

# INSTITUTO POLITÉCNICO NACIONAL

# CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN



## **Actividad:**

A . /1	.1 . 1		.1	•	.1 . 1	TD*4
Analisis	aei	reporte	ae	pasajeros	aei	1 itanic.

# Alumno: Francisco Javier Requena Fragoso Curos: Introducción a la ciencia de datos con Python

**Docente:** 

Alan Badillo Salas

Fecha:

15/07/2023

### Introducción

El hundimiento del Titanic en 1912 fue uno de los eventos más trágicos y conocidos en la historia de la navegación. Este desastre marítimo ha sido objeto de numerosos estudios y análisis a lo largo de los años. En este informe, utilizaremos la biblioteca de Python "pandas" para realizar un análisis de los datos de los pasajeros del Titanic. La elección de "pandas" se tomó debido a que es una poderosa herramienta de manipulación y análisis de datos que nos permite explorar y extraer información valiosa de conjuntos de datos complejos, la cual nos permitirá realizar operaciones de filtrado, agrupación y cálculo de estadísticas utilizando un conjunto de datos del Titanic que contiene información sobre los pasajeros, como su género, edad, clase de pasaje y si lograron sobrevivir al naufragio.

Este análisis nos permitirá examinar las tasas de supervivencia en diferentes grupos demográficos y clases de pasaje, así como calcular probabilidades condicionales de acuerdo con algunas características y utilizando la regla de Bayes.

A medida que avancemos en este informe, exploraremos diferentes aspectos relacionados con el género, la edad y la clase de pasaje, proporcionando una visión integral de las circunstancias que rodearon a los supervivientes del Titanic.

### Justificación

El análisis de los datos de los supervivientes del Titanic no solo es un ejercicio interesante desde el punto de vista histórico, sino que también proporciona valiosas lecciones sobre los factores que influyeron en las tasas de supervivencia.

La realización de un análisis detallado de los datos de los supervivientes del Titanic utilizando la biblioteca Pandas tiene como objetivo arrojar luz sobre estos factores y proporcionar una comprensión más profunda de las circunstancias que rodearon a los pasajeros en ese fatídico viaje. Al explorar y examinar los datos de manera sistemática, podremos descubrir patrones y tendencias significativas que nos ayudarán a comprender mejor cómo ciertos factores influyeron en las posibilidades de supervivencia que a su vez nos permitirá obtener una comprensión más profunda de los factores que pueden haber influido en las posibilidades de sobrevivir al desastre.

### **Desarrollo**

Para iniciar con el análisis de los datos primero se necesita obtener la información en este caso el conjunto utilizado fue proporcionado por el profesor una vez descargada esta información ahora si puede ser utilizada. Aunque antes de esto se hizo un análisis visual rápido acerca de las columnas y valores posibles que contenían los registros. Una vez observado esto entonces se empezó a diseñar el programa que obtendría la información relevante que en este caso fueron los siguientes datos:

### • Total de supervivientes

Para obtener el total de supervivientes se creo una lista con los registros de los pasajeros que sobrevivieron ya que esta podría ser utilizada posteriormente para otros cálculos para ello se creo un filtro que devolvía True con los pasajeros que sobrevivieron y posteriormente con el método .loc() podíamos hacer el filtro de estos registros. Y a este data set le aplicábamos el método .shape[0] para obtener el total de pasajeros que sobrevivieron.

```
#Creamos una lista que tenga solo los pasajeros que sobrevivieron
#Para ello primero encontramos las posicicones de las filas donde la
condición se cumple
filtroInfoSurvived = data['Survived'] == 1
#Ahora con el método loc obtenemos los supervivientes pasando la lista
anterior que tiene las filas que cumplieron la condición
Supervivientes = data.loc[filtroInfoSurvived]
#Con la función shape obtenemos la cantidad de supervivientes en nuestro
nuevo dataset
TotalSupervivientes = Supervivientes.shape[0]
```

Para realizar el cálculo de supervivientes creamos una función lambda que de una dataset condicional cuanta las condiciones True que se cumplan esto utilizando .sum()

```
#Creamos una función lambda que cuente los valores True de un dataframe usando la función sum()

ContadorTrue = lambda dataframe: dataframe.sum()
```

