

StudentPerformance

September 16, 2025

```
[2]: # 1. Install required packages
      # Run this once in your environment
      # pip install kaggle pandas sqlite3
```

```
import os
import pandas as pd
import sqlite3
```

```
[3]: # 2. Download Kaggle dataset using Kaggle API
      # First ensure: You have your Kaggle API token file (kaggle.json) in ~/.kaggle/
      # Kaggle dataset identifier: spscientist/students-performance-in-exams

      #os.system('kaggle datasets download -d spscientist/
      ↪students-performance-in-exams -p ./ --unzip')
```

```
[1]: #print(os.getcwd())
```

```
[5]: # 3. Load CSV into pandas
      df = pd.read_csv('../StudentsPerformance.csv')
```

```
[6]: # 4. Load DataFrame to SQLite
      conn = sqlite3.connect('students_performance.db')
      df.to_sql('students_performance', conn, index=False, if_exists='replace')
```

```
[6]: 1000
```

```
[9]: # 5. Example SQL Queries and Analysis
      def query_and_print(sql):
          result = pd.read_sql(sql, conn)
          print(result, '\n')

      analysis_queries = {
          'Average scores by gender': '''
              SELECT gender, AVG("math score") AS avg_math, AVG("reading score") AS_
              ↪avg_reading, AVG("writing score") AS avg_writing
              FROM students_performance
              GROUP BY gender;
```

```

'''
'Impact of test preparation': '''
    SELECT "test preparation course", AVG("math score") AS avg_math,
    ↪AVG("reading score") AS avg_reading, AVG("writing score") AS avg_writing
    FROM students_performance
    GROUP BY "test preparation course";

'''
'Parental education influence': '''
    SELECT "parental level of education", AVG("math score") AS avg_math,
    ↪AVG("reading score") AS avg_reading, AVG("writing score") AS avg_writing
    FROM students_performance
    GROUP BY "parental level of education"
    ORDER BY avg_math DESC;

'''
'Lunch type analysis': '''
    SELECT lunch, AVG("math score") AS avg_math, AVG("reading score") AS
    ↪avg_reading, AVG("writing score") AS avg_writing
    FROM students_performance
    GROUP BY lunch;

'''
'High achievers by ethnicity': '''
    SELECT "race/ethnicity", COUNT(*) AS num_high_achievers
    FROM students_performance
    WHERE "math score" > 80 AND "reading score" > 80 AND "writing score" >
    ↪80
    GROUP BY "race/ethnicity"
    ORDER BY num_high_achievers DESC;

'''
}

```

```

[10]: for desc, sql in analysis_queries.items():
        print(f'-- {desc} --')
        query_and_print(sql)

conn.close()

```

```

-- Average scores by gender --
gender  avg_math  avg_reading  avg_writing
0 female  63.633205   72.608108   72.467181
1  male   68.728216   65.473029   63.311203

-- Impact of test preparation --
test preparation course  avg_math  avg_reading  avg_writing
0          completed  69.695531   73.893855   74.418994
1              none  64.077882   66.534268   64.504673

```

```
-- Parental education influence --
parental level of education  avg_math  avg_reading  avg_writing
0      master's degree  69.745763    75.372881    75.677966
1      bachelor's degree  69.389831    73.000000    73.381356
2      associate's degree  67.882883    70.927928    69.896396
3      some college  67.128319    69.460177    68.840708
4      some high school  63.497207    66.938547    64.888268
5      high school  62.137755    64.704082    62.448980

-- Lunch type analysis --
      lunch  avg_math  avg_reading  avg_writing
0 free/reduced  58.921127    64.653521    63.022535
1      standard  70.034109    71.654264    70.823256

-- High achievers by ethnicity --
race/ethnicity  num_high_achievers
0      group C                34
1      group E                29
2      group D                26
3      group B                14
4      group A                 7
```

1

Strategies for Improvement

1. Encourage test preparation courses, as these have a strong positive impact.
2. Develop support programs for students with less educated parents.
3. Address lunch and nutrition disparities, as these impact attention and performance.
4. Implement targeted interventions for lower-performing demographic groups to close achievement gaps

[]: