

Migración de clientes prepago como estrategia para la fidelización y una oportunidad para rentabilizar.

Realizado por:

Yvers Caruzo Linares



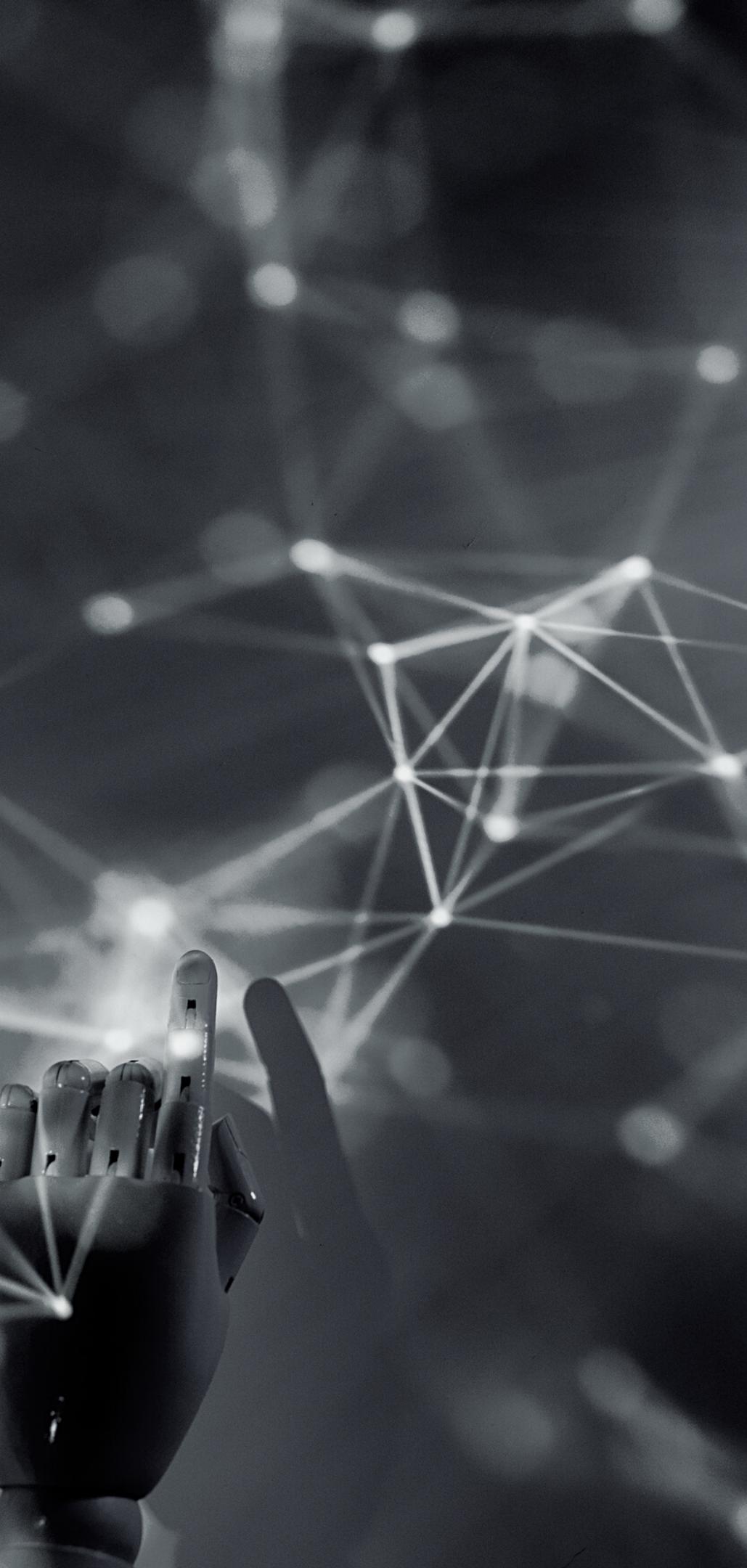
Agenda

- 
- 1 **Contexto académico**
 - 2 **Contexto Empresarial y de la iniciativa de negocio**
 - 3 **Preguntas de investigación**
 - 4 **Metadata - Variables de interés**
 - 5 **Análisis exploratorio multinivel**
 - 6 **Modelo / rendimientos / validaciones**
 - 7 **Insights y recomendaciones**

Contexto académico

Como parte de este proyecto de ciencia de datos se trabajo en varias fases:

- 1.- **ETL**, desde las plataformas de almacenamiento de logs.
- 2.- **Análisis descriptivo** e identificación de tendencias y variables de interés.
- 3.- **Modelado y validaciones.**
- 4.- **Conclusiones y propuestas.**



