

**Applied Scripting Languages**

**Assignment Part 1: Selecting a Data Set**

**Student ID:** A00291307

(To facilitate anonymous marking, do not include your name in this report).

# Introduction

**Name of Data Set:** Udemy Courses

**Source (URL):** https://www.kaggle.com/tanaypatare/udemy-courses-analysis/data

**Provide a brief description of what your dataset represents:**

The data represents a list of Udemy Courses from the year 2011 to 2017.

The data has a course title as well as a primary key of course id to uniquely identify each course.

The data has been primarily divided based on 4 subjects Web Development, Graphic Design, Musical Instruments and Business Finance. There is a secondary category for the data based on the level of the course.

The data also contains the price, duration, subscribers, and reviews for the course.

# Data

**Number of records/lines/rows:** 3678

**Number of fields/columns:** 10

**Complete the following table, providing details of each field/column.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Field/Column**  **Name** | **Description** | **Python**  **Data Type** | **Example** |
| 1 | course\_id | A primary key with which each of the courses in the list can be uniquely identified | str | 1282064 |
| 2 | course\_title | The name by which a course was added to the list. | str | Angular 4 Front to Back |
| 3 | price | The price for which the course is listed (A price of 0 would imply a free course) | int | 20 |
| 4 | num\_subscribers | The number of learners who have subscribed to the course. | int | 3697 |
| 5 | num\_reviews | The number of reviews available about the course. | int | 195 |
| 6 | num\_lectures | The number of video lectures in the course. | int | 48 |
| 7 | level | The level of experience for which the course is recommended. (4 default master data available Beginner, Expert, Intermediate and All Levels) | str | All Levels |
| 8 | content\_duration | The number of hours required to complete the course | float | 3.5 |
| 9 | published\_timestamp | The DateTime stamp at which the course was published on Udemy | str | 2011-07-09T05:43:31Z |
| 10 | subject | The subject on which the course is based on (4 default master data available Web Development, Graphic Design, Musical Instruments and Business Finance) | str | Web Development |

**Provide a screenshot of an excerpt of the data.**

Table

Description automatically generated with medium confidence

# Reason for Selection

**Describe why you selected this data set. Why is it important or of interest to you?**

In today’s world where you can learn anything and everything from the internet through paid as well as free content, I hope this data would give insights into how websites like Udemy has influenced professionals or people who’d like to pursue a future career in a particular industry.

As for me, I have always tried upskilling myself in a wide variety of subjects, to keep myself up to date in the industry and to have a bigger picture of how a whole software system works. Online-contents and courses have been immensely helpful to me in this aspect, especially the ease at which these things were available, how organised these courses were and how cheap these actually are for the level of knowledge it provides.

In conclusion, the data that I’ve picked is of great interest to me personally as well as it’d give a lot of insights into how online education has affected the world in the past few years.

# Proposed Analysis

**Describe what information/answers you wish to extract from the data set.**

1. What among the 4 subjects Web Development, Musical Instruments, Graphic Design and Business Finance is a more popular career to pursue.
2. What do people use these courses for? A high-level basic insight into the subject or do pursuers aim to get an advanced level of knowledge from these courses? (Using analysis of the level column)
3. How many people try to help the community out by providing helpful reviews, after pursuing the course? (Comparing the number of subscribers to the number of reviews available)
4. Are people more interested in pursuing longer more explanatory courses or smaller basic information courses? (Comparing content duration to the number of subscribers)
5. Do people prefer free or paid courses? If people prefer paid courses, what range of prices do most people expect?
6. What year did platforms like Udemy gain popularity?

(Suggested Word Count: 500 words).