



Оптимальный выбор местоположения нового офиса банка генетическим алгоритмом

**ХАКАТОН
СПБГУ X Альфа-Банк**

г. Санкт-Петербург
2022

Навигаторы 2000

Кузнецова Дарья
Латышенко Ульяна
Бевз Роман
Гордюшин Денис

Подготовка исходных данных

- ☐ Эталонные факторы рассматривались в пределах одного города – Санкт-Петербурга.
- ☐ Очищаем датафреймы от дубликатов по номеру **64-bit H3 index** гексогона.
- ☐ Определение. Район - сумма всех гексагонов, центр которых лежит внутри изохроны.

Принцип отбора параметров

- ☐ **Плотность населения** в районе
- ☐ **Количество банкоматов** (отделений) Альфа-банка в районе
- ☐ **Количество банков:** банков — конкурентов в районе
- ☐ **Транспортная доступность:** остановок общественного транспорта внутри района

Целевой параметр и ограничения

- ❑ В рамках упрощения модели взяты первые 20 банкоматов с максимальным количеством транзакций АБ.
- ❑ Нормировали данные параметров.
- ❑ Взяты значения средних и моды.

Ограничения – категория банка

Название	Расшифровка
Общий доступ	category1
ЗП	category2
Отделение	category3
Самоинкассация	category4

Объясняющие факторы

- Взяли 20 наибольших значений для таргета и эталонные (средние и мода) по городу Санкт-Петербургу

Численность населения в районе	Число банкоматов Альфа-банка	Число отделений банков конкурентов	Число остановок общественного транспорта в районе
0.53809037	0.75	1	0.83673469
0.46977648	0.875	0.90410959	0.7755102
0.53048637	0.375	0.35616438	0.42857143
0.61041277	0.25	0.21917808	0.46938776
0.54838861	0.875	0.10958904	0.26530612
0.60582222	0.5	0.2739726	0.24489796
...