splunk®>

 python

Google
 colab

 pandas

 NumPy

 seaborn

 scikit
learn



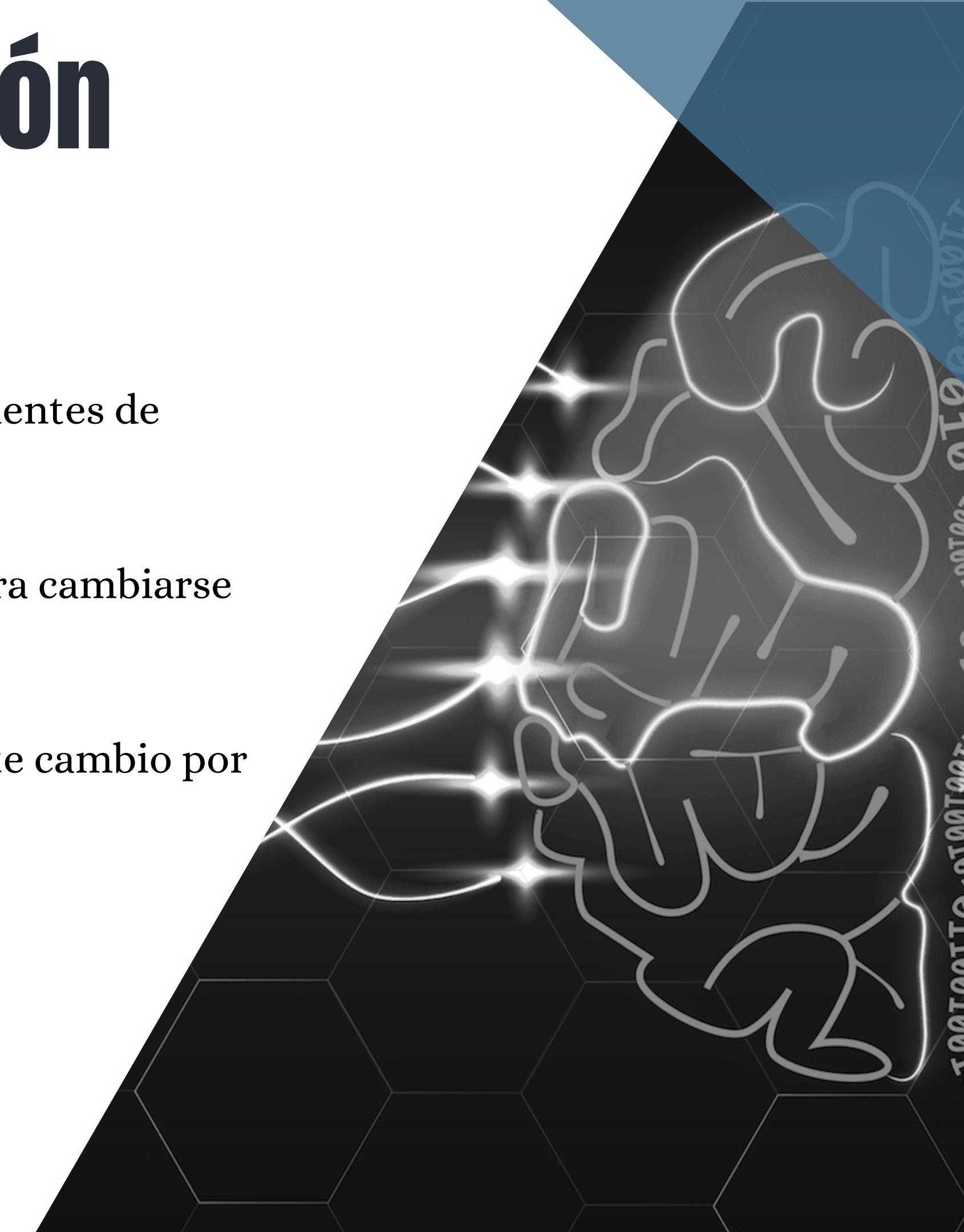
Contexto empresarial y del negocio

- Empresa del sector de telecomunicaciones, con más de 50 años en el mercado chileno.
- Busca acercar las infinitas oportunidades de la tecnología, con responsabilidad y con centro en el cliente.
- Adaptación a nuevas necesidades y servicios, entregando soluciones y beneficios con el objetivo de promover una alta fidelización del cliente.
- Apartir de las oportunidades que presenta el mercado actual, en ese marco según (Subtel) el 30% de los clientes abonados con lineas activas corresponden a clientes prepago, esto es un segmento de un poco más de 7MM de lineas que pertenecen a un potencial mercado que permitiría incrementar el segmento de clientes con cuenta.

Preguntas de investigación

Lo que pretendemos responder

- 1.- Cómo se viene comportando la iniciativa de migración de clientes de servicios prepago al mundo postpago.
- 2.- Cómo es el viaje del cliente al que se le presenta la oferta para cambiarse del tipo de servicio.
- 3.- Es posible predecir la cantidad de clientes que realizaran este cambio por tipo de cliente.



Metadata: Variables de interés

Definiciones

compuesto por 10 variables de interés que componen el proceso de captura de datos que se ha desarrollado para esta iniciativa.

1.- FECHA: corresponde a la fecha en la que se mostró la oferta al cliente.

2.- Tipo_Cliente: define si el cliente que se perfiló para la oferta posee registro de identidad o no, esto significa que como el mercado abordado es el mercado prepago el cliente no está obligado a registrar sus datos, sin embargo existe un % interesante que si lo realiza.

3.- Oferta: define el plan ofertado.

4.- Escucha_Oferita_Mig: Indica si el cliente escuchó o no la oferta mostrada.

5.- Acepta_Mig: Recoge si el cliente al que se le mostró la oferta la aceptó e inicia los pasos para tomar el plan ofertado.

6.- Valida_Rut: Indica si el cliente realizó el proceso para validar su identidad en el sistema automático.

7.- EvalCredito: Recoge a los clientes que asan la evaluación comercial posterior, ya que de entrada existe un perfilamiento del cliente, sin embargo esta evaluación se realiza a la luz de los datos del sistema financiero.

