



# IBM Data Analyst Capstone

Grattan Welch  
July 29, 2022



# OUTLINE

---



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

# EXECUTIVE SUMMARY

---



- Summary of methodologies
  - Data collection and web scraping
  - Exploratory Data Analysis (EDA) / Data wrangling
  - EDA using data visualization
  - EDA using SQL
  - Interactive dashboard with Cognos Dashboard Embedded
- Summary of all results
  - Exploratory Data Analysis findings
  - Interactive visual analytics dashboard
  - Javascript and HTML/CSS are number 1 and 2 respectively for current year and next
  - Database trends are unstable
  - Washington DC is a hotspot for technology job postings

# INTRODUCTION

---



- Get the number of job openings using the GitHub Jobs API for technologies and locations
- Extract information from the given web site
- Explore and wrangle data
- Identify technology usage and trends using SQL, EDA and data visualizations
- Analyze demographic information from survey respondents





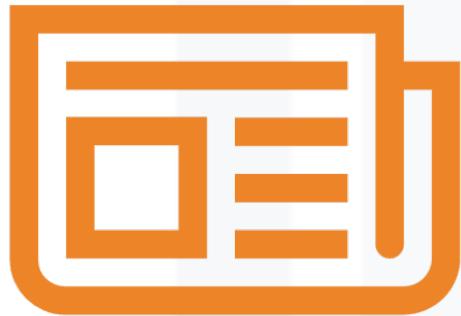
*Data Analyst Capstone*

# Methodology



# METHODOLOGY

## *Executive Summary*



- Data collection methodology:
  - Github Jobs REST API
  - Web scraping
- Perform data wrangling
  - Data cleaning and imputation of missing values
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Cognos Dashboard Embedded

# Data Collection

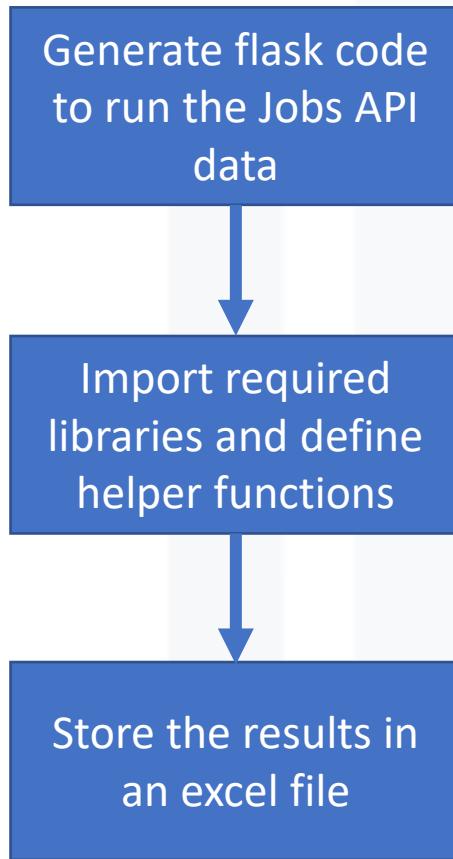
---



- Two datasets were collected
- Requested and parsed the Github JSON jobs data using the GET request
- Used BeautifulSoup to scrape the name of the programming language and average annual salary from web site
- Create Excel and CSV files from extracted data
- Explore and wrangle data
- Deal with duplicate and missing values
- Normalize data in a dataset



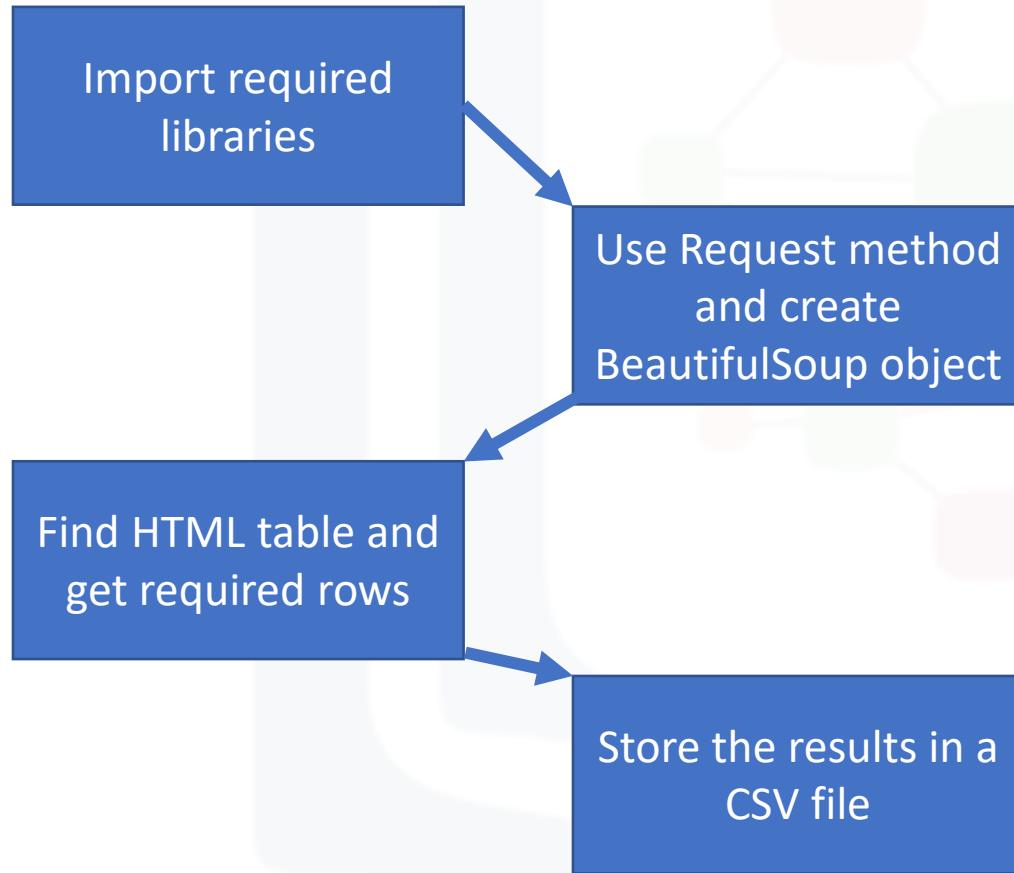
# Data Collection Using Github Jobs API



Data collection with REST API calls

[https://github.com/onestopbb/IBMDATAAnalystCapstone/blob/master/Collecting\\_Jobs%20data%20Using%20API-Questions.ipynb](https://github.com/onestopbb/IBMDATAAnalystCapstone/blob/master/Collecting_Jobs%20data%20Using%20API-Questions.ipynb)

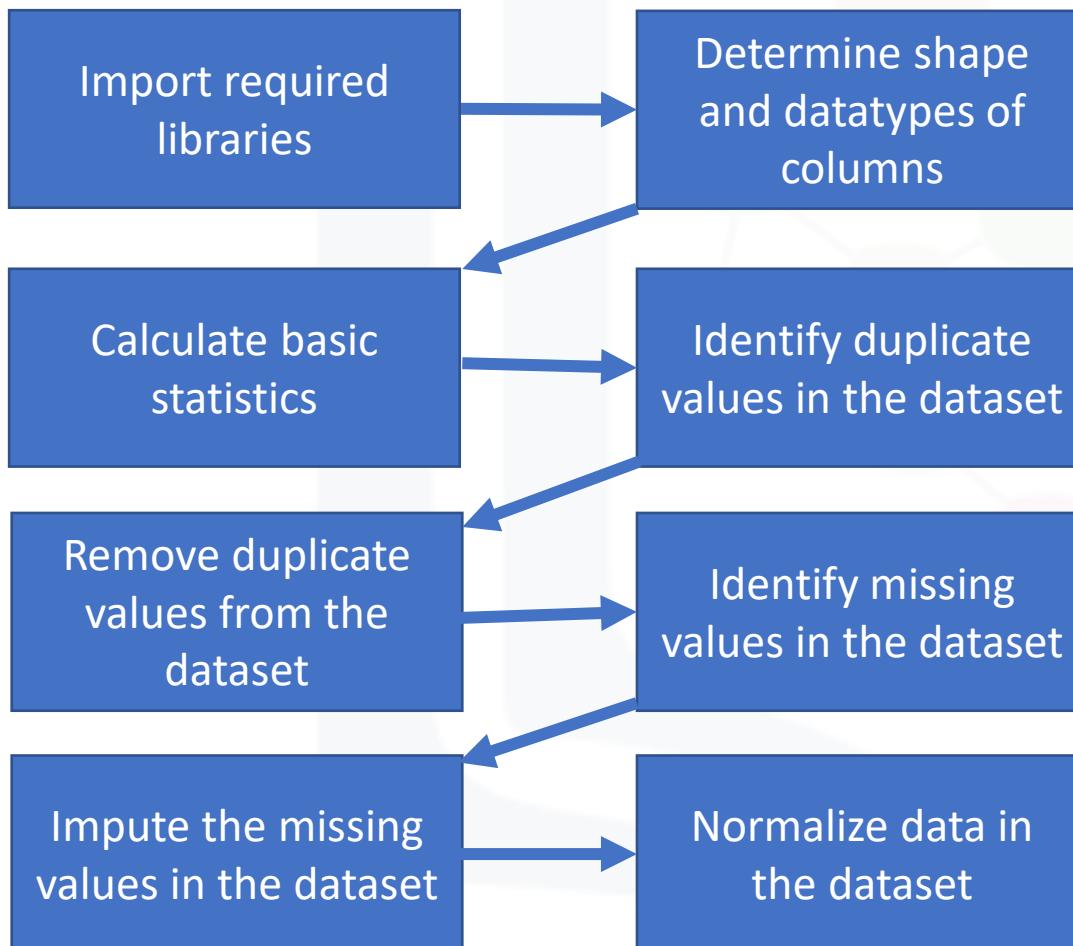
# Data Collection Using Web Scraping



Web scraping process

[https://github.com/onestopbb/IBMDATAAnalystCapstone  
blob/master/Web-Scraping-Lab.ipynb](https://github.com/onestopbb/IBMDATAAnalystCapstone/blob/master/Web-Scraping-Lab.ipynb)

