

IBM DATA ANALYST CAPSTONE PROJECT



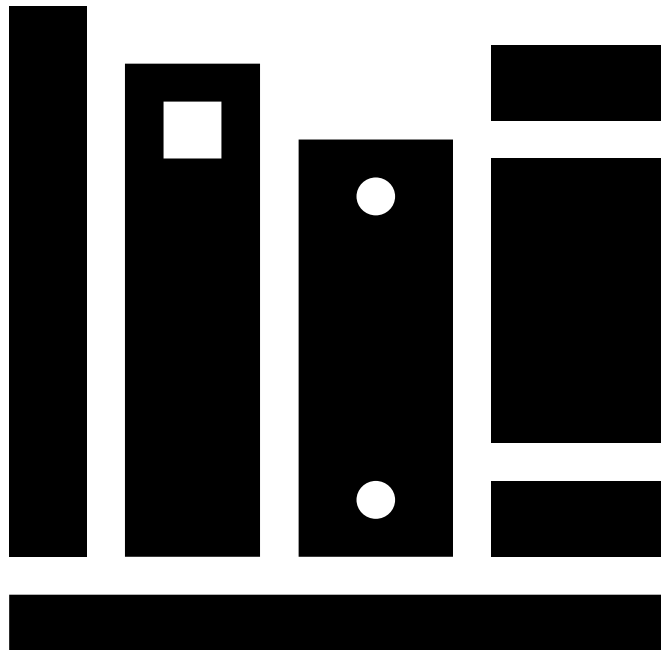
INSIGHTS AND TRENDS IN DATA ANALYSIS

2019 AND NEXT YEAR

Beatriz Duarte

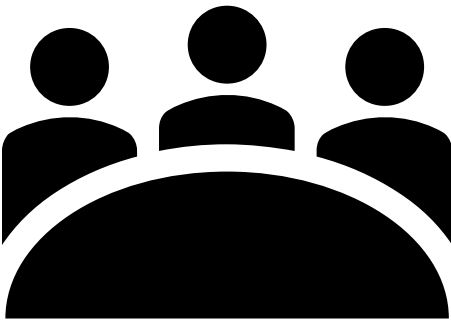
2024-10-06

OUTLINE



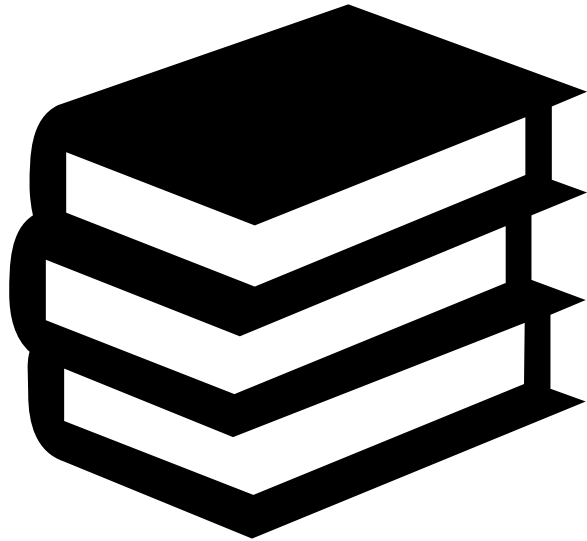
- Executive Summary
- Introduction
- Methodology
- Results
 - Visualization – Charts
 - Dashboard
- Discussion
 - Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY



- The present and the future demands adaptability, innovative thinking and a proactive approach to learning, ensure that professionals remain competitive and equipped to meet emerging challenges and opportunities in technology
- In 2019 a stack overflow survey was carried out to provide valuable insights into the current landscape of technology usage, revealing key trends in programming languages, databases, frameworks and platforms
- The data was subjected to exploratory analysis using Python and IBM Cognos Analytics and it can be visualized in charts, graphs and dashboards
- The results showed that JavaScript remains the most widely used programming language, dominating web development, while Python is rapidly growing in popularity. In the realm of databases, MySQL continues to lead, but PostgreSQL is increasingly favored for its advanced capabilities
- The most dominant framework is jQuery and platforms are Linux and Windows
- Overall, according to gender the respondents are mainly males and according to countries are from United States, ranging between 24-30 years old and have a bachelor's degree
- Adapting to change through continuous education and networking is crucial, as is staying informed about industry trends and innovations to always get the best results

INTRODUCTION



OVERVIEW OF THE 2019 STACK OVERFLOW DEVELOPER SURVEY AND PREDICTION FOR THE NEXT YEAR

- Top 10 Trends of the Current Year and Next Year:
 - [Programming languages](#)
 - [Databases](#)
 - [Web Frames](#)
- Current platforms used in data analysis and the desired platforms for the next year
- Percentage of respondents according gender
- Count of the respondents by Country and count by Age
- Count of respondents by Gender and count by Formal Education Level

METHODOLOGY

Data collection Period:

- The survey was open from January 23 to February 14, 2019

Response Rate and Sample Size:

- Over 90,000 responses from developers around the world (The large sample size gives significant insights across a variety of demographics, job roles, and geographies)

Survey Distribution:

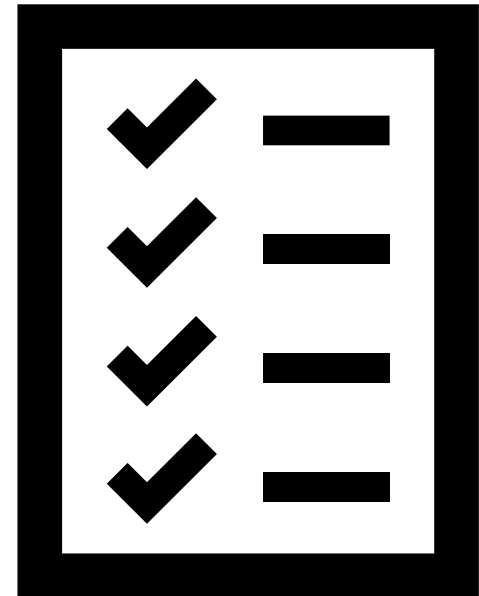
- Multiple channels, including direct emails to users of Stack Overflow and its related sites. It was also promoted through blogs and social media channels to reach a wider audience

Target Audience:

- The survey was open to anyone who identifies as a developer, working in software development or related fields. It also encouraged participation from people in various development roles (like software engineers, data scientists, system administrators, and more)

Data Analysis and Reporting:

- IBM Cognos Analytics and Python
- Charts, Graphs and Dashboards



RESULTS

Top Programming Languages (Current Year and Trends for Next Year):

- JavaScript (most popular), Python (rapidly growing)

Top Databases (Current Year and Trends for Next Year):

