Mining the Trend of the Most Frequently Used Programming Languages and Tools of the Last Decade

CSPB 4502
Data Mining Project

Group 3

Ning Chih Chang

Description

- This is preliminary trend analysis of the most frequently used programming languages and development tools of the last decade (from 2015 to 2024).
- Example questions that could be answered by this project:
  - What are the most frequently used language of the last decade? Could we find the trend of languages that has changed the popularity of the past decade?
  - What are the most used programming languages for Full-stack Web Developers in the last decade? Would the result be the same for other occupations such as Mobile Developers or Data Scientists?
  - Are certain languages and tools associated?

Datasets

- **Stackoverflow Developer Survey 2015 – 2024** (10 sets of surveys in total)
- Downloaded from https://survey.stackoverflow.co/
- After dropping Nan values in programming languages and development tools related attributes, there are **634,598** valid data for analysis.
- The final eight attributes of this study are :

Occupation (nominal), Country (nominal), Education (nominal), Years _Coding (ordinal), OS (nominal), Dev_Env (nominal), Year (numeric-interval) and Tools (nominal).

Data preparation work

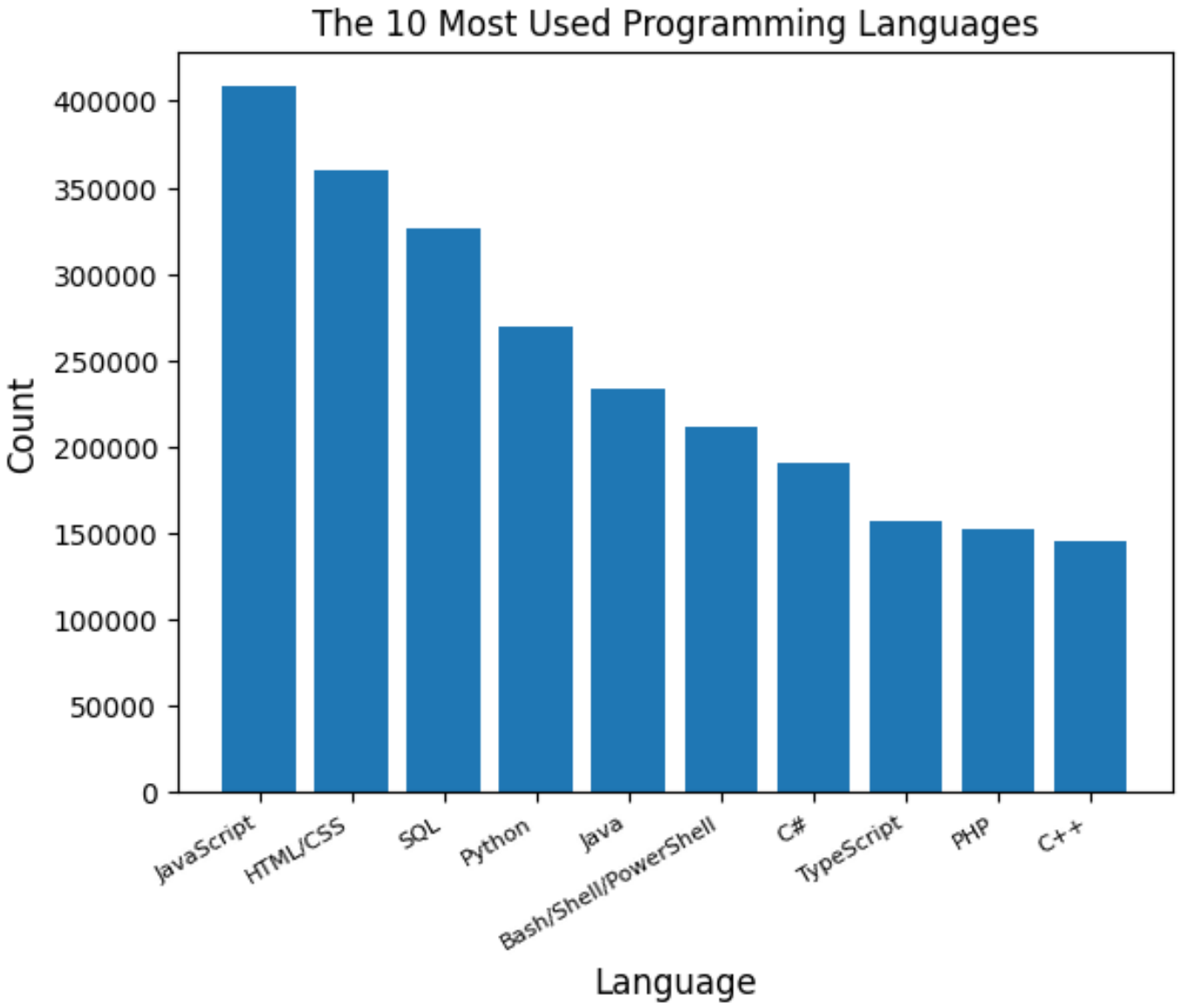
- Data cleaning: drop Nan values in programming languages and development tools related attributes.
- Data transformation: performed data discretization to create a consistent attribute type for the integrated dataset
- Data integration: match the shared attributes and remove irrelevant attributes among 10 different surveys
- Tool used: Python, Pandas, NumPy and Matplotlib

