M2.859_20211_A9-Enunciado_gbonillas

January 12, 2022

M2.859 · Visualización de datos · Práctica, Parte 2

2021-1 · Máster universitario en Ciencia de datos (Data science)

Estudios de Informática, Multimedia y Telecomunicación

1 A9: Práctica Final (parte 2) - Wrangling data

El **wrangling data** es el proceso de transformar y mapear datos de un formulario de datos "sin procesar" a otro formato con la intención de hacerlo más apropiado y valioso para una variedad de propósitos posteriores, como el análisis. El objetivo del wrangling data es garantizar la calidad y la utilidad de los datos. Los analistas de datos suelen pasar la mayor parte de su tiempo en el proceso de disputa de datos en comparación con el análisis real de los datos.

El proceso de wrangling data puede incluir más manipulación, visualización de datos, agregación de datos, entrenamiento de un modelo estadístico, así como muchos otros usos potenciales. El wrangling data normalmente sigue un conjunto de pasos generales que comienzan con la extracción de los datos sin procesar de la fuente de datos, "removiendo" los datos sin procesar (por ejemplo, clasificación) o analizando los datos en estructuras de datos predefinidas y, finalmente, depositando el contenido resultante en un sumidero de datos para almacenamiento y uso futuro.

Para ello vamos a necesitar las siguientes librerías:

[4]:
```
from six import StringIO

from IPython.display import Image
from sklearn import datasets
from sklearn.decomposition import PCA
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.metrics import accuracy_score, confusion_matrix,␣
 ↪classification_report
from sklearn.tree import DecisionTreeClassifier, export_graphviz
import pydotplus
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
import pandas as pd
```

```
pd.set_option('display.max_columns', None)
```

2 1. Carga del conjunto de datos (1 punto)

Se ha seleccionado un conjunto de datos desde el portal Stack Overflow Annual Developer Survey, que examina todos los aspectos de la experiencia de los programadores de la comunidad (Stack Overflow), desde la satisfacción profesional y la búsqueda de empleo hasta la educación y las opiniones sobre el software de código abierto; y los resultados se publican en la siguiente URL: https://insights.stackoverflow.com/survey.

En este portal se encuentran publicados los resultados de los últimos 11 años. Para los fines de la práctica final de esta asignatura se usará el dataset del año 2021, cuyo link de descarga es: https://info.stackoverflowsolutions.com/rs/719-EMH-566/images/stack-overflow-developer-survey-2021.zip.

[3]:
```
so2021_df = pd.read_csv('survey_results_public.csv', header=0)
so2021_df.sample(5)
```

[3]:
```
       ResponseId                                         MainBranch  \
14490       14491  I am not primarily a developer, but I write co…
76987       76988                     I am a developer by profession
38902       38903                     I am a developer by profession
78069       78070                     I am a developer by profession
477           478                     I am a developer by profession

                                             Employment  \
14490                                Student, full-time
76987                                Employed full-time
38902  Independent contractor, freelancer, or self-em…
78069                Not employed, but looking for work
477                                  Employed full-time

                                                Country US_State UK_Country  \
14490                                             Chile      NaN        NaN
76987                          United States of America  Indiana        NaN
38902                          United States of America    Texas        NaN
78069                                        Bangladesh      NaN        NaN
477    United Kingdom of Great Britain and Northern I…      NaN    England

                                                EdLevel     Age1stCode  \
14490    Master’s degree (M.A., M.S., M.Eng., MBA, etc.)  11 - 17 years
76987  Secondary school (e.g. American high school, G…  11 - 17 years
38902       Bachelor’s degree (B.A., B.S., B.Eng., etc.)  11 - 17 years
78069                         Primary/elementary school  11 - 17 years
477      Master’s degree (M.A., M.S., M.Eng., MBA, etc.)  18 - 24 years

                                              LearnCode YearsCode  \
14490                            Other (please specify):         7
76987  Other online resources (ex: videos, blogs, etc…        18
38902                                             School        13
78069  Coding Bootcamp;Other online resources (ex: vi…         5
477                              Other (please specify):         7

      YearsCodePro                                            DevType  \
14490          NaN                                                NaN
76987            7                              Developer, full-stack
38902            8                              Developer, full-stack
78069          NaN  Developer, mobile;Developer, front-end;Develop…
477              6  Developer, back-end;DevOps specialist;System a…

                                                 OrgSize  \
14490                                                NaN
76987                           10,000 or more employees
38902  Just me - I am a freelancer, sole proprietor, …
78069                                                NaN
477                                   20 to 99 employees

                        Currency  CompTotal CompFreq  \
14490                        NaN        NaN      NaN
76987  USD\tUnited States dollar   200000.0   Yearly
38902  USD\tUnited States dollar    50000.0   Yearly
78069                        NaN        NaN      NaN
477          GBP\tPound sterling    41950.0   Yearly

                                  LanguageHaveWorkedWith  \
14490                              Java;Node.js;Python;R
76987                             Clojure;JavaScript;SQL
38902  HTML/CSS;Java;JavaScript;Node.js;PHP;Python;SQ…
78069         HTML/CSS;JavaScript;Node.js;PHP;Python;SQL
477                           Bash/Shell;Java;Python;SQL

                                  LanguageWantToWorkWith  \
14490                                             Python
76987                            Clojure;JavaScript;Rust
38902     HTML/CSS;JavaScript;Node.js;PHP;SQL;TypeScript
78069  C;C#;C++;HTML/CSS;Java;JavaScript;Kotlin;Node…
477                            Bash/Shell;C#;Java;Python

              DatabaseHaveWorkedWith  \
14490                            NaN
76987                  Elasticsearch
38902                 Firebase;MySQL
78069  Firebase;MongoDB;MySQL;Oracle
477                            MySQL
```

```
                                  DatabaseWantToWorkWith  \
14490                                                NaN
76987                                                NaN
38902                                     Firebase;MySQL
78069  Firebase;MariaDB;Microsoft SQL Server;MongoDB;…
477                                 Microsoft SQL Server

          PlatformHaveWorkedWith  \
14490                        NaN
76987                        AWS
38902  AWS;Google Cloud Platform
78069                     Heroku
477                          AWS

                                  PlatformWantToWorkWith  \
14490                                                NaN
76987                                                NaN
38902                                                AWS
78069  AWS;DigitalOcean;Google Cloud Platform;Heroku;…
477                                                  AWS

                          WebframeHaveWorkedWith  \
14490                                        NaN
76987                                   React.js
38902  Angular;Angular.js;Express;jQuery;Laravel
78069     Express;jQuery;Laravel;React.js;Vue.js
477                                          NaN

                                  WebframeWantToWorkWith  \
14490                                                NaN
76987                                           React.js
38902          Angular;Angular.js;Express;jQuery;Laravel
78069  Angular.js;Django;Express;jQuery;Laravel;React…
477                                                  NaN

              MiscTechHaveWorkedWith MiscTechWantToWorkWith  \
14490  Keras;NumPy;Pandas;TensorFlow           NumPy;Pandas
76987                            NaN                    NaN
38902                        Cordova                Cordova
78069                            NaN                    NaN
477                              NaN                    NaN

      ToolsTechHaveWorkedWith ToolsTechWantToWorkWith  \
14490                     Git                     Git
76987                     NaN                     NaN
38902                     Git                     Git
```

```
78069                     Git                     Git
477                Docker;Git              Docker;Git

                            NEWCollabToolsHaveWorkedWith  \
14490    IPython/Jupyter;Sublime Text;Visual Studio Code   
76987                                 Emacs;Sublime Text   
38902  Android Studio;Notepad++;Sublime Text;TextMate…   
78069  Android Studio;Notepad++;Sublime Text;Visual S…   
477                 Eclipse;Notepad++;Visual Studio Code   

                            NEWCollabToolsWantToWorkWith        OpSys  \
14490                       IPython/Jupyter;Sublime Text  Linux-based   
76987                                 Emacs;Sublime Text        MacOS   
38902            Android Studio;Visual Studio Code;Xcode        MacOS   
78069  Android Studio;Notepad++;PyCharm;Visual Studio…      Windows   
477                 Eclipse;Notepad++;Visual Studio Code      Windows   

                                                NEWStuck  \
14490  Visit Stack Overflow;Go for a walk or other ph…   
76987  Go for a walk or other physical activity;Googl…   
38902                                          Google it   
78069  Call a coworker or friend;Visit Stack Overflow…   
477                       Visit Stack Overflow;Google it   

                                              NEWSOSites  \
14490                      Stack Overflow;Stack Exchange   
76987                      Stack Overflow;Stack Exchange   
38902                      Stack Overflow;Stack Exchange   
78069  Stack Overflow;Stack Overflow for Teams (priva…   
477                                       Stack Overflow   

                               SOVisitFreq SOAccount  \
14490                Daily or almost daily       Yes   
76987  Less than once per month or monthly        No   
38902                Daily or almost daily       Yes   
78069               Multiple times per day       Yes   
477                 Multiple times per day       Yes   

                                              SOPartFreq           SOComm  \
14490                               A few times per week  Yes, definitely   
76987                                                NaN   No, not at all   
38902  I have never participated in Q&A on Stack Over…   No, not really   
78069                             Multiple times per day  Yes, definitely   
477                  Less than once per month or monthly   No, not really   

      NEWOtherComms                 Age Gender Trans                Sexuality  \
14490            No     25-34 years old    Man    No  Straight / Heterosexual   
```

```
76987           Yes     25-34 years old    Man    No  Straight / Heterosexual
38902            No     25-34 years old    Man    No  Straight / Heterosexual
78069            No  Under 18 years old    Man   Yes  Straight / Heterosexual
477              No     25-34 years old    Man    No        Prefer not to say

                          Ethnicity      Accessibility       MentalHealth  \
14490        Hispanic or Latino/a/x  None of the above  None of the above
76987  White or of European descent  None of the above  None of the above
38902                   South Asian  None of the above  None of the above
78069             Prefer not to say  Prefer not to say  Prefer not to say
477    White or of European descent  None of the above  None of the above

                SurveyLength                  SurveyEase  ConvertedCompYearly
14490  Appropriate in length                        Easy                  NaN
76987               Too long                        Easy             200000.0
38902  Appropriate in length                        Easy              50000.0
78069  Appropriate in length  Neither easy nor difficult                  NaN
477    Appropriate in length                        Easy              54224.0
```

Selección de variables: se realiza la selección de todas las variables del dataset que servirán para responder a todas las cuestiones planteadas en la primera parte de la práctica:

```
[8]: so2021_data = so2021_df[['MainBranch', 'Employment', 'Country', 'EdLevel',␣
      ↪'Age1stCode', 'YearsCode', 'YearsCodePro', 'DevType', 'CompTotal',␣
      ↪'LanguageHaveWorkedWith', 'DatabaseHaveWorkedWith',␣
      ↪'PlatformHaveWorkedWith', 'WebframeHaveWorkedWith',␣
      ↪'MiscTechHaveWorkedWith', 'ToolsTechHaveWorkedWith',␣
      ↪'NEWCollabToolsHaveWorkedWith', 'OpSys', 'Age', 'Gender', 'Trans',␣
      ↪'Ethnicity', 'MentalHealth', 'ConvertedCompYearly']]
     so2021_data.head(5)
```

```
[8]:                                           MainBranch  \
     0                      I am a developer by profession
     1              I am a student who is learning to code
     2  I am not primarily a developer, but I write co…
     3                      I am a developer by profession
     4                      I am a developer by profession

                                              Employment  \
     0  Independent contractor, freelancer, or self-em…
     1                                Student, full-time
     2                                Student, full-time
     3                                Employed full-time
     4  Independent contractor, freelancer, or self-em…

                                                 Country  \
     0                                          Slovakia
     1                                       Netherlands
```

```
2                                 Russian Federation
3                                            Austria
4  United Kingdom of Great Britain and Northern I…

                                             EdLevel     Age1stCode YearsCode  \
0  Secondary school (e.g. American high school, G…  18 - 24 years       NaN
1       Bachelor’s degree (B.A., B.S., B.Eng., etc.)  11 - 17 years         7
2       Bachelor’s degree (B.A., B.S., B.Eng., etc.)  11 - 17 years       NaN
3    Master’s degree (M.A., M.S., M.Eng., MBA, etc.)  11 - 17 years       NaN
4    Master’s degree (M.A., M.S., M.Eng., MBA, etc.)   5 - 10 years        17

  YearsCodePro                                            DevType  CompTotal  \
0          NaN                                  Developer, mobile     4800.0
1          NaN                                                NaN        NaN
2          NaN                                                NaN        NaN
3          NaN                               Developer, front-end        NaN
4           10  Developer, desktop or enterprise applications;…        NaN

                          LanguageHaveWorkedWith  \
0  C++;HTML/CSS;JavaScript;Objective-C;PHP;Swift
1                              JavaScript;Python
2                       Assembly;C;Python;R;Rust
3                          JavaScript;TypeScript
4                 Bash/Shell;HTML/CSS;Python;SQL

           DatabaseHaveWorkedWith PlatformHaveWorkedWith  \
0               PostgreSQL;SQLite                    NaN
1                      PostgreSQL                    NaN
2                          SQLite                 Heroku
3                             NaN                    NaN
4  Elasticsearch;PostgreSQL;Redis                    NaN

  WebframeHaveWorkedWith                 MiscTechHaveWorkedWith  \
0        Laravel;Symfony                                    NaN
1   Angular;Flask;Vue.js                                Cordova
2                  Flask  NumPy;Pandas;TensorFlow;Torch/PyTorch
3         Angular;jQuery                                    NaN
4                  Flask       Apache Spark;Hadoop;NumPy;Pandas

      ToolsTechHaveWorkedWith  \
0                         NaN
1             Docker;Git;Yarn
2                         NaN
3                         NaN
4  Docker;Git;Kubernetes;Yarn

                        NEWCollabToolsHaveWorkedWith        OpSys  \
0                                    PHPStorm;Xcode        MacOS
1           Android Studio;IntelliJ;Notepad++;PyCharm      Windows
2  IPython/Jupyter;PyCharm;RStudio;Sublime Text;V…        MacOS
3                                                 NaN      Windows
4          Atom;IPython/Jupyter;Notepad++;PyCharm;Vim  Linux-based

               Age Gender Trans                     Ethnicity  \
0  25-34 years old    Man    No  White or of European descent
1  18-24 years old    Man    No  White or of European descent
2  18-24 years old    Man    No             Prefer not to say
3  35-44 years old    Man    No  White or of European descent
4  25-34 years old    Man    No  White or of European descent

        MentalHealth  ConvertedCompYearly
0  None of the above              62268.0
1  None of the above                  NaN
2  None of the above                  NaN
3                NaN                  NaN
4                NaN                  NaN
```

[11]:
```
so2021_data.shape
```

[11]:
```
(83439, 23)
```

[12]:
```
so2021_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 83439 entries, 0 to 83438
Data columns (total 23 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   MainBranch                   83439 non-null  object
 1   Employment                   83323 non-null  object
 2   Country                      83439 non-null  object
 3   EdLevel                      83126 non-null  object
 4   Age1stCode                   83243 non-null  object
 5   YearsCode                    81641 non-null  object
 6   YearsCodePro                 61216 non-null  object
 7   DevType                      66484 non-null  object
 8   CompTotal                    47183 non-null  float64
 9   LanguageHaveWorkedWith       82357 non-null  object
 10  DatabaseHaveWorkedWith       69546 non-null  object
 11  PlatformHaveWorkedWith       52135 non-null  object
 12  WebframeHaveWorkedWith       61707 non-null  object
 13  MiscTechHaveWorkedWith       47055 non-null  object
 14  ToolsTechHaveWorkedWith      72537 non-null  object
 15  NEWCollabToolsHaveWorkedWith  81234 non-null  object
 16  OpSys                        83294 non-null  object
 17  Age                         82407 non-null  object
 18  Gender                      82286 non-null  object
 19  Trans                       80678 non-null  object
 20  Ethnicity                   79464 non-null  object
 21  MentalHealth                76920 non-null  object
 22  ConvertedCompYearly         46844 non-null  float64
dtypes: float64(2), object(21)
memory usage: 14.6+ MB
```

```
[13]: so2021_data.isnull().values.any() #valores perdidos en dataset
```

```
[13]: True
```

```
[14]: so2021_data.isnull().any() # valores perdidos por columnas en el dataset
```

```
[14]: MainBranch                      False
      Employment                       True
      Country                         False
      EdLevel                          True
      Age1stCode                       True
      YearsCode                        True
      YearsCodePro                     True
      DevType                          True
      CompTotal                        True
      LanguageHaveWorkedWith           True
      DatabaseHaveWorkedWith           True
      PlatformHaveWorkedWith           True
      WebframeHaveWorkedWith           True
      MiscTechHaveWorkedWith           True
      ToolsTechHaveWorkedWith          True
      NEWCollabToolsHaveWorkedWith     True
      OpSys                            True
      Age                              True
      Gender                           True
      Trans                            True
      Ethnicity                        True
      MentalHealth                     True
      ConvertedCompYearly              True
      dtype: bool
```

```
[157]: data = so2021_data.dropna()
```

```
[158]: data.isnull().any() # valores perdidos por columnas en el dataset
```

```
[158]: MainBranch                      False
       Employment                      False
       Country                         False
       EdLevel                         False
```

```
Age1stCode                      False
YearsCode                       False
YearsCodePro                    False
DevType                         False
CompTotal                       False
LanguageHaveWorkedWith          False
DatabaseHaveWorkedWith          False
PlatformHaveWorkedWith          False
WebframeHaveWorkedWith          False
MiscTechHaveWorkedWith          False
ToolsTechHaveWorkedWith         False
NEWCollabToolsHaveWorkedWith    False
OpSys                           False
Age                             False
Gender                          False
Trans                           False
Ethnicity                       False
MentalHealth                    False
ConvertedCompYearly             False
dtype: bool
```

[159]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 15173 entries, 45 to 83437
Data columns (total 23 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   MainBranch                    15173 non-null  object
 1   Employment                    15173 non-null  object
 2   Country                       15173 non-null  object
 3   EdLevel                       15173 non-null  object
 4   Age1stCode                    15173 non-null  object
 5   YearsCode                     15173 non-null  object
 6   YearsCodePro                  15173 non-null  object
 7   DevType                       15173 non-null  object
 8   CompTotal                     15173 non-null  float64
 9   LanguageHaveWorkedWith        15173 non-null  object
 10  DatabaseHaveWorkedWith        15173 non-null  object
 11  PlatformHaveWorkedWith        15173 non-null  object
 12  WebframeHaveWorkedWith        15173 non-null  object
 13  MiscTechHaveWorkedWith        15173 non-null  object
 14  ToolsTechHaveWorkedWith       15173 non-null  object
 15  NEWCollabToolsHaveWorkedWith  15173 non-null  object
 16  OpSys                         15173 non-null  object
 17  Age                           15173 non-null  object
 18  Gender                        15173 non-null  object
 19  Trans                         15173 non-null  object
 20  Ethnicity                     15173 non-null  object
 21  MentalHealth                  15173 non-null  object
 22  ConvertedCompYearly           15173 non-null  float64
dtypes: float64(2), object(21)
memory usage: 2.8+ MB
```

[160]:

```
data.head()
```

[160]:

```
                                           MainBranch  \
45                     I am a developer by profession
50                     I am a developer by profession
58  I am not primarily a developer, but I write co…
64  I am not primarily a developer, but I write co…
76  I am not primarily a developer, but I write co…

                                           Employment  \
45                                 Employed full-time
50                                 Employed full-time
58                                 Employed full-time
64  Independent contractor, freelancer, or self-em…
76                                 Employed full-time

                     Country                                          EdLevel  \
45                    Brazil     Bachelor’s degree (B.A., B.S., B.Eng., etc.)
50                    Greece     Bachelor’s degree (B.A., B.S., B.Eng., etc.)
58        Russian Federation               Professional degree (JD, MD, etc.)
64  United States of America  Master’s degree (M.A., M.S., M.Eng., MBA, etc.)
76                    Poland     Bachelor’s degree (B.A., B.S., B.Eng., etc.)

