23-3-2022

Manual

Edgar Asael Martínez

Tania Sayuri Guizado Hernández | A01640092
Ricardo Jiménez Ureña | A01636825 Carlos
David Blanco | A01633323
Rafael Jiménez | A01637850
Carlos Estrada Ceballos | A01638214

DIGITAL TOOLS: THE ART OF ANALYTICS – TC1002S.100

For this exercise, we chose the statistics from the radar on COVID tests.

With the pandas library, read the CSV file and convert the Federal data to float.

```
In [1]: import pandas as pd
        #import numpy as np
        from matplotlib import pyplot as plt
        #import random
```

```
In [2]: df = pd.read_csv("01.01.2022_da_radar_pruebas_covid19.csv")
        del df['Unnamed: 6']
```

```
In [3]: df['Federal'] = df['Federal'].str.replace(',','')
```

```
In [4]: df["Federal"] = df.Federal.astype(float)
```

In the next line, we use the head function to view the first 10 data

```
In [5]: df.head(10)
```

Out[5]:

	Fecha	Resultado	Federal	U de G	Privado	Rapidas
0	17/04/2020	Confirmados	190.0	0	18	Na
1	17/04/2020	Descartados	2601.0	110	271	Na
2	17/04/2020	Sospechosos	573.0	18	Na	Na
3	18/04/2020	Confirmados	196.0	0	18	Na
4	18/04/2020	Descartados	2666.0	110	271	Na
5	18/04/2020	Sospechosos	547.0	18	Na	Na
6	19/04/2020	Confirmados	198.0	0	18	Na
7	19/04/2020	Descartados	2805.0	110	271	Na
8	19/04/2020	Sospechosos	372.0	0	Na	Na
9	20/04/2020	Confirmados	216.0	0	18	Na

Looking at the table that produced the code, the variables that you are going to consider are:

- **Date - 1875 rows × 1 column**
  - Day/Month/Year
- **Result - 1875 rows × 2 column**
  - Confirmed: COVID tests with positive results
  - Discarded: COVID tests with negative results
  - Suspects: COVID tests with a likelihood of positive or negative
- **Federal - 1875 rows × 3 column**
  - Number of tests
- **U de G - 1875 rows × 4 column**
  - Number of tests
- **Privado - 1822 rows × 5 column**
  - Number of tests
- **QuickTest 1822 rows × 6 column**
  - Number of tests

With the next line, identify that there are 1875 total records.

```
In [6]: len(df.index)
Out[6]: 1875
```

The line where we request information by column, realize that the private and quicktest variables have 1822 points, while the others have 1875.

```
In [7]: df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1875 entries, 0 to 1874
Data columns (total 6 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Fecha      1875 non-null   object
 1   Resultado  1875 non-null   object
 2   Federal    1875 non-null   float64
 3   U de G     1875 non-null   object
 4   Privado    1822 non-null   object
 5   Rapidas    1822 non-null   object
dtypes: float64(1), object(5)
memory usage: 88.0+ KB
```

Mean:

Mean

```
In [8]: print('Mean: ', df["Federal"].mean())
Media:  60256.4832
```

Median:

Median

```
In [9]: print('Median:', df["Federal"].median())
Mediana: 29800.0
```

Standard deviation:

Standard deviation

```
In [10]: print('Standard deviation: ', df["Federal"].std())
Desviación estándar:  59558.50207360816
```

Conclusiones:

The results given by the data analysis, we could see that there is a general average of 60,256 tests among the three variables (confirmed, discarded, and suspect), in addition to this, we can see that there are many outliers in the data, which tells us that there is a great variation in the amount of tests applied per day, depending on the phase and emergency and availability of these, as well as the positive results give us a general idea of when the most critical phase of the pandemic was and, in turn, the discarded tests give us a general overview of which period of the pandemic had a reduction in the number of infections.

Git: https://github.com/sayuriGui/AnaliticsChallenge.git