Initialization

Set paths

Determinamos os caminhos, para salvar e processar documentos, para utilizar pacotes e dicionários:

```
In[ ]:= Clear[path, path2, path3, path4, path5]
        path = "C:\\Users\\aleja\\Google Drive\\ObservatorioPesquisas2022-2023\\datasets\\";
        path2 = "C:\\Users\\aleja\\Google Drive\\ObservatorioPesquisas2022-2023\\processing\\";
        path3 = "C:\\Users\\aleja\\Google Drive\\ObservatorioPesquisas2022-2023\\packages\\";
        path4 =
          "C:\\Users\\aleja\\Google Drive\\ObservatorioPesquisas2022-2023\\diccionaries\\";
        path5 = "C:\\Users\\aleja\\Google Drive\\ObservatorioPesquisas2022-2023\\pictures\\";
```

Importando os datasets

```
Clear[authors, fullauthors]
authors = {"MF_ES", "MF_EN", "MF_PT"};
```

Importamos o csv e o chamamos data[i], em que i são as autoras

```
In[ ]:= Clear[data]
        Do[data[i] = Import[path <> i <> ".csv"], {i, authors}];
```

Eliminamos todas as linhas sem informação

```
In[ ]:= Clear[data1]
        Do[data1[i] = Select[data[i], Not[# == {}] &];
          Print[i <> " → " <> ToString[Length[data[i]]]], {i, authors}];
```

```
MF_ES → 301787

MF_EN → 181711

MF_PT → 2092874
```

```
In[ ]:= Clear[data2]
        Do[data2[i] = Drop[data1[i], 1], {i, authors}];
```

Dentre todos os dados, escolhemos o texto, a linguagem e a data.

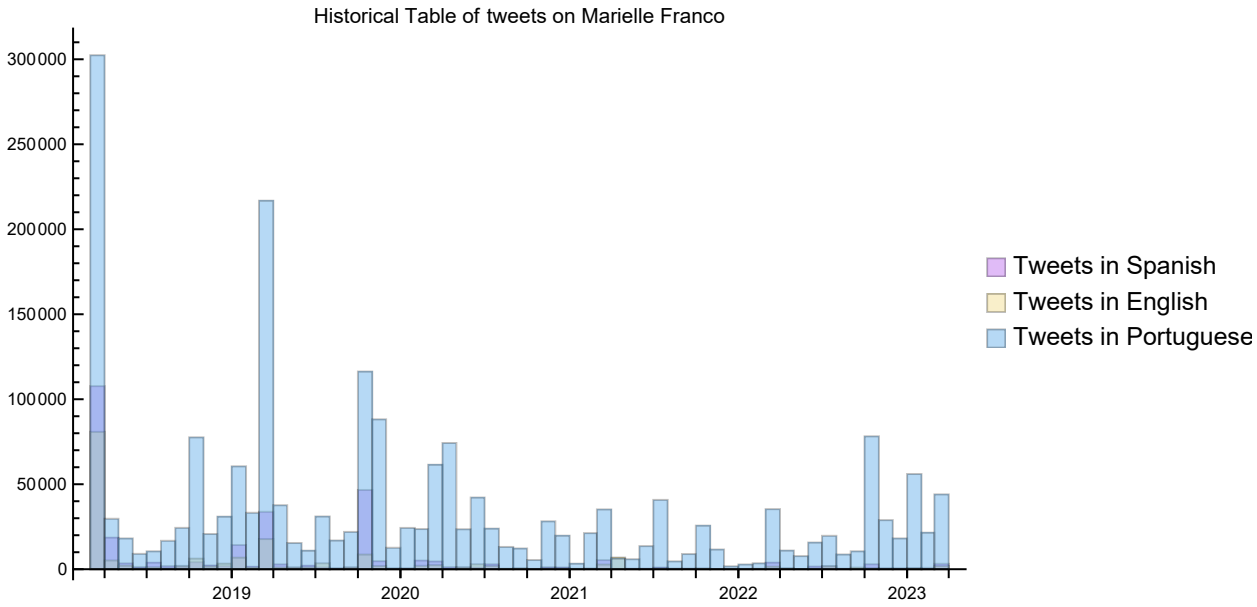
```
In[ ]:= Clear[data3]
        Do[data3[i] = data2[i][[All, {2, 4, 8}]], {i, authors}];
```

