

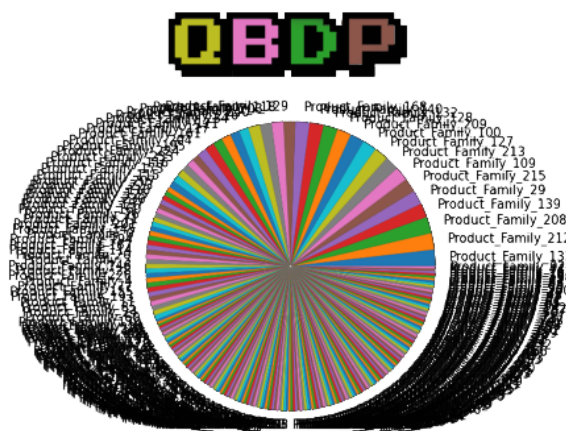
# UBA - Facultad de Ingeniería

Departamento de Computación

Organización de Datos (75.06)

## Trabajo Práctico 1

2do cuatrimestre - 2020



GRUPO: Que Buena Data Papá		
Alumno	Padrón	Mail
Xifro, Juan Bautista	101717	jxifro@fi.uba.ar
Rojas, Mateo	104985	mrojas@fi.uba.ar
Re, Gabriel	105095	gre@fi.uba.ar
Locatelli, Santiago	104107	slocatelli@fi.uba.ar

# Índice

<b>1. Introducción</b>	<b>3</b>
<b>2. Dataset</b>	<b>4</b>
2.1. Manipulación de Datos . . . . .	6
<b>3. Análisis de Años</b>	<b>7</b>
3.1. Porcentaje por años . . . . .	7
3.2. Análisis 2017 y 2018 . . . . .	9
<b>4. Análisis de Región y Territorio</b>	<b>11</b>
4.1. Porcentaje Stages por Región . . . . .	11
4.2. Territorios con más oportunidades . . . . .	13
4.3. Porcentaje Stages por Territorio . . . . .	15
4.4. Territorios con más manejos de montos . . . . .	16
4.5. Porcentaje Stages por Quarters . . . . .	22
<b>5. Análisis de Productos</b>	<b>24</b>
5.1. Tipos de productos especificados . . . . .	25
5.1.1. Distintos Tipos de productos especificados . . . . .	25
5.1.2. Productos más frecuentados . . . . .	27
5.1.3. Caracterización de los productos . . . . .	28
5.2. Tipos de productos no especificados . . . . .	32
5.2.1. Productos más frecuentados . . . . .	34
5.2.2. Caracterización de los productos . . . . .	35
5.3. Conclusión Productos . . . . .	39
<b>6. Análisis de Marcas</b>	<b>40</b>
6.1. Incidencia de Marcas Por Región . . . . .	40
6.2. Oportunidades de Marcas . . . . .	40
6.3. Análisis de Marcas Especificadas . . . . .	42
6.4. Análisis de Marcas No Especificadas . . . . .	42
6.5. Marcas: Conclusión . . . . .	44
<b>7. Análisis de Divisas</b>	<b>45</b>
7.1. Incidencia de Divisas . . . . .	45
7.2. Divisas: Dólar . . . . .	45
7.3. Divisas: Yen . . . . .	46

7.4. Divisas: Euro . . . . .	47
7.5. Divisas: Conclusión . . . . .	47
<b>8. Análisis de Precios Convertidos</b>	<b>48</b>
8.1. Análisis general de precios convertidos . . . . .	48
8.1.1. Análisis de oportunidades de mayor precio . . . . .	48
8.1.2. Análisis de oportunidades de menor precio . . . . .	49
8.2. Precio Universal: Marcas . . . . .	49
8.2.1. Marcas: Mayor Precio Universal . . . . .	50
8.2.2. Marcas: Menor Precio Universal . . . . .	50
8.3. Precio Universal: Productos . . . . .	51
8.3.1. Productos: Mayor Precio Universal . . . . .	51
8.3.2. Productos: Menor Precio Universal . . . . .	52
8.4. Precio Universal: Conclusión . . . . .	53
<b>9. Análisis de Permisos</b>	<b>54</b>
9.1. Permisos . . . . .	54
9.1.1. Aprobación de Envío . . . . .	54
9.1.2. Aprobación Burocrática . . . . .	56
9.2. Permisos de las Oportunidades por Región . . . . .	58
9.3. Permisos: Juntando todo . . . . .	58
9.4. Permisos: Conclusión . . . . .	59
<b>10. Conclusión</b>	<b>61</b>
<b>11. Github</b>	<b>62</b>

## 1. Introducción

Se propone analizar en el presente informe los datos brindados por el dataset, el cual tiene datos sobre la empresa "Frío Frío", dicha empresa es B2B ("Business To Business") y está destinada a la venta e instalación de equipos de aire acondicionado para grandes superficies. A continuación se expone la documentación de la resolución del trabajo práctico de la materia Organización de Datos. El trabajo tiene como objetivo la exploración, revisión y análisis de los datos dados por la cátedra para poder predecir (con cierto grado de error) la probabilidad de éxito de cada oportunidad comercial, y así mejorar el rendimiento y optimizar el esfuerzo de los vendedores.

Dentro de la documentación se encontrarán distintas conclusiones a las que se llegan luego de varias exploraciones sobre el dataset y cómo se llegaron a las mismas. Nos concentramos específicamente en los **Stages** ya que estos son el indicador del estado de cada oportunidad y el análisis lo hacemos sobre las Regiones, Territorios, Permisos, Divisas, Montos, Productos, Categorías y Marcas.

El lenguaje que se utilizó fue Python, acompañado por distintas bibliotecas como: Pandas, Numpy, Seaborn, Matplotlib, WordCloud.

Antes de comenzar vamos a hablar un poco del Dataset en el cual vamos a trabajar.

## 2. Dataset

Primero y principal, lo que hicimos fue analizar como están conformados nuestros datos. De esta forma tenemos un conocimiento general sobre lo que vamos a trabajar. El dataset a analizar pesa **662.1 KB** . Cuenta con **16947 Filas** y **52 Columnas**.

Cada fila representa una oportunidad, cada oportunidad consiste en un proyecto de venta o instalación de equipos para un cliente. Y cada columna tiene distinta información sobre el data set, las cuales son:

- **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 (Categorica).
- **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 (Categorica).
- **Total\_Taxable\_Amount:** monto gravado total (Decimal).
- **Stage:** variable target. Estado de la oportunidad (Categorica).
- **Prod\_Category\_A:** categoría 'A' del producto (Categorica).
- **TRF:** Toneladas de refrigeración (Entero). Es una unidad de potencia.

## 2.1. Manipulación de Datos

En esta sección comentamos las modificaciones que aplicamos sobre el dataset para poder trabajar sobre el mismo de forma más cómoda y eficiente.

- Los territorios que estaban como **None** y tenían como región Japón, los catalogamos como **Japón**.
- Las variables que estaban como **None**, fueron modificadas a conveniencia, siendo renombradas como **No Especificado** o **Sin Especificar**.
- Para la variable **Total Amount**, decidimos hacer la **conversión a dólares** y así tener un precio universal.

Dicho esto podemos comenzar con el análisis.

### 3. Análisis de Años

#### 3.1. Porcentaje por años

Vamos a arrancar analizando las oportunidades ganadas por año.

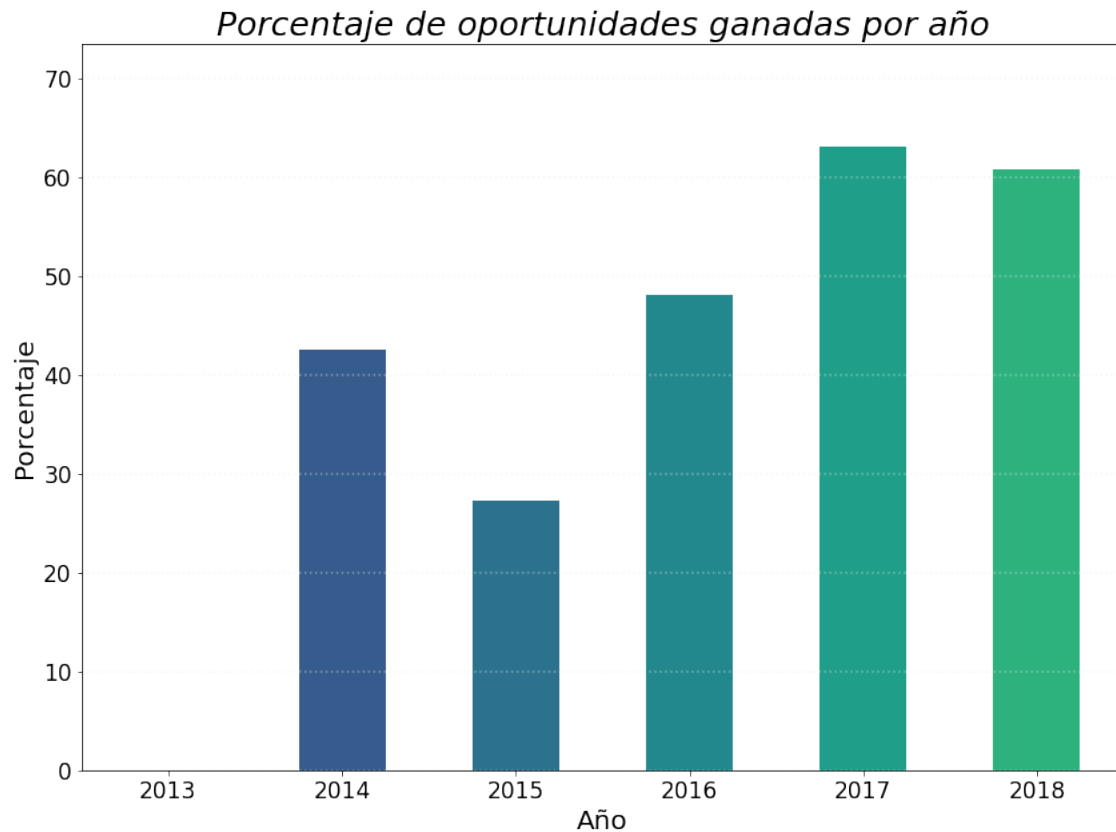


Figura 1

El gráfico nos muestra como el año que tiene un porcentaje más alto de oportunidades ganadas es el año 2017, y como en el 2013 resulta que hay un 0 % de porcentaje, esto en si resulta bastante extraño, por lo que vamos a utilizar otro método para analizar.





Figura 2

Teniendo ambos gráficos podemos asegurar una mayor fidelidad de la representación de los datos.

### Conclusión

Nos permitimos concluir que el hecho de que el año 2013 tenga 0 %, se atribuye a que la cantidad de oportunidades totales es muy baja, y estas solamente fueron oportunidades que terminaron en 'Closed Lost'. Por lo que realmente no se puede concluir algo sobre el año 2013. Y para el año 2014 también, en el primero gráfico parecía tener un porcentaje por debajo de la mitad, pero sucede lo mismo que con el año 2013.

Para **los años 2017 y 2018**, el porcentaje de oportunidades ganadas es considerablemente alto, y la cantidad de oportunidades totales también lo es. Esto nos da la seguridad para poder afirmar que estos años **fueron los mejores en cuanto a porcentaje de oportunidades ganadas**.

Pero, ¿Por qué será que en los años 2017 y 2018 les fue tan bien? Vamos a analizarlo.

### 3.2. Análisis 2017 y 2018

Empezamos viendo el monto total perteneciente a las oportunidades que terminaron en Closed Lost vs Closed Won.

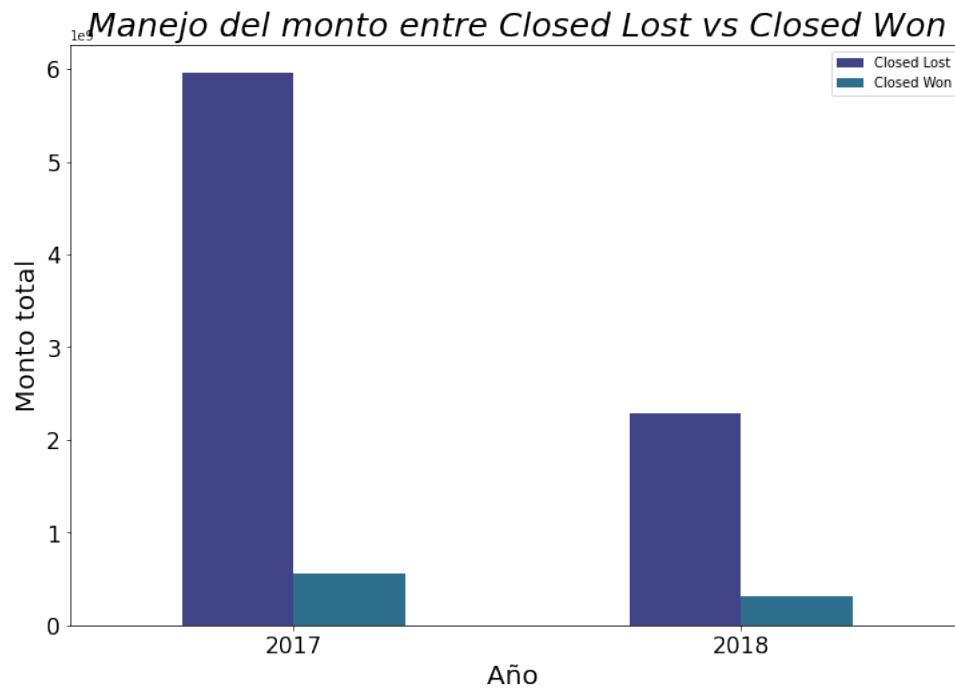


Figura 3

En ambos años vemos como la suma de los montos de las oportunidades tienen una gran diferencia, las que terminaron en 'Closed Lost' suman mucho más que las que terminaron en 'Closed Won'. Pero, también puede ser que este hecho sea así porque la cantidad de oportunidades que terminaron en 'Closed Lost' es mucho mayor, pero teniendo en cuenta el gráfico que compara las densidades (2), vemos que es completamente al revés, es decir, **hay muchas más oportunidades que terminaron en 'Closed Won', pero la suma total de los montos es mucho menor a la suma de las oportunidades 'Closed Lost'**.

Seguimos pensando, ¿Que más puede afectar? Consideraremos a la región como un factor clave, entonces veamos como afecta esta. Mas adelante se realizará un análisis más general.

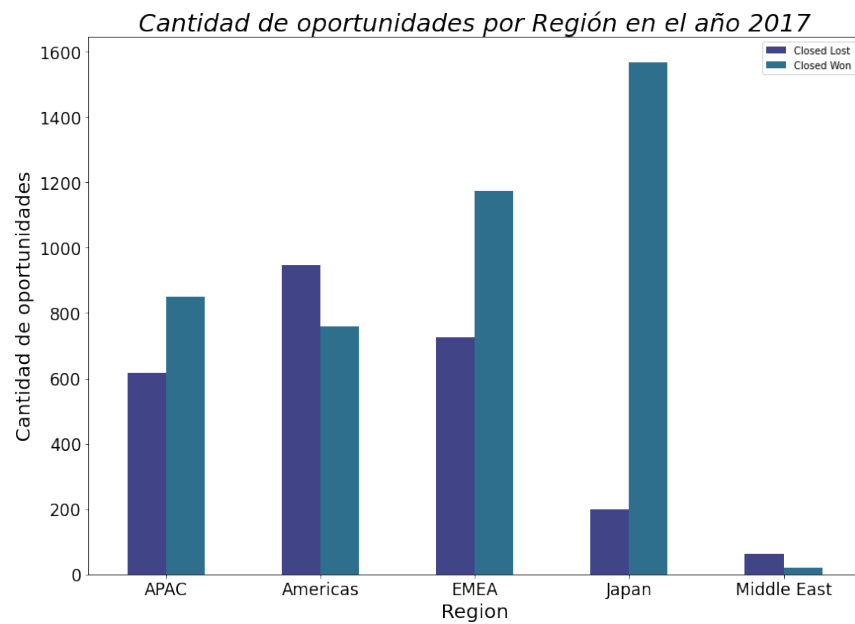


Figura 4

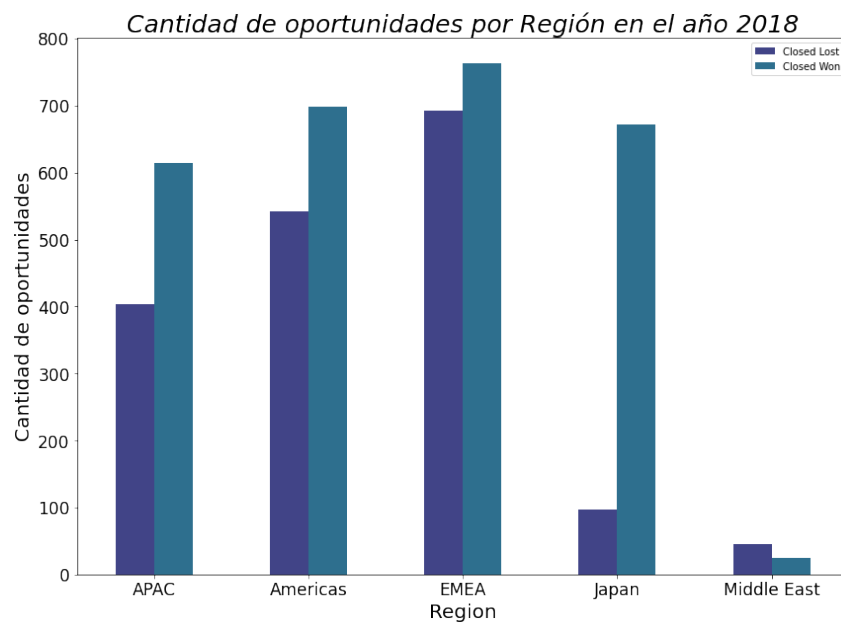


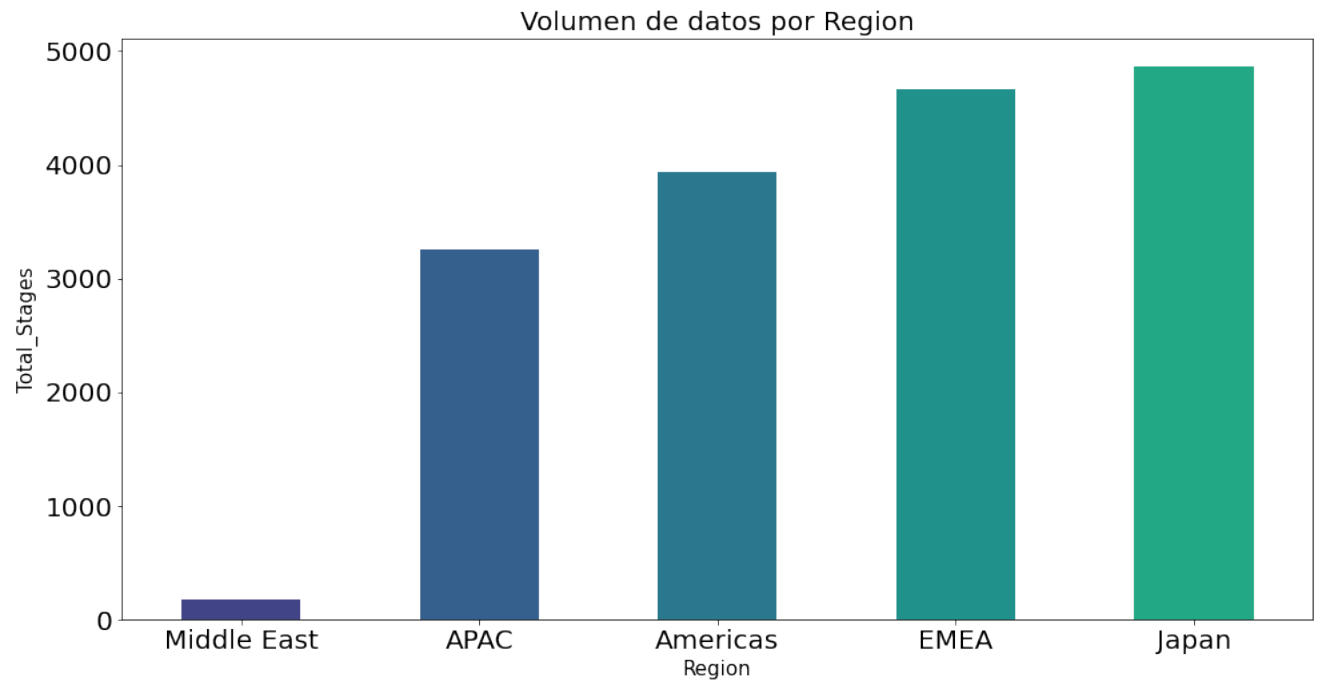
Figura 5

Viendo ambos gráficos, llegamos a la conclusión de que es **la región de Japón** quien realmente genera tal éxito en estos años. Por otro lado, notamos que la región de **Medio Oriente** es todo lo contrario a Japón.

