

Final-Definitivo-Informe-Proyecto

Juan Pablo Serrano Rodríguez

Departamento del Valle del Cauca, Servicio Nacional de Aprendizaje

Limpieza de Datos

Ing. Luis Armando Amaya Quiroga

08-10-2020

INTRODUCCION

El propósito de este trabajo es mostrar la producción de café en Colombia durante los últimos años y mirar si la producción y precio de este ha subido o bajado en tiempos recientes, mediante el uso de gráficas y tablas se plasmaran los datos con respecto a la producción y uso de tierras.

Producción de café 2018-2019

URL: <https://federaciondecafeteros.org/wp/estadisticas-cafeteras/>

Producción de café en 2019 (Sacos 60 kg)

Fuente: Federación de cafeteros

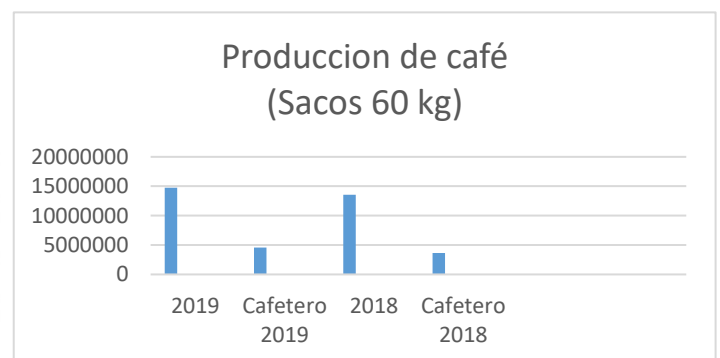
Ene -Dic 2019	14.752.000
Ene Dic 2018	13.557.000
Variación	9%

La producción de café del 2019 cerró con 14,7 millones de sacos de 60 kilos vendidos con un porcentaje del 9% más a comparación del año 2018 que fue de 13,5 Millones de sacos de 60 kilos.

Producción de café en año cafetero 2018-2019

Producción de café año cafetero (Sacos 60 kg)

Oct 2019 - Dic 2019	4.555.000
Oct 2018 Dic 2018	3.669.000
Variación	24%



La producción del año cafetero en el 2019 fue mayor a en un 24% a comparación del año cafetero del 2018

URL: <https://federaciondecafeteros.org/wp/estadisticas-cafeteras/>

Fuente: Federación de cafeteros

Producción de café en diciembre 2018-2019



La producción del café en diciembre del 2019 fue mayor al 2018 en un 31%

URL: <https://federaciondefeferos.org/wp/estadisticas-cafeferas/>

Fuente: Federación de cafeteros

El Precio interno del café Mensual entre los años 2018-2019-Promedio

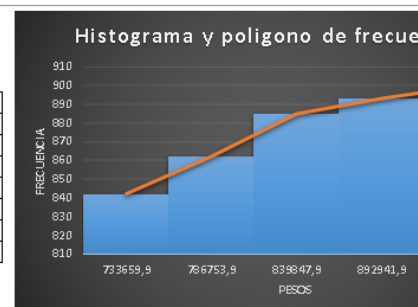
El dato que se muestra es un promedio del precio del café diario

Precio interno base de compra del café colombiano - promedio Mensual
Pesos por carga de 125 kg. de café pergamino seco

dic-17	757.968	
ene-18	763.903	Columna1
feb-18	745.031	
mar-18	729.855	Media 764288,927
abr-18	715.325	Error típico 14929,3138
may-18	754.210	Mediana 745715,625
jun-18	746.400	Moda #N/A
jul-18	717.839	Desviación e 73138,402
ago-18	705.065	Varianza de 5349225853
sep-18	686.933	Curtosis 3,79699077
oct-18	796.774	Coefficiente c 1,72478646
nov-18	804.283	Rango 318562,366
dic-18	727.645	Mínimo 680566,667
ene-19	727.274	Máximo 999129,032
feb-19	708.089	Suma 18342934,2
mar-19	690.581	Cuenta 24
abr-19	680.567	
may-19	724.065	
jun-19	779.917	
jul-19	796.484	
ago-19	798.935	
sep-19	815.450	

Intervalo 6
Amplitud 53093,728

	Intervalos	Grupos	Frecuencia
0	Li	Ls	
1	680567	733660	733659,9 842
2	733660	786754	786753,9 862
3	786754	839848	839847,9 885
4	839848	892942	892941,9 900
5	892942	946035	946034,9 904
6	946035	999129	999128,9 904



Como se aprecia en la gráfica mientras más avanza más alta esta

URL: <https://federaciondecafeteros.org/wp/estadisticas-cafeteras/>

Fuente: Federación de cafeteros

Precio externo del café colombiano-Promedio Mensual

Centavos de dólar por libra de 453,6 gr

Precio externo del café colombiano (exdock)* - promedio mensual

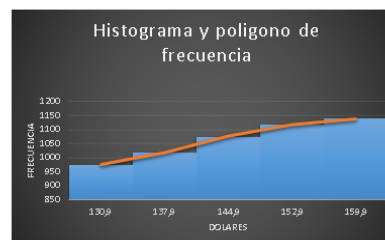
Centavos de dólar por libra de 453,6 gr. Excelso

