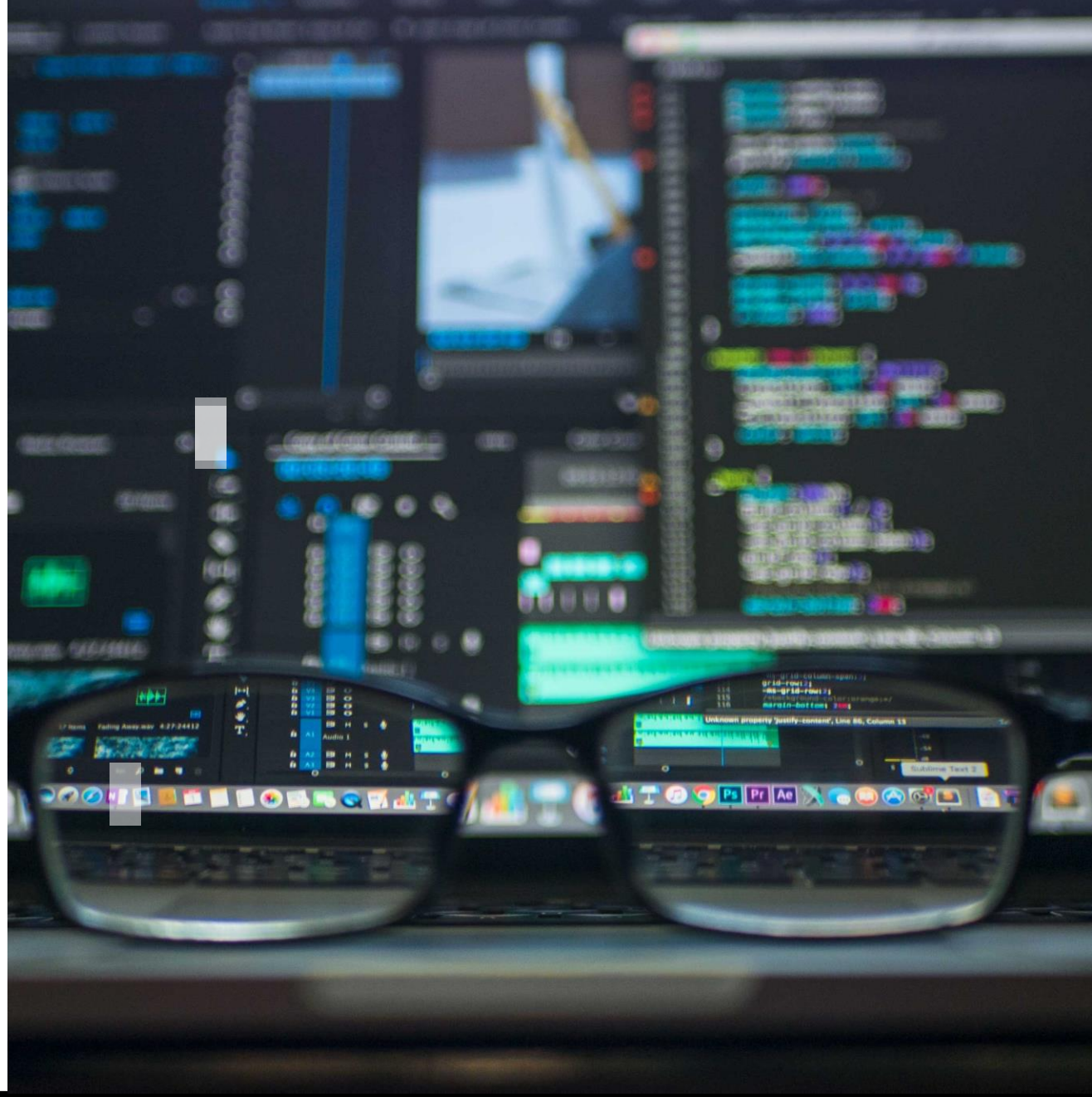
An Exploration Of Data Analyst Trends

Bekki Connolly

August 2021



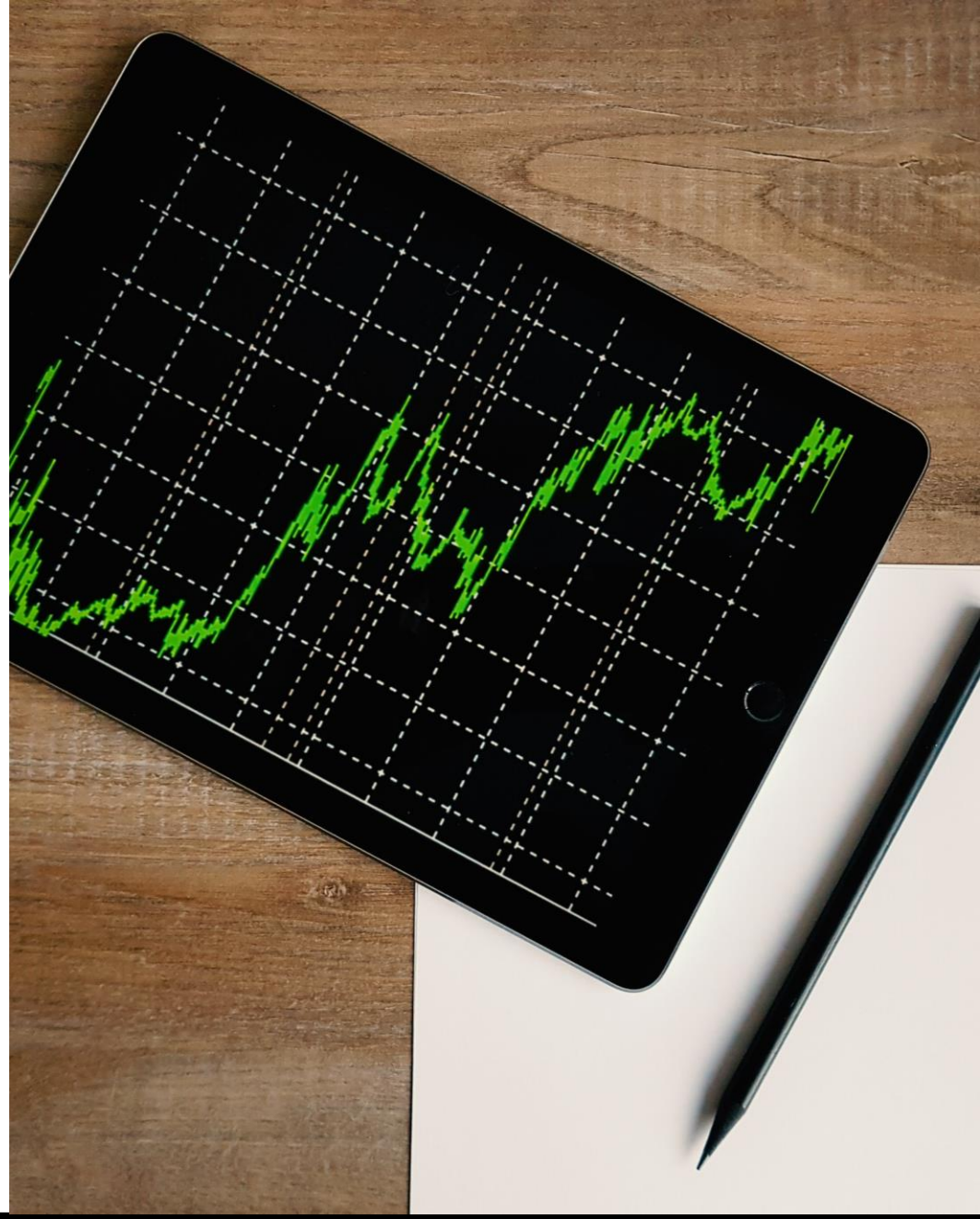
OUTLINE

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

EXECUTIVE SUMMARY

This research project explores trends within the data analyst career path to give a view of the industry to technology professionals, organisations and recruiters.

- The industry is fast paced with constantly shifting trends. It will be important for professionals and companies to monitor trends to stay relevant. For example, MySQL is currently most widely used but other databases are becoming more popular in the industry such as PostgreSQL and MongoDB.
- While knowing many languages and databases could be an advantage, focus should be on JavaScript, Python, Typescript, PostgreSQL and MongoDB as these are set to be important in the future of data analysis.
- While there is a current skew towards those with a Bachelor's degree, 17% have lower qualifications which shows that relevant practical skills are more important that a degree.
- There is a current gender bias towards men (93.7% of respondents) in the industry that needs to be addressed by companies and recruiters. The industry is quite young but higher salaries for older employees push up averages so to stay attractive to talent, companies should offer competitive compensation.



INTRODUCTION

This research project aims to explore trends within the Data Analyst career path with a focus around the most popular languages and databases used. Insight gained from this report will provide a guide to the current and future trends about in-demand skills within the industry. It will also investigate the demographics of workers within the industry and their salaries.

This report will help:

- Technology professionals
  - Learn about the current landscape of the industry
  - Gain knowledge about the best skills to enhance their career
  - See how they currently stack up against industry expectations

- Organisations
  - Understand if they are in line with industry norms
  - Develop programs within their workplace to keep talent up to date
  - Stay relevant within the industry

- Recruiters
  - Find the best talent based on industry trends
  - Understand if current job specifications are relevant

METHODOLOGY

- In order to find reliable data for this report, data was collected from:
  - GitHub Jobs API – to find job listings by programming language
  - A webscraping from this table to collate average salaries
  - Using Pandas to pull data from this CSV file to find insight such as the average age of people who use code within their job role
  - M4 Data Source & M2 Survey Data Source - used to create visuals to explore averages and correlation
- Tools/Platforms used:
  - Jupyter Notebooks – with Python
  - Microsoft Excel – with xlsx, csv and json
  - IBM Cognos Dashboard
- Packages used:
  - Pandas
  - NumPy
  - BeautifulSoup
  - Requests
  - Matplotlib
  - Seaborn
  - SQLite

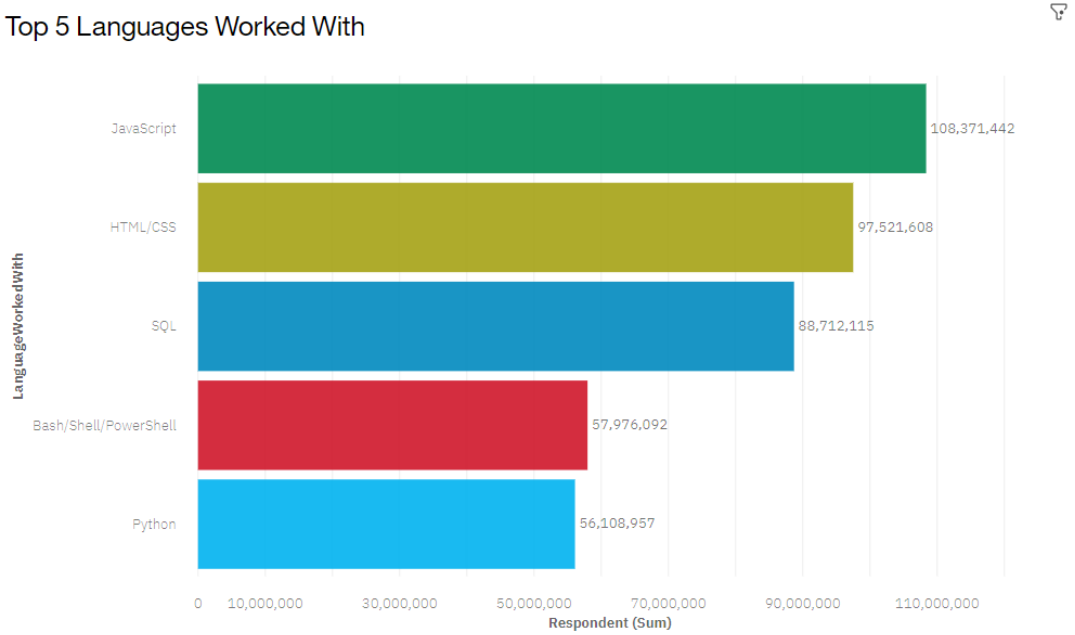
RESULTS

From the various data sources, this project identifies the current trends pertaining to programming languages and databases along with predictions for future trends. This will allow stakeholders to ensure they develop their skills or the skills within their business to keep up to date with the industry.

