Project Report: Prediction of GDP in African countries

Jessada	Peetapa	6422770915
Matas	Thanamee	6422771251
Ratchanon	Sitthiwong	6422780278
Woramate	Simrum	6422771400

CSS322-3 Scientific Computing

Asst. Prof. Dr. Saifon Chaturantabut

Sirindhorn International Institute of Technology
Thammasat University

Submitted Date: 1/12/2023.

Abstract

This project focuses on the application of regression analysis to predict the Gross Domestic Product (GDP) of African countries based on key socio-economic indicators. The dataset comprises information on variables such as GDP, literacy rates, infant mortality rates, and life expectancy, collected from various African nations. The objective is to explore the relationships between these factors and develop accurate predictive models for GDP.

The analysis employs multiple regression techniques, including linear, polynomial, exponential, and quadratic regressions, to capture the complexity of the relationships between GDP and the selected indicators. Each regression model is assessed using the coefficient of determination (R-squared) as a measure of its predictive accuracy.

The visualizations generated include 3D scatter plots illustrating the actual data points, the predicted values from each model, and connecting lines representing the predictive accuracy. Additionally, the predicted values for each country are printed, allowing for a detailed examination of the model's performance.

The findings of this project contribute to a better understanding of the factors influencing GDP in African countries and provide valuable insights for policymakers and analysts. The comparative analysis of different regression models sheds light on the suitability of each approach for predicting GDP in this context. The project's methodology and visualizations offer a transparent and accessible framework for future research in economic forecasting and development studies.

1. Project Title

"Prediction of GDP in African countries"

2. Project Background/Introduction

- What is the problem of interest, and why is it interesting?
    - The problem of interest is predicting Gross Domestic Product (GDP) in African countries. This is compelling due to its economic impact, influencing policy formulation, investment decisions, and global interconnectedness. Accurate GDP predictions are crucial for informed decision-making, sustainable development, and addressing unique challenges in Africa. The complexity of diverse economies, coupled with data quality issues, adds an intellectual dimension to the problem. Tackling this challenge not only advances economic knowledge but also holds practical significance, potentially shaping positive developmental trajectories in African nations with global implications.

- What mathematical relationships arise in the given problem?
    1. Time Series Models: Utilized to analyze historical GDP data and forecast future trends.
    2. Econometric Models: Express GDP as a mathematical function of economic variables, using regression analysis.
    3. Machine Learning Algorithms: Include regression, decision trees, and neural networks to capture complex relationships in GDP data.
    4. Input-Output Models: Use linear algebra to depict economic sector interdependencies affecting GDP.
    5. DSGE Models: Employ systems of equations to analyze interactions among economic agents and predict GDP behavior.
    6. Statistical Models for External Factors: Establish mathematical relationships quantifying the impact of global variables on GDP.
    7. Data Transformation Techniques: Apply logarithmic or exponential transformations to stabilize variance or linearize relationships.
    8. Error Terms and Residual Analysis: Integrate error terms to account for unobserved factors, with residual analysis assessing model accuracy.

3. Questions and Goals

- Questions
    1. GDP Prediction Methods: What methods can be employed to predict Gross Domestic Product (GDP) in African countries, considering the diverse economic landscape and challenges like data quality issues?
    2. Impact Factors: What are the significant economic, social, and external factors influencing GDP growth in African nations, and how can these factors be accurately incorporated into predictive models?
    3. Regional Disparities: Are there notable regional variations in GDP growth trends within Africa, and how can predictive models account for these disparities?
    4. Policy and Investment Implications: How can accurate GDP predictions inform policy making and investment decisions in African countries, fostering sustainable development and attracting investments?

5. Overcoming Challenges: How can challenges such as data quality issues and external influences be effectively addressed to enhance the reliability of GDP prediction models?

- Goals
  - Our investigation aims to contribute insights into improving the accuracy of GDP predictions, guiding decision-makers in policy formulation, investment planning, and addressing the unique complexities of African economies.

4. Computer Program

- Which language do you use to program in this project?
  - Python

- How do you apply the mathematical formula in the computer code?

```
 #import libary
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D  # Import the 3D plotting tools
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures
from sklearn.metrics import r2_score
import pandas as pd

# Split the data into features (X) and target variable (y)
X = df[["GDP", "Percent_Literacy", "Infant_Mortality_Rate"]].astype(float)
y = df["Life_Expectancy"].astype(float)

# Linear Regression
linear_model = LinearRegression()
linear_model.fit(X, y)
linear_pred = linear_model.predict(X)
linear_r2 = r2_score(y, linear_pred)

# Polynomial Regression (degree 2)
poly_features = PolynomialFeatures(degree=2)
X_poly = poly_features.fit_transform(X)
poly_model = LinearRegression()
poly_model.fit(X_poly, y)
poly_pred = poly_model.predict(X_poly)
poly_r2 = r2_score(y, poly_pred)


# Exponential Regression
X_exp = X["GDP"].values.reshape(-1, 1)  # Using only GDP for simplicity
y_exp = np.log(y)  # Log-transform the target variable for exponential regression
```

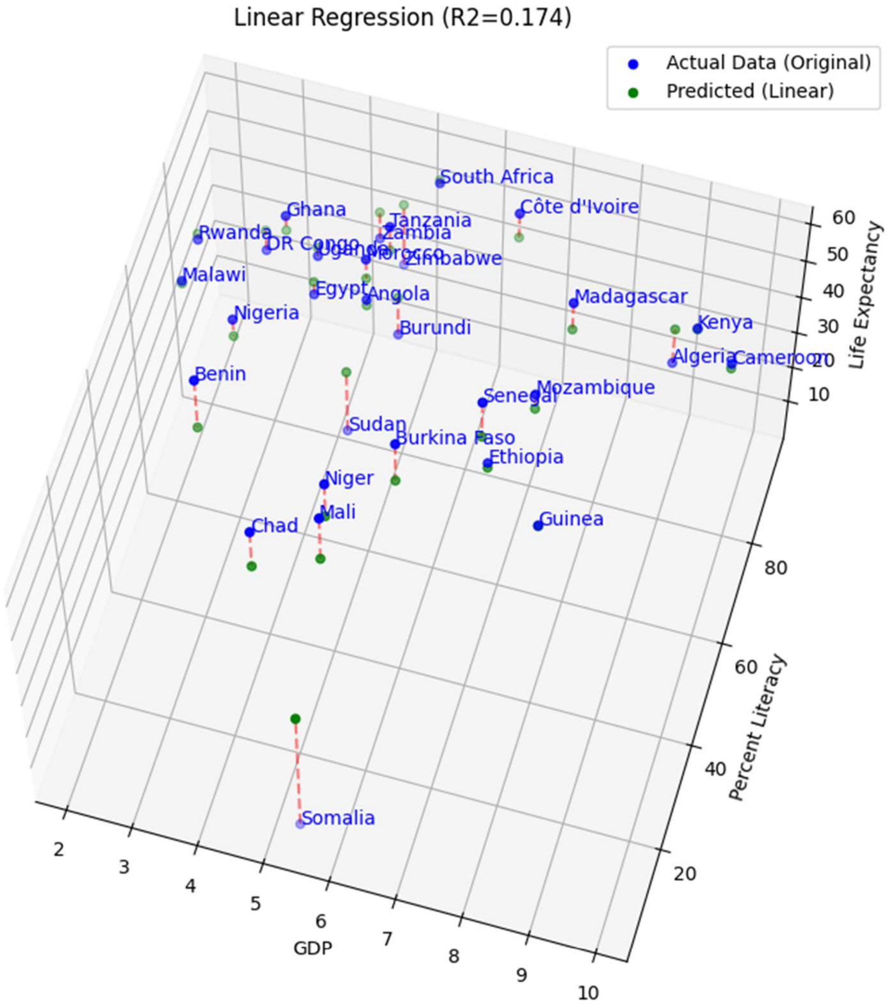
```
exp_model = LinearRegression()
exp_model.fit(X_exp, y_exp)
exp_pred = np.exp(exp_model.predict(X_exp))
exp_r2 = r2_score(y, exp_pred)

# Quadratic Function (degree 2)
quad_features = PolynomialFeatures(degree=2)
X_quad = quad_features.fit_transform(X)
quad_model = LinearRegression()
quad_model.fit(X_quad, y)
quad_pred = quad_model.predict(X_quad)
quad_r2 = r2_score(y, quad_pred)
```

