

EjercicioTitanic_Kaleb

February 15, 2024

Actividad M2.1 Google Colab - Operaciones con dataframes

Ricardo Kaleb Flores Alfonso, A01198716

```
[ ]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

1 Lectura de Datos

```
[ ]: from google.colab import drive
drive.mount('/content/gdrive')
```

Drive already mounted at /content/gdrive; to attempt to forcibly remount, call drive.mount("/content/gdrive", force_remount=True).

```
[ ]: df = pd.read_csv('/content/gdrive/MyDrive/Pandas/Act2/titanic.csv')
df.head(10)
```

```
[ ]: PassengerId  Survived  Pclass  \
0             1         0         3
1             2         1         1
2             3         1         3
3             4         1         1
4             5         0         3
5             6         0         3
6             7         0         1
7             8         0         3
8             9         1         3
9            10         1         2
```

```

                                Name    Sex  Age  SibSp  \
0                Braund, Mr. Owen Harris  male  22.0     1
1  Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0     1
2                Heikkinen, Miss. Laina  female  26.0     0
3  Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0     1
4                Allen, Mr. William Henry   male  35.0     0
```

5		Moran, Mr. James	male	NaN	0
6		McCarthy, Mr. Timothy J	male	54.0	0
7		Palsson, Master. Gosta Leonard	male	2.0	3
8	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)		female	27.0	0
9	Nasser, Mrs. Nicholas (Adele Achem)		female	14.0	1

	Parch	Ticket	Fare	Cabin	Embarked
0	0	A/5 21171	7.2500	NaN	S
1	0	PC 17599	71.2833	C85	C
2	0	STON/O2. 3101282	7.9250	NaN	S
3	0	113803	53.1000	C123	S
4	0	373450	8.0500	NaN	S
5	0	330877	8.4583	NaN	Q
6	0	17463	51.8625	E46	S
7	1	349909	21.0750	NaN	S
8	2	347742	11.1333	NaN	S
9	0	237736	30.0708	NaN	C

2 Descripción de Variables

```
[ ]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   PassengerId     891 non-null   int64
1   Survived        891 non-null   int64
2   Pclass         891 non-null   int64
3   Name            891 non-null   object
4   Sex             891 non-null   object
5   Age            714 non-null   float64
6   SibSp           891 non-null   int64
7   Parch          891 non-null   int64
8   Ticket          891 non-null   object
9   Fare           891 non-null   float64
10  Cabin           204 non-null   object
11  Embarked        889 non-null   object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

```
[ ]: df.isna().sum()
```

```
[ ]: PassengerId    0
      Survived      0
      Pclass       0
```

```

Name          0
Sex           0
Age          177
SibSp         0
Parch         0
Ticket        0
Fare          0
Cabin        687
Embarked      2
dtype: int64

```

PassengerId int64 numerica

Survived int64 categorica

Pclass int64 categorica

Name object categorica

Sex object categorica

Age float64 numerica

SibSp int64 categorica

Parch int64 categorica

Ticket object categorica

Fare float64 numerica

Cabin object categorica

Embarked object categorica

3 Exploración y Análisis de Datos

```
[ ]: df.describe()
```

```

[ ]:
   PassengerId  Survived  Pclass    Age  SibSp  \
count  891.000000   891.000000   891.000000  714.000000  891.000000
mean    446.000000     0.383838     2.308642   29.699118    0.523008
std    257.353842     0.486592     0.836071   14.526497    1.102743
min      1.000000     0.000000     1.000000    0.420000    0.000000
25%    223.500000     0.000000     2.000000   20.125000    0.000000
50%    446.000000     0.000000     3.000000   28.000000    0.000000
75%    668.500000     1.000000     3.000000   38.000000    1.000000
max    891.000000     1.000000     3.000000   80.000000    8.000000

      Parch    Fare
count  891.000000  891.000000
mean     0.381594   32.204208

```

```
std      0.806057    49.693429
min      0.000000     0.000000
25%      0.000000     7.910400
50%      0.000000    14.454200
75%      0.000000    31.000000
max      6.000000   512.329200
```

La tarifa más barata cuesta 0, lo cual significa que fue gratis.

