Introduction:

The objective of this analysis is to gain valuable business insights from the World Development Data. The dataset encompasses various economic, demographic, and developmental indicators across different countries and regions. Focus is on extracting meaningful information to inform decision-making and potentially identify patterns or trends.

What & Why:

1.Descriptive Statistics:

The initial exploration involves descriptive statistics to understand the central tendencies, variations, and distributions of key variables. These statistics provide a baseline understanding of the dataset, guiding subsequent analyses.

2. Boxplot of GDP by Region:

A boxplot is generated to visualize the distribution of Gross Domestic Product (GDP) across different regions. This helps identify potential variations and outliers in economic performance.

3. T-Test for Regional Comparison:

A t-test is conducted to compare the GDP between regions 'Asia' and 'Americas'. The resulting t-statistic and p-value provide insights into whether there is a significant difference in economic performance between the two regions.

4. A/B Testing with Paired T-Test:

A/B testing is performed to assess the impact of a policy change on GDP. The paired t-test compares GDP before and after the policy change in the years 2000 and 2020.

5. Linear Regression:

Linear regression is employed to model the relationship between GDP and selected features ('PopDens' and 'FertRate'). The Mean Squared Error (MSE) is used to evaluate the model's performance.

6. K-Means Clustering:

K-means clustering is applied to group countries based on demographic and economic features. The resulting clusters and their characteristics provide insights into potential similarities or disparities among nations.

Conclusion:

In conclusion, this multifaceted analysis has provided valuable insights into the World Development Data. Descriptive statistics offered an overview, the boxplot highlighted regional

disparities, t-tests gauged economic differences, A/B testing assessed policy impact, linear regression modeled relationships, and k-means clustering identified country clusters. Each approach was chosen judiciously based on the dataset's nature and the business questions at hand, contributing to a comprehensive understanding of the global economic landscape.

Key Business Insights:

Descriptive Statistics:
- The dataset consists of 23 columns covering demographic and economic variables.
- Population-related variables such as 'PopTotal' and 'PopDens' exhibit substantial variation, with a wide range of values.
- Economic indicators like 'GDP' have a considerable standard deviation, indicating heterogeneity in global economic development.

Regional Comparison:
- A t-test was conducted to compare the GDP of countries in Asia and the Americas.
- The result suggests no significant difference in GDP between these regions (T-statistic: -0.994, P-value: 0.320).

A/B Testing:
- An A/B test was performed to assess the impact of a policy change on GDP between the years 2000 and 2020.
- The results indicate no statistically significant difference in GDP before and after the policy change (T-statistic: -1.723, P-value: 0.086).

Linear Regression:
- A linear regression model was built to predict GDP based on population density ('PopDens') and fertility rate ('FertRate').
- The model's performance was evaluated using Mean Squared Error, which resulted in an unexpectedly high value (2.68e+24), suggesting potential issues with the model or the choice of features.

K-means Clustering:
- K-means clustering was applied to group countries based on demographic and economic features ('PopDens', 'GDP', 'LifeExpBirth').
- Three clusters were identified, each exhibiting distinct characteristics in terms of economic indicators and demographic features.

Regional Disparities: Significant variations in GDP exist between different regions, indicating diverse economic performances.

Policy Impact Assessment: A/B testing suggests a potential impact of the policy change on GDP, though the result is not statistically significant at a conventional threshold.

Predictive Modeling: The linear regression model provides a basis for predicting GDP based on population density and fertility rate, offering insights for economic planning.

Country Clustering: K-means clustering reveals distinct groups of countries based on demographic and economic features, aiding in targeted policy and business strategies.