



75.06 ORGANIZACIÓN DE DATOS

FACULTAD DE INGENIERÍA, UNIVERSIDAD DE BUENOS AIRES

Trabajo Práctico 1: Análisis Exploratorio

Padrón	Alumno	Dirección de correo
83351	Egozcue, Sebastian	segozcue@gmail.com
81845	Funes, Federico Alejandro	fafunes@fi.uba.ar
101456	Pérez Andrade, Violeta	viperez@fi.uba.ar
99139	Sinisi, Fernando	fsinisi@fi.uba.ar

Índice

1. Introducción	2
2. Breve descripción de los datos	2
2.1. Columnas	2
2.2. Correlación entre Columnas	6
2.2.1. Correlación entre columnas de montos	7
2.2.2. Correlación entre columnas de aprobación	7
3. Análisis geográfico	8
3.1. Región	8
3.2. Territorios	14
3.3. País	17
4. Análisis temporal	19
5. Oportunidades	26
5.1. Cantidad	26
5.2. Ítem	28
5.2.1. Montos de los items	29
5.2.2. Categorías de los items	32
5.3. Montos	34
5.4. Vendedores	35
5.5. Categorías	37
6. Insights	39
6.1. Insights por región	39
6.2. Insights por moneda	41
7. Conclusiones	43

1. Introducción

En el presente informe se busca analizar los datos de la empresa "Frío Frío", dedicada a la venta e instalación de equipos de aire acondicionado para grandes superficies. Para ello, se nos proporcionó el archivo *Entrenamieto_ECI_2020.csv*. Cada fila representa un ítem vendido para una 'oportunidad' la cual 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 ser una oportunidad que resultó exitosa o no. El objetivo principal de este informe es realizar un análisis exploratorio abarcativo.

Finalmente, con lo descubierto en el análisis exploratorio se buscará llegar a conclusiones sobre las relaciones que hay entre las distintas columnas para luego en el segundo trabajo práctico poder predecir, con ayuda de este análisis, la probabilidad de que una oportunidad sea exitosa.

Link al repositorio: <https://github.com/FernandoSinisi/Orga-Datos-2C-2020>

2. Breve descripción de los datos

El set de datos brindado cuenta con 16947 registros, clasificados en 52 columnas con información de ventas comprendidas entre los años 2013 y 2018.

2.1. Columnas

De las 52 columnas, 8 corresponden a datos del tipo fecha y 3 más están relacionadas con las mismas:

Columna	Cantidad de Registros con datos validos
Account_Created_Date	16947
Opportunity_Created_Date	16947
Last_Activity	0
Quote_Expiry_Date	12322
Last_Modified_Date	16947
Planned_Delivery_Start_Date	16947
Planned_Delivery_End_Date	16872
Month	16947
Delivery_Quarter	16947
Delivery_Year	16947
Actual_Delivery_Date	0

Figura 1: Columnas con datos temporales

Las columnas que se pueden apreciar en la figura corresponden a lo siguiente:

- Account Created Date: fecha de creación de la cuenta del cliente.
- Opportunity Created Date: fecha de creación de la oportunidad comercial
- Last Activity: fecha de la última actividad. En esta columna todos los valores son *NaT*
- Quote Expiry Date: fecha de vencimiento del presupuesto
- Last Modified Date: fecha de ultima modificación en la oportunidad
- Planned Delivery Start Date: límite inferior del rango previsto para la fecha de entrega
- Planned Delivery End Date: límite superior del rango previsto para la fecha de entrega
- Month: mes-año de Planned Delivery Start Date
- Delivery Quarter: trimestre de Planned Delivery Start Date
- Delivery Year: año de Planned Delivery Start Date
- Actual Delivery Date: fecha real de la entrega. En esta columna todos los valores son *NaT*

Como podemos observar las columnas '*Actual_Delivery_Date*' y '*Last_Activity*' no tienen datos y tanto '*Planned_Delivery_End_Date*' como '*Quote_Expiry_Date*' tienen algunos datos nulos o inválidos

Otras 10 columnas están relacionadas con montos o monedas:

Columna	Cantidad de Registros con datos validos
Price	356
Currency	16947
Total_Amount	16888
Total_Amount_Currency	16947
ASP	13738
ASP_Currency	16947
ASP_(converted)	13738
ASP_(converted)_Currency	16947
Total_Taxable_Amount	16947
Total_Taxable_Amount_Currency	16947

Figura 2: Columnas con datos monetarios

Las columnas de la tabla representan:

- Price: precio. De esta columna, la mayoría de los valores son *None* (15982) *Other*(609)
- Currency: moneda. De esta columna la mayoría de los valores son *None*(16052)
- Total Amount: monto total
- Total Amount Currency: moneda del monto total
- ASP: (Average Selling Price) precio promedio a la venta
- ASP Currency: moneda del precio promedio
- ASP (converted): precio promedio a la venta convertido a otra moneda
- ASP_(converted)_Currency: moneda del precio promedio convertido en la variable
- Total_Taxable_Amount: monto gravado total
- Total_Taxable_Amount_Currency: moneda del monto gravado total

Observando la tabla podemos darnos cuenta que solo 6 de las 10 columnas tienen valores para todos los registros, quedando columnas como '*Price*' o '*Total_Amount*' con valores inválidos en registros.

Luego tenemos 3 columnas las cuales brindan información sobre la locación de la oportunidad.

Columna	Cantidad de Registros con datos validos
Region	16947
Territory	11948
Billing_Country	16920

Figura 3: Columnas con datos geográficos

Estas columnas representan:

- Region: región de la oportunidad
- Territory: territorio comercial de la oportunidad. La mayoría de los valores de esta columna son *None* (4999).
- Billing_Country: país donde se emite la factura

Como se puede observar, la única columna la cual cuenta con datos para todos los registros es la columna '*Region*'

Por último, las columnas restantes cuentan con valores categóricos, los cuales brindan información propia del ítem de la oportunidad en particular, algunas de estas son:

- Source: fuente de creación de la oportunidad. La mayoría de los valores de esta columna son *None* (9497).
- Opportunity_Name: nombre de la oportunidad
- Quote_Type: tipo de presupuesto
- Delivery_Terms: términos de entrega
- Brand: marca del producto. La mayoría de los valores son *None* (15911).
- Product_Type: tipo de producto. La mayoría de los valores son *None* (15935).
- Size: tamaño del producto. La mayoría de los valores son *None* (15967).
- Product_Family: familia de producto
- Product_Name: nombre del producto
- Prod_Category_A: categoría 'A' del producto
- Product_Category_B: categoría 'B' del producto. La mayoría de los valores son *None* (15928).
- Stage: variable target. Estado de la oportunidad

Columna	Cantidad de Registros con datos validos
Source	7450
Opportunity_Type	16947
Quote_Type	16947
Delivery_Terms	16947
Brand	1036
Product_Type	1012
Size	980
Product_Family	16947
Product_Name	16947
Prod_Category_A	16947
Product_Category_B	1019
Stage	16947

Figura 4: Columnas con datos categóricos

Podemos ver en la figura 4, que muchas de estas columnas tienen registros con valores que no son validos. Si tenemos 16947 registros y solo tenemos por ejemplo 980 con valores en la columna 'Size', vamos a tener que tenerlo muy en cuenta al realizar el análisis.

2.2. Correlación entre Columnas

En la siguiente sección se analizaran la correlación entre las columnas que tienen datos numéricos y binarios. Se dejo fuera del análisis el ID (numérico) ya que no hace referencia a una cantidad sino es un identificador de registro.

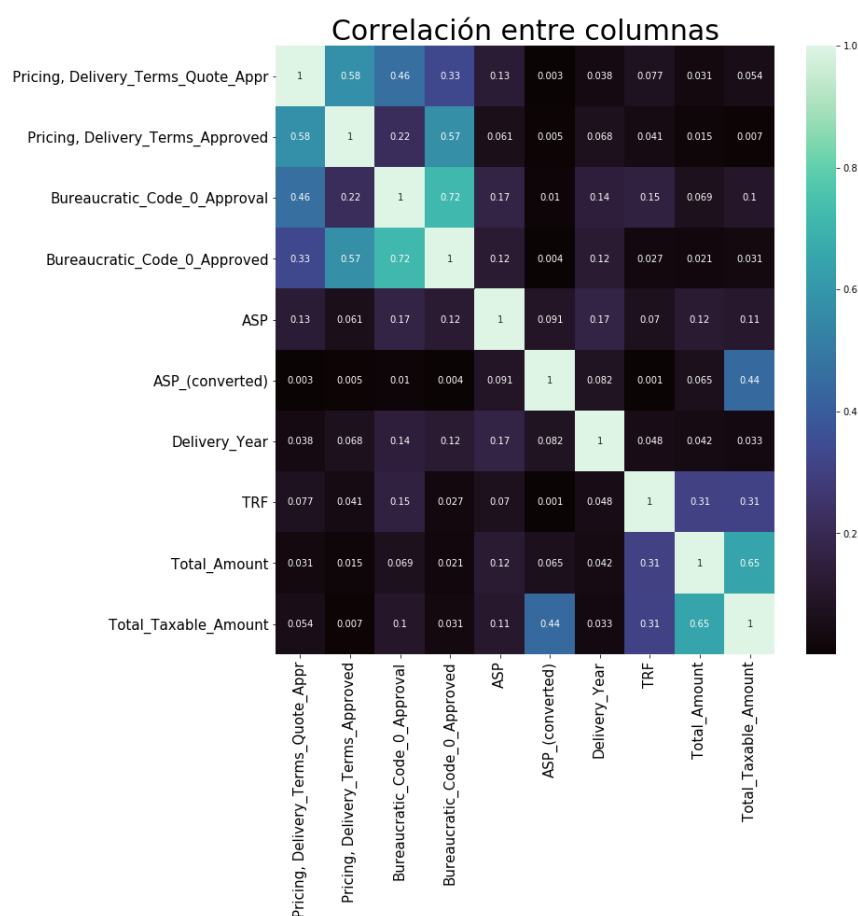


Figura 5: Correlación entre columnas

Se pueden distinguir 2 pequeños grupos de correlaciones entre todas las columnas numéricas, los cuales se analizarán a continuación:

2.2.1. Correlación entre columnas de montos

Se puede notar que existe una correlación entre las columnas *'Total_Amount'* y *'Total_Taxable_Amount'*, algo de esperarse ya que si estuvieran en la misma moneda, la suma de cada valor de la columna *'Total_Amount'* para una oportunidad debería ser el valor de la columna *'Total_Taxable_Amount'* de esa oportunidad.

También podemos observar que existe otra correlación menos fuerte entre la columna *'TRF'* y *'ASP_(converted)'* con las columnas antes mencionadas. Esto se debe a que la columna *'ASP_(converted)'* es el precio por unidad de frigoría en moneda dolar y si lo multiplicamos por los valores de la columna *'TRF'*, que representan las toneladas de frigorías obtengo el valor de *'Total_Taxable_Amount'*

2.2.2. Correlación entre columnas de aprobación

Observando la parte superior izquierda del heatmap vemos muy claramente que hay una correlación entre 4 columnas: *'Pricing_Delivery_Terms_Quote_Approval'*, *'Pricing_Delivery_Terms_Approved'*, *'Bureaucratic_Code_0_Approval'* y *'Bureaucratic_Code_0_Approved'*.

Esto era de esperarse ya que las columnas tienen valores binarios para definir si necesitan o no aprobación y si obtuvieron o no la aprobación, sería normal que si un ítem/oportunidad no necesita aprobación de algún tipo la columna de si obtuvo la aprobación sea consistente, y viceversa.

3. Análisis geográfico

En la siguiente sección se analizarán los datos de las columnas: *'Region'*, *'Territory'* y *'Billing_Country'*, que se podrían considerar relacionadas con el aspecto geográfico de la oportunidad.

3.1. Región

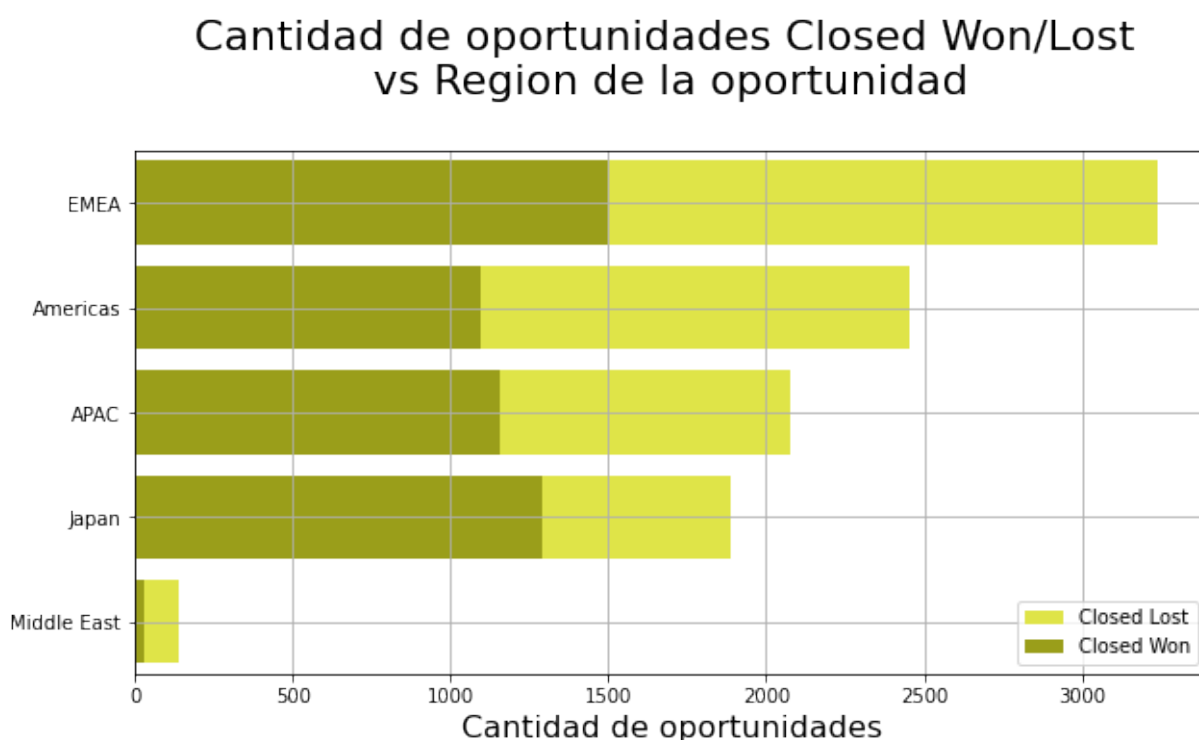


Figura 6: Closed Won por territorios

Como se puede ver en las figuras 6, tanto la región *'EMEA'* como *'Americas'* son las que más cantidad de oportunidades tienen y su proporción favorece a las pérdidas. También ocurre lo mismo con *'Middle East'* aunque en este caso son muy pocas las oportunidades como para considerarla representativa. En cambio las regiones *'APAC'* y *'Japan'* tienen una cantidad considerable de oportunidades pero su relación Closed Won y Closed Lost las favorece.

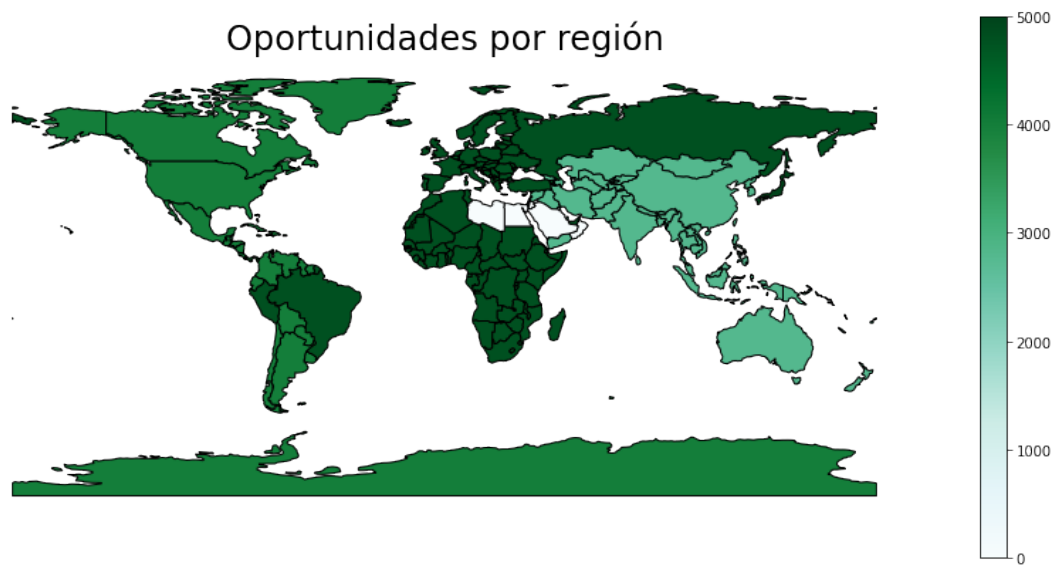


Figura 7: Distribución en el mapa de la cantidad de oportunidades

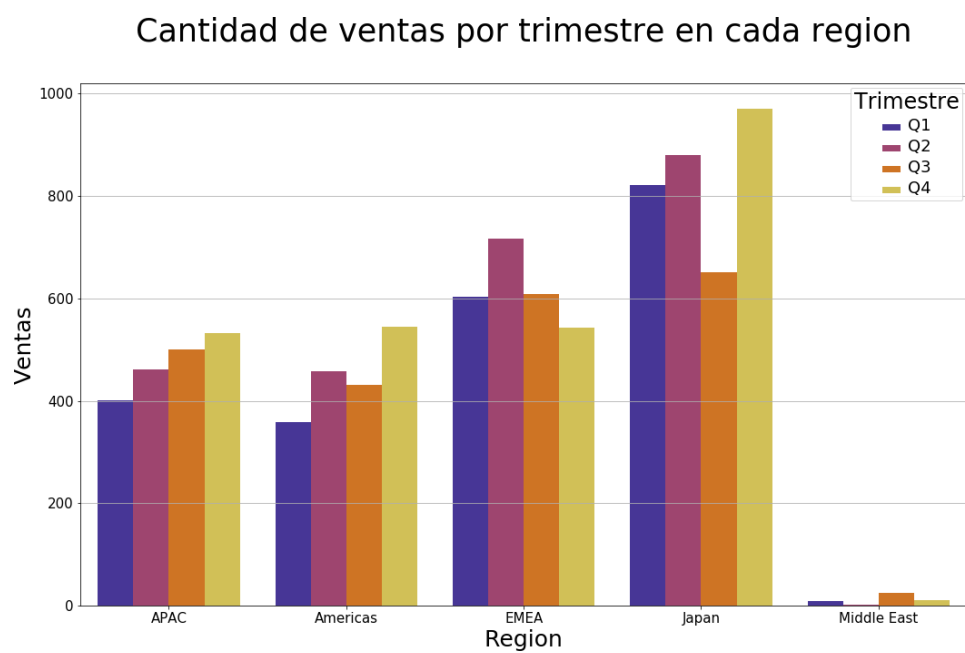


Figura 8: Ventas por trimestre de cada región

Si vemos las cantidad de ventas por región discriminadas por trimestre, vemos que en 3 de las 5 regiones el Q4 es el que mas ventas representa y que si descartamos la región 'Middle East' (por tener muy pocas ventas en comparación) en las otras 4 regiones el Q1 tiene menos ventas que el Q2.

