THE HUMAN FREEDOM INDEX DATA

Dataset and Domain

Data Dictionary :-

THE HUMAN FREEDOM INDEX measures economic freedoms such as the freedom to trade or to use sound money, and it captures the degree to which people are free to enjoy the major freedoms often referred to as civil liberties—freedom of speech, religion, association, and assembly— in the countries in the survey. In addition, it includes indicators on rule of law, crime and violence, freedom of movement, and legal discrimination against same-sex relationships. We also include nine variables pertaining to women-specific freedoms that are found in various categories of the index.

Variable categorization (count of numeric and categorical):

The Human Freedom Index presents a broad measure of human freedom, understood as the absence of coercive constraint. It uses 79 distinct indicators of personal and economic freedom in the following areas:

- Rule of Law
- Security and Safety
- Movement
- Religion
- Association, Assembly, and Civil Society
- Expression and Information
- Identity and Relationships
- Size of Government
- Legal System and Property Rights
- Access to Sound Money
- Freedom to Trade Internationally
- Regulation of Credit, Labor, and Business

There are 123 columns in the data out of which 3 objects, 1 integer and 119 floats data types are present.

- Pre Processing Data Analysis (count of missing/ null values, redundant columns, etc.) –

```
pf_identity_legal                      85.939643
pf_identity_divorce                    59.876543
pf_association_prof_establish          50.137174
pf_religion_estop_establish            50.137174
pf_association_sport_operate           50.137174
pf_association_sport_establish         50.137174
pf_association_prof_operate            50.137174
pf_religion_estop_operate              50.137174
pf_association_political_operate       50.137174
pf_association_political_establish     50.137174
pf_rol_criminal                        39.643347
pf_rol_civil                           39.643347
pf_rol_procedural                      39.643347
pf_ss_women_inheritance_daughters      37.105624
pf_ss_women_inheritance_widows         37.105624
pf_identity_parental_divorce           36.694102
pf_identity_parental_marriage          36.694102
```

