A

Project Report on

**Udemy Courses Analysis**

**(Core Module: 5)**

Submitted in partial fulfillment of completion of the course

Advanced Diploma in IT, Networking and Cloud Computing

Submitted by:

**HENA DHAWA**

**SUMIT KAYAL**

**RAHUL KUMAR**

Under Guidance of:

**Arpita Roy (Edunet)**

**Sayanti Manna (Edunet)**

|  |  |  |
| --- | --- | --- |
| IBM-Logo - Chicago Innovation | DGT introduces high end diploma courses - digitalLEARNING Magazine | Edunet Foundation-Delhi- CSR Organization profile |

Year 2023

**PROJECT OVERVIEW**

Abstract

Acknowledgement

Team Composition and Workload Division

Table of Contents

1. Introduction to Problem
2. Proposed Solution
3. Requirements

4.1 Technology Stack

4.2 Hardware

4.3 Software

4.4 Deployment Environment

5. User Requirements

6. Implementation Details

7. Testing

8. Deployment

9. Future Scope

10. Conclusion

Appendix A Screenshot of Project

References

**ABSTRACT**

Now a days there are many online learning platforms and Udemy is one of the most popular among them. Udemy is an online learning platform that offers a wide variety of courses on a multitude of topics. Udemy courses are available in over 65 languages and are taken by students from over 180 countries.

This project aims to conduct a comprehensive analysis of Udemy course data to gain insights into the trends, preferences, and effectiveness of online learning experiences. With the increasing popularity of online education platforms, understanding the dynamics of Udemy courses becomes crucial for both learners and instructors. The primary objectives include identifying popular course categories, evaluating user engagement patterns, and assessing factors influencing course ratings.

The methodology involves collecting and processing data from a diverse range of Udemy courses, considering variables such as course duration, instructor credentials, and learner demographics. Advanced statistical techniques will be employed to uncover correlations and trends within the dataset. Additionally, sentiment analysis will be applied to learner reviews to gauge overall satisfaction and identify areas for improvement.

The project's outcomes are expected to provide valuable insights for learners seeking courses, instructors aiming to optimize their content, and Udemy as a platform to enhance its offerings. The analysis may reveal patterns in successful courses, optimal course durations, and factors influencing learner satisfaction. Ultimately, this project aims to contribute to the ongoing discourse on effective online education and provide actionable recommendations for both learners and content creators.

**ACKNOWLEDGEMENT**

We would like to express our sincere gratitude to all those who have supported and contributed to the successful completion of this project. Your assistance and encouragement have been invaluable throughout this journey.

First and foremost, we want to thank my project supervisors, Arpita Roy(Edunet) and Sayanti Manna(Edunet), for their guidance, expertise, and unwavering support. Their insights and feedback have been instrumental in shaping the direction of this project.

We would also like to extend my appreciation to my fellow classmates who provided valuable input, shared resources, and engaged in stimulating discussions that enriched the project. Your collaborative spirit was a driving force behind our achievements.

Furthermore, we want to acknowledge our friends and family for their patience, understanding, and encouragement throughout this endeavour. Your support provided the motivation we needed to see this project through to its completion.

Last but not least, we are grateful to the entire faculty and staff of NSTI Howrah for providing a conducive learning environment and the necessary resources to undertake this project.

This project has been a rewarding learning experience, and we are thankful for the collective efforts of everyone involved. Your support has been instrumental in making this project a reality.

Thank you all for being a part of this journey.

Rahul Kumar, Sumit Kayal, Hena Dhawa

NSTI Howrah

14/11/2023

**TEAM COMPOSITION AND WORKLOAD DIVISION**

The team composition for the "Udemy Course Analysis" project can vary depending on the project's complexity, goals, and scope. This is the team composition:

**TEAM COMPOSITION**

1. Sumit Kayal (Leader):

* Responsible for overseeing the overall project coordination, planning, and communication Ensures that the project stays on schedule and meets objectives, Coordinates with other team members to gather requirements and documentation.

