DATA ANALYTICS REPORT

LATEST TREND IN TECHOLOGY SECTOR

Rohit chandran V

6TH SEPTEMBER 2024

OUTLINE

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

EXECUTIVE SUMMARY

Objective: This report aims to identify trends in emerging techs among developers from different age groups.

Methodology: Data was collected from various sources and analyzed using statistical methods.

Results:

- gathered data on top programming languages databases and other emerging technologies.
- Inferred key insights based on the Expletory Analysis conducted.

Conclusion:

The technology industry is witnessing significant shifts in the popularity of programming languages and databases. Staying updated with these trends is crucial for developers, organizations, and educators to remain competitive and innovative. Developers' demographics doesn’t play much impact in the changing trends.

INTRODUCTION

The technology industry is rapidly evolving, with new programming languages and databases emerging regularly. organization regularly analyzes data to help identify future skill requirements.

Objective: This report presents the results of a data analytics project aimed at identifying trends for this year's report on emerging skills.

Scope: Actual Dataset comes from various sources such as job postings, Training portals & Surveys. Only a subset of the original data set is used for statistical Analysis. Analysis may only reflect only next few years trend. The target audience for this report is the organization's stakeholders, including executives, managers, and team leaders.

Purpose: To identify future skill requirements.

Questions for analysis

- What are the top programming languages in demand?
- What are the top database skills in demand?
- What are the popular IDEs?

METHODOLOGY

- The dataset used in this assignment comes from the following source:
" https://stackoverflow.blog/2019/04/09/the-2019-stack-overflow-developer-survey-results-are-in ".

- Data will be stored using SQLite in the form rows & columns.
- Preprocessing: Removed duplicates, Replaced missing value with mean value and Removed outliers and normalized data using Python.
- Analytics methods: A combination of descriptive, Exploratory Data Analysis (EDA) and Inferential Analysis were undertaken.

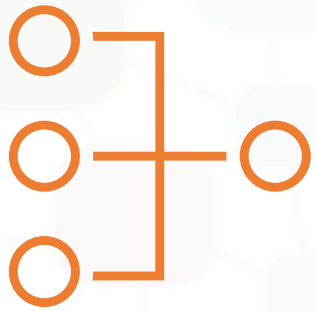
- Results will be displayed using graphs and charts and key insights were inferred from results and henceforth visualized using Cognos Analytics.

RESULTS

Respondent

11.4K

Respondent



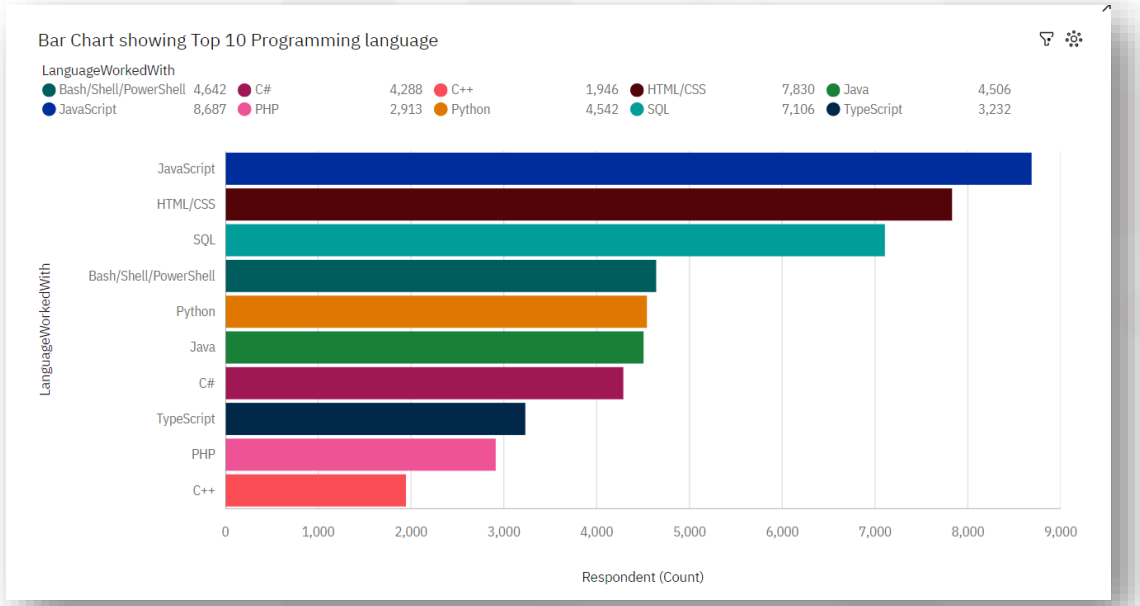
AGE	
count	11111
mean	30.77
std	7.393
min	16
25%	25
50%	29
75%	35
max	99

GENDER COUNT	
Men	10480
Women	731
Others	846

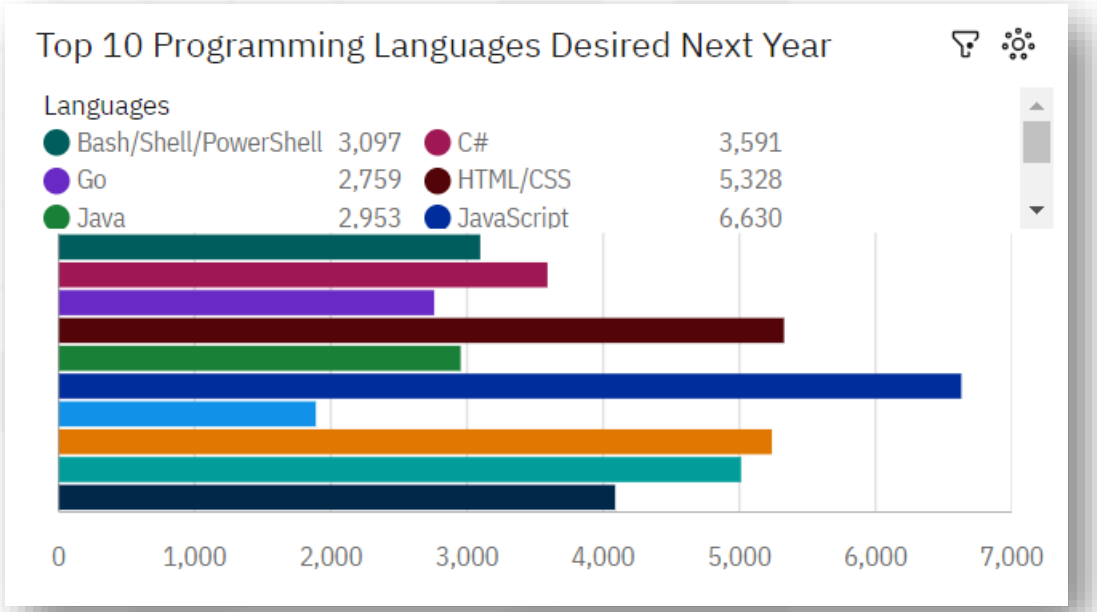
Note: Too many blank spaces among categorial variables which may affect overall accuracy in predicting Top 10 languages & Databases.

