

Week 11 Final Project Diary

Wong Zi Xin

2023-10-30

Week 9 Diary

1. What is the topic that you have finalised?

The final topic that I have chosen is to create a data story and analyse air transportation data and information in the region of Europe. In the context of my analysis, the definition of Europe refers to the countries in the European Union to date: EU-27 countries. In particular, I will look at EU-27 countries with the most number of passengers and analyse which regions do most people travel to, and observe trends over the years. I chose this topic as I am very interested in travelling, and am curious to derive insights about how the air transport industry was impacted and has picked up again in a post COVID world.

2. What are the data sources that you have curated so far?

I mainly curated my data sources from Kaggle, using datasets containing information about the number of air transport passengers carried by country and datasets containing informaon on the list of airports and airlines globally.

Global datasets:

<https://www.kaggle.com/datasets/tjkynrer/global-air-transport-data>

<https://www.kaggle.com/datasets/thedevastator/global-air-transportation-network-mapping-the-wo>

<https://www.kaggle.com/datasets/johnmwega/trends-and-insights-of-global-tourism>

Datasets specifically looking at Europe:

<https://www.kaggle.com/datasets/gpreda/passengers-air-transport-in-europe>

<https://data.europa.eu/data/datasets/38mt9yvpq2fhg7wgwqf13q?locale=en>

Week 10 Diary

1. What is the question you are going to answer?

How have air transportation trends in the EU-27 changed over time?

2. Why is this an important question?

According to the International Air Transport Association (IATA), air travel is one of the most important modes of transportation as the aviation industry contributes significantly to global GDP by facilitating global trade, business, tourism and more. With the outbreak of the COVID-19 pandemic IATA revealed the aviation industry suffered a loss of \$118 billion in 2020, but with the gradual revival of air travel post-COVID, insights into air travel can better inform strategies to foster economic recovery. Europe was selected as the focus region as according to the United Nations World Tourism Organisation (UNWTO), Europe is the world's top tourist destination.

Sources: <https://www.iata.org/en/iata-repository/publications/economic-reports/aviation-economic-benefits/>

<https://www.iata.org/en/iata-repository/publications/economic-reports/understanding-the-pandemics-impact-on-the-aviation-value-chain/>

<https://www.unwto.org/impact-assessment-of-the-covid-19-outbreak-on-international-tourism>

3. Which rows and columns of the dataset will be used to answer this question?

Columns that are useful to answer this question will be geo (for the country's name), TIME_PERIOD (to represent the corresponding year) and OBS_VALUE (to represent the total number of passengers). All rows are useful as they represent unique data entries of each country by year.

europe air passenger data 2022													
DATAFLOW	LAST UPDATE	freq	unit	tra_mess	tra_cov	schedule	geo	TIME PERIOD	OBS VALUE	OBS FLAG			
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	AT	2011	25137612				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	AT	2012	25965977				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	AT	2013	26749724				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	AT	2014	263776576				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	AT	2015	26754007				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	AT	2016	27181511				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	AT	2017	28327279				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	AT	2018	31138417				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	AT	2019	35644188				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	AT	2020	91168431				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	AT	2021	11105564				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	AT	2022	26381180				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	BA	2021	987059				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	BE	2011	25102695				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	BE	2012	259119515				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	BE	2013	26389927				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	BE	2014	26776258				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	BE	2015	30958841				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	BE	2016	30115832				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	BE	2017	33365493				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	BE	2018	34506309				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	BE	2019	35385188				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	BE	2020	9465828				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	BE	2021	13500020				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	BE	2022	27873992				
ESTAT:TTR00012(1:0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	BG	2011	6652007				

Screenshot of a portion of dataset used

4. Include the challenges and errors that you faced and how you overcame them.

The main dataset that I am using, as seen in the screenshot provided when answering the above question, is not displayed in a very organised and tidy format. The data for each country for each year are all displayed as separate rows. This would make it difficult to create visualisation plots on R, therefore, the first thing I did was to tidy the dataset.

Firstly, I copied over only variables needed (geo, TIME_PERIOD and OBS_VALUE) into a new Excel sheet. Then, I used the pivot table function in Excel to reorganise the data to make it tidy.