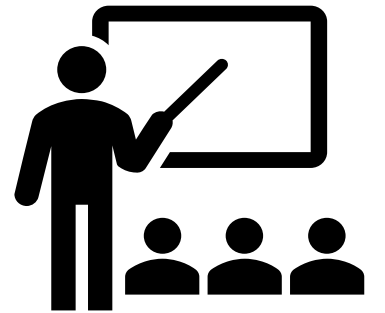
- MySQL was the most widely used database in 2019 but PostgreSQL is the most desired database for the next year

Top Web Frameworks (Current Year and Trends for Next Year):

- jQuery was on the top in 2019 but the most desired web framework for the next year is React.js, with high demand for developers using it

Current Platforms Used in Data Analysis and Desired Platforms for Next Year:

- Linux, Windows, Docker and AWS are the most current used platforms, meanwhile on the top of the most desired for the next year are still Linux and Docker



RESULTS - CONTINUED

Percentage of Respondents According to Gender:

- Only 6.5% of respondents identifying as women

Count of Respondents by Country:

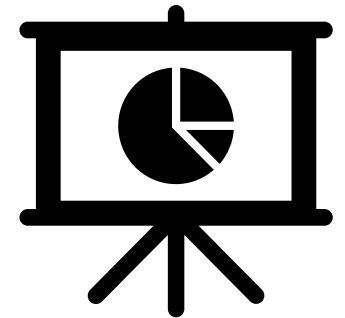
- The highest number of respondents came from the United States, followed by India, Germany and United Kingdom

Count of Respondents by Age:

- The age group with the most respondents falls within the 24-30 range

Count of Respondents by Gender, Classified by Formal Education Level

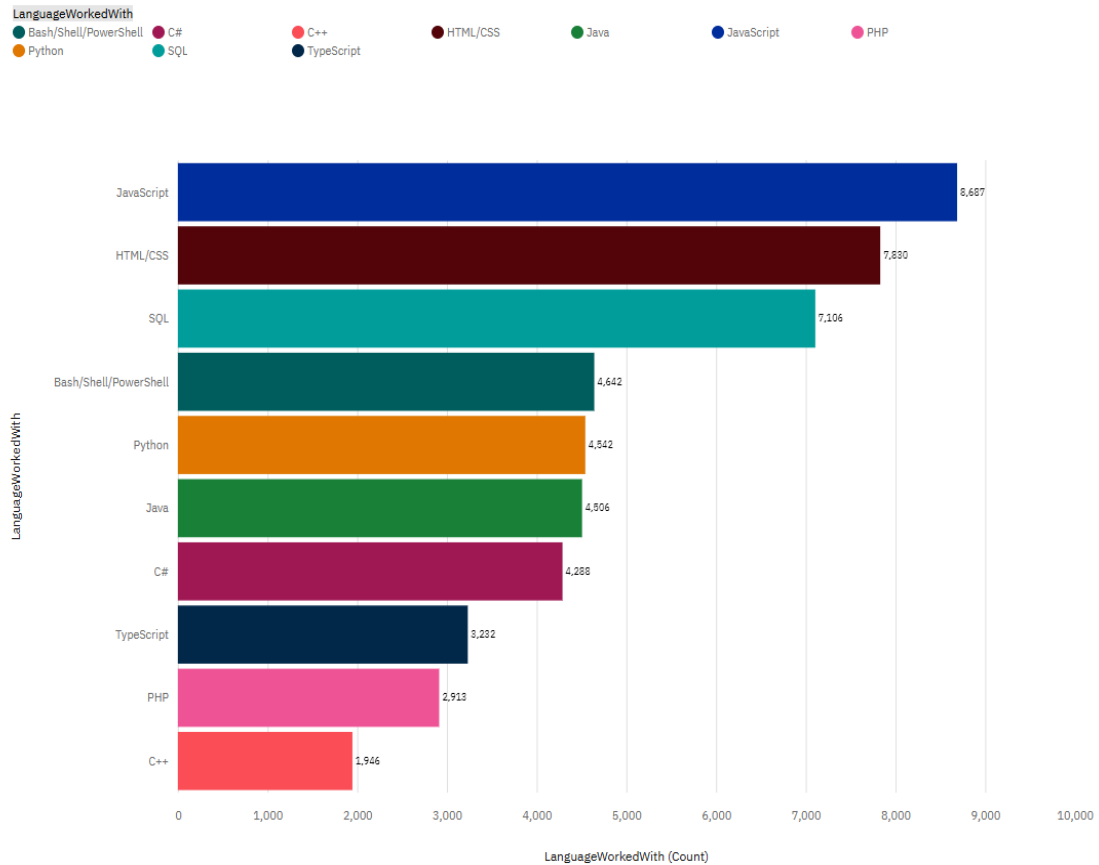
- A majority of male respondents reported having completed at least a Bachelor's degree and female respondents showed similar educational levels, though a smaller percentage in proportion