Analysis

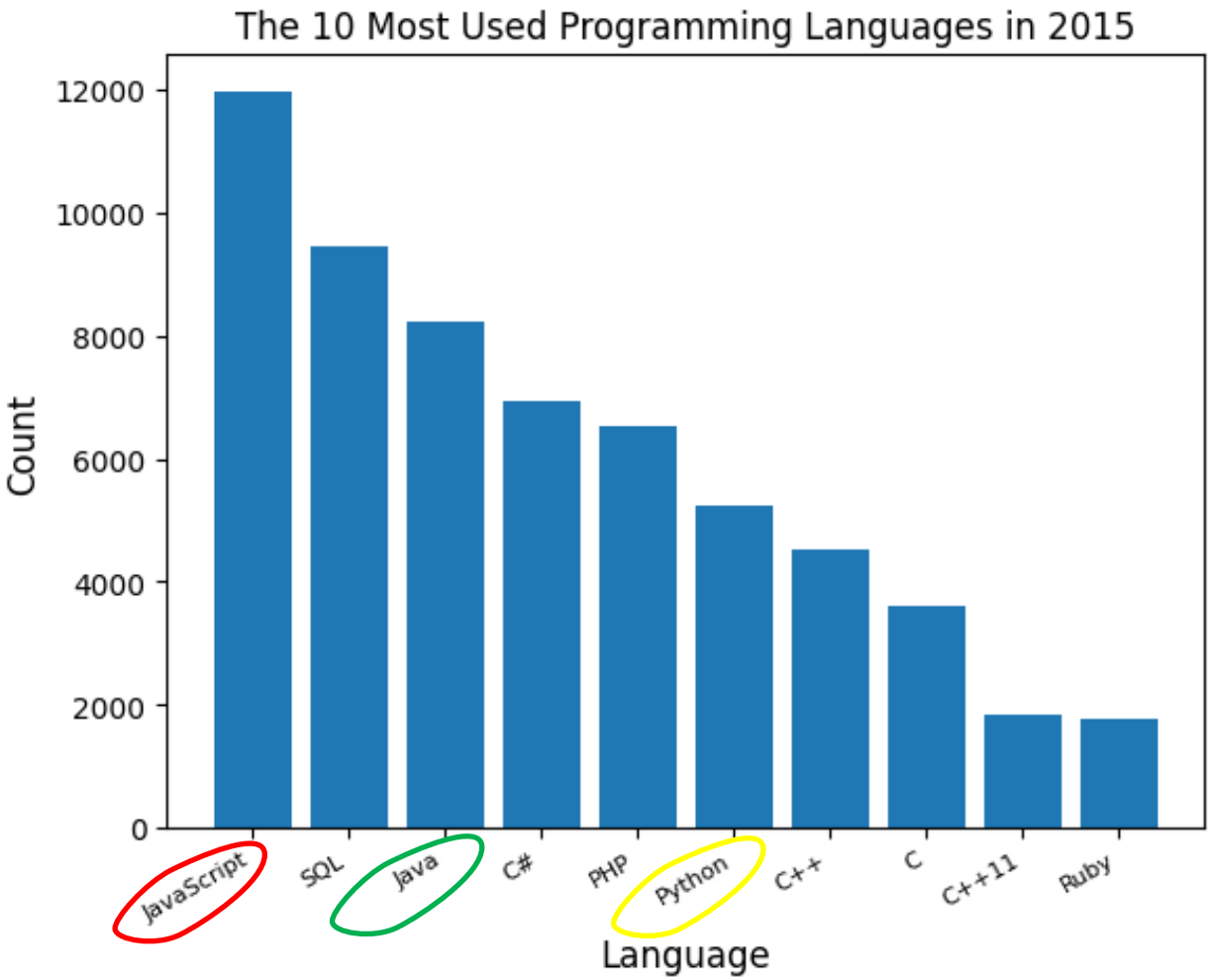
- Applied **frequency count** to the language related attributes to derive the most used languages by years of coding, occupation, country and year.
- Applied **Apriori algorithm with 20% support level and 60% confidence level** to find the frequent patterns of languages and tool

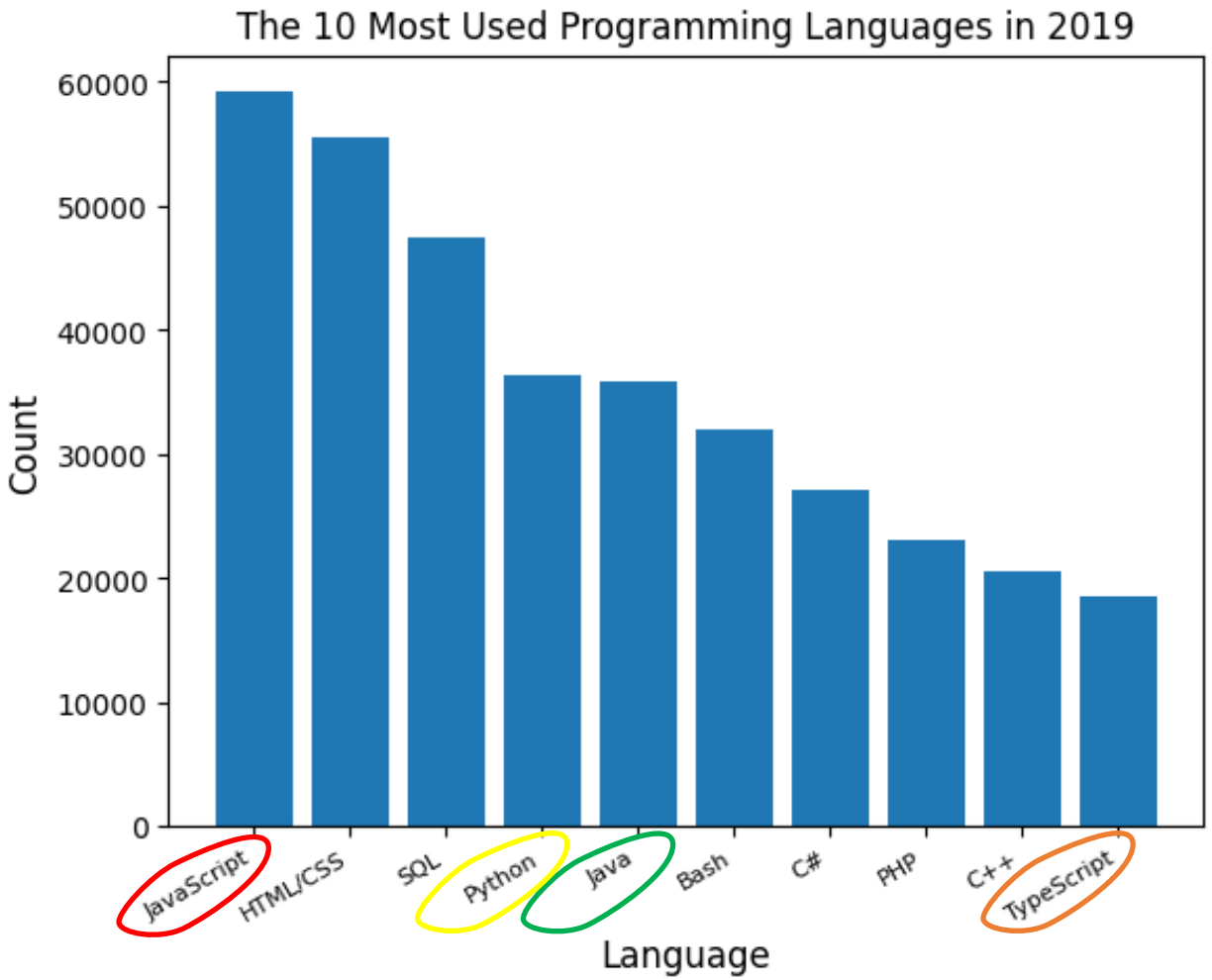
Result

- JavaScript is the first most used programming language in the past 10 years.
- Then followed by HTML/CSS and SQL
- Mostly web application related languages

