



**STEVENS**  
INSTITUTE of TECHNOLOGY  
THE INNOVATION UNIVERSITY®

# Analyze People Migration Data using Python

Mehul Kumar Jain  
[mjain17@stevens.edu](mailto:mjain17@stevens.edu)

SSE



# Contents



- ☐ Project Goals and Conditions
- ☐ CRISP
  - ☐ Business Understanding
  - ☐ Data Understanding
  - ☐ Data Preparation
  - ☐ Data Representation
- ☐ Practical Results – Conclusions

# Project Goals and Conditions



- Human migration is the movement by people from one place to another, particularly different countries, with the intention of settling temporarily or permanently in the new location. It typically involves movements over long distances and from one country or region to another.
- Migrations are the essence of peopling the world and they are believed to have begun approximately 2 million years ago with the early expansions out of Africa by Homo erectus.
- In more modern times, migrations have often been driven by necessity, that could include forced displacement (in various forms such as deportation, slave trade, trafficking in human beings) and flight (war refugees, ethnic cleansing).

# Project Goals and Conditions



- This study is focused on analyzing the data to get insights on the migration dynamics, addressing questions such as:
  - What are the major flows of migration (meaning from where to where)?
  - What are the dynamics of migration by Income, Regional Development and Geographic location?
  - What is distribution in the total number of migrations from the years 1990 to 2020?
- This study will address the above questions for cumulative values for 3 decades (1990-2020).

# Contents



- ☐ Project Goals and Conditions
- ☐ CRISP
  - ☐ Business Understanding
  - ☐ Data Understanding
  - ☐ Data Preparation
  - ☐ Data Representation
- ☐ Practical Results – Conclusions

# Business Understanding



- To analyze the flow of human migration around the world between the years 1990 and 2019.
- To observe any trends or patterns of human migration in the previous years
- Since the dataset has some missing values for individual countries, this study focuses on extracting cumulative information.

# Data Understanding



- The data is published by United Nations, and it covers the following years – 1990, 1995, 2000, 2005, 2010, 2015 and 2019 with some years missing.
- Total number of columns – 241
- Total number of columns - 1983
- The dataset has the Origin of the migration positioned along columns and the Destination of the migration positioned along rows.
- The destination regions are split into various sub-categories based on income, development process stage and geographic regions.

# Data Understanding



- More developed regions comprise Europe, Northern America, Australia, New Zealand, and Japan.
- Less developed regions comprise all regions of Africa, Asia (excluding Japan), Latin America and the Caribbean plus Melanesia, Micronesia and Polynesia.
- Some of the least developed countries are Afghanistan, Angola, Bangladesh, Benin, Bhutan and Cambodia.





# Data Preparation

- **Methodology**

- Reading the excel file "UN\_Migrations.csv" using the Python module Pandas
- Dropping the unnecessary columns like "Sort Order", "Notes", "Code" and "Type of Data"
- Renaming the columns to appropriate names
- Dropping the rows and columns with empty cells and replacing missing values with "0"
- Correcting the data types of all the columns
- Gathering information for cumulative data
- Creating data frames for each decade – 1990-2000, 2000-2010, 2010-2020 and analyzing the data

# Data Representation



- Snapshot of the original data frame –

	Year	Destination	Total	Other South	Other North	Afghanistan	Albania	Algeria	American Samoa	Andorra	...	Unnamed: 279	Unnamed: 280	Unnamed: 281
0	1990.0	WORLD	153,011,473	6,548,526	2,366,800	6,823,350	180,284	921,727	2,041	3,792	...	NaN	NaN	NaN
1	1990.0	UN development groups	..	..	..	..	..	..	..	..	...	NaN	NaN	NaN
2	1990.0	More developed regions	82,767,216	3,385,103	1,077,179	119,386	177,986	867,015	1,027	3,737	...	NaN	NaN	NaN
3	1990.0	Less developed regions	70,244,257	3,163,423	1,289,621	6,703,964	2,298	54,712	1,014	55	...	NaN	NaN	NaN
4	1990.0	Least developed countries	11,060,221	482,753	239,756	0	0	5,622	0	0	...	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2727	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
2728	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
2729	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
2730	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
2731	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN

2732 rows x 285 columns

# Data Representation



- Snapshot of the original data frame after data cleaning –

	Year	Destination	Total	Other South	Other North	Afghanistan	Albania	Algeria	American Samoa	Andorra	...	Uruguay	Uzbekistan	Vanuatu
0	1990	WORLD	153011473	6548526	2366800	6823350	180284	921727	2041	3792	...	237486	1428020	5060
2	1990	More developed regions	82767216	3385103	1077179	119386	177986	867015	1027	3737	...	56838	1078563	1017
3	1990	Less developed regions	70244257	3163423	1289621	6703964	2298	54712	1014	55	...	180648	349457	4043
4	1990	Least developed countries	11060221	482753	239756	0	0	5622	0	0	...	286	2027	9
5	1990	Less developed regions excluding least develop...	59184036	2680670	1049865	6703964	2298	49090	1014	55	...	180362	347430	4034
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1713	2019	Asia	83559197	2486931	1559115	4432460	4913	46313	0	0	...	6263	453346	0
1714	2019	Europe	82304539	311208	868939	488944	1080239	1759610	43	7818	...	97451	1455576	1137
1715	2019	Latin America and the Caribbean	11673288	220861	88624	493	344	1318	0	99	...	192260	56	233
1716	2019	Northern America	58647822	2584771	16148	135162	115508	89772	0	0	...	325459	66188	0
1717	2019	Oceania	8927925	74580	89844	62408	3872	3068	1328	31	...	11246	2147	5976

119 rows x 237 columns

# Contents



- ☐ Project Goals and Conditions
- ☐ CRISP
  - ☐ Business Understanding
  - ☐ Data Understanding
  - ☐ Data Preparation
  - ☐ Data Representation
- ☐ Practical Results – Conclusions



# Major Flows of Migration for each Decade

# For the decade 1990 – 2000:



		Other South	Other North	Afghanistan	Albania	Algeria	American Samoa	Andorra	Angola	Anguilla	Antigua and Barbuda	Argentina	Armenia
Destination	Year												
Less developed regions	2005	3034205	1478637	3593735	4605	72371	1908	57	486722	2013	8977	280205	211473
	2000	3015091	1196350	4383284	4571	63168	2022	50	698597	1931	8783	257964	230740
More developed regions	2005	2697921	729178	304921	981829	1527511	680	5721	229784	1040	28564	528204	686613
High-income countries	2005	3353485	1116199	518295	889591	1564872	139	5734	229419	2118	34741	644944	128211
Middle-income countries	2000	1397355	852717	4197991	76260	28013	3051	40	349555	554	1702	156822	769011
	2005	1692405	861808	3367099	76643	28269	2449	44	192076	365	1609	162967	769876
Asia	2005	1915295	1080968	3592951	3846	38784	0	0	2042	0	0	38513	210015
High-income countries	2000	2808880	1162405	391093	748153	1005272	135	5584	213384	2006	31023	398264	107252
More developed regions	2000	2154380	1013370	223080	819842	974880	1164	5574	213717	1163	25064	297623	645730
Oceania	2005	58905	58663	20671	2270	2792	2028	30	450	0	30	13959	1110
	2000	54117	53514	13294	1530	2556	2142	30	420	0	30	12393	980

The above snapshot highlights the information about the major flows of migration for the decade 1990 – 2000. We can see that the top destination for migration from Afghanistan is Less Developed regions. The detailed information for each country is available in the “.ipynb” file.



# For the decade 2000 – 2010:

		Other South	Other North	Afghanistan	Albania	Algeria	American Samoa	Andorra	Angola	Anguilla	Antigua and Barbuda	Argentina	Armenia
Destination	Year												
Less developed regions	1990	3163423	1289621	6703984	2298	54712	1014	55	857561	1507	7192	206861	245538
More developed regions	1995	2350290	1135796	171042	497670	918261	1115	4140	186922	850	19730	250039	658802
High-income countries	1995	2878313	1112768	329622	437627	943933	103	4150	186499	1676	25293	337980	92075
Middle-income countries	1995	1522623	1060066	3374115	63515	30779	2535	42	204576	539	1582	143510	809576
Less developed regions	1995	3049178	1204305	3545035	3472	61443	1523	52	501738	1716	7952	231957	243070
Middle-income countries	1990	1877678	987475	6544932	48836	30513	1973	44	174224	515	1487	132407	822011
High-income countries	1990	3803597	1223239	269933	131448	886589	68	3748	166885	1364	19773	297255	77381
Oceania	1990	34250	11579	2885	984	1783	1078	333	310	0	21	13574	388
More developed regions	1990	3385103	1077179	119386	177986	867015	1027	3737	167381	540	14561	223308	654111
Oceania	1995	67404	15814	8382	1315	2208	1617	181	370	0	15	13057	710

The above snapshot highlights the information about the major flows of migration for the decade 2000 – 2010. We can see that the top destination for migration from Afghanistan is now changed to More Developed regions. The detailed information for each country is available in the “.ipynb” file.