2. Hena Dhawa (Data Analysts):

* Responsible for data collection, cleaning, and preprocessing.
* Use Python libraries such as Pandas, NumPy for data manipulation and analysis.
* Creates visualizations using libraries like Matplotlib, Seaborn, or Plotly.
* Transforms data insights into visually appealing and informative charts and graphs.

3. Rahul Kumar(Testers/QA Team):

* Tests the code for bugs and ensures that the analysis results are accurate.
* Validates the consistency of the data and the reliability of the code.

**WORKLOAD DIVISION**

Detail how the workload is divided among the team members, highlighting their responsibilities and tasks:

1. Sumit Kayal: Project Planning and Conceptualization
2. Hena Dhawa: Data Analyzing and data visualization
3. Rahul Kumar: Testing and quality assurance

Documentation and Report Writing:

Lead the documentation process, with contributions from all team members for their respective areas of expertise.

**INTRODUCTION TO PROBLEM**

The primary objective is to explore and understand the Udemy dataset, which encompasses a diverse range of courses. Key problem areas and questions we aim to address include:

1. Course Popularity:

* What factors contribute to the popularity of a Udemy course?
* Can we identify trends in terms of the most popular course categories or topics?

2. User Engagement:

* How engaged are users with Udemy courses?
* Are there specific features or attributes of a course that correlate with higher user engagement?

3. Pricing Analysis:

* Is there a relationship between course pricing and enrolment rates?
* How do discounts or promotions affect user enrolment?

4. Instructor Influence:

* Do certain instructors have a higher success rate in terms of course completion or positive reviews?
* What characteristics of an instructor or their teaching style contribute to course success?

**PROPOSED SOLUTION**

The solution to the Udemy Course Analysis project involves a series of steps, utilizing Python and Jupyter Notebooks for data analysis and visualization.

1. Data Collection:

Obtain Udemy course dataset.

2. Data Cleaning and Pre-processing:

Clean, handle missing values, and standardize data.

3. Exploratory Data Analysis (EDA):

Visualize and analyse key dataset characteristics.

4. Course Popularity Analysis:

Identify factors influencing course popularity.

5. User Engagement Analysis:

Explore user engagement factors and patterns.

6. Pricing Analysis:

Analyse pricing impact on enrolment.

7. Machine Learning:

Apply ML models for predictive analysis.

**REQUIREMENTS**

**TECHNOLOGY STACK**

Programming Language:

1. Python:

* Core language for data analysis, manipulation, and visualization.

Data Analysis and Visualization:

1. Jupyter Notebooks:

* Interactive and collaborative environment for data analysis.

Libraries and Frameworks:

1. Pandas:

* Data manipulation and analysis.

1. NumPy:

* Numerical operations on data.

1. Matplotlib, Seaborn:

* Data visualization.

1. Scikit learn, TensorFlow:

* Machine learning models.

**HARDWARE**

* CPU and RAM
* Storage: Adequate storage space for the database, application code, and uploaded files.
* Network Infrastructure: Reliable internet connectivity, firewalls, and security measures to protect the system.

**SOFTWARE**

* Operating System
* Python Environment
* Jupyter Notebook
* Web Server

**DEPLOYMENT ENVIRONMENT**

1. Local Development:

* Team members can initially work on their local machines.

**USER REQUIREMENTS**

* Mobile Accessibility or PCs
* Internet Access
* Supported Browser
* require specific software tools or applications depending on the content. If you are taking a programming course, you might need a specific integrated development environment (IDE).
* Payment Method: While Udemy offers many free courses, some courses may require payment. Users need a valid payment method (credit card, PayPal, etc.) to enrol in paid courses.

**IMPLEMENTATION DETAILS**

The implementation of the Udemy Course Analysis project involves several steps, including data loading, cleaning, analysis, and visualization.