* Resultado de la ponderación de los precios de los 6 días anteriores

ago-17	156,46	
sep-17	151,57	
oct-17	145,01	
nov-17	143,85	
dic-17	142,51	
ene-18	143,48	Media 133,1929
feb-18	142,35	Error típico 2,9641896
mar-18	139,79	Mediana 130,6246
abr-18	139,00	Moda #N/A
may-18	140,50	Desviación 10,268254
jun-18	139,17	Varianza de 105,43704
jul-18	134,85	Curtosis 3,9461992
ago-18	130,99	Coefficiente 1,8556727
sep-18	126,53	Rango 37,030399
oct-18	138,80	Mínimo 123,13484
nov-18	140,75	Máximo 160,16524
dic-18	129,62	Suma 1598,3148
ene-19	128,68	Cuenta 12
feb-19	128,97	
mar-19	125,47	
abr-19	124,07	
may-19	123,13	
jun-19	132,77	
jul-19	138,41	
ago-19	129,89	

Intervalo 5
Amplitud 7,4060799

	Intervalos	Grupos	Frecuencia
	Li	Ls	
1	123	131	130,9 974
2	131	138	137,9 1016
3	138	145	144,9 1075
4	145	153	152,9 1119
5	153	160	159,9 1139



Como se aprecia en la gráfica mientras más avanza más alta esta

URL: <https://federaciondecafeteros.org/wp/estadisticas-cafeteras/>

Fuente: Federación de cafeteros

Precio externo del café colombiano-Promedio Anual Civil

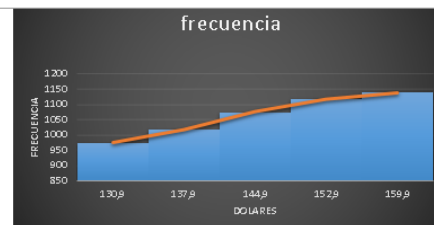
Precio externo del café colombiano (exdock)* - promedio mensual

Centavos de dólar por libra de 453,6 gr. Excelso

* Resultado de la ponderación de los precios de los 6 días anteriores

ene-18	143,48	
feb-18	142,35	
mar-18	139,79	
abr-18	139,00	Media 133,1929
may-18	140,50	Error típico 2,9641896
jun-18	139,17	Mediana 130,6246
jul-18	134,85	Moda #N/A
ago-18	130,99	Desviación 10,268254
sep-18	126,53	Varianza de 105,43704
oct-18	138,80	Curtosis 3,9461992
nov-18	140,75	Coefficiente 1,8556727
dic-18	129,62	Rango 37,030399
ene-19	128,68	Mínimo 123,13484
feb-19	128,97	Máximo 160,16524
mar-19	125,47	Suma 1598,3148
abr-19	124,07	Cuenta 12
may-19	123,13	
jun-19	132,77	
jul-19	138,41	
ago-19	129,89	
sep-19	131,36	
oct-19	131,88	
nov-19	143,52	
dic-19	160,17	
ene-20	150,80	

	Intervalos	Grupos	Frecuencia
	Li	Ls	
1	123	131	130,9 974
2	131	138	137,9 1016
3	138	145	144,9 1075
4	145	153	152,9 1119
5	153	160	159,9 1139



Como se aprecia en la gráfica mientras más avanza más alta esta

URL:<https://federaciondecafeteros.org/wp/estadisticas-cafeteras/>

Fuente: Federación de cafeteros

Producción mensual del café

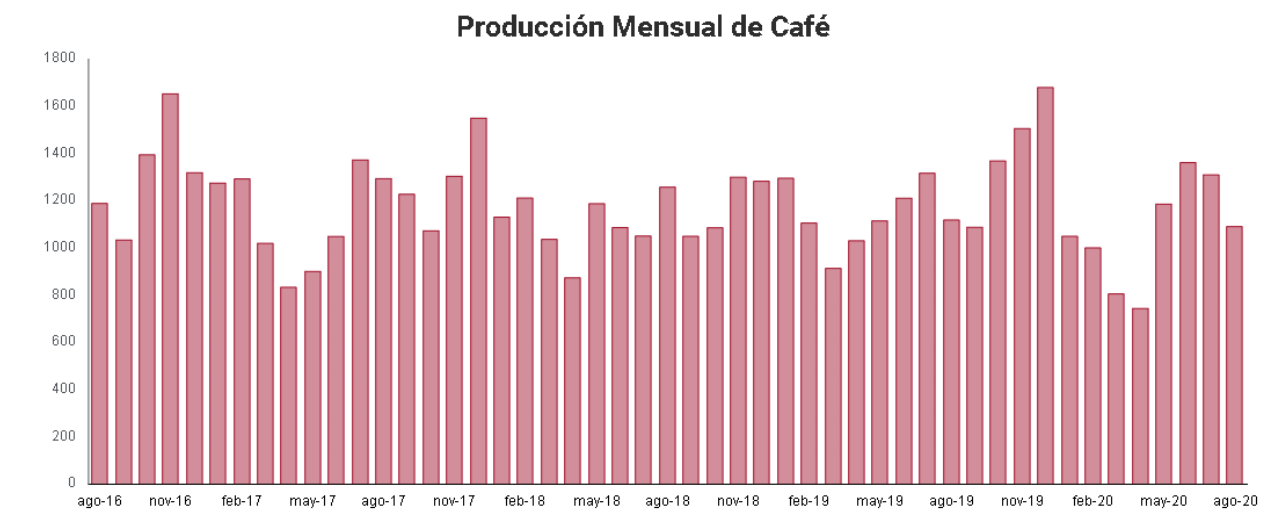


Ilustración 1 Producción del café

La producción más alta se dio durante noviembre del 2019 y la más baja durante mayo del 2020

URL:<https://federaciondecafeteros.org/wp/estadisticas-cafeteras/>

Fuente: Federación de cafeteros

Precio interno diario del café en Colombia

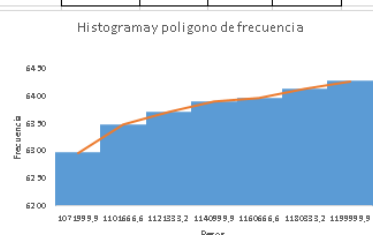
Precio Interno del Café Colombiano

Definición: Precio interno base de compra del FoNC por carga de 125 Kg. de café pergamino seco.

záb-01-año-20	\$	1.140.000
dam-02-año-20	\$	1.140.000
jun-03-año-20	\$	1.140.000
mar-04-año-20	\$	1.170.000
may-05-año-20	\$	1.165.000
jun-06-año-20	\$	1.123.000
vía-07-año-20	\$	1.115.000
záb-08-año-20	\$	1.115.000
dam-09-año-20	\$	1.115.000
jun-10-año-20	\$	1.105.000
mar-11-año-20	\$	1.042.000
may-12-año-20	\$	1.091.000
jun-13-año-20	\$	1.125.000
vía-14-año-20	\$	1.120.000
záb-15-año-20	\$	1.120.000
dam-16-año-20	\$	1.120.000
jun-17-año-20	\$	1.120.000
mar-18-año-20	\$	1.153.000
may-19-año-20	\$	1.130.000
jun-20-año-20	\$	1.141.000
vía-21-año-20	\$	1.151.000
záb-22-año-20	\$	1.153.000
dam-23-año-20	\$	1.153.000
jun-24-año-20	\$	1.147.000
mar-25-año-20	\$	1.142.000
may-26-año-20	\$	1.174.000
jun-27-año-20	\$	1.170.000
vía-28-año-20	\$	1.110.000
záb-29-año-20	\$	1.110.000
dam-30-año-20	\$	1.110.000
jun-31-año-20	\$	1.200.000

Resumen Agosto	
Media	1143967,742
Error típico	5407,975785
Mediana	1141000
Moda	1115000
Desviación estándar	30110,33474
Varianza de la muestra	906632258,1
Curtosis	-0,778795196
Coefficiente de asimetría	-0,068654951
Rango	118000
Mínimo	1082000
Máximo	1200000
Suma	35463000
Cuenta	31

Intervalo	6			
Amplitud	19666,66667			
	Intervalo	Grupo	Frecuencia	
0	Li	Lr		
1	1082000	1101666,67	1071999,9	6297
2	1101666,667	1121333,3	1101666,6	6348
3	1121333,3	1141000,0	1121333,2	6370
4	1141000,0	1160666,7	1140999,9	6390
5	1160666,7	1180333,3	1160666,6	6396
6	1180333,3	1200000,0	1180333,2	6412
7	1200000	1219666,7	1199999,9	6427



En el histograma podemos observar que mientras más avanza la gráfica más alta esta y la información está basada en el precio interno diario del mes de agosto en el año 2019.

URL: <https://federaciondecafeteros.org/wp/estadisticas-cafeteras/>

Fuente: Federación de cafeteros

Precio externo del café Colombiano

Precio externo del café colombiano (exdock)* - promedio mensual

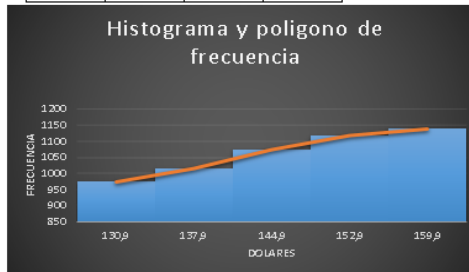
Centavos de dólar por libra de 453.6 gr. Excelso

* Resultado de la ponderación de los precios de los 6 días anteriores

		Dólares
ene-19	128,68	
feb-19	128,97	
mar-19	125,47	Media 133,1929
abr-19	124,07	Error típico 2,9641896
may-19	123,13	Mediana 130,6246
jun-19	132,77	Moda #N/A
jul-19	138,41	Desviación 10,268254
ago-19	129,89	Varianza de 105,43704
sep-19	131,36	Curtosis 3,9461992
oct-19	131,88	Coefficiente 1,8556727
nov-19	143,52	Rango 37,030399
dic-19	160,17	Mínimo 123,13484
ene-20	150,80	Máximo 160,16524
feb-20	144,13	Suma 1598,3148
mar-20	157,31	Cuenta 12
abr-20	164,62	
may-20	155,82	
jun-20	149,24	
jul-20	151,29	
ago-20	167,63	
sep-20		
oct-20		
nov-20		
dic-20		

Intervalo 5
Amplitud 7,4060799

	Intervalos	Grupos	Frecuencia
Li	Ls		
1	123	131	130,9 974
2	131	138	137,9 1016
3	138	145	144,9 1075
4	145	153	152,9 1119
5	153	160	159,9 1139



En el histograma podemos observar que mientras más avanza la gráfica más alta esta y los datos usada para la gráfica están basados en el precio interno mensual del año 2019.

