**The Udemy Project**

INSY 661 - Database and Distributed Systems

# Business Overview

***1a. An Overview of Online Learning Platforms***

Major online learning platforms like Udemy connect millions of students with expert instructors through large catalogs of on-demand video courses. These platforms provide a global solution for skills development through online coursework.

The global online learning platform industry has experienced tremendous growth in recent years. The Online Education Market Size was valued at USD 38.5 billion in 2021 and it is projected to grow to USD 197.79 billion by 2030, indicating strong ongoing demand. (Source: Market Research Update, 2022, <https://www.marketresearchfuture.com/reports/online-education-market-3957>). During the pandemic in 2020, enrollments within the United States alone increased substantially, reaching over 30 million students – a 19% increase compared to 2019. (Source: Market Research Update, 2022, <https://www.marketresearchfuture.com/reports/online-education-market-3957>).

Typically, these platforms encourage users to explore their diverse course catalogs spanning dozens of topic areas. Courses are organized systematically by categories, language, skill levels, price, and other filters to allow easy browsing. Instructors range from individual experts to faculty at top universities. Detailed profiles provide insight into teaching styles, content, and student reviews.

Central to the online learning experience is the video courses. Through lecture videos, quizzes, and interactive components, courses deliver immersive and practical learning experiences. Students have unlimited access to self-paced materials on their own schedules. For flexibility, courses can also be downloaded for offline mobile viewing.

***1b. Specific Examples of Online Learning Platforms***

To understand the online learning platforms better, Table 1 (seen below) provides a high-level overview of four prominent platforms in the online education sector: Udemy, Coursera, edX, and Codecademy.

**Table 1. High-Level Overview of Four Online Learning Platforms**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Description** | **Founded** | **Vocation** | **Learners** |
| **Udemy** | An open marketplace through which anyone can create and take courses. It offers a large variety of topics from technology skills to entertainment. | February 2010 by Eren Bali, Oktay Caglar, Gagan Biyani | For-profit | 30 million |
| **Coursera** | A platform that offers university-style courses provided by universities or organizations world-wide. | April 2012 by Daphne Koller, Andrew Ng | For-profit | 40 million |
| **edX** | A platform that offers university-style courses and provides traditional instruction education. | May 2012 by Anant Agarwal, Chris Terman, Piotr Mitros | Non-profit | 18 million |
| **Codecademy** | An educational company that teaches coding in different programming languages. | August 2011 by Zach Sims, Ryan Bubinski | Non-profit | 45 million |

(Source: Cornejo-Velazquez, Eduardo et al. 2020, <https://eric.ed.gov/?id=EJ1245290>)

**Udemy**, founded in 2010, has established itself as the largest online learning marketplace in terms of scale. With over 30 million registered learners to date, Udemy operates a highly accessible and open marketplace model. This allows anyone to create and teach courses on a wide variety of topics spanning business, technology, creative skills, and more. By focusing on an open ecosystem and vast course catalog, Udemy has been tremendously successful at attracting students and instructors worldwide. On the other hand, **Coursera**, launched two years later in 2012, takes a more curated approach centered around university-level coursework. Partnering with top academic institutions from around the globe, Coursera provides learners access to a library of over 5,000 individual courses and full degree programs. With its emphasis on high-quality, credentials-focused content from renowned universities, Coursera has grown to become the second-largest platform with 40 million users enrolled to date. **edX**, also founded in 2012, shares Coursera's university-partnered model but operates as a non-profit organization. While smaller in size currently at 18 million users, edX collaborates with elite institutions like MIT, Harvard, and UC Berkeley to offer open online courses and master's degree programs. Its non-profit structure allows edX to focus on advancing accessible online learning through technology. The youngest platform featured is **Codecademy**, which launched in 2011, centered specifically around teaching various coding languages and technologies. Taking a niche, skills-based approach, Codecademy stands out for rapidly achieving over 45 million learners through its engaging, interactive course formats focused solely on programming instruction.

