Data Analyst Capstone Project

Syed Bokhari

10-09-2022

Table Of Contents

OUTLINE

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

EXECUTIVE SUMMARY

The world of technology is advancing rapidly. New technologies are appearing on the horizon at breakneck speed. Terms such as Artificial Intelligence, Internet-of-Things and Deep Learning were relatively unknown just 10 years ago but are now commonplace in the tech world. The total number of smart devices is increasing so rapidly that the number of devices connected to the internet in 2030 is expected to be 500 billion. (TechJury, 2022). These devices will generate massive amounts of data. Already in 2025, the world is expected to produce 463 exabytes of data. (TechJury, 2022).

In this context, it is difficult for students and new entrants to decide which technology is the most useful to learn about. There are a plethora of options and there are several technologies that are hyped as the next big thing. This project aims to cut through all the hype and provide some analysis grounded in actual data. It makes use of the Stack Overflow Developer Survey 2019 dataset which is open-source for its analysis.

EXECUTIVE SUMMARY

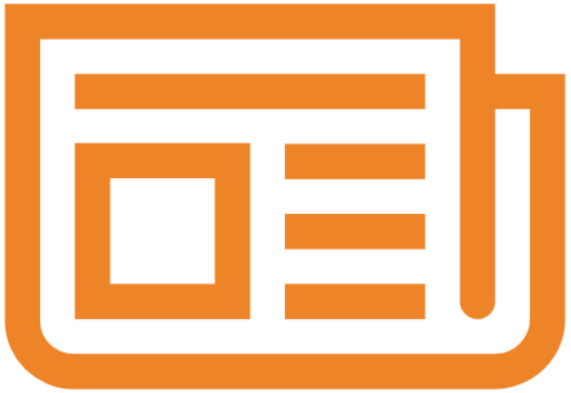
The project makes use of the dataset to draw visualizations and identify trends. It makes use of data on popularity and desirability of Programming Languages, Databases, Platforms, Web Frameworks and Miscellaneous Technologies. The visualizations make apparent the trends in all these technologies and certain patterns become clear. Next, a Dashboard of a subset of the data is created. The dashboard is used to draw some important insights about the dataset and the demographics of the respondents to the survey that comprises the dataset. All this analysis is subsequently combined to draw some important insights about the trends in the technology sector.

The analysis showed that languages related to Web Development were the most popular and desirable. It also showed the popularity of SQL for databases but with caveats. And it indicated rising popularity and desirability for Python. It also noted the educated male dominated nature of technology sector with a relatively young cohort.

INTRODUCTION

The world of technology grows exponentially. New technologies are gaining popularity while old technologies are adapting to the needs of the technology landscape.
These rapid changes are daunting for new entrants and students who find themselves uncertain about which technology skills to acquire. This project aims to shed light on some of the current trends and possible future trends in the tech world. It makes use of the popular developer website, Stack Overflow's Developer Survey for 2019 (which was made open-source) to draw insights. Data visualization is used to make the trends apparent.
The project analyses trends in Programming Languages, Databases, Web Frameworks, Platforms and Miscellaneous Technologies as well as future trends based on survey results. The results of the analysis are then used to identify future skill requirements as well as potential pitfalls about interpreting the results.

METHODOLOGY



Performed data collection with API, webscraping

Performed data wrangling on Survey dataset

Performed EDA

Used SQL and Plotly Express to generate visualization for further analysis

Built an IBM Cognos dashboard

METHODOLOGY-DATA COLLECTION (API)

- Data for API was collected from: https://www.kaggle.com/datasets/promptcloud/jobs-on-naukricom
- Data from the API was collected using the Python responses library which provides function for accessing URLs by making use of HTTP methods.
- Pandas library was sued for data manipulation.
- The URL of the API was used to get data in JSON format. Subsequently, a custom function was built to access the URL using Responses library, and get the required data in JSON format .
- The data was then stored as an excel file by making use of the openpyxl library alongside the Pandas library.
- A visualization made with this data is in the Appendix section.

METHODOLOGY-DATA COLLECTION (WEBSCRAPING)

- For webscraping, data was acquired from:
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DA0321EN-SkillsNetwork/labs/datasets/Programming_Languages.html
- For webscraping, the requests library along with the BeautifulSoup library from bs4 was used.
- The URL was accessed and data collected using the requests library.
- The BeautifulSoup library along with the html5lib parser was used to collect the data from the webpage.
- The csv library was used to write the data into a .csv file.
- A visualization made with this data is in the Appendix section.

METHODOLOGY-DATA WRANGLING-DUPLICATES

- The Pandas library in Python was used for Data Wrangling.
- The data was acquired from the link:
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DA0321EN-SkillsNetwork/LargeData/m1_survey_data.csv
- First, duplicated values were identified by using the Pandas .duplicated() method on the dataframe. There were a total of 154 duplicated records.
- The .drop_duplicates(inplace=True) method with the inplace parameter set to True to make changes to the dataframe itself was used to remove the duplicates from the data.
- The .duplicated() method and the .unique() methodwere used to verify the removal of duplicates.

METHODOLOGY-DATA WRANGLING-MISSING VALUES

- Next, missing values were tackled.
- The .isnull().sum().sum() method was used to find the total number of missing values in all columns.
- Then isnull.sum() method was used to find missing values in columns WorkLoc, Country and EdLevel each.
- The .value_counts() method was used to find the values used and their total counts in the WorkLoc column. The majority value was noted.
- The .fillna(inplace=True) method with the inplace parameter set to True was used to replace all missing values in the WorkLoc column with the majority value identified before.
- The isnull.sum() method was used to confirm all missing values from the WorkLoc column were removed.

METHODOLOGY-DATA WRANGLING-NORMALIZATION

- Next, normalization was performed.
- There are two columns in the data, CompFreq and CompTotal which give the total compensation for respondents when used together.
- Due to varying values of these two columns, a standard Compensation was difficult to calculate for respondents.
- The .value_counts() method was used to find the different values in CompFreq column.
- A custom function was created to apply different values to the CompTotal column depending on the CompFreq column. The results were stored in a Pandas Series. Missing values were replaced with a 0.
- The Series was added to the data as a new column called Normalized Annual Compensation.

METHODOLOGY-DATA WRANGLING-OUTLIERS

- Next, outliers were handled.
- A box plot was used to model the ConvertedComp column using MatPlotLib, Seaborn and PlotlyExpress. The resulting plot showed several points lying outside the 1st and 3rd quartiles of the data.
- The interquartile range was calculated using the .quantile() method with 25 and 75 as inputs. The resulting values were subtracted for the range.
- The upper and lower bounds of the data were found using the formula and the values calculated previously.
- The upper and lower outliers lying outside the calculated upper and lower bounds were found.
- The outliers were eliminated using the .clip() method. The results were saved in the data.

METHODOLOGY-EXPLORATORY DATA ANALYSIS

- An Exploratory Data Analysis (EDA) was conducted to understand the data in the dataset.
- The .describe() method was used to find summary statistics for various columns in the dataset.
- Various columns were plotted and the results were used to perform some basic analysis of the columns.
- Distributions of values in various numeric columns were plotted using Plotly Express to get a sense for the data.
- The .corrwith() method was used to see the correlation of columns with each other.
- The plots have been added to the Appendix section.

METHODOLOGY-SQL AND PLOTLY

- The sqlite3 library along with Pandas library were used to query a sqlite database for the data from the survey.
- A conn object was created using sqlite3 to connect to the survey data database. Then the Pandas method .read_sql_query(Query, conn) was used to send SQL queries to the database.
- The data from the survey had been divided into multiple tables in the sqlite Database. The tables were listed for ease of access.
- Subsequently, queries were used on various tables to extract the information deemed necessary for the project.
- The data acquired from the queries was used to create plots of trends in various areas of technology using Plotly Express. Those are shared in the results section.

METHODOLOGY-COGNOS DASHBOARD

- In order to create a Dashboard, a subset of the survey data was used:
- The Cognos Dashboard was created by logging in to IBM Cloud and using the Cognos Dashboard there.
- The dashboard comprised of three different tabs, each representing a different set of visualizations.
- The Current Technology Usage tab uses a Bar Chart to display the 10 most popular Programming Languages, a Column Chart to display the 10 most popular Databases, a Word Chart to display the 10 most popular Platforms and a Hierarchy Bubble Chart to display the 10 most popular Web Frameworks.

METHODOLOGY-COGNOS DASHBOARD

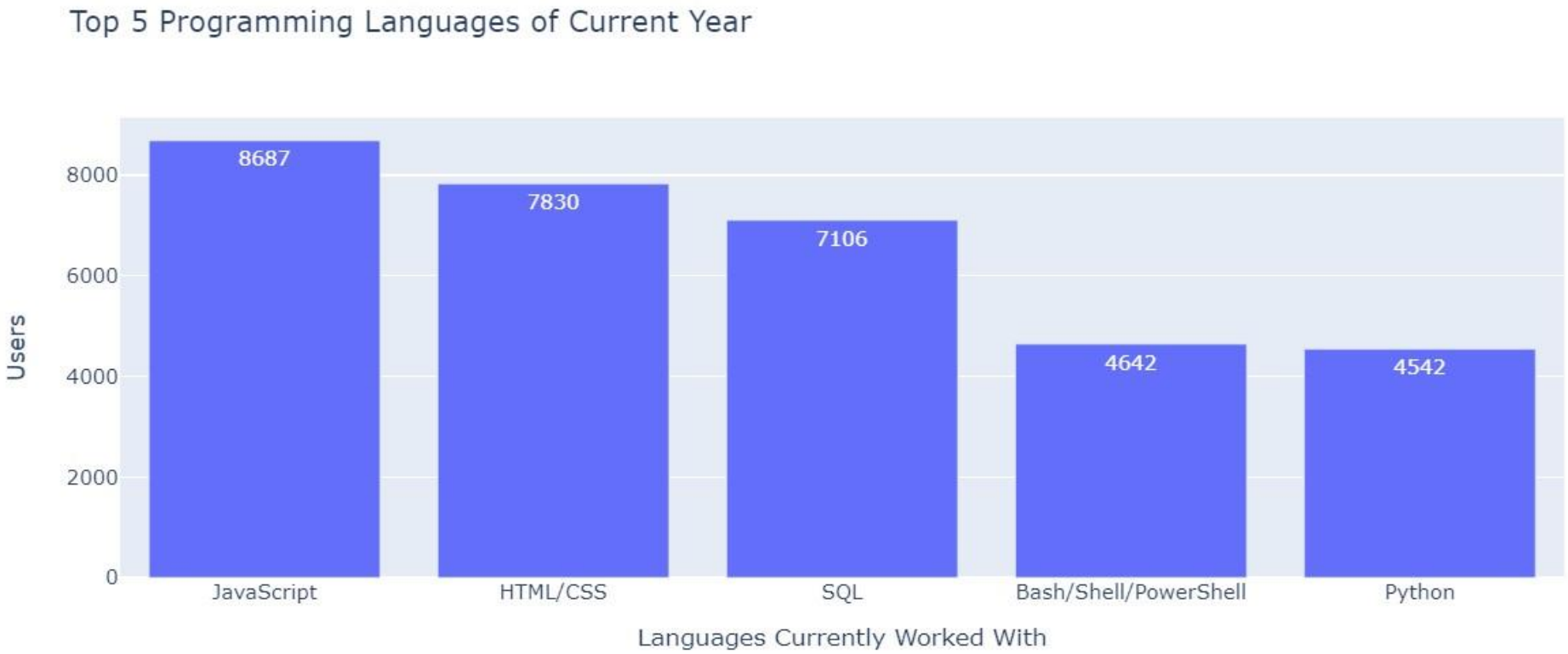
- The Future Technology Trend tab uses a Bar Chart to display the 10 most desirable Programming Languages, a Column Chart to display the 10 most desirable Databases, a Tree Map to represent the 10 most desirable Platforms and a Hierarchy Bubble Chart to represent the 10 most desirable Web Frameworks.
- The Demographics tab displays the demographic information of survey respondents by using a Pie Chart to represent the Gender of respondents, a Choropleth Map to represent the Country of respondents, a Line Chart to represent the Age of the respondents and a Horizontal Bar Chart to represent the Formal Education Level of participants.