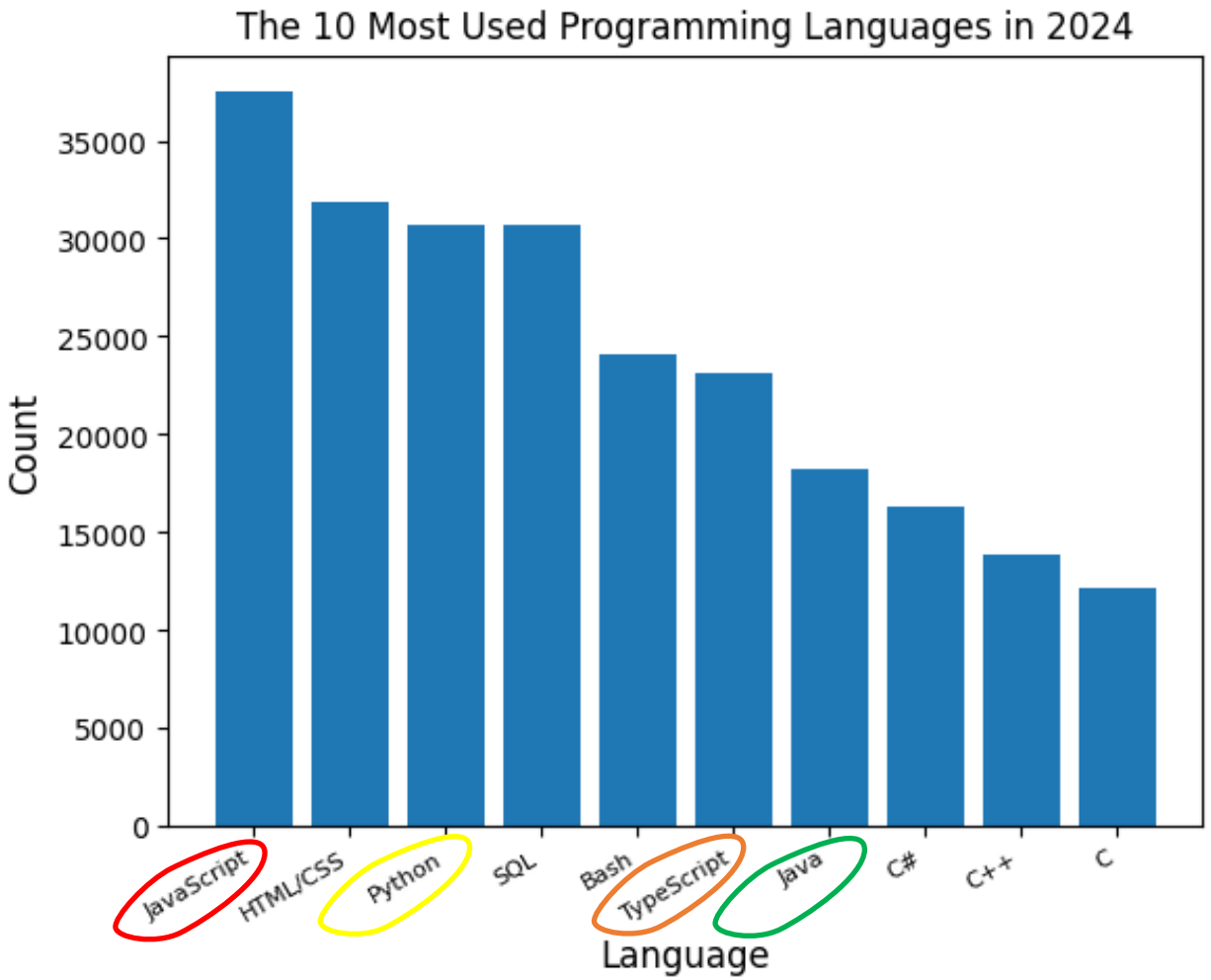


Result


The 10 Most Used Programming Languages in 2015
Count
12000
10000
8000
6000
4000
2000
0
JavaScript
SQL
Java
C#
PHP
Python
C++
C
C++11
Ruby
Language