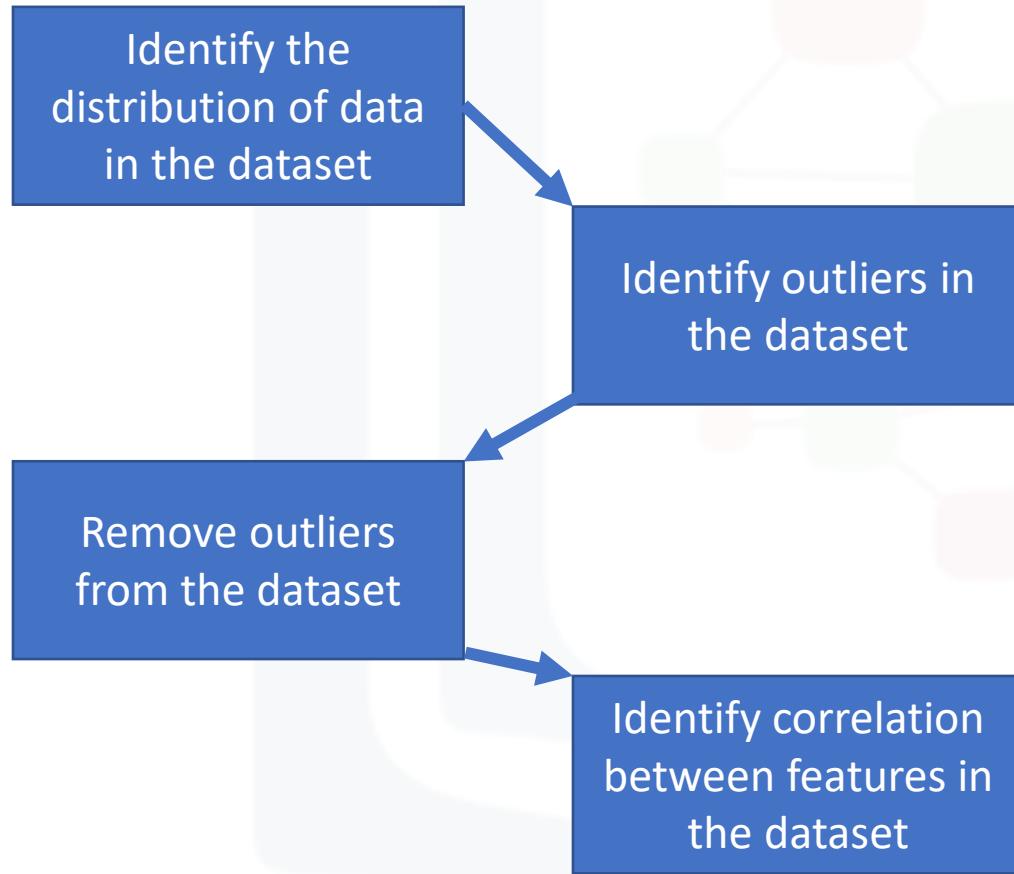
# Data Exploration and Wrangling



## Perform Exploratory Data Analysis and Wrangling

- <https://github.com/onestopbb/IBMDataAnalystCapstone/blob/master/M1ExploreDataSet-lab.ipynb>
- <https://github.com/onestopbb/IBMDataAnalystCapstone/blob/master/M2DataWrangling-lab.ipynb>

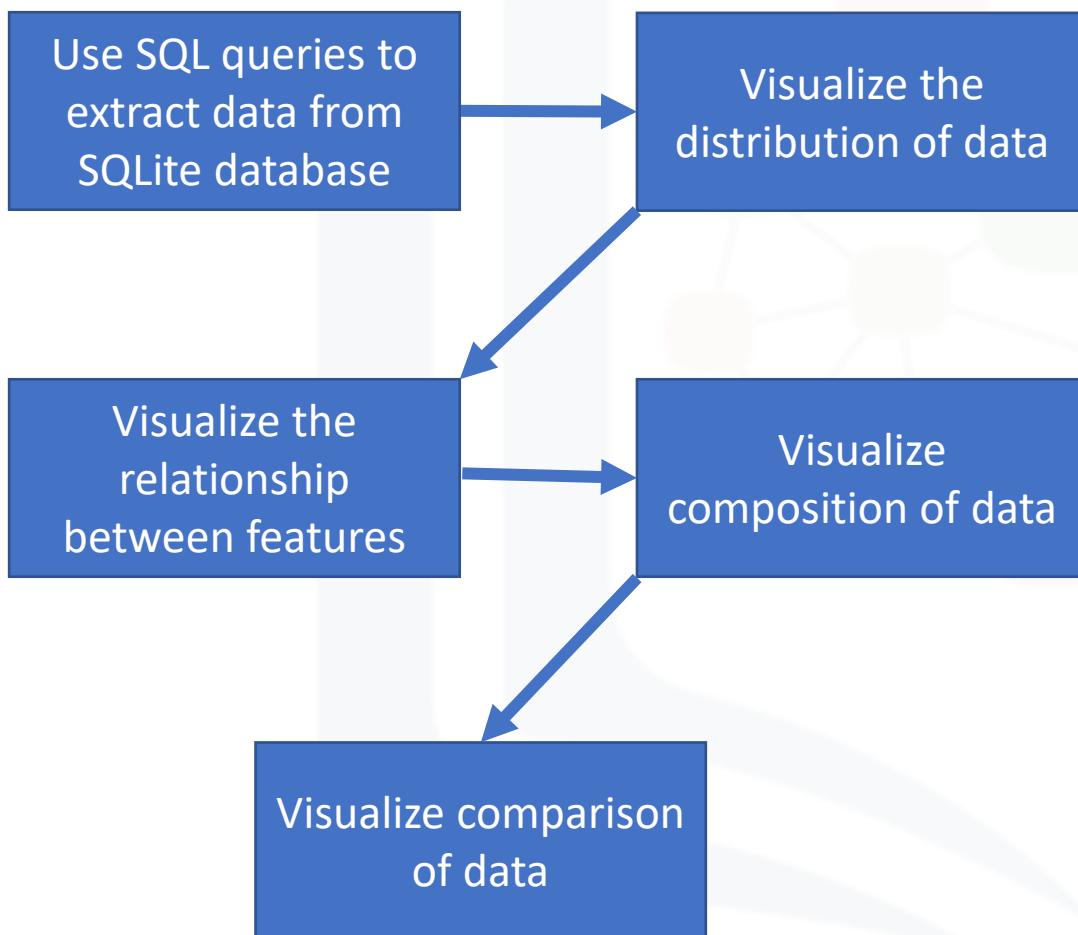
# Exploratory Data Analysis



Perform Exploratory Data Analysis (EDA)

<https://github.com/onestopbb/IBMDATAAnalystCapstone/blob/master/M3ExploratoryDataAnalysis-lab.ipynb>

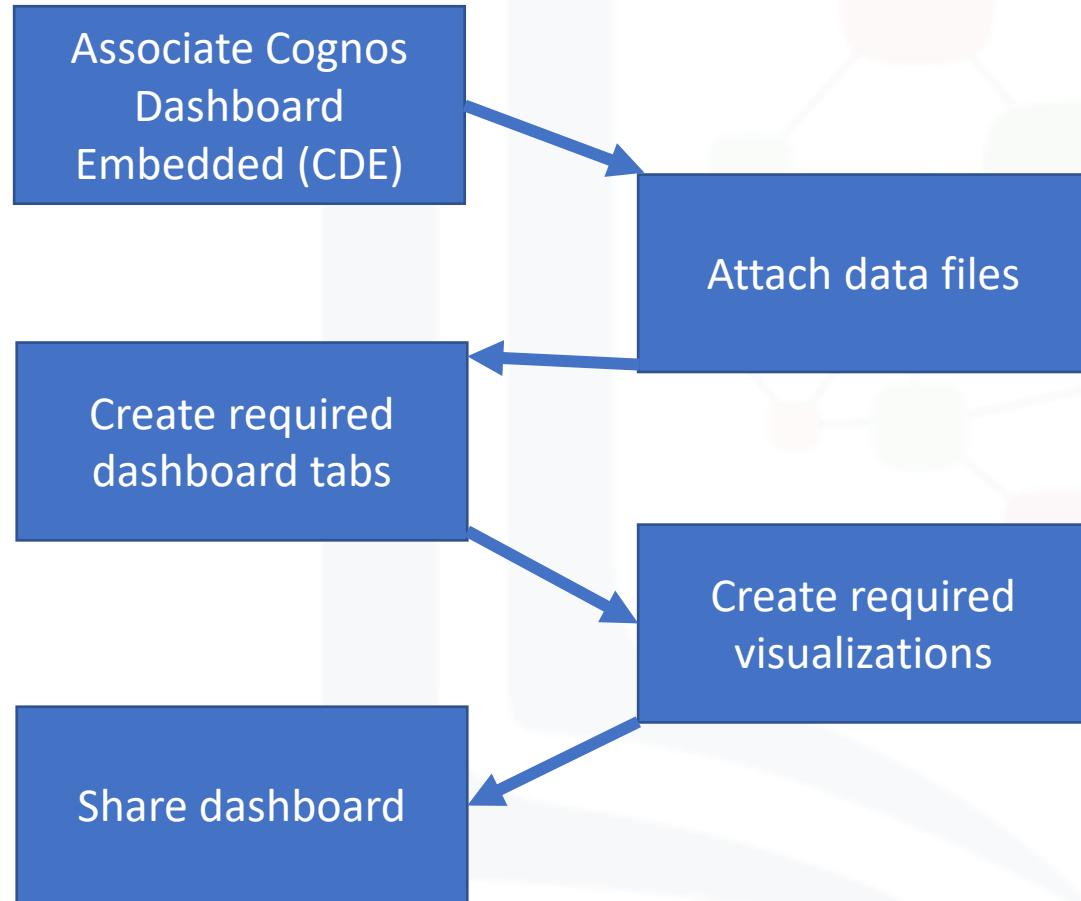
# Data Visualization



## Data Visualization

[https://github.com/onestopbb/IBMDataAnalystCapstone/  
blob/master/M4DataVisualization-lab\\_LOCAL.ipynb](https://github.com/onestopbb/IBMDataAnalystCapstone/blob/master/M4DataVisualization-lab_LOCAL.ipynb)

