

Taller de Data Analytics

Código: <https://github.com/Yen-US/Data-Analytics-in-Python>

1. Instalar Anaconda

- a) Descargar el instalador de Anaconda:
<https://www.anaconda.com/products/individual#windows>
- b) Abrir el instalador.
- c) Click en "Siguiente".
- d) Seleccionar "Solo para mí" y click en "Siguiente".
- e) Seleccionar la carpeta de instalación donde se guardará en Anaconda y click en "Siguiente".
- f) Click en "Instalar".
- g) Click en "Siguiente".
- h) Opcional: Instalar PyCharm para Anaconda y click en "Siguiente".
- i) Para verificar la instalación escribir "Anaconda Navigator" en el menú de inicio de Windows.

2. Instalar *Pandas Profiling*

- a) Abrir la consola de Anaconda.
- b) Ejecutar el comando: `conda install -c conda-forge pandas-profiling`

3. Realizar los *imports*

- a)

```
1 #Data Analysis imports
2 import pandas as pd
3 import pandas_profiling as pp
```
- b)

```
5 # Visualization Imports
6 import seaborn as sns
7 color = sns.color_palette()
8 import plotly.offline as py
9 py.init_notebook_mode(connected=True)
10 import plotly.graph_objs as go
11 import plotly.express as px
```

4. Conseguir el dataset

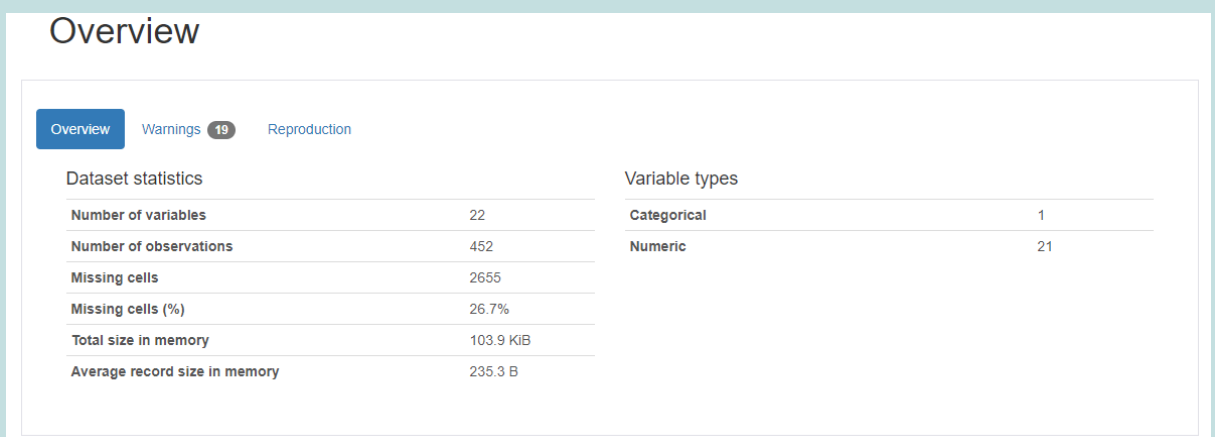
- a) Los datasets pueden obtenerse de múltiples sitios, hay que buscar en páginas de confianza y en formatos compatibles. Por ejemplo CSV. Ejemplo de página: <https://www.stats.govt.nz/large-datasets/csv-files-for-download/>