Fazemos uma tabela histórica com os tweets

```
In[ ]:= Clear[dateES, datePT, dateEN]
        dateES = Map[DateObject, data3["MF_ES"][[All, 1]]];
        dateEN = Map[DateObject, data3["MF_EN"][[All, 1]]];
        datePT = Map[DateObject, data3["MF_PT"][[All, 1]]];
```

In[ ]:=
```
TableDate =
 DateHistogram[{dateES, dateEN, datePT}, "Month", ChartStyle → "Pastel", ImageSize → 500,
   PlotLabel → "Historical Table of tweets on Marielle Franco", ChartLegends →
    {"Tweets in Spanish", "Tweets in English", "Tweets in Portuguese"}] // TableForm
```

Out[ ]//TableForm=



In[ ]:=
```
Export["TABLEENG.png", TableDate]
```

Out[ ]=

TABLEENG.png

In[ ]:=
```
Clear[languages, tweets]
languages = {"es", "pt", "en"};
tweets["es"] = Select[data3["MF_ES"], #[[3]] == "es" &];
tweets["pt"] = Select[data3["MF_PT"], #[[3]] == "pt" &];
tweets["en"] = Select[data3["MF_EN"], #[[3]] == "en" &];
Table[Length[tweets[j]], {j, languages}]
```

