



ANALYZING CURRENT AND EMERGING TRENDS IN IT INDUSTRY

Tomasz Tomaszewski

17 April 2021

Table of contents



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

EXECUTIVE SUMMARY



AIM OF ANALYSIS:

- 1) Identification of trends in IT industry.
- 2) Analyzing demographics and gender gaps.

KEY FINDINGS:

- 1) Top current skills:
 - Language: Javascript, HTML/CSS, SQL,
 - Database: MySQL, Microsoft SQL Server, PostgreSQL
- 2) Top desired skills:
 - Language: Javascript, HTML/CSS, Python
 - Database: PostgreSQL, MongoDB, Redis
- 3) Emerging trends: winners are Python and PostgreSQL
- 4) Demographic data:
 - overrepresentation of male gender,
 - survey biased towards USA market.

INTRODUCTION



- Report will present up to date analysis of technologies currently used and desired in future.
- The key questions to answer are:
What is current and future top languages demand?
What is current and future top databases demand?
Does any gender dominance exist in industry?
Can other criteria affect analysis?
- Report is aimed to give insights to set up future business strategy in order to stay competitive in fast changing environment.
- Report is intended for use by : IT executives, Human Resources, IT Training Management.

METHODOLOGY

DATA COLLECTION



1) Collecting Job Data from GitHub Jobs API

Collected data stored into an excel spreadsheet.

2) Webscraping data with Python from following URL

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DA0321EN-SkillsNetwork/labs/datasets/Programming_Languages.html

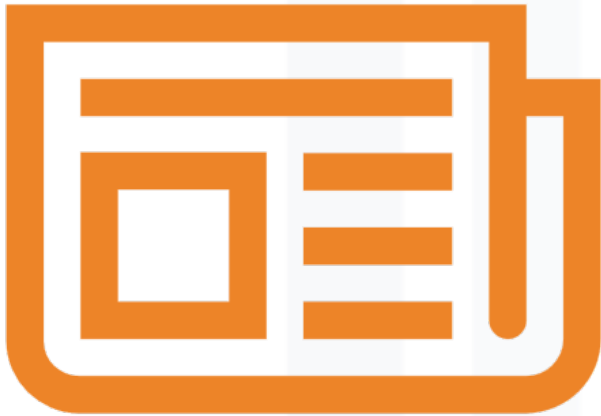
Extracting name of the programming language and average annual salary.

Collected data stored into an CSV format file.

3) Loading data set of Stack Overflow survey given by IBM in form of CSV file and RDBMS sqlite file.

METHODOLOGY

DATA EXPLORATION



Exploration of data set of Stack Overflow survey based on following link given by IBM:

```
dataset_url = "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DA0321EN-SkillsNetwork/LargeData/m1_survey_data.csv"
```

DATA WRANGLING

Identification and removal of duplicate values in the dataset.
Identification of missing values in the dataset.
Imputation the missing values in the dataset.
Normalization of data in the dataset.

METHODOLOGY



EXPLORATORY DATA ANALYSIS

Finding out the distribution of data, presence of outliers.
Determining the correlation between different columns in the dataset.

DATA VISUALISATION

Distribution of data.
Relationship between two features.
Composition of data.
Comparison of data.

RESULTS

- **Webscraping data with Python**
- Stored in CSV file.

Language ▼	Average Annual Salary USD ▼
Swift	130 801,00
Python	114 383,00
C++	113 865,00
Javascript	110 981,00
Java	101 013,00
Go	94 082,00
R	92 037,00
C#	88 726,00
SQL	84 793,00
PHP	84 727,00

- Collecting job data from GitHub Jobs API
- Stored into an excel spreadsheet.

Language ▼	Job Openings ▼
C	301
Java	132
JavaScript	102
Python	82
Scala	61
SQL Server	42
C#	26
PostgreSQL	22
C++	21
MySQL Server	16
MongoDB	16
Oracle	12

RESULTS

- Survey data exploration in Python based on dataset from IBM website.

Respondent	MainBranch	Hobbyist	OpenSourcer	OpenSource	Employment	Country	Student	EdLevel	UndergradMajor	...	WelcomeChange	SONewContent	Age	Gender	Trans	Sexuality	Ethnicity
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	Tech articles written by other developers;Indu...	22.0	Man	No	Straight / Heterosexual	White or of European descent
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	NaN	23.0	Man	No	Bisexual	White or of European descent
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	Tech articles written by other developers;Cour...	28.0	Man	No	Straight / Heterosexual	White or of European descent
3	16	I am a developer by profession	Yes	Never	The quality of OSS and closed source software ...	Employed full-time	United Kingdom	No	Master's degree (MA, MS, M.Eng., MBA, etc.)	NaN	Just as welcome now as I felt last year	Tech articles written by other developers;Indu...	26.0	Man	No	Straight / Heterosexual	White or of European descent
4	17	I am a developer by profession	Yes	Less than once a month but more than once per ...	The quality of OSS and closed source software ...	Employed full-time	Australia	No	Bachelor's degree (BA, BS, B.Eng., etc.)	Computer science, computer engineering, or sof...	Just as welcome now as I felt last year	Tech articles written by other developers;Indu...	29.0	Man	No	Straight / Heterosexual	Hispanic or Latino/Latina;Multiracial

5 rows x 18 columns

RESULTS

- Exploratory data analysis based on dataset from IBM website.