También podemos observar que las regiones '*EMEA*' y '*Americas*' a pesar de tener mas oportunidades que la región '*Japan*' (ver figura 6) representan menos ventas o mejor dicho productos por oportunidad. A partir de esto nos preguntamos que región es mas rentable, es decir obtuvo más ganancia.

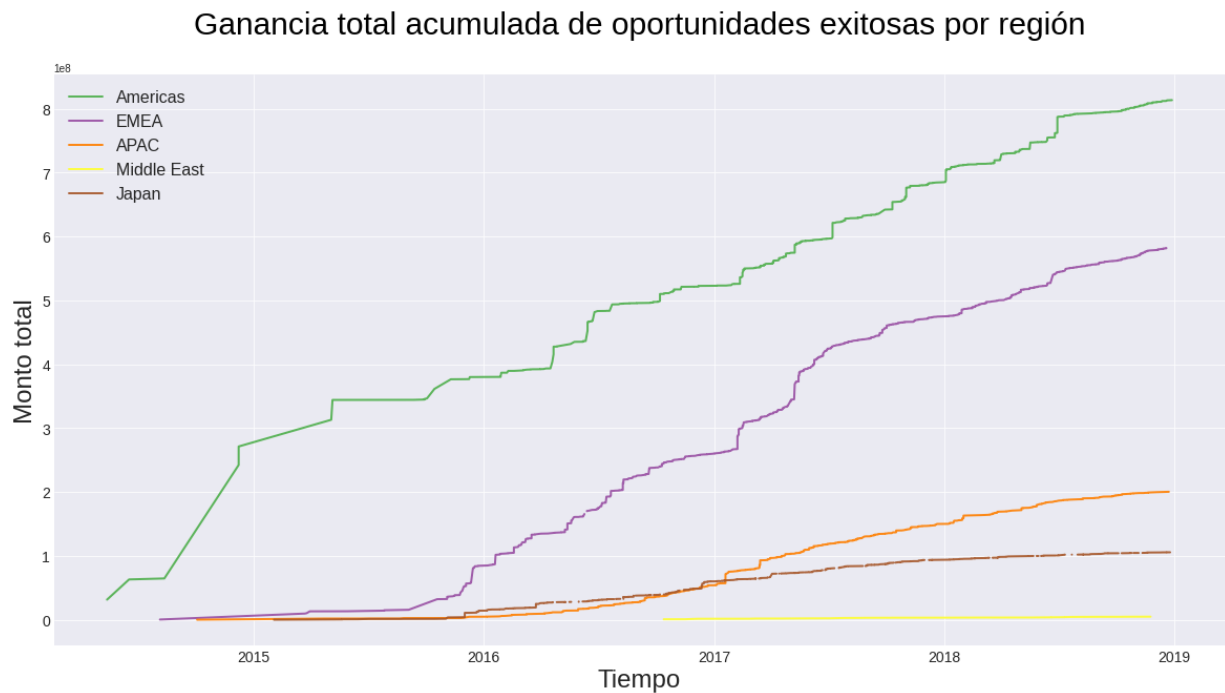


Figura 9: Ganancia neta por región

Finalmente podemos ver que si solo consideramos las oportunidades exitosas (generan ganancia) las regiones '*EMEA*' y '*Americas*' son las mejores. También podemos ver que a partir de 2017 la región *APAC* le saca un diferencia a la *Japan* que se mantiene hasta el 2019 (momento hasta el cual tenemos información).

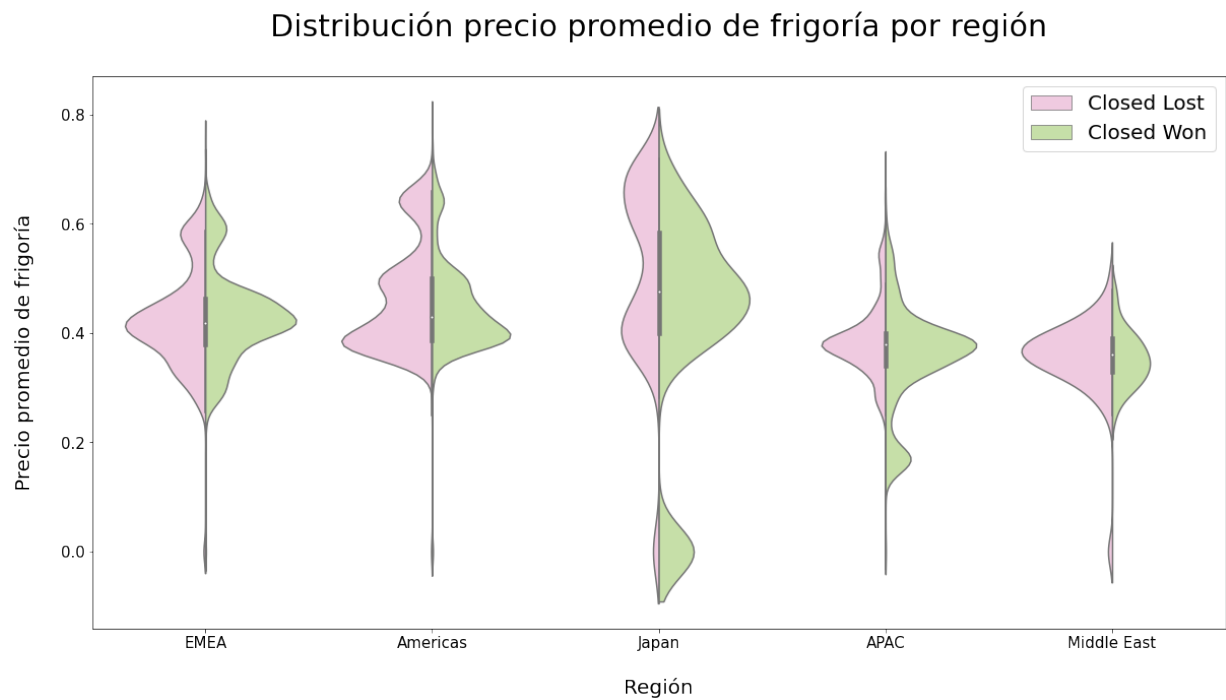


Figura 10: Distribución precio frigoría por región

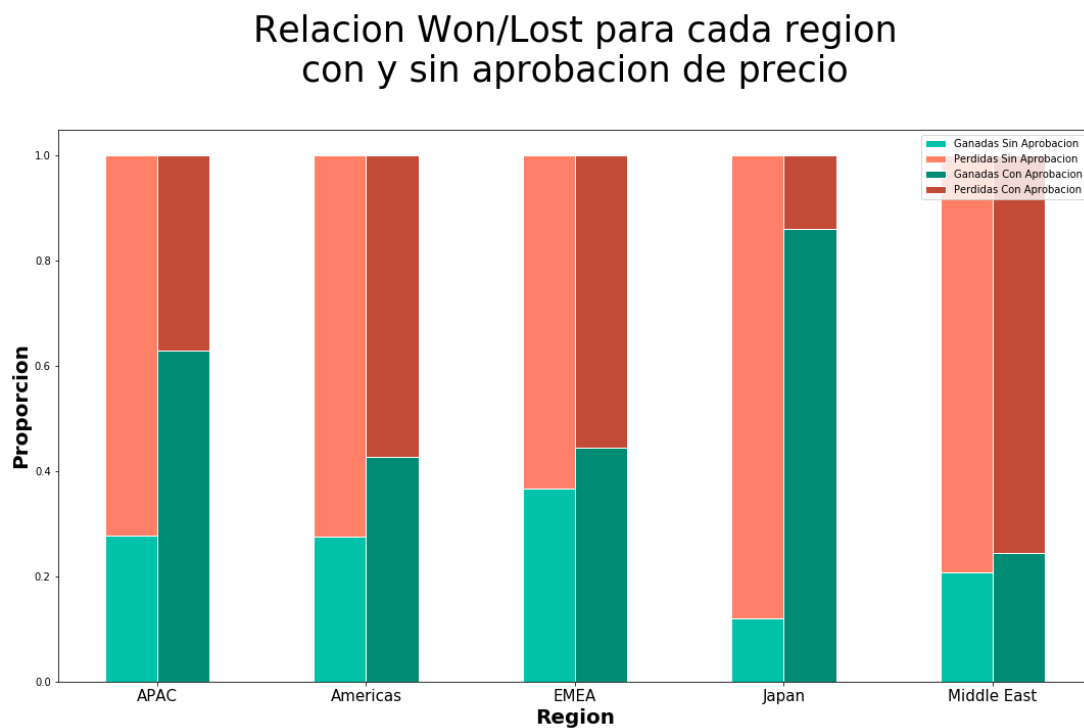


Figura 11: Éxito/fracaso según aprobación de precio y envío por región

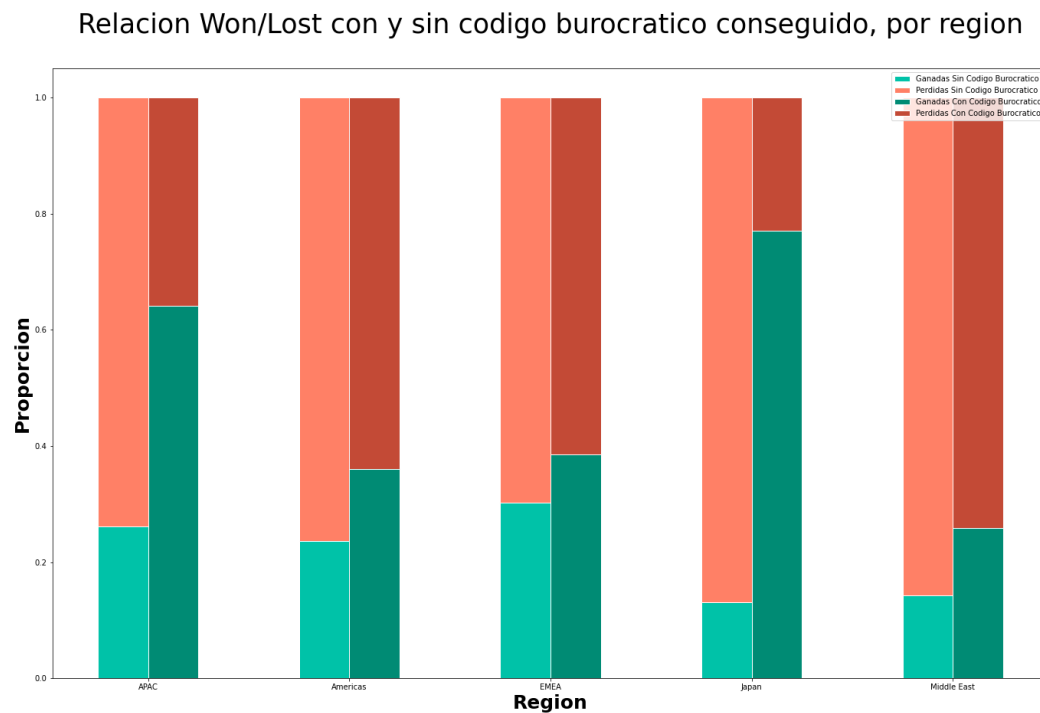


Figura 12: Éxito/fracaso según aprobación de código burocrático por región

Analizando los gráficos de las figuras 11 y 12 podemos ver que la proporción de oportunidades ganadas, con relación a las perdidas, aumenta cuando el permiso es otorgado. En especial en Japón donde se da un aumento considerable. Esta relación se da tanto para los permisos de importe y entrega, como los de código burocrático.

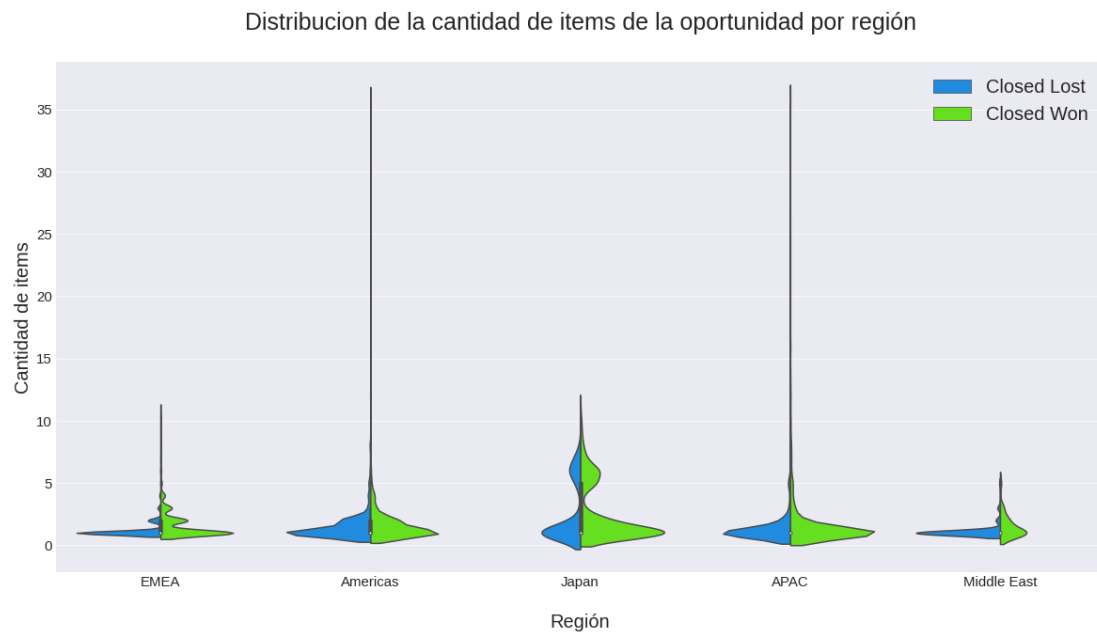


Figura 13: Distribución cantidad de items de la oportunidad por región

Dado que hay muy pocas oportunidades con mas de 10 items se realizó el mismo gráfico pero acotando las oportunidades a las que tienen menos de 8 items para poder apreciar mejor las distribuciones

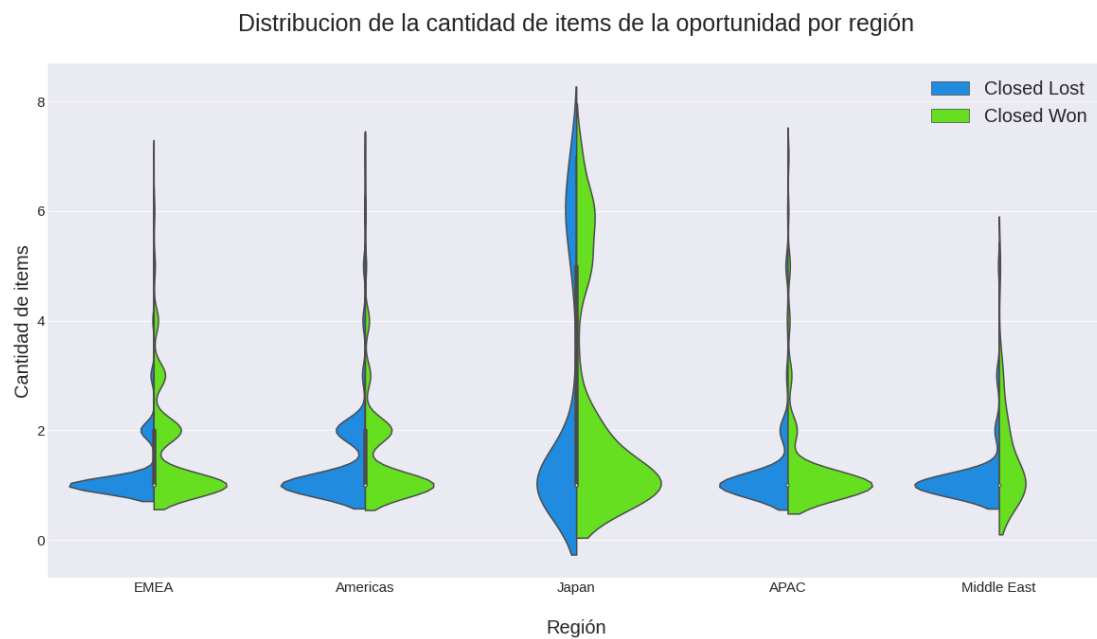


Figura 14: Distribución cantidad de items de la oportunidad por región

3.2. Territorios

Como vimos en la descripción de los datos, la columna territorio tiene casi 5000 registros sin información, por ende para analizar particularmente los territorios descartamos los mismos.