RESULTS

Trends in Top 5 most popular and Top 5 most desirable Programming Languages, Databases, Platforms, Web Frameworks and Miscellaneous Technologies were visualized from the collected data for the purposes of Analysis. Those plots, the resulting Findings, and the Implications from the findings are shared in the following section.

PROGRAMMING LANGUAGE TRENDS

Current Year



PROGRAMMING LANGUAGE TRENDS

Next Year

PROGRAMMING LANGUAGE TRENDS - FINDINGS

- A comparison of the current and future trends in Programming Languages indicates that JavaScript and HTML/CSS are the top two languages respectively among working professionals and as languages that respondents wish to learn.
- SQL is the 3rd most popular language being used by professionals but it is only the 4th most desired language by respondents for learning.
- Python is the 5th most popular language amongst working professionals but it is the 3rd most desired language for learning by respondents.
- Typescript is the 5th most desired language for learning while Bash/Shell is the 4th most popular language among working professionals.

PROGAMMING LANGUAGE TRENDS - IMPLICATIONS

–The popularity of JavaScript and HTML/CSS indicates that web development is the most popular use of programming languages as both these languages are used extensively for creating web based applications. Given the growth in internet users and the increasing number and types of devices connected to it, these two languages are likely to remain in the top two for some time as these developments will only encourage more development of web based applications.

– The popularity of learning Typescript is also related to growth in internet users and devices connected to it as it is in effect, an enhanced version of JavaScript with additional features for application development. Its popularity is expected to rise.

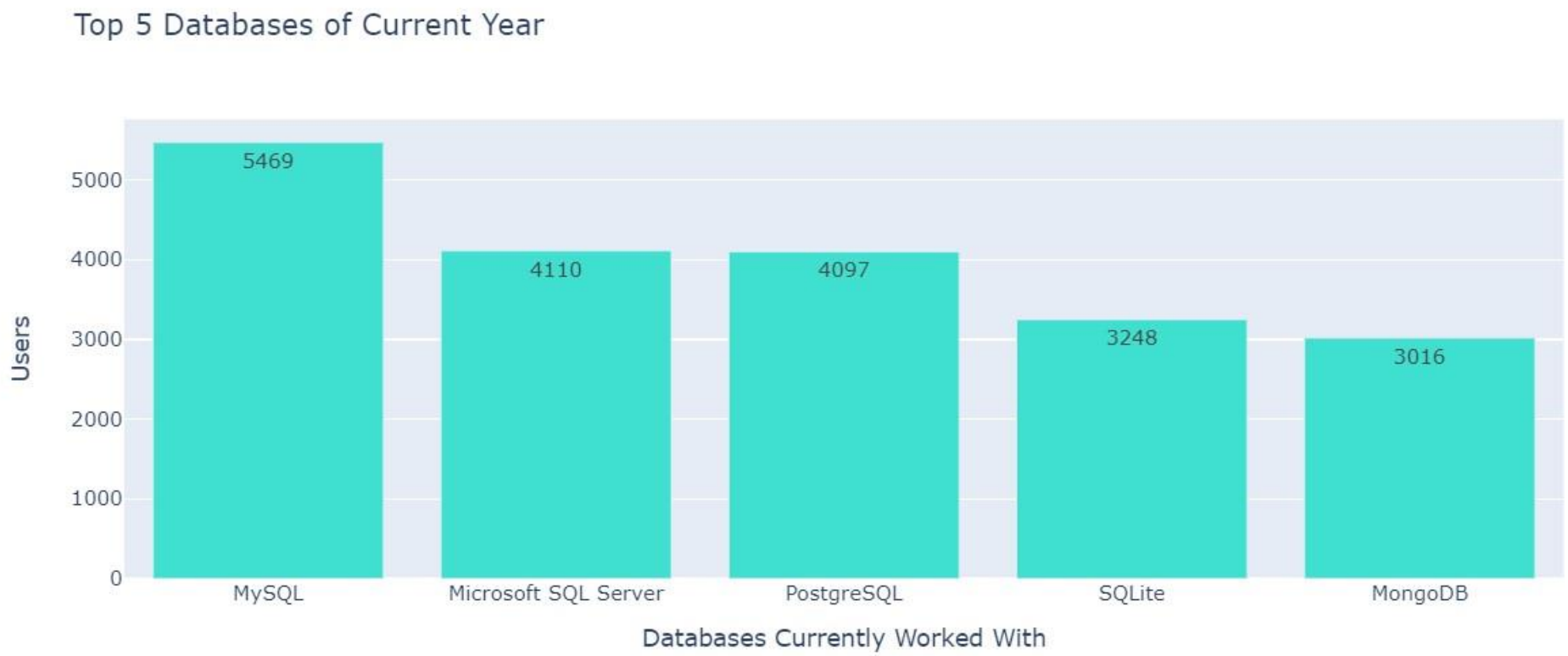
–Bash/Shell is mostly used in Linux based systems for administration and scripting and is directly tied to use of Linux as a platform.

PROGRAMMING LANGUAGE TRENDS - IMPLICATIONS

- SQL is used for Querying databases so its popularity makes sense. It will remain popular because no other language can be used to query databases. Despite recent popularity in NoSQL databases, most data is still stored in SQL friendly databases. On the other hand, the exponential growth in data with the growth in Big Data technologies and smart devices means that the need for storing data in such databases and the demand for professionals who can query these databases for this data will continue to grow.
- Python's popularity and the desire to learn it should not come as a surprise since it is a very user friendly programming language and easily scalable which makes it handy for development. It also has excellent capabilities for Data Analysis and Data Science with the Pandas, Numpy and Sci-kit Libraries and is very handy for Machine Learning tasks. It should rise in popularity in the future.

DATABASE TRENDS

Current Year



DATABASE TRENDS

Next Year

DATABASE TRENDS - FINDINGS

–The top 4 most popular databases among working Database professionals are MySQL, Microsoft SQL Server, PostgreSQL and SQLite.

–The 5th most popular database among working Database professionals is MongoDB, a NoSQL database.

– The most desired database for learning among respondents is PostgreSQL. The 4th most desired database for learning is MySQL.

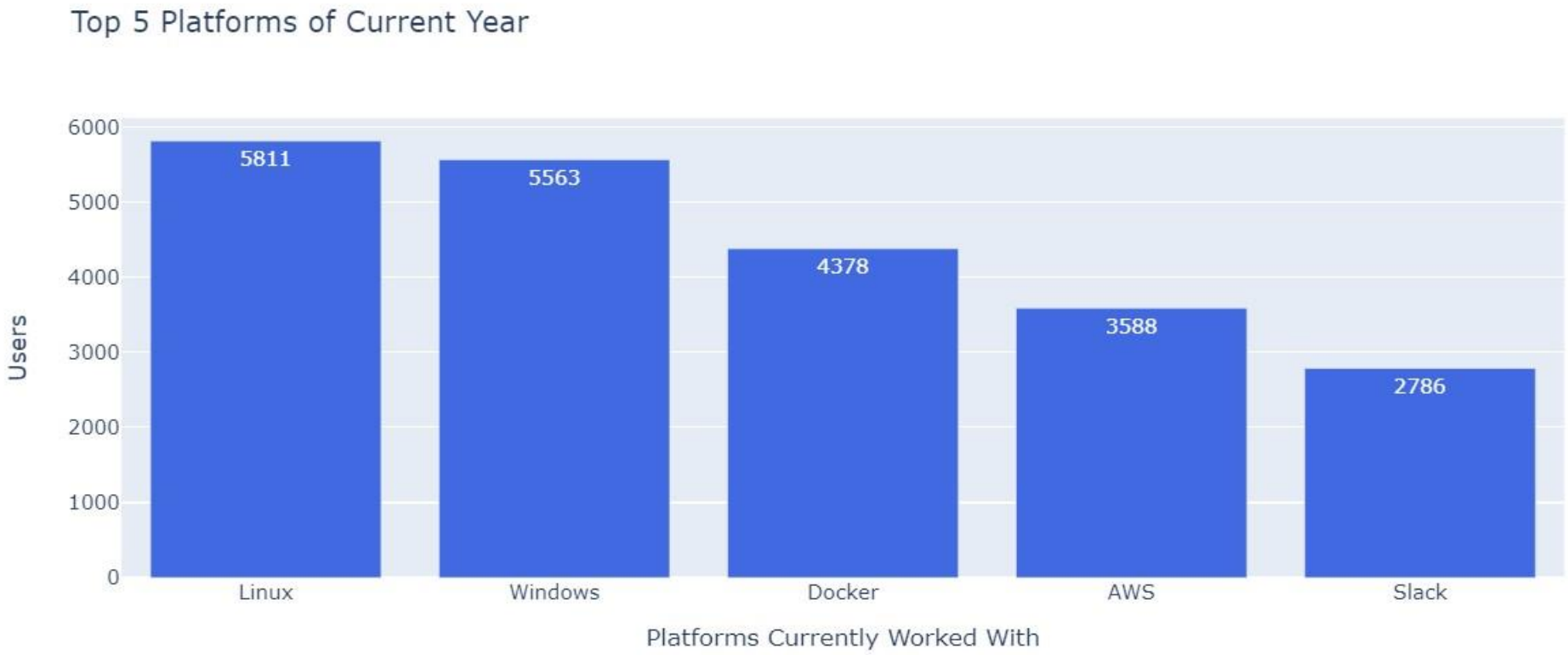
–The 2nd most desired database for learning is MongoDB. The 3rd and 5th most desired databases for learning are Redis and Elasticsearch. All 3 are NoSQL databases.

DATABASE TRENDS - IMPLICATIONS

- The findings clearly indicate that amongst working professionals, SQL based databases are popular. The top 4 out of 5 most popular databases are SQL databases. This indicates that SQL will remain popular for sometime at least and is likely to grow in demand as data storage and data querying needs grow with the increased use of smart devices and the Internet.
- Interestingly, while the top most desired database for learning and the 4th most desired database for learning are both SQL databases, the 2nd , 3rd and 5th most desired databases for learning are NoSQL databases. This indicates that while SQL will remain popular for some time, NoSQL databases might possibly supplant it in the future. Whether this turns out to be the case though is still not certain, and will warrant closer inspection of the trends. As it is SQL still remains popular and it is unlikely that current SQL databases will switch to NoSQL databases.

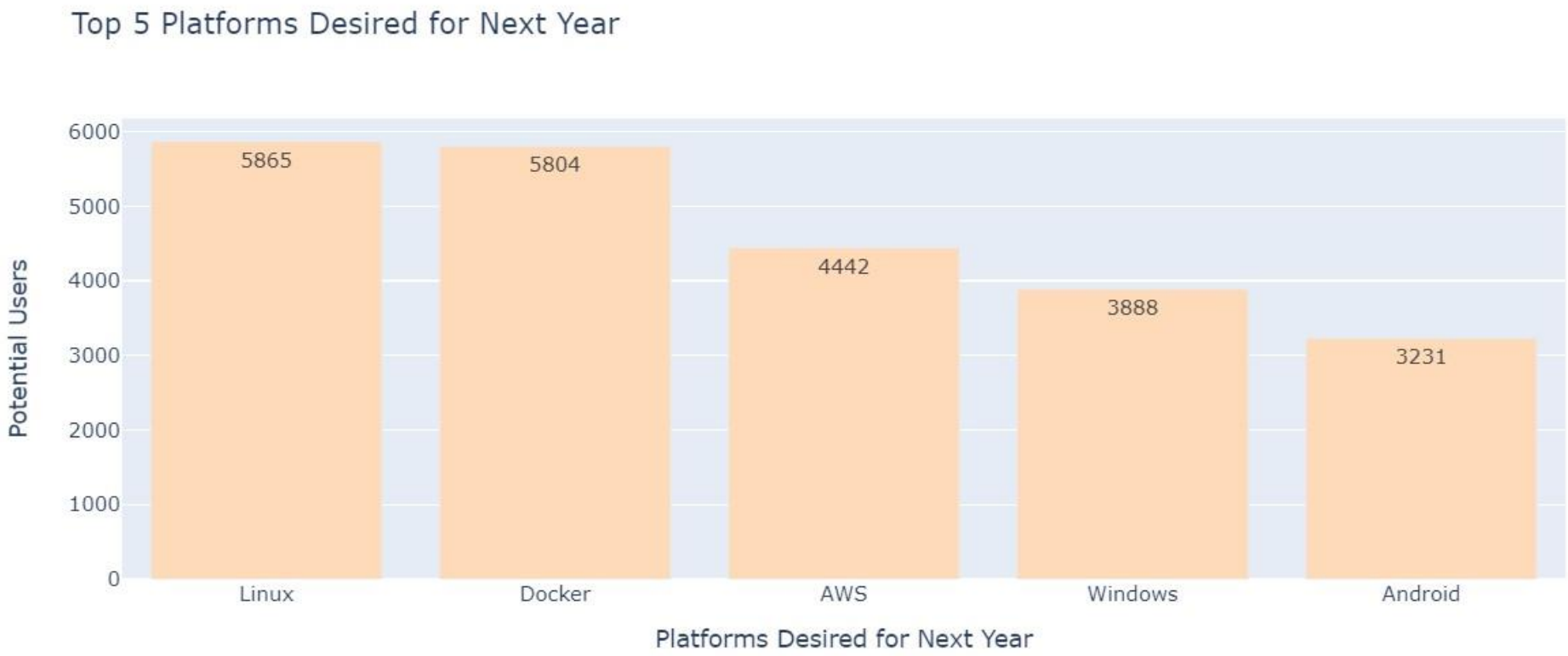
PLATFORM TRENDS

Current Year



PLATFORM TRENDS

Next Year



PLATFORM TRENDS - FINDINGS

- The most popular platform among working professionals is Linux while Windows is the 2nd. Linux is the most desired platform for learning also but Windows is the 4th most desired platform for learning.
- Docker is the 3rd most popular platform with working professionals and the 2nd most popular for learning.
- AWS (Amazon Web Services) is the 4th most popular platform with working professionals and 3rd most popular for learning.
- Slack is the 5th most popular platform with working professionals but Android is the 5th most desired platform for learning.

PLATFORM TRENDS - IMPLICATIONS

–The Linux is most popular with developers for working and for learning is not surprising because it is very customizable and thus offers considerable facility for development work. Its popularity is directly tied to the usage of Bash/Shell as a programming language.

–Similarly, while Windows remains the 2nd most popular platform for working professionals, it is only the 4th most desirable for learning. This is because the advent of new platforms with features favoring specific applications has reduced popularity of Windows.

–AWS is 4th most popular and 3rd most desirable for learning because it offers cloud based development which is gaining popularity amongst developers.

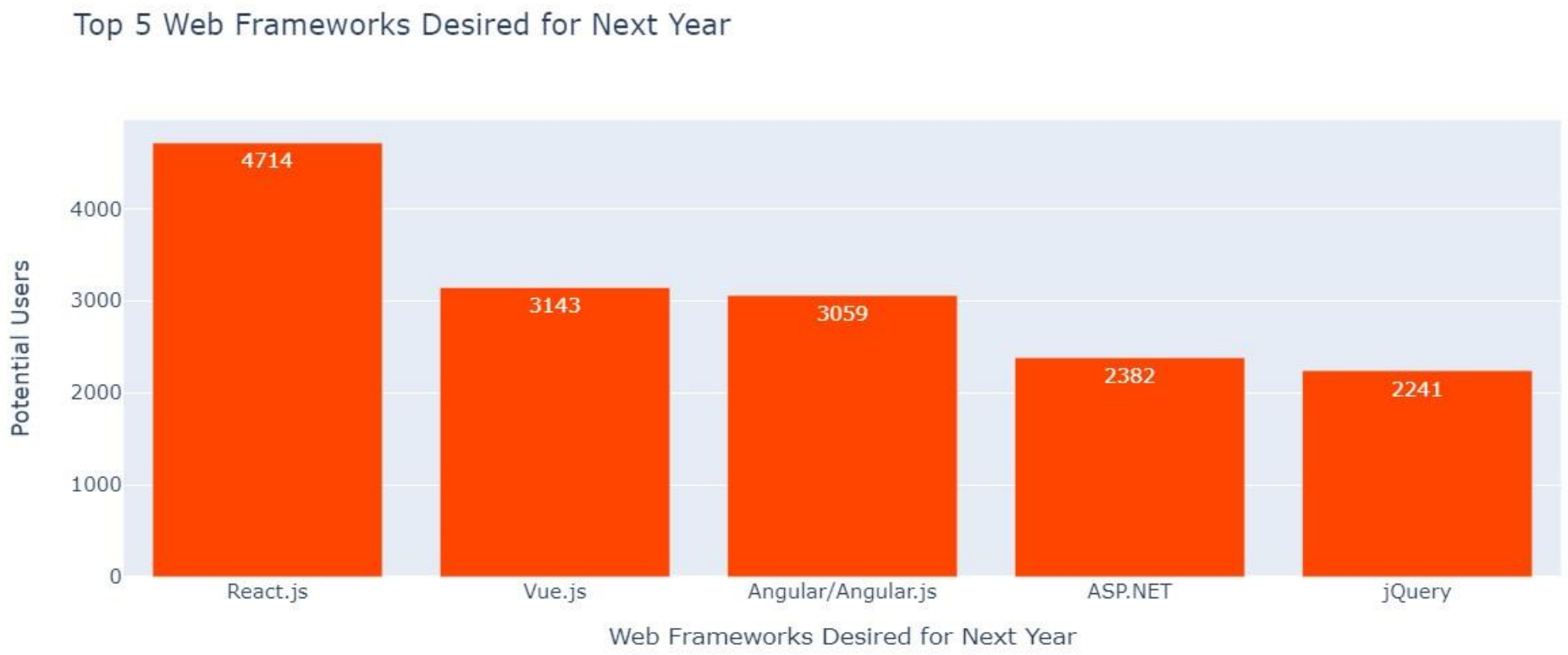
–Docker is the 3rd most popular platform with professionals and 2nd most desirable to learn because it offers tools for working with Servers and the Cloud. This makes it very versatile to use for developers and must be behind its popularity.

WEB FRAMEWORKS TRENDS

Current Year

WEB FRAMEWORKS TRENDS

Next Year



WEB FRAMEWORKS TRENDS – FINDINGS AND IMPLICATIONS

–A look at the most popular Web Frameworks and the Web Frameworks desired for learning indicates that top 4 out of 5 Web Frameworks both popular with working professionals and desirable for learning are based on JavaScript.

–The only exception is ASP.NET which uses C#.

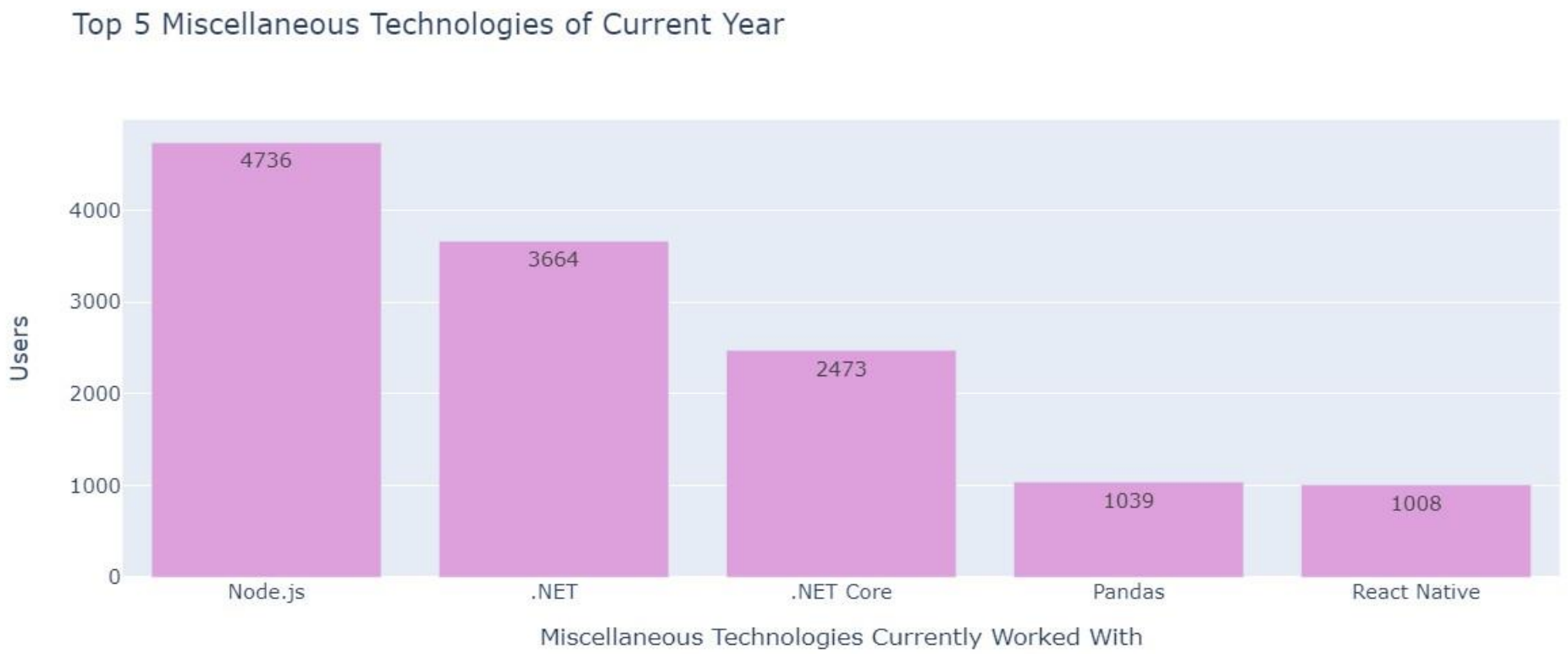
–Angular, a version of Angular.js can be used with TypeScript also.

–This by and large confirms the findings from Programming Language trends. JavaScript is the language for Web Development so Web Frameworks utilizing JavaScript also remain popular.

–Popularity of Angular may be related with the rise in popularity of TypeScript.

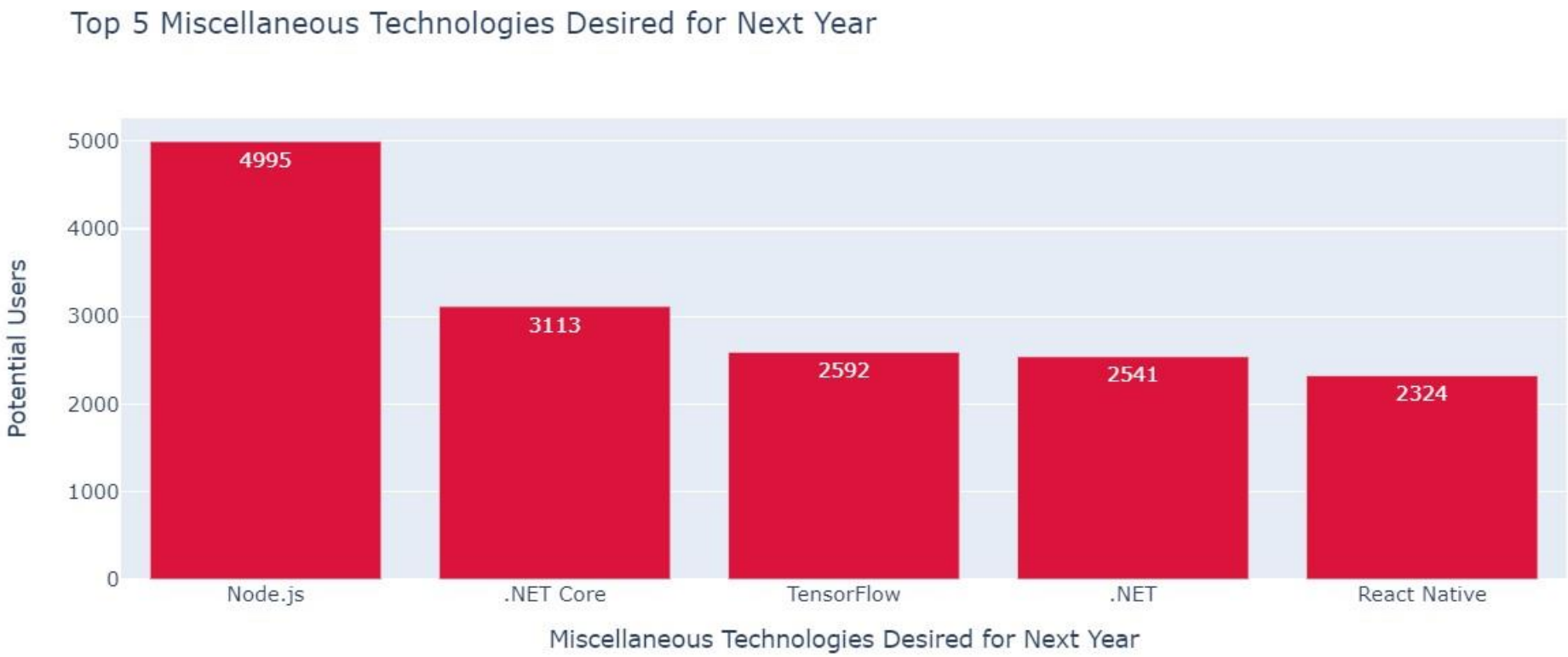
MISCELLANOUS TECHNOLOGY TRENDS

Current Year



MISCELLANOUS TECHNOLOGY TRENDS

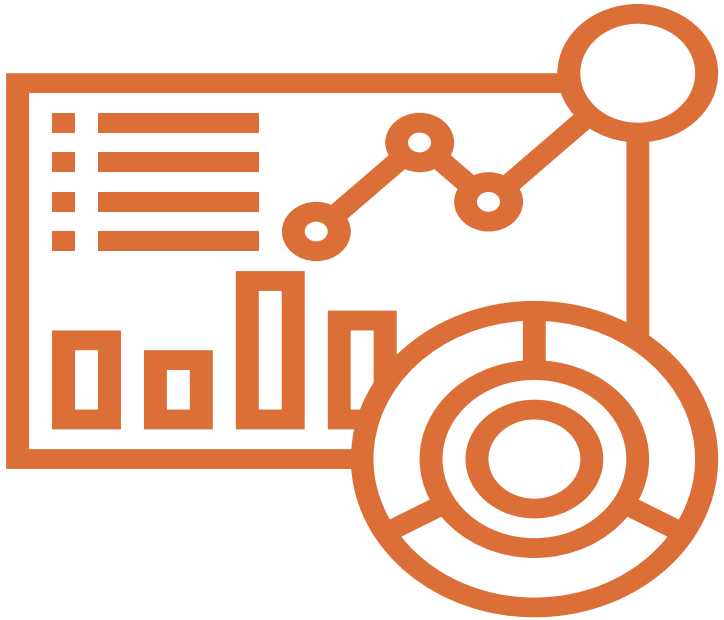
Next Year



MISCELLANOUS TECHNOLOGY TRENDS – FINDINGS AND IMPLICATIONS

- A look at the trends in miscellaneous technologies shows that Python library Pandas is 4th most popular technology with working developers. Similarly, Python based library TensorFlow is 3rd most desirable technology for learning. These libraries have uses in Machine Learning (Pandas is useful for data analysis and manipulation) and may feed into the popularity of Python as a Programming Language.
- Node.js, React native and .NET technologies form the bulk of the rest of technologies that are both popular and desirable. All these technologies are used in Web Development. This also confirms findings from trends in Programming Languages indicating continued popularity of Web Development and by extension, JavaScript.

DASHBOARD

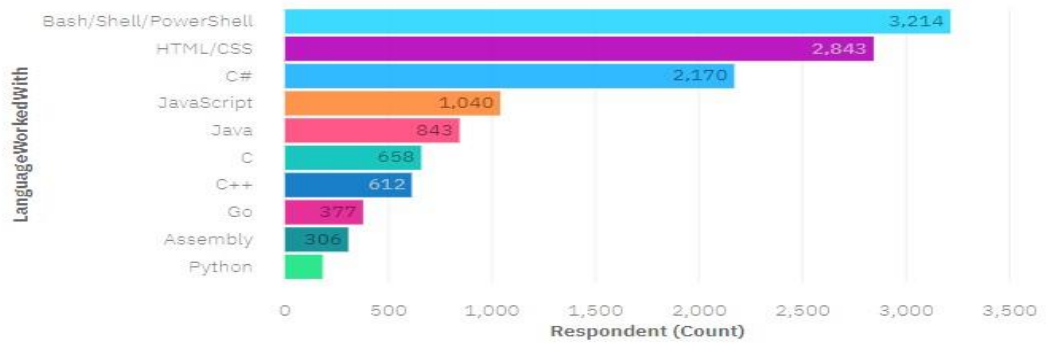


To better understand the trends for current and future technologies in different areas, as well as to understand better the demographics of the community generating these trends, a dashboard was created using Cognos Analytics to visualize various aspects of interest. The dashboard used a subset of the original dataset. The permanent link to the dashboard is given below:
https://eu-de.dataplatform.cloud.ibm.com/dashboards/cacd97fe-b475-408f-8f9f-d27c9b2db610/view/0406d4261cbb0bc714c9c8e4079028022930230eb5bb870281d17b495b602797f33f4098c879495fdb430666a7ed4658c0

DASHBOARD TAB 1

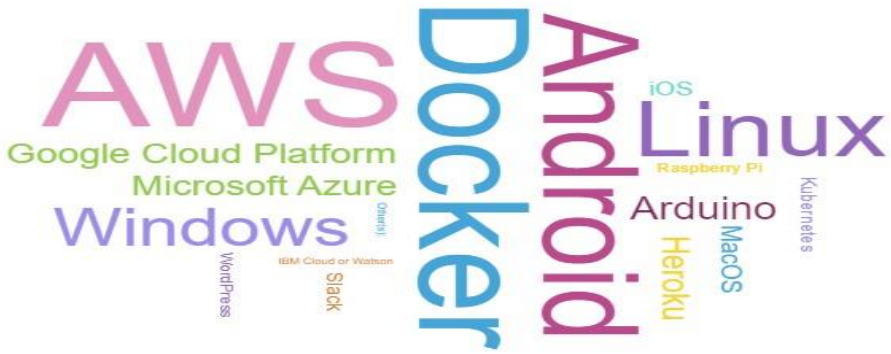
Current Technology Usage | Future Technology Trend | Demographics

Top 10 Languages Worked With

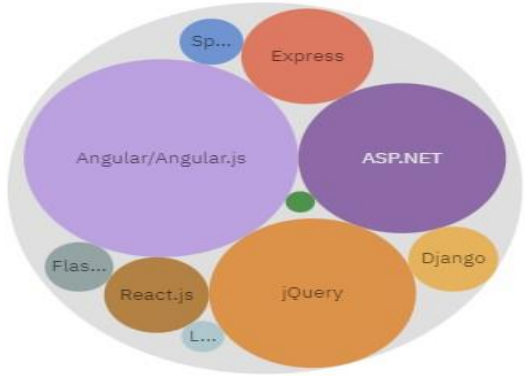


Top 10 Databases Worked With

Platforms Worked With



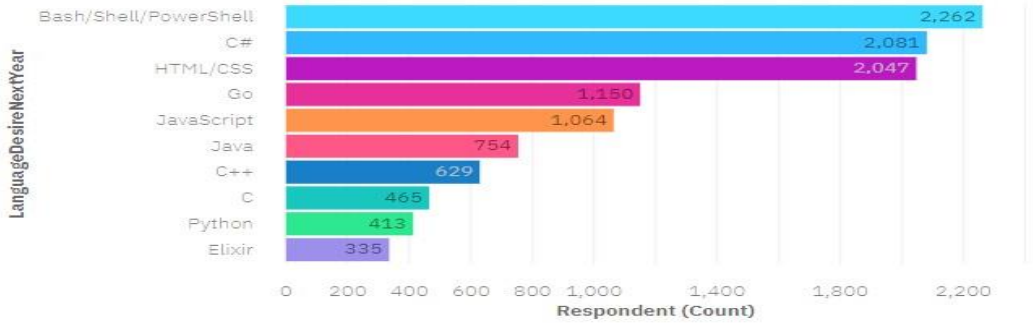
Top 10 Web Frameworks Worked With



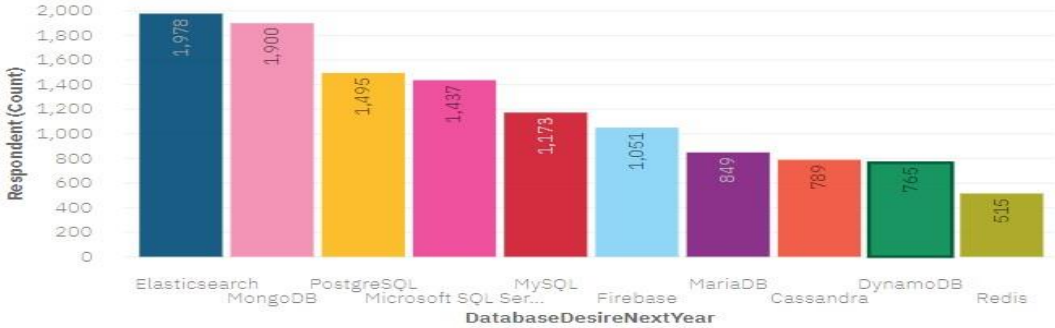
DASHBOARD TAB 2

Current Technology Usage | Future Technology Trend | Demographics

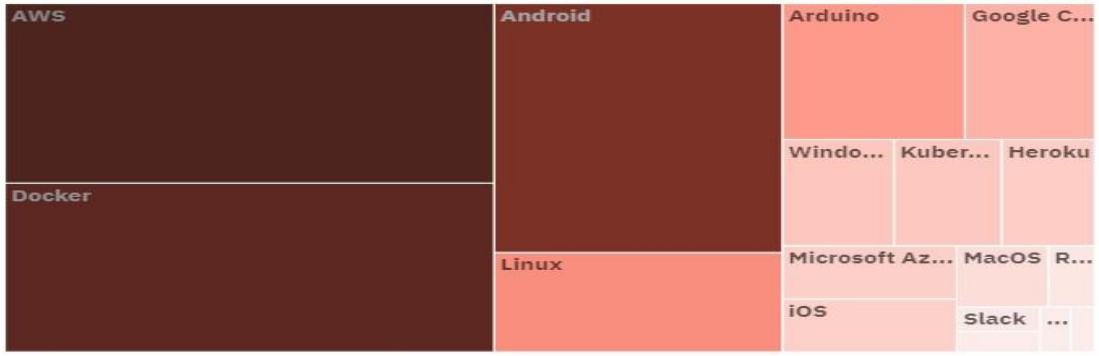
Top 10 Languages Desired for Next Year



Top 10 Databases Desired for Next Year

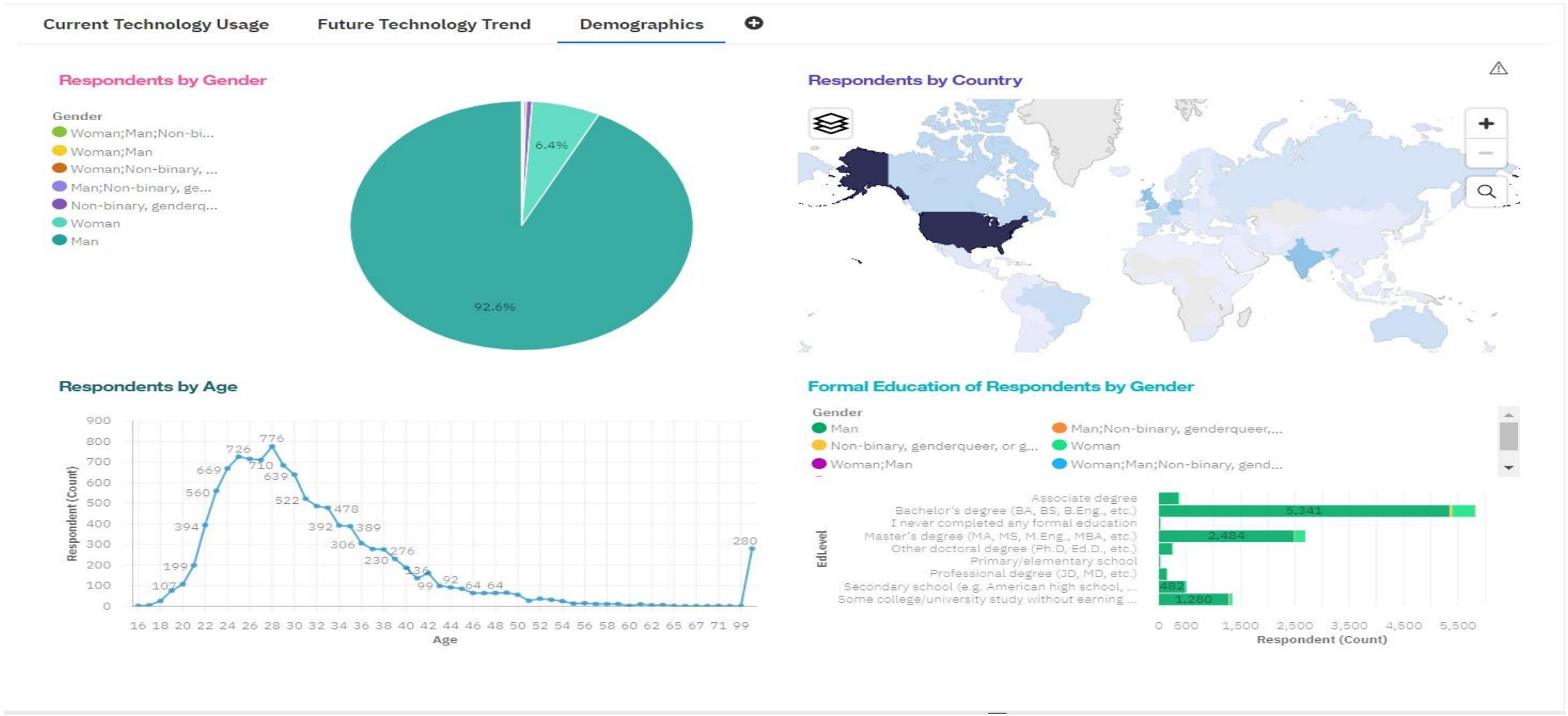


Platforms Desired for Next Year



Top 10 Web Frameworks Desired for Next Year

DASHBOARD TAB 3



DISCUSSION - DASHBOARD

The Dashboard and the visualizations together provide some interesting points to ponder. An important distinction must be made between the two. The dashboard uses a smaller subset of the survey data therefore its findings cannot be generalized. But the dashboard provides some interesting information about the demographics of the survey.

- One key point to note is that the vast majority of respondents seem to come from the United States, Great Britain, Germany and other European countries. Only India has a notable contribution from the developing world. China, the 2nd largest economy in the world and Japan, the 4th largest economy in

DISCUSSION - DASHBOARD

the world do not have much representation in the survey. So these results must be taken while keeping in mind this bias.

- The vast majority of the respondents to the survey are male. Females make up less than 7% of the total respondents. This makes for a sad commentary on the state of STEM education in general and on the developer world in particular. One hopes that future trends will see greater participation from females in development and the survey.
- The majority of the respondents to the survey are between 20 to 42 years old. This bodes well for development as it seems to hold considerable promise for the younger generation. It should see growth in the years

DISCUSSION - DASHBOARD

to come if this trend holds, spurring development of new technologies and possibly leading to shifts in trends. As an aside, the jump at 99 years appears to be an outlier.

- The vast majority of the respondents held a Bachelors Degree as their education level while the 2nd largest group held a Masters degree. This indicates that a formal education background is considered necessary by most respondents to be a developer. This is confirmed by the relatively 3rd largest group having some college experience. Will this trend shift in the coming years as online modes of education that do not require formal schooling but focus on specific skill development become popular?

DISCUSSION - DASHBOARD

- Finally, the trends displayed for the subset of survey data in the dashboard vary from the trends displayed in the prior visualizations. Perhaps they are more reflective of the trend among developers in the North American and European region? This could have implications. As it is, Python is only 10th in popularity and 9th in desirability to learn according to these results while SQL is not in the Top 10 at all. JavaScript and CSS/HTML are in the top 5 though.

OVERALL FINDINGS & IMPLICATIONS

- The survey results indicate that Web Development appears to be the most in-demand technology as JavaScript and CSS/HTML remain the top two most popular and desired Programming languages. The 5th most desired programming language, TypeScript is also related to Web Development.
- Web Frameworks indicate the continuing popularity of JavaScript for Web Development as 4 out of 5 most popular and desirable Frameworks are based on JavaScript.
- Among Miscellaneous Technologies, many Web Development technologies are listed. The most popular and desirable technology is Node.js, which is based on JavaScript, reinforcing its popularity. The 5th most popular and desirable technology is React Native, also based on JavaScript.

OVERALL FINDINGS & IMPLICATIONS

- The rising popularity of cloud based platform like Amazon Web Services (AWS) and those that can be integrated with cloud like Docker (or Docker Cloud) indicates the rising popularity of cloud development.
- SQL, because of its role in querying data from databases remains popular as well and it is the 3rd most popular language. It is also the 4th most desirable language.
- 4 out of 5 most popular databases are based on SQL which reinforces the trend for its popularity. However, many Database professionals desire to learn NoSQL databases as opposed to SQL ones. Of the 5 most desirable databases, only 1st and 4th are SQL based. The 2nd, 3rd and 5th are NoSQL databases.

OVERALL FINDINGS & IMPLICATIONS

- Python is rising in popularity and desirability. This may be due to its intuitive syntax, easy to learn nature and scalability. It also has good libraries for Machine learning which might be driving its popularity up as well. As it is, it is the 5th most popular language. It is also the 3rd most desirable language.
- Among miscellaneous technologies, Python library Pandas is the 4th most popular technology which is used for Data Manipulation and Data Analysis. This feeds into Python's popularity as well.
- Furthermore, the 3rd most desirable technology, TensorFlow, is a machine learning platform based partly on Python. It interfaces with Python really well and many of its models are written in Python also. This too contributes to Python's popularity.
- Finally, as cloud based Platforms gain in popularity, Bash/Shell will decline in popularity, allowing Python to become at least 4th most popular programming language.

OVERALL FINDINGS & IMPLICATIONS: FINE POINTS

- The popularity and desirability of technologies should be taken differently. Popularity is based on continued usage and thus is more robust. Desirability is based on various factors. Some of these may be influenced by trends as well as what appears popular. Desirability may not necessarily translate into future popularity. Developers may not necessarily act on their desires of learning technologies. So desirability should be taken with a grain of salt.
- Type of technology is also important to consider. Databases, for instance are not chosen by developers, but by companies and other players in the Database sub-sector. Most SQL databases are unlikely to switch over to NoSQL databases. Thus, even if many developers for Database may desire to learn something (like NoSQL), they are less likely to be working with that technology since other players determine the technology used in Databases. This is true for other technology sub-sectors as well to a certain extent.

