**License Info : -**

This project is licensed under the terms of the Apache License 2.0 as <https://tinyurl.com/y85g3ct7>

**About Data Set**

First, i have changed the format of data set as :

Changed the column names as short and compact

Removed the null values

Removed float values

Removed all special characters from course title and site

Removed date column

This data set contains 3473 rows and 12 columns about the information related to Udemy Courses. It contains attributes - sno,id,coursetitle,site,ispaid,price,subscribers,reviews,lectures,level,duration,subject

**The details of the data set are given below:**

**Format:**

The data set contains the following information about different Udemy Courses, they are divided by ‘,’ in the data file:

sno,id,coursetitle,site,ispaid,price,subscribers,reviews,lectures,level,duration,subject - attributes

**Attribute Information**

|  |  |
| --- | --- |
| **Attribute Name** | **Description** |
| Sno | A integer value for serial number |
| Id | A integer value for id’s |
| Coursetitle | A string value for course title |
| Site | A string value for the course site |
| Ispaid | A string value if course is paid or not in true or false |
| Price | A integer value for the course price |
| Subscribers | A integer for the subscribed users for each course |
| Reviews | A integer value for the reviews given by the students |
| Lectures | A integer for the no of lectures to be there for each course |
| Level | A string for kind of difficulty levels |
| Duration | A integer for the duration in hours to complete the course |
| Subject | A string for each subject there are different courses |

**Reference**

<https://www.kaggle.com/andrewmvd/udemy-courses>

Also I have shared my data set sample as DataSample.txt

**Big data project implemented by Aniket**