



## IDC4U Interdisciplinary Studies: AI and Business Innovation

### Project Step 5: - Planning data sources for your startup company

#### How do you plan to collect your data? (60-120 words)

Learning21 aims to enhance its AI-driven educational platform through two data collection methods: collecting our own data and using public datasets. We gather direct data from our educational activities to analyze student engagement and teaching effectiveness. Additionally, we utilize public datasets, which provide broader educational statistics and learning patterns. This combination allows us to benchmark our internal data against external trends, significantly improving our AI algorithms and understanding of learning contexts.

#### Where will your data be stored? (60-120 words)

We are considering using cloud services to host our platform, with options including Amazon Web Services, Google Cloud Platform, and IBM Cloud. We will analyze these services to determine the most suitable one for our needs. Additionally, we will utilize a Learning Management System (LMS) to support our learning solutions. Most of our data, including course contents, student learning logs, assignments, exams, student marks, and teaching feedback, will be stored in this LMS. Public datasets used for analysis will be downloaded and stored in publicly available data storage services such as Google Drive, OneDrive, and Dropbox.

#### Do you intend to use public datasets for your business analysis and planning? If so, what types of public datasets are you interested in? (100-200 words)

Yes, we do intend to use public datasets for our business analysis and planning, particularly in two main areas: student learning analytics and student enrollment data. Public datasets related to student demographics, such as geographical distribution, will help us analyze market segments and refine our marketing strategies. We are also interested in broader educational data that can offer insights into student engagement, performance, behavioral patterns, and feedback. These include:

Engagement Data: Public records of how students interact with educational platforms and resources.

Performance Data: Aggregated results from standardized tests and assessments available in public educational databases.

Behavioral Data: Studies and reports on student behavior within educational environments, accessible through educational research databases.

Feedback Data: Surveys and feedback aggregated by educational agencies and organizations.

Attendance Data: Reports and studies that correlate student attendance with academic outcomes.

By using these public datasets, we aim to compare our educational offerings with traditional services and competitors' solutions, which will guide us in enhancing our educational strategies and solutions.



## THE ERINDALE ACADEMY

www.teacademy.ca | Tel: 905-232-1576 | Email: info@teacademy.ca  
1576 Dundas Street West, Mississauga ON L5C 1E5

### **Why are these particular datasets valuable to your business? (100-200 words)**

There are some valuable public datasets available for our business, primarily sourced from Kaggle. we access various education-related datasets such as Student Study Performance, Student Attitude and Behavior, Student School Attendance, Student Marks, and Student Stress Factors, along with datasets predicting students' dropout rates and academic success. These datasets are crucial as they enable us to explore learning patterns, performance metrics, and related factors influencing student outcomes. By analyzing these data, we can identify trends, predict future educational challenges, and tailor our interventions more effectively.

Learning21 will provide Massive Open Online Courses (MOOCs), a format that revolutionized the learning landscape earlier this decade. Popular MOOC platforms include Coursera, Udacity, edX, Kaggle Learn, Datacamp, Fast.AI, DataQuest, and TheSchool.AI. Learner feedback is crucial for analyzing learner satisfaction. Learning21 will utilize public datasets to conduct sentiment analysis on learners' feedback. This analysis will help assess the quality of MOOCs compared to traditional education and among various MOOC providers. Kaggle offers several relevant datasets for free, such as "The MOOC Wars" series. One dataset, "The MOOC Wars: Kaggle's Perspective," includes 42.56 MB of data from nearly 16,000 responses to question 39.

Additionally, we plan to explore other data sources like Amazon Datasets, the UCI Machine Learning Repository, and Google's Dataset Search Engine to further enrich our analytical capabilities.

### **How do you plan to access or obtain these datasets? (60-120 words)**

We plan to access these datasets primarily through platforms that offer them for free or at an affordable membership fee. Specifically, datasets hosted on Kaggle are available for free download. This accessibility allows us to efficiently gather the data necessary for our analysis without significant cost barriers.