

# COMERCIO INTERNACIONAL DE COLOMBIA: EXPORTACIONES E IMPORTACIONES DE COLOMBIA DE 1962 A 2017

Final Data Triathlon | Olimpiadas | Platzi

Diego Lesmes

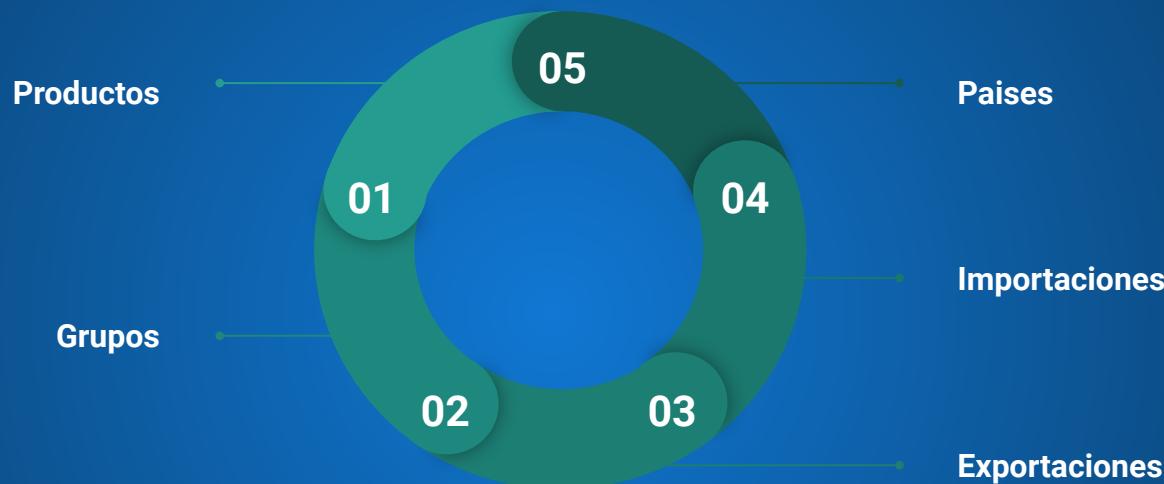
[El Observatorio de la Complejidad Económica](#)



# CONJUNTOS DE DATOS



# CONJUNTO DE INFORMACIÓN Y LIBRERÍAS



# INFORMACIÓN EXTERNA

Ge<sup>O</sup>Py



# DOMADO DE DATOS



# DIAGRAMA ENTIDAD RELACIÓN

## Transacciones

```
imports

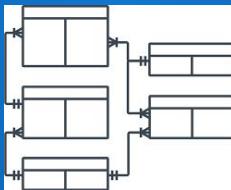
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 775902 entries, 0 to 775901
Data columns (total 7 columns):
 #   Column       Non-Null Count   Dtype  
--- 
 0   Unnamed: 0    775902 non-null    int64  
 1   year         775902 non-null    int64  
 2   origin        775902 non-null    object  
 3   dest          775902 non-null    object  
 4   sitc4         775902 non-null    int64  
 5   export_val    731962 non-null    float64 
 6   import_val    719568 non-null    float64 
dtypes: float64(2), int64(3), object(2)
memory usage: 41.4+ MB
```

```
exports

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 799406 entries, 0 to 799405
Data columns (total 7 columns):
 #   Column       Non-Null Count   Dtype  
--- 
 0   Unnamed: 0    799406 non-null    int64  
 1   year         799406 non-null    int64  
 2   origin        799406 non-null    object  
 3   dest          799406 non-null    object  
 4   sitc4         799406 non-null    int64  
 5   export_val    715852 non-null    float64 
 6   import_val    756158 non-null    float64 
dtypes: float64(2), int64(3), object(2)
memory usage: 42.7+ MB
```

## Product

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 988 entries, 0 to 987
Data columns (total 4 columns):
 #   Column       Non-Null Count   Dtype  
--- 
 0   Unnamed: 0    988 non-null    int64  
 1   id           988 non-null    object  
 2   sitc          988 non-null    int64  
 3   name          988 non-null    object  
dtypes: int64(2), object(2)
memory usage: 31.0+ KB
```



## countrys

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 263 entries, 0 to 262
Data columns (total 4 columns):
 #   Column       Non-Null Count   Dtype  
--- 
 0   Unnamed: 0    263 non-null    int64  
 1   id           263 non-null    object  
 2   id_3char      263 non-null    object  
 3   name          263 non-null    object  
dtypes: int64(1), object(3)
memory usage: 8.3+ KB
```

## groups

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 65 entries, 0 to 64
Data columns (total 3 columns):
 #   Column       Non-Null Count   Dtype  
--- 
 0   Unnamed: 0    65 non-null    int64  
 1   id           65 non-null    int64  
 2   category     65 non-null    object  
dtypes: int64(2), object(1)
memory usage: 1.6+ KB
```



# HITO #1

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1575308 entries, 0 to 1575307
Data columns (total 14 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   year        1575308 non-null   int64  
 1   origin       1575308 non-null   object  
 2   dest         1575308 non-null   object  
 3   sitc         1575308 non-null   int64  
 4   export_val   1447814 non-null   float64 
 5   import_val   1475726 non-null   float64 
 6   Transaction  1575308 non-null   object  
 7   name_origin  1574258 non-null   object  
 8   name_dest    1571389 non-null   object  
 9   name_product 1575308 non-null   object  
 10  lat_origin   1572816 non-null   float64 
 11  lon_origin   1572816 non-null   float64 
 12  lat_dest     1572781 non-null   float64 
 13  lon_dest     1572781 non-null   float64 
dtypes: float64(6), int64(2), object(6)
memory usage: 180.3+ MB
```



# VALORES FALTANTES

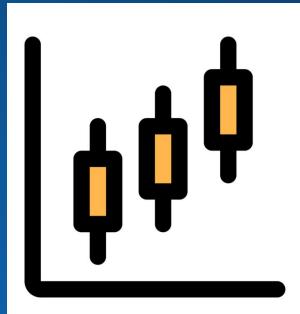


A treemap visualization illustrating the distribution of transaction rows across various variables. The vertical axis on the left represents the number of rows, ranging from 0 at the top to 1531565 at the bottom. The horizontal axis represents different variables: year, origin, dest, sitc, export\_val, import\_val, Transaction, name\_origin, name\_dest, name\_product, lat\_origin, lon\_origin, lat\_dest, and lon\_dest. The size of each rectangular cell corresponds to the count of rows for that specific variable combination. The largest categories are 'name\_origin' and 'name\_dest', which together account for the majority of the rows.

# ESTADÍSTICAS



# ESTADÍSTICAS



## Numerical data

	count	mean	std	min	25%	50%	75%	max
year	1575308.0	1.997857e+03	1.417671e+01	1962.000000	1989.000000	2001.000000	2009.000000	2.017000e+03
sitc	1575308.0	6.222229e+03	2.339567e+03	10.000000	5514.000000	6760.000000	7723.000000	9.710000e+03
export_val	1447814.0	1.074322e+06	3.221065e+07	0.000000	0.000000	3308.000000	73794.750000	1.385914e+10
import_val	1475726.0	1.106083e+06	3.493362e+07	0.000000	0.000000	3344.000000	73000.000000	1.580013e+10
lat_origin	1572816.0	1.435434e+01	2.157828e+01	-41.500083	2.889443	2.889443	32.647531	7.761923e+01
lon_origin	1572816.0	-4.255188e+01	5.870840e+01	-176.204224	-73.783892	-73.783892	-3.276575	1.791583e+02
lat_dest	1572781.0	1.470021e+01	2.164618e+01	-41.500083	2.889443	2.889443	35.000074	7.761923e+01
lon_dest	1572781.0	-4.116957e+01	5.965435e+01	-176.204224	-73.783892	-73.783892	1.888334	1.791583e+02



# ESTADÍSTICAS

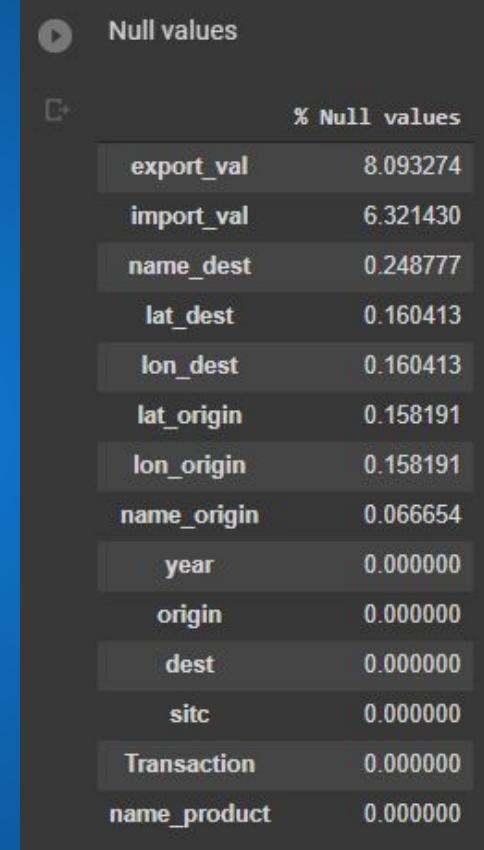
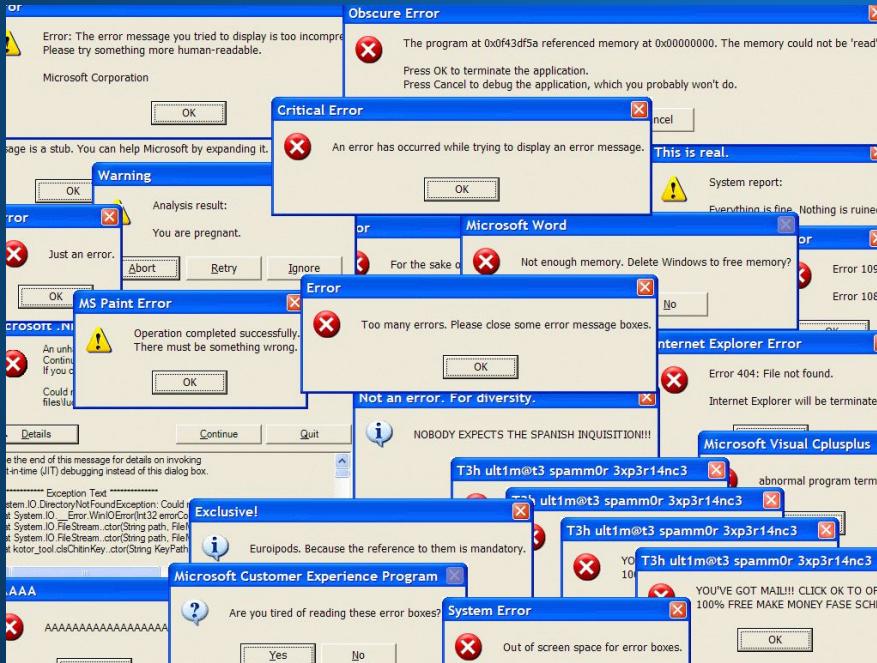