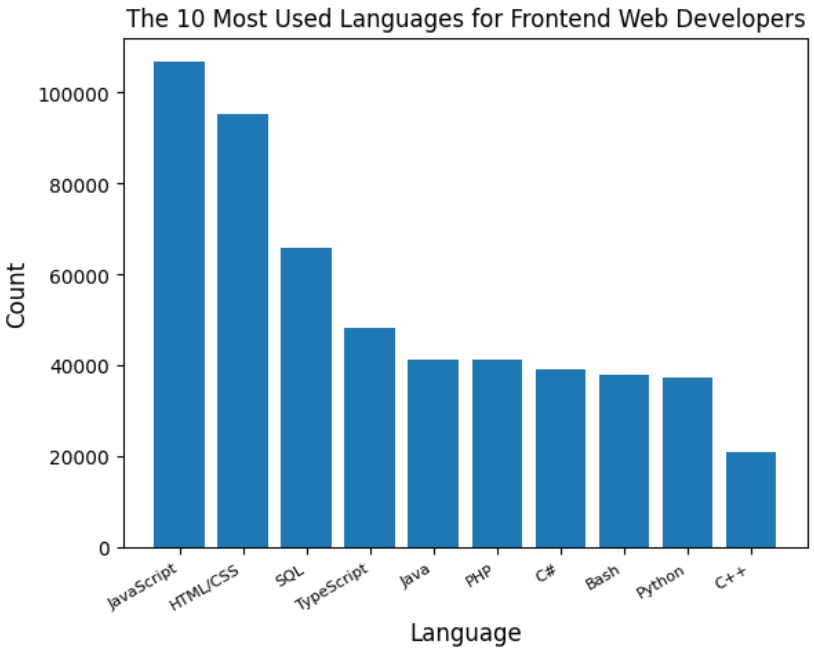

The 10 Most Used Programming Languages in 2019
Count
60000
50000
40000
30000
20000
10000
0
JavaScript
HTML/CSS
SQL
Python
Java
Bash
C#
PHP
C++
TypeScript
Language

