En el archivo *Datos\_Banco.csv* están los datos de 4521 clientes que fueron seleccionados en la última campaña masiva de tarjetas de crédito de un banco colombiano. Asesores de la entidad se contactaron con los clientes y les ofrecieron una nueva tarjeta de crédito con beneficios en la tasa de interés y estos decidieron si la tomaban o no. El diccionario de variables es:

1. **Edad.** Edad en años del cliente.
2. **Ocupación.** Ocupación del cliente.
3. **Est\_civil.** Estado civil del cliente.
4. **Nivel\_educativo.** Máximo nivel escolar alcanzado por el cliente.
5. **Contrato\_meses.** Duración del último o actual contrato de trabajo del cliente.
6. **Cred\_hipotecario.** Si el cliente tiene o no crédito hipotecario vigente con el banco.
7. **Cred\_personal.** Si el cliente tiene o no crédito personal vigente con el banco.
8. **Medio\_contacto.** Medio por el cual se contactó al cliente.
9. **Día.** Día del mes en el que se realizó el último contacto con el cliente.
10. **Mes.** Mes del año en el que se realizó el último contacto con el cliente.
11. **Duración\_seg.** Duración en segundos del último contacto con el cliente.
12. **Aceptación\_TC.** Si el cliente aceptó o no la tarjeta de crédito en la campaña mencionada.

Ustedes fueron contratados como consultores por el banco para que, con base en los datos, hagan un **perfilamiento** general de los clientes que fueron escogidos para ofrecerles la tarjeta de crédito con el beneficio y, además, identifiquen características comunes de aquellos que aceptaron la tarjeta y características comunes de los que no. Para ello, deben usar herramientas descriptivas sobre los datos empleando el software R. El objetivo es aplicar distintas técnicas que permitan alcanzar el objetivo planteado.

Metodología de entrega:

1. Deben resolver en el documento las preguntas que se presentan más adelante, convertir este archivo a formato PDF, guardarlo indicando sus apellidos en el nombre del archivo y enviarlo vía correo en respuesta al mail en el que se envía este documento y los datos.
2. En dicho correo también deben adjuntar el link al notebook de Colab o el archivo html que les genera R. Tengan en cuenta que, en caso de trabajar con Colab, el notebook debe estar público y debe poderse comentar. Además, una vez enviado, ustedes no deben ejecutar ninguna celda pues esto invalidará todo el trabajo y tendrán una nota de 0.0 en el taller.
3. El plazo máximo de entrega es el **miércoles 15 de mayo a las 10:00 PM (diez de la noche)**. Cualquier entrega que llegue luego de esta hora se calificará con 0.0.

**Resuelva las siguientes preguntas en el espacio a continuación e incluya salidas de R (pueden ser pantallazos donde también se vea la línea de código de la salida) que soporte sus respuestas. Justifique en cada caso.**

1. (5) Clasifique las 12 variables que componen el conjunto de datos en escalares o cualitativas y, dentro de cada nivel, determine si son continuas/discretas o nominales/ordinales.

**Rta.**

1. (8) Usando las funciones **describe()** y **describeBy**() calcule estadísticos de tendencia central, dispersión, posición y forma para las variables escalares. Analícelos para la base completa y en términos comparativos para los clientes que aceptaron la tarjeta de crédito y para los que no la aceptaron. Si lo requiere use alguna función adicional para hallar algún estadístico específico. ¿Obtiene algún *insight* que le permita empezar a diferenciar a los clientes que sí aceptan el producto de los que no?

**Rta.**

1. (7) Usando tablas de contingencia absoluta, de contingencia condicional y gráficos de barras adecuados analice si existe alguna dependencia entre la tenencia de créditos (personal y/o hipotecario) y la aceptación o rechazo de la tarjeta de crédito ofrecida. ¿Obtiene algún *insight* que le permita caracterizar a los clientes que aceptan la tarjeta de crédito? Explique.

**Rta.**

1. (20) Indique el valor de verdad de las siguientes afirmaciones (V o F). Sustente con salidas del software y analice en cada caso.
   1. (5) La mayoría de las personas casadas en la base de datos tienen ocupación informal.

**Rta.**

* 1. (5) Es más probable que un cliente acepte la tarjeta de crédito si se le contacta en abril que si se le contacta en marzo.

**Rta.**

* 1. (5) La cantidad de meses del último o del actual contrato de trabajo tiene un comportamiento simétrico independientemente de si el cliente aceptó o no la tarjeta de crédito.

**Rta.**

* 1. (5) La edad promedio de las personas en la base de datos cuyo nivel educativo es bachillerato supera los 35 años.

**Rta.**

1. (10) Escriba dos párrafos de máximo 10 líneas cada uno en el que resuma sus hallazgos de la exploración. Céntrese en mencionar los atributos que diferencian a los clientes que aceptan la tarjeta de los que no la aceptan y, con base en esas características, sugiera estrategias al banco para aumentar la tasa de aceptación del producto de crédito. Incluya hallazgos adicionales a los encontrados en los puntos anteriores.

**Rta.**

1. (+5) Realice un análisis descriptivo bivariado entre las variables *edad* y *duración\_seg* incluyendo diagramas de dispersión y cálculo y análisis de coeficiente de correlación lineal. Explique lo que obtiene. ¿Hay alguna relación lineal entre la duración del contacto y la edad del cliente contactado?

**Rta.**