## 4. Análisis de Región y Territorio

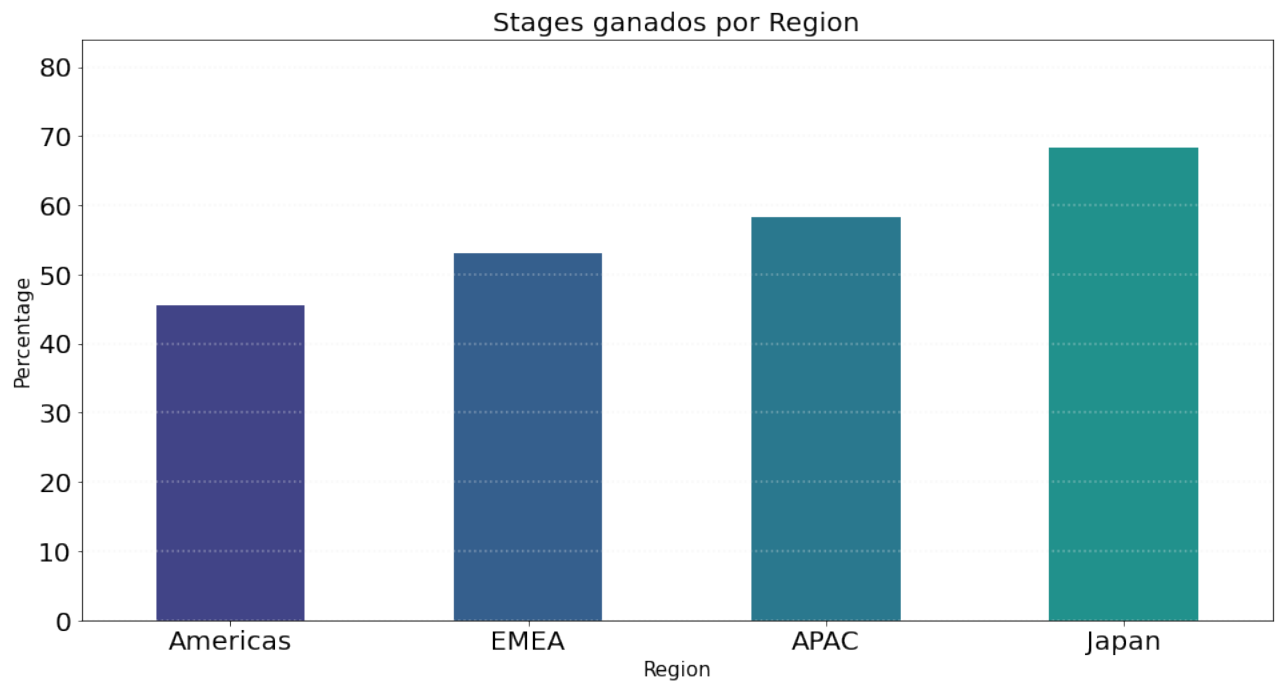
### 4.1. Porcentaje Stages por Región

El análisis comienza mirando el volumen de datos que hay respecto del *Total Stages* por Región.



Como podemos ver Middle East tiene muy poca cantidad, por lo tanto para el siguiente análisis no lo tomaremos en cuenta.

Continuamos analizando el porcentaje de las oportunidades cerradas exitosamente (Closed Won, dentro de la columna Stage) que hay por región. Este porcentaje esta hecho en base de la cantidad de *Closed Won* + *Closed Lost*, las demás Stages como *Proposal*, *Negotiation* y *Qualification* fueron descartadas de este análisis y no entran en el total de Stages, por la poca cantidad de datos en comparación a los anteriores.



Podemos ver que el porcentaje en **Américas** es menor del 50 % y que **Japón** es la región con mejor porcentaje y mayor *Total Stages*. También vemos que **APAC** siendo el que menos *Total Stages* tiene, es la segunda región con mejor porcentaje.

## 4.2. Territorios con más oportunidades

En esta sección vamos a comenzar a analizar las oportunidades. Empezamos analizando cuales son los 10 territorios con más oportunidades.

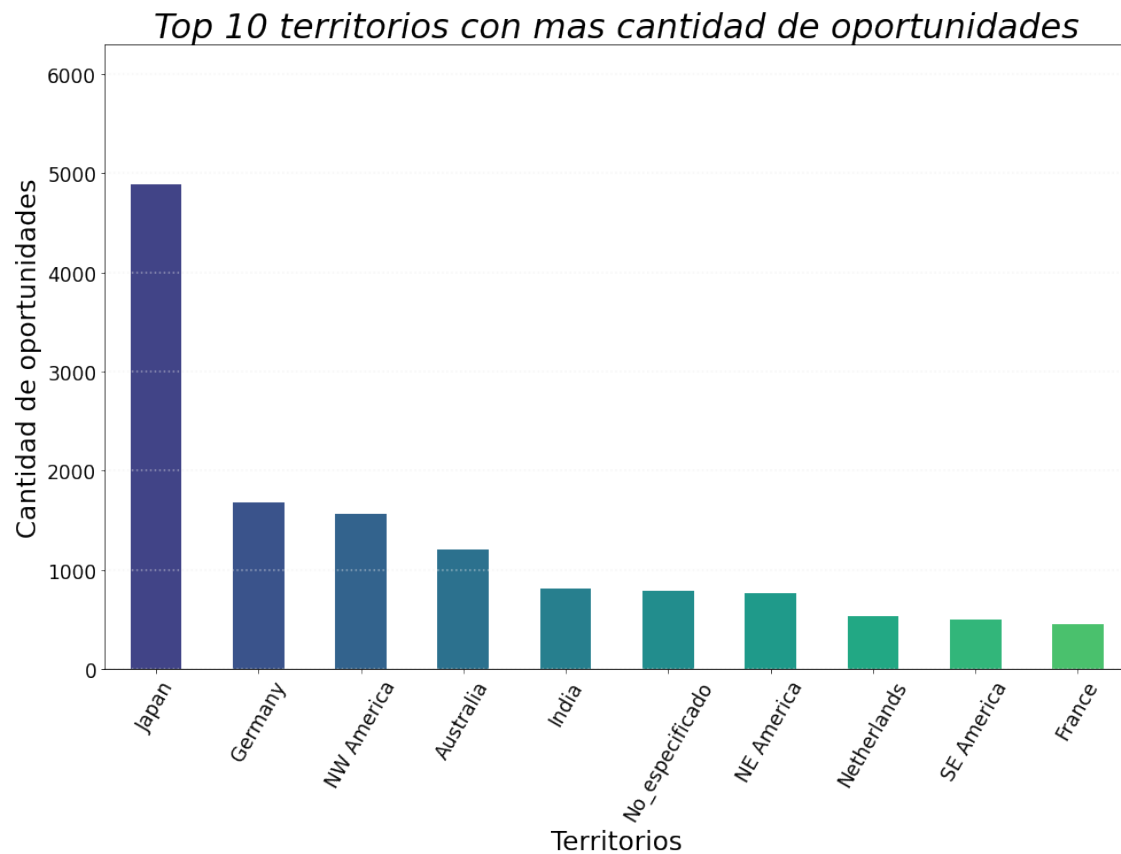


Figura 6

Como podemos ver en el gráfico, **el territorio con más oportunidades es Japón** por una gran diferencia, de mas de 3000 oportunidades al segundo territorio, que viene a ser Alemania. Después no hay una gran diferencia que valga la pena mencionar. Pero, podemos notar que **hay una gran cantidad de oportunidades (No\_especificado)**, de las cuales no se conoce el territorio. Veamos que pasa si ignoramos los no especificados.

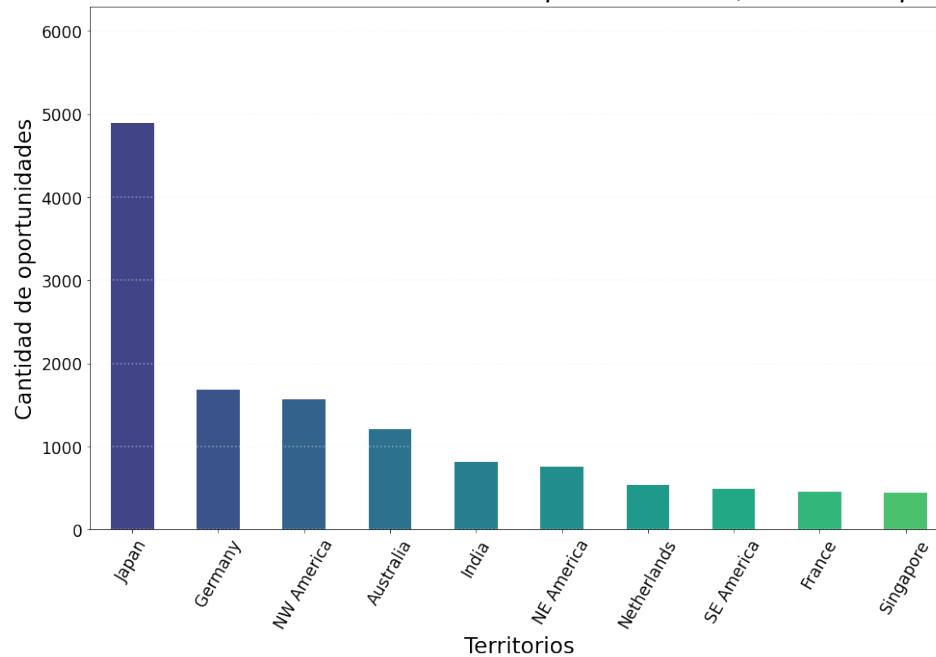
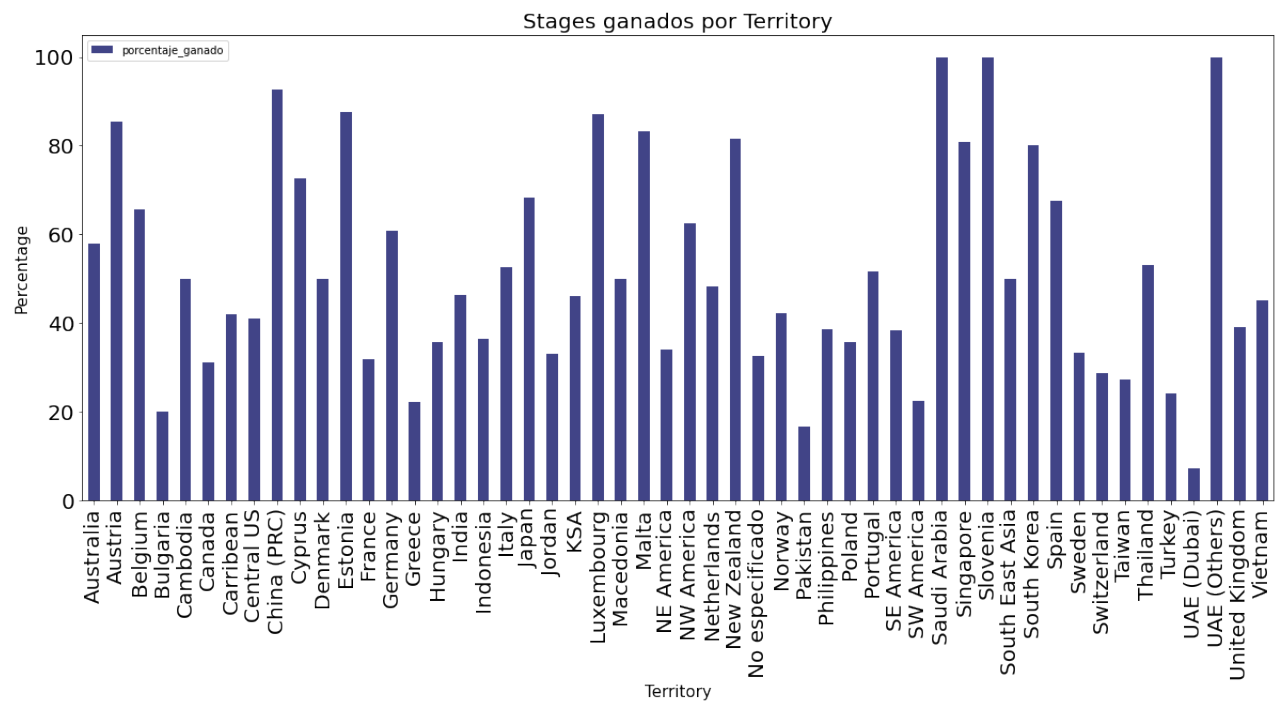
*Top 10 territorios con mas cantidad de oportunidades, sin "No especificados"*

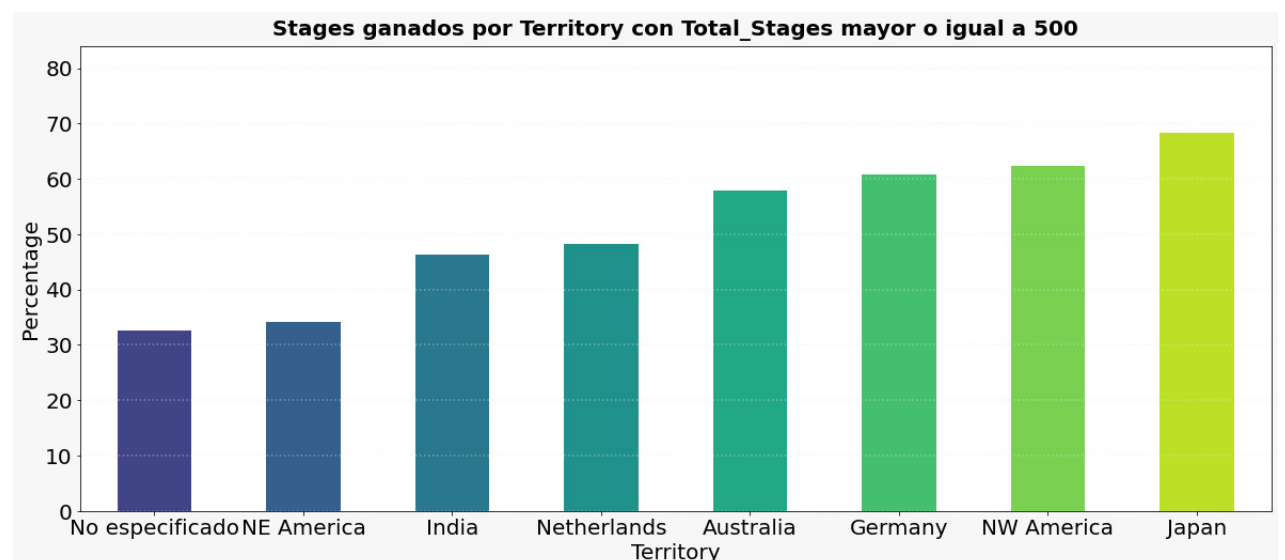
Figura 7

Gracias a que ignoramos los territorios no especificados, podemos ver que Singapur entra en el top10 de territorios con más oportunidades, entonces tenemos como una conclusión que **los 10 territorios con más oportunidades son: Japón, Alemania, América Noroeste, Australia, India, América Noreste, Holanda, América Sureste, Francia y Singapur.**

### 4.3. Porcentaje Stages por Territorio



Vemos que en este caso la cantidad de territorios es mucho mayor a la de las regiones. Para que el análisis tenga algún sentido, reducimos los territorios a los que tengan un *Total Stages* mayor o igual a 500



Vemos que **Japón** sigue siendo el ganador, mientras que *India*, *NE America*, *No especificado* no superan el 50 %. Si bien Australia no parece destacar veamos que pasa en el siguiente análisis.



#### 4.4. Territorios con más manejos de montos

En esta parte vamos a ver que sucede si buscamos los 10 países con mayor Monto Total.

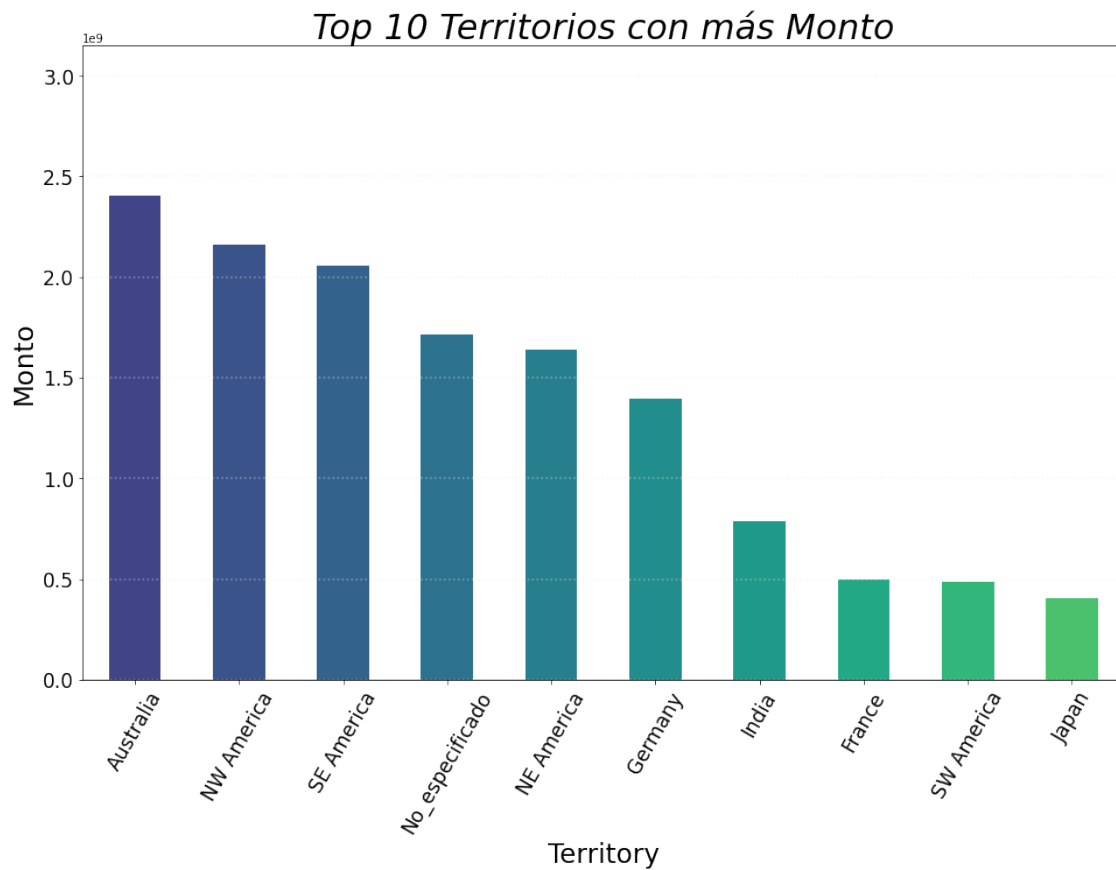


Figura 8

Podemos ver que el gráfico está bastante escalonado, y que el territorio con mayor monto es Australia, sin una gran diferencia, seguido de los territorios: América Noroeste, América Sureste, América Noreste, Alemania, India, Francia, América Suroeste y Japón. También, notemos que hay una gran manejo de Monto por parte de los territorios no especificados, veamos que pasa si ignoramos estos.

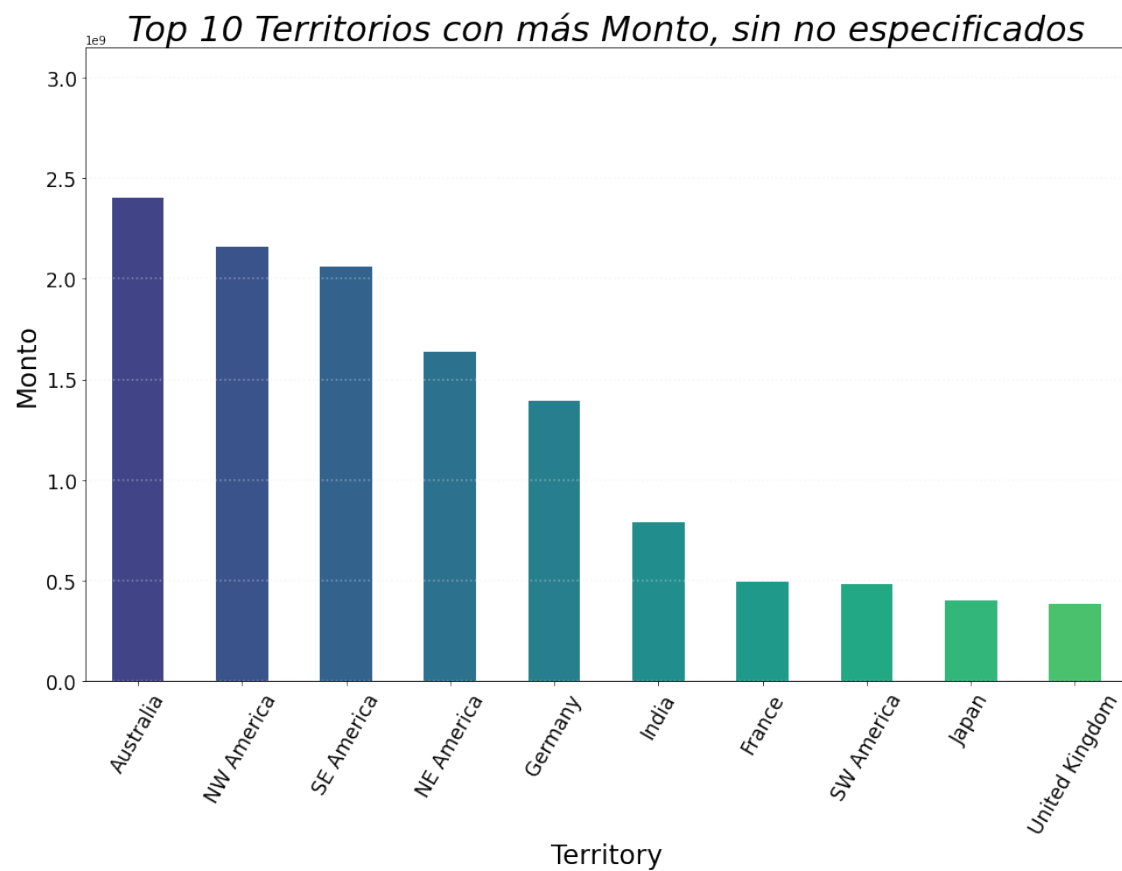


Figura 9

Al ignorar los no especificados, el territorio que entra en el top 10 es Reino Unido. Estos gráficos son sin diferenciar entre las stages, es por eso, que vamos a ver que pasa si buscamos los 10 territorios con mayor monto solo de las stages 'Closed Won'.

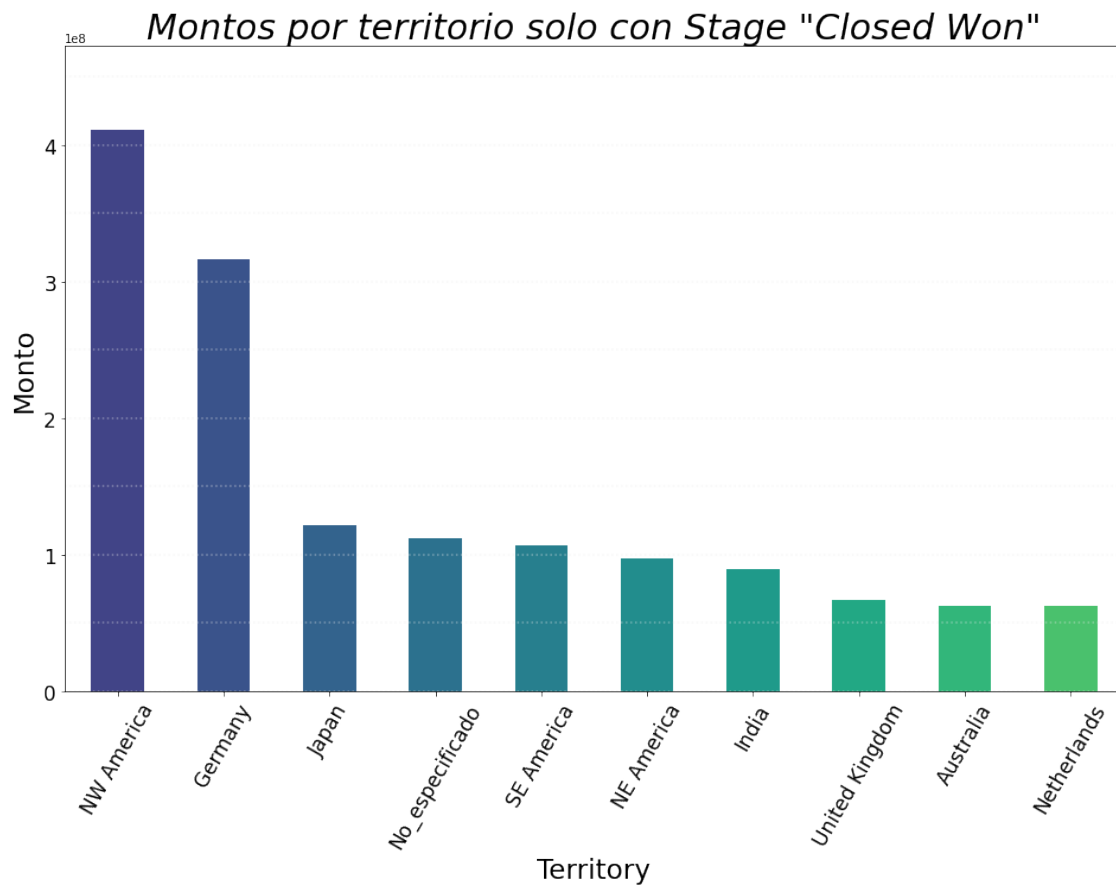


Figura 10

Es importante ver como Australia paso de estar primero a estar noveno, esto debe significar que una gran parte del monto debe pertenecer a la stage de 'Closed Lost'. Entonces, vemos que pasa si buscamos los 10 territorios con mayor monto en esta stage.

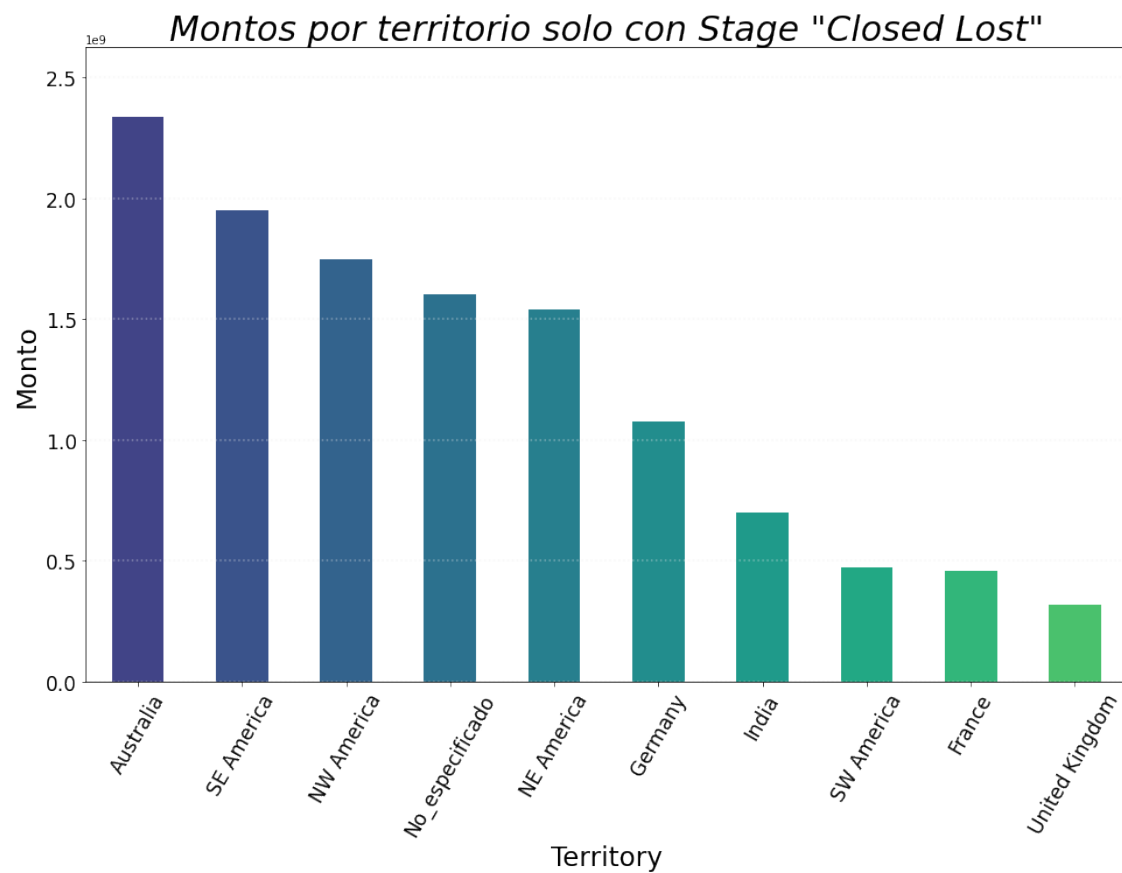


Figura 11

Con este gráfico podemos ver como Australia mantiene el primer puesto, y viendo las escalas entre este gráfico y el gráfico *'Top 10 Territorios con más Monto'*(8) notamos que la cantidad no varía mucho. Entonces proponemos ver que es lo que pasa en específico con Australia, en cuanto a los montos.

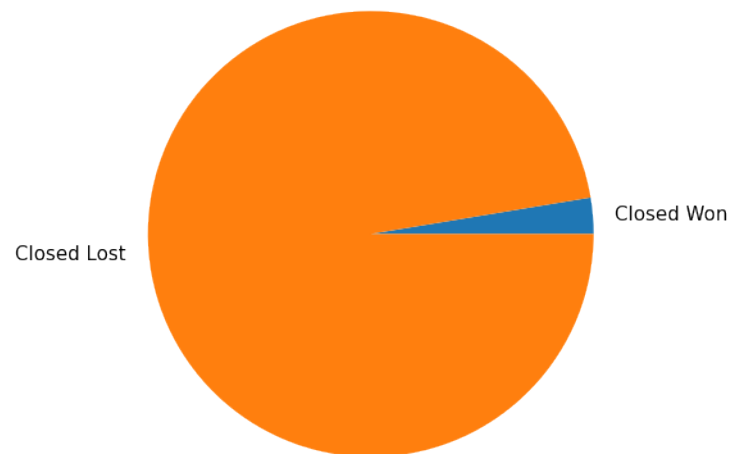
*Comparación entre Montos de Closed Won y Closed Lost de Australia*

Figura 12

Podemos ver como el monto que pertenece a la stage 'Closed Lost' es mucho mayor al monto que pertenece a la stage 'Closed Won', lo cual tiene sentido teniendo en cuenta los otros gráficos, pero ahora surgió una nueva duda. ¿Que pasa si vemos la cantidad de oportunidades de Australia que terminaron en 'Closed Lost' y 'Closed Won'?

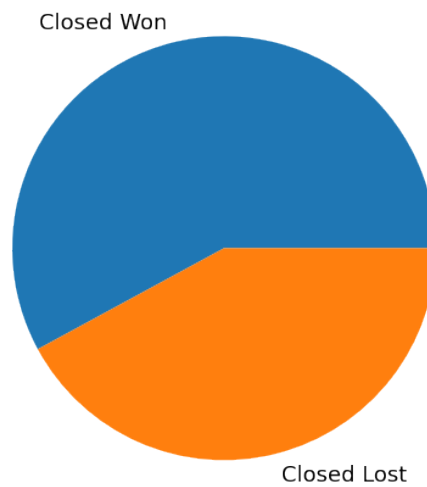
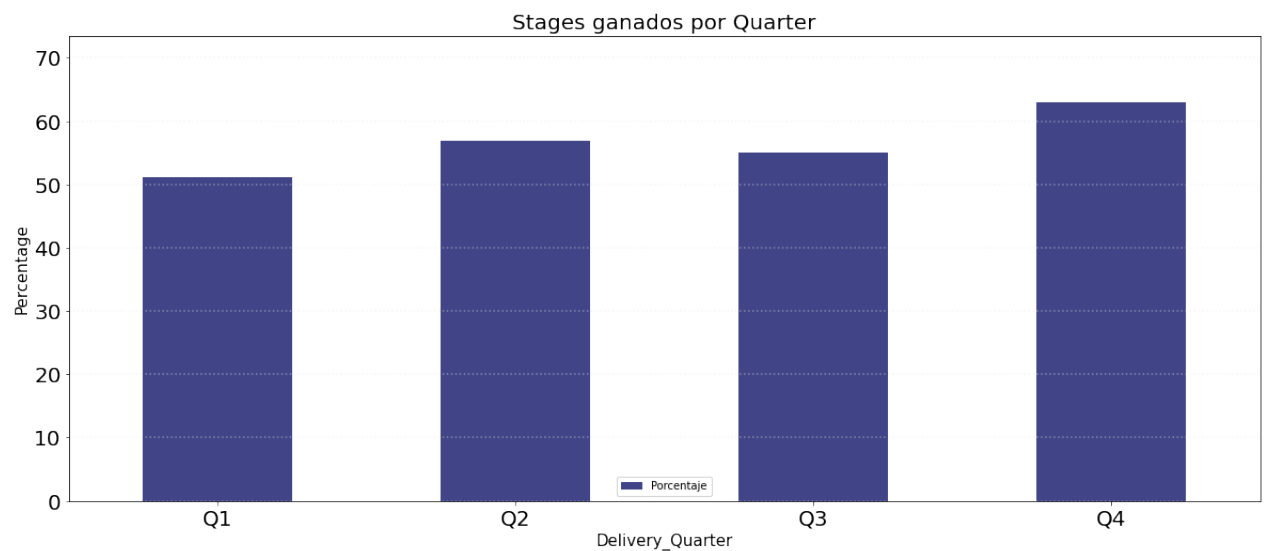
*Comparación entre cantidades de Closed Won y Closed Lost de Australia*

Figura 13

Es interesante el ver que la cantidad de oportunidades que terminaron en 'Closed Won' es considerablemente mayor a las 'Closed Lost', pero aun así el monto acumulado de las oportunidades que terminaron en 'Closed Lost' es mucho mayor al monto acumulado por las oportunidades que terminaron 'Closed Won'. Y tenemos como conclusión, que **en Australia las oportunidades que tengan un monto pequeño tiene una probabilidad mucho mayor de terminar en 'Closed Won'** que las oportunidades de montos grandes.

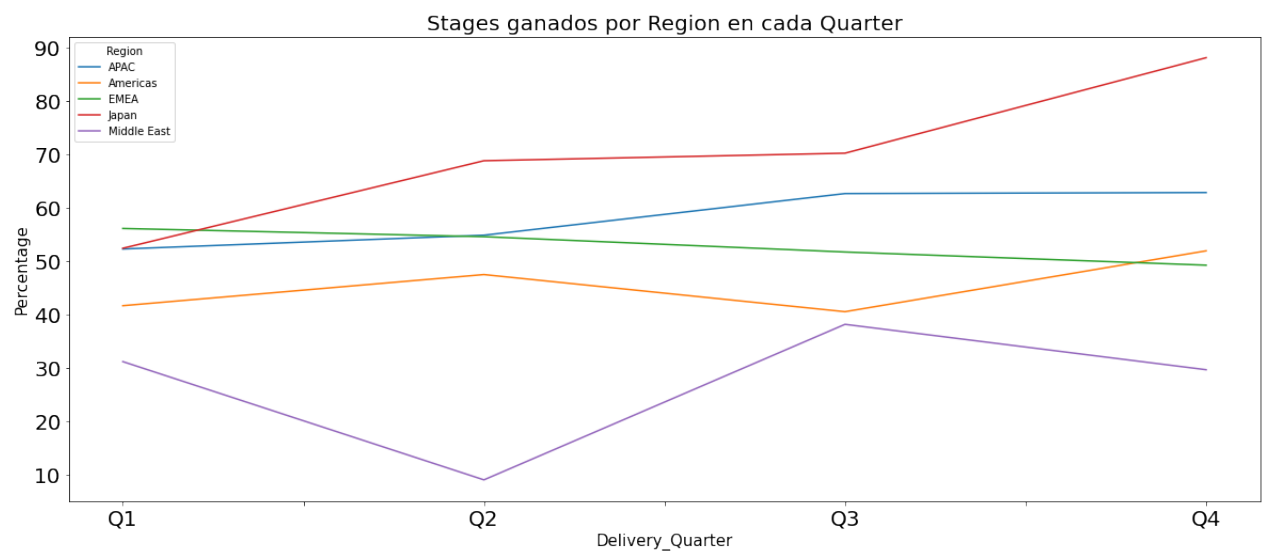
#### 4.5. Porcentaje Stages por Quarters

Seguimos con el análisis porcentual de los Stages pero ahora en los Quarters. La idea de este análisis es ver si hay alguna relación en los tratos que se cierran exitosamente según en que Quarter lo hagan.

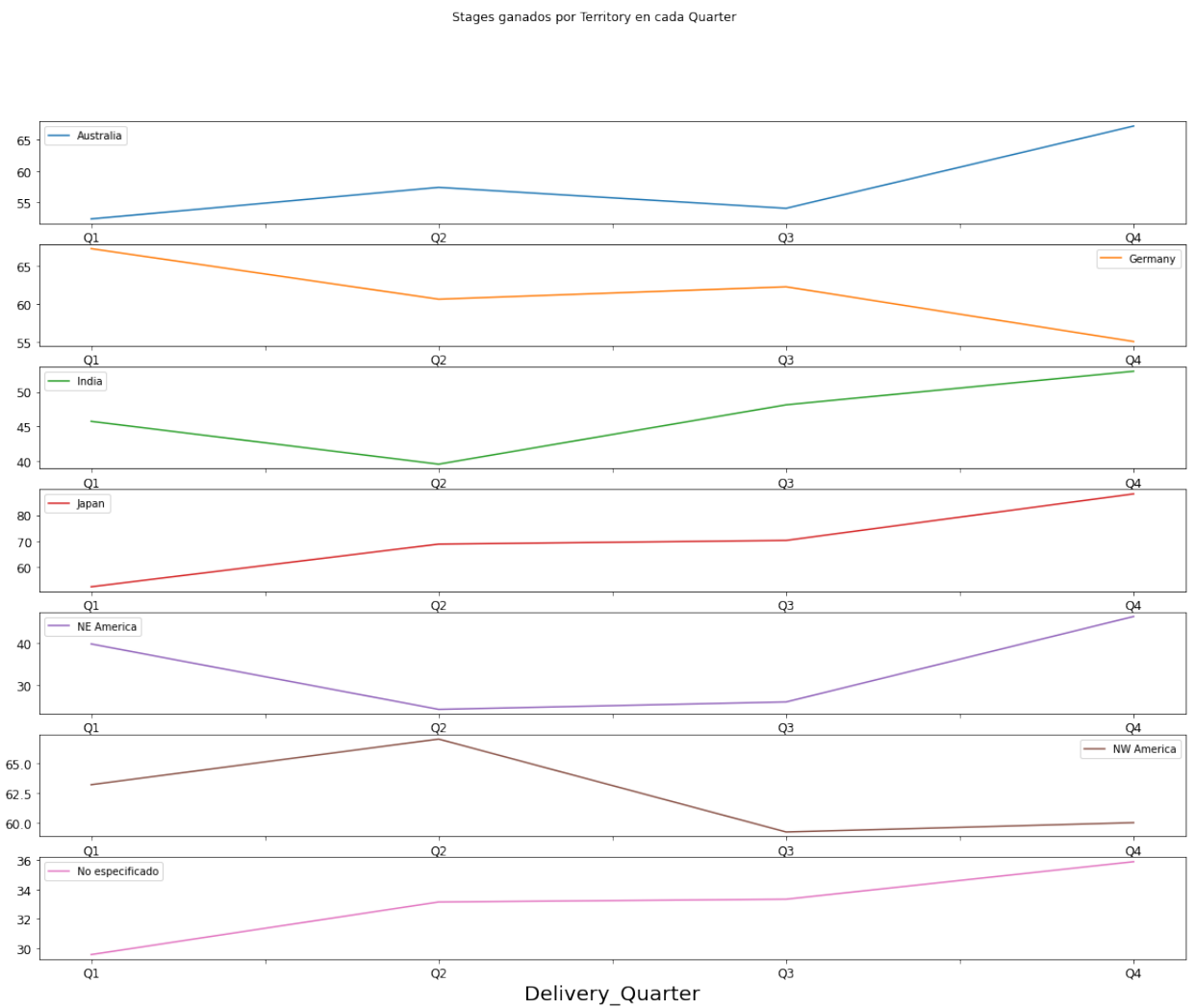


Si bien hay ligeras diferencias, todas superan el 50 %

Ahora queremos ver el porcentaje de las oportunidades cerradas exitosamente según las regiones en cada Quarter.