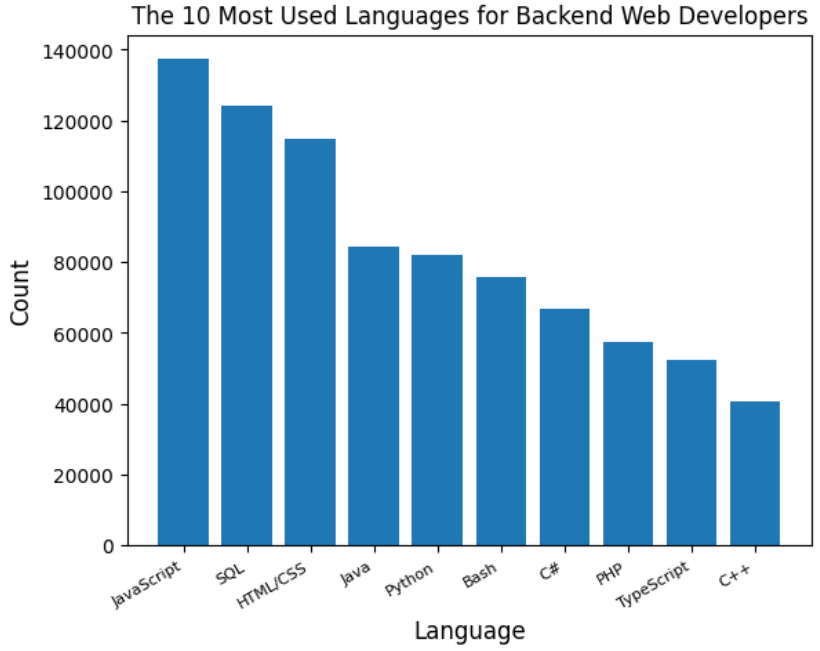
Result

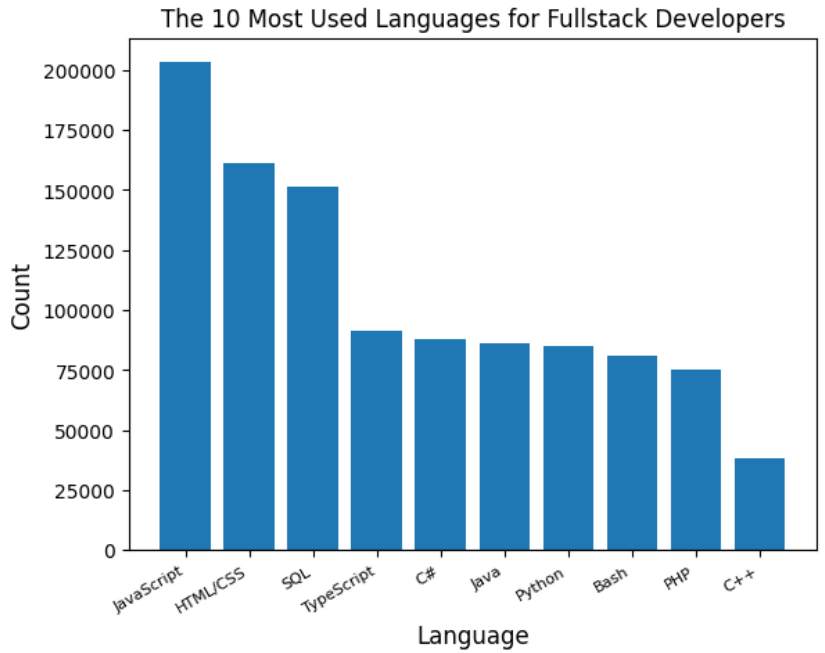


- JavaScript is still the first most used language in 2015, 2019 and 2024
- Python and TypeScript moved to a higher position from 2015 to 2024
- Java moved to a lower position from 2015 to 2024