       Age1stCode YearsCode YearsCodePro  \
45  11 - 17 years        22           15
50  18 - 24 years        12            6
58  11 - 17 years         5            3
64  11 - 17 years         6            5
76  11 - 17 years        12            8

                                              DevType  CompTotal  \
45  Developer, desktop or enterprise applications;…    22000.0
50                              Developer, full-stack     2000.0
58  Developer, full-stack;Data scientist or machin…   120000.0
64  Developer, front-end;Developer, desktop or ent…   500000.0
76  Developer, front-end;Developer, full-stack;Dev…    15000.0

                               LanguageHaveWorkedWith  \
45        C#;C++;JavaScript;PowerShell;SQL;TypeScript
50  C#;HTML/CSS;JavaScript;Node.js;PowerShell;Type…
58          Bash/Shell;HTML/CSS;JavaScript;Python;SQL
```

```
64                         HTML/CSS;JavaScript;Python
76  Bash/Shell;C#;Dart;Delphi;Go;HTML/CSS;Java;Jav…

                               DatabaseHaveWorkedWith  \
45              Microsoft SQL Server;PostgreSQL;Redis
50  Couchbase;MariaDB;Microsoft SQL Server;MongoDB…
58                                             Oracle
64                                              MySQL
76  Firebase;Microsoft SQL Server;MongoDB;MySQL;Po…

                   PlatformHaveWorkedWith  \
45                 Heroku;Microsoft Azure
50       AWS;DigitalOcean;Microsoft Azure
58                                 Heroku
64                                    AWS
76  Google Cloud Platform;Microsoft Azure

                               WebframeHaveWorkedWith  \
45                               ASP.NET Core ;React.js
50       Angular;ASP.NET;ASP.NET Core ;Express;Svelte
58                                Django;FastAPI;Vue.js
64                                                Flask
76  Angular;Angular.js;ASP.NET;ASP.NET Core ;Djang…

                               MiscTechHaveWorkedWith ToolsTechHaveWorkedWith  \
45                                   .NET Core / .NET 5   Docker;Git;Kubernetes
50                    .NET Framework;.NET Core / .NET 5       Docker;Kubernetes
58                           NumPy;Pandas;Torch/PyTorch              Docker;Git
64                                               Pandas                     Git
76  .NET Framework;.NET Core / .NET 5;Apache Spark…     Docker;Git;Unity 3D

                         NEWCollabToolsHaveWorkedWith        OpSys  \
45         Notepad++;Visual Studio;Visual Studio Code      Windows
50         Notepad++;Visual Studio;Visual Studio Code      Windows
58                 IPython/Jupyter;Visual Studio Code  Linux-based
64                     Notepad++;PyCharm;Sublime Text  Linux-based
76  Android Studio;Eclipse;NetBeans;Notepad++;Visu…  Linux-based

                Age Gender Trans                     Ethnicity  \
45  35-44 years old    Man    No  White or of European descent
50  25-34 years old    Man    No  White or of European descent
58  25-34 years old    Man    No  White or of European descent
64  35-44 years old    Man    No  White or of European descent
76  25-34 years old    Man    No  White or of European descent

                                         MentalHealth  ConvertedCompYearly
45  I have a mood or emotional disorder (e.g. depr…              60480.0
50                                  None of the above              25944.0
58                                  None of the above              22644.0
64                                  None of the above             500000.0
76                                  None of the above              45564.0
```

[174]:

```
data.to_csv('data.csv', index=False)
```

[293]:

```
data_test = data.copy()
```

[295]:

```
data_test.to_csv('data_test.csv', index=False)
```

Variable Ethnicity:.

[296]:

```
from re import search

def choose_ethnia(cell_ethnia):
    val_ethnia_exceptions = ["I don't know", "Or, in your own words:"]

    if cell_ethnia == "I don't know;Or, in your own words:":
        return val_ethnia_exceptions[0]

    if search(";", cell_ethnia):
        row_ethnia_values = cell_ethnia.split(';', 5)
        first_val = row_ethnia_values[0]

        if first_val not in val_ethnia_exceptions:
            return first_val

        if len(row_ethnia_values) > 1:
            if row_ethnia_values[1] not in val_ethnia_exceptions:
                return row_ethnia_values[1]

        if len(row_ethnia_values) > 2:
            if row_ethnia_values[2] not in val_ethnia_exceptions:
                return row_ethnia_values[2]
    else:
        return cell_ethnia
```

[297]:

```
data_test['Ethnicity'] = data_test['Ethnicity'].apply(choose_ethnia)
```

[299]:

```
data_test.drop(index=data_test[data_test['Ethnicity'] == 'Or, in your own words:
↪'].index, inplace=True)
```

[300]:

```
data_test.drop(index=data_test[data_test['Ethnicity'] == 'Prefer not to say'].
↪index, inplace=True)
```

[301]:

```
data_test['Ethnicity'].drop_duplicates().sort_values()
```

[301]:
```
7670                                             Biracial
109                           Black or of African descent
122                                            East Asian
201                               Hispanic or Latino/a/x
465                                          I don't know
4719    Indigenous (such as Native American, Pacific I…
188                                        Middle Eastern
318                                           Multiracial
243                                           South Asian
186                                       Southeast Asian
45                          White or of European descent
Name: Ethnicity, dtype: object
```

[302]:
```
data_test['Ethnicity'] = data_test['Ethnicity'].replace(['Black or of African␣
 ↪descent'], 'Negro')
data_test['Ethnicity'] = data_test['Ethnicity'].replace(['East Asian'],␣
 ↪'Asiatico del este')
data_test['Ethnicity'] = data_test['Ethnicity'].replace(['Hispanic or Latino/a/
 ↪x'], 'Latino')
data_test['Ethnicity'] = data_test['Ethnicity'].replace(["I don't know"], 'No␣
 ↪Definido')
data_test['Ethnicity'] = data_test['Ethnicity'].replace(['Indigenous (such as␣
 ↪Native American, Pacific Islander, or Indigenous Australian)'], 'Indigena')
data_test['Ethnicity'] = data_test['Ethnicity'].replace(['Middle Eastern'],␣
 ↪'Medio Oriente')
data_test['Ethnicity'] = data_test['Ethnicity'].replace(['South Asian'],␣
 ↪'Asiatico del Sur')
data_test['Ethnicity'] = data_test['Ethnicity'].replace(['Southeast Asian'],␣
 ↪'Asiatico del Sudeste')
data_test['Ethnicity'] = data_test['Ethnicity'].replace(['White or of European␣
 ↪descent'], 'Blanco o Europeo')
```

[303]:
```
data_test['Ethnicity'].drop_duplicates().sort_values()
```

[303]:
```
186     Asiatico del Sudeste
243         Asiatico del Sur
122        Asiatico del este
7670                Biracial
45          Blanco o Europeo
4719                Indigena
201                   Latino
188            Medio Oriente
318              Multiracial
109                    Negro
465              No Definido
Name: Ethnicity, dtype: object
```

```
[304]: data_test.to_csv('data_test.csv', index=False)
```

Variable Employment:.

```
[305]: data_test['Employment'].drop_duplicates().sort_values()
```

```
[305]: 45                                     Employed full-time
       83                                     Employed part-time
       64    Independent contractor, freelancer, or self-em…
       Name: Employment, dtype: object
```

```
[306]: data_test['Employment'] = data_test['Employment'].replace(['Employed␣
        ↪full-time'], 'Tiempo completo')
       data_test['Employment'] = data_test['Employment'].replace(['Employed␣
        ↪part-time'], 'Tiempo parcial')
       data_test['Employment'] = data_test['Employment'].replace(['Independent␣
        ↪contractor, freelancer, or self-employed'], 'Independiete')
```

```
[307]: data_test['Employment'].drop_duplicates().sort_values()
```

```
[307]: 64       Independiete
       45    Tiempo completo
       83     Tiempo parcial
       Name: Employment, dtype: object
```

Variable EdLevel:.

```
[308]: data_test['EdLevel'].drop_duplicates().sort_values()
```

```
[308]: 130                  Associate degree (A.A., A.S., etc.)
       45          Bachelor’s degree (B.A., B.S., B.Eng., etc.)
       64       Master’s degree (M.A., M.S., M.Eng., MBA, etc.)
       77            Other doctoral degree (Ph.D., Ed.D., etc.)
       731                            Primary/elementary school
       58                    Professional degree (JD, MD, etc.)
       380    Secondary school (e.g. American high school, G…
       110    Some college/university study without earning …
       86                                        Something else
       Name: EdLevel, dtype: object
```

```
[309]: data_test['EdLevel'] = data_test['EdLevel'].replace(['Associate degree (A.A., A.
        ↪S., etc.)'], 'Grado Asociado')
       data_test['EdLevel'] = data_test['EdLevel'].replace(['Bachelor’s degree (B.A.,␣
        ↪B.S., B.Eng., etc.)'], 'Licenciatura')
       data_test['EdLevel'] = data_test['EdLevel'].replace(['Master’s degree (M.A., M.
        ↪S., M.Eng., MBA, etc.)'], 'Master')
       data_test['EdLevel'] = data_test['EdLevel'].replace(['Other doctoral degree (Ph.
        ↪D., Ed.D., etc.)'], 'Doctorado')
```

```
data_test['EdLevel'] = data_test['EdLevel'].replace(['Primary/elementary␣
 ↪school'], 'Primaria')
data_test['EdLevel'] = data_test['EdLevel'].replace(['Professional degree (JD,␣
 ↪MD, etc.)'], 'Grado Profesional')
data_test['EdLevel'] = data_test['EdLevel'].replace(['Secondary school (e.g.␣
 ↪American high school, German Realschule or Gymnasium, etc.)'], 'Secundaria')
data_test['EdLevel'] = data_test['EdLevel'].replace(['Some college/university␣
 ↪study without earning a degree'], 'Estudios sin grado')
data_test['EdLevel'] = data_test['EdLevel'].replace(['Something else'], 'Otro')
```

[310]:

```
data_test['EdLevel'].drop_duplicates().sort_values()
```

[310]:

```
77              Doctorado
110    Estudios sin grado
130        Grado Asociado
58      Grado Profesional
45           Licenciatura
64                 Master
86                   Otro
731              Primaria
380            Secundaria
Name: EdLevel, dtype: object
```

[311]:

```
data_test.to_csv('data_test.csv', index=False)
```

Variable DevType:.

[312]:

```
data_test['DevType'].drop_duplicates().sort_values()
```

[312]:

```
6113                                    Academic researcher
1413             Academic researcher;Data or business analyst
9267         Academic researcher;Database administrator;Dev…
37989        Academic researcher;Database administrator;Sci…
8378                    Academic researcher;DevOps specialist
                               …
27480                                       Student;Educator
23376                           Student;System administrator
77415                  Student;System administrator;Educator
1317                                    System administrator
14465                   System administrator;Product manager
Name: DevType, Length: 3374, dtype: object
```

[313]:

```
from re import search

def choose_devtype(cell_devtype):
    val_devtype_exceptions = ["Other (please specify):"]

    if cell_devtype == "Other (please specify):":
        return val_devtype_exceptions[0]

    if search(";", cell_devtype):
        row_devtype_values = cell_devtype.split(';', 10)
        first_val = row_devtype_values[0]

        if first_val not in val_devtype_exceptions:
            return first_val

        if len(row_devtype_values) > 1:
            if row_devtype_values[1] not in val_devtype_exceptions:
                return row_devtype_values[1]

    else:
        return cell_devtype
```

```
[314]: data_test['DevType'] = data_test['DevType'].apply(choose_devtype)
```

```
[315]: data_test['DevType'].head()
```

```
[315]: 45    Developer, desktop or enterprise applications
       50                            Developer, full-stack
       58                            Developer, full-stack
       64                             Developer, front-end
       76                             Developer, front-end
       Name: DevType, dtype: object
```

```
[316]: data_test['DevType'].drop_duplicates().sort_values()
```

```
[316]: 1160                               Academic researcher
       4752                          Data or business analyst
       77       Data scientist or machine learning specialist
       6237                            Database administrator
       21365                                         Designer
       4288                                 DevOps specialist
       1348                             Developer, QA or test
       86                               Developer, back-end
       45       Developer, desktop or enterprise applications
       2517       Developer, embedded applications or devices
       64                              Developer, front-end
       50                             Developer, full-stack
       690                        Developer, game or graphics
       100                                 Developer, mobile
       15061                                         Educator
       114                                    Engineer, data
       3419                        Engineer, site reliability
       942                                Engineering manager
```

```
17240                   Marketing or sales professional
249                             Other (please specify):
28419                                   Product manager
5724                                          Scientist
710                Senior Executive (C-Suite, VP, etc.)
9664                                            Student
1317                               System administrator
Name: DevType, dtype: object
```

[317]:

```
data_test['DevType'].value_counts()
```

[317]:

```
Developer, full-stack                             4416
Developer, front-end                              2903
Developer, mobile                                 2798
Developer, back-end                               1484
Developer, desktop or enterprise applications     1096
Engineer, data                                     595
Data scientist or machine learning specialist      408
Other (please specify):                            137
Engineering manager                                126
DevOps specialist                                  107
Senior Executive (C-Suite, VP, etc.)                74
Academic researcher                                 62
Developer, QA or test                               59
Data or business analyst                            48
Developer, embedded applications or devices         41
System administrator                                25
Engineer, site reliability                          24
Product manager                                     23
Database administrator                              20
Student                                             16
Developer, game or graphics                         15
Scientist                                           13
Designer                                            10
Educator                                            10
Marketing or sales professional                      7
Name: DevType, dtype: int64
```

[318]:

```
data_test['DevType'] = data_test['DevType'].replace(['Developer, full-stack'],␣
 ↪'Desarrollador full-stack')
data_test['DevType'] = data_test['DevType'].replace(['Developer, front-end'],␣
 ↪'Desarrollador front-end')
data_test['DevType'] = data_test['DevType'].replace(['Developer, mobile'],␣
 ↪'Desarrollador móvil')
data_test['DevType'] = data_test['DevType'].replace(['Developer, back-end'],␣
 ↪'Desarrollador back-end')
```

```
data_test['DevType'] = data_test['DevType'].replace(['Developer, desktop or␣
↪enterprise applications'], 'Desarrollador Escritorio')
data_test['DevType'] = data_test['DevType'].replace(['Engineer, data'],␣
↪'Ingeniero de datos')
data_test['DevType'] = data_test['DevType'].replace(['Data scientist or machine␣
↪learning specialist'], 'Cientifico de datos')
data_test['DevType'] = data_test['DevType'].replace(['Other (please specify):
↪'], 'Otro')
data_test['DevType'] = data_test['DevType'].replace(['Engineering manager'],␣
↪'Manager de Ingeniería')
data_test['DevType'] = data_test['DevType'].replace(['DevOps specialist'],␣
↪'Especialista en DevOps')
data_test['DevType'] = data_test['DevType'].replace(['Senior Executive␣
↪(C-Suite, VP, etc.)'], 'Ejecutivo Senior')
data_test['DevType'] = data_test['DevType'].replace(['Academic researcher'],␣
↪'Investigador Académico')
data_test['DevType'] = data_test['DevType'].replace(['Developer, QA or test'],␣
↪'Desarrollador de QA o Test')
data_test['DevType'] = data_test['DevType'].replace(['Data or business␣
↪analyst'], 'Analista de datos o negocio')
data_test['DevType'] = data_test['DevType'].replace(['Developer, embedded␣
↪applications or devices'], 'Desarrollador de aplicaciones embebidas')
data_test['DevType'] = data_test['DevType'].replace(['System administrator'],␣
↪'Administrador de sistemas')
data_test['DevType'] = data_test['DevType'].replace(['Engineer, site␣
↪reliability'], 'Ingeniero de confiabilidad del sitio')
data_test['DevType'] = data_test['DevType'].replace(['Product manager'],␣
↪'Gerente de producto')
data_test['DevType'] = data_test['DevType'].replace(['Database administrator'],␣
↪'Administrador de base de datos')
data_test['DevType'] = data_test['DevType'].replace(['Student'], 'Estudiante')
data_test['DevType'] = data_test['DevType'].replace(['Developer, game or␣
↪graphics'], 'Desarrollador de juegos o gráfico')
data_test['DevType'] = data_test['DevType'].replace(['Scientist'], 'Científico')
data_test['DevType'] = data_test['DevType'].replace(['Designer'], 'Diseñador')
data_test['DevType'] = data_test['DevType'].replace(['Educator'], 'Educador')
data_test['DevType'] = data_test['DevType'].replace(['Marketing or sales␣
↪professional'], 'Profesional en Marketing o ventas')
```

```
[319]: data_test['DevType'].drop_duplicates().sort_values()
```

```
[319]: 6237             Administrador de base de datos
       1317                  Administrador de sistemas
       4752                Analista de datos o negocio
       77                          Cientifico de datos
       5724                                 Científico
```

```
45                      Desarrollador Escritorio
86                        Desarrollador back-end
1348                  Desarrollador de QA o Test
2517     Desarrollador de aplicaciones embebidas
690            Desarrollador de juegos o gráfico
64                       Desarrollador front-end
50                      Desarrollador full-stack
100                          Desarrollador móvil
21365                                  Diseñador
15061                                   Educador
710                             Ejecutivo Senior
4288                      Especialista en DevOps
9664                                  Estudiante
28419                        Gerente de producto
3419        Ingeniero de confiabilidad del sitio
114                           Ingeniero de datos
1160                      Investigador Académico
942                        Manager de Ingeniería
249                                         Otro
17240          Profesional en Marketing o ventas
Name: DevType, dtype: object
```

```
[320]: data_test['DevType'].value_counts()
```

```
[320]: Desarrollador full-stack                   4416
       Desarrollador front-end                    2903
       Desarrollador móvil                        2798
       Desarrollador back-end                     1484
       Desarrollador Escritorio                   1096
       Ingeniero de datos                          595
       Cientifico de datos                         408
       Otro                                        137
       Manager de Ingeniería                       126
       Especialista en DevOps                      107
       Ejecutivo Senior                             74
       Investigador Académico                       62
       Desarrollador de QA o Test                   59
       Analista de datos o negocio                  48
       Desarrollador de aplicaciones embebidas      41
       Administrador de sistemas                    25
       Ingeniero de confiabilidad del sitio         24
       Gerente de producto                          23
       Administrador de base de datos               20
       Estudiante                                   16
       Desarrollador de juegos o gráfico            15
       Científico                                   13
       Diseñador                                    10
```

```
Educador                                         10
Profesional en Marketing o ventas                 7
Name: DevType, dtype: int64
```

Variable MainBranch:

[321]:
```
data_test['MainBranch'].drop_duplicates().sort_values()
```

[321]:
```
45                         I am a developer by profession
58    I am not primarily a developer, but I write co…
Name: MainBranch, dtype: object
```

[322]:
```
data_test['MainBranch'] = data_test['MainBranch'].replace(['I am a developer by␣
 ↪profession'], 'Desarrollador Profesional')
data_test['MainBranch'] = data_test['MainBranch'].replace(['I am not primarily␣
 ↪a developer, but I write code sometimes as part of my work'], 'Desarrollador␣
 ↪ocasional')
```

[323]:
```
data_test['MainBranch'].drop_duplicates().sort_values()
```

[323]:
```
45    Desarrollador Profesional
58      Desarrollador ocasional
Name: MainBranch, dtype: object
```

[324]:
```
data_test.to_csv('data_test.csv', index=False)
```

Variable Age1stCode:

[325]:
```
data_test['Age1stCode'].drop_duplicates().sort_values()
```

[325]:
```
45                11 - 17 years
50                18 - 24 years
222               25 - 34 years
751               35 - 44 years
2371              45 - 54 years
77                 5 - 10 years
2225              55 - 64 years
37610      Older than 64 years
188       Younger than 5 years
Name: Age1stCode, dtype: object
```

[326]:
```
data_test['Age1stCode'].value_counts()
```

[326]:
```
11 - 17 years           8018
18 - 24 years           3408
5 - 10 years            2018
25 - 34 years            639
35 - 44 years            219
Younger than 5 years     105
```

```
45 - 54 years             85
55 - 64 years             24
Older than 64 years        1
Name: Age1stCode, dtype: int64
```

[327]:

```
data_test['Age1stCode'] = data_test['Age1stCode'].replace(['11 - 17 years'],␣
 ↪'11-17')
data_test['Age1stCode'] = data_test['Age1stCode'].replace(['18 - 24 years'],␣
 ↪'18-24')
data_test['Age1stCode'] = data_test['Age1stCode'].replace(['25 - 34 years'],␣
 ↪'25-34')
data_test['Age1stCode'] = data_test['Age1stCode'].replace(['35 - 44 years'],␣
 ↪'35-44')
data_test['Age1stCode'] = data_test['Age1stCode'].replace(['45 - 54 years'],␣
 ↪'45-54')
data_test['Age1stCode'] = data_test['Age1stCode'].replace(['5 - 10 years'],␣
 ↪'5-10')
data_test['Age1stCode'] = data_test['Age1stCode'].replace(['55 - 64 years'],␣
 ↪'55-64')
data_test['Age1stCode'] = data_test['Age1stCode'].replace(['Older than 64␣
 ↪years'], '> 64')
data_test['Age1stCode'] = data_test['Age1stCode'].replace(['Younger than 5␣
 ↪years'], '< 5')
```

[328]:

```
data_test['Age1stCode'].value_counts()
```

[328]:

```
11-17    8018
18-24    3408
5-10     2018
25-34     639
35-44     219
< 5       105
45-54      85
55-64      24
> 64        1
Name: Age1stCode, dtype: int64
```

Variable YearsCode:

[333]:

```
data_test['YearsCode'] = data_test['YearsCode'].replace(['More than 50 years'],␣
 ↪50)
data_test['YearsCode'] = data_test['YearsCode'].replace(['Less than 1 year'], 1)
```

Variable YearsCodePro:

[334]:

```
data_test['YearsCodePro'] = data_test['YearsCodePro'].replace(['More than 50␣
 ↪years'], 50)
```

```
data_test['YearsCodePro'] = data_test['YearsCodePro'].replace(['Less than 1␣
↪year'], 1)
```

Variable OpSys:

[335]:
```
data_test['OpSys'].value_counts()
```

[335]:
```
Windows                               6770
MacOS                                 4255
Linux-based                           2912
Windows Subsystem for Linux (WSL)      523
Other (please specify):                 47
BSD                                     10
Name: OpSys, dtype: int64
```

[336]:
```
data_test['OpSys'] = data_test['OpSys'].replace(['Windows Subsystem for Linux␣
↪(WSL)'], 'Windows')
data_test['OpSys'] = data_test['OpSys'].replace(['Linux-based'], 'Linux')
data_test['OpSys'] = data_test['OpSys'].replace(['Other (please specify)'],␣
↪'Otro')
```

[337]:
```
data_test['OpSys'].value_counts()
```

[337]:
```
Windows                    7293
MacOS                      4255
Linux                      2912
Other (please specify):      47
BSD                          10
Name: OpSys, dtype: int64
```

Variable Age:

[338]:
```
data_test['Age'].value_counts()
```

[338]:
```
25-34 years old       7275
35-44 years old       3361
18-24 years old       2602
45-54 years old        957
55-64 years old        255
Under 18 years old      35
65 years or older       26
Prefer not to say        6
Name: Age, dtype: int64
```

[339]:
```
data_test['Age'] = data_test['Age'].replace(['25-34 years old'], '25-34')
data_test['Age'] = data_test['Age'].replace(['35-44 years old'], '35-44')
data_test['Age'] = data_test['Age'].replace(['18-24 years old'], '18-24')
data_test['Age'] = data_test['Age'].replace(['45-54 years old'], '45-54')
```

```
data_test['Age'] = data_test['Age'].replace(['55-64 years old'], '55-64')
data_test['Age'] = data_test['Age'].replace(['Under 18 years old'], '< 18')
data_test['Age'] = data_test['Age'].replace(['65 years or older'], '>= 65')
data_test['Age'] = data_test['Age'].replace(['Prefer not to say'], 'No␣
 ↪definido')
```

[362]:

```
data_test['Age'] = data_test['Age'].replace(['25-34 years old'], '25-34')
```

[363]:

```
data_test['Age'].value_counts()
```

[363]:

```
25-34          7275
35-44          3361
18-24          2602
45-54           957
55-64           255
< 18             35
>= 65            26
No definido       6
Name: Age, dtype: int64
```

Variable Gender:

[341]:

```
data_test['Gender'].value_counts()
```

[341]:

```
Man
13748
Woman
502
Non-binary, genderqueer, or gender non-conforming
94
Prefer not to say
53
Man;Non-binary, genderqueer, or gender non-conforming
37
Man;Or, in your own words:
27
Or, in your own words:
25
Woman;Non-binary, genderqueer, or gender non-conforming
19
Man;Woman
5
Man;Woman;Non-binary, genderqueer, or gender non-conforming;Or, in your own
words:        3
Non-binary, genderqueer, or gender non-conforming;Or, in your own words:
2
Man;Woman;Non-binary, genderqueer, or gender non-conforming
2
```

```
Name: Gender, dtype: int64
```

```
[342]: data_test['Gender'] = data_test['Gender'].replace(['Man'], 'Hombre')
       data_test['Gender'] = data_test['Gender'].replace(['Woman'], 'Mujer')
       data_test['Gender'] = data_test['Gender'].replace(['Non-binary, genderqueer, or␣
        ↪gender non-conforming'], 'No binario u otro')
       data_test['Gender'] = data_test['Gender'].replace(['Man;Non-binary,␣
        ↪genderqueer, or gender non-conforming'], 'No binario u otro')
       data_test['Gender'] = data_test['Gender'].replace(['Man;Or, in your own words:
        ↪'], 'Hombre')
       data_test['Gender'] = data_test['Gender'].replace(['Or, in your own words:'],␣
        ↪'No definido')
       data_test['Gender'] = data_test['Gender'].replace(['Woman;Non-binary,␣
        ↪genderqueer, or gender non-conforming'], 'No binario u otro')
       data_test['Gender'] = data_test['Gender'].replace(['Man;Woman'], 'No definido')
       data_test['Gender'] = data_test['Gender'].replace(['Man;Woman;Non-binary,␣
        ↪genderqueer, or gender non-conforming;Or, in your own words:'], 'No binario␣
        ↪u otro')
       data_test['Gender'] = data_test['Gender'].replace(['Non-binary, genderqueer, or␣
        ↪gender non-conforming;Or, in your own words:'], 'No binario u otro')
       data_test['Gender'] = data_test['Gender'].replace(['Man;Woman;Non-binary,␣
        ↪genderqueer, or gender non-conforming'], 'No binario u otro')
```

```
[344]: data_test['Gender'] = data_test['Gender'].replace(['Prefer not to say'], 'No␣
        ↪definido')
```

```
[350]: data_test['Gender'].value_counts()
```

```
[350]: Hombre               13775
       Mujer                  502
       No binario u otro      157
       No definido             83
       Name: Gender, dtype: int64
```

Variable Trans:

```
[349]: data_test['Trans'].value_counts()
```

```
[349]: No                          14262
       Yes                           110
       Prefer not to say              88
       Or, in your own words:         57
       Name: Trans, dtype: int64
```

```
[351]: data_test['Trans'] = data_test['Trans'].replace(['Yes'], 'Si')
       data_test['Trans'] = data_test['Trans'].replace(['Prefer not to say'], 'No␣
        ↪definido')
```

```
data_test['Trans'] = data_test['Trans'].replace(['Or, in your own words:'], 'No␣
↪definido')
```

[352]:
```
data_test['Trans'].value_counts()
```

[352]:
```
No             14262
No definido      145
Si               110
Name: Trans, dtype: int64
```

Variable MentalHealth:

[353]:
```
data_test['MentalHealth'].value_counts()
```

[353]:
```
None of the above
10924
I have a concentration and/or memory disorder (e.g. ADHD)
627
I have an anxiety disorder
605
I have a mood or emotional disorder (e.g. depression, bipolar disorder)
442
Prefer not to say
396
I have a mood or emotional disorder (e.g. depression, bipolar disorder);I have
an anxiety disorder
369
I have autism / an autism spectrum disorder (e.g. Asperger's)
206
I have a concentration and/or memory disorder (e.g. ADHD);I have a mood or
emotional disorder (e.g. depression, bipolar disorder);I have an anxiety
disorder
191
Or, in your own words:
142
I have a concentration and/or memory disorder (e.g. ADHD);I have a mood or
emotional disorder (e.g. depression, bipolar disorder)
137
I have a concentration and/or memory disorder (e.g. ADHD);I have an anxiety
disorder
131
I have a concentration and/or memory disorder (e.g. ADHD);I have autism / an
autism spectrum disorder (e.g. Asperger's)
72
I have a concentration and/or memory disorder (e.g. ADHD);I have a mood or
emotional disorder (e.g. depression, bipolar disorder);I have an anxiety
disorder;I have autism / an autism spectrum disorder (e.g. Asperger's)
63
```

```
I have a mood or emotional disorder (e.g. depression, bipolar disorder);I have
autism / an autism spectrum disorder (e.g. Asperger's)
48
I have a mood or emotional disorder (e.g. depression, bipolar disorder);I have
an anxiety disorder;I have autism / an autism spectrum disorder (e.g.
Asperger's)
39
I have an anxiety disorder;I have autism / an autism spectrum disorder (e.g.
Asperger's)
27
I have a concentration and/or memory disorder (e.g. ADHD);I have a mood or
emotional disorder (e.g. depression, bipolar disorder);I have autism / an autism
spectrum disorder (e.g. Asperger's)
24
I have a concentration and/or memory disorder (e.g. ADHD);I have an anxiety
disorder;I have autism / an autism spectrum disorder (e.g. Asperger's)
22
I have a concentration and/or memory disorder (e.g. ADHD);Or, in your own words:
18
I have a concentration and/or memory disorder (e.g. ADHD);I have a mood or
emotional disorder (e.g. depression, bipolar disorder);Or, in your own words:
7
I have a mood or emotional disorder (e.g. depression, bipolar disorder);Or, in
your own words:
6
I have an anxiety disorder;Or, in your own words:
5
I have a mood or emotional disorder (e.g. depression, bipolar disorder);I have
an anxiety disorder;Or, in your own words:
4
I have a concentration and/or memory disorder (e.g. ADHD);I have a mood or
emotional disorder (e.g. depression, bipolar disorder);I have an anxiety
disorder;Or, in your own words:
4
I have a concentration and/or memory disorder (e.g. ADHD);I have an anxiety
disorder;Or, in your own words:
3
I have a concentration and/or memory disorder (e.g. ADHD);I have autism / an
autism spectrum disorder (e.g. Asperger's);Or, in your own words:
2
I have a concentration and/or memory disorder (e.g. ADHD);I have a mood or
emotional disorder (e.g. depression, bipolar disorder);I have an anxiety
disorder;I have autism / an autism spectrum disorder (e.g. Asperger's);Or, in
your own words:        1
I have a concentration and/or memory disorder (e.g. ADHD);I have a mood or
emotional disorder (e.g. depression, bipolar disorder);I have autism / an autism
spectrum disorder (e.g. Asperger's);Or, in your own words:
```

```
1
I have autism / an autism spectrum disorder (e.g. Asperger's);Or, in your own
words:
1
Name: MentalHealth, dtype: int64
```

[356]:

```
from re import search

def choose_mental_health(cell_mental_health):
    val_mental_health_exceptions = ["Or, in your own words:"]

    if cell_mental_health == "Or, in your own words:":
        return val_mental_health_exceptions[0]

    if search(";", cell_mental_health):
        row_mental_health_values = cell_mental_health.split(';', 10)
        first_val = row_mental_health_values[0]

        return first_val
    else:
        return cell_mental_health
```

[357]:

```
data_test['MentalHealth'] = data_test['MentalHealth'].
 ↪apply(choose_mental_health)
```

[358]:

```
data_test['MentalHealth'].value_counts()
```

[358]:

```
None of the above                                                       10924
I have a concentration and/or memory disorder (e.g. ADHD)                1303
I have a mood or emotional disorder (e.g. depression, bipolar disorder)   908
I have an anxiety disorder                                                637
Prefer not to say                                                         396
I have autism / an autism spectrum disorder (e.g. Asperger's)             207
Or, in your own words:                                                    142
Name: MentalHealth, dtype: int64
```

[359]:

```
data_test['MentalHealth'] = data_test['MentalHealth'].replace(['None of the␣
 ↪above'], 'Ninguna de las mencionadas')
data_test['MentalHealth'] = data_test['MentalHealth'].replace(['I have a␣
 ↪concentration and/or memory disorder (e.g. ADHD)'], 'Desorden de␣
 ↪concentración o memoria')
data_test['MentalHealth'] = data_test['MentalHealth'].replace(['I have a mood␣
 ↪or emotional disorder (e.g. depression, bipolar disorder)'], 'Desorden␣
 ↪emocional')
data_test['MentalHealth'] = data_test['MentalHealth'].replace(['I have an␣
 ↪anxiety disorder'], 'Desorden de ansiedad')
```

```
data_test['MentalHealth'] = data_test['MentalHealth'].replace(['Prefer not to␣
↪say'], 'No definido')
data_test['MentalHealth'] = data_test['MentalHealth'].replace(["I have autism /␣
↪an autism spectrum disorder (e.g. Asperger's)"], 'Tipo de autismo')
data_test['MentalHealth'] = data_test['MentalHealth'].replace(['Or, in your own␣
↪words:'], 'No definido')
```

```
[360]: data_test['MentalHealth'].value_counts()
```

```
[360]: Ninguna de las mencionadas             10924
       Desorden de concentración o memoria     1303
       Desorden emocional                       908
       Desorden de ansiedad                     637
       No definido                              538
       Tipo de autismo                          207
       Name: MentalHealth, dtype: int64
```

3 2. Selección de campos para subdatasets

Se seleccionarán los campos adecuados para responder a cada una de las cuestiones que se plantearon en la primera parte de la práctica.

3.0.1 2.1. Según la autodeterminación de la etnia, ¿Qué etnia tiene un mayor sueldo anual?

Se seleccionarán los campos adecuados para responder a esta pregunta

```
[366]: data_etnia = data_test[['Country', 'Ethnicity', 'ConvertedCompYearly']]
       data_etnia.head()
```

```
[366]:                      Country         Ethnicity  ConvertedCompYearly
       45                     Brazil  Blanco o Europeo              60480.0
       50                     Greece  Blanco o Europeo              25944.0
       58         Russian Federation  Blanco o Europeo              22644.0
       64   United States of America  Blanco o Europeo             500000.0
       76                     Poland  Blanco o Europeo              45564.0
```

```
[509]: df_data_etnia = data_etnia.copy()
```

```
[512]: def remove_outliers(df, q=0.05):
           upper = df.quantile(1-q)
           lower = df.quantile(q)
           mask = (df < upper) & (df > lower)
           return mask

       mask = remove_outliers(df_data_etnia['ConvertedCompYearly'], 0.1)
```

```
print(df_data_etnia[mask])
```

```
                          Country             Ethnicity  ConvertedCompYearly
45                         Brazil      Blanco o Europeo              60480.0
50                         Greece      Blanco o Europeo              25944.0
58             Russian Federation      Blanco o Europeo              22644.0
76                         Poland      Blanco o Europeo              45564.0
77                         Canada      Blanco o Europeo             151263.0
…                             …                   …                   …
83425                     Finland      Blanco o Europeo              19452.0
83428                      Brazil                Latino              41232.0
83431                    Pakistan  Asiatico del Sudeste              11676.0
83432                      Canada     Asiatico del este              80169.0
83436  United States of America      Blanco o Europeo              90000.0