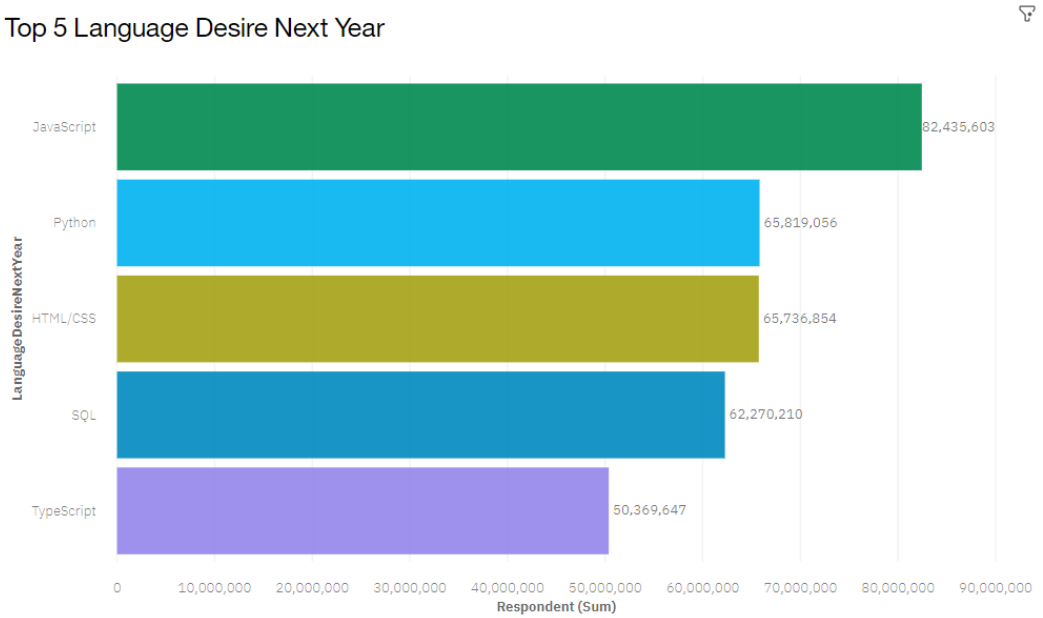
It also highlights the average salaries, demographics and education levels of professionals within the field. This presents a profile of the average employee within the field to create effective job roles but also identify gaps within the industry. Salary analysis can also help to ensure that companies are offering the correct salaries to attract the best candidates.

PROGRAMMING LANGUAGE TRENDS

Current Year



Next Year



PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings:

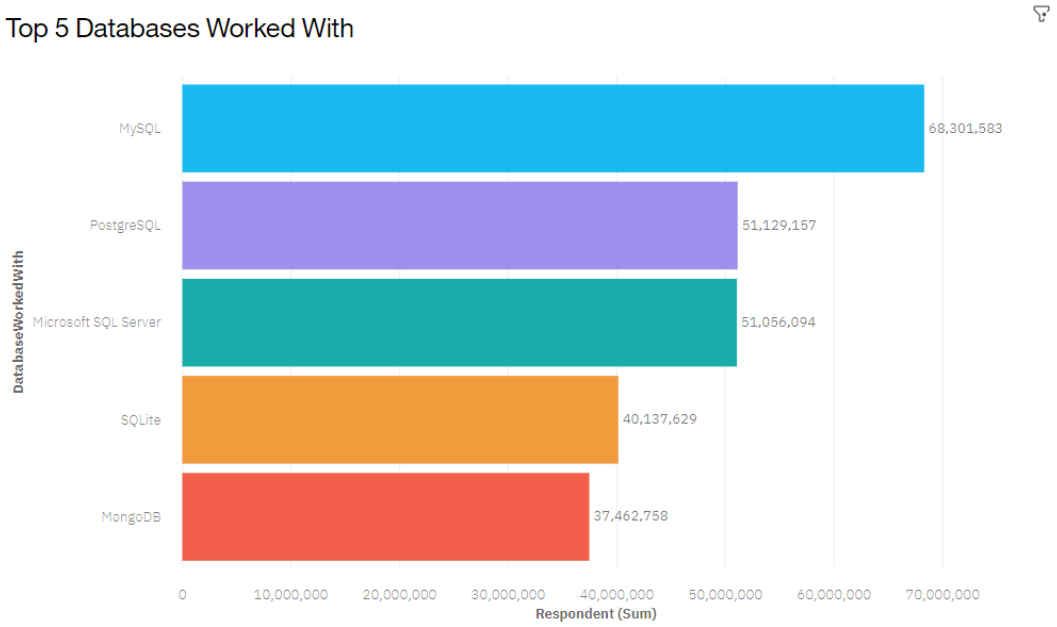
- JavaScript is currently the programming language favoured by professionals and it is set to remain the most popular language next year.
- Bash/Shell/PowerShell is within the top 5 used for this year, however, professionals have less desire to use it in the future.
- TypeScript is within the top 10 languages used this year, but its desirability is within the top 5 for next year.
- Python is already within the top 5 but looks to become even more popular in the next year.

Implications:

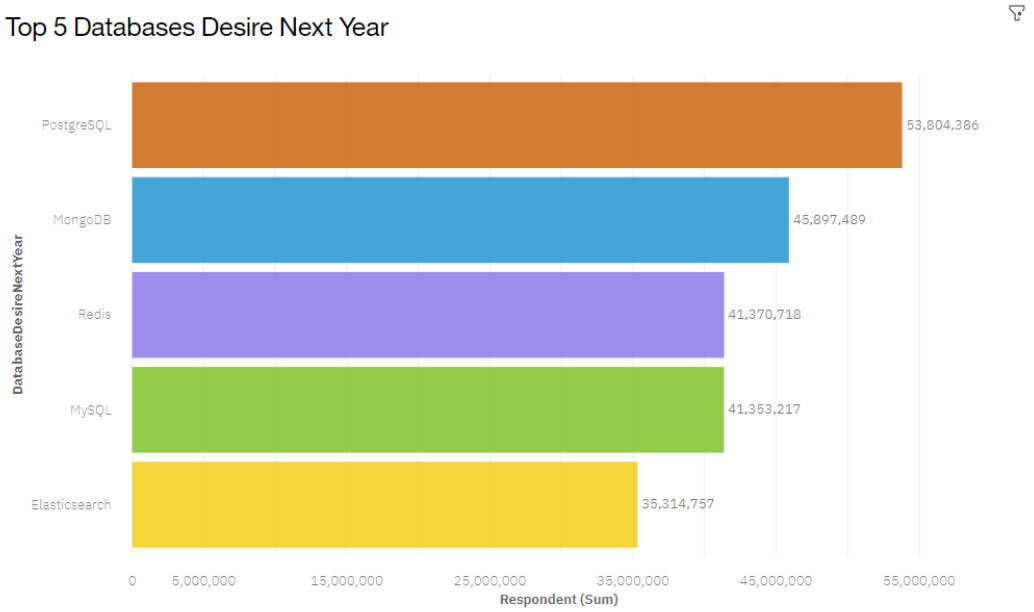
- It is important for professionals to continue to develop their Javascript skills and stay up to date with all new developments. Organisations should look to keep this as their main language used.
- There should be less focus around skills in Bash/Shell/PowerShell, however, as it is still within the top 10 desired languages it is still important to make sure that a basic knowledge is maintained within
- With TypeScript and Python becoming more desirable, organisations and individuals should look to develop skills in both languages as they may see an industry preference shift with job vacancies advertised more towards those languages.