**STEP1**- Firstly, I have to download csv file from Kaggle (<https://kaggle.com/datasets/andrewmvd/udemy-courses> ).

**STEP 2-** Secondly, I have to exact file.

**STEP 3-** I have to open JUPYTER NOTEBOOK.

**STEP 4 -** I have to import all necessary libraries like NumPy, pandas, math plot, seaborn etc.

**STEP 5 -** Then, I have to read the whole csv file by this command [df = pd. read\_csv (udemy\_courses.csv')

print(df)]

**STEP 6 -** Then I have Explore the data to get a sense of its structure and content.

#to find the rows and columns of the dataframe.

print(df.shape)

#### #To find what are the column index of the dataframe

print(df.columns)

#### #to find the datatype of the CSV file

print(df.dtypes)

#### # To find the information about the file like storage, datatypes, class etc.

df.info()

#### #to see the last 10 data from the dataframe

print(df.tail(n=10))

#### # it take all value present in csv file method returns a DataFrame object where all the values are replaced with a Boolean value True for NULL values, and otherwise False

df.isnull().values.any()

**STEP 7 -** I have to show different charts and graph for understanding that in my csv file the necessary queries of the students and faculties about the course or subject.

**STEP 8 -** Correlation matrix heat map. So, we have to check the courses relation with other columns. It shows the relation with all.

**STEP 9 -** I have to update and improve platform based on user because day by day the demand of courses is increasing.

**STEP 10 –** I have also created a dashboard so that we can check the different changes according to the subject present.

**TESTING**

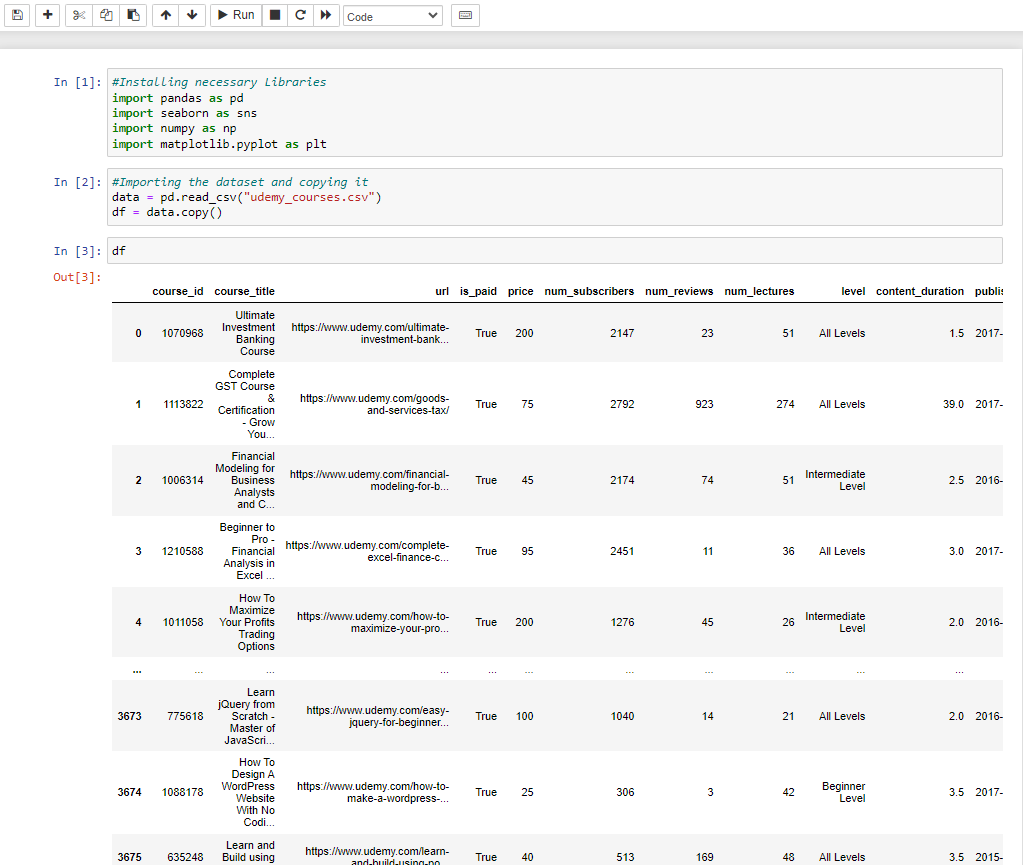


Image1: testing if all the data are imported.