No vemos nada anormal. Repetimos esto mismo pero esta vez con los territorios.



Podemos resaltar que **Germany** es el único territorio que decae en el ultimo cuatrimestre, fuera de esto no vemos ninguna anormalidad. Cabe aclarar que muchos de los territorios no tenían datos para todos los Quarters, tomamos la decisión de eliminarlos de este análisis.



## 5. Análisis de Productos

En este análisis vamos a entrar en profundidad sobre como se dividen los productos, en base a tipos, categorías, familias, etc. Para ello se utilizaron ciertos datos y estos dieron paso a dos situaciones:

- Un análisis que contenga únicamente a los productos que tienen tipo especificado .
- Otro análisis que contenga únicamente a los productos que no tienen tipo especificado .

¿Por qué se hace esto? Al realizar el análisis notamos que muchos de los productos tienen datos sin especificar y creemos que a la hora de comprar un producto, aires acondicionados en este caso, es muy importante saber de que tipo de producto se trata. Por otro lado, esto nos facilitó investigar por separado, ya que predominan los productos que no son especificados, en caso de realizar el análisis sin separar los tipos de productos perderíamos información importante ya que los tipos de productos especificados serían opacados por los que no.

Empezamos hablando de tipos de productos especificados y no especificados, veamos como se plasman en el gráfico estos datos.

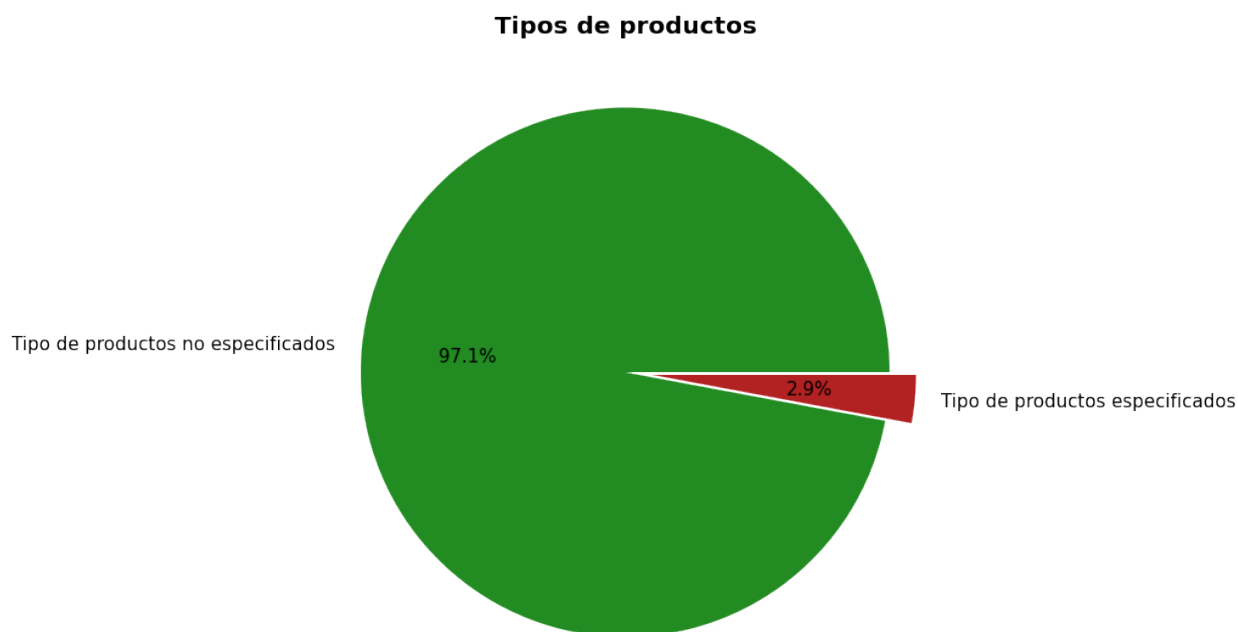


Figura 14: Tipos de productos especificados y no especificados.

Como se puede observar, de un total de 16947 productos (**100 %**), 497 tienen su tipo de producto (**2,9 %**) especificado y 16450 (**97,1 %**) son no especificados. Hay una gran diferencia entre ambos, por lo tanto hay que ser cuidadosos a la hora de realizar el análisis ya que no

podemos comparar de forma directa por la gran diferencia de datos. Comenzamos con lo dicho en las siguientes secciones.

## 5.1. Tipos de productos especificados

En esta sección se analizan aquellos productos que cuentan con un tipo de producto especificado, es decir, que el dataset brinda información sobre a que tipo hace referencia el producto, en nuestro caso podríamos verlo como distintos tipos de aires acondicionados. Como por ejemplo: Aires acondicionados portátiles, de ventana, frío y calor, etc.

¿ Por qué nos interesa saber de que tipo de producto estamos tratando?

Sencillo, queremos saber cuales son los distintos tipos de productos que dominan el mercado, o cuales son aquellos que son más frecuentados.

### 5.1.1. Distintos Tipos de productos especificados

#### Cantidad de productos por tipo de productos especificados

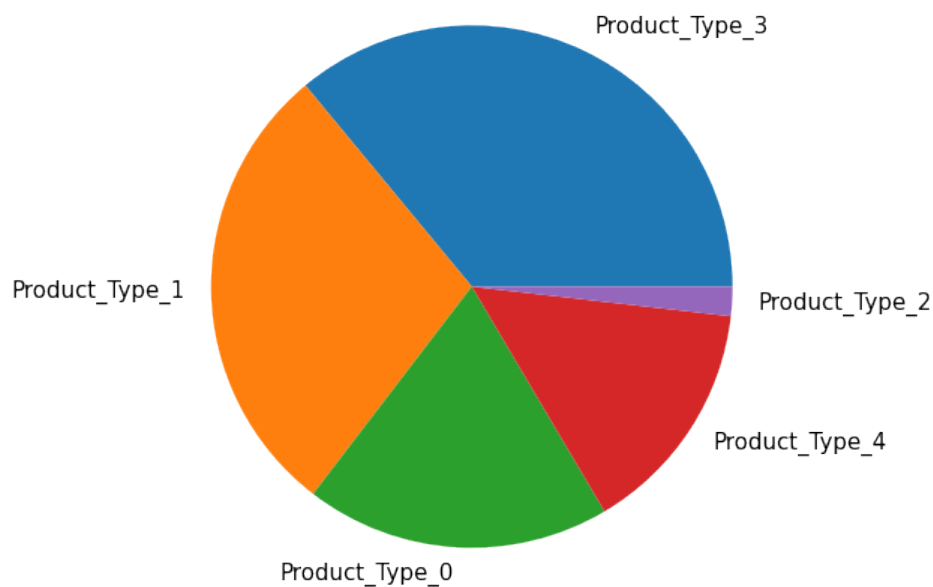


Figura 15: Cantidad de productos por tipo de producto.

Podemos observar que contamos con una poca variedad de tipos de productos, pero que se encuentran bien distribuidos.

Veamos un poco más específico como se vería en un gráfico de barras.

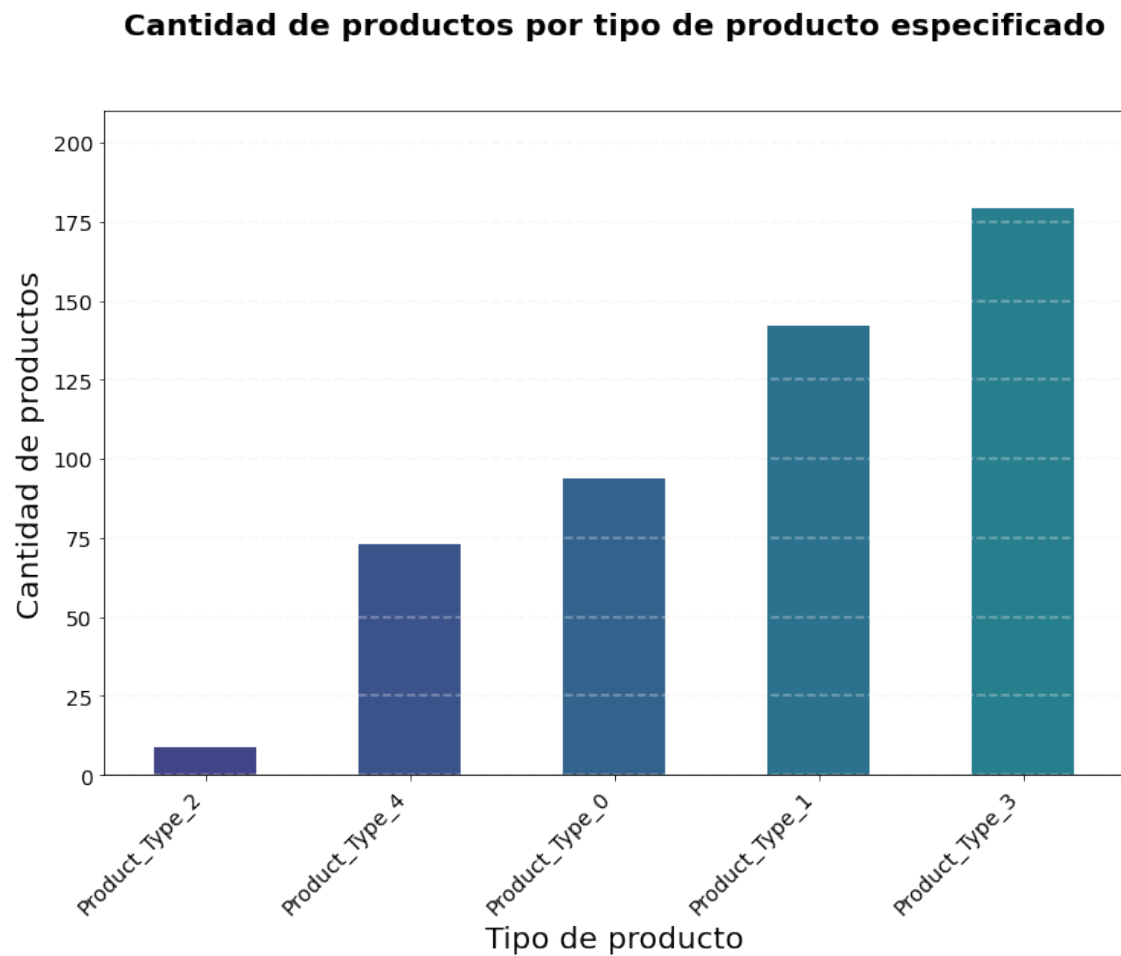


Figura 16: Cantidad de productos por tipo de producto.

Se puede observar claramente que el tipo de producto 3 es el que domina el mercado dentro de los tipo de productos especificados pero no por mucho, no hay mucha de diferencia entre cantidades. El único que no influye mucho dentro de este gráfico es el tipo de producto 2.

### 5.1.2. Productos más frecuentados

¿Por qué nos interesan?

Es uno de los datos principales, ya que cada oportunidad es sobre un producto específico. Queremos averiguar cuales son aquellos productos que Dominan en el mercado sobre los demás, estos productos son las principales movilizaciones del mercado, por lo tanto es importante saber de cuales se trata.

Para visualizar esto vamos a usar wordclouds donde el tamaño de los elementos representa su frecuencia, al verse más grandes quiere decir que son frecuentados más veces que los demás. De esta forma sabremos cuales son los productos más frecuentados.

### Productos más frecuentados por tipo de productos especificados

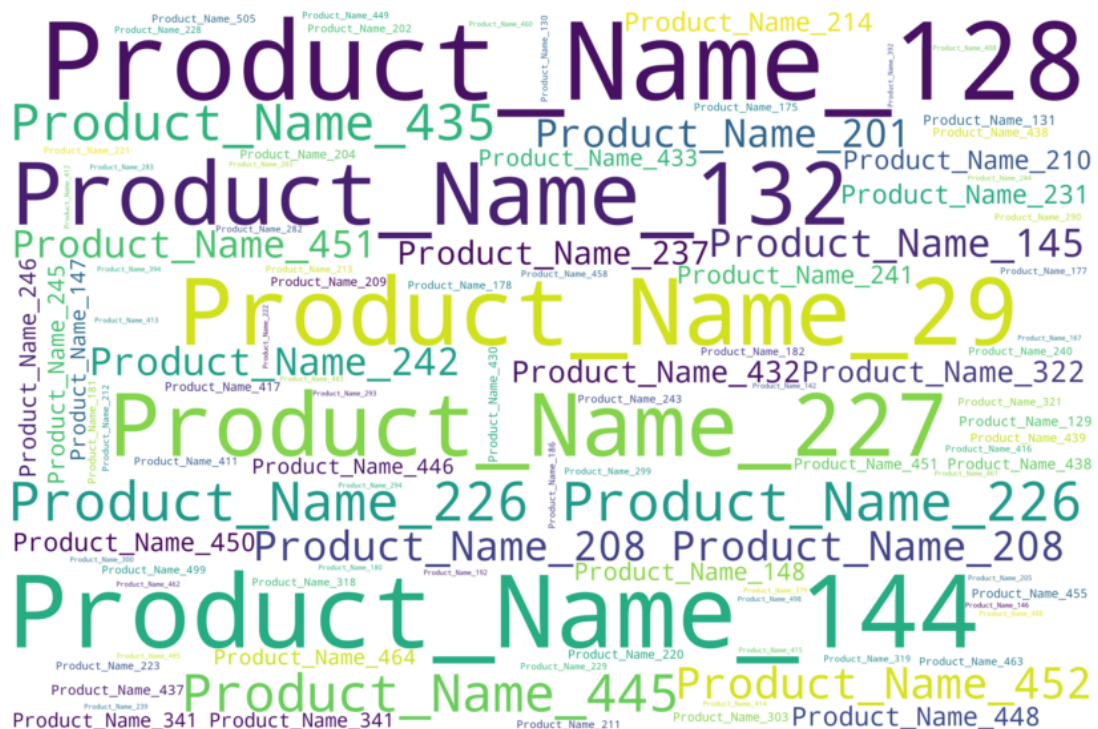


Figura 17: Productos más frecuentados.

Podemos observar que los productos más frecuentados son los productos:

128, 132, 29, 227 y 144.

Dentro de los tipo de producto especificados estos son sus productos más vendidos.

En relación con el gráfico de los tipo de productos podemos deducir que aquellos productos que son más frecuentados serán de los tipo de productos con mayor cantidad de productos.

### 5.1.3. Caracterización de los productos

Los productos pueden dividirse por familias y por distintas categorías, a y b, primero vamos a analizar como se dividen en cuanto a las familias y luego como son en base a las categorías. Planteamos unas preguntas: ¿Habrà variedad en cuanto a las familias y las categorías o será un monopolio? ¿Habrán categorías o familias sin especificar?

#### Top 10 mayor cantidad de productos por familia en tipos de productos especificados

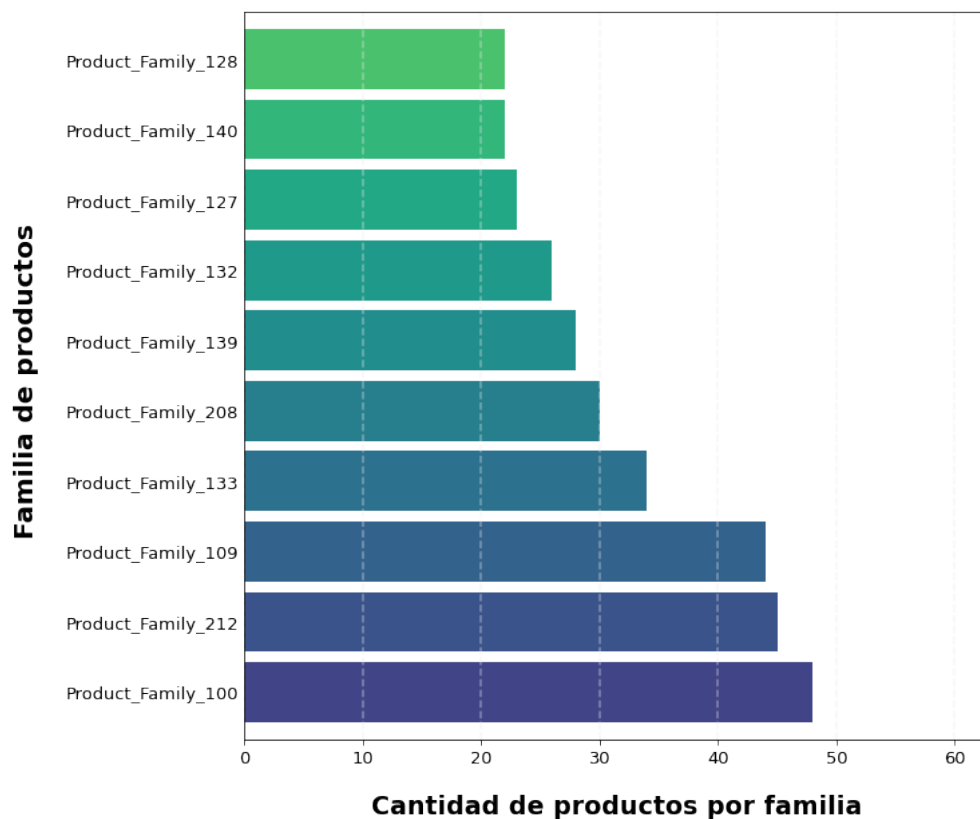


Figura 18: Top 10 mayor cantidad de productos por familia.

Este gráfico nos interesó bastante, ya que manejando una pequeña cantidad de datos que haya bastante variedad de familias, y que dentro de todo, estén bien distribuidas en cuanto a las cantidades. No hay una familia que sea el monopolio del mercado. ¿Serán las mismas subcategorías cuando se incluyan los productos tipos de productos sin especificar?

