Developer Survey Analysis

Iuliia Geslin, a Coursera student

IBM Data Analyst Capstone Project

April 2023

IBM Developer

SKILLS NETWORK

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts:
    - Programming language trends
    - Database trends
  - Dashboard:
    - Current technology usage
    - Future technology trends
    - Demographics
- Discussion
- Findings & Implications
- Conclusion
- Appendix:
  - Appendix 1: Correlation coefficients between number of languages/databases and converted compensation
  - Appendix 2: Job postings
  - Appendix 3: Popular languages

EXECUTIVE SUMMARY

- This work analyses current coding trends and attempts to predict how the market may develop over the next year. Two main sources have been used:
  - survey of software professionals conducted by StackOverflow
  - dataset of job offers published on www.naukri.com
- Top 5 most used programming languages this year: JavaScript, HTML/CSS, SQL, Bash/Shell/PowerShell, Python
- Top 5 most attractive languages to learn next year: JavaScript, HTML/CSS, Python, SQL, TypeScript
- Top 5 most used databases this year: MySQL, Microsoft SQL Server, PostgreSQL, SQLite, MongoDB
- Top 5 most attractive databases to learn next year: PostgreSQL, MongoDB, Redis, MySQL, Elasticsearch
- 93,7% of survey respondents are men; average age of responders is 28

INTRODUCTION

- This work covers current coding trends. It is based on analysis of open-sourced information, including developers survey and job market statistics.
- My goal was to identify the most popular technologies to understand the current state of the market and predict demand for the next year. The analysis covers:
  - popular programming languages,
  - popular databases,
  - popular platforms,
  - popular frames.
- The hypothesis to be tested was a possible correlation between salary and the diversity of languages/databases worked with.
- The democratic characteristics of the respondents were analysed in order to try to build a typical developer portrait, but also to assess the reach, inclusiveness and representativeness of the survey. I looked at:
  - gender,
  - country of residence,
  - age,
  - level of education.

METHODOLOGY

- Stack Overflow, a popular website for developers, conducted an online survey of software professionals around the world. The survey data was later open sourced here under a ODbL: Open Database License.
  - The actual dataset has about 90,000 responses.
  - A randomised subset of about 1/10th of the original data set has been analyzed.
- Another dataset helped to complete the results with statistics on job postings and salaries for different programming languages in different American cities.
- The data has been collected and analyzed using Python for data analysis, including:
  - Beautiful Soup
  - pandas library
  - matplotlib
  - Seaborn
- IBM technologies has been helpful:
  - IBM Cognos Analytics
  - IBM Cloud Pak for Data
  - IBM Watson Studio

RESULTS

Here I'll share graphs:

- Top 5 most widely used programming languages this year
- Top 5 chosen programming languages to learn next
- Top 5 most widely used databases this year
- Top 5 most mentioned databases to learn next year

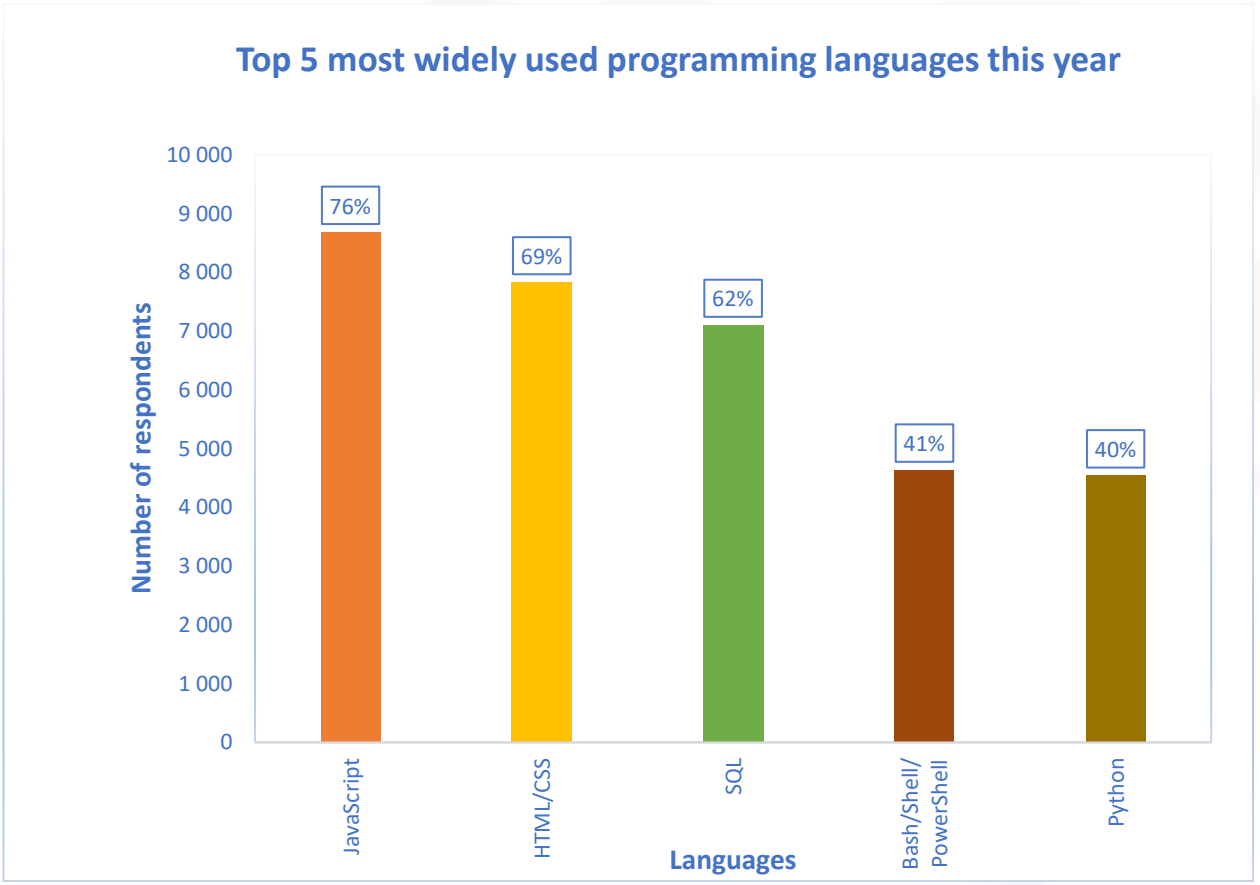
Then we'll move toward on to testing the hypothesis about the impact of the number of languages and databases known by developers on the average salary.

Finally, we'll look at a dashboard summarizing current technology usage, future technology trends and the demographics of the survey.

