



## Data Analyst Capstone Project

[LinkedIn](#)

---

Julián Uribe Castañeda

Coursera

14 de febrero de 2023

# Outline

- (✓) Executive Summary
- (✓) Introduction
- (✓) Methodology
- (✓) Results
- (✓) Popular Languages
- (✓) Job Postings
- (✓) Programming Language Trends
- (✓) Database Trends
- (✓) Dashboard
- (✓) Discussion
- (✓) Conclusions
- (✓) Appendix



# Executive Summary



# Executive Summary

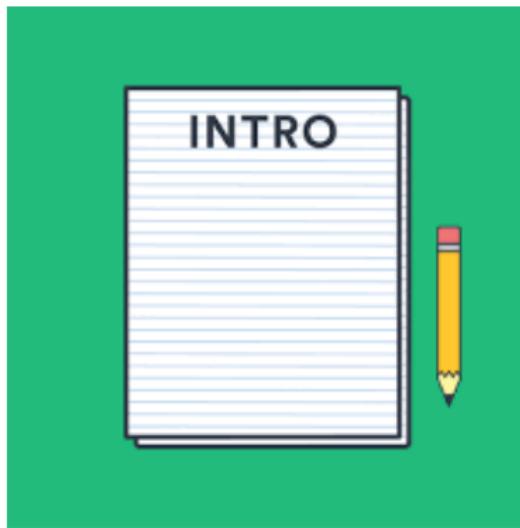
## ***Summary of methodologies***

- (✓) Data collection
- (✓) Data Wrangling
- (✓) Exploratory Data Analysis
- (✓) Data Visualization

## ***Summary of all results***

- (✓) Data analysis results
- (✓) Data visualization results

# Introduction



# Introduction

Collect the best programming skills that are most in demand from various sources, including:

- (✓) Job offers
- (✓) Training portals
- (✓) Surveys

Once the data is collected, it begins to analyze and identify important factors such as:

- (✓) Major programming languages in demand
- (✓) Top database skills in demand
- (✓) Popular IDEs

Once this is complete, that data will be prepared for analysis using data dispute techniques.

# Methodology



# Methodology

## Data Collection

The main objective in this part is the collection and exploration of data. This task is done as follows:

- (✓) The Data is collected via API and webscraping
- (✓) The obtained data set is explored (using some important Jupyter Notebook libraries)

## Data Wrangling

This part removes errors and combining complex data sets to make them more accessible and easier to analyze. Some of the tasks we have in this methodology are:

- (✓) Finding Missing Values
- (✓) Determine Missing Values
- (✓) Finding Duplicates
- (✓) Removing Duplicates
- (✓) Normalizing Data

## Exploratory Data Analysis

Here the data provided is analyzed and its main characteristics are summarized. For example some important tasks in this methodology are:

- (✓) Finding the outliers
- (✓) Finding the correlation between some variables

## Data Visualization

Data visualization is the representation of data through use of common graphics, such as charts, plots, infographics, and even animations. These visual displays of information communicate complex data relationships and data-driven insights in a way that is easy to understand.

# Results



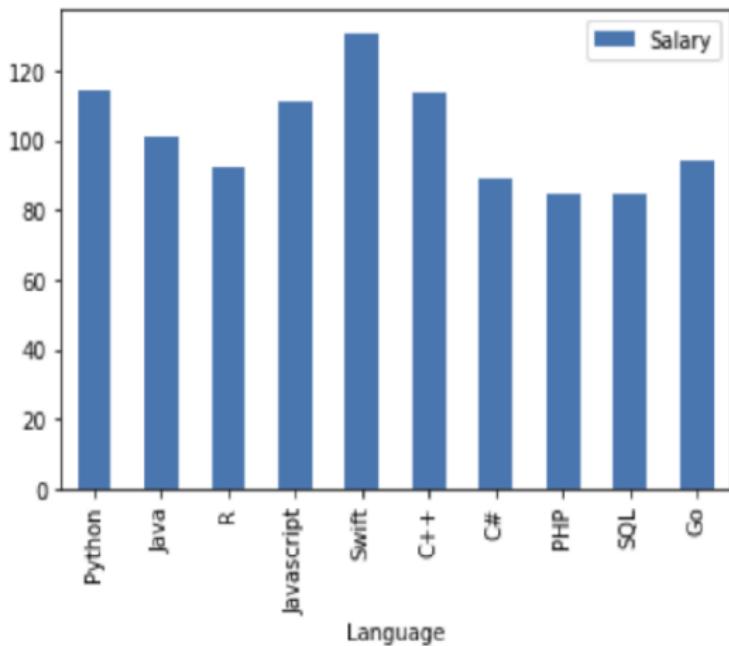
All the lab answers are given in the following link.