Ahora procedemos a analizar los productos en base a la subcategoría B. Dentro de nuestro análisis descubrimos que hay subcategorías B que no están especificadas. Entonces volvemos a partir en 2 nuestros gráficos, especialmente para observar cuantos datos volvemos a perder por no tener ciertas características especificadas. Primero veamos las categorías b especificadas

### Top 10 mayor cantidad de productos por categoria B en tipos de productos especificados

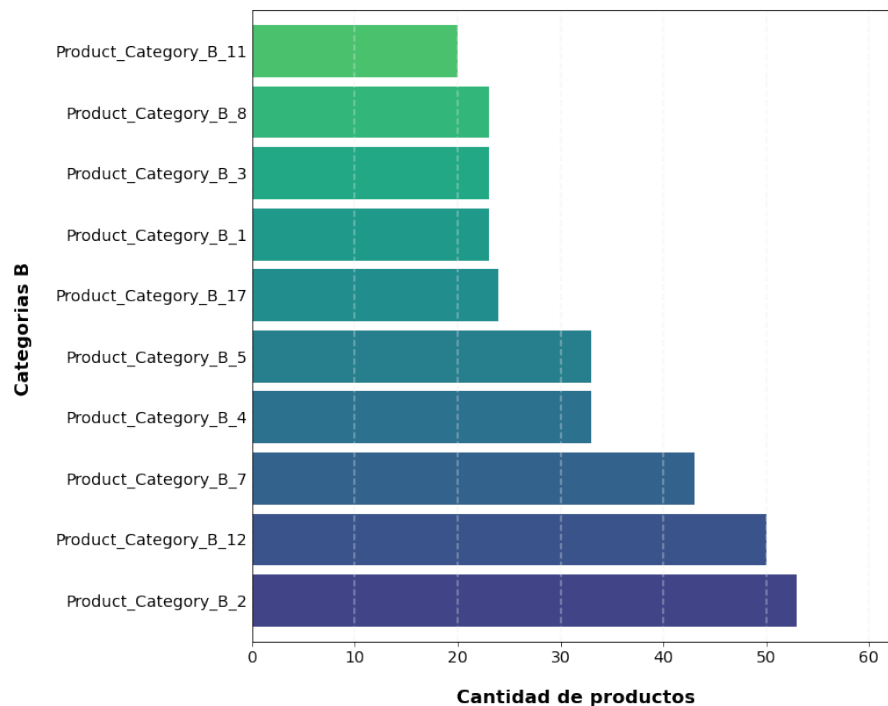


Figura 19: Top 10 mayor cantidad de productos por subcategoría b especificada.

Volvemos a sacar la misma conclusión, no hay una subcategoría que domine el mercado. Hay bastante variedad de subcategorías b para la pequeña cantidad de datos que hay. Volvemos a preguntarnos. ¿Serán las mismas subcategorías cuando se incluyan los demás productos?

Continuamos con el análisis de las categorías b, en este caso categorías b no especificadas.

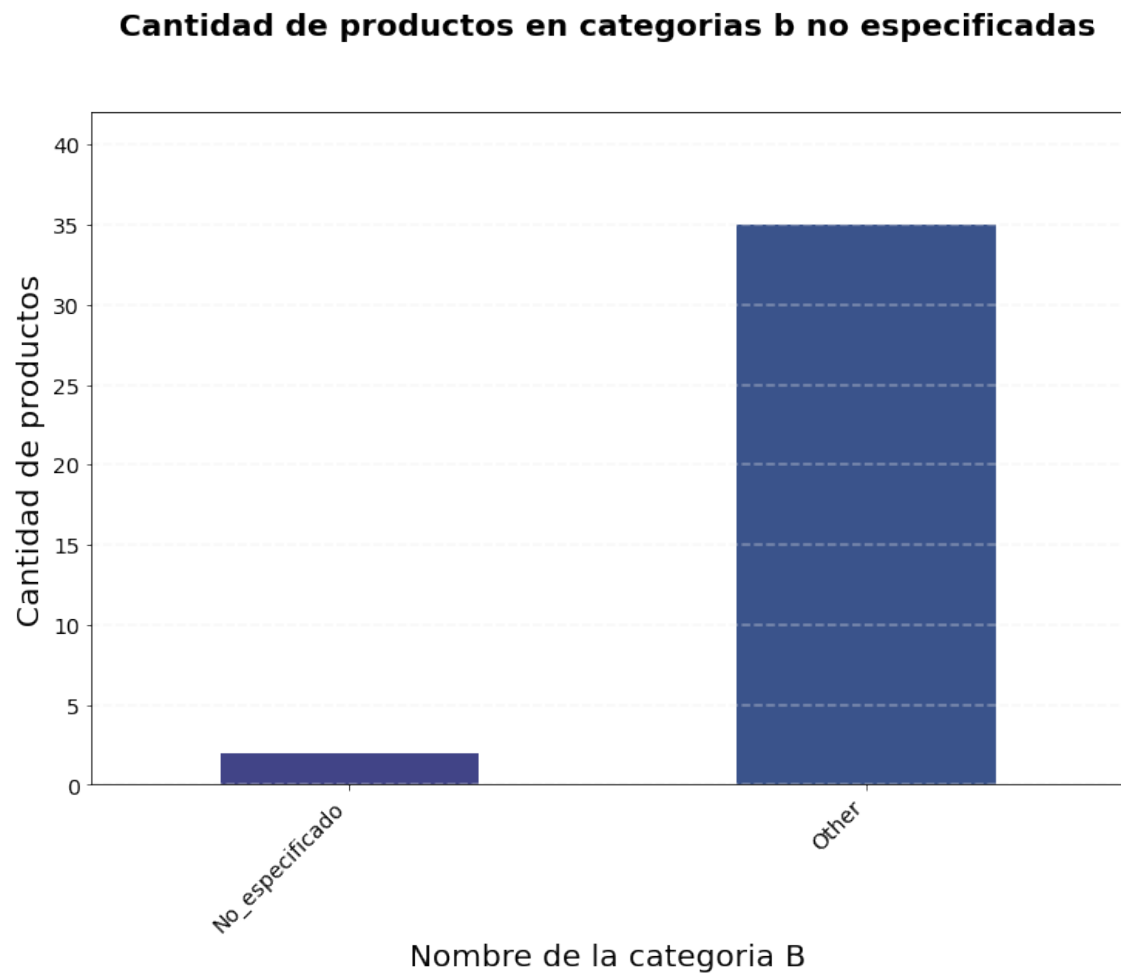


Figura 20: Cantidad de productos con categoría b sin especificar.

Observamos que casi un 10 por ciento de nuestros productos no tienen categoría b especificada. Destacamos que a medida que filtramos (para obtener datos más específicos) dentro de los productos vamos perdiendo gran cantidad de datos.

Ahora procedemos a analizar los productos en base a la subcategoría A.

### Cantidad de productos en categorías a

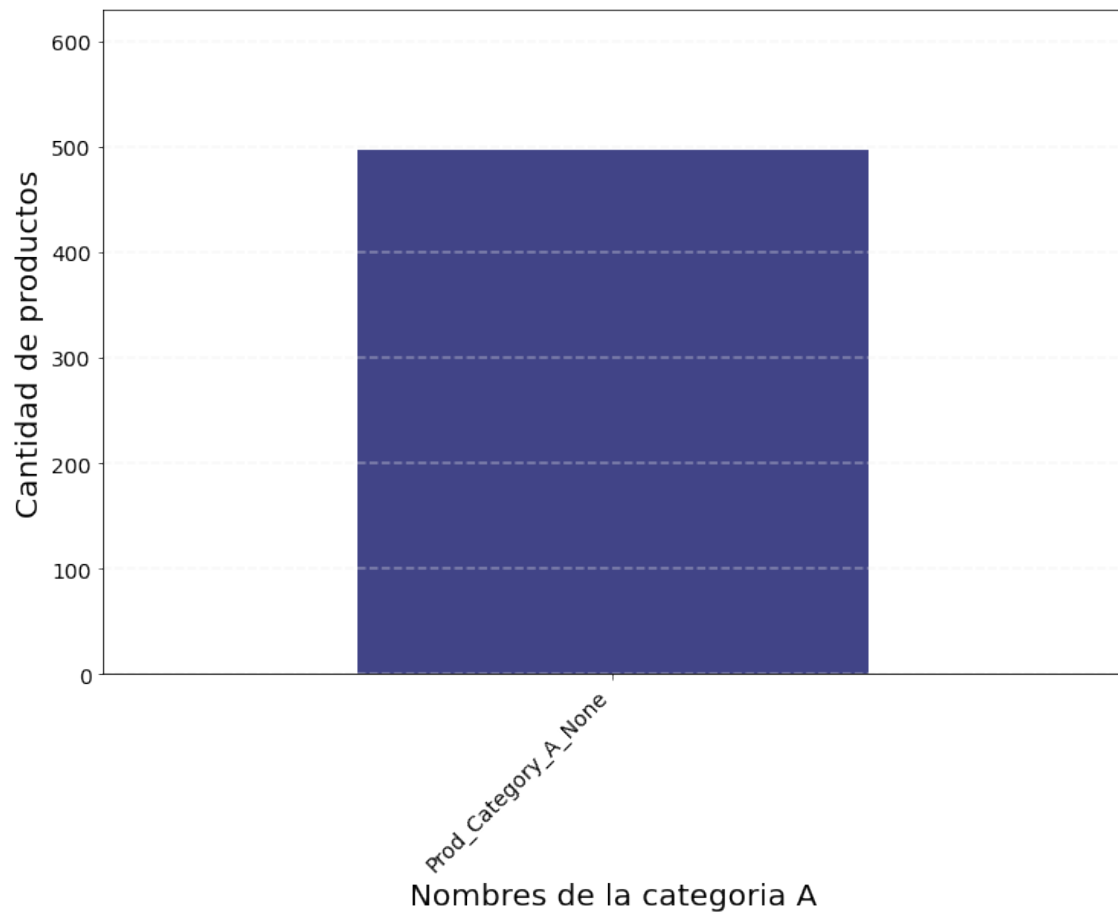


Figura 21: Productos más frecuentados.

A comparación de los gráficos anteriores, se puede observar fácilmente que solo hay una categoría A. La cual es la que domina el mercado en cuanto a un monopolio dentro de los tipos de productos especificados. Pero recordemos que estamos hablando de una pequeña cantidad de productos, más específicamente un 2,9 por ciento del total de productos. ¿Seguirá dominando el mercado cuando incluyamos el análisis de los tipo de productos no especificados?



## 5.2. Tipos de productos no especificados

Como continuación de la sección anterior, seguimos analizando los tipos de productos pero en este caso aquellos que no tienen tipo de producto especificado. Como hemos dicho anteriormente los tipos de productos no especificados son mayoría dentro de este dataset. ¿Encontraremos familias y/o categorías especificadas o no habrá ningún dato concreto? ¿Habrá alguna categoría de productos que domine el mercado? ¿Y alguna familia?

A continuación veremos los distintos tipos de productos no especificados y que cantidades manejamos.

Cantidad de productos por Tipo de productos (no especificado)

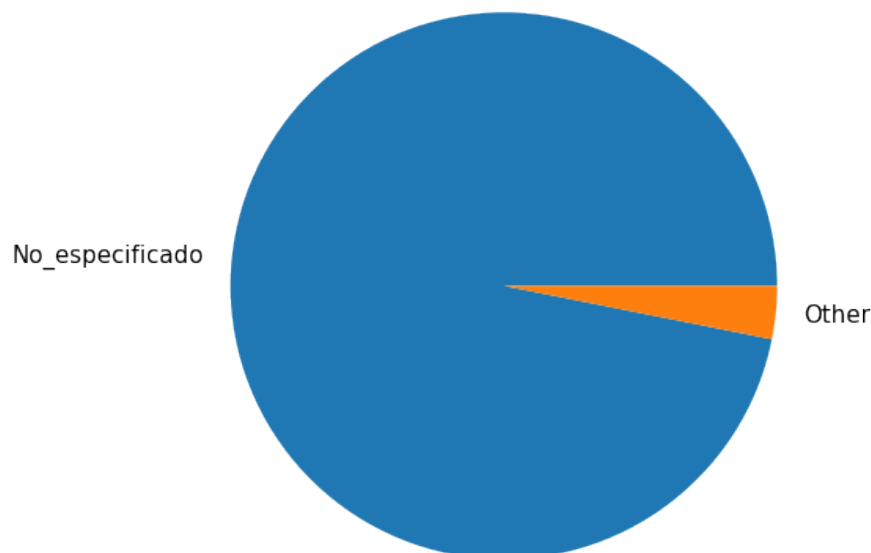


Figura 22: Tipos de productos no especificados.

Como podemos observar solo tenemos dos tipos de productos no especificados, pero a comparación de antes, tenemos mucha mayor cantidad de datos, como se podrá observar en el siguiente gráfico.

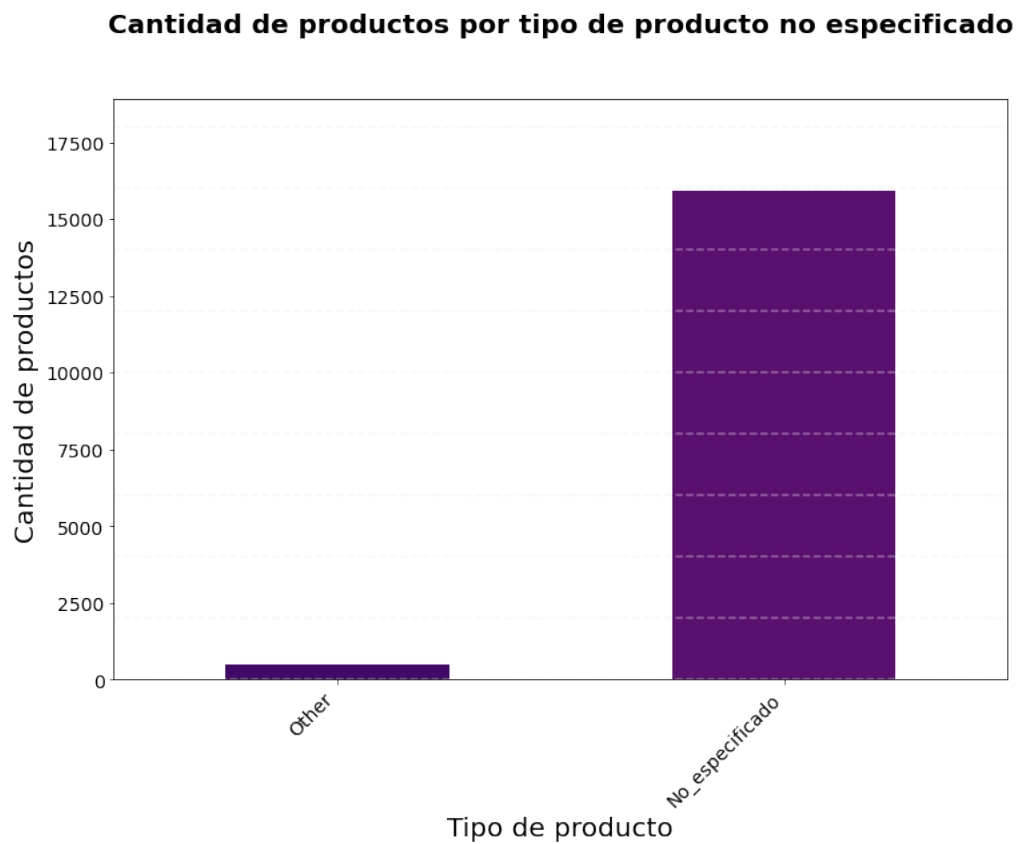


Figura 23: Cantidad de tipos de productos no especificados.

Como se podrá ver, a comparación de los tipos de productos especificados, estamos manejando grandes cantidades de productos. Por lo tanto fue una buena decisión separar los tipos de productos en dos análisis diferentes.

Ahora nuestro desafío es ver como influye esto en nuestro análisis comparado con lo visto anteriormente.

### 5.2.1. Productos más frecuentados

En esta sección vamos a realizar el mismo análisis que hicimos en los tipos de productos especificados pero para los tipo de productos no especificados.

### Productos más frecuentados en tipos de productos no especificados

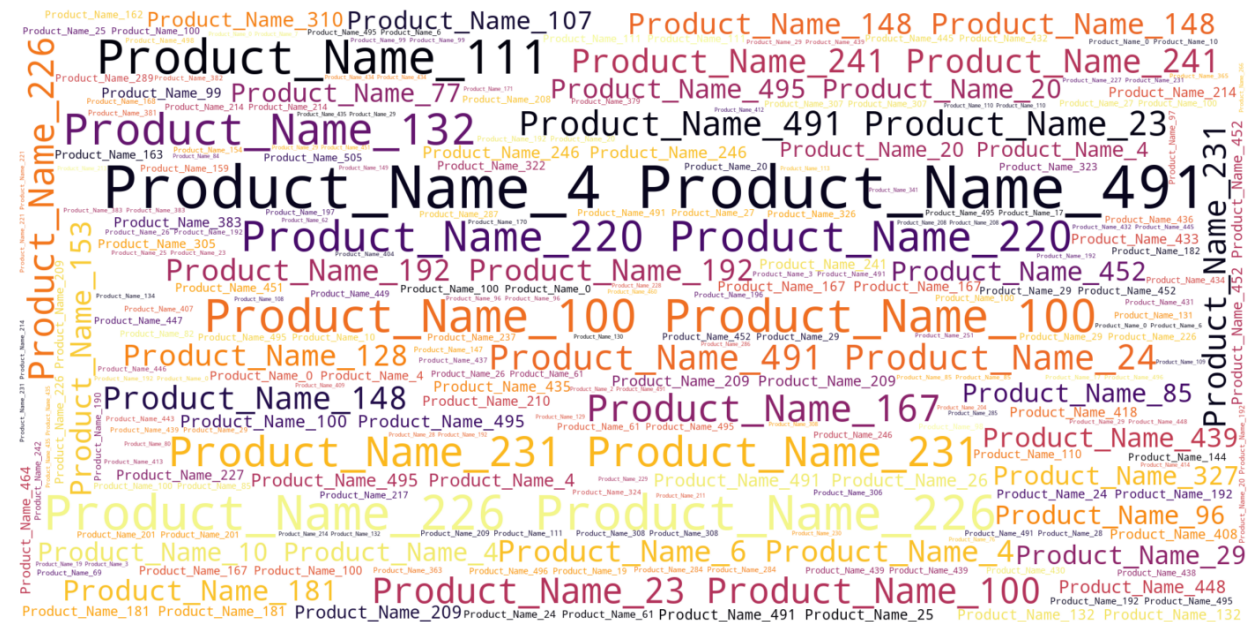


Figura 24: Productos mas frecuentados.

Como se puede ver tenemos mucha más variedad de productos, recordemos que anteriormente en los tipos de productos especificados eran 497 productos, acá estamos visualizando una cantidad de 16450 productos. Por lo tanto aquellos que son más grandes, no se podrían comparar con los mismos productos que se encuentran dentro del tipo de producto especificado, ya que seguramente debe contar con mayor cantidad de ocurrencias.

### 5.2.2. Caracterización de los productos

Continuamos con el mismo orden del análisis de tipo de producto especificado, planteamos unas preguntas antes de comenzar con esta sección. ¿Habrá variedad de familias y subcategorías de productos o tendremos una que domine el mercado?

#### Top 10 mayor cantidad de productos por familia en tipos de productos no especificados

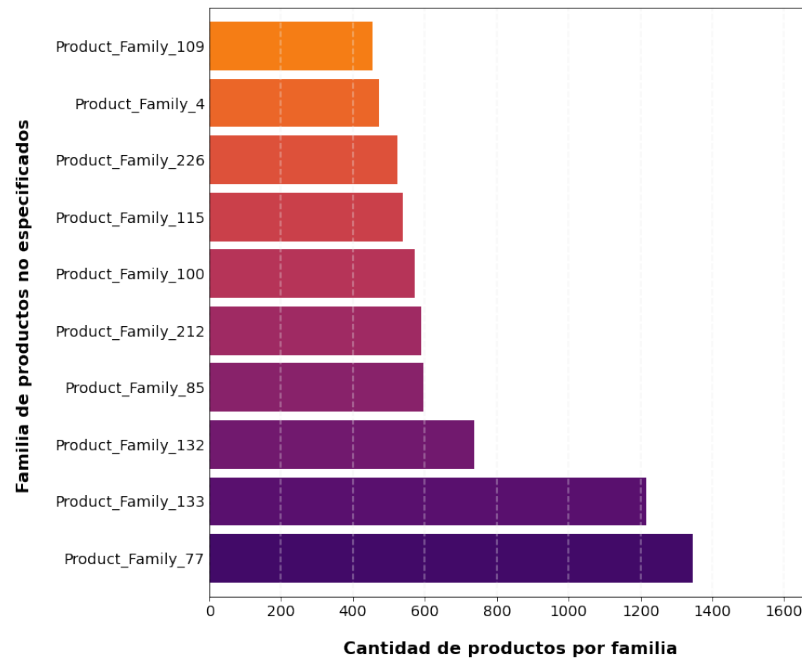


Figura 25: Top 10 Familias con mayor cantidad de productos.

