Richest People in the world and GDP

Alessandro Belotti
Alessandro Castagna
Ilaria Salvatori

1. Do countries with more billionaires have higher average GDP?

2. Who are the billionaires with the highest correlation between their worth trend and their Region/State GDP trend?

3. What is the percentage of women billionaires in the top 10 richest (with the highest average GDP) countries?

Key concepts

NUTS nomenclature of territorial units statistics

**In the italian meaning
NUTS 1: State
NUTS 2: Regions
NUTS 3: Provinces**

NUTS 2

GDP per capita

Monetary value of all finished goods and services made within a country during a specific period divided by the population.

**Countries
USA States
European regions**

California

Billionaires wealth

Information about billionaires

Name, City, Industries, birth date, sex, ecc.

Data acquisition

Data about:
- Billionaires
- Residence

Information	Acquistion technique	File name
Information about each billionaire	API from Forbes	true.json
Billionaires worth trend	Scraping from Forbes	all_worth_trend_bill.json
Residence State of each american billionaire	Creation of a dictionaire from information in the true.json	American_bill_byProvince.json
Residence Region of each european billionaire	Scraping from GeonNames	European_bill_byProvince.json

true.json

```
[{'rank': 1,
  'listUri': 'rtb',
  'finalWorth': 214258.803,
  'personName': 'Bernard Arnault & family',
  'city': 'Paris',
  'source': 'LVMH',
  'industries': ['Fashion & Retail'],
  'countryOfCitizenship': 'France',
  'imageExists': True,
  'gender': 'M',
  'birthDate': -657244800000,
  'lastName': 'Arnault',
  'wealthList': False,
  'estWorthPrev': 214206.544307,
  'familyList': False,
  'squareImage': '//specials-images.forbesimg.com/
cropX2=4000&cropY1=1209&cropY2=5212',
  'bioSuppress': False},
```

European_bill_byProvince.json

```
Bernard Arnault & family, Île-de-France Paris
Francoise Bettencourt Meyers & family, Île-de-France Paris
Amancio Ortega, Galicia A Coruña
Dieter Schwarz, Baden-Wurttemberg Regierungsbezirk Stuttgart
FranÃ§ois Pinault & family, Île-de-France Paris
Giovanni Ferrero, Brussels Capital Bruxelles-Capitale
```

all_worth_trend_bill.json

```
Bernard Arnault & family {'2018': 72.0, '2019': 76.0, '2020': 76.0, '2021': 150.0}
Elon Musk {'2018': 19.9, '2019': 22.3, '2020': 24.6, '2021': 151.0}
Jeff Bezos {'2018': 112.0, '2019': 131.0, '2020': 113.0, '2021': 177.0}
Larry Ellison {'2018': 58.5, '2019': 62.5, '2020': 59.0, '2021': 93.0}
Bill Gates {'2018': 90.0, '2019': 96.5, '2020': 98.0, '2021': 124.0}
Warren Buffett {'2018': 84.0, '2019': 82.5, '2020': 67.5, '2021': 96.0}
Steve Ballmer {'2018': 38.4, '2019': 41.2, '2020': 52.7, '2021': 68.7}
Carlos Slim Helu & family {'2018': 67.1, '2019': 64.0, '2020': 52.1, '2021': 62.8}
Michael Bloomberg {'2018': 50.0, '2019': 55.5, '2020': 48.0, '2021': 59.0}
```

American_bill_byProvince.json

```
Elon Musk,Texas
Jeff Bezos,Washington
Larry Ellison,Hawaii
Bill Gates,Washington
Warren Buffett,Nebraska
Steve Ballmer,Washington
```

Data about:
- GDP
- additional tables

Information	Acquistion technique	File name
GDP per capita for each European Region	Download	european_GDP_procapite.csv
GDP per capita for each American State	Download	US_states_GDP.csv
GDP per capita for each State in the world	Download	world_GDP.csv
American population during 2022	Download	Population_us.csv
Information about Countries and continents of the world	Download	Countries_and_continents.csv

european_GDP_procapite.csv

	TIME	2014	2015	2016	2017	2018	2019	2020	2021
1	Région de Bruxelles-Capitale/Brussels Hoofdste...	57,800	59,500	59,400	60,600	61,600	63,400	60,900	66,200
2	Prov. Antwerpen	37,600	39,100	39,900	40,900	41,900	43,400	43,000	46,900
3	Prov. Limburg (BE)	26,600	26,900	27,400	28,300	29,100	29,700	29,200	32,300
4	Prov. Oost-Vlaanderen	29,100	30,400	31,000	32,000	32,600	33,500	33,000	36,300
5	Prov. Vlaams-Brabant	34,500	35,800	36,700	37,600	38,700	39,900	36,500	39,800
...	...	...	...	...	...	...	...	...	...

