STUDENT PERFORMANCE INSIGHTS DASHBOARD

Week 1 – Data Understanding & Import

Data Miners

DATA MINERS

Contents

Project Overview

The goal of this project is to analyze student performance data to uncover trends and insights, with the aim of building an interactive dashboard.

This week's objective includes understanding data types, inspecting the dataset structure, and importing the data into MySQL Workbench and Databricks.

Dataset Information:

- **Source**: Kaggle: Students Performance in Exams
- **Schema Name** (MySQL): side_project
- **Table Name**: students_performance
- **Rows**:1000
- **Columns**: 8

Column Data Types

COLUMN_NAME	DATA_TYPE	CHARACTER_MAXIMUM_LENGTH	NUMERIC_PRECISION	IS_NULLABLE	COLUMN_KEY
gender	text	65535	NULL	YES	
race/ethnicity	text	65535	NULL	YES	
parental level of education	text	65535	NULL	YES	
lunch	text	65535	NULL	YES	
test preparation course	text	65535	NULL	YES	
math score	int	NULL	10	YES	
reading score	int	NULL	10	YES	
writing score	int	NULL	10	YES	

The schema overview shows that all categorical fields (gender, race/ethnicity, parental level of education, lunch, and test preparation course) are stored as **TEXT** types, each allowing NULL values. The score fields (math, reading, and writing) are defined as **INT** with standard numeric precision and allow NULLs. No primary keys or default values are set, indicating a simple, flat dataset structure imported directly from the source.

Data Preview

gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
female	group B	bachelor's degree	standard	none	72	72	74
female	group C	some college	standard	completed	69	90	88
female	group B	master's degree	standard	none	90	95	93
male	group A	associate's degree	free/reduced	none	47	57	44
male	group C	some college	standard	none	76	78	75
female	group B	associate's degree	standard	none	71	83	78
female	group B	some college	standard	completed	88	95	92
male	group B	some college	free/reduced	none	40	43	39
male	group D	high school	free/reduced	completed	64	64	67
female	group B	high school	free/reduced	none	38	60	50

This table displays a sample of student performance records, combining demographic attributes with exam scores. Each row represents one student and includes fields for gender, race/ethnicity, parental education level, lunch type, and whether a test preparation course was completed. Academic performance is captured through three numerical columns: math, reading, and writing scores (ranging from 0 to 100). This preview illustrates the dataset's structure, five categorical variables and three numerical variables and highlights potential relationships, such as higher scores among students who completed test preparation courses and differences linked to lunch type or parental education.

Categorical Columns Summary

column_name	value	count
gender	female	518
gender	male	482
race/ethnicity	group B	190
race/ethnicity	group C	319
race/ethnicity	group A	89
race/ethnicity	group D	262
race/ethnicity	group E	140
parental level of education	bachelor's degree	118
parental level of education	some college	226
parental level of education	master's degree	59
parental level of education	associate's degree	222
parental level of education	high school	196
parental level of education	some high school	179
lunch	standard	645
lunch	free/reduced	355
test preparation course	none	642
test preparation course	completed	358

The table shows the frequency of values for each categorical column in the dataset. There are five main categories:

- Gender: 518 female, 482 male (almost balanced).
- Race/Ethnicity: Group C is the largest (319), followed by Group D (262), Group B (190), Group E (140), and Group A (89).
- Parental Education: Most common levels are “some college” (226) and “associate’s degree” (222), while “master’s degree” is least common (59).
- Lunch Type: 645 students have standard lunch; 355 have free/reduced lunch.
- Test Preparation: 642 students did not complete the course; 358 completed it.

This summary highlights the distribution of demographic and contextual factors that can influence performance analysis.

Numerical Columns Summary

column_name	min_value	max_value	avg_value	std_dev
math score	0	100	66.0890	15.155496659628154
reading score	17	100	69.1690	14.592890015346521
writing score	10	100	68.0540	15.188057281956775

This table shows key statistics for the three numerical columns: math, reading, and writing scores. Each subject has a score range from low to high, along with average performance and variability:

- Math Score: Min0, max100, average 66.09, standard deviation ~15.16.
- Reading Score: Min17, max100, average 69.17, standard deviation ~14.59.
- Writing Score: Min10, max100, average 68.05, standard deviation ~15.19.

Overall, students perform slightly better in reading and writing than in math, and all subjects show similar variability (around 15 points), indicating moderate spread in scores.

Missing Values Check

	column_name	missing_count
▶	gender	0
	race/ethnicity	0
	parental level ...	0
	lunch	0
	test preparatio...	0
	math score	0
	reading score	0
	writing score	0

The table shows that there are no missing values across all columns in the dataset. Each field, gender, race/ethnicity, parental level of education, lunch, test preparation course, math score, reading score, and writing score, has a missing count of 0. This means the dataset is complete and ready for analysis without requiring data cleaning or imputation.

Observations / Notes

Data Completeness: The dataset has 1,000 rows and 8 columns with no missing values, so it's ready for analysis.

Data Structure:

- 5 categorical columns: gender, race/ethnicity, parental education, lunch, test preparation.
- 3 numerical columns: math, reading, and writing scores.
  This mix supports demographic and performance analysis.

Categorical Overview:

- Gender is almost balanced.
- Race/ethnicity is diverse, with Groups C and D most common.
- Parental education varies; "some college" and "associate's degree" are most frequent.
- Lunch type and test prep completion may influence performance.

Performance Trends:

- Average scores: Math (66), Reading (69), Writing (68).
- Reading and writing slightly outperform math.
- Variability is moderate across all subjects.

Insights for Dashboard:

- Compare scores by test prep and lunch type.
- Check links between parental education and performance.
- Analyze patterns by race/ethnicity and gender.

Technical Notes:

- Schema allows NULL and lacks a primary key; adding an ID and constraints is recommended.
- Data is clean and ready for visualization and modelling.