



Report For The Course Work of Data Analytics and Visualization.

Name: Nabin Pyakurel

Module Leader: Siman Giri

Instructor: Ronit Shrestha and Shiv Kumar Yadav

Data Exploration

Dataset Loading

```
df = pd.read_csv('/content/drive/MyDrive/Dataset/World_Development_Indicators.csv')
```

```
df.shape
```

```
(18359, 24)
```

Data Filtering

```
df.isnull().sum()
```

```
0
```

```
Country Name 3
```

```
Country Code 5
```

```
Series Name 5
```

```
Series Code 5
```

```
2005 [YR2005] 5
```

```
2006 [YR2006] 5
```

```
2007 [YR2007] 5
```

```
2008 [YR2008] 5
```

```
2009 [YR2009] 5
```

```
2010 [YR2010] 5
```

```
2011 [YR2011] 5
```

```
2012 [YR2012] 5
```

```
2013 [YR2013] 5
```

```
2014 [YR2014] 5
```

```
2015 [YR2015] 5
```

```
2016 [YR2016] 5
```

```
2017 [YR2017] 5
```

```
2018 [YR2018] 5
```

```
2019 [YR2019] 5
```

```
2020 [YR2020] 5
```

```
2021 [YR2021] 5
```

```
2022 [YR2022] 5
```

```
2023 [YR2023] 5
```

```
2024 [YR2024] 5
```

```
dtype: int64
```

```
df_clean=df.dropna()
```

```
df_clean.shape
```

```
(18354, 24)
```

```
df.columns
```

```
Index(['Country Name', 'Country Code', 'Series Name', 'Series Code',  
      '2005 [YR2005]', '2006 [YR2006]', '2007 [YR2007]', '2008 [YR2008]',  
      '2009 [YR2009]', '2010 [YR2010]', '2011 [YR2011]', '2012 [YR2012]',  
      '2013 [YR2013]', '2014 [YR2014]', '2015 [YR2015]', '2016 [YR2016]',  
      '2017 [YR2017]', '2018 [YR2018]', '2019 [YR2019]', '2020 [YR2020]',  
      '2021 [YR2021]', '2022 [YR2022]', '2023 [YR2023]', '2024 [YR2024]'],  
      dtype='object')
```

```
df_long = df_clean.melt(id_vars=['Country Name', 'Country Code', 'Series Name', 'Series Code'],  
                        var_name='Year',  
                        value_name='Value')
```

```
df_long['Year'] = df_long['Year'].str.extract(r'(\d{4})').astype(int)  
  
df_long['Value'] = pd.to_numeric(df_long['Value'], errors='coerce')
```

```
# Filter for the relevant indicator  
political_stability_df = df_long[df_long['Series Name'] == 'Political Stability and Absence of Violence/Terrorism: Estimate'].copy()  
  
display(political_stability_df.head())
```

```
# Pivot the data for easier analysis  
political_stability_pivot = political_stability_df.pivot_table(index='Country Name',  
                                                              columns='Year',  
                                                              values='Value')  
  
# Filter for the relevant indicator (FDI)  
fdi_df = df_long[df_long['Series Name'] == 'Foreign direct investment, net inflows (% of GDP)'].copy()  
  
display(fdi_df.head())
```

	Country Name	Country Code	Series Name	Series Code	Year	Value
48	Afghanistan	AFG	Foreign direct investment, net inflows (% of GDP)	BX.KLT.DINV.WD.GD.ZS	2005	4.368673
117	Albania	ALB	Foreign direct investment, net inflows (% of GDP)	BX.KLT.DINV.WD.GD.ZS	2005	3.178998
186	Algeria	DZA	Foreign direct investment, net inflows (% of GDP)	BX.KLT.DINV.WD.GD.ZS	2005	1.079903
255	American Samoa	ASM	Foreign direct investment, net inflows (% of GDP)	BX.KLT.DINV.WD.GD.ZS	2005	NaN
324	Andorra	AND	Foreign direct investment, net inflows (% of GDP)	BX.KLT.DINV.WD.GD.ZS	2005	NaN

```
# Pivot the data for easier analysis  
fdi_pivot = fdi_df.pivot_table(index='Country Name',  
                               columns='Year',  
                               values='Value')
```

To Find min, max, mean, std and also to check for outliers or unusual values.

```
political_stability_pivot.describe()
```

	Year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
count		201.000000	202.000000	202.000000	203.000000	204.000000	204.000000	205.000000	205.000000	205.000000	205.000000	205.000000	205.000000	205.000000	205.000000	205.000000	205.000000	205.000000	205.000000	205.000000
mean		-0.022641	-0.023472	-0.022597	-0.020549	-0.027940	-0.033424	-0.031282	-0.030006	-0.028475	-0.024850	-0.026591	-0.025961	-0.025669	-0.034222	-0.034531	-0.035217	-0.035258	-0.035477	-0.027372
std		1.004189	1.003161	1.003880	1.005384	1.007303	1.000941	0.998179	0.999821	1.001067	1.001481	0.999693	1.000762	1.000878	1.001815	0.999925	0.998697	0.998937	0.998603	1.000556
min		-2.705590	-2.826402	-3.228497	-3.280517	-3.312951	-3.130971	-3.083847	-2.860755	-2.750687	-2.748824	-2.965177	-2.906036	-2.934317	-2.996235	-2.771014	-2.711848	-2.727339	-2.776898	-2.750607
25%		-0.782739	-0.736241	-0.661469	-0.657102	-0.665882	-0.718850	-0.708380	-0.680454	-0.665950	-0.643627	-0.619956	-0.628554	-0.642760	-0.640142	-0.577341	-0.610105	-0.627814	-0.615027	-0.521786
50%		0.085603	0.101034	0.183067	0.118725	0.105565	0.051369	0.021105	0.079325	0.065314	0.049136	0.035213	0.052190	0.048514	0.013894	0.027658	0.016718	0.035331	0.043973	0.069302
75%		0.875413	0.859174	0.787584	0.811134	0.806966	0.798065	0.912176	0.893706	0.894792	0.806274	0.847117	0.834964	0.783372	0.823102	0.801589	0.837047	0.832015	0.789451	0.763501
max		1.594947	1.501111	1.488697	1.511940	1.553164	1.645201	1.940006	1.934407	1.931953	1.924541	1.947082	1.964211	1.927584	1.936804	1.887363	1.910595	1.877587	1.660742	1.625601

fdi_pivot.describe()

	Year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
count		238.000000	239.000000	241.000000	242.000000	242.000000	241.000000	243.000000	243.000000	244.000000	242.000000	241.000000	241.000000	241.000000	241.000000	239.000000	239.000000	240.000000	239.000000	236.000000	136.000000
mean		5.944341	14.659727	13.993642	9.612242	10.573739	10.555918	11.182638	10.842080	8.200703	9.840260	17.739125	8.485800	8.457469	0.670109	7.773958	2.671286	6.147533	2.203376	3.153753	4.236123
std		22.684422	84.575417	68.292233	35.744738	54.761654	51.158453	57.064424	50.918764	51.432083	66.745861	137.224384	62.855916	45.143238	92.097184	35.082736	25.980596	28.701426	30.929106	12.692372	15.805484
min		-8.804936	-6.221490	-57.532314	-6.549732	-3.482045	-15.885438	-5.178040	-12.033180	-5.655618	-8.261019	-7.418557	-37.172653	-41.650995	-1303.108267	-13.674036	-296.013199	-15.121443	-444.706890	-71.719467	-14.708707
25%		1.518191	2.065121	2.511457	2.520335	1.525264	1.687783	1.866232	1.580224	1.434747	1.389289	1.249618	1.397506	1.537521	1.187104	1.362335	0.663284	1.436251	1.221365	0.717786	0.701312
50%		2.921040	3.677271	4.495417	3.824474	2.746755	2.930410	3.218660	2.791457	2.850557	2.552040	2.701035	2.739903	2.651415	2.314587	2.388496	1.704829	2.706077	2.259507	1.903532	1.740657
75%		5.680209	7.232493	8.654718	7.961528	5.450169	5.592659	6.173205	5.948393	4.786671	4.379341	4.930208	5.021740	4.550256	3.941183	4.248640	3.658095	4.651217	4.507132	3.753439	3.710631
max		341.007652	1114.933386	728.065060	426.427221	664.573767	587.830616	743.700703	540.366946	792.319802	1009.031051	1709.827232	972.693539	602.374526	452.221040	431.788470	191.860528	433.758961	126.084088	113.224329	174.827688

```
# Rename columns before merging
political_stability_df = political_stability_df.rename(columns={'Value': 'Political_Stability'})
fdi_df = fdi_df.rename(columns={'Value': 'FDI'})

merged_df = political_stability_df.merge(fdi_df, on=['Country Name','Country Code','Year'], suffixes=('_stability','_fdi'))
```

```
merged_df.isnull().sum()
```

	0
Country Name	0
Country Code	0
Series Name_stability	0
Series Code_stability	0
Year	0
Political_Stability	1439
Series Name_fdi	0
Series Code_fdi	0
FDI	610

dtype: int64

```
merged_df['Political_Stability'] = merged_df['Political_Stability'].fillna(
    merged_df['Political_Stability'].mean()
)
merged_df['FDI'] = merged_df['FDI'].fillna(
    merged_df['FDI'].mean()
)

merged_df.isnull().sum()
```

