



# IBM Data Analyst Capstone project : Sharing the story of analysis on Emerging technologies and trends

Rutvi Gajipara

November 27th, 2023

# OUTLINE



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

# EXECUTIVE SUMMARY

- ▶ To be competitive in the global IT sector, It's essential to keep up with the ever-changing technologies. The report uses data analytics to highlight current and projected trends in the need for skills related to programming languages, databases and other technologies. It also studies the demographics of professionals in the technology sector.
- ▶ Data was gathered from a Stack overflow survey, IBM site, GitHub job postings. It was collected, Cleaned, and subjected to exploratory analysis and visualized on dashboards.

# INTRODUCTION

- This presentation report uses data analytics highlights current and projected trends in the need for skills related to programming languages, databases, platforms and web frames.
- The following inquiries were investigated using the data :
  1. Which programming languages are most in demand today?
  2. What are the most in-demand skills?
  3. What popular IDEs or Web frames are there?
- The target audience for this research are IT professionals, HR managers, and anybody else with an interest in the IT sector who want to learn about the top on-demand IT skills which will be relevant in the future.

# METHODOLOGY

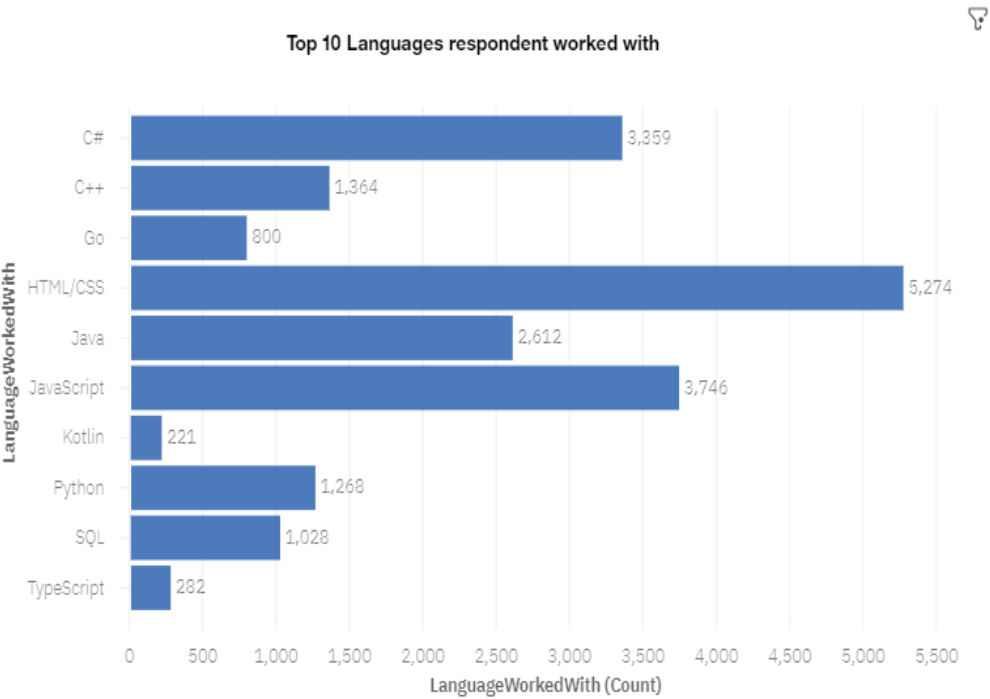
- ▶ Data in several formats, such as the number of jobs currently available for different technologies and for different places, were gathered using the GitHub jobs API on python.
- ▶ To obtain the names of the programming languages and their yearly wages, the IBM website was scrapped.
- ▶ Python was used to clean and analyze the data. To access the distribution of data, the presence of outliers and the correlation between columns, and exploratory data analysis was carried out.
- ▶ Charts, Graphs and Dashboards were created using Python and Cognos Analytics to visualize the data.