Como vemos claramente hay una gran diferencia en cuanto a las cantidades que tienen las familias de productos, en comparación con las de los productos especificados. Pero sigue existiendo variedad en cuanto a las familias de productos. De este lado del análisis solo varió en cuanto a cantidades de productos por familia y la cantidad de familias (Recordemos que trabajamos con muchos más datos).

Seguimos con la subcategoría b. Primero con las subcategorías b especificadas y luego con aquellas que no estén especificadas.

### Top 10 mayor cantidad de productos por categoria b en tipos de productos no especificados

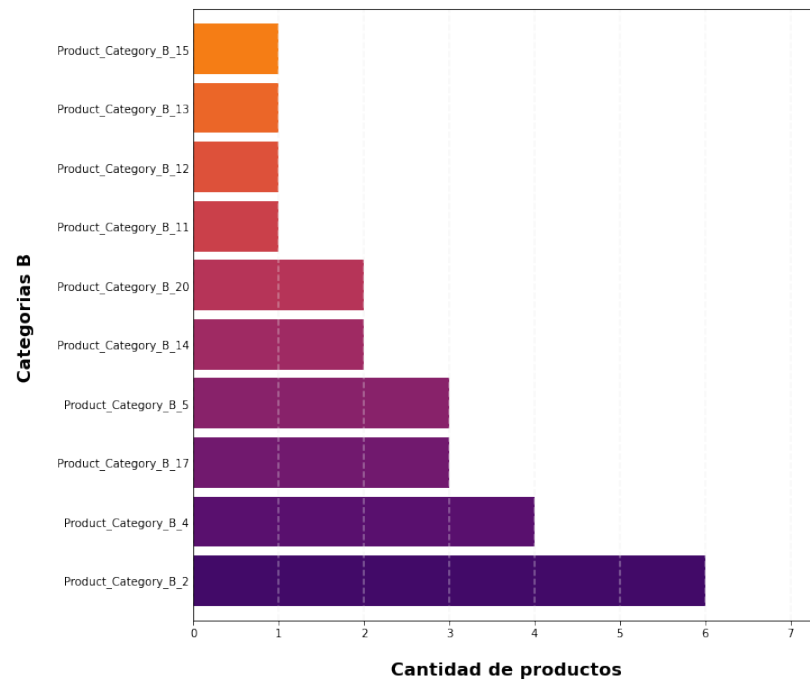


Figura 26: Top 10 Categorías b especificadas con mayor cantidad de productos.

Observamos que las categorías b especificadas cuentan con muy poca cantidad de productos, esto nos estaría indicando que vamos a tener casi el total de categorías b sin especificar. Dentro de los tipos de productos sin especificar tenemos muy pocos productos con datos concretos.

Continuamos con las subcategorías b sin especificar.

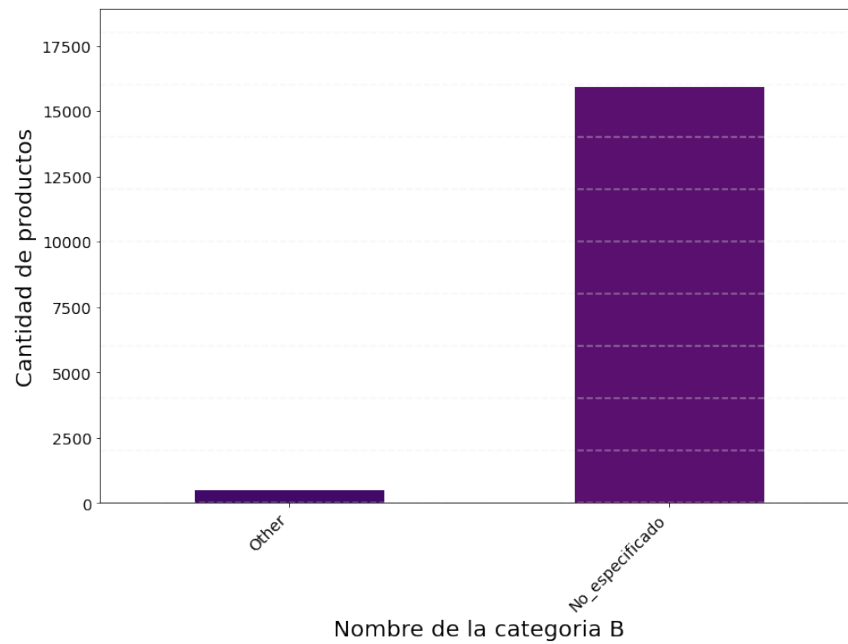
**Cantidad de productos en categorías b no especificadas en tipos de productos no especificados**

Figura 27: Cantidad de productos en categorías b no especificadas.

Claramente se puede ver que las categorías b no especificadas dominan el mercado sobre las demás categorías, incluyendo a las categorías de los tipos de productos especificados, ya que la gran mayoría de los productos tienen categoría b no especificada. Cada vez tenemos menos datos para especificar.

Por ultimo vemos las categorías a. ¿Habrá variedad o seguiremos con el monopolio?

### Cantidad de productos en categorías a

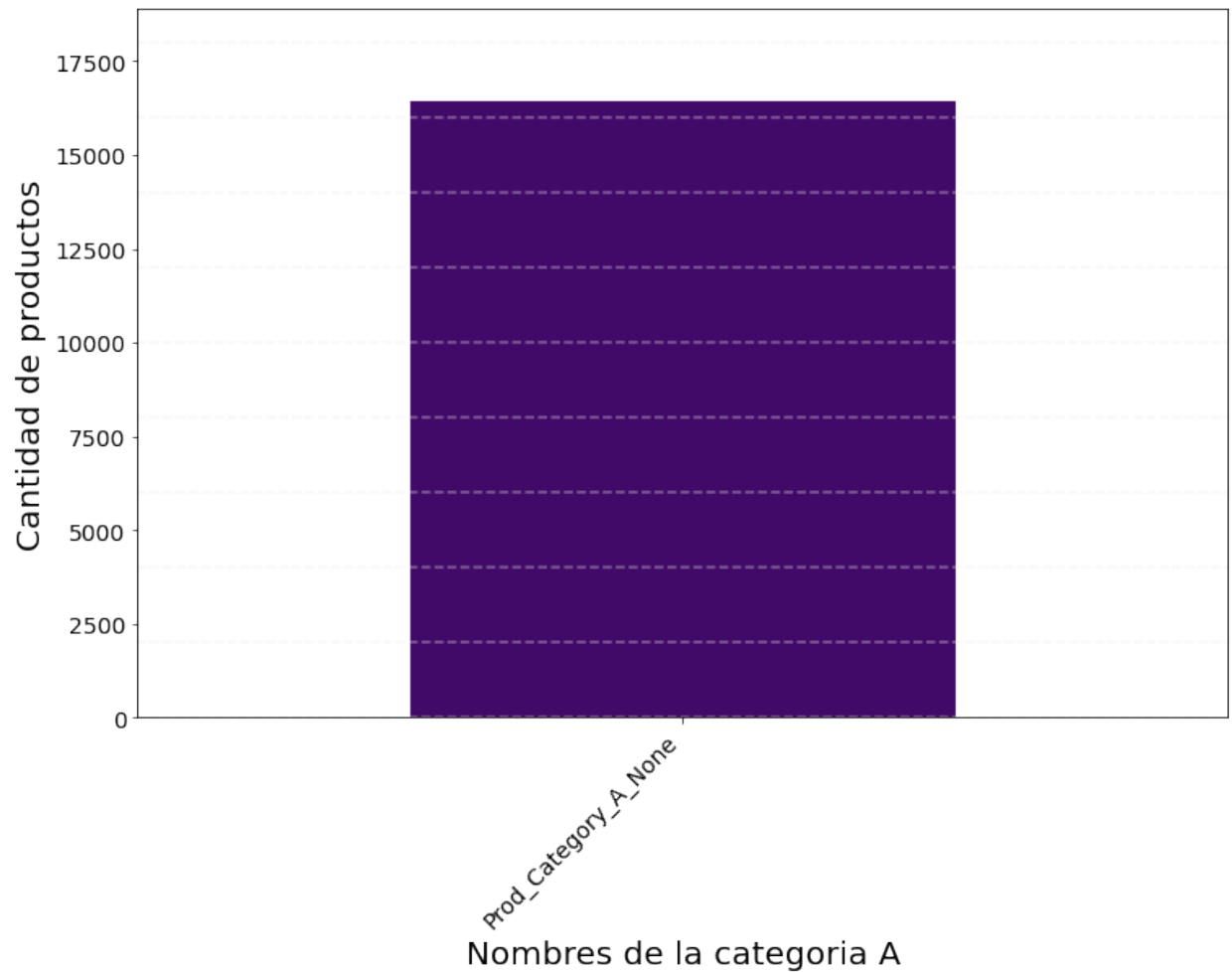


Figura 28: Cantidad de productos en categorías a.

Como se puede notar en el gráfico solo existe una categoría a para los productos. Por lo tanto tenemos un monopolio dentro de las **Categorías a** tanto para los tipos de productos especificados como para los que no.

### 5.3. Conclusión Productos

Con los distintos análisis que hicimos nos vamos dando cuenta que a medida que vamos filtrando, mientras más específico es el producto en cuanto a sus tipos, categorías y familias encontramos menor cantidad de productos, esto quiere decir, que realmente sabemos datos específicos en una pequeñísima cantidad de productos. Ya filtrando por los tipos de productos perdemos un **97,1 %** de los productos (16450 para ser más específicos). Y luego si dentro de estos volvemos a filtrar por lo dicho anteriormente notaremos que volvemos a tener pérdida de información.

Contamos con mucha variedad de **Familias de productos** y de datos concretos en cuanto a los dos análisis, pero cuando hablamos de subcategorías en estas predominan los datos sin especificar nuevamente, en la **Categoría a** hay un monopolio y en la **Categoría b** se podría decir casi lo mismo ya que existe en su gran mayoría la subcategoría b sin especificar.

No está demás decir que los tipo de productos sin especificar junto con subcategorías a y b sin especificar son aquellos que tienen grandes cantidades de productos dentro del mercado.



## 6. Análisis de Marcas

### 6.1. Incidencia de Marcas Por Región

Un aspecto importante a tener en cuenta, es la marca de los productos ofrecidos en cada oportunidad. Es por este motivo que vamos a profundizarnos en esta variable de nuestro set de datos. Para comenzar con su respectivo análisis, podemos ver su incidencia respecto a las regiones.

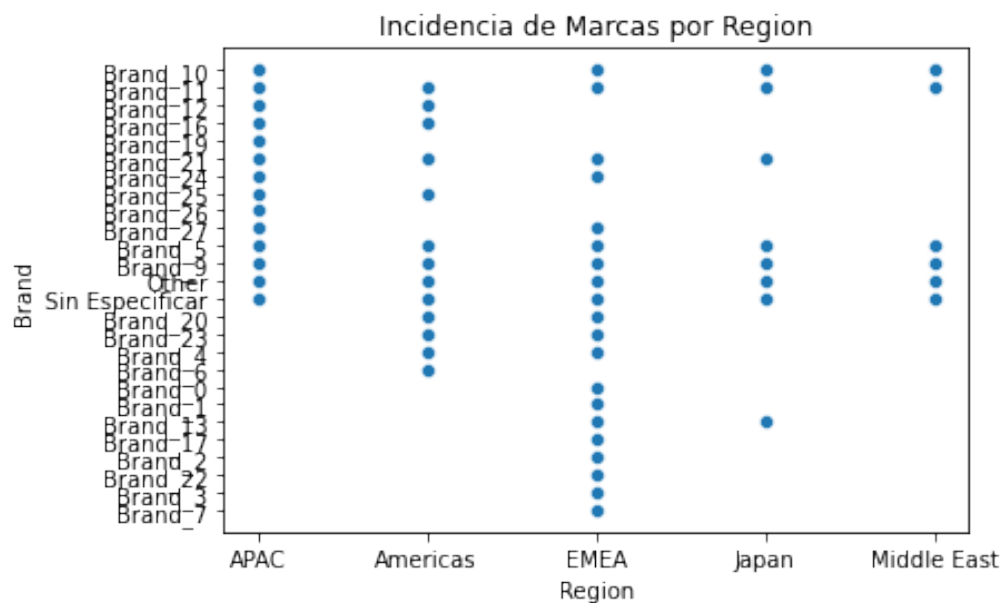


Figura 29: Incidencia de marcas por región.

A simple vista, parece una relación normal. Cuando vemos mejor nuestro gráfico, notamos algo: Las marcas 'Sin Especificar' y 'Other' aparecen en todas las regiones. La Marca 9 y 5 también, pero cuando profundicemos, notaremos una característica especial de las marcas especificadas.

### 6.2. Oportunidades de Marcas

Para generalizar el análisis de las marcas, veremos que tantas oportunidades tuvo cada marca. A continuación, un gráfico que nos muestra la cantidad de oportunidades por marca.



Figura 30: Cantidad de oportunidades de todas las marcas.

Aquí es donde nos aparece un problema, las marcas sin especificar se roban todo el protagonismo. Debido a esto, conviene hacer un análisis por separado. Así que, filtrando por las marcas especificadas (las marcas de tipo *Other* también fueron filtradas), he aquí las oportunidades por marcas especificadas.

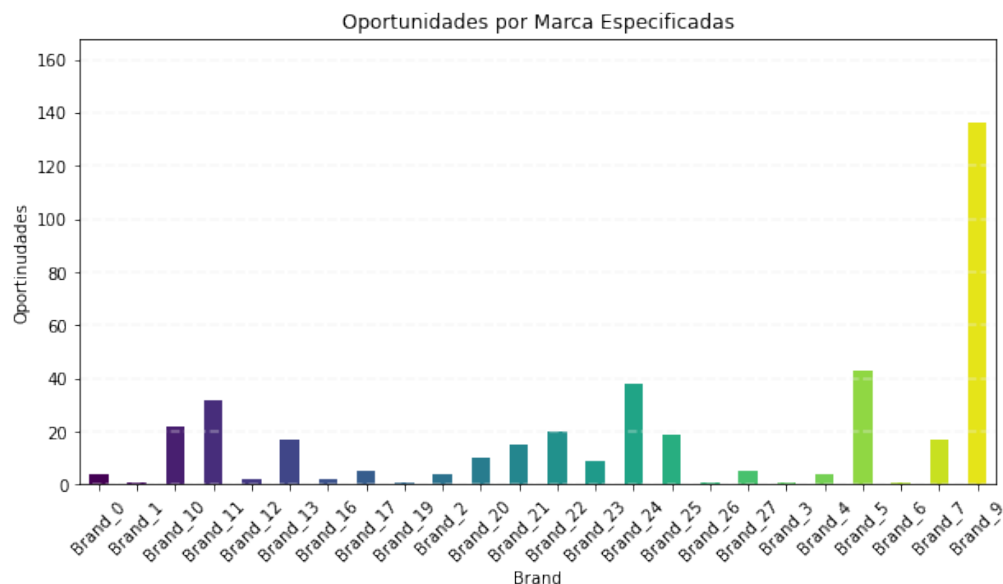


Figura 31: Cantidad de oportunidades de todas las marcas especificadas.

Ahora si podemos ver mucho mejor que esta pasando. Sin embargo, las oportunidades de las

marcas especificadas apenas rozan las 400 oportunidades (por sobre mas de 16000), dejando al descubierto una gran conclusión: Las marcas no especificadas dominan las oportunidades.

### 6.3. Análisis de Marcas Especificadas

En el momento en que nos toco analizar las oportunidades con marcas especificadas, nos toparamos con una gran sorpresa: absolutamente todas las oportunidades que contenían una marca especificada culminaron en 'Closed Lost'. A raíz de esto, como queremos recuperar algo de información, vamos a ver cual fue la peor marca: la que mas oportunidades tuvo y perdió. A continuación, el gráfico.



Figura 32: Cantidad de oportunidades perdidas de las marcas.

Tenemos entonces, que de todas las marcas, la marca 'Brand 9' fue la que más oportunidades perdió. Seguramente, en una eventual oportunidad nueva, podremos casi confirmar que si tiene marca especificada, las posibilidades de que se llegue a un acuerdo, son pocas.

### 6.4. Análisis de Marcas No Especificadas

Ahora analizaremos las marcas Sin Especificar, pero para eso debemos partir sobre la siguiente información que detectamos:

- Las oportunidades de la marca 'Other' fueron todas 'Closed Lost'
- Todas las oportunidades que terminaron en 'Close Won', son de marca Sin Especificar.
- No todas las oportunidades de marca Sin Especificar terminaron en 'Closed Won'.

Con esta información mostramos la cantidad de Lost y Won para la marca Sin Especificar.



Figura 33: Stages de la Marca Sin Especificar.

Y por ultimo queremos ver que pasa con las oportunidades que no son ni Lost ni Won, que como se imaginaran a este punto, son todas de marca Sin Especificar.

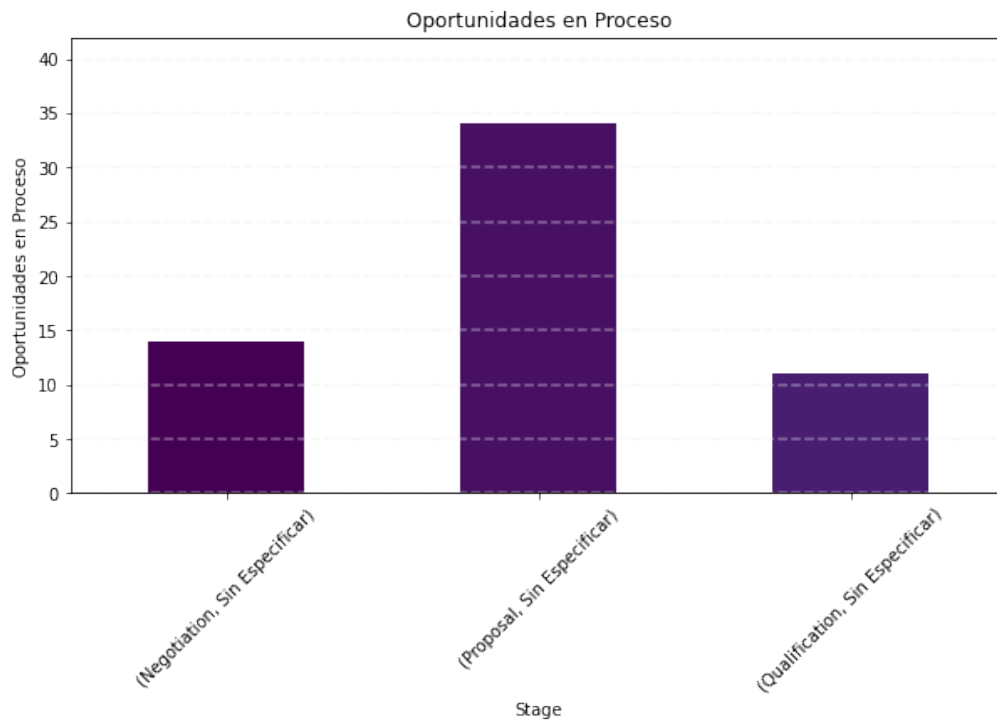


Figura 34: Oportunidades en Proceso.

## 6.5. Marcas: Conclusión

Luego del arduo análisis con respecto a las marcas de los productos en nuestras oportunidades, podemos observar que el mercado de las oportunidades ganadas, esta conquistado en su totalidad por la marca Sin Especificar. Debido a esto, concluimos que si una oportunidad tiene esta marca entre sus datos, las probabilidades de que esta triunfe son altísimas. Por otro lado e inversamente proporcional, si la marca es otra que Sin Especificar, lo mas probable es que la venta falle.

