

Министерство науки и высшего образования Российской Федерации

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО
ОБРАЗОВАНИЯ**

**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО»
(Университет ИТМО)**

Факультет **Инфокоммуникационных технологий**

Образовательная программа **Интеллектуальные системы в гуманитарной сфере**

Направление подготовки **45.03.04 Интеллектуальные системы в гуманитарной сфере**

КУРСОВОЙ ПРОЕКТ

по дисциплине «Наука о данных»

на тему: «Сравнительный анализ степени озеленения городского пространства в г. Санкт-Петербурге»

Обучающиеся:

Алексеев Павел Алексеевич, группа К33421,

Смирнов Тимур Олегович, группа К33421

Работа выполнена с оценкой _____

Преподаватель (и): Гусарова Наталья Федоровна

_____ (подпись)

_____ (подпись)

Дата _____

Санкт-Петербург

2022

Оглавление

<i>ВВЕДЕНИЕ</i>	3
<i>Бизнес-анализ (Business Understanding)</i>	4
<i>Анализ данных (Data Understanding)</i>	5
I. Набор данных “Зеленые насаждения общего пользования городского значения в Санкт-Петербурге” (Датасет №1)	5
II. Набор данных “Зеленые насаждения общего пользования местного значения в Санкт-Петербурге” (Датасет №2)	6
<i>Подготовка данных (Data Preparation)</i>	7
I. Набор данных “Зеленые насаждения общего пользования городского значения в Санкт-Петербурге” (Датасет №1)	7
II. Набор данных “Зеленые насаждения общего пользования местного значения в Санкт-Петербурге” (Датасет №2)	8
<i>Моделирование (Modeling)</i>	10
<i>Заключение</i>	15
<i>Список Источников</i>	16
<i>Приложение</i>	17

ВВЕДЕНИЕ

Даже быстрое сравнение норм озеленения на одного человека в столице и у нас наталкивает на то, что проблема есть: в Москве эта норма составляет 24 м²/чел, а в Санкт-Петербурге такая же норма для центральных районов - 6 м²/чел.

На сегодняшний день трудно назвать Санкт-Петербург зеленым городом, но так было не всегда.

После распада СССР деревья начали массово вырубать. И, к огромному сожалению, Санкт-Петербург перестал быть зеленым городом. В 1991 году была изменена система бюджетирования и ликвидирован отдельный орган в городской власти, отвечающий за озеленение.

Стало выгодным выделять средства на конкретные действия. Сейчас властями принято закупать цветы по шесть раз в год вместо необходимого ухода за ними, многократно подстригать газоны и кронировать деревья.

Часто это делать нельзя – деревья начинают умирать и их вырубают. Такая схема распределения бюджета направлена на получение личной выгоды, а не на сохранение облика города.

За последние 30 лет не было озеленено ни одной новой улицы.

Бизнес-анализ (Business Understanding)

На данный момент большинство жителей, города Санкт-Петербурга, недовольны качеством, а также количеством озелененных территории в городе.

Цель данной курсовой работы – показать степень озеленения районов Санкт-Петербурга, а также проверить гипотезу того, что наш город мало озеленен.

Объектом исследования являются наборы данных: «Зеленые насаждения общего пользования городского значения в Санкт-Петербурге (Архивный) (Версия №1 от 22.04.2021)» и «Зеленые насаждения общего пользования местного значения в Санкт-Петербурге (Архивный) (Версия №1 от 22.04.2021)»

За основу были взяты метрики Всемирной Организации Здравоохранения (далее ВОЗ) о нормах озеленения городов:

- 50 м² городских зеленых насаждений на одного жителя;
- Плохими по условиям озеленения считаются города, где растительность занимает менее 10 % площади города, хорошими – 40–60 %;
- Норма зелёных насаждений общего пользования для крупных городов — 21 м² на одного человека.

Анализ данных (Data Understanding)

Данные для исследования были собраны с сайта “Системы классификаторов Санкт-Петербурга”, который содержит более 200 различных наборов данных и предоставляет открытый доступ к ним. Для анализа использовался набор данных в формате CSV.

I. Набор данных “Зеленые насаждения общего пользования городского значения в Санкт-Петербурге” (Датасет №1)

Данный набор данных содержит 4923 строки и 7 столбцов с информацией о расположении зеленых насаждений (см. Рисунок 1).

Описание:

- Район (Район города)
- Номер территории (Условный номер объекта на территории Санкт-Петербурга)
- Адрес (Адрес объекта внутриквартального озеленения)
- Ориентировочная (Площадь объекта внутриквартального озеленения, ориентировочная, га)
- Основание (Реквизиты документа, утверждающего перечень внутриквартального озеленения)
- Муниципальное (Наименование муниципального образования)
- Дата внесения в базу (Дата внесения объекта в базу данных)

	Район	Номер территории	Адрес	Ориентировочная	Основание	Муниципальное	Дата внесения в базу
0	Адмиралтейский	32-1-1	сквер б/н на наб. кан. Грибоедова, д. 164	0.16	Закон Санкт-Петербурга от 19.09.2007 № 430-85 ...	муниципальный округ Коломна	NaN
1	Адмиралтейский	32-1-2	сквер б/н на наб. кан. Грибоедова, д. 166	0.09	Закон Санкт-Петербурга от 19.09.2007 № 430-85 ...	муниципальный округ Коломна	NaN
2	Адмиралтейский	32-1-3	сквер б/н на Английском пр., д. 17-19	0.16	Закон Санкт-Петербурга от 19.09.2007 № 430-85 ...	муниципальный округ Коломна	NaN
3	Адмиралтейский	32-1-4	сквер б/н в пер. Макаренко, д. 9	0.21	Закон Санкт-Петербурга от 19.09.2007 № 430-85 ...	муниципальный округ Коломна	NaN
4	Адмиралтейский	32-1-5	сквер б/н на Садовой ул., д. 109	0.07	Закон Санкт-Петербурга от 19.09.2007 № 430-85 ...	муниципальный округ Коломна	NaN

Рисунок 1 – Данные из датасета №1

II. Набор данных “Зеленые насаждения общего пользования местного значения в Санкт-Петербурге” (Датасет №2)

Данный набор данных содержит 2073 строки и 5 столбцов с информацией о расположении зеленых насаждений (см. Рисунок 2).

Описание:

- Район города (Указание района города, где находится объект зеленых насаждений общего пользования)
- Номер по списку (Номер объекта в соответствии с Приложением к Закону Санкт-Петербурга от 12.02.2014 № 69-16 "О внесении изменений в Закон Санкт-Петербурга О зеленых насаждениях общего пользования")
- Наименование (Наименование объекта зеленых насаждений общего пользования и описание его местоположения)
- Площадь, га (Ориентировочная площадь объекта зеленых насаждений общего пользования в соответствии с Законом Санкт-Петербурга от 12.02.2014 № 69-16 "О внесении изменений в Закон Санкт-Петербурга О зеленых насаждениях общего пользования")
- Основание (Основание для включения объекта в перечень зеленых насаждений общего пользования Санкт-Петербурга)

	Район города	Номер по списку	Наименование	Площадь, га	Основание
0	Кировский	5046	сквер б/н на ул.Васи Алексеева, д.16	0.63	Закон Санкт-Петербурга от 19.09.2007 № 430-85 ...
1	Кировский	5058	сквер б/н на Двинской ул., д.16, корп.1	0.42	Закон Санкт-Петербурга от 19.09.2007 № 430-85 ...
2	Кировский	5061	Молвинский сад на Промышленной ул. между д.17 ...	4.45	Закон Санкт-Петербурга от 19.09.2007 № 430-85 ...
3	Кировский	5065	Кировский сквер на Кировской пл.	1.13	Закон Санкт-Петербурга от 19.09.2007 № 430-85 ...
4	Кировский	5066	сад Девятого Января между ул.Маршала Говорова,...	11.30	Закон Санкт-Петербурга от 19.09.2007 № 430-85 ...

Рисунок 2 – Данные из датасета №2

Подготовка данных (Data Preparation)

Цель этапа – подготовить обучающую выборку для использования в моделировании.

I. Набор данных “Зеленые насаждения общего пользования городского значения в Санкт-Петербурге” (Датасет №1)

Как можно заметить, такие столбцы как "Номер территории" (не понадобится в последующем анализе), "Основание" (имеет только 2 уникальных значения и не понадобится в последующем анализе)(см. Рисунок 3) и "Дата внесения в базу" (все значения пустые) не имеют должной ценности, поэтому было принято решение их удалить (см. Рисунок 4).

```
data_local['Основание'].unique()

array(['Закон Санкт-Петербурга от 19.09.2007 № 430-85 «О зеленых насаждениях общего пользования»',
      nan,
      'Закон Санкт-Петербурга от 19.09.2007 № 430-85 «О зеленых насаждениях общего пользования»79-59'],
      dtype=object)
```

Рисунок 3 – Уникальные значения столбца "Основание"

```
data_local = data_local.drop(columns=["Номер территории", "Основание", "Дата внесения в базу", "Муниципальное"]).rename(columns={'Ориентировочная': 'Площадь'})
data_local
```

	Район	Адрес	Площадь
0	Адмиралтейский	сквер б/н на наб. кан. Грибоедова, д. 164	0.1600
1	Адмиралтейский	сквер б/н на наб. кан. Грибоедова, д. 166	0.0900
2	Адмиралтейский	сквер б/н на Английском пр., д. 17-19	0.1600
3	Адмиралтейский	сквер б/н в пер. Макаренко, д. 9	0.2100
4	Адмиралтейский	сквер б/н на Садовой ул., д. 109	0.0700
...
4918	Петроградский	сквер б/н на Лахтинской ул. у д. 34 по ул. Ленина	0.0960
4919	Красносельский	сквер б/н севернее д. 15, корп. 2, по Петергоф...	0.4633
4920	Колпинский	сквер б/н восточнее д. 3 по ул. Максима Горько...	0.3715
4921	Кировский	сквер б/н восточнее д. 15 и д. 17 по ул. Марша...	0.1583
4922	Невский	сквер б/н северо-западнее д. 3, корп. 2, по ул...	0.0800

4923 rows x 3 columns

Рисунок 4 – Датасет №1 с удаленными столбцами

II. Набор данных “Зеленые насаждения общего пользования местного значения в Санкт-Петербурге” (Датасет №2)

Как можно заметить, такие столбцы как "Номер по списку" (не понадобится в последующем анализе) и "Основание" (имеет только 1 уникальное значение) (см. Рисунок 5) не имеют должной ценности, поэтому было принято решение их удалить (см. Рисунок 6).

```
data_urban['Основание'].unique()

array(['Закон Санкт-Петербурга от 19.09.2007 № 430-85 «О зеленых насаждениях общего пользования»'],
      dtype=object)
```

Рисунок 5 – Уникальные значения столбца "Основание"

```
data_urban = data_urban.drop(columns=["Номер по списку", "Основание"]).rename(columns={'Площадь, га': 'Площадь', 'Наименование': 'Адрес', 'Район города': 'Район'})
data_urban
```

	Район	Адрес	Площадь
0	Кировский	сквер б/н на ул.Васи Алексеева, д.16	0.6300
1	Кировский	сквер б/н на Двинской ул., д.16, корп.1	0.4200
2	Кировский	Молвинский сад на Промышленной ул. между д.17 ...	4.4500
3	Кировский	Кировский сквер на Кировской пл.	1.1300
4	Кировский	сад Девятого Января между ул.Маршала Говорова,...	11.3000
...
2068	Центральный	бульвар б/н по наб. Обводного кан. от Кременчу...	0.4860
2069	Центральный	бульвар б/н на Синопской наб. у Большеохтинско...	0.6060
2070	Центральный	бульвар б/н на ул. Черняховского	0.5840
2071	Центральный	Калужский сквер в Калужском пер., д.9	0.4900
2072	Петроградский	сквер б/н северо-западнее д.7 по Петровской косе	5.0455

2073 rows x 3 columns

Рисунок 6 – Датасет №2 с удаленными столбцами

III. Объединение и очистка данных

Объединим Датасет №1 и Датасет №2 для дальнейшего анализа (см. Рисунок 7):


```
data = pd.concat([data_urban, data_local])
print(data.shape)
data.head()
```

(6996, 3)

	Район	Адрес	Площадь
0	Кировский	сквер б/н на ул.Васи Алексеева, д.16	0.63
1	Кировский	сквер б/н на Двинской ул., д.16, корп.1	0.42
2	Кировский	Молвинский сад на Промышленной ул. между д.17 ...	4.45
3	Кировский	Кировский сквер на Кировской пл.	1.13
4	Кировский	сад Девятого Января между ул.Маршала Говорова,...	11.30

Рисунок 7 – Объединили данные

Была создана функция нормализации адреса, а также применили ее к данным (см. Рисунок 8):

```
data['Адрес'] = data['Адрес'].apply(lambda x: clear_urban(x))
data.head(10)
```

	Район	Адрес	Площадь
0	Кировский	санкт петербург улица Васи Алексеева 16	0.63
1	Кировский	санкт петербург Двинская улица 16 корп.1	0.42
2	Кировский	санкт петербург 19	4.45
3	Кировский	санкт петербург Кировский сквер Кировская площ...	1.13
4	Кировский	санкт петербург проспект Стачек	11.30
5	Кировский	санкт петербург Охотничьего переулок	0.71
6	Кировский	санкт петербург улица Корнеева	1.95
7	Кировский	санкт петербург сад Кирьяново проспект Стачек ...	2.76
8	Кировский	санкт петербург Автовская улица	0.90
9	Кировский	санкт петербург улица Новостроек	3.82

Рисунок 8 – Очистили датасет

Моделирование (Modeling)

Для сбора геоданных была написана функция `custom_geocoder`, которая работает с адресами из наших наборов. Данная функция возвращает широту и долготу для конкретного адреса (см. Рисунок 9).

```
def custom_geocoder(address):  
    try:  
        dataframe = geocode(address, provider="nominatim", user_agent='my_request')  
        point = dataframe.geometry.iloc[0]  
        return pd.Series({'Latitude': point.y, 'Longitude': point.x})  
    except:  
        return pd.Series({'Latitude': None, 'Longitude': None})
```

Рисунок 9 – Функция `custom_geocoder`

Применим нашу функцию (Рисунок 10):

```
data[['latitude', 'longitude', 'Полный адрес']] = data['Адрес'].apply(lambda x: custom_geocoder(x))
```

	Район	Адрес	Площадь	latitude	longitude	Полный адрес
0	Кировский	санкт петербург улица Васи Алексеева 16	0.6300	59.879151	30.270631	16, улица Васи Алексеева, Автово, округ Автово...
1	Кировский	санкт петербург Двинская улица 16 корп.1	0.4200	NaN	NaN	None
2	Кировский	санкт петербург 19	4.4500	59.944531	30.491583	Санкт-Петербург, 19, проспект Наставников, окр...
3	Кировский	санкт петербург Кировский сквер Кировская площ...	1.1300	NaN	NaN	None
4	Кировский	санкт петербург проспект Стачек	11.3000	59.859766	30.255499	проспект Стачек, округ Княжево, Санкт-Петербур...
...
4918	Петроградский	санкт петербург 34 улица Ленина	0.0960	60.083230	30.251004	34, улица Ленина, Михайловка, Парголово, Санкт...
4919	Красносельский	санкт петербург 15 Петергофское шоссе	0.4633	59.848354	30.186739	15 к1, Петергофское шоссе, Южно-Приморский окр...
4920	Колпинский	санкт петербург 3 улица Максима Горького	0.3715	59.766997	30.128784	3, улица Максима Горького, Горелово, округ Гор...
4921	Кировский	санкт петербург 17 улица Маршала Говорова	0.1583	59.879954	30.268895	17, улица Маршала Говорова, Автово, округ Авто...
4922	Невский	санкт петербург 3 улица Ворошилова	0.0800	59.931836	30.463749	Продукты, 31 к3, улица Ворошилова, округ Право...

6996 rows x 6 columns

Рисунок 10 – Датасет с `custom_geocoder()`

Для отрисовки карты воспользуемся функциями создания границ районов Санкт-Петербурга. Они берут за основу данные проекта OpenStreetMap, население и общую площадь района. Полученный после преобразований датафрейм можно посмотреть на рисунке 11.

	id_района	Название	Население, чел	Общая площадь, м2	geometry	Количество зеленых насаждений, шт	Площадь зеленых насаждений, %	Площадь зеленых насаждений на человека, м2
1	1114193	Адмиралтейский	158253	13820000.0	MULTIPOLYGON (((30.25023 59.90129, 30.25112 59...	118	18.536881	16.187984
2	1114252	Василеостровский	209733	21470000.0	MULTIPOLYGON (((30.18468 59.95472, 30.18449 59...	409	5.159082	5.281262
3	1114354	Выборгский	533808	115520000.0	MULTIPOLYGON (((30.07276 60.09727, 30.08512 60...	312	5.798543	12.548476
4	1114806	Калининский	546908	40180000.0	MULTIPOLYGON (((30.34405 59.97749, 30.34419 59...	610	11.313123	8.311477

