

# Intelligent Analytical Data System

Gulnara Bektemyssova, Raisa Uskenbayeva, Yerbolat Kerimbay, Ainur Kerimbay and  
Saniya Kabdrgalinova

International Information Technologies University, 050040, Almaty, Kazakhstan  
g.bektemisova@gmail.com

**Abstract.** This article discusses the development of the intelligent – analytical system aiming to analyze the immigration of the population of the Republic of Kazakhstan in recent years. The results showed a greater outflow of the population. To determine the causes of this process and analyze them the intelligent – analytical system has been developed. The system uses Big Data technology that allows us to compare and contrast the real-time information from various data sources. Because of testing the real-time pilot system, “The dynamics of change” was obtained based on “age and gender”, and “the external immigration” of the population as well as “the distribution of the population and the internal migration” taking place in the city of Nur-Sultan.

**Keywords:** intelligent systems, analytics, immigration, registration of citizens, big data, data analysis.

## 1 Introduction

### 1.1 The statistical reliability analysis

By July 1, 2019, the number of residents of Kazakhstan amounted to 18.27 million people. In 90% of cases, Kazakhstan relocates to the countries of the Commonwealth of Independent States (hereinafter referred to as the CIS). The migration of the population in the post-Soviet space takes place mainly in the form of labor migration or repatriation. [1]. Many factors contributed to the increase in dynamics. First of all, the economic one: the able-bodied population moves to the centers of economic growth to increase their income, professional growth, and self-realization.

Most often Kazakhstanis move to the Russian Federation - 15.07 thousand people in January-June 2018, or 87.8% of the total flow of external migrants. Germany is in the second place, where 1,097 people went to permanent residence (hereinafter referred to as permanent residence), or 6.4% of the total number of those who left. To the United States for six months, 158 people moved from Kazakhstan; over the same period of the previous year, the number of

emigrants to the United States of America was 131. For the same period, 116 Kazakhstanis changed their citizenship to Belarusian. In the first half of this year, 74 Kazakhstanis moved to the Kyrgyz Republic. Canada closes the top five most popular areas of external migration, where 66 people moved to a permanent place of residence for six months. Since labor migration requires higher education as a competitive advantage, qualified specialists prevail among emigrants. At the same time, the decision to change the place of residence is made by the adult population, who mainly have work experience and qualifications required in the host country.

In addition, ethnic Kazakhs are leaving the country. In six months, 735 Kazakhstanis moved from the country abroad. Out of these, 517 preferred the countries of the Commonwealth, 218 moved to other states. Most of the Kazakhstanis left the city of Almaty - 110 people, Astana - 100 people, and the Almaty region - 73 people.

The top ten ethnic groups whose representatives left Kazakhstan in the first half of 2018 included Tatars - 404 people, Poles - 264 people, Belarusians - 208 people, Koreans - 104 people, Chechens - 74 people and Uzbeks - 58 people. The influx of migrants to Kazakhstan is provided due to ethnic Kazakhs, it is more than half. In the first half of 2018, 3.46 thousand Kazakhs arrived in the country, of which 67.1% were from the CIS countries. In January-June 2018, 1,441 Russians, 169 Ukrainians, 161 Azerbaijanis, and 148 Uzbeks arrived in Kazakhstan.

We carried out a comparative analysis of the information data of the Federal Service for Statistics of the Russian Federation (hereinafter - the RF) and the Statistics Committee of the Ministry of National Economy of the Republic of Kazakhstan (hereinafter - the MNE of RK) on the citizens' migration. The results of the analysis show data inconsistency. For 5 months of 2018, the Ministry of Internal Affairs of the Republic of Kazakhstan (hereinafter referred to as MIA of RK) issued documents for 39,891 Kazakhstanis leaving for permanent residence [LSM reports]. For comparison, for the whole of 2017, according to the Committee on Statistics, fewer people left Kazakhstan - 37,725 people. At the end of 2018, the emigration rate increased by 2 times. The largest number of travel documents was issued in Kostanay, East Kazakhstan, Karaganda, Pavlodar regions and Almaty. Departing immigrant to Russia, Germany, Poland, the USA, Canada, and Israel.