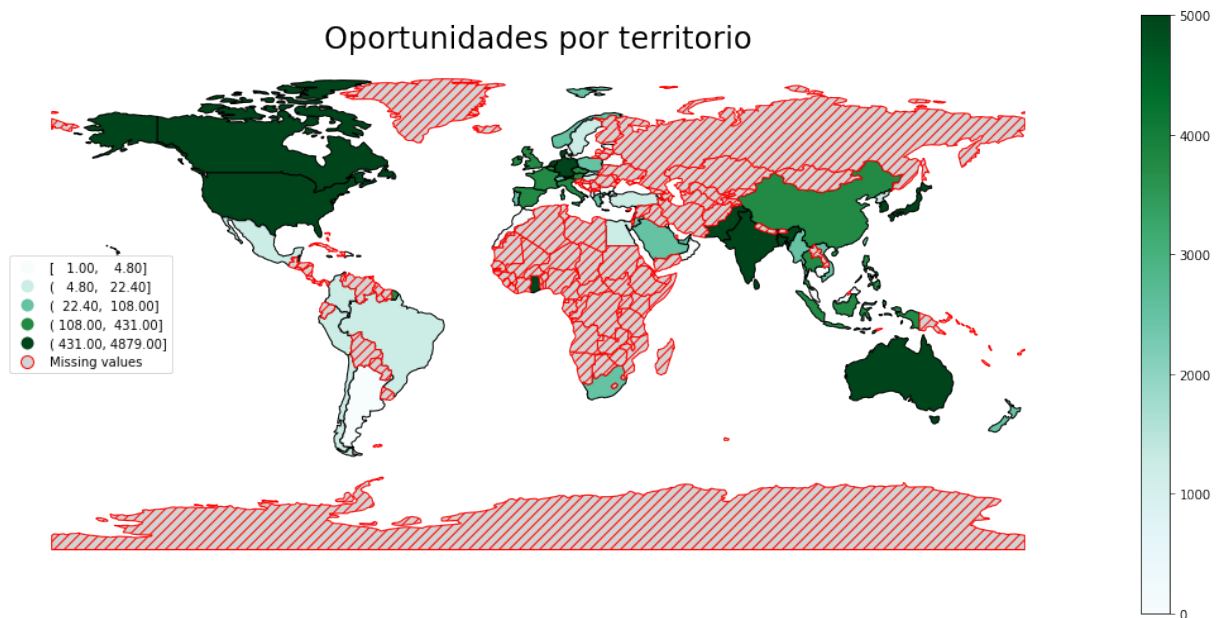


Figura 15: Región de mapa sin valores de territorios

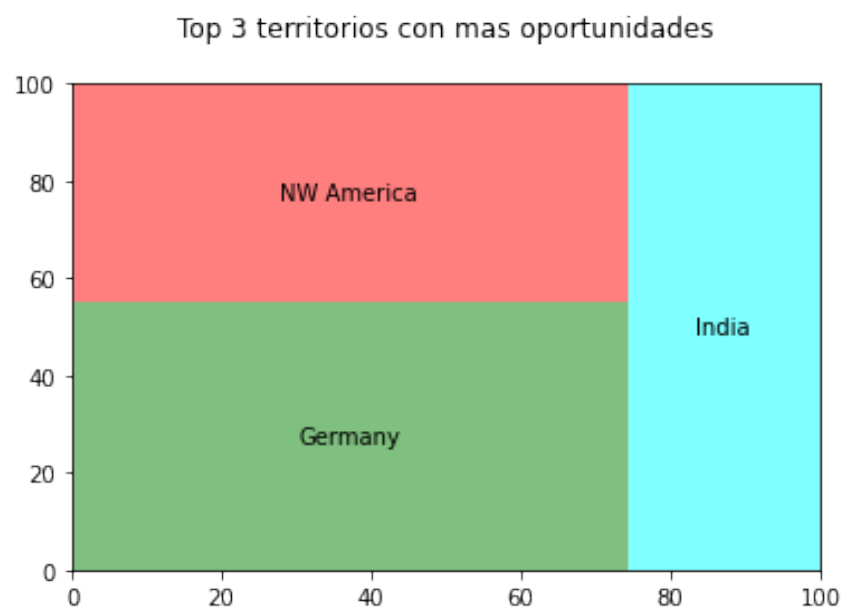


Figura 16: Top 3 territorios con mayor cantidad de oportunidades

Al observar la figura 15 podemos ver que del mapa hay muchos pais/territorios de los cuales no se tienen datos, tanto por no haber tenido oportunidades en los mismos como por faltantes de datos en sus oportunidades.

Por la figura 16, vemos que los territorios NW America y Germany son los que predominan en demasía, por lo tanto se realizará el mismo gráfico con mas territorios y excluyendo a estos dos.

Top territorios con mas oportunidades sin NW America y Germany

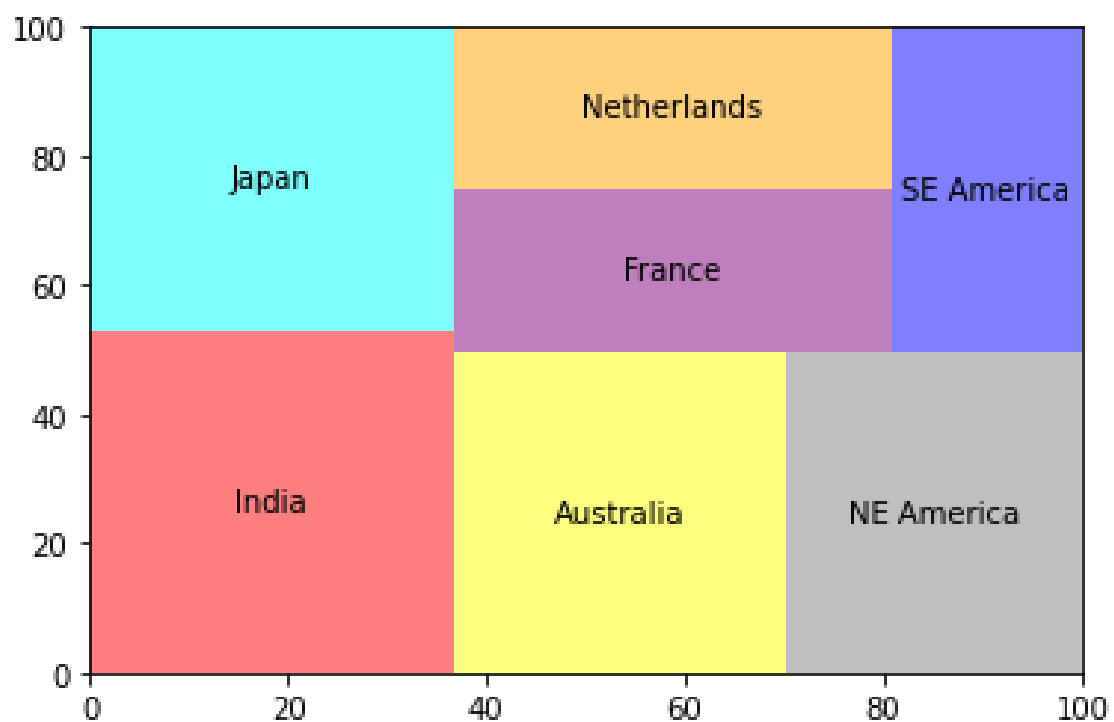


Figura 17: territorios con mas oportunidades

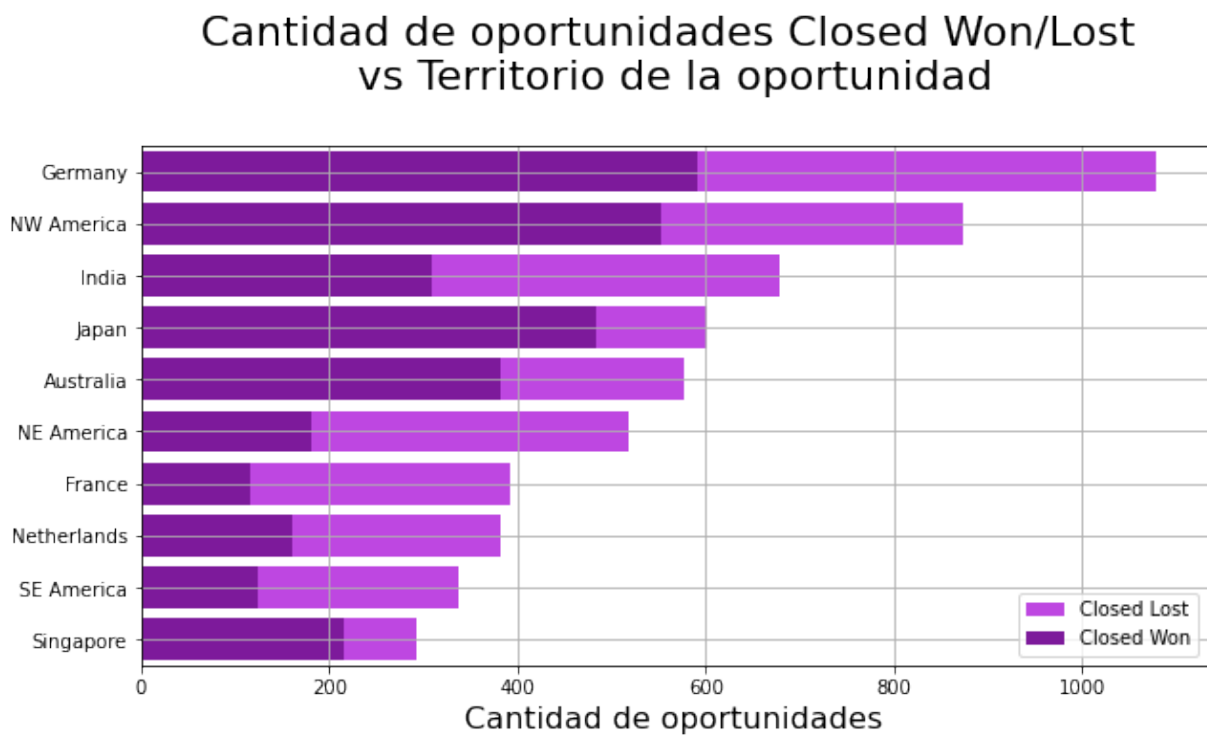


Figura 18: Closed Won por territorios

Como vemos por más que Germany y NW America sean los territorios que mas cantidad de oportunidades tienen, la relación de closed won y closed lost es mejor en regiones como Japan o Australia.

3.3. País

En esta sección se analizará la variable de país pero no de la oportunidad sino de donde se emitió la factura.



Figura 19: Mapa oportunidades

Como algunos países, en comparación con las demás, tiene muy pocas o ninguna oportunidad, no se puede apreciar la diferencia entre ellos, por lo tanto realizó la misma visualización pero con escala logarítmica.

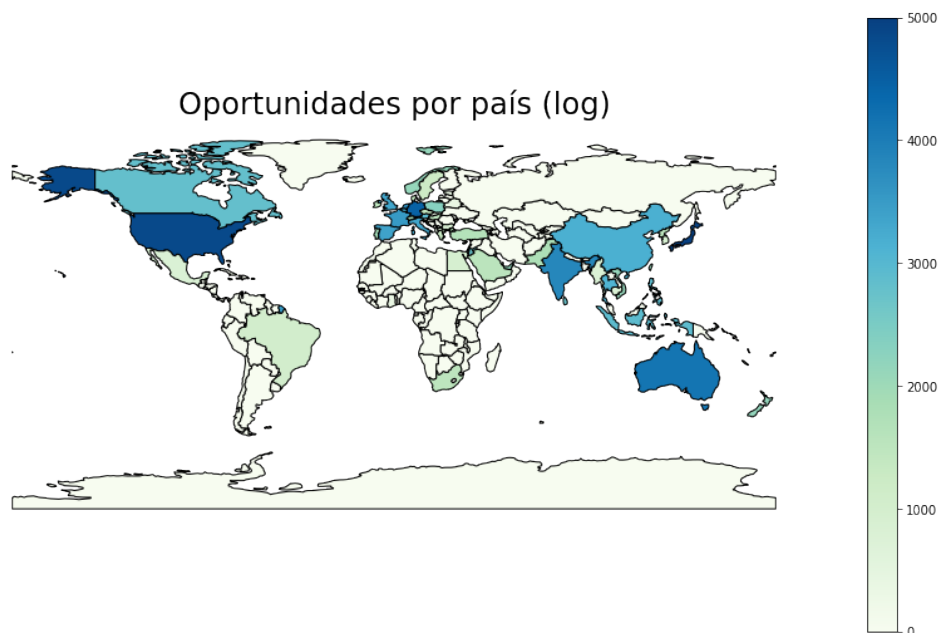


Figura 20: Mapa oportunidades escala log

Finalmente tomamos una muestra de los 10 países con más facturas emitidas y vemos que Japan y United States son los países donde más se emitió, aunque se ve una diferencia muy grande cuando se los clasifica en Closed Won/lost.

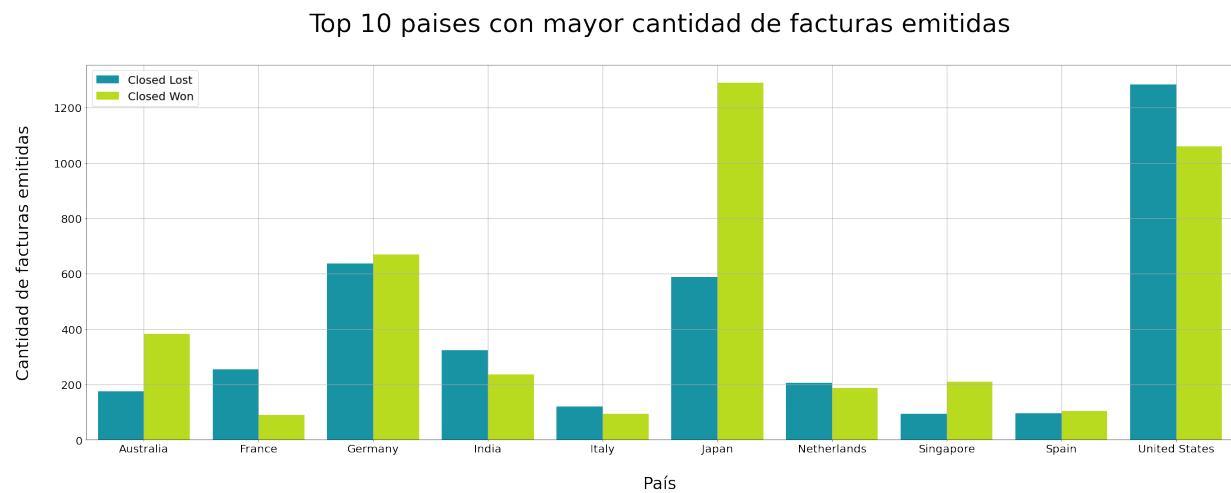


Figura 21: Top 10 países con mayor cantidad de facturas emitidas

4. Análisis temporal

En esta sección se analizará el aspecto temporal de los datos, como fechas de creación de oportunidades, fecha de delivery, fecha de creación de las cuentas, etc.

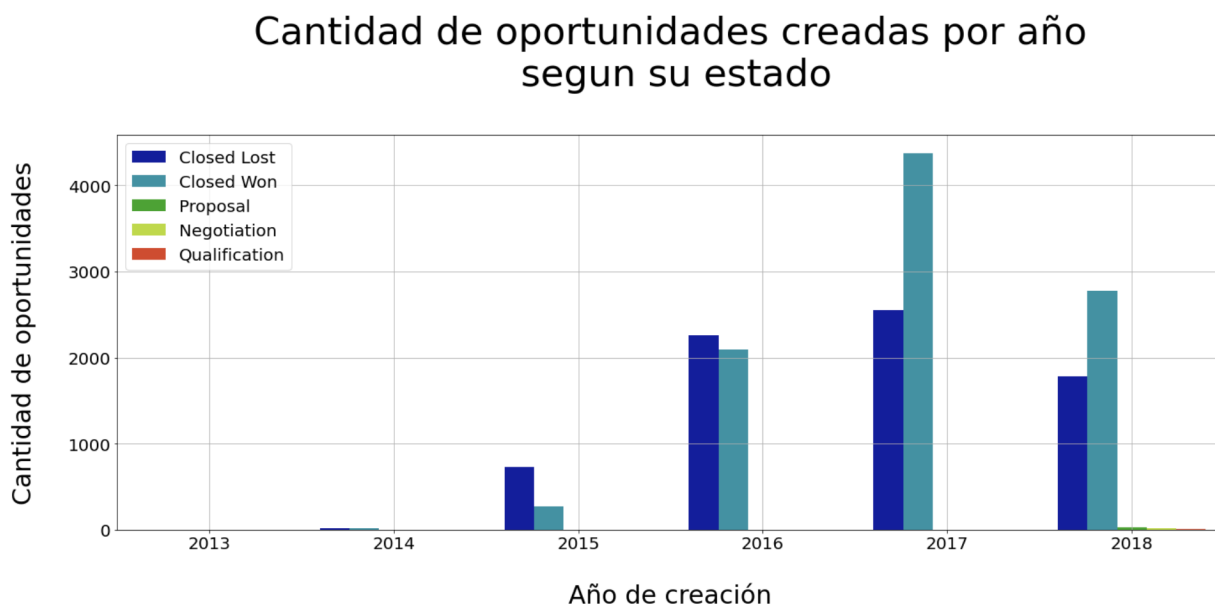


Figura 22: Cantidad de oportunidades creadas por año

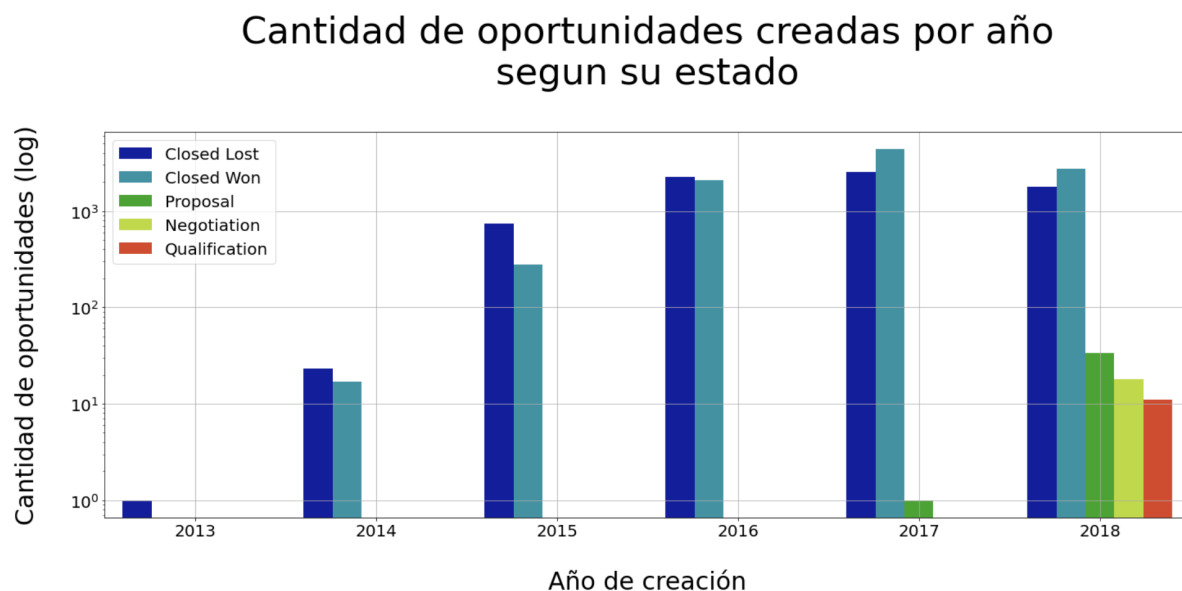


Figura 23: Cantidad de oportunidades creadas por año en escala logarítmica

Como podemos ver el año con mayor creación de oportunidades es el 2017 y tanto para 2013 y 2014 no podemos apreciar la cantidad en la figura 22, por esto se volvió a realizar la visualización pero con escala logarítmica. De esta manera podemos apreciar mejor la comparación entre 2013 y 2014 y los valores para las oportunidades que no están closed won o closed lost

