Алматы қаласы
Цифрландыру
Басқармасы

Almaty
Data
Lake

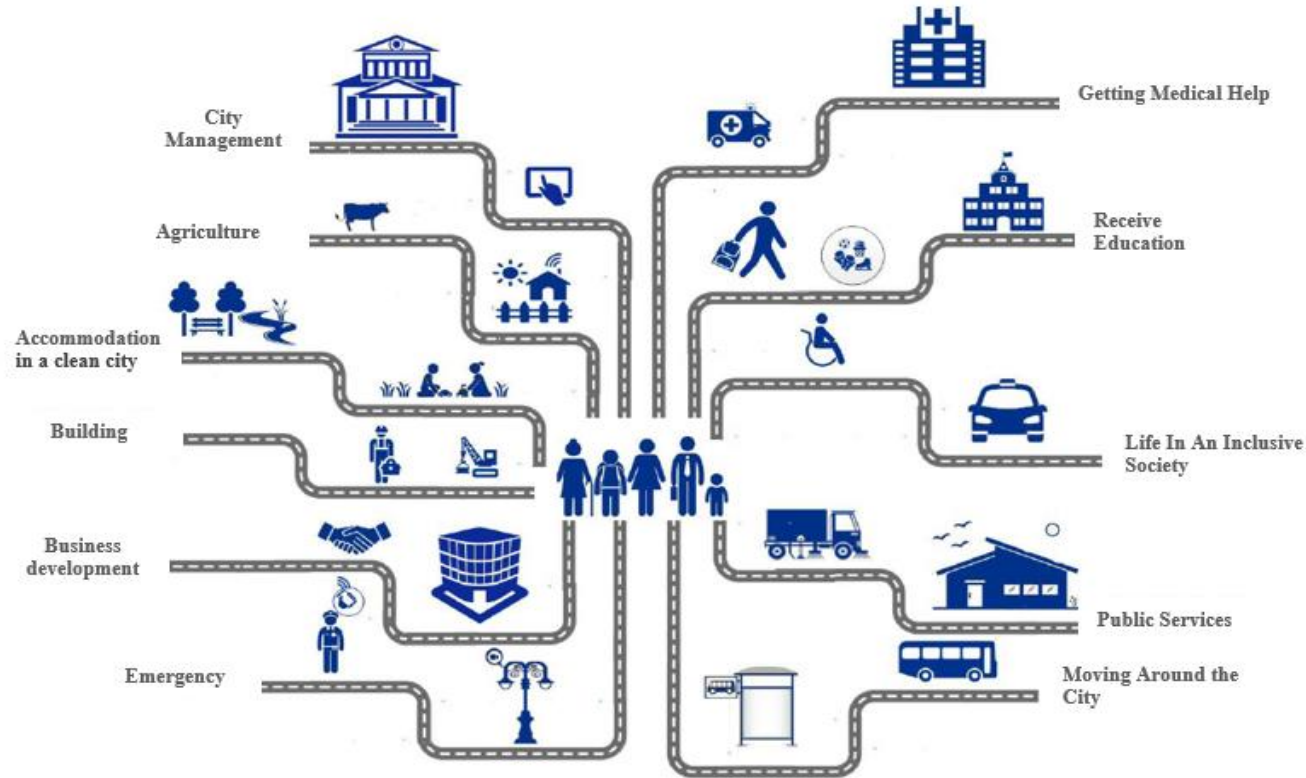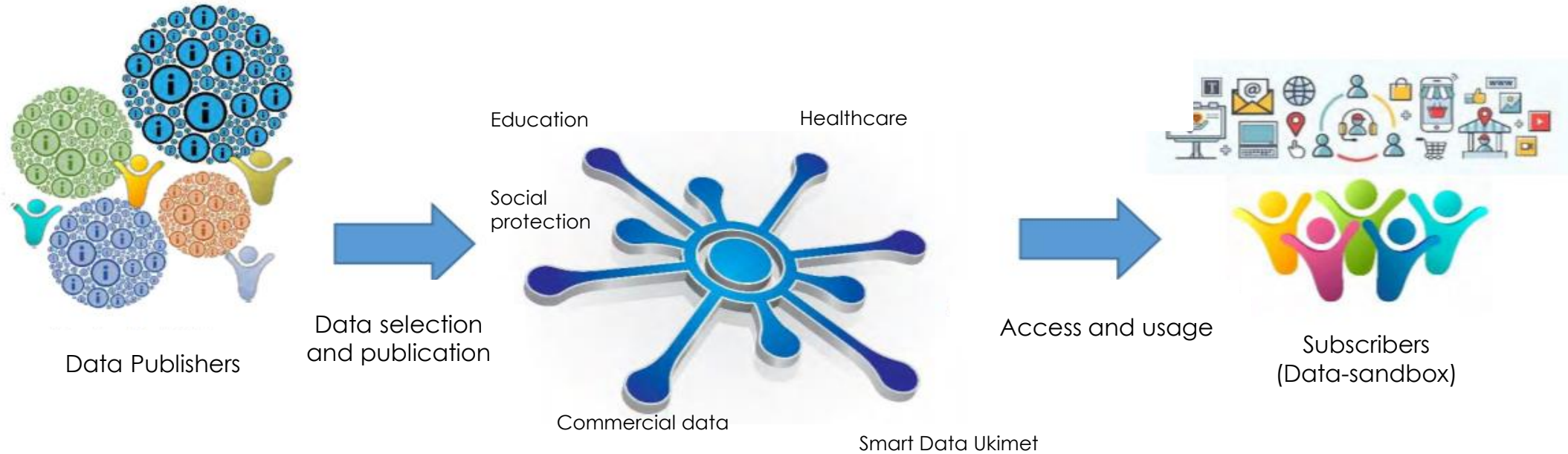# Almaty Data Lake: strategy, principles, achievements and future plans

Agenda:

1. Almay Data Lake overview
2. The Singapore Story
3. Almaty Data Lake development strategy
4. 2022 : analytical cases, development of pro-active services
5. Plans for 2023-2025

Data as an asset to deliver on Smart Almaty strategic goals

- Empower senior officials to make **data-driven decisions**

- Surface city data for business ecosystem to develop a vibrant **digital business/economy**

- Make **moving around the city** easy and eco-friendly

- Ensure **physical security** of citizens, businesses and assets

- Make Almaty an **ecologically friendly city**

Almaty Data Lake

Education

Healthcare

Social protection

Data selection and publication

Data Publishers

Commercial data

Smart Data Ukimet

Access and usage

Subscribers (Data-sandbox)

What has changed:

✓ A project on data integration from one public and one commercial databases enabled utility providers to start filling the long-term gap of lost revenue (**around 5 mln USD per year**);

✓ **Social cluster** established: Education, Healthcare, Social protection and utilities data;

✓ In place technical opportunities for deployment of **value-added proactive services** for Almaty residents.

Future plans:

1) Set up of **Economic cluster** in 2022-2023 and **Land Management** cluster in 2024-2025;

2) **Launch** of proactive services**;**

3) **Development of data-sandbox** – most requested data for city development;

4) Creation of data-based **value-added IT-products** to foster innovation and digital transformation.

Almaty city with 'Almaty Data Lake' project entered finals of the international competition IEEE Smart Cities Awards Contest 2022

An opportunity to tell and show Almaty Data Lake at 2022 IEEE International Smart Cities Conference, which took place 26-29 September 2022 in Paphos, Cyprus. The conference brings together practitioners, urban policymakers and administrators, infrastructure operators, industry representatives and researchers to present technologies and applications, share their experiences and opinions on current and future smart city applications.

https://attend.ieee.org/isc2-2022/sc-awards-contest-results-2

IEEE Smart Cities recognizes the
City of Almaty, Kazakhstan
as one of the finalists of the
2022 IEEE Smart Cities Award
with the project
'Almaty Data Lake'
(Category of Emerging Economy)

Gustavo Giannattasio
IEEE Smart Cities
Contest Chair

Georges Zissis
IEEE Smart Cities
Chair

Wei-Jen Lee
IEEE Smart Cities
Education Cmte. Chair

IEEE Smart Cities Award 2022 — Finalist

* IEEE is an international non-profit association of technical professionals, a world leader in the development of standards for radioelectronics, electrical engineering, and computer system and networking hardware

3

The Singapore experience in building 'digital government' :

- **'Smart Nation'** programme was launched by Prime-Minister Lee Hsien Loong in **November 2014**.

- **Vision**: Improve people's life and create new, ICT-based opportunities. All deployed programmes and projects should address day-to-day citizens' issues.

- **Data have value when they interlink with one logic and cover all areas of human's life**.

- Data are **transparent and accessible** (data.gov.sg, LTA DataMall, SingStat и OneMap).

- A core data model is based on **citizens' requests for services**. Data cleaning and catalogue structuring are performed **upon data model maturity and growing interest to data usage**.

- Working with data should bring **real benefits**.

- Set up of **Data Exchange Platform** as a **Single Source of Truth** to increase trust level from business and society.

- **Data as a Service**

The Singapore story

KEY MILESTONES FOR STRATEGIC NATIONAL PROJECTS

National Digital Identity (NDI)
- SingPass Mobile Token MyInfo extended to more Govt and Private Sector services
- NDI platform operational Digital signatures for paperless transactions
- Widespread adoption of NDI

E-Payments
- PayNow
- Common Payment QR Standard
- E-payments in Hawker Centres
- Account-Based Ticketing for Public Transport
- Rollout of 25,000 Unified Point-of-Sales terminals

Smart Nation Sensor Platform (SNSP)
- Wireless Sensor Network deployment in Orchard Road and selected areas
- Smart Connected Lamp Post trial
- City-level sensor data available for industry and public

Smart Urban Mobility
- More timely bus arrivals with common public bus fleet management system
- Trial on-demand bus services
- Deploy AVs to enhance public transport (Earliest)

Moments of Life (MOL)
- MOL app for families with young children
- MOL approach extended to more life moments
- Expanded suite of services for families with young children

2017 2018 2019 2020 2021 2022

- **National Digital Identity** (NDI)provides comfortable and secure digital transactions between citizens and enterprises, with online identification
- **Smart Nation Sensor Platform**(SNSP) allows smart connection of IoT sensors;
- **Smart Urban Mobility** deploys AI for better possibilities of public transport, via data analytics and autonomic transport devices;
- **Moments of Life** (MOL) unites government services and information via single platform.
- **MyInfo** — services for SingPass, enables users to avoid multi-time filling-in of personal data for every digital transaction.

My Info: Personal Data Platform

Screencap from MyInfo website

❖ Shared by Government for Trusted Commercial Partners.

❖ Speeds up the Economic Transaction, e.g. Opening Bank Account, Opening Trading Account, etc.

❖ Data from Trusted Source (the Government).

❖ Created greater convenience for its residents.

❖ Extend of information shared by MyInfo covers Contact, Education, Salary, etc.

Singapore: progressive built of data hubs

- ❖ MyInfo built on Data Sharing initiative started in the 80s.
- ❖ Similar to Almaty, started with the first data hub focusing on people - PeopleInfo.
- ❖ Then a 2nd data hub is implemented, targeting Business.
- ❖ Now internal to government there are 7 data hubs covering:
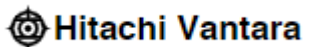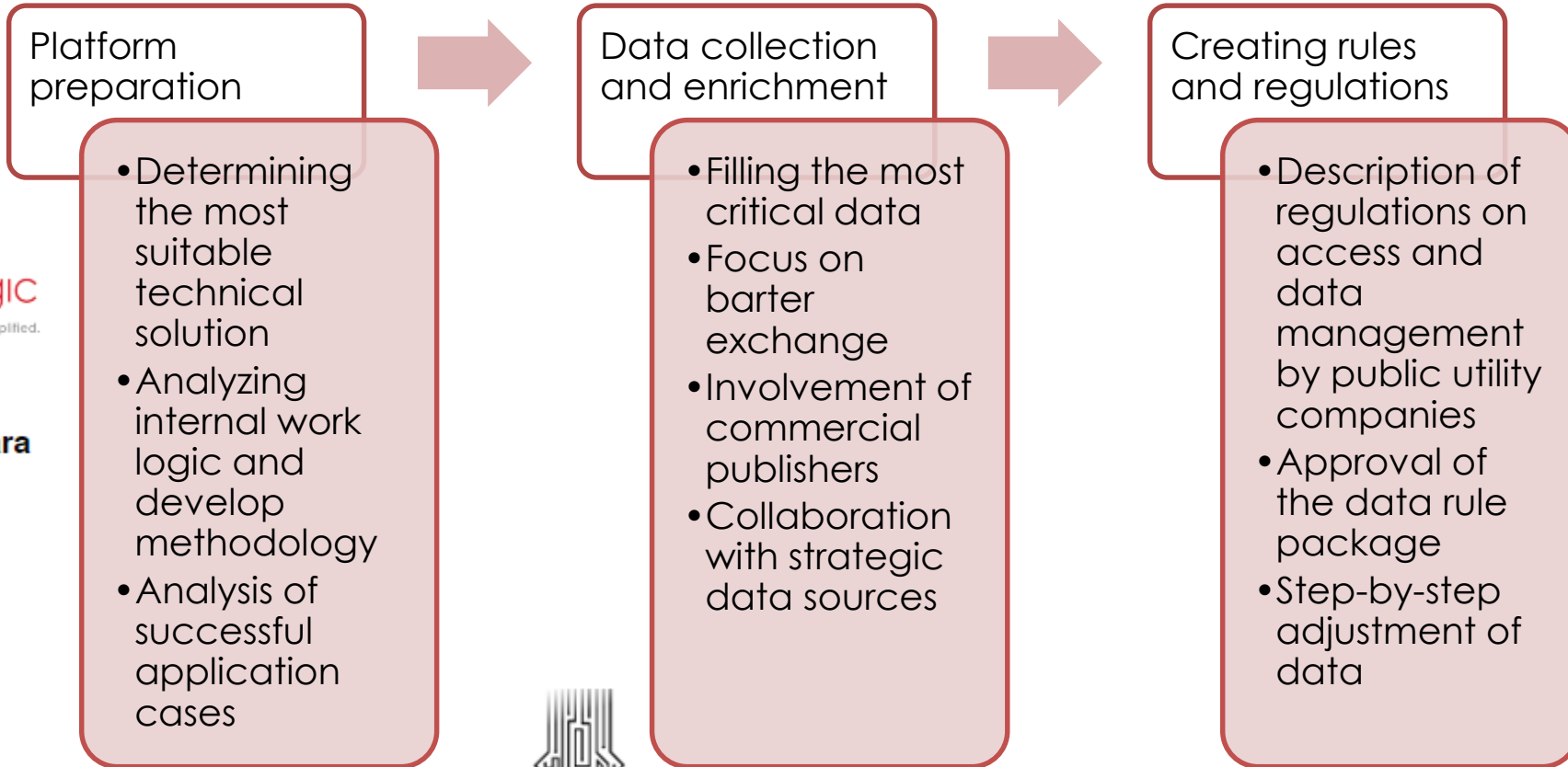 Education, Healthcare, Social Development, Land, Vehicle etc.

CrimsonLogic and Smart Government

03   SUSTAIN – PROACTIVE COMMUNITY ENGAGEMENTS

02   EXECUTE – PROVIDE CORE APPLICATIONS

01   PLAN – PEOPLE CENTERRED DATA HUBS FOR DATA DRIVEN SMART NATION

We are part of the Singapore Digitalisation Journey.
We build systems that focus on bringing Citizen and Economical values.

Development Strategy of Almaty Data Lake:

Platform preparation

- Determining the most suitable technical solution
- Analyzing internal work logic and develop methodology
- Analysis of successful application cases

Data collection and enrichment

- Filling the most critical data
- Focus on barter exchange
- Involvement of commercial publishers
- Collaboration with strategic data sources

Creating rules and regulations

- Description of regulations on access and data management by public utility companies
- Approval of the data rule package
- Step-by-step adjustment of data

Regulations signed 07/2022

Rules are ready, under approval 10/2022

NBK
Data Factory

Бюро национальной статистики
Агентства по стратегическому
планированию и реформам
Республики Казахстан

Progressive built of data hubs

Data integration from Education, Healthcare, Social Protection – people's migration, youth employment, etc

Analysis of household economic activities and city enterprises

Social cluster 2021-2023

Economic cluster 2022-2024

Citizen 360°

Data integration of Almaty city and oblast – development of Almaty agglomeration

Land Management 2024-2026

Logistics cluster 2023-2025

Analysis of in-city produced goods, its movement, circular economy

Data governance principles

1. **Dream Big Start Small –** specific cases and indicators (for each case: goal, objectives, expected results, KPIs.

2. **Focus on cases that are beneficial to ALL participants** (the main focus - socio-economic development of the city. Akimat acts as a partner, without the use of pressure levers).

3. **Refinement of business and operational processes** (bringing to unified standards across the city, signing of regulations and agreements, synchronization with meta-data national directories, adaptation of different technical standards)

4. **Data cleaning before onboarding to the exchange platform** (publishers responsibility for data completeness, purity, legitimacy, ETL on the publisher side, automatic discarding of inappropriate data).
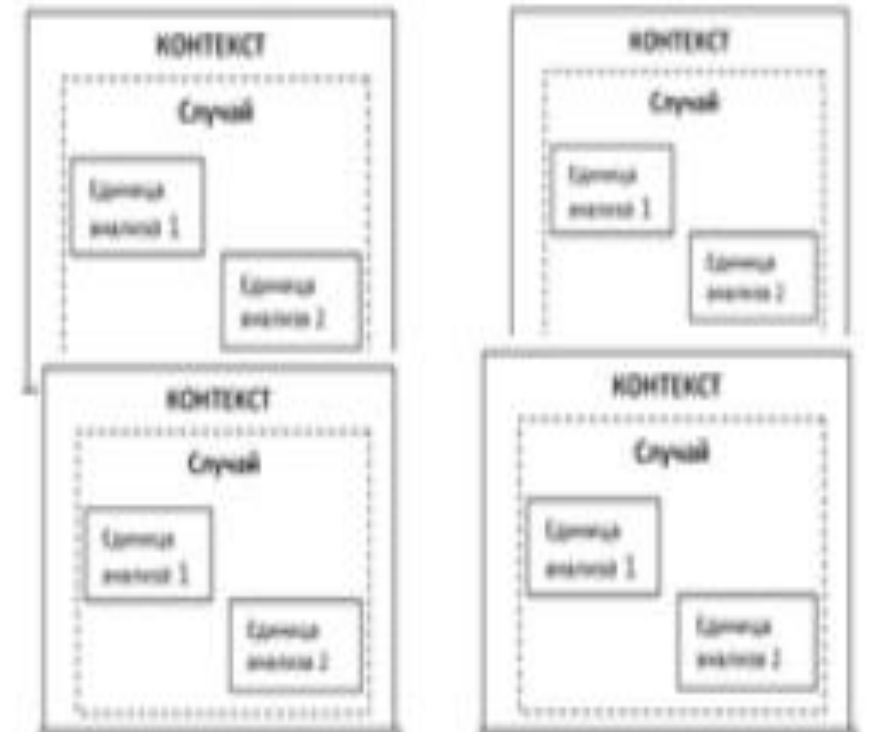
Data governance principles (2)

5. **Less data is better** (first discuss the necessary tables and attributes, then test the data samples, only then build a harmonized cluster data model).

6. **Data clustering** (Citizen 360° Social Cluster, Economy Cluster , Logistics Cluster, Land Management Cluster).

7. **Integration and cross-enrichment of data** (only practical application of integrated data shows completeness and purity of data in different IS).

8. **Creation of value-added services** (example - automatic calculation on DXP the number of socially vulnerable people at a particular address).

Data analytics in ADL is based on the principles of **case study**, which consists of several steps**:**

1. Identification of the **most indicative cases** in each of the key areas;

2. Determination of case/problem solving performance indicators (**KPIs**);

3. **Verification of case results** using different units of analysis;

4. **Sharing the results of different cases** for an in-depth and comprehensive analysis of social and economic phenomena.

A multiple case study approach with different units of analysis (Yin, 2014).

1) Digitalize a process to obtain a preferential public transport card «Onay» for students and schoolchildren of Almaty city (***coverage over 570 thousand***) – jointly with Transport Holding of Almaty city.

2) Automate calculation of citizens from a specific social group at a particular address, for potential reimbursement of communal utilities payments (***coverage over 305 thousand***) – together with Alseco and Almaty Su.

Analytical case on availability of Almaty medical organizations (primary healthcare, PHC)

Алматы қаласы
Цифрландыру
Басқармасы

Almaty
Data
Lake

Objectives:

✓ Comparative analysis of several databases on healthcare in Almaty city

✓ Consider districts and medical organizations with the largest number of attached population.

✓ To identify "abnormal" cases, i.e. short-term (less than 1 year), attachments to medical institutions institutions according to MZVitrina (Smart Data Ukimet).

✓ Compare data on the attached population with registrations in the city over 14 years old.

✓ Analyze the availability of current medical organizations for residents of Almaty by cartographic method.

✓ Use cluster analysis for predictive analytics in order to identify the locations of new necessary medical facilities.