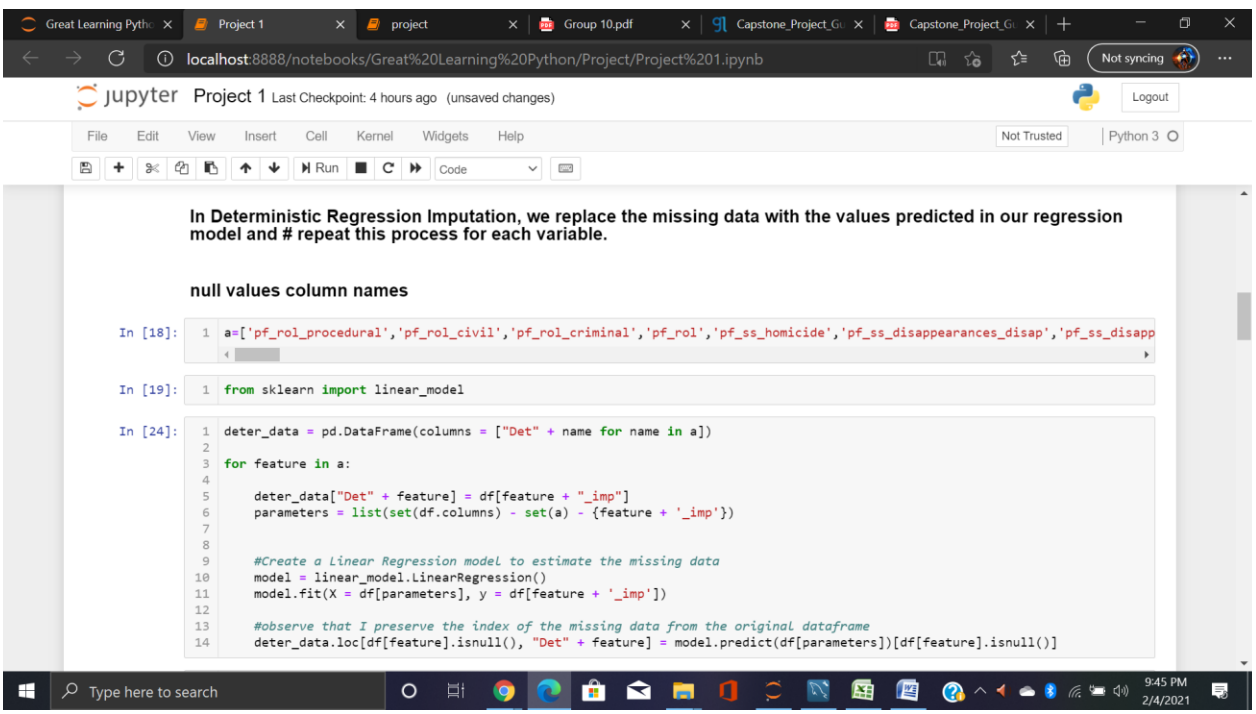
```
pf_expression_cable                    22.976680
pf_expression_newspapers               22.976680
pf_expression_internet                 22.565158
pf_association_association             22.565158
pf_religion_estop                      22.565158
pf_association                         22.565158
pf_association_sport                   22.565158
pf_association_prof                    22.565158
pf_association_political               22.565158
pf_association_assembly                22.565158
ef_legal_integrity                     18.998628
ef_government_tax_payroll              13.237311
pf_ss_disappearances_organized         12.277092
ef_regulation_business_bribes          12.002743
ef_regulation_credit_ownership         11.796982
pf_ss_women_fgm                        11.796982
ef_regulation_labor_firing             11.728395
ef_regulation_labor_bargain            11.659808
ef_trade_regulatory_nontariff          11.659808
ef_legal_protection                    11.591221
ef_legal_crime                         11.591221
ef_legal_police                        11.591221
ef_regulation_business_adm             11.591221
ef_trade_tariffs_revenue               11.591221
ef_legal_judicial                      11.454047
ef_trade_movement_foreign              11.248285
ef_government_transfers                10.973937
pf_movement_women                       9.670782
ef_government_tax_income                8.504801
ef_government_tax                       8.504801
pf_ss_women_missing                     8.230453
pf_ss_women_inheritance                 8.161866
ef_regulation_labor_dismissal           7.544582
ef_government_enterprises               7.133059
ef_regulation_business_bureaucracy      6.995885
ef_regulation_business_licensing        6.927298
ef_legal_restrictions                   6.858711
pf_ss_women                             6.858711
ef_regulation_credit_interest           6.858711
pf_identity_parental                    6.858711
pf_movement_foreign                     6.721536
pf_movement_domestic                    6.721536
pf_religion_harassment                  6.447188
pf_religion_restrictions                6.447188
ef_trade_tariffs_mean                   6.310014
ef_regulation_labor_minwage             6.241427
ef_trade_tariffs_sd                     6.241427
ef_legal_enforcement                    6.172840
pf_religion                             6.172840
ef_regulation_business_compliance       6.172840
ef_trade_regulatory_compliance          6.172840
ef_regulation_business_start            6.172840
ef_trade_movement_capital               6.104252
pf_ss_disappearances_disap              6.104252
ef_regulation_labor_hours               6.035665
ef_trade_black                          5.967078
ef_trade_movement_visit                 5.829904
ef_trade_tariffs                        5.829904
ef_regulation_labor                     5.761317
ef_trade_regulatory                     5.761317
ef_regulation_business                  5.761317
pf_identity_sex_male                    5.692730
ef_money                                5.624143
ef_trade                                5.555556
ef_regulation_labor_conscription        5.555556
pf_rol                                  5.486968
pf_ss_disappearances_fatalities         5.486968
pf_movement                             5.486968
pf_ss_homicide                          5.486968
pf_ss_disappearances_violent            5.486968
pf_ss_disappearances_injuries           5.486968
pf_ss                                   5.486968
pf_ss_disappearances                    5.486968
hf_quartile                             5.486968
pf_score                                5.486968
pf_expression_killed                    5.486968
ef_legal_courts                         5.486968
hf_score                                5.486968
ef_rank                                 5.486968
ef_score                                5.486968
```

```
ef_regulation                      5.486968
ef_regulation_credit               5.486968
ef_trade_movement                  5.486968
ef_money_currency                  5.486968
pf_expression_jailed               5.486968
ef_legal_military                  5.486968
ef_legal                           5.486968
ef_government                      5.486968
pf_identity_sex                    5.486968
pf_expression_influence            5.486968
pf_rank                            5.486968
pf_expression_control              5.486968
pf_expression                      5.486968
pf_identity_sex_female             5.486968
hf_rank                            5.486968
pf_identity                        5.486968
ef_regulation_credit_private       4.938272
ef_money_inflation                 4.938272
ef_money_sd                        4.938272
ef_money_growth                    4.801097
ef_government_consumption          4.526749
ef_legal_gender                    1.646091
region                             0.000000
countries                          0.000000
ISO_code                           0.000000
year                               0.000000
```

