

Escritura de un reporte de texto con las métricas del Titanic

Curso de Python Científico

Profesor: Alan Badillo Salas

Julio, 2023

Aburto Alcudia Alfredo
alfredo.abac92@gmail.com
5576430540

Introducción

Como parte del curso de Python científico en específico la sección de análisis de datos, se me ha encomendado la ilustre tarea de diseccionar el enigmático conjunto de datos del Titanic, una calamidad histórica que resuena con una profunda intriga hasta el día de hoy. En ese fatídico día, el RMS Titanic, una maravilla de la ingeniería humana se encontró con su prematura desaparición, enviando ondas de choque por todo el mundo y grabando un capítulo sombrío en la historia marítima. Ahora, mientras nos embarcamos en una odisea de revelaciones numéricas, buscando descubrir la esencia misma de la supervivencia en medio del caos se encuentra una pregunta tentadora: ¿Qué múltiples variables gobernaron la danza etérea de la vida y la muerte a bordo de este barco condenado? Nuestra valiente búsqueda abarca un intrincado tapiz de métricas fundamentales:

1. El conteo de sobrevivientes: Al sumergirnos en las profundidades de nuestro análisis se emerge iluminado con el etéreo recuento de supervivientes en medio de la calamitosa tempestad, un testimonio del indomable espíritu humano en medio del abismo.

2. Géneros en medio de la agitación: atraviase los corredores laberínticos de datos, donde examinamos meticulosamente las tasas de supervivencia en un panorama estratificado por género. He aquí, la desconcertante interacción de los destinos masculino y femenino, entrelazados con las duras mareas del destino.

3. Estratos socioeconómicos revelados: asciende en la jerarquía de los pasajeros del Titanic mientras desentrañamos el impacto de la posición social en la supervivencia. Desde los confines de élite de la grandeza de primera clase hasta las fatigas del escalón de tercera clase, sea testigo de la incesante batalla por la supervivencia.

Al comenzar esta odisea, prepárate para dejarte seducir por el nexo esotérico del género, el estatus social y los caprichosos vientos de la supervivencia. El conjunto de datos del Titanic oculta sus secretos con la mayor sutileza, y nosotros, los intrépidos viajeros de datos, estamos decididos a desentrañar sus verdades veladas.

Con una perplejidad enigmática, nos sumergimos en las profundidades de las intrincadas relaciones de datos, mientras que una explosión estimulante impregna nuestra narrativa con una sinfonía de oraciones de variada longitud. Más allá del ámbito de los meros dígitos y cifras, nuestro viaje iluminará la condición humana en medio del abismo de la historia, preservando el legado de aquellos que navegaron en el fatídico viaje del Titanic.

Justificación

En los anales de la historia, el desastre del Titanic se erige como un emblema de la resiliencia y vulnerabilidad humana, evocando un profundo sentido de curiosidad y empatía. Mientras nos encontramos al borde del precipicio para desentrañar sus misterios, nos vemos obligados tanto por un sentido del deber histórico como por una sed de conocimiento para aventurarnos en las profundidades del conjunto de datos del Titanic. Este viaje trasciende una mera búsqueda académica; es un esfuerzo ardiente para conmemorar las almas perdidas.

La importancia de comprender los factores que influyeron en la supervivencia a bordo del Titanic se extiende más allá del ámbito de las estadísticas. Al iluminar la interacción del género, la clase socioeconómica y las probabilidades de supervivencia, otorgamos a la historia un tapiz vívido de luchas y triunfos humanos. Cada vida representada por un punto de datos es un conmovedor recordatorio del delicado equilibrio entre la fortuna y el destino, el coraje y la fragilidad, en medio del caótico telón de fondo de una catástrofe marítima.