# Building a Dashboard with Cognos Dashboard Embedded (CDE)



## Build Dashboard Using CDE

<https://eu-de.dataplatform.cloud.ibm.com/dashboards/000f76bd-c291-4c3e-ae08-7087b6232ccf/view/413cfa253bae6cc011bcd4e407cf28017c65705ab2bbd506d4d67b490b357997a96945c2c82d4a0ede400465f2be100c9c>

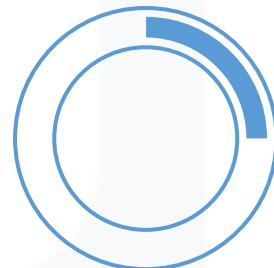
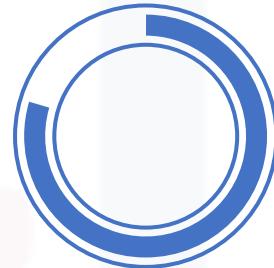
*Data Analyst Capstone*



# RESULTS

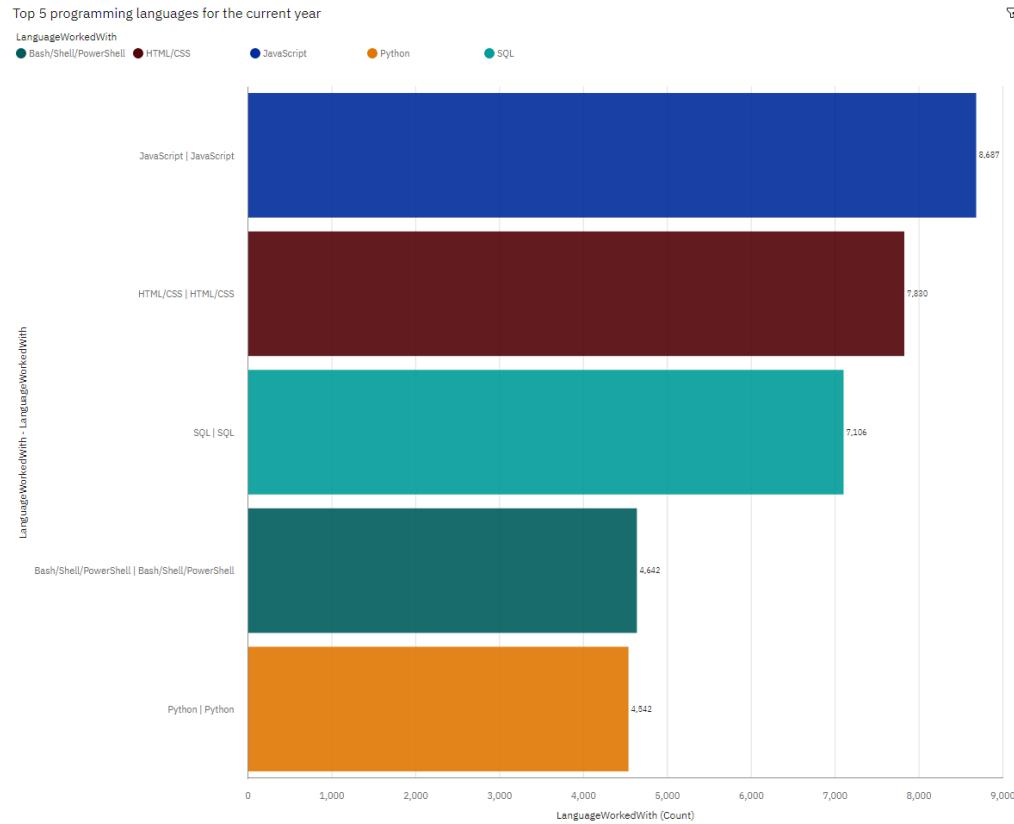
---

- Programming language trends identified
- Database trends identified
- Interactive analytics via dashboard presented

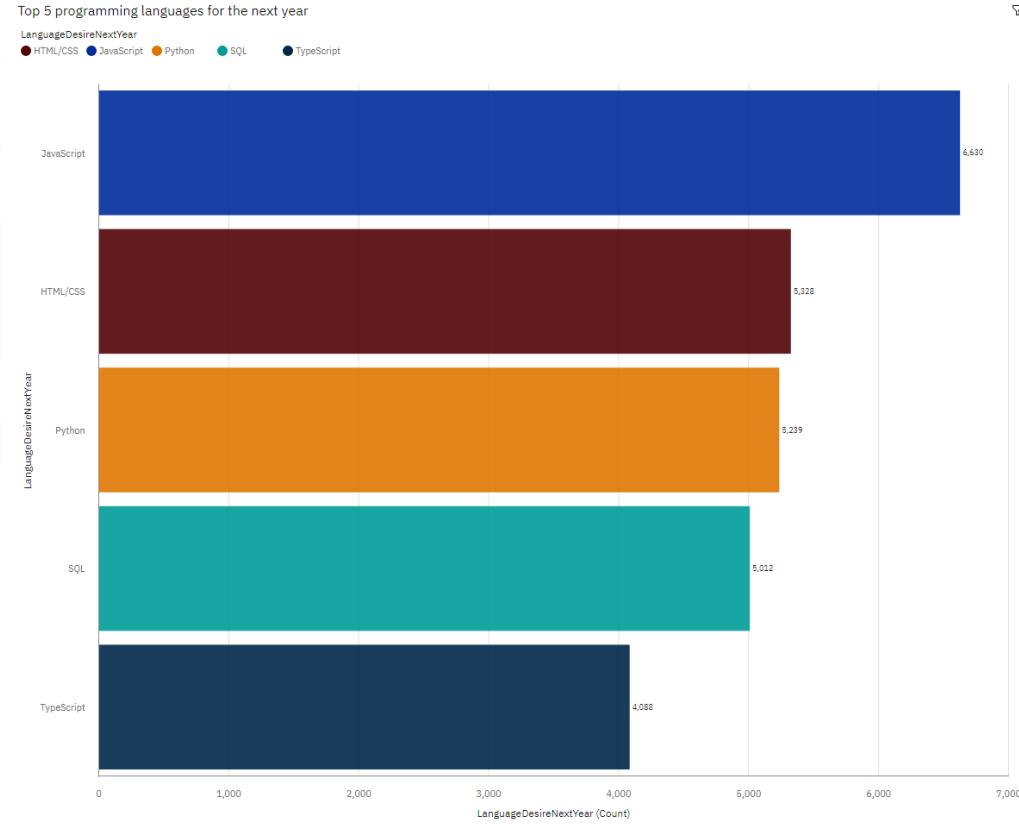


# PROGRAMMING LANGUAGE TRENDS

Bar chart of top 5 programming languages for the current year



Bar chart of top 5 programming languages for the next year



# PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

---

## Findings

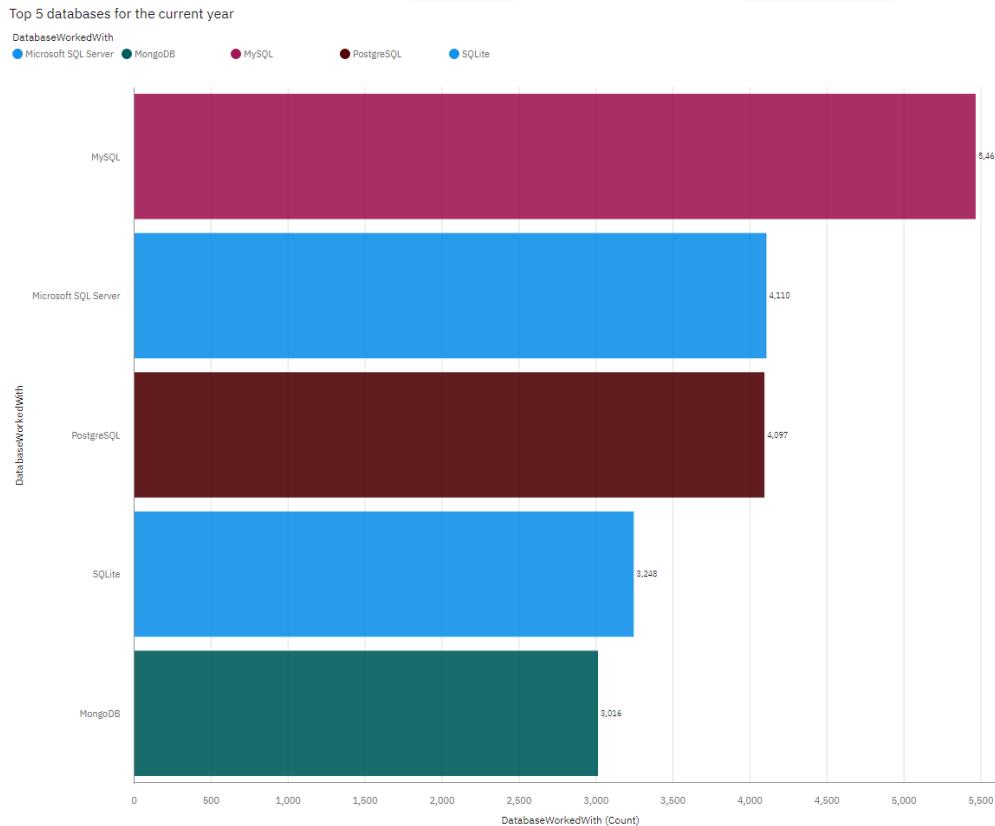
- Javascript and HTML/CSS are number 1 and 2 respectively for current year and next
- Python and SQL remain on top 5 list for both periods
- TypeScript is pushing out Bash/Shell/PowerShell

## Implications

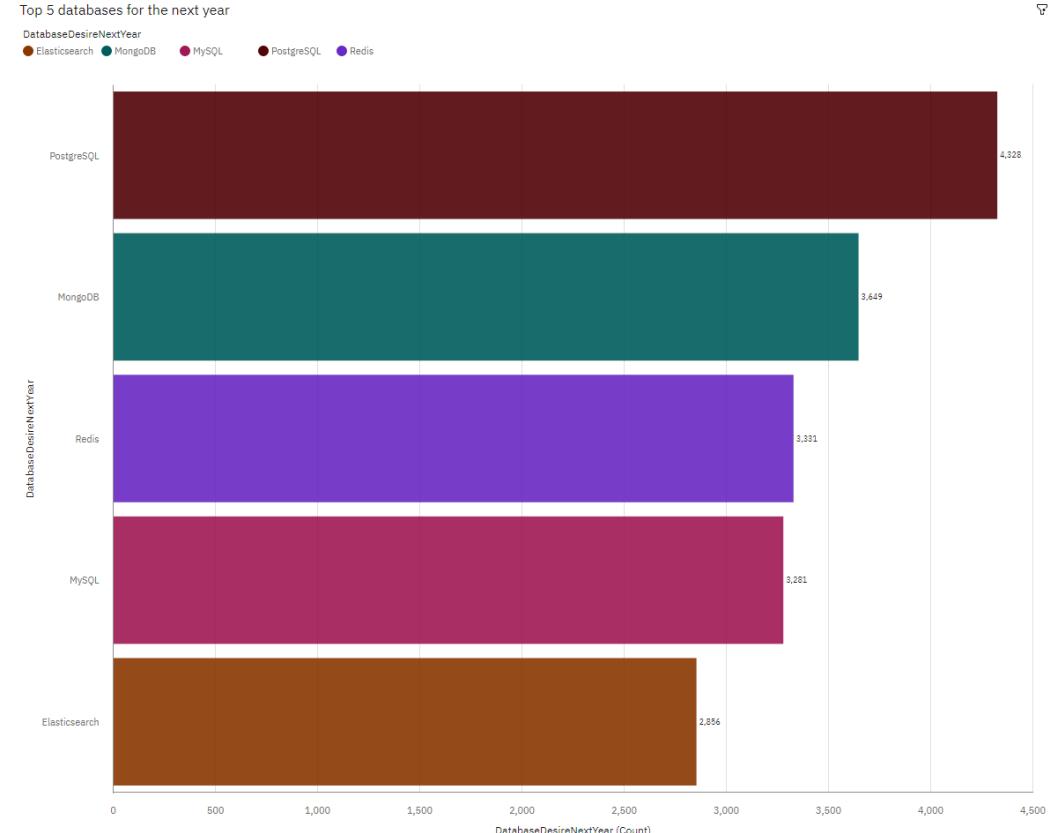
- Most of the current programming remaining popular the next year implies a continued high demand for them
- Becoming in newer languages like TypeScript could potentially be lucrative

# DATABASE TRENDS

Bar chart of top 5 databases for the **current year**



Bar chart of top 5 databases for the **next year**



# DATABASE TRENDS - FINDINGS & IMPLICATIONS

---

## Findings

- MySQL, PostgreSQL, and MongoDB remain on future list, but with fluctuations in rank
- Microsoft SQL Server and SQLite replaced by Redis and Elasticsearch

## Implications

- Learning the traditional open-sourced RDBMS MySQL, PostgreSQL, and MongoDB is still recommended
- Proprietary systems are losing popularity
- Cloud-based RDMS like Redis should continue to gain rank with a push to the cloud

# DASHBOARD

---



<https://eu-de.dataplatform.cloud.ibm.com/dashboards/000f76bd-c291-4c3e-ae08-7087b6232ccf/view/413cfa253bae6cc011bcd4e407cf28017c65705ab2bbd506d4d67b490b357997a96945c2c82d4a0ede400465f2be100c9c>



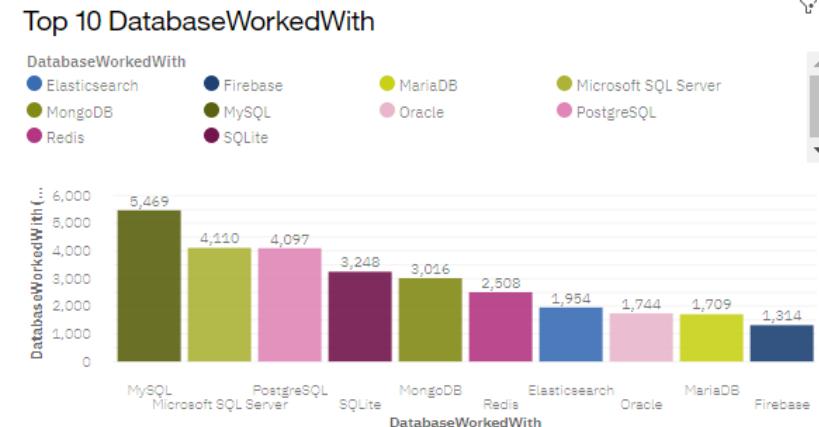
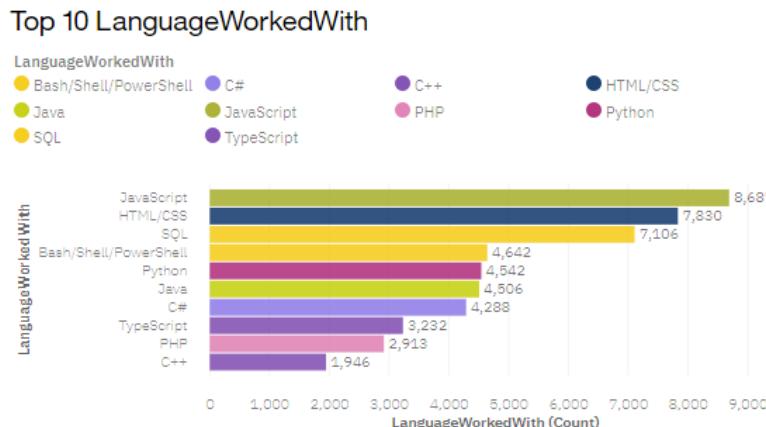
# DASHBOARD TAB 1

# *Current Technology Usage*

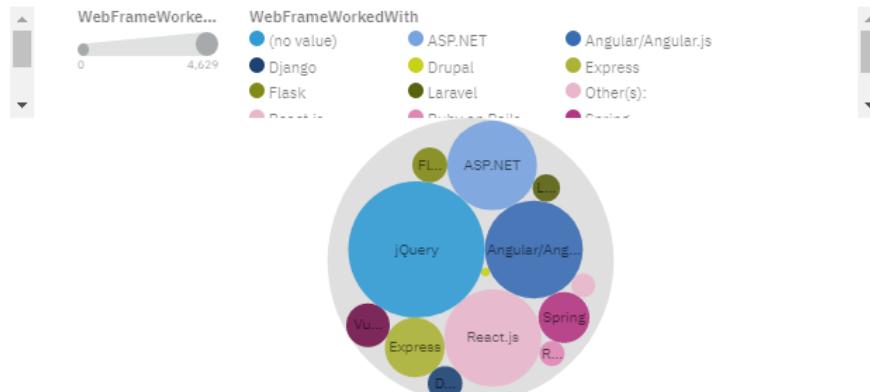
## Current Technology Usage

## Future Technology Trend

## Demographics

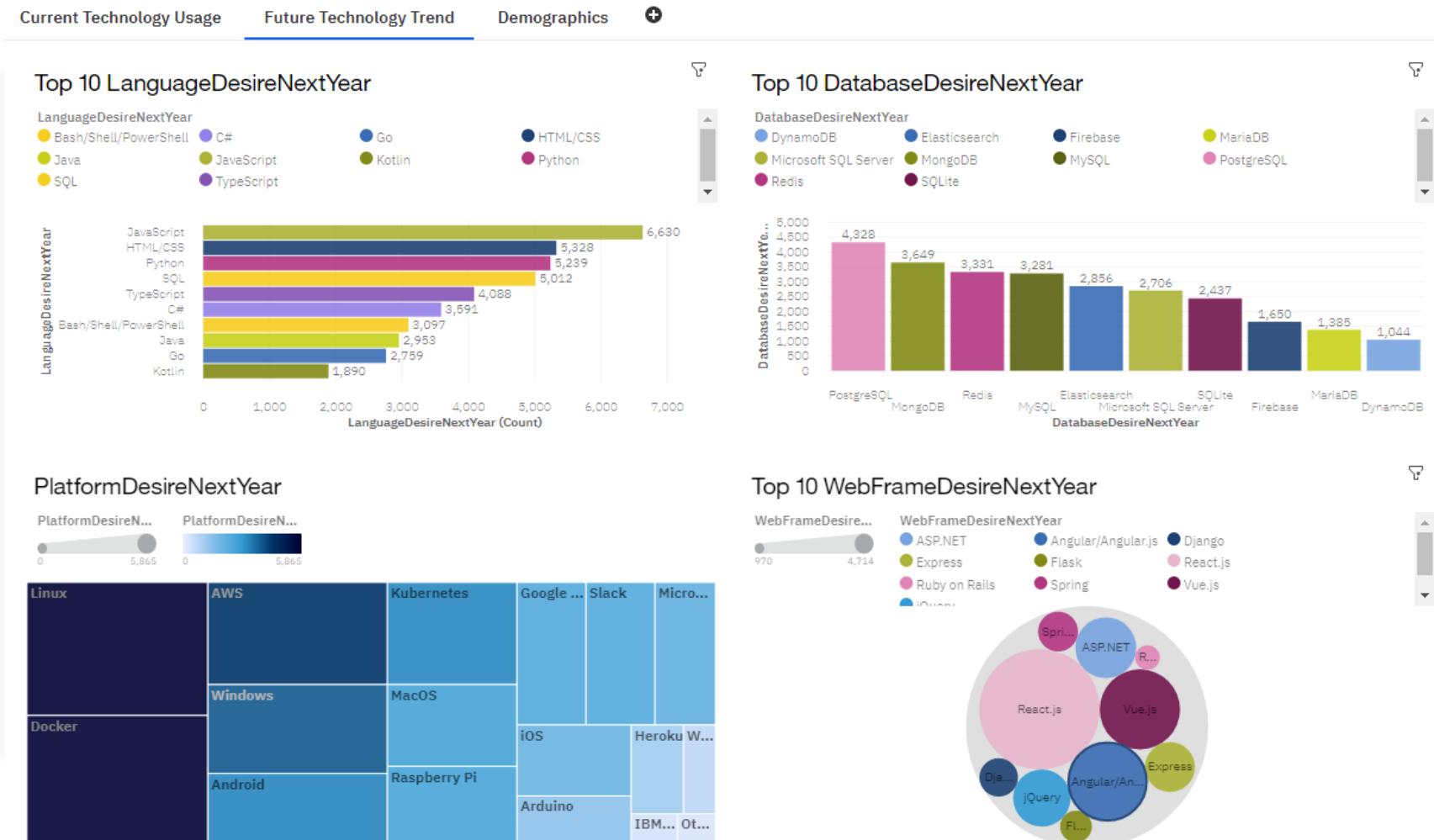


## PlatformWorkedWith



# DASHBOARD TAB 2

## Future Technology Trend



# DASHBOARD TAB 3

## Demographics

Current Technology Usage

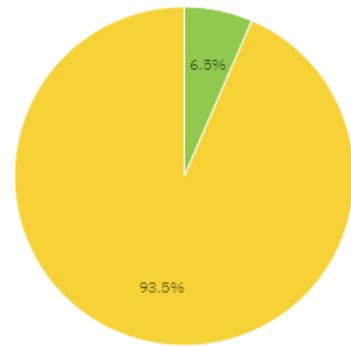
Future Technology Trend

Demographics

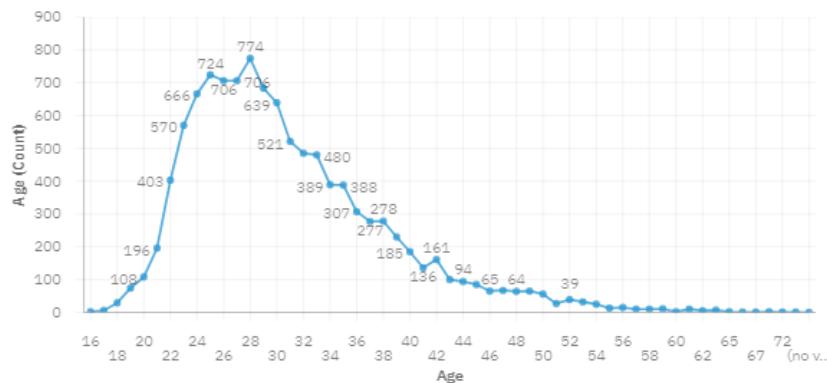


Respondent classified by Gender

Gender  
Woman Man

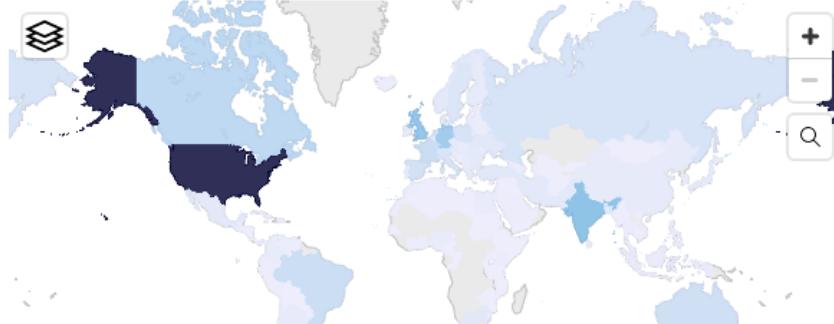


Respondent Count by Age



Respondent Count for Countries

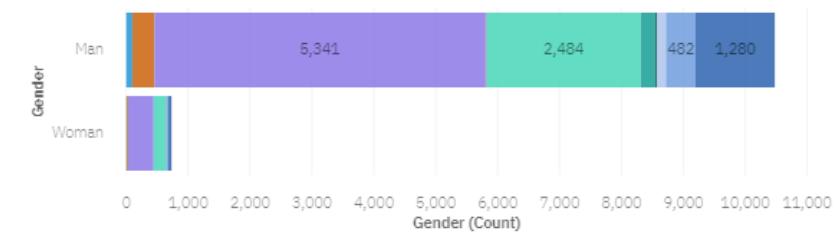
Respondent (Cou...  
1 3,058

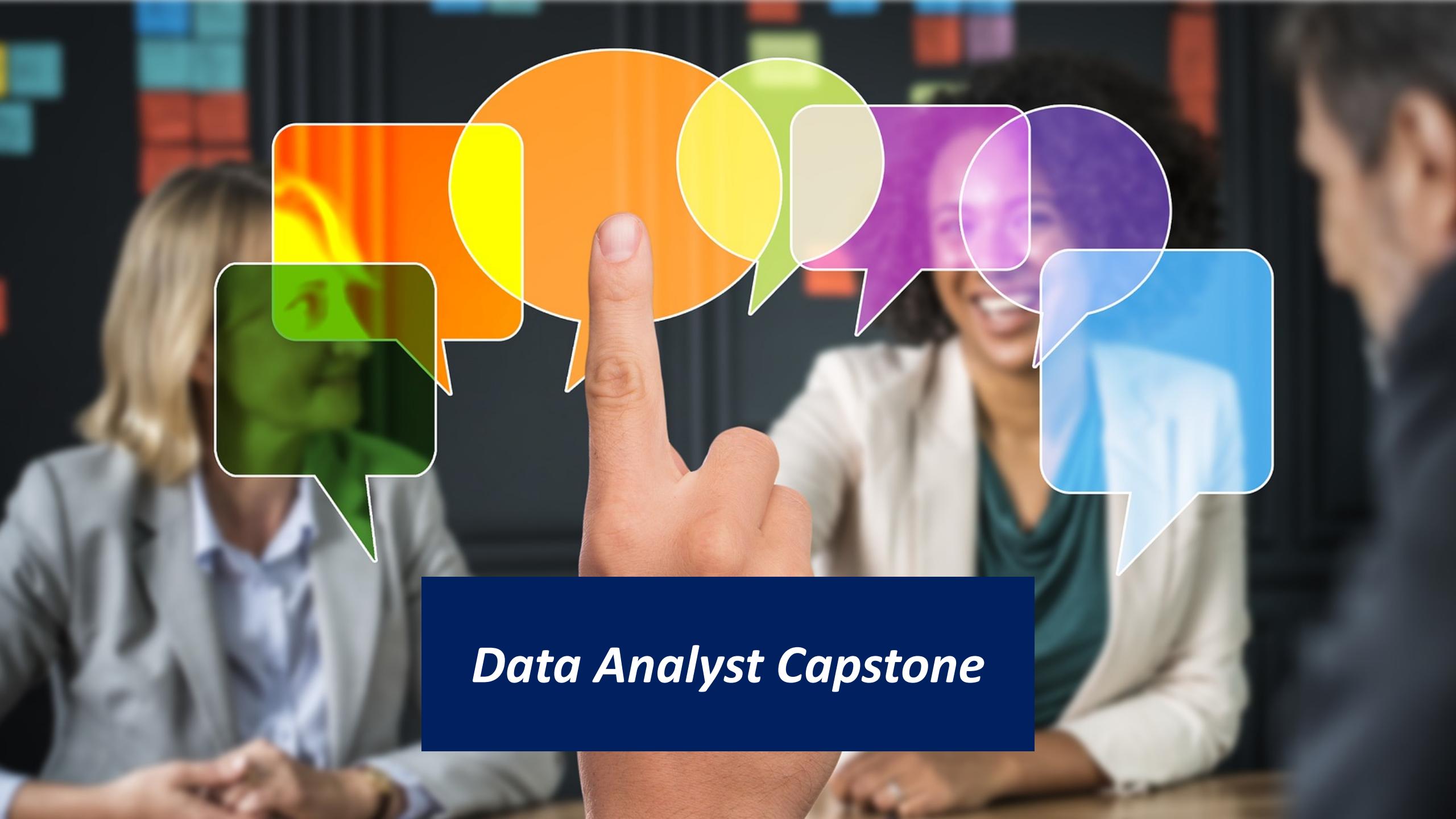


Respondent Count by Gender, classified by Formal Education

Level

EdLevel  
● (no value) ● Associate degree  
● I never completed any formal ... ● Bachelor's degree (BA, BS, B.E...  
● Primary/elementary school ● Master's degree (MA, MS, M.E...  
● Professional degree (JD, MD, E... ● Other doctoral degree (Ph.D, E...  
● Secondary school (e.g. Americ...