Data source:

• Population attached to PHC (RPN) from Smart Data Ukimet (ИС «Регистр прикреплённого населения» МЗ РК);
• Almaty People Database (АИС «Население»);
• MZVitrina (Smart Data Ukimet);
• Damumed data (private company);
• Almaty city Healthcare department(УОЗ);
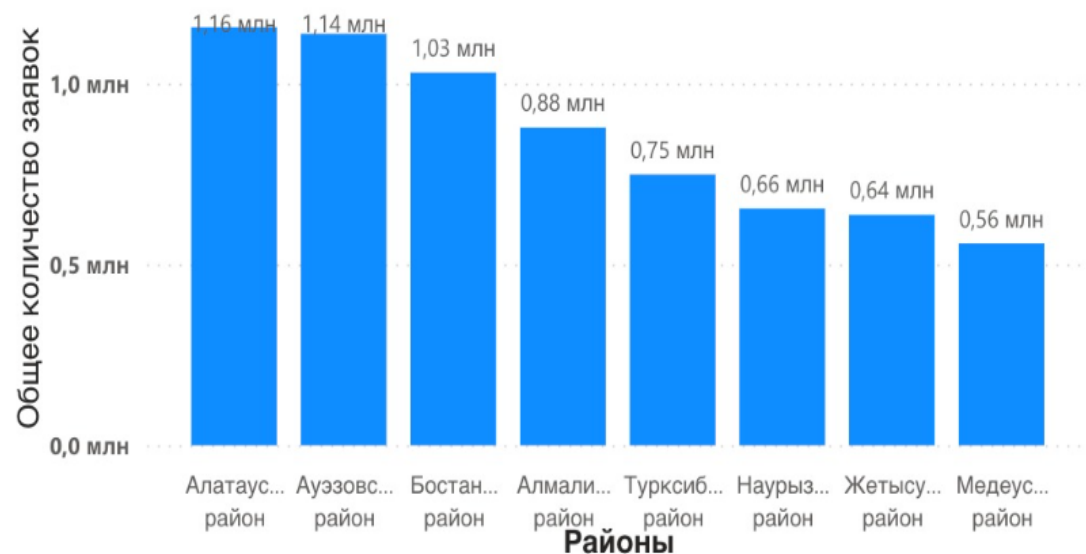• NextGIS (roads);

Legal sources:

• Orders of Ministry of Healthcare Kazakhstan

Comparison of actual attachments (RPN) vs applications (SDU)

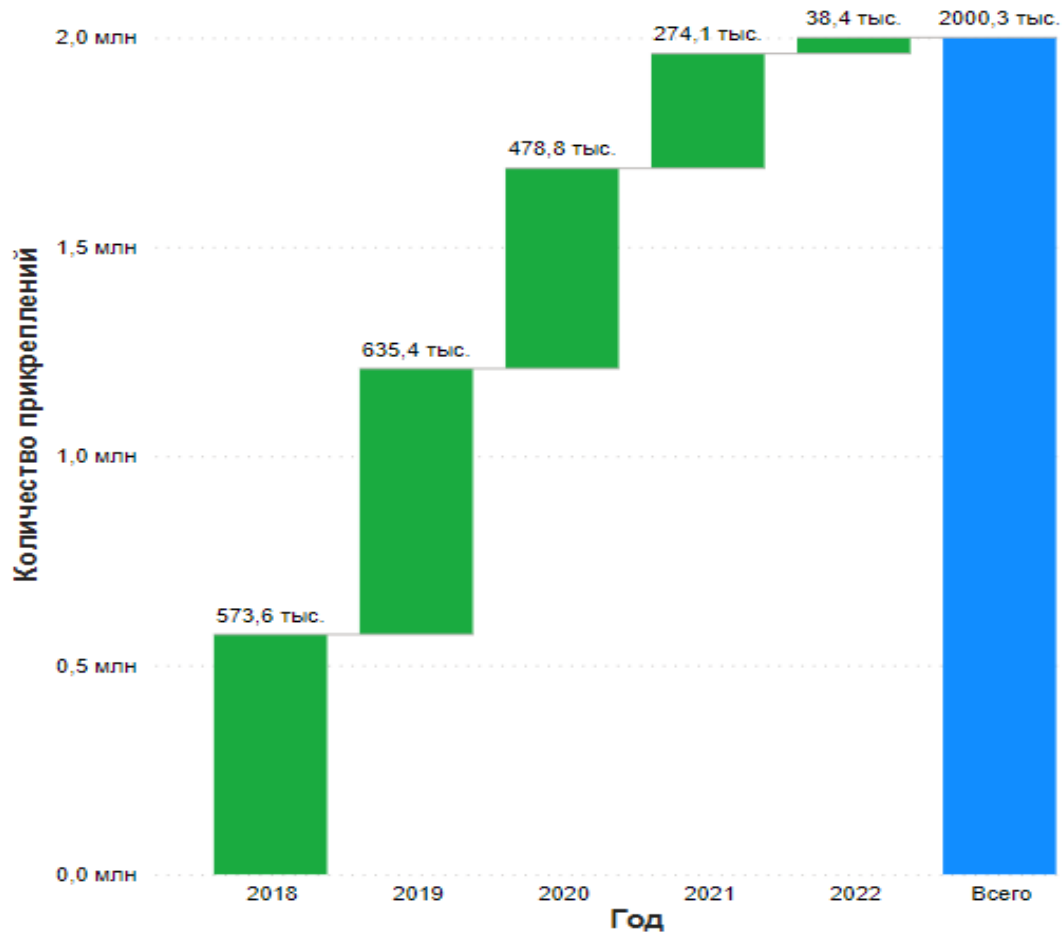RPN 2018-2022 г.
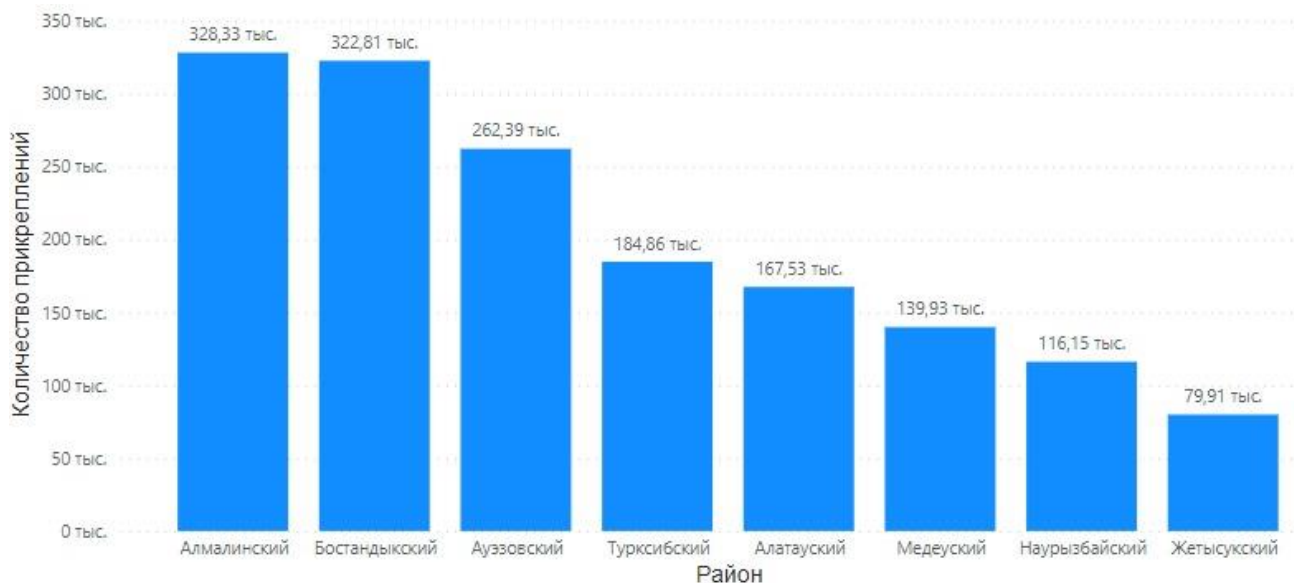
MZVItrina 2018-2022 г.

H: High number of applications is overload for reception offices

Source: MZVitrina, RPN

Dynamics of short-term attachments to medical organizations (MZvitrina 01.01.2018-09.06.2022)

Количество прикреплений менее одного года

In 2018-2022 there were **3,2 mln attachments** registered in MZVitrina.

Out of them, short-term attachments (less than 1 year) made **2 mln records, two thirds of total records.**

H: High number of short-term attachments imposes the over-load for receptions and doctors

Source: MZVitrina

Data on attachment of over-14 year-population, by administrative districts (RPN vs AlmatyPeopleDatabase)

The attachment of a person to PHC organizations is the basis for the provision of PHC and is carried out in accordance with the Order of the Minister of Health of the Republic of Kazakhstan dated November 13, 2020 No. KR DSM-194/2020:
1) Territorial accessibility of PHC (20-minute walking distance within the administrative district of the city);Free choice of the organization of PHC;
2) Free choice of a specialist (therapist, pediatrician, GP).

Районы прикреплений в разрезе районов мест жительства
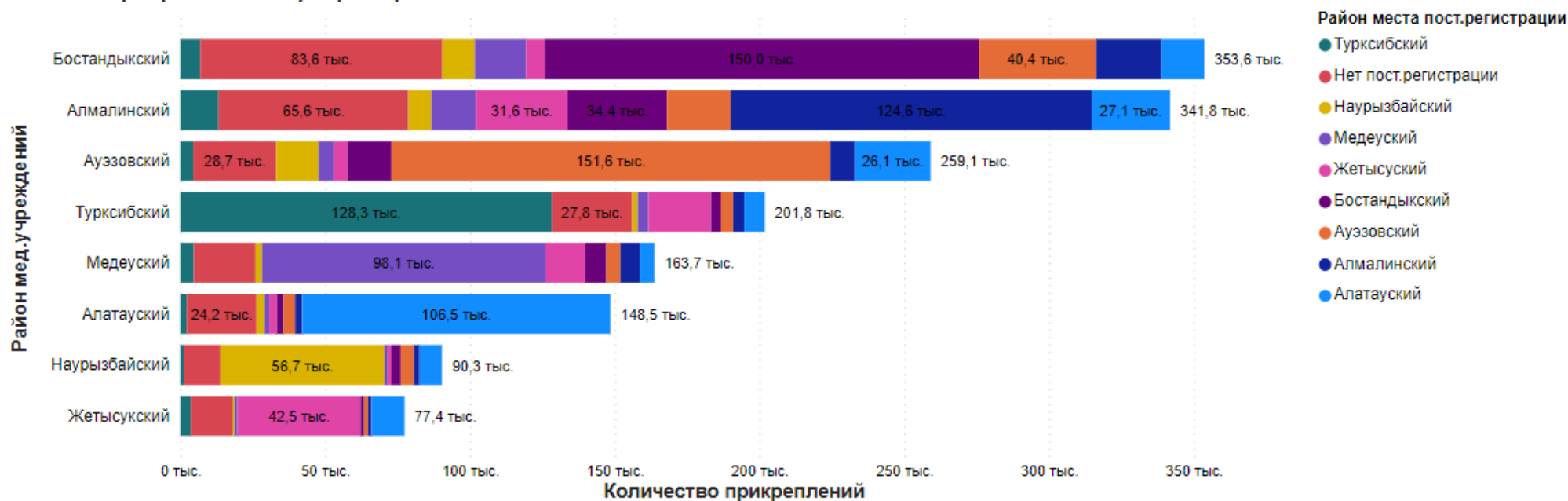
The method comprises the following main steps:

1. Calculation of the parameters of a standard polyclinic, namely its load: in this case, the limits of 5,000 and 30,000 people per PHC organization are taken(limits described in the Rules for PHC centers in the city).
2. Application of clustering to the population grid (Scenario 1)
3. Visualization of walking distance in 20 minutes from the cluster mass centers
4. Application of PHC institutions and allocation of cells not covered by 20-minute pedestrian accessibility of these objects
5. Clustering of uncovered cells (Scenario 2)
6. Visualization of walking distance in 20 minutes from real and simulated objects

Scenario 1: Modeling the "ideal" distribution of PHC organizations in the city without taking into account the location of existing PHC objects.

Scenario 2: Simulation of missing PHC organizations, taking into account the location of existing PHC objects
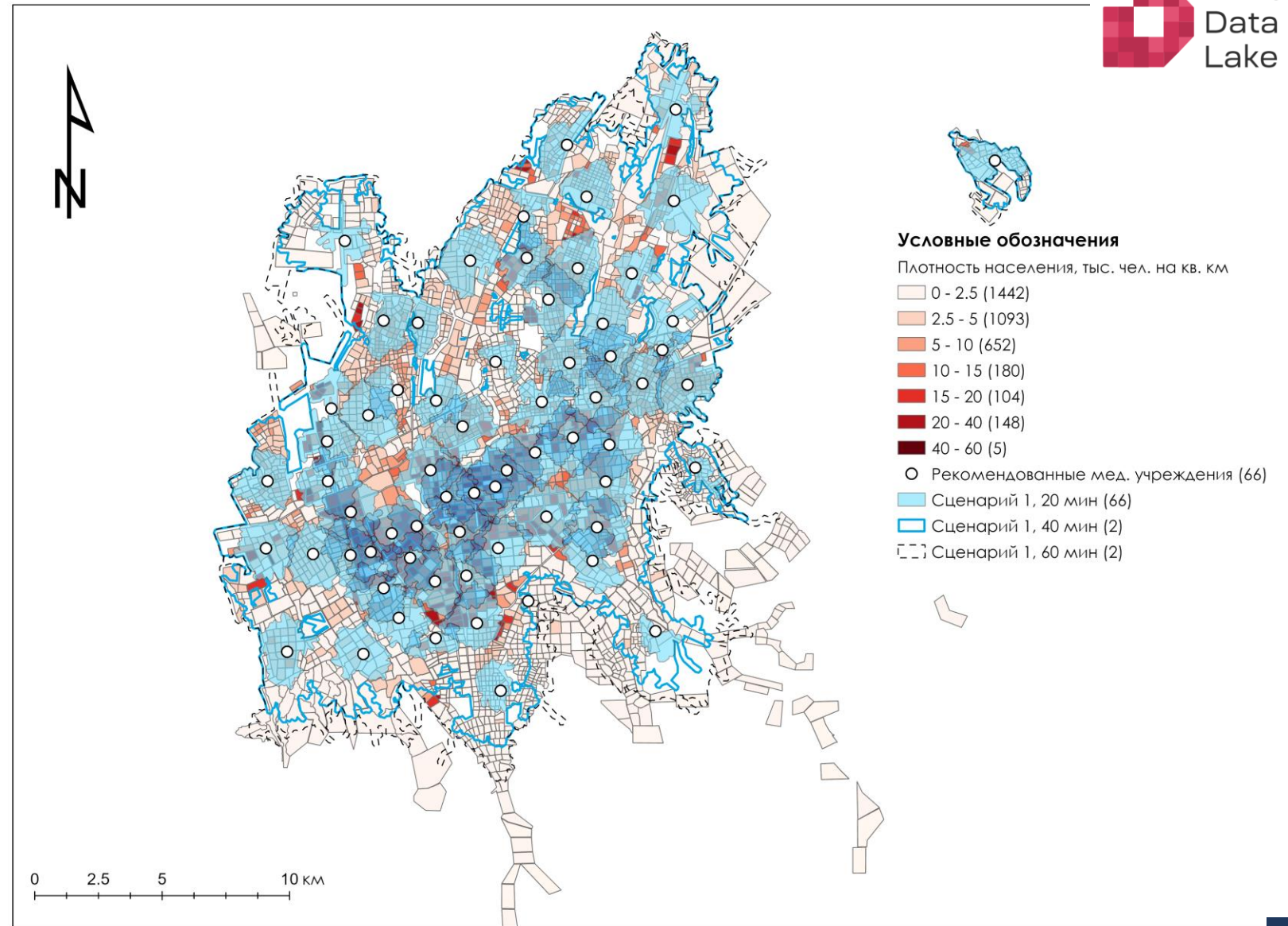
20