URL: <https://federaciondefeferos.org/wp/estadisticas-cafeferas/>

Fuente: Federación de cafeteros

CONDICIONES IDEALES PARA EL CULTIVO

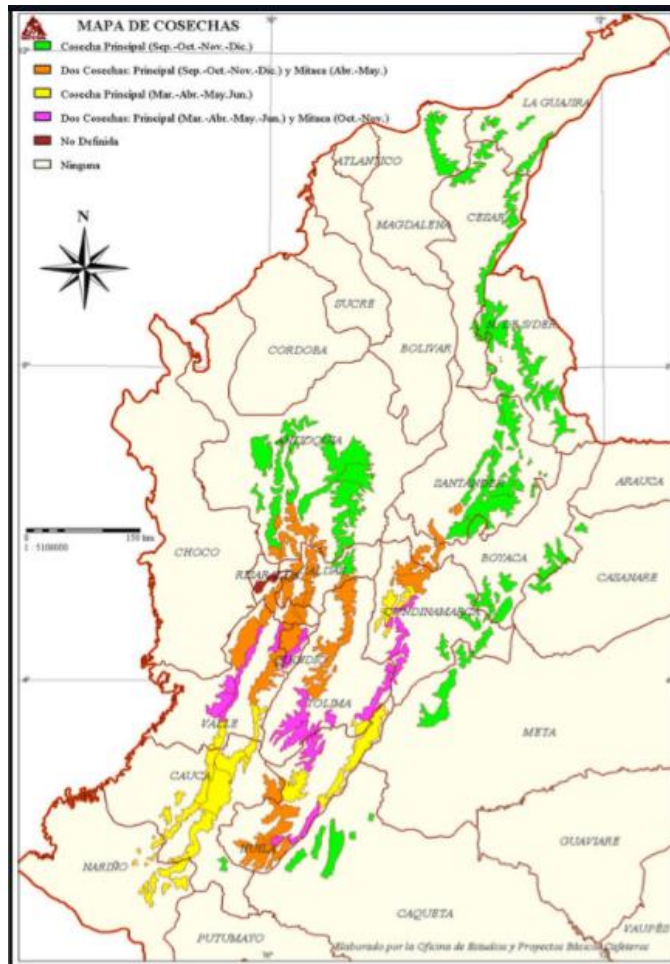


Para los cultivos de café el ambiente debe de estar a 17 y 23 grados

URL: <https://www.cafedecolombia.com/particulares/regiones-cafeteras/>

Fuente: Café de Colombia

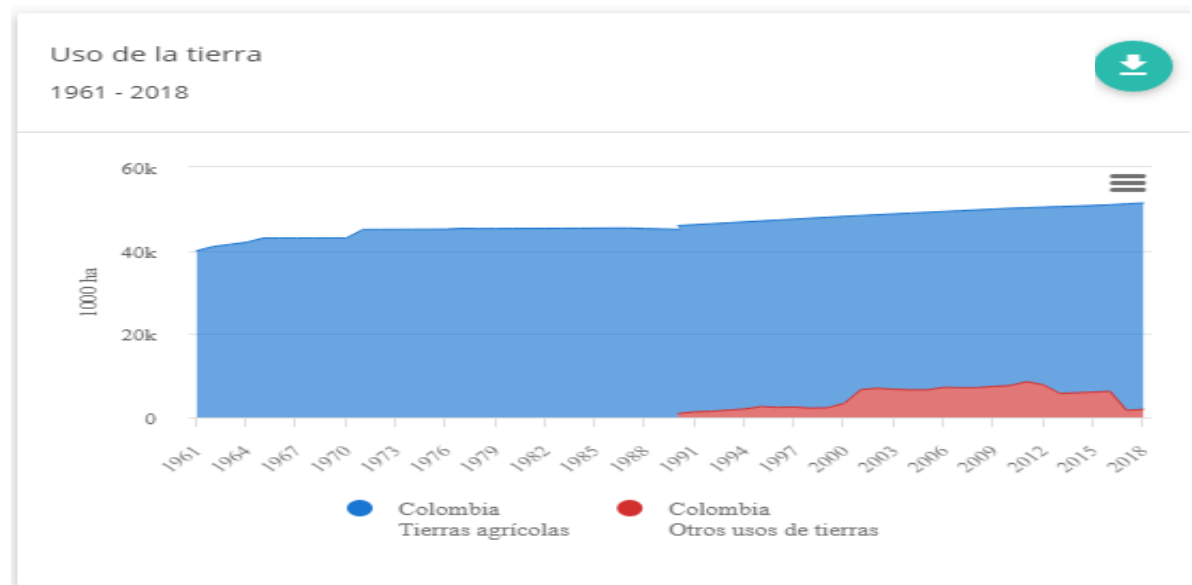
Mapa de Cosechas



URL: <https://cnccolombia.com.co/index.php/cafe-colombiano/regiones-cafeteras>

Fuente: Compañía Nacional del café

Uso de las Tierras



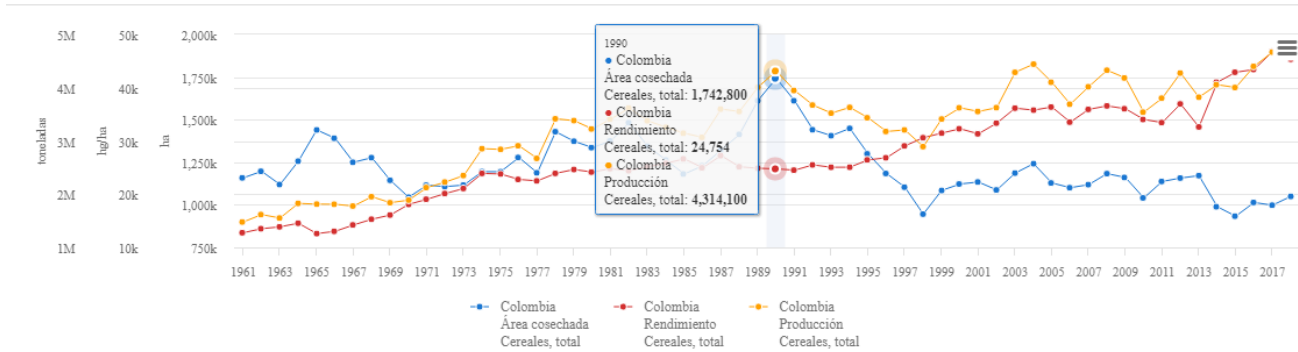
En la grafica se muestra que en colombia, la tierra se suele usar mas para la cosecha

URL: <http://www.fao.org/faostat/es/?#country/44>

Fuente: FAOSTAT

Cereales Total Produccion

Cereales, Total Producción
1961 - 2018



En los últimos años se ha mostrado una disminución en el área de cosechas al contrario de la producción y el rendimiento de la cosecha que fueron altas sus estadísticas.