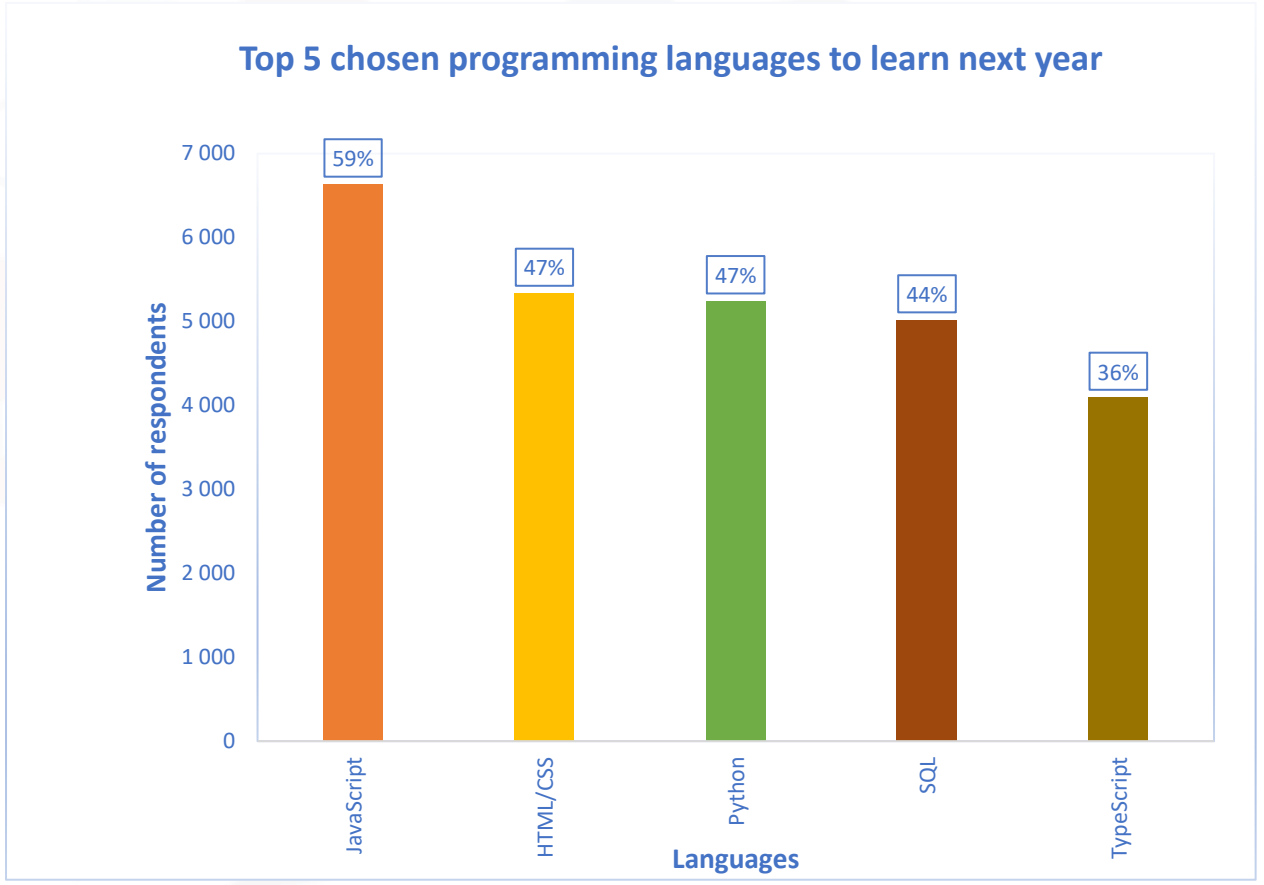
Note that the percentages in the graphs add up to more than 100% because the questions related to used and desired languages and databases are checkbox-type questions that is allow respondents to select multiple responses.

PROGRAMMING LANGUAGE TRENDS

Current Year



Next Year



IBM Developer

SKILLS NETWORK

PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

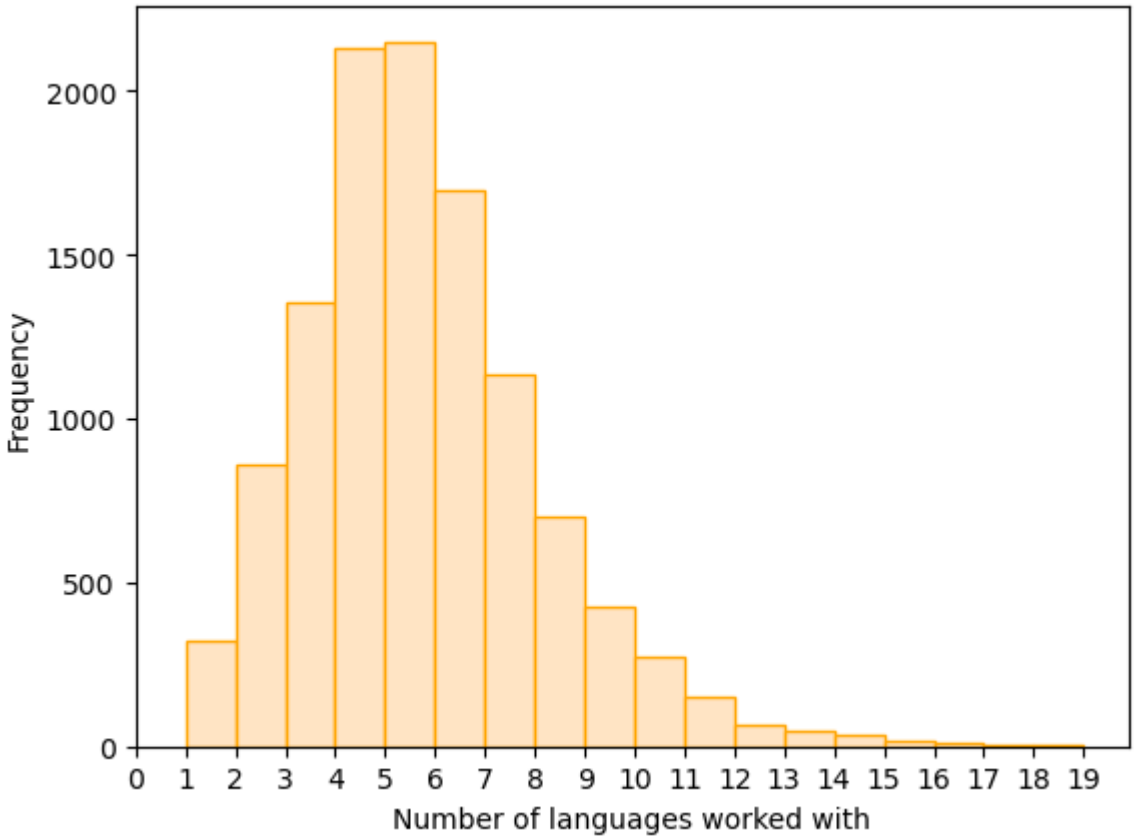
Findings

- JavaScript is by far the most popular programming language and continues to attracting new users
- SQL and Python are growing fast with 62% and 40% of respondents respectively using them this year, 44% and 47% respectively intending to learn them next year
- TypeScript is emerging with 36% of developers expressing desire to learn it next year
- Bash/Shell/PowerShell are widely used (41% of respondents) but are not among the most attractive languages to learn next year