Cognos IBM

# Popular Languages

## Popular Languages

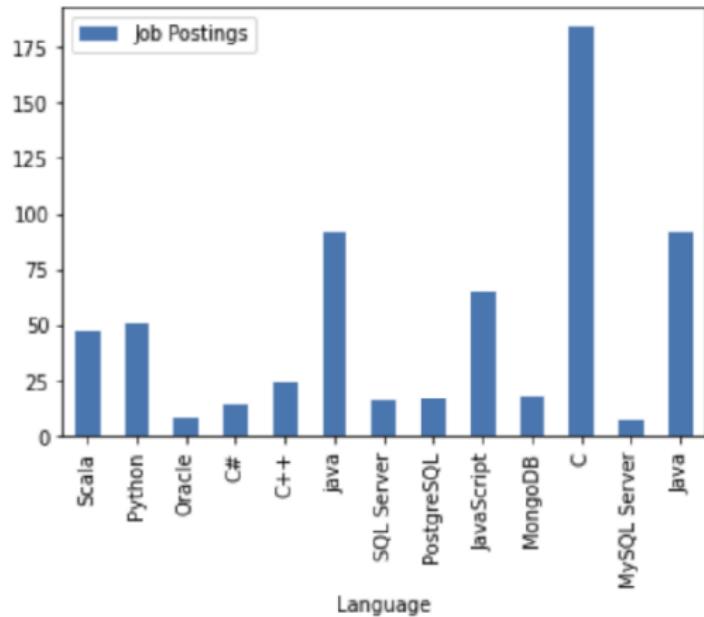
	Language	Salary	+
0	Python	114.383	
1	Java	101.013	
2	R	92.037	
3	Javascript	110.981	
4	Swift	130.801	
5	C++	113.865	
6	C#	88.726	
7	PHP	84.727	
8	SQL	84.793	
9	Go	94.082	



# Job Postings

## Job Postings

	Language	Job Postings
0	Scala	47
1	Python	51
2	Oracle	8
3	C#	14
4	C++	24
5	java	92
6	SQL Server	16
7	PostgreSQL	17
8	JavaScript	65
9	MongoDB	18
10	C	184
11	MySQL Server	7
12	Java	92

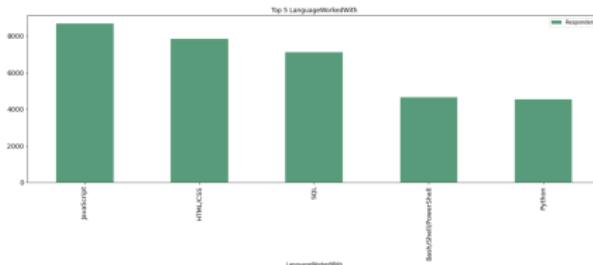


# Programming Language Trend

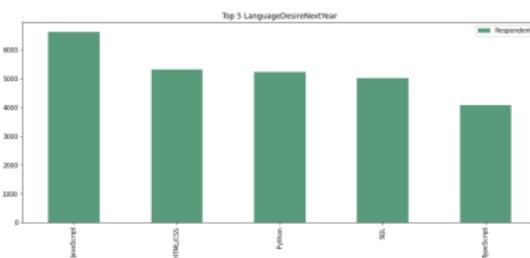


# Programming Language Trends

## Current Year



## Next Year



## Findings

- (✓) The most demanded programming language are JavaScript and HTML/CSS.
- (✓) The popularity of Python and TypeScript are growing.

## Implications

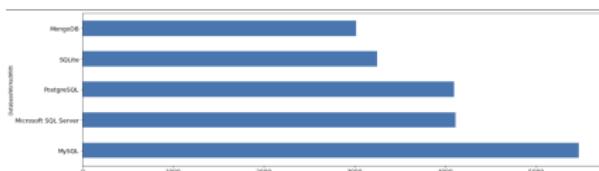
- (✓) JavaScript and HTML/CSS are programming languages that are in high demand and are consistent.
- (✓) Due to the growth of Python it is possible that in a few years this programming language will surpass JavaScript.

