

```
In [97]: import pandas as pd
import cufflinks as cf
import seaborn as sn
from IPython.display import display, HTML

cf.set_config_file(sharing="public", theme="space", offline=True)
```

```
In [98]: # Importing dataframes
df_internet = pd.read_csv("final.csv")
df_gdp = pd.read_csv("world_gdp.csv")
```

The data above were retrieved from kaggle:

- df\_internet = <https://www.kaggle.com/datasets/ashishraut64/internet-users>
- df\_gdp = <https://www.kaggle.com/datasets/tmishinev/world-country-gdp-19602021>

```
In [99]: df_internet.sample(5)
```

```
Out[99]:
```

Unnamed: 0	Entity	Code	Year	Cellular Subscription	Internet Users(%)	No. of Internet Users	Broadband Subscription	
7624	7624	Sudan	SDN	2020	80.264084	28.400000	12621097	0.065639
6388	6388	Peru	PER	1989	0.000000	0.000000	0	0.000000
3684	3684	India	IND	2013	69.196991	12.300000	158809242	1.165457
5490	5490	Morocco	MAR	1993	0.025712	0.000000	0	0.000000
2909	2909	Gambia	GMB	1997	0.394327	0.051311	676	0.000000

```
In [100]: df_gdp.sample(5)
```

```
Out[100]:
```

	Country Name	Country Code	year	GDP_USD	GDP_per_capita_USD
1739	Lithuania	LTU	1966	NaN	NaN
10927	Bulgaria	BGR	2001	1.418350e+10	1770.913534
6543	Malta	MLT	1984	1.101829e+09	3332.885357
11941	Timor-Leste	TLS	2004	4.407646e+08	453.513131
11075	Malaysia	MYS	2001	9.278395e+10	3913.429386

```
In [101]: # Renaming column for merging dataframes and reference
df_gdp.rename(columns={"year": "Year",
                        "Country Code": "Country_Code"}, inplace=True)

df_internet.rename(columns={"Code": "Country_Code"}, inplace=True)
```

```
# Merge dataframes
df_main = df_internet.merge(df_gdp[["Year", "Country_Code", "GDP_USD", "GDP_per_capita_USD"]],
df_main.sample(5)
```

```
Out[101]:
```

Unnamed: 0	Entity	Country_Code	Year	Cellular Subscription	Internet Users(%)	No. of Internet Users	Broadband Subscription	GDP_USD	GDP_per_capita_USD
3329	3329	Guinea-Bissau	GNB	1985	0.000000	0.000000	0	0.000000	1.43
3551	3551	Hong Kong	HKG	2003	109.287064	52.200043	3594737	18.855419	1.61
830	830	Bermuda	BMU	1982	0.000000	0.000000	0	0.000000	7.85
3263	3263	Guatemala	GTM	2000	7.354303	0.712333	83599	0.000000	1.92
6053	6053	North Macedonia	MKD	2002	17.831081	17.330000	357819	0.000000	4.01

```
In [102]: # Filter ASEAN Countries
asean_list = ["IDN", "KHM", "SGP", "VNM", "MMR", "THA", "BRN", "LAO", "MYS", "PHL"]
df_asean = df_main[df_main["Country_Code"].isin(asean_list)]
df_asean.sample(5)
```

```
Out[102]:
```

Unnamed: 0	Entity	Country_Code	Year	Cellular Subscription	Internet Users(%)	No. of Internet Users	Broadband Subscription	GDP_USD	GDP_per_capita_USD
6420	6420	Philippines	PHL	1980	0.000000	0.000000	0	0.000000	3.0
5570	5570	Myanmar	MMR	1991	0.000000	0.000000	0	0.000000	2.0
7906	7906	Thailand	THA	2015	149.811157	39.316128	27637039	9.065043	4.0
3731	3731	Indonesia	IDN	2019	126.106911	47.690647	128565820	3.800219	1.0
1100	1100	Brunei	BRN	1987	0.000000	0.000000	0	0.000000	2.0

```
In [103]: # checking for null values
df_asean.isnull().sum()
```

```
Out[103]:
```

