

Week 10 Final Project Diary

Wong Zi Xin

2023-10-23

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 global air transportation data and information. In particular, I will look at countries with the most air transport of passengers and try to analyse which regions do most people travel to, which regions have the most airports 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 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 information on the list of airports and airlines globally.

Global datasets:

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

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

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

Dataset specifically looking at Europe, if I decide to further scope my analysis to a specific region:

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

<https://data.europa.eu/data/datasets/38mt9yvpq2fhg7wwgqf13q?locale=en>

Week 10 Diary

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

How have air transportation trends in the region of Europe evolved over the years, especially in a post COVID world?

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_meas	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	25749724			
ESTAT:TTR00012(1.0)	16/10/23 11:00:00	A	PAS	PAS_CRD	TOTAL	TOT	AT	2014	26378676			
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	9168431			
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	987659			
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	25919515			
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	28776258			
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	33260493			
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	27873892			
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	Column Labels	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	Grand Total
Row Labels														
AT		25137612	25965977	25749724	26378676	26754007	27181511	28327279	31138417	35644188	9168431	11105564	26381180	298932566
BA												987659		987659
BE		25102695	25919515	26389927	28776258	30958841	30115832	33260493	34506309	35385188	9465828	13500020	27873892	321254798
BG		6652007	6819103	7079292	7520697	7610949	9324217	11092651	12137714	11713068	3729017	5047877	8807502	97534094
CH		41439848	43236086	44217568	46127426	48026375	50505492	53564943	56139549	57194328	16006811	19109708	42424182	517992316
CY		7190387	7328300	7011437	7328546	7590787	8961817	10238913	10927101	11261410	2270577	5099704	8613471	93822450
CZ		12650532	11742352	11891812	12079873	12672004	13672362	16245554	17638221	18767088	3821372	4755160	11532650	147668980
DE		175316076	178591103	180783188	186445814	193364340	200687293	212389343	22422361	226764086	57795978	73597370	155302643	2064031685
DK		25808321	26532730	27450623	29015133	30095505	32763142	33261214	34701139	34780127	8658654	10817817	26649573	320542978
EE		1907569	2202427	1958565	2019806	2160978	2214989	2635145	2995528	3258003	857837	1292941	2731365	26235153
EL		33770739	32082336	34023934	39117833	42096402	45543371	50170728	54258826	56088527	17341192	32245559	57893929	494633776
ES		165153230	159771261	157731973	165354382	174652503	193872037	209824089	220611429	228262372	57797305	91898241	199571203	2024500025
EU27_2020		730569231	734860381	746100398	781202599	819698948	871695782	908854476	996295411	1035119832	276758108	373609763	819837926	914889895
FI		16374398	16458815	16565391	17171931	17479246	18099954	20054947	22173530	23287929	5428622	4554497	13812577	191461837
FR		126013257	129764462	132762875	136360671	140867569	145280602	154096485	161991179	168726788	50724011	66033809	136560938	1549182646
HR		4989236	5422632	5722447	6140797	6571698	7475463	8843053	9731294	10623239	1943547	4458400	9415321	81337127
HU		8884837	8429843	8441319	9054848	10228352	11660366	13350029	15176493	16700750	3962687	4665369	12383512	122948405
IE		23362989	23594089	24603640	26310826	29545020	32595709	34271771	36345005	37947510	8268297	9097359	32405890	318348005
IS		2462894	2740691	3199266	3853614	4847288	6801814	8318734	10166386	7584197	1527633	2437139	6463479	60403135
IT		116226667	116029388	115279105	121164587	127665221	134477781	144306325	153352444	160667939	40405355	59709143	132425719	1421709674
LT		2691991	3166628	3482358	3798110	4227389	4787561	5246101	6254178	6504685	1804500	2464603	5333890	49761994
LU		1836780	1893991	2169327	2433966	2651751	2984242	3564730	3988804	4365569	1426310	2002903	4057247	33356620
LV		5098360	4754530	4782257	4802282	5145856	5384160	6077854	7037070	7785726	1995133	2336134	5368369	60567731
ME												1986852		1986852
MK												2628801		2628801
												2152746		2152746
MT		3506814	3650347	4032029	4290032	4619557	5080446	6007731	6805817	7318357	1752445	2547912	5861597	55473084
NL		53895259	55680299	58077271	60963003	64570938	70317995	76240304	79644163	81192507	23594783	29062583	61289771	714548819
NO		32403522	34592225	36686364	37603195	37505062	37727546	38739778	40383105	40348437	13216883	14136316	32562643	395550066
PL		20635672	21800765	23274484	25714422	28907439	32266861	37684668	43767548	46942771	13825460	18893812	39347542	353061444
PT		27579707	28186254	29694146	32560621	36005814	40930044	47673057	51018598	55007894	16548993	22347692	57081723	444635453
RO		9687456	9674226	10016933	10907487	12580711	15153719	17934774	21565865	6633447	10384613	19535951	16384624	16384624
RS														
SE		29732247	30350849	31443225	32766043	34011263	35952558	38456213	38945096	37614763	9317677	10798009	25038812	354426755
SI		1358792	1167877	1265766	1307128	1436003	1404152	1682133	1810567	1719039	287787	419346	968811	14827401
SK		1808187	1563197	1557149	1671290	1943656	2168261	2402651	2794094	2839787	500604	642078	1942598	21825922
UK		201536753	20367015	210468980	220022122	232270437	248688873	264629454	272190155	277432380				2130486169
Grand Total		1940870998	1957039604	1993921773	2090264018	2200784362	2353855648	2538298567	2687118655	2781871120	670004952	916286346	1993529416	24123845459