PROGRAMMING LANGUAGE TRENDS

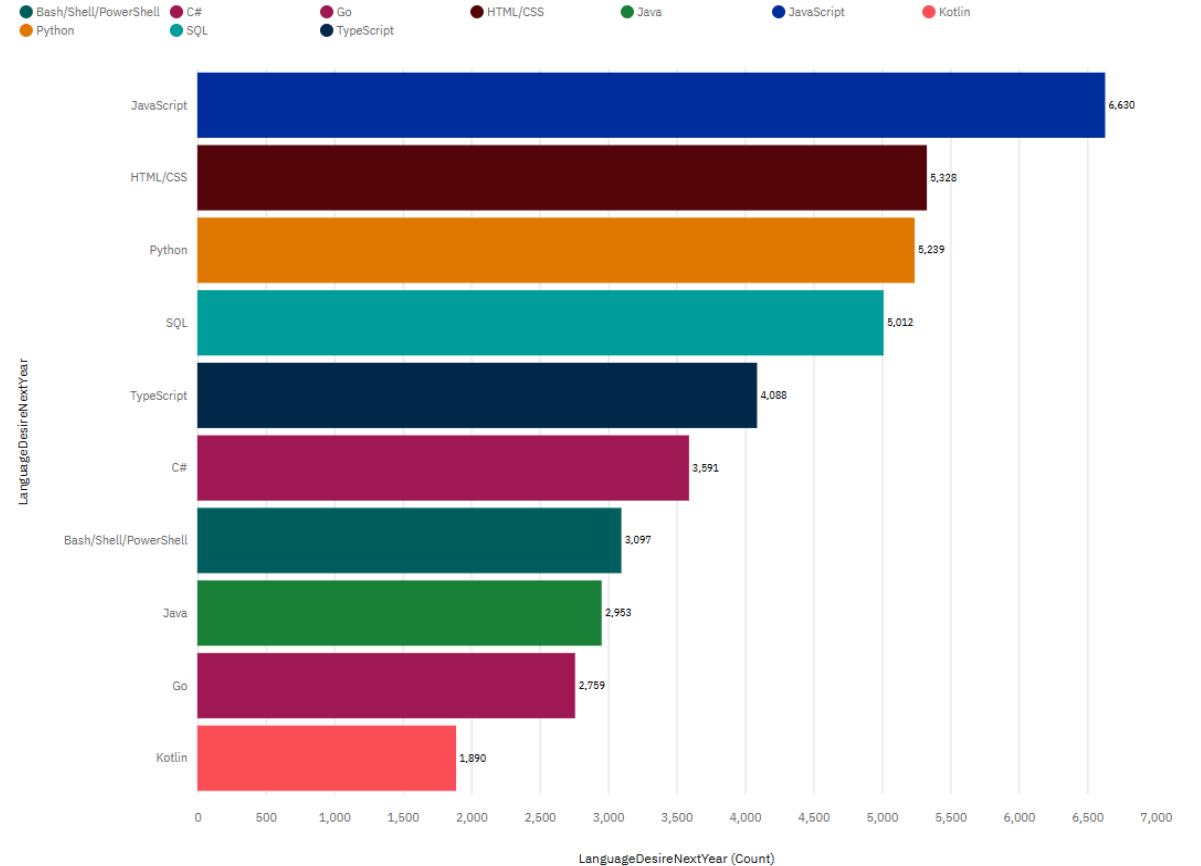
Current Year

Top 10 Languages Worked With



Next Year

Top 10 Languages Desired for the Next Year



PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

- The most currently used languages are JavaScript, HTML/CSS, SQL, Shell languages and Python
- The most used languages for the next year will be JavaScript, HTML/CSS, Python, SQL and Typescript
- Python will be more desired than SQL on the next year

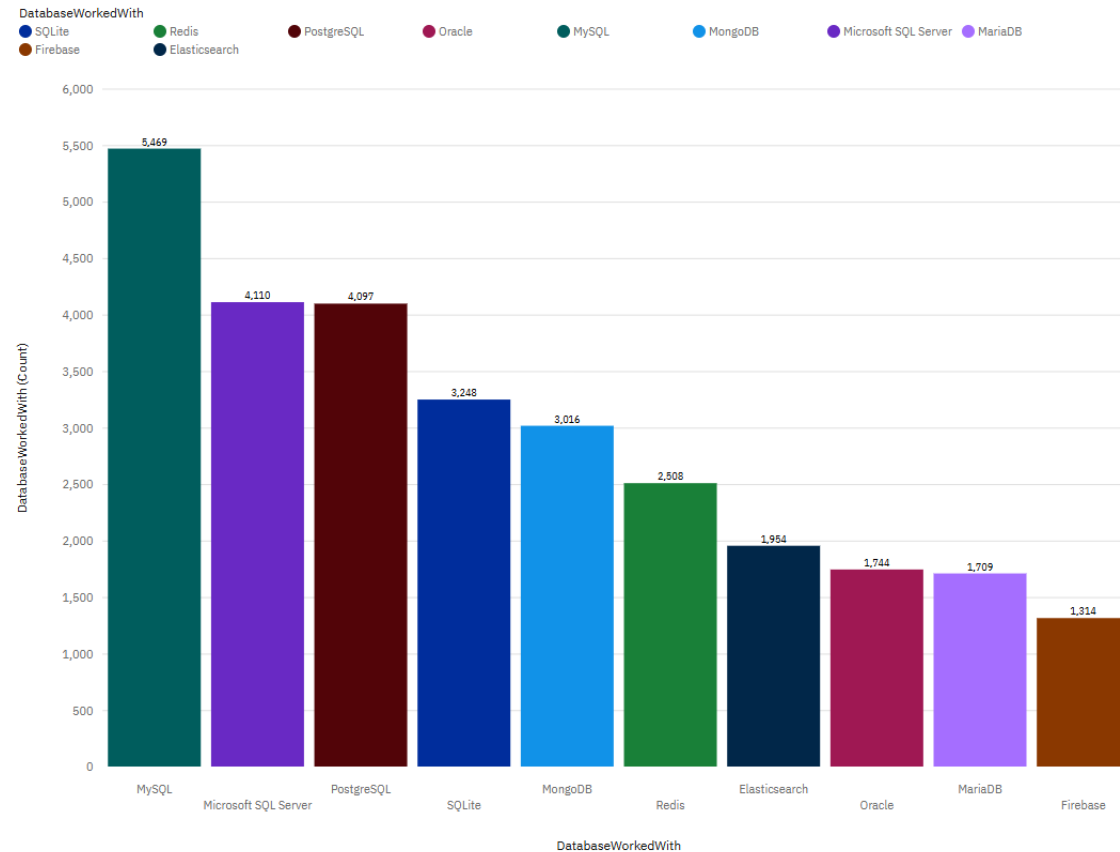
Implications

- JavaScript's and HTML/CSS's dominance indicates the continued growth of web technologies and full-stack development
- Python's growth suggests its importance in cutting-edge fields like AI, reinforcing the demand for developers with data-oriented skills
- SQL's large and active community of users ensures continued innovation, support and resources, making it easy to find tools, libraries, and frameworks that optimize database interactions