	Respondent	CompTotal	ConvertedComp	WorkWeekHrs	CodeRevHrs	Age
count	11398.000000	1.058900e+04	1.058200e+04	11276.000000	8972.000000	11111.000000
mean	12490.392437	7.570477e+05	1.315967e+05	42.064606	4.781071	30.778895
std	7235.461999	9.705598e+06	2.947865e+05	24.672741	4.567060	7.393686
min	4.000000	0.000000e+00	0.000000e+00	3.000000	0.000000	16.000000
25%	6264.250000	2.500000e+04	2.686800e+04	40.000000	2.000000	25.000000
50%	12484.000000	6.500000e+04	5.774500e+04	40.000000	4.000000	29.000000
75%	18784.750000	1.200000e+05	1.000000e+05	43.000000	5.000000	35.000000
max	25142.000000	7.000000e+08	2.000000e+06	1012.000000	99.000000	99.000000

RESULTS

- **Visualisation** based on data set presented in the form of a RDBMS.
- SQL queries used to extract the data.

```
In [183]: # getting table 'DatabaseDesireNextYear' into dataframe
QUERY = """
SELECT DatabaseDesireNextYear, COUNT(*) AS ValueCount FROM DatabaseDesireNextYear GROUP BY DatabaseDesireNextYear ORDER BY ValueCount DESC LIMIT 5;
"""

#SELECT working_area, COUNT(*) FROM agents GROUP BY working_area;
# the read_sql_query runs the sql query and returns the data as a dataframe
df3= pd.read_sql_query(QUERY,conn)
df3
```

Out[183]:

	DatabaseDesireNextYear	ValueCount
0	PostgreSQL	4328
1	MongoDB	3649
2	Redis	3331
3	MySQL	3281
4	Elasticsearch	2856

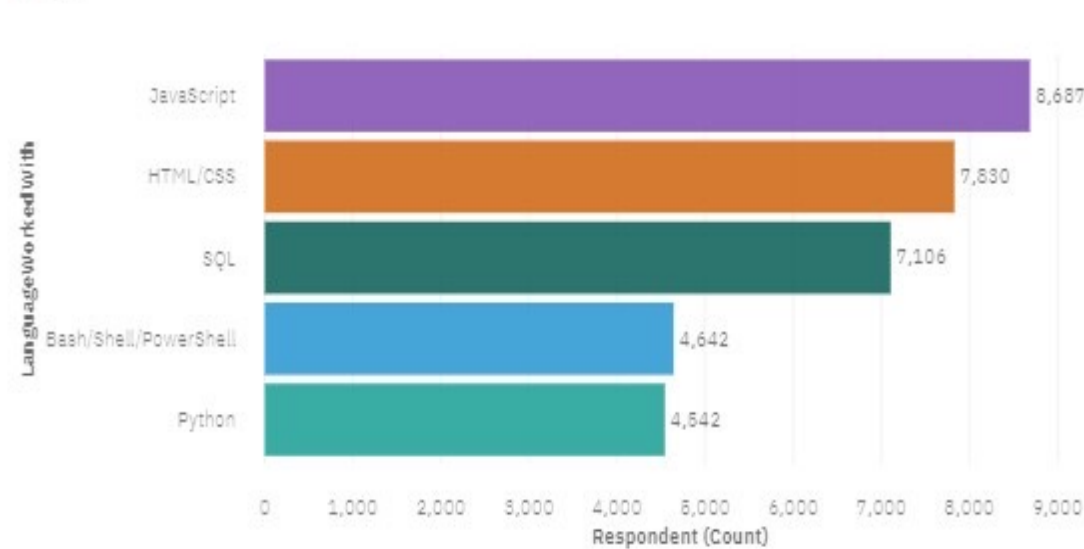
PROGRAMMING LANGUAGE TRENDS

Current Year

Top 5 Programming Languages Current Year

Top 10 Programming Languages Current Year

● Bash/Shell/PowerShell ● HTML/CSS ● JavaScript ● Python ● SQL

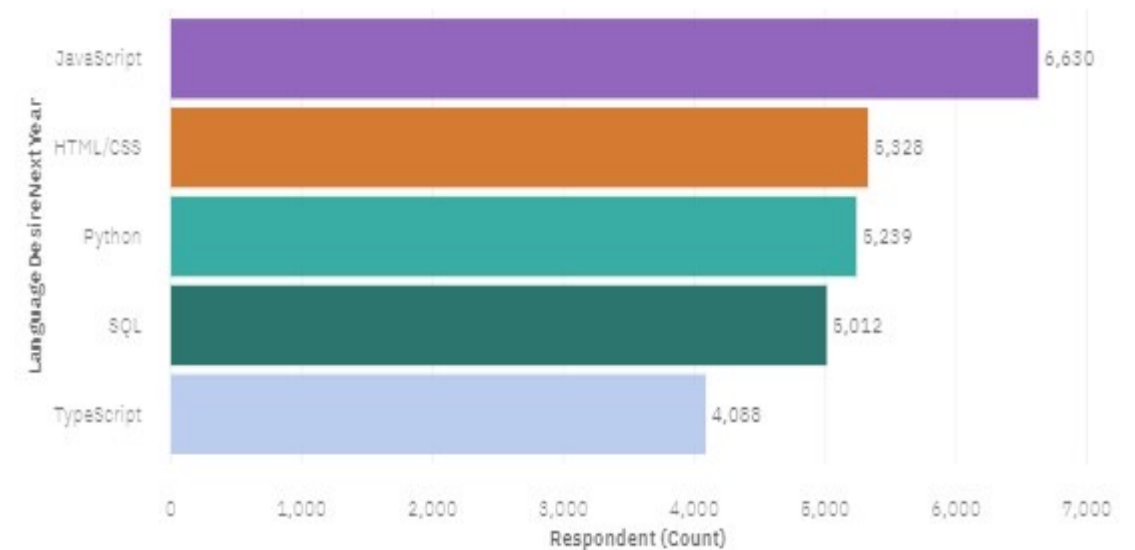


Next Year

Top 5 Programming Languages Next Year

Language Desire Next Year

● HTML/CSS ● JavaScript ● Python ● SQL ● TypeScript



PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

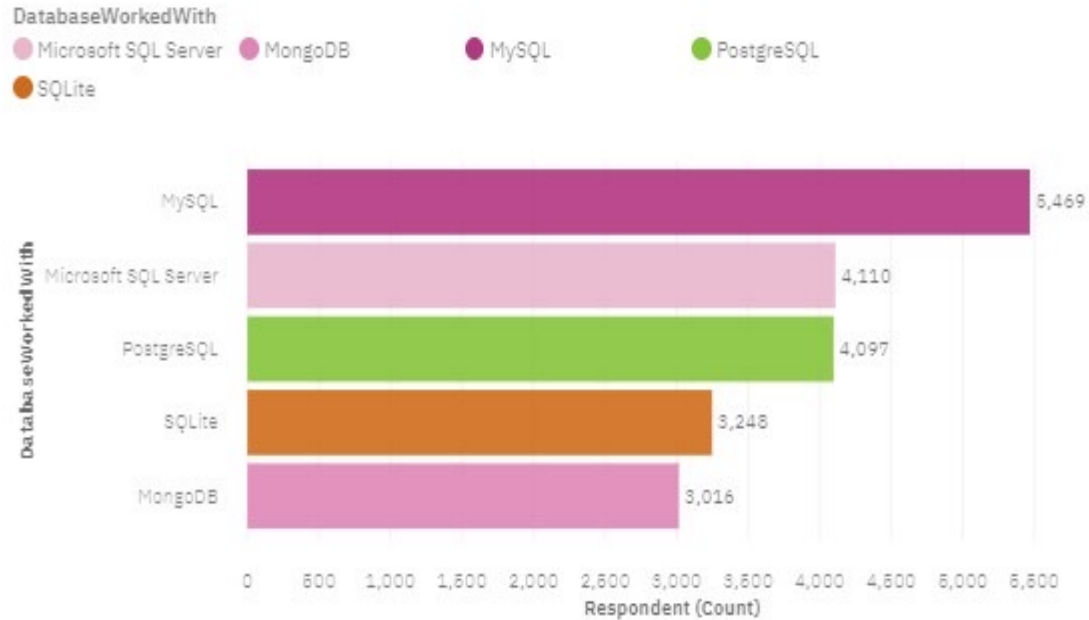
- Javascript, HTML/CSS and SQL are currently most used programming languages.
- JavaScript and HTML/CSS will remain the most popular language next year.
- Python grows in popularity and becomes 3rd most desired language to learn.
- Popularity of languages is highly diversified.

Implications

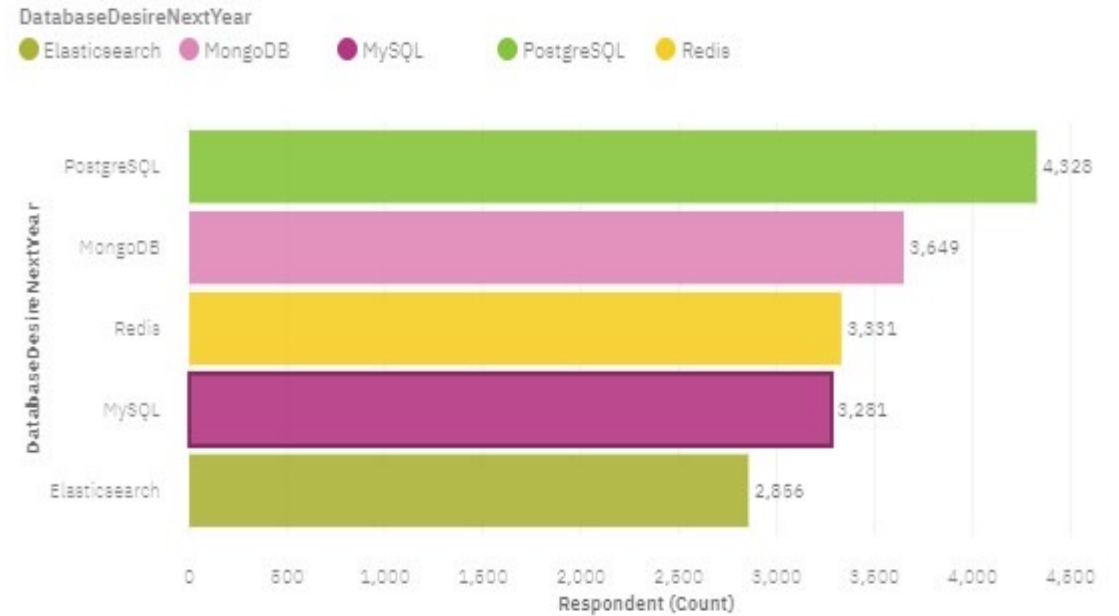
- JavaScript and HTML/CSS skills should still be in high priority for recruiting new personnel.
- If Python growing in popularity continues, this language might become top desired language in future.
- Hiring and training professionals fluent in python will help company stay up to date with market trends.
- Python could be a good choice as it is often used in Artificial Intelligence and Machine Learning Technologies which are in rising demand.