El promedio de la edad fue 29.7 años.

```
[ ]: df = df.astype({'Survived': 'object', 'Pclass': 'object'})
df.describe(include='object')
```

```
[ ]:      Survived  Pclass      Name  Sex  Ticket  Cabin \
count      891      891      891  891      891      204
unique        2        3      891    2      681      147
top           0        3  Braund, Mr. Owen Harris  male  347082  B96 B98
freq         549      491          1  577          7          4

      Embarked
count      889
unique        3
top          S
freq        644
```

Distribución de frecuencias de embarked

```
[ ]: df.Embarked.value_counts()
```

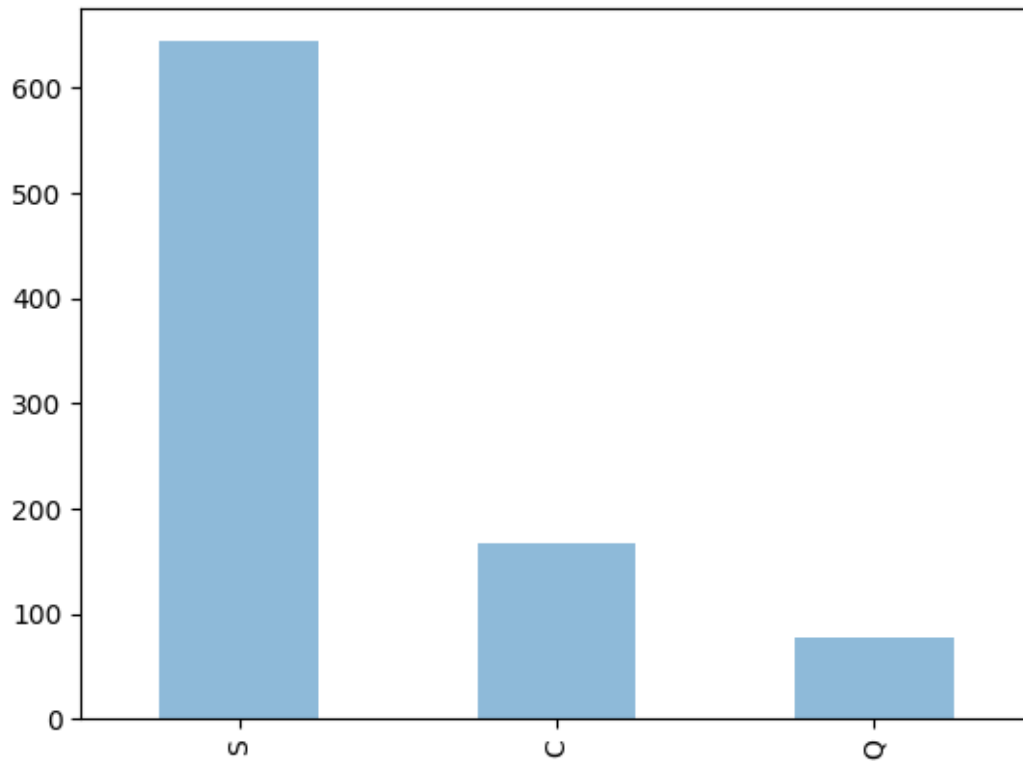
```
[ ]: S    644
     C    168
     Q     77
     Name: Embarked, dtype: int64
```

```
[ ]: fig = plt.figure(figsize=(18,6))
```

<Figure size 1800x600 with 0 Axes>

```
[ ]: df.Embarked.value_counts().plot(kind="bar",alpha=0.5)
```

```
[ ]: <Axes: >
```



```
[ ]: df.embark
```

Pasajero con id 149

```
[ ]: df.iloc[148]
```

```
[ ]: PassengerId      149
      Survived        0
      Pclass         2
      Name      Navratil, Mr. Michel ("Louis M Hoffman")
      Sex          male
      Age         36.5
      SibSp         0
      Parch         2
      Ticket      230080
      Fare        26.0
      Cabin        F2
      Embarked      S
      Name: 148, dtype: object
```