**DEPLOYMENT**

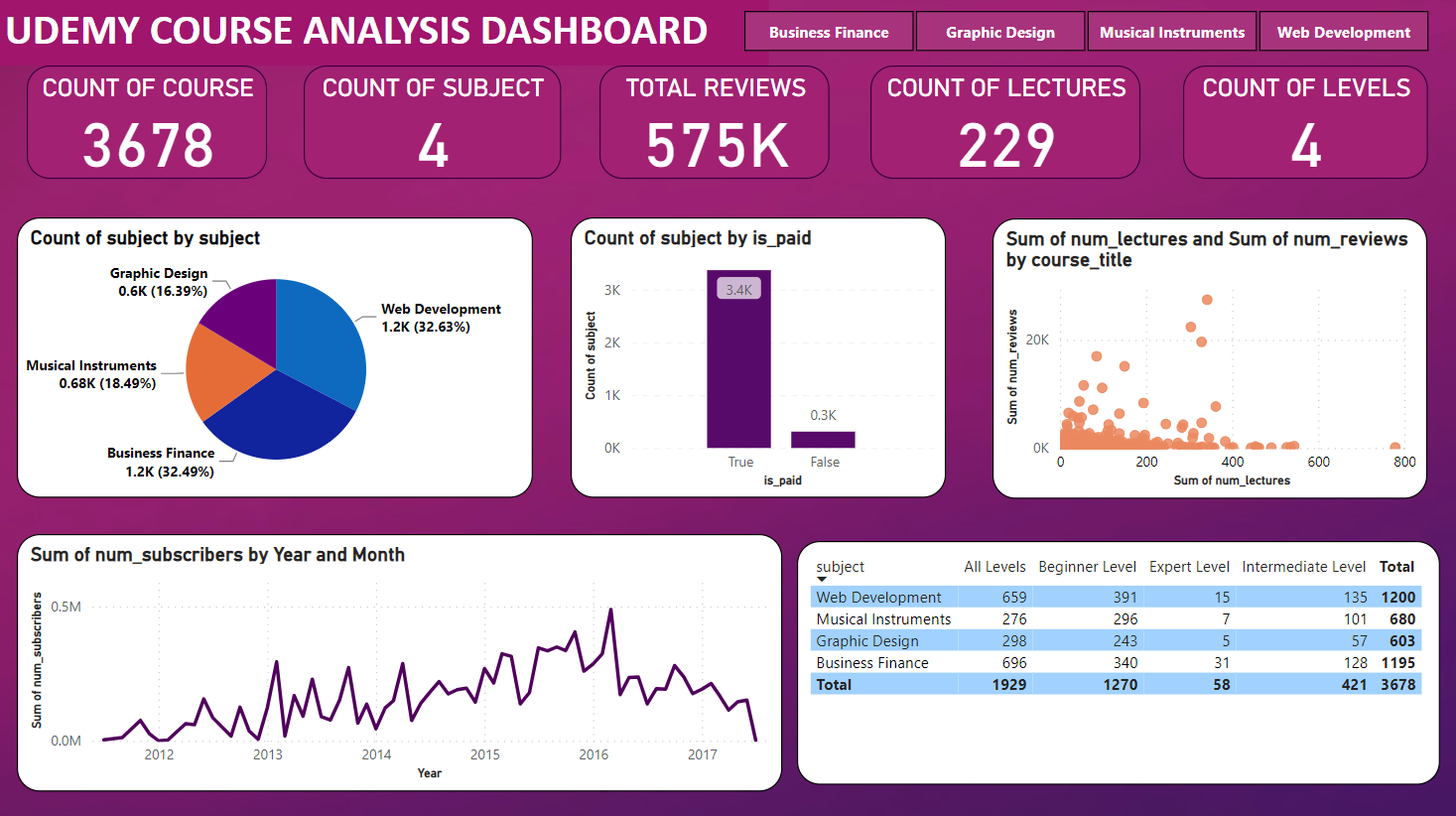


Image2: Dashboard of Udemy course analysis.

**FUTURE SCOPE**

The Udemy Course Analysis project provides a foundation for ongoing development and expansion. Here are potential future directions and enhancements:

1. Real Time Data Integration:

Explore ways to integrate real time data feeds to provide up to the minute insights into course trends and user behaviors.

2. Enhanced Machine Learning Models:

Improve and expand machine learning models to predict course success, user engagement, or recommend personalized courses based on user preferences.

3. User Feedback and Ratings Analysis:

Implement sentiment analysis on user reviews to gain a deeper understanding of user satisfaction and identify areas for improvement in courses.

4. Dynamic Visualizations:

Create interactive and dynamic visualizations that allow users to explore data dynamically, zoom in on specific time periods, or filter based on various criteria.

5. Predictive Analytics:

Implement advanced predictive analytics to forecast future trends in the online education landscape, helping course creators and platforms stay ahead of the curve.

9. Multi-Platform Support:

Extend the project to support multiple online learning platforms, allowing for a broader analysis of the online education ecosystem.

The future scope of the project is expansive and can be tailored based on emerging technologies, user feedback, and the evolving landscape of online education. Continuous improvement and adaptation to new trends will be crucial for keeping the project relevant and valuable.

**CONCLUSION**

The Udemy Course Analysis project has successfully provided valuable insights into the dynamics of the online learning platform. Through a comprehensive approach to data collection, cleaning, analysis, and visualization, the team has uncovered trends and patterns that offer meaningful implications for various stakeholders, including course creators, instructors, and Udemy itself.

Key Findings:

1. Course Popularity Factors:

Identified key factors influencing the popularity of Udemy courses, including the number of reviews, instructor experience, and pricing.

2. User Engagement Patterns:

Explored user engagement patterns, revealing correlations between course categories and user reviews.

3. Pricing Impact:

Analyzed the impact of pricing on enrollment rates, providing insights into optimal pricing strategies for course creators.