5. Crear el reporte con Pandas

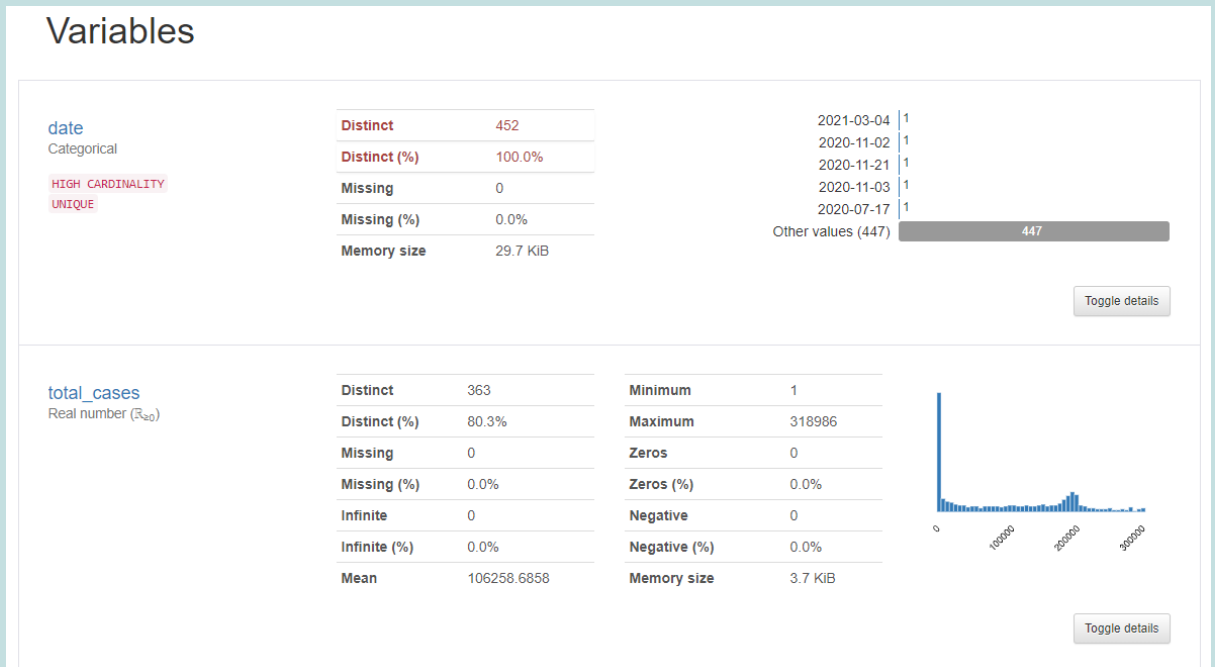
```
13 #Data analysis report creation
14 df = pd.read_csv('DataSet.csv') #Import the CSV dataset file
15 #With pandas profiling creete the Data report
16 profile = pp.ProfileReport(df, title="Pandas Profiling Report", minimal=True, explorative=True)
17 profile.to_file("CRCovidReport.html") #Save the date report in a HTML file
```

a)

6. Identificar la información y variables del reporte



a)



b)

7. Asignar la data a los ejes de plotly

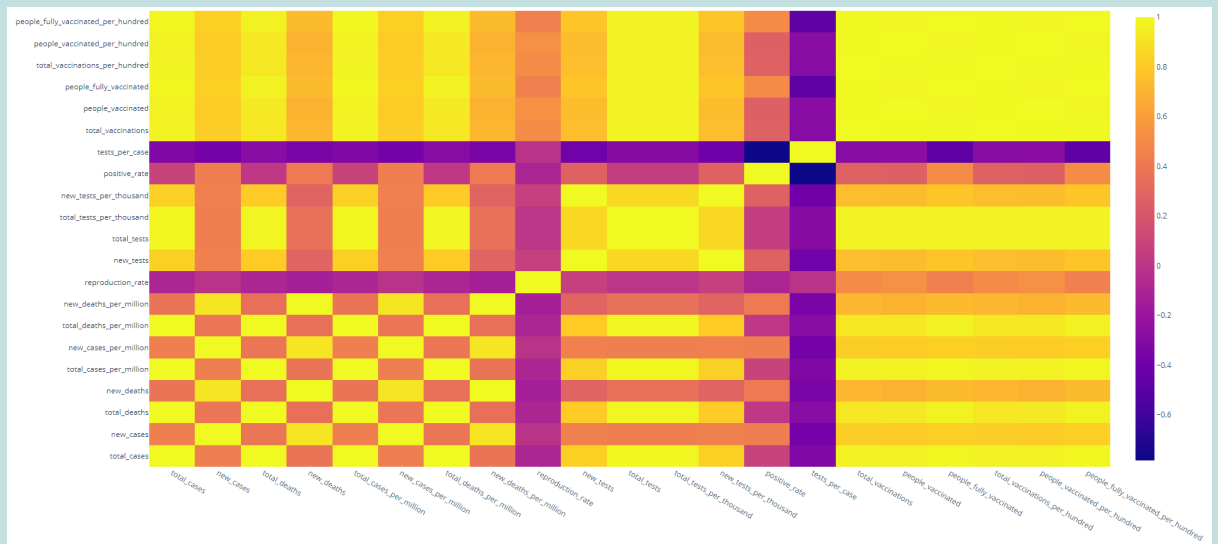
```
19 #From the analysis report, creation of the variables
20 def df_to_plotly(df):
21     #Take the dataset information and assign it to the Plotly axes
22     return {'z': df.values.tolist(),
23            'x': df.columns.tolist(),
24            'y': df.index.tolist() }
```

