

```
[20]: #import libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
plt.rcParams['figure.figsize']=(20,8)

In [2]: #importing the database
df=pd.read_csv('Suicides in India 2001-2012.csv')

In [3]: df

Out[3]:
   State  Year  Type_code  Type  Gender  Age_group  Total
0  A & N Islands  2001  Causes  Illness (Aids/STD)  Female  0-14  0
1  A & N Islands  2001  Causes  Bankruptcy or Sudden change in Economic  Female  0-14  0
2  A & N Islands  2001  Causes  Cancellation/Non-Settlement of Marriage  Female  0-14  0
3  A & N Islands  2001  Causes  Physical Abuse (Rape/Incest Etc.)  Female  0-14  0
4  A & N Islands  2001  Causes  Dowry Dispute  Female  0-14  0
...  ...  ...  ...  ...  ...  ...
237514  West Bengal  2012  Social_Status  Separated  Male  0-100+  149
237515  West Bengal  2012  Social_Status  Widowed/Widower  Male  0-100+  233
237516  West Bengal  2012  Social_Status  Married  Male  0-100+  5451
237517  West Bengal  2012  Social_Status  Divorcee  Male  0-100+  189
237518  West Bengal  2012  Social_Status  Never Married  Male  0-100+  2658

237519 rows x 7 columns

In [5]: df.shape

Out[5]:
(237519, 7)

In [6]: df.isnull().sum()

Out[6]:
State      0
Year       0
Type_code  0
Type       0
Gender     0
Age_group  0
Total      0
dtype: int64

In [7]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 237519 entries, 0 to 237518
Data columns (total 7 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   State       237519 non-null  object
1   Year        237519 non-null  int64
2   Type_code   237519 non-null  object
3   Type        237519 non-null  object
4   Gender      237519 non-null  object
5   Age_group   237519 non-null  object
6   Total       237519 non-null  int64
dtypes: int64(2), object(5)
memory usage: 12.7+ MB

In [8]: df.describe()

Out[8]:
   Year  Total
count 237519.000000  237519.000000
mean    2006.500448    55934477
std      3.452240    792.749038
min      2001.000000    0.000000
25%      2004.000000    0.000000
50%      2007.000000    0.000000
75%      2010.000000    6.000000
max      2012.000000   63343.000000

In [11]: df['Type'].unique()

Out[11]:
array(['Illness (Aids/STD)', 'Bankruptcy or Sudden change in Economic',
      'Cancellation/Non-Settlement of Marriage',
      'Physical Abuse (Rape/Incest Etc.)', 'Dowry Dispute',
      'Family Problems', 'Ideological Causes/Hero Worshipping',
      'Other Prolonged Illness', 'Property Dispute',
      'Fall in Social Reputation', 'Illegitimate Pregnancy',
      'Failure in Examination', 'Insanity/Mental Illness',
      'Love Affairs', 'Professional/Career Problem', 'Divorcee',
      'Drug Abuse/Addiction', 'Not having Children (Barrenness/Impotency',
      'Causes Not known', 'Unemployment',
      'Other Causes (Please Specify)', 'Poverty', 'Death of Dear Person',
      'Cancer', 'Suspected/Illicit Relation', 'Paralysis', 'Diploma',
      'No Education', 'Post Graduate and Above', 'Middle', 'Graduate',
      'Hr. Secondary/Intermediate/Pre-University', 'Primary',
      'Matriculate/Secondary', 'By Consuming Insecticides', 'By Hanging',
      'By Jumping from (other sites)', 'By touching electric wires',
      'By Machine', 'By Fire/Self Immolation',
      'By Jumping off Moving Vehicles/Trains',
      'By Other means (Please specify)', 'By Self Infliction of injury',
      'By Over Alcoholism', 'By Consuming Other Poison',
      'By coming under running vehicles/trains',
      'By Overdose of sleeping pills', 'By Jumping from (Building)',
      'By Drowning', 'By Fire-Arms', 'Retired Person', 'Unemployed',
      'Public Sector Undertaking', 'Service (Private)', 'Married',
      'Self-employed (Business activity)', 'Professional Activity',
      'Student', 'Others (Please Specify)',