Categorical data

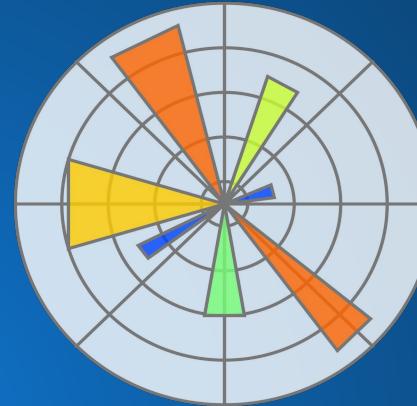


	count	unique	top	freq
origin	1575308	245	col	801096
dest	1575308	250	col	777592
Transaction	1575308	2	export	799406
name_origin	1574258	241	Colombia	801096
name_dest	1571389	243	Colombia	777592
name_product	1575308	936	Coffee	6433

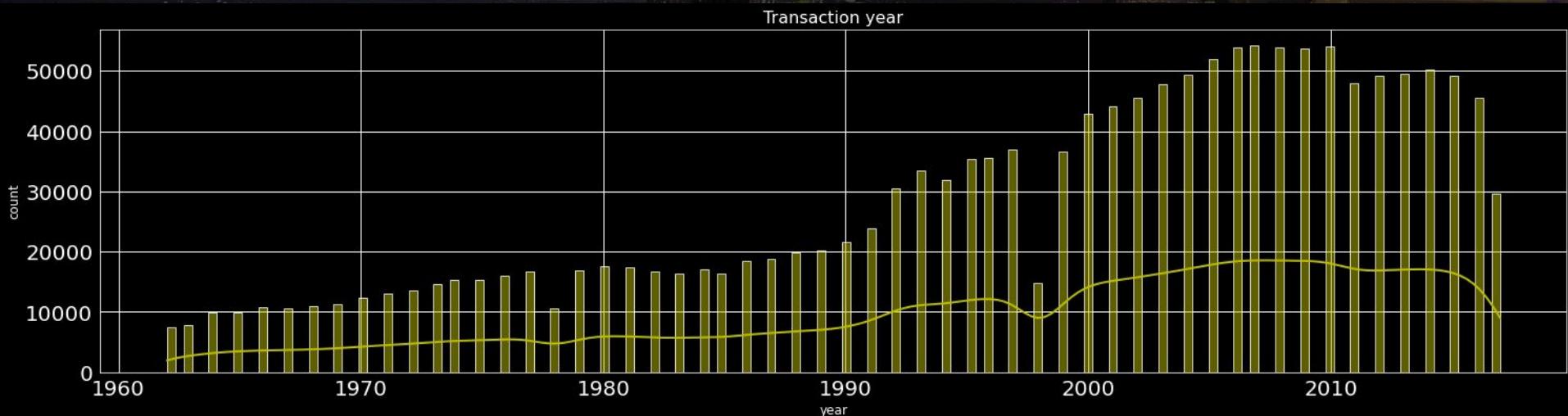
# DATOS FALTANTES



# VISUALIZACIONES

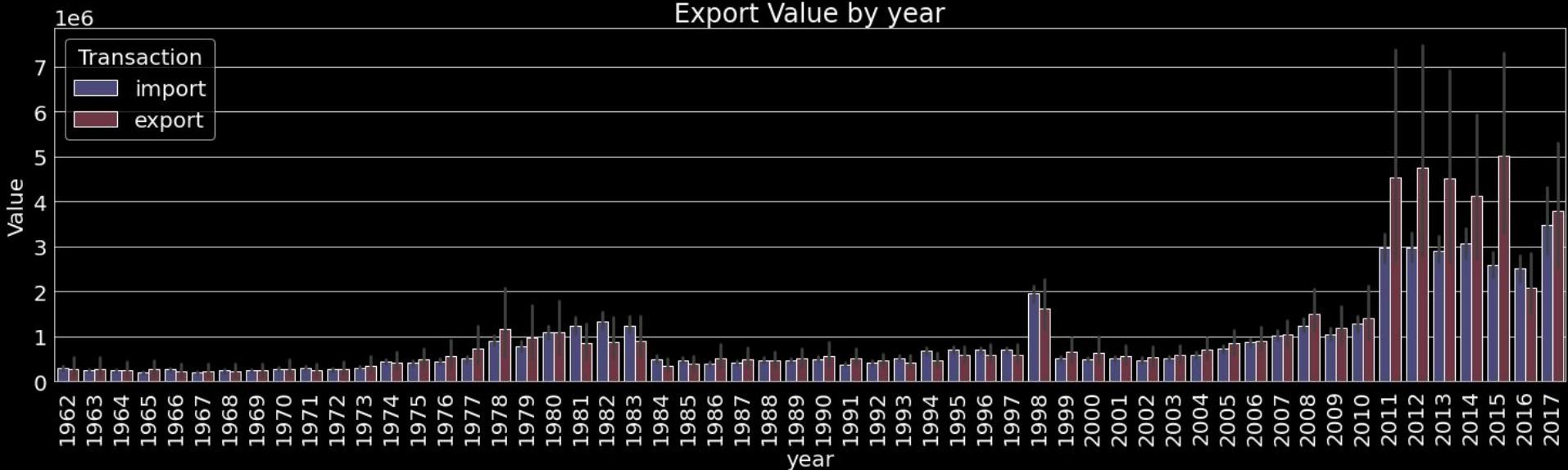


