In [1]:
```
# EJERCICIO PRÁCTICA PARA EL ANÁLISIS DEL DATAFRAME DE LA PRODUCCIÓN DE CAFÉ EN C
# ELABORÓ Ing. Luis Armando Amaya Q.
# Se incluyen en este ejercicio práctico algunas instrucciones vistas, puede util
# Recuerde incluir todos los resultados que considere importantes y necesarios
# También la presentación de resultados mediante gráficos
# EL DATAFRAME SE ENCUENTRA EN LA CARPETA del drive en Google creada subcartpeta
# TE LO DEJADO LISTO PARA QUE LO PUEDAS UTILIZAR CON TUS CONOCIMIENTOS Y CREATIVI
# Este ejercicio le permite enriquecer su conocimiento y aportar en la construcci
# IMPORTATE: Debes DOCUMENTAR CADA LÍNEA CON UN COMENTARIO utilizando el simbolo
# IMPORTANTE: UTILICE SUS CONOCIMIENTOS Y CREATIVIDAD PARA "REALIZAR MINERIA DE D
# ESTE EJERICIO ES UNA GUÍA, NO UN INFORME DETALLADO, POR FAVOR, UTILIZARLO COMO
# SOLAMENTE USTED MISMO SE PONE LOS LÍMITES, ASI QUE DEBES EXIGIRTE A TI MISMO.
```

In [1]:
```
# Recuerda que en INTERNET hay MUCHAS más instrucciones que puedes APLICAR en tu

# Solamente te he mostrado algunas instrucciones importantes, pero hay muchísimas
```

In [2]:
```
import pandas as pd
# Importar la libreria PANDAS
```

In [ ]:
```

```

In [3]:
```
pd.read_csv("PRODUCCIONc.csv")
#Lectura del Dataframe
```

Out[3]:

	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 [4]:
```
produccion_df=pd.read_csv("produccionc.csv")
# Asignación del nombre del Dataframe
```

In [5]:
```
produccion_df
# listado general del Dataframe produccion
```

Out[5]:

	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 [6]:
```
type(produccion_df)
```

Out[6]:
```
pandas.core.frame.DataFrame
```

In [7]:
```
produccion_df.dtypes
```

Out[7]:
```
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 [8]:
```
pd.unique(produccion_df['Anio'])
#indica los valores de los años en el dataframe
```

Out[8]:
```
array([2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017,
       2018], dtype=int64)
```

In [9]:
```
pd.unique(produccion_df['Departamento'])
#indica los departamentos en el dataframe
```

Out[9]:
```
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 [10]:
```
pd.unique(produccion_df['Producto'])
#indica los productos que están en el dataframe
```

Out[10]:
```
array(['CAFE'], dtype=object)
```

In [11]:
```
pd.unique(produccion_df['Area (ha)'])
```

Out[11]:
```
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',
       '17,686.00', '2,783.00', '27,806.40', '19,339.31', '42.00',
       '21,109.83', '45,588.03', '33,947.15', '90,904.48', '69,456.71',
       '109,755.50', '659.04', '9,289.05', '60,264.29', '2,905.84',
       '2,232.94', '74,105.64', '25,106.39', '125.01', '36,189.18',
       '118,200.88', '5,750.70', '17,016.72', '2,483.43', '32,136.51',
       '25,332.45', '24.27', '21,203.03', '39,615.60', '38,613.68',
       '97,308.81', '53,481.02', '110,115.86', '936.34', '9,834.39',
       '59,757.18', '3,074.92', '2,599.43', '77,068.46', '26,138.58',
       '136.88', '33,623.54', '128,273.15', '6,078.64', '18,533.11',
       '2,739.71', '33,608.32', '23,724.20', '101.16', '21,462.81',
       '40,154.46', '40,733.20', '100,832.91', '56,035.94', '109,649.61',
       '1,065.07', '10,461.85', '58,376.40', '3,410.56', '2,752.31',
       '77,405.83', '25,948.50', '137.47', '34,101.49', '130,452.40',
       '5,631.53', '17,996.31', '2,922.21', '33,490.93', '22,940.64',
       '128.65', '21,491.21', '41,732.03', '42,679.11', '103,368.73',
       '54,938.79', '105,666.60', '1,065.97', '10,181.80', '56,022.04',
       '3,392.22', '2,671.04', '78,421.95', '25,530.59', '134.96',
       '33,214.17', '126,052.15', '5,531.20', '17,745.80', '2,924.89',
       '32,750.16', '21,520.64', '20,041.70', '40,472.26', '41,387.79',
       '100,328.77', '52,648.25', '99,311.53', '1,137.42', '9,598.33',
       '51,854.59', '3,408.69', '2,436.63', '80,289.56', '25,158.80',
       '125.67', '30,894.16', '122,575.76', '5,340.80', '18,129.50',
       '2,926.85', '33,639.55', '21,409.77', '209.29', '17,699.67',
       '37,334.16', '42,327.26', '96,018.89', '51,470.86', '98,038.15',
       '1,182.13', '9,653.45', '50,762.22', '3,485.24', '2,360.55',
       '82,085.54', '23,915.45', '140.33', '29,085.24', '122,002.46',
       '4,810.97', '17,414.32', '2,761.01', '33,465.54', '20,873.04',
       '209.93', '16,374.73', '35,874.73', '42,269.07', '97,304.04',
       '48,305.31'], dtype=object)
```

In [12]:
```
pd.unique(produccion_df['Produccion (ton)'])
```

Out[12]:
```
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',
       '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',
       '3,434.30', '14,096.05', '2,133.10', '28,077.94', '12,214.54',
       '48.40', '18,030.13', '36,989.43', '23,271.89', '85,027.49',
       '61,190.55', '102,403.24', '395.07', '5,591.05', '58,634.19',
       '2,188.92', '1,338.56', '56,303.92', '15,050.27', '105.93',
       '24,993.74', '115,874.98', '3,447.31', '10,200.84', '1,650.41',
       '28,606.96', '15,185.79', '16.87', '20,599.27', '39,073.92',
       '30,227.02', '77,215.36', '42,948.40', '111,452.91', '606.93',
       '6,364.41', '62,869.38', '2,503.81', '1,688.60', '63,365.76',
       '16,935.63', '125.42', '25,118.55', '135,971.20', '3,923.80',
       '12,012.98', '1,950.84', '32,321.56', '15,108.55', '76.04',
       '22,518.42', '42,719.53', '34,512.79', '86,453.62', '49,799.28',
       '120,365.77', '1,089.74', '9,501.54', '67,231.37', '3,749.27',
       '2,626.73', '83,626.44', '22,240.81', '158.20', '31,165.15',
       '145,168.10', '4,317.50', '16,691.31', '3,206.35', '36,607.56',
       '20,267.64', '124.67', '24,694.56', '47,215.69', '47,304.16',
       '105,563.88', '57,583.56', '119,970.68', '1,128.32', '9,583.80',
       '66,661.14', '3,861.63', '2,638.88', '87,642.49', '22,649.03',
       '160.62', '31,413.34', '145,154.42', '4,387.19', '17,031.09',
       '3,322.42', '37,020.90', '19,590.10', '23,791.30', '47,357.02',
       '47,512.36', '105,976.19', '57,067.08', '140,398.62', '748.97',
       '7,638.99', '68,668.20', '5,108.33', '1,747.51', '97,922.49',
       '16,628.14', '158.85', '33,943.39', '133,787.95', '3,516.80',
       '11,937.90', '4,013.11', '35,004.18', '23,409.44', '282.18',
       '18,792.05', '46,779.71', '54,908.68', '94,556.71', '51,687.80',
       '141,898.91', '734.91', '7,780.34', '68,670.96', '5,280.40',
       '1,629.25', '102,147.00', '14,943.62', '181.42', '32,580.24',
       '136,161.86', '2,990.91', '10,826.24', '3,877.62', '35,679.42',
       '23,471.69', '289.50', '17,739.03', '45,918.75', '55,918.71',
       '97,451.31', '49,667.88'], dtype=object)
```