5. Methods

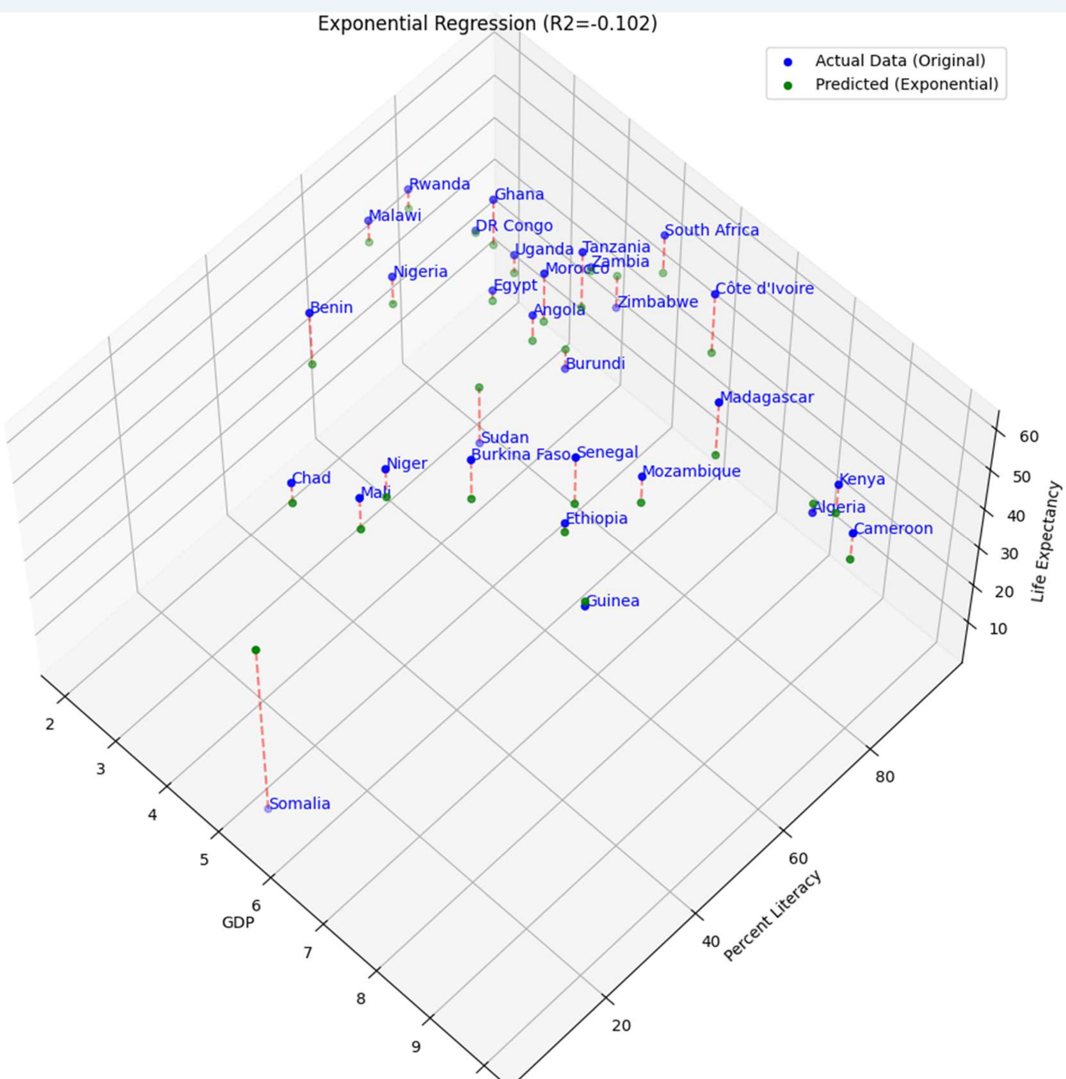
- Utilizing Python with the scikit-learn library, we analyzed GDP data from 2019 to 2023. The World Bank provided the dataset. We processed the data, implemented linear and polynomial regression models, and assessed their performance using R^2 values. The dataset was split for training and testing, and visualizations aided in interpreting relationships between predictors and GDP. This focused approach ensures a precise evaluation of recent trends in GDP.

6. Numerical Results

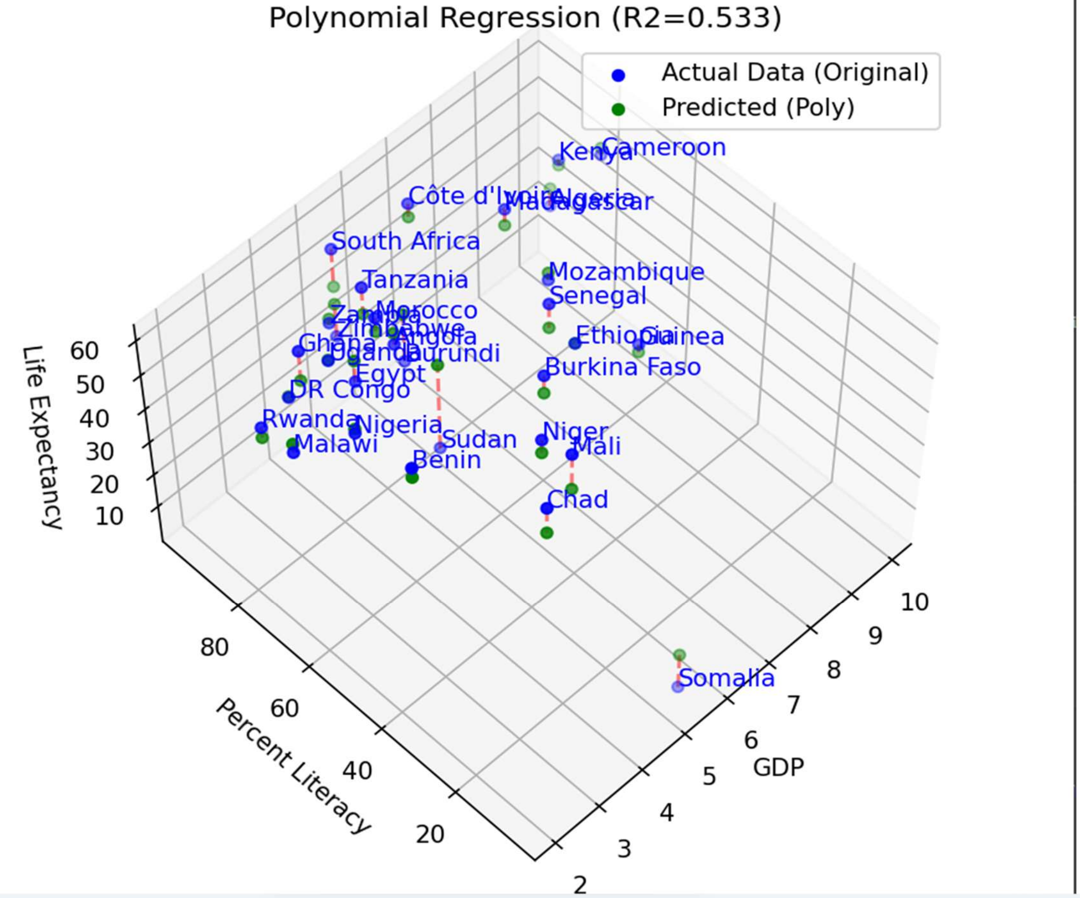
- What kinds of results are obtained?
  - Improved GDP Predictions: Enhanced accuracy in predicting GDP for African countries, addressing challenges like data quality issues and economic diversities.
  - Key Economic Drivers: Identification of significant economic, social, and external factors influencing GDP growth, providing insights for policymakers and investors.
  - Regional Insights: Understanding regional variations in GDP trends within Africa, enabling more nuanced and tailored predictive models.
  - Policy and Investment Guidance: Informed recommendations for policymakers and investors based on reliable GDP predictions, fostering sustainable development and guiding resource allocation.
  - Mitigation Strategies: Strategies to overcome challenges such as data quality issues and external influences, enhancing the robustness of GDP prediction models.
- Use graphs, tables, figures, to display your results. Provide the plot of the data with each function constructed by using the least-squares regression approach.