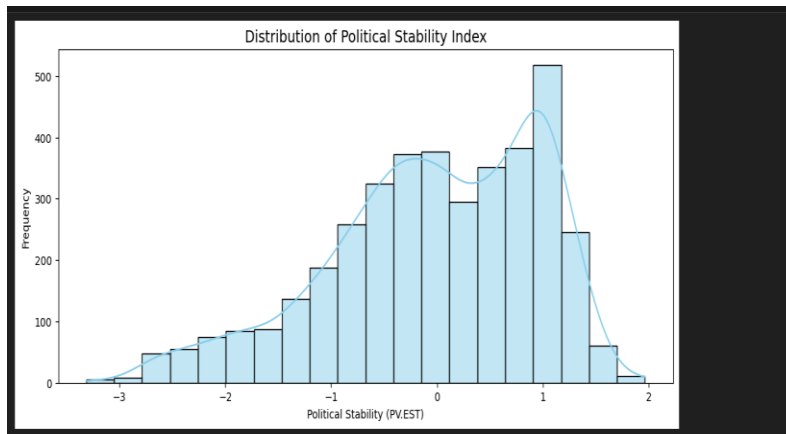
```
merged_df.to_csv("merged_df.csv", index=False)
import os
print(os.listdir()) # check if file is there

['.config', 'drive', 'merged_df.csv', 'sample_data']
```

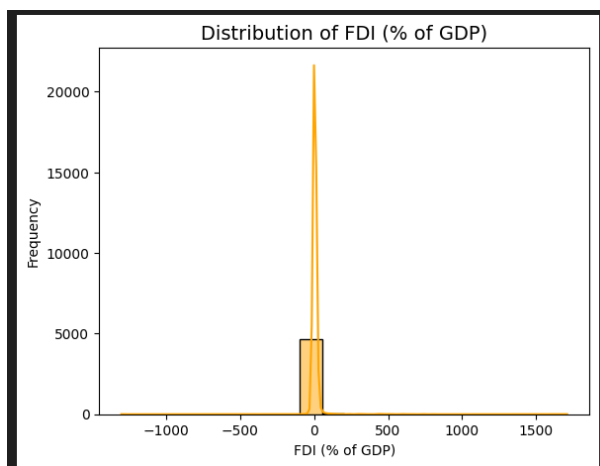
Data Analysis Key Findings

- The median political stability scores and the median FDI percentages varied significantly among the selected countries (United States, China, India, Brazil, Germany, Nigeria, Nepal).
- Some countries exhibited a wider spread in political stability scores, indicating higher volatility, while others were more stable. Similarly, the spread of FDI percentages varied, suggesting different levels of consistency in attracting foreign investment.
- Outliers were observed in both political stability and FDI data for some countries, indicating periods of unusual fluctuations.
- While a simple linear relationship was not immediately apparent, some countries with higher median political stability appeared to have higher or more consistent median FDI, and vice versa. However, exceptions were noted, suggesting other factors influence the relationship.

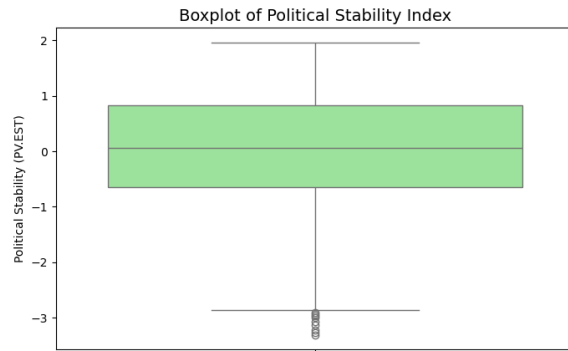
Data Visualization



The histogram above shows the distribution of the Political Stability Index. The distribution appears roughly symmetrical, centered around zero, which indicates that political stability is evenly distributed across the countries and years in the dataset.

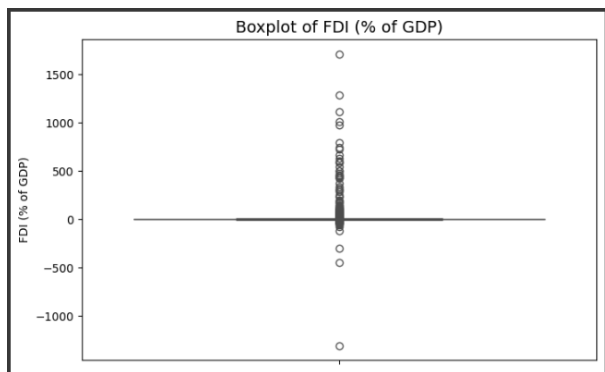


The histogram above shows the distribution of FDI (% of GDP). The distribution is heavily skewed to the right, with a large number of observations clustered near zero and a long tail extending to the right. This indicates that while most countries have relatively low FDI, a few countries attract exceptionally high levels of foreign investment.