DATABASE TRENDS

Current Year



Next Year



DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- MySQL is current the most popular database used but desire drops to the 4th most popular for the coming year.
- PostgreSQL is only currently the second most popular database used but it's the most desirable for next year.
- MongoDB sees a rise in desirability from the 5th most popular to the 2nd most popular.
- SQLite is currently in the top 5 databases but is not in the top 5 desired for next year, while Redis is not currently in the top 5 used but is in the top 3 most desired for next year.

Implications:

- While MySQL is currently the most widely used, we are likely to see a slight shift away from use. Businesses should aim to ensure that their teams are versed in the databases becoming more popular in the industry, such as PostgreSQL and MongoDB.
- As SQLite becomes less desirable to professionals, it could show a shift away from organisations using it. This could mean that there is less focus on it being a part of person specifications in the future.
- As preferences change, businesses will need to monitor these trends and make sure that they have the capacity to change the databases that they use if the industry shifts.

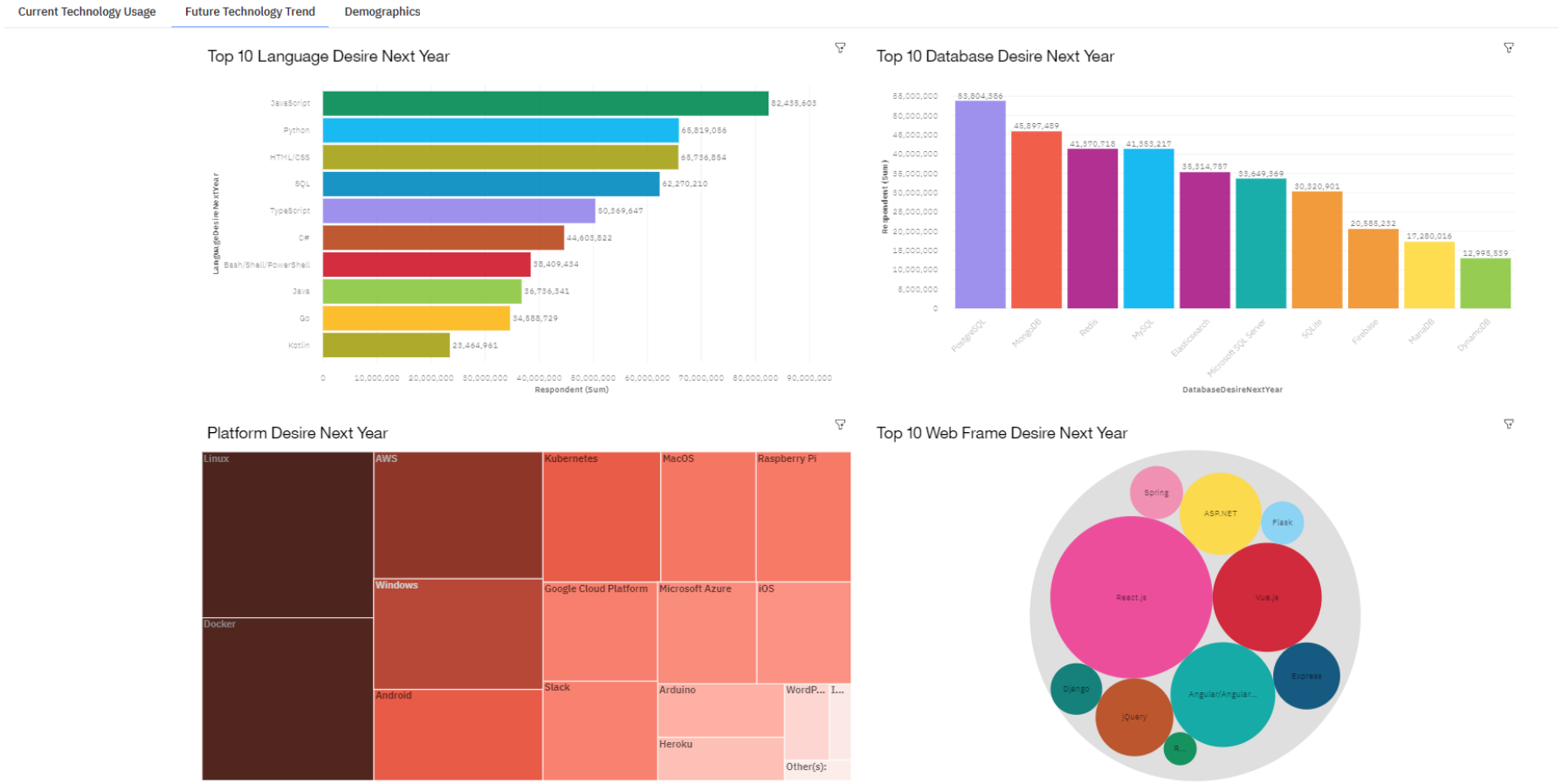
DASHBOARD

For an interactive version of the Cognos Analytics Dashboard referred to the in the next few slides, please visit:

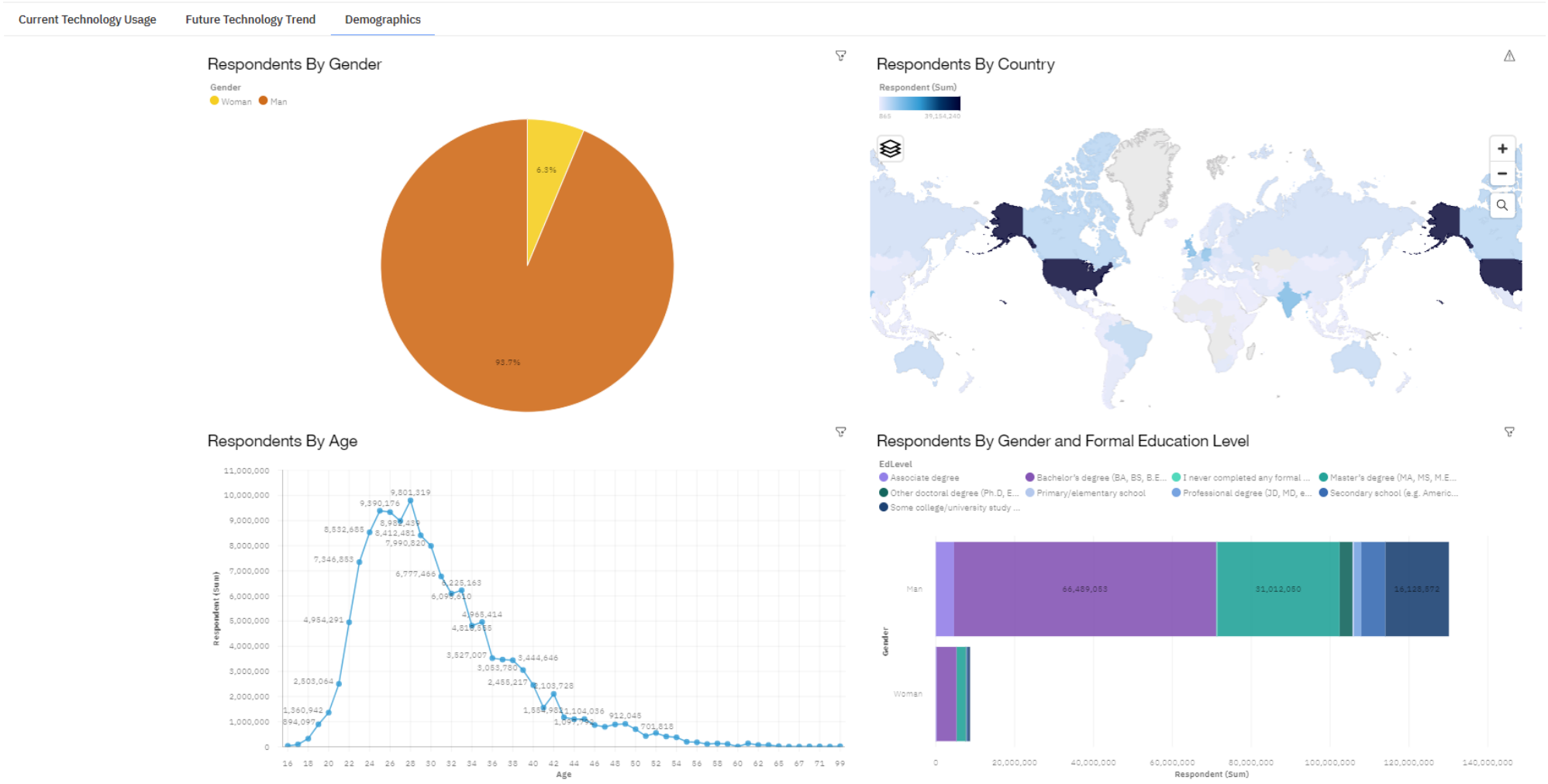
https://eu-gb.dataplatform.cloud.ibm.com/dashboards/fb087fd3-fc39-445c-a9b1-ece6e3fa0a95/view/0337d71f04963fee67b4d0e4079f2b037b3e275ebabb815081d07b495c637197a86c1491c82f43538c170032f4e44108cc

TAB 1: CURRENT TECHNOLOGY USAGE

TAB 2: FUTURE TECHNOLOGY TREND



TAB 3: DEMOGRAPHICS



DISCUSSION

Demographics:

The majority of the professionals in the dataset are males (93.7%) with an age skew of 22-35. While the small amount of females can be an indicator that the industry is male-dominated, it could also indicate a bias in the data that may potentially need addressing for future surveys.

While we do see a high proportion of respondents having a Bachelor's Degree, it is important to note that 17% of the respondents have lower formal qualification levels – which proves that, while a degree is preferable, you can gain entry to the industry without investing in attaining a degree.

OVERALL FINDINGS & IMPLICATIONS

Findings

- There is a lot of movement in the industry between which programming language and database skills people already possess compared to what they would like to develop in the future.
- There are a higher number of respondents concentrated in the USA. This suggests a higher demand for data roles in the States.
- The industry median is 29 with the median compensation coming in at $57,745. This number is below the average salaries per language which suggests that there are a lot of people who are being salaried lower than they should be considering their skillset. Average compensation rises with age with a large gap between 25 year old analysts and 60 year old analysts which could also explain this variance.
- There is a high proportion of men compared to women with a skew towards people with higher education.

Implications

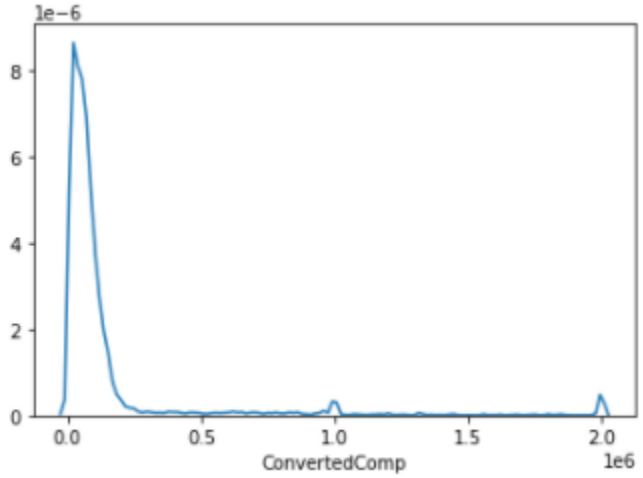
- With high average salaries across the board regardless of language, it is worth learning multiple programming languages to get a more prestigious role in data.
- Python is one of the most popular languages – and is predicted to stay that way – with the second highest average salary and the fourth highest number of jobs being posted; this indicates that it could be an important language to know for career progression. JavaScript also presents a high desirability case.
- To attract the best talent, companies and recruiters should look to compensate within industry standards, or they may lose employees to those who are willing to pay more. They should also constantly develop their current talent to stay up to date with the industry standards and make sure they do not fall behind.
- Women are currently underrepresented which could indicate that employers and recruiters aren't currently appealing themselves to women.

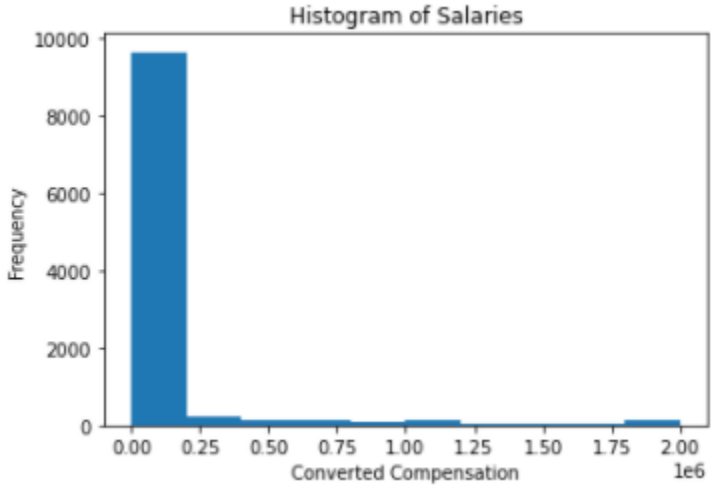
CONCLUSION

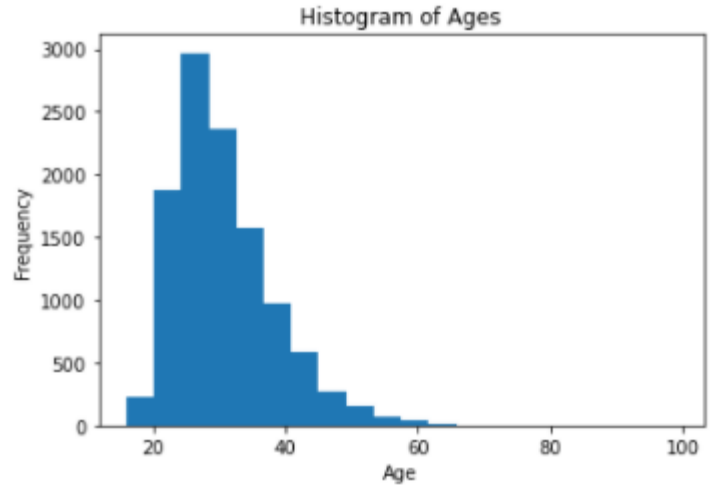
- The most popular languages and databases used within data analysis are constantly changing. It would be prudent for all stakeholders to constantly monitor data trends.
- Anyone looking to find work in the data analyst field should look to develop their programming language and database skills in as many languages and databases as they can. Priority should be given to learning Javascript, Python, Typescript, PostgreSQL and MongoDB as these are set to be important in the future of data analysis.
- In order to stay up to date with this fast-moving industry, businesses should aim to constantly develop their current employees, hire new employees with wider language knowledge and keep their systems up to date.
- In order to find and employ the best people within the industry, recruiters will need to make sure that job postings advertise competitive salaries, focus on practical language and database skills rather than education and use language that attracts more women to the field.

APPENDIX

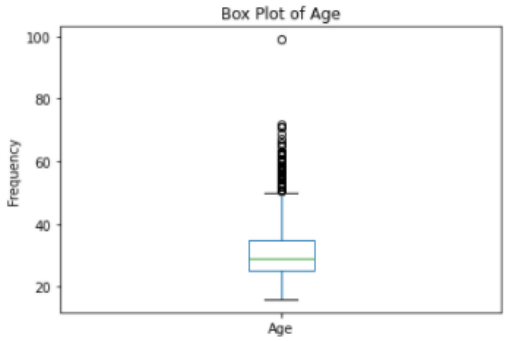
M2 Survey Data Source

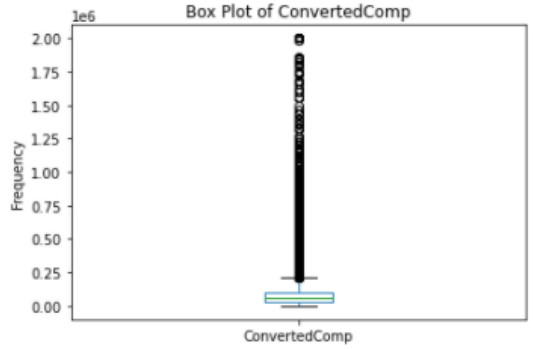






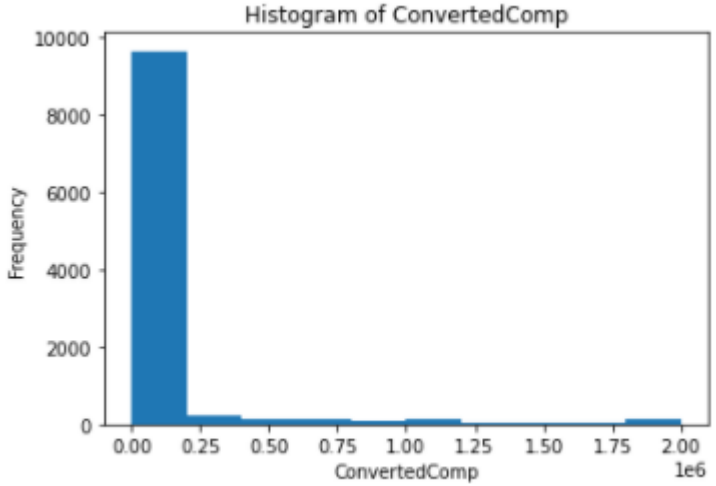

```
Gender Value Count:  7
Man                                                            9650
Woman                                                           694
Non-binary, genderqueer, or gender non-conforming                56
Man;Non-binary, genderqueer, or gender non-conforming            26
Woman;Non-binary, genderqueer, or gender non-conforming          14
Woman;Man                                                         9
Woman;Man;Non-binary, genderqueer, or gender non-conforming       2
Name: Gender, dtype: int64
9650 reponders gave the response Man
```