US_states_GDP.csv

	GeoName	2014	2015	2016	2017	2018	2019	2020	2021
0	United States	16932051.0	17390295.0	17680274.0	18076651.0	18609078.0	19036052.0	18509143.0	19609812.0
1	Alabama	189886.3	191335.2	194283.8	196974.9	200372.6	203432.7	199880.8	209979.3
2	Alaska	54188.2	54740.8	54246.6	54278.7	53327.0	53433.8	50705.2	50869.4
3	Arizona	276948.9	282577.0	291275.2	303606.1	314827.5	325395.3	327178.0	347656.0
4	Arkansas	111734.5	112351.0	112798.1	113850.2	115885.2	117126.2	117268.2	123347.3
5	California	2256054.7	2357452.9	2427894.6	2538204.0	2644061.2	2729225.8	2667220.9	2874730.8
6	Colorado	298655.3	312409.7	318953.4	329913.3	342733.2	358438.5	353345.2	373763.3

world_GDP.csv

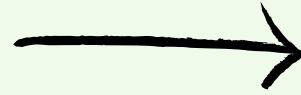|	Country Name	2014	2015	2016	2017	2018	2019	2020	2021
0	Aruba	36846.848290	37343.912960	37583.840320	38865.188200	41679.238190	42501.641550	34971.009910	42698.359870
1	Africa Eastern and Southern	3470.550675	3498.125396	3591.099809	3635.564622	3724.868177	3777.972368	3621.058106	3839.470058
2	Afghanistan	2110.829568	2128.125938	2023.834656	2096.093111	2109.929296	2167.704111	2076.138380	1665.805842
3	Africa Western and Central	4143.177054	4075.994459	3998.951876	4045.303263	4160.540965	4264.731035	4174.504565	4409.450961
4	Angola	8123.048065	7274.090475	7027.146634	7216.061373	7042.923829	6881.076241	6362.636076	6491.125578
...	...	...	...	...	...	...	...	...	...

Data cleaning and pre-enrichment

true.json
- adding continent and residence_state
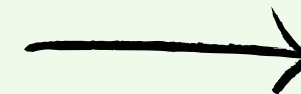- bug fixing of some residence_state like Russia or Slovakia

{'uri': 'bernard-arnault', 'rank': 1, 'listUri': 'rtb', 'finalWorth': 238188
.732, 'personName': 'Bernard Arnault & family', 'city': 'Paris', 'source': '
LVMH', 'industries': ['Fashion & Retail'], 'countryOfCitizenship': 'France',
'imageExists': True, 'gender': 'M', 'birthDate': -657244800000, 'lastName':
'Arnault', 'wealthList': False, 'estWorthPrev': 237579.9297, 'familyList': F
alse, 'squareImage': '//specials-images.forbesimg.com/imageserve/5dc05518ca4
25400079c659f/416x416.jpg?background=000000&cropX1=0&cropX2=4000&cropY1=1209
&cropY2=5212', 'bioSuppress': False, 'continent': 'EU', 'residence_state': '
France'}

world_GDP.csv
- round GDP at second decimal place
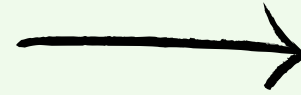- mean replacement for missing values

US_GDP.csv
- round GDP at second decimal place
- GDP per capita computation with american population of 2022

Unnamed: 0	GeoName	2014	2015	2016	2017	2018	2019	2020	2021
0	United States	50995.31	52375.43	53248.78	54442.57	56046.11	57332.06	55745.13	59060.09
1	Alabama	37602.39	37889.31	38473.21	39006.12	39678.95	40284.93	39581.56	41581.33
2	Alaska	73807.58	74560.26	73887.13	73930.85	72634.58	72780.05	69063.53	69287.18
3	Arizona	38121.62	38896.32	40093.62	41790.95	43335.56	44790.20	45035.59	47854.35
4	Arkansas	36898.94	37102.53	37250.18	37597.63	38269.66	38679.49	38726.38	40733.93
5	California	57636.24	60226.69	62026.29	64844.41	67548.78	69724.51	68140.45	73441.78
6	Colorado	51392.19	53759.03	54885.06	56771.03	58977.06	61679.60	60803.16	64316.67