Result

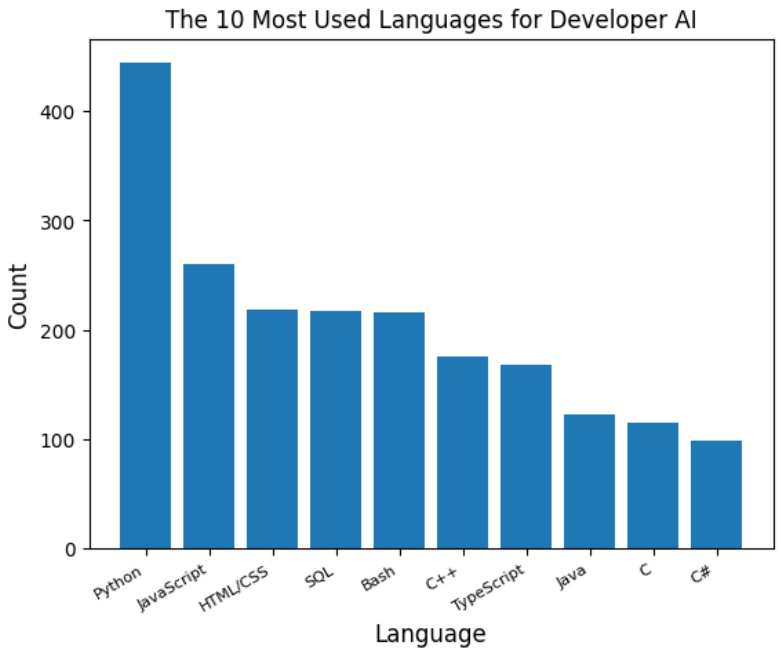


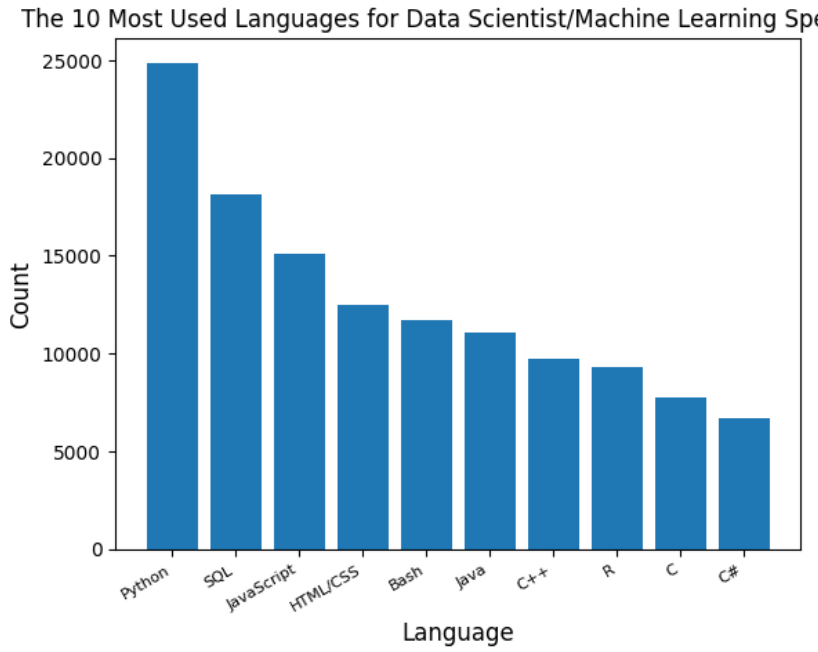


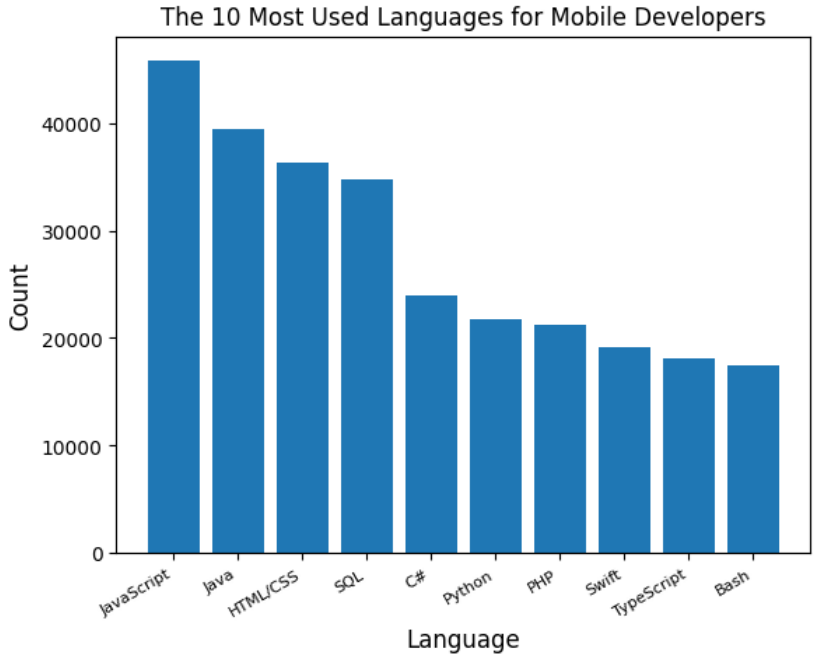


- The most used languages differ among various occupations.
- JavaScript, HTML/CSS and SQL are the top three most used languages for Web Developers.

