**AIML ASSIGNMENT**

**SOLUTION**

In [3]: **import** pandas **as** pd

In [11]: df **=** pd**.**read\_csv('amazon.csv',encoding**=**'latin1')

In [13]: df

Out[13]: **year state month number date**

**0** 1998 Acre Janeiro 0.0 1998-01-01

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **1** | 1999 | Acre | Janeiro | 0.0 | 1999-01-01 |
| **2** | 2000 | Acre | Janeiro | 0.0 | 2000-01-01 |
| **3** | 2001 | Acre | Janeiro | 0.0 | 2001-01-01 |
| **4** | 2002 | Acre | Janeiro | 0.0 | 2002-01-01 |
| **...** | ... | ... | ... | ... | ... |
| **6449** | 2012 | Tocantins | Dezembro | 128.0 | 2012-01-01 |
| **6450** | 2013 | Tocantins | Dezembro | 85.0 | 2013-01-01 |
| **6451** | 2014 | Tocantins | Dezembro | 223.0 | 2014-01-01 |
| **6452** | 2015 | Tocantins | Dezembro | 373.0 | 2015-01-01 |
| **6453** | 2016 | Tocantins | Dezembro | 119.0 | 2016-01-01 |

6454 rows × 5 columns

**1.Display Top 5 Rows of The Dataset.**

In [16]: df**.**head(5)

Out[16]: **year state month number date**

**0** 1998 Acre Janeiro 0.0 1998-01-01

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **1** | 1999 | Acre | Janeiro | 0.0 | 1999-01-01 |
| **2** | 2000 | Acre | Janeiro | 0.0 | 2000-01-01 |
| **3** | 2001 | Acre | Janeiro | 0.0 | 2001-01-01 |
| **4** | 2002 | Acre | Janeiro | 0.0 | 2002-01-01 |

**2.Check Last 5 Rows.**

|  |
| --- |
| df**.**tail(5) |

In [19]:

Out[19]: **year state month number date**

**6449** 2012 Tocantins Dezembro 128.0 2012-01-01

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **6450** | 2013 | Tocantins | Dezembro | 85.0 | 2013-01-01 |
| **6451** | 2014 | Tocantins | Dezembro | 223.0 | 2014-01-01 |
| **6452** | 2015 | Tocantins | Dezembro | 373.0 | 2015-01-01 |
| **6453** | 2016 | Tocantins | Dezembro | 119.0 | 2016-01-01 |

**3.Find Shape of Our Dataset (Number of Rows and Number of Columns).**

|  |
| --- |
| x **=** df**.**shape |

In [22]:

|  |
| --- |
| print("The number of rows are:",x[0]) print("The number of columns are:",x[1]) |

In [24]:

The number of rows are: 6454

The number of columns are: 5

**4.Getting Information About Our Dataset Like Total Number Rows, Total**

**Number of Columns, Datatypes of Each Column and Memory Requirement.**

|  |
| --- |
| df**.**info() |

In [27]:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 6454 entries, 0 to 6453 Data columns (total 5 columns):

# Column Non-Null Count Dtype --- ------ -------------- ----- 0 year 6454 non-null int64

1. state 6454 non-null object
2. month 6454 non-null object
3. number 6454 non-null float64 4 date 6454 non-null object dtypes: float64(1), int64(1), object(3) memory usage: 252.2+ KB

**5.Check For Duplicate Data and Drop Them.**

|  |
| --- |
| df**.**duplicated() |

In [30]:

|  |  |
| --- | --- |
| Out[30]: | 1. False 2. False 3. False 4. False 5. False ... 6449 False 6. False 7. False 8. False 9. False   Length: 6454, dtype: bool |

In [32]: df**.**drop\_duplicates()

Out[32]: **year state month number date**

**0** 1998 Acre Janeiro 0.0 1998-01-01

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **1** | 1999 | Acre | Janeiro | 0.0 | 1999-01-01 |
| **2** | 2000 | Acre | Janeiro | 0.0 | 2000-01-01 |
| **3** | 2001 | Acre | Janeiro | 0.0 | 2001-01-01 |
| **4** | 2002 | Acre | Janeiro | 0.0 | 2002-01-01 |
| **...** | ... | ... | ... | ... | ... |
| **6449** | 2012 | Tocantins | Dezembro | 128.0 | 2012-01-01 |
| **6450** | 2013 | Tocantins | Dezembro | 85.0 | 2013-01-01 |
| **6451** | 2014 | Tocantins | Dezembro | 223.0 | 2014-01-01 |
| **6452** | 2015 | Tocantins | Dezembro | 373.0 | 2015-01-01 |
| **6453** | 2016 | Tocantins | Dezembro | 119.0 | 2016-01-01 |

6422 rows × 5 columns

**6.Check Null Values in The Dataset.**

In [35]: df**.**isnull()**.**sum()

Out[35]: year 0 state 0 month 0 number 0 date 0 dtype: int64

**7.Get Overall Statistics About the Dataframe.**

In [38]: df**.**describe()

Out[38]: **year number**

**count** 6454.000000 6454.000000

|  |  |  |
| --- | --- | --- |
| **mean** | 2007.461729 | 108.293163 |
| **std** | 5.746654 | 190.812242 |
| **min** | 1998.000000 | 0.000000 |
| **25%** | 2002.000000 | 3.000000 |
| **50%** | 2007.000000 | 24.000000 |
| **75%** | 2012.000000 | 113.000000 |
| **max** | 2017.000000 | 998.000000 |

**8.Rename Month Names to English.**

In [41]: df['month']**.**unique()

Out[41]: array(['Janeiro', 'Fevereiro', 'Março', 'Abril', 'Maio', 'Junho', 'Julho',

'Agosto', 'Setembro', 'Outubro', 'Novembro', 'Dezembro'], dtype=object)

In [43]: df['month'] **=** df['month']**.**replace({'Janeiro':'January', 'Fevereiro':'Feburary', 'Ma 'Julho':'July','Agosto':'August', 'Setembro':'Se

'Novembro':'November', 'Dezembro':'December'})

In [45]: df['month']**.**unique()

Out[45]: array(['January', 'Feburary', 'March', 'April', 'May', 'June', 'July', 'August', 'September', 'October', 'November', 'December'], dtype=object)

**9.Total Number of Fires Registered.**

In [48]: df['number']**.**sum()

Out[48]: 698924.073

**10.In Which Month Maximum Number of Forest Fires Were Reported?**

In [51]: max\_fires **=** df**.**groupby('month')['number']**.**sum()

In [53]: max\_fires

Out[53]: month

April 28188.770

August 88050.435

December 57535.480

Feburary 30848.050

January 47747.844

July 92326.113

June 56010.675

March 30717.405

May 34731.363

November 85508.054

October 88681.579

September 58578.305

|  |
| --- |
| max\_fire\_month **=** max\_fires**.**idxmax() max\_fire **=** max\_fires**.**max() print("The month with maximum fire is:",max\_fire\_month) print("The total number of fire in that month is:",max\_fire) |

Name: number, dtype: float64 In [55]:

The month with maximum fire is: July

The total number of fire in that month is: 92326.113

**11.In Which Year Maximum Number of Forest Fires Was Reported?**

In [58]: max\_fire\_year **=** df**.**groupby('year')['number']**.**sum()

In [60]: max\_fire\_year

|  |  |
| --- | --- |
| Out[60]: | year   1. 20013.971 2. 26882.821 3. 27351.251 4. 29071.612 5. 37390.600 6. 42760.674 7. 38453.163 8. 35004.965 9. 33832.161 10. 33037.413 11. 29378.964 12. 39117.178 13. 37037.449 14. 34633.545 15. 40084.860 16. 35146.118 17. 39621.183 18. 41208.292 19. 42212.229 20. 36685.624   Name: number, dtype: float64 |

In [62]: max\_fire\_year\_name **=** max\_fire\_year**.**idxmax() max\_fires **=** max\_fire\_year**.**max()

print("The year with the maximum number of fire is:",max\_fire\_year\_name) print("The maximum number of fire in that year is:",max\_fires)

