# Отчет по проекту “Сделай мир чище”

***Анализ рынка клининговых услуг в регионах России***

## Введение

***Цель проекта:*** *Создать сервис, предоставляющий предпринимателям аналитику по рынку клининговых услуг для оптимизации цен, выявления перспективных регионов и оценки конкуренции.*

***Целевая аудитория:*** *-* ***B2B:*** *Малые и средние бизнесы, франчайзинговые компании.  
-* ***Проблемы ЦА:*** *- Отсутствие данных для установки конкурентных цен.  
- Сложности в оценке платежеспособности клиентов и уровня конкуренции.  
- Недостаток информации о региональных особенностях рынка.*

***Основные задачи:*** *1. Сбор данных из открытых источников (Яндекс.Услуги, Росстат).  
2. Анализ цен, отзывов и географии услуг.  
3. Построение портрета целевой аудитории.  
4. Визуализация данных для принятия решений.*

## Методология

### Источники данных:

1. ***Статический парсинг*** *(Яндекс.Услуги, данные Росстата).*
2. ***Динамический парсинг*** *(Selenium).*
3. ***API Яндекс.Карт*** *для геокодирования регионов.*

### Инструменты:

* ***Python-библиотеки:*** *Selenium, pandas, Plotly, Matplotlib.*
* ***ООП:*** *Классы для обработки данных (ServiceDataProcessor, GeoDataEnhancer), визуализации (DataVisualizer).*
* ***Визуализация:*** *Интерактивные карты, динамические графики, дашборды.*

## Ход работ

### 1. Сбор данных

***Источники и методы:*** *-* ***Росстат:*** *Загружены данные по индексу физического объема бытовых услуг (2017–2025 гг.) для анализа спроса.  
-* ***Яндекс.Услуги:*** *Динамический парсинг через undetected\_chromedriver для сбора данных по 10 топовым городам (406 записей).  
-* ***API Яндекс.Карт:*** *Получены координаты регионов для визуализации.*

***Проблемы и решения:*** *-* ***Капча на Яндекс.Услугах:*** *Реализовано ожидание и ручное решение для обхода блокировок.  
-* ***Неоднородные форматы цен:*** *Регулярные выражения для преобразования “₽/м²” в фиксированную стоимость (предположение: 50 м²).*

### 2. Обработка данных

***Этапы:*** *1.* ***Очистка:*** *- Удаление дубликатов (например, повторяющиеся записи по Архангельской области).  
- Замена пропущенных значений нулями в данных Росстата.  
2.* ***Сопоставление регионов:*** *- Использование fuzzywuzzy для автоматического матчинга городов из датасета GitHub (https://github.com/hflabs/city/).  
- Ручная корректировка для регионов с низким score (например, “Ненецкий автономный округ”).  
3.* ***Расчет показателей:*** *- Платежеспособность: Взвешенные расходы на душу населения с учетом временного тренда (2017–2025 гг.).*

***Улучшение точности:*** *- Добавление весовых коэффициентов для учета динамики спроса.  
- Фильтрация городов с покрытием данных менее 2000 дней.*

### 3. Анализ и коррекция

***Промежуточные результаты:*** *-* ***Таблица Росстата:*** *Выявлены аномалии (например, резкий рост расходов в Белгородской области в 2017 г.).  
-* ***Геокодирование:*** *Ошибки в координатах для малых регионов (исправлены через ручной ввод).*

***Корректировки:*** *- Исключение данных с нулевыми значениями расходов.  
- Агрегация услуг в две категории: “Уборка квартир” и “Уборка после ремонта”.*

### 4. Визуализация

***Инструменты:*** *-* ***Интерактивная карта (Plotly):*** *Отображение динамики расходов по годам.  
-* ***Графики (Matplotlib):*** *Сравнение средних цен в топ-10 городах.*

***Оптимизация:*** *- Добавление анимации для карты (2017–2025 гг.).  
- Нормализация шкалы расходов для устранения выбросов.*

***Комментарий:*** *Ход работ демонстрирует итеративный подход: от сырых данных — через очистку, анализ и коррекцию — к финальным инсайтам. Ключевые улучшения:  
1. Учет временных трендов через взвешенные метрики.  
2. Упрощение интерпретации за счет категоризации услуг.*

*Это позволило сократить погрешность расчетов и обеспечить бизнес-рекомендации, привязанные к реальным рыночным условиям.*

## Результаты анализа

### 1. География и платежеспособность ЦА

* **Топ-10 городов по покупательной способности:**  
  | Город | Средние расходы (руб/чел) |  
  |———————-|—————————|  
  | Ростов-на-Дону | 1.97 млрд |  
  | Новосибирск | 1.65 млрд |  
  | Екатеринбург | 1.27 млрд |
* **Регионы с высоким спросом:** Центральный, Приволжский, Южный федеральные округа.
* **Общий портрет ЦА**- это люди платежеспособного возраста (25-60 лет), мужчины и женщины, которые делают ремонт или слишком заняты работой/бытовыми делами, с хорошим достатком (высокой покупательной способностью).

### 2. Анализ цен на услуги

* ***Средняя стоимость уборки квартир:***
  + *Москва: 2 500–3 000 руб.*
  + *Ростов-на-Дону: 1 500–2 000 руб.*
* ***Уборка после ремонта:***
  + *Санкт-Петербург: 4 000–5 000 руб.*

### 3. Конкуренция

* ***Основные конкуренты:*** *Мелкие локальные компании (80% рынка), реже федеральные сети.*
* ***Ключевые отличия сервиса:***
  + *Данные по платежеспособности покупателей по всем регионам.*
  + *Рейтинг компаний на основе отзывов, рейтинга и проверенных документов.*

## Выводы и рекомендации

### Стратегия выхода на рынок:

1. ***Фокус на топ-*3 *городов*** *(Ростов-на-Дону, Новосибирск, Екатеринбург).*
2. ***Ценообразование:*** *Гибкие тарифы с учетом региональных особенностей.*
3. ***Партнерства:*** *Интеграция с франчайзинговыми сетями для анализа издержек.*

### Преимущества сервиса:

* ***Для предпринимателей:***
  + *Снижение рисков за счет анализа конкурентов и ЦА.*
  + *Оптимизация логистики (API Яндекс.Карт).*
* ***Для франчайзеров:***
  + *Оценка рентабельности новых филиалов.*

## Итог

*Проект “Сделай мир чище” предоставляет предпринимателям инструменты для анализа рынка, минимизации рисков и максимизации прибыли.*