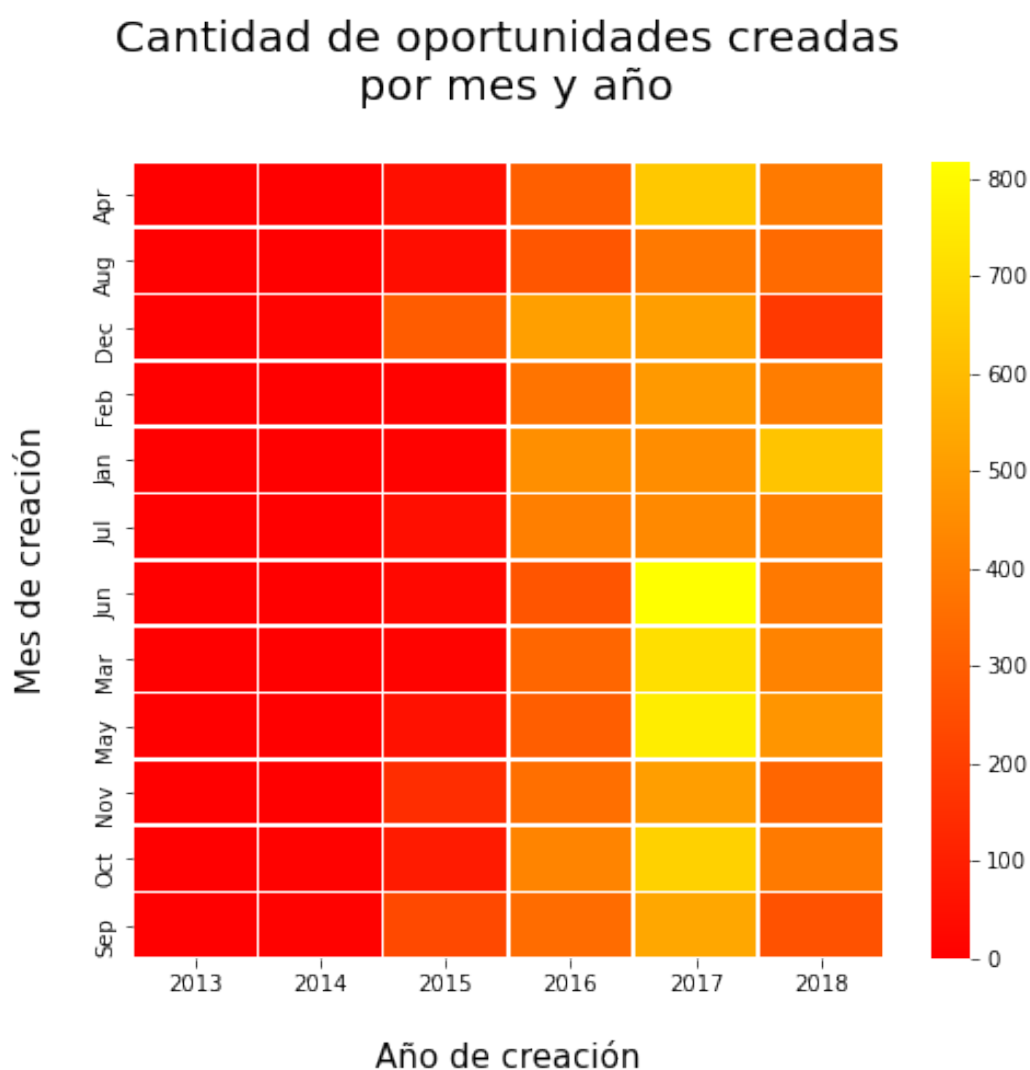


Figura 24: Cantidad de oportunidades creadas por mes y año

Afirmando lo que se mencionó anteriormente, el año 2017 fue el mejor y dentro de ese año el mes de Junio seguido del mes Mayo.

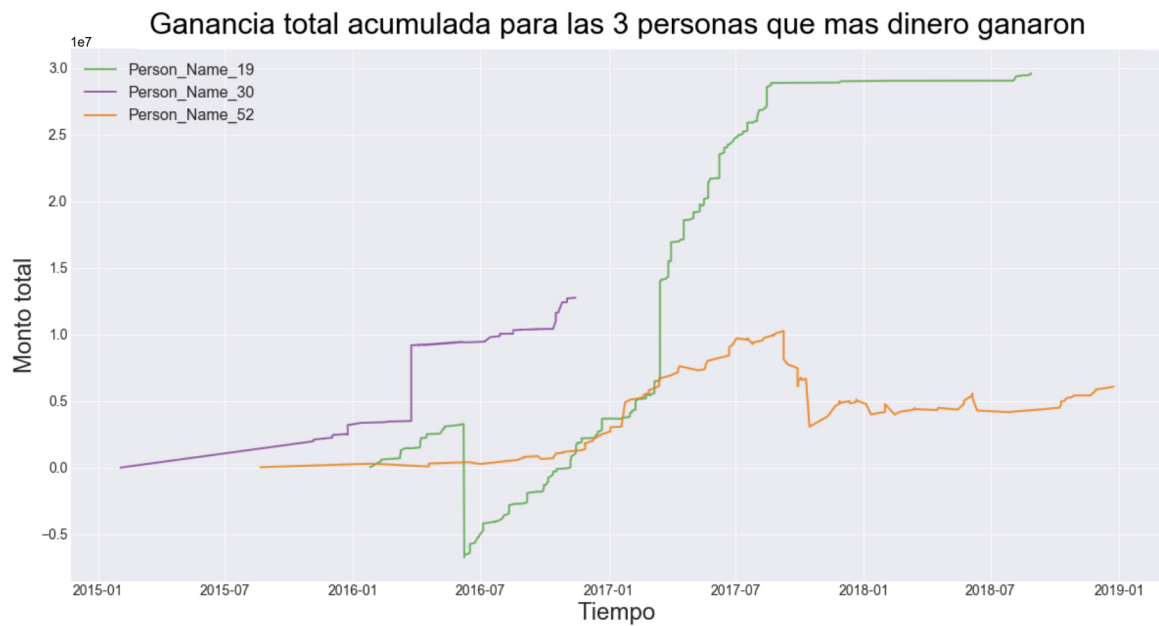


Figura 25: Ganancia acumulada para los 3 vendedores con mas ventas

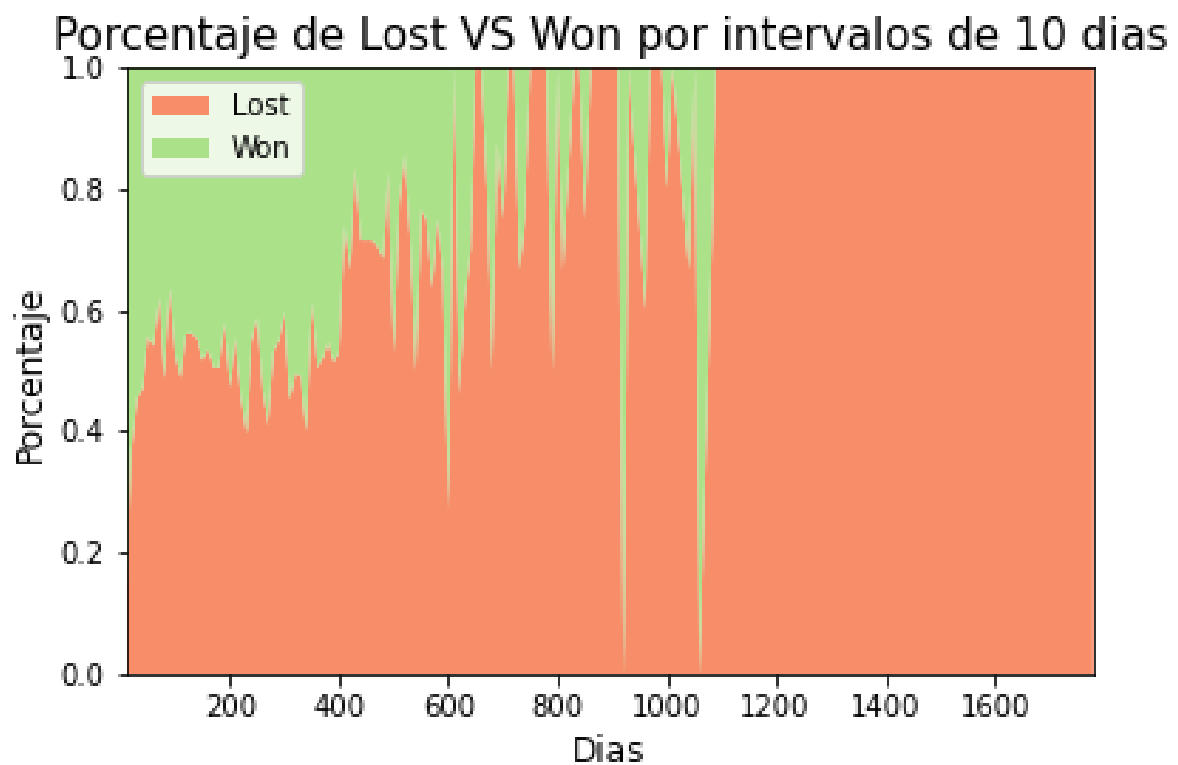


Figura 26: Porcentaje de Lost VS Won por intervalos de 10 días

En este gráfico podemos ver que a medida que aumenta la cantidad de días transcurridos de una oportunidad, el porcentaje de Lost va aumentando, por lo tanto cuanto mas pasa el tiempo menor probabilidad de concretar la oportunidad.

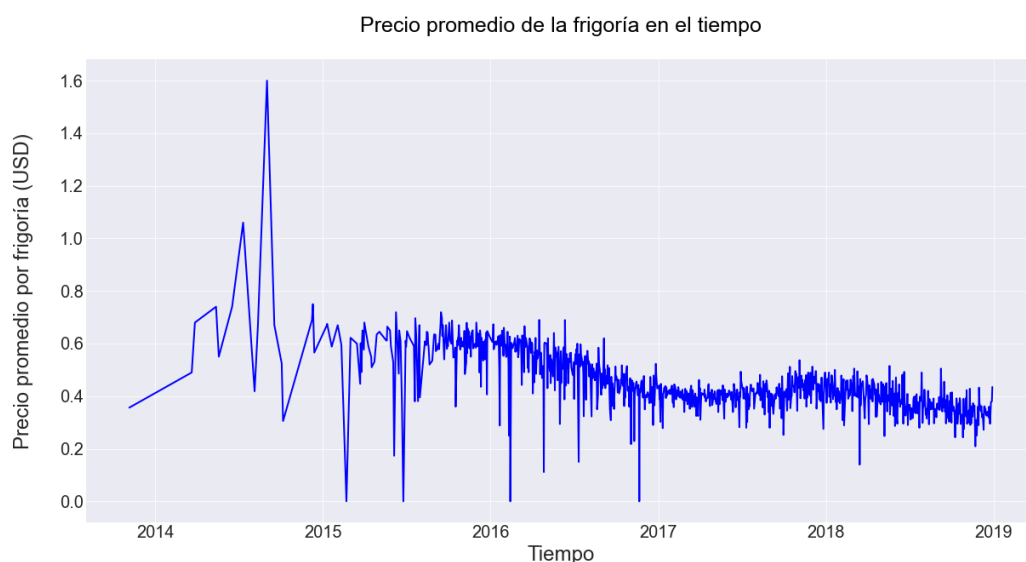


Figura 27: Precio de frigoría en el tiempo

Vemos que a lo largo del tiempo el precio promedio de frigoría en moneda dolar estadounidense fue disminuyendo levemente a razón de unos centavos al cabo de 4 años aproximadamente.

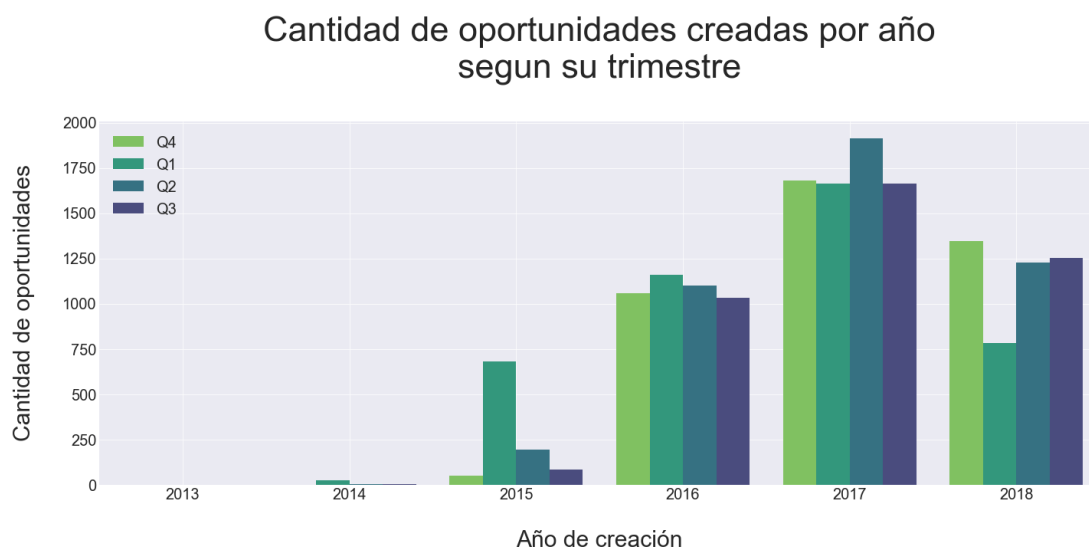


Figura 28: Cantidad de oportunidades al año por trimestre

Como se puede ver en el gráfico, con el pasar de los años aumentaron la cantidad de oportunidades, hasta el año 2017. Luego en el año 2018 se puede observar que hubo un pequeño descenso.

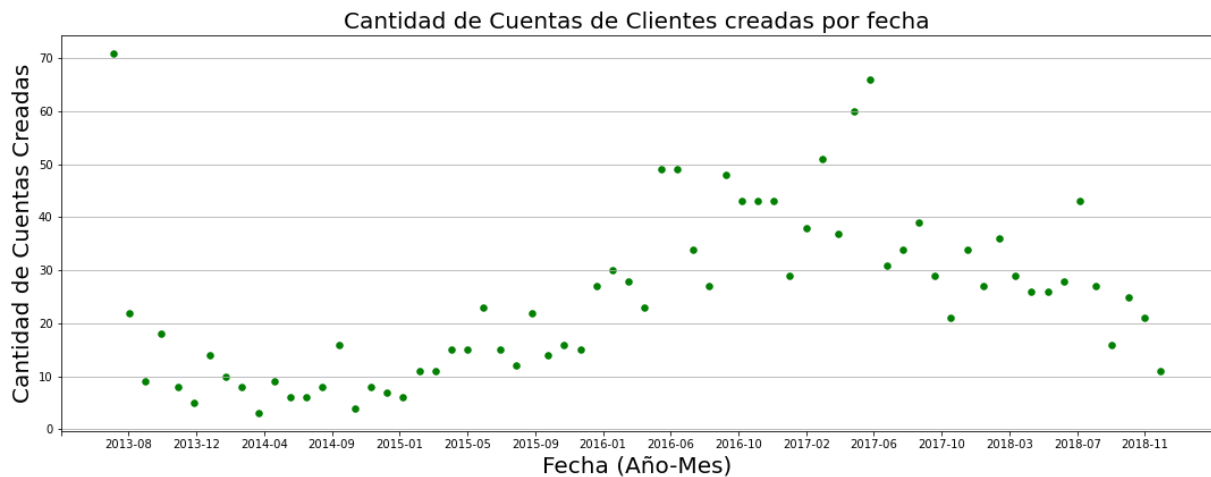


Figura 29: Cuentas creadas en el tiempo

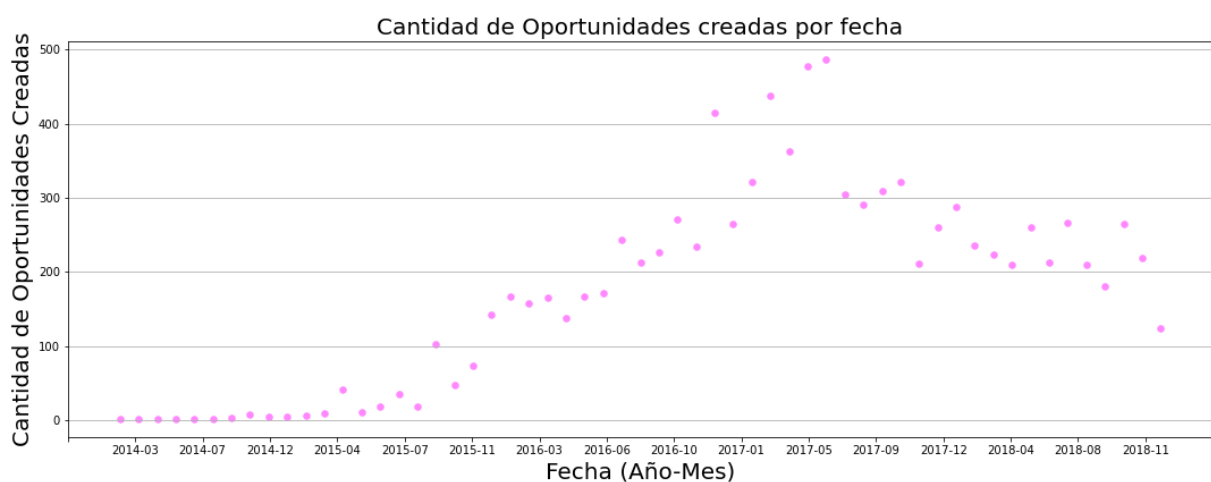


Figura 30: Oportunidades creadas en el tiempo

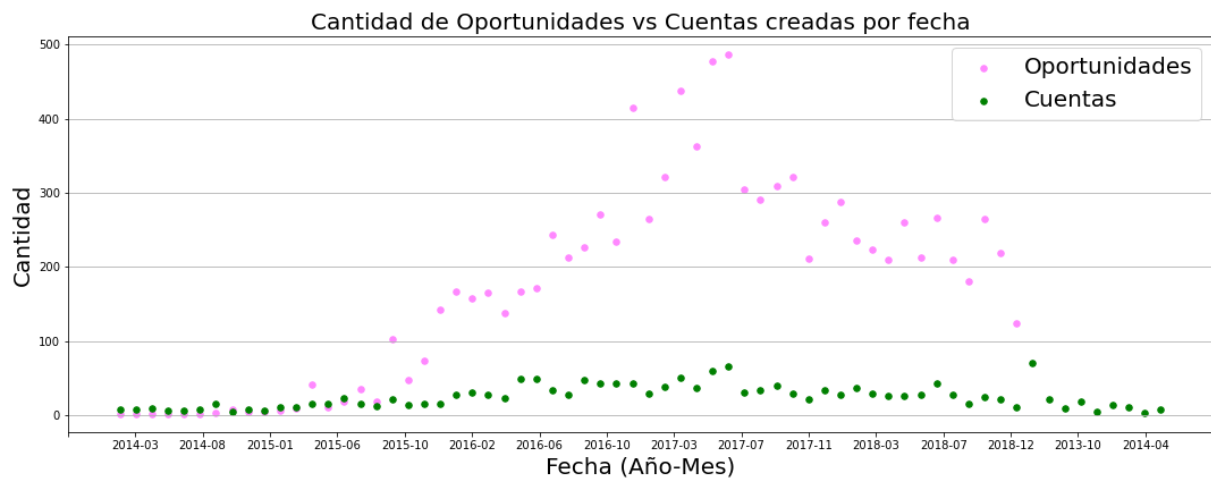


Figura 31: Oportunidades vs Cuentas creadas en el tiempo

Podemos observar que las oportunidades fueron creciendo desde el año 2014 llegando a un pico a mediados de 2017, correspondiéndose también con las fechas de mayor creación de cuentas cliente.

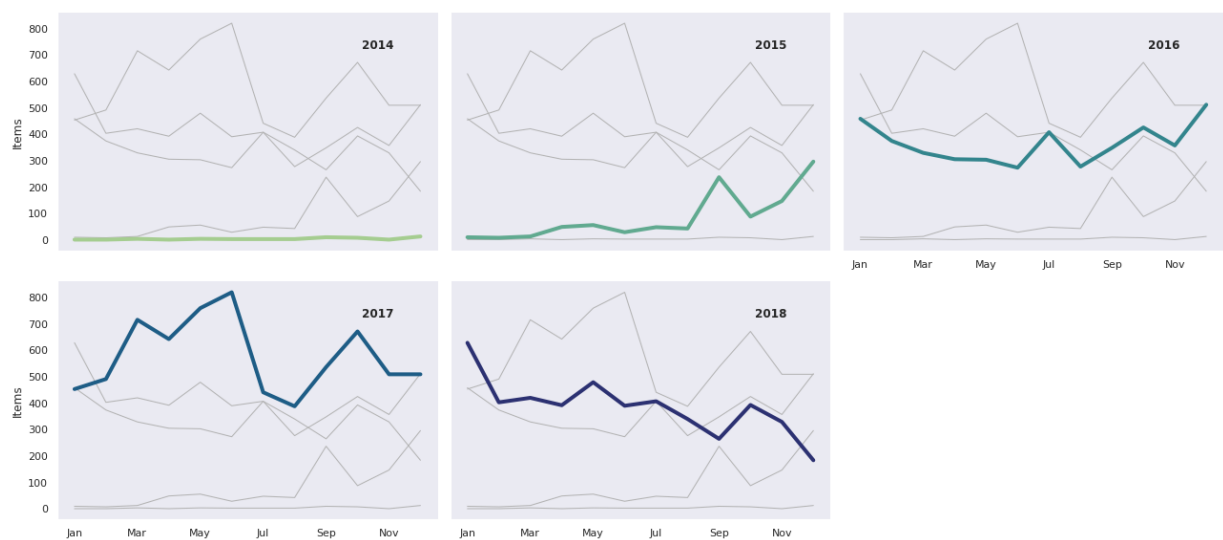


Figura 32: Cantidad de items por año

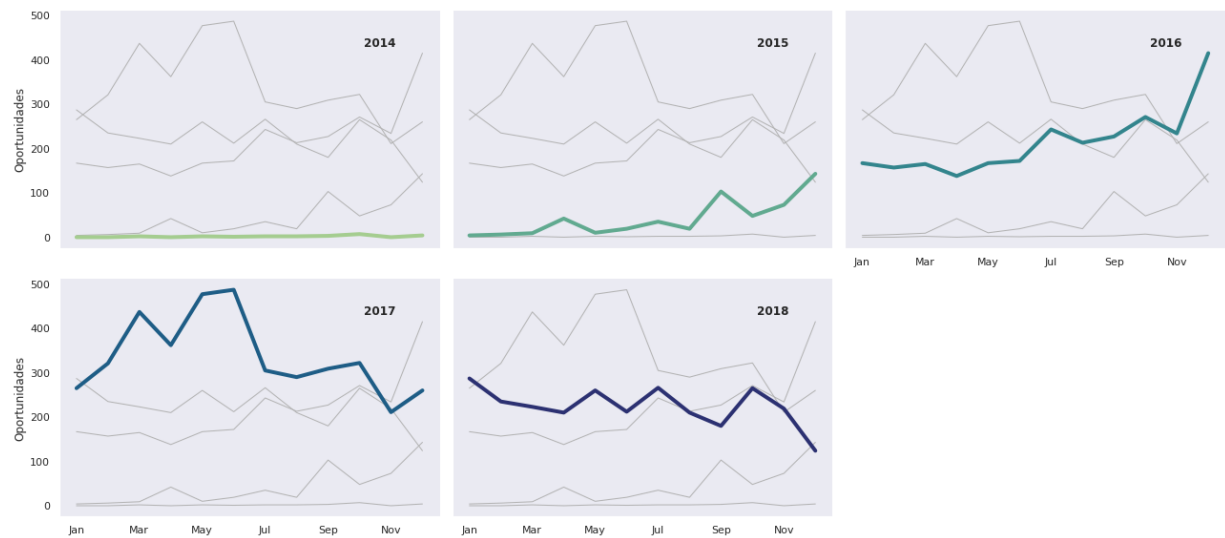


Figura 33: Cantidad de oportunidades por año

Podemos observar que la relación entre cantidad de oportunidades y cantidad de items por mes de cada año se mantiene proporcional, es decir a medida que aumenta la cantidad de oportunidades aumenta la cantidad de items en una relación mayor a 1:1

5. Oportunidades

Sabiendo que cada registro representa un producto/ítem dentro de una oportunidad, quisimos conocer cuantas oportunidades tenemos entre nuestros datos y de las mismas cuantas fueron finalizadas exitosamente y cuantas no.

5.1. Cantidad



Figura 34: Oportunidades según su estado

Como la cantidad de oportunidades con estados '*Closed Won*' y '*Closed Lost*' son tan grandes comparadas con las cantidades de los demás estados, realizamos la misma visualización pero con escala logarítmica para poder apreciar mejor los datos.

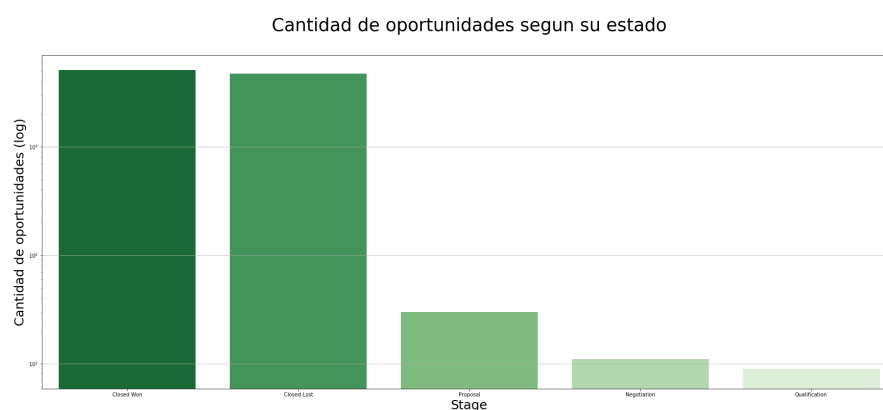


Figura 35: Oportunidades según su estado en escala logarítmica

Por último para saber que porcentaje representan dentro del total de 9841 oportunidades realizamos las siguientes visualizaciones.

Porcentaje de oportunidades según su tipo

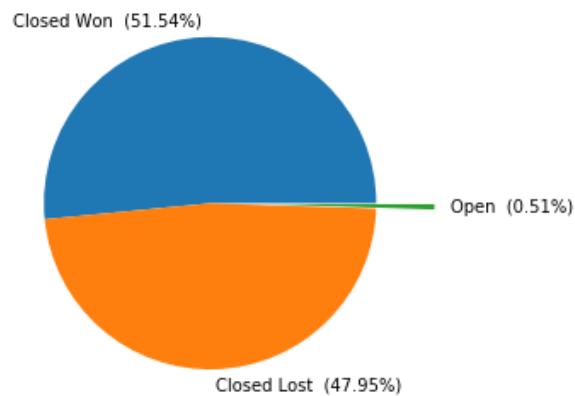
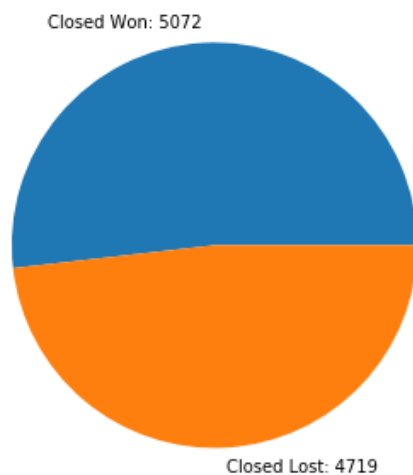


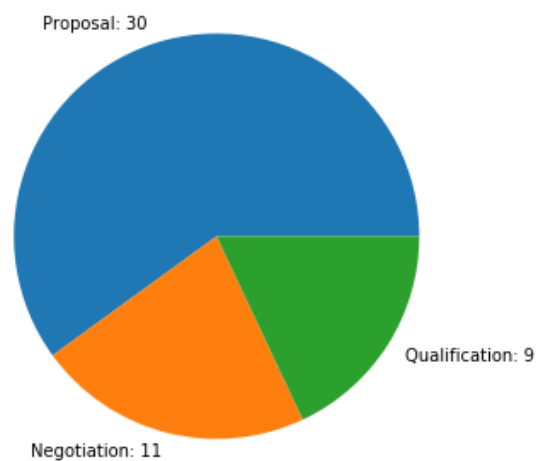
Figura 36: Porcentaje de oportunidades según su estado

Cantidad de oportunidades cerradas según su tipo



(a) closed

Cantidad de oportunidades abiertas según su tipo



(b) open

Figura 37: Cantidad de oportunidades según su estado

Como vemos más del 99% están en un estado final, es decir '*Closed Won*' o '*Closed Lost*'

5.2. Ítem

Ahora si tenemos 16947 registros, que cada uno de ellos representa un ítem de la oportunidad y tenemos solo 9841 oportunidades, podemos concluir que hay oportunidades que tienen mas de un ítem cada una.

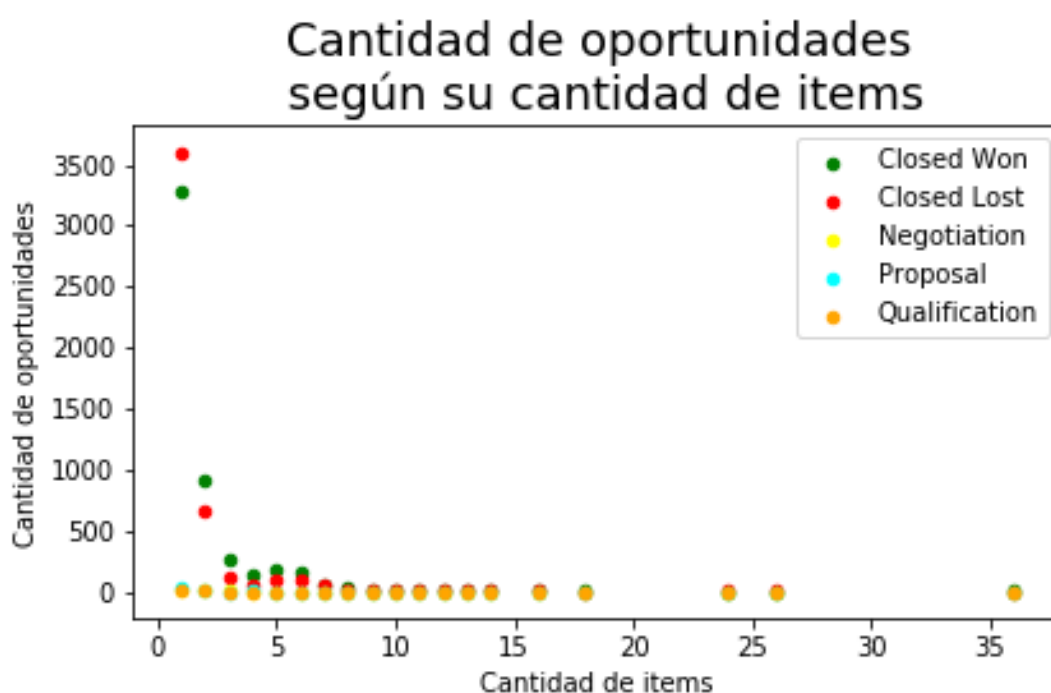


Figura 38: Cantidad de oportunidades según cantidad de ítem por oportunidad

A partir de las ultimas figuras podemos ver como se distribuyen los ítems según su estado, manteniendo la misma relación que las oportunidades prácticamente, es decir todos los ítems son en su mayoría de oportunidades que tienen estados '*Closed Won*' o '*Closed Lost*'.

Si miramos la figura 9 podemos observar que tenemos más de 3500 oportunidades '*Closed Lost*' y más de 3000 en estado '*Closed Won*' con un solo ítem. Ya a partir de 3 ítem por oportunidad no pasamos de las 500 oportunidades y es muy raro encontrar oportunidades con más de 10 ítem.

5.2.1. Montos de los items

Para el análisis de los montos de los items necesitamos convertir los valores a la misma moneda.

