

Organización de Datos

TP N°1: Análisis Exploratorio

2°C 2020

Equipo The Devs Army:

Nicolas Hazan	100849
Miguel Angel Fonzalida	86125
Agustin Tardá	100932
Kevin Ariel Gadacz	104531

Repositorio:

<https://github.com/miguelAfonzalida/TP1-OrgaDatos>



Índice

Índice	2
Resumen	3
Descripción general del dataset	3
Pre Procesamiento	6
Stage	7
Región	9
Hemisferio	13
Burocracia	15
Territorio	21
Estaciones del año	33
Toneladas de Refrigeración	35
Vendedores Responsables de la Oportunidad	37
Monto Total	40
Oportunidades	47
Conclusiones	50

Resumen

En este informe vamos a analizar un set de datos que contiene información sobre distintas oportunidades, con el fin de próximamente generar un modelo de predicción que pueda clasificar si cada una de las oportunidades termina en éxito o fracaso. En esta primera instancia realizaremos un análisis exploratorio del dataset con el propósito de encontrar información relevante y descartar información irrelevante. Durante el desarrollo mostraremos distintas visualizaciones dónde se podrán apreciar distintas conclusiones sobre el conjunto de datos. El informe va a estar dividido en secciones con visualizaciones, su correspondiente análisis y conclusión del mismo.

Descripción general del dataset

El dataset cuenta con unas 16947 filas y 52 columnas las cuales se detallan a continuación:

- **ID:** id único del registro (Entero).
- **Región:** región de la oportunidad (Categórica).
- **Territory:** territorio comercial de la oportunidad (Categórica).
- **Pricing, Delivery_Terms_Quote_Approval:** variable que denomina si la oportunidad necesita aprobación especial de su precio total y los términos de la entrega (Binaria).
- **Pricing, Delivery_Terms_Approved:** variable que denomina si la oportunidad obtuvo aprobación especial de su precio total y los términos de la entrega (Binaria).
- **Bureaucratic_Code_0_Approval:** variable que denomina si la oportunidad necesita el código burocrático 0 (Binaria).
- **Bureaucratic_Code_0_Approved:** variable que denomina si la oportunidad obtuvo el código burocrático 0 (Binaria).
- **Submitted_for_Approval:** variable que denomina si fue entregada la oportunidad para la aprobación (Binaria).
- **Bureaucratic_Code:** códigos burocráticos que obtuvo la oportunidad (Categórica).
- **Account_Created_Date:** fecha de creación de la cuenta del cliente (Datetime).
- **Source:** fuente de creación de la oportunidad (Categórica).
- **Billing_Country:** país donde se emite la factura (Categórica).
- **Account_Name:** nombre de la cuenta del cliente (Categórica).
- **Opportunity_Name:** nombre de la oportunidad (Categórica).
- **Opportunity_ID:** id de la oportunidad (Entero).

- **Sales_Contract_No:** número de contrato (Entero).
- **Account_Owner:** vendedor del equipo comercial responsable de la cuenta cliente (Categórica).
- **Opportunity_Owner:** vendedor del equipo comercial responsable de la oportunidad comercial (Categórica).
- **Account_Type:** tipo de cuenta cliente (Categórica).
- **Opportunity_Type:** tipo de oportunidad (Categórica).
- **Quote_Type:** tipo de presupuesto (Categórica).
- **Delivery_Terms:** términos de entrega (Categórica).
- **Opportunity_Created_Date:** fecha de creación de la oportunidad comercial (Datetime).
- **Brand:** marca del producto (Categórica).
- **Product_Type:** tipo de producto (Categórica).
- **Size:** tamaño del producto (Categórica).
- **Product_Category_B:** categoría 'B' del producto (Categórica).
- **Price:** precio (Decimal).
- **Currency:** moneda (Categórica).
- **Last_Activity:** fecha de la última actividad (Datetime).
- **Quote_Expiry_Date:** fecha de vencimiento del presupuesto (Datetime).
- **Last_Modified_Date:** fecha de ultima modificación en la oportunidad (Datetime).
- **Last_Modified_By:** usuario responsable de la última modificación en la oportunidad (Categórica).
- **Product_Family:** familia de producto (Categórica).
- **Product_Name:** nombre del producto (Categórica).
- **ASP_Currency:** moneda del precio promedio (Categórica).
- **ASP:** (Average Selling Price) precio promedio a la venta (Decimal).
- **ASP_(converted)_Currency:** moneda del precio promedio convertido en la variable (Categórica)
- **ASP_(converted):** precio promedio a la venta convertido a otra moneda (Decimal).
- **Planned_Delivery_Start_Date:** límite inferior del rango previsto para la fecha de entrega (Datetime).
- **Planned_Delivery_End_Date:** límite superior del rango previsto para la fecha de entrega (Datetime).
- **Month:** mes-año de Planned_Delivery_Start_Date (Fecha).
- **Delivery_Quarter:** trimestre de Planned_Delivery_Start_Date (Categórica).
- **Delivery_Year:** año de Planned_Delivery_Start_Date (Fecha).
- **Actual_Delivery_Date:** fecha real de la entrega (Datetime).
- **Total_Power:** potencia del producto (Entero).
- **Total_Amount_Currency:** moneda del monto total (Decimal).



- **Total_Amount:** monto total (Decimal).
- **Total_Taxable_Amount_Currency:** moneda del monto gravado total (Categórica).
- **Total_Taxable_Amount:** monto gravado total (Decimal).
- **Stage:** variable target. Estado de la oportunidad (Categórica).
- **Prod_Category_A:** categoría 'A' del producto (Categórica).
- **TRF:** Toneladas de refrigeración (Entero). Es una unidad de potencia.

En las siguientes secciones desarrollaremos el análisis de cada uno de los campos contenidos en el dataset.

Pre Procesamiento

- En un primer análisis se pudo ver que hay **16947** registros en el set de datos entre los cuales existen duplicados (tienen todos los campos con el mismo valor excepto por la comuna “ID”). Se pudo detectar 619 registros duplicados, de los cuales nos quedamos con solo la primera aparición de los mismos, por lo cual se terminaron filtrando 448. Al hacer esta limpieza observamos que seguimos conservando la mismas cantidades de oportunidades (**9841**) pero la cantidad total de registros pasa de **16947** a **19499** por lo tanto se eliminaron un **2,64%** de los registros iniciales.
- Luego para la limpieza de regiones y territorios se observó casos en que Japón estaba como Región y en su Territory figuraba None, por lo tanto se decidió asignarle a esos casos Japón como Territory y como Región a Asia-Pacífico (APAC).
- Para la Región que aparecía como Américas se modificó para que pase a llamarse América.
- Para la región que aparecía como EMEA se modificó para que pase a llamarse Europa.
- Para la región que aparecía como APAC se modificó para que pase a llamarse Asia-Pacífico.
- Para la región que aparecía como Middle East se modificó para que pase a llamarse Oriente Medio.
- También se corrigieron asignaciones de Territorios a sus correspondientes Regiones para que los datos sean congruentes.
- Se clasificaron los territorios según el Hemisferio al que pertenecen, para así poder analizar el set de datos teniendo en cuenta este nuevo atributo. Además esta clasificación nos ayudará luego a calcular la estación del año en que se crearon las oportunidades para cada cliente para cada territorio.
- Se agregaron las estaciones del año en las que se crearon las oportunidades y para las cuales se tuvo en cuenta el Hemisferio perteneciente a las mismas, como se indicó anteriormente.

Stage

Stage es nuestra variable Target, la cual nos indica el estado de la oportunidad, es decir, si la oportunidad resultó en éxito o no (Closed Won o Closed Lost respectivamente).

Se observó cómo además de los estados que indican si una oportunidad resultó exitosa o no (estados finales), tenemos otros 3 estados (Proposal, Negotiation y Qualification) que corresponden a estados intermedios que tienen las oportunidades dentro del pipeline que desarrolla la empresa.

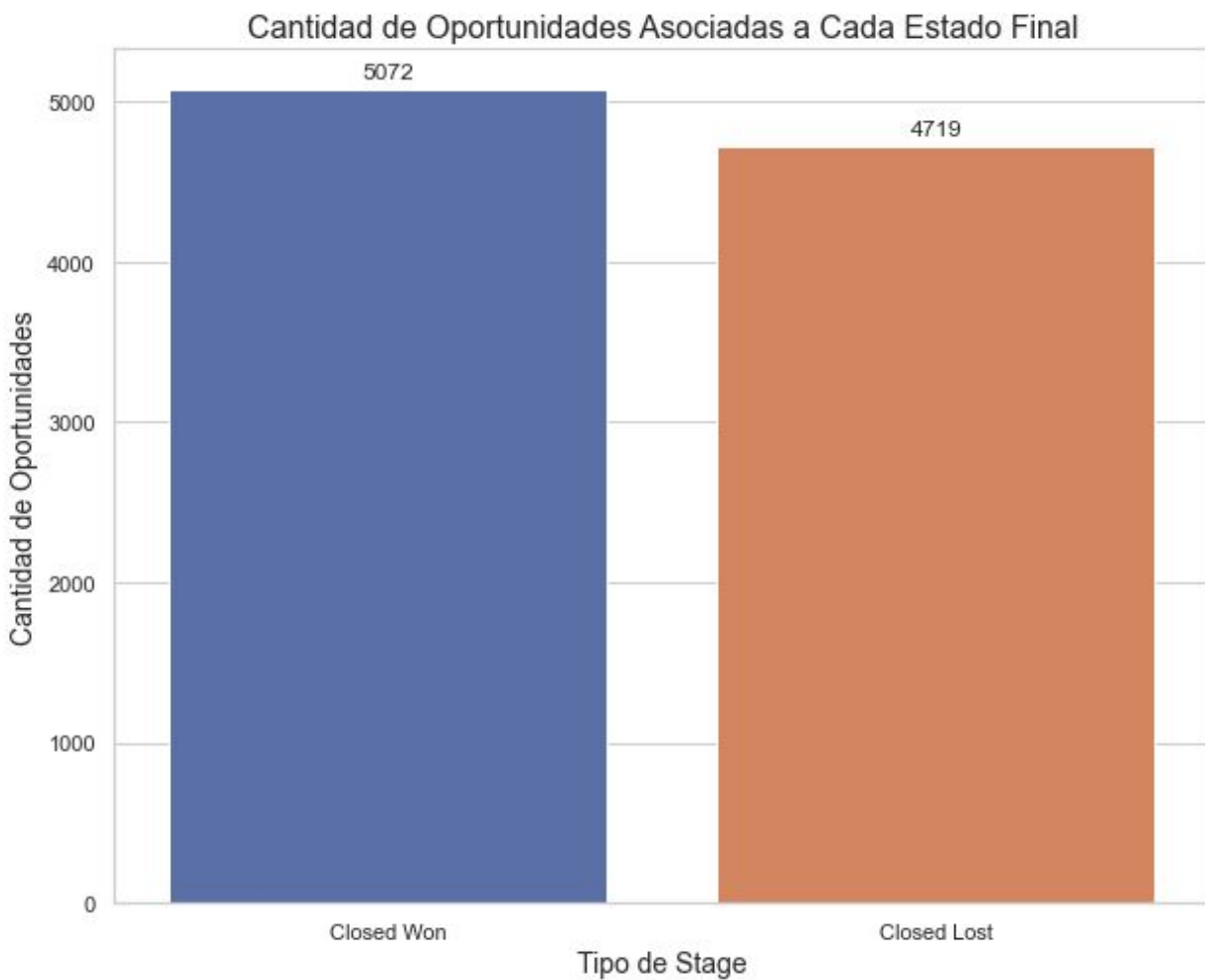
Como con esta variable se estará luego tratando de predecir la “Probabilidad de éxito” para cada oportunidad, o sea ¿Cuál es la probabilidad de que la oportunidad se convierta en un caso Closed Won? La primera pregunta que surge es qué decisión se debe tomar con dichos estados? Al no tener una resolución, debemos asignar arbitrariamente algún estado final? ¿Debemos filtrar estos casos?

Luego de analizar esos registros con estos estados intermedios y verificar que corresponden a oportunidades que no tuvieron un desenlace exitoso o no y teniendo en cuenta que corresponden a muy pocos casos, se tomó la decisión de filtrar estos registros y quedarnos solamente con los que tienen un estado final definido.

Otro punto interesante a tener en cuenta es que se pueden desarrollar distintas operaciones dentro de una misma oportunidad, es decir que para una misma oportunidad podemos tener varios ítems (registros).

Dicho esto nos resulta primeramente analizar:

¿Cómo se distribuyen las oportunidades en función de si las mismas resultaron o no exitosas?



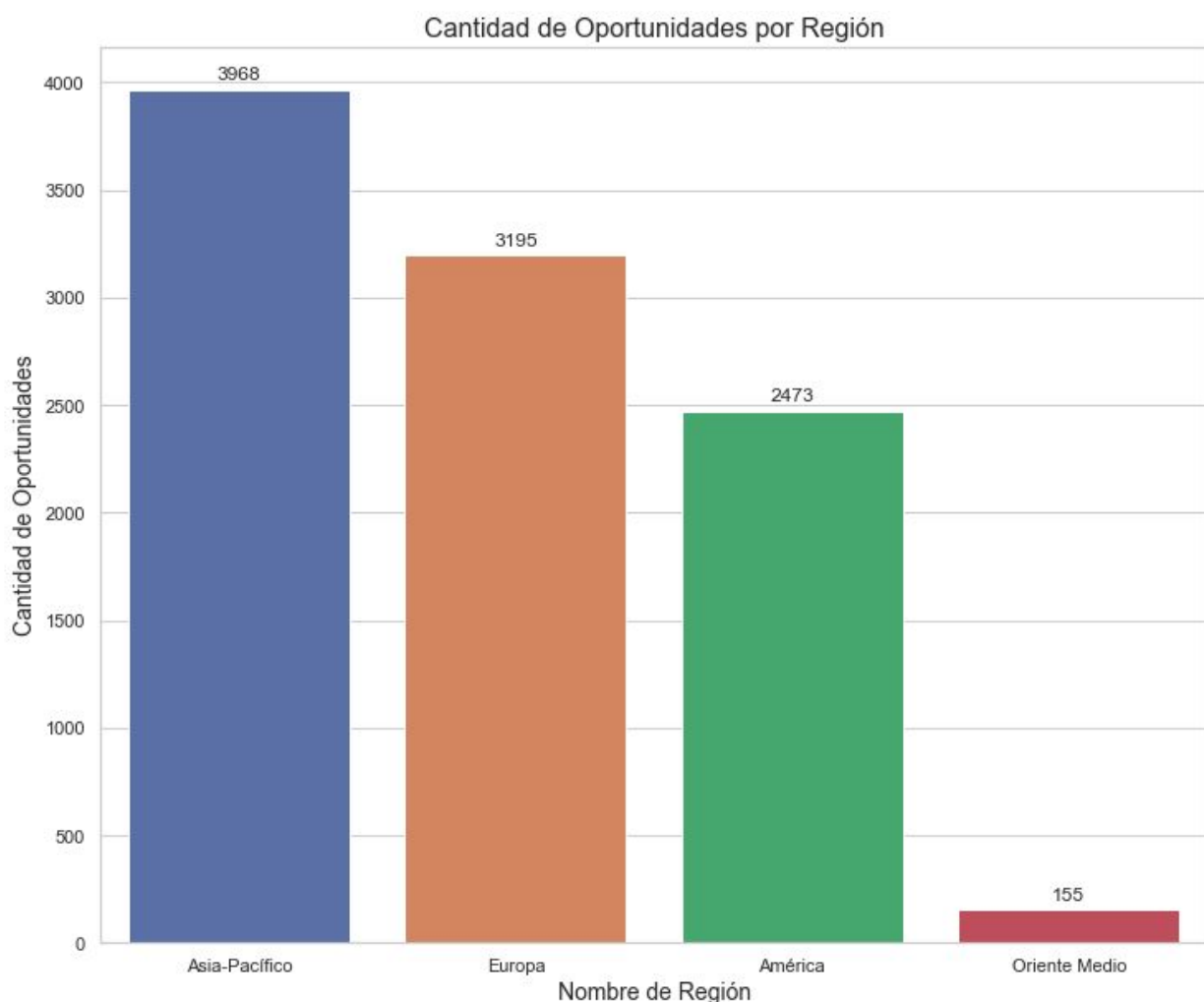
En esta visualización se puede ver como queda definida la cantidad de oportunidades según el estado final (Closed Won y Closed Lost)

Se observa que en el set de datos predominan las oportunidades que resultaron exitosas con 5072 casos (51.8%) frente a las No exitosas con 4719 casos (48.2%)

Región

La primer pregunta interesante que nos hacemos frente al análisis de la variable Región es:

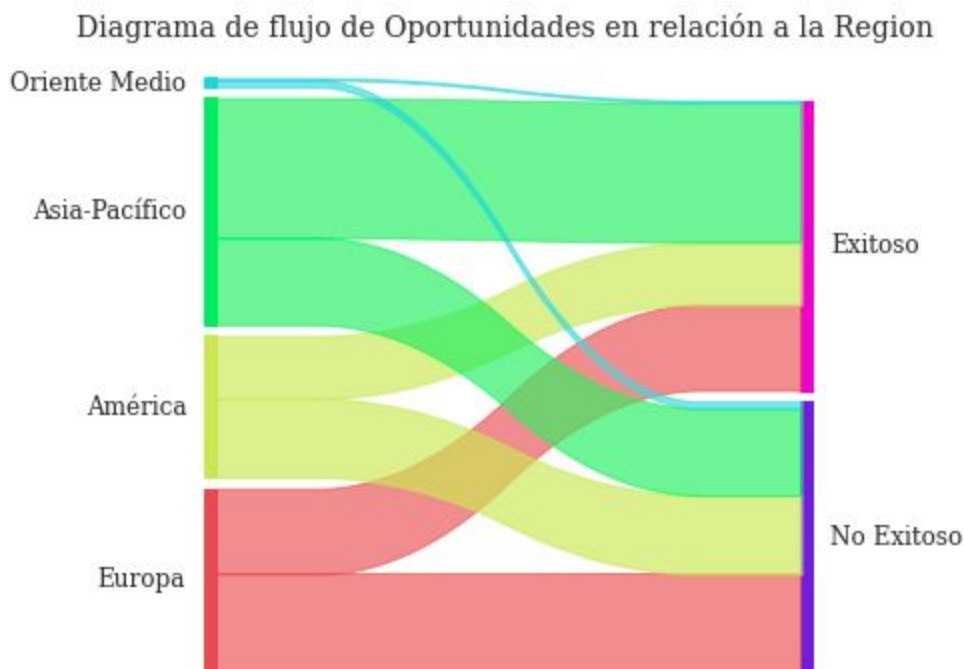
¿Cuántas regiones distintas tenemos en nuestro set de datos y cómo se distribuyen la cantidad de oportunidades en las mismas?



Como se puede ver la Región Asia-Pacífico es la predominante con 3968 oportunidades, es importante destacar que el valor alcanzado por esta Región, tiene mucho que ver con lo realizado en el preprocesamiento de los datos. A esta Región se le asignó Japón como territorio perteneciente. Este dato no es menor ya que Japón representa el 48% del total de oportunidades de la Región Asia-Pacífico y un 19% del

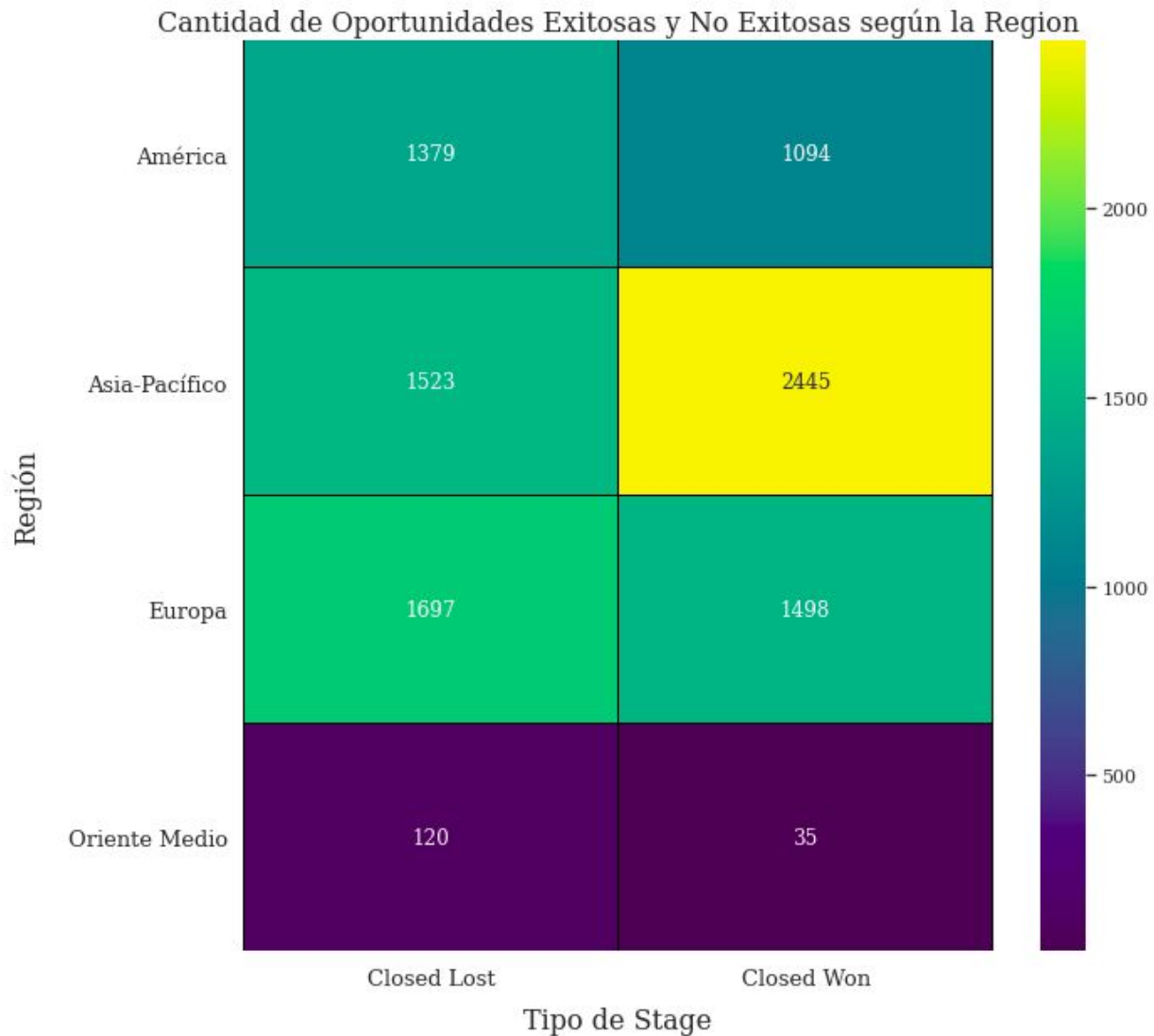
total de oportunidades globales. Las Regiones que le siguen son Europa y América en cantidad de oportunidades con valores 3195 y 2473 respectivamente. Finalmente Oriente Medio es la región que presenta menos oportunidades para la empresa con solamente 155.

¿Cómo se distribuyen las oportunidades exitosas y no exitosas según las distintas Regiones?



Asia-Pacífico, que como vimos es la Región que más oportunidades presenta y también es la que tiene mayor cantidad de casos exitosos y no exitosos como era de esperar. Pero notamos que su flujo hacia los casos exitosos en dicha región es mayor que a los no exitosos, mientras que las demás Regiones a simple vista presentan una distribución bastante similar para casos Exitosos y No Exitosos.

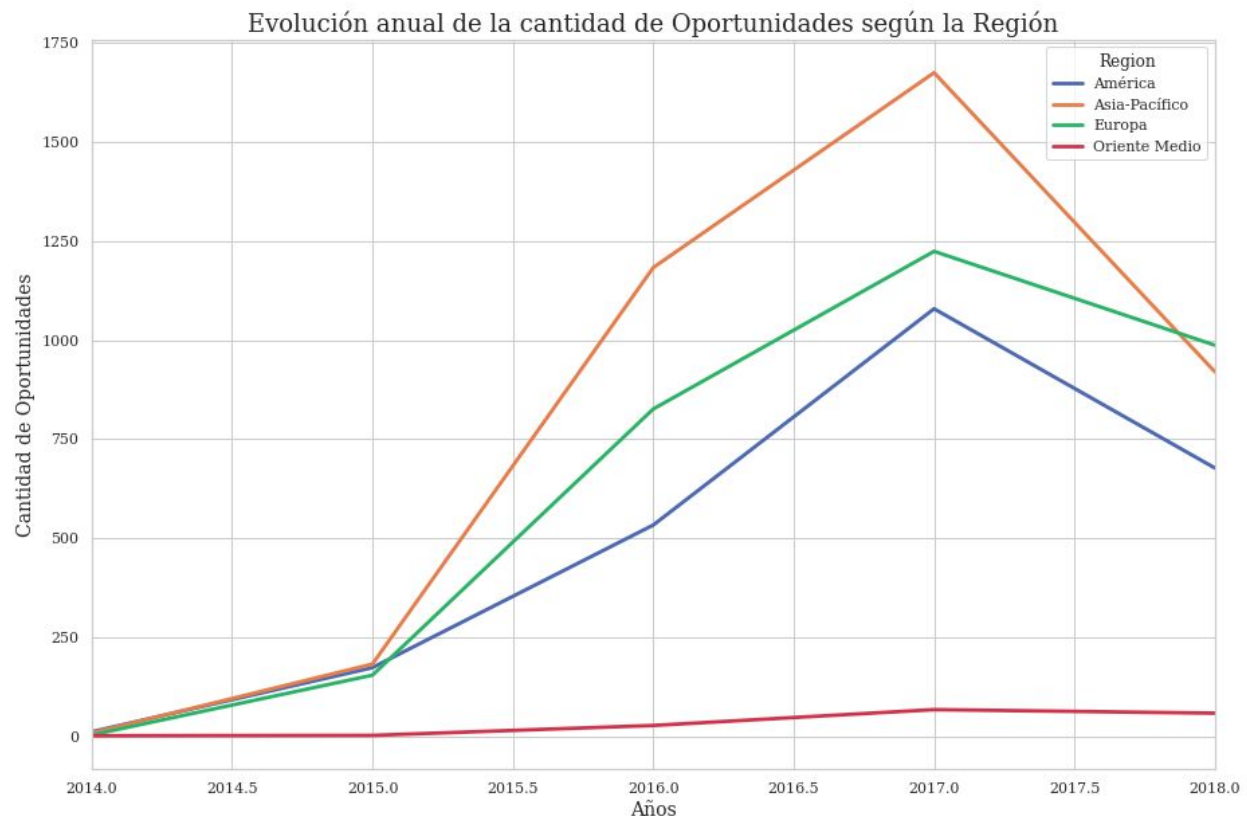
Si se quisiera cuantificar esas distribuciones de Oportunidades en Exitosas o No Exitosas según cada Región. ¿Cuáles son esas cantidades?



Al cuantificar los valores de casos Exitosos y No Exitosos asociados a cada región podemos ver varias cosas interesantes. Para Asia-Pacífico, no sólo las cantidades exitosas son mayores a las no exitosas sino que representan una alta cantidad del total de oportunidades de dicha región, más precisamente un 62%. Europa contrariamente presenta mayor cantidad de oportunidades fracasadas en relación a su total de oportunidades (1498 casos exitosos frente a 1697 no exitosos) un 53% aproximadamente. Para las dos Regiones restantes (América y Oriente Medio) los

casos no exitosos también son los que predominan en el total de oportunidades de dichas regiones.

¿Cómo es la evolución anual de la cantidad de Oportunidades según la Región?



Oriente Medio en general presenta un comportamiento prácticamente constante con muy bajas cantidades de oportunidades, las cuales tuvieron un tenue incremento entre 2016 y 2017.

Las regiones América, Europa y Asia-Pacífico de 2014 a 2015 notamos que presentan un incremento en cantidad de oportunidades. Asia-Pacífico presenta un marcado incremento (mayor pendiente) que las dos primeras. América y Europa en este período tienen una pendiente similar, con la primera Región levemente por encima de la segunda.

Asia-Pacífico continúa con el crecimiento en cantidad de oportunidades de 2015 a 2017 siendo más notorio su incremento en el período 2015-2016 que en 2016-2017. Luego

de 2017 a 2018 ocurre una brusca caída, terminando con cantidades inferiores a Europa.

América y Europa también continúan con el crecimiento en cantidad de oportunidades de 2015 a 2017 pero ésta última supera a la primera antes de la mitad del 2015 con una pendiente positiva constante hasta el 2017. América por su lado de 2015 a 2016 tuvo una menor pendiente positiva pero de 2016 a 2017 muestra una mayor recuperación. Luego, de 2017 al 2018 ambas Regiones presentan pendiente negativas denotando caídas en las oportunidades, aunque no tan bruscas como Asia-Pacífico.

Hemisferio

¿Cómo se distribuyen las oportunidades exitosas y no exitosas según los distintos Hemisferios?

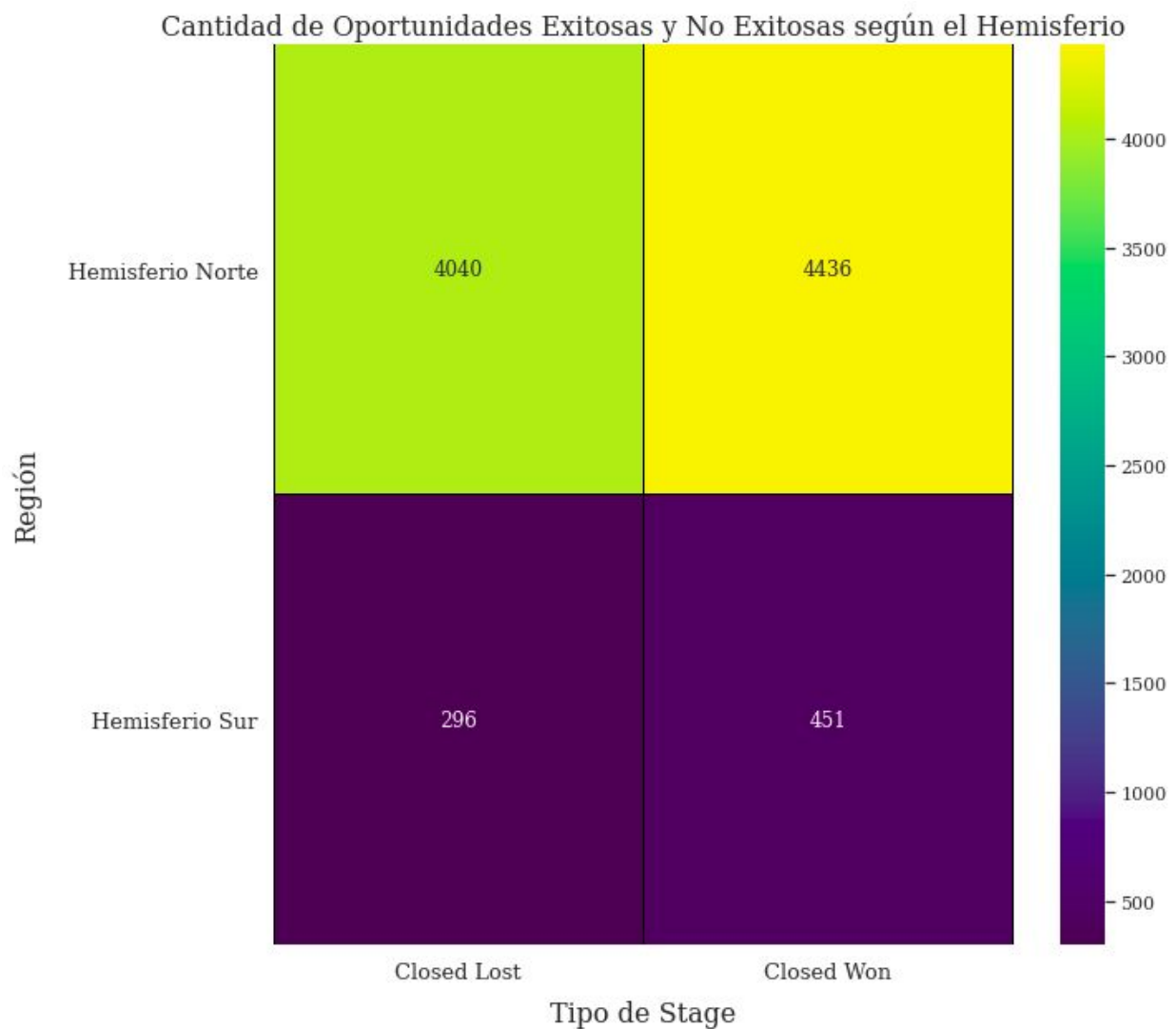
Diagrama de flujo de Oportunidades en relación al Hemisferio



Tomando los hemisferios norte y sur, se puede observar a simple vista que el caudal mayor por una diferencia importante es de oportunidades realizadas en el hemisferio

norte, donde la diferencia entre las oportunidades que se realizaron con éxito y las que fracasaron es muy similar. De forma similar, en una proporción mucho menor, en el hemisferio sur se puede observar a simple vista que la relación entre éxito/fracaso es muy similar, aunque se nota que las operaciones exitosas son las que predominaron.

Si quiero cuantificar esas distribuciones de Oportunidades en Exitosas o No Exitosas según cada Hemisferio. ¿Cuáles son esas cantidades?



Del mismo modo que se visualizaba en el gráfico anterior, ahora podemos visualizar con un heap map y tener una mayor noción de las cantidades. Se puede observar que

para ambos hemisferios la cantidad de operaciones fracasadas resultó menor que la cantidad de operaciones exitosas.

Burocracia

¿Qué sucede con las peticiones de burocracia?

Pricing, ms_Quote_Appr	Delivery_Terms_Quote_Appr	Pricing, Delivery_Terms_Approved	Bureaucratic_Code_0_Approval	Bureaucratic_Code_0_Approved	Submitted_for_Approval	Bureaucratic_Code
1		1	1	1	0	Bureaucratic_Code_4
0		0	0	0	0	Bureaucratic_Code_4
0		0	0	0	0	Bureaucratic_Code_4
1		0	1	0	0	Bureaucratic_Code_5
1		0	1	0	0	Bureaucratic_Code_5

Hay distintas columnas a tener en cuenta para este apartado.

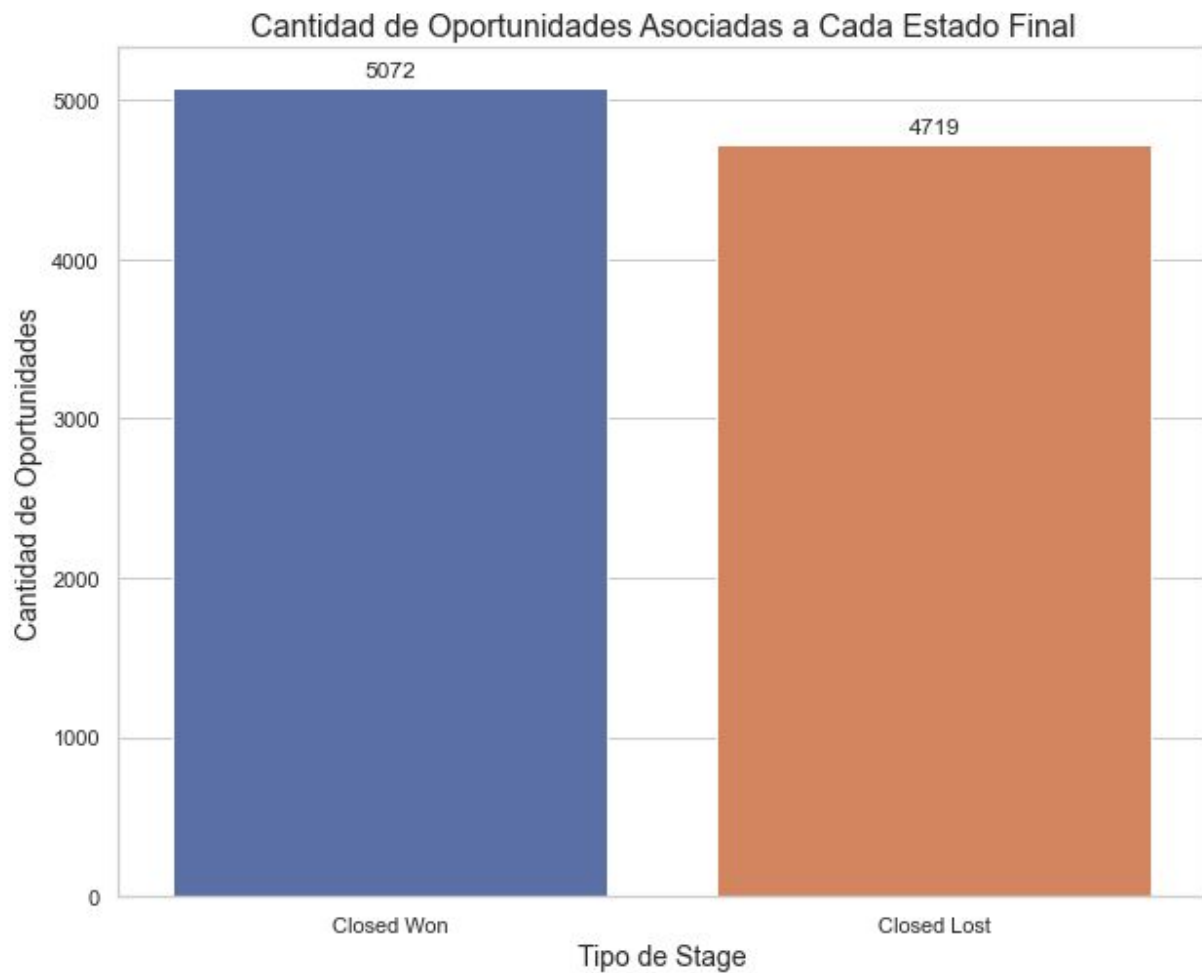
Pricing, Delivery_Terms_Quote_Appr y **Bureaucratic_Code_0_Approval** son respectivamente campos donde se solicita una aprobación de envío y precio, y del código burocrático 0.

Pricing, Delivery_Terms_Approved y **Bureaucratic_Code_0_Approved** son respectivamente los campos donde se aprueba, o no, las peticiones mencionadas anteriormente.

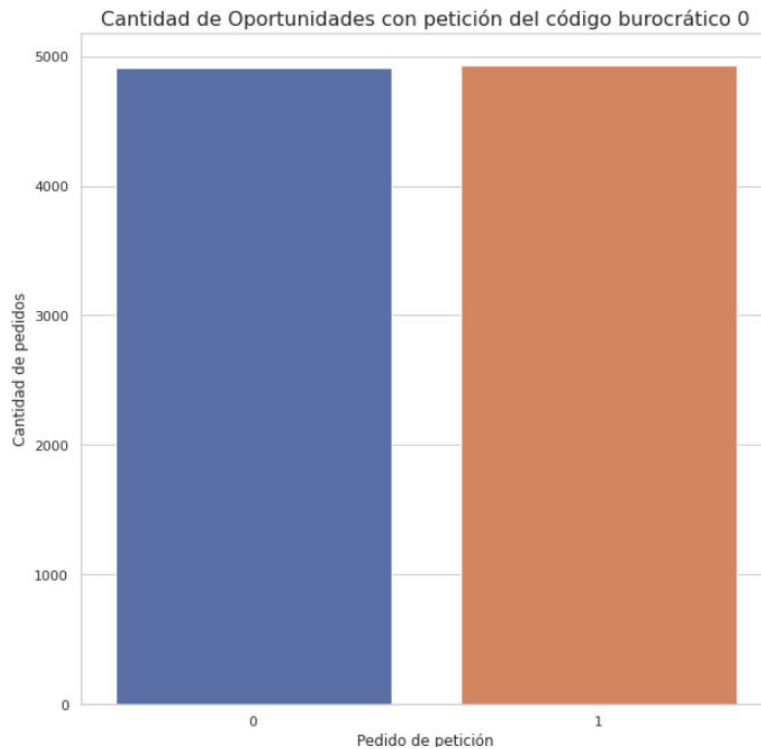
Por último hay 2 columnas curiosas que son **Submitted_for_Approval** y **Bureaucratic_Code**. Realizando una investigación sobre ellas podemos observar que la primera, la “petición para aprobación” no nos da ninguna información, ya que todos los registros tienen un *False*, no aporta nada. En cuanto a la columna del código burocrático tampoco podemos rescatar mucho, ya que el código va desde el 0 hasta el 6, pero sin importar qué código tiene (solamente 50 registros tienen el código burocrático 0) es independiente si tuvieron petición o no para las columnas de código burocrático 0 o la aprobación de envío y precio, por lo que no se puede obtener una

información relevante con cualquiera de estas 2 columnas que aporte para saber en que resulta la oportunidad.

Como se mencionó anteriormente, y se vuelve a resaltar en este apartado, se muestra el gráfico de oportunidades filtrado por una única operación por oportunidad, ya que será relevante.

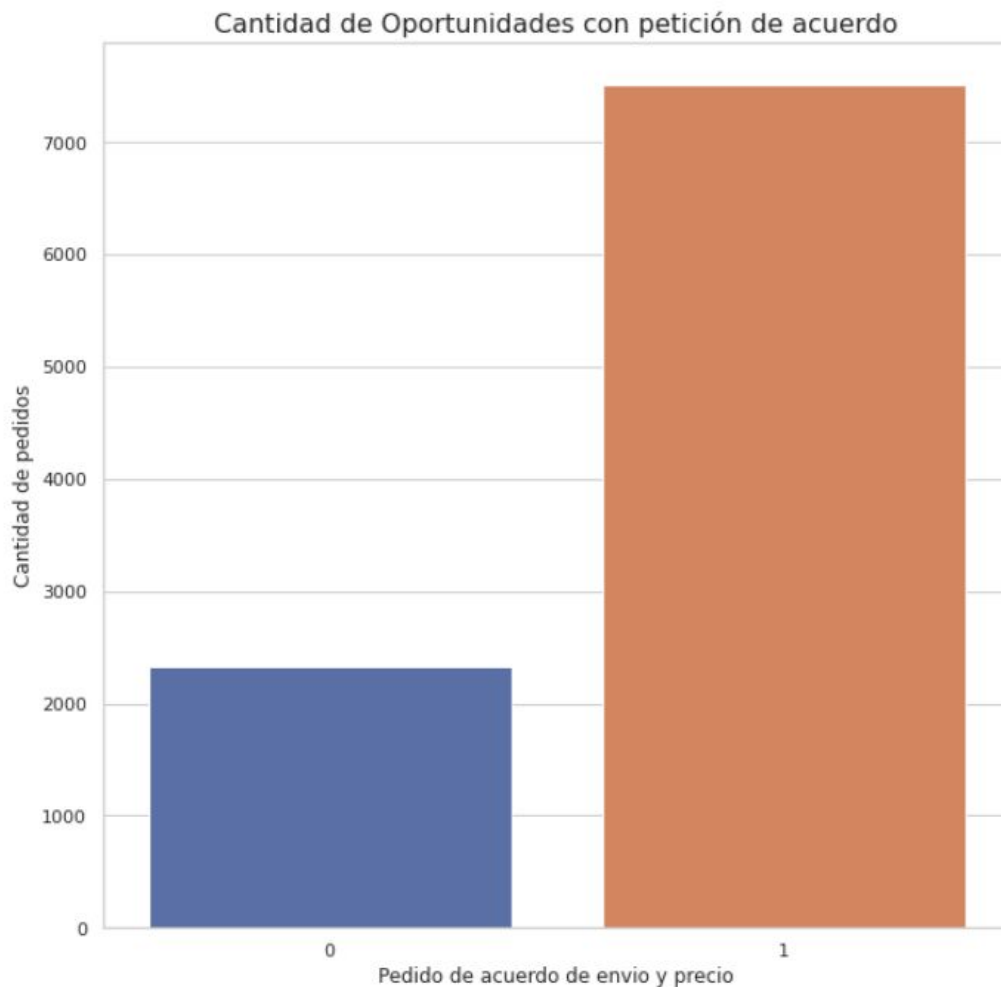


Veamos ahora lo que sucede con las distintas peticiones. Primero veamos el código burocrático 0.



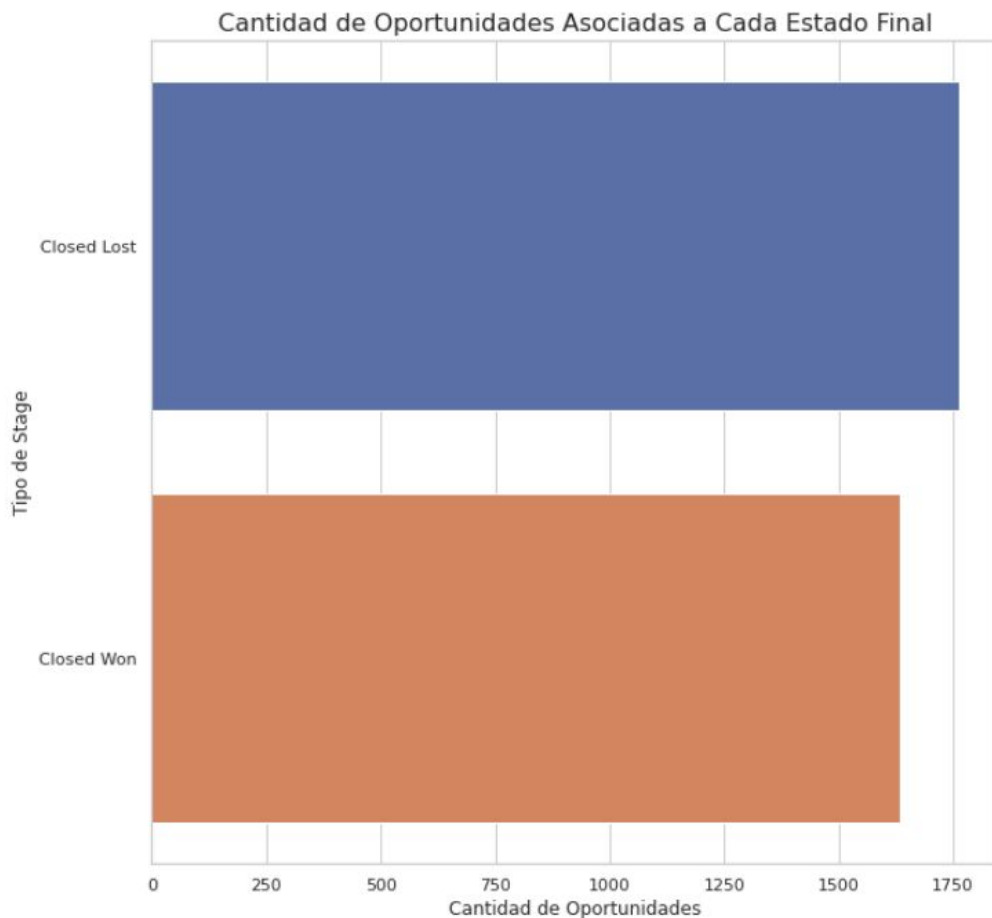
Podemos observar que es prácticamente la misma cantidad que solicitó el código burocrático 0 con respecto a la cantidad que no lo pidió. Siendo específicos, fueron 4933 peticiones contra 4908 que no se pidieron. Teniendo en cuenta que la cantidad de operaciones exitosas es 5072, ya tenemos un primer indicio de que la petición de aprobación no implica que el pedido termine siendo exitoso, y veremos que lo mismo sucede del lado contrario. No es necesario pedir la aprobación para que la operación termine en éxito.

Del mismo modo veamos que sucede con el pedido de términos de envío y precio.



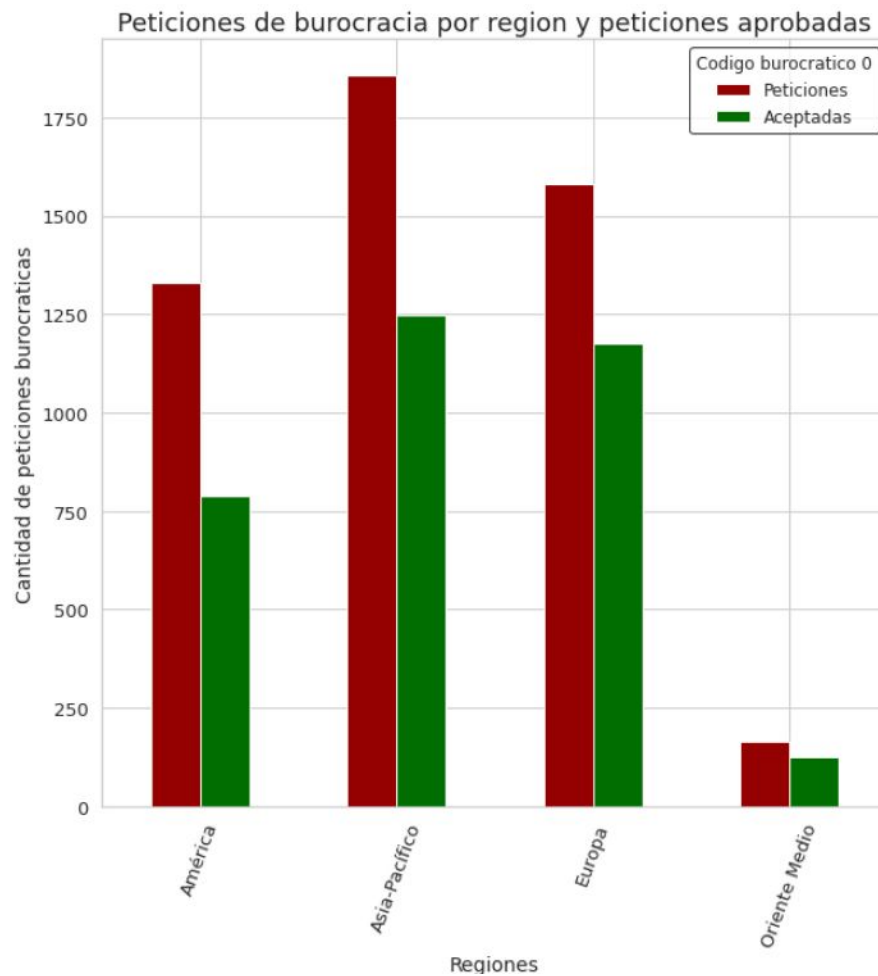
En este caso veamos que si hay una clara tendencia a pedir los términos de envío y precios, por lo que ahora veamos que sucede luego de pedir estas peticiones.

Tomemos ahora una condición de que estas 4 columnas, o atributos, se cumplan. Es decir, se pide la aprobación y esta se aprueba, tanto para los términos como para el código burocrático. Esto resulta en el siguiente gráfico.



Podemos observar que tener estas condiciones aprobadas no implica ya que, en este caso, se redujo considerablemente la cantidad de oportunidades. De las 3395 oportunidades, hay una mayoría que finalizó en fracaso, en particular 1762, contra 1633 oportunidades que terminaron en éxito. Esto nos indica que tener las aprobaciones necesarias, así como el acuerdo previo de envío y precio no implica ni da una tendencia a tener en cuenta para reconocer si la operación termina en éxito o no.

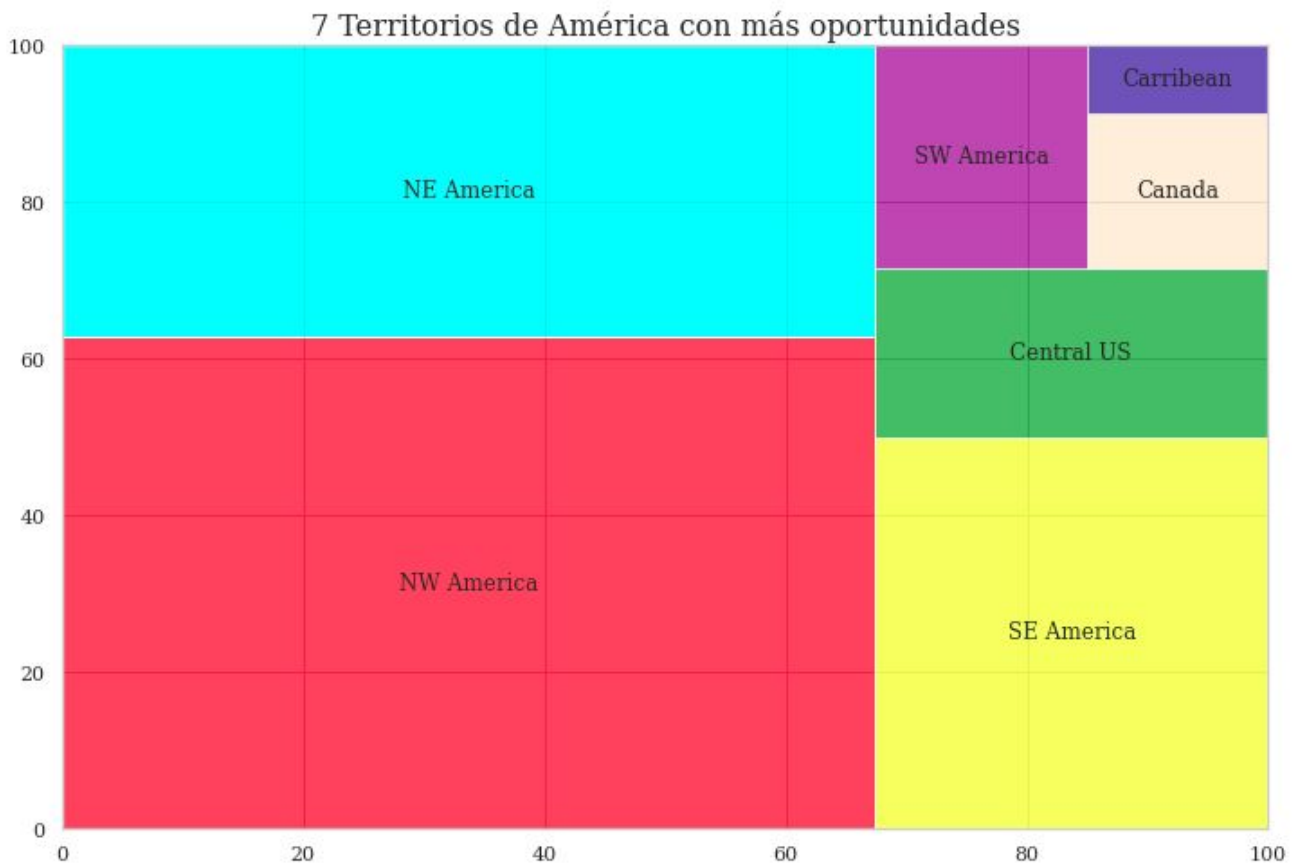
Veamos ahora qué sucede en cada región, ya que puede suceder que en alguna región se requiera el código burocrático 0 con mayor obligación que en otra, y veamos también, para cada una de ellas, cuántas peticiones fueron aprobadas.