# For the decade 2010 – 2020:

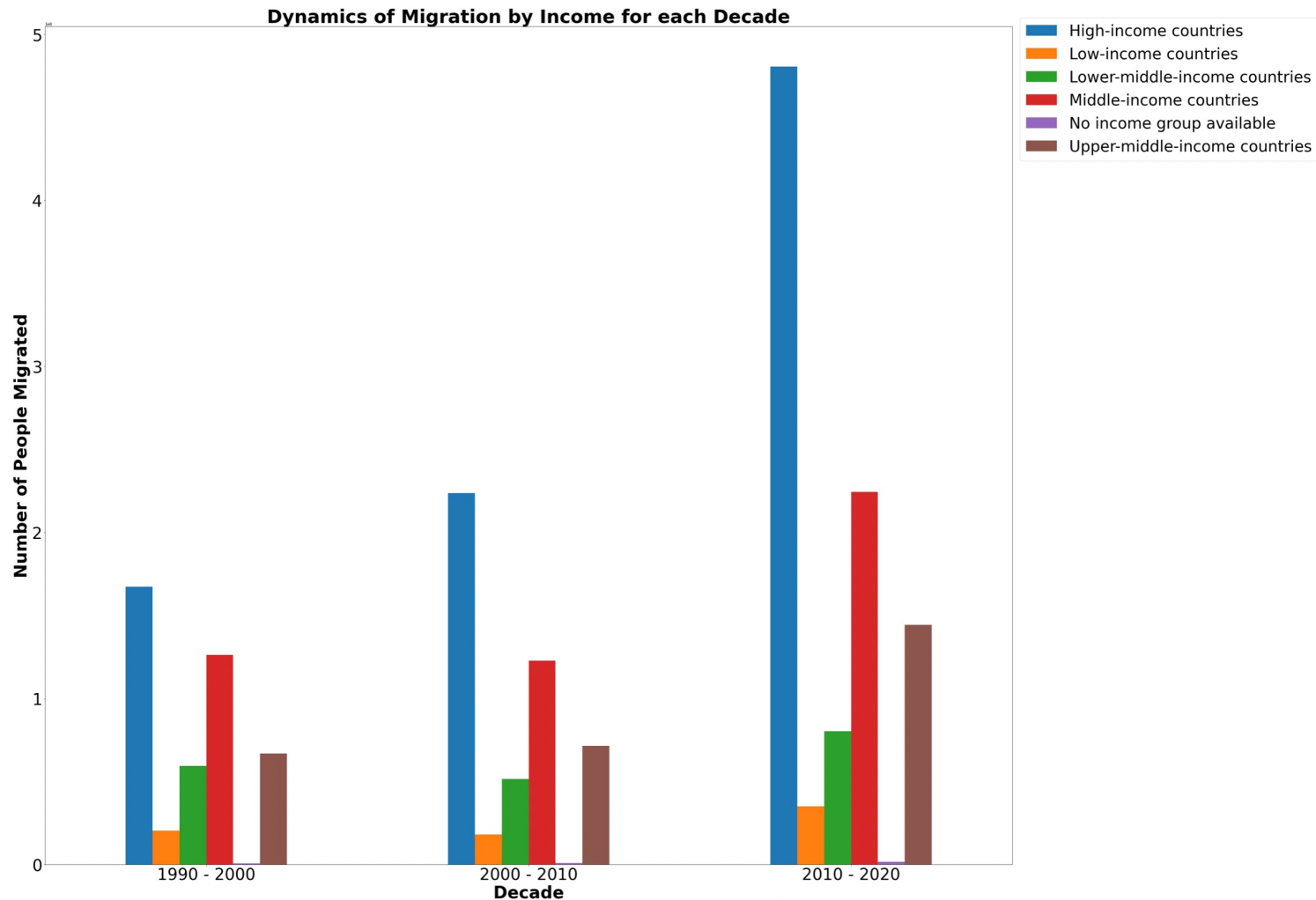
		Other South	Other North	Afghanistan	Albania	Algeria	American Samoa	Andorra	Angola	Anguilla	Antigua and Barbuda	Argentina	Armenia
Destination	Year												
Less developed regions	2019	4835514	2416729	4434242	7413	93657	1774	1265	416982	1646	12737	348657	186661
	2010	3422478	1876202	4693910	5539	78640	1995	914	451002	1563	8115	297452	198413
More developed regions	2019	3114911	991191	686514	1199619	1851127	43	7849	244608	1061	42352	664757	778187
High-income countries	2019	4135205	1552536	1160945	1126602	1895453	1	7886	244314	2491	49035	807631	197331
Middle-income countries	2010	1862617	1029378	4392970	78913	28572	1714	54	152676	178	1549	173163	768285
Less developed regions	2015	4797719	2338048	4503169	5994	92882	1803	1137	408370	1609	12781	318985	191500
Asia	2019	2486931	1559115	4432460	4913	46313	0	0	78	0	0	41974	185296
Oceania	2019	74580	89844	62408	3872	3068	1328	31	592	0	63	19402	1629
Middle-income countries	2019	3182596	1680139	3945911	78991	36676	1370	62	195663	183	1659	205189	767316
High-income countries	2015	3851384	1487567	863575	1051094	1791603	121	7614	235940	2355	46370	754620	175221
More developed regions	2010	2550329	724441	361895	1114766	1577813	164	7012	209054	803	33316	640635	718937
High-income countries	2010	3336493	1332920	647230	1040355	1622567	121	7030	208666	2163	39750	764480	148867

The above snapshot highlights the information about the major flows of migration for the decade 2010 – 2020. We can see that the top destination for migration from Afghanistan is now again changed to Less Developed regions. The detailed information for each country is available in the “.ipynb” file.





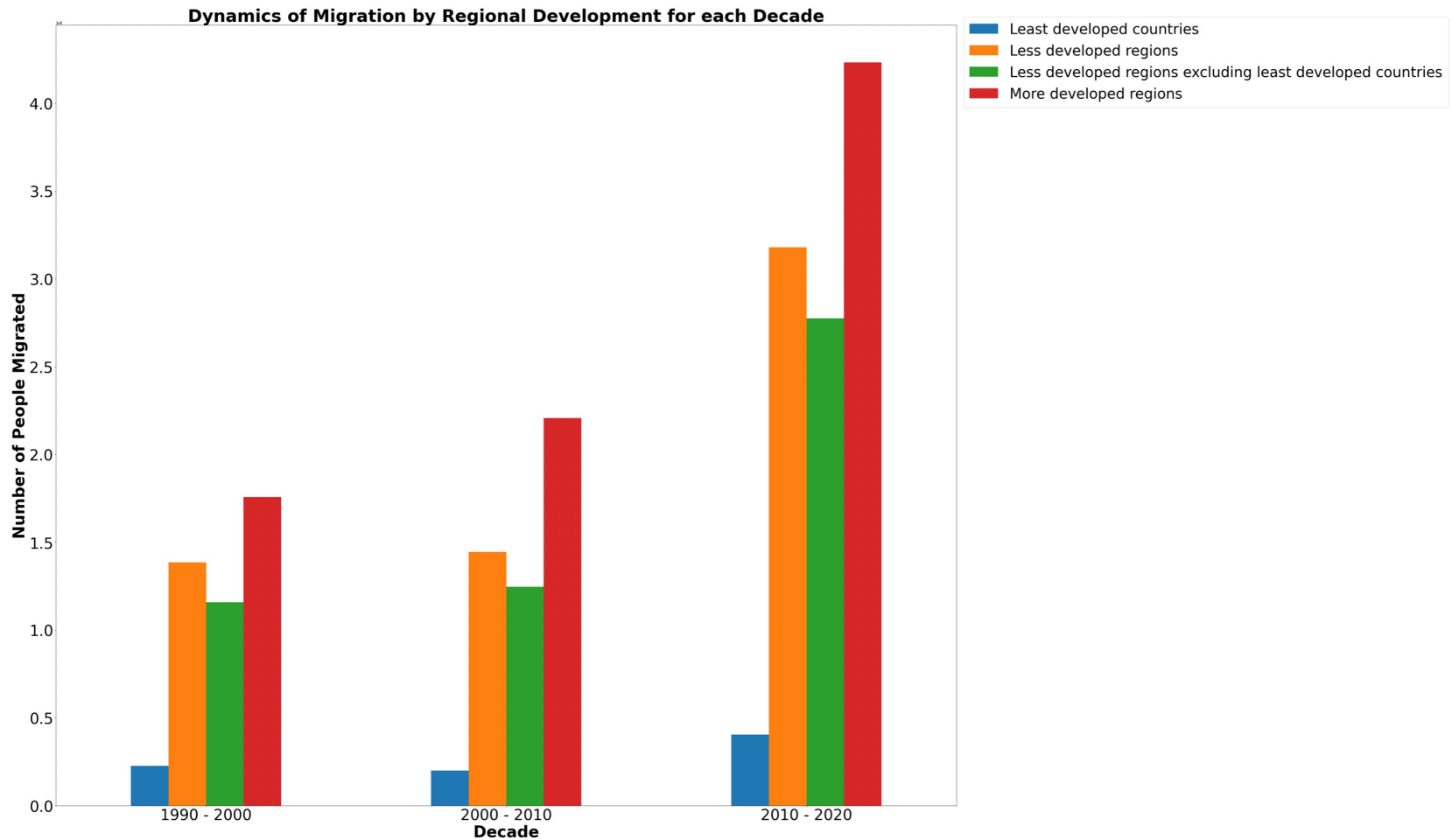
# Dynamics of Migration by Income for each Decade



From the above bar graph, we can conclude that in every decade most of the people are migrating to High-Income countries and the number of people migrating to these countries is drastically increasing each decade.



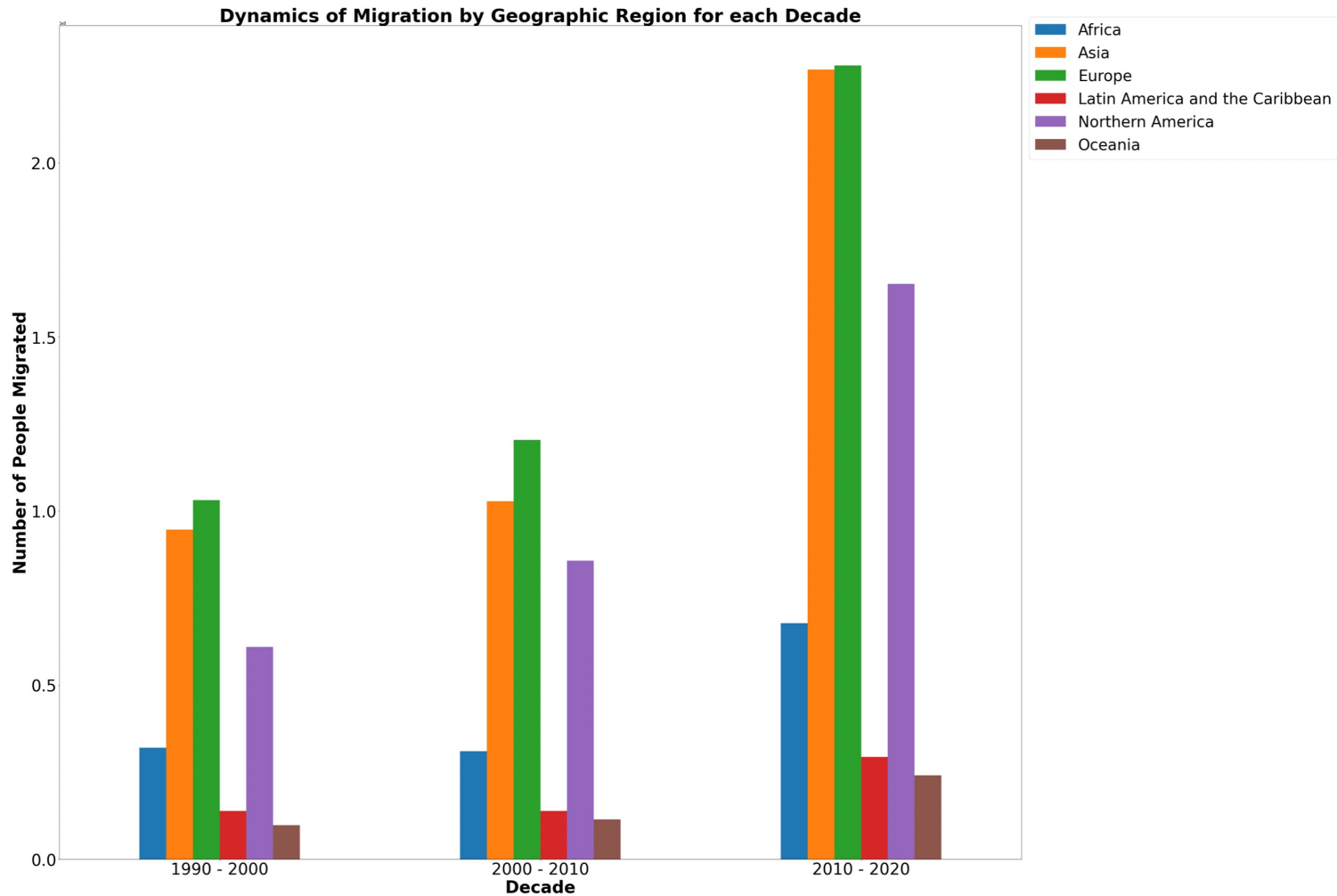
# **Dynamics of Migration by Regional Development for each Decade**



From the above bar graph, we can conclude that in every decade most of the people are migrating to More Developed regions and the number of people migrating to these countries is drastically increasing each decade.