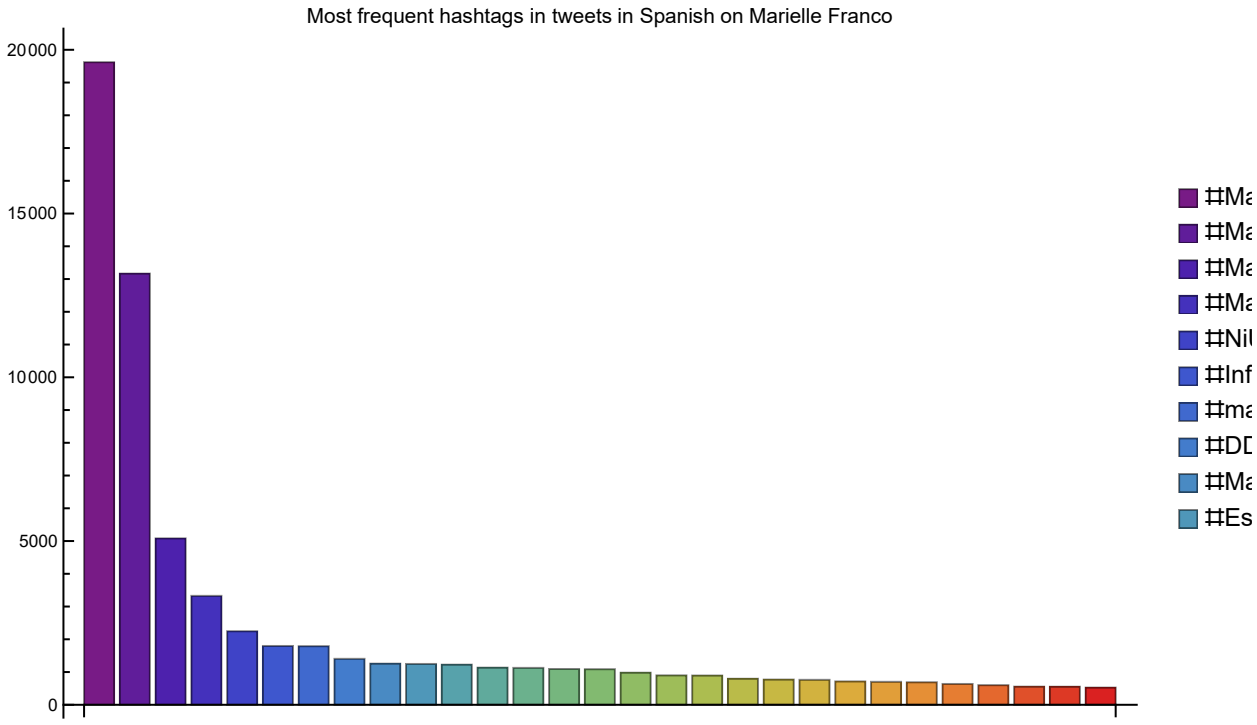
Out[ ]=

{301 786, 2 092 873, 181 710}

In[ ]:=
```
Clear[hashtags]
hashtags[text_] := Reverse[SortBy[Tally[Flatten[StringCases[text,
       RegularExpression["#([a-z0-9]|[À-ü]|_)+"], IgnoreCase → True]]], Last]];
```

```
In[ ]:= BarChart[Drop[Take[hashtags[tweets["es"]][[All, 2]], 30], {2}][[All, 2]], ImageSize → 600,
 ChartLegends → Drop[Take[hashtags[tweets["es"]][[All, 2]], 30], {2}][[All, 1]],
 ChartStyle → "Rainbow",
 PlotLabel → Style["Most frequent hashtags in tweets in Spanish on Marielle Franco"],
 ChartLayout → "Grouped", RotateLabel → True]
```

Out[ ]=



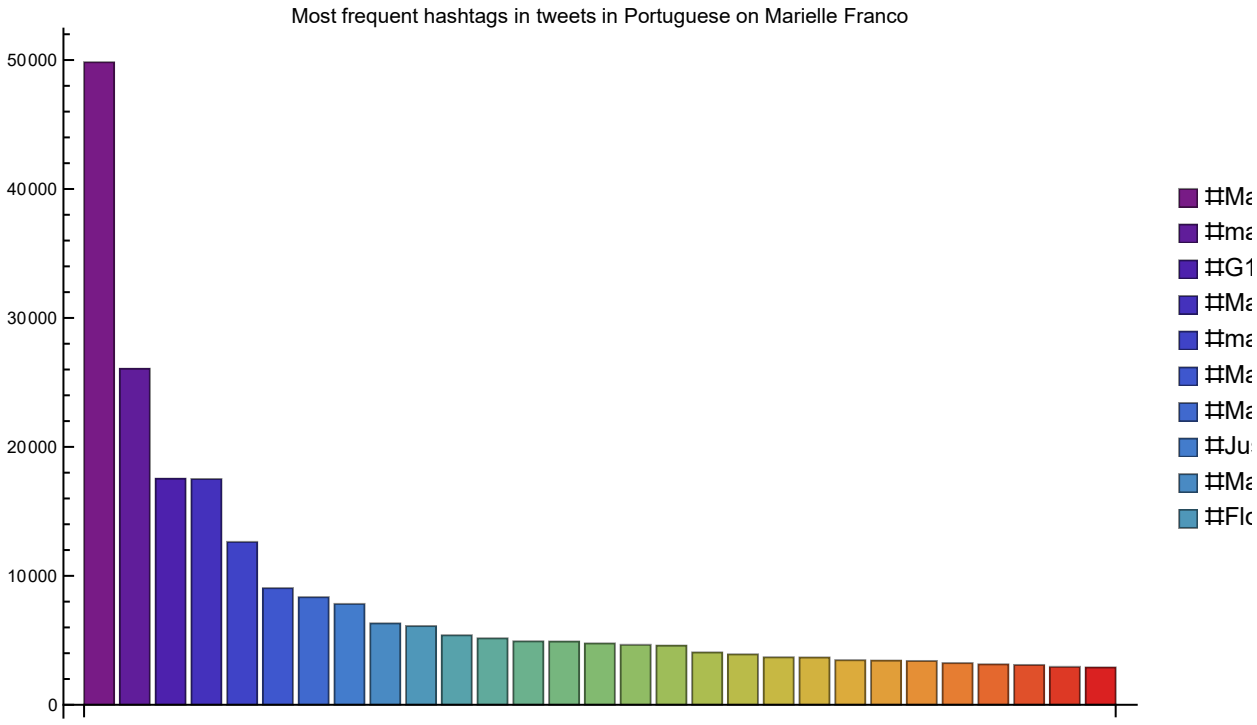

```
In[ ]:= Export[path5 <> "BarChartMFES.png", %]
```

