



UNIVERSIDAD DE BUENOS AIRES

FACULTAD DE INGENIERÍA

## 75.06 - ORGANIZACIÓN DE DATOS

TRABAJO PRACTICO 1

Segundo cuatrimestre de 2020

Chávez Cabanillas, José E.	96467	jchavez@fi.uba.ar
Tadini, Santiago	104439	stadini@fi.uba.ar
Fedele Edo, María Cecilia	102573	mfedele@fi.uba.ar
Hojman de la Rosa, Joaquin Guido	102264	jhojman@fi.uba.ar

# Índice

<b>1. Introducción</b>	<b>3</b>
1.1. Objetivo . . . . .	3
1.2. Repositorio . . . . .	3
1.3. Análisis inicial de los datos . . . . .	3
1.4. Introducción al historial de ventas de la empresa . . . . .	4
1.5. Precio y estado, ¿Asunto separado? . . . . .	7
<b>2. Region</b>	<b>11</b>
2.1. Identificación de las regiones . . . . .	11
2.2. Éxitos y fracasos . . . . .	12
2.3. Japón . . . . .	13
2.4. Dinero ganado y perdido . . . . .	14
2.5. Impacto del clima . . . . .	15
<b>3. Vendedores</b>	<b>18</b>
3.1. Éxitos y fracasos de los vendedores . . . . .	18
3.2. Dinero recaudado por los vendedores . . . . .	20
3.3. Conclusiones ¿Quién es el mejor vendedor? . . . . .	24
3.4. Premios consuelo: Mejores vendedores por región . . . . .	25
3.5. Apéndice infográfico . . . . .	28
<b>4. Clientes</b>	<b>30</b>
4.1. Clientes de la empresa . . . . .	30
4.2. ¿Son los clientes que más compraron aquellos que más gastaron? . . . . .	30
4.3. Conclusiones, menciones, agradecimientos y algo más . . . . .	31
<b>5. Familia de Productos</b>	<b>33</b>
5.1. Análisis Global . . . . .	33
5.1.1. Ganancias en Ventas . . . . .	33
5.1.2. Cantidades de Ventas . . . . .	34
5.2. Análisis por Región . . . . .	35
5.2.1. Ganancias en Ventas . . . . .	36
5.2.2. Cantidades de Ventas . . . . .	39
5.3. Conclusiones . . . . .	43
<b>6. Aprobación</b>	<b>44</b>
6.1. ¿Cuántas oportunidades necesitan aprobación? . . . . .	44
6.2. Oportunidades que no necesitan aprobación . . . . .	45
6.2.1. Análisis por región . . . . .	46
6.2.2. Análisis por precio . . . . .	46
6.3. Oportunidades que necesitan aprobación . . . . .	47
6.3.1. Análisis por región . . . . .	49
6.3.2. Análisis por precio . . . . .	51

<b>7. Potencia</b>	<b>52</b>
7.1. Análisis inicial de los datos . . . . .	52
7.2. Si el producto es más caro, ¿es mas potente? . . . . .	52
7.3. ¿Cuan potentes son los productos potentes? . . . . .	53
7.4. Conclusiones finales . . . . .	55
<b>8. Source</b>	<b>57</b>
8.1. ¿Qué son las Sources? . . . . .	57
8.2. Sources Declaradas . . . . .	57
8.2.1. Contra cantidad de oportunidades . . . . .	57
8.2.2. Contra Stage . . . . .	58
8.2.3. Conta el Dinero aportado . . . . .	60
8.2.4. Uniendo todo . . . . .	61
<b>9. Productos</b>	<b>63</b>
9.1. Productos más vendidos . . . . .	63
9.2. Los productos más vendidos, ¿son los más exitosos? . . . . .	64
<b>10.Comentarios finales</b>	<b>66</b>

# 1. Introducción

Les presentamos a la empresa “Frío Frío”, dedicada a la venta e instalación de equipos de aire acondicionado para grandes superficies. Al ser una empresa B2B (“Business To Business”), es esencial para ellos optimizar los esfuerzos de los representantes comerciales, ayudándolos a priorizar las oportunidades en el pipeline.

Una “oportunidad” consiste en un proyecto de venta o instalación de equipos para un cliente. La venta se estructura alrededor de TRF (Toneladas de refrigeración) y puede estar compuesta por varios productos distintos. El “pipeline” hace referencia al flujo de oportunidades prospecto que la empresa está desarrollando. El equipo comercial asigna a distintos momentos, para cada oportunidad, un estado en la negociación.

El dataset cuenta con información de cada oportunidad, como por ejemplo información sobre el vendedor a cargo de la venta, información geográfica de los clientes, TRF pedidas, fecha prevista de entrega de los equipos, etc.

## 1.1. Objetivo

El TP consiste en realizar un análisis exploratorio de los datos provistos con el objetivo de determinar características y variables importantes, descubrir insights interesantes, y analizar la estructura de los mismos.

A lo largo del presente informe se darán una serie de tópicos generales los cuales engloban los análisis y conclusiones que fuimos extrayendo de los datos presentados.

## 1.2. Repositorio

El código generado para ayudarnos a realizar el análisis de los datos y el presente informe puede encontrarse en:

*[github.com/joaquinhoman/Entrenamiento\\_ECI\\_2020](https://github.com/joaquinhoman/Entrenamiento_ECI_2020)*

## 1.3. Análisis inicial de los datos

Comenzamos nuestro análisis de datos contando la cantidad de filas y columnas, las cuales son 16.947 y 41, respectivamente, y notando algunas características del set, como por ejemplo que pesa 3,5 MB, o que casi todas (más del 99 por ciento) de las ventas que aparecen en la base de datos ya finalizaron, ya sea como éxito o como fracaso.

Luego lo que hicimos fue limpiar ciertas columnas que necesitaban retoques. De esta manera casteamos todas las columnas al tipo de dato indicado, e incluso modificamos globalmente ciertos valores de tipo string. Creamos una nueva columna llamada USD Total Amount que contiene todos los precios en dólares. Tuvimos que eliminar algunas columnas por diversos motivos, estas son:

- Submitted for Approval (Un solo valor = 0)
- Brand (Muchos nulos = 15911)
- Product Type (Muchos nulos = 15935)

- Size (Muchos nulos = 15967)
- Product Category B (Muchos nulos = 15928)
- Price (Muchos nulos = 15982)
- Currency (Muchos nulos = 16052)
- Last Activity (Todos nulos)
- ASP (converted) Currency (Un solo valor = USD)
- Actual Delivery Date (Todos nulos)
- Prod Category A (Un solo valor = "Prod Category A None")

Antes de analizar los que consideramos son los tópicos ejes del set, y de cualquier gran empresa que se dedique a las ventas, daremos algunos análisis "iniciales".

Primero definiremos nuestras variables mas importantes, y consideramos que ellas son:

- Monto total de venta del producto, o conocido mas coloquialmente como "Precio". Por supuesto, el objetivo principal de una empresa es vender, por lo que el precio y consecuentemente el dinero ganado es un factor clave en cualquier análisis.
- Estado de la venta: Se refiere a si la venta ya terminó (y en ese caso si fue un éxito, o un fracaso) o si aun están en alguna fase previa de negociación. Casi todas las ventas del set ya concluyeron, por lo que casi siempre obviaremos las que aún no terminaron.

La gran mayoría de los análisis que incluiremos en el presente informe tratan sobre las diversas fluctuaciones de estas dos variables. Además cabe mencionar que esta empresa tiene operaciones a nivel global, por lo que en muchos casos también hablaremos de regiones, haciendo análisis particulares para cada una.

Antes de dar paso a los análisis principales sobre los que ya comentamos antes, haremos en esta sección introductoria algunos análisis previos que nos ayudaran a comprender mejor los datos con los que nos iremos encontrando más tarde.

## 1.4. Introducción al historial de ventas de la empresa

Lo primero que nos interesa saber para comenzar con nuestro análisis es como le fue en ventas a la empresa durante su existencia. La empresa comenzó a operar en el año 2016 y cuenta con operaciones hasta el año 2020. En esta subsección analizaremos los éxitos y fracasos en ventas durante ese período.

Comenzamos con una gráfica que muestra el estado de todas las ventas que ha tenido la empresa desde su creación.

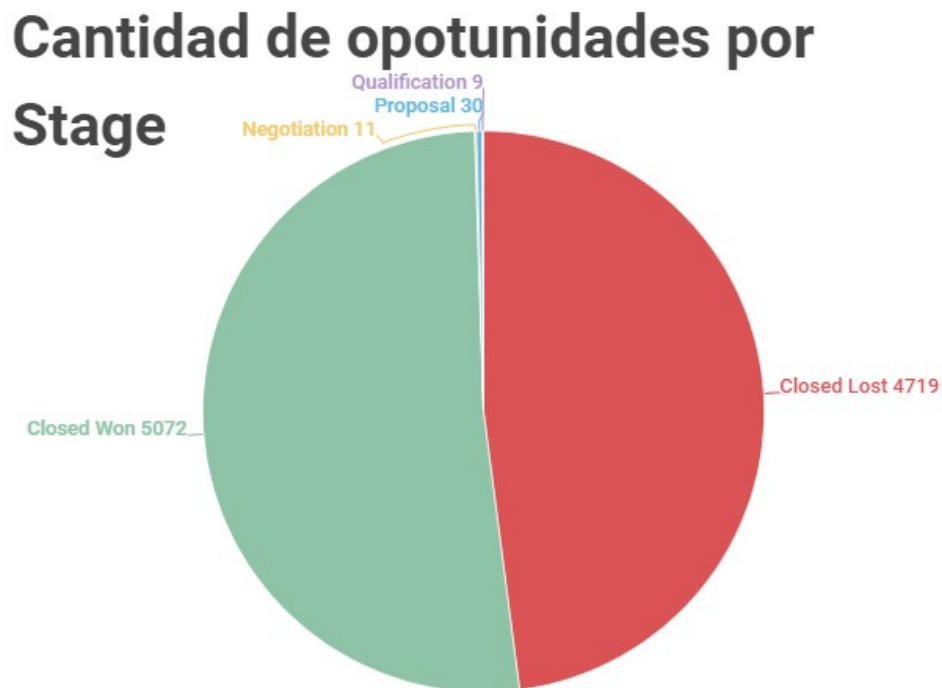


Figura 1: Cantidad de oportunidades por Stage

Podemos ver fácilmente lo comentado antes, la gran mayoría de las ventas ya finalizaron. Dentro de las mismas, notamos que un poco más del 50 por ciento tuvo un cierre exitoso, mientras que algo menos de ese valor fracasó. El número exacto de cada estado es el siguiente:

- Closed Won = 5072
- Closed Lost = 4719
- Negotiation = 11
- Proposal = 30
- Qualification = 9

Presentados los posibles estados de una venta, se mostrará un gráfico que muestra la evolución de los éxitos a través del tiempo, tomando como intervalo las fechas comentadas más arriba.

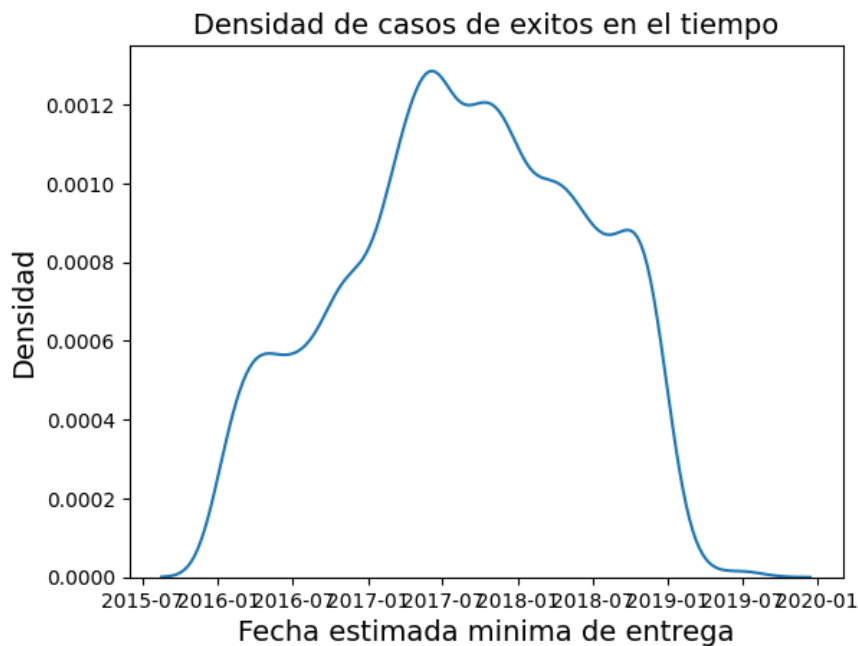


Figura 2: Densidad de casos de éxito en el tiempo

Podemos notar a simple vista que al principio de su "Historia" la empresa tenía una baja densidad de ventas, la cual fue aumentando hasta llegar a un pico aproximadamente a mitades de 2017, y luego fue descendiendo bruscamente hasta tener una densidad muy baja en el presente año 2020. Tuvo algún repunte en 2019 pero no fue suficiente y la situación actual de ventas parece ser muy mala.

Repetimos el mismo gráfico pero para ventas fallidas en función del tiempo.

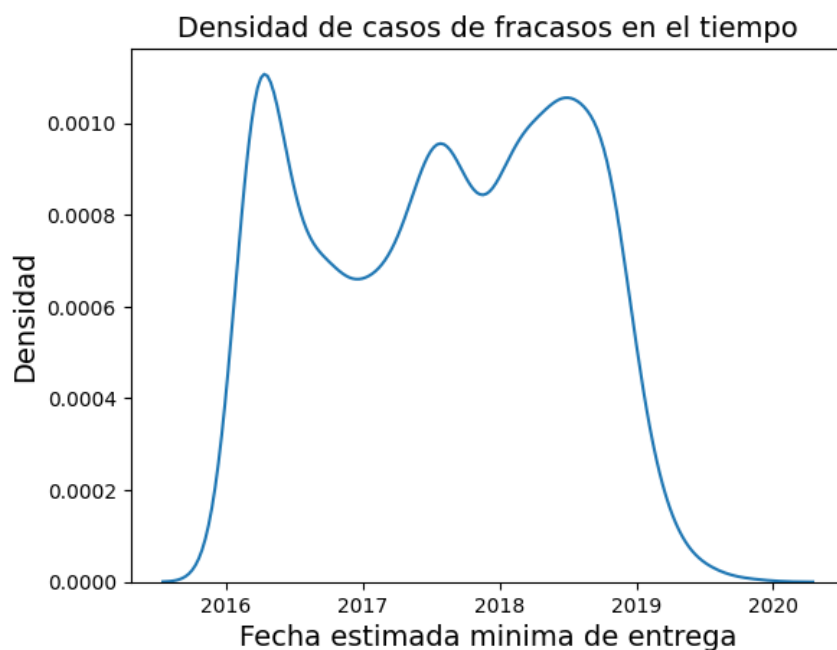


Figura 3: Densidad de casos de fracaso en el tiempo

En esta gráfica notamos que el mayor pico de fracasos fue durante el inicio de las operaciones de la empresa, luego experimentaron una baja que coincide con la mejor época de ventas (2017) y a continuación volvió a subir, hasta casi superar el pico en 2019. Luego en el corriente año 2020 vemos que casi no hubo fracasos.

Si miramos los dos gráficos vemos que parecen ser, en cierto modo, complementarios. Es decir, cuando hubo muchos éxitos, hubo pocos fracasos, y cuando hubo muchos fracasos hubo pocos éxitos.

Es de destacar que en 2020 no parece haber nada de ninguno de los estados, por lo que podríamos inferir que actualmente la empresa no está iniciando operaciones, o está iniciando muy pocas.

En una gráfica más podemos ver los dos casos anteriores superpuestos, y vemos como parecen alternar entre épocas de muchos éxitos y épocas de muchos fracasos. Podemos ver los éxitos en azul, y los fracasos en rojo.

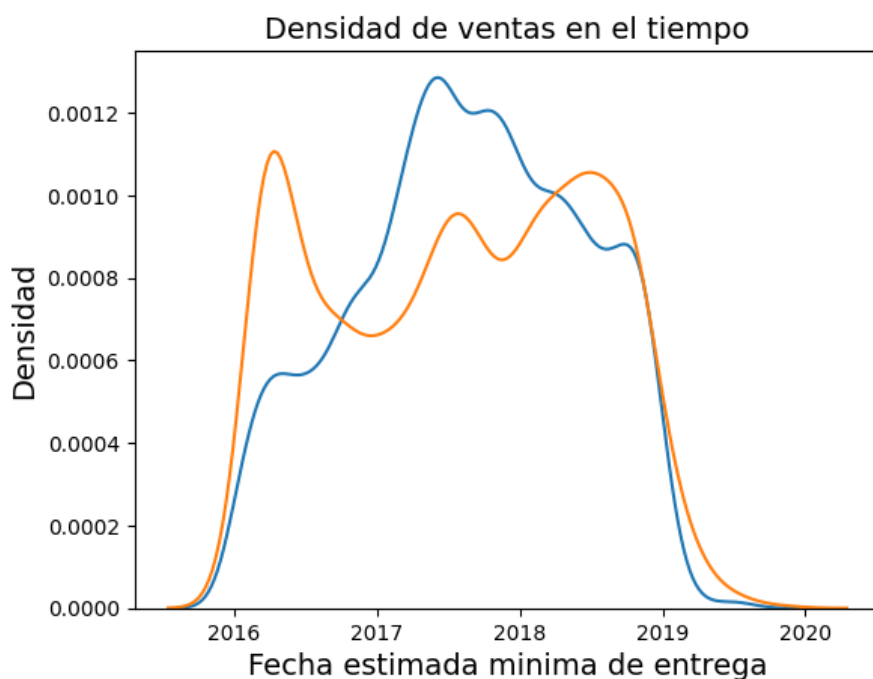


Figura 4: Densidad de ventas en el tiempo

### 1.5. Precio y estado, ¿Asunto separado?

Para concluir esta primera sección introductoria y volcarnos de lleno al análisis del set de datos, queremos saber la respuesta a la primera pregunta que podríamos hacernos. ¿El precio de la venta está relacionado con el estado final de la negociación? ¿Cuan relacionados están? Nos interesa saber si el precio es un factor determinante a la hora de cerrar una venta.

La hipótesis inicial con la que nos manejamos es que el precio y el estado están relacionados, fuertemente y de forma lineal.

¿Qué quiere decir que estén relacionados? Nos referimos a que mientras más caro sea el producto, más difícil es que esa venta se haya concretado de forma exitosa. Por lo que esperamos muchas ventas exitosas con bajos precios y muchas ventas fallidas con altos precios.



Ya detallado el asunto, nos disponemos a analizar cómo se distribuyen los éxitos y los fracasos en función del precio.

En este caso analizaremos oportunidades, es decir que agruparemos ventas por oportunidad, sumando los precios de cada artículo, y trabajando con precios en dólares. Cabe destacar que este análisis no discrimina por regiones, sino que es global.

Entonces luego de una serie de operaciones logramos obtener una gráfica de dispersión donde se ven tanto las ventas exitosas como sus precios en dólares.

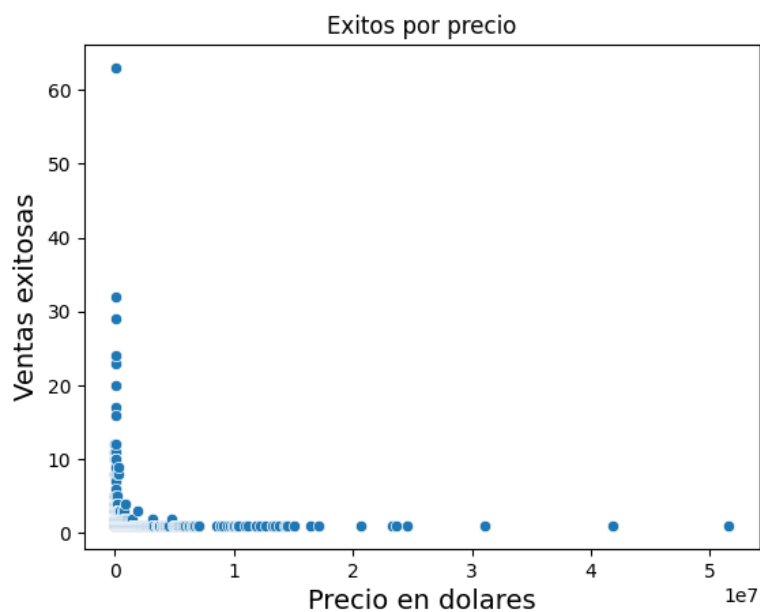


Figura 5: Dispersión de exitos y precios

Análogamente graficamos el precio en dolares y las ventas fallidas.

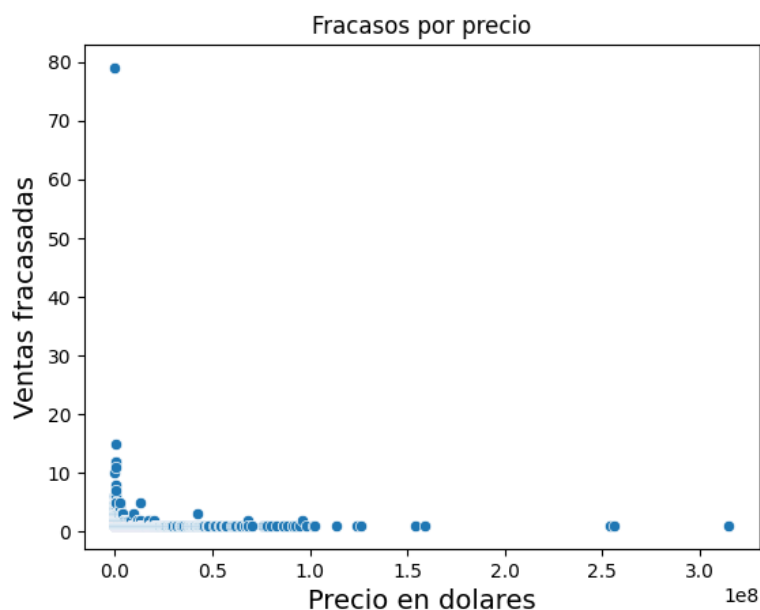


Figura 6: Dispersión de fracasos y precios

Mirando ambas graficas es fácil notar que son extremadamente similares entre ellas. Ambas parecen tener la misma cantidad de éxitos y fracasos en los precios más usuales, que suelen ser "bajos". Es verdad que hay fracasos con precios altos... pero también hay éxitos con precios altos. A priori parece ser que el estado no depende del precio, al menos no únicamente.

Para ahondar en este tema antes de descartar nuestra hipótesis, realizaremos algunos análisis más. Por ejemplo, en la siguiente figura, vemos cada una de las oportunidades como un círculo, agrupadas todas aquellas que son ganancia y así tambien las que son pérdidas. Vemos como hay más oportunidades ganadas que perdidas, tal como habíamos analizado antes.

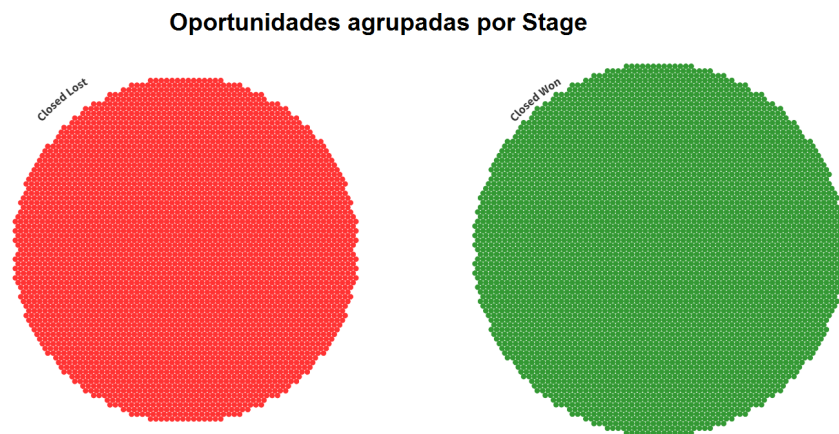


Figura 7: Oportunidades agrupadas según su estado final

Sin embargo, cuando el tamaño de cada oportunidad representada por un círculo, se modifica a un tamaño relativo a su precio, se puede ver que las oportunidades perdidas son generalmente más grandes que las ganadas, y la agrupación total de las perdidas es más de 6 veces más grande que la de las ganancias. Esto nos lleva a pensar que el precio de la oportunidad influye en su estado y respalda la hipótesis inicial que dice que mientras más cara una oportunidad, tenderá más a ser una pérdida.

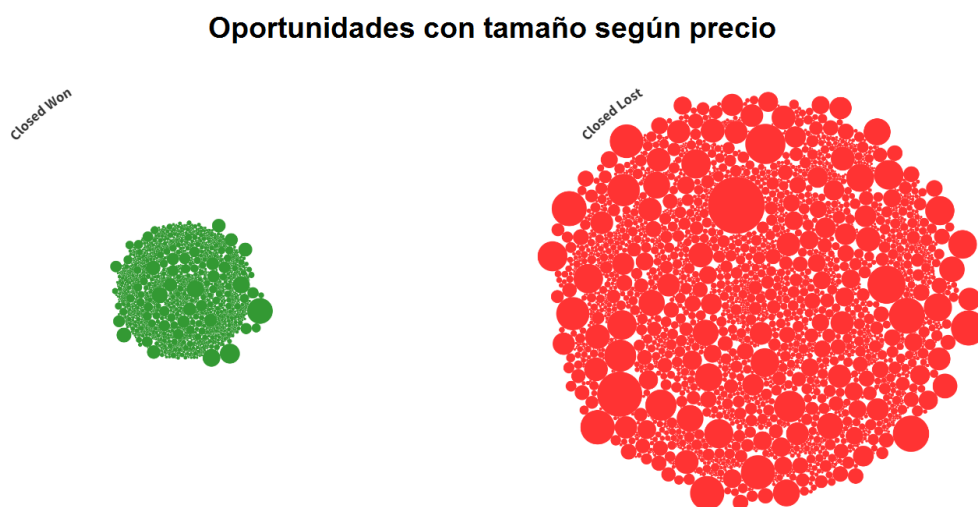


Figura 8: Oportunidades agrupadas según su estado y tamaño relativo a su precio

Finalmente, se dividió a las oportunidades por rango de precio, para tener una visión más clara de como se relaciona el precio y el estado. Medir la cantidad de ganancias en cada rango no tendría sentido, ya que la gran mayoría de las oportunidades pertenece a los rangos de menor cantidad de dinero. En este gráfico se ve como las oportunidades que menor precio tienen, son mas propensas a generar ganancias, mientras que las oportunidades con más valor, tienden a ser pérdidas. El quiebre se ve a partir del medio millón, cuando deja de ser más probable obtener una ganancia, y pasa a ser más probable obtener una pérdida.

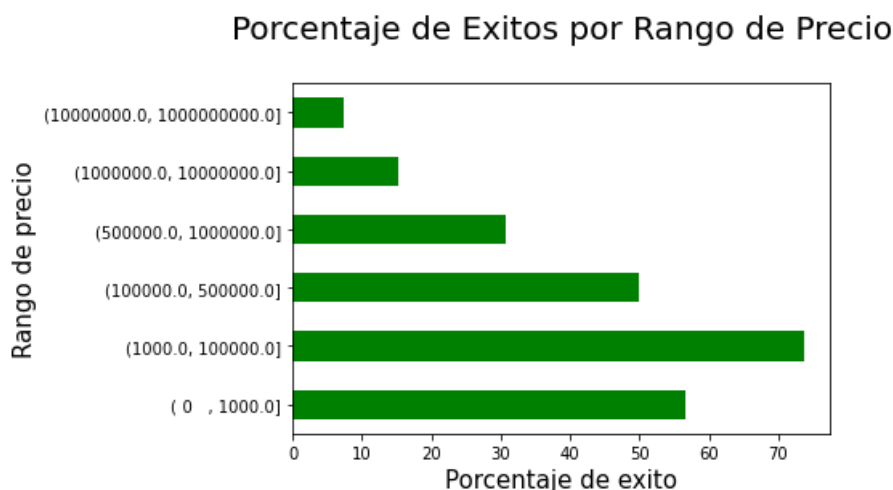


Figura 9: Porcentaje de éxito por rango de precio

Ahora sí podemos concluir que el precio y el estado están altamente relacionados. No linealmente como suponíamos al principio, ya que entre 0 y 1000 vemos un peor porcentaje que entre 1000 y 100000, pero sí con una clara tendencia a poder decir que las oportunidades más baratas tienen mayor probabilidad a ser ganancia y las oportunidades más caras tienden a ser pérdidas.

## 2. Region

### 2.1. Identificación de las regiones

Continuamos investigando el set de datos, a partir del cual observamos que tenemos cinco regiones donde la empresa tuvo operaciones, las cuales a continuación ubicaremos en el mapa. Observación: se colocaron sólo los países en los que la empresa tiene cobertura.

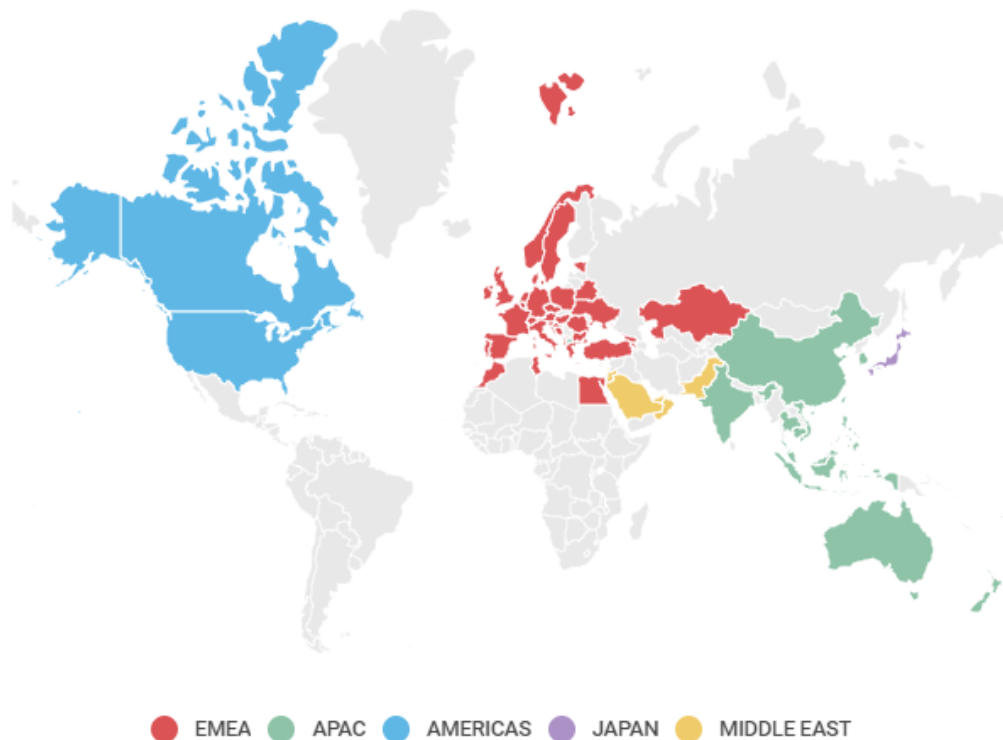


Figura 10: Map World regiones

Seguidamente, describiremos brevemente cada una:

- Americas: (América del Norte) esta región consiste principalmente en los países de Canadá y Estados Unidos.
- Japan: esta región corresponde únicamente al país de Japón.
- Middle East: traducido como Medio Oriente, los países en los que trabaja la empresa que componen esta región, mencionando los más grandes, son Arabia Saudita, Pakistán, Oman, entre otros.
- EMEA: es una designación geográfica cuyas siglas en inglés son Europa, Medio Oriente y África. Si bien el continente entero de África pertenece a esta unión, la empresa no trabaja en todos sus países.
- APAC: son las siglas de Asia-Pacífico, y abarca la parte del mundo dentro o cerca del océano Pacífico occidental.

## 2.2. Éxitos y fracasos

Analizaremos la cantidad de productos vendidos exitosamente y la cantidad de productos con oportunidad de venta pero que al final fallaron, es decir, terminaron el Closed Lost.

### Cantidad de éxitos y fracasos por región

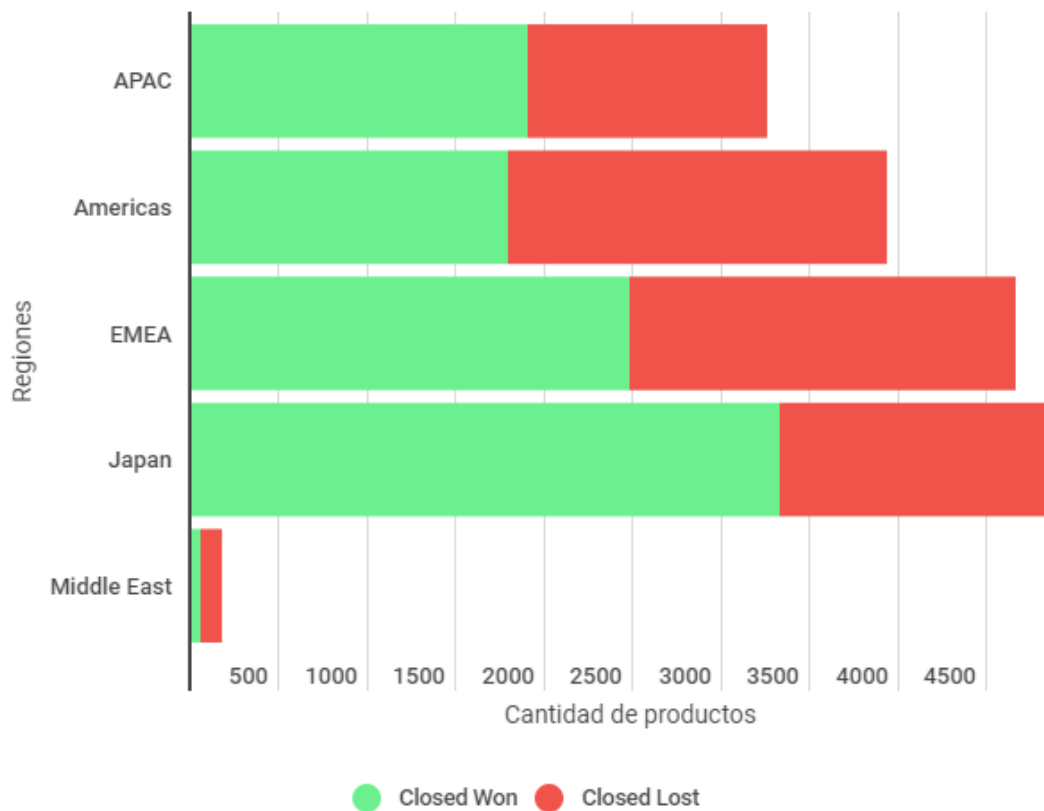


Figura 11: Cantidad de éxitos y fracasos por región

Observando el gráfico podemos hacer un primer análisis de cada región:

- Middle East: esta es la región con menos operaciones de todas, y evidencia más fracasos que éxitos, aunque no parece ser un número preocupante.
- APAC: evidencia una posición media en ambos gráficos, con más éxitos que fracasos.
- Japan: esta es la región con más éxitos de todas, es decir que aquí se vendieron más productos que en ninguna otra región. En cuanto a fracasos, aparece tercera.
- Americas: esta región evidencia el doble de fracasos que de éxitos, por lo cual es la región con la peor tasa de éxito/fracaso.
- EMEA: es la región con más fracasos, pero aun así evidencia un número superior de éxitos, siendo la segunda en este tópico.

Por lo tanto, hay más productos efectivamente vendidos que aquellos productos que terminaron en una oportunidad fallida. Esto es porque en APAC, EMEA y Japan se ve claramente que hay más éxitos, y las otras dos regiones, si bien hay más fracasos, la diferencia es suficientemente chica para poder ser compensada.

Adicionalmente, podemos observar que sólo en Japón hay más actividad comercial que en el resto de las regiones que son conjuntos de países. Es por eso que en el siguiente apartado analizaremos en profundidad este territorio.

## 2.3. Japón

Dado que Japón es con creces el país con más oportunidades de la empresa, realizaremos un análisis de algunos de los principales ítems aplicados a Japón.

### Stage de cada oportunidad en Japón

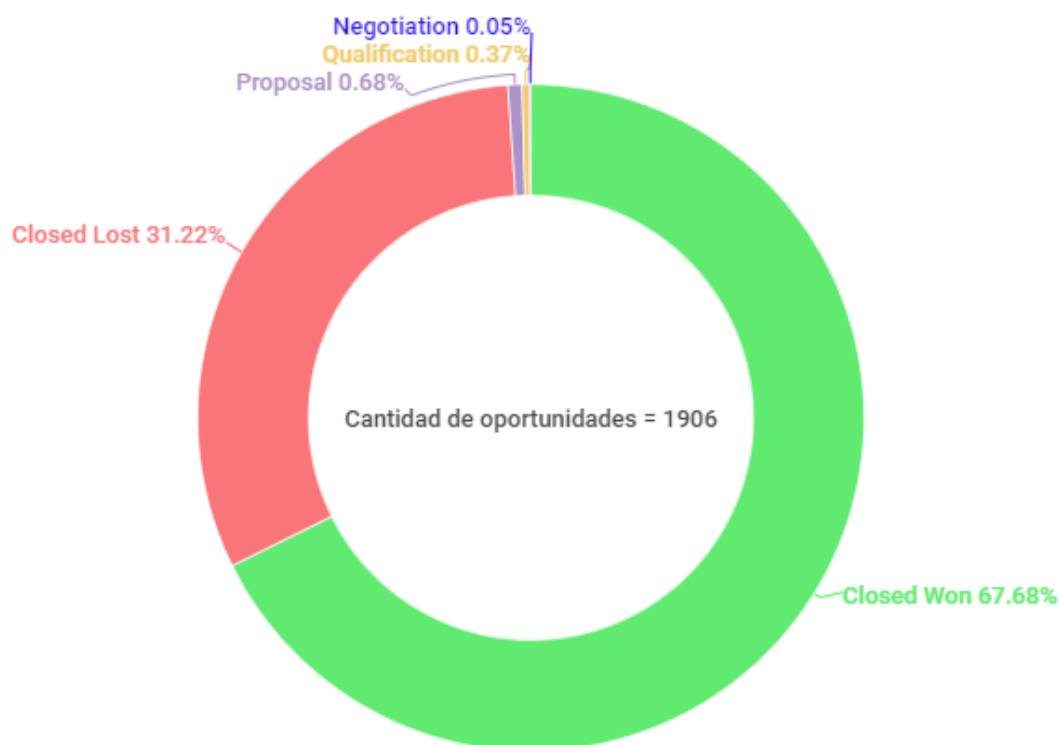


Figura 12: Stage de cada oportunidad en Japón

Vemos primero que la cantidad de oportunidades dentro de Japón en un estado que no es de finalización es muy chiquita, siendo ésta apenas más del 1%. Luego observamos que las oportunidades Closed Lost no llegan a las 600 mientras que las ganadas son, por amplia mayoría, la cantidad superior.

Luego analizando la variable Account.Name, pudimos calcular que en Japón se encuentran 96 clientes de la empresa, que es el 14% del total de la empresa, de los cuales 58 son clientes positivos, es decir, cuyas oportunidades cierran con Closed Won. Teniendo en cuenta solamente

a estos 58 clientes que aportan dinero a la empresa, calculamos el promedio de gasto por cliente en Japón, cuyo resultado de U\$S 2085818.32 por usuario.

Por último, calculamos también los cinco productos más vendidos del territorio, a partir de lo cual llegamos a aquellos que figuran en el siguiente gráfico, con sus respectivas cantidades.

## Top 5 de productos más vendidos

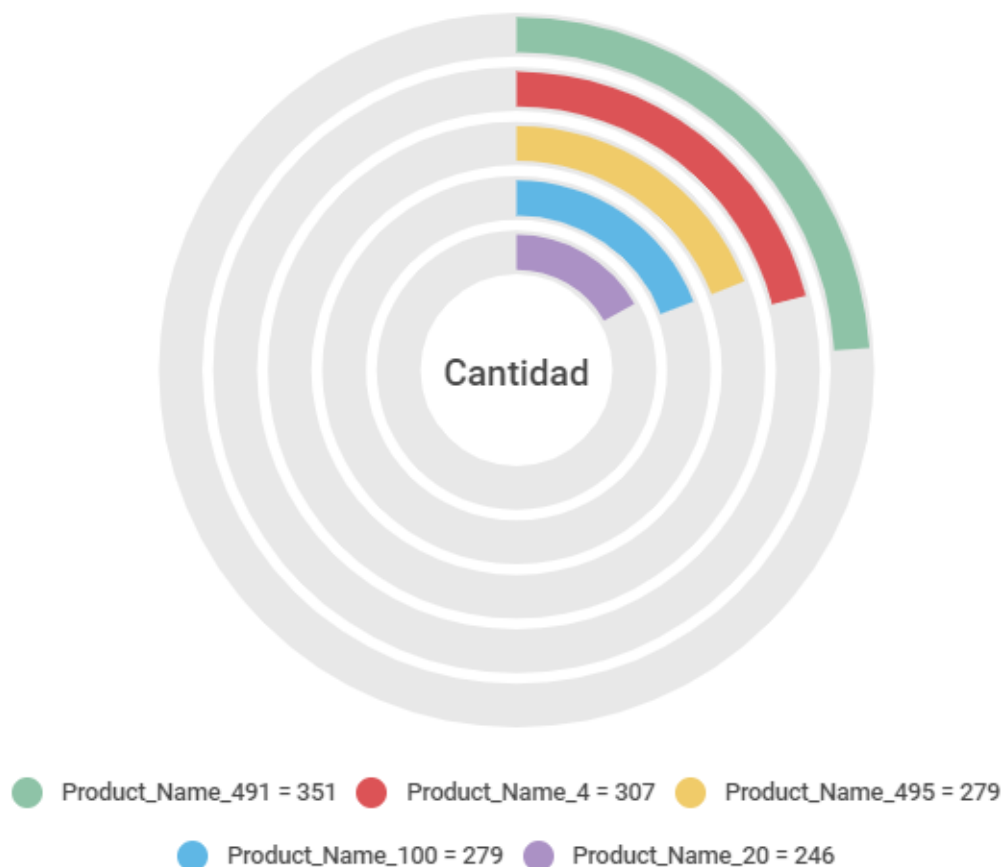


Figura 13: Top 5 productos más vendidos en Japón

## 2.4. Dinero ganado y perdido

Analizaremos las ganancias y pérdidas que se generaron en cada una de las regiones, con el objetivo de establecer una comparación entre las mismas.

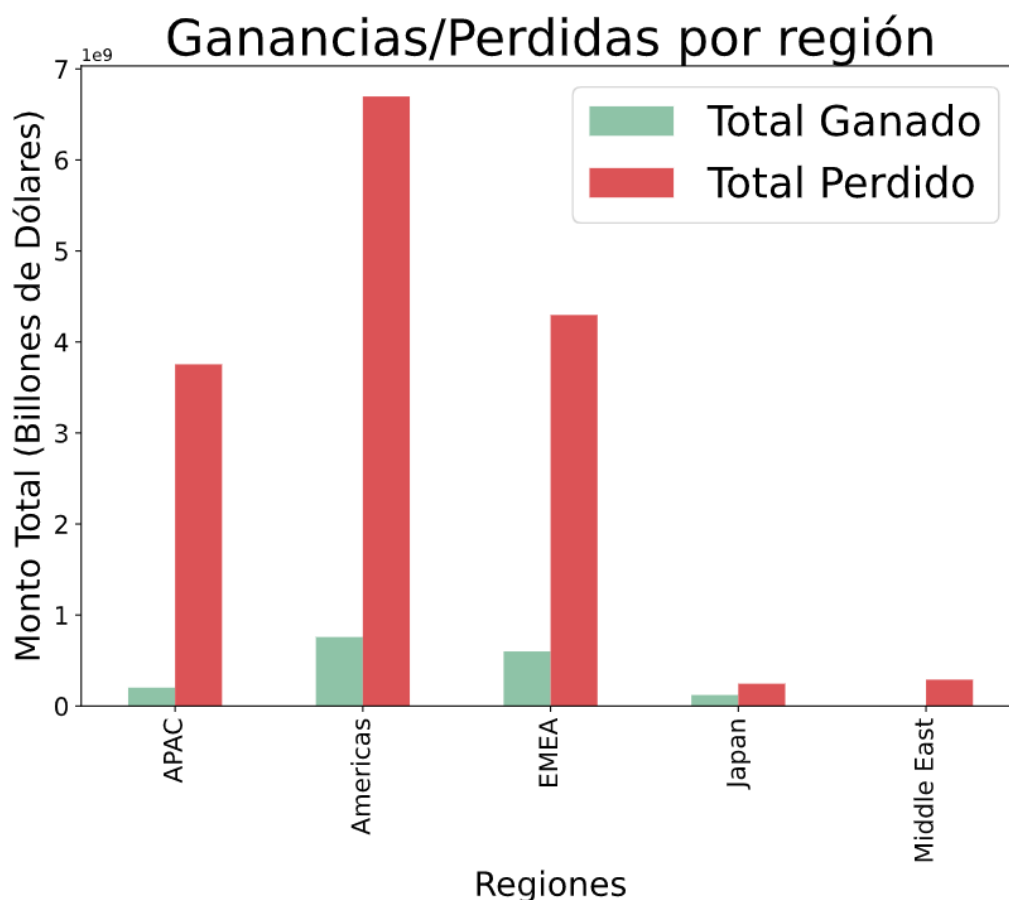


Figura 14: Ganancias y pérdidas por región

De la figura anterior, se puede ver que cada región ha perdido mucho más de lo que han generado de ganancias. Continuando con lo dicho en la sección de “Éxitos y fracasos”, vemos que, si bien la cantidad de éxitos es mayor a la de fracasos, los montos ingresados son menores a los de oportunidades perdidas.

## 2.5. Impacto del clima

En esta sección del trabajo veremos el impacto de los distintos climas de las regiones al rendimiento de la empresa ‘Frío Frío’.

Para cada gran región, analizamos qué países abarcan para tener una mejor idea de los climas de cada lugar. A partir de ello determinamos qué meses corresponden a cada estación del año, fijándonos principalmente en qué hemisferio se encuentran (ver Figura 10), y realizamos los cálculos correspondientes.

A continuación, mencionaremos el trabajo aplicado a cada región para determinar las estaciones de cada una.

- Americas (América del Norte), Japan, Middle East y EMEA (los países de esta última región en los que trabaja la empresa se encuentran todos sobre la línea del Ecuador): se encuentran en el hemisferio norte y así las estaciones se aproximan a una distribución como se muestra a continuación.



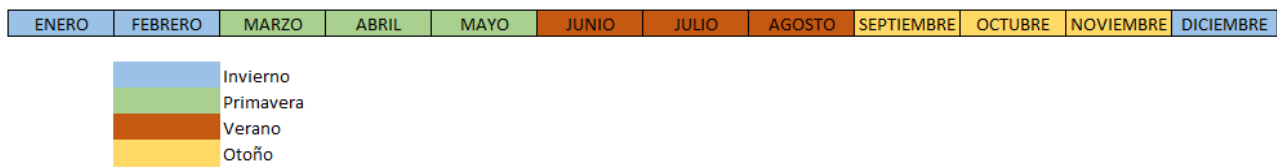


Figura 15: Distribución de las estaciones del año en el hemisferio norte

• APAC: como se puede observar en el world map, todos los países están cerca del Ecuador, menos China y Corea del Sur (Norte) y Australia y Nueva Zelanda (Sur). Analizando el set de datos llegamos a la conclusión que esta región se puede dividir en tres sectores, para los cuales se aclara la proporción de ventas ganadas en esos países para la región:

- Norte (11.29 %): China y Corea del Sur
- Sur (38.34 %): Australia y Nueva Zelanda
- Sudeste asiático (50.37 %): Países cerca del Ecuador (por ejemplo, Indonesia, Singapur, etcétera)

Con el objeto de simplificar el análisis, descartamos el sector norte de APAC (China y Corea del Sur) por abarcar menos del 15 % de ventas ganadas en la región. En consecuencia, quedan para el análisis los sectores de Australia y Nueva Zelanda y Sudeste asiático.

El sector de Australia y Nueva Zelanda se encuentra en el hemisferio sur, entonces las estaciones quedan distribuidas de la siguiente forma:

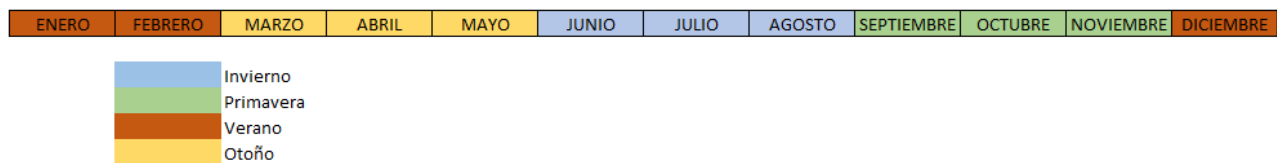


Figura 16: Distribución de las estaciones del año en Australia y Nueva Zelanda

El sector de Sudeste asiático, al estar cerca del Ecuador, tienen un clima especial. Según pudimos averiguar, las estaciones de esa zona se distribuyen a lo largo del año de la siguiente forma:

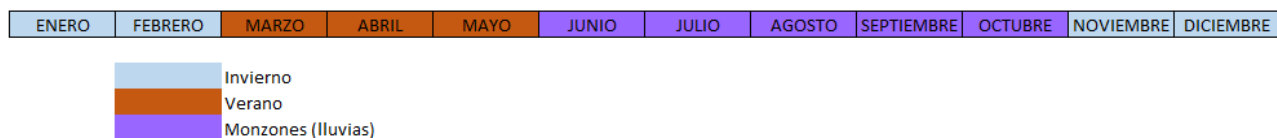


Figura 17: Distribución de las estaciones del año en el Sudeste asiático

NOTA: un monzón es un viento estacional propio del océano Índico y el sur de Asia. Suelen acarrear lluvias torrenciales y fuertes inundaciones.

Una vez considerados los climas de cada zona, analizamos la cantidad de ventas ganadas y el promedio de los precios de los productos por temporada, para ver cómo influye la estación en ambos conceptos. A continuación, se grafica lo mencionado para cada región:

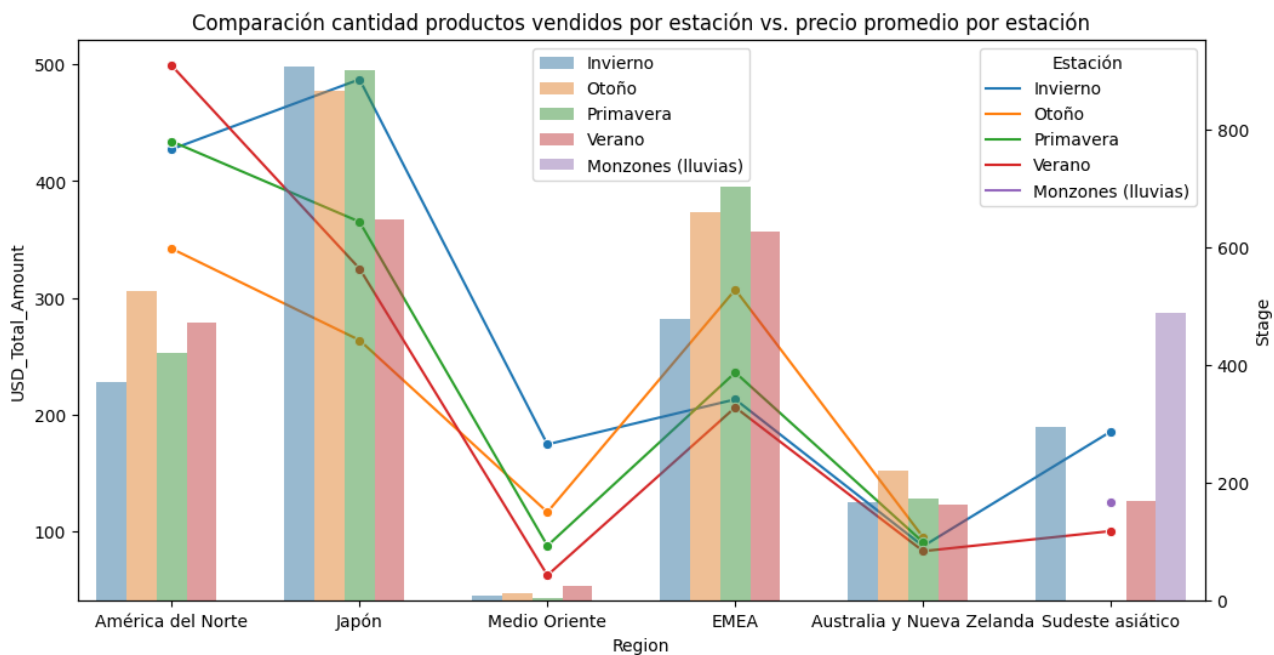


Figura 18: Comparación cantidad de productos vendidos por estación contra el promedio de los precios por estación

Del gráfico anterior, podemos concluir primero que en cinco regiones de las seis el verano no es la época de mayores ventas, contrariamente a lo que esperábamos al tratarse de una empresa de comercialización de aires acondicionados. Además, vemos que Otoño es una estación con presencia similar en todas las zonas.