URL: <http://www.fao.org/faostat/es/?#country/44>

Fuente: FAOSTAT

Análisis de Jupyter

```
In [1]: import pandas as pd
# Se importa la biblioteca pandas
```

```
In [2]: pd.read_csv("PRODUCCION.csv")
# Muestra la tabla
```

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

```
In [3]: Produccion_df=pd.read_csv("PRODUCCION.csv")
# Asignacion de la variable para el Dataframe
```

```
In [4]: Produccion_df
# Listado del dataframe
```

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
...
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 x 8 columns

```
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
# Muestra la estructura del dataframe y el tipo de elemento de cada campo
```

```
Out[6]: Año                               int64
Departamento                          object
Producto                              object
Área (ha)                             object
Producción (ton)                       object
Rendimiento (ha/ton)                  float64
Producción Nacional (ton)              float64
Área Nacional (ha)                    float64
dtype: object
```

```
In [7]: Produccion_df.columns
# Muestra las columnas del dataframe
```

```
Out[7]: Index(['Año', 'Departamento', 'Producto', 'Área (ha)', 'Producción (ton)',
              'Rendimiento (ha/ton)', 'Producción Nacional (ton)',
              'Área Nacional (ha)'],
              dtype='object')
```

```
In [8]: Produccion_df.shape
# Muestra la cantidad de columnas y filas del dataframe
```

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

```
In [9]: pd.unique(Produccion_df['Año'])
# Muestra los valores del Año y el tipo de manrea horizontal
```

```
Out[9]: array([2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017,
              2018], dtype=int64)
```

```
In [10]: pd.unique(Produccion_df['Departamento'])
# Muestra los valores del Departamento y el tipo de manrea horizontal
```

```
Out[10]: array(['ANTIOQUIA', 'BOLIVAR', 'BOYACA', 'CALDAS', 'CAQUETA', 'CASANARE',
               'CAUCA', 'CESAR', 'CHOCO', 'CUNDINAMARCA', 'HUILA', 'LA GUAJIRA',
               'MAGDALENA', 'META', 'NARIÑO', 'NORTE DE SANTANDER', 'PUTUMAYO',
               'QUINDIO', 'RISARALDA', 'SANTANDER', 'TOLIMA', 'VALLE DEL CAUCA',
               'ARAUCA', 'GUAVIARE'], dtype=object)
```

```
In [11]: pd.unique(Produccion_df['Producto'])
# Muestra los valores del Producto y el tipo de manrea horizontal
```

```
Out[11]: array(['CAFE'], dtype=object)
```

```
In [12]: pd.unique(Produccion_df['Área (ha)'])
# Muestra los valores del Área(ha) y el tipo de manrea horizontal
```

```
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',
               '55,015.00', '33,011.00', '37,175.05', '70,000.24', '5,143.00',
```



```
In [13]: pd.unique(Produccion_df['Produccion (ton)'])
# Muestra los valores del Produccion (Ton) y el tipo de manrea hizontal
```

```
Out[13]: array(['120,500.00', '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',
                '62,711.00', '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',
                '652.50', '4,981.59', '54,115.96', '2,446.38', '1,718.25',
                '50,588.14', '19,994.35', '140.00', '30,786.41', '85,212.64',
```

```
In [14]: pd.unique(Produccion_df['Rendimiento (ha/ton)'])
# Muestra los valores del Rendimiento (ha/ton) y el tipo de manrea hizontal
```

```
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)'])
# Muestra los valores del Produccion nacional (ton) y el tipo de manrea hizontal
```

```
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, 5.700e+00, 1.700e+00
```

```
In [16]: pd.unique(Produccion_df['Area Nacional (ha)'])
# Muestra los valores del Area Nacional (ha) y el tipo de manrea hizontal
```

```
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, 0.000e-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, 0.710e+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,
                2.720e+00, 2.970e+00, 6.420e+00, 4.780e+00, 1.280e+01, 9.780e+00,
                1.422e+01, 9.000e-02, 1.200e+00, 7.810e+00, 3.800e-01, 9.600e+00,
                3.250e+00, 4.690e+00, 1.531e+01, 7.500e-01, 2.200e+00, 3.200e-01,
                4.160e+00, 3.280e+00, 2.750e+00, 5.130e+00, 5.000e+00, 1.261e+01,
                6.930e+00, 1.384e+01, 1.240e+00, 7.510e+00, 0.690e+00, 3.290e+00,
                4.230e+00, 1.612e+01, 7.600e-01, 4.220e+00, 2.980e+00, 2.700e+00,
                5.050e+00, 5.120e+00, 1.267e+01, 7.040e+00, 1.369e+01, 1.300e-01,
```

```
In [17]: Produccion_df['Anio']
# Muestra Los valores del campo y el tipo de forma vertical
```

```
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['Departamento']
# Muestra Los valores del departamento y tipo de forma vertical
```

```
Out[18]: 0      ANTIOQUIA
1      BOLIVAR
2      BOYACA
3      CALDAS
4      CAQUETA
...
261    QUINDIO
262    RISARALDA
263    SANTANDER
264    TOLIMA
265    VALLE DEL CAUCA
Name: Departamento, Length: 266, dtype: object
```

```
In [19]: Produccion_df['Departamento'] + Produccion_df['Produccion (ton)']
# Muestra Los valores de Departamento y el produccion con tipo
```

```
Out[19]: 0      ANTIOQUIA120,500.80
1      BOLIVAR446.00
2      BOYACA9,683.10
3      CALDAS92,815.00
4      CAQUETA2,134.00
...
261    QUINDIO17,739.03
262    RISARALDA45,918.75
263    SANTANDER55,918.71
264    TOLIMA97,451.31
265    VALLE DEL CAUCA49,667.88
Length: 266, dtype: object
```

```
In [20]: Produccion_df['Produccion (ton)'] + Produccion_df['Producto']
# Muestra Los valores de produccion y el Producto con tipo
```

```
Out[20]: 0      120,500.80CAFE
1      446.00CAFE
2      9,683.10CAFE
3      92,815.00CAFE
4      2,134.00CAFE
...
261    17,739.03CAFE
262    45,918.75CAFE
263    55,918.71CAFE
264    97,451.31CAFE
265    49,667.88CAFE
Length: 266, dtype: object
```

```
In [21]: Produccion_df['Departamento'], Produccion_df['Produccion (ton)']
# Muestra Los valores de Departamento y el produccion con tipo por separado
```

```
Out[21]: (0      ANTIOQUIA
1      BOLIVAR
2      BOYACA
3      CALDAS
4      CAQUETA
...
261    QUINDIO
262    RISARALDA
263    SANTANDER
264    TOLIMA
265    VALLE DEL CAUCA
Name: Departamento, Length: 266, dtype: object,
0      120,500.80
1      446.00
2      9,683.10
3      92,815.00
4      2,134.00
...
261    17,739.03
262    45,918.75
263    55,918.71
264    97,451.31
265    49,667.88
Name: Produccion (ton), Length: 266, dtype: object)
```

```
In [22]: Produccion_df['Produccion (ton)'], Produccion_df['Producto']
# Muestra los valores de produccion y el Producto con tipo por separado
```

```
Out[22]: (0      120,500.80
1         446.00
2      9,683.10
3    92,815.00
4      2,134.00
...
261    17,739.03
262    45,918.75
263    55,918.71
264    97,451.31
265    49,667.88
Name: Produccion (ton), Length: 266, dtype: object,
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 [23]: Produccion_df.describe()
# Muestra los valores del dataframe de forma detallada y estadística
```

```
Out[23]:
```