Рисунок 11 – Результат функции отрисовки границ

Далее была построена карта озеленения территорий относительно площади озеленения района (в %):

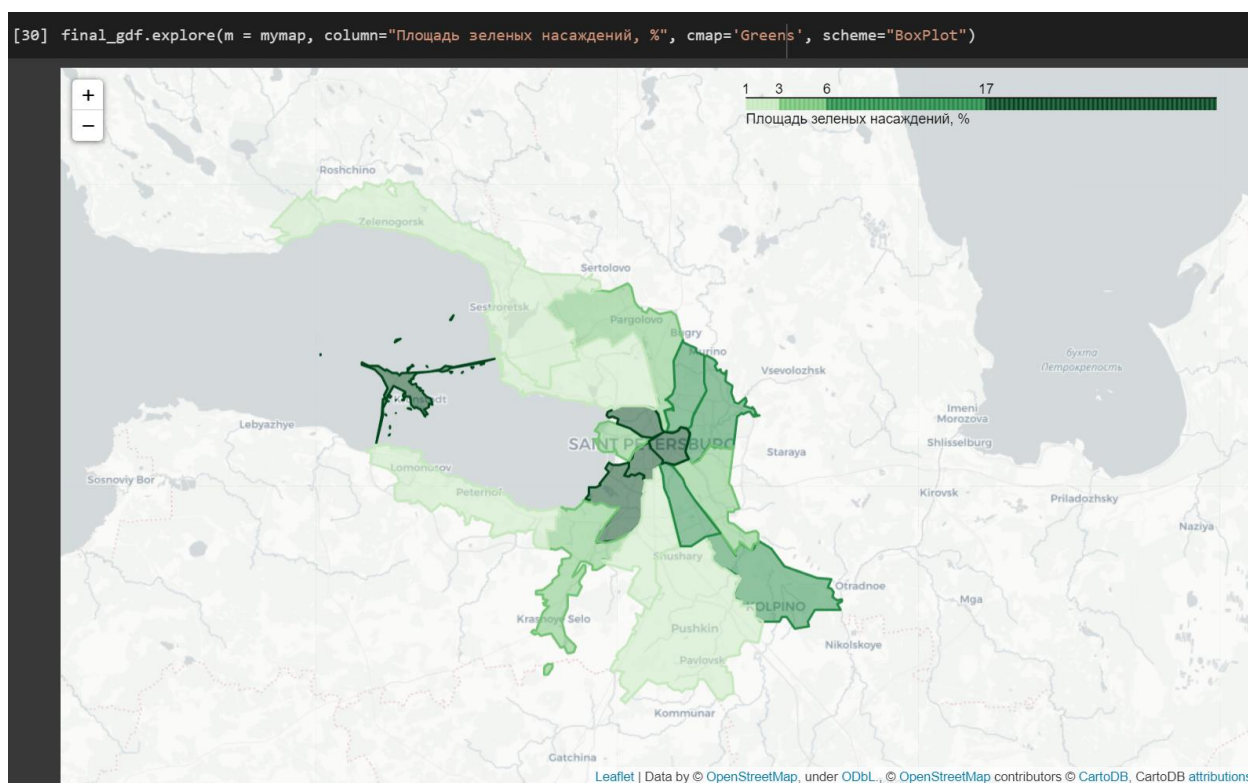


Рисунок 12 – Карта озеленения территорий относительно площади озеленения района (в %)

Как можно заметить, многие районы не так озеленены, как может показаться в жизни. Это связано с тем, что большая часть зеленых территорий в окраинных районах не входит в перечень Зеленых насаждений общего пользования.