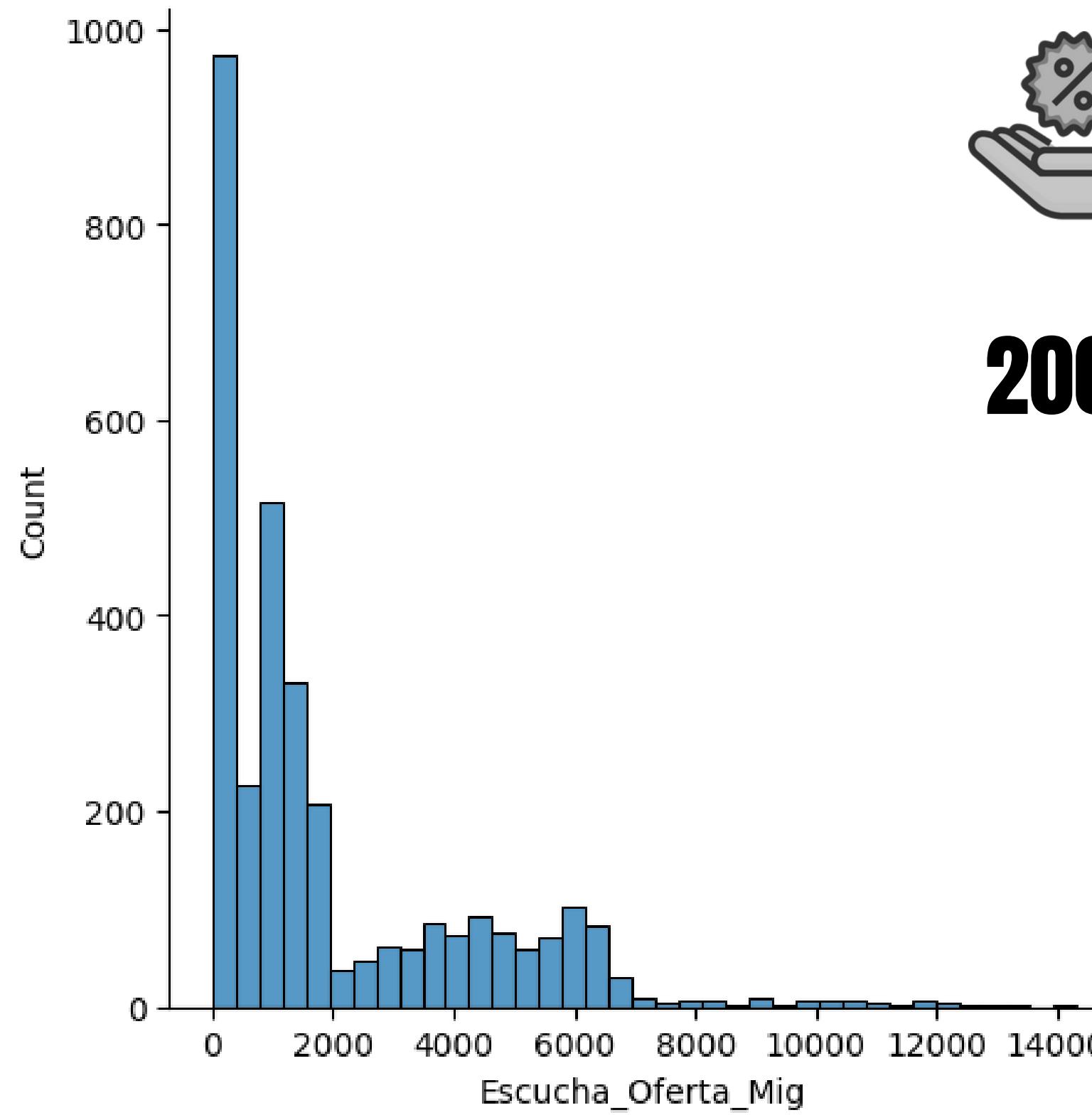
8.- T&C: Indica si el cliente aceptó los términos y condiciones.

9.- migradoOK: Muestra a los clientes que lograron realizar el cambio y contratación de manera exitosa.

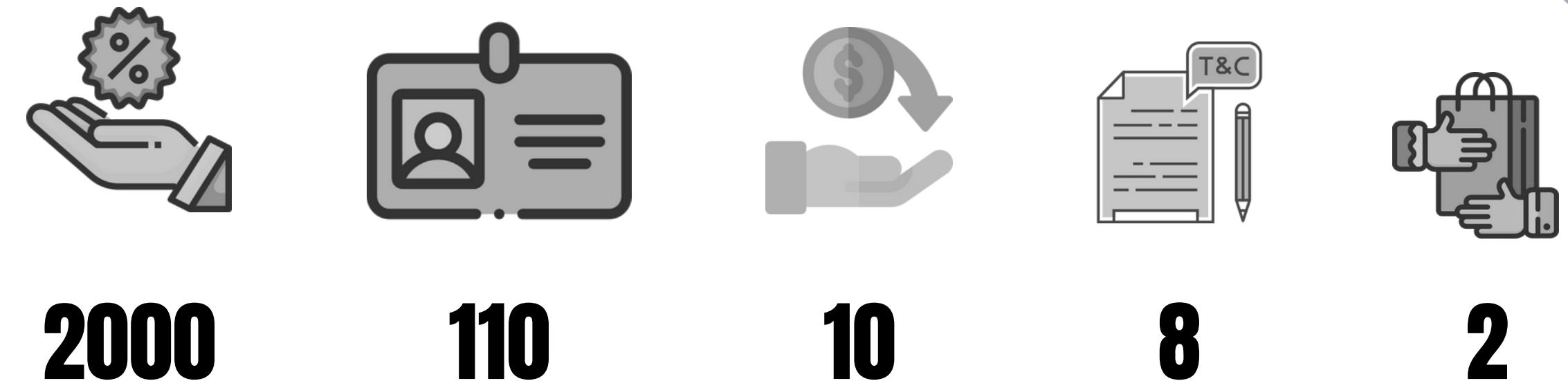
10.- Envio_RPA: Muestra los clientes que con algún error del sistema automático se envían a un derivador para poder retomar el proceso pero esta vez con el soporte de un Back.



