Astana tourism statistics
on the basis of mobile data positioning

Experimental Statistics
Working Paper №2024-04

Astana 2024

About experimental statistics

Experimental statistics are indicators and tools that are in the testing phase and are not part of official statistics. Indicators and tools are based on innovative methods and/or new data sources. Given the collection of feedback, approaches and results may change as the final methodology takes shape.

Both the maturity and quality of the results differ from those of official statistics, especially with regard to harmonization, coverage and methodology. Nevertheless, these results offer new and interesting ways of looking at various statistical topics that may later become official statistics.

Thus, experimental statistics are designed to test new approaches and collect feedback from users and, therefore, cannot be used as official statistics.

Within the framework of experimental statistics the confidentiality of used primary statistical data is preserved, according to the paragraphs 1-7 of the article 8 «Provision of guarantees of confidentiality and protection of submitted data» of the Law of the Republic of Kazakhstan "On state statistics" from March 19, 2010 No. 257-IV.

1. Key points

The use of cellular operator data to generate tourism statistics presents a number of opportunities for their analysis and application:

- movement analysis: data on subscriber movements can help to identify popular tourist destinations. By tracking the location of phones at different time intervals, it is possible to determine the density of tourist flow in different regions;

- determining the length of stay: using data on the time spent in a particular location, it is possible to estimate the average length of stay of tourists in different locations. This will help in analyzing the preferences and interests of visitors, as well as the number of days of stay;

- estimation of tourist congestion: using travel data, it is possible to estimate the congestion of different tourist facilities at different periods of time. This will help in managing tourist flows and optimizing infrastructure.

- tourist classification: mobile positioning data can be used to identify key socio-demographic groups, categorize by current location and create a tourist portrait;

- based on historical data on tourist movements, models can be developed to forecast demand for services and goods in different regions. This will help businesses and government agencies to plan their activities and resources;

- seasonality monitoring: analyzing travel data will identify seasonal trends in tourist activity and demand for various services. This will help to manage resources more effectively and adapt the offer to the changing needs of tourists will help more

2. Introduction

The Bureau of National Statistics of the Agency on Strategic Planning and Reforms of the Republic of Kazakhstan within the framework of implementation of Section 2 of the Roadmap of development of state statistics and national data ecosystem for 2023-2025 carries out works on implementation and use of alternative sources in the formation of statistical information for users. Within the framework of this direction works on obtaining data from cellular operators and formation of tourism statistics on the basis of mobile positioning data are carried out.

In international statistical practice mobile positioning data are an additional source of data and are widely used to obtain social and economic indicators and to refine information for timely and objective population statistics. Obtaining data from cellular operators will allow the formation of relevant and complete official statistical information on a wide range of industries, one of which is tourism. The United States of America, France, Spain, China, South Korea, Indonesia and Georgia have successful examples of using cellular data to collect tourism statistics, helping them to better manage the tourism sector and make informed decisions on industry and tourism development.

One of the key aspects of using mobile operator data is to monitor the flow of tourists, analyze their behavior and make strategic decisions in tourism through mobile positioning technologies. For example, data on tourists' movements can be used to identify popular tourist routes and destinations, and to forecast demand for tourism sector services.

The use of mobile operator data allows for more accurate analysis of tourist flow, identifying trends and predicting demand for tourism services. This in turn helps public and private entities to make more informed decisions in the field of tourism industry development and improving the quality of tourist services, as well as to assess the impact of tourism on the economy.

This paper presents a comparative analysis of data obtained from the mobile operator KCell *(whose market share is 29.9%)*, with information from the information system «eQonaq», as well as official statistical information generated on the basis of national statistical observations «Report on the activities of accommodation facilities» *(index - 2-tourism, frequency - quarterly)* and «Household Travel Expenditure Survey Questionnaire» *(index H-050, frequency - annual).*

3. Methodology

The main source of mobile positioning information is the load on cellular operators' base stations that record the movements of cellular subscribers, i.e. active positioning.

Mobile positioning data is cell phone event data that includes subscriber ID, time attribute, and location.

Active positioning data is generated by determining the location of the mobile subscriber using device-centric and network-centric methods, as well as via satellite (e.g., GPS). These techniques are used either to provide services at a specific location or in response to national regulations requiring the collection of highly accurate location data. Active positioning capabilities are common in operator systems in developing countries and are often only used on a case-by-case basis.

A cell phone event is considered as an action initiated by or intended for a subscriber or mobile device. Most of these events generate *passive positioning data*. When these events occur, they are registered by different network entities. For example, when a mobile device initiates a location zone update, a database such as a VLR stores a new location area code (LAC) for the mobile device. Similarly, a location is generated each time a subscriber utilizes services such as incoming and/or outgoing calls (CDR), incoming and/or outgoing messaging, or internet access (IPDR). Event data may also come from sensing or signaling source devices and include CDRs and IPDRs, as well as from other technical events such as location zone updates or other operational actions that generate an event together with a link to a cell tower. For each such event, the mobile network collects data related to the event, which means that the list of characteristics is often event-dependent. Moreover, only the cellular network can record some characteristics, which also makes their presence specific to a particular cellular network.

The home location is determined for each subscriber and is based on which cell the cell phone is in at night or when it is first switched on in the morning. This is done to identify recurring patterns and improve accuracy in determining home location. Many employees do not use the address specified in personal contracts, so this information is considered less reliable for home location.