Out[ ]=

```
C:\Users\aleja\Google Drive\ObservatorioPesquisas2022-2023\pictures\BarChartMFES.png
```

In[ ]:=
```
BarChart[Drop[Take[hashtags[tweets["pt"]][[All, 2]], 30], {2}][[All, 2]], ImageSize → 600,
 ChartLegends → Drop[Take[hashtags[tweets["pt"]][[All, 2]], 30], {2}][[All, 1]],
 ChartStyle → "Rainbow",
 PlotLabel → Style["Most frequent hashtags in tweets in Portuguese on Marielle Franco"],
 ChartLayout → "Grouped", RotateLabel → True]
```

Out[ ]=



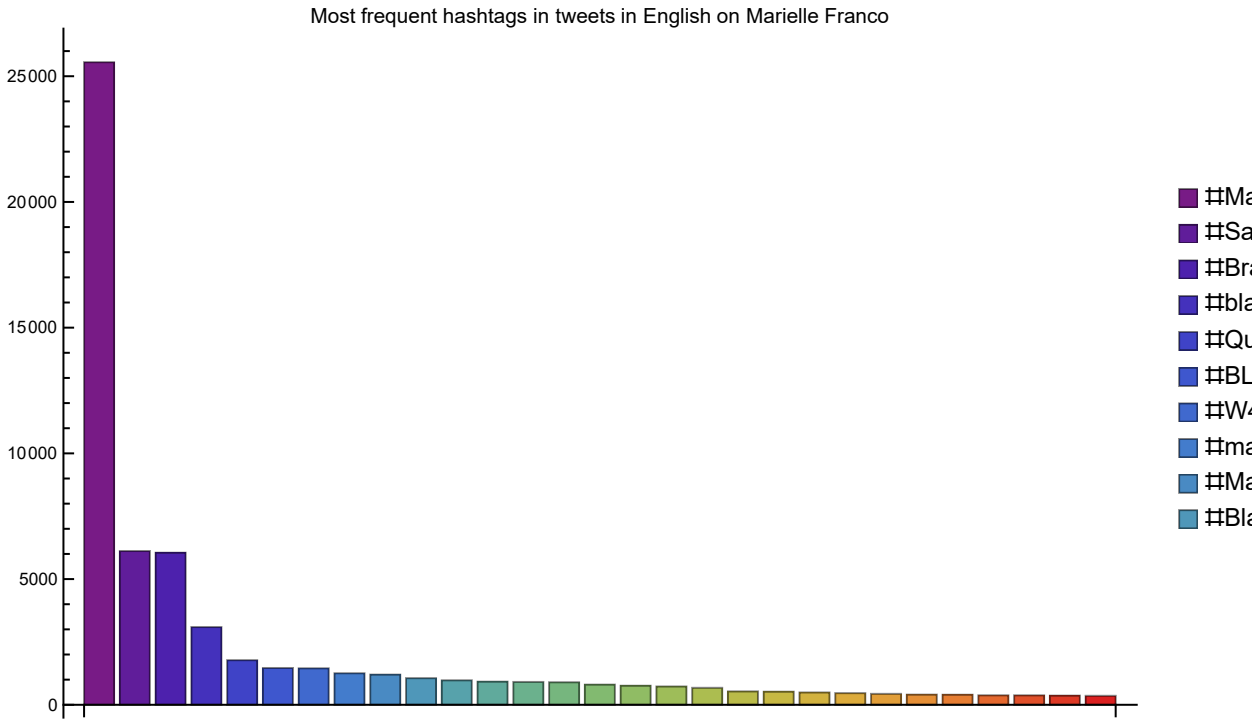
In[ ]:=
```
Export[path5 <> "BarChartMFPT.png", %]
```