european_GDP_procapite.csv

- round GDP at second decimal place
- change EUR/USD
- fix unity measure from thousand of euros
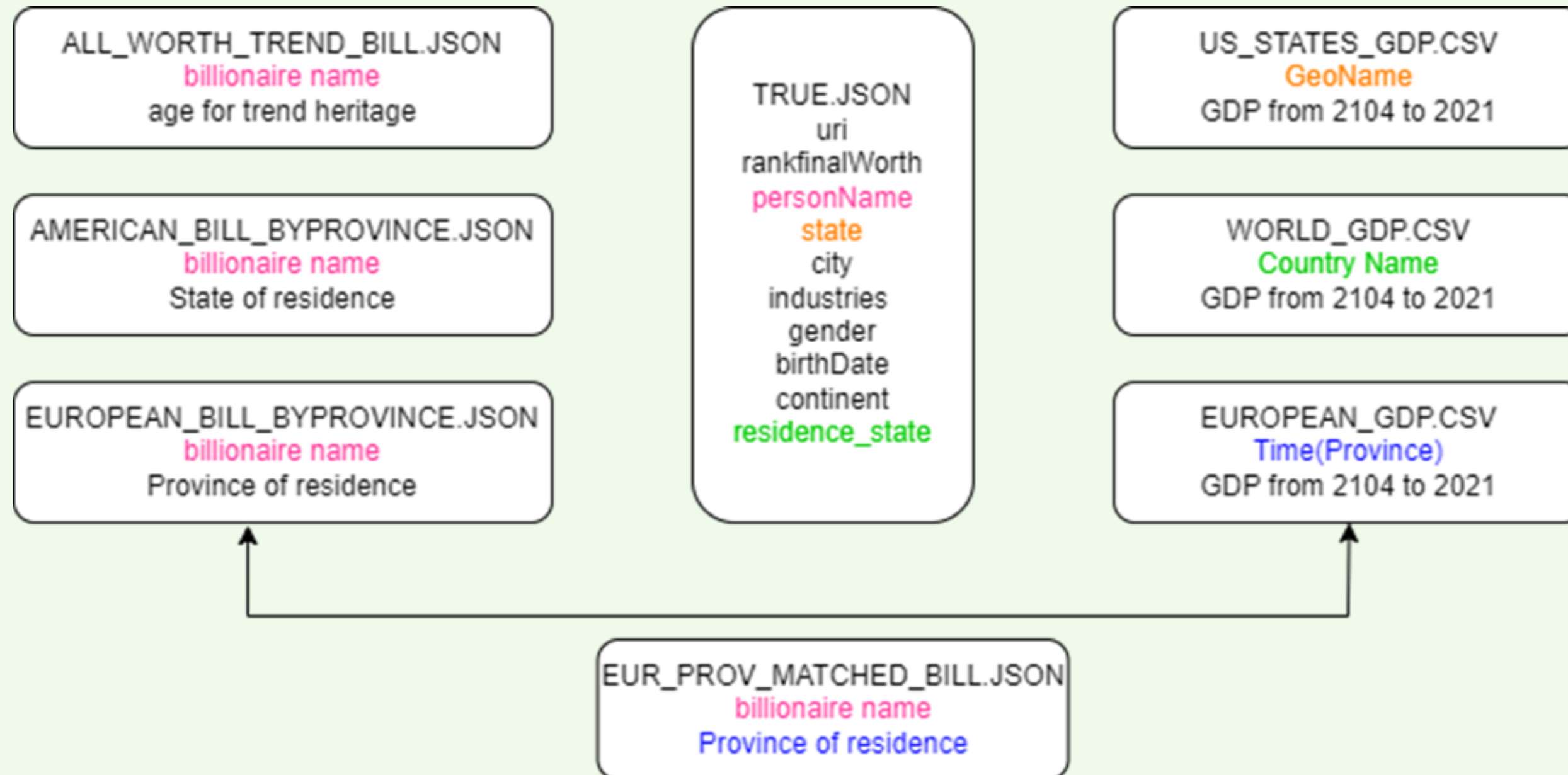- mean replacement for missing values

