Metro Train systems around the world

Date- April 8, 2022

Author- Sanjay Chauhan

Introduction

Metro rails network is widely spreading across the country at a rapid pace. From metropolitan cities to tier-2 cities the network is getting mammoth and it's not just in India rather it is pervasive all around the world. But it is not necessary that every country is heading towards the same destination, there could be some differences according to the country's fundamental position, requirements, etc. For instance, there could be differences in the timeline of network development and other phases and so on. This project comes with all the questions and answers regarding metro systems around the globe. The report consists of data analysis, and presentation through visualization to answer all those questions.

In [1]:
```
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns
```

Data processing

Let's see the data

In [2]:
```
df=pd.read_csv("metro_countries_cities.csv")
```

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

Out[3]:

	city	country	name	year	year_last_expansion	stations	length_km	annual_ridership_mill	region
0	Algiers	Algeria	Algiers Metro	2011	2018	19	185	45.30	africa
1	Buenos Aires	Argentina	Buenos Aires Underground	1913	2019	90	567	74.00	latin_america
2	Yerevan	Armenia	Yerevan Metro	1981	1996	10	134	10.75	asia
3	Sydney	Australia	Sydney Metro	2019	2019	13	36	12.90	australia
4	Vienna	Austria	Vienna U-Bahn	1976	2017	98	833	459.80	europe

The column "region" has some inconsistent names within it so it needs to get rectified for better presentation.

In [4]:
```
df.region=df.region.str.replace("_"," ")
```

In [5]:
```
df.head()
```

Out[5]:

	city	country	name	year	year_last_expansion	stations	length_km	annual_ridership_mill	region
0	Algiers	Algeria	Algiers Metro	2011	2018	19	185	45.30	africa
1	Buenos Aires	Argentina	Buenos Aires Underground	1913	2019	90	567	74.00	latin america
2	Yerevan	Armenia	Yerevan Metro	1981	1996	10	134	10.75	asia
3	Sydney	Australia	Sydney Metro	2019	2019	13	36	12.90	australia
4	Vienna	Austria	Vienna U-Bahn	1976	2017	98	833	459.80	europe

In [6]:
```
df.shape
```

Out[6]:(198, 9)

In [7]:
```
df.loc[df.duplicated(),:]
```

Out[7]:

	city	country	name	year	year_last_expansion	stations	length_km	annual_ridership_mill	region

In [8]:
```
df.isna().sum()
```

Out[8]:
```
city                  0
country               0
name                  0
year                  0
year_last_expansion   0
stations              0
length_km             0
annual_ridership_mill 0
region                0
dtype: int64
```

Raw data does not contain any wrong datatype neither it has inconsistent columns. Data duplication also checked and it does not have duplicated elements as well. After processing and rechecking the data it is clean enough for analysis phase.

Data analysis and visualization

1. Countries with most number of metro systems

In [9]:
```
top_countries=df.country.value_counts().head(10)
top_countries=top_countries.to_frame(name="number_of_systems")
top_countries.index.name="country"
top_countries.reset_index(inplace=True)
```

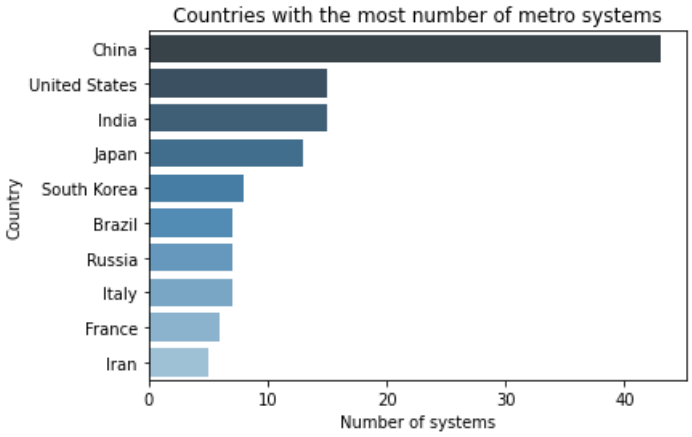
In [10]:
```
top_countries.head()
```

Out[10]:

	country	number_of_systems
0	China	43
1	United States	15
2	India	15
3	Japan	13
4	South Korea	8