# REGISTROS POR AÑO

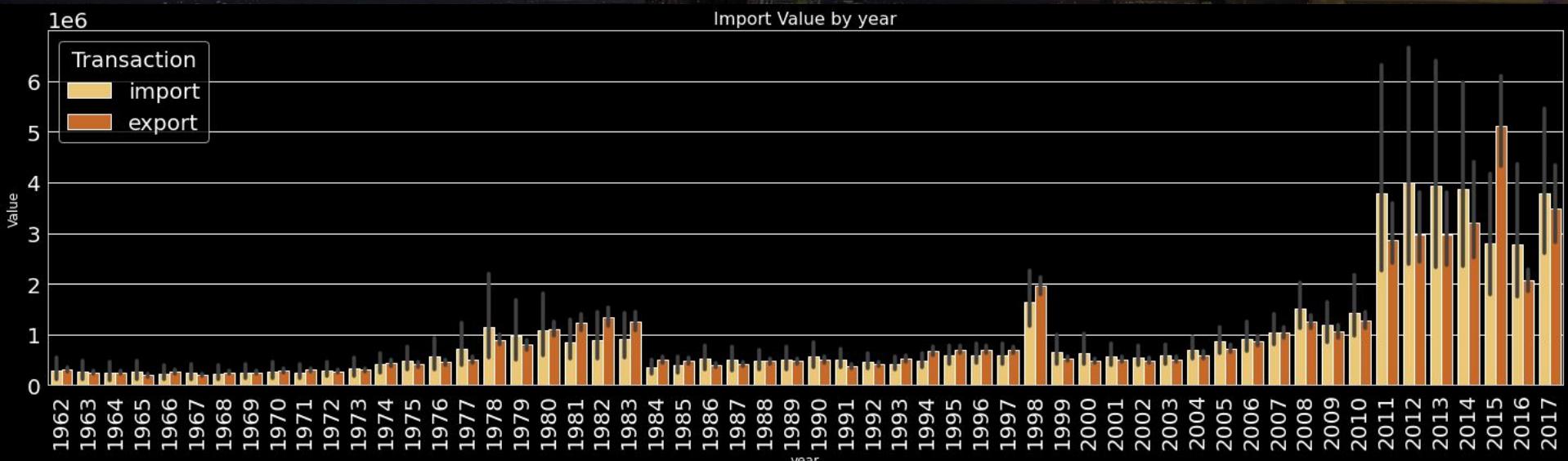


# VALOR EXPORTACIONES

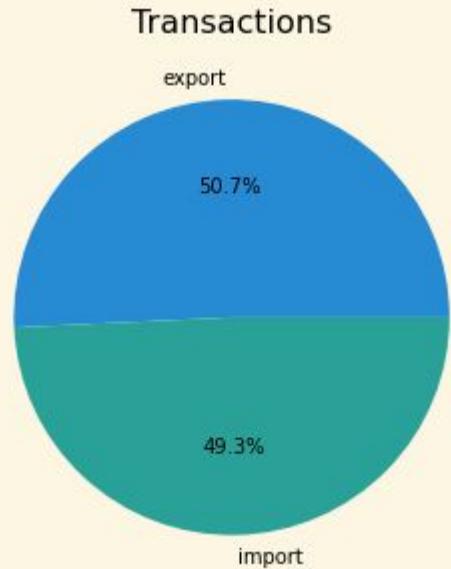
Export Value by year



# VALOR IMPORTACIONES

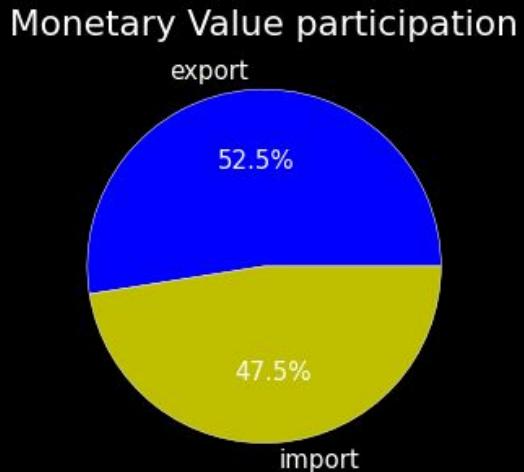


# REGISTROS



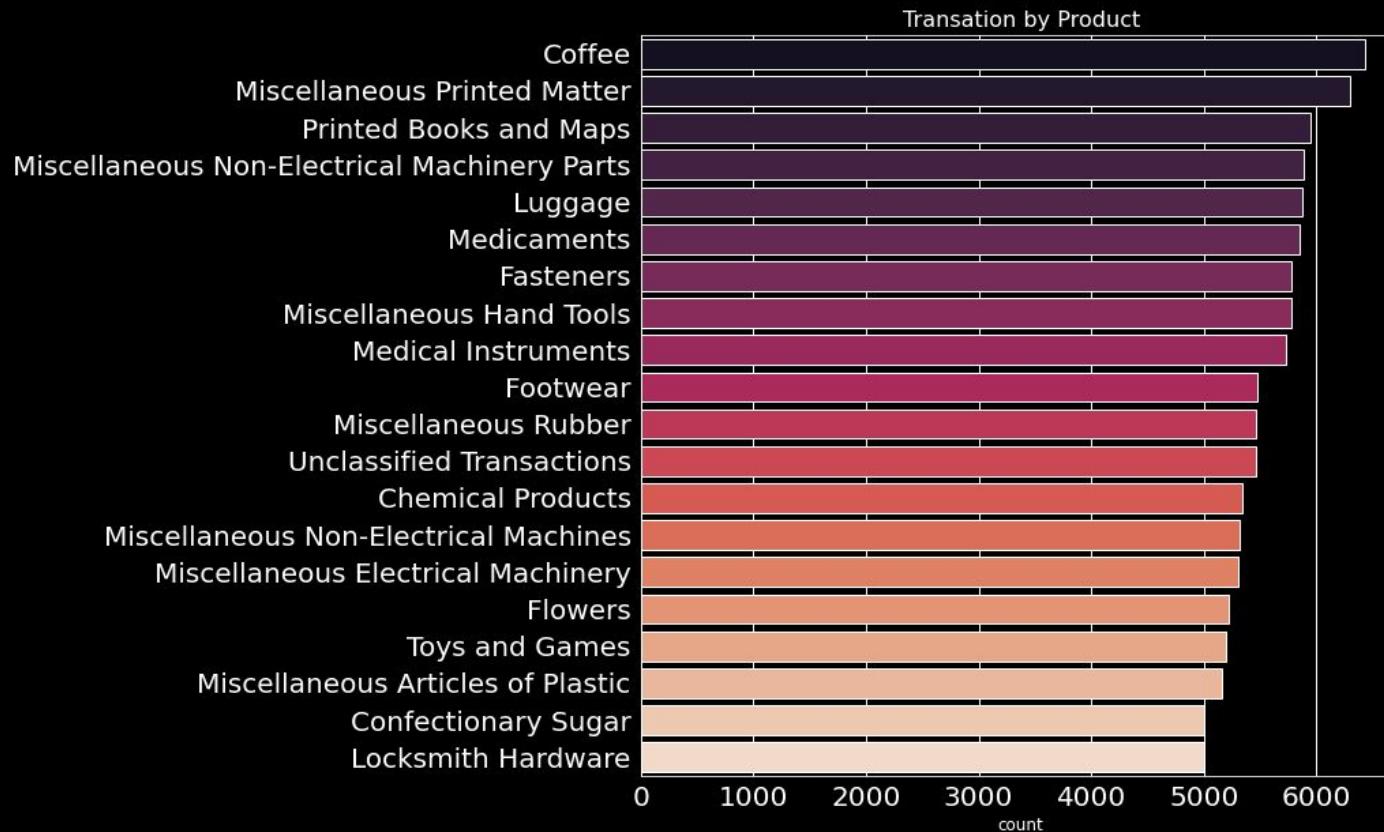
THERE ARE 0.7% MORE RECORDS FROM EXPORT DATA

# DOLARES

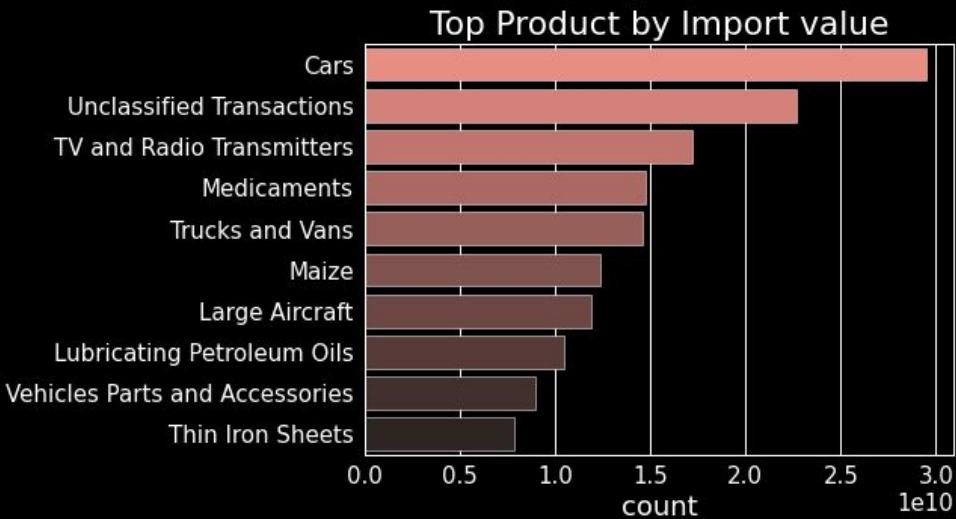
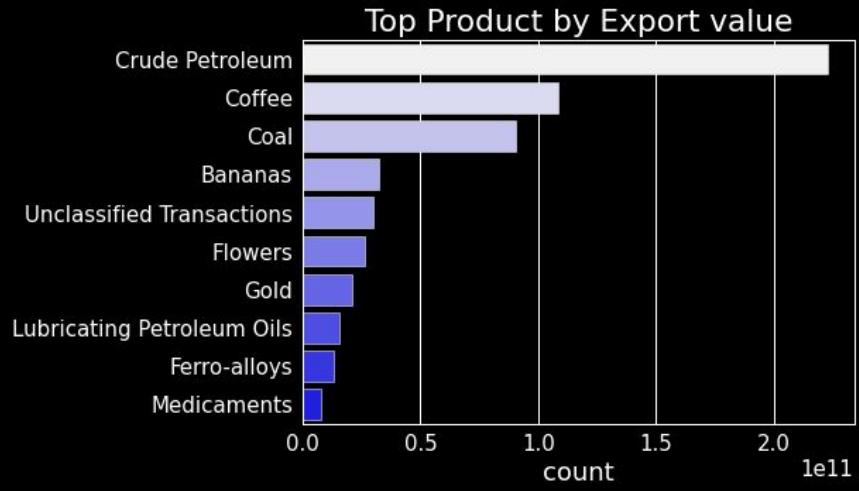


