Data Collection and Preprocessing Phase

Date	10 July 2024
Team ID	739779
Project Title	Predictive Modelling for H1b Visa Approval Using Machine Learning
Maximum Marks	2 Marks

Data Quality Report Template

The Data Quality Report Template will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Data Source: Customer Order Database

Data Source	Data Quality Issue	Severity	Resolution Plan
Customer Order Table	Missing values in the 'Order Date' field.	High	Implement a mandatory check for 'Order Date' field before data entry
	Inconsistent formatting in the 'ID' field (e.g., mix of numbers and letters).	Medium	Apply a standard format and use a script to reformat existing entries.
	Outdated 'Shipping Address' information.	Low	Regularly update addresses through automated address validation APIs.
	Duplicate entries	High	Develop a duplication script to identify and remove duplicate records.

	Invalid Payment Method	Medium	Validate Payment Method against a predefined set of acceptable values

Data Source: H1b Visa Approval Dataset

Data Source	Data Quality Issue	Severity	Resolution Plan
H1b Dataset	Missing values in job_title field	High	Implement a mandatory check for job_title field before data entry.
	Incorrect approval_status values	High	Validate approval_status values against a predefined set of acceptable values.
	Inconsistent formatting in salary field (e.g., $/hour)	Medium	Apply standard format and use a script to reformat existing entries.