

PROGRAMACION para CIENCIA DE DATOS

EXAMEN FINAL INTEGRADOR

Tomando como base el Ejercicio práctico realizado en el TP 5 con el lenguaje R, mejorar el mismo incorporando más herramientas y complejidad. Es decir este trabajo final se enfocará solamente en el lenguaje R. En los requisitos del trabajo hay algunos que requieren una investigación de su lado (identificados como “investigar y aplicar”).

Si por cualquier motivo desea cambiar el tema de su trabajo y que no tenga relación con el TP 5 mencionado, puede hacerlo, pero yo le recomiendo que sí lo haga a partir de lo ya desarrollado. El ejercicio debe poseer tanto su enunciado claro como su programación en R. El trabajo es totalmente individual.

Poner en el programa una primera línea de comentario con su Apellido y Nombre. Subir 1 solo archivo PDF que contenga muy clara y prolijamente el enunciado, el código y muestras de ejecución satisfactoria del programa. Se valorará prolijidad y dedicación en la presentación, no solo en la capacidad de codificación. Si desea, sume más archivos que haya utlizado.

El programa debe contener al menos:

* Uso de Dataset.
* Importación desde un archivo CSV.
* Visualización de datos en gráficos con colores.
* Aplicar un Gráfico de Clústeres (investigar y aplicar)
* Al menos un análisis numérico con Frecuencias - Al menos un análisis numérico con Descriptivos - Uso de Correlaciones.

(Tener en cuenta que en el TP había que usar alguno de estos últimos 3, acá hay que utilizar los 3).

* Aplicar Regresión Lineal (investigar y aplicar)
* Aplicar Tablas de Contingencia (investigar y aplicar)

Cantidad de líneas de código mínimo: 250.

Haciendo uso del lenguaje R y el dataset generado en el TP N° 3, realice lo siguiente:

1. Importe y lectura del csv.
2. Cree funciones para ayudarle a optimizar su trabajo.
3. Filtrar los teléfonos que tienen soporte para NFC.
4. Calcular el precio mínimo, promedio y máximo de los teléfonos con soporte para NFC.
5. Graficar el TOP 10 de celulares con mayor precio.
6. Graficar el TOP 5 de celulares con menor precio
7. Graficar la distribución de los tipos de memorias.
8. Graficar el TOP 5 de baterías más utilizadas.
9. Realiza un análisis descriptivo de la columna ‘precio’ y compara los resultados con los del punto 4.
10. Realiza un análisis numérico con frecuencias de la columna ‘ram’.

Fuentes posibles de investigación:

Gráfico de Clústeres https://rpubs.com/lhromeroj/analisisdeclusterR https://www.youtube.com/watch?v=7AFuL-1Q8eg

https://medium.com/@gloria.yantas.alcantara/tenemos-mucho-encom%C3%BAn-introducci%C3%B3n-al-cluster-analysis-con-r-c8fdedab7206 Regresión Lineal

 https://rpubs.com/joser/RegresionSimple https://www.datacamp.com/es/tutorial/linear-regression-R https://estadisticamente.com/como-hacer-regresion-lineal-en-r/ Tablas de Contingencia https://rpubs.com/gustavomtzv/690223 https://www.datacamp.com/es/tutorial/contingency-tables-r https://martinpaladino.github.io/rsociales/analisis-de-datos-categoricos.html

¡¡EXITOS!!

Ing. Mario Martínez