Tidy dataset

After doing so, I realised there were missing values for certain years for some of the countries which I highlighted in yellow. I tried to resolve this issue by searching for other datasets to fill in the empty data entries, but was unable to do so. EU27_2020 is also highlighted in yellow despite not having missing values, as it represents the European Union consisting of 27 member countries from the year 2020 onwards. This means that the data for EU27_2020 consists of different countries data, hence I excluded it when carrying out my preliminary analyses which is focused on individual countries and trends over individual countries across the years. To proceed with my data analysis, I then decided to exclude the countries with missing data (BA, ME, MK, RS and UK) and EU27_2020.

Country	Year													Grand Total
	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022		
AT	25137612	25969577	25749724	26378676	26754007	27185111	28327279	31138417	35644188	9168341	11105564	26381180	298932566	
BE	25102695	25915915	26399972	28776258	30958841	30115832	33260493	34356308	35385188	9465828	13500020	28773882	321254788	
BG	6813907	6813907	6813907	6813907	6813907	6813907	6813907	6813907	6813907	6813907	6813907	6813907	78603907	
CH	4139848	4323086	44217568	46127426	48026375	50505492	53566493	56139549	57194328	16008611	19109708	42424182	517992316	
CY	7190387	7328300	7011437	7328546	7590787	8961817	10238913	10927101	11264110	2270577	5099704	8613471	93822450	
CZ	12605532	11742352	11898182	12079873	12672004	13672362	16245554	17838221	18676088	3821372	4755610	11526350	147668980	
DE	175316076	178591103	180783188	186445814	193936430	200687293	212389343	222423631	226764086	7759378	73597370	155302640	2060431685	
DK	25808321	26533720	27459632	29015133	30965505	32763142	332621614	34701139	34780127	8658564	10817817	26644937	320542978	
EE	23295229	23295229	23295229	23295229	23295229	23295229	23295229	23295229	23295229	23295229	23295229	23295229	274232952	
ES	33707739	32082336	34029394	39117833	40906450	45543371	50170728	54258826	56808522	17341192	32455559	75893929	944333376	
FI	16532308	159771261	157731973	165354382	174652033	193872037	209824089	220611429	228263272	57797305	91898241	1995712030	2024500025	
FR	15173498	16488831	16565391	17171931	17479246	18099954	20054947	22173350	23289729	5428622	4554497	13812577	191461837	
FR	126013257	129764462	132768275	136360671	140867586	145280602	154096845	161991179	168727688	5072401	66038309	136560398	1549182646	
HR	4899236	5426232	5722447	6147097	6571698	7475463	8843053	9731294	10623329	1943547	4458400	9415321	81327127	
HU	8848837	8429483	8441319	9054848	10228352	11680366	13350529	15176493	16700750	3962687	48565369	123935312	122940485	
IS	23295229	23295229	23295229	23295229	23295229	23295229	23295229	23295229	23295229	23295229	23295229	23295229	274232952	
IT	2428894	2740691	3199266	3853614	4847288	6801814	8183734	10166386	7584197	1527633	2437139	6463479	60403315	
LT	116226667	116023988	115279105	121164587	127665221	134477781	144306325	153352444	160667398	44045355	55709143	132425719	1421709674	
LU	2691991	3166628	3428358	3769111	4227389	4787561	5246101	6254178	5064865	1804500	2464603	5338390	49761994	
LV	1836780	1893991	2169327	2433966	2651175	2984242	3554730	3988804	4365569	1426810	2200923	4057247	33356620	
LU	5989360	7454530	4782527	4802282	5145856	5384160	6077854	7037070	7785726	1995313	2336134	5386839	60567371	
RO	3506814	3650347	4073357	4290032	4619557	5080446	6007731	6806877	7318357	1752445	2547012	5861597	55473084	
NO	53895292	55680209	58077271	60963033	64570933	70371795	76240304	79644163	81192507	2359478	29082583	61296777	714548819	
NL	32403522	34529225	36688364	37603195	37350352	37725578	38739778	40030105	40348437	13216883	14136316	32586243	395550066	
PL	20635672	21800765	23274484	25714827	29400344	32266861	37684668	43767554	46942771	13825460	18893812	39374574	533061444	
PT	27579707	28186254	29694146	32506621	36005814	40930004	47673057	51018598	55007894	16548993	22374692	57081723	444634543	
RO	9687456	9674226	10016933	10907487	12580711	15153719	17943714	19809642	21565865	6633447	10384613	19535951	163884824	
SE	39732247	30350849	31443322	32786043	34011263	35952558	38452733	38945096	37616473	9317677	10798008	25038812	534426755	
SK	11673922	11673922	13027488	13027488	13027488	13027488	13027488	13027488	13027488	13027488	13027488	13027488	158292922	
SK	1808187	1563197	1557149	1671290	1943656	2158261	2402651	2794094	2839787	500604	642078	1942568	18235522	
Grand Total	194087098	195730936	199392173	2090264018	2200784362	2353585648	2538289567	268711855	278171820	67000495	91628636	1993529164	212384554	