INFORME FINAL-PROYECTO

Leidy Katerine Benavides Pacheco

Limpieza de Datos

Ing. Luis Armando Amaya Quiroga

Servicio Nacional de Aprendizaje SENA

Cali-valle del cauca

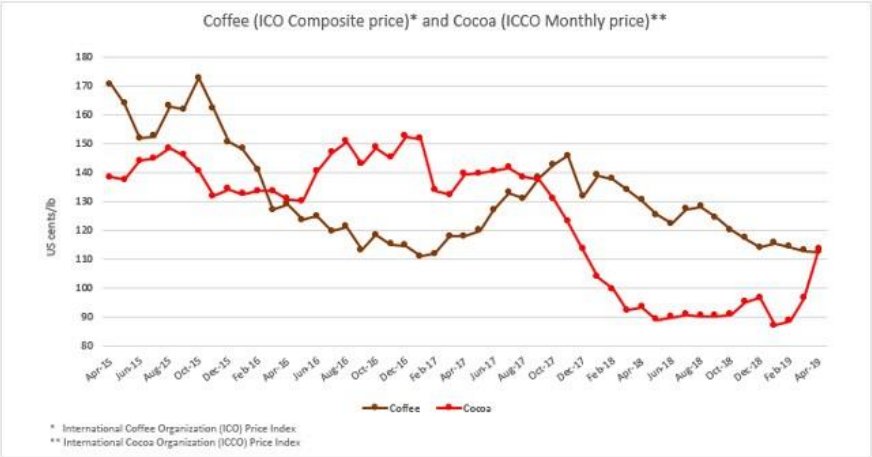
2020

INTRODUCCION

En el presente trabajo se evaluara la producción del café en Colombia donde se hace apreciaciones a nivel mundial y se representa por medio de gráficos para darle una comparación frente a los últimos años de producción de este producto y la importancia de dar a conocer el café colombiano que lo hacen uno de los productos más importantes en la competencia mundial de los productores de café, por su producción, exportación, calidad y variedad que ha posicionado de manera favorable a Colombia en los últimos tiempos.

Investigación sobre el Café

CULTIVOS DE CAFÉ FAOST



Café y cacao

Precios internacionales deprimidos del café: información sobre la naturaleza de la caída de los precios: este breve análisis describe los factores que están detrás de los bajos precios internacionales del café.

Cacao tuvo decadencias entre abril 2018 hasta octubre y también a principios de febrero 2019

A cambio con el café su punto máximo fue en octubre 2015

Fuente: www.faostat

Url: http://www.fao.org/economic/est/est-commodities/cafe-y-cacao/es/

PRODUCCIÓN Y EXPORTACIÓN DE CAFÉ DE COLOMBIA, EN SU MÁXIMO DE UN AÑO

2019-Colombia, mayor productor de café arábico suave lavado, seguirá trabajando por mantener la caficultura joven y productiva y profundizará sus esfuerzos en mejorar la calidad del café, para continuar conquistando nichos de alto valor y por esta vía mejorar la rentabilidad de la actividad cafetera.

Al cierre del 2019, el valor de la cosecha cafetera fue de 7,2 billones de pesos, un 15,8% más frente a los 6,2 billones de 2018, recursos que van directamente a dinamizar la economía de los más de 600 municipios cafeteros del país.

2018- El mes pasado se caracterizó por la subida del dólar en los mercados globales. Ese y otros factores como los mayores paises y el pico de cosecha en buena parte del país contribuyeron al aumento de la producción y las exportaciones de café.

En noviembre, según la Federación Nacional de Cafeteros, la producción de café de Colombia alcanzó un millón 300 mil sacos, cifra similar al millón 304 mil sacos cosechados en el mismo mes de 2017.

Producción de café–Noviembre 2018
(Sacos 60 kg)

Noviembre 2018	1.300.000
Noviembre 2017	1.304.000
Variación	-0,3%

Producción de café en 2019
(Sacos 60 kg)

Ene -Dic 2019	14.752.000
Ene Dic 2018	13.557.000
Variación	9%

Producción de café – Diciembre
(Sacos 60 kg)

Diciembre 2019	1.680.000
Diciembre 2018	1.283.000
Variación	31%

Producción de café–Año corrido
(Sacos 60 kg)

Enero–Noviembre 2018	12.274.000
Enero–Noviembre 2017	12.644.000
Variación	-2,9%

Producción de café-Últimos 12 meses
(Sacos 60 kg)

Diciembre 2017-Noviembre 2018	13.824.000
Diciembre 2016-Noviembre 2017	13.963.000
Variación	-1,0%

Fuente:www.federaciondecafeteros.org

Url: https://federaciondecafeteros.org/wp/listado-noticias/produccion-de-cafede-colombia-cerro-el-2019-en-148-millones-de-sacos/

Se trata de la cifra de producción más alta desde diciembre de 2017 cuando fue de un millón 550 mil sacos.

Eso representó un importante aumento frente a octubre pasado cuando se produjeron un millón 86 mil sacos.

De otro lado, a producción en lo corrido de 2018 (enero-noviembre) rozó los 12,3 millones de sacos de 60 kg, 2,9% menos frente a los más de 12,6 millones de sacos cosechados en los 11 primeros meses de 2017.

En los últimos 12 meses (diciembre 2017-noviembre 2018), la producción de café superó 13,8 millones de sacos, apenas 1% menos frente a los casi 14 millones de sacos producidos en igual periodo anterior.

Y en los primeros dos meses del año cafetero 2018-2019 (octubre-noviembre), se produjeron 2,39 millones de sacos, 0,4% más que los 2,38 millones de sacos producidos en igual lapso del año cafetero anterior.

Exportaciones

Las exportaciones de café en noviembre con respecto al mismo mes de 2017 aumentaron 7,9%, de un millón 161 mil sacos a un millón 253 mil sacos de 60 kg.

En lo corrido del año, Colombia exportó casi 11,5 millones de sacos, 2,3% menos frente a los casi 11,8 millones de sacos exportados entre enero y noviembre de 2017.

En los últimos 12 meses (diciembre 2017–noviembre 2018), las exportaciones de café superaron 12,7 millones de sacos, 3,8% menos frente a los 13,2 millones de sacos registrados en el mismo periodo anterior.

Y en los primeros dos meses del año cafetero, las exportaciones del grano superaron los 2,3 millones de sacos, 1,1% más frente a los casi 2,3 millones de sacos en el mismo periodo anterior.

Exportación de café–Noviembre de 2018
(Sacos 60 kg)

Noviembre 2018	1.253.000
Noviembre 2017	1.161.000
Variación	7,9%

Exportaciones de café – Diciembre
(Sacos 60 kg)

Diciembre 2019	1.378.000
Diciembre 2018	1.283.000
Variación	7%

Exportaciones de café año cafetero
(Sacos 60 kg)

Oct 2019 - Dic 2019	3.783.000
Oct 2018 Dic 2018	3.574.000
Variación	6%

Exportación de café–Año corrido
(Sacos 60 kg)

Enero-Noviembre 2018	11.488.000
Enero-Noviembre 2017	11.761.000
Variación	-2,3%

Exportación de café-Últimos 12 meses
(Sacos 60 kg)

Diciembre 2017-Noviembre 2018	12.710.000
Diciembre 2016-Noviembre 2017	13.211.000
Variación	-3,8%

Fuente:www.federaciondecafeteros.org

Url: https://federaciondecafeteros.org/wp/listado-noticias/produccion-de-cafede-colombia-cerro-el-2019-en-148-millones-de-sacos/

Mercados para la exportación del café colombiano

Colombia es un país reconocido en el mundo por ser el productor del café más suave, una característica que de entrada le permitirá sobresalir sobre otros competidores fuertes en el **mercado internacional.** La existencia del gremio cafetero, que está organizado en el país con el fin de regular y mejorar las prácticas en el cultivo del café, también garantiza, de alguna manera, que la **producción de café colombiano** que adquieren los **compradores internacionales** sea de calidad. De hecho, las grandes empresas colombianas cuentan con certificaciones que así lo reconocen

Los **exportadores colombianos** deberán innovar con nuevos sabores, aromas, texturas y variedades, aprovechando que hay un interés en la salud y el bienestar por parte de los consumidores. En cuanto al café instantáneo, en el mercado hay una demanda por los cafés con sabores como amaretto, ron, brandy, crema irlandesa, vainilla, maple, caramelo y chocolate, pues el café ya no es considerado únicamente como una bebida energizante: quienes lo compran están buscando calidad, variedad y frescura que pueden llevar al **café colombiano** al nivel de una bebida gourmet.

Fuente: www.colombiatrade.com.co

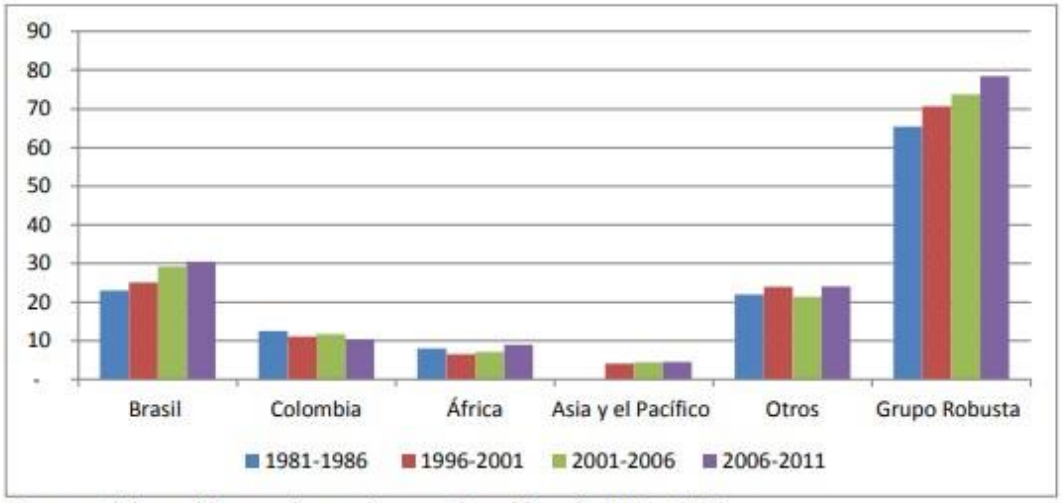
Url: https://www.colombiatrade.com.co/noticias/como-aprovecharoportunidades-para-exportar-cafe-mercados-internacionales

En el 2015 las exportaciones de cafés especiales alcanzaron los US$2.809,9 millones. Entre los países que más importan este producto se cuentan Estados Unidos con US$1.151,9 millones (41 % de participación), Japón con US$289,6 millones (10,3 %), Alemania con US$227,3 millones (8,1 %), Bélgica con US$214,8 millones (7,6 %) y Canadá con US$166,8 millones (5,9 %).

Fuente: www.compradores.procolombia.co

Url: https://compradores.procolombia.co/es/explore-oportunidades/caf-sespeciales-0

Gráfica 2. Producción de arábica anual (millones de sacos)



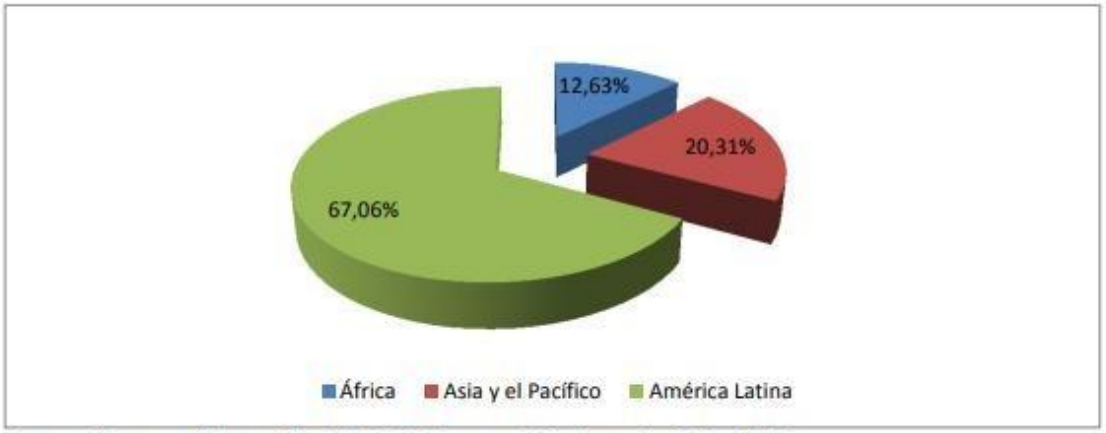
Fuente: elaboración propia con base en las cifras de ICO e ITC.

Como se aprecia en la Gráfica 2, la producción de la variedad arábica ha crecido (19,64%), al comparar la producción de 1981-1986 y 2006-2011, lo cual puede explicarse por el aumento de la producción en Brasil, África, y Asia y el Pacífico. En Colombia la producción de la variedad arábica descendió (16,94%), durante el mismo período.

Fuente: www.sic.gov.co

Url:https://www.sic.gov.co/recursos_user/documentos/promocion_competencia/Estudios_Economicos/Estudios_Economicos/Estudios_Mercado/Estudiosectori alCafe.pdf

Gráfica 5. Consumo doméstico de países productores de café (año cafetero 2010/2011, Miles de sacos)



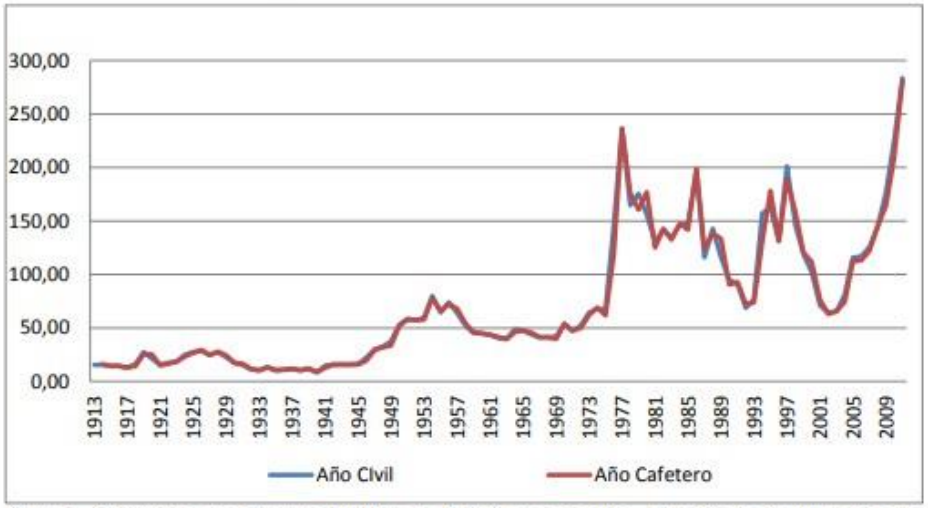
Fuente: Fuente: elaboración GEE-SIC a partir de datos de ICO e ITC

América Latina es la región del mundo que más café produce y también la que más café consume (Gráfica 5).

Fuente: www.sic.gov.co

Url:https://www.sic.gov.co/recursos_user/documentos/promocion_competencia/Estudios_Economicos/Estudios_Economicos/Estudios_Mercado/Estudiosectori alCafe.pdf

Gráfica 7. Evolución del precio externo del café colombiano (1913-2011)



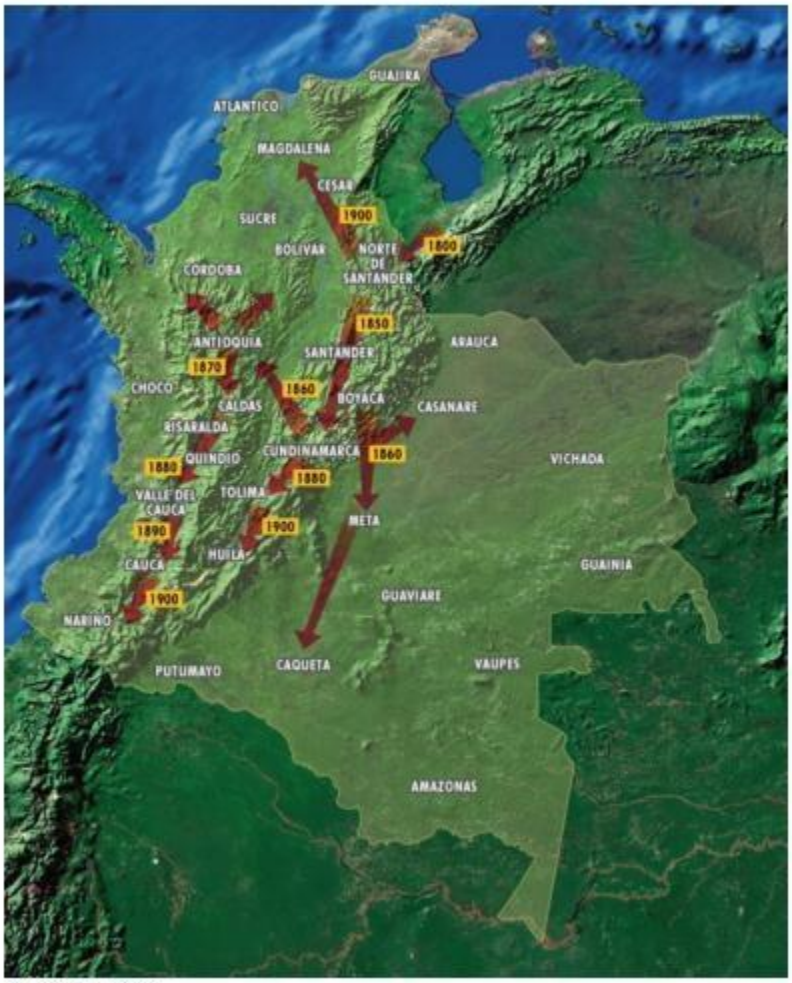
Nota 1. Centavos de dólar por libra de 453.6 gr. Excelso. Resultado de la ponderación de los precios de los 6 días anteriores
Fuente: GEE-SIC a partir de cifras de la Federación Nacional de Cafeteros.

En la Gráfica 7 se muestra la evolución del precio externo del café colombiano entre 1913 y 2011. Como se mencionó anteriormente, a partir de 1989 se rompe el pacto de cuotas, lo que explica la tendencia decreciente y el aumento de la volatilidad. El precio más bajo del café colombiano se obtuvo en 2002, al alcanzar 63,08 centavos de dólar por libra. A partir de ese momento se evidencia una tendencia creciente, obteniendo en 2011 280.74 centavos de dólar por libra.