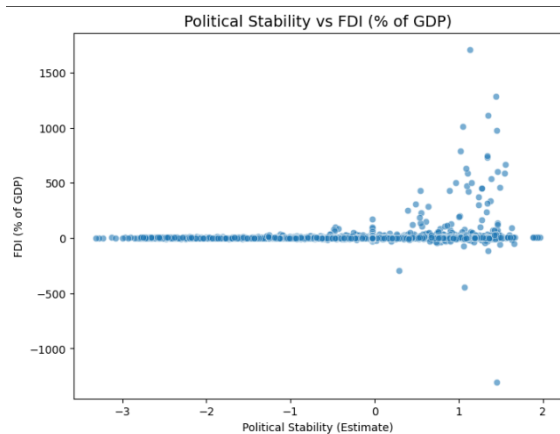
Political Stability Distribution

This boxplot shows the distribution of the Political Stability Index across different countries and years. From the plot, it appears that the political stability scores vary, and there are some potential **outliers**, particularly on the lower end. Since our research question is "**Is political stability linked to higher foreign investment?**", this chart helps visualize the distribution and spread of our independent variable, which is relevant for our analysis.

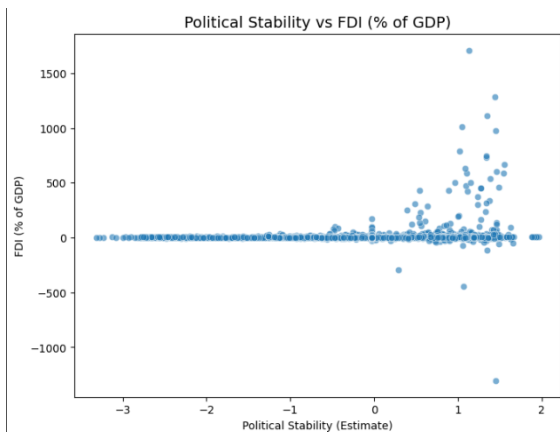


FDI Distribution

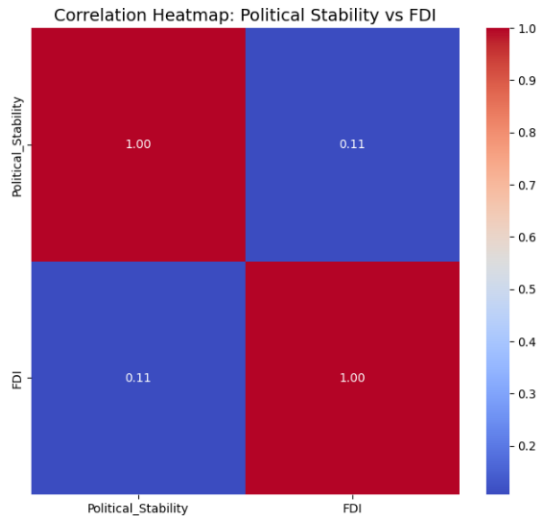
This boxplot shows the distribution of Foreign Direct Investment (FDI) (% of GDP) across different countries and years. From the plot, it appears that FDI varies significantly, with a large number of **outliers**, particularly on the higher end.



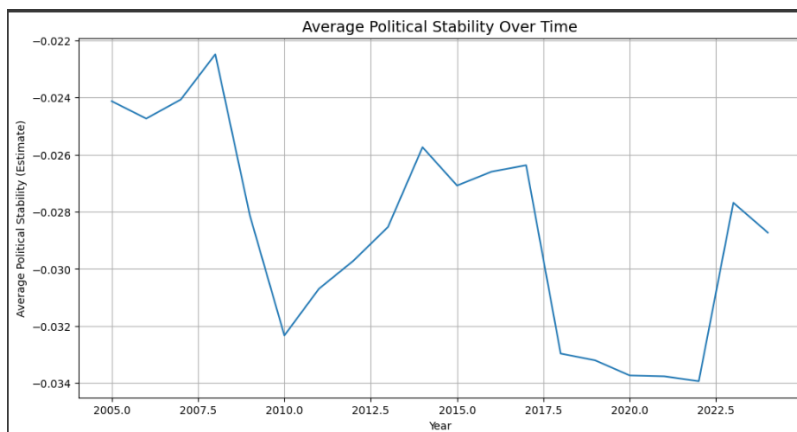
This scatter plot visualizes the relationship between Political Stability and Foreign Direct Investment (FDI) (% of GDP). Each point represents a country-year observation. The plot shows a general trend where higher political stability (moving to the right on the x-axis) is associated with higher FDI (moving upwards on the y-axis), but there is a significant amount of scatter, indicating that the relationship is not perfectly linear. Notably, there are many **outliers** with very high FDI values, particularly at higher levels of political stability. These outliers suggest that while stability may be a factor, other influences can lead to exceptionally high foreign investment in certain cases.