	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 [24]: Produccion_df['Anio'].describe()
# Muestra los valores del anio de forma detallada y estadística
```

```
Out[24]: count    266.000000
mean    2012.469925
std      3.443484
min     2007.000000
25%     2010.000000
50%     2012.000000
75%     2015.000000
max     2018.000000
Name: Anio, dtype: float64
```

```
In [25]: Produccion_df['Area (ha)'].describe()
# Muestra los valores del Area (ha) de forma detallada y estadística
```

```
Out[25]: count      266
unique      261
top        850.00
freq         2
Name: Area (ha), dtype: object
```

```
In [26]: Produccion_df['Rendimiento (ha/ton)'].describe()
# Muestra los valores del Rendimiento (ha/ton) de forma detallada y estadística
```

```
Out[26]: 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 [27]: Produccion_df['Produccion Nacional (ton)'].describe()  
# Muestra los valores del año de forma detallada y estadística
```

```
Out[27]: count      266.000000  
mean         4.511316  
std          4.950568  
min          0.000000  
25%         0.352500  
50%         2.720000  
75%         7.147500  
max         18.670000  
Name: Produccion Nacional (ton), dtype: float64
```

```
In [28]: Produccion_df['Año'].min()  
# Muestra el valor mínimo de la columna
```

```
Out[28]: 2007
```

```
In [29]: Produccion_df['Año'].max()  
# Muestra el valor máximo de la columna
```

```
Out[29]: 2018
```

```
In [30]: Produccion_df['Año'].count()  
# Cuenta cuántos datos hay en cada columna
```

```
Out[30]: 266
```

```
In [31]: Produccion_df['Año'].mean()  
# indica el promedio de cada
```

```
Out[31]: 2012.46992481203
```

```
In [32]: Produccion_df['Rendimiento (ha/ton)'].min()
```

```
Out[32]: 0.0
```

```
In [33]: Produccion_df['Rendimiento (ha/ton)'].max()
```

```
Out[33]: 2.0
```

```
In [34]: Produccion_df['Rendimiento (ha/ton)'].count()
```

```
Out[34]: 266
```

```
In [35]: Produccion_df['Rendimiento (ha/ton)'].mean()
```

```
Out[35]: 0.9364285714285712
```

```
In [36]: Produccion_df['Produccion Nacional (ton)'].min()
```

```
Out[36]: 0.0
```

```
In [37]: Produccion_df['Produccion Nacional (ton)'].max()
```

```
Out[37]: 18.67
```

```
In [38]: Produccion_df['Produccion Nacional (ton)'].count()
```

```
Out[38]: 266
```

```
In [39]: Produccion_df['Produccion Nacional (ton)'].mean()
```

```
Out[39]: 4.511315789473683
```

```
In [40]: Produccion_df['Area Nacional (ha)'].min()
```

```
Out[40]: 0.0
```

```
In [41]: Produccion_df['Area Nacional (ha)'].max()
```

```
Out[41]: 16.43
```

```
In [42]: Produccion_df['Area Nacional (ha)'].count()
```

```
Out[42]: 266
```

```
In [43]: Produccion_df['Area Nacional (ha)'].mean()
```

```
Out[43]: 4.511203007510795
```

```
In [44]: Produccion_df.groupby('Anio')['Produccion (ton)'].count()[2012]
# Agrupa los datos del anio y cuenta las de produccion que sean igual a 2012
```

```
Out[44]: 23
```

```
In [45]: Produccion_df.groupby('Departamento')['Produccion (ton)'].count()['VALLE DEL CAUCA']
```

