

Examen-1-Parcial-MODEL-2.pdf



Eng_dades



Visualització de Dades



3º Grado en Ingeniería de Datos



Escuela de Ingeniería
Universidad Autónoma de Barcelona

antes



**Descarga sin publi
con 1 coin**



Después

WUOLAH



Importante

Puedo eliminar la publi de este documento con 1 coin

¿Cómo consigo coins? → Plan Turbo: barato
→ Planes pro: más coins

pierdo espacio



Necesito concentración

ali ali ooh
esto con 1 coin me
lo quito yo...

WUOLAH

Visualització de Dades (Enginyeria de Dades – EE - UAB)
Examen Primer Parcial – 29 Març 2022
MODEL 2

Nom i Cognoms: _____

NIU: _____ Grup de Matrícula: _____

PARTE 1 (3 pt)

Dataset: *titanic.csv*

1.1. (1 pt) Abre el fichero. ¿Qué tipo de atributo son: *Sex*, *Age*, *Survived*, *Pclass* y *Fare*?
¿Qué atributo es el key (clave primaria) del dataset?

RESPOSTA:

1.2. (2 pt) Queremos saber si la clase en la que viajaban los pasajeros (*Pclass*) influye en la probabilidad de sobrevivir (*Survived*). Haz una gráfica que permita visualizar esa relación y razona tu elección de acuerdo al framework Datos/Tareas/Codificación (Cuántos atributos usas y de qué tipo son; qué tarea ayuda a llevar a cabo la gráfica; marcas y canales empleados).

Sube la imagen de la gráfica debidamente anotada junto a la respuesta.

RESPOSTA:

PARTE 2 (2 pt)

Dataset: *life_expectancy.csv*

Crea un subset con todos los países para el año 2011, que contenga los valores de *Male* y *Female* (usando pipes).

2.1. ¿Cual sería la forma óptima de visualizar la diferencia en Esperanza de Vida entre Male y Female por país? Razona tu respuesta en relación al framework Datos/Tareas/Codificación. (1 pt)

RESPOSTA:

2.2. Haz la gráfica y sube una imagen de los 20 países con la mayor diferencia entre hombres y mujeres. La gráfica debe tener todas las etiquetas y anotaciones necesarias para ser legible y comprensible. (1 pt)

RESPOSTA:

PARTE 3 (5 pt)

Dataframe: iris de R

Iris proporciona las medidas (en cm) de las variables longitud y anchura de los sépalos y de los pétalos respectivamente para 50 flores de cada una de las 3 especies de Iris (150 en total). Las especies de iris son: la Versicolor, la Virginica i la Setosa.



Iris es un *dataframe* con 150 observaciones (filas) y 5 variables (columnas) cuyos nombres son: Sepal.Length, Sepal.Width, Petal.Length, Petal.Width y Species. Los valores de las variables referentes a las respectivas longitudes y anchuras (métricas) están en centímetros.

NOTA: En los ejercicios de esta parte, hacer uso de las *pipes* y añadir al gráfico un título y etiquetas personalizadas en los ejes x e y.

3.1 (2 pt)

- Haz un gráfico que muestre la distribución de la longitud de los pétalos por especie. Deben verse claramente los cuartiles y la mediana. (0.75 puntos)
- Haz un multipanel en el que se muestren las distribuciones de las 4 métricas (una en cada panel) para las diferentes especies. En el momento de mostrar las distribuciones, utiliza un tipo de gráfico distinto al que hayas utilizado en el apartado anterior. (1.25 puntos)

RESPOSTA:

3.2. Data massage (1.5 pt)

- ¿Cuál es la longitud máxima y mínima de los sépalos? (0.25pts)
- Normaliza la anchura del sépalo por su media y por su mediana respectivamente y muestra ambas distribuciones en un mismo gráfico. (1.25 pt)

RESPOSTA:

3.3. (1.5pt) Haz un gráfico para cada especie en el que se vea la forma de la distribución de la amplitud del sépalo. Preséntalos de forma que se puedan comparar entre ellos.

RESPOSTA: