

Proyecto de investigación.
Exportaciones de Colombia año 2022.

Seminario de investigación 1

Docente:
Wilmer Lopez

Primera entrega: conocimientos de estadísticas

Estudiantes:
Jose Luis Gomez
Marcela Cepeda

Maestría Analítica de datos
Facultad de ingeniería
Universidad Central
1 de junio de 2023

Introducción

Colombia un país de diversas fuentes de riqueza se encuentra en el puesto 61 del ranking de los 191 países exportadores, (Datosmacro.com, s.f.); Su constante crecimiento y evolución en sus exportaciones juegan un papel crucial para el desarrollo económico social y territorial, es por ello que decidimos realizar un completo análisis de los datos de las exportaciones del año 2022, buscando así, comprender el comportamiento del mercado nacional y porque no internacional.

Este estudio proporciona el conocimiento de diferentes variables que contribuyen a la toma de decisiones, considera una interpretación casuística y completa de la base tomada de la página del DANE <https://microdatos.dane.gov.co/index.php/catalog/472/data-dictionary> con el que identificamos las fortalezas y debilidades de la economía colombiana en términos de su capacidad de exportación, lo que permitiría a las empresas y al gobierno tomar medidas para mejorar la competitividad del país en el mercado global.

Escogimos este proyecto porque nos llamó la atención la fuente de la información, la calidad, las variables que contiene y el resultado que se puede generar de una excelente interpretación, sabemos que nos faltara muchos mas por abarcar e interpretar mas sin embargo buscamos e realizamos la interpretación de las principales preguntas que se pueden hacer los lectores que quieran conocer el detalle y características de las exportaciones en Colombia de la vigencia 2022.

Nuestro objetivo personal y educativo en este escrito es describir las diferentes variables y conclusiones que sacamos de los datos recolectados, mostrar los diferentes comportamientos de las variables y deducir bajo las estadísticas, modelos y graficas entre otros vistos en clase bajo el lenguaje Python.

Esperamos sea de su mayor interes...

Datos Generales de la tabla:

| | |
|----------------------|---|
| Fuente Base: | https://www.dane.gov.co/index.php/estadisticas-por-tema/comercio-internacional/exportaciones |
| Notebook | Exportaciones_2022 |
| Dataframe: | df_Exportaciones_2022 |
| Total, de registros: | 15.752.736 |
| Total, Filas | 555.812 |
| Total, Columnas: | 28 |

Clasificación de datos:

| Categóricos | |
|-------------|----------------------------|
| Ordinales | Nominales |
| Fecha | Modalidad Exportación |
| | Ciudad Aduana |
| | País destino |
| | Medio de transporte |
| | Unidad de medida comercial |

| Numéricos | |
|--------------------|--------------------|
| Continuos | Discretos |
| Total, Kilos | Cantidad de medida |
| Valor dólares | |
| Valor Pesos | |
| Valor de aranceles | |
| Valor Fletes | |
| Valores seguros | |
| Precio cantidad | |

Ordenamiento de datos:

1. En el primer notebook realizado en el mes de abril para la primera exposición reemplazamos los títulos, dejándolos mas precisos para cada variable, dejamos únicamente información de 1 año (2022), leímos cada una de las variables identificando las principales, así:

| | | | |
|-----------------|--------------------------|---------------|-----------------------------|
| FECH | Año | REGIM | Régimen |
| ADUA | Cod_Aduana | MODAD | Modalidad_Expo |
| PAIS | Cod_País_Destino | FINALID | Forma_de_Pago |
| COD_PAIS | Pais_Destino | CER_ORI1 | Tipo_Certificado |
| COD_SAL | Cod_Lugar_salida | SISESP | Sistema_Especiales |
| DPTO2 | Dpto_Procedencia | POSAR | Posicion_Arancelaria |
| VIA | Transporte | DPTO1 | Origen_Posicion |
| BANDERA | Nacionalida_Transporte | UNID | Cod_Unidad_Comercial |
| CANTI | Cantidad_Unidades | PBK | Kilos_Brutos |
| PNK | Kilos_Netos | FOBDOL | Valor dolares |
| FOBPES | Valor_Pesos | AGRENA | Valor_Agregado |
| FLETES | FLETES | SEGURO | SEGURO |
| OTROSG | Otros_gastos | NIT | Nit |
| RAZ_SIAL | Razón_Social | | |

Descripción campos

Año: Año y mes de Registro

Pais_Destino: Corresponde al código de descripción del país al que se envía la mercancía.

Lugar Salida: Corresponde a la ciudad de salida de la mercancía

Régimen: Código del régimen (acorde con CAN)

Modalidad Expo: modalidad de la exportación Son las diferentes formas para la salida de las mercancías de acuerdo con la finalidad que le quiera dar el exportador. Están definidas en el Decreto 1165 de 2019 y se clasifican en tres grupos Definitiva, Temporal y Especial.

Tipo Certificado: Tipo de certificado de origen documentos físicos que constituyen los certificados, existen tres grandes categorías: certificados de origen para exportación, certificados de origen para importación, y certificados de origen para tránsito.

Sistemas Especiales: Es un mecanismo que permite importar temporalmente bienes con exención, suspensión total o parcial de derechos e impuestos a la importación o con el diferimiento del pago del IVA, única y exclusivamente cuando éstos han de ser incorporados a través de un proceso industrial, o en un bien final con destino a la exportación.

Posición Arancelaria: La clasificación arancelaria consiste en la asignación de un código numérico a las mercancías que se denomina subpartida arancelaria, teniendo como presupuesto fundamental el conocimiento de sus características físicas, químicas y técnicas, entre otras, en el marco de la Nomenclatura Arancelaria vigente.

Origen Posición: Departamento de origen.

Cod_Unidad_Comercial: unidad comercial de medida en Valores.

Unidad_Comercial: unidad comercial de medida.

Cantidad_Unidades: Cantidad de unidades exportadas.

Kilos_Brutos: Es el peso de la mercancía incluyendo los empaques, tanto interiores como exteriores que se utilizan para evitar el deterioro o proveer algún aislamiento especial, así como aquellas envolturas o elementos de presentación que comúnmente acompañan la mercancía para su venta directa.

Kilos_Netos: El peso de la mercancía, excluyendo los empaques que no sean necesarios para su venta directa.

Valor_Dolares: Corresponde al precio de venta de los bienes embarcados a otros países, puestos en el medio de transporte, sin incluir valor de seguro y fletes. Este valor que inicialmente se expresa en dólares americanos se traduce al valor FOB en pesos colombianos, empleando la tasa promedio de cambio del mercado correspondiente al mes de análisis.

Valor_Pesos: Corresponde al precio de venta de los bienes embarcados a otros países, puestos en el medio de transporte, sin incluir valor de seguro y fletes. Este valor que inicialmente se expresa en dólares americanos se traduce al valor FOB en pesos colombianos, empleando la tasa promedio de cambio del mercado correspondiente al mes de análisis.

Valor_Agregado: Valores adicionales por el envío.

Fletes: Tarifa básica pactada entre el transportador y el usuario del servicio. Por este valor el primero se compromete a trasladar la mercancía desde un punto de origen hasta el de destino acordado previamente.

Ciudad_Aduana: Ciudad origen de la mercancía.

Categorización de Datos:

Creamos nuestra variable objetivo, de ella sacamos las diferentes categorizaciones para describir sus comportamientos y comparaciones con otras variables importantes de la base, como lo fueron:

| | |
|----------------------------|---|
| Aduana_producto_MOD | <pre>def categorizar_fila(row): if row['Aduana_producto'] == 'Rosas frescas': return 'Rosas frescas' if row['Aduana_producto'] == 'claveles frescos, cortados para ramos o adornos': return 'claveles frescos, cortados para ramos o adornos' if row['Aduana_producto'] == 'Hortensias': return 'Hortensias' if row['Aduana_producto'] == 'Cafés sin tostar': return 'Cafés sin tostar' if row['Aduana_producto'] == 'flores y capullos frescos': return 'flores y capullos frescos' else: return 'OTROS'</pre> |
|----------------------------|---|

| | |
|-------------|---|
| SALIDA_MOD | <pre>def categorizar_fila(row): if row['Lugar_salida'] == 'BOG': return 'BOGOTA' if row['Lugar_salida'] == 'CTG': return 'CARTAGENA' if row['Lugar_salida'] == 'BUN': return 'BUENAVENTURA' else: return 'OTROS'</pre> |
| PRODEXP_MOD | <pre>def categorizar_fila(row): if row['Posicion_Arancelaria'] == '2709000000': return 'Aceites de petroleo' if row['Posicion_Arancelaria'] == '2701120010 ': return 'Hullas termicas' if row['Posicion_Arancelaria'] == '901119000': return 'Cafes sin tostar' if row['Posicion_Arancelaria'] == '7108120000': return 'Oro' else: return 'OTROS'</pre> |

Se agrego a las variables la información clara y concisa de las de diferentes variables que venían codificadas de acuerdo a el manejo la data, por ejemplo, en Ciudades_aduana venia con códigos 1,2,3,4... el cual se homologaron a los nombres de la ciudad correspondiente:

| df_ciudades_aduana | Cod_Aduana | Ciudad_Aduana |
|--------------------|------------|----------------------------------|
| | 1 | Armenia |
| | 2 | Barranquilla |
| | 3 | Especial de Aduanas de Bogotá |
| | 4 | Bucaramanga |
| | 5 | Cali |
| | 6 | Especial de Aduanas de Cartagena |
| | 7 | Cúcuta |
| | 10 | Manizales |
| | 11 | Medellín |
| | 16 | Pereira |
| | 19 | Santa Marta |

