**UNIVERSIDAD FRANCISCO MARROQUÍN**

**FACULTAD DE CIENCIAS ECONÓMICAS**

**DATA SCIENCE FOR FINANCE**

**MARIA ISABEL AVILA RIGALT**

**PROYECTO FINAL**

**Cartera de clientes de tarjeta**

**Caso de negocio:**

Los ejecutivos de Eurobank no han llegado a metas en los últimos meses y tienen la presión de colocar un total de 2,000 tarjetas nuevas para el cierre de 2024. Quieren saber si aún es posible que lleguen con la tendencia actual y cuáles son las probabilidades de fallar para evaluar medidas correctivas. También les preocupa cuántos clientes puedan cancelar su tarjeta en el siguiente año y que el monto promedio de los saldos pueda bajar debido a la crisis económica de la región. Eurobank los ha contratado para evaluar estos riesgos.

**Instrucciones generales:**

El proyecto consiste en proyectar las colocaciones de tarjetas que hará el banco siguiendo la tendencia de crecimiento que lleva desde hace unos años y dar una respuesta a la gerencia sobre si aún es posible que lleguen a la meta de las 2,000 tarjetas nuevas o necesitan una nueva estrategia. Además, le solicitan indicar cuales son los clientes que muy probablemente cancelarán su tarjeta en el siguiente año, para estimar cuanto bajaría el saldo total de la cartera. Finalmente, debe identificar la distribución de los saldos de los clientes actuales para evaluar el riesgo de que disminuya.

Trabajarán con información de clientes del banco (cartera\_historica.csv y cartera\_actual.csv) e información histórica de colocación de tarjetas (tarjetas.csv).

**Requerimientos:**

1. Desarrollar un modelo ARIMA para proyectar las colocaciones de crédito al cierre de 2024 y las metas que serían realistas para los ejecutivos de ventas cada mes, siguiendo la estrategia actual.
   1. El resultado "Best model: ARIMA(1,0,1)(0,1,1)[12] with drift" indica que el mejor modelo ajustado para los datos de la serie de tiempo es un modelo ARIMA con los siguientes componentes:
   2. Orden (p, d, q) = (1, 0, 1): Esto significa que el modelo incluye un t√©rmino autorregresivo de orden 1 (AR(1)) y un t√©rmino de media m√≥vil de orden 1 (MA(1)). El valor 1 en el lugar del componente de diferenciaci√≥n (d) indica que no se aplic√≥ una diferenciaci√≥n en la serie de tiempo.
   3. Componente de estacionalidad (P, D, Q, S) = (0, 1, 1, 12): Este modelo ARIMA tambi√©n incluye un componente estacional, indicado por la presencia de los valores (0, 1, 1, 12). Esto significa que se aplic√≥ una diferencia estacional de orden 1 (D = 1) y un t√©rmino de media m√≥vil estacional de orden 1 (Q = 1) con una frecuencia de estacionalidad de 12 (S = 12), lo que indica una estacionalidad mensual.
   4. "with drift": Esto indica que se ha incluido un t√©rmino de deriva o tendencia lineal en el modelo para capturar cualquier patr√≥n de tendencia presente en los datos.
   5. En resumen, el modelo ARIMA(1,0,1)(0,1,1)[12] with drift representa una combinaci√≥n de componentes autorregresivos, de media m√≥vil y estacionales, con la adici√≥n de un t√©rmino de deriva. Este modelo fue seleccionado como el mejor modelo ajustado para los datos de la serie de tiempo, utilizando el criterio de informaci√≥n de Akaike (AIC) para evaluar el rendimiento de diferentes modelos.
2. A partir de los resultados del inciso “a” responda las siguientes consultas de gerencia:
   1. ¿Llegarán a la meta si continúan con la estrategia actual?

* No llegan.
  1. ¿Cuánto es lo menos que colocarían al 95% de confianza?
  2. ¿Recomendaría seguir con la misma estrategia o mejorarla?

1. Desarrollar un modelo de clasificación (con regresión logística, árbol de decisión o bosque aleatorio) utilizando los datos históricos de la cartera de tarjetahabientes (cartera\_historica.csv) y aplicarlo a los datos de los 1,000 clientes actuales (cartera\_actual.csv). Debe predecir cuántos clientes cancelarán y responder a las siguientes preguntas:
   1. ¿Qué porcentaje de los clientes actuales cancelarán?
   2. ¿Cuál es el perfil de los clientes que cancelarán? ¿Cómo se diferencia del resto? Sustentar su respuesta con al menos 3 gráficas de exploración de la cartera de clientes que no cancelan y también de la cartera que cancela de la base de cartera\_historica.csv.
   3. Cada integrante del grupo debe entregar además un Rscript con otras 2 gráficas de exploración, interpretarlas y dar su recomendación para prevenir cancelaciones en Eurobank (un párrafo, basado en datos).
   4. ¿Cuánto de saldo perderían por los clientes que probablemente cancelarán según los resultados de su algoritmo de clasificación?

**Entregables:**

1. Código de R por grupo con su procedimiento. Cualquier transformación aplicada al set de datos debe aparecer en el código y debe poderse correr completo y que de las respuestas que colocaron en su análisis.
2. Presentación con las respuestas a todos los puntos anteriores, asegúrese de incluir la gráfica del modelo ARIMA y proyecciones de ventas, cancelaciones esperadas, la descripción y visualización de la cartera de clientes que cancelan vs los activos visualización de saldos simulados y su VAR, más sus recomendaciones para Eurobank.

**Rúbrica:**

* Modelo ARIMA de pronóstico de ventas: 8
* Modelo de clasificación de cancelaciones: 8
* Exploración de cancelaciones individual: 4
* Presentación final: 5
* Evaluación oral individual: 5

**Total: 30 puntos netos**

**Nota:** recuerde que también habrá una **evaluación oral individual el día de la presentación** que vale 5 netos y un **examen teórico el día miércoles 12** con valor de 10 netos para completar los 40 netos del final.

**Fecha de entrega:** jueves 13 de julio antes de la clase.

**Presentaciones y evaluación oral individual:** jueves 13 de julio a las 7 am.