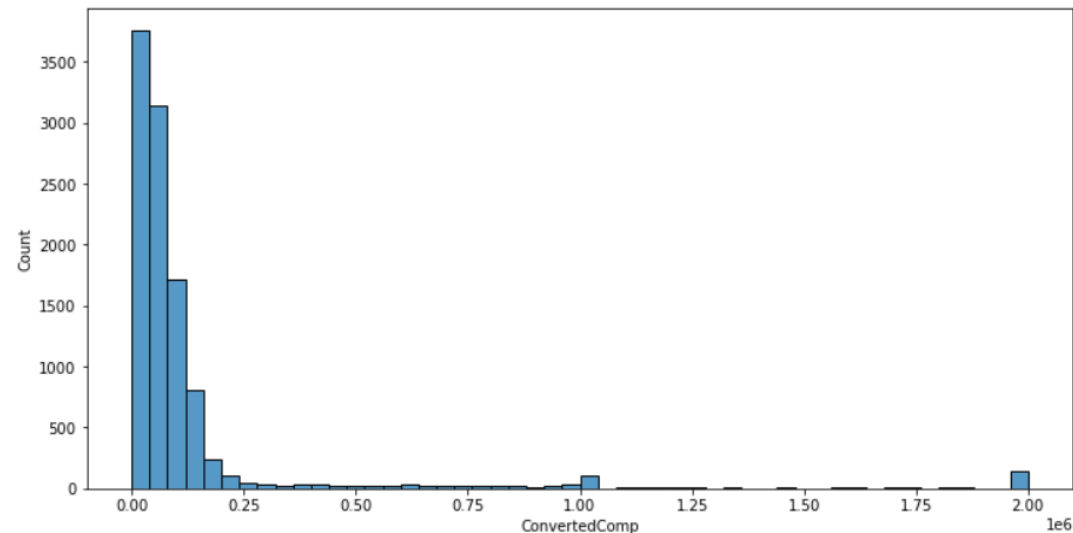
- JavaScript is the currently leading language, and it will at least remain so for the next following year.
- Python is the fastest growing language, and it has a potential to surpass JavaScript in the future.
- MySQL database is currently leading, however, its usage is expected to decline significantly by the following year
- Elasticsearch is expected to gain momentum to up its rank in the Database Market. Nevertheless, Postgres SQL will become the leading Database in the following year.

APPENDIX

- Survey Data:
 - !wget https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DA0321EN-SkillsNetwork/LargeData/m4_survey_data.sqlite

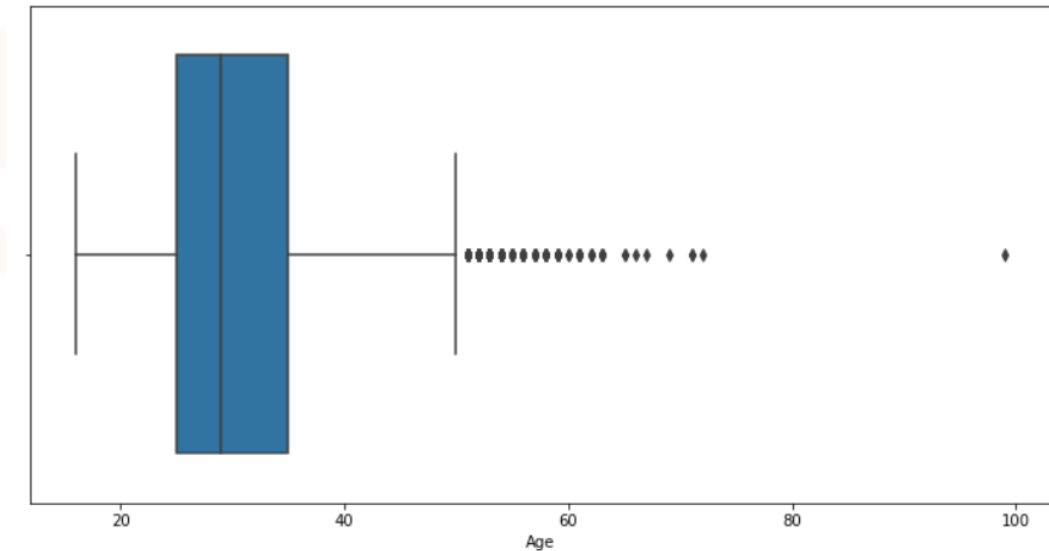
```
plt.figure(figsize=(12,6))
sns.histplot(data=df, x='ConvertedComp', bins=50)
```

<AxesSubplot: xlabel='ConvertedComp', ylabel='Count'>



```
plt.figure(figsize=(12,6))
sns.boxplot(data=df, x='Age')
```

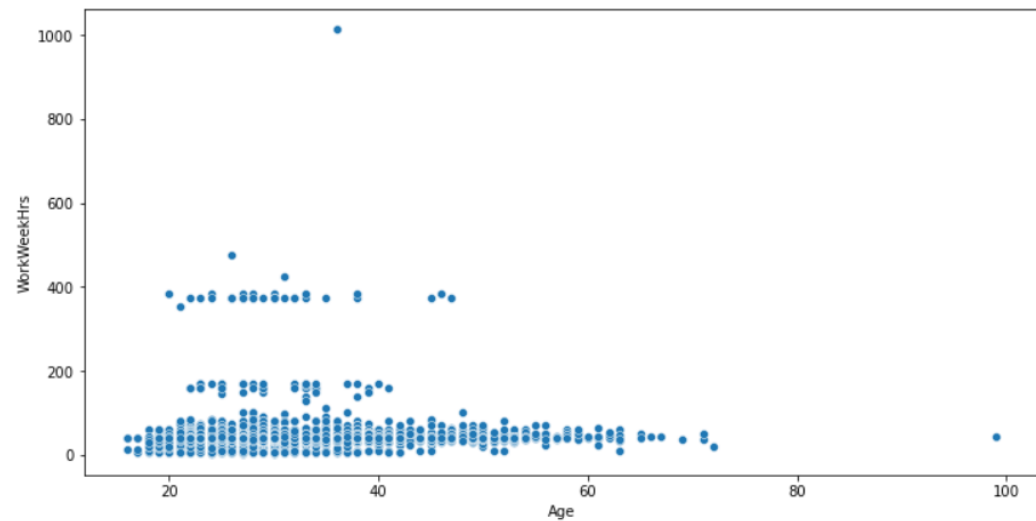
<AxesSubplot: xlabel='Age'>



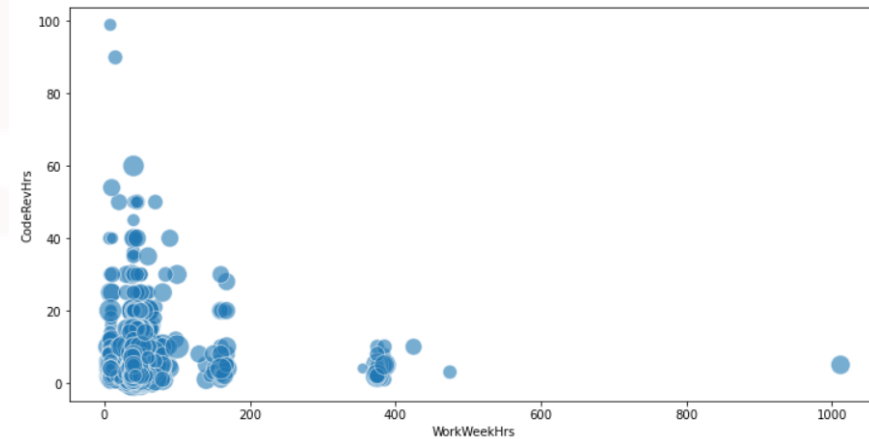
APPENDIX cont1

```
plt.figure(figsize=(12,6))
sns.scatterplot(data=df, x='Age', y='WorkWeekHrs')
```

<AxesSubplot:xlabel='Age', ylabel='WorkWeekHrs'>

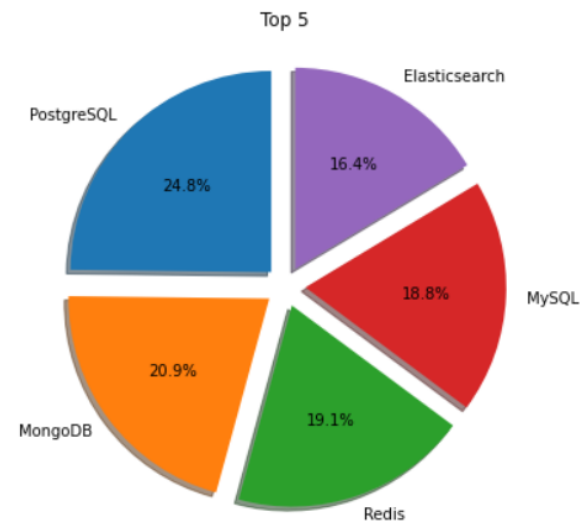


```
plt.figure(figsize=(12,6))
sns.scatterplot(data=df, x='WorkWeekHrs', y='CodeRevHrs', size='Age', alpha=0.6, legend=False, sizes=(20,1000))
plt.show()
```