DATABASE TRENDS

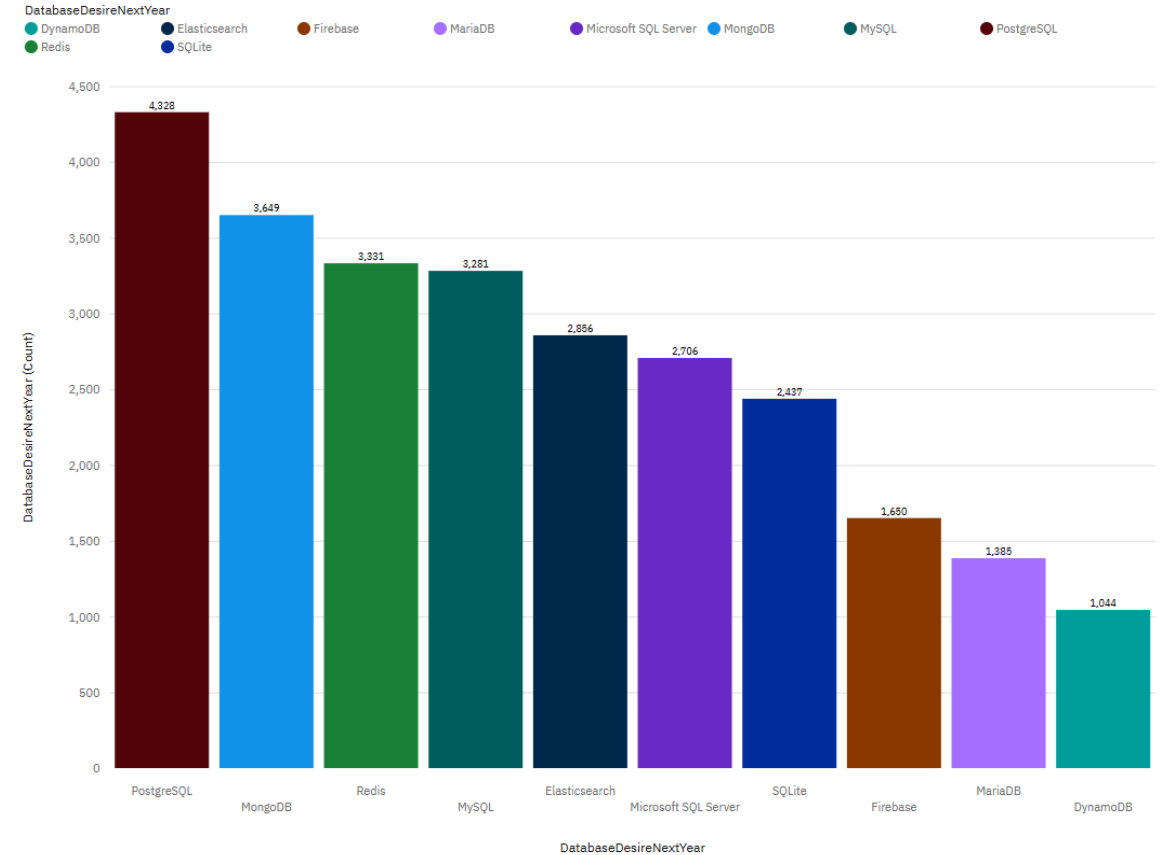
Current Year

Top 10 Databases Worked With



Next Year

Top 10 Databases Desired for the Next Year



DATABASE TRENDS - FINDINGS & IMPLICATIONS

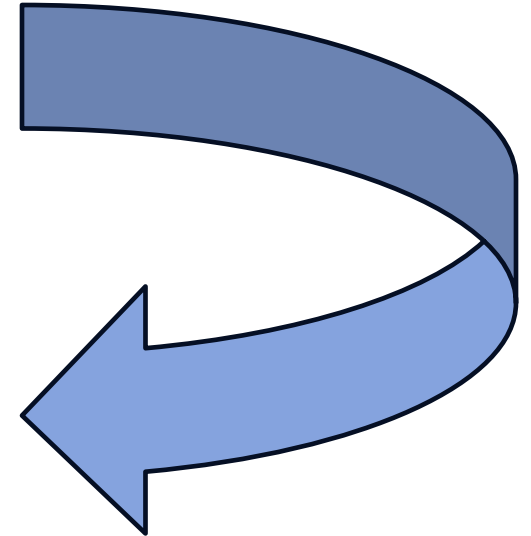
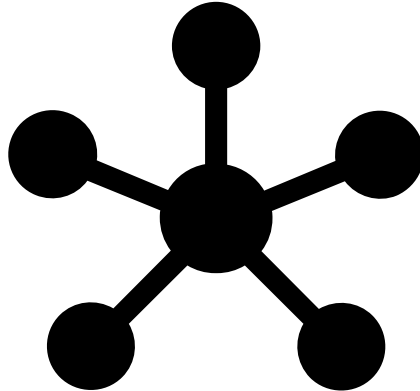
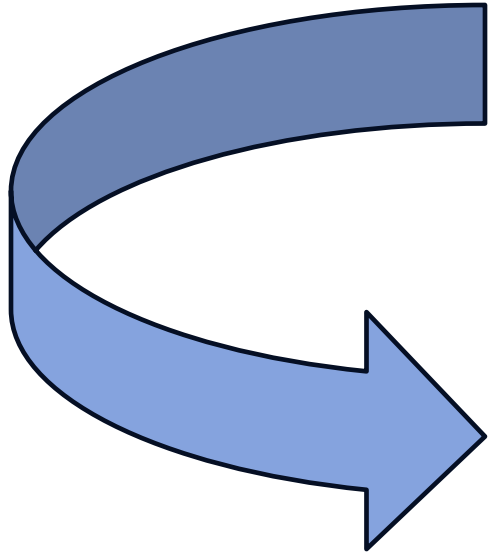
Findings

- MySQL was the most commonly used database technology, followed by Microsoft SQL Server, PostgreSQL and SQLite
- PostgreSQL saw continued growth in popularity, often praised for its advanced features, performance, and support for complex queries
- MongoDB and Redis were the most popular NoSQL databases

Implications

- MySQL and PostgreSQL's leading roles highlights the continued preference for open-source databases, particularly in web development, startups, and projects where cost-effectiveness and flexibility are important
- The rise of MongoDB and Redis implies that developers are increasingly building applications that handle unstructured data
- The growing interest in cloud-based databases suggests a shift towards managed database services that offer scalability, backup and high availability without requiring significant infrastructure management

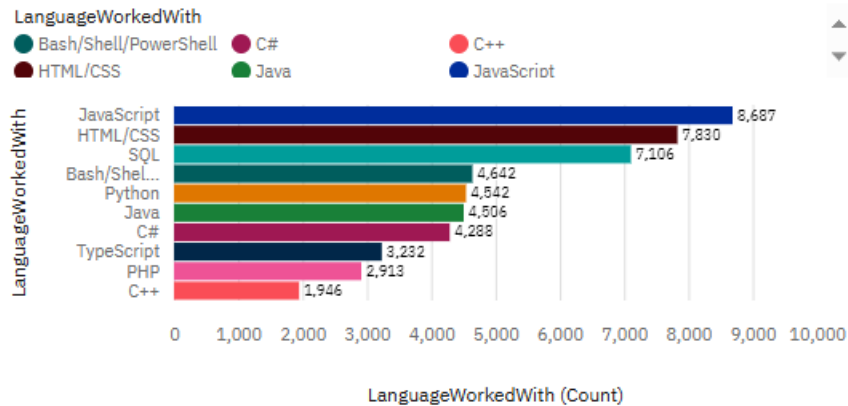
DASHBOARD - THE GITHUB LINK OF THE COGNOS DASHBOARD