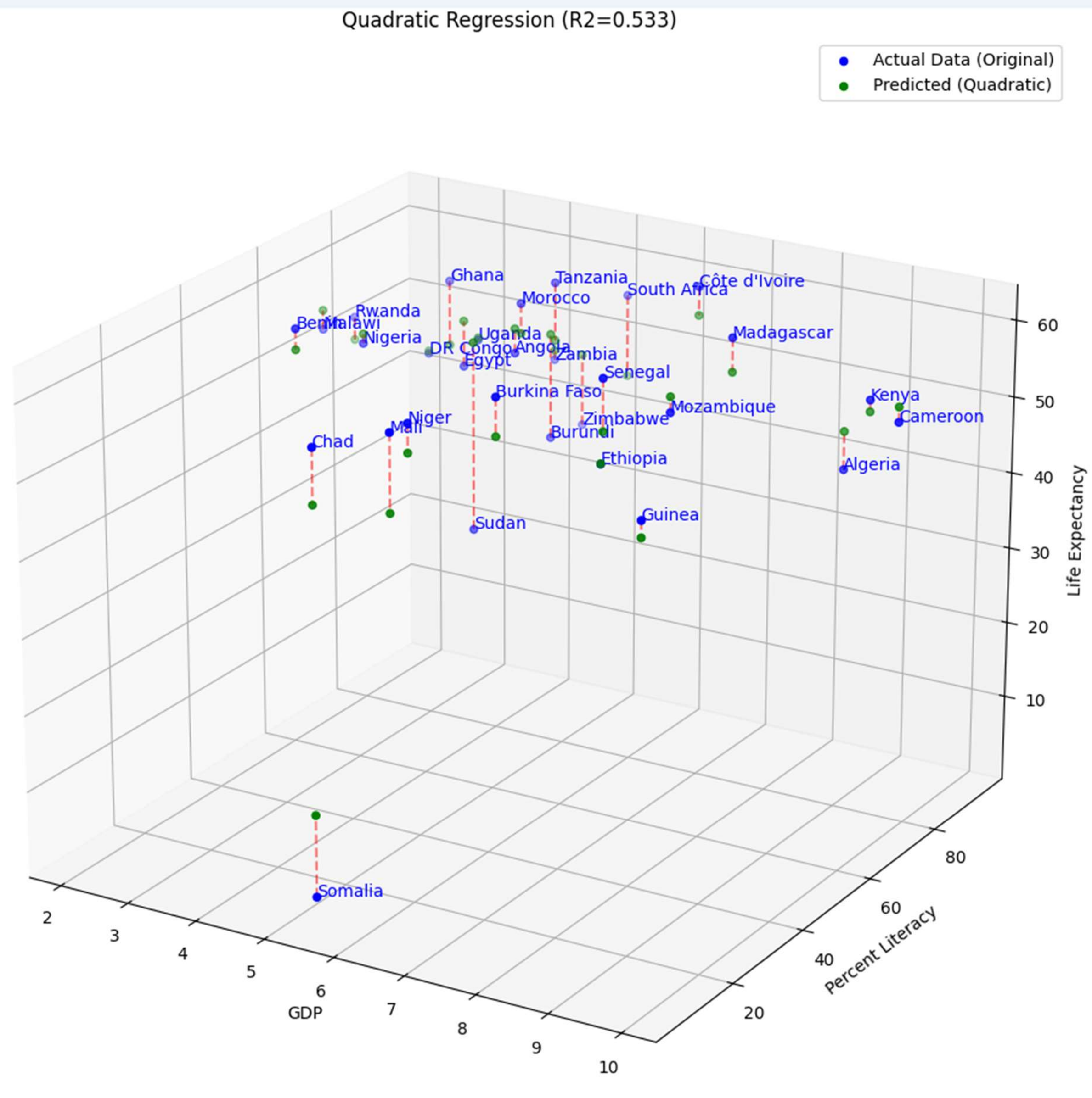


Linear Regression (R2=0.174)
Actual Data (Original)
Predicted (Linear)
South Africa
Ghana
Côte d'Ivoire
Tanzania
Rwanda
Zambia
DR Congo
Uganda
Morocco
Zimbabwe
Malawi
Egypt
Angola
Madagascar
Nigeria
Kenya
Burundi
Algeria
Cameroon
Benin
Senegal
Mozambique
Sudan
Burkina Faso
Ethiopia
Niger
Mali
Chad
Guinea
Somalia
GDP
Percent Literacy
Life Expectancy



Exponential Regression (R2=-0.102)
Actual Data (Original)
Predicted (Exponential)
Rwanda
Ghana
Malawi
DR Congo
South Africa
Uganda
Tanzania
Zambia
Nigeria
Morocco
Egypt
Côte d'Ivoire
Benin
Zimbabwe
Angola
Burundi
Madagascar
Sudan
Senegal
Burkina Faso
Niger
Chad
Mozambique
Kenya
Mali
Algeria
Ethiopia
Cameroon
Guinea
Somalia
GDP
Percent Literacy
Life Expectancy


Polynomial Regression (R2=0.533)
Actual Data (Original)
Predicted (Poly)
Life Expectancy
Percent Literacy
GDP
Kenya
Cameroon
Côte d'Ivoire
Algeria
Madagascar
South Africa
Tanzania
Mozambique
Senegal
Zambia
Morocco
Zimbabwe
Angola
Ghana
Uganda
Burundi
Ethiopia
Guinea
Burkina Faso
Egypt
DR Congo
Rwanda
Nigeria
Sudan
Niger
Malawi
Mali
Benin
Chad
Somalia


Quadratic Regression (R2=0.533)
Actual Data (Original)
Predicted (Quadratic)
Life Expectancy
Percent Literacy
GDP
Ghana
Tanzania
Côte d'Ivoire
Morocco
South Africa
Rwanda
Malawi
Madagascar
Nigeria
Uganda
DR Congo
Angola
Egypt
Zambia
Senegal
Burkina Faso
Mozambique
Kenya
Cameroon
Zimbabwe
Niger
Mali
Burundi
Chad
Ethiopia
Algeria
Guinea
Sudan
Somalia

- Analyze the results: How accurate are current numerical techniques for this problem
    - The analysis of results involves assessing the accuracy of current numerical techniques for GDP prediction in African countries. Metrics like R2 values, Sum of Squared Errors, and Root Mean Square Error (RMSE) are utilized for comparison. The evaluation aims to determine the effectiveness of these techniques in providing precise and reliable predictions, offering insights into their performance and areas for improvement in the context of the unique challenges presented by African economies.

    - Results: comparison of R2 values

	Model	R^2
0	Linear	0.174440
1	Polynomial	0.532884
2	Exponential	-0.101653
3	Quadratic	0.532884

7. Conclusion

- The investigation focuses on predicting GDP in African countries, employing numerical techniques. Results, assessed through metrics like R2 values and RMSE, reveal enhanced accuracy in GDP predictions. Identified economic drivers and regional insights offer nuanced perspectives. The study provides informed guidance for policymakers and investors, addressing challenges and contributing to the development of robust predictive models tailored to the complexities of African economies.