Work location is more difficult to model than residential location. In this case, the location given is the location where the cell phone is located during normal working hours on normal working days (Monday through Friday). The method used to determine the work location also requires that the cell phone subscriber travels at least twice a week for a four-week period. He also needs to spend a sufficient amount of time there.

Operator subscribers whose home and work locations do not coincide with the analyzed region are assumed to be internal tourists. For example, a subscriber lives in Karaganda, but visited Astana. In this case, home and work locations are analyzed based on the subscriber's behavior over the last 60 days with weekly updates.

Subscribers of non-Kazakhstani cellular operators roaming within the territory of the Republic of Kazakhstan are assumed to be **external tourists**. However, this category does not include citizens of other countries who purchase SIM cards of Kazakhstani operators.

Allocation by country of stay is based on the processing of technical parameters transmitted by the subscriber's SIM card MobileCountryCode (MCC).

4. Results

Internal visitors to Astana

According to the mobile operator KCell for the first three months of 2024 in Astana was recorded 542.3 thousand internal visitors from the regions of the Republic of Kazakhstan. At the same time, according to the results of the national statistical observation "Report on the activities of places of accommodation" *(index - 2-tourism, periodicity - quarterly)* during the first quarter in Astana accommodation places served 294.5 thousand people, which is 1.8 times less than the data of the cellular operator. It is important to note that statistical surveys do not take into account self-organized visitors staying with relatives and friends, while the data of mobile operators include all visitors, regardless of the type of their accommodation.

According to mobile operator KCell's data, the highest number of internal visitors was recorded in January and amounted to 186.3 thousand people or 34.4% of the total number of visitors for the quarter.

Number of internal visitors to Astana *

person
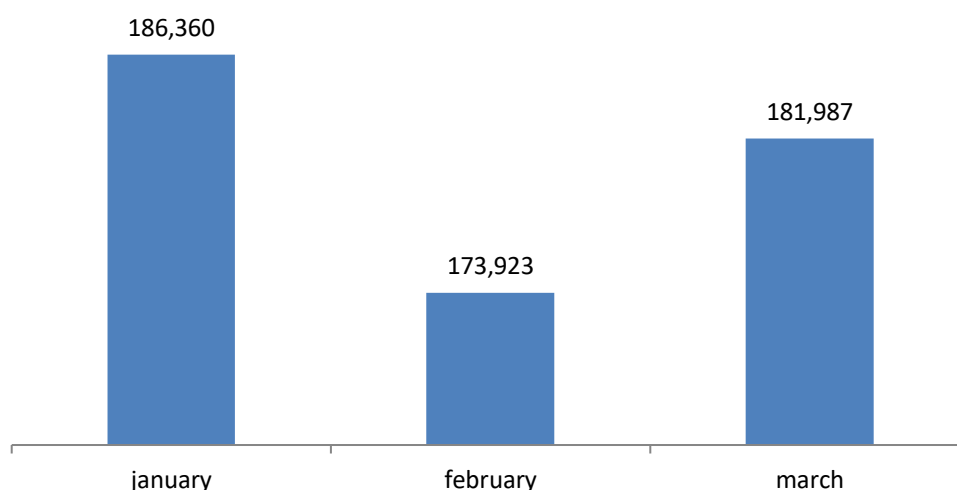
* *According to the mobile operator KCell.*

In February, there was a decrease in tourist flow to 173.9 thousand people or by 6.7%. In March, there was a slight increase in the number of arriving visitors to 182 thousand people, which is probably due to the increased attractiveness of the capital in connection with various events and the number of weekends and holidays in honor of Nauryz celebration.
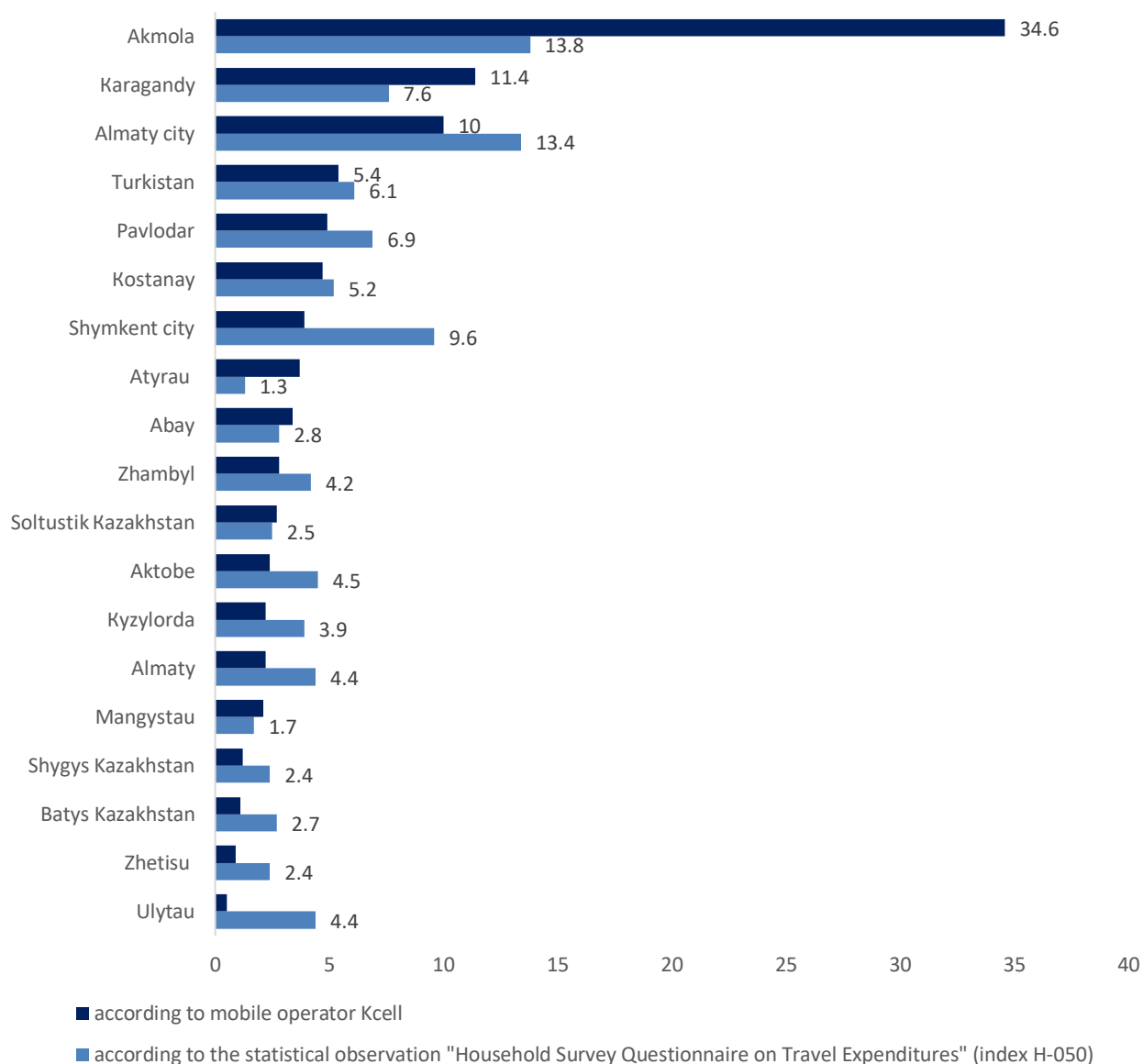
The key regions-donors of internal tourists for Astana are the neighboring regions and large cities.

