

Práctica 101 Escribir un reporte de texto con las métricas del Titanic

Alumno: José Luis Domínguez Hernández
Profesor asignado: Alan Badillo Salas

(julio de 2023)

Introducción

El RMS Titanic, fue un transatlántico británico (el mayor barco de pasajeros del mundo al finalizar su construcción), que naufragó en las aguas del océano Atlántico, durante la noche del 14 abril de 1912 mientras realizaba su viaje inaugural desde Southampton, Reino Unido a Nueva York, Estados Unidos, chocó con un iceberg, provocando daños irreparables en la nave. En el hundimiento murieron 1496 personas, de las 2208 que iban a bordo, lo que convierte a esta catástrofe en uno de los mayores naufragios de la historia, ocurridos en tiempos de paz.

En el presente trabajo se presenta un análisis de los datos de los pasajeros. La información se obtuvo de dirección reportada en bibliografía como datos del Titanic

1. Información disponible

Se tiene la información de 891 pasajeros y 12 atributos asociados. La información es incompleta. Faltan 177 datos de la edad, 687 datos de cabina y 2 datos de embarcación. En esta práctica, no incluye completar la información.

La propiedad shape del objeto, nos informa del número de y columnas

```
data.shape
```

(891,12)

La función value_counts() da información sobre los datos. Mediante esta instrucción se obtiene la siguiente información:

El conteo de sobrevivientes se obtuvo directamente de los datos con sum de pandas.

```
print('Sobrevivientes \n{} \n '.format(data['Survived'].value_counts()))
print('Sexo \n {} \n '.format(data['Sex'].value_counts()))
print('Clase \n {} \n '.format(data['Pclass'].value_counts()))
```

```
Sobrevivientes
0    549
1    342
Name: Survived, dtype: int64
```

```
Sexo
male    577
female  314
Name: Sex, dtype: int64
```

```
Clase
3    491
1    216
2    184
Name: Pclass, dtype: int64
```

```

print('Total de sobrevivientes = {}'.format(data["Survived"].eq(1).sum()))
print('Total de sobrevivientes mujeres = {}'.format((data["Survived"].eq(1) & data["Sex"].eq('female')).sum()))
print('Total de sobrevivientes hombres = {}'.format((data["Survived"].eq(1) & data["Sex"].eq('male')).sum()))
print('Total de sobrevivientes mayores a 50 años= {}'.format((data["Survived"].eq(1) & data["Age"].gt(50)).sum()))
print('Total de no sobrevivientes = {}'.format(data["Survived"].eq(0).sum()))
print('Total de no sobrevivientes mujeres = {}'.format((data["Survived"].eq(0) & data["Sex"].eq('female')).sum()))

```

```

Total de sobrevivientes = 342
Total de sobrevivientes mujeres = 233
Total de sobrevivientes hombres = 109
Total de sobrevivientes mayores a 50 años= 22
Total de no sobrevivientes = 549
Total de no sobrevivientes mujeres = 81

```

Tanto el conteo de sobrevivientes en cada una de las clases como el cálculo de la tasa por clase, se automatizó mediante la programación de dos funciones como se muestra en la siguiente imagen.

```

def clase(n):
    N = data["Pclass"].eq(n).sum()
    return (N)

def clase_mujer(n):
    N = data["Pclass"].eq(n).sum()
    Ns = (data["Sex"].eq('female') & data["Pclass"].eq(n)).sum()
    return (Ns/N*100)

print('Número de sobrevivientes por clase \n')
for i in data.Pclass.unique():
    print(' clase {} = {}'.format(i,clase(i)))

print('\n tasa de sobrevivientes por clase siendo mujer en porcentaje \n')

for i in data.Pclass.unique():
    print(' clase {} = {:.2f} %'.format(i,clase_mujer(i)))

```

Número de sobrevivientes por clase

```

clase 3 = 491
clase 1 = 216
clase 2 = 184

```

tasa de sobrevivientes por clase siendo mujer en porcentaje

```

clase 3 = 29.33 %
clase 1 = 43.52 %
clase 2 = 41.30 %

```

```

M_sobrevive = (data["Sex"].eq('female') & data["Survived"].eq(1)).sum()
H_sobrevive = (data["Sex"].eq('male') & data["Survived"].eq(1)).sum()
M_s_Clasel = (data["Sex"].eq('female') & data["Survived"].eq(1) & data["Pclass"]
M_Clasel = (data["Sex"].eq('female') & data["Pclass"].eq(1)).sum()

M = data["Sex"].eq('female').sum()
H = data["Sex"].eq('male').sum()
total = data["Sex"].eq('female').sum() + data["Sex"].eq('male').sum()

print('Probabilidad de mujer que sobrevive {:.2f}'.format(M_sobrevive/total*100))
print('Probabilidad de ser mujer {:.2f}'.format(M/total*100))
print('Probabilidad de sobrevivir dado que se es mujer (usar regla de Bayes) {:.
    .format(M_sobrevive/M*100)
    )
print('Probabilidad de sobrevivir dado que se es hombre (usar regla de Bayes) {:.
    .format(H_sobrevive/H*100)
    )
print('Probabilidad de sobrevivir dado que se es hombre (usar regla de Bayes) {:.
    .format(M_s_Clasel /M_Clasel*100)
    )

```

```

Probabilidad de mujer que sobrevive 26.15
Probabilidad de ser mujer 35.24
Probabilidad de sobrevivir dado que se es mujer (usar regla de Bayes) 74.20
Probabilidad de sobrevivir dado que se es hombre (usar regla de Bayes) 18.89
Probabilidad de sobrevivir dado que se es hombre (usar regla de Bayes) 96.81

```

Las probabilidades se obtienen aplicando el teorema de Bayes, como a continuación se muestra:

Conclusiones

Se llevo a cabo el análisis de los datos resolviendo las preguntas solicitadas como examen.

Bibliografía

datos del titanic

https://raw.githubusercontent.com/mrBronnWow/Curso_Beginners/main/1_Dataset_titanic/train.csv