Table 1 shows the comparison of the data of the Federal Statistics Service of the Russian Federation and the data of the Statistics Committee of the Ministry of National Economy of the Republic of Kazakhstan. In the results of the tabular data, there is a difference in numbers by almost 0.5 million people [2].

**Table 1.** Comparison of Russian and Kazakhstan data on emigration.

Year	Arrived from RK to the RF, according to RK	Left RK according to data from RK	Departed from the RF to RK, according to the RF	Arrived in RK in total according to RK
2000	124 903	155 749	17 913	47 442
2001	65 226	141 710	15 186	53 548
2002	55 706	120 223	13 939	58 211
2003	29 552	73 890	14 017	65 584
2004	40 150	65 530	12 504	68 319
2005	51 945	52 139	12 437	74 807
2006	38 606	33 690	11 948	66 731
2007	40 258	42 435	10 211	53 397
2008	39 964	45 287	7 483	46 404
2009	38 830	33 985	7 232	41 511
2010	27 862	26 541	7 329	42 057
2011	36 474	32 920	6 176	38 016
2012	45 506	29 722	8 843	28 296
2013	51 958	24 384	11 802	24 105
2014	59 142	28 946	18 328	16 784
2015	65 750	30 047	30 983	16 581
2016	69 356	34 900	32 226	13 755
2017		37 725		15 595

The analysis results show that the data are not consistent. In recent years, data began to vary much more significantly. From 2005 to 2016, according to Russian data, 565.6 thousand Kazakhstanis moved to the Russian Federation for permanent residence. At the same time, Kazakhstan believes that only 415 thousand citizens have left. If we assume that 70% of those who decided to leave Kazakhstan move to Russia, then the total emigration from 2005 to 2016 will be 200 thousand people more than they think in Kazakhstan. And if we assume that Russian data is only 55% of the total number of those who left, then the number of people who left Kazakhstan is approximately 200-600 thousand people, but no one took this into account and did not include it in official statistics.

If we talk about emigration from Russia to Kazakhstan, discrepancies are just as easy, since 2014, the number of emigrants from Russia to Kazakhstan, according to Russian data, has begun to exceed the data on general immigration to the country. There is a reason for large double-checks of data. Moreover, it is so not only in terms of Russia. Migration data is one of the most obvious examples that cast doubt on official Kazakhstani statistics. These data differ significantly from the figures that are statistics of other countries. Moreover, they diverge at

times.

For example, the Committee on Statistics of Kazakhstan indicates that 48.7 thousand immigrants from Russia arrived in the country in 2009-2017, and 272.1 thousand people left for the neighboring country (Russia is the number one destination among Kazakh emigrants). At the same time, the Russian Committee on Statistics provides other data: 161.8 thousand people left for Kazakhstan, according to his information, and 466.6 thousand people moved from Kazakhstan. Thus, the data of Kazakhstan and Russian statisticians differ by 2-3 times. At the same time, Russian statistics indicate that the immigration from Kazakhstan, has grown significantly in the last three years, Kazakhstan statistics confirm this trend. However, Roskomstat's numbers are twice as high as those of the Kazakhstan Statistical Office are.

There are similar differences in data, for example, with Canada. According to the Canadian Government's open data portal, 2,618 people immigrated to Canada from Kazakhstan in 2009-2015. Data for the same period of the Kazakhstan Committee on Statistics give a figure of exactly 1,000 people [3].

There is the five-fold discrepancy in the data on the number of emigrants in the USA. According to the Migration Policy Institute, in 2010-2017, 7,634 people legally moved from Kazakhstan to the United States, the Committee on Statistics of Kazakhstan cite a figure of 1,643 people.

Theoretically, the statistics themselves can be so distorted by migrants themselves, who do not formally formalize permanent residence. However, the differences in numbers are so significant that this is hardly the only reason.

