



PROBLEM STATEMENT

Udemy Courses Data Analysis



PROBLEM STATEMENT

The objective of this analysis is to perform an Exploratory Data Analysis (EDA) on a dataset containing information about apps available on the Google Play Store.

The goal is to extract valuable insights that can guide app developers, marketers, and businesses in understanding app performance, user preferences, and industry trends.

Specifically, the analysis focuses on evaluating app characteristics such as ratings, reviews, pricing models, and categories, as well as identifying any correlations or patterns that might influence user behavior and app success.

BUSINESS PROBLEM OVERVIEW

The dataset provided contains various attributes of Udemy courses, including details about course pricing, number of subscribers, reviews, lectures, and the level of difficulty. The primary goal of this analysis is to derive insights into the distribution and popularity of courses, with a specific focus on paid and free courses, and the performance of Python-related courses.

Another important aspect of this analysis is identifying the most popular courses, with a focus on the course with the highest number of subscribers. A list of the top 10 most popular courses will be created based on subscriber count.

In addition, the analysis will examine which courses have the highest number of reviews, providing insights into student engagement and satisfaction. Further, the relationship between course pricing and the number of reviews will be explored to determine if there is any correlation between the two factors.



Udemy courses are online learning modules offered on the Udemy platform, which is a widely used marketplace for education. These courses span a broad range of subjects, from programming and web development to personal growth and creative arts. The code provided analyzes various attributes of these courses, offering insights into their characteristics and performance.

The course title is used to identify the course and categorize it by its content. Every course is also assigned a unique course ID, which helps differentiate one course from another in the system. One of the key attributes is whether the course is paid or free. Paid courses are marked as True in the dataset, while free courses are marked as False. This classification helps in analyzing the distribution and popularity of paid versus free courses.

The price attribute further complements this by showing the cost of paid courses and highlighting whether there's any correlation between course price and other factors like the number of reviews.

Another important metric is the number of subscribers, which represents how many learners are enrolled in a given course. This helps indicate the popularity of a course. Along with the number of subscribers, the number of reviews is also analyzed, as reviews are a direct reflection of student engagement and satisfaction.

The dataset also includes information about the number of lectures in each course, which can be used to compare the depth of content between different courses, especially when comparing paid versus free options. Additionally, the level of the course (e.g., Beginner, Intermediate, or All Levels) provides insight into the intended audience and difficulty of the course, which is crucial for understanding its appeal to different types of learners.

UNDERSTANDING & DEFINING DATASET

PROJECT PIPELINE

The project pipeline can be briefly summarized in the following steps:

- Data Understanding: Here, we need to load the data and understand the features present in it. This would help us choose the features that we will need for your final model.
- Exploratory data analytics (EDA): Normally, in this step, we need to perform univariate and bivariate analyses of the data, followed by feature transformations, if necessary. For the current data set, because Gaussian variables are used, we do not need to perform Z-scaling. However, you can check if there is any skewness in the data and try to mitigate it, as it might cause problems during the model-building phase.





THANK YOU