PROGRAMMING LANGUAGE TRENDS

Current Year



Next Year

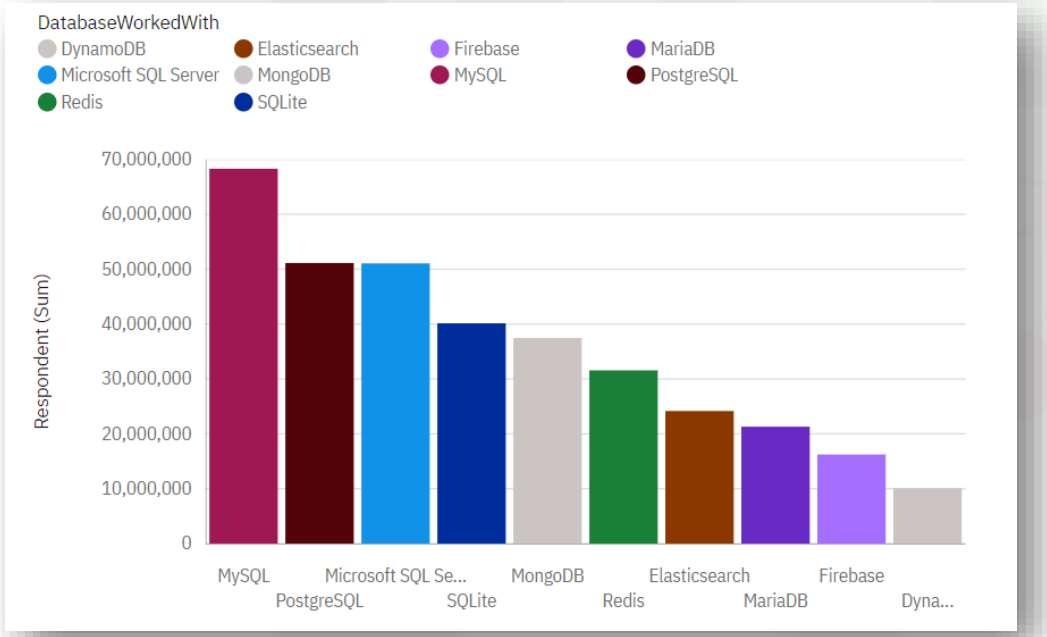


PROGRAMMING LANGUAGE TRENDS-FINDINGS & IMPLICATIONS

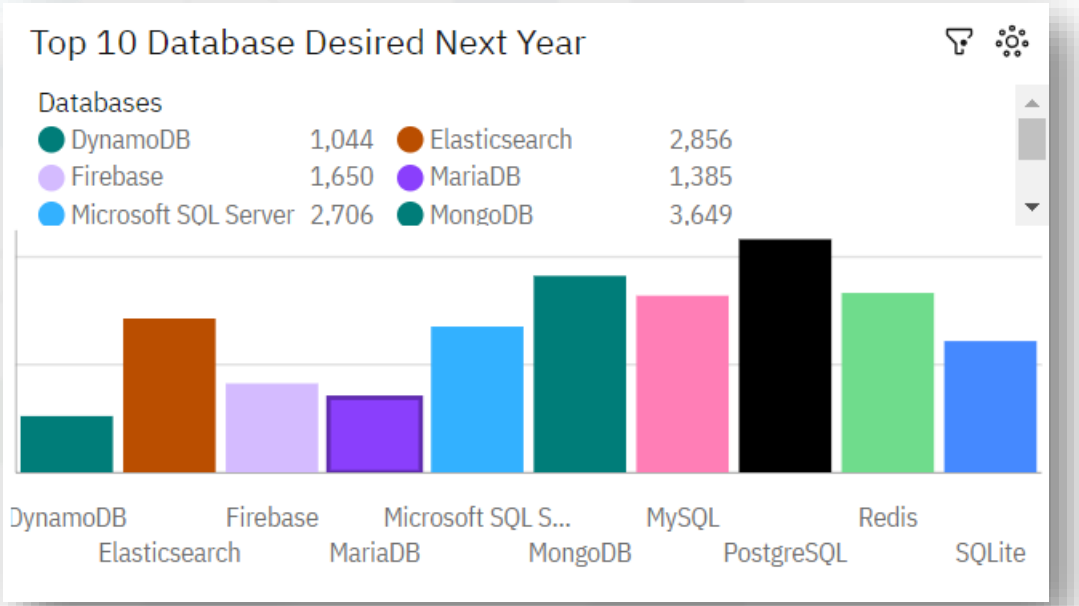
FINDINGS	IMPLICATIONS
No Change in trend among Top two languages between current & next year.	Most Respondents who worked with JavaScript & HTML/CSS are likely to continue them in future.
Rise in no of Python & Typescript developers. Also in the second chart, both rose to a better position compared to previous year.	Python & Typescript shall have more demand than any other languages in the technological industry.
In the chart that display languages desired next year, two new languages have grabbed a position among the top 10 languages.	Go & Kotlin could be the next big thing in the market among emerging programming languages.

DATABASE TRENDS

Current Year



Next Year



DATABASE TRENDS-FINDINGS & IMPLICATIONS

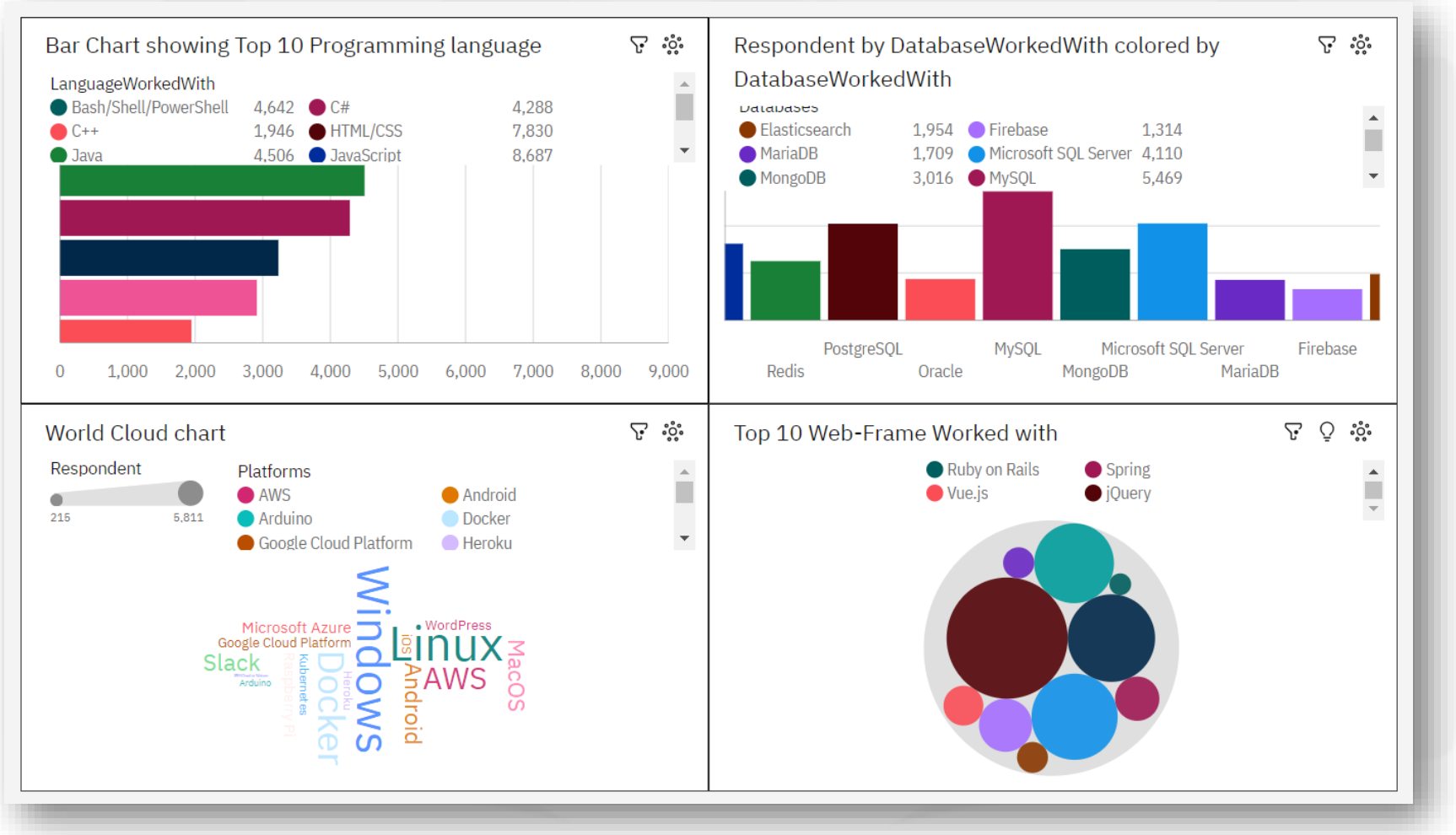
FINDINGS	IMPLICATIONS
Dip in MySQL while rise in MongoDB & PostgreSQL .	Developers using MySQL may shift to MongoDB or PostgreSQL among traditional databases available.
No change in top 10 Database.	Can't expect any new Database to pose as a threat to current market leaders.
Significant rise among developers who prefer Redis & Elasticsearch.	Developers are shifting to a better technology that provides high scalability, reliability, has better user cases over traditional databases.