The above data is percentage of null values present in the each columns

To treat them we have used Deterministic Regression Imputation

In Deterministic Regression Imputation, we replace the missing data with the values predicted in our regression model and repeat this process for each variable.



All the columns are filled with values predicted by Deterministic Linear Regression Imputer. After applying the imputer method every feature has 0 null value in it.

- Project Statement:

Primary Goal: Predicting Human freedom score, rank and quartile based on Personal Freedom score and Economic freedom score for all countries.

Secondary Goal: To analyse India's position for consecutive years from 2008 - 2017 and suggest which attributes to focus on to improve its position for upcoming years.

Complexity involved:

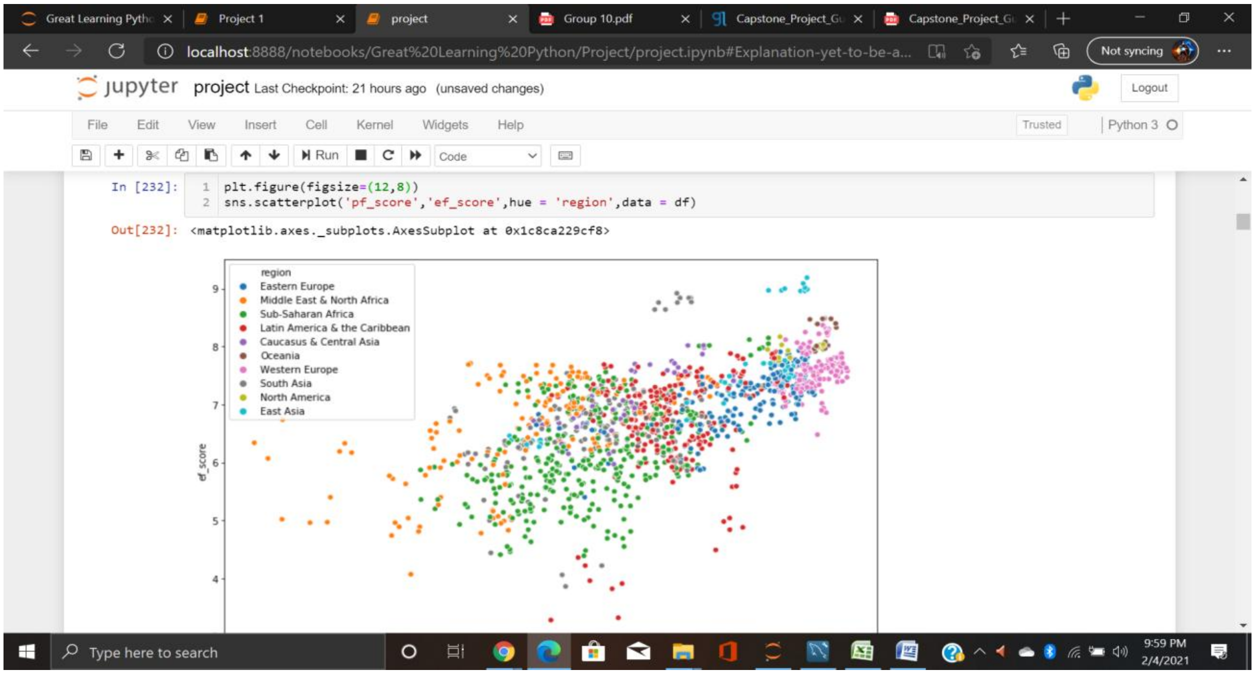
- Eliminating Features based on multi collinearity was the biggest challenge we encountered in this project.
- Imputing null values for the dataset also posed a challenge. It was achieved through Deterministic Linear Regression Imputer.
- Finding relationship between variables when there is 123 columns present in the data was hard.