Pasajeros de primera clase

```
[ ]: df[["Name", "Pclass"]].loc[df["Pclass"]== 1].sort_values(by="Name")
```

```
[ ]:                                     Name Pclass
730                Allen, Miss. Elisabeth Walton      1
305                Allison, Master. Hudson Trevor      1
297                Allison, Miss. Helen Loraine       1
498 Allison, Mrs. Hudson J C (Bessie Waldo Daniels)   1
460                Anderson, Mr. Harry                1
..                "" "" "" "" "" "" "" "" "" "" ""
155                Williams, Mr. Charles Duane        1
351                Williams-Lambert, Mr. Fletcher Fellows 1
55                Woolner, Mr. Hugh                  1
555                Wright, Mr. George                1
325                Young, Miss. Marie Grice           1
```

[216 rows x 2 columns]

Total pagado por los pasajeros

```
[ ]: df['Fare'].sum()
```

```
[ ]: 28693.9493
```

Subconjunto de sobrevivientes

```
[ ]:
```

```
[ ]: PassengerId Survived Pclass \
1      2      1      1
2      3      1      3
3      4      1      1
8      9      1      3
9     10      1      2
..     ..     ..     ..
875    876      1      3
879    880      1      1
880    881      1      2
887    888      1      1
889    890      1      1
```

```

Name      Sex  Age  SibSp \
1  Cumings, Mrs. John Bradley (Florence Briggs Th... female 38.0      1
2                Heikkinen, Miss. Laina  female 26.0      0
3      Futrelle, Mrs. Jacques Heath (Lily May Peel) female 35.0      1
8  Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg) female 27.0      0
9      Nasser, Mrs. Nicholas (Adele Achem)  female 14.0      1
..     ..     ..     ..
875                Najib, Miss. Adele Kiamie "Jane" female 15.0      0
879  Potter, Mrs. Thomas Jr (Lily Alexenia Wilson) female 56.0      0
880  Shelley, Mrs. William (Imanita Parrish Hall)  female 25.0      0
```

887	Graham, Miss. Margaret Edith	female	19.0	0
889	Behr, Mr. Karl Howell	male	26.0	0

	Parch	Ticket	Fare	Cabin	Embarked
1	0	PC 17599	71.2833	C85	C
2	0	STON/O2. 3101282	7.9250	NaN	S
3	0	113803	53.1000	C123	S
8	2	347742	11.1333	NaN	S
9	0	237736	30.0708	NaN	C
..
875	0	2667	7.2250	NaN	C
879	1	11767	83.1583	C50	C
880	1	230433	26.0000	NaN	S
887	0	112053	30.0000	B42	S
889	0	111369	30.0000	C148	C

[342 rows x 12 columns]

```
[ ]: survived = df.loc[df["Survived"]==1]
      passedAway = df.loc[df["Survived"]==0]
```

```
[ ]: survived.head()
```

```
[ ]: PassengerId Survived Pclass \
1      2          1          1
2      3          1          3
3      4          1          1
8      9          1          3
9     10          1          2
```

	Name	Sex	Age	SibSp	\
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	
2	Heikkinen, Miss. Laina	female	26.0	0	
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	
8	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	
9	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	

	Parch	Ticket	Fare	Cabin	Embarked
1	0	PC 17599	71.2833	C85	C
2	0	STON/O2. 3101282	7.9250	NaN	S
3	0	113803	53.1000	C123	S
8	2	347742	11.1333	NaN	S
9	0	237736	30.0708	NaN	C

```
[ ]: passedAway.head()
```

```
[ ]: PassengerId Survived Pclass Name Sex Age \
0 1 0 3 Braund, Mr. Owen Harris male 22.0
4 5 0 3 Allen, Mr. William Henry male 35.0
5 6 0 3 Moran, Mr. James male NaN
6 7 0 1 McCarthy, Mr. Timothy J male 54.0
7 8 0 3 Palsson, Master. Gosta Leonard male 2.0
```

```
SibSp Parch Ticket Fare Cabin Embarked
0 1 0 A/5 21171 7.2500 NaN S
4 0 0 373450 8.0500 NaN S
5 0 0 330877 8.4583 NaN Q
6 0 0 17463 51.8625 E46 S
7 3 1 349909 21.0750 NaN S
```

Agrega tu nombre aqui: Raúl Correa Ocañas