Central Asian Scientific Journal

VOL 4(32)
2026

ASTANA

ISSN: 3135-3061

Электронный научный журнал «Central Asian Scientific Journal»

Central Asian Scientific Journal

выпуск №4 (32), октябрь 2026 г.
Том 1
Основан в 2021 году (издается ежеквартально)

зарегистрирован в Комитете информации Министерства информации и общественного развития Республики Казахстан №KZ77VPY00147053 от 17.04.2026 г.

Taqyryptyq baǵyt:

- Pedagogıkalyq, qoǵamdyq-áleýmettik, tehnıkalyq, ekonomıkalyq jáne zań ǵylymdary
- Aqparattyq-komýnıkasıalyq tehnologıalar
- Teorıalyq jáne ǵylymı-praktıkalyq ǵylymı zertteýler

Тематическая направленность:

- Педагогические, общественно-социальные, технические, экономические и юридические науки
- Информационно-коммуникационные технологии
- Теоретические и научно-практические научные исследования

Thematic focus:

- Pedagogical, socio-political, technical, economic, and legal sciences
- Information and communication technologies
- Theoretical and scientific-practical research

Jarıalanatyn aqparattyń, dáıeksózderdiń jáne ózge de baıandamalardyń durystyǵy úshin avtor jaýapty bolady

За достоверность публикуемой информации, цитат и иных изложений ответственность несет автор

The author is responsible for the accuracy of the published information, quotes, and other statements.

ISSN: 3135-3061

"Central Asian Scientific Journal" elektrondyq ǵylymı jýrnaly aqparattyq agenttigi

№4 (32), 2026 j
Shyǵarý jıiligi –jylyna 4 nómir
2021 j. bastap shyǵady

Qazaqstan Respýblıkasy Aqparat jáne qoǵamdyq damý mınıstrliginiń 17.04.2026 j. №KZ77VPY00147053 aqparat komıtetinde tirkelgen.

JK DOC, 010000, Qazaqstan Respýblıkasy, Astana q.

Информационное агентство Электронный научный журнал «Central Asian Scientific Journal»

№4 (32), 2026 г.
Периодичность – 4 номера в год
Выходит с 2021 года

Зарегистрирован в Комитете информации Министерства информации и общественного развития Республики Казахстан №KZ77VPY00147053 от 17.04.2026

ИП DOC, 010000, Республика Казахстан, г. Астана

Information Agency Electronic scientific Journal "Central Asian Scientific Journal"

No.4 (32), 2026
Periodicity: 4 issues per year
Since 2021

Registered with the Information Committee of the Ministry of Information and Public Development of the Republic of Kazakhstan No. KZ77VPY00147053 dated 17.04.2026.

IP DOC, 010000, Kazakhstan, Astana

МАЗМҰНЫ – СОДЕРЖАНИЕ – CONTENT

ӘЛЕУМЕТТІК-ЭКОНОМИКАЛЫҚ ЖӘНЕ ҚҰҚЫҚ ҒЫЛЫМДАРЫ - СОЦИАЛЬНО-ЭКОНОМИЧЕСКИЕ И ЮРИДИЧЕСКИЕ НАУКИ - SOCIO- ECONOMIC AND LEGAL SCIENCES

Төлегенов Тимур Сейфуллаулы

ДИНАМИКА ТУРИСТСКИХ ПОТОКОВ И ЭФФЕКТИВНОСТЬ ИНФРАСТРУКТУРЫ ТУРКЕСТАНА: ОТ ЭКСТЕНСИВНОГО РОСТА К СМАРТ-УПРАВЛЕНИЮ 4

ПӘНДЕР АРАЛЫҚ ҒЫЛЫМДАР – МЕЖДИСЦИПЛИНАРНЫЕ НАУКИ – INTERDISCIPLINARY SCIENCES

Albiyev Asser

HOW MUCH OF AN ARTICLE COUNT IS HUMAN WORK? DECOMPOSING THE CREATION OF FIVE WIKIPEDIAS, 2010–2025..... 9

Amangeldy Nurbolat

HOW MUCH OF KAZAKHSTAN'S URBAN BUILT STOCK PREDATES ITS SEISMIC CODES, AND WHERE DOES THAT STOCK SIT RELATIVE TO HAZARD? 19

ӘЛЕУМЕТТІК-ЭКОНОМИКАЛЫҚ ЖӘНЕ ҚҰҚЫҚ ҒЫЛЫМДАРЫ - СОЦИАЛЬНО-ЭКОНОМИЧЕСКИЕ И ЮРИДИЧЕСКИЕ НАУКИ - SOCIO-ECONOMIC AND LEGAL SCIENCES

УДК 338.487

Төлегенов Тимур Сейфуллаулы
Магистрант 2 курса
Образовательная программа «Менеджмент в туризме»
НАО «Международный университет туризма и гостеприимства»
(г. Туркестан, Казахстан)

ДИНАМИКА ТУРИСТСКИХ ПОТОКОВ И ЭФФЕКТИВНОСТЬ ИНФРАСТРУКТУРЫ ТУРКЕСТАНА: ОТ ЭКСТЕНСИВНОГО РОСТА К СМАРТ-УПРАВЛЕНИЮ

Аннотация: В статье исследуется динамика и структурные изменения туристских потоков в быстрорастущей культурно-познавательной дестинации на примере города Туркестан. На основе данных, официально предоставленных руководством многофункционального туристского комплекса «Karavansaray», проведен сравнительный макроанализ за 2024–2025 годы. Исследование базируется на оценке относительных показателей — средней загрузки средств размещения (Occupancy) и географической структуры прибытий, что позволило соблюсти принципы коммерческой конфиденциальности. Выявлено увеличение доли иностранных посетителей и устойчивый рост загрузки инфраструктурного ядра (с 38,9% до 47,7%). Обоснована необходимость внедрения дестинационных управляющих организаций (DMO) и технологий смарт-туризма для предотвращения овертуризма.

Ключевые слова: туристская дестинация, туристический поток, инфраструктура гостеприимства, Туркестан, комплекс Karavansaray, управление дестинацией (DMO), смарт-туризм, загрузка номерного фонда.

Развитие индустрии туризма в Республике Казахстан на современном этапе характеризуется уверенным переходом от формирования базовой инфраструктуры к необходимости комплексного и интеллектуального управления туристскими территориями. Особое, стратегически важное место в этом процессе занимает город Туркестан — признанная духовная и историческая столица страны. Завершение масштабных инфраструктурных проектов, включая строительство международного аэропорта, обновление городских транспортных артерий и, в первую очередь, запуск многофункционального туристского комплекса «Karavansaray», обеспечило городу статус одной из наиболее динамично развивающихся дестинаций всей Центральной Азии. Однако столь резкий рост популярности направления закономерно порождает классическую проблему территориального менеджмента: возникает дисбаланс между постоянно растущим туристским потоком и реальной пропускной способностью отдельных культурно-исторических объектов (в частности, мавзолея Ходжи Ахмеда Ясави и исторического центра).

В современной научной литературе эффективность туристской дестинации традиционно оценивается через призму комплексного взаимодействия ее ключевых

элементов. Согласно классической концепции интегрированного туристского продукта Д. Бухалиса, любая дестинация представляет собой сложную социально-экономическую систему, состоящую из достопримечательностей, транспортной доступности и средств размещения [1]. В свою очередь, Р. Батлер в своей фундаментальной теории жизненного цикла туристской территории (Tourism Area Life Cycle, TALC) утверждает, что неконтролируемый, стихийный рост туристических прибытий неизбежно приводит дестинацию к стадии стагнации и последующего упадка, если управляющими органами своевременно не внедряются эффективные механизмы пространственного и административного регулирования потоков [2]. Данные теоретические подходы сегодня дополняются актуальными трендами, задаваемыми Всемирной туристской организацией (UNWTO), которые категорично подчеркивают необходимость перехода территорий к модели «смарт-дестинаций», где управление потоками базируется на анализе больших данных (Big Data) и глубокой координации всех участников рынка через профильные организации по управлению дестинацией (DMO) [3, 4].

