Developing Integration Policy for Refugees through Mobile Phone Data Analysis: A Study on Türk Telekom Customers

Ibrahim Zincir[1], Tohid Ahmed Rana[2], Ayselin Yıldız*[3] and Dilaver Arıkan Açar[4]

[1] Yaşar University, Department of Computer Engineering, Izmir, Turkey
[2] Yaşar University, Department of Computer Programming, Izmir, Turkey
[3] Yaşar University, Department of International Relations, Izmir, Turkey
[4] Yaşar University, Department of International Relations, Izmir, Turkey
*ayselin.yildiz@yasar.edu.tr

Abstract. Managing migration and developing effective integration policies require reliable, updated information and comparable statistical data. However, traditional data sources such as public censuses, poor official statistics or geographically limited field studies do not provide safe and updated migration data on a highly dynamic phenomenon. Mobile phone data as a new and alternative source offers higher quality, comparable, up-to-date and better data that can ensure increased support for statistical systems of policy makers. This paper aims to bring relevant insight for policy makers in Turkey and possibly abroad for better integration of refugees by analyzing the database based on anonymized mobile Call Detail Record of phone calls and SMS messages of Türk Telekom customers. The results show that mobile phone data is able to identify approximate location of the refugees, their mobility trends and activities over time along with their general communication patterns. The algorithms used in this study are able to differentiate each individual caller and their location information based on the communication patterns derived from the data. The results provide policy makers a handy alternative data source to get information and verify their statistics related to refugees and develop better integration policies by optimizing limited welfare resources effectively.

Keywords: Data Analysis, Social Integration, Refugees, Mobile Phone Data.

1 Introduction

Turkey, located on the Eastern Mediterranean migration route[1], has become a major "transit" and "destination" country for many migrants, asylum-seekers and refugees in the recent years (İçduygu 2013, Kaiser and Kaya 2015). As İçduygu and Sert (2009) notes, Turkey's location between the politically and economically unstable East and

[1] Eastern Mediterranean migration route begins in Asia, Central Asia, Middle East or the Horn of Africa and ends in Cyprus, Greece or Bulgaria via Turkey both by sea and by land.

the prosperous West makes it both an obvious transit route and an attractive destination in itself for migrants (İçduygu and Sert 2009). Although this is not a new fact, following the migration crisis of 2015, when over a million irregular migrants and refugees have reached Europe mostly from Syria, Africa and South Asia (IOM, 2015), Turkey has started to attract more attention from the international society in terms of its migration policies. European Border and Coast Guard Agency, also known as FRONTEX, reported 885,386 migrants of the nearly one million migrants had reached the European Union (EU) via the Eastern Mediterranean route in 2015 (FRONTEX, 2016). However, beyond 2015 migration crisis, Turkey has always been considered as a country where migration "is" and "will" be a significant issue of concern. As İçduygu (2005) notes, as long as the ongoing political turmoil and violence persist in conflicted neighbouring areas, people will continue to leave their homelands to search for prosperity, security and protection from persecution. Thus, Turkey is always exposed to new waves of especially irregular migration and asylum seekers. These movements have implications and challenges not only for social, economic and political dynamics in Turkey but also for the neighboring regions.

Currently, Turkey hosts the world's largest number of refugees; more than 3.5 million Syrians and additionally 346,800 refugees and asylum-seekers of various nationalities (DGMM, 2018a, UNHCR 2018). Beyond being a transit country, as a destination country where millions of refugees are residing, Turkey develops its own integration policy, which is currently lacking, to better serve not only the migrants and refugees but also to keep the solidarity and social acceptance among the host society. In this context, policy makers and practitioners require to benefit from relevant and actual data in order to develop well targeted and efficient policies on; a) managing migration and refugee movements through/to Turkey, b) develop an inclusive and comprehensive integration policy for the settled migrants.

Accordingly, new data sources such as mobile phone calls, internet searches, or interaction on social media provide the policy makers with more timely and evidence based data on migration. As it was also widely agreed at UN High-level Dialogue on Migration and Development (2013) and also at the Report of the High-Level Panel of Eminent Persons on the Post-2015 Development Agenda (2013), policy makers need better data on migration and development. Lack of data often leads to inefficient public policies on migration management and integration due to miscalculations about the scale and impact of migration. The problem is not only the lack of robust migration data but also how to confirm and update the data that comes from poor official statistics. Especially in the migration field, dynamic people movements have to be quickly reflected in the statistics in order to develop effective policies.

This report aims to provide an insight for policy makers to demonstrate how mobile phone data can be utilized to facilitate the development of better migration policies in Turkey. It focuses on the integration aspect of Syrian refugees in Turkey and specifically prioritizes to contribute the themes on "integration" and to some extent on "unemployment". The report uses the large dataset which is composed of anonymous

mobile phone data and made available for the Data for Refugees (D4R) Challenge (Salah et al.). D4R Challenge is initiated by Türk Telekom in partnership with the Turkish Academic and Research Council (TÜBİTAK) and Boğaziçi University.

The first section of the report provides brief information on the current nature of refugees in Turkey and the main challenges of integration. The second part explains the main problem and specific aim of this research. The third part explains the methodology and data analysis. The final part evaluates the research findings and concludes with some suggestions for policy makers. It also provides insights for further academic research that can be improved with additional data utilization.

2 Refugees in Turkey and Challenges of Integration

In the last decades, mobility of people through and towards Turkey have become more diversified in terms of the changing dynamics in types, flows and destinations. In this context, Turkey emerges as a major significance and also a concern in terms of its migration policies that are highly influenced by the high levels of irregular migration and asylum flows through its territory. Currently, Syrians constitute the highest number of refugees in Turkey and their socio-economic integration refers to one of the primary challenging cases of migration management in Turkey. Every progressive step and ad-hoc practices on Syrians' integration contributes to Turkey's wider integration policy, which is currently evolving.

After the civil war in Syria that has erupted in 2011, Turkey followed an open-border policy and provided protection to all destitute Syrians[2]. As of 2018, among 3.5 million registered Syrians, only 6 percent of whom are sheltered in 25 camps in 10 provinces of Turkey while 3.3 million Syrians are living dispersed around various locations throughout the country (DGMM 2018b). In addition to Syrians, 346,800 refugees and asylum-seekers from other countries are provided international protection in Turkey (UNHCR, 2018). These include people from various countries such as Iraq, Afghanistan, Iran, Somalia, Pakistan, Yemen, Eritrea and Palestine. Syrians are registered under "temporary protection status" and granted free access to education, health services, social aid and right to work. These rights and benefits have been extended to everyone from all nationalities granted international protection in Turkey. In this context, Turkey has been generous in providing humanitarian protection and aid

[2] Due to keeping its geographical limitation to 1951 Geneva Convention, Turkey does not grant refugee status for the asylum applicants from non-European countries of origin. According to Regulation No. 29153 on Temporary Protection (2014 Regulation), which was enacted in line with Article 91 of the Law Number 6458 on Foreigners and International Protection, Syrians in Turkey are offered a group-based "temporary protection". The regulation applies to all Syrian nationals, refugees as well as stateless persons from Syria seeking international protection, including those without identification documents. It ensures *non-refoulment,* and grants right to legal stay as well as free basic health care, education and social assistance for registered Syrians.

for not only Syrians but all refugees, even when the numbers have become overwhelming.

One of the biggest challenges of migration management in Turkey is the lack of overarching official "integration" policy. Turkey currently develops its integration policy based on its ad-hoc humanitarian practices in particular that are developed for Syrians. The temporary stay of Syrians in Turkey is prolonged and a high number of them started to permanently settle in Turkey. Most studies report that the majority of Syrian refugees are inclined to stay in Turkey for the long term, since even if the conflict in Syria ends, peace and stability will not be restored immediately (Yıldız and Uzgören 2016, Kirişci 2014, Erdoğan 2015). Actually, even before the Syrian case, Turkey has been always exposed to new waves of refugees, which would certainly bear additional socio-economic challenges. Thus, a sustainable integration policy including in all the best practices being implemented in the fields of education, health, employment and social integration is an urgent necessity for Turkey. In doing so, the existing legislative loopholes and administrative shortcomings could be well determined and resolved. Despite Turkey's long-lasting experience on migration movements, the institutionalization of the policy area is highly dependent on the provision and regular tracking of "fast changing" and "hardly obtained" accurate statistics due to the very mobile nature of the issue.

3 Main Problem and Aim of the Research

This paper aims to demonstrate how data mining can be utilized to develop solutions and offer benefits for social integration of refugees in Turkey. Accordingly, by using different datasets, it attempts to provide an analysis based on mobile phone usage of randomly selected in total 20 anonymous Türk Telekom customers who are composed of 10 refugees and 10 non-refugees (Salah et al.).

Within the limits of the available data, firstly, the analysis focuses on mainly two features; namely "location" and "timing" of the calls. Upon these two features, by employing basic statistics to the datasets modified by the project team, it tries to map out the accurate residence, mobility story and the possible employment status of the refugees. This new information provides a unique comparable input for the statistical records of policy makers of mainly two institutions; Turkish Ministry of Interior Directorate General for Migration Management (DGMM) and Turkish Ministry of Labour, Social Services and Family.

Secondly, through using five algorithms (Decision Tree, Ripper, K-Nearest Neighbours, Hoeffding Tree and Naive Bayes) and implementing them on the sample group, the analysis figures out both refugee and non-refugee customers' behaviour of mobile phone usage. Depending upon the users' tracked habit of mobile phone usage, the algorithms proved to be able to not only differentiate refugee and non-refugee users, but also to identify and confirm the individual caller behaviour and to find out

the exact ID of the user with more than 90 per cent success rate. This would also help to get better result from a noisy data which identifies the "refugee users" as a heterogeneous group composed of refugees, foreigners with residence/work permit, international/exchange students, and the ones registered with temporary status.

The main problem that motivates this research stems from the fact that a significant number of Syrians are eventually mobile in Turkey and they do not live in the cities and districts that they are first registered. As a result, there are shortcomings of current migration statistics and there exists an information gap concerning the provincial residency of these mobile refugees, which hinders the policy makers' planning, decision making, efficient allocation of available resources and managing welfare services. For example, most of the Syrians who had been first registered at the border cities just after they entered Turkey, moved to different cities in search for employment and better living conditions. This creates challenges for the public institutions in terms of managing their resources and offering sustainable and adequate public services for the refugees. According to the 2014 Regulation, Syrians can only benefit from public services such as education, access to free health services and social aids in the cities that they are registered (with few exceptions on specific criteria). Therefore, public institutions, civil society organizations and local authorities develop their policies and manage the allocation of resources on the basis of number of registered Syrians in their district. As long as Syrians do not transfer their registrations from one city to another, which is not a common and easy case due to some administrative difficulties, the domestic movement of Syrians creates unexpected burden and aggregation on some public institutions in specific locations. The update of statistics is possible through Syrians' own initiatives to register themselves or transfer their registrations. Consequently, the statistics of DGMM on the exact number of Syrians by province (DGMM, 2018) could have an alternative accurate comparable data to confirm their numbers through such a mobile data analysis with to some extent negligible deviation. In addition to this, despite the fact that DGMM does not share the official number of refugees by district but only by province/city, the mobile data analysis provides a general estimation about the location of refugees for academics and researchers. As a result, based on the problem of determining and confirming the real number of refugees in a specific location, the data mining analysis of the mobile phone calls might facilitate the development of public policies and allocation of welfare resources for the better social integration Syrians in Turkey.

As the second problem, the number of employed Syrians in the labour market of Turkey can only be estimated. This is mainly because Syrians are mostly working as unregistered, as part of informal economy. The Regulation on Provision of Work Permits for People under Temporary Protection, introduced in January 2016, allows Syrians registered under temporary protection to get work permits. The number of issued work permits is reported as 20,968 in 2017 (Turkish Ministry of Labour, Social Services and Family, 2018). The appraised efforts of Turkey to ensure working rights for refugees, the ones under international protection and Syrians under temporary protection is noteworthy which also reflects an inclusive approach of Turkey's evolv-

ing integration policy. Nevertheless, the total number of issued work permits for Syrians has remained rather low with regards to the fact that there are more than 2 million Syrians in Turkey who are at working age. As of September 2018, the number of Syrians who are at the working age (15-64) is 2.051,186 (DGMM, 2018). However it is estimated that almost 800,000 Syrians are working in Turkey as part of informal economy mainly in the sectors of agriculture, textile, construction and services.

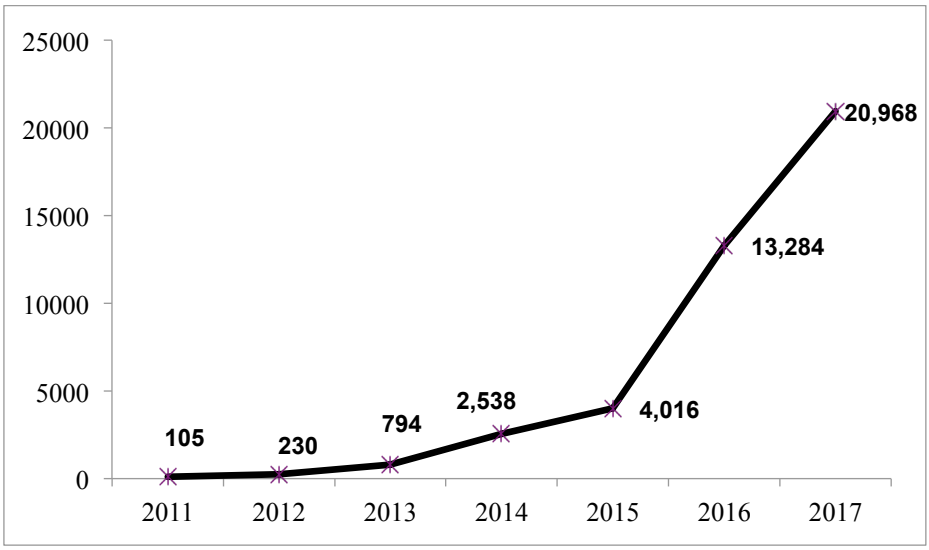

Fig. 1. Number of work permits issued for Syrians in Turkey.
Source: Data compiled for the study from Turkish Ministry of Labour, Social Services and Family Statistics, 2018.

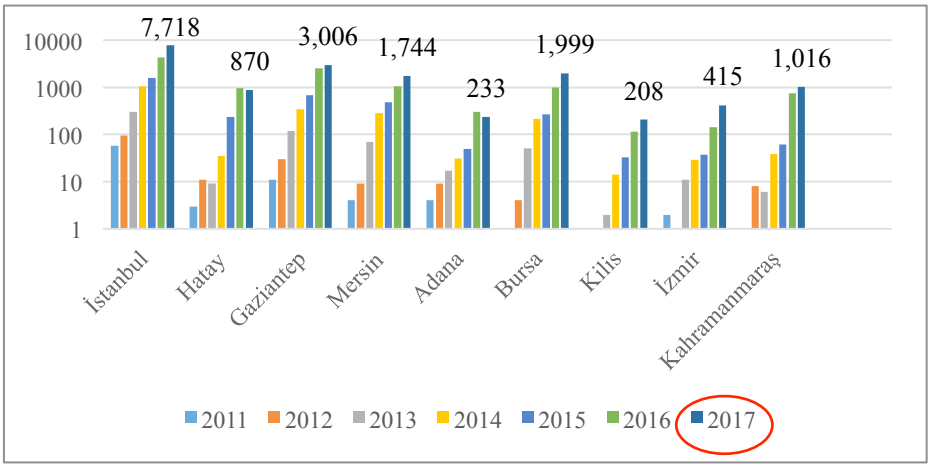

Fig. 2. The number of work permits granted to Syrian nationals by top-ten cities in Turkey. Source: Data compiled for the study from Turkish Ministry of Labour, Social Services and Family Statistics, 2018.

Conducting an analysis on the call hours and the locations of the mobile users can provide a rough idea about where the users spent their day and night times. Although not necessarily to be considered as the working place, the analysis can provide a general hint for the possible status of the refugee, whether unemployed or employed, spending specific hours of the day in a different location, or move back-and-forward from one city to another as seasonal worker. It also provides a picture on patterns of changing conditions, expectations and way of life of refugees. If the gender information of the user is also provided, the data might offer a broader interpretation of the employment status. For example, in Torbalı district of Izmir, it is a crucial problem that many Syrians are working as undocumented labours in agriculture and they are subject to labour exploitation. They live in informal encampment and they are mostly coming to Torbalı from border cities in search for seasonal work. The mobile phone data can at this point provide a rather accurate estimation about counting the uncountable, unregistered Syrian workers.

In this context, this paper aims to provide some insights for the policy makers to help decision-making and policy planning concerning refugee mobility and optimizing the public services for their better social integration. In doing so, it offers a comprehensive analysis based on understanding the communication activities of users and interactions among;

a) Users' mobile phone usage habits,
b) Population structures (distribution of refugees) in specific geographic locations,
c) Timing of the calls of the selected population.

4 Methodology and Data Analysis

This analysis utilizes the large-scale mobile Call Detail Record (CDR) database, which is composed of the anonymized mobile CDR of phone calls and SMS messages of Türk Telekom customers. The dataset is created from 992,457 customers of Türk Telekom, of which 184.949 are tagged as "refugees", and 807,508 as Turkish citizens (Salah et al. p. 3). Keeping in mind the limitation that, it is not certainly possible to address a particular CDR belongs to a refugee or not, based on the patterns of aggregated records (Salah et al.), for each modified dataset, the analysis used 20 anonymous user data which is composed of 10 refugee and 10 non-refugee data. The users are randomly selected from Dataset 2 and Dataset 3.

The analysis is conducted through four steps over the given data in Dataset 2 and Dataset 3;

1) Data normalization,
2) Modifying 6 datasets by merging and associating different data in Database 2 (incoming and outgoing voice call activity, incoming and outgoing SMS activity) and Database 3 (incoming and outgoing voice call activity),
3) Implementing machine learning algorithms,
4) Evaluating the results.

The research findings are concluded by employing two different analytical approaches. First, "basic statistics" is used and the data of users' locations and time stamps of their calls are analyzed in order to understand the communication activities of the refugees. Accordingly, the data is modified to understand the mobile phone usage characteristics of the refugees. First, the time stamp value of the data is taken and then from this value three new columns are generated;

1) The value of day of the month (1,2,3,…,30,31)
2) The value of month of the year (1,2,3,…,11,12)
3) The value of hourly intervals (0,1,2,3,…,22,23)

As another new input which improves datasets, the time period for one day is grouped/split into 3 intervals; 00:00-8:00; 08:00-19:00; 19:00-24:00.

The findings are interpreted and some conclusions have been drawn through a statistical process. Additionally, WEKA data mining software is used to analyse the datasets. The results give an idea about the "time bounded mobility activity" and "time-bounded location" information about the refugees. For this purpose, both Dataset 2 (includes voice and SMS data of a particular group of users who are observed for a period of 2 weeks) and Dataset 3 are used.

In the second approach, "machine learning algorithms" (Decision Tree, Ripper, K-Nearest Neighbours, Hoeffding Tree and Naive Bayes) are applied over the modified datasets, which provide an analysis on the mobile data usage habits/patterns of the given users and then identify the caller IDs and district IDs of the users. The district ID identification is only possible for Dataset 3.

The analysis is conducted by a team of four researchers; two from computer engineering and two from international relations field. Working in collaboration under the UNESCO Chair on International Migration at Yasar University, the research employs a multidisciplinary approach, which is highly lacking in migration studies. In this respect, the expertise of both computer engineers and migration experts are merged and pooled in to provide feasible solutions to the problems of migration management by using available but idle large-scale data.

4.1 Limitations of the Data

The dataset defines the caller "refugee" with reference to the caller's identity card number starting with 98 or 99. In Turkey, not only the ones with international protection status or refugees but all foreigners, regular migrants are provided with these official ID cards by the Turkish government and they are given ID numbers starting with 98 and 99. Thus, the datasets include such biases of a heterogeneous group, which requires different integration policy approaches with different priorities and needs. For example, according to DGMM, as of September 2018, 716,494 residence permits have been issued for foreigners in Turkey (DGMM 2018c). Roughly, this means foreigners who are legally residing, working or studying in Turkey constitutes 19% of the group addressed as "refugees" in the given dataset. Therefore, beyond the fact that mobile market share of Türk Telekom is only 24,7% in Turkey across all operators (Bilgi Teknolojileri ve İletişim Kurumu 2017) the main limitation exists with the representativeness of the data where "refugee" customers are defined in a heterogeneous group of foreigners. As a result this report does not intend to be representative, rather it contributes to the integration policies for refugees through sampling and aspires to present how data mining can be useful for policy making.

Another limitation is since the ID numbers of the mobile phones of the users in the given databases are not identical, the analysis does not able the researchers to connect databases. In Dataset 2, concerning both incoming/outgoing voice and SMS activity the callers are not the same users. This hinders to figure out the users' behaviour of mobile phone communication activities by using enriched data.

4.2 Data Normalization and Modifying Datasets

First, from the original dataset, which is composed of three different callee prefixes, "refugee", "non-refugee" and "unknown", the "unknown" users have been removed. The analysis is based on six new sub-samples generated by the project team; four of them are created from Dataset 2 and two of them from Dataset 3. After sub-sampling, duplications have been removed and both datasets have been modified for the analysis.

In Dataset 2, which is composed of "incoming voice call, outgoing voice call, incoming SMS and outgoing SMS) are modified by adding the city and district data to the dataset. The features in these subsets are "CALLER_ID, SITE_ID, DAY_OF_CALL, MONTH_OF_CALL, HOUR_OF_CALL". The four new datasets are prepared by including the city, district data (acquired from BTS_ID).

In Dataset 3, the data of "incoming voice call" and "outgoing voice call" are modified again by adding city, base station and time stamp data. In total these six datasets are modified for the analysis by removing the duplications. The features in these subsets are CALLER_ID, ID, CITY_ID, DAY_OF_CALL, MONTH_OF_CALL, HOUR_OF_CALL.

5 Evaluation

5.1 Volume of the Communication Activity of Refugees

One of the findings revealed through WEKA data mining tool obviously demonstrates that, the total volume of communication activity of refugees is significantly lower than the non-refugee mobile phone users. The data reveals that while non-refugee users have in total 223,360 communication activities, this remains at only 15,599 activities for refugees.

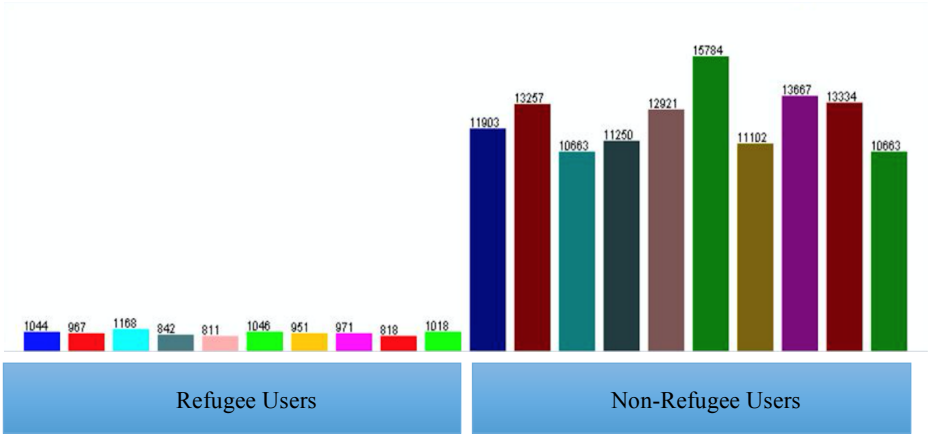

Fig. 3. Outgoing Voice Call (Dataset 3).

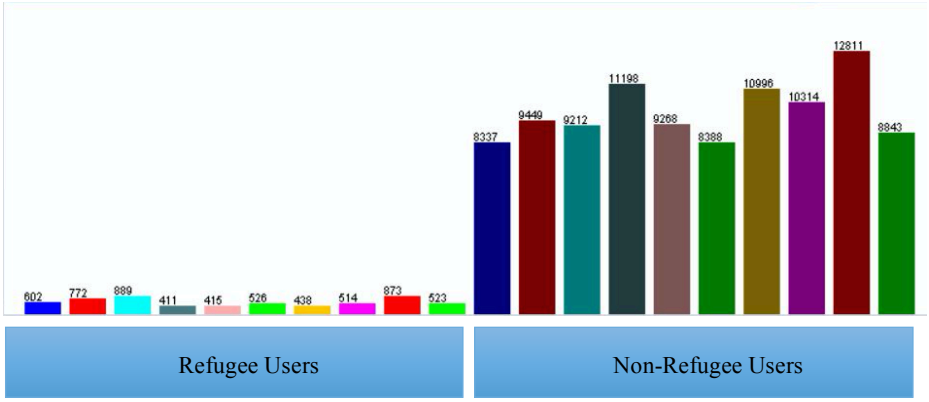

Fig. 4. Incoming Voice Call (Dataset 3).

If further data is made available in terms of the status of called user as well whether he/she is also refugee or not, the analysis might provide insights on social communication network among refugee and host population. Then, the data is able to identify

the social integration of refugees through their social communication networks and the interactions with the host society.

Dataset 2 is analysed in terms of SMS activity and it demonstrated that while refugees had 4,703 SMS activity (2,442 SMS in and 4,703 SMS out), non-refugee users had 14,871 SMS activity in total (7,892 SMS in and 19,574 SMS out). This means 24% of all SMS activities belong refugee users.

Table 1. Refugee Users SMS in (Dataset 2).

Count of HOUR_OF_SMS_IN	Column Labels																						
Day hours	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	Grand Total
ADANA	10	8	6	3								5	2	7	4	14	5	3	8	6	6	11	13	7	118
ANKARA	15	8						1	8	12	13	11	7	17	14	10	18	19	26	22	34	29	33	29	326
ANTALYA	3	3	2	4	3								2	2	2		2	1	3	1	1	5	4	6	44
GAZIANTEP	62	39	12					2	4	14	34	79	103	94	108	116	124	111	102	81	76	80	90	69	1400
ICEL	25	17	8	9	3	4	10	30	10	13	28	22	44	23	18	32	27	36	22	27	19	36	36	55	554
Grand Total	115	75	28	16	6	4	10	33	22	39	75	117	158	143	146	172	176	170	161	137	136	161	176	166	2442

Table 2. Non-refugee Users SMS in (Dataset 2).

Count of HOUR_OF_SMS_IN	Column Labels																								
Day hours	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	Grand Total
BALIKESIR								10	1	2	1	1	5												20
BURSA	244	146	64	37	14	12	31	86	94	114	123	119	166	215	313	343	353	354	282	402	423	457	589	447	5428
ISTANBUL	38	26	16	2	2				2	16	29	26	19	22	37	54	40	76	49	51	35	47	38	49	674
IZMIR	137	105	89	87	56	38	53	56	37	36	28	43	49	54	64	72	77	84	86	110	99	96	104	97	1757
MANISA						5	1	2					1	4											13
Grand Total	419	277	169	126	72	55	85	154	134	168	181	189	240	295	414	469	470	514	417	563	557	600	731	593	7892

Table 3. Refugee Users SMS out (Dataset 2).

Count of HOUR_OF_SMS_OUT	Column Labels																								
Day hours	0	1	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	Grand Total
ANKARA	15	8			1	8	13	14	11	7	19	15	10	18	20	26	23	33	29	32	28	330
ANTALYA						2	18	45	60	44	46	45	51	53	57	32	32	31	33	30	16	595
GAZIANTEP	33	9		1	4	9	17	31	63	69	73	81	70	81	50	58	46	50	50	56	39	890
ICEL	7		6	23	28	6	11	9	13	36	13	11	19	27	48	27	31	12	36	36	47	446
Grand Total	55	17	6	24	33	25	59	99	147	156	151	152	150	179	175	143	132	126	148	154	130	2261

Table 4. Non-refugee Users SMS out (Dataset 2).

Count of HOUR_OF_SMS_OUT	Column Labels																						
Day hours	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	Grand Total
ADANA	12					1	5	8	18	36	22	14	12	14	32	24	30	64	44	56	62	46	70	46	616
AFYON																					6	13	11		30
BILECIK																		2	7						9
BURSA	168	90	36	20	4	3	35	27	16	46	50	63	99	148	213	159	241	228	152	260	225	393	459	275	3410
ESKISEHIR																			1						1
GAZIANTEP	20	9		1		4	3	9	33	69	66	79	79	49	78	61	77	67	76	56	101	88	86	82	1193
HATAY	2	1	2				2				2	3	2	7	6	4	4	3	3	3	3	2	3	3	55
IZMIR	80	56	22	22	18	14	38	24	32	28	4	20	8	22	18	26	50	74	78	104	110	62	98	80	1088
KAYSERI	2					2	2	4	9	7	14	8	20	13	16	18	11	5	11	7	3	4	3	6	165
KILIS														1			3	3	2	2	14	2			27
KMARAS												5	4												9
KONYA	1																						6	2	9
KUTAHYA																			3	8	4				15
MANISA													2	2											4
OSMANIYE							3																		3
SIVAS	9	2			1	6	19	26	24	28	7	17	12	12	14	22	32	19	18	19	15	18	13	12	345
Grand Total	294	158	60	43	23	30	107	98	132	214	165	209	238	268	377	314	448	465	395	515	543	628	749	506	6979

When the voice call and SMS activity ratios are compared it could be observed that refuges are more intended to use SMS activity than voice calling as seen in Table 5 below.

Table 5. Preferred communication activity among refugee and non-refugee customers.

Communication Activity	Refugees	Non-refugees
Voice call	7%	93%
SMS	24%	76%

5.2 Interpreting Communication and Mobility Activity Based on Location and Time

The basic statistics analysis demonstrates the communication activity of refugees by examining their daily mobile phone usage habits (over 24 hours) which gives an idea about some patterns. It provides location information which helps to understand the mobility activity of users. The data on location and time (space-time behaviour) allows the policy makers to identify more precisely where and when the refugees undertake their daily activities. Moreover, it presents an accurate geographic information about the residency of the refugees. The data also demonstrates the daily mobility activity on location basis. The daily mobility can be traced over day and night time period. Within the limitations of the data, for this research the information is limited to 2 weeks, however if the data is extended with the access to one-year period, the whole trajectory of a refugee can be better traced. It might provide information for mobility story and change of location of refugees.

Table 6 demonstrates the mobile phone usage of a selected refugee customer between 1-30 January 2017. The table shows how many times the refugee called someone/somewhere, from where these calls are done and during which hours of the day these calls have taken place. It can be interpreted that the selected refugee high proba-

bly lives in Oğuzeli district of Gaziantep referring to 430 calls made from Oğuzeli. The highest number of calls made during night time and early in the morning (20:00-00:00 and 00:00-07:00) have taken place in Oğuzeli. Table 5, as a complementary data, also supports the fact that the refugee most probably lives in Oğuzeli as 202 calls have been received in Oğuzeli with the highest at night time as well.

Table 6. Number of calls out of a refugee customer in one month (Dataset 3).

Count of HOUR_OF_CALL	Column Labels																		
Day hours	0	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	Grand Total
GAZIANTEP	**6**	**16**	**32**	**31**	**14**	**19**	**29**	**20**	**19**	**24**	**14**	**44**	**49**	**67**	**56**	**75**	**31**	**33**	**579**
ISLAHIYE				3		1													4
KARKAMIS				1				1					2						4
NURDAGI			1				3												4
OGUZELI	6	12	17	21	10	12	15	5	11	5	6	31	35	57	52	71	31	33	430
SAHINBEY		3	13	6	2	1		4	1	10	2	3	4	6	4	2			61
SEHITKAMIL		1	1		2	5	11	10	7	9	6	10	8	4		2			76
HATAY				**2**	**1**	**1**													**4**
ANTAKYA					1														1
HASSA				2		1													3
KILIS			**11**	**43**	**45**	**60**	**24**	**41**	**47**	**45**	**44**	**44**	**24**	**25**	**5**	**2**	**1**		**461**
ELBEYLI			4	26	36	46	20	33	44	41	34	31	9	5					329
POLATELI			1	3			1	1					2	1					9
SEHIRMERKEZI-KILIS			6	14	9	14	3	7	3	4	10	13	13	19	5	2	1		123
Grand Total	**6**	**16**	**43**	**76**	**60**	**80**	**53**	**61**	**66**	**69**	**58**	**88**	**73**	**92**	**61**	**77**	**32**	**33**	**1044**

Table 7. Number of received calls of a refugee customer in one month (Dataset 3).

Count of HOUR_OF_CALL	Column Labels																		
Day Hours	0	1	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	Grand Total
GAZIANTEP	**3**	**1**	**7**	**18**	**19**	**10**	**14**	**22**	**20**	**20**	**14**	**24**	**12**	**23**	**22**	**29**	**37**	**26**	**17**	**338**
ISLAHIYE				1			3													4
KARKAMIS									1											1
NURDAGI				2				3												5
OGUZELI	3	1	6	5	8	7	6	7	5	7	2	10	9	19	19	17	28	26	17	202
SAHINBEY				8	8				5	2	4		1	1	2	9	9			49
SEHITKAMIL			1	2	3	3	5	12	9	11	8	14	2	3	1	3				77
HATAY					**1**	**6**	**3**													**10**
ANTAKYA						6	1													7
HASSA					1		1													2
KIRIKHAN							1													1
KILIS				**7**	**26**	**26**	**34**	**31**	**18**	**15**	**23**	**23**	**20**	**16**	**12**	**3**				**254**
ELBEYLI					3	19	25	27	16	13	19	21	14	3	3					163
POLATELI				1	1								1	1		1				5
SEHIRMERKEZI-KILIS				6	22	7	9	4	2	2	4	2	5	12	9	2				86
Grand Total	**3**	**1**	**7**	**25**	**46**	**42**	**51**	**53**	**38**	**35**	**37**	**47**	**32**	**39**	**34**	**32**	**37**	**26**	**17**	**602**

When the communication activity of this selected refugee is traced over according to the each day of January 2017, the mobility trajectory of him/her becomes much clear.

The following map demonstrates the mobility of the selected refugee based on the number and order of his/her usage of mobile phone. The mobility story is interpreted by taking into account the location and timing of the calls done and received.

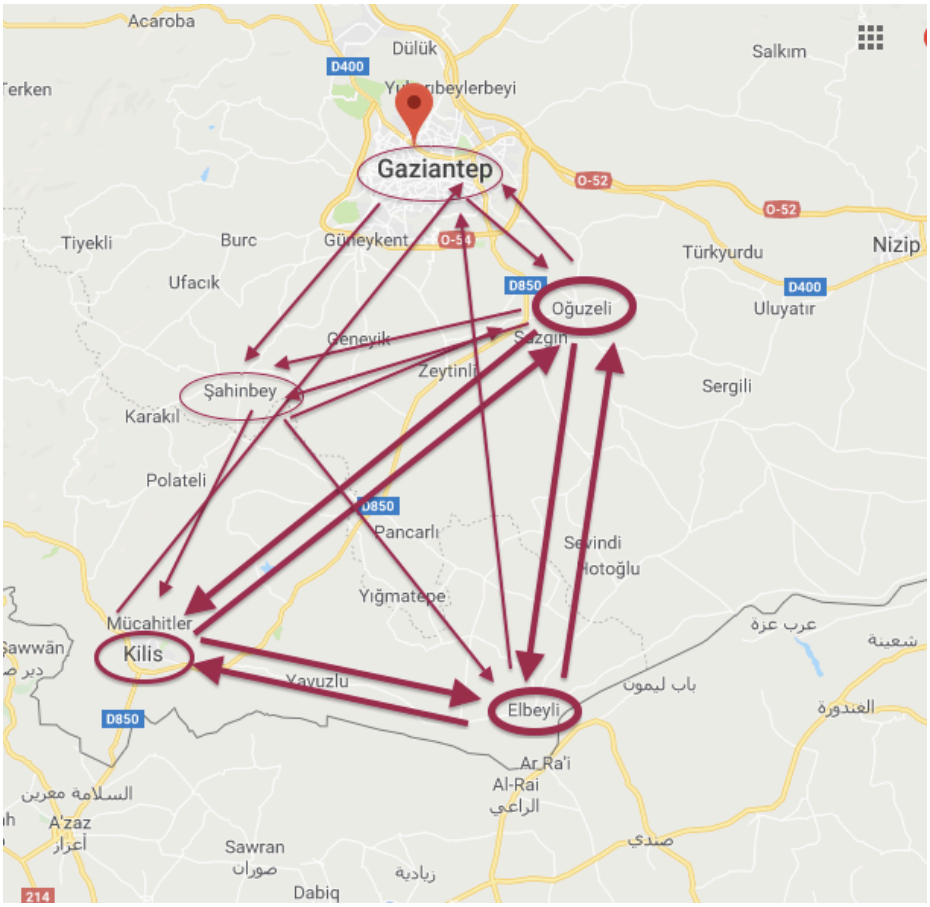

Fig. 5. Original project data sample implemented on Google map.

The mobility trajectory of the refugee on hourly and daily basis gives the idea that the refugee resides in Oğuzeli and he/she spent day time (working hours slot) in Elbeyli. It seems the refugee travels regularly to Elbeyli mostly over Kilis city center and sometimes over Şahinbey. The refugee also had some visits back and forward between Gaziantep/ Şehitkamil and Oğuzeli. Although not certain, it is probable that the refugee might be working in Elbeyli and residing in Oğuzeli according to his/her phone calls (out/in) during specific times of the day over one month.

This method can be applied to all refugee users by merging and interpreting the location and time information acquired from data analysis. By also using GIS tools, further studies are possible to map out the residency, possible employment location

and the route of mobility if exists. These constitute very crucial information to clarify the accurate statistics by considering the existing informality of mobility activities in order to plan better integration policies.

5.3 Identifying the "Caller" and the "Location" by Using Machine Learning Algorithms

Based on the calling habits of the people with reference to the location and timing of the communication activities, some machine learning algorithms are successfully able to identify the caller or the district of the caller. While conducting the data mining over 6 datasets, some of the widely used algorithms are implemented and five of them achieved the best results with high success rates. These algorithms are Decision Tree, Ripper, K-Nearest Neighbours (KNN), Hoeffding Tree and Naive Bayes.

The final results obtained are presented in the tables below. In order to decide the most effective approach to understand data, Total Accuracy (TA), True Positive (TP), True Negative (TN), ROC Area and Precision (PR) rates are taken into consideration to analyse the data.

In modified Dataset 2 and Dataset 3, the algorithms are implemented to identify the callers (CALLER_ID) by examining their autonomous patterns depending upon location of activity and hourly usage. The following algorithms achieved more than 90% accuracy rate in identifying the CALLER_ID.

Table 8. Call in Caller ID (Dataset 2)

D2CALL_IN - CALLER_ID	Decision Tree	Ripper	KNN	Hoeffding Tree	Naive Bayes
TA	0,96944	0,993284	0,9665	0,725531	0,987669
TP	0,97	0,993	0,966	0,726	0,988
PR	0,971	0,994	0,967	0,726	0,988

Table 9. Call out Caller ID (Dataset 2)

D2CALL_OUT - CALLER_ID	Decision Tree	Ripper	KNN	Hoeffding Tree	Naive Bayes
TA	0,857796	0,90625	0,920187	0,53811	0,900479
TP	0,858	0,906	0,92	0,538	0,9
PR	0,862	0,897	0,92	0,545	0,9

Table 10. SMS in Caller ID (Dataset 2)

D2SMS_IN - CALLER_ID	Decision Tree	Ripper	KNN	Hoeffding Tree	Naive Bayes
TA	0,92636	0,956454	0,971163	0,736307	0,98684
TP	0,926	0,956	0,971	0,736	0,987
PR	0,932	0,957	0,971	0,758	0,987

Table 11. SMS out Caller ID (Dataset 2)

SMS_OUT - CALLER_ID	Decision Tree	Ripper	KNN	Hoeffding Tree	Naive Bayes
TA	0,937987	0,902056	0,966991	0,74632	0,979437
TP	0,938	0,902	0,967	0,746	0,979
PR	0,943	0,904	0,968	0,747	0,98

Table 12. Call in Caller ID (Dataset 3)

D3CALL_IN - CALLER_ID	Decision Tree	Ripper	KNN	Hoeffding Tree	Naive Bayes
TA	0,971378	0,968257	0,971149	0,961414	0,963685
TP	0,971	0,968	0,971	0,961	0,964
PR	0,971	0,968	0,971	0,962	0,964

Table 13. Call out Caller ID (Dataset 3)

D3CALL_OUT - CALLER_ID	Decision Tree	Ripper	KNN	Hoeffding Tree	Naive Bayes
TA	0,9287	0,9151	0,940334	0,904375	0,902698
TP	0,929	0,902	0,94	0,904	0,903
PR	0,93	0,91	0,94	0,908	0,903

In modified Dataset 3, the algorithms are able to find out the district (BTS_ID**)** of the caller through analysing the communication pattern of the user based on user and time of activity.

Table 14. Call in BTS ID (Dataset 3)

D3CALL_ IN	Decision Tree	Ripper	KNN	Hoeffding Tree	Naive Bayes
TA	0,88073	0,87062	0,896659	0,842173	0,842468
TP	0,881	0,875	0,897	0,842	0,842
PR	0,851	0,868	0,89	0,806	0,809

Table 15. Call out BTS ID (Dataset 3)

D3CALL_ OUT	Decision Tree	Ripper	KNN	Hoeffding Tree	Naive Bayes
TA	0,849612	0,835421	0,884923	0,786645	0,790475
TP	0,85	0,821	0,885	0,787	0,79
PR	0,848	0,831	0,881	0,758	0,765

This analysis provides a unique contribution to the policy makers of migration in terms of identifying the population structures base on each individual caller characteristics and also the location. The algorithms point out the fact that every caller has specific characteristics/behavior of mobile phone usage and some patterns/commonalities can be drawn for specific groups of population. Beyond differentiating refugee groups among a heterogeneous group of foreigners with divergent needs for integration and different mobile phone usage habits, the algorithms proved to identify each individual caller and their locations. Accordingly, algorithms help the policy makers to verify their statistics in terms of eliminating discrepancies by matching the number of population and their exact location. Moreover, it presents some insights about the communication behavior of refugee groups which can foster their integration by helping the policy makers in terms of offering new ways of communication to reach, guide and support their target group with specially designed tools for mobile phones.

6 Concluding Remarks

Turkey is not only a transit but also a destination country for many migrants, asylum-seekers and refugees who are searching for a better life and humanitarian protection.

Therefore, Turkey is always exposed to new waves of migration movements since it is part of a migration system that is spread over a large geographical area including Europe, Middle East and Africa. Referring to its long history of hosting many migrants and providing protection to many refugee populations, Turkey's integration policy has been developing in recent years. While hosting more than 3.5 million refugees in Turkey in the last 7 years, Turkey once more proved not only its ability to manage such a high number of mass movement, but also the hospitality and tolerance among its host population. However, as the temporary stay of Syrians has been prolonged, there are risks of increasing unrest and decreasing social acceptance among the host society which is strongly linked with the need for a well-established, inclusive and comprehensive integration policy.

Accordingly, not only to cater migrants and refugees but also to keep the peaceful co-existence and social acceptance among the host society, integration policies should manage the allocation of resources properly. Needs and challenges in education, health, housing, employment and social provisions should be carefully determined and planned while policies are being developed. In that respect, accurate information and statistics play significant role since migration involves many informalities concerning the movement of people, change of location and also engagement in labor market. Policy makers who are in dire need of accurate statistics are dependent to their official initiatives, refugees' own engagements with the authorities or some small-scale field work to confirm and update their information and statistics. However, new data sources such as mobile phone data provides a unique contribution to verify and also to provide new data for policy makers.

This research which utilizes the mobile phone data provided by Türk Telekom (Salah et al.), provides some insights for policy makers for better integration policies. Although the study is not representative due to some limitation of data and time, the sample studied demonstrates a methodology that can be utilized for large group of refugee populations as well. The research findings and some policy suggestions drawn from the analysis can be listed as follows:

1. Mobile phone data clearly helps the policy makers to understand how many refugees are actually living in specific locations. This is a crucial new information because as DGMM has provincial/city level directorates that keep the number of refugees in their area of responsibility (exceptions are Antalya, Hatay, Istanbul and Şanlıurfa) the analysis of mobile phone data could not only provide at city level but also at district level information about the actual number of refugees spread in a city. This is also an important information for local authorities such as municipalities. The data provides a unique new input to verify the official statistics and keep track of records of refugees in terms of their movements, residence, work and public service preferences.

2. The detailed analysis of mobile phone data modified according to time and location demonstrates the mobility story of refugees and their main places of visit. The

traced information can give an idea about where the refugee is actually living, whether he/she moved to another place, and even whether he/she might be working depending upon the repeating or regular mobile phone usage hours and location information. Interpretation on employment status is a very important input because it is known that more than 80 % of Syrians who are in the labor market are working as unregistered. The information about the age and gender distribution of this group is also not existing. Thus, the mobile phone data occurs as a trustable source of information to clarify the actual situation in informal labor market which constitutes the very important aspect of integration policies and livelihood practices. If the data is extended with age and gender information the methodology provides a better picture on the socio-demographic mapping of the refugee population and also explains better their mobility trajectory. At this point, data on internet usage (as it is widely accepted that refugees use mostly internet) and could be useful to understand the needs of refugees for better integration policies.

3. Although not representative but the data analysis drawn from the sample reveals that in comparison to non-refugee population, refugees tend to use SMS communication activity. The volume of their voice calls is significantly lower than the non-refugee callers. The reasons need to be explored in a further study but some of the reasons might be related with refugees' low level of income, limited social network in Turkey and low level of daily activity that requires or prioritizes communication.

4. Machine learning algorithms of Decision Tree, Ripper, K-Nearest Neighbours, Hoeffding Tree and Naive Bayes proved to differentiate the individual caller ID and also location depending upon the mobile phone usage pattern of the refugees and non-refugees. This finding is useful to determine the actual living location of refugees and help to verify statistics related to residences and mobility of refugees. Moreover, it ensures information on communication patterns of refugees which provides the policy makers some hints about the ways of effective communication to support and reach out to refugees with better policies.

Based on the conclusions drawn from this data analysis which worked on a sample taken from the given dataset, the study demonstrates that mobile phone data can be used to determine the exact location, mobility trajectory and patterns of communication of large refugee populations. Moreover, a deeper analysis developed by interrelating location and time of the calls presents some insights to interpret possible employment status of the refugee population. Considering the limitations with given data and limited time, the analysis does not come up with ambitious and grandiose policy suggestions but it provides a new methodology which clearly demonstrates how mobile phone data can be utilized for verifying, contributing and updating the actual official statistics on refugees in terms of identification of their location, time-bounded mobility activity and mobile phone communication patterns.

Finally, it is worth mentioning that further relevant research in the field of migration and refugee studies could be possible and feasible if the anonymous data shared is extended with gender and age info of the caller, information on the called number (refugee or not, individual or institution etc.) and duration of the call. Needless to say, these new features would make it more possible for the policy makers to remain relevant and develop well-targeted efficient integration policies.

References

1. Bilgi Teknolojileri ve İletişim Kurumu [Information and Communications Technologies Authority-Turkey], Üç Aylık Pazar Verileri Raporu, 2017 Yılı 1. Çeyrek. Sektörel Araştırma ve Strateji Geliştirme Dairesi Başkanlığı, (June 2017).
2. DGMM 2018a (Turkish Ministry of Interior Directorate General of Migration Management), [International Protection] http://www.goc.gov.tr/icerik6/uluslararasi-koruma_363_378_4712_icerik, last accessed 2018/08/05.
3. DGMM 2018b (Turkish Ministry of Interior Directorate General of Migration Management), [Temporary Protection] http://www.goc.gov.tr/icerik3/gecici-koruma_363_378_4713, last accessed 2018/08/05.
4. DGMM 2018c (Turkish Ministry of Interior Directorate General of Migration Management), [Residence Permits] http://www.goc.gov.tr/icerik3/ikamet-izinleri_363_378_4709, last accessed 2018/08/05.
5. Erdoğan, M.: Türkiye'deki Suriyeliler Toplumsal Kabul ve Uyum [Syrians in Turkey: Social Acceptance and Integration]. İstanbul Bilgi Üniversitesi Yayınları, Istanbul (2015).
6. Frontex [European Agency for the Management of Operational Cooperation at the External Borders of the Member States of the European Union] Risk Analysis Report 2016 (2016). doi:10.2819/416783, last accessed 2018/08/10.
7. International Organization for Migration (IOM): Irregular migrant, refugee arrivals in Europe top one million in 2015: IOM. [Press Release] (22 December 2015), https://www.iom.int/news/irregular-migrant-refugee-arrivals-europe-top-one-million-2015-iom, last accessed 2018/08/14.
8. İçduygu, A.: The politics of international migratory regimes: Transit migration flows in Turkey. International Social Science Journal 52 (165), 357–367 (2000).
9. İçduygu, A., Sert, D.: Country profile Turkey. Focus Migration 5, (April 2009).
10. İçduygu, A.: Turkey and international migration 2012–13. Migration Research Center at Koç University Report, Istanbul (November 2013).
11. Kaiser, B., Kaya, A.: Transformation of migration and asylum policies in Turkey. In: Güney A., Tekin, A. (eds.) The Europeanization of Turkish public policies. Routledge, New York (2015).
12. Kirişçi, K.: Syrian refugees and Turkey's challenges: Going beyond hospitality. Brookings, Washington, D.C. (2014). https://www.brookings.edu/wp-content/uploads/2016/06/Syrian-Refugees-and-Turkeys-Challenges-May-14-2014.pdf, last accessed 2018/08/14.
13. Regulation on Provision of Work Permits for People under Temporary Protection [Geçici Koruma Sağlanan yabancıların Çalışma İzinlerine Dair Yönetmelik], T.C. Resmi Gazete [The Turkish Official Gazette] 2016/8375, 11 January 2016, http://www.resmigazete.gov.tr/eskiler/2016/01/20160115-23.pdf, last accessed 2018/08/10.

14. Salah, A.A., Pentland, A., Lepri, B., Letouzé, E., Vinck, P., de Montjoye, Y.A., Dong, X. and Dağdelen, Ö., 2018. Data for Refugees: The D4R Challenge on Mobility of Syrian Refugees in Turkey. arXiv preprint arXiv:1807.00523.
15. The Report of the High-Level Panel of Eminent Persons on the Post-2015 Development Agenda: A New Global Partnership: Eradicate Poverty and Transform Economies through Sustainable Development (2013), https://www.un.org/sg/sites/www.un.org.sg/files/files/HLP_P2015_Report.pdf, last accessed 2018/08/10.
16. Turkish Ministry of Labour, Social Services and Family Statistics 2018, http://cibs.csgb.gov.tr/, last accessed 2018/08/06.
17. UN High-level Dialogue on International Migration and Development (HLD), United Nations General Assembly Background Paper 3, http://www.un.org/en/ga/68/meetings/migration/pdf/HLD%20RT3%20background%20paper_final%20version%209.9.2013.pdf, last accessed 2018/08/14.
18. UNHCR Turkey: Key Facts and Figures June 2018, http://reporting.unhcr.org/sites/default/files/UNHCR%20Turkey%20Facts%20%26%20Figures%20-%20June%202018.pdf, last accessed 2018/08/10.