In [13]:
```
pd.unique(produccion_df['Rendimiento (ha/ton)'])
```

Out[13]:
```
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 [14]:
```
pd.unique(produccion_df['Produccion Nacional (ton)'])
```

Out[14]:
```
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, 3.900e-01, 3.200e-01, 6.500e+00, 1.720e+00,
       5.120e+00, 1.330e+01, 2.080e+00, 4.000e-01, 3.760e+00, 1.930e+00,
       3.250e+00, 7.660e+00, 3.450e+00, 8.320e+00, 1.022e+01, 1.462e+01,
       1.000e-01, 7.900e-01, 8.630e+00, 2.700e-01, 8.070e+00, 3.190e+00,
       4.910e+00, 1.360e+01, 5.500e-01, 2.250e+00, 3.400e-01, 4.480e+00,
       1.950e+00, 2.880e+00, 5.900e+00, 1.357e+01, 9.760e+00, 1.570e+01,
       6.000e-02, 8.600e-01, 8.990e+00, 2.100e-01, 2.310e+00, 1.777e+01,
       5.300e-01, 1.560e+00, 4.390e+00, 2.330e+00, 5.990e+00, 4.640e+00,
       1.184e+01, 6.590e+00, 1.530e+01, 8.700e-01, 2.300e-01, 8.700e+00,
       1.867e+01, 5.400e-01, 1.650e+00, 4.440e+00, 2.070e+00, 3.090e+00,
       5.860e+00, 4.740e+00, 1.187e+01, 6.840e+00, 1.415e+01, 1.300e-01,
       1.120e+00, 7.900e+00, 4.400e-01, 9.830e+00, 2.620e+00, 3.660e+00,
       1.707e+01, 5.100e-01, 1.960e+00, 3.800e-01, 4.300e+00, 2.380e+00,
       2.900e+00, 5.550e+00, 5.560e+00, 1.241e+01, 6.770e+00, 1.405e+01,
       7.810e+00, 4.500e-01, 1.026e+01, 2.650e+00, 3.680e+00, 1.700e+01,
       1.990e+00, 4.340e+00, 2.290e+00, 2.790e+00, 6.680e+00, 1.649e+01,
       9.000e-01, 8.060e+00, 6.000e-01, 1.150e+01, 3.990e+00, 1.571e+01,
       4.100e-01, 1.400e+00, 4.700e-01, 4.110e+00, 2.750e+00, 3.000e-02,
       2.210e+00, 5.490e+00, 6.450e+00, 1.110e+01, 6.070e+00, 1.658e+01,
       8.020e+00, 6.200e-01, 1.900e-01, 1.194e+01, 3.810e+00, 1.591e+01,
       3.500e-01, 1.260e+00, 4.170e+00, 2.740e+00, 5.370e+00, 6.530e+00,
       1.139e+01])
```

In [15]:
```
pd.unique(produccion_df['Area Nacional (ha)'])
```

Out[15]:
```
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,
       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, 9.690e+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,
       1.310e+00, 7.290e+00, 4.300e-01, 9.660e+00, 3.240e+00, 4.260e+00,
       1.628e+01, 7.000e-01, 2.250e+00, 4.180e+00, 2.860e+00, 2.680e+00,
       5.210e+00, 5.330e+00, 1.290e+01, 6.860e+00, 1.359e+01, 1.400e-01,
       7.200e+00, 4.400e-01, 1.008e+01, 4.270e+00, 1.621e+01, 7.100e-01,
       4.210e+00, 2.770e+00, 5.200e+00, 5.320e+00, 6.770e+00, 1.318e+01,
       1.500e-01, 6.880e+00, 4.500e-01, 1.066e+01, 3.340e+00, 4.100e+00,
       1.627e+01, 2.410e+00, 4.470e+00, 2.840e+00, 3.000e-02, 2.350e+00,
       4.960e+00, 5.620e+00, 1.275e+01, 6.830e+00, 1.321e+01, 1.600e-01,
       1.300e+00, 6.840e+00, 4.700e-01, 3.920e+00, 1.643e+01, 6.500e-01,
       3.700e-01, 2.810e+00, 2.210e+00, 4.830e+00, 5.690e+00, 1.311e+01,
       6.510e+00])
```

In [16]:
```
produccion_df['Anio'].min()
```

Out[16]:
```
2007
```

In [17]:
```
produccion_df['Anio'].max()
```

Out[17]:
```
2018
```

In [18]:
```
produccion_df['Area (ha)'].min()
```

Out[18]:
```
'0.00'
```

In [20]:
```
produccion_df['Area (ha)'].max()+" Hectarea"
```

Out[20]:
```
'99,311.53 Hectarea'
```

In [21]:
```
produccion_df['Rendimiento (ha/ton)'].min()
```

Out[21]:
```
0.0
```

In [22]:
```
produccion_df['Rendimiento (ha/ton)'].max()
```

Out[22]:
```
2.0
```

In [23]:
```
produccion_df['Anio'].isnull()
```

Out[23]:
```
0      False
1      False
2      False
3      False
4      False
       ...
261    False
262    False
263    False
264    False
265    False
Name: Anio, Length: 266, dtype: bool
```

In [24]:
```
produccion_df['Area (ha)'].isnull()
```

Out[24]:
```
0      False
1      False
2      False
3      False
4      False
       ...
261    False
262    False
263    False
264    False
265    False
Name: Area (ha), Length: 266, dtype: bool
```

In [25]:
```
produccion_df['Rendimiento (ha/ton)'].isnull()
```

Out[25]:
```
0      False
1      False
2      False
3      False
4      False
       ...
261    False
262    False
263    False
264    False
265    False
Name: Rendimiento (ha/ton), Length: 266, dtype: bool
```

In [26]:
```
produccion_df['Rendimiento (ha/ton)'].isnull().sum()
```

Out[26]: 0

In [27]:
```
produccion_df['Area (ha)'].isnull().sum()
```

Out[27]: 0

In [28]:
```
produccion_grouped_Anio=produccion_df.groupby("Anio").sum()
produccion_grouped_Anio
```

Out[28]:

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

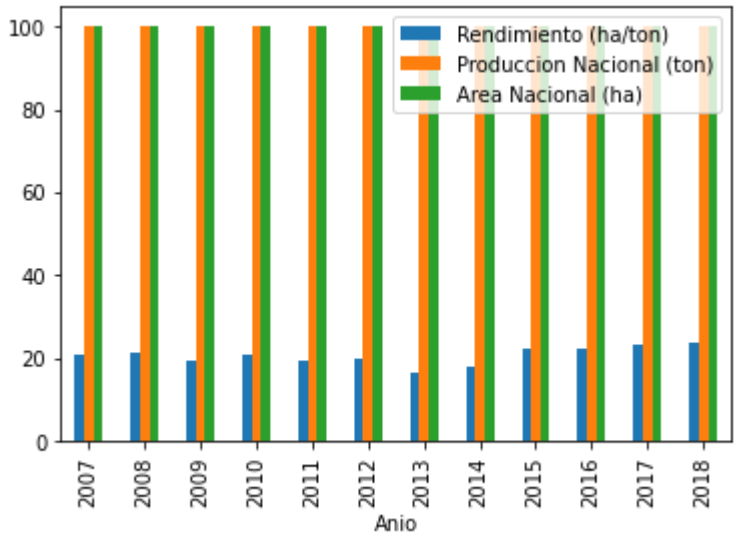
In [29]:
```
produccion_grouped_Anio2=produccion_df.groupby("Anio").sum()
produccion_grouped_Anio2
```

Out[29]:

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 [30]:
```
import numpy as np
import re
import sys
%matplotlib inline
produccion_grouped_Anio2.plot(kind='bar')
```

Out[30]: <matplotlib.axes._subplots.AxesSubplot at 0x191578af670>



In [31]:
```
produccion_grouped_Departamento=produccion_df.groupby(["Anio","Departamento"]).su
produccion_grouped_Departamento
```

Out[31]:

		Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
Anio	**Departamento**			
2007	**ANTIOQUIA**	1.07	14.54	14.66
	BOLIVAR	0.89	0.05	0.07
	BOYACA	0.85	1.17	1.48
	CALDAS	1.18	11.20	10.23
	CAQUETA	0.93	0.26	0.30
...	**...**	...	...	...
2018	**QUINDIO**	1.08	2.07	2.21
	RISARALDA	1.28	5.37	4.83
	SANTANDER	1.32	6.53	5.69
	TOLIMA	1.00	11.39	13.11
	VALLE DEL CAUCA	1.03	5.80	6.51

266 rows × 3 columns

In [32]:
```
produccion_grouped_Departamento_Rendimiento=produccion_df.groupby(["Anio","Depart
produccion_grouped_Departamento_Rendimiento
```

Out[32]:

			Producto	Area (ha)	Produccion (ton)	Produccion Nacional (ton)	Area Nacional (ha)
Anio	**Departamento**	**Rendimiento (ha/ton)**					
2007	**ANTIOQUIA**	**1.07**	CAFE	112,343.60	120,500.80	14.54	14.66
	BOLIVAR	**0.89**	CAFE	502.00	446.00	0.05	0.07
	BOYACA	**0.85**	CAFE	11,374.50	9,683.10	1.17	1.48
	CALDAS	**1.18**	CAFE	78,393.65	92,815.00	11.20	10.23
	CAQUETA	**0.93**	CAFE	2,295.00	2,134.00	0.26	0.30
...	**...**	**...**	...	...	...	...	...