In [11]:
```
sns.barplot(x='number_of_systems',y='country',data=top_countries,palette='Blues_r_d')
plt.xlabel("Number of systems")
plt.ylabel("Country")
plt.title("Countries with the most number of metro systems")
```

Out[11]:Text(0.5, 1.0, 'Countries with the most number of metro systems')



2. A comparison and analysis of top rated countries

In [12]:
```
cntry_gwth= df[["country","year"]]
```

In [13]:
```
cntry_gwth=cntry_gwth.value_counts()
```

In [14]:
```
cntry_gwth=cntry_gwth.to_frame(name="Number of systems")
cntry_gwth.reset_index(inplace=True)
```

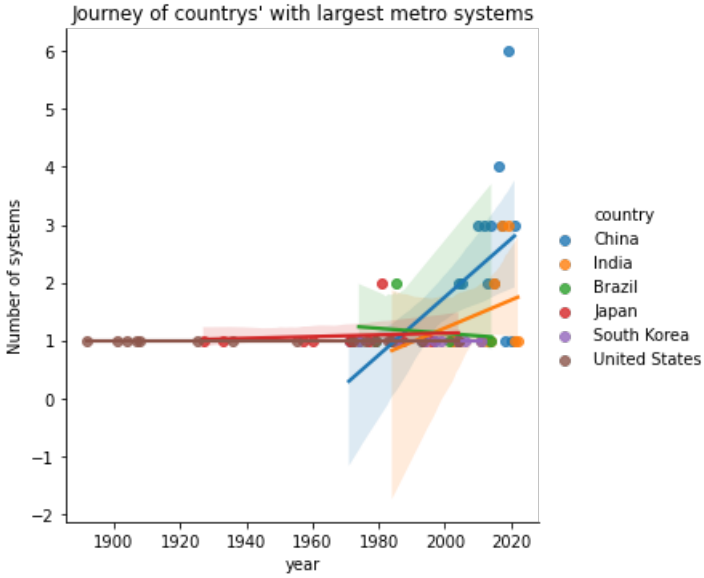
In [37]:
```
cntry_gwth=cntry_gwth[cntry_gwth.country.isin(["China","India","United States","Japan","South Korea","Brazil"])]
cntry_gwth.head()
```

Out[37]:

	country	year	Number of systems
0	China	2019	6
1	China	2016	4
2	China	2010	3
3	China	2012	3
4	China	2014	3

In [16]:
```
sns.lmplot(x="year",y="Number of systems",data=cntry_gwth,hue="country")
plt.title("Journey of countrys' with largest metro systems")
```

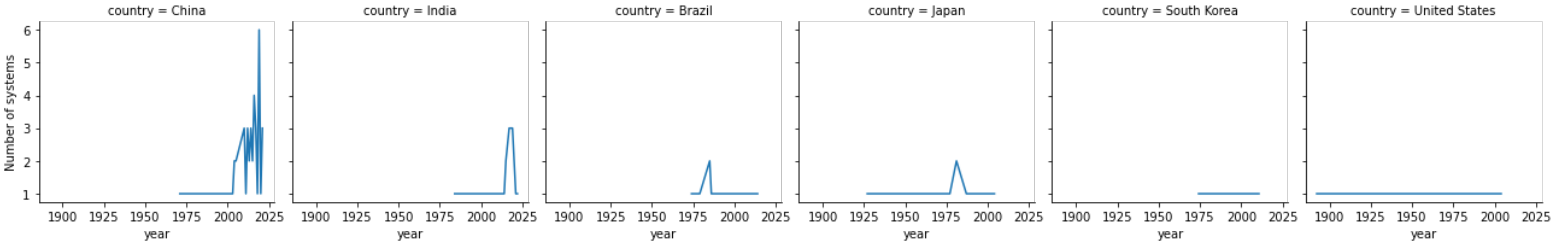
Out[16]:Text(0.5, 1.0, "Journey of countrys' with largest metro systems")



