

# Proyecto Final

**Valor 30%**

## 1. Descripción

Se debe plantear un problema de negocio (real o ficticio) que pueda resolverse mediante técnicas de *visto a lo largo de la especialidad*, ya sean predictivas o descriptivas.

Tanto el problema como su solución inicial serán realizados siguiendo la **metodología CRISP-DM** vista en clase. Se documentarán las fases y actividades realizadas.

Para el desarrollo de las actividades podrá utilizar el lenguaje de **programación R** o bien **Python**. El *dashboard* deberá contener información relevante del *dataset* utilizado y se implementará con **Tableau Desktop**.

## 2. Objetivos

- Plantear un problema de negocio para ser resuelto mediante técnicas de minería de datos.
- Aplicar *parcialmente* la metodología CRISP-DM para resolver el caso
- Aplicar principios de visualización de información para presentar mediante un *dashboard* los resultados de los modelos creados.

## 3. Fases y actividades por desarrollar:

- **Fase de Entendimiento del negocio**
  - Determinar los objetivos de negocio
    - Objetivos de negocio
    - Criterios de éxito (en términos de negocio)
  - Determinar los objetivos de minería de datos
    - Objetivos de minería de datos
    - Criterios de éxito desde la perspectiva de minería de datos
- **Fase de Entendimiento de los datos**
  - Descripción de los datos (descripción general de los datos a utilizar, cantidad de registros, tipos de datos, etc.)
  - Exploración de los datos (debe explorar todas las variables o la mayoría)
  - Verificación de la calidad de datos (sobre todos los datos)
- **Fase de Preparación de los datos**
  - Selección de los datos
  - Limpieza de los datos
  - Construcción de nuevos datos (atributos)
  - Transformaciones aplicadas a los datos

Si la preparación de los datos la realiza en un lenguaje de programación distinto a R, debe documentar los cambios o transformaciones aplicadas a los datos. Aporte *screenshots* y agréguelos como figuras en el documento **RMarkdown**.

- **Fase de modelado**
  - Selección de las técnicas a utilizar
  - Construcción del modelo
    - Selección de los parámetros (justificar su selección)
    - Ejecución
    - Descripción de los modelos obtenidos
  - Evaluación de los modelos
    - Describir los resultados obtenidos con cada modelo
    - Seleccionar el mejor modelo

#### 4. Diseño y construcción del *dashboard*

- Integrar los datos de análisis obtenidos a partir del dataset en un *dashboard* que facilite el análisis de dicha información.
- Publique el *dashboard* en **Tableau Public** y agregue el enlace al documento **RMarkdown**.
- Agregue el enlace al documento **RMarkdown** y exportarlo a **html** o **pdf**

#### 5. Aspectos administrativos

- Un archivo **Rmd** con el código fuente del proyecto y la salida en un archivo **html** o **pdf**, y deben contener:
  - Nombre del estudiante
  - Descripción y resultados de las diferentes fases CRISP-DM; debe contener el código, el resultado de ejecutar el código y las explicaciones o descripciones que correspondan.
  - Enlace al dashboard en Tableau Public
- **Forma de trabajo:** en grupos de 1 a 2 integrantes
- **Forma de entrega:** colocar los archivos **Rmd** y **html** en un archivo comprimido, subirlo al TEC Digital en la sección de **proyectos**.
- **Fecha de entrega:** La última lección del curso (semana 8) será para la presentación de sus trabajos.

#### 6. Evaluación

- Se dispondrá de 25 minutos por grupo de trabajo para exponer el proyecto.
- El proyecto debe ser 100% original del grupo de trabajo.
- A continuación, el detalle de la rúbrica:

<b>Fase de entendimiento del negocio</b>	<b>8</b>
Objetivos de negocio	2
Criterios de éxito de negocio	2
Objetivos de minería de datos	2
Criterios de éxito (perspectiva de minería de datos)	2
<b>Fase de entendimiento de los datos</b>	<b>12</b>
Descripción de los datos	4
Exploración de los datos	4
Verificación de la calidad de datos	4
<b>Fase de Preparación de los datos</b>	<b>15</b>
Selección de los datos	4
Limpieza de los datos	3
Construcción de nuevos datos	3
Transformaciones aplicadas	5
<b>Fase de modelado</b>	<b>35</b>
Selección de técnicas	3
Construcción de los modelos	12
Descripción de los modelos	10
Evaluación de los modelos y selección	10
<b>Dashboard</b>	<b>20</b>
<b>Exposición</b>	<b>10</b>
<b>TOTAL</b>	<b>100</b>

## 7. Referencias

- Datasets Kaggle <https://www.kaggle.com/datasets>