[11611 rows x 3 columns]
```

[513]:
```
df_data_etnia_no_outliers = df_data_etnia[mask]
```

[517]:
```
df_data_etnia_no_outliers = df_data_etnia_no_outliers.copy()
```

[519]:
```
df_data_etnia_no_outliers['ConvertedCompYearlyCategorical'] = 'ALTO'
df_data_etnia_no_outliers.loc[(df_data_etnia_no_outliers['ConvertedCompYearly']␣
 ↪>= 0) & (df_data_etnia_no_outliers['ConvertedCompYearly'] <= 32747),␣
 ↪'ConvertedCompYearlyCategorical'] = 'BAJO'
df_data_etnia_no_outliers.loc[(df_data_etnia_no_outliers['ConvertedCompYearly']␣
 ↪> 32747) & (df_data_etnia_no_outliers['ConvertedCompYearly'] <= 90000),␣
 ↪'ConvertedCompYearlyCategorical'] = 'MEDIO'

print(df_data_etnia_no_outliers)
```

```
                          Country             Ethnicity  ConvertedCompYearly  \
45                         Brazil      Blanco o Europeo              60480.0
50                         Greece      Blanco o Europeo              25944.0
58             Russian Federation      Blanco o Europeo              22644.0
76                         Poland      Blanco o Europeo              45564.0
77                         Canada      Blanco o Europeo             151263.0
…                             …                   …                   …
83425                     Finland      Blanco o Europeo              19452.0
83428                      Brazil                Latino              41232.0
83431                    Pakistan  Asiatico del Sudeste              11676.0
83432                      Canada     Asiatico del este              80169.0
83436  United States of America      Blanco o Europeo              90000.0

      ConvertedCompYearlyCategorical
45                             MEDIO
50                              BAJO
58                              BAJO
```

```
76                              MEDIO
77                               ALTO
…                               …
83425                            BAJO
83428                           MEDIO
83431                            BAJO
83432                           MEDIO
83436                           MEDIO

[11611 rows x 4 columns]
```

[520]:
```
df_data_etnia_alto =␣
 ↪df_data_etnia_no_outliers[df_data_etnia_no_outliers['ConvertedCompYearlyCategorical']␣
 ↪== 'ALTO']
```

[521]:
```
df_data_etnia_alto = df_data_etnia_alto[['Ethnicity',␣
 ↪'ConvertedCompYearlyCategorical']]
```

[523]:
```
df_flourish = df_data_etnia_alto['Ethnicity'].value_counts().to_frame('counts').
 ↪reset_index()
```

[524]:
```
df_flourish
```

[524]:
```
                   index  counts
0       Blanco o Europeo    2413
1                 Latino     119
2        Asiatico del Sur     97
3           Medio Oriente     75
4       Asiatico del este     51
5                  Negro      44
6    Asiatico del Sudeste     43
7            Multiracial      23
8            No Definido      15
9               Biracial       9
10              Indigena       3
```

[525]:
```
df_flourish.to_csv('001_df_flourish.csv', index=False)
```

[443]:
```
df_data_etnia_alto.to_csv('001_df_data_etnia_alto.csv', index=False)
```

[439]:
```
df_data_etnia.to_csv('001_data_etnia_categorical.csv', index=False)
```

[367]:
```
data_etnia.to_csv('001_data_etnia.csv', index=False)
```

3.0.2 2.2. ¿Cuáles son los porcentajes de programadores que trabajan a tiempo completo, medio tiempo o freelance?

Se seleccionarán los campos adecuados para responder a esta pregunta

[368]:

```
data_time_work_dev = data_test[['Country', 'Employment', 'ConvertedCompYearly',␣
↪'EdLevel', 'Age']]
data_time_work_dev.head()
```

[368]:

```
                     Country       Employment  ConvertedCompYearly  \
45                    Brazil  Tiempo completo              60480.0
50                    Greece  Tiempo completo              25944.0
58        Russian Federation  Tiempo completo              22644.0
64  United States of America     Independiete             500000.0
76                    Poland  Tiempo completo              45564.0

             EdLevel    Age
45      Licenciatura  35-44
50      Licenciatura  25-34
58  Grado Profesional  25-34
64            Master  35-44
76      Licenciatura  25-34
```

[448]:

```
df_flourish_002 = data_time_work_dev['Employment'].value_counts().
↪to_frame('counts').reset_index()
```

[449]:

```
df_flourish_002
```

[449]:

```
             index  counts
0  Tiempo completo   12402
1     Independiete    1678
2   Tiempo parcial     437
```

[454]:

```
df_flourish_002['counts'] = (df_flourish_002['counts'] * 100 ) /␣
↪data_time_work_dev.shape[0]
```

[455]:

```
df_flourish_002
```

[455]:

```
             index     counts
0  Tiempo completo  85.430874
1     Independiete  11.558862
2   Tiempo parcial   3.010264
```

[456]:

```
df_flourish_002['counts'] = df_flourish_002['counts'].round(2)
```

[457]:

```
df_flourish_002
```

[457]:

```
             index  counts
0  Tiempo completo   85.43
1     Independiete   11.56
2   Tiempo parcial    3.01
```

[458]:

```
df_flourish_002.to_csv('002_df_flourish.csv', index=False)
```

3.0.3 2.3. ¿Cuáles son los países con mayor número de programadores profesionales que son activos en la comunidad Stack Overflow?

Se seleccionarán los campos adecuados para responder a esta pregunta

[370]:
```
data_pro_dev_active_so = data_test[['Country', 'Employment', 'MainBranch',␣
 ↪'EdLevel', 'DevType', 'Age']]
data_pro_dev_active_so.head()
```

[370]:
```
                     Country       Employment                 MainBranch  \
45                    Brazil  Tiempo completo  Desarrollador Profesional
50                    Greece  Tiempo completo  Desarrollador Profesional
58        Russian Federation  Tiempo completo    Desarrollador ocasional
64  United States of America     Independiete    Desarrollador ocasional
76                    Poland  Tiempo completo    Desarrollador ocasional

              EdLevel                    DevType    Age
45       Licenciatura  Desarrollador Escritorio  35-44
50       Licenciatura  Desarrollador full-stack  25-34
58  Grado Profesional  Desarrollador full-stack  25-34
64             Master   Desarrollador front-end  35-44
76       Licenciatura   Desarrollador front-end  25-34
```

[464]:
```
df_flourish_003 = data_pro_dev_active_so['Country'].value_counts().
 ↪sort_values(ascending=False).head(10)
```

[477]:
```
df_flourish_003 = df_flourish_003.to_frame()
```

[482]:
```
df_flourish_003 = df_flourish_003.reset_index()
df_flourish_003.columns = ["País", "# Programadores Profesionales"]
```

[485]:
```
df_flourish_003.to_csv('003_df_flourish_003.csv', index=False)
```

3.0.4 2.4. ¿Cuál es el nivel educativo que mayores ingresos registra entre los encuestados?

Se seleccionarán los campos adecuados para responder a esta pregunta

[495]:
```
data_edlevel_income = data_test[['ConvertedCompYearly', 'EdLevel']]
data_edlevel_income.head()
```

[495]:
```
    ConvertedCompYearly            EdLevel
45              60480.0       Licenciatura
50              25944.0       Licenciatura
58              22644.0  Grado Profesional
64             500000.0             Master
76              45564.0       Licenciatura
```

[501]:
```
df_data_edlevel_income = data_edlevel_income.copy()
```

[502]:
```
def remove_outliers(df, q=0.05):
    upper = df.quantile(1-q)
    lower = df.quantile(q)
    mask = (df < upper) & (df > lower)
    return mask

mask = remove_outliers(df_data_edlevel_income['ConvertedCompYearly'], 0.1)

print(df_data_edlevel_income[mask])
```

```
       ConvertedCompYearly           EdLevel
45                 60480.0      Licenciatura
50                 25944.0      Licenciatura
58                 22644.0  Grado Profesional
76                 45564.0      Licenciatura
77                151263.0         Doctorado
…                      …                 …
83425              19452.0        Secundaria
83428              41232.0            Master
83431              11676.0      Licenciatura
83432              80169.0      Licenciatura
83436              90000.0        Secundaria

[11611 rows x 2 columns]
```

[503]:
```
df_data_edlevel_income = df_data_edlevel_income[mask]
```

[505]:
```
df_data_edlevel_income['ConvertedCompYearlyCategorical'] = 'ALTO'
df_data_edlevel_income.loc[(df_data_edlevel_income['ConvertedCompYearly'] >= 0)␣
 ↪& (df_data_edlevel_income['ConvertedCompYearly'] <= 32747),␣
 ↪'ConvertedCompYearlyCategorical'] = 'BAJO'
df_data_edlevel_income.loc[(df_data_edlevel_income['ConvertedCompYearly'] >␣
 ↪32747) & (df_data_edlevel_income['ConvertedCompYearly'] <= 90000),␣
 ↪'ConvertedCompYearlyCategorical'] = 'MEDIO'

print(df_data_edlevel_income)
```

```
       ConvertedCompYearly           EdLevel ConvertedCompYearlyCategorical
45                 60480.0      Licenciatura                          MEDIO
50                 25944.0      Licenciatura                           BAJO
58                 22644.0  Grado Profesional                          BAJO
76                 45564.0      Licenciatura                          MEDIO
77                151263.0         Doctorado                           ALTO
…                      …                 …                            …
83425              19452.0        Secundaria                           BAJO
83428              41232.0            Master                          MEDIO
83431              11676.0      Licenciatura                           BAJO
83432              80169.0      Licenciatura                          MEDIO
```

```
83436                 90000.0         Secundaria                                MEDIO

[11611 rows x 3 columns]
```

[506]:
```
df_data_edlevel_income =␣
↪df_data_edlevel_income[df_data_edlevel_income['ConvertedCompYearlyCategorical']␣
↪== 'ALTO']
```

[507]:
```
df_data_edlevel_income = df_data_edlevel_income[['EdLevel',␣
↪'ConvertedCompYearlyCategorical']]
```

[527]:
```
df_flourish_004 = df_data_edlevel_income['EdLevel'].value_counts().
↪to_frame('counts').reset_index()
```

[528]:
```
df_flourish_004
```

[528]:
```
                index  counts
0        Licenciatura    1481
1              Master     715
2  Estudios sin grado     356
3      Grado Asociado     117
4           Doctorado      96
5          Secundaria      80
6    Grado Profesional     21
7            Primaria      13
8                Otro      13
```

[529]:
```
df_flourish_004.to_csv('004_df_flourish.csv', index=False)
```

3.0.5 2.5. ¿Existe brecha salarial entre hombres y mujeres u otros géneros?, y de ¿Cuánto es la diferencia? ¿Cuáles son los peores países en cuanto a brecha salarial? ¿Cuáles son los países que han reducido esta brecha salarial entre programadores?

Se seleccionarán los campos adecuados para responder a esta pregunta

[585]:
```
data_wage_gap = data_test[['Country', 'ConvertedCompYearly', 'Gender']]
data_wage_gap.head()
```

[585]:
```
                     Country  ConvertedCompYearly  Gender
45                    Brazil              60480.0  Hombre
50                    Greece              25944.0  Hombre
58        Russian Federation              22644.0  Hombre
64  United States of America             500000.0  Hombre
76                    Poland              45564.0  Hombre
```

[587]:
```
df_data_wage_gap = data_wage_gap.copy()
```

[588]:

```
def remove_outliers(df, q=0.05):
    upper = df.quantile(1-q)
    lower = df.quantile(q)
    mask = (df < upper) & (df > lower)
    return mask

mask = remove_outliers(df_data_wage_gap['ConvertedCompYearly'], 0.1)

print(df_data_wage_gap[mask])
```

```
                        Country  ConvertedCompYearly  Gender
45                       Brazil              60480.0  Hombre
50                       Greece              25944.0  Hombre
58           Russian Federation              22644.0  Hombre
76                       Poland              45564.0  Hombre
77                       Canada             151263.0  Hombre
…                         …                  …     …
83425                   Finland              19452.0  Hombre
83428                    Brazil              41232.0  Hombre
83431                  Pakistan              11676.0  Hombre
83432                    Canada              80169.0   Mujer
83436  United States of America              90000.0  Hombre

[11611 rows x 3 columns]
```

[589]:

```
df_data_wage_gap = df_data_wage_gap[mask]
```

[591]:

```
df_data_wage_gap['ConvertedCompYearlyCategorical'] = 'ALTO'
df_data_wage_gap.loc[(df_data_wage_gap['ConvertedCompYearly'] >= 0) &␣
 ↪(df_data_wage_gap['ConvertedCompYearly'] <= 32747),␣
 ↪'ConvertedCompYearlyCategorical'] = 'BAJO'
df_data_wage_gap.loc[(df_data_wage_gap['ConvertedCompYearly'] > 32747) &␣
 ↪(df_data_wage_gap['ConvertedCompYearly'] <= 90000),␣
 ↪'ConvertedCompYearlyCategorical'] = 'MEDIO'

print(df_data_wage_gap)
```

```
                        Country  ConvertedCompYearly  Gender  \
45                       Brazil              60480.0  Hombre
50                       Greece              25944.0  Hombre
58           Russian Federation              22644.0  Hombre
76                       Poland              45564.0  Hombre
77                       Canada             151263.0  Hombre
…                         …                  …     …
83425                   Finland              19452.0  Hombre
83428                    Brazil              41232.0  Hombre
83431                  Pakistan              11676.0  Hombre
83432                    Canada              80169.0   Mujer
```

```
83436  United States of America            90000.0  Hombre

      ConvertedCompYearlyCategorical
45                             MEDIO
50                              BAJO
58                              BAJO
76                             MEDIO
77                              ALTO
…                                …
83425                           BAJO
83428                          MEDIO
83431                           BAJO
83432                          MEDIO
83436                          MEDIO

[11611 rows x 4 columns]
```

[592]:
```
df_data_wage_gap =␣
 ↪df_data_wage_gap[df_data_wage_gap['ConvertedCompYearlyCategorical'].
 ↪isin(['ALTO', 'MEDIO'])]
```

[593]:
```
df_data_wage_gap = df_data_wage_gap[['Country', 'Gender',␣
 ↪'ConvertedCompYearlyCategorical']]
```

[595]:
```
df_data_wage_gap.to_csv('005_df_data_wage_gap.csv', index=False)
```

[572]:
```
df_data_wage_gap['ConvertedCompYearlyCategorical'].drop_duplicates().
 ↪sort_values()
```

[572]:
```
77     ALTO
45    MEDIO
Name: ConvertedCompYearlyCategorical, dtype: object
```

[573]:
```
df_data_wage_gap['Gender'].drop_duplicates().sort_values()
```

[573]:
```
45                   Hombre
264                   Mujer
702     No binario u otro
2559           No definido
Name: Gender, dtype: object
```

[574]:
```
df_data_wage_gap['Country'].drop_duplicates().sort_values()
```

[574]:
```
27198                             Afghanistan
54847                                 Albania
25364                                 Algeria
34843                                 Andorra
289                                 Argentina
```

```
                        …
128                   United States of America
1759                                   Uruguay
44422    Venezuela, Bolivarian Republic of…
10617                                 Viet Nam
27638                                   Zambia
Name: Country, Length: 126, dtype: object
```

```
[575]: df_data_wage_gap1 = df_data_wage_gap.copy()
```

```
[615]: df_flourish_005 = df_data_wage_gap1.groupby(['Country', 'Gender']).size().
        ↪unstack(fill_value=0).sort_values('Hombre')
```

```
[616]: df_flourish_005 = df_flourish_005.apply(lambda x: pd.concat([x.head(40), x.
        ↪tail(5)]))
```

```
[609]: df_flourish_005.to_csv('005_flourish_data.csv', index=True)
```

3.0.6 2.6. ¿Cuáles son los ingresos promedios según los rangos de edad? ¿Cuál es el rango de edad con el mejor y peor ingreso?

Se seleccionarán los campos adecuados para responder a esta pregunta

```
[618]: data_age_income = data_test[['ConvertedCompYearly', 'Age']]
       data_age_income.head()
```

```
[618]:     ConvertedCompYearly    Age
       45              60480.0  35-44
       50              25944.0  25-34
       58              22644.0  25-34
       64             500000.0  35-44
       76              45564.0  25-34
```

```
[619]: df_data_age_income = data_age_income.copy()
```

```
[620]: def remove_outliers(df, q=0.05):
           upper = df.quantile(1-q)
           lower = df.quantile(q)
           mask = (df < upper) & (df > lower)
           return mask

       mask = remove_outliers(df_data_age_income['ConvertedCompYearly'], 0.1)

       print(df_data_age_income[mask])
```

```
       ConvertedCompYearly    Age
45                 60480.0  35-44
50                 25944.0  25-34
58                 22644.0  25-34
```

```
76                    45564.0   25-34
77                   151263.0   35-44
…                         …     …
83425                 19452.0   18-24
83428                 41232.0   25-34
83431                 11676.0   18-24
83432                 80169.0   18-24
83436                 90000.0   25-34

[11611 rows x 2 columns]
```

[621]:
```
df_data_age_income = df_data_age_income[mask]
```

[625]:
```
df_data_age_income1 = df_data_age_income.copy()
```

[643]:
```
df_data_age_income1.to_csv('006_df_data_age_income1.csv', index=False)
```

[627]:
```
grouped_df = df_data_age_income1.groupby("Age")

average_df = grouped_df.mean()
```

[628]:
```
average_df
```

[628]:
```
             ConvertedCompYearly
Age
18-24                43758.228943
25-34                60962.367068
35-44                76911.641812
45-54                87229.578231
55-64               100102.974874
< 18                 39841.117647
>= 65                95988.611111
No definido          77170.666667
```

[644]:
```
df_flourish_006 = average_df.copy()
```

[646]:
```
df_flourish_006.to_csv('006_df_flourish_006.csv', index=True)
```

3.0.7 2.7. ¿Cuáles son las tecnologías que permiten tener un mejor ingreso salarial anual?

Se seleccionarán los campos adecuados para responder a esta pregunta

[754]:

```
data_techs_best_income1 = data_test[['ConvertedCompYearly',␣
 ↪'LanguageHaveWorkedWith', 'DatabaseHaveWorkedWith',␣
 ↪'PlatformHaveWorkedWith', 'WebframeHaveWorkedWith',␣
 ↪'MiscTechHaveWorkedWith',  'ToolsTechHaveWorkedWith',␣
 ↪'NEWCollabToolsHaveWorkedWith']]
data_techs_best_income1.head()
```

[754]:

```
    ConvertedCompYearly                             LanguageHaveWorkedWith  \
45              60480.0        C#;C++;JavaScript;PowerShell;SQL;TypeScript
50              25944.0  C#;HTML/CSS;JavaScript;Node.js;PowerShell;Type…
58              22644.0          Bash/Shell;HTML/CSS;JavaScript;Python;SQL
64             500000.0                         HTML/CSS;JavaScript;Python
76              45564.0  Bash/Shell;C#;Dart;Delphi;Go;HTML/CSS;Java;Jav…

                               DatabaseHaveWorkedWith  \
45              Microsoft SQL Server;PostgreSQL;Redis
50  Couchbase;MariaDB;Microsoft SQL Server;MongoDB…
58                                             Oracle
64                                              MySQL
76  Firebase;Microsoft SQL Server;MongoDB;MySQL;Po…

                   PlatformHaveWorkedWith  \
45                 Heroku;Microsoft Azure
50       AWS;DigitalOcean;Microsoft Azure
58                                 Heroku
64                                    AWS
76  Google Cloud Platform;Microsoft Azure

                               WebframeHaveWorkedWith  \
45                             ASP.NET Core ;React.js
50       Angular;ASP.NET;ASP.NET Core ;Express;Svelte
58                              Django;FastAPI;Vue.js
64                                              Flask
76  Angular;Angular.js;ASP.NET;ASP.NET Core ;Djang…

                               MiscTechHaveWorkedWith ToolsTechHaveWorkedWith  \
45                                 .NET Core / .NET 5   Docker;Git;Kubernetes
50                  .NET Framework;.NET Core / .NET 5       Docker;Kubernetes
58                         NumPy;Pandas;Torch/PyTorch              Docker;Git
64                                             Pandas                     Git
76  .NET Framework;.NET Core / .NET 5;Apache Spark…     Docker;Git;Unity 3D

                         NEWCollabToolsHaveWorkedWith
45         Notepad++;Visual Studio;Visual Studio Code
50         Notepad++;Visual Studio;Visual Studio Code
58                 IPython/Jupyter;Visual Studio Code
64                     Notepad++;PyCharm;Sublime Text
```

```
76   Android Studio;Eclipse;NetBeans;Notepad++;Visu…
```

[755]:

```
data_techs_best_income1['AllTechs'] =␣
 ↪data_techs_best_income1['LanguageHaveWorkedWith'].map(str) + ';' +␣
 ↪data_techs_best_income1['DatabaseHaveWorkedWith'].map(str) + ';' +␣
 ↪data_techs_best_income1['PlatformHaveWorkedWith'].map(str) + ';' +␣
 ↪data_techs_best_income1['WebframeHaveWorkedWith'].map(str) + ';' +␣
 ↪data_techs_best_income1['MiscTechHaveWorkedWith'].map(str) + ';' +␣
 ↪data_techs_best_income1['ToolsTechHaveWorkedWith'].map(str) + ';' +␣
 ↪data_techs_best_income1['NEWCollabToolsHaveWorkedWith'].map(str)
print (data_techs_best_income1)
```

```
       ConvertedCompYearly                             LanguageHaveWorkedWith  \
45                 60480.0        C#;C++;JavaScript;PowerShell;SQL;TypeScript
50                 25944.0  C#;HTML/CSS;JavaScript;Node.js;PowerShell;Type…
58                 22644.0          Bash/Shell;HTML/CSS;JavaScript;Python;SQL
64                500000.0                         HTML/CSS;JavaScript;Python
76                 45564.0  Bash/Shell;C#;Dart;Delphi;Go;HTML/CSS;Java;Jav…
…                    …                                                 …
83428              41232.0                      Bash/Shell;Node.js;TypeScript
83431              11676.0  C#;Dart;HTML/CSS;Java;JavaScript;Kotlin;Node.j…
83432              80169.0                                               Ruby
83436              90000.0                                 Groovy;Java;Python
83437             816816.0               Bash/Shell;JavaScript;Node.js;Python

                                  DatabaseHaveWorkedWith  \
45                 Microsoft SQL Server;PostgreSQL;Redis
50     Couchbase;MariaDB;Microsoft SQL Server;MongoDB…
58                                                Oracle
64                                                 MySQL
76     Firebase;Microsoft SQL Server;MongoDB;MySQL;Po…
…                                                  …
83428             Elasticsearch;MongoDB;PostgreSQL;Redis
83431                              Firebase;MySQL;SQLite
83432                                   MySQL;PostgreSQL
83436    DynamoDB;Elasticsearch;MongoDB;PostgreSQL;Redis
83437   Cassandra;Elasticsearch;MongoDB;PostgreSQL;Redis

                      PlatformHaveWorkedWith  \
45                    Heroku;Microsoft Azure
50          AWS;DigitalOcean;Microsoft Azure
58                                    Heroku
64                                       AWS
76     Google Cloud Platform;Microsoft Azure
…                                      …
83428              AWS;Google Cloud Platform
83431                  Google Cloud Platform
83432           Google Cloud Platform;Heroku
```

```
83436              AWS;Google Cloud Platform
83437                                 Heroku

                                  WebframeHaveWorkedWith  \
45                                ASP.NET Core ;React.js
50          Angular;ASP.NET;ASP.NET Core ;Express;Svelte
58                                 Django;FastAPI;Vue.js
64                                                 Flask
76     Angular;Angular.js;ASP.NET;ASP.NET Core ;Djang…
…                                                   …
83428                                           React.js
83431                                       Flask;jQuery
83432               Flask;React.js;Ruby on Rails;Vue.js
83436                                      FastAPI;Flask
83437                     Django;Express;Flask;React.js

                                  MiscTechHaveWorkedWith  \
45                                    .NET Core / .NET 5
50                     .NET Framework;.NET Core / .NET 5
58                            NumPy;Pandas;Torch/PyTorch
64                                                Pandas
76     .NET Framework;.NET Core / .NET 5;Apache Spark…
…                                                   …
83428                                       React Native
83431                                            Flutter
83432             NumPy;Pandas;TensorFlow;Torch/PyTorch
83436                          Hadoop;Keras;NumPy;Pandas
83437             NumPy;Pandas;TensorFlow;Torch/PyTorch

            ToolsTechHaveWorkedWith  \
45              Docker;Git;Kubernetes
50                  Docker;Kubernetes
58                         Docker;Git
64                                Git
76                Docker;Git;Unity 3D
…                               …
83428      Docker;Git;Terraform;Yarn
83431                             Git
83432     Docker;Git;Kubernetes;Yarn
83436  Ansible;Docker;Git;Terraform
83437  Ansible;Docker;Git;Terraform

                            NEWCollabToolsHaveWorkedWith  \
45              Notepad++;Visual Studio;Visual Studio Code
50              Notepad++;Visual Studio;Visual Studio Code
58                      IPython/Jupyter;Visual Studio Code
64                          Notepad++;PyCharm;Sublime Text
76     Android Studio;Eclipse;NetBeans;Notepad++;Visu…
```

```
…                                                    …
83428                          Visual Studio Code;Webstorm
83431   Android Studio;IntelliJ;IPython/Jupyter;Notepa…
83432         Atom;IPython/Jupyter;Vim;Visual Studio Code
83436   Android Studio;Eclipse;IntelliJ;IPython/Jupyte…
83437                                 PyCharm;Sublime Text

                                                 AllTechs
45      C#;C++;JavaScript;PowerShell;SQL;TypeScript;Mi…
50      C#;HTML/CSS;JavaScript;Node.js;PowerShell;Type…
58      Bash/Shell;HTML/CSS;JavaScript;Python;SQL;Orac…
64      HTML/CSS;JavaScript;Python;MySQL;AWS;Flask;Pan…
76      Bash/Shell;C#;Dart;Delphi;Go;HTML/CSS;Java;Jav…
…                                                    …
83428   Bash/Shell;Node.js;TypeScript;Elasticsearch;Mo…
83431   C#;Dart;HTML/CSS;Java;JavaScript;Kotlin;Node.j…
83432   Ruby;MySQL;PostgreSQL;Google Cloud Platform;He…
83436   Groovy;Java;Python;DynamoDB;Elasticsearch;Mong…
83437   Bash/Shell;JavaScript;Node.js;Python;Cassandra…

[14517 rows x 9 columns]
```

```
C:\Users\GPBONI~1\AppData\Local\Temp/ipykernel_9952/782511894.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  data_techs_best_income1['AllTechs'] =
data_techs_best_income1['LanguageHaveWorkedWith'].map(str) + ';' +
data_techs_best_income1['DatabaseHaveWorkedWith'].map(str) + ';' +
data_techs_best_income1['PlatformHaveWorkedWith'].map(str) + ';' +
data_techs_best_income1['WebframeHaveWorkedWith'].map(str) + ';' +
data_techs_best_income1['MiscTechHaveWorkedWith'].map(str) + ';' +
data_techs_best_income1['ToolsTechHaveWorkedWith'].map(str) + ';' +
data_techs_best_income1['NEWCollabToolsHaveWorkedWith'].map(str)
```

[757]:

```
df_data_techs_best_income = data_techs_best_income1[['ConvertedCompYearly',␣
 ↪'AllTechs']].copy()
```

[759]:

```
df_data_techs_best_income1 = df_data_techs_best_income.copy()
```

[760]:

```
def remove_outliers(df, q=0.05):
    upper = df.quantile(1-q)
    lower = df.quantile(q)
    mask = (df < upper) & (df > lower)
    return mask
```

```
mask = remove_outliers(df_data_techs_best_income1['ConvertedCompYearly'], 0.1)

print(df_data_techs_best_income1[mask])
```

```
       ConvertedCompYearly                                           AllTechs
45                 60480.0  C#;C++;JavaScript;PowerShell;SQL;TypeScript;Mi…
50                 25944.0  C#;HTML/CSS;JavaScript;Node.js;PowerShell;Type…
58                 22644.0  Bash/Shell;HTML/CSS;JavaScript;Python;SQL;Orac…
76                 45564.0  Bash/Shell;C#;Dart;Delphi;Go;HTML/CSS;Java;Jav…
77                151263.0  HTML/CSS;Python;R;DynamoDB;AWS;Flask;Keras;Num…
…                      …                                                  …
83425              19452.0  HTML/CSS;JavaScript;Node.js;TypeScript;DynamoD…
83428              41232.0  Bash/Shell;Node.js;TypeScript;Elasticsearch;Mo…
83431              11676.0  C#;Dart;HTML/CSS;Java;JavaScript;Kotlin;Node.j…
83432              80169.0  Ruby;MySQL;PostgreSQL;Google Cloud Platform;He…
83436              90000.0  Groovy;Java;Python;DynamoDB;Elasticsearch;Mong…

[11611 rows x 2 columns]
```

[761]:
```
df_data_techs_best_income1 = df_data_techs_best_income1[mask]
```

[762]:
```
df_data_techs_best_income1['ConvertedCompYearlyCategorical'] = 'ALTO'
df_data_techs_best_income1.
 ↪loc[(df_data_techs_best_income1['ConvertedCompYearly'] >= 0) &␣
 ↪(df_data_techs_best_income1['ConvertedCompYearly'] <= 32747),␣
 ↪'ConvertedCompYearlyCategorical'] = 'BAJO'
df_data_techs_best_income1.
 ↪loc[(df_data_techs_best_income1['ConvertedCompYearly'] > 32747) &␣
 ↪(df_data_techs_best_income1['ConvertedCompYearly'] <= 90000),␣
 ↪'ConvertedCompYearlyCategorical'] = 'MEDIO'

print(df_data_techs_best_income1)
```

```
       ConvertedCompYearly                                           AllTechs  \
45                 60480.0  C#;C++;JavaScript;PowerShell;SQL;TypeScript;Mi…
50                 25944.0  C#;HTML/CSS;JavaScript;Node.js;PowerShell;Type…
58                 22644.0  Bash/Shell;HTML/CSS;JavaScript;Python;SQL;Orac…
76                 45564.0  Bash/Shell;C#;Dart;Delphi;Go;HTML/CSS;Java;Jav…
77                151263.0  HTML/CSS;Python;R;DynamoDB;AWS;Flask;Keras;Num…
…                      …                                                  …
83425              19452.0  HTML/CSS;JavaScript;Node.js;TypeScript;DynamoD…
83428              41232.0  Bash/Shell;Node.js;TypeScript;Elasticsearch;Mo…
83431              11676.0  C#;Dart;HTML/CSS;Java;JavaScript;Kotlin;Node.j…
83432              80169.0  Ruby;MySQL;PostgreSQL;Google Cloud Platform;He…
83436              90000.0  Groovy;Java;Python;DynamoDB;Elasticsearch;Mong…

      ConvertedCompYearlyCategorical
```

```
45                             MEDIO
50                              BAJO
58                              BAJO
76                             MEDIO
77                              ALTO
…                              …
83425                           BAJO
83428                          MEDIO
83431                           BAJO
83432                          MEDIO
83436                          MEDIO

[11611 rows x 3 columns]
```

```
[763]: df_data_techs_best_income1 =␣
        ↪df_data_techs_best_income1[df_data_techs_best_income1['ConvertedCompYearlyCategorical'].
        ↪isin(['ALTO', 'MEDIO'])]
```

```
[765]: df_data_techs_best_income1['AllTechs'] = df_data_techs_best_income1['AllTechs'].
        ↪str.replace(' ', '')
```

```
[766]: df_data_techs_best_income1['AllTechs'] = df_data_techs_best_income1['AllTechs'].
        ↪str.replace(';', ' ')
```

```
[768]: df_counts = df_data_techs_best_income1['AllTechs'].str.split(expand=True).
        ↪stack().value_counts().rename_axis('Tech').reset_index(name='Count')
```

```
[771]: df_counts.head(10)
```

```
[771]:                Tech  Count
        0               Git   8300
        1  VisualStudioCode   7131
        2        JavaScript   7057
        3            Docker   5879
        4          HTML/CSS   5821
        5               SQL   5699
        6               AWS   5066
        7                C#   4717
        8        TypeScript   4531
        9      VisualStudio   4497
```

```
[772]: df_data_techs_best_income_007 = df_counts.head(10)
```

```
[773]: df_data_techs_best_income_007.to_csv('007_df_data_techs_best_income.csv',␣
        ↪index=False)
```

3.0.8 2.8. ¿Cuántas tecnologías en promedio domina un programador profesional?

Se seleccionarán los campos adecuados para responder a esta pregunta

[792]:
```
data_techs_dev_pro1 = data_test[['DevType', 'LanguageHaveWorkedWith',␣
 ↪'DatabaseHaveWorkedWith', 'PlatformHaveWorkedWith',␣
 ↪'WebframeHaveWorkedWith', 'MiscTechHaveWorkedWith', ␣
 ↪'ToolsTechHaveWorkedWith', 'NEWCollabToolsHaveWorkedWith']]
data_techs_dev_pro1.head()
```

[792]:
```
                     DevType  \
45  Desarrollador Escritorio
50  Desarrollador full-stack
58  Desarrollador full-stack
64   Desarrollador front-end
76   Desarrollador front-end

                               LanguageHaveWorkedWith  \
45        C#;C++;JavaScript;PowerShell;SQL;TypeScript
50  C#;HTML/CSS;JavaScript;Node.js;PowerShell;Type…
58          Bash/Shell;HTML/CSS;JavaScript;Python;SQL
64                         HTML/CSS;JavaScript;Python
76  Bash/Shell;C#;Dart;Delphi;Go;HTML/CSS;Java;Jav…

                               DatabaseHaveWorkedWith  \
45              Microsoft SQL Server;PostgreSQL;Redis
50  Couchbase;MariaDB;Microsoft SQL Server;MongoDB…
58                                             Oracle
64                                              MySQL
76  Firebase;Microsoft SQL Server;MongoDB;MySQL;Po…

                   PlatformHaveWorkedWith  \
45                 Heroku;Microsoft Azure
50       AWS;DigitalOcean;Microsoft Azure
58                                 Heroku
64                                    AWS
76  Google Cloud Platform;Microsoft Azure

                               WebframeHaveWorkedWith  \
45                             ASP.NET Core ;React.js
50       Angular;ASP.NET;ASP.NET Core ;Express;Svelte
58                              Django;FastAPI;Vue.js
64                                              Flask
76  Angular;Angular.js;ASP.NET;ASP.NET Core ;Djang…

                               MiscTechHaveWorkedWith ToolsTechHaveWorkedWith  \
45                                 .NET Core / .NET 5   Docker;Git;Kubernetes
50                  .NET Framework;.NET Core / .NET 5       Docker;Kubernetes
```

```
58                          NumPy;Pandas;Torch/PyTorch              Docker;Git
64                                                  Pandas                     Git
76   .NET Framework;.NET Core / .NET 5;Apache Spark…     Docker;Git;Unity 3D

                            NEWCollabToolsHaveWorkedWith
45         Notepad++;Visual Studio;Visual Studio Code
50         Notepad++;Visual Studio;Visual Studio Code
58                 IPython/Jupyter;Visual Studio Code
64                     Notepad++;PyCharm;Sublime Text
76   Android Studio;Eclipse;NetBeans;Notepad++;Visu…
```

[793]:
```
data_techs_dev_pro1['AllTechs'] = data_techs_dev_pro1['LanguageHaveWorkedWith'].
 ↪map(str) + ';' + data_techs_dev_pro1['DatabaseHaveWorkedWith'].map(str) + ';
 ↪' + data_techs_dev_pro1['PlatformHaveWorkedWith'].map(str) + ';' +␣
 ↪data_techs_dev_pro1['WebframeHaveWorkedWith'].map(str) + ';' +␣
 ↪data_techs_dev_pro1['MiscTechHaveWorkedWith'].map(str) + ';' +␣
 ↪data_techs_best_income1['ToolsTechHaveWorkedWith'].map(str) + ';' +␣
 ↪data_techs_dev_pro1['NEWCollabToolsHaveWorkedWith'].map(str)
print (data_techs_dev_pro1)
```

```
                          DevType  \
45      Desarrollador Escritorio
50      Desarrollador full-stack
58      Desarrollador full-stack
64       Desarrollador front-end
76       Desarrollador front-end
…                             …
83428           Ejecutivo Senior
83431        Desarrollador móvil
83432     Desarrollador back-end
83436         Cientifico de datos
83437     Desarrollador back-end

                                    LanguageHaveWorkedWith  \
45            C#;C++;JavaScript;PowerShell;SQL;TypeScript
50       C#;HTML/CSS;JavaScript;Node.js;PowerShell;Type…
58              Bash/Shell;HTML/CSS;JavaScript;Python;SQL
64                              HTML/CSS;JavaScript;Python
76       Bash/Shell;C#;Dart;Delphi;Go;HTML/CSS;Java;Jav…
…                                                    …
83428                         Bash/Shell;Node.js;TypeScript
83431   C#;Dart;HTML/CSS;Java;JavaScript;Kotlin;Node.j…
83432                                                  Ruby
83436                                    Groovy;Java;Python
83437                 Bash/Shell;JavaScript;Node.js;Python

                                    DatabaseHaveWorkedWith  \
45                  Microsoft SQL Server;PostgreSQL;Redis
```

```
50     Couchbase;MariaDB;Microsoft SQL Server;MongoDB…
58                                                Oracle
64                                                 MySQL
76     Firebase;Microsoft SQL Server;MongoDB;MySQL;Po…
…                                                  …
83428             Elasticsearch;MongoDB;PostgreSQL;Redis
83431                              Firebase;MySQL;SQLite
83432                                   MySQL;PostgreSQL
83436    DynamoDB;Elasticsearch;MongoDB;PostgreSQL;Redis
83437   Cassandra;Elasticsearch;MongoDB;PostgreSQL;Redis

                      PlatformHaveWorkedWith  \
45                    Heroku;Microsoft Azure
50          AWS;DigitalOcean;Microsoft Azure
58                                    Heroku
64                                       AWS
76     Google Cloud Platform;Microsoft Azure
…                                      …
83428              AWS;Google Cloud Platform
83431                  Google Cloud Platform
83432           Google Cloud Platform;Heroku
83436              AWS;Google Cloud Platform
83437                                 Heroku

                                  WebframeHaveWorkedWith  \
45                                ASP.NET Core ;React.js
50          Angular;ASP.NET;ASP.NET Core ;Express;Svelte
58                                 Django;FastAPI;Vue.js
64                                                 Flask
76     Angular;Angular.js;ASP.NET;ASP.NET Core ;Djang…
…                                                  …
83428                                           React.js
83431                                       Flask;jQuery
83432                Flask;React.js;Ruby on Rails;Vue.js
83436                                      FastAPI;Flask
83437                      Django;Express;Flask;React.js

                                  MiscTechHaveWorkedWith  \
45                                    .NET Core / .NET 5
50                     .NET Framework;.NET Core / .NET 5
58                            NumPy;Pandas;Torch/PyTorch
64                                                Pandas
76     .NET Framework;.NET Core / .NET 5;Apache Spark…
…                                                  …
83428                                       React Native
83431                                            Flutter
83432              NumPy;Pandas;TensorFlow;Torch/PyTorch
83436                          Hadoop;Keras;NumPy;Pandas
```

```
83437             NumPy;Pandas;TensorFlow;Torch/PyTorch

            ToolsTechHaveWorkedWith  \
45            Docker;Git;Kubernetes
50                Docker;Kubernetes
58                       Docker;Git
64                              Git
76              Docker;Git;Unity 3D
…                             …
83428     Docker;Git;Terraform;Yarn
83431                           Git
83432    Docker;Git;Kubernetes;Yarn
83436  Ansible;Docker;Git;Terraform
83437  Ansible;Docker;Git;Terraform

                           NEWCollabToolsHaveWorkedWith  \
45           Notepad++;Visual Studio;Visual Studio Code
50           Notepad++;Visual Studio;Visual Studio Code
58                   IPython/Jupyter;Visual Studio Code
64                       Notepad++;PyCharm;Sublime Text
76     Android Studio;Eclipse;NetBeans;Notepad++;Visu…
…                                                 …
83428                       Visual Studio Code;Webstorm
83431  Android Studio;IntelliJ;IPython/Jupyter;Notepa…
83432       Atom;IPython/Jupyter;Vim;Visual Studio Code
83436  Android Studio;Eclipse;IntelliJ;IPython/Jupyte…
83437                              PyCharm;Sublime Text

                                               AllTechs
45     C#;C++;JavaScript;PowerShell;SQL;TypeScript;Mi…
50     C#;HTML/CSS;JavaScript;Node.js;PowerShell;Type…
58     Bash/Shell;HTML/CSS;JavaScript;Python;SQL;Orac…
64     HTML/CSS;JavaScript;Python;MySQL;AWS;Flask;Pan…
76     Bash/Shell;C#;Dart;Delphi;Go;HTML/CSS;Java;Jav…
…                                                 …
83428  Bash/Shell;Node.js;TypeScript;Elasticsearch;Mo…
83431  C#;Dart;HTML/CSS;Java;JavaScript;Kotlin;Node.j…
83432  Ruby;MySQL;PostgreSQL;Google Cloud Platform;He…
83436  Groovy;Java;Python;DynamoDB;Elasticsearch;Mong…
83437  Bash/Shell;JavaScript;Node.js;Python;Cassandra…

[14517 rows x 9 columns]

C:\Users\GPBONI~1\AppData\Local\Temp/ipykernel_9952/1321581082.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  data_techs_dev_pro1['AllTechs'] =
data_techs_dev_pro1['LanguageHaveWorkedWith'].map(str) + ';' +
data_techs_dev_pro1['DatabaseHaveWorkedWith'].map(str) + ';' +
data_techs_dev_pro1['PlatformHaveWorkedWith'].map(str) + ';' +
data_techs_dev_pro1['WebframeHaveWorkedWith'].map(str) + ';' +
data_techs_dev_pro1['MiscTechHaveWorkedWith'].map(str) + ';' +
data_techs_best_income1['ToolsTechHaveWorkedWith'].map(str) + ';' +
data_techs_dev_pro1['NEWCollabToolsHaveWorkedWith'].map(str)
```

[794]:

```python
df_data_techs_dev_pro = data_techs_dev_pro1[['DevType', 'AllTechs']].copy()
```

[796]:

```python
df_data_techs_dev_pro = df_data_techs_dev_pro[df_data_techs_dev_pro['DevType'].
 ↪isin(['Desarrollador full-stack', 'Desarrollador front-end', 'Desarrollador␣
 ↪móvil', 'Desarrollador back-end', 'Desarrollador Escritorio', 'Desarrollador␣
 ↪de QA o Test', 'Desarrollador de aplicaciones embebidas', 'Administrador de␣
 ↪base de datos', 'Desarrollador de juegos o gráfico'])]
```

[797]:

```python
df_data_techs_dev_pro.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 12832 entries, 45 to 83437
Data columns (total 2 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   DevType   12832 non-null  object
 1   AllTechs  12832 non-null  object
dtypes: object(2)
memory usage: 300.8+ KB
```

[858]:

```python
df_data_techs_dev_pro1 = df_data_techs_dev_pro.copy()
```

[860]:

```python
df_data_techs_dev_pro1.to_csv('008_df_data_techs_dev_pro1.csv', index=True)
```

[866]:

```python
def convert_row_to_list(lst):
    return lst.split(';')
```

[867]:

```python
df_data_techs_dev_pro1['ListTechs'] = df_data_techs_dev_pro1['AllTechs'].
 ↪apply(convert_row_to_list)
```

[868]:

```python
df_data_techs_dev_pro1['LenListTechs'] = df_data_techs_dev_pro1['ListTechs'].
 ↪map(len)
```

[871]:

```python
df_flourish_008 = df_data_techs_dev_pro1[['DevType', 'LenListTechs']].copy()
df_flourish_008
```

[871]:
```
                          DevType  LenListTechs
45     Desarrollador Escritorio            20
50     Desarrollador full-stack            30
58     Desarrollador full-stack            17
64      Desarrollador front-end            11
76      Desarrollador front-end            50
…                           …           …
83423  Desarrollador full-stack            26
83425  Desarrollador full-stack            14
83431       Desarrollador móvil            28
83432    Desarrollador back-end            21
83437    Desarrollador back-end            24

[12832 rows x 2 columns]
```

[879]:
```
grouped_df = df_flourish_008.groupby("DevType")

average_df_008 = round(grouped_df.mean())
```

[882]:
```
df_flourish_008 = average_df_008.copy()
```

[884]:
```
df_flourish_008.to_csv('008_df_flourish_008.csv', index=True)
```

3.0.9 2.9. ¿En qué rango de edad se inició la mayoría de los programadores en la programación?

Se seleccionarán los campos adecuados para responder a esta pregunta

[886]:
```
data_age1stcode_dev_pro1 = data_test[['Age1stCode']]
data_age1stcode_dev_pro1.head()
```

[886]:
```
   Age1stCode
45      11-17
50      18-24
58      11-17
64      11-17
76      11-17
```

[888]:
```
data_age1stcode_dev_pro1 = data_age1stcode_dev_pro1['Age1stCode'].
 ↪value_counts().to_frame('counts').reset_index()
```

[891]:
```
data_age1stcode_dev_pro1.to_csv('009_flourish_data.csv', index=False)
```

3.0.10 2.10. ¿Cuántos años como programadores se requiere para obtener un ingreso salarial alto?

Se seleccionarán los campos adecuados para responder a esta pregunta

[929]:

```
data_yearscode_high_income1 = data_test[['ConvertedCompYearly', 'YearsCode']]
data_yearscode_high_income1.head()
```

[929]:

```
    ConvertedCompYearly YearsCode
45              60480.0        22
50              25944.0        12
58              22644.0         5
64             500000.0         6
76              45564.0        12
```

[930]:

```
df_data_yearscode_high_income = data_yearscode_high_income1.copy()
```

[931]:

```
def remove_outliers(df, q=0.05):
    upper = df.quantile(1-q)
    lower = df.quantile(q)
    mask = (df < upper) & (df > lower)
    return mask

mask = remove_outliers(df_data_yearscode_high_income['ConvertedCompYearly'], 0.
 ↪1)

print(df_data_yearscode_high_income[mask])
```

```
       ConvertedCompYearly YearsCode
45                 60480.0        22
50                 25944.0        12
58                 22644.0         5
76                 45564.0        12
77                151263.0        10
…                      …        …
83425              19452.0         5
83428              41232.0        12
83431              11676.0         9
83432              80169.0         5
83436              90000.0        10

[11611 rows x 2 columns]
```

[932]:

```
df_data_yearscode_high_income = df_data_yearscode_high_income[mask]
```

[933]:

```
df_data_yearscode_high_income['ConvertedCompYearlyCategorical'] = 'ALTO'

df_data_yearscode_high_income.
 ↪loc[(df_data_yearscode_high_income['ConvertedCompYearly'] >= 0) &␣
 ↪(df_data_yearscode_high_income['ConvertedCompYearly'] <= 32747),␣
 ↪'ConvertedCompYearlyCategorical'] = 'BAJO'
```

```
df_data_yearscode_high_income.
 ↪loc[(df_data_yearscode_high_income['ConvertedCompYearly'] > 32747) &␣
 ↪(df_data_yearscode_high_income['ConvertedCompYearly'] <= 90000),␣
 ↪'ConvertedCompYearlyCategorical'] = 'MEDIO'

print(df_data_yearscode_high_income)
```

```
       ConvertedCompYearly YearsCode ConvertedCompYearlyCategorical
45                 60480.0        22                          MEDIO
50                 25944.0        12                           BAJO
58                 22644.0         5                           BAJO
76                 45564.0        12                          MEDIO
77                151263.0        10                           ALTO
…                      …       …                             …
83425              19452.0         5                           BAJO
83428              41232.0        12                          MEDIO
83431              11676.0         9                           BAJO
83432              80169.0         5                          MEDIO
83436              90000.0        10                          MEDIO

[11611 rows x 3 columns]
```

[971]:
```
df_data_yearscode_high_income.to_csv('010_df_flourish.csv', index=False)
```

[953]:
```
df_data_yearscode_high_income['ConvertedCompYearlyCategorical'].value_counts()
```

[953]:
```
MEDIO    5816
BAJO     2903
ALTO     2892
Name: ConvertedCompYearlyCategorical, dtype: int64
```

[972]:
```
df_flourish_010 = df_data_yearscode_high_income[['YearsCode',␣
 ↪'ConvertedCompYearlyCategorical']].copy()

df_flourish_010.head()
```

[972]:
```
   YearsCode ConvertedCompYearlyCategorical
45        22                          MEDIO
50        12                           BAJO
58         5                           BAJO
76        12                          MEDIO
77        10                           ALTO
```

[974]:
```
df_flourish_010['YearsCode'] = pd.to_numeric(df_flourish_010['YearsCode'])
```

[975]:
```
df_flourish_010.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
Int64Index: 11611 entries, 45 to 83436
Data columns (total 2 columns):
 #   Column                         Non-Null Count  Dtype
---  ------                         --------------  -----
 0   YearsCode                      11611 non-null  int64
 1   ConvertedCompYearlyCategorical 11611 non-null  object
dtypes: int64(1), object(1)
memory usage: 530.2+ KB
```

[976]:

```
grouped_df_010 = df_flourish_010.groupby("ConvertedCompYearlyCategorical")

average_df_010 = round(grouped_df_010.mean())
```

[977]:

```
average_df_010
```

[977]:

```
                                YearsCode
ConvertedCompYearlyCategorical
ALTO                                 19.0
BAJO                                 10.0
MEDIO                                14.0
```

[978]:

```
average_df_010.to_csv('010_flourish_data.csv', index=True)
```

3.0.11 2.11. ¿Cuáles son los perfiles que registran los mejores ingresos?

Se seleccionarán los campos adecuados para responder a esta pregunta

[979]:

```
data_profiles_dev_high_income1 = data_test[['ConvertedCompYearly', 'DevType']].
 ↪copy()
data_profiles_dev_high_income1.head()
```

[979]:

```
    ConvertedCompYearly                    DevType
45              60480.0  Desarrollador Escritorio
50              25944.0  Desarrollador full-stack
58              22644.0  Desarrollador full-stack
64             500000.0   Desarrollador front-end
76              45564.0   Desarrollador front-end
```

[1010]:

```
df_data_profiles_dev_high_income = data_profiles_dev_high_income1.copy()
```

[1011]:

```
def remove_outliers(df, q=0.05):
    upper = df.quantile(1-q)
    lower = df.quantile(q)
    mask = (df < upper) & (df > lower)
    return mask

mask = remove_outliers(df_data_profiles_dev_high_income['ConvertedCompYearly'],␣
 ↪0.1)
```

```
print(df_data_profiles_dev_high_income[mask])
```

```
       ConvertedCompYearly                     DevType
45                 60480.0  Desarrollador Escritorio
50                 25944.0  Desarrollador full-stack
58                 22644.0  Desarrollador full-stack
76                 45564.0   Desarrollador front-end
77                151263.0       Cientifico de datos
…                      …                       …
83425              19452.0  Desarrollador full-stack
83428              41232.0          Ejecutivo Senior
83431              11676.0       Desarrollador móvil
83432              80169.0     Desarrollador back-end
83436              90000.0       Cientifico de datos

[11611 rows x 2 columns]
```

[1012]:
```
df_data_profiles_dev_high_income = df_data_profiles_dev_high_income[mask]
```

[1013]:
```
df_data_profiles_dev_high_income['ConvertedCompYearlyCategorical'] = 'ALTO'

df_data_profiles_dev_high_income.
 ↪loc[(df_data_profiles_dev_high_income['ConvertedCompYearly'] >= 0) &␣
 ↪(df_data_profiles_dev_high_income['ConvertedCompYearly'] <= 32747),␣
 ↪'ConvertedCompYearlyCategorical'] = 'BAJO'
df_data_profiles_dev_high_income.
 ↪loc[(df_data_profiles_dev_high_income['ConvertedCompYearly'] > 32747) &␣
 ↪(df_data_profiles_dev_high_income['ConvertedCompYearly'] <= 90000),␣
 ↪'ConvertedCompYearlyCategorical'] = 'MEDIO'

print(df_data_profiles_dev_high_income)
```

```
       ConvertedCompYearly                     DevType  \
45                 60480.0  Desarrollador Escritorio
50                 25944.0  Desarrollador full-stack
58                 22644.0  Desarrollador full-stack
76                 45564.0   Desarrollador front-end
77                151263.0       Cientifico de datos
…                      …                       …
83425              19452.0  Desarrollador full-stack
83428              41232.0          Ejecutivo Senior
83431              11676.0       Desarrollador móvil
83432              80169.0     Desarrollador back-end
83436              90000.0       Cientifico de datos

      ConvertedCompYearlyCategorical
45                             MEDIO
```

```
50                               BAJO
58                               BAJO
76                              MEDIO
77                               ALTO
…                               …
83425                            BAJO
83428                           MEDIO
83431                            BAJO
83432                           MEDIO
83436                           MEDIO

[11611 rows x 3 columns]
```

[1015]:
```
df_data_profiles_dev_high_income['ConvertedCompYearlyCategorical'].
 ↪value_counts()
```

[1015]:
```
MEDIO    5816
BAJO     2903
ALTO     2892
Name: ConvertedCompYearlyCategorical, dtype: int64
```

[1016]:
```
df_flourish_011 = df_data_profiles_dev_high_income[['DevType',␣
 ↪'ConvertedCompYearlyCategorical']].copy()
```

[1018]:
```
df_flourish_011 =␣
 ↪df_flourish_011[df_flourish_011['ConvertedCompYearlyCategorical'].
 ↪isin(['ALTO'])]
```

[1019]:
```
df_flourish_011.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2892 entries, 77 to 83372
Data columns (total 2 columns):
 #   Column                          Non-Null Count  Dtype 
---  ------                          --------------  ----- 
 0   DevType                         2892 non-null   object
 1   ConvertedCompYearlyCategorical  2892 non-null   object
dtypes: object(2)
memory usage: 67.8+ KB
```

[1021]:
```
df_data_flourish_011 = df_flourish_011['DevType'].value_counts().
 ↪to_frame('counts').reset_index()
```

[1023]:
```
df_data_flourish_011 = df_data_flourish_011.head(10)
```

[1024]:
```
df_data_flourish_011
```

[1024]:

```
                       index  counts
0  Desarrollador full-stack     981
1   Desarrollador front-end     539
2       Desarrollador móvil     380
3    Desarrollador back-end     302
4  Desarrollador Escritorio     262
5         Ingeniero de datos    142
6        Cientifico de datos     70
7      Manager de Ingeniería     48
8                       Otro     36
9    Especialista en DevOps      32
```

[1025]:

```
df_data_flourish_011.to_csv('011_flourish_data.csv', index=False)
```

3.0.12 2.12. ¿Cuáles son las 10 tecnologías más usadas entre los programadores por países?

Se seleccionarán los campos adecuados para responder a esta pregunta

[390]:

```
data_10_techs_popular_dev_countries = data_test[['Country',␣
 ↪'LanguageHaveWorkedWith', 'DatabaseHaveWorkedWith',␣
 ↪'PlatformHaveWorkedWith', 'WebframeHaveWorkedWith',␣
 ↪'MiscTechHaveWorkedWith',  'ToolsTechHaveWorkedWith',␣
 ↪'NEWCollabToolsHaveWorkedWith']]
data_10_techs_popular_dev_countries.head()
```

[390]:

```
                     Country  \
45                    Brazil
50                    Greece
58        Russian Federation
64  United States of America
76                    Poland

                              LanguageHaveWorkedWith  \
45        C#;C++;JavaScript;PowerShell;SQL;TypeScript
50  C#;HTML/CSS;JavaScript;Node.js;PowerShell;Type…
58          Bash/Shell;HTML/CSS;JavaScript;Python;SQL
64                        HTML/CSS;JavaScript;Python
76  Bash/Shell;C#;Dart;Delphi;Go;HTML/CSS;Java;Jav…

                              DatabaseHaveWorkedWith  \
45              Microsoft SQL Server;PostgreSQL;Redis
50  Couchbase;MariaDB;Microsoft SQL Server;MongoDB…
58                                             Oracle
64                                              MySQL
76  Firebase;Microsoft SQL Server;MongoDB;MySQL;Po…

                  PlatformHaveWorkedWith  \
45                 Heroku;Microsoft Azure
50       AWS;DigitalOcean;Microsoft Azure
58                                 Heroku
64                                    AWS
76  Google Cloud Platform;Microsoft Azure

                               WebframeHaveWorkedWith  \
45                             ASP.NET Core ;React.js
50       Angular;ASP.NET;ASP.NET Core ;Express;Svelte
58                              Django;FastAPI;Vue.js
64                                              Flask
76  Angular;Angular.js;ASP.NET;ASP.NET Core ;Djang…

                               MiscTechHaveWorkedWith ToolsTechHaveWorkedWith  \
45                                 .NET Core / .NET 5   Docker;Git;Kubernetes
50                  .NET Framework;.NET Core / .NET 5       Docker;Kubernetes
58                         NumPy;Pandas;Torch/PyTorch              Docker;Git
64                                             Pandas                     Git
76  .NET Framework;.NET Core / .NET 5;Apache Spark…     Docker;Git;Unity 3D

                         NEWCollabToolsHaveWorkedWith
45         Notepad++;Visual Studio;Visual Studio Code
50         Notepad++;Visual Studio;Visual Studio Code
58                 IPython/Jupyter;Visual Studio Code
64                     Notepad++;PyCharm;Sublime Text
76  Android Studio;Eclipse;NetBeans;Notepad++;Visu…
```

[1029]:

```
data_10_techs_popular_dev_countries['AllTechs'] =␣
↪data_10_techs_popular_dev_countries['LanguageHaveWorkedWith'].map(str) + ';'␣
↪+ data_10_techs_popular_dev_countries['DatabaseHaveWorkedWith'].map(str) + ';
↪' + data_10_techs_popular_dev_countries['PlatformHaveWorkedWith'].map(str) +␣
↪';' + data_10_techs_popular_dev_countries['WebframeHaveWorkedWith'].map(str)␣
↪+ ';' + data_10_techs_popular_dev_countries['MiscTechHaveWorkedWith'].
↪map(str) + ';' +␣
↪data_10_techs_popular_dev_countries['ToolsTechHaveWorkedWith'].map(str) + ';
↪' + data_10_techs_popular_dev_countries['NEWCollabToolsHaveWorkedWith'].
↪map(str)
print (data_10_techs_popular_dev_countries)
```

```
                        Country  \
45                       Brazil
50                       Greece
58           Russian Federation
64     United States of America
76                       Poland
…                          …
83428                    Brazil
83431                  Pakistan
```

```
83432                    Canada
83436  United States of America
83437                    Canada

                                  LanguageHaveWorkedWith  \
45           C#;C++;JavaScript;PowerShell;SQL;TypeScript
50     C#;HTML/CSS;JavaScript;Node.js;PowerShell;Type…
58             Bash/Shell;HTML/CSS;JavaScript;Python;SQL
64                                 HTML/CSS;JavaScript;Python
76     Bash/Shell;C#;Dart;Delphi;Go;HTML/CSS;Java;Jav…
…                                                  …
83428                        Bash/Shell;Node.js;TypeScript
83431  C#;Dart;HTML/CSS;Java;JavaScript;Kotlin;Node.j…
83432                                                Ruby
83436                                  Groovy;Java;Python
83437               Bash/Shell;JavaScript;Node.js;Python

                                  DatabaseHaveWorkedWith  \
45                 Microsoft SQL Server;PostgreSQL;Redis
50     Couchbase;MariaDB;Microsoft SQL Server;MongoDB…
58                                                Oracle
64                                                 MySQL
76     Firebase;Microsoft SQL Server;MongoDB;MySQL;Po…
…                                                  …
83428             Elasticsearch;MongoDB;PostgreSQL;Redis
83431                              Firebase;MySQL;SQLite
83432                                   MySQL;PostgreSQL
83436    DynamoDB;Elasticsearch;MongoDB;PostgreSQL;Redis
83437   Cassandra;Elasticsearch;MongoDB;PostgreSQL;Redis

                      PlatformHaveWorkedWith  \
45                      Heroku;Microsoft Azure
50          AWS;DigitalOcean;Microsoft Azure
58                                      Heroku
64                                         AWS
76       Google Cloud Platform;Microsoft Azure
…                                      …
83428              AWS;Google Cloud Platform
83431                  Google Cloud Platform
83432           Google Cloud Platform;Heroku
83436              AWS;Google Cloud Platform
83437                                 Heroku

                                  WebframeHaveWorkedWith  \
45                                  ASP.NET Core ;React.js
50          Angular;ASP.NET;ASP.NET Core ;Express;Svelte
58                                 Django;FastAPI;Vue.js
64                                                 Flask
```

```
76     Angular;Angular.js;ASP.NET;ASP.NET Core ;Djang…
…                                                  …
83428                                          React.js
83431                                      Flask;jQuery
83432               Flask;React.js;Ruby on Rails;Vue.js
83436                                     FastAPI;Flask
83437                     Django;Express;Flask;React.js

                                 MiscTechHaveWorkedWith  \
45                                   .NET Core / .NET 5
50                    .NET Framework;.NET Core / .NET 5
58                           NumPy;Pandas;Torch/PyTorch
64                                               Pandas
76     .NET Framework;.NET Core / .NET 5;Apache Spark…
…                                                  …
83428                                      React Native
83431                                           Flutter
83432             NumPy;Pandas;TensorFlow;Torch/PyTorch
83436                         Hadoop;Keras;NumPy;Pandas
83437             NumPy;Pandas;TensorFlow;Torch/PyTorch

            ToolsTechHaveWorkedWith  \
45            Docker;Git;Kubernetes
50                Docker;Kubernetes
58                       Docker;Git
64                              Git
76              Docker;Git;Unity 3D
…                             …
83428     Docker;Git;Terraform;Yarn
83431                           Git
83432    Docker;Git;Kubernetes;Yarn
83436  Ansible;Docker;Git;Terraform
83437  Ansible;Docker;Git;Terraform

                           NEWCollabToolsHaveWorkedWith  \
45            Notepad++;Visual Studio;Visual Studio Code
50            Notepad++;Visual Studio;Visual Studio Code
58                    IPython/Jupyter;Visual Studio Code
64                        Notepad++;PyCharm;Sublime Text
76     Android Studio;Eclipse;NetBeans;Notepad++;Visu…
…                                                  …
83428                          Visual Studio Code;Webstorm
83431  Android Studio;IntelliJ;IPython/Jupyter;Notepa…
83432        Atom;IPython/Jupyter;Vim;Visual Studio Code
83436  Android Studio;Eclipse;IntelliJ;IPython/Jupyte…
83437                               PyCharm;Sublime Text

                                               AllTechs
```

```
45     C#;C++;JavaScript;PowerShell;SQL;TypeScript;Mi…
50     C#;HTML/CSS;JavaScript;Node.js;PowerShell;Type…
58     Bash/Shell;HTML/CSS;JavaScript;Python;SQL;Orac…
64     HTML/CSS;JavaScript;Python;MySQL;AWS;Flask;Pan…
76     Bash/Shell;C#;Dart;Delphi;Go;HTML/CSS;Java;Jav…
…                                                   …
83428  Bash/Shell;Node.js;TypeScript;Elasticsearch;Mo…
83431  C#;Dart;HTML/CSS;Java;JavaScript;Kotlin;Node.j…
83432  Ruby;MySQL;PostgreSQL;Google Cloud Platform;He…
83436  Groovy;Java;Python;DynamoDB;Elasticsearch;Mong…
83437  Bash/Shell;JavaScript;Node.js;Python;Cassandra…

[14517 rows x 9 columns]
```

```
C:\Users\GPBONI~1\AppData\Local\Temp/ipykernel_9952/1489135702.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  data_10_techs_popular_dev_countries['AllTechs'] =
data_10_techs_popular_dev_countries['LanguageHaveWorkedWith'].map(str) + ';' +
data_10_techs_popular_dev_countries['DatabaseHaveWorkedWith'].map(str) + ';' +
data_10_techs_popular_dev_countries['PlatformHaveWorkedWith'].map(str) + ';' +
data_10_techs_popular_dev_countries['WebframeHaveWorkedWith'].map(str) + ';' +
data_10_techs_popular_dev_countries['MiscTechHaveWorkedWith'].map(str) + ';' +
data_10_techs_popular_dev_countries['ToolsTechHaveWorkedWith'].map(str) + ';' +
data_10_techs_popular_dev_countries['NEWCollabToolsHaveWorkedWith'].map(str)
```

[1030]:

```
df_data_10_techs_popular_dev_countries =␣
 ↪data_10_techs_popular_dev_countries[['Country', 'AllTechs']].copy()
```

[1031]:

```
df_data_10_techs_popular_dev_countries.head()
```

[1031]:

```
                     Country  \
45                    Brazil
50                    Greece
58        Russian Federation
64  United States of America
76                    Poland

                                           AllTechs
45  C#;C++;JavaScript;PowerShell;SQL;TypeScript;Mi…
50  C#;HTML/CSS;JavaScript;Node.js;PowerShell;Type…
58  Bash/Shell;HTML/CSS;JavaScript;Python;SQL;Orac…
64  HTML/CSS;JavaScript;Python;MySQL;AWS;Flask;Pan…
76  Bash/Shell;C#;Dart;Delphi;Go;HTML/CSS;Java;Jav…
```

```
[1032]: df_data_10_techs_popular_dev_countries['AllTechs'] =␣
         ↪df_data_10_techs_popular_dev_countries['AllTechs'].str.replace(' ', '')
```

```
[1033]: df_data_10_techs_popular_dev_countries['AllTechs'] =␣
         ↪df_data_10_techs_popular_dev_countries['AllTechs'].str.replace(';', ' ')
```

```
[1034]: df_counts = df_data_10_techs_popular_dev_countries['AllTechs'].str.
         ↪split(expand=True).stack().value_counts().rename_axis('Tech').
         ↪reset_index(name='Count')
```

```
[1035]: df_counts
```

```
[1035]:                  Tech  Count
        0                 Git  13828
        1      VisualStudioCode  12030
        2          JavaScript  11779
        3            HTML/CSS   9714
        4              Docker   9296
        ..                ...    ...
        120            Erlang    128
        121            Pulumi    121
        122             COBOL     91
        123           Crystal     87
        124               APL     45

        [125 rows x 2 columns]
```

```
 [391]: data_10_techs_popular_dev_countries.
         ↪to_csv('012_data_10_techs_popular_dev_countries.csv', index=False)
```

3.0.13 2.13. ¿Cuáles el sistema operativo más usado entre los encuestados?

Se seleccionarán los campos adecuados para responder a esta pregunta

```
[1036]: df_data_so_devs = data_test[['OpSys']].copy()
```

```
[1038]: df_data_so_devs.tail()
```

```
[1038]:          OpSys
        83428    MacOS
        83431  Windows
        83432    MacOS
        83436  Windows
        83437    MacOS
```

```
[1039]: df_data_so_devs['OpSys'].drop_duplicates().sort_values()
```

```
[1039]: 7037                          BSD
        58                          Linux
        77                          MacOS
        464      Other (please specify):
        45                        Windows
        Name: OpSys, dtype: object
```

```
[1042]: df_data_so_devs['OpSys'] = df_data_so_devs['OpSys'].replace(['Other (please␣
         ↪specify):'], 'Otro')
```

```
[1043]: df_data_so_devs['OpSys'].value_counts()
```

```
[1043]: Windows    7293
        MacOS      4255
        Linux      2912
        Otro         47
        BSD          10
        Name: OpSys, dtype: int64
```

```
[1045]: df_counts = df_data_so_devs['OpSys'].str.split(expand=True).stack().
         ↪value_counts().rename_axis('OS').reset_index(name='Count')
```

```
[1046]: df_counts
```

```
[1046]:         OS  Count
        0  Windows   7293
        1    MacOS   4255
        2    Linux   2912
        3     Otro     47
        4      BSD     10
```

```
[1047]: df_counts.to_csv('013_flourish_data.csv', index=False)
```

3.0.14 2.14. ¿Qué proporción de programadores tiene algún desorden mental por país?

Se seleccionarán los campos adecuados para responder a esta pregunta

```
 [394]: data_devs_mental_health_countries = data_test[['Country', 'MentalHealth']]
        data_devs_mental_health_countries.head()
```

```
 [394]:                      Country                  MentalHealth
        45                    Brazil            Desorden emocional
        50                    Greece  Ninguna de las mencionadas
        58        Russian Federation  Ninguna de las mencionadas
        64  United States of America  Ninguna de las mencionadas
        76                    Poland  Ninguna de las mencionadas
```

```
[1048]: data_devs_mental_health_countries['MentalHealth'].value_counts()
```

```
[1048]: Ninguna de las mencionadas              10924
        Desorden de concentración o memoria     1303
        Desorden emocional                       908
        Desorden de ansiedad                     637
        No definido                              538
        Tipo de autismo                          207
        Name: MentalHealth, dtype: int64
```

```
[1100]: df_data_devs_mental_health_countries = data_devs_mental_health_countries.copy()
```

```
[1101]: df_data_devs_mental_health_countries =␣
         ↪df_data_devs_mental_health_countries[df_data_devs_mental_health_countries['MentalHealth'].
         ↪isin(['Desorden de concentración o memoria', 'Desorden emocional', 'Desorden␣
         ↪de ansiedad', 'Tipo de autismo'])]
```

```
[1103]: df_data_devs_mental_health_countries.head()
```

```
[1103]:                       Country                         MentalHealth
        45                     Brazil                   Desorden emocional
        96                    Germany                   Desorden emocional
        129  United States of America                      Tipo de autismo
        199  United States of America  Desorden de concentración o memoria
        213        Russian Federation                 Desorden de ansiedad
```

```
[1091]: df_data_flourish_014 = df_data_devs_mental_health_countries['Country'].
         ↪value_counts().to_frame('counts').reset_index()
```

```
[1095]: df_data_flourish_014 = df_data_flourish_014.head(10)
        df_data_flourish_014
```

```
[1095]:                                              index  counts
        0                         United States of America    1027
        1  United Kingdom of Great Britain and Northern I…     206
        2                                           Brazil     194
        3                                           Canada     140
        4                                            India     134
        5                                          Germany     112
        6                                        Australia      85
        7                                      Netherlands      84
        8                                           Poland      57
        9                                           Turkey      52
```

```
[1111]:
```

```
df_data_flourish_014_best_ten =␣
 ↪df_data_devs_mental_health_countries[df_data_devs_mental_health_countries['Country'].
 ↪isin(['United States of America', 'United Kingdom of Great Britain and␣
 ↪Northern Ireland', 'Brazil', 'Canada', 'India', 'Germany', 'Australia',␣
 ↪'Netherlands', 'Poland', 'Turkey'])]
```

[1135]:

```
df = df_data_flourish_014_best_ten.copy()
```

[1136]:

```
df
```

[1136]:

```
                        Country                          MentalHealth
45                       Brazil                    Desorden emocional
96                      Germany                    Desorden emocional
129    United States of America                      Tipo de autismo
199    United States of America  Desorden de concentración o memoria
237                     Germany                    Desorden emocional
…                           …                                    …
83319                   Germany  Desorden de concentración o memoria
83342  United States of America                    Desorden emocional
83347                    Brazil                  Desorden de ansiedad
83370                    Brazil  Desorden de concentración o memoria
83437                    Canada                    Desorden emocional

[2091 rows x 2 columns]
```

[1138]:

```
df1 = pd.crosstab(df['Country'], df['MentalHealth'])
df1
```

[1138]:

```
MentalHealth                                       Desorden de concentración o
memoria  \
Country
Australia
36
Brazil
67
Canada
71
Germany
38
India
42
Netherlands
42
Poland
14
Turkey
19
```

```
United Kingdom of Great Britain and Northern Ir…
64
United States of America
573

MentalHealth                                        Desorden emocional  \
Country
Australia                                                           25
Brazil                                                              52
Canada                                                              30
Germany                                                             43
India                                                               51
Netherlands                                                         15
Poland                                                              26
Turkey                                                              19
United Kingdom of Great Britain and Northern Ir…                  77
United States of America                                           261

MentalHealth                                        Desorden de ansiedad  \
Country
Australia                                                             16
Brazil                                                                64
Canada                                                                30
Germany                                                               21
India                                                                 38
Netherlands                                                           11
Poland                                                                 7
Turkey                                                                14
United Kingdom of Great Britain and Northern Ir…                    42
United States of America                                             144

MentalHealth                                        Tipo de autismo
Country
Australia                                                         8
Brazil                                                           11
Canada                                                            9
Germany                                                          10
India                                                             3
Netherlands                                                      16
Poland                                                           10
Turkey                                                            0
United Kingdom of Great Britain and Northern Ir…               23
United States of America                                         49
```

```
[1066]: (df_data_devs_mental_health_countries.groupby(['Country', 'MentalHealth']).
         ↪size()
           .sort_values(ascending=False)
```

```
    .reset_index(name='count')
    .drop_duplicates(subset='Country'))
```

```
[1066]:
                                               Country  \
0                             United States of America
3    United Kingdom of Great Britain and Northern I…
4                                               Canada
5                                               Brazil
9                                                India
..                                                  …
295                                             Kuwait
298                                         Luxembourg
299                                             Malawi
300                                           Maldives
301                                             Zambia

                            MentalHealth  count
0    Desorden de concentración o memoria    573
3                     Desorden emocional     77
4    Desorden de concentración o memoria     71
5    Desorden de concentración o memoria     67
9                     Desorden emocional     51
..                                    …    …
295  Desorden de concentración o memoria      1
298                 Desorden de ansiedad      1
299  Desorden de concentración o memoria      1
300                   Desorden emocional      1
301  Desorden de concentración o memoria      1

