



УНИВЕРСИТЕТ ИТМО

Курсовая работа

“Исследование плотности установки камер уличного видеонаблюдения на объектах недвижимого фонда, предоставляемых субъектам малого и среднего предпринимательства в Санкт-Петербурге в соответствии с CRISP-DM”

Колганова Анастасия, К33422

Орлова Любовь, К33422

# Business Understanding

**МЫ** - компания предоставляющая услуги по установки наружных камер видеонаблюдения для малого и среднего бизнесов.

**НАША ЦЕЛЬ:** спрогнозировать в каком районе Санкт-Петербурга предлагаемые нами услуги будут наиболее востребованными. В каких районах нужно размещать рекламу? Какова максимально возможная прибыль с каждого из районов города?

**КРИТЕРИЙ УСПЕХА:** ROI (Return on Investment) - коэффициент окупаемости или возврата всех инвестиций. Рассчитывается ROI по формуле:

$$ROI = \frac{\text{ДОХОД} - \text{ЗАТРАТЫ}}{\text{ЗАТРАТЫ}} \times 100\%$$



# Situation normal

## РЕСУРСЫ:

- специалисты по анализу данных Любовь и Анастасия
- два датасета с портала “Система классификаторов СПб”
- Google Collab

## ПРЕДПОЛОЖЕНИЕ:

Совместить датасеты по столбцу адреса;

=> районы, в которых объектов недвижимого фонда > установленных камер уличного видеонаблюдения

## ТРЕБОВАНИЯ:

> 1000 строк в объединенном датасете

## РИСКИ:

- нестандартизированный адрес
- недостаточная выборка
- неправильная стратегия моделирования

# ROI

Наша компания планирует оснащать районы линейно, за год полностью покрыв весь спрос.

За установку камер мы получаем:  $A * 4000$

За обслуживание камер мы получаем:  $A * 280 * \sum_{i=1}^{12} i$

Траты на рекламу за год =  $A * 500 * \sum_{i=1}^{12} i$

$$ROI = \frac{(A * 4000 + A * 280 * 6) - A * 500 * 6}{A * 500 * 6} * 100\% = 89\%$$

- окупаемость за первый год. После этого у нас только чистая прибыль с обслуживания камер =  $(280 * A)$  в месяц.

# Data Understanding: exploration №1

Распределение объектов с установленными видеокамерами по районам.

```
[4] df_cameras['district'].value_counts(dropna=False)[:5]
```

Калининский	1146
Центральный	800
Петроградский	789
Приморский	737
Невский	671

Name: district, dtype: int64

# Data Understanding: exploration №2

Распределение объектов недвижимого фонда по районам.

```
[14] df_business['district'].value_counts(dropna=False)[:5]
```

```

Центральный      1028
Адмиралтейский   427
Петроградский    244
Невский          169
Московский       152
Name: district, dtype: int64

```

Суммарная площадь объектов недвижимого фонда по районам.

```
[15] df_business.groupby('district')['area'].agg('sum')
```

```

district
Адмиралтейский      60829.9
Василеостровский    13297.6
Выборгский           7831.8
Калининский          5646.2
Кировский            13092.8
Колпинский           2521.1
Красногвардейский    17005.4
Красносельский       5147.3
Кронштадтский        789.6
Курортный            2333.6
Московский           19932.2
Невский              28798.0
Петроградский        30228.0
Петродворцовый       4846.8
Приморский           3060.0
Пушкинский           5352.5
Фрунзенский          12938.3
Центральный          164305.5
Name: area, dtype: float64