2018	**QUINDIO**	**1.08**	CAFE	16,374.73	17,739.03	2.07	2.21
	RISARALDA	**1.28**	CAFE	35,874.73	45,918.75	5.37	4.83
	SANTANDER	**1.32**	CAFE	42,269.07	55,918.71	6.53	5.69

In [53]:
```
produccion_df['Produccion (ton)'].count()
# cuenta el numero de registros en el dataframe para el campo de la Producción
```

Out[53]: 266

In [54]:
```
produccion_df['Anio'].count()
# cuenta el numero de registros en el dataframe para el campo del año
```

Out[54]: 266

In [55]:
```
produccion_grouped_Departamento1=produccion_df.groupby(["Anio","Departamento"]).s
produccion_grouped_Departamento1
```

Out[55]:
```
Rendimiento (ha/ton)          2.00
Produccion Nacional (ton)    18.67
Area Nacional (ha)           16.43
dtype: float64
```

In [56]:
```
produccion_grouped_Departamento=produccion_df.groupby(["Anio","Departamento"]).su
produccion_grouped_Departamento
```

Out[56]:
```
Rendimiento (ha/ton)         0.0
Produccion Nacional (ton)    0.0
Area Nacional (ha)           0.0
dtype: float64
```

In [57]:
```
produccion_grouped_Departamento=produccion_df.groupby(["Anio","Departamento"]).de
produccion_grouped_Departamento
```

Out[57]:

		Rendimiento (ha/ton)								Produccion Nacional (ton)				
		count	mean	std	min	25%	50%	75%	max	count	mean	...	75%	ma
Anio	**Departamento**													
2007	**ANTIOQUIA**	1.0	1.07	NaN	1.07	1.07	1.07	1.07	1.07	1.0	14.54	...	14.54	14.
	BOLIVAR	1.0	0.89	NaN	0.89	0.89	0.89	0.89	0.89	1.0	0.05	...	0.05	0.
	BOYACA	1.0	0.85	NaN	0.85	0.85	0.85	0.85	0.85	1.0	1.17	...	1.17	1.
	CALDAS	1.0	1.18	NaN	1.18	1.18	1.18	1.18	1.18	1.0	11.20	...	11.20	11.
	CAQUETA	1.0	0.93	NaN	0.93	0.93	0.93	0.93	0.93	1.0	0.26	...	0.26	0.
...	**...**	...	...	...	...	...	...	...	...	...	...	...	...	
2018	**QUINDIO**	1.0	1.08	NaN	1.08	1.08	1.08	1.08	1.08	1.0	2.07	...	2.07	2.
	RISARALDA	1.0	1.28	NaN	1.28	1.28	1.28	1.28	1.28	1.0	5.37	...	5.37	5.
	SANTANDER	1.0	1.32	NaN	1.32	1.32	1.32	1.32	1.32	1.0	6.53	...	6.53	6.
	TOLIMA	1.0	1.00	NaN	1.00	1.00	1.00	1.00	1.00	1.0	11.39	...	11.39	11.
	VALLE DEL CAUCA	1.0	1.03	NaN	1.03	1.03	1.03	1.03	1.03	1.0	5.80	...	5.80	5.

266 rows × 24 columns

In [58]:
```
produccion_Departamento3=produccion_df.groupby(["Anio","Departamento", "Area Naci
produccion_Departamento3
```

Out[58]:

			Rendimiento (ha/ton)								Produccion Nacion		
			count	mean	std	min	25%	50%	75%	max	count	mean	std
Anio	Departamento	Area Nacional (ha)											
2007	ANTIOQUIA	14.66	1.0	1.07	NaN	1.07	1.07	1.07	1.07	1.07	1.0	14.54	NaN
	BOLIVAR	0.07	1.0	0.89	NaN	0.89	0.89	0.89	0.89	0.89	1.0	0.05	NaN
	BOYACA	1.48	1.0	0.85	NaN	0.85	0.85	0.85	0.85	0.85	1.0	1.17	NaN
	CALDAS	10.23	1.0	1.18	NaN	1.18	1.18	1.18	1.18	1.18	1.0	11.20	NaN
	CAQUETA	0.30	1.0	0.93	NaN	0.93	0.93	0.93	0.93	0.93	1.0	0.26	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...
2018	QUINDIO	2.21	1.0	1.08	NaN	1.08	1.08	1.08	1.08	1.08	1.0	2.07	NaN
	RISARALDA	4.83	1.0	1.28	NaN	1.28	1.28	1.28	1.28	1.28	1.0	5.37	NaN
	SANTANDER	5.69	1.0	1.32	NaN	1.32	1.32	1.32	1.32	1.32	1.0	6.53	NaN
	TOLIMA	13.11	1.0	1.00	NaN	1.00	1.00	1.00	1.00	1.00	1.0	11.39	NaN
	VALLE DEL CAUCA	6.51	1.0	1.03	NaN	1.03	1.03	1.03	1.03	1.03	1.0	5.80	NaN

266 rows × 16 columns

In [59]:
```
produccion_Anio=produccion_df.groupby(["Anio"]).describe()
produccion_Anio
```

Out[59]:

	Rendimiento (ha/ton)								Produccion Nacional (ton)				
	count	mean	std	min	25%	50%	75%	max	count	mean	...	75%	m
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
2008	22.0	0.982727	0.322670	0.45	0.7775	0.905	1.2000	1.79	22.0	4.545455	...	7.7600	15
2009	22.0	0.881364	0.264652	0.30	0.7600	0.930	1.1125	1.21	22.0	4.545455	...	7.3425	14
2010	23.0	0.906087	0.324692	0.00	0.7050	0.960	1.1250	1.52	23.0	4.348261	...	7.3550	15
2011	23.0	0.854348	0.238305	0.47	0.6100	0.900	1.0550	1.20	23.0	4.348696	...	7.0800	18
2012	23.0	0.858696	0.329618	0.00	0.7450	0.830	0.9150	2.00	23.0	4.347391	...	6.9850	14
2013	22.0	0.759545	0.145421	0.60	0.6000	0.755	0.8800	0.99	22.0	4.545455	...	6.4400	17
2014	22.0	0.822273	0.157629	0.64	0.6500	0.815	0.9500	1.06	22.0	4.545455	...	6.5950	18
2015	22.0	1.024545	0.110096	0.77	0.9350	1.065	1.1075	1.15	22.0	4.544545	...	6.4675	17
2016	21.0	1.063810	0.116725	0.79	0.9600	1.120	1.1500	1.19	21.0	4.761429	...	6.6800	17
2017	22.0	1.068182	0.272443	0.66	0.8450	1.090	1.2900	1.50	22.0	4.545909	...	6.3550	16
2018	22.0	1.079545	0.296672	0.62	0.8575	1.120	1.3125	1.52	22.0	4.545455	...	6.3475	16

12 rows × 24 columns

In [60]:
```
produccion_Anio2_Rendimiento=produccion_df.groupby(["Rendimiento (ha/ton)"]).desc
produccion_Anio2_Rendimiento
```

Out[60]:

	Anio								Produccion Nacional		
	count	mean	std	min	25%	50%	75%	max	count	mean	...
Rendimiento (ha/ton)											
0.00	2.0	2011.0	1.414214	2010.0	2010.50	2011.0	2011.50	2012.0	2.0	0.000	...
0.30	1.0	2009.0	NaN	2009.0	2009.00	2009.0	2009.00	2009.0	1.0	1.440	...
0.38	1.0	2009.0	NaN	2009.0	2009.00	2009.0	2009.00	2009.0	1.0	0.040	...
0.45	2.0	2007.5	0.707107	2007.0	2007.25	2007.5	2007.75	2008.0	2.0	1.640	...
0.47	1.0	2011.0	NaN	2011.0	2011.00	2011.0	2011.00	2011.0	1.0	0.300	...
...	...	...	...	...	...	...	...	...	...	...	...
1.50	1.0	2017.0	NaN	2017.0	2017.00	2017.0	2017.00	2017.0	1.0	0.600	...
1.52	2.0	2014.0	5.656854	2010.0	2012.00	2014.0	2016.00	2018.0	2.0	4.935	...
1.53	1.0	2007.0	NaN	2007.0	2007.00	2007.0	2007.00	2007.0	1.0	8.790	...
1.79	1.0	2008.0	NaN	2008.0	2008.00	2008.0	2008.00	2008.0	1.0	9.440	...
2.00	1.0	2012.0	NaN	2012.0	2012.00	2012.0	2012.00	2012.0	1.0	0.020	...

94 rows × 24 columns

In [61]:
```
produccion_df.describe()
# Indica datos estadísticos generales del dataframe produccion desde el año 2007
```

Out[61]:

	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 [62]:
```
produccion_df["Produccion Nacional (ton)"].describe()
# Indica datos estadísticos generales para la Producción nacional del dataframe p
```

Out[62]:
```
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 [63]:
```
produccion_df["Area Nacional (ha)"].describe()
# Indica datos estadísticos generales para el Area Nacional del dataframe produc
```