According to KCell mobile operator data for the first quarter of 2024, the largest number of internal visitors came from Akmola region (34.6%), Karaganda region (11.4%) and Almaty city (10%).

At the same time, according to the data of the national statistical observation «Questionnaire of household survey on travel expenses» (*index - H-050, periodicity - annual)* the main share falls on Akmola region (13.8%), Almaty city (13.4%), Shymkent city (9.6%) and Karaganda region (9.6%).

Share of internal visitors to Astana by region of arrival

percentage

Region	Kcell	H-050
Akmola	34.6	13.8
Karagandy	11.4	7.6
Almaty city	10	13.4
Turkistan	5.4	6.1
Pavlodar		6.9
Kostanay		5.2
Shymkent city		9.6
Atyrau		1.3
Abay		2.8
Zhambyl		4.2
Soltustik Kazakhstan		2.5
Aktobe		4.5
Kyzylorda		3.9
Almaty		4.4
Mangystau		1.7
Shygys Kazakhstan		2.4
Batys Kazakhstan		2.7
Zhetisu		2.4
Ulytau		4.4

■ according to mobile operator Kcell

■ according to the statistical observation "Household Survey Questionnaire on Travel Expenditures" (index H-050)
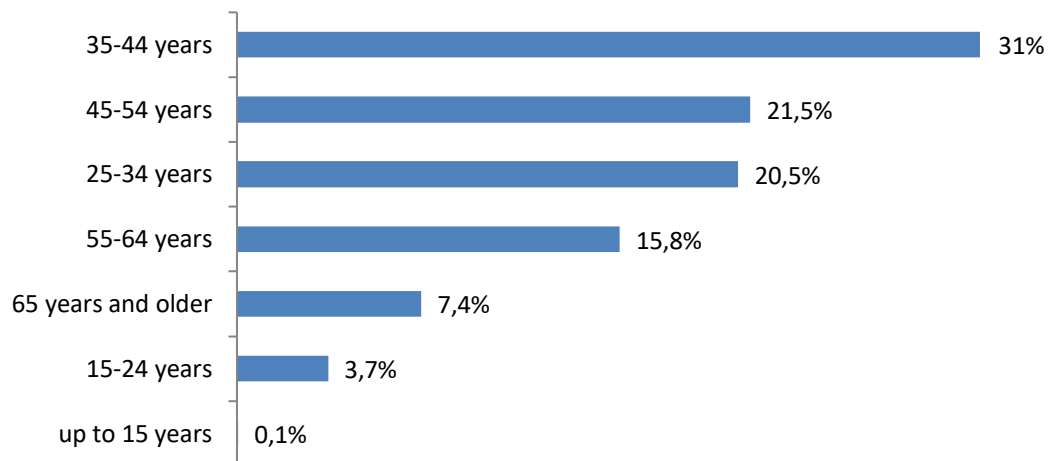
It should be noted that mobile positioning data allow to analyze more accurately and quickly the gender and age structure of internal visitors to the capital, which is a significant advantage compared to the nationwide statistical observation «Household Travel Expenditure Survey Questionnaire» *(index - H-050),* for which information is formed only on an annual basis.

