

Curso: BD -142 Exploración de datos Grupo 1

Nota

obtenida:

Prueba: II Prueba parcial

Puntos

Puntaje total de la Prueba: 100 puntos

obtenidos:

Porcentaje de la Prueba: 20%

Tiempo disponible: 90 minutos (1.5 horas)

Porcentaje

Fecha de aplicación: 10 noviembre 2025

obtenido:

Profesor: Ronald Córdoba Vega

Estudiante: \_\_\_\_\_ Cédula: \_\_\_\_\_

### **Instrucciones.**

- Sea claro y ordenado al dar sus respuestas.
- Debe entregar en el CUCVirtual un archivo en Word con su **nombre completo\_IIParcial\_Teorico.docx** que contenga portada y sus respuestas para la parte teórica de su prueba en la carpeta IIParcial\_Parte\_teorica.
- La parte práctica debe subir al CUCVirtual.
  - En la carpeta IIParcial\_Parte\_Practica, un archivo con sus iniciales ejemplo:  
**RCV\_IIParcial\_Practico.html** que incluya portada, el código y los resultados de lo solicitado.
  - Entregar este enunciado con su nombre y cédula al profesor.
- Lea cuidadosamente las instrucciones y las preguntas de su examen, si tiene alguna duda dispondrá de 10 minutos antes de iniciar el examen para realizar consultas de forma al profesor.

### **Parte teórica (50 pts. - 5 pts. c/u)**

- 1) ¿Qué transforma la minería de datos?
- 2) ¿Qué permite hacer el conocimiento extraído?
- 3) ¿Qué etapas incluye el proceso de minería de datos?
- 4) ¿Qué sectores aplican minería de datos?
- 5) ¿Qué implica la comprensión del negocio?
- 6) ¿Qué se hace en la comprensión de los datos?
- 7) ¿Qué es la ciencia de datos?
- 8) ¿Qué se hace en el análisis exploratorio (EDA)?
- 9) ¿Qué hace un científico de datos?
- 10) ¿Qué técnicas claves se aplican?
- 11) ¿Qué es un individuo?
- 12) ¿Qué es el código disyuntivo completo?
- 13) Explique Ciencia de datos vs. ingeniería de datos.
- 14) ¿Qué desafíos enfrenta la ciencia de datos?
- 15) ¿Qué tipos de variables existen?

### **Parte práctica (50 pts.)**

- A. Cargue el set de datos "SAheart.csv" en RStudio.
- B. Realice un resumen que contenga lo siguiente:(50 pts.)
  1. Estadísticas básica
  2. Histograma
  3. Boxplot
  4. Barras
  5. QQline y qqnorm
  6. Shapiro