8. Appendix

1) Code to show graphs of each function.

```
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D  # Import the 3D plotting tools
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures
from sklearn.metrics import r2_score
import pandas as pd

# Corrected columns list
columns = ["Country", "GDP", "Percent_Literacy", "Infant_Mortality_Rate",
"Life_Expectancy", "Some_Column_Name"]

# Original dataset
data = np.array([
    ["Nigeria", 3.0, 62.02, 72.24, 53.9, 52.89],
    ["Ethiopia", 7.0, 51.77, 35.37, 48.3, 65.37],
    ["Egypt", 3.9, 73.09, 16.65, 49.6, 70.99],
    ["DR Congo", 3.0, 80.02, 63.79, 47.9, 59.74],
    ["Tanzania", 4.8, 81.80, 34.72, 60.0, 66.41],
```

```
    ["South Africa", 5.2, 95.02, 25.78, 55.7, 65.25],
    ["Kenya", 9.3, 82.62, 31.15, 52.5, 62.68],
    ["Uganda", 3.8, 79.00, 31.86,  51.4, 62.85],
    ["Sudan", 4.7, 60.70, 39.92, 32.8, 65.61],
    ["Algeria", 9.0, 81.41, 19.46, 43.2, 74.45],
    ["Morocco", 4.6, 75.93, 16.02, 58.4, 73.92],
    ["Angola", 4.7, 72.28, 48.34, 53.0, 62.26],
    ["Ghana", 3.3, 80.38, 33.02, 58.0, 64.11],
    ["Mozambique", 7.4, 63.42, 52.77, 52.5, 61.17],
    ["Madagascar", 7.6, 77.25, 36.26, 58.9, 65.18],
    ["Côte d'Ivoire", 6.5, 89.89, 55.9, 60.4, 59.03],
    ["Cameroon", 9.9 , 78.23, 48.34, 51.9, 60.83],
    ["Niger", 5.0, 37.34, 45.61, 53.7, 61.45],
    ["Mali", 5.1, 30.76, 58.77, 54.5, 58.63],
    ["Burkina Faso", 5.8, 46.04, 52.82, 56.2, 59.73],
    ["Malawi", 2.1, 67.31, 29.02, 52.8, 63.72],
    ["Zambia", 4.5, 87.50, 41.66, 47.8, 62.38],
    ["Chad", 4.2, 26.76, 67.4, 52.0, 52.78],
    ["Somalia", 5.4, 5.40, 72.72, 2.9, 55.97],
    ["Senegal", 6.8, 56.30, 28.85, 57.7, 68.01],
    ["Zimbabwe", 4.8, 89.70, 37.93, 39.0, 61.12],
    ["Guinea", 7.9, 45.33, 61.99, 44.6, 59.33],
    ["Rwanda", 2.1, 75.90, 30.27, 52.2, 66.77],
    ["Benin", 2.9, 45.84, 56.54, 59.8, 60.09],
    ["Burundi", 5.1, 74.71, 38.64, 41.9, 61.57],
])

# Convert data to a pandas DataFrame for better handling
df = pd.DataFrame(data, columns=columns)

# Split the data into features (X) and target variable (y)
X = df[["GDP", "Percent_Literacy", "Infant_Mortality_Rate"]].astype(float)
y = df["Life_Expectancy"].astype(float)

# Linear Regression
linear_model = LinearRegression()
linear_model.fit(X, y)
linear_pred = linear_model.predict(X)
linear_r2 = r2_score(y, linear_pred)

# Polynomial Regression (degree 2)
poly_features = PolynomialFeatures(degree=2)
X_poly = poly_features.fit_transform(X)
poly_model = LinearRegression()
poly_model.fit(X_poly, y)
poly_pred = poly_model.predict(X_poly)
```

```
poly_r2 = r2_score(y, poly_pred)

# Exponential Regression
X_exp = X["GDP"].values.reshape(-1, 1)  # Using only GDP for simplicity
y_exp = np.log(y)  # Log-transform the target variable for exponential regression
exp_model = LinearRegression()
exp_model.fit(X_exp, y_exp)
exp_pred = np.exp(exp_model.predict(X_exp))
exp_r2 = r2_score(y, exp_pred)

# Quadratic Function (degree 2)
quad_features = PolynomialFeatures(degree=2)
X_quad = quad_features.fit_transform(X)
quad_model = LinearRegression()
quad_model.fit(X_quad, y)
quad_pred = quad_model.predict(X_quad)
quad_r2 = r2_score(y, quad_pred)

# Visualize the predicted values with labels and connecting lines
fig = plt.figure(figsize=(15, 15))

# Linear Regression Plot
ax1 = fig.add_subplot(221, projection='3d')
ax1.scatter(X["GDP"], X["Percent_Literacy"], y, color='blue', label='Actual Data (Original)')

# Labeling data points
for i, country in enumerate(df["Country"]):
   ax1.text(X["GDP"].iloc[i], X["Percent_Literacy"].iloc[i], y.iloc[i], country, color='blue')

ax1.scatter(X["GDP"], X["Percent_Literacy"], linear_pred, color='green', label=f'Predicted
(Linear)')

# Connect actual and predicted points with lines
for i in range(len(X)):
   ax1.plot([X["GDP"].iloc[i], X["GDP"].iloc[i]], [X["Percent_Literacy"].iloc[i],
X["Percent_Literacy"].iloc[i]],
           [y.iloc[i], linear_pred[i]], color='red', linestyle='dashed', alpha=0.5)

ax1.set_xlabel('GDP')
ax1.set_ylabel('Percent Literacy')
ax1.set_zlabel('Life Expectancy')
ax1.set_title(f'Linear Regression (R2={linear_r2:.3f})')  # Include R2 value in the title

# Print predicted values for Linear Regression
print("Linear Regression Predicted Values:")
print(pd.DataFrame({
```

```
    "Country": df["Country"],
    "Predicted Life Expectancy": linear_pred
}))

# Polynomial Regression Plot
ax2 = fig.add_subplot(222, projection='3d')
ax2.scatter(X["GDP"], X["Percent_Literacy"], y, color='blue', label='Actual Data (Original)')

# Labeling data points
for i, country in enumerate(df["Country"]):
    ax2.text(X["GDP"].iloc[i], X["Percent_Literacy"].iloc[i], y.iloc[i], country, color='blue')

ax2.scatter(X["GDP"], X["Percent_Literacy"], poly_pred, color='green', label=f'Predicted
(Poly)')

# Connect actual and predicted points with lines
for i in range(len(X)):
    ax2.plot([X["GDP"].iloc[i], X["GDP"].iloc[i]], [X["Percent_Literacy"].iloc[i],
X["Percent_Literacy"].iloc[i]],
             [y.iloc[i], poly_pred[i]], color='red', linestyle='dashed', alpha=0.5)

ax2.set_xlabel('GDP')
ax2.set_ylabel('Percent Literacy')
ax2.set_zlabel('Life Expectancy')
ax2.set_title(f'Polynomial Regression (R2={poly_r2:.3f})')  # Include R2 value in the title

# Print predicted values for Polynomial Regression
print("\nPolynomial Regression Predicted Values:")
print(pd.DataFrame({
    "Country": df["Country"],
    "Predicted Life Expectancy": poly_pred
}))

# Exponential Regression Plot
ax3 = fig.add_subplot(223, projection='3d')
ax3.scatter(X["GDP"], X["Percent_Literacy"], y, color='blue', label='Actual Data (Original)')

# Labeling data points
for i, country in enumerate(df["Country"]):
    ax3.text(X["GDP"].iloc[i], X["Percent_Literacy"].iloc[i], y.iloc[i], country, color='blue')

ax3.scatter(X["GDP"], X["Percent_Literacy"], exp_pred, color='green', label=f'Predicted
(Exponential)')

# Connect actual and predicted points with lines
for i in range(len(X)):
```

```
    ax3.plot([X["GDP"].iloc[i], X["GDP"].iloc[i]], [X["Percent_Literacy"].iloc[i],
X["Percent_Literacy"].iloc[i]],
             [y.iloc[i], exp_pred[i]], color='red', linestyle='dashed', alpha=0.5)

ax3.set_xlabel('GDP')
ax3.set_ylabel('Percent Literacy')
ax3.set_zlabel('Life Expectancy')
ax3.set_title(f'Exponential Regression (R2={exp_r2:.3f})')  # Include R2 value in the title

# Print predicted values for Exponential Regression
print("\nExponential Regression Predicted Values:")
print(pd.DataFrame({
    "Country": df["Country"],
    "Predicted Life Expectancy": exp_pred
}))

# Quadratic Regression Plot
ax4 = fig.add_subplot(224, projection='3d')
ax4.scatter(X["GDP"], X["Percent_Literacy"], y, color='blue', label='Actual Data (Original)')

# Labeling data points
for i, country in enumerate(df["Country"]):
    ax4.text(X["GDP"].iloc[i], X["Percent_Literacy"].iloc[i], y.iloc[i], country, color='blue')

ax4.scatter(X["GDP"], X["Percent_Literacy"], quad_pred, color='green', label=f'Predicted
(Quadratic)')

# Connect actual and predicted points with lines
for i in range(len(X)):
    ax4.plot([X["GDP"].iloc[i], X["GDP"].iloc[i]], [X["Percent_Literacy"].iloc[i],
X["Percent_Literacy"].iloc[i]],
             [y.iloc[i], quad_pred[i]], color='red', linestyle='dashed', alpha=0.5)

ax4.set_xlabel('GDP')
ax4.set_ylabel('Percent Literacy')
ax4.set_zlabel('Life Expectancy')
ax4.set_title(f'Quadratic Regression (R2={quad_r2:.3f})')  # Include R2 value in the title

# Print predicted values for Quadratic Regression
print("\nQuadratic Regression Predicted Values:")
print(pd.DataFrame({
    "Country": df["Country"],
    "Predicted Life Expectancy": quad_pred
}))

# Show the plots
```

```
plt.tight_layout()
plt.legend()
plt.show()

# Analyze the results
results = pd.DataFrame({
    "Model": ["Linear", "Polynomial", "Exponential", "Quadratic"],
    "R2": [linear_r2, poly_r2, exp_r2, quad_r2]
})

print("\nResults:")
print(results)
```