OVERALL FINDINGS & IMPLICATIONS: DEMOGRAPHICS

–The demographic information of the survey respondents indicates some important information. For one, it appears that the majority of the respondents are from the United States, United Kingdom, Germany and other European countries. Asian countries except for India have little representation.

–The majority of the respondents are Male with barely 7% respondents identifying as Female. This is a concerning trend and it is hoped that steps are taken to make the Technology sector more inclusive and friendly for Women.

–The majority of the respondents are between 20 to 42 years old. This indicates that technology sector interests the younger generation. This bodes well for the future and there should be more growth in this sector in the future.

–Most respondents have formal education till at least Bachelors. Whether this trend holds as more online resources make technology easier to learn is a trend to keep an eye on.

CONCLUSION

In this project, data from Stack Overflow Developer Survey 2019 was used. The project made use of visualizations created from the data to model the trends in Programming Languages, Databases, Platforms, Web Frameworks and Miscellaneous Technologies. It also made use of demographic information from the survey to shed light on important characteristics of the participants in the survey and through them the demographic nature of the technology world.

The results from the project indicated strong support for JavaScript and HTML/CSS. Both of these Programming Languages were considered the most popular and the most desirable to learn by the participants. This indicated a strong trend in Web Development as both these languages are used for Web Development. The results also indicated strong desirability for learning TypeScript, an extended version of JavaScript used in Application Development as well. These trends were reinforced by domination of JavaScript based frameworks in Web Frameworks. Miscellaneous Technologies such as Node.js involving JavaScript are also popular. Thus, learning JavaScript the most important future skill for Web Developers.

CONCLUSION

Among Databases, SQL remains the most popular language and 3rd most popular programming language. Most popular Databases are SQL based. Although there are strong trends of Database developers showing a desire to learn NoSQL databases, it seems unlikely that SQL will wane in popularity in the near future as data generated is increasing exponentially and the concurrent need to store and query this data will require SQL as established databases are unlikely to switch to NoSQL databases. Thus, learning SQL is an important future skill for developers wishing to interact with and work on databases.

Among programming languages, Python is the 5th most popular language and 3rd most desirable language. This reflects the ease of use of Python, its scalability, use in Machine Learning and its ubiquitous presence in the technology sector. Furthermore, various popular and desirable Miscellaneous Technologies such as Pandas and Tensorflow are based wholly or partly on Python, contributing to its popularity. Thus, programming in Python is a very useful skill for future developers to learn. Finally, with popularity of cloud based platforms rising, Bash/Shell is declining with Python likely to take its spot as 4th most popular Programming Language at least.

CONCLUSION

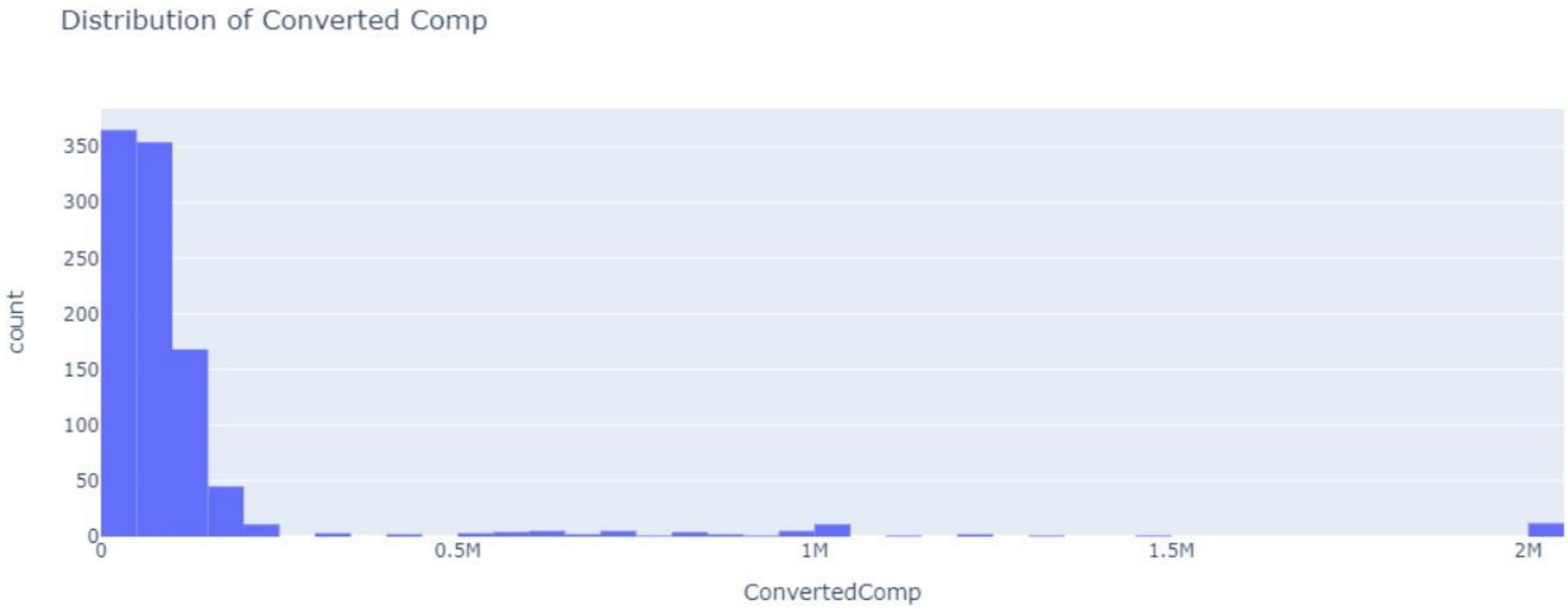
Finally, a few points became clear from the demographic information of the survey participants:

- Most respondents, about 93%, are male. This means that women are not particularly represented in the technology sector which is concerning and might merit some thinking.
- Most of the responses are from participants in North America and Europe. Asia, with the exception of India, appears to be underrepresented.
- Most respondents are young, between 20 and 42 years old, which means technology is popular among the younger generation and this sector should see more growth in the future.
- Most respondents have a formal education till at least a Bachelors degree. Whether increased use of online resources for learning technology will change this is a something to keep an eye on.

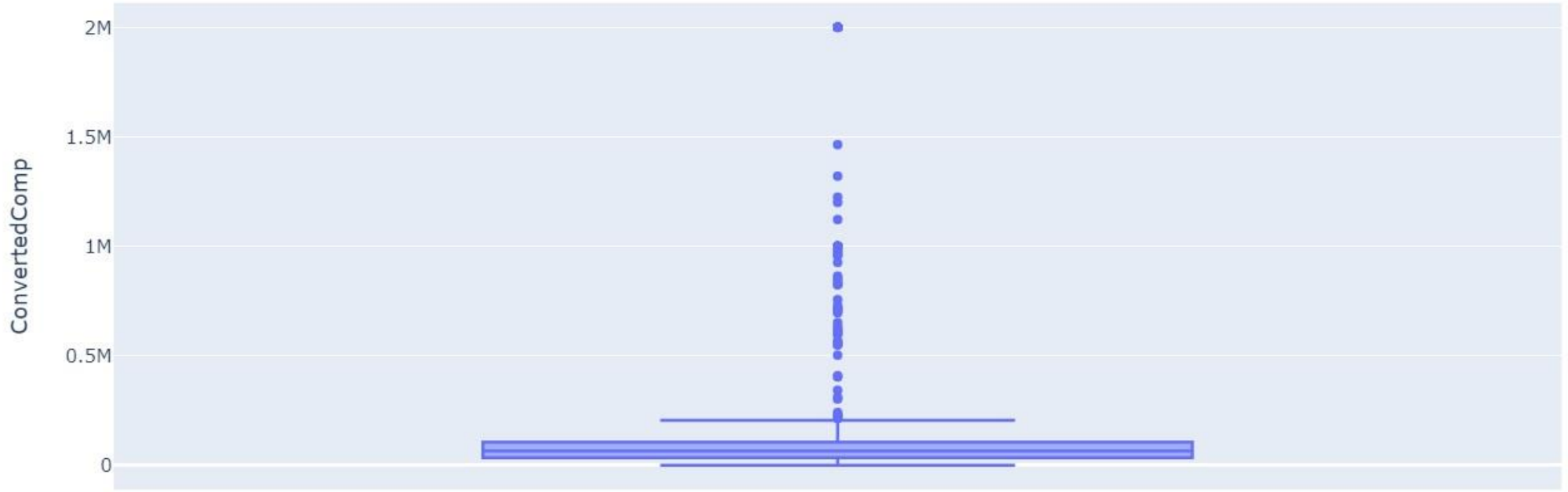
APPENDIX

Includes several additional charts created during the analysis phase of the project as well as charts from alternate data collected through other means such as APIs and Webscraping.

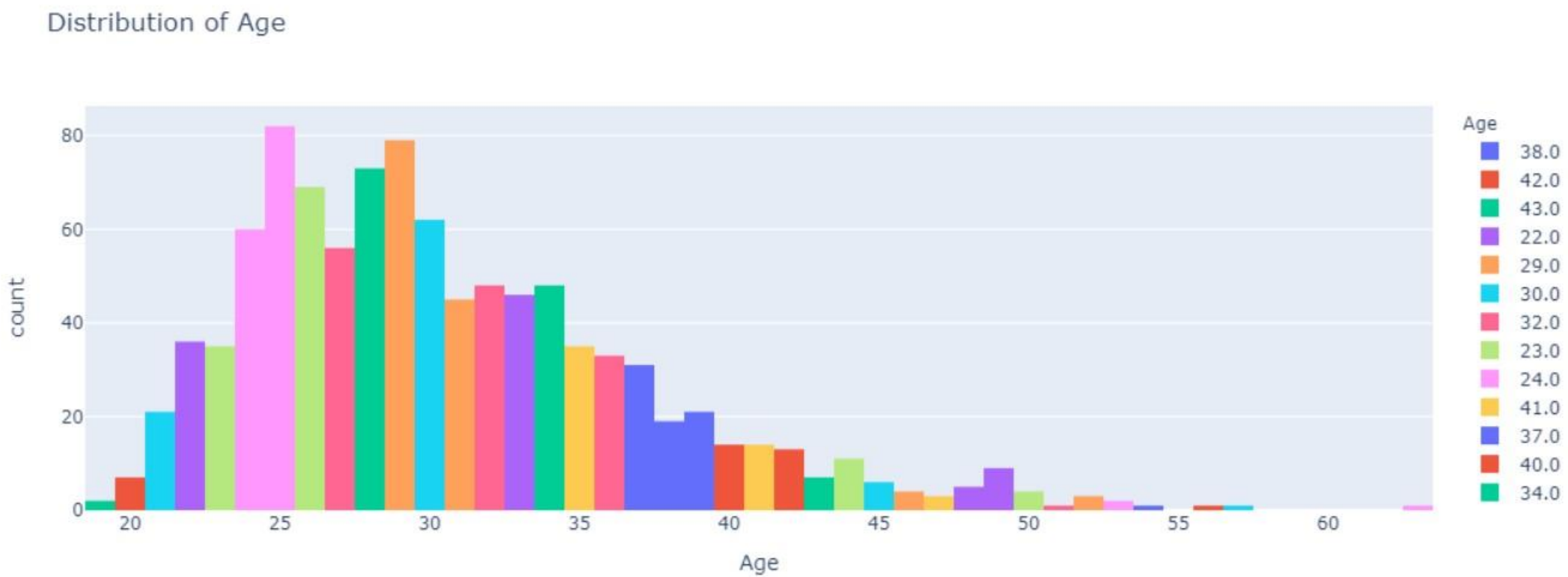
APPENDIX-COMPENSATION DISTRIBUTION



APPENDIX-COMPENSATION DISTRIBUTION

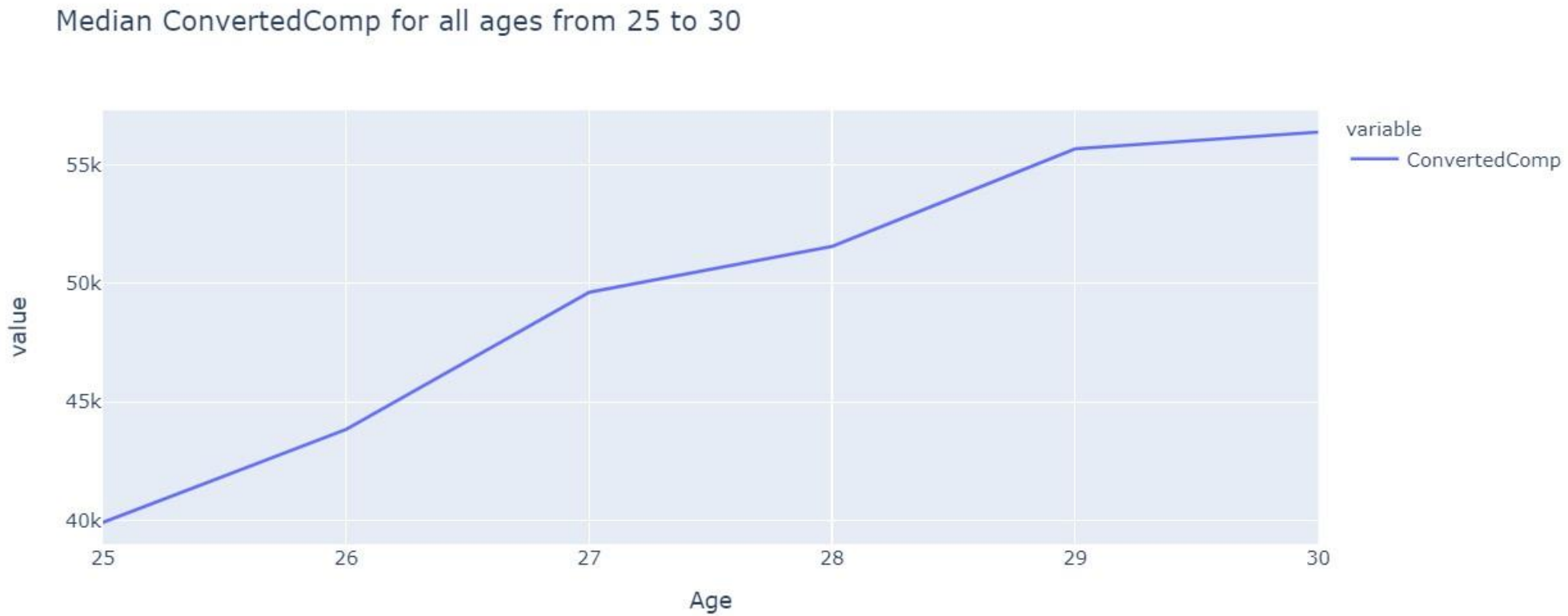


APPENDIX-AGE DISTRIBUTION

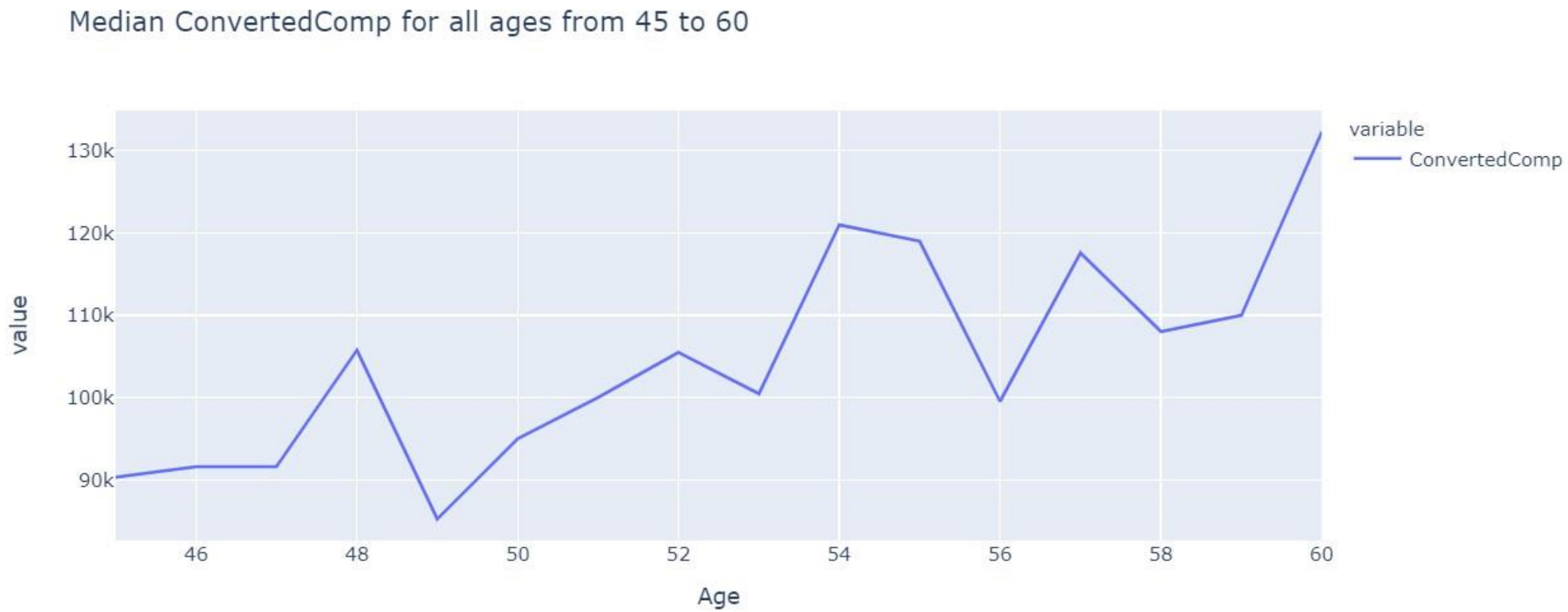


APPENDIX-AGE DISTRIBUTION

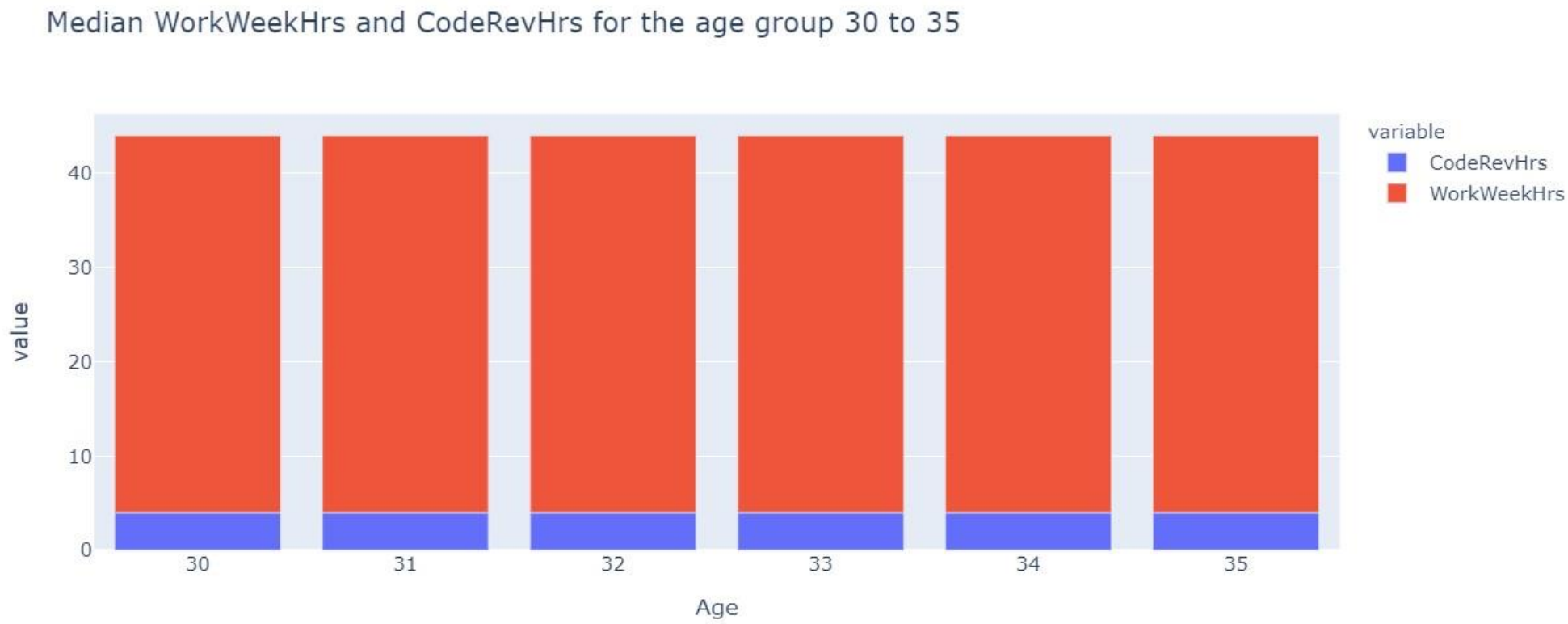
APPENDIX-COMPENSATION VS AGE



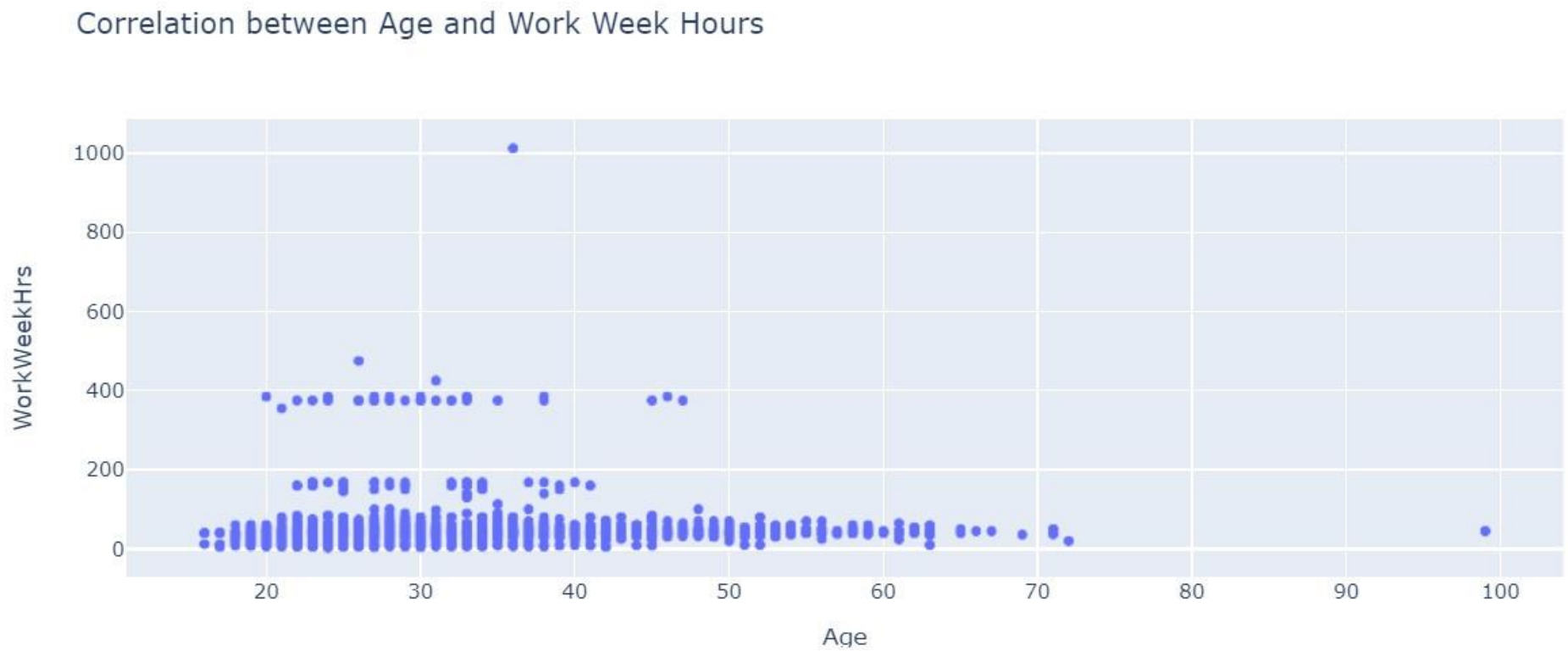
APPENDIX-COMPENSATION VS AGE



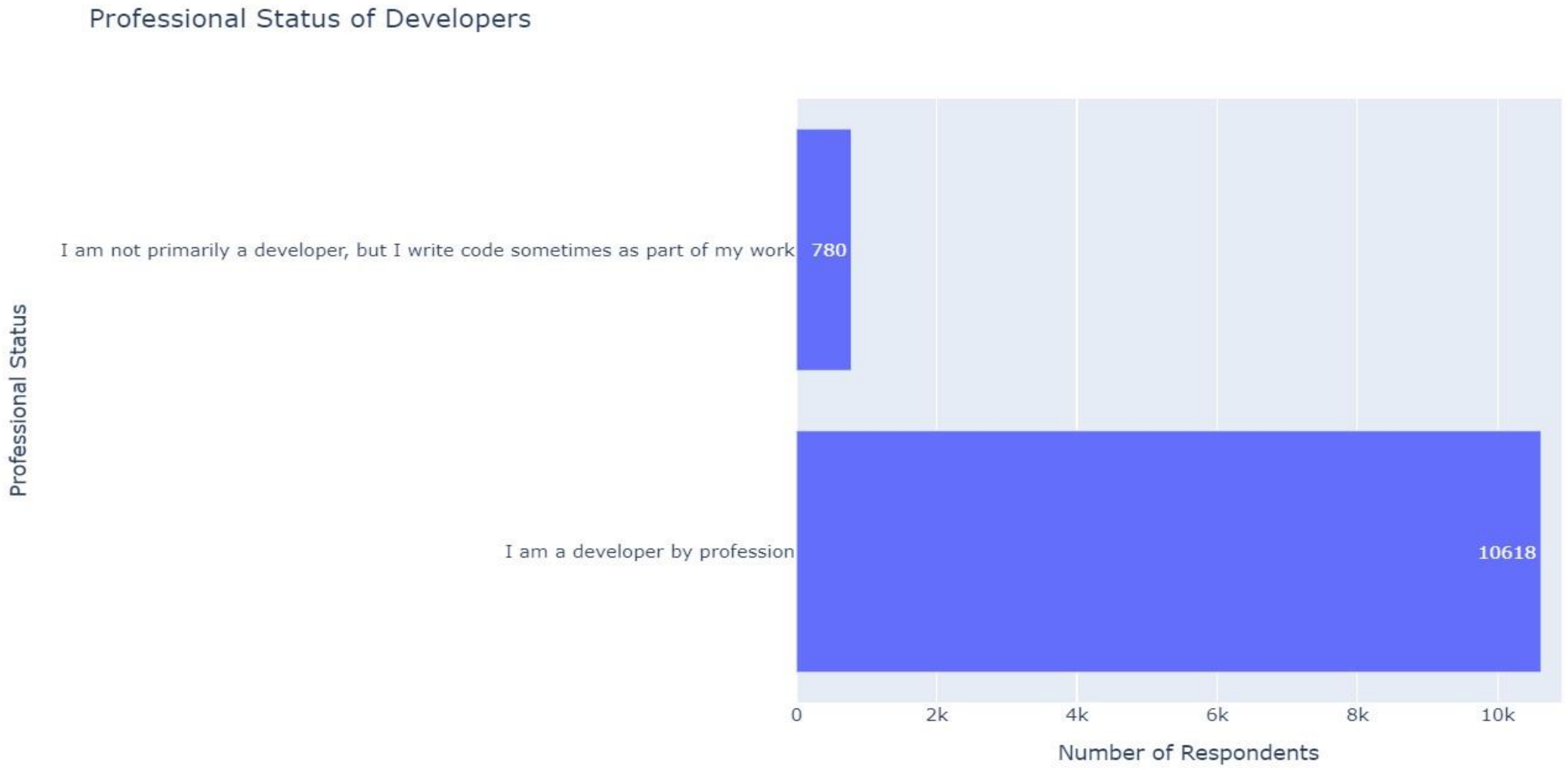
APPENDIX-DEVELOPER WORK DISTRIBUTION



APPENDIX-AGE VS WEEKLY WORK HOURS



APPENDIX–DEVELOPER PROFESSIONAL STATUS

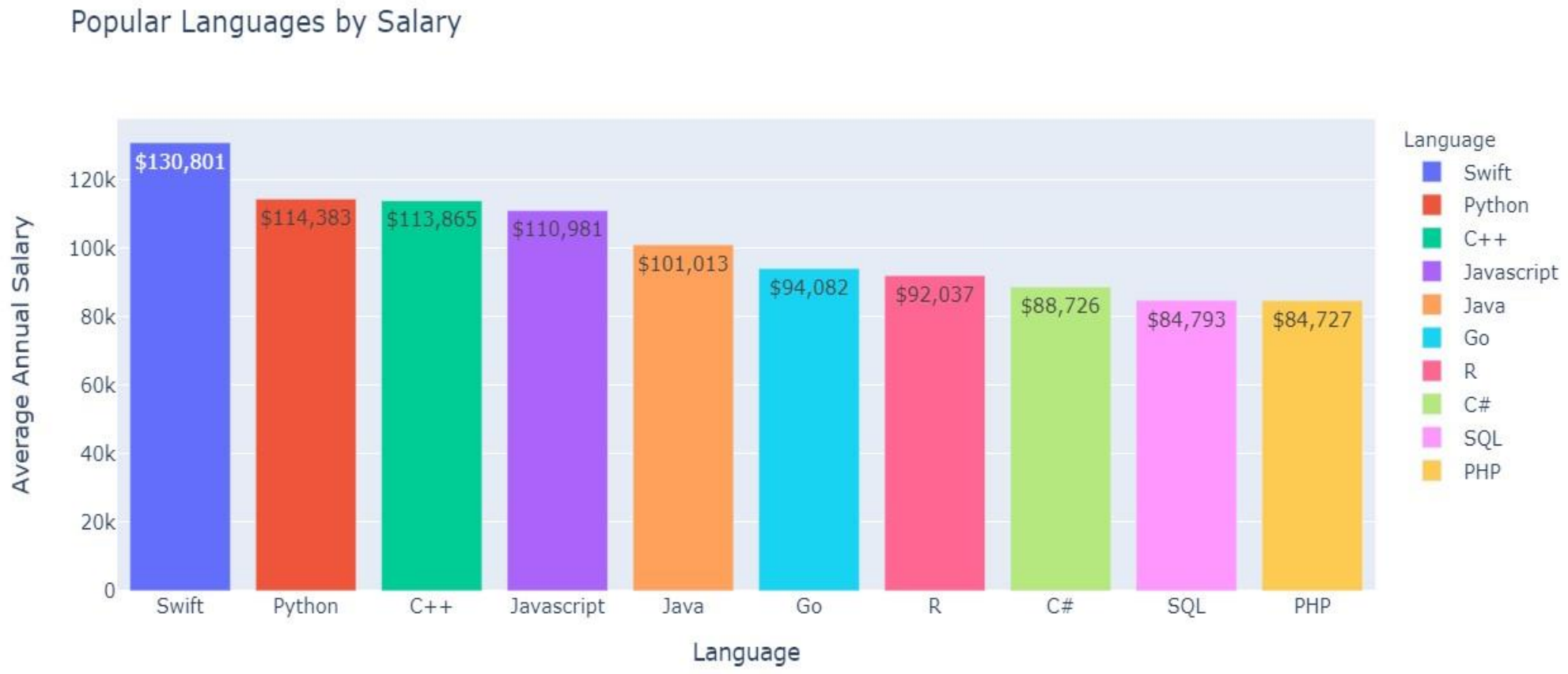


APPENDIX-API AND WEBSCRAPING DATA

- In addition to the dataset used for analysis, data from webscraping and API was also acquired.
- The first graph uses API data to compare the number of tech jobs in major US cities. Interestingly, Washington DC beats out all other cities in this regard.
- The second graph uses Webscraped data to rate Programming Languages according to the Average Annual Salary of the programmers for that language. Interestingly, Swift (used for iOS development) developers are the most highly paid.
- Python developers rank as the 2nd most highly paid. JavaScript developers rank as the 4th most highly paid. But the difference of salary between these two is negligible (less than 5%).
- SQL developers are the 9th most highly paid.

APPENDIX-API DATA-JOB POSTINGS

APPENDIX-WEBSCRAPING DATA-POPULAR LANGUAGES



REFERENCES

- (Techjury, 2022), “How Fast Is Technology Advancing in 2022?” *Techjury*, https://techjury.net/blog/how-fast-is-technology-growing/. Accessed 10 Sept. 2022.