*Data Analyst Capstone*

# DISCUSSION

---



# OVERALL FINDINGS & IMPLICATIONS

---

## Findings

- Programming language trends are relatively stable
- Database trends are somewhat unstable
- Linux, Docker, AWS, and Windows are among the top 5 current and future platforms
- The vast majority of respondents (93.5%) were male
- Most respondents have a Bachelor's degree, and were in the 20 to 35 age group

## Implications

- There should be a push to encourage more females to enter technology field
- There may be a push to cloud-based RDBMS
- There may also be a push to multipurpose open-sourced tools such as Elasticsearch
- The platform usage trend appears to be stable
- Post-graduate degrees in tech are probably not needed by most persons

# *Data Analyst Capstone*

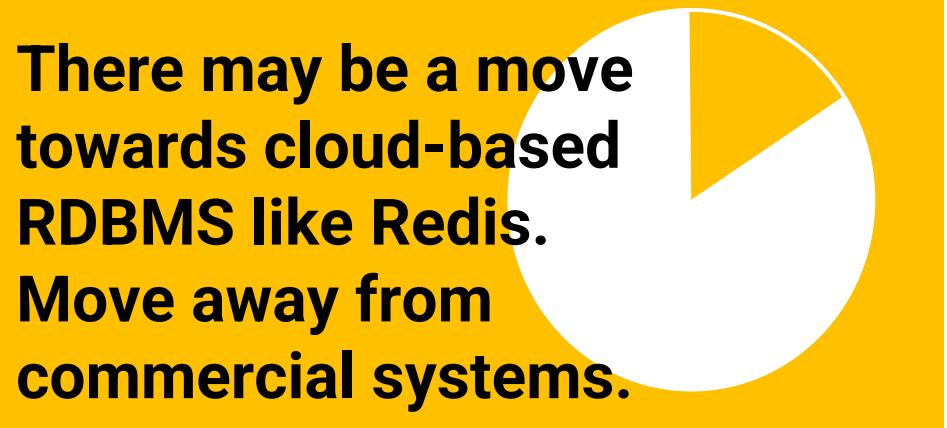


# CONCLUSION

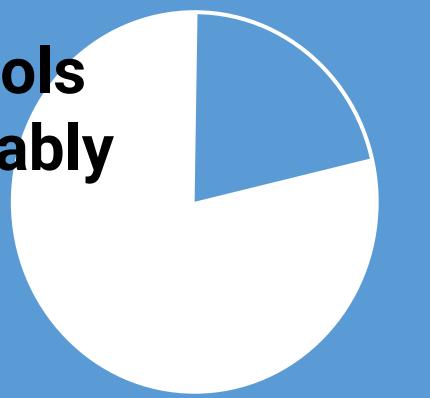
Respondents were over 90% males. More women should be encouraged.



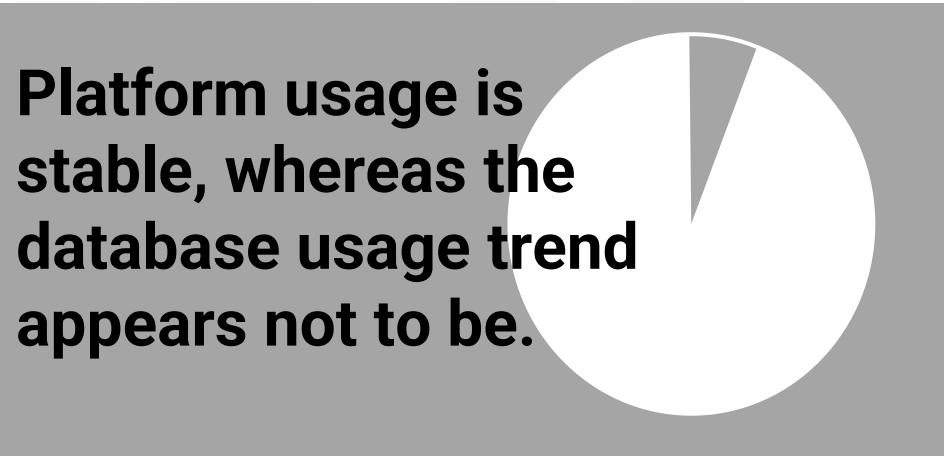
There may be a move towards cloud-based RDBMS like Redis. Move away from commercial systems.



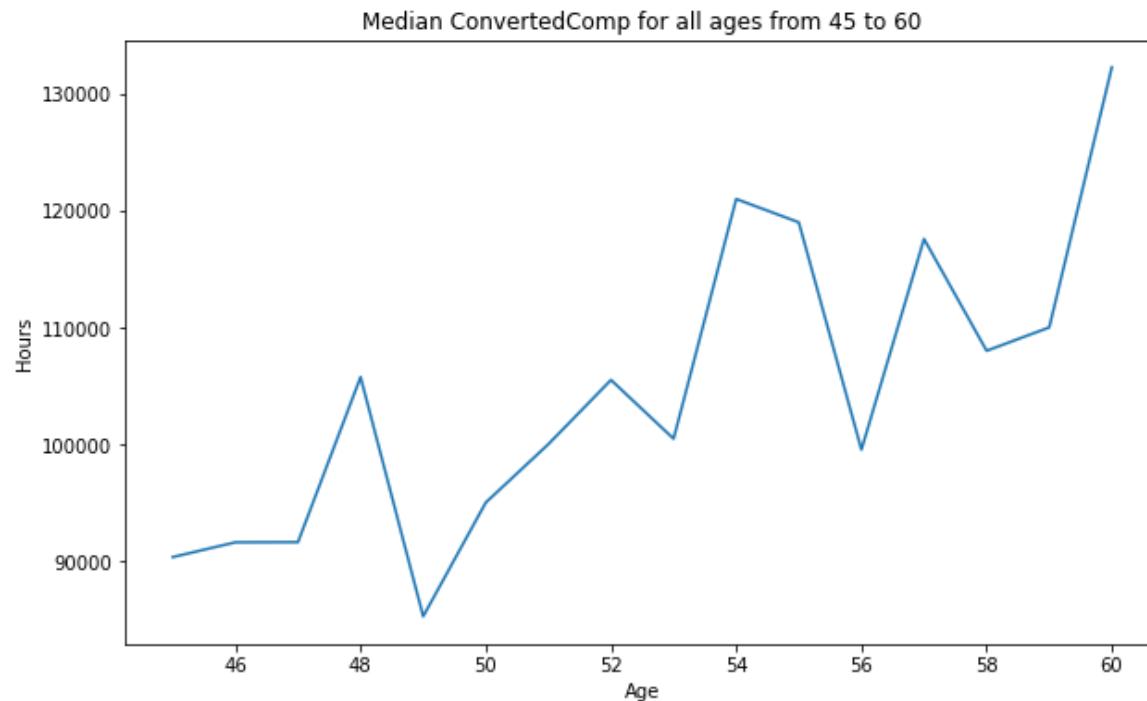
Open-sourced tools are and will probably become more commonly used.



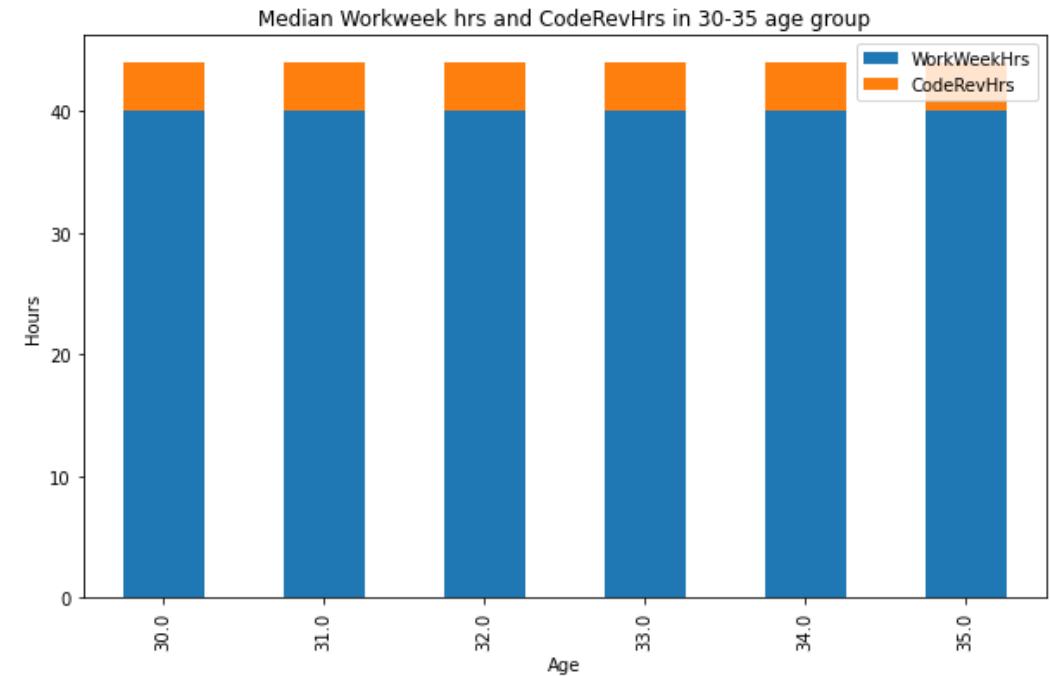
Platform usage is stable, whereas the database usage trend appears not to be.