Проверка данных на предположения об автокорреляции и мультиколлинеарности

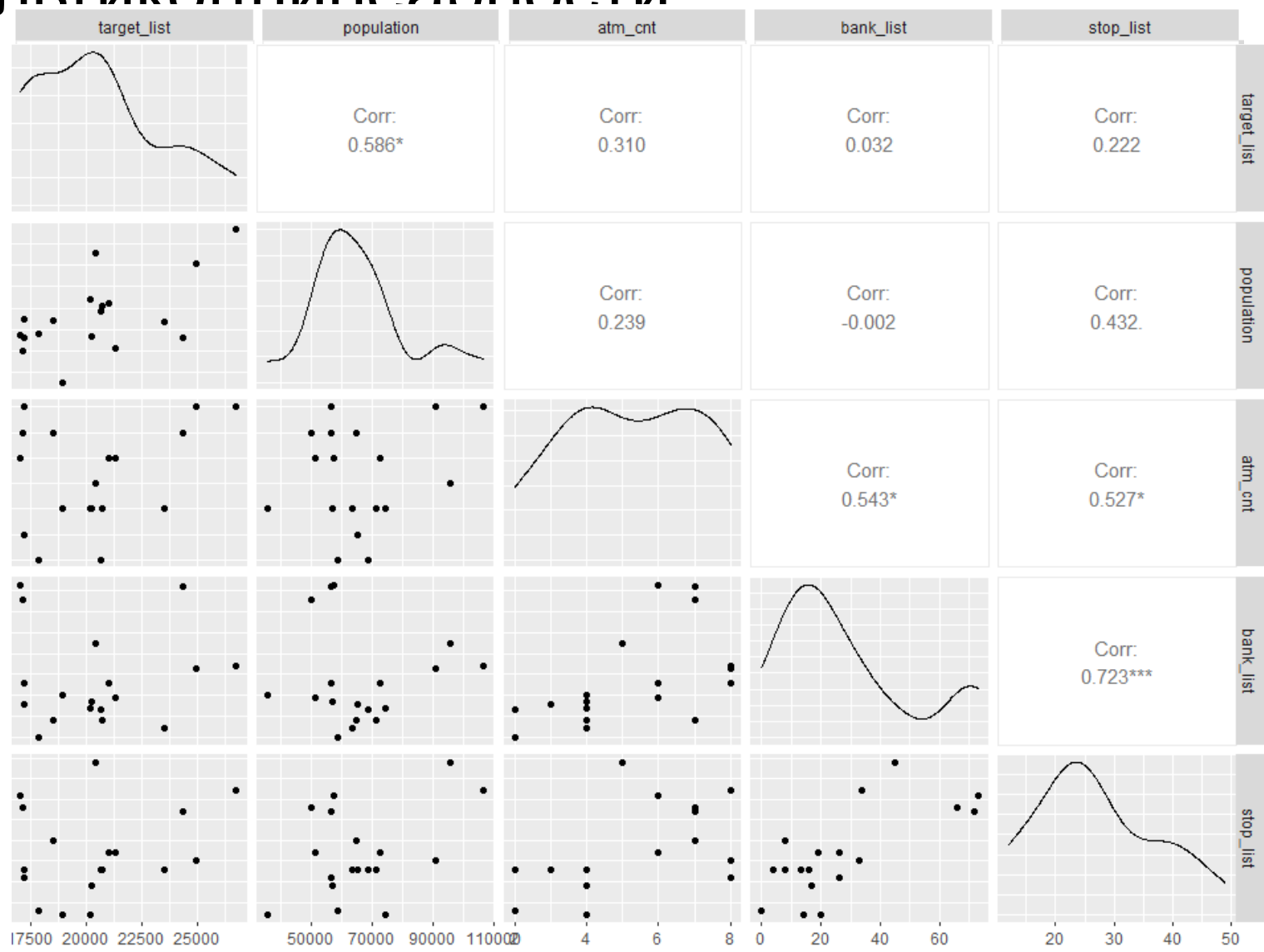
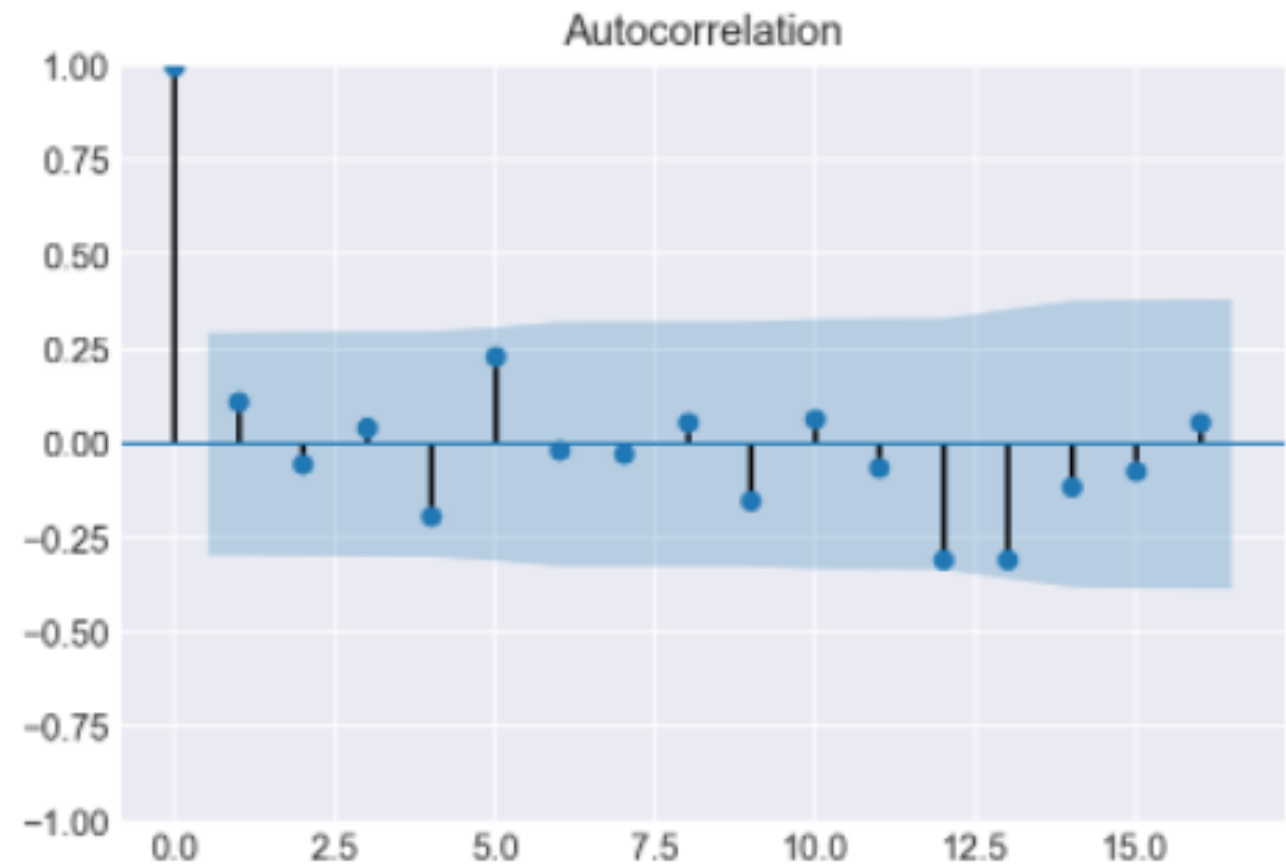