## **2 Integrated information system**

### **2.1 Tasks and functions solved in the system management practice**

Today in Kazakhstan there is an Integrated Information System for Population Service Centers (hereinafter - IIS PSC, system) - this is an information system designed to automate the process of providing public services to the population through Population Centers (hereinafter - PSC, Center).

The main idea of creating a public service center is to provide comfortable and high-quality conditions for the provision of public services to the population based on the "single window" principle. Its idea to become a single access point and to manage information on the provision of services and organize developed monitoring and control over the implementation of regulations.

The introduction of the system has greatly simplified the processes of interaction between the population and state bodies. Because of that, the bureaucratic delays arising from the provision of public services as well as the number of government bodies that must be visited by the applicant have been reduced; and the efficiency

of operators has increased due to the implementation of a performance indicator monitoring system.

Today in the republic, 313 public service centers provide 332 state services that are included in the registration of internal and external immigration. Currently, the creation of a public service center was a positive shift in the provision of public services and has become a kind of platform for e-government, which has allowed the integration of different information systems in one database.

Despite the operation of the PSC system, the statistics on external immigration of the Committee on Statistics of the Ministry of National Economy of the Republic of Kazakhstan do not correspond with the statistics of those countries where citizens of Kazakhstan are much more immigrant [4].

In connection with the above problems, we decided to develop an intellectual and analytical system for registering the immigration movement of citizens of the Republic of Kazakhstan.

The purposes of this system are to create a single database of 313 public service centers of the Republic of Kazakhstan on internal and external immigration and to develop intelligent reporting in real-time according to different immigration criteria while using big data processing technologies.

The use of big data analysis makes it possible in the future to analyze critical regions of immigration, to determine the reasons for immigration and the ability of the state to take preventive measures before immigrating citizens.

The intelligent – analytical system (hereinafter - IAS, system) is necessary to create a single access point to personal data and statistical reports, as well as to improve monitoring of the processes of registration of the population of the Republic of Kazakhstan on immigration.

Thus, the main tasks and functions of the System in practice are:

- the formation of intelligent reporting forms on existing data;
- providing interested parties with the necessary information to prevent risks;
- optimization of data input and data processing.

The development of the System provides for the implementation of the following functions and tasks:

- providing a mechanism for implementing automated data collection from various Information systems of State bodies (hereinafter - IS GO), by presenting the functionality of existing systems in the form of services and the interaction of these services among themselves, i.e. data collection may involve the interaction of many services;
- providing tools and methods for generating analytical information with further uploading to external systems;
- differentiation of access rights to personal data of a citizen;
- providing a standardized way of exchanging information that facilitates the interaction between service endpoints.

## 2.2 Analysis of existing control and management practices

Intelligent-analytical systems (hereinafter - IAS) will be called a large class of IP designed for analytical processing of large amounts of information that implies the presence of its source (s) [5], [6].

Thus, such an IP should include at least two large subsystems: for collecting information and processing it. The task of analytical information processing is secondary to the tasks of collecting and efficient data organization.

Therefore, two classes of systems are distinguished (Table 2):

- Data processing systems (DPS);
- Decision support systems (DSS).

Modern DPSs are developed using the OLTP methodology - On-Line Transaction Processing – the operational transaction processing; and DSS - following the OLAP methodology - On-Line Analytical Processing – the operational analytical processing. OLAP implementation is discussed in more details in [9].

**Table 2.** Comparison of Russian and Kazakhstan data on emigration.

Characteristic	DPS	DSS
The update frequency	High frequency, in small portions	Low frequency, in large portions
Data sources	Mostly internal	Mostly external
The volume of data stored	Hundred megabytes, gigabytes	Gigabytes and terabytes
the data age	Current (for the period from several months to one year)	Current and historical (over a period of several years, tens of years)
Purpose	Capture, search and transform data	Storage of detailed and aggregated historical data, analytical processing, forecasting and modeling

It should be noted that in the analysis and design of DPS, the general principles of designing systems using databases (DB) are used.