Share of internal visitors to Astana by age groups

percentage

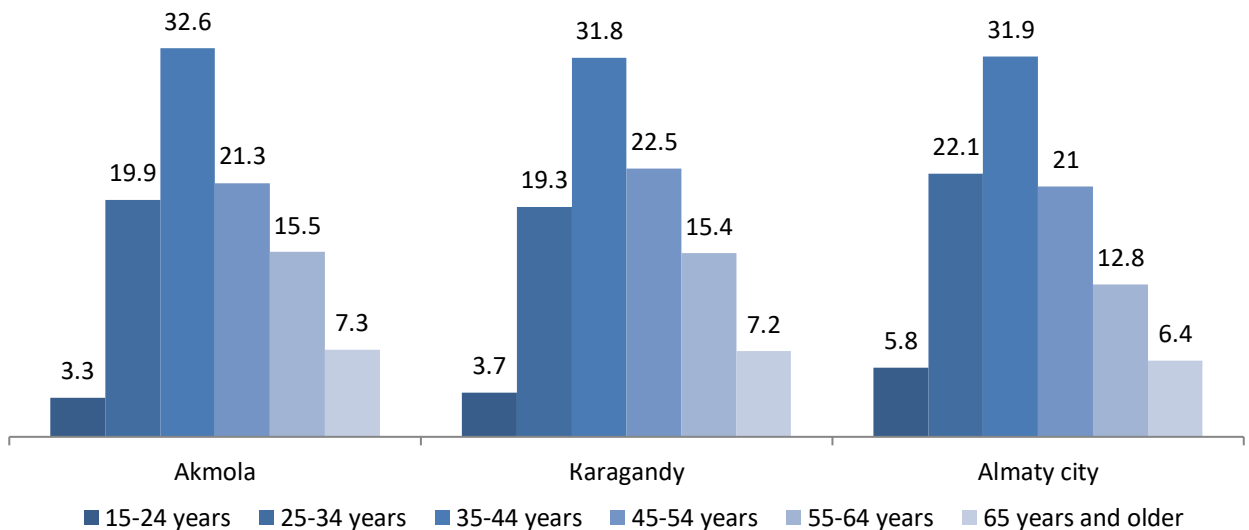| Age group	Percentage
35-44 years	31%
45-54 years	21,5%
25-34 years	20,5%
55-64 years	15,8%
65 years and older	7,4%
15-24 years	3,7%
up to 15 years	0,1%

According to the data of mobile operator KCell, the largest share in the age structure of internal visitors to Astana is occupied by persons aged 35-44 years, which amounted to 31% of the total number of flow.

Following 35-44 year olds in the top 3 age groups are 45-54 year olds and 25-34 year olds, with a slight difference in the share of internal visitors, 21.5% and 20.5% respectively.

Share of internal visitors to Astana by age groups

percentage

	Akmola	Karagandy	Almaty city
15-24 years	3.3	3.7	5.8
25-34 years	19.9	19.3	22.1
35-44 years	32.6	31.8	31.9
45-54 years	21.3	22.5	21
55-64 years	15.5	15.4	12.8
65 years and older	7.3	7.2	6.4

The largest share among the arriving visitors to Astana are persons aged 35-44 years and 45-54 years, which indicates the similarity in tourist preferences and interests of these age groups, as well as the similarity of the nature of tourist activity of donor regions.
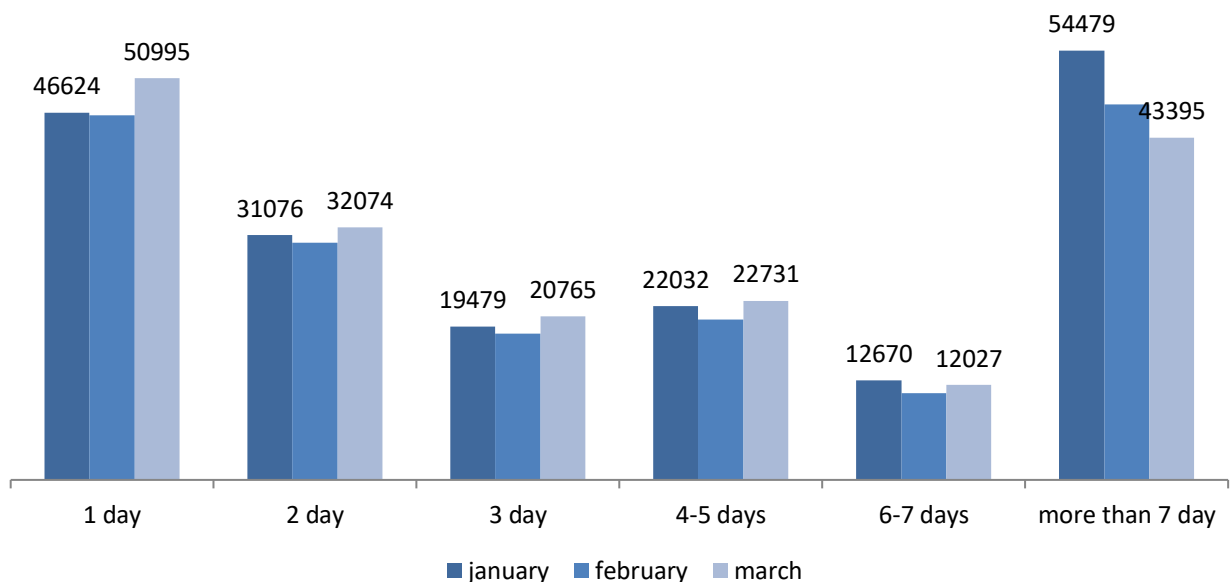
**Share of internal visitors to Astana by
length of stay**

