|  |  |  |
| --- | --- | --- |
| https://www.cuc.ac.cr/app/cms/www/images/logo_cuc_color.png | **Colegio Universitario de Cartago** |  |
| **BIG DATA (BD)** |  |

**I Cuatrimestre 2025**

**Lenguaje de Minería de Datos**

**BD-123**

**Proyecto 10%**

**Fecha de entrega: viernes 25 de abril 2025 (presentación en clase)**

**INSTRUCCIONES**:

- El proyecto consiste en realizar un análisis exploratorio en R con las siguientes pautas:

1. El estudiante debe aportar el set de datos: puede tomar cualquier set de datos de interés, puede ser un set de datos numérico, categórico o ambos. Plantee al menos 3 hipótesis o preguntas de lo que quisiera resolver o encontrar en los datos.
2. Realice limpieza de datos en caso de ser necesario.
3. Obtenga las estadísticas básicas del set de datos: moda, media, mediana.
4. Realice al menos 2 correlaciones entre las variables del set de datos, muestre gráficos y explique con sus palabras cual es la correlación que existe.
5. Identifique los valores atípicos(outliers) por medio del grafico Box Plot, explique su análisis.
6. Aplique 2 algoritmos: no supervisado clustering y cualquier otro a escogencia visto en clase para encontrar algún patrón y/o predicción entre los datos. Explique su análisis.

Consideraciones

* Entregar un documento en PDF con el resultado para cada uno de los puntos anteriores, debe asumir que usted es contratado y debe mostrar un informe a la gerencia del resultado del análisis exploratorio y la aplicación de algortimos, debe llevar recomendaciones y conclusiones.
* Entregar el script de R.
* Grupos de 2 personas máximo.
* Debe presentar el proyecto en clase por medio de una presentación.
* Puntos extra: Generar informe en Markdown por medio de la librería R Markdown y exportarlo a formato HTML, subir proyecto a RPUBS (plataforma de publicación gratuita proporcionada por RStudio)