Scenario 1

Spatially limited multidimensional clustering with lower and upper limits of 5 and 30 thousand people per 1 object was applied to the population grid, as a result, 66 objects were built in the centers of cluster masses.

It is noticeable that in the central part of the city the objects are located denser due to the high population density.

On the map – the population density, location and coverage of pedestrian accessibility in 20 minutes of the proposed PHC objects.

Условные обозначения

Плотность населения, тыс. чел. на кв. км
- 0 - 2.5 (1442)
- 2.5 - 5 (1093)
- 5 - 10 (652)
- 10 - 15 (180)
- 15 - 20 (104)
- 20 - 40 (148)
- 40 - 60 (5)
- ○ Рекомендованные мед. учреждения (66)
- Сценарий 1, 20 мин (66)
- Сценарий 1, 40 мин (2)
- Сценарий 1, 60 мин (2)

0   2.5   5   10 км

Source: AlmatyPeopleDatabase, Healthcare Dpt

21

Алматы каласы
Цифрландыру
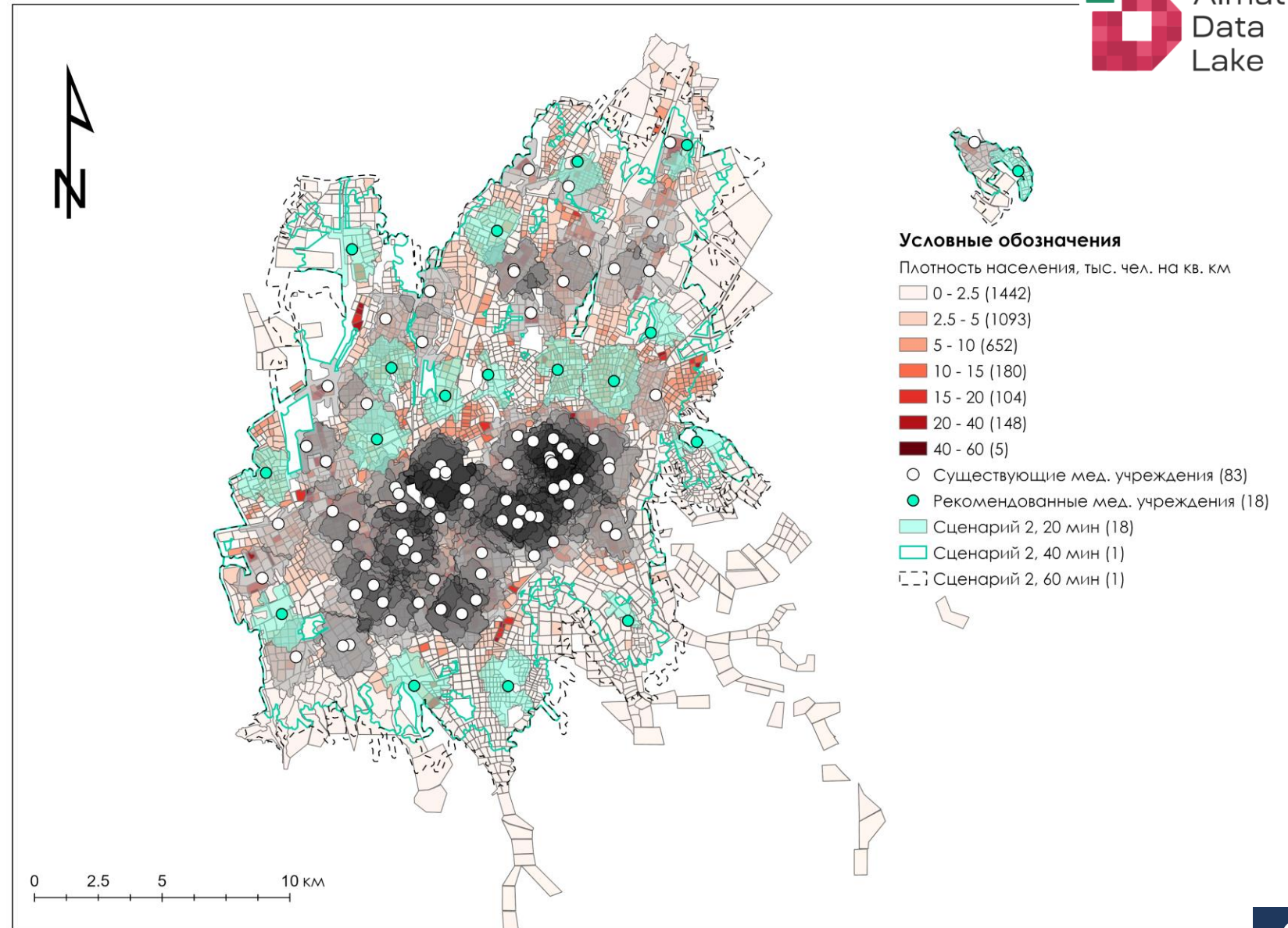Баскармасы

Almaty
Data
Lake

Scenario 2

83 existing PHC organizations were added to the map.

Spatially limited multidimensional clustering with lower and upper limits of 5 and 30 thousand was applied to the population grid outside the pedestrian accessibility (20 min) of PHC organizations. As a result, **18 objects were built** in the centers of mass of clusters.

On the map – population density and comparison of coverage of current and suggested PHC organizations.

Условные обозначения

Плотность населения, тыс. чел. на кв. км
- 0 - 2.5 (1442)
- 2.5 - 5 (1093)
- 5 - 10 (652)
- 10 - 15 (180)
- 15 - 20 (104)
- 20 - 40 (148)
- 40 - 60 (5)
- ○ Существующие мед. учреждения (83)
- ● Рекомендованные мед. учреждения (18)
- Сценарий 2, 20 мин (18)
- Сценарий 2, 40 мин (1)
- Сценарий 2, 60 мин (1)

0   2.5   5   10 км

22

Source: AlmatyPeopleDatabase, Healthcare Dpt

Analysis of labour migration of college and university graduates of Almaty city (2020-2021)

Objectives:

- ✓ Spatial analysis of student distribution of within the city districts.
- ✓ Comparative analysis of methods of financing graduate education.
- ✓ Estimating demand for graduates in the labor market of Almaty.
- ✓ In-depth analysis of employment in IT market of Almaty.
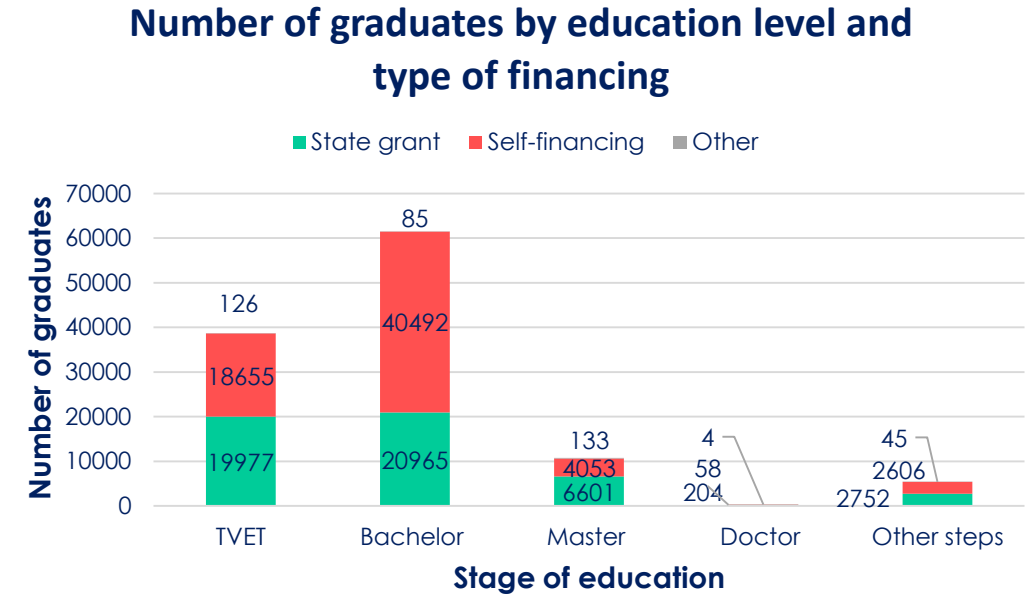
Data source:

- National Educational Database (NEDB), 7.06.2022 and 25.12.2022 (from Smart Data Ukimet)
- AIS "AlmatyPeopleDatabase", 01.12.2022
- AIS "Pension payments and benefits (E-market)", 30.06.2021 - 30.06.2022 (from Smart Data Ukimet)
- "Register of the attached population" of the Ministry of Health of the Republic of Kazakhstan, 25.05.22 and 7.12.2022 (from Smart Data Ukimet)
- Data from the Bureau of National Statistics of the Agency for Strategic Planning and Reforms of the Republic of Kazakhstan (stat.gov.kz)
- Report of the Labor Resources Center of the Ministry of Labor and Social Protection of the Population (the CENTER of the Ministry of Labor and Social Protection of the Population)
- Smart Data Ukimet data on Suleyman Demirel University, 16.08.2022

Classification of Almaty college and university graduates by type of financing of education for 2020-2021

Stage of education	Number of graduates	Number of graduates with state scholarship	Share of graduates with state scholarship out of all graduates
Secondary vocational education	38 605*	19 977	51,75%
Bachelor	63 894	20 965	32,81%
Master	11 134	6 601	59,29%
Doctor	288	204	70,83%
Master of Business Administration	211	0	0
Doctor of Business Administration	10	0	0%
Residency	1 420	1125	79,23%
Internship	3 338	1616	48,41%
Pedagogical retraining	24	1	4,17%
Preparatory department for foreign citizens	13	10	76,92%
Total	118 937	50 499	42,38%

"Other steps" include:
- *Doctor of Business Administration*
- *Pedagogical retraining*
- *Internship*
- *Residency*
- *Preparatory Department for foreign citizens*

* The total number of graduates does not match due to repetitions and the number of students in NEDB

Number of graduates by education level and type of financing

"State grants" include:
- Grant from the Ministry of Education and Science of the Republic of Kazakhstan
- Grant from other ministries of the Republic of Kazakhstan
- Republican budget and local budget
- Scholarships from local authorities
- Scholarships from district mayors

"Other" includes:
- Scholarship under interstate agreements
- Director's scholarship
- Enterprise grant
- Grant from the receiving or sending university
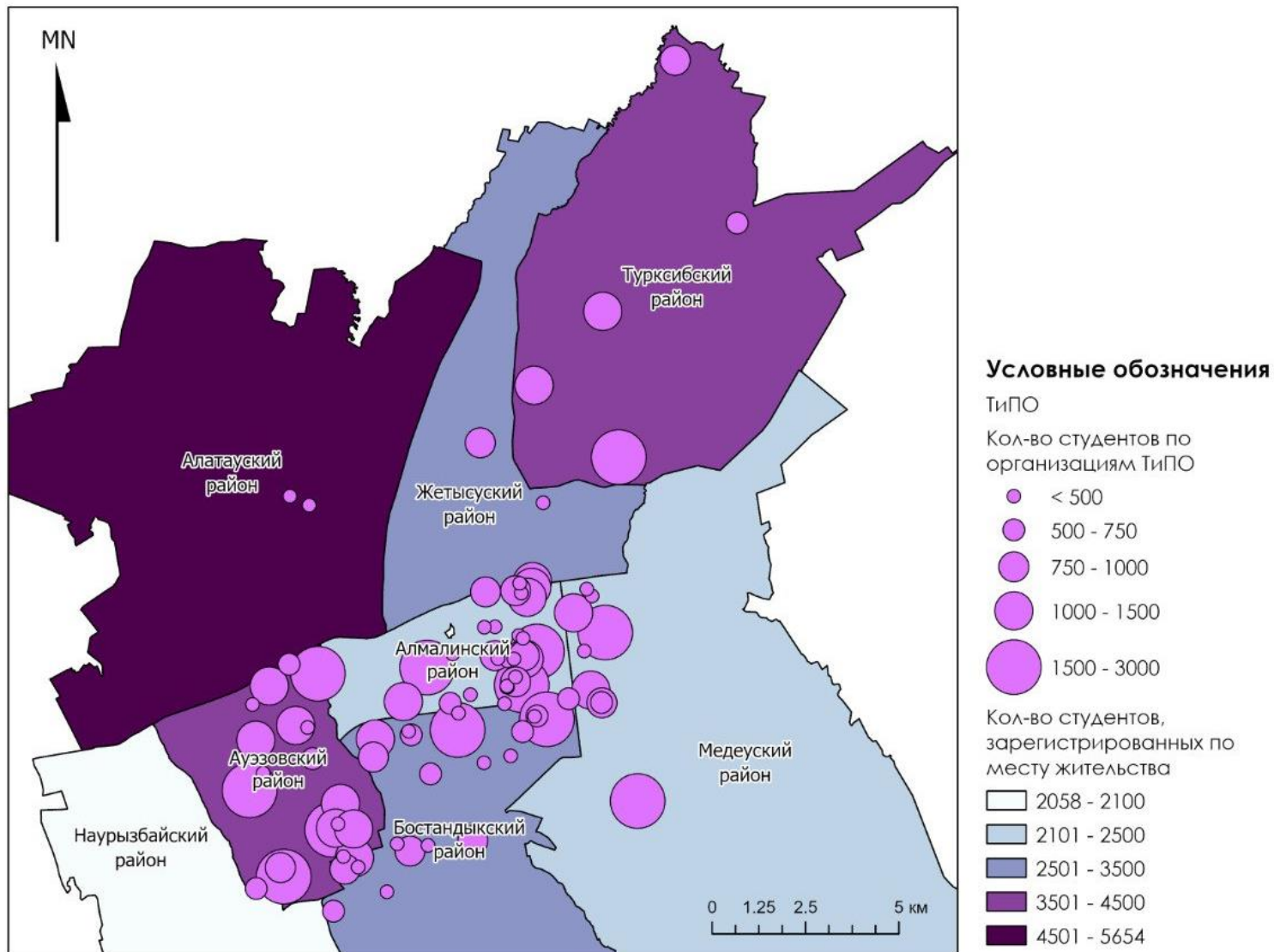- State educational loan

"Self-financing" includes:
- Student funds

The level of education of 2 273 graduates is not specified,
The type of funding for 10 518 graduates is not specified

Source: NEDB

24

# Distribution of active college students by registration districts

The map shows 91 colleges and the boundaries of 8 districts of Almaty city.