The year with the maximum number of fire is: 2003

The maximum number of fire in that year is: 42760.674

**12.In Which State Maximum Number of Forest Fires Was Reported?**

In [65]: max\_fire\_state **=** df**.**groupby('state')['number']**.**max()

In [67]: max\_fire\_state

|  |  |
| --- | --- |
| Out[67]: | state  Acre 960.0  Alagoas 162.0  Amapa 969.0  Amazonas 998.0  Bahia 995.0  Ceara 995.0  Distrito Federal 196.0  Espirito Santo 307.0  Goias 943.0  Maranhao 972.0  Mato Grosso 979.0  Minas Gerais 959.0  Paraiba 987.0  Pará 982.0  Pernambuco 859.0  Piau 943.0  Rio 885.0  Rondonia 969.0  Roraima 820.0  Santa Catarina 765.0  Sao Paulo 981.0  Sergipe 198.0  Tocantins 989.0  Name: number, dtype: float64 |

In [69]: max\_fire\_state\_name **=** max\_fire\_state**.**idxmax() max\_fire **=** max\_fire\_state**.**max() print("The state with the maximum number of fire is:",max\_fire\_state\_name) print("The maximum number of fire in that state is:",max\_fire)

The state with the maximum number of fire is: Amazonas

The maximum number of fire in that state is: 998.0

**13.Find Total Number of Fires Were Reported in Amazonas.**

|  |
| --- |
| fire\_amazonas **=** df[df['state']**==**'Amazonas'] |

In [72]:

|  |
| --- |
| total**=**fire\_amazonas['number']**.**sum() |

In [74]:

|  |
| --- |
| total |

In [76]:

Out[76]: 30650.129

**14.Display Number of Fires Were Reported in Amazonas (Year-Wise).**

|  |
| --- |
| amazons\_fire\_year **=** df**.**groupby('year')['number']**.**sum() |

In [79]:

|  |
| --- |
| print("The number of fire in Amazons year-wise is:\n",amazons\_fire\_year) |

In [81]:

The number of fire in Amazons year-wise is:

year

1. 20013.971
2. 26882.821
3. 27351.251
4. 29071.612
5. 37390.600
6. 42760.674
7. 38453.163
8. 35004.965
9. 33832.161
10. 33037.413
11. 29378.964
12. 39117.178
13. 37037.449
14. 34633.545
15. 40084.860
16. 35146.118
17. 39621.183
18. 41208.292
19. 42212.229
20. 36685.624

Name: number, dtype: float64

**15.Display Number of Fires Were Reported in Amazonas (Day-Wise).**

|  |
| --- |
| amazonas\_fire\_date **=** df**.**groupby('date')['number']**.**sum() |

In [84]:

|  |
| --- |
| amazonas\_fire\_date |

In [86]:

Out[86]: date

1998-01-01 20013.971

1999-01-01 26882.821

2000-01-01 27351.251

2001-01-01 29071.612

2002-01-01 37390.600

2003-01-01 42760.674

2004-01-01 38453.163

2005-01-01 35004.965

2006-01-01 33832.161

2007-01-01 33037.413

2008-01-01 29378.964

2009-01-01 39117.178

2010-01-01 37037.449

2011-01-01 34633.545

2012-01-01 40084.860

2013-01-01 35146.118

2014-01-01 39621.183

2015-01-01 41208.292

2016-01-01 42212.229

2017-01-01 36685.624

Name: number, dtype: float64

**16.Find Total Number of Fires Were Reported In 2015 And Visualize Data Based on Each ‘Month’.**

In [89]: fire\_report **=** df[df['year']**==**2015]

