Libraries Import

```
In [1]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns
        import warnings
        warnings.simplefilter("ignore")
```

Data Import

```
In [2]: df = pd.read_csv('internet_users.csv')
```

```
In [3]: df.head()
```

Out[3]:

	entity	code	year	users	share
0	Afghanistan	AFG	1991	0	0.000000
1	Afghanistan	AFG	1992	0	0.000000
2	Afghanistan	AFG	1993	0	0.000000
3	Afghanistan	AFG	1994	0	0.000000
4	Afghanistan	AFG	2001	930	0.004723

Exploratory Data Analysis

```
In [4]: avg_share_by_country = df.groupby('entity')['share'].mean()
```

```
In [5]: top_5_countries = avg_share_by_country.nlargest(5)
```

```
In [6]: print("Top 5 Countries with the Highest Internet Use (by Population Share) from 1990-2020:")
        print(top_5_countries)

        Top 5 Countries with the Highest Internet Use (by Population Share) from 1990-2020:
        entity
        Kosovo     86.668315
        Curacao    64.991360
        Iceland    64.677995
        Norway     62.985167
        Sweden     61.373734
        Name: share, dtype: float64
```

Analysis: Kosovo emerges as the leading nation with the highest percentage of population using the internet, closely trailed by Curacao, Iceland, Norway, and Sweden.

```python
In [7]: filtered_data = df[(df['year'] == 2020) & (df['code'].notna()) & (df['code'] != 'OWID_WRL')]
```

```python
In [8]: total_users_by_country = filtered_data.groupby('entity')['users'].sum().reset_index()
```

```python
In [9]: top_5_countries = total_users_by_country.nlargest(5, 'users')
```

```python
In [10]: print("Top 5 Countries with the Highest Internet Use in 2020 :")
         print(top_5_countries[['entity', 'users']])
```

```
Top 5 Countries with the Highest Internet Use in 2020 :
              entity       users
30             China  1003218650
64             India   600446441
144    United States   305371298
19            Brazil   173419624
65         Indonesia   146059763
```

```python
In [11]: import plotly.express as px
```

```python
In [12]: top_5 = df[(df['code'].notna()) & (df['code'] != 'OWID_WRL')]
```

```python
In [13]: top_5_entities = top_5.groupby('entity')['users'].sum().nlargest(5).index
```

```python
In [14]: top_5_data = top_5[top_5['entity'].isin(top_5_entities)]
```

```python
In [15]: top_5_data = top_5[top_5['entity'].isin(top_5_entities)]
```

```
In [16]: fig = px.line(top_5_data, x='year', y='users', color='entity',
                        title='Top 5 Internet Users Countries Over Time',
                        labels={'users': 'Number of Internet Users', 'year': 'Year'})
         fig.show()
```