Unnamed: 0	0
Entity	0
Country_Code	0
Year	0
Cellular Subscription	0
Internet Users(%)	0
No. of Internet Users	0
Broadband Subscription	0
GDP_USD	22
GDP_per_capita_USD	22
dtype:	int64

```
In [104]: # finding null locations
temp = df_asean.isnull().any(axis=1)
temp[temp].index
```

```
Out[104]:
```

Index([1256, 1257, 1258, 1259, 1260, 1261, 1262, 1263, 1264, 1265, 1266, 1267, 1268, 4262, 4263, 4264, 4265, 8662, 8663, 8664, 8665, 8666],	
dtype='int64')	

```
In [105]: # reviewing null values
df_asean.loc[[1256, 1257, 1258, 1259, 1260, 1261, 1262, 1263, 1264, 1265, 1266, 1267, 1268, 4262, 4263, 4264, 4265, 8662, 8663, 8664, 8665, 8666], :]
```

```
Out[105]:
```

Unnamed: 0	Entity	Country_Code	Year	Cellular Subscription	Internet Users(%)	No. of Internet Users	Broadband Subscription	GDP_USD	GDP_per_capita_USD
1256	1256	Cambodia	KHM	1980	0.0	0.0	0	0.0	
1257	1257	Cambodia	KHM	1981	0.0	0.0	0	0.0	
1258	1258	Cambodia	KHM	1982	0.0	0.0	0	0.0	
1259	1259	Cambodia	KHM	1983	0.0	0.0	0	0.0	
1260	1260	Cambodia	KHM	1984	0.0	0.0	0	0.0	
1261	1261	Cambodia	KHM	1985	0.0	0.0	0	0.0	
1262	1262	Cambodia	KHM	1986	0.0	0.0	0	0.0	
1263	1263	Cambodia	KHM	1987	0.0	0.0	0	0.0	
1264	1264	Cambodia	KHM	1988	0.0	0.0	0	0.0	
1265	1265	Cambodia	KHM	1989	0.0	0.0	0	0.0	
1266	1266	Cambodia	KHM	1990	0.0	0.0	0	0.0	
1267	1267	Cambodia	KHM	1991	0.0	0.0	0	0.0	
1268	1268	Cambodia	KHM	1992	0.0	0.0	0	0.0	
4262	4262	Laos	LAO	1980	0.0	0.0	0	0.0	
4263	4263	Laos	LAO	1981	0.0	0.0	0	0.0	
4264	4264	Laos	LAO	1982	0.0	0.0	0	0.0	
4265	4265	Laos	LAO	1983	0.0	0.0	0	0.0	
8662	8662	Vietnam	VNM	1980	0.0	0.0	0	0.0	
8663	8663	Vietnam	VNM	1981	0.0	0.0	0	0.0	
8664	8664	Vietnam	VNM	1982	0.0	0.0	0	0.0	
8665	8665	Vietnam	VNM	1983	0.0	0.0	0	0.0	
8666	8666	Vietnam	VNM	1984	0.0	0.0	0	0.0	

```
In [106]: df_asean.dropna(inplace=True)
```

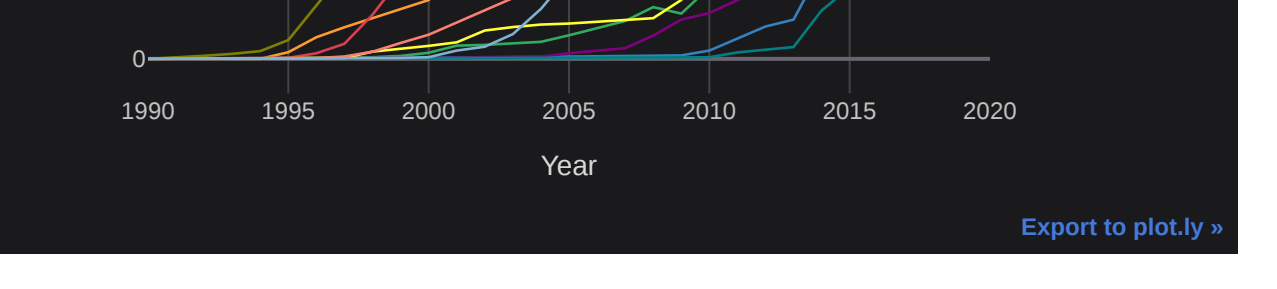
```
In [107]: # filter year
df_asean = df_asean[df_asean["Year"] > 1989]
```

