



**BNP PARIBAS  
CARDIF**

Prueba Técnica  
Científico de Datos  
Gerencia de Data Analytics  
BNP Paribas Cardif  
Enero 2022

## Instrucciones Generales

Esta prueba está diseñada para conocer tus habilidades técnicas en análisis de datos. No se espera que logres llegar a un entregable óptimo, pero sí que demuestres tu capacidad de análisis, conocimientos en herramientas y metodologías de analytics.

Puedes hacer cualquier tipo de suposición que consideres necesaria y mencionar cualquier inconveniente que se te presente. Por favor, adjuntar cualquier código utilizado para la realización de la prueba la única condición es que este código debe ser realizado en Python.

Ten en cuenta que para el desarrollo del caso puedes usar cualquier herramienta que consideres útil y el procedimiento debe ser entendible para el calificador de la prueba. Se calificará basado en la creatividad de la solución, la calidad del análisis y conclusiones y la manera en que los resultados son presentados.

Junto con este enunciado se te serán entregadas dos bases de datos para la realización de la prueba: *df\_train.csv*, *df\_test.csv* y *sample\_envio.csv*

## Entregables

La prueba consiste de tres partes, descritas a continuación y para su solución debes presentar:

- Un informe con tu procedimiento y código detallado en forma de un Jupyter Notebook, donde se resuelvan las primeras dos partes de la prueba. Se te pedirá que sustentas tu código el día de la entrevista, así que asegurate de hacerlo claro y ordenado
- Un archivo .csv con las predicciones de tu modelo para la base *df\_test.csv* que te fue entregada. Muy importante que tu archivo contenga el id del cliente y su respectiva predicción, para que pueda ser revisado (usa el archivo *sample\_envio.csv* para ver un ejemplo de como debes mandar tus predicciones)
- Durante la ultima parte de la prueba se te pedira realizar una presentación que resuma tu trabajo, la cual, también, se te pedirá exponer durante tu entrevista con el equipo de Analytics.

De esta manera, debes mandar antes de la fecha y hora mencionada en el correo los tres archivos: el jupyter notebook, el csv y la presentación.

# 1. Modelamiento y Analisis de Datos

Durante este primer semestre del 2022, Cardif tiene planeado sacar al mercado un nuevo producto de seguros vehiculares. Para iniciar, se piensa ofrecer este nuevo producto únicamente a clientes que tengan vigente una póliza de vida con el grupo.

Con esto en mente y con el fin de garantizar el éxito del nuevo producto, el equipo de telemarketing te ha solicitado ayuda para determinar las características principales que hacen que nuestros clientes estén dispuestos a comprar la póliza de vehículos.

Para solucionar este problema se te proporciono una base de datos de clientes del grupo con información proveniente de un estudio de mercado. En el cuadro 1 puedes encontrar una descripción detallada de las variables contenidas en el Dataset.

Se espera que tu solución contenga, como mínimo, los siguientes puntos:

- Limpieza y análisis exploratorio de los datos. ¿Qué variables consideras relevantes para la tarea? ¿por qué?
- Construcción de un modelo analítico para la solución del problema. Discutir ¿Por qué se seleccionó dicha aproximación? ¿Qué otras metodologías se consideraron, ¿Qué tan valido es el modelo? y ¿Qué problemas puede tener?
- Análisis crítico de como Cardif puede utilizar tus principales hallazgos para maximizar las probabilidades de éxito del nuevo producto.

Variable	Definition
id	Id único por cada cliente
Gender	Género del cliente
Age	Edad del cliente
Driving_License	0 : Cliente no tiene licencia, 1 : Cliente tiene licencia
Region_Code	Código único que indica la región del cliente
Previously_Insured	1 : Cliente ya tiene un seguro vehicular, 0 : Cliente no tiene un seguro vehicular
Vehicle_Age	Edad del vehiculo
Vehicle_Damage	1 : Cliente ha dañado su vehiculo en el pasado. 0 : Cliente no ha dañado su vehiculo en el pasado.
Annual_Premium	La prima total que el cliente pagara al año
PolicySalesChannel	Codigo del canal por el cual se contacta al cliente ie. Agentes, Via Mail, Via Teléfono, En oficina, etc.
Vintage	Número de días que el cliente ha estado asociado con la compañía
Response	1 : Cliente esta interesado en el seguro, 0 : Cliente no esta interesado en el seguro

Cuadro 1: Variables contenidas en el Dataset

Adicional a tratar estos puntos, una vez tengas un modelo desarrollado, deberas generar y adjuntar las predicciones para la base de datos *df\_test.csv*

## 2. Diseño de Experimentos

Buenas noticias, tu modelo de venta de seguros vehiculares es todo un éxito, o al menos es lo que parece por su desempeño de acuerdo con tu análisis. Al haber pasado todas las pruebas controladas es hora de ponerlo a prueba con clientes reales.

Si se empieza a utilizar únicamente el modelo, esto representa un cambio importante en la manera como el departamento de telemarketing prioriza y establece comunicación con los clientes. Por esta razón, tu objetivo es diseñar un plan de evaluación que le permita al departamento de telemarketing tomar la mejor decisión para guiar su proceso de ventas (seguir con su metodología actual o utilizar el modelo)

Para esto te encargan la construcción de un informe en el cual deberás:

- Proponer y definir la métrica principal mediante la cual mediría el éxito del modelo en la práctica? ¿A qué otras métricas podría ser útil hacerles seguimiento?
- Definir un plan de evaluación que permita definir si utilizar el modelo construido presenta una mejora sobre la metodología actual de ventas. ¿Cómo balancearías la necesidad de entregar resultados en el corto plazo con la validez estadística de los mismos?
- Explicar cómo trasladarías los resultados del plan de evaluación a la toma de la decisión de implementar en definitiva el modelo o seguir con la metodología actual, como guía del proceso de ventas.

## 3. Presentación de Resultados

Después de todo tu análisis el modelo que desarrollaste resultó ser una mejora significativa al compararse con el proceso de ventas que se venía utilizando. Por esta razón, la junta directiva del grupo está muy interesada en conocer tu procedimiento, solución e interpretación de los resultados obtenidos. Para esto, debes realizar una presentación ejecutiva en la cual presentarás un resumen y explicación de tu trabajo.

Es muy importante que tengas en cuenta que esta presentación estará dirigida a un público que no posee los conocimientos técnicos de análisis de datos que tú posees, pero igualmente debe ser capaz de transmitir toda la información relevante. Debes ser capaz de presentar tu procedimiento de una manera que cualquier persona podría entender. Concéntrate en presentar tus hallazgos acerca de la naturaleza del problema, oportunidades accionables de mejora y resultados.