Top 5 Internet Users Countries Over Time
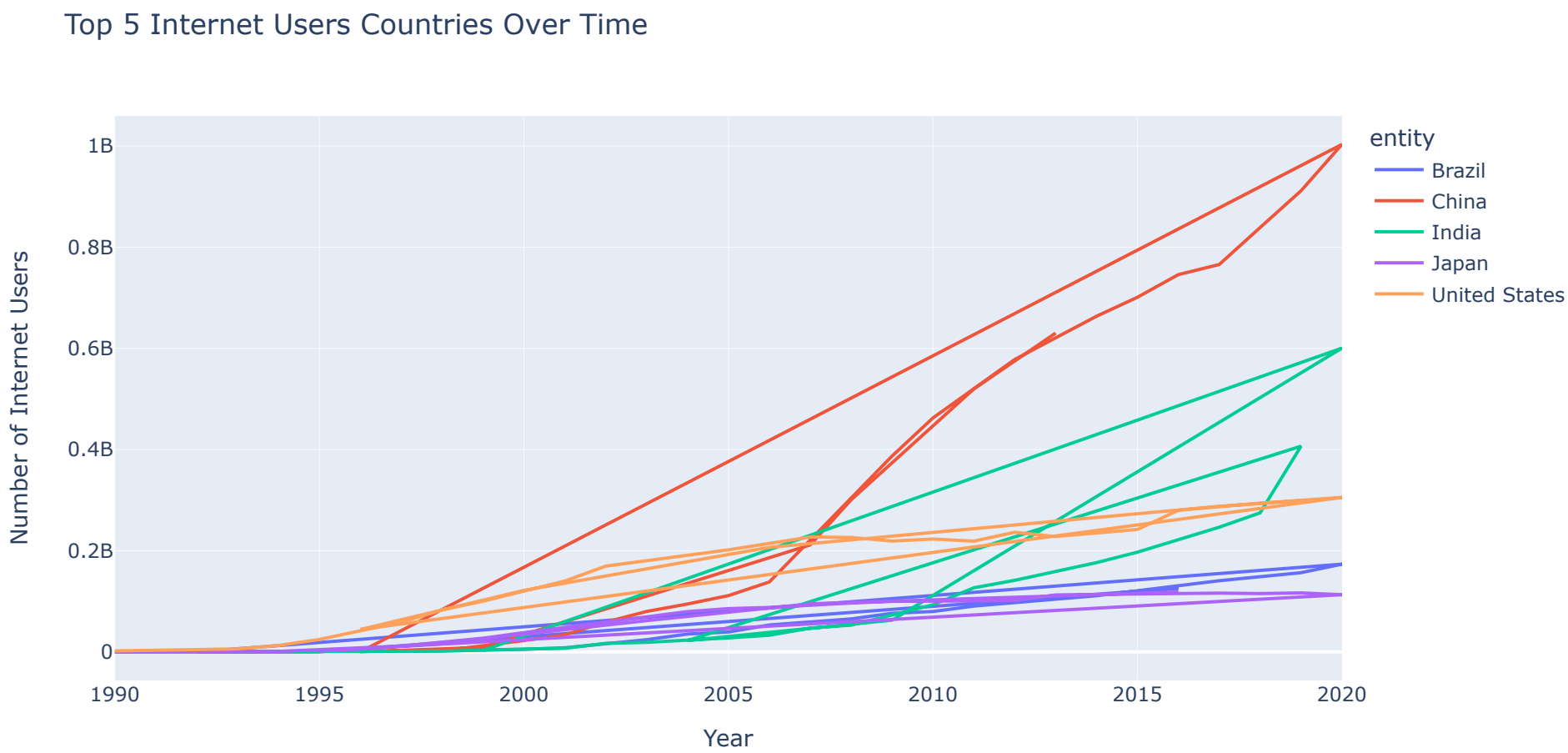
Analysis: The data reveals a significant shift in internet usage trends. Post-1998, China surpassed the United States, securing its position as the leading internet user country. Notably, India claimed the second spot on the list. In the latest data from 2020, China maintains its dominance, with India, the United States, Brazil, and Japan following suit in internet usage rankings. This evolution highlights the dynamic landscape of global internet adoption over the years.

```
In [17]: df_2020 = df[df['year'] == 2020]
```

```
In [18]: selected_columns = ['users', 'share']
```

```
In [19]: correlation_matrix = df_2020[selected_columns].corr()
```

```
In [20]: plt.figure(figsize=(8, 6))
         sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f", linewidths=.5)
         plt.title('Correlation Heatmap for Internet Usage and Population Share in 2020')
         plt.show()
```