- Total de supervivientes mujeres
- Total de supervivientes hombres
- Total de supervivientes mayores a 18 años
- Total de supervivientes menores a 18 años
- Total de supervivientes mayores a 50 años

Con la función lambda que creamos anteriormente calculamos todos los puntos anteriores

```
    #Utilizando la función lambda mandamos la lista para obtener la siguiente información
    TotalSupervivientesHombres = ContadorTrue(Supervivientes['Sex']=='male')
    TotalSupervivientesMujeres = ContadorTrue(Supervivientes['Sex']=='female')
    TotalSupervivientes_Plus18 = ContadorTrue(Supervivientes['Age']>18)
    TotalSupervivientes_Less18 = ContadorTrue(Supervivientes['Age']<18)</li>
    TotalSupervivientes_Plus50 = ContadorTrue(Supervivientes['Age']>50)
```

Ahora para calcular el total de defunciones creamos nuevamente un nuevo dataset que tenga los registros de las personas que no lograron sobrevivir para ello lo hacemos de la misma manera que el primer dataset de supervivientes:

```
#De la misma manera creamos una lista de los pasajeros que no
lograron sobrevivir
filtroInfoNoSurvived = data['Survived'] == 0
Defunciones = data.loc[filtroInfoNoSurvived]
```

Ahora con la misma función lambda que creamos obtenemos la siguiente información.

- Total de muertos
- Total de muertos mujeres
- Total de muertos hombres

Que se puede observar en el código de la siguiente manera:

```
#Nuevamente utilizando la función lambda mandamos la lista para obtener la siguiente información

TotalDefunciones = ContadorTrue(data['Survived'] == 0)

TotalDefuncionesHombres = ContadorTrue(Defunciones['Sex']=='male')

TotalDefuncionesMujeres = ContadorTrue(Defunciones['Sex']=='female')
```

Para calcular la tasa de supervivencia aplicamos filtros de acuerdo a cada clase y calculamos las tasas de supervivencia.

- Tasa de supervivencia de la clase 1
- Tasa de supervivencia de la clase 2
- Tasa de supervivencia de la clase 3
- Tasa de supervivencia de la clase 1 siendo mujer
- Tasa de supervivencia de la clase 2 siendo mujer
- Tasa de supervivencia de la clase 3 siendo mujer

```
SupervivientesC1 = ContadorTrue(Supervivientes['Pclass']==1)

DefuncionesC1 = ContadorTrue(Defunciones['Pclass']==1)

#Obtenemos la tasa con una regla de 3 solo utilizando la población de la clase A = 1

TasaSupervivenciaC1 = (SupervivientesC1*100) / (SupervivientesC1 + DefuncionesC1)

SupervivientesC2 = ContadorTrue(Supervivientes['Pclass']==2)

DefuncionesC2 = ContadorTrue(Defunciones['Pclass']==2)

#Obtenemos la tasa con una regla de 3 solo utilizando la población de la clase B = 2

TasaSupervivenciaC2 = (SupervivientesC2*100) / (SupervivientesC2 + DefuncionesC2)

SupervivientesC3 = ContadorTrue(Supervivientes['Pclass']==3)

#Obtenemos la tasa con una regla de 3 solo utilizando la población de la clase C = 3

TasaSupervivenciaC3 = (SupervivientesC3*100) / (SupervivientesC3 + DefuncionesC3)

#Supervivientes mujeres de la clase A obtenemos una lista que cumpla las condiciones que sea de la clas 1 y que sea mujer y aplicamos un len a la lista resultante para saber:

SupervivientesMujeresC1 = len(Supervivientes[(Supervivientes['Pclass']==1) & (Supervivientes['Sex']=='female')])

TasaSupervivenciaMujeresC1 = (SupervivientesMujeresC1*100) / (SupervivientesC1 + DefuncionesC1)
```

```
#Supervivientes mujeres de la clase B obtenemos una lista que cumpla las condiciones que sea de la clas 1 y que sea mujer y aplicamos un len a la lista resultante para saber:

SupervivientesMujeresC2 = len(Supervivientes[(Supervivientes['Pclass']==2) & (Supervivientes['Sex']=='female')])

TasaSupervivenciaMujeresC2 = (SupervivientesMujeresC2*100) / (SupervivientesC2 + DefuncionesC2)

#Supervivientes mujeres de la clase C obtenemos una lista que cumpla las condiciones que sea de la clas 1 y que sea mujer y aplicamos un len a la lista resultante para saber:

SupervivientesMujeresC3 = len(Supervivientes[(Supervivientes['Pclass']==3) & (Supervivientes['Sex']=='female')])

TasaSupervivenciaMujeresC3 = (SupervivientesMujeresC3*100) / (SupervivientesC3 + DefuncionesC3)
```

