

Introducción a la Ciencia de Datos - 2C 2022

Guía extra de Trabajos Prácticos N° 2

Les proponemos comenzar a analizar dos set de datos ubicados en el Campus de la materia a partir de una serie de consignas asociadas a cada uno..

Parte 2. Análisis dataset titanic

En el Campus de la materia van a encontrar el dataset titanic que contiene una muestra de los pasajeros que viajaban en el RMS Titanic en abril del año 1912.

1. Realizar un análisis de los y las tripulantes del barco elaborando un gráfico de barras para cada caso. Estudiar según:
 - a. Sexo
 - b. Cantidad de hermanos y padres o hijos en el barco
 - c. Puerto donde embarcaron
 - d. Clase del boleto
 - e. Hombre/Mujer/Niñx (variable “who”)
2. Mostrar en un gráfico de barras la cantidad de personas que viajaban solas y las que viajaban acompañadas desagregando por las variables “class” y “who”.
3. Realizar un gráfico de barras que indique la cantidad de pasajeros hombres, mujeres y niñxs (de forma desagregada) que viajaban en cada clase.
4. Hacer el mismo gráfico de 3. pero ahora expresado en porcentajes.
5. ¿Cuál fue el tipo de población que tuvo una mayor fracción de supervivencia teniendo en cuenta la variable “who” y “class”? Realizar el gráfico en el cual se puede encontrar esa respuesta.
6. Estudiar el porcentaje de sobrevivientes para cada ciudad de embarque. Hay una ciudad que tiene aproximadamente un 20% más de sobrevivientes que las otras dos. ¿Cómo pueden explicar esto?
Pista: leer al final de la guía.
7. ¿Los hermanos se hunden o nadan juntos? Hacer un gráfico donde se muestre la supervivencia según la cantidad de hermanos que tenía cada pasajero. ¿Qué se observa?

Pista Parte 2. 6.: Estudiar el porcentaje poblacional en la ciudad en cuestión y compararlo con el de las otras ciudades, teniendo en cuenta lo observado en el punto 5.