**APPENDIX A**

**SCREENSHOT OF PROJECT**

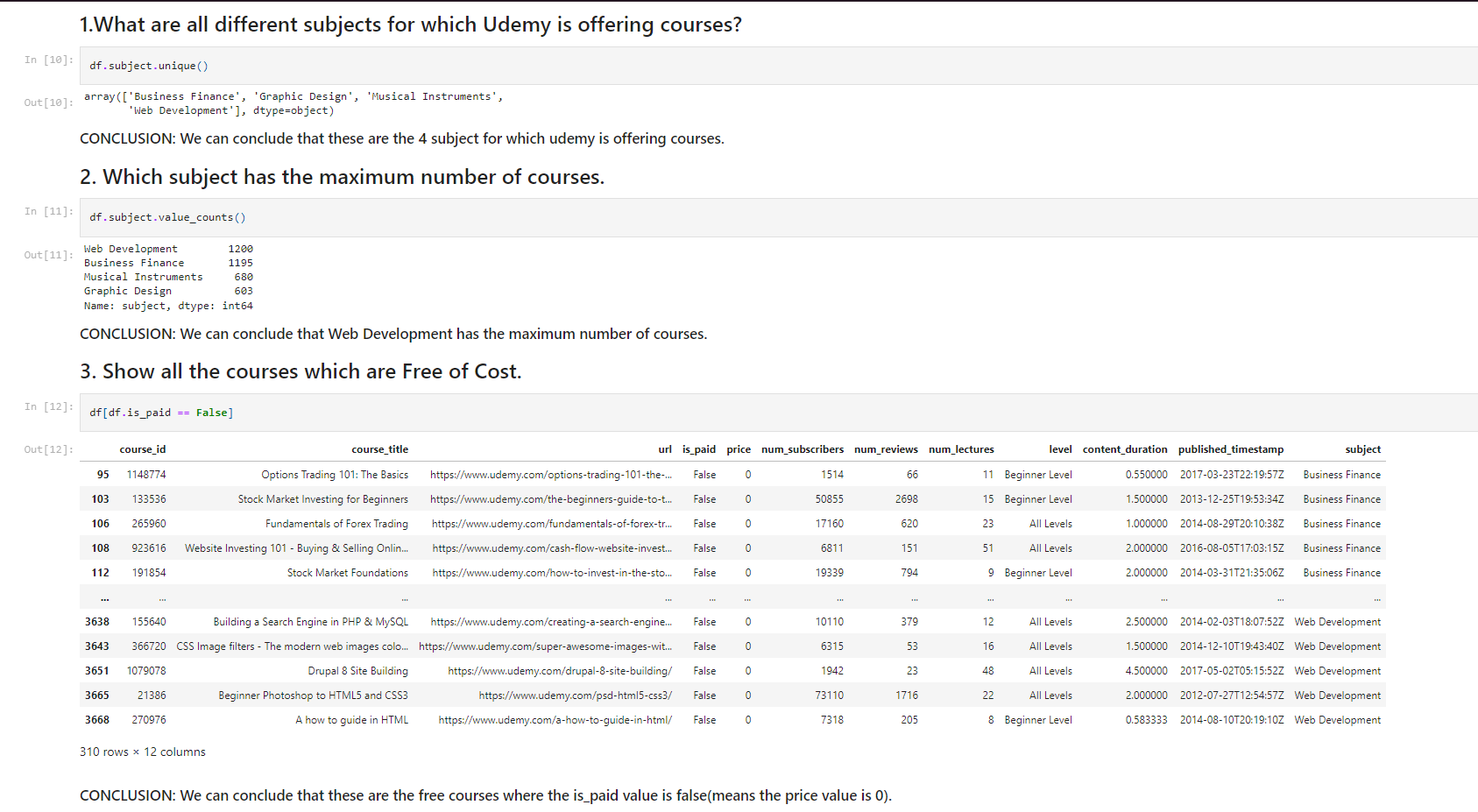


Image3: Extracting different data from the uploaded DataFrame.