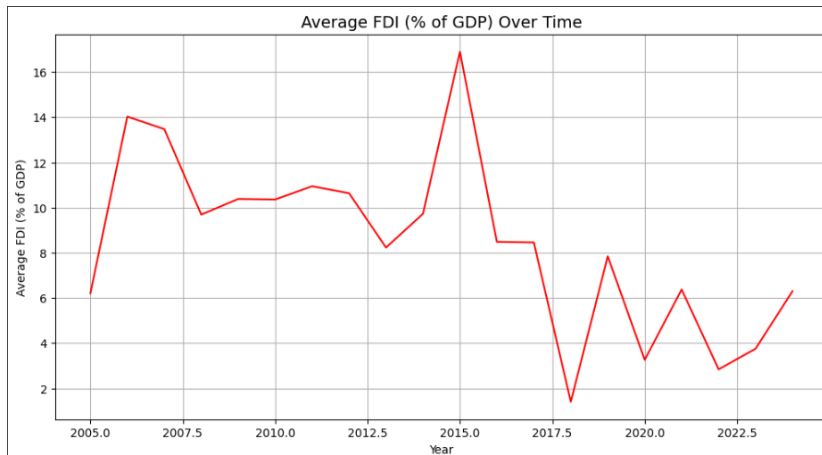
This scatter plot visualizes the relationship between Political Stability and Foreign Direct Investment (FDI) (% of GDP). Each point represents a country-year observation. The plot shows a general trend where higher political stability (moving to the right on the x-axis) is associated with higher FDI (moving upwards on the y-axis), but there is a significant amount of scatter, indicating that the relationship is not perfectly linear. Notably, there are many **outliers** with very high FDI values, particularly at higher levels of political stability. These outliers suggest that while stability may be a factor, other influences can lead to exceptionally high foreign investment in certain cases.



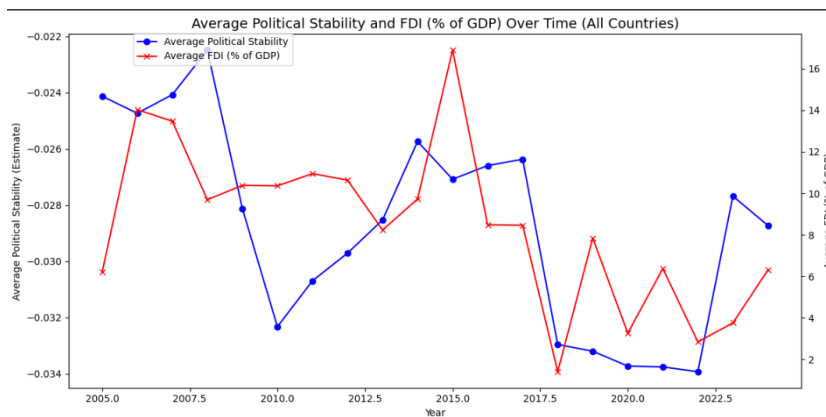
This heatmap shows the correlation matrix between Political Stability and FDI (% of GDP). The value in each cell represents the Pearson correlation coefficient between the two variables. A value close to 1 indicates a strong positive correlation, a value close to -1 indicates a strong negative correlation, and a value close to 0 indicates a weak correlation. This heatmap helps visualize the strength and direction of the linear relationship between political stability and FDI.



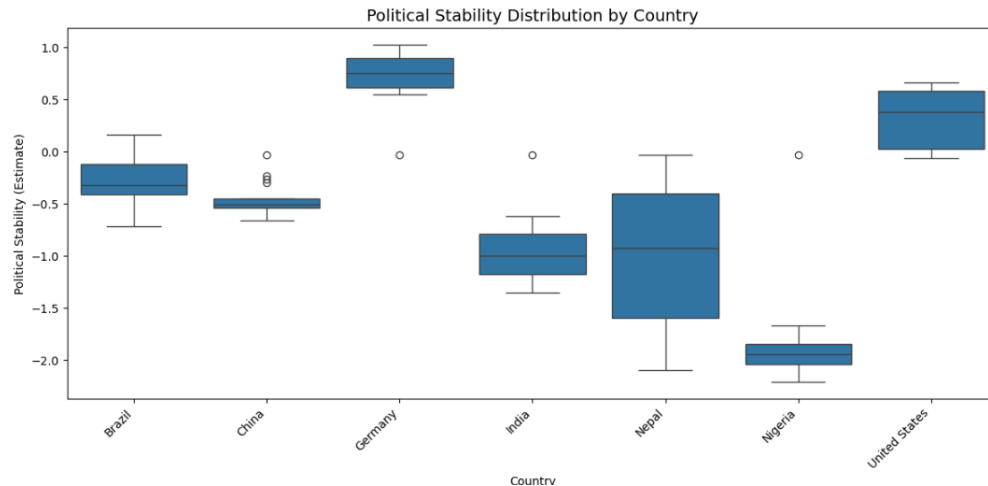
This line plot shows the trend of the average Political Stability Index across all countries over time. It allows us to observe how political stability has changed year by year, identifying any general patterns or significant shifts in stability over the period.



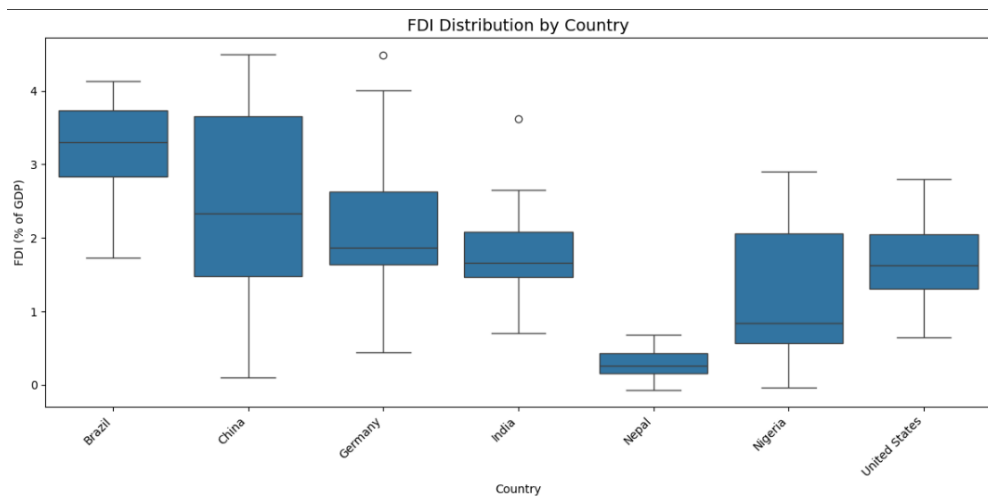
This line plot shows the trend of the average Foreign Direct Investment (FDI) (% of GDP) across all countries over time. It allows us to observe how FDI has changed year by year, identifying any general patterns or significant shifts in foreign investment over the period.



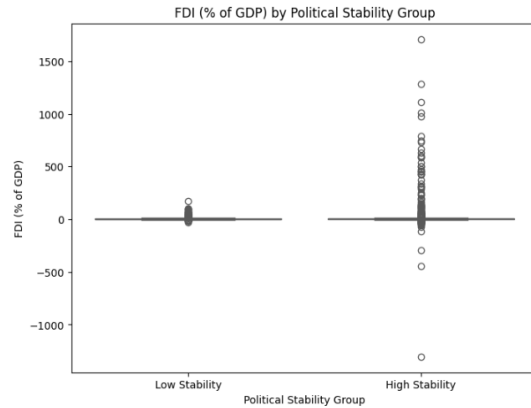
This dual-axis line plot shows the average Political Stability Index and the average Foreign Direct Investment (FDI) (% of GDP) across all countries over time on the same figure, using separate y-axes to accommodate their different scales. The blue line represents the average political stability, and the red line represents the average FDI. By visualizing both trends together, we can observe if there are any co-movements or divergences between political stability and FDI over the years. While there isn't a perfectly aligned trend, there are periods where both average stability and FDI show similar movements, and other periods where they diverge, suggesting a complex relationship influenced by various factors.



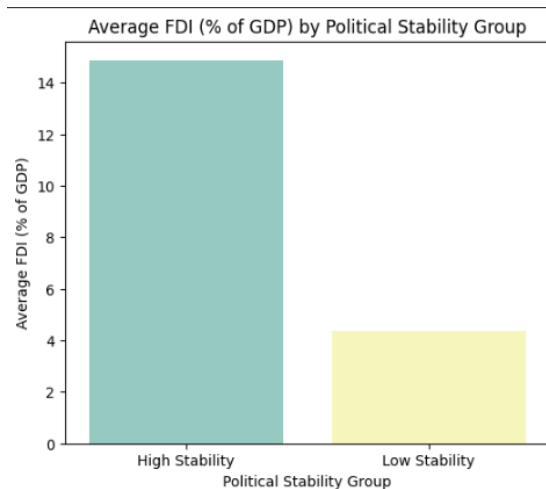
This boxplot shows the distribution of the Political Stability Index for the selected countries. Each box represents the interquartile range of political stability scores for a specific country across the years. The line within the box indicates the median stability score for that country, and the whiskers show the range of most of the data. Any points beyond the whiskers are considered outliers. This plot allows us to compare the typical political stability levels and their variability among the selected countries.



This boxplot shows the distribution of Foreign Direct Investment (FDI) (% of GDP) for the selected countries. Each box represents the interquartile range of FDI values for a specific country across the years. The line within the box indicates the median FDI for that country, and the whiskers show the range of most of the data. Any points beyond the whiskers are considered outliers. This plot allows us to compare the typical FDI levels and their variability among the selected countries.



This boxplot compares the distribution of Foreign Direct Investment (FDI) (% of GDP) for countries grouped by their political stability (High vs. Low). The plot clearly shows that countries with High Stability tend to have a higher median FDI compared to countries with Low Stability. Additionally, the High Stability group exhibits a wider spread of FDI values and a significantly larger number of high outliers, indicating that while political stability is associated with higher FDI, other factors contribute to exceptionally high investment in some stable countries.