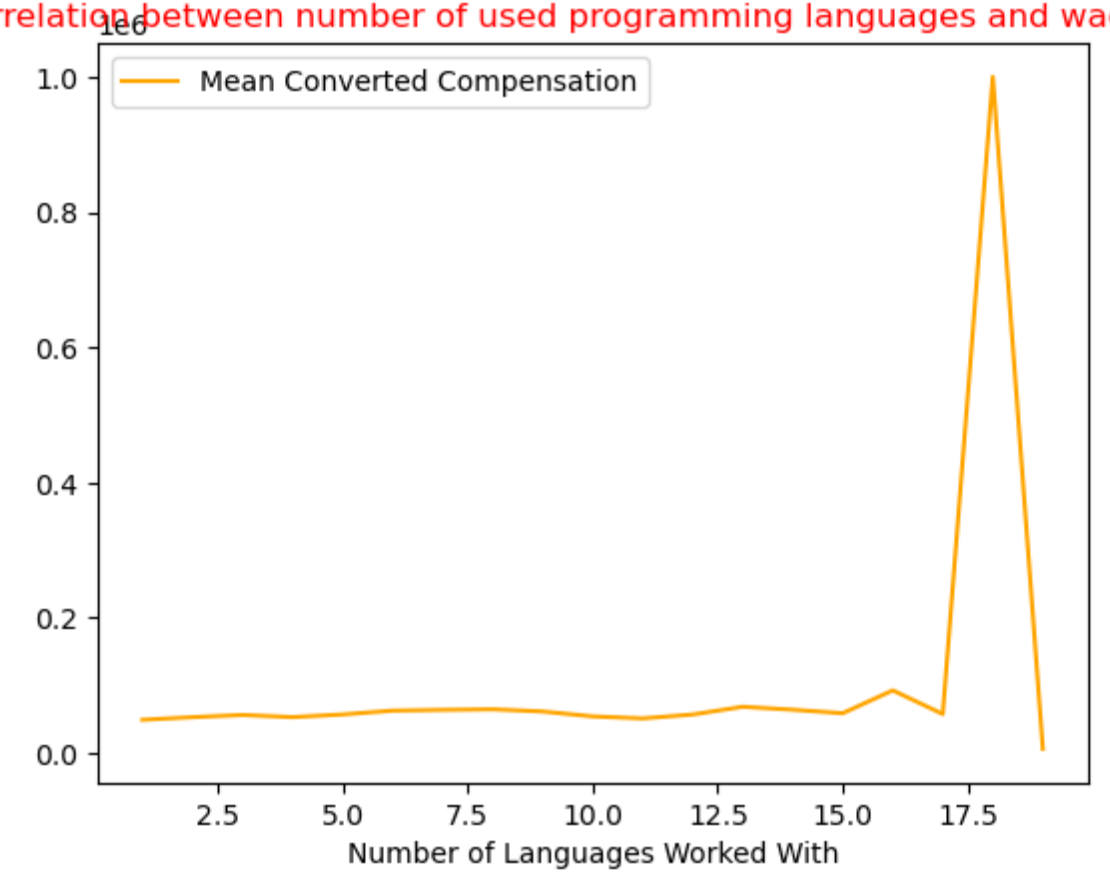
Implications

- JavaScript is set to continue its growth, and it is well worth learning for those who want to remain competitive in the job market
- Since SQL and Python are widely used for data analysis, their popularity may indirectly indicate an explosion in the data analysis field
- TypeScript, a superset of JavaScript, owes its growth to theavaScript's popularity and will be in demand next year
- Bash's, Shell's and PowerShell's are likely to see their growth slow down next year

TESTING POSSIBLE CORRELATION BETWEEN NUMBER OF PROGRAMMING LANGUAGES AND AVERAGE SALARY



Correlation between number of used programming languages and wage level



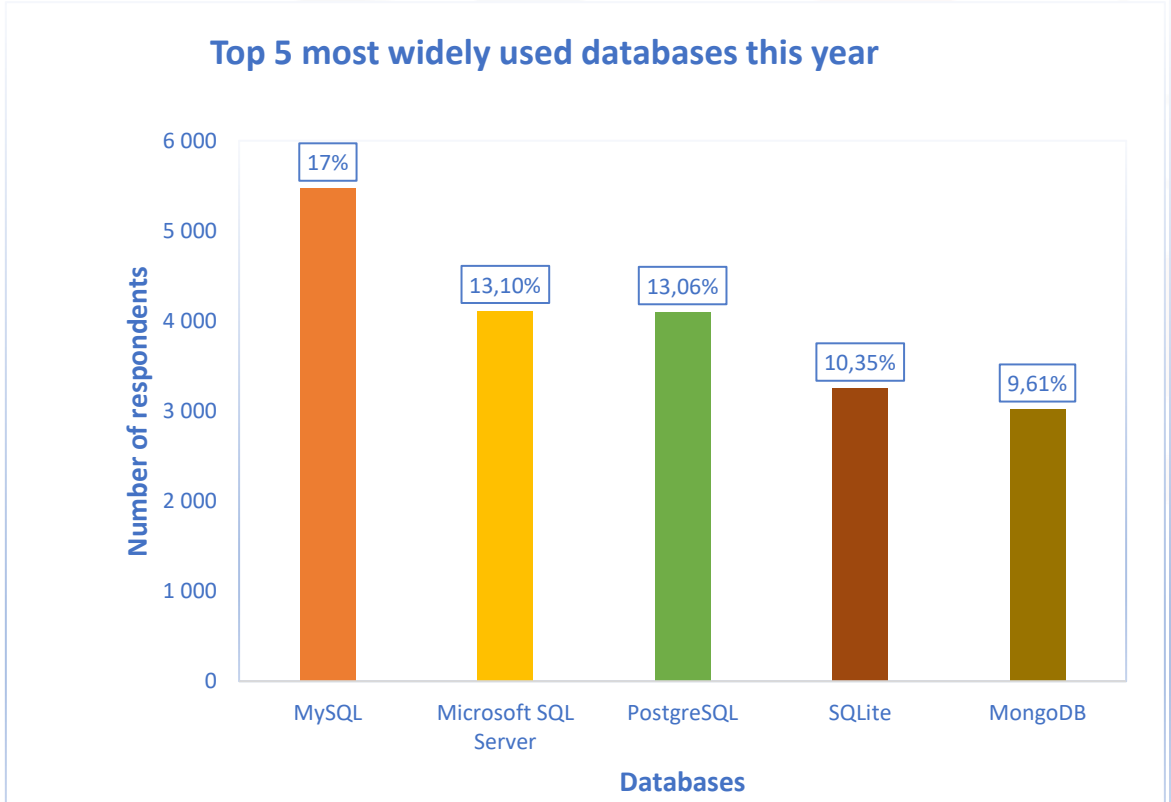
HYPOTHESIS TESTED: FAILED

- Most of developers code in several languages, with half of respondents using between 4 and 6 different programming languages
- No correlation was found between number of the languages and salary (correlation coefficient shown in the appendix 1)

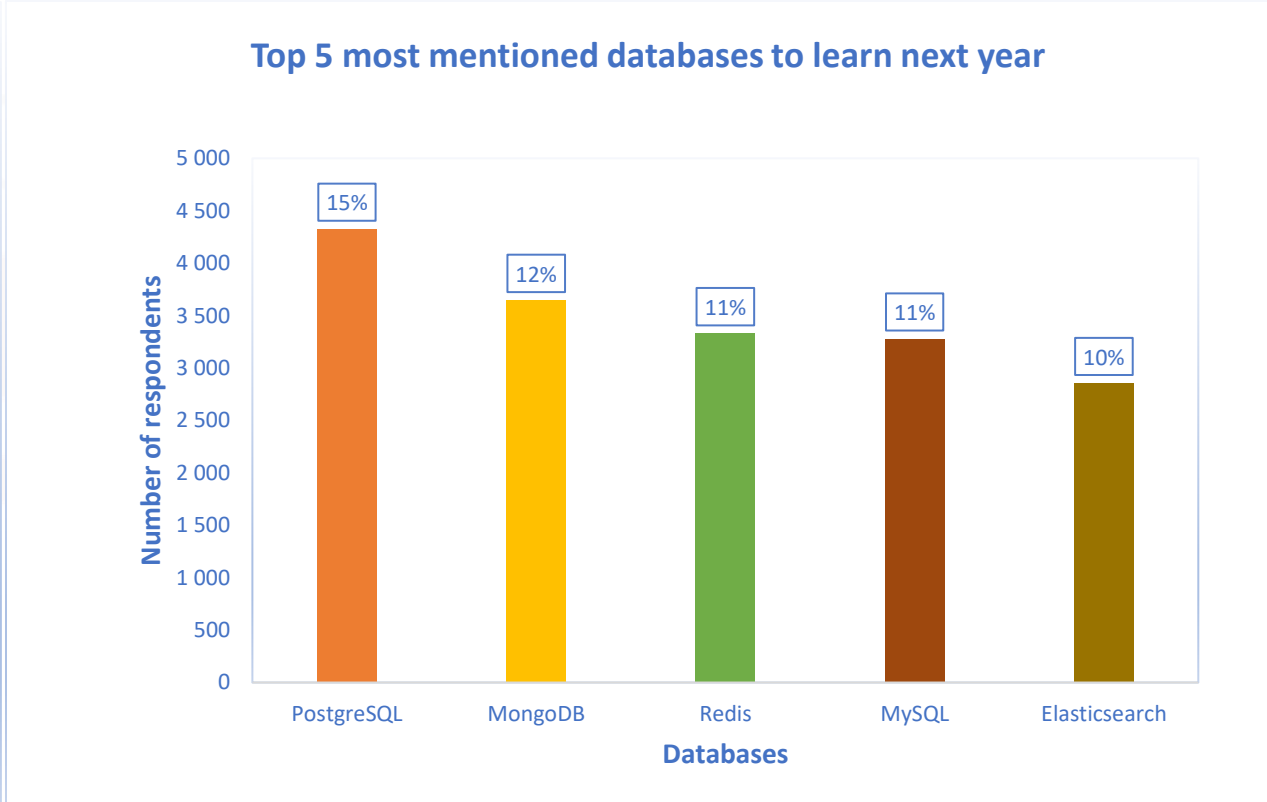
Knowing different programming languages can be an advantage for developers, although a solid knowledge of just one language seems to be enough to get a job