      'Farming/Agriculture Activity', 'Service (Government)', 'Married',
      'Separated', 'Widowed/Widower', 'Divorcee', 'Never Married',
      'Bankruptcy or Sudden change in Economic Status',
      'Not Having Children (Barrenness/Impotency', 'By Other means',
      dtype=object)

In [13]: df.groupby('Year')['Total'].sum().to_frame()

Out[13]:
   Year  Total
2001   976464
2002   993648
2003   997622
2004  1023137
2005  1025201
2006  1062991
2007  1103667
2008  1125082
2009  1144033
2010  1211322
2011  1219499
2012  1189068

In [14]: suicidesperyear=df.groupby('Year')['Total'].sum()
suicidesperyear

Out[14]:
Year
2001    976464
2002    993648
2003    997622
2004   1023137
2005   1025201
2006   1062991
2007   1103667
2008   1125082
2009   1144033
2010   1211322
2011   1219499
2012   1189068
Name: Total, dtype: int64

In [21]: sns.barplot(x=suicidesperyear.index, y=suicidesperyear, data=df.set(title= "TOTAL SUICIDES PER YEAR"))

Out[21]:
[Text(0.5, 1.0, 'TOTAL SUICIDES PER YEAR')]

TOTAL SUIDES PER YEAR

1200000
1000000
800000
600000
400000
200000
0
2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012

In [27]: #suicided due to unemployment
suicidesduetomp=df[df['Type']=='Unemployment'].groupby('Year')['Total'].sum().sort_values(ascending=False)
sns.barplot(x=suicidesduetomp.index, y=suicidesduetomp)

Out[27]:
<AxesSubplot: xlabel='Year', ylabel='Total'>

250000
200000
150000
100000
50000
0
2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012

In [29]: #suicide causes
suicauses=df.groupby('Type')['Total'].sum().sort_values(ascending=False)
suicauses

Out[29]:
Type
Married      3065322
Primary      1088481
Middle       1028913
No Education  965271
Never Married 954903
...
Bankruptcy or Sudden change in Economic Status 2455
Illegitimate Pregnancy 2494
Ideological Causes/Hero Worshipping 2118
By Machine 1661
Professional Childen (Barrenness/Impotency 1766
Name: Total, Length: 69, dtype: int64

In [31]: sns.barplot(x=suicauses.index[:5], y=suicauses[:5])

Out[31]:
<AxesSubplot: xlabel='Type', ylabel='Total'>

3500000
3000000
2500000
2000000
1500000
1000000
500000
0
Married Primary Middle No Education Never Married

In [33]: #gender plays any role
df['Gender'].value_counts()

Out[33]:
Male    118879
Female  118640
Name: Gender, dtype: int64

In [36]: plt.pie(df['Gender'].value_counts(),labels=['Males','Females'],autopct='%1.0f%%')

Out[36]:
((<matplotlib.patches.Wedge at 0x27899145550>,
  <matplotlib.patches.Wedge at 0x27899145c70a>),
 (Text(0.001786668802412718, -1.099986259252269, 'Males'),
  Text(0.001786668802412718, -1.099986259252269, 'Females')),
 (Text(-0.0009483637528589488, 0.599999250504692, '50%'),
  Text(0.0009483637528589754, -0.599999250504692, '50%'))

Males
50%
Females
50%

In [38]: #age group mostlikely to suicides
df[df['Type']=='Married']['Gender'].value_counts()

Out[38]:
Gender
Male    456
Female  456
Name: Gender, dtype: int64

In [39]: df.head()

Out[39]:
   State  Year  Type_code  Type  Gender  Age_group  Total
0  A & N Islands  2001  Causes  Illness (Aids/STD)  Female  0-14  0
1  A & N Islands  2001  Causes  Bankruptcy or Sudden change in Economic  Female  0-14  0
2  A & N Islands  2001  Causes  Cancellation/Non-Settlement of Marriage  Female  0-14  0
3  A & N Islands  2001  Causes  Physical Abuse (Rape/Incest Etc.)  Female  0-14  0
4  A & N Islands  2001  Causes  Dowry Dispute  Female  0-14  0

In [42]: suibysagegrp=df['Age_group'].value_counts()
suibysagegrp

Out[42]:
15-29    45223
30-44    45193
45-59    45146
60-74    45074
0-14     45007
0-100+   11856
Name: Age_group, dtype: int64

In [43]: sns.barplot(x=suibysagegrp.index, y=suibysagegrp)

Out[43]:
<AxesSubplot: xlabel='Age_group'>

40000
30000
20000
10000
0
15-29 30-44 45-59 60+ 0-14 0-100+

