**Описание данных для Хакатона СберИндекса**

В этом документе термин «муниципальные образования» сокращён до МО. Набор данных состоит из следующих частей:

1. Потребительские расходы на уровне МО: 8\_consumption.parquet
2. Индекс доступности рынков на уровне МО: 1\_market\_access.parquet
3. Данные Росстата
   1. Население МО: 2\_bdmo\_population.parquet
   2. Миграция по МО: 3\_bdmo\_migration.parquet
   3. Заработная плата по МО: 4\_bdmo\_salary.parquet
4. Автодорожные связи между МО: 5\_connection.parquet
5. Справочник МО: <https://sberindex.ru/ru/research/dataset-borders-and-changes-of-municipalities>
6. Сеточные данные о плотности населения: <https://geoportal.hse.ru/portal/home/item.html?id=9ee9d4ad2b124f82949f9f061e0b42c9>

# 1. Потребительские расходы на уровне МО

Таблица 8\_consumption.parquet содержит следующие поля:

| **Поле** | **Содержание** |
| --- | --- |
| territory\_id | Код территории (МО в постоянных границах) |
| date | Год и месяц |
| category | Категория потребительских расходов |
| consumption | Оценка средних безналичных потребительских расходов в текущем месяце жителей МО на основе моделей СберИндекса на транзакционных данных (руб.) |

Охват по времени: с января 2023 г. по декабрь 2024 г. включительно.

Охват по МО: отсутствуют около 500 МО (точное количество варьируется от месяца к месяцу), для которых оценка не превышает необходимый порог качества.

Включены следующие категории:

* Категория "*Продовольствие*" охватывает расходы на покупку продуктов питания для домашнего приготовления и потребления. В данную категорию входят транзакции в супермаркетах, гипермаркетах, продуктовых магазинах у дома, на рынках и ярмарках, а также расходы на алкогольных напитки и табачные изделия.
* Категория "*Здоровье*" включает в себя покупки лекарственных препаратов, медицинских изделий и сопутствующих товаров, отпускаемых по рецепту и без рецепта, а также широкий спектр услуг по диагностике, лечению и профилактике заболеваний, оказываемых медицинскими учреждениями и специалистами.
* Категория "*Общественное питание*" охватывает расходы на услуги предприятий, занимающихся приготовлением и продажей готовых блюд и напитков для немедленного потребления на месте или на вынос. В данную категорию входят транзакции в ресторанах, кафе, барах, фастфуд-заведениях, столовых, а также в службах доставки готовой еды. Сюда относятся как покупки отдельных блюд и напитков, так и оплата кейтеринговых услуг для банкетов и корпоративных мероприятий.
* Категория "*Транспорт*" включает в себя расходы на услуги городского и пригородного общественного транспорта, такси, каршеринг, расходы на топливо и услуги по ремонту, техническому обслуживанию и уходу за автотранспортными средствами. Также включены расходы на услуги по перевозке, хранению и доставке различных видов отправлений. Не включает авиа- и ж/д билеты.
* Категория "*Маркетплейсы*" охватывает транзакции на онлайн-платформах, которые предоставляют площадку для продажи товаров и услуг множеством различных продавцов. Маркетплейсы обычно предлагают широкий ассортимент товаров, включая одежду, электронику, бытовую технику, книги и др. Транзакции в специализированных интернет-магазинах относятся к соответствующим категориям.
* "*Все категории*" включает вышеперечисленные и иные категории.

# 2. Индекс доступности рынков на уровне МО

Таблица 1\_market\_access.parquet содержит следующие поля:

| **Поле** | **Содержание** |
| --- | --- |
| territory\_id | Код территории (МО в постоянных границах) |
| market\_access | Индекс доступности рынков в 2024 году |

Охват по МО: отсутствуют 22 МО без постоянного автодорожного сообщения с сетью автомобильных дорог или без паромного сообщения с материковой частью России.

Индекс доступности рынков показывает относительный потенциальный объём внешнего рынка, доступный в МО. Более высокое значение индекса говорит о большем потенциале. Рассчитывается по формуле:

где МА – показатель доступности рынков для муниципалитета *o*, N – население муниципалитета *d*, *τ* – расстояние от муниципалитета *o* до муниципалитета *d.* Далее значения были нормированы от 0 до 1 000.

# 3. Данные Росстата

# 3.1. Население МО

Таблица 2\_bdmo\_population.parquet содержит следующие поля:

| **Поле** | **Содержание** |
| --- | --- |
| territory\_id | Код территории (муниципального образования в постоянных границах) |
| year | Год, за который опубликованы данные |
| period | Периодичность (год) |
| age | Возрастная группа |
| gender | Пол |
| value | Численность всего населения на 1 января текущего года, человек |

Охват по времени: с 2023 г. по 2024 г. включительно.

Охват по МО: полностью отсутствуют данные по г. Санкт-Петербург и г. Севастополь. По остальным МО могут отсутствовать значения в некоторых возрастных группах.

# 3.3. Миграция по МО

Таблица 3\_bdmo\_migration.parquet содержит следующие поля:

| **Поле** | **Содержание** |
| --- | --- |
| territory\_id | Код территории (муниципального образования в постоянных границах) |
| year | Год, за который опубликованы данные |
| period | Периодичность (год) |
| age | Возрастная группа |
| gender | Пол |
| value | Миграция, всего |

Охват по времени: 2023 г.

Охват по МО: могут отсутствовать значения в некоторых возрастных группах.

# 3.4. Заработная плата по МО

Таблица 4\_bdmo\_salary.parquet содержит следующие поля:

| **Поле** | **Содержание** |
| --- | --- |
| territory\_id | Код территории (муниципального образования в постоянных границах) |
| year | Год, за который опубликованы данные |
| period | Периодичность (кварталы) |
| okved\_name | Наименование ОКВЭДа |
| okved\_letter | Буквенное обозначение ОКВЭДа |
| value | Среднемесячная заработная плата работников организаций (без субъектов малого предпринимательства) |

Охват по времени: с 1го квартала 2023 г. по 4й квартал 2024 г. включительно.

Охват по МО: по некоторым муниципалитетам могут отсутствовать данные в отдельных ОКВЭДах.

# 4. Автодорожные связи между МО

Таблица 5\_connection.parquet содержит следующие поля:

| **Поле** | **Содержание** |
| --- | --- |
| territory\_id\_x | Код территории (МО в постоянных границах) отправления |
| territory\_id\_y | Код территории (МО в постоянных границах) прибытия |
| distance | Расстояние (км) на 31 декабря 2024 года |

Охват по МО: отсутствуют 22 МО без автодорожного сообщения.

Расстояние рассчитывается по графу автодорог (OSM) от центра одного МО до центра другого. В таблице для любых пар **x** и **y** расстояние рассчитывалось лишь в одну сторону (**x** → **y**), предполагая, что обратный путь (**y** → **x**) такой же.

# 5. Справочник МО

Данные и их описание доступны по ссылке: <https://sberindex.ru/ru/research/dataset-borders-and-changes-of-municipalities>

# 6. Сеточные данные о плотности населения

Данные и их описание доступны по ссылке: <https://geoportal.hse.ru/portal/home/item.html?id=9ee9d4ad2b124f82949f9f061e0b42c9>