APPENDIX cont2

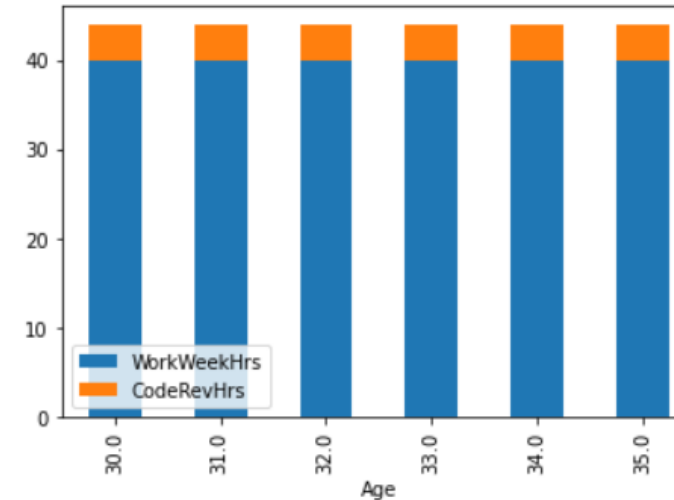
```
plt.figure(figsize=(12,6))
df.set_index('DatabaseDesireNextYear', inplace=True)
lab = df.index_
explode_list = [0.1, 0.1, 0.1, 0.1, 0.1]
sizes = df.iloc[:,0]
plt.pie(sizes, labels=lab, startangle=90, shadow=True, autopct='%1.1f%%', explode=explode_list)
plt.title('Top 5')
plt.show()
```



```
plt.figure(figsize=(12,6))
df.plot(kind='bar', stacked=True)
```

