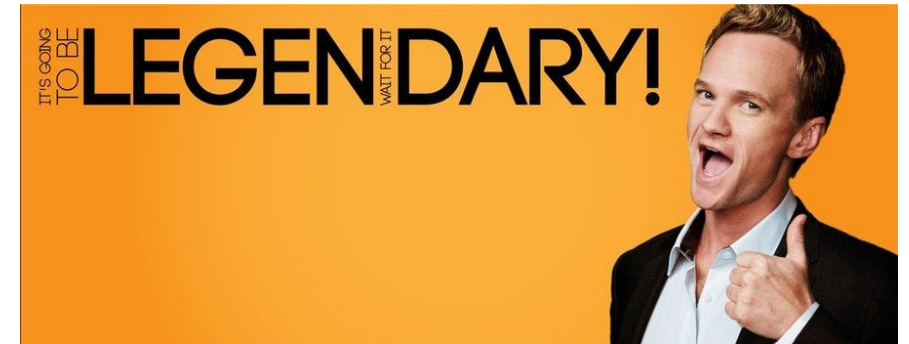


# ANÁLISIS DE CLIENTES DEL



¿QUIÉNES ABANDONAN AL BANCO MÁS ASOMBROSO?



# Agenda

- Contexto y audiencia
- Preguntas de interés
- EDA
- Resumen y primeros insights

# Contexto y audiencia

Como dicen muchos experimentados, es importante captar nuevos clientes, pero aún más importante es poder retenerlos, ya que recuperarlos requiere un esfuerzo mucho mayor.

Bajo este concepto, es imprescindible que en Goliath National Bank entendamos el comportamiento de nuestros clientes actuales y que semejanzas existen con los que se dieron de baja para seguir siendo legen... wait for it.

Para eso es analizará una muestra de 10127 clientes entre los que hay actuales y dados de baja.

Con el presente análisis trataremos de explicar y entender que variables son relevantes a la hora de detectar posibles bajas y, en consecuencia, poder actuar proactivamente sobre las mismas. CHALLENGE ACCEPTED



# Contexto y audiencia

## Contexto empresarial

- Sabemos que somos el banco más asombroso de todos, por eso no podemos permitirnos perder clientes.
- Para esto, el equipo de data science, llevó a cabo un análisis exhaustivo donde se proyectará a lo largo de la presentación.

## Problema comercial

- Tan como lo explicábamos anteriormente, no podemos dejar de ser el banco más asombroso de todos y para eso debemos evitar que nuestros clientes migren hacia otros bancos, entendiendo por qué se dan de baja y actuando sobre dichas variables.

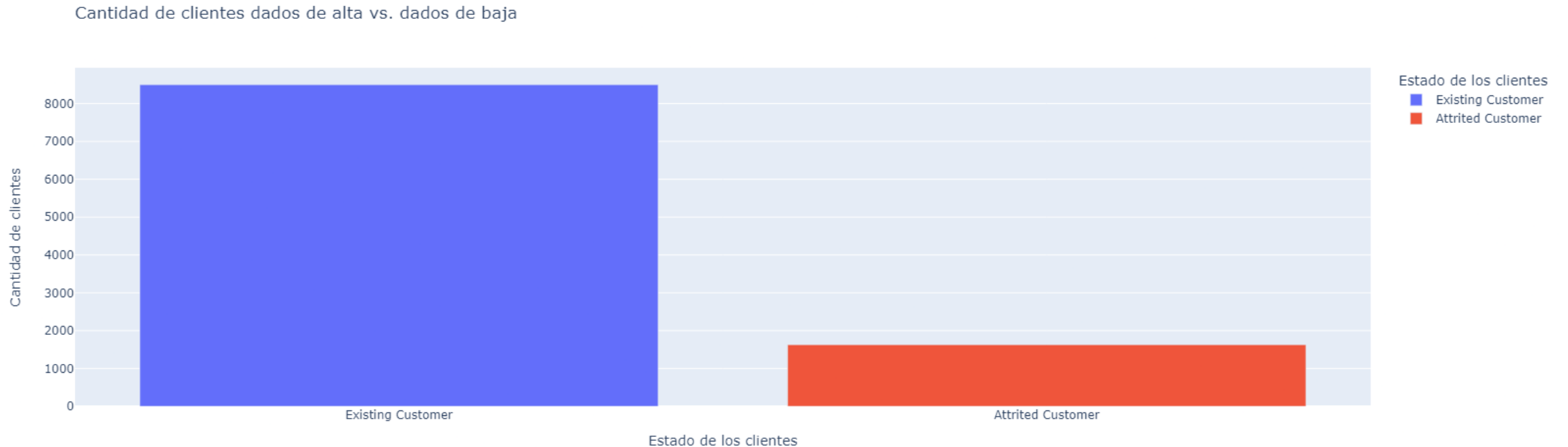
## Contexto analítico

- El archivo a trabajar es un CSV que contiene datos de 10127 clientes en donde, además de encontrar etiquetados cuales ya se encuentran dados de bajo, podemos analizar desde ingresos, gastos, tiempo desde su alta, estado civil, educación, etc.

## ¿Por qué los clientes se dan de baja?

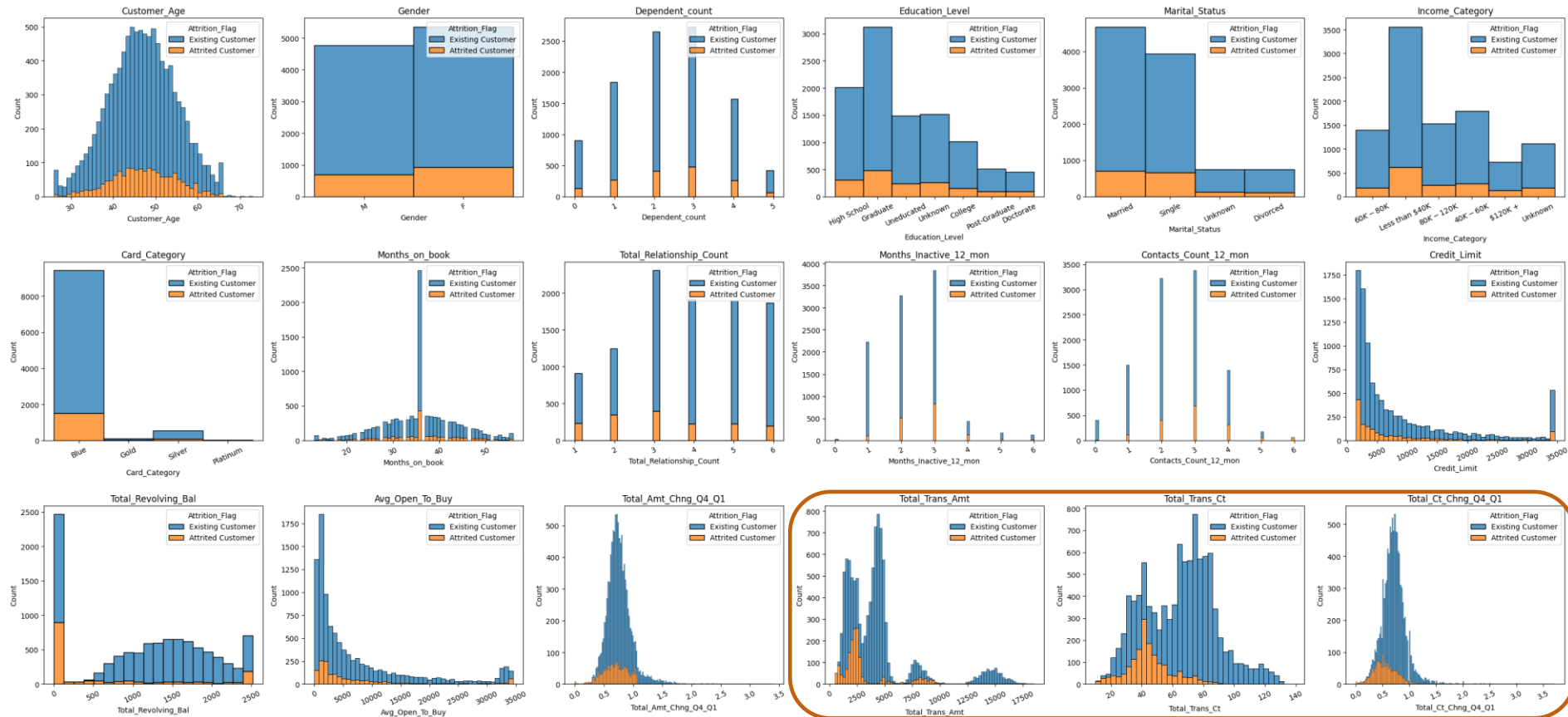
- ¿Se pueden observar comportamientos diferentes tanto en las variables categóricas como en las numéricas?
- ¿Cuáles variables poseen comportamientos diferentes entre los clientes que se dan de baja y los que no?
- ¿Qué valores adoptan las variables que permiten diferenciar entre "Attrited Customer" y "Existing Customer"?
- ¿Qué modelo es mejor para detectar posibles bajas? (futuras entregas)

# ¿Cómo es nuestra base de datos?



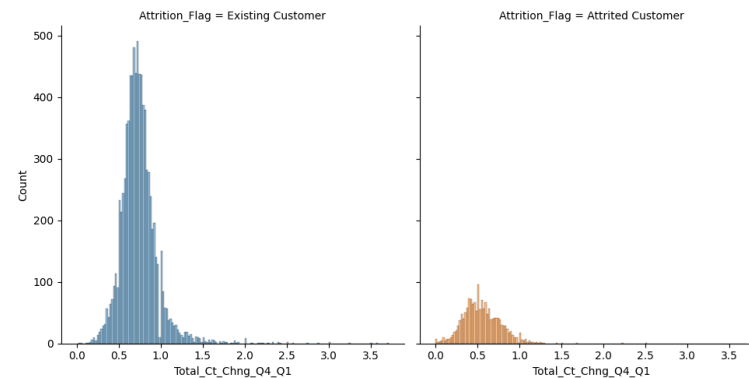
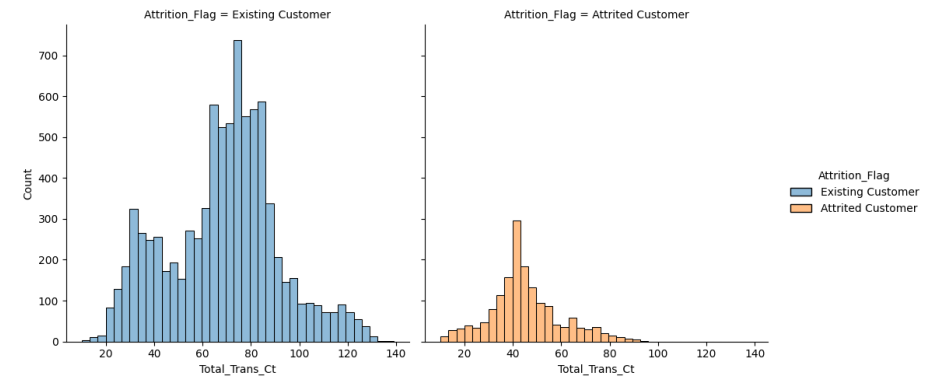
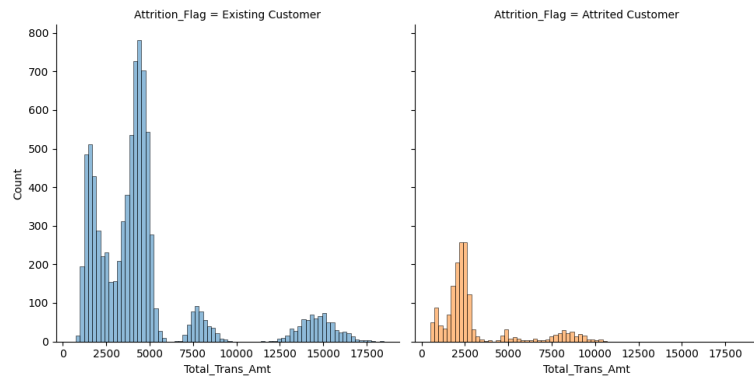
- Se aprecia una clara diferencia entra la cantidad de datos de una categoría y otro
- Esto puede afectar a la hora de realizar un modelo de predicción por lo que nivelaremos los datos de cada uno más adelante

# ¿Cómo es nuestra base de datos?



- Entre todas las variables que tenemos, llaman la atención 3 en particular: “Total\_Trans\_Amt”, “Total\_Trans\_Ct” y “Total\_Ct\_Chng\_Q4\_Q1” ya que parecen tener comportamientos diferentes en sus medias

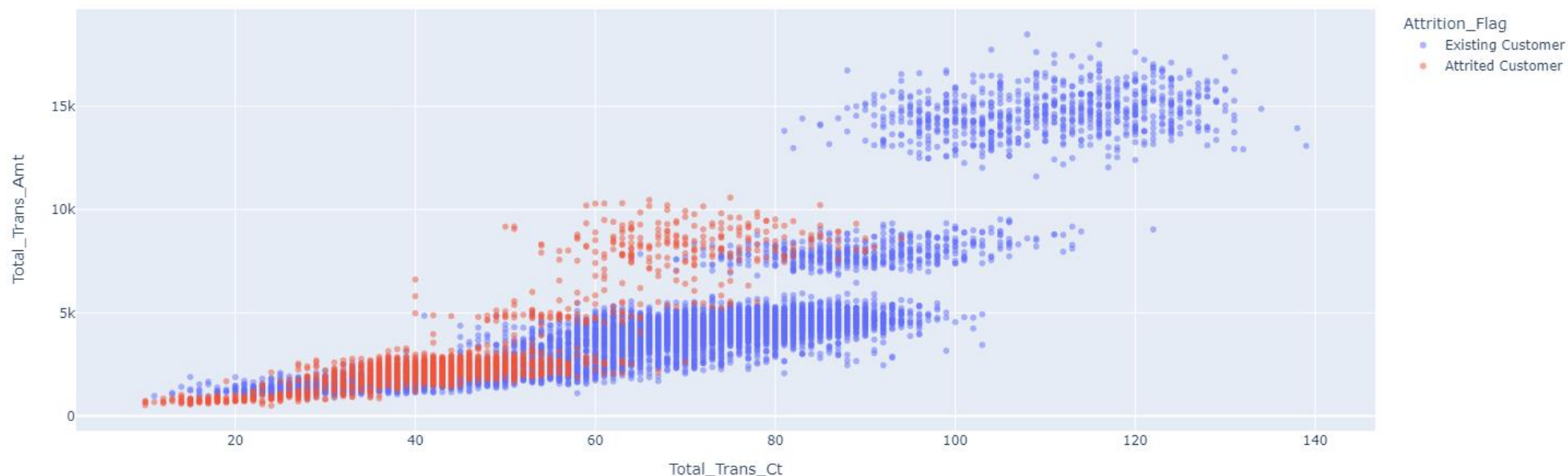
# ¿Cómo es nuestra base de datos?



- Haciendo una especie de “zoom” sobre las susodichas variables, vemos que tenemos comportamientos diferentes en sus medias según son clientes que aún permanecen o abandonaron por lo que ya tenemos 3 potenciales variables de interés



# Cantidad de Transacciones vs. Monto de Transacciones



- Realizando un Scatter sobre la cantidad de transacciones y el monto total, vemos que los clientes que han abandonado la institución se contienen en el cuadrante inferior a las 100 transacciones y montos menores a 10k., por lo que podemos asegurar con bastante certeza que aquellos clientes que no se encuentren en dicho cuadrante van a continuar en la institución

# Test de hipótesis – Comparación de medias

- Variables donde se rechaza la  $H_0$ : “las medias de la distribución de clientes que se mantienen son iguales a la de aquellos que se dieron de baja”

- Total\_Relationship\_Count

- p-valor: 0.0

- Months\_Inactive\_12\_mon

- p-valor: 0.0

- Contacts\_Count\_12\_mon

- p-valor: 0.0

- Credit\_Limit

- p-valor: 0.016

- Total\_Revolving\_Bal

- p-valor: 0.0

- Total\_Amt\_Chng\_Q4\_Q1

- p-valor: 0.0

- Total\_Trans\_Amt

- p-valor: 0.0

- Total\_Trans\_Ct

- p-valor: 0.0

- Total\_Ct\_Chng\_Q4\_Q1

- p-valor: 0.0

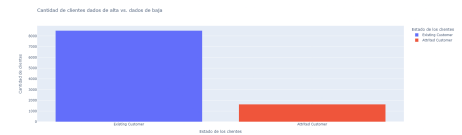
# Correlación de datos

	Client_ID	Customer_Age	Dependent_count	Months_on_book	Total_Relationship_Count	Months_Inactive_12_mon	Contacts_Count_12_mon	Credit_Limit	Total_Revolving_Bal	Avg_Open_To_Buy	Total_Amt_Chng_Q4_Q1	Total_Trans_Amt	Total_Trans_Ct	Total_Ct_Chng_Q4_Q1
Total_Ct_Chng_Q4_Q1	0.01	-0.01	0.01	-0.01	0.04	-0.04	-0.09	-0.0	0.09	-0.01	0.38	0.09	0.11	1.0
Total_Trans_Ct	-0.0	-0.07	0.05	-0.05	-0.24	-0.04	-0.15	0.08	0.06	0.07	0.01	0.81	1.0	0.11
Total_Trans_Amt	-0.02	-0.05	0.03	-0.04	-0.35	-0.04	-0.11	0.17	0.06	0.17	0.04	1.0	0.81	0.09
Total_Amt_Chng_Q4_Q1	0.02	-0.06	-0.04	-0.05	0.05	-0.03	-0.02	0.01	0.06	0.01	1.0	0.04	0.01	0.38
Avg_Open_To_Buy	0.01	0.0	0.07	0.01	-0.07	-0.02	0.03	1.0	-0.05	1.0	0.01	0.17	0.07	-0.01
Total_Revolving_Bal	0.0	0.01	-0.0	0.01	0.01	-0.04	-0.05	0.04	1.0	-0.05	0.06	0.06	0.06	0.09
Credit_Limit	0.01	0.0	0.07	0.01	-0.07	-0.02	0.02	1.0	0.04	1.0	0.01	0.17	0.08	-0.0
Contacts_Count_12_mon	0.01	-0.02	-0.04	-0.01	0.06	0.03	1.0	0.02	-0.05	0.03	-0.02	-0.11	-0.15	-0.09
Months_Inactive_12_mon	0.01	0.05	-0.01	0.07	-0.0	1.0	0.03	-0.02	-0.04	-0.02	-0.03	-0.04	-0.04	-0.04
Total_Relationship_Count	0.01	-0.01	-0.04	-0.01	1.0	-0.0	0.06	-0.07	0.01	-0.07	0.05	-0.35	-0.24	0.04
Months_on_book	0.13	0.79	-0.1	1.0	-0.01	0.07	-0.01	0.01	0.01	0.01	-0.05	-0.04	-0.05	-0.01
Dependent_count	0.01	-0.12	1.0	-0.1	-0.04	-0.01	-0.04	0.07	-0.0	0.07	-0.04	0.03	0.05	0.01
Customer_Age	0.01	1.0	-0.12	0.79	-0.01	0.05	-0.02	0.0	0.01	0.0	-0.06	-0.05	-0.07	-0.01
Client_ID	1.0	0.01	0.01	0.13	0.01	0.01	0.01	0.01	0.0	0.01	0.02	-0.02	-0.0	0.01

- Se observan relaciones muy fuertes entre “Total\_Trans\_Ct” y “Total\_Trans\_Amt” por lo que seguramente estén ligadas
- Tenemos una alta correlación, como puede llegar a ser lógico, entre “Months\_on\_book” y “Customer\_Age”
- Existe una correlación perfecta entre “Credit\_Limit” y “Avg\_Open\_To\_Buy” por lo que podrían ser redundantes incluir ambas

# Resumen y primeros insights

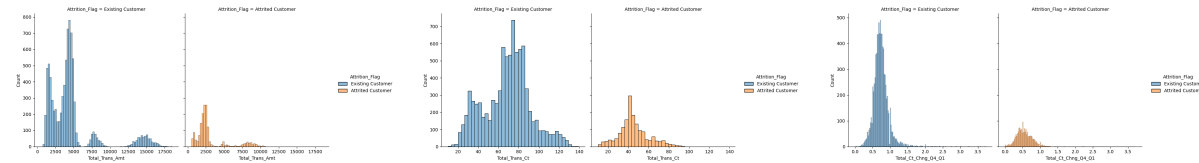
- Tenemos una base de datos con una **amplia diferencia entre clientes que aún siguen activos (84%) y los que se dieron de baja (16%)**



- Si queremos crear modelos predictivos, antes habrá que nivelar dichas cantidades

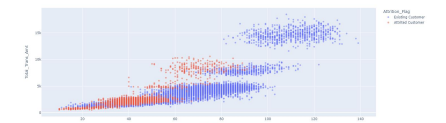
- Hay 3 variables que destacan por tener **medias diferentes según analicemos el estado de los clientes:**

- “Total\_Trans\_Amt” (4650 vs. 3095)
- “Total\_Trans\_Ct” (70 vs. 45)
- “Total\_Ct\_Chng\_Q4\_Q1” (0.75 vs 0.55)



- A estas 3 variables detectadas en el EDA se le suman otras 6 que se detectaron mediante la prueba de hipótesis presentando como  $H_0: \mu_i = \mu_j$  (dichas variables adicionales son: “Total\_Relationship\_Count”, “Credit\_Limit”, “Total\_Revolving\_Bal” y “Total\_Amt\_Chng\_Q4\_Q1”)

- Cuando comparamos la cantidad y el monto de las transacciones podemos observar que **ningún cliente que haya realizado más de 100 transacciones y montos mayores a 10k ha abandonado la institución**



- Se observa una **fuerte relación** entre 3 pares de variables:

- “Total\_Trans\_Ct” - “Total\_Trans\_Amt” (0.81)
- “Months\_on\_book” - “Customer\_Age” (0.79)
- “Credit\_Limit” - “Avg\_Open\_To\_Buy” (1.00)