In [47]: #states having highest count
suinistatesasc=df.groupby('State')['Total'].sum().sort_values(ascending=False)
suinistatesasc

Out[47]:
State
Total (All India) 2911862
Total (States) 288026
Maharashtra 901945
West Bengal 849936
Tamil Nadu 818691
Andhra Pradesh 814059
Karnataka 734825
Kerala 538946
Madhya Pradesh 451535
Gujarat 330858
Chhattisgarh 302354
Odisha 267234
Rajasthan 255134
Uttar Pradesh 233352
Assam 172276
Haryana 147176
Delhi (Ut) 84272
Total (Uts) 58366
Jharkhand 49720
Punjab 46350
Bihar 46214
Tripura 45965
Puducherry 32144
Himachal Pradesh 26562
Uttarakhand 18496
Goa 17363
Jammu & Kashmir 14821
Sikkim 9606
A & N Islands 8109
Arunachal Pradesh 6633
Meghalaya 5415
Chandigarh 5164
Mizoram 4154
D & N Haveli 3430
Manipur 2102
Nagaland 1728
Daman & Diu 1391
Lakshadweep 50
Name: Total, dtype: int64

In [45]: df.head()

Out[45]:
   State  Year  Type_code  Type  Gender  Age_group  Total
0  A & N Islands  2001  Causes  Illness (Aids/STD)  Female  0-14  0
1  A & N Islands  2001  Causes  Bankruptcy or Sudden change in Economic  Female  0-14  0
2  A & N Islands  2001  Causes  Cancellation/Non-Settlement of Marriage  Female  0-14  0
3  A & N Islands  2001  Causes  Physical Abuse (Rape/Incest Etc.)  Female  0-14  0
4  A & N Islands  2001  Causes  Dowry Dispute  Female  0-14  0

In [51]: sns.barplot(x=suinistatesasc.index[:9], y=suinistatesasc[:9])

Out[51]:
<AxesSubplot: xlabel='State', ylabel='Total'>

800000
600000
400000
200000
0
Maharashtra West Bengal Tamil Nadu Andhra Pradesh Karnataka Kerala Madhya Pradesh

In [52]: suinistatesdesc=df.groupby('State')['Total'].sum().sort_values(ascending=True)
suinistatesdesc

Out[52]:
State
Lakshadweep 50
Daman & Diu 1391
Nagaland 1728
Manipur 2102
D & N Haveli 3430
Mizoram 4154
Chandigarh 5164
Meghalaya 5415
Arunachal Pradesh 6633
A & N Islands 8109
Sikkim 9606
Jammu & Kashmir 14821
Goa 17363
Uttarakhand 18496
Himachal Pradesh 26562
Tripura 45965
Puducherry 32144
Bihar 46214
Punjab 46350
Jharkhand 49720
Total (Uts) 58366
Delhi (Ut) 84272
Haryana 147176
Assam 172276
Uttar Pradesh 233352
Rajasthan 255134
Odisha 267234
Chhattisgarh 302354
Madhya Pradesh 451535
Kerala 538946
Karnataka 734825
Andhra Pradesh 814059
Tamil Nadu 818691
West Bengal 849936
Maharashtra 901945
Total (States) 288026
Total (All India) 2911862
Name: Total, dtype: int64

In [54]: sns.barplot(x=suinistatesdesc.index[:9], y=suinistatesdesc[:9])

Out[54]:
<AxesSubplot: xlabel='State', ylabel='Total'>

6000
4000
2000
0
Lakshadweep Daman & Diu Nagaland Manipur D & N Haveli Mizoram Chandigarh Meghalaya Arunachal Pradesh

In [55]: df.head()

Out[55]:
   State  Year  Type_code  Type  Gender  Age_group  Total
0  A & N Islands  2001  Causes  Illness (Aids/STD)  Female  0-14  0
1  A & N Islands  2001  Causes  Bankruptcy or Sudden change in Economic  Female  0-14  0
2  A & N Islands  2001  Causes  Cancellation/Non-Settlement of Marriage  Female  0-14  0
3  A & N Islands  2001  Causes  Physical Abuse (Rape/Incest Etc.)  Female  0-14  0
4  A & N Islands  2001  Causes  Dowry Dispute  Female  0-14  0

In [66]: df['Age_group'].unique()
youth=df[df['Age_group']=='0-14']
ygy=youth.groupby('Type')['Total'].sum().sort_values(ascending=False)
ygy

Out[66]:
Type
By Hanging 176914
By Consuming Insecticides 155539
House Wife 127984
Family Problems 125681
By Consuming Insecticides 104929
Causes Not known 90839
By Consuming Other Poison 85190
Other Causes (Please Specify) 81003
Student 70838
By Fire/Self Immolation 57408
Farming/Agriculture Activity 53741
Other Prolonged Illness 52458
Unemployed 47123
By Other means (please specify) 44724
Service (Private) 39512
By Drowning 37282
Love Affairs 34773
Insanity/Mental Illness 28482
Failure in Examination 25932
Dowry Dispute 23211
Self-employed (Business activity) 20544