In [91]: fire\_report

Out[91]: **year state month number date**

**17** 2015 Acre January 1.000 2015-01-01

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **37** | 2015 | Acre | Feburary | 2.000 | 2015-01-01 |
| **57** | 2015 | Acre | March | 2.000 | 2015-01-01 |
| **77** | 2015 | Acre | April | 3.000 | 2015-01-01 |
| **97** | 2015 | Acre | May | 2.000 | 2015-01-01 |
| **...** | ... | ... | ... | ... | ... |
| **6372** | 2015 | Tocantins | August | 2.540 | 2015-01-01 |
| **6392** | 2015 | Tocantins | September | 5.510 | 2015-01-01 |
| **6412** | 2015 | Tocantins | October | 4.844 | 2015-01-01 |
| **6432** | 2015 | Tocantins | November | 833.000 | 2015-01-01 |
| **6452** | 2015 | Tocantins | December | 373.000 | 2015-01-01 |

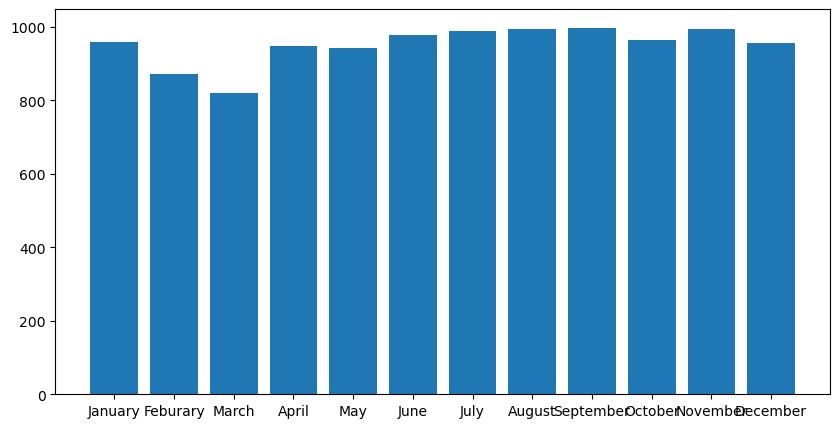
324 rows × 5 columns

In [93]: fire\_report**.**groupby('month')['number']**.**sum()

|  |  |
| --- | --- |
| Out[93]: | month  April 2573.000  August 4363.125  December 4088.522  Feburary 2309.000  January 4635.000  July 4364.392  June 3260.552  March 2202.000  May 2384.000  November 4034.518  October 4499.525  September 2494.658  Name: number, dtype: float64 |

In [95]: **import** matplotlib.pyplot **as** plt

In [96]: plt**.**figure(figsize**=**(10,5)) plt**.**bar(df['month'],df['number']) plt**.**show()



**17.Find Average Number of Fires Were Reported from Highest to Lowest (State-Wise).**

|  |
| --- |
| avg\_fire **=** df**.**groupby('state')['number']**.**mean()**.**sort\_values(ascending**=False**) |

In [98]:

|  |
| --- |
| avg\_fire |

In [100…

Out[100… state

Sao Paulo 213.896226

Mato Grosso 201.351523

Bahia 187.222703

Piau 158.174674

Goias 157.721841

Minas Gerais 156.800243

Tocantins 141.037176

Amazonas 128.243218

Ceara 127.314071

Paraiba 109.698573

Maranhao 105.142808

Pará 102.561272

Pernambuco 102.502092

Roraima 102.029598

Santa Catarina 101.924067

Amapa 91.345506

Rondonia 84.876272

Acre 77.255356

Rio 62.985865

Espirito Santo 27.389121

Alagoas 19.350000

Distrito Federal 14.899582

Sergipe 13.543933

Name: number, dtype: float64

**18.To Find the State Names Where Fires Were Reported In 'dec' Month.**

In [102… fire\_reported **=** df[df['month']**==**'December']['state']**.**unique()

In [103… fire\_reported

Out[103… array(['Acre', 'Alagoas', 'Amapa', 'Amazonas', 'Bahia', 'Ceara',

'Distrito Federal', 'Espirito Santo', 'Goias', 'Maranhao',

'Mato Grosso', 'Minas Gerais', 'Pará', 'Paraiba', 'Pernambuco',

'Piau', 'Rio', 'Rondonia', 'Roraima', 'Santa Catarina',

'Sao Paulo', 'Sergipe', 'Tocantins'], dtype=object)