DATABASE TRENDS

Top 5 Databases Worked Current Year



Top 5 Database Desire Next Year



DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- MySQL is most popular database.
- Most of the Top 5 databases currently used are SQL type databases.
- PostgreSQL stays high both in currently used databases in use and database desire next year.
- Popularity of databases is diversified and in constant change.

Implications

- NoSQL databases are growing in popularity which can impact current company database systems.
- Adaptation to changing environment should be considered.
- Training employees in rising technologies should be prioritized.
- SQL database will be less in demand in future if trend continues.

DASHBOARD



Dashboard contains interactive visualization of current and desired technology usage, together with brief overview of demographics. Data set based on stack overflow survey.

[Link to dashboard](#)

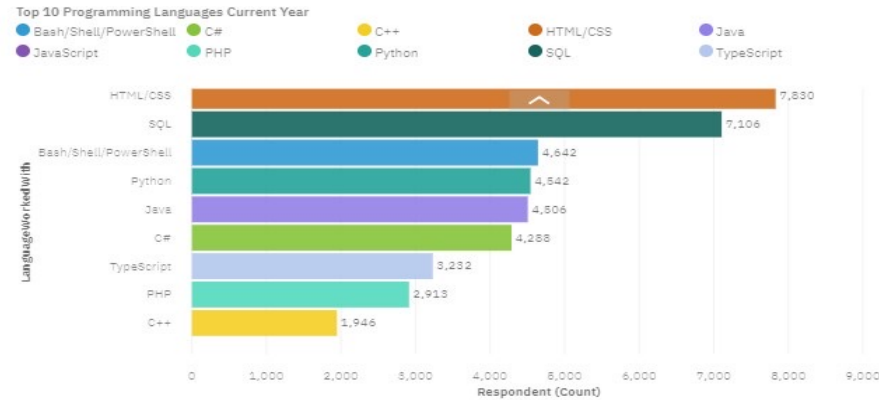
In case link is not directing you to dashboard please copy following url:

<https://eu-gb.dataplatform.cloud.ibm.com/dashboards/299d2960-8417-448e-8c9d-9c4b1f7f57c9/view/4c65c73a24e908ca68f1d4e407c8780f7a33240eb4bb835182d47b4906317497a9691494c82a1d0cd9435730a0e8140c9d>

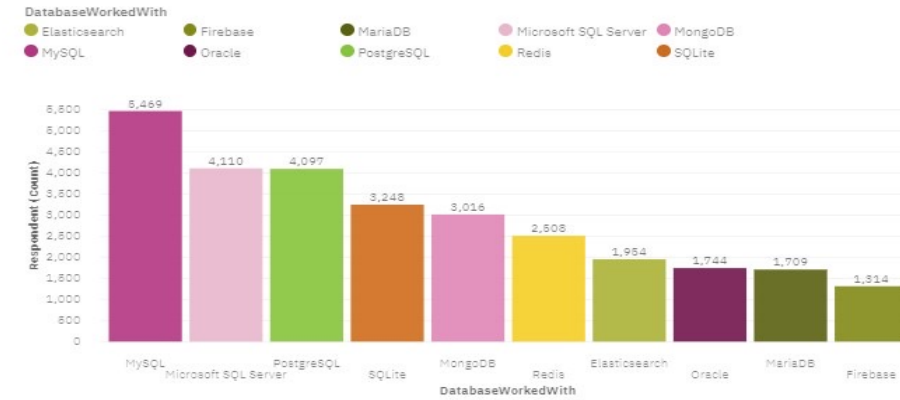
DASHBOARD TAB 1 – Current Technology Usage

Current Technology Usage Future Technology Trend Demographics

Top 10 Programming Languages Current Year



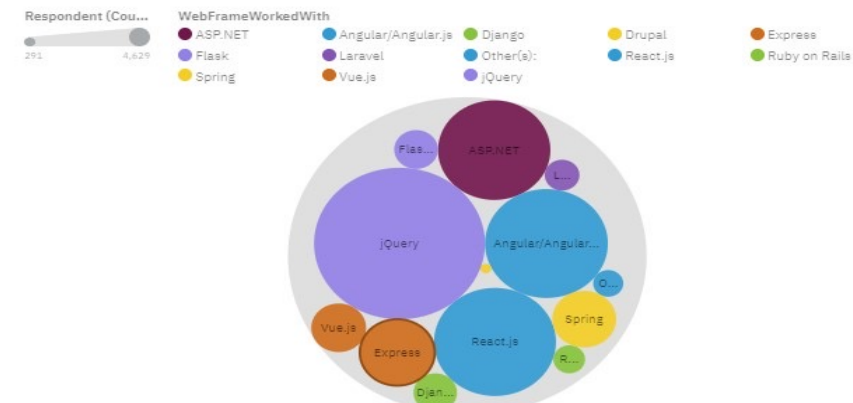
Top 10 Databases Worked Current Year



Platforms Worked With Current Year



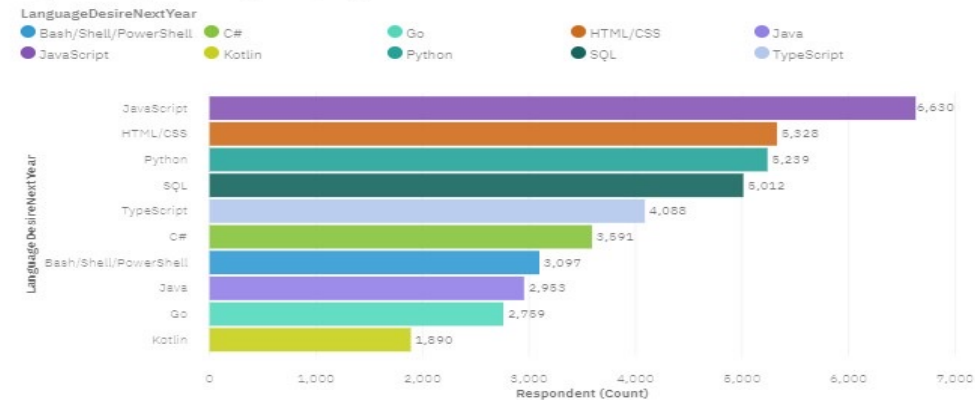
Top 10 Web Framework Current Year



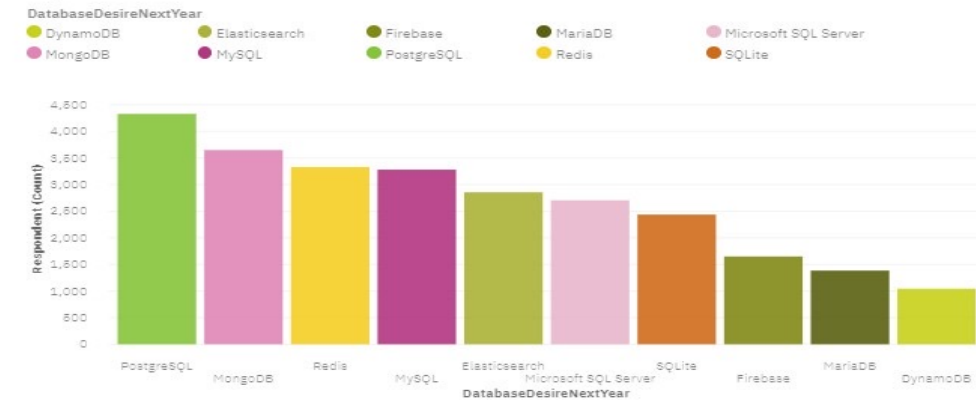
DASHBOARD TAB 2 – Future Technology Trend