Fuente: www.sic.gov.co

Url:https://www.sic.gov.co/recursos_user/documentos/promocion_competencia/Estudios_Economicos/Estudios_Economicos/Estudios_Mercado/Estudiosectori alCafe.pdf

Mapa 3. Arribo y expansión del café en Colombia



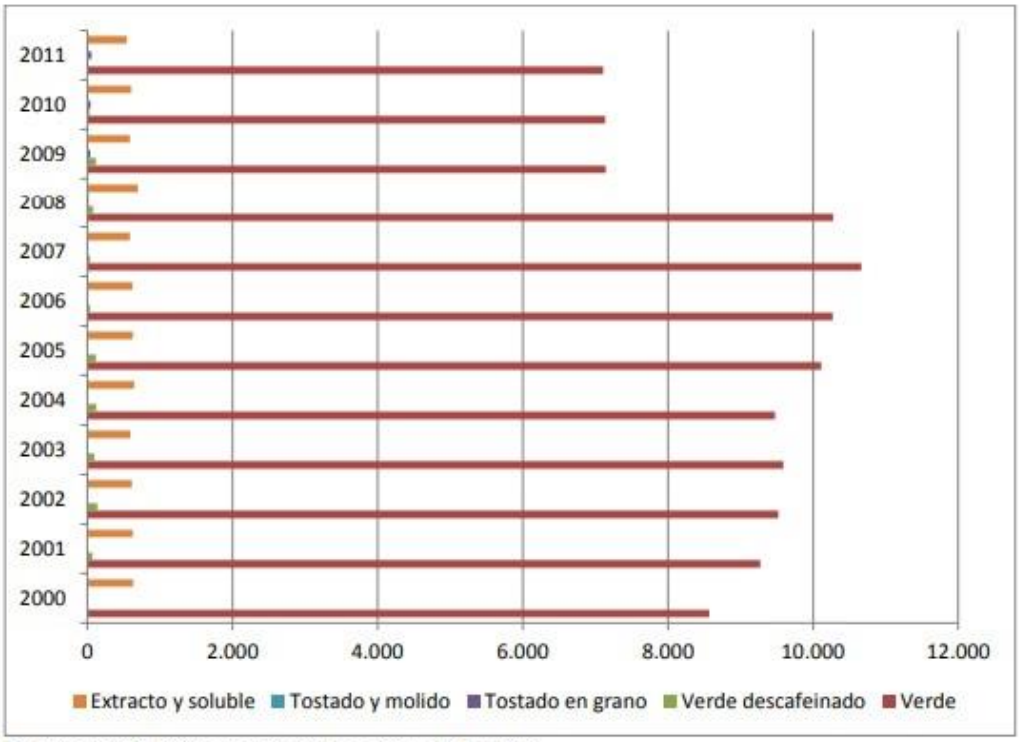
Fuente: FNC

De acuerdo con la Federación Nacional de Cafeteros (FNC), los primeros cultivos de café se ubicaron en Santander y Norte de Santander. En 1835 se realizó la primera producción comercial que correspondió a 2.560 sacos y fue exportada desde Cúcuta. Para 1850 los cultivos se expandieron hacia Cundinamarca, Antioquia y el viejo Caldas

Fuente: www.sic.gov.co

Url:https://www.sic.gov.co/recursos_user/documentos/promocion_competencia/Estudios_Economicos/Estudios_Economicos/Estudios_Mercado/Estudiosectori alCafe.pdf

Gráfica 18. Volumen de las exportaciones colombianas de café según tipo – anual. Miles de sacos de 60 Kg de café verde equivalente

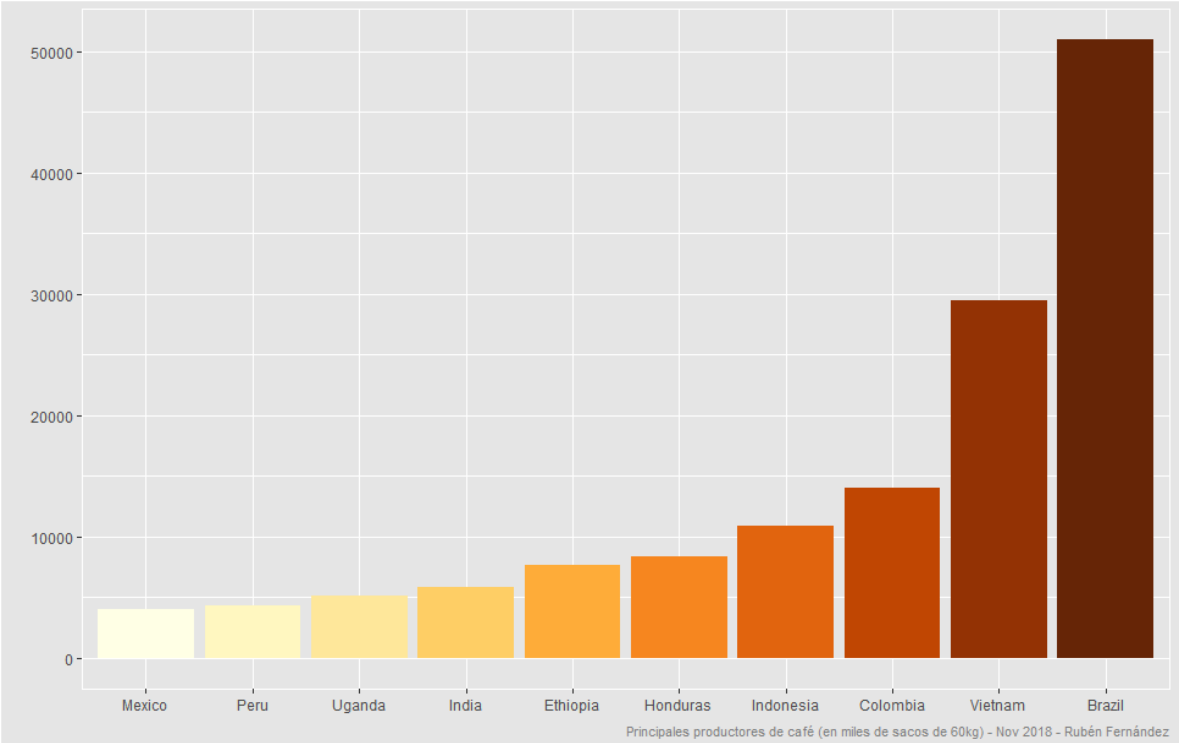


Fuente: GEE-SIC a partir de las cifras de la FNC.

De acuerdo con la Gráfica 18, el tipo de café que más exportó Colombia entre 2000 y 2011, fue el café verde; sin embargo, se evidencia que desde 2009 el volumen de café verde exportado se ha reducido, llegando a niveles inferiores con respecto al comienzo de la década. No obstante, debe considerarse que: "gracias al mayor número de caficultores vinculados a los programas de cafés especiales, las exportaciones de estos cafés crecieron 29% respecto al 2010"

Fuente: www.sic.gov.co

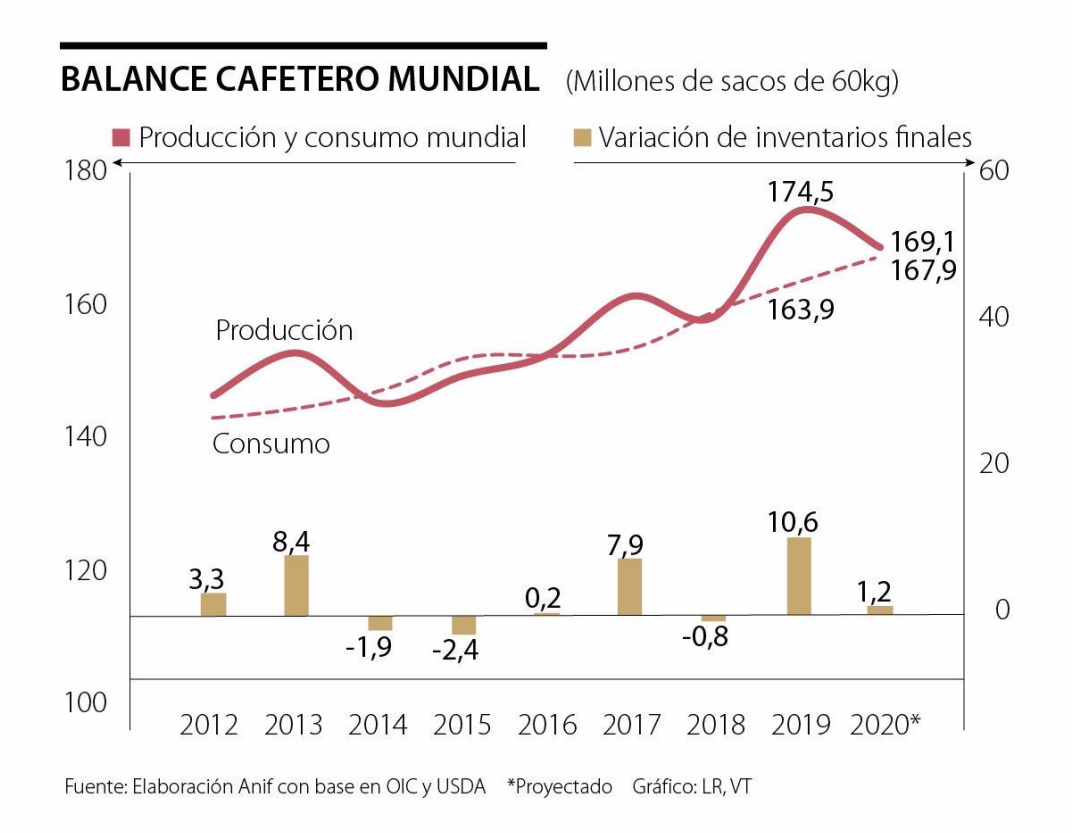
Url:https://www.sic.gov.co/recursos_user/documentos/promocion_competencia/Estudios_Economicos/Estudios_Economicos/Estudios_Mercado/Estudiosectori alCafe.pdf



Mayores productores de café (miles de sacos de 60kg). Cosecha 2017-18.

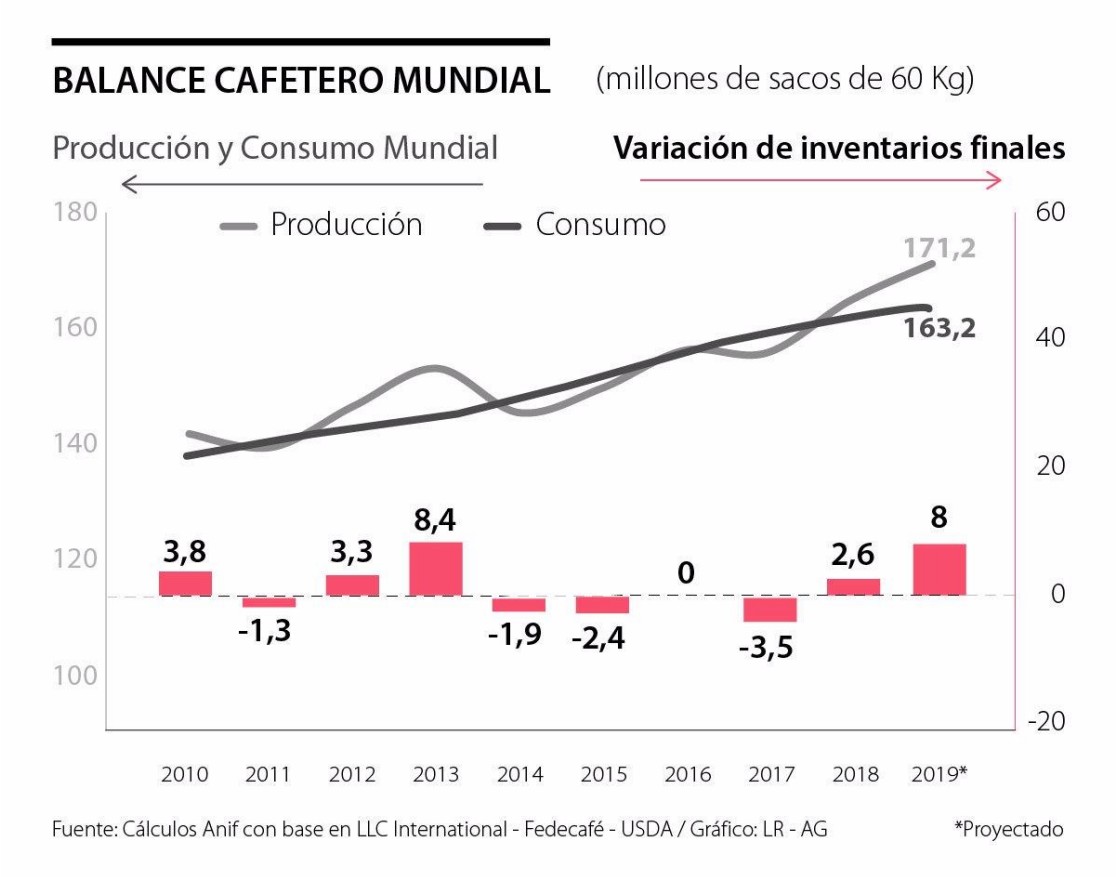
Fuente: **https://quecafe.info/mayores-productores-de-cafe-en-el-mundo/**

Colombia ocupa 8 puesto sobre el rendimiento de café a mediados de nov 2018.Por el contrario Perú produce anualmente 4,3 millones de sacos (258 mil Tm) y ocupa el segundo lugar a nivel mundial como productor y exportador de café orgánico.



Fuente: https://www.larepublica.co/analisis/sergio-clavijo-500041/panorama-cafetero-2019-2020-2920631

A lo largo del año cafetero 2018-2019, el precio internacional del café se mantuvo en niveles históricamente bajos debido a dichos excesos de oferta. En efecto, el precio internacional del grano se ubicó en US$0.98 en septiembre de 2019 (similar a un año atrás).



Fuente: https://www.larepublica.co/analisis/sergio-clavijo-500041/la-crisis-de-rentabilidad-cafetera-y-sus-perspectivas-2844218

Es evidente que detrás de la reciente caída en los precios está la mayor producción mundial de café, impulsada por Brasil (59 millones de sacos en 2018 vs. 51 millones en 2017), Vietnam (30 millones vs. 29 millones) e Indonesia (11 millones vs. 10 millones). Esto estaría incrementando la producción mundial a niveles récord de 171 millones de sacos, muy superior a los 163 millones de consumo aparente (ver gráfico adjunto). Así pues, todo parece indicar que el mercado mundial está sobre-ofrecido, induciendo esto a sustituciones de las variedades que se transan en el Contrato C de la Bolsa de New York.

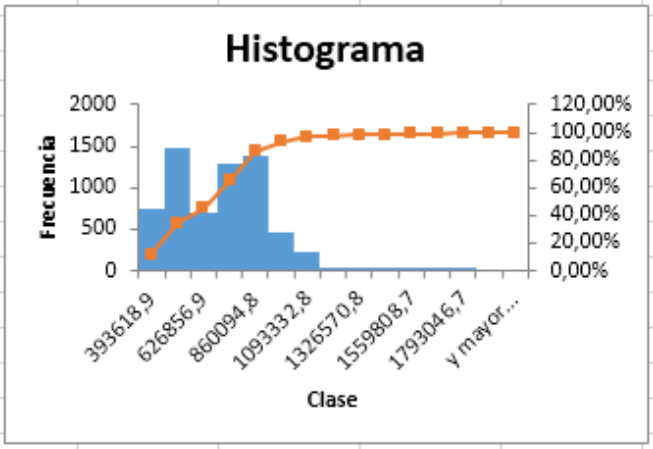
Análisis estadístico en Excel

La estadística de análisis descriptivo que se usó con los datos de la producción del café, se aplicó en cada tabla este tipo de metodología proporciono un enfoque para confeccionar un resumen de información que dan los datos de una muestra. Es decir, su meta es hacer síntesis de la información para arrojar precisión, sencillez y aclarar y ordenar los datos.

Estadísticas tomadas de precios, y producción de café

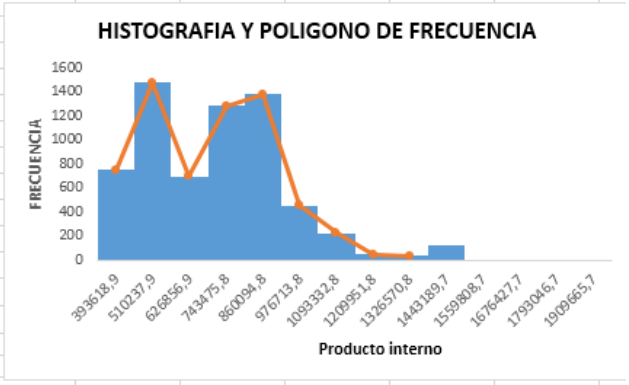
- **Precio interno diario**

Clase	*Frecuencia*	*% acumulado*
393618,9	747	11,57%
510237,9	1476	34,44%
626856,9	693	45,18%
743475,8	1284	65,08%
860094,8	1377	86,41%
976713,8	454	93,45%
1093332,8	220	96,85%
1209951,8	47	97,58%
1326570,8	32	98,08%
1443189,7	27	98,50%
1559808,7	27	98,92%
1676427,7	27	99,33%
1793046,7	27	99,75%
1909665,7	16	100,00%
y mayor...	0	100,00%



Número de clases	14
Tamaño de clases	116618,9834

	Intervalo		Grupos	Frecuencia
0	LI	Ls		
1	277000	393619,0	393618,9	747
2	393619,0	510238,0	510237,9	1476
3	510238,0	626857,0	626856,9	693
4	626857,0	743475,9	743475,8	1284
5	743475,9	860094,9	860094,8	1377
6	860094,9	976713,9	976713,8	454
7	976713,9	1093332,9	1093332,8	220
8	1093332,9	1209951,9	1209951,8	47
9	1209951,9	1326570,9	1326570,8	32
10	1326570,9	1443189,8	1443189,7	124
11	1443189,8	1559808,8	1559808,7	#N/A
12	1559808,8	1676427,8	1676427,7	#N/A
13	1676427,8	1793046,8	1793046,7	#N/A
14	1793046,8	1909665,8	1909665,7	#N/A

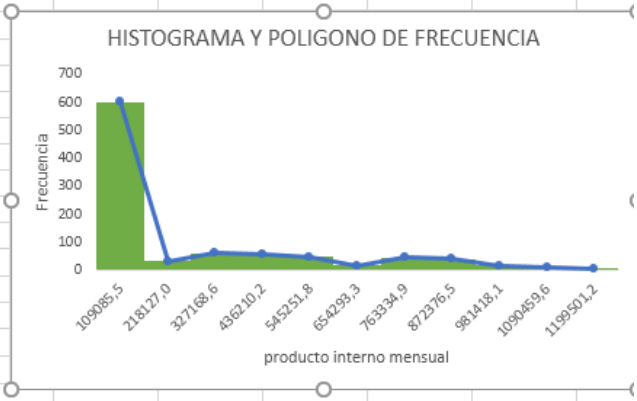


Fuente:https://federaciondecafeteros.org/wp/estadisticas-cafeteras/

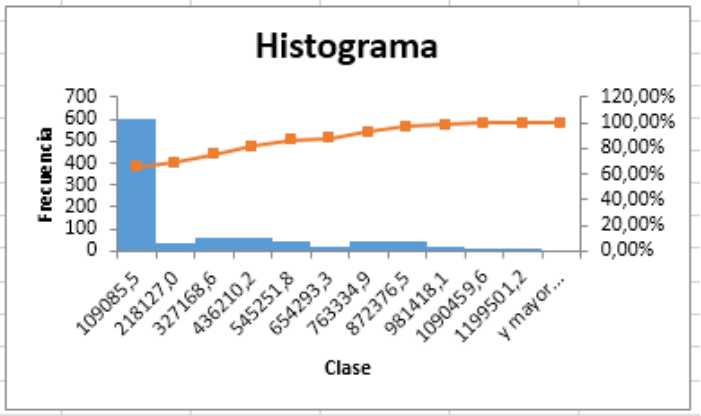
- **Precio interno mensual**

Número de clases	11
Tamaño de clases	109041,57

	Intervalo		Grupos	Frecuencia
0	LI	Ls		
1	44	109085,6	109085,5	601
2	109085,6	218127,1	218127,0	31
3	218127,1	327168,7	327168,6	59
4	327168,7	436210,3	436210,2	57
5	436210,3	545251,9	545251,8	46
6	545251,9	654293,4	654293,3	15
7	654293,4	763335,0	763334,9	44
8	763335,0	872376,6	872376,5	40
9	872376,6	981418,2	981418,1	16
10	981418,2	1090459,7	1090459,6	8
11	1090459,7	1199501,3	1199501,2	3



Clase	Frecuencia	% acumulado
109085,5	601	65,33%
218127,0	31	68,70%
327168,6	59	75,11%
436210,2	57	81,30%
545251,8	46	86,30%
654293,3	15	87,93%
763334,9	44	92,72%
872376,5	40	97,07%
981418,1	16	98,80%
1090459,6	8	99,67%
1199501,2	3	100,00%
y mayor...	0	100,00%

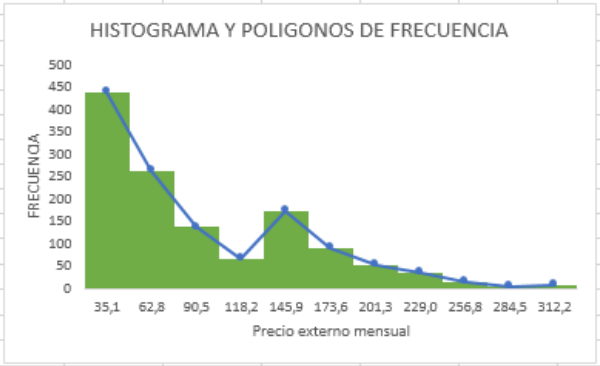


Fuente:https://federaciondecafeteros.org/wp/estadisticas-cafeteras/

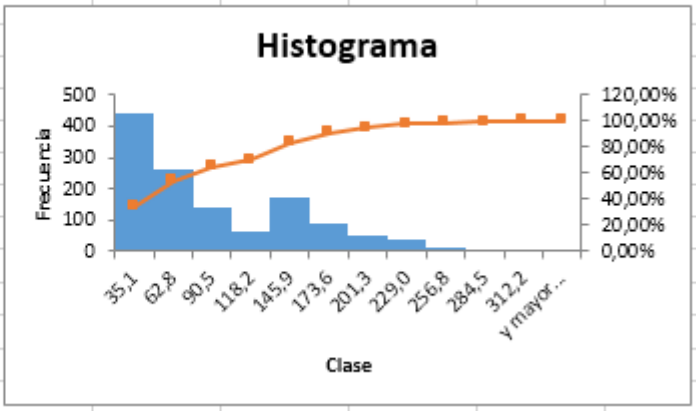
- **Precio de exportación mensual**

Número de clases	11	
Tamaño de clase	27,707943	

	Intervalo		Grupos	Frecuencia
0	LI	Ls		
1	7,48	35,2	35,1	439
2	35,2	62,9	62,8	263
3	62,9	90,6	90,5	138
4	90,6	118,3	118,2	67
5	118,3	146,0	145,9	174
6	146,0	173,7	173,6	91
7	173,7	201,4	201,3	53
8	201,4	229,1	229,0	36
9	229,1	256,9	256,8	15
10	256,9	284,6	284,5	6
11	284,6	312,3	312,2	9



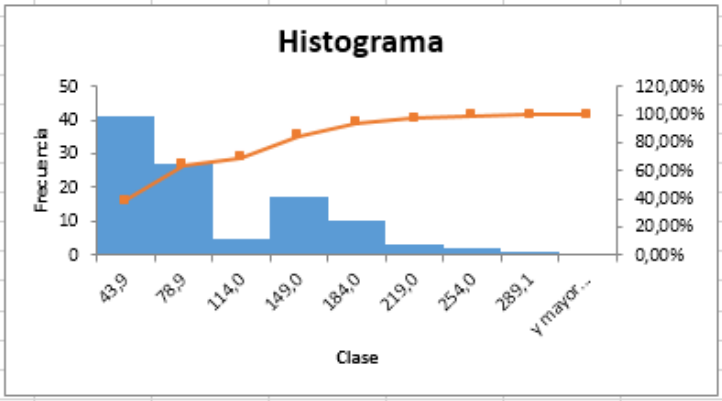
Clase	Frecuencia	% acumulado
35,1	439	33,98%
62,8	263	54,33%
90,5	138	65,02%
118,2	67	70,20%
145,9	174	83,67%
173,6	91	90,71%
201,3	53	94,81%
229,0	36	97,60%
256,8	15	98,76%
284,5	6	99,23%
312,2	9	99,92%
y mayor...	1	100,00%



Fuente:https://federaciondecafeteros.org/wp/estadisticas-cafeteras/

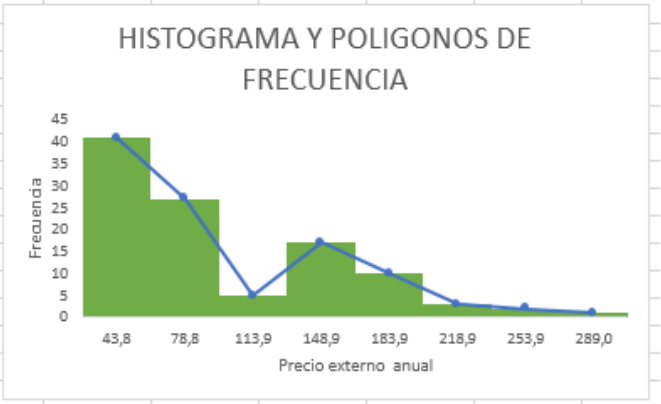
- **Precio de exportación Anual Cafetero**

Clase	*Frecuencia*	*% acumulado*
43,9	41	38,68%
78,9	27	64,15%
114,0	5	68,87%
149,0	17	84,91%
184,0	10	94,34%
219,0	3	97,17%
254,0	2	99,06%
289,1	1	100,00%
y mayor...	0	100,00%



Número de clases	8
Tamaño de clases	35,019258

	Intervalo		Grupos	Frecuencia
0	LI	Ls		
1	8,90583333	43,9	43,8	41
2	43,9	78,9	78,8	27
3	78,9	114,0	113,9	5
4	114,0	149,0	148,9	17
5	149,0	184,0	183,9	10
6	184,0	219,0	218,9	3
7	219,0	254,0	253,9	2
8	254,0	289,1	289,0	1



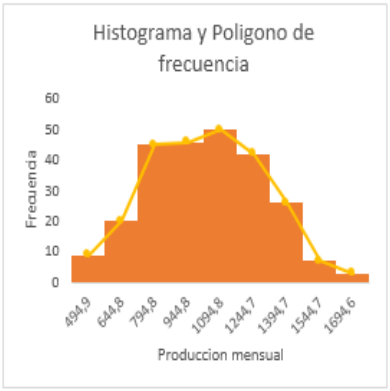
Fuente:https://federaciondecafeteros.org/wp/estadisticas-cafeteras/

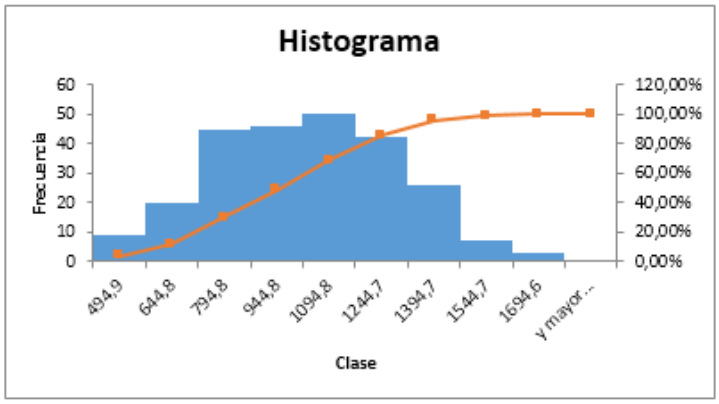
- **Producción mensual registrada**

Produccion	
Media	959,8548387
Error típico	16,79627635
Mediana	966,5
Moda	770
Desviación	264,5080245
Varianza de	69964,49504
Curtosis	-0,474064951
Coeficiente	0,0629848
Rango	1335
Mínimo	345
Máximo	1680
Suma	238044
Cuenta	248

Número de clases 9
Tamaño de clases 149,9715

	Intervalo		Grupos	Frecuencia
0	LI	Ls		
1	345	495,0	494,9	9
2	495,0	644,9	644,8	20
3	644,9	794,9	794,8	45,00
4	794,9	944,9	944,8	46
5	944,9	1094,9	1094,8	50
6	1094,9	1244,8	1244,7	42,00
7	1244,8	1394,8	1394,7	26
8	1394,8	1544,8	1544,7	7,00
9	1544,8	1694,7	1694,6	3,00





Clase	*Frecuencia*	*% acumulado*
494,9	9	3,63%
644,8	20	11,69%
794,8	45	29,84%
944,8	46	48,39%
1094,8	50	68,55%
1244,7	42	85,48%
1394,7	26	95,97%
1544,7	7	98,79%
1694,6	3	100,00%
y mayor...	0	100,00%

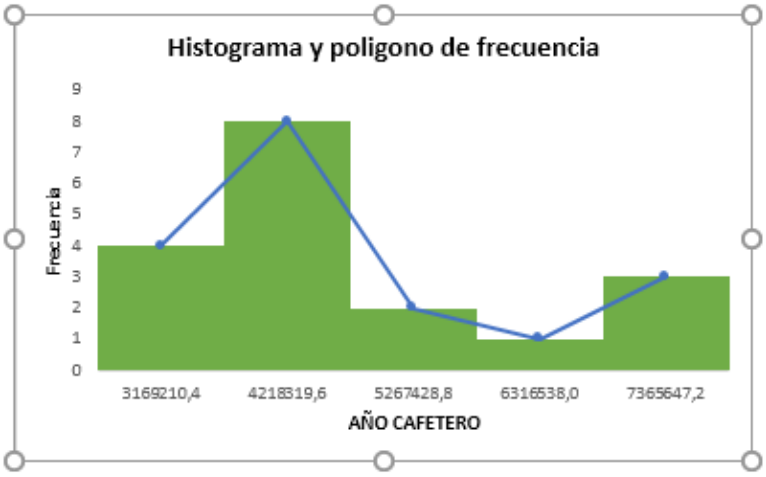
Fuente:https://federaciondecafeteros.org/wp/estadisticas-cafeteras/

- **Valor cosecha**

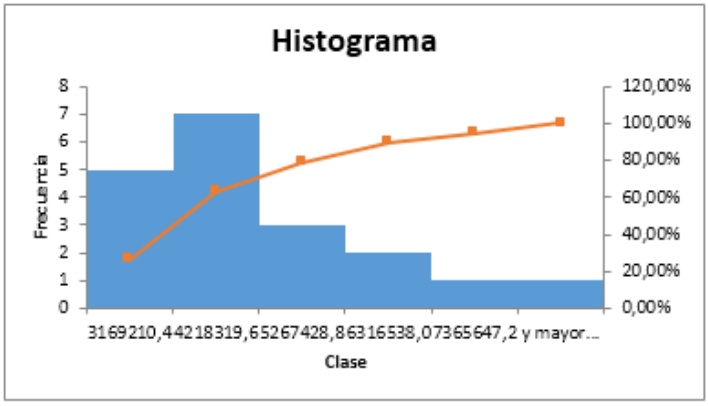
Años cafeteros	
Media	4261583,774
Error típico	394430,5277
Mediana	3719387
Moda	#N/A
Desviación	1719282,811
Varianza de	2,95593E+12
Curtosis	-0,311167708
Coeficiente	0,656105465
Rango	6052722,242
Mínimo	2009660
Máximo	8062382,242
Suma	80970091,7
Cuenta	19

Número de clases	5
Tamaño de clases	1159550,5

	Intervalo		Grupos	Frecuencia
0	LI	Ls		
1	2009660	3169210,5	3169210,4	4
2	3169210,5	4218319,7	4218319,6	8
3	4218319,7	5267428,9	5267428,8	2
4	5267428,9	6316538,1	6316538,0	1
5	6316538,1	7365647,3	7365647,2	3



Clase	*Frecuencia*	*% acumulado*
3169210,4	5	26,32%
4218319,6	7	63,16%
5267428,8	3	78,95%
6316538,0	2	89,47%
7365647,2	1	94,74%
y mayor...	1	100,00%



Fuente:https://federaciondecafeteros.org/wp/estadisticas-cafeteras/

Análisis de Jupyter

In [1]:
```
import pandas as pd
# importa o carga la libreria para manejo de dataframes denominada pandas y se asigna el alias pd
```

In [2]:
```
pd.read_csv("produccion.csv")
# se indica que lea o carge el dataframe (archivo csv) produccion.csv
```

Out[2]:

	Anio	Departamento	Producto	Area (ha)	Produccion (ton)	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
0	2007	ANTIOQUIA	CAFE	112,343.60	120,500.80	1.07	14.54	14.66
1	2007	BOLIVAR	CAFE	502.00	446.00	0.89	0.05	0.07
2	2007	BOYACA	CAFE	11,374.50	9,683.10	0.85	1.17	1.48
3	2007	CALDAS	CAFE	78,393.65	92,815.00	1.18	11.20	10.23
4	2007	CAQUETA	CAFE	2,295.00	2,134.00	0.93	0.26	0.30
...	...	...	...	...	...	...	...	...
261	2018	QUINDIO	CAFE	16,374.73	17,739.03	1.08	2.07	2.21
262	2018	RISARALDA	CAFE	35,874.73	45,918.75	1.28	5.37	4.83
263	2018	SANTANDER	CAFE	42,269.07	55,918.71	1.32	6.53	5.69
264	2018	TOLIMA	CAFE	97,304.04	97,451.31	1.00	11.39	13.11
265	2018	VALLE DEL CAUCA	CAFE	48,305.31	49,667.88	1.03	5.80	6.51

266 rows × 8 columns

In [3]:
```
produccion_df=pd.read_csv("produccion.csv")
# el datadrame completo anterior se almacena en la variable produccion_df
```

In [4]:
```
produccion_df
# Ahora se indica que liste el contenido del dataframe produccion_df
```

Out[4]:

	Anio	Departamento	Producto	Area (ha)	Produccion (ton)	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
0	2007	ANTIOQUIA	CAFE	112,343.60	120,500.80	1.07	14.54	14.66
1	2007	BOLIVAR	CAFE	502.00	446.00	0.89	0.05	0.07
2	2007	BOYACA	CAFE	11,374.50	9,683.10	0.85	1.17	1.48
3	2007	CALDAS	CAFE	78,393.65	92,815.00	1.18	11.20	10.23
4	2007	CAQUETA	CAFE	2,295.00	2,134.00	0.93	0.26	0.30

In [5]:
```
type(produccion_df)
# Se describe la estructura y el tipo del dataframe utilizado
```

Out[5]:
```
pandas.core.frame.DataFrame
```

In [6]:
```
produccion_df.dtypes
# Se describe la estructura y el tipo cada elemento o campo del dataframe utilizado
```

Out[6]:
```
Anio                          int64
Departamento                 object
Producto                     object
Area (ha)                    object
Produccion (ton)             object
Rendimiento (ha/ton)        float64
Produccion Nacional (ton)   float64
Area Nacional (ha)          float64
dtype: object
```

In [7]:
```
produccion_df.columns
# Se describe los nombres de cada una de las columnas o campos del dataframe utilizado
```

Out[7]:
```
Index(['Anio', 'Departamento', 'Producto', 'Area (ha)', 'Produccion (ton)',
       'Rendimiento (ha/ton)', 'Produccion Nacional (ton)',
       'Area Nacional (ha)'],
      dtype='object')
```

In [8]:
```
produccion_df.shape
# Se describe la cantidad de filas y luego de las columnas del dataframe utilizado
```

Out[8]:
```
(266, 8)
```

In [9]:
```
pd.unique(produccion_df['Producto'])
# Se describe los valores del campo Nombres y el tipo
```

Out[9]:
```
array(['CAFE'], dtype=object)
```

In [10]:
```
pd.unique(produccion_df['Anio'])
# Se describe los valores del campo Nombres y el tipo
```

Out[10]:
```
array([2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017,
       2018], dtype=int64)
```

In [11]:
```
pd.unique(produccion_df['Departamento'])
# Se describe los valores del campo Nombres y el tipo
```

In [12]:
```
pd.unique(produccion_df['Area (ha)'])
# Se describe los valores del campo Nombres y el tipo
```

Out[12]:
```
array(['112,343.60', '502.00', '11,374.50', '78,393.65', '2,295.00',
       '2,605.00', '53,471.00', '23,172.00', '290.00', '43,017.30',
       '89,661.56', '4,785.00', '17,506.00', '2,048.00', '24,458.50',
       '30,171.84', '35.00', '19,904.00', '47,689.25', '34,406.67',
       '91,679.10', '76,667.80', '114,694.00', '572.00', '10,778.50',
       '74,897.00', '2,735.00', '2,149.00', '56,208.00', '23,198.00',
       '90.00', '43,633.35', '89,131.20', '4,553.00', '17,521.00',
       '2,146.00', '25,582.00', '31.00', '19,571.00', '47,227.00',
       '34,169.37', '86,829.20', '72,419.00', '112,420.20', '770.00',
       '10,672.50', '73,083.00', '2,332.00', '1,904.00', '57,860.00',
       '23,420.00', '70.00', '43,475.84', '86,726.78', '4,488.00',
       '17,036.00', '2,216.00', '26,467.20', '33,552.58', '23.00',
       '19,052.00', '45,428.00', '37,985.90', '88,667.00', '67,001.30',
       '111,602.71', '0.00', '850.00', '9,427.00', '72,240.58',
       '2,536.00', '2,198.00', '55,162.00', '22,489.50', '157.50',
       '44,264.16', '87,139.53', '4,207.00', '17,000.00', '2,326.00',
       '23,504.05', '30,731.96', '24.00', '18,159.00', '47,308.00',
       '39,000.64', '84,658.70', '69,332.10', '106,419.57', '10.00',
       '8,441.74', '66,331.61', '2,810.00', '2,081.50', '54,246.42',
       '22,350.00', '37,478.87', '78,792.21', '4,100.00', '16,577.00',
       '2,578.00', '24,263.80', '21,520.45', '40.00', '20,139.30',
       '44,733.64', '37,282.04', '93,145.35', '68,038.40', '112,221.14',
       '870.00', '6,698.20', '54,871.88', '2,882.50', '2,322.00',
       '56,825.00', '22,911.00', '37,175.06', '79,809.34', '5,143.00',
```

In [13]:
```
pd.unique(produccion_df['Produccion (ton)'])
# Se describe los valores del campo Nombres y el tipo
```

Out[13]:
```
array(['120,500.80', '446.00', '9,683.10', '92,815.00', '2,134.00',
       '2,048.40', '51,348.00', '13,278.50', '205.90', '33,729.14',
       '129,052.51', '2,958.70', '14,005.00', '1,617.20', '31,770.05',
       '13,593.24', '34.00', '25,426.00', '72,842.55', '29,469.52',
       '112,322.38', '69,618.24', '113,505.20', '711.00', '9,547.30',
       '86,884.00', '2,469.00', '1,388.13', '48,073.00', '13,841.45',
       '68.00', '78,254.77', '131,316.47', '2,328.90', '14,017.00',
       '1,656.96', '31,262.50', '13,593.25', '35.60', '23,669.00',
       '60,079.00', '29,016.75', '101,201.88', '65,666.43', '103,703.00',
       '292.60', '8,567.97', '81,668.22', '2,332.00', '2,079.70',
       '47,221.00', '12,770.00', '78.75', '37,118.07', '104,609.42',
       '2,340.40', '13,412.80', '1,672.60', '27,487.71', '10,221.69',
       '26.70', '21,985.00', '53,648.00', '26,311.61', '88,633.10',
       '62,711.08', '121,253.38', '0.00', '510.00', '7,083.07',
       '95,957.90', '2,902.50', '2,564.86', '45,113.00', '13,276.08',
       '98.00', '37,214.80', '104,336.56', '2,393.00', '13,600.00',
       '2,221.90', '24,594.10', '22,111.65', '21,065.00', '72,091.00',
       '27,094.16', '94,230.20', '69,496.65', '115,267.98', '12.00',
       '5,643.39', '78,805.87', '2,528.40', '2,023.50', '41,645.39',
       '11,035.85', '32,780.35', '85,150.66', '1,933.00', '13,301.60',
       '2,533.75', '24,073.95', '12,332.00', '45.80', '20,814.11',
       '49,042.31', '22,089.82', '53,288.42', '65,475.63', '91,621.30',
```

In [14]:
```
pd.unique(produccion_df['Rendimiento (ha/ton)'])
# Se describe los valores del campo Nombres y el tipo
```

Out[14]:
```
array([1.07, 0.89, 0.85, 1.18, 0.93, 0.79, 0.96, 0.57, 0.71, 0.78, 1.44,
       0.62, 0.8 , 1.3 , 0.45, 0.97, 1.28, 1.53, 0.86, 1.23, 0.91, 0.99,
       1.24, 1.16, 0.9 , 0.65, 0.6 , 0.76, 1.79, 1.47, 0.51, 0.77, 1.22,
       1.15, 1.21, 1.27, 1.17, 0.92, 0.38, 1.12, 1.  , 1.09, 0.82, 0.55,
       1.13, 0.52, 0.75, 1.04, 0.3 , 0.69, 0.94, 0.  , 1.33, 1.14, 0.59,
       0.84, 1.2 , 1.05, 0.72, 1.11, 1.52, 1.08, 0.67, 1.19, 0.49, 0.87,
       0.47, 0.98, 1.03, 1.1 , 0.74, 2.  , 0.83, 1.01, 0.63, 0.81, 0.88,
       0.66, 0.7 , 1.06, 0.64, 1.02, 0.95, 1.41, 1.32, 1.5 , 1.26, 1.37,
       1.35, 1.25, 1.45, 1.29, 1.4 , 1.38])
```

In [15]:
```
pd.unique(produccion_df['Produccion Nacional (ton)'])
# Se describe los valores del campo Nombres y el tipo
```

Out[15]:
```
array([1.454e+01, 5.000e-02, 1.170e+00, 1.120e+01, 2.600e-01, 2.500e-01,
       6.190e+00, 1.600e+00, 2.000e-02, 4.070e+00, 1.557e+01, 3.600e-01,
       1.690e+00, 2.000e-01, 3.830e+00, 1.640e+00, 0.000e+00, 3.070e+00,
       8.790e+00, 3.560e+00, 1.355e+01, 8.400e+00, 1.370e+01, 9.000e-02,
       1.150e+00, 1.049e+01, 3.000e-01, 1.700e-01, 5.800e+00, 1.670e+00,
       1.000e-02, 9.440e+00, 1.585e+01, 2.800e-01, 3.770e+00, 2.860e+00,
       7.250e+00, 3.500e+00, 1.221e+01, 7.930e+00, 1.463e+01, 4.000e-02,
       1.210e+00, 1.152e+01, 3.300e-01, 2.900e-01, 6.660e+00, 1.800e+00,
       5.240e+00, 1.476e+01, 1.890e+00, 2.400e-01, 3.880e+00, 1.440e+00,
       3.100e+00, 7.570e+00, 3.710e+00, 1.250e+01, 8.850e+00, 1.556e+01,
       7.000e-02, 9.100e-01, 1.231e+01, 3.700e-01, 5.790e+00, 1.700e+00,
       4.780e+00, 1.339e+01, 3.100e-01, 1.750e+00, 3.160e+00, 2.840e+00,
       2.700e+00, 9.250e+00, 3.480e+00, 1.209e+01, 8.920e+00, 1.800e+01,
       8.000e-02, 8.800e-01, 3.800e-01, 3.200e-01, 6.500e+00, 1.720e+00
```

In [16]:
```
pd.unique(produccion_df['Area Nacional (ha)'])
# Se describe los valores del campo Nombres y el tipo
```

Out[16]:
```
array([1.466e+01, 7.000e-02, 1.480e+00, 1.023e+01, 3.000e-01, 3.400e-01,
       6.980e+00, 3.020e+00, 4.000e-02, 5.610e+00, 1.170e+01, 6.200e-01,
       2.280e+00, 2.700e-01, 3.190e+00, 3.940e+00, 0.000e+00, 2.600e+00,
       6.220e+00, 4.490e+00, 1.196e+01, 1.000e+01, 1.513e+01, 8.000e-02,
       1.420e+00, 9.880e+00, 3.600e-01, 2.800e-01, 7.410e+00, 3.060e+00,
       1.000e-02, 5.750e+00, 1.175e+01, 6.000e-01, 2.310e+00, 3.370e+00,
       3.980e+00, 2.580e+00, 6.230e+00, 4.510e+00, 1.145e+01, 9.550e+00,
       1.490e+01, 1.000e-01, 1.410e+00, 9.680e+00, 3.100e-01, 2.500e-01,
       7.670e+00, 3.100e+00, 5.760e+00, 1.149e+01, 5.900e-01, 2.260e+00,
       2.900e-01, 3.510e+00, 4.450e+00, 2.520e+00, 6.020e+00, 5.030e+00,
       8.880e+00, 1.499e+01, 1.100e-01, 1.270e+00, 9.710e+00, 2.000e-02,
       5.950e+00, 1.171e+01, 5.700e-01, 3.160e+00, 4.130e+00, 2.440e+00,
       6.360e+00, 5.240e+00, 1.137e+01, 9.310e+00, 1.494e+01, 1.200e-01,
       1.180e+00, 3.900e-01, 7.610e+00, 3.140e+00, 5.260e+00, 1.106e+01,
       5.800e-01, 2.330e+00, 3.410e+00, 2.830e+00, 6.280e+00, 5.230e+00,
       1.308e+01, 1.580e+01, 9.400e-01, 7.720e+00, 4.100e-01, 3.300e-01,
       8.000e+00, 3.220e+00, 1.123e+01, 7.200e-01, 2.490e+00, 3.910e+00,
```

In [17]:
```
produccion_df['Anio']
# Se describe los valores del campo Nombres y el tipo
```

Out[17]:
```
0      2007
1      2007
2      2007
3      2007
4      2007
       ...
261    2018
262    2018
263    2018
264    2018
265    2018
Name: Anio, Length: 266, dtype: int64
```

In [18]:
```
produccion_df['Producto']
# Se describe los valores del campo Nombres y el tipo
```

Out[18]:
```
0      CAFE
1      CAFE
2      CAFE
3      CAFE
4      CAFE
       ...
261    CAFE
262    CAFE
263    CAFE
264    CAFE
265    CAFE
Name: Producto, Length: 266, dtype: object
```

In [19]:
```
produccion_df['Departamento']+ produccion_df['Producto']
# Se describe los valores del campo Departamento y del Producto y el tipo
```

Out[19]:
```
0            ANTIOQUIACAFE
1              BOLIVARCAFE
2               BOYACACAFE
3               CALDASCAFE
4              CAQUETACAFE
               ...
261            QUINDIOCAFE
262          RISARALDACAFE
263          SANTANDERCAFE
264             TOLIMACAFE
265    VALLE DEL CAUCACAFE
Length: 266, dtype: object
```

In [20]:
```
produccion_df['Anio'],produccion_df['Producto']
# Se describe los valores del campo Anio y el Producto pero separados
```

Out[20]:
```
(0      2007
 1      2007
 2      2007
 3      2007
 4      2007
        ...
 261    2018
 262    2018
 263    2018
 264    2018
 265    2018
Name: Anio, Length: 266, dtype: int64,
0      CAFE
1      CAFE
2      CAFE
3      CAFE
4      CAFE
       ...
261    CAFE
262    CAFE
263    CAFE
264    CAFE
265    CAFE
Name: Producto, Length: 266, dtype: object)
```

In [22]:
```
produccion_df['Anio'].min()
# Se describe el valor minimo del año del Dataframe
```

Out[22]:
```
2007
```

In [23]:
```
produccion_df['Anio'].max()
# Se describe el valor Maximo del año del Dataframe
```

Out[23]:
```
2018
```

In [24]:
```
produccion_df['Produccion (ton)'].min()
# Se describe el valor minimo de la produccion del Dataframe
```

Out[24]:
```
'0.00'
```

In [25]:
```
produccion_df['Produccion (ton)'].max()
# Se describe el valor minimo de la produccion del Dataframe
```

Out[25]:
```
'98.00'
```

In [26]:
```
produccion_df['Produccion Nacional (ton)'].min()
# Se describe el valor minimo de la produccion nacional del Dataframe
```

Out[26]:
```
0.0
```

In [27]:
```
produccion_df['Produccion Nacional (ton)'].max()
# Se describe el valor minimo de la produccion nacional del Dataframe
```

Out[27]:
```
18.67
```

In [28]:
```
produccion_df['Departamento'].count()
# Se describe el total de Departamentos almacenados en el Dataframe
```

Out[28]:
```
266
```

In [29]:
```
produccion_df['Anio'].isnull()
#el método isnull devuelve una serie booleana que almacena True para siempre y False para un valor No nulo.
```

Out[29]:
```
0      False
1      False
2      False
3      False
4      False
       ...
261    False
262    False
263    False
264    False
```

In [31]:
```
produccion_df['Area (ha)'].isnull().sum()
```

Out[31]:
```
0
```

In [32]:
```
produccion_df['Rendimiento (ha/ton)'].isnull().sum()
```

Out[32]:
```
0
```

In [33]:
```
produccion_grouped_Producto=produccion_df.groupby("Producto").sum()
produccion_grouped_Producto
```

Out[33]:

Producto	Anio	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
CAFE	535317	249.09	1200.01	1199.98

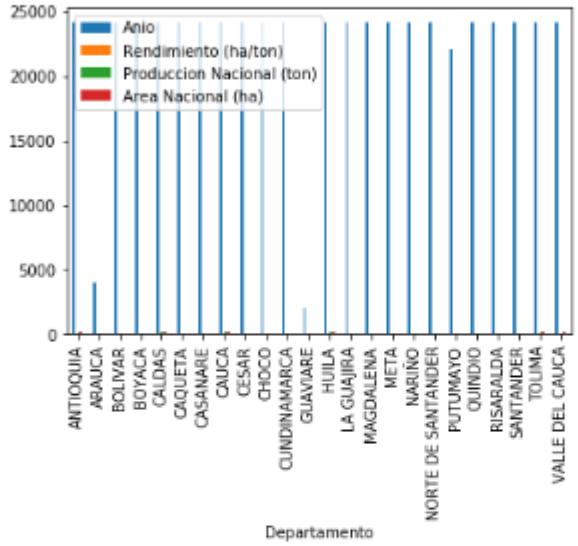
In [34]:
```
produccion_grouped_Departamento=produccion_df.groupby("Departamento").sum()
produccion_grouped_Departamento
```

Out[34]:

Departamento	Anio	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
ANTIOQUIA	24150	13.01	183.32	172.15
ARAUCA	4021	1.20	0.00	0.00
BOLIVAR	24150	9.07	1.01	1.39
BOYACA	24150	9.41	11.89	15.33
CALDAS	24150	14.00	115.87	100.06
CAQUETA	24150	12.54	4.83	4.67
CASANARE	24150	10.01	3.09	3.73
CAUCA	24150	11.36	99.87	105.83
CESAR	24150	7.95	25.29	38.18
CHOCO	24150	12.50	0.21	0.23
CUNDINAMARCA	24150	11.48	55.98	59.03
GUAVIARE	2012	0.00	0.00	0.00
HUILA	24150	13.98	188.60	165.56
LA GUAJIRA	24150	7.45	4.98	7.96
MAGDALENA	24150	9.21	21.17	27.77
META	24150	11.40	3.88	4.06

In [35]:
```
import numpy as np
import re
import sys
%matplotlib inline
produccion_grouped_Departamento.plot(kind='bar')
```

Out[35]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x92af948>
```



In [36]:
```
produccion_grouped_Producto=produccion_df.groupby(["Anio","Producto"]).sum()
produccion_grouped_Producto
```

Out[36]:

		Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
Anio	**Producto**			
2007	**CAFE**	20.91	100.01	100.00
2008	**CAFE**	21.62	100.00	99.99
2009	**CAFE**	19.39	100.00	99.98
2010	**CAFE**	20.84	100.01	100.00
2011	**CAFE**	19.65	100.02	100.00
2012	**CAFE**	19.75	99.99	100.00
2013	**CAFE**	16.71	100.00	99.99
2014	**CAFE**	18.09	100.00	100.00

In [37]:
```
produccion_grouped_Producto_Produccion=produccion_df.groupby(["Departamento","Producto","Produccion (ton)"]).sum()
produccion_grouped_Producto_Produccion
```

Out[37]:

Departamento	Producto	Produccion (ton)	Anio	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
ANTIOQUIA	CAFE	102,403.24	2013	0.93	15.70	14.22
		103,703.00	2009	0.92	14.63	14.90
		111,452.91	2014	1.01	15.30	13.84
		113,505.20	2008	0.99	13.70	15.13
		115,267.98	2011	1.08	18.00	14.94
...	...	...	...	...	...	...
VALLE DEL CAUCA	CAFE	62,711.08	2009	0.94	8.85	8.88
		65,475.63	2011	0.96	10.22	9.55
		65,666.43	2008	0.91	7.93	9.55
		69,496.65	2010	1.00	8.92	9.31
		69,618.24	2007	0.91	8.40	10.00

263 rows × 4 columns

In [38]:
```
produccion_grouped_Producto1=produccion_df.groupby(["Departamento","Producto"]).sum()
produccion_grouped_Producto1
```

Out[38]:

Departamento	Producto	Anio	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
ANTIOQUIA	CAFE	24150	13.01	183.32	172.15
ARAUCA	CAFE	4021	1.20	0.00	0.00
BOLIVAR	CAFE	24150	9.07	1.01	1.39
BOYACA	CAFE	24150	9.41	11.89	15.33
CALDAS	CAFE	24150	14.00	115.87	100.06
CAQUETA	CAFE	24150	12.54	4.83	4.67
CASANARE	CAFE	24150	10.01	3.09	3.73
CAUCA	CAFE	24150	11.36	99.87	105.83
CESAR	CAFE	24150	7.95	25.29	38.18
CHOCO	CAFE	24150	12.50	0.21	0.23

In [42]:
```
produccion_df.head()
# describe el principio del Dataframe, por si es muy grande facilita ver solo el comienzo
```

Out[42]:

	Anio	Departamento	Producto	Area (ha)	Produccion (ton)	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
0	2007	ANTIOQUIA	CAFE	112,343.60	120,500.80	1.07	14.54	14.66
1	2007	BOLIVAR	CAFE	502.00	448.00	0.89	0.05	0.07
2	2007	BOYACA	CAFE	11,374.50	9,683.10	0.85	1.17	1.48
3	2007	CALDAS	CAFE	78,393.65	92,815.00	1.18	11.20	10.23
4	2007	CAQUETA	CAFE	2,295.00	2,134.00	0.93	0.26	0.30

In [43]:
```
produccion_df.tail()
# describe el final del Dataframe, por si es muy grande facilita ver solo la parte final
```

Out[43]:

	Anio	Departamento	Producto	Area (ha)	Produccion (ton)	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
261	2018	QUINDIO	CAFE	16,374.73	17,739.03	1.08	2.07	2.21
262	2018	RISARALDA	CAFE	35,874.73	45,918.75	1.28	5.37	4.83
263	2018	SANTANDER	CAFE	42,269.07	55,918.71	1.32	6.53	5.69
264	2018	TOLIMA	CAFE	97,304.04	97,451.31	1.00	11.39	13.11
265	2018	VALLE DEL CAUCA	CAFE	48,305.31	49,667.88	1.03	5.80	6.51

In [44]:
```
produccion_grouped_Anio=produccion_df.groupby(["Producto","Anio"]).describe()
produccion_grouped_Anio
```

Out[44]:

		Rendimiento (ha/ton)								Produccion Nacional (ton)		...			Area Nacional (ha)					
		count	mean	std	min	25%	50%	75%	max	count	mean	...	75%	max	count	mean	std	min	25%	50
Producto	Anio																			
CAFE	2007	22.0	0.950455	0.279566	0.45	0.7900	0.900	1.1525	1.53	22.0	4.545909	...	7.8475	15.57	22.0	4.545455	4.544143	0.00	0.4100	3.1
	2008	22.0	0.982727	0.322670	0.45	0.7775	0.905	1.2000	1.79	22.0	4.545455	...	7.7600	15.85	22.0	4.545000	4.529220	0.00	0.4200	3.2
	2009	22.0	0.881364	0.264652	0.30	0.7600	0.930	1.1125	1.21	22.0	4.545455	...	7.3425	14.76	22.0	4.544545	4.470076	0.00	0.3800	3.3
	2010	23.0	0.906087	0.324692	0.00	0.7050	0.960	1.1250	1.52	23.0	4.348261	...	7.3550	15.56	23.0	4.347826	4.497794	0.00	0.3250	3.0
	2011	23.0	0.854348	0.238305	0.47	0.6100	0.900	1.0550	1.20	23.0	4.348696	...	7.0800	18.00	23.0	4.347826	4.563870	0.00	0.3750	3.0
	2012	23.0	0.858696	0.329618	0.00	0.7450	0.830	0.9150	2.00	23.0	4.347391	...	6.9850	14.62	23.0	4.347826	4.599389	0.00	0.4000	2.9
	2013	22.0	0.759545	0.145421	0.60	0.6000	0.755	0.8800	0.99	22.0	4.545455	...	6.4400	17.77	22.0	4.545000	4.725951	0.00	0.4725	3.2
	2014	22.0	0.822273	0.157629	0.64	0.6500	0.815	0.9500	1.06	22.0	4.545455	...	6.5950	18.67	22.0	4.545455	4.778870	0.01	0.4825	3.1
	2015	22.0	1.024545	0.110096	0.77	0.9350	1.065	1.1075	1.15	22.0	4.544545	...	6.4675	17.07	22.0	4.545455	4.793782	0.02	0.4975	3.0
	2016	21.0	1.063810	0.118725	0.79	0.9600	1.120	1.1500	1.19	21.0	4.761429	...	6.6800	17.00	21.0	4.761905	4.802241	0.02	0.7100	3.2

In [49]:
```
produccion_Producto2=produccion_df.groupby(["Produccion (ton)","Producto", "Rendimiento (ha/ton)"]).describe()
produccion_Producto2
```

Out[49]:

			Anio								Produccion Nacional (ton)					Area Nacional (ha)			
			count	mean	std	min	25%	50%	75%	max	count	mean	...	75%	max	count	mean	std	
Produccion (ton)	Producto	Rendimiento (ha/ton)																	
0.00	CAFE	0.00	2.0	2011.0	1.414214	2010.0	2010.50	2011.0	2011.50	2012.0	2.0	0.000	...	0.0000	0.00	2.0	0.00	0.0	
1,089.74	CAFE	1.02	1.0	2015.0	NaN	2015.0	2015.00	2015.0	2015.00	2015.0	1.0	0.130	...	0.1300	0.13	1.0	0.13	NaN	
1,128.32	CAFE	1.06	1.0	2016.0	NaN	2016.0	2016.00	2016.0	2016.00	2016.0	1.0	0.130	...	0.1300	0.13	1.0	0.14	NaN	
1,338.56	CAFE	0.60	1.0	2013.0	NaN	2013.0	2013.00	2013.0	2013.00	2013.0	1.0	0.210	...	0.2100	0.21	1.0	0.29	NaN	
1,388.13	CAFE	0.65	1.0	2008.0	NaN	2008.0	2008.00	2008.0	2008.00	2008.0	1.0	0.170	...	0.1700	0.17	1.0	0.28	NaN	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
94,556.71	CAFE	0.98	1.0	2017.0	NaN	2017.0	2017.00	2017.0	2017.00	2017.0	1.0	11.100	...	11.1000	11.10	1.0	12.75	NaN	
95,957.90	CAFE	1.33	1.0	2010.0	NaN	2010.0	2010.00	2010.0	2010.00	2010.0	1.0	12.310	...	12.3100	12.31	1.0	9.71	NaN	
97,451.31	CAFE	1.00	1.0	2018.0	NaN	2018.0	2018.00	2018.0	2018.00	2018.0	1.0	11.390	...	11.3900	11.39	1.0	13.11	NaN	
97,922.49	CAFE	1.22	1.0	2017.0	NaN	2017.0	2017.00	2017.0	2017.00	2017.0	1.0	11.500	...	11.5000	11.50	1.0	10.66	NaN	
98.00	CAFE	0.62	2.0	2010.5	0.707107	2010.0	2010.25	2010.5	2010.75	2011.0	2.0	0.015	...	0.0175	0.02	2.0	0.02	0.0	

263 rows × 24 columns

In [50]:
```
produccion_Anio=produccion_df.groupby(["Anio"]).describe()
produccion_Anio
```

Out[50]:

	Rendimiento (ha/ton)								Produccion Nacional (ton)					Area Nacional (ha)					
	count	mean	std	min	25%	50%	75%	max	count	mean	...	75%	max	count	mean	std	min	25%	50%	75%
Anio																				
2007	22.0	0.950455	0.279566	0.45	0.7900	0.900	1.1525	1.53	22.0	4.545909	...	7.8475	15.57	22.0	4.545455	4.544143	0.00	0.4100	3.105	6.7900
2008	22.0	0.982727	0.322670	0.45	0.7775	0.905	1.2000	1.79	22.0	4.545455	...	7.7600	15.85	22.0	4.545000	4.529220	0.00	0.4200	3.215	7.1150
2009	22.0	0.881364	0.264652	0.30	0.7600	0.930	1.1125	1.21	22.0	4.545455	...	7.3425	14.76	22.0	4.544545	4.470076	0.00	0.3800	3.305	7.2575
2010	23.0	0.906087	0.324692	0.00	0.7050	0.960	1.1250	1.52	23.0	4.348261	...	7.3550	15.56	23.0	4.347826	4.497794	0.00	0.3250	3.020	6.8850
2011	23.0	0.854348	0.238305	0.47	0.6100	0.900	1.0550	1.20	23.0	4.348696	...	7.0800	18.00	23.0	4.347826	4.563870	0.00	0.3750	3.020	6.9450
2012	23.0	0.858696	0.329618	0.00	0.7450	0.830	0.9150	2.00	23.0	4.347391	...	6.9850	14.62	23.0	4.347826	4.599389	0.00	0.4000	2.970	7.0700
2013	22.0	0.759545	0.145421	0.60	0.6000	0.755	0.8800	0.99	22.0	4.545455	...	6.4400	17.77	22.0	4.545000	4.725951	0.00	0.4725	3.265	6.4800
2014	22.0	0.822273	0.157629	0.64	0.6500	0.815	0.9500	1.06	22.0	4.545455	...	6.5950	18.67	22.0	4.545455	4.778870	0.01	0.4825	3.135	6.5600

In [53]:
```
produccion_Anio_Produccion=produccion_df.groupby(["Produccion Nacional (ton)"]).describe()
produccion_Anio_Produccion
```

Out[53]:

	Anio								Rendimiento (ha/ton)					Area Nacional (ha)				
	count	mean	std	min	25%	50%	75%	max	count	mean	...	75%	max	count	mean	std	min	25%
Produccion Nacional (ton)																		
0.00	8.0	2010.000000	2.000000	2007.0	2008.75	2010.0	2011.25	2013.0	8.0	0.786250	...	1.1525	1.20	8.0	0.000000	0.000000	0.00	0.
0.01	7.0	2011.285714	2.563480	2008.0	2009.50	2011.0	2013.00	2015.0	7.0	0.932857	...	1.1400	1.15	7.0	0.012857	0.004880	0.01	0.
0.02	9.0	2013.666667	3.391165	2007.0	2012.00	2014.0	2016.00	2018.0	9.0	1.110000	...	1.2600	2.00	9.0	0.021111	0.007817	0.01	0.
0.03	2.0	2017.500000	0.707107	2017.0	2017.25	2017.5	2017.75	2018.0	2.0	1.365000	...	1.3725	1.38	2.0	0.030000	0.000000	0.03	0.
0.04	1.0	2009.000000	NaN	2009.0	2009.00	2009.0	2009.00	2009.0	1.0	0.380000	...	0.3800	0.38	1.0	0.100000	NaN	0.10	0.
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
17.00	1.0	2016.000000	NaN	2016.0	2016.00	2016.0	2016.00	2016.0	1.0	1.150000	...	1.1500	1.15	1.0	16.210000	NaN	16.21	16.
17.07	1.0	2015.000000	NaN	2015.0	2015.00	2015.0	2015.00	2015.0	1.0	1.110000	...	1.1100	1.11	1.0	16.280000	NaN	16.28	16.
17.77	1.0	2013.000000	NaN	2013.0	2013.00	2013.0	2013.00	2013.0	1.0	0.980000	...	0.9800	0.98	1.0	15.310000	NaN	15.31	15.
18.00	1.0	2011.000000	NaN	2011.0	2011.00	2011.0	2011.00	2011.0	1.0	1.080000	...	1.0800	1.08	1.0	14.940000	NaN	14.94	14.
18.67	1.0	2014.000000	NaN	2014.0	2014.00	2014.0	2014.00	2014.0	1.0	1.060000	...	1.0600	1.06	1.0	16.120000	NaN	16.12	16.

205 rows × 24 columns

In [54]:
```
produccion_df.describe()
# Indica datos estadísticos generales del dataframe produccion desde el año 2007
```

Out[54]:

	Anio	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
count	266.000000	266.000000	266.000000	266.000000
mean	2012.469925	0.936429	4.511316	4.511203
std	3.443484	0.267129	4.950568	4.565865
min	2007.000000	0.000000	0.000000	0.000000
25%	2010.000000	0.750000	0.352500	0.390000
50%	2012.000000	0.940000	2.720000	3.120000
75%	2015.000000	1.120000	7.147500	6.875000
max	2018.000000	2.000000	18.670000	16.430000

In [55]:
```
produccion_df["Rendimiento (ha/ton)"].describe()
# Indica datos estadísticos generales para la Producción nacional del dataframe produccion
```

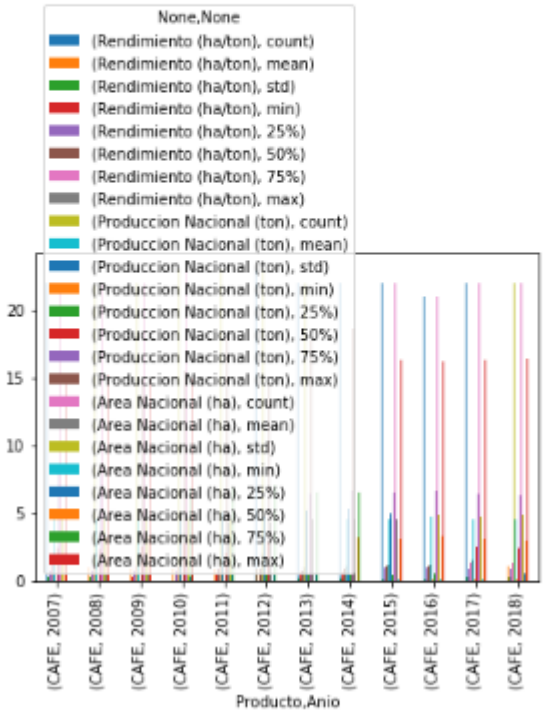
Out[55]:
```
count    266.000000
mean       0.936429
std        0.267129
min        0.000000
25%        0.750000
50%        0.940000
75%        1.120000
max        2.000000
Name: Rendimiento (ha/ton), dtype: float64
```

In [56]:
```
produccion_df.describe()
produccion_df.mean()
# Indica el promedio del dataframe produccion para Rendimiento, Produccion y el Área Nacional
```

Out[56]:
```
Anio                        2012.469925
Rendimiento (ha/ton)           0.936429
Produccion Nacional (ton)      4.511316
Area Nacional (ha)             4.511203
dtype: float64
```

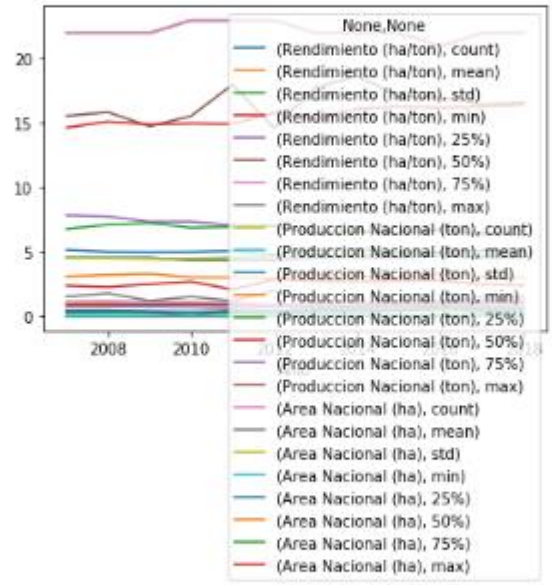
In [59]:
```
import numpy as np
import re
import sys
%matplotlib inline
produccion_grouped_Anio.plot(kind='bar')
```

Out[59]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x990eb48>
```



In [64]:
```
import numpy as np
import re
import sys
%matplotlib inline
produccion_Anio.plot(kind='line')
```

Out[64]:
```
<matplotlib.axes._subplots.AxesSubplot at 0xb1f4808>
```

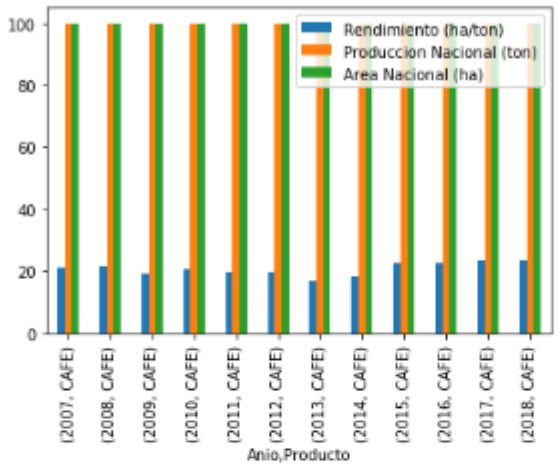


In [65]:
```
produccion_df.duplicated().sum()
#Registros que esten duplicados
```

Out[65]:
```
0
```

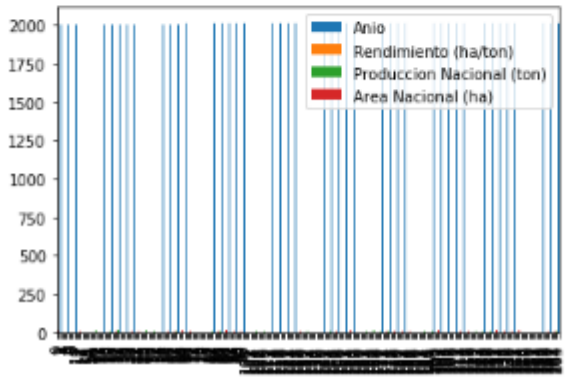
In [69]:
```
produccion_grouped_Producto.plot(kind='bar')
```

Out[69]:
```
<matplotlib.axes._subplots.AxesSubplot at 0xbe6ff08>
```



In [78]:
```
# Construcción del gráfico produccion por departamento año tipo lineas
%matplotlib inline
produccion_grouped_Departamento.plot(kind='bar')
```

Out[78]:
```
<matplotlib.axes._subplots.AxesSubplot at 0xcfa8208>
```



In [79]:
```
grouped_data = produccion_df.groupby("Producto")
d=grouped_data.describe().mean()
print (d)
```

```
Anio                        count     266.000000
                            mean     2012.469925
                            std         3.443484
                            min      2007.000000
                            25%      2010.000000
                            50%      2012.000000
                            75%      2015.000000
                            max      2018.000000
Rendimiento (ha/ton)        count     266.000000
                            mean        0.936429
                            std         0.267129
                            min         0.000000
                            25%         0.750000
                            50%         0.940000
                            75%         1.120000
                            max         2.000000
Produccion Nacional (ton)   count     266.000000
                            mean        4.511316
                            std         4.950568
                            min         0.000000
                            25%         0.352500
                            50%         2.720000
                            75%         7.147500
                            max        18.670000
Area Nacional (ha)          count     266.000000
                            mean        4.511203
                            std         4.565865
                            min         0.000000
                            25%         0.390000
                            50%         3.120000
                            75%         6.875000
                            max        16.430000
dtype: float64
```

In [81]:
```
Departamento_Choco=produccion_df.loc[produccion_df["Departamento"]=="CHOCO"]
print (Departamento_Choco)
# Indica los resultados estadísticos por año para el Departamento Seleccionado
```

```
     Anio Departamento Producto Area (ha) Produccion (ton)  \
8    2007        CHOCO     CAFE    290.00           205.90
30   2008        CHOCO     CAFE     90.00            68.00
52   2009        CHOCO     CAFE     70.00            78.75
75   2010        CHOCO     CAFE    157.50            98.00
98   2011        CHOCO     CAFE    157.50            98.00
120  2012        CHOCO     CAFE     70.00           140.00
143  2013        CHOCO     CAFE    125.01           105.93
165  2014        CHOCO     CAFE    136.88           125.42
187  2015        CHOCO     CAFE    137.47           158.20
209  2016        CHOCO     CAFE    134.96           160.62
230  2017        CHOCO     CAFE    125.67           158.85
252  2018        CHOCO     CAFE    140.33           181.42

     Rendimiento (ha/ton)  Produccion Nacional (ton)  Area Nacional (ha)
8                    0.71                       0.02                0.04
30                   0.76                       0.01                0.01
52                   1.13                       0.01                0.01
75                   0.62                       0.01                0.02
98                   0.62                       0.02                0.02
120                  2.00                       0.02                0.01
143                  0.85                       0.02                0.02
165                  0.92                       0.02                0.02
187                  1.15                       0.02                0.02
209                  1.19                       0.02                0.02
230                  1.26                       0.02                0.02
252                  1.29                       0.02                0.02
```

In [82]:
```
Departamento_Antioquia=produccion_df.loc[produccion_df["Departamento"]=="ANTIOQUIA"]
print (Departamento_Antioquia)
# Indica los resultados estadísticos por año para el Departamento Seleccionado
```

```
     Anio Departamento Producto   Area (ha) Produccion (ton)  \
0    2007    ANTIOQUIA     CAFE  112,343.60       120,500.80
22   2008    ANTIOQUIA     CAFE  114,694.00       113,505.20
44   2009    ANTIOQUIA     CAFE  112,420.20       103,703.00
66   2010    ANTIOQUIA     CAFE  111,602.71       121,253.38
89   2011    ANTIOQUIA     CAFE  106,419.57       115,267.98
112  2012    ANTIOQUIA     CAFE  112,221.14        91,621.30
135  2013    ANTIOQUIA     CAFE  109,755.50       102,403.24
157  2014    ANTIOQUIA     CAFE  110,115.86       111,452.91
```

In [88]:
```
Estadística_Anio2012=produccion_df.loc[produccion_df["Anio"]== 2012]
print (Estadística_Anio2012)
# Indica los resultados estadísticos por departamento para el año 2012
```

	Anio	Departamento	Producto	Area (ha)	Produccion (ton)	\
112	2012	ANTIOQUIA	CAFE	112,221.14	91,621.30	
113	2012	BOLIVAR	CAFE	870.00	652.50	
114	2012	BOYACA	CAFE	6,698.20	4,981.59	
115	2012	CALDAS	CAFE	54,871.88	54,115.96	
116	2012	CAQUETA	CAFE	2,882.50	2,446.38	
117	2012	CASANARE	CAFE	2,322.00	1,718.25	
118	2012	CAUCA	CAFE	56,825.00	50,588.14	
119	2012	CESAR	CAFE	22,911.00	19,994.35	
120	2012	CHOCO	CAFE	70.00	140.00	
121	2012	CUNDINAMARCA	CAFE	37,175.06	30,786.41	
122	2012	GUAVIARE	CAFE	0.00	0.00	
123	2012	HUILA	CAFE	79,809.34	85,212.64	
124	2012	LA GUAJIRA	CAFE	5,143.00	3,434.30	
125	2012	MAGDALENA	CAFE	17,686.00	14,096.05	
126	2012	META	CAFE	2,783.00	2,133.10	
127	2012	NARIÑO	CAFE	27,806.40	28,077.94	
128	2012	NORTE DE SANTANDER	CAFE	19,339.31	12,214.54	
129	2012	PUTUMAYO	CAFE	42.00	48.40	
130	2012	QUINDIO	CAFE	21,109.83	18,030.13	
131	2012	RISARALDA	CAFE	45,588.03	36,989.43	
132	2012	SANTANDER	CAFE	33,947.15	23,271.89	
133	2012	TOLIMA	CAFE	90,904.48	85,027.49	
134	2012	VALLE DEL CAUCA	CAFE	69,456.71	61,190.55	

	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
112	0.82	14.62	15.80
113	0.75	0.10	0.12
114	0.74	0.79	0.94
115	0.99	8.63	7.72
116	0.85	0.39	0.41
117	0.74	0.27	0.33
118	0.89	8.07	8.00
119	0.87	3.19	3.22
120	2.00	0.02	0.01

In [89]:
```
produccion_df[0:25]
#Lista los primeros 25 elementos del dataframe
```

Out[89]:

	Anio	Departamento	Producto	Area (ha)	Produccion (ton)	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
0	2007	ANTIOQUIA	CAFE	112,343.60	120,500.80	1.07	14.54	14.66
1	2007	BOLIVAR	CAFE	502.00	446.00	0.89	0.05	0.07
2	2007	BOYACA	CAFE	11,374.50	9,683.10	0.85	1.17	1.48
3	2007	CALDAS	CAFE	78,393.65	92,815.00	1.18	11.20	10.23
4	2007	CAQUETA	CAFE	2,295.00	2,134.00	0.93	0.26	0.30
5	2007	CASANARE	CAFE	2,605.00	2,048.40	0.79	0.25	0.34
6	2007	CAUCA	CAFE	53,471.00	51,348.00	0.96	6.19	6.98
7	2007	CESAR	CAFE	23,172.00	13,278.50	0.57	1.60	3.02
8	2007	CHOCO	CAFE	290.00	205.90	0.71	0.02	0.04
9	2007	CUNDINAMARCA	CAFE	43,017.30	33,729.14	0.78	4.07	5.61
10	2007	HUILA	CAFE	89,661.56	129,052.51	1.44	15.57	11.70
11	2007	LA GUAJIRA	CAFE	4,785.00	2,958.70	0.62	0.36	0.62
12	2007	MAGDALENA	CAFE	17,506.00	14,005.00	0.80	1.69	2.28
13	2007	META	CAFE	2,048.00	1,617.20	0.79	0.20	0.27
14	2007	NARIÑO	CAFE	24,458.50	31,770.05	1.30	3.83	3.19
15	2007	NORTE DE SANTANDER	CAFE	30,171.84	13,593.24	0.45	1.64	3.94
16	2007	PUTUMAYO	CAFE	35.00	34.00	0.97	0.00	0.00
17	2007	QUINDIO	CAFE	19,904.00	25,426.00	1.28	3.07	2.60
18	2007	RISARALDA	CAFE	47,689.25	72,842.55	1.53	8.79	6.22
19	2007	SANTANDER	CAFE	34,406.67	29,469.52	0.86	3.56	4.49
20	2007	TOLIMA	CAFE	91,679.10	112,322.38	1.23	13.55	11.96
21	2007	VALLE DEL CAUCA	CAFE	76,667.80	69,618.24	0.91	8.40	10.00
22	2008	ANTIOQUIA	CAFE	114,694.00	113,505.20	0.99	13.70	15.13
23	2008	BOLIVAR	CAFE	572.00	711.00	1.24	0.09	0.08
24	2008	BOYACA	CAFE	10,778.50	9,547.30	0.89	1.15	1.42

In [91]:
```
produccion_df[150:210]
#Lista los elementos desde el 150 al 210 del dataframe
```

Out[91]:

	Anio	Departamento	Producto	Area (ha)	Produccion (ton)	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
150	2013	NORTE DE SANTANDER	CAFE	25,332.45	15,185.79	0.60	2.33	3.28
151	2013	PUTUMAYO	CAFE	24.27	16.87	0.70	0.00	0.00
152	2013	QUINDIO	CAFE	21,203.03	20,599.27	0.97	3.16	2.75
153	2013	RISARALDA	CAFE	39,615.60	39,073.92	0.99	5.99	5.13
154	2013	SANTANDER	CAFE	38,613.68	30,227.02	0.78	4.64	5.00
155	2013	TOLIMA	CAFE	97,308.81	77,215.36	0.79	11.84	12.61
156	2013	VALLE DEL CAUCA	CAFE	53,481.02	42,948.40	0.80	6.59	6.93
157	2014	ANTIOQUIA	CAFE	110,115.86	111,452.91	1.01	15.30	13.84
158	2014	BOLIVAR	CAFE	936.34	606.93	0.65	0.08	0.12
159	2014	BOYACA	CAFE	9,834.39	6,364.41	0.65	0.87	1.24
160	2014	CALDAS	CAFE	59,757.18	62,869.38	1.05	8.63	7.51
161	2014	CAQUETA	CAFE	3,074.92	2,503.81	0.81	0.34	0.39
162	2014	CASANARE	CAFE	2,599.43	1,688.60	0.65	0.23	0.33
163	2014	CAUCA	CAFE	77,068.46	63,365.76	0.82	8.70	9.69
164	2014	CESAR	CAFE	26,138.58	16,935.63	0.65	2.33	3.29
165	2014	CHOCO	CAFE	136.88	125.42	0.92	0.02	0.02
166	2014	CUNDINAMARCA	CAFE	33,623.54	25,118.55	0.75	3.45	4.23
167	2014	HUILA	CAFE	128,273.15	135,971.20	1.06	18.67	16.12
168	2014	LA GUAJIRA	CAFE	6,078.64	3,923.80	0.65	0.54	0.76
169	2014	MAGDALENA	CAFE	18,533.11	12,012.98	0.65	1.65	2.33
170	2014	META	CAFE	2,739.71	1,950.84	0.71	0.27	0.34

In [99]:
```
produccion_df["Rendimiento (ha/ton)"].describe()
# Indica datos estadísticos generales para el Rendimiento del dataframe produccion
```

Out[99]:
```
count    266.000000
mean       0.936429
std        0.267129
min        0.000000
25%        0.750000
50%        0.940000
75%        1.120000
max        2.000000
Name: Rendimiento (ha/ton), dtype: float64
```

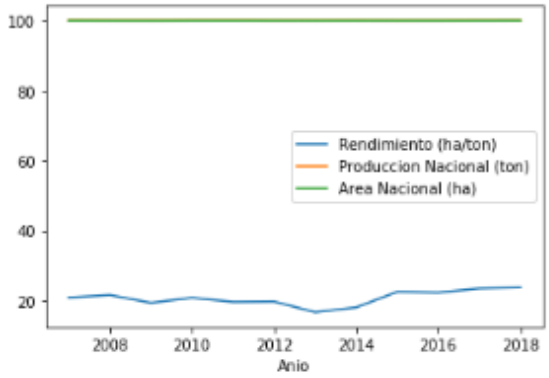
In [101]:
```
produccion_grouped_Anio5=produccion_df.groupby("Anio").sum()
produccion_grouped_Anio
```

Out[101]:

Anio	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
2007	20.91	100.01	100.00
2008	21.62	100.00	99.99
2009	19.39	100.00	99.98
2010	20.84	100.01	100.00
2011	19.65	100.02	100.00
2012	19.75	99.99	100.00
2013	16.71	100.00	99.99
2014	18.09	100.00	100.00
2015	22.54	99.98	100.00
2016	22.34	99.99	100.00
2017	23.50	100.01	100.00
2018	23.75	100.00	100.02

In [103]:
```
import numpy as np
import re
import sys
%matplotlib inline
produccion_grouped_Anio5.plot(kind='line')
```

Out[103]:
```
<matplotlib.axes._subplots.AxesSubplot at 0xefc5688>
```



In [109]:
```
produccion_grouped_Anio6=produccion_df.groupby("Anio")["Departamento"].sum()
produccion_grouped_Anio6
```

Out[109]:
```
Anio
2007    ANTIOQUIABOLIVARBOYACACALDASCAQUETACASANARECAU...
2008    ANTIOQUIABOLIVARBOYACACALDASCAQUETACASANARECAU...
2009    ANTIOQUIABOLIVARBOYACACALDASCAQUETACASANARECAU...
2010    ANTIOQUIAARAUCABOLIVARBOYACACALDASCAQUETACASAN...
2011    ANTIOQUIAARAUCABOLIVARBOYACACALDASCAQUETACASAN...
2012    ANTIOQUIABOLIVARBOYACACALDASCAQUETACASANARECAU...
2013    ANTIOQUIABOLIVARBOYACACALDASCAQUETACASANARECAU...
2014    ANTIOQUIABOLIVARBOYACACALDASCAQUETACASANARECAU...
2015    ANTIOQUIABOLIVARBOYACACALDASCAQUETACASANARECAU...
2016    ANTIOQUIABOLIVARBOYACACALDASCAQUETACASANARECAU...
2017    ANTIOQUIABOLIVARBOYACACALDASCAQUETACASANARECAU...
2018    ANTIOQUIABOLIVARBOYACACALDASCAQUETACASANARECAU...
Name: Departamento, dtype: object
```

In [114]:
```
produccion_grouped_Rendimiento4=produccion_df.groupby("Rendimiento (ha/ton)").sum()
produccion_grouped_Rendimiento4
```

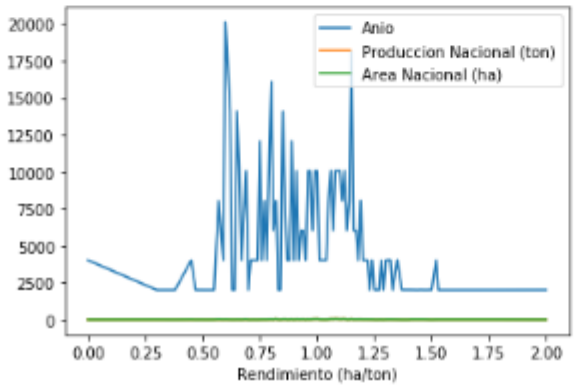
Out[114]:

Rendimiento (ha/ton)	Anio	Produccion Nacional (ton)	Area Nacional (ha)
0.00	4022	0.00	0.00
0.30	2009	1.44	4.45
0.38	2009	0.04	0.10
0.45	4015	3.28	7.92
0.47	2011	0.30	0.58
...	...	...	...
1.50	2017	0.60	0.45
1.52	4028	9.87	6.83
1.53	2007	8.79	6.22
1.79	2008	9.44	5.75
2.00	2012	0.02	0.01

94 rows × 3 columns

In [117]:
```
import numpy as np
import re
import sys
%matplotlib inline
produccion_grouped_Rendimiento4.plot(kind='line')
```

Out[117]: <matplotlib.axes._subplots.AxesSubplot at 0xef83e48>



In [123]:
```
Grupos_Departamentos=produccion_df.groupby("Anio")["Departamento"].count()
print (Grupos_Departamentos)
# Indica la cantidad dedepartamentos incluidos o analizados en cada uno de los años
```

```
Anio
2007    22
2008    22
2009    22
2010    23
2011    23
2012    23
2013    22
2014    22
2015    22
2016    21
2017    22
2018    22
Name: Departamento, dtype: int64
```

In [124]:
```
departamentos_counts = produccion_df.groupby("Departamento")["Produccion (ton)"].count()
print(departamentos_counts)
# Permite verificar y contar para cada uno de los Departamentos las distintas var
# Se encuentra que algunos departamentos tienen otros valores diferentes a los 12
```

```
Departamento
ANTIOQUIA             12
ARAUCA                 2
BOLIVAR               12
BOYACA                12
CALDAS                12
CAQUETA               12
CASANARE              12
CAUCA                 12
CESAR                 12
CHOCO                 12
CUNDINAMARCA          12
GUAVIARE               1
HUILA                 12
LA GUAJIRA            12
MAGDALENA             12
META                  12
NARIÑO                12
NORTE DE SANTANDER    12
PUTUMAYO              11
QUINDIO               12
RISARALDA             12
SANTANDER             12
TOLIMA                12
VALLE DEL CAUCA       12
Name: Produccion (ton), dtype: int64
```

In [130]:
```
pd.isnull(produccion_df)
```

Out[130]:

	Anio	Departamento	Producto	Area (ha)	Produccion (ton)	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...
261	False	False	False	False	False	False	False	False
262	False	False	False	False	False	False	False	False
263	False	False	False	False	False	False	False	False
264	False	False	False	False	False	False	False	False
265	False	False	False	False	False	False	False	False

266 rows × 8 columns

In [133]:
```
produccion_df['Area Nacional (ha)'].mean()
# indica el promedio de la Area Nacional
```

Out[133]:
```
4.511203007518795
```

In [134]:
```
produccion_df.groupby('Departamento')['Departamento'].count()['CAUCA']
# agrupa los datos por Departamento y cuenta las ciudades que sean igual a cauca
```

Out[134]:
```
12
```

In [137]:
```
produccion_df.groupby('Departamento')['Departamento'].count()
# agrupa los datos por Edad y describe la cantidad de cada Edad
```

Out[137]:
```
Departamento
ANTIOQUIA          12
ARAUCA              2
BOLIVAR            12
BOYACA             12
CALDAS             12
CAQUETA            12
CASANARE           12
CAUCA              12
CESAR              12
CHOCO              12
CUNDINAMARCA       12
GUAVIARE            1
```

In [139]:
```
produccion_df.groupby('Departamento')['Departamento'].count()[12]
# agrupa los datos por Departamento y cuenta las Departamentos que sean igual 12
```

Out[139]: 12

In [141]:
```
produccion_df['Produccion Nacional (ton)'].std()
#describe la desviación estándar de la produccion
```

Out[141]: 4.950567735489969

In [142]:
```
produccion_df['Rendimiento (ha/ton)'].std()
#describe la desviación estándar del rendimiento
```

Out[142]: 0.26712944458019805

In [143]:
```
# Datos agrupados por sexo
grouped_data = produccion_df.groupby('Departamento')
```

In [144]:
```
# Estadísticas para todas las columnas numéricas por sexo
grouped_data.describe()
# Regresa la media de cada columna numérica por sexo
grouped_data.mean()
```

Out[144]:

Departamento	Anio	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
ANTIOQUIA	2012.500000	1.084167	15.276667	14.345833
ARAUCA	2010.500000	0.600000	0.000000	0.000000
BOLIVAR	2012.500000	0.755833	0.084167	0.115833
BOYACA	2012.500000	0.784167	0.990833	1.277500
CALDAS	2012.500000	1.166667	9.655833	8.338333
CAQUETA	2012.500000	1.045000	0.402500	0.389167
CASANARE	2012.500000	0.834167	0.257500	0.310833
CAUCA	2012.500000	0.946667	8.322500	8.819167
CESAR	2012.500000	0.662500	2.107500	3.181667
CHOCO	2012.500000	1.041667	0.017500	0.019167
CUNDINAMARCA	2012.500000	0.956667	4.665000	4.919167
GUAVIARE	2012.000000	0.000000	0.000000	0.000000
HUILA	2012.500000	1.165000	15.716667	13.796667
LA GUAJIRA	2012.500000	0.620833	0.415000	0.663333
MAGDALENA	2012.500000	0.767500	1.764167	2.314167

Pandas profiling

In [176]:
```
# USO DE PANDAS PROFILING
# Instructor Ing. Luis Armando Amaya Q.
import pandas as pd
import numpy as np
from pandas_profiling import ProfileReport
profile=ProfileReport(produccion_df, title='CAFE', html ={'style': {'full_width':
profile
#NOTA IMPORTANTE
# LA DOS SIGUIENTES INSTRUCCIONES, CREAN UN INFORME EN FORMATO HTML
# DEBE BUSCARLO EN SU COMPUTADOR CON EL NOMBRE:---> ANALISIS EXPLORATORIO CADE_PA
# LUEGO DE ENCONTRAR LA CARPETA ---> Producción_Cafe  <------------------
# PARA ABRIR EL INFORME DEBE HACER CLIC SOBRE EL ARCHIVO LLAMADO----------->your
# RECUERDE: -----> LA DOS SIGUIENTES INSTRUCCIONES, CREAN UN INFORME EN FORMATO H
# TAMBIÉN LE SUBÍ TODA LA CARPETA AL DRIVE CON TODOS ESTOS INFORMES, LLAMADA ----
#profile2=profile
#profile2.to_file("ANALISIS EXPLORATORIO CAFE_PANDAS.html")
```

PRODUCCION DE CAFE EN COLOMBIA

Overview | Variables | Interactions | Correlations | Missing values | Sample

Overview

Overview | Warnings 9 | Reproduction

Dataset statistics

Number of variables	8
Number of observations	266
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	16.8 KiB
Average record size in memory	64.5 B

Variable types

CAT	4
NUM	4

PRODUCCION DE CAFE EN COLOMBIA

Overview | Variables | Interactions | Correlations | Missing values | Sample

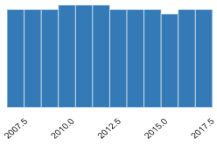
Variables

Anio

Real number (ℝ≥0)

Distinct	12
Distinct (%)	4.5%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%

Mean	2012.469925
Minimum	2007
Maximum	2018
Zeros	0
Zeros (%)	0.0%
Memory size	2.1 KiB



Toggle details

Departamento

Categorical

Distinct	24
Distinct (%)	9.0%
Missing	0
Missing (%)	0.0%
Memory size	2.1 KiB

PRODUCCION DE CAFE EN COLOMBIA

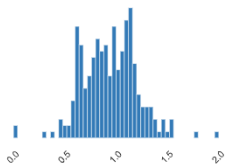
Overview | Variables | Interactions | Correlations | Missing values | Sample

Rendimiento (ha/ton)

Real number (ℝ≥0)

Distinct	94
Distinct (%)	35.3%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%

Mean	0.9364285714
Minimum	0
Maximum	2
Zeros	2
Zeros (%)	0.8%
Memory size	2.1 KiB



Toggle details

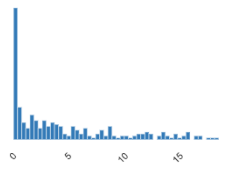
Produccion Nacional (ton)

Real number (ℝ≥0)

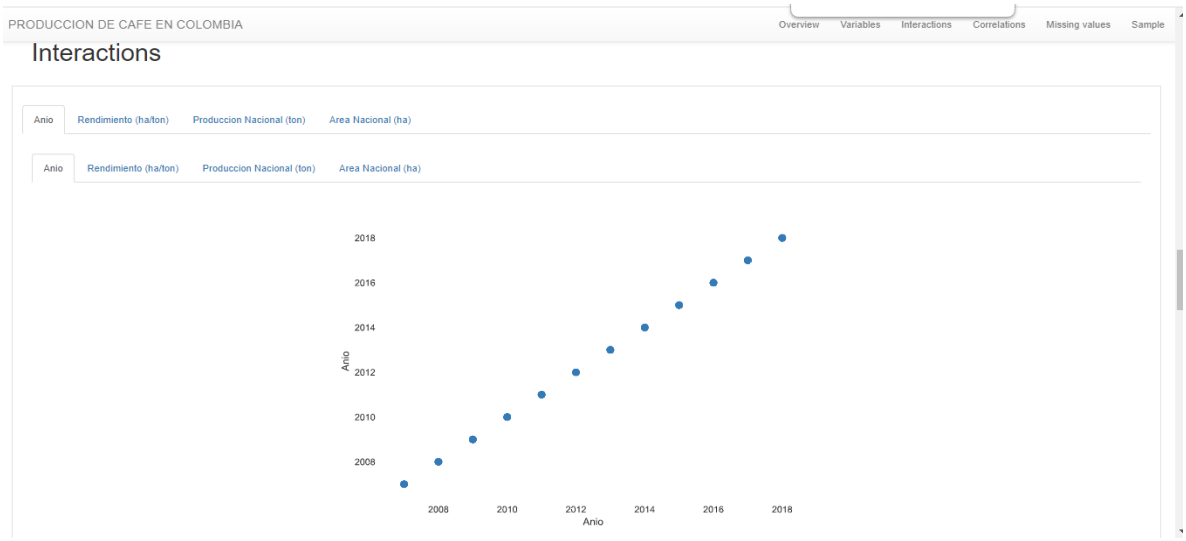
HIGH CORRELATION
ZEROS

Distinct	205
Distinct (%)	77.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%

Mean	4.511315789
Minimum	0
Maximum	18.67
Zeros	8
Zeros (%)	3.0%
Memory size	2.1 KiB



Toggle details



Modelo predictivo

In [164]:
```
produccion_df.describe()
# Información estadístico del Dataframe para las variables
```

Out[164]:

	Anio	Rendimiento (ha/ton)	Produoolon Nacional (ton)	Area Nacional (ha)
ocunt	266.000000	266.000000	266.000000	266.000000
mean	2012.469925	0.936429	4.511316	4.511203
std	3.443484	0.267129	4.950568	4.565865
min	2007.000000	0.000000	0.000000	0.000000
26%	2010.000000	0.750000	0.352500	0.390000
60%	2012.000000	0.940000	2.720000	3.120000
76%	2015.000000	1.120000	7.147500	6.875000
max	2018.000000	2.000000	18.670000	16.430000

