1. **Descriptive Analysis**:
   * Calculate summary statistics such as mean, median, minimum, and maximum cost of living values.
   * Visualize the data using histograms, box plots, or bar charts to get a sense of the distribution.
2. **Ranking Analysis**:
   * Explore the global rankings and identify the top and bottom countries in terms of the cost of living.
   * Compare the rankings to see how different countries compare to each other.
3. **Time Series Analysis**:
   * If you have data for multiple years, you can analyze trends and changes in the cost of living over time.
   * Plot a line chart to visualize how the cost of living has changed in each country over the years.
4. **Correlation Analysis**:
   * If you have additional datasets, you can analyze the correlation between the cost of living and other factors such as GDP, inflation rates, or income levels.
   * Calculate correlation coefficients to quantify the relationships.
5. **Geospatial Analysis**:
   * If you have geographic data, you can map the cost of living values for different countries.
   * Analyze if there are regional patterns or clusters in the cost of living.
6. **Comparative Analysis**:
   * Compare the cost of living in these countries to a reference point (e.g., the cost of living in the United States) to see how they differ.
   * Explore what factors contribute to the cost of living differences.
7. **Forecasting**:
   * If you have data for multiple years, you can build time series forecasting models to predict future cost of living values for these countries.
8. **Outlier Detection**:
   * Identify if there are any outliers in the dataset, which may indicate unusual cost of living values for a particular country.
9. **Machine Learning**:
   * If you have additional data, you can build predictive models to understand what factors influence the cost of living and make predictions.
10. **Policy Analysis**:
    * Use the data to inform policy decisions or recommendations related to cost of living adjustments, minimum wage, or economic policies.

The type of analysis you choose should align with your research questions and goals. Additionally, you may want to consider using data from multiple years to track changes over time and gain deeper insights into the cost of living in these countries.

**DATA VISUALIZATION**

1. **Bar Charts**:
   * Create a bar chart to visualize the global rank for each country. This can help you see the ranking distribution.
   * Create a bar chart to visualize the cost of living for each country. This can help you compare the cost of living across countries.
2. **Histograms**:
   * Create histograms to show the distribution of global ranks and cost of living values. This can help you understand the frequency of different ranks and cost of living ranges.
3. **Scatter Plots**:
   * Create a scatter plot with global rank on one axis and cost of living on the other. This can help you explore the relationship between these two variables.
4. **Box Plots**:
   * Create box plots to visualize the distribution of global ranks and cost of living. Box plots show quartiles, medians, and potential outliers.
5. **Line Charts**:
   * If you have data for multiple years, you can create line charts to visualize how global ranks and cost of living have changed over time.
6. **Heatmaps**:
   * Create a heatmap to show the correlation between global rank and cost of living. This can help you see if there's a relationship between these two variables.
7. **Geospatial Maps**:
   * If you have geographic data, you can create a map to visualize the cost of living on a world map. You can use color coding to represent different cost of living levels.
8. **Pair Plots**:
   * Create pair plots to visualize pairwise relationships between multiple variables, including global rank and cost of living.
9. **Violin Plots**:
   * Violin plots combine a box plot with a kernel density estimation and can be useful for visualizing the distribution of cost of living and global ranks.
10. **Word Clouds**:
    * If you have textual data related to the countries (e.g., country names, descriptions), you can create a word cloud to highlight frequently occurring words.
11. **Data Preparation**: Prepare your dataset by cleaning, preprocessing, and transforming the data as needed. This includes handling missing values, encoding categorical variables, and splitting the data into training and testing sets.
12. **Feature Selection**: Identify the most relevant features (independent variables) that could influence the cost of living. You can use feature selection techniques to choose the most important variables.
13. **Model Selection**: Choose an appropriate machine learning model for your task. For predictive modeling, common choices include linear regression, decision trees, random forests, support vector machines, or neural networks. The choice depends on the nature of your data and the problem you are trying to solve.
14. **Model Training**: Train your selected model using the training dataset. The model will learn the relationships between the features and the cost of living.
15. **Model Evaluation**: Evaluate the performance of your model using the testing dataset. Common evaluation metrics for regression tasks include Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R2).
16. **Model Tuning**: If necessary, tune the hyperparameters of your model to improve its performance. You can use techniques like grid search or random search for hyperparameter tuning.
17. **Interpretation**: Analyze the model to understand which factors have the most significant influence on the cost of living. This can provide valuable insights into the drivers of cost of living.
18. **Prediction**: Once you have a trained and validated model, you can use it to make predictions on new or unseen data. For example, you can predict the cost of living for a new set of countries or regions.
19. **Model Deployment**: If you want to make the model available for practical use, you can deploy it as a web service or integrate it into your applications.

In this example, we create a world map using **folium** and add markers for each country with color-coded circles based on their cost of living levels. The **categorize\_cost\_of\_living** function assigns countries to "High," "Moderate," or "Low" cost of living categories based on your defined thresholds. The map is saved as an HTML file (**cost\_of\_living\_map.html**) that you can open in a web browser to view the interactive map.

1. **Linear Regression**:
   * **Nature**: Linear regression models assume a linear relationship between the independent variables and the target variable. They are simple and interpretable.
   * **Use Case**: Use linear regression when you suspect a linear relationship between the features (e.g., "Global rank") and the cost of living.
2. **Decision Trees**:
   * **Nature**: Decision trees are versatile and can capture nonlinear relationships. They partition the feature space into regions and make predictions based on the majority class or mean within each region.
   * **Use Case**: Decision trees can work well for complex relationships when the cost of living prediction might have multiple decision boundaries.
3. **Random Forests**:
   * **Nature**: Random forests are an ensemble of decision trees. They provide robust predictions by averaging or voting across multiple decision trees, reducing overfitting.
   * **Use Case**: Random forests are effective for capturing complex interactions and reducing the risk of overfitting. They are a popular choice for regression tasks.
4. **Support Vector Machines (SVM)**:
   * **Nature**: SVMs aim to find the hyperplane that best separates data into different classes. In regression (SVR), they find a hyperplane that best fits the data.
   * **Use Case**: SVMs can be useful when there's a need to capture complex relationships in a high-dimensional space.
5. **Neural Networks (Deep Learning)**:
   * **Nature**: Deep learning models, such as neural networks, can capture highly complex and nonlinear relationships. They consist of multiple layers of neurons.
   * **Use Case**: Use deep learning when dealing with large datasets and complex relationships. However, they require more data and computational resources.

The choice of the model should be based on a combination of factors, including the complexity of the problem, the amount of data available, and the interpretability of the model. You can start with a simple model like linear regression and gradually explore more complex models if needed. Additionally, you should evaluate the model's performance using appropriate metrics (e.g., Mean Absolute Error, Mean Squared Error) and consider techniques like cross-validation to assess generalization performance.

Support Vector Machines (SVMs) are versatile machine learning models that can be used for both classification and regression tasks. In the context of regression (Support Vector Regression or SVR), they aim to find a hyperplane that best fits the data. Here's how you can proceed with building and using an SVR model to predict the cost of living based on the "Global rank" feature:

Top of Form