Current Technology Usage Future Technology Trend Demographics +

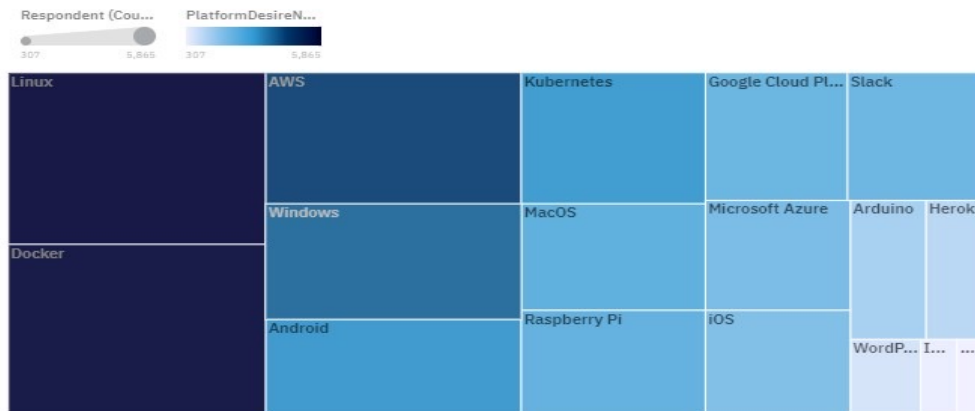
Top 10 Programming Languages Next Year



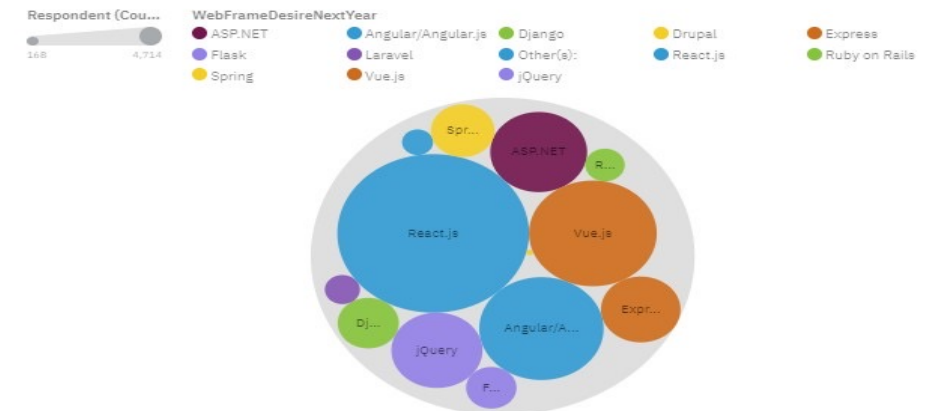
Top 10 Database Desire Next Year



Platform Desire to Learn Next Year



Top 10 Web Frame Desire Next Year



DASHBOARD TAB 3 - Demographics

Current Technology Usage

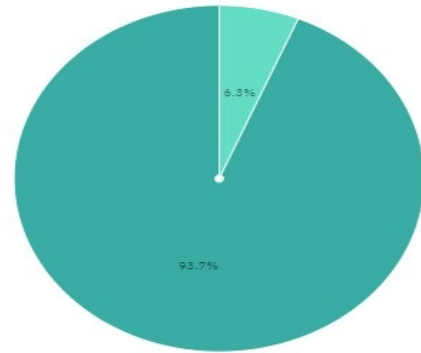
Future Technology Trend

Demographics



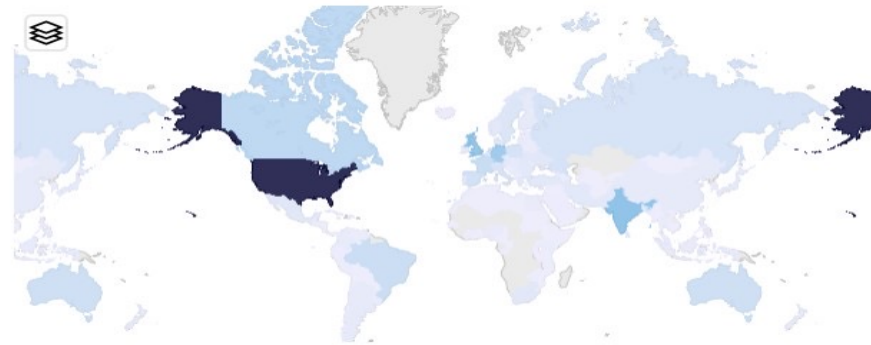
Respondent Classified By Gender

Gender
Woman Man

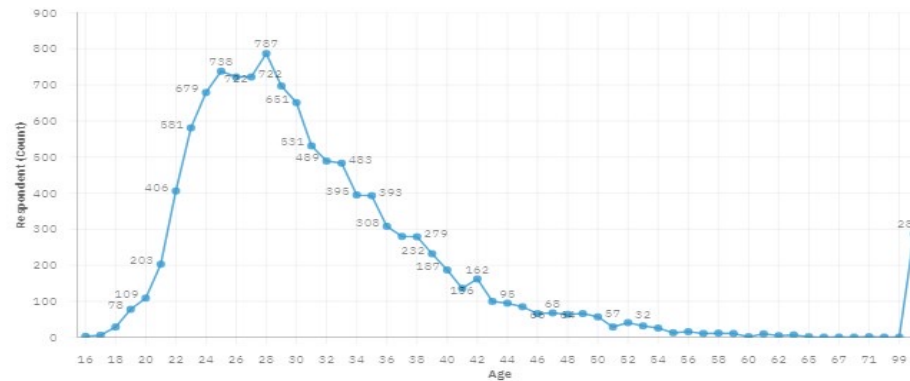


Respondent Count for Countries

Respondent (Cou...
1 3,127

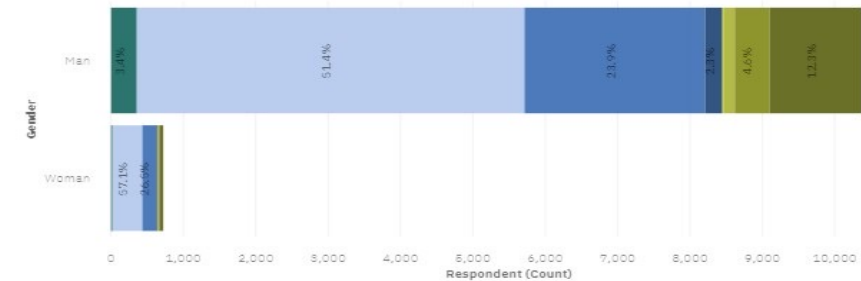


Respondent Count by Age



Respondent Count by Gender Classified by Education Level

EdLevel
Associate degree Bachelor's degree (BA, BS, B.E... I never completed any formal ...
Master's degree (MA, MS, M.E... Other doctoral degree (Ph.D, E... Primary/elementary school
Professional degree (JD, MD, e... Secondary school (e.g. Americ... Some college/university study ...



DISCUSSION



- What can be other sources of data?
- What is the correlation between data from survey (age, education level, working hours, salary etc).
- Most respondents were from USA. Does it affect analysis?
- Should we worry about under representation of women in IT?
- Are you surprised with results?

OVERALL FINDINGS & IMPLICATIONS

Findings

- JavaScript and HTML/CSS are popular and this popularity should stay stable in near future.
- PostgreSQL database is popular and rising in demand.
- Changes in trends visible, especially in the field of databases technologies.
- Fast changing technology calls for flexibility and quick reaction.