<AxesSubplot:xlabel='Age'>

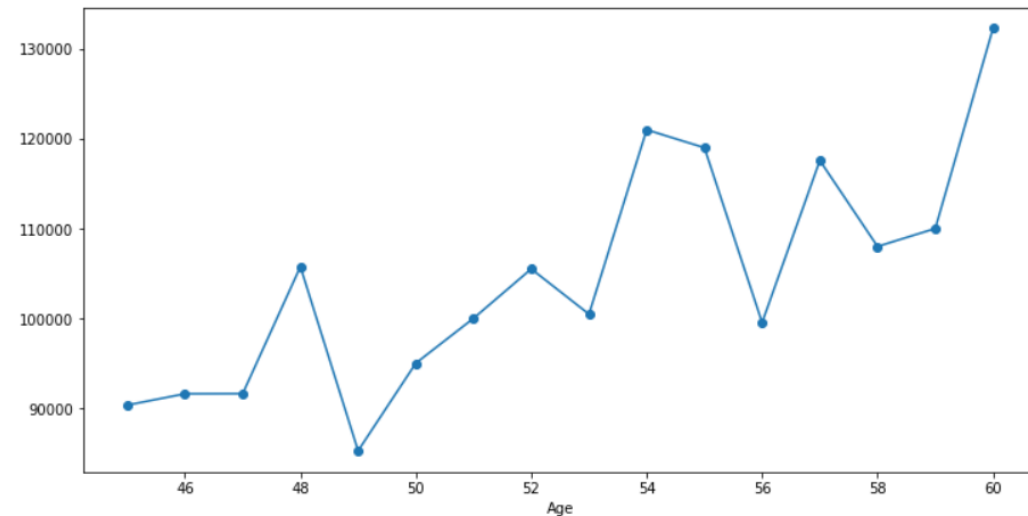
<Figure size 864x432 with 0 Axes>



APPENDIX cont3

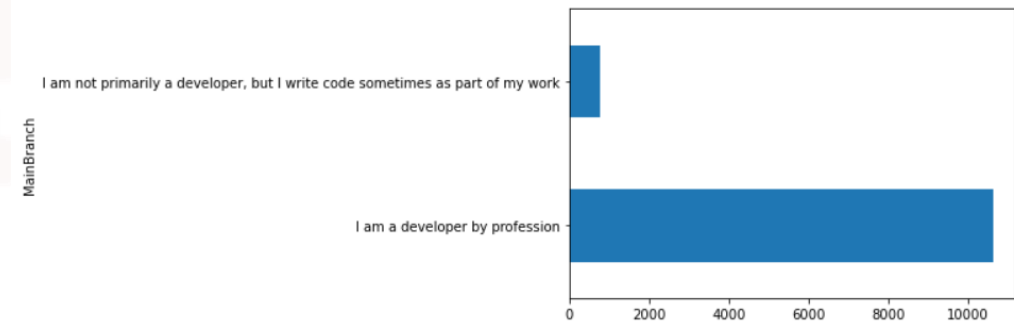
```
plt.figure(figsize=(12,6))  
df.plot(linestyle="--", marker='o')
```

<AxesSubplot: xlabel='Age'>

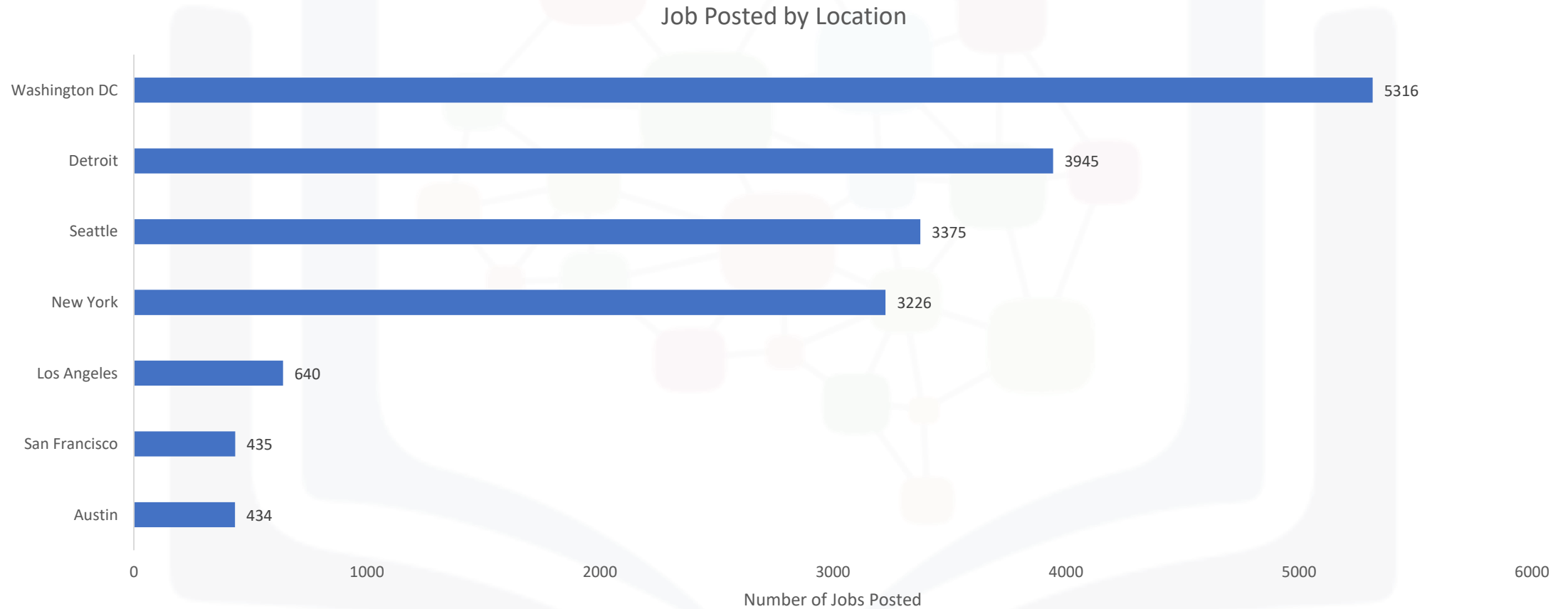


```
df.plot.barh()
```

<AxesSubplot: ylabel='MainBranch'>



JOB POSTINGS



POPULAR LANGUAGES

