1. Load dataset from - https://covid.ourworldindata.org/data/owid-covid-data.csv

2. Subset only those rows that have "India" in the "location" column(This subsetted dataframe has to be used for modelling)

3. Univariate Analysis:

   a. Draw histograms of each numerical variable

   b. Find mean, median and mode of each column

4. Bivariate Analysis:

   a. Draw scatter plots of each numerical column versus one another

   b. Draw line plots of each numerical column versus one another

5. Handle Missing values:

   a. If there are null values in numerical column, replace the null values by the mean of that column

   b. If there are null values in categorical column, replace the null values by the mode of that column

   c. If more than 50%the values in a column are null, then drop that entire column

6. Convert date column to ordinal

   a. Code:

      import datetime as dt

      df["date"]=pd.to_datetime(df["date"])

      df["date"]=df["date"].map(dt.datetime.toordinal)

7. Drop all categorical columns

8. Select "total_cases" column as the target variable

9. Select the other columns as the features(the "date" column has to be in the features)

10. Perform train-test split

11. Modelling:

    a. Linear Regression

    b. Random Forest Regressor

12. Get accuracy

13. Predict Total case for a new date

NOTE: To convert anytime back from ordinal to date-time use the following sample code:

from datetime import datetime

ordinal value = 733828 # This is an example

dt = datetime.fromordinal(ordinal value)