Implications

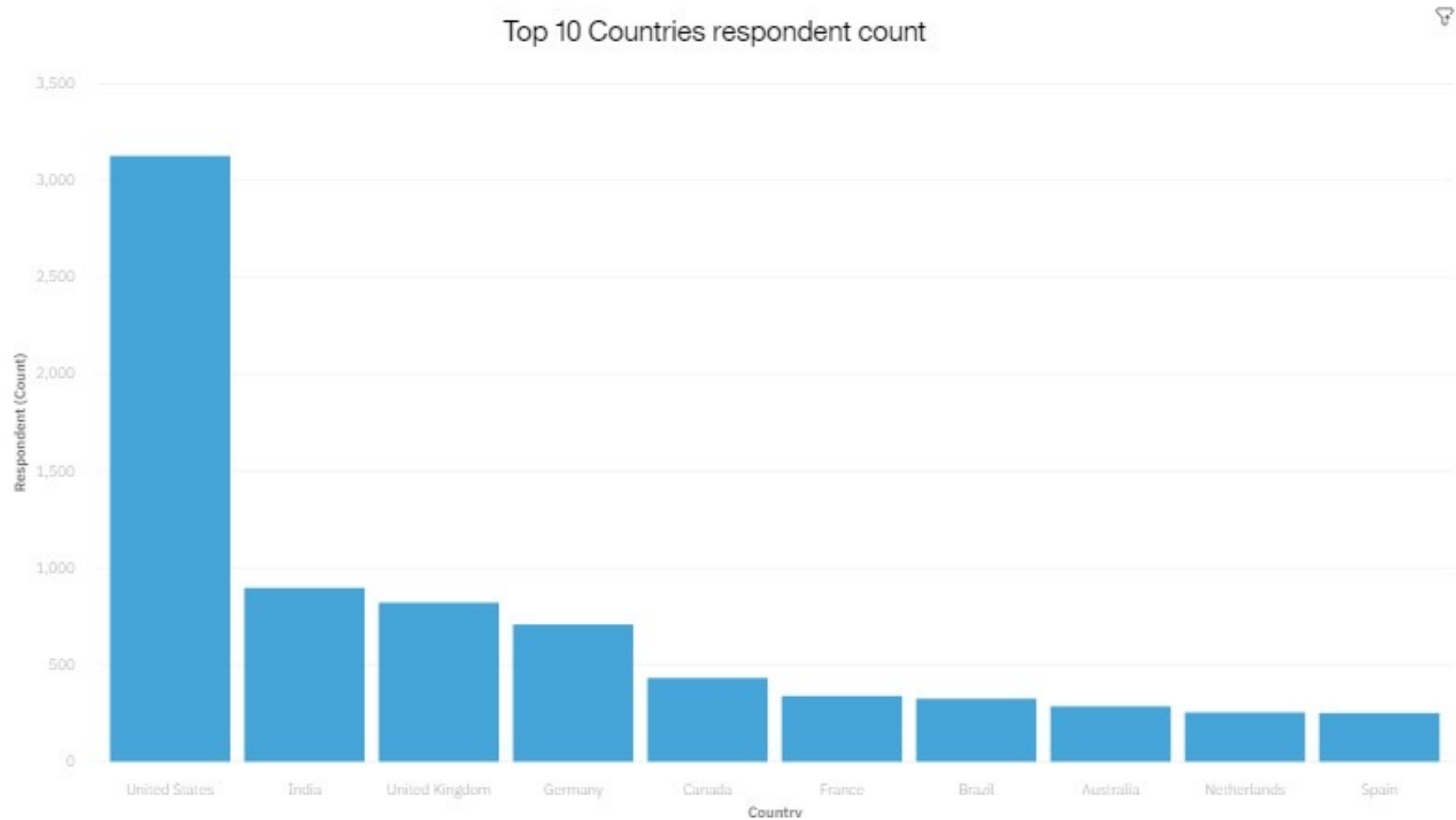
- JavaScript and HTML/CSS should remain priority while we should research for usage of other languages.
- Focus on PostgreSQL while doing research for other databases.
- Adaptation to changing trends required in order to stay competitive.
- Impact on training and hiring.

