Assignment 2

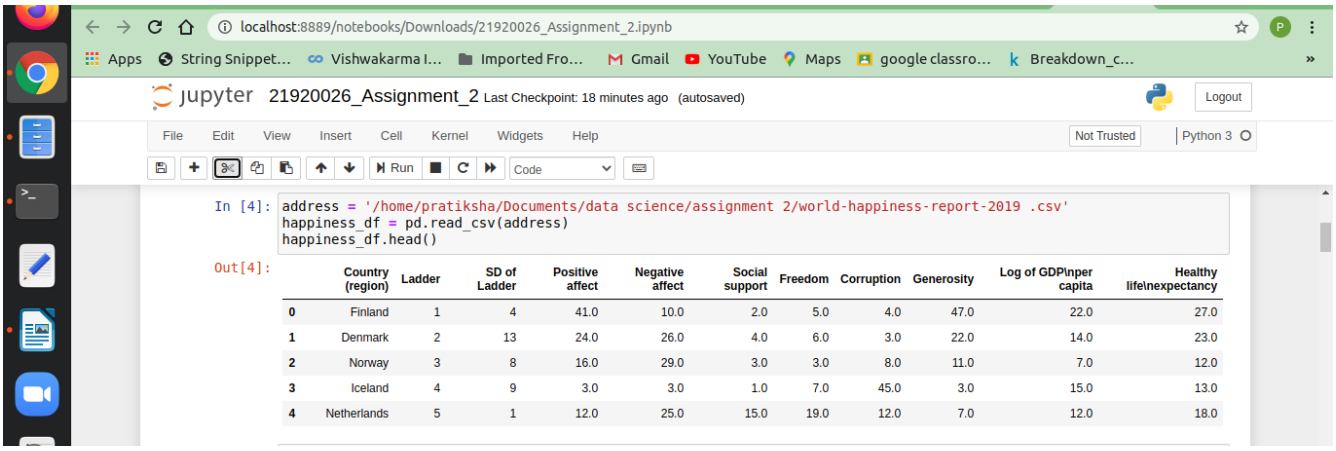
Name:Pratiksha Gaikwad
Gr.No: 21920026
RollNo: 321066
TY-Comp Batch: A3

Aim : Perform the following operations using Python on the data sets Compute and display summary statistics for each feature available in the dataset. (eg. minimum value, maximum value, mean, range, standard deviation, variance and percentiles) ·Data Visualization-Create a histogram for each feature in the dataset to illustrate the feature distributions., Data cleaning, Data integration, Data transformation

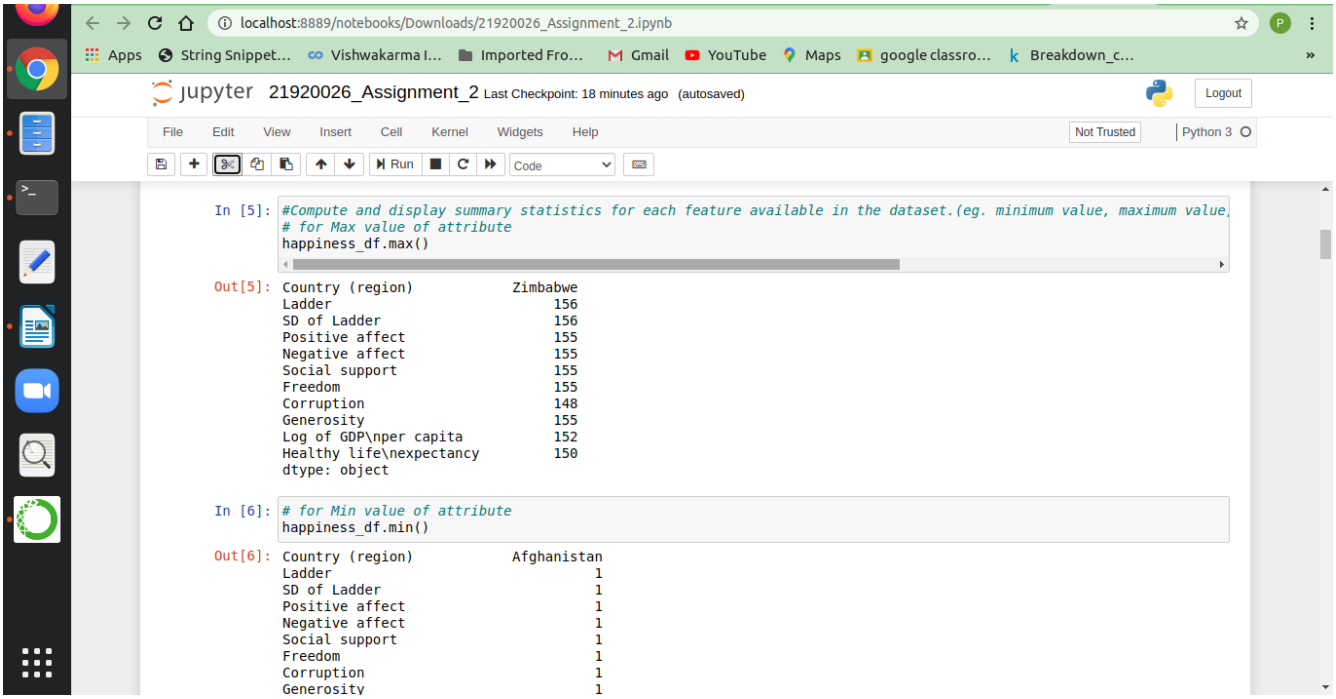
Outcome:

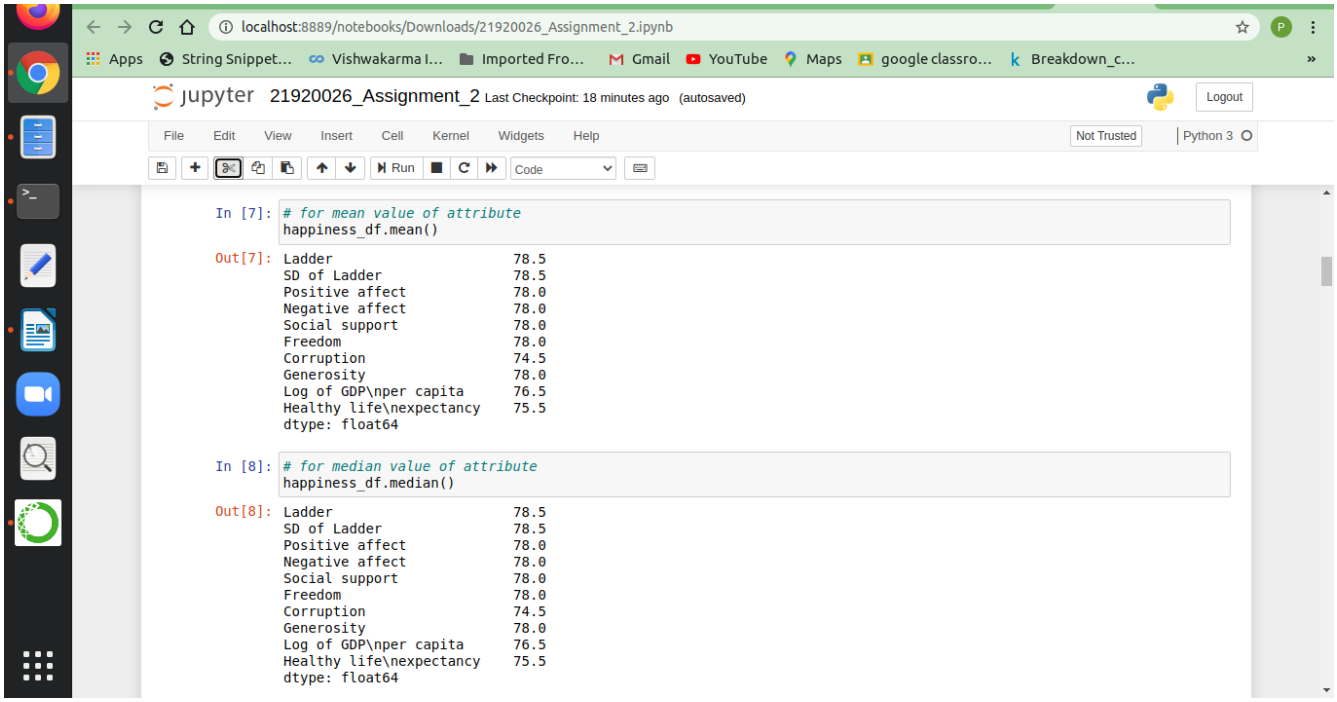
Dataset Name: World Happiness Report 2019
Overview of Data :
Each column of data has the next description.

