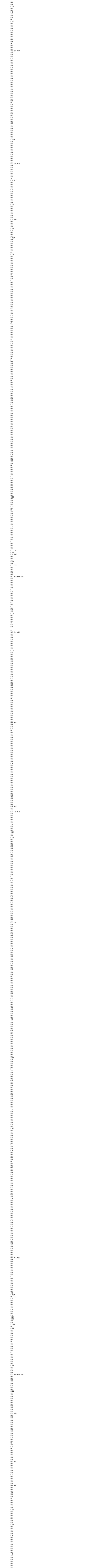
lf = d	4	1	1 Futrelle	, Mrs. Jacques	Heath (Lily	(l 0 . 0	1 0	STON/O2. 3101282		C123	
1		3	sengerId ", axis=1	Allen, Mr. Wi	illiam Henry Name S n Harris m	female 35.0 male 35.0 Sex Age Sil ale 22.0 ale 38.0	0 0	373450 Ticket A/5 21171 PC 17599	8.0500 Fare 7.2500 71.2833	NaN Cabin Eml NaN C85	barl
2 3 4 386 387 388 389	1 0 0 1 0 1 0 0	3 2 1 3 John 1 3	Mrs. Jacques A Graham	Heikkinen, Mis Heath (Lily Ma Illen, Mr. Williar Montvila, Rev. n, Miss. Margar atherine Helen Behr, Mr. Kar Dooley, Mr.	ay Peel) fem m Henry m Juozas m ret Edith fem "Carrie" fem I Howell m	ale 35.0 ale 35.0 ale 27.0	0 0 1 0 0 0 0 0 0 0 1 2 0 0 0 0	112053 W./C. 6607	53.1000 8.0500 13.0000 30.0000 23.4500 30.0000	NaN C123 NaN NaN B42 NaN C148 NaN	
Explo		Data A	Mo Graham, Mi , Miss. Cathe	Nantvila, Rev. Judiss. Margaret Erine Helen "Car	ame Sex ozas male dith female rrie" female	19.0 0 NaN 1	0 2: 0 1: 2 W./C.	Ficket Fare 11536 13.00 12053 30.00 6607 23.45 11369 30.00	Cabin Em NaN B42 NaN C148	barked S S S C	
	oe 11) cribe() Survived	3 Pclass	Age	Dooley, Mr. Pat	Parch	Fare	0 3	70376 7.75	NaN	Q	
mean std min 25% 50% 75% max	91.000000 0.383838 0.486592 0.000000 0.000000 1.000000 1.000000	891.000000 2.308642 0.836071 1.000000 2.000000 3.000000 3.000000 3.000000	714.000000 29.699118 14.526497 0.420000 20.125000 28.000000 38.000000	891.000000 0.523008 1.102743 0.000000 0.0000000 1.0000000 8.0000000	0.381594 0.806057 0.000000 0.000000 0.000000	891.000000 32.204208 49.693429 0.000000 7.910400 14.454200 31.000000 512.329200					
y aquitions of the second seco	ui vemos dull().sum(ed 0 0 0 177 0 0 0	cuantas co		(Valores que		tan					
abin mbark type:	687 ed 2 int64 intvalue_c										
or cal	abin, Ler Dina <mark>in</mark> df Int(cabina	ngth: 147,	dtype: i	nt64							



nan nan C92 nan nan nan D28 nan nan nan E17 nan nan nan nan D17 nan nan nan nan A24 nan nan nan D35 B51 B53 B55 nan nan nan nan nan nan C50 nan nan nan nan nan nan nan B42 nan C148 nan In [13]: df.info() <class 'pandas.core.frame.DataFrame'> RangeIndex: 891 entries, 0 to 890 Data columns (total 11 columns): Column Non-Null Count Dtype 0 Survived 891 non-null int64 891 non-null 1 Pclass int64 2 Name 891 non-null object 3 Sex 891 non-null object 4 Age 714 non-null float64 5 891 non-null int64 SibSp 6 Parch 891 non-null int64 7 Ticket 891 non-null object 8 Fare 891 non-null float64 9 204 non-null Cabin object 10 Embarked 889 non-null object dtypes: float64(2), int64(4), object(5) memory usage: 76.7+ KB In [14]: df.Survived.value_counts() Out[14]: 0 1 342 Name: Survived, dtype: int64 In [15]: df.Survived.value_counts().plot(kind="bar") plt.show() 500 400 300 200 100 0 0 ¿Cómo seleccionar información concreta de nuestro dataset? Forma 1 In [16]: df["Age"].head() Out[16]: 0 22.0 1 38.0 2 26.0 3 35.0 35.0 Name: Age, dtype: float64 Forma 2 In [17]: df.Age.head() Out[17]: 0 22.0 38.0 1 26.0 3 35.0 35.0 Name: Age, dtype: float64 Forma 3 In [18]: df[["Age"]].head() Out[18]: Age **0** 22.0 **1** 38.0 **2** 26.0 **3** 35.0 4 35.0 Crosstab In [19]: pd.crosstab(df.Sex, df.Survived) Out[19]: Survived Sex female 233 male 468 109 pd.crosstab(df.Sex, df.Survived).plot(kind="bar") plt.show() Survived 0 1 400 300 200 100 Sex Conclusión: • La