13.12.2020

D.Tristram
Ironhack Bootcamp Data-Analytics Final Project

Documentation Final Project socio-económic Impact of Coronavirus

Doc of work done by Excel

1. I checked all corona related indicators on Eurostats:
https://ec.europa.eu/eurostat/web/covid-19/data

Which ones I could use. My first approach was using quarterly measured indicators. Decided which to choose and built up a time panel database. But after discussing the strength a model could have with the few datapoints I had decided to build a new database.

Database 2 uses monthly collected indicators. Decided to use the following indicators in Analysis:

Monthly Indicators	Section	Measurement	Series from - to
1_Inflation Rate	Inflation and prices	Index, 2015=100, monthly rate of change	1_2010 - 09_2020
2_Rent Prices	Inflation and prices	Index, 2015 = 100	1_2010 - 09_2020
3_Balance of payments by country	International trade	Value in million euro	1_2010 - 09_2020
4_Member States EU27 trade by product group	international trade	Balance for values / ration for indexes	1_2010 - 09_2020
5_Sentiment indicators_economy	other economic and business indicators	composite indicator	1_2010 - 09_2020
6_Unemployment rates_total (%)	Labour market	percentage of active population	1_2010 - 09_2020
7_Youth_unemployment rates_under25 (%)	Labour market	percentage of active population	1_2010 - 09_2020
8_Slaughtering in slaughterhouses	Agriculture	thousand tons	1_2010 - 09_2020
Economic sentiment indicator			
Index-indicator			
The economic sentiment indicator (ESI) is a composite indicator produced by the Directorate General for Economic and Financial Affairs (DG ECFIN) of the European Commission. Its objective is to track GDP growth at Member states, EU and euro area levels. The ESI is a weighted average of the balances of replies to selected questions addressed to firms in five sectors covered by the EU Business and Consumer Surveys and to consumers. The sectors covered are industry (weight 40 %), services (30 %), consumers (20 %), retail (5 %) and construction (5 %). Balances are constructed as the difference between the percentages of respondents giving positive and negative replies. EU and euro-area aggregates are calculated on the basis of the national results and seasonally adjusted. The ESI is scaled to a long-term mean of 100 and a standard deviation of 10. Thus, values above 100 indicate above-average economic sentiment and vice versa. Data are seasonally adjusted (SA).			
Further details on the construction of the ESI can be found in the user guide of the Joint Harmonised EU Programme of Business and Consumer Surveys.			
Source Indicators: Eurostat, Corona related indicators			
Queried: 11.12.2020			

2. Downloaded data tables as .tsv file and imported them to Excel.

3. Major prework/data manipulation for the time-panel analysis did in Excel,
The imported tables looked like this when raw:

freq	currency	bop_item	sector10	sectpart	stk_flow	partner	geo\TIME_PI	2010-01	2010-02	2010-03	2010-04	2010-05	2010-06
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	AT	: c	: c	: c	: c	: c	: c
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	BE	-653	-343	759	5733	-110	1406
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	BG	-253,5	87,7	-252,8	-204,9	-56	-69,7
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	CY	:	:	:	:	:	:
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	CZ	625	362,5	49,1	454,2	-1207,2	-252,2
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	DE	7106	10990	17052	11609	5256	13110
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	DK	287,1	814,8	1190,5	1416,8	1037	1482,5
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	EE	9,6	4,1	-50,7	18	-7,6	6,8
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	EL	-2975	-2866	-3634	-2487	348	-1719
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	ES	-4934	-5024	-3135	-3520	-4073	-1759
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	FI	177,6	-86,1	-477,8	-401,7	-754,5	662,9
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	FR	-3659	-1206	369	540	-7362	-1931
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	HR	:	:	:	:	:	:
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	HU	5,9	-26,1	80,1	177,7	-7,5	176,8
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	IE	:	:	:	:	:	:
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	IT	-6740	-5517	-5570	-2375	-5395	-4168
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	LT	-12,3	-19,4	-20	-67,1	100,4	64,3
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	LU	476	491	-51	130	1020	-908
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	LV	91	42	139	83	50	60
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	MK	:	:	:	:	:	:
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	MT	-23,7	:	:	:	:	77,5
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	NL	: c	: c	: c	: c	: c	: c
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	PL	-1389,6	-257,6	-1897,4	-778,5	-700,1	-1673,6
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	PT	-1663	-1789	-1515	-1062	-2153	-2674
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	RO	-95,4	-673,1	-900,3	-635,5	-981,6	-970,4
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	RS	-86	-260	-259	-91	-144	-243
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	SE	1767,1	1292,8	2714,8	1711,8	1879,6	2237,5
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	SI	16,7	-113,8	97,7	21,6	-68,4	-11,8
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	SK	-233,7	-130,8	-170,1	-120,2	-146,4	-198,1
M	MIO_EUR	CA	S1	S1	BAL	WRL_REST	UK	: c	: c	: c	: c	: c	: c

a) checked the data tables and values to get a feeling about the data
b) deleted columns freq, currency, bop_item, sector_10, sectorpart, stk_flow, partner
c) changed column name geo/time period to country.
d) cleaned the tables. deleted all letters and colons out of the numeric fields. Had to do this in different steps as datatype changed when simply replaced.

4. When data was clean to work with, I had a look on the time series, and countries I could use for Analysis. As seen in graphic above in some table some countries had missing data, or no data at all.

Decided to use time series of data from January 2010 – September 2020 on following 13 countries in Analysis:

Country Code	Country
BE	Belgium
BG	Bulgaria
CZ	Ckekzia
DE	Germany
DK	Denmark
EE	Estonia
EL	Greece
ES	Spain
FI	Finland
FR	France
HU	Hungary
IT	Italia
LT	Lithuania

The countries were randomly chosen by availability of data. Decided to go for 13 countries because 13th is my birthday. And I didn´t want to spend too much time on creating a time series panel table.

For each of the 13 countries there exist 8 indicators and 129 data points for each indicator, leads to a total datapoints for modelbuilding in machine learning analysis of:

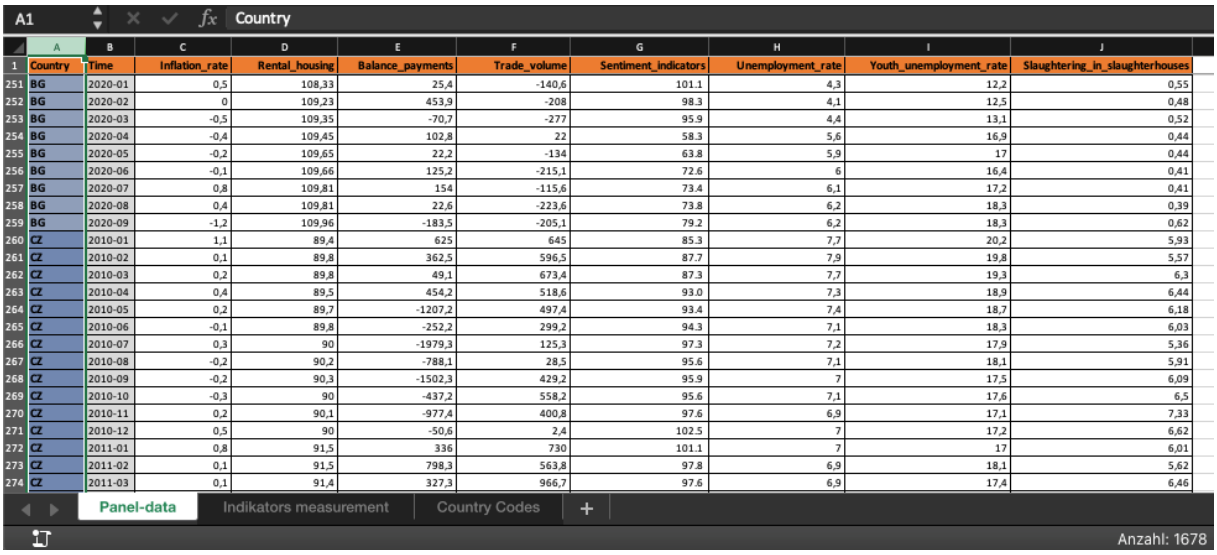
Total: 13.416

5. Next step was to transpose each table to following format:

Country	BE	BG	CZ	DE	DK	EE	EL	ES	FI	FR	HU	IT	LT
2010-01	19,8	0,35	5,93	93,91	10,7	0,63	4,47	47,14	6,2	124,58	2,24	95,76	2,5
2010-02	18,54	0,37	5,57	90,23	9,8	0,6	4,36	44,9	6,37	111,31	2,38	84,85	2,62
2010-03	22,01	0,38	6,3	111,8	11,5	0,78	4,37	50,86	7,53	136,85	2,67	91,7	3,4
2010-04	21,27	0,38	6,44	93,41	10,1	0,73	4,79	47,25	5,02	118,45	2,27	71,48	3,33
2010-05	21,54	0,38	6,18	86,47	10,4	0,74	4,97	50,64	5,47	129,13	2,1	83,13	3,23
2010-06	20,83	0,31	6,03	90,34	11,6	0,66	5,1	52,33	7,01	123,71	2,1	86,97	3,64
2010-07	19,07	0,35	5,36	79,7	7,1	0,67	5,3	51,17	6,83	115,4	2,01	86,08	3,84
2010-08	22,49	0,28	5,91	100,97	12,1	0,8	5,17	53,37	7,27	139,6	2,35	87,74	3,75
2010-09	22,75	0,37	6,09	108,53	12,4	0,93	4,96	51,23	7,54	131,29	2,4	97,29	4,53
2010-10	23,4	0,41	6,5	105,28	11,4	0,86	4,82	50,87	7,41	121,75	2,05	96,13	5,03
2010-11	26,05	0,41	7,33	121,57	13,8	0,84	4,62	49,82	8,02	137,39	2,31	95,55	3,57
2010-12	25,39	0,46	6,62	104,52	10,3	0,72	5,11	57,01	7,48	129,79	2,23	98,74	3,19

6. After finishing transposing the tables I built a time-panel-master.xlsx, where I did import the data in. Next page is seen a section of final version time-panel-master.xlsx to be imported in Python to be analyzed by descriptive, inferential, and advanced analytical models.

Section of time-panel-master.xlsx



	A	B	C	D	E	F	G	H	I	J
1	Country	Time	Inflation_rate	Rental_housing	Balance_payments	Trade_volume	Sentiment_indicators	Unemployment_rate	Youth_unemployment_rate	Slaughtering_in_slaughterhouses
251	BG	2020-01	0,5	108,33	25,4	-140,6	101.1	4,3	12,2	0,55
252	BG	2020-02	0	109,23	453,9	-208	98.3	4,1	12,5	0,48
253	BG	2020-03	-0,5	109,35	-70,7	-277	95.9	4,4	13,1	0,52
254	BG	2020-04	-0,4	109,45	102,8	22	58.3	5,6	16,9	0,44
255	BG	2020-05	-0,2	109,65	22,2	-134	63.8	5,9	17	0,44
256	BG	2020-06	-0,1	109,66	125,2	-215,1	72.6	6	16,4	0,41
257	BG	2020-07	0,8	109,81	154	-115,6	73.4	6,1	17,2	0,41
258	BG	2020-08	0,4	109,81	22,6	-223,6	73.8	6,2	18,3	0,39
259	BG	2020-09	-1,2	109,96	-183,5	-205,1	79.2	6,2	18,3	0,62
260	CZ	2010-01	1,1	89,4	625	645	85.3	7,7	20,2	5,93
261	CZ	2010-02	0,1	89,8	362,5	596,5	87.7	7,9	19,8	5,57
262	CZ	2010-03	0,2	89,8	49,1	673,4	87.3	7,7	19,3	6,3
263	CZ	2010-04	0,4	89,5	454,2	518,6	93.0	7,3	18,9	6,44
264	CZ	2010-05	0,2	89,7	-1207,2	497,4	93.4	7,4	18,7	6,18
265	CZ	2010-06	-0,1	89,8	-252,2	299,2	94.3	7,1	18,3	6,03
266	CZ	2010-07	0,3	90	-1979,3	125,3	97.3	7,2	17,9	5,36
267	CZ	2010-08	-0,2	90,2	-788,1	28,5	95.6	7,1	18,1	5,91
268	CZ	2010-09	-0,2	90,3	-1502,3	429,2	95.9	7	17,5	6,09
269	CZ	2010-10	-0,3	90	-437,2	558,2	95.6	7,1	17,6	6,5
270	CZ	2010-11	0,2	90,1	-977,4	400,8	97.6	6,9	17,1	7,33
271	CZ	2010-12	0,5	90	-50,6	2,4	102.5	7	17,2	6,62
272	CZ	2011-01	0,8	91,5	336	730	101.1	7	17	6,01
273	CZ	2011-02	0,1	91,5	798,3	563,8	97.8	6,9	18,1	5,62
274	CZ	2011-03	0,1	91,4	327,3	966,7	97.6	6,9	17,4	6,46