```

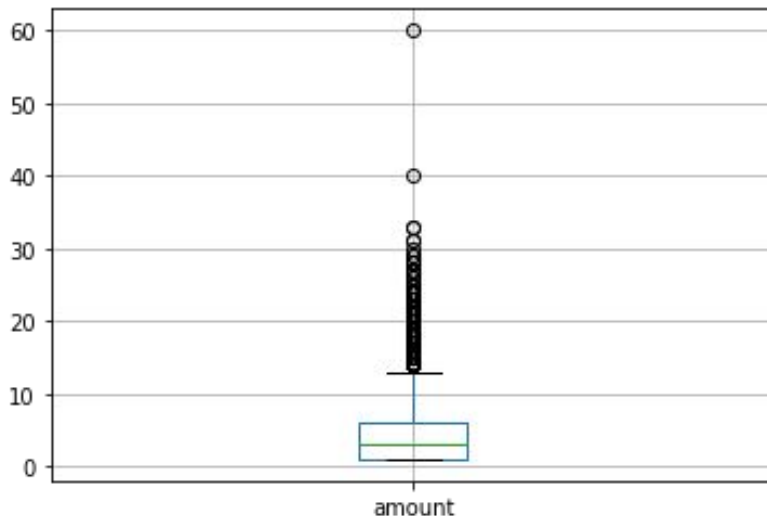
# Data Understanding: гипотезы

Так как в Центральном и Адмиралтейском районах суммарная площадь объектов недвижимого фонда максимальная, можно сделать предположение, что из всех объектов, предоставляемых малому и среднему предпринимательству, в этих районах максимальное их количество должно иметь уже установленные видеокамеры. Можно предположить, что в этих районах услуги по установке и обслуживанию камер уличного видеонаблюдения будут не актуальны.

В Петроградском, Невском, Московском, Кировском и Василеостровском районах сравнительно много объектов недвижимого фонда, при этом лишь Петроградский и Невский район попали в топ-5 по количеству установленных камер. Вероятно, наши услуги будут пользоваться спросом в Московском, Кировском и Василеостровском районах.

# Data Understanding: Verify data quality №1

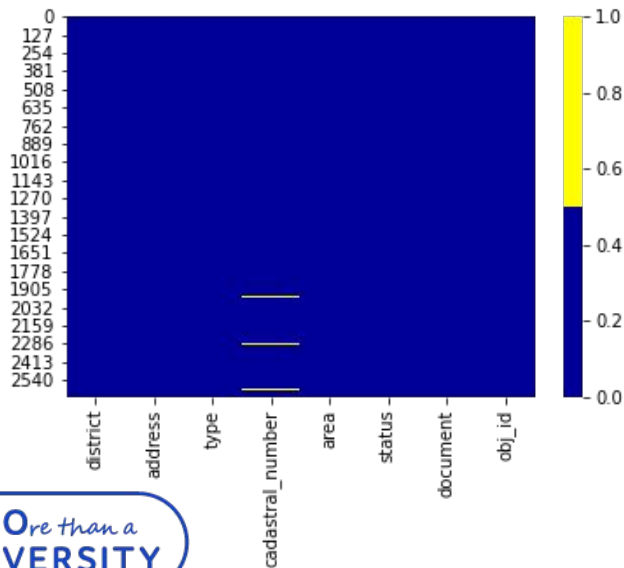
В колонке с количеством установленных камер обнаружен выброс.





# Data Understanding: Verify data quality №2

Пропущенные данные в колонке  
cadastral\_number.



Обнаружены дубликаты.

```
[22] # отбрасываем неуникальные строки
      df_bus_dedupped = df_business.drop('obj_id', axis=1).drop_duplicates()

      # сравниваем формы старого и нового наборов
      print(df_business.shape)
      print(df_bus_dedupped.shape)

      (2655, 8)
      (2653, 7)
```

# Data Preparation: output №1

`training_data` - адреса объектов недвижимого фонда, находящихся в собственности Санкт-Петербурга, предназначенных для предоставления субъектам малого и среднего предпринимательства в Санкт-Петербурге, по которым уже установлены уличные камеры видеонаблюдения.

```
[54] training_data.head()
```

	district	index	city	street	house	type	area	amount
0	Адмиралтейский	190000	санкт-петербург	набережная канала адмиралтейского	7	помещение	143.2	2
1	Адмиралтейский	190068	санкт-петербург	проспект английский	40	помещение	335.3	4
3	Адмиралтейский	190000	санкт-петербург	улица большая морская	40	помещение	3420.9	1
4	Адмиралтейский	191180	санкт-петербург	переулок большой казачий	1/61	помещение	266.8	3
5	Адмиралтейский	191180	санкт-петербург	переулок большой казачий	1/61	помещение	204.3	3