Out[ ]=

```
C:\Users\aleja\Google Drive\ObservatorioPesquisas2022-2023\pictures\BarChartMFPT.png
```

```
In[ ]:= BarChart[Drop[Take[hashtags[tweets["en"]][[All, 2]], 30], {2}][[All, 2]], ImageSize → 600,
 ChartLegends → Drop[Take[hashtags[tweets["en"]][[All, 2]], 30], {2}][[All, 1]],
 ChartStyle → "Rainbow",
 PlotLabel → Style["Most frequent hashtags in tweets in English on Marielle Franco"],
 ChartLayout → "Grouped", RotateLabel → True]
```

Out[ ]=




```
In[ ]:= Export[path5 <> "BarChartMFEN.png", %]
```

Out[ ]=

```
C:\Users\aleja\Google Drive\ObservatorioPesquisas2022-2023\pictures\BarChartMFEN.png
```

```
In[ ]:= Clear[usernames]
usernames[text_] := Reverse[SortBy[Tally[Flatten[StringCases[text,
        RegularExpression["@([a-z0-9]|[À-ü]|_)+"], IgnoreCase → True]]], Last]];
```

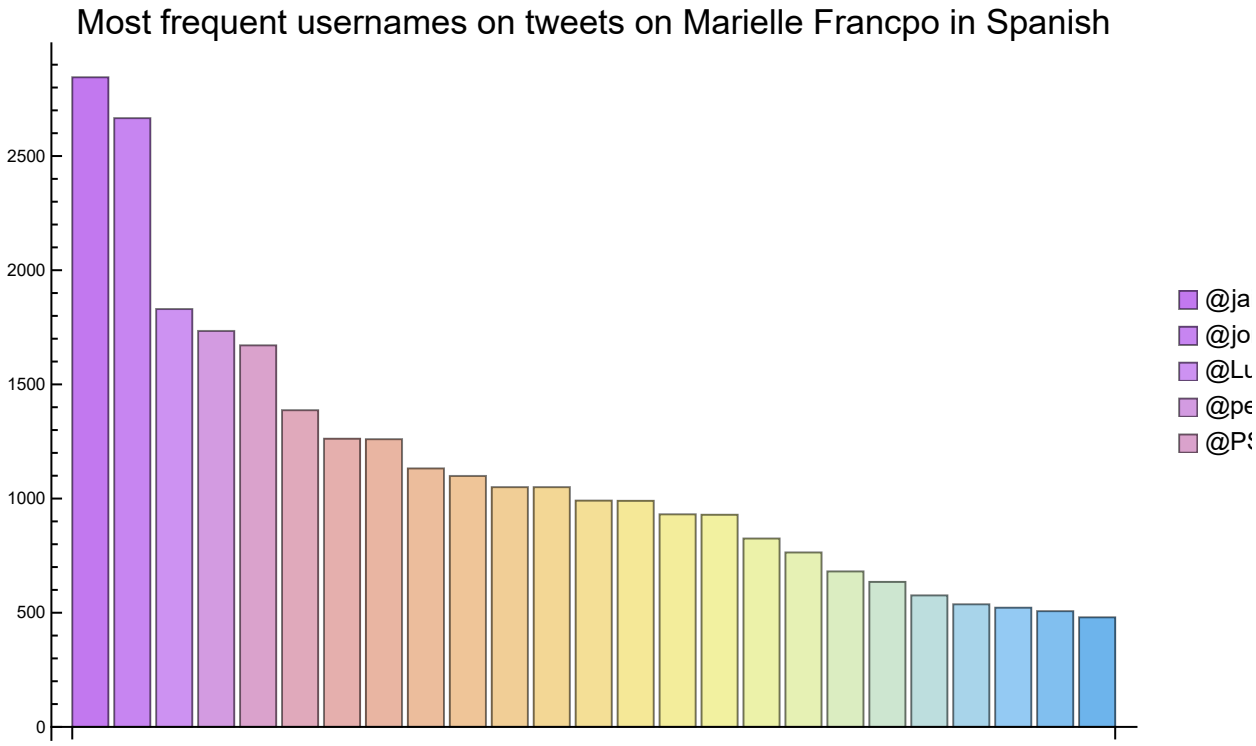
```
In[ ]:= Clear[us, US]
us = Take[usernames[tweets["es"][[All, 2]], 25][[All, 1]]
US = us /. {"@LunaGamp" → "@■■■■", "@jbellolio" → "@■■■■", "@dimash_official" → "@■■■■",
    "@PrakritiMaduro" → "@■■■■", "@agustin_gut" → "@■■■■", "@Cuauhtemoc_1521" → "@■■■■",
    "@trasnochocult" → "@■■■■", "@scastaldi9" → "@■■■■", "@culturacolectiv" → "@■■■■",
    "@DramaturgiaMex" → "@■■■■", "@dianadep1" → "@■■■■", "@albertopetro2" → "@■■■■"};
```