Y para el cálculo de probabilidades igual se utilizan mas filtros y los valores utilizados en las formulas para el calculo de probabilidad condicional utilizando la regla de Bayes

- Probabilidad de sobrevivir dado que se es mujer (usar regla de Bayes)
- Probabilidad de sobrevivir dado que se es hombre (usar regla de Bayes)
- Probabilidad de sobrevivir dado que se es mujer y está en la clase 1 (usar regla de Bayes compuesta)

```
#Calculamos la probabilidad de sobrevivir para todos

ProbabilidadSobrevivir = ((TotalSupervivientes*100) / TotalPasajeros)

print(ProbabilidadSobrevivir)

#Calculamos la probabilidad de ser mujer en el total de supervivientes

ProbabilidadMujerSobrevive = ((ContadorTrue(Supervivientes['Sex']=='female')*100) /

TotalSupervivientes)

#Calculamos la probabilidad de ser mujer que en el total de las defunciones

ProbabilidadMujerNoSobrevive = ((ContadorTrue(Defunciones['Sex']=='female')*100) / TotalDefunciones)

#Calculamos la probabilidad conficional con la regla de Bayes

ProbabilidadSobrevivirDadoMujer = ((ProbabilidadMujerSobrevive * ProbabilidadSobrevivir) /
((ProbabilidadMujerSobrevive * ProbabilidadSobrevivir) + ((100-

ProbabilidadSobrevivir)*ProbabilidadMujerNoSobrevive)))*100

#Calculamos la probabilidad de ser hombre en el total de supervivientes

ProbabilidadHombreSobrevive = ((ContadorTrue(Supervivientes['Sex']=='male')*100) / TotalDefunciones)

#Calculamos la probabilidad de ser mujer en el total de defunciones

ProbabilidadHombreNoSobrevivir = ((ContadorTrue(Defunciones['Sex']=='male')*100) / TotalDefunciones)
```

```
# Calculamos la probabilidad de sobrevivir dado que se es hombre utilizando la regla de Bayes

ProbabilidadSobrevivirDadoHombre = (ProbabilidadHombreSobrevive * ProbabilidadSobrevivir) /

((ProbabilidadHombreSobrevive * ProbabilidadSobrevivir) + ((100-

ProbabilidadSobrevivir)*ProbabilidadHombreNoSobrevivir))*100

## Calculamos la probabilidad de ser mujer y estar en la clase 1 que sovrevivio

ProbabilidadMujerSobreviveClase1 = ContadorTrue((Supervivientes['Sex'] == 'female') &

(Supervivientes['Pclass'] == 1)) / TotalSupervivientes

## Calculamos la probabilidad de ser mujer y estar en la clase 1 que no sobrevivio

ProbabilidadMujerNoSobreviveClase1 = ContadorTrue((Defunciones['Sex'] == 'female') &

(Defunciones['Pclass'] == 1)) / TotalDefunciones

# Calculamos la probabilidad de sobrevivir dado que se es mujer de la clase 1 utilizando la regla de Bayes

ProbabilidadSobrevivirDadoMujerClase1 = (ProbabilidadMujerSobreviveClase1 * ProbabilidadSobrevivir)

/ ((ProbabilidadMujerSobreviveClase1 * ProbabilidadSobrevivir) + ((100-

ProbabilidadSobrevivir)*ProbabilidadMujerNoSobreviveClase1))*100
```

Por ultimo utilizando las variables creadas y el metodo print() y .fomat() desplegamos la información obtenida en pantalla.

```
print("Supervivientes\n")
print("Total de pasajeros: {:d}".format(TotalPasajeros))
print("Sobrevivientes totales: {:d}".format(TotalSupervivientes))
print(\Sobrevivientes - Hombres: {:d}\".format(\TotalSupervivientesHombres))
print(\"Sobrevivientes - Hombres: {:d}\".format(\TotalSupervivientesHombres))
print(\"Sobrevivientes - Mujeres: {:d}\".format(\TotalSupervivientesMujeres))
print(\"Sobrevivientes - Mayores 18: {:d}\".format(\TotalSupervivientes_Plus18))
print(\"Sobrevivientes - Menores 18: {:d}\".format(\TotalSupervivientes_Less18))
print(\"Sobrevivientes - Mayores 50: {:d}\".format(\TotalSupervivientes_Plus50))
print("Fallecimientos totales: {:d}".format(TotalDefunciones))
print("Fallecidos - Hombres: {:d}".format(TotalDefuncionesHombres))
print("Fallecidos - Mujeres: {:d}".format(TotalDefuncionesMujeres))
print("-----
print("Tasas de superviviencia\n")
print("Tasa de superviviencia de la clase 1: {:.2f}% ".format(TasaSupervivenciaC1))
print("Tasa de superviviencia de la clase 2: {:.2f}% ".format(TasaSupervivenciaC2))
print("Tasa de superviviencia de la clase 3: {:.2f}% ".format(TasaSupervivenciaC3))
print("\nTasas de superviviencia mujeres:\n")
print("Tasa de superviviencia de la clase 1 - Mujeres: {:.2f}% ".format(TasaSupervivenciaMujeresC1))
print("Tasa de superviviencia de la clase 2 - Mujeres: {:.2f}% ".format(TasaSupervivenciaMujeresC2))
print("Tasa de superviviencia de la clase 3 - Mujeres: {:.2f}% ".format(TasaSupervivenciaMujeresC3))
 print("---
print("Probabilidad de superviviencia:\n")
print("Probabilidad de sobrevivir dado que se es mujer: {:.2f}%".format(ProbabilidadSobrevivirDadoMujer))
print("Probabilidad de sobrevivir dado que se es hombre: {:.2f}%".format(ProbabilidadSobrevivirDadoHombre))
 print("Probabilidad de sobrevivir dado que se es mujer y es de la clase 1 : {:.2f}%".format(ProbabilidadSobrevivirDadoMujerClase1)
```

# Capturas de pantalla (Resultados)

```
Reporte de supervivientes Titanic
Supervivientes
Total de pasajeros: 891
Sobrevivientes totales: 342
Sobrevivientes - Hombres: 109
Sobrevivientes - Mujeres: 233
Sobrevivientes - Mayores 18: 220
Sobrevivientes - Menores 18: 61
Sobrevivientes - Mayores 50: 22
Fallecidos
Fallecimientos totales: 549
Fallecidos - Hombres: 468
Fallecidos - Mujeres: 81
Tasas de superviviencia
Tasa de superviviencia de la clase 1: 62.96%
Tasa de superviviencia de la clase 2: 47.28%
Tasa de superviviencia de la clase 3: 24.24%
Tasas de superviviencia mujeres:
Tasa de superviviencia de la clase 1 - Mujeres: 42.13%
Tasa de superviviencia de la clase 2 - Mujeres: 38.04%
Tasa de superviviencia de la clase 3 - Mujeres: 14.66%
Probabilidad de superviviencia:
Probabilidad de sobrevivir dado que se es mujer: 74.20%
Probabilidad de sobrevivir dado que se es hombre: 18.89%
Probabilidad de sobrevivir dado que se es mujer y es de la clase 1 : 96.81%
```

1 Resultados del análisis de datos

### **Conclusiones**

El análisis de los datos de supervivientes del Titanic utilizando la librería Pandas en Python nos ha permitido obtener diversas conclusiones significativas. Se identificó que la cantidad total de pasajeros analizados fue una cantidad considerable considerando más de un 50% de fallecimientos del total de pasajeros además de que se observó una clara diferencia en las tasas de supervivencia entre hombres y mujeres, donde las mujeres tuvieron una probabilidad de sobrevivir significativamente mayor en comparación con los hombres. Esto indica que el género fue un factor determinante en la probabilidad de supervivencia en el desastre del Titanic y esto era debido a los protocolos de emergencia que tenían al momento de un incidente en donde se consideraba mas a las mujeres y los niños.