| df_Dpto_Procedencia | <table> <tr> <th>Dpto_Procedencia</th><th>Origen_Procedencia</th></tr> <tr><td>1</td><td>PETROLEO Y DERIVADOS</td></tr> <tr><td>5</td><td>ANTIOQUIA</td></tr> <tr><td>8</td><td>ATLANTICO</td></tr> <tr><td>11</td><td>BOGOTA</td></tr> <tr><td>13</td><td>BOLIVAR</td></tr> <tr><td>15</td><td>BOYACA</td></tr> <tr><td>17</td><td>CALDAS</td></tr> <tr><td>18</td><td>CAQUETA</td></tr> <tr><td>19</td><td>CAUCA</td></tr> <tr><td>20</td><td>CESAR</td></tr> <tr><td>23</td><td>CORDOBA</td></tr> <tr><td>25</td><td>CUNDINAMARCA</td></tr> <tr><td>27</td><td>CHOCO</td></tr> </table> | Dpto_Procedencia | Origen_Procedencia | 1 | PETROLEO Y DERIVADOS | 5 | ANTIOQUIA | 8 | ATLANTICO | 11 | BOGOTA | 13 | BOLIVAR | 15 | BOYACA | 17 | CALDAS | 18 | CAQUETA | 19 | CAUCA | 20 | CESAR | 23 | CORDOBA | 25 | CUNDINAMARCA | 27 | CHOCO |
|----------------------|---|----------------------|--------------------|------------|----------------------|------------|------------------|-----------|-----------------|------------|--------|------------|--------------------|-----|------------|----|---|----|---|----|-------------------------|----|---|----|---------------------------------------|----|--------------|----|-------|
| Dpto_Procedencia | Origen_Procedencia | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | PETROLEO Y DERIVADOS | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 5 | ANTIOQUIA | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 8 | ATLANTICO | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 11 | BOGOTA | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 13 | BOLIVAR | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 15 | BOYACA | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 17 | CALDAS | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 18 | CAQUETA | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 19 | CAUCA | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 20 | CESAR | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 23 | CORDOBA | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 25 | CUNDINAMARCA | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 27 | CHOCO | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| df_Medio_transporte | <table> <tr> <th>Transporte</th><th>Medio_Transporte</th></tr> <tr><td>1</td><td>MARITIMO</td></tr> <tr><td>2</td><td>FERREO</td></tr> <tr><td>3</td><td>TERRESTRE</td></tr> <tr><td>4</td><td>AEREO</td></tr> <tr><td>5</td><td>CORREO</td></tr> <tr><td>6</td><td>MULTIMODAL</td></tr> <tr><td>7</td><td>INSTALACIONES DE TRANSPORTE FIJAS (TUBERIAS, C...</td></tr> <tr><td>8</td><td>VIAS NAVEGABLES INTERIORES</td></tr> <tr><td>9</td><td>OTRO MODO DE TRANSPORTE</td></tr> </table> | Transporte | Medio_Transporte | 1 | MARITIMO | 2 | FERREO | 3 | TERRESTRE | 4 | AEREO | 5 | CORREO | 6 | MULTIMODAL | 7 | INSTALACIONES DE TRANSPORTE FIJAS (TUBERIAS, C... | 8 | VIAS NAVEGABLES INTERIORES | 9 | OTRO MODO DE TRANSPORTE | | | | | | | | |
| Transporte | Medio_Transporte | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | MARITIMO | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | FERREO | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3 | TERRESTRE | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4 | AEREO | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 5 | CORREO | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 6 | MULTIMODAL | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 7 | INSTALACIONES DE TRANSPORTE FIJAS (TUBERIAS, C... | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 8 | VIAS NAVEGABLES INTERIORES | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 9 | OTRO MODO DE TRANSPORTE | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| df_Aduana_producto | <table> <tr> <th>Posicion_Arancelaria</th><th>Aduana_producto</th></tr> <tr><td>2709000000</td><td>Aceites de petróleo</td></tr> <tr><td>2701120010</td><td>Hullas térmicas.</td></tr> <tr><td>901119000</td><td>café sin tostar</td></tr> <tr><td>7108120000</td><td>Oro</td></tr> <tr><td>2710192100</td><td>Gasóleos (gasóleo)</td></tr> <tr><td>...</td><td>...</td></tr> <tr><td>94</td><td>Muebles</td></tr> <tr><td>95</td><td>Juguetes, artículos para recreo, deporte; part...</td></tr> <tr><td>96</td><td>Manufacturas diversas</td></tr> <tr><td>97</td><td>Objetos de arte, de colección o de antigüedad</td></tr> <tr><td>98</td><td>Disposiciones de tratamiento especial</td></tr> </table> | Posicion_Arancelaria | Aduana_producto | 2709000000 | Aceites de petróleo | 2701120010 | Hullas térmicas. | 901119000 | café sin tostar | 7108120000 | Oro | 2710192100 | Gasóleos (gasóleo) | ... | ... | 94 | Muebles | 95 | Juguetes, artículos para recreo, deporte; part... | 96 | Manufacturas diversas | 97 | Objetos de arte, de colección o de antigüedad | 98 | Disposiciones de tratamiento especial | | | | |
| Posicion_Arancelaria | Aduana_producto | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2709000000 | Aceites de petróleo | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2701120010 | Hullas térmicas. | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 901119000 | café sin tostar | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 7108120000 | Oro | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2710192100 | Gasóleos (gasóleo) | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ... | ... | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 94 | Muebles | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 95 | Juguetes, artículos para recreo, deporte; part... | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 96 | Manufacturas diversas | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 97 | Objetos de arte, de colección o de antigüedad | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 98 | Disposiciones de tratamiento especial | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Variable Objetivo:

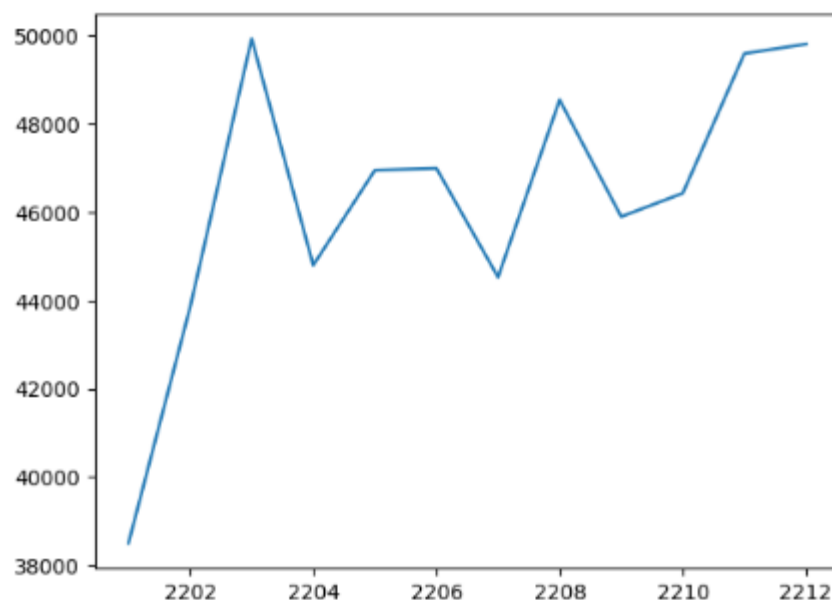
Luego de conocer la base e identificar estas variables principales hallamos nuestra variable objetivo el cual está compuesta por el valor en dólares dividida por la cantidad de la unidad de medida, para hallar el valor dólar por unidad:

```
1  ## Variable Objetivo
2  df['Precio_Cantidades'] = df['Valor dolares'] / df['Cantidad_Unidades']
```

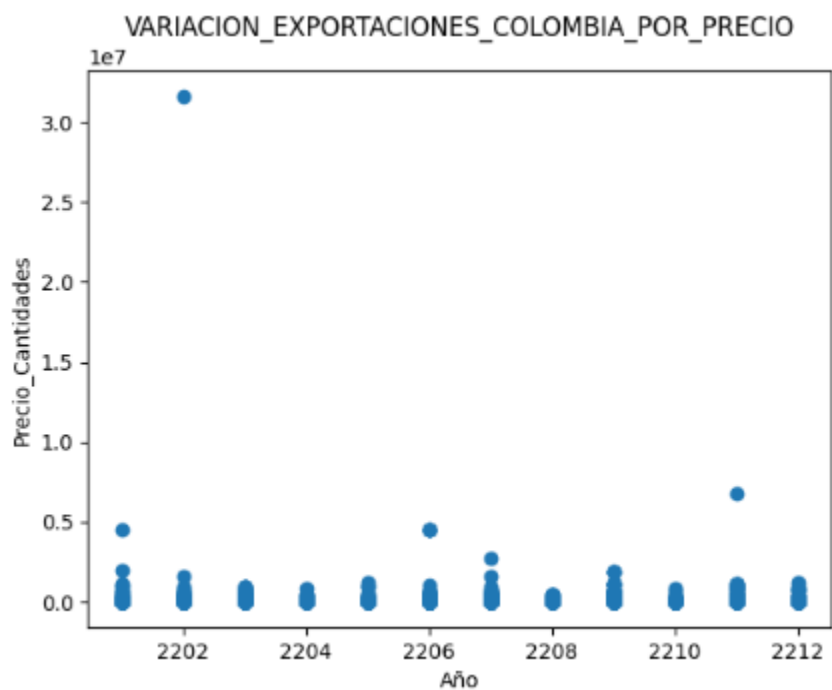
La descripción de nuestra variable objetivo es:

| | | | |
|-----------------|---------------------------|---------------------|----------------------|
| Máximo | 31.625.000 dólares x und. | Percentil 25 | 1.09 dólares x und. |
| Mínimo | 0 dólares x und. | Percentil 50 | 4.45 dólares x und. |
| Media | 977 dólares x und. | Percentil 75 | 14.89 dólares x und. |
| Estándar | 47.968.22 dólares x und. | cantidad | 555.812 filas |

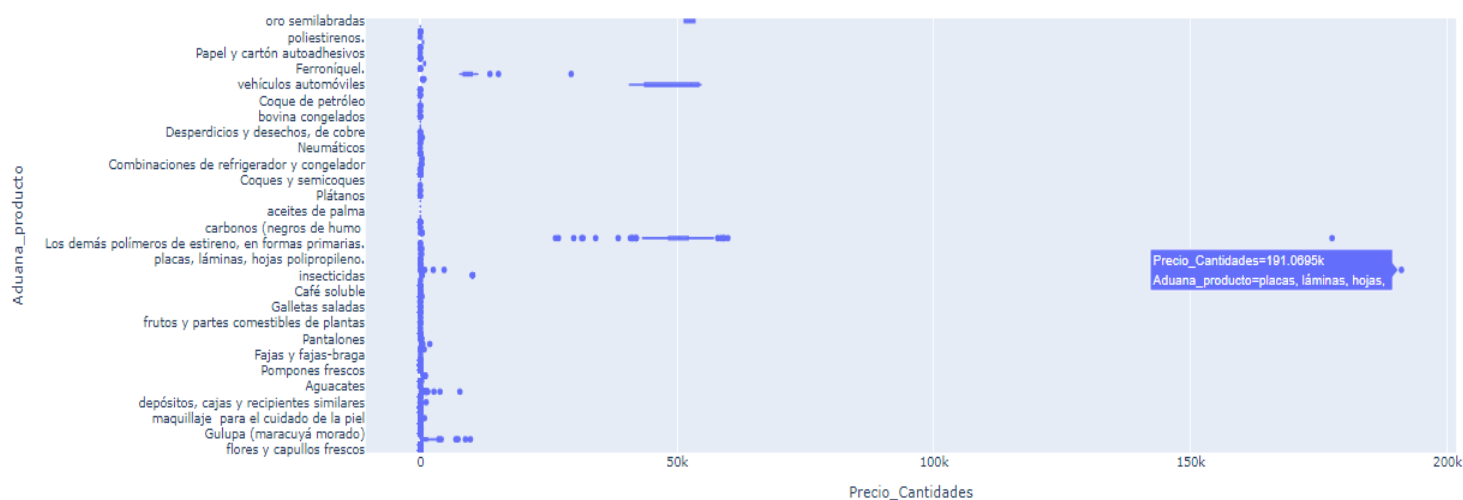
Las exportaciones comienzan a incrementar a partir mes de marzo del año 2022, esto puede darse a que después del mes de diciembre baja la volumetría y la cantidad de producción y a partir del mes de marzo inicia la volatilidad de cantidades de veces que se exportan:



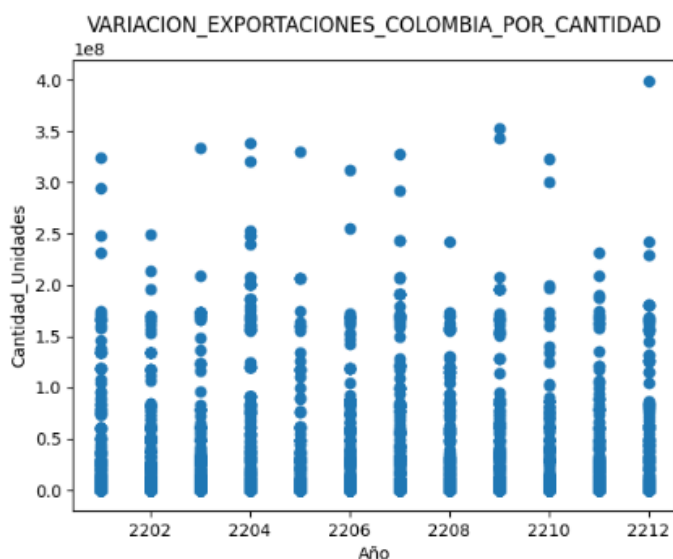
En detalle de nuestra variable objetivo (Precio_Cantidad) vemos que tenemos un dato extremo de los demás, el precio oscila entre los 31mil dólares por Metro lineal en el mes de febrero; decimos metro lineal dado que corresponde al producto Placas laminas y hojas polipropileno.



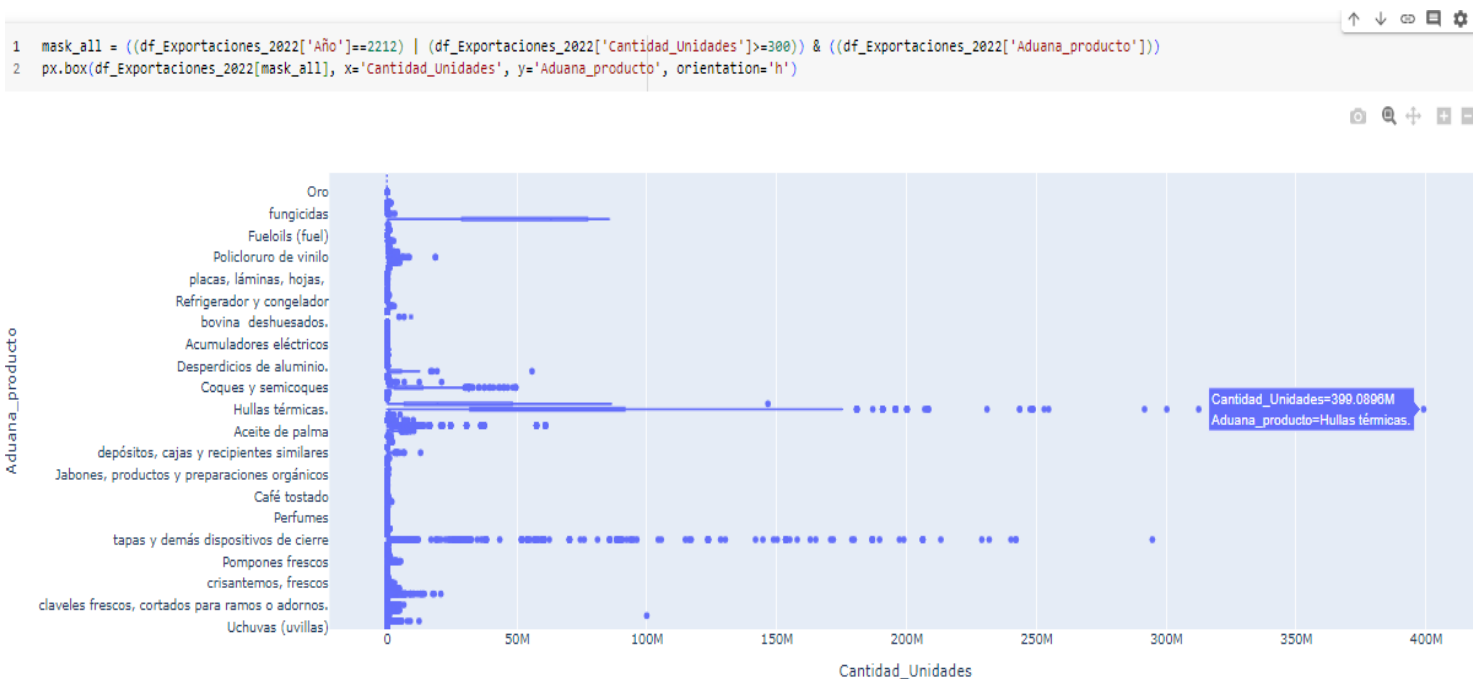
```
1 mask_all = ((df_Exportaciones_2022['Año']==2202) | (df_Exportaciones_2022['Precio_Cantidades']>=30625000.0)) & ((df_Exportaciones_2022['Aduana_producto']))
2 px.box(df_Exportaciones_2022[mask_all], x='Precio_Cantidades', y='Aduana_producto', orientation='h')
```



a nuestro anterior estudio nos surgió la duda de cual es el producto con mayor cantidad de medida exportado para la vigencia 2022; este se da en el mes de diciembre:



El producto con mayor unidad de medida (399mil Kilogramos) exportados, corresponde a las Hullas térmicas, una roca sedimentaria orgánica, un tipo de carbón mineral que contiene entre un 80 y un 90 % de carbono. Es dura y quebradiza, estratificada, de color negro y brillo mate o también graso (Wikipedia.org, s.f.):

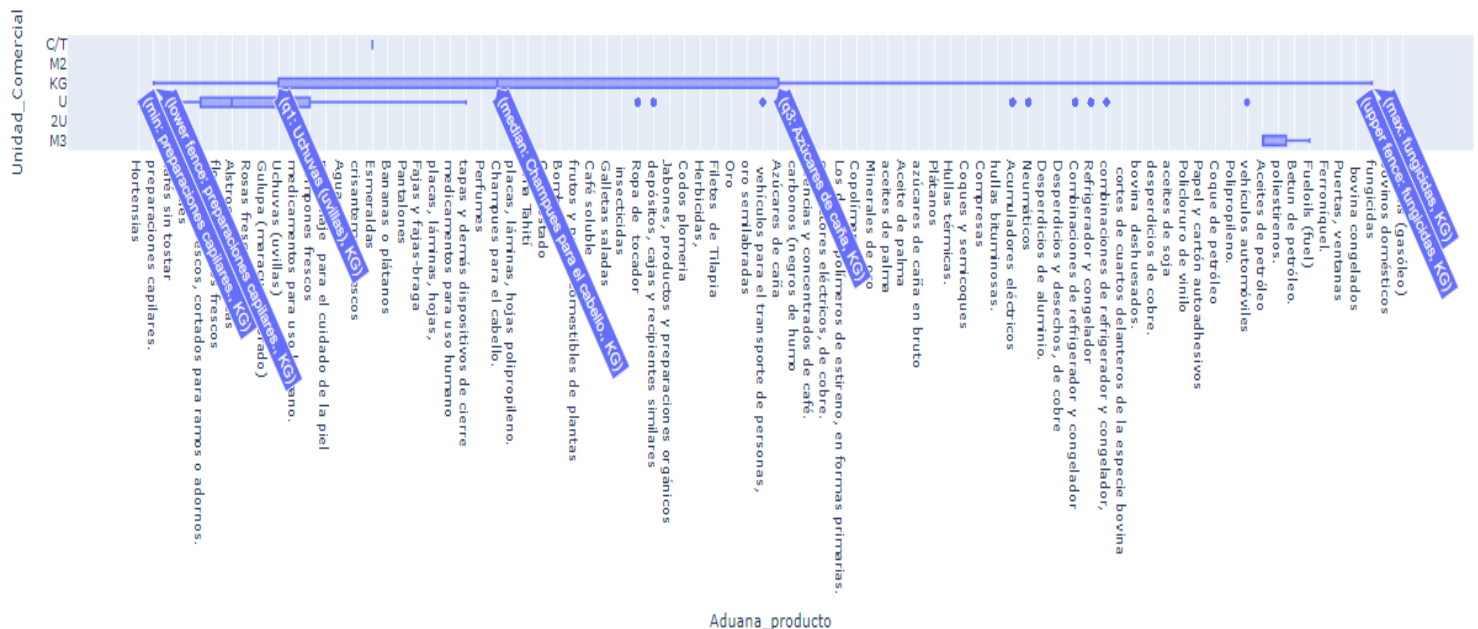


Hay tres variedades:

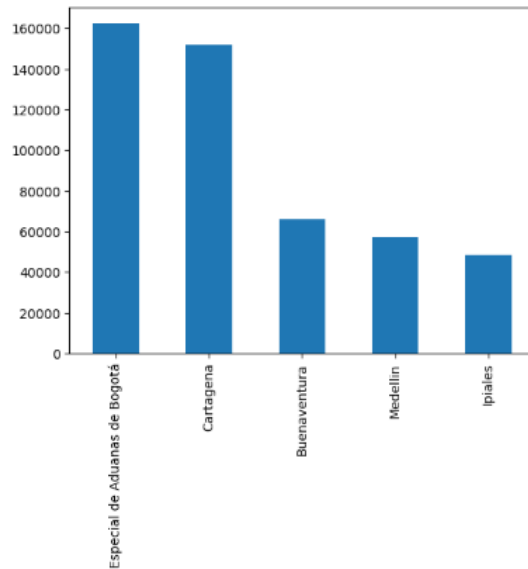
- Hulla grasa, antiguamente al destilarla se obtenía gas de alumbrado.
- Hulla magra o seca, que se emplea como combustible. Su aspecto presenta bandas mate.
- Hulla semiseca: Es alternada con bandas brillantes.