DATABASE TRENDS

Current Year



Next Year



DATABASE TRENDS - FINDINGS & IMPLICATIONS

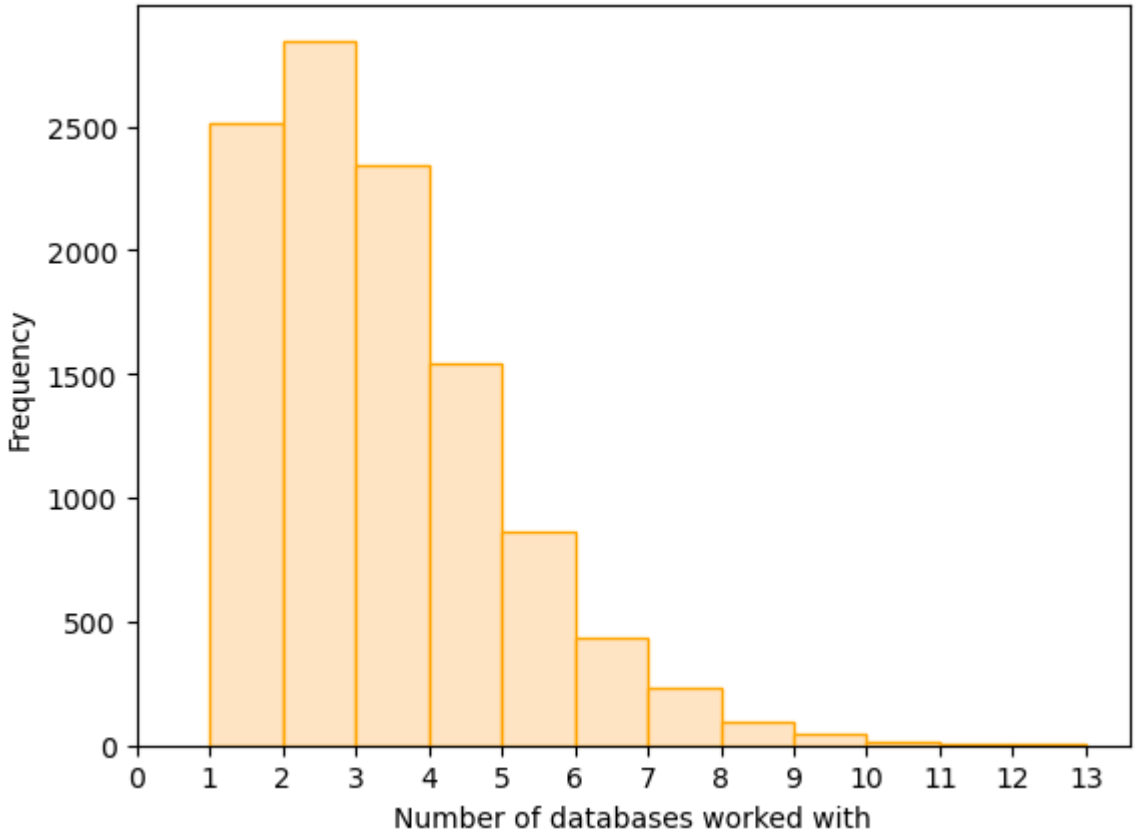
Findings

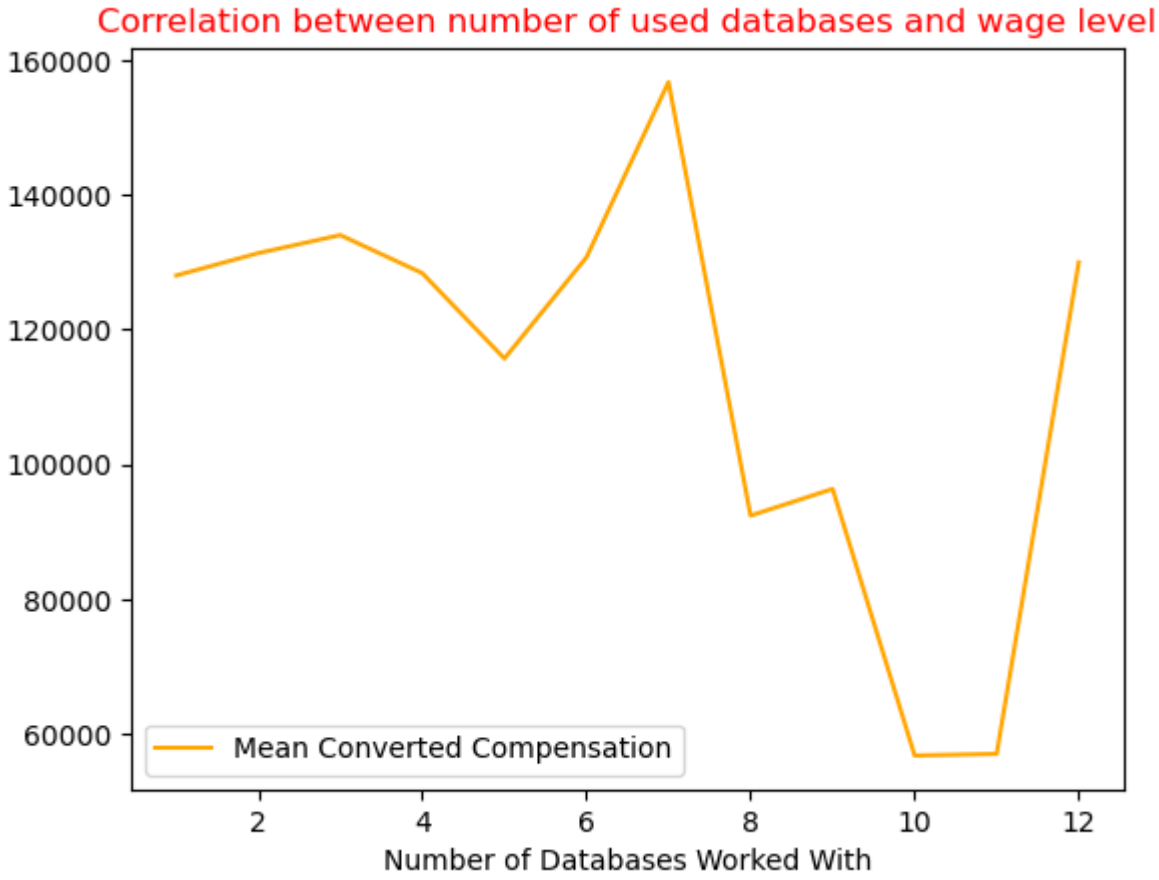
- MySQL is the most popular database (used by 17% of respondents)
- PostgreSQL and MongoDB are already in the top 5 this year and their popularity is growing as more developers are willing to learn these databases
- Microsoft SQL Server is widely used (13,1% of respondents this year) but fails to attract new users
- Redis and ElasticSearch seem to be emerging as more developers are willing to learn them next year

Implications

- PostgreSQL database will be in high demand next year, followed by MongoDB
- MySQL will remain very popular
- Microsoft SQL Server and other less popular databases may cede some of their market to newly emerging databases such as Redis and ElasticSearch

TESTING POSSIBLE CORRELATION BETWEEN NUMBER OF DATABASES AND AVERAGE SALARY





HYPOTHESIS TESTED: FAILED

- Almost a quarter of respondents use only 1 database, although most of them use 2 and many use 3 different databases.
- No correlation was found between the number of databases and salary (correlation coefficient shown in appendix 1)

Using only one or two different databases is very common for developers and is enough to get a well paid job.

DASHBOARD

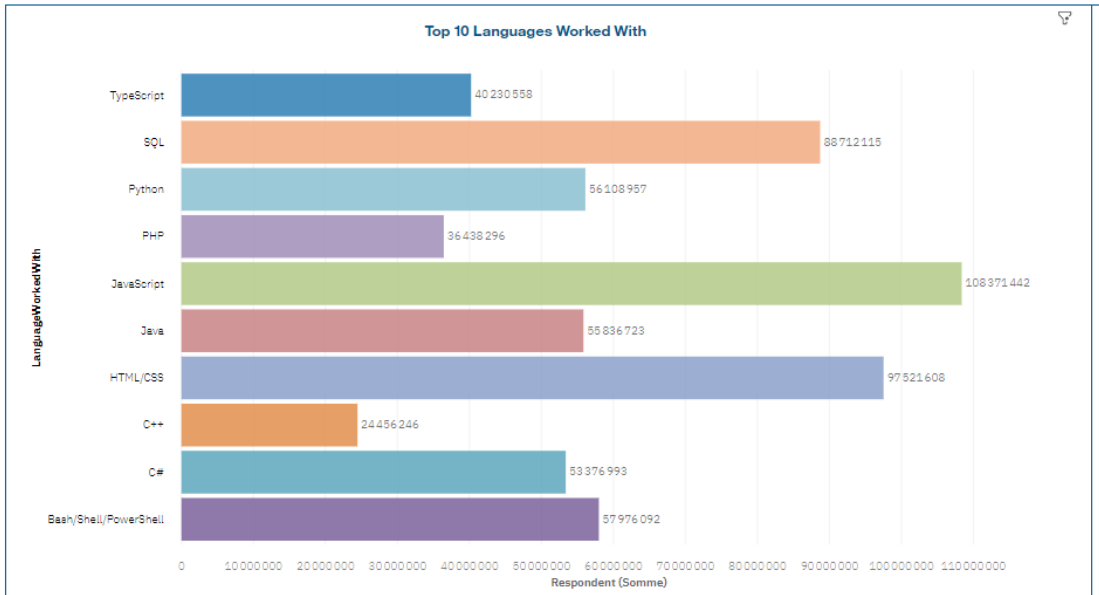
https://dataplatform.cloud.ibm.com/dashboards/a054f524-93d6-4444-824e-11450854e6fd/view/500ade1738ab099f11e0b1e4079978532b352359b3bbd750d4847b495a607697a8681ac4c82c4a0bd9160c62f6ef105b9b

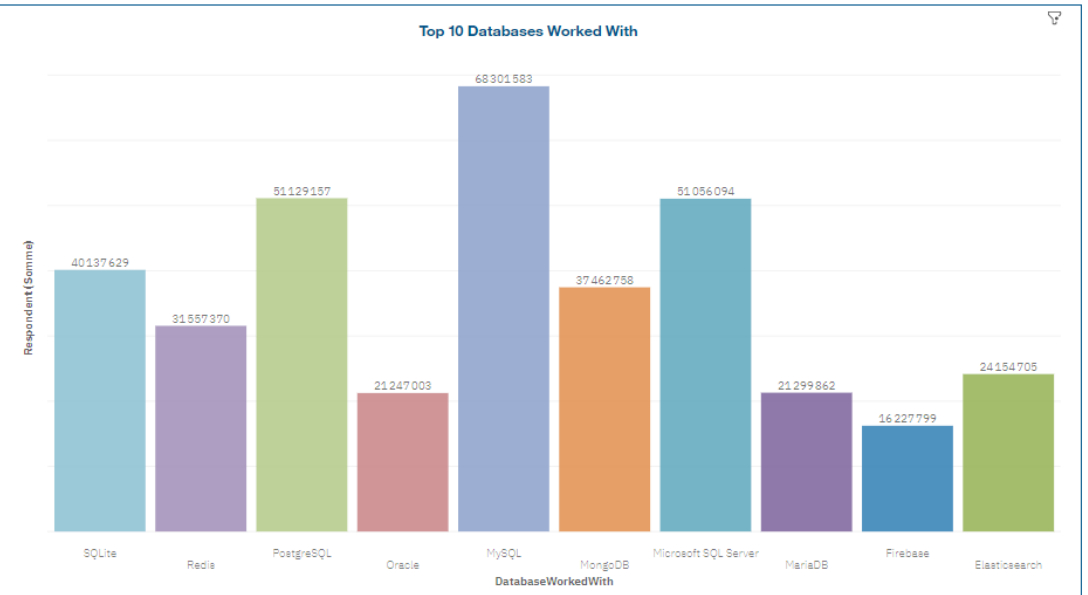
DASHBOARD TAB 1

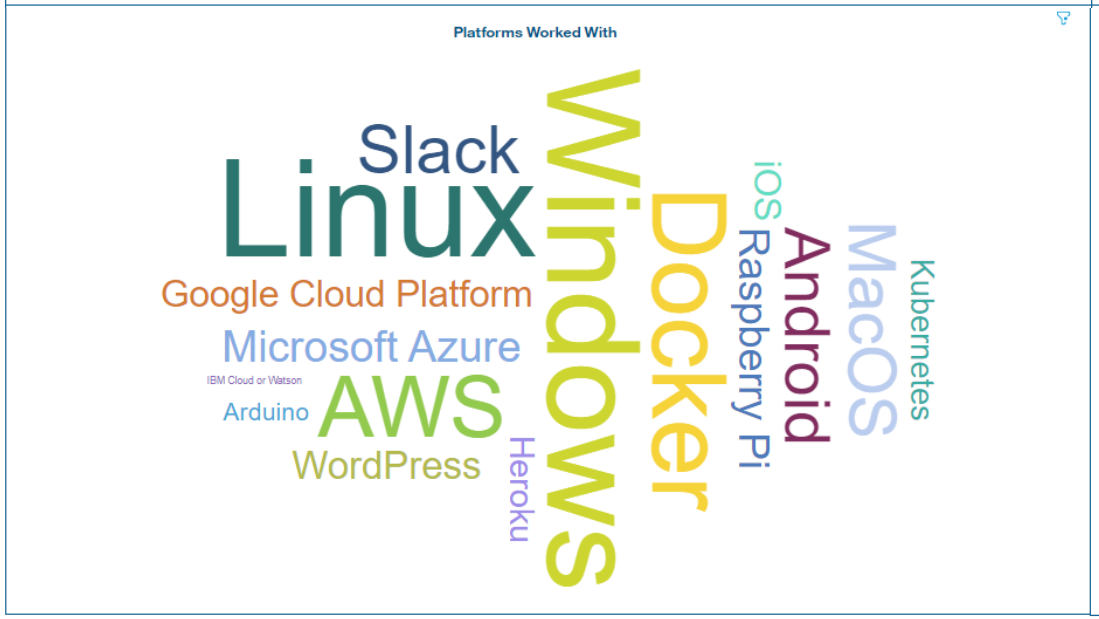
Tous les onglets — Aucun filtre n'a été appliqué.

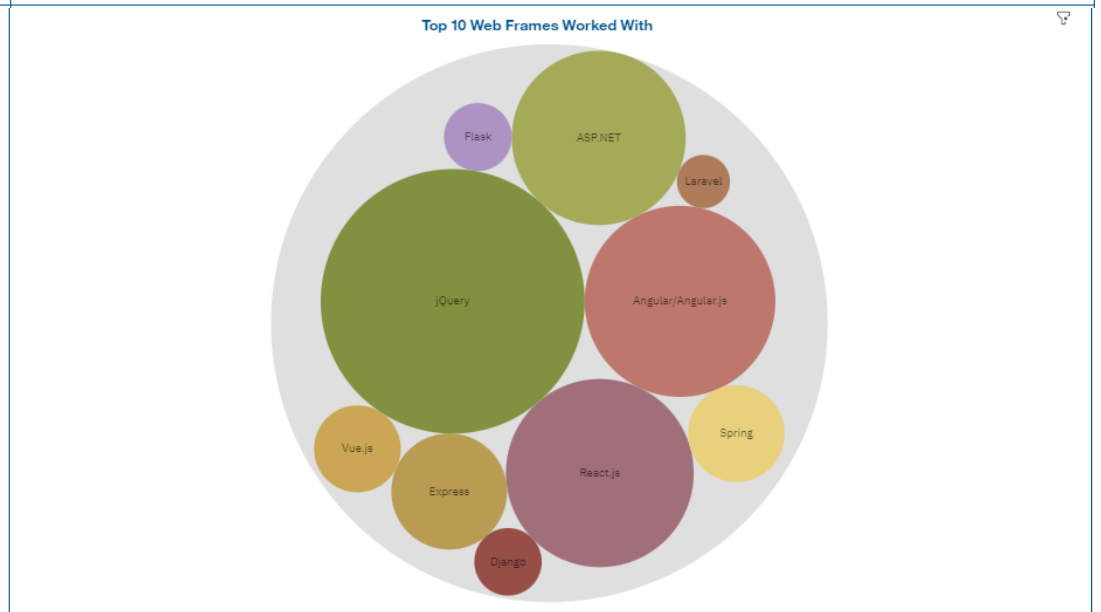
Cet onglet — Aucun filtre n'a été appliqué.

Current Technology Usage | Future Technology Trend | Demographics









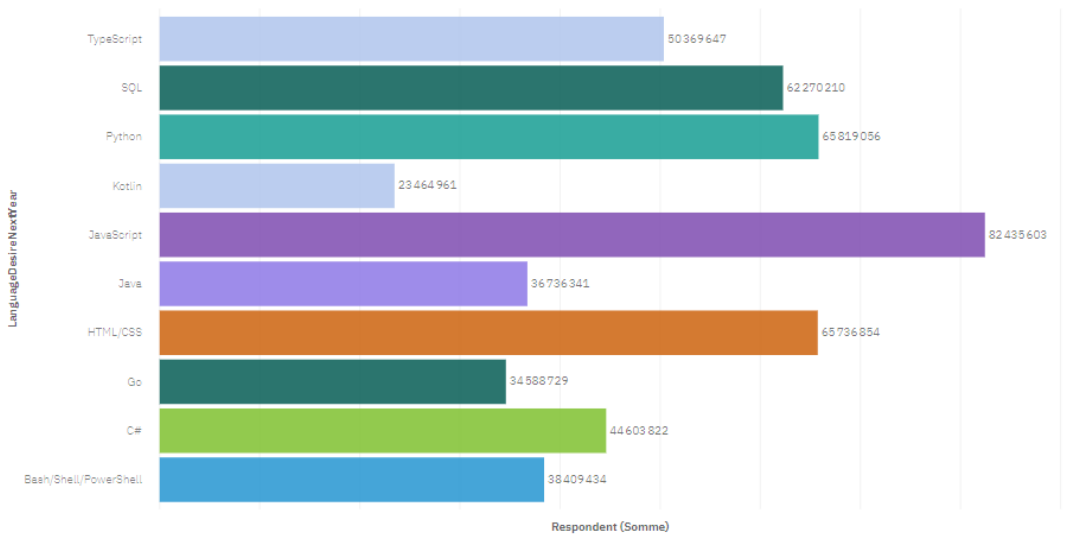
DASHBOARD TAB 2

Tous les onglets — Aucun filtre n'a été appliqué.

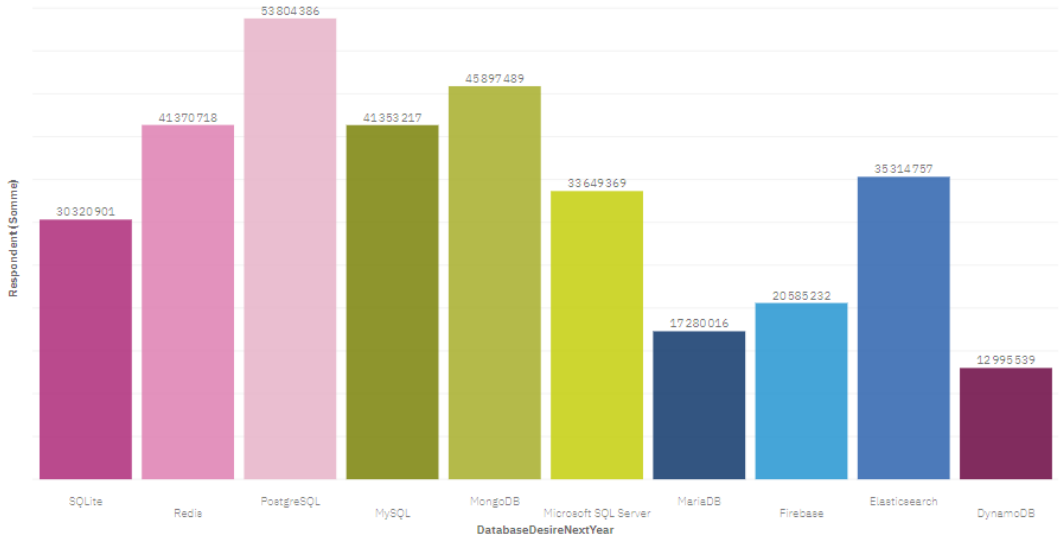
Cet onglet — Aucun filtre n'a été appliqué.

Current Technology Usage | Future Technology Trend | Demographics

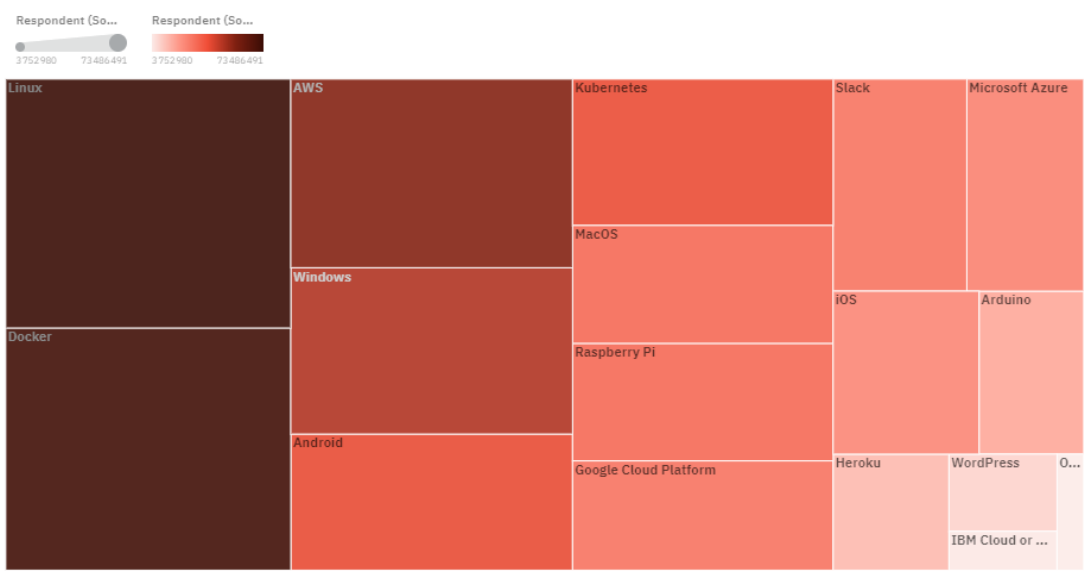
Top 10 Languages Desired Next Year



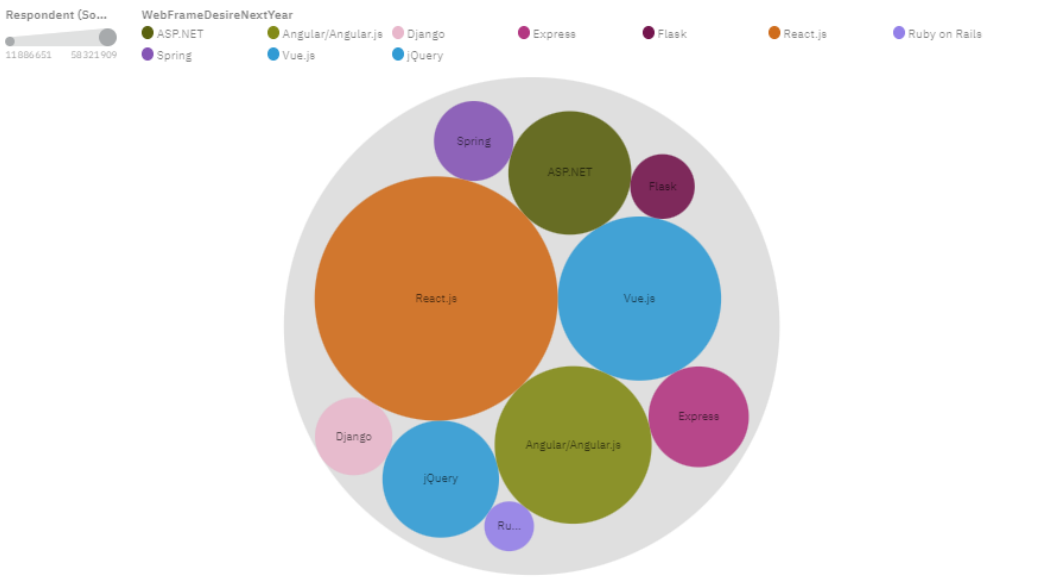
Top 10 Databases Desired Next Year



Platforms Desired Next Year



Top 10 Web Frames Desired Next Year



IBM

DASHBOARD TAB 3

Tous les onglets

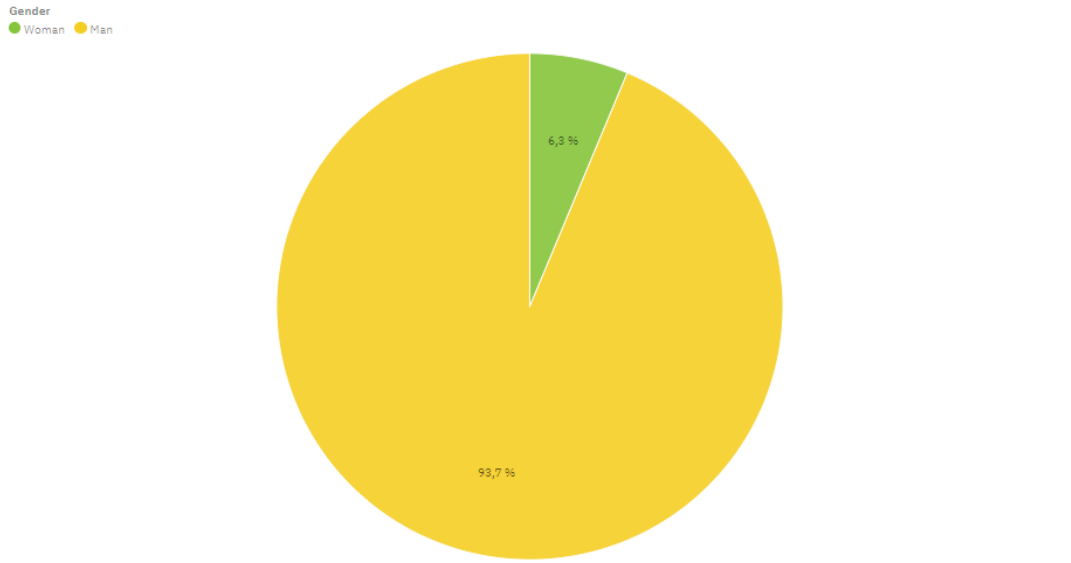
Aucun filtre n'a été appliqué.