Adicionalmente, para cada región podemos observar lo siguiente:

En Américas u América del Norte, la estación del año con menores precios es Otoño, que a su vez es cuando más compras se realizan. En Japón, la época de mayor cantidad de ventas es a su vez la de mayores precios. En Medio Oriente, como es esperado, la época con mayor cantidad de ventas es el verano, que a su vez coincide con ser la época de menor promedio de los precios de ventas. En EMEA, la época de mayores precios es una de las épocas de mayores precios, Otoño. Algo similar pasa con Australia y Nueva Zelanda, pero, en este caso, Otoño coincide en ser la de mayor cantidad de productos vendidos. Por último, en el Sudeste Asiático, Monzones es la época de mayores ventas, teniendo esto sentido ya que la región se encuentra cerca del Ecuador, y por lo tanto tiende a ser de templada a calurosa, y es una época húmeda.

### 3. Vendedores

En esta sección nos interesará analizar el comportamiento de los vendedores de la empresa en cuanto a tópicos como estado de las ventas en las que participaron o dinero que recaudaron. Queremos ciertos datos como "quien Vendió más", "quien Vendió menos", "ganancia promedio", etc.

Esperamos luego de un detallado análisis poder dirimir quien es el "Mejor vendedor" de la empresa, teniendo en cuenta todos los datos con los que nos vayamos encontrando. Analizaremos particularmente el comportamiento de los vendedores frente al dinero que hayan recaudado con sus ventas y la cantidad de éxitos y fracasos que hayan tenido cada vez que tuvieron una posible venta.

Los supuestos iniciales que hacemos sobre lo que esperamos de un "Mejor vendedor" de una empresa es que debe tener un alto índice de éxitos y un bajo índice de fracasos, es decir que debe fallar poco y vender mucho, y, en consecuencia a esto, debe juntar mucho dinero con sus ventas, dinero que por supuesto es ganancia para la empresa. En resumidas cuentas: el mejor vendedor de la empresa debe tener el mayor promedio "dinero/intentos de venta".

Lo primero que podríamos preguntarnos es, ¿cuántos vendedores tiene la empresa? La respuesta es que son 53 vendedores diferentes quienes alguna vez participaron en al menos un intento de venta.

#### 3.1. Éxitos y fracasos de los vendedores

Primero nos enfocaremos en cuantos éxitos y fracasos tiene cada vendedor en su haber. Por supuesto, solo nos interesan las ventas finalizadas, por lo que descartamos aquellas que están en una fase previa.

Un veloz análisis nos permite saber que el vendedor con más ventas exitosas es la Persona 50, y el vendedor con más ventas fallidas es... la Persona 50. Es bastante extraño, habíamos dicho que nuestro mejor vendedor debe ser aquel que tuviera más ventas exitosas y además pocas fracasadas. Entonces, ¿Persona 50 es el mejor vendedor de la empresa? ¿O es más bien candidato a ser el peor? Por el momento dejamos la pregunta abierta, será respondida luego.

Ya comentado el caso de la persona con más éxitos, y el de la persona con más fracasos (curiosamente el mismo vendedor), queremos saber cómo es el estado del resto de los vendedores de la empresa. Lo visualizamos más cómodamente en los siguientes gráficos.

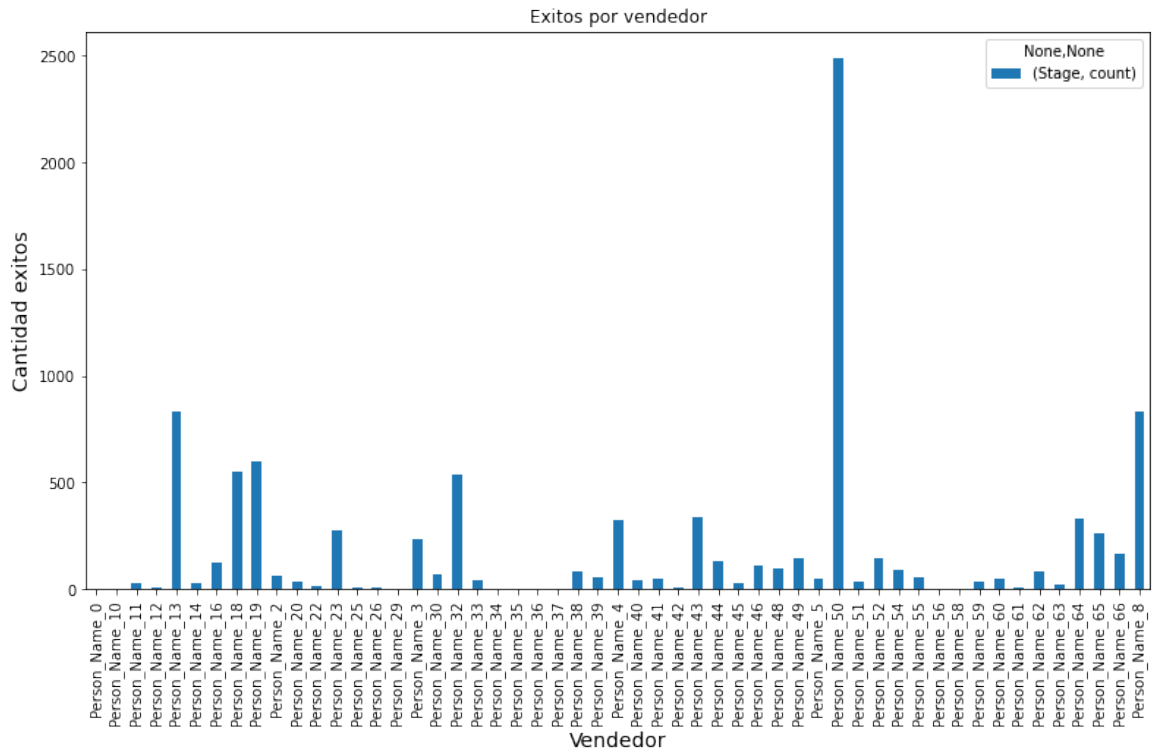


Figura 19: Éxitos por vendedor

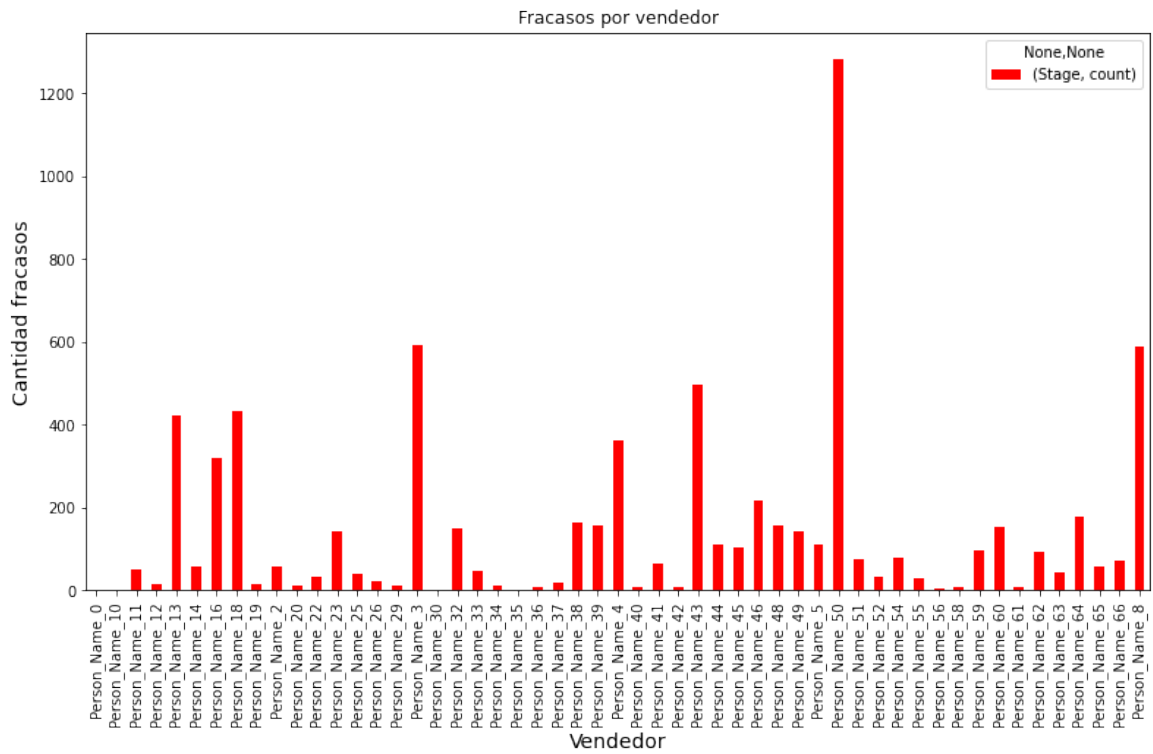


Figura 20: Fracasos por vendedor

Volvemos a graficarlos con una escala más chica a fin de apreciar esta vez la diferencia entre

éxitos y fracasos por cada vendedor. No lo hacemos todo en el mismo gráfico porque la cantidad de vendedores x 2 haría prácticamente ilegible una visualización conjunta. Cuidado, no están en la misma escala de cantidades.

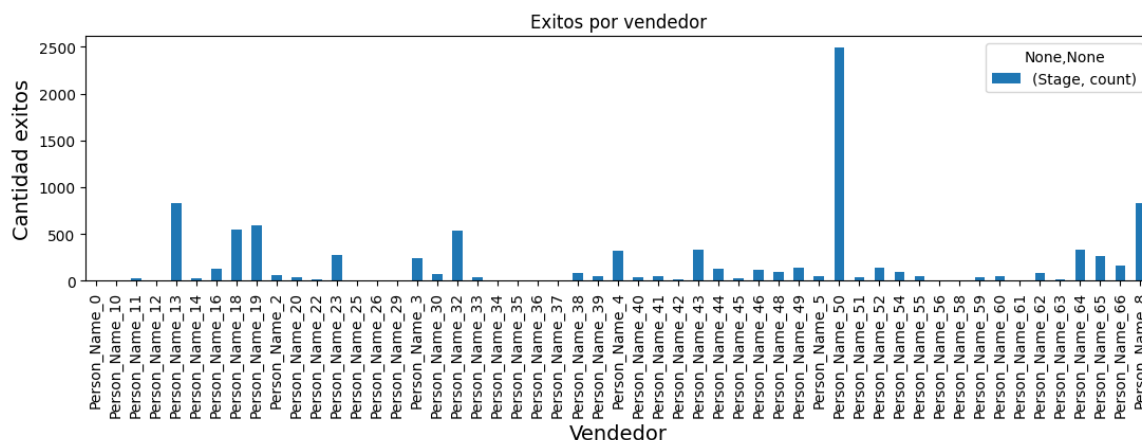


Figura 21: Éxitos por vendedor

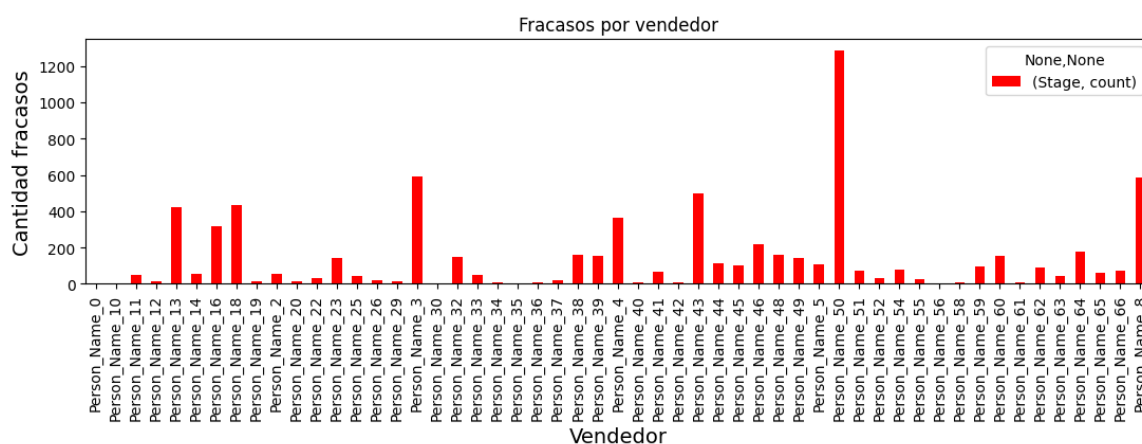


Figura 22: Fracasos por vendedor

Lo que podemos inferir de estas gráficas es que en líneas generales los vendedores suelen tener un número similar de éxitos y fracasos. Ya mencionamos antes el caso del "Persona 50" que es el Mejor vendedor y el Peor vendedor simultáneamente. Sin embargo, trataremos a la "Persona 50", por ahora, como el Mejor vendedor ya que suponemos más importantes los éxitos que los fracasos. Nos reportará más beneficio una persona que venda 10000 productos y fracase en otros 10000, que una que fracase en 0 y venda 100. La "Persona 43" fracasó en 496 ventas... pero vendió 336, que es lo importante. De todas formas, destacamos la performance de Persona 19 y Persona 32, cuyo número de éxitos es bastante más elevado que el de fracaso, sin embargo, esos éxitos aún quedan lejos de nuestro actual "Mejor Vendedor".

### 3.2. Dinero recaudado por los vendedores

Seguimos tratando de encontrar a nuestro mejor vendedor. Para ello ahora calcularemos quiénes son los vendedores que más dinero le reportaron a la empresa con sus ventas, y cuál

tiene el mejor promedio de ventas. Trabajamos con montos en dólares para todos los vendedores, y lógicamente solo tenemos en cuenta las ventas que cerraron con Closed Won, descartando ahora, además de las que no terminaron, las que fracasaron.

Primero calculamos y comentamos algunos datos interesantes que pudimos extraer:

- El total acumulado por todos los vendedores, sumando cada venta de cada uno de ellos es algo más de 1685 millones de dolares.
- El promedio de ganancia por venta por artículo para todos los vendedores por artículo vendido, es decir el total acumulado (calculado en el ítem anterior) dividido los intentos totales de todos los vendedores sumados es aproximadamente 176 mil dolares.
- El producto que más caro se vendió en toda la historia de la empresa fue vendido por más de 27 millones de dolares, y quien logró esto fue "Persona 12".
- Análogamente el producto que más barato se vendió costó cero dólares, y hay varios casos como este. No interpretamos esto como un error del set de datos, sino que asumimos que puede haber sido un regalo en alguna venta, o una promoción de artículos, 2x1, etc.

Lo siguiente que hicimos fue calcular el promedio de venta por producto por vendedor, y obtuvimos que el vendedor con mejor promedio es "Persona 12" (luego de saber que fue el que vendió el artículo más caro, nos lo podemos esperar), y el vendedor con peor promedio es "Persona 35", quien tiene 0 de promedio. Asumimos que es un nuevo vendedor que aún no logro vender ningún artículo.

Luego sumamos las ganancias de cada venta para cada vendedor, a fin de saber quién le hizo ganar más dinero a la empresa. El vendedor que más dinero acumuló en ganancias por ventas fue Persona 8. Cabe mencionar que quizás esperamos que Persona 50 estuviera en este honorífico lugar (ya que fue el que más artículos vendió), pero él ni siquiera figura en el top 10...

Ahora que tenemos la suma de ventas por artículo y el promedio de ventas por artículo, ambos agrupados por vendedor, podemos hacer unos gráficos para tratar de extraer algunas conclusiones.

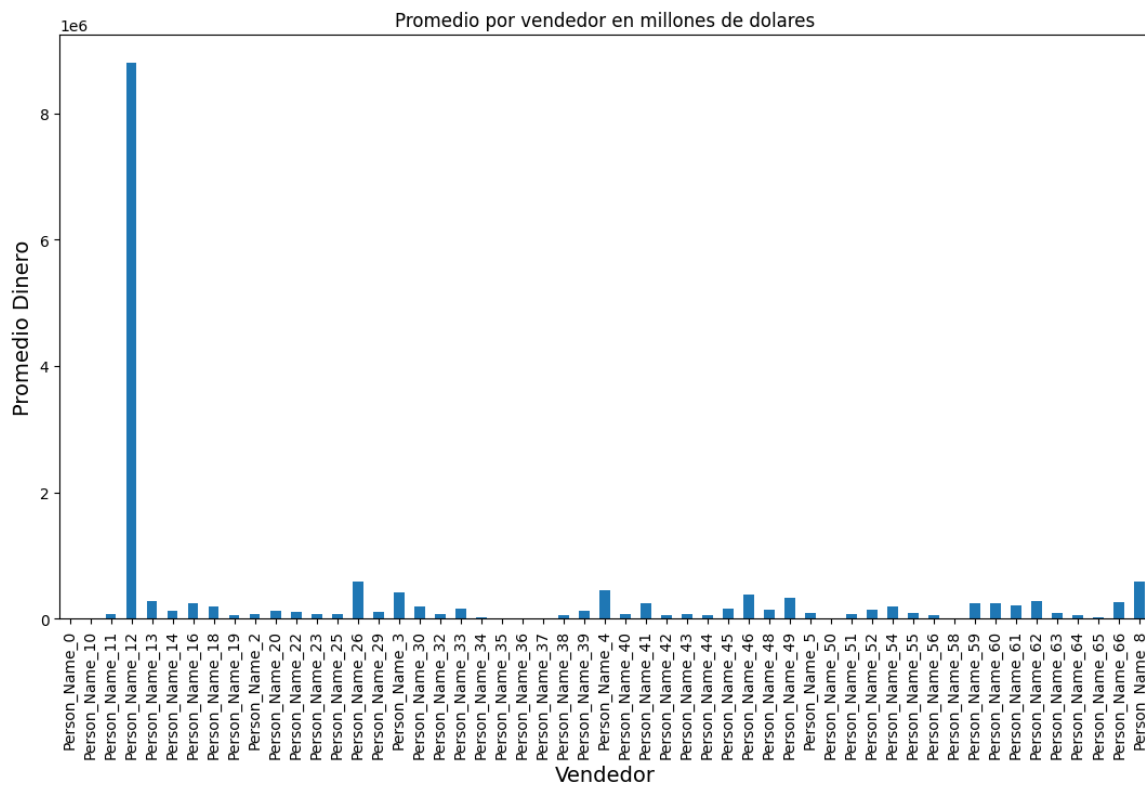


Figura 23: Promedio por vendedor en millones de dolares

Vemos que solo unos pocos se salen de la norma y superan ampliamente a los demás. Es decir, podemos observar claramente dos grupos: Los que tienen un promedio muy bajo y los que tienen un promedio muy alto. Cabe destacar que es la minoría quien se encuentra dentro de este grupo de altos.

¿Y el promedio general para todos los vendedores que habíamos calculado antes?

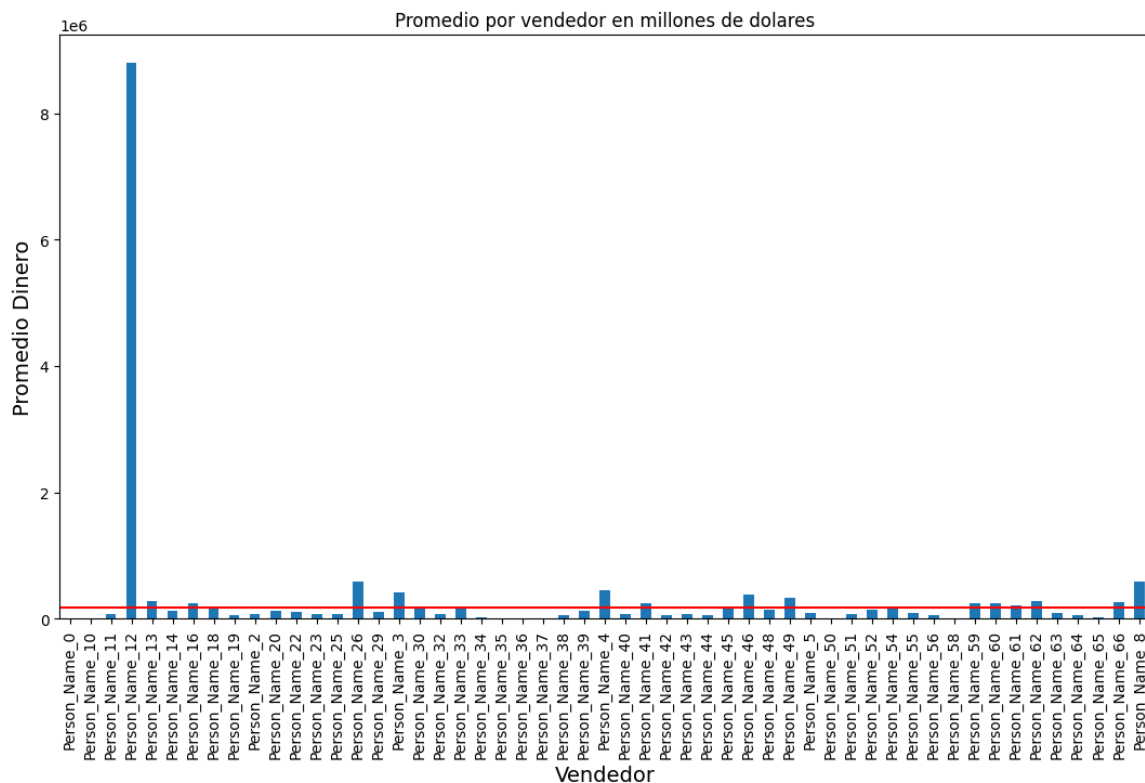


Figura 24: Promedio por vendedor en millones de dolares

Se cumple lo que esperábamos: unos pocos dentro del grupo de "los mejores", y muchos dentro del grupo de "los peores". Tan solo 15 vendedores están arriba de la línea, y es tanto el dinero recaudado por ellos que elevan el promedio de forma que muy pocos, ellos mismos, lo superan.

¿Y nuestro mejor vendedor? ¿La Persona 50 que tenía más ventas cerradas que nadie? Sorprendentemente lo encontramos por debajo del promedio, y ubicado entre "Los peores"... ¿Seguimos diciendo que es el mejor?

Hacemos ahora el gráfico para los vendedores que más dinero recaudaron, nos preguntamos si será análogo a los gráficos del promedio.



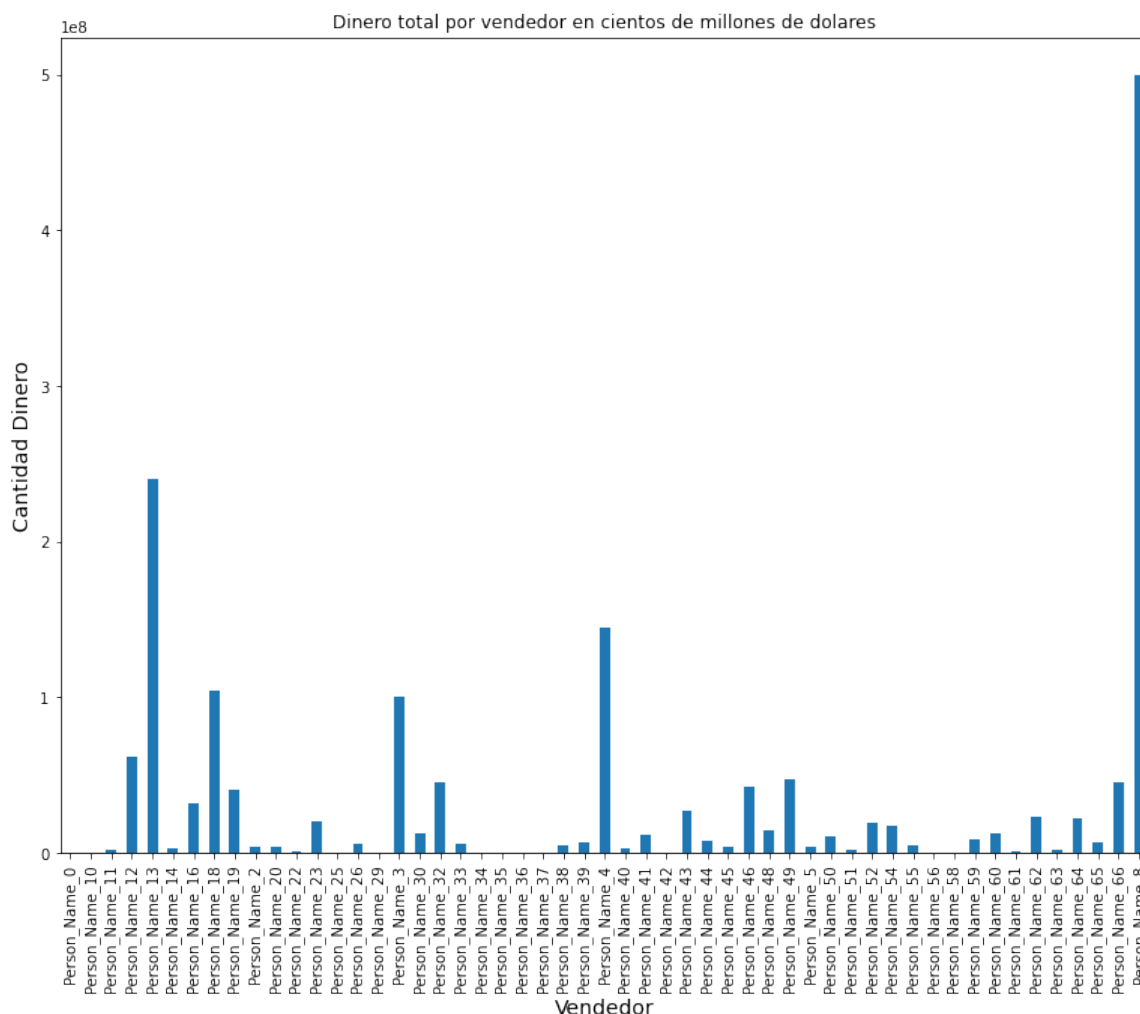


Figura 25: Dinero total por vendedor en cientos de millones de dolares

En este grafico vemos que, por mucho, la Persona 8 es el vendedor que más dinero le hizo ganar a la empresa. Además, esta persona era el segundo mejor vendedor en cuanto a promedios, y ocupaba un lugar bastante alto entre los que más ventas había cerrado. Sin embargo, no fue él quien había vendido el producto más caro (era la Persona 12, quien era el que tenía mejor promedio).

¿Y nuestro mejor vendedor, la Persona 50? Luego de haber ocupado un lugar entre los peores en promedios... Vuelve a ocupar un lugar bastante discreto también en este gráfico.

### 3.3. Conclusiones ¿Quién es el mejor vendedor?

En conclusión, por ocupar el mejor lugar en precio acumulado por ventas y segundo lugar en promedio, diremos que la Persona 8 es el "Mejor vendedor" de la empresa.

Decimos que la Persona 50, quien antes habíamos considerado el mejor, ni siquiera está ahora entre "los mejores". Sí destacamos su, seguramente, amplia trayectoria por ser el vendedor que más productos cerró.

Finalmente, podemos decir que no hay una relación directa entre ventas cerradas y suma de ganancias, aunque ciertos vendedores evidencien una alta posición en las dos categorías. Sí

hay muchos, la mayoría, que aparecen bastante abajo en cada tópico.

Otra conclusión a la que podemos llegar es que en realidad el promedio es más una curiosidad que algo que pueda importar mucho en una empresa. Antes habíamos dicho que no nos importaba las veces que el vendedor falla, sino las que acierta. El promedio es el mismo caso. En todo caso nos importan las ventas que la persona cerró y el dinero que le trajo a la empresa, más que cuánto por cada vez promediada. Más aun si suponemos que un vendedor joven que recién inicia fallará más y recaudará menos que un vendedor experto con años en el negocio. Es decir, en todo caso tendría más sentido un promedio por etapas en la vida laboral del vendedor, separados en años o en cantidad de ventas.

### 3.4. Premios consuelo: Mejores vendedores por región

En la sección anterior logramos discernir que el vendedor número 1 de la empresa es "Persona 8". Ahora también queremos saber quién es el mejor vendedor por región. Por supuesto y sin lugar a duda, Persona 8 también obtendrá el título de Mejor vendedor de su región, pero veamos quien más logra alguna distinción.

Nuevamente, tomamos como válidas las ventas que ya han terminado en Closed Won y todos los precios en dólares. Sabemos que hay 5 regiones: APAC, Américas, EMEA, Japón y Middle East. Analizamos cada una de ellas por separado.

Agrupando por vendedor y sumando los montos obtenidos por cada uno obtenemos lo siguiente:

- El mejor vendedor de APAC es "Person 49".
- El mejor vendedor de Americas es "Person 8".
- El mejor vendedor de EMEA es "Person 13".
- El mejor vendedor de Japan es "Person 66"
- El mejor vendedor de Middle East es "Person 5"

Finalmente mostramos con unos gráficos el comportamiento de todos los vendedores por región.

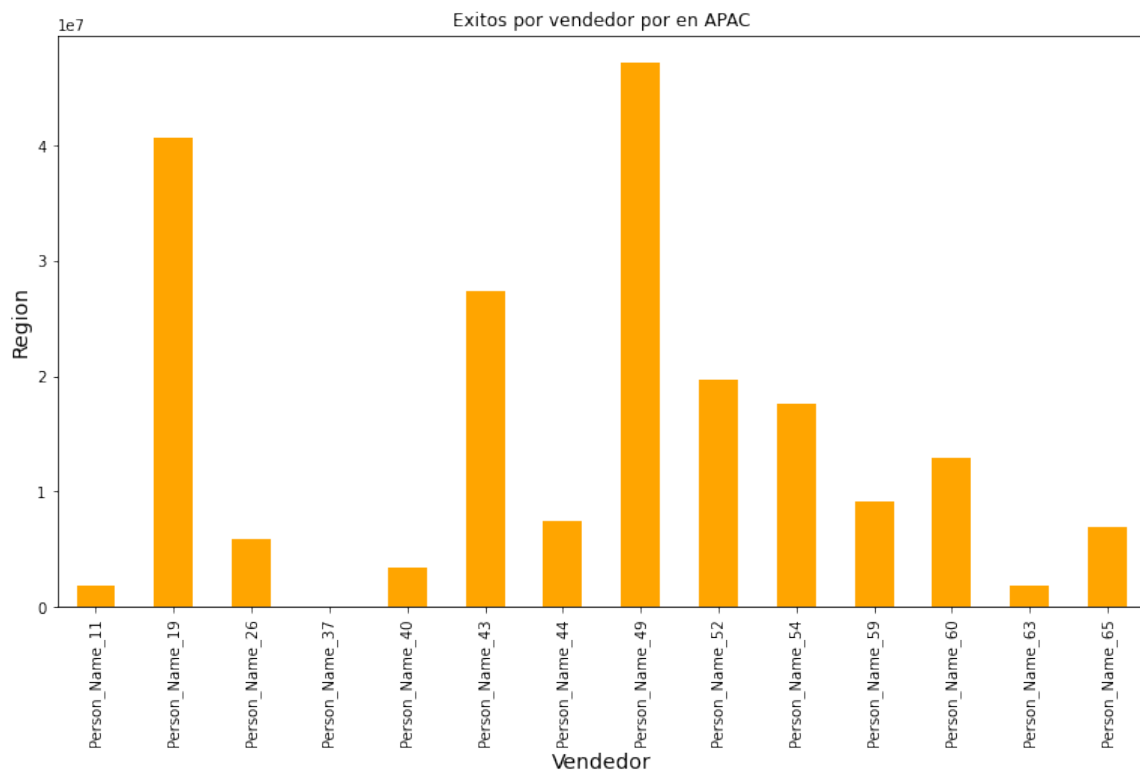


Figura 26: APAC: Ganancias por vendedor en decenas de millones de dolares.

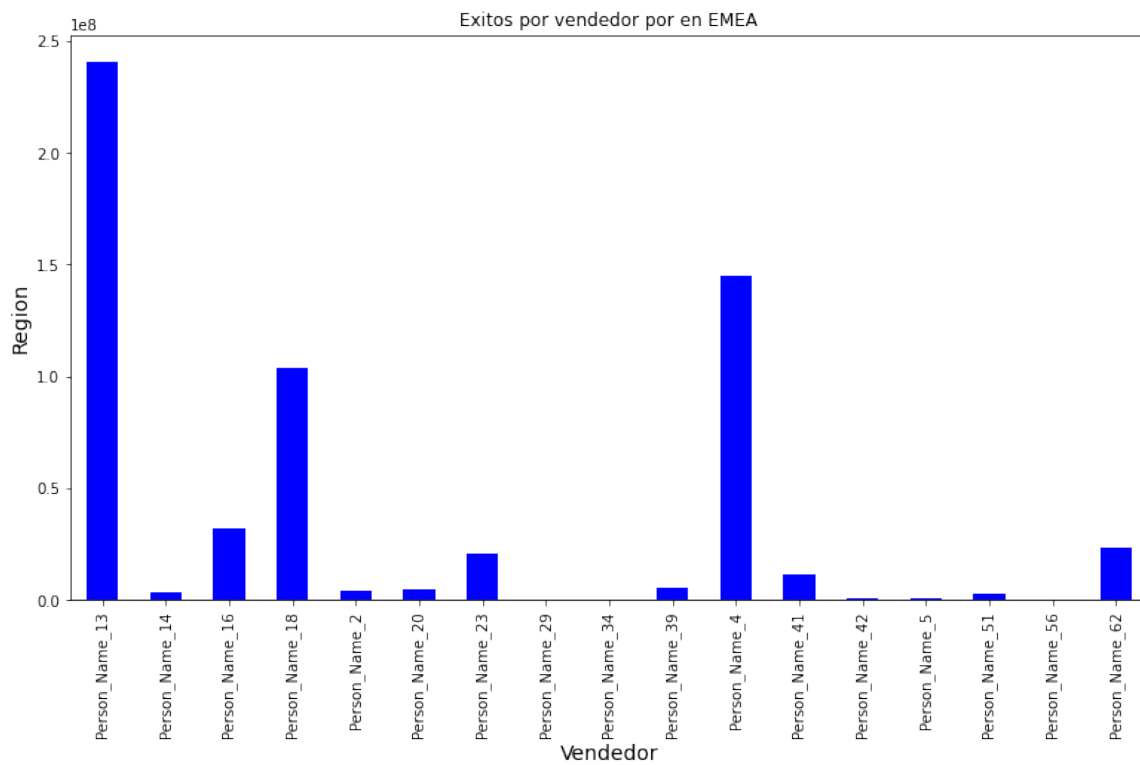


Figura 27: Americas: Ganancias por vendedor en cientos de millones de dolares.

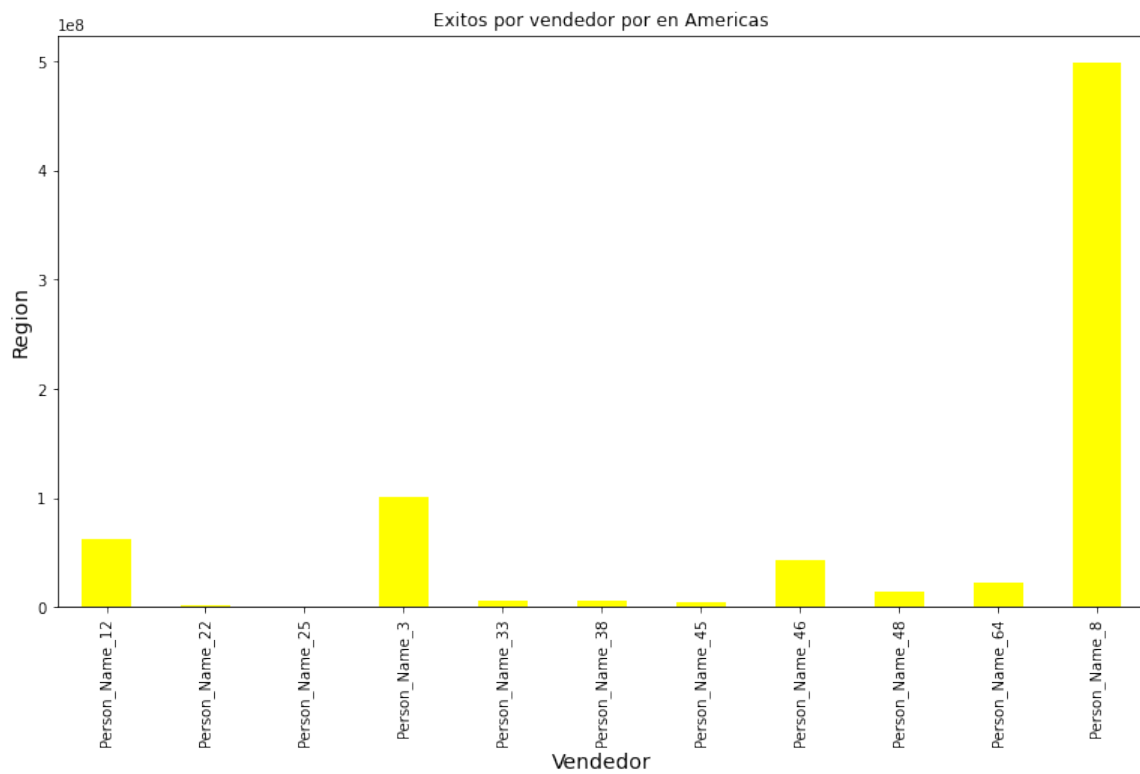


Figura 28: EMEA: Ganancias por vendedor en cientos de millones de dolares.

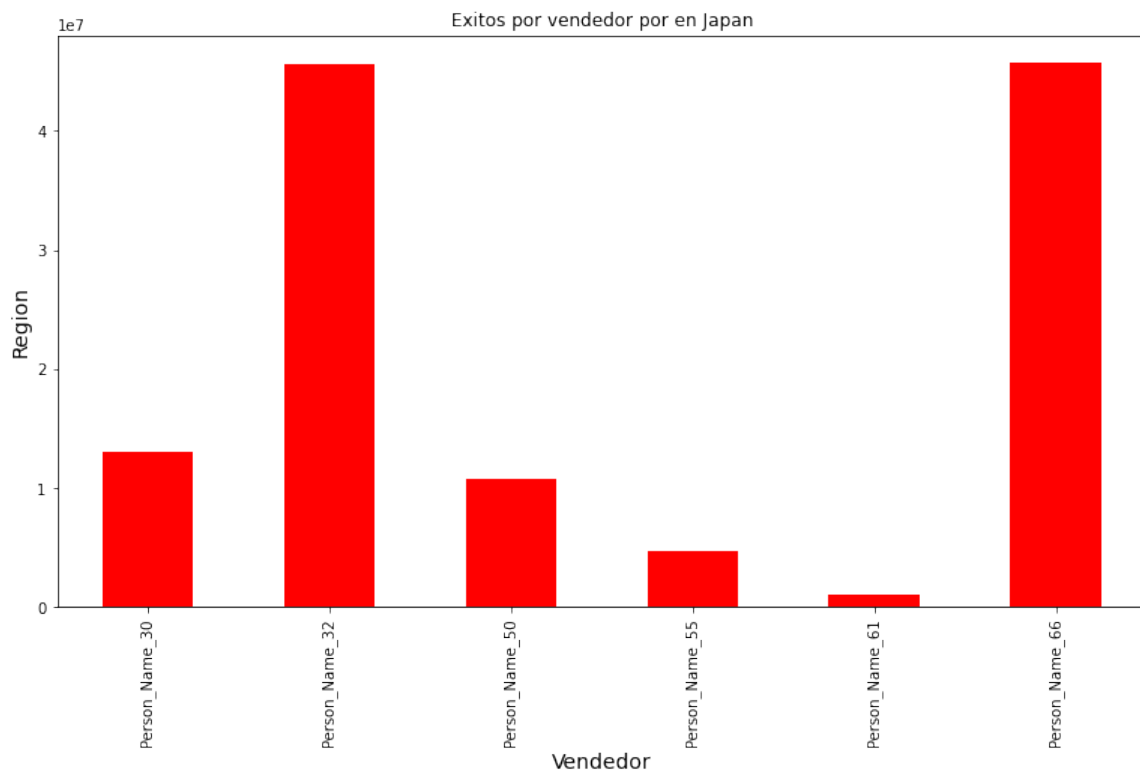


Figura 29: JAPAN: Ganancias por vendedor en decenas de millones de dolares.

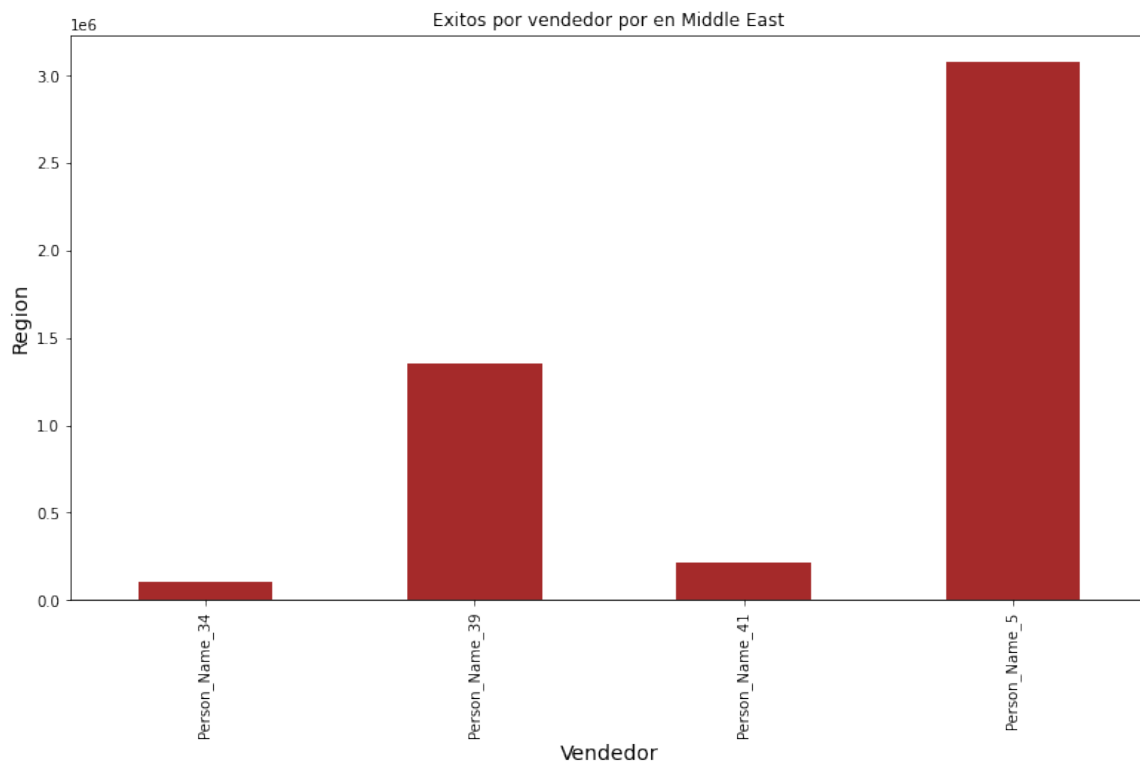


Figura 30: Middle East: Ganancias por vendedor en millones de dolares.

### 3.5. Apéndice infográfico

Para terminar esta sección, se presenta una infografía sobre nuestros queridos vendedores, presentando más detalladamente números y montos que, si bien estaban en los gráficos, no los detallamos en palabras.

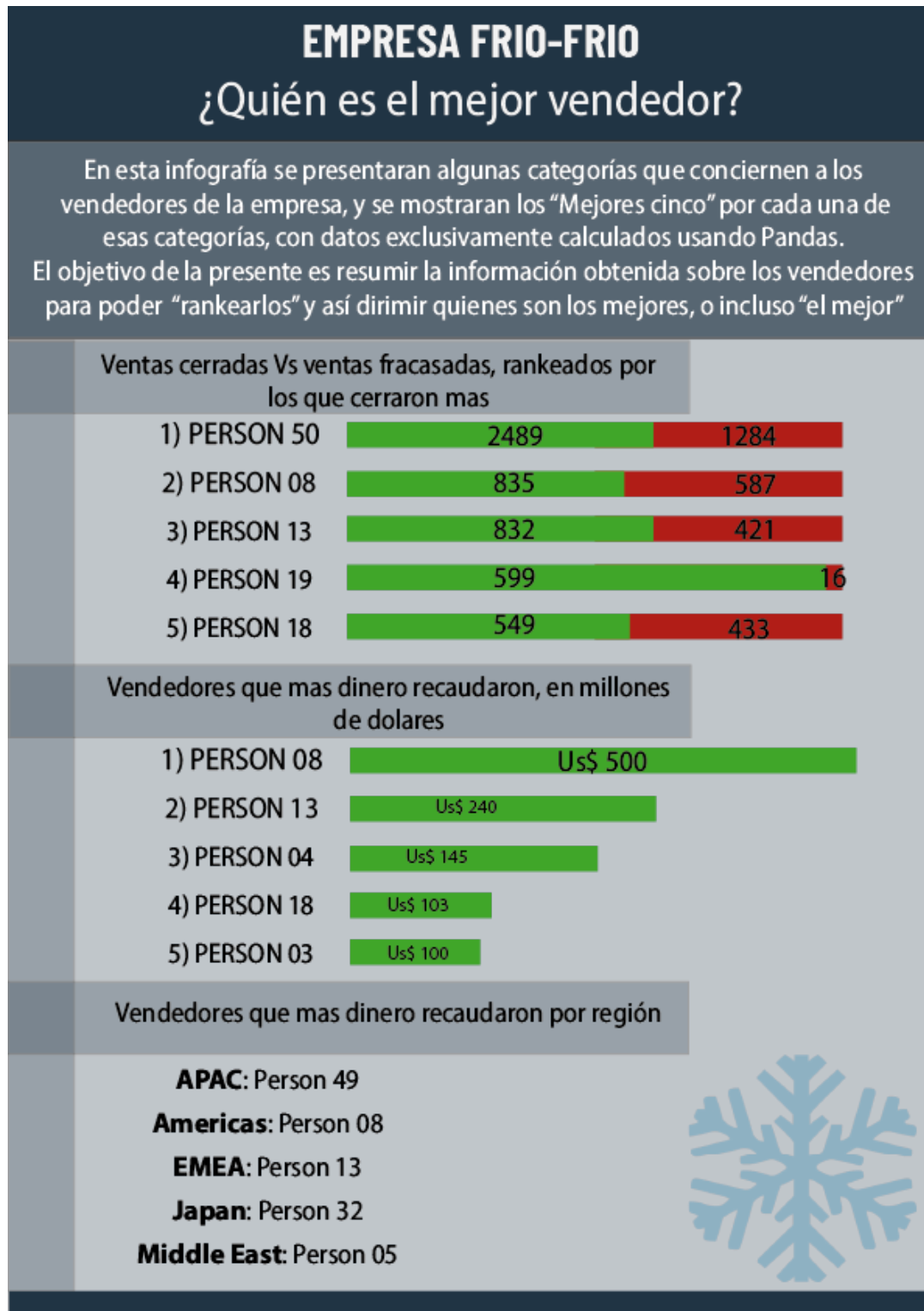


Figura 31: Infografía sobre el comportamiento de los vendedores en diversas categorías

## 4. Clientes

En la sección anterior hicimos un análisis acerca del comportamiento de los vendedores de la empresa (y sí, logramos familiarizarnos e incluso encariñarnos con muchos de ellos). Ahora nos toca mirar la otra cara de la moneda: los clientes.

Analizaremos una serie de tópicos y luego plantearemos una pregunta análoga a la que nos hicimos con los vendedores: ¿los clientes que más veces compraron son los que le reportaron más dinero a la empresa? Recordamos que esto no era así con los vendedores, es decir los vendedores que más venden no eran (no siempre) los que recaudaban más. Finalmente entregaremos nuestro premio al "Mejor cliente".

Por supuesto sólo analizamos ventas exitosas y todos los montos en dólares.

### 4.1. Clientes de la empresa

En total la empresa reportó exactamente 664 clientes que conforman las más de 16.000 ventas que tienen en su haber. Un rápido cálculo nos dice que el promedio de clientes por vendedor es entre 12 y 13, aunque pronto veremos que una cosa es un cliente que compra una vez y otra son los clientes frecuentes", por lo que un promedio así no tiene mucha razón de ser.

A modo de aclaración, solamente llamamos clientes a aquellas personas que compraron al menos un producto de la empresa, es decir que no consideramos clientes a las personas que no compraron nada, aunque aparezcan en la base.

A modo de curiosidad exponemos que el cliente que compró el producto más caro es Name 1723 y gastó más de 27 millones de dólares en dicho producto. Luego Name 1723 ocupa también el tercer puesto dentro de personas que compraron el producto más caro. El segundo lugar, cuarto y quinto también los ocupa una misma persona, Name 1662. ¿Alguno de ellos serán el Mejor cliente?

### 4.2. ¿Son los clientes que más compraron aquellos que más gastaron?

Llegó la hora de responder la pregunta planteada en el inicio de la sección. Queríamos asumir que sí, pero el caso de los vendedores nos enseñó a no dar cosas por sentado hasta ver los hechos. Y los hechos se ven mejor en gráficos.

Antes de mostrar la respuesta comentamos que para esta parte del análisis solamente tomamos en cuenta a los clientes frecuentes de la empresa. Definimos como frecuentes a los clientes que compraron más de 30 productos, estos son exactamente 51 clientes, nuestros clientes favoritos podríamos llamarlos.

Agrupamos por cliente y, por un lado, sumamos la cantidad de veces que compraron, y por el otro sumamos los montos que gastaron en sus productos. Ordenamos descendientemente y obtenemos la siguiente gráfica, con la cantidad de compra en azul y el dinero total gastado en rojo.

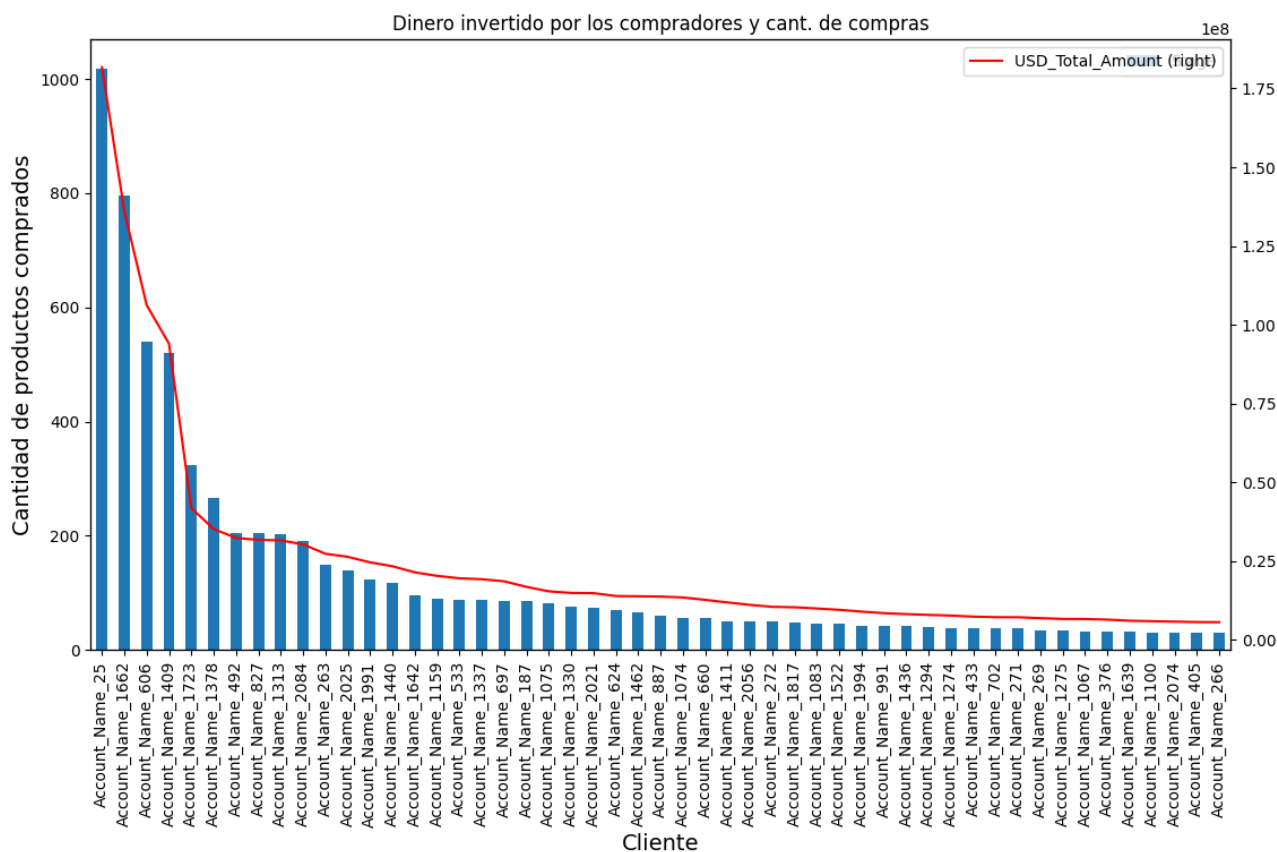


Figura 32: Dinero gastado por cliente y cantidad de compras

### 4.3. Conclusiones, menciones, agradecimientos y algo más

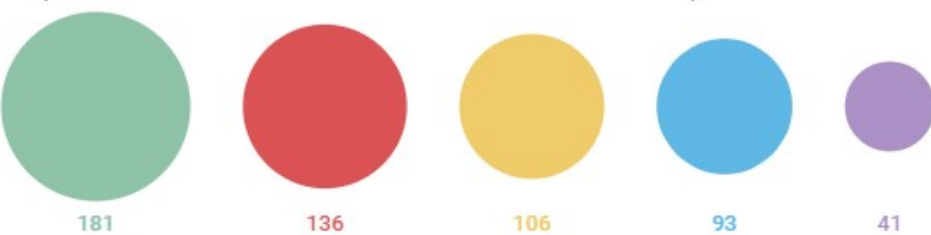
Como podríamos haber esperado, existe una relación directamente proporcional entre el dinero invertido por los clientes en la empresa y la cantidad de productos que compraron. Es decir, los que compraron más, gastaron más, así como los que compraron menos gastaron menos. Contrario al caso de los vendedores, efectivamente existe una relación que podríamos considerar prácticamente lineal.

Llego la hora de otorgar la mención a "Mejor cliente", que corresponde a Name 25, quien invierte en la empresa nada menos que 175 millones de dólares con sus compras, muchas gracias Name 25. También otorgamos menciones honorificas a los clientes Name 1723 y Name 1662 quienes habían sido los que compraron los cinco productos más caros. Ellos ocupan la quinta y la segunda posición, respectivamente, en el ranking de "Mejores clientes". Quienes completan el top 5 son Name 606 (tercero) y Name 1409 (cuarto).

Para cerrar esta sección, además de agradecerles a nuestros clientes su lealtad para con la empresa, presentaremos dos gráficas más. La primera muestra a los clientes que más dinero gastaron en la empresa, y la segunda muestra a los 5 clientes que más gastaron y cuales fueron esos montos.



**Top 5 clientes que mas gastaron  
(en millones de dolares):**



32

## 5. Familia de Productos

Siguiendo con el análisis, queremos analizar la posibilidad de encontrar una familia de productos líder en ventas y ganancias, cabe descartar que esta no será una tarea sencilla dado que, para este análisis, nos enfocaremos tanto regionalmente como globalmente, para poder determinar un ganador absoluto.

Para poder seleccionar a dicho líder en ventas y ganancias, debemos tener en cuenta algunos criterios para que la elección sea un poco más "justa", básicamente buscamos que el ganador tenga una relación bastante baja entre las ganancias y pérdidas, así como también tenga ventas. Estos criterios se aplicarán tanto a nivel global como regionalmente.

### 5.1. Análisis Global

Enfocado en el mercado global, cada familia de productos se destaca tanto por sus ganancias, pérdidas y cantidad de ventas. Como primer análisis nos enfocaremos en sus ganancias, luego para terminar este análisis cerraremos con las ventas que tiene cada familia y así determinar si estas repercuten al determinar el líder del mercado.

#### 5.1.1. Ganancias en Ventas

Para este análisis, y dada la cantidad de familias de productos competidores, nos enfocaremos en un TOP 10, analizando detenidamente sus principales ganancias y pérdidas. En la figura siguiente, se puede ver que la Familia de Productos 77, tiene la mayor Ganancia en ventas a nivel Global (\$ 177 666 400.00), sin embargo, vemos que tiene una pérdida en ganancias bastante considerable (\$ 997 046 000.00), la pregunta es ¿se le puede considerar líder del mercado?.

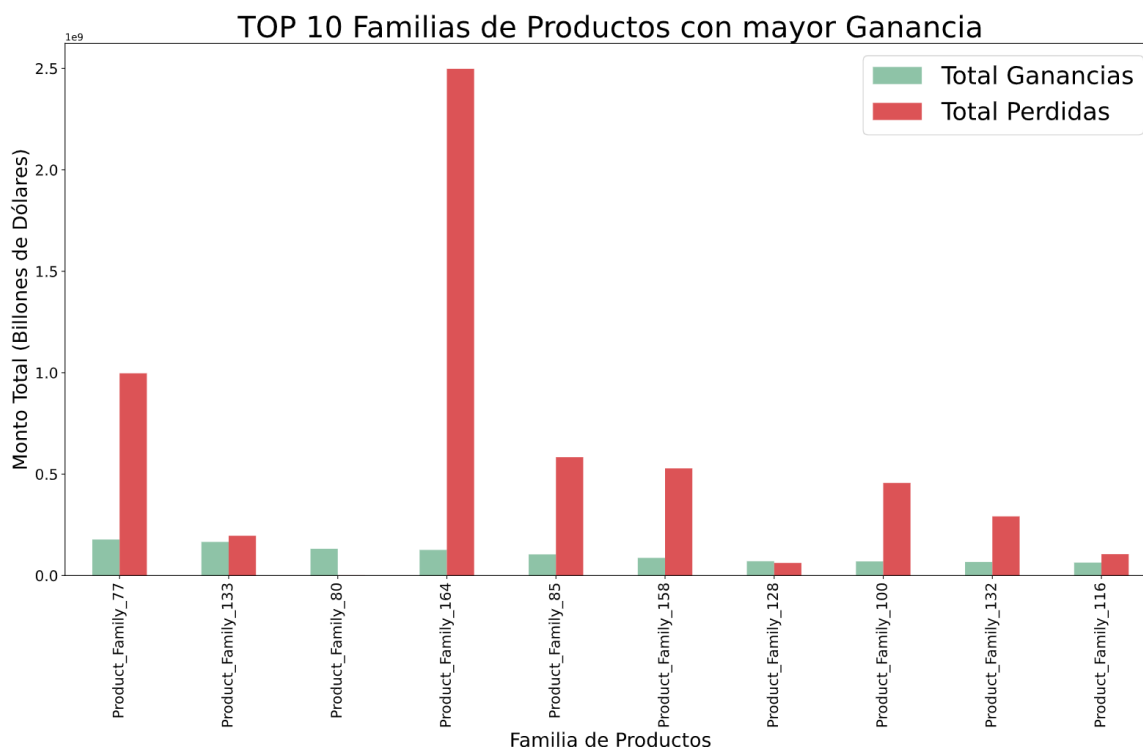


Figura 35: TOP 10 - Familia de Productos con mayor Ganancia

Como vemos, esto nos dejó una gran pregunta, y para tratar de responder dicha pregunta, presentaremos otra figura con el TOP 10 de familia de productos con mayor pérdidas en ganancias, y veremos si la Familia de productos 77 con sus pérdidas está a un paso de ser el líder de mercado.

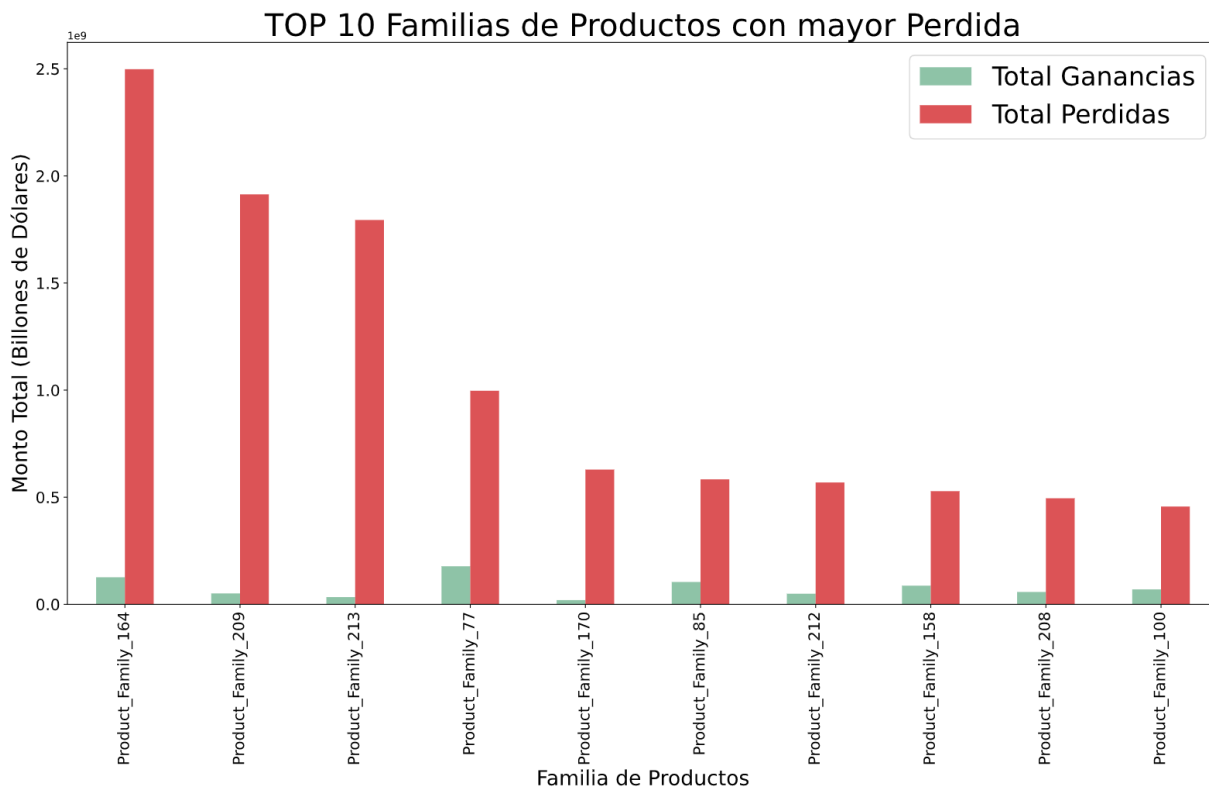


Figura 36: TOP 10 - Familia de Productos con mayores Pérdidas

Al parecer a la Familia 77 no le va tan bien, dado que ocupa el 4º lugar de este top de pérdidas, sin embargo podemos destacar que dentro de ese top es quien tiene mayor ganancias en ventas. Con estas figuras podemos decir que la Familia 77 está a un paso de convertirse en el líder del mercado global. En la siguiente sección veremos si la cantidad de ventas influye para elegir el líder de mercado.

### 5.1.2. Cantidades de Ventas

Para empezar con esta sección, primero haremos mención que la Familia de productos 77 está a un paso de convertirse en el líder de ventas global; veremos si con la siguiente figura podemos concluir y darle finalmente ese título. Al apreciar y comprender la figura siguiente nos damos con la sorpresa que la Familia 77 no es quien tiene una mayor cantidad de ventas exitosas, por el contrario es quien ocupa el 2º de este TOP.

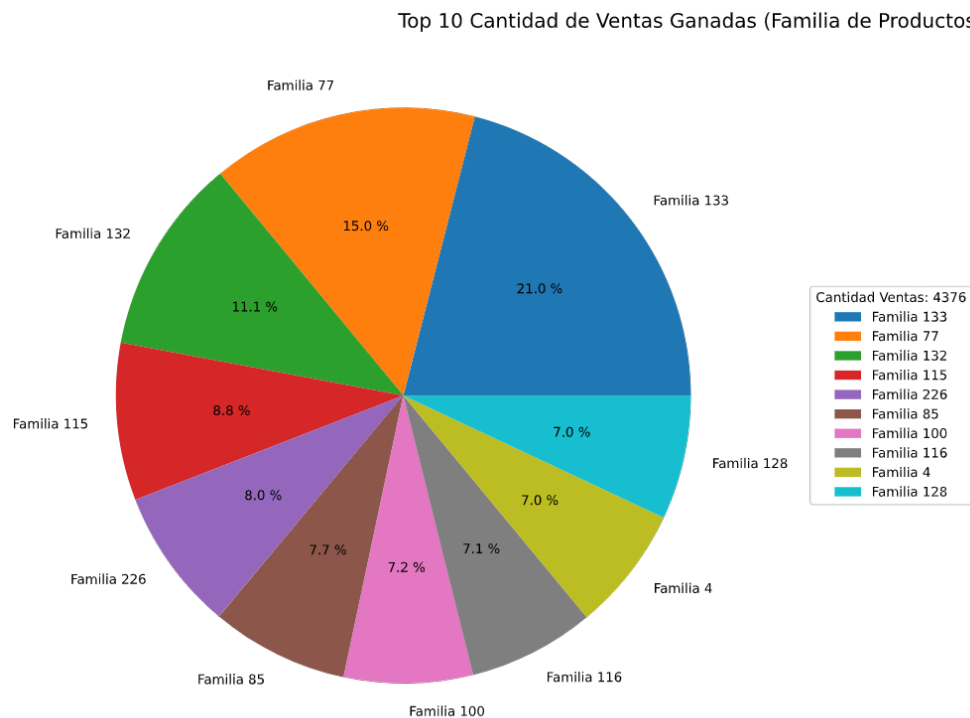


Figura 37: TOP 10 - Familia de Productos con mayor Cantidad de Ventas

Ahora nos encontramos con un dilema: "¿generar ganancias es lo mismo que vender mucho?". Por lo que se pudo apreciar este no es el caso. Pero podemos decir que la familia que dio la sorpresa y es la número 1 en ventas es la familia 133. Dicha familia se encontraba en el 2º lugar de nuestro TOP 10 en ganancias en ventas (\$165 821 100.00), además debemos resaltar que las pérdidas (\$196 199 500.00) que se muestran en dicha Figura 35, no son tantas si lo comparamos con la Familia 77. Además, la Familia 133 no aparece en el TOP 10 de familias con mayores pérdida.

Con todo esto podemos decir que tenemos una familia líder en el mercado Global, la Familia 133, y esto se debe a su gran cantidad de ventas y ganancias generadas, además de sus pocas pérdidas.

Aun así, esta Familia 133, es también líder en el mercado regional, esto lo veremos en la siguiente sección, dentro de la cual haremos un análisis detallado por región y veremos qué sucede.

## 5.2. Análisis por Región

Para empezar esta sección debemos recordar como se distribuyen las regiones en el mundo, para lo cual recordaremos la Figura 10, de una sección anterior (ver Sección 2), para poder comprender el análisis siguiente.

Al igual que con el análisis global, muchas familias de productos destacan en diferentes aspectos dentro del mercado global. Veamos si ahora podemos encontrar un líder de mercado por región y ver si esto influyó de alguna manera a convertir en el líder de mercado global a la Familia 133.

### 5.2.1. Ganancias en Ventas

Como hicimos anteriormente, primero analizaremos las ganancias y pérdidas que se generaron en cada una de las regiones, esto desde un lado más global, por ahora sin tener en cuenta las familias de productos, esto simplemente para ver una comparativa de las ganancias y pérdidas que se establecieron en cada región.

De la figura 14, mostrada en la sección 2, se puede ver que cada región ha perdido más de lo que han generado las ganancias, veremos como esto afecta a cada familia de productos, para lo cual se presentarán varios gráficos de cada región mostrando una comparativa entre sus ganancias y pérdidas. Cabe destacar que dichos gráficos serán TOP 10 por cada región dado que, como mencionamos, estamos en la búsqueda del líder del mercado regional.

## Ventas Totales - Región América

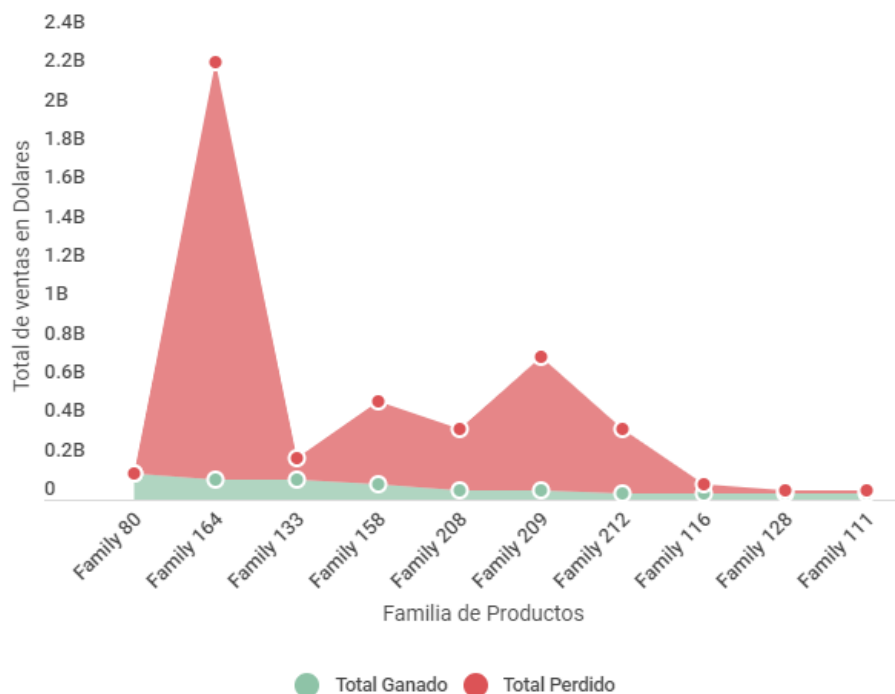


Figura 38: TOP 10 Ventas Totales - Región América

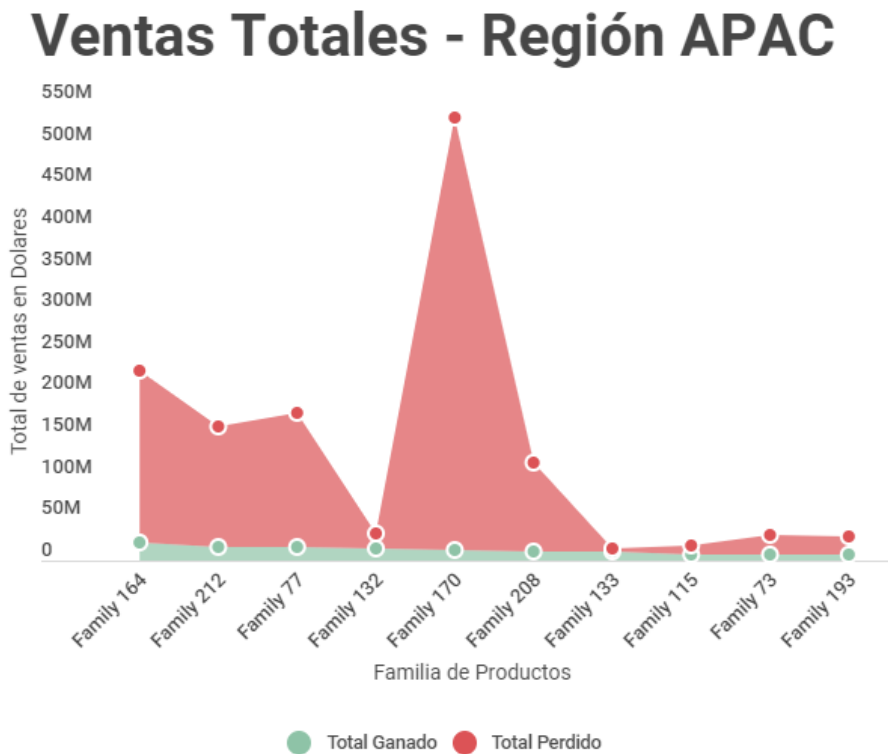


Figura 39: TOP 10 Ventas Totales - Región APAC

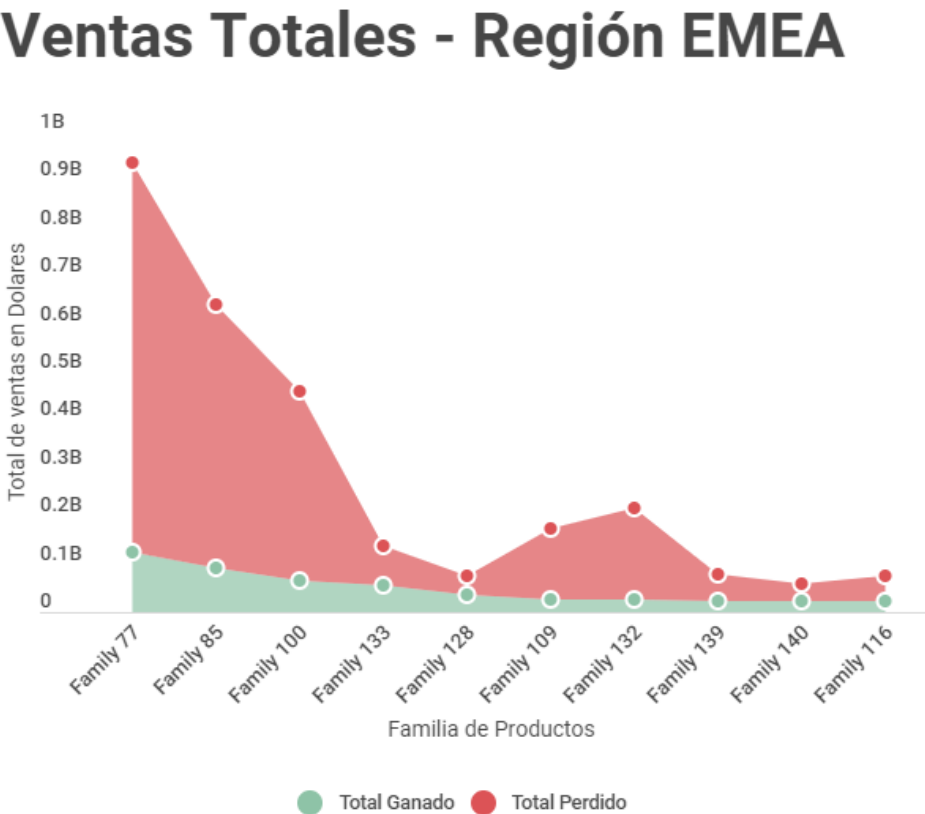


Figura 40: TOP 10 Ventas Totales - Región EMEA

## Ventas Totales- Región JAPAN

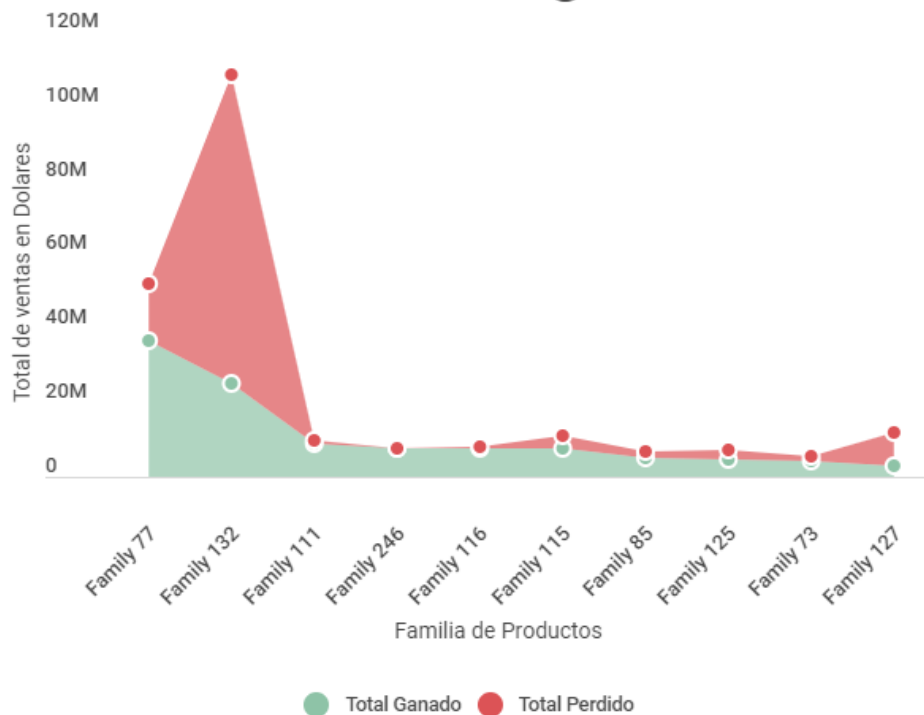


Figura 41: TOP 10 Ventas Totales - Región Japan

## Ventas Totales - Región Medio Este

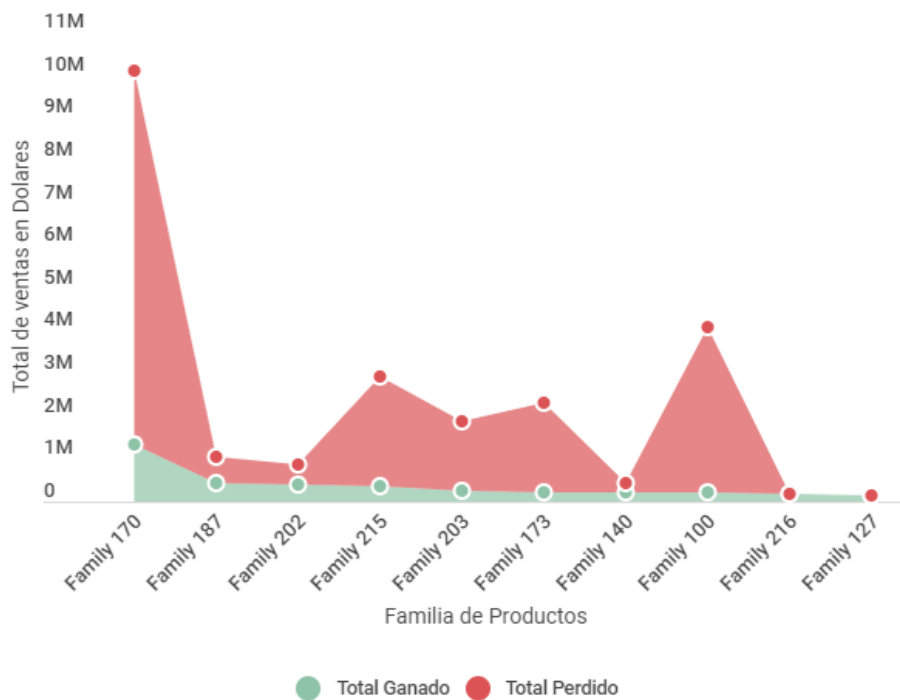


Figura 42: TOP 10 Ventas Totales - Región Medio Este

Las figuras anteriores demuestran una vez más que no podemos sacar conclusiones apresuradas y decir que una familia de productos es líder del mercado, pues se ve claramente que hay mucha diferencia en cuanto a ganancias generadas por cada familia. En resumen, tenemos que para la región EMEA la que obtuvo mayor ganancias es la anteriormente mencionada Familia 77, a esto se suma que la Familia 133 también aparece en dicha región. Cabe resaltar que ambas familias también tienen pérdidas en sus ganancias, así que aun no hay nada concluyente. Por otro lado, las demás regiones tienen líderes distintos a excepción de la región de Japan que nuevamente tiene a la Familia 77 como cabeza dentro del TOP 10, quedando de la siguiente manera:

- EMEA: Familia 77 (1º) - Familia 133 (4º)
- América: Familia 80 (1º) - Familia 133(3º)
- Japan: Familia 77 (1º)
- Medio Este: Familia 170 (1º)
- APAC: Familia 164 (1º) - Familia 77 (3º) - Familia 133 (7º)

De la pequeña lista anterior se ve que que la Familia 77 genera muchas ganancias en 3 regiones, pero son muchas también las pérdidas que sufre. Por otro lado, a pesar de no generar tantas ganancias como la Familia 77, la Familia 133 también está presente en 3 regiones distintas y más aún las pérdidas que tiene, respecto a sus ganancias siempre son mínimas, por ahora la familia 133 tiene ventaja a convertirse en líder del mercado en cada una de las regiones en las que se encuentra, siendo solamente 3 regiones.

Puede parecer que tengamos una preferencia hacia la Familia 133, sin embargo estamos buscando un líder que genere ganancias con pocas pérdidas. Notamos que dentro de las figuras mostradas con anterioridad hay otras familias que tienen pocas o cero pérdidas, sin embargo las ganancias generadas están a otra escala, con lo cual no llegan a tener el nivel de mención como se esperaría.

