Capstone project

Stage 1

By : group 6

Introduction

The Capstone Project focuses on data management and governance within a Learning Management System (LMS).

- The LMS was initially developed quickly to meet urgent online learning needs.
- The system lacks proper reporting, automation, and governance, leading to inefficiencies.
- This project aims to enhance data integrity, improve analytics, and establish governance frameworks.

Project Goals

The project aims to build a well-structured, efficient, and scalable LMS by addressing the following:

Data cleaning and consistency → Process data to ensure accuracy and reliability by removing duplicates and ensuring consistent formatting.

Advanced analytics & reporting → Utilize Power BI/Tableau for real-time insights.

Data governance policies → Enhance security, privacy, and consistency of data

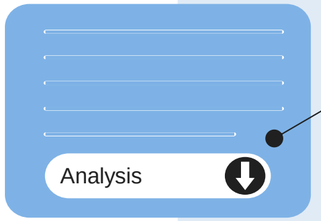
Data Understanding & Source

- Attendance 34 files: Records of student attendance (CSV, TSV files).
- Center: Data related to educational centers.
- Change New Joiners: Information on newly joined students.
- Change Requests: Various change request data.
- Cohort Assignments: Data on student group allocations.
- Cohort Schedule: Schedules for student cohorts.
- Student Status: Tracking the status of each student in the system.

Data Preprocessing

Data Preprocessing

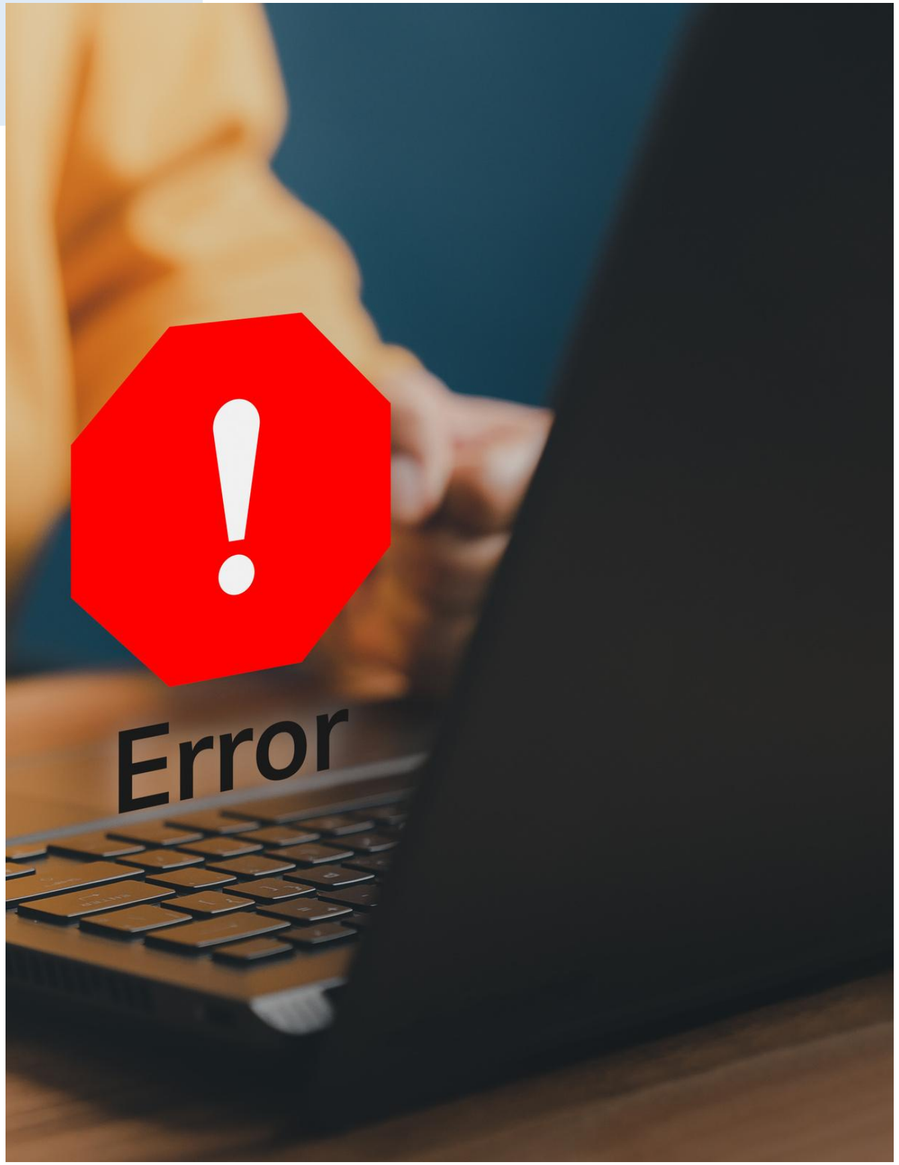
- Data Transformation : The data was transferred to MySQL for organization and analysis , For attendance records, we used Python to merge all logs into a single file then uploaded into the SQL database.
- Removing duplicates : using DISTINCT
- Handling missing values :Filling in or replacing missing values to prevent errors Keeping them empty if required.
- Data formatting: Using DATE_FORMAT for date standardization.
- Removing unnecessary spaces using TRIM.
- Data merging: Combining data from multiple sources with UNION.