RESULTS

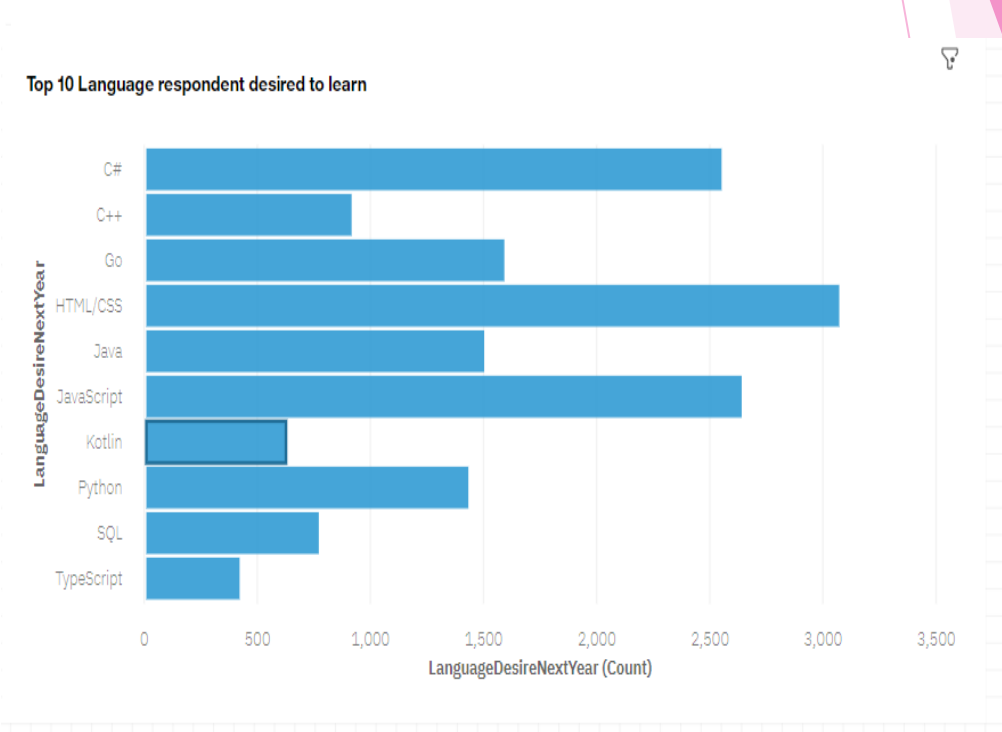


# PROGRAMMING LANGUAGE TRENDS

## Current Year



## Next Year



# PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

## Findings

- ▶ JavaScript, HTML/CSS, SQL and Python are the most used languages currently
- ▶ JavaScript, HTML/CSS, Python, and Typescript will be the most used languages next year and future years
- ▶ Python will have more demand than SQL next year.

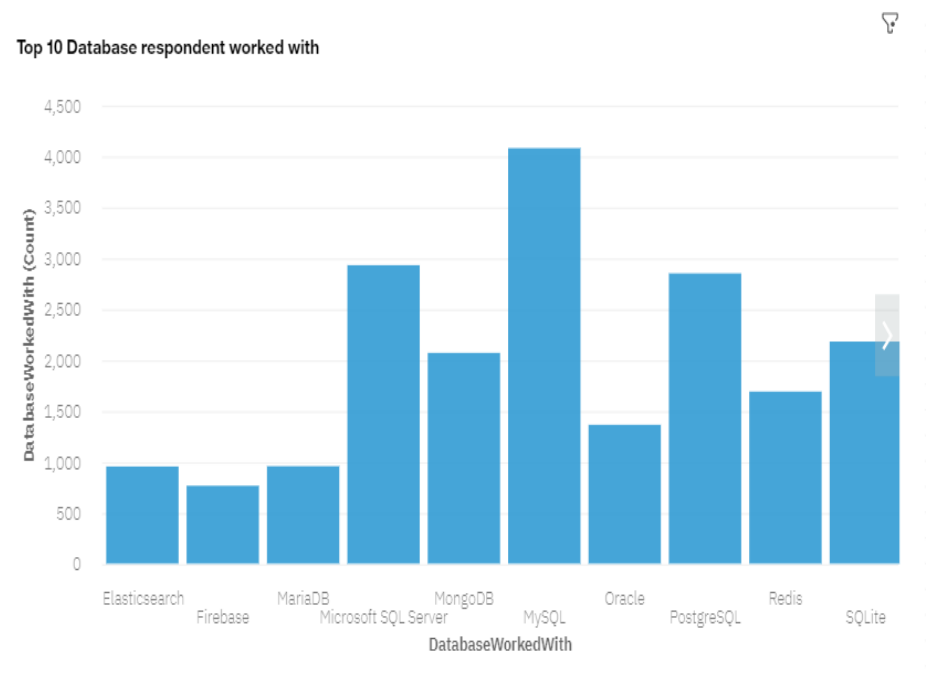
## Implications

- ▶ Web development as a tech skill has the highest demand, especially as Typescript is getting viral
- ▶ Python is gaining more and more traction due to the increase in demand for AI and ML skills
- ▶ SQL is the most relevant language for Data professionals.

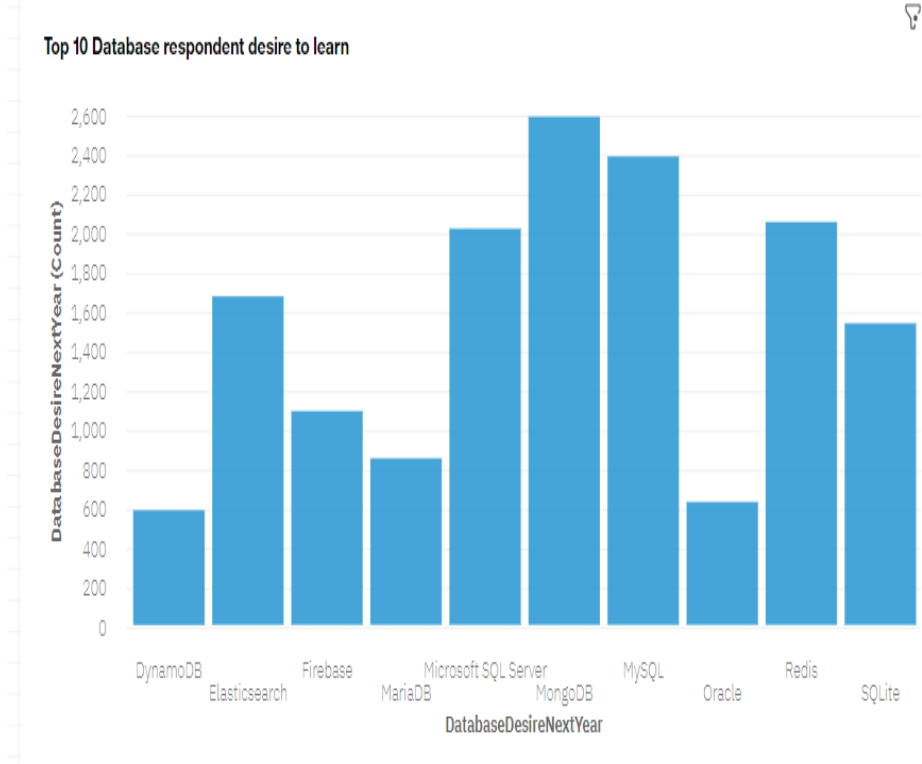


# DATABASE TRENDS

## Current Year



## Next Year



# DATABASE TRENDS - FINDINGS & IMPLICATIONS

## Findings

- ▶ MySQL, MS SQL server, Postgre SQL, SQLite and MongoDB are the top 5 most used databases at the moment.
- ▶ However, MongoDB, Redis and Elasticsearch are projected to become more popular in the future.
- ▶ Redis and Elasticsearch are relatively new tools and are set to gain more traction in the IT Space.

## Implications

- ▶ SQL is still a top tool to watch out for in data specialists.
- ▶ Companies still prefer open-source databases.
- ▶ Oracle SQL was not among top 5. It is losing relevance as time pass.

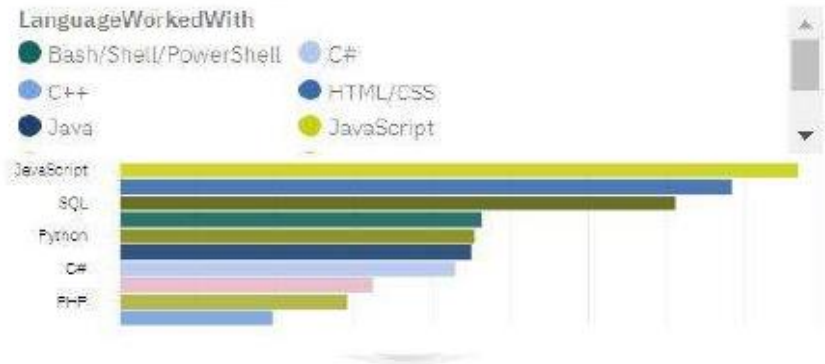
# DASHBOARD



<https://dataplatfrom.cloud.ibm.com/dashboards/81f87faa-6187-48fb-acf4-134789249059/view/5b23de0214bf71e21eb5c0e4079b255329372c59e1bbd005808d7b4908667297a8614692c82c430edf475635a7e41b5cc9>

# DASHBOARD TAB 1

## Top 10 LanguageWorkedWith



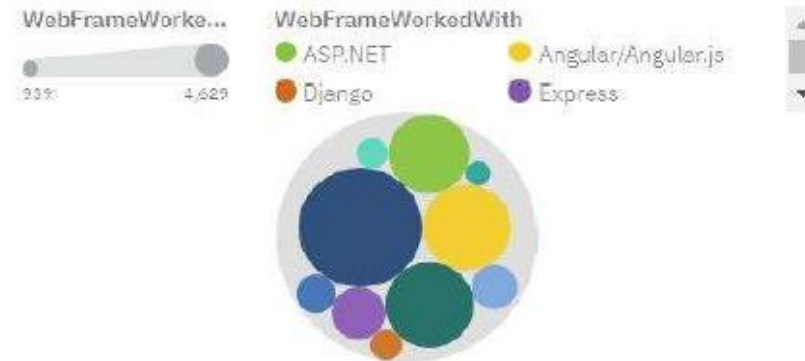
## Top 10 DatabaseWorkedWith



## PlatformWorkedWith

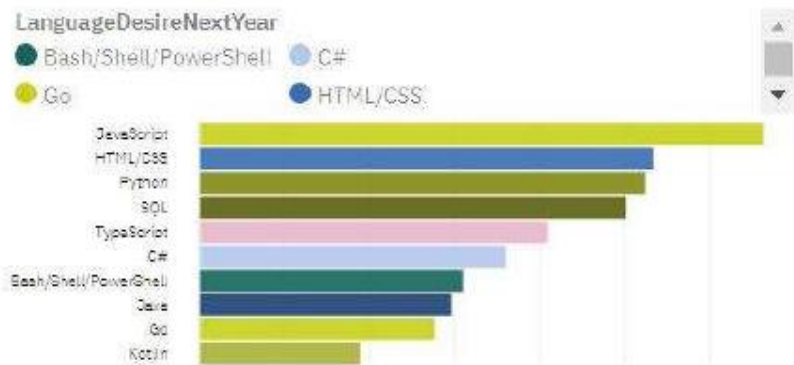


## Top 10 WebframeWorkedWith

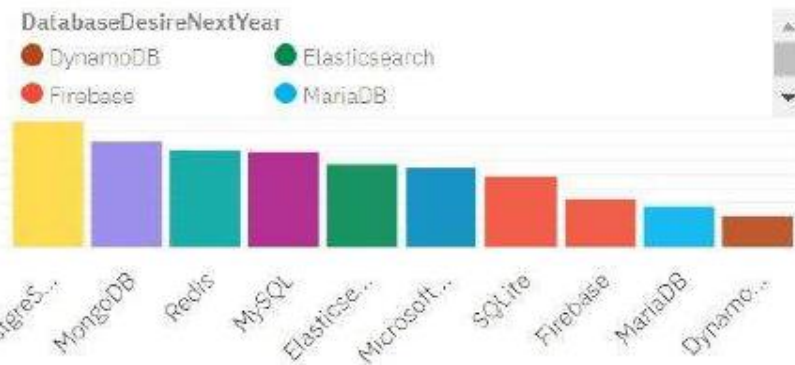


# DASHBOARD TAB 2

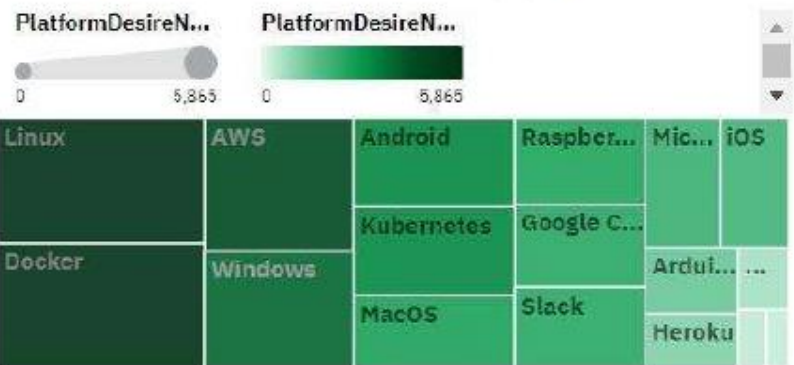
## Top 10 Language Desire Next Year



## Top 10 Most Desired Databases



## Most desired Platforms next year



## Most Desired Web frames Next Year



# DASHBOARD TAB 3

Percentage of Respondents by gender

Gender  
Woman Man

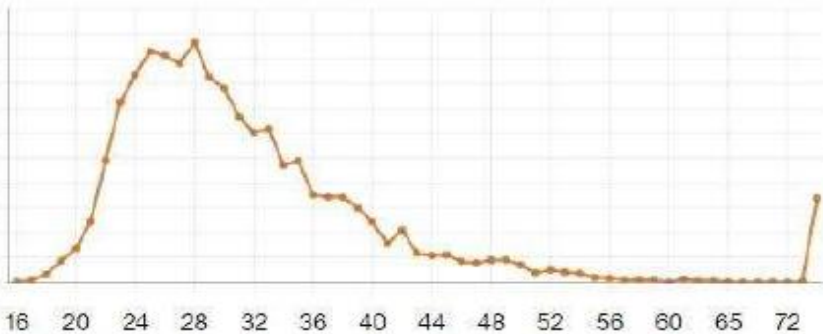


Respondents by Country

Respondent (Sum)



Respondents by Age



Respondents by Formal Education Level

Education level  
(no value)  
Associate degree



# DISCUSSION



- Upskilling in the tech sector
- How do we close the wide gender gap in the tech sector?
- The increasing demand for mobile development as Kotlin is getting popular
- How relevant will Oracle SQL still be in the future?

# OVERALL FINDINGS & IMPLICATIONS

## Findings

- ▶ Most people in IT field, have a Bachelor's degree
- ▶ The tech sector is filled with majorly young people under 40 years
- ▶ Most respondents want to learn Postgre SQL and React JS next year

## Implications

- ▶ It is important for Data professionals to develop proficiencies in NoSQL in addition to SQL databases.
- ▶ Web development is still a very lucrative skill
- ▶ Less developed countries need more access to tech training and education.



# CONCLUSION

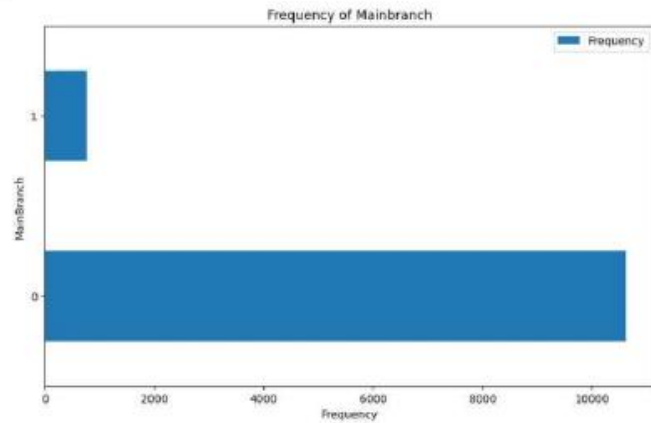


It is expedient to stay updated in the tech sector as the trends keep changing over time.

# APPENDIX

	MainBranch	Frequency
0	I am a developer by profession	10010
1	I am not primarily a developer, but I write code	780

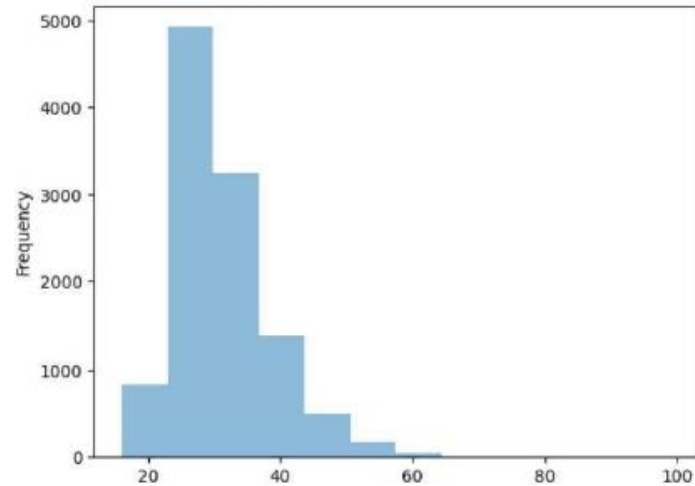
```
bar_df.plot(kind='bar', figsize=(10,6))
plt.xlabel('Frequency')
plt.ylabel('MainBranch')
plt.title('Frequency of Mainbranch')
plt.show()
```



Plot a histogram of the column 'Age'.