percentage

more than 7 day	26.8
1 day	26.5
2 day	17.2
4-5 days	12
3 day	10.8
6-7 days	6.6
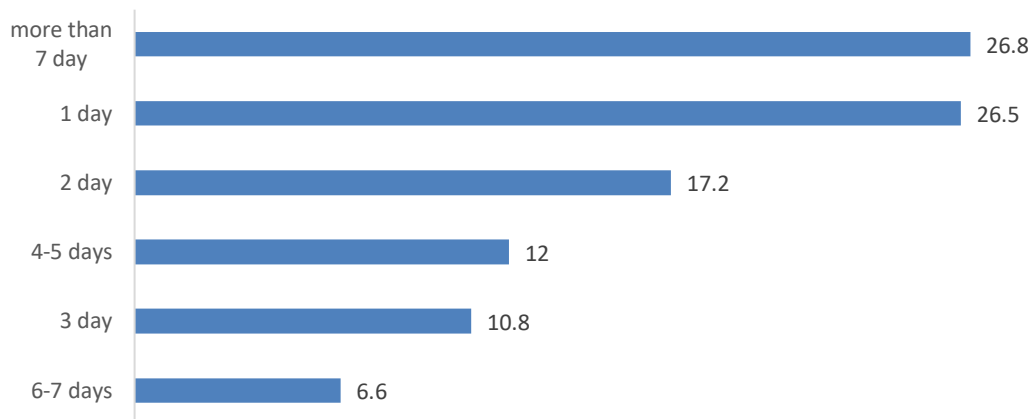
According to the duration of stay in Astana, the largest share is occupied by persons who arrived for more than 7 days, amounting to 26.8% of the total number of internal visitors. With a small gap with a share of 26.5% are those who arrived for 1 day, which is associated with business visits, excursions or one-day events.

**Number of internal visitors to Astana by
length of stay**

person

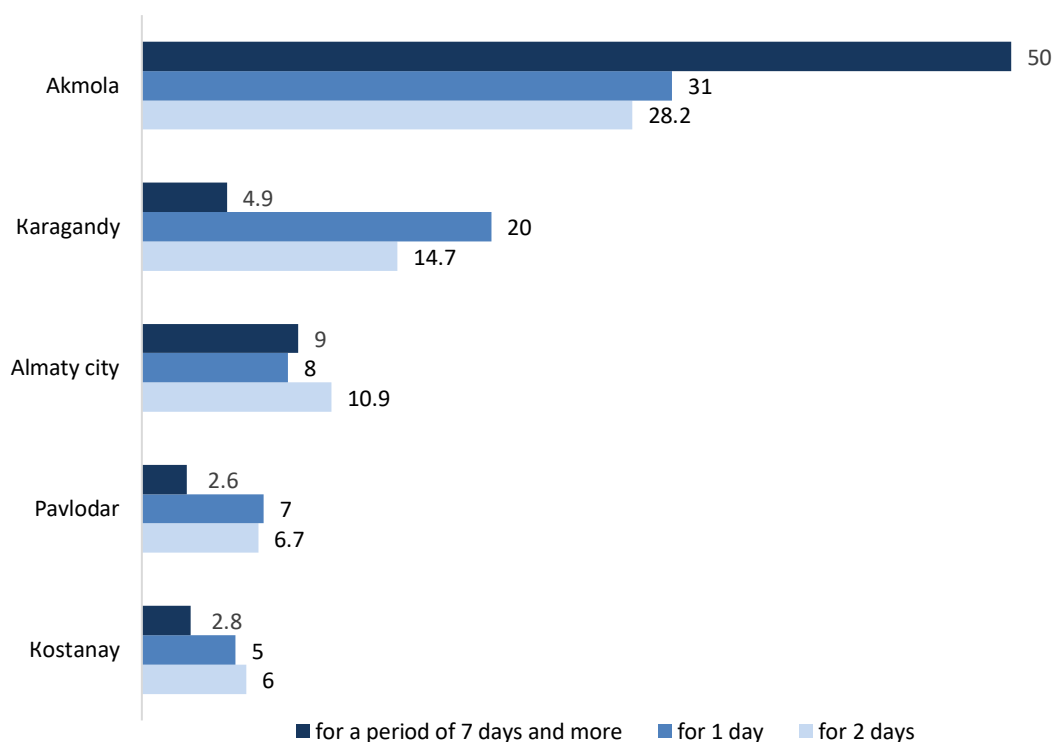|	1 day	2 day	3 day	4-5 days	6-7 days	more than 7 day
january	46624	31076	19479	22032	12670	54479
march	50995	32074	20765	22731	12027	43395

■ january ■ february ■ march

Compared to January in March, there is a gradual increase in the number of internal tourist arrivals, except for those staying in Astana for more than 7 days, whose number decreased from 54.5 thousand people in January to 43.4 thousand people in March 2024.

Analyzing the data on the top 5 regions where visitors arrive from for different periods of stay, we can identify interesting characteristics of visitors depending on the duration of the trip.

Top-5 regions by internal visitors arriving in Astana by the longest duration of stay

percentage

	for a period of 7 days and more	for 1 day	for 2 days
Akmola	50	31	28.2
Karagandy	4.9	20	14.7
Almaty city	9	8	10.9
Pavlodar	2.6	7	6.7
Kostanay	2.8	5	6

The capital for a stay of 7 days or more is most preferred by visitors from Akmola region (50%). The second place among donor regions is occupied by Almaty city (9%), followed by Turkestan region (6%), Karaganda region (4.9%) and Shymkent city (3.6%).