A Separate comparison among top countries

In [36]:
```
x=sns.FacetGrid(data=cntry_gwth,col="country")
x.map(sns.lineplot,"year","Number of systems")
```

Out[36]:<seaborn.axisgrid.FacetGrid at 0x27b7a9a5210>



3. Oldest metro systems of the world

In [18]:
```
oldest_metros= df.sort_values(by="year").head(10)
```

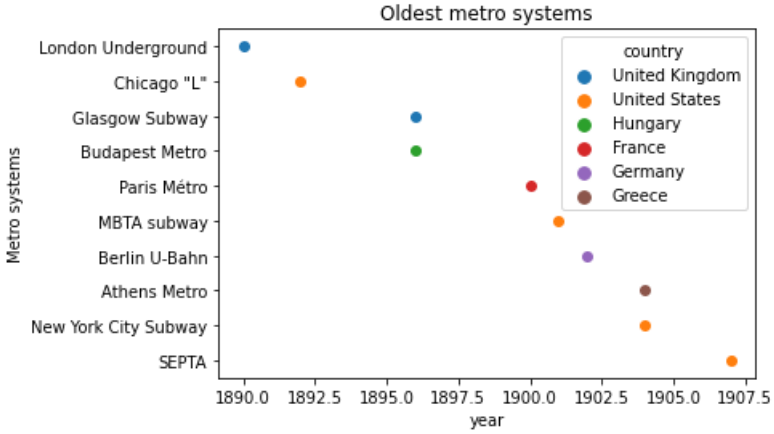
In [19]:
```
oldest_metros
```

Out[19]:

	city	country	name	year	year_last_expansion	stations	length_km	annual_ridership_mill	region
178	London	United Kingdom	London Underground	1890	2021	272	4052	296.0	europe
183	Chicago	United States	Chicago "L"	1892	2017	145	1654	76.0	north america
177	Glasgow	United Kingdom	Glasgow Subway	1896	1896	15	104	12.7	europe
81	Budapest	Hungary	Budapest Metro	1896	2014	48	382	232.8	europe
72	Paris	France	Paris Métro	1900	2022	306	2251	753.0	europe
182	Boston	United States	MBTA subway	1901	2014	51	61	57.5	north america
76	Berlin	Germany	Berlin U-Bahn	1902	2021	175	1478	596.0	europe
80	Athens	Greece	Athens Metro	1904	2020	64	887	259.2	europe
187	New York City	United States	New York City Subway	1904	2017	424	399	639.5	north america
190	Philadelphia	United States	SEPTA	1907	1973	75	591	37.7	north america

In [20]:
```
sns.swarmplot(x="year",y="name",data=oldest_metros,hue="country",size=7)
plt.title("Oldest metro systems")
plt.ylabel("Metro systems")
```

Out[20]:Text(0, 0.5, 'Metro systems')



4. Metro systems with most number of stations

In [21]:
```
most_stns=df.sort_values(by="stations",ascending=False).head(10)
```

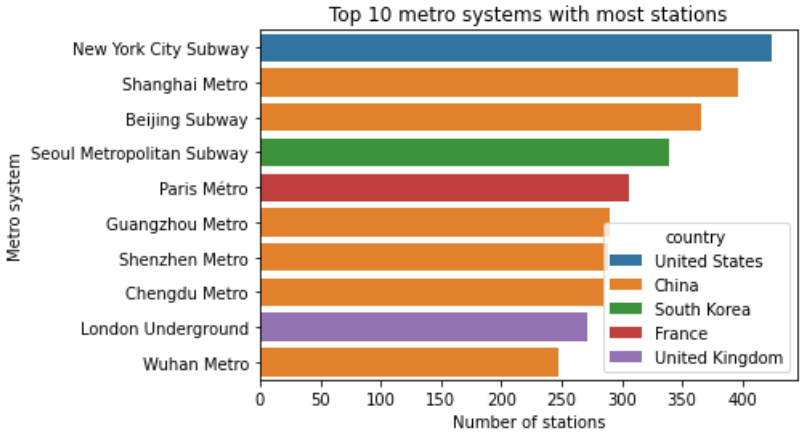
In [22]:
```
most_stns.head()
```

Out[22]:

	city	country	name	year	year_last_expansion	stations	length_km	annual_ridership_mill	region
187	New York City	United States	New York City Subway	1904	2017	424	399	639.50	north america
46	Shanghai	China	Shanghai Metro	1993	2021	396	803	2834.69	asia
20	Beijing	China	Beijing Subway	1971	2021	366	762	2292.65	asia
130	Seoul	South Korea	Seoul Metropolitan Subway	1974	2021	339	3649	2127.20	asia
72	Paris	France	Paris Métro	1900	2022	306	2251	753.00	europe