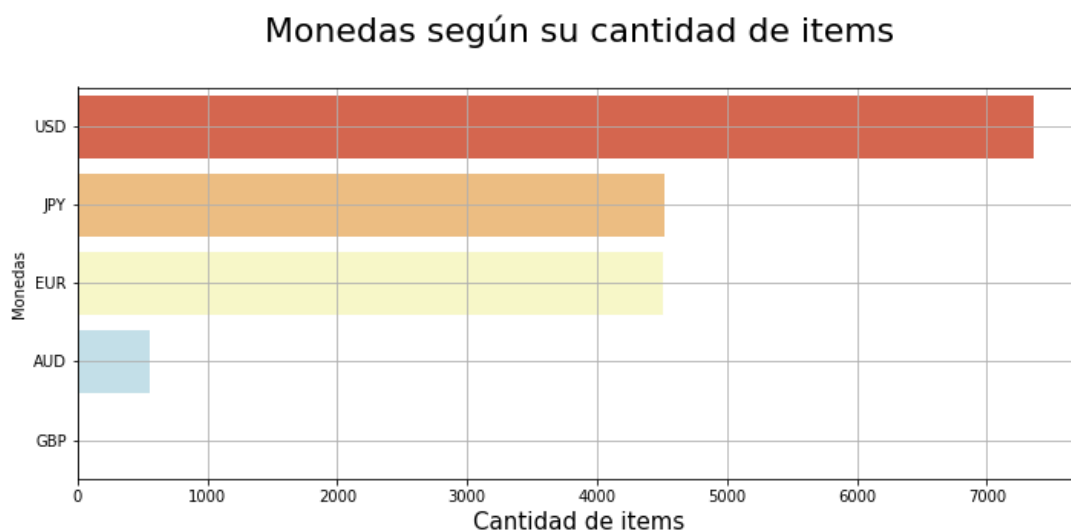


Figura 39: Monedas según sus cantidad de items

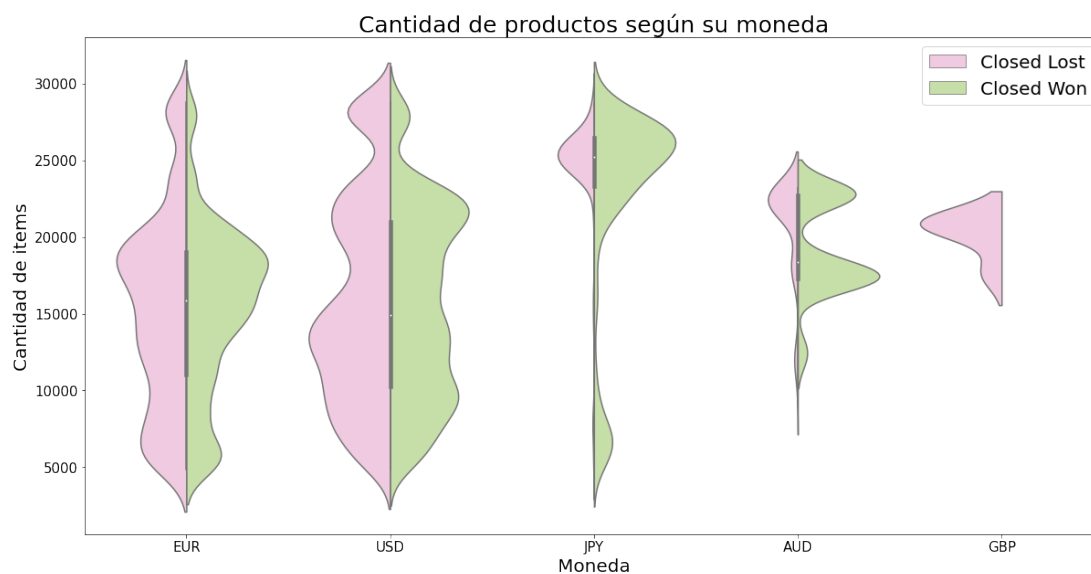


Figura 40: Monedas según sus items en estado exitoso o no

Por las figuras 39 y 40 podemos apreciar que tenemos montos con 5 monedas distintas y que la que predomina es dolar estadounidense (USD) por lo tanto realizaremos una conversión de los montos a esta moneda en los casos que sea posible.

Comparacion de Top 10 items con mayor monto según su estado

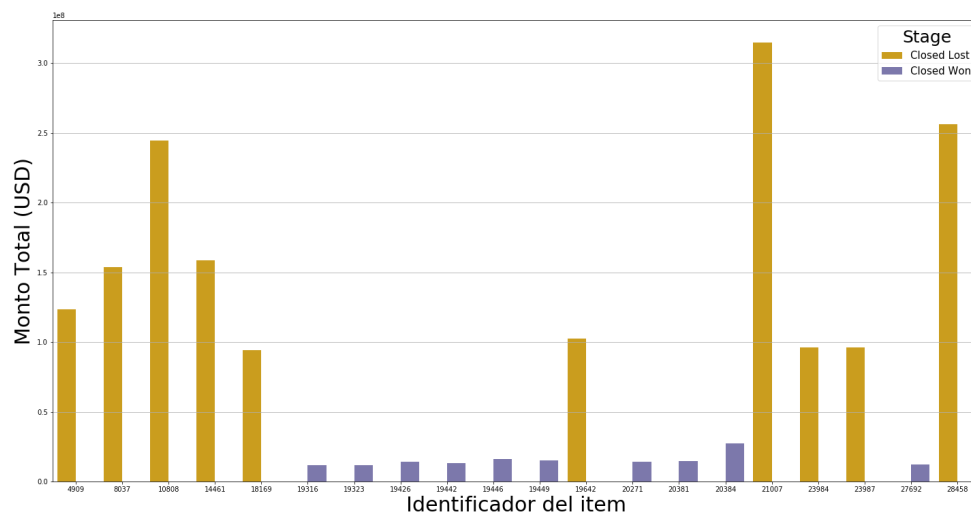


Figura 41: Items con mayor monto

Relacion entre monto del item y su estado

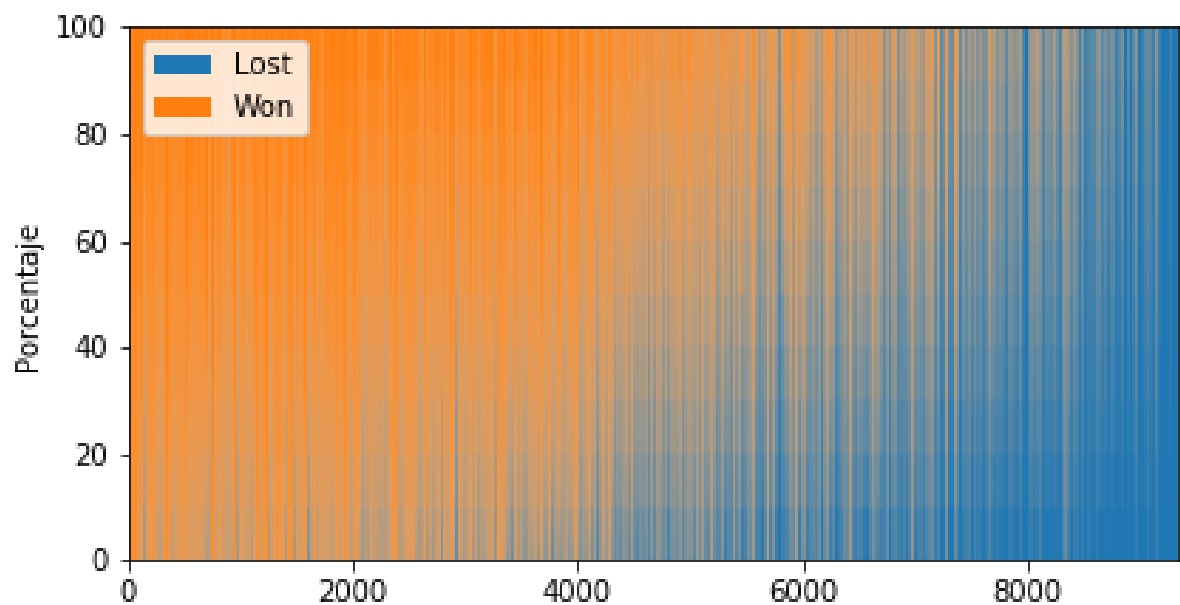


Figura 42: Relación de estados con monto del ítem

Apreciando las figuras 41 y 42 vemos que los items de oportunidades no exitosas tiene mayor monto, y además que a medida que el aumenta el monto tiende a crecer la probabilidad de no éxito.

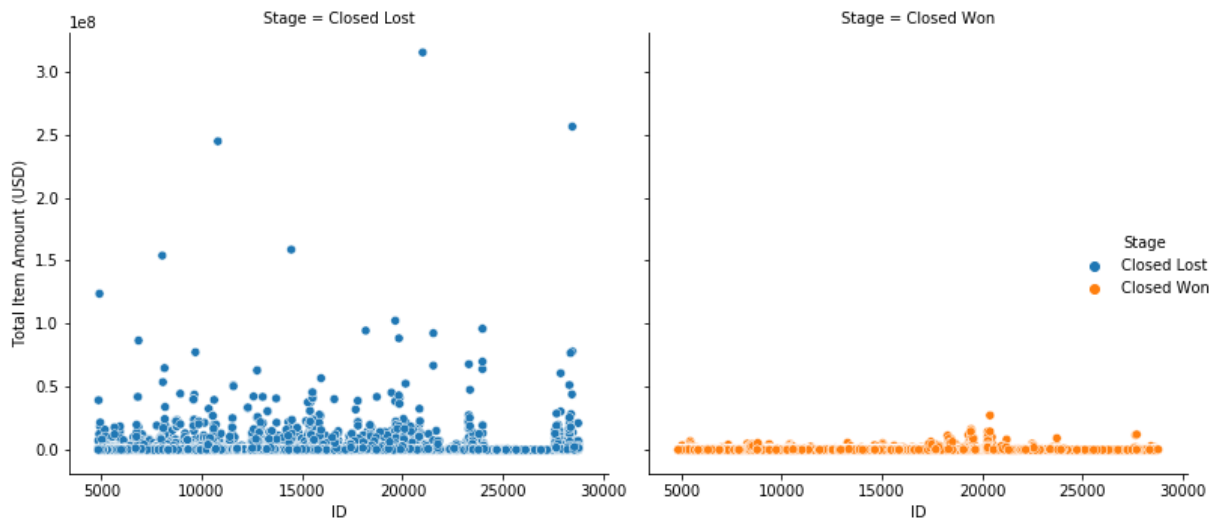


Figura 43: Distribución de items según sus montos

Como se puede observar los items de las oportunidades a medida que aumentan su valor tienden a aumentar su probabilidad de quedar '*Closed Lost*'

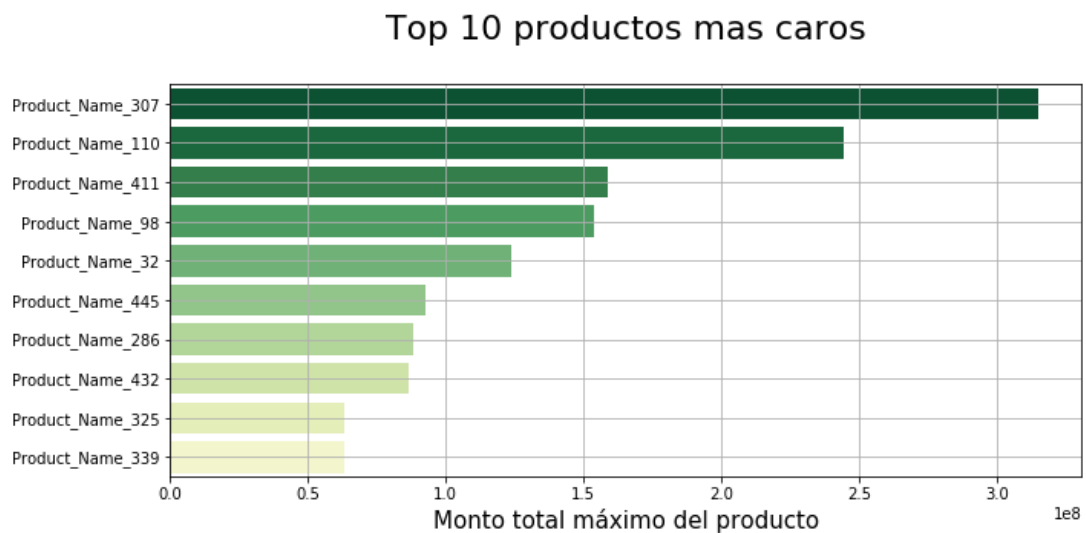


Figura 44: Productos con mayor monto

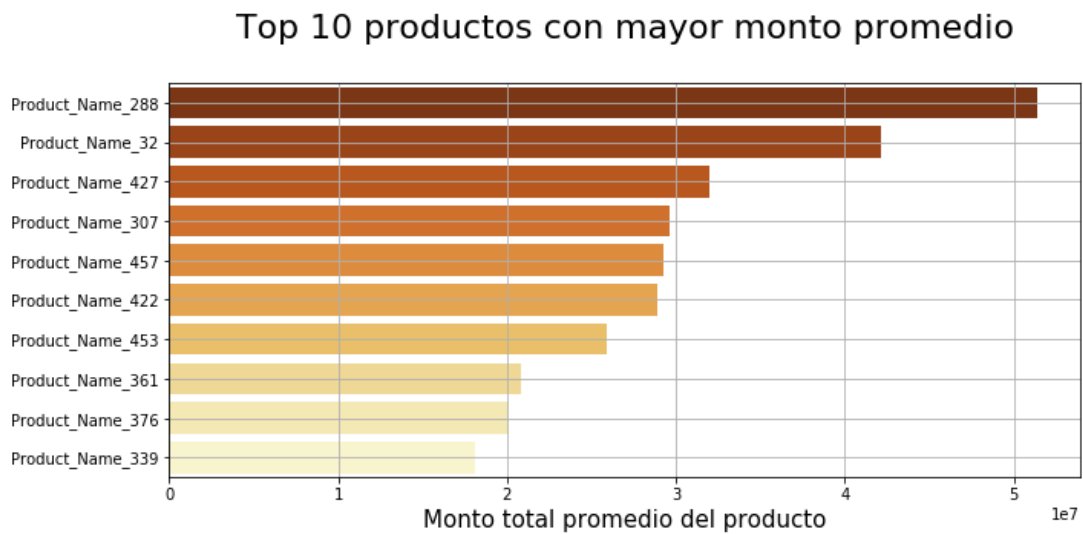


Figura 45: Productos con mayor monto promedio

5.2.2. Categorías de los items

Los productos mas populares



(a) items

Las familias de productos mas populares



(b) familias

Figura 46: Popularidad de items

Los productos mas populares no se encuentran en el top 10 de los productos mas caros (en valor total o promedio).

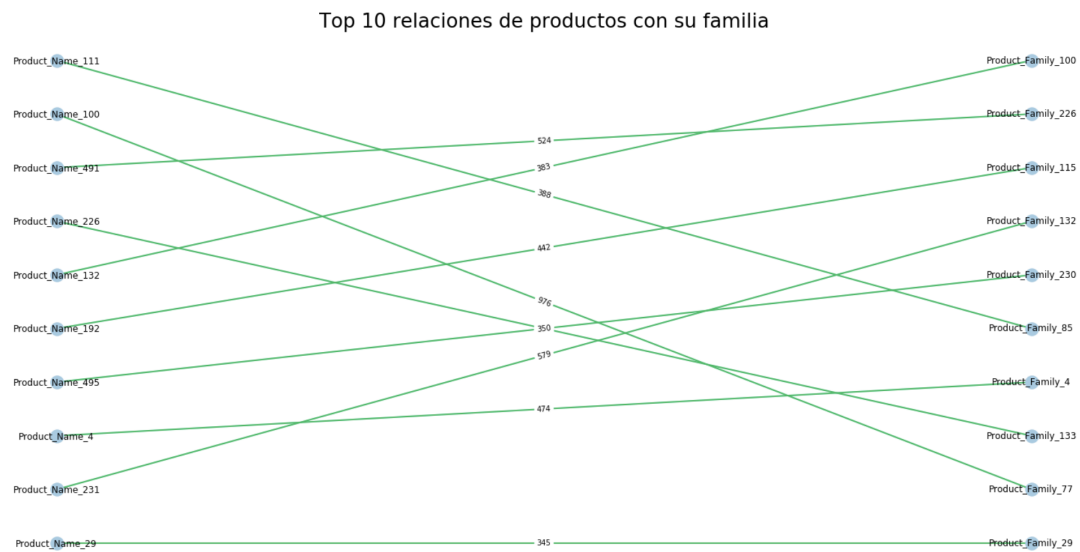


Figura 47: Top 10 relación familia y producto

Ademas podemos concluir que los productos mas populares coinciden con las familias mas populares, es decir, si un producto estuvo presente en muchas oportunidades, su familia se mantuvo igual para todas ellas haciendo que esa familia sea también muy popular.

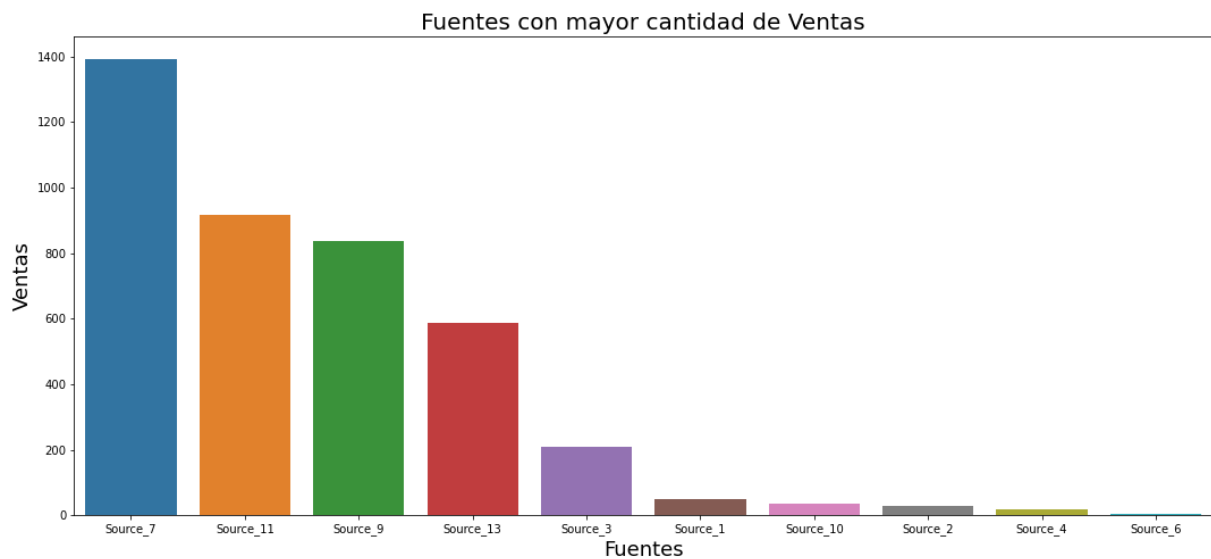


Figura 48: Fuentes con mayor cantidad de Ventas

5.3. Montos

En la siguiente visualización se puede observar el monto total de las oportunidades en USD en función de la cantidad de frigorías tanto para las oportunidades con stage '*Closed Won*', como '*Closed Lost*' separadas según si la oportunidad obtuvo o no aprobación especial de su precio total y los términos de la entrega.