Целью данного исследования является сравнительный анализ структурных изменений туристского потока и эффективности загрузки ключевой туристской инфраструктуры Туркестана за 2024–2025 годы, а также разработка практических рекомендаций по оптимизации управления данной дестинацией в условиях растущего спроса.

Для достижения поставленной цели нами был применен комплекс научных методов: дескриптивный анализ (на основе агрегированных данных о туристских прибытиях), метод сравнительного статистического анализа (сопоставление операционных показателей 2024 и 2025 годов) и метод концептуального моделирования процессов управления. В качестве основной эмпирической базы исследования выступили статистические отчеты, официально предоставленные для проведения научного анализа администрацией многофункционального туристского комплекса «Karavansaray» (в том числе данные отеля Karavansaray Hotels Turkistan). Важно подчеркнуть, что с целью строгого соблюдения коммерческой тайны предприятия и принципов академической этики, в исследовании полностью исключены абсолютные финансовые показатели выручки (Total Revenue) и точные объемы реализованного номерного фонда. Весь анализ базируется исключительно на относительных метриках: процентной операционной загрузке (Occupancy) и процентном соотношении географических рынков-источников.

Анализ географической структуры туристского потока позволил выявить существенные сдвиги в рыночном позиционировании Туркестана. В 2024 году поток носил ярко выраженный внутренний, локальный характер. На долю граждан Республики Казахстан приходилось абсолютное большинство прибытий, тогда как международный сегмент был представлен преимущественно транзитными пассажирами или узкими целевыми паломническими группами из Турции, Узбекистана и Китая. Однако данные за 2025 год демонстрируют качественную трансформацию дестинации и значительный, статистически подтвержденный рост доли въездного туризма.

Согласно проанализированной статистике, предоставленной комплексом «Karavansaray», в 2025 году Туркестан значительно усилил свои позиции на международной арене туризма. Доля туристов из Турции выросла практически вдвое (составив более 11% от общего объема ключевых рынков), а доля китайских посетителей, проявляющих растущий интерес к историческому наследию Шелкового пути,

увеличилась до 4,9%. Также наблюдается устойчивый и планомерный спрос со стороны стран Центральной Азии (в первую очередь Узбекистана и Кыргызстана). Подобная интернационализация туристического потока свидетельствует о том, что Туркестан плавно переходит из статуса исключительно локальной паломнической святыни в категорию полноценного международного культурно-познавательного хаба.

Этот геоэкономический переход накладывает совершенно новые, повышенные требования на инфраструктуру гостеприимства всего региона. Рост числа иностранных посетителей требует немедленной адаптации городской сервисной среды: повсеместного внедрения многоязычной навигации, повышения международных стандартов обслуживания в сегменте общественного питания (Food & Beverage), а также диверсификации экскурсионных продуктов, которые должны быть понятны и интересны не только внутреннему туристу, обладающему базовым историческим контекстом, но и гостям из дальнего зарубежья.

Помимо глубоких структурных изменений потока, важнейшим макроэкономическим индикатором эффективности дестинации выступает степень операционной утилизации (загрузки) ее номерного фонда. Сравнительный анализ относительных показателей продемонстрировал уверенную и стабильную позитивную динамику. Если в 2024 году средняя годовая операционная загрузка (Occupancy) качественных средств размещения на примере анализируемого комплекса составляла 38,93%, то по итогам 2025 года этот показатель вырос до 47,71% (прирост составил существенные 8,78 процентных пункта). Данный факт эмпирически подтверждает, что построенная в городе масштабная инфраструктура (в первую очередь комплекс Karavansaray) не простаивает, а активно интегрируется в туристское потребление и успешно абсорбирует растущий рыночный спрос.

Тем не менее, детальный помесячный анализ выявляет наличие сильно выраженной фактора сезонности. Пиковые нагрузки традиционно и ожидаемо приходятся на весенние месяцы (март-апрель) в связи с празднованием государственного праздника Наурыз и наиболее благоприятными климатическими условиями региона, а также на осенний период (сентябрь-октябрь). Именно в эти периоды возникает серьезный риск кратковременного локального овертуризма, который на практике сопровождается образованием длительных очередей у ключевых достопримечательностей, пиковой нагрузкой на транспортную и санитарно-гигиеническую логистику, а также временным снижением общего качества предоставляемого сервиса.

Экстенсивный путь развития (подразумевающий рост исключительно за счет наращивания бюджетов на маркетинг и непрерывного строительства новых отелей) на данном этапе своего жизненного цикла исчерпывает себя. Дестинации Туркестана требуется качественный скачок и переход к интенсивному смарт-управлению. На основе проведенного теоретико-эмпирического анализа мы формулируем следующие научно-практические рекомендации, направленные на повышение общей эффективности туристской дестинации:

1. Институционализация координации через создание единого DMO. Туркестану критически необходима профессиональная локальная организация по управлению дестинацией (Destination Management Organization), которая объединит разрозненные усилия местного акимата, дирекции государственного историко-культурного заповедника-музея «Хазрет Султан», отельеров (включая менеджмент флагманского

комплекса «Karavansaray»), ресторанного бизнеса и туроператоров. Создание такой структуры позволит синхронизировать городской календарь событий, унифицировать маркетинговые послания на внешние рынки и применять единые, высокие стандарты качества обслуживания туристов на всей территории города.

2. Внедрение технологий смарт-мониторинга и предиктивной аналитики. Для эффективной борьбы с сезонными и событийными перегрузками инфраструктуры целесообразно использование систем геотрекинга и формирования тепловых карт плотности толпы на основе обезличенных данных мобильных операторов и систем городской сети Wi-Fi. Прямое информирование туристов в режиме реального времени о критической загруженности Мавзолея Ходжи Ахмеда Ясави через специализированные мобильные приложения (Smart City App) позволит заблаговременно перенаправлять туристический поток на альтернативные локации дестинации (многофункциональные развлекательные центры, новые интерактивные музеи, этно-аулы, гастрономические улицы).

3. Развитие стратегии пространственного и временного распределения маршрутов. Для решения проблемы «краткосрочных визитов» и увеличения продолжительности пребывания туристов (чтобы среднестатистический гость оставался на 2-3 ночи вместо быстрого транзитного визита на 1 световой день) необходимо формировать логически связанные многодневные маршруты. Такие туристические пакеты должны органично объединять центральное историческое ядро города Туркестан с мощными периферийными археологическими, культурными и природными локациями всего Туркестанского региона (включая городища Отрар и Сауран, природный потенциал Сайрам-Угамского государственного национального природного парка и другие объекты).

В заключение следует с уверенностью отметить, что туристская дестинация Туркестана демонстрирует исключительно высокие темпы развития и в настоящий момент успешно проходит критически важную стадию консолидации на внутреннем и растущем международном туристских рынках. Документально подтвержденное увеличение операционной загрузки номерного фонда и начавшаяся диверсификация географии посетителей являются сильными позитивными трендами. Однако для обеспечения долгосрочной, надежной устойчивости и, что самое главное, сохранения сакральной аутентичности бесценного культурного наследия, городу жизненно необходимо ускоренное внедрение современных цифровых инструментов управления потоками и глубокая, ежедневная координация действий всех стейкхолдеров индустрии гостеприимства.

Список литературы:

1. Buhalis D. Marketing the competitive destination of the future // Tourism Management. — 2000. — Vol. 21, No. 1. — P. 97–116.

2. Butler R.W. The concept of a tourist area cycle of evolution: Implications for management of resources // Canadian Geographer. — 1980. — Vol. 24, No. 1. — P. 5–12.