# **Dynamics of Migration by Geographic Region for each Decade**

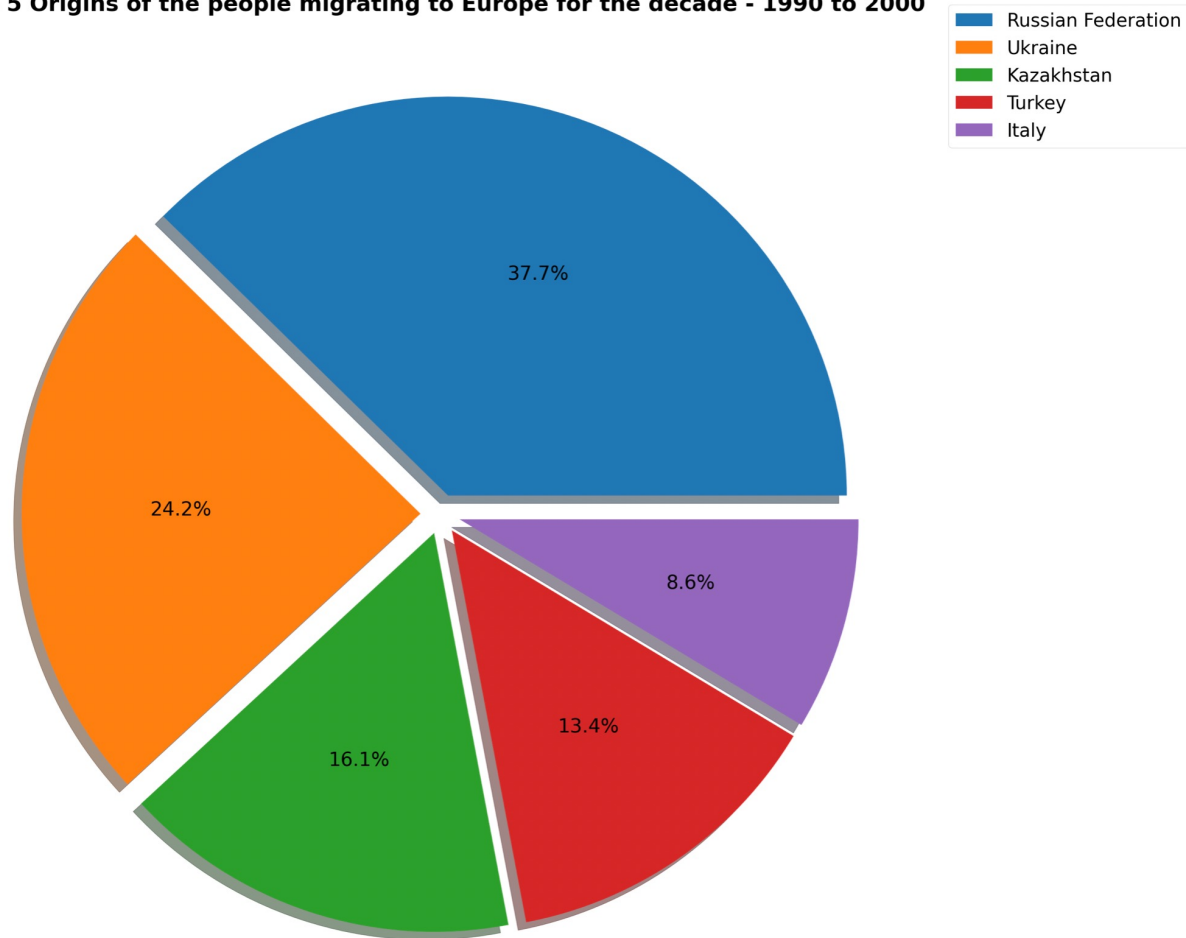


From the above bar graph, we can conclude that the major geographic destinations for migration are Europe and Asia, and with each decade the number of people migrating is increasing.



- Since the major geographic region for migration has been Europe, I wanted to find out the top 5 origins of the people migrating to Europe for each decade.
- The next few slides will give us the detailed information about the people migrating to Europe.

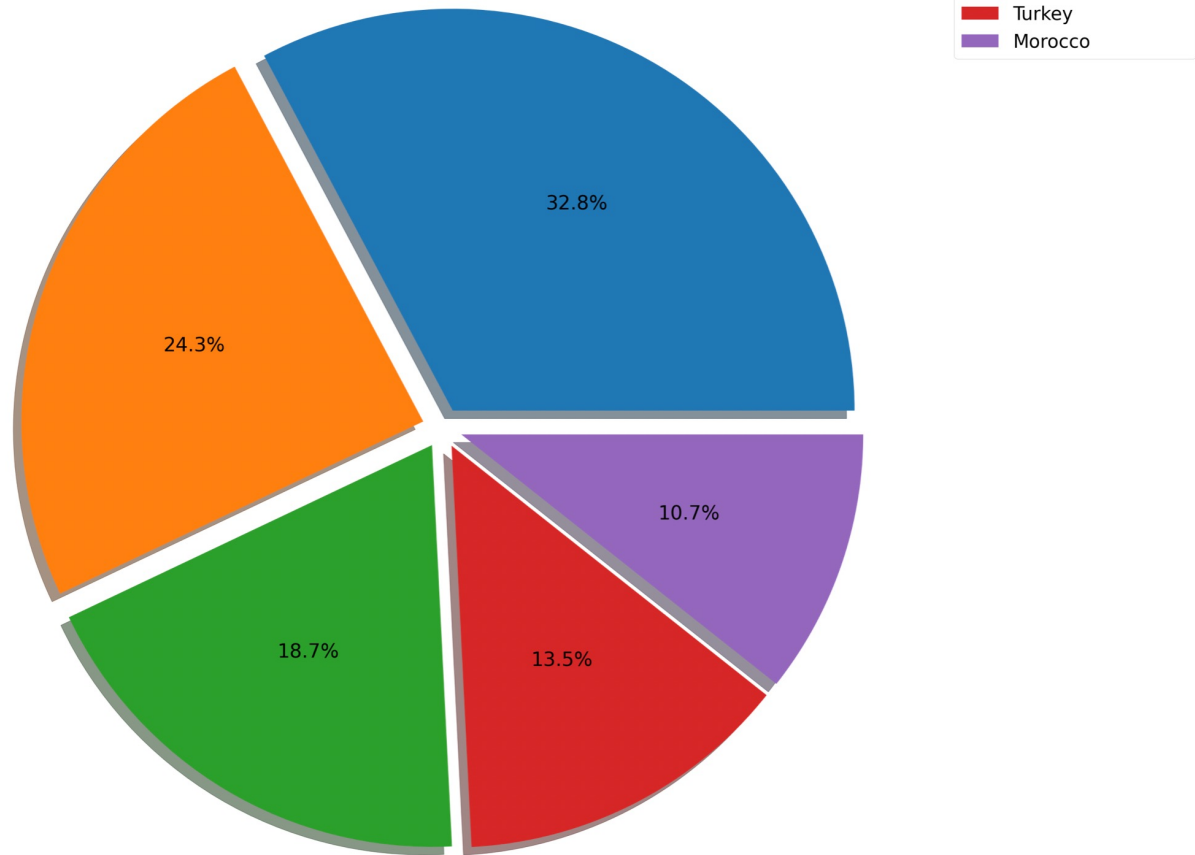
**Top 5 Origins of the people migrating to Europe for the decade - 1990 to 2000**



From the above visualization, we can conclude that majority of the people have migrated from Russian Federation for the decade – 1990 to 2000.

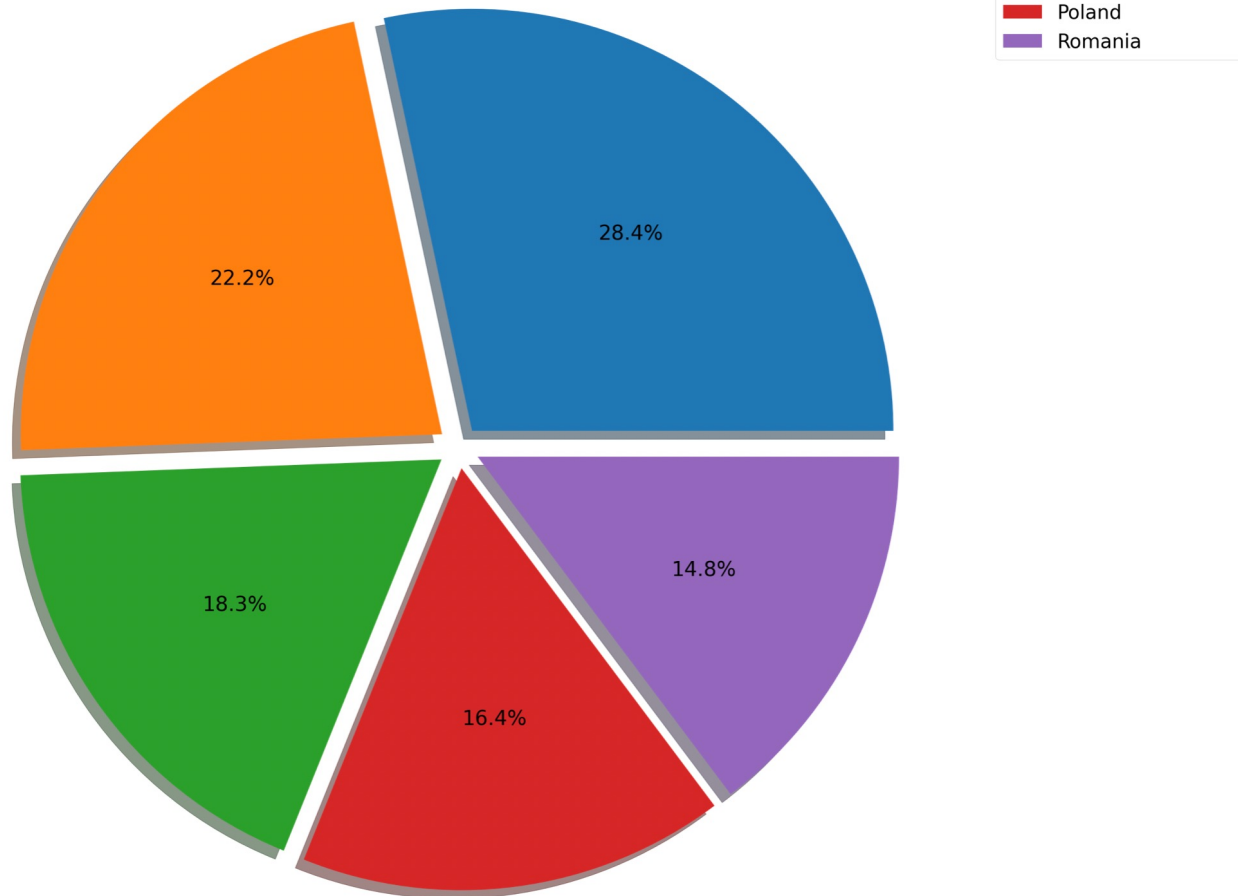


Top 5 Origins of the people migrating to Europe for the decade - 2000 to 2010



From the above visualization, we can conclude that majority of the people have migrated from Russian Federation for the decade – 2000 to 2010. But the percentage of people migrating has been decreased.

### Top 5 Origins of the people migrating to Europe for the decade - 2010 to 2020

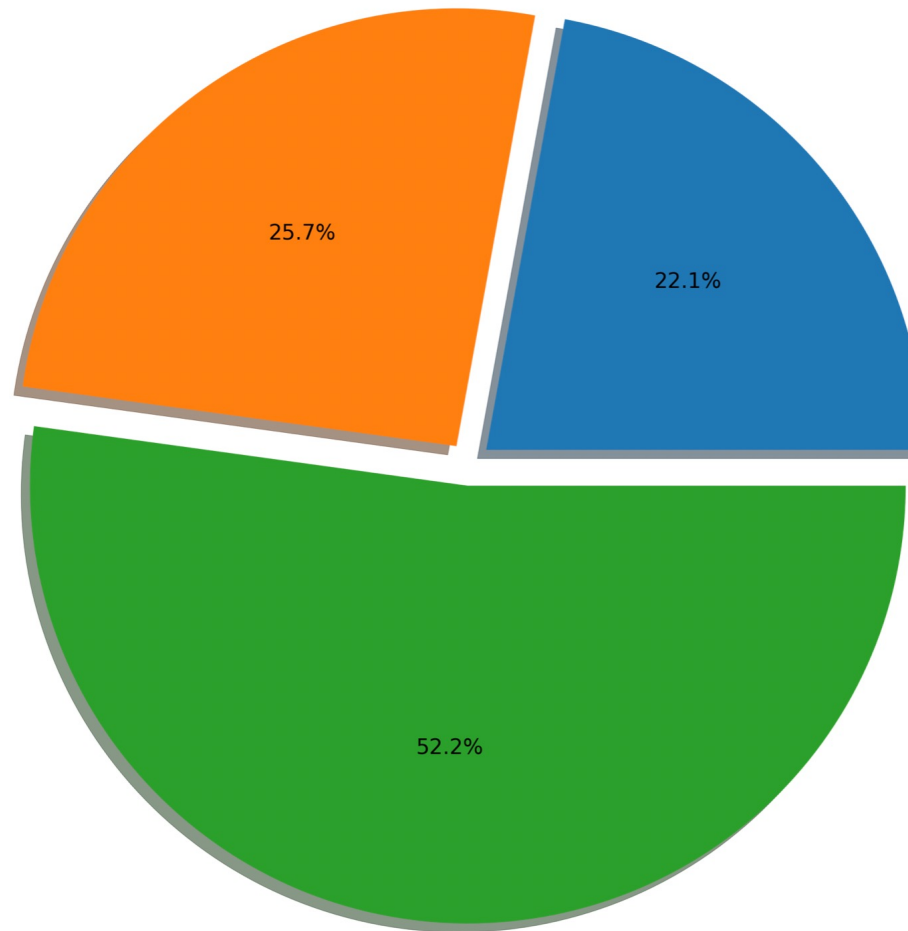


From the above visualization, we can conclude that the top origin for the people migrating to Europe is still Russian Federation for the decade – 2010 to 2020, but the percentage of people migrating has been further dropped. But the number of people migrating from Poland has increased in this decade.



# Total Migrations in the World for each Decade

## Total Migrations in the World for each decade



From the above visualization, we can conclude that the total number of migrations has been almost doubled in the years – 2010 to 2020, when compared to the years – 1990 to 2010.



# Thank You