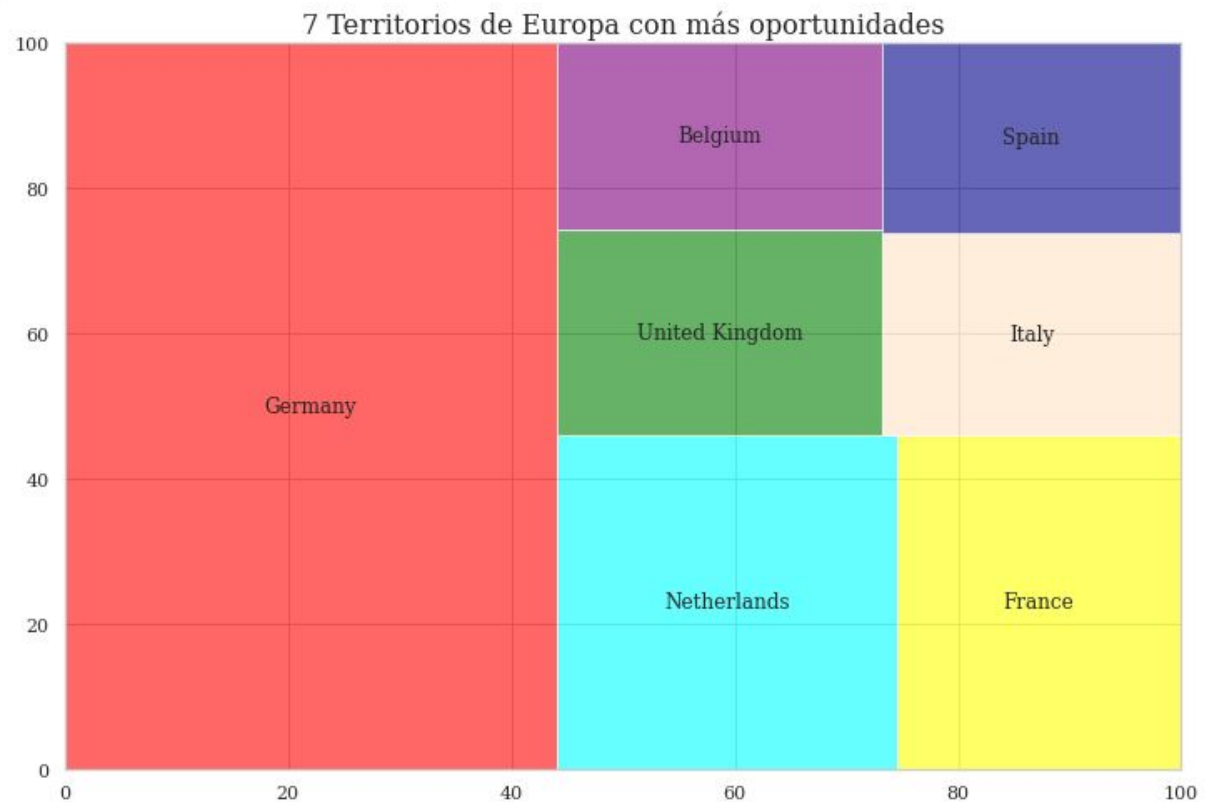
Podemos observar que el caudal de oportunidades va acorde a lo ya mencionado anteriormente en cuanto a la cantidad, donde la región de Asia-Pacífico predomina, y en todas las regiones la relación de petición-aprobación es bastante similar. En el caso de Oriente Medio, aunque sea un caudal menor, se puede ver que la relación petición-aprobación es muy grande, de un 76% para ser exactos.

Territorio

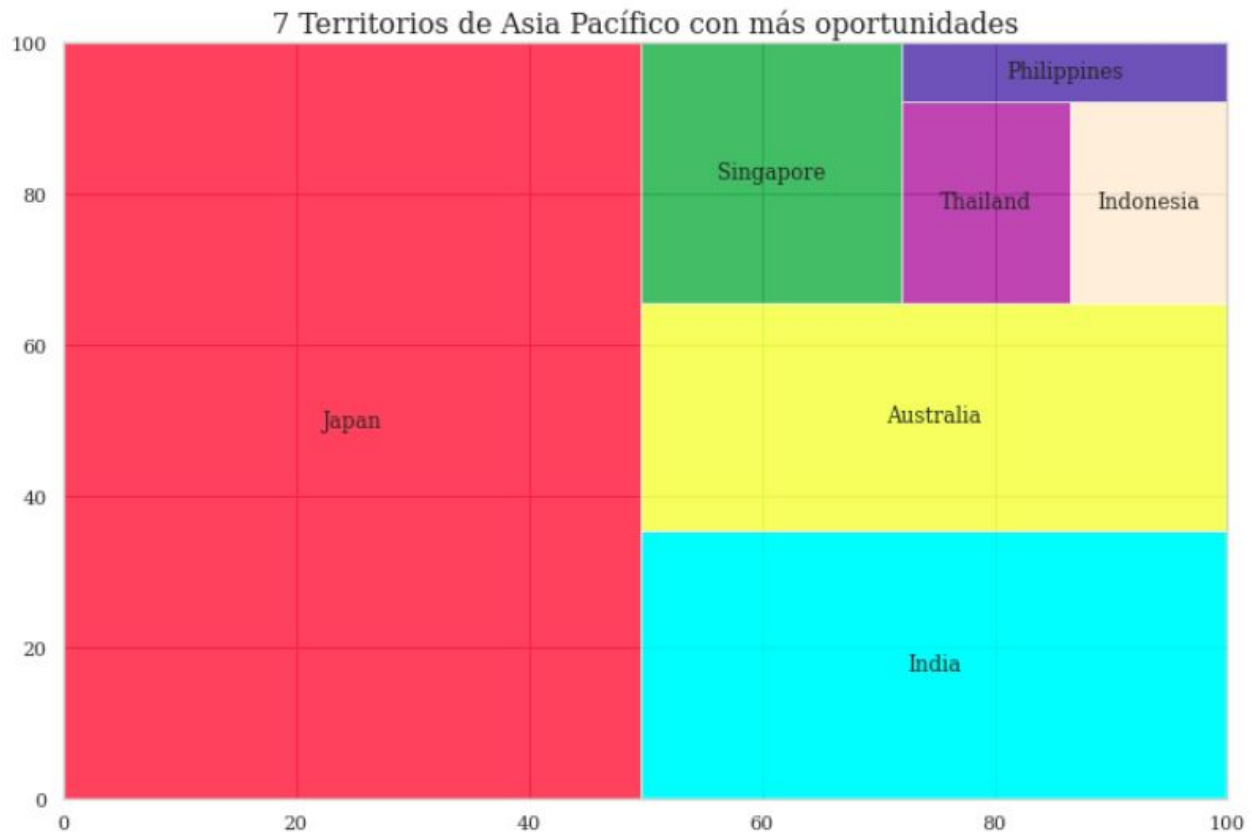
¿Cuáles son los 7 Territorios más populares para cada Región?



Comencemos por la región de América. En este gráfico Treemap notamos a simple vista el caudal de oportunidades realizadas en los 7 territorios más populares de la región. Se puede observar que el mayor porcentaje lo tiene NW America seguido de NE America, entre ambos territorios tenemos concentrado casi el 70% de las oportunidades del top7. También destacamos a SE America que del 30% restante ocupa un 15%.



En Europa podemos destacar a Germany claramente como el más popular abarcando aproximadamente el 45% de las oportunidades. Del 55% restante tenemos que las proporciones de Netherlands y Francia abarcan aproximadamente un 27,5% (la mitad de 55%) y los 4 territorios restantes del top 7 de esta región se reparten el otro 27,5% en similares proporciones.

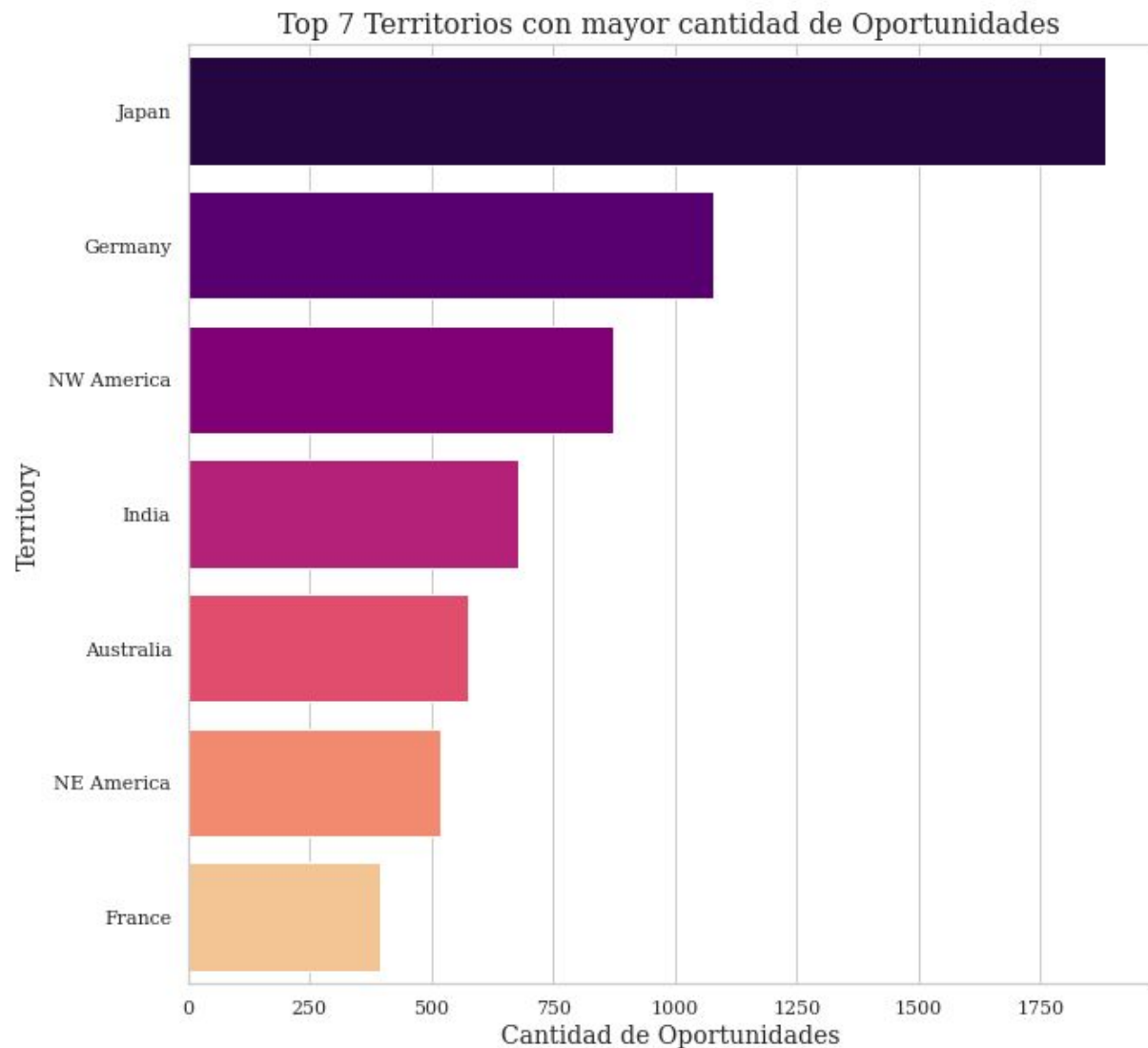


En la región de Asia Pacífico podemos observar que Japón es el claro dominante en cantidad de oportunidades con un porcentaje aproximado del 50%. Del 50% restante, presentan un gran caudal India y Australia ocupando aproximadamente un 33%. El 17% restante se distribuye entre Singapore, Indonesia, Thailand y Philippines.



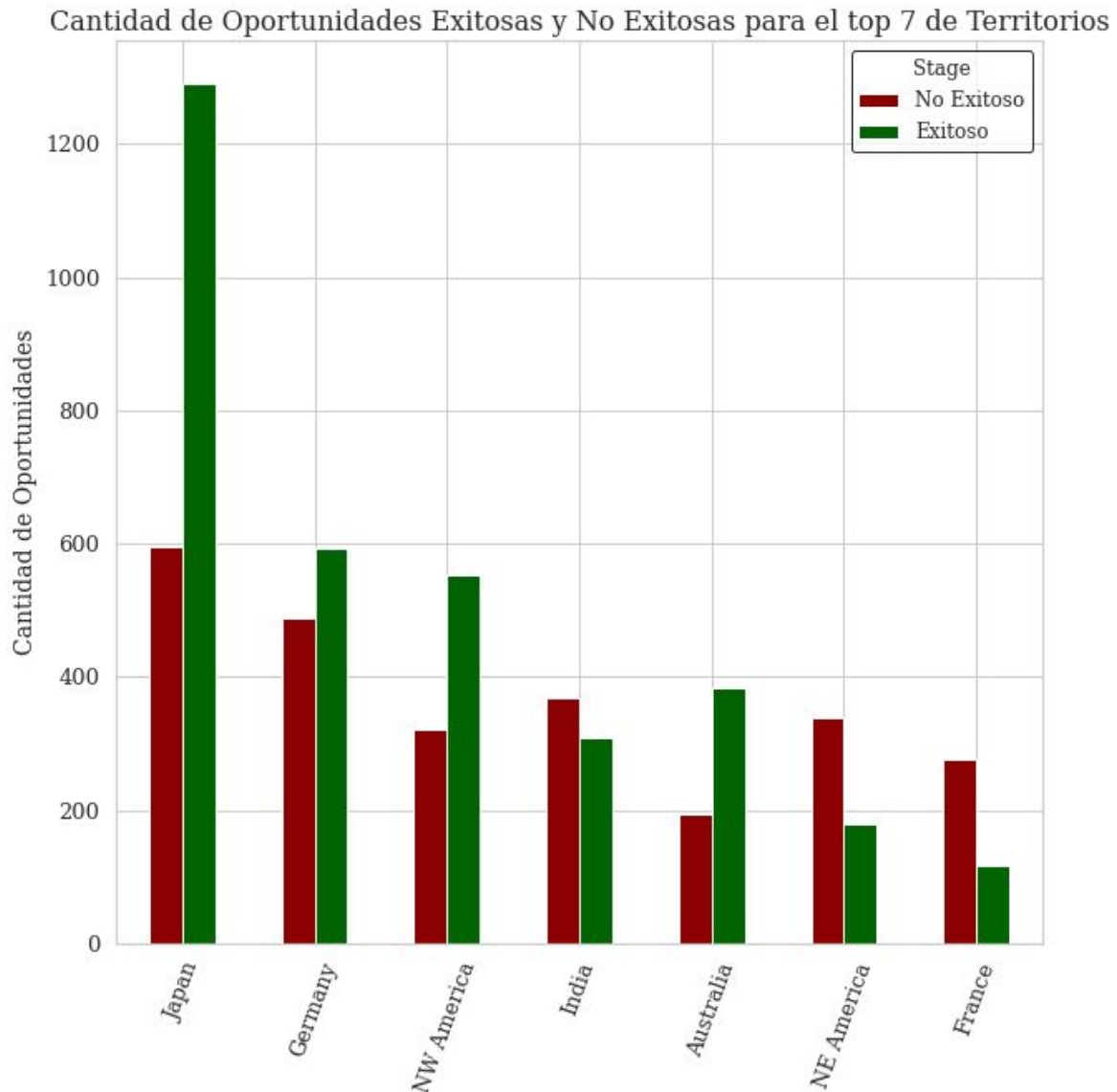
En Oriente medio tenemos a Jordan claramente como el más popular abarcando algo más del 60% de las oportunidades, seguido por Dubai por un 20% aproximadamente, y el tercero más popular siendo Arabia Saudita con un 10% aproximado.

Independientemente de la Región. ¿Cuál es el Top 7 de Territorios más populares y sus correspondientes cantidades de Oportunidades?



Podemos ver que Japón es por lejos el territorio que abarca la mayor cantidad de oportunidades. Esto era de esperar ya que es el territorio más popular de Asia Pacífico que a su vez es la región más popular.

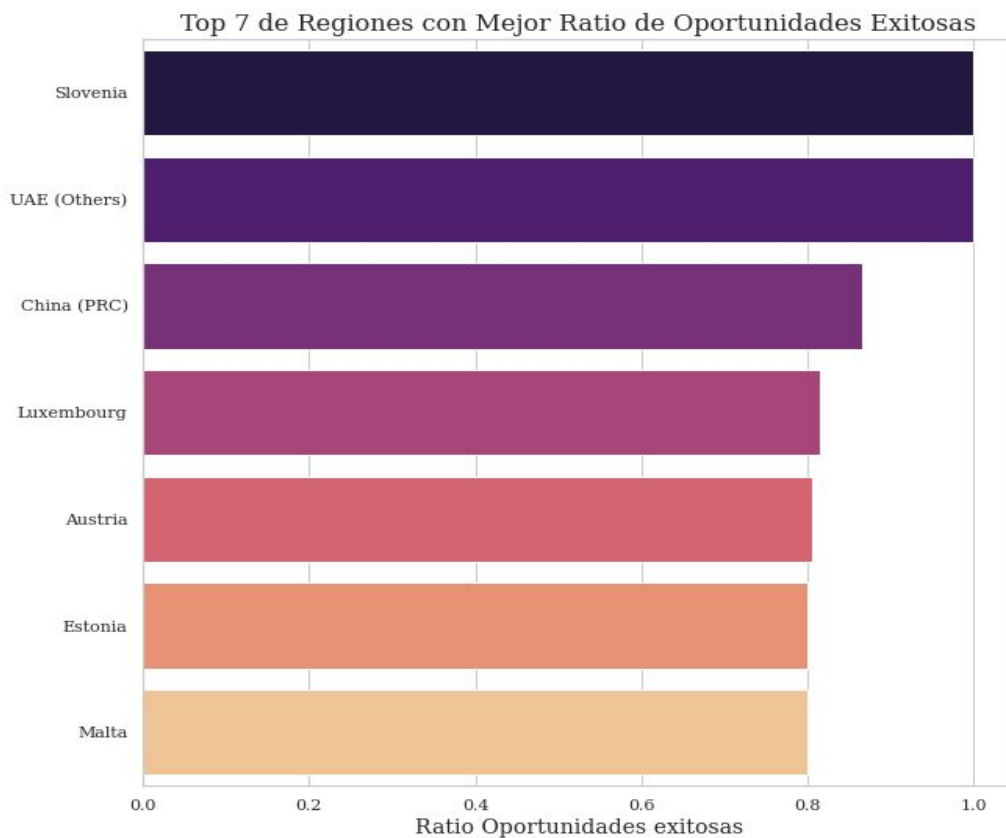
Para los 7 Territorios con mayor cantidad de Oportunidades. ¿Cuántas de ellas son exitosas y no exitosas?



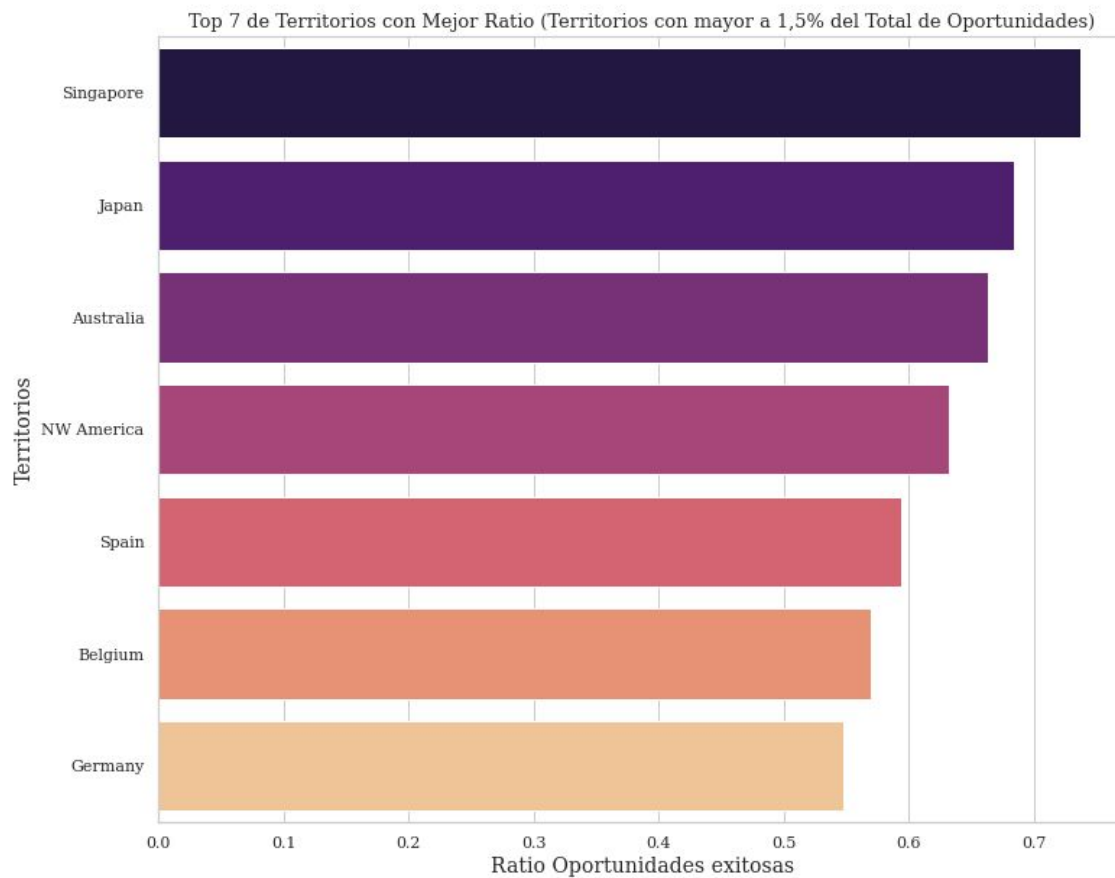
En este gráfico se aprecia que, dentro del Top 7, los tres primeros Territorios con mayor cantidad de oportunidades y Australia presentan una mayor cantidad de casos

Exitosos que no Exitosos. Destacándose Japón con una cantidad de casos Exitosos de más del doble que no Exitosos.

Si prestamos atención un poco más al gráfico nos damos cuenta de que, si quisiéramos listar por ejemplo los dos países (en cantidad de Oportunidades) más exitosos, estos serían: Japón, Germany. Igualmente si quisiéramos listar los dos países menos exitosos (en cantidad de Oportunidades) nuevamente éstos serían: Japón, Germany. Esto pasa porque estamos analizando cantidades absolutas, con lo cual, para poder sacar mejores conclusiones sobre dichos Territorios sería más conveniente realizar un análisis proporcional (relativo) que contemple por ejemplo la cantidad de casos Exitosos sobre el total de casos para cada Territorio y luego sí listar esos 7 Territorios con mejor Ratio Total.



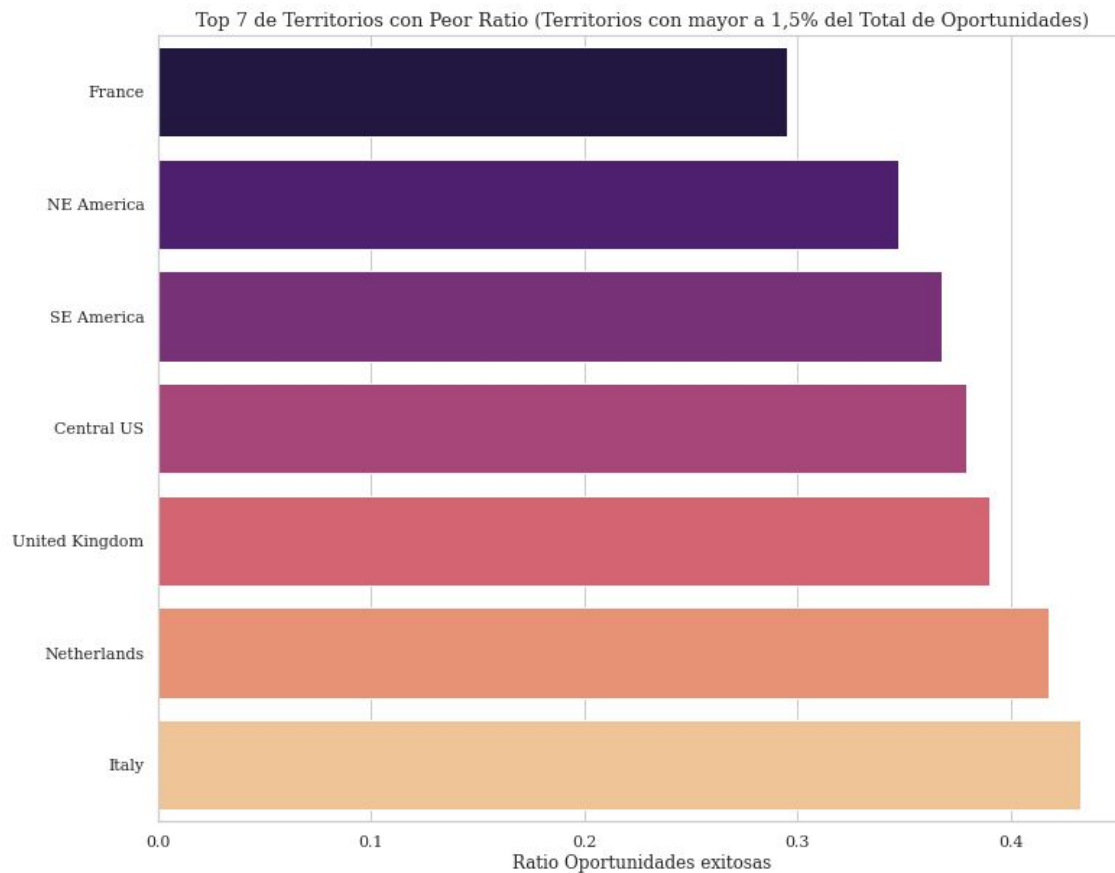
Además para que este análisis no caiga en el problema de la ecuación más peligrosa del mundo, debemos tomar ciertas restricciones, como por ejemplo que en ese Top 7 de mejores Ratios tengamos Territorios con una cantidad de Oportunidades mayor al 1,5% del Total de Oportunidades que presenta la empresa. Con lo cual obtenemos:



Se comprueba que nuestras sospechas de que no solamente tener mejor ratio significa ser el más exitoso sino que debemos tener en cuenta además un mínimo de oportunidades para que nuestras comparaciones sean válidas ya que por ejemplo Slovenia si tuvo 2 oportunidades y si ambas fueron exitosas tiene un porcentaje de éxito del 100%. De esta manera vemos Singapore es el más exitoso con un ratio mayor

a 0,7, seguidos por Japón, Australia y NW América con un ratio mayor a 0,6 y Spain, Belgium y Germany con un ratio mayor a 0,5.

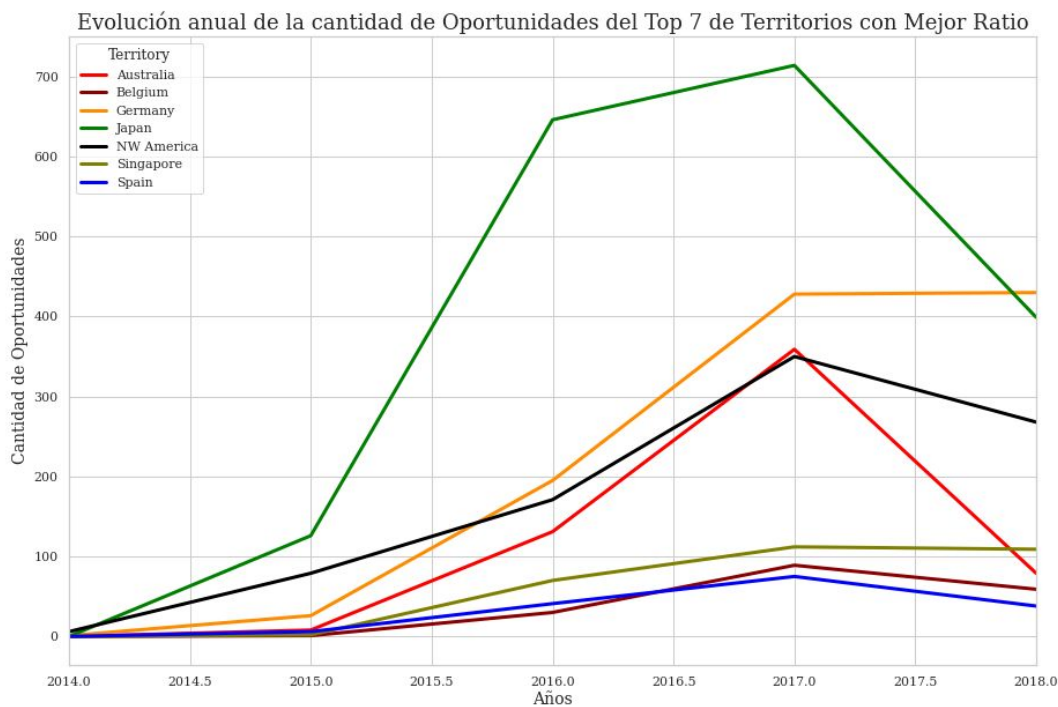
También nos puede interesar los territorios con peor Ratio:



Es interesante ver que tanto Francia como NE America que pertenecen al top 7 de territorios con mayor cantidad de oportunidades también son las que peor ratio de éxito presentan.

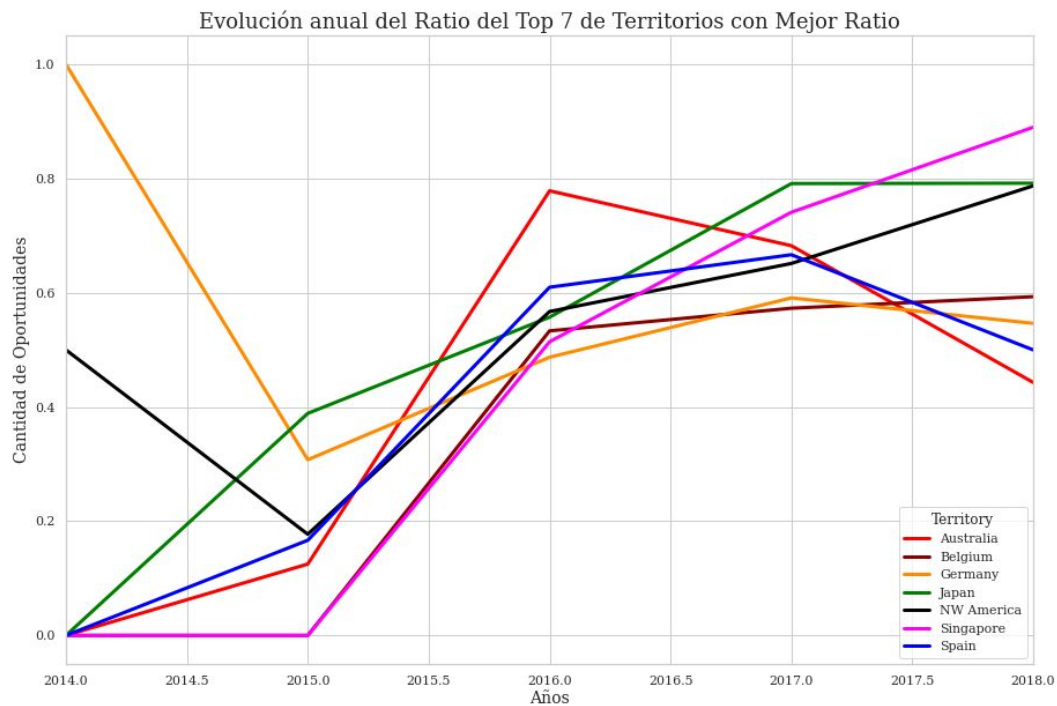
Los 7 Territorios con mejor Ratio ¿Cómo fue su evolución anual?

Comencemos viendo Los 7 Territorios con mejor Ratio y su evolución en cantidad de Oportunidades:



En este gráfico vemos que la mayoría de los territorios con mejor ratio presentan un crecimiento sostenido de 2015 a 2017, destacándose Japón como el de mayor incremento (notar que la evolución temporal de oportunidades de Japón es prácticamente la misma que la analizada anteriormente en el apartado regiones para Asia Pacífico, mostrando la preponderancia de este territorio en dicha región). Luego de 2017 a 2018 se aprecia una caída en cantidad de oportunidades en casi todos los territorios salvo Singapore y Germany que presentan un amesetamiento. También algo interesante es ver como poco antes del 2018 Germany para a ser el territorio con mayor cantidad de oportunidades.

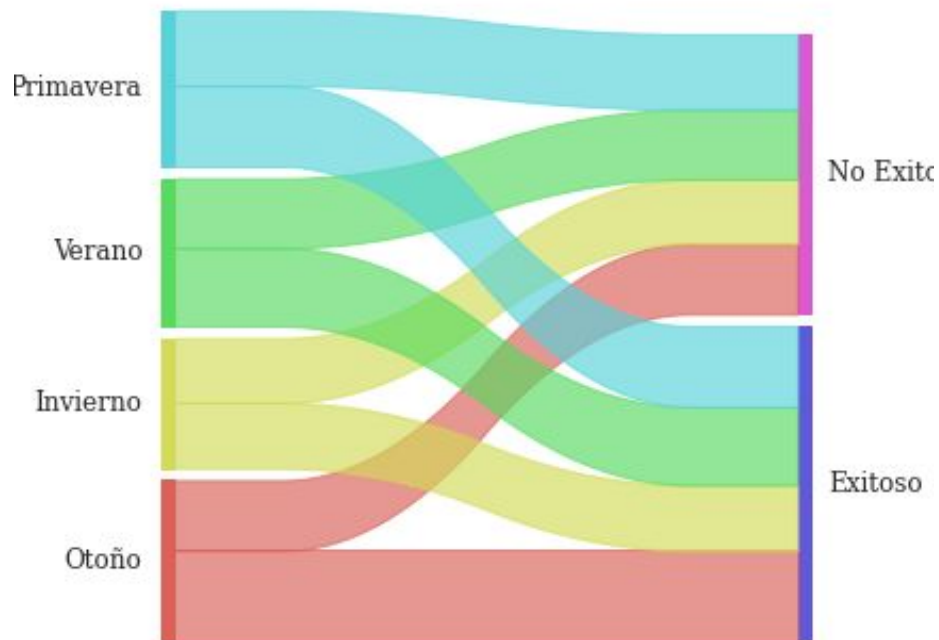
¿Cómo evolucionó el Ratio de estos 7 territorios con mejor ratio?

