

INTRODUCTION TO DATASCIENCE USING PYTHON

BY YUVASHREE.N

BTECH IT

AGENDA:

* **To understand technique such as lambdas and manipulating CSV files.**
* **To describe common python functionality and features used for data Science.**
* **To query data Frame structure for the cleaning and processing.**

**PANDAS LIBRARY**

* **Tools for reading and writing data between in-memory data structure and different formats.**
* **Intelligent data alignment and integrated handling of missing data.**
* **Aggregating or transforming data with a powerful group by engine allowing split-apply-combine operations on data sets.**
* **High performance merging and joining of datasets.**
* **Columns can be inserted and deleted from data structures for size mutability.**

DATAFRAMES

* **DATAFRAME is a two dimensional data structure with columns of potentially different types.**
* **It is like a spread sheet or a sql table ,or a directory of series object.**
* **It is generally the most commonly used pandas object.**
* **Like series, data frame accepts many different kinds of input.**

****

COVID-19 VACCINE ANALYSIS

**OBJECTIVES:**

* **To learn the steps,needed to be taken to prepare the data for an analysis.**
* **To learn how to look at the data to find a good measure to establish the analysis based upon.**
* **To learn to visualize the result of analysis.**

**PROBLEM:**

**Is there any relationship between the spread of the coronavirus and how happy people living in that country are?**

**LIBRARIES ,DATASET AND OBSERVATIONS**

**LIBRARIES: Pandas, numPy, seaborn**

**FEATURES OF COVID\_19 DATA SET:**

**Province:** State

**Lat/long:** latitude and longitude dates.

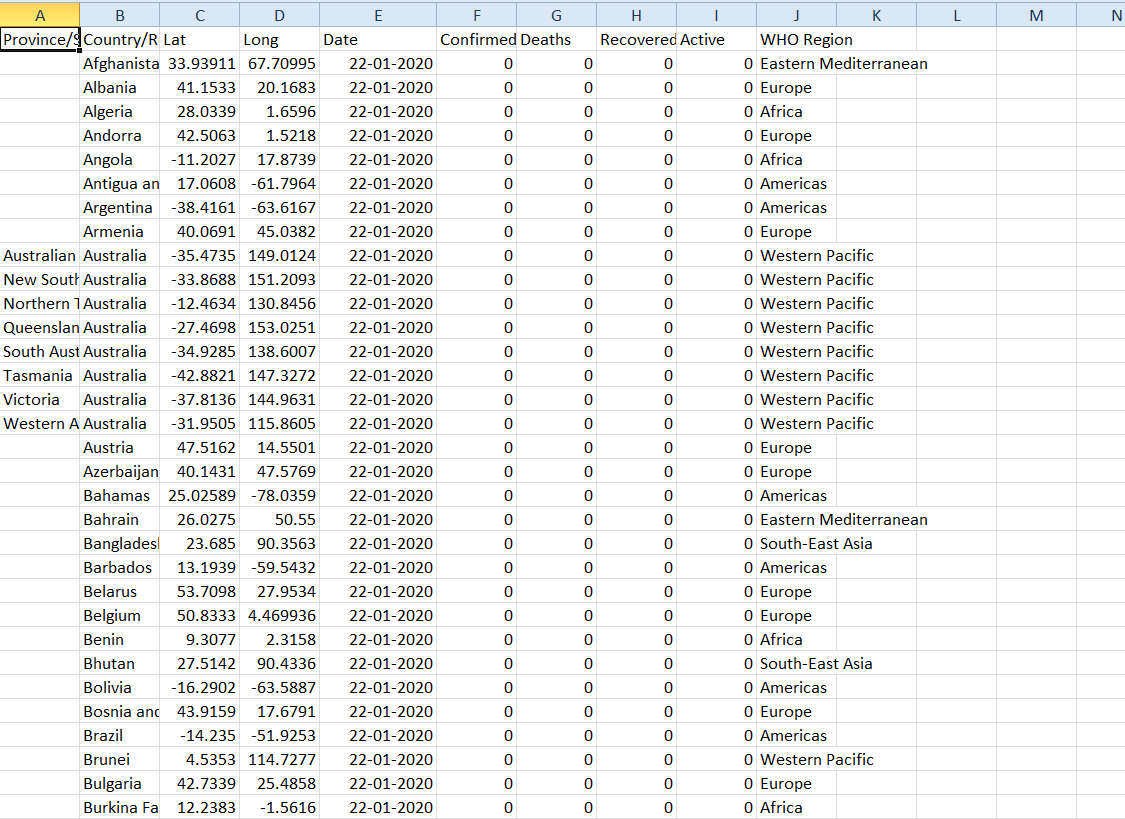
**FEATURES OF HAPPINESS DATA SET:**

**Overall rank:** Rank of the country on the basis of happiness

**Score:** Happiness score of the country/region

**GDP, Social Support, Healthy life, Freedom to make life choices etc**

**DATASET:**

****

**PREPARATION OF DATA FOR CONSUMPTION:**

* **Cleaning data:** removing or modifying data that is incorrect, incomplete, irrelevant, duplicated, or improperly formatted.
* **Grouping:** Grouping the related data.
* **Preparation of data**: Creating maximum infection rate column for storing the infection rate for each country.