# ETL Case:

Our learning team would like to consume data from a Learning Management System. They would like to see which users enrolled into which courses.

1. Using the following LMS documentation <https://canvas.instructure.com/doc/api/>, can you please explain how you could retrieve course data for users?
2. What are the development steps for the ETL?
3. What other data could be interesting to retrieve and why?

I have answered questions 1, 2 and 3 through a user manual that covers the entire workflow that I will present. https://miro.com/app/board/uXjVM8o4ENU=/?share\_link\_id=295313621447

# Data Visualization Case:

A recruitment leader feels like they could be more efficient in their recruitment process and would like to understand what step to focus-on first. Create a dashboard that would help her answer this question that can be used in the future to optimize their operations.

Dataset: “case recruitment dataset.xlsx”

1. Requirements Gathering: Understand the business requirements and data sources. Identify the data to be extracted, transformed, and loaded, as well as any specific data quality or integration requirements.
2. Data Source Analysis: Analyze the structure, format, and quality of the source data. Identify any data inconsistencies, missing values, or data quality issues that need to be addressed during the transformation process.
3. Extraction: Extract the data from the source systems or files. This may involve querying databases, accessing APIs, reading flat files, or any other method of retrieving data.
4. Data Cleaning and Validation: Cleanse the extracted data by handling missing values, removing duplicates, and performing data validation checks. This step ensures that the data is accurate, consistent, and conforms to the defined rules and standards.
5. Transformation: Apply transformations to the extracted data to convert it into the desired format for analysis or loading into the target system. This may involve data mapping, aggregation, filtering, joining, or any other data manipulation operations.
6. Data Loading: Load the transformed data into the target system, such as a database, data warehouse, or analytics platform. The loading process may include creating tables, defining schemas, and inserting or updating records.
7. Data Quality Assurance: Perform data quality checks and validations on the loaded data to ensure its accuracy and integrity. This step involves verifying data completeness, consistency, and conformity to the expected standards.
8. Error Handling and Logging: Implement mechanisms to handle errors, exceptions, and data inconsistencies encountered during the ETL process. Logging and error tracking help in identifying and resolving issues efficiently.
9. Scheduling and Automation: Set up a scheduling mechanism to run the ETL process at regular intervals or as per the defined schedule. Automation ensures that the data is kept up to date and the ETL pipeline operates reliably.
10. Testing and Validation: Conduct thorough testing to validate the ETL pipeline's functionality, data accuracy, and performance. Perform end-to-end testing, unit testing, and integration testing to identify and rectify any issues.
11. Documentation: Document the ETL process, including data flow diagrams, transformation rules, and technical specifications. Documentation helps in understanding and maintaining the ETL pipeline in the future.
12. Monitoring and Maintenance: Set up monitoring mechanisms to track the performance, data quality, and health of the ETL pipeline. Regular maintenance and updates may be required to accommodate changes in data sources, transformation rules, or target systems.
13. User Demographics: Retrieving demographic information about the users, such as age, gender, location, educational background, or job title, can provide insights into the characteristics of the learner population. This information can be useful for analyzing learning patterns, personalizing course offerings, or identifying target audiences for specific courses.
14. Course Completion Data: Tracking course completion data, including the percentage of users who completed each course, the average completion time, or the completion rate over time, can help evaluate the effectiveness and engagement of the courses. It provides insights into learner behavior, course difficulty, or potential areas for improvement in the course design or content.
15. Assessment Results: Capturing assessment results, such as quizzes, tests, or assignments, allows for analyzing the performance and learning outcomes of the users. It helps identify areas where learners excel or struggle, evaluate the effectiveness of assessments, and make data-driven decisions for course improvements or individualized feedback.
16. User Interactions: Collecting data on user interactions within the learning management system, such as login patterns, course access frequency, time spent on each course page, or discussion forum participation, can provide valuable insights into user engagement and learning behaviors. Analyzing these interactions can help identify active learners, popular course materials, or areas where learners need additional support.
17. Course Feedback and Ratings: Gathering feedback and ratings from users about their learning experiences, course satisfaction, or specific course elements can provide valuable insights into the quality and effectiveness of the courses. It helps identify strengths and weaknesses, areas for improvement, and can guide future course development or modifications.
18. Learning Analytics: Incorporating learning analytics data, such as clickstream data, social network analysis, or sentiment analysis from discussion forums, can provide deeper insights into user behavior, collaboration patterns, or sentiment trends. These advanced analytics techniques help identify hidden patterns, predict learner outcomes, or personalize learning experiences.
19. Instructor Data: Including information about instructors, such as their qualifications, teaching experience, or user ratings, can provide additional context for understanding course quality and learner satisfaction. It allows for analyzing the impact of instructor characteristics on course outcomes or identifying highly effective instructors.