Project Outcome – Commercial, Academic or Social value:

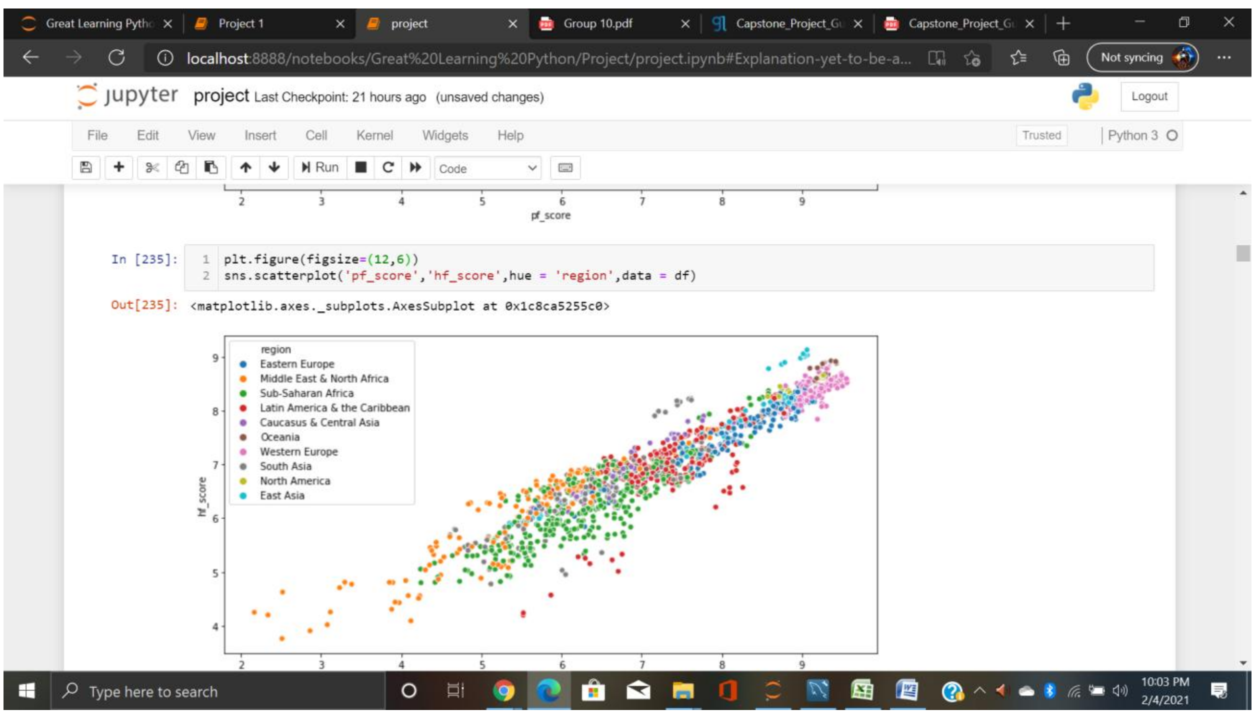
Through this project we have found why India's score is consistently bad in the Human Freedom Index and we have suggested what to do to improve India's score in upcoming reports.Data Exploration (EDA):