DASHBOARD

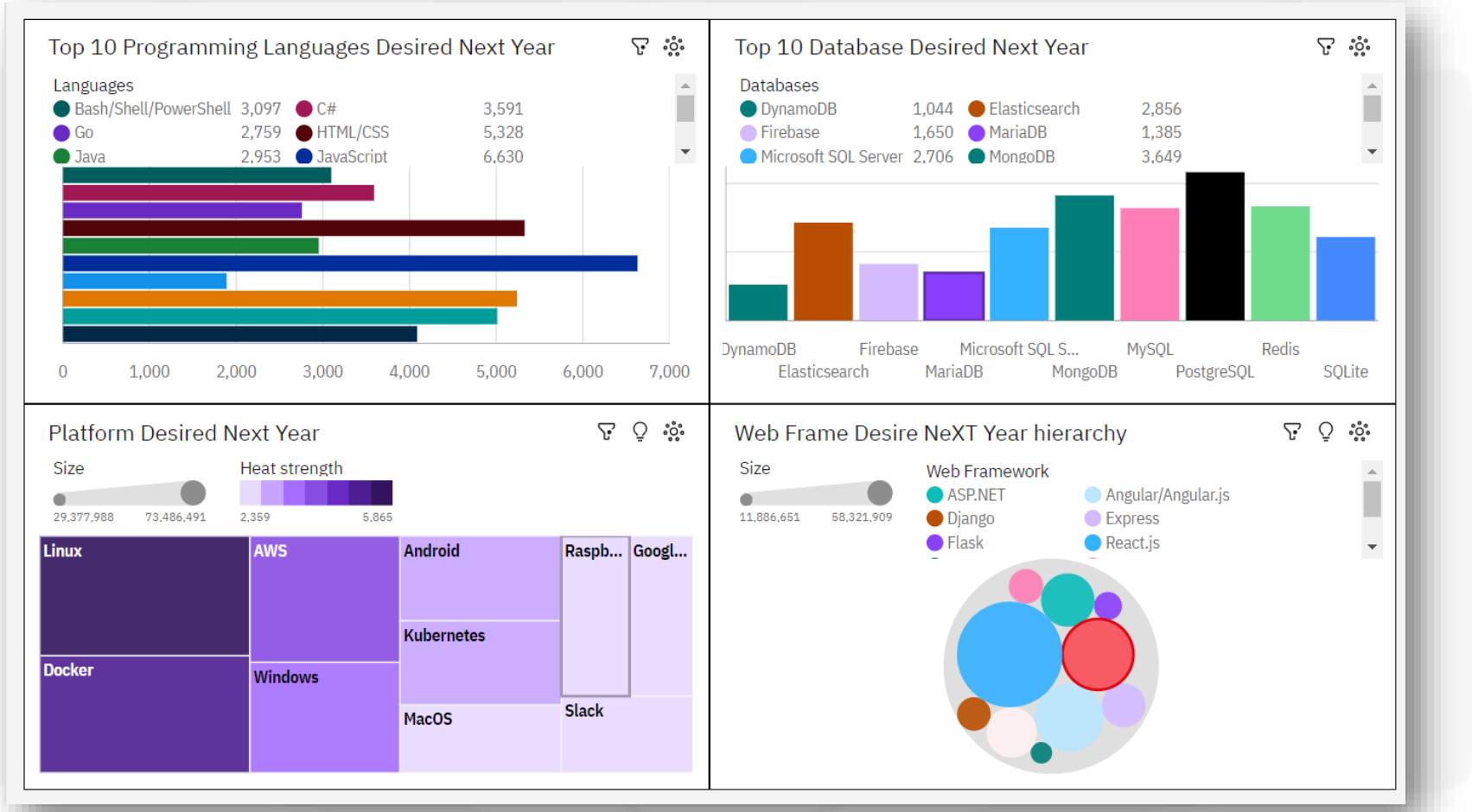
GitHub link to Cognos Analytics Dashboard:

https://github.com/rohitchandran97/CognosAnalytics_Dashboard/blob/ce249cd51d418f680ddae215fefd06fc94e971dd/New%20dashboard.pdf

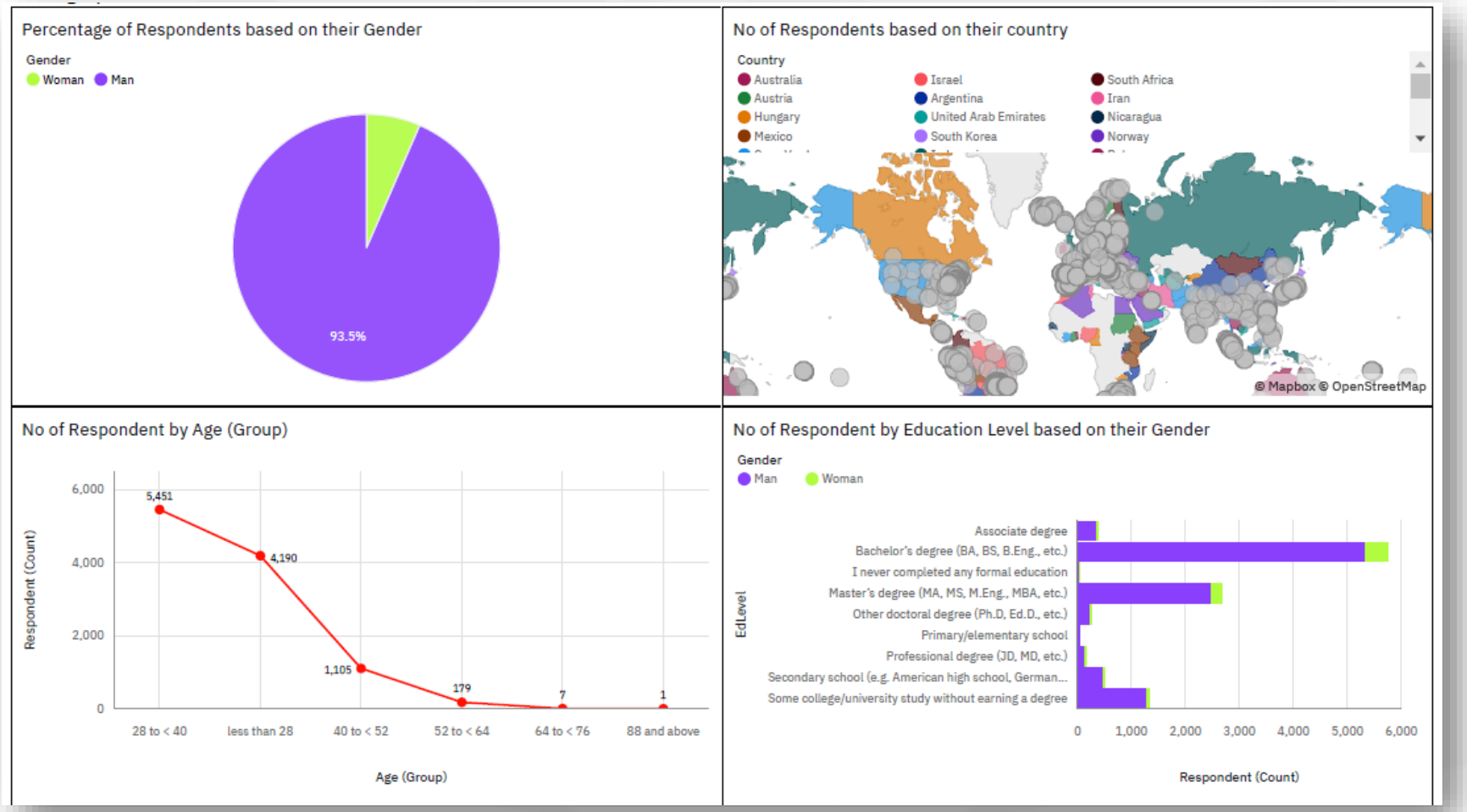
DASHBOARD TAB 1

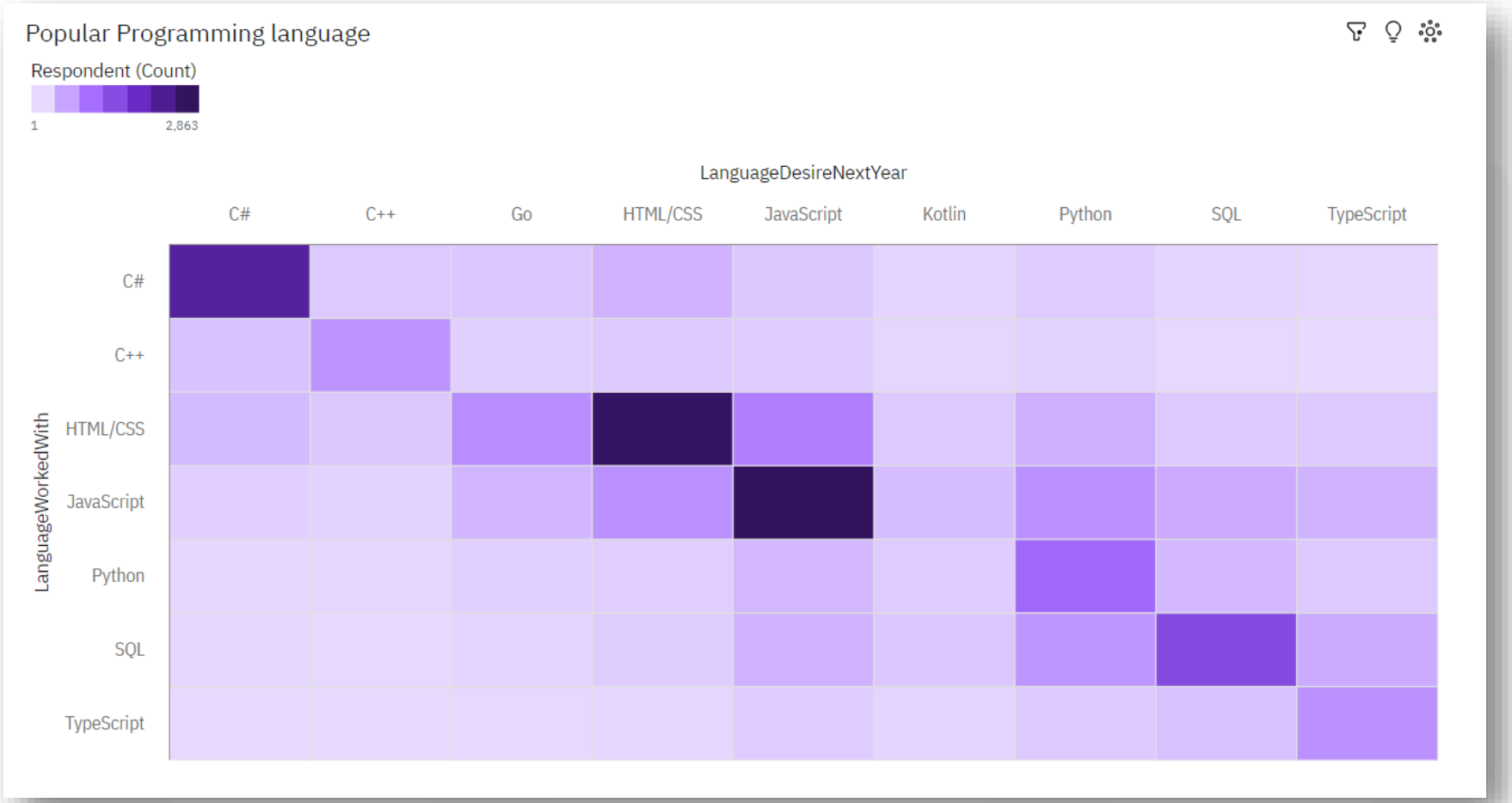


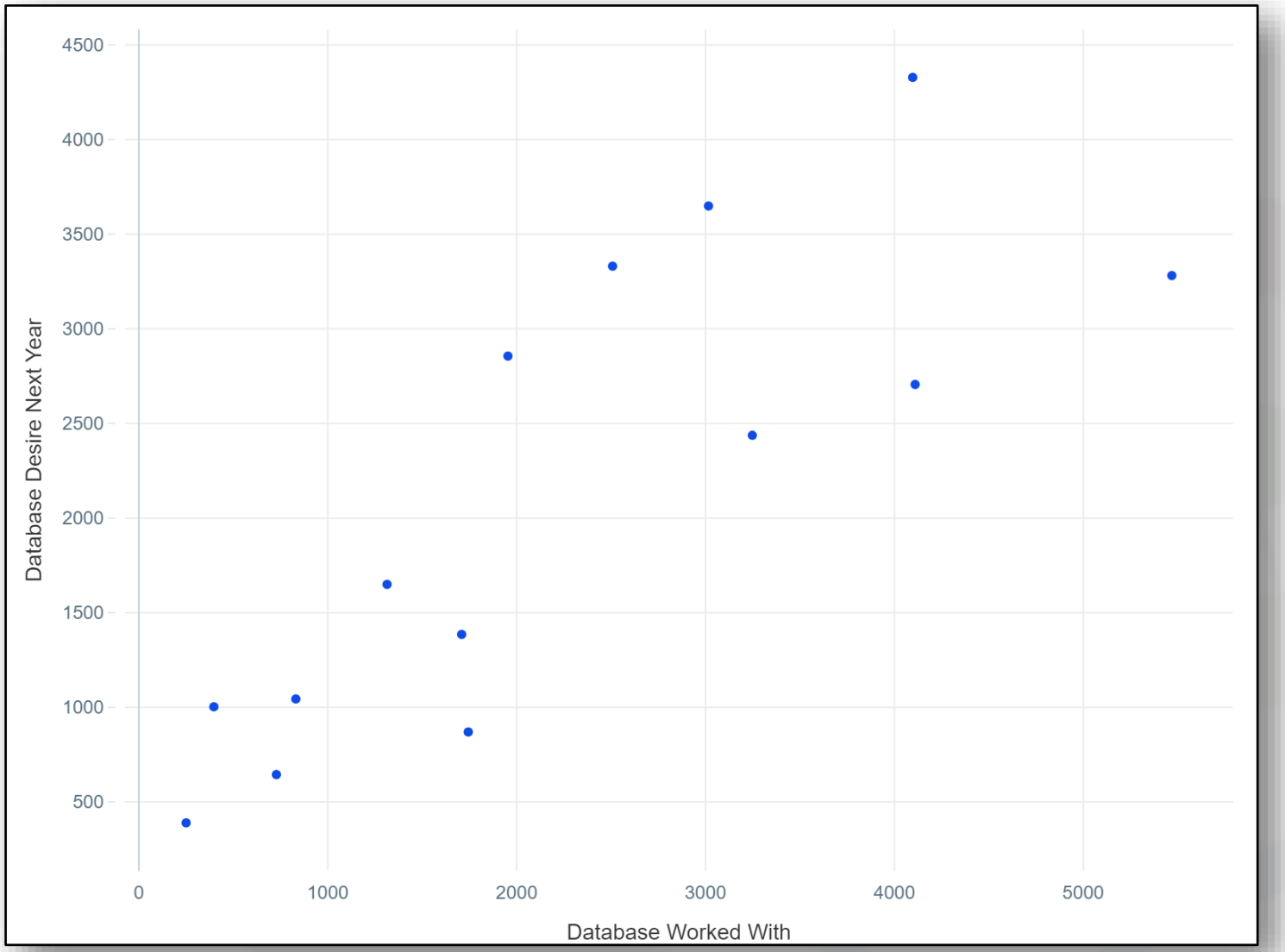
DASHBOARD TAB 2



DASHBOARD TAB 3

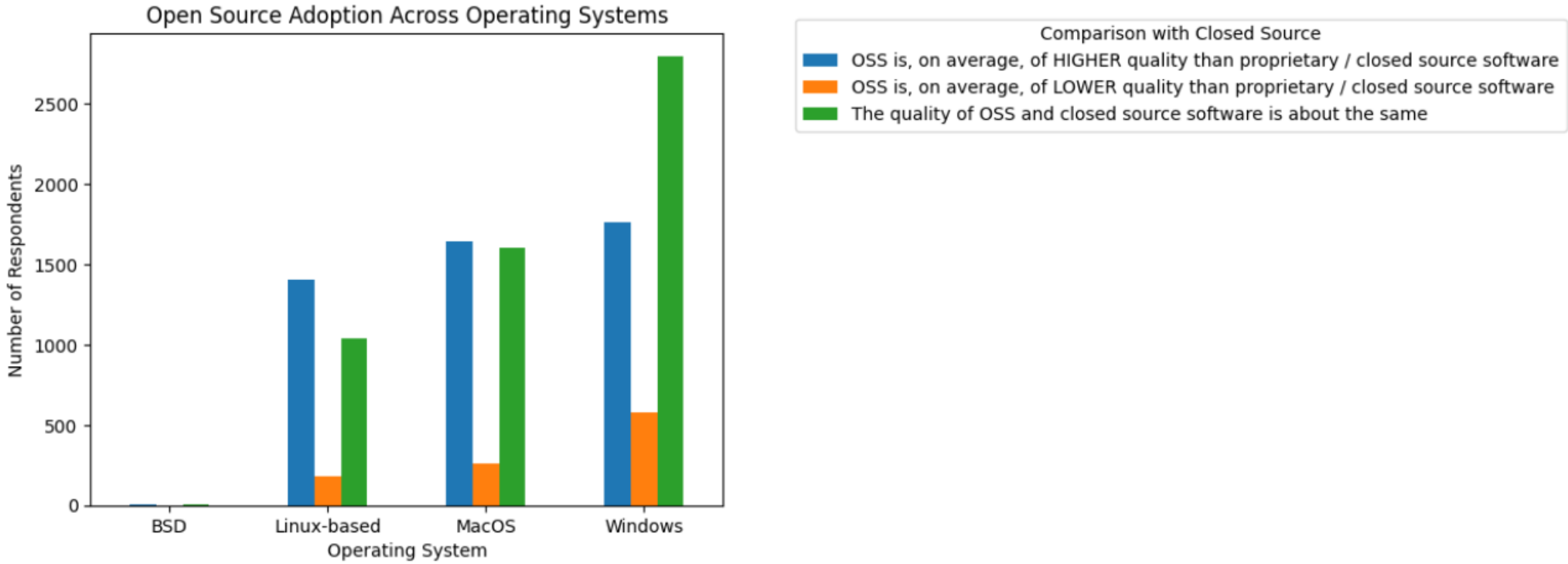


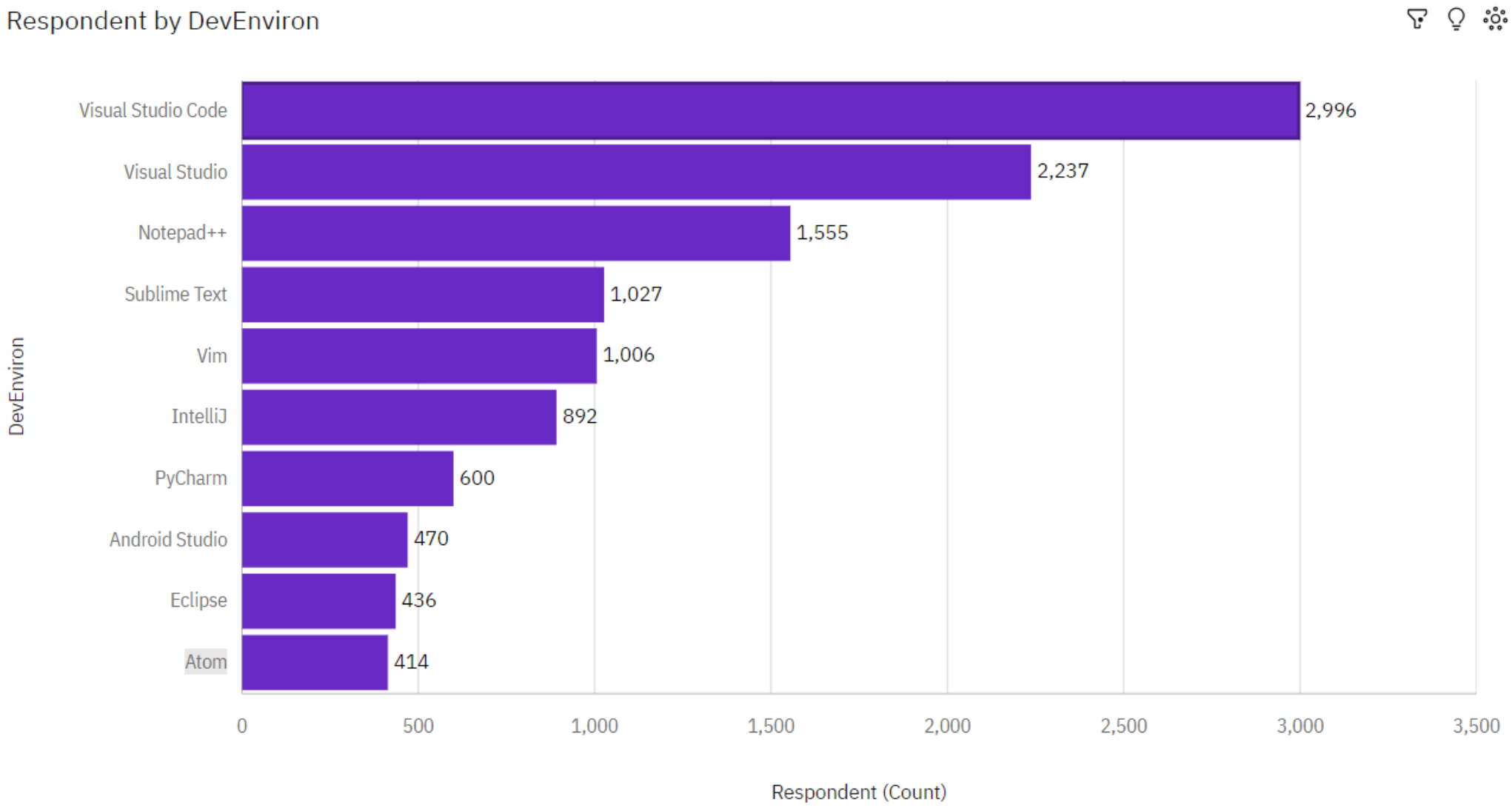

Popular Programming language
Respondent (Count)
1
2,863
LanguageDesireNextYear
C#
C++
Go
HTML/CSS
JavaScript
Kotlin
Python
SQL
TypeScript
LanguageWorkedWith
C#
C++
HTML/CSS
JavaScript
Python
SQL
TypeScript



Scatter plot to show correlation between database worked with and database desired Next Year.

Positive correlation may be due to the possibility that some developers may have choose same database in both years.


Open Source Adoption Across Operating Systems
Number of Respondents
0
500
1000
1500
2000
2500
BSD
Linux-based
MacOS
Windows
Operating System
Comparison with Closed Source
OSS is, on average, of HIGHER quality than proprietary / closed source software
OSS is, on average, of LOWER quality than proprietary / closed source software
The quality of OSS and closed source software is about the same


Respondent by DevEnviron
Visual Studio Code
Visual Studio
Notepad++
Sublime Text
Vim
IntelliJ
PyCharm
Android Studio
Eclipse
Atom
2,996
2,237
1,555
1,027
1,006
892
600
470
436
414
DevEnviron
0
500
1,000
1,500
2,000
2,500
3,000
3,500
Respondent (Count)

Inferential Statistics

Correlation Analysis:

- Positive correlation between top 10 programming languages in current & next year (r = 0.7)
- Positive correlation between Top 10 Database in current & Next year (r = 0.637).

CORRELATION BETWEEN AGE AND ALL OTHER NUMERICAL COLUMNS	
Respondent	0.002180
CompTotal	0.006337
ConvertedComp	0.401821
WorkWeekHrs	0.032032
CodeRevHrs	0.012878

DISCUSSION

- The number of respondents who have chosen JavaScript in both Years is 2,863, which makes up approximately 25.12% of the total respondents hence most JavaScript Developers will continue with JavaScript especially with frameworks like React and Vue.
- Java's decline is due to the rise of more modern languages and frameworks.
- MySQL popularity predicts that SQL remains dominant due to its reliability and widespread use.
- NoSQL databases(MongoDB) are gaining traction for their flexibility and scalability.

- Analyzing these two sets of data can also reveal insights into market trends and community support. If a database shows a decline in current usage but an increase in desired usage, it may indicate that developers are responding to new trends or community recommendation.

- The trends indicate a shift towards languages and databases that support modern development needs, such as AI, big data, and scalable web applications. The rise of NoSQL databases suggests a need for skills in handling unstructured data and scalable systems.

- The reasons behind shifts in language preference can be influenced by various factors such as job demand, community support, or new frameworks. Analyzing these two charts together may provide insights into the motivations behind developers' language choices.

OVERALL FINDINGS & IMPLICATIONS

FINDINGS	IMPLICATIONS
Rise in both Python and PostgreSQL in their respective charts	More Scope for in AI, Data Science related sector.
Man is the most frequently occurring category of **Gender** with a count of **10,480** items with **Respondent** values (**91.9** % of the total).	A lack of opportunities or resources can prevent individuals of other genders from participating.
Windows platform popularity fell to 3rd position in the Platform desired next year	Linus & Docker will be the most popular among upcoming developers.
The language that gained the most respondents is go with a gain of 1645 respondents	Go is the most popular among emerging programming languages.

CONCLUSION

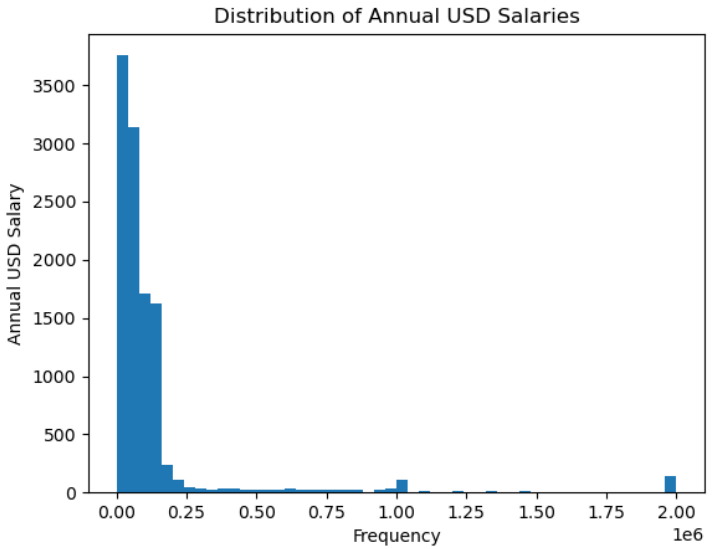
The analysis shows significant shifts in the popularity of programming languages and databases. These trends will likely influence the demand for specific skills in the tech industry. Developers should focus on learning Python and NoSQL databases to stay relevant.

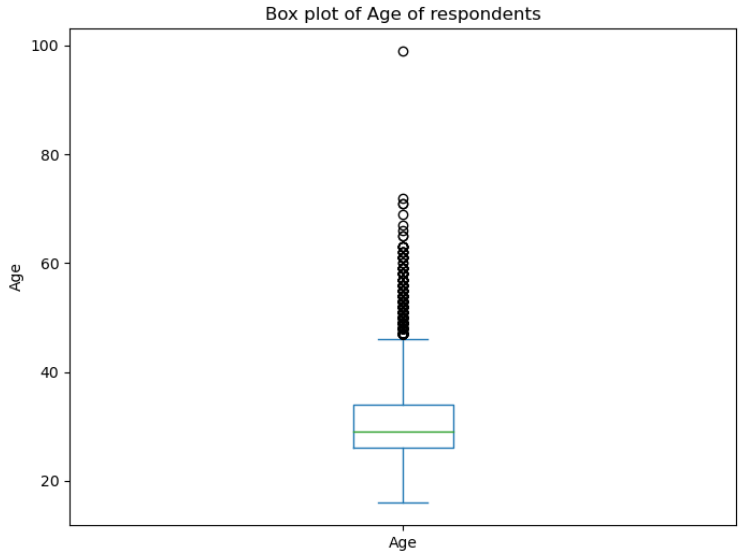
Recommendation

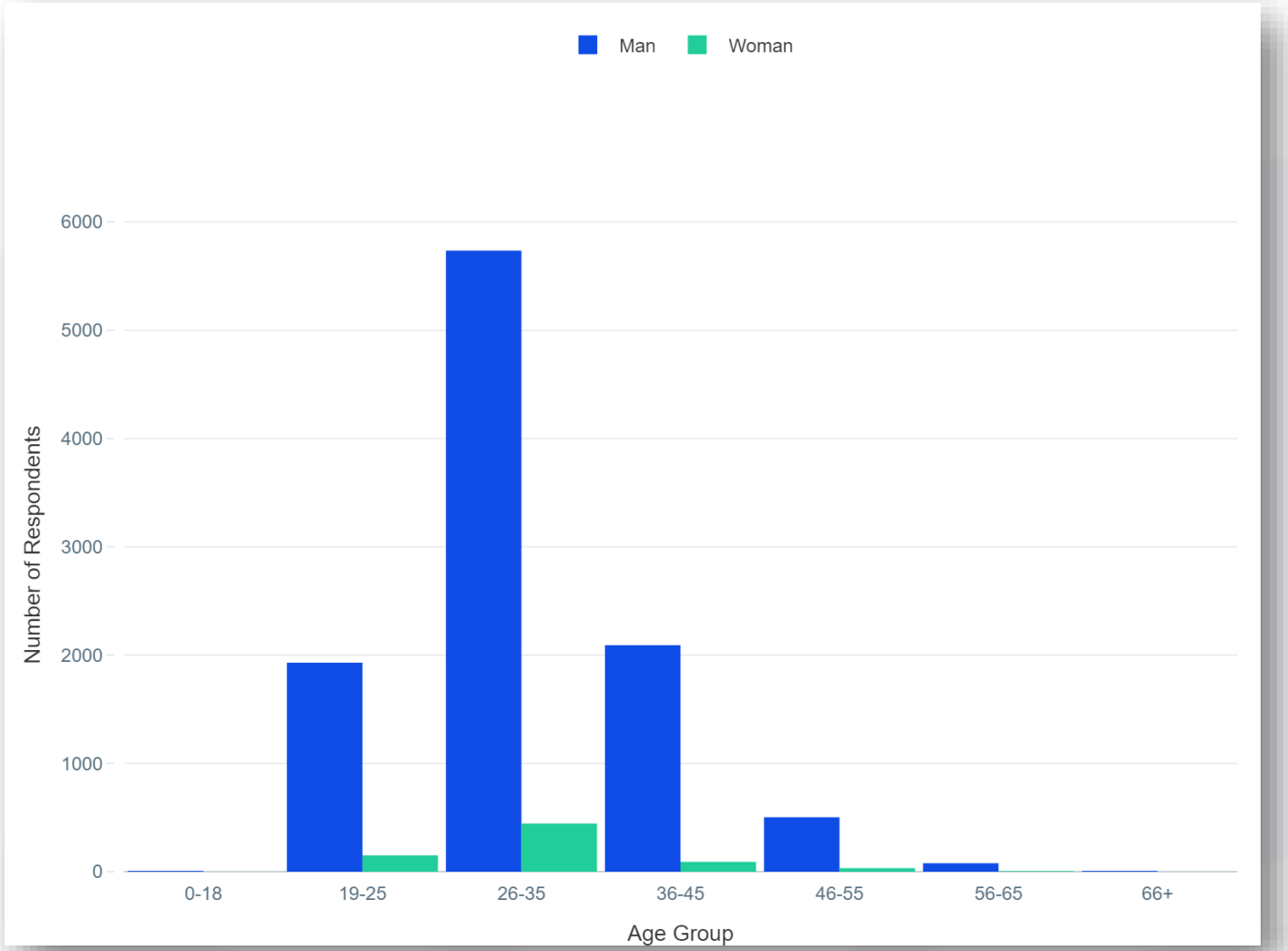
- For Developers: Invest time in learning Python and NoSQL databases.
- For Organizations: Encourage training programs focused on emerging technologies.