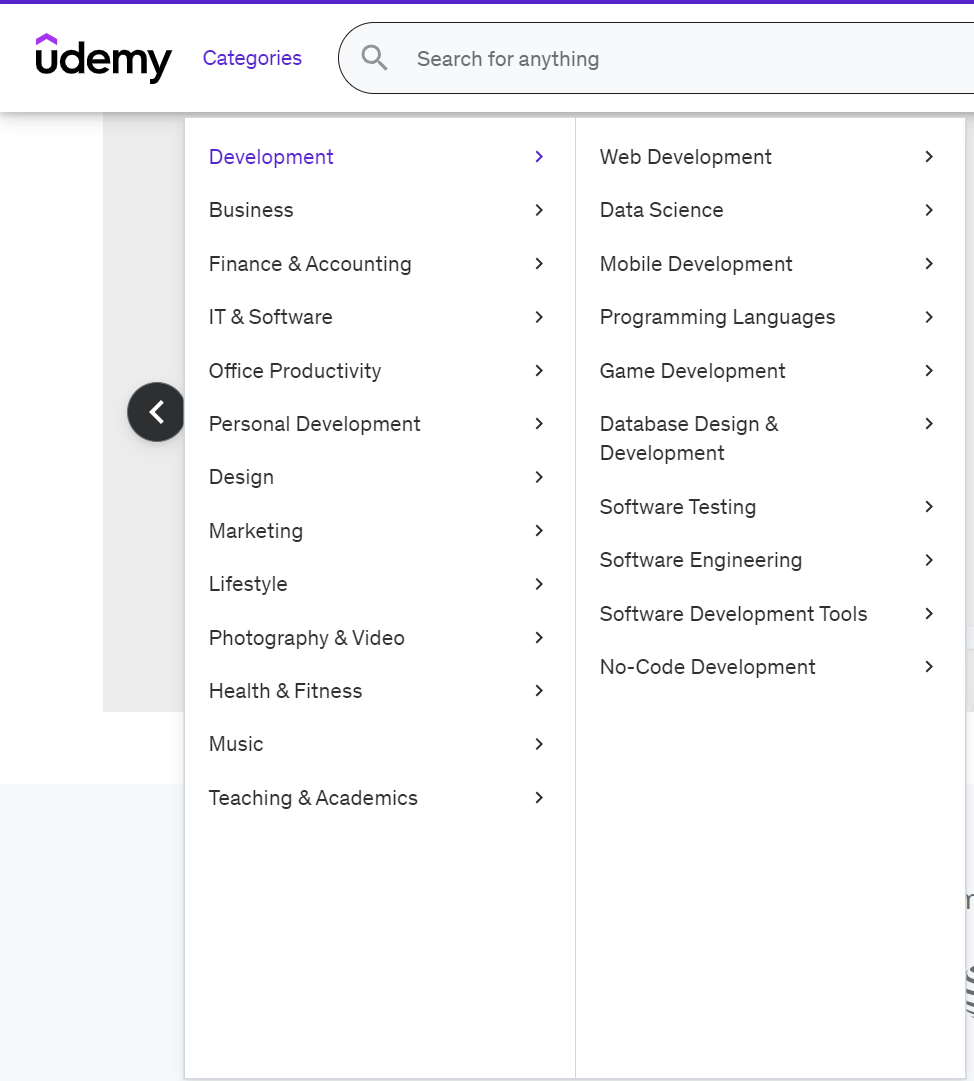
Overall, this table highlights the growth and diversity that exists across four leading online learning platforms, whether focused on open markets, university partnerships, non-profit status, or specific skill areas like coding. In summary, it is evident that each has capitalized on the boom in digital education through distinct models attuned to learners' varied needs and interests.

***1c. Our Project***

For the purposes of this database project, we have chosen **Udemy** as our main point of reference for the online learning platform. As one of the largest online learning marketplaces in terms of scale and number of learners, Udemy has established itself as a leader in the industry through its highly accessible and open marketplace model. This allows for a vast catalog of courses spanning a wide variety of topics to be created and taught by anyone. Through its focus on an open ecosystem approach, Udemy has been tremendously successful in attracting both students and instructors globally. Given Udemy's prominence and leadership position supported by its open platform and rich course availability, it provides an excellent case study for us to model our database design and objectives.