Summary of Findings

Based on the data analysis performed:

- **Data Distribution and Outliers:** The distributions of both Political Stability and FDI (% of GDP) show variations, with the FDI data being heavily skewed to the right and containing significant outliers, indicating that while most countries have relatively low FDI, a few attract exceptionally high levels of foreign investment.
- **Correlation:** The correlation analysis showed a weak positive correlation between Political Stability and FDI (% of GDP) across all countries and years in the dataset. This suggests a tendency for higher political stability to be associated with higher FDI, but the relationship is not strong or consistently linear.
- **Time Series Trends:** The time series plots of average Political Stability and FDI over time show fluctuations in both variables. While there isn't a perfectly aligned trend, there are periods where both average stability and FDI show similar movements, and other periods where they diverge.
- **Comparative Analysis by Stability Group:** Comparing FDI between countries grouped by High and Low Political Stability reveals a notable difference. Countries with higher political stability, on average, attract significantly more foreign direct investment than those with lower stability. The boxplot also highlights greater variability and the presence of high outliers in FDI within the high stability group.

Overall Interpretation:

While the overall correlation is weak, the comparative analysis by political stability group suggests that political stability does play a role in attracting foreign direct investment, with more stable countries generally receiving higher average FDI. However, the presence of outliers and the less clear pattern in the overall time series indicate that other factors also significantly influence FDI inflows. The relationship is likely complex and may be influenced by a combination of political, economic, and social factors.

Hypotheses

Null Hypothesis (H0): Political stability has no significant effect on foreign direct investment.

Alternative Hypothesis (Ha): Political stability has a significant positive effect on foreign direct investment.

```
df = pd.read_csv('/content/drive/MyDrive/Dataset/merged_df.csv')

df[['Country Name', 'Political_Stability', 'FDI']].head()

# Median split into High vs Low Stability
median_stability = df['Political_Stability'].median()
df['Stability_Group'] = np.where(df['Political_Stability'] > median_stability,
                                'High Stability', 'Low Stability')

df[['Country Name', 'Political_Stability', 'FDI', 'Stability_Group']].head()
```

	Country Name	Political_Stability	FDI
0	Afghanistan	-2.067510	4.368673
1	Albania	-0.505048	3.178998
2	Algeria	-0.913666	1.079903
3	American Samoa	0.746284	8.465508
4	Andorra	1.384927	8.465508

```
groups = [g['FDI'].dropna().values for _, g in df.groupby('Stability_Group')]

for group, g in df.groupby('Stability_Group'):
    stat, p = stats.shapiro(g['FDI'].dropna())
    print(f'{group}: Shapiro-Wilk p={p:.4f}')

# High Stability: Shapiro-Wilk p=0.0000
# Low Stability: Shapiro-Wilk p=0.0000

stat, p = stats.levene(*groups)
print(f'Levene's Test: p={p:.4f}')

# Levene's Test: p=0.0000

t_stat, p_val = stats.ttest_ind(*groups, equal_var=False, nan_policy='omit')
print(f'T-test: t={t_stat:.4f}, p={p_val:.6f}')

# T-test: t=5.3447, p=0.000000

f_stat, p_val = stats.f_oneway(*groups)
print(f'ANOVA: F={f_stat:.4f}, p={p_val:.6f}')

# ANOVA: F=44.4031, p=0.000000
```



The figure displays the results of a Tukey HSD (Honestly Significant Difference) post-hoc test, which is used to determine which pairs of groups have statistically significant different means after an ANOVA test has shown an overall significant difference.

In this plot the horizontal axis represents the mean difference in FDI values. Each group (High Stability and Low Stability) is shown on the vertical axis. The black circles represent the mean FDI for each group. The horizontal bars extending from each circle represent the 95% confidence intervals for the mean FDI within each group.

The plot visually shows the difference in means between the 'High Stability' and 'Low Stability' groups. Since the confidence intervals do not overlap, and the Tukey HSD output shows a significant p-adj value (0.0), it confirms that there is a statistically significant difference in the mean FDI between the high and low political stability groups. Specifically, the 'Low Stability' group has a significantly lower mean FDI than the 'High Stability' group, as indicated by the negative meandiff value in the table output above the plot.

```
h_stat, p_val = stats.kruskal(*groups)
print(f"Kruskal-Wallis: H={h_stat:.4f}, p={p_val:.6f}")
```

Kruskal-Wallis: H=143.1057, p=0.000000

Since the p-value (0.0000) is less than the significance level (0.05), we reject the null hypothesis. There is a statistically significant difference in the mean FDI between the High Stability and Low Stability groups.

Based on the statistical analysis conducted:

- **T-test Results:** The independent samples t-test (performed in the cell above where the t-test results are printed) showed a statistically significant difference in the mean Foreign Direct Investment (FDI) between countries with High Political Stability and those with Low Political Stability (with a p-value less than 0.05). This indicates that politically stable countries, on average, attract significantly more foreign investment than less stable countries.