Построим график площади озеленения районов (в %):

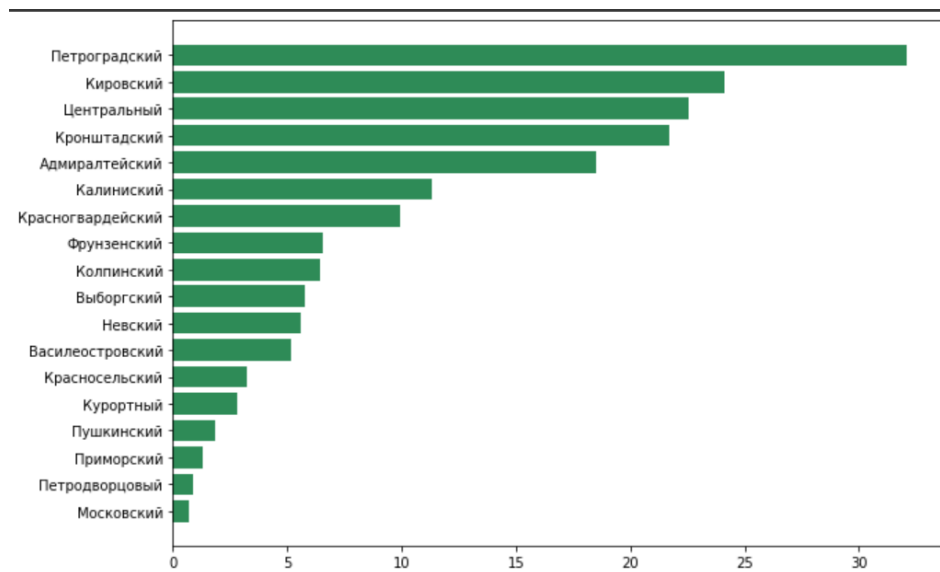


Рисунок 13 – График площади озеленения районов (в %)

Самыми зелеными являются Петроградский, Кировский и Центральный районы. Однако данная метрика не является достаточно показательной, так как не берет во внимание Зеленые насаждения ограниченного пользования и специального назначения.

В таком случае построим карту озеленения районов относительно площади зеленых насаждений на одного жителя (в м^2):

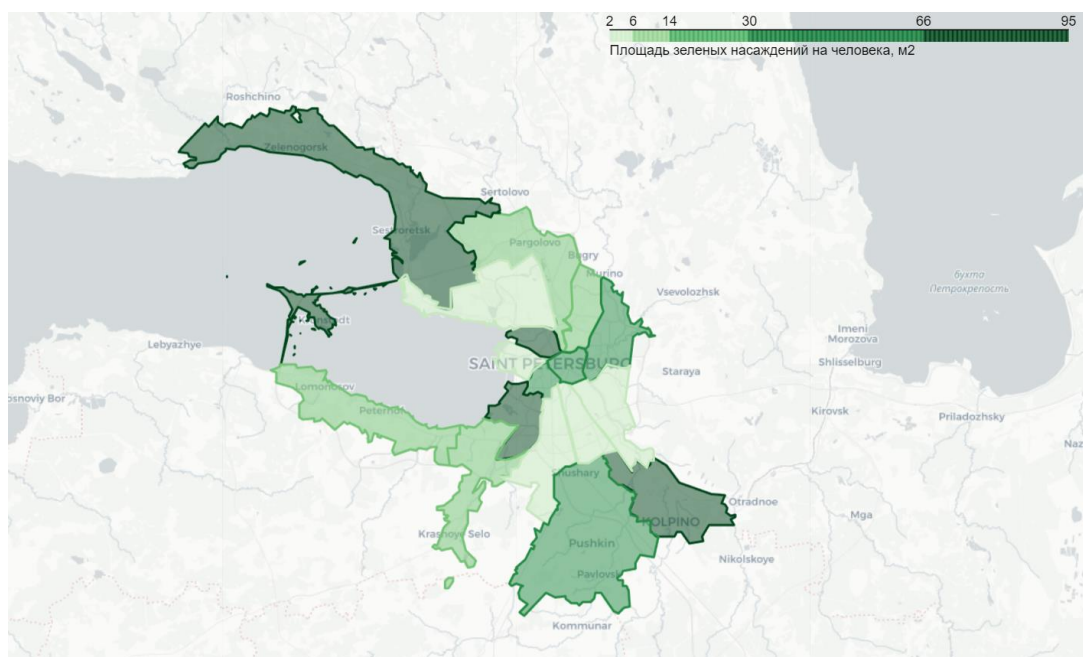


Рисунок 14 – Карта озеленения районов относительно площади зеленых насаждений на одного жителя (в м^2)