A través de un análisis de datos, llevamos la antorcha del recuerdo, asegurando que la memoria colectiva de esta tragedia perdure de generación en generación. Más allá de la búsqueda de respuestas, nuestro viaje está impulsado por la creencia inquebrantable de que, al comprender el pasado, podemos dar forma a un futuro más compasivo. Al reconocer las disparidades en las tasas de supervivencia y sus causas subyacentes, obtenemos información sobre el impacto de las normas sociales y los roles de género en la experiencia humana durante tiempos de crisis.

Además, el conjunto de datos del Titanic sirve como un potente recordatorio de la importancia de la preparación y la seguridad en los tiempos modernos. Nuestro análisis puede inspirar políticas y prácticas que mejoren la seguridad de las generaciones actuales y futuras en sus viajes a través de aguas peligrosas. Al aprender del pasado, nos esforzamos por forjar un mundo en el que menos tragedias reflejen la triste historia del Titanic.

La sinfonía de complejidad y explosión en nuestro análisis de datos garantiza una comprensión de los matices multifacéticos que rodean la supervivencia. Aceptamos esta diversidad en nuestra narrativa, ya que refleja la miríada de emociones evocadas por la historia del Titanic: tristeza, esperanza, desconcierto y coraje, todas entremezcladas para crear una profunda historia humana.

Resolución del problema

▼ Resolución del problema

▼ 1. Adquisición de datos

1. Importamos la libreria Pandas, siendo esta libreria la que nos permitira adquirir los datos desde un data set con formato CSV.

```
1 import pandas as pd
```

2. Leemos el CSV ttitulado "Titanic.csv" ingresando la url donde existe dicho dataset y lo almacenamos en un DataFrame nombrado data

```
1 data = pd.read_csv("https://raw.githubusercontent.com/dragonnomada/pycien-2023/main/datasets/titanic.csv")
```

3. Visualizamos una parte del DataSet para incorporar una idea general de la estructura de los datos.

```
1 data.sample(20)
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
673	674	1	2	Wilhelms, Mr. Charles	male	31.0	0	0	244270	13.0000	NaN	S
510	511	1	3	Daly, Mr. Eugene Patrick	male	29.0	0	0	382651	7.7500	NaN	Q
17	18	1	2	Williams, Mr. Charles Eugene	male	NaN	0	0	244373	13.0000	NaN	S
73	74	0	3	Chronopoulos, Mr. Apostolos	male	26.0	1	0	2680	14.4542	NaN	C
138	139	0	3	Osen, Mr. Olaf Elon	male	16.0	0	0	7534	9.2167	NaN	S
56	57	1	2	Rugg, Miss. Emily	female	21.0	0	0	C.A. 31026	10.5000	NaN	S
425	426	0	3	Wiseman, Mr. Phillippe	male	NaN	0	0	A/4. 34244	7.2500	NaN	S
750	751	1	2	Wells, Miss. Joan	female	4.0	1	1	29103	23.0000	NaN	S
553	554	1	3	Leeni, Mr. Fahim ("Philip Zenni")	male	22.0	0	0	2620	7.2250	NaN	C
571	572	1	1	Appleton, Mrs. Edward Dale (Charlotte Lamson)	female	53.0	2	0	11769	51.4792	C101	S
814	815	0	3	Tomlin, Mr. Ernest Portage	male	30.5	0	0	364499	8.0500	NaN	S
274	275	1	3	Healy, Miss. Hanora "Nora"	female	NaN	0	0	370375	7.7500	NaN	Q
737	738	1	1	Lesurer, Mr. Gustave J	male	35.0	0	0	PC 17755	512.3292	B101	C
423	424	0	3	Danbom, Mrs. Ernst Gilbert (Anna Sigrid Maria ...	female	28.0	1	1	347080	14.4000	NaN	S
474	475	0	3	Strandberg, Miss. Ida Sofia	female	22.0	0	0	7553	9.8375	NaN	S
191	192	0	2	Carbines, Mr. William	male	19.0	0	0	28424	13.0000	NaN	S
374	375	0	3	Palsson, Miss. Stina Viola	female	3.0	3	1	349909	21.0750	NaN	S
388	389	0	3	Sadlier, Mr. Matthew	male	NaN	0	0	367655	7.7292	NaN	Q

4. Obtenemos información de los tipos de datos contenidos en cada categoria del dataframe