Metadata: Variables de interés

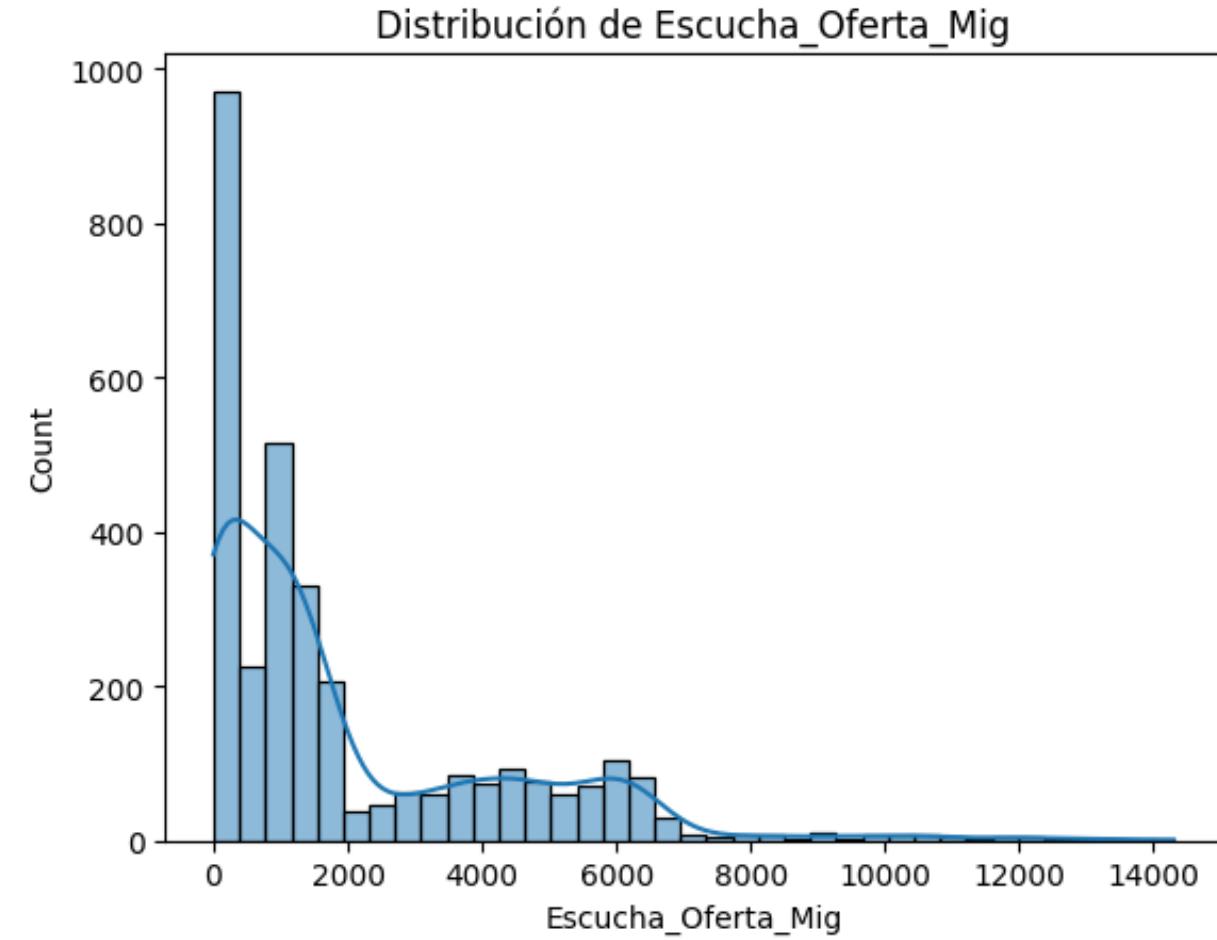