# APPENDIX – Charts created during analysis



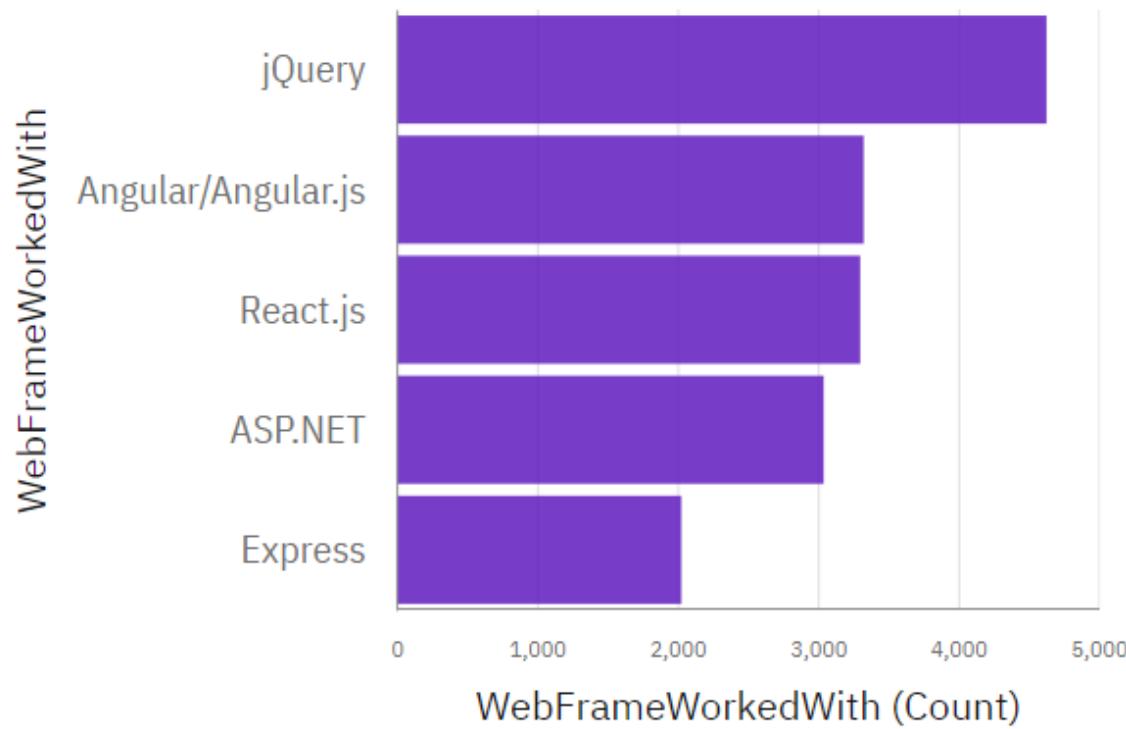
Median ConvertedComp for all ages from 45 to 60



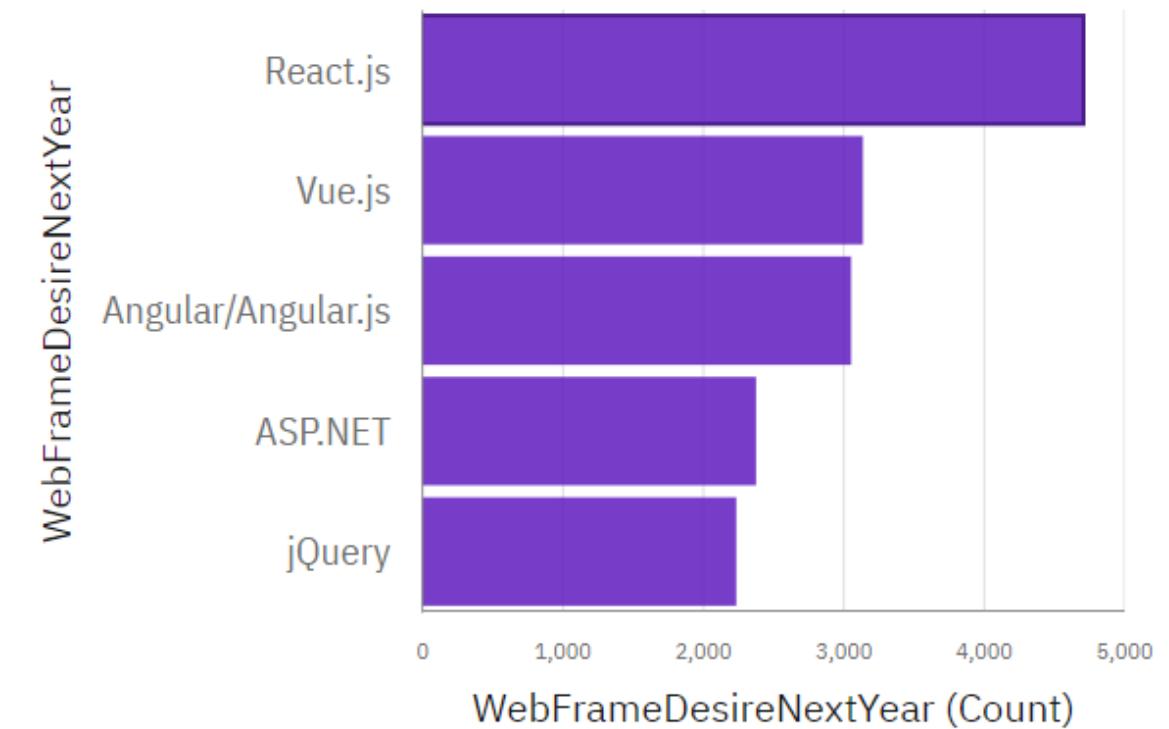
Median Workweek hrs and CodeRevHrs in 30-35 age group