## 2.3 The system requirements in general

In general, the system meets the following requirements:

- Efficiency - to ensure prompt and timely implementation of the required functions and user services;
- Scalability - the ability to expand, if necessary, functional capabilities and the availability of a mechanism to ensure the normal functioning of the System with an increase in the number of users, growth in the volume of data, the number of processed documents, and etc.;

- Fault tolerance - resistance to user errors and various technical failures;
- Component structure - the ability to build the System based on the component structure, providing maximum flexibility and, if necessary, not to develop software to solve typical problems anew, but to use existing standard solutions;
- Technical means - ensuring reliable and an uninterrupted operation of the System and the correct completion of the work in the event of an emergency;
- Modularity - providing an optimal starting configuration with the possibility of phased development of the System in the future, the ability to customize to changing requirements during an operation;
- Data recovery after accidents - the availability of tools to restore databases in case of their destruction in emergencies and technical server failures.

## **2.4 System modules and their purposes**

The system is deployed using a centralized deployment topology. The system is installed in one data center (hereinafter referred to as the DC) and all offices must access the System through the channel of the State authorities.

The system consists of two work areas:

- The workspace of the operator;
- The workspace of the analyst.

The workspace of the operator contains 4 Modules:

- Entering information on registration of citizens of the Republic of Kazakhstan;
- Entering information on deregistration;
- Entering information about a citizen;
- Entering information about the property.

The workspace of the analyst is designed to generate statistical data from existing information in the database. This work area should be located in a limited access circuit and be accessible only to employees of the Public Service Coordination Department and the Public Services Monitoring Department.

## **2.5 Prospects for the development and modernization of the system**

When creating the System, development prospects and the possibility of its subsequent modernization are provided. The following areas of development are provided as well:

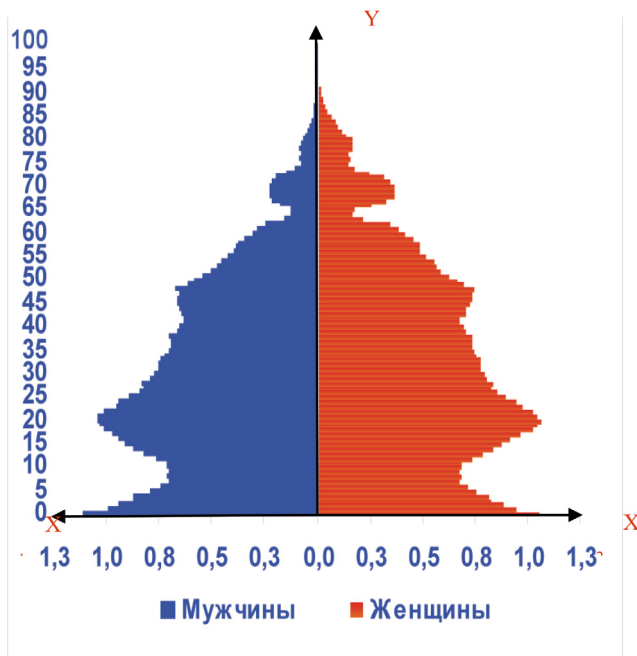
- The possibility of expanding the number of users of the System, expanding application functions and tasks, and increasing the amount of information accumulated in the System should be ensured by observing the requirements for scalability and openness of the System;
- The possibility of prospects for integrating the System with internal IS GO;
- The ability to custom data analytics [7].

## 2.6 System analytics

To test the intellectual and analytical system, we obtained the data of 1,006,574 (one million six thousand five hundred and seventy-four) citizens from the public service center of Astana City for 2018 [8]. Since the data of citizens is confidential, the data of past years are used. Because of testing the system, analytics was obtained according to the following criteria:

- employed population – 789,741 people (the employed population indicator is calculated from the total number of employed citizens aged 15 and older based on the current date);
- the number of unemployed is 150,456 people (number of unemployed citizens aged 15 and older with the status of “unemployed” at the date of formation);
- the number of disabled people (an indicator of children aged 0 to 14, as well as adults 65 and older, whose status is “does not have a job”).

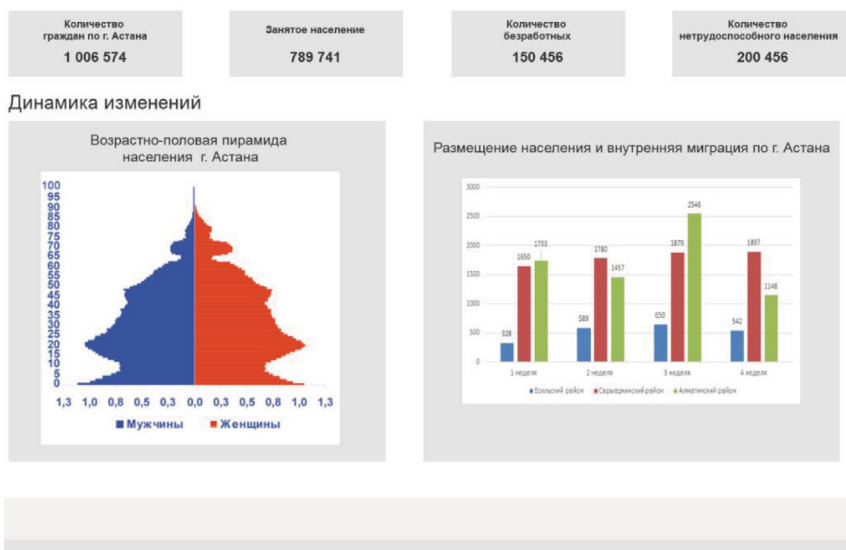
As a result, analytical reports are also generated on the “level of education”, the type of activity and on income, etc. Analytics shows statistics on indicators for the current date of its formation. The chart values are plotted in real-time with current values for the current date without specifying parameters. The analytics results are shown in Fig. 1.



**Fig. 1.** A figure caption is always placed below the illustration. Short captions are centered, while long ones are justified. The macro button chooses the correct format automatically.



## Diagram “The age and gender pyramid”



**Fig. 2.** The age and gender pyramid of migration

**Charting technique:** The age and gender pyramid is a graphical representation of the distribution of the population by gender and age, used to characterize the gender and age composition of the population.

The age and gender composition of the population is the ratio of age and gender groups – the population of people of the same age. This is the main element of the age structure of the population. The age and gender pyramid is built in groups with an aging step of 5 years.

The diagram for age and gender pyramid, shows the number of people by gender and age with an interval of 5 years, see Figure 5.

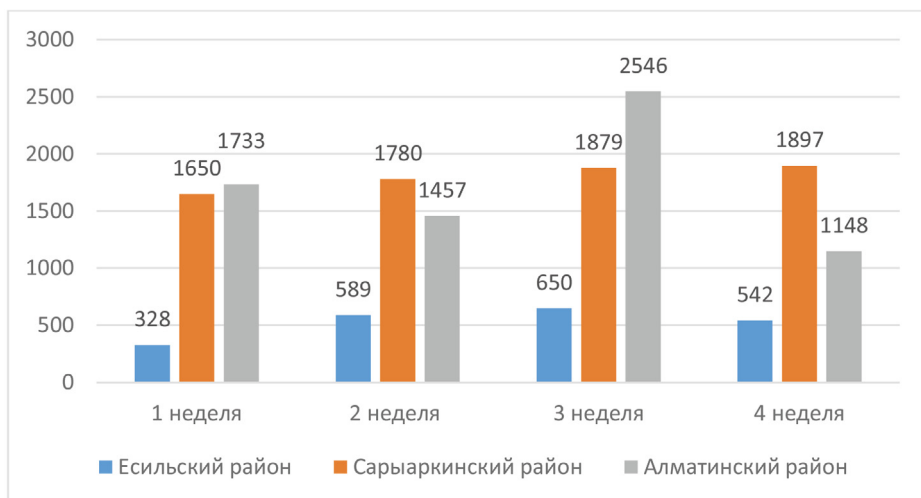
When constructing a chart data on the number of men and women should be taken into account in the ratio of 1:1000.

This chart should be based on data from the Register of “Individuals”.

### **Chart “The population distribution and the internal migration in Astana”**

This chart includes data on two indicators: population distribution in Astana and internal migration by districts of Astana.

The diagram is built based on the number of citizens by city districts in the context of the last 4 weeks to view the dynamics of internal migration by districts of Astana, see Figure 6.



**Fig. 3.** Diagram “The population distribution and the internal migration in Astana”

**Charting technique:** a chart is built based on the ratio of the total number of citizens in the district that is current at the current date to the territory of each district of Astana.

To view the dynamics of the population distribution by regions, the System should take into account data for a month divided by 4 weeks.

### 3 Conclusion

This article discusses the solution to the problem of analysis and accounting of population immigration in the Republic of Kazakhstan. We have analyzed the external and internal immigration of the population over the past five years – 2013-2018. Today in the Republic of Kazakhstan, there are Population Service Centers where internal and external immigration of the population is registered. This information is merged into a common database in the city of Nur Sultan.

The results of our analytical study showed that information on the immigration of the population of Kazakhstan and Russian statisticians diverge 2-3 times. At the same time, Russian statistics indicate that the immigration from Kazakhstan has grown significantly in the last three years. Kazakhstan statistics confirm this trend. However, Roskomstat's numbers are twice as high as those of the Kazakhstan Statistical Office are.

Similar differences are there in data, for example, with Canada. According to the Canadian Government's open data portal, 2,618 people immigrated to Canada from Kazakhstan in 2009-2015. Data for the same period of the Kazakhstan Committee on Statistics give a figure of exactly 1,000 people.

There is the five-fold discrepancy in the data on the number of immigrants to the USA. According to the Migration Policy Institute, in 2010-2017, 7,634 people legally moved from Kazakhstan to the United States, the Committee on Statistics of Kazakhstan cite a figure of 1,643 people.

To solve the problem, we proposed the development of an intelligent information system of an intelligent-analytical system for analyzing the immigration of the population of the Republic of Kazakhstan. Currently, this system has been developed and a test version of the system has been launched. To test the intelligent-analytical system, the data of 1,006,574 (one million six thousand five hundred and seventy-four) citizens from the public service center of Astana City for 2018 was obtained. The test was successful. As a result, we received “The dynamics of change” and “The population distribution and the internal migration in Astana” in real-time.

In the future, our task is to use deep learning algorithms, and according to the results of the analysis, determine the factors that affect immigration and develop algorithms for the early prevention of immigration.

## References

1. Zhussupova, A. <https://kursiv.kz/news/tendencii-i-issledovaniya/2018-08/migraciya-kak-ugroza-ekonomicheskoy-bezopasnosti-pochemu>, 2018. [in the Internet].
2. Toktarov, E. <https://kursiv.kz/news/tendencii-i-issledovaniya/2018-08/migraciya-kak-ugroza-ekonomicheskoy-bezopasnosti-pochemu>, 2018. [in the Internet].
3. Beissebaev, K. <https://kursiv.kz/news/obschestvo/2019-02/kazakhstancy-prodolzhayut-pokidat-rodinu-kolichestvo-emigrantov-iz>, 2019.
4. Hulten, J. Building Intelligent Systems, Moscow: Black and White, 2019.
5. Yevgeniyev, G.B. Intellectual project systems, Moscow: MSTU named after Bauman, 2012.
6. Zhussupova, A. <https://kursiv.kz/news/obschestvo/2019-02/kazakhstancy-prodolzhayut-pokidat-rodinu-kolichestvo-emigrantov-iz>, [in the Internet].
7. Rassel, P. S. Artificial Intelligence. The Modern Approach, Moscow: Williams, 2007.
8. Bektemyssova, G. Zhuanysh, I. The development of a web application for the automatic analysis of the tonality of texts based on machine learning methods. In International Conference on Control, Automation and Systems, pp. 101-105, IEEE, (2018).
9. Sarsimbaeva S., Dimitrov V., Multidimensional Data Analysis with OLAP, Proceedings of the Information Systems and Grid Technologies, Sofia, Bulgaria, November 16-17, 2018, online, <http://ceur-ws.org/Vol-2464/paper10.pdf>.