AS WE CAN SEE HISTORICALLY COLOMBIA HAVE MORE EXPORTS THAN IMPORTS, AND THAT IS A GOOD BALANCE SIGNAL

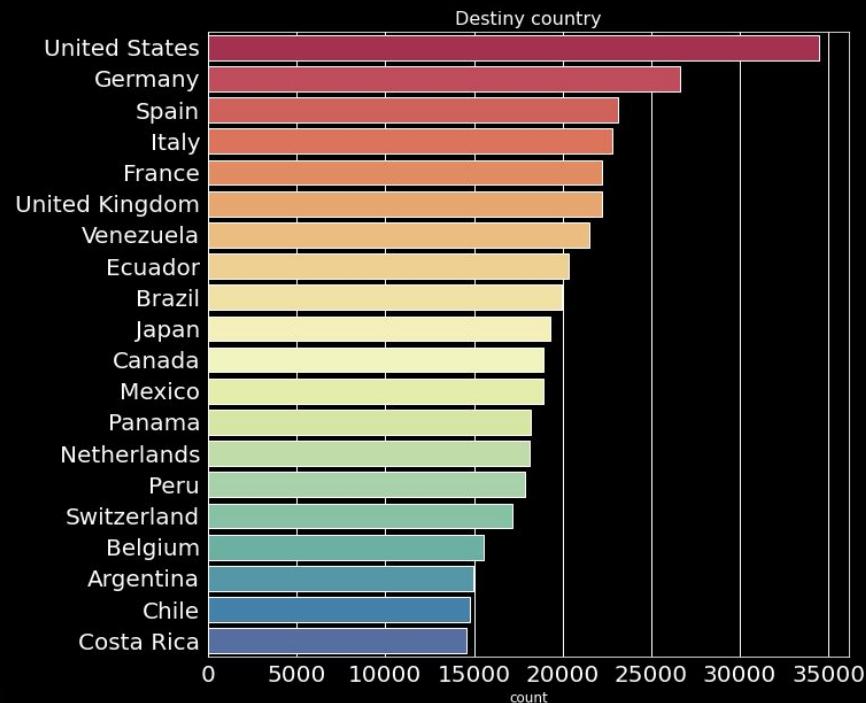
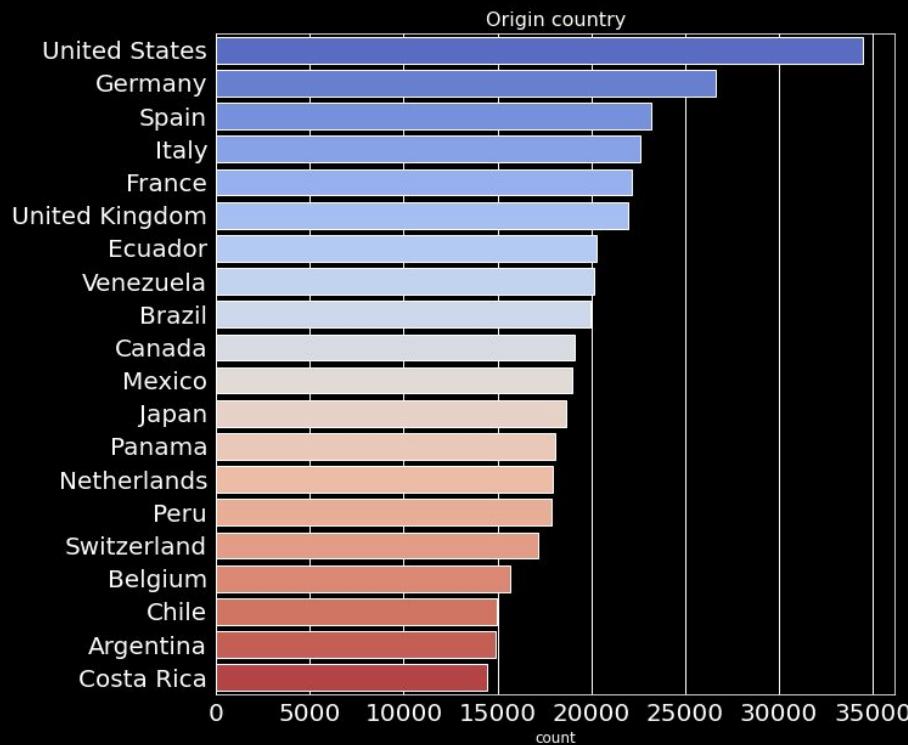
# #REGISTROS POR PRODUCTO



# #VALOR POR PRODUCTO

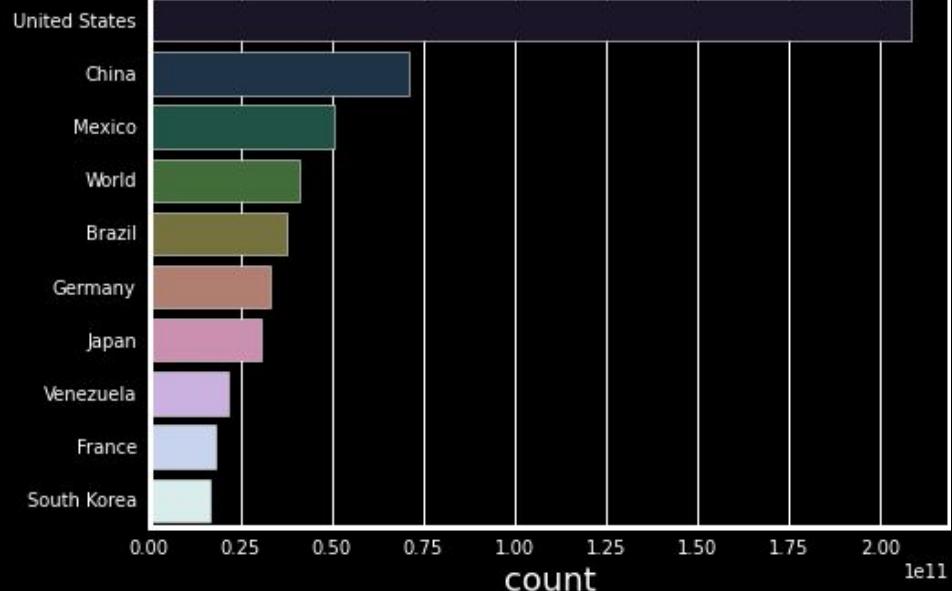


# #REGISTROS ORIGEN/DESTINO

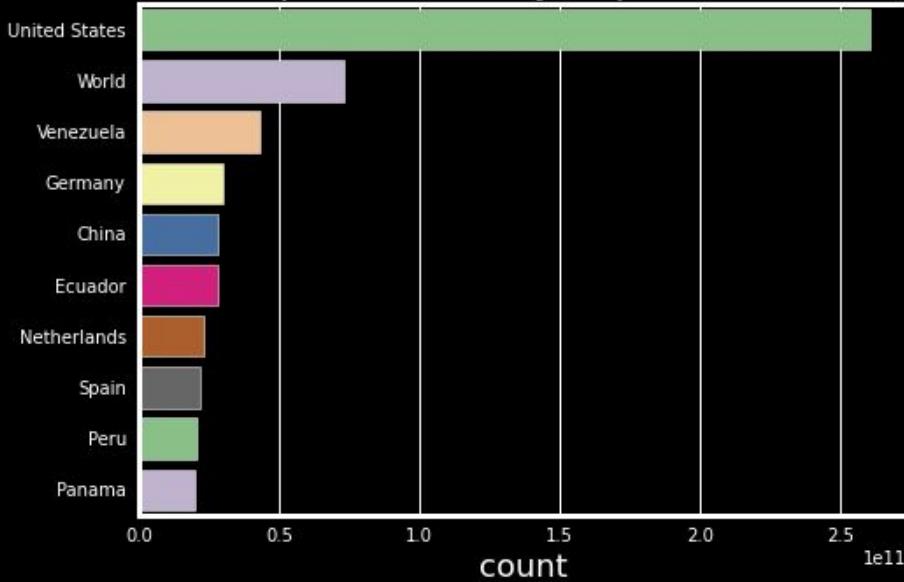


# #VALOR IMPORT/EXPORT

Top Countries by Import value

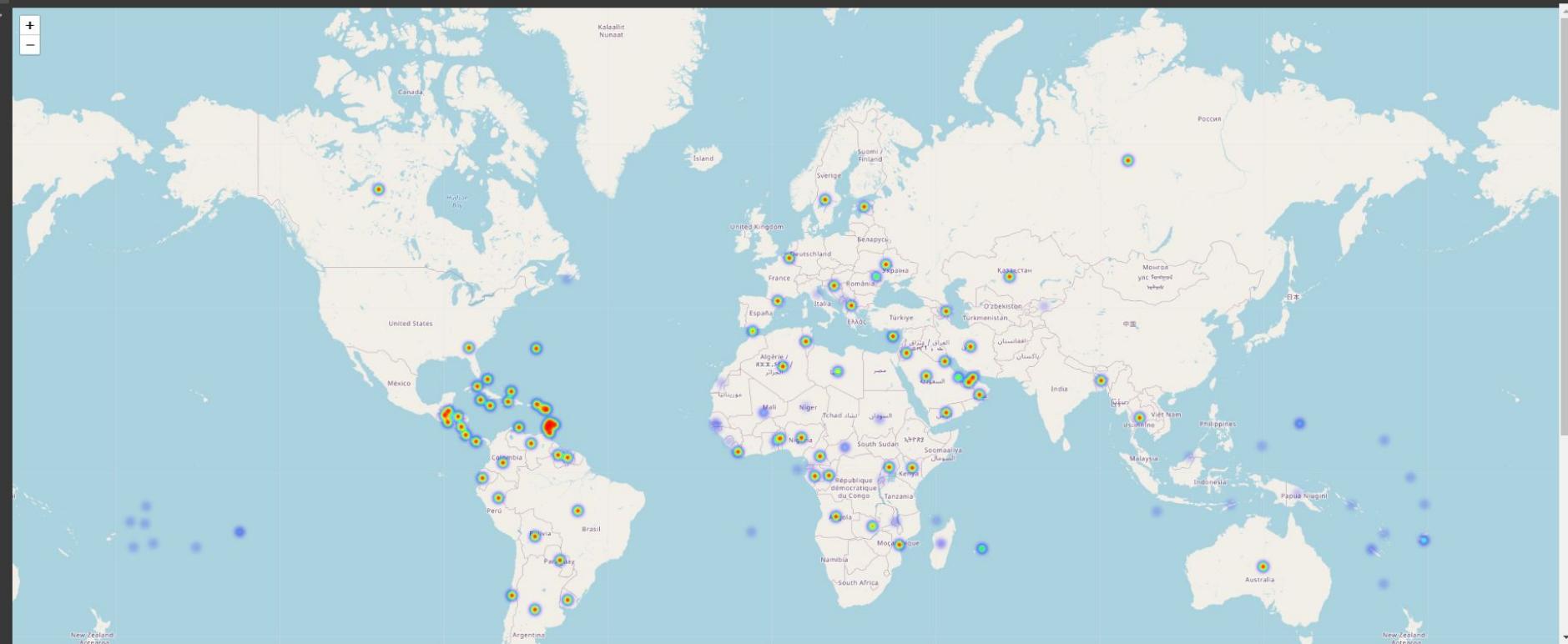


Top Countries by Export value



# DESTINO DE LAS EXPORTACIONES

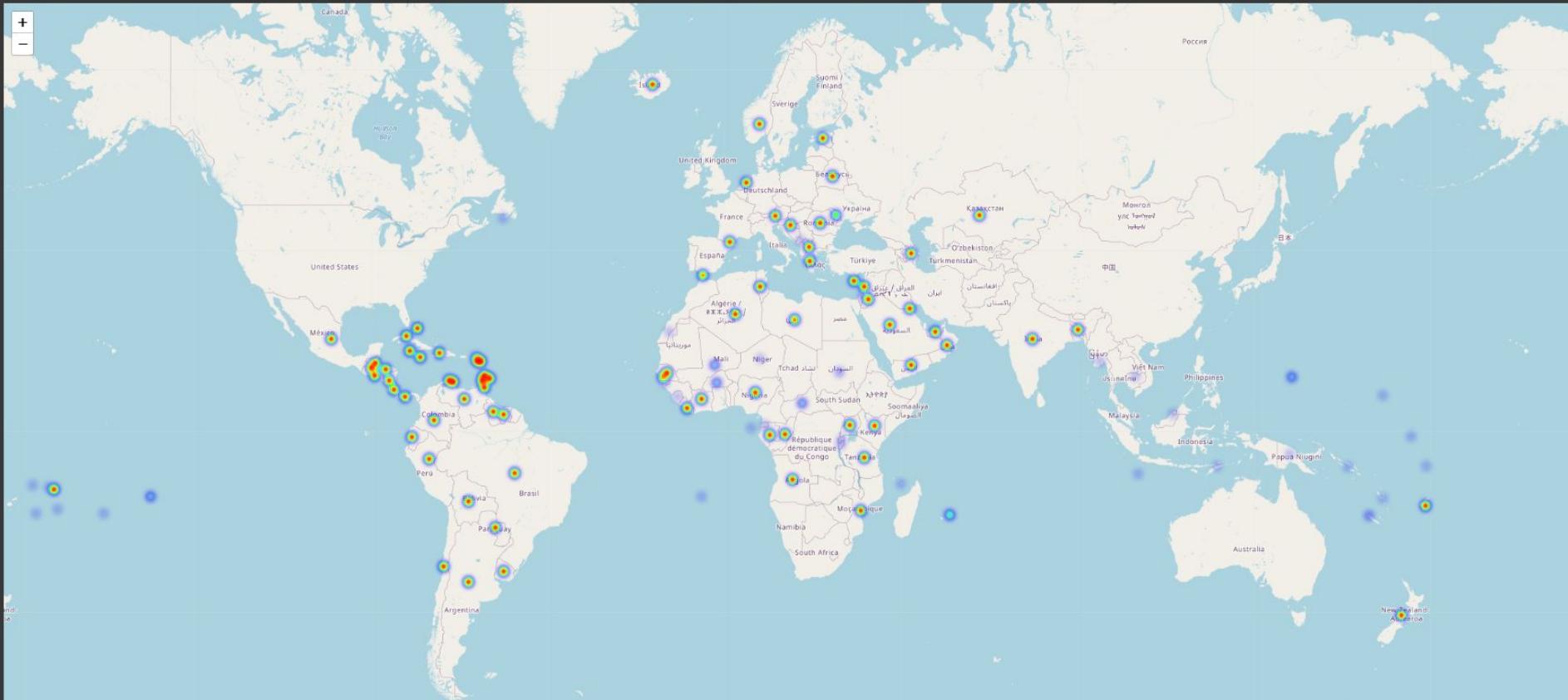
### • 21st Century Exports Destiny



*We can see where the exports goes all around the world, as more red the point higher the value of the transactions*

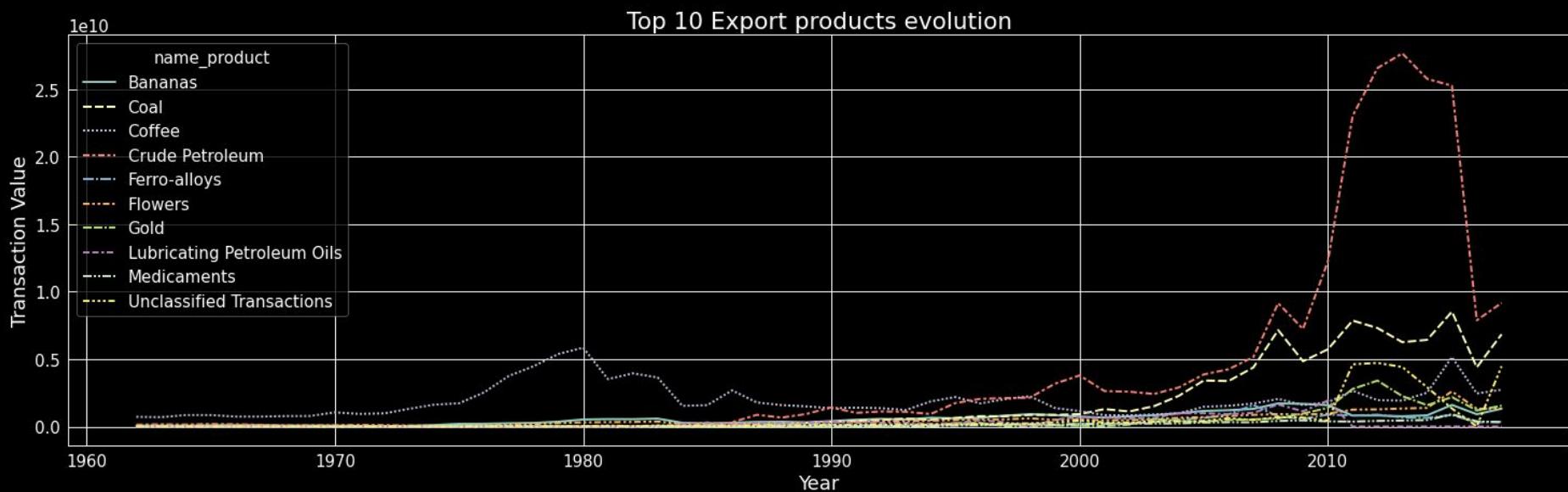
# ORIGEN DE LAS IMPORTACIONES

• 21st Century Import Origins

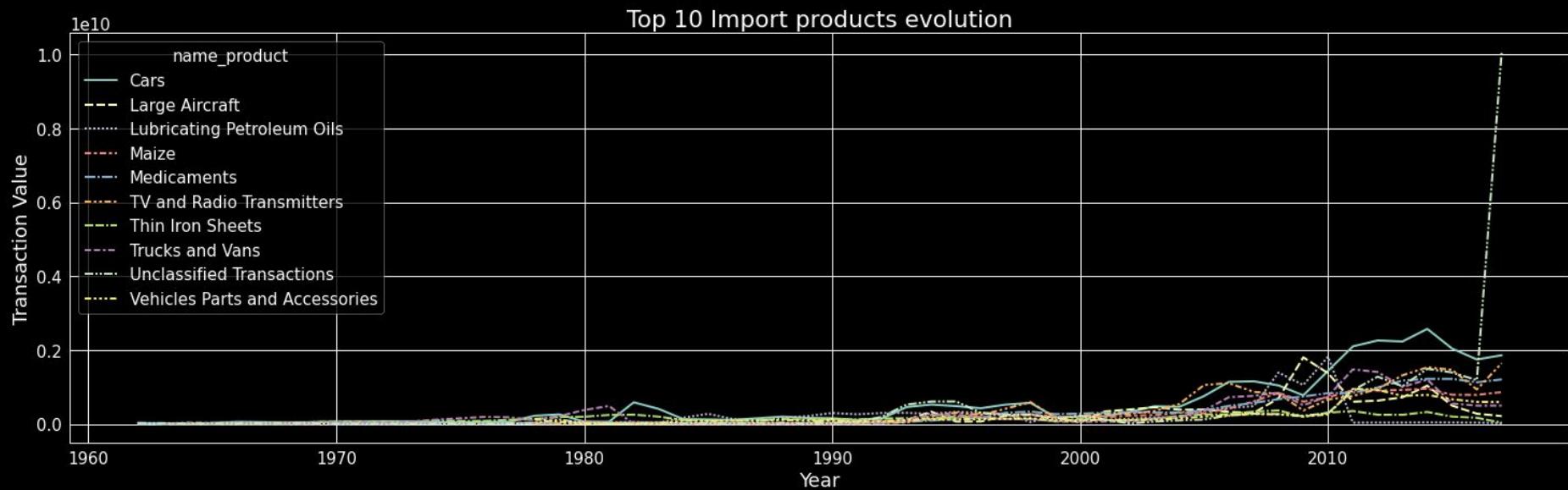


We can see where the imports comes from all around the world, as more red the point higher the value of the transactions

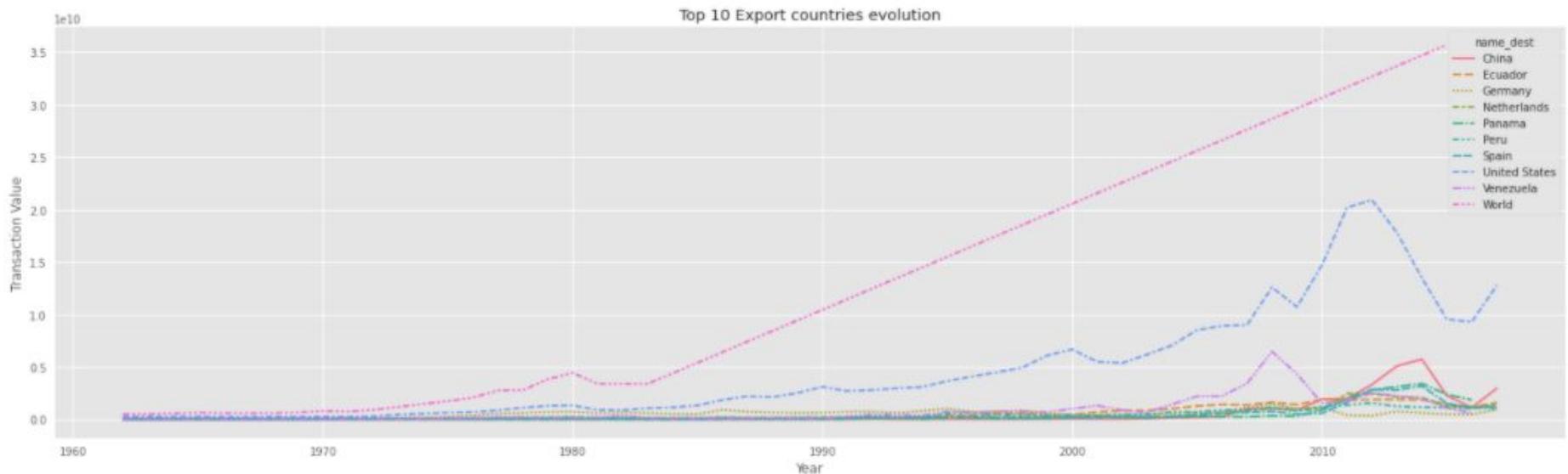
# HISTORIA VALOR DE LOS PRODUCTOS EXPORTADOS



# HISTORIA VALOR DE LOS PRODUCTOS IMPORTADOS

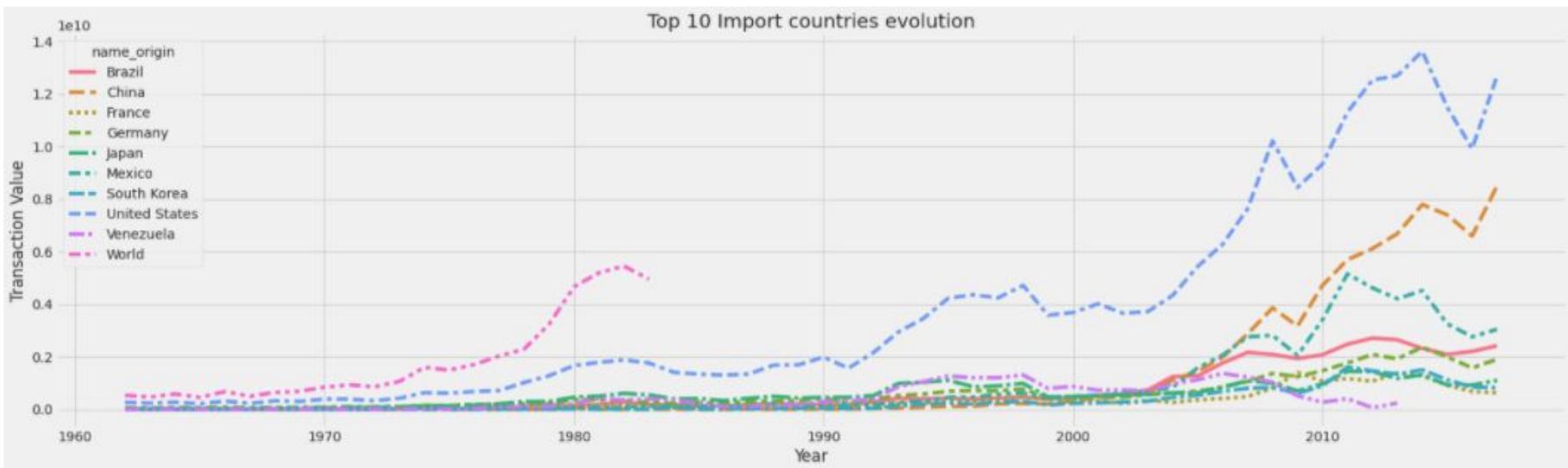


# HISTORIA DESTINO EXPORTACIONES



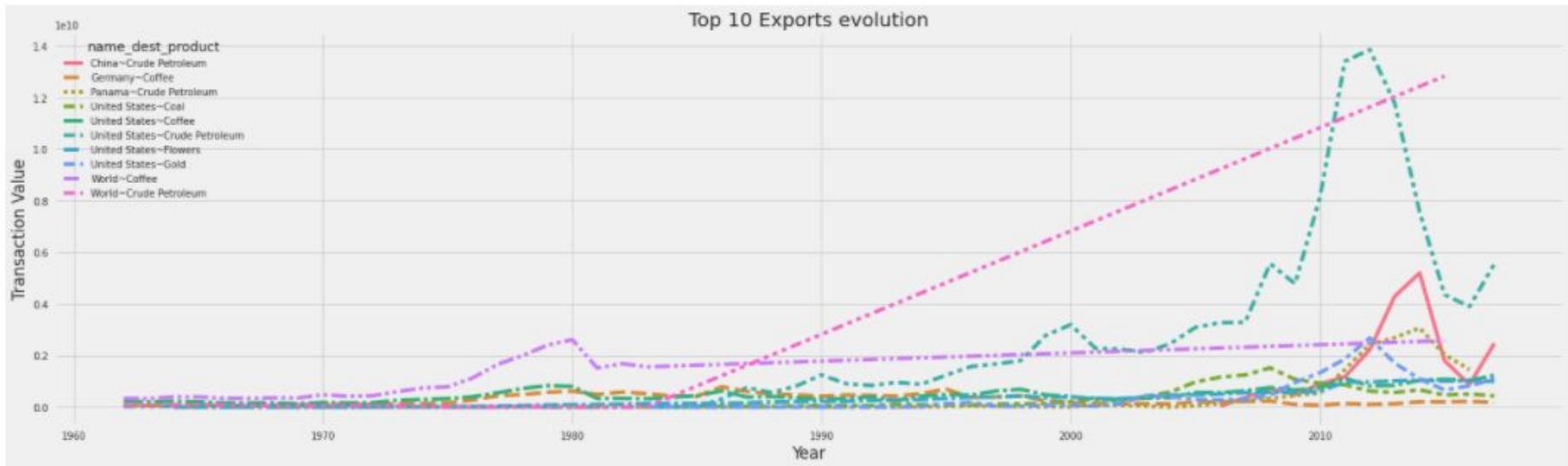
We can see how since the eighties the export to the World has been rising like linearly, followed by USA and China in the last 10 years

# HISTORIA ORIGEN IMPORTACIONES



Here we can see that the records of World imports finished in 1983 and how USA is the best exporter for Colombia and in the last ten years China and Mexico have been leading the import values to Colombia

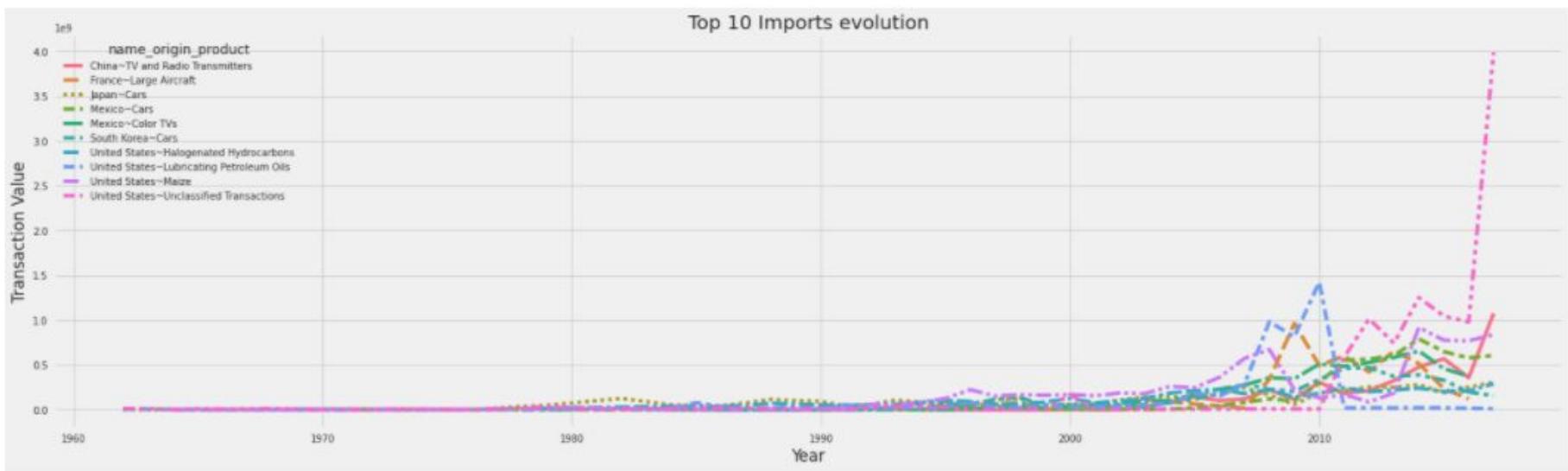
# HISTORIA DESTINO DE PRODUCTOS EXPORTADOS



In this graphic we can see exactly historically which products we sell to which countries with the highest value transactions, the Crude Petroleum to USA and the world, and coffee to the world

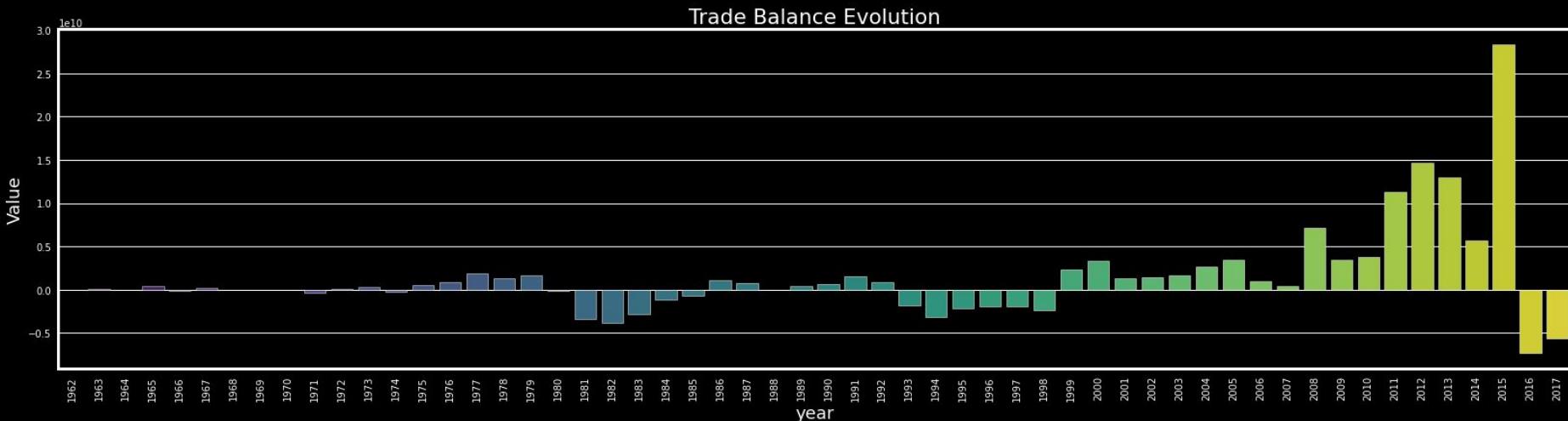
In the last ten years we have Crude Petroleum sold to China and Panama and Gold to USA

# HISTORIA ORIGEN DE PRODUCTOS IMPORTADOS

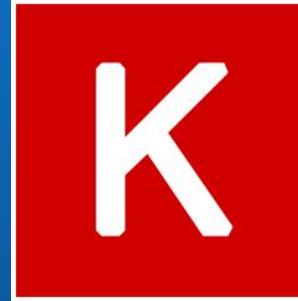


*The Imports have been so different along the history, we have bought Lubricating Petroleum Oils to United States, Large Aircraft to France, Cars to Mexico and TV and Radio Transmitters to China*

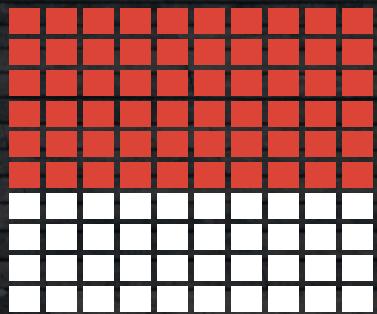
# HISTORIA BALANZA COMERCIAL



# MODELOS



# CONJUNTO DE ENTRENAMIENTO

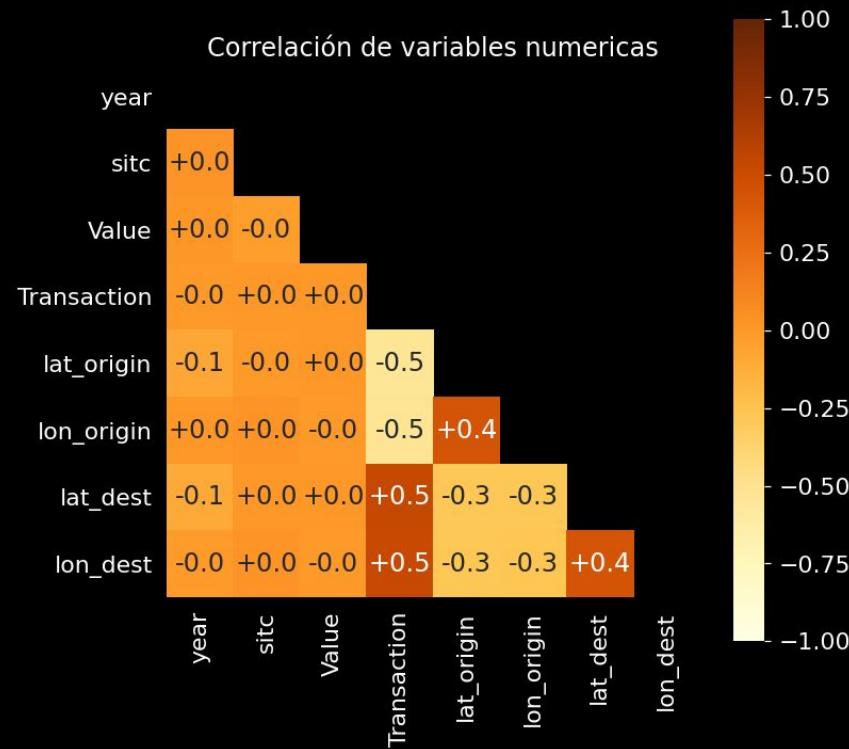


▶ Numerical data set for training

	year	sitc	value	transaction	lat_origin	lon_origin	lat_dest	lon_dest
775902	1962	712	0.0	1	2.889443	-73.783892	7.989737	-5.567946
775903	1962	1210	12000.0	1	2.889443	-73.783892	7.989737	-5.567946
775904	1962	2925	0.0	1	2.889443	-73.783892	-2.981434	23.822264
775905	1962	2925	0.0	1	2.889443	-73.783892	-0.726433	15.641915
775906	1962	1210	112000.0	1	2.889443	-73.783892	28.000027	2.999983
...	...	...	...	...	...	...	...	...
775896	2017	6635	26526.5	0	-32.875555	-56.020153	2.889443	-73.783892
775898	2017	8924	4285.5	0	-32.875555	-56.020153	2.889443	-73.783892
775899	2017	8928	25667.5	0	-32.875555	-56.020153	2.889443	-73.783892
775900	2017	8972	5306.0	0	-32.875555	-56.020153	2.889443	-73.783892
775901	2017	9310	10063053.0	0	-32.875555	-56.020153	2.889443	-73.783892

1430423 rows × 8 columns

# CORRELACIONES



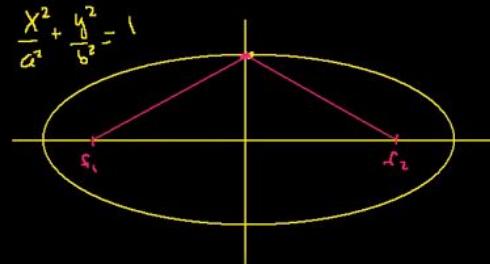
# REGRESIÓN LINEAL

OLS Regression Results						
Dep. Variable:	Value	R-squared:	0.001			
Model:	OLS	Adj. R-squared:	0.001			
Method:	Least Squares	F-statistic:	26.96			
Date:	Tue, 09 Mar 2021	Prob (F-statistic):	2.89e-37			
Time:	04:52:01	Log-Likelihood:	-5.5083e+06			
No. Observations:	286084	AIC:	1.102e+07			
Df Residuals:	286076	BIC:	1.102e+07			
Df Model:	7					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	-1.115e+08	1.52e+07	-7.346	0.000	-1.41e+08	-8.18e+07
year	5.747e+04	7596.371	7.566	0.000	4.26e+04	7.24e+04
sitc	-479.6545	44.206	-10.850	0.000	-566.298	-393.011
Transaction	1.139e+05	3.13e+05	0.364	0.716	-4.99e+05	7.27e+05
lat_origin	1.366e+04	5843.622	2.338	0.019	2208.899	2.51e+04
lon_origin	-5404.6829	2153.004	-2.510	0.012	-9624.512	-1184.854
lat_dest	1.694e+04	5920.842	2.861	0.004	5332.728	2.85e+04
lon_dest	-4495.8773	2163.549	-2.078	0.038	-8736.374	-255.381
Omnibus:	1398897.358	Durbin-Watson:	2.000			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	27702796069333.508			
Skew:	199.863	Prob(JB):	0.00			
Kurtosis:	48209.512	Cond. No.	1.01e+06			

Warnings:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 1.01e+06. This might indicate that there are strong multicollinearity or other numerical problems.