## Southeast Asia Internet User Growth

```
In [108]: # pivot table
a = df_asean.pivot(index="Year", columns="Entity", values="Internet Users(%)")

# fill 0 values
a.loc[2018, "Cambodia"] = a.loc[2018, :].median()
a.loc[2019, "Cambodia"] = a.loc[2019, :].median()
a.loc[2020, "Cambodia"] = a.loc[2020, :].median()

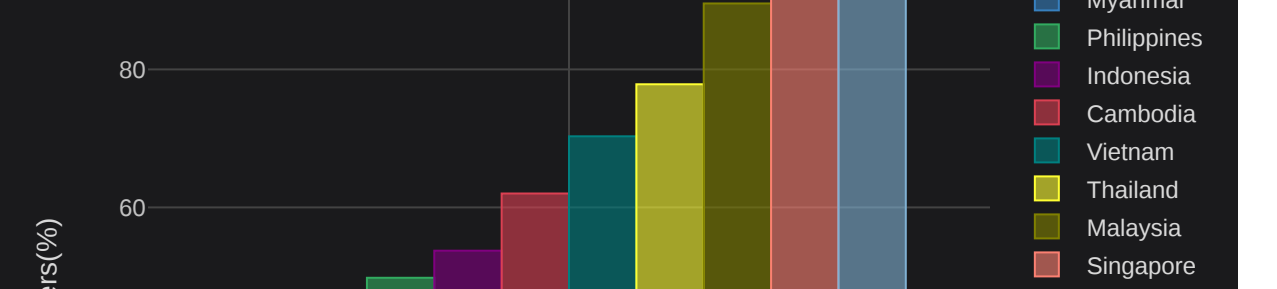
a.plot(kind="line", xTitle="Year", yTitle="Users(%)")
```



Countries that were early adopters of the internet, such as Brunei, Singapore, and Malaysia, generally have a higher percentage of internet users than others. Additionally, in 2015, there was a significant increase in user percentages in each country. There is a possibility that this sharp increase was caused by several factors, such as increased smartphone adoption, the rise of social media, and e-commerce growth.

## Internet User Percentage Rank in Southeast Asia (2020)

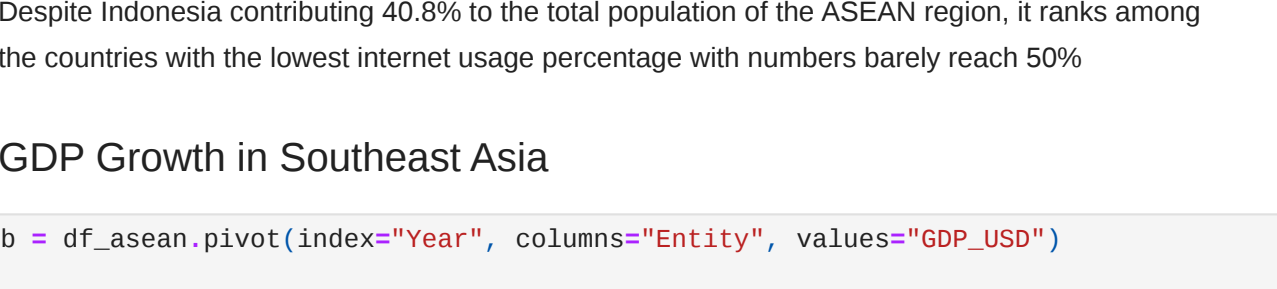
```
In [109]: a[a.index.isin([2020]).sort_values(2020, axis=1).plot(kind="bar", xTitle="Year",
```



Despite Indonesia contributing 40.8% to the total population of the ASEAN region, it ranks among the countries with the lowest internet usage percentage with numbers barely reach 50%

