Dana Alshanbari

DS210 Final Project Writeup

A. Project Overview

Goal: To see if the amount of time a student studies correlates with their academic performance, and what the distribution of study hours and exam grades looks like.

Dataset:

- Source: Kaggle Dataset
- Size: 10,000 student rows
- Included directly in project as: **"student_performance_large_dataset.csv"**
- Small sample dataset (9 rows): **"test_data.csv"**

B. Data Processing

- Loaded using a custom **read_data** function in **data.rs**
- Used the csv and serde crates to deserialize data directly into a Student struct
- Selected relevant fields:
    - Study_Hours_per_Week (as f32)
    - Exam_Score (%) (as f32, renamed to score)
- Invalid or incomplete rows were skipped automatically

C. Code Structure

Modules:

- **data.rs** – Loads and parses the CSV dataset
- **analysis.rs** – Performs binning and plots a correlation graph

Key Structs:

- **Student**
    - Represents: One student's study and exam data
    - Fields:
        - study_hours_per_week: f32
        - score: f32

Key Functions:

1. **read_data**
    - Purpose: Load and parse student data from CSV
    - Logic: Uses serde to deserialize and filters out invalid records
2. **avg_scores_bins**
    - Purpose: Bins students by study hour ranges and computes average scores
    - Output: Vector of (study hour bin, average score)
3. **plot_hours_vs_score**
    - Purpose: Plots a line chart using Plotters showing score vs. study hour bins
    - Output: Saves **"hours_vs_scores.png"** in the root directory

D. Tests

cargo test Output:

running 3 tests
test analysis::tests::test_avg_score ... ok
test data::tests::test_load_data ... ok
test analysis::tests::test_plot ... ok

test result: ok. 3 passed; 0 failed; 0 ignored; 0 measured; 0 filtered out; finished in 0.31s

Test Descriptions:

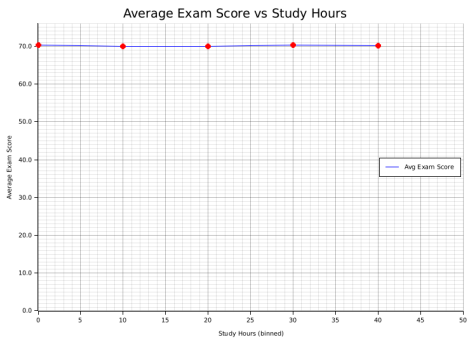
1. **test_avg_score**
    - Confirms **avg_scores_bins** computes expected bin averages using sample CSV
2. **test_load_data**
    - Checks that all rows in the test CSV load correctly and one sample row has the right values
3. **test_plot**
    - Verifies that **plot_hours_vs_score** runs without error and generates the graph file

E. Results

Sample Output:

Chart saved to hours_vs_scores.png

Chart Output:



Interpretation:

There is little variation in average scores across study hour bins. This suggests that study hours alone might not be a reliable/strong indicator of exam performance in this dataset.

F. Usage Instructions

Build & run the project:

- **cargo run** or **cargo --release**

Output:

- A chart image will be saved to the root directory as: **“hours_vs_scores.png”**

Expected Runtime:

- Both **cargo run** and **--release** complete in a few seconds.

G. AI Assistance Disclosure

- I used ChatGPT for code debugging and fixing test logic.
- Outside source (not AI): used Plotters online guide to create my chart.