# Data Preparation: output №2

`predict_data` - адреса объектов недвижимого фонда, находящихся в собственности Санкт-Петербурга, предназначенных для предоставления субъектам малого и среднего предпринимательства в Санкт-Петербурге, по которым уличные камеры видеонаблюдения не установлены.

```
[55] predict_data.head()
```

	district	index	city	street	house	type	area
2	Центральный	191025	санкт-петербург	улица маяковского	2/94	помещение	100.8
277	Невский	192029	санкт-петербург	улица седова	37	помещение	86.7
345	Петроградский	197198	санкт-петербург	улица зверинская	18	помещение	39.0
348	Петроградский	197198	санкт-петербург	улица шамшева	5/50	помещение	94.8
349	Петроградский	197110	санкт-петербург	улица большая зеленина	22	помещение	44.6

# Моделирование

Задачи:

- с помощью полученного совмещенного датасета камер и зданий под аренду сделать предсказание относительно максимального количества камер в районах;
- рассчитать возможную прибыль от установки видеокамер.

В качестве "эталонных" были взяты Адмиралтейский и Центральный районы. В этих двух районах, благодаря визуализации, хорошо видно, что плотность покрытия сдаваемой площади камерами значительно больше.

# Моделирование: рейтинг прибыльных районов

## Центральный

Максимальная спрос на камеры в районе (шт.): 426.0

Максимальная прибыль от установки (₽): 1704000.0

Максимальная прибыль от обслуживания (₽/мес.): 119280.0

## Невский

Максимальная спрос на камеры в районе (шт.): 146.0

Максимальная прибыль от установки (₽): 584000.0

Максимальная прибыль от обслуживания (₽/мес.): 40880.0

## Василеостровский

Максимальная спрос на камеры в районе (шт.): 86.0

Максимальная прибыль от установки (₽): 344000.0

Максимальная прибыль от обслуживания (₽/мес.): 24080.0

## Петродворцовый

Максимальная спрос на камеры в районе (шт.): 43.0

Максимальная прибыль от установки (₽): 172000.0

Максимальная прибыль от обслуживания (₽/мес.): 12040.0

## Пушкинский

Максимальная спрос на камеры в районе (шт.): 33.0

Максимальная прибыль от установки (₽): 132000.0

Максимальная прибыль от обслуживания (₽/мес.): 9240.0

## Кронштадтский

Максимальная спрос на камеры в районе (шт.): 8.0

Максимальная прибыль от установки (₽): 32000.0

Максимальная прибыль от обслуживания (₽/мес.): 2240.0

## Курортный

Максимальная спрос на камеры в районе (шт.): 30.0

Максимальная прибыль от установки (₽): 120000.0

Максимальная прибыль от обслуживания (₽/мес.): 8400.0

## Колпинский

Максимальная спрос на камеры в районе (шт.): 8.0

Максимальная прибыль от установки (₽): 32000.0

Максимальная прибыль от обслуживания (₽/мес.): 2240.0

# Моделирование

## Результат

Таким образом получаем топ 3 потенциально наиболее доходных районов для нашего бизнеса (без учета затрат):