Out[ ]=

```
{@jairbolsonaro, @jornalnacional, @LulaOficial, @periodicopausa, @PSOLOficial,
 @dilmabr, @MidiaNINJA, @jeanwyllys_real, @JusticaGovBR, @monica_benicio,
 @MarielleFranco, @LunaGamp, @Pagina12, @izquierdadiario, @teleSURtv,
 @mariellefranco, @todonoticias, @amnistiaespana, @MichelTemer, @jbellolio,
 @LulapeloBrasil, @aniellefranco, @desalambre, @InfoNodal, @eldiarioes}
```

In[ ]:=
```
BarChart[Take[usernames[tweets["es"]⟦All, 2⟧], 25]⟦All, 2⟧,
 ImageSize → 600, ChartLegends → US, ChartStyle → "Pastel", PlotLabel →
  Style["Most frequent usernames on tweets on Marielle Francpo in Spanish", 18]]
```

Out[ ]=



In[ ]:=
```
Export[path5 <> "BarChartUMFES.png", %]
```

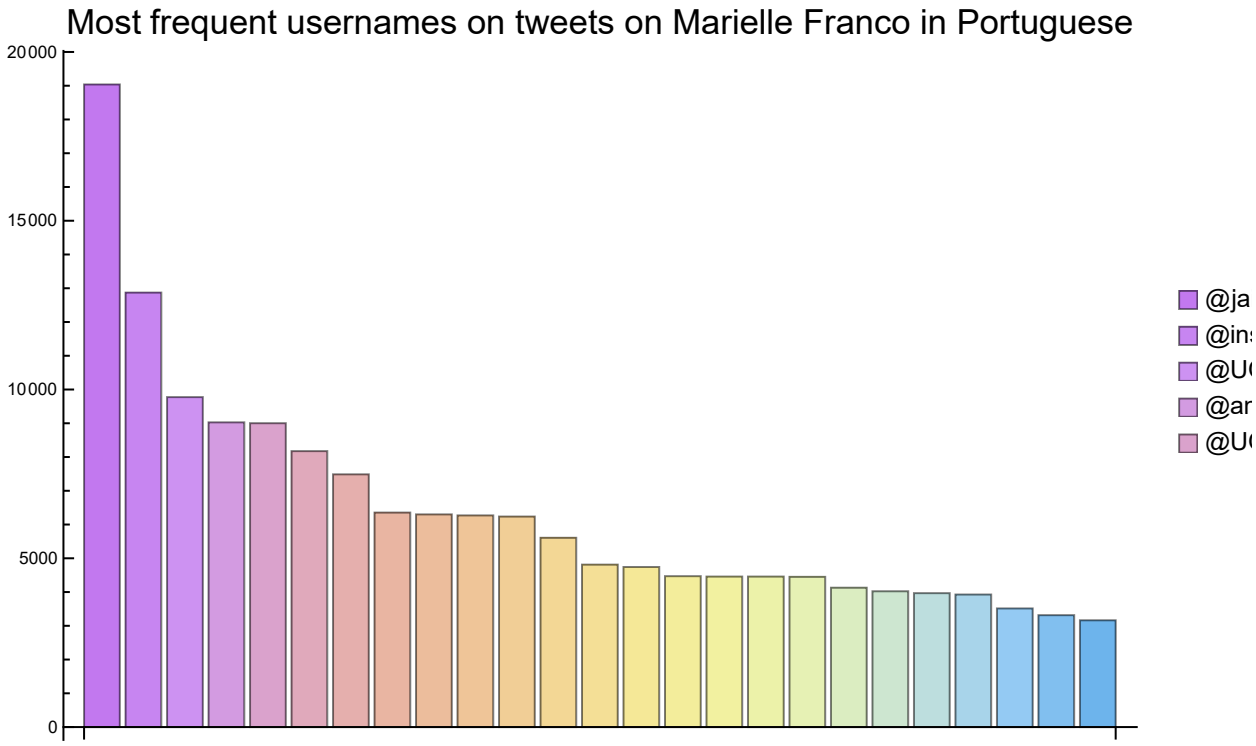
Out[ ]=
```
C:\Users\aleja\Google Drive\ObservatorioPesquisas2022-2023\pictures\BarChartUMFES.png
```

In[ ]:=
```
Clear[us, US]
us = Take[usernames[tweets["pt"]⟦All, 2⟧], 25]⟦All, 1⟧
US =
  us /. {"@zehdeabreu" → "@■■■■", "@Rconstantino" → "@■■■■", "@dimash_official" → "@■■■■",
    "@PrakritiMaduro" → "@■■■■", "@agustin_gut" → "@■■■■", "@Cuauhtemoc_1521" → "@■■■■",
    "@trasnochocult" → "@■■■■", "@scastaldi9" → "@■■■■", "@culturacolectiv" → "@■■■■",
    "@DramaturgiaMex" → "@■■■■", "@dianadep1" → "@■■■■", "@albertopetro2" → "@■■■■"};
```

Out[ ]=
```
{@jairbolsonaro, @inst_marielle, @UOL, @anielle_franco, @UOLNoticias,
 @YouTube, @aniellefranco, @DefesaGovBr, @MarceloFreixo, @LulaOficial,
 @SF_Moro, @ricardostuckert, @FlavioDino, @zehdeabreu, @Estadao,
 @mariellefranco, @katyperry, @lindberghfarias, @JornalOGlobo, @CarlosBolsonaro,
 @Rconstantino, @felipeneto, @monica_benicio, @wilsonwitzel, @anistiabrasil}
```

In[ ]:=
```
BarChart[Take[usernames[tweets["pt"]⟦All, 2⟧], 25]⟦All, 2⟧,
 ImageSize → 600, ChartLegends → US, ChartStyle → "Pastel", PlotLabel →
  Style["Most frequent usernames on tweets on Marielle Franco in Portuguese", 18]]
```

Out[ ]=



In[ ]:=
```
Export[path5 <> "BarChartUMFPT.png", %]
```

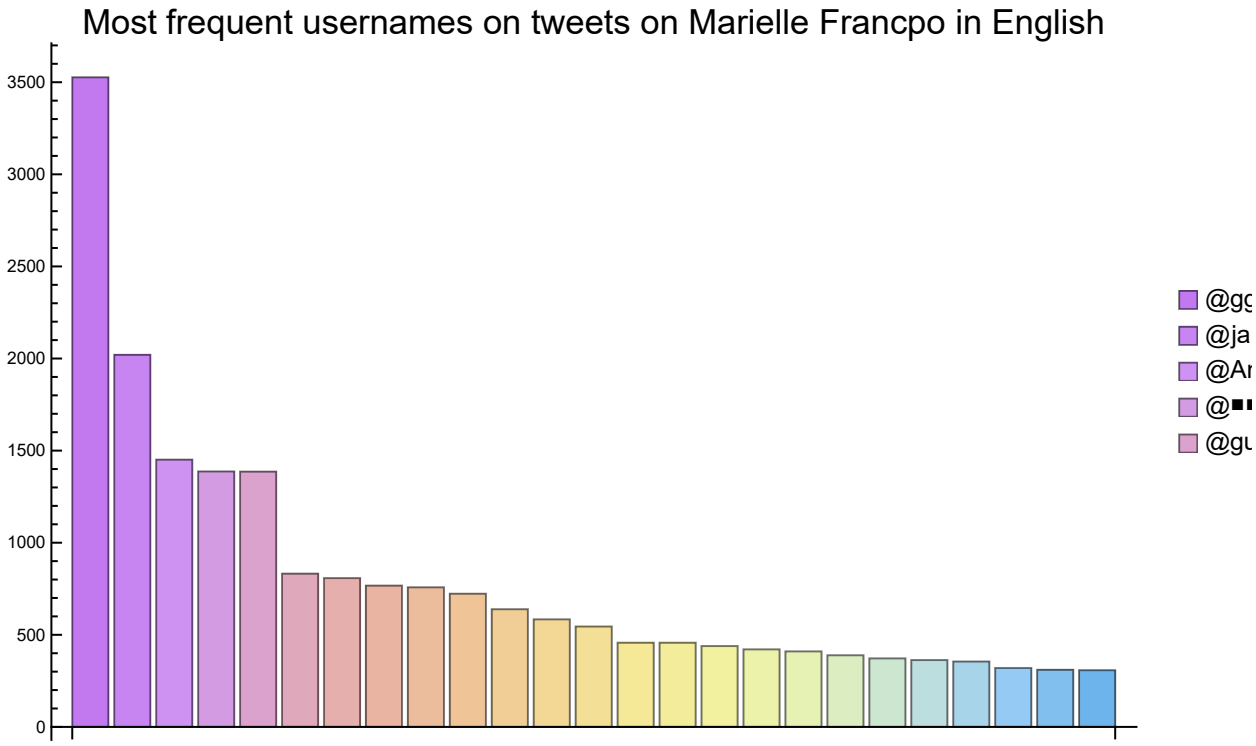
Out[ ]=
```
C:\Users\aleja\Google Drive\ObservatorioPesquisas2022-2023\pictures\BarChartUMFPT.png
```

In[ ]:=
```
Clear[us, US]
us = Take[usernames[tweets["en"]⟦All, 2⟧], 25]⟦All, 1⟧
US =
  us /. {"@ShaunKing" → "@■■■■", "@MrDavidAWilson" → "@■■■■", "@davidmirandario" → "@■■■■",
    "@bydorianadiaz" → "@■■■■", "@drkeishakhan" → "@■■■■", "@KennedyAlencar" → "@■■■■",
    "@feminine_harbor" → "@■■■■", "@profsassy" → "@■■■■", "@KiaLCaldwell" → "@■■■■",
    "@KiaLCaldwell" → "@■■■■", "@dianadep1" → "@■■■■", "@albertopetro2" → "@■■■■"};
```

Out[ ]=
```
{@ggreenwald, @jairbolsonaro, @AmnestyUK, @ShaunKing, @guardian,
 @TheRoot, @davidmirandario, @YouTube, @TheInterceptBr, @MrDavidAWilson,
 @YourAnonCentral, @samiabomfim, @monica_benicio, @theintercept, @mariellefranco,
 @bydorianadiaz, @drkeishakhan, @KennedyAlencar, @Independent, @feminine_harbor,
 @profsassy, @renatasouzario, @AOC, @KiaLCaldwell, @aniellefranco}
```

In[ ]:=
```
BarChart[Take[usernames[tweets["en"]⟦All, 2⟧], 25]⟦All, 2⟧,
 ImageSize → 600, ChartLegends → US, ChartStyle → "Pastel", PlotLabel →
  Style["Most frequent usernames on tweets on Marielle Francpo in English", 18]]
```

Out[ ]=



In[ ]:=
```
Export[path5 <> "BarChartUMFEN.png", %]
```

Out[ ]=

```
C:\Users\aleja\Google Drive\ObservatorioPesquisas2022-2023\pictures\BarChartUMFEN.png
```