CONCLUSION



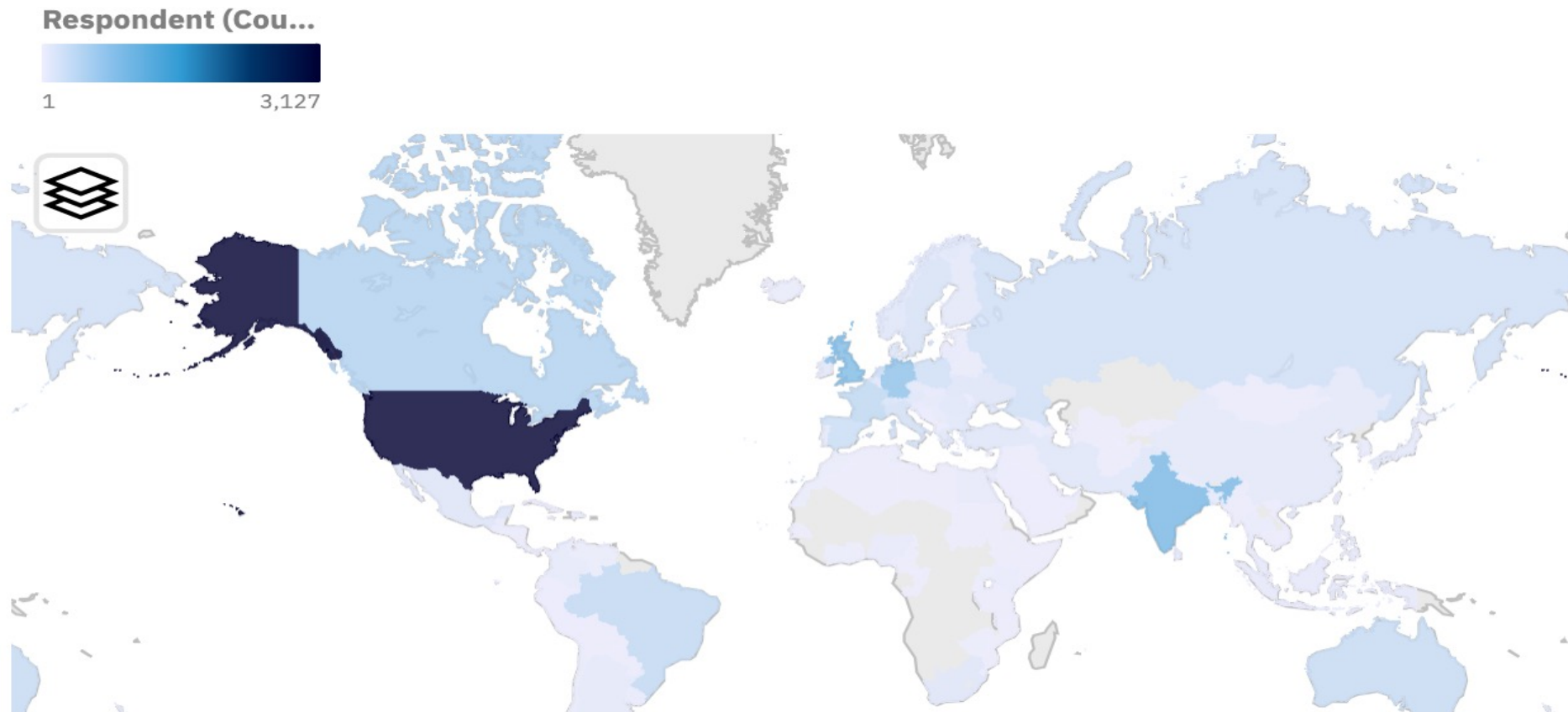
- Top languages to focus on: Javascript, HTML/CSS, Python .
- Top databases to focus on: PostgreSQL, and rising in demand databases.
- Technology popular currently can be taken over by other technology in rising demand.
- Adaptation to fast moving and fast changing environment required.
- Constant analysis and strategy adjustment vital to be in the run in this competitive and demanding industry.
- Consideration regarding gender balancing.
- Required research and updates regarding incorporation of other technologies like Machine Learning or Artificial Intelligence.

APPENDIX

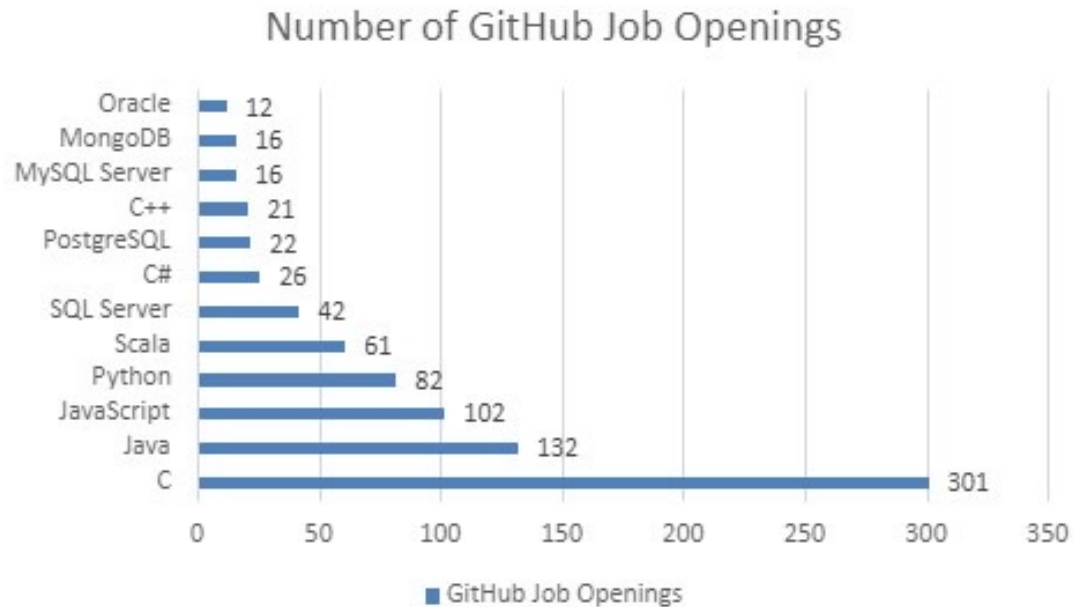


APPENDIX

Respondent Count for Countries



GITHUB JOB POSTINGS



- GitHub job postings are highest for C and Java.
- C and Java are not in Top 5 languages neither for current year or as desired skill for next years.

POPULAR LANGUAGES



Software professionals working with Swift and Python receive highest gratification.

Swift and C++ are not the top choice programming languages neither current or next year.

Python on the other hand is one of the most in demand languages and among of most well-paid.