Также построим график:

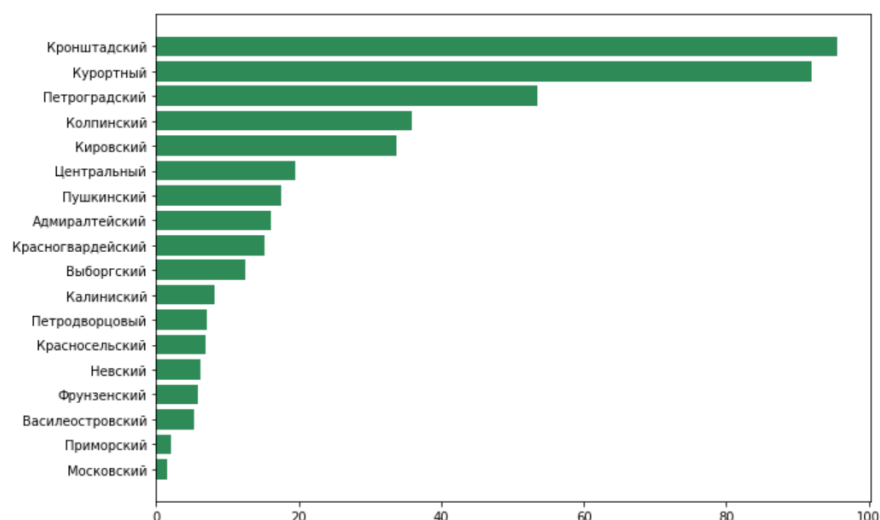


Рисунок 15 – График озеленения районов относительно площади зеленых насаждений на одного жителя (в м²)

Данная метрика является более информативной с той точки зрения, что она представлена в нормах ВОЗ по озеленению городских пространств. По площади зеленых насаждений общего пользования на одного жителя (в м²) в норму (21 м² на жителя) укладывается целых 5 районов: Кронштадтский, Курортный, Петроградский, Колпинский и Кировский.

Для наглядности построим карту зеленых насаждений Санкт-Петербурга:

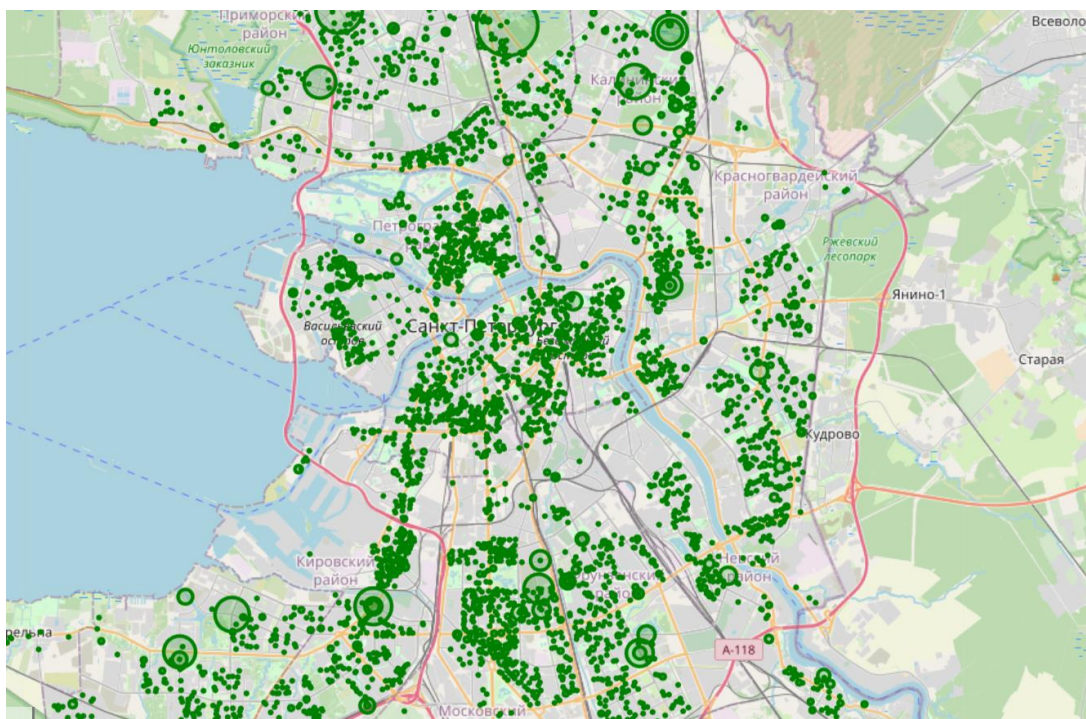


Рисунок 16 – Карта зеленых насаждений Санкт-Петербурга

На карте отчетливо видно знаменитый «Серый пояс Петербурга», также выделяются промышленные районы и кладбища. Размер кругов зависит от площади территорий и рассчитывается по формуле:

$$R = \sqrt{\frac{S}{\pi}}$$

Все карты интерактивные. Их можно посмотреть в Google Colab, доступном в Приложении 1

Заключение

Цели курсового проекта были достигнуты.

В ходе работы была проведена предобработка данных. Были найдены географические данные зеленых насаждений с помощью предобработанных адресов.

На основе анализа были сделаны следующие выводы:

1. Самыми озелененными районами являются Петроградский (благодаря Каменному, Крестовскому и Петровскому о-вам), Кронштадтский и Кировский
2. Для более четкого определения степени озеленения районов нужно также брать во внимание зеленые насаждения ограниченного пользования и специального назначения
3. Норму, установленную ВОЗ, по площади зеленых насаждений общего пользования на одного жителя (21 м² для крупных городов) соблюдают 5 районов: Кронштадтский, Курортный, Петроградский, Колпинский и Кировский

После проделанных исследований, действительно можно утверждать, что Санкт-Петербург не является достаточно озелененным городом. Проблема с озеленением стоит остро особенно в районах промышленной близости, а также в районах малоиспользуемых культурных объектах, таких как стадионы, площади и пустыри.

Нами были изучены методы создания карт с помощью языка Python и его библиотек. Была изучена информация, касаемая типов озеленения, его законного основания и метрики оценки озеленения в городах.

Список Источников

1. Зеленые насаждения общего пользования местного значения в Санкт-Петербурге (Архивный) (Версия №1 от 22.04.2021) [Электронный ресурс]. 2022. URL: https://classif.gov.spb.ru/irsi/7832000076-znop-mestnogo-znacheniya/structure_version/471/?page=8&per_page=10.
2. Зеленые насаждения общего пользования городского значения в Санкт-Петербурге (Архивный) (Версия №1 от 22.04.2021) [Электронный ресурс]. 2022. URL: https://classif.gov.spb.ru/irsi/7832000076-znop-gorodskogo-znacheniya/structure_version/472/?page=6&per_page=10
3. Pandas Documentation: [Электронный ресурс]. 2022. URL: <https://pandas.pydata.org/docs/index.html>
4. GeoPandas Documentation: [Электронный ресурс]. 2022. URL: <https://geopandas.org/en/stable/index.html>
5. Геоаналитика с помощью Python и открытых данных: пошаговое руководство // Хабр: [Электронный ресурс]. 2021. URL: <https://habr.com/ru/post/579838/>
6. CRISP-DM: проверенная методология для Data Scientist-ов \\\ Хабр [Электронный ресурс]. 2017. URL: <https://habr.com/ru/company/lanit/blog/328858/>
7. Проблема озеленения Петербурга: [Электронный ресурс]. 2021. URL: <https://vk.com/@snovazeleniypeterburg-problema-ozeleneniya-peterburga>
8. Зеленый Петербург: [Электронный ресурс]. 2021. URL: <https://www.ozelenimvse.ru/>
9. Отрисовка карты с использованием Geopandas: [Google Colab]. 2022. URL: <https://colab.research.google.com/drive/1lnw2Pkg-7ebvQLmrVMkxcucAF8NGXRsh?usp=sharing>
10. Закон Санкт-Петербурга О зеленых насаждениях общего пользования. [Электронный ресурс]. 2022. URL: <https://docs.cntd.ru/document/8458668>

Приложение 1

Google collab с кодом для исследования:

<https://colab.research.google.com/drive/1DDUrIkrdXtSTjiILt2UMESPssWWEyv1O?usp=sharing#scrollTo=1HmCeIDbjQrn>