Graduation Project: Data Analysis of Udemy Courses

Project Overview:

This project aims to apply data analysis techniques to a real-world dataset (Udemy Courses) obtained from Kaggle. Students will explore trends, identify insights, and potentially make recommendations for instructors or the Udemy platform itself.

Learning Objectives:

- Data Cleaning and Wrangling: Handle missing values, identify outliers, and transform data for analysis.
- Exploratory Data Analysis (EDA): Uncover patterns, relationships between variables, and summarize key characteristics of the data.
- Data Visualization: Create clear and informative visualizations to communicate insights effectively.

Project Dataset:

You will use the Udemy Courses dataset available on Kaggle:
https://www.kaggle.com/datasets/jilkothari/finance-accounting-courses-udemy-13k-course

Project Steps:

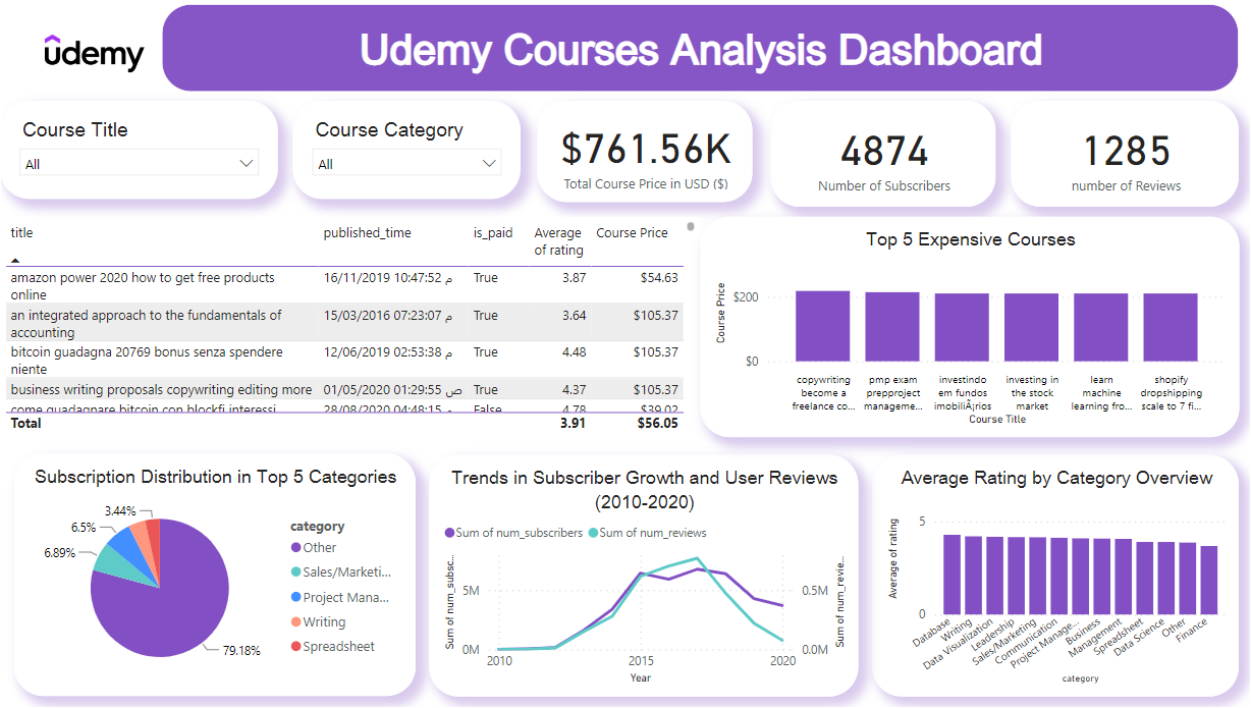
1. **Data Acquisition:**
    - Download the Udemy Courses dataset from Kaggle in a suitable format (e.g., CSV).

2. **Exploratory Data Analysis (EDA) & Data Cleaning (using chosen tools):**
    - Load the data using Python (Jupyter Notebook).
    - Identify and remove unnecessary columns.
    - Convert currency to USD for consistency in analysis.
    - Calculate discount percentage (if price and original price columns exist).
    - Impute missing values using appropriate techniques.
    - Categorize the courses and store the categories in a column.
    - Perform general data cleaning tasks (correct inconsistencies, typos, outliers).
        - Hint: points falling more than 2-3 standard deviations away from the mean are considered as outliers. Use box plot to confirm.

3. **Save the Cleaned Data in a MySQL Database using python libraries.**

4. **Data Analysis (using SQL Queries):**
    - Sort and identify top courses by:
        - Rating
        - Number of subscribers (if available)
        - Discount percentage (created during preprocessing)
    - Identify the most expensive courses.
    - Group the data by category and calculate:
        - Number of courses in each category
        - Average number of reviews per category
    - Analyze discounts:
        - Calculate average discount percentage offered per category.
    - Analyze price by paid/free status (if available):
        - Calculate the median price for courses in each category.

5. **Create a Tableau Dashboard the resembles the following Picture.**



Deliverables:

- A written report summarizing the data cleaning process, key findings from the analysis, and visualizations that effectively communicate insights.
- A presentation deck showcasing the project's findings and recommendations.
- Python code or scripts used for data analysis.
- Tableau dashboard for data visualization

Evaluation Criteria:

- Clarity and effectiveness of data cleaning procedures.
- Depth and accuracy of data analysis.
- Quality and informativeness of data visualizations.
- Cohesiveness and communication of findings in the report and presentation.