National College of Ireland

National College of Ireland

Project Submission Sheet – 2022/2023

Student Name:	Piyush Dhawad, Sarthak Bhatnagar, Tejali Gangane, Usama Hanif
Student ID:	x21236933, x21185352, x21148872, x22108831
Programme:	MSc in Data Analytics **Year:** 2022-23
Module:	Business Intelligence and Business Analytics
Lecturer:	Vikas Sahni
Submission Due Date:	05-12-2022
Project Title:	Business Analysis of 365 Data Science
Word Count:	2656

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
<u>ALL</u> internet material must be referenced in the references section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature:	**Piyush Dhawad, Sarthak Bhatnagar, Tejali Gangane, Usama Hanif**
Date:	05-12-2022

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. Projects should be submitted to your Programme Coordinator.
3. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
4. You must ensure that all projects are submitted to your Programme Coordinator on or before the required submission date. **Late submissions will incur penalties.**
5. All projects must be submitted and passed in order to successfully complete the year. **Any project/assignment not submitted will be marked as a fail.**

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Business Analysis of 365 Data Science

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*—The company 365 Data Science is a learning platform that provides courses related to Data Science. Ever since the pandemic, online learning has been preferred by people over traditional physical classes. This has been extremely beneficial as they can watch courses or learn anything from the comfort of their homes or from anywhere they want. The company has provided access to data that holds information about the courses they provide and the students that are registered in those courses. The aim is to analyse and get insights concerning the way these students engage with the platform and identify the areas in which the company can improve. To achieve this a visualization tool called PowerBI is used. This report deals with the analysis of the data provided and suggestions to improve based on the insights generated.**

***Index Terms*—Business, Power BI, Online Learning Platform, Analysis, Report**

I. Introduction

Several forms of technological platforms support the online education system. In recent years, the trend of online learning systems has increased significantly because it integrates student management, learning resources, lesson delivery, student evaluation, etc. under one system. In online learning platforms, students can take online course videos and can then evaluate their knowledge by attempting quizzes, and exams and tracking their progress. One of the main reasons the online education system is becoming so popular is that students can take lessons whenever they want or are available to study/work. Students can attempt the exams more than once. The online education system has various advantages when compared to traditional physical classes. For example, online students have the flexibility of time and place to access the course resources and ask questions in the student hub. Mostly those people who work part-time or full-time and want to continue their education can take full advantage of an online education system.

Generally, an online learning system is almost effective as a traditional physical format. On the basis of the medium through which the online lectures are delivered, the system can be synchronous or asynchronous. The former includes mediums like video conferences, and live chats while the latter includes email and discussion forums. In both settings of the system, the role of the instructor has changed from lecturer to facilitator. Currently, the main focus of the e-learning system is the asynchronous approach. The main focus of this approach is to study the relationship between students' behaviour in an online learning system and the evaluation of their performance and provide ways and strategies on improving them.

II. Development of Solution

A. Source of Data

The data used for the analysis was provided by 365 Data Science as part of their challenge campaign. This challenge provides data for the participants to use different data analysis techniques to get a deeper understanding of the field. Their main aim is to build and improve the skills of the participants in the Data Science domain. As a part of this challenge, they have provided around 11 data tables. The data is from the dates 1st January 2022 to 20th October 2022. These tables are related to each other on the basis of common attributes. For information related to students, the common attribute among multiple tables is *student_id* and for details related to the courses, the common attribute is *course_id*.

These tables hold information about students, purchases made by them, the number of courses they watched and the ratings the courses have received, performance on quizzes and exams, information about quizzes and exams, the way students engage with the platform and questions asked by students.

B. Data Pre-processing

Before implementing the solution, a system needs to be formed, this system will have specific requirements for processing the data efficiently. This involves data pre-processing in order to make the raw data more efficient for the system and correct for any irregularities. Generally, we can find that pre-processing of data contributes a huge role in the accuracy of the final solution that we are aiming for. Furthermore, data pre-processing will also help our understanding of the data structure.

The process involved the following things -

- *Dealing with NULL or NA values*: Most of the datasets have some irregularities which might have no readings at times. These values will affect the solution and the trends if not taken care of. General methods to deal with the null value or NA value are either eliminating the entire reading associated with it or replacing the NA value with the appropriate placeholder value or taking the median or mean of the entire variable wherever applicable. We have

then searched the null values in the data set in order to treat it. After extensive analysis of the data, it was found that the data did not have any such values,

- *Assigning valid data types*: The attributes of the dataset, when it was loaded into PowerBI had *text* as the default data type, which was later changed to suitable types wherever required.
- *Checking the data consistency*: The data were checked to ensure that a consistent range of values is present across the attributes.

III. Development of System

There are multiple systems available in the industry today which can help us demonstrate the dashboard. We have chosen Power BI as a Business Intelligence platform purely due to its wide adaptation across industries. Not only it is a comprehensive tool which provides better ways for analysis and has AI-driven suggestions. Furthermore, on the organisational level, it is very easy to communicate the findings of deeper analysis across the board. The dashboard information that is been shown can be automated in future to show the real-time readings which can be helpful to track real-time development or show the impact of changes that are happening. Considering all these advantages it has made us easy to integrate the Dataset with the Power BI. The wide range of file format support means that employees can have the freedom to work with their preferred settings and bring the best with the implementation of PowerBI tools.

A. Architecture

Fig 1. gives the overview of the architecture flow of the whole process. The data provided by the company was in the form of CSV files. This data was loaded into a visualization tool called PowerBI by selecting the *Get data* option. Once the data was loaded into PowerBI, the data was checked for any missing values and the data type of the attribute was corrected. After that, the data was further used to uncover insights and was analysed to get an overview of the company and identify where it stood in terms of company progress.

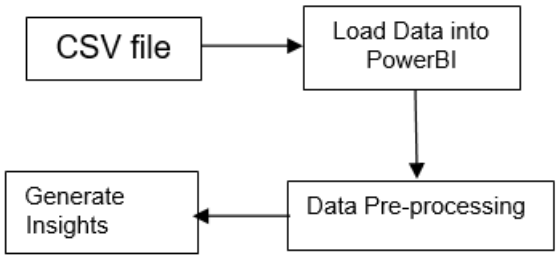

Fig. 1. Architecture

B. Benefits of the Solution

Having a system in place is extremely beneficial for a company to ensure a smooth flow of the process. Here are a few benefits of the system incorporated:

- Data is read from the CSV file stored at a local path and further loaded into PowerBI which is easy and efficient. As once the data is loaded, the data preparation and analysis can be directly done in the tool itself. If the data changes at some point or needs to be updated, the file just needs to be changed at the mentioned local path and the PowerBI will read the data from the previously supplied path.
- Cleaning the data in PowerBI is extremely easy as it provides various direct options to use and perform the task.

IV. Data Management

A. Relationships

Fig 2. shows the data model created for the data provided by the company. There exists multiple relationships between the tables.

- The table *student_info* has one to many relationship with *student_purchases*, *student_engagement*, *student_hub_questions*, *course_ratings*, *student_learning*, *student_quiz*, *student_exams*
- The table *course_info* has one to many relationship with the tables *course_ratings*, *student_quizzes*
- The table *quiz_info* has one to many relationship with the table *student_quizzes*

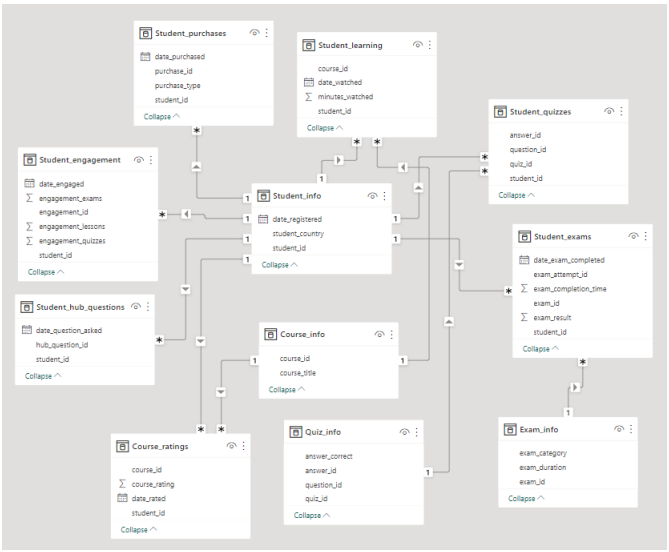

Fig. 2. Data Model in PowerBI

B. Layouts

Fig 3. shows the layout of the dashboard built to represent the insights concerning the data for the company. The first row of the dashboard represents the individual values for

the number of students, courses, the average rating provided for the course and the number of engagements. The second row represents the visuals that show the relationship between multiple attributes. For example, the first visual from the second row represents the popularity of the courses based on the number of ratings received by the course. The last row represents different visuals which show either the individual attributes or the relationship between attributes.

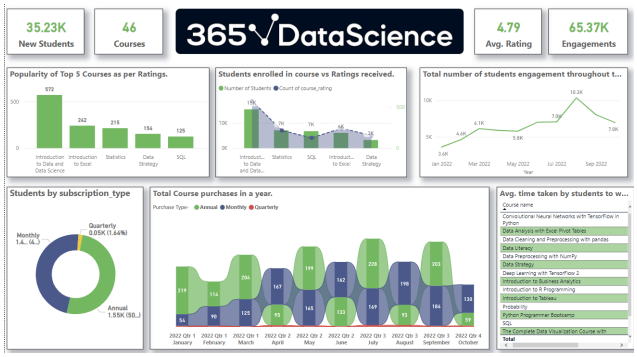

Fig. 3. Layout

V. Solution Implementation

A. Analysis of data using dashboard in PowerBI

The analysis of the data was done in a visualization tool called PowerBI. This analysis was done to uncover the hidden information present within the data which may not be directly visible to the naked eye. From Fig. 3 it can be observed that there are more than 35K new students who registered in the year 2022. The number of courses is 46 along with the average rating for all courses together is 4.79. The total number of engagements performed by students till October 2022 is more than 65K.

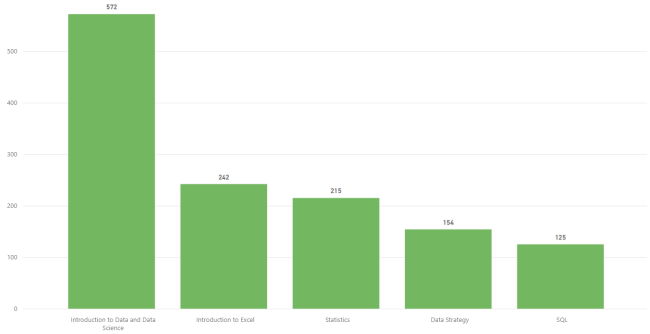

Fig. 4. Popularity of Top 5 courses as per ratings

Fig 4 indicates the top 5 courses that are popular among students on the basis of the number of ratings received by each course. It can be seen that the course *Introduction to Data and Data Science* is the one with the highest rating count of 572, which is further followed by *Introduction to Excel* with a rating count of 242 and along with some other courses like *Statistics*, *Data Strategy*, *SQL*.

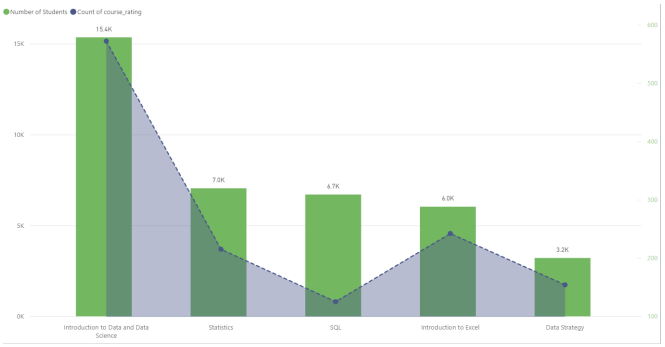

Fig. 5. Number of students enrolled in course vs ratings received

Fig. 5 represents the difference between the number of students that were enrolled in a course with respect to the number of ratings received by the course. It can be observed that the course with a higher rating count of 572 (as observed previously) has more than 15k registrations. That is, there is a significant difference between the enrollment of the course and the ratings provided. Similar is the case for other top courses, where the count of ratings received by each course is less than the number of students registered for that course.

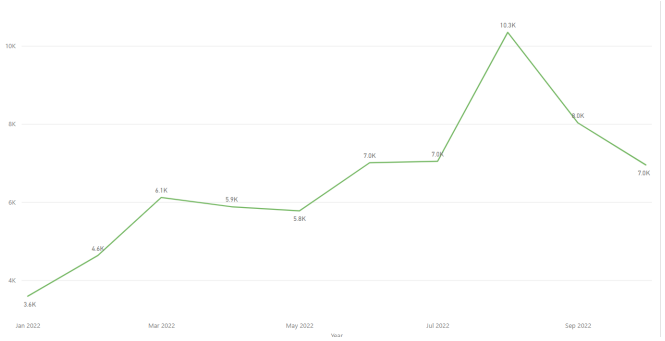

Fig. 6. Total Number of student engagements throughout the year

Fig. 6 shows us the number of times the students engaged with the platform throughout the year 2022. Starting with January the engagement was more than 3K and kept increasing till March. However, the engagement slightly dropped after it till May 2022. But later, the engagement seems to have risen till August After that the engagement kept dropping slowly till October.

Fig. 7 indicates the percentage and number of students subscribing to different types of purchases. It is clear that most of the students have purchased *Monthly* and *Annual* subscriptions as compared to *Quarterly*. There is a slight difference of a bit more than 3% between the percentage of students buying monthly and annual subscriptions. However, there are an extremely low number of students purchasing *Quarterly* subscriptions.

The ribbon graph in Fig 8. represents the total number of course purchases in the year 2022 which comprise subscription types ie. annual, monthly and quarterly. It is observed that in the starting month annual subscribed students are 219 which

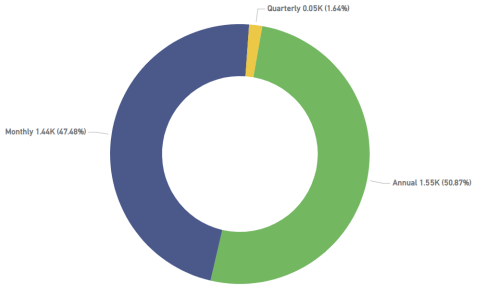

Fig. 7. Students by subscription type

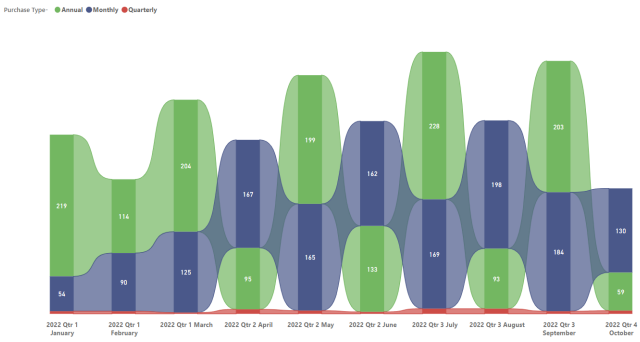

Fig. 8. Total Course Purchases in a year

is far more as compared to monthly subscribed students i.e., 54. As the month passes by we noticed both the subscription types are somewhat equally subscribed with little up and down values.

Moreover, it is discovered that the annual subscription is maximum in July due to the Data science Hero campaign competition which is very popular in the market and minimum in the month of October whereas the monthly subscription type is almost constant from April to October. Furthermore, Quarterly subscription is negligible with regards to other subscription types.

Course name	Average of minutes_watched
Convolutional Neural Networks with TensorFlow in Python	34.67
Data Analysis with Excel Pivot Tables	35.80
Data Cleaning and Preprocessing with pandas	37.21
Data Literacy	37.08
Data Preprocessing with NumPy	48.13
Data Strategy	32.05
Deep Learning with TensorFlow 2	31.25
Introduction to Business Analytics	36.59
Introduction to R Programming	39.79
Introduction to Tableau	44.33
Probability	28.75
Python Programmer Bootcamp	33.53
SQL	35.04
The Complete Data Visualization Course with Python, R, Tableau, and Excel	54.16
Time Series Analysis with Python	29.90
Total	**36.17**

Fig. 9. Average time taken by students to watch courses

Fig. 9 represents the table mentioning details about the average number of minutes watched by students for every course. It is evident that the most watched course is *The Complete Data Visualization Course with Python, R, Tableau and Excel*

VI. Solutions for the Business

After careful analysis of the data from the company, there are a few strategies that the company could implement to ensure they get more registrations and purchases, leading to revenue growth.

A. Internal - Within the organisation

The best way to grow the company at the base level is to ensure that all the internal processes are working fine. Make sure, that all the teams are timely trained and updated about the latest changes to the platform, to ensure they are aware of any queries that might be raised to them by the students and are able to solve them effectively.

There are a few changes that can be made within the organisation:

- Introducing Data Science mentors for respective courses. This will benefit students as they can reach out to the respective mentors for any queries related to the course or any project the course might be associated with.
- Providing categorised Q&A hub for each course. Make sure there exists a separate tab for Q&A under each course and ensure the answers are verified by the mentors of respective courses.
- Ensuring that the Social media team is active on LinkedIn and Instagram to promote more about the courses. The promotion or advertisement to be done should include small clips from the courses to let people get a glimpse of it.
- Conducting interviews of the employees who are from the Data Science domain to get the latest industry updates in the field and share them on the website.
- Holding live sessions of past students who have been able to land a job in the Data Science domain with the help of courses provided by the company 365DataScience and to be shared on the company website.
- Making a dashboard available for students on the website to check their metrics concerning course completion and overall progress.

B. External - For Students

Ensuring that students get the value or outcome they deserve is an important aspect of the growth of the company. This will ensure the students will benefit enough from the company which may further lead to them referring other people they know:

There are a few things that can be ensured to make sure the students are always satisfied with the platform:

- Ensuring students are awarded some incentive based on the number of minutes they watched, quizzes or exams they completed. This will help them stay motivated to learn and use the platform more. For example, awarding points after completing each video or lesson of the course, which can later be redeemed for a chance to have a

conversation for 10 minutes with an industry specialist from the same domain.

- A referral system should be introduced for students who have been registered for a longer time and are consistent or active on the platform. For example, students who have been active on the platform for more than 15 days, get a chance to refer the course or platform to their friend and return get awarded points.
- Discounts for students who have participated in the challenges hosted by the company and special discounts for students performing exceptionally well in these challenges.

VII. Conclusion

The company 365DataScience has been running great so far. However, there are certain things or changes the company needs to integrate to ensure it can improve and get better. After careful analysis of the data provided by the company, certain suggestions were made which can be beneficial for the company. The company has a scope to grow more over the future years as, since the pandemic, online learning is becoming a booming industry and is preferable by most people.

References

[1] Data is taken from 365 Data Science Challenge - https://365datascience.com/news/365-learning-data-challenge/