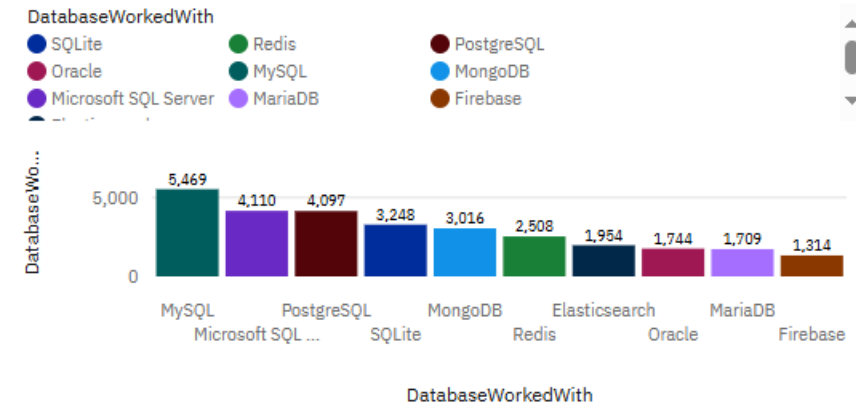
https://eu1.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=.my_folders%2FBuilding%2Ba%2Bdashboard%2Bwith%2BIBM%2BCognos%2BAnalytics&action=view&mode=dashboard&subView=model000001921f95ba25_00000003

DASHBOARD – CURRENT TECHNOLOGY USAGE

Top 10 Languages Worked With



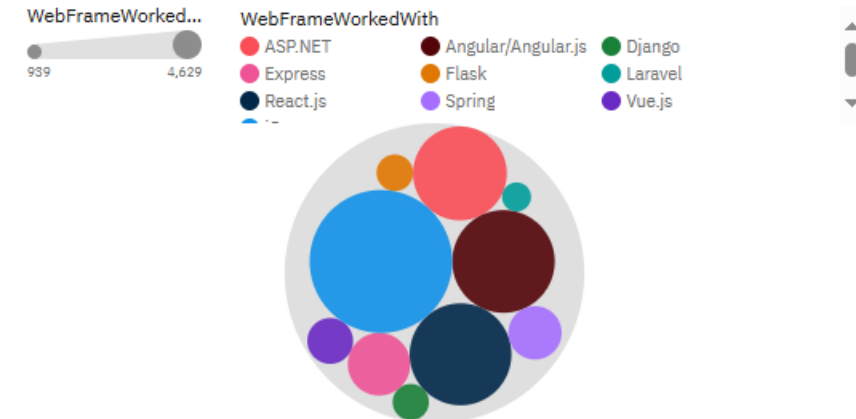
Top 10 Databases Worked With



Platforms Worked With

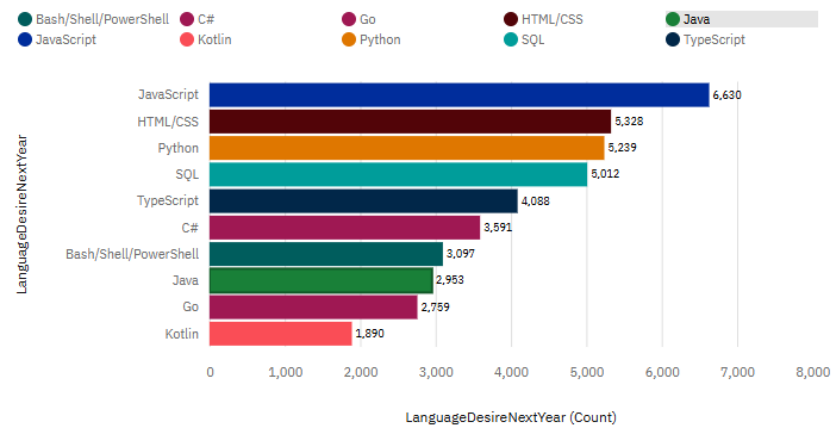


Top 10 Web Frame Worked With

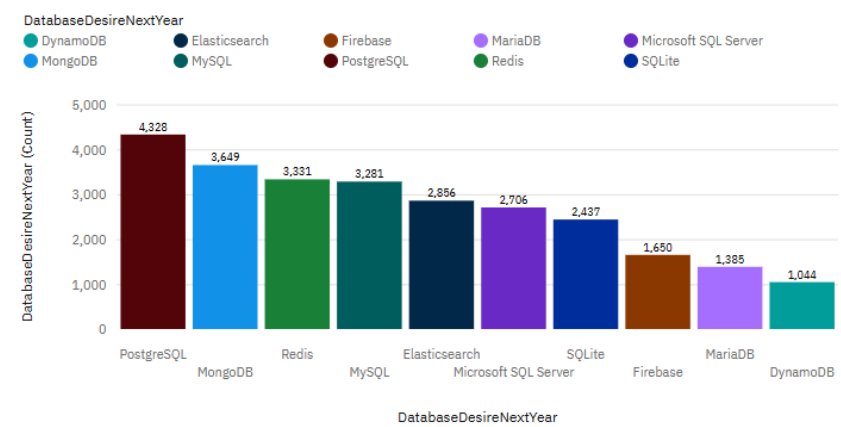


DASHBOARD – FUTURE TECHNOLOGY TREND

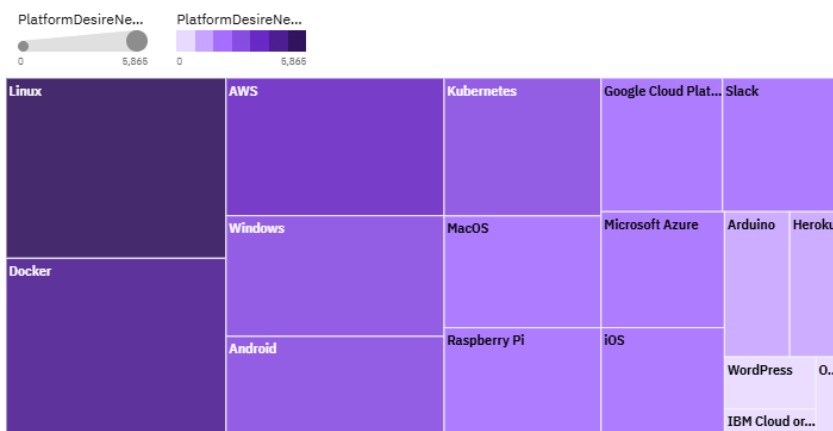
Top 10 Languages Desired for the Next Year



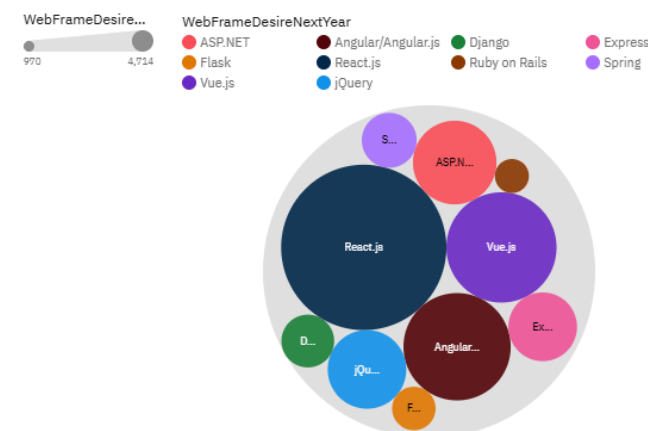
Top 10 Databases Desired for the Next Year