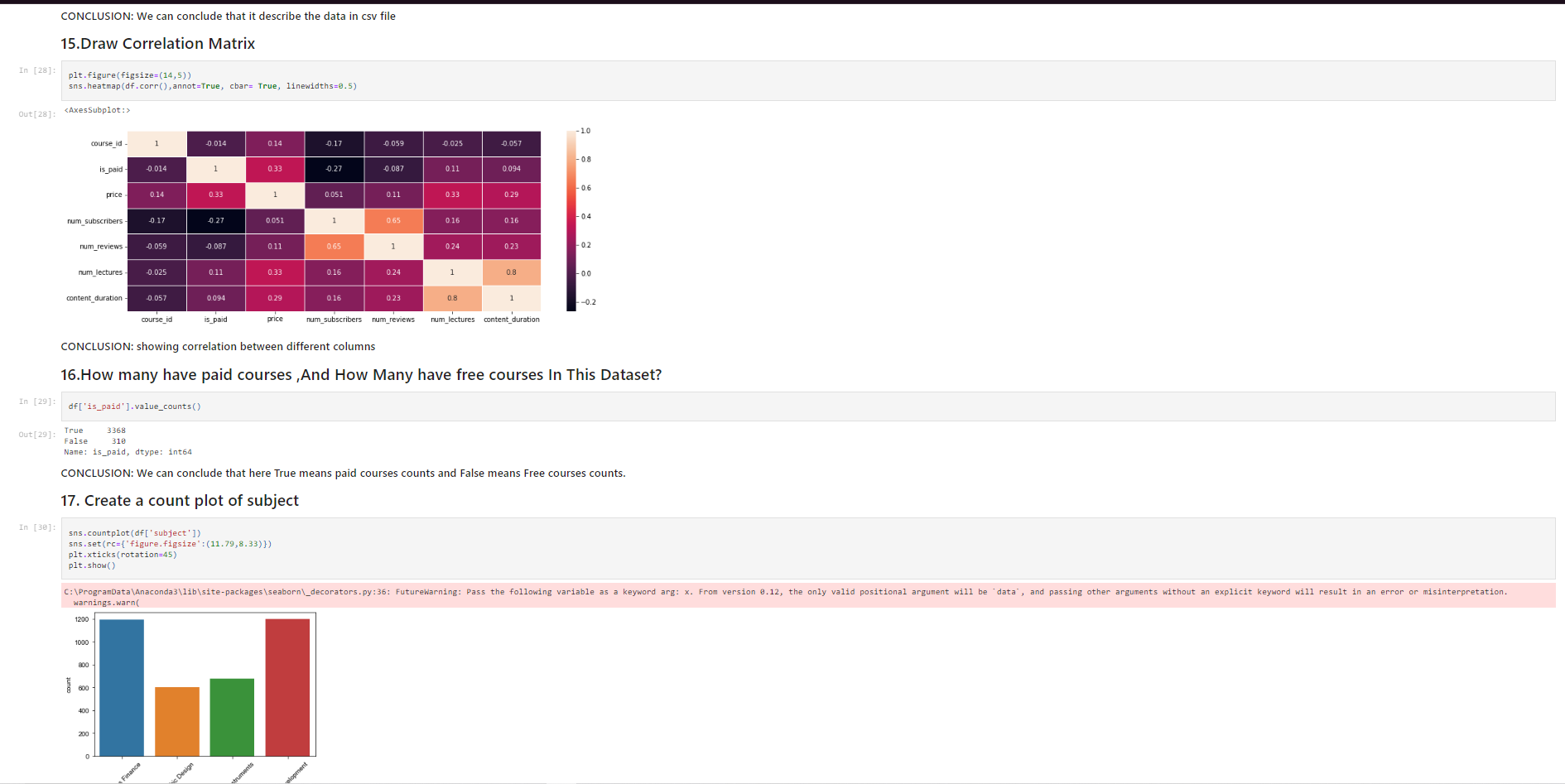


Image4: creating different plots and correlation matrix.

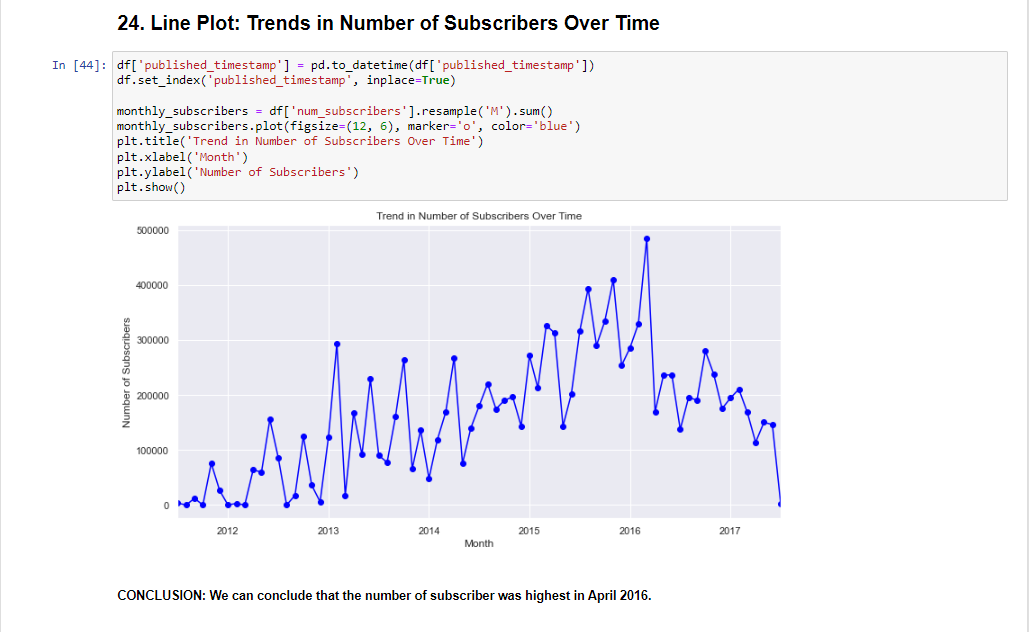


Image5: Line Plot: Trends in number of subscribers over time.

