Software Developers Survey Analysis

Presented by: D Tadesse

BSc Mathematics & Statistics, https://github.com/datadesse

Last updated: 06 Aug 2023

London

TABLE OF CONTENTS

- **Executive Summary**
- **Introduction**
- **Methodology**
- **Data Collection & Process**
- **Results**
  - **Visualization – Charts & tables**
  - **Dashboard**
- **Discussion**
  - **Findings & Implications**
- **Conclusion**
- **Appendix**

EXECUTIVE SUMMARY

In this analysis I will provide evidences & suggestions to students where to work, which programing languages and database to study to remain competitive in the job market.

My research will answer the following questions:

- Why **TypeScript and Python** are popular to study multiple languages & **Python is best** to study one language

- Why **Elasticsearch & Redis** are popular databases and recommended to study

- Why I suggest to work in some selected **countries**

INTRODUCTION

<u>Objective of this analysis is:</u>

To identify programming languages and databases, job opportunities & make suggestions to students

<u>Problems to:</u>

1. Identify most popular programming languages, databases, average annual salary, countries where job opportunities available then present the result on dashboard

2. Suggest what to study to remain competitive in the job market and where to work

METHODOLOGY

Data source:

1. **Demographic & technologies CSV files** from the IBM website **to answer problem 1**

    • Actual data set was **90000 rows** from **Stack Over** only **11552 rows** of randomised data provided by IBM.

2. Extract names and average annual salary of programming languages from IBM website **to answer problem 1 & 2**

3. Extract programming languages job posting data from IBM website using Jobs API **to answer problem 1 & 2**

    • **Kaggle.com** was the **original source** of the data in **CSV format** but was converted into **Json** format **by IBM**

DATA COLLECTION & PROCESS

Demographic & Technologies csv files downloaded from IBM website then filtered out top 10 most popular Languages & Databases

Web scrapping: Python's **BeautifulSoup**, **Requests, Pandas & Json** libraries used to scrape job data using Jobs API , names and average annual salary of programming language from IBM website, filtered out & converted the required data then created charts, tables and dashboards
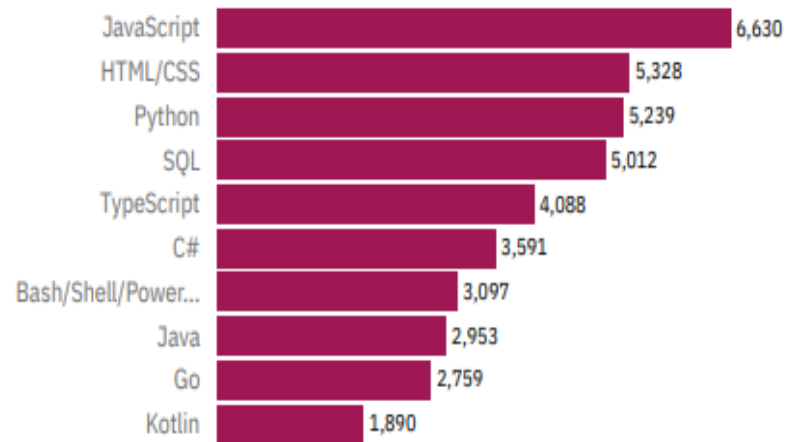
PROGRAMMING LANGUAGES TRENDS

Fig 1

Top 10 Languages Current Year
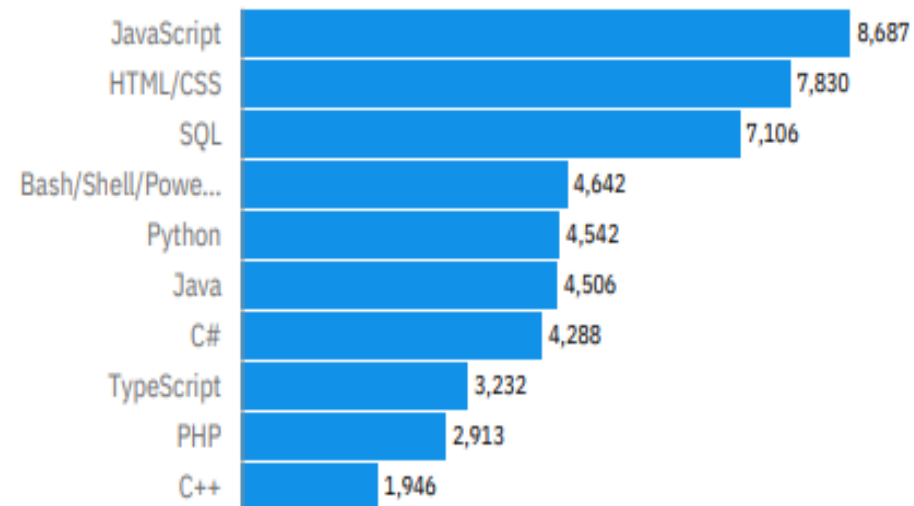
Top 10 Languages Next Year

Fig 1 indicates that the demand for **Python** & **TypeScript** increased suggesting that students would benefit if they study these languages

PROGRAMMING LANGUAGES COMPARISON

Fig 2

languages	LanguageWorkedWith	LanguageDesireNextYear	differences	pct_change
TypeScript	3232	4088.0	856.0	26.0%
Python	4542	5239.0	697.0	15.0%
PHP	2913	0.0	0.0	-100.0%
C++	1946	0.0	0.0	-100.0%
C#	4288	3591.0	-697.0	-16.0%
Bash/Shell/PowerShell	4642	3097.0	-1545.0	-33.0%
Java	4506	2953.0	-1553.0	-34.0%
JavaScript	8687	6630.0	-2057.0	-24.0%
SQL	7106	5012.0	-2094.0	-29.0%
HTML/CSS	7830	5328.0	-2502.0	-32.0%

Fig 2 indicates that the demand for **TypeScript** and **Python** increased by **26% and 15%** respectively suggesting that **their popularity grew**

PROGRAMMING LANGUAGE–FINDINGS & IMPLICATIONS

Findings

- Despite **TypeScript** has low take up grew by **26%**

- **Interest for Python** increased by **15%**

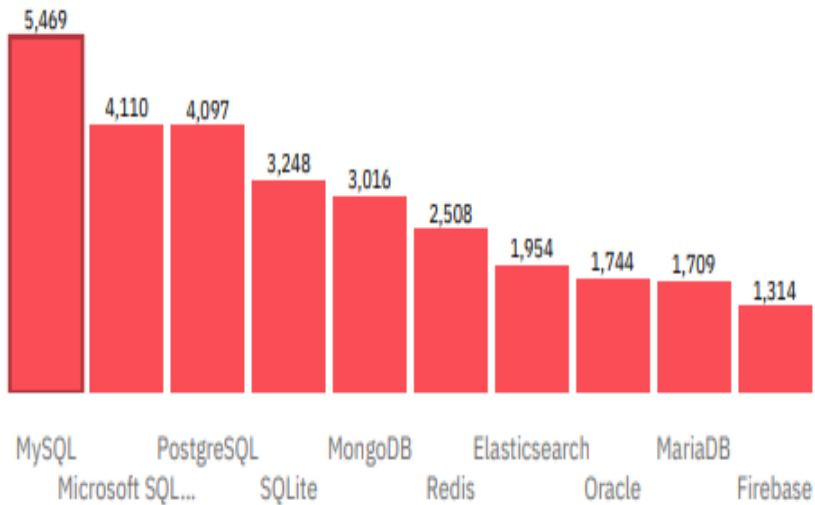- **HTML/CSS & JavaScript** have high take up **but** declined by **-32%** **and -24% respectively**