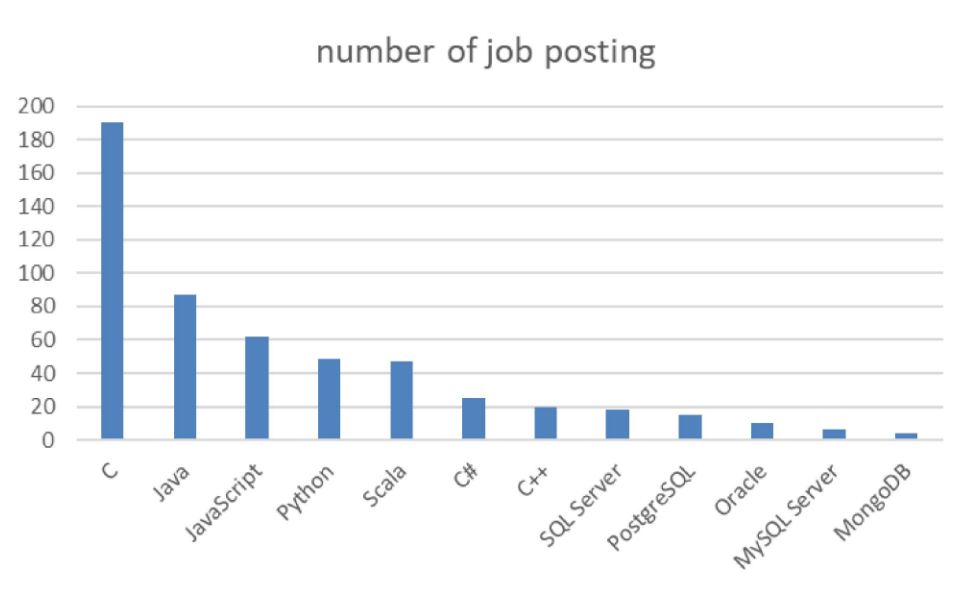
```
df['Age'].plot.hist(bins=12, alpha=0.5) #
```

<AxesSubplot: ylabel='Frequency'>



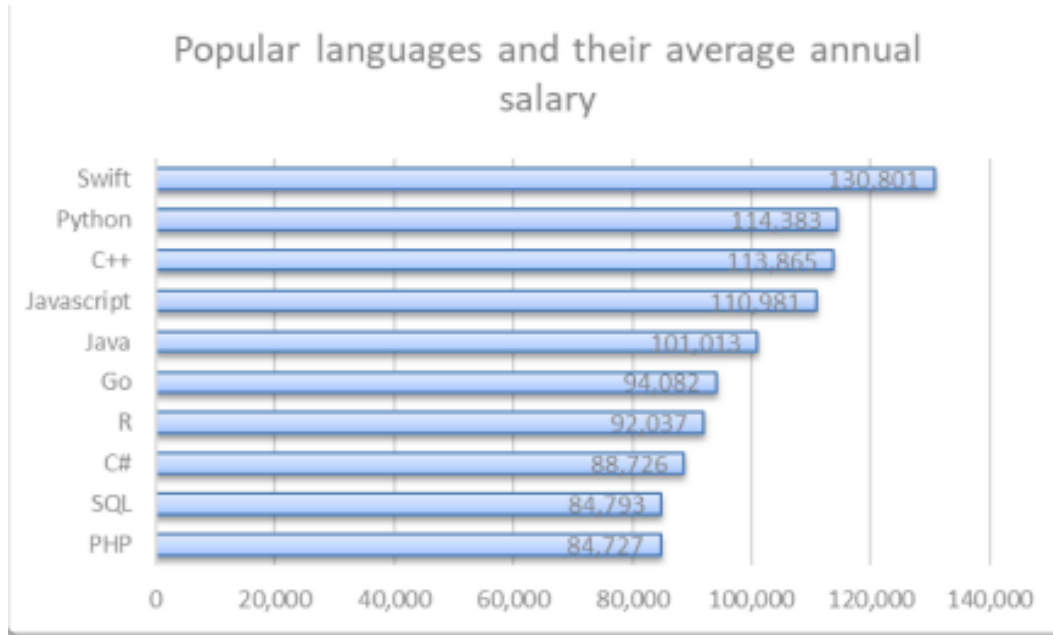
Respondent	0.004041
CompTotal	0.006970
ConvertedComp	0.105386
WorkWeekHrs	0.036518
CodeRevHrs	-0.020469
Age	1.000000
Name: Age, dtype: float64	

# JOB POSTINGS



This bar chart represent the number of job postings as per the programming languages.

# POPULAR LANGUAGES



This bar chart display popular languages and their annual salary. The data was collected through web scrapping the GitHub jobs data and saved in csv file.