

## Clapstone Project Final Dashboard

Marco Antonio Ramírez Godínez


February 2024

### Project Overview

1. This final project aims to analyze overall usage of languages and databases among IT community.
2. The data was collected from a survey in which IT experts were asked about preferences and usages of tools and programming languages. This project analyzes preferences in order to understand what is going on in the future (1 year approach).
3. The workflow for this assignment will be: Utilize python and sql to get the data (.csv file), clean the data, do some EDA analysis and then get key visualizations. Finally, we are going to create a dashboard in Power Bi for further analysis.
4. This dashboard is the last part of the project and consists of screenshots of the dashboard panel, as wells as M code and DAX scripts. Also, a brief explanation is being added of how this dashboard would be enhanced to support real time data. Txt files of m code and dax code has been added to github platform for easy manipulation.

## Part 1. M code


Since data was already cleaned in python and it was ready to be exploited, I only got the csv file right from the source folder then I changed some data types manually.

 Editor avanzado

Clapston\_Project\_Final



```
let
    Source = Csv.Document(File.Contents("C:\Users\c405424\Documents\Clapston_Project_Final.csv"),[Delimiter=";", Columns=20, Encoding=1252, QuoteStyle=QuoteStyle.None]),
    #"Promoted Headers" = Table.PromoteHeaders(Source, {PromoteAllScalars=true}),
    #"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"", Int64.Type},
    {"Respondent", type text},
    {"LanguageWorkedWith", type text},
    {"LanguageDesireNextYear", type text},
    {"DatabaseWorkedWith", type text},
    {"DatabaseDesireNextYear", type text},
    {"PlatformWorkedWith", type text},
    {"PlatformDesireNextYear", type text},
    {"WebFrameWorkedWith", type text},
    {"WebFrameDesireNextYear", type text},
    {"DevEnviron", type text},
    {"Country", type text},
    {"YearsCode", Int64.Type},
    {"CompTotal", Int64.Type},
    {"CompFreq", type text},
    {"ConvertedComp", Int64.Type},
    {"Age", Int64.Type},
    {"Gender", type text},
    {"Dependents", type text},
    {"FinalComp", type number}})
in
    #"Changed Type"
```

 No se han detectado errores de sintaxis.

## Part 2. DAX code

I used 8 measures to being able to count respondents' answers and filter out "no answers" label. Then, I used it to create RANK() function in which Top 10 respondents' answers would be capsuled. This helped me to later filter top 10 in both current and future context.

## 1. Aggregation functions to summarize answer counts:

```
db_current =
CALCULATE(COUNT(Clapston_Project_Final[Respondent]),FILTER(Clapston_Project_Final,Clapston_Project_Final[DatabaseWorkedWith] <> "No answer"))
db_future =
CALCULATE(COUNT(Clapston_Project_Final[Respondent]),FILTER(Clapston_Project_Final,Clapston_Project_Final[DatabaseDesireNextYear] <> "No answer"))
language_current =
CALCULATE(COUNT(Clapston_Project_Final[Respondent]),FILTER(Clapston_Project_Final,Clapston_Project_Final[LanguageWorkedWith] <> "No answer"))
language_future =
CALCULATE(COUNT(Clapston_Project_Final[Respondent]),FILTER(Clapston_Project_Final,Clapston_Project_Final[LanguageDesireNextYear] <> "No answer"))
```

## 2. Ranking functions to filter top 10:

```
ranking_10_db =
RANKX(ALL(Clapston_Project_Final[DatabaseWorkedWith]),[db_current],,DESC,Skip)
ranking_10_future_db =
RANKX(ALL(Clapston_Project_Final[DatabaseDesireNextYear]),[db_future],,DESC,Skip)
ranking_10_future_language =
RANKX(ALL(Clapston_Project_Final[LanguageDesireNextYear]),[language_future],,DESC,Skip)
ranking_10__current_language =
RANKX(ALL(Clapston_Project_Final[LanguageWorkedWith]),[language_current],,DESC,Skip)
```

Part 3. Dashboard Panel

- Dashboard consists of 3 kinds of plots: bar plot, Treemap and Horizontal bar plot.
- Colors are set to red for most popular languages bar plots and blue for most demanded languages bar plots. Treemaps are set to default colors.
- All plots have been added with labels.
- Everything can be dynamically filtered not only with Country and Gender details, but also with 2 buttons for current and future context data.
- Additional cards with count of users are located in the left-center of the dashboard.



#### Part 4. Enhancing dashboard

Some of the key points that were not aboard in this final project were:

- This project is not thought to be a real time data repository due to the limitations in its design. In other words, from the source to the data modeling, there could be problems handling changes in data structure. ¿How can we solve it? Including dimension tables and incorporating a star schema, also changing the source M code to be cumulative.
- There is also a need to analyze whether country and gender affects preferences among users.