Cet onglet

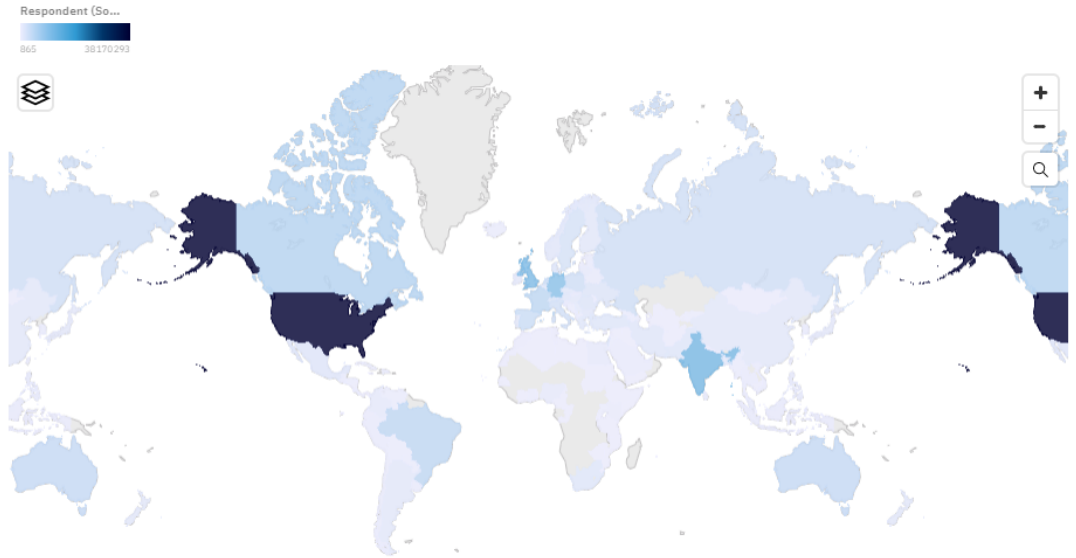
Gender
Man, Woman 2

Current Technology Usage | Future Technology Trend | Demographics

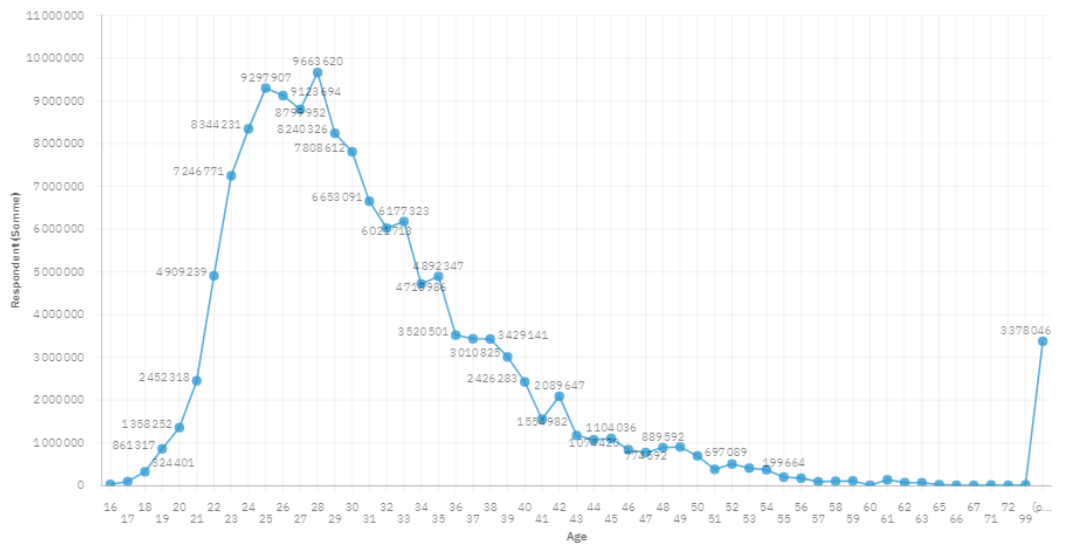
Respondent classified by Gender



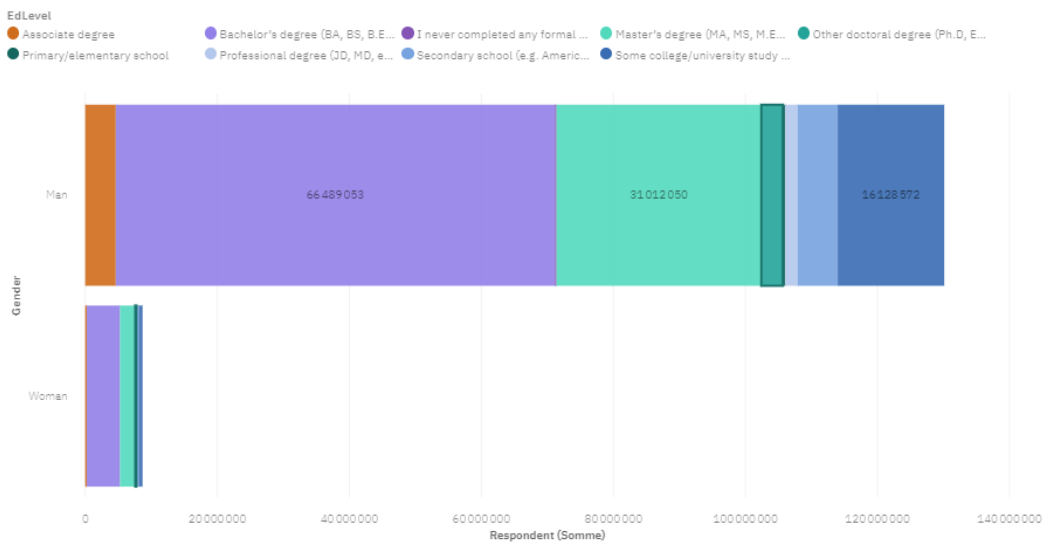
Respondent Count for Countries



Respondent Count by Age



Respondent Count by Gender, classified by Formal Education Level



IBM

powered by IBM Cloud Pak for Data

OVERALL FINDINGS & IMPLICATIONS

Findings

- JavaScript remains a leader among programming languages
- Python is leading among the languages developers want to learn next year
- MySQL is the most popular database
- PostgreSQL is the most desired database to learn next year
- Linux, Windows and Docker are the most popular platforms
- ¡Query is the most popular web frame

Implications

The most popular technologies continue to attract new users and learners, while developers are ready to learn new skills

DISCUSSION

- Do you think the survey is representative enough? Are women so under-represented in the IT sector, or do they participate in surveys? If so, why?
- Developers between 25 and 28 years old participate in the survey more readily, but should we try to collect more data about other age groups?
- Will Python overtake JavaScript next year?
- Will PostgreSQL overtake MySQL next year?

CONCLUSION

- The coding sector is very diverse and rapidly changing in terms of the technologies available
- It remains a very male-dominated field
- Many developers work with a range of different technologies
- Developers are willing to learn new languages, databases and platforms in the coming year

APPENDIX 1: CORRELATION COEFFICIENTS BETWEEN NUMBER OF LANGUAGES/DATABASES AND CONVERTED COMPENSATION

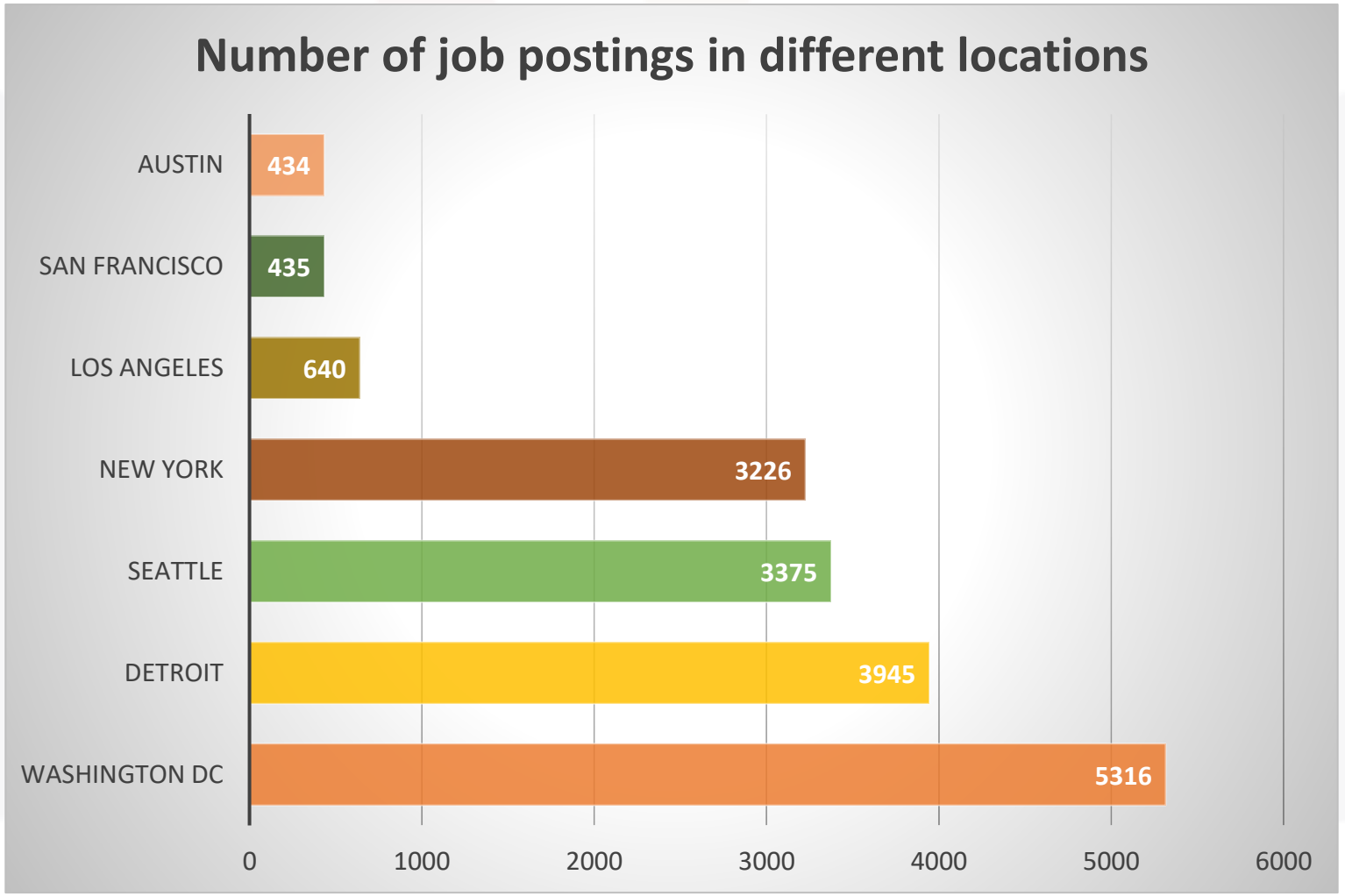
```
merged_df = pd.merge(dataframe, dataframe2, on="Respondent")
print(merged_df)
merged_df.corr()['ConvertedComp']
```

```
       Respondent  NumberOfLanguages  ConvertedComp
0             207                  1        57349.0
1             224                  1       233000.0
2             383                  1       140000.0
3             402                  1       110509.0
4             426                  1            NaN
...           ...                ...            ...
11382       10544                 17         8700.0
11383       15378                 17       200000.0
11384       21543                 17         4080.0
11385        2482                 18      1000000.0
11386       10245                 19         6563.0

[11387 rows x 3 columns]
Respondent          0.002101
NumberOfLanguages   0.022890
ConvertedComp       1.000000
Name: ConvertedComp, dtype: float64
```

```
merged_df2 = pd.merge(dataframe1, dataframe2, on="Respondent")
print(merged_df2)
merged_df2.corr()['ConvertedComp']
```

```
       Respondent  NumberOfDatabases  ConvertedComp
0              25                  1         6300.0
1              38                  1        51312.0
2              70                  1        75000.0
3              74                  1       114575.0
4              75                  1        17491.0
...           ...                ...            ...
10940       10222                 11        14300.0
10941       10982                 11        30935.0
10942       22329                 11       120000.0
10943       11365                 12       120000.0
10944       21878                 12       140000.0

[10945 rows x 3 columns]
Respondent          -0.000049
NumberOfDatabases   -0.006128
ConvertedComp        1.000000
Name: ConvertedComp, dtype: float64
```

APPENDIX 2: JOB POSTINGS



APPENDIX 3: POPULAR LANGUAGES