3. Management of Tourism Flows in Destinations. – Madrid: UN Tourism (UNWTO), 2024. – 112 p.

4. Ivars-Baidal J.A., Celdrán-Bernabeu M.A., Mazón J.N. Smart destinations and the evolution of ICTs: a new scenario for destination management? // Current Issues in Tourism. – 2022. – Vol. 25(10). – P. 1581-1600.

5. Об утверждении Концепции развития туристской отрасли Республики Казахстан на 2023-2029 годы: Постановление Правительства Республики Казахстан от 28 марта 2023 года № 262. – Текст: электронный // Электронное правительство РК.

ПӘНДЕР АРАЛЫҚ ҒЫЛЫМДАР – МЕЖДИСЦИПЛИНАРНЫЕ НАУКИ – INTERDISCIPLINARY SCIENCES

UDC 81'272:004.738.5

Albiyev Asser

Grade 11 student
Quantum STEM School
(Astana, Kazakhstan)

HOW MUCH OF AN ARTICLE COUNT IS HUMAN WORK? DECOMPOSING THE CREATION OF FIVE WIKIPEDIAS, 2010–2025

Abstract: The number of articles in a language's Wikipedia is routinely quoted as evidence of that language's presence online, including in a 2026 study of Kazakh digital language policy that reports 241,858 articles for the Kazakh edition. An article count is a count of pages, not of human work, and pages can be created automatically. Using the Wikimedia Analytics API, this study decomposes every content page created in the Kazakh, Uzbek, Kyrgyz, Azerbaijani, Tatar and Russian editions between January 2010 and August 2025 by the class of account that created it. Accounts carrying the platform's bot flag created 50.4 per cent of the Kazakh pages, 64.1 per cent of Kyrgyz and 90.0 per cent of Tatar, against 0.3 per cent of Azerbaijani. The flag is then shown to be insufficient. A burst rule — a month in which human-flagged page creation is at least ten times its own trailing median — identifies four Kazakh months, thirteen Uzbek and ten Kyrgyz, and none at all in Azerbaijani, Tatar or Russian. In those months a newly created page receives 2.5 to 3.8 subsequent edits against 17.1 to 21.8 in ordinary months, and carries roughly half the net bytes, which is the signature of automated stub creation by accounts the flag does not catch. Treating the burst excess as automation lowers the human share of Kazakh creation from 48.3 to 27.2 per cent. The same rule run as a placebo on already-flagged bot creation recovers 87 to 97 per cent of it in the three editions where it fires, so the signature it detects is automation rather than volume. Ranking the five non-Russian editions by articles gives Tatar first and Azerbaijani fourth; ranking them by human-created pages reverses this, and Azerbaijani leads on both corrected and uncorrected counts. Two findings are specific to Kazakh: no flagged bot has created an article there since 2012, and the human share of edits rose from 17.2 to 95.1 per cent over the period, but this rise is a collapse in bot activity rather than growth in human activity, and the core editing community averaged 44.2 people a month during the 2011 campaign against 37.1 in 2024. An article count should not be read as a measure of human work without this decomposition, and the platform's own bot flag is not sufficient to perform it.

Keywords: Wikipedia; Kazakh language; bots; automated content; digital language vitality; Central Asia; article count

1. Introduction

Wikipedia keeps a public counter of how many articles each of its language editions contains, and that counter has become a shorthand for how well a language is doing online. The habit reaches into the research literature. A 2026 study of Kazakh digital linguistic sovereignty

lists the Kazakh Wikipedia among its indicators of the language's digital presence, reporting that "the section comprised 241,858 articles covering a wide range of subjects, with approximately 2 million edits and a membership of over 10,000 registered participants" [1]. The count is used there exactly as it is used in public argument: as a quantity of Kazakh-language knowledge that exists.

The counter does not measure that. It measures pages. A page can be written by a person over several sittings, or generated by a program from a database in a fraction of a second, and the counter treats the two identically. Where automated creation has been heavy this is not a small correction. The Cebuano edition of Wikipedia is the extreme case: it is the largest non-English edition by article count, and almost the whole of it was produced by a single article-creating program [11].

Kazakhstan has a specific reason to care. In 2011 the foundation WikiBilim ran a campaign to grow the Kazakh edition, and the printed Kazakh National Encyclopedia donated its content to it [12]; a Wikimedia Foundation post from July 2012 records that "Today we have more than 130,000 articles and more than 200 active editors" [12]. Fifteen years later that campaign is still inside the counter that gets quoted. Nobody has said how much of it.

This paper asks a narrow, answerable question: of the content pages created in a language edition between 2010 and 2025, what share was created by a human being? We ask it for Kazakh and for four comparison editions — Uzbek, Kyrgyz and Azerbaijani, which face broadly similar circumstances, and Tatar, which is known for heavy automated creation — with Russian as a large reference.

The first answer is the obvious one, and it is already substantial: on the platform's own bot flag, half of the Kazakh pages created since 2010 were created by bots. The second answer is the one we think matters more. The flag is not sufficient. Mass creation also runs through ordinary, unflagged accounts, and it leaves a measurable trace that the flag misses. Correcting for that trace roughly halves the human share again, and it changes which edition sits at the top of the table.

We report one further result that cuts against a simple pessimistic reading. Kazakh bot creation stopped in 2012 and has not resumed, and the human share of Kazakh editing has risen steadily since. We also show why that rise is weaker evidence than it looks.

2. Background and related work

Automated accounts have been part of Wikipedia almost since its beginning, and the research on them is mostly about what they do rather than how much of the encyclopedia they built. Geiger describes bots as socio-technical actors embedded in the community's rules rather than as mere tools [7]. Halfaker and Riedl distinguish bots proper from "cyborgs" — human editors working through semi-automated tools that combine computational speed with human judgement [6]. That distinction is central to this paper: a cyborg edits from a human account, and no bot flag is attached to it. Tsvetkova and colleagues show that bots interact with each other in patterned ways over long periods [5], and Zheng and colleagues build a taxonomy of bot functions on the English edition and classify 1,601 of them [8], establishing that "bot" is not one kind of thing.

The work closest to ours in method is Steiner's, which monitors real-time edit activity across all language editions and reports that Wikipedias and Wikidata together are edited roughly 50 per cent by bots and 23 per cent by anonymous users [3, 4]. Steiner measures edits.

We measure page creation, which is the quantity the article counter actually accumulates, and we measure it per edition over fifteen years rather than in a live window.

Reeves and Simperl review the literature on automated content generation for Wikipedia [9]. It is a review of generation techniques and evaluation methods; it does not provide a per-edition accounting of how many existing articles were machine-made, which is the gap this paper addresses.

On the other side sits the literature that treats online presence as a measure of a language's health. Kornai argues that most of the world's languages will not make the transition to digital use at all, and that the digital domain is where language vitality is now decided [2]. That framing is what gives an article count its rhetorical force, and it is also what makes an uncorrected count dangerous: if the count is the vitality measure, then a language can appear to be thriving on the strength of work no speaker did.

To our knowledge no study decomposes the article creation of the Central Asian Turkic Wikipedia editions by the class of account responsible, and none tests whether the platform's own bot flag is adequate for such a decomposition. Both are what we do here.

3. Data

All series come from the Wikimedia Analytics REST API [10], which is public, requires no account and no key, and exposes monthly aggregates of the MediaWiki history. We use six editions — kk, uz, ky, az, tt and ru — over the 188 months from January 2010 to August 2025, the last month available at the time of the pull on 23 September 2026.