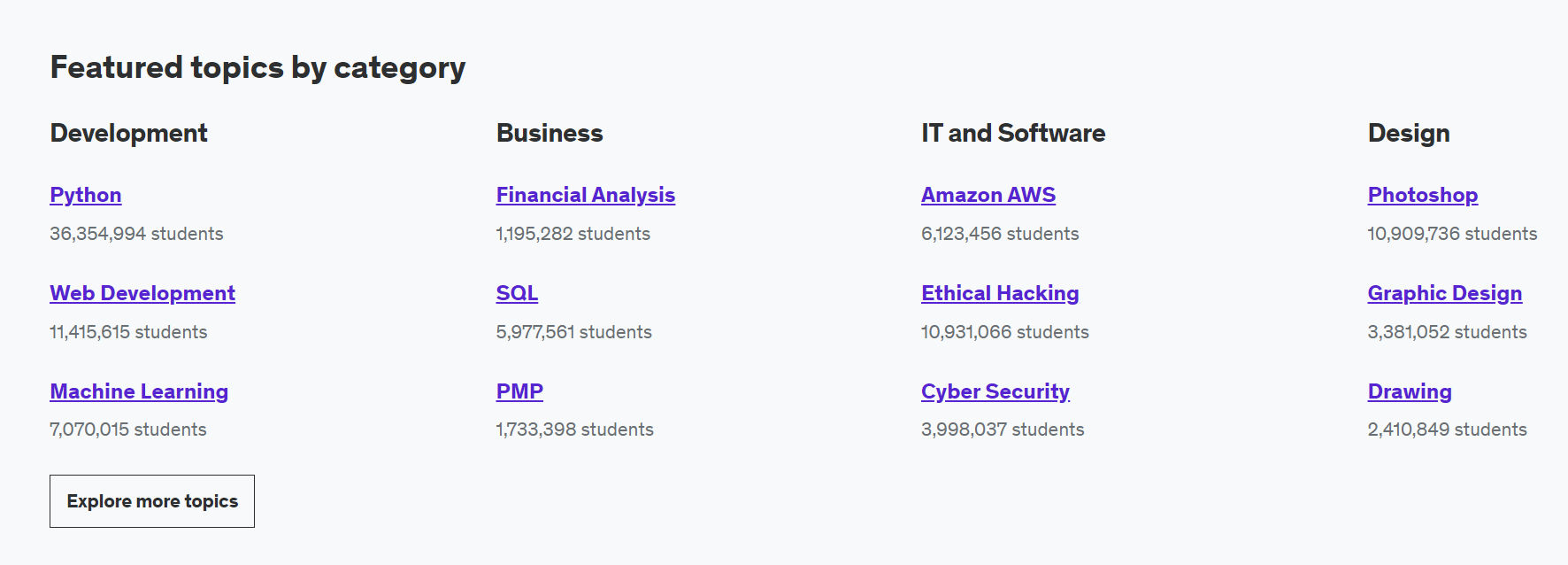
Here are some sample figures illustrating the functionalities of the Udemy website, which we used as our point of reference.

**Figure 1. List of Categories and the Search Bar on the Udemy Website**



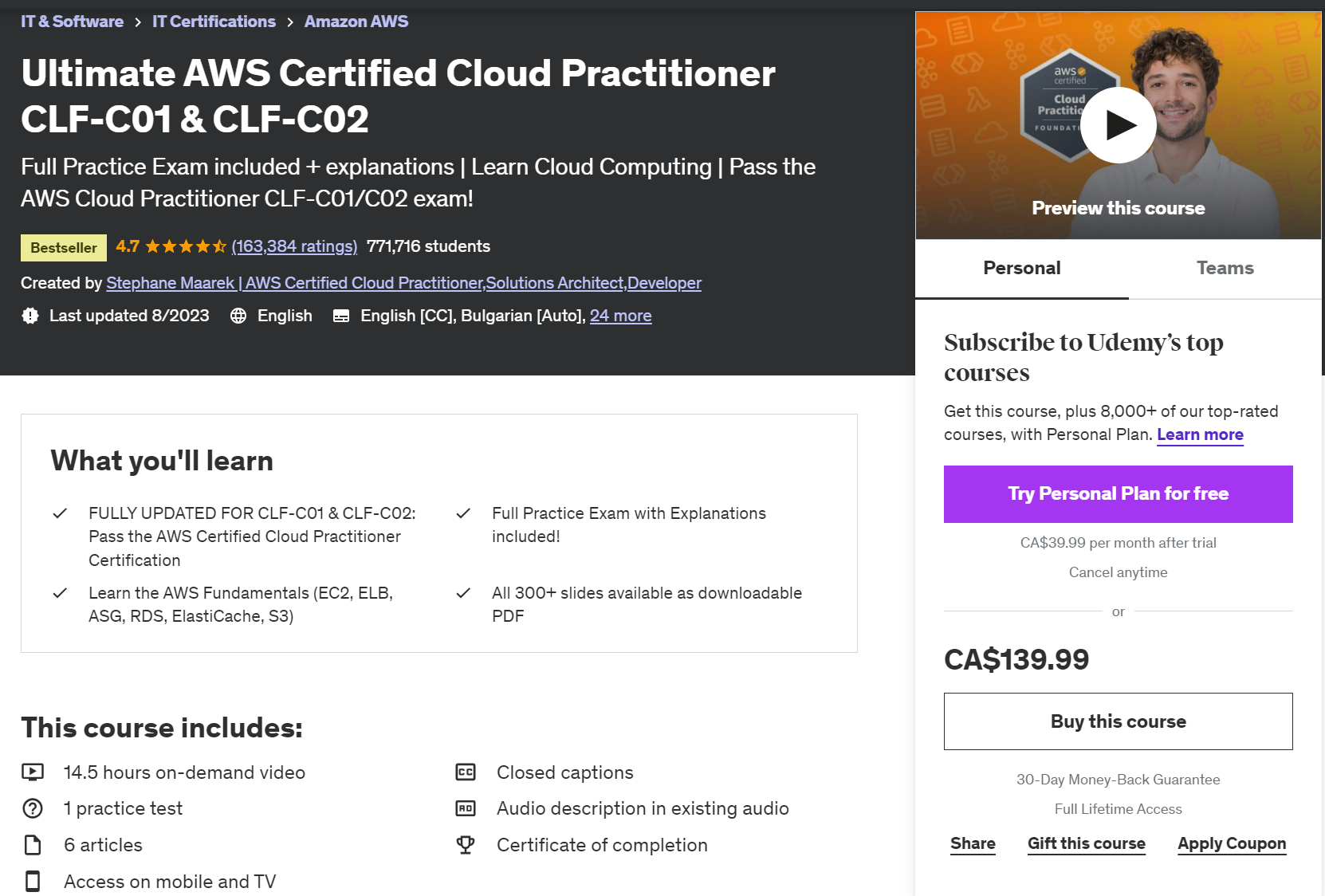
(Source: <https://www.udemy.com/>)

**Figure 2. Featured topics by category on the Udemy website**



(Source: <https://www.udemy.com/>)

**Figure 3. Sample Course Information on the Udemy Website**



(Source: <https://www.udemy.com/course/aws-certified-cloud-practitioner-new/>)

Our online learning platform aims to provide a comprehensive and engaging education experience for students and instructors worldwide. Central to our mission is maintaining flexible datasets on key entities like courses, learners, categories, and instructors. This integrated data infrastructure supports objectives around customized searching, personalized recommendations, and dynamic listings organized by criteria.

Additionally, we seek to generate actionable insights through reporting on metrics such as course popularity trends, correlated relationships between user attributes and choices, and optimal pricing strategies. Performance tracking covers indicators such as enrollments, completion rates, progress over time, and sharing effectiveness.

Our dynamic web interfaces aim to simplify user experiences through targeted homepages. Community remains a priority through functions like student reviews aiding discovery.

In general, our goals center on continuous optimization. Regular evaluation of indicators will afford data-driven refinements to content, promotions, programs, and platform elements. In this way, we strive to maintain engagement, maximize outcomes, and address the increasing global demand for online learning opportunities. For more information, below are our mission statement and mission objectives.

# Mission Statement

The purpose of our database system is to maintain data that supports the learning and teaching experience for all users and provides analytics to optimize business operations.

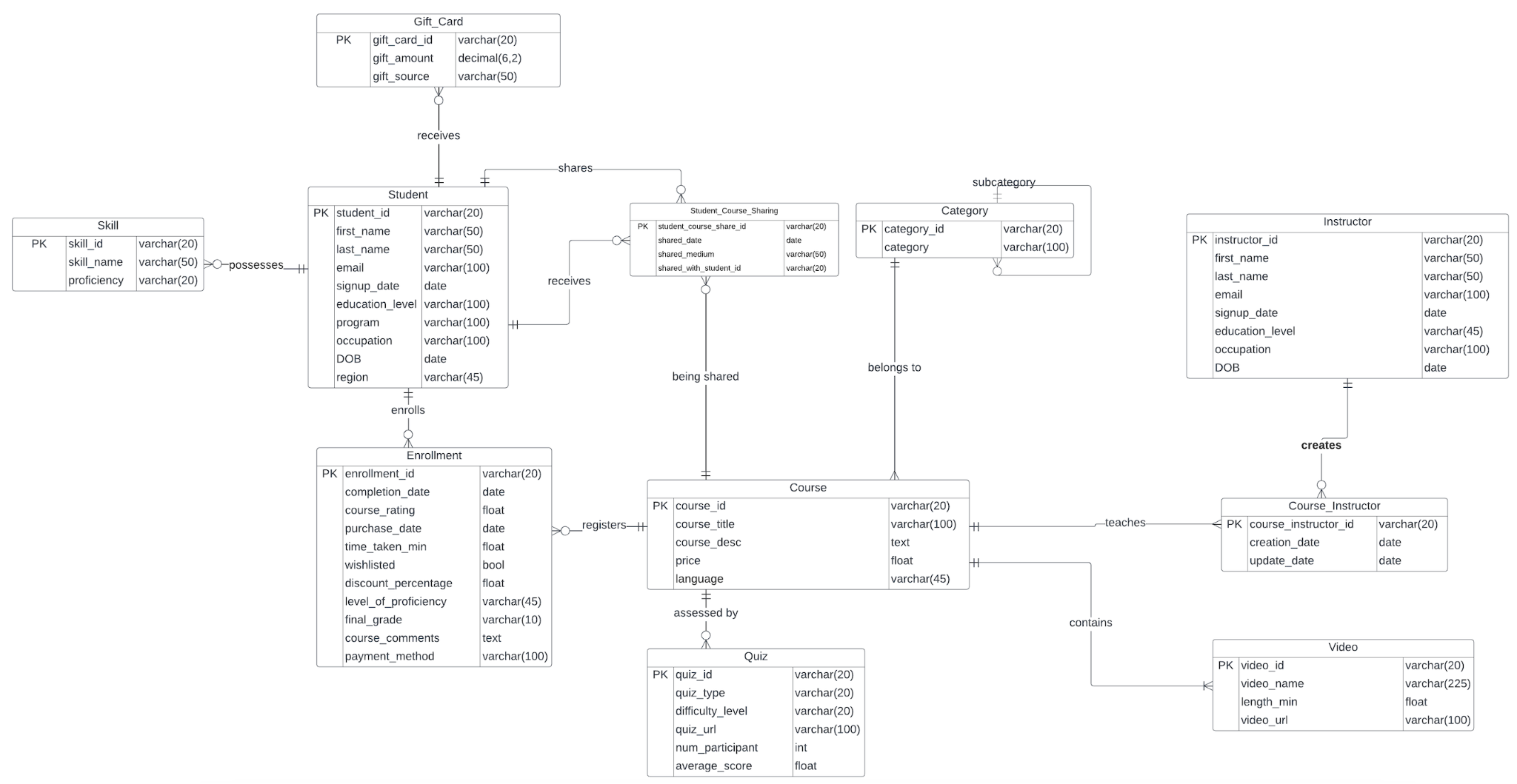
# Mission Objectives

Our mission objectives are the following:

1. To maintain (enter, update, and delete) data on courses, students, categories, instructors, enrollments, videos, quizzes, gift cards, and skills.
2. To perform targeted searches on courses, instructors, and categories based on keywords and filters.
3. To report on course popularity metrics like bestsellers and trending courses.
4. To create targeted content for people with similar demographics.
5. To create category pages with dynamic course listings based on diverse parameters.
6. To create dynamic webpages for instructors with their most essential information.
7. To create a dynamic dashboard for students to see their progress.
8. To track student retention rate.
9. To report on student insights like correlations between education/occupation and course choices.
10. To report on the price sensitivity and its relationship to enrollment and completion rates among different course price points and student age groups.
11. To track the conversion of free to paid users.
12. To track sharing activity and impact on enrollments across different mediums.
13. To track changing enrollment patterns and rates throughout the year to inform the planning of promotional events and catalog updates.
14. To report on educational outcomes like completion rates versus video lengths.
15. To report on retention and churn trends over time by customer segment to pinpoint at-risk cohorts for targeted retention programs.
16. To track the conversion rate of students who were recommended a class into actual enrollments to evaluate the effectiveness of recommendation strategies.
17. To report on the top bestselling paid course(s) within each category based on total enrollment rankings, displaying all courses that share the number one rank if total enrollments are equal
18. To report on customer segments and purchasing behavior to prioritize retention programs.
19. To track course enrollments, completions, and average time taken.
20. To report on instructor diversity metrics like the number of regions and categories taught.

# ERD

**Figure 4 – Entity-Relationship Diagram**



**Data Dictionaries**

**Description of Entities**

**Table 2 – description of entities**

|  |  |  |  |
| --- | --- | --- | --- |
| entity name | description | aliases | occurrence |
| course | entity that represents the online courses that are uploaded on the platform | course | one course can contain up to multiple quizzes and videos, can be taught by multiple instructors and belongs to only one category. |
| student | entity that represents each student that registers on the platform | student | one student can register for up to multiple classes, can share up to multiple classes and can buy more than one gift card |
| category | entity that contains a descriptive of each category | category | one category can have multiple subcategories. each course is unique under a category. for instance, the sql class is only under the analytics category, and nowhere else |
| instructor | entity that represents the instructor for each course | instructor | one instructor can teach multiple courses, and one course can be taught by multiple instructors. |
| video | entity that represents the video linked to the uploaded course | video | one course can contain multiple videos |
| quiz | entity that represents the quiz linked to the uploaded course | quiz | one course can contain multiple quizzes |
| gift card | entity that represents gift cards applicable to a student's account balance | gift\_card | one student can buy multiple gift cards |
| sharing | entity that contains the information relative to course sharing among students | student\_course\_sharing | one student can share multiple courses with many other students |
| enrollment | entity that contains the enrollment information of each student for each course | enrollment | one student can enroll for multiple courses |
| course instructor | entity that links the instructor to the course | course\_instructor | one instructor can teach multiple courses, and one course can be taught by multiple instructors. |
| skill | entity that represents the skills linked to the a student | skill | One student can have zero to many skills |

# Description of Attributes

**Table 3 – description of attributes**

|  |  |  |  |
| --- | --- | --- | --- |
| entity name | attributes | description | data type |
| course | course\_id course\_title course\_desc price language | course identifier course title course description price of the course language of the course | varchar(20) varchar(100) text float varchar(45) |
| student | student\_id first\_name last\_name email signup\_date education\_level program occupation dob region | student identifier student's first name student's last name student's email student's signup date highest education level field of studies job date of birth region of the world | varchar(20) varchar(50) varchar(50) varchar(100) date varchar(100) varchar(100) varchar(100) date varchar(45) |
| category | category\_id category | category identifier name of the category | varchar(20) varchar(100) |
| instructor | instructor\_id first\_name last\_name email signup\_date education\_level occupation dob | instructor identifier instructor's first name instructor's last name instructor's email instructor's signup date instructor's education instructor's job instructor's date of birth | varchar(20) varchar(50) varchar(50) varchar(100) date varchar(45) varchar(100) date |
| video | video\_id video\_name length\_min video\_url | video identifier video's title video's length url of the video | varchar(20) varchar(255) float varchar(100) |
| quiz | quiz\_id quiz\_type difficulty\_level quiz\_url | quiz identifier quiz type difficulty level url of the quiz | varchar(20) varchar(20) varchar(20) varchar(100) |
| gift\_card | gift\_card\_id gift\_amount gift\_source | identifier for the gift card amount of the gift source of the gift | varchar(20) decimal(6,2) varchar(50) |
| student\_course\_sharing | student\_course\_share\_id shared\_date shared\_medium | identifier of the share date the course has been shared sharing medium | varchar(20) date varchar(50) |
| enrollment | enrollment\_id completion\_date course\_rating purchase\_date time\_taken\_min wishlisted discount\_percent prior\_proficiency final\_grade  course\_comments  payment\_method | enrollment identifier date the course has been completed rating date the course was purchased time taken to complete the course was the course wishlisted? discount percentage prior proficiency in said course obtained final grade | varchar(20) date float date float bool float varchar(45) varchar(10)  text  varchar |
| course\_instructor | course\_instructor\_id creation\_date update\_date  shared\_with\_student\_id | identifier linking course and instructor date a course has been created date a course has been updated  The student which receives the shared course | varchar(20) date date  varchar(20) |
| skill | Skill\_id  Skill\_name  proficiency |  | varchar(20)  varchar(50)  varchar(20) |

# 

# Relational/Logical Model

**Category** (category\_id, category)

Primary Key: category\_id

**Course** (course\_id, category\_id, course\_title, course\_desc, price, language)

Primary Key: course\_id

Foreign Key: category\_id references Category(category\_id)

**Video** (video\_id, course\_id, video\_name, length\_min, video\_url)

Primary Key: video\_id

Foreign Key: course\_id references Course(course\_id)

**Quiz** (quiz\_id, course\_id, quiz\_type, difficulty\_level, quiz\_url)

Primary Key: quiz\_id

Foreign Key: course\_id references Course(course\_id)

**Instructor** (instructor\_id, first\_name, last\_name, email, signup\_date, education\_level, occupation, dob)

Primary Key: instructor\_id

**Course\_Instructor** (course\_instructor\_id, course\_id, instructor\_id, creation\_date, update\_date)

Primary Key: course\_instructor\_id

Foreign Keys: course\_id references Course(course\_id), instructor\_id references Instructor(instructor\_id)

**Student** (student\_id, first\_name, last\_name, email, signup\_date, education\_level, program, occupation, dob, region)

Primary Key: student\_id

**Enrollment** (enrollment\_id, student\_id, course\_id, completion\_date, course\_rating, purchase\_date, time\_taken\_min, wishlisted, discount\_percent, prior\_proficiency, final\_grade, course\_comments, payment\_method)

Primary Key: enrollment\_id

Foreign Keys: student\_id references Student(student\_id), course\_id references Course(course\_id)

**Student\_Course\_Sharing** (student\_course\_share\_id, student\_id, course\_id, shared\_date, shared\_medium, shared\_with\_student\_id)

Primary Key: student\_course\_share\_id

Foreign Keys: student\_id references Student(student\_id)

Shared\_with\_student\_id references Student(student\_id)

course\_id references Course(course\_id)

**Gift\_Card** (gift\_card\_id, student\_id, gift\_amount, gift\_source)

Primary Key: gift\_card\_id

Foreign Key: student\_id references Student(student\_id)

**Skill** (skill\_id, student\_id, skill\_name, proficiency)

Primary Key: skill\_id

Foreign Key: student\_id references Student(student\_id)

**Figure 5 – Logical Model**