a)

8. Crear el gráfico de correlación

```
26 #Creation of the correlation matrix graphic with pandas
27 dfNew = df.corr()
28 fig1 = go.Figure(data=go.Heatmap(df_to_plotly(dfNew)))
29 with open('CorrelationMatrix.html', 'w') as f1: #Exporting the graphic to a HTML
30     print(fig1.to_html(), file=f1)
```

a)

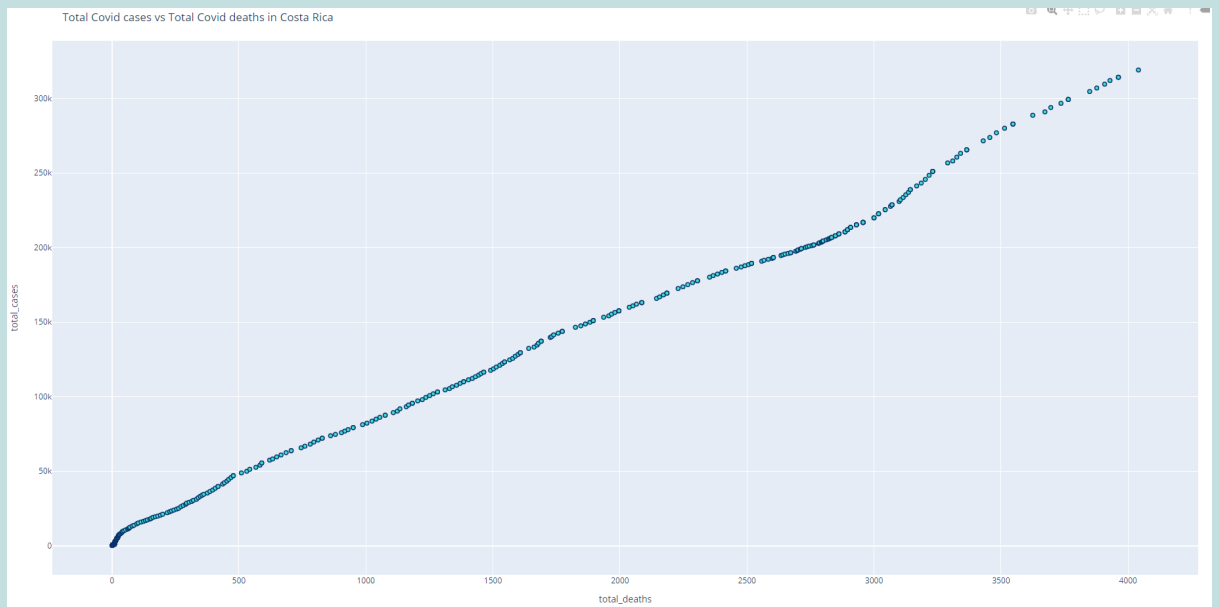


b)

9. Creación de un gráfico de puntos con dos variables

```
32 #Dot graphic creation total_deaths vs total_cases
33 fig2 = px.scatter(df, x='total_deaths', y='total_cases')
34 fig2.update_traces(marker_color='turquoise', marker_line_color='rgb(8,48,107)',
35                    marker_line_width=1.5)
36 fig2.update_layout(title_text='Total Covid cases vs Total Covid deaths in Costa Rica')
37 with open('DeathsVsCases.html', 'w') as f2:
38     print(fig2.to_html(), file=f2)
```

a)



b)