En la data cada uno de los productos detallan la unidad de medida, para toda la vigencia 2022 las exportaciones con mayor cantidad de medida se dan por Kilogramos con un dato máximo (extremo) en los productos como fungicidas, con un cuartil 75 en azúcares de caña, una media en los champées para cabello, el cuartil 25 con uchuvras y un mínimo con los productos de preparaciones capilares:

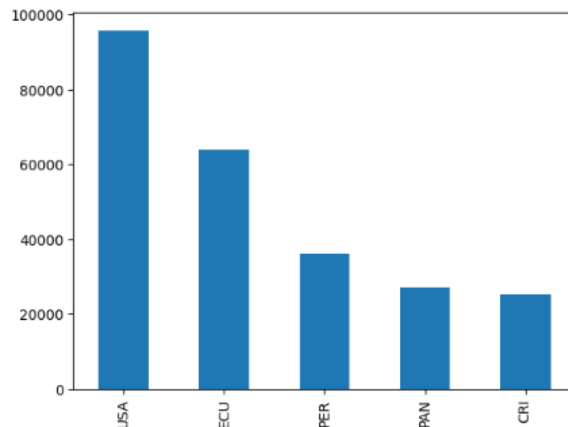


La ciudad que más legaliza la exportación de mercancías se centra en Bogotá, con un aproximado de 160mil legalizaciones para la vigencia del 2022, supera las cantidades de legalización de las ciudades portuarias como Cartagena o buenaventura. Una opción de este comportamiento es que en Bogotá se encuentran la mayoría de parques industriales y zonas francas:

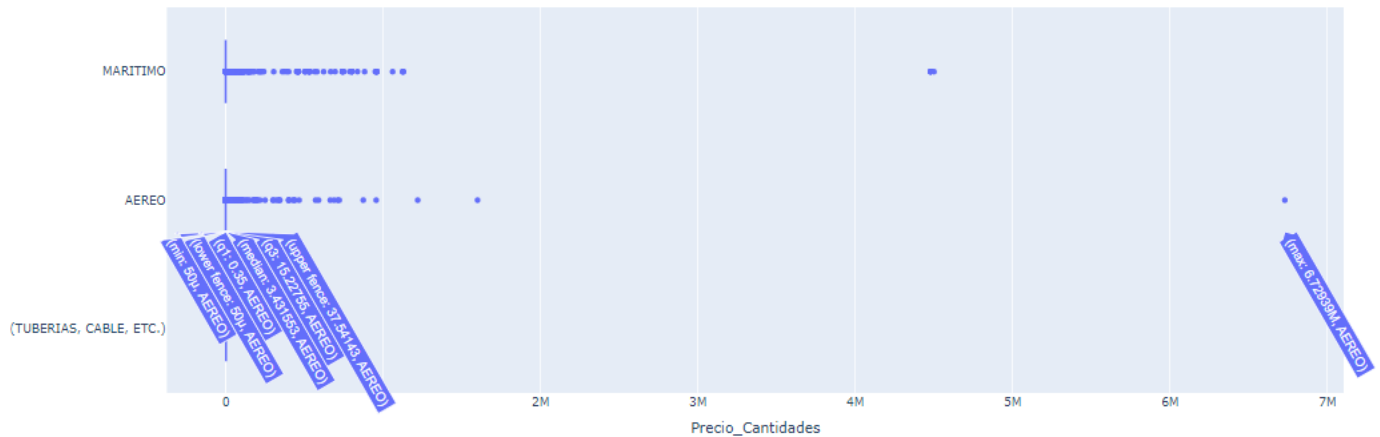


El país destino donde proveemos con mayor frecuencia nuestros productos es Estados Unidos, desde el año 2018 somos sus principales proveedores de todo tipo de flores, esto es posible ya que nuestro país por su cultura y su riqueza natural es segundo en floricultor después de Holanda.

En el 2017 Colombia contaba con un total de 7.700 hectáreas de producción de flores hoy en día cuanta con un 15% más hectáreas enfocadas al incremento de indicadores de exportación.



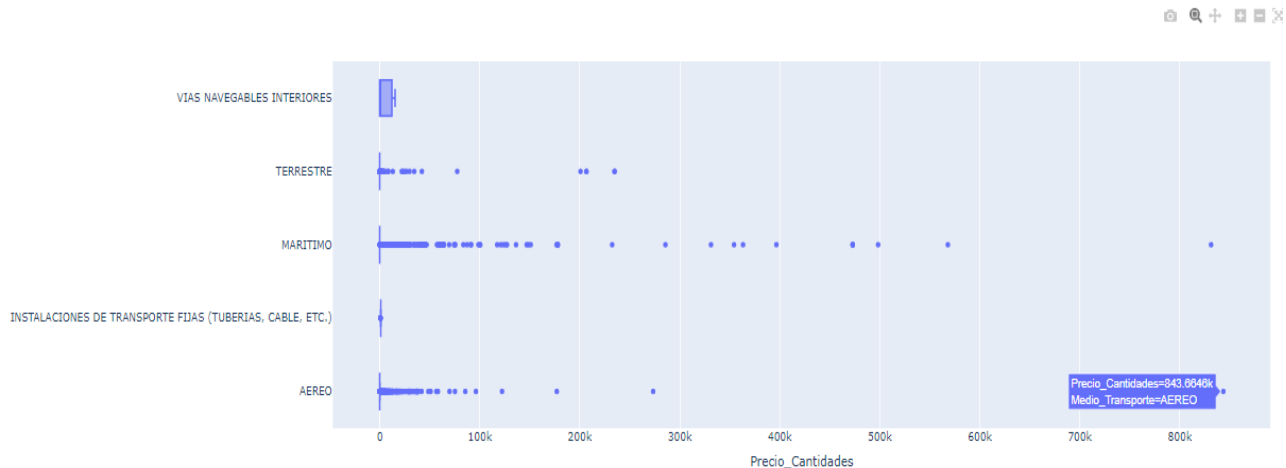
Teniendo en cuenta lo anterior iniciamos a indagar mas en nuestra data, buscando así entrar en detalle de cada una de las variables, e iniciamos con un filtro del país estados unidos siendo a donde con mas frecuencia exportamos, queriendo conocer el medio de transporte que más usado comparado con nuestra variable objetivo (precio_cantidad) y así logramos conocer el medio de transporte en el cual transportamos el producto más costoso por unidad, siendo el: Aereo



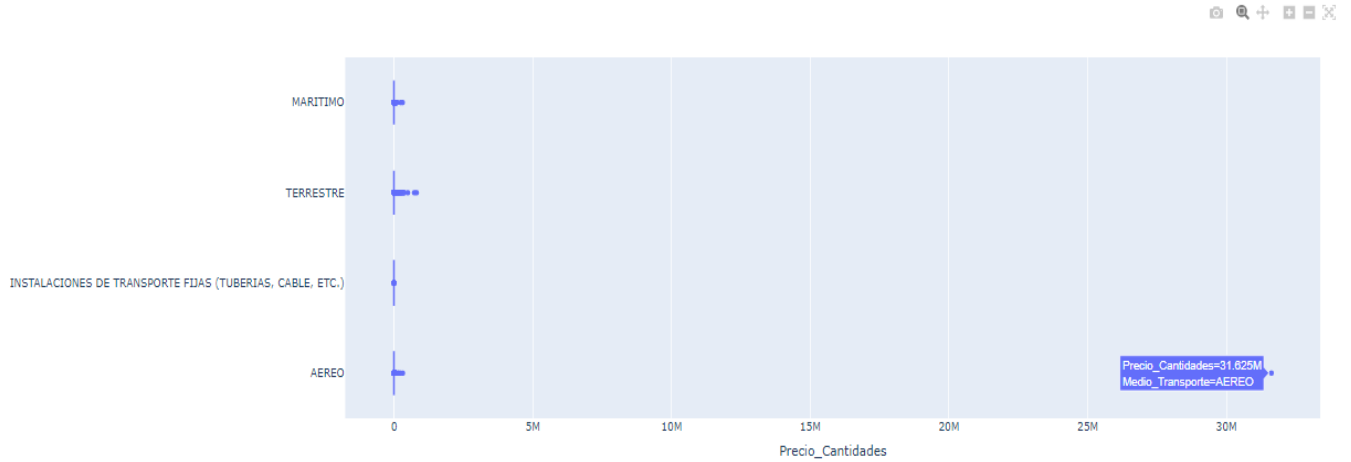
Recordemos que anteriormente hablábamos de un preferido Marítimo, en este ejemplo estamos visualizando únicamente exportaciones para los estados Unidos, y encontramos el un dato extremo de 6.7 dólares por unidad de medida; un percentil 75 de 15.22 dólares por unidad de medida; una media de 3.43 dólares por unidad de medida.

Ahora bien, quisimos hacer las mismas comparaciones con los siguientes países con mayor frecuencia de exportación, Perú (843 dólar por unidad) y Ecuador (31.6 dólares por unidad), encontrando que el aéreo sigue siendo pionero en el costo por unidad de medida al momento de exportar:

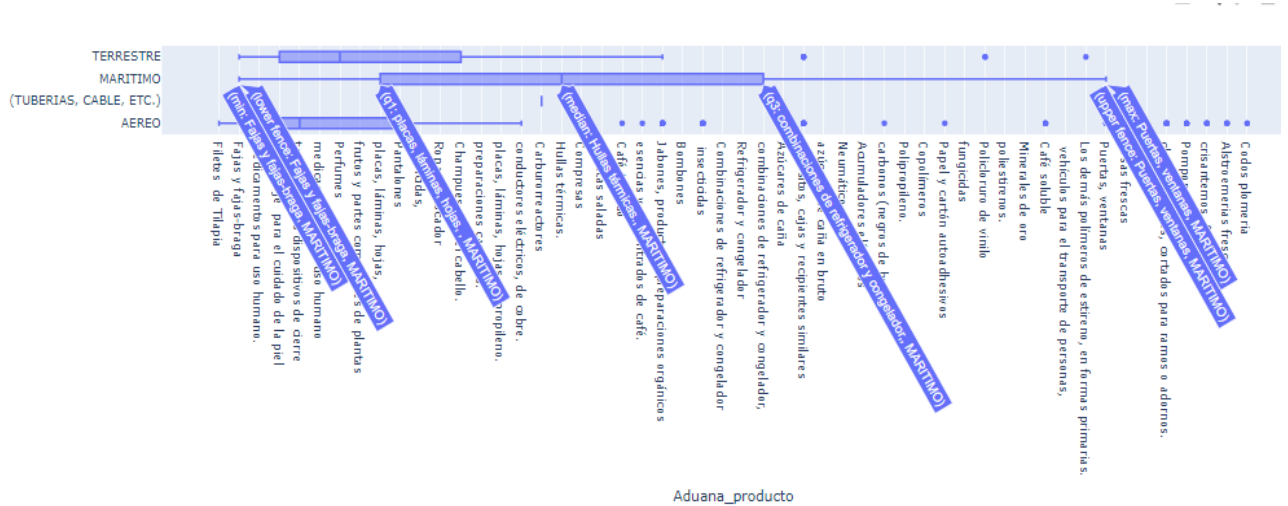
```
mask2 = df_Exportaciones_2022['Pais_Destino']=='PER'
px.box(df_Exportaciones_2022[mask2], x='Precio_Cantidades', y='Medio_Transporte', orientation='h')
```



```
mask1 = df_Exportaciones_2022['Pais_Destino']=='ECU'
px.box(df_Exportaciones_2022[mask1], x='Precio_Cantidades', y='Medio_Transporte', orientation='h')
```