Among those arriving in the capital for one day, Akmola region is also the most popular donor region, accounting for 31% of the total number of those arriving for this period. The second place is occupied by Karaganda region (20%), followed by Almaty city (8%), Pavlodar region (7%) and Kostanay region (5%).

In contrast to one-day trips, the most popular donor regions for visiting the capital for 2 days remain Akmola region (28%) and Karaganda region (15%). The third place with a share of 11% is occupied by Almaty city, followed by Pavlodar (7%) and Kostanay regions (6%).

Gender and age structure of internal visitors to Astana city

According to the data of mobile operator KCell for the first three months of 2024 in the capital of Kazakhstan arrived 324.8 thousand men, or 59.9% and 217.5 thousand women, or 40.1% of the total number of internal visitors. A similar picture is observed in all regions of the country.
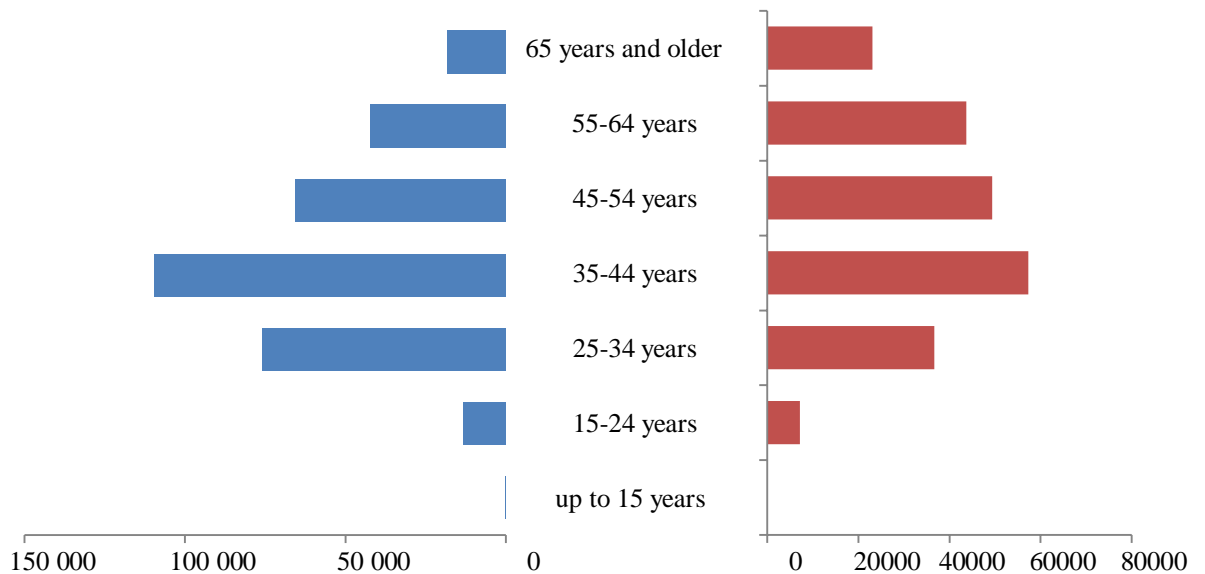
Number of internal visitors to Astana by gender

person
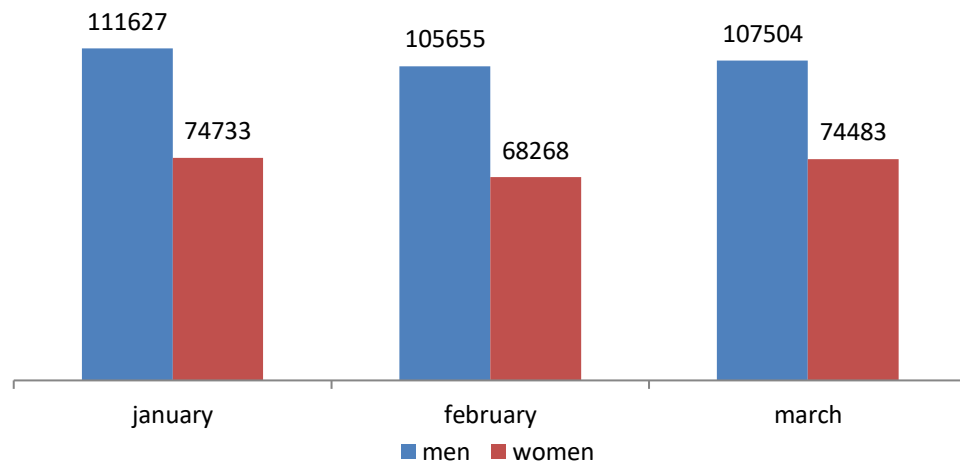
During the first three months of 2024, there are no significant differences in the sex and age structure.

Gender and age structure of internal visitors to Astana city

person

In all age groups up to 55 years of age, the number of men prevails over the number of women. The largest gender gap is observed in the age groups 25-34 years, where the share of men in the total number of the age group amounted to 67.4%, and 35-44 years - 65.6%.

## Duration of stay of internal visitors to Astana by gender

percentage
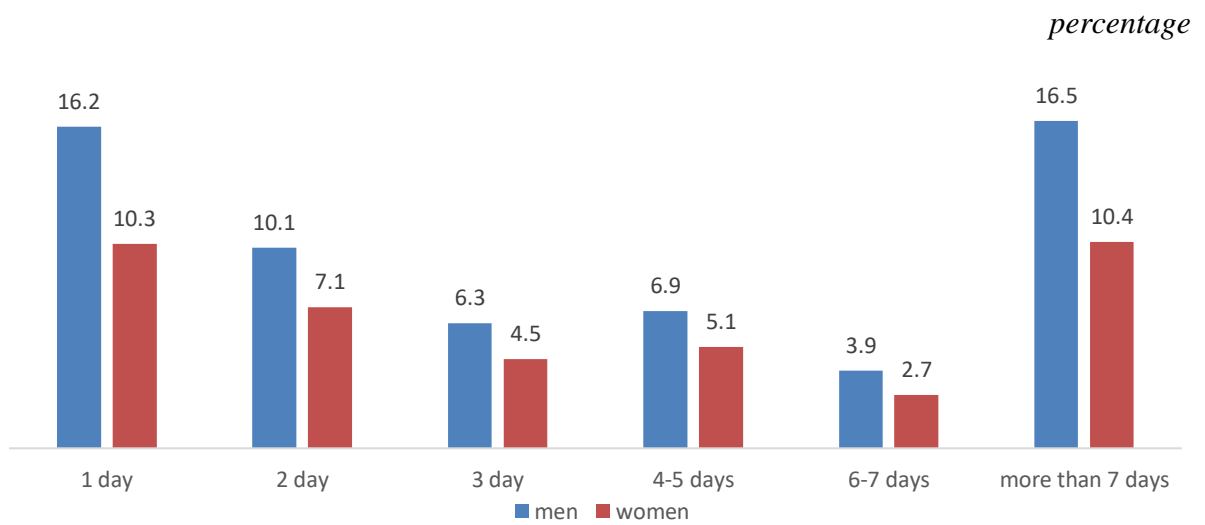
In terms of the length of stay in the capital, the largest proportion were men staying for more than 7 days and 1 day, together accounting for 32.7% of the total number of internal visitors.

Въездные иностранные посетители г.Астана

According to the data of mobile operator KCell for February-March 2024, the capital of Kazakhstan was visited by 113.1 thousand foreign tourists, of which 59.7% fell on March.

In general, for February-March this year Astana was visited by 66.9 thousand people (59.2%) from the Russian Federation, 7.7 thousand people (6.8%) from China, 5.6 thousand people (5%) from Turkey, 3.7 thousand people (3.2%) from Estonia and 3.4 thousand people (3%) from Germany.
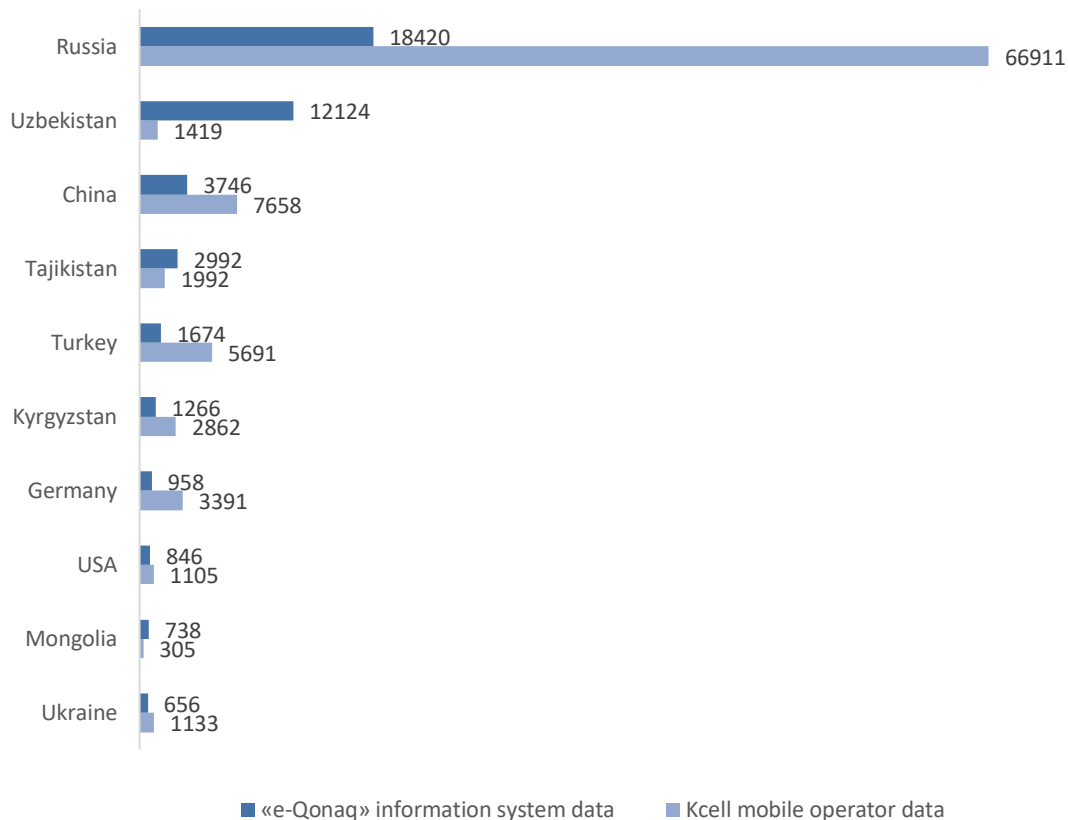
Top 10 countries by the number of arriving visitors to Astana in February-March 2024 according to the mobile operator KCell.

persentage
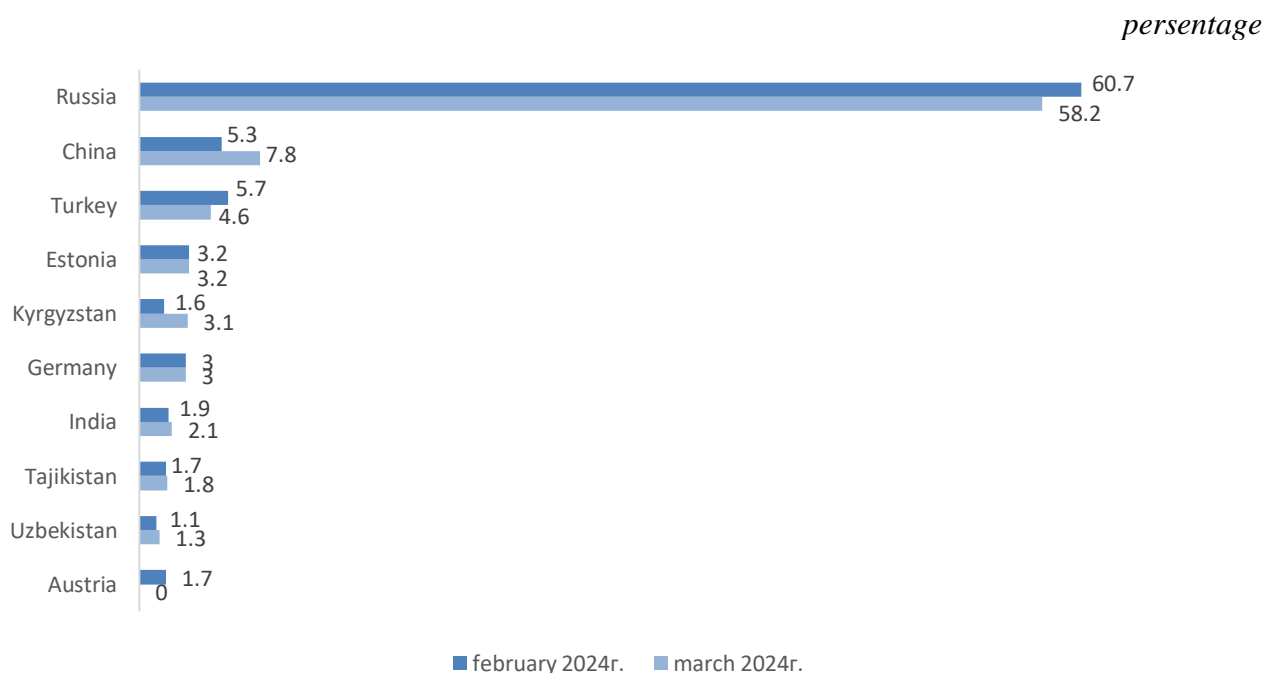
For two months, Russia, China and Turkey remain the constant leaders in the number of foreign tourists arriving in Astana.

**Comparison of the number of arrivals in Astana city
visitors by top-10 countries of IS «e-Qonaq» with data of mobile
operator KCell for January-March 2024.**

person

Country	«e-Qonaq» information system data	Kcell mobile operator data
Russia	18420	66911
Uzbekistan	12124	1419
China	3746	7658
Tajikistan	2992	1992
Turkey	1674	5691
Kyrgyzstan	1266	2862
Germany	958	3391
USA	846	1105
Mongolia	738	305
Ukraine	656	1133

■ «e-Qonaq» information system data   ■ Kcell mobile operator data

IS «eQonaq» is an electronic system of accounting and registration of tourists, which was introduced to simplify the process of registration of tourists' stay in the country and increase the transparency of data on tourist flow. The number of foreign visitors to the capital city according to KCell mobile operator data for some countries exceeds the number of foreign tourists according to eQonaq IS by times. This difference is due to the fact that the range of data collected by mobile operators is more extensive and diverse, as these data include information on calls, SMS, mobile Internet sessions, geolocation and service usage. These data can cover a wider range of activities and provide more detailed information on visitor behavior, while in eQonaq IS, tourist counting and registration is based on data from accommodation facilities (hotels, inns, hostels, etc.).

5. Conclusions

Mobile positioning data is a valuable source of information about the gender and age structure of visitors, their portrait, geography of accommodation, preferences in the choice of tourist destinations. Analyzing these data sets allows for more accurate identification of popular tourist routes and destinations, as well as forecasting demand for tourism sector services.

Mobile operators have access to a huge amount of data about their subscribers, including information about every call, message and internet activity. This enables a wide range of data, including network, infrastructure, and user devices, providing a more comprehensive view of user behavior and activity.

Overall, mobile positioning data from cellular operators represents an effective and promising tool for analyzing and managing tourist flow and developing the tourism industry.

The comparative analysis showed that the structure and trends of KCell data are similar to official statistics and to e-Qonaq IS data. This indicates the reliability and validity of mobile positioning data obtained from mobile operators.

Thus, it is recommended to expand the list of data sources and actively use alternative sources to improve the accuracy of analysis and prompt decision-making.

The diversity of information sources will help to improve the analysis of tourists' needs, identify new trends and adapt strategies in real time, which, in turn, will allow for more effective planning of tourism policies and informed decision-making that contributes to the development of the tourism industry in the country.