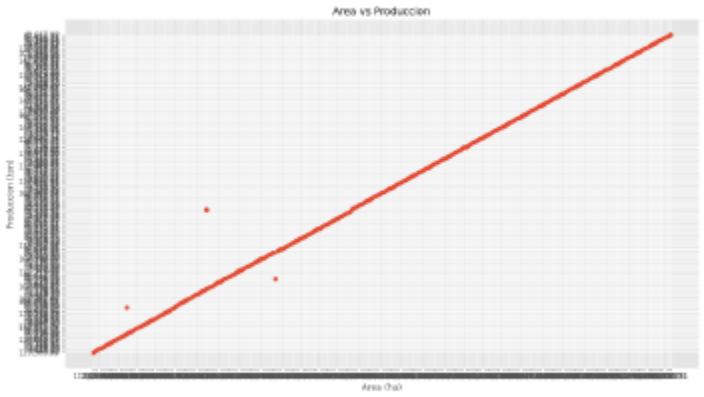
```
# estas instrucciones aun no las voy a emplear, por eso están con el simbolo #
#arreglo=list(produccion_df.columns)
#produccion1_df= produccion_df[arreglo[2:len(arreglo)]]
#produccion1_df
#print(produccion1_df)
#arreglo2.describe()
```

In [165]:
```
# Obtenemos información de los tipos de las variables del Dataframe o DataSet
produccion_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 266 entries, 0 to 265
Data columns (total 8 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Anio                        266 non-null    int64
 1   Departamento                266 non-null    object
 2   Producto                    266 non-null    object
 3   Area (ha)                   266 non-null    object
 4   Produccion (ton)            266 non-null    object
 5   Rendimiento (ha/ton)        266 non-null    float64
 6   Produccion Nacional (ton)   266 non-null    float64
 7   Area Nacional (ha)          266 non-null    float64
dtypes: float64(3), int64(1), object(4)
memory usage: 16.8+ KB
```

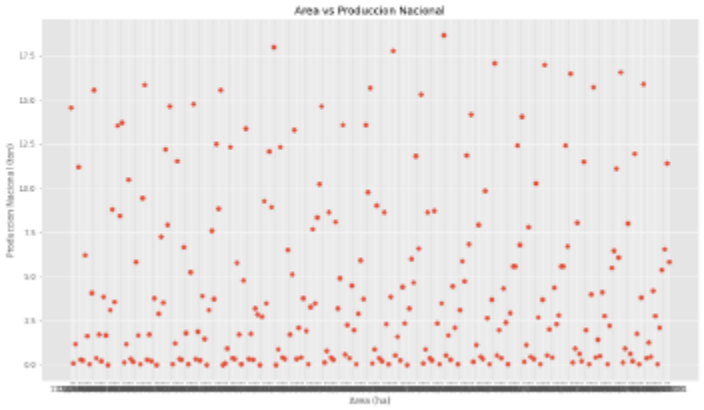
In [47]:
```
# Gráfico del comportamiento del Area versus Produccion.
plt.scatter(produccion_df['Area (ha)'],produccion_df['Produccion (ton)'])
plt.title('Area vs Produccion')
plt.xlabel('Area (ha)')
plt.ylabel("Produccion (ton)")
```

Out[47]: Text(0, 0.5, 'Produccion (ton)')



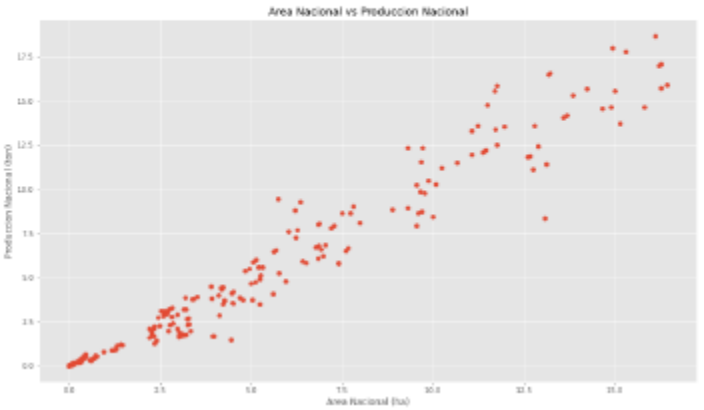
In [40]:
```
# Gráfico del comportamiento del Area versus Produccion Nacional
plt.scatter(produccion_df['Area (ha)'],produccion_df['Produccion Nacional (ton)']
plt.title('Area vs Produccion Nacional')
plt.xlabel('Area (ha)')
plt.ylabel("Produccion Nacional (ton)")
```

Out[40]:
```
Text(0, 0.5, 'Produccion Nacional (ton)')
```



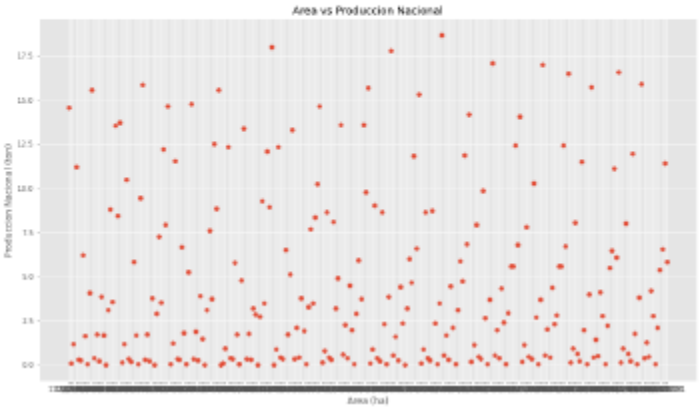
In [49]:
```
# Gráfico del comportamiento del Area Nacional versus Produccion Nacional
plt.scatter(produccion_df['Area Nacional (ha)'],produccion_df['Produccion Nacione
plt.title('Area Nacional vs Produccion Nacional')
plt.xlabel('Area Nacional (ha)')
plt.ylabel("Produccion Nacional (ton)")
```

Out[49]:
```
Text(0, 0.5, 'Produccion Nacional (ton)')
```



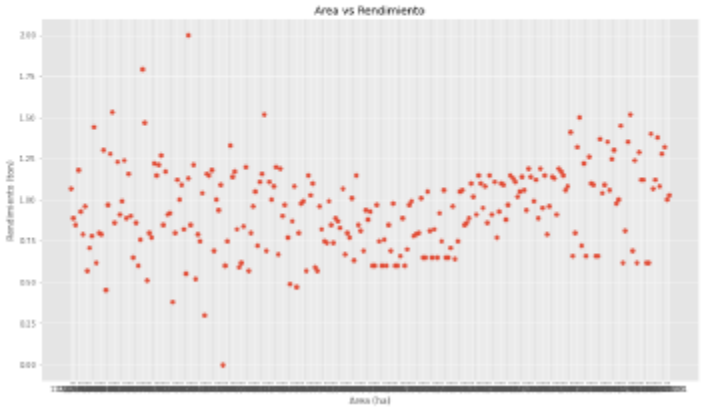
In [41]:
```
# Gráfico del comportamiento del Area versus Produccion Nacional
plt.scatter(produccion_df['Area (ha)'],produccion_df['Produccion Nacional (ton)'])
plt.title('Area vs Produccion Nacional')
plt.xlabel('Area (ha)')
plt.ylabel("Produccion Nacional (ton)")
```

Out[41]:
```
Text(0, 0.5, 'Produccion Nacional (ton)')
```



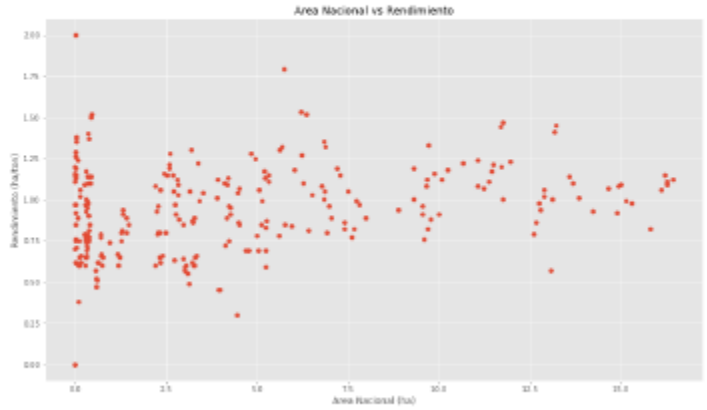
In [43]:
```
# Gráfico del comportamiento del Area versus Rendimiento
plt.scatter(produccion_df['Area (ha)'],produccion_df['Rendimiento (ha/ton)'])
plt.title('Area vs Rendimiento')
plt.xlabel('Area (ha)')
plt.ylabel("Rendimiento (ton)")
```

Out[43]:
```
Text(0, 0.5, 'Rendimiento (ton)')
```



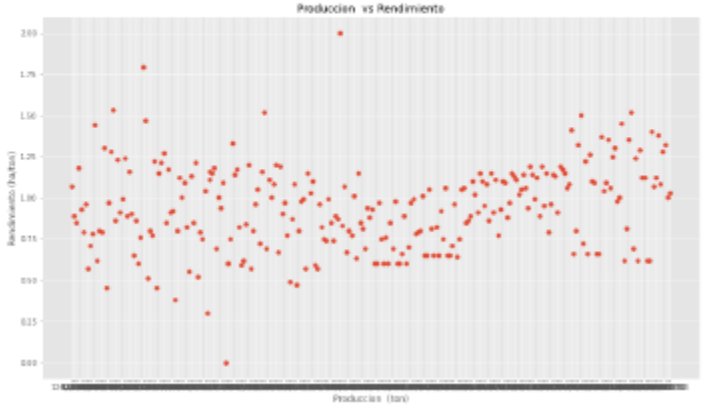
In [44]:
```
# Gráfico del comportamiento del Area Nacional versus Rendimiento
plt.scatter(produccion_df['Area Nacional (ha)'],produccion_df['Rendimiento (ha/to
plt.title('Area Nacional vs Rendimiento')
plt.xlabel('Area Nacional (ha)')
plt.ylabel("Rendimiento (ha/ton)")
```

Out[44]:
```
Text(0, 0.5, 'Rendimiento (ha/ton)')
```



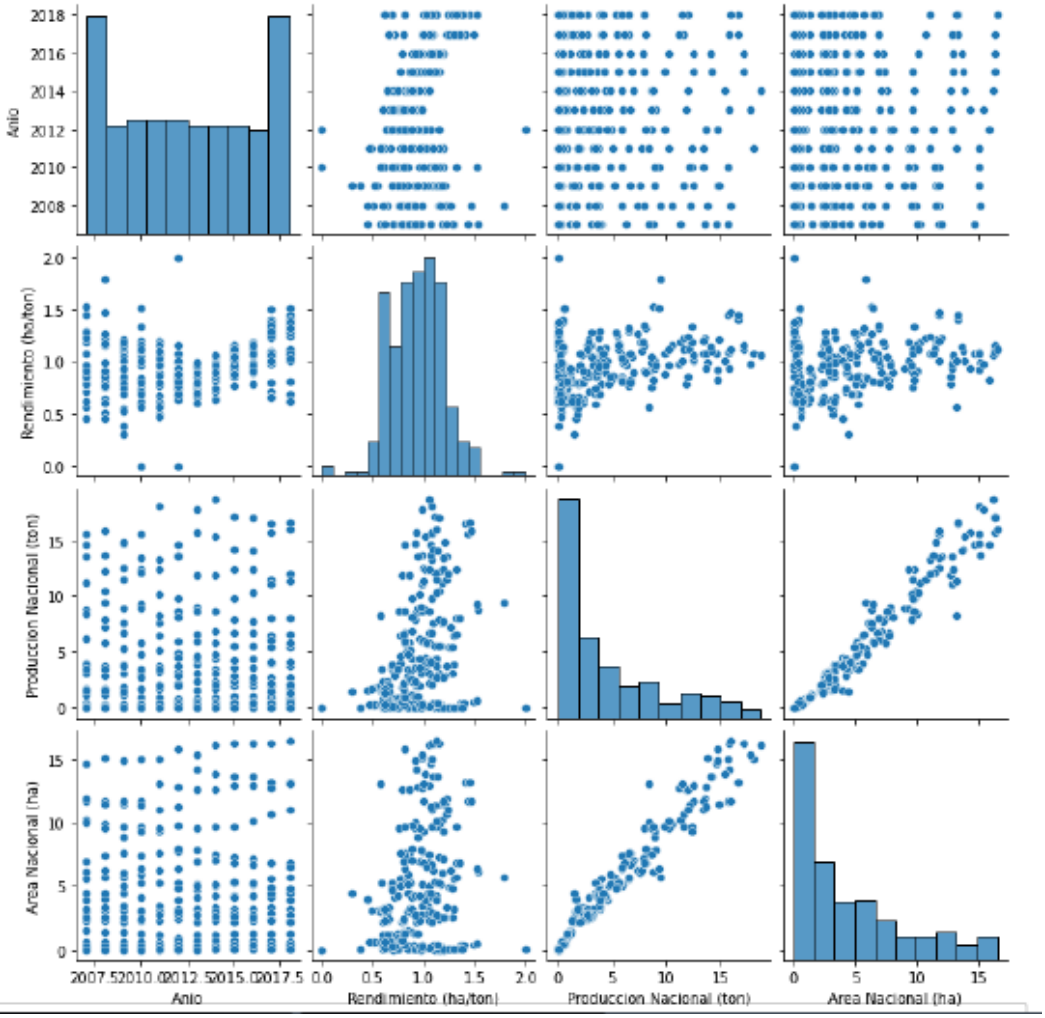
In [45]:
```
# Gráfico del comportamiento de la Produccion versus Rendimiento
plt.scatter(produccion_df['Produccion (ton)'], produccion_df['Rendimiento (ha/tor
plt.title('Produccion  vs Rendimiento')
plt.xlabel('Produccion  (ton)')
plt.ylabel("Rendimiento (ha/ton)")
```

Out[45]:
```
Text(0, 0.5, 'Rendimiento (ha/ton)')
```



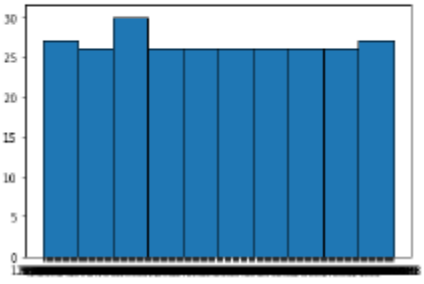
```
[20]: import seaborn as sns #Esta libreria permite construir gráficos muy
      particulares sns.pairplot(produccion_df)
```

```
Out[20]: <seaborn.axisgrid.PairGrid at 0x243b3938d60>
```



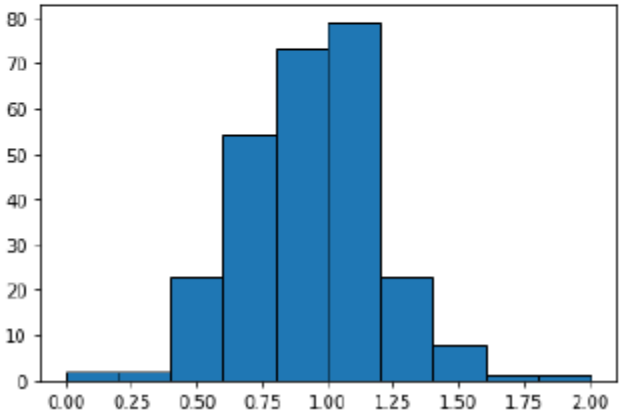

```
In [170]: # Histograma de La Produccion de Café
          plt.hist(produccion_df['Produccion (ton)'], edgecolor='black', linewidth=1)
```

```
Out[170]: (array([27., 26., 30., 26., 26., 26., 26., 26., 26., 27.]),
           array([  0. ,  26.1,  52.2,  78.3, 104.4, 130.5, 156.6, 182.7, 208.8,
                  234.9, 261. ]),
           <a list of 10 Patch objects>)
```

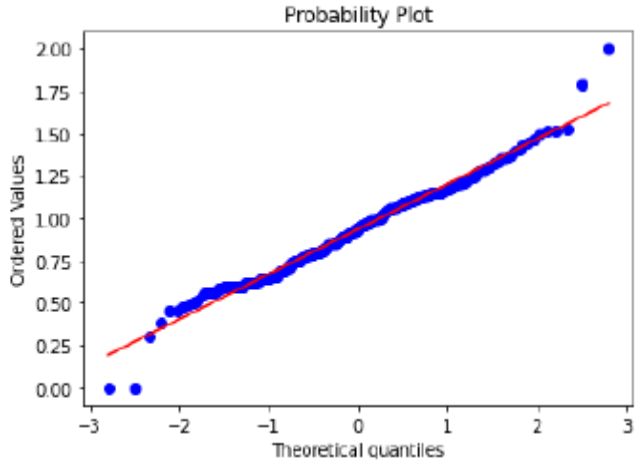


In [22]:
```
# Histograma del rendimiento del Café
plt.hist(produccion_df['Rendimiento (ha/ton)'], edgecolor='black',
linewidth=1)
```

Out[22]:
```
(array([ 2.,  2., 23., 54., 73., 79., 23.,  8.,  1.,  1.]),
array([0. , 0.2, 0.4, 0.6, 0.8, 1. , 1.2, 1.4, 1.6, 1.8, 2.
]),
 <a list of 10 Patch objects>)
```



[23]:
```
# Además, para corroborar la anterior distribucion normal, podemos construir
# el gráfico QUANTILE-QUANTILE NORMAL
# si los puntos están muy cerca a la linea recta, indica que los valores tienen u
import pylab
import scipy.stats as stats #librerias para construir estos tipos de graficos
stats.probplot(produccion_df['Rendimiento (ha/ton)'], dist= 'norm', plot = pylab)
pylab.show()
```



In [24]:
```
# importar la liberia shapiro para realizar el TEST DE SHAPIRO WILK,
# el test de Shapiro Wilk CONFIRMA EFECTIVAMENTE la correlacion entre
las variabl from scipy.stats import shapiro
estadistico,p_value =shapiro(produccion_df['Rendimiento (ha/ton)'])
print('Estadística=%.3f, El Valor de: p_value=%.3f' %
(estadistico,p_value))
# Si el valor entregado en la variable P_VALUE es MENOR a 0.05,
# indica que si existe una distribucion normal y correlacion
# entre las variables
```

```
Estadística=0.983, El Valor de: p_value=0.003
```

[25]:
```
# valores correlacion de
Spearman import numpy as np
produccion_correlacion_spearman = produccion_df.corr(method='spearman')
produccion_correlacion_spearman
# los valores del COEFICIENTE DE SPEARMAN cercanos a cero o inferiores
a (+-)(0.4
# indica que las variables no tienen correlacion
# los valores del COEFICIENTE DE SPEARMAN mayores a (+-)(0.4)
```

	Nacional Anio	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area (ha)
Anio	1.000000	0.180205	0.037725	0.023246
Rendimiento (ha/ton)	0.180205	1.000000	0.366952	0.264041
Produccion Nacional (ton)	0.037725	0.366952	1.000000	0.986380
Area Nacional (ha)	0.023246	0.264041	0.986380	1.000000

In [26]:
```
# valores correlacion de
Pearson import numpy as np
produccion_correlacion_pearson = produccion_df.corr(method='pearson')
produccion_correlacion_pearson
```

Out[26]:

	Nacional Anio	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area (ha)
Anio	1.000000	0.173474	0.007957	0.008715
Rendimiento (ha/ton)	0.173474	1.000000	0.385570	0.280677
Produccion Nacional (ton)	0.007957	0.385570	1.000000	0.978409
Area Nacional (ha)	0.008715	0.280677	0.978409	1.000000