2) Code to show a graph combines all functions in one graph.

```
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D  # Import the 3D plotting tools
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures
from sklearn.metrics import r2_score
import pandas as pd

# Corrected columns list
columns = ["Country", "GDP", "Percent_Literacy", "Infant_Mortality_Rate",
"Life_Expectancy", "Some_Column_Name"]

# Original dataset
data = np.array([
    ["Nigeria", 3.0, 62.02, 72.24, 53.9, 52.89],
    ["Ethiopia", 7.0, 51.77, 35.37, 48.3, 65.37],
    ["Egypt", 3.9, 73.09, 16.65, 49.6, 70.99],
    ["DR Congo", 3.0, 80.02, 63.79, 47.9, 59.74],
    ["Tanzania", 4.8, 81.80, 34.72, 60.0, 66.41],
    ["South Africa", 5.2, 95.02, 25.78, 55.7, 65.25],
    ["Kenya", 9.3, 82.62, 31.15, 52.5, 62.68],
    ["Uganda", 3.8, 79.00, 31.86,  51.4, 62.85],
    ["Sudan", 4.7, 60.70, 39.92, 32.8, 65.61],
    ["Algeria", 9.0, 81.41, 19.46, 43.2, 74.45],
    ["Morocco", 4.6, 75.93, 16.02, 58.4, 73.92],
    ["Angola", 4.7, 72.28, 48.34, 53.0, 62.26],
    ["Ghana", 3.3, 80.38, 33.02, 58.0, 64.11],
    ["Mozambique", 7.4, 63.42, 52.77, 52.5, 61.17],
    ["Madagascar", 7.6, 77.25, 36.26, 58.9, 65.18],
    ["Côte d'Ivoire", 6.5, 89.89, 55.9, 60.4, 59.03],
    ["Cameroon", 9.9 , 78.23, 48.34, 51.9, 60.83],
```

```
    ["Niger", 5.0, 37.34, 45.61, 53.7, 61.45],
    ["Mali", 5.1, 30.76, 58.77, 54.5, 58.63],
    ["Burkina Faso", 5.8, 46.04, 52.82, 56.2, 59.73],
    ["Malawi", 2.1, 67.31, 29.02, 52.8, 63.72],
    ["Zambia", 4.5, 87.50, 41.66, 47.8, 62.38],
    ["Chad", 4.2, 26.76, 67.4, 52.0, 52.78],
    ["Somalia", 5.4, 5.40, 72.72, 2.9, 55.97],
    ["Senegal", 6.8, 56.30, 28.85, 57.7, 68.01],
    ["Zimbabwe", 4.8, 89.70, 37.93, 39.0, 61.12],
    ["Guinea", 7.9, 45.33, 61.99, 44.6, 59.33],
    ["Rwanda", 2.1, 75.90, 30.27, 52.2, 66.77],
    ["Benin", 2.9, 45.84, 56.54, 59.8, 60.09],
    ["Burundi", 5.1, 74.71, 38.64, 41.9, 61.57],
])

# Convert data to a pandas DataFrame for better handling
df = pd.DataFrame(data, columns=columns)

# Split the data into features (X) and target variable (y)
X = df[["GDP", "Percent_Literacy", "Infant_Mortality_Rate"]].astype(float)
y = df["Life_Expectancy"].astype(float)

# Linear Regression
linear_model = LinearRegression()
linear_model.fit(X, y)
linear_pred = linear_model.predict(X)
linear_r2 = r2_score(y, linear_pred)

# Polynomial Regression (degree 2)
poly_features = PolynomialFeatures(degree=2)
X_poly = poly_features.fit_transform(X)
poly_model = LinearRegression()
poly_model.fit(X_poly, y)
poly_pred = poly_model.predict(X_poly)
poly_r2 = r2_score(y, poly_pred)

# Exponential Regression
X_exp = X["GDP"].values.reshape(-1, 1)  # Using only GDP for simplicity
y_exp = np.log(y)  # Log-transform the target variable for exponential regression
exp_model = LinearRegression()
exp_model.fit(X_exp, y_exp)
exp_pred = np.exp(exp_model.predict(X_exp))
exp_r2 = r2_score(y, exp_pred)

# Quadratic Function (degree 2)
quad_features = PolynomialFeatures(degree=2)
```

```
X_quad = quad_features.fit_transform(X)
quad_model = LinearRegression()
quad_model.fit(X_quad, y)
quad_pred = quad_model.predict(X_quad)
quad_r2 = r2_score(y, quad_pred)

# Visualize the predicted values with labels and connecting lines
fig = plt.figure(figsize=(15, 15))

# Combined 3D Plot
ax_combined = fig.add_subplot(111, projection='3d')
ax_combined.scatter(X["GDP"], X["Percent_Literacy"], y, color='blue', label='Actual Data
(Original)')

# Labeling data points
for i, country in enumerate(df["Country"]):
    ax_combined.text(X["GDP"].iloc[i], X["Percent_Literacy"].iloc[i], y.iloc[i], country,
color='blue')

# Predicted values from different models
ax_combined.scatter(X["GDP"], X["Percent_Literacy"], linear_pred, color='green',
label=f'Predicted (Linear)')
ax_combined.scatter(X["GDP"], X["Percent_Literacy"], poly_pred, color='orange',
label=f'Predicted (Polynomial)')
ax_combined.scatter(X["GDP"], X["Percent_Literacy"], exp_pred, color='red', label=f'Predicted
(Exponential)')
ax_combined.scatter(X["GDP"], X["Percent_Literacy"], quad_pred, color='purple',
label=f'Predicted (Quadratic)')

# Connect actual and predicted points with lines
for i in range(len(X)):
    ax_combined.plot([X["GDP"].iloc[i], X["GDP"].iloc[i]], [X["Percent_Literacy"].iloc[i],
X["Percent_Literacy"].iloc[i]],
                     [y.iloc[i], linear_pred[i]], color='red', linestyle='dashed', alpha=0.5)
    ax_combined.plot([X["GDP"].iloc[i], X["GDP"].iloc[i]], [X["Percent_Literacy"].iloc[i],
X["Percent_Literacy"].iloc[i]],
                     [y.iloc[i], poly_pred[i]], color='red', linestyle='dashed', alpha=0.5)
    ax_combined.plot([X["GDP"].iloc[i], X["GDP"].iloc[i]], [X["Percent_Literacy"].iloc[i],
X["Percent_Literacy"].iloc[i]],
                     [y.iloc[i], exp_pred[i]], color='red', linestyle='dashed', alpha=0.5)
    ax_combined.plot([X["GDP"].iloc[i], X["GDP"].iloc[i]], [X["Percent_Literacy"].iloc[i],
X["Percent_Literacy"].iloc[i]],
                     [y.iloc[i], quad_pred[i]], color='red', linestyle='dashed', alpha=0.5)

ax_combined.set_xlabel('GDP')
ax_combined.set_ylabel('Percent Literacy')
```

```
ax_combined.set_zlabel('Life Expectancy')
ax_combined.set_title('Combined Predictions')

# Show the plots
plt.tight_layout()
plt.legend()
plt.show()

# Analyze the results
results = pd.DataFrame({
    "Model": ["Linear", "Polynomial", "Exponential", "Quadratic"],
    "R2": [linear_r2, poly_r2, exp_r2, quad_r2]
})

print("Results:")
print(results)
```

3)Actual data

```
columns = ["Country", "GDP", "Percent_Literacy", "Infant_Mortality_Rate","Life_Expectancy"]
Nigeria, 3.0, 62.02, 72.24, 53.9, 52.89
Ethiopia, 7.0, 51.77, 35.37, 48.3, 65.37
Egypt, 3.9, 73.09, 16.65, 49.6, 70.99
DR Congo, 3.0, 80.02, 63.79, 47.9, 59.74
Tanzania, 4.8, 81.80, 34.72, 60.0, 66.41
South Africa, 5.2, 95.02, 25.78, 55.7, 65.25
Kenya, 9.3, 82.62, 31.15, 52.5, 62.68
Uganda, 3.8, 79.00, 31.86, 51.4, 62.85
Sudan, 4.7, 60.70, 39.92, 32.8, 65.61
Algeria, 9.0, 81.41, 19.46, 43.2, 74.45
Morocco, 4.6, 75.93, 16.02, 58.4, 73.92
Angola, 4.7, 72.28, 48.34, 53.0, 62.26
Ghana, 3.3, 80.38, 33.02, 58.0, 64.11
Mozambique, 7.4, 63.42, 52.77, 52.5, 61.17
Madagascar, 7.6, 77.25, 36.26, 58.9, 65.18
Côte d'Ivoire, 6.5, 89.89, 55.9, 60.4, 59.03
Cameroon, 9.9, 78.23, 48.34, 51.9, 60.83
Niger, 5.0, 37.34, 45.61, 53.7, 61.45
Mali, 5.1, 30.76, 58.77, 54.5, 58.63
Burkina Faso, 5.8, 46.04, 52.82, 56.2, 59.73
Malawi, 2.1, 67.31, 29.02, 52.8, 63.72
Zambia, 4.5, 87.50, 41.66, 47.8, 62.38
Chad, 4.2, 26.76, 67.4, 52.0, 52.78
Somalia, 5.4, 5.40, 72.72, 2.9, 55.97
Senegal, 6.8, 56.30, 28.85, 57.7, 68.01
Zimbabwe, 4.8, 89.70, 37.93, 39.0, 61.12
```