# APPENDIX – Web Frame Worked Trends

WebFrameWorkedWith

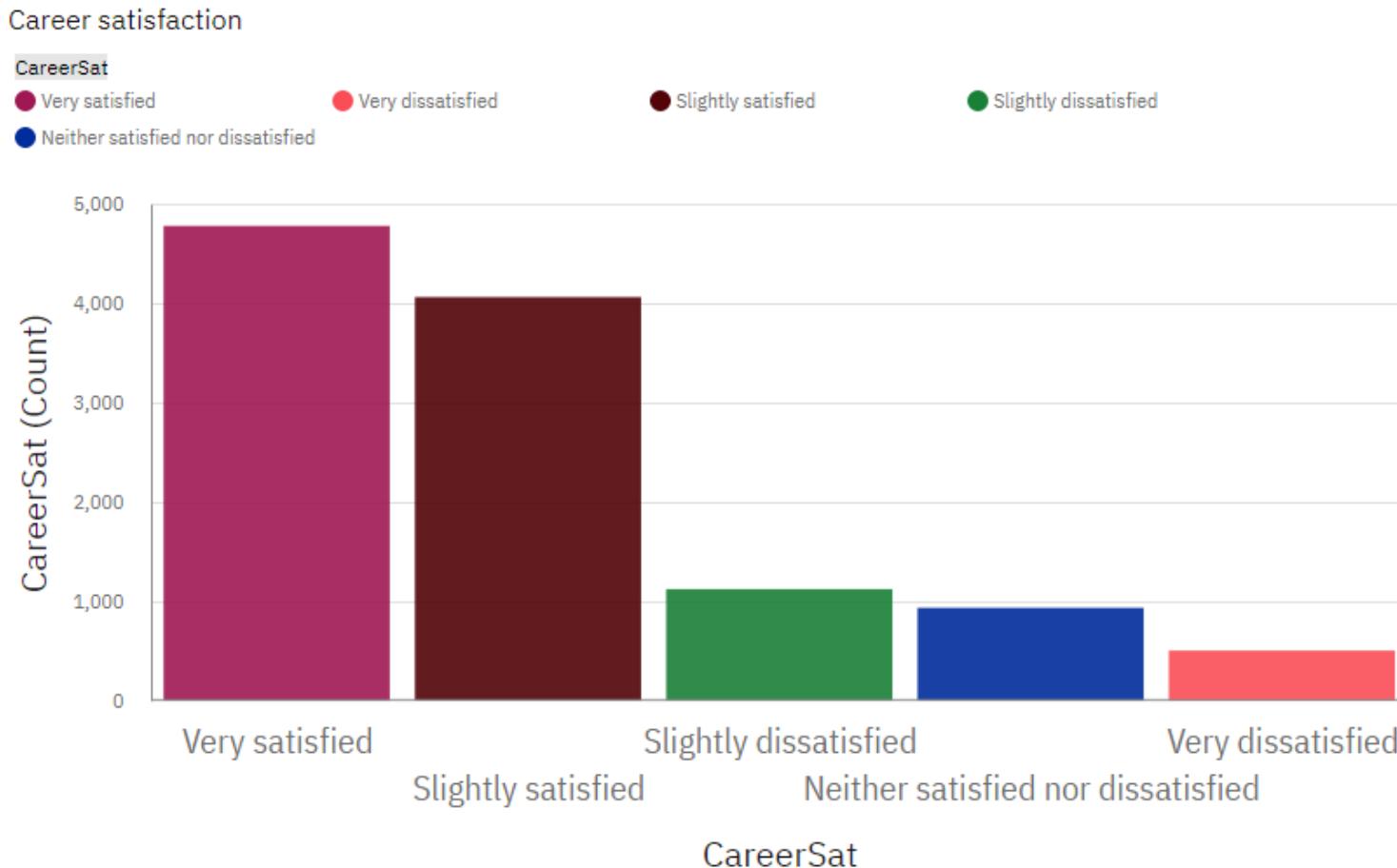


WebFrameDesireNextYear



The web frame worked with trend is very unstable

# APPENDIX – Career Satisfaction

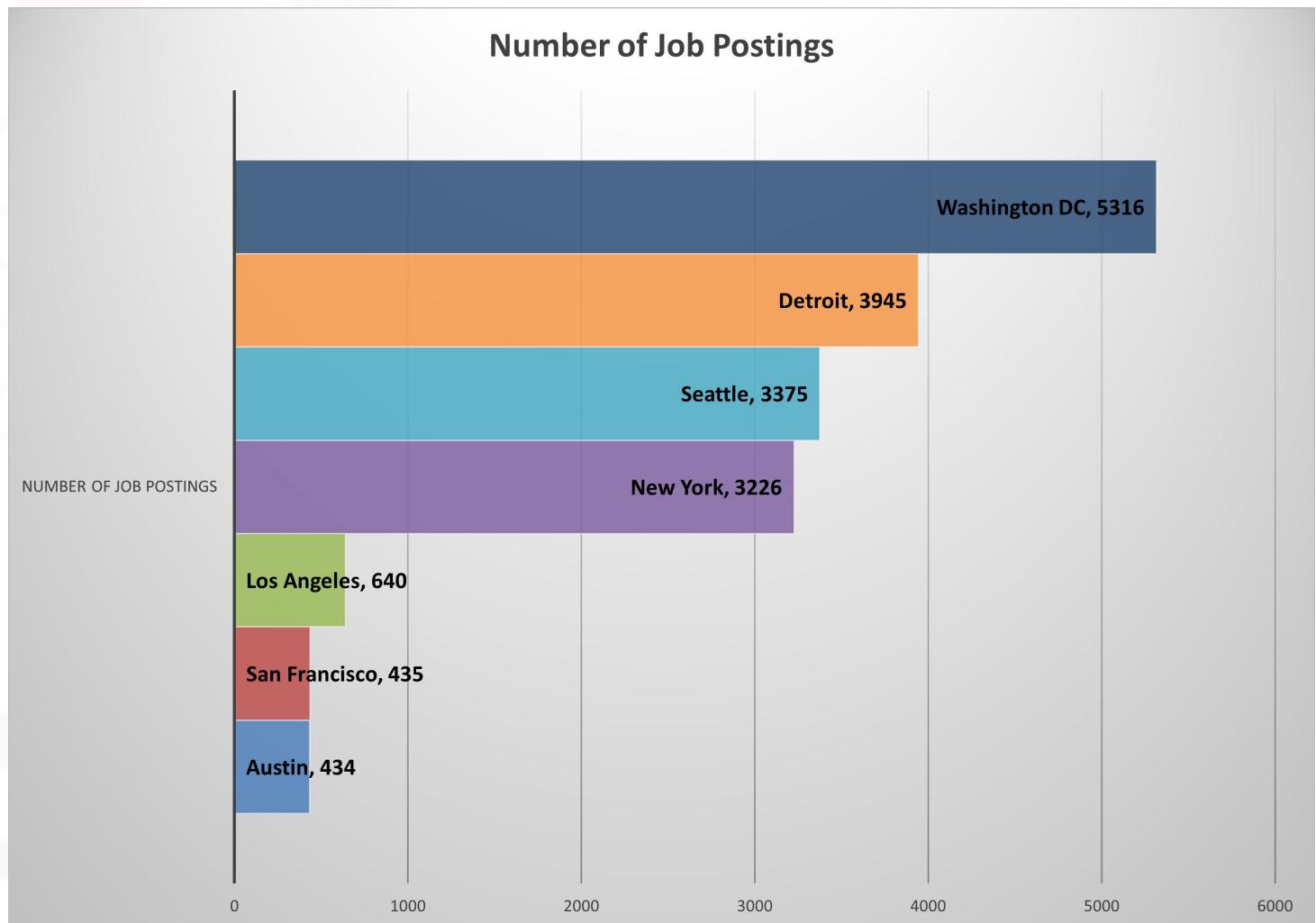


Most respondents  
are very satisfied  
with their jobs

# JOB POSTINGS

Washington DC is a hotspot for job postings

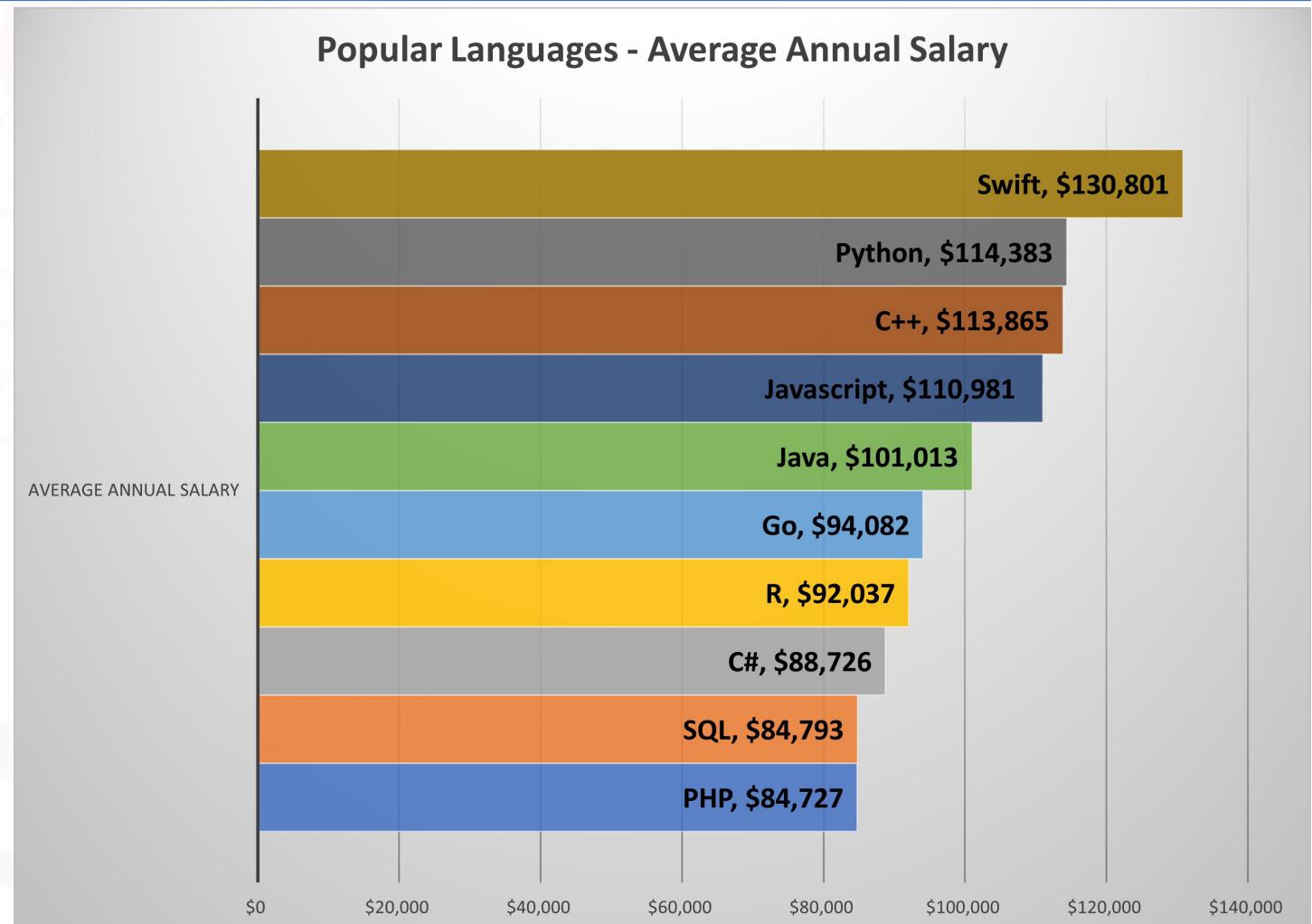
<https://github.com/onestopbb/IBMDATAAnalystCapstone/blob/master/Collecting%20Jobs%20data%20Using%20API-Questions.ipynb>



# POPULAR LANGUAGES

Swift, Python, and C++ are the three most popular programming languages based on the average annual salary.

<https://github.com/onestopbb/IBMDataAnalystCapstone/blob/master/Web-Scraping-Lab.ipynb>



Thank you!