Sum of OBS_VALUE Row Labels	Column Labels	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	Grand Total
AT	25137612	25965977	25749724	26378676	26754007	27181511	28327279	31138417	35644188	9168431	11105564	26381180	298932566	
BA	987059												987059	
BE	25102695	259119515	26389927	26776258	30958841	30115832	33260493	34506309	35385188	9465828	13500020	27873992	321254788	
BG	6652007	6819103	7079292	7520697	7610849	7520697	7610849	7520697	7520697	7520697	7520697	7520697	7520697	
CH	41439848	43296396	44271598	46127456	48026375	50054942	56139549	57194326	19009811	19109708	42454182	517992318		
CY	7190387	7282900	72745663	7459663	7459663	7459663	7459663	7459663	7459663	7459663	7459663	7459663	7459663	
CZ	12650532	11742362	11891812	12079873	12672004	13672362	16245554	17832221	16767088	3821372	4755160	11532650	14766890	
DE	175316076	178591103	180783188	186445814	193360430	200687293	212883443	222422361	226764086	57795978	73597370	155302643	2064031685	
DK	26808321	26532730	27459663	29015133	30065505	32763142	33261214	34701139	34780127	8659854	10817817	26648972	32542978	
EE	1907569	2020427	1958665	2019006	2100978	2514889	2636145	2698428	3038003	807637	1260941	2731365	26230153	
EL	33770739	32082338	34023934	36117833	42096402	45543371	50170728	54258826	60885527	17341192	32245559	57893929	494633076	
ES	165153230	159771261	157731973	165354362	174652503	193872037	209824089	220611429	228262372	57797305	9188241	198571203	2054002025	
FR	126013257	129764462	132762875	1363602671	140867569	1450967659	1540987659	161991179	168726788	50724011	66032609	135656038	1549182646	
FI	16374398	16468815	16565391	17171931	17479246	18099954	20054947	22173530	23287929	5428622	4554467	13812577	191461837	
HR	129013267	129764462	132762875	1363602671	140867569	1450967659	1540987659	161991179	168726788	50724011	66032609	135656038	1549182646	
HU	8884837	8429843	8413119	9054848	10228352	11600306	13350029	15176493	16700750	3962697	4062369	12395512	122944005	
IE	23302889	23594089	24603040	26311625	28071422	29266861	32763142	34701139	34780127	8659854	10817817	26648972	32542978	
IS	2460934	2750991	3195056	3653614	4847288	6801814	8318734	10166366	7984197	1267033	2427139	6463479	6463479	
IT	11622667	11602398	11527915	121196597	127695221	13447781	1450987659	1540987659	161991179	168726788	50724011	66032609	135656038	
LT	26911991	3160628	3462358	3798110	4227389	4767561	5246101	6254178	6504685	1804500	2464003	5333890	48761934	
LU	1836780	189091	216937	243366	2695121	2884242	3305329	3808844	4306569	1402310	2029093	407247	33595520	
LV	5098360	4754330	4782527	4802282	5145856	5384160	60077854	7037070	7785276	1995133	2336134	5368369	60567731	
ME	4482843	4149843	8413119	9054848	10228352	11660366	13350029	15176493	16700750	3962697	4062369	12395512	122944005	
MT	3506814	3650347	4032029	4290032	4619557	5080446	6007731	6805817	7318357	1752445	2547912	5861597	56473084	
NL	53895292	55880209	58077271	62693003	64570938	70317395	76240304	79941631	81192507	23594783	29082883	61289771	714548819	
NO	35403532	34692526	36696354	37903195	37903502	37775546	38739778	40330105	40348437	13316983	14136316	32562643	395590306	
PL	257579707	28186254	29641146	32550651	36005814	40930044	47673057	51018598	55007894	16548993	22347692	57081723	444934454	
PT	9687456	9674226	10016933	10907487	12580711	15153719	17934774	19809642	21555895	6633447	10384613	19535951	163884824	
RO	29732247	30356949	31443225	32766043	34011263	35952558	38456213	38945096	37614763	9317677	10798009	25038812	25426755	
SE	1358792	1167877	1265766	1307128	1436003	1404152	1682133	1810567	1719039	287787	149346	968811	14827401	
SI	1808187	1563197	1557149	1594556	1582615	2158261	2402651	2794094	2839787	500064	642078	1942568	21823522	
SK	20881268	1563197	1557149	1594556	1582615	2158261	2402651	2794094	2839787	500064	642078	1942568	21823522	
Grand Total	1940670998	1957036004	1950962173	20002064018	2069685889	2130346645	2253282825	2302182567	237811120	670000652	91608346	190529416	24123845936	

Tidy dataset

After doing so, I highlighted data for countries which did not belong in the EU-27, as I would need to exclude these data for the purpose of my analysis. Furthermore, some of the countries highlighted contain missing values, therefore I removed these countries from the final dataset to be used before I began my analysis

Country	Year	2011	2012	2013	2014	2015
---------	------	------	------	------	------	------