Q1 Academic Honesty

0 Points

It is a violation of the Academic Integrity Code to look at any reference material other than your textbook and lecture notes, or to give inappropriate help to someone or to receive unauthorized aid by someone in person or electronically via messaging apps such as WhatsApp. Academic Integrity is expected of all students of Hacettepe University at all times, whether in the presence or absence of members of the faculty. Do NOT sign nor take this exam if you do not agree with the honor code.

Understanding this, I declare I shall not give, use or receive unauthorized aid in this examination.

Signature (Specify your name and surname as your signature)

Mehmet Taha USTA MTUSTA

Q2 Data Science

100 Points

You are given a dataset that you are familiar from the lecture already. The dataset contains immigration data on Canada, and is available as an Excel file (Figure 1).

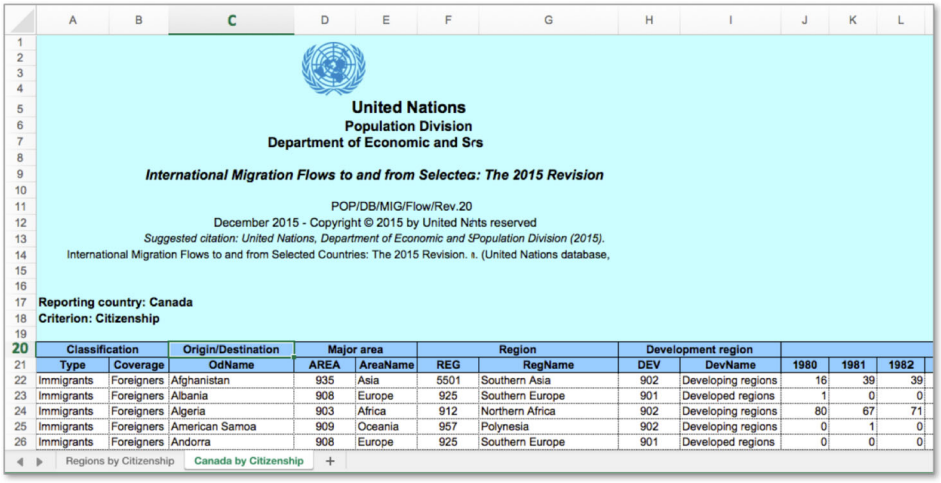
Assume that you skipped twenty rows and retrieved remaining part of the "Canada by Citizenship" sheet into a Pandas dataframe named **df_original** (Figure 2). Each row represents a country and contains metadata about the country such as where it is located geographically, and whether it is developing or developed. Each row also contains numerical figures of annual immigration from that country to Canada from 1980 to 2013.

Further assume that you need to process and convert your **df_original** dataframe into the **df_canada** dataframe seen in Figure 3 to perform your analysis task as a datascientist. In the processed dataset "df_canada", the country name becomes the index of each row. This should make querying specific countries easier. Also, an extra column (**Total**) is added to represent the cumulative sum of

annual immigration from each country from 1980 to 2013. So for Afghanistan, it is 58,639, total, and for Albania it is 15,699, and so on.

Note: You can refer to lecture notes for the immigration data on Canada. It is available on the course's web site as well.

Figure 1. Immigration Data on Canada by UN as an Excel file



United Nations
Population Division
Department of Economic and Srs

International Migration Flows to and from Selectea: The 2015 Revision

POP/DB/MIG/Flow/Rev.20
December 2015 - Copyright © 2015 by United Nıtıs reserved
Suggested citation: United Nations, Department of Economic and Population Division (2015).
International Migration Flows to and from Selected Countries: The 2015 Revision. s. (United Nations database,

Reporting country: Canada
Criterion: Citizenship

Classification		Origin/Destination	Major area		Region		Development region				
Type	Coverage	OdName	AREA	AreaName	REG	RegName	DEV	DevName	1980	1981	1982
Immigrants	Foreigners	Afghanistan	935	Asia	5501	Southern Asia	902	Developing regions	16	39	39
Immigrants	Foreigners	Albania	908	Europe	925	Southern Europe	901	Developed regions	1	0	0
Immigrants	Foreigners	Algeria	903	Africa	912	Northern Africa	902	Developing regions	80	67	71
Immigrants	Foreigners	American Samoa	909	Oceania	957	Polynesia	902	Developing regions	0	1	0
Immigrants	Foreigners	Andorra	908	Europe	925	Southern Europe	901	Developed regions	0	0	0

Regions by Citizenship | Canada by Citizenship | +

Figure 2. First five rows of the df_original dataframe (i.e., output of df_original.head())

	Type	Coverage	OdName	AREA	AreaName	REG	RegName	DEV	DevName	1980	...	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
0	Immigrants	Foreigners	Afghanistan	935	Asia	5501	Southern Asia	902	Developing regions	16	...	2978	3436	3009	2652	2111	1746	1758	2203	2635	2004
1	Immigrants	Foreigners	Albania	908	Europe	925	Southern Europe	901	Developed regions	1	...	1450	1223	856	702	560	716	561	539	620	603
2	Immigrants	Foreigners	Algeria	903	Africa	912	Northern Africa	902	Developing regions	80	...	3616	3626	4807	3623	4005	5393	4752	4325	3774	4331
3	Immigrants	Foreigners	American Samoa	909	Oceania	957	Polynesia	902	Developing regions	0	...	0	0	1	0	0	0	0	0	0	0
4	Immigrants	Foreigners	Andorra	908	Europe	925	Southern Europe	901	Developed regions	0	...	0	0	1	1	0	0	0	0	1	1

Figure 3. Pre-processed dataset (i.e., output of df_canada.head())

	Continent	Region	DevName	1980	1981	1982	1983	1984	1985	1986	...	2005	2006	2007	2008	2009	2010	2011	2012	2013	Total
Country																					
Afghanistan	Asia	Southern Asia	Developing regions	16	39	39	47	71	340	496	...	3436	3009	2652	2111	1746	1758	2203	2635	2004	58639
Albania	Europe	Southern Europe	Developed regions	1	0	0	0	0	0	1	...	1223	856	702	560	716	561	539	620	603	15699
Algeria	Africa	Northern Africa	Developing regions	80	67	71	69	63	44	69	...	3626	4807	3623	4005	5393	4752	4325	3774	4331	69439
American Samoa	Oceania	Polynesia	Developing regions	0	1	0	0	0	0	0	...	0	1	0	0	0	0	0	0	0	6
Andorra	Europe	Southern Europe	Developed regions	0	0	0	0	0	0	2	...	0	1	1	0	0	0	0	1	1	15

Please provide Python codes to obtain "df_canada" out of "df_original".

```
...
```

Short Exam 5

GRADED

STUDENT
MEHMET TAHA USTA

TOTAL POINTS
0 / 100 pts

QUESTION 1
Academic Honesty — **0** / 0 pts

QUESTION 2
Data Science — **0** / 100 pts