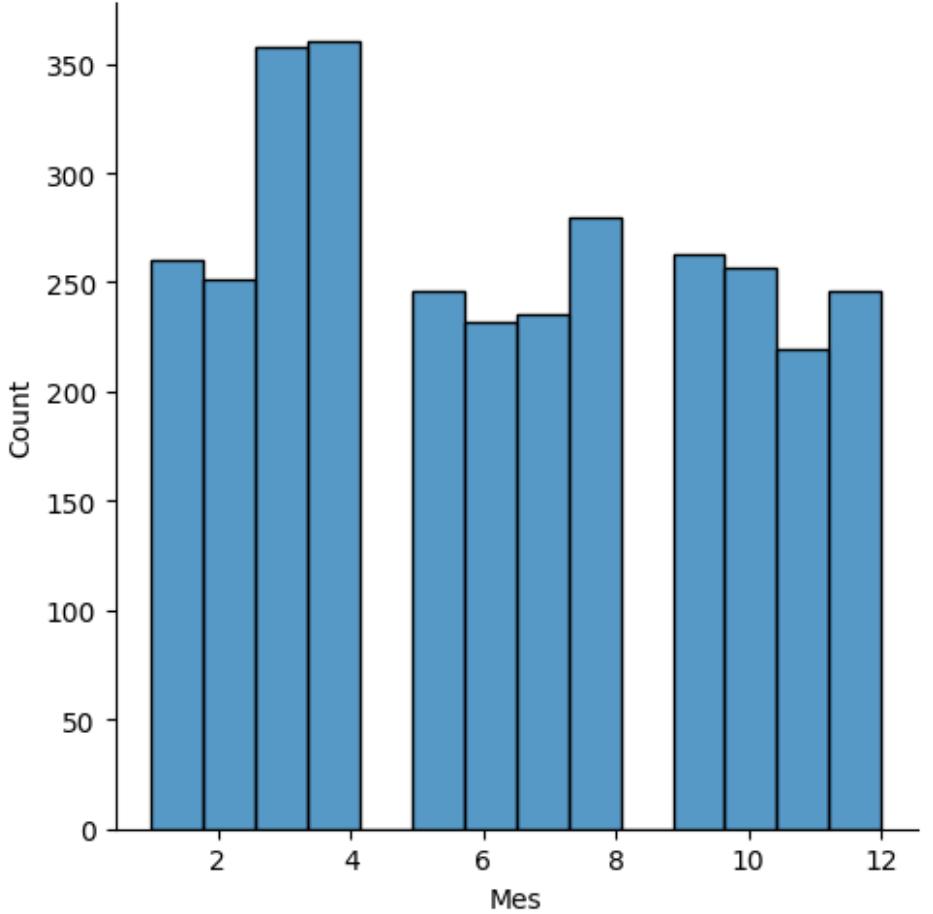
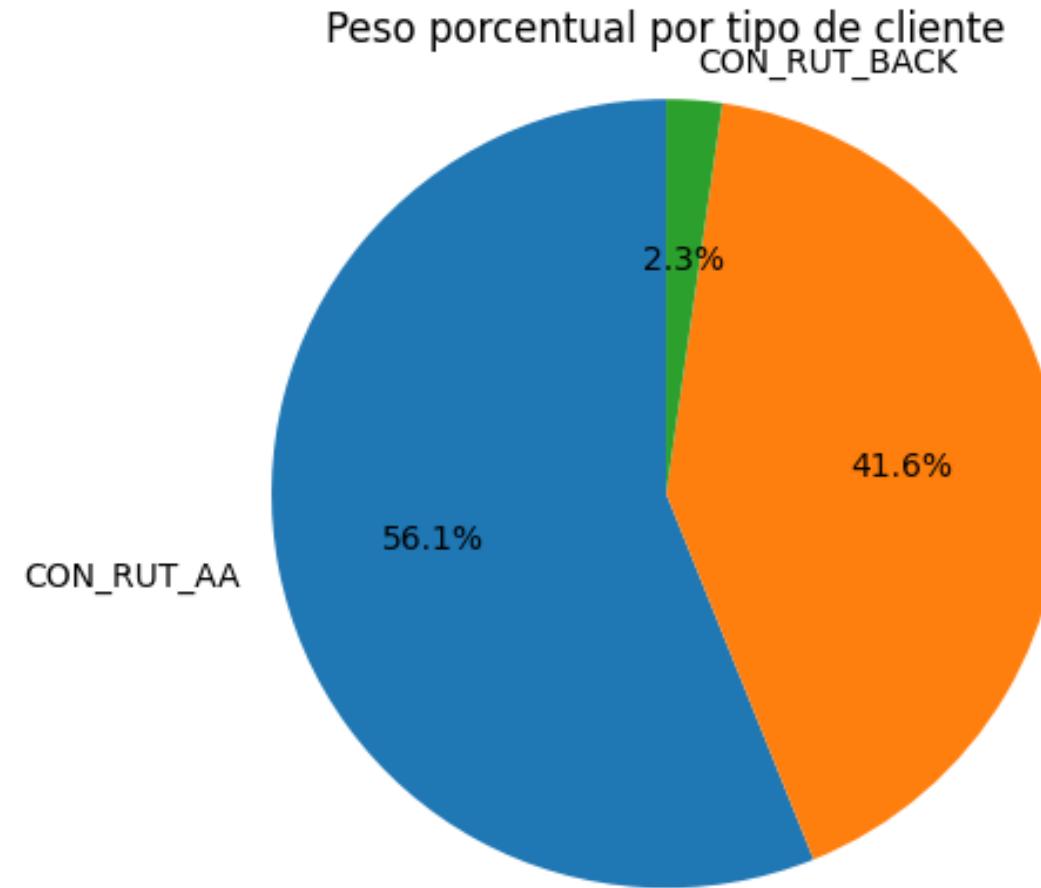