Más allá de lo mencionado anteriormente, hay nuevas familias de productos que aparecen en estas regiones y es importante saber que dependiendo la región están generando ganancias, como en la región de América, siendo la Familia 80 la cual generó una ganancia de \$131 849 700.00 lo destacable es que no tuvo pérdida, por otro lado en la región de Medio Este, tenemos a la Familia 170 que con una ganancia de \$1 350 963.20, lastimosamente tiene pérdidas (\$8 732 547.20), que son bastante considerables.

En la siguiente sección analizaremos qué pasa con las cantidades de ventas que se generan en cada región y si esto nos permite llegar a una conclusión de quién es líder de mercado en cada una de las regiones.

### 5.2.2. Cantidades de Ventas

Veamos si en esta sección podemos tener un panorama de cómo se establecen definitivamente los líderes para cada región. Para ello presentaremos figuras que corresponden nuevamente al TOP 10 de cantidad de ventas por cada región.



### Porcentaje de Ventas - Región América

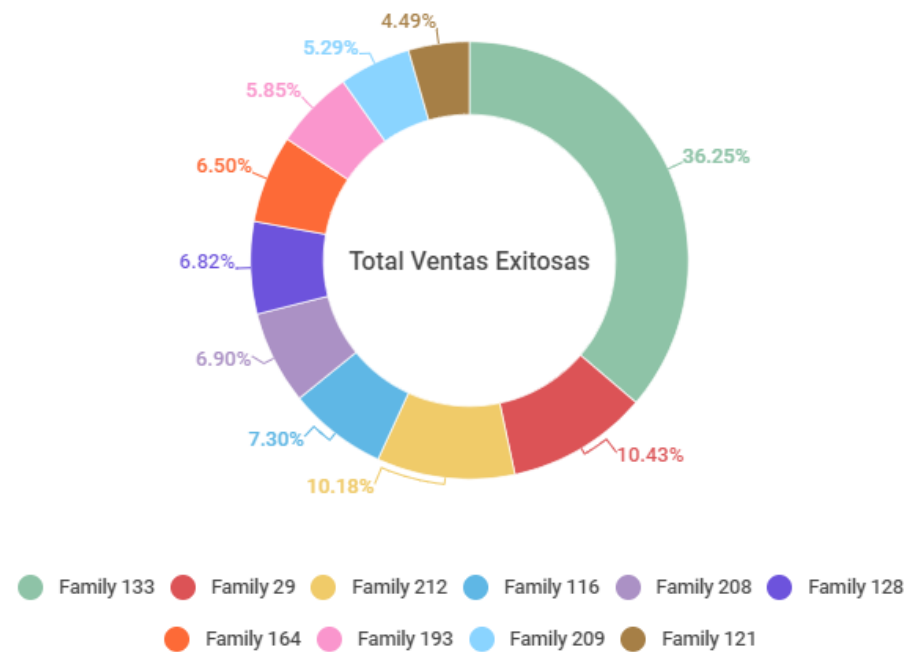


Figura 43: TOP 10 Cantidad de Ventas - Región América

### Porcentaje de Ventas - Región APAC

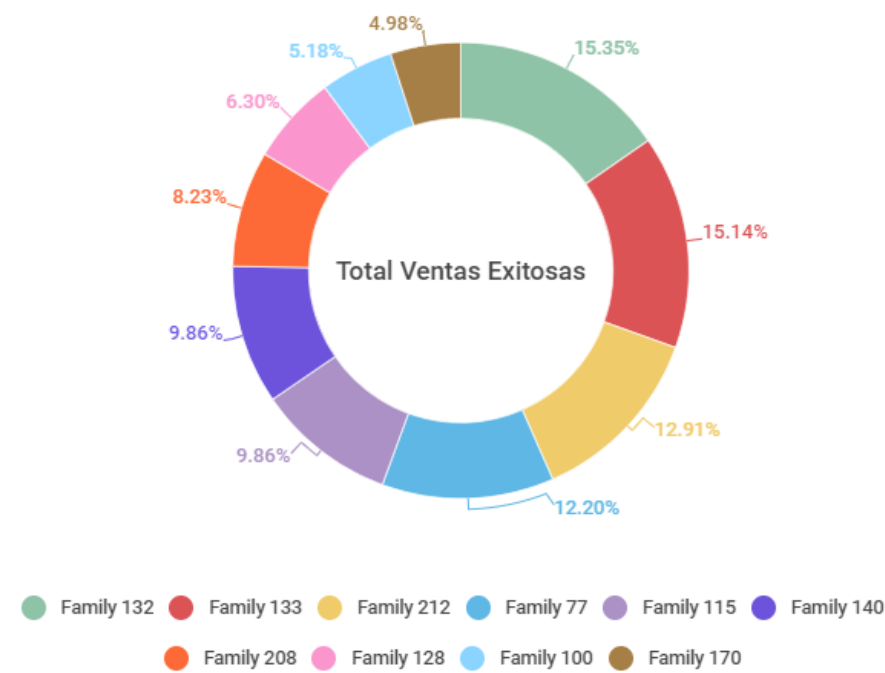


Figura 44: TOP 10 Cantidad de Ventas - Región APAC

Porcentaje de Ventas - Región Japan

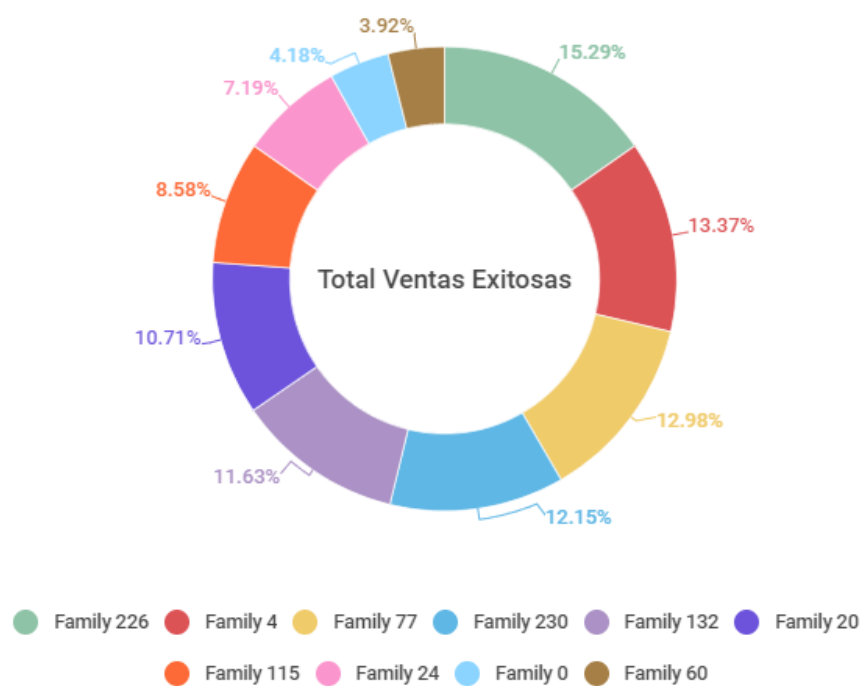


Figura 45: TOP 10 Cantidad de Ventas - Región Japan

Porcentaje de Ventas - Región EMEA

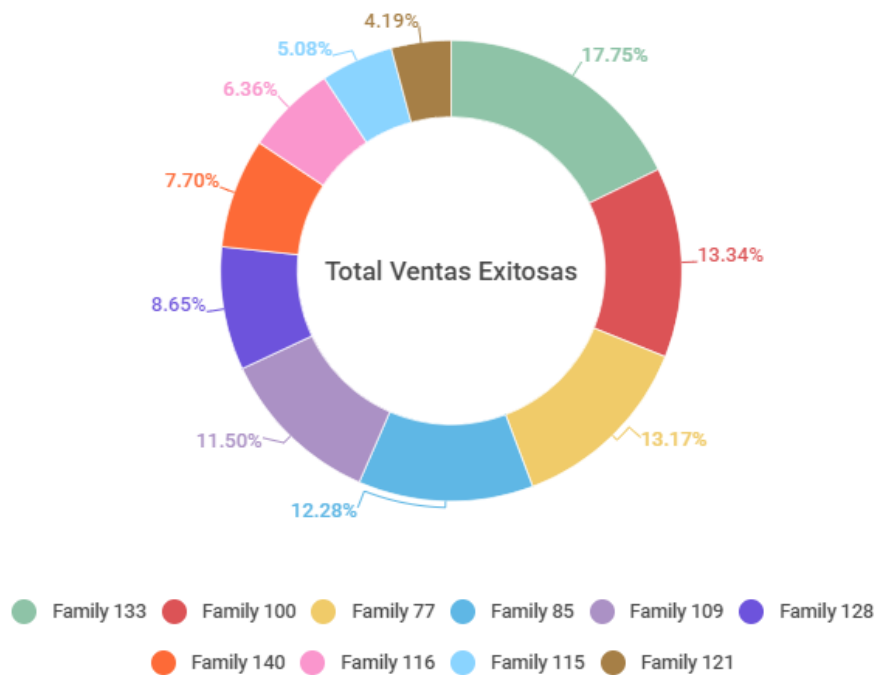


Figura 46: TOP 10 Cantidad de Ventas - Región EMEA

## Porcentaje de Ventas - Región Middle East

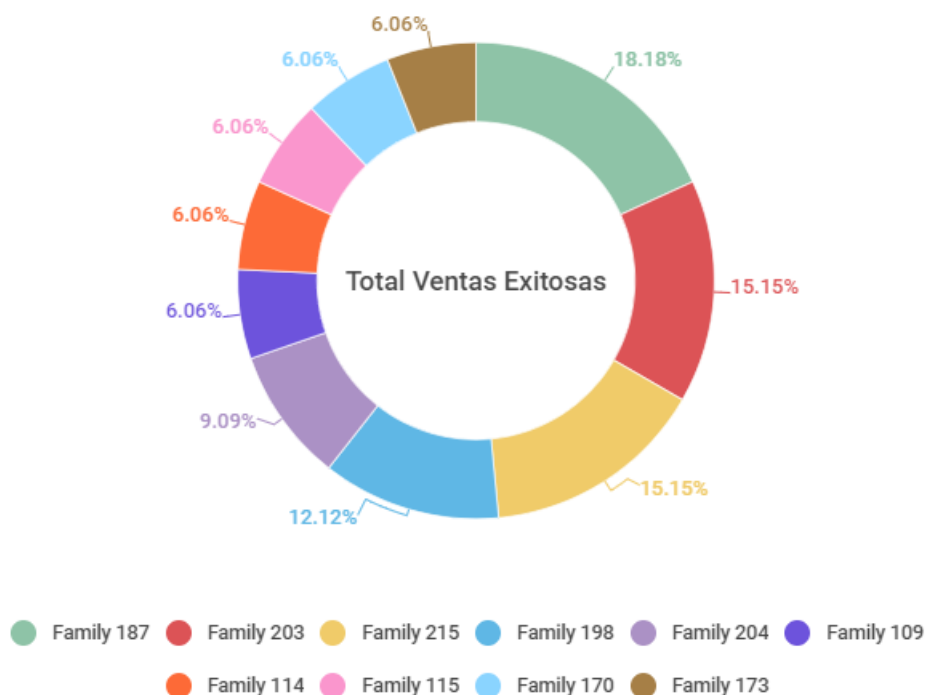


Figura 47: TOP 10 Cantidad de Ventas - Región Medio Este

Para poder llegar a una conclusión que pueda satisfacer todos nuestros criterios de análisis, hablaremos un poco más en detalle de lo que sucedió en cada una de la regiones.

- **Región América:** dijimos ya que la Familia 133 cumplía con nuestros criterios de análisis siendo uno de los candidatos para líder en 3 regiones, esta es una de ellas, con el 36.25 % de la cantidad de ventas dentro de nuestro TOP 10. Claramente esta familia es el líder del mercado regional.
- **Región APAC:** En esta región tenemos un líder en cuanto a cantidad de ventas, la Familia 132, con un 15.35 %, seguido muy cerca por la Familia 133, que encabeza nuestra lista de TOP 10. Sin embargo, como ya habíamos dicho anteriormente no debemos sacar conclusiones apresuradas, dado que la Familia 132 tiene ganancias y pérdidas aceptables a nuestro criterio, con lo cual, el líder de esta región es en definitiva la Familia 132.
- **Región Japan:** Para esta región encabezando el TOP tenemos a la familia 226 con un 15.29 % de ventas exitosas, sin embargo no aparece como una familia que haya generado ganancias a destacar. Por otro lado, otra familia que sí apareció es la ya mencionada Familia 77 en el 3º lugar del TOP con un 12.98 %, con esto podemos darla como ganadora de líder de ventas en esta región.
- **Región EMEA:** Nuevamente encontramos que la Familia 133 encabeza el TOP de ventas con 17.75 % en otra región; también aparece nuevamente la Familia 77 con un 13.17 % de ventas exitosas. Sin embargo, al analizar los gráficos para llegar a una conclusión de quien es el líder de ventas en esta región, nos encontramos con otra familia de la cual

debemos hacer una importante mención, la Familia 128 con el 8.65 % de ventas exitosas. Si bien es cierto que destaca por la poca diferencia entre pérdida y ganancia que tuvo en sus ventas mostrado en la figura 40, esto no le sirvió del todo para llevarse el liderato de esta región, debido al poco porcentaje de ventas exitosas que tuvo. Para terminar con esta región, debemos entonces entregar ese liderato a la Familia 133 nuevamente.

- Región Medio Este: Esta es una región con familias bastante alejadas de las mencionadas con anterioridad. Aquí debemos destacar la presencia de quien encabeza nuestro TOP de cantidad de ventas, la Familia 187 que con 18.18 % de ventas exitosas logró esta posición, sumado a que ya se posicionaba en 2º lugar de nuestro TOP de ganancias (ver Figura 42), la podemos considerar como líder de ventas de esta región.

### 5.3. Conclusiones

Después de todo este análisis, podemos decir que no hay un único líder, si bien es cierto que a nivel global la Familia 133 destaca de manera importante tanto por la ganancias generadas y las cantidades de ventas exitosas que tuvo. De la misma a nivel regional, dentro de las regiones de América y EMEA, no debemos dejar de nombrar a otras familias que realmente destacaron tanto global como regionalmente, como la Familia 77, que era el TOP 1 en ganancias globales pero sus pérdidas no compensaban dichas ganancias y a su vez es líder en la región de Japan, la Familia 80, que fue la 3º familia con mayor ganancias (ver Figura 35) que a su vez destacó de la misma forma en la región de América siendo el TOP 1 de ganancias generadas y aun así esto no le alcanzó para ser el líder en dicha región, y la Familia 187, que se convirtió en el líder de ventas en la región de Medio Este.

Con todo esto, podemos concluir que no existe un líder absoluto de ventas, dado que hay distintos factores que contribuyen al posicionamiento de una familia de productos en el mercado; uno de estos factores se puede ver reflejado a nivel regional, donde vemos mucha variación con respecto a las ganancias que se generaron (sin contar las pérdidas que hubieron), así como también a la cantidad de ventas que se hicieron, esto hace destacar a familias de productos a nivel regional, pero no a nivel global. Claro que esto no aplica a todas las familias de productos, porque hubo familias que claramente destacaron en ambos casos (global y regional), pero nos permitió ver claramente como se distribuyeron las ventas.

## 6. Aprobación

Algunas oportunidades necesitan aprobación especial de su precio total y los términos de entrega. Realizamos un análisis sobre este índice, investigando qué varía en oportunidades que necesitan aprobación, cuáles de ellas fueron aprobadas y finalmente, cuántas se ganaron y cuántas se perdieron. Para lograr unos resultados más reales, para cada cálculo se agrupó cada una de las ventas por oportunidad, de manera que una sola oportunidad vale como una y no vale más si se trata de una oportunidad con más productos. También, se ignoraron los resultados de 'Stage' que no son 'Closed Won' o 'Closed Lost' ya que son muy pocos (ver Figura 1) y entorpecen los datos más que dar información adicional.

### 6.1. ¿Cuántas oportunidades necesitan aprobación?

Se observa en el siguiente gráfico que lo más común en una oportunidad es que necesite aprobación. Aproximadamente, 3 de cada 4 oportunidades necesitan aprobación. De las 9841 oportunidades, solamente 2327 no necesitaron aprobación. Las otras 7514 oportunidades necesitaron ser aprobadas.



Figura 48: Porcentaje de Oportunidades que necesitan aprobación

Realizamos un análisis de oportunidades que necesitan aprobación por precio total que tiene. Vemos como las oportunidades más caras son las que siempre necesitan aprobación. Aquellas oportunidades que no necesitan aprobación no pasan de los 10 millones de dólares. Todas las otras oportunidades mayores a este precio necesitan aprobación. Sin embargo, también hay muchas oportunidades con precios bajos que necesitan aprobación. Tener un precio bajo no define la necesidad de aprobación, pero tener un precio alto sí.

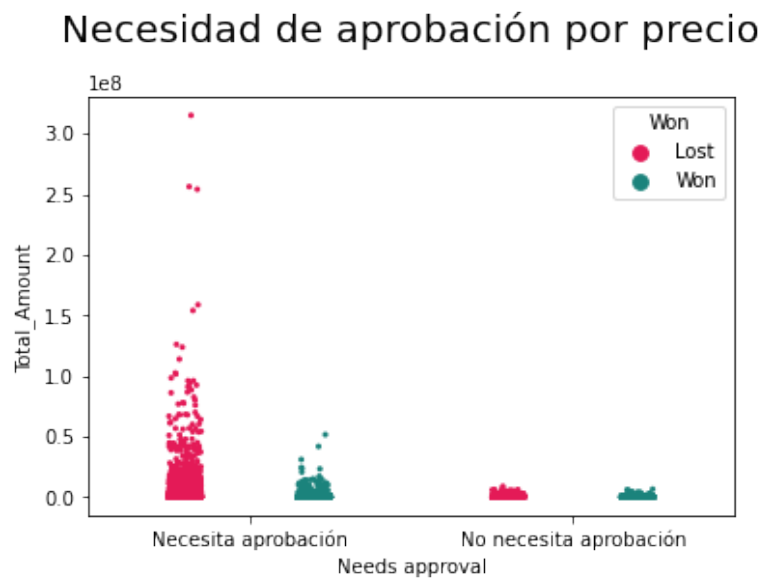


Figura 49: Distribución de necesidad de aprobación por precio en 100 millones de dolares

## 6.2. Oportunidades que no necesitan aprobación

Algunas oportunidades no requieren de una aprobación especial. Analizamos estas oportunidades y su tendencia a ser una oportunidad ganada o perdida. El siguiente gráfico muestra como esa tendencia apunta hacia el lado de las ganancias. De las 2327 oportunidades, 1447 son ganancias, mientras que 880 son pérdidas.

### Oportunidades que no necesitan aprobación

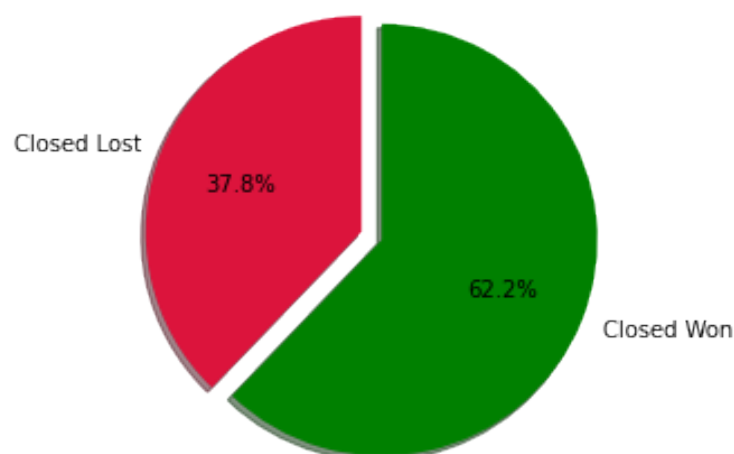


Figura 50: Porcentaje de ganancias en las oportunidades que no necesitan aprobación

Podríamos aquí obtener una conclusión un poco temprana, la cual muestra que aquellas oportunidades que no necesitan aprobación tienden a ser oportunidades ganadas. El hecho de no necesitar aprobación podría darnos una primera idea de si la oportunidad será una ganancia o una pérdida. Seguiremos con el análisis para poder fundamentar nuestra primer hipótesis.

### 6.2.1. Análisis por región

Analizamos el porcentaje de ganancia en cada región por separado, para aquellas oportunidades las cuales no necesitan aprobación. Lo primero que podemos ver es que la región Middle East no está presente. Esto quiere decir que la región Middle East no tiene ninguna oportunidad que no necesite aprobación, todas sus oportunidades precisaron de ella. Lo segundo que podemos ver es que las regiones APAC, Americas y EMEA se mantienen muy cerca del 62 %, siendo esta última la única que baja de esta marca hasta el 58 %. En cambio, Japón supera ampliamente la tasa general y la supera por casi 18 puntos, llegando a 79,8 %.

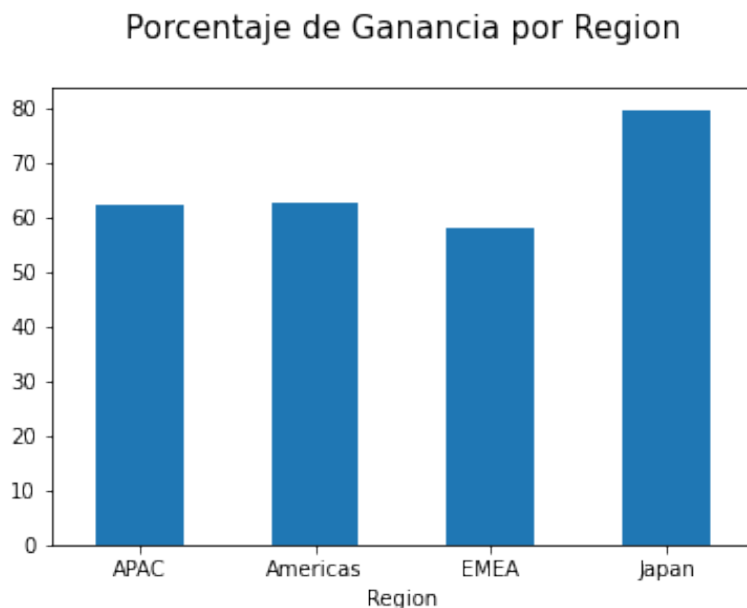


Figura 51: Porcentaje de ganancias por región

Nuestra primera hipótesis se mantiene en pie. Cuando una oportunidad no necesita aprobación, es más probable que termine siendo ganancia. Ahora también, sabemos que esa probabilidad varía en la región en donde ocurre la oportunidad. En APAC, Americas y EMEA la probabilidad es cercana a 0.6, cuando en Japón la probabilidad es mucho más alta. Es decir, cuando una oportunidad no necesite aprobación, con fijarnos en la región obtendremos mucha mas información. Al mismo tiempo, si la oportunidad proviene de la región Middle East, no tendremos información fundamentada sobre esa oportunidad. Podríamos asumir que el porcentaje será similar al resto de las regiones, ya que no se ve tanta diferencia entre las mismas, pero no podemos afirmarlo con certeza.

### 6.2.2. Análisis por precio

Las oportunidades que no necesitan aprobación son de un precio menor comparado al resto. Todas las oportunidades están a menos de 10 millones de dolares. En el gráfico vemos como se distribuye el estado final con relación al precio. Recordemos que la gran mayoría de las oportunidades tienen un precio entre 0 y medio millón de dolares. Ambas columnas tienen muchos casos entre estos precios. La gran diferencia se ve cuando los precios pasan el medio millón. Cuando esto ocurre, se ve una densidad mucho mayor de puntos en la columna de pérdidas que en la columna de las ganancias.

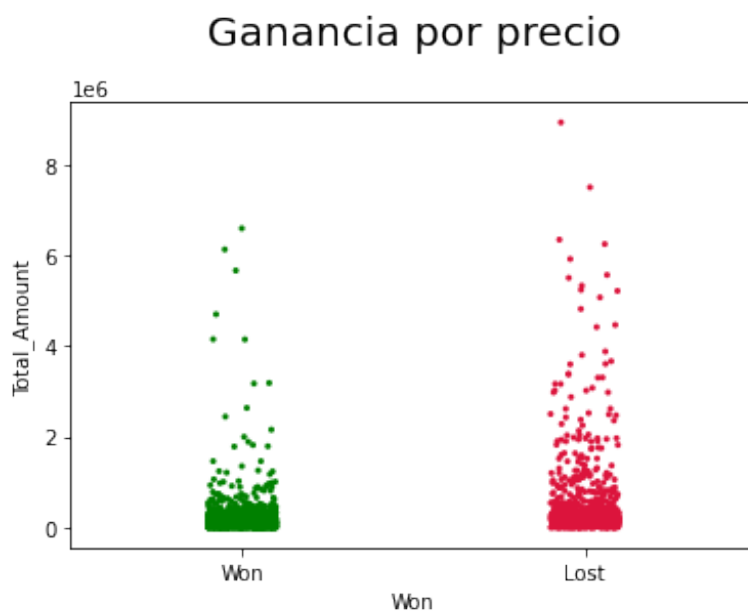


Figura 52: Densidad de ganancia y pérdida por precio en millones de dolares

Esto muestra como cuando un precio es más caro, tiende a ser una pérdida. Esta es una conclusión que hicimos en general, independientemente de su necesidad de aprobación, por lo que no logramos obtener nuevas conclusiones de este análisis, solamente reforzar las conclusiones previamente realizadas.

### 6.3. Oportunidades que necesitan aprobación

Finalizado el análisis de las oportunidades que no necesitan aprobación, seguiremos con las oportunidades que sí necesitan. Estas oportunidades pueden luego ser aprobadas o no. Una vez aprobadas o desaprobadas, también tienen la posibilidad de ser ganancias o pérdidas. Analizaremos entonces la probabilidad de que sean aprobadas, y la probabilidad de que estas luego terminen siendo ganancias. El siguiente gráfico muestra la probabilidad de que una oportunidad sea aprobada. De las 7514 oportunidades, solo 2098 no fueron aprobadas. Las otras 5416 se aprobaron. Es decir, cuando una oportunidad necesita aprobación, es aproximadamente 3 veces más probable que se apruebe a que se desapruebe.



### Porcentaje de aprobación

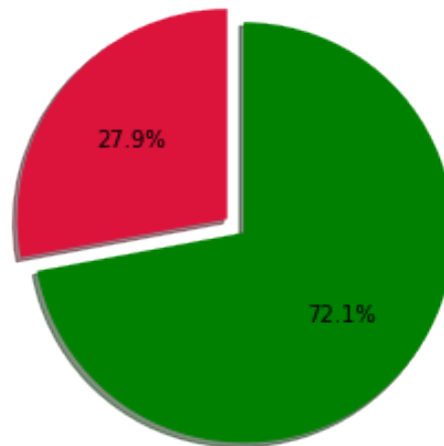


Figura 53: Porcentaje de aprobación para oportunidades que las necesitan

Al mismo tiempo, estas oportunidades que necesitan aprobación, también obtienen un valor de ganancia o pérdida. Vemos en el siguiente gráfico como esta vez la mayoría de las oportunidades son pérdidas. De las 7514, 3839 fueron pérdidas, mientras que 3625 fueron ganadas. Las restantes 50 se encuentran en cualquiera de los otros valores de 'Stage'. Aquí no podemos hacer muchas conclusiones, ya que la diferencia es muy pequeña y es más parecido a ser equiprobable la ganancia y la pérdida.