# Database Trend

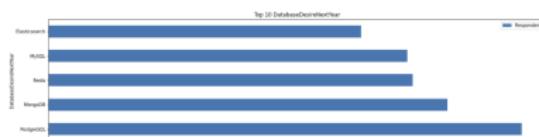


# Database Trends

## Current Year



## Next Year



## Findings

- (✓) PostgreSQL is in high demand and is growing very quickly. The community has as a priority to learn this database management system.
- (✓) Demand for MySQL is decreasing very rapidly, while MongoDB and Redis are growing.

## Implications

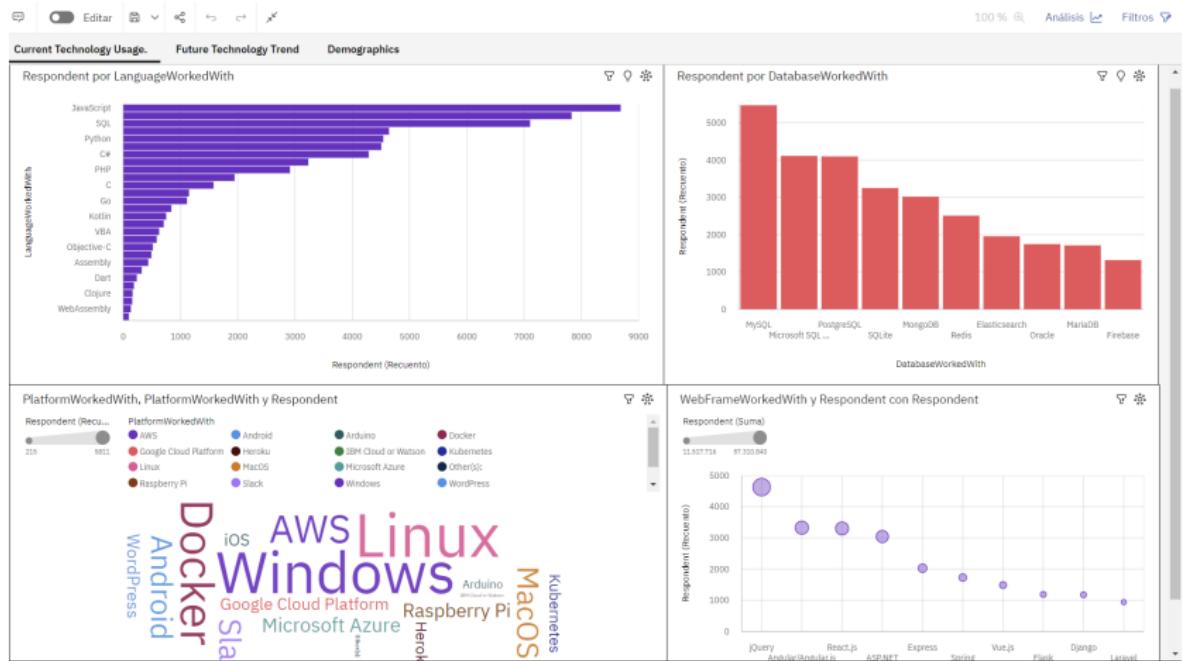
- (✓) PostgreSQL and MongoDB are in high demand and are very consistent, with PostgreSQL being the database management system that the community most wants to understand and update.
- (✓) the preferred database management system is open source.