Promedio mes



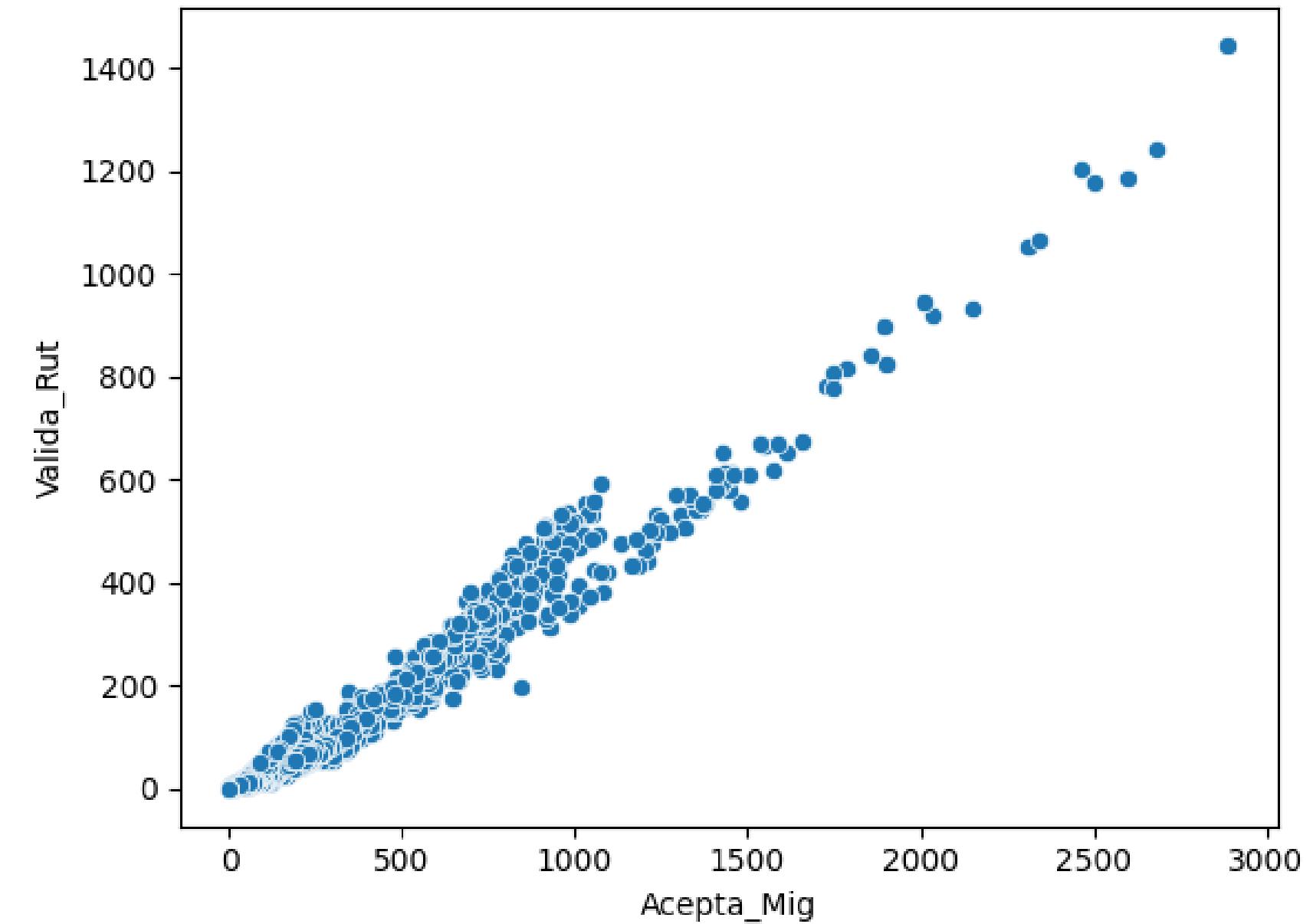
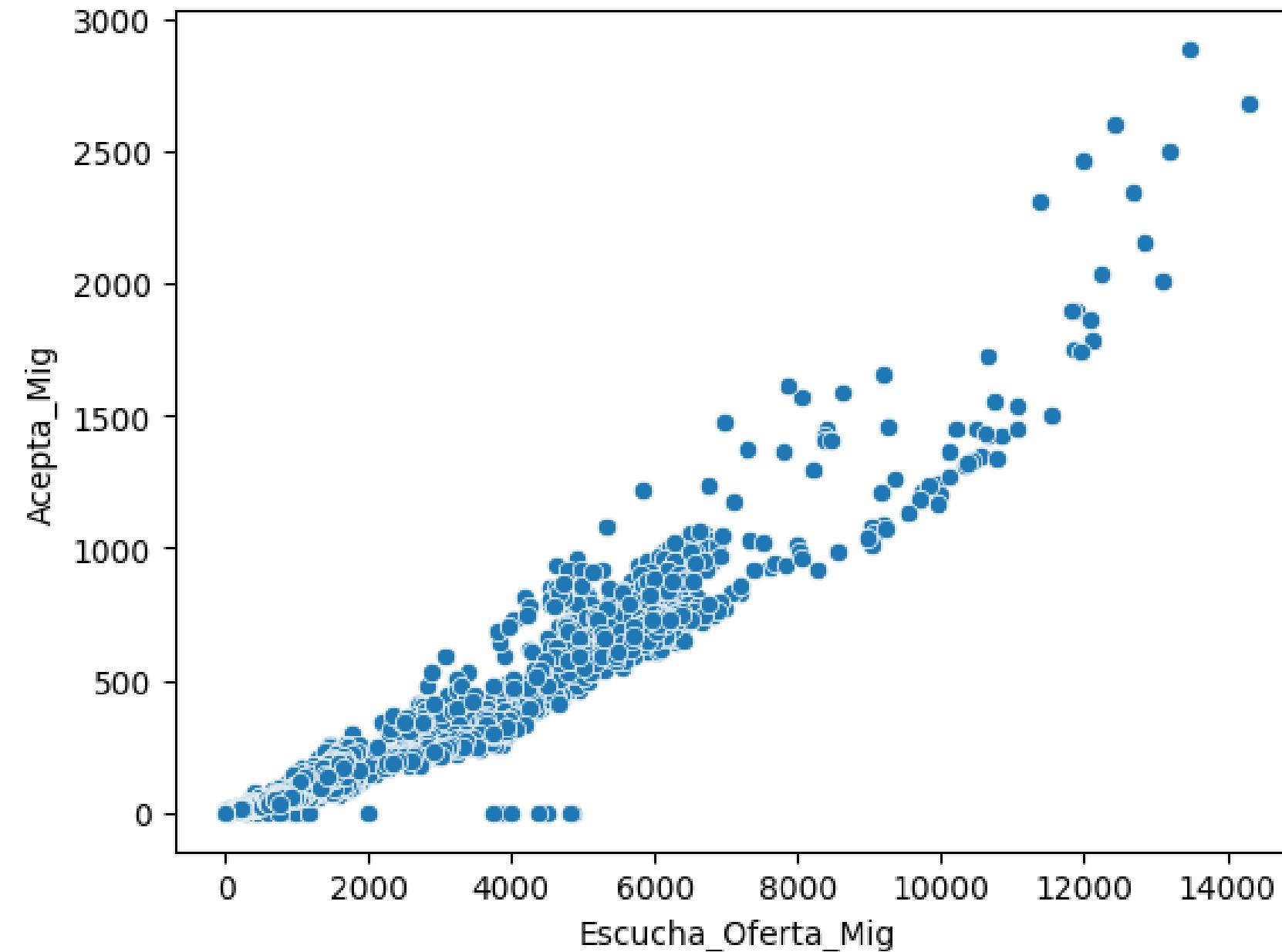
| | count | mean | std | min | 25% | 50% | 75% | max |
|--------------------------------|--------|-------------|-------------|-----|-------|--------|---------|---------|
| <code>Escucha_Ofera_Mig</code> | 3134.0 | 2000.500319 | 2308.366397 | 1.0 | 310.0 | 1098.5 | 3271.00 | 14300.0 |
| <code>Acepta_Mig</code> | 3134.0 | 230.028717 | 325.639745 | 0.0 | 20.0 | 101.0 | 294.75 | 2886.0 |
| <code>Valida_Rut</code> | 2677.0 | 109.908853 | 152.706052 | 1.0 | 14.0 | 45.0 | 131.00 | 1445.0 |
| <code>EvalCredito</code> | 3146.0 | 10.083916 | 15.732136 | 0.0 | 0.0 | 3.0 | 13.00 | 109.0 |
| <code>T&C</code> | 3154.0 | 8.474001 | 13.578522 | 0.0 | 0.0 | 3.0 | 10.00 | 93.0 |
| <code>migradoOK</code> | 3207.0 | 1.498285 | 4.715808 | 0.0 | 0.0 | 0.0 | 1.00 | 82.0 |
| <code>Envio_RPA</code> | 3179.0 | 5.586662 | 12.207546 | 0.0 | 0.0 | 0.0 | 5.00 | 93.0 |

Análisis exploratorio de datos - Univariado



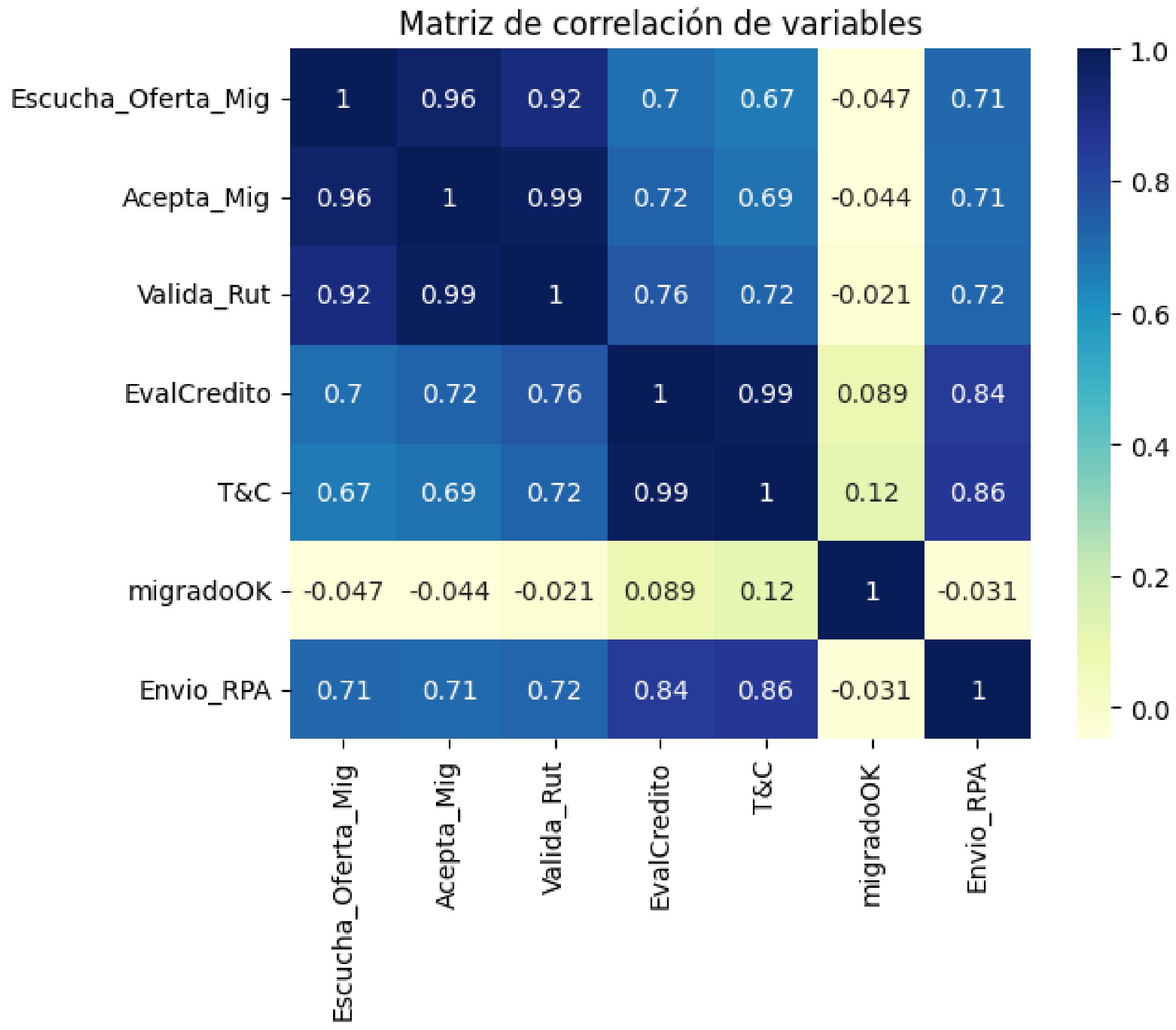
- los clientes con identificación corresponden al 58% del total de clientes con oferta disponible, esto indica que lograr mayor performance y tracción de la oferta significa incrementar la base de clientes perfilados con identificación, esto ademas refuerza el primer punto observado, si simplificamos el flujo se podrá tener un mejor performance de la iniciativa, dada la perdida de clientes a medida de que se tienen más fases en el proceso de migración.
- Se observa que el primer Q del año analizado contó con mayor base de clientes perfilados, este tema se abordaría con mercado para atomizar la oferta de manera más balanceada durante el año de gestión.
- Se observa la presencia de un skew a la izquierda indicando saturación de casos alejados de la media.

Análisis bivariado



- Se aprecian relaciones lineales para las variables descritas.
- Cambios en la relación de variables sobre todo en las de migración Ok e Identidad.

Análisis multivariado

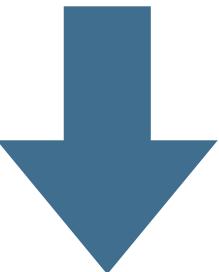


- Se aprecian los más altos niveles de correlación en las variables de *Escucha, Aceptación y Validación de identidad.*
- *Evaluación de crédito y TyC actúan como subproceso según lo visto en la matriz.*
- El comportamiento de las migraciones Ok lo muestra como puna variable aislada si miramos las relaciones con el resto de variables del proceso.

Desarrollo del modelo

Random Forest

- *Rondas I - II*
 - *Variables sintéticas (**Momentos del viaje del cliente**)*
 - Rendimientos - Crossvalidation



El modelo presenta un accuracy entorno al **0.74**, visto el *Random Forest - Cross-Validation Accuracy*: **0.7393541034980973**, lo que muestra una correspondencia entre ambos modelos.

```
[ ] df_prueba
df_prueba = pd.read_csv('https://raw.githubusercontent.com/ageron/patterns/master/datasets/telecom_churn.csv')
df_prueba.info()
df_prueba.head()

# Evaluamos el rendimiento del modelo a partir solo de las variables sintéticas
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)

Accuracy: 0.7420382165605095

#obtenemos el reporte de clasificación y la matriz de confusión
print("Classification Report:")
print(classification_report(y_test, y_pred))

print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))

[[219  5  2  0  0  0  0  0  1  1  0  0  1  0  0  0  0  0  0  0]
 [ 0  1  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 11  5  3  0  1  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 7  4  2  1  1  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 2  1  0  1  1  1  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 1  3  0  1  2  0  0  2  0  1  0  0  0  1  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 2  0  0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0]]
```

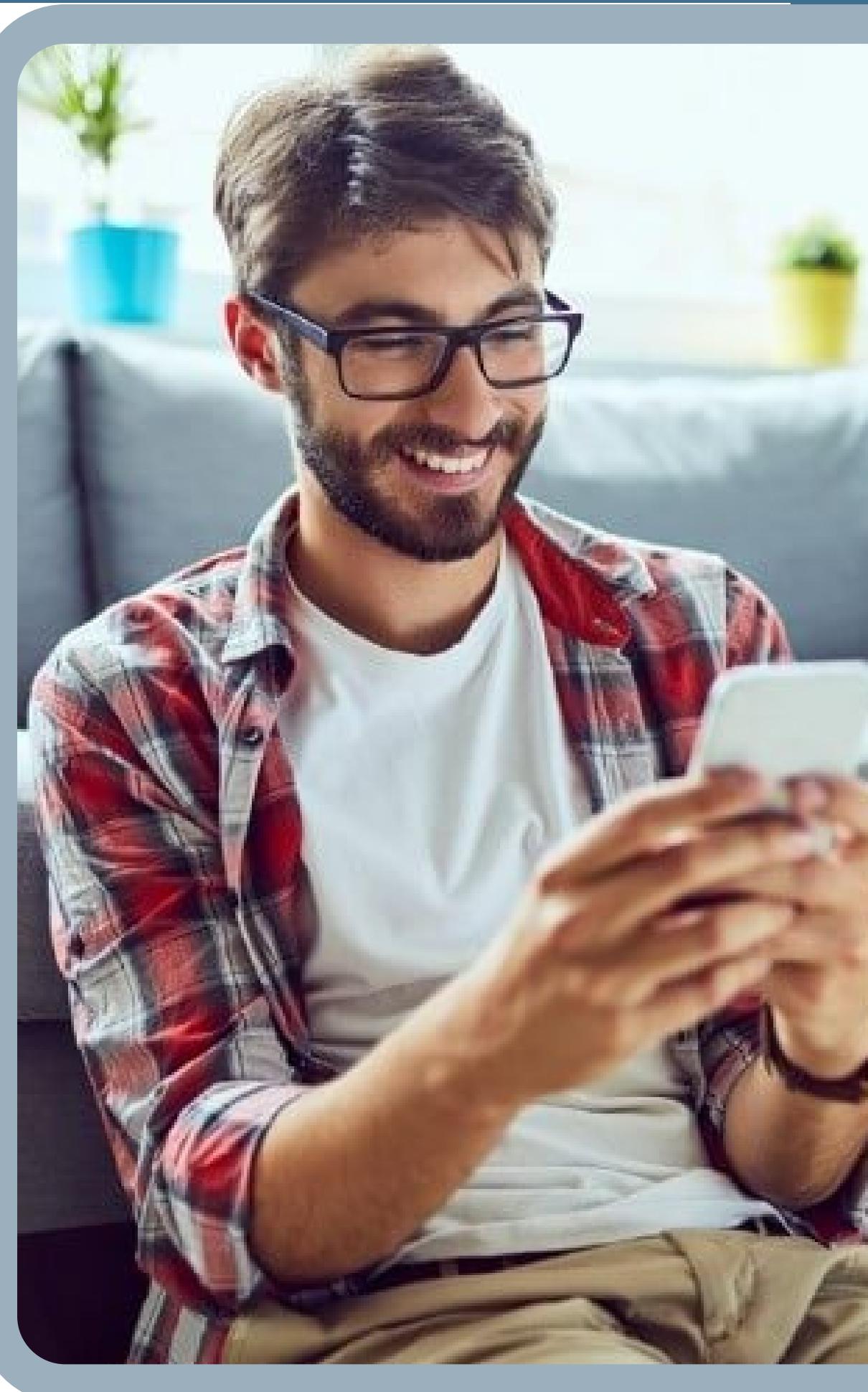
¿Qué indica el modelo? Conclusiones

- A partir de la creación de variables sintéticas (que englobaron los momentos del flujo de migraciones) el modelo mejoró considerablemente, complementado con la selección del modelo recomendado y luego del uso del **Bias-Variance tradeoff**, se obtuvieron las siguientes métricas de desempeño:
- 1. para la **Precisión** el 87%. Esto indica que del total de ejemplos que el modelo predijo como "o", el 87% realmente pertenecían a la clase "o".
- 2. **Recall**: En este caso, el recall para la clase "o" es del 95%. Es decir, el modelo identificó correctamente el 95% de todos los ejemplos que realmente pertenecían a la clase "o".
- 3. **F1-score**: El F1-score para la clase "o" resultó del 90%. Esto es considerado en la mayoría de los casos, como un buen rendimiento, evidenciando que el modelo está obteniendo un buen equilibrio entre precisión y recall para la clase "o".
- 4. Para la columna de **support** se muestra el número total de ejemplos en la clase "o" en el conjunto de datos de prueba, esto indica en términos de investigación un criterio de saturación claro.



Recomendaciones para mejora de la iniciativa

- Se observa durante el análisis que las variables iniciales del proceso Escucha oferta, Acepta y validación de identidad comparten relaciones lineales, lo que permitiría la proyección del comportamiento de la aceptación de la oferta y cuyo performance mejoraría si se realizan mejoras a nivel de CX.
- con la variable validación de identidad se presenta una oportunidad comercial, ya que ese 41% de clientes que no entregan ni validan podrían recibir un tratamiento de plan inicial, que presente un nivel de riesgo razonable, permita construir historial al cliente e ir permitiendo el upgrade de los servicios contratados.
- La conversión de la iniciativa no llega al 1% lo que supone la necesidad de revisar tanto el perfilamiento de la base, así como la oferta entregada, lo que podría complementarse con un estudio de clúster de clientes para construir esta oferta más segmentada y que promueva la mejora de la tasa de conversión de esta iniciativa.
- Tomar en consideración el comportamiento del Q1, parecería constituir el momento de mayor permeabilidad del cliente ante ofertas de este tipo.



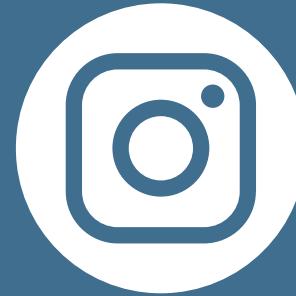
Para contactos y recomendaciones de mejora....



[Yvers Jesus Caruzo Linares](#)



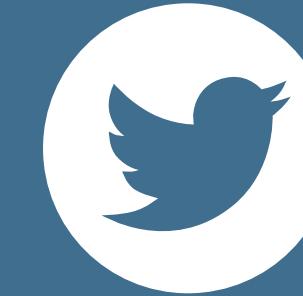
yverscaruso@gmail.com



[Yvers Caruso](#)



[yverscaruso](#)



[@yversjcl](#)

