

# Introducción a Data Science en Python

Utilizando el paquete Pandas para Python, realiza un libro de Jupiter que analice los datos del CSV proporcionado con información sobre los pasajeros del Titanic.

Carga los datos del CSV en un dataset.

```
In [ ]: # Lo primero sera instalar pandas con pip install pandas
import pandas as p

titanic3 = p.read_csv('titanic3.csv')
print(titanic3.columns)
```

```
Index(['pclass', 'survived', 'name', 'sex', 'age', 'sibsp', 'parch', 'ticket',
      'fare', 'cabin', 'embarked', 'boat', 'body', 'home.dest'],
      dtype='object')
```

Previsualiza las primeras filas del dataset.

```
In [ ]: titanic_dataset = p.read_csv("titanic3.csv", nrows=5)
print(titanic_dataset)
```

	pclass	survived	name	sex	
0	1	1	Allen, Miss. Elisabeth Walton	female	
1	1	1	Allison, Master. Hudson Trevor	male	
2	1	0	Allison, Miss. Helen Loraine	female	
3	1	0	Allison, Mr. Hudson Joshua Creighton	male	
4	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	

	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	
0	29.0000	0	0	24160	211.3375	B5	S	2.0	NaN	
1	0.9167	1	2	113781	151.5500	C22 C26	S	11.0	NaN	
2	2.0000	1	2	113781	151.5500	C22 C26	S	NaN	NaN	
3	30.0000	1	2	113781	151.5500	C22 C26	S	NaN	135.0	
4	25.0000	1	2	113781	151.5500	C22 C26	S	NaN	NaN	

	home.dest
0	St Louis, MO
1	Montreal, PQ / Chesterville, ON
2	Montreal, PQ / Chesterville, ON
3	Montreal, PQ / Chesterville, ON
4	Montreal, PQ / Chesterville, ON

Calcula la edad media de los pasajeros.

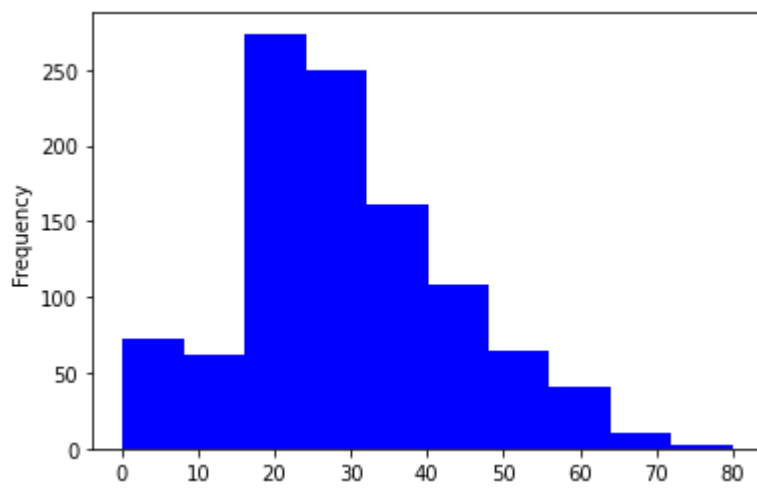
```
In [ ]: edad_media = titanic3['age'].mean()
print(edad_media)
```

```
29.8811345124283
```

Realiza un histograma sobre la distribución de edades de los pasajeros.

```
In [ ]: # Instalar antes matplotlib con pip install matplotlib
edades = titanic3['age']
histograma = edades.plot.hist(
    x="Pasajeros", y="Edad", color="blue", orientation="vertical")
print(histograma)
```

```
AxesSubplot(0.125,0.125;0.775x0.755)
```



Calcula el número de pasajeros con billetes de primera, segunda o tercera clase.

```
In [ ]: pasajeros_primera_clase = titanic3.loc[titanic3['pclass'] == 1].index.size
pasajeros_segunda_clase = titanic3.loc[titanic3['pclass'] == 2].index.size
pasajeros_tercera_clase = titanic3.loc[titanic3['pclass'] == 3].index.size
pasajeros = titanic3['ticket'].index.size

print("Pasajeros de Primera clase:", {pasajeros_primera_clase},
      "\nPasajeros de Segunda clase:", {pasajeros_segunda_clase},
      "\nPasajeros de Tercera clase:", {pasajeros_tercera_clase})
print("Total de Pasajeros:", {pasajeros})
```

```
Pasajeros de Primera clase: {323}
Pasajeros de Segunda clase: {277}
Pasajeros de Tercera clase: {709}
Total de Pasajeros: {1309}
```

Realiza una tabla pivot con la tasa de supervivencia en función del tipo de billete.

```
In [ ]: titanic3.pivot_table('survived', 'pclass')
```

```
Out[ ]:      survived
```

pclass	survived
1	0.619195
2	0.429603
3	0.255289