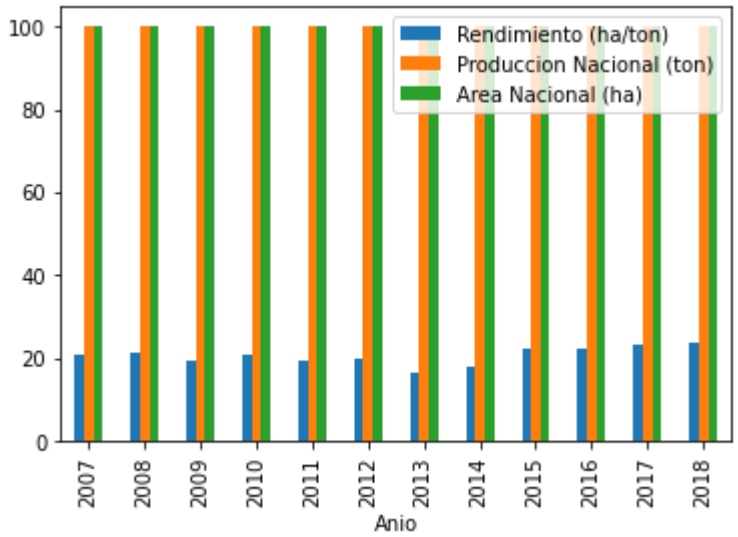
Out[63]:
```
count    266.000000
mean       4.511203
std        4.565865
min        0.000000
25%        0.390000
50%        3.120000
75%        6.875000
max       16.430000
Name: Area Nacional (ha), dtype: float64
```

In [64]:
```
produccion_df.describe()
produccion_df.mean()
# Indica el promedio del dataframe produccion para Rendimiento, Produccion y el A
```

Out[64]:
```
Anio                         2012.469925
Rendimiento (ha/ton)            0.936429
Produccion Nacional (ton)       4.511316
Area Nacional (ha)              4.511203
dtype: float64
```

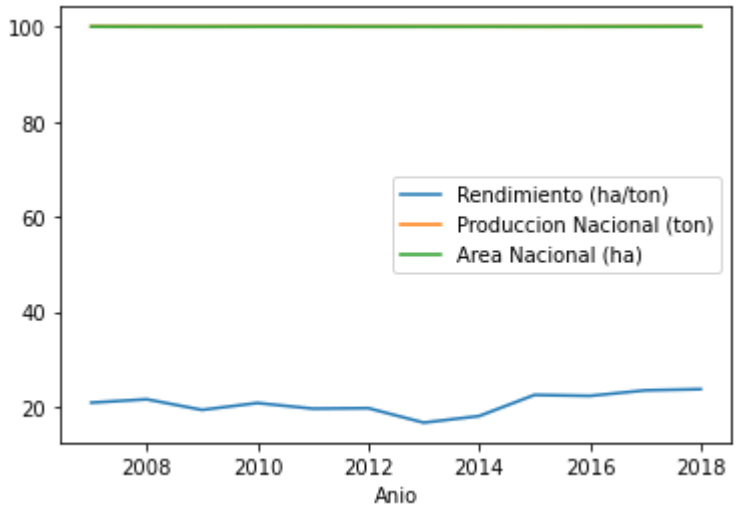
In [65]:
```
import numpy as np
import re
import sys
%matplotlib inline
produccion_grouped_Anio2.plot(kind='bar')
```

Out[65]: <matplotlib.axes._subplots.AxesSubplot at 0x19158226fa0>



In [66]:
```
import numpy as np
import re
import sys
%matplotlib inline
produccion_grouped_Anio2.plot(kind='line')
```

Out[66]: <matplotlib.axes._subplots.AxesSubplot at 0x191597935e0>

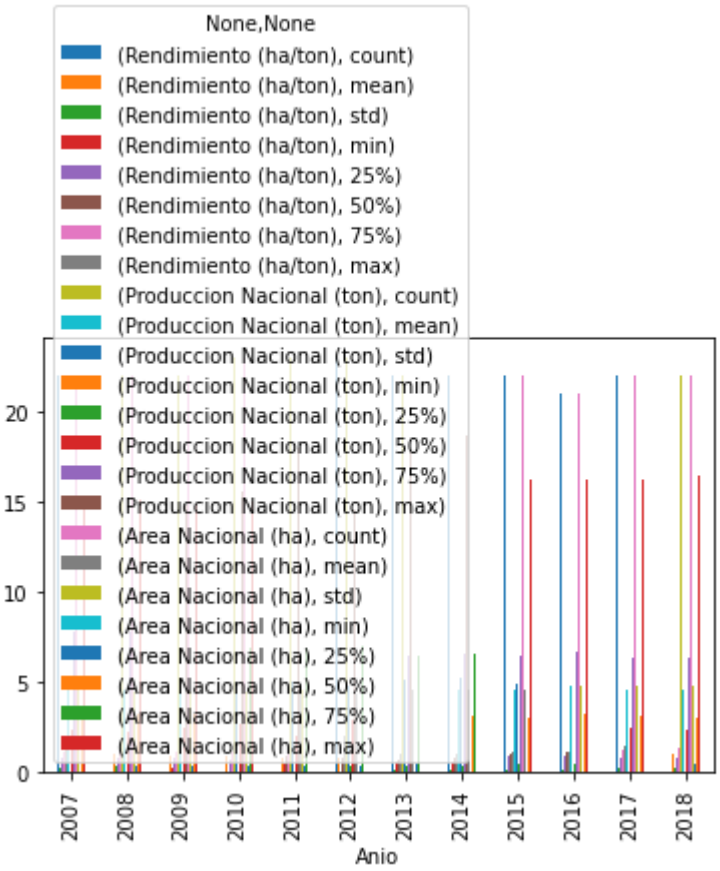


In [67]:
```
produccion_df.duplicated().sum()
#Registros que esten duplicados
```

Out[67]: 0

In [68]:
```
# Construcción del gráfico produccion por año tipo lineas
import numpy as np
import re
import sys
%matplotlib inline
produccion_Anio.plot(kind='bar')
```

Out[68]: <matplotlib.axes._subplots.AxesSubplot at 0x19159817880>

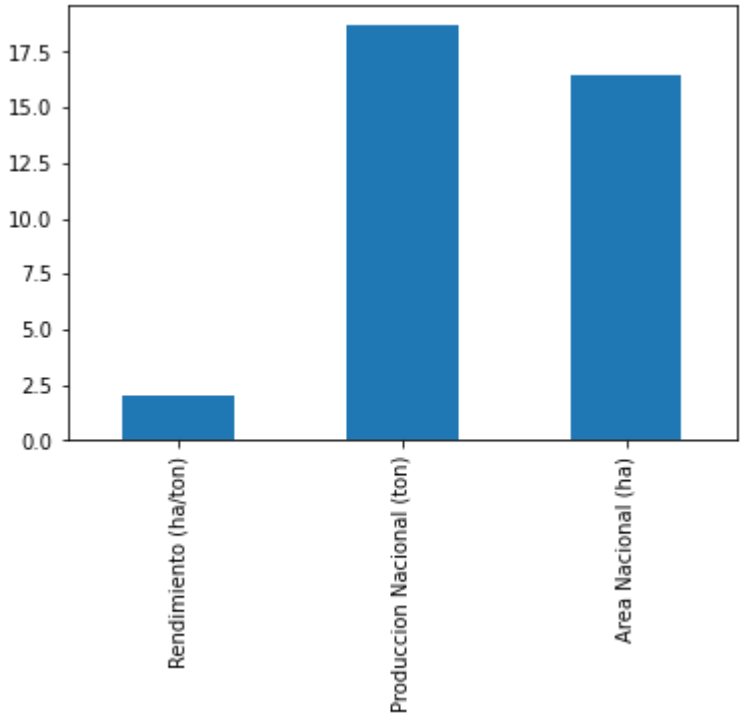


In [69]:
```
produccion_grouped_Departamento1=produccion_df.groupby(["Anio","Departamento"]).s
produccion_grouped_Departamento1
```

Out[69]:
```
Rendimiento (ha/ton)          2.00
Produccion Nacional (ton)    18.67
Area Nacional (ha)           16.43
dtype: float64
```

In [70]:
```
produccion_grouped_Departamento1.plot(kind='bar')
```

Out[70]: <matplotlib.axes._subplots.AxesSubplot at 0x19158340af0>



In [71]:
```
produccion_grouped_Departamento1.plot(kind='pie')
```

Out[71]: <matplotlib.axes._subplots.AxesSubplot at 0x19158323310>

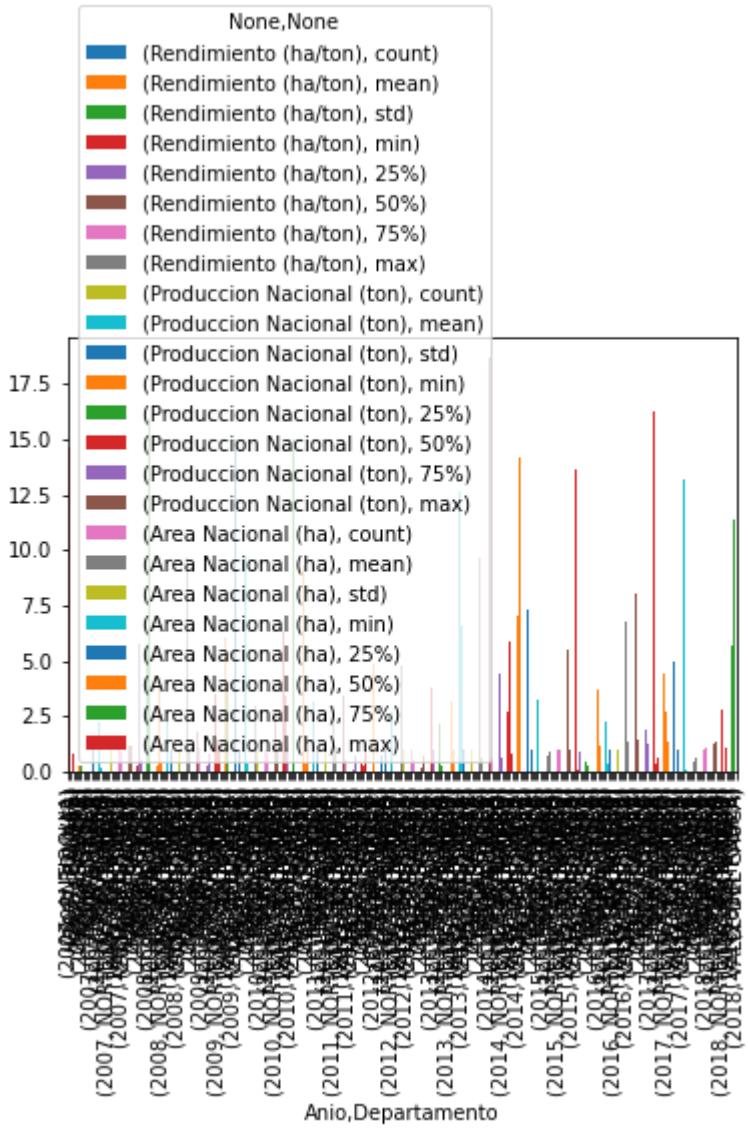
In [18]:
```
#  Construcción del gráfico Rendimiento por año tipo lineas
%matplotlib inline
produccion_Anio_Rendimiento.plot(kind='bar')
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-18-cde4741a05ea> in <module>
      1 #  Construcción del gráfico Rendimiento por año tipo lineas
      2 get_ipython().run_line_magic('matplotlib', 'inline')
----> 3 produccion_Anio_Rendimiento.plot(kind='bar')

NameError: name 'produccion_Anio_Rendimiento' is not defined
```

In [73]:
```
# Construcción del gráfico produccion por departamento año tipo lineas
%matplotlib inline
produccion_grouped_Departamento.plot(kind='bar')
```

Out[73]: <matplotlib.axes._subplots.AxesSubplot at 0x19158283ee0>



In [74]:
```
grouped_data = produccion_df.groupby("Departamento")
z=grouped_data.describe().mean()
print (z)
```

```
Anio                          count      11.083333
                              mean     2012.382576
                              std         3.479313
                              min      2007.333333
                              25%      2009.854167
                              50%      2012.375000
                              75%      2014.895833
                              max      2017.458333
Rendimiento (ha/ton)          count      11.083333
                              mean        0.889467
                              std         0.216119
                              min         0.620833
                              25%         0.769167
                              50%         0.863750
                              75%         0.986771
                              max         1.235417
Produccion Nacional (ton)     count      11.083333
                              mean        4.166733
                              std         0.719931
                              min         3.261250
                              25%         3.687812
                              50%         4.031667
                              75%         4.614271
                              max         5.387500
Area Nacional (ha)            count      11.083333
                              mean        4.166632
                              std         0.511340
                              min         3.537500
                              25%         3.758229
                              50%         4.136042
                              75%         4.588854
                              max         4.838333
dtype: float64
```

In [75]:
```
departamentos_counts = produccion_df.groupby("Departamento")["Producto"].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: Producto, dtype: int64
```

In [76]:
```
Grupos_Departamentos=produccion_df.groupby("Anio")["Departamento"].count()
print (Grupos_Departamentos)
# Indica la cantidad de Departamentos incluidos o analizados en cada uno de los A
```

```
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 [77]:
```
Departamento_Meta=produccion_df.loc[produccion_df["Departamento"]=="META"]
print (Departamento_Meta)
# Indica los resultados estadísticos por año para el Departamento Seleccionado
```

```
     Anio Departamento Producto Area (ha) Produccion (ton)  \
13   2007         META     CAFE  2,048.00         1,617.20
35   2008         META     CAFE  2,146.00         1,656.96
57   2009         META     CAFE  2,216.00         1,672.60
80   2010         META     CAFE  2,326.00         2,221.90
103  2011         META     CAFE  2,578.00         2,533.75
126  2012         META     CAFE  2,783.00         2,133.10
148  2013         META     CAFE  2,483.43         1,650.41
170  2014         META     CAFE  2,739.71         1,950.84
192  2015         META     CAFE  2,922.21         3,206.35
214  2016         META     CAFE  2,924.89         3,322.42
235  2017         META     CAFE  2,926.85         4,013.11
257  2018         META     CAFE  2,761.01         3,877.62

     Rendimiento (ha/ton)  Produccion Nacional (ton)  Area Nacional (ha)
13                   0.79                       0.20                0.27
35                   0.77                       0.20                0.28
57                   0.75                       0.24                0.29
80                   0.96                       0.29                0.31
103                  0.98                       0.40                0.36
126                  0.77                       0.34                0.39
148                  0.66                       0.25                0.32
170                  0.71                       0.27                0.34
192                  1.10                       0.38                0.36
214                  1.14                       0.39                0.38
235                  1.37                       0.47                0.39
257                  1.40                       0.45                0.37
```

In [78]:
```
Departamento_QUINDIO=produccion_df.loc[produccion_df["Departamento"]=="QUINDIO"]
print (Departamento_QUINDIO)
# Indica los resultados estadísticos por año para el Departamento Seleccionado
```

```
     Anio Departamento Producto  Area (ha) Produccion (ton)  \
17   2007      QUINDIO     CAFE  19,904.00        25,426.00
39   2008      QUINDIO     CAFE  19,571.00        23,669.00
61   2009      QUINDIO     CAFE  19,052.00        21,985.00
84   2010      QUINDIO     CAFE  18,159.00        21,065.00
107  2011      QUINDIO     CAFE  20,139.30        20,814.11
130  2012      QUINDIO     CAFE  21,109.83        18,030.13
152  2013      QUINDIO     CAFE  21,203.03        20,599.27
174  2014      QUINDIO     CAFE  21,462.81        22,518.42
196  2015      QUINDIO     CAFE  21,491.21        24,694.56
217  2016      QUINDIO     CAFE  20,041.70        23,791.30
239  2017      QUINDIO     CAFE  17,699.67        18,792.05
261  2018      QUINDIO     CAFE  16,374.73        17,739.03

     Rendimiento (ha/ton)  Produccion Nacional (ton)  Area Nacional (ha)
17                   1.28                       3.07                2.60
39                   1.21                       2.86                2.58
61                   1.15                       3.10                2.52
84                   1.16                       2.70                2.44
107                  1.03                       3.25                2.83
130                  0.85                       2.88                2.97
152                  0.97                       3.16                2.75
174                  1.05                       3.09                2.70
196                  1.15                       2.90                2.68
217                  1.19                       2.79                2.58
239                  1.06                       2.21                2.35
261                  1.08                       2.07                2.21
```

In [79]:
```
Estadística_Anio2015=produccion_df.loc[produccion_df["Anio"]== 2015]
print (Estadística_Anio2015)
# Indica los resultados estadísticos por departamento para el año 2015
```

	Anio	Departamento	Producto	Area (ha)	Produccion (ton)	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
179	2015	ANTIOQUIA	CAFE	109,649.61	120,365.77	1.10	14.15	13.69
180	2015	BOLIVAR	CAFE	1,065.07	1,089.74	1.02	0.13	0.13
181	2015	BOYACA	CAFE	10,461.85	9,501.54	0.91	1.12	1.31
182	2015	CALDAS	CAFE	58,376.40	67,231.37	1.15	7.90	7.29
183	2015	CAQUETA	CAFE	3,410.56	3,749.27	1.10	0.44	0.43
184	2015	CASANARE	CAFE	2,752.31	2,626.73	0.95	0.31	0.34
185	2015	CAUCA	CAFE	77,405.83	83,626.44	1.08	9.83	9.66
186	2015	CESAR	CAFE	25,948.50	22,240.81	0.86	2.62	3.24
187	2015	CHOCO	CAFE	137.47	158.20	1.15	0.02	0.02
188	2015	CUNDINAMARCA	CAFE	34,101.49	31,165.15	0.91	3.66	4.26
189	2015	HUILA	CAFE	130,452.40	145,168.10	1.11	17.07	16.28
190	2015	LA GUAJIRA	CAFE	5,631.53	4,317.50	0.77	0.51	0.70
191	2015	MAGDALENA	CAFE	17,996.31	16,691.31	0.93	1.96	2.25
192	2015	META	CAFE	2,922.21	3,206.35	1.10	0.38	0.36
193	2015	NARIÑO	CAFE	33,490.93	36,607.56	1.09	4.30	4.18
194	2015	NORTE DE SANTANDER	CAFE	22,940.64	20,267.64	0.88	2.38	2.86
195	2015	PUTUMAYO	CAFE	128.65	124.67	0.97	0.01	0.02
196	2015	QUINDIO	CAFE	21,491.21	24,694.56	1.15	2.90	2.68
197	2015	RISARALDA	CAFE	41,732.03	47,215.69	1.13	5.55	5.21
198	2015	SANTANDER	CAFE	42,679.11	47,304.16	1.11	5.56	5.33
199	2015	TOLIMA	CAFE	103,368.73	105,563.88	1.02	12.41	12.90
200	2015	VALLE DEL CAUCA	CAFE	54,938.79	57,583.56	1.05	6.77	6.86

In [80]:
```
Estadística_Anio2018=produccion_df.loc[produccion_df["Anio"]== 2018]
print (Estadística_Anio2018)
# Indica los resultados estadísticos por departamento para el año 2018
```

	Anio	Departamento	Producto	Area (ha)	Produccion (ton)	\
244	2018	ANTIOQUIA	CAFE	98,038.15	141,898.91	
245	2018	BOLIVAR	CAFE	1,182.13	734.91	
246	2018	BOYACA	CAFE	9,653.45	7,780.34	
247	2018	CALDAS	CAFE	50,762.22	68,670.96	
248	2018	CAQUETA	CAFE	3,485.24	5,280.40	
249	2018	CASANARE	CAFE	2,360.55	1,629.25	
250	2018	CAUCA	CAFE	82,085.54	102,147.00	
251	2018	CESAR	CAFE	23,915.45	14,943.62	
252	2018	CHOCO	CAFE	140.33	181.42	
253	2018	CUNDINAMARCA	CAFE	29,085.24	32,580.24	
254	2018	HUILA	CAFE	122,002.46	136,161.86	
255	2018	LA GUAJIRA	CAFE	4,810.97	2,990.91	
256	2018	MAGDALENA	CAFE	17,414.32	10,826.24	
257	2018	META	CAFE	2,761.01	3,877.62	
258	2018	NARIÑO	CAFE	33,465.54	35,679.42	
259	2018	NORTE DE SANTANDER	CAFE	20,873.04	23,471.69	
260	2018	PUTUMAYO	CAFE	209.93	289.50	
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	48,305.31	49,667.88	

	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
244	1.45	16.58	13.21
245	0.62	0.09	0.16
246	0.81	0.91	1.30
247	1.35	8.02	6.84
248	1.52	0.62	0.47
249	0.69	0.19	0.32
250	1.24	11.94	11.06
251	0.62	1.75	3.22
252	1.29	0.02	0.02
253	1.12	3.81	3.92
254	1.12	15.91	16.43
255	0.62	0.35	0.65
256	0.62	1.26	2.35
257	1.40	0.45	0.37
258	1.07	4.17	4.51
259	1.12	2.74	2.81
260	1.38	0.03	0.03
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

In [81]:
```
produccion_df[0:10]
#lista los primeros 10 elementos del dataframe
```

Out[81]:

	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

In [82]:
```
produccion_df[11:30]
#lista los elementos desde el 11 al 30 del dataframe
```

Out[82]:

	Anio	Departamento	Producto	Area (ha)	Produccion (ton)	Rendimiento (ha/ton)	Produccion Nacional (ton)	Area Nacional (ha)
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
25	2008	CALDAS	CAFE	74,897.00	86,884.00	1.16	10.49	9.88
26	2008	CAQUETA	CAFE	2,735.00	2,469.00	0.90	0.30	0.36
27	2008	CASANARE	CAFE	2,149.00	1,388.13	0.65	0.17	0.28
28	2008	CAUCA	CAFE	56,208.00	48,073.00	0.86	5.80	7.41
29	2008	CESAR	CAFE	23,198.00	13,841.45	0.60	1.67	3.06

In [83]:
```
# Recuerda que en INTERNET hay MUCHAS más instrucciones que puedes APLICAR en tu

# Solamente te he mostrado algunas instrucciones importantes, pero hay muchísimas
```

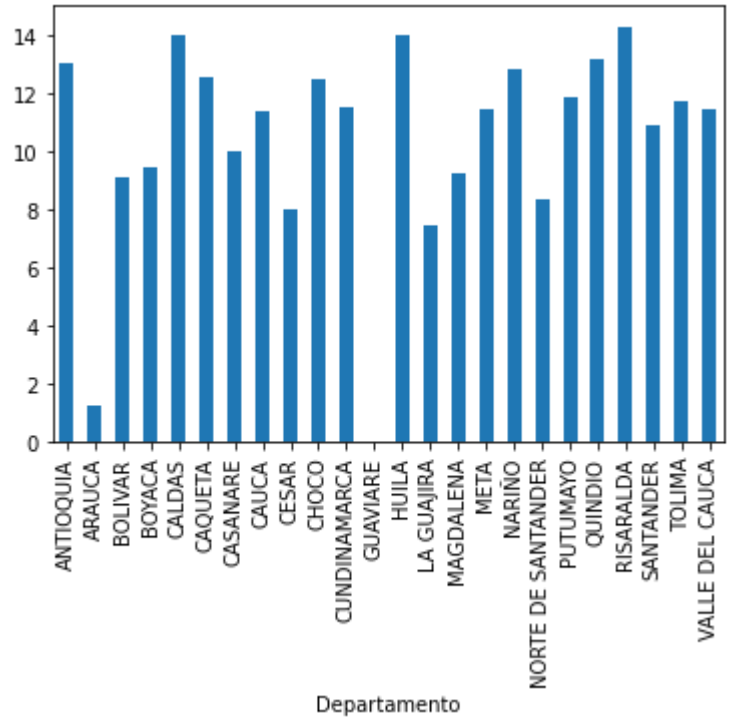
In [84]:
```
# ANIMO, ESFUERZATE. SOLO TU MISMO TE IMPONES LOS LIMITES, DEBES VENCER TODOS LOS
# INSTRUCTOR : Ing. Luis Armando Amaya Quiroga
```

In [85]:
```
Grupos_Departamentos_Rendimiento=produccion_df.groupby("Departamento")["Rendimier
Grupos_Departamentos_Rendimiento
# Indica el Rendimiento total del café en hectareas por toneladas (ha/ton) de cac
```

Out[85]:
```
Departamento
ANTIOQUIA             13.01
ARAUCA                 1.20
BOLIVAR                9.07
BOYACA                 9.41
CALDAS                14.00
CAQUETA               12.54
CASANARE              10.01
CAUCA                 11.36
CESAR                  7.95
CHOCO                 12.50
CUNDINAMARCA          11.48
GUAVIARE               0.00
HUILA                 13.98
LA GUAJIRA             7.45
MAGDALENA              9.21
META                  11.40
NARIÑO                12.79
NORTE DE SANTANDER     8.36
PUTUMAYO              11.84
QUINDIO               13.18
RISARALDA             14.29
SANTANDER             10.88
TOLIMA                11.73
VALLE DEL CAUCA       11.45
Name: Rendimiento (ha/ton), dtype: float64
```

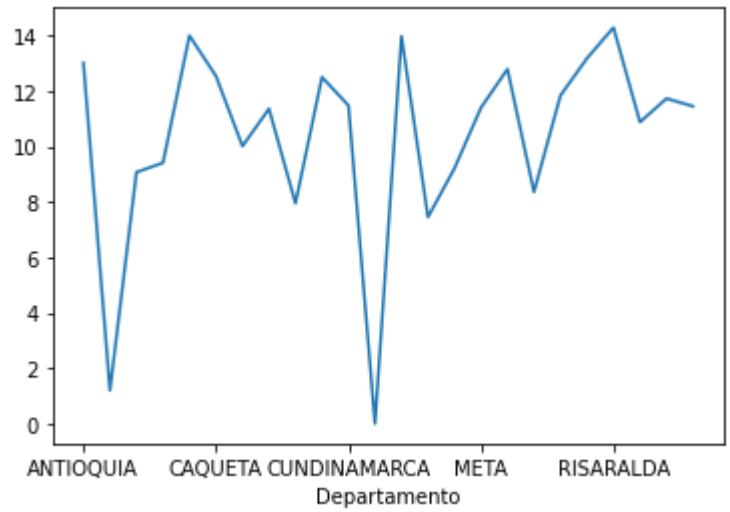
In [86]:
```
Grupos_Departamentos_Rendimiento.plot(kind='bar')
```

Out[86]: <matplotlib.axes._subplots.AxesSubplot at 0x1915d1ab250>



In [87]:
```
Grupos_Departamentos_Rendimiento.plot(kind='line')
```

Out[87]: <matplotlib.axes._subplots.AxesSubplot at 0x1915c7f0730>



In [88]:
```
Grupos_Departamentos_Produccion=produccion_df.groupby("Departamento")["Produccior
Grupos_Departamentos_Produccion
# Indica la Produccion total del café en hectareas por toneladas (ha/ton) de cadc
```

Out[88]:
```
Departamento
ANTIOQUIA             120,500.80113,505.20103,703.00121,253.38115,26...
ARAUCA                                                        0.0012.00
BOLIVAR               446.00711.00292.60510.00510.00652.50395.07606....
BOYACA                9,683.109,547.308,567.977,083.075,643.394,981....
CALDAS                92,815.0086,884.0081,668.2295,957.9078,805.875...
CAQUETA               2,134.002,469.002,332.002,902.502,528.402,446....
CASANARE              2,048.401,388.132,079.702,564.862,023.501,718....
CAUCA                 51,348.0048,073.0047,221.0045,113.0041,645.395...
CESAR                 13,278.5013,841.4512,770.0013,276.0811,035.851...
CHOCO                 205.9068.0078.7598.0098.00140.00105.93125.4215...
CUNDINAMARCA          33,729.1478,254.7737,118.0737,214.8032,780.353...
GUAVIARE                                                        0.00
HUILA                 129,052.51131,316.47104,609.42104,336.5685,150...
LA GUAJIRA            2,958.702,328.902,340.402,393.001,933.003,434....
MAGDALENA             14,005.0014,017.0013,412.8013,600.0013,301.601...
META                  1,617.201,656.961,672.602,221.902,533.752,133....
NARIÑO                31,770.0531,262.5027,487.7124,594.1024,073.952...
NORTE DE SANTANDER    13,593.2413,593.2510,221.6922,111.6512,332.001...
PUTUMAYO              34.0035.6026.7026.7045.8048.4016.8776.04124.67...
QUINDIO               25,426.0023,669.0021,985.0021,065.0020,814.111...
RISARALDA             72,842.5560,079.0053,648.0072,091.0049,042.313...
SANTANDER             29,469.5229,016.7526,311.6127,094.1622,089.822...
TOLIMA                112,322.38101,201.8888,633.1094,230.2053,288.4...
VALLE DEL CAUCA       69,618.2465,666.4362,711.0869,496.6565,475.636...
Name: Produccion (ton), dtype: object
```

In [89]:
```
produccion_df["Rendimiento (ha/ton)"].describe()
# Indica datos estadísticos generales para el Rendimiento del dataframe produccic
```

Out[89]:
```
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 [90]:
```
produccion_df["Produccion (ton)"].describe()
# Indica datos estadísticos generales para ls Produccion del dataframe produccior
```

Out[90]:
```
count       266
unique      262
top       98.00
freq          2
Name: Produccion (ton), dtype: object
```

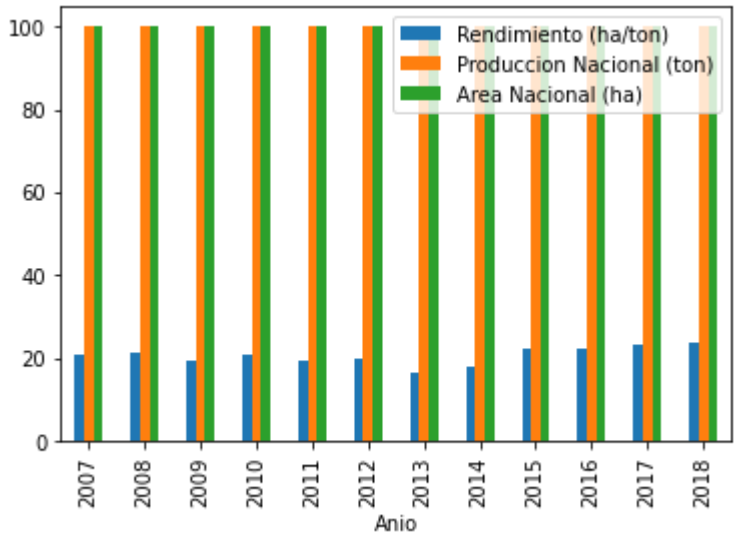
In [91]:
```
produccion_grouped_Anio4=produccion_df.groupby("Anio").sum()
produccion_grouped_Anio4
```

Out[91]:

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

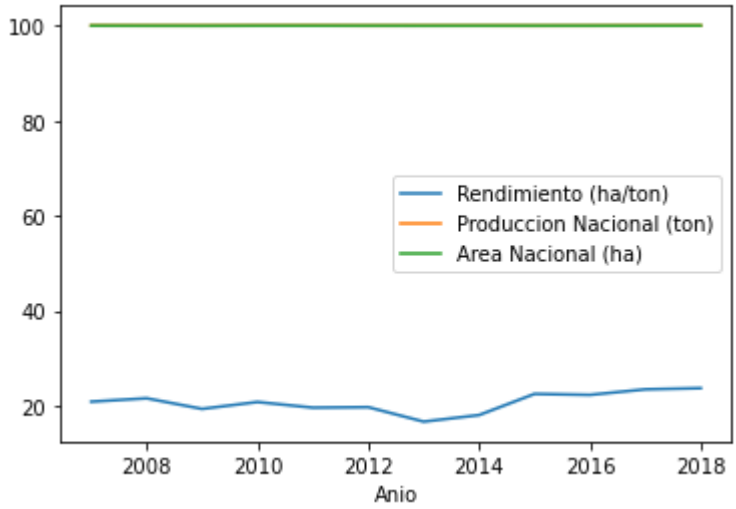
In [92]:
```
import numpy as np
import re
import sys
%matplotlib inline
produccion_grouped_Anio4.plot(kind='bar')
```