## GDP Growth in Southeast Asia

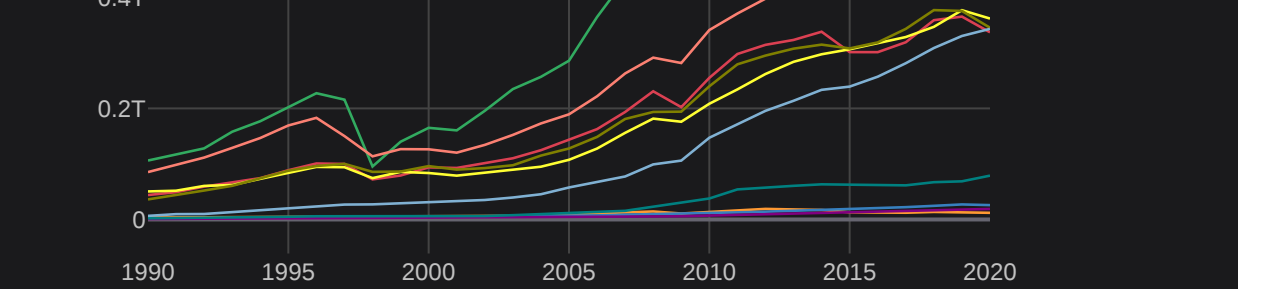
```
In [110]: b = df_asean.pivot(index="Year", columns="Entity", values="GDP_USD")
b.plot(kind="line", yTitle="GDP in USD", xTitle="Year")
```



Although Indonesia has the highest GDP, only 53.73% of the population in Indonesia uses the internet. This data may indicate that there is a significant economic disparity in Indonesia or a low level of digital literacy.

## GDP per Capita In Southeast Asia

```
In [111]: c = df_asean.pivot(index="Year", columns="Entity", values="GDP_per_capita_USD")
c[c.index.isin([2020]).sort_values(2020, axis=1).plot(kind="bar", yTitle="GDP/Ca
```



By examining the GDP per Capita and internet User Percentage (2020) barchart, we can conclude that countries with a high GDP per capita tend to have higher internet user percentage. Also, the chart above confirms the hypothesis about the economic disparity that is happening in Indonesia is true

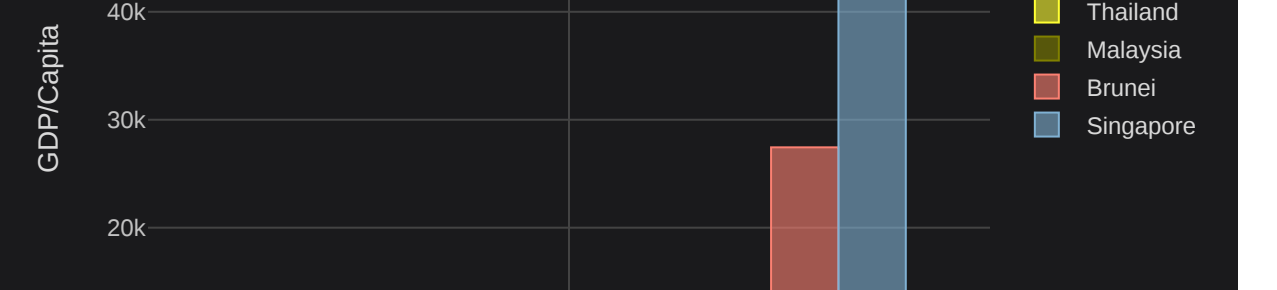
## Linear Regression Model

```
In [112]: # removing outliers
max_limit = df_asean["Internet Users(%)"].mean() + 3 * df_asean["Internet Users(%)"]
min_limit = df_asean["Internet Users(%)"].mean() - 3 * df_asean["Internet Users(%)"]

df_asean2 = df_asean[(df_asean["Internet Users(%)"] < max_limit) & (df_asean["Inte
```

```
In [113]: sm.regplot(data=df_asean2, x="Internet Users(%)", y="GDP_per_capita_USD")
```

```
Out[113]: <Axes: xlabel='Internet Users(%)', ylabel='GDP_per_capita_USD'>
```



The model above shows that the GDP per capita of a country has a significant influence on the percentage of internet users in that respective country.

```
In [ ]:
```