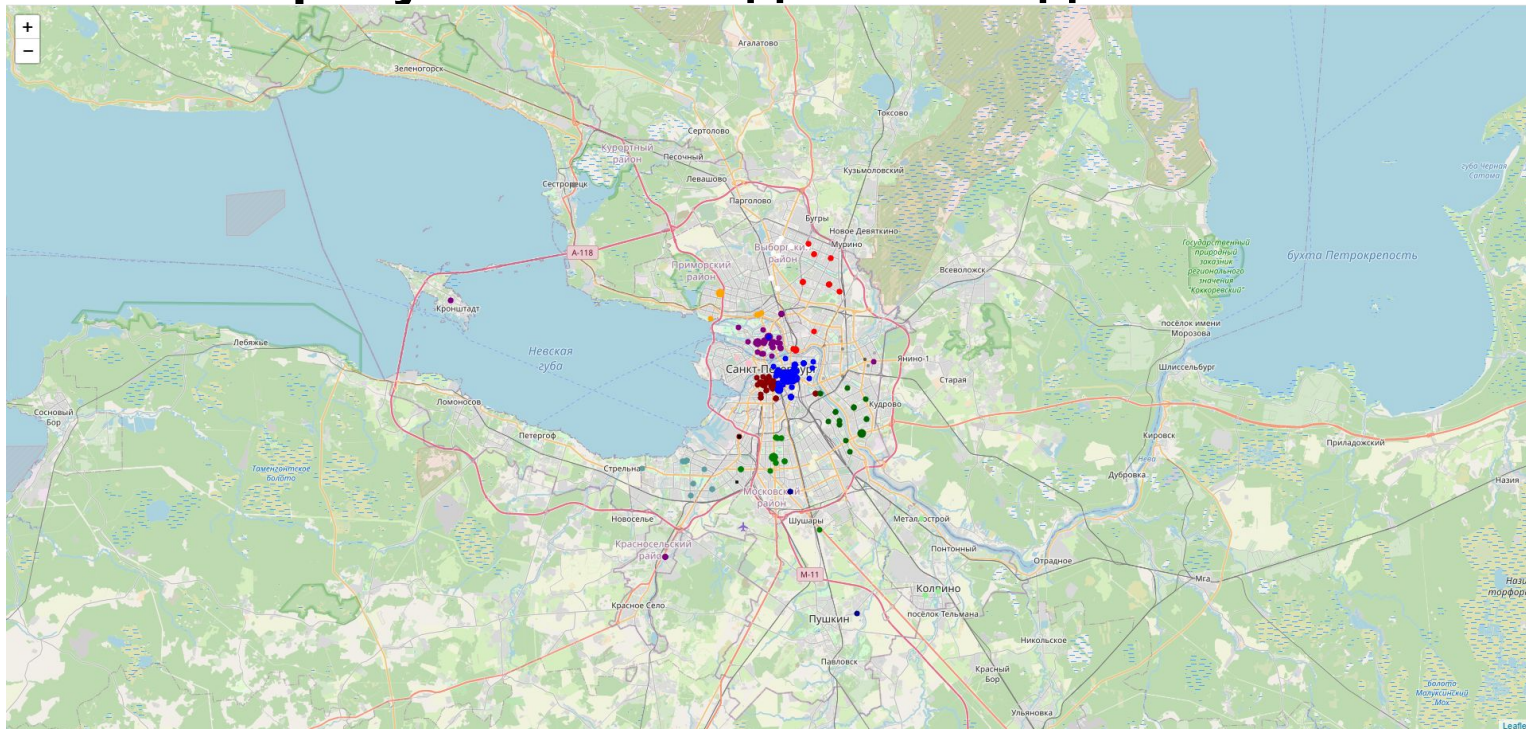
**Центральный** (2'419'680₽/год с установкой)