In [23]:
```
sns.barplot(x="stations",y="name",data=most_stns,hue="country",dodge=False)
plt.xlabel("Number of stations")
plt.ylabel("Metro system")
plt.title("Top 10 metro systems with most stations")
```

Out[23]:Text(0.5, 1.0, 'Top 10 metro systems with most stations')



5. Ranking of metro systems on the bases of annual ridership

In [24]:
```
annual_rides=df.sort_values(by="annual_ridership_mill",ascending=False).head(10)
```

In [25]:
```
annual_rides
```

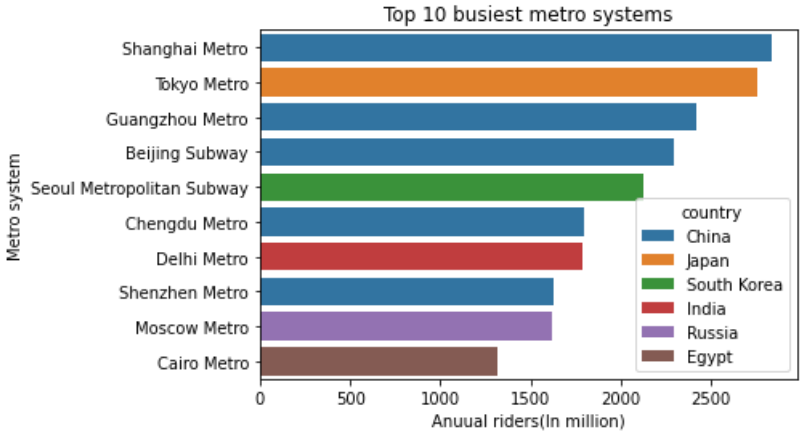
Out[25]:

	city	country	name	year	year_last_expansion	stations	length_km	annual_ridership_mill	region
46	Shanghai	China	Shanghai Metro	1993	2021	396	803	2834.69	asia
119	Tokyo	Japan	Tokyo Metro	1927	2020	142	1951	2757.40	asia
30	Guangzhou	China	Guangzhou Metro	1997	2021	290	5894	2415.60	asia
20	Beijing	China	Beijing Subway	1971	2021	366	762	2292.65	asia
130	Seoul	South Korea	Seoul Metropolitan Subway	1974	2021	339	3649	2127.20	asia
24	Chengdu	China	Chengdu Metro	2010	2020	285	5192	1800.00	asia
85	Delhi	India	Delhi Metro	2002	2021	230	34812	1790.00	asia
49	Shenzhen	China	Shenzhen Metro	2004	2021	288	419	1626.73	asia
150	Moscow	Russia	Moscow Metro	1935	2021	203	4357	1618.20	asia
67	Cairo	Egypt	Cairo Metro	1987	2020	71	894	1314.00	africa

In [26]:
```
sns.barplot(x="annual_ridership_mill",y="name",data=annual_rides,hue="country",dodge=False)
plt.xlabel("Anuual riders(In million)")
```

```
plt.ylabel("Metro system")
plt.title("Top 10 busiest metro systems")
```

Out[26]:Text(0.5, 1.0, 'Top 10 busiest metro systems')



6. Yearly progress of metro systems around the globe

In [27]:
```
yrly_progression=df.year.value_counts()
yrly_progression=yrly_progression.to_frame(name="Number of systems")
yrly_progression.index.name="Year"
yrly_progression.reset_index(inplace=True)
```