- Most colleges are located in Almaly (29), the northern part of Bostandyk (24), and Auezov (17) districts.

- However, most of college students registered at the place of residents live in the Alatau district (21.4% or 5654 students out of 26,460), which has only 2 institutions out of 91.

- **The most balanced district in terms of the number of students registered at the place of residence and studying in colleges is the Auezov district.**

H: Opening colleges in Nauryzbay and Alatau districts in line with the districts' main economic activities.

Source: NEBD, AlmatyPeopleDatabase
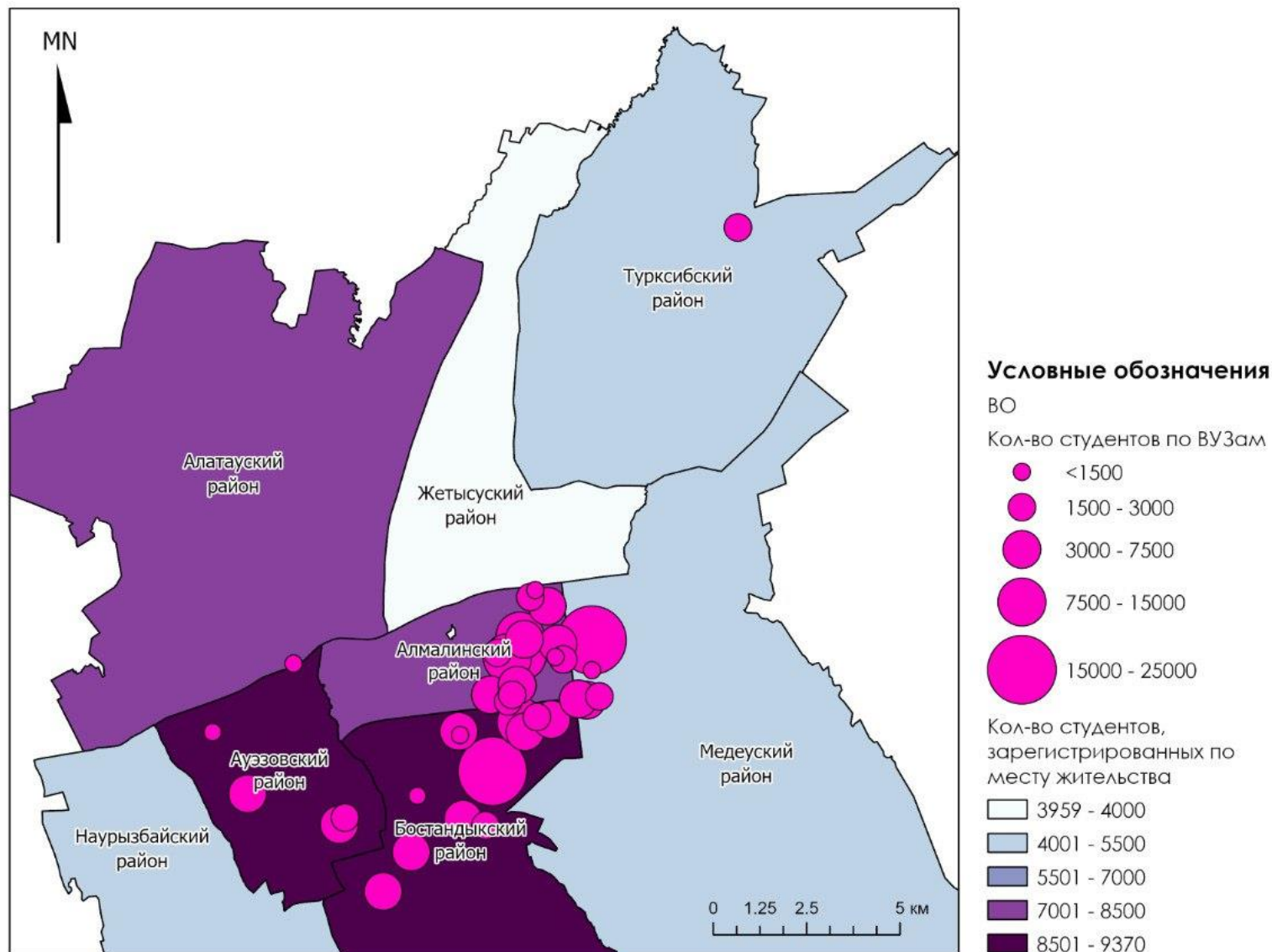
The map shows 40 universities and the borders of 8 districts of Almaty city.

- Most HE institutions are located in Almaly (16) and the northern part of Bostandyk (13) districts.

- There is a greater imbalance in the distribution of universities within the city compared to colleges.

- Many university students registered at the place of residents live in Alatau (15.9% or 8459), where there is only 1 institution out of a total of 40, and Auezov districts (17.2% or 9163 students out of 53,266), which houses 4 universities.

- **The most balanced district in terms of the number of registered and studying students in university is the Bostandyk district (17.6% of registrations and 13 universities).**

H: Opening colleges in accordance with the profile of the universities and main economic specialization of the districts

Source: NEBD, AlmatyPeopleDatabase

26

WorldSkills Kazakhstan

The mission of the international "WorldSkills" organization is to improve the quality of technical and vocational education and develop the skills of qualified specialists to achieve economic growth through certification and assessment of students' abilities.

WorldSkills goals in Kazakhstan:

1. Identification of the best world practices in the professional training of young people and their introduction into the TVET system of Kazakhstan;
2. Stimulation of innovative processes and introduction of new learning technologies in the TVET system of Kazakhstan;
3. Development of the expert community in Kazakhstan through regular methodological assistance to improve curricula;
4. Raising the status of a working person and the prestige of technical professions.

At the moment, only 5 out of 89 colleges in Almaty have passed the assessment of student skills from WorldSkills Kazakhstan:

- "Almaty College of Polygraphy" (publishing)
- "College of the tourism and hospitality industry" (assistant administrator, waiter)
- "Almaty Automotive College" (automotive mechanic)
- "Almaty Construction and Technical College" (master builder of a wide profile)
- "Almaty College of Passenger Transport and Technologies" (hairdresser-fashion designer)

Considering that there are usually at least 5 specializations in each college, the proportion of TVET specializations in Almaty that pass the skills assessment is extremely low.

An increase in this share can lead to an increase not only in the quality of education, but also in the prestige of college graduates among employers.

Source: worldskills.kz

Thank you!