```
Out[45]: 12
```

```
In [46]: Produccion_df.groupby('Anio')['Produccion (ton)'].count()
# Agrupa los datos del anio y muestra la cantidad de produccion (ton)
```

```
Out[46]: 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: Produccion (ton), dtype: int64
```

```
In [47]: Produccion_df.groupby('Departamento')['Produccion (ton)'].count()
```

```
Out[47]: 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 [48]: Produccion_df.groupby('Departamento')['Anio'].count()
```

```
Out[48]: 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: Anio, dtype: int64
```

```
In [49]: Producto_Counts=Produccion_df.groupby('Departamento')['Producto'].count()
print(Producto_Counts)
# Cuenta el departamento por la produccion y luego imprime los resultados
```

```
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: Producto, dtype: int64
```

```
In [50]: Produccion_df.head(20)
# Muestra las primeras 20 filas del dataframe
```

```
Out[50]:
```

	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	13,276.50	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

```
In [51]: Produccion_df.tail(20)
# Muestra las ultimas 20 filas del dataframe
```

```
Out[51]:
```

	Anio	Departamento	Producto	Area (ha)	Produccion (ton)	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
246	2018	BOYACA	CAFE	9,653.45	7,780.34	0.81	0.91	1.30
247	2018	CALDAS	CAFE	50,762.22	68,670.96	1.35	8.02	6.84
248	2018	CAQUETA	CAFE	3,485.24	5,280.40	1.52	0.62	0.47
249	2018	CASANARE	CAFE	2,360.55	1,629.25	0.69	0.19	0.32
250	2018	CAUCA	CAFE	82,085.54	102,147.00	1.24	11.94	11.06
251	2018	CESAR	CAFE	23,915.45	14,943.62	0.62	1.75	3.22
252	2018	CHOCO	CAFE	140.33	181.42	1.29	0.02	0.02
253	2018	CUNDINAMARCA	CAFE	29,085.24	32,580.24	1.12	3.81	3.92
254	2018	HUILA	CAFE	122,002.46	136,161.86	1.12	15.91	16.43
255	2018	LA GUAJIRA	CAFE	4,810.97	2,990.91	0.62	0.35	0.65
256	2018	MAGDALENA	CAFE	17,414.32	10,826.24	0.62	1.26	2.35
257	2018	META	CAFE	2,761.01	3,877.62	1.40	0.45	0.37
258	2018	NARIÑO	CAFE	33,465.54	35,679.42	1.07	4.17	4.51
259	2018	NORTE DE SANTANDER	CAFE	20,673.04	23,471.69	1.12	2.74	2.81
260	2018	PUTUMAYO	CAFE	209.93	289.50	1.38	0.03	0.03
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

```
In [52]: Produccion_df['Produccion Nacional (ton)'].std()
# Describe la desviacion estandar
```

```
Out[52]: 4.950567735489969
```

```
In [53]: Anio_grouped=Produccion_df.groupby('Anio')
print(Anio_grouped)

<pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000000086674C8>
```

```
In [54]: Anio_grouped.describe()
```

Out[54]:

	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	...	6.9650	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	0.98	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
2015	22.0	1.024545	0.110096	0.77	0.9350	1.065	1.1075	1.15	22.0	4.545455	...	6.4675	17.07	22.0	4.545455	4.793782	0.02	0.4975	3.050	6.4775
2016	21.0	1.063810	0.116725	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	3.280	6.7700
2017	22.0	1.068182	0.272443	0.66	0.8450	1.090	1.2900	1.50	22.0	4.545909	...	6.3550	16.49	22.0	4.545455	4.781849	0.02	0.5150	3.090	6.5275
2018	22.0	1.079545	0.296672	0.62	0.8575	1.120	1.3125	1.52	22.0	4.545455	...	6.3475	16.58	22.0	4.546364	4.858793	0.02	0.5150	3.015	6.3050

17 rows x 21 columns

```
In [55]: Anio_grouped.mean()
```

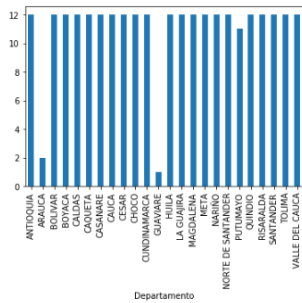
Out[55]:

	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
Anio			
2007	0.950455	4.545909	4.545455
2008	0.982727	4.545455	4.545000
2009	0.881364	4.545455	4.544545
2010	0.906087	4.348261	4.347826
2011	0.854348	4.348696	4.347826
2012	0.858696	4.347391	4.347826
2013	0.759545	4.545455	4.545000
2014	0.822273	4.545455	4.545455
2015	1.024545	4.545455	4.545455
2016	1.063810	4.761429	4.761905
2017	1.068182	4.545909	4.545455
2018	1.079545	4.545455	4.546364

```
In [56]: %matplotlib inline
# Se carga la libreria para graficos de pandas denominada matplotlib
```

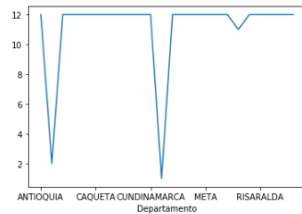
```
In [57]: Producto_Counts.plot(kind='bar')
```

Out[57]: <matplotlib.axes._subplots.AxesSubplot at 0x950ec08>

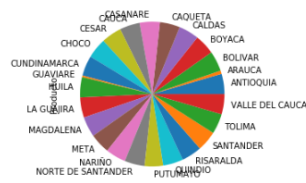


```
In [58]: Producto_Counts.plot(kind='line')
```

Out[58]: <matplotlib.axes._subplots.AxesSubplot at 0x963f988>



```
In [59]: Producto_Counts.plot(kind='pie')
Out[59]: <matplotlib.axes._subplots.AxesSubplot at 0x96cde08>
```