### Oportunidades que necesitan aprobación

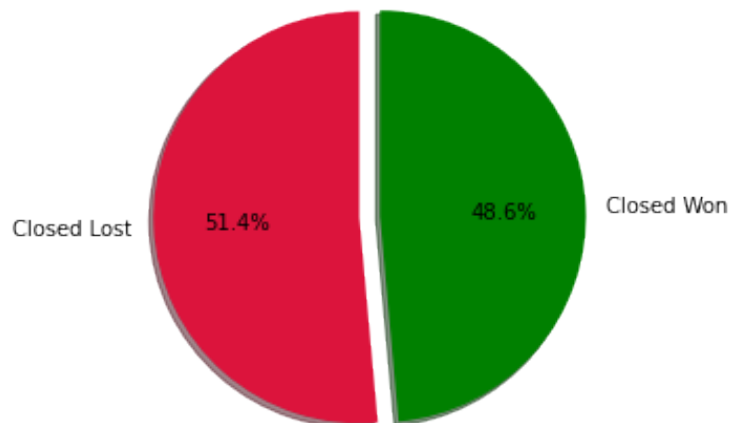


Figura 54: Porcentaje de ganancia para oportunidades que necesitan aprobación

Por más que no obtengamos ninguna conclusión hasta el momento, ahondaremos en el análisis de estas oportunidades. Nos preguntamos si tiene alguna relevancia la aprobación de la oportunidad con su estado final. El siguiente gráfico muestra como, a pesar de que ganancias y pérdidas tengan aproximadamente los mismos números, vemos como la gran mayoría de las oportunidades ganadas provienen de las oportunidades aprobadas, mientras que muy pocas provienen de las oportunidades desaprobadas. En cuanto a las oportunidades aprobadas, 3066

fueron ganadas y 2318 fueron perdidas. Por más que haya una leve tendencia a que estas oportunidades sean ganadas, no podemos tener ninguna conclusión muy fuerte. Sin embargo, de las oportunidades desaprobadas, 1521 fueron perdidas, mientras que solamente 559 fueron ganadas. Aquí sí podemos obtener una conclusión más fuerte, la cual nos indica que cuando una oportunidad es rechazada, lo más probable es que termine siendo una pérdida.



Figura 55: Ganancia de oportunidad según su aprobación

Hasta el momento, pudimos concluir que, cuando una oportunidad es desaprobada, lo más probable es que termine siendo una pérdida. También concluimos que cuando una oportunidad es aprobada, tiende levemente a ser ganancia, aunque esta conclusión no tiene la misma fuerza que la primera. Para profundizar más en el análisis, agregamos como variable la región y el monto total de la oportunidad.

### 6.3.1. Análisis por región

Haremos un análisis sobre como varía en su resultado final la aprobación de la oportunidad en cada una de las regiones. Separaremos el análisis en las oportunidades que fueron aprobadas y las que fueron desaprobadas ya que en el estudio anterior vimos como cuando una oportunidad es desaprobada, lo mas probable es que termine siendo una perdida. Vemos en el gráfico que esta hipótesis se mantiene para cada una de las regiones, en algunas hay mas diferencia que en otras, pero en todas es mas probable obtener una ganancia cuando la oportunidad se aprueba que cuando no.

## Porcentaje de Ganancia por Aprobación por Region

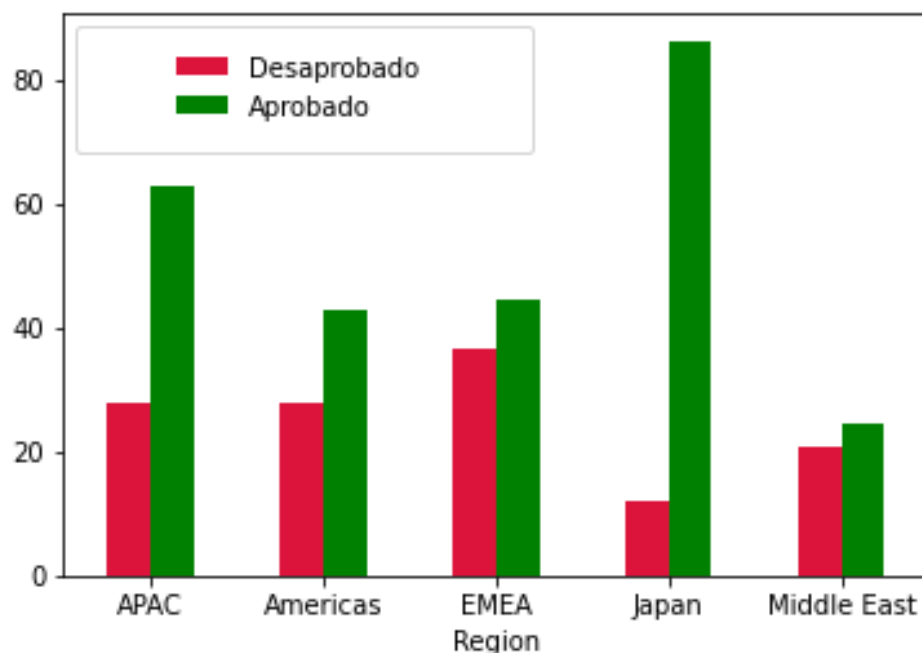


Figura 56: Ganancia de oportunidad según su aprobación por región

Haremos un análisis de cada una de las regiones:

- APAC: Es una de las regiones en donde vemos más diferencia entre las oportunidades aprobadas y desaprobadas. De las 976 oportunidades aprobadas, 614 fueron ganadas. Es decir, el 63 % de las oportunidades aprobadas son ganancias. Con un porcentaje de ganancia de 28 %, las 423 oportunidades desaprobadas, solamente 118 fueron ganancia.
- Americas: Vemos una diferencia respecto a la aprobación de las oportunidades, aunque menor a la que veíamos en APAC. Las oportunidades desaprobadas muestran un porcentaje de ganancia de 28 %, con 171 oportunidades ganadas de 619. Las oportunidades aprobadas superan en porcentaje y en cantidad a las desaprobadas con 43 % y 484 de 1133. Ambos porcentajes son menores al 50 %, lo que muestra que en esta región, cuando una oportunidad necesita aprobación, lo más probable es que termine siendo una pérdida. Si también tenemos en cuenta la aprobación de la misma, podemos tener una aproximación mucho más cercana sobre lo que va a ser su estado final.
- EMEA: Tiene números muy parecidos a Americas. El porcentaje de ganancia de aquellas oportunidades que fueron aprobadas solo difiere en un 2 %. Sin embargo, el porcentaje de ganancia de oportunidades desaprobadas es mayor, con un porcentaje de 37 %. EMEA es la región en la que más oportunidades necesitan aprobación, con un total de 2458 oportunidades. Entre estas, 1881 fueron aprobadas, de las cuales se ganaron 837, y 577 desaprobadas, de las cuales se ganaron 212. En esta región, al igual que en Americas, ambos porcentajes son menores a 50 %, aunque están más cerca de esta marca que los porcentajes de Americas, por lo que una oportunidad que necesita aprobación tiende a ser pérdida, aunque más débilmente.

- Japan: Aquí podemos ver la mayor diferencia. Las oportunidades aprobadas y desaprobadas tienen una diferencia de 74 %, con 86 % y 12 % respectivamente. En esta región podemos obtener conclusiones muy fuertes. Cuando una oportunidad necesita aprobación, su aprobación tiene gran influencia en su estado final. Cuando una oportunidad es desaprobada, es muy probable que termine siendo una pérdida. De los 437 casos, solamente 53 pudieron ser ganancia. Al mismo tiempo, si una oportunidad es aprobada, lo más probable es que termine siendo una ganancia. De los 1280 casos, solo 177 fueron pérdidas. Esto quiere decir que a la hora de analizar una oportunidad de Japón, es muy importante verificar su estado de aprobación en caso de que necesite, ya que su aprobación influye de gran manera en su estado final.
- Middle East: Esta región tiene el peor porcentaje de ganancia para oportunidades aprobadas, con un 25 %. Las oportunidades desaprobadas tienen un porcentaje de ganancia de 21 %. Estos números muestran como cuando una oportunidad necesita aprobación, lo mas probable es que no sea una ganancia. De todas formas, la cantidad de oportunidades que hay en Middle East son muy pocas comparando con el resto de las regiones, por lo que quita fuerza a esta conclusión. Solamente 138 en total, con 24 desaprobadas y 114 aprobadas.

### 6.3.2. Análisis por precio

Hicimos un análisis sobre aquellas oportunidades que necesitan aprobación, cómo variaba su estado final por su precio y por si fueron aprobadas o no. Como podemos ver en el gráfico, cuando una oportunidad es aprobada, no tiene mucha relación con su precio, ya que las columnas de Aprobado y Desaprobado son muy similares. Nuevamente, podemos ver que la mayoría de las oportunidades con precios elevados respecto de las demás, son oportunidades que se pierden, dándole más fuerza a nuestra previa conclusión.

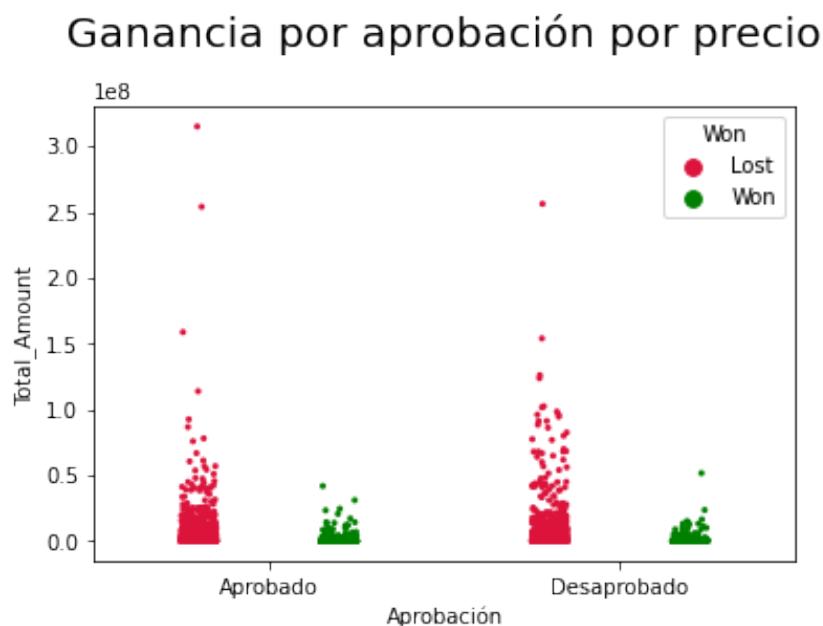


Figura 57: Densidad de ganancia por aprobación por precio en 100 millones de dólares

## 7. Potencia

Nos interesa analizar algunos datos sobre la Potencia de un producto, a fin de extraer algunas conclusiones. Trataremos de llegar a estas conclusiones mediante una serie de cálculos y gráficos.

Las hipótesis iniciales que hacemos son que todos los productos de la empresa tendrán una potencia asociada, ya que la empresa vende Aires Acondicionados. Además, suponemos que el precio de un producto varía de forma directamente proporcional con su potencia.

Como siempre, trabajamos con precios en dólares. Las potencias están medidas en TRF (Toneladas de refrigeración). No discriminamos por estado de la venta en este análisis, ya que nos importa el precio que tuvo el producto en esa oportunidad, no el estado final de la operación.

### 7.1. Análisis inicial de los datos

Lo primero que notamos analizando los datos es que hay muchas filas que tienen potencia cero. No suponemos esto un error, sino que simplemente se trata de productos en los que la potencia no aplica.

Por ejemplo: Un aire acondicionado tiene potencia, un soporte para ponerlo en la pared no la tiene. Asumimos entonces que la empresa vende productos que no tienen potencia asociada, por lo que descartamos una de las hipótesis planteadas al inicio de la sección. Nos interesan todas las ventas de la empresa, por lo tanto, no removemos aquellas sin potencia.

Algunos datos sueltos sobre la Potencia son:

- La potencia máxima de un producto de la empresa es de 500 TRF.
- Por supuesto la potencia mínima es 0, pero si pedimos un dato mayor a 0, es decir un producto con potencia, la mínima potencia es 1 TRF.
- El promedio de potencias si tenemos en cuenta todos los productos, incluyendo aquellos de "potencia cero", es aproximadamente 2.3 TRF.

### 7.2. Si el producto es más caro, ¿es mas potente?

Ya analizado el set, queremos saber cómo están relacionados el precio de un producto y su potencia. Como se mencionó, esperamos que los más caros sean los más potentes.

Directamente graficamos la potencia y el precio para quizás sacar una rápida conclusión e irnos temprano a almorzar.

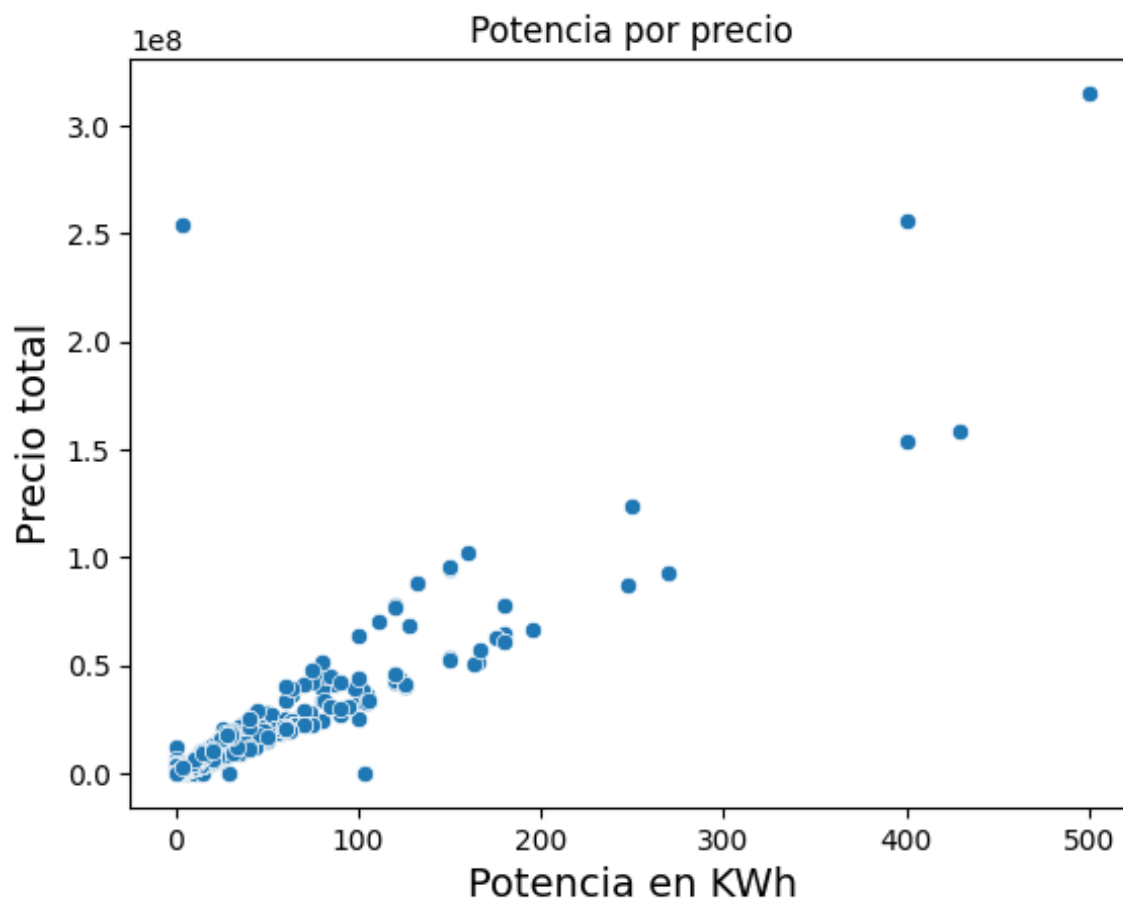


Figura 58: Potencia por precio, en cientos de millones de dólares

Tal como esperábamos en nuestra hipótesis inicial, efectivamente de esta visualización podemos inferir que, en líneas generales, los productos aumentan su precio según la potencia. Es decir, a mayor potencia, mayor precio, y además tienen una relación prácticamente lineal.

También vemos que la mayor cantidad de ventas se agrupan en torno a una potencia de hasta 200 TRF, siendo su punto de mayor densidad el de 0 TRF.

Sin embargo, también notamos algún producto de relativa baja potencia con precio muy alto. Al ser un solo producto, no nos importa mucho por lo que descartamos su análisis.

Respondemos entonces la pregunta que tiene la Subsección por nombre afirmando que Sí, los productos más potentes, son más caros, y los menos potentes, más baratos.

### 7.3. ¿Cuan potentes son los productos potentes?

La Subsección anterior fue muy gratificante en cuanto a conclusiones. Rápidamente detectamos que la potencia y el precio están relacionados. Sin embargo, hay algo no nos deja conciliar el sueño: en nuestra anterior visualización notamos que la mayor parte de los productos se agrupan en torno a... potencia cero.

Enseguida vemos cuantos productos tiene potencia asociada, y la respuesta es impactante: tan solo 1/3 de los productos de la empresa tienen una potencia asociada. Lo que termina de echar por los suelos nuestra hipótesis inicial de que todos los productos de la empresa tienen una potencia.

Como aun nos interesa analizar la distribución de potencia de nuestros productos, removemos del próximo análisis las potencias iguales a cero, es decir nos quedamos con productos con potencia asociada. Ahora calculamos nuevamente el promedio de potencia y la respuesta es aproximadamente 7 TRF.

De la gráfica de la Subsección anterior también notamos que la gran mayoría de potencias se agrupan entre cero (descartado) y cien TRF. Por ello solo analizaremos productos con potencias en ese intervalo, es decir

$$0 < Pot(TRF) \leq 100$$

Descartando entonces aquellos productos que no cumplen la inecuación anterior notamos que solos nos quedamos con aproximadamente 1/3 de los productos totales de la empresa, que también era el resultado que obtuvimos cuando descartamos solo aquellos de potencia cero, y por ello decimos ahora que muy pocos productos son los que tienen potencia mayor a 100 TRF.

Ahora nos preguntamos, dentro de los productos con potencias entre cero y cien, como se distribuye la misma. Para responder realizamos un gráfico de densidad.

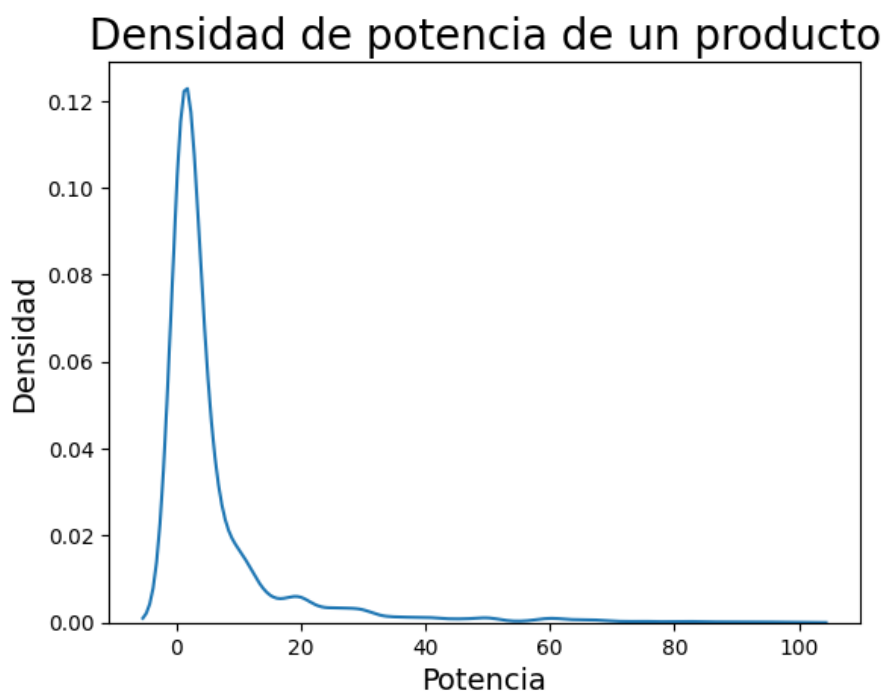


Figura 59: Densidad de potencia

De este nuevo gráfico inferimos lo esperado: la mayor densidad de potencias se agrupa en torno al cero, es decir con baja potencia cercana al cero. Recordemos que habíamos quitado las potencias cero para este análisis, y también las mayores a cien. ¿Por qué esperamos este resultado? Porque cuando calculamos el promedio, éste nos había dado aproximadamente 7, es decir bastante más cercano al 0 que al 500, que era la potencia máxima.

Achicamos aún mas el intervalo y calculamos el mismo grafico para

$$0 < Pot(TRF) \leq 10$$

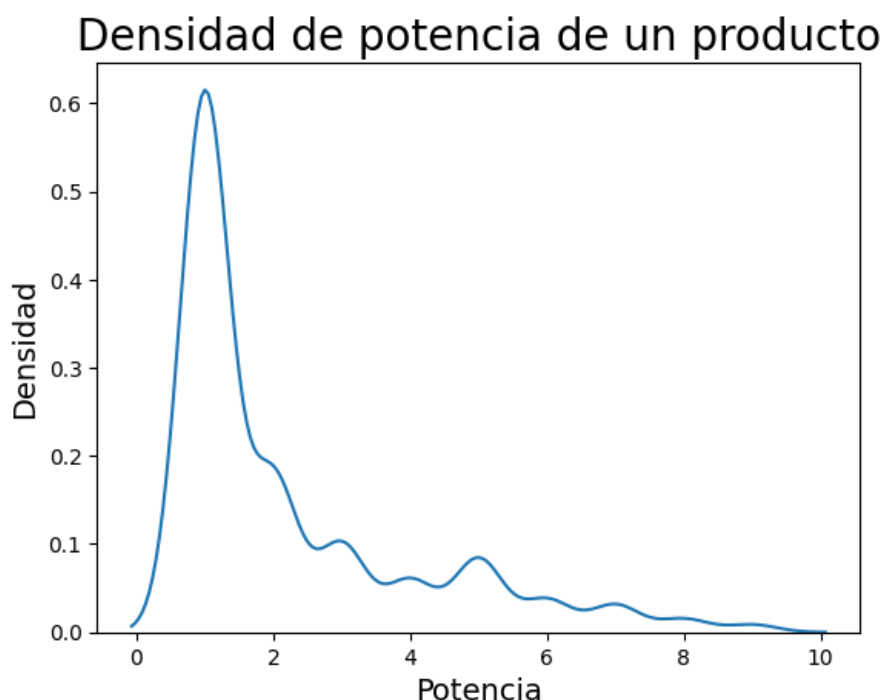


Figura 60: Densidad de potencia

De este gráfico mucho más pequeño en escala, vemos que la mayoría de las potencias toma como valor 1 TRF, por lo que nuestra conclusión final es que la mayoría de los productos con potencia asociada tienen baja potencia.

Es fácil notar que más de la mitad de los productos con potencias distintas de cero tienen potencia = 1 TRF.

## 7.4. Conclusiones finales

Al principio de la sección tenemos dos hipótesis.

La primera era que todos los productos de la empresa tendrán una potencia asociada, hipótesis que descartamos rápidamente, y no solo eso, sino que además logramos descubrir que las potencias de aquellos productos que tienen una... son muy bajas. De hecho, la mayoría tienen la potencia máxima, que es 1 TRF. Aproximadamente  $1/3$  de los productos tienen potencia cero,  $1/3$  tienen potencia 1 TRF, y el  $1/3$  restante tienen potencias entre 2 y 500.

La segunda hipótesis que tenemos era que la potencia aumentaba con el precio. Esta hipótesis fue rápida y fácilmente demostrada. Agregamos una última gráfica de esto que se ve con mucha claridad.



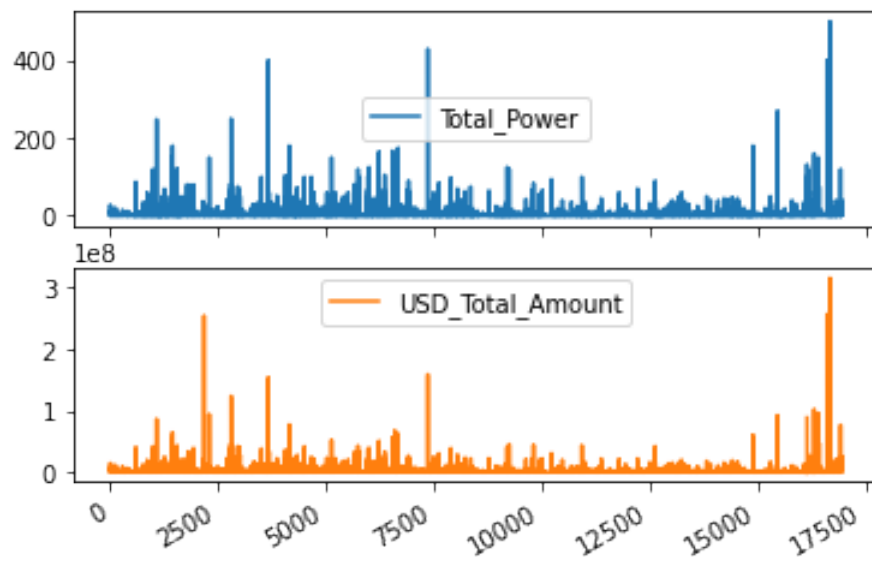


Figura 61: Potencia y precio, notar que son distintas escalas

## 8. Source

### 8.1. ¿Qué son las Sources?

La Source es la fuente de creación de la oportunidad. Se puede entender, por ejemplo, como la campaña que origina dicha oportunidad. Siguiendo con este ejemplo, sus valores posibles podrían ser publicidad física, publicidad digital, recomendación, contacto por terceros, entre otras.