## 7. Análisis de Divisas

### 7.1. Incidencia de Divisas

Un aspecto que nos pareció interesante analizar fue el uso de las divisas de nuestras oportunidades. Al haber oportunidades alrededor de todo mundo, es normal que diferentes tipos de monedas sean usadas. A continuación, un gráfico que nos muestra la incidencia de distintas monedas en nuestras oportunidades.

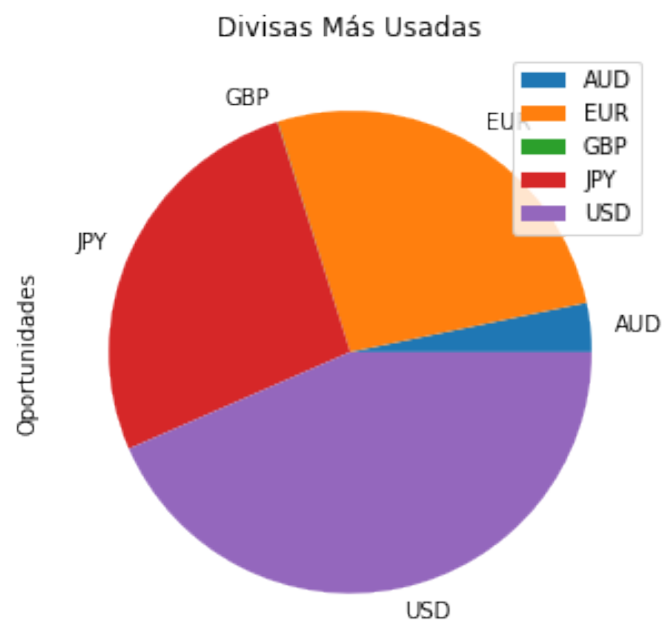


Figura 35: Incidencia de divisas.

Como vemos, las mas usadas son los dolares, los euros y los yenes. Esto está estrictamente relacionado con la incidencia de las regiones. En lo que sigue, haremos un breve análisis de las oportunidades de cada divisa.

### 7.2. Divisas: Dólar

A continuación, los resultados finales de las oportunidades que usaron como moneda principal el dólar.



Figura 36: Stages del Dólar.

Observamos que perdió mas de las que gano, sin embargo, la diferencia es muy poca, al mismo tiempo las oportunidades son una cantidad considerable.

### 7.3. Divisas: Yen

Ahora veremos los resultados finales de las oportunidades que usaron como moneda principal los yenes.



Figura 37: Stages del Yen.

Aquí vemos que ganó mas de las que perdió, la diferencia es muy grande, al mismo tiempo las

oportunidades son menos a las del dolar.

#### 7.4. Divisas: Euro

Por último, veremos como le fue a las oportunidades en euros.

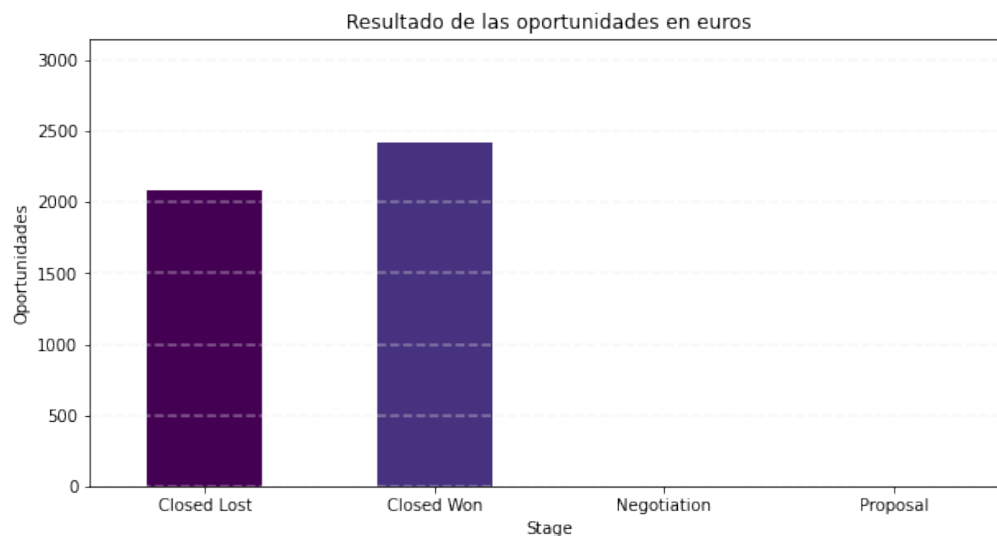


Figura 38: Stages del Euro.

Observamos que ganó mas de las que perdió, sin embargo, la diferencia es un poco mas considerable que el dolar pero menos que el yen, al mismo tiempo las oportunidades son una cantidad menor al dolar y al yen.

#### 7.5. Divisas: Conclusión

Al analizar los resultados de las divisas mas usadas, obtenemos información interesante. Si bien el dolar estadounidense es el mas usado, este perdió mas oportunidades de las que ganó. A su vez, el euro y el yen tienen una cantidad de oportunidades muy parecida. El euro y el yen ganaron mas de las que perdieron. Esto nos genera una relación bastante extraña. La conclusión a la que llegamos es que necesitamos mas parámetros de comparación, solo la divisa no nos da tanta información.



## 8. Análisis de Precios Convertidos

### 8.1. Análisis general de precios convertidos

Al ya haber visto que se usan diferentes divisas para las oportunidades, tal vez sea mejor convertir todo en una misma moneda para hacer un análisis mas exacto de nuestro set de datos. Así es como buscamos las conversiones y obtuvimos los precios en una misma moneda de las oportunidades. En este apartado se busca tener un vistazo general de las oportunidades en base a su precio universal. Arbitrariamente, separamos las oportunidades en base a su precio, veamos sus resultados.

#### 8.1.1. Análisis de oportunidades de mayor precio

Decidimos que nuestro precio mínimo para ser considerado mayor precio es de 100000.

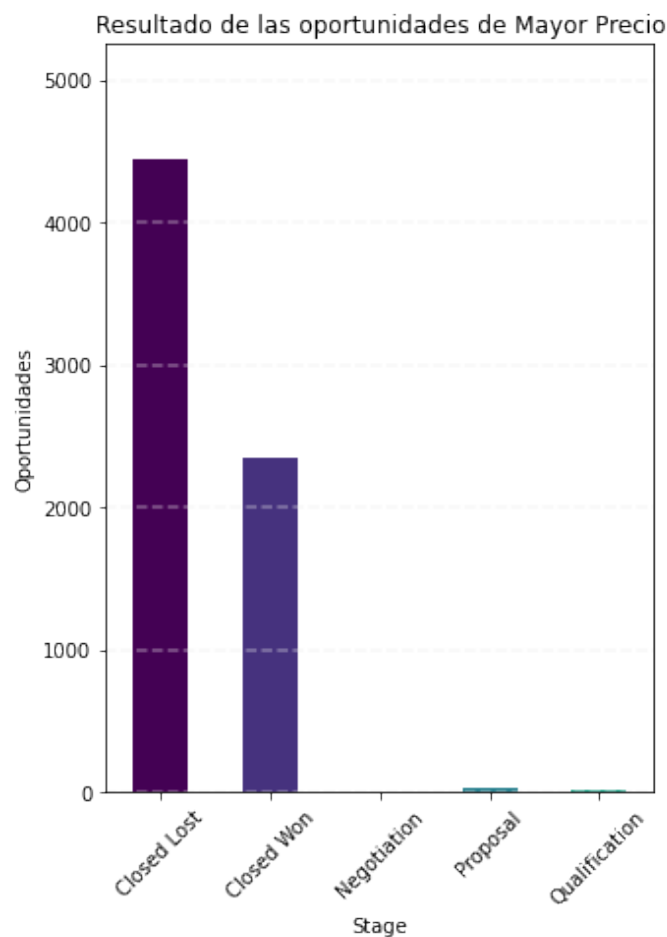


Figura 39: Incidencia de marcas por región.

Observamos que casi todas las oportunidades son Lost o Won. Vemos que las perdidas casi

doblan la cantidad de las ganadas.

### 8.1.2. Análisis de oportunidades de menor precio

Decidimos que nuestro precio máximo para ser considerado menor precio es de 100.

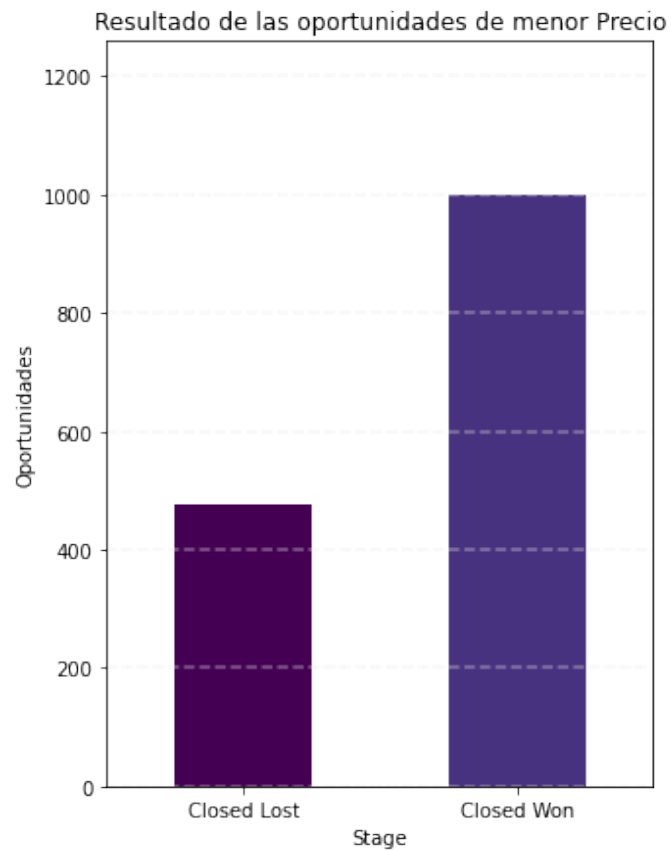


Figura 40: Incidencia de marcas por región.

Vemos que las ganadas casi doblan la cantidad de las pérdidas. Ambos análisis nos sirven para una conclusión muy sólida: Mientras mas bajo sea el precio universal, mayores posibilidades hay de que triunfe.

## 8.2. Precio Universal: Marcas

Ahora queremos poner en practica este concepto de 'Precio Universal' con un ejemplo menos abstracto. Haremos un análisis de las marcas mas concurridas en las oportunidades de mayor y menor precio.

### 8.2.1. Marcas: Mayor Precio Universal

Notamos que las marcas con mayor concurrencia en las oportunidades de mayor precio son: 'Sin Especificar', 'Otro', 'Brand 9', 'Brand 24', 'Brand 5'. Analicemos, en general, el desempeño en conjunto de estas.

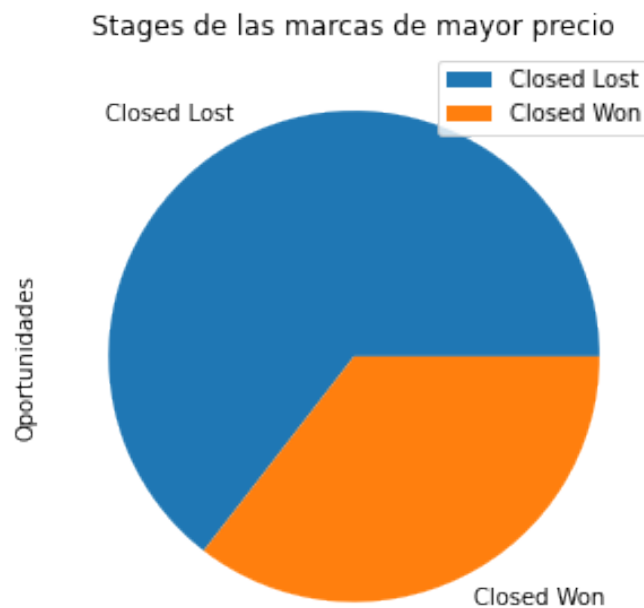


Figura 41: Stages de las marcas de mayor precio.

Como esperábamos y como ha sucedido en el análisis general, las oportunidades perdidas dominan las estadísticas.

### 8.2.2. Marcas: Menor Precio Universal

Notamos que las marcas con mayor concurrencia en las oportunidades de menor precio son solo 'Sin Especificar' y 'Otro'. Analicemos, en general, el desempeño en conjunto de estas.

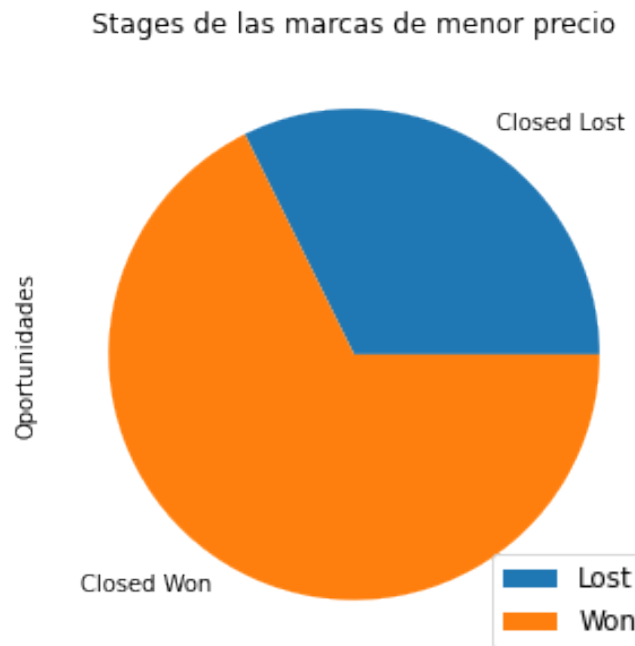


Figura 42: Stages de las marcas de menor precio.

Nuevamente concuerda con el análisis general, las oportunidades ganadas dominan las estadísticas. La relación entre las de mayor y menor precio por marca, es casi proporcional.

### 8.3. Precio Universal: Productos

Continuando con la ejemplificación del Precio Universal, haremos un análisis de los productos mas concurridos en las oportunidades de mayor y menor precio.

#### 8.3.1. Productos: Mayor Precio Universal

Notamos que los productos con mayor concurrencia en las oportunidades de mayor precio son Product Name 220, 226, 100, 96, 448. Analicemos, en general, el desempeño en conjunto de estos.

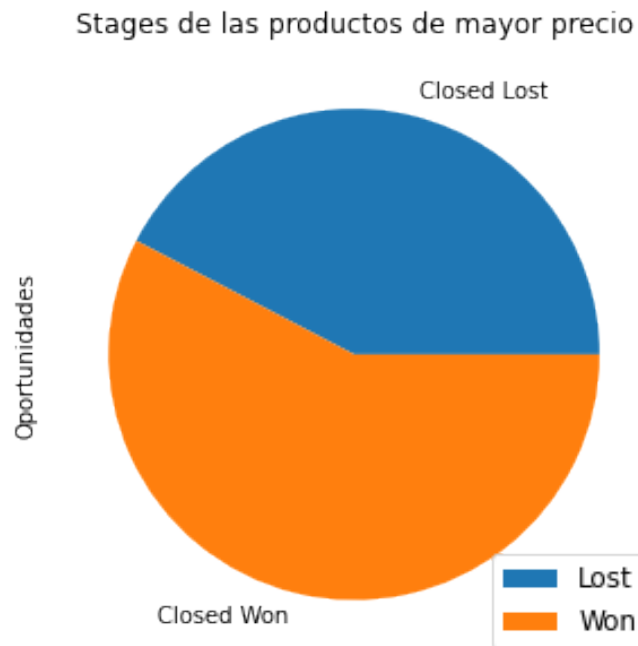


Figura 43: Stages de los productos de mayor precio.

Sorprendentemente hay mas ganados que perdidos, sin embargo la diferencia no es tan abismal como antes. Veamos si podemos deducir algo terminando con este pequeño análisis.

### 8.3.2. Productos: Menor Precio Universal

Notamos que los productos con mayor concurrencia en las oportunidades de menor precio son Product Name 491, 4, 20, 0, 17. Analicemos, en general, el desempeño en conjunto de estos.

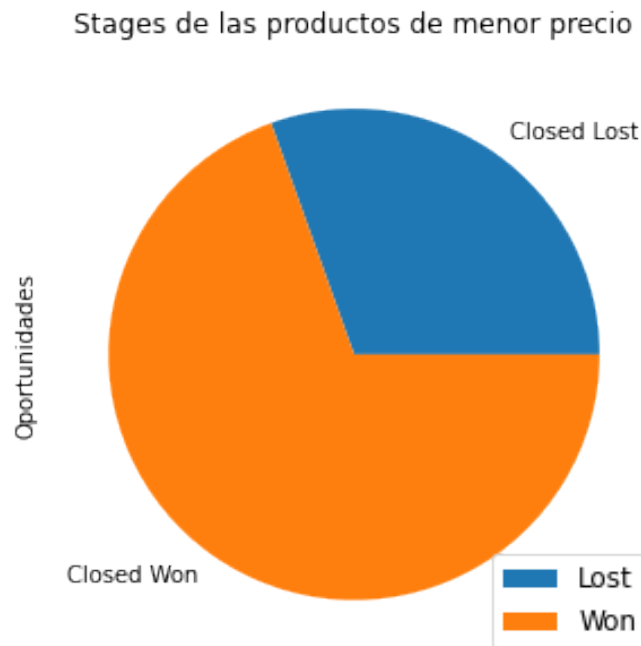


Figura 44: Stages de los productos de menor precio.

Aquí vemos como se cumple la relación original, los de menor precio vendieron mucho mas y perdieron mucho menos. Aunque el anterior no cumplió, parecemos estar en condiciones para dar una conclusión consciza.

#### 8.4. Precio Universal: Conclusión

Una vez terminado y ejemplificado nuestro análisis de precios en una divisa universal, la conclusión es muy clara: Mientras más baratas sean nuestras oportunidades, mas probabilidades tienen de ser un Closed Won. Esto se ve reflejado en los estudios de caso con respecto a las marcas y los productos, por eso vemos que la relación es muy parecida.

## 9. Análisis de Permisos

### 9.1. Permisos

Existen dos tipos de permiso: Permiso de Envío y Permiso Burocrático. A su vez, cada oportunidad puede necesitar o no estas aprobaciones, así como también pueden obtenerlas o no. Nos pareció importante analizar esta sección para ver como influían en los resultado finales. Procedemos a analizar las aprobaciones por separado.

#### 9.1.1. Aprobación de Envío

Veremos como les fue a las oportunidades, basándonos en las aprobaciones de envío.

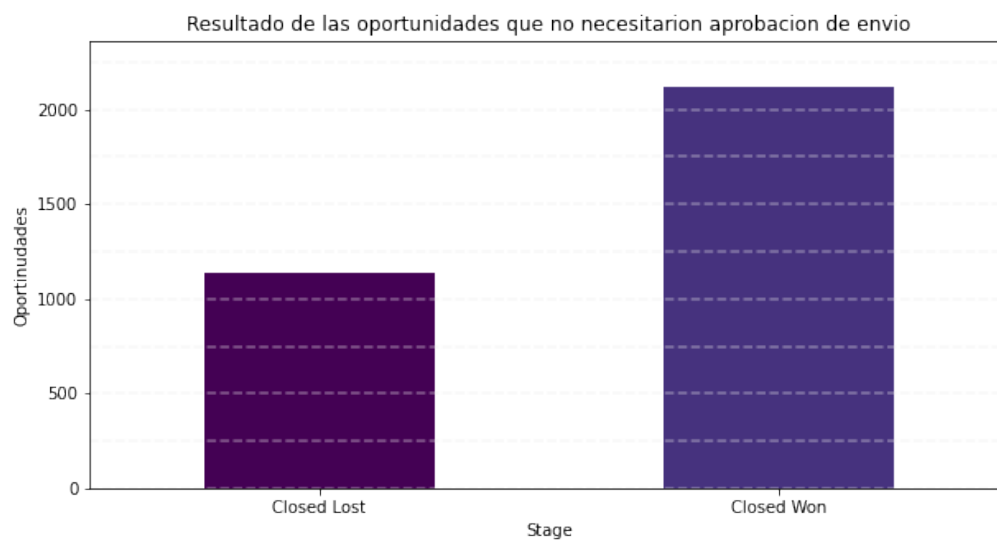


Figura 45: Oportunidades sin permisos de envío.