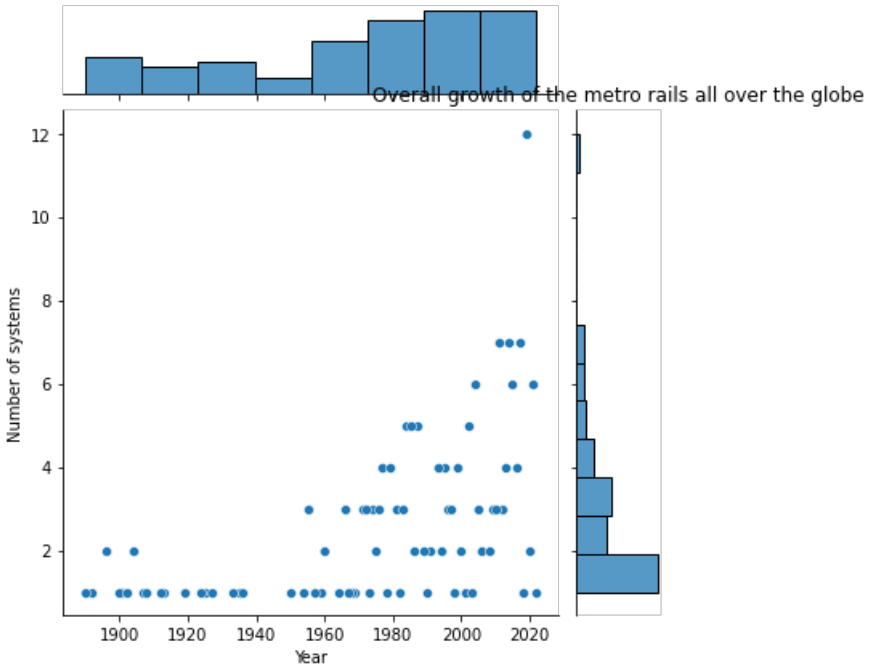
In [38]:
```
yrly_progression.head()
```

Out[38]:

	Year	Number of systems
0	2019	12
1	2011	7
2	2014	7
3	2017	7
4	2004	6

In [29]:
```
sns.jointplot(x="Year",y="Number of systems",data=yrly_progression,kind="scatter")
plt.title("Overall growth of the metro rails all over the globe")
```

Out[29]:Text(0.5, 1.0, 'Overall growth of the metro rails all over the globe')



7. Regions with the number of metro systems at present

In [30]:
```
top_region= df.region.value_counts()
top_region=top_region.to_frame(name="Number of systems")
top_region.index.name="Region"
top_region.reset_index(inplace=True)
```

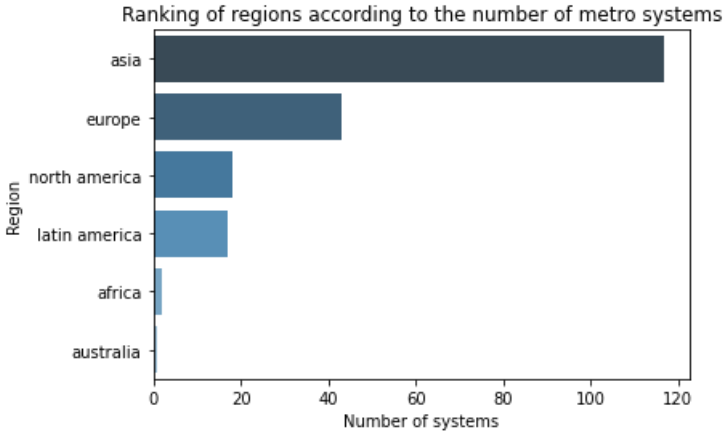
In [31]:
```
top_region
```

Out[31]:

	Region	Number of systems
0	asia	117
1	europe	43
2	north america	18
3	latin america	17
4	africa	2
5	australia	1

In [32]:
```
sns.barplot(x="Number of systems",y="Region",data=top_region,palette="Blues_r_d")
plt.title("Ranking of regions according to the number of metro systems")
```

Out[32]:Text(0.5, 1.0, 'Ranking of regions according to the number of metro systems')



8. Timeline of the Growth of regions

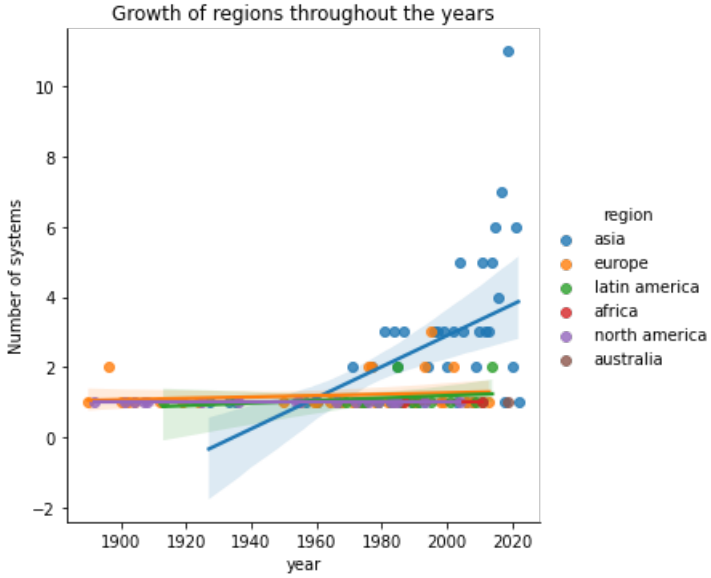
In [39]:
```
reg_gwth=df[["region","year"]]
reg_gwth=reg_gwth.value_counts()
reg_gwth=reg_gwth.to_frame(name="Number of systems")
reg_gwth.reset_index(inplace=True)
reg_gwth.head()
```

Out[39]:

	region	year	Number of systems
0	asia	2019	11
1	asia	2017	7
2	asia	2021	6
3	asia	2015	6
4	asia	2011	5

In [34]:
```
sns.lmplot(x="year",y="Number of systems",data=reg_gwth,hue="region")
plt.title("Growth of regions throughout the years")
```

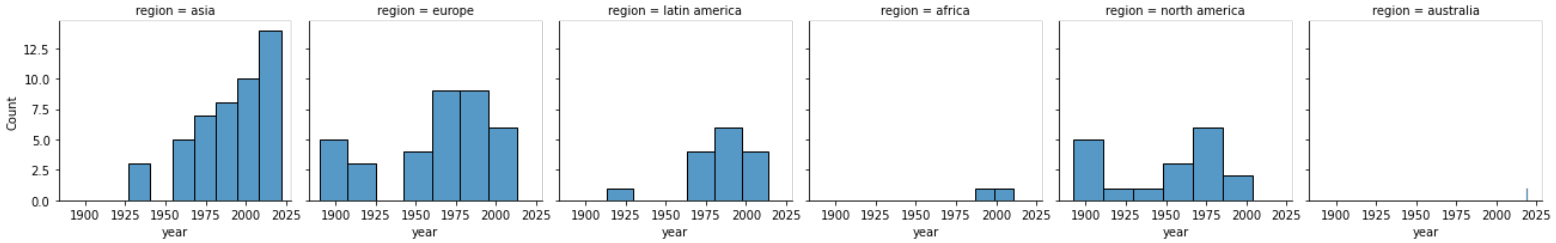
Out[34]:Text(0.5, 1.0, 'Growth of regions throughout the years')



The timeline of development of metro systems in each region

In [45]:
```
y=sns.FacetGrid(data=reg_gwth,col="region")
y.map(sns.histplot,"year")
```

Out[45]:<seaborn.axisgrid.FacetGrid at 0x27b7dcb50f0>



Conclusion

If we closely monitor the analysis we will find that after 2000 most of the developed countries have stopped working further in the development of metro train networks. There are many reasons behind that such as an early start of the developmental phase of networks in the early 1900s, decreasing population in the first world, and enough metro systems looking at their population and country size. Meanwhile, if we look at the developing countries primarily China and India, we will find that these countries started working on their systems after 1970 and that could be one reason behind their constant and massive growth in the 2000s.

China stands out with 43 metro rail systems and still working on it and India also working continuously. Both of these countries are major contributors to the reason why Asia ranks first among all in the ranking of regions with the most metro systems. The rise of China and India after 1980 is the major highlight of the analysis.