```
In [60]: Produccion_df.index
# Muestra la cantidad de datos por columnas
Out[60]: RangeIndex(start=0, stop=266, step=1)
```

```
In [61]: Produccion_df.size
# Muestra el numero de elementos del dataframe
Out[61]: 2128
```

```
In [62]: Produccion_df.info
#Devuelve información sobre el dataframe
```

```
Out[62]: <bound method DataFrame.info of
0 2007 ANTIOQUIA CAFE 112,343.60 120,500.00
1 2007 BOLIVAR CAFE 502.00 446.00
2 2007 BOYACA CAFE 11,374.50 9,683.10
3 2007 CALDAS CAFE 78,393.65 92,815.00
4 2007 CAQUETA CAFE 2,295.00 2,134.00
.. ..
261 2018 QUINDIO CAFE 16,374.73 17,739.03
262 2018 RISARALDA CAFE 35,874.73 45,918.75
263 2018 SANTANDER CAFE 42,269.07 55,918.71
264 2018 TOLIMA CAFE 97,304.04 97,451.31
265 2018 VALLE DEL CAUCA CAFE 40,305.31 49,667.88

Rendimiento (ha/ton) Produccion Nacional (ton) Area Nacional (ha)
0 1.07 14.54 14.66
1 0.89 0.05 0.07
2 0.85 1.17 1.48
3 1.18 11.20 10.23
4 0.93 0.26 0.30
.. ..
261 1.08 2.07 2.21
262 1.28 5.37 4.83
263 1.32 6.53 5.69
264 1.00 11.39 13.11
265 1.03 5.80 6.51

[266 rows x 8 columns]>
```

```
In [63]: Produccion_df.iloc[1,:32]
# Devuelve un dataframe con los elementos de las filas de la lista y de las columnas de la lista
```

```
Out[63]: Anio 2007
Departamento BOLIVAR
Producto CAFE
Area (ha) 502.00
Produccion (ton) 446.00
Rendimiento (ha/ton) 0.89
Produccion Nacional (ton) 0.05
Area Nacional (ha) 0.07
Name: 1, dtype: object
```

```
In [64]: Produccion_df.loc[31,'Departamento']
# Devuelve un dataframe con los elementos que se encuentra en la fila 32 con los nombres de la lista Departamento
```

```
Out[64]: 0 ANTIOQUIA
1 BOLIVAR
2 BOYACA
3 CALDAS
4 CAQUETA
5 CASANARE
6 CAUCA
7 CESAR
8 CHOCO
9 CUNDINAMARCA
10 HUILA
11 LA GUAJIRA
12 MAGDALENA
13 META
14 NARIÑO
15 NORTE DE SANTANDER
16 PUTUMAYO
17 QUINDIO
18 RISARALDA
19 SANTANDER
20 TOLIMA
21 VALLE DEL CAUCA
22 ANTIOQUIA
23 BOLIVAR
24 BOYACA
25 CALDAS
26 CAQUETA
27 CASANARE
28 CAUCA
```


In [65]: `Produccion_df.isnull().sum()`
Muestra si en el datrame hay un valor vacio

Out[65]:

Anio	0
Departamento	0
Producto	0
Area (ha)	0
Produccion (ton)	0
Rendimiento (ha/ton)	0
Produccion Nacional (ton)	0
Area Nacional (ha)	0

dtype: int64

In [66]: `Produccion_df.duplicated().sum()`
Muestra si el dataframe presenta un valor duplicado

Out[66]: 0

In [67]: `Produccion_df[1:40]`
Muestra Los valores de La fila 1 hasta La fila 40

Out[67]:

	Anio	Departamento	Producto	Area (ha)	Produccion (ton)	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
1	2007	BOLIVAR	CAFE	502.00	446.00	0.89	0.05	0.07
2	2007	BOYACA	CAFE	11,374.50	9,663.10	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	QUIINDIO	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

CONCLUSION

El café es importante en la economía ya que este se exporta a una gran cantidad de países y además es un producto que casi nunca falta en los hogares colombianos

BIBLIOGRAFIA

• Holmes, E., Dawn. Big Data: Una breve introducción. Disponible en: <https://ebookcentral-proquest-com.bdigital.sena.edu.co/lib/senavirtualsp/detail.action?docID=5756204>

• García, J., Bustamante. Ciencia de Datos. Editorial Alfaomega. Disponible en: <https://www-alfaomegacloud-com.bdigital.sena.edu.co/reader/ciencia-de-datos?location=13>

• Redondo, G., Carlos. El programa R, herramienta clave en investigación. Disponible en: <https://ebookcentral-proquest-com.bdigital.sena.edu.co/lib/senavirtualsp/detail.action?docID=5214094&query=rstudio>

• Marques, F. R en profundidad Programación, gráficos y estadística. Editorial AlfaOmega. Disponible en: https://www-alfaomegacloud-com.bdigital.sena.edu.co/auth/ip?intended_url=https://www-lfaomegacloudcom.bdigital.sena.edu.co/reader/r-en-profundidad?location=2

• Charte, O., Francisco. Análisis exploratorio y visualización de datos con R. Disponible en: <http://www.fcharte.com/libros/ExploraVisualizaConR-Fcharte.pdf>

• Gil, J., Carlos. R para profesionales de los datos: Una introducción. Disponible en: https://www.datanalytics.com/libro_r/main.pdf

• Monserat, G., Alcina. Big Data: gestión y explotación de grandes volúmenes de datos. Disponible en: <https://ebookcentral-proquest-com.bdigital.sena.edu.co/lib/senavirtualsp/detail.action?docID=5308389&query=BIG+DATA+CON+R>

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

Etapas

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