Shamiri Assessments Insights and Conclusions

Kevien Otieno Mulambe

2025-02-22

```python
# Importing the required libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

```python
# Loading the data
df = pd.read_excel("Data.xlsx")

# preview of the dataset
df.head
```

```
## <bound method NDFrame.head of          School  ...                      Category
## 0     Nomantown  ...                     Nomantown
## 1    Neighbor A  ...               Close neighbors
## 2    Neighbor B  ...               Close neighbors
## 3    Neighbor C  ...               Close neighbors
## 4   Developed A  ...   Developed African economies
## 5   Developed B  ...   Developed African economies
## 6   Developed C  ...   Developed African economies
## 7     Similar A  ...             Similar economies
## 8     Similar B  ...             Similar economies
## 9     Similar C  ...             Similar economies
##
## [10 rows x 6 columns]>
```

Data Cleaning

Missing values

```python
# Lets check for missing values
df.isnull().sum()
```

```
## School                                            0
## Students per school                               0
## Students per teacher                              0
## Government spending on education per student (US $)   0
## Average international assessment score            0
## Category                                          0
## dtype: int64
```

Data types

```
df.info()
```

```
## <class 'pandas.core.frame.DataFrame'>
## RangeIndex: 10 entries, 0 to 9
## Data columns (total 6 columns):
##  #   Column                                              Non-Null Count  Dtype
## ---  ------                                              --------------  -----
##  0   School                                              10 non-null     object
##  1   Students per school                                 10 non-null     object
##  2   Students per teacher                                10 non-null     object
##  3   Government spending on education per student (US $) 10 non-null     object
##  4   Average international assessment score              10 non-null     object
##  5   Category                                            10 non-null     object
## dtypes: object(6)
## memory usage: 612.0+ bytes
```

```
# Converting some column datatypes to int
df["Students per school"] = df["Students per school"].astype(int)
df["Students per teacher"] = df["Students per teacher"].astype(int)
df["Government spending on education per student (US $)"] = df["Government spending on education per st
df["Average international assessment score"] = df["Average international assessment score"].astype(int)
df["School"] = df["School"].astype(str)

# check
df.info()
```

```
## <class 'pandas.core.frame.DataFrame'>
## RangeIndex: 10 entries, 0 to 9
## Data columns (total 6 columns):
##  #   Column                                              Non-Null Count  Dtype
## ---  ------                                              --------------  -----
##  0   School                                              10 non-null     object
##  1   Students per school                                 10 non-null     int64
##  2   Students per teacher                                10 non-null     int64
##  3   Government spending on education per student (US $) 10 non-null     int64
##  4   Average international assessment score              10 non-null     int64
##  5   Category                                            10 non-null     object
## dtypes: int64(4), object(2)
## memory usage: 612.0+ bytes
```

```
# Renaming columns
df.rename(columns={
    "School": "School",
    "Government spending on education per student (US $)": "Spending per Student",
    "Average international assessment score": "Assessment Score"
}, inplace=True)
```

Exploratory Data Analysis

Spending per Student vs. Assessment Score

```python
plt.figure(figsize=(12,6))
sns.scatterplot(x=df["Spending per Student"], y=df["Assessment Score"], hue=df["Category"], palette=["Re
sns.regplot(data=df, x="Spending per Student", y="Assessment Score", scatter=False, color='red')
plt.title("Spending per Student vs. Assessment Score")
plt.xlabel("Spending per Student (USD)")
plt.ylabel("Assessment Score")
plt.grid(True)
plt.show()
```
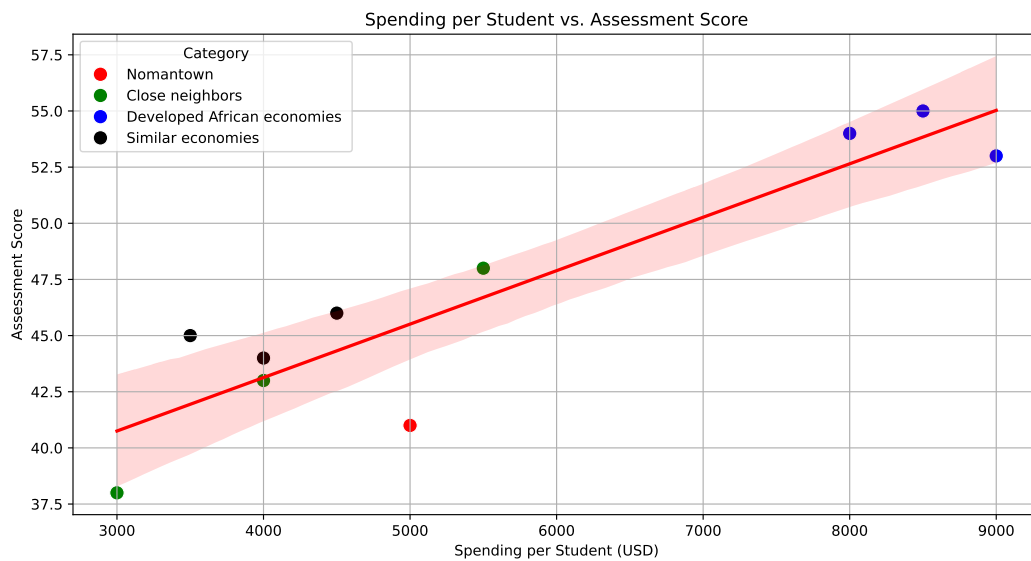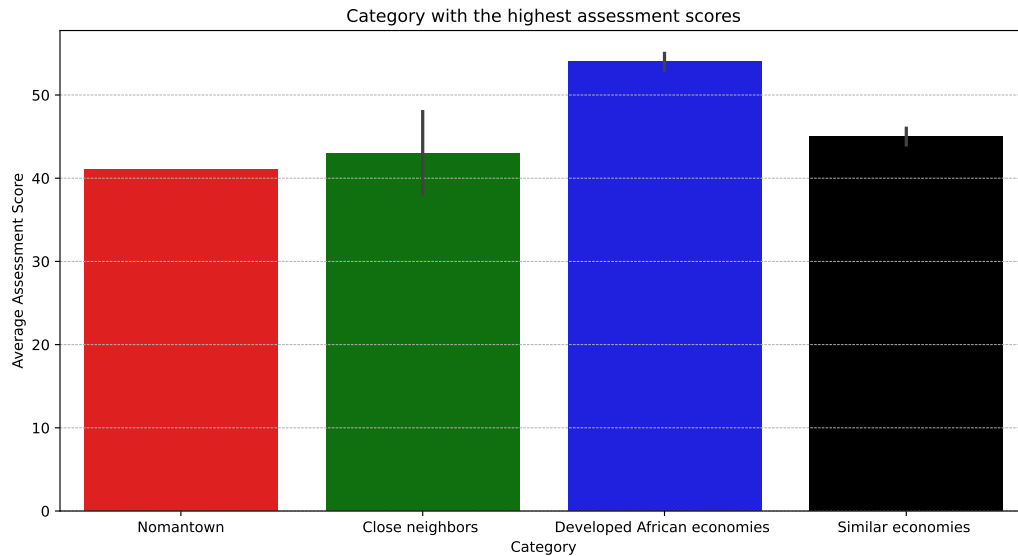

Insights

- From the graph, we can see the higher the **spending per student(USD)**, the higher the **assessment score** which can only be seen among the **Developed African Economies**.

```python
# Extracting the needed columns
df_spending_assessment = df[["Category", "Spending per Student", "Assessment Score", "Students per scho

# Grouping by Category
df_spending_assessment = df_spending_assessment.groupby("Category").mean().reset_index().sort_values("As

# Visualization
plt.figure(figsize = (12, 6))
sns.barplot(x= df["Category"], y= df["Assessment Score"], hue = df["Category"], palette=["Red", "Green"
plt.title("Category with the highest assessment scores")
plt.xlabel("Category")
plt.ylabel("Average Assessment Score")
```

```
plt.grid(axis = "y", linewidth = 0.5, linestyle="--")
plt.show()
```


Category with the highest assessment scores

Insights

- From the graph, we can see that **Developed African Economies** have on average the highest **assessment scores**.

Spending per Student and Assessment Scores

```
# Extracting columns needed
df_melt = df[["School", "Spending per Student", "Assessment Score"]]

# Convert to long format
df_melted = df_melt.melt(id_vars=["School"], var_name="Metric", value_name="Value")

plt.figure(figsize=(14, 7))
sns.barplot(data=df_melted, x="School", y="Value", hue="Metric", palette=["Blue", "green"], dodge=True)
plt.title("Comparison of Spending per Student and Assessment Scores")
plt.xticks(rotation=45, ha="right")
```

```
## ([0, 1, 2, 3, 4, 5, 6, 7, 8, 9], [Text(0, 0, 'Nomantown'), Text(1, 0, 'Neighbor A'), Text(2, 0, 'Nei
```

```
plt.ylabel("Value")
plt.grid(axis="y")
plt.show()
```

Comparison of Spending per Student and Assessment Scores
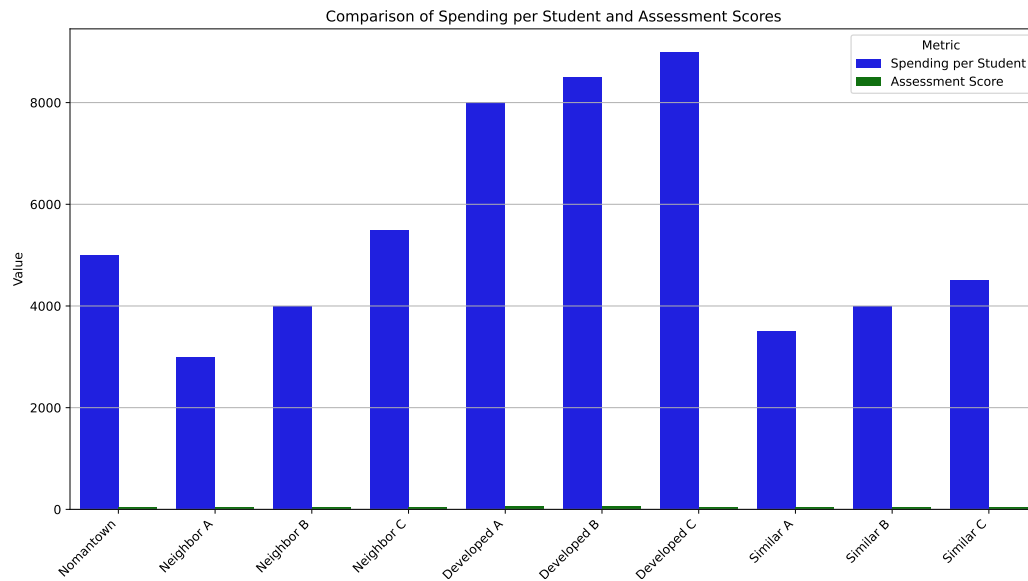
Insights

- From the graph, we can see that School **Developed C** receives the highest level of funding of **9000** which can also be reflected by its high **assessment score** of **53**.
- Similarly School **Neighbor A** receives the lowest level of funding of **3000** which is refelcted with an **assessment score** of **38**.

Student-Teacher Ratio vs. Assessment Score

```python
plt.figure(figsize=(12,6))
sns.lineplot(data=df, x="Students per teacher", y="Assessment Score", marker="o", color="green")
plt.title("Student-Teacher Ratio vs. Assessment Score")
plt.xlabel("Students per Teacher")
plt.ylabel("Assessment Score")
plt.grid(True)
plt.show()
```

Student-Teacher Ratio vs. Assessment Score
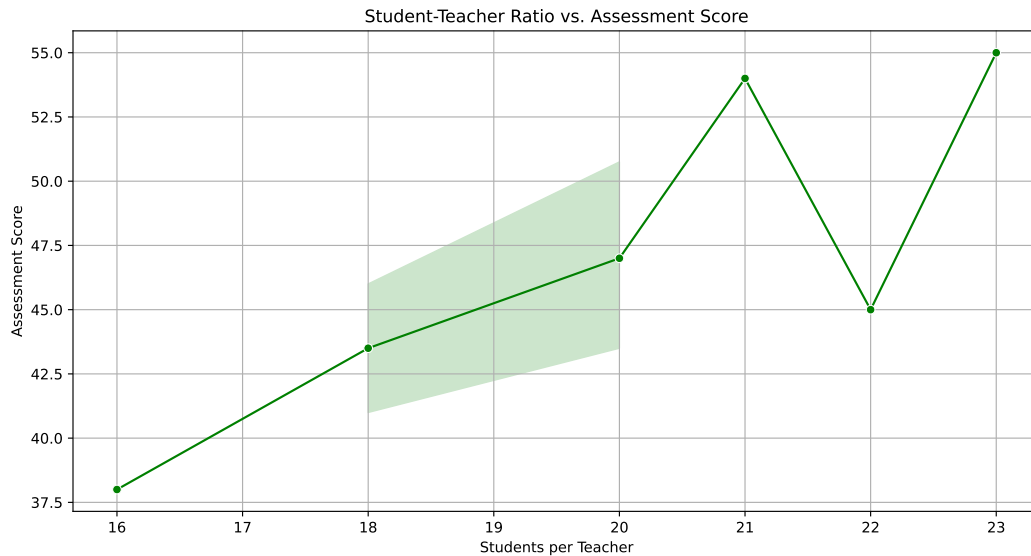
Insights

- From the graph, we can see the higher the **Student to teacher** the higher the **assessment score**, meaning the assessment score is not determined by only Government funding but also Student to teacher ratio.

Advanced Analysis

Correlation Analysis