Out[92]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1915cf67610>
```



In [93]:
```
import numpy as np
import re
import sys
%matplotlib inline
produccion_grouped_Anio4.plot(kind='line')
```

Out[93]: <matplotlib.axes._subplots.AxesSubplot at 0x1915ec63070>



In [94]:
```
produccion_grouped_Anio5=produccion_df.groupby("Anio")["Departamento"].sum()
produccion_grouped_Anio5
```

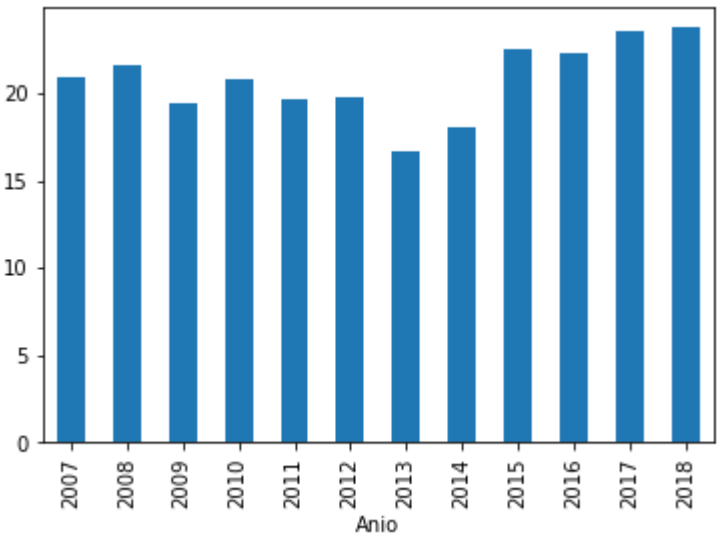
Out[94]:
```
Anio
2007    ANTIOQUIABOLIVARBOYACACALDASCAQUETACASANARECAU...
2008    ANTIOQUIABOLIVARBOYACACALDASCAQUETACASANARECAU...
2009    ANTIOQUIABOLIVARBOYACACALDASCAQUETACASANARECAU...
2010    ANTIOQUIAARAUCABOLIVARBOYACACALDASCAQUETACASAN...
2011    ANTIOQUIAARAUCABOLIVARBOYACACALDASCAQUETACASAN...
2012    ANTIOQUIABOLIVARBOYACACALDASCAQUETACASANARECAU...
2013    ANTIOQUIABOLIVARBOYACACALDASCAQUETACASANARECAU...
2014    ANTIOQUIABOLIVARBOYACACALDASCAQUETACASANARECAU...
2015    ANTIOQUIABOLIVARBOYACACALDASCAQUETACASANARECAU...
2016    ANTIOQUIABOLIVARBOYACACALDASCAQUETACASANARECAU...
2017    ANTIOQUIABOLIVARBOYACACALDASCAQUETACASANARECAU...
2018    ANTIOQUIABOLIVARBOYACACALDASCAQUETACASANARECAU...
Name: Departamento, dtype: object
```

In [95]:
```
Grupos_Departamentos_Rendimiento4=produccion_df.groupby("Anio")["Rendimiento (ha/
Grupos_Departamentos_Rendimiento4
# Indica el Rendimiento total del café en hectareas por toneladas (ha/ton) por aí
```

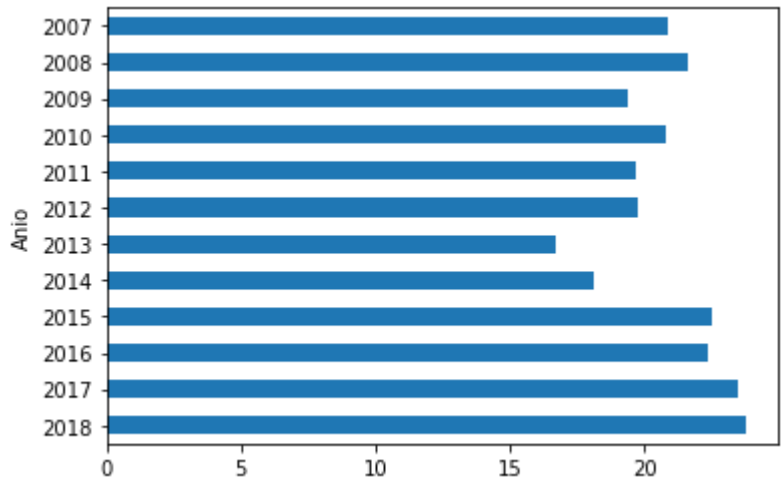
Out[95]:
```
Anio
2007    20.91
2008    21.62
2009    19.39
2010    20.84
2011    19.65
2012    19.75
2013    16.71
2014    18.09
2015    22.54
2016    22.34
2017    23.50
2018    23.75
Name: Rendimiento (ha/ton), dtype: float64
```

In [96]:
```
import numpy as np
import re
import sys
%matplotlib inline
Grupos_Departamentos_Rendimiento4.plot(kind='bar')
```

Out[96]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1915f23b6d0>
```



In [97]:
```
#pd.value_counts(produccion_df['Departamento']).plot(kind='barh').invert_yaxis()
Grupos_Departamentos_Rendimiento4.plot(kind='barh').invert_yaxis()
```

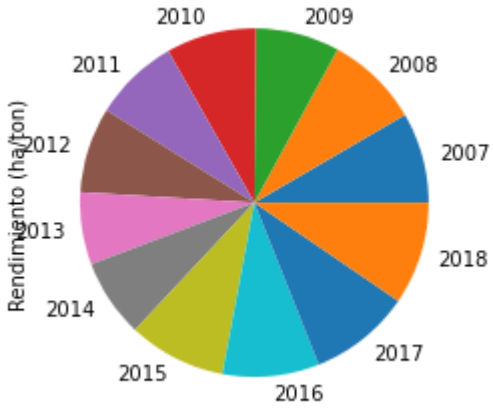


In [98]:
```
import numpy as np
import re
import sys
%matplotlib inline
Grupos_Departamentos_Rendimiento4.plot(kind='line')
```

Out[98]: <matplotlib.axes._subplots.AxesSubplot at 0x1915f49f0a0>

In [99]:
```
import numpy as np
import re
import sys
%matplotlib inline
Grupos_Departamentos_Rendimiento4.plot(kind='pie')
```

Out[99]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1915f272e20>
```



In [100]:
```
produccion_df.isnull()
#verificar si existen valores nulos
```

Out[100]:

	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 [101]:
```
produccion_df.isnull().sum()
## Totales de los valores nulos en el dataframe
```

Out[101]:
```
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
```

```
describe(produccion_df[,6:18])
```

In [102]:
```
produccion_df.tail()
```

Out[102]:

	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 [103]:
```
produccion_df.describe()
```

Out[103]:

	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 [104]:
```
produccion_df
```

Out[104]:

	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 [ ]:
```
```

In [ ]:
```
#pip install pandas-profiling
```

In [106]:
```
produccion_df
```

Out[106]:

	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 [16]:
```
produccion_df
```

Out[16]:

	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 [29]:
```
# 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")
```

Summarize dataset: 100% 22/22 [00:37<00:00, 1.69s/it, Completed]

Generate report structure: 100% 1/1 [00:03<00:00, 3.37s/it]

Render HTML: 100% 1/1 [00:00<00:00, 1.39it/s]

Export report to file: 100% 1/1 [00:00<00:00, 22.79it/s]

In [30]:
```
# La siguiente instrucción despliega el informe aquí en JUPYTER NOTEBOOK
profile
```

Overview

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

NUM	4
CAT	4

Warnings

`Producto` has constant value "266"	**Constant**
`Area (ha)` has a high cardinality: 261 distinct values	**High cardinality**
`Produccion (ton)` has a high cardinality: 262 distinct values	**High cardinality**
`Area Nacional (ha)` is highly correlated with `Produccion`	**High correlation**

Out[30]:

In [ ]: