



“Introduction to pandas ”

What is Pandas and what is it used for?
How to install Pandas?



What is Pandas?





What is Pandas in Python?

Pandas is a Python library that specializes in data management, analysis and processing. To do this, it is based on the data structures or arrays of the NumPy library.





How to install Pandas?

Use the command `py -m pip --version` to see if you have the PIP package manager installed. Use the command `py -m pip install numpy` to install NumPy and finally use `py -m pip install pandas` to install Pandas.



How to install Pandas?



Step	Command	Description
0	<code>py --version</code>	(Optional) This is just to ensure that Python is installed and to know which version is installed. If you are sure you have Python, this step is not necessary.
1	<code>py -m pip --version</code>	PIP is the most popular package manager for Python. In this step you make sure you have it installed and know the installed version.
2	<code>py -m pip install --upgrade pip</code>	(Optional) Update the PIP version just in case.
3	<code>py -m pip install numpy</code>	Use PIP to install NumPy, the library Pandas relies on for data structures.
4	<code>py -m pip install pandas</code>	Install Pandas with PIP. You could just run this command from the start, but you better make sure you have the other packages installed so you don't run into any errors.
5	<code>py -m pip install matplotlib</code>	(Optional) Install Matplotlib, a library that uses Pandas in the <code>plot()</code> function to create graphs from the data being processed.