Guinea, 7.9, 45.33, 61.99, 44.6, 59.33
Rwanda, 2.1, 75.90, 30.27, 52.2, 66.77
Benin, 2.9, 45.84, 56.54, 59.8, 60.09
Burundi, 5.1, 74.71, 38.64, 41.9, 61.57

Linear Regression Predicted Values:

	Country	Predicted Life Expectancy
0	Nigeria	49.354436
1	Ethiopia	46.885578
2	Egypt	52.957126
3	DR Congo	53.210174
4	Tanzania	53.767267
5	South Africa	56.526558
6	Kenya	52.257783
7	Uganda	53.695859
8	Sudan	49.435486
9	Algeria	52.504030
10	Morocco	53.265559
11	Angola	51.479230
12	Ghana	54.132871
13	Mozambique	48.501495
14	Madagascar	51.700409
15	Côte d'Ivoire	54.038310
16	Cameroon	50.602433
17	Niger	44.478879
18	Mali	42.712249
19	Burkina Faso	45.669037
20	Malawi	52.128820
21	Zambia	54.804171
22	Chad	41.999679
23	Somalia	37.096448
24	Senegal	48.073787
25	Zimbabwe	55.241415
26	Guinea	44.405146
27	Rwanda	53.802346
28	Benin	46.662570
29	Burundi	52.110849

Polynomial Regression Predicted Values:

	Country	Predicted Life Expectancy
0	Nigeria	55.137116
1	Ethiopia	48.338092
2	Egypt	55.591256
3	DR Congo	48.225634
4	Tanzania	52.388128
5	South Africa	44.920994
6	Kenya	50.970157
7	Uganda	51.670816
8	Sudan	57.380889
9	Algeria	48.180346
10	Morocco	54.479890
11	Angola	56.191501
12	Ghana	49.471845
13	Mozambique	54.546862
14	Madagascar	54.429427
15	Côte d'Ivoire	56.625531
16	Cameroon	53.870708
17	Niger	49.879350
18	Mali	44.156455
19	Burkina Faso	51.117384
20	Malawi	55.314833
21	Zambia	49.025752
22	Chad	44.604887
23	Somalia	13.527344
24	Senegal	50.851827
25	Zimbabwe	48.393676
26	Guinea	42.381610
27	Rwanda	49.217679
28	Benin	57.083879
29	Burundi	55.526133

Exponential Regression Predicted Values:

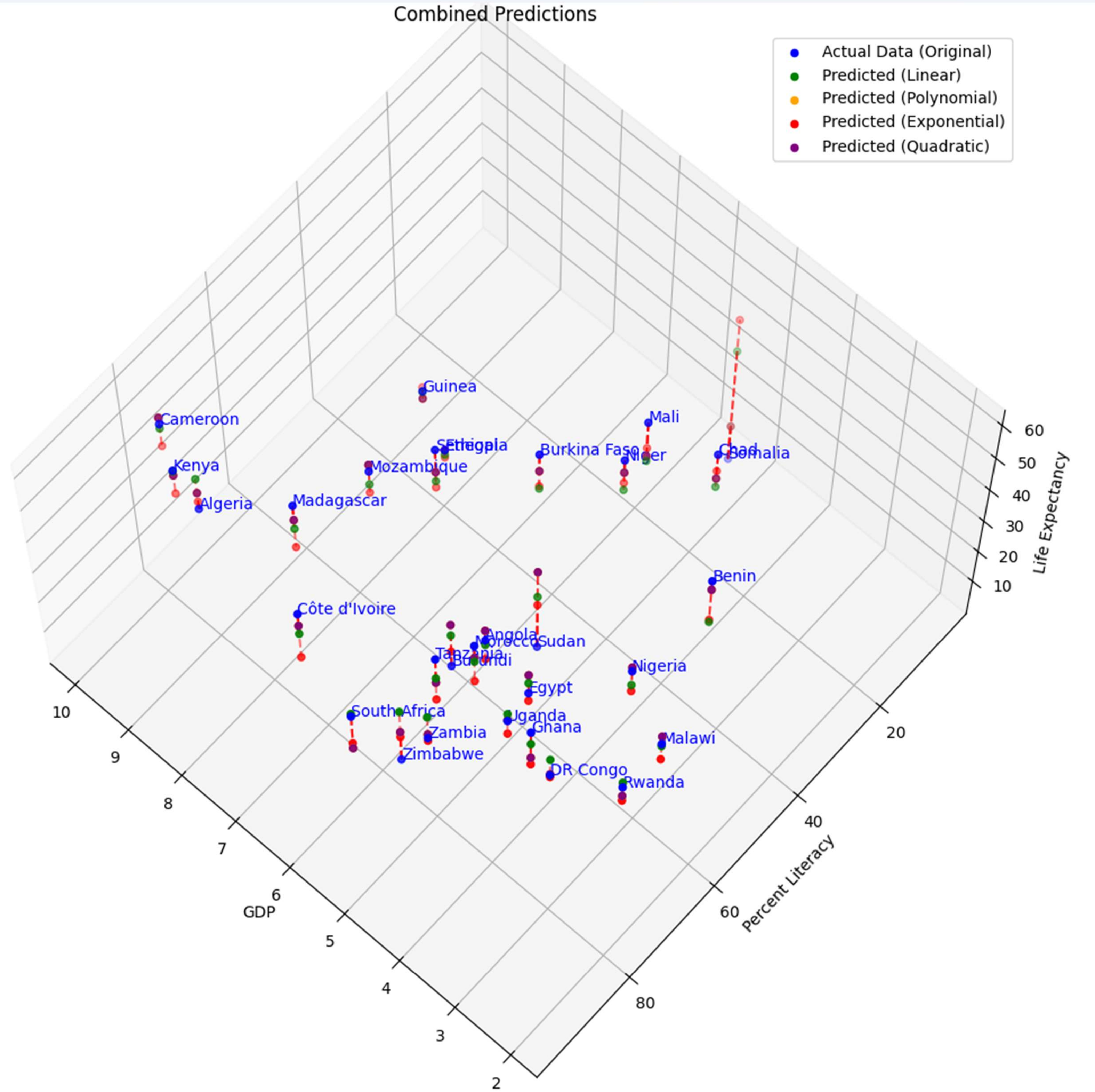
	Country	Predicted Life Expectancy
0	Nigeria	47.338629
1	Ethiopia	46.078451
2	Egypt	47.052117
3	DR Congo	47.338629
4	Tanzania	46.767340
5	South Africa	46.641326
6	Kenya	45.369096
7	Uganda	47.083866
8	Sudan	46.798896
9	Algeria	45.460998
10	Morocco	46.830474