```
1 data.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
```

▼ Limpieza de los datos

Como podemos observar, los resultados del método `info()` y `sample()` tenemos ciertos puntos importantes a notar y corregir los cuáles se enlistan a continuación:

1. Columnas en inglés.
2. El género esta definido con su nombre en inglés.
3. Existen valores de tipo NaN para edad, cabina, embarque.

Las siguientes lineas de código buscaran "limpiar" los datos para poder obtener información del dataset de manera más eficiente.

▼ Columnas en inglés

1. Cambiamos los encabezados de las columnas a nuestra conveniencia

```
1 data.columns = ["PASAJERO_ID", "ESTADO_SUPERVIVENCIA", "CLASE_BOLETO", "NOMBRE",
2                 "GENERO", "EDAD", "NUMERO_HERMANOS", "NUMERO_PADRES",
3                 "TICKET_ID", "PRECIO_TICKET", "CABINA", "MUELLE"]
4
5 # Mostramos data
6 data
```

	PASAJERO_ID	ESTADO_SUPERVIVENCIA	CLASE_BOLETO	NOMBRE	GENERO	EDAD	NUMERO_HERMANOS	NUMERO_PADRES	TICKET_ID	PRECIO_TIC
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0
...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7

891 rows × 12 columns

▼ Definición del género en inglés

2. Cambiamos los textos "female" y "male" a "FEMENINO" y "MASCULINO", correspondientemente.

```
1 # Usando el método map() buscamos en la columna "GENERO" del dataframe data las
2 # palabras "female" y "male" para después reemplazarlas
3 data['GENERO'] = data['GENERO'].map({
4     "female" : "FEMENINO",
5     "male"   : "MASCULINO"
6 })
```

```
7
8 # Mostramos data
9 data
```

	PASAJERO_ID	ESTADO_SUPERVIVENCIA	CLASE_BOLETO	NOMBRE	GENERO	EDAD	NUMERO_HERMANOS	NUMERO_PADRES	TICKET_ID	PRECIO
0	1	0	3	Braund, Mr. Owen Harris	MASCULINO	22.0	1	0	A/5 21171	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	FEMENINO	38.0	1	0	PC 17599	
2	3	1	3	Heikkinen, Miss. Laina	FEMENINO	26.0	0	0	STON/O2. 3101282	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	FEMENINO	35.0	1	0	113803	
4	5	0	3	Allen, Mr. William Henry	MASCULINO	35.0	0	0	373450	
...
886	887	0	2	Montvila, Rev. Juozas	MASCULINO	27.0	0	0	211536	
887	888	1	1	Graham, Miss. Margaret Edith	FEMENINO	19.0	0	0	112053	
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	FEMENINO	NaN	1	2	W./C. 6607	
889	890	1	1	Behr, Mr. Karl Howell	MASCULINO	26.0	0	0	111369	
890	891	0	3	Dooley, Mr. Patrick	MASCULINO	32.0	0	0	370376	

891 rows × 12 columns

Valores de tipo NaN para columna "EDAD"

Como se ha mostrado contamos con 891 registros en el dataframe y como se observa con el método info. Tenemos para la columna edad un total de 714 valores diferentes a NaN dejando asi **177** registros sin valor. Representando estos valores el **19.86 %** de los datos totales. Con estas consideraciones en mente se hace la siguiente aclaración *los resultados presentados que requieran el uso de la edad no toman en cuenta la muestra de 891 personas sino utilizara la muestra referente a 177 personas*

▼ Obtención de las variables requeridas para el reporte

▼ Total de supervivientes

```
1 # Sobrevive es una serie booleana que evalua si los valores en la columna
2 # "ESTADO_SUPERVIVENCIA" son exactamente iguales a 1
3 sobrevive = data['ESTADO_SUPERVIVENCIA'] == 1
4
5 print(sobrevive)
6
7
8
```

0	False
1	True
2	True
3	True
4	False
...	...
886	False
887	True
888	False

```
889     True
890     False
Name: ESTADO_SUPERVIVENCIA, Length: 891, dtype: bool

1 # Luego utilizando el metodo sum contamos el número de valores TRUE
2 #almacenados en "sobrevive"
3 supervivientes_total = sobrevive.sum()
4
5 # Mostramos el total de sobrevivientes
6 print(f'El total de sobrevivientes es: {supervivientes_total}')
```

El total de sobrevivientes es: 342

```
1 # Para unificar el procedimiento creamos una función

1
2 def conteo(data_frame, columna, condicion):
3     '''
4     Cuenta el número de coincidencias que cumplen una condición en una columna
5     de un DataFrame.
6
7     Args:
8         data_frame : Es el DataFrame donde se desea buscar una coincidencia
9         columna: Es el eje en el que se quiere buscar la coincidencia
10        condicion : Es la condición que se desea evaluar
11    '''
12    coincidencias = data_frame[condicion][columna]
13    return coincidencias.count()
14
15 # Ejemplo de llamada a la función con una condición simple
16 sobrevivientes = conteo(data, 'ESTADO_SUPERVIVENCIA',
17                          data['ESTADO_SUPERVIVENCIA'] == 1)
18
19 print(f'Total de sobrevivientes es: {sobrevivientes}')
```

Total de sobrevivientes es: 342

```
1 # Ejemplo de llamada a la función con una condición compuesta
2
3 sobrevivientes_mujeres = conteo(data, 'ESTADO_SUPERVIVENCIA',
4 (data['ESTADO_SUPERVIVENCIA'] == 1) & (data['GENERO'] == "FEMENINO"))
5
6 print(f'Total de sobrevivientes mujeres es: {sobrevivientes_mujeres}')
```

Total de sobrevivientes mujeres es: 233

Creamos la mayoría del reporte haciendo una única llamada a la función

```
1 def reporte(data_frame):
2     '''
3     Imprime un reporte con datos extraidos de un data frame
4
5     Args:
6         data_frame : DataFrame de donde se extraera toda la información
7     '''
8
9     # Headers
10    print('{:^45}'.format("Métricas del Titanic"))
11    print("+-----+")
12    print('{:^30} {:^10}|'.format('Parametro','Valor'))
13    print("| ----- |")
14
15
16    # Total de sobrevivientes
17    sobrevivientes = conteo(data, 'ESTADO_SUPERVIVENCIA',
18                          data['ESTADO_SUPERVIVENCIA'] == 1)
19
20    print('{:^30} {:^10}|'.format('Total de sobrevivientes', sobrevivientes))
21    print("| ----- |")
22
23    # Total de supervivientes mujeres
24    sobrevivientes_mujeres = conteo(data, 'ESTADO_SUPERVIVENCIA',
25 (data['ESTADO_SUPERVIVENCIA'] == 1) & (data['GENERO'] == "FEMENINO"))
26
27    print('{:^30} {:^10}|'.format('Sobrevivientes mujeres',
28 sobrevivientes_mujeres))
29    print("| ----- |")
30
31    # Total de supervivientes hombres
32    sobrevivientes_hombres = conteo(data, 'ESTADO_SUPERVIVENCIA',
33 (data['ESTADO_SUPERVIVENCIA'] == 1) & (data['GENERO'] == "MASCULINO"))
34
35    print('{:^30} {:^10}|'.format('Sobrevivientes hombres',
36 sobrevivientes_hombres))
37    print("| ----- |")
38
39    # Total de supervivientes mayores a 18 años
```



```
40 s_mayor_edad = conteo(data, 'ESTADO_SUPERVIVENCIA',
41 (data['ESTADO_SUPERVIVENCIA'] == 1) & (data['EDAD'] > 18))
42
43 print('|{: ^30} {: ^10}|'.format('Sobrevivientes +18', s_mayor_edad))
44 print("| ----- |")
45
46 # Total de supervivientes menores a 18 años
47 s_menor_edad = conteo(data, 'ESTADO_SUPERVIVENCIA',
48 (data['ESTADO_SUPERVIVENCIA'] == 1) & (data['EDAD'] < 18))
49
50 print('|{: ^30} {: ^10}|'.format('Sobrevivientes -18', s_menor_edad))
51 print("| ----- |")
52
53 # Total de supervivientes mayores a 50 años
54 s_mayor_cincuenta = conteo(data, 'ESTADO_SUPERVIVENCIA',
55 (data['ESTADO_SUPERVIVENCIA'] == 1) & (data['EDAD'] > 50))
56
57 print('|{: ^30} {: ^10}|'.format('Sobrevivientes +50', s_mayor_cincuenta))
58 print("| ----- |")
59
60 # Total de muertos
61 muertos_total = conteo(data, 'ESTADO_SUPERVIVENCIA',
62 (data['ESTADO_SUPERVIVENCIA'] == 0))
63
64 print('|{: ^30} {: ^10}|'.format('Total de fallecidos', muertos_total))
65 print("| ----- |")
66
67 # Total de muertos mujeres
68 muertos_mujeres = conteo(data, 'ESTADO_SUPERVIVENCIA',
69 (data['ESTADO_SUPERVIVENCIA'] == 0) & (data['GENERO'] == 'FEMENINO'))
70
71 print('|{: ^30} {: ^10}|'.format('Mujeres Fallecidas', muertos_mujeres))
72 print("| ----- |")
73
74 # Total de muertos hombres
75 muertos_hombres = conteo(data, 'ESTADO_SUPERVIVENCIA',
76 (data['ESTADO_SUPERVIVENCIA'] == 0) & (data['GENERO'] == 'MASCULINO'))
77
78 print('|{: ^30} {: ^10}|'.format('Hombres Fallecidos', muertos_hombres))
79 print("| ----- |")
80
81 # Tasa de supervivencia de la clase 1
82 s_c1 = conteo(data, 'CLASE_BOLETO', (data['CLASE_BOLETO'] == 1) &
83 (data['ESTADO_SUPERVIVENCIA'] == 1))
84 clase_001 = conteo(data, 'CLASE_BOLETO', data['CLASE_BOLETO'] == 1)
85
86 tasa_sup_cls_001 = (s_c1 / clase_001)
87
88 print('|{: ^30} {: ^10.2f}|'.format('Tasa de Supervivencia clase 1',
89 tasa_sup_cls_001))
90 print("| ----- |")
91
92 # Tasa de supervivencia de la clase 1 siendo mujer
93
94 clase_0m1 = conteo(data, 'CLASE_BOLETO', (data['CLASE_BOLETO'] == 1) &
95 (data['GENERO'] == 'FEMENINO') & (data['ESTADO_SUPERVIVENCIA'] == 1))
96 tasa_sup_cls_0m1 = (clase_0m1 / clase_001)
97
98 print('|{: ^30} {: ^10.2f}|'.format('" " " " " siendo mujer',
99 tasa_sup_cls_0m1))
100 print("| ----- |")
101
102 # Tasa de supervivencia de la clase 2
103 s_c2 = conteo(data, 'CLASE_BOLETO', (data['CLASE_BOLETO'] == 2) &
104 (data['ESTADO_SUPERVIVENCIA'] == 1))
105 clase_002 = conteo(data, 'CLASE_BOLETO', data['CLASE_BOLETO'] == 2)
106 tasa_sup_cls_002 = (s_c2 / clase_002)
107
108 print('|{: ^30} {: ^10.2f}|'.format('Tasa de Supervivencia clase 2',
109 tasa_sup_cls_002))
110 print("| ----- |")
111
112 # Tasa de supervivencia de la clase 2 siendo mujer
113 clase_0m2 = conteo(data, 'CLASE_BOLETO', (data['CLASE_BOLETO'] == 2) &
114 (data['GENERO'] == 'FEMENINO') & (data['ESTADO_SUPERVIVENCIA'] == 1))
115 tasa_sup_cls_0m2 = (clase_0m2 / clase_002)
116
117 print('|{: ^30} {: ^10.2f}|'.format('" " " " " siendo mujer',
118 tasa_sup_cls_0m2))
119 print("| ----- |")
120
121
122
123 # Tasa de supervivencia de la clase 3
124 s_c3 = conteo(data, 'CLASE_BOLETO', (data['CLASE_BOLETO'] == 3) &
125 (data['ESTADO_SUPERVIVENCIA'] == 1))
126 clase_003 = conteo(data, 'CLASE_BOLETO', data['CLASE_BOLETO'] == 3)
127 tasa_sup_cls_003 = (s_c3 / clase_003)
128
129 print('|{: ^30} {: ^10.2f}|'.format('Tasa de Supervivencia clase 3',
130 tasa_sup_cls_003))
```

```
131     print("| ----- |")
132
133     # Tasa de supervivencia de la clase 3 siendo mujer
134     clase_0m3 = conteo(data, 'CLASE_BOLETO', (data['CLASE_BOLETO'] == 3) &
135         (data['GENERO'] == 'FEMENINO') & (data['ESTADO_SUPERVIVENCIA'] == 1))
136     tasa_sup_cls_0m3 = (clase_0m3 / clase_003)
137
138     print('|{: ^30} {: ^10.2f}|'.format('" " " " " siendo mujer',
139                                         tasa_sup_cls_0m3))
140     print("| ----- |")
141
142     # Probabilidad previa de ser mujer
143     mujeres = conteo(data, 'GENERO', data['GENERO'] == 'FEMENINO')
144     total_personas = len(data)
145     prob_mujer = mujeres / total_personas
146
147     # Probabilidad previa de ser hombre
148     prob_hombre = 1 - prob_mujer
149
150     # Probabilidad de sobrevivir dado que se es mujer
151     s_mujeres = conteo(data, 'ESTADO_SUPERVIVENCIA', (data['ESTADO_SUPERVIVENCIA'] == 1) & (data['GENERO'] == 'FEMENINO'))
152     prob_s_mujer = s_mujeres / mujeres
153
154     print('|{: ^30} {: ^10.2f}|'.format('P(S) dado que se es mujer',
155                                         prob_s_mujer))
156
157     print("| ----- |")
158     # Probabilidad de sobrevivir dado que se es hombre
159     s_hombres = conteo(data, 'ESTADO_SUPERVIVENCIA', (data['ESTADO_SUPERVIVENCIA'] == 1) & (data['GENERO'] == 'MASCULINO'))
160     prob_s_hombre = s_hombres / (total_personas - mujeres)
161
162     print('|{: ^30} {: ^10.2f}|'.format('P(S) dado que se es hombre',
163                                         prob_s_hombre))
164     print("| ----- |")
165
166     # Probabilidad de sobrevivir dado que se es mujer y está en la clase 1
167     # (usar regla de Bayes compuesta)
168     clase_001 = conteo(data, 'CLASE_BOLETO', data['CLASE_BOLETO'] == 1)
169     clase_0m1 = conteo(data, 'CLASE_BOLETO', (data['CLASE_BOLETO'] == 1) & (data['GENERO'] == 'FEMENINO') & (data['ESTADO_SUPERVIVENCIA']
170     prob_mujer_c1 = clase_0m1 / clase_001
171
172     s_mujeres_c1 = conteo(data, 'ESTADO_SUPERVIVENCIA', (data['ESTADO_SUPERVIVENCIA'] == 1) & (data['GENERO'] == 'FEMENINO') & (data['CL
173     prob_s_mujer_c1 = s_mujeres_c1 / clase_0m1
174
175     # Probabilidad de sobrevivir siendo mujer y en la clase 1 (usar regla de Bayes compuesta)
176     prob_s_f_y_c1 = prob_s_mujer_c1 * prob_mujer_c1
177
178     print('|{: ^30} {: ^10.2f}|'.format('P(S) siendo mujer en clase 1', prob_s_f_y_c1))
179
180
181     print("+-----+")
182
183 reporte(data)
```

Métricas del Titanic

Parametro	Valor
Total de sobrevivientes	342
Sobrevivientes mujeres	233
Sobrevivientes hombres	109
Sobrevivientes +18	220
Sobrevivientes -18	61
Sobrevivientes +50	22
Total de fallecidos	549
Mujeres Fallecidas	81
Hombres Fallecidos	468
Tasa de Supervivencia clase 1	0.63
" " " " " siendo mujer	0.42
Tasa de Supervivencia clase 2	0.47
" " " " " siendo mujer	0.38
Tasa de Supervivencia clase 3	0.24
" " " " " siendo mujer	0.15
P(S) dado que se es mujer	0.74
P(S) dado que se es hombre	0.19
P(S) siendo mujer en clase 1	0.42

1



Conclusión

En conclusión, el viaje a través del conjunto de datos del Titanic no estuvo exento de desafíos y demandas de un mayor esfuerzo. La implementación de una función de conteo de conjuntos de datos, aunque necesaria, reveló las complejidades involucradas en el manejo de datos históricos de tal magnitud. Subrayó la necesidad de una atención meticulosa a los detalles y una búsqueda incesante de la precisión.

De hecho, este análisis iluminó la dificultad inherente de establecer correlaciones entre la información numérica y las narrativas profundas que residen dentro de los datos. Los números por sí solos no pueden encapsular el coraje, el sacrificio y la resistencia exhibidos por los pasajeros a bordo del Titanic. Nos corresponde a nosotros, como analistas, trascender el ámbito numérico y profundizar en las historias humanas que se desarrollan detrás de cada punto de datos. Debemos esforzarnos por infundir empatía y contexto en nuestras interpretaciones, cerrando la brecha entre las estadísticas y la profunda experiencia humana.

Además, aunque hemos vislumbrado la posible correlación entre el género, la clase socioeconómica y la supervivencia, se justifica una investigación más profunda. Se recomienda encarecidamente un proyecto integral para identificar variables que ofrezcan una visión más profunda de la intrincada red de relaciones entre la clase socioeconómica y la supervivencia. Descubrir tales conexiones puede arrojar luz sobre las estructuras sociales históricas y brindar lecciones valiosas para la dinámica social contemporánea.

Además, para mejorar la solidez de los análisis futuros, es vital buscar fuentes de datos con una mayor cantidad de registros e información más completa. Al expandir el conjunto de datos, podemos desbloquear nuevas dimensiones de comprensión y descubrir patrones que podrían haber permanecido ocultos en nuestras limitaciones actuales.

A medida que nos adentramos en las profundidades de la narrativa del Titanic, queda claro que este evento histórico trasciende el mero análisis numérico. Nos llama a empatizar con la condición humana, rindiendo homenaje a la memoria de quienes se embarcaron en ese fatídico viaje. Esto, a su vez, nos recuerda la responsabilidad que conlleva el análisis de datos: recordar que cada punto de datos representa a un individuo único, una vida que se cruzó con el destino en medio de las aguas heladas.

Más allá de las complejidades técnicas, un análisis del Titanic nos implora reflexionar sobre las lecciones que ofrece. Nos insta a priorizar la seguridad, la preparación y la compasión en todos los aspectos de la vida. Nos recuerda que bajo el barniz de las jerarquías sociales y los roles de género, la esencia de la humanidad se mantiene resistente frente a la adversidad.