For each edition and each month we pulled four series, each broken out by the four editor classes the API distinguishes (`user`, `group-bot`, `name-bot`, `anonymous`): content-namespace edits, newly created content pages, distinct editors, and net bytes added. We also pulled human editors split by monthly activity level, and the live article count from each edition's own MediaWiki statistics endpoint. The API returns only months in which a series is non-zero; an absent month is read as zero, which matters for bot series that run in short campaigns.

Two properties of the classes should be stated plainly. `group-bot` means the account held the bot user group; `name-bot` means the account's name contains "bot" without that group; `anonymous` means an unregistered editor. The classification is a property of the account, not of the behaviour, and Section 5.2 is about what that costs.

The window covers most but not all of each edition's stock. Pages created in the window number 97 per cent of the live Kazakh article count, 99 per cent of Kyrgyz, 85 per cent of Azerbaijani, 81 per cent of Uzbek, 75 per cent of Russian and 70 per cent of Tatar. The residual is pages created before 2010 net of pages deleted since. We therefore state our primary results as a decomposition of the 2010–2025 creation flow, which is fully measured, and treat any statement about the present stock as an extrapolation that carries the coverage ratio with it.

4. Method

The first step is arithmetic: for each edition, sum created content pages over the window by editor class and express each class as a share.

The second step is the substantive one. If mass creation also runs through unflagged human accounts, the class shares understate automation, and the understatement will be worst exactly where the campaigns were largest. We need a way to detect automation from behaviour rather than from account status.

We use a burst rule. For each edition and each month from the thirteenth onward, compute the median of human-flagged page creation over the preceding twelve months. A month is a creation burst when human-flagged creation is at least *K* times that trailing median and also clears an absolute floor of 300 pages, which prevents the rule from firing on the ordinary noise of a very small edition. The burst excess is the month's creation minus its own trailing median — that is, the part that is not explained by the edition's own recent human pace. We treat the burst excess as automation. Because a genuine human surge would also be caught, the excess is an upper bound on unflagged automated creation, and the flagged-bot count is a lower bound; the truth lies between, and we report both.

We set *K* = 10 for the main results and sweep *K* over 5, 10, 20 and 50, and the floor over 100, 300 and 1,000, so that no conclusion rests on a chosen threshold.

A rule of this kind must be shown to detect automation rather than simply volume, so we subject it to three checks. First, two behavioural corroborators computed independently of the rule: the number of subsequent edits a newly created page receives, and the net bytes added per new page, both compared between burst and ordinary months. Automated stub creation should produce pages that are barely revisited and short. Second, a placebo: the identical rule applied to creation that is *already* flagged as bot. If the burst shape is the automation signature, it should recover most flagged-bot creation; if it merely flags large months, it proves nothing. Third, the editions where we expect no unflagged campaigns — Azerbaijani, Russian — function as negative controls, and the rule is expected to stay silent there.

Finally, for the question of whether the human share is growing, we aggregate to years and fit ordinary least squares of the annual human share of edits on the year, reporting the slope, its *t* statistic and the fit. Sixteen annual points is a small sample and we treat weak *t* statistics as weak evidence rather than as evidence of absence.

5. Results

5.1 What the article counts are made of

Table 1 — Composition of content-page creation, January 2010 to August 2025.

"Bot" is the platform's own flag (group-bot plus name-bot). Core human editors is the mean monthly count of accounts making 25 or more content edits, over the last twelve months of the window. Articles per core editor divides the live article count by that mean.

Edition	Live articles	Created 2010–2025	Coverage	Bot share of creation	Bot share of edits	Core human editors / month	Articles per core editor
Kazakh (kk)	245,273	238,743	97%	50.4%	49.9%	34	7,232
Uzbek (uz)	361,810	292,212	81%	38.7%	57.3%	58	6,211
Kyrgyz (ky)	76,549	75,853	99%	64.1%	36.7%	7	11,067
Azerbaijani (az)	217,531	185,357	85%	0.3%	25.1%	73	2,973
Tatar (tt)	709,127	499,864	70%	90.0%	87.1%	8	94,550
Russian (ru)	2,118,986	1,589,009	75%	3.6%	19.3%	1,071	1,979

Half of everything the Kazakh edition created since 2010 was created by flagged bots: 120,333 pages against 115,427 from human-flagged accounts and 2,983 from unregistered editors. Kyrgyz is higher at 64.1 per cent and Tatar higher still at 90.0 per cent. Uzbek, at 38.7 per cent, looks the healthiest of the four.

The outlier runs the other way. The Azerbaijani edition has 217,531 articles, within eleven per cent of Kazakh, and flagged bots created 533 of the 185,357 pages it added in the window — three tenths of one per cent. Two editions of comparable size, built in almost opposite ways.

The last column is the crudest possible sanity check and it is already informative. Tatar carries 94,550 live articles for every core human editor active in the last year. Kazakh carries 7,232, Azerbaijani 2,973, Russian 1,979. An encyclopedia is maintained, not just created, and these ratios describe how much material sits behind each person available to maintain it.

5.2 The flag is not enough

Table 1 rests entirely on account status. Section 2 gives the reason to distrust it: semi-automated editing from ordinary accounts is a recognised category [6], and it carries no flag.

The Kazakh series shows the problem immediately. Human-flagged accounts created a median of 271 pages a month across the window. In June 2011 they created 18,448, in July 21,972, in August 9,456. Those three months are 229, 243 and 63 times the trailing twelve-month median, and they alone account for 49,876 pages.

Table 2 — Burst detection on human-flagged page creation, K = 10, floor = 300 pages.

Burst excess is creation above the month's own trailing twelve-month median. The last two columns are computed independently of the rule.

Edition	Burst months	Burst excess pages	Excess as share of human creation	Human share, flag only	Human share, corrected	Later edits per new page, burst	Later edits per new page, ordinary
Kazakh (kk)	4	50,498	43.7%	48.3%	27.2%	2.8	17.1
Uzbek (uz)	13	92,530	52.1%	60.7%	29.1%	3.8	19.6
Kyrgyz (ky)	10	13,755	52.3%	34.7%	16.6%	2.5	21.8
Azerbaijani (az)	0	0	0.0%	97.8%	97.8%	—	21.9
Tatar (tt)	0	0	0.0%	9.7%	9.7%	—	10.9
Russian (ru)	0	0	0.0%	89.7%	89.7%	—	39.5

The rule fires in three editions and stays silent in three. Where it fires, it fires hard: the excess is 43.7 per cent of all human-flagged Kazakh creation, 52.1 per cent of Uzbek and 52.3 per cent of Kyrgyz. Treating that excess as automation cuts the human share of Kazakh creation from 48.3 to 27.2 per cent, of Uzbek from 60.7 to 29.1 and of Kyrgyz from 34.7 to 16.6.

The last two columns are the reason to believe this is not an artefact of the rule. They are not part of the rule and could have come out flat. In Kazakh burst months a newly created page goes on to receive 2.8 later edits; in ordinary Kazakh months, 17.1. The same contrast appears in Uzbek (3.8 against 19.6) and Kyrgyz (2.5 against 21.8). Pages created in burst months are

made once and left. Net bytes point the same way, if less sharply: 2,950 bytes per new page in Kazakh burst months against 6,953 in ordinary ones, 2,870 against 7,679 in Kyrgyz, 3,745 against 4,738 in Uzbek.

The Uzbek bursts are recent and sustained — eight months across 2022, and June 2024 alone with 37,456 pages against a trailing median of 1,174. The Kazakh bursts are all in 2011, and they coincide with the documented WikiBilim campaign and the Kazakh National Encyclopedia donation of that year [12]. We take that coincidence as corroboration that the rule detects a real event, not as an identification of who performed it, which the aggregate data cannot establish.

5.3 Does the detector point the right way?

Table 3 — Placebo. The identical burst rule applied to creation that already carries the bot flag.

A rule that detects automation should recover most of it; a rule that merely detects large months would not.

Edition	Pages created by flagged bots	Falling in months the rule flags	Share recovered
Kazakh (kk)	120,333	105,291	87%
Uzbek (uz)	113,174	51,456	45%
Kyrgyz (ky)	48,555	47,319	97%
Azerbaijani (az)	248	0	0%
Tatar (tt)	440,707	389,384	88%
Russian (ru)	15,121	7,161	47%

Run against creation we already know to be automated, the rule recovers 87 per cent of it in Kazakh, 97 per cent in Kyrgyz and 88 per cent in Tatar. It recovers nothing in Azerbaijani, where flagged bots created only 248 pages in fifteen years and never enough in one month to clear the floor — the correct answer, not a failure. Uzbek and Russian sit near half, meaning their flagged bot work is more evenly spread and less campaign-shaped; where that is so, our correction is conservative, since steady automation does not burst.

Taken with the two behavioural corroborators in Table 2, this is the evidence that the burst shape is the automation shape. It is not proof that every flagged month was automated, and we do not claim it: the excess remains an upper bound.

Table 4 — Corrected human share of content-page creation under different burst thresholds K, floor held at 300.

Varying the floor over 100, 300 and 1,000 pages moves Kazakh not at all (27.2% throughout), Uzbek from 29.0% to 29.8% and Kyrgyz from 16.0% to 20.2%.

K	Kazakh	Uzbek	Kyrgyz	Azerbaijani	Tatar	Russian
5	24.8%	24.8%	15.6%	87.4%	9.7%	89.7%
10	27.2%	29.1%	16.6%	97.8%	9.7%	89.7%
20	27.6%	38.4%	17.9%	97.8%	9.7%	89.7%
50	27.6%	56.8%	28.4%	97.8%	9.7%	89.7%

The sensitivity sweep separates a robust conclusion from a fragile one, and we report both. The Kazakh correction is robust: across a tenfold range of *K* the corrected human share moves only between 24.8 and 27.6 per cent, and changing the floor does not move it at all. The

Uzbek correction is not: at *K* = 5 it gives 24.8 per cent and at *K* = 50 it gives 56.8 per cent. The Uzbek campaigns are long enough to raise their own trailing median, so how much of them the rule catches depends on where the threshold sits. We therefore claim a level for Kazakh and only a direction for Uzbek. Kyrgyz sits in between, ranging from 15.6 to 28.4 per cent, and Azerbaijani flips only at the loosest threshold, where a single 2020 month is caught.

5.4 The league table, three ways

Table 5 — Five editions ranked by the headline article count and by human-created pages in the 2010–2025 window, before and after the burst correction. Russian excluded as the reference case.

Rank	By article count	By human-created pages (flag only)	By human-created pages (corrected)
1	Tatar — 709,127	Azerbaijani — 181,308	Azerbaijani — 181,308
2	Uzbek — 361,810	Uzbek — 177,468	Uzbek — 84,938
3	Kazakh — 245,273	Kazakh — 115,427	Kazakh — 64,930
4	Azerbaijani — 217,531	Tatar — 48,525	Tatar — 48,525
5	Kyrgyz — 76,549	Kyrgyz — 26,312	Kyrgyz — 12,557

Ranking by the public counter puts Tatar first and Azerbaijani last but one. Ranking by pages a person created inverts that: Azerbaijani first, Tatar fourth, and the gap between first and fourth is a factor of 3.7 rather than the factor of 3.3 by which Tatar appeared to lead. Azerbaijani leads on both the uncorrected and the corrected count, and it is the only edition whose position is unchanged by the correction — because there is nothing there to correct.

For Kazakh the two counts differ by a factor of 3.8: 245,273 articles, and 64,930 pages created by a human being under the corrected reading — 115,427 if only the flag is trusted. Both figures are for the 2010–2025 window; the live stock includes pre-2010 pages, which for Kazakh is a small residual given the 97 per cent coverage.

5.5 Is the human share growing?

Table 6 — Annual human share of content edits regressed on year, 2010–2025, sixteen annual observations. 2025 covers January to August only.

Edition	Slope, pp per year	t	R^2	Human share of edits, 2010	Human share of edits, 2025
Kazakh (kk)	+4.80	4.21	0.56	17.2%	95.1%
Uzbek (uz)	+4.45	2.94	0.38	2.3%	75.2%
Kyrgyz (ky)	+2.03	1.81	0.19	29.8%	86.9%
Azerbaijani (az)	+1.86	2.00	0.22	55.9%	83.6%
Tatar (tt)	−1.55	−1.34	0.11	26.3%	17.8%
Russian (ru)	−0.11	−0.30	0.01	60.1%	46.0%

On edits the Kazakh trend is the strongest in the set: +4.80 percentage points a year, t = 4.21, and a rise from 17.2 per cent in 2010 to 95.1 per cent in the first eight months of 2025. Taken alone this reads as a decisively human edition.

The annual series says why that reading is too generous. Kazakh bot edits fell from 278,875 in 2011 and 436,448 in 2012 to 2,980 in 2025. Kazakh human edits went 5,422 in 2010, 204,444 in 2011, 80,828 in 2016, 102,969 in 2024, and 57,687 in the eight months of 2025. The human share rose overwhelmingly because the denominator collapsed. Human editing has not grown since the middle of the decade; the bots left.

The same point holds for creation, in Kazakh's favour. No flagged bot has created a Kazakh content page since 2012: the 120,333 bot-created pages are 49,468 from 2011 and 70,865 from 2012 and nothing after. Every Kazakh article created in the last thirteen years was created from a human-flagged account, and no month since 2011 trips the burst rule. Whatever the stock is made of, the recent flow is human.

The size of the community that produces that flow is the sobering part. The mean monthly count of Kazakh accounts making 25 or more content edits was 4.0 in 2010, 44.2 in the campaign year of 2011, fell to 12.8 by 2017, and recovered to 37.1 in 2024 and 36.0 in the first eight months of 2025. We fitted both a linear and a quadratic trend to these sixteen points and neither is distinguishable from noise ($t = 1.45$ and $t = 1.61$ respectively), so we report no trend and state the levels instead. The level is the finding: fourteen years after the campaign, the core Kazakh editing community has not returned to the size it reached during it.

The following is the author's own account, written before seeing the results of this study (Telegram, 20 and 23 September 2026), lightly edited and translated.

I open the English Wikipedia for almost everything. It has a large community that edits often, so it is usually the most current, and its articles are the most complete. My schoolwork is in English too, so if I need to quote something it is easiest to take it from there. My classmates do the same, except the ones whose English is weak; they use the Russian Wikipedia. The Kazakh Wikipedia we open very rarely, because there is not much in it. The one exception is when we need something from the history of Kazakhstan, because for that the other languages have little. Comparing three articles myself showed the same thing from the other side. On Astana the Russian article had 19 sections against 14 in Kazakh and was more recent, and the Kazakh one looked translated. On Mathematics the Russian article had 15 sections against 5, but the Kazakh one read as if it had been written in Kazakh rather than translated. On Gennady Golovkin the Russian article had 16 sections against 8, and the Kazakh article was again original rather than translated, with more attention to his state awards where the Russian one gave more space to international ones.

The following is the author's reaction after reading the results (Telegram, 25 September 2026), lightly edited and translated.

What surprised me most is that since 2013 not a single article has been created by a bot. I also did not expect that the rise in the human share of editing is not explained by more people arriving, but by bots leaving. And the most interesting thing is that the number of editors has actually fallen compared with 2011.

6. Limitations

The editor classes are properties of accounts. An unflagged account running a script is counted as human until the burst rule catches it, and a flagged bot performing hand-guided work is counted as a bot throughout. The burst rule narrows the first error and does nothing about the second.