Ahora, quisimos conocer cuáles son los productos que exportamos a los diferentes países a los que les proveemos, y el medio que usamos para ello, e iniciamos con Perú:



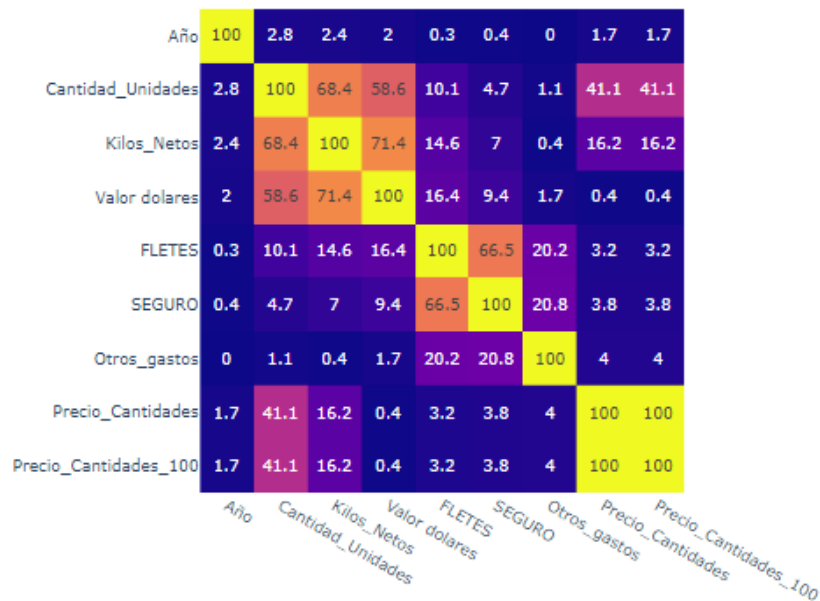
Aquí identificamos que usamos el medio de transporte marítimo para exportar los siguientes productos principales hacia Perú:

| | |
|--------------|--|
| Mayor | Puertas, ventanas |
| Percentil 75 | Combinaciones de refrigeradores y congeladores |
| Media | Hullas térmicas |
| Mínimos | Fajas y fajas-braza |



Ahora damos a conocer el comportamiento para Estados Unidos...

- A medida que la cantidad de unidades de medida, aumenta así mismo el precio de unidad de medida.
- A medida que aumentan los seguros aumentan el valor de los fletes.
- A medida que aumentan los kilos netos aumenta el valor en dólares.
- A medida que aumentan las cantidades aumenta el valor en dólares.



Sección de cortes:

Después de este conocimiento descriptivo, continuamos con el estudio del Dataframe con los cortes de la variable objetivo (Precio_Cantidad), para ello iniciamos haciendo los cortes de la siguiente manera:

| Cortes | Cantidades |
|---------|------------------|
| Corte 1 | De 0 a 1 dólar |
| Corte 2 | De 1 a 2 dólares |
| Corte 3 | De 2 a 3 dólares |

| Cortes | Cantidades |
|---------|------------------|
| Corte 4 | De 4 a 5 dólares |
| Corte 5 | De 5 a 6 dólares |
| Corte 6 | De 6 a 9 dólares |

| Cortes | Cantidades |
|---------|---------------------|
| Corte 7 | De 9 a 21 dólares |
| Corte 8 | De 21 a 100 dólares |
| Corte 9 | De 100 a máximo |

Aquí observamos que el percentil 75 se concentra en el corte No. 1 de 0 a 1 dólares:

```
(0, 1]      134170
(1, 2]      53689
(2, 3]      43847
(3, 6]      84297
(6, 9]      52327
(9, 21]     73885
(21, 100]   61639
(100, 31625000] 51958
Name: Precio_Cantidades_cortes, dtype: int64
```

Estos cortes los comparamos con la variable medio de transporte identificando la cantidad de veces que se utiliza el medio de transporte con cada corte, por ejemplo, el corte No. 1 que corresponde de 0 a 1 dólar por unidad vendida se utilizo exportando con instalaciones fijas (tuberías o cables) solo 12 veces en el año 2022.

| Medio_Transporte | AEREO | INSTALACIONES DE TRANSPORTE FIJAS (TUBERIAS, CABLE, ETC.) | MARITIMO | TERRESTRE | VIAS NAVEGABLES INTERIORES |
|--------------------------|-------|---|----------|-----------|----------------------------|
| Precio_Cantidades_cortes | | | | | |
| (0, 1] | 69883 | 12 | 54696 | 9578 | 1 |
| (1, 2] | 10715 | 0 | 35190 | 7784 | 0 |
| (2, 3] | 9946 | 0 | 27286 | 6615 | 0 |
| (3, 6] | 24155 | 0 | 47922 | 12219 | 1 |
| (6, 9] | 17382 | 0 | 27935 | 7010 | 0 |
| (9, 21] | 34524 | 0 | 27260 | 12101 | 0 |
| (21, 100] | 32300 | 0 | 21182 | 8155 | 2 |
| (100, 31625000] | 21509 | 1494 | 23366 | 5585 | 4 |

En este mismo precio de 0 a 1 dólares por unidad de medida exportada, es contratado cada uno de los medios de transporte, la relación anterior también nos deja ver un salto en el corte 4 de 3 a 6 dólares la cantidad de veces contratadas los tres principales medios de transporte (Aéreo, Marítimo y terrestre) aumenta.

Mostramos los tres principales medios, buscando identificar de cada uno los cortes que priman, entonces:

| Medio | Corte - Variable objetivo |
|-----------|---------------------------|
| Aéreo | Corte 1, 6 y 7 |
| Marítimo | Corte 1, 2 y 4 |
| Terrestre | Corte 4, 6 y 1 |

| Medio_Transporte | AEREO | MARITIMO | TERRESTRE |
|--------------------------|-------|----------|-----------|
| Precio_Cantidades_cortes | | | |
| (0, 1] | 69883 | 54696 | 9578 |
| (1, 2] | 10715 | 35190 | 7784 |
| (2, 3] | 9946 | 27286 | 6615 |
| (3, 6] | 24155 | 47922 | 12219 |
| (6, 9] | 17382 | 27935 | 7010 |
| (9, 21] | 34524 | 27260 | 12101 |
| (21, 100] | 32300 | 21182 | 8155 |
| (100, 31625000] | 21509 | 23366 | 5585 |

Contraste de hipótesis:

Para nuestra hipótesis nos vamos a centrar en el producto con mayor frecuencia de exportación, las flores, también vamos a tomar el mes donde mayor movimiento tubo este producto que fue en el mes de febrero y por último la ciudad aduana que legaliza esta exportación Cartagena; con la anterior información comparada con nuestra variable objetivo (precio_cantidad) conseguimos el siguiente test:

| | T | dof | alternative | p-val | CI99% | cohen-d | BF10 | power |
|--------|----------|-------|-------------|---------|--|----------|-------|----------|
| T-test | 1.932992 | 12406 | two-sided | 0.05326 | [858.7961530633505, 1861.707413727958] | 0.017354 | 0.066 | 0.489231 |

Que traduce el siguiente test estadístico:

| | |
|---------------------------------------|----------------------------|
| Limite seleccionado | 60 |
| Cantidad de registros del filtro | 12.406 registros |
| Probabilidad | 5.32% |
| Valor que está por debajo de la media | 858 dólares a 1861 dólares |
| Fuerza relativa de las diferencias | 1.73% |
| Fuerza | 48.92 |

En conclusión, decimos que existen evidencias suficientes para rechazar que las muestras proceden de poblaciones con misma media. (Rodrigo, 2021)

Hicimos el mismo ejercicio con la ciudad aduana Bogotá, también donde concluimos rechazo porque la probabilidad de crecimiento es poca:

| | T | dof | alternative | p-val | CI95% | cohen-d | BF10 | power |
|--------|-----------|--------------|-------------|---------|-------------------|----------|------|----------|
| T-test | -2.356853 | 20917.685424 | two-sided | 0.01844 | [-993.93, -91.35] | 0.029935 | 0.23 | 0.652218 |

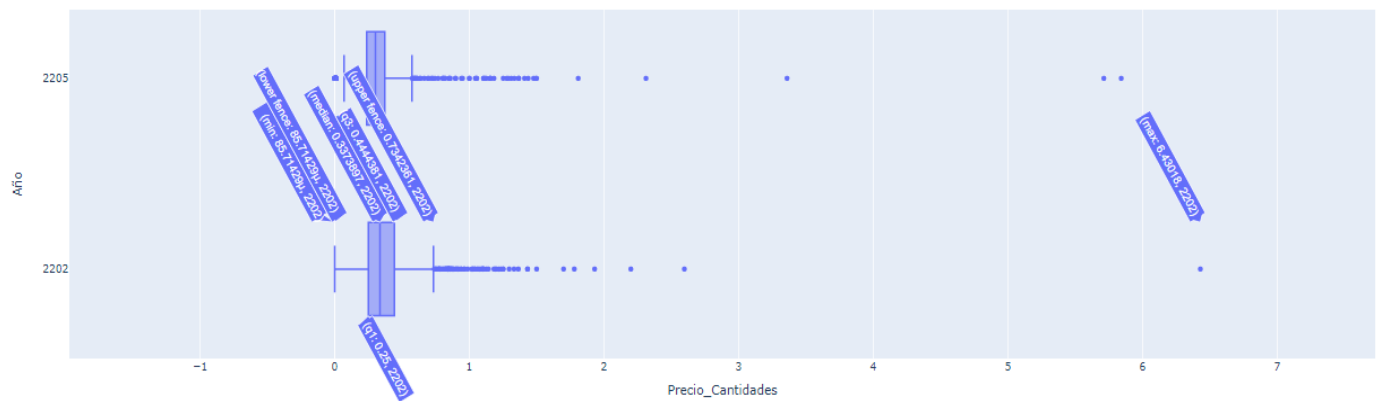
De lo anterior evaluamos diferencias en tras las variables, tomando como filtro los medes de febrero y mayo, por dos productos aduana; “Rosas frescas” y “Hortensias” comparadas con nuestra variable objetivo:

| | T | dof | alternative | p-val | CI95% | cohen-d | BF10 | power |
|--------|-----------|-------------|-------------|----------|---------------|----------|---------|----------|
| T-test | -4.460936 | 1379.487779 | two-sided | 0.000009 | [-0.26, -0.1] | 0.187279 | 923.418 | 0.989071 |

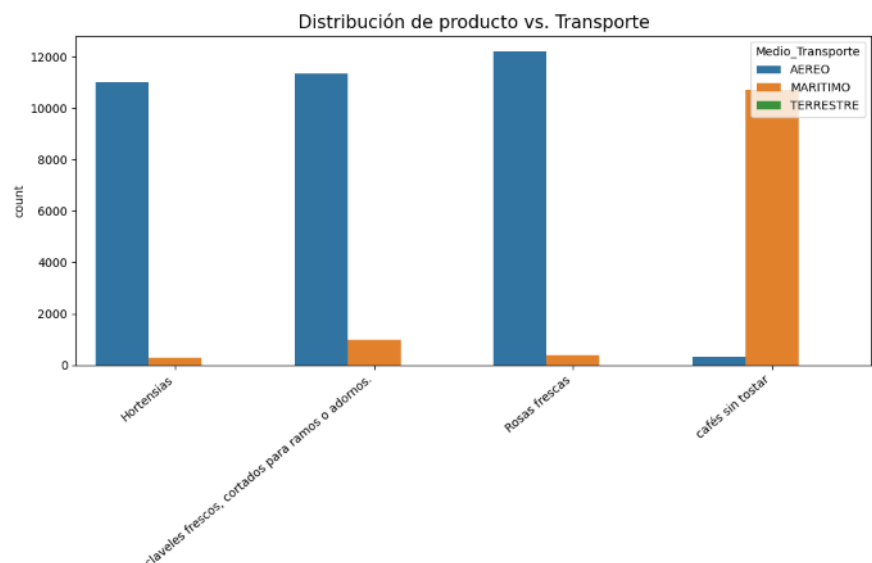
| | T | dof | alternative | p-val | CI95% | cohen-d | BF10 | power |
|--------|-----------|------|-------------|----------|---------------|----------|---------|----------|
| T-test | -4.254756 | 2070 | two-sided | 0.000022 | [-0.26, -0.1] | 0.187279 | 381.348 | 0.989071 |

Podemos concluir que el mes de febrero del 2022 es el mes con mayor exportación en precio * cantidad, con una media de 0.33 dólar por unidad exportada, aclaramos que escogimos estos dos meses dado que corresponden a el mes de san Valentín y el mes de las mamas.

```
1 mask_all = ((DF_MEDIOS_TRANSPORTE['Año']==2205) | (DF_MEDIOS_TRANSPORTE['Año']==2202)) & ((DF_MEDIOS_TRANSPORTE['Aduana_producto']=='Rosas frescas'))
2 px.box(DF_MEDIOS_TRANSPORTE[mask_all], x='Precio_Cantidades', y='Año', orientation='h')
```

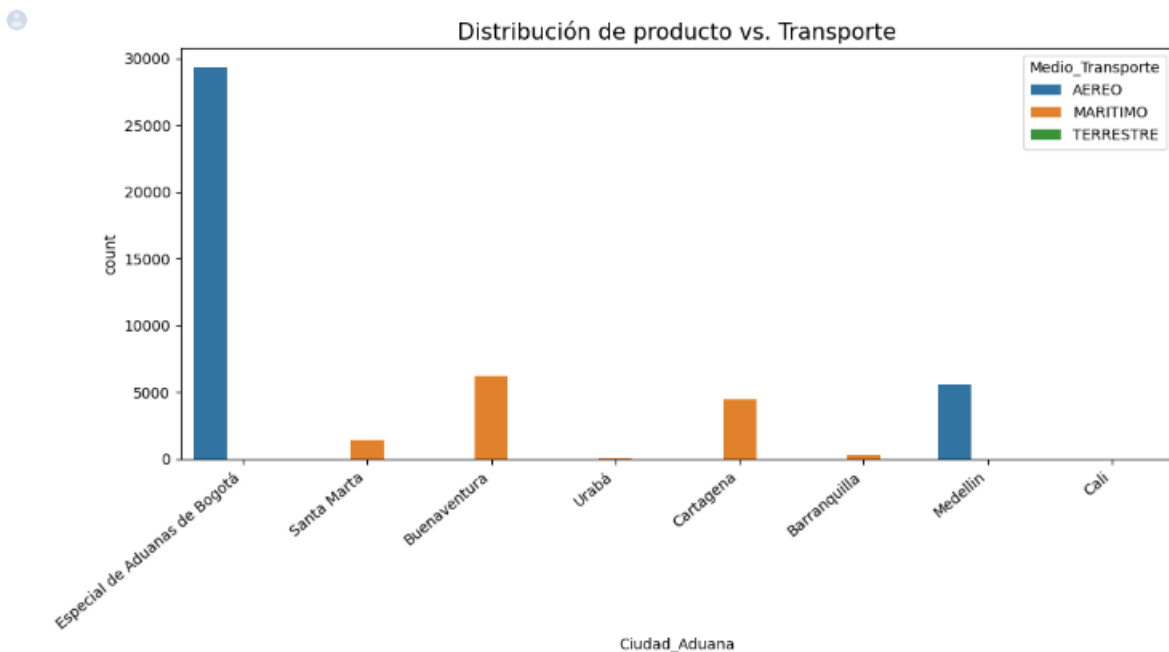


De la data también comparamos como se comporta el medio de transporte con los 5 productos con mayor frecuencia de exportación y con la siguiente grafica ilustramos el comportamiento de las misma.



Como lo muestra la distribución el producto que mas se exporta en modo marítimo es café sin tostar.

Ahora revisemos el comportamiento de la distribución del medio de transporte con la ciudad de legalización, como lo es muy común en las dos principales ciudades (Bogotá y Medellín) el medio de transporte utilizado por su geografía es el Aéreo; caso contrario para las ciudades portuarias (Santa marta, Buenaventura, Urabá, Cartagena, Barranquilla) el medio de transporte utilizado en el marítimo:



Modelo de Regresión Lineal:

Con el fin de desarrollar el modelo de Regresión, filtramos los 4 productos con mayor exportación en el año 2022

```
df_Exportaciones = df_Exportaciones[(df_Exportaciones['Aduana_producto'] == 'Rosas frescas') | (df_Exportaciones['Aduana_producto'] == 'claveles frescos, corta
```

```
df_Exportaciones["Aduana_producto"].value_counts(#normalize=True)
```

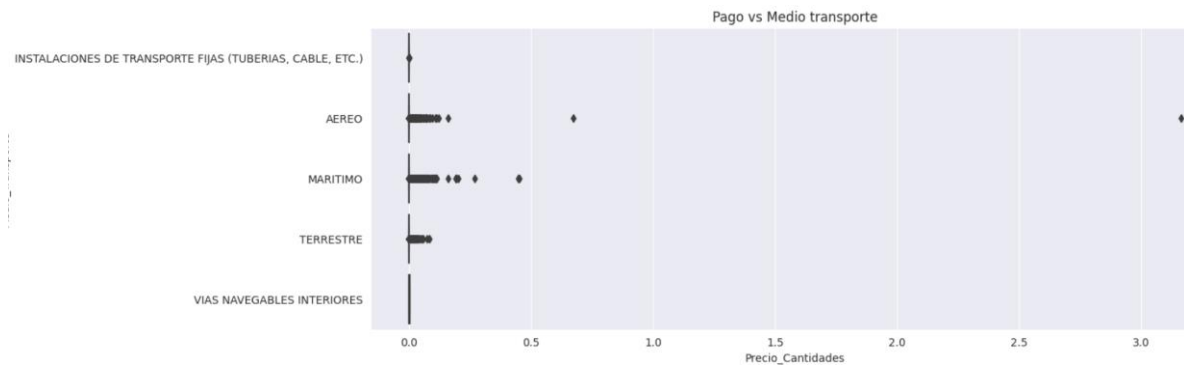
```
Rosas frescas          12580
claveles frescos, cortados para ramos o adornos.  12319
Hortensias             11293
flores y capullos frescos    9289
Name: Aduana_producto, dtype: int64
```

Adicional comprobamos los registros únicos de nuestro DataFrame

```
for column in df_Exportaciones.select_dtypes("object").columns:
    print(column,":",df_Exportaciones[column].unique())
    print("=*80")
```

```
=====
Lugar_salida : ['BOG' 'SMR' 'BUN' 'IPI' 'TRB' 'CTG' 'BAQ' 'MDE']
=====
Unidad_Comercial : ['U']
=====
Ciudad_Aduana : ['Especial de Aduanas de Bogotá' 'Santa Marta' 'Buenaventura' 'Ipiates'
'Urabá' 'Cartagena' 'Barranquilla' 'Medellín']
=====
Origen_Procedencia : ['ANTIOQUIA' 'BOGOTA' 'CUNDINAMARCA' 'VALLE DEL CAUCA' 'ATLANTICO'
'BOYACA' 'RISARALDA' 'CALDAS' 'QUINDIO' 'BOLIVAR' 'GUANIA']
=====
Medio_Transporte : ['AEREO' 'MARITIMO' 'TERRESTRE']
=====
Aduana_producto : ['flores y capullos frescos' 'Hortensias'
'claveles frescos, cortados para ramos o adornos.' 'Rosas frescas']
=====
Nac_Transporte : [nan 'Afganistán' '7 Caimán, Islas' '0 ZFPE Colmotores'
'1 ZFPE Tablemac S' '2 ZFPE Fundación' '3 ZFPE Aceites Ci'
'4 ZFPE Puerto Bri' '5 ZFPE Clínica Ca' 'Bahamas' '2 Svalbard Y Jan'
'0 Surinam' '3 Swazilandia' '6 Tailandia' '4 Tayikistán' 'Alemania'
'5 Dominica' '2 Dinamarca' '9 Ecuador' 'Antigua y Barbu' '1 Libano'
'4 Liberia' '8 Libia' 'Argentina' '1 ZFP Palermo.' '3 ZFP FEMSA'
'4 ZFPE SYKES.' '5 ZFPE Sociedad P' '6 ZFPE Termotasaj'
'7 ZFPE Sociedad P' '8 ZFPE GETCOM.']
=====
```

Para desarrollar el modelo identificamos contra nuestra variable objetivo PRECIO CANTIDADES, como ha sido el canal de transporte más utilizado durante el año 2022, donde identificamos

