# **ADTA 5410**

**Applications and Deployment of Advanced Analytics**

**Assignment-3**

1. This question involves the use of simple linear regression on the [OWID data set](https://covid.ourworldindata.org/data/owid-covid-data.csv).
2. (8 points) Use the pd.read\_csv(…., index\_col="…") function to read the data into Python. Call the loaded data covid. Set the date columns as index, using index\_col="…". Make sure that you have the directory set to the correct location for the data. Check the first five rows of the data.
3. (8 points) How many rows are in this data? How many columns?
4. (8 points) Check if all of your columns have the same number of data, using .info().
5. Fill the missing data as follows:
   1. (8 points) Fill any missing values (NaN) in the "total\_cases" column with 0, using covid['total\_cases']=covid['total\_cases'].fillna(value=0)
   2. (8 points) Drop all the rows with missing gdp\_per\_capita and stringency index, using

covid\_clean = covid.dropna(subset=['gdp\_per\_capita','stringency\_index'])

* 1. Check the number of rows and columns again.

1. (8 points) Use the sm.OLS() function to perform a simple linear regression with “total\_cases” as the response and the following variables (see the table below) as the predictors. Print the regression output and comment on the result.
2. (8 points) Comment on the signs and the significance of the relationships between the predictors and the response variable.
3. (8 points) What percent of the response variable is explained by the predictors? Comment on R-squared.
4. (8 points) What is the predicted “total\_cases” associated with the value given in the value column?
5. (8 points) Plot scatter plots of all variables (the response and each of the predictors) using pd.plotting.scatter\_matrix(). Comment on the results.

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Predictor Variable | Value | Countries |
| *Biniam Abebe, Cynthia Ani* | Stringency index, gdp per capita | 50, $40,000 | Asia |
| *Akhila Chilukuri, Anthony Cisneros* | Stringency index, gdp per capita | 50, $40,000 | Europe |
| [*Saicharan Goud Gunegari*](https://unt.instructure.com/courses/106763/users/357805)*, Tharuni Dusakanti* | Stringency index, gdp per capita | 50, $40,000 | N. America |
| *Keerthi Erram, Briaunna Fontaine-Deth* | Stringency index, gdp per capita | 50, $40,000 | Africa |
| *Prathyusha Gogineni, Niranjan Guda* | Stringency index, gdp per capita | 50, $40,000 | Ocenia |
| *Kaladhar Gundla, Kristin Hassett* | Stringency index, gdp per capita | 50, $40,000 | S. America |
| *Liz Jones, Gowtham Kavvadi* | Stringency index, gdp per capita | 50, $40,000 | All the data |
| *Supriya Kolluru, Shravani Kuragayala* | Stringency index, gdp per capita | 80, $60,000 | Asia |
| *Ian Lynch, Surya Sai Srikanth Machimchetty* | Stringency index, gdp per capita | 80, $60,000 | Europe |
| *Chloe Marshall, G. H. M. Naga Venkata* | Stringency index, gdp per capita | 80, $60,000 | N. America |
| *Shivani Muvva, Chariteash Narra* | Stringency index, gdp per capita | 80, $60,000 | Africa |
| *Trevor Nolen, Nikhila Pinnapureddy* | Stringency index, gdp per capita | 80, $60,000 | Ocenia |
| *Hima Sindhuja Pushadapu, Kayla Reynard* | Stringency index, gdp per capita | 80, $60,000 | S. America |
| *V. Sanath Kumar Suram, Vaishnavi Thalla* | Stringency index, gdp per capita | 80, $60,000 | All the data |
| *Krishna Vamsi Uppala, Hari Babu Uppari* | Stringency index, gdp per capita | 50, $40,000 | Asia |

1. (BONUS- Up to 10% additional) Create your own question and provide the answer.