**Preguntas finales prueba técnica.**

**Pablo Andrés Figueroa Chamorro.**

1) ¿La gráfica presentada arriba es congruente con respecto al título, subtitulo, ejes, etc? ¿Por qué no? ¿Qué le añadiría?

En términos generales, la congruencia de la gráfica con el título, subtítulo, eje, etc, es aceptable. Sin embargo, respecto a los ejes, podría modificarse las etiquetas para que la gráfica sea más clara, concretamente, podría utilizarse “Número de Tarjetas Canceladas" en el eje Y y "Mes" en el eje X. Además, con el objetivo de facilitar la lectura de la gráfica podría modificarse la escala del eje Y y presentar los datos en miles.

2) ¿Cuál es crecimiento mensual promedio de las tarjetas canceladas?

El crecimiento promedio absoluto corresponde a 32128.67, se lo obtiene a través de:

mean(diff(data\_grafica$sum))

Por otro lado, la tasa de crecimiento promedio equivale a 14.28849% y se obtiene con:

mean(diff(data\_grafica$sum) / head(data\_grafica$sum, -1))

3) ¿Cuál fue el Establecimiento de Crédito que más tarjetas canceló durante todo el horizonte de tiempo?

El establecimiento que más tarjetas canceló fue Tuya, se llega a esta conclusión utilizando:

data2 <- data1 %>% group\_by(Entidad) %>% summarise(sum = sum(canceladas))

data2$Entidad[which.max(data2$sum)]

4) El cliente le pregunta ¿Cuál es la razón de está tendencia? Detalle los pasos a seguir de su metodología, entre más robusta sea, mejor.

En primer lugar, para responder a la pregunta planteadaes necesario aumentar la cantidad de periodos observados, puesto que revisando únicamente 4 meses es realmente comlejo llegar a una solución más precisa y robusta robusta. Para esto puede consultarse la fuente de la que originalmente se obtuvo los datos (SFC).

Ahora bien, una vez ya se tiene determinada la variable endógena, es necesario determinar las exógenas. Un buen inicio es revisar variables macroeconómicas, como puede ser la tasa desempleo, ya que Un aumento en el desempleo puede reducir la capacidad de las personas para pagar sus deudas y llevarlos a cancelar ciertos productos financieros como las tarjetas. La tasa de interés de colocación es una opción interesante, ya que estas, al ser muy altas, pueden desincentivar el uso de tarjetas de crédito. Ingreso y/o consumo promedio, puesto que pueden modificar las necesidades en términos de pagos de la población y, por ende, inducir a un aumento en la cancelación de tarjetas en caso de ser muy bajo.

Por otro lado, también es necesario añadir variables que modelen factores del sistema financiero y gubernamental. Por ejemplo, podría incluirse, en forma de variable dicotómica, la introducción de nuevos productos financieros al mercado, como depósitos de bajo monto, que cubran necesidades de pago similares a las de las tarjetas. Además, también podría analizarse índices de morosidad, puesto que puede inducir a las personas a cancelar productos por su incapacidad de pagarlos. Por otro lado, puede incluirse variables dicotómicas que modelen la introducción de cambios regulatorios o estímulos económicos relacionados con el uso de tarjetas.

Finalmente, también podría incluirse variables dicotómicas para determinados periodos del año para analizar posibles fluctuaciones causadas por determinados eventos, como pueden ser navidad, vacaciones, entre otros.

Una vez recopilados los datos, es necesario seleccionar un modelo econométrico que permita localizar las relaciones de causalidad que solicita el cliente. Es por ello que modelos ARIMA o de machine learning pueden no ser los ideales, ya que estos no están diseñados para buscar causalidad. Una buena opción, dada su congruencia con los objetivos planteados y flexibilidad, puede ser un modelo de regresión dinámica con variables exógenas, como el ARDL.

Para utilizar el modelo ARDL es necesario cargar el paquete "ARDL". el modelo se crearía mediante los siguientes códigos:

modelo1 <- auto\_ardl(Y ~ X, data = data, max\_order = c(5, 5), selection = "AIC")

modelo2 <- auto\_ardl(Y ~ X, data = data, max\_order = c(5, 5), selection = "BIC")

Lo ideal en este punto es que la cantidad de regresores de la variable endógena y de las variables exógenas coincidan en ambos modelos, sin embargo, en caso de que no lo hagan, lo más conveniente sería dar prioridad al modelo seleccionado y estimado por el criterio BIC, ya que este favorece más la parsimonia que el AIC.

Una vez especificado el modelo, puede verse los resultados a través de:

summary(modelo2)

Y la cointegración a largo plazo con:

coint\_eq(modelo2)

Ahora bien, para validar los resultados del modelo, es necesario aplicar ciertas pruebas estadísticas. Para ello es necesario utilizar los paquetes "lmtest" y "car". Para corroborar la no presencia de cambios estructurales, se realiza el test CUSUM:

cusum <- cusum(modelo2)

plot(cusum)

En caso de la línea salga de los límites establecidos por la prueba, entonces puede ser necesario incluir una dummy adicional que modele el cambio estructural. Ahora bien, Para la autocorrelación de errores se puede emplear la prueba de Breusch-Godfrey o la prueba de Durbin-Watson:

bgtest(modelo2)

dwtest(modelo2)

En este caso un p-valor bajo puede indicar presencia de autocorrelación en los errores, por lo que puede ser necesario ajustar la especificación del modelo. Por otro lado, la heteroceasticidad puede ser probada con el test de Breusch-Pagan:

bptest(modelo2)

En caso de detectar heteroceasticidad a través de un p-valor bajo, es necesario estimar el modelo a través de GLS o máxima verosimilitud. A continuación, se aplica la prueba VIF:

vif(modelo2)

En caso de que la prueba arroje un número alto, puede ser mayor o igual a 10, es necesario revisar la especificación del modelo y eliminar alguna variable o integrarla con otra.

Finalmente, después de implementar los ajustes sugeridos por los resultados de las pruebas estadísticas, se puede concluir que las estimaciones del modelo son apropiadas para explicar la tendencia observada en la gráfica inicial en lo que a cancelación de tarjetas respecta.