- Relationship between variables

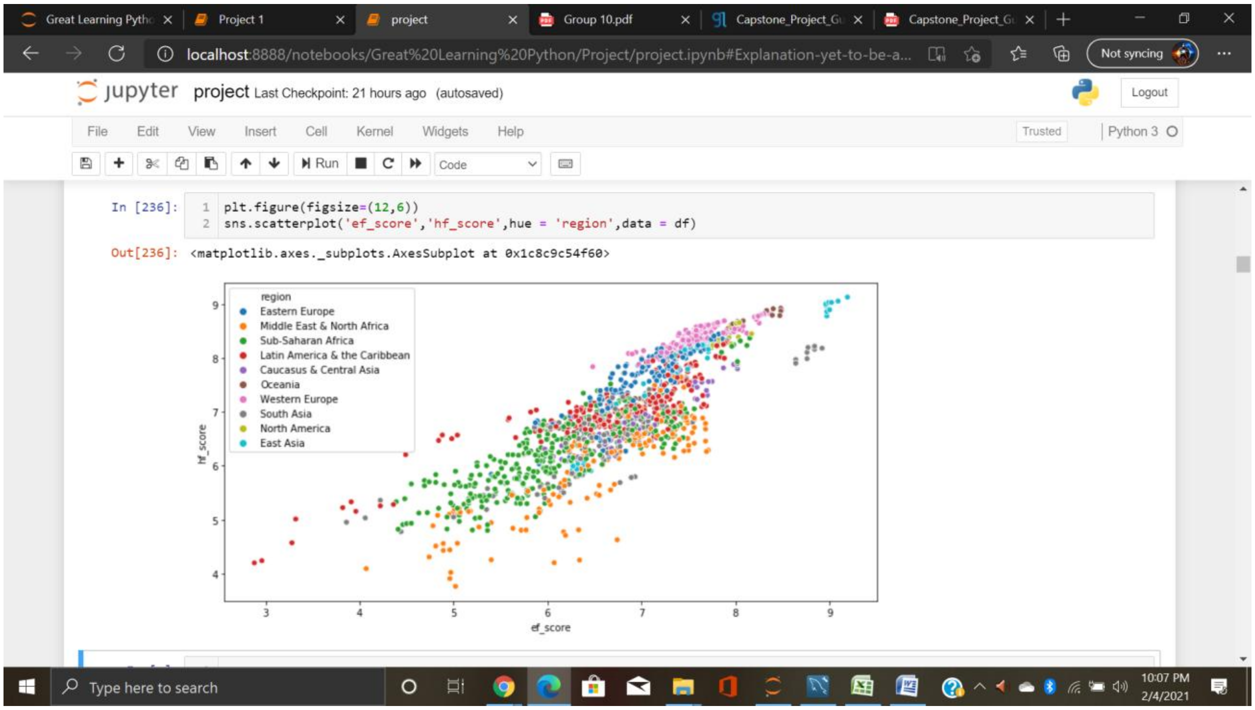
The variables pf_score(Personal Freedom score) and ef_score(Economic Freedom score) both constitute(average) to hf_score(Human Freedom Score). Below plot shows how pf_score and ef_Score is affecting each other based on each regions.



We can clearly see East asia has both high pf_score and ef_score.



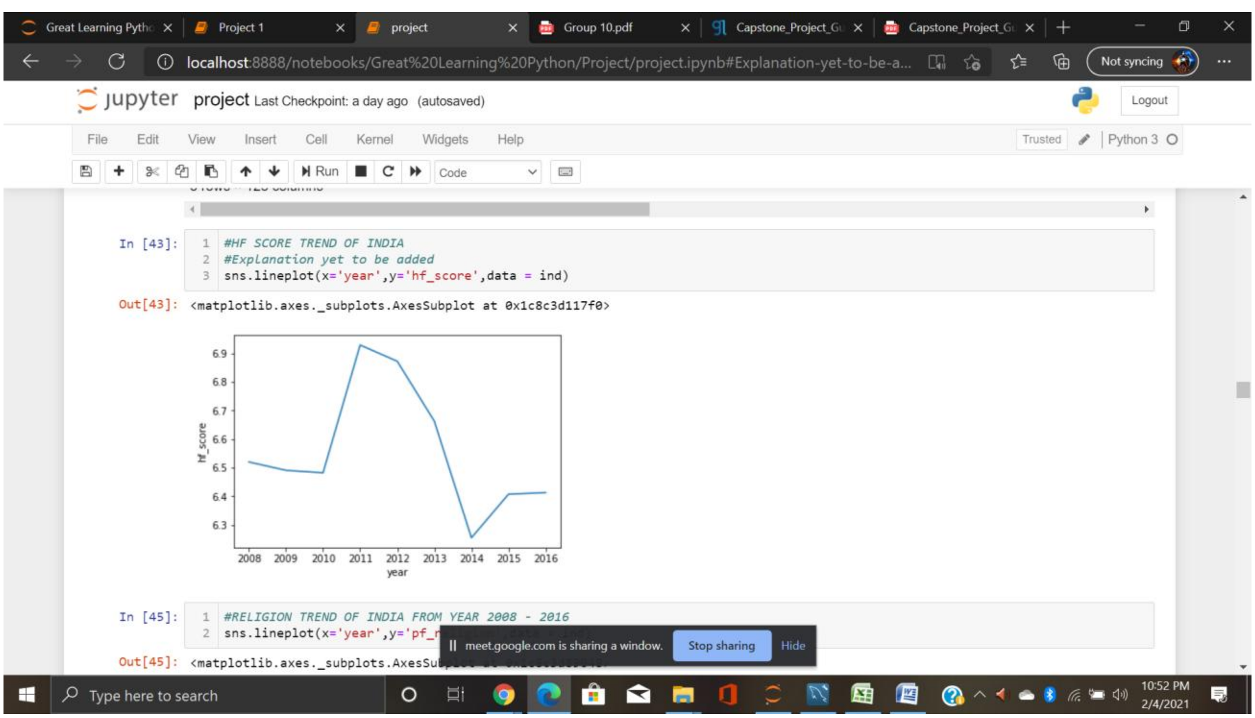
We can see how Western Europe and East Asia has high values of both personal and Human Freedom score. This is obvious given the countries present in these regions has high GDP, High Per Capita Income and High Standard of Living.



The Economic Freedom Score and Human Freedom score for East Asia is high given their ease of business doing laws.

Surprising thing is South Asia has also high score of Economic score. In further enquiry it was found that Bangladesh and Srilanka's liberal policies when it comes to manufacturing and Trading respectively impacted the score of the region.

INDIA'S POSITION:



India's position was gradual until 2010 when it had sudden spike after 2011 there was a sudden drop. India' position gradulally picked up after 2014.

The reason for India's drop was India has got very less score for religious freedom since 2012 and got drastic less score from 2015 to 2018 .

And with respect to another Identity and Relationships it got very less score in 2014 and it also maintained less score til 2018 ...

And with respect to Rule of Law eventough India maintained its score throughout these years but it is very less compared to other countries

These 2 majorly caused India Rank to fall

- Check for:

- **<u>Distribution of variables</u>**

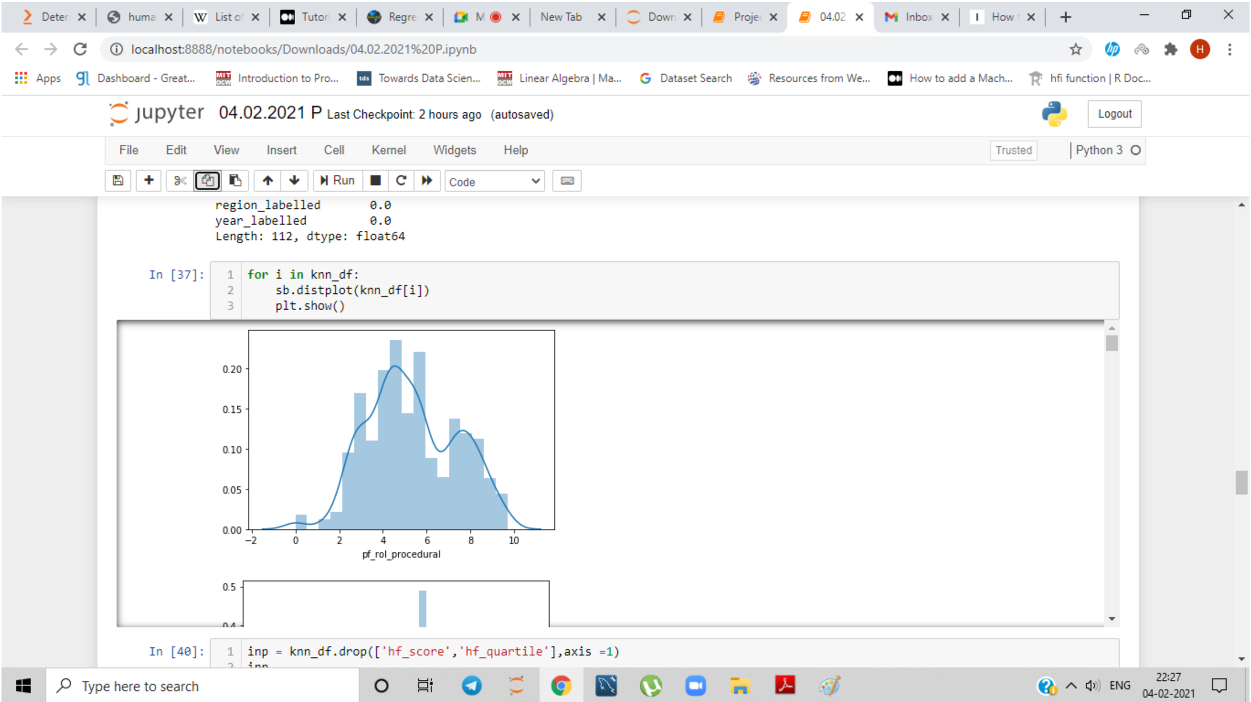
All the variables present in the feartures has score between 1 to 10 with many variables heavily skewed. The rank columns has data ranging from 1 to 162. It requires transformation.