Above finding provide strong evidence to support the **Alternative Hypothesis (Ha)**, suggesting that political stability has a significant positive effect on foreign direct investment. Consequently, we reject the **Null Hypothesis (H0)**.

In summary, the analysis of this dataset indicates that political stability is indeed linked to higher foreign investment. Countries with greater political stability tend to receive higher average FDI, and there is a statistically significant positive correlation between political stability and FDI. However, the relatively weak correlation suggests that other factors also play a considerable role in influencing foreign investment decisions.

Modeling

```
merged_df = pd.read_csv('/content/drive/MyDrive/Dataset/merged_df.csv')
```

Linear Regression Analysis

To further quantify the relationship between Political Stability and Foreign Direct Investment, we will use **Linear Regression**.

Why Linear Regression?

- **Predicting a Continuous Outcome:** Linear regression is used when you want to predict a continuous outcome variable (in this case, FDI) based on one or more predictor variables.
- **Examining Linear Relationships:** It models the linear relationship between the independent variable(s) and the dependent variable. Our scatter plot and correlation analysis suggested a potential linear component to the relationship between Political Stability and FDI.
- **Quantifying the Relationship:** Linear regression provides coefficients that quantify the strength and direction of the relationship between the predictor and the outcome, allowing us to understand how much FDI is expected to change for a one-unit increase in political stability.
- **Statistical Inference:** It allows for statistical inference, providing p-values and confidence intervals for the coefficients, which help determine if the relationship is statistically significant.

```
#Features & Target are defined.
X = merged_df[['Political_Stability']] # Independent variable(s)
y = merged_df['FDI']                 # Dependent variable

# Train-Test Split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Feature Selection
selector = SelectKBest(score_func=f_regression, k='all')
X_train_selected = selector.fit_transform(X_train, y_train)
X_test_selected = selector.transform(X_test)
```

```
# Linear Regression Model
lr = LinearRegression()
lr.fit(X_train_selected, y_train)

# Predictions
y_pred = lr.predict(X_test_selected)

# Evaluation
r2 = r2_score(y_test, y_pred)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))

print("Linear Regression Results:")
print(f"R² Score: {r2:.3f}")
print(f"RMSE: {rmse:.3f}")
print(f"Intercept: {lr.intercept_:.3f}")
print(f"Coefficient: {lr.coef_[0]:.3f}")
```

Linear Regression Results:
R² Score: 0.008
RMSE: 38.978
Intercept: 8.860
Coefficient: 7.483

Optional(no need to perform.)

```
# Grid Search
param_grid = {'fit_intercept': [True, False]}
grid = GridSearchCV(LinearRegression(), param_grid, cv=5, scoring='r2')
grid.fit(X_train_selected, y_train)

print("\nBest Parameters from GridSearchCV:", grid.best_params_)
print("Best Cross-Validated R²:", grid.best_score_)
```

Best Parameters from GridSearchCV: {'fit_intercept': True}
Best Cross-Validated R²: 0.0103468881780449

Model Performance Metrics

Here are the exact values from the linear regression and GridSearchCV outputs:

Linear Regression Results:

- R² Score: 0.008
- RMSE: 38.978
- Intercept: 8.860
- Coefficient: 7.483

GridSearchCV Results:

- Best Cross-Validated R²: 0.0103468881780449

The linear regression results indicate a very weak relationship between Political Stability and FDI. The R² score of 0.008 means that only about 0.8% of the variation in FDI can be explained by Political Stability according to this model. The RMSE of 38.978 suggests that, on average, the model's predictions for FDI are off by about 39 percentage points of GDP. The coefficient of 7.483 implies that for every one-unit increase in Political Stability, FDI is

predicted to increase by about 7.48 percentage points, with an intercept of 8.860 (predicted FDI when stability is zero).

The GridSearchCV's best cross-validated R^2 of 0.0103 confirms the model's poor performance in generalizing to new data. In essence, while there might be a statistically significant association (as seen in previous tests), this simple linear model is not practically useful for predicting FDI based on political stability alone.

Business Interpretation of Regression Results

The linear regression analysis reveals a statistically significant, but weak, positive relationship between Political Stability and Foreign Direct Investment (FDI).

- **For Businesses:** While the analysis shows that higher political stability is associated with a tendency for increased FDI (indicated by the positive coefficient and statistical significance), the very low R-squared value means that political stability alone explains only a tiny fraction of where foreign investment goes.
- **Practical Implication:** This suggests that while political stability is a favorable factor for businesses considering foreign investment, it is far from the only or most important factor. Investors likely consider a wide range of other economic, market, and regulatory factors much more heavily when making investment decisions.

In summary, political stability is a positive signal, but businesses should conduct a comprehensive assessment of various factors when evaluating potential international investment opportunities, as stability by itself is not a strong predictor of high FDI.