

# Proyecto Final de Bases de Datos

## Comunidad Virtual

### Curso de Bases de Datos

## Objetivo general

Desarrollar una base de datos funcional para gestionar información de una comunidad virtual utilizando herramientas en RStudio. El proyecto incluye tareas de diseño, implementación, manipulación de datos y análisis básico. Se debe investigar el uso de la herramienta RMarkdown para su realización.

## Descripción del proyecto

1. **Contexto:** Una comunidad virtual de estudiantes desea registrar información sobre sus integrantes y actividades. Los datos incluyen:
  - Información personal: Nombre, edad, género, correo electrónico.
  - Actividades realizadas: Título de la actividad, fecha, duración, tipo (académica, recreativa, social).
  - Interacciones: Relación entre estudiantes y actividades (¿quién participó en qué actividad?).
2. **Entregables:**
  - Diseño de la base de datos.
  - Archivo R con el código de creación, llenado y manipulación de la base de datos.
  - Reporte final en PDF generado en RMarkdown, incluyendo visualizaciones y análisis básicos.

## Fases del proyecto

### Semana 1: Diseño y creación de la base de datos

- **Diseño:**
  - Identificar las tablas necesarias y definir las relaciones entre ellas.
  - Crear un diagrama de la estructura de la base de datos.

- **Implementación en R:**

- Escribir el código para crear las tablas en formato `data.frame`.
- Llenar las tablas con datos simulados (al menos 50 estudiantes y 100 actividades).

## **Semana 2: Manipulación, análisis y predicción**

- **Manipulación:**

- Consultas básicas, como:
  1. Lista de estudiantes que participaron en más de 5 actividades.
  2. Actividades más populares por tipo.
  3. Porcentaje de estudiantes por género en cada tipo de actividad.
- Agregar columnas calculadas, como la duración promedio de actividades por estudiante.

- **Visualización:**

- Generar gráficos (barras, pastel, líneas) para mostrar tendencias clave.

- **Análisis predictivo y cálculo de probabilidades:**

- **Preparación de datos:**

- \* Crear una tabla resumen que combine las características del estudiante con su historial de actividades.
- \* Definir una variable objetivo (por ejemplo, “Participa en actividades recreativas” como un valor binario).

- **Modelo predictivo:**

- \* Implementar un modelo de regresión logística utilizando el paquete `caret` o `glm` en R.
- \* Entrenar el modelo y calcular las probabilidades de participación para cada estudiante.

- **Evaluación del modelo:**

- \* Dividir los datos en conjunto de entrenamiento y prueba.
- \* Reportar métricas como la precisión, el recall y el F1-score.

- **Resultados y visualizaciones:**

- \* Mostrar las probabilidades estimadas en un gráfico (por ejemplo, histograma o diagrama de cajas).
- \* Realizar un análisis de cómo las características demográficas afectan la probabilidad de participación.

- **Reporte:**

- Crear un reporte en RMarkdown explicando el proceso, incluyendo:
  - \* Introducción y contexto.
  - \* Diseño de la base de datos.
  - \* Resultados de las consultas y análisis.
  - \* Conclusiones.

## Evaluación

1. **Documentación (20%)**: Presentación clara y completa de los datos y procesos.
2. **Código en R (50%)**: Funcionalidad, limpieza y cumplimiento de las especificaciones.
3. **Reporte final (30%)**: Claridad, contenido y calidad visual.