- **<u>Multi-collinearity</u>**

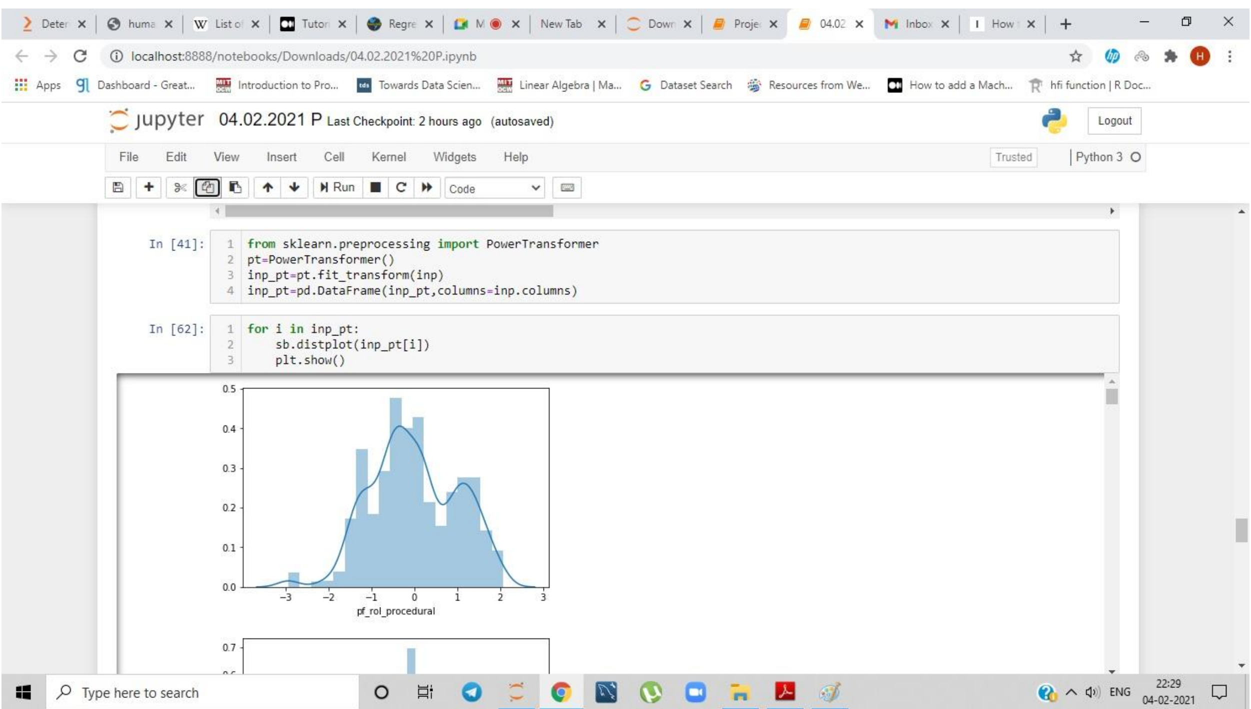
All the features show high multicollinearity which is shown by VIF values and correlation tables.

The Problem of Multicollinearity is solved by using Principal Component Analysis in our Project on which we got 58 components.

Before Transformation:

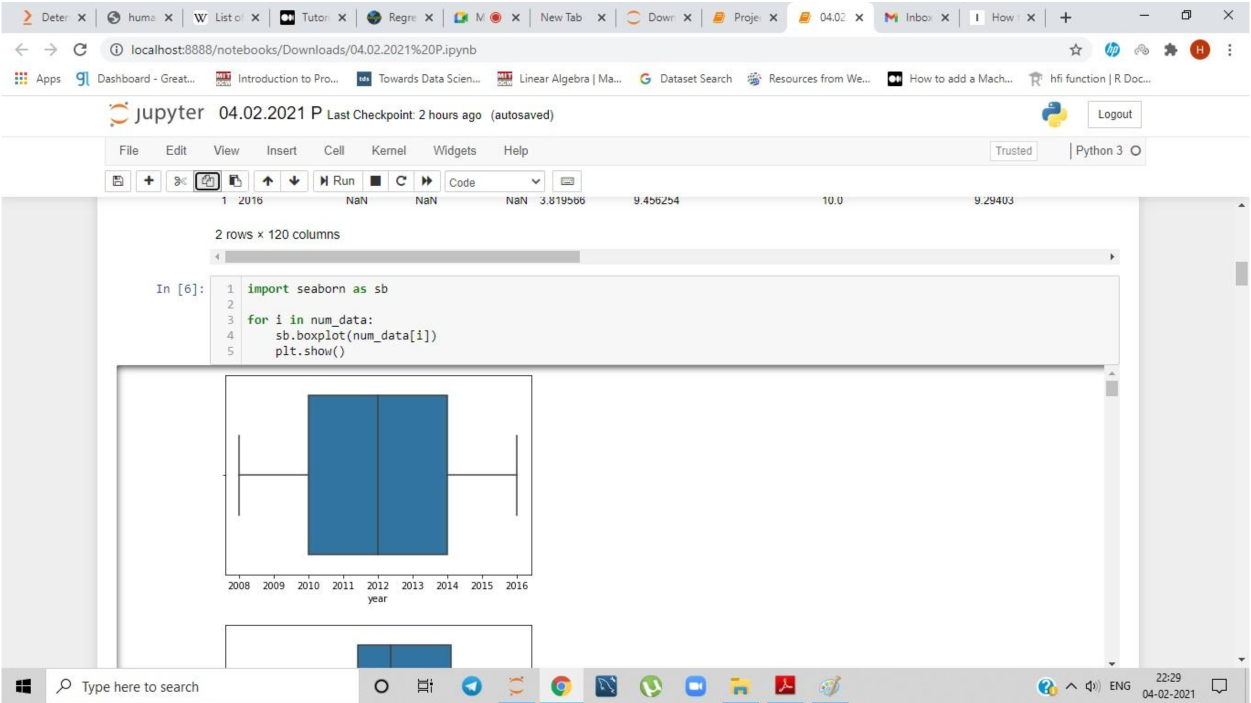


After Transformation:



- ## Presence of outliers and its treatment:

The presence of outliers in this dataset are all real since the variables score is based on scoring. Each country value is real. So we are not able to classify them as outliers. They just showed heavy skewing which has been taken care by Power transformation.



Feature Engineering :

- Whether any transformations required –

Yes transformation is required since we have three columns hf_rank ,ef_rank and pf_rank where the variables range from 1 to 162 whereas other features has data from 1 to 10.

This was achieved using Power Transformation.

```
In [59]:
from sklearn.preprocessing import PowerTransformer
pt=PowerTransformer()
inp_pt=pt.fit_transform(inp_sc)
inp_pt=pd.DataFrame(inp_pt,columns=inp.columns)
```