Question / Problem Statement:

This is a data set which contains all the courses that published between 2011-07-09 to 2017-07-06 for a Udemy courses, It aims to analyze the behavior of the coach with the highest rating and the most requested courses, price of the courses

- Who is the coach with the best highest rating, and the most subscriptions to his training courses?
- Who is the most teacher who has done training courses?
- What are the most requested courses based on the number of subscriptions?
- What is the most expensive and cheapest price for the course you want to participate in?

Data Description:

- course_id: digit integer number uniquely, assigned to each courses cannot be NULL
- course_title: It is a text, for each course that cannot be NULL
- url: text, for each course that cannot be NULL
- is_paid: boolean
- price: integer number, price per course cannot be NULL
- num_subscribers: integer number
- num_reviews: integer number
- num_lectures: text, how many lectures in the course
- level: text, course level
- content_duration: float, How many hours in the course
- published_timestamp: Dates, day and time when each course was generated cannot be NULL
- subject: text, name of courses

No. of rows: 3679
No. of column: 12

Tools:

- Programs: Python , SQL
- Libraries: Pandas, Matplotlib
- Function:
- Plots:

MVP Goal:

The goal of this project is to better understand which courses are most important in driving revenue, most outstanding teachers and Knowing the price level of the courses according to the highest or lowest