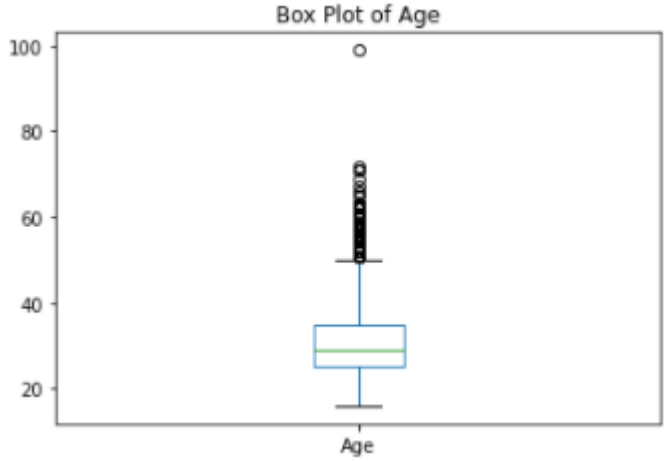


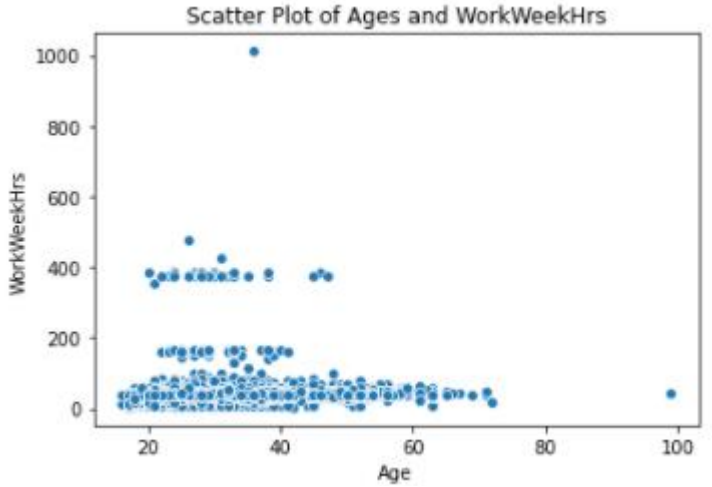


APPENDIX

M4 Survey Data Source

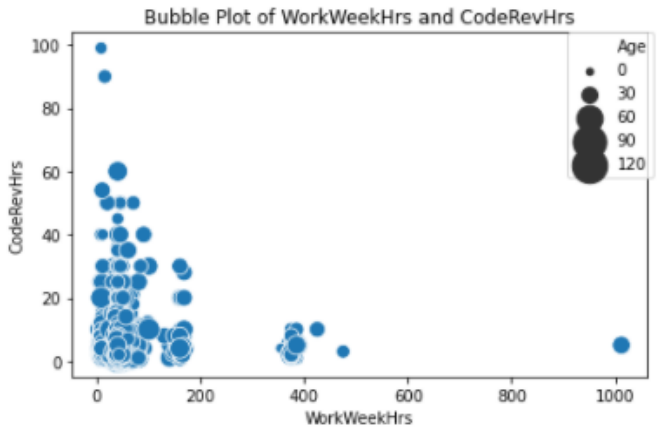


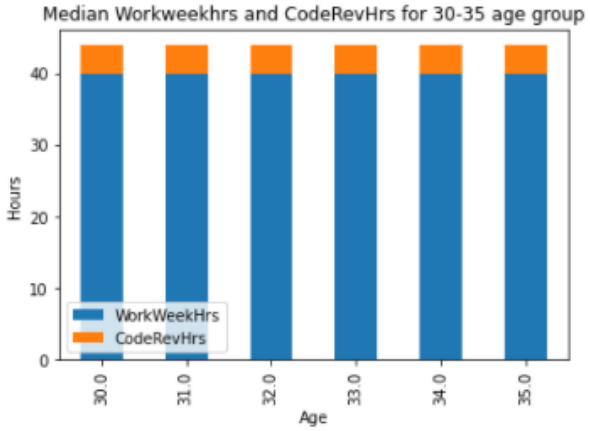





```
df.corr()["Age"]
```

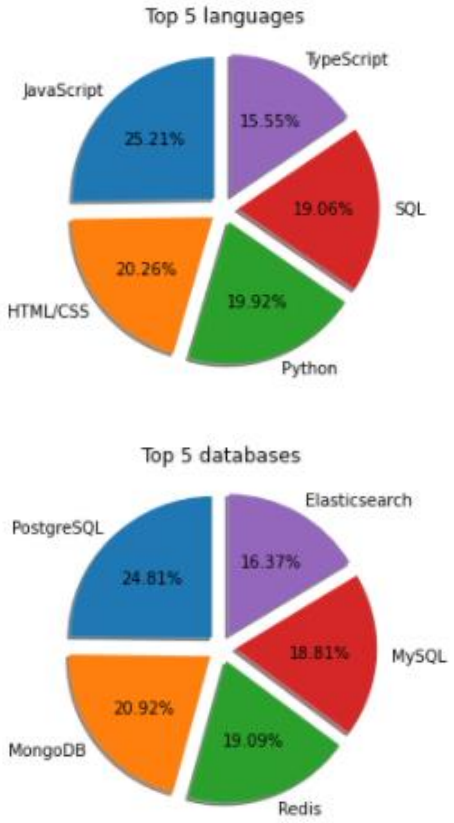
```
index             0.004177
Respondent        0.004048
CompTotal         0.006966
ConvertedComp     0.105412
WorkWeekHrs       0.036517
CodeRevHrs       -0.020498
Age               1.000000
Name: Age, dtype: float64
```





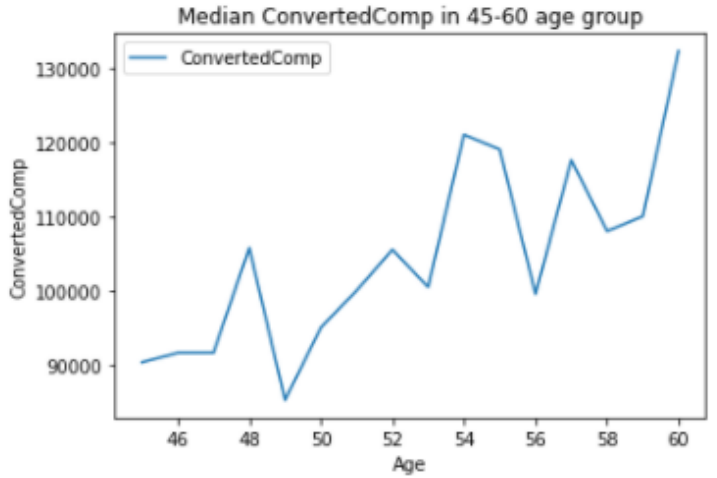
APPENDIX

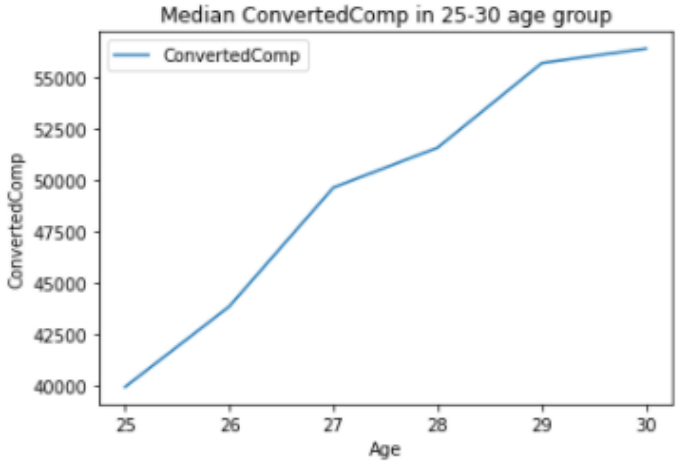
M4 Survey Data Source



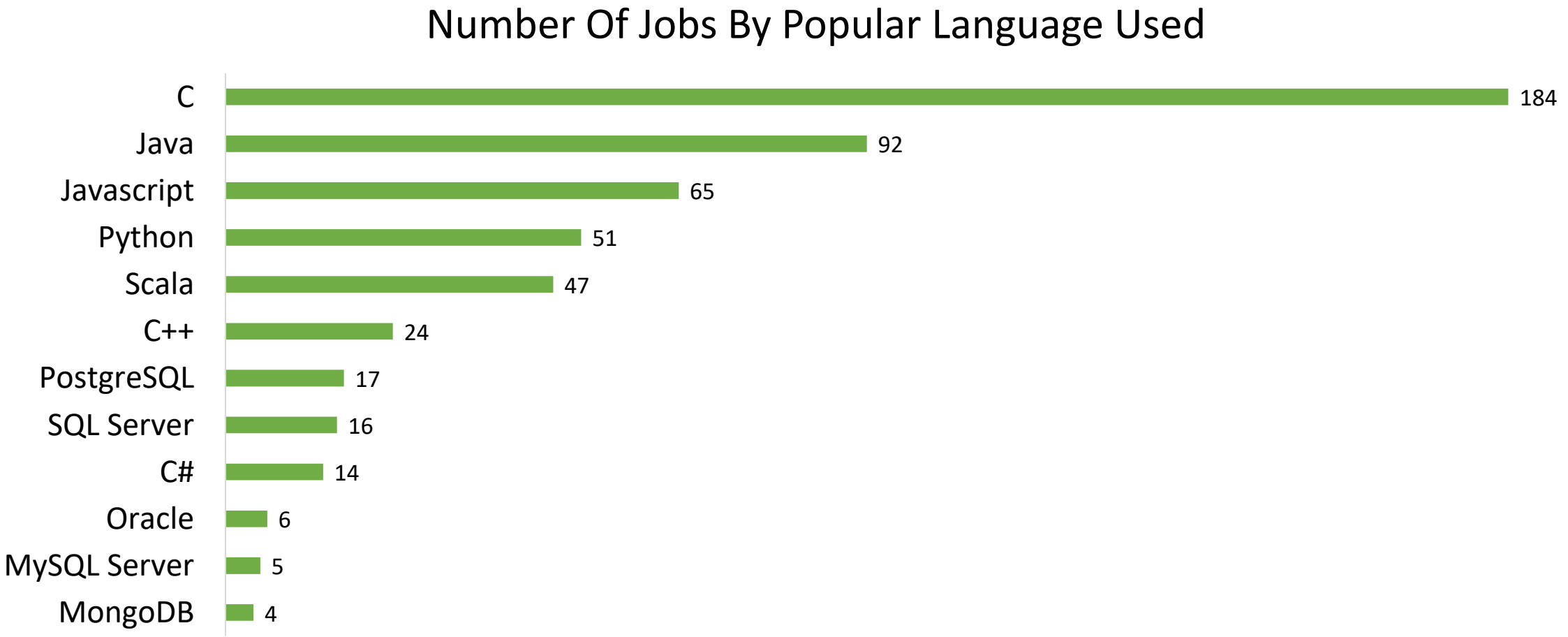
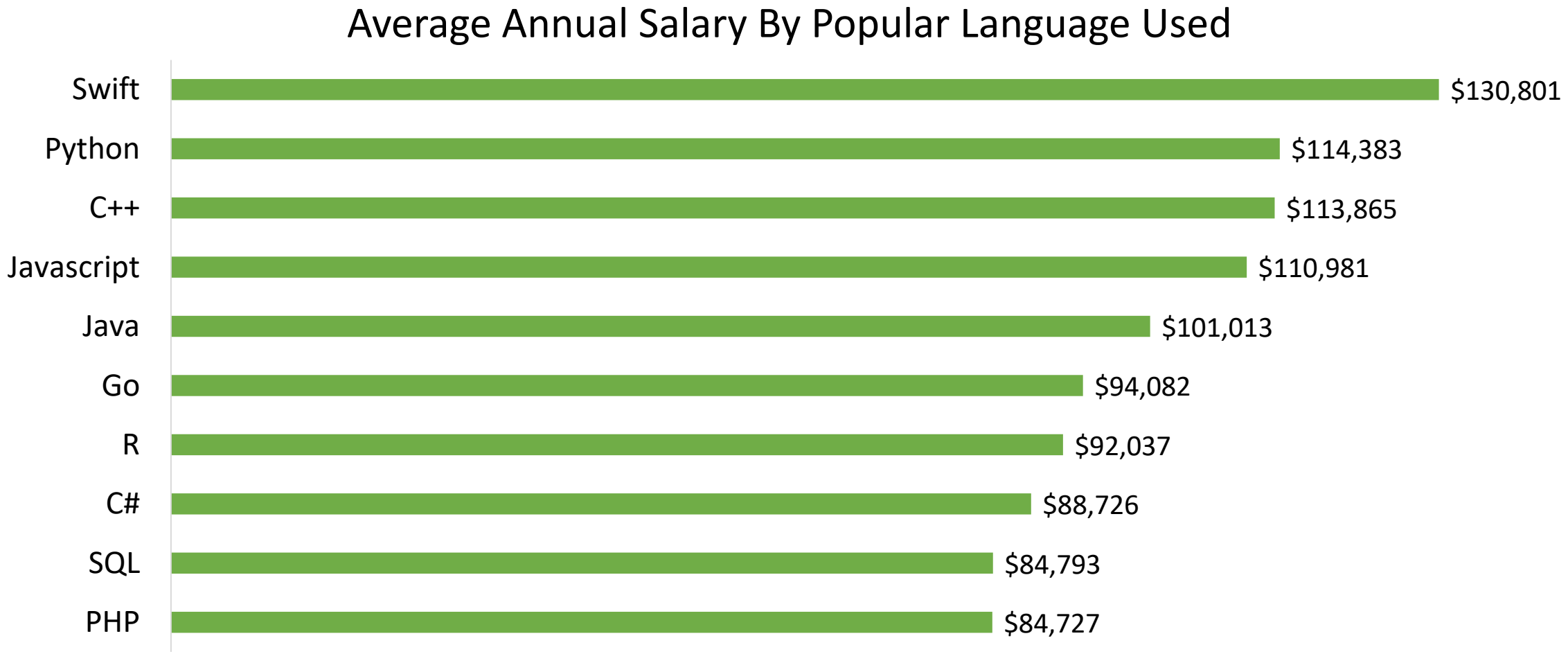
	Respondent
DatabaseWorkedWith	
Microsoft SQL Server	845
MySQL	474
PostgreSQL	351
SQLite	192
MongoDB	170
Oracle	162
Other(s):	94
Firebase	90
MariaDB	44
DynamoDB	28
Elasticsearch	27
Redis	20
Cassandra	10
Couchbase	8

	Respondent
DevType	
Developer, full-stack	6928
Developer, back-end	6290
Developer, front-end	3920
Developer, desktop or enterprise applications	2575
Developer, mobile	1959
DevOps specialist	1639
Database administrator	1413
System administrator	1202
Designer	988
Developer, QA or test	911
Developer, embedded applications or devices	854
Engineer, data	832
Data scientist or machine learning specialist	803
Data or business analyst	802
Student	766
Academic researcher	556
Educator	514
Product manager	480
Developer, game or graphics	472
Engineer, site reliability	449
Engineering manager	386
Scientist	354
Senior executive/VP	160
Marketing or sales professional	61





APPENDIX: GITHUB JOB POSTINGS



APPENDIX: POPULAR LANGUAGES