**Василеостровский** (829'280/год с установкой)

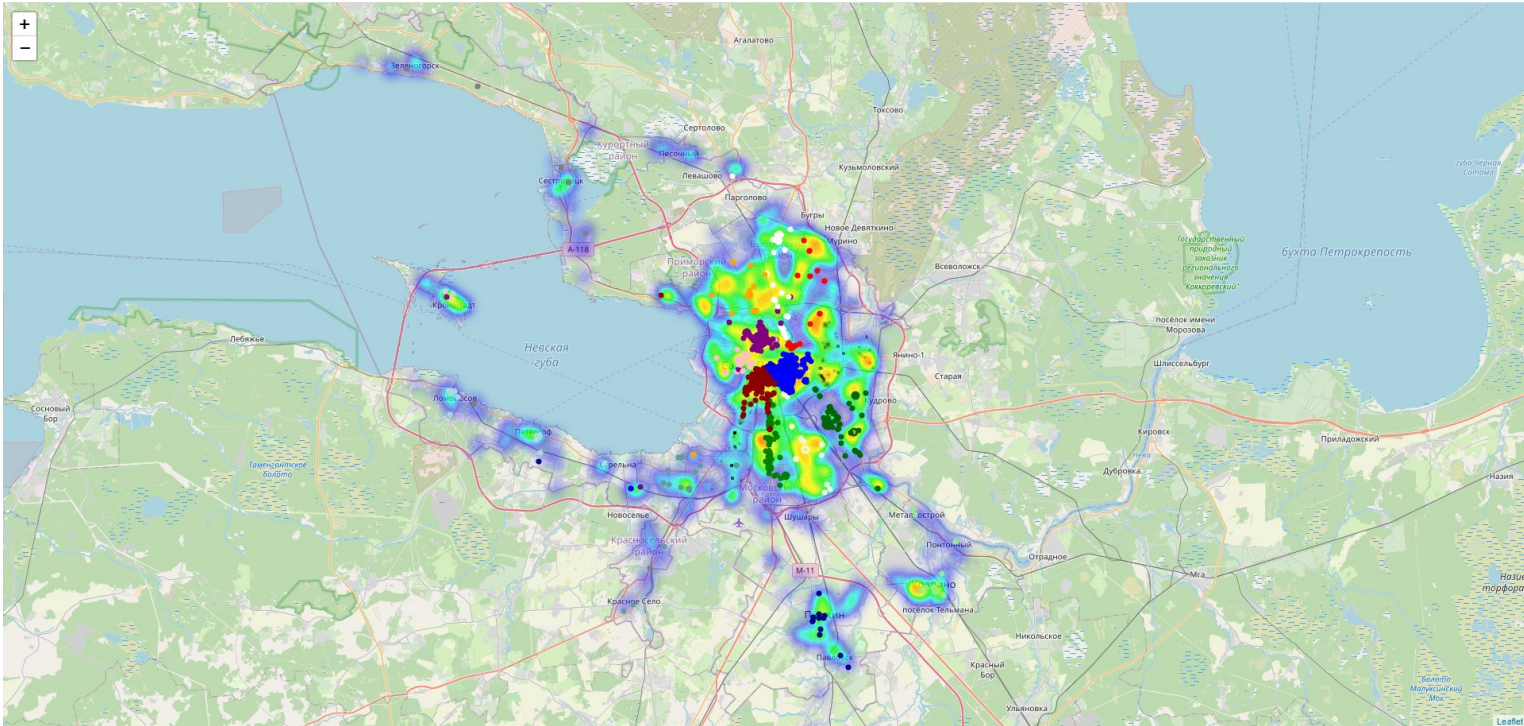
**Петродворцовый** (488'480/год с установкой)



# Визуализация: объекты недвижимого фонда, где уже установлены камеры уличного видеонаблюдения



**Визуализация: объекты недвижимого фонда, где камеры не установлены, теплокарта по установленным камерам.**





# Визуализация

Наибольшая концентрация объектов с уже установленными камерами достигается в Центральном и Адмиралтейском районах города. Именно эти два района и можно считать “эталонными” по плотности камерного покрытия сдаваемой площади.

Примечательно, что в Невском районе расположено много объектов под сдачу (темно-зеленые точки), но камер настолько мало, что на этом участке теплокарта прозрачная. Вероятно, в этом районе услуги по установке камер уличного видеонаблюдения будут наиболее востребованы.

# Выводы

Наша гипотеза о том, что в Центральном и Адмиралтейском районе плотность установленных камер по адресам объектов недвижимого фонда, предоставляемых малому и среднему бизнесу, наибольшая, подтвердилась. Тем не менее, предположение о том, что в Центральном районе услуги по установке камер не являются актуальными, оказалось неверным. Согласно результатам моделирования **Центральный район - потенциально наиболее доходный район из всех для нашего бизнеса.**

Гипотеза о спросе на наши услуги в Московском, Кировском и Василеостровском районах оказалась правдива лишь частично. **Василеостровский район занял второе место в списке перспективных районов**, в то время как Московский и Кировский вовсе не учитывались в анализе в связи с большей плотностью покрытия видеонаблюдением, что вероятно связано с меньшей доступной под аренду площадью.

# Ссылки на Google Collab

Второй и третий этап CRISP-DM

Моделирование и визуализация

# Спасибо за внимание!

[www.ifmo.ru](http://www.ifmo.ru)

ITsMO *re than a*  
UNIVERSITY