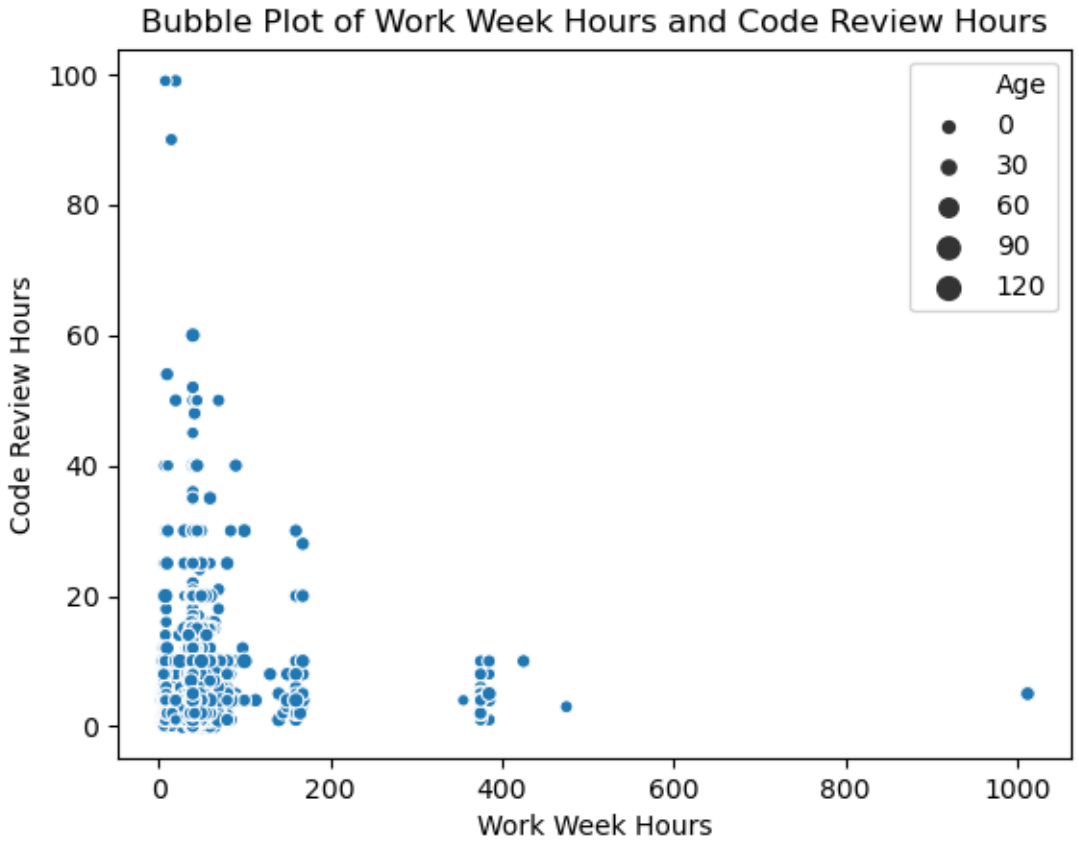
APPENDIX

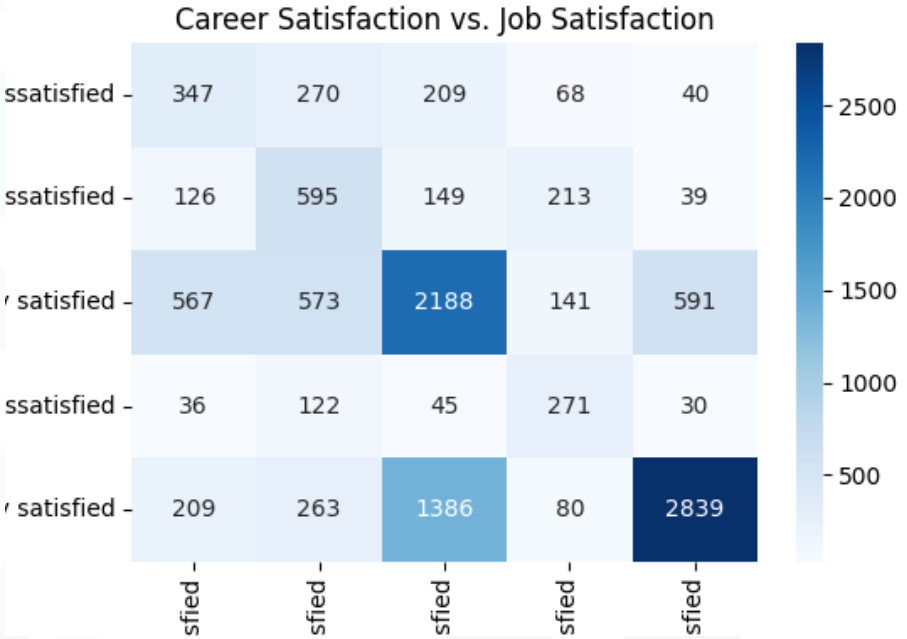
Note: Quality of Dataset is not good enough due to presence of too many ambiguous data points. Eg “ C” & “ C Programming points to same Programming language. Also “ C “ being a single word may push the machine to fetch words containing letter “ c “ and display under C Programming.