By coming under running vehicles/trains 14395
Unemployment 12505
Professional Activity 11230
Public Sector Undertaking 9528
Poverty 8450
Suspected/Illicit Relation 7752
Cancellation/Non-Settlement of Marriage 7492
Drug Abuse/Addiction 5926
Service (Government) 5903
Bankruptcy or Sudden change in Economic 4888
Fall in Social Reputation 4399
Property Dispute 4127
By touching electric wires 3687
Professional/Career Problem 3510
By Other means 3263
By Over Alcoholism 3050
By Overdose of sleeping pills 2957
By Jumping from (other sites) 2828
Not having Children (Barrenness/Impotency 2791
By Jumping off Moving Vehicles/Trains 2702
By Jumping from (Building) 2633
By Fire-Arms 2519
Illness (Aids/STD) 2185
Physical Abuse (Rape/Incest Etc.) 1847
Illegitimate Pregnancy 1769
Divorcee 1557
Paralysis 1121
Cancer 956
Ideological Causes/Hero Worshipping 859
By Machine 626
Bankruptcy or Sudden change in Economic Status 486
Not having Children (Barrenness/Impotency 300
Retired Person 294
Name: Total, dtype: int64

In [67]: sns.barplot(x=ygy.index[:9], y=ygy[:9])
plt.xticks(rotation=40)

Out[67]:
(array([0, 1, 2, 3, 4, 5, 6, 7, 8]),
 [Text(1, 0, 'By Hanging'),
  Text(2, 0, 'Other Causes (Please Specify)'),
  Text(1, 0, 'House Wife'),
  Text(3, 0, 'Family Problems'),
  Text(4, 0, 'By Consuming Insecticides'),
  Text(5, 0, 'Causes Not known'),
  Text(6, 0, 'By Consuming Other Poison'),
  Text(7, 0, 'Other Causes (Please Specify)'),
  Text(8, 0, 'Student')])

175000
150000
125000
100000
75000
50000
25000
0
By Hanging Other Causes (Please Specify) House Wife Family Problems By Consuming Insecticides Causes Not known By Consuming Other Poison Other Causes (Please Specify) Student

In [68]: #least suicide cases
df['Age_group'].unique()
youth=df[df['Age_group']=='0-14']
ygy=youth.groupby('Type')['Total'].sum().sort_values(ascending=True)
ygy

Out[68]:
Type
Retired Person 294
Not Having Children (Barrenness/Impotency 300
Bankruptcy or Sudden change in Economic Status 486
By Machine 626
Ideological Causes/Hero Worshipping 859
Cancer 956
Paralysis 1121
Divorcee 1557
Illegitimate Pregnancy 1769
By Self Infliction of injury 1847
Physical Abuse (Rape/Incest Etc.) 2185
Illness (Aids/STD) 2519
By Fire-Arms 2533
By Jumping from (Building) 2702
By Jumping off Moving Vehicles/Trains 2791
Not having Children (Barrenness/Impotency 2828
By Jumping from (Other sites) 2957
By Overdose of sleeping pills 2989
By Over Alcoholism 3050
By Other means 3263
Professional/Career Problem 3510
Death of Dear Person 3687
Love Affairs 37282
Insanity/Mental Illness 34773
Love Affairs 34773
By Drowning 37282
By Other means (please specify) 44724
Unemployed 47123
Other Prolonged Illness 52458
Farming/Agriculture Activity 53741
By Fire/Self Immolation 57408
Student 70838
Other Causes (Please Specify) 81003
By Consuming Other Poison 85190
Causes Not known 90839
By Consuming Insecticides 104929
Family Problems 125681
House Wife 127984
Others (Please Specify) 155539
By Hanging 176914
Name: Total, dtype: int64

In [69]: sns.barplot(x=ygy.index[:9], y=ygy[:9])
plt.xticks(rotation=40)

Out[69]:
(array([0, 1, 2, 3, 4, 5, 6, 7, 8]),
 [Text(1, 0, 'Retired Person'),
  Text(2, 0, 'Not having Children (Barrenness/Impotency)'),
  Text(3, 0, 'Bankruptcy or Sudden change in Economic Status'),
  Text(4, 0, 'Ideological Causes/Hero Worshipping'),
  Text(5, 0, 'Cancer'),
  Text(6, 0, 'Paralysis'),
  Text(7, 0, 'Divorcee'),
  Text(8, 0, 'Illegitimate Pregnancy')])

1750
1500
1250
1000
750
500
250
0
Retired Person Not having Children (Barrenness/Impotency Bankruptcy or Sudden change in Economic Status By Machine Ideological Causes/Hero Worshipping Causes Not known By Consuming Other Poison Other Causes (Please Specify) Student

In [ ]:
```