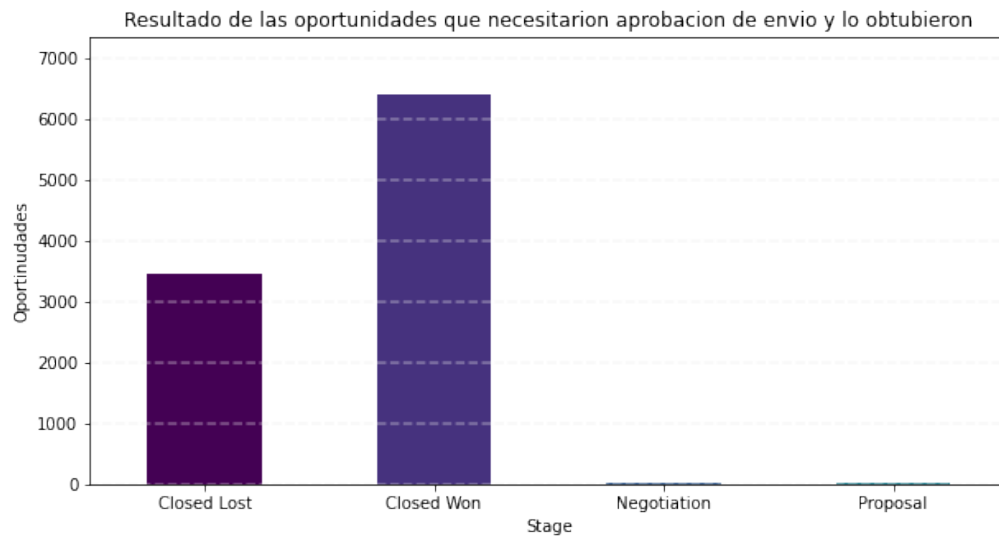


Figura 46: Oportunidades con permisos de envío aprobado.

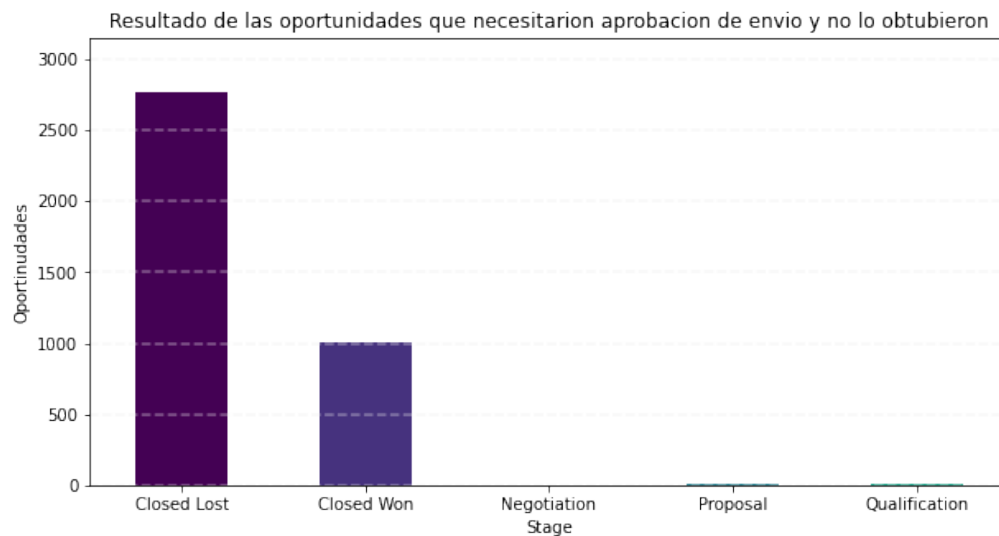


Figura 47: Oportunidades con permisos de envío no aprobado.

En lo que respecta a oportunidades sin necesidad de aprobación de envío, hay una considerable diferencia entre ganados y perdidos, vemos un saldo positivo. En los que necesitaron aprobación la situación cambia. Si lo obtienen, como lógicamente se esperarí, la cantidad de ganados sera mayor, así como también vemos que los que no lo obtuvieron hasta tienen más perdidos que ganados. Una breve conclusión sería decir que si no necesita aprobación de envío, le irá relativamente bien, pero si la necesita, es necesario que lo obtenga para triunfar.



### 9.1.2. Aprobación Burocrática

Veremos como les fue a las oportunidades, basándonos en las aprobaciones burocráticas.

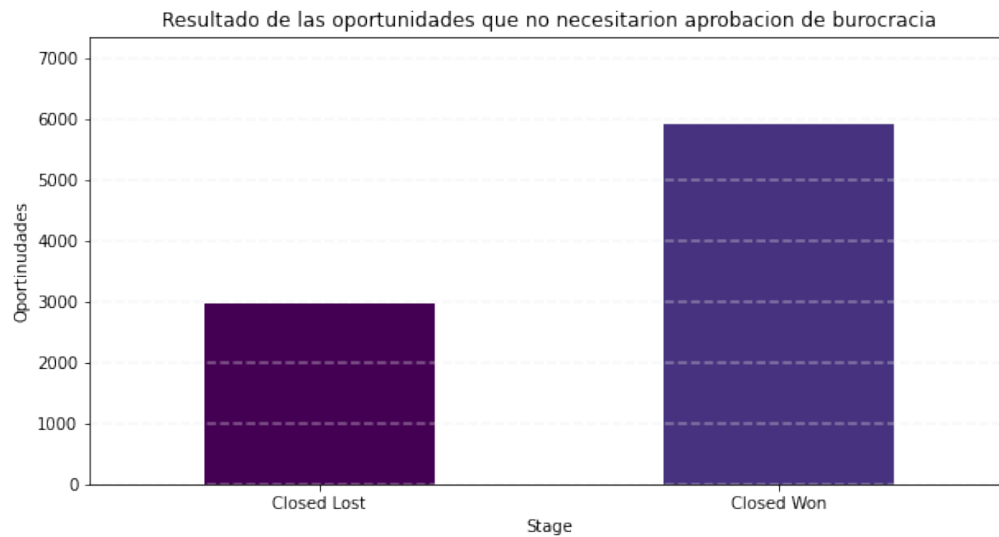


Figura 48: Oportunidades sin permisos burocráticos.

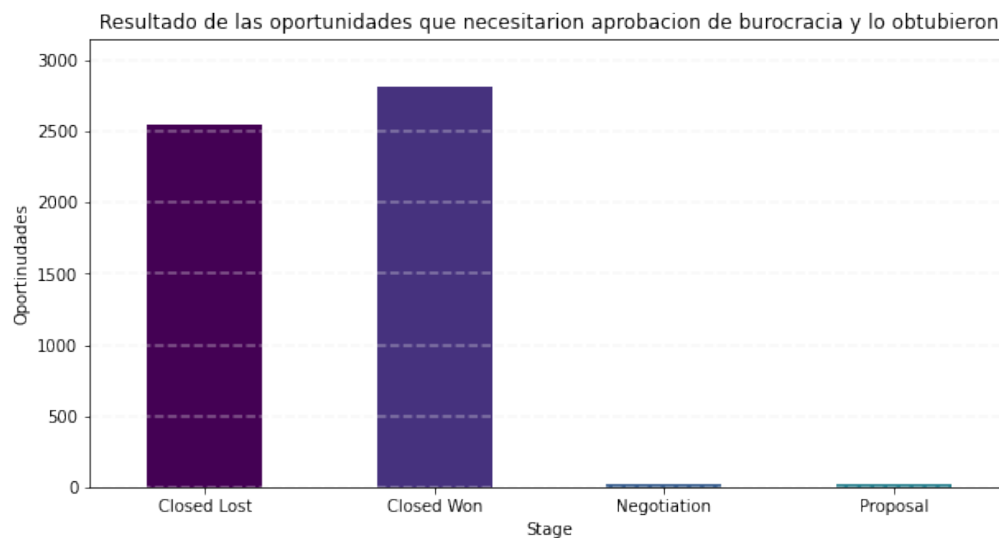


Figura 49: Oportunidades con permisos burocráticos aprobados.

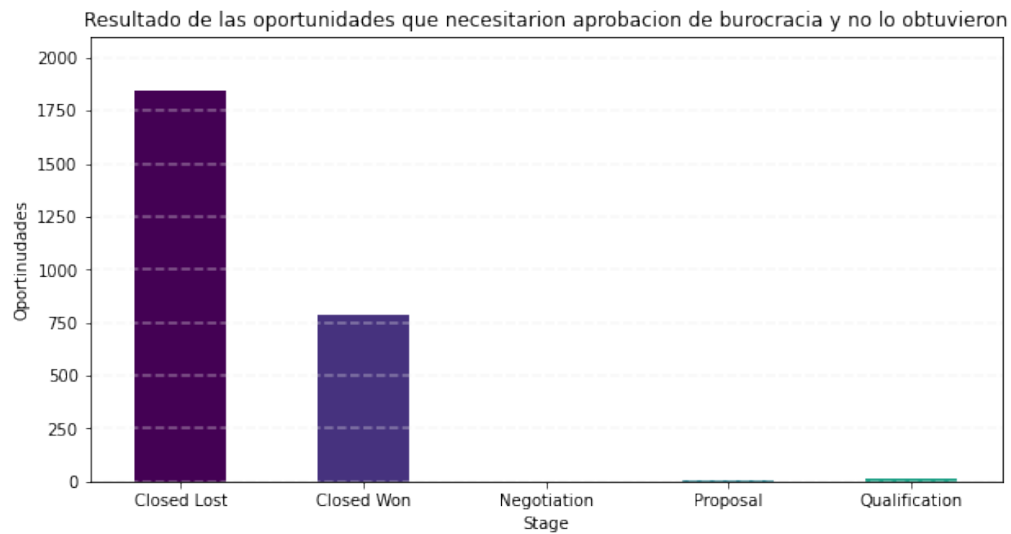


Figura 50: Oportunidades con permisos burocráticos no aprobado.

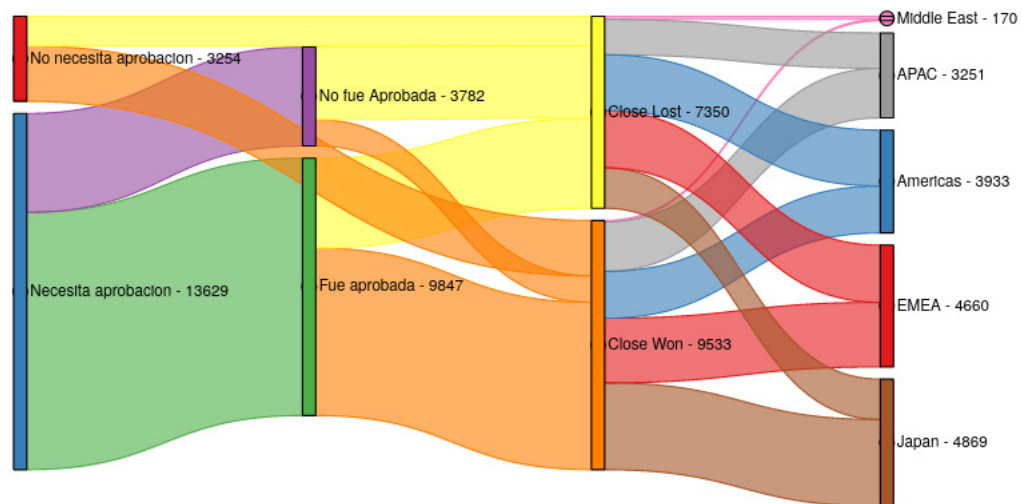
En las oportunidades sin necesidad de aprobación burocrática vemos una considerable diferencia entre ganados y perdidos, dominando las oportunidades triunfantes. En los que necesitaron, si lo obtienen la cantidad de ganados sera mayor ( igual al de envío), pero no tanto, así como también vemos que los que no lo obtuvieron hasta tienen más perdidos que ganados. La conclusión sería muy parecida a la de los envíos. Si no necesita aprobación burocrática, le irá relativamente bien, no tanto como el anterior, pero si la necesita, es necesario que lo obtenga para triunfar.

## 9.2. Permisos de las Oportunidades por Región

Creemos que los Stages son un indicador muy importante para analizar los datos. Por eso ahora vamos a analizar los distintos estados de las oportunidades según la región pero considerando los permisos de los mismos.

Para esto se utilizo un Sankey chart, donde se puede ver bien las distintas transiciones de las oportunidades. Nuevamente no se han tenido en cuenta los Stages *Proposal*, *Negotiation* y *Qualification* por la misma razón de que son pocos datos para este análisis.

### Estados de las oportunidades por region considerando permisos



Podemos ver que una gran cantidad de oportunidades que necesitaban aprobación y no la obtuvieron, terminaron en fracaso. En cambio, las que sí fueron aprobadas, en su mayoría concretaron. También vemos que en las que no necesitaban aprobación, a pesar de no ser muchas, la mayoría (casi el doble) concretaron el trato.

## 9.3. Permisos: Juntando todo

Ahora nos interesa hacer un breve análisis de las aprobaciones combinadas. Partimos del envío aprobado, ya que esa característica fue la mas positiva con respecto al envío. Con esta base, analizamos la parte burocrática.

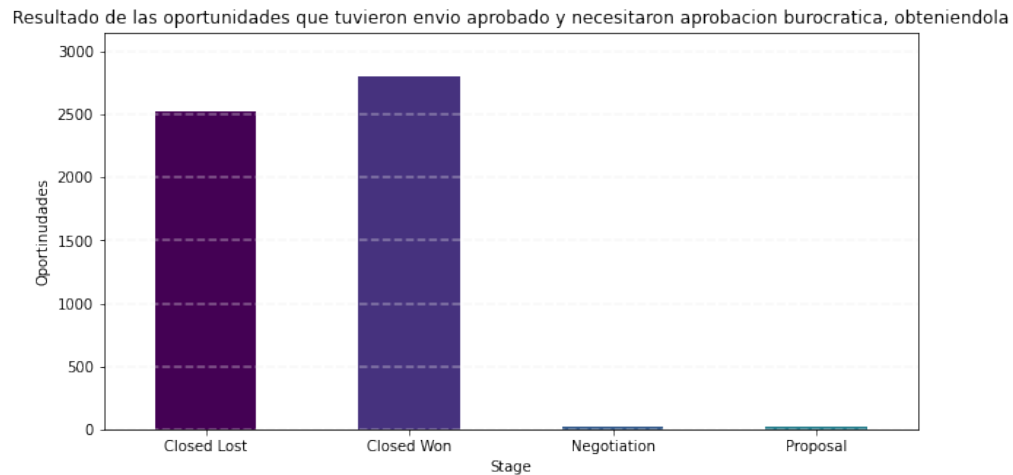


Figura 51: Oportunidades con envío y permisos burocráticos.

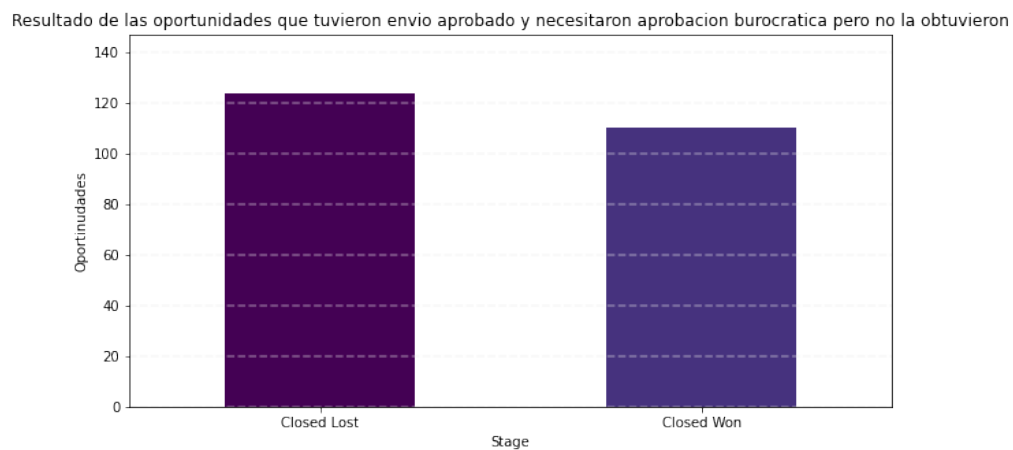


Figura 52: Oportunidades con envío pero sin permisos burocráticos.

Nuevamente vemos un resultado lógico, pero con cantidades no esperadas: cuando tienen ambas aprobaciones, se destacan los ganados pero no por tanto y cuando no tiene el permiso burocrático, se destacan los perdidos pero no por tanto. Evidentemente, el permiso de envío tiene más peso que el burocrático.

#### 9.4. Permisos: Conclusión

Como dedujimos antes, vemos que la aprobación de envío pesa mas que la burocrática. Debido a esto, si la oportunidad necesita aprobación de envío y la tiene, probablemente sea una venta exitosa. Sumando la aprobación burocrática, si la tiene, la oportunidad sumara probabilidades de triunfar. En lo que respecta a las que no necesitan aprobaciones, las posibilidades son balanceadas

y tal vez necesitemos mas parámetros para determinar su destino. Por otro lado, si tiene permisos rechazados, la oportunidad tiene altas probabilidades de fracasar.

## 10. Conclusión

Englobando, llegamos a deducir varias razones por las cuales una oportunidad cierra exitosamente.

Entre ellas está si la oportunidad necesita **permisos**, si lo necesita y recibe aprobación, tanto para los permisos de **envío** (9.1.1) como los **burocráticos**(9.1.2), aumenta sus probabilidades de cerrar exitosamente, de lo contrario, ocurrirá lo opuesto. Si no necesita aprobación, su porcentaje también se mantendrá del lado del éxito, pero en menor proporción.

Una conclusión interesante es que el *Total Amount* es inversamente proporcional al cierre exitoso de las oportunidades (si el *Total Amount* es muy alto menos probabilidad de éxito tendrá). Esto es lo que le pasa al territorio de **Australia** (4.4)

Otra conclusión es que si surge una nueva oportunidad el Territorio o Región de donde proviene probablemente sea de *Japón* (en ambos casos) y si proviene de ahí, su probabilidad de éxito será buena. Por lo contrario si la oportunidad proviene de la región *Americas* y/o de los territorios NE America, No especificado o India su probabilidad de éxito será mala. Cabe aclarar que solo este indicador no define el éxito o no de una oportunidad pero si afecta a su probabilidad y se debe tener en cuenta.

Observando los productos nos encontramos que en el mercado hay pocos que cuentan con datos especificados, especialmente en las categorías y los tipos, la mayoría de los productos no tienen ninguno de estos campos especificados. Aquellos datos que si cuentan con información detallada, son opacados por aquellos que no, por la gran diferencia en cuanto a las cantidades que se encuentran dentro del mercado. Dentro de las categorías, en la **Categoría A** existe un monopolio, y en la **Categoría B** se puede ver claramente como las no especificadas están por encima de todas las demás. Pero cuando hablamos de **Familias de Productos** contamos con información detallada, por lo tanto lo consideramos una forma de filtrar en cuanto a búsqueda, ya que no encontramos datos nulos que predominen y alteren esta forma de búsqueda.

Por otro lado, en las marcas llegamos a la siguiente conclusión: Aquellas sin especificar arrasan las estadísticas, con un 100 por ciento de efectividad y una amplia mayoría en cuanto la incidencia. Si su marca es Sin Especificar, es muy probable que se concrete.

Sobre las divisas no hay un claro ganador, vemos una gran variedad de resultados a medida que cambian las monedas.

Gracias a este arduo análisis, y teniendo en cuenta la cantidad de campos cubiertos en este, creemos estar en condiciones de saber predecir, con cierto margen de error, como resultará una eventual nueva oportunidad.

## 11. Github

Dejamos el enlace al Github del TP1. <https://github.com/BautistaXifro/Datos>