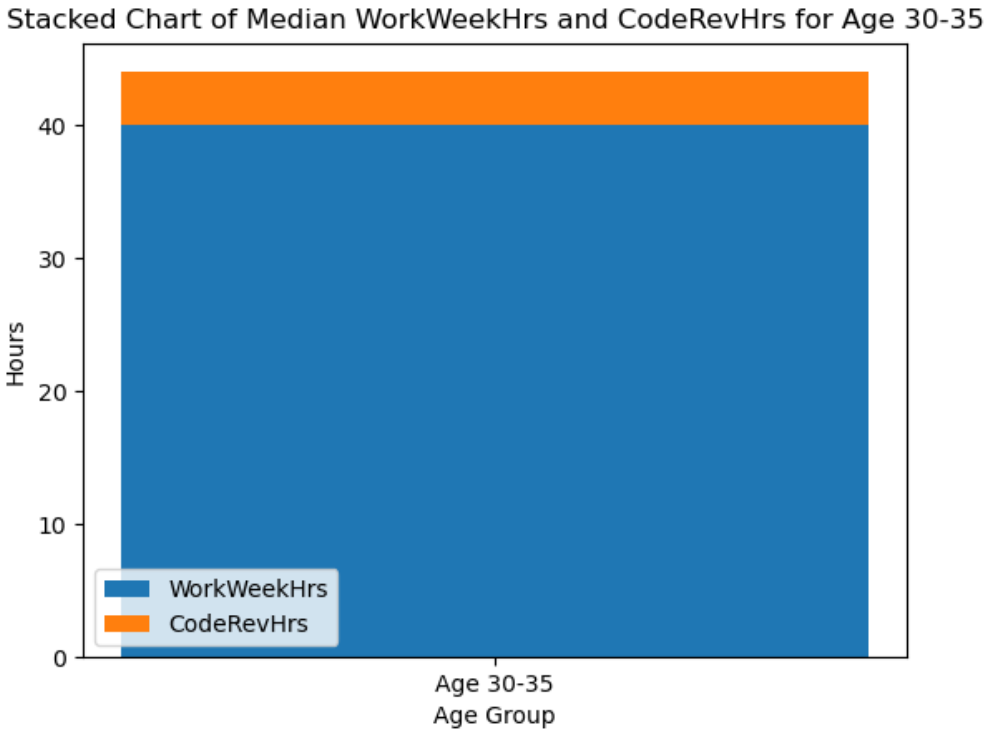


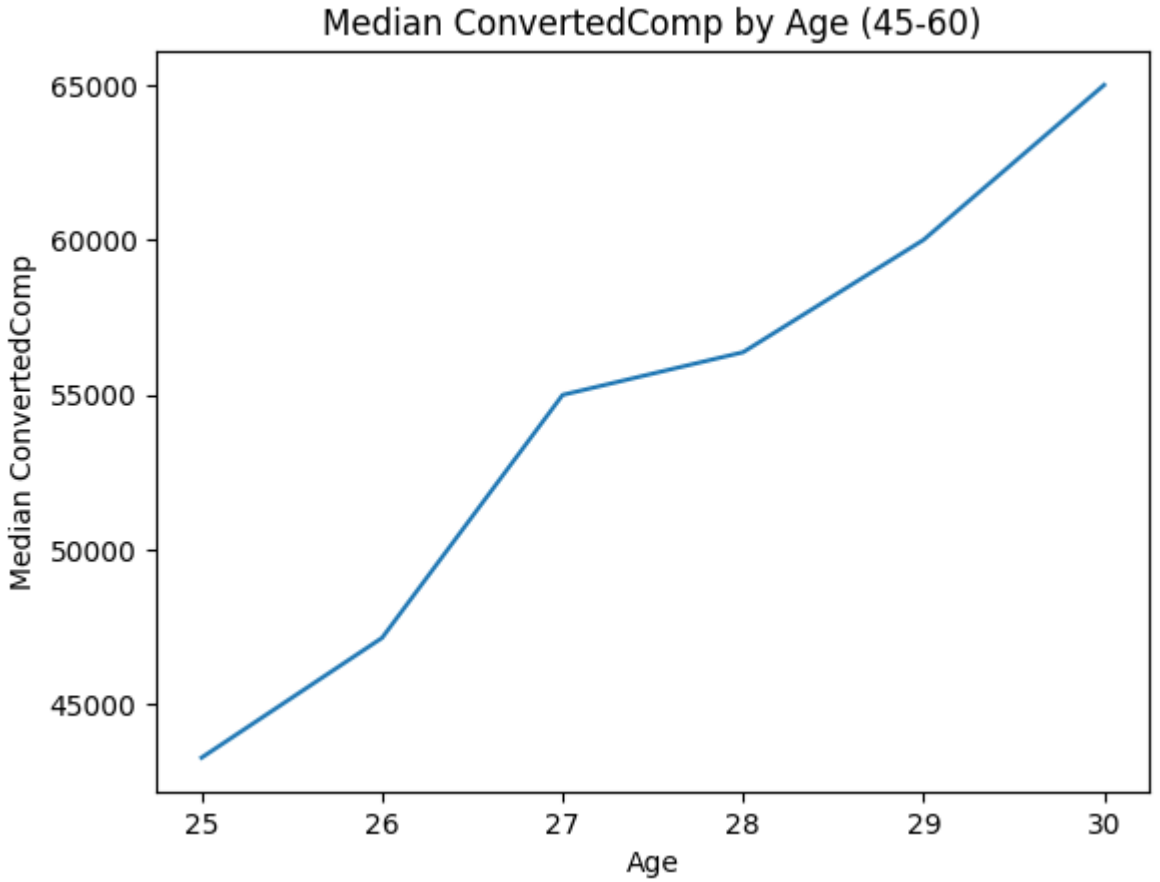



Man
Woman
6000
5000
4000
3000
2000
1000
0
Number of Respondents
0-18
19-25
26-35
36-45
46-55
56-65
66+
Age Group


Bubble Plot of Work Week Hours and Code Review Hours
Age
0
30
60
90
120
Code Review Hours
Work Week Hours
0
20
40
60
80
100
0
200
400
600
800
1000


Career Satisfaction vs. Job Satisfaction
ssatisfied
ssatisfied
satisfied
ssatisfied
satisfied
347 270 209 68 40
126 595 149 213 39
567 573 2188 141 591
36 122 45 271 30
209 263 1386 80 2839
sfied
sfied
sfied
sfied
sfied
2500
2000
1500
1000
500


Stacked Chart of Median WorkWeekHrs and CodeRevHrs for Age 30-35
Hours
0
10
20
30
40
WorkWeekHrs
CodeRevHrs
Age 30-35
Age Group


Median ConvertedComp by Age (45-60)
Median ConvertedComp
45000
50000
55000
60000
65000
25
26
27
28
29
30
Age

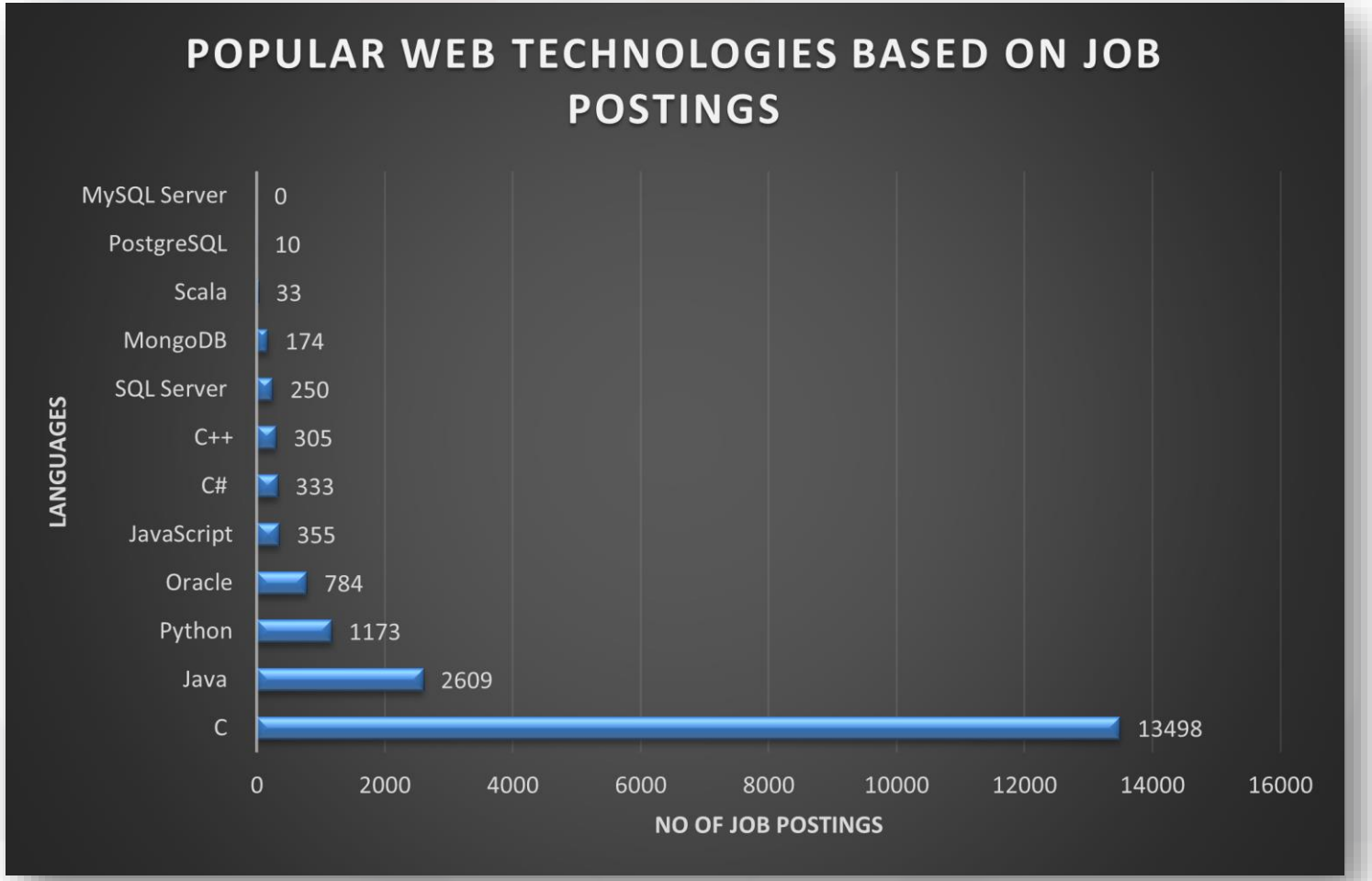
Python program to Create a pie chart of the top 5 databases that respondents wish to learn next year.

```
QUERY =  """
SELECT * FROM DatabaseDesireNextYear
"""
df1 = pd.read_sql_query(QUERY, conn)
df_DB = df1.groupby('DatabaseDesireNextYear', axis=0).count()
df_DB.head(20)
top = df_DB.nlargest(5, 'Respondent')
top['Respondent'].plot(kind='pie',
                       figsize=(15, 6),
                       autopct='%1.1f%%',
                       startangle=90,
                       pctdistance=0.5)

plt.title('Pie chart for most popular Data Base', y=1.12)
plt.axis('equal')


plt.legend(labels=df_DB.index, loc='upper left')


plt.show()
```

Python program to find correlation between `Age` and all other numerical columns.

```
import pandas as pd
from pyodide.http import pyfetch

async def download(url, filename):
    response = await pyfetch(url)
    if response.status == 200:
        with open(filename, "wb") as f:
            f.write(await response.bytes())
file_path = "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DA0321EN-
SkillsNetwork/LargeData/m2_survey_data.csv"
await download(file_path, "m2_survey_data.csv")
file_name="m2_survey_data.csv"
df = pd.read_csv(file_name)
%pip install seaborn

import seaborn as sns
import matplotlib.pyplot as plt

df = df[(df['ConvertedComp'] >= lower_bound) & (df['ConvertedComp'] <= upper_bound)]
numerical_cols = df.select_dtypes(include=[int, float])

# calculated the correlation between Age and all other numerical columns
correlations = numerical_cols.corrwith(df['Age'])
print(correlations)
```

JOB POSTINGS



POPULAR LANGUAGES