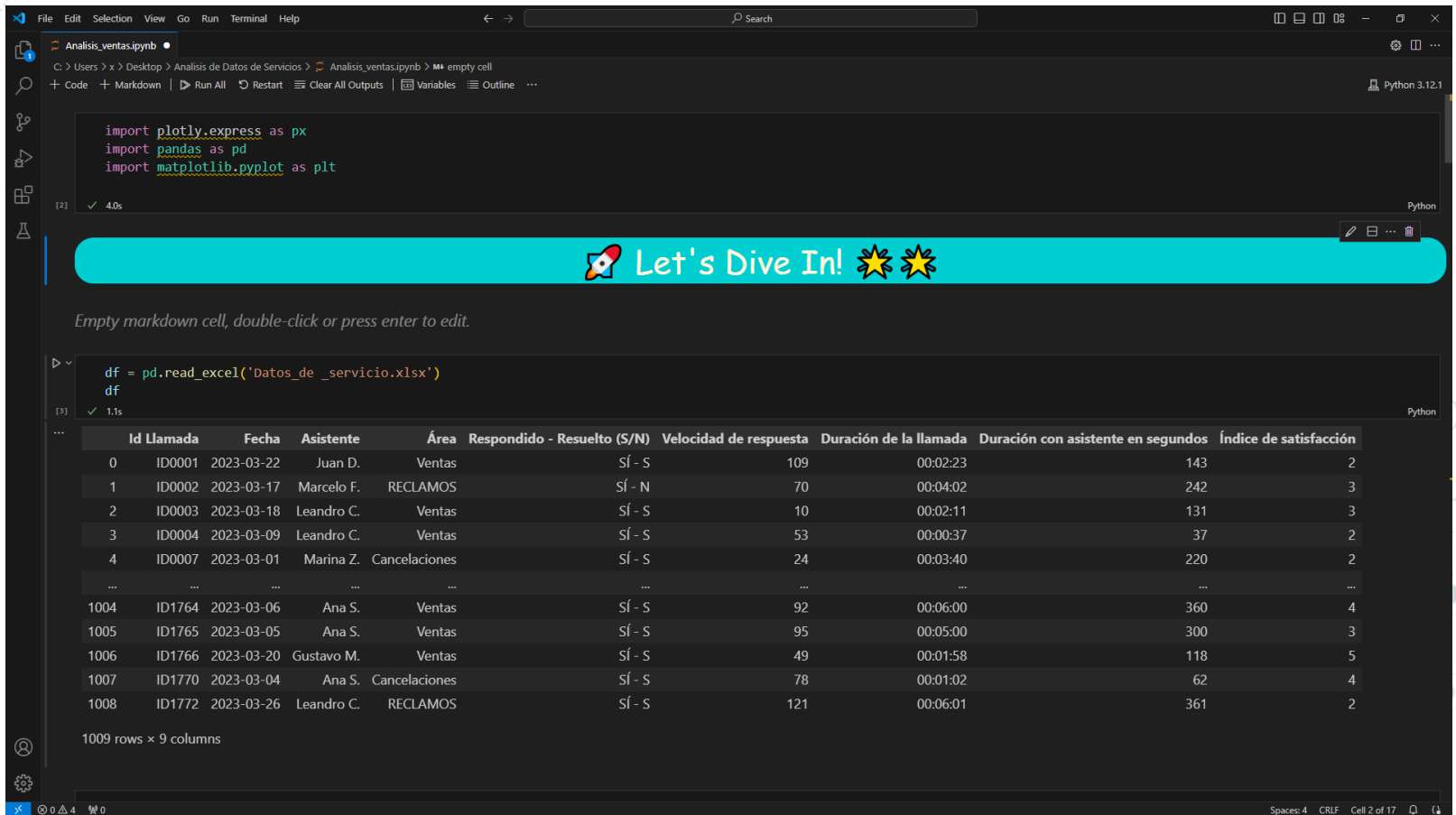


First we are going to import the libraries we need, such as pandas, matplotlib.
We are going to read our file, in this case it is called "Service_Data.xlsx", as we do with:

df - Es el dataframe

df=pd.read_Excel('Datos_de_servicios.xlsx')

pd.read - Lee el archivo de Excel



The screenshot shows a Jupyter Notebook window with the following content:

```
import plotly.express as px
import pandas as pd
import matplotlib.pyplot as plt
```

Below the code, there is a blue banner with a rocket icon and the text "Let's Dive In! 🌟🌟".

Underneath the banner, there is a Python code cell that has been executed, showing the output of the `pd.read_excel` function:

```
df = pd.read_excel('Datos_de_servicio.xlsx')
df
```

The output is a table with 1009 rows and 9 columns. The first few rows are:

	ID Llamada	Fecha	Asistente	Área	Respondido - Resuelto (S/N)	Velocidad de respuesta	Duración de la llamada	Duración con asistente en segundos	Índice de satisfacción
0	ID0001	2023-03-22	Juan D.	Ventas	Sí - S	109	00:02:23	143	2
1	ID0002	2023-03-17	Marcelo F.	RECLAMOS	Sí - N	70	00:04:02	242	3
2	ID0003	2023-03-18	Leandro C.	Ventas	Sí - S	10	00:02:11	131	3
3	ID0004	2023-03-09	Leandro C.	Ventas	Sí - S	53	00:00:37	37	2
4	ID0007	2023-03-01	Marina Z.	Cancelaciones	Sí - S	24	00:03:40	220	2

The table continues with rows 1004 to 1008, and then indicates 1009 rows in total.

We analyze the data we have in the file and we want to obtain the data.
In this case we are going to analyze the top 10 complaints. How do we do it:

Name of the column and variable where the data is stored

Name of the column to be analyzed

The screenshot shows a Jupyter Notebook interface with a dark theme. The code cell contains the following Python code:

```
reclamos= df[df['Área']=='RECLAMOS']
reclamos
top_10_Aasistente_cancelados = reclamos['Asistente'].value_counts()[:10]
top_10_Aasistente_cancelados.to_frame('Cantidad de reclamos')
```

Below the code, the output is displayed as a table titled "Cantidad de reclamos". The table has two columns: "Asistente" and a numerical value. The top 10 rows are shown.

Asistente	
Juan D.	43
Leandro C.	29
Gustavo M.	28
Ana S.	24
Paulo G.	23
Marina Z.	23
Marcelo F.	21

Four orange arrows point from text labels to specific parts of the code and output:

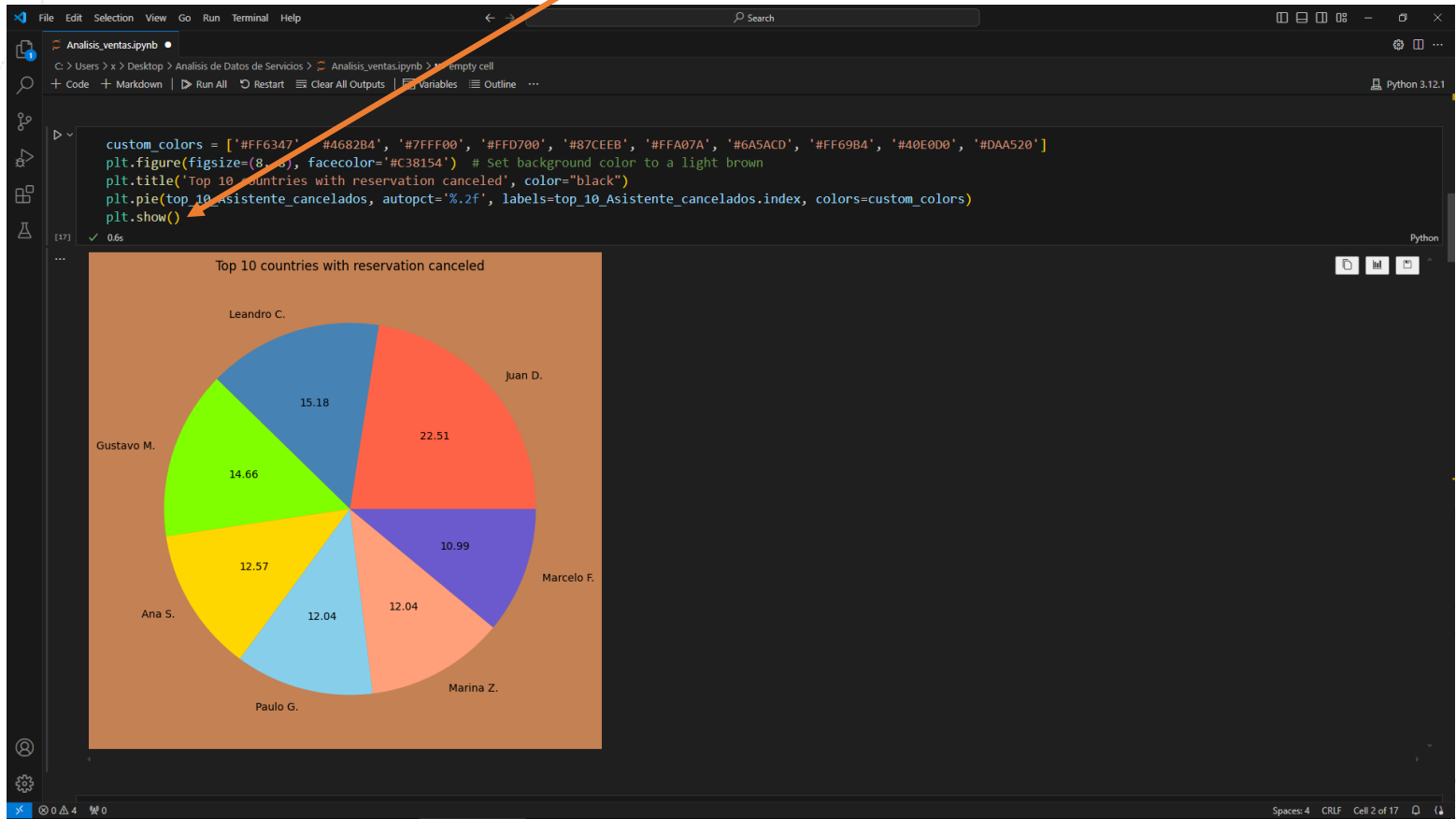
- From "Name of the column and variable where the data is stored" to the variable `reclamos` in the first line of code.
- From "Name of the column to be analyzed" to the column name `'Asistente'` in the third line of code.
- From "The 10 Names of the Attendees" to the "Asistente" column header in the output table.
- From "A variable is created where we store the Top 10." to the variable `top_10_Aasistente_cancelados` in the third line of code.

The 10 Names of the Attendees

A variable is created where we store the Top 10.

After analyzing the data, we visualize it. We use matplotlib. To visualize the graph we write:

plt.show()





THANKS



<https://github.com/Glaramos>