Correlation Heatmap for Internet Usage and Population Share in 2020
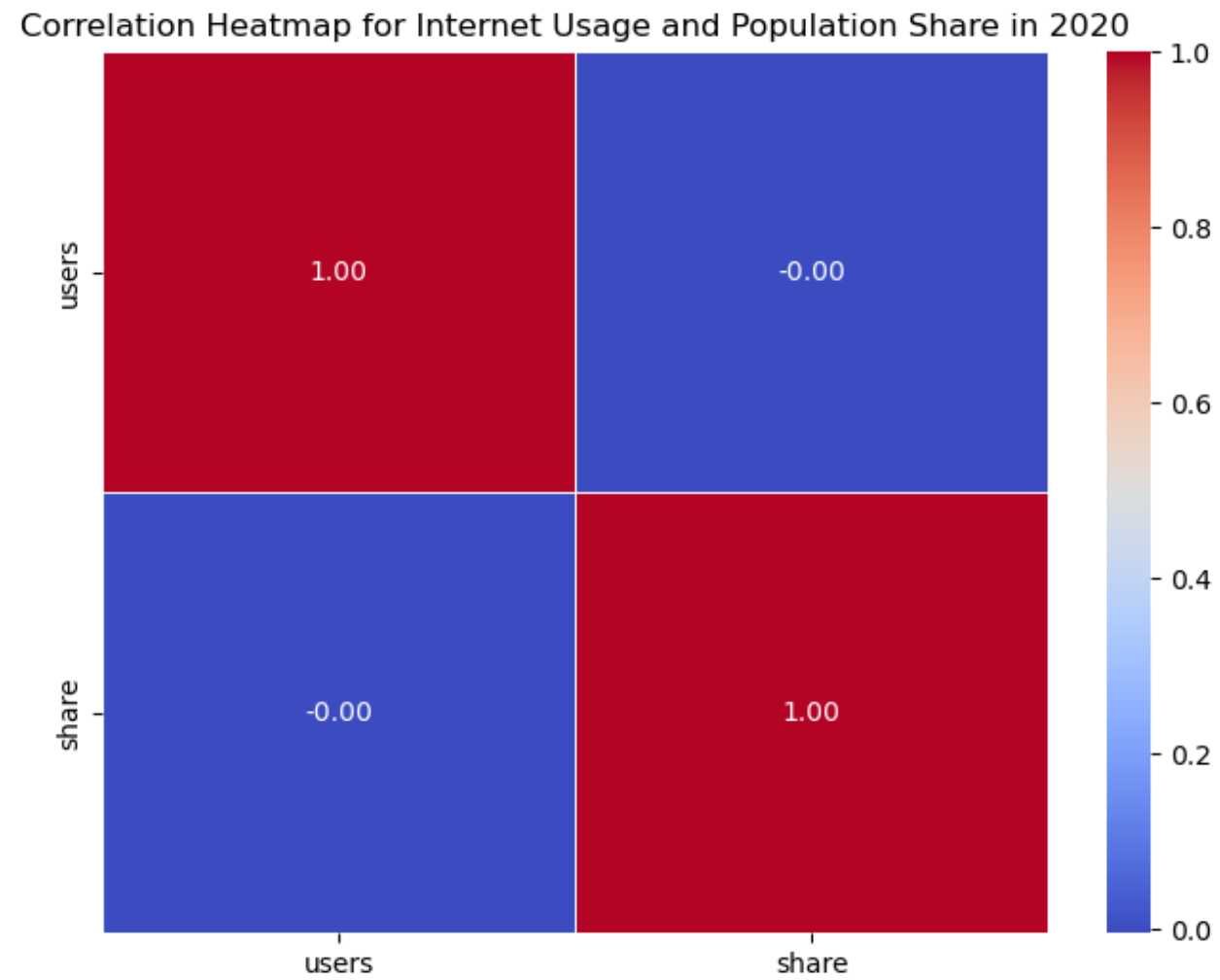

```
         pip install geopandas
```

```
In [21]: import geopandas as gpd
```

```
In [22]: world = gpd.read_file(gpd.datasets.get_path('naturalearth_lowres'))
```

```
In [23]: merged_data = world.merge(df, how='left', left_on='iso_a3', right_on='code')
```

```
In [24]: year_to_visualize = 2020
         filtered_data = merged_data[merged_data['year'] == year_to_visualize]
```

```
In [25]: fig, ax = plt.subplots(1, 1, figsize=(15, 10))
         filtered_data.plot(column='share', cmap='Blues', linewidth=0.8, ax=ax, edgecolor='0.8', legend=True)
         ax.set_title(f'Internet Usage Around the World in {year_to_visualize}')
         ax.set_xlabel('Longitude')
         ax.set_ylabel('Latitude')
         plt.show()
```
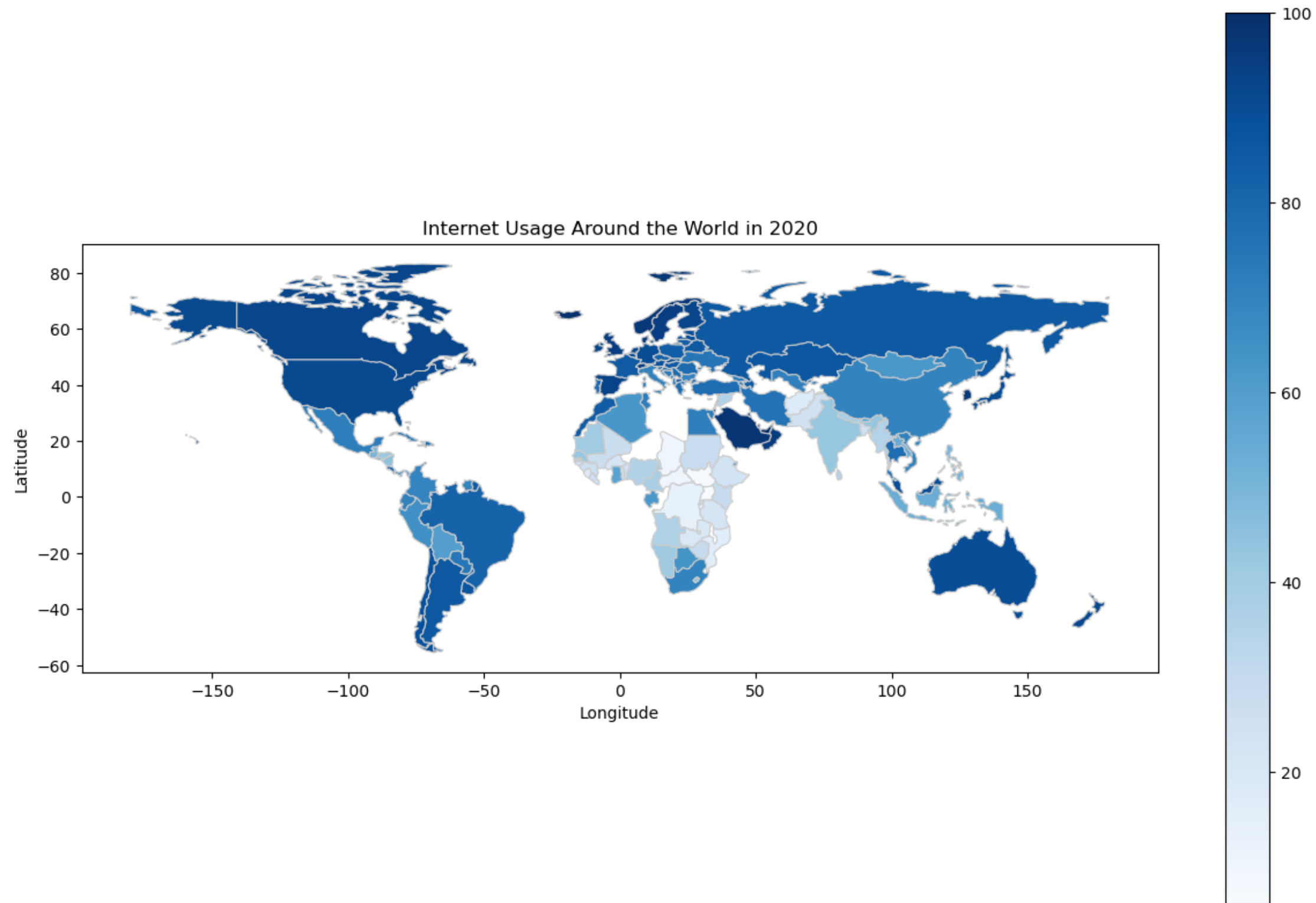
Observation: A clear pattern emerges as we observe internet usage across continents. In Asia, Saudi Arabia, Taiwan, Japan, and Malaysia stand out. Oceania is represented by Australia and New Zealand. In Europe, countries like the United Kingdom, Ireland, Iceland, Finland, Norway, Sweden, and Spain exhibit higher internet usage. North America, including the United States and Canada, demonstrates notable usage. Meanwhile, South America sees prominent internet activity in Brazil, Argentina, and Chile. This analysis underscores the varying degrees of internet adoption across different regions, reflecting the global connectivity landscape.

Conclusion:

After the above analysis I came to the conclusion that China, India, United States, Brazil, and Japan are the countries where there is a big opportunity to roll out our services.

Reason Why Service Provider Should Focus On These Countries:

Internet providers may prioritize offering services in China, India, the United States, Japan, and Brazil for several compelling reasons:

1) Large Population
2) Economic Potential
3) Technological Adoption
4) Urbanization and Connectivity
5) Growing Middle Class
6) Business Opportunities
7) Government Initiatives
8) Competitive Landscape

What if there is no internet for the whole day

No Internet

Top 10 countries where a day without internet costs the most

Country	Cost
United States	$11.0Bn
China	$9.9Bn
United Kingdom	$3.3Bn
Japan	$2.7Bn
Germany	$1.5Bn
India	$1.4Bn
South Koreo	$1.3Bn
France	$0.9Bn
Italy	$0.6Bn
Canada	$0.6Bn

If an internet outage occurs just for a single day, it would cost $43 billion globally.

Source: NetBlocks, 2023

moneyfront