The burst excess is an upper bound on unflagged automation, not an estimate of it. A genuine coordinated human effort — a school assignment, an edit-a-thon, a translation drive —

produces the same shape in creation counts, and the behavioural corroborators reduce but do not eliminate that ambiguity: a hurried human stub is also short and also rarely revisited. The honest reading of Kazakh is a range, 27.2 to 48.3 per cent human creation, with our evidence favouring the lower half.

Steady automation is invisible to the rule by construction. An account creating two hundred pages a month for a decade never exceeds its own trailing median and is never flagged. The Uzbek and Russian placebo results (45 and 47 per cent) indicate that some flagged bot work in those editions has exactly this shape, so our corrections there are conservative.

The window begins in January 2010 and covers 70 to 99 per cent of each edition's live stock. The Tatar figure of 70 per cent is the weakest, so Tatar's position in Table 5 is the least secure in the set.

We measure creation and edits, not quality. A human-written page can be worse than a well-generated one, and nothing here should be read as a claim about the content of any article. The quality question would require article-level assessment, which this design does not attempt.

Finally, we do not identify any account, campaign or organisation. The attribution of the 2011 Kazakh bursts to the documented WikiBilim campaign rests on the coincidence of dates with a published record [12], not on account-level evidence, and should be treated as consistent with the data rather than established by it.

7. Conclusion

Of the content pages the Kazakh Wikipedia gained between 2010 and 2025, flagged bots created 50.4 per cent. Reading the burst signature as automation, the share created by a human being falls to 27.2 per cent, and that estimate holds across a tenfold range of the detection threshold. The same exercise puts the Kyrgyz figure at 16.6 per cent and the Tatar figure at 9.7 per cent, while the Azerbaijani edition, of near-identical size to the Kazakh one, was built almost entirely by hand.

The methodological result is the transferable one. The platform's own bot flag is a lower bound on automation and a poor one in exactly the editions where the question matters most, because mass creation also runs through ordinary accounts. The correction is cheap: monthly creation counts by editor class, a trailing-median rule, and two behavioural cross-checks that are already in the same API. It is falsifiable, it stays silent on the editions where nothing happened, and it recovers 87 to 97 per cent of known bot creation when used as a placebo.

For the specific argument this paper started from, the finding cuts both ways and we report both directions. Roughly half to three quarters of the Kazakh article counter is not human work, so the counter should not be quoted as a measure of Kazakh-language knowledge production without this decomposition [1]. But no bot has created a Kazakh article since 2012, the human share of editing has risen every few years since, and the recent flow is entirely human. What has not grown is the number of people doing it: the core Kazakh editing community is smaller today than during the 2011 campaign. The honest summary is that Kazakh Wikipedia's problem is no longer its bots. It is that roughly three dozen people a month are maintaining a quarter of a million articles, half of which somebody else's program wrote fifteen years ago.

Data availability and replication

All data are public and require no account. The pull script (`pull.py`), the analysis script (`analysis.py`) and the complete machine-readable results (`results.json`) accompany this manuscript. Running `pull.py` then `analysis.py` reproduces every number in every table from the Wikimedia Analytics API. The raw API payloads as retrieved on 23 September 2026 are

included so that the tables can be reproduced without network access, and so that any later revision of the underlying data lake is detectable by comparison.

Declarations

Funding: none. Conflicts of interest: none. Ethics: the study uses only aggregate public counts; no personal data, no individual accounts and no human subjects are involved. Use of AI tools: the author used AI assistance for data retrieval code, statistical computation and language editing; the research question, the interpretation of results and the conclusions are the author's own, and every reported number traces to the saved scripts and payloads above.

References:

1. Zharkynbekova S., Ayupova G., Galiyeva B., Shakhputova Z., Zabrodskaja A. Digital linguistic sovereignty and Kazakh language policy in the digital environment // Frontiers in Psychology. 2026. DOI: 10.3389/fpsyg.2026.1858650.

2. Kornai A. Digital Language Death // PLoS ONE. 2013. Vol. 8, No. 10. e77056. DOI: 10.1371/journal.pone.0077056.

3. Steiner T. Bots vs. Wikipedians, Anons vs. Logged-Ins // Proceedings of the 23rd International Conference on World Wide Web Companion (WWW '14 Companion). 2014. arXiv:1402.0412.

4. Steiner T. Bots vs. Wikipedians, Anons vs. Logged-Ins (Redux): A Global Study of Edit Activity on Wikipedia and Wikidata // Proceedings of the International Symposium on Open Collaboration (OpenSym '14). 2014. DOI: 10.1145/2641580.2641613.

5. Tsvetkova M., García-Gavilanes R., Floridi L., Yasseri T. Even good bots fight: The case of Wikipedia // PLoS ONE. 2017. Vol. 12, No. 2. e0171774. DOI: 10.1371/journal.pone.0171774.

6. Halfaker A., Riedl J. Bots and Cyborgs: Wikipedia's Immune System // Computer. 2012. Vol. 45, No. 3. P. 79–82.

7. Geiger R. S. The Lives of Bots // Lovink G., Tkacz N. (eds.) Critical Point of View: A Wikipedia Reader. Amsterdam: Institute of Network Cultures, 2011.

8. Zheng L., Albano C. M., Vora N. M., Mai F., Nickerson J. V. The Roles Bots Play in Wikipedia // Proceedings of the ACM on Human-Computer Interaction. 2019. Vol. 3, CSCW, Article 215. DOI: 10.1145/3359317.

9. Reeves N., Simperl E. Machines in the Margins: A Systematic Review of Automated Content Generation for Wikipedia. 2025. arXiv:2509.22443.

10. Wikimedia Foundation. Analytics Query Service (Wikistats 2) REST API. URL: https://wikimedia.org/api/rest_v1/ (accessed 23.09.2026).

11. Wikipedia's largest non-English version was created by a bot. Generative AI poses new problems // ABC News. 15.04.2025. URL: https://www.abc.net.au/news/science/2025-04-15/wikipedia-cebuano-lsjbot-ai-article-generation-non-english/105123090 (accessed 23.09.2026).

12. Davis L. Class assignment inspired Wikipedian of the Year to grow Kazakh Wikipedia // Wikimedia Foundation, Diff. 05.07.2012. URL: https://diff.wikimedia.org/2012/07/05/class-assignment-inspired-wikipedian-of-the-year-to-grow-kazakh-wikipedia/ (accessed 23.09.2026).

UDC 699.841:711.4

Amangeldy Nurbolat

Grade 12 student
Nazarbayev Intellectual School of Physics and Mathematics, Shymkent
(Shymkent, Kazakhstan)

HOW MUCH OF KAZAKHSTAN'S URBAN BUILT STOCK PREDATES ITS SEISMIC CODES, AND WHERE DOES THAT STOCK SIT RELATIVE TO HAZARD?

Abstract: Kazakhstan introduced modern seismic design requirements in 1982, revised them in 2006 and replaced them with a Eurocode-harmonised code in 2017, yet no published study reports how much of the country's present urban built-up area was laid down before those rules existed, or how that older stock is distributed across seismic hazard. This study answers both questions for all 32 Kazakhstani urban centres using only open global data: the European Commission's GHS-AGE R2025A product, which records the epoch in which each 100 m cell first became built-up, masked by the GHS Urban Centre Database polygons, crossed with peak ground acceleration from the Global Earthquake Model hazard map and with the database's own macroseismic intensity classes. Of 180,745 hectares of urban built-up area, 47.9 per cent dates from before 1980 and 58.5 per cent from before 1990. Nine centres sit at a peak ground acceleration of 0.15 g or above; they hold 37.4 per cent of the national urban built-up area and 29.2 per cent of the pre-1980 stock. Exposure, defined as pre-code area multiplied by hazard, is extremely concentrated: the top three centres account for 79.1 per cent of the national total and Almaty alone for 61 per cent, 5.2 times the next city. The relationship between hazard and the age of the stock runs in the opposite direction to intuition: across the 32 centres a higher hazard is associated with a *lower* pre-1980 share (Spearman −0.53; regression coefficient on log hazard −0.028, $p = 0.002$, controlling for city size), because the aseismic northern cities are the oldest by share while the high-hazard south has been rebuilt fastest. The practical implication is that a national seismic-retrofit priority cannot be set from either variable alone: the stock that is both old and shaken is a small, identifiable set of places, led by Almaty.

Keywords: seismic risk; built environment; construction era; Kazakhstan; remote sensing; GHS-AGE; urban exposure

Introduction

Shymkent is an old city, and it has many old apartment blocks that may not withstand a strong earthquake. The district where my grandparents live is literally called the Old Town. Many of its private houses are now being demolished and replaced by new residential complexes. Near my school there is a residential complex where the walls are visibly crumbling and there are large cracks, which worries me for the people who live there and for everyone who walks past. At the same time new blocks are going up in the city centre, such as Turan, Asar 2 and Nursat 2.

An earthquake does not damage a city uniformly. What it damages depends on how the buildings were designed, and that in turn depends on when they were built, because seismic

design requirements change over time and apply only to what is built after them. Kazakhstan's requirements have three dates. The Soviet code SNiP II-7-81, *Construction in seismic regions*, came into force on 1 January 1982. It was replaced for independent Kazakhstan by SNiP RK 2.03-30-2006, and that in turn by SP RK 2.03-30-2017, *Construction in seismic zones*, in force from 20 December 2017 and harmonised with Eurocode 8. Everything standing today that was built before 1982 was therefore designed to a weaker standard than everything built after it, and in many cases to no explicit seismic standard at all.

The hazard itself is unevenly distributed. The south-east of the country, along the Tian Shan front, carries peak ground accelerations above 0.4 g, comparable to the most seismically active parts of Central Asia, while the northern and western plains are effectively aseismic. Almaty, the largest city, sits in the highest-hazard zone; Astana, the capital, does not.

Both facts are known. What has not been published is the intersection: how much of the built-up area of each Kazakhstani city dates from before the codes, and whether the cities with the most pre-code construction are the cities that shake. That intersection is what determines where retrofit money would do most good, and it can now be estimated without a single field survey, because the European Commission's Joint Research Centre publishes a global satellite-derived product recording the epoch in which every 100 m cell first became built-up, and the Global Earthquake Model publishes a global hazard map on a comparable grid.

This study crosses those two products for all 32 Kazakhstani urban centres and asks three questions. What share of the present urban built stock predates the 1982 code? How is that pre-code stock distributed across hazard, and which cities concentrate it? And does the age of a city's stock relate to its hazard at all, once city size is accounted for?

Materials and Methods

Unit of analysis. The 32 Kazakhstani urban centres defined by the GHS Urban Centre Database R2024A (GHS-UCDB V1.2). The urban centre, rather than the administrative city boundary, is used because it is defined consistently worldwide from population and built-up density, so the comparison between cities does not depend on where an administrative line happens to fall. Each centre carries a population for 2025 and a macroseismic intensity class from the same database.

Construction era. GHS-AGE R2025A at 100 m resolution records, for every built-up cell, the five-year epoch in which it first became built-up, from before 1975 through 2015–2020. Each cell is one hectare. The product was masked by each urban-centre polygon and the cells tallied by epoch class, giving for every city the built-up area in each era. Four era bands are used, mapped onto Kazakhstan's code dates at the nearest class boundary: pre-code (before 1980, approximating the 1982 code), Soviet-code era (1980–2005), first Kazakh code (2005–2015, approximating 2006–2017) and Eurocode-based (2015–2020, approximating 2018 onward). The mapping is imperfect by up to two years at each boundary because the satellite epochs do not fall on the code years; a sensitivity analysis shifting every boundary one class later is reported.

Hazard. Peak ground acceleration with a 10 per cent probability of exceedance in 50 years, on rock, from the Global Earthquake Model global seismic hazard map v2023.1, averaged over each urban-centre polygon. The database's macroseismic intensity class is carried alongside as an ordinal check. Almaty is the only Kazakhstani centre that the database splits across two intensity classes, 7 and 8, with 1.09 million residents in the lower and 1.22 million in the higher.

Measures. For each city: total built-up area in hectares; the share of that area in each era band; and an exposure index defined as pre-code built-up area multiplied by hazard, E = A_pre1980 × PGA, in hectare-g. The index is deliberately simple and is not a loss estimate: it is a ranking device that answers "where is there the most old building in the places that shake hardest".

Statistics. The association between hazard and the age of the stock is reported three ways: Pearson and Spearman correlations between PGA and the pre-1980 share, and an ordinary-least-squares regression of the pre-1980 share on log hazard and log population with heteroskedasticity-robust (HC1) standard errors. With n = 32 the regression is reported as a descriptive association, not as a causal estimate, and the rank correlation is given because it does not assume linearity.

Decision rules, fixed before the analysis was run. Exposure is called "concentrated" only if the top three centres hold at least half of the national total. The hazard-versus-age relationship is claimed only if the regression coefficient on log hazard and the Spearman correlation are both significant at 5 per cent and agree in sign; otherwise it is reported as no association.

What is not claimed. GHS-AGE records the surface age of built-up land, not the structural type, material or code compliance of any building; the paper therefore speaks of "built-up surface dating from era X", never of "buildings that violate code Y". The product records first appearance, so demolition and replacement on the same footprint are invisible and the pre-code share is, if anything, an overestimate of what physically stands from that era. The GEM hazard grid is coarse relative to a city, so within-city variation is captured only through the UCDB two-class split for Almaty. Nothing after the product's last observed imagery, about 2018, is asserted.

Reproducibility. The per-city class tallies, the analysis script (Python, standard library only) and the full numerical output are released with this paper.

Results

The national picture. The 32 urban centres hold 180,745 hectares of built-up area. Of that, 47.9 per cent dates from before 1980 and 58.5 per cent from before 1990. Just under half of urban Kazakhstan, by surface, therefore predates the first modern seismic code.

Where the hazard is. Nine of the 32 centres sit at a polygon-mean peak ground acceleration of 0.15 g or above. Together they hold 37.4 per cent of the national urban built-up area and 29.2 per cent of the national pre-1980 stock. The remaining 23 centres, including the capital, are at or below that threshold, and the 22 centres below 0.05 g are effectively aseismic.

Table 1 – The ten Kazakhstani urban centres with the highest seismic exposure of pre-code built stock, 2026. Exposure index = pre-1980 built-up area × peak ground acceleration.

Rank	Urban centre	Built-up, ha	Pre-1980 share, %	PGA, g	Exposure index, ha·g	Population 2025
1	Almaty	30,313	43.8	0.399	5,300	2,315,538
2	Shymkent	21,598	28.5	0.165	1,017	1,239,663
3	Taraz	8,675	56.2	0.106	517	234,775
4	Taldykorgan	3,609	43.2	0.223	347	146,374
5	Esik	1,408	40.8	0.437	251	53,719
6	Zharkent	1,683	58.5	0.250	246	54,229
7	Sayram	2,472	29.4	0.167	122	71,026

8	Aqsukent	2,954	23.9	0.163	115	65,489
9	Jetisay	1,303	49.8	0.171	111	61,507
10	Ust-Kamenogorsk	5,516	54.6	0.036	108	240,226

Exposure is concentrated, and it is concentrated in one city. The national exposure total is 8,643 hectare-g. The top three centres hold 79.1 per cent of it and the top five 86.0 per cent, so the pre-registered threshold for "concentrated" is passed comfortably. Almaty alone accounts for 61 per cent, an exposure index 5.2 times that of Shymkent, the next city. Almaty holds 13,283 hectares of pre-1980 built-up area, 15.3 per cent of the national pre-1980 stock, and it sits at the highest hazard of any large Kazakhstani city. No other city combines the two.

Age runs against hazard, not with it. Across the 32 centres, higher hazard is associated with a *lower* pre-1980 share. The Spearman correlation is −0.53 and the Pearson correlation −0.45; in the regression, the coefficient on log hazard is −0.028 (p = 0.002) while city size is irrelevant (log population, p = 0.965), with an R^2 of 0.28. Both pre-registered conditions are met and they agree in sign, so the association is reported as real. The mechanism is visible in the extremes: the nine high-hazard centres average 38.3 per cent pre-1980 stock, the 22 aseismic centres 54.8 per cent. The oldest cities by share are Kostanay (73.4 per cent), Oral (67.1), Kokshetau (67.0), Semey (65.4) and Atyrau (64.2), all of them in the aseismic north and west. The fastest-renewing large cities are Shymkent, where 38.7 per cent of the built-up area appeared after 2005, Aktau (36.7) and Astana (32.2).

Table 2 – The hazard-versus-age association under three definitions of the pre-code era. All specifications control for log population with HC1 robust standard errors, n = 32.

Specification	Coefficient on log PGA	p	Spearman	R^2
Pre-1980 share (baseline)	−0.0278	0.002	−0.53	0.28
Pre-1990 share	−0.0228	0.006	−0.49	0.24
Boundary shifted one class later (pre-1985)	−0.0234	0.003	−0.48	—

Sensitivity. Moving the era boundary one class later, to pre-1985, leaves the association intact (coefficient −0.023, p = 0.003; Spearman −0.48), as does using the pre-1990 share instead (−0.023, p = 0.006). The finding is therefore not an artefact of where the code year is mapped onto the satellite epochs.

Almaty split by intensity. Almaty is the single centre the database divides between intensity classes 7 and 8, with 1.22 million of its 2.32 million residents in the higher class. The city's overall pre-1980 share, 43.8 per cent, is above the average for high-hazard centres, and its absolute pre-code area exceeds that of the next nine high-hazard centres combined. Resolving the era composition of the two intensity bands separately requires the intensity polygons rather than the aggregate class counts and is the natural extension of this work.

Discussion

Half of urban Kazakhstan, by built surface, predates the country's first modern seismic code, and that fact alone says little about risk, because most of that old surface stands where the

ground does not shake. The distribution matters more than the total, and the distribution turns out to be lopsided in two ways at once.

First, it is lopsided between cities. Three centres hold four-fifths of the national exposure and one city holds three-fifths of it. A national retrofit programme that treated all old building as equally urgent would spend most of its money in Kostanay, Oral and Semey, the cities with the oldest stock by share, where the hazard is close to zero. The exposure ranking points instead at a short list led by Almaty, with Shymkent and Taraz behind it and several small southern towns, Esik and Zharkent among them, appearing far above their population weight because they are both old and close to the Tian Shan front.

Second, it is lopsided in a direction that is easy to get backwards. The high-hazard south is not the country's oldest built environment; it is among its newest, because Shymkent, Taldykorgan and Turkestan have been rebuilt and extended rapidly since 2005, while the aseismic north has grown slowly and kept what it had. The negative association between hazard and age is therefore not a comforting result. It means that the aggregate national statistic, "half the stock is pre-code", is a poor guide to anything, and that the cities which most need attention are the exceptions to it, above all Almaty, which combines the highest hazard among large cities with an absolute pre-code area unmatched anywhere in the country.

The limits of the method should be read alongside these results. The satellite product sees the first appearance of built-up surface, not buildings; a 1960s footprint rebuilt in 2015 still counts as pre-1980, which biases the pre-code share upward and makes this an upper bound rather than a point estimate. Hazard is taken as a single value per city from a grid coarser than a city, so within-city differences, which are exactly what a municipal engineer needs, are beyond this design except for the Almaty split. And code dates are mapped to five-year epochs, so the boundary is approximate by up to two years, though the sensitivity analysis shows the conclusion does not hinge on it.

The epoch assignment was not checked against dated buildings on the ground. Comparing the satellite epoch of named buildings with their known construction years is the most direct next test of the upper-bound bias described above, and is left for further work.

Conclusion

Across Kazakhstan's 32 urban centres, 47.9 per cent of the 180,745 hectares of built-up area predates the 1982 seismic code and 58.5 per cent predates 1990. Nine centres sit at a peak ground acceleration of 0.15 g or above and hold 29.2 per cent of that pre-code stock. Exposure, measured as pre-code area times hazard, is concentrated to an unusual degree: 79.1 per cent of the national total in three cities and 61 per cent in Almaty alone, 5.2 times the next city. Across cities, hazard and the age of the stock are negatively associated (Spearman −0.53; −0.028 on log hazard, $p = 0.002$), because the aseismic north is the oldest by share while the high-hazard south has renewed fastest, and this holds under every alternative definition of the pre-code era tested. The policy reading is narrow and, for that reason, usable: the question "where is Kazakhstan's seismically exposed old building" has a short answer with a named list at the top of it, and that list can be produced for any country from open global data in a day, without a field survey.

Acknowledgments

Built-up epoch data are published by the European Commission Joint Research Centre (GHS-AGE R2025A, GHS Urban Centre Database R2024A) and the hazard map by the Global Earthquake Model Foundation (Global Seismic Hazard Map v2023.1); both are openly licensed

and available to any reader. The analysis script and the full numerical output are released with this paper. The author declares no conflict of interest and received no funding for this work.

Use of AI tools. The author used generative AI tools to help write the analysis code, run the computations and edit the English text. The research direction, the observations of Shymkent in the author's own account and the choice of question are the author's. Every number reported here is produced by the released analysis script from the published JRC and GEM data; none was generated by a language model.

References:

1. Pesaresi M., Politis P. GHS built-up surface and age of the built environment, multitemporal (1975–2030). European Commission, Joint Research Centre (JRC), 2025. https://doi.org/10.2905/9F06F36F-4B11-47EC-ABB0-4F8B7B1D72EA

2. Florczyk A. J., Corbane C., Ehrlich D. et al. GHS Urban Centre Database, multitemporal and multidimensional attributes. European Commission, Joint Research Centre (JRC), 2024. https://data.jrc.ec.europa.eu/dataset/ghs-ucdb

3. Pagani M., Garcia-Pelaez J., Gee R. et al. The 2023 global seismic hazard map of the Global Earthquake Model. GEM Foundation, Zenodo, 2023. https://doi.org/10.5281/zenodo.8409647

4. SNiP II-7-81. Construction in seismic regions. USSR State Committee for Construction, in force from 1 January 1982.

5. SP RK 2.03-30-2017. Construction in seismic zones. Committee for Construction and Housing and Communal Services, Republic of Kazakhstan, in force from 20 December 2017.

6. Silva V., Amo-Oduro D., Calderon A. et al. Development of a global seismic risk model. Earthquake Spectra, 2020, vol. 36, no. 1_suppl, pp. 372–394. https://doi.org/10.1177/8755293019899953

7. Albiyev A. Decoupling gas flaring from oil production in Kazakhstan: a field-level satellite analysis, 2012–2025. Central Asian Scientific Journal, 2026, no. 2(30), vol. 3, p. 133.

Электронный научный журнал «Central Asian Scientific Journal»

Редактор: Байдильдинов Т.Ж.
Комп.верстка: Хусаинов Е.М.

Электронный научный журнал «Central Asian Scientific Journal»
-2026-34(32)-Астана-ИП DOC
Зарегистрировано и выдано свидетельство
Министерством Информации и Общественного Развития РК
№KZ77VPY00147053 от 17.04.2026 г.
ISSN: 3135-3061

За достоверность публикуемой информации, цитат и иных изложений ответственность несет автор