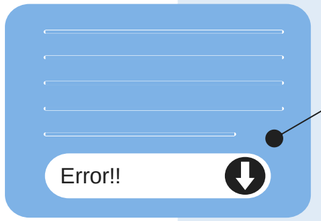


Data analysis

Data analysis

- Some SQL queries were used to extract data and generate new tables for further analysis in Power BI.
- In Power BI, DAX language, measures, and filters were applied to highlight key insights and statistics that we found particularly noteworthy.
- Power BI was used to establish relationships between tables, utilizing student_id as the primary key, along with other relationships such as level and more.





Problem management

Problem

Managing attendance and LMS data requires a structured governance framework to ensure integration, security, analytics, and cost efficiency. Challenges include data consistency, reporting accuracy, and scalability.

Solution

1.Data Integration & Management – Centralized database with ETL pipelines for clean, unified data.
2.Analytics & Reporting – Dashboards tracking student count, withdrawals, enrollment, attendance, and completion rates.
3.Ad-hoc Reporting for CEO – Interactive dashboard with filters for dynamic analysis of attendance trends.

Problem

Sending and sharing the file among team members causes errors and issues related to verifying versions and data formatting.

Solution

Standardize the file name and path, and use consistent naming conventions to avoid conflicts or errors when sharing data among team members.

Inconsistent date data leads to difficulties in correctly analyzing and processing the data.

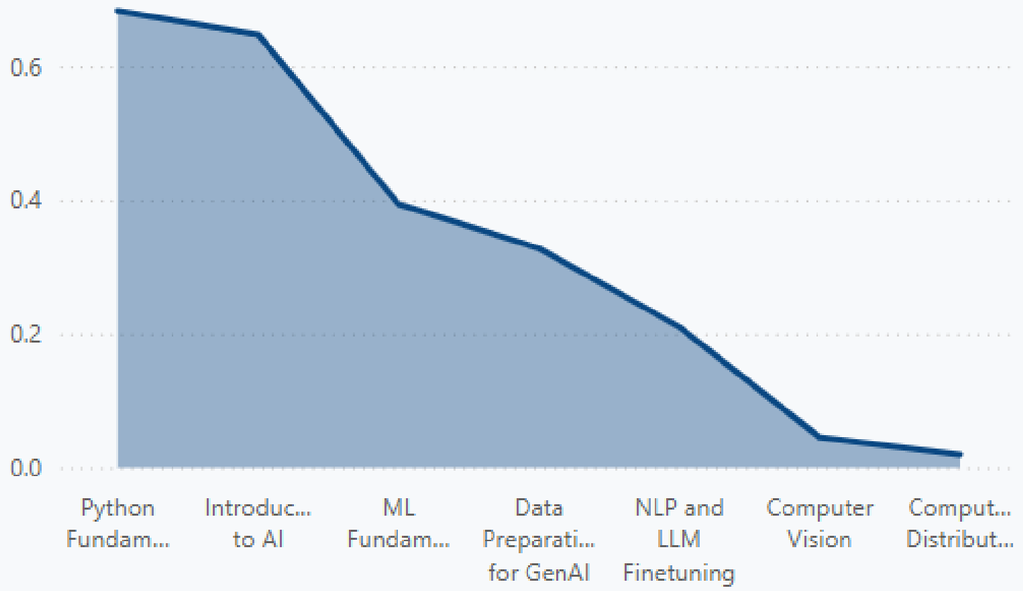
Convert the dates to a string variable first, then reformat them into a unified date format to ensure data consistency.

