

## Colegio Universitario de Cartago

### Tecnologías de Información – Big Data BD-142

Examen Final, Exploración de Datos.

Profesor Ronald Córdoba Vega, fecha de entrega 8 de diciembre 2025 en el CUCVirtual, antes de las 6pm. La defensa del análisis se realizará ese mismo día en horario de clase.

Rúbrica:

Set de datos 7%, descripción completa del Set de datos original a utilizar.

Documento HTML 15%, Debe incluir la lectura de los datos, limpieza de éstos y un EDA completo para los mismos.

Defensa 5%, Todos deben participar en la defensa del proyecto. Deben aportar un archivo con su presentación final en PowerPoint.

Participación del foro 3%, es necesario su participación en esta actividad para obtener los puntos indicados. Esto es individual.

Total 30%

Deben realizar el análisis del set de datos aprobado para este efecto, incluir en el análisis:

1. Carga datos
2. Descripciones de las variables, estadísticas
3. Detalles sobre las mejores variables.
4. visualización de gráficos
5. código Jambú
  - Enlace donde pueden encontrar un ejemplo: [https://rstudio-pubs-static.s3.amazonaws.com/334041\\_fddc996d240a4533baba961f9fe46687.html](https://rstudio-pubs-static.s3.amazonaws.com/334041_fddc996d240a4533baba961f9fe46687.html)
6. Dendogramas
7. PCA
  - Enlace donde pueden encontrar un ejemplo: [https://rpubs.com/cristina\\_gil/pca](https://rpubs.com/cristina_gil/pca)
8. clusterización k-means
  - Enlace donde pueden encontrar un ejemplo: <https://rpubs.com/orodriguez/13318>

**Colegio Universitario de Cartago**  
**Tecnologías de Información – Big Data BD-142**

**A- Grupo #1**

402240016	MONTERO VARGAS LUIS DIEGO
305270837	ROJAS LOPEZ NADIN JAZMIN
305590965	COTO GUEVARA ROBERTO
305100511	CORDERO NAVARRO DENIS FAURICIO
Set de datos	Product-Sales-Region.xlsx

**B- Grupo #2**

305630546	VIQUEZ GUZMAN MARIANA
305650504	DELGADO LOAIZA MARIA PAUBLA
305480836	CHINCHILLA SAENZ VALERIA DE LOS ANGELES
303650023	BRENES LORIA FABIAN
Set de datos	global_disaster_response_2018_2024.csv

**C- Grupo #3**

116040603	MENDEZ PEREZ MARIANA
207260607	RODRIGUEZ JIMENEZ CLARET CHARLOTTE
119780239	VIQUEZ RIVERA MARISOL
305600228	ZAMORA REYES STEVEN PAUL
Set de datos	ObesityDataSet_raw_and_data_sinthetic.csv