365 Data Science Specification Report

Piyush Dhawad
MSc in Data Analytics
National College of Ireland
Dublin, Ireland
x21236933@student.ncirl.ie

Sarthak Bhatnagar
MSc in Data Analytics
National College of Ireland
Dublin, Ireland
x21185352@student.ncirl.ie

Tejali Gangane
MSc in Data Analytics
National College of Ireland
Dublin, Ireland
x21148872@student.ncirl.ie

Usama Hanif
MSc in Data Analytics
National College of Ireland
Dublin, Ireland
x22108831@student.ncirl.ie

***Abstract*—In this project, we will analyze an online learning platform and will provide an effective report that will significantly enhance the platform's performance. From the knowledge gained from this report, admins can use different techniques and strategies to understand the learning behaviour of the student and to deal with the competition in the market. During implementation, we design the system's flow process and get the dataset from 365 Data Science. We use PowerBI to produce the visualizations of our data and build an effective report and dashboard. As business analysts, our main aim is to identify and analyse the factors that contribute towards the growth of the business. To find out the popular courses which could be measured through watch time and the number of ratings provided, and how we can increase the purchase of online courses from 365 data science platform so that the company can succeed in reaching its goal.**

***Index Terms*—Dashboard, Power BI, Platform, Data Analytics, Reporting**

I. Background Information

A group of friends turned to start-up enthusiasts to begin their journey in 2017 with the mission to build a platform that provides the world's most flexible data science training resources online. 365 Data Science is an online teaching company which primarily focuses on the training of students pursuing Data Science. It provides high-quality data science courses which will help students (who are soon to be data scientists) in building market-ready skills. Their courses are designed in such a way that helps anyone new to this domain, in building the skills; set from beginner level to professional level regardless of their background education and experience. As the demand for the Data related fields has been growing tremendously 365 Data Science aids students and working professionals to begin their careers as data analysts, data scientists and business analysts. As traditional degrees are expensive, time-consuming and have tight schedules, the 365 Data Science mission is to close the gap by providing flexible high quality and affordable training that a student can take whenever they want. The mission of 365 Data Science is to motivate individuals all over the world to grab new opportunities and meet market needs. This will not only help students to increase the list of attributes in their skill set but will also help them practice skills that were learnt in the course, by implementing everything on real-world datasets.

The 365 Data Science shared a vision for success. They have a top-tier team that is committed to helping everyone in reaching their full potential. They believe that excellence in every lesson of the course should be the top priority.

II. Marketplace

The 365 Data Science operates in an Online Education Marketplace. As online education has become very popular it is attracting a large number of teachers and learners across the globe. Typically, the online education marketplace allows the user to become an expert by completing online courses and acquiring the certificates which they can show in their CVs. The main goal of online courses is to provide a platform to users through which they can communicate with the experts and get knowledge. The main mission of the online education marketplace is to improve the lives of the users through learning and encourage everyone to utilize their knowledge and expertise to create and share education and to enrol themselves to acquire knowledge from the courses.

After the Covid-19 pandemic, there has been a significant increase in the online education marketplace. As there was a lockdown and the physical classes were cancelled, all the education systems were converted to online. The trend of online education increased, and many companies have started to provide online courses. The 365 Data science provides high-quality online courses which enable students to learn from anywhere and gain market-ready skills.

This gives students enough time to keep learning along with other activities they might have. Traditional learning methods where teaching/training was performed in classes led to additional time investment on the side of teachers/students and other parties concerning the time taken to travel to the location. Ever since the introduction of online learning, people have preferred more remote learning options to in-person learning.

Academic leaders up to 77 per cent are said to believe that online classes tend to have the upper hand over in-person classes.

III. Related Work

Research on students' behaviour and performance evaluation in online learning has produced mixed results.

Xiangmin Zhang in his research [1] investigated how the final grades of students can be measured according to their behaviour and learning performance. He studied the four online classes of information technology in the library of the US. He analyzed the final grades of students at both general and class levels using data gathered from all the classes. The results of his research show that if the student had no access to the lectures and educational material weekly their final grades were badly impacted. On the other hand, in one class it was seen that if the frequency of accessing the lectures was high their final grades turned out to be very good.

Anna Sun and Xiufang Chen in their research paper [2], review the 47 published research papers on online learning systems. Their main focus was to provide practical suggestions on how online learning platforms can improve themselves. Based on their results, a well-designed content the course, an effective learning environment can motivate the students, professional training of the faculty on how to deliver an online lecture more effectively and intellectual interaction between the instructor and students can increase the overall productivity of the students in an online education system.

IV. Scope of Process

The 365 Data Science is an online learning platform and over the past 2 years, the company has reached over 10 million active students learning from their platform. They have scaled quicker than anticipated and need a team of BI Analysts to oversee the scaling process how are they doing so far concerning the students that are enrolled and on basis of this how much more they can improve. This information will help them make better decisions concerning their platform, to ensure they can reach more people across the globe. As BI analysts, our responsibility is to uncover hidden information from the data. Below mentioned are a few of the things that will be uncovered -

- Identifying which courses are popular among students with the help of student registrations.
- Understanding the trend of students engaging with the training platform throughout the year; by - lessons, quizzes or exams.
- Recognizing the courses that are popular with the comparison of students' enrollments with the number of ratings received for the courses.
- Comprehending the percentage of students choosing different subscription types; like - monthly, quarterly and annual.
- Total purchase made throughout the year concerning purchase type and the increase in percentage every month.

All the above processes will be performed in a visualization tool called PowerBI, and the report and dashboard will be created as a final deliverable through which the company 365 Data Science can take advantage and improve its business model.

V. Design

The data used to perform the analysis of the company was taken from the 365 Data Science (online learning platform) which includes information about courses and students. This data was provided to the public as part of *The 365 Learning Data Challenge*. The aim of the challenge is to improve the data skills of the participant including but not limited to - the development of analytical skills, a chance to work on real-world data, etc. The task at hand is to take the role of a BI analyst and gather the information related to the company's success so far and identify the things that lack or could be improved. To further understand the relationship between the data, a data model is created along with the ER diagram.

A. Customer Integration

The customers from the company point of view for the 365 data science learning platform are the users looking to get trained in data science. These are the students who will be accessing the training throughout their subscriptions. Certain things ensured that these students get a seamless experience. This can be done by making sure that timely feedback through ratings is taken from the students concerning the courses they have undertaken, the quizzes they do, and the exams they take. Apart from that, the way the customer is registered in the system is also taken into consideration. This will be helpful to store the information about the customer i.e., the student, which includes details like - the unique id assigned to the student, the country the student is from and the date the student registered into the system.

Along with that, considering customer experience is an essential part of growing the business. A better customer experience helps in promoting customer loyalty and leads to further referrals by these customers to other people they know.

The introduction of data science mentors for students would be extremely beneficial as students will receive guidelines from the mentors concerning their respective courses. And whenever a student gives any kind of exam or quiz, as a part of the result, they will receive suggestion topics for the areas they lack knowledge in i.e., the area in which they received less scores along with detailed feedback.

B. Data Model

Data Modeling is a method to represent the communication between different data tables.
As we can see in Fig. 1, our data model has a *student info* table which has a unique id for each student, which is further connected to almost all the tables. This connection will help us build an overview of insights. Taking into consideration, the table *student info*, it has attributes like *student id*, *student*

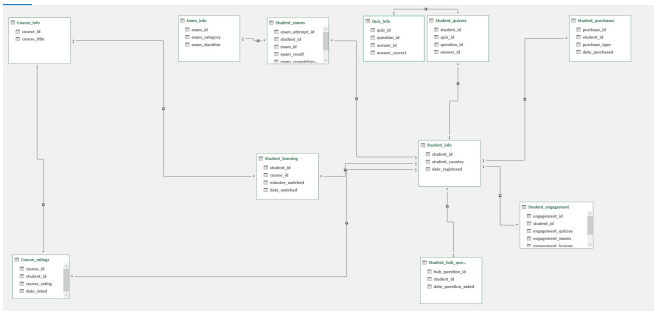

Fig. 1. Figure

country, *date registered*. The attribute *student id* can be seen forming a relationship with most of the tables. One of the tables the *student id* is connected to the table *student learning* which could be helpful in filtering the information regarding the maximum number of minutes the courses were watched in a particular country.
As from the data model we can see the connections clearly and produce the reports and dashboards accordingly.

C. ER Diagram

An entity relationship diagram is a type of flowchart which describes how entities are related to each other within a system. We have created the following ER diagram to show how our data are related to each other.

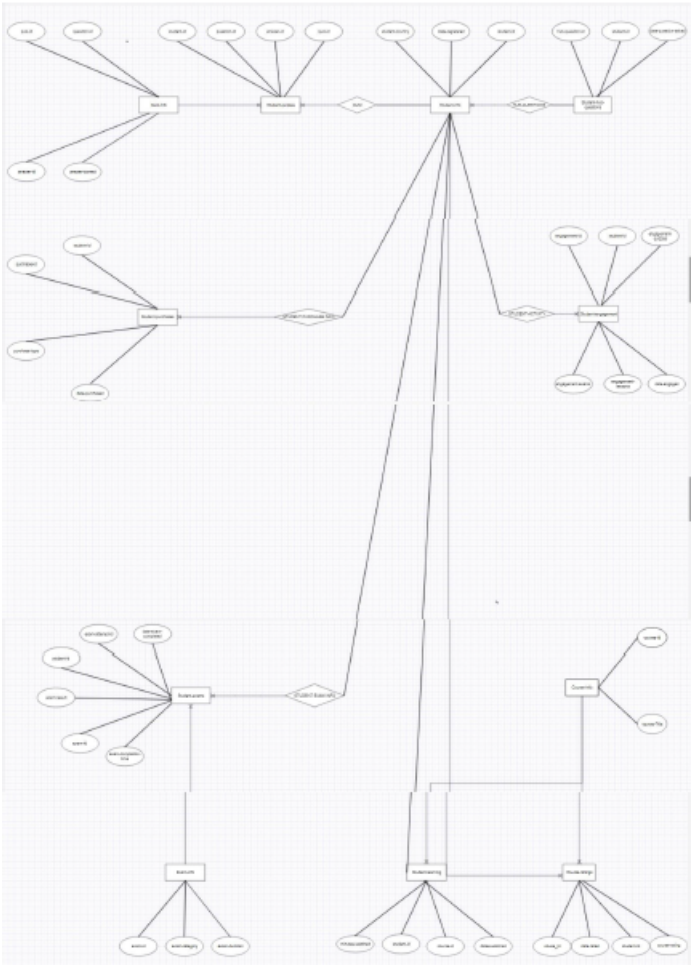

Fig. 2. ER Diagram

The entity relationship diagram (Fig. 2) shows the simplified view of our database schema. From the entity relationship diagram we can see the entity *student info* has *student country*, *date registered*, and *student id* attributes and has the mandatory relationship with *hub questions*, *quiz*, *student activity*, *student purchase info* and *student exam info* because as the student makes any activity like purchasing, taking an exam or asking a question, the student's engagement data will be updated. The exam info entity has a direct relationship with the student exams which is related to the student info because as the student takes the exam, their attempt, duration and date information will be updated. The course info has a relation with course rating which is connected to student info through student learning because as a student gives a rating or watches the course, the date the course was rated or course rating will be updated accordingly.

D. Data Dictionary

The company provided access to multiple tables as a part of their *learning challenge*. This data was used for analysis and holds information about students and courses. Fig. 3 shows the tables, the number of records they had and the information stored in these tables.

- *student_info* focuses on information about the students.
- *student_purchases* contains information about the type of subscriptions taken by the students.
- *student_learning* consists of information regarding the courses students have watched, the duration and the date they have watched on.
- *course_info* contains the information about the courses available on the platform for the students.
- *course_ratings* consists of details regarding the ratings received by the courses.
- *student_quizzes* stores details about the performances on quizzes given by the students.
- *quiz_info* gives the details about the quizzes available on the platform.
- *student_exams* holds information about the performance on the exams given by the students.
- *exam_info* holds the detailed information about the exams that are available on the company's platform
- *student_engagement* consists of information about en-

Table	Number of Records	Attributes
student_info	35,230	student_id
		student_country
		date_registered
student_purchases	3,041	purchase_id
		student_id
		purchase_type
		date_purchased
student_learning	64,535	student_id
		course_id
		minutes_watched
		date_watched
course_info	46	course_id
		course_title
course_ratings	2,500	course_id
		student_id
		course_rating
		date_rated
student_quizzes	1,47,029	student_id
		quiz_id
		question_id
		answer_id
quiz_info	4,741	quiz_id
		question_id
		answer_id
		answer_correct
student_exams	34,030	exam_attempt_id
		student_id
		exam_id
		exam_result
		exam_completion_time
		date_exam_completed
exam_info	156	exam_id
		exam_category
		exam_duration
student_engagement	65,371	engagement_id
		student_id
		engagement_quizzes
		engagement_exams
		engagement_lessons
		date_engaged
student_hub_questions	827	hub_question_id
		student_id
		date_question_asked

Fig. 3. Data Dictionary

gagements concerning lessons, exams and quizzes done by students.

- *student_hub_questions* holds information about the doubts or questions asked by students in the Q&A hub of the platform.

REFERENCES

[1] Xiangmin Zhang (Fall 2016) An Analysis of Online Students' Behaviors on Course Sites and the Effect on Learning Performance: A Case Study of Four LIS Online Classes Journal of Education for Library and Information Science, Vol. 57, No. 4 (Fall 2016), pp. 255-270

[2] Sun, A., & Chen, X. (2016). Online education and its effective practice: A research review. Journal of Information Technology Education: Research, 15, 157-190

[3] Data and data dictionary taken from 365 Data Science Challenge - https://365datascience.com/news/365-learning-data-challenge/

[4] ER model has been designed from Gliffy - https://www.gliffy.com/