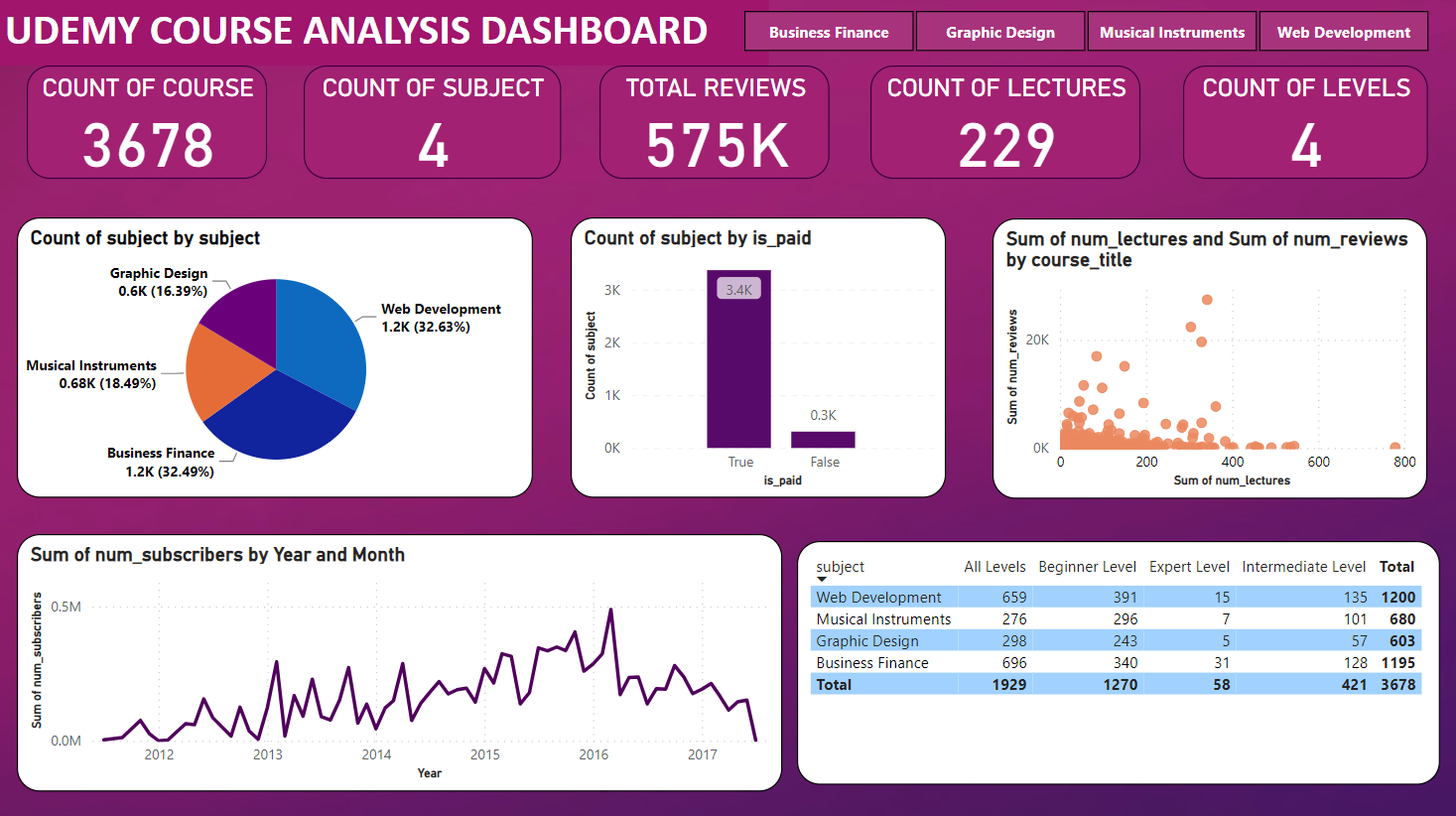


Image6: Dashboard of Udemy course analysis.

**REFERENCES**

[1] <https://www.youtube.com/watch?v=AO5uhxa1R84>

[2] <https://www.w3schools.com/python/numpy/numpy_intro.asp>

[3] <https://seaborn.pydata.org/>