Desired Platform for the Next Year



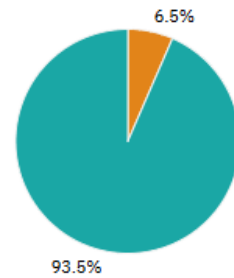
Top 10 Desired Web Frame for the Next Year



DASHBOARD - DEMOGRAPHICS

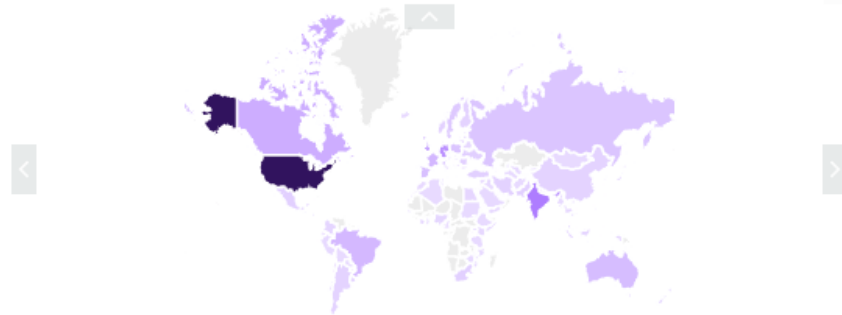
Respondent Classified by Gender

Gender
Woman Man

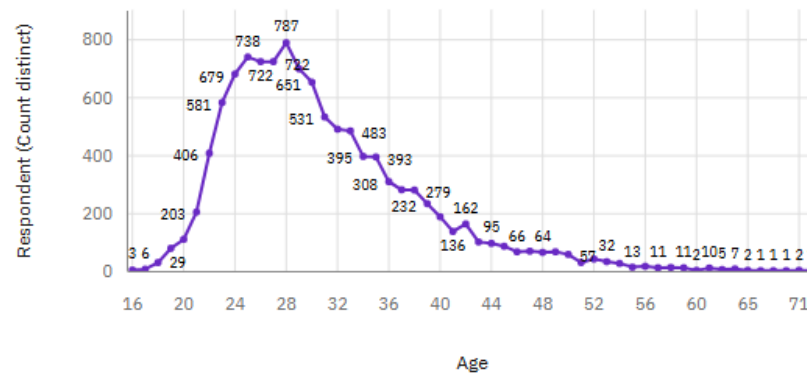


Respondent Count for Countries

Respondent (Count)
1 3,127

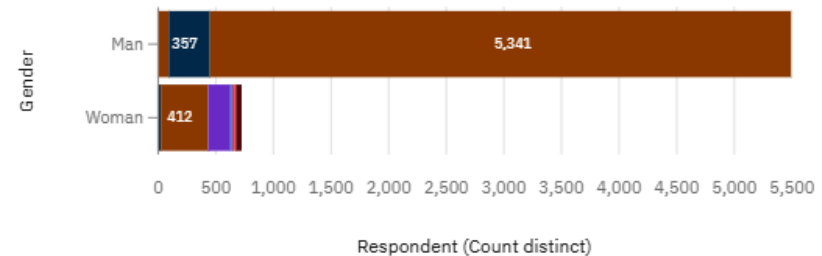


Respondent Count by Age

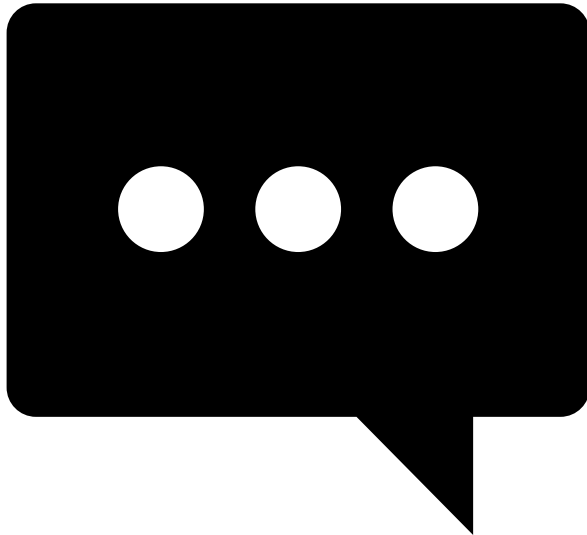


Respondent Count by Gender, Classified by Education Level

EdLevel
(no value) Associate degree
Bachelor's degree (BA, BS, B.Eng.... I never completed any formal edu...
Master's degree (MA, MS, M.Eng.,.... Other doctoral degree (Ph.D, Ed.D....

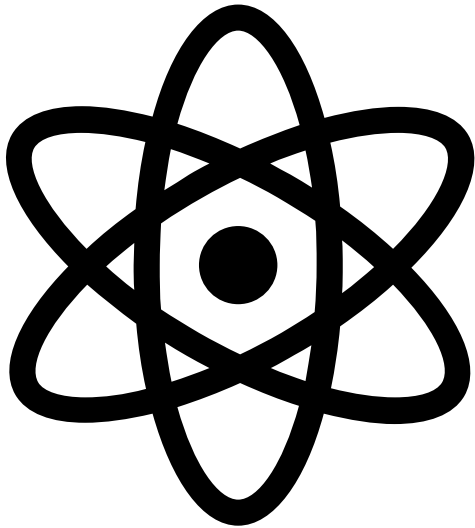


DISCUSSION



- Impact of technology trends: Rise of Python, DevOps culture, importance of open-source
- Developers emphasize the importance of cloud skills in modern software development and infrastructure
- A significant point of discussion is the lack of diversity in the tech industry, with a majority of respondents being male. There's growing awareness and efforts to encourage underrepresented groups to pursue careers in technology
- Overall, the survey sparks discussions about how developers should adapt to the evolving technology landscape, focusing on the tools, platforms, and languages that are shaping the future of software development

CONCLUSION



- Overall, the 2019 database trends point to a future where open-source, flexible and cloud-native databases are increasingly favored, reflecting the evolving needs of modern applications
- By focusing on these areas, we can position ourselves to remain competitive and adaptable in the evolving tech landscape

APPENDIX

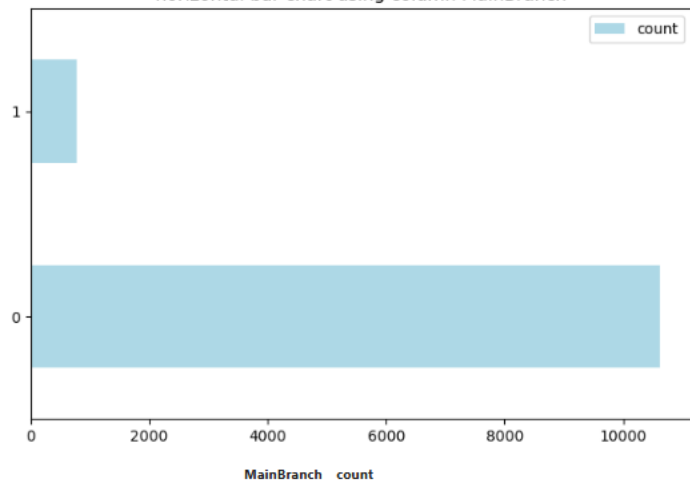
Bar Chart

Create a horizontal bar chart using column `MainBranch`.

```
# your code goes here
QUERY="""
SELECT MainBranch, COUNT(*) as count
FROM master
GROUP BY MainBranch
"""
df_bar= pd.read_sql_query(QUERY,conn)

df_bar.plot(kind='barh', figsize=(8,5), color='lightblue')
plt.title('horizontal bar chart using column MainBranch')
plt.show()
df_bar
```

horizontal bar chart using column MainBranch



0	I am a developer by profession	10618
1	I am not primarily a developer, but I write co...	780

Pie Charts

Create a pie chart of the top 5 databases that respondents wish to learn next year. Label the pie chart with database names. Display percentages of each database on the pie chart.

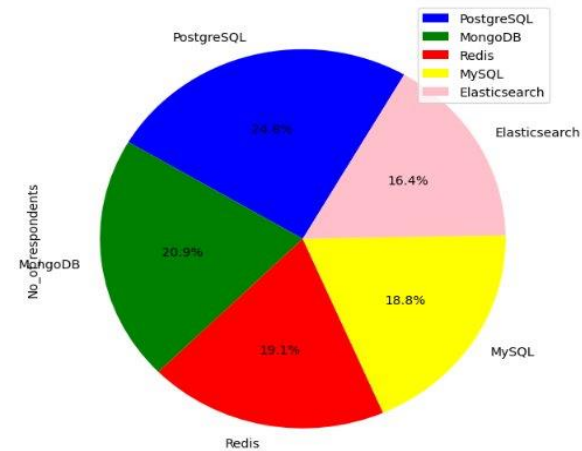
```
# your code goes here
Query5 = """
SELECT DatabaseDesireNextYear as Database, count(*) as No_of_respondents
FROM DatabaseDesireNextYear
GROUP BY DatabaseDesireNextYear
ORDER BY No_of_respondents DESC LIMIT 5"""

db_df = pd.read_sql_query(Query5, conn)
print(db_df)

db_df.plot(kind = 'pie', labels = db_df['Database'], y = 'No_of_respondents',
figsize = (7,10), #dimension of the chart window
autopct='%1.1f%%', #autopct displays percentages in pie chart. so .1f% rounds to 1 decimal place.
colors = ['blue', 'green', 'red', 'yellow', 'pink'], #colours of the pie chart items
startangle = 60) #start Angle implies that we can rotate the pie chart according to the degree angle we specify
plt.title('Top 5 databases that respondents wish to learn next year')
plt.show()
```

	Database	No_of_respondents
0	PostgreSQL	4328
1	MongoDB	3649
2	Redis	3331
3	MySQL	3281
4	Elasticsearch	2856

Top 5 databases that respondents wish to learn next year



APPENDIX - CONTINUED

Plot a histogram of the column `Age`.

```
# your code goes here
plt.figure(figsize=(10,5))
sns.distplot(a=df["Age"],bins=20,kde=False)
plt.show()
```

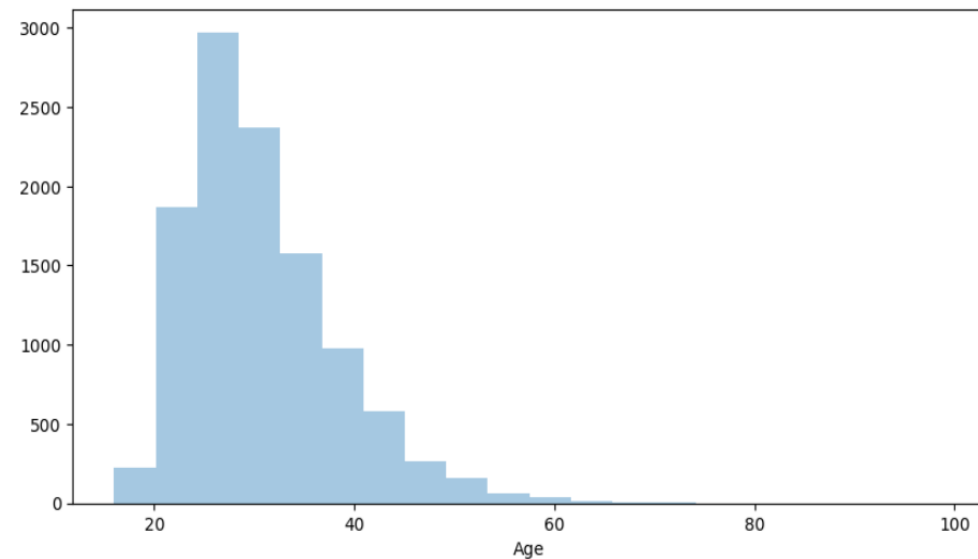
<ipython-input-14-1221b7f502a6>:3: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(a=df["Age"],bins=20,kde=False)
```



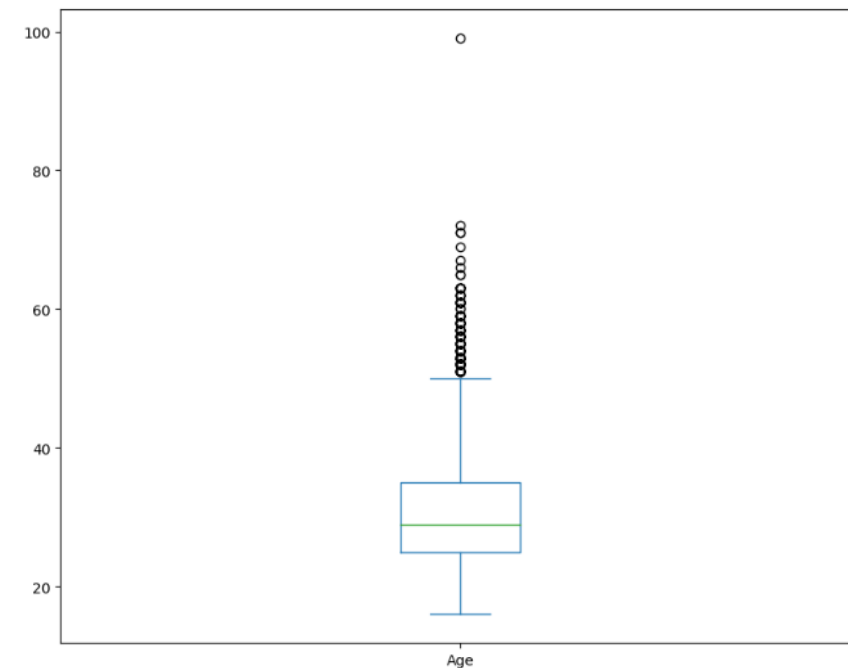
Box Plots

Plot a box plot of `Age`.

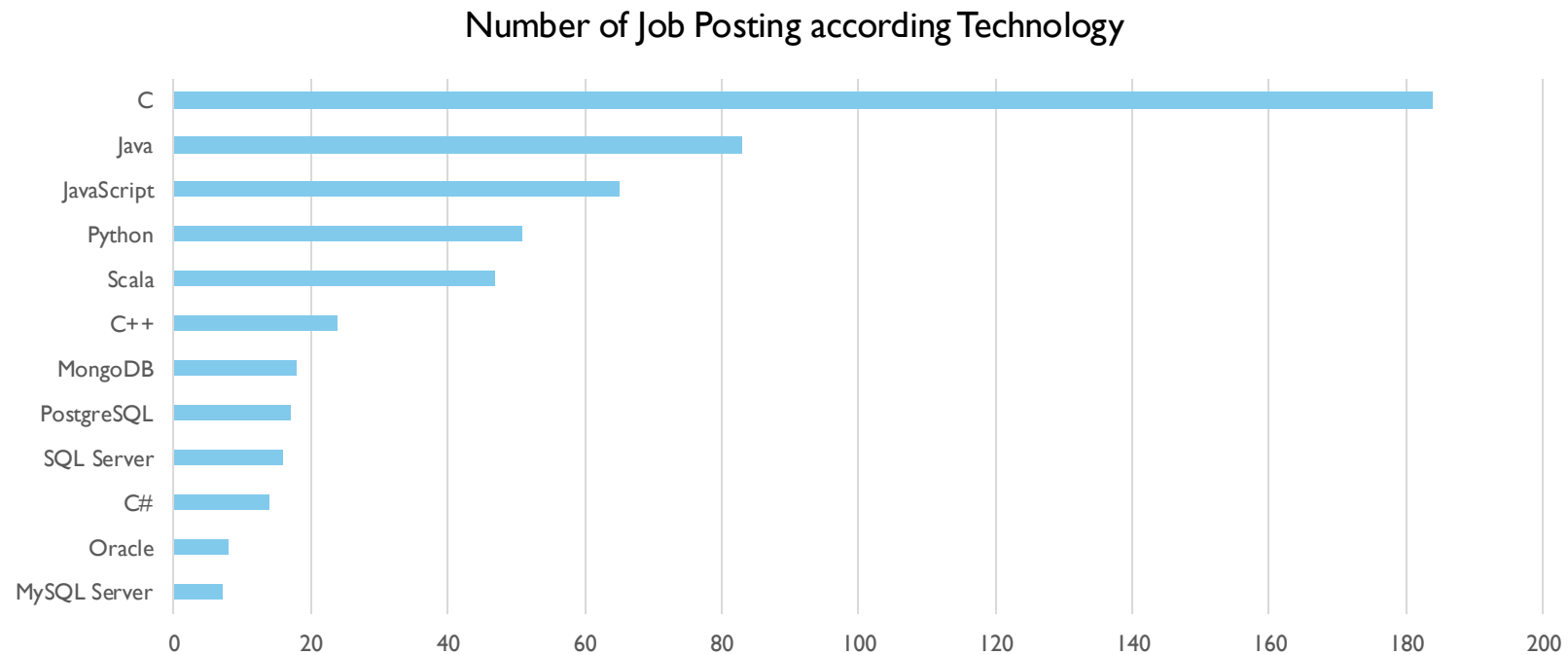
```
# your code goes here
Query2 = """
SELECT Age
FROM master
"""

Age = pd.read_sql_query(Query2, conn)
Age.plot(kind='box', figsize=(10, 8))
```

<AxesSubplot:>

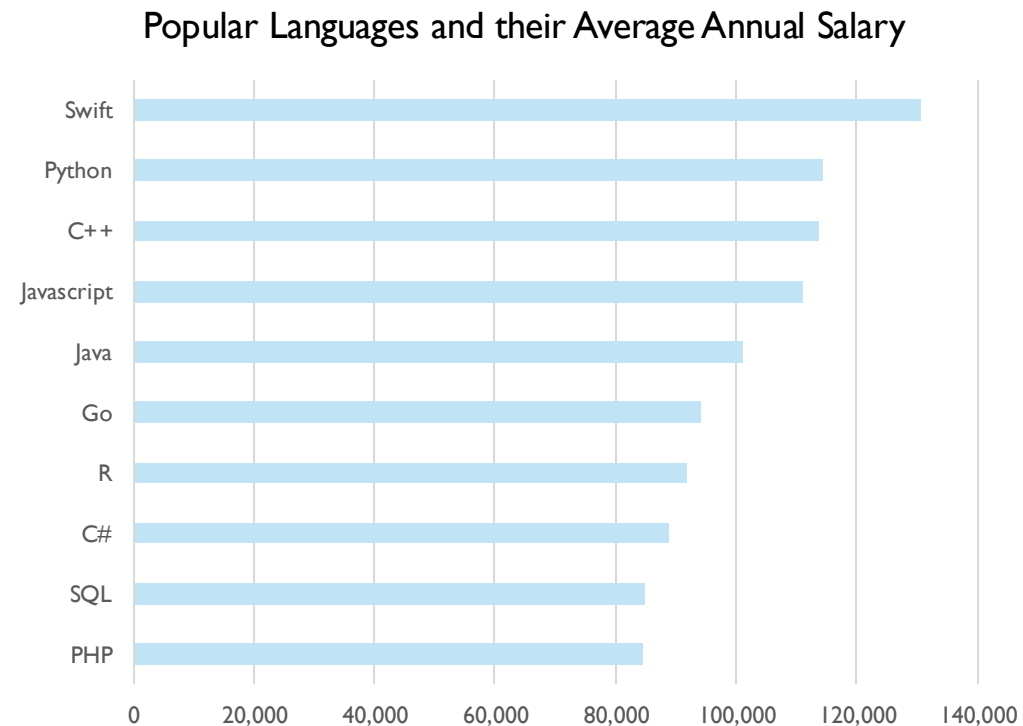


JOB POSTINGS



Bar chart presenting the job posting data in a descending order according to Technology

POPULAR LANGUAGES



Bar chart presenting the popular languages in a descending order according the Average Annual Salary