mayoría de las mujeres sobreviven. • La mayoría de los hombres NO sobrevivieron pd.crosstab(df.Sex, df.Survived).plot(kind="bar", stacked=True) plt.show() 600 Survived 0 500 400 300 200 100 female Sex Conclusión: • Hay más hombres que mujeres, es casi el doble. pd.crosstab(df.Pclass, df.Survived) Survived Out[22]: **Pclass** 80 136 97 87 2 **3** 372 119 In [23]: pd.crosstab(df.Pclass, df.Survived).plot(kind="bar") plt.show() Survived 350 1 300 250 200 150 100 50 0 Pclass Conclusión: • La mayoría de los que NO sobrevivieron eran de la 3ª clase groupby In [24]: df.groupby("Sex").Survived.value_counts() Out[24]: Sex Survived 233 female 0 81 468 0 male 109 Name: Survived, dtype: int64 df.groupby("Sex").Survived.value_counts().plot(kind="bar") plt.show() 400 300 200 100 (female, 0) (male, 0) Sex, Survived por filtrado Selecciono aquellas filas donde Pclass == 1 • Me creo un dataframe de la misma forma que tenía antes # Una forma... In [26]: In [27]: df_sex_uno = df[df.Pclass == 1] df_sex_uno.head() Survived Pclass Sex Age SibSp Parch **Ticket** Fare Cabin Embarked Out[27]: Name 1 Cumings, Mrs. John Bradley (Florence Briggs Th... С 1 female 38.0 1 0 PC 17599 71.2833 C85 3 1 1 Futrelle, Mrs. Jacques Heath (Lily May Peel) 0 113803 53.1000 C123 S 35.0 1 6 0 1 0 17463 51.8625 S McCarthy, Mr. Timothy J male 54.0 0 E46 11 1 Bonnell, Miss. Elizabeth 58.0 0 0 113783 26.5500 C103 S female 23 1 1 Sloper, Mr. William Thompson 0 0 113788 35.5000 Α6 S male 28.0 In [28]: # Otra forma... In [29]: df_sex_crosstab = df[df.Pclass == 1]["Survived"] df_sex_crosstab.head() Out[29]: 1 1 3 1 6 0 11 1 Name: Survived, dtype: int64 Ejemplos de creación de dataframes In [30]: df_sobreviven_todos = df[df["Survived"] == 1] df_sobreviven_ninguno = df[df["Survived"] == 0] $hombres_sobrevivieron = df[(df["Survived"] == 1) & (df["Sex"] == "male")]$ $hombres_no_sobrevivieron = df[(df["Survived"] == 0) & (df["Sex"] == "male")]$ $mujeres_sobrevivieron = df[(df["Survived"] == 1) & (df["Sex"] == "female")]$ mujeres no sobrevivieron = df[(df["Survived"] == 0) & (df["Sex"] == "female")] In [31]: df_sobreviven_todos.head() Survived Pclass Name Sex Age SibSp Parch **Ticket** Fare Cabin Embarked Out[31]: Cumings, Mrs. John Bradley (Florence Briggs 1 1 1 0 PC 17599 71.2833 C85 С female 38.0 1 STON/O2. 2 1 3 0 0 Heikkinen, Miss. Laina female 26.0 7.9250 NaN S 3101282 3 1 1 Futrelle, Mrs. Jacques Heath (Lily May Peel) female 35.0 0 113803 53.1000 C123 S Johnson, Mrs. Oscar W (Elisabeth Vilhelmina 8 3 0 2 S female 27.0 347742 11.1333 NaN 9 1 2 1 0 237736 30.0708 С Nasser, Mrs. Nicholas (Adele Achem) female 14.0 NaN df_sobreviven_todos.Survived.value_counts(3) In [32]: Out[32]: 1 1.0 Name: Survived, dtype: float64 In [33]: | df_sobreviven_ninguno.head() Survived Pclass Name Sex Age SibSp Parch **Ticket** Fare Cabin Embarked Out[33]: 0 0 3 Braund, Mr. Owen Harris male 22.0 A/5 21171 7.2500 NaN S 4 0 3 Allen, Mr. William Henry male 35.0 0 373450 8.0500 NaN S 5 0 3 0 330877 Q Moran, Mr. James male NaN 8.4583 NaN 6 S 1 McCarthy, Mr. Timothy J male 54.0 0 17463 51.8625 E46 7 S 0 3 Palsson, Master. Gosta Leonard male 2.0 3 1 349909 21.0750 NaN In [34]: df_sobreviven_ninguno.Survived.value_counts(3) Out[34]: 0 1.0 Name: Survived, dtype: float64 In [35]: hombres_sobrevivieron.head() Survived Pclass Out[35]: Name Sex Age SibSp Parch **Ticket** Fare Cabin **Embarked** 17 1 Williams, Mr. Charles Eugene male NaN 0 244373 13.0000 NaN S 21 1 2 0 248698 13.0000 D56 Beesley, Mr. Lawrence 34.0 male 1 1 Sloper, Mr. William Thompson male 0 113788 35.5000 S 2677 С 36 Mamee, Mr. Hanna male NaN 7.2292 NaN 1 Woolner, Mr. Hugh male 19947 35.5000 S In [36]: hombres__no_sobrevivieron.head() Survived Pclass Age SibSp **Ticket** Cabin Embarked Name Sex Parch Fare Out[36]: 0 0 7.2500 S 3 Braund, Mr. Owen Harris male 22.0 0 A/5 21171 NaN 4 0 3 0 S Allen, Mr. William Henry male 35.0 373450 8.0500 NaN 5 0 3 0 Q Moran, Mr. James male NaN 0 330877 8.4583 NaN S 0 0 1 McCarthy, Mr. Timothy J 54.0 17463 51.8625 E46 7 349909 21.0750 S 3 Palsson, Master. Gosta Leonard male 2.0 NaN mujeres_sobrevivieron.head() Age SibSp Parch Fare Cabin Embarked Out[37]: Survived Pclass Name Ticket Cumings, Mrs. John Bradley (Florence Briggs 1 0 PC 17599 71.2833 C85 female 38.0 1 С Th... STON/O2. 2 1 3 0 0 S Heikkinen, Miss. Laina female 26.0 7.9250 NaN 3101282 3 0 S 1 Futrelle, Mrs. Jacques Heath (Lily May Peel) 1 113803 53.1000 C123 1 female 35.0 Johnson, Mrs. Oscar W (Elisabeth Vilhelmina 8 1 3 0 2 S female 27.0 347742 11.1333 NaN 9 1 2 Nasser, Mrs. Nicholas (Adele Achem) 0 237736 30.0708 С female 14.0 In [38]: mujeres_no_sobrevivieron.head() Survived Pclass SibSp Parch **Ticket** Fare Cabin Embarked Out[38]: Name Sex Age 0 S 14 3 Vestrom, Miss. Hulda Amanda Adolfina 14.0 0 0 350406 7.8542 NaN female 0 S 18 Vander Planke, Mrs. Julius (Emelia Maria Vande... 31.0 1 345763 18.0000 NaN 0 3 3 S 24 Palsson, Miss. Torborg Danira female 8.0 349909 21.0750 NaN 0 3 Vander Planke, Miss. Augusta Maria 2 S 38 female 18.0 345764 18.0000 NaN 40 0 3 Ahlin, Mrs. Johan (Johanna Persdotter Larsson) 1 S female 40.0 7546 9.4750 NaN Obtenemos información de los gráficos In [39]: df.head() Out[39]: **Embarked Pclass** SibSp Parch Cabin Survived Name Sex Age Ticket Fare 0 0 3 Braund, Mr. Owen Harris S 0 male 22.0 1 A/5 21171 7.2500 NaN Cumings, Mrs. John Bradley (Florence Briggs 1 1 1 0 C85 female 38.0 1 PC 17599 71.2833 С STON/O2. 2 S 1 3 0 0 Heikkinen, Miss. Laina female 26.0 7.9250 NaN 3101282 3 Futrelle, Mrs. Jacques Heath (Lily May Peel) female 113803 53.1000 C123 S S 4 0 3 Allen, Mr. William Henry 0 0 8.0500 35.0 373450 NaN male In [40]: opciones = ["Pclass", "Sex", "Embarked"] for opcion in opciones: pd.crosstab(df[opcion], df.Survived).plot(kind="bar") plt.show() Survived 350 0 1 300 250 200 150 100 50 0 **Pclass** Survived 0 1 400 300 200 100 0 female Sex Survived 0 400 1 350 300 250 200 150 100 50 0 O Embarked **SEABORN** In [41]: sns.catplot(x="Sex", y="Survived", hue="Pclass", kind="point", height=4, aspect=2, data=df) plt.show() 1.0 0.8 Survived 0.6 Pclass 2 0.4 3 0.2 male female Sex In [42]: sns.catplot(x="Pclass", y="Survived", hue="Sex", kind="point", col="Embarked", data=df) plt.show() Embarked = S Embarked = CEmbarked = Q 1.0 0.8 male 0.2 0.0 Pclass Algunas conclusiones: • Nos fijamos en la gráfica de la izquierda, embarked="S" --> las mujeres de 3 clase que embarcaron en "S" fallecieron muchas en comparación con 1 y 2 clase, pese a ello sobrevivieron algo más que los hombres