Además, se analizó la influencia de la clase en las tasas de supervivencia. Se encontró que los pasajeros de primera clase tuvieron una tasa de supervivencia más alta en comparación con las clases inferiores. Durante el proceso de análisis, se enfrentaron desafíos al manipular y filtrar los datos para obtener la información específica requerida. Sin embargo, gracias al uso de la librería Pandas, se superaron estos obstáculos y se logró realizar el análisis de manera efectiva.

En resumen, este análisis destaca la importancia del género y la clase en las probabilidades de supervivencia en el desastre del Titanic. Los resultados obtenidos proporcionan información valiosa sobre los factores que influyeron en la supervivencia de los pasajeros y podrían ayudar a comprender mejor la historia y las circunstancias de ese trágico evento.

# Anexos (Código):

```
import pandas as pd
data = pd.read_csv('./dataset/titanic.csv')
ContadorTrue = lambda dataframe: dataframe.sum()
TotalPasajeros = len(data['Survived'])
filtroInfoSurvived = data['Survived'] == 1
Supervivientes = data.loc[filtroInfoSurvived]
TotalSupervivientes = Supervivientes.shape[0]
TotalSupervivientesHombres = ContadorTrue(Supervivientes['Sex']=='male')
TotalSupervivientesMujeres = ContadorTrue(Supervivientes['Sex']=='female')
TotalSupervivientes Plus18 = ContadorTrue(Supervivientes['Age']>18)
TotalSupervivientes_Less18 = ContadorTrue(Supervivientes['Age']<18)</pre>
TotalSupervivientes_Plus50 = ContadorTrue(Supervivientes['Age']>50)
filtroInfoNoSurvived = data['Survived'] == 0
Defunciones = data.loc[filtroInfoNoSurvived]
TotalDefunciones = ContadorTrue(data['Survived'] == 0)
TotalDefuncionesHombres = ContadorTrue(Defunciones['Sex']=='male')
TotalDefuncionesMujeres = ContadorTrue(Defunciones['Sex']=='female')
```

```
SupervivientesC1 = ContadorTrue(Supervivientes['Pclass']==1)
DefuncionesC1 = ContadorTrue(Defunciones['Pclass']==1)
TasaSupervivenciaC1 = (SupervivientesC1*100) / (SupervivientesC1 + DefuncionesC1)
SupervivientesC2 = ContadorTrue(Supervivientes['Pclass']==2)
DefuncionesC2 = ContadorTrue(Defunciones['Pclass']==2)
TasaSupervivenciaC2 = (SupervivientesC2*100) / (SupervivientesC2 + DefuncionesC2)
SupervivientesC3 = ContadorTrue(Supervivientes['Pclass']==3)
DefuncionesC3 = ContadorTrue(Defunciones['Pclass']==3)
TasaSupervivenciaC3 = (SupervivientesC3*100) / (SupervivientesC3 + DefuncionesC3)
SupervivientesMujeresC1 = len(Supervivientes[(Supervivientes['Pclass']==1) &
(Supervivientes['Sex']=='female')])
TasaSupervivenciaMujeresC1 = (SupervivientesMujeresC1*100) / (SupervivientesC1 + DefuncionesC1)
SupervivientesMujeresC2 = len(Supervivientes[(Supervivientes['Pclass']==2) &
(Supervivientes['Sex']=='female')])
TasaSupervivenciaMujeresC2 = (SupervivientesMujeresC2*100) / (SupervivientesC2 + DefuncionesC2)
SupervivientesMujeresC3 = len(Supervivientes[(Supervivientes['Pclass']==3) &
(Supervivientes['Sex']=='female')])
TasaSupervivenciaMujeresC3 = (SupervivientesMujeresC3*100) / (SupervivientesC3 + DefuncionesC3)
ProbabilidadSobrevivir = ((TotalSupervivientes*100) / TotalPasajeros)
ProbabilidadMujerSobrevive = ((ContadorTrue(Supervivientes['Sex']=='female')*100) /
TotalSupervivientes)
```

```
ProbabilidadMujerNoSobrevive = ((ContadorTrue(Defunciones['Sex']=='female')*100) / TotalDefunciones)
ProbabilidadSobrevivirDadoMujer = ((ProbabilidadMujerSobrevive * ProbabilidadSobrevivir) /
((ProbabilidadMujerSobrevive * ProbabilidadSobrevivir) + ((100-
ProbabilidadSobrevivir)*ProbabilidadMujerNoSobrevive)))*100
ProbabilidadHombreSobrevive = ((ContadorTrue(Supervivientes['Sex']=='male')*100) /
TotalSupervivientes)
ProbabilidadHombreNoSobrevivir = ((ContadorTrue(Defunciones['Sex']=='male')*100) / TotalDefunciones)
ProbabilidadSobrevivirDadoHombre = (ProbabilidadHombreSobrevive * ProbabilidadSobrevivir) /
((ProbabilidadHombreSobrevive * ProbabilidadSobrevivir) + ((100-
ProbabilidadSobrevivir)*ProbabilidadHombreNoSobrevivir))*100
ProbabilidadMujerSobreviveClase1 = ContadorTrue((Supervivientes['Sex'] == 'female') &
(Supervivientes['Pclass'] == 1)) / TotalSupervivientes
ProbabilidadMujerNoSobreviveClase1 = ContadorTrue((Defunciones['Sex'] == 'female') &
(Defunciones['Pclass'] == 1)) / TotalDefunciones
ProbabilidadSobrevivirDadoMujerClase1 = (ProbabilidadMujerSobreviveClase1 * ProbabilidadSobrevivir)
/ ((ProbabilidadMujerSobreviveClase1 * ProbabilidadSobrevivir) + ((100-
ProbabilidadSobrevivir)*ProbabilidadMujerNoSobreviveClase1))*100
print("-----")
print("
                 Reporte de supervivientes Titanic")
print("-----")
print("Supervivientes\n")
print("Total de pasajeros: {:d}".format(TotalPasajeros))
print("Sobrevivientes totales: {:d}".format(TotalSupervivientes))
print("Sobrevivientes - Hombres: {:d}".format(TotalSupervivientesHombres))
print("Sobrevivientes - Mujeres: {:d}".format(TotalSupervivientesMujeres))
print("Sobrevivientes - Mayores 18: {:d}".format(TotalSupervivientes_Plus18))
print("Sobrevivientes - Menores 18: {:d}".format(TotalSupervivientes Less18))
```

```
print("Sobrevivientes - Mayores 50: {:d}".format(TotalSupervivientes_Plus50))
print("-----")
print("Fallecidos\n")
print("Fallecimientos totales: {:d}".format(TotalDefunciones))
print("Fallecidos - Hombres: {:d}".format(TotalDefuncionesHombres))
print("Fallecidos - Mujeres: {:d}".format(TotalDefuncionesMujeres))
print("-----")
print("Tasas de superviviencia\n")
print("Tasa de superviviencia de la clase 1: {:.2f}% ".format(TasaSupervivenciaC1))
print("Tasa de superviviencia de la clase 2: {:.2f}% ".format(TasaSupervivenciaC2))
print("Tasa de superviviencia de la clase 3: {:.2f}% ".format(TasaSupervivenciaC3))
print("\nTasas de superviviencia mujeres:\n")
print("Tasa de superviviencia de la clase 1 - Mujeres: {:.2f}% ".format(TasaSupervivenciaMujeresC1))
print("Tasa de superviviencia de la clase 2 - Mujeres: {:.2f}% ".format(TasaSupervivenciaMujeresC2))
print("Tasa de superviviencia de la clase 3 - Mujeres: {:.2f}% ".format(TasaSupervivenciaMujeresC3))
print("-----")
print("Probabilidad de superviviencia:\n")
print("Probabilidad de sobrevivir dado que se es mujer:
{:.2f}%".format(ProbabilidadSobrevivirDadoMujer))
print("Probabilidad de sobrevivir dado que se es hombre:
{:.2f}%".format(ProbabilidadSobrevivirDadoHombre))
print("Probabilidad de sobrevivir dado que se es mujer y es de la clase 1 :
{:.2f}%".format(ProbabilidadSobrevivirDadoMujerClase1))
```