Implications

- **TypeScript and Python** become very popular

- **HTML/CSS & JavaScript** still dominate but their futurity is in question

DATABASE TRENDS

Fig 3

Top 10 Database Current Year
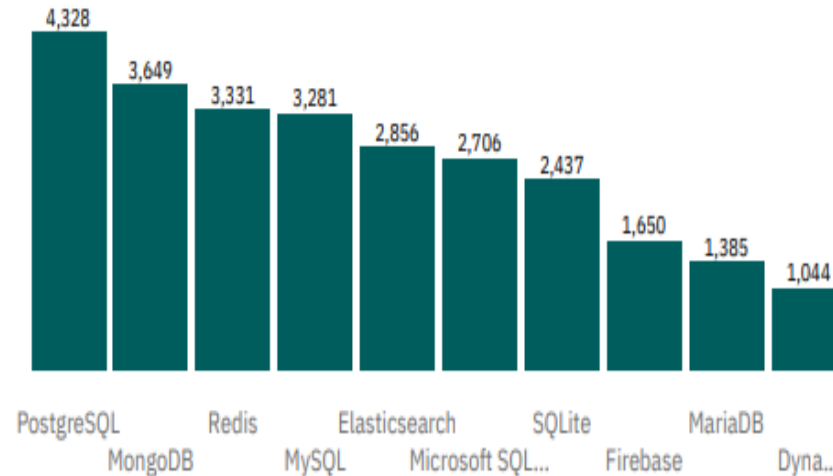

Top 10 Database Next Year

Fig 3 shows that growing interest for **Elasticsearch, Redis, Firebase, MongoDB, PostgreSQL increased** suggesting that students would benefit if they study one or two of these databases

DATABASE COMPARISON

Fig 4

databases	DatabaseWorkedWith	DatabaseDesireNextYear	differences	pct_change
Elasticsearch	1954	2856.0	902.0	46.0%
Redis	2508	3331.0	823.0	33.0%
MongoDB	3016	3649.0	633.0	21.0%
Firebase	1314	1650.0	336.0	26.0%
PostgreSQL	4097	4328.0	231.0	6.0%
MariaDB	1709	1385.0	-324.0	-19.0%
SQLite	3248	2437.0	-811.0	-25.0%
Microsoft SQL Server	4110	2706.0	-1404.0	-34.0%
Oracle	1744	0.0	-1744.0	-100.0%
MySQL	5469	3281.0	-2188.0	-40.0%

Fig 4 indicates that:

- Interest for **Elasticsearch, Redis, Firebase, MongoDB, PostgreSQL** increased

- **Elasticsearch** increased by a significant amount 46% suggesting that its **popularity grew**

DATABASE FINDINGS & IMPLICATIONS

Findings

- The interest for **Elasticsearch** & **Redis** increased by a

  significant amount 46% & 33% respectively

- **Oracle and MySQL** declined by huge amounts

  -**100% and -40%** respectively despite high take up

Implications

- **Elasticsearch & Redis** are most popular databases

- **Oracle and MySQL** dropping by a significant

  amount indication of their future uncertainty**?**

JOB ADVERT Vs LANGUAGES COMPARISON

Fig 5

	languages	total_jobs_posted	Top_Languages_Current_Year
0	C	13498	1578.0
3	Java	2609	4506.0
5	Python	1173	4542.0
7	Oracle	784	0.0
4	JavaScript	355	8687.0
1	C#	333	4288.0
2	C++	305	1946.0
8	SQL Server	250	0.0
11	MongoDB	174	0.0
6	Scala	33	0.0
10	PostgreSQL	10	0.0
9	MySQL Server	0	0.0

Based on **Fig 5** findings **C, Java, Python & JavaScript** have higher job vacancies and demand by employers
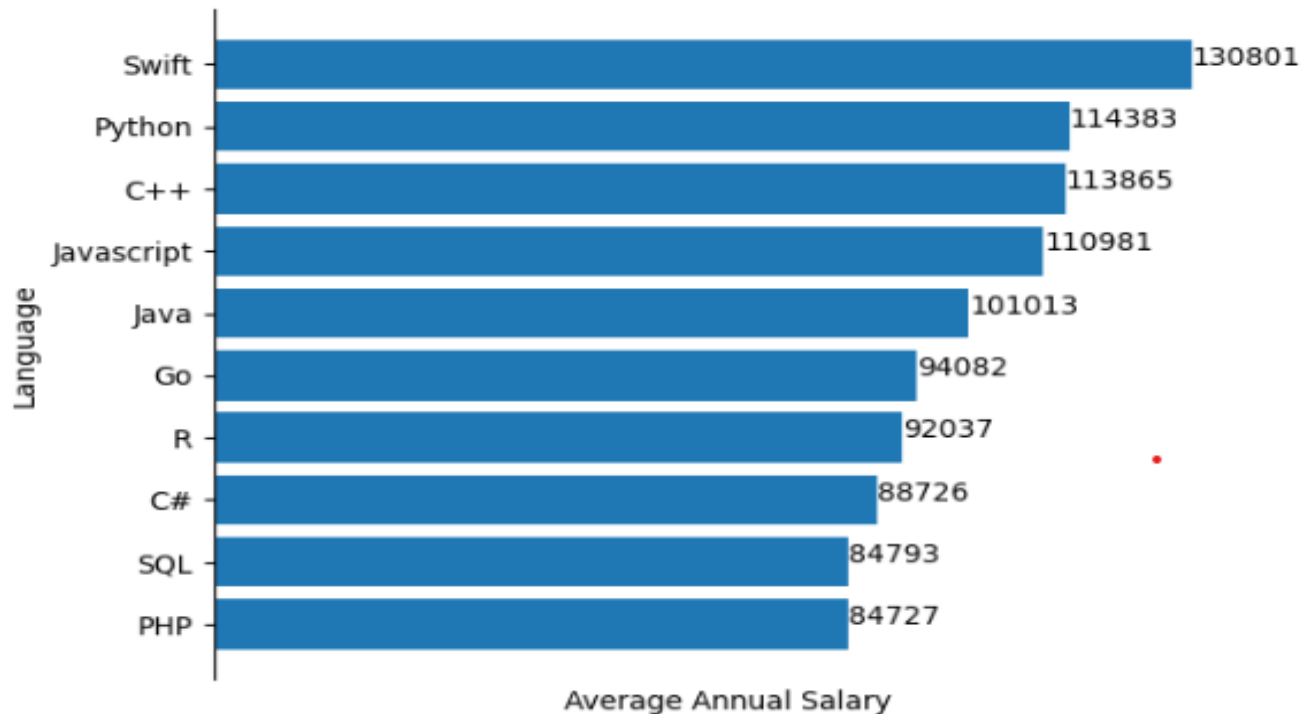
Job Postings by Number of Jobs

Fig 6
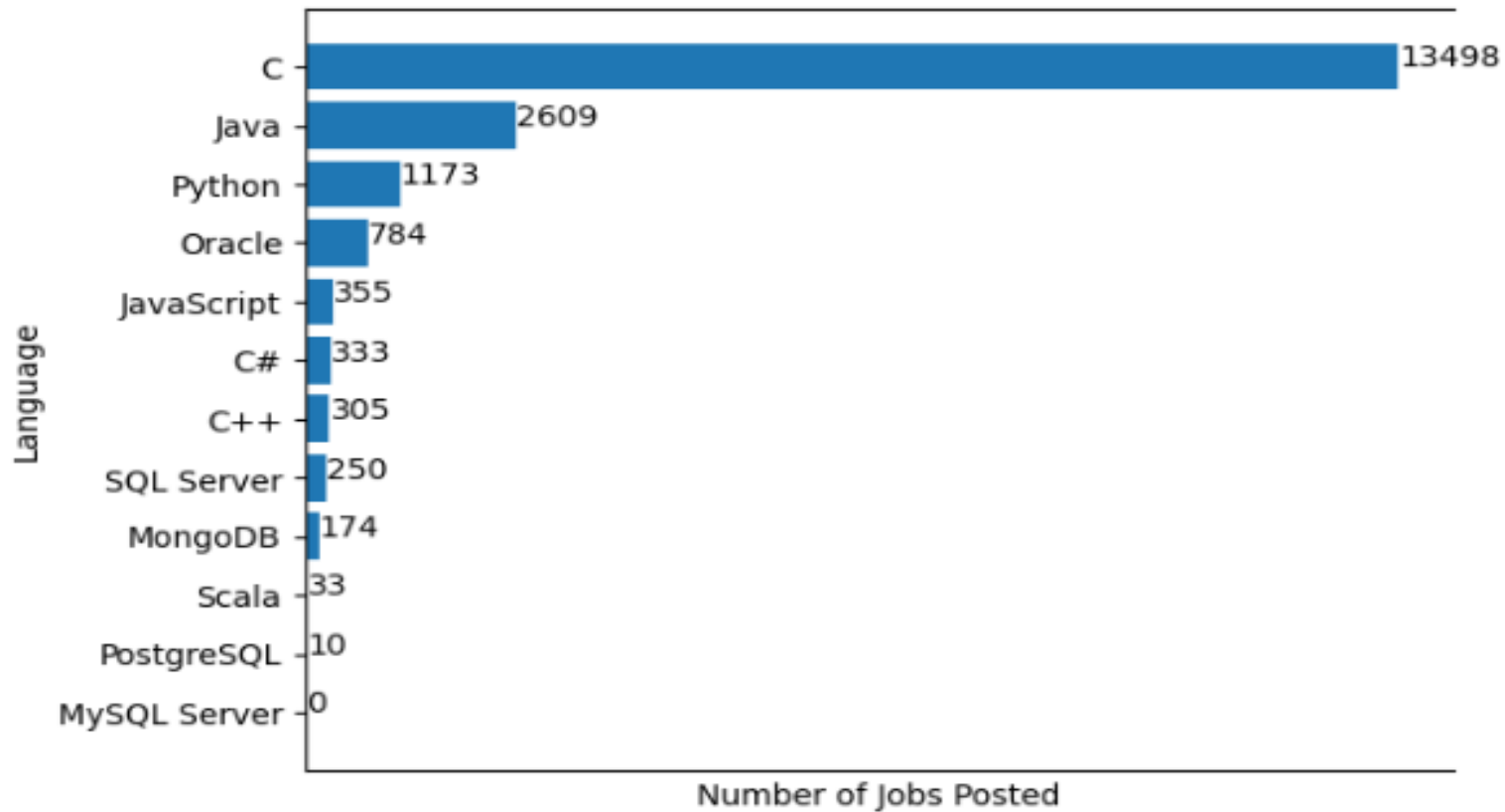
Fig 6 indicates that **C, Java, Python & JavaScript** programming languages have more job vacancies

Languages by Descending Order of Salary

Fig 7

Fig 7 shows that **Swift, Python, C++, JavaScript and Java** are **well paid** languages with **salaries over 100K**.
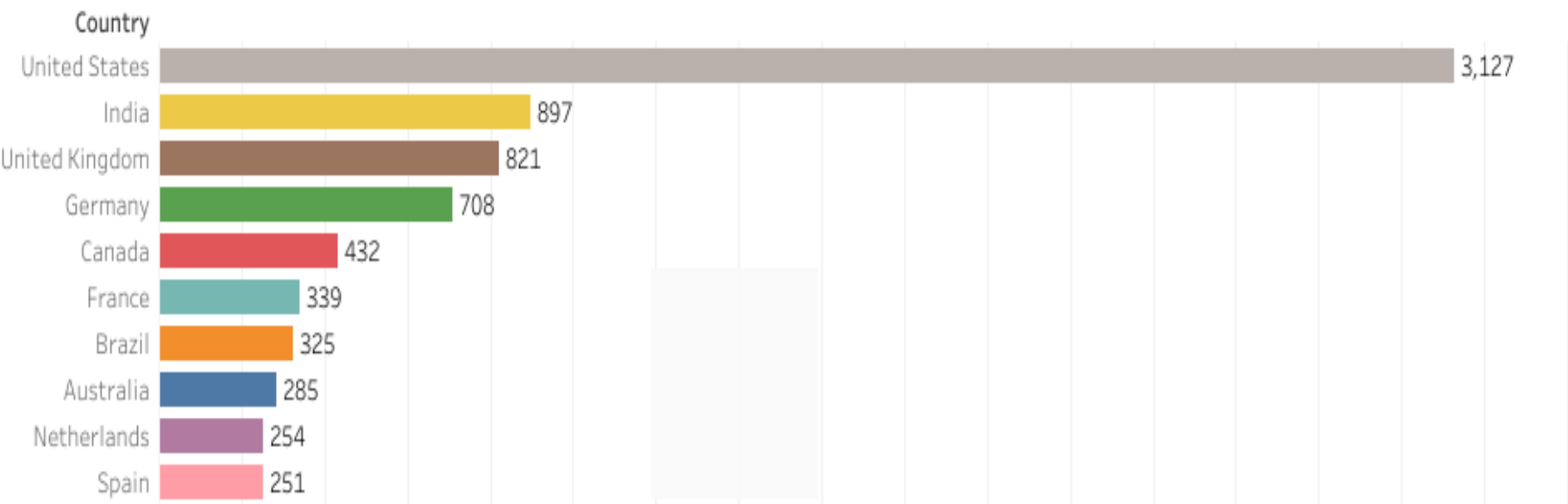
RESPONDENT COUNT BY COUNTRY

Fig 8

Country

Country	Count
United States	3,127
India	897
United Kingdom	821
Germany	708
Canada	432
France	339
Brazil	325
Australia	285
Netherlands	254
Spain	251

Fig 8 indicates that the **US, India, UK & Germany have** highest number participant indicating that **students would** have higher job opportunities if they worked in the these countries
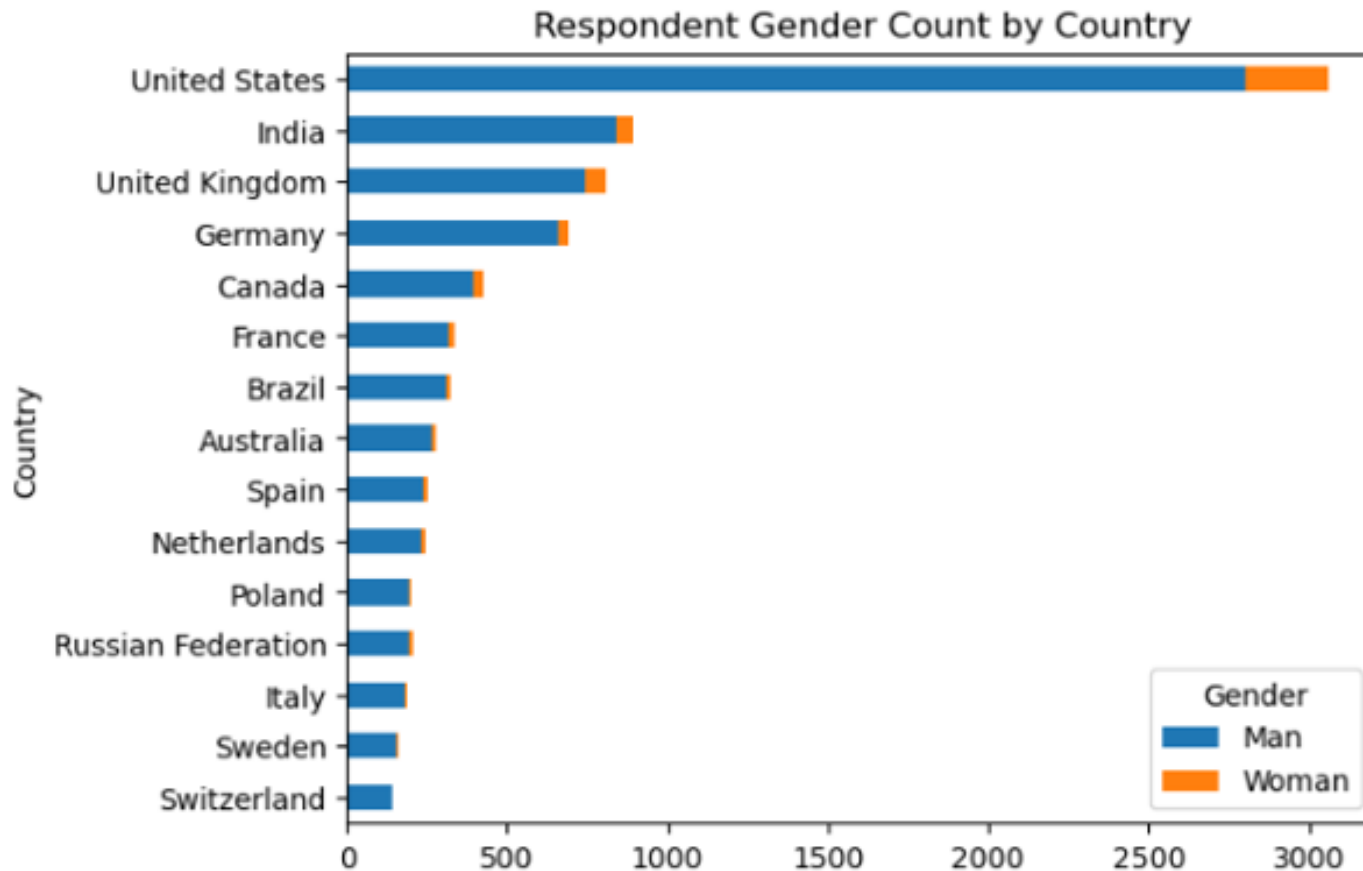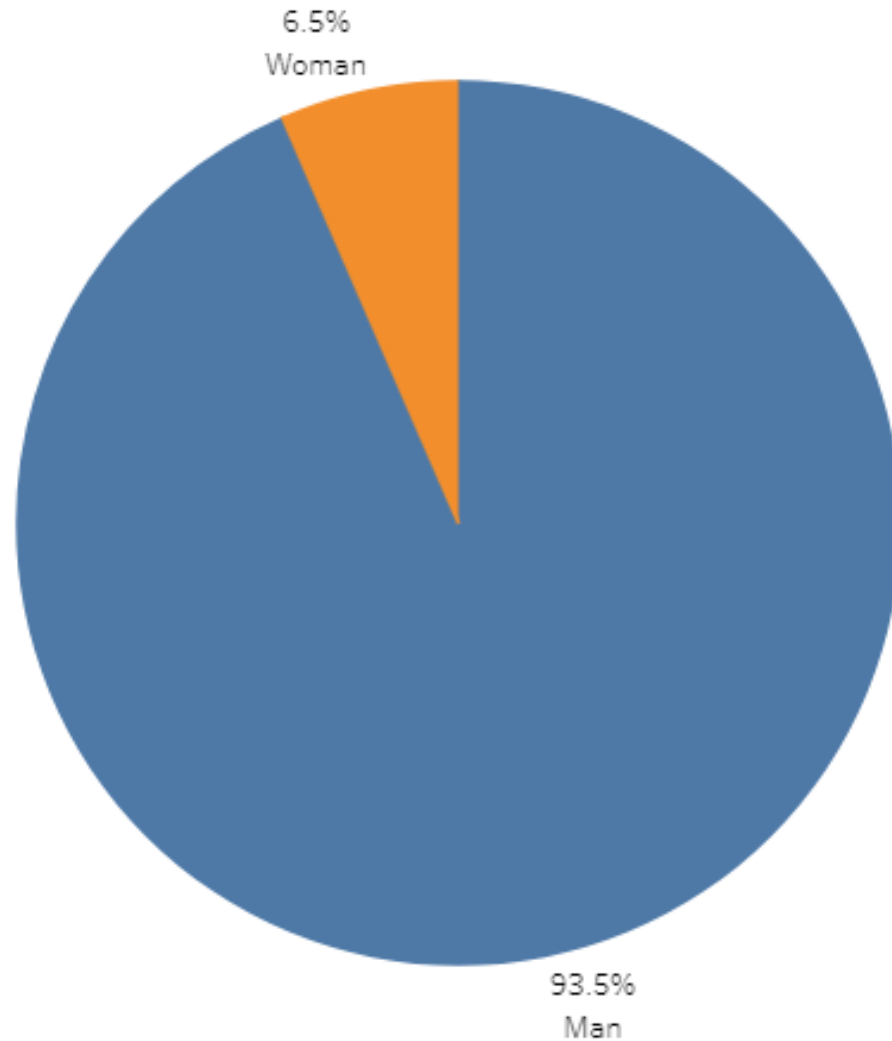
GENDER COUNT BY COUNTRY

Fig 9

Respondent Gender Count by Country

Fig 9 shows that **women have** significantly

lower participation in the survey **than men**.