En nuestro set de datos existen oportunidades que pueden tomar uno de 13 valores posibles de Sources declaradas, y otro tanto con sources No declaradas. Se muestra a continuación un gráfico que indica el porcentaje entre fuentes declaradas y no declaradas:

### Porcentaje de oportunidades con sources declaradas y no declaradas

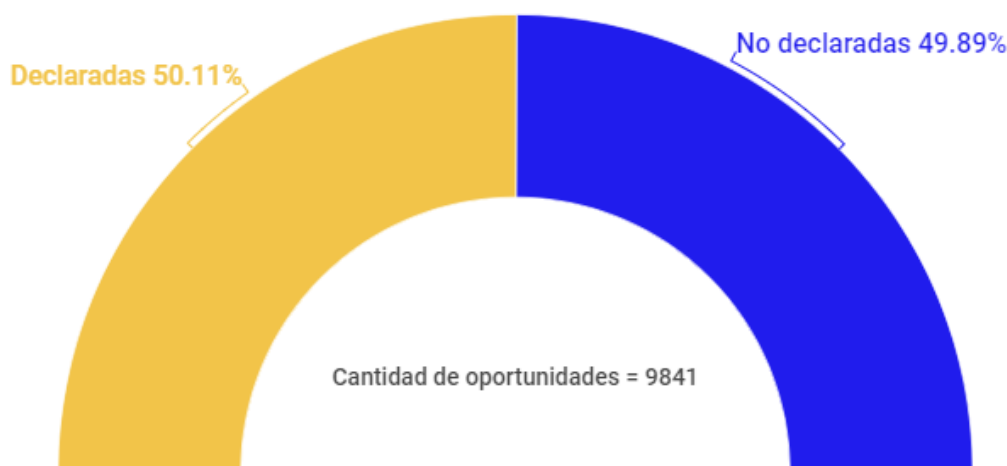


Figura 62: Oportunidades con sources declaradas y no declaradas

### 8.2. Sources Declaradas

A partir de ahora vamos a trabajar con aquellas fuentes declaradas para así poder analizar su importancia e impacto en las ventas.

#### 8.2.1. Contra cantidad de oportunidades

Primero analizaremos cuántas oportunidades trae cada fuente, y a partir de ello determinaremos cuál es la que más oportunidades aporta. Para ello, primero verificamos que cada oportunidad proviene de una sola Source.

La siguiente figura es una tabla donde, ordenado de mayor a menor, se muestran la cantidad de oportunidades que corresponden a cada Source.

	Cantidad_de_oportunidades
Source_7	2069
Source_9	873
Source_11	847
Source_13	536
Source_3	305
Source_10	89
Source_1	88
Source_2	68
Source_4	32
Source_6	17
Source_0	4
Source_12	2
Source_5	1

Figura 63: Cantidad de oportunidades por Source declarada

Se puede ver que es Source 7 la que mayor proporción de oportunidades trae.

### 8.2.2. Contra Stage

Ahora vamos a analizar si hay alguna fuente que tenga más tendencia a aportar ventas ganadas. Para eso primero debemos realizar una serie de observaciones previas.

En primer lugar, verificamos que cada oportunidad corresponde a un solo Stage. Luego, descompondremos los datos de la tabla de la Figura 63 para separar, para cada Source, las oportunidades según su Stage. Este análisis se puede observar a continuación:

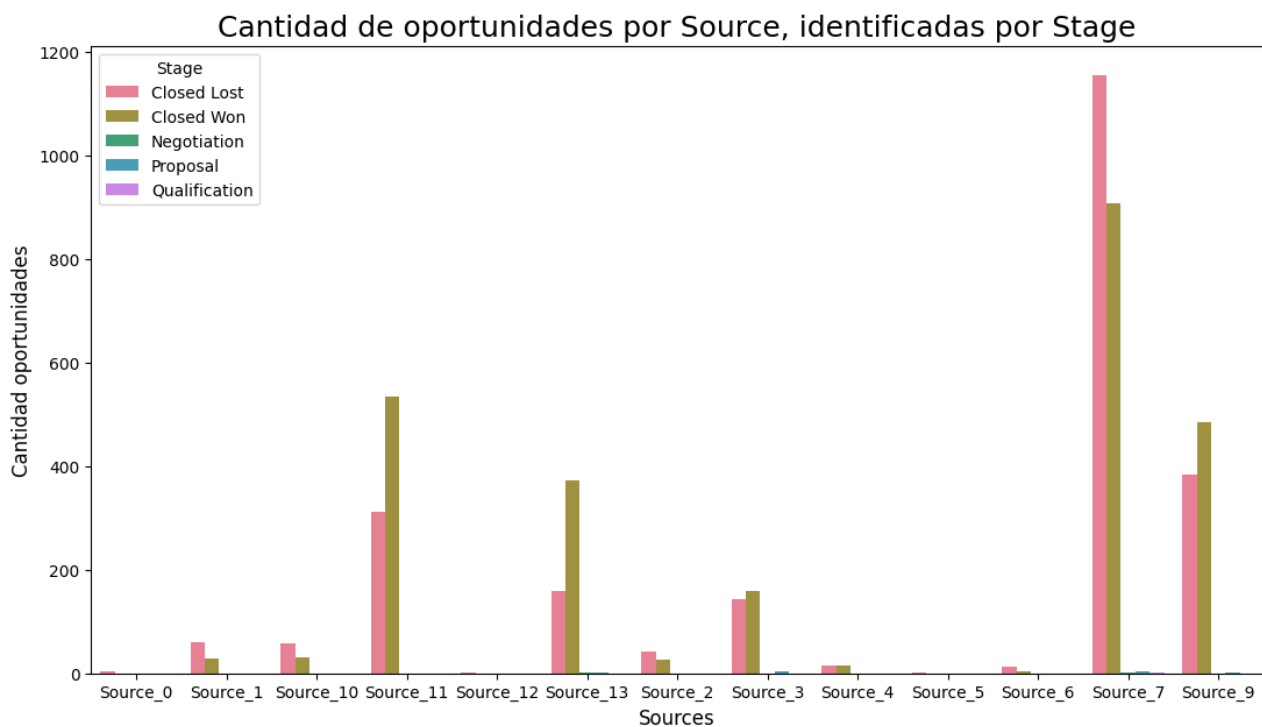


Figura 64: Cantidad de oportunidades por Source, identificadas por Stage

De acuerdo con lo visto en la Figura 1, ya vimos que la cantidad de oportunidades en las categorías de Negotiation, Proposal y Qualification son muy inferiores. Por lo tanto, nos centraremos en ver las cantidades para los estados de Closed Won y Closed Lost.

Podemos ver como si bien la Source 7 es la que más oportunidades aporta, la mayoría de ellas son perdidas, por lo tanto, su calidad de oportunidades no es tan buena. Pasa lo contrario en casos como Source 9, Source 11 o Source 13, que, siendo las siguientes fuentes con más oportunidades aportadas, tienen más oportunidades Closed Won que Closed Lost.

Nos enfocamos ahora en las Closed Won, que son las oportunidades que verdaderamente aportan a la empresa, ya que son las que permiten el ingreso de dinero.

La figura a continuación muestra, de mayor a menor, las cantidades de oportunidades ganadas para las distintas Sources.

### Porcentajes de cantidad Closed Won de cada Source

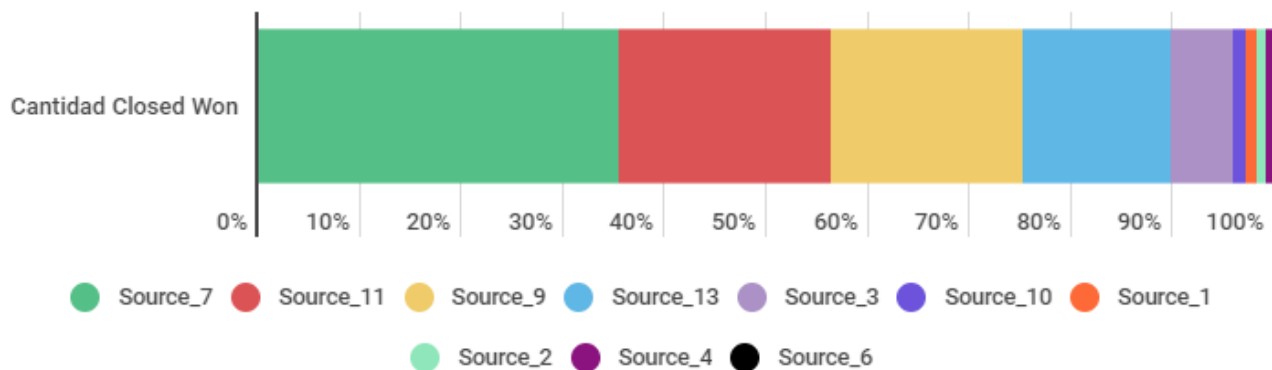


Figura 65: Porcentajes de cantidad de oportunidades Closed Won para cada Source

Se puede contemplar que, por bastante diferencia, Source 7 (35.34 %) es la fuente con más ventas ganadas, cuando la siguiente fuente con mayor cantidad de Closed Won es Source 11 con el 20.86 %; mientras que Source 6 es la que menos, con el 0.19 %.

#### 8.2.3. Conta el Dinero aportado

En esta sección veremos cuánto dinero aporta cada source, es decir, el total de los montos de las oportunidades ganadas.

### Dinero aportado por Source

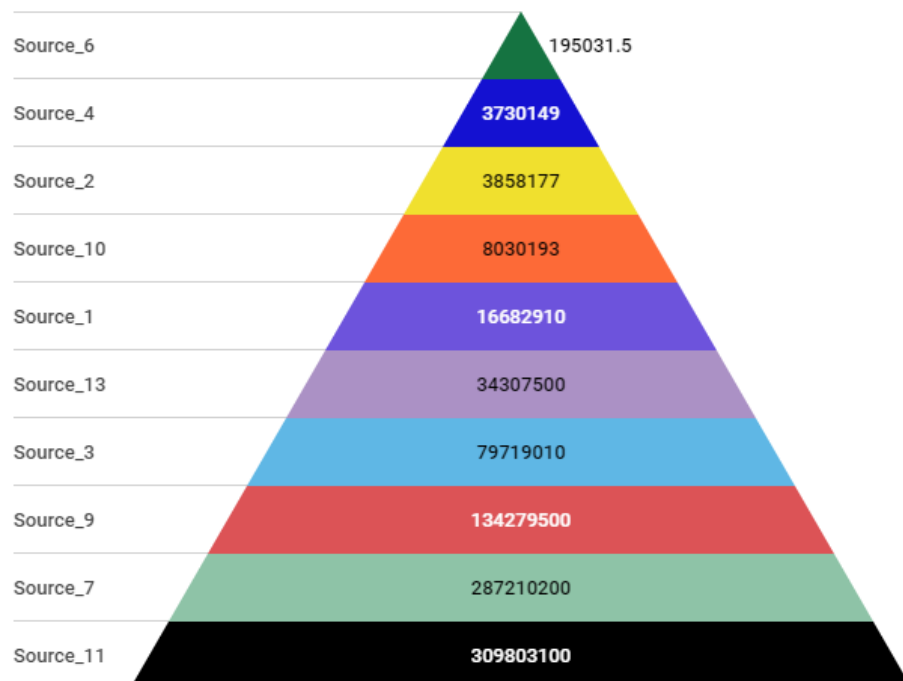


Figura 66: Dinero aportado por Source

Ahí se puede ver que es Source 11 la que más dinero aporta a la empresa con el 35.29 %. Le sigue Source 7 con 287210200 dólares aportados, que representan el 32.72 %. En los últimos escalones de la pirámide tenemos las fuentes con menos aportes, Source 4 y Source 6, con el 0.42 % y el 0.02 % respectivamente.

#### 8.2.4. Uniendo todo

En el siguiente diagrama se unirán los ítems profundizados antes, comparando el dinero ganado a partir de cada source contra la cantidad de oportunidades ganadas aportadas por cada source:

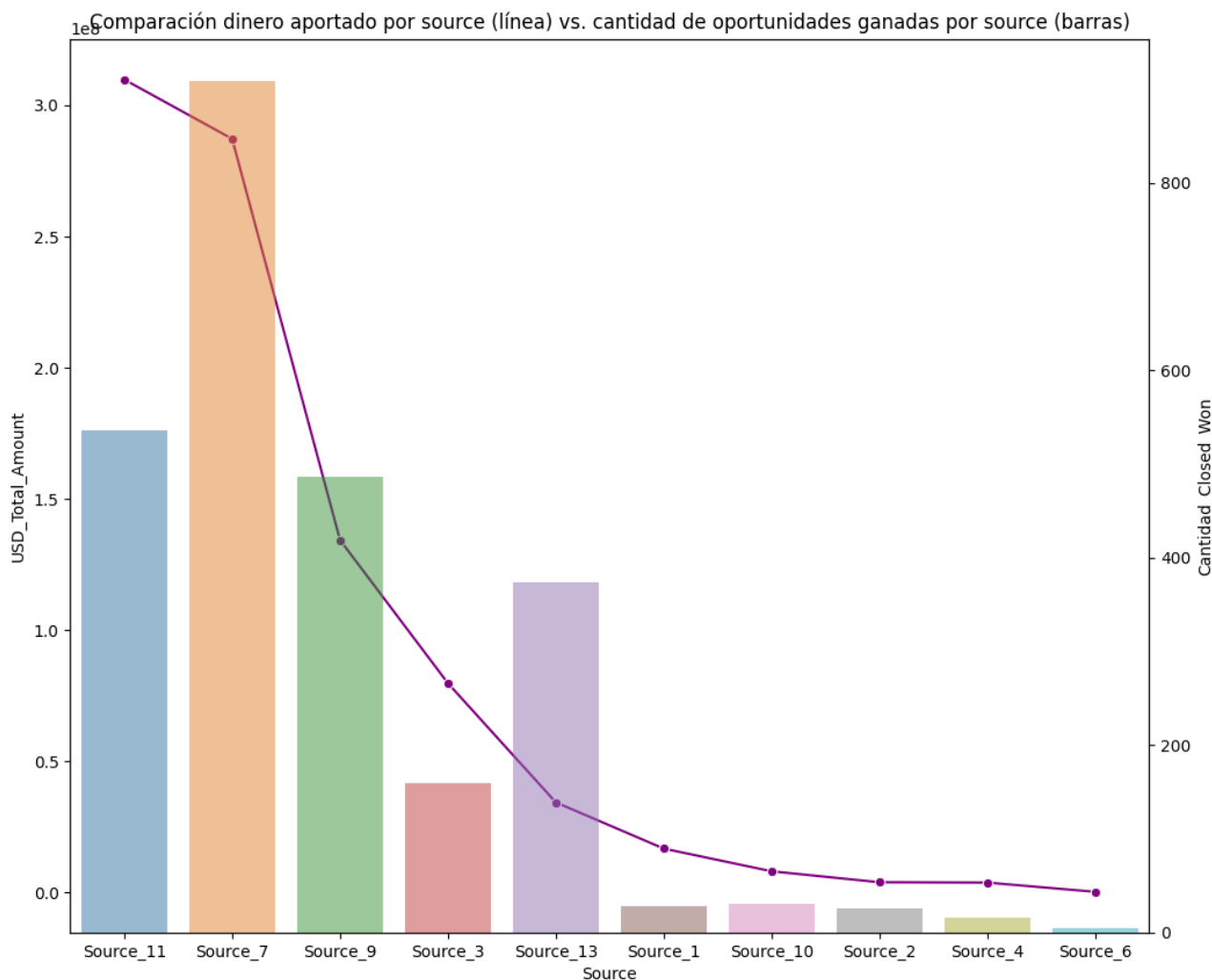


Figura 67: Comparación dinero aportado por source contra cantidad de oportunidades ganadas por source

En este gráfico vemos que a pesar de que la Source 11 aporta menos cantidad de ventas ganadas, el caudal de dinero aportado es mayor, es decir que pese a ser menos ventas, aporta más plata. En cambio, Source 7 es la fuente que más aporta en cantidad de ventas Closed Won, pero no es la que más dinero aporta a la empresa. Esto está relacionado con la calidad de ventas de Source 7, mencionado luego de la Figura 64.

Algo similar pasa con Source 3, que si bien no aporta mucha cantidad de ventas ganadas, aporta más dinero que otra fuente como Source 13, que si bien aporta mayor cantidad de ventas, también aporta menor cantidad de dinero.

Sobre las últimas fuentes del gráfico, al mismo tiempo que desciende la cantidad de oportunidades, desciende el dinero aportado.

## 9. Productos

En esta sección, haremos un análisis de cada producto individualmente. Analizaremos cuáles fueron los productos más vendidos, y así mismo, un análisis de sus porcentajes de éxito. A priori, esperamos que los productos más vendidos tengan una relación con las familias más vendidas, por ejemplo, que los top 5 productos más vendidos pertenezcan a las top 5 familias más vendidas.

### 9.1. Productos más vendidos

Como dijimos anteriormente, esperamos que los productos más vendidos pertenezcan a las familias más vendidas. En el gráfico podemos ver cuales son los productos más vendidos, empezando por Product 226 con 677 ventas, 100 más que su seguidor más cercano. El quinto producto más vendido es el Product 491, con 351 ventas, casi la mitad que las del Product 226.

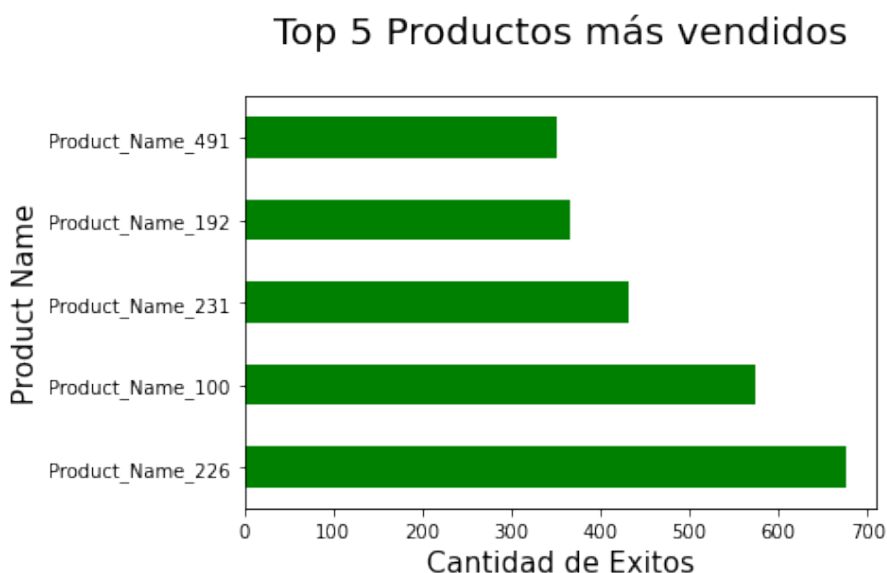


Figura 68: Top 5 productos con mayor cantidad de éxitos

Los productos más vendidos pertenecen a las siguientes familias:

- Product 226 pertenece a la Family 133
- Product 100 pertenece a la Family 77
- Product 231 pertenece a la Family 132
- Product 192 pertenece a la Family 115
- Product 491 pertenece a la Family 226

Como era de esperarse, las 5 familias pertenecen a las familias más vendidas.



## 9.2. Los productos más vendidos, ¿son los más exitosos?

Haremos un análisis de porcentaje de éxito de los productos. Queremos ver por qué los productos más vendidos son los más vendidos. ¿Tienen un alto porcentaje de éxito o simplemente tuvieron más intentos de venta?

Para realizar el análisis, solo se tuvieron en cuenta aquellos productos que tienen más de 50 intentos de venta, ya que algunos productos tienen 2 ventas exitosas y 0 fracasos, dando un porcentaje de 100 %. No sería justo juzgar productos con mayor cantidad de ventas y exigirles un porcentaje similar a productos con tan pocas ventas, por lo que se eligió la barrera de 50 intentos de venta totales.

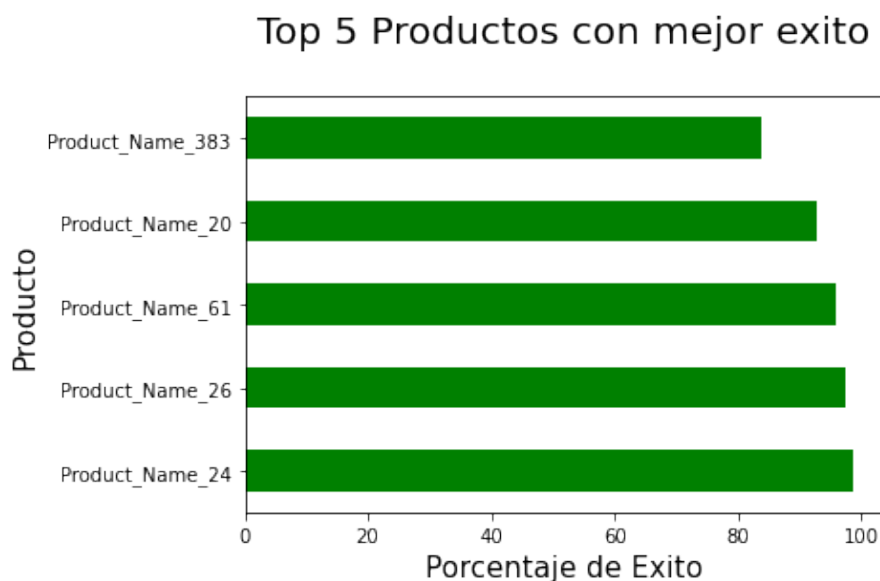


Figura 69: Top 5 productos con mayor porcentaje de éxitos

Como podemos ver, ninguno de los productos con mayor cantidad de ventas pertenece a los productos con mejor porcentaje. Esto indica que los productos con más ventas no tienen ese título por tener un buen porcentaje de éxito, sino que su alta cantidad de ventas proviene de una alta cantidad de intentos de venta.

Finalmente, haremos un análisis de aquellos productos con peor porcentaje de venta. Productos que a pesar de su gran cantidad de intentos, muy pocas veces fueron capaces de ser exitosos.

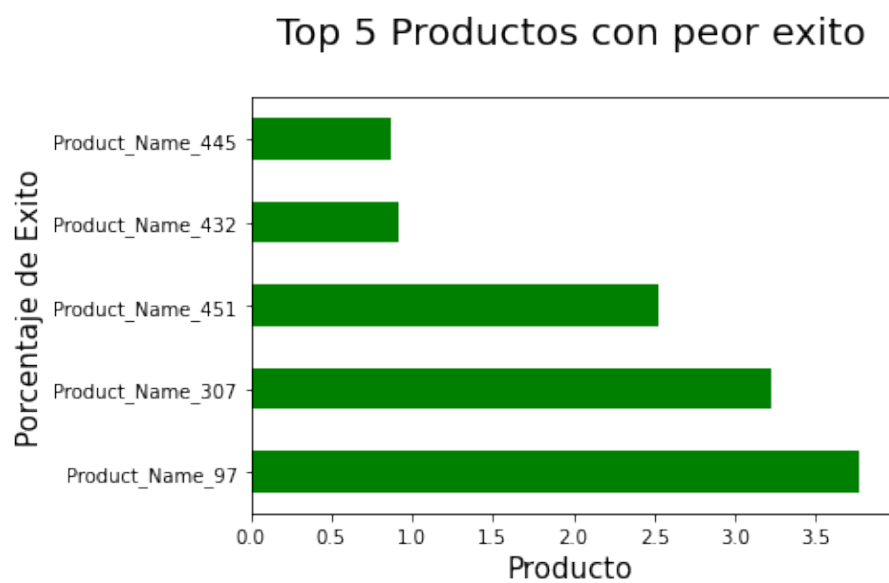


Figura 70: Top 5 productos con menor porcentaje de éxitos

## 10. Comentarios finales

El objetivo del primer trabajo práctico del curso de Organización de Datos de la Facultad de Ingeniería fue realizar, usando Python Pandas, un análisis y visualización de datos de una base de datos dada. A nivel personal consideramos que el trabajo fue altamente provechoso como primera aproximación a lo que es la Ciencia de datos, ya que utilizamos herramientas modernas de análisis de datos para extraer información y realizar consultas en lo que fue un trabajo en gran medida didáctico y que por las características del mismo nos permitió una gran libertad de acción para trabajar con el set de datos dado.

En cuanto al trabajo en sí mismo, consideramos como una ventaja el tamaño de la base de datos, en cuanto a registros y en cuanto a columnas. Tener tantas columnas nos dio la posibilidad de realizar diversos análisis que abarcaron diferentes categorías, ya expuestas en las secciones anteriores. Nos ocurrió que hubo categorías que considerábamos esenciales para un determinado análisis, y luego no lo fueron, y por el contrario hubo categorías que a priori consideramos descartar, pero más tarde nos dimos cuenta que explicaban o complementaban varios factores importantes de alguna otra categoría.

Además, la gran cantidad de columnas hacen al análisis más extenso, lo que permite ahondar mucho más profundamente en los datos. Por ejemplo, en nuestro trabajo analizamos muchas veces el comportamiento de la variable Estado. Gracias a la alta cantidad de categorías contra las que comparamos esta variable, podemos predecir más fácilmente el resultado de una negociación futura ya que contamos con un gran historial de datos de, justamente, los estados. Si no tuviéramos, por ejemplo, una categoría Clientes, no sabríamos si la persona que está interesada en hacer una compra tiene un historial que lo hace tender a comprar los productos, o a no comprarlos, o a comprarlos dependiendo la estación del año, etc.

A lo que nos referimos con el párrafo anterior es que consideramos que todos los análisis hechos en el trabajo nos dejan en una buena posición para encarar el segundo trabajo práctico de la materia que consiste en, como se comentó antes, predecir el estado de futuras negociaciones. Toda la información recabada es útil, necesaria y de gran valor para que los futuros algoritmos predictores puedan hacer correctamente su trabajo.

Por otra parte, las desventajas que encontramos también están referidas al set de datos: por un lado, había una buena cantidad de columnas que no aportaban información, sea porque aparecía un único y mismo valor en todos sus registros, porque estaban casi vacías, o incluso en algunos casos estaban completamente vacías. Sin embargo, lo que más inconvenientes trajo a la hora de analizar los datos fueron los nombres propios. Vendedores, clientes, productos, familias, sources y más tenían nombres directamente genéricos. Este asunto no perturba el análisis en sí (ya que el mismo, por supuesto, no discrimina nombres), pero sí lo vuelve algo tedioso ya que los nombres se confunden y olvidan rápidamente (de nuevo, por supuesto, al ojo humano, no al algoritmo), además de que consideramos que empeoran la calidad de los gráficos, sobre todo los que dimensionan palabras (nombres propios, en esos casos) por la cantidad de apariciones.

Bien, hasta acá llegamos. Las conclusiones de cada sección se encuentran al final de ellas, por lo que decidimos no volver a incluirlas aquí, priorizando algunos comentarios generales, sobre el trabajo y sobre los datos, que son los que expusimos en los párrafos anteriores. Así que, por el momento, esto fue todo.

El código completo que utilizamos para el desarrollo de este trabajo se encuentra en su correspondiente repositorio online, y el link para acceder al mismo está incluido en la sección 1.2 . En caso de querer mirarlo, recomendamos acompañar su lectura con la correspondiente

sección del informe a fin de facilitar la comprensión sobre lo que se está haciendo.

Por último, en nombre de todo el equipo de desarrollo, agradecemos profundamente por la atención brindada durante la lectura del presente informe.

*“In God we trust, all others must bring data” William Edwards Deming (1900 - 1993)*

**FIN**