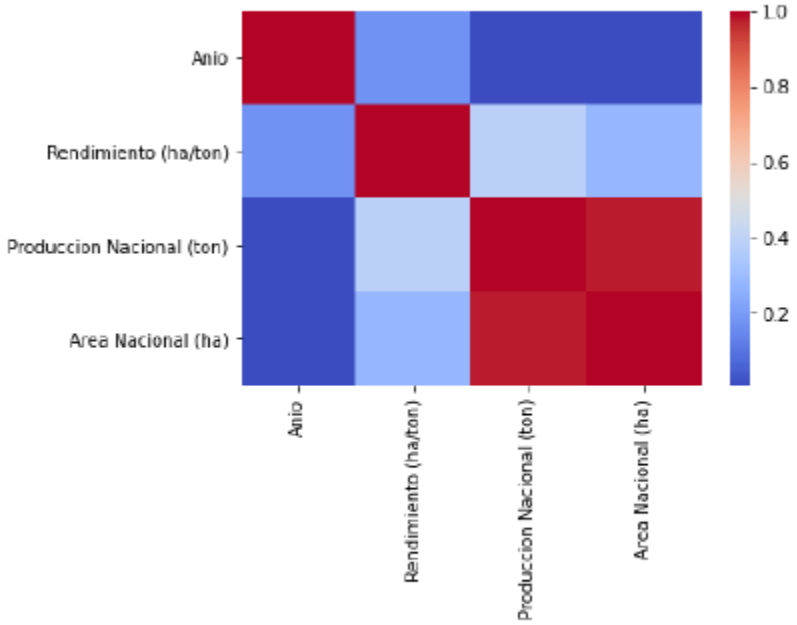
In [27]:
```
# valores correlacion de
Kendall import numpy as np
produccion_correlacion_kendall = produccion_df.corr(method='kendall')
produccion_correlacion_kendall
```

Out[27]:

Nacional Anio	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area (ha)
Anio	1.000000	0.140836	0.026879	0.016567
Rendimiento (ha/ton)	0.140836	1.000000	0.265165	0.186979
Produccion Nacional (ton)	0.026879	0.265165	1.000000	0.909233
Area Nacional (ha)	0.016567	0.186979	0.909233	1.000000

[28]:
```
# Generacion de mapa de calor para observar fácilmente las variables
correlaciona # las rojas son correlaciones fuertes positivas y las
azules correlaciones negati import seaborn as sns # esta libreria
permite crear gráficos estadísticos
sns.heatmap(produccion_correlacion_pearson,
xticklabels=produccion_correlacion_pearson.columns,
yticklabels=produccion_correlacion_pearson.columns, cmap='coolwarm'
)
```

Out[28]: <matplotlib.axes._subplots.AxesSubplot at 0x243b5e19be0>



In [33]:
```
# Imports necesarios
import numpy as np
import pandas as pd
import seaborn as sb
import matplotlib.pyplot as plt
%matplotlib inline
from mpl_toolkits.mplot3d import Axes3D
from matplotlib import cm
plt.rcParams['figure.figsize'] = (16, 9)
plt.style.use('ggplot')
from sklearn import linear_model
from sklearn.metrics import mean_squared_error, r2_score
```

In [35]:
```
produccion_df.shape  #Nos indica un dataframe de 266 registros con 8 variables o
```

Out[35]: (266, 8)

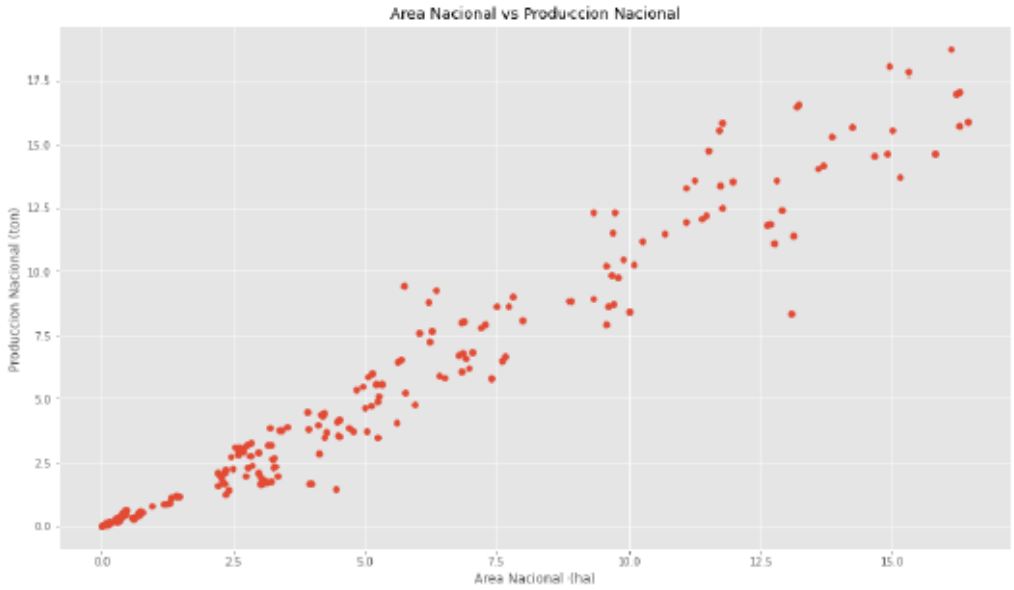
In [36]:
```
produccion_df.describe()
```

Out[36]:

	Anio	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
count	266.000000	266.000000	266.000000	266.000000
mean	2012.469925	0.936429	4.511316	4.511203
std	3.443484	0.267129	4.950568	4.565865
min	2007.000000	0.000000	0.000000	0.000000
25%	2010.000000	0.750000	0.352500	0.390000
50%	2012.000000	0.940000	2.720000	3.120000
75%	2015.000000	1.120000	7.147500	6.875000
max	2018.000000	2.000000	18.670000	16.430000

In [88]:
```
# Gráfico de dispersion del comportamiento del Area Nacional versus Produccion Na
plt.scatter(produccion_df['Area Nacional (ha)'],produccion_df['Produccion Nacional
plt.title('Area Nacional vs Produccion Nacional')
plt.xlabel('Area Nacional (ha)')
plt.ylabel("Produccion Nacional (ton)")
```

Out[88]:
```
Text(0, 0.5, 'Produccion Nacional (ton)')
```



In [ ]:
```
#CONSTRUCCION DEL MODELO PREDICTIVO UTILIZANDO EL METODO DE REGRESION LINEAL
```

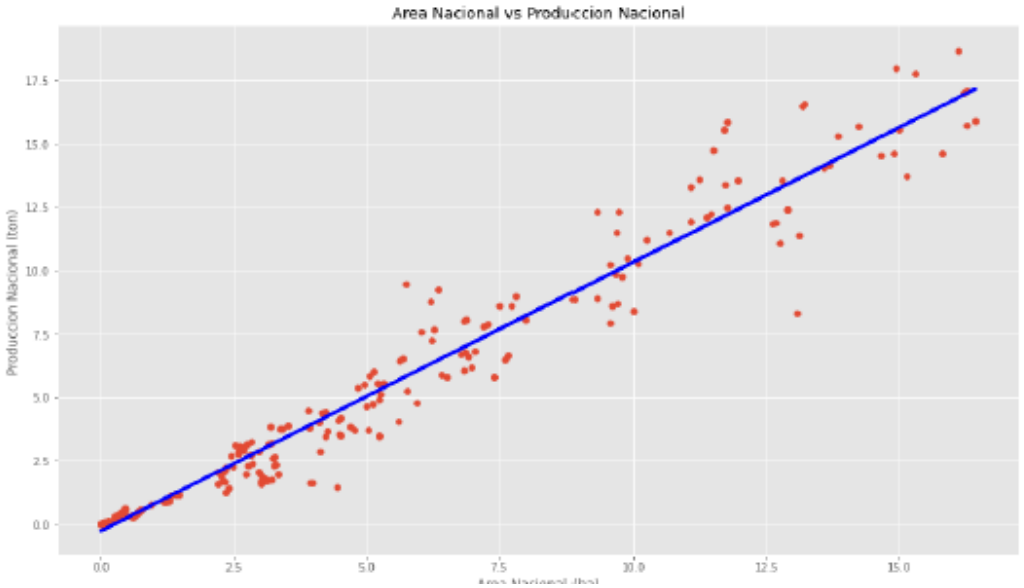
In [89]:
```
# Iniciamos el proceso para determinar el modelo de regresion lineal, de la anali
# Asignamos a nuestra variable de entrada X (En este caso corresponde al Area Nac
# Asignamos a la variable dependiente Y (En este caso corresponde a la Produccion
dataX =produccion_df[["Area Nacional (ha)"]]
X_train = np.array(dataX)
y_train = produccion_df['Produccion Nacional (ton)'].values
```

In [90]:
```
# Creamos la función objeto para determinar la Regresión Lineal Y= mX+bo
regr = linear_model.LinearRegression()

# Entrenamos nuestro modelo de regresion lineal, con la siguiente función
regr.fit(X_train, y_train)

# Hacemos las predicciones segun el modelo de regresion lineal
y_pred = regr.predict(X_train)

# Ahora imprimimos los resultados obtenidos
# Vemos el valor de la pendiente, osea la variable m, el coeficiente de la variab
print('Valor de la tangente (m) o Coefficients:=======> ', regr.coef_)
# Ahora el valor de la constante bo, es decir el valor donde la recta corta el ej
print('Valor de la constante o Independent term: =====>', regr.intercept_)
# Se imprime el Error Cuadrado Medio
print("Error cuadrado medio o Mean squared error:=====> %.2f " % mean_squared_err
# Puntaje o valor de Varianza. El mejor puntaje es un 1.0
print('valor de la varianza o Variance score:=========> %.2f' % r2_score(y_train,
```

```
Valor de la tangente (m) o Coefficients:=======>  [1.06084584]
Valor de la constante o Independent term: =====> -0.2743751434833559
Error cuadrado medio o Mean squared error:=====> 1.04
valor de la varianza o Variance score:=========> 0.96
```

In [91]:

```
# Gráfico de dispersion del comportamiento del Area Nacional versus Produccion Na
plt.scatter(produccion_df['Area Nacional (ha)'],produccion_df['Produccion Nacional
plt.title('Area Nacional vs Produccion Nacional')
plt.xlabel('Area Nacional (ha)')
plt.ylabel("Produccion Nacional (ton)")

# A continuación se grafica en colo azul, la funcion lineal obtenida a partir del
plt.plot(X_train[:,0], y_pred, color='blue', linewidth=3)
plt.show()
```



In [92]:

```
# Ahora vamos a predecir utilizando la función obtenida, la producción nacional (
# Queremos predecir cuántos toneladas de producción nacional de Café vamos a obte
# según nuestro modelo, hacemos:
produccion_obtenida = regr.predict([[2]])
print('Estimación de la Producción Nacional del Café en toneladas===>%.3f' %produ
```

```
Estimación de la Producción Nacional del Café en toneladas===>1.847
```

In [93]:

```
# Ahora vamos a predecir utilizando la función obtenida, la producción nacional (
# Queremos predecir cuántos toneladas de producción nacional de Café vamos a obte
# según nuestro modelo, hacemos:
produccion_obtenida = regr.predict([[2.5]])
print('Estimación de la Producción Nacional del Café en toneladas===>%.3f' %produ
```

```
Estimación de la Producción Nacional del Café en toneladas===>2.378
```

In [94]:
```
# Ahora vamos a predecir utilizando la función obtenida, la producción nacional (
# Queremos predecir cuántos toneladas de producción nacional de Café vamos a obte
# según nuestro modelo, hacemos:
produccion_obtenida = regr.predict([[8]])
print('Estimación de la Producción Nacional del Café en toneladas===>%.3f' %produ
```

```
Estimación de la Producción Nacional del Café en toneladas===>8.212
```

In [95]:
```
# Ahora vamos a predecir utilizando la función obtenida, la producción nacional (
# Queremos predecir cuántos toneladas de producción nacional de Café vamos a obte
# según nuestro modelo, hacemos:
produccion_obtenida = regr.predict([[11]])
print('Estimación de la Producción Nacional del Café en toneladas===>%.3f' %produ
```

```
Estimación de la Producción Nacional del Café en toneladas===>11.395
```

In [96]:
```
# Ahora vamos a predecir utilizando la función obtenida, la producción nacional (
# Queremos predecir cuántos toneladas de producción nacional de Café vamos a obte
# según nuestro modelo, hacemos:
produccion_obtenida = regr.predict([[15]])
print('Estimación de la Producción Nacional del Café en toneladas===>%.3f' %produ
```

```
Estimación de la Producción Nacional del Café en toneladas===>15.638
```

In [97]:
```
# Ahora vamos a predecir utilizando la función obtenida, la producción nacional (
# Queremos predecir cuántos toneladas de producción nacional de Café vamos a obte
# según nuestro modelo, hacemos:
produccion_obtenida = regr.predict([[35]])
print('Estimación de la Producción Nacional del Café en toneladas===>%.3f' %produ
```

```
Estimación de la Producción Nacional del Café en toneladas===>36.855
```

In [107]:
```
# Creamos la función objeto para determinar la Regresión Lineal Y= mX+bo
regr = linear_model.LinearRegression()

# Entrenamos nuestro modelo de regresion lineal, con la siguiente función
regr.fit(X_train, y_train)

# Hacemos las predicciones segun el modelo de regresion lineal
y_pred = regr.predict(X_train)

# Ahora imprimimos los resultados obtenidos
# Vemos el valor de la pendiente, osea la variable m, el coeficiente de la variab
print('Valor de la tangente (m) o Coefficients:=======> ', regr.coef_)
# Ahora el valor de la constante bo, es decir el valor donde la recta corta el ej
print('Valor de la constante o Independent term: =====>', regr.intercept_)
# Se imprime el Error Cuadrado Medio
print("Error cuadrado medio o Mean squared error:=====> %.2f " % mean_squared_err
# Puntaje o valor de Varianza. El mejor puntaje es un 1.0
print('valor de la varianza o Variance score:=========> %.2f' % r2_score(y_train,
```

```
Valor de la tangente (m) o Coefficients:=======>  [0.01642121]
Valor de la constante o Independent term: =====> 0.8623491800082033
Error cuadrado medio o Mean squared error:=====> 0.07
valor de la varianza o Variance score:=========> 0.08
```

In [108]:
```
# Gráfico de dispersion del comportamiento del Area Nacional versus Produccion Na
plt.scatter(produccion_df['Area Nacional (ha)'],produccion_df['Rendimiento (ha/to
plt.title('Area Nacional vs Produccion Nacional')
plt.xlabel('Area Nacional (ha)')
plt.ylabel("Rendimiento (ha/ton)")

# A continuación se grafica en colo azul, la funcion lineal obtenida a partir del
plt.plot(X_train[:,0], y_pred, color='blue', linewidth=3)
plt.show()
```

In [110]:
```
# Ahora vamos a predecir utilizando la función obtenida, el RENDIMIENTO (ha/ton):
# Queremos predecir el rendimiento de la producción nacional de Café vamos a obte
# según nuestro modelo, hacemos:
produccion_obtenida = regr.predict([[2]])
print('Estimación del rendimiento del Café en (hectareas/toneladas)===>%.3f' %prc
```

```
Estimación del rendimiento del Café en (hectareas/toneladas)===>0.895
```

In [111]:
```
# Ahora vamos a predecir utilizando la función obtenida, el RENDIMIENTO (ha/ton):
# Queremos predecir el rendimiento de la producción nacional de Café vamos a obte
# según nuestro modelo, hacemos:
produccion_obtenida = regr.predict([[6]])
print('Estimación del rendimiento del Café en (hectareas/toneladas)===>%.3f' %prc
```

```
Estimación del rendimiento del Café en (hectareas/toneladas)===>0.961
```

In [112]:
```
# Ahora vamos a predecir utilizando la función obtenida, el RENDIMIENTO (ha/ton):
# Queremos predecir el rendimiento de la producción nacional de Café vamos a obte
# según nuestro modelo, hacemos:
produccion_obtenida = regr.predict([[11]])
print('Estimación del rendimiento del Café en (hectareas/toneladas)===>%.3f' %prc
```

```
Estimación del rendimiento del Café en (hectareas/toneladas)===>1.043
```

In [113]:
```
# Ahora vamos a predecir utilizando la función obtenida, el RENDIMIENTO (ha/ton):
# Queremos predecir el rendimiento de la producción nacional de Café vamos a obte
# según nuestro modelo, hacemos:
produccion_obtenida = regr.predict([[26]])
print('Estimación del rendimiento del Café en (hectareas/toneladas)===>%.3f' %prc
```

```
Estimación del rendimiento del Café en (hectareas/toneladas)===>1.289
```

CONCLUSION

- El Dataframe trata sobre Producción del café que consta de los productos y su escala de producción en los últimos años, datos que nos sirvieron para determinar y realizar un análisis de la café en Colombia y así explicar todo lo relacionado con este producto, es su fuerte impacto e influencia que tiene para la economía colombiana y como sus dinámicas de productividad han sabido mantenerse al margen, a pesar de sus diferentes crisis.
- Para finalizar se pudo esclarecer el manejo del Dataframe y ampliar nuestro conocimiento en cuanto análisis de la información de cualquier archivo de datos que se requiera para que se convierta a fracciones de información solo solicitada.

BIBLIOGRAFIA

https://www.sic.gov.co/recursos_user/documentos/promocion_competencia/Estudios_Economicos/Estudios_Economicos/Estudios_Mercado/EstudiosectorialCafe.pdf

https://federaciondecafeteros.org/wp/listado-noticias/produccion-de-cafe-de-colombia-cerroel-2019-en-148-millones-de-sacos/

https://www.valoraanalitik.com/2018/12/04/produccion-y-exportacion-de-cafe-de-colombiaen-su-maximo-de-un-ano/

https://www.colombiatrade.com.co/noticias/como-aprovechar-oportunidades-para-exportarcafe-mercados-internacionales https://compradores.procolombia.co/es/explore-oportunidades/caf-s-especiales-0

https://repository.udem.edu.co/bitstream/handle/11407/304/Plan%20exportador%20de%20caf%E9%20especial%20suave%20colombiano%20tostado%20y%20molido%20a%20mercados%20internacionales.pdf?sequence=1

fao.org/3/a-i5137s.pdf – GRAFICAS INTERESANTES

https://www.datos.gov.co/es/Agricultura-y-Desarrollo-Rural/Evaluaci-n-de-Tierras-con-finesAgr-colas-para-el-/ggaa-6f3s

https://www.datos.gov.co/en/Agricultura-y-Desarrollo-Rural/Evaluaci-n-de-tierras-con-finesagr-colas-para-el-/h4ms-ukui

https://www.datos.gov.co/en/Agricultura-y-Desarrollo-Rural/TIPO-DE-CULTIVOS-DE-LASVEREDAS-DE-HERRAN/uvav-hgca

https://www.datos.gov.co/en/Agricultura-y-Desarrollo-Rural/Superficie-Sembrada-porhectareas-con-cultivos-per/v4ub-9eme

GLOSARIO

Análisis

Big Data

Calidad de los Datos

Ciencia de Datos

Conocimiento

CRISP-DM

CRM

Cualitativo

Cuantitativo

Dato

Deep Data

Descriptiva

Diagnóstico

Eficacia

Eficiencia

Enfoque

Estadística

Etapa

ETL

Información

Informática

Inteligencia de Negocios

Investigación

Jupyter Notebook

KDD

Minería de Datos

Modelo

Negocio

Predictiva

Prescriptiva

Proceso

Productividad

Python

Rendimiento

Rentabilidad

RStudio

Síntesis

Sistema Operativo

Smart Data

Toma de decisiones

Validación

Variable

Café Arábica