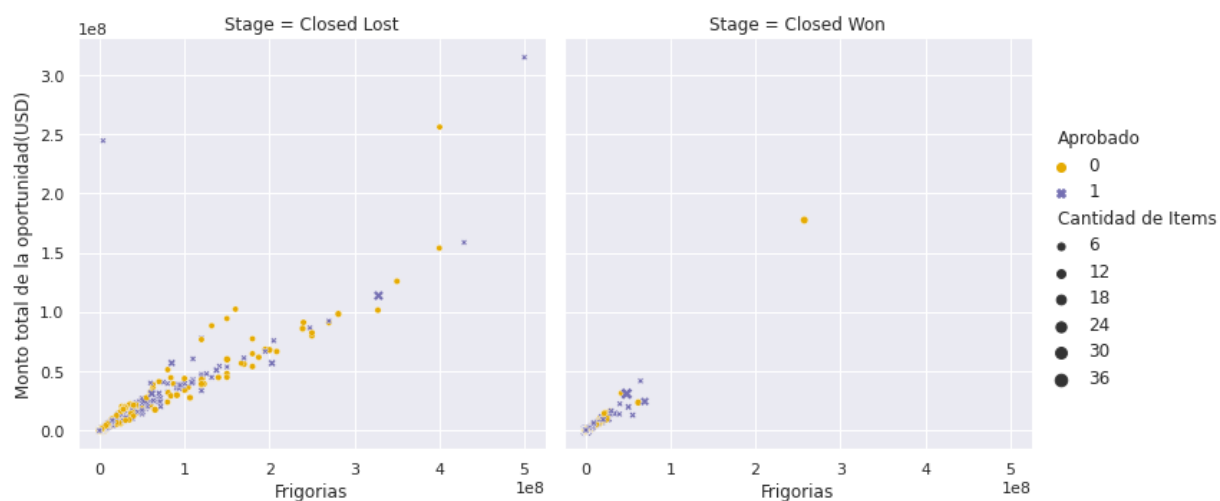


Figura 49: Monto total en función de cantidad de frigorías

Como la concentración de puntos se da para las oportunidades con pocas frigorías y menor monto (esquina inferior izquierda) y no se logra apreciar la distribución de buena manera, se realizó otro gráfico pero haciendo énfasis en dicha zona.

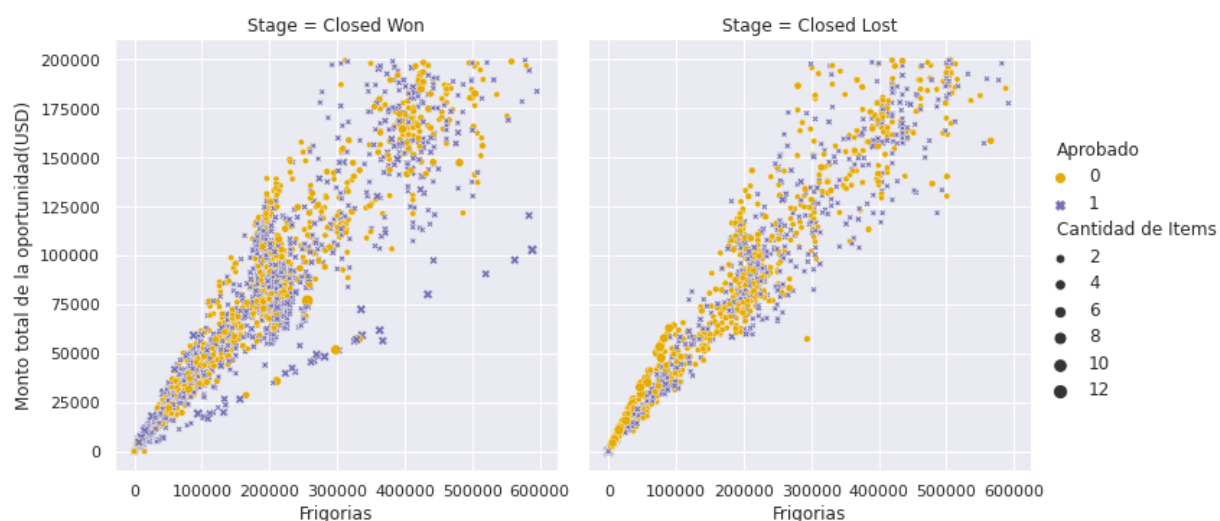


Figura 50: Zoom Monto total en función de cantidad de frigorías

Como se puede apreciar en la figura, el monto total y la cantidad de frigorías mantienen una relación lineal, a medida que crecen las frigorías aumenta el monto total casi proporcionalmente. Por otra parte, también se puede apreciar que para las oportunidades cuyo stage es '*Closed Won*', la mayoría de estas obtuvieron la aprobación de términos y precio.

5.4. Vendedores

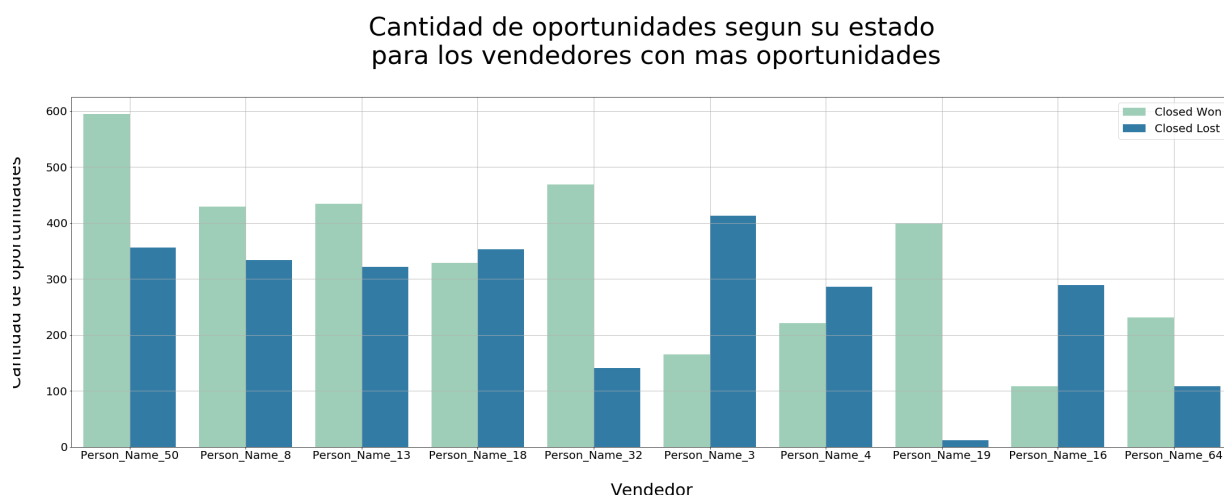


Figura 51: Cantidad de oportunidad según su estado para los 10 mayores vendedores

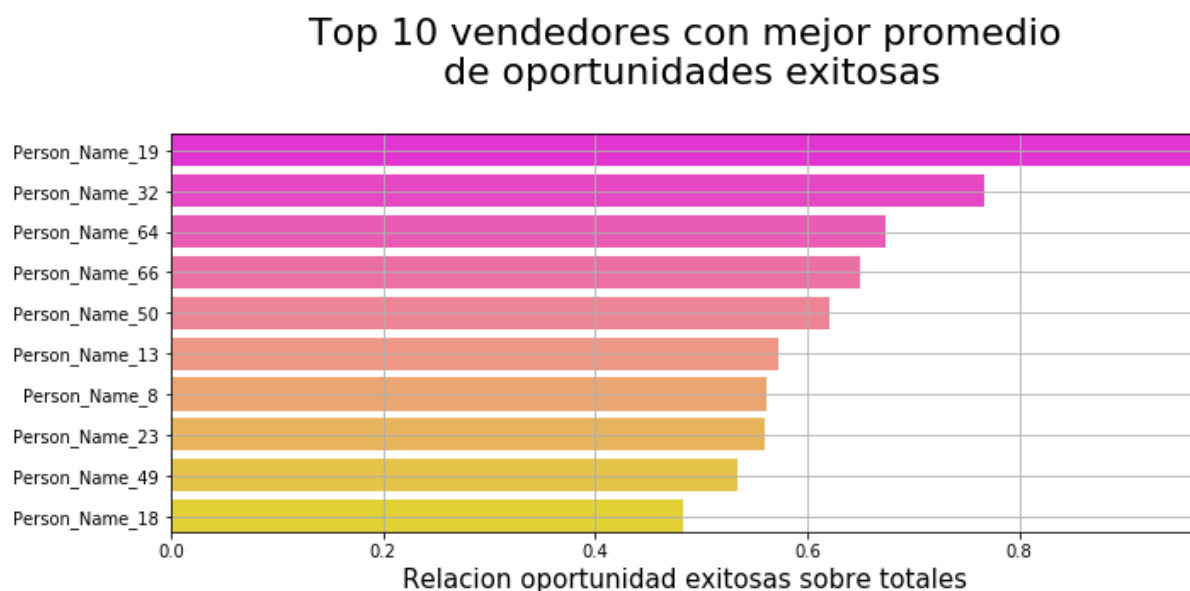


Figura 52: Top 10 de los vendedores comercial con mejor performance

De los vendedores con mejor porcentaje de éxito, la gran mayoría están dentro también del top 10 de los vendedores con mas oportunidades.



Figura 53: Popularidad de items

De la figura 53 vemos que es mejor que el responsable de la cuenta comercial sea distinto del responsable de la cuenta del cliente.

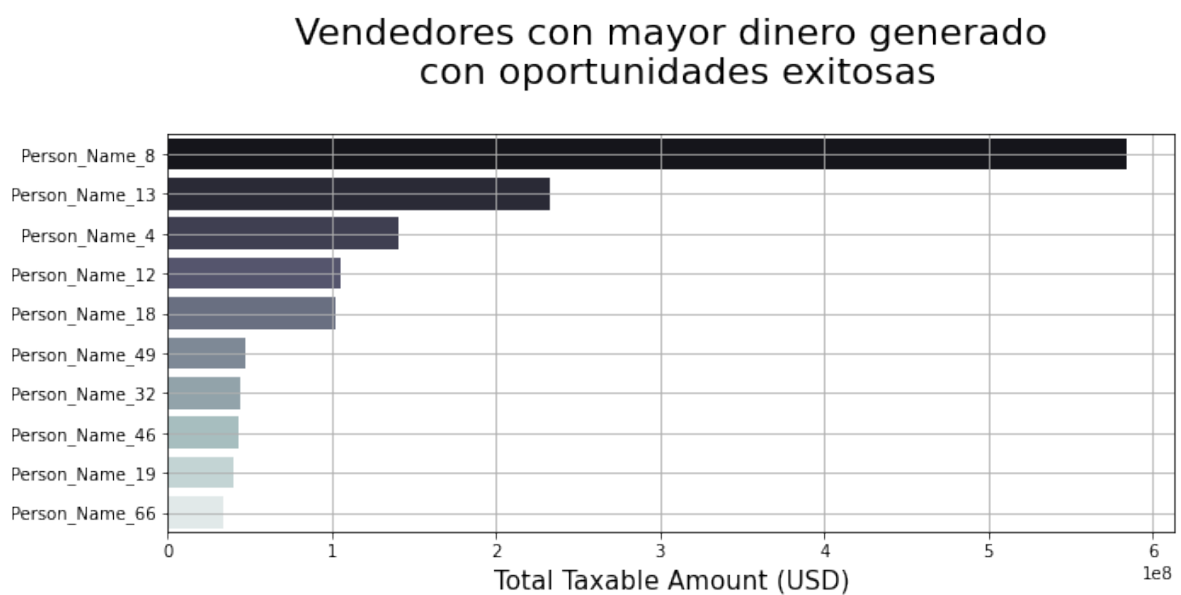


Figura 54: Vendedores responsable de cuenta comercial con mayor dinero generado

En cantidad de dinero generado podemos apreciar que '*Person_Name_8*' saca una gran ventaja respecto del resto.

5.5. Categorías

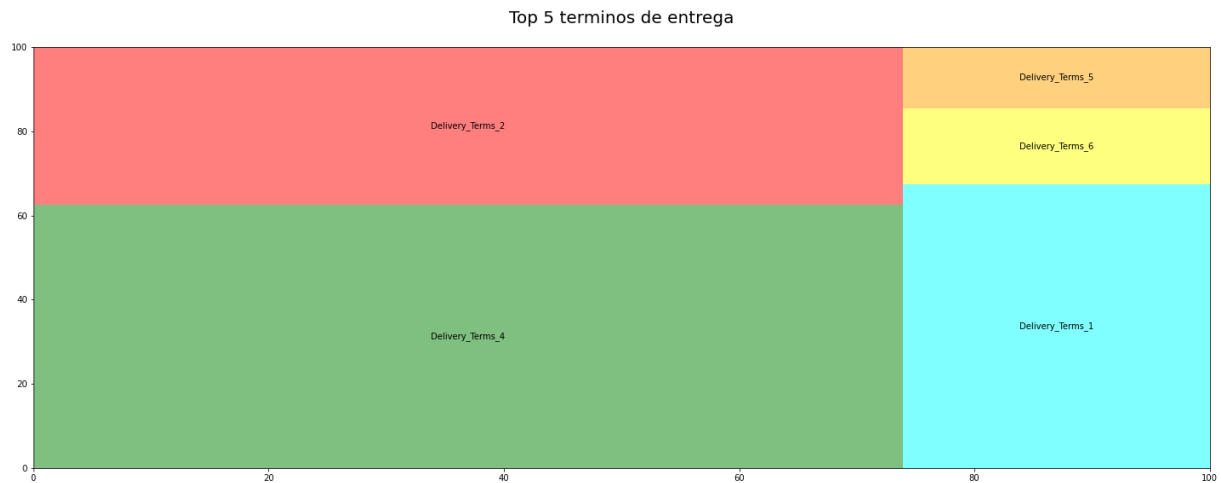


Figura 55: Top 5 términos de entrega

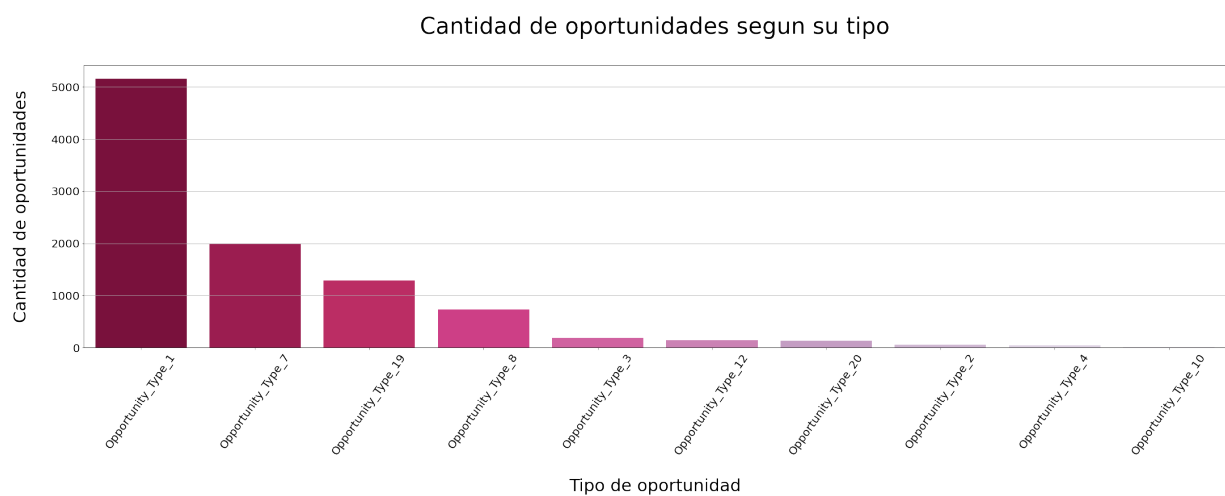


Figura 56: Top tipos de oportunidad

Como los tipos de oportunidad están concentrados casi en su totalidad en 4 posibles valores, decidimos cambiar la escala para ver mejor la comparación entre los tipos de oportunidad.

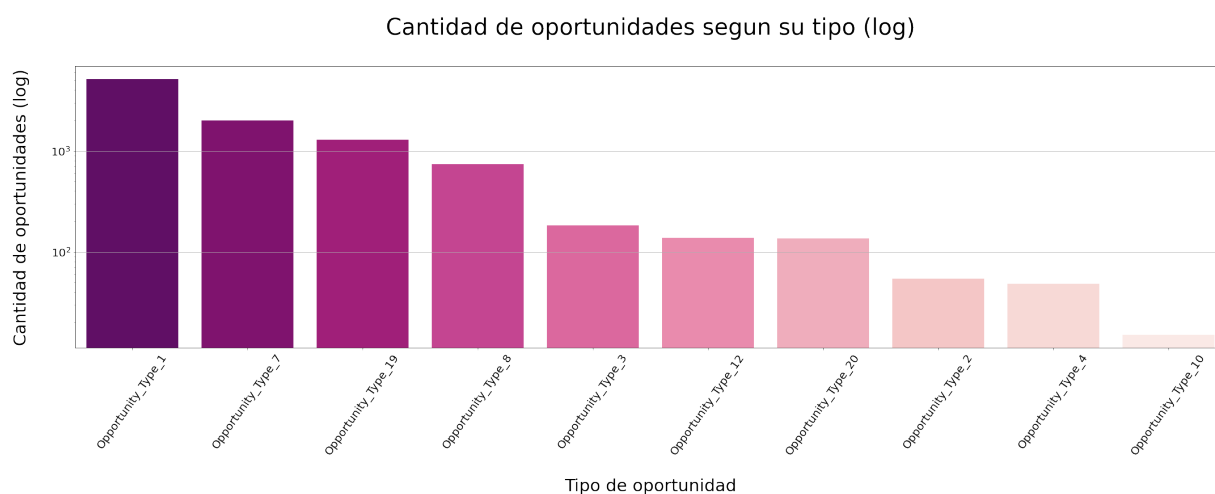


Figura 57: Top tipos de oportunidad (log)

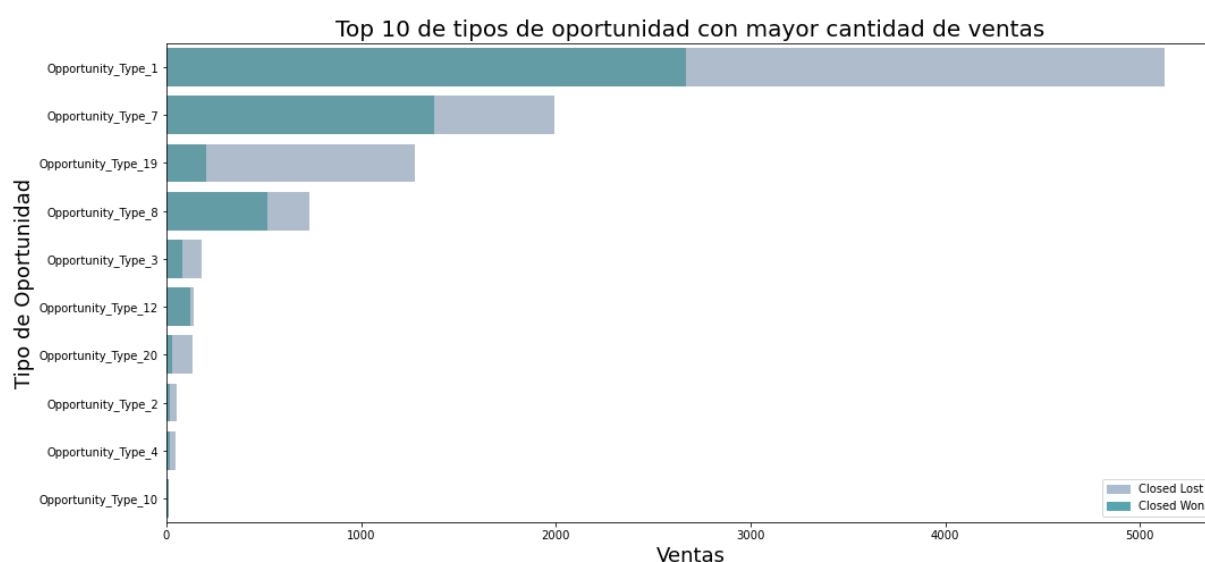


Figura 58: Top 10 tipos de oportunidad con mayor cantidad de ventas exitosas

Finalmente podemos ver que el tipo *Opportunity_Type_1*, a pesar de ser la que más cantidad tiene, su relación won/lost es casi del 50%, pero los tipos *Opportunity_Type_7* y *Opportunity_Type_8*, teniendo menos de la mitad de las oportunidades de la del tipo *Opportunity_Type_1*, su proporción de won/lost es mejor.

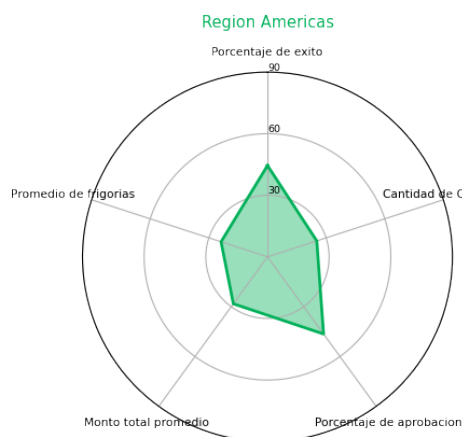
6. Insights

En esta sección analizaremos ciertas variables categóricas en varios aspectos:

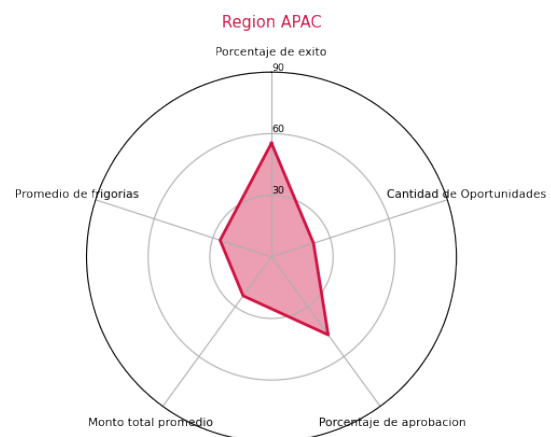
- Porcentaje de éxito: Representa la cantidad de oportunidades que finalizaron exitosas sobre el total de las mismas.
- Porcentaje de aprobación: Representa la cantidad de oportunidades que obtuvieron aprobación de los términos sobre el total de las oportunidades.
- Cantidad de oportunidades: Es la cantidad total de oportunidades que tienen asociada la variable categórica a analizar.
- Promedio de frigorías: Representa el promedio de frigorías para este tipo de oportunidades (normalizado).
- Monto total Promedio: Representa el valor del monto promedio en moneda dolar de las oportunidades (normalizado).

6.1. Insights por región

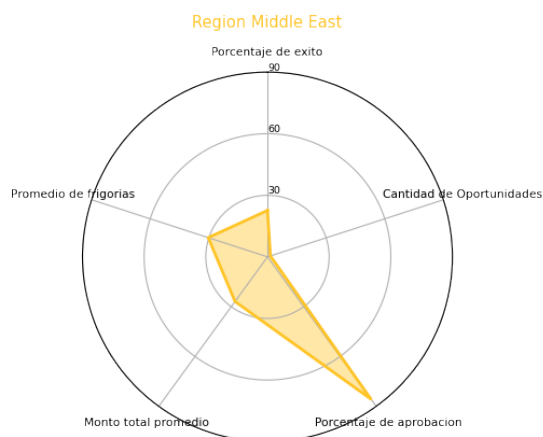
Elegimos la región como una variable a analizar ya que tenemos todas las oportunidades asociadas a una región, sin datos faltantes y solo manejamos 5 valores posibles, dándonos la posibilidad de poder compararlos con los insights descriptos anteriormente.



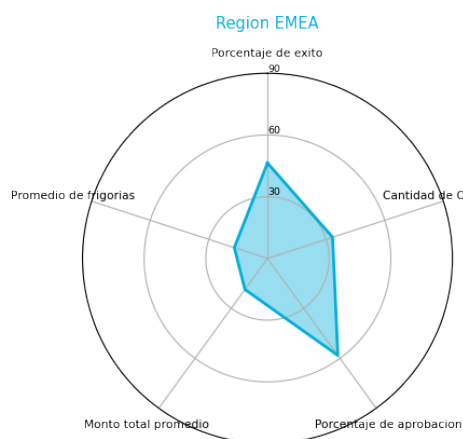
(a) Insights Americas



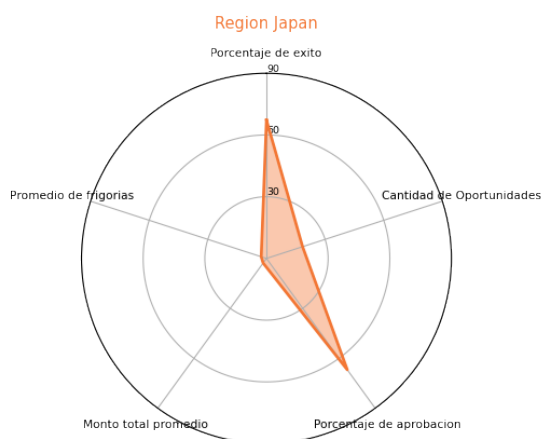
(b) Insights APAC



(a) Insights Middle East



(a) Insights EMEA



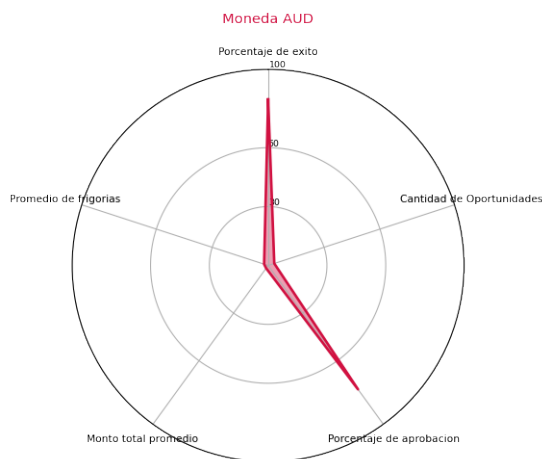
(b) Insights Japan

Viendo la comparación de las 5 figuras podemos ver que las regiones APAC y Americas son similares en todos los aspectos. La región de Middle West es la que mejor porcentaje de aprobación de los términos tiene pero viendo su cantidad de oportunidades no sería tan representativo. Finalmente comparando todos los insights respecto al porcentaje de éxito saca una ventaja la región de JAPAN, aunque se puede apreciar que su promedio de frigorías y montos es más bajo que el del resto de las regiones.

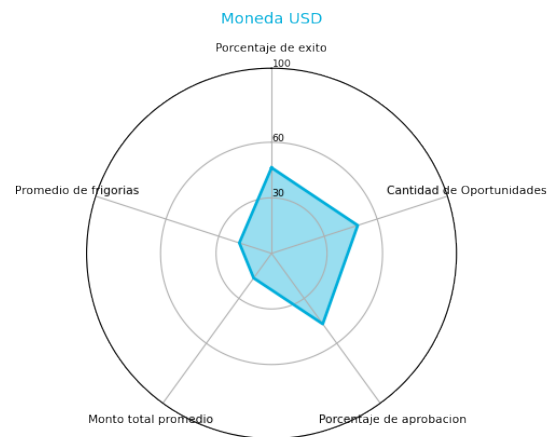
6.2. Insights por moneda

Decidimos analizar la moneda del monto de la frigoría promedio ya que tenemos todas las oportunidades asociadas a una moneda, y al igual que las regiones solo tienen 5 valores posibles:

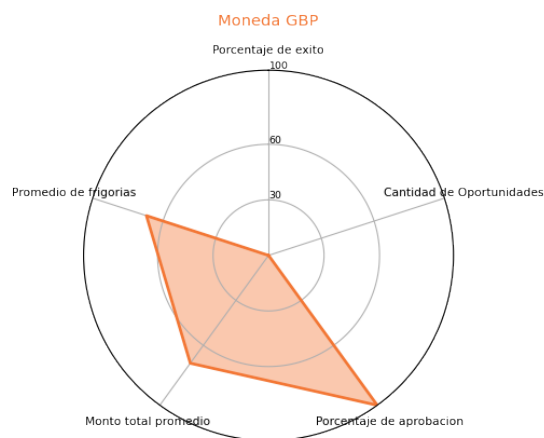
- USD: Dolar Estadounidense
- AUD: Dolar Australiano
- GBP: Libras Esterlina
- EUR: Euro
- JPY: Yen Japonés



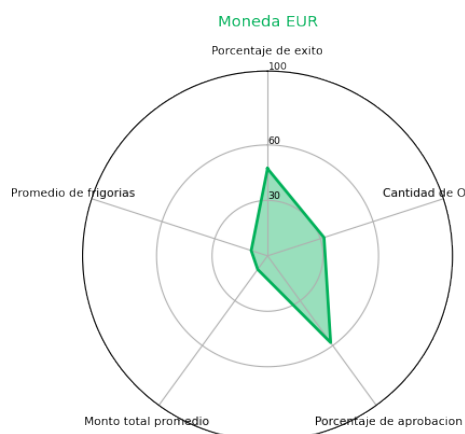
(a) Insights AUD



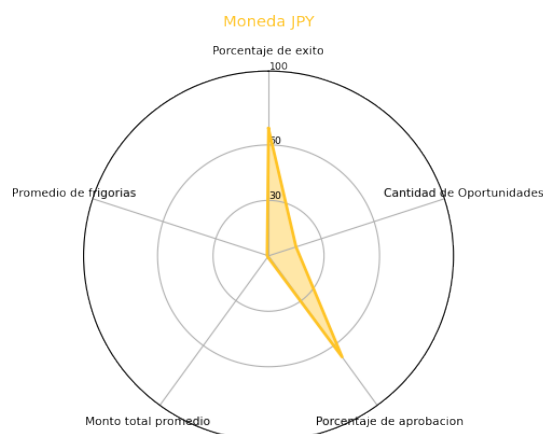
(b) Insights USD



(a) Insights GBP



(a) Insights EUR



(b) Insights JPY

Viendo la comparación de las 5 figuras podemos ver que el dolar Australiano y la libra esterlina no tienen una cantidad de oportunidades suficiente como para considerarse representativas. Por la poca cantidad de las oportunidades, una (GBP) llega al 100 % de aprobación y casi 0 % éxito y la otra (AUD) tiene gran porcentaje de éxito y aprobación pero nulo monto total y promedio de frigorías. El Yen Japonés tiene la misma distribución que la región Japan y era de esperarse ya que en esa región predomina la moneda del país. Por lo tanto, las monedas que predominan son el euro, dolar Estadounidense y el yen japones (en menor medida). El porcentaje de éxito entre el euro y el dolar estadounidense es muy parecido pero el dolar tiene mas cantidad de oportunidades por lo tanto lo consideramos como la mejor moneda para las oportunidades (justamente hay una columna que tiene el valor promedio de la frigoría en una moneda unificada y la elegida fue USD)

7. Conclusiones

En base al análisis exploratorio realizado se pudo llegar a las siguientes conclusiones:

Si bien algunas columnas tenían un gran porcentaje de información faltante y otras no contenían ninguna, el conjunto de datos era lo suficientemente grande como para poder hacer un análisis bastante completo.

Para mayor claridad este trabajo fue dividido en distintas secciones las cuales se centran en la investigación de un grupo específico de características.

- Del análisis geográfico podemos concluir que tanto EMEA como Americas fueron las regiones que mas oportunidades crearon, siendo dentro de estas Alemania y NW América los territorios que acumulan la mayoría de las mismas. A diferencia del resultado por cantidad de oportunidades, se puede observar que las regiones con mejor rendimiento fueron APAC y Japon, dado que obtuvieron mayor cantidad de ventas en relación al total de oportunidades generadas. Otra cosa que podemos observar es que la región de Middle East tiene muy pocas oportunidades en comparación con las otras regiones. Esto hace que sea mas difícil obtener conclusiones de esta región en particular, como se vió a lo largo del trabajo en donde gran parte de las conclusiones se centran en las otras regiones.

En particular si analizamos el momento del año en el cual se da el pico de ventas, vemos que en general el cuarto trimestre es el mejor salvo en la región de EMEA donde el segundo cuatrimestre es en el que se da las mayor cantidad de ventas.

Dada la imposibilidad de concluir qué región fue la mejor debido a la disparidad entre cantidad de oportunidades y rendimiento de las mismas, y aprovechando que en todas las regiones el precio promedio de la frigoría es muy similar, se hizo un análisis de la ganancia total, dando por resultado que las regiones con mayor cantidad de oportunidades terminan siendo las que mas ganancias obtuvieron (Americas y EMEA).

También se puede concluir que para todas las regiones, en mayor o menor proporción, se cumple que si una oportunidad fue aprobada en términos de precio y entrega u obtuvo la aprobación de los códigos burocráticos, tiene mas probabilidad de finalizar exitosamente.

- Al realizar un análisis temporal de los datos podemos ver que el 2017 (especialmente los meses de Mayo y Junio) fue el año que mayor oportunidades se crearon, seguido por el 2018 y luego el 2016. Así mismo, el 2017 fue el año que mayor cantidad de oportunidades se cerraron de manera exitosa. Además, en 2017, no solo hubo incremento de oportunidades y ventas sino también un pico de captación de clientes a mediados de este mismo año. En la misma linea, y haciendo un análisis de las personas que mas ganancias acumularon, vemos como la persona 'Person Name 19' tuvo un

incremento muy grande en su facturación en este año, siendo esta la persona que mas facturó. Con estos datos podemos concluir que desde la creación de la empresa, el 2017 fue el año mas exitoso, teniendo una pequeña merma en el año siguiente.

Por otro lado, observando ciertos datos a través del tiempo, encontramos algunas tendencias que podrían resultar de interés:

- El precio de las frigorías, se mantuvo estable y con una leve tendencia a la baja a lo largo del tiempo.
- Cuanto mayor es la cantidad de días que transcurren desde la creación de una oportunidad hasta su conclusión, la proporción de éxito disminuye. Con lo cual se podría decir que a medida que la oportunidad permanece mas tiempo abierto, la probabilidad de que esta se cierre de manera exitosa disminuye.
- A pesar de que en la mayoría de las oportunidades el vendedor responsable de la cuenta comercial es el mismo que el vendedor responsable de la cuenta cliente, tiene mejor porcentaje de éxito cuando la oportunidad tiene responsables distintos.
- En cuanto a los vendedores responsables de la cuenta comercial, que uno tenga muchas oportunidades o un porcentaje de éxito muy grande, no garantiza que sea el que mas ganancias obtuvo.
- El éxito de una oportunidad esta relacionado con el monto y frigorías de la misma, de manera que si la cantidad de frigorías aumenta, aumenta su costo y por ende si es mas costosa, tiene mas probabilidad de no finalizar exitosa.