# Dashboard

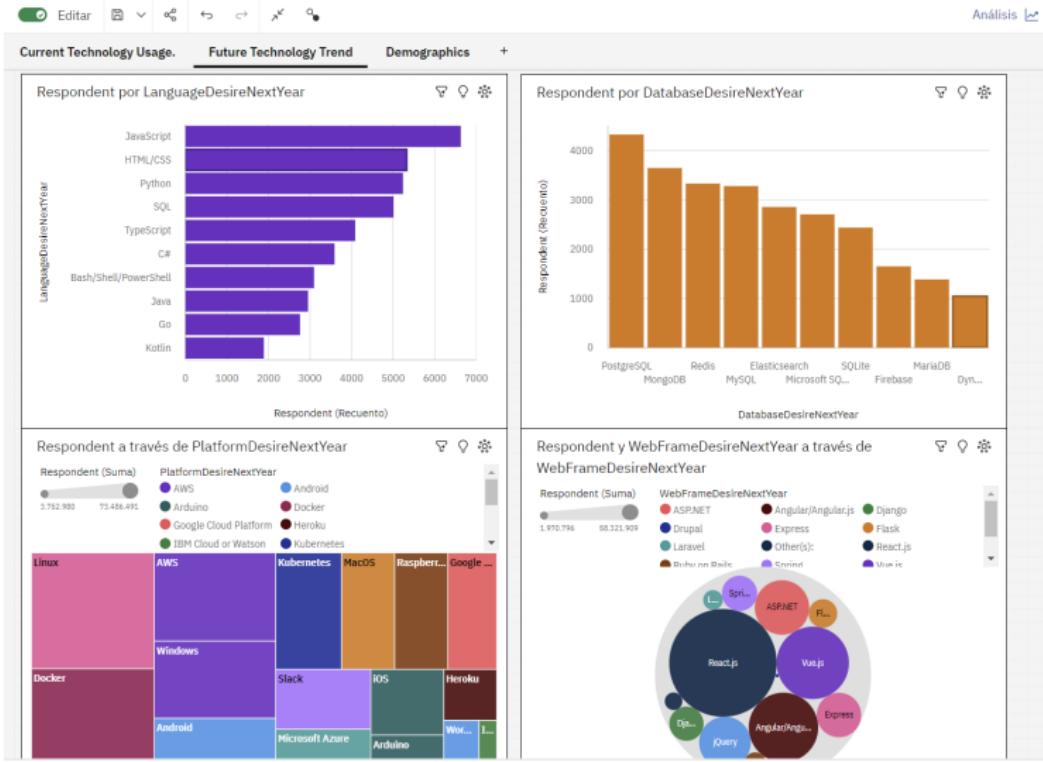


Link GitHub- IBM Cognos

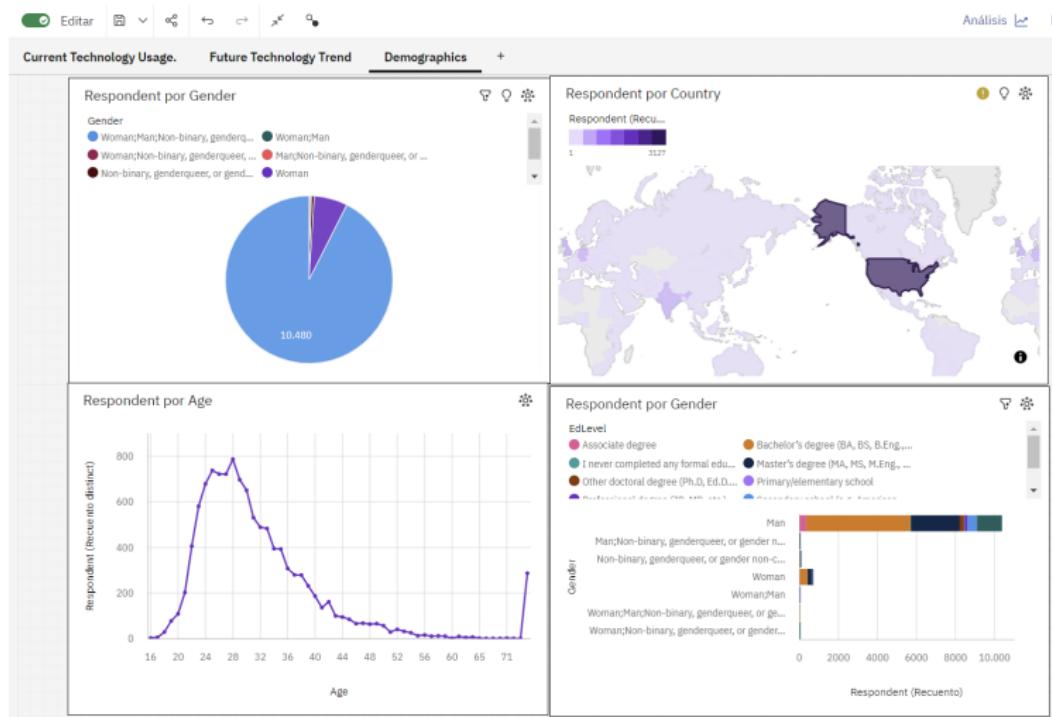
## Dashboard Tab 1



## Dashboard Tab 2



## Dashboard Tab 3



# Discussion

## Discussion

- (✓) New technologies make programmers constantly update themselves. This means learning a new programming language that can be more versatile than others in some particular aspect or dealing with new database query languages to stay ahead of the industry.
- (✓) The programming industry needs to be more flexible with gender differences.

# Conclusions

## Conclusions

- (✓) The most popular programming language is JavaScript.
- (✓) The most popular database are PostgreSQL and MySQL.
- (✓) The information technology field needs gender diversity.
- (✓) New technologies lead to varied trends and demands.

# Appendix

## Appendix

In the following link is the IBM notebooks that i used in this capstone project.

[IBM notebooks - GitHub](#)