TIME	2014	2015	2016	2017	2018	2019	2020	2021
Région de Bruxelles-Capitale/Brussels Hoofdste...	62447.12	64283.80	64175.76	65472.24	66552.64	68497.36	65796.36	71522.48
Prov. Antwerpen	40623.04	42243.64	43107.96	44188.36	45268.76	46889.36	46457.20	50670.76
Prov. Limburg (BE)	28738.64	29062.76	29602.96	30575.32	31439.64	32087.88	31547.68	34896.92
Prov. Oost-Vlaanderen	31439.64	32844.16	33492.40	34572.80	35221.04	36193.40	35653.20	39218.52
Prov. Vlaams-Brabant	37273.80	38678.32	39650.68	40623.04	41811.48	43107.96	39434.60	42999.92
....	....	....	....	....	....	....	....	....

european_bill_byProvince.json

- English translation of the European region names
- removing of some special characters ('-', '/')

Schema matching

1.Unidecode function

```
{'nome': 'FranÃ§ois Pinault & family', 'region': 'Île de France'}
```

⬇

{'nome': 'FranASSois Pinault & family', 'region': 'Île de France'}

2.Adding labels to the attributes in the following dataset:
- eur_prov_matched_bill.json
- all_worth_trend_bill.json
- american_bill_byProvince.json

```
{'Elon Musk': 'Texas',
 'Jeff Bezos': 'Washington',
 'Larry Ellison': 'Hawaii',
 'Bill Gates': 'Washington',
 'Warren Buffett': 'Nebraska',
 'Steve Ballmer': 'Washington',
```

➡

```
{"nome": "Elon Musk", "country": "Texas"},
{"nome": "Jeff Bezos", "country": "Washington"},
{"nome": "Larry Ellison", "country": "Hawaii"},
{"nome": "Bill Gates", "country": "Washington"},
{"nome": "Warren Buffett", "country": "Nebraska"},
{"nome": "Steve Ballmer", "country": "Washington"}
```
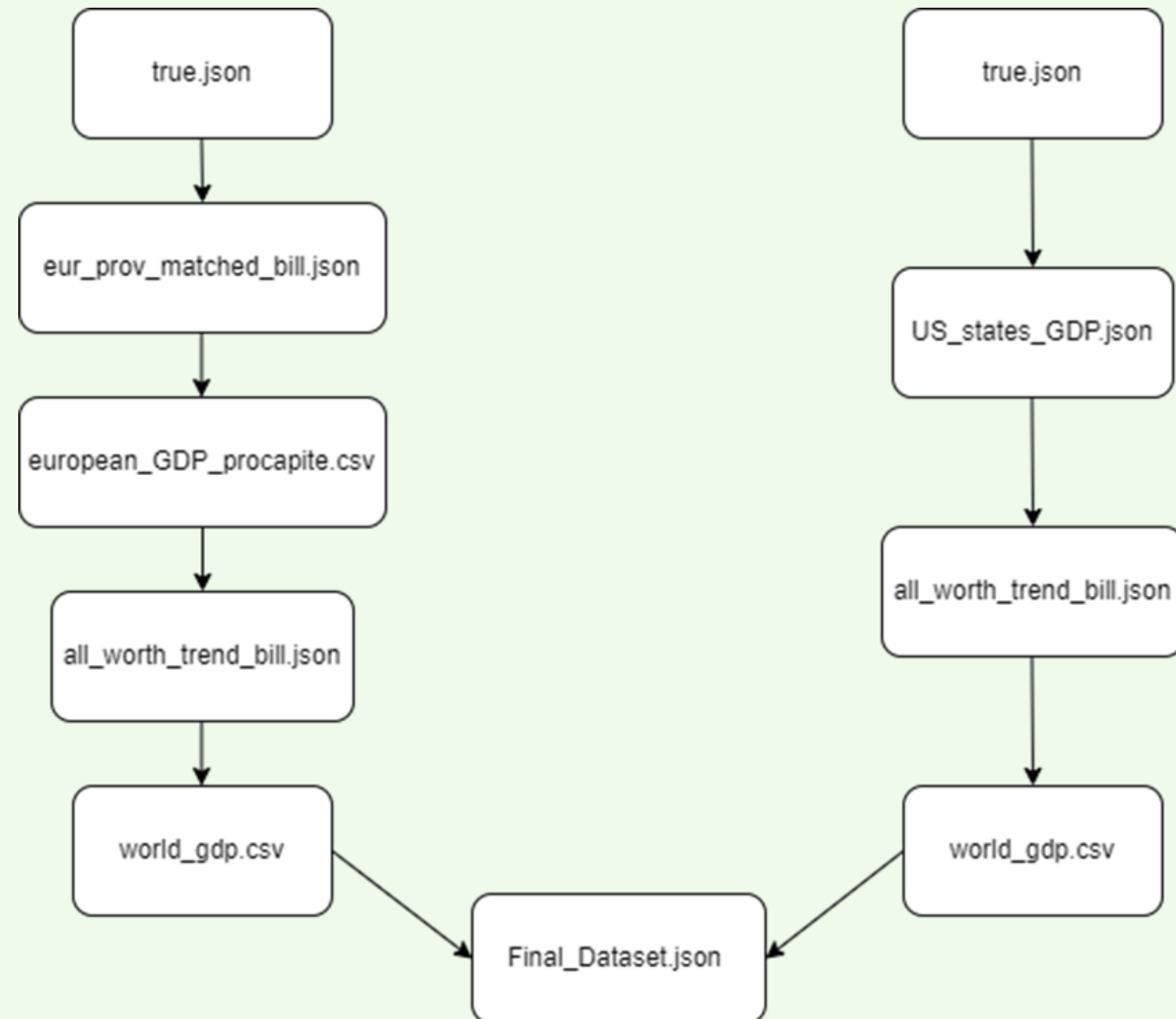
Data storage

Information in the Final_Dataset.json
- characteristics of the billionaire
- Trend of the GDP of the residence State/Region
- Trend of the billionaire's worth
- Trend of the GDP of the residence Country

Why Document Based model?
- Forbes API input
- Null values in billionaires worth trend
- High flexibility level scheme

true.json → eur_prov_matched_bill.json → european_GDP_procapite.csv → all_worth_trend_bill.json → world_gdp.csv → Final_Dataset.json

true.json → US_states_GDP.json → all_worth_trend_bill.json → world_gdp.csv → Final_Dataset.json

Data quality

- Accuracy

Dataset	Length	Accuracy
European_GDP_procapite.csv	307	the currency transformation has been computed with the current EUR/USD change value (no change value for each year)
European_prov_matched_bill.json	493	63 Russian billionaires won't be matched with the European GDP dataset because in this dataset Russia was not take into account (here were considered UE Countries)
US_states_GDP.csv	52	The population dataset used for computing the per capita GDP is referred only to 2022

* si tratta di missing valuessolo riferiti a Regioni Europee di cui non disponiamo del valore del GDP, come Russia, Svizzera, Monaco, ecc.

- Completeness

Dataset	Length	Missing values	% of missing values
world_GDP.csv	264	-	0%
US_states_GDP.csv	52	-	0%
European_GDP_procapite.csv	307	-	0%
true.json	2624	133	5%
European_bill_byProvince.json	493	-	0%
American_bill_byProvince.json	753	-	0%
All_worth_trend_bill.json	1089	59	5.4%
European_prov_matched_bill.json	493	180	36.5%
finalGDP_world_bill.json	1228	155	12,6%*

- Consistency

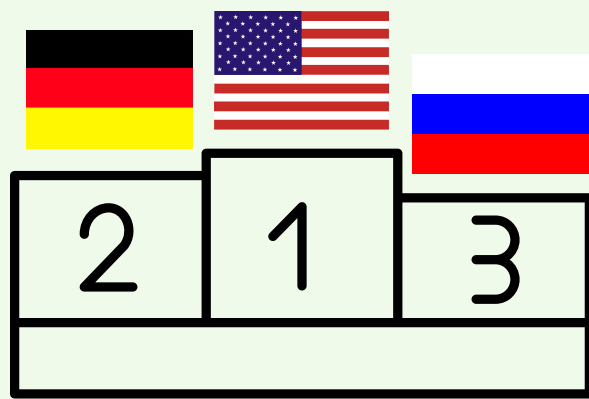| Dataset	Consistency
world_GDP.csv	Rounded values at 2 decimal numbers
US_states_GDP.csv	It wasn't per capita, so we compute this value by dividing it by population
European_GDP_procapite.csv	The values are in euro, we change the currency in USD, and from thousands to unitary values
true.json	For further enrichment we have changed the semantic value of these States of residence: Russia and Slovakia
European_bill_byProvince.json	The European provincies are not written in the same format of the European GDP dataset