Result







- Python is the 1st most used languages for Data Scientists/Machine Learning Specialists and AI Developers
- Java is the 2nd most used languages for Mobile Developers
- SQL has appeared numerously in the top five most used languages of our analysis

Result

- Frequent items and association rules for 2015 survey

[‘JavaScript, ‘PHP’]

Confidence (PHP => JavaScript) = 0.77

- Frequent items and association rules for 2016 survey

[‘JavaScript, ‘PHP’]

Confidence (PHP => JavaScript) = 0.77

Result

- Frequent items and association rules for 2017 survey (partial)

[‘AngularJS’, ‘JavaScript’]

Confidence (AngularJS=>JavaScript) = 0.91

[JavaScript, Node.js]

Confidence (Node.js => JavaScript) = 0.92

- Frequent items and association rules for 2018 survey (partial)

[‘Angular’, ‘CSS’, ‘HTML’]

Confidence (Angular, CSS => HTML) = 0.98

Confidence (Angular, HTML => CSS) = 0.96

[‘Angular’, ‘HTML’, ‘JavaScript’]

Confidence (Angular, HTML => JavaScript) = 0.95

Confidence (Angular, JavaScript => HTML) = 0.90

Result

- Frequent items and association rules for 2019 survey (partial)

[‘JavaScript’, ‘Node.js’]

Confidence (Node.js => JavaScript) = 0.93

[‘JavaScript’, ‘React.js’]

Confidence (React.js => JavaScript) = 0.94

[‘JavaScript’, ‘jQuery’]

Confidence (jQuery => JavaScript) = 0.92

- Frequent items and association rules for 2020 survey (partial)

[‘JavaScript’, ‘Node.js’]

Confidence (Node.js => JavaScript) = 0.93

[‘JavaScript’, ‘React.js’]

Confidence (React.js => JavaScript) = 0.93

[‘JavaScript’, ‘jQuery’]

Confidence (jQuery => JavaScript) = 0.93

Result

- Frequent items and association rules for 2021 survey (partial)

[‘AWS’, ‘Docker’, ‘Git’]

Confidence (AWS, Docker => Git) = 0.94

[‘AWS’, ‘Git’, ‘JavaScript’]

Confidence (AWS, JavaScript => Git) = 0.91

- Frequent items and association rules for 2022 survey (partial)

[‘JavaScript’, ‘Node.js’]

Confidence (Node.js => JavaScript) = 0.91

[‘JavaScript’, ‘React.js’]

Confidence (React.js => JavaScript) = 0.90

[‘JavaScript’, ‘jQuery’]

Confidence (jQuery => JavaScript) = 0.92

Result

- Frequent items and association rules for 2023 survey (partial)

[‘ChatGPT’, ‘GitHub Copilot’]

Confidence (GitHub Copilot => ChatGPT) = 0.86

[‘JavaScript’, ‘Node.js’]

Confidence (Node.js => JavaScript) = 0.91

[‘JavaScript’, ‘React’]

Confidence (React => JavaScript) = 0.89

- Frequent items and association rules for 2024 survey (partial)

[‘ChatGPT’, ‘GitHub Copilot’]

Confidence (GitHub Copilot => ChatGPT) = 0.82

[‘JavaScript’, ‘Node.js’]

Confidence (Node.js => JavaScript) = 0.90

[‘Pip’, ‘Python’]

Confidence (Pip => Python) = 0.88

Application

- Create and customize study guide for CS/IT related major students to help them enter work force based on their career path decision.
- Provide insights for developers who want to switch career path and/or strengthen their skillset.
- Identifying new popular languages
  - Help educators design suitable courses
  - Businesses keep up with the latest technology
  - Good opportunity for researchers to investigate the cause
- Give directions to future research