Диаграмма парной корреляции параметров

Автокорреляция (таргет)



Мультиколлинеарность

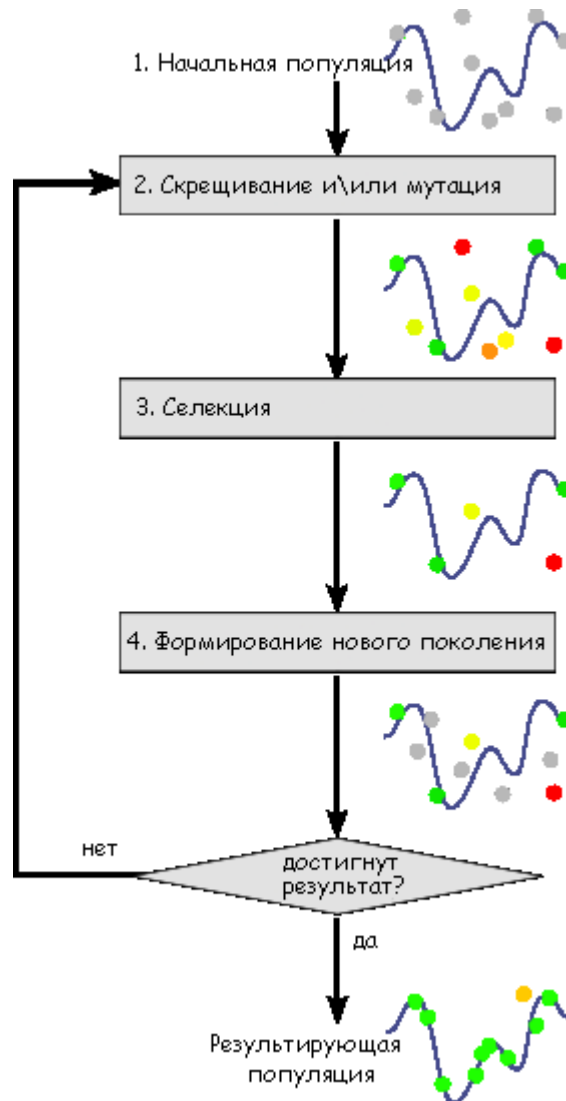
Обнаружение мультиколлинеарности - наличие линейной зависимости между объясняющими переменными (факторами) регрессионной модели. При этом различают полную коллинеарность, которая означает наличие функциональной (тождественной) линейной зависимости и частичную или просто мультиколлинеарность — наличие сильной корреляции между факторами.

```
from statsmodels.stats.outliers_influence import variance_inflation_factor

vif = [variance_inflation_factor(X_constant, i) for i in range(X_constant.shape[1])]