[122 rows x 3 columns]
```

```
[1074]: df_flourish_data_014 = (df_data_devs_mental_health_countries.
         ↪groupby(['Country', 'MentalHealth']).size()
          .sort_values(ascending=False)
          .reset_index(name='count'))
```

```
[1077]: df_flourish_data_014 = df_flourish_data_014.sort_values('Country')
```

```
[1094]: df_data_flourish_014.head(10).to_csv('014_flourish_data_014.csv', index=False)
```

```
[1140]: df1.to_csv('014_flourish_data_014.csv', index=True)
```

3.0.15 2.15. ¿Cuáles son los países que tienen los mejores sueldos entre los programadores?

Se seleccionarán los campos adecuados para responder a esta pregunta

```
[1141]: df_best_incomes_countries = data_test[['Country', 'ConvertedCompYearly']].copy()
```

[1142]:

```
df_best_incomes_countries
```

[1142]:

```
                        Country  ConvertedCompYearly
45                       Brazil              60480.0
50                       Greece              25944.0
58           Russian Federation              22644.0
64     United States of America             500000.0
76                       Poland              45564.0
…                           …                  …
83428                    Brazil              41232.0
83431                  Pakistan              11676.0
83432                    Canada              80169.0
83436  United States of America              90000.0
83437                    Canada             816816.0

[14517 rows x 2 columns]
```

[1143]:

```
def remove_outliers(df, q=0.05):
    upper = df.quantile(1-q)
    lower = df.quantile(q)
    mask = (df < upper) & (df > lower)
    return mask

mask = remove_outliers(df_best_incomes_countries['ConvertedCompYearly'], 0.1)

print(df_best_incomes_countries[mask])
```

```
                        Country  ConvertedCompYearly
45                       Brazil              60480.0
50                       Greece              25944.0
58           Russian Federation              22644.0
76                       Poland              45564.0
77                       Canada             151263.0
…                           …                  …
83425                   Finland              19452.0
83428                    Brazil              41232.0
83431                  Pakistan              11676.0
83432                    Canada              80169.0
83436  United States of America              90000.0

[11611 rows x 2 columns]
```

[1145]:

```
df_best_incomes_countries_no_outliers = df_best_incomes_countries[mask]
```

[1146]:

```
df_best_incomes_countries_no_outliers1 = df_best_incomes_countries_no_outliers.
 ↪copy()
```

```
[1148]: df_best_incomes_countries_no_outliers1['ConvertedCompYearlyCategorical'] =␣
         ↪'ALTO'
        df_best_incomes_countries_no_outliers1.
         ↪loc[(df_best_incomes_countries_no_outliers1['ConvertedCompYearly'] >= 0) &␣
         ↪(df_best_incomes_countries_no_outliers1['ConvertedCompYearly'] <= 32747),␣
         ↪'ConvertedCompYearlyCategorical'] = 'BAJO'
        df_best_incomes_countries_no_outliers1.
         ↪loc[(df_best_incomes_countries_no_outliers1['ConvertedCompYearly'] > 32747)␣
         ↪& (df_best_incomes_countries_no_outliers1['ConvertedCompYearly'] <= 90000),␣
         ↪'ConvertedCompYearlyCategorical'] = 'MEDIO'

        print(df_best_incomes_countries_no_outliers1)
```

```
                        Country  ConvertedCompYearly  \
45                       Brazil              60480.0
50                       Greece              25944.0
58           Russian Federation              22644.0
76                       Poland              45564.0
77                       Canada             151263.0
…                           …                  …
83425                   Finland              19452.0
83428                    Brazil              41232.0
83431                  Pakistan              11676.0
83432                    Canada              80169.0
83436  United States of America              90000.0

      ConvertedCompYearlyCategorical
45                             MEDIO
50                              BAJO
58                              BAJO
76                             MEDIO
77                              ALTO
…                             …
83425                           BAJO
83428                          MEDIO
83431                           BAJO
83432                          MEDIO
83436                          MEDIO

[11611 rows x 3 columns]
```

```
[1149]: df_best_incomes_countries_no_outliers1['ConvertedCompYearlyCategorical'].
         ↪value_counts()
```

```
[1149]: MEDIO    5816
        BAJO     2903
        ALTO     2892
```

```
Name: ConvertedCompYearlyCategorical, dtype: int64
```

[1151]:
```
df_best_incomes_countries_alto =␣
 ↪df_best_incomes_countries_no_outliers1[df_best_incomes_countries_no_outliers1['ConvertedCom
 ↪== 'ALTO']
```

[1152]:
```
df_alto = df_best_incomes_countries_alto[['Country',␣
 ↪'ConvertedCompYearlyCategorical']].copy()
```

[1154]:
```
df_flourish_015 = df_alto['Country'].value_counts().to_frame('counts').
 ↪reset_index()
```

[1156]:
```
df_flourish_015.head(10)
```

[1156]:
```
                                                  index  counts
0                              United States of America    1547
1  United Kingdom of Great Britain and Northern I…     244
2                                                Canada     166
3                                               Germany     107
4                                             Australia     106
5                                                Israel      82
6                                           Switzerland      81
7                                               Denmark      57
8                                           Netherlands      40
9                                                France      36
```

[1157]:
```
df_flourish_015.head(10).to_csv('015_flourish_data.csv', index=False)
```

3.0.16 2.16. ¿Cuáles son los 10 lenguajes de programación más usados entre los programadores?

Se seleccionarán los campos adecuados para responder a esta pregunta

[1161]:
```
df_10_prog_languages_devs = data_test[['LanguageHaveWorkedWith']].copy()
df_10_prog_languages_devs.head()
```

[1161]:
```
                          LanguageHaveWorkedWith
45        C#;C++;JavaScript;PowerShell;SQL;TypeScript
50   C#;HTML/CSS;JavaScript;Node.js;PowerShell;Type…
58          Bash/Shell;HTML/CSS;JavaScript;Python;SQL
64                         HTML/CSS;JavaScript;Python
76   Bash/Shell;C#;Dart;Delphi;Go;HTML/CSS;Java;Jav…
```

[1162]:
```
df_10_prog_languages_devs['LanguageHaveWorkedWith'] =␣
 ↪df_10_prog_languages_devs['LanguageHaveWorkedWith'].str.replace(';', ' ')
```

[1163]:

```
df_counts_016 = df_10_prog_languages_devs['LanguageHaveWorkedWith'].str.
 ↪split(expand=True).stack().value_counts().rename_axis('Languages').
 ↪reset_index(name='Count')
```

```
[1164]: df_counts_016.head(10)
```

```
[1164]:     Languages  Count
        0  JavaScript  11779
        1    HTML/CSS   9714
        2         SQL   9294
        3          C#   7318
        4  TypeScript   7261
        5      Python   7225
        6     Node.js   7066
        7        Java   4855
        8  Bash/Shell   4574
        9         PHP   3524
```

```
[1165]: df_counts_016.head(10).to_csv('016_flourish_data.csv', index=False)
```

3.0.17 2.17. ¿Cuáles son las bases de datos más usadas entre los programadores?

Se seleccionarán los campos adecuados para responder a esta pregunta

```
[1171]: df_10_databases = data_test[['DatabaseHaveWorkedWith']].copy()
        df_10_databases.head()
```

```
[1171]:                               DatabaseHaveWorkedWith
        45              Microsoft SQL Server;PostgreSQL;Redis
        50  Couchbase;MariaDB;Microsoft SQL Server;MongoDB…
        58                                             Oracle
        64                                              MySQL
        76  Firebase;Microsoft SQL Server;MongoDB;MySQL;Po…
```

```
[1172]: df_10_databases['DatabaseHaveWorkedWith'] =␣
         ↪df_10_databases['DatabaseHaveWorkedWith'].str.replace(' ', '')
```

```
[1173]: df_10_databases['DatabaseHaveWorkedWith'] =␣
         ↪df_10_databases['DatabaseHaveWorkedWith'].str.replace(';', ' ')
```

```
[1174]: df_counts_017 = df_10_databases['DatabaseHaveWorkedWith'].str.
         ↪split(expand=True).stack().value_counts().rename_axis('Databases').
         ↪reset_index(name='Count')
```

```
[1175]: df_counts_017.head(10)
```

```
[1175]:             Databases  Count
        0          PostgreSQL   7163
```

```
1               MySQL   7150
2  MicrosoftSQLServer   6553
3              SQLite   5442
4             MongoDB   5107
5               Redis   4507
6            Firebase   3032
7       Elasticsearch   2890
8             MariaDB   2704
9              Oracle   1921
```

```
[1176]: df_counts_017.head(10).to_csv('017_flourish_data.csv', index=False)
```

3.0.18 2.18. ¿Cuáles son las plataformas más usadas entre los programadores?

Se seleccionarán los campos adecuados para responder a esta pregunta

```
[1177]: df_10_platforms = data_test[['PlatformHaveWorkedWith']].copy()
        df_10_platforms.head()
```

```
[1177]:                   PlatformHaveWorkedWith
        45                Heroku;Microsoft Azure
        50       AWS;DigitalOcean;Microsoft Azure
        58                                Heroku
        64                                   AWS
        76  Google Cloud Platform;Microsoft Azure
```

```
[1178]: df_10_platforms['PlatformHaveWorkedWith'] =␣
         ↪df_10_platforms['PlatformHaveWorkedWith'].str.replace(' ', '')
```

```
[1179]: df_10_platforms['PlatformHaveWorkedWith'] =␣
         ↪df_10_platforms['PlatformHaveWorkedWith'].str.replace(';', ' ')
```

```
[1181]: df_counts_018 = df_10_platforms['PlatformHaveWorkedWith'].str.
         ↪split(expand=True).stack().value_counts().rename_axis('Platform').
         ↪reset_index(name='Count')
```

```
[1182]: df_counts_018.head(10)
```

```
[1182]:                   Platform  Count
        0                      AWS   8348
        1           MicrosoftAzure   6738
        2      GoogleCloudPlatform   4710
        3                   Heroku   3182
        4             DigitalOcean   2829
        5         IBMCloudorWatson    350
        6  OracleCloudInfrastructure  212
```

```
[1183]: df_counts_018.to_csv('018_flourish_data.csv', index=False)
```

3.0.19 2.19. ¿Cuáles son los frameworks web más usados entre los programadores?

Se seleccionarán los campos adecuados para responder a esta pregunta

[1185]:
```
df_10_web_frameworks = data_test[['WebframeHaveWorkedWith']].copy()
df_10_web_frameworks.head()
```

[1185]:
```
                               WebframeHaveWorkedWith
45                               ASP.NET Core ;React.js
50         Angular;ASP.NET;ASP.NET Core ;Express;Svelte
58                                Django;FastAPI;Vue.js
64                                                Flask
76  Angular;Angular.js;ASP.NET;ASP.NET Core ;Djang…
```

[1186]:
```
df_10_web_frameworks['WebframeHaveWorkedWith'] =␣
 ↪df_10_web_frameworks['WebframeHaveWorkedWith'].str.replace(' ', '')
```

[1187]:
```
df_10_web_frameworks['WebframeHaveWorkedWith'] =␣
 ↪df_10_web_frameworks['WebframeHaveWorkedWith'].str.replace(';', ' ')
```

[1188]:
```
df_counts_019 = df_10_web_frameworks['WebframeHaveWorkedWith'].str.
 ↪split(expand=True).stack().value_counts().rename_axis('Web framework').
 ↪reset_index(name='Count')
```

[1189]:
```
df_counts_019.head(10)
```

[1189]:
```
  Web framework  Count
0      React.js   6745
1        jQuery   5391
2    ASP.NETCore   5304
3       Angular   4506
4       ASP.NET   4169
5       Express   4048
6        Vue.js   3141
7         Flask   2873
8    Angular.js   2349
9        Django   2273
```

[1190]:
```
df_counts_019.to_csv('019_flourish_data.csv', index=False)
```

3.0.20 2.20. ¿Cuáles son las herramientas tecnológicas más usadas entre los programadores?

Se seleccionarán los campos adecuados para responder a esta pregunta

[1192]:
```
df_10_data_misc_techs = data_test[['MiscTechHaveWorkedWith',␣
 ↪'ToolsTechHaveWorkedWith']].copy()
df_10_data_misc_techs.head()
```

```
[1192]:                              MiscTechHaveWorkedWith ToolsTechHaveWorkedWith
        45                               .NET Core / .NET 5   Docker;Git;Kubernetes
        50                .NET Framework;.NET Core / .NET 5       Docker;Kubernetes
        58                       NumPy;Pandas;Torch/PyTorch              Docker;Git
        64                                           Pandas                     Git
        76  .NET Framework;.NET Core / .NET 5;Apache Spark…     Docker;Git;Unity 3D
```

```
[1193]: df_10_data_misc_techs['AllMiscTechs'] =␣
         ↪df_10_data_misc_techs['MiscTechHaveWorkedWith'].map(str) + ';' +␣
         ↪df_10_data_misc_techs['ToolsTechHaveWorkedWith'].map(str)
```

```
[1194]: df_10_data_misc_techs.head()
```

```
[1194]:                              MiscTechHaveWorkedWith ToolsTechHaveWorkedWith  \
        45                               .NET Core / .NET 5   Docker;Git;Kubernetes
        50                .NET Framework;.NET Core / .NET 5       Docker;Kubernetes
        58                       NumPy;Pandas;Torch/PyTorch              Docker;Git
        64                                           Pandas                     Git
        76  .NET Framework;.NET Core / .NET 5;Apache Spark…     Docker;Git;Unity 3D

                                               AllMiscTechs
        45           .NET Core / .NET 5;Docker;Git;Kubernetes
        50  .NET Framework;.NET Core / .NET 5;Docker;Kuber…
        58              NumPy;Pandas;Torch/PyTorch;Docker;Git
        64                                       Pandas;Git
        76  .NET Framework;.NET Core / .NET 5;Apache Spark…
```

```
[1197]: df_10_data_misc_techs['AllMiscTechs'] = df_10_data_misc_techs['AllMiscTechs'].
         ↪str.replace(' ', '')
```

```
[1198]: df_10_data_misc_techs['AllMiscTechs'] = df_10_data_misc_techs['AllMiscTechs'].
         ↪str.replace(';', ' ')
```

```
[1200]: df_counts_020 = df_10_data_misc_techs['AllMiscTechs'].str.split(expand=True).
         ↪stack().value_counts().rename_axis('Tecnología').reset_index(name='#␣
         ↪Programadores')
```

```
[1201]: df_counts_020.head(10)
```

```
[1201]:       Tecnología  # Programadores
        0            Git            13828
        1         Docker             9296
        2   .NETCore/.NET5           6046
        3    .NETFramework           5697
        4          NumPy             3807
        5     Kubernetes             3709
        6         Pandas             3634
        7           Yarn             3617
```

```
8     ReactNative             2960
9      TensorFlow             2068
```

```
[1202]: df_counts_020.head(10).to_csv('020_flourish_data.csv', index=False)
```

3.0.21 2.21. ¿Cuáles son las herramientas colaborativas más usadas entre programadores?

Se seleccionarán los campos adecuados para responder a esta pregunta

```
[1204]: df_10_colab = data_test[['NEWCollabToolsHaveWorkedWith']].copy()
        df_10_colab.head()
```

```
[1204]:                       NEWCollabToolsHaveWorkedWith
        45         Notepad++;Visual Studio;Visual Studio Code
        50         Notepad++;Visual Studio;Visual Studio Code
        58                 IPython/Jupyter;Visual Studio Code
        64                     Notepad++;PyCharm;Sublime Text
        76  Android Studio;Eclipse;NetBeans;Notepad++;Visu…
```

```
[1205]: df_10_colab['NEWCollabToolsHaveWorkedWith'] =␣
         ↪df_10_colab['NEWCollabToolsHaveWorkedWith'].str.replace(' ', '')
```

```
[1206]: df_10_colab['NEWCollabToolsHaveWorkedWith'] =␣
         ↪df_10_colab['NEWCollabToolsHaveWorkedWith'].str.replace(';', ' ')
```

```
[1207]: df_counts_021 = df_10_colab['NEWCollabToolsHaveWorkedWith'].str.
         ↪split(expand=True).stack().value_counts().rename_axis('Herramienta␣
         ↪Colaborativa').reset_index(name='# Programadores')
```

```
[1208]: df_counts_021.head(10)
```

```
[1208]:   Herramienta Colaborativa  # Programadores
        0         VisualStudioCode            12030
        1             VisualStudio             7183
        2                Notepad++             4987
        3            AndroidStudio             4291
        4                 IntelliJ             4242
        5                      Vim             3773
        6             SublimeText              3080
        7                  PyCharm             3024
        8                    Xcode             2602
        9                  Eclipse             2176
```

```
[1209]: df_counts_021.head(10).to_csv('021_flourish_data.csv', index=False)
```

3.0.22 2.22. ¿Cuáles son los países con mayor número de programadores trabajando a tiempo completo?

Se seleccionarán los campos adecuados para responder a esta pregunta

```
[1210]: df_fulltime_employment = data_test[['Country', 'Employment']].copy()
        df_fulltime_employment.head()
```

```
[1210]:                      Country       Employment
        45                    Brazil  Tiempo completo
        50                    Greece  Tiempo completo
        58        Russian Federation  Tiempo completo
        64  United States of America     Independiete
        76                    Poland  Tiempo completo
```

```
[1213]: df_fulltime_employment.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 14517 entries, 45 to 83437
Data columns (total 2 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Country     14517 non-null  object
 1   Employment  14517 non-null  object
dtypes: object(2)
memory usage: 856.3+ KB
```

```
[1211]: df_fulltime_only = df_fulltime_employment[df_fulltime_employment['Employment']␣
         ↪== 'Tiempo completo']
```

```
[1212]: df_fulltime_only.head()
```

```
[1212]:                Country       Employment
        45              Brazil  Tiempo completo
        50              Greece  Tiempo completo
        58  Russian Federation  Tiempo completo
        76              Poland  Tiempo completo
        77              Canada  Tiempo completo
```

```
[1215]: df_flourish_022 = df_fulltime_only['Country'].value_counts().to_frame('#␣
         ↪Programadores').reset_index()
```

```
[1216]: df_flourish_022.head(10)
```

```
[1216]:                                                 index  # Programadores
        0                            United States of America             2947
        1                                               India              984
        2  United Kingdom of Great Britain and Northern I…             859
        3                                             Germany              611
```

```
4                                              Brazil              502
5                                              Canada              499
6                                           Australia              308
7                                              France              303
8                                               Spain              279
9                                         Netherlands              258
```

```
[1218]: df_flourish_022.head(10).to_csv('022_flourish_data.csv', index=False)
```