- Temporal dimension

Dataset	Temporal quality
European_GDP_procapite.csv	The most up-to-date value of GDP refers to 2021, not 2022
US_states_GDP.csv	The most up-to-date value of GDP refers to 2021, not 2022
All_worth_trend_bill.json	Some billionaires have no worth trend or partial trend in the period 2014/2021
finalGDP_world_bill.json	The most up-to-date value of GDP refers to 2021, not 2022. Some billionaires have no worth trend or partial trend in the period 2014/2021

Conclusion

1) **Do countries with more billionaires have higher average GDP?**

Country Name	Rank for number of billionaires	Number of billionaires	Mean of GDP per capita
Russian Federation	3	93	27711.724
Spain	8	24	38466.799
Ukrain	15	6	12150.101

Country Name	Rank for number of billionaires	Number of billionaires	Mean of GDP per capita
Liechtenstein	26	1	80366.791
Luxembourg	25	1	116373.078
Ireland	17	4	12150.101

BUT

	N_Billionaire	Country_Name	Mean_GDP
0	751	United States	61235.421410
1	97	Germany	53131.493705
2	93	Russian Federation	27711.724122
3	64	Switzerland	70229.963014
4	56	Italy	41607.775795
5	31	France	45564.990925
6	24	Sweden	52924.849920
7	24	Spain	38466.799735
8	14	Monaco	190606.729464
9	10	Netherlands	56097.806949
10	9	Austria	54848.999427
11	8	Norway	66631.125173
12	7	Denmark	55940.659083
13	7	Finland	48207.960091
14	6	Ukraine	12150.101198
15	5	Poland	31339.643860
16	4	Ireland	80366.790925
17	3	Belgium	51506.217821
18	3	Romania	28148.195549
19	2	Cyprus	38364.745608
20	2	Slovak Republic	31125.482252
21	2	Greece	28769.391272
22	1	Albania	13190.426124
23	1	Portugal	33276.407383
24	1	Luxembourg	116373.078013
25	1	Liechtenstein	169148.765476
26	1	Andorra	41088.504733
27	1	Hungary	30906.060742

2) **Who are the billionaires with the highest correlation between their worth trend and their Region/State GDP trend?**

1 = positive correlation ✔

-1 = negative correlation ✘

	Billionaire	Pearson correaltion
0	Daniel Ek	0.997998
1	Bert Beveridge	0.988241
2	Herbert Wertheim	0.982886
3	Jeff Bezos	0.981290
4	Michael Bloomberg	0.981247
5	Arturo Moreno	0.979405
6	Laurie Tisch	0.977823
7	Mark Cuban	0.975748
8	Robert F. Smith	0.975006
9	Reinhold Schmieding	0.973664

	Billionaire	Pearson correaltion
776	Michael Platt	-0.812828
777	Martin Haefner	-0.818783
778	Forrest Preston	-0.819838
779	Ronald Perelman	-0.838403
780	Jaroslav Hascak & family	-0.855138
781	Bernd Freier	-0.863694
782	Ivar Tollefsen	-0.867990
783	Gabriella Meister	-0.881738
784	Heikki Kyostila	-0.887301
785	James France	-0.904857

The most influential billionaire is Daniel Ek, with a correlation value of his worth and the General Domestic Product of his country (Düsseldorf) of 0.997998.
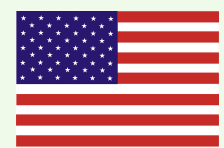This correlation coefficient is based on the following values:
1) General Domestic Product of Düsseldorf: {'2019': 55856.68, '2020': 55208.44, '2021': 59746.12}.
2) Daniel Ek's worth trend: {'2019': 2.2, '2020': 2, '2021': 4.6} (each values is expres in billions of dollars).

3) **What is the percentage of women billionaires in the top 10 richest (with the highest average GDP) countries?**

	State	Average GDP per capita	Women percentage	N women billionaire	N billionaire
0	Monaco	190606.73	21.43	3	14
1	Liechtenstein	169148.77	0.00	0	1
2	Luxembourg	116373.08	100.00	1	1
3	Ireland	80366.79	0.00	0	4
4	Switzerland	70229.96	17.19	11	64
5	Norway	66631.13	25.00	2	8
6	United States	61235.42	13.18	99	751
7	Netherlands	56097.81	0.00	0	10
8	Denmark	55940.66	42.86	3	7
9	Austria	54849.00	11.11	1	9

751 billionaires
only 13% women