

Data Dictionary

Online Course Engagement Analytics

Project Overview

This document describes the cleaned and transformed dataset created for the Online Course Engagement Analytics project. The data is structured using a **star schema**, with one central fact table and three dimension tables for optimal Business Intelligence reporting.

1. Fact Table: fact_engagement.csv

Core Engagement Metrics

Each row represents a single user's activity in a single course.

Column Name	Data Type	Description
UserSK (FK)	Integer	Foreign Key: A unique identifier for a user. Links to the dim_user table.
CourseSK (FK)	Integer	Foreign Key: A unique identifier for a course category. Links to the dim_course table.
DeviceSK (FK)	Integer	Foreign Key: A unique identifier for a device type. Links to the dim_device table.
TimeSpentOnCourse	Float	Total time in hours that the user spent on the course.
NumberOfVideosWatched	Integer	Total count of videos watched by the user.
NumberOfQuizzesTaken	Integer	Total count of quizzes attempted by the user.
QuizScores	Float	Average score achieved by the user on quizzes (e.g., 85.5).
CompletionRate	Float	Percentage of course content completed by the user (e.g., 75.0 for 75%).
EngagementScore	Float	Normalized score ranging from 0 to 1, representing overall user activity. Calculated from time spent, videos watched, and quizzes taken.
CourseCompletion	Integer	Binary flag indicating course completion status. 1 = Completed, 0 = Not Completed.

2. Dimension Table: dim_user.csv

User Lookup

Each row represents a unique user.

Column Name	Data Type	Description
UserSK (PK)	Integer	Surrogate Key: The primary unique identifier for a user. This is the main identifier.
UserID	Integer	Business Key: The original user ID from the raw data.

3. Dimension Table: dim_course.csv

Course Category Lookup

Each row represents a unique course category.

Column Name	Data Type	Description
CourseSK (PK)	Integer	Surrogate Key: The primary unique identifier for each course category. This is the main identifier.
CategoryName	String	The name of the course category (e.g., "Programming", "Health", "Arts").

4. Dimension Table: dim_device.csv

Device Type Lookup

Each row represents a unique device type.

Column Name	Data Type	Description
DeviceSK (PK)	Integer	Surrogate Key: The primary unique identifier for each device type. This is the main identifier.
DeviceType	Integer	The original numeric code for the device type (0 or 1).
DeviceTypeDesc	String	Human-readable description of the device type: "Desktop" or "Mobile".

Key Abbreviations:

- PK = Primary Key
- FK = Foreign Key
- SK = Surrogate Key