

TP 1 – EXPLORACIÓN, LIMPIEZA Y ANÁLISIS DE DATOS

El presente trabajo se deberá realizar en Python.

Las consignas son las siguientes:

- 1) *Explorar* el dataset del Titanic. Explicar la cantidad de filas y columnas que tiene el mismo, cuantas son descriptivas, cuantas son numéricas. ¿Existen valores vacíos?
Responder las siguientes preguntas:
 - ¿Cuánta gente sobrevivió y cuánta gente no?
 - ¿Cuántos hombres y cuántas mujeres se tiene en la base de datos?
- 2) *Proceso de limpieza*. Todo dataset debe ser limpiado y preparado para realizar un análisis. Por lo general existen columnas vacías, filas con datos faltantes, columnas de mayor interés que otras. ¿Que trabajo realizarías? ¿Existen valores faltantes? Completar los “vacíos”, removerlos o en caso de incluirlos explicar por qué se dejan en el dataset.
- 3) *Visualización y análisis de los datos*. Si quisiéramos predecir que personas sobrevivieron y que personas no sobrevivieron al accidente, ¿Con qué variables te quedarías? Explicar. ¿Existe correlación entre algunas variables numéricas?. Realizar al menos 3 gráficos que sirvan para entender mejor la base de datos y que permitan luego realizar un análisis más en detalle (¿Te animás a mejorar el formato del gráfico?)
- 4) **(OPCIONAL)**. ¿Hay relación entre el título de cada pasajero y su supervivencia? ¿El capitán sobrevive al accidente? Responder a partir de los datos.
Pista: Cada pasajero tiene un título asociado. Por ejemplo, el título del capitán es “Capt.”

El trabajo práctico se deberá entregar con:

- Un informe o presentación (PDF o PPT) de lo realizado. Ser concreto pero incluir todo el análisis realizado.
- El código realizado en Python

Comentario: No quedarse solo con las consignas presentadas. Todo análisis extra que se realice será tenido en cuenta.

Explicación de las variables

Variable	Definition	Key
survival	Survival	0 = No, 1 = Yes
pclass	Ticket class	1 = 1st, 2 = 2nd, 3 = 3rd
sex	Sex	
Age	Age in years	
sibsp	# of siblings / spouses aboard the Titanic	
parch	# of parents / children aboard the Titanic	
ticket	Ticket number	
fare	Passenger fare	
cabin	Cabin number	
embarked	Port of Embarkation	C = Cherbourg, Q = Queenstown, S = Southampton