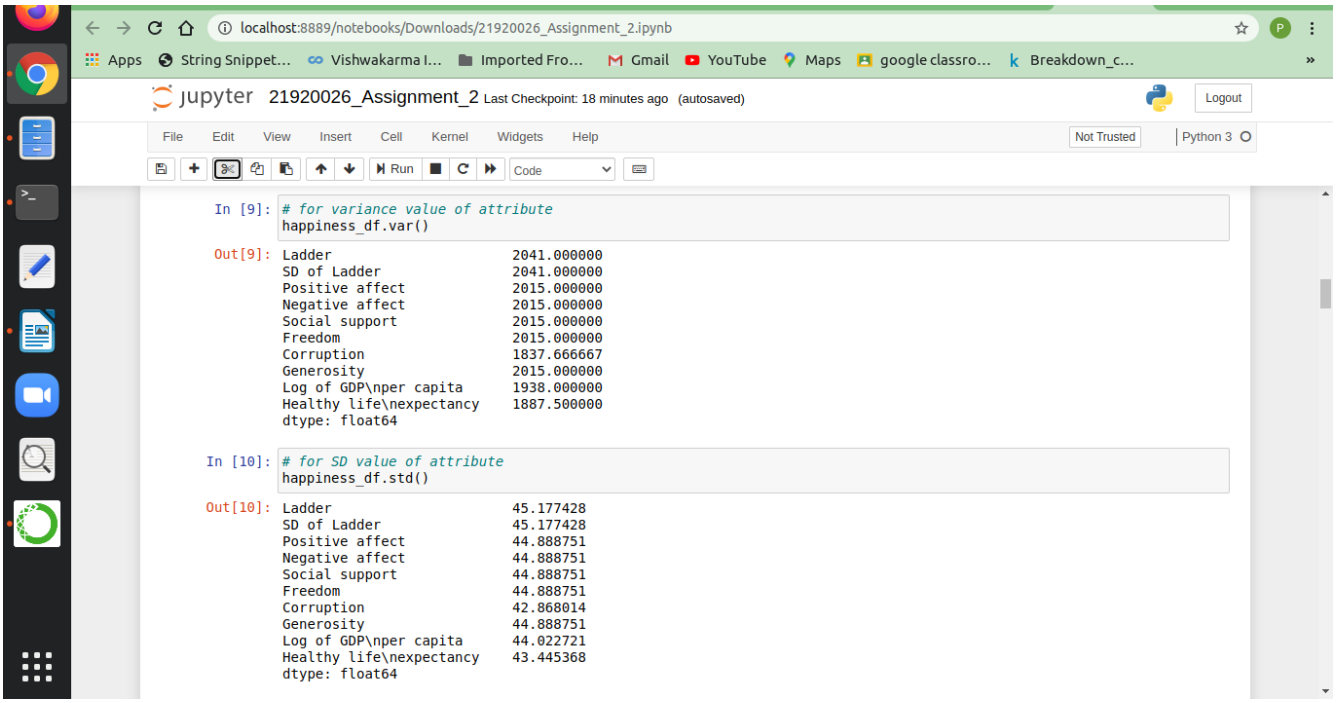
1. **Country (region)** : Name of the country.
2. **Ladder:** is a measure of life satisfaction.
3. **SD of Ladder:** Standard deviation of the ladder.
4. **Positive affect:** Measure of positive emotion.
5. **Negative affect:** Measure of negative emotion.
6. **Social support** :The extent to which Social support contributed to the calculation of the Happiness Score.
7. **Freedom:** The extent to which Freedom contributed to the calculation of the Happiness Score.
8. **Corruption:** The extent to which Perception of Corruption contributes to Happiness Score.
9. **Generosity:** The extent to which Generosity contributed to the calculation of the Happiness Score.
10. **Log of GDP per capita:** The extent to which GDP contributes to the calculation of the Happiness Score.
11. **Healthy life expectancy:** The extent to which Life expectancy contributed to the calculation of the Happiness Score.

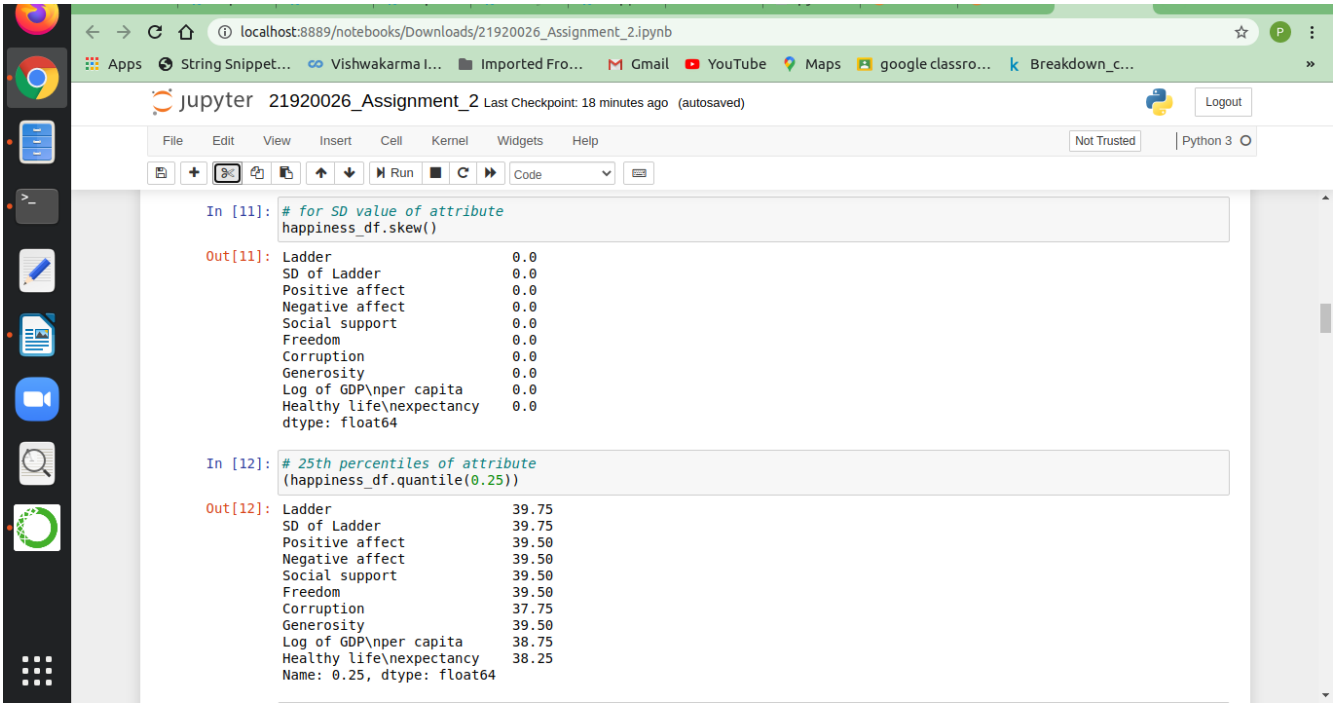


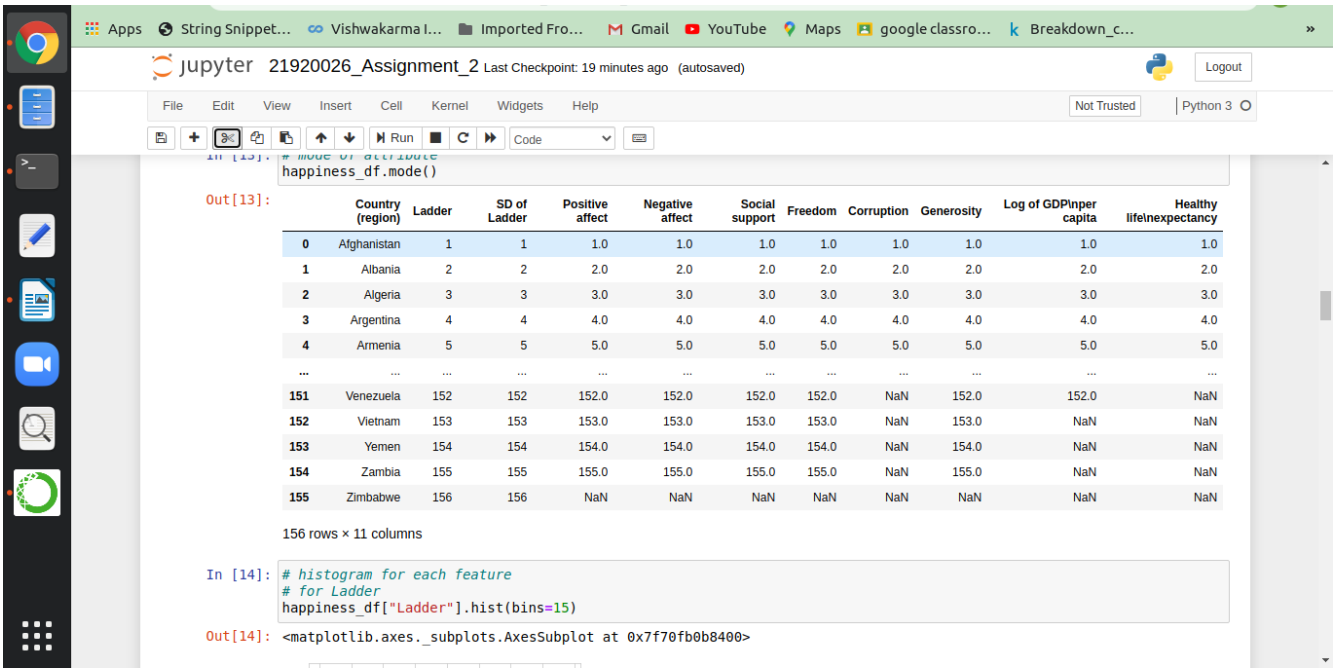
	Country (region)	Ladder	SD of Ladder	Positive affect	Negative affect	Social support	Freedom	Corruption	Generosity	Log of GDP\nper capita	Healthy life\nexpectancy
0	Finland	1	4	41.0	10.0	2.0	5.0	4.0	47.0	22.0	27.0
1	Denmark	2	13	24.0	26.0	4.0	6.0	3.0	22.0	14.0	23.0
2	Norway	3	8	16.0	29.0	3.0	3.0	8.0	11.0	7.0	12.0
3	Iceland	4	9	3.0	3.0	1.0	7.0	45.0	3.0	15.0	13.0
4	Netherlands	5	1	12.0	25.0	15.0	19.0	12.0	7.0	12.0	18.0

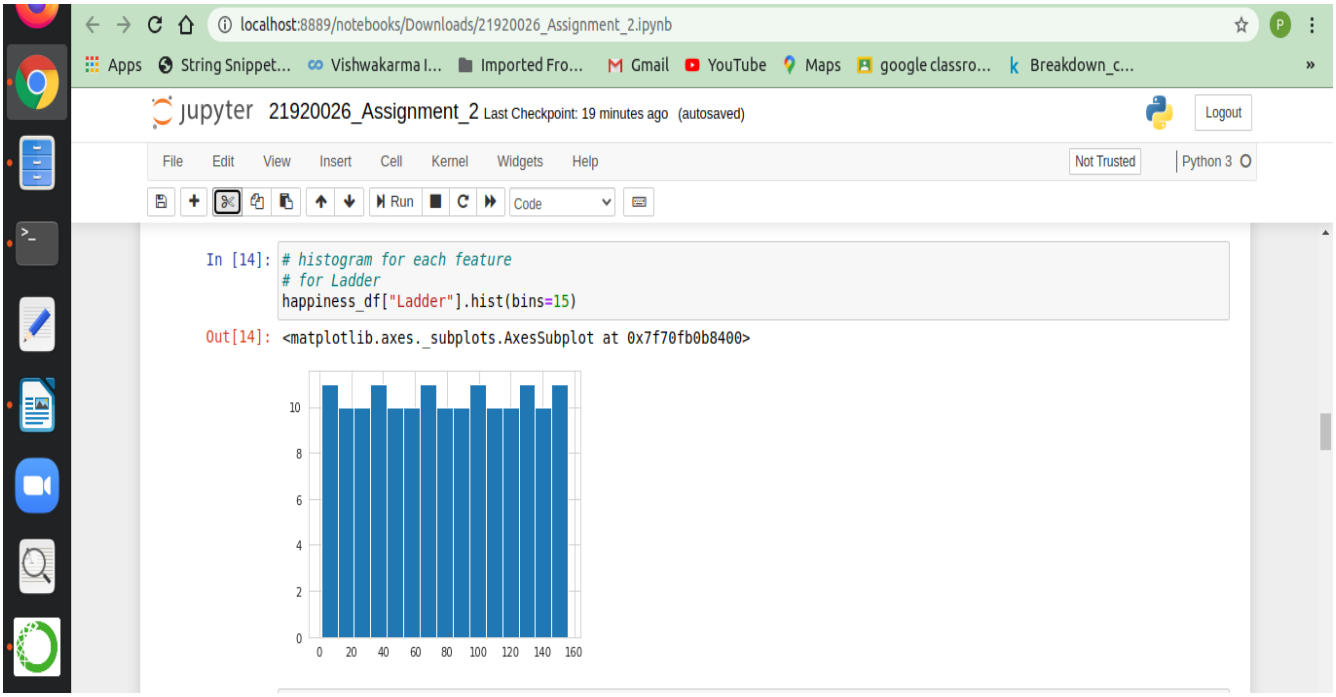


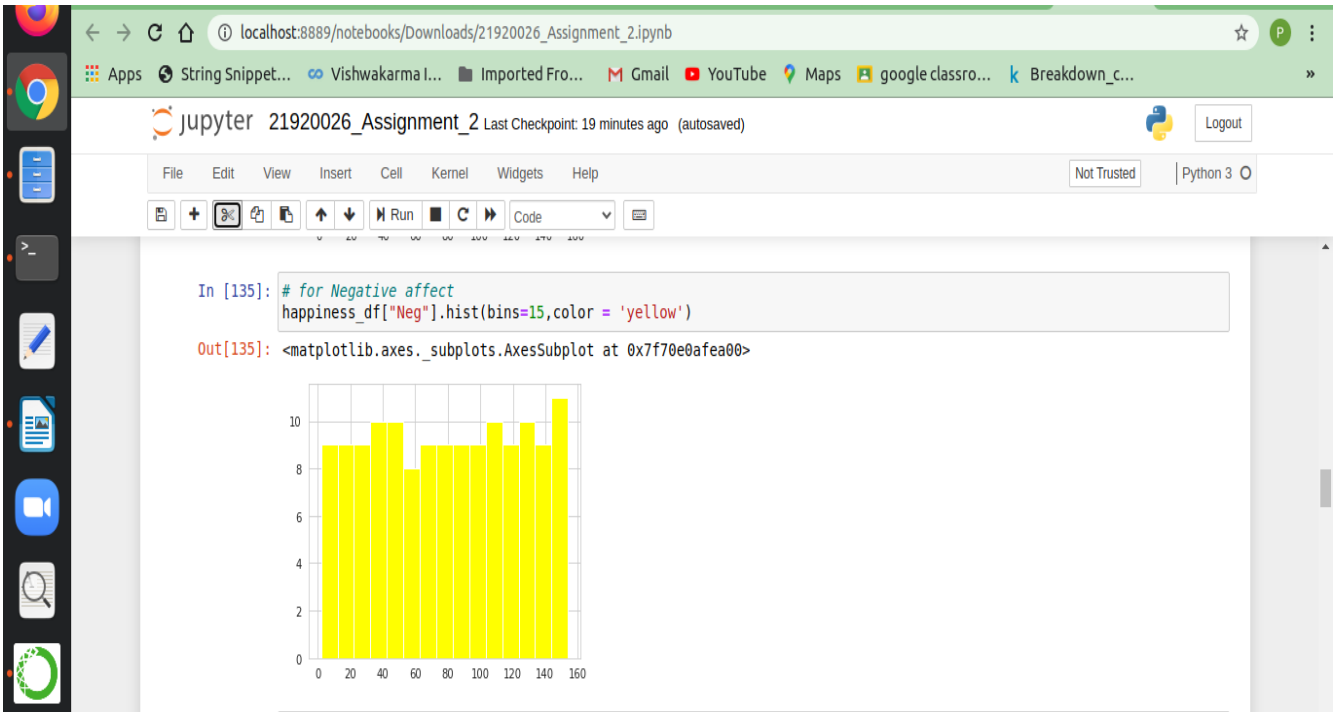

localhost:8889/notebooks/Downloads/21920026_Assignment_2.ipynb
jupyter 21920026_Assignment_2 Last Checkpoint: 18 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3
Run
Code
In [7]:
# for mean value of attribute
happiness_df.mean()
Out[7]:
Ladder 78.5
SD of Ladder 78.5
Positive affect 78.0
Negative affect 78.0
Social support 78.0
Freedom 78.0
Corruption 74.5
Generosity 78.0
Log of GDP\nper capita 76.5
Healthy life\nexpectancy 75.5
dtype: float64
In [8]:
# for median value of attribute
happiness_df.median()
Out[8]:
Ladder 78.5
SD of Ladder 78.5
Positive affect 78.0
Negative affect 78.0
Social support 78.0
Freedom 78.0
Corruption 74.5
Generosity 78.0
Log of GDP\nper capita 76.5
Healthy life\nexpectancy 75.5
dtype: float64

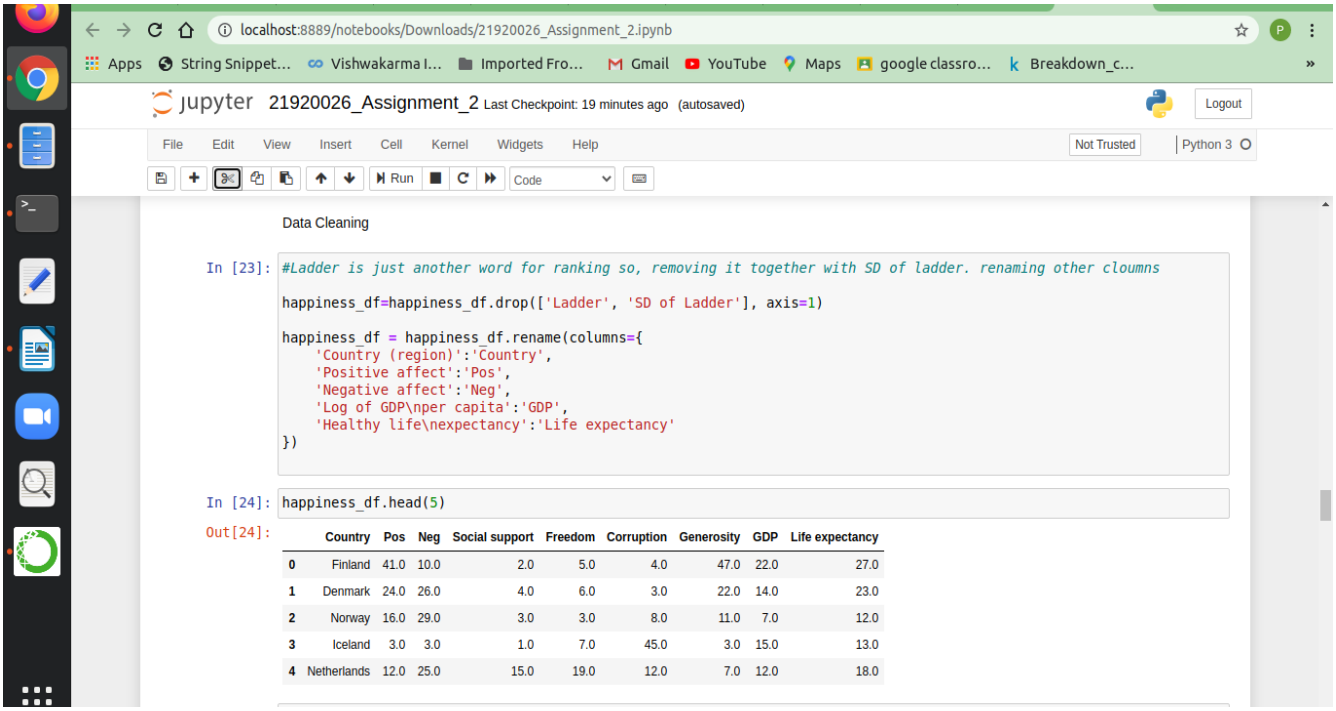

localhost:8889/notebooks/Downloads/21920026_Assignment_2.ipynb
jupyter 21920026_Assignment_2 Last Checkpoint: 18 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3
Run
Code
In [9]:
# for variance value of attribute
happiness_df.var()
Out[9]:
Ladder 2041.000000
SD of Ladder 2041.000000
Positive affect 2015.000000
Negative affect 2015.000000
Social support 2015.000000
Freedom 2015.000000
Corruption 1837.666667
Generosity 2015.000000
Log of GDP\nper capita 1938.000000
Healthy life\nexpectancy 1887.500000
dtype: float64
In [10]:
# for SD value of attribute
happiness_df.std()
Out[10]:
Ladder 45.177428
SD of Ladder 45.177428
Positive affect 44.888751
Negative affect 44.888751
Social support 44.888751
Freedom 44.888751
Corruption 42.868014
Generosity 44.888751
Log of GDP\nper capita 44.022721
Healthy life\nexpectancy 43.445368
dtype: float64


localhost:8889/notebooks/Downloads/21920026_Assignment_2.ipynb
Apps String Snippet... Vishwakarma I... Imported Fro... Gmail YouTube Maps google classro... Breakdown_c...
jupyter 21920026_Assignment_2 Last Checkpoint: 18 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3
Run
Code
In [11]:
# for SD value of attribute
happiness_df.skew()
Out[11]:
Ladder 0.0
SD of Ladder 0.0
Positive affect 0.0
Negative affect 0.0
Social support 0.0
Freedom 0.0
Corruption 0.0
Generosity 0.0
Log of GDP\nper capita 0.0
Healthy life\nexpectancy 0.0
dtype: float64
In [12]:
25th percentiles of attribute
(happiness_df.quantile(0.25))
Out[12]:
Ladder 39.75
SD of Ladder 39.75
Positive affect 39.50
Negative affect 39.50
Social support 39.50
Freedom 39.50
Corruption 37.75
Generosity 39.50
Log of GDP\nper capita 38.75
Healthy life\nexpectancy 38.25
Name: 0.25, dtype: float64


Apps String Snippet... Vishwakarma I... Imported Fro... Gmail YouTube Maps google classro... Breakdown_c...
jupyter 21920026_Assignment_2 Last Checkpoint: 19 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3
Run
Code
happiness_df.mode()
Out[13]:
Country (region) Ladder SD of Ladder Positive affect Negative affect Social support Freedom Corruption Generosity Log of GDP\nper capita Healthy life\nexpectancy
0 Afghanistan 1 1 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1 Albania 2 2 2.0 2.0 2.0 2.0 2.0 2.0 2.0 2.0
2 Algeria 3 3 3.0 3.0 3.0 3.0 3.0 3.0 3.0 3.0
3 Argentina 4 4 4.0 4.0 4.0 4.0 4.0 4.0 4.0 4.0
4 Armenia 5 5 5.0 5.0 5.0 5.0 5.0 5.0 5.0 5.0
... ...
151 Venezuela 152 152 152.0 152.0 152.0 152.0 NaN 152.0 152.0 NaN
152 Vietnam 153 153 153.0 153.0 153.0 153.0 NaN 153.0 NaN NaN
153 Yemen 154 154 154.0 154.0 154.0 154.0 NaN 154.0 NaN NaN
154 Zambia 155 155 155.0 155.0 155.0 155.0 NaN 155.0 NaN NaN
155 Zimbabwe 156 156 NaN NaN NaN NaN NaN NaN NaN NaN
156 rows × 11 columns
In [14]:
histogram for each feature
for Ladder
happiness_df["Ladder"].hist(bins=15)
Out[14]:
<matplotlib.axes._subplots.AxesSubplot at 0x7f70fb0b8400>


localhost:8889/notebooks/Downloads/21920026_Assignment_2.ipynb
Apps
String Snippet...
Vishwakarma I...
Imported Fro...
Gmail
YouTube
Maps
google classro...
Breakdown_c...
jupyter 21920026_Assignment_2 Last Checkpoint: 19 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3
Run
Code
In [14]:
histogram for each feature
for Ladder
happiness_df["Ladder"].hist(bins=15)
Out[14]:
<matplotlib.axes._subplots.AxesSubplot at 0x7f70fb0b8400>
0 2 4 6 8 10
0 20 40 60 80 100 120 140 160


localhost:8889/notebooks/Downloads/21920026_Assignment_2.ipynb
Apps
String Snippet...
Vishwakarma I...
Imported Fro...
Gmail
YouTube
Maps
google classro...
Breakdown_c...
jupyter 21920026_Assignment_2 Last Checkpoint: 19 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3
Run
Code
In [135]:
for Negative affect
happiness_df["Neg"].hist(bins=15,color = 'yellow')
Out[135]:
<matplotlib.axes._subplots.AxesSubplot at 0x7f70e0afea00>
0 2 4 6 8 10
0 20 40 60 80 100 120 140 160



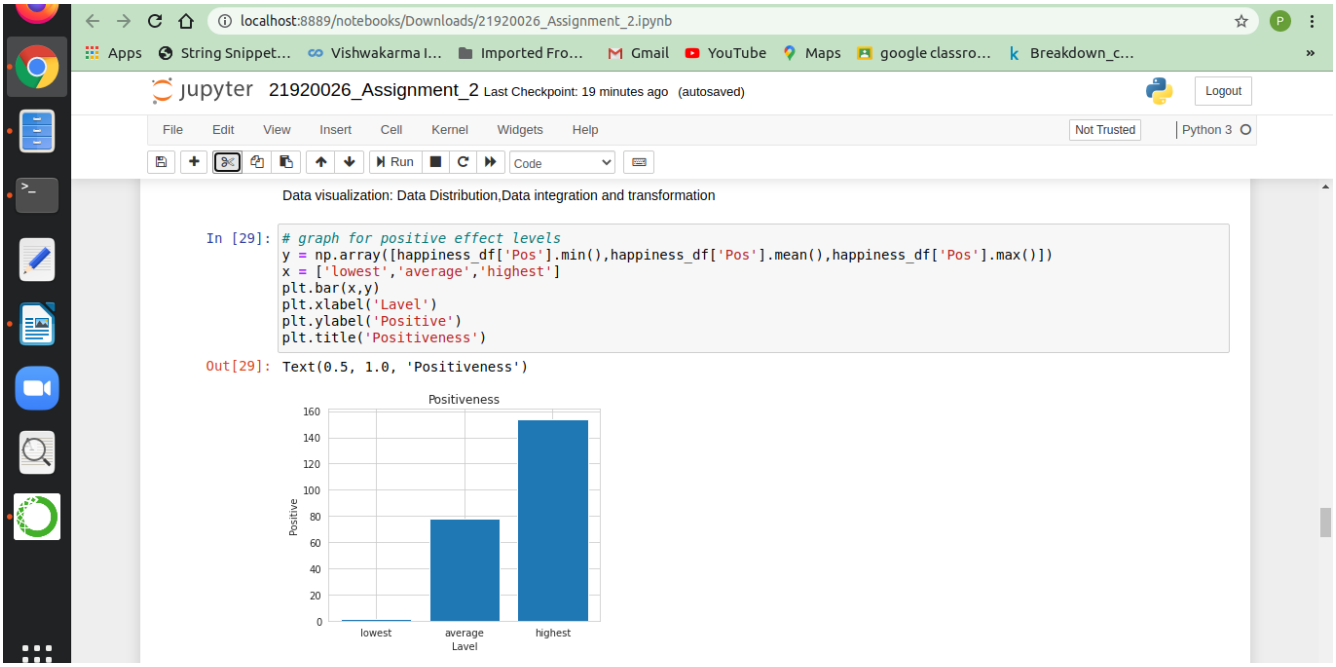
Data Cleaning

In [23]:
```
#Ladder is just another word for ranking so, removing it together with SD of ladder. renaming other cloumns

happiness_df=happiness_df.drop(['Ladder', 'SD of Ladder'], axis=1)

happiness_df = happiness_df.rename(columns={
    'Country (region)':'Country',
    'Positive affect':'Pos',
    'Negative affect':'Neg',
    'Log of GDP\nper capita':'GDP',
    'Healthy life\nexpectancy':'Life expectancy'
})
```

In [24]:
```
happiness_df.head(5)
```

Out[24]:

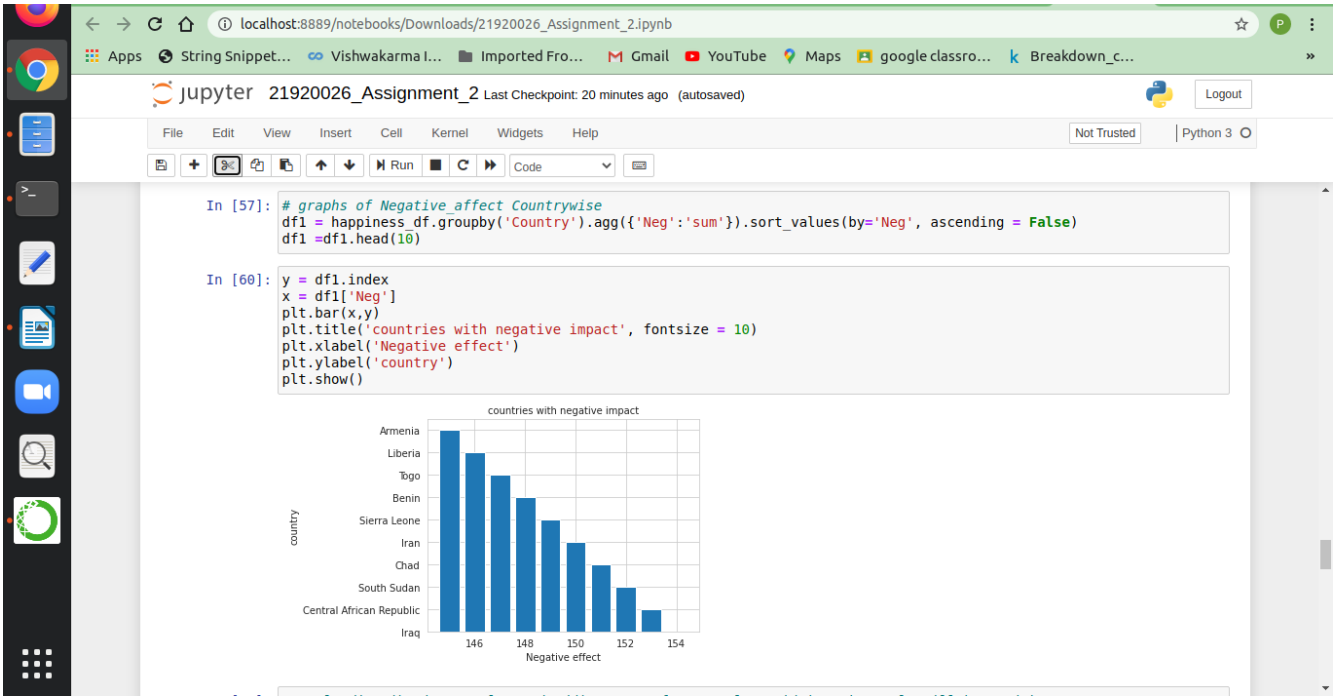
	Country	Pos	Neg	Social support	Freedom	Corruption	Generosity	GDP	Life expectancy
0	Finland	41.0	10.0	2.0	5.0	4.0	47.0	22.0	27.0
1	Denmark	24.0	26.0	4.0	6.0	3.0	22.0	14.0	23.0
2	Norway	16.0	29.0	3.0	3.0	8.0	11.0	7.0	12.0
3	Iceland	3.0	3.0	1.0	7.0	45.0	3.0	15.0	13.0
4	Netherlands	12.0	25.0	15.0	19.0	12.0	7.0	12.0	18.0

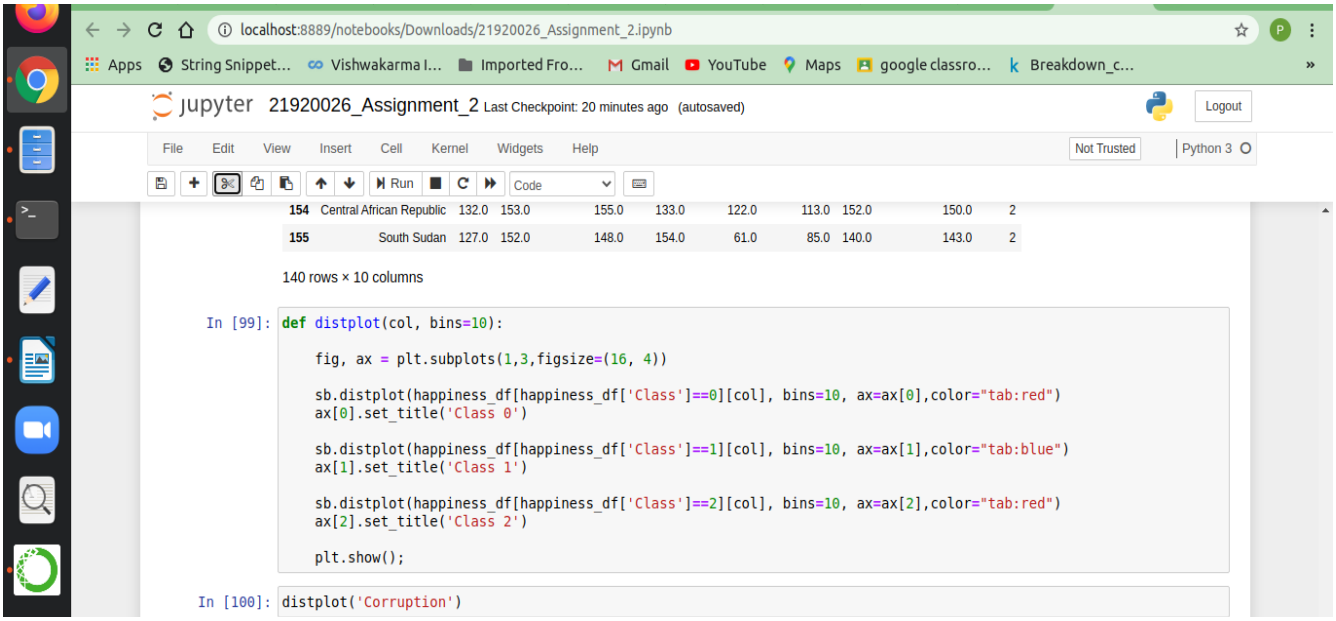


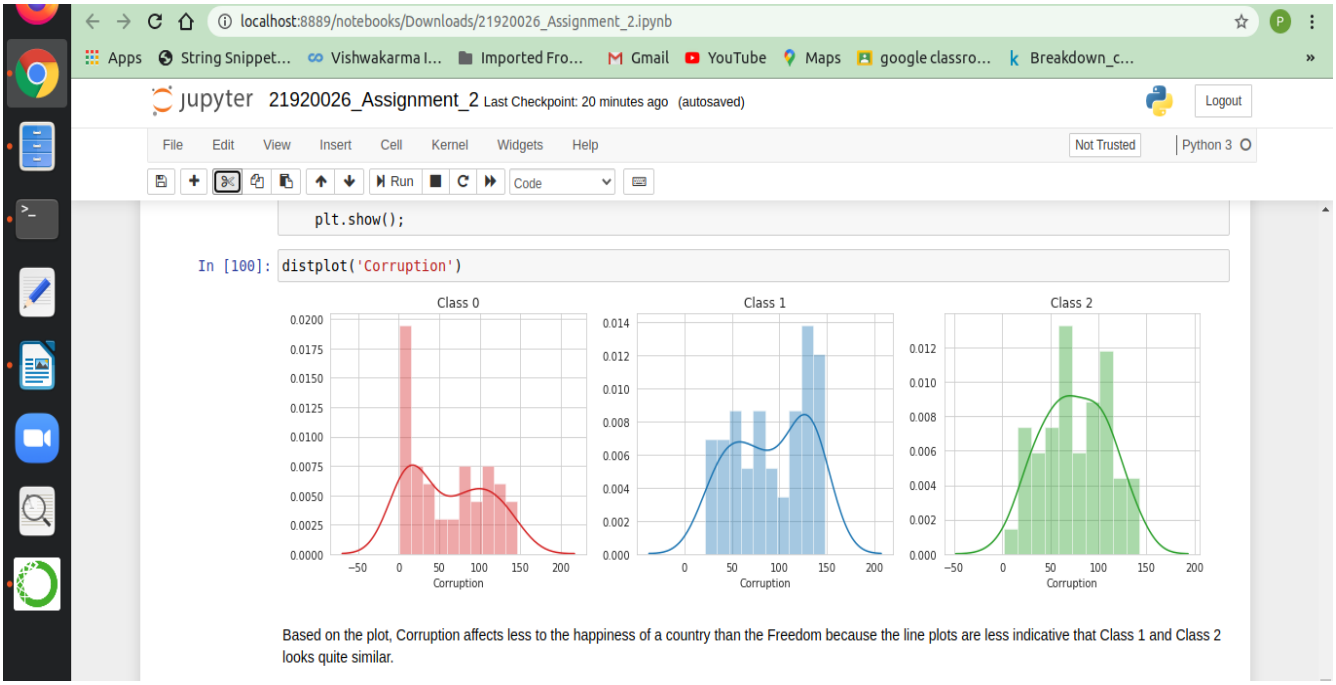
Data visualization: Data Distribution,Data integration and transformation

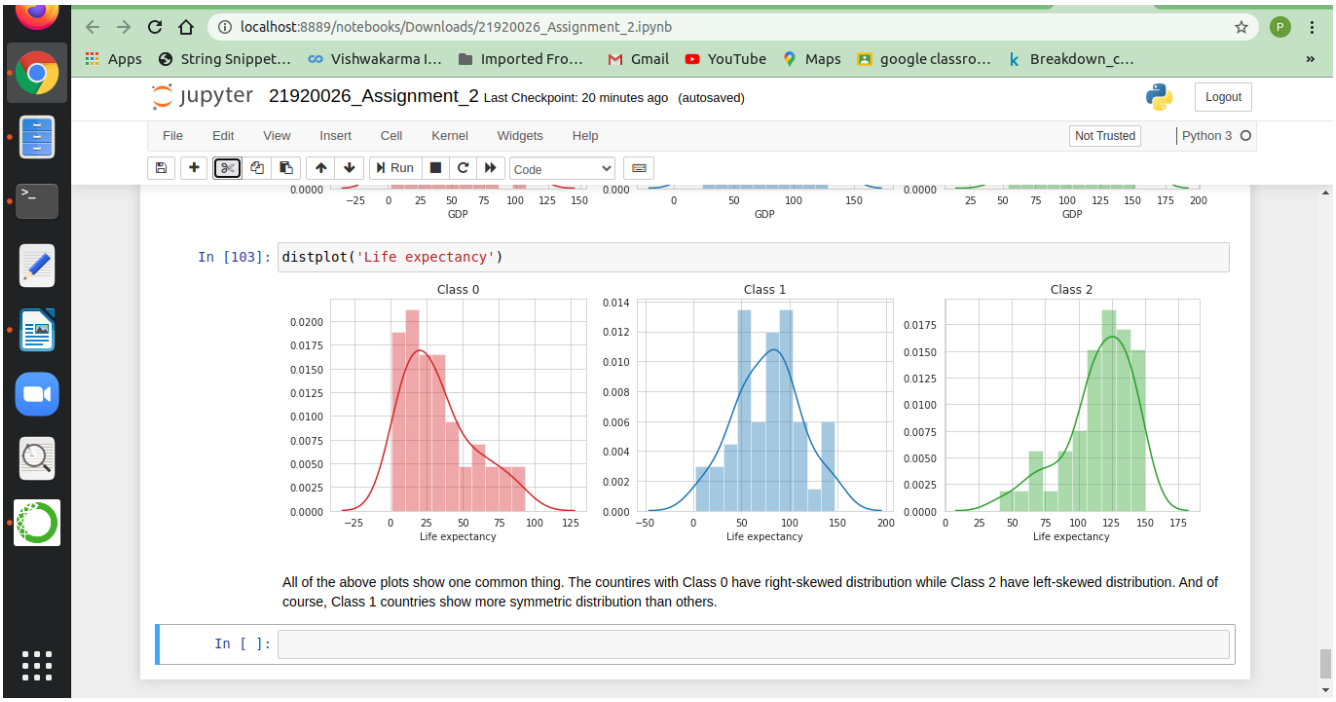
In [29]:
```
# graph for positive effect levels
y = np.array([happiness_df['Pos'].min(),happiness_df['Pos'].mean(),happiness_df['Pos'].max()])
x = ['lowest','average','highest']
plt.bar(x,y)
plt.xlabel('Lavel')
plt.ylabel('Positive')
plt.title('Positiveness')
```

Out[29]: Text(0.5, 1.0, 'Positiveness')


localhost:8889/notebooks/Downloads/21920026_Assignment_2.ipynb
Apps
String Snippet...
Vishwakarma I...
Imported Fro...
Gmail
YouTube
Maps
google classro...
Breakdown_c...
jupyter 21920026_Assignment_2 Last Checkpoint: 20 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3
Run
Code
In [57]:
# graphs of Negative_affect Countrywise
df1 = happiness_df.groupby('Country').agg({'Neg':'sum'}).sort_values(by='Neg', ascending = False)
df1 =df1.head(10)
In [60]:
y = df1.index
x = df1['Neg']
plt.bar(x,y)
plt.title('countries with negative impact', fontsize = 10)
plt.xlabel('Negative effect')
plt.ylabel('country')
plt.show()
countries with negative impact
Armenia
Liberia
Togo
Benin
Sierra Leone
Iran
Chad
South Sudan
Central African Republic
Iraq
country
146
148
150
152
154
Negative effect

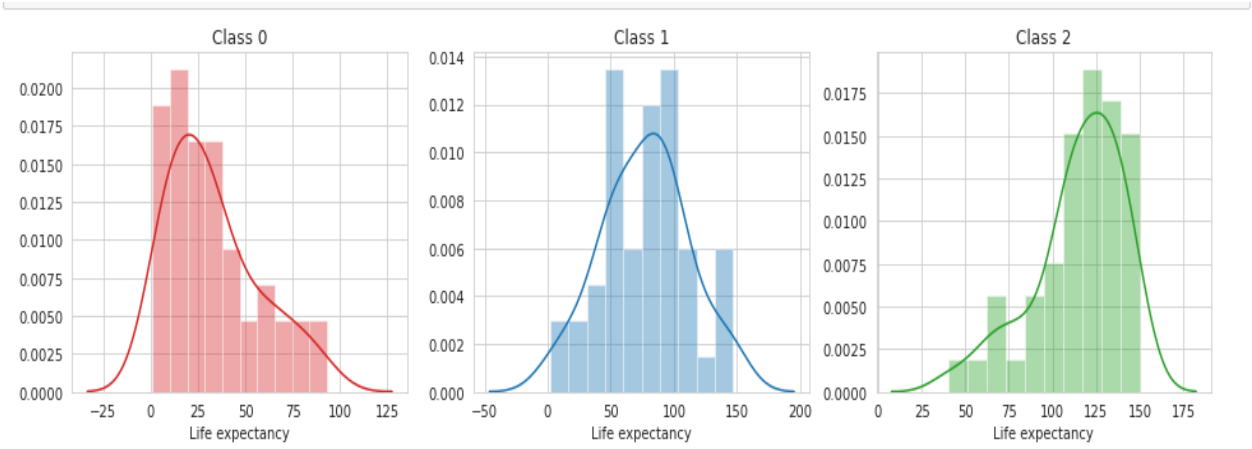

localhost:8889/notebooks/Downloads/21920026_Assignment_2.ipynb
Apps
String Snippet...
Vishwakarma I...
Imported Fro...
Gmail
YouTube
Maps
google classro...
Breakdown_c...
jupyter 21920026_Assignment_2 Last Checkpoint: 20 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3
Run
Code
154 Central African Republic 132.0 153.0 155.0 133.0 122.0 113.0 152.0 150.0 2
155 South Sudan 127.0 152.0 148.0 154.0 61.0 85.0 140.0 143.0 2
140 rows × 10 columns
In [99]:
def distplot(col, bins=10):
fig, ax = plt.subplots(1,3,figsize=(16, 4))
sb.distplot(happiness_df[happiness_df['Class']==0][col], bins=10, ax=ax[0],color="tab:red")
ax[0].set_title('Class 0')
sb.distplot(happiness_df[happiness_df['Class']==1][col], bins=10, ax=ax[1],color="tab:blue")
ax[1].set_title('Class 1')
sb.distplot(happiness_df[happiness_df['Class']==2][col], bins=10, ax=ax[2],color="tab:red")
ax[2].set_title('Class 2')
plt.show();
In [100]:
distplot('Corruption')


localhost:8889/notebooks/Downloads/21920026_Assignment_2.ipynb
jupyter 21920026_Assignment_2 Last Checkpoint: 20 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3
Run
Code
plt.show();
In [100]: distplot('Corruption')
Class 0
Class 1
Class 2
Corruption
Corruption
Corruption
Based on the plot, Corruption affects less to the happiness of a country than the Freedom because the line plots are less indicative that Class 1 and Class 2 looks quite similar.


localhost:8889/notebooks/Downloads/21920026_Assignment_2.ipynb
jupyter 21920026_Assignment_2 Last Checkpoint: 20 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3
Run
Code
GDP
GDP
GDP
In [103]: distplot('Life expectancy')
Class 0
Class 1
Class 2
Life expectancy
Life expectancy
Life expectancy
All of the above plots show one common thing. The countires with Class 0 have right-skewed distribution while Class 2 have left-skewed distribution. And of course, Class 1 countries show more symmetric distribution than others.
In [ ]:

Description:

- summary statistics for each feature in the dataset minimum value, maximum value, mean, range, standard deviation,variance.
- Using min(), max(), sum(), var(), std(), mean(), mode(), median() function summary of statistics is calculated as shown in above screenshots.
- For data visualization histogram, barchart etc graph are drawn for each feature in dataset.
- In Data Cleaning cloumn name of dataset are changed in convinent form using rename(), missing values are droped using dropna() and unnecessary features are droped i.e. Ladder and SD of Laddder.
- Countries in dataset are divided in 0,1 and 2 (0=happy, 1=nutral ,2=sad/not less happy)classes to measure happiness of country. Inferances are drawn from visualization as shown in following fig.



All of the above plots show one common thing. The countires with Class 0 have right-skewed distribution while Class 2 have left-skewed distribution. And of course, Class 1 countries show more symmetric distribution than others.

Interpretation:

- Class 2 (less happy/ sad country) have left-skewed means there is a long tail in negative direction on the number line so the countries which are less happy/sad having less Life expectancy, low Social support and more Log of GDP.
- Simillary, The countires with Class 0 (happy country) have right-skewed distribution which means countries are happy having more Life expectancy, mare Social support and less Log of GDP.
- ranked 1 happiest country in world is Finland.
- Happiest countries among each group (class 0,1 and 2) are Finland, South Korea, Laos respectively.
- Corruption affects less to the happiness of a country than the Freedom .