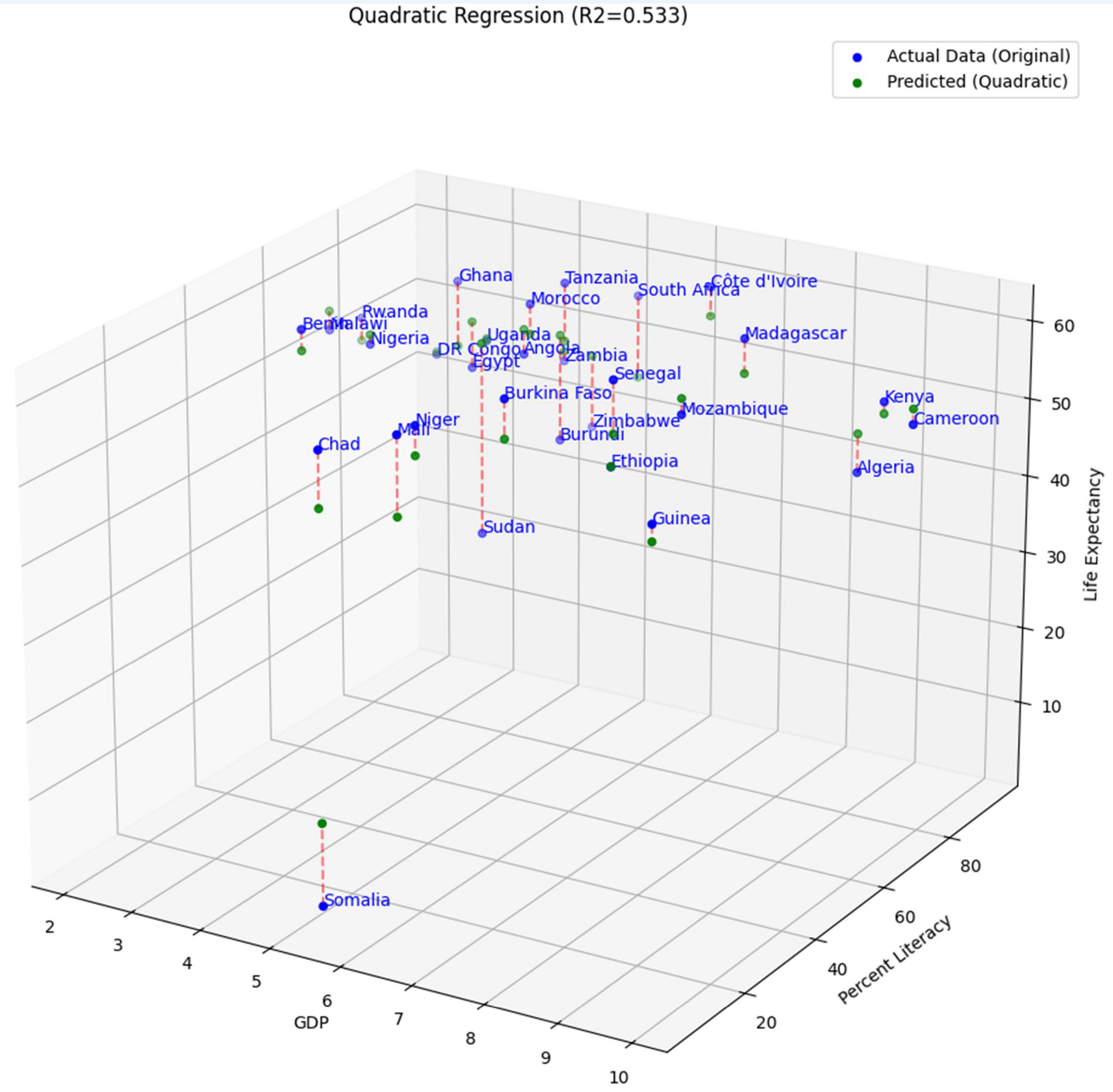
11	Angola	46.798896
12	Ghana	47.242932
13	Mozambique	45.954293
14	Madagascar	45.892340
15	Côte d'Ivoire	46.234120
16	Cameroon	45.185850
17	Niger	46.704290
18	Mali	46.672797
19	Burkina Faso	46.452941
20	Malawi	47.626886
21	Zambia	46.862074
22	Chad	46.956999
23	Somalia	46.578446
24	Senegal	46.140656
25	Zimbabwe	46.767340
26	Guinea	45.799566
27	Rwanda	47.626886
28	Benin	47.370572
29	Burundi	46.672797

Quadratic Regression Predicted Values:

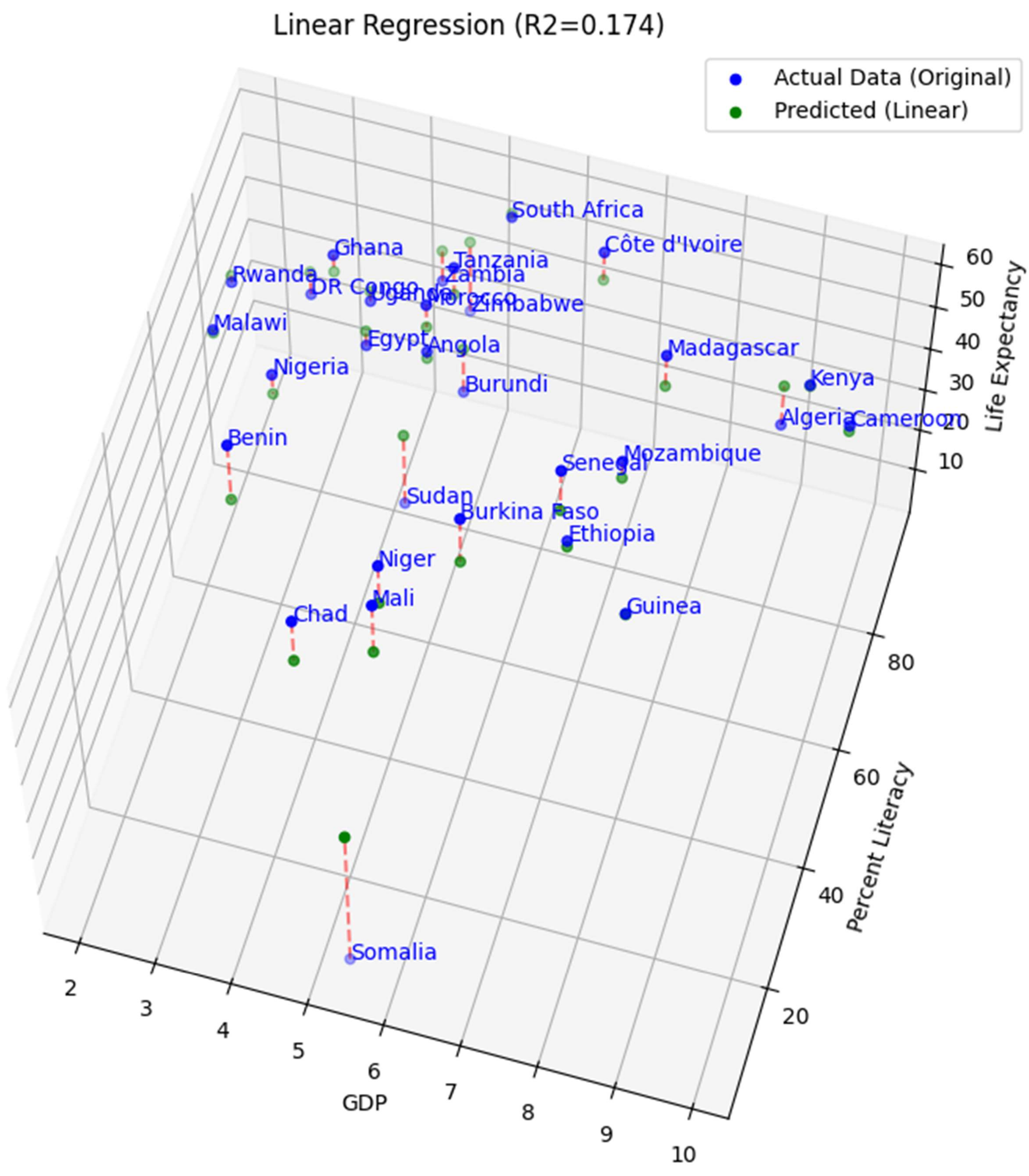
	Country	Predicted Life Expectancy
0	Nigeria	55.137116
1	Ethiopia	48.338092
2	Egypt	55.591256
3	DR Congo	48.225634
4	Tanzania	52.388128
5	South Africa	44.920994
6	Kenya	50.970157
7	Uganda	51.670816
8	Sudan	57.380889
9	Algeria	48.180346
10	Morocco	54.479890
11	Angola	56.191501
12	Ghana	49.471845
13	Mozambique	54.546862
14	Madagascar	54.429427
15	Côte d'Ivoire	56.625531
16	Cameroon	53.870708
17	Niger	49.879350
18	Mali	44.156455
19	Burkina Faso	51.117384
20	Malawi	55.314833
21	Zambia	49.025752
22	Chad	44.604887
23	Somalia	13.527344

24	Senegal	50.851827
25	Zimbabwe	48.393676
26	Guinea	42.381610
27	Rwanda	49.217679
28	Benin	57.083879
29	Burundi	55.526133