Conclusion

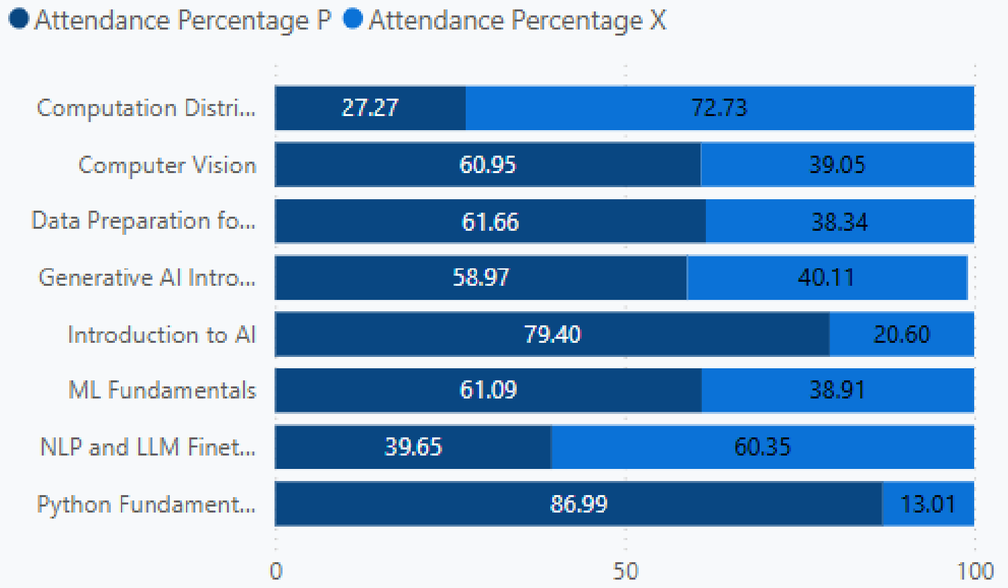
- These analyses contributed to enhancing data quality and supporting data-driven decision-making.
- Factors influencing attendance and dropout rates were identified.
- This data can be used to improve training programs and enhance the learning experience.
- Trends and patterns in student performance were uncovered, helping to identify areas in need of improvement.
- These analyses provide insights into the effectiveness of teaching strategies used in the training courses.

DATA OVERSIGHT

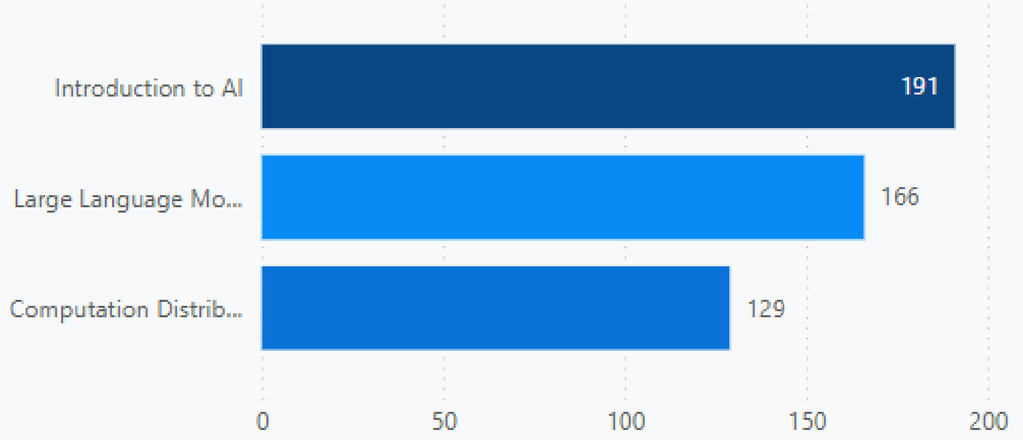
COMPLETION RATE PER STUDENT BY COURSE NAME



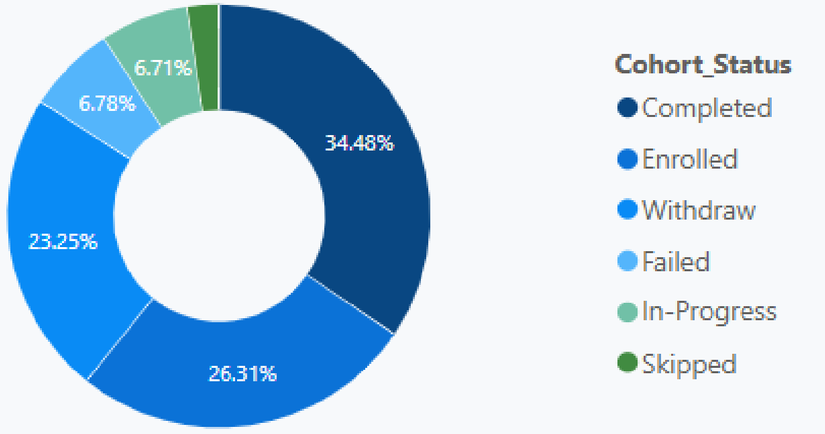
ATTENDNCE PERCENTAGE P & X FOR EACH COURSE



NUMBER OF GRADUATION FOR EACH COURSES



DISTRIBUTION OF STUDENTS BY COHORT STATUS



ENROLLMENT NUMBER

1448

STUDENTS NUMBER

427

COURSES NUMBER

9

PRESENT PERCENTAGE

62.21

ABSENCE PERCENTAGE

37.64

Recommendations

- Holidays should be taken into account before scheduling new courses, as they may impact attendance and completion rates.
- A high dropout rate and frequent absences were observed in certain courses, such as Distribution Computation, possibly due to the course's difficulty or the workload of its content.
- Conversely, courses like Python and Introduction to AI showed high completion rates and low absenteeism, which may be attributed to their engaging content or the high demand for these fields.