# ANÁLISIS COMPONENTES PRINCIPALES



	Variable	Autovalor	Participacion	Acumulado
0	year	2.846423e+21	1.000000e+02	100.0
1	sitc	7.943518e+12	2.790702e-07	100.0
2	Value	6.403370e+09	2.249620e-10	100.0
3	Transaction	3.520487e+09	1.236811e-10	100.0
4	lat_dest	5.603693e+08	1.968679e-11	100.0
5	lon_origin	4.639822e+08	1.630053e-11	100.0
6	lat_origin	2.598734e+08	9.129821e-12	100.0
7	lon_dest	1.583099e+05	5.561714e-15	100.0

# RED NEURONAL

## Setting Min Max Scaling

	count	mean	std	min	25%	50%	75%	max
0	1430423.0	0.63	0.25	0.0	0.47	0.69	0.82	1.0
1	1430423.0	0.64	0.24	0.0	0.57	0.69	0.80	1.0
2	1430423.0	0.00	0.00	0.0	0.00	0.00	0.00	1.0
3	1430423.0	0.47	0.18	0.0	0.37	0.37	0.60	1.0
4	1430423.0	0.37	0.16	0.0	0.29	0.29	0.49	1.0

```
#@title Setting variables
best_variables = ['year',
                  'sitc',
                  'Value',
                  'lat_dest',
                  'lon_origin'
                 ]
df=colombian_Transactions_numerical[best_variables]
```

## Model

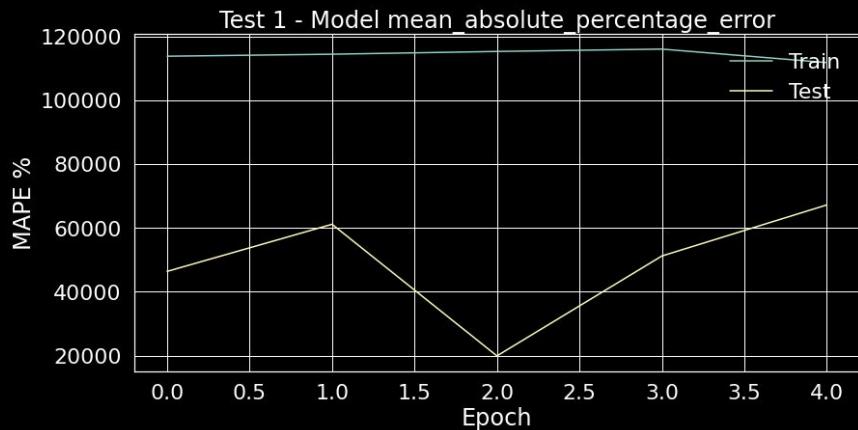
### Model: "sequential"

Layer (type)	output shape	Param #
dense (Dense)	(None, 8)	40
dense_1 (Dense)	(None, 4)	36
dense_2 (Dense)	(None, 2)	10
dense_3 (Dense)	(None, 1)	3

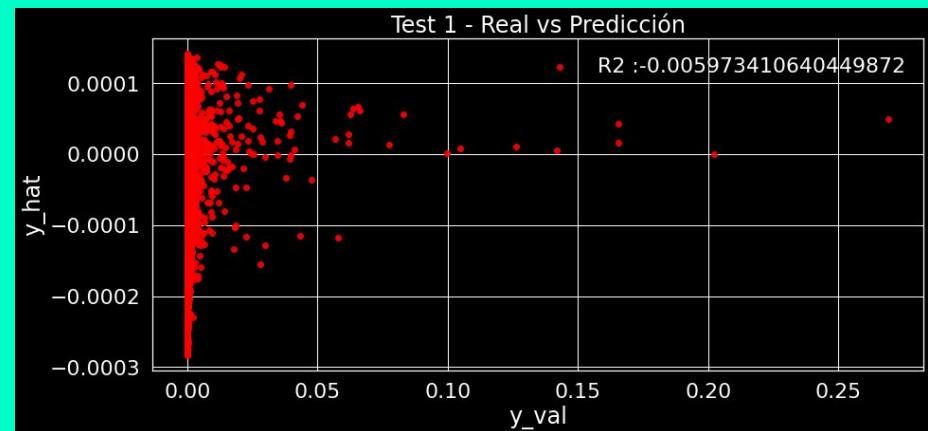
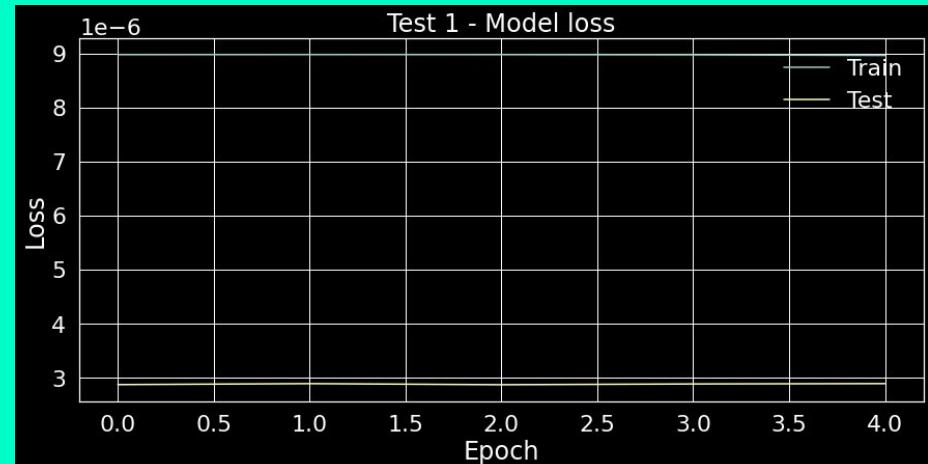
Total params: 89  
Trainable params: 89  
Non-trainable params: 0

None

# RED NEURONAL - 1ER ENTRENAMIENTO

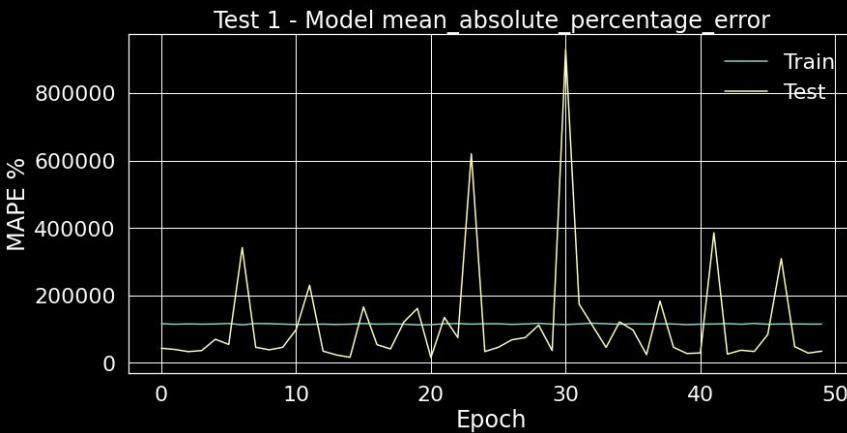


```
8941/8941 [=====] - 8s 845us/step - loss:  
7.0024e-06 - mean_absolute_percentage_error: 67348.2891  
loss : 7.002386610110989e-06  
mean_absolute_percentage_error : 67348.2890625
```

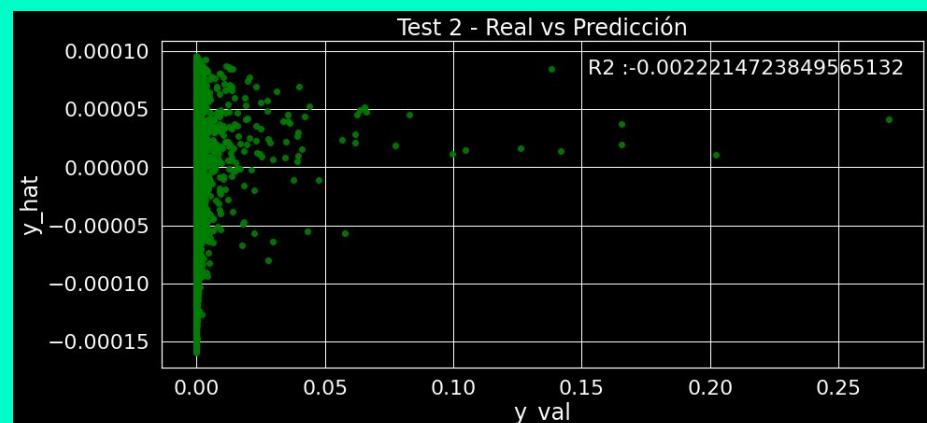
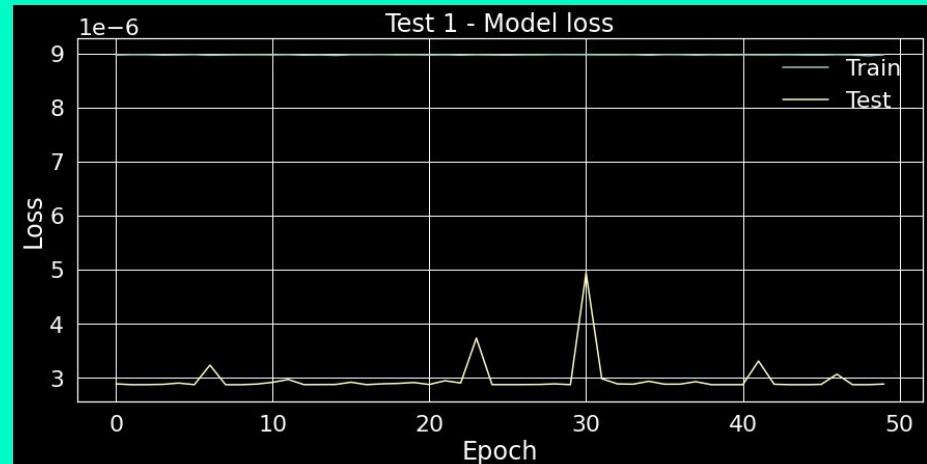


# RED NEURONAL - 2D0

## ENTRENAMIENTO



```
8941/8941 [=====] - 7s 740us/step - loss:  
6.9915e-06 - mean_absolute_percentage_error: 34551.4258  
loss : 6.991492682573153e-06  
mean_absolute_percentage_error : 34551.42578125
```



```
[ '¡Thank you!'.format() for _ in range(0,1000) ]
```

# COMERCIO INTERNACIONAL DE COLOMBIA: EXPORTACIONES E IMPORTACIONES DE COLOMBIA DE 1962 A 2017



Final Data Triathlon | Olimpiadas | Platzi

Diego Lesmes



[https://github.com/DLesmes/Data\\_Triathlon\\_2021](https://github.com/DLesmes/Data_Triathlon_2021)