RESPONDENTS COUNT BY GENDER

Fig 10

6.5%
Woman

93.5%
Man

Fig 10 shows that the number of **women 731 (6.5%)** who took part in the survey was **significantly lower** than **men 10480 (93.5%).** This result indicates that the field is **dominated by men**.

DASHBOARD

I was unable to generate Cognos dashboard link due to my IBM account issues. Alternatively I created similar charts using **Python on Jupyter Notebook and Dashboards using Tableau Public.**

My codes & dashboards for this analysis can be found on my GitHub page link below:

https://github.com/datadesse/IBM_Data_Analyst_Capstone_Project/tree/main

COGNOS DASHBOARD

Fig 11

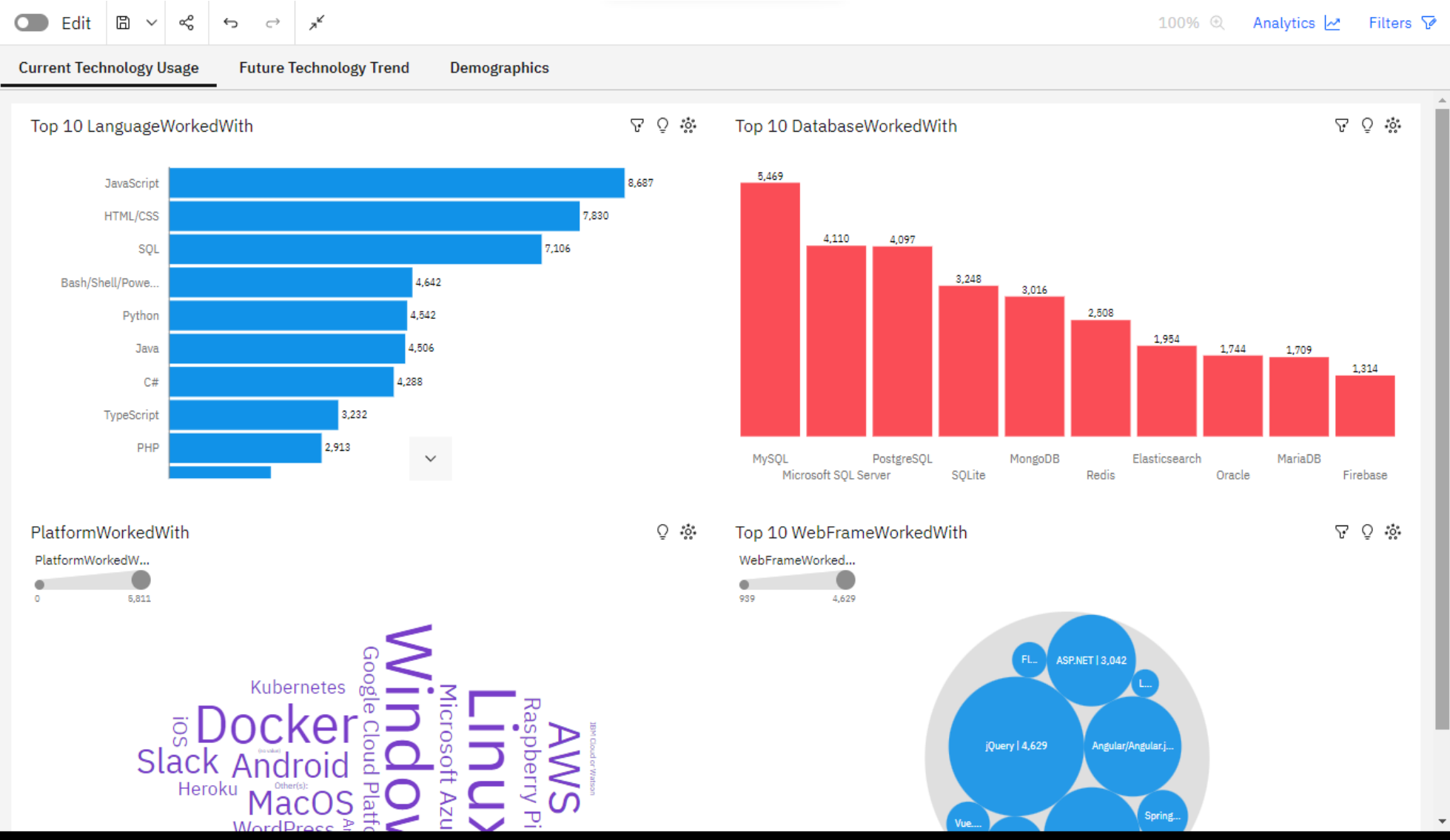**Fig 11** shows Databases, Languages, Platform and Web Frame of the Next Year using Congos Dashboard

COGNOS DASHBOARD

Fig 12
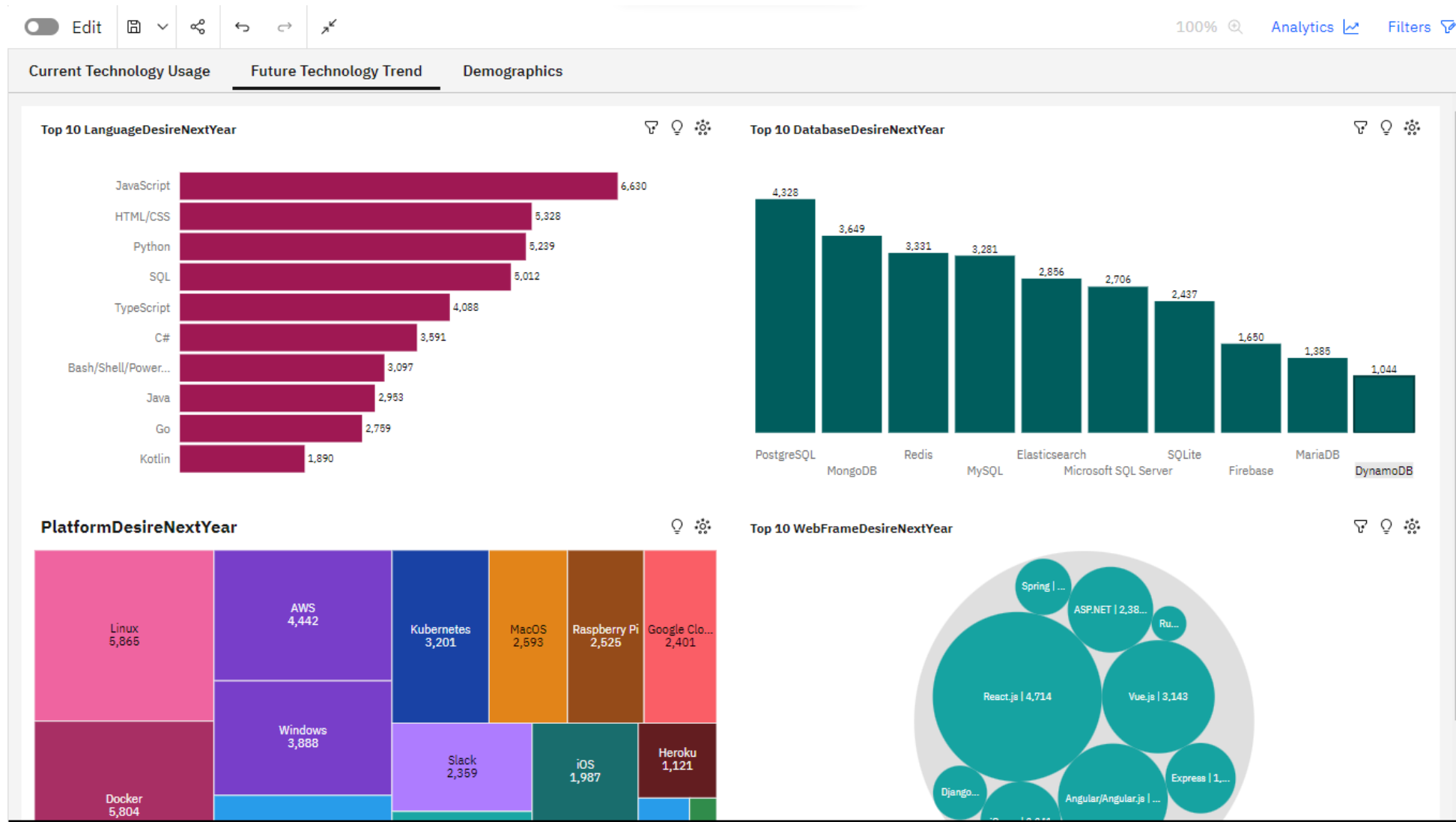
Fig 12 shows

Databases, Languages,

Platform and Web

Frame Current Year

COGNOS DASHBOARD

Fig 13
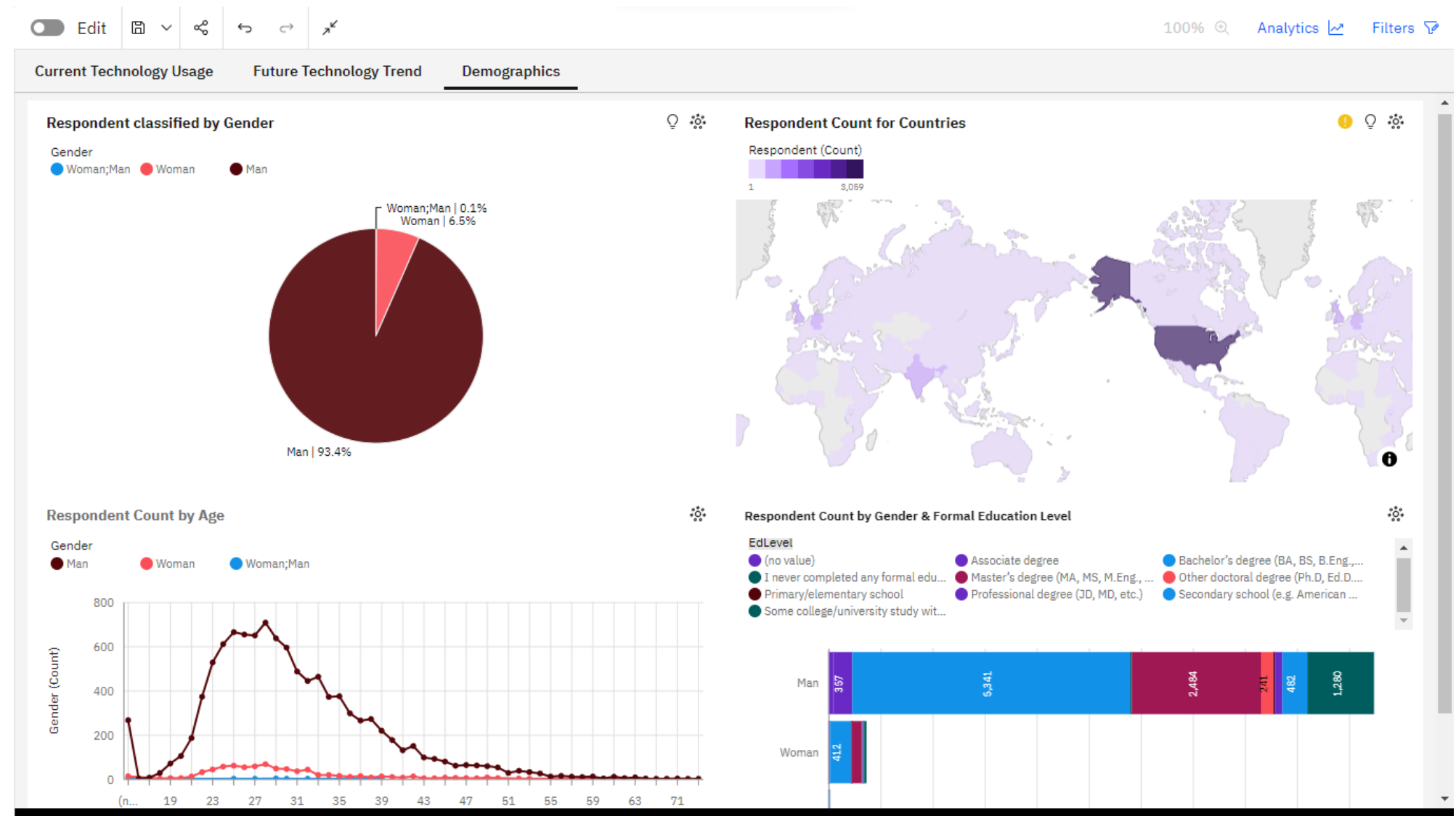
Fig 13 shows demography of the respondents'- Gender, Age, Education Level and Country using Congos Dashboard

TABLEAU DASHBOARD

Fig 14

Demographics

Respondent Count by Countries

Gender count by Education Level

Respondent Count by Gender

Gender

Man — 1,280 | 2,484 | 5,341

Woman

Woman 6.3%

Man 93.7%

Fig 14 shows demography of the respondents'- Gender, Education Level and Country using Tableau Dashboard

TABLEAU DASHBOARD

Fig 15

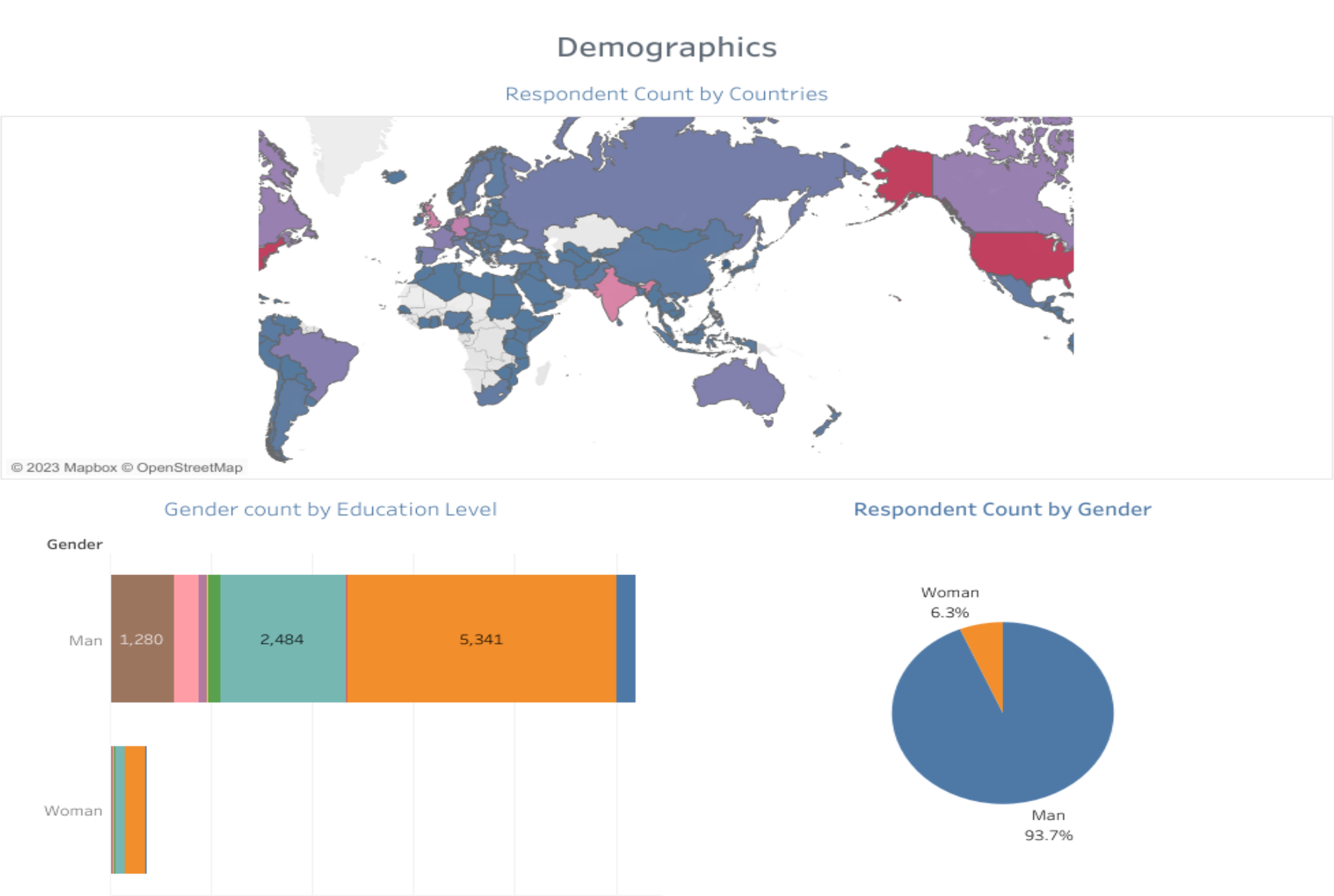## Current Technology Usage

Top 10 Current Database

Database Worked With

MySQL	5,469
Microsoft SQL Server	4,110
PostgreSQL	4,097
SQLite	3,248
MongoDB	3,016
Redis	2,508
Elasticsearch	1,954
Oracle	1,744
MariaDB	1,709
Firebase	1,314

Top 10 Current Languges

Language Worked With

JavaScript	8,687
HTML/CSS	7,830
SQL	7,106
Bash/Shell/PowerShell	4,642
Python	4,542
Java	4,506
C#	4,288
TypeScript	3,232
PHP	2,913
C++	1,946
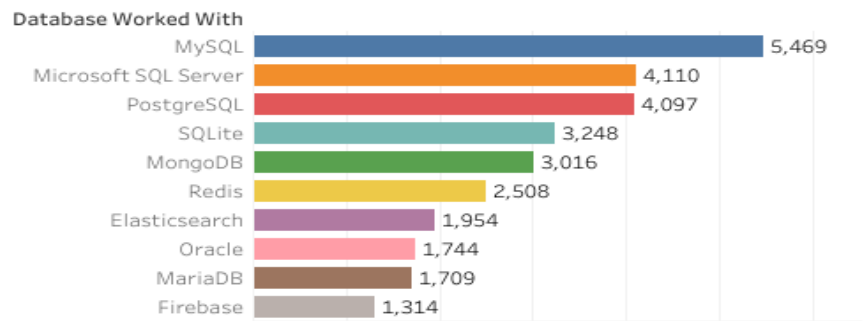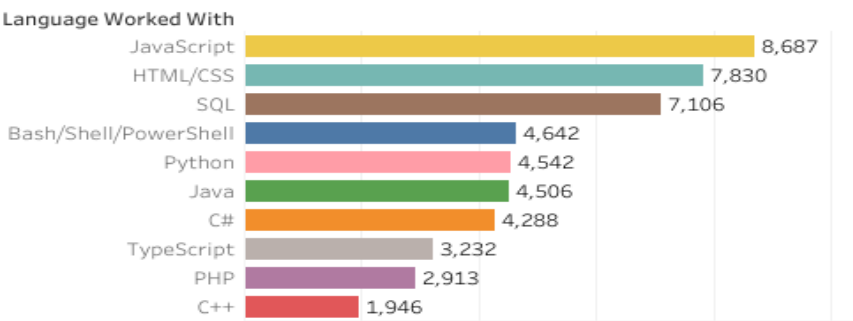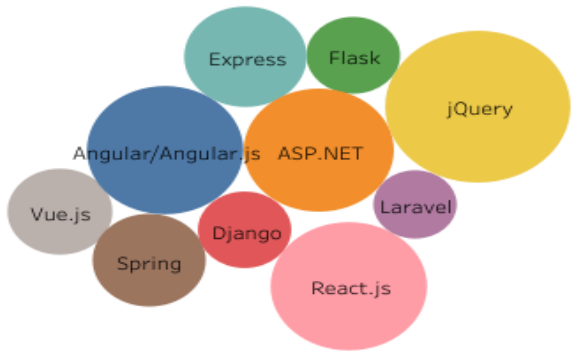
Current WebFrame

Current Platform

Fig 15 shows **Current Technology Used of Databases, Languages, Platform and Web Frame using Tableau Dashboard**

TABLEAU DASHBOARD

Fig 16

Future Technology Trend

Top 10 Languges Next year

Language Desire Next ..
Language	Value
JavaScript	6,630
HTML/CSS	5,328
Python	5,239
SQL	5,012
TypeScript	4,088
C#	3,591
Bash/Shell/PowerShell	3,097
Java	2,953
Go	2,759
Kotlin	1,890

Top 10 Database Next Year

Database Desire Next ..
Database	Value
PostgreSQL	4,328
MongoDB	3,649
Redis	3,331
MySQL	3,281
Elasticsearch	2,856
Microsoft SQL Server	2,706
SQLite	2,437
Firebase	1,650
MariaDB	1,385
DynamoDB	1,044

Top 10 WebFrame Next Year

Express, Flask, Drupal, jQuery, Angular/Angular.js, ASP.NET, Laravel, Django, Other(s), React.js

Platform Next Year

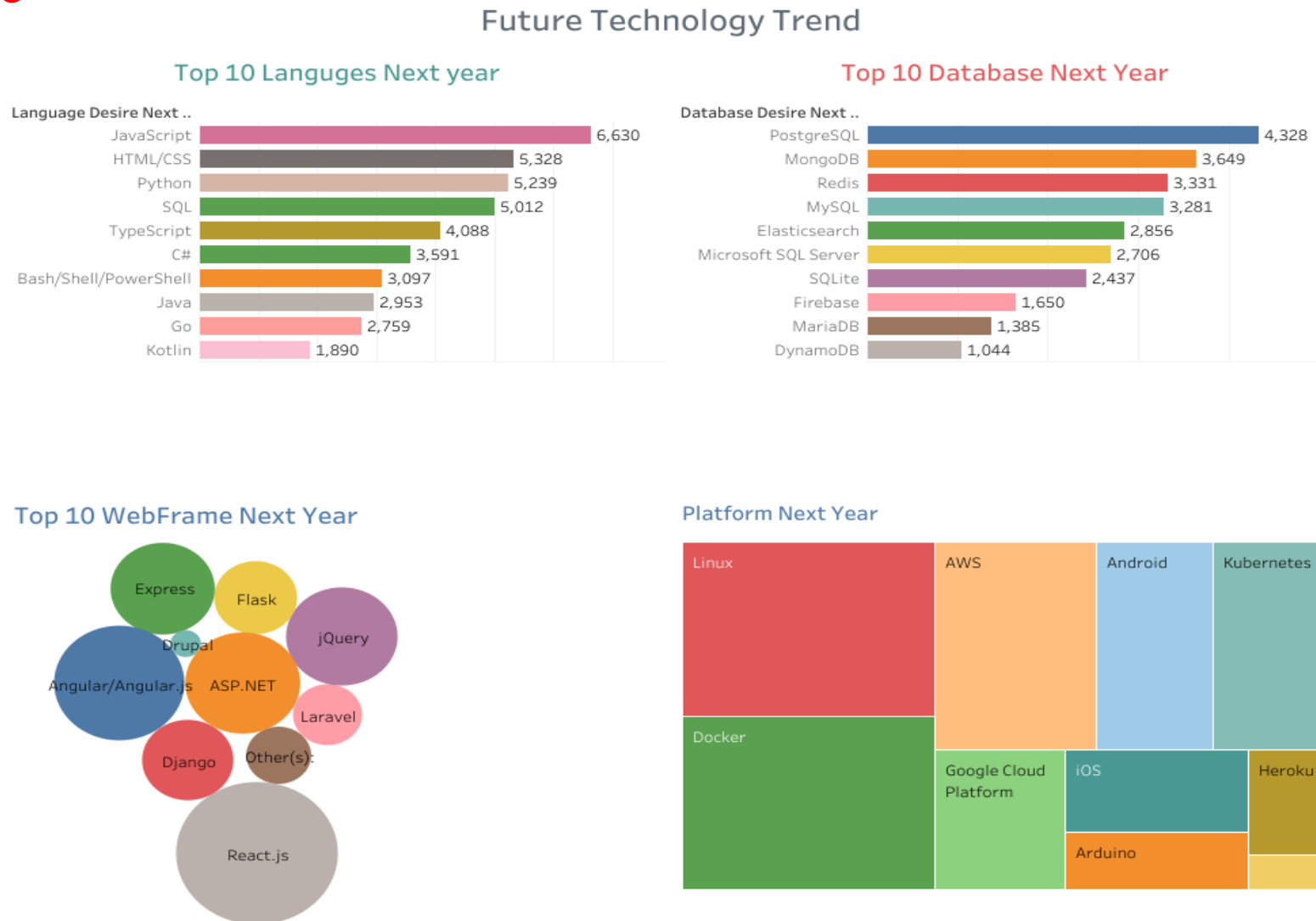Linux, AWS, Android, Kubernetes, Docker, Google Cloud Platform, iOS, Heroku, Arduino

Fig 16 shows **the Future Technology trend of Databases, Languages, Platform and Web Frame using Tableau Dashboard**

OVERALL FINDINGS & IMPLICATIONS

Findings

- **Databases –Elasticsearch 46%, Redis 33%, Firebase 26%, MongoDB 21%** increased by good amount showing growing interest for them

- **Languages**- Despite **TypeScript** has low take up grew by **26%, Python** increased by **15%,**

- **US(3127**), **India(897), UK(821)** & **Germany(708)** have highest number of participants.

Implications

- **Elasticsearch & Redis databases** have higher demands than the rest

- **TypeScript and Python** become most popular and leading programing languages

- Students will have **higher job opportunities** if they worked in the **US, India, UK & Germany.**

RESULTS

According to my findings to remain competitive in the job market students would benefit if they study*:*

- **TypeScript and Python multiple languages** or **Python** is best to study one **Language**

- **Elasticsearch & Redis multiple** databases or **Elasticsearch** to study one **Database**

My investigation suggest that **Students would be better off** if they worked in the **US, India, UK & Germany**

DISCUSSION

One or multiple languages/databases to study?

- For students who don't have the time & resources I suggest that they study **Python** (one language) as its popularity is growing fast & becoming supreme.

- Depending on the field students want to work **Elasticsearch** is best to study one database.

- **Overall** if students study **multiple languages (TypeScript and Python)** and **databases** (**Elasticsearch & Redis**) will **stand out** and become **competitive in the job market.**

CONCLUSION

- **Python** is top programming language with highest job vacancies, demand by employers as well as excellent salary over 100K.

- **TypeScript** and **Python** popularity increased by a good amount so students would benefit if they study these languages

- **Elasticsearch** & **Redis** databases popularity increased by a good amount indicating that growing interest suggesting that students would benefit if they study these databases.

- **US, India, UK, & Germany** are countries with highest number of participant who took part in the survey suggesting that students have highest job opportunities if they worked in these countries.

Limitations

- The actual Demographic & Technologies data set from Stack Over was 90000 rows but only 11552 rows of randomised data provided by IBM for this analysis purpose. Conclusions drawn here may not reflect the real world scenario therefore further analysis is required on larger data.

Suggestions:

❖ Reach out more women during data collection to increase their number taking part in the survey

❖ Encourage more women to take up programing languages & databases to grow their contribution in the field.

Appendix

Data source

- **Demographic & technologies CSV files** (90,000 rows) collected & produced by Stack Over from 2019 survey they conducted with professional software developers but from this only 11552 rows of randomised data (**IBM)**

- **Programming languages job posting data**- Kaggle.com was the original source in CSV format but IBM converted it into Json format for this analysis purpose

Gender Gap

- From 11,552 software developers who took part in the survey the number of women respondents were only 731 (6.5%) whereas men 10480 (93.5%).

- Women's contribution was significantly lower than men. This result indicates that men dominated the software developers' field and is crucial to closing the gender gap.

Acknowledgements and credits to:

- Stack Overflow for conducting the survey, collecting, producing and making the the dataset available open source.

- IBM & Coursera for providing IBM Data Analyst Professional Certificate course & facilitating resources for my analysis.

Thank You