**PROYECTO DE CURSO: PRIMER BANCO DE SEBASTOPOL**

Usted es un analista financiero recientemente contratado en el Primer Banco de Sebastopol, un banco consolidado y líder en la República del mismo nombre. Acaba de salir de una reunión y las palabras de su jefe siguen dando vuelta en su cabeza:

“La próxima semana me reuniré con la alta gerencia para discutir la estrategia financiera. Hay un gran desacuerdo dentro del grupo sobre si deberíamos comenzar a cobrar una mensualidad a nuestros clientes por el uso del canal de banca en línea o si deberíamos a ofrecer incentivos a los clientes, como reembolsos y tarifas de servicio más bajas para fomentar un mayor uso. El debate realmente gira en torno a si los clientes en línea son realmente más rentables. ¿por qué no pasas el fin de semana revisando los datos relevantes de año pasado? Reunámonos el lunes por la mañana para discutir sus hallazgos.

A ud. le entusiasma la industria de la banca minorista, pero aún no lo entendía del todo. Hace una hora, miró un gráfico sobre la rentabilidad de los clientes que, entre otras cosas, demostraba que más de la mitad de los cinco millones de clientes de Sebastopol no eran rentables. Siendo consciente de que aún le falta mucho por entender, se dirige a la oficina de Rosa Hernández, quien era la analista con mayor experiencia dentro del grupo y estaba seguro que le ayudaría.

Ud. le planteo a Rosa su idea original de comparar los saldos de las cuentas de los clientes con y sin banca en línea. Rosa sonrió, y le explicó:

“Los saldos son solo una parte de la historia. Si está interesado en saber qué tan rentables son los clientes, mira directamente las ganancias. A grandes rasgos, calculamos la rentabilidad del cliente en Sebastopol con una fórmula bastante sencilla:

(Saldo en cuentas de depósito) \* (Margen de interés neto) + (Comisiones) + (Intereses de préstamos) - (Costo de servicio)

Hernandez continuó,

Como notarás, los niveles de saldo realmente solo capturan una parte del valor general del cliente para el banco. Si te centras únicamente en los saldos, se perderán componentes importantes de los ingresos, como las comisiones, e ignorarás los costos de servicio. Por eso te recomiendo que utilices la rentabilidad por cliente como tu variable de análisis. Se que en el departamento de Banca de Personas han estado recopilando cierta información y si les explicas para que la quieres talvez quieran compartirtela.”

Entusiasmado por la conversación con Rosa, ud. envió un correo al departamento de Banca de Personas copiando a su jefe y obtuvo los datos solicitados en un archivo con extensión \*.csv. En el correo, el analista le compartió los siguientes datos:

“… En el correo adjunto encontrarás una muestra representativa de nuestros clientes, para ser exactos 31,264 observaciones. Te detallo la descripción de las variables a continuación:

1. Ganancia: Rentabilidad por cliente
2. Online: variable dummy, cero si no es usuario, uno si es usuario
3. CategoríaEdad: Categorías por edad del cliente en años 1 menor de 15 años, 2 entre 15-24, 3 entre 25-34, 4 entre 35-44, 5 entre 45-54, 6 entre 55-64 y 7 para mayores de 65.
4. CategoríaIngresos: Categorías por nivel de ingresos anuales del cliente. 1 para menos de $5,000, 2 entre 5,001 y 10,000, 3 entre 10,001 y 20,000, 4 entre 20,001 y 30,000, 5 entre 30,001 y 40,000, 6 entre 40,001 y 50,000, 7 entre 50,001 y 60,000, 8 entre 60,001 y 70,000 y 9 para más de 70,000.
5. Permanencia: Número de años que el cliente ha sido cliente del banco.
6. Municipio: Municipio de residencia del cliente.

Favor nota que hay algunos valores faltantes.”

Con la base y la descripción de las variables, ud. pensaba que tendría buenos insumos para cumplir con el pedido de su jefe.

Preguntas guía

1. Estime la rentabilidad promedio para los clientes con y sin banca en línea. ¿Son diferentes? ¿Es esta diferencia estadísticamente significativa?
2. Corra una regresión con Ganancia como variable dependiente y Online y CategoríaEdad como variables independientes. ¿Es el coeficiente Online estadísticamente significativo? ¿Por qué R no utilizó todos los datos disponibles? ¿Hay una diferencia estadísticamente significativa entre las ganacias de los clientes con y sin datos faltantes en CategoríaEdad? Si la hay, ¿puede utilizar sus resultados para sacar conclusiones?
3. Cree nuevas categorías para los datos faltantes tanto en CategoríaEdad como CategoríaIngresos y corra una regresión incluyendo las variables de control. ¿Cómo interpreta cada uno de los coeficientes? Luego de controlar por todas las variables demográficas disponibles, ¿es el coeficiente de Online estadísticamente significativo? ¿Qué conclusión le dará a su jefe en torno a la rentabilidad de los clientes de banca online?