



---

## Organización de Datos - 75.06/95.58

---

### Trabajo práctico 1

### Análisis exploratorio de datos

2° cuatrimestre 2020

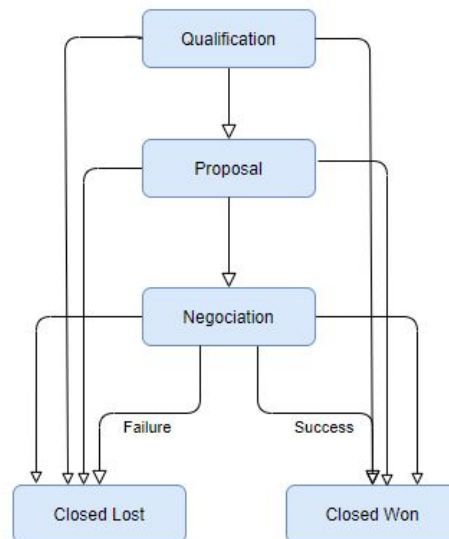
Integrantes	
Nombre	Padrón
Nicolás De Giácomo	99702
Esteban Peña	100102
Figueroa Matías	92498

# Índice

<b>1. Introducción general</b>	<b>3</b>
<b>2. Análisis de datos</b>	<b>4</b>
2.1 Tecnologías utilizadas	4
2.2 Estructura de los datos	4
2.3 Filtrado de datos	5
2.4 Oportunidades abiertas y cerradas por mes	5
2.4.1 Introducción	5
2.4.2 Análisis	5
2.4.3 Conclusión	6
2.5 Oportunidades ganadas y perdidas por estación	6
2.5.1 Introducción	6
2.5.2 Análisis	7
2.5.3 Conclusión	8
2.6 Oportunidades ganadas y perdidas por región	8
2.6.1 Introducción	8
2.6.2 Análisis	8
2.6.3 Conclusión	9
2.7 Frecuencia y éxito de los tipos de oportunidades	9
2.7.1 Introducción	9
2.7.2 Análisis	9
2.7.3 Conclusion	10
2.8 Éxito por vendedor	11
2.8.1 Introducción	11
2.8.2 Análisis	11
2.8.3 Conlusión	11
2.9 Oportunidades por país	12
2.9.1 Introducción	12
2.9.2 Análisis	12
2.9.3 Conlusión	15
2.10 Aprobación y éxito	15
2.10.1 Introducción	15
2.10.2 Análisis	15
2.10.3 Conlusión	17
2.11 Éxito por cliente	17
2.11.1 Introducción	17
2.11.1 Análisis	17
2.11.1 Conclusión	19
<b>3. Conclusiones generales</b>	<b>20</b>

# 1. Introducción general

Se tienen datos de una empresa B2B (Business To Business), donde los representantes comerciales son los encargados de llevar a cabo las ventas, llamadas *oportunidades*. Cada *oportunidad*, hace referencia a un proyecto de venta que cuenta con uno o más productos a ser vendidos. La misma tiene dos estados finales: *Closed Won* o *Closed Lost*, que hacen referencia al éxito o fracaso de dicha oportunidad, respectivamente. Los estados intermedios de las oportunidades se pueden apreciar en el siguiente gráfico.



Para concluir la introducción, es importante remarcar el objetivo principal de este análisis: determinar características, insights y variables importantes al momento de predecir el éxito o fracaso de una *oportunidad*.

Lo que se está buscando, es la construcción de un esquema que nos permita calcular una *probabilidad de éxito* para nuevas oportunidades. Es decir, la probabilidad de que una nueva oportunidad tenga como estado final *Closed Won*.

Es por esto, que la gran mayoría de nuestros apartados (si no todos) están apuntados a descubrir variables que tengan alto impacto en el destino de dichas oportunidades.

## 2. Análisis de datos

### 2.1 Tecnologías utilizadas

Para el proyecto se utilizó Python3, ayudándose con las siguientes librerías:

- Pandas y NumPy para el análisis de datos
- Matplotlib, Seaborn y Plotly para los gráficos y visualizaciones.
- Geopy para obtener datos necesarios para graficar en mapas.

Para aportar aún más a las visualizaciones y gráficos, se utilizó la plataforma web [Infogram](#).

Para el control de versiones se utilizó [Github](#), integrado con Google Colab.

### 2.2 Estructura de los datos

Se cuenta con un CSV donde cada registro corresponde a la (posible) venta de un producto. Entre los campos destacados, se encuentran:

- ID: id único del registro.
- Región: región de la oportunidad.
- Billing\_Country: país donde se emite la factura.
- Opportunity\_ID: id de la oportunidad.
- Opportunity\_Owner: vendedor del equipo comercial responsable de la oportunidad comercial.
- Opportunity\_Created\_Date: fecha de creación de la oportunidad.
- Opportunity\_Type: tipo de oportunidad.
- Last\_Modified\_Date: fecha de última modificación de la oportunidad.
- Stage: estado de la oportunidad.
- Pricing, Delivery\_Terms\_Quote\_Appr: Indica la necesidad de una oportunidad de recibir aprobación por el superior al vendedor asignado.
- Account\_Name: ID de cuenta del vendedor.
- Account\_Type: Tipo de cuenta del vendedor

También se cuenta con un CSV que contiene información sobre cada país.

Entre los campos destacados, se encuentran:

- Territory: nombre del país.
- region: región a la que pertenece.
- Hemisphere: hemisferio al que pertenece.

## 2.3 Filtrado de datos

Dependiendo el caso particular de análisis, se llevaron a cabo diferentes tipos de filtrado, por ejemplo:

- ❑ Al momento de contar oportunidades, se utiliza un filtro por `Opportunity_ID`, para evitar que oportunidades con más de un producto modifiquen el número total de oportunidades.
- ❑ Al momento de contar oportunidades abiertas y cerradas, se utiliza un filtro por `Stage`, para trabajar únicamente con las oportunidades en estados *Closed Won* o *Closed Lost*.

## 2.4 Oportunidades abiertas y cerradas por mes

### 2.4.1 Introducción

Un aspecto a analizar tiene que ver con el rendimiento de los representantes comerciales y la distribución de las oportunidades a lo largo del año.

¿Hay algún mes con mayor creación o cerrado de oportunidades?

¿La diferencia neta entre las oportunidades creadas y cerradas crece, decrece o se mantiene constante?

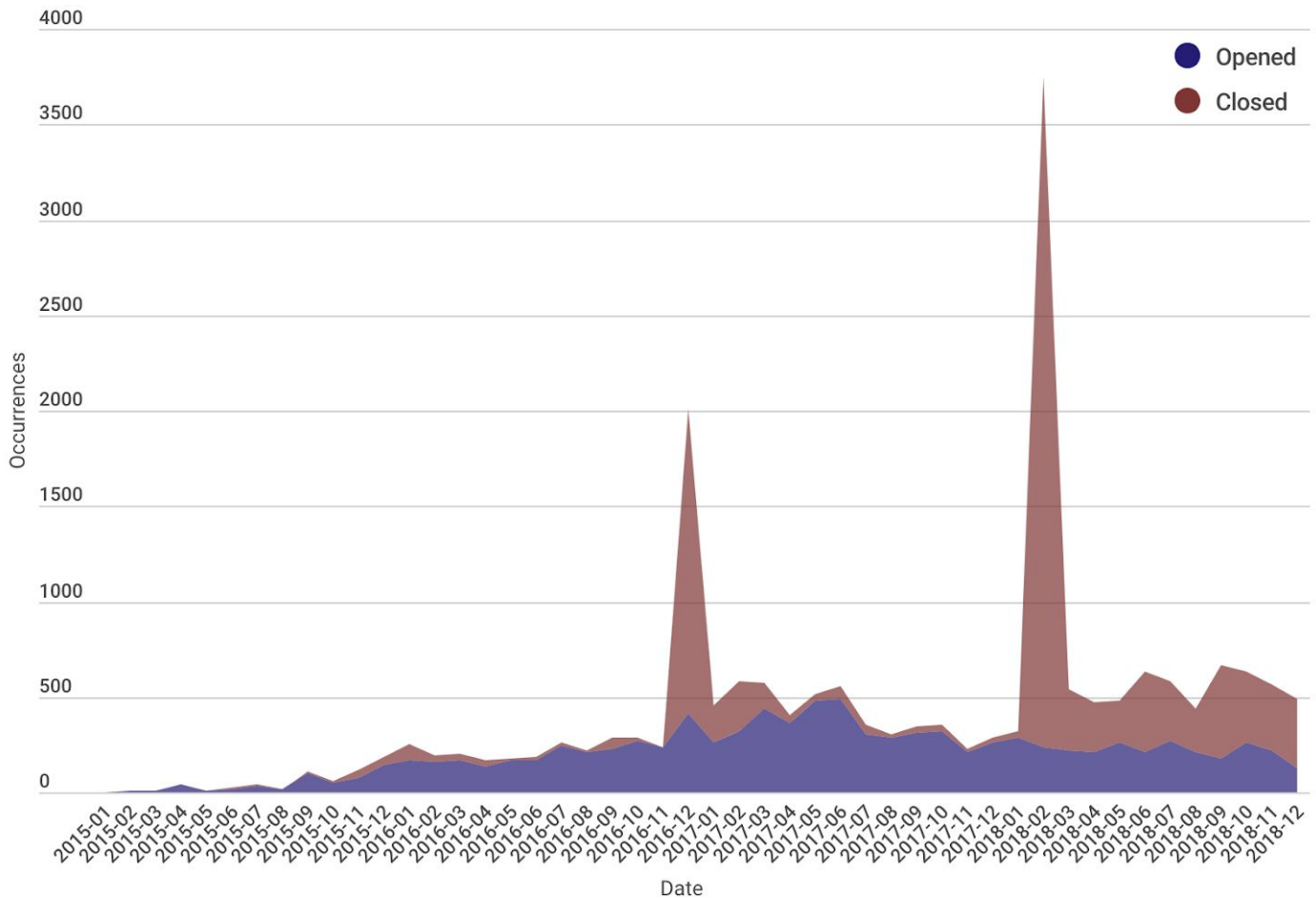
### 2.4.2 Análisis

Se utiliza el campo `Opportunity_Created_Date` agrupado por mes para graficar las oportunidades creadas.

Se utiliza el campo `Last_Modified_Date` agrupado por mes para graficar las oportunidades cerradas. Para esto se asume que, cuando una oportunidad pasa a algún estado cerrado (*Closed Won* o *Closed Lost*), no sufrirá nuevas modificaciones. De ese modo el campo mencionado anteriormente estaría marcando la fecha de cerrado de la oportunidad.

También se hace un filtrado por `Opportunity_ID`, de modo que cada oportunidad solo sume una vez al total.

El gráfico que describe la distribución es el siguiente.



### 2.4.3 Conclusión

En el gráfico se observan los siguientes puntos:

- En el último tiempo, se observa cómo se cierran más oportunidades de las que se abren. Esto podría indicar que la empresa en este momento tiene recursos suficientes para lidiar con la creación de más oportunidades.
- Dos anomalías son claramente visibles en los meses 12-2016 y 02-2018, donde se observa que la cantidad de oportunidades cerradas son mucho mayor a las abiertas. Estas oportunidades cerradas corresponden a las que no fueron cerradas en los meses posteriores al que fueron creadas.

## 2.5 Oportunidades ganadas y perdidas por estación

### 2.5.1 Introducción

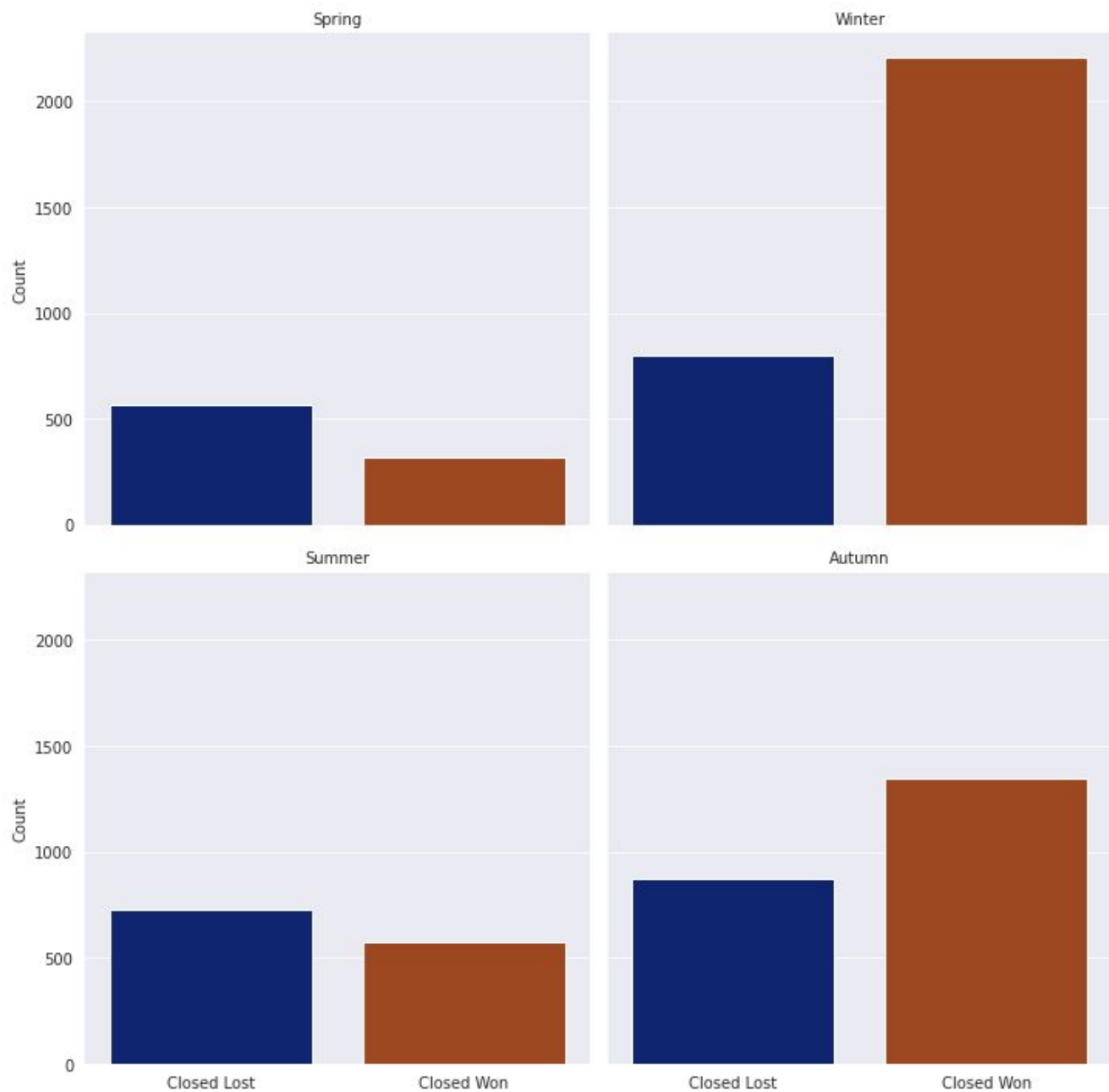
La empresa “Frío Frío” se dedica a la venta e instalación de equipos de aire acondicionado para grandes superficies. Suena lógico, entonces, pensar en la posibilidad de que las

estaciones del año (o más precisamente la variación de temperatura promedio) tengan un efecto en las ventas de esta empresa.

### 2.5.2 Análisis

Utilizamos el CSV con información de los países para identificar el hemisferio al que pertenece cada uno. Esto nos permite deducir correctamente la estación del año en la que se encuentran al momento de cerrar una oportunidad.

De esta manera podemos graficar la relación entre los casos cerrados como Closed Won y los Closed Lost para cada estación.



### 2.5.3 Conclusión

Como se puede observar en el gráfico, contrariamente a lo que se podría haber esperado, los casos cerrados como `Closed Won` aumentan en otoño y más considerablemente en invierno.

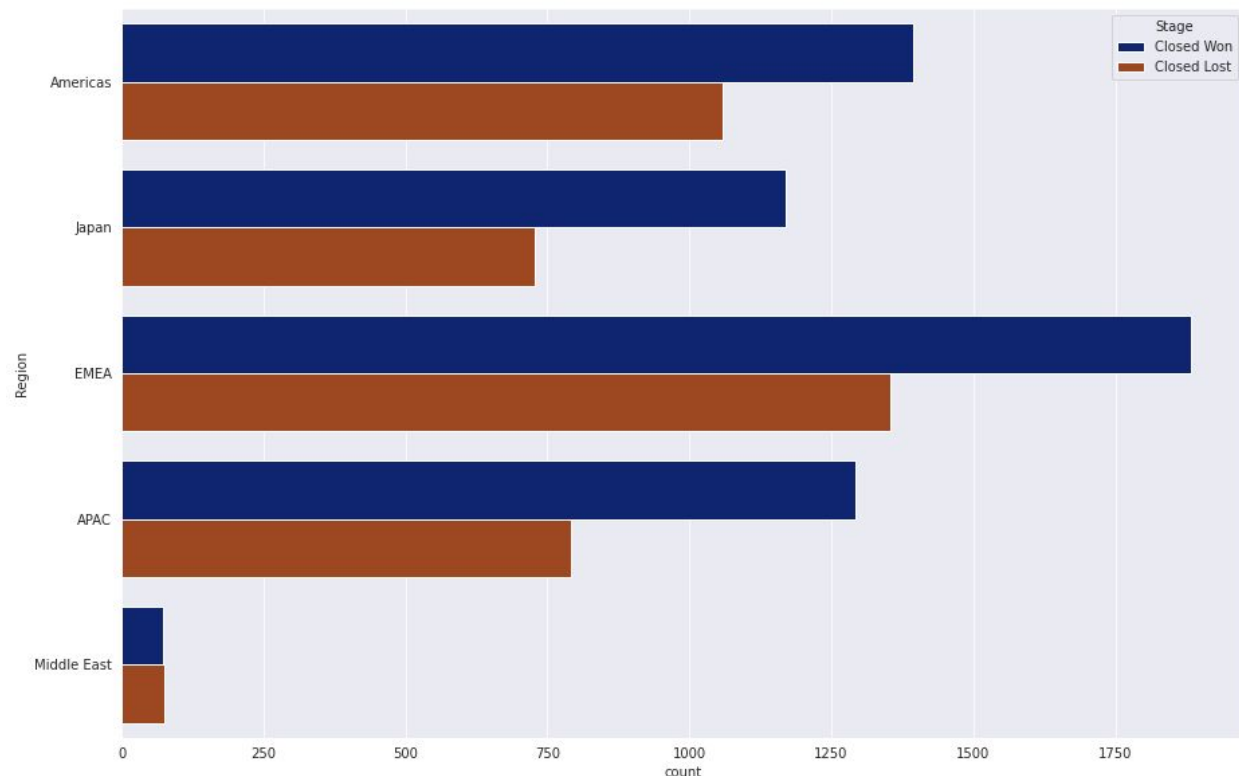
## 2.6 Oportunidades ganadas y perdidas por región

### 2.6.1 Introducción

En este apartado se realizó un análisis respecto de la variación de las oportunidades dependiendo de la región en donde se realizó la venta.

### 2.6.2 Análisis

Luego de filtrar las oportunidades que no estaban cerradas, y de quedarnos con solo un producto por oportunidad, se procedió a realizar un gráfico que muestre claramente cuáles eran las regiones que más aportan en cuanto al número de oportunidades y cómo se distribuía su porcentaje de éxito.





### 2.6.3 Conclusión

Como se puede observar en el gráfico, se destacan los siguientes puntos:

- La región correspondiente al Middle East aporta pocas oportunidades, pero aquellas oportunidades aportadas tienen un muy alto porcentaje de conversión a `Closed Won`.
- Las regiones de Americas, APAC y Japan tienen un aporte muy parecido en cuanto al número de oportunidades ganadas. Sin embargo, si observamos con cuidado, podemos ver que Americas presenta una cantidad de oportunidades perdidas significativamente mayor.
- La región de EMEA es la que más oportunidades aporta, tanto `Closed Won` como `Closed Lost`.

## 2.7 Frecuencia y éxito de los tipos de oportunidades

### 2.7.1 Introducción

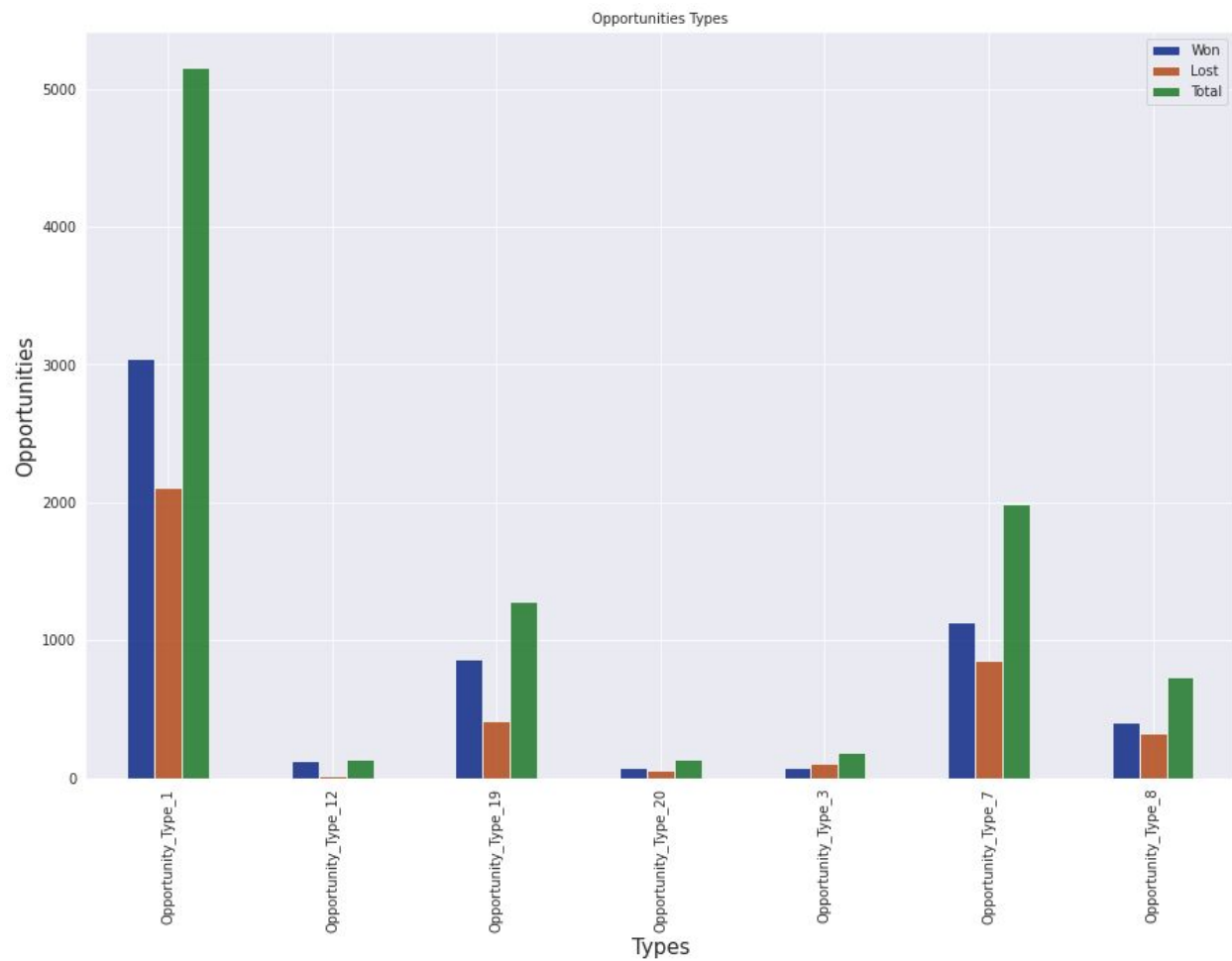
En este punto analizaremos si hay algunos tipos de oportunidades más frecuentes que otros o más fáciles de cerrar exitosamente.

### 2.7.2 Análisis

Lo primero que hacemos es concentrarnos en las oportunidades cerradas. Y como en casos anteriores, descartamos duplicados de oportunidades.

Una vez pasado el filtro creamos dos tablas, agrupando por tipo de oportunidad, en una el total de casos cerrados exitosos, y en otra los perdidos.

Agrupamos la información y notamos que hay tipos de oportunidades muy poco frecuentes, por lo que decidimos, para un mejor análisis y visualización, sacar los tipos cuyo total de cierres esté por debajo del 1% total de cierres de todas las oportunidades.



### 2.7.3 Conclusion

Se puede notar que hay cuatro tipos de oportunidades que son los más frecuentes, estas son las oportunidades de tipo 1, 19, 7 y 8.

Dentro de estas, las oportunidades de tipo 19 son las que mejor promedio de cierre exitoso tienen, cercano al 70%.

Con respecto a los menos frecuentes, tenemos un caso muy particular: el tipo 12 tiene un casi 90% de cierres exitosos.

En conclusión, en caso de tener que seleccionar tipos de oportunidades para potenciar, sería conveniente concentrarse en los tipos 19 y 12.

## 2.8 Éxito por vendedor

### 2.8.1 Introducción

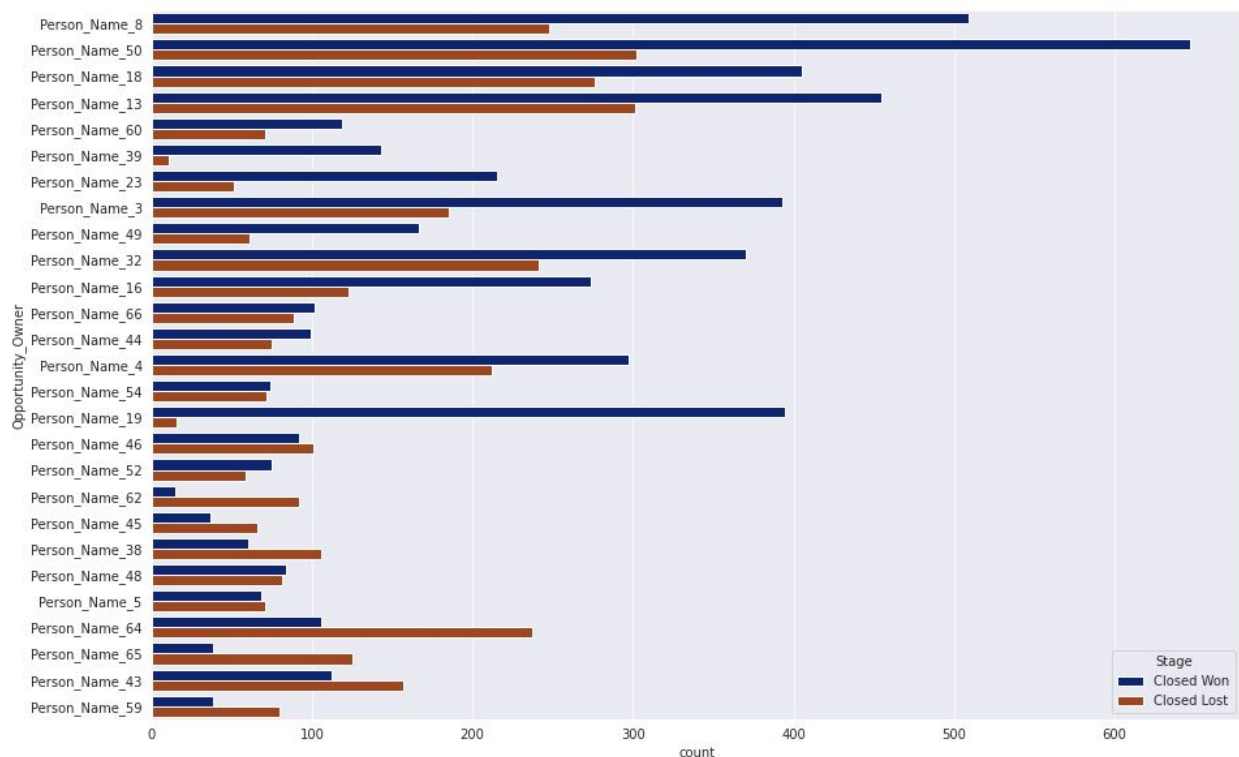
Cada oportunidad es llevada a cabo por un representante comercial, y este representante se hace cargo de la oportunidad durante toda la “vida” de la misma.

Por este motivo, consideramos una métrica esencial a tener en cuenta al éxito de cada vendedor, considerando “éxito” como la capacidad de cerrar una oportunidad como “Closed Won”.

### 2.8.2 Análisis

Para este análisis primero se calculó el total de oportunidades cerradas por vendedor, para así poder descartar a los vendedores con menos de 100 ventas, evitando caer en el error descrito por la ecuación de De Moivre. El número de ventas límite fue elegido teniendo en cuenta el promedio de ventas de todos los vendedores.

Se procedió entonces a graficar las oportunidades cerradas por vendedor, distinguiendo entre Closed Won y Closed Lost.



### 2.8.3 Conclusión

En el gráfico se puede observar que los vendedores pueden ser divididos en tres categorías dependiendo de su rendimiento:

- Buen rendimiento
  - ◆ El vendedor posee más oportunidades ganadas que perdidas.
  - ◆ Ejemplo: 8, 50, 18, 13
- Rendimiento medio
  - ◆ La diferencia entre las oportunidades ganadas y perdidas del vendedor no es significativa.
  - ◆ Ejemplo: Vendedores 48, 5, 66, 44.
- Rendimiento bajo
  - ◆ El vendedor presenta un mayor número de ventas perdidas.
  - ◆ Ejemplo: Vendedores 59, 65, 64, 38

Es pertinente hacer ciertas menciones especiales para vendedores específicos. Este es el caso de los vendedores 19 y 39, cuyo promedio de éxito ronda el 90%.

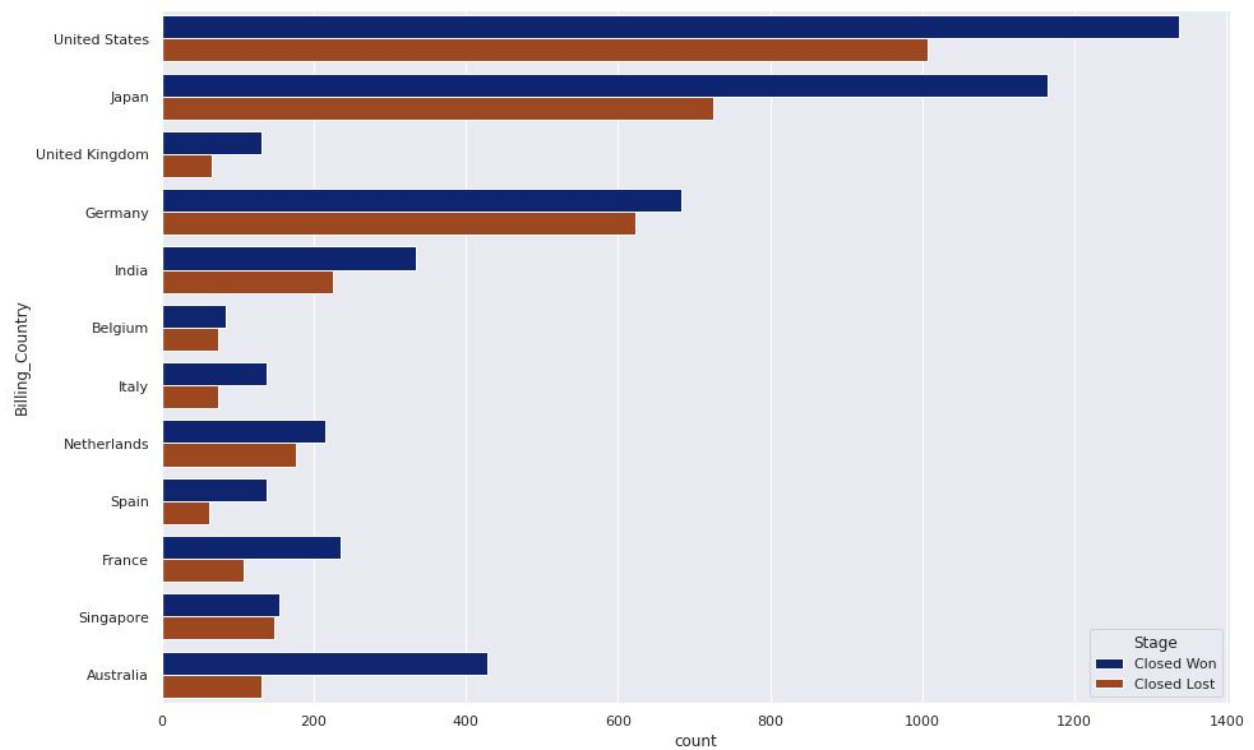
## 2.9 Oportunidades por país

### 2.9.1 Introducción

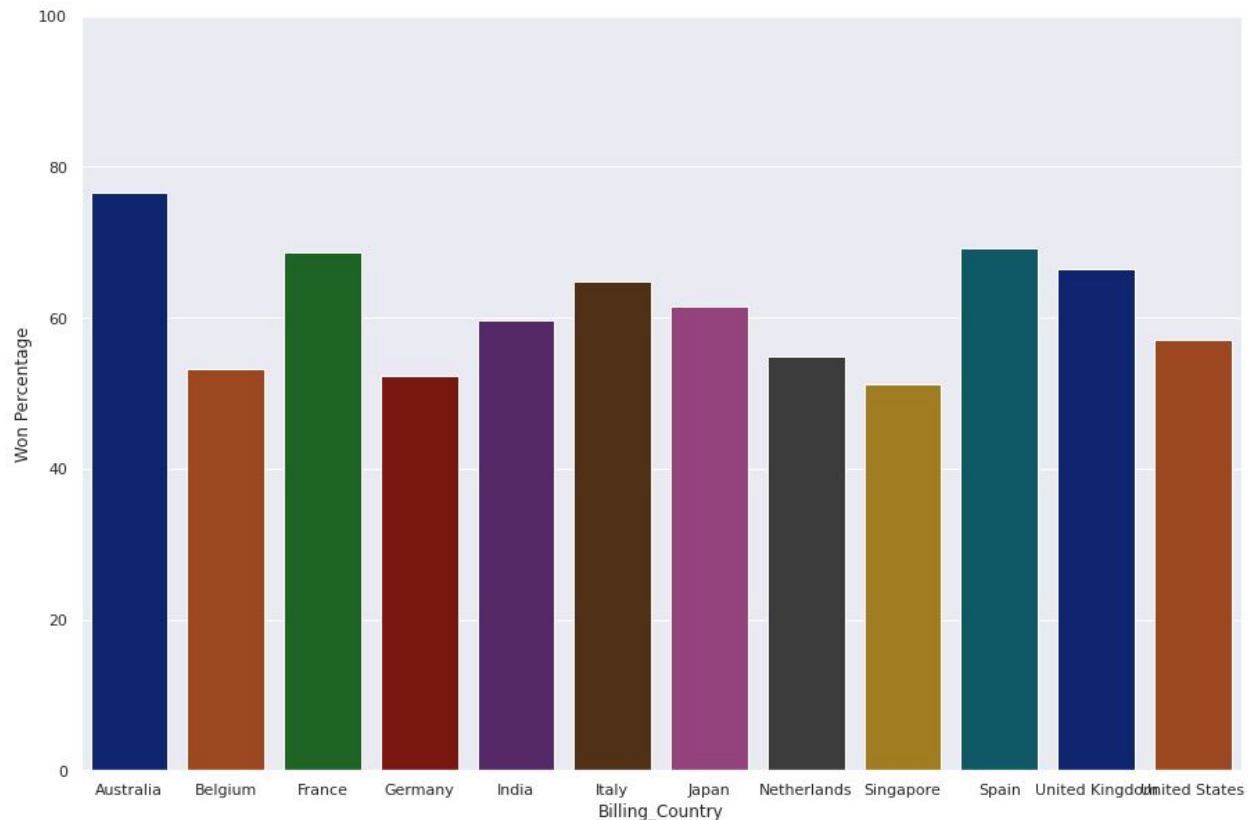
Luego del análisis por región de la sección 2.6, se quiso investigar más profunda y granularmente el porcentaje de éxito y su variación dependiendo del territorio. Es por esto que se procedió a realizar un análisis teniendo en cuenta el país asociado a cada oportunidad

### 2.9.2 Análisis

Para comenzar, se realizó un sencillo análisis sobre la cantidad de oportunidades ganadas y perdidas por país. Para poder llevar a cabo este análisis de manera efectiva se utilizó solamente a los veinte países con más oportunidades totales. El siguiente gráfico muestra esta estadística.



Las diferencias se pueden observar claramente, pero para realizar predicciones puede ser conveniente trabajar con probabilidades. Es por esto que se procedió a realizar un gráfico de las mismas.



Para concluir, cuando se trabaja con países o territorios, las distribuciones características terminan de hacer sentido cuando son presentadas en un mapa. Entonces se decidió unir estas estadísticas previas de la siguiente forma.



En el mapa se observa con un círculo color bordó a la cantidad neta de oportunidades (ganadas + perdidas), y sobre este círculo se presenta otro círculo de color rojo claro que representa el porcentaje de estas oportunidades que son Closed Won.

Por simplicidad, en el mapa se marcaron los cinco países con mayores oportunidades creadas. Gracias a este mapa se pueden apreciar las diferencias entre las cantidades netas y porcentuales de una mejor manera.

### 2.9.3 Conclusión

Las siguientes conclusiones, fueron determinadas a partir de los gráficos anteriores.

- Del top 5 de países Australia se encuentra en último lugar en cuanto a oportunidades cerradas. Sin embargo, es el mejor país en cuanto a porcentaje de conversión, cuyo valor está cercano al 80%.
- Estados Unidos y Japón son, por lejos, los que más oportunidades generaron. Si su porcentaje de conversión no fuese relativamente bajo (60% aproximado), su contribución sería una parte aún más significativa de todo el ecosistema.
- En segundo puesto en cuanto a porcentaje de conversión se encuentra España. Sin embargo, su contribución pasa desapercibida debido al bajo número de oportunidades creadas.

## 2.10 Aprobación y éxito

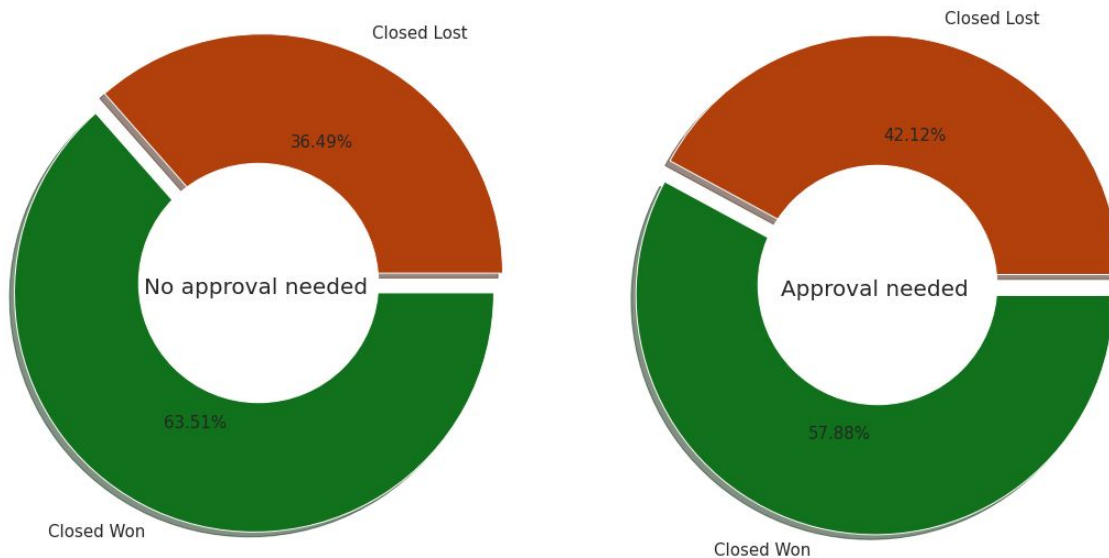
### 2.10.1 Introducción

En este apartado se busca establecer una relación entre la necesidad de aprobación de una oportunidad y su éxito. Existe una posibilidad de que la espera por aprobación termine en un decremento en la posibilidad de éxito de la oportunidad.

### 2.10.2 Análisis

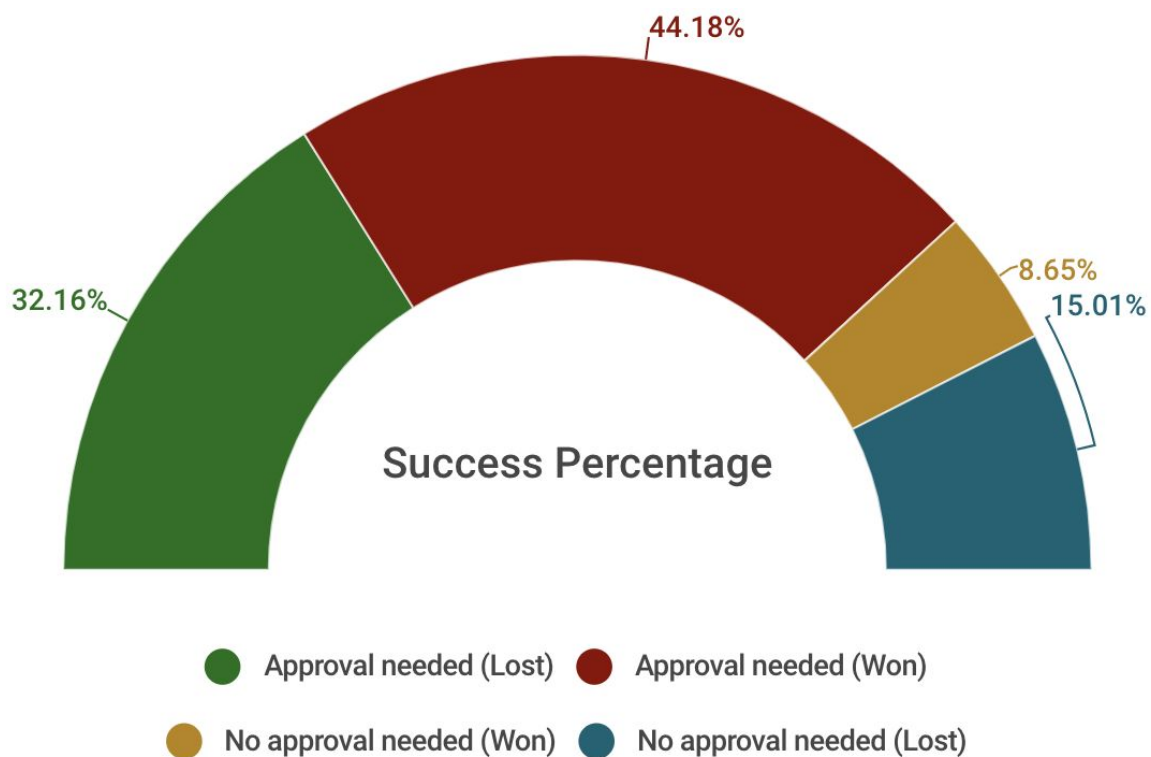
Para el siguiente análisis se utilizaron todas las oportunidades con un estado cerrado, y estas fueron agrupadas por su ID, para evitar así que las oportunidades con varios productos afecten el resultado.

En el siguiente gráfico se pueden ver los porcentajes de éxito para las oportunidades que requieren o no requieren aprobación.



Utilizando los gráficos de arriba, es difícil ver cuál es el porcentaje final del total que acaba en cada una de las cuatro opciones.

Es por este motivo que, antes de concluir este análisis, se opta por realizar un gráfico más. Un gráfico que muestre los porcentajes, pero calculados contra el total de las oportunidades.





### 2.10.3 Conclusión

Como se puede observar en la comparación de los gráficos, la diferencia entre la probabilidad de éxito es pequeña, rondando el 6%. Si bien es algo que podría tomarse en cuenta para predecir el resultado de una oportunidad, se concluye que su peso dentro del esquema general debería ser muy inferior.

## 2.11 Éxito por cliente

### 2.11.1 Introducción

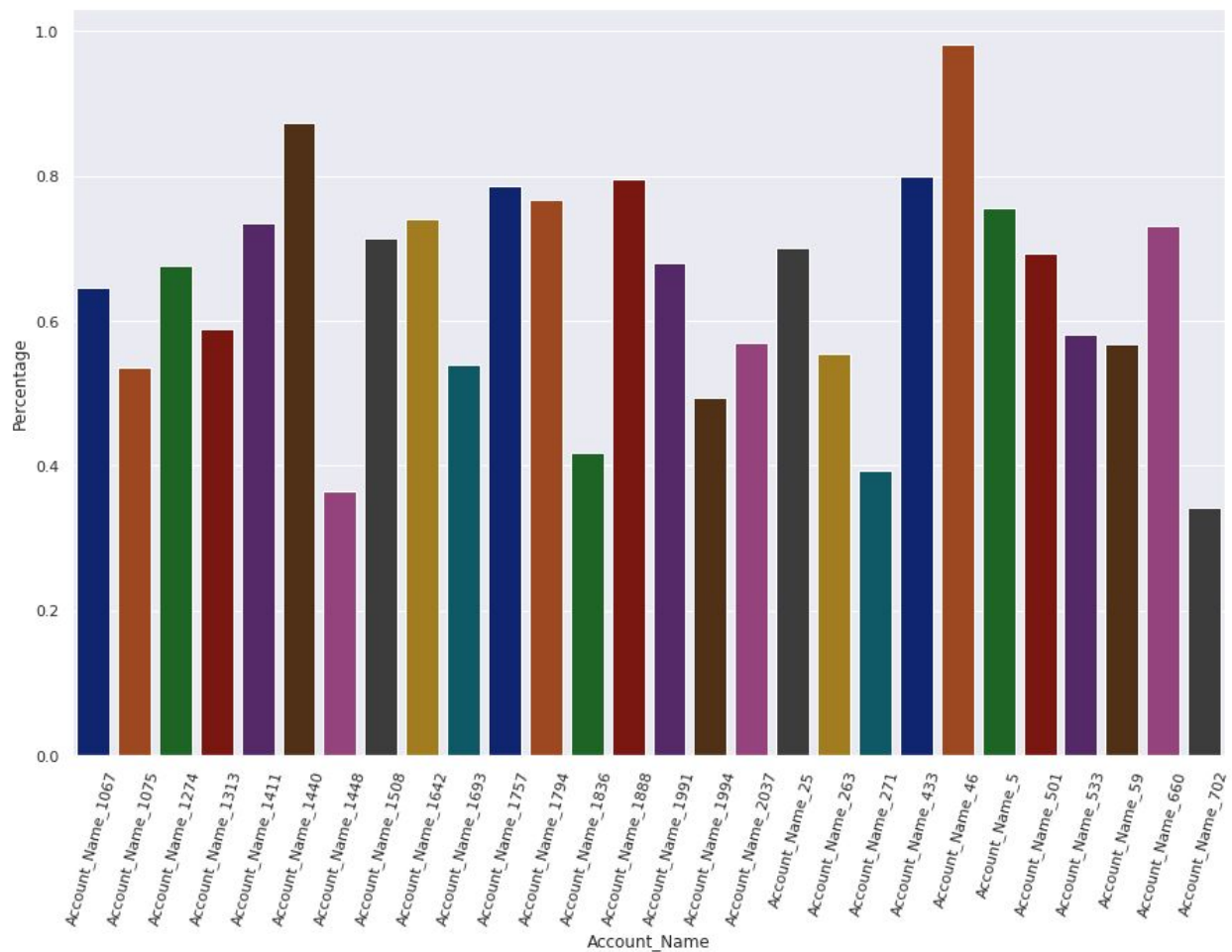
En este apartado se busca analizar el impacto de los diferentes clientes en el éxito o fracaso de sus respectivas compras. ¿Existirán los clientes conflictivos?

### 2.11.2 Análisis

Como en los otros apartados, nos vamos a concentrar en contar las oportunidades netas (ignorando la cantidad de productos) y en aquellas cuyo Stage sea Close Won o Close Lost.

El comienzo es bastante obvio. A continuación se presenta un diagrama que muestra el porcentaje de éxito para los clientes con más de 50 ventas.

Una vez más, elegimos un límite inferior para no caer en el error descubierto por Moivre.

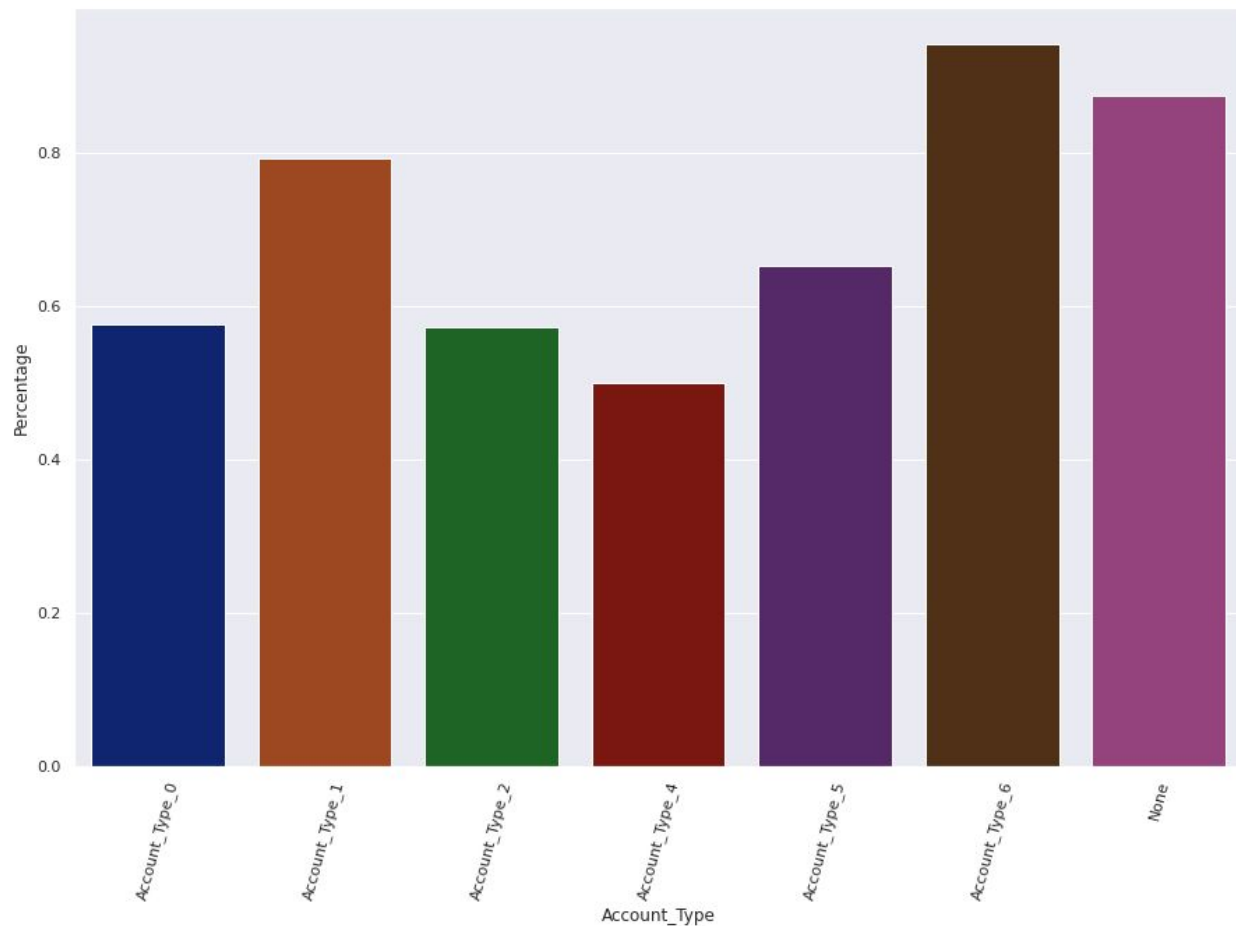


Por otro lado, analizando los datos se observa una alternativa: el `Account_Type`.

Relevando el set de datos encontramos que existen 6 tipos de cuentas distintos, y de que existen compradores sin ningún tipo de cuenta asociado. Igualmente, a estos últimos también los vamos a incluir en el análisis.

¿Acaso el tipo de cuenta afecta la probabilidad de éxito?

La respuesta a esa pregunta se encuentra en el gráfico a continuación.



### 2.11.3 Conclusión

Las conclusiones de este apartado se pueden dividir en dos partes, ya que se realizaron dos análisis distintos.

→ Conclusiones por `Account_Name`

- ◆ Observamos que se destacan los clientes número 1440, 433 y 46, los tres superan el 80% de oportunidades cerradas con éxito. Sin embargo, el comprador número 46 merece una mención especial, ya que su promedio de éxito roza el 100%
- ◆ Por el lado de los clientes conflictivos, tenemos a los clientes con número 1448, 271, 1836 y 702. Todos por debajo del 40% de éxito.

→ Conclusiones por `Account_Type`

- ◆ Se puede observar que los tipos de cuentas de los clientes parecen dividirse en dos claros grupos:
  - Por un lado los que superan (o rozan) el 80%. Estos son los tipos de cuenta 1, 6 y None.
  - Por otro lado, los que están por debajo (o rozan) el 60%. En este grupo hacen su aparición los tipos 0, 2, 4, 5.

### 3. Conclusiones generales

Analizando las conclusiones de los diferentes apartados, se busca hallar conclusiones generales sobre las variables que afectan el resultado final de una oportunidad.

El objetivo final es poder realizar una predicción sobre las nuevas oportunidades, y para esto, consideramos que el análisis de las siguientes variables son claves.

- El impacto del vendedor no puede ser ignorado, habiendo una gran diferencia apreciable entre las oportunidades asignadas a uno u otro vendedor.
- Cierta tipo de oportunidades tienen un porcentaje de éxito significativamente mayor, esta variable también es importante y debería ser tomada en cuenta.
- No todas las estaciones modifican en gran medida el resultado de las oportunidades. Sin embargo, hay una estación que presenta una gran diferencia en el número de oportunidades ganadas con respecto al resto: el invierno.
- Dos países en particular (Estados Unidos y Japón) se llevan la mayor parte del mercado.
- Australia es un país en donde las oportunidades tienen un muy alto porcentaje de conversión.
- La diferenciación por nivel de aprobación de una oportunidad es despreciable con respecto a la variación detectada en otros apartados.
- Los clientes tienen un impacto directo sobre el éxito de la oportunidad y son una variable a tener en cuenta. Tanto a través de los `Account_Type`, como a través de sus respectivos `Account_Name`.

Estas fueron las conclusiones generales. De ser necesario profundizar, más particularidades se encuentran dispersas por los distintos apartados.

Así concluye nuestro análisis exploratorio del set de datos de esta empresa. Nos preparamos a continuación, con toda la información recaudada en este trabajo práctico, para idear un esquema que nos permita predecir el destino de una oportunidad al instante en que es creada, o incluso antes.