# **Project Title:**

Dự đoán hành vi người dùng nhằm tối ưu hóa trải nghiệm học tập trên nền tảng trực tuyến

# **Purpose and Outcome:**

* **Purpose:** Tăng mức độ hoàn thành khóa học của học viên khi tham gia trải nghiệm học tập trên nền tảng trực tuyến
* **Outcome:** Hướng đến việc tối ưu hóa trải nghiệm học tập trực tuyến của học viên.

# **Dataset:**

* **Source:** Predict Online Course Engagement Dataset:

<https://www.kaggle.com/datasets/rabieelkharoua/predict-online-course-engagement-dataset>

* **Description:** Bộ dữ liệu này thu thập số liệu về mức độ tương tác của người dùng từ nền tảng khóa học trực tuyến, tạo điều kiện phân tích các yếu tố ảnh hưởng đến việc hoàn thành khóa học. Bộ dữ liệu này bao gồm thông tin nhân khẩu học của người dùng, dữ liệu cụ thể về khóa học và số liệu về mức độ tương tác.
* **Structure:**
  + UserID: Mã định danh duy nhất cho mỗi người dùng
  + CourseCategory: Danh mục khóa học mà người dùng đã tham gia \(ví dụ: Lập trình, Kinh doanh, Nghệ thuật)
  + TimeSpentOnCourse: Tổng thời gian người dùng dành cho khóa học, tính bằng giờ
  + NumberOfVideosWatched: Tổng số video người dùng đã xem
  + NumberOfQuizzesTaken: Tổng số bài kiểm tra người dùng đã làm
  + QuizScores: Điểm trung bình người dùng đạt được trong các bài kiểm tra (tính theo phần trăm)
  + CompletionRate: Tỷ lệ phần trăm nội dung khóa học người dùng đã hoàn thành
  + DeviceType: Loại thiết bị người dùng sử dụng (Loại thiết bị: Máy tính để bàn **(0)** hoặc Điện thoại di động **(1)**)
  + ***CourseCompletion (Target Variable)***: Trạng thái hoàn thành khóa học (**(0)**: Chưa hoàn thành, **(1)**: Đã hoàn thành)

# **Công cụ:**

* Power Bi: Trực quan hóa dữ liệu
* Python trong Google Colab: Dùng để làm sạch, Phân tích khám phá, Xây dựng và đánh giá mô hình Logistic, Trực quan hóa dữ liệu.

# **Các bước làm:**

* Data Cleaning (Làm sạch dữ liệu): Xử lý giá trị thiếu, sửa kiểu dữ liệu và loại bỏ các bản ghi trùng lặp.
* EDA (Phân tích dữ liệu khám phá): Tạo thống kê tóm tắt, trực quan hóa phân phối và khám phá mối quan hệ giữa các biến.
* Analysis (Phân tích): Thực hiện phân tích tương quan, xây dựng và đánh giá mô hình hồi quy logistic, và xác định các yếu tố chính tác động đến việc hoàn thành khóa học.
* Visualization (Trực quan hóa): Tạo các hình ảnh trực quan như biểu đồ cột, biểu đồ hộp (boxplot) để hỗ trợ phân tích.
* Data Storytelling (Kể chuyện dữ liệu): Trình bày các phát hiện với một câu chuyện giải thích các yếu tố chính dẫn đến sự rời bỏ của khách hàng và các khuyến nghị để giảm tỷ lệ rời bỏ.

# **Kết quả:**

* Học viên hoàn thành nhiều bài kiểm tra hơn sẽ có tỷ lệ hoàn thành khóa học cao hơn.
* Học viên có điểm kiểm tra cao hơn sẽ có tỷ lệ hoàn thành khóa học cao hơn.
* Học viên có tỷ lệ hoàn thành các phần trong khóa học cao hơn sẽ có tỷ lệ hoàn thành toàn bộ khóa học cao hơn.
* Học viên dành nhiều thời gian cho khóa học hơn sẽ có tỷ lệ hoàn thành khóa học cao hơn.
* Học viên xem nhiều video hơn sẽ có tỷ lệ hoàn thành khóa học cao hơn.
* Học viên làm nhiều bài kiểm tra hơn có thể dành ít thời gian hơn cho việc xem lại tài liệu khóa học hoặc video.

**Khuyến nghị:**

* **Tập trung vào các yếu tố ảnh hưởng mạnh đến việc hoàn thành khóa học:**
  + Khuyến khích học viên làm nhiều bài kiểm tra hơn bằng cách cung cấp các bài kiểm tra thường xuyên, bài kiểm tra thực hành và các hoạt động tương tác.
  + Cung cấp phản hồi chi tiết về bài kiểm tra, tài liệu ôn tập và các buổi hỗ trợ để giúp học viên cải thiện điểm số.
  + Theo dõi tiến độ học tập của học viên và cung cấp hỗ trợ kịp thời cho những người có nguy cơ bỏ học.
* **Khuyến khích xem video và dành thời gian cho khóa học:**
  + Cung cấp các video hấp dẫn, dễ hiểu và có liên quan đến nội dung khóa học.
  + Khuyến khích học viên dành đủ thời gian cho việc học tập bằng cách cung cấp lịch trình học tập, nhắc nhở và các công cụ quản lý thời gian.