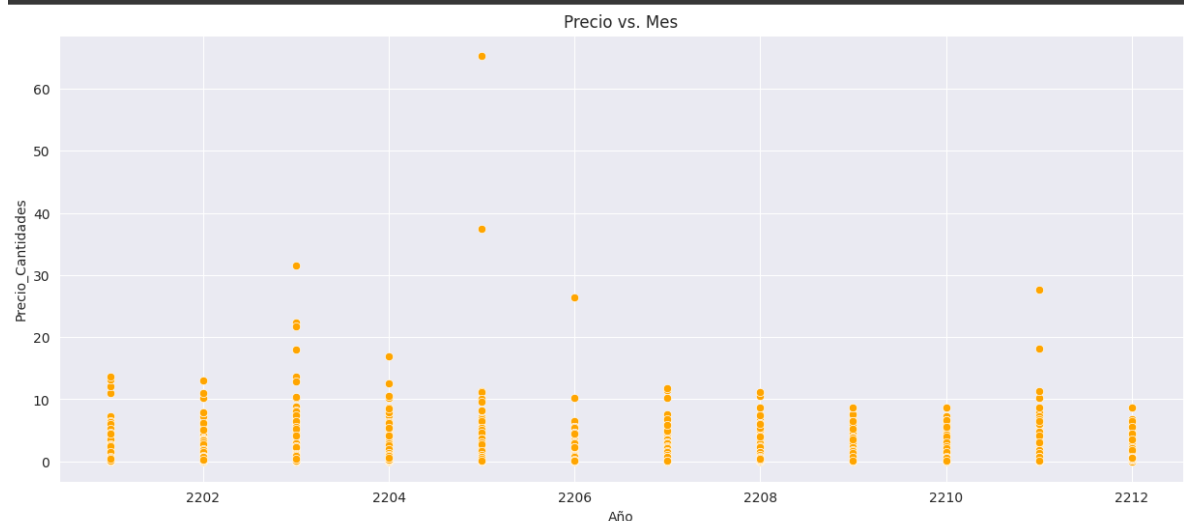


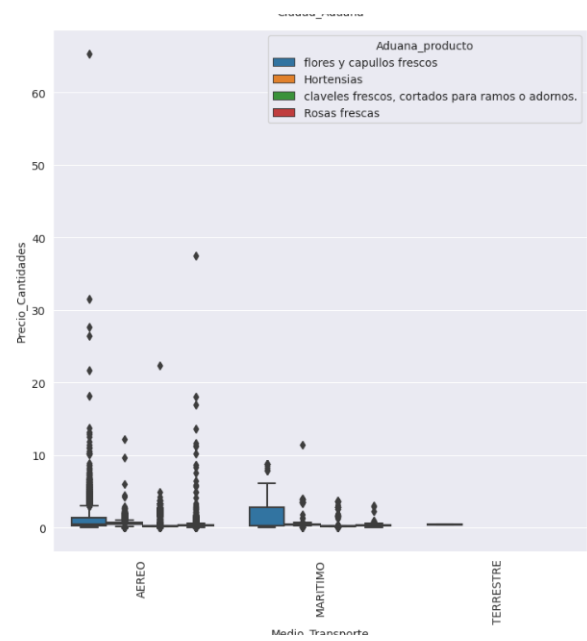
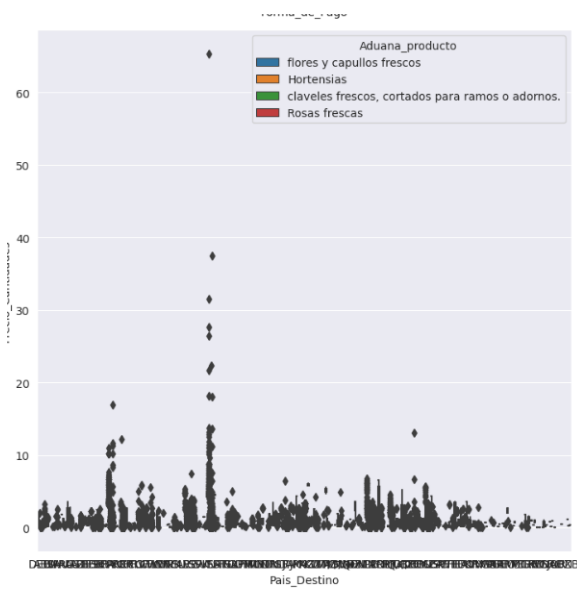
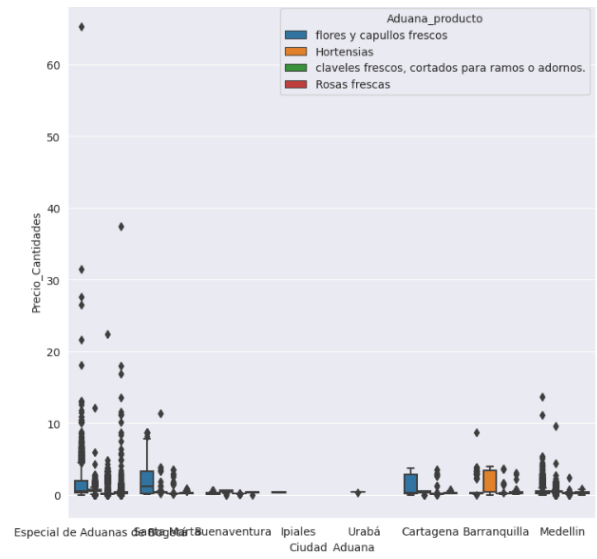
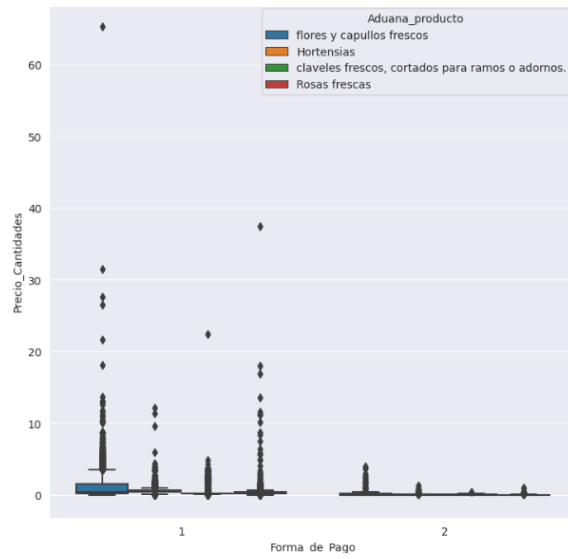
```
plt.figure(figsize=(14,5))
sns.boxplot(y = "Aduana_producto", x = "Precio_Cantidades", data=df_Exportaciones)
plt.title("Pay vs Gender")
plt.show()
```



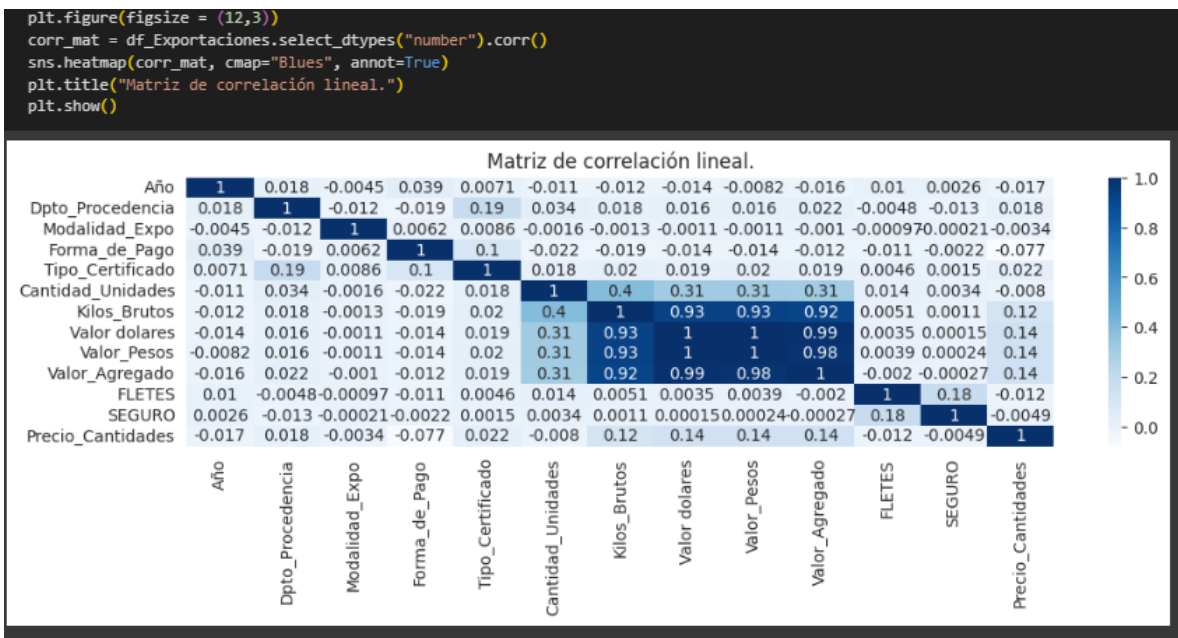
Con el anterior ejercicio podemos identificar que los 5 productos con mayores exportaciones se exportan por canal AEREO , por unidad los precios más altos se presentaron durante el mes de Mayo

```
plt.figure(figsize = (15,6))
sns.scatterplot(data = df_Exportaciones, x = "Año", y = "Precio_Cantidades", color="orange",)
plt.title("Precio vs. Mes")
plt.show()
```





Adicional podemos visualizar que la mayor exportacion se realiza desde Bogotá con pagos anticipados y por canal aereo



Presentamos la matriz de correlación lineal en la cual identificamos que el precio en dólares por cantidad, está asociado al peso en Kilos brutos de la mercancía transportada y es independiente a la forma de pago.

```
## Modelo de regresión lineal simple.
formulal = "Precio_Cantidades ~ Aduana_producto"
modell = smf.ols(formula = formulal, data = df_Exportaciones).fit()
print(modell.summary())
```

```

OLS Regression Results
=====
Dep. Variable:      Precio_Cantidades      R-squared:      0.109
Model:              OLS                    Adj. R-squared: 0.109
Method:              Least Squares         F-statistic:    1859.
Date:                Tue, 06 Jun 2023      Prob (F-statistic): 0.00
Time:                01:26:42              Log-Likelihood: -58284.
No. Observations:    45481                AIC:          1.166e+05
Df Residuals:        45477                BIC:          1.166e+05
Df Model:            3
Covariance Type:     nonrobust
=====
                    coef      std err      t      P>|t|      [0.025      0.975]
-----
Intercept          0.5741      0.008     69.995    0.000      0.558      0.590
Aduana_producto[T.Rosas frescas] -0.2108      0.011    -18.661    0.000     -0.233     -0.189
Aduana_producto[T.claveles frescos, cortados para ramos o adornos.] -0.2991      0.011    -26.336    0.000     -0.321     -0.277
Aduana_producto[T.flores y capullos frescos] 0.5310      0.012     43.491    0.000      0.507      0.555
=====
Omnibus:            93228.233      Durbin-Watson:    1.778
Prob(Omnibus):      0.000          Jarque-Bera (JB): 1316816205.657
Skew:               16.641          Prob(JB):         0.00
Kurtosis:           835.927          Cond. No.         4.82
=====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

```

model_MF = smf.ols("Precio_Cantidades ~ Aduana_producto", data = df_Exportaciones)
fitted_MF = model_MF.fit()
print(fitted_MF.summary())

```

```

=====
                        OLS Regression Results
=====
Dep. Variable:          Precio_Cantidades      R-squared:                0.109
Model:                  OLS                   Adj. R-squared:           0.109
Method:                 Least Squares         F-statistic:             1859.
Date:                   Tue, 06 Jun 2023       Prob (F-statistic):      0.00
Time:                   01:28:10              Log-Likelihood:          -58284.
No. Observations:       45481                AIC:                    1.166e+05
Df Residuals:           45477                BIC:                    1.166e+05
Df Model:               3
Covariance Type:        nonrobust
=====

```

| | coef | std err | t | P> t | [0.025 | 0.975] |
|---|---------|---------|---------|-------|--------|--------|
| Intercept | 0.5741 | 0.008 | 69.995 | 0.000 | 0.558 | 0.590 |
| Aduana_producto[T.Rosas frescas] | -0.2108 | 0.011 | -18.661 | 0.000 | -0.233 | -0.189 |
| Aduana_producto[T.claveles frescos, cortados para ramos o adornos.] | -0.2991 | 0.011 | -26.336 | 0.000 | -0.321 | -0.277 |
| Aduana_producto[T.flores y capullos frescos] | 0.5310 | 0.012 | 43.491 | 0.000 | 0.507 | 0.555 |

```

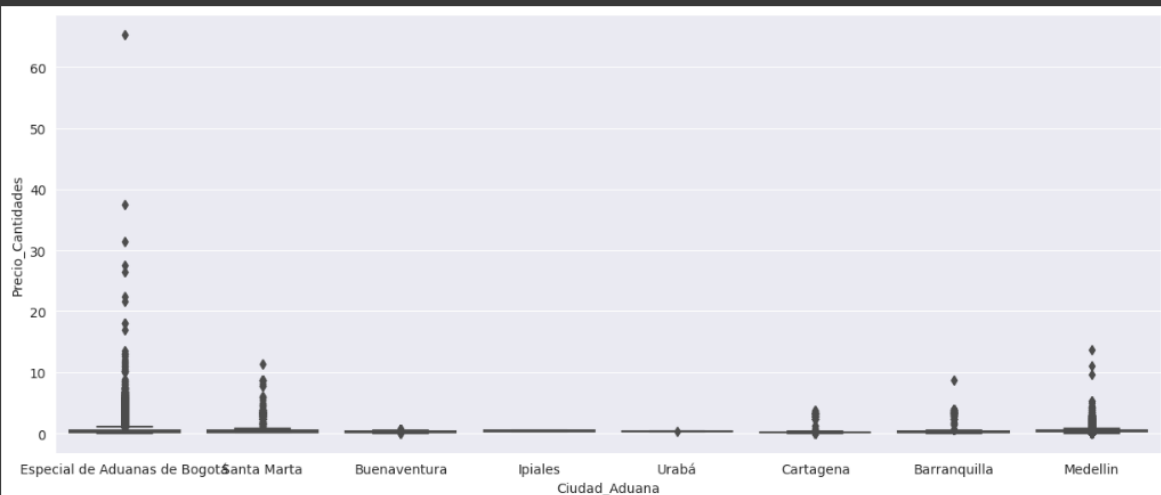
=====
Omnibus:                 93228.233      Durbin-Watson:           1.778
Prob(Omnibus):           0.000          Jarque-Bera (JB):        1316816205.657
Skew:                    16.641          Prob(JB):                0.00
Kurtosis:                 835.927        Cond. No.                 4.82
=====
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```

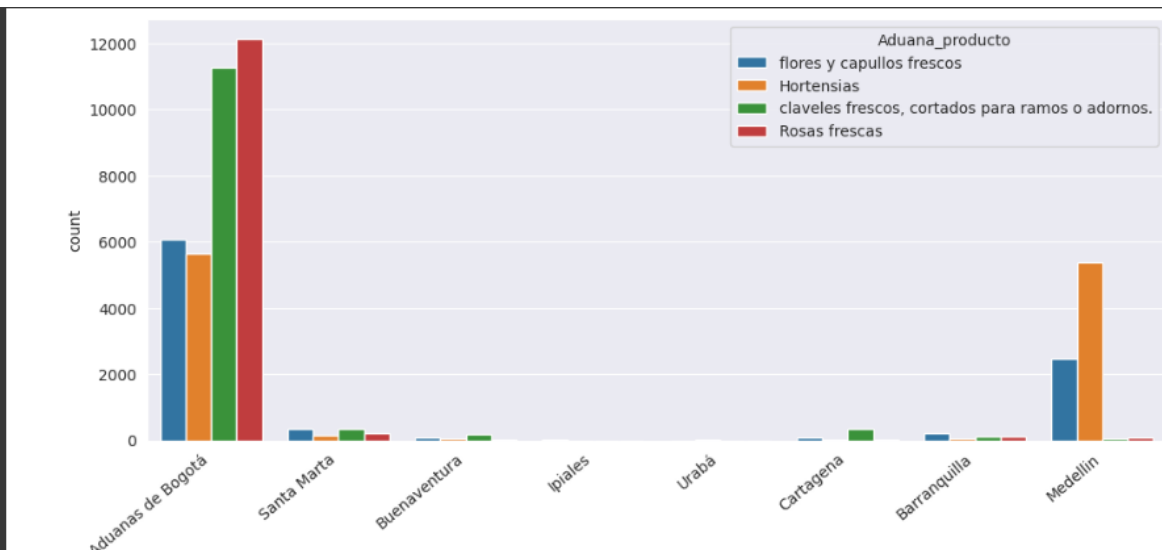
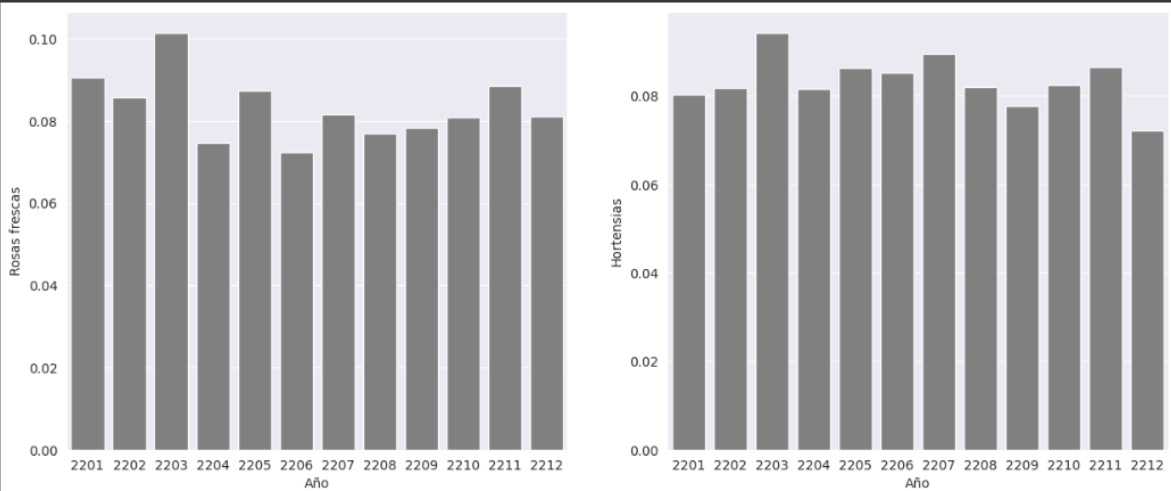
```

sns.boxplot(data = df_Exportaciones, y = "Precio_Cantidades", x = "Ciudad_Aduana", color = "cyan")
plt.show()

```



```
ct = pd.crosstab(df_Exportaciones["Aduana_producto"], df_Exportaciones["Año"], normalize="index")
fig, ax = plt.subplots(1,2, figsize = (15,6))
sns.barplot(x = ct.columns, y = ct.loc["Rosas frescas"], ax = ax[0], color = "grey")
sns.barplot(x = ct.columns, y = ct.loc["Hortensias"], ax = ax[1], color = "grey")
plt.show()
```



Utilizamos las técnicas de regresión lineal para determinar si existía o no diferencias en los precios de exportación de los distintos tipos de flores. Modelamos el efecto de varias variables de entrada (en este caso, Ciudad de Destino, Año, Kilos y Medio de transporte) para explicar la variación observada de una variable de salida (en este caso, Precio). Observamos el coeficiente R^2 de nuestros modelos lineales para ayudarnos a medir qué porcentaje de la variación observada en el Precio se explica por las variables de entrada.

Vimos que la diferencia en el precio medio de los 4 diferentes productos de exportación son estadísticamente significativa. Una exploración más profunda de los datos sugirió que un

gran impulsor de esta diferencia se debía en parte a la fecha que se realizaba la exportación, mostrando picos en los meses de Febrero, Marzo, Mayo, Agosto y Diciembre.

Adicional con el fin de garantizar la calidad del producto el canal AEREO genera un incremento en el precio de la flor

```

Dep. Variable:    Precio_Cantidades    R-squared:        0.147
Model:            OLS                  Adj. R-squared:    0.147
Method:           Least Squares        F-statistic:       560.2
Date:             Tue, 06 Jun 2023      Prob (F-statistic): 0.00
Time:             01:39:49              Log-Likelihood:    -57295.
No. Observations: 45481                AIC:               1.146e+05
Df Residuals:     45466                BIC:               1.148e+05
Df Model:         14
Covariance Type:  nonrobust

=====
                    coef    std err          t      P>|t|      [0.025    0.975]
-----
Intercept                10.7571         2.692         3.996     0.000         5.481    16.034
Medio_Transporte[T.MARITIMO]  0.2321         0.854         0.272     0.786        -1.441     1.906
Medio_Transporte[T.TERRESTRE] -0.2358         0.437        -0.539     0.590        -1.092     0.621
Ciudad_Aduana[T.Buenaventura] -0.2091         0.061        -3.426     0.001        -0.329    -0.089
Ciudad_Aduana[T.Cartagena]    0.1029         0.054         1.908     0.056        -0.003     0.209
Ciudad_Aduana[T.Especial de Aduanas de Bogotá] 0.3610         0.853         0.423     0.672        -1.311     2.033
Ciudad_Aduana[T.Ipiales]      -0.2358         0.437        -0.539     0.590        -1.092     0.621
Ciudad_Aduana[T.Medellin]     0.0041         0.853         0.005     0.996        -1.668     1.676
Ciudad_Aduana[T.Santa Marta]  0.3112         0.045         6.877     0.000         0.222     0.400
Ciudad_Aduana[T.Urabá]        -0.2426         0.272        -0.891     0.373        -0.776     0.291
Aduana_producto[T.Rosas frescas] -0.3884         0.013       -31.041     0.000        -0.413    -0.364
Aduana_producto[T.claveles frescos, cortados para ramos o adornos.] -0.4701         0.013       -37.332     0.000        -0.495    -0.445
Aduana_producto[T.flores y capullos frescos]  0.4528         0.012       36.779     0.000         0.429     0.477
Año                        -0.0045         0.001        -3.859     0.000        -0.007    -0.002
Forma_de_Pago              -0.5306         0.024       -22.258     0.000        -0.577    -0.484
Kilos_Brutos                1.676e-06      7.69e-08      21.790     0.000      1.53e-06     1.83e-06
=====

Omnibus:            95076.381    Durbin-Watson:      1.848
Prob(Omnibus):      0.000      Jarque-Bera (JB):   1529154021.955
Skew:               17.406      Prob(JB):           0.00
Kurtosis:           900.615      Cond. No.           2.74e+20
=====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 1.67e-27. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.

```