```python
# Extracting columns needed
correlation_matrix = df[["Students per school", "Students per teacher", "Spending per Student", "Assessm

# calculating the correlation matrix
correlation_matrix = correlation_matrix.corr()

# Visualization
plt.figure(figsize = (14, 7))
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm", fmt=".2f", linewidths=0.5)
plt.title("Correlation Heatmap")
plt.show()
```

Correlation Heatmap
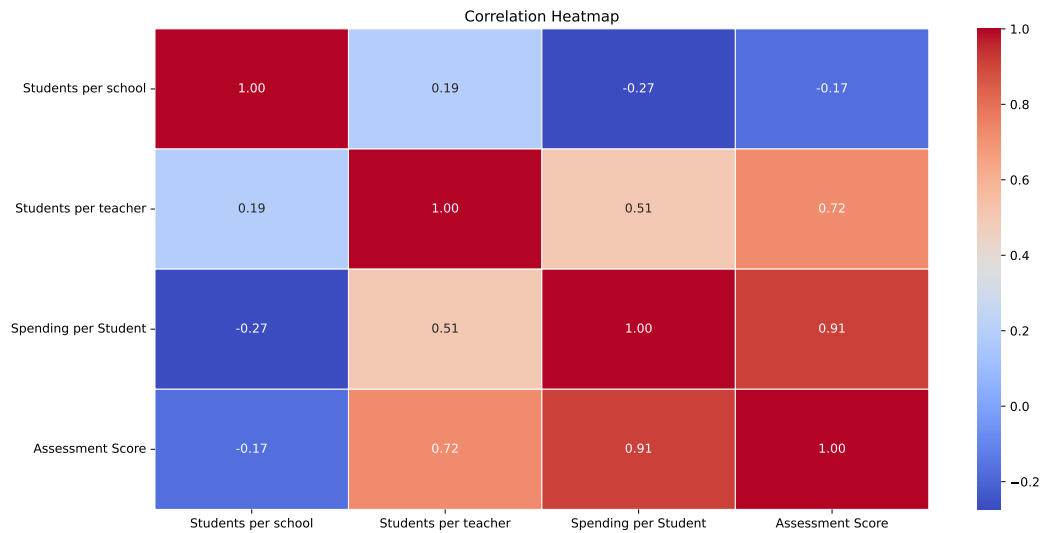
Insights

- Higher spending correlates with higher assessment scores
- Developed African economies receive the highest funding with an average of **8500** and also having the highest assessment scores with an average of **54**.
- The higher the students per teacher the higher the assessment score, this can be seen from the Developed African economies schools and Similar economies.
- Although government funding has an impact on the assessment score it is not completely since schools in similar economies received on average the lowest government funding but still has the second best average assessment score of **45** which might have been due to a high students per teacher.
- Nomantown's spending is moderate, but its assessment scores are relatively low, suggesting potential inefficiencies

Conclusions

- **Higher Spending Generally Leads to Higher Assessment Scores:** There is a positive correlation between government spending per student and assessment performance, indicating that increased funding often translates to better academic outcomes.
- **Developed African Economies Receive the Highest Funding and Achieve the Best Scores:** Schools in Developed African economies receive the highest average funding of **$8,500** per student and also achieve the highest average assessment score of **54**, reinforcing the link between investment in education and performance.
- **Higher Student-to-Teacher Ratios Are Associated with Higher Scores:** Contrary to common assumptions, a higher students-per-teacher ratio appears to be linked to better assessment scores. This is evident in Developed African economies and Similar economies, suggesting that teacher quality, instructional methods, or peer-driven learning may play a significant role in student success.
- **Government Funding Is Not the Only Determining Factor for Performance:** Schools in Similar economies receive the lowest average government funding but still achieve the second-highest assessment score **45**. This suggests that other factors—such as teacher effectiveness, school management, or alternative funding sources—contribute to educational outcomes.