de 1 clase del mismo puerto. • los hombres con mayor porcentaje e supervivencia embarcaron en "C" • Los hombres con menor porcentaje de supervivencia embarcaron en "Q" • Vemos nuevamente como la mayoría de las mujeres sobreviveron, pero no los hombres. Edad y supervivencia In [43]: # me creo una figura fig = plt.figure(figsize=(16,6)) # 3 subplots # 1 fila 3 columnas - gráfica 1 $ax1 = fig.add_subplot(131)$ # 1 fila 3 columnas - gráfica 2 $ax2 = fig.add_subplot(132)$ # 1 fila 3 columnas - gráfica 3 $ax3 = fig.add_subplot(133)$ # violinplot sns.violinplot(x="Embarked", y="Age", hue="Survived", data=df, ax=ax1) sns.violinplot(x="Pclass", y="Age", hue="Survived", data=df, ax=ax2) sns.violinplot(x="Sex", y="Age", hue="Survived", data=df, ax=ax3) plt.show() 0 0 0 ____1 ___1 1 80 80 80 60 60 60 40 9g 40 40 Age 20 20 20 0 0 0 -20 S C Q 2 male female Embarked Hago un split== True, para mostrarlo más visual In [44]: # me creo una figura fig = plt.figure(figsize=(16,6)) # 3 subplots # 1 fila 3 columnas - gráfica 1 $ax1 = fig.add_subplot(131)$ # 1 fila 3 columnas - gráfica 2 $ax2 = fig.add_subplot(132)$ # 1 fila 3 columnas - gráfica 3 $ax3 = fig.add_subplot(133)$ # violinplot sns.violinplot(x="Embarked", y="Age", hue="Survived", data=df, split=True, ax=ax1) sns.violinplot(x="Pclass", y="Age", hue="Survived", data=df, split=True, ax=ax2) sns.violinplot(x="Sex", y="Age", hue="Survived", data=df, split=True, ax=ax3) plt.show() Survived Survived Survived 0 0 ____1 80 80 80 60 60 60 40 Age 9g 40 96 40 20 20 20 0 0 0 -20S Ċ female Embarked Pclass Sex Conclusiones: • EMBARKED y Age: La gente de unos 18-35 años de Q SI sobrevivieron mayoritariamente, (no todos) no hay porcentajes mayoritarios significativos en las otras 2 embarcaciones ■ En Q embarcaron bastantes niños los cuales no sobrevivieron. · PCLASS y Age: De la 2ª clase sobre todo y la 3 sobrevivieron la mayoria de sus niños • Sex y Age: Hay mas ancianos que ancianas Los jovenes (varón) menores de 20 años en general sobrevivieron pero no las mujeres In [45]: df.Age.describe() 714.000000 Out[45]: count 29.699118 mean std 14.526497 min 0.420000 25% 20.125000 50% 28.000000 75% 38.000000 80.000000 max Name: Age, dtype: float64 Heapmap In [46]: plt.figure(figsize=(8,8)) sns.heatmap(df.corr(), annot=True) plt.show() /tmp/ipykernel_7659/644827889.py:2: FutureWarning: The default value of numeric_only in DataFrame.corr is depre cated. In a future version, it will default to False. Select only valid columns or specify the value of numeric _only to silence this warning. sns.heatmap(df.corr(), annot=True) - 1.0 Survived 1 -0.34-0.077-0.035 0.082 0.26 - 0.8 Pclass -0.55 -0.34 1 -0.37 0.083 0.018 - 0.6 - 0.4 -0.077 1 -0.31 -0.37 -0.190.096 - 0.2 -0.035 0.083 -0.311 0.41 0.16 - 0.0 Parch 0.082 0.018 -0.190.41 0.22 1 -0.2-0.40.26 -0.55 0.096 0.16 0.22 1 **Pclass** Survived SibSp Age Parch Fare barplot sns.barplot(x="Pclass", y="Survived", data=df) plt.show() 0.7 0.6 0.5 0.4 0.3 0.2 0.1 0.0 3 1 2 **Pclass** In [48]: def funcion_graficas(feat): plt.subplot(2, 1, 1) df.groupby(feat).Survived.value_counts().plot(kind="bar") plt.figure(figsize=(12,8)) plt.subplot(2, 1, 2) sns.barplot(x=feat, y="Survived", data=df) plt.show() In [49]: funcion_graficas("Pclass")