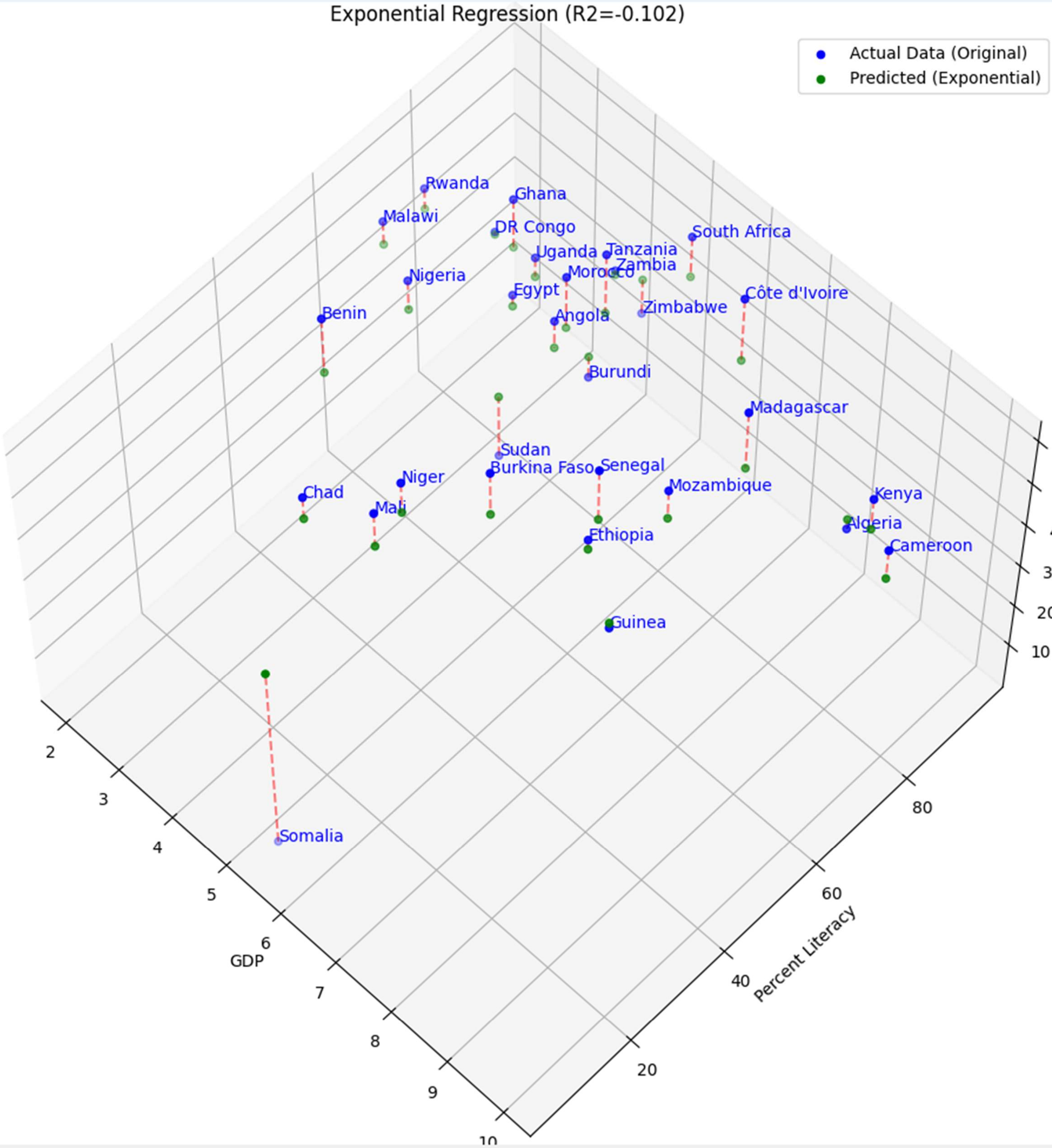




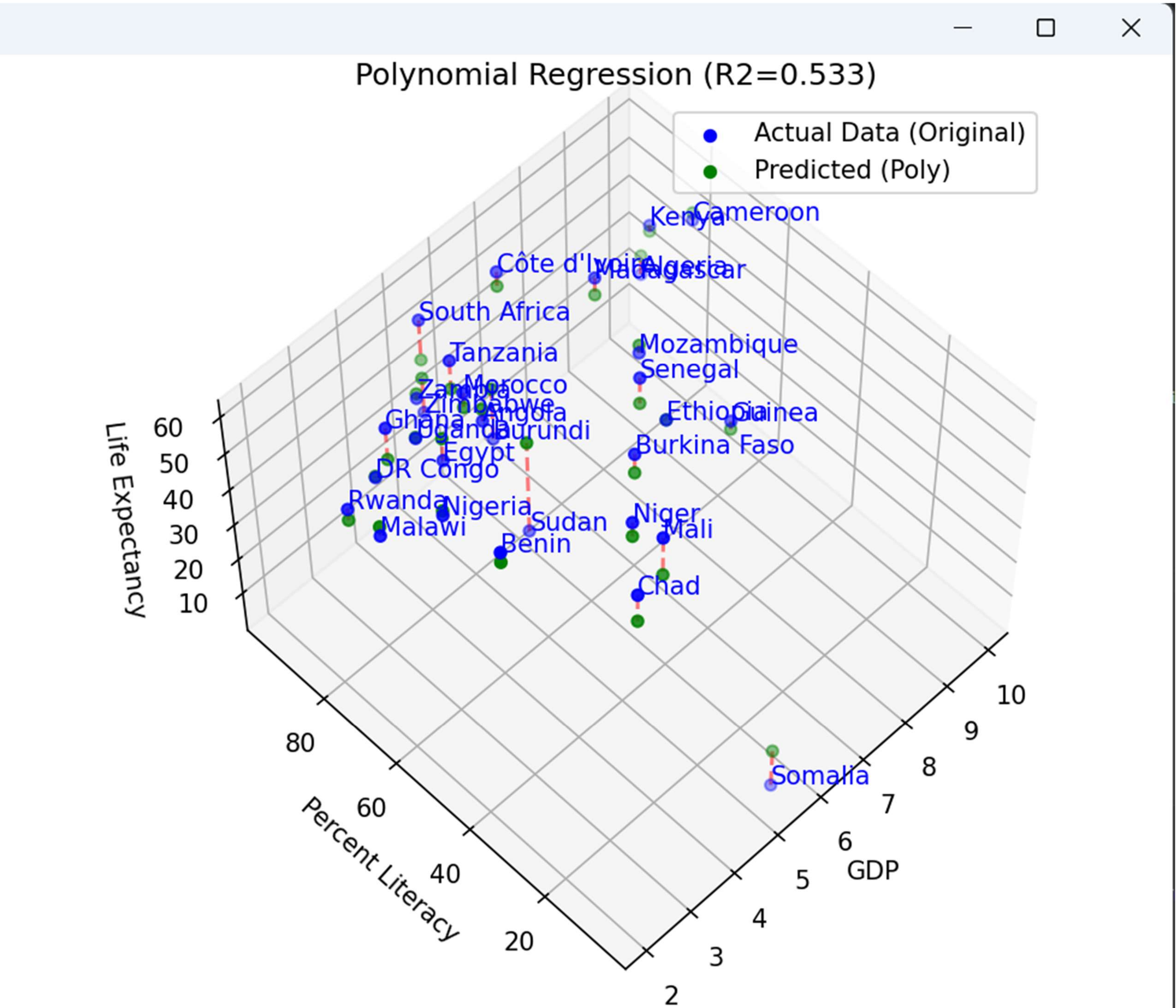
Quadratic Regression (R2=0.533)
Actual Data (Original)
Predicted (Quadratic)
Ghana
Tanzania
Côte d'Ivoire
Morocco
South Africa
Rwanda
Benin
Malawi
Nigeria
Uganda
DR Congo
Angola
Egypt
Zambia
Madagascar
Senegal
Burkina Faso
Mozambique
Kenya
Cameroon
Niger
Mali
Zimbabwe
Burundi
Chad
Ethiopia
Algeria
Guinea
Sudan
Somalia
GDP
Percent Literacy
Life Expectancy
2
3
4
5
6
7
8
9
10
20
40
60
80
10
20
30
40
50
60



Linear Regression (R2=0.174)
Actual Data (Original)
Predicted (Linear)
South Africa
Ghana
Côte d'Ivoire
Tanzania
Zambia
Rwanda
DR Congo
Uganda
Morocco
Zimbabwe
Malawi
Egypt
Angola
Madagascar
Nigeria
Burundi
Kenya
Algeria
Cameroon
Benin
Mozambique
Senegal
Sudan
Burkina Faso
Ethiopia
Niger
Mali
Chad
Guinea
Somalia
GDP
Percent Literacy
Life Expectancy



Exponential Regression (R2=-0.102)
Actual Data (Original)
Predicted (Exponential)
Rwanda
Ghana
Malawi
DR Congo
South Africa
Uganda
Tanzania
Zambia
Morocco
Nigeria
Egypt
Zimbabwe
Côte d'Ivoire
Benin
Angola
Burundi
Madagascar
Sudan
Senegal
Burkina Faso
Niger
Mozambique
Chad
Mali
Kenya
Algeria
Ethiopia
Cameroon
Guinea
Somalia
GDP
Percent Literacy
Life Expectancy
2
3
4
5
6
7
8
9
10
20
40
60
80
10
20
30
40
50
60



9. Reference

GDP Rating

Country Rankings: World & Global Economy Rankings on Economic freedom. (n.d.). Index of Economic Freedom. https://www.heritage.org/index/ranking

Mortality Rates

Infant mortality rate by country 2023. (n.d.). https://worldpopulationreview.com/country-rankings/infant-mortality-rate-by-country

Literacy Rates
(Country) Literacy Rate 2000-2019. (2019). Macrotrends.net. https://www.macrotrends.net/countries/KEN/kenya/literacy-rate (This is one of countries that reference on the same url but change country name)

United Nation. (2022). Statistical Annex. https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/WESP2022_ANNEX.pdf