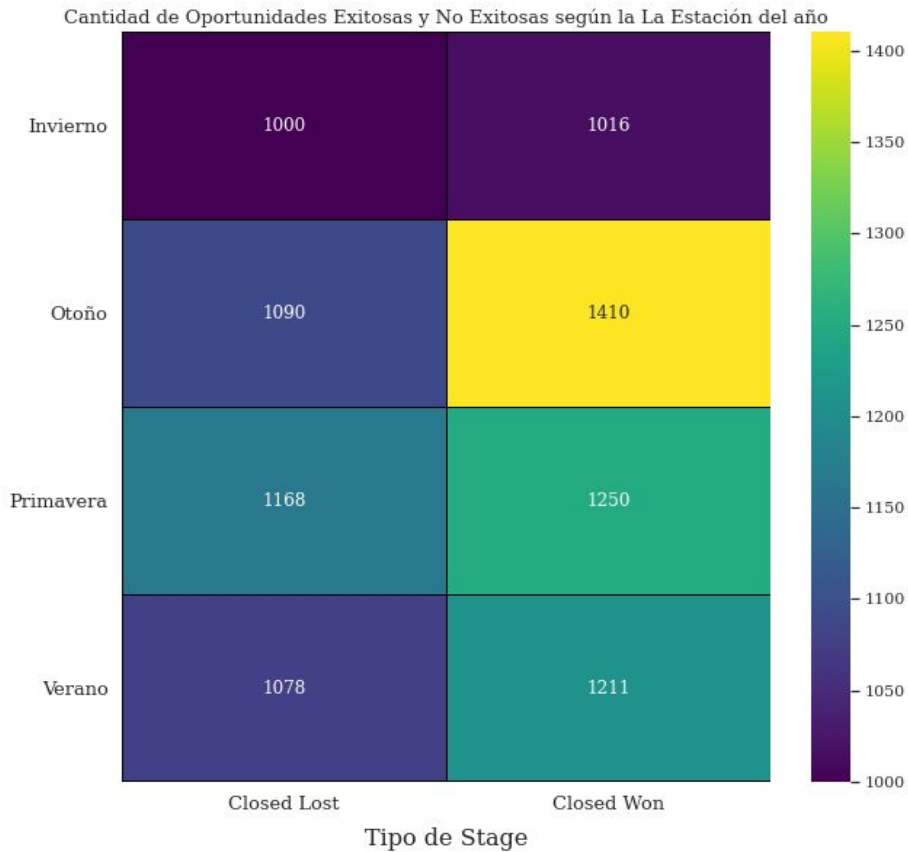


Notar cómo Singapore antes de la mitad del 2017 pasa a ser el territorio con mejor ratio superando a Japón que de 2017 a 2018 muestra un amesetamiento en su ratio y que también está próximo a ser superado por NW América que muestra una tendencia favorable en su ratio desde 2017.

Estaciones del año

Diagrama de flujo de Oportunidades en relación a la Estación del año

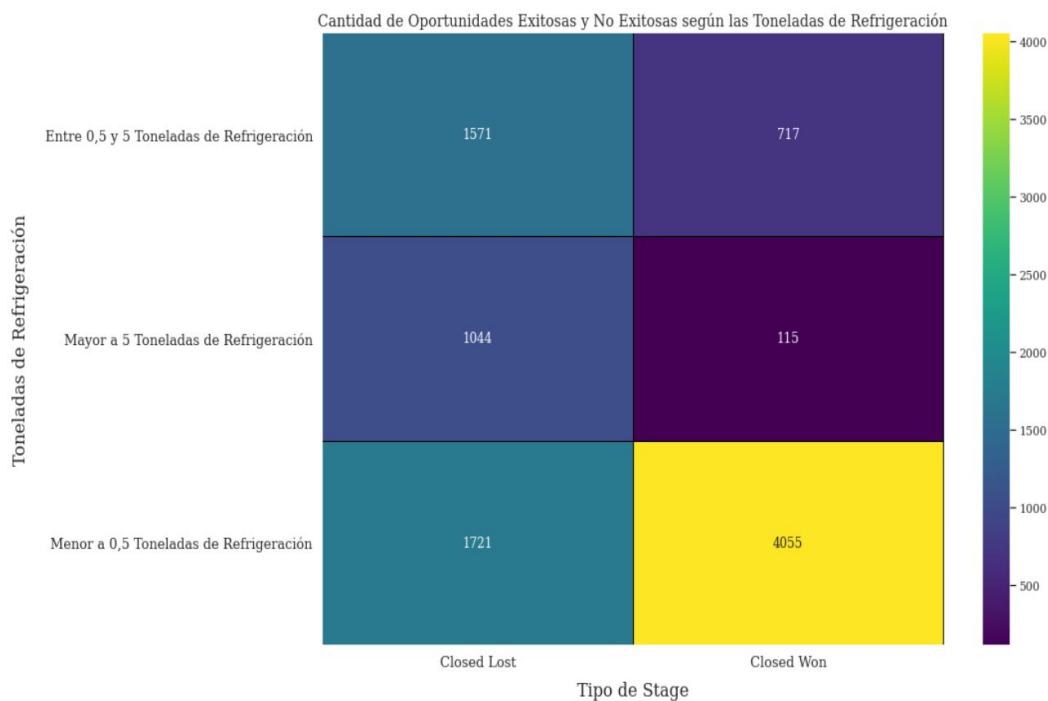
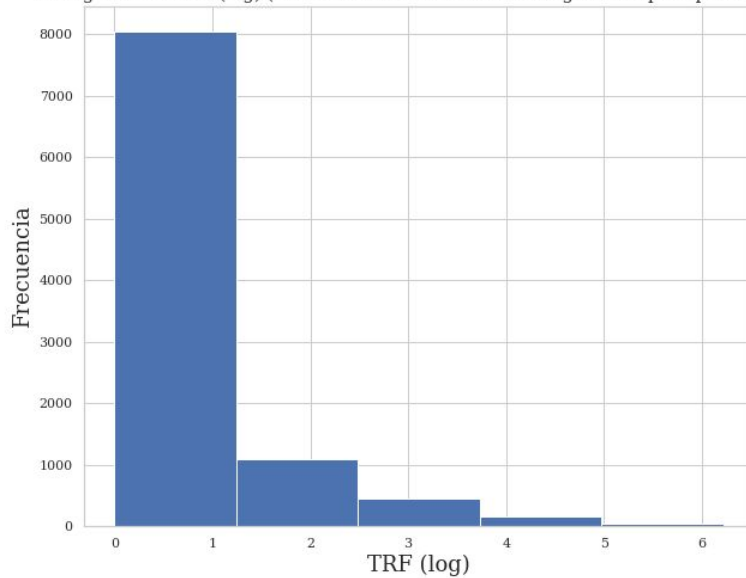




Respecto a las estaciones del año vemos que las estaciones Otoño y Primavera se destacan como las que mayor cantidad de oportunidades presentan, siendo Otoño la más popular y mostrando además ser la que presenta mayores oportunidades exitosas con un 57% frente a un 53% de Verano, 51% en Primavera y un 50% en invierno.

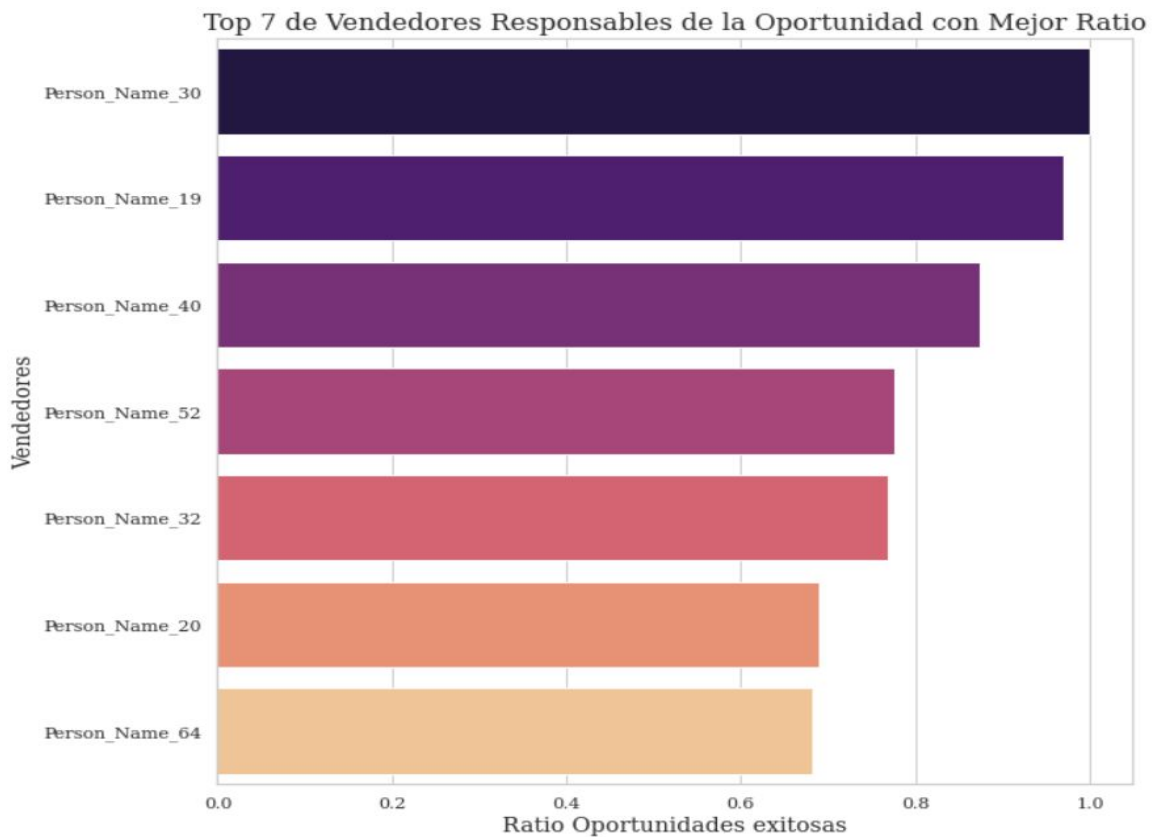
Toneladas de Refrigeración

Histograma de TRF (log) (Cantidad de Toneladas de Refrigeración por oportunidad)

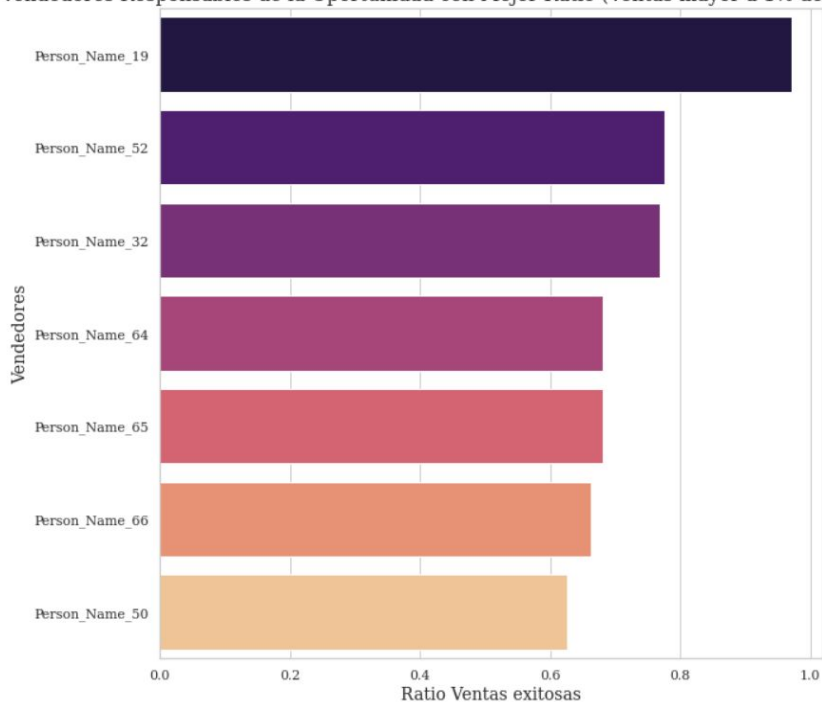


Se observa que las oportunidades se dan mayormente para valores menores a 0,5 toneladas de refrigeración. El éxito de las oportunidades va disminuyendo conforme aumentan las toneladas de refrigeración.

Vendedores Responsables de la Oportunidad



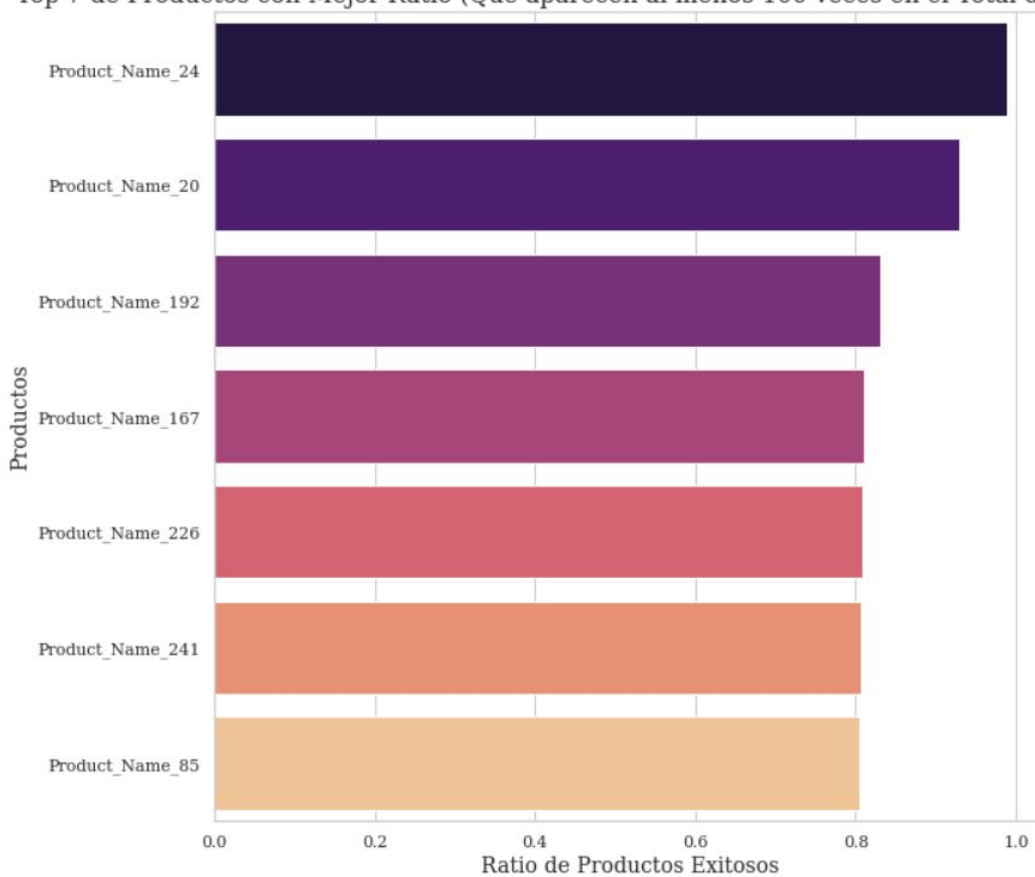
Top 7 de Vendedores Responsables de la Oportunidad con Mejor Ratio (Ventas mayor a 1% del Total de Oportunidades)



Se aprecia nuevamente que el vendedor con mejor ratio (Person_Name_30) absoluto no es en realidad el mejor tomando ciertas restricciones, es más ni siquiera está en el top 7 de mejores vendedores. Cabe destacar que el 1% tomado como filtro mínimo de ventas representa unas 100 ventas. Es muy destacable el ratio que presenta el vendedor Person_Name_19 con un ratio altísimo superior al 0,9.

Productos

Top 7 de Productos con Mejor Ratio (Que aparecen al menos 100 veces en el Total de oportunidades)



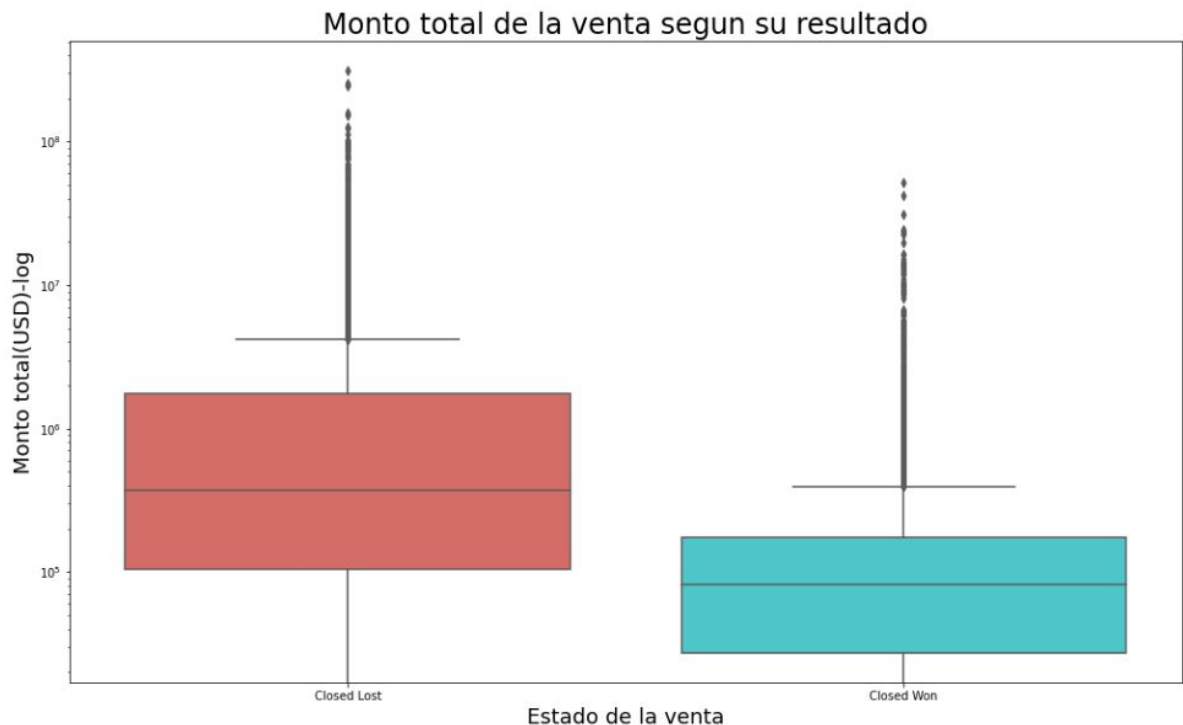
En cuanto a los productos podemos destacar al Product_Name_24 con un ratio cercano a 1, Product_Name_20 con ratio mayor a 0,9. Los 5 restantes muestran un interesante ratio mayor o igual a 0,8.

Monto Total

Para analizar el monto total de cada oportunidad de venta se realizó la sumatoria de todos los montos totales asociados a un mismo Opportunity-ID, luego se realizó una conversión de moneda a USD utilizando el ASP y el ASP(converted), por lo tanto todos los gráficos siguientes se realizaron comparando montos en igual moneda.

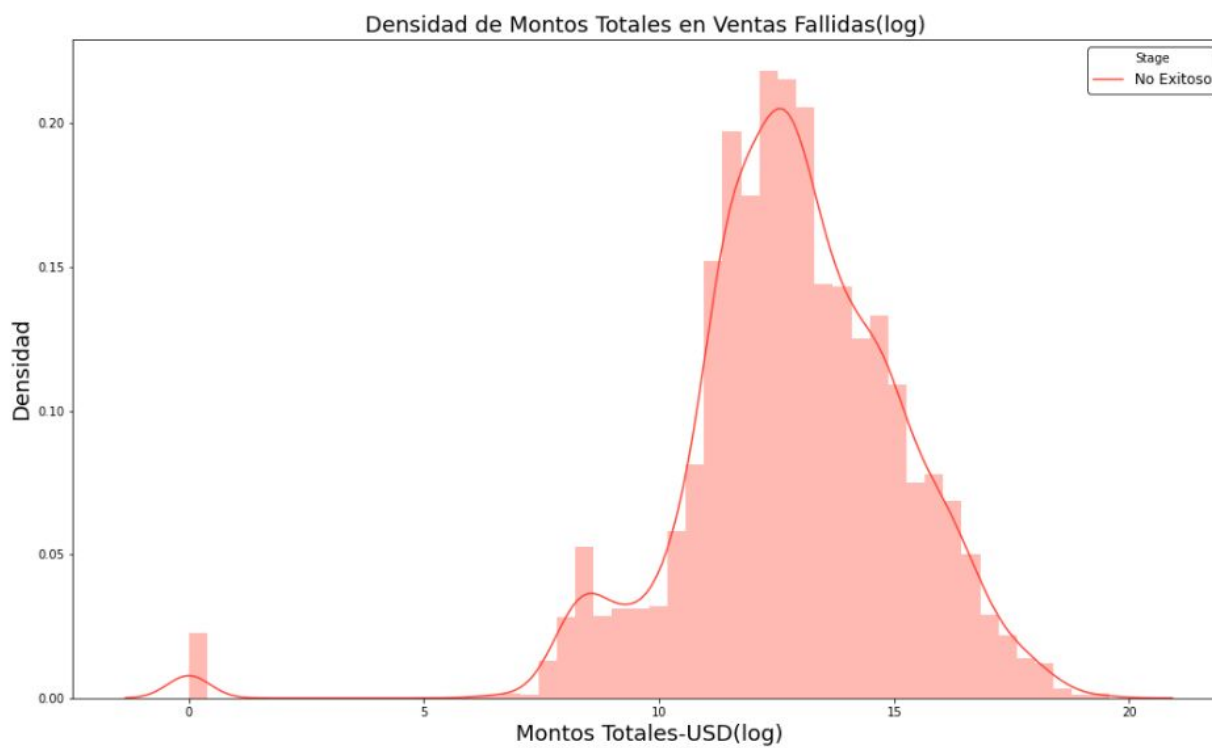
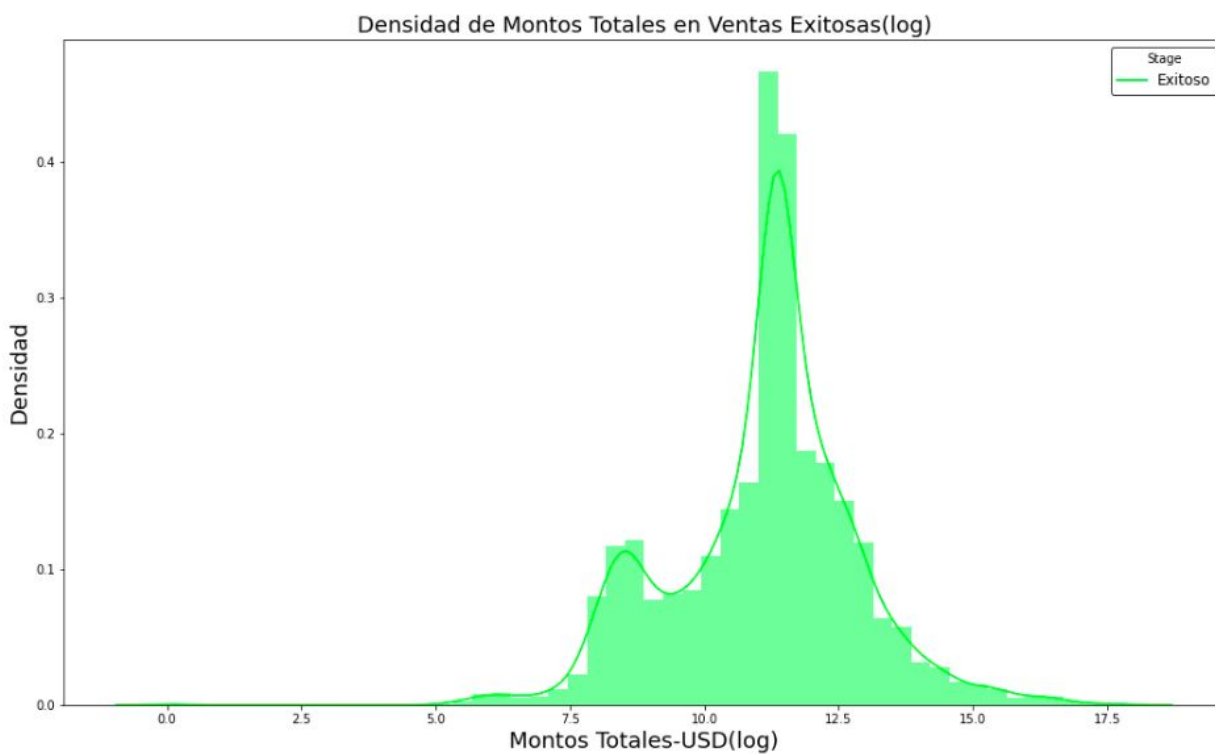
¿Cuál es la relación entre el monto total de una oportunidad y su resultado ?

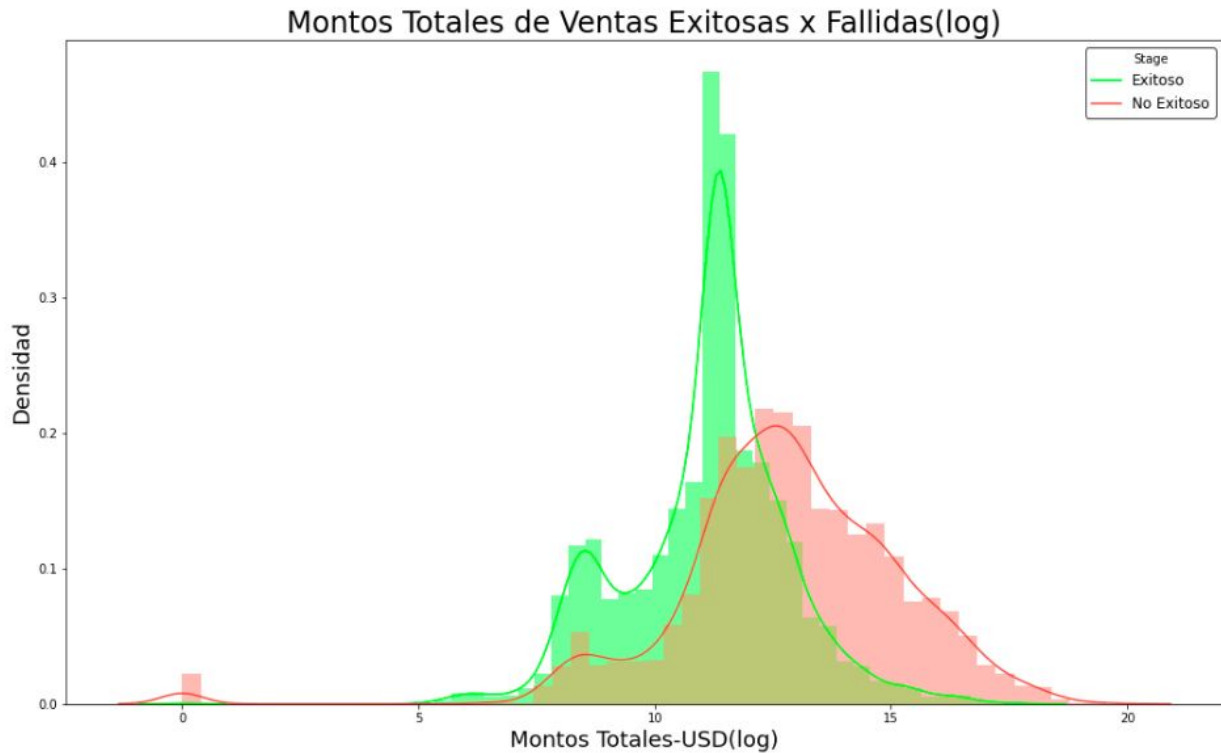
El siguiente gráfico permite visualizar de forma rápida como las ventas exitosas tienden a ser de menor valor que las ventas fallidas. Por otro lado, la distribución de casos fallidos según el monto total es mucho mayor mientras que las exitosas son más acotadas.



¿Cual es la distribución de los montos totales de las oportunidades de venta ?

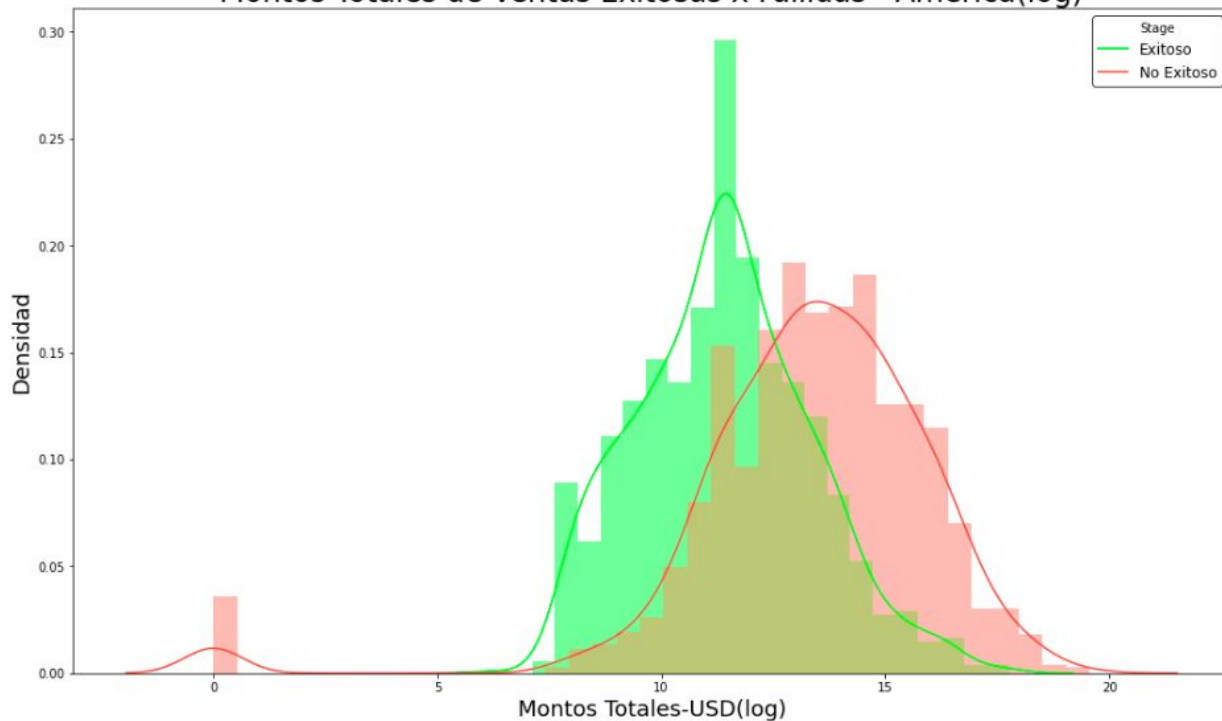
Los siguientes gráficos representan la densidad de los casos exitosos y fallidos según el monto total en escala logarítmica, primero para todas las oportunidades en total y luego dividido por regiones.



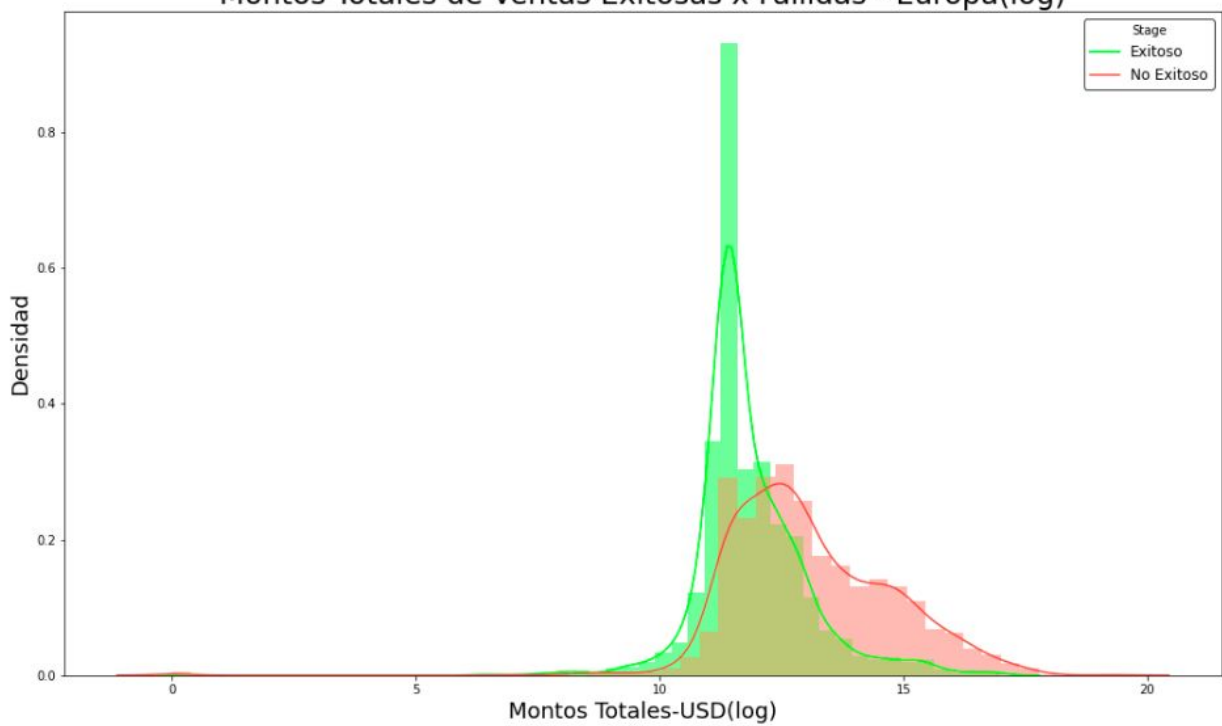


Gracias a este gráfico se puede ver que la mayoría de los casos exitosos se encuentran entre 50.000 y 200.000 USD, hasta este pico predominan las ventas exitosas mientras que a partir de 300.000USD las ventas exitosas decaen rápidamente y a partir de los 500.000USD las ventas fallidas superan ampliamente a las exitosas. De esto se deduce que, cuando se superan los 200.000USD, cuanto mayor sea el monto de la venta la probabilidad de que sea exitosa disminuye rápidamente.

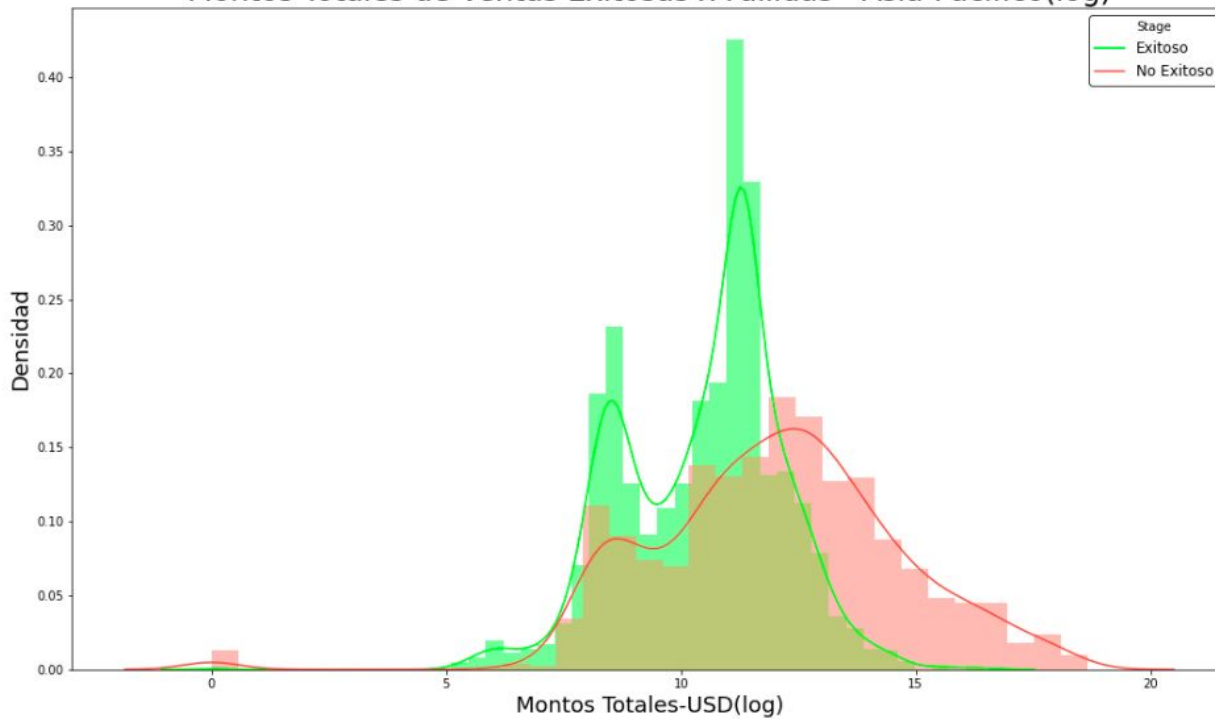
Montos Totales de Ventas Exitosas x Fallidas - América(log)



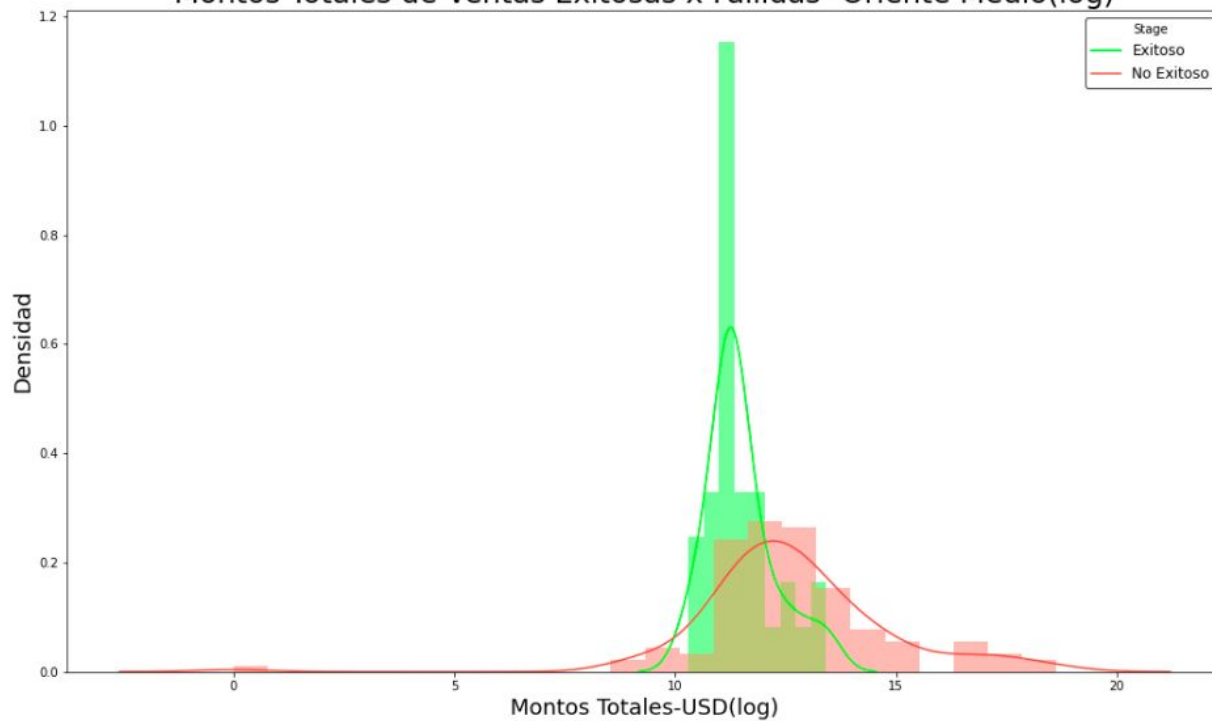
Montos Totales de Ventas Exitosas x Fallidas - Europa(log)



Montos Totales de Ventas Exitosas x Fallidas - Asia-Pacífico(log)



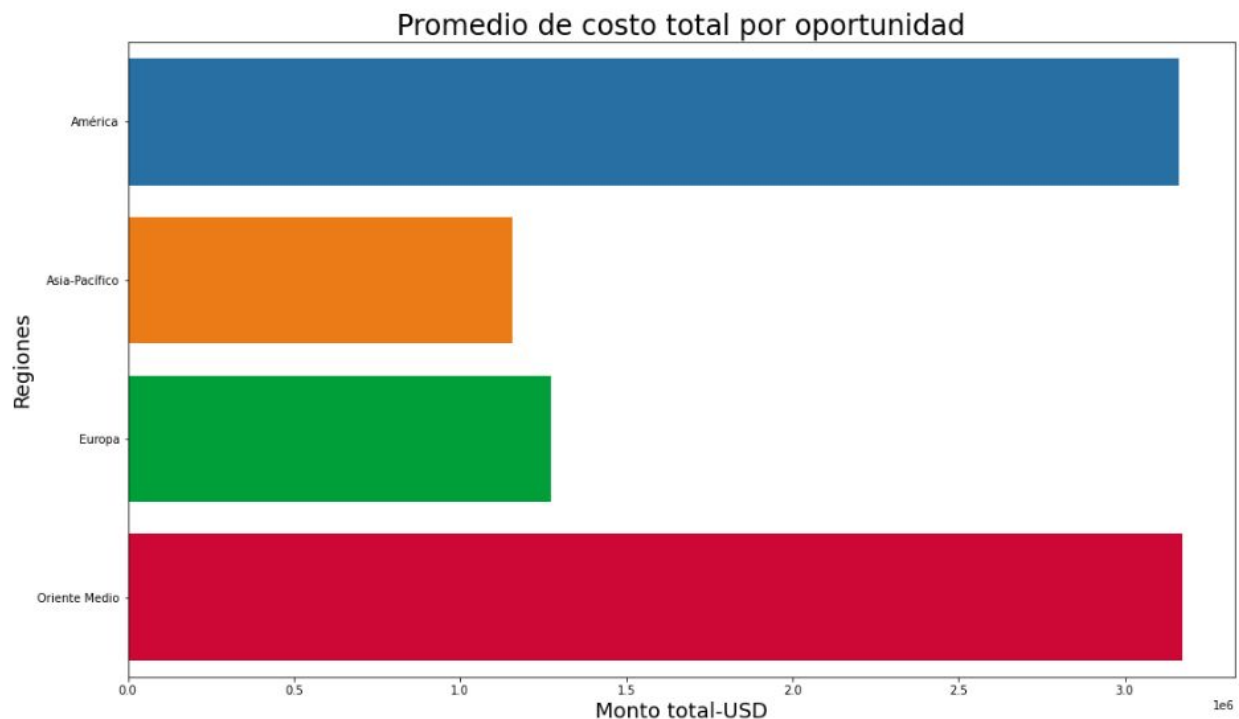
Montos Totales de Ventas Exitosas x Fallidas -Oriente Medio(log)

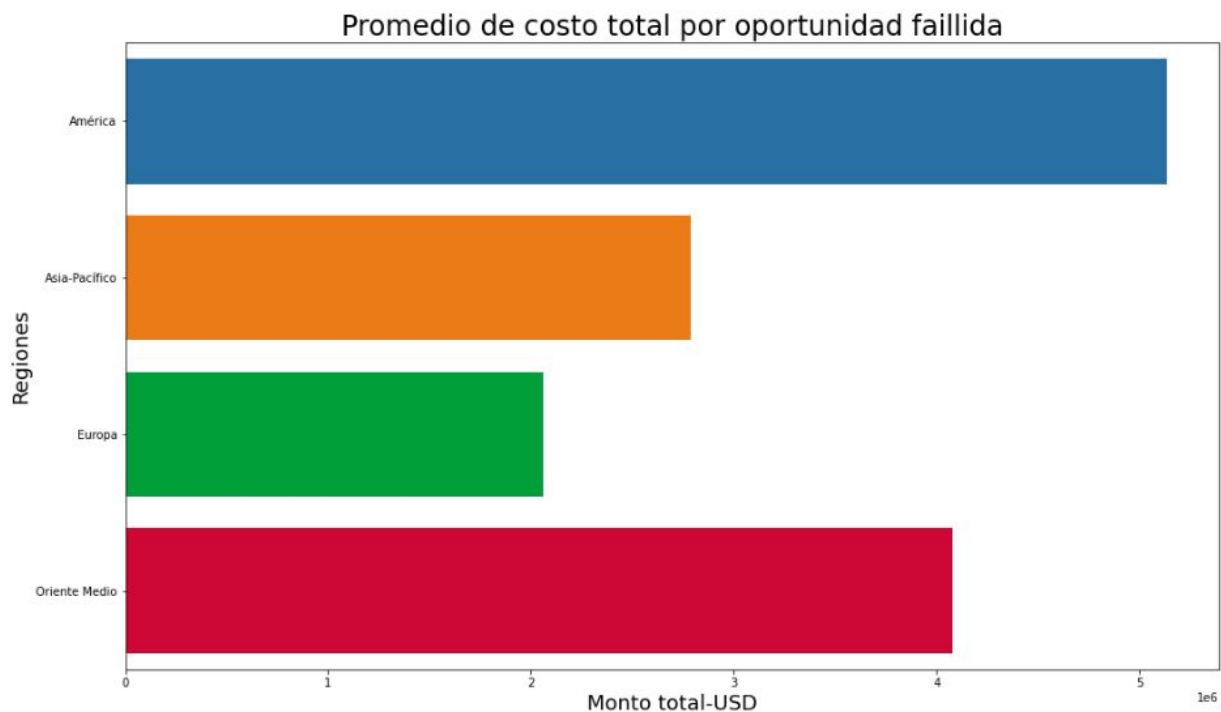
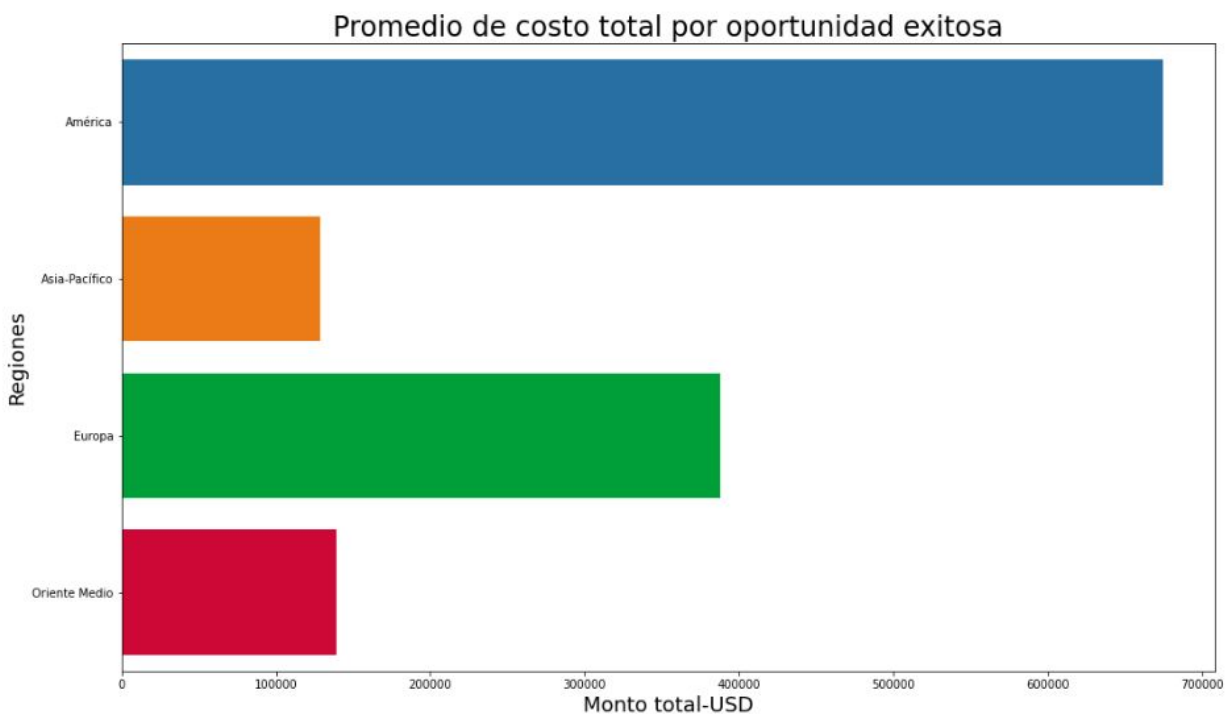


Para todas las regiones se mantiene el pico de densidad de casos y el predominio de los casos fallidos después de este. Se observa que en general para cada región la cantidad de ventas exitosas y fallidas son similares, presentado este corrimiento en base al monto total, resaltando que en América es donde se produce una mayor diferencia de casos exitosos para ventas menores a 300.000USD. Otra observación para resaltar es el doble pico que tiene la región de Asia-Pacífico donde ocurren picos de ventas exitosas entre 2.000-10.000USD Y 50.000-200.000USD. Y por último se denota que las regiones de América y Asia-Pacífico tienen un rango de monto total para las ventas exitosas más amplio que las regiones de Europa y Oriente Medio.

¿Cual es el promedio de monto total de las oportunidades por región ?

Los siguientes gráficos representan el monto total promedio por oportunidad en función de las Regiones, dividido en todas las oportunidades, las oportunidades exitosas y las oportunidades fallidas.





Se puede observar en los gráficos que en las regiones donde se tienen oportunidades de mayor valor es en América y Oriente Medio, sin embargo el promedio de montos

totales de casos exitosos para Oriente Medio es mucho menor(esto mismo también se ve reflejado en los histogramas anteriores). También se observa que en las regiones donde triunfan más las oportunidades de mayor valor son Europa y América, mientras que en Asia-Pacífico y Oriente Medio el promedio es mucho menor.

Oportunidades

Cada fila del dataframe es un ítem de una oportunidad, por lo tanto se agrupó por Opportunity_Name y se sumó los Total_Amount esperando llegar a Total_Taxable_Amount. A continuación se muestran los resultados de cuando esto se cumple.



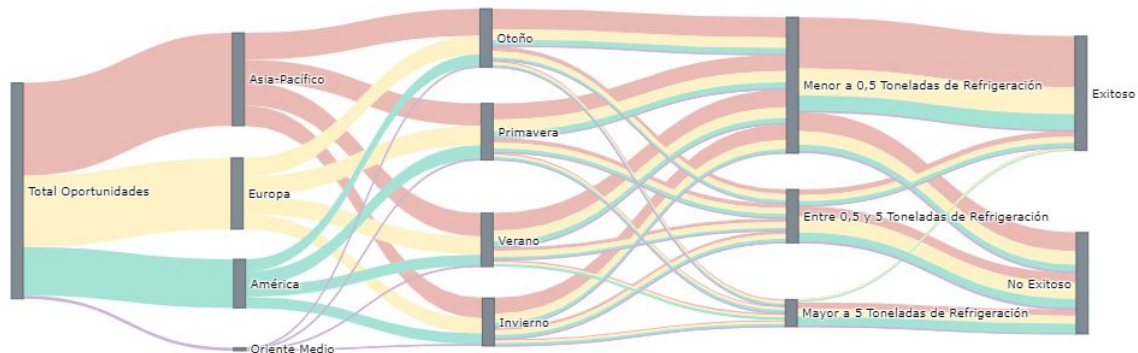
Además para los casos donde esto no se cumple se analizó los casos donde el total_taxable_amount es 0 y se ven los resultados en la siguiente visualización



Combinación de varias variables

¿Cómo se distribuyen los flujos de oportunidades, para cada región, estación del año, Toneladas de Refrigeración y Stage?

Distribución de las Oportunidades Según la Región, La Estación del Año, Toneladas de Refrigeración y Si resultó o No Exitosa



Se observa que las distribuciones de oportunidades son bastante similares para cada estación, destacándose otoño en Asia Pacífico con un mayor flujo. También se aprecia que la mayoría de las regiones optan por oportunidades menores a 0,5 Toneladas de refrigeración y que además esta elección se traduce en una mayor cantidad de casos exitosos que no exitosos, a medida que aumentan las toneladas de refrigeración las oportunidades aumentan su flujo hacia casos no exitosos.

Conclusiones

Hemos hecho un resumen por sector para tener una idea que se puede esperar para los distintos apartados. En base a los distintos parámetros, el objetivo es encontrar una tendencia o alguna razón por la cual una oportunidad puede resultar en éxito o en fracaso.

Tengamos en cuenta algunos puntos interesantes para determinar qué sucedió con el avance de las oportunidades:

- Aprobación burocrática: Como se mencionó anteriormente, teniendo en cuenta que una oportunidad pidió la aprobación burocrática del código 0, se aprobó, se acordó un término de envío y precio y se aprobó, la tendencia lleva a decir que la oportunidad termine en fracaso con una probabilidad del 51.8%.

En cuanto a las Toneladas Refrigeración podemos enunciar lo siguiente:

- Las oportunidades más exitosas se presentan para valores bajos, menores a 0,5 Toneladas. A medida que las oportunidades optan por mayores toneladas descenden los casos exitosos.

En cuanto a las estaciones del año:

- Otoño la más popular y mostrando además ser la que presenta mayores oportunidades exitosas con un 57% frente a un 53% de Verano, 51% en Primavera y un 50% en invierno.

En cuanto a Territorios (tomando como valor mínimo de oportunidades un 1,5%)

- Singapore, Japón y Australia son los que presentan un mayor ratio exitoso.
- Francia, NE América y SE América son los que peor ratio presentan.

En cuanto a los Productos:

- Product_Name_24 y Product_Name_20 son los de mejor ratio para la empresa.

En cuanto a los Vendedores (ventas mayores a 100 oportunidades distintas):

- Se destaca el vendedor Person_Name_19 con un altísimo ratio de éxito superior al 0,9.

Teniendo en cuenta el monto total de las oportunidades con respecto a las regiones y al éxito o fracaso de dicha oportunidad concluimos que:

- La curva de oportunidades fallidas supera a la de oportunidades exitosas a partir de los 200.000USD, esto significa que pasado este punto la probabilidad de que la oportunidad sea exitosa es menor a que resulte fallida, y esta probabilidad decae rápidamente cuanto mayor es el monto total.
- El éxito de los casos está concentrado entre los valores 2.000USD y 200.000USD, teniendo su mayor éxito entre los 50.000USD y 200.000USD.
- Las regiones donde mayor probabilidad de éxito tienen las oportunidades de mayor valor son América y Europa teniendo un promedio de valor de ventas exitosas de 400.000USD y 650.000USD.
- El rango de oportunidades exitosas es mayor en América y Asia-Pacífico, presentando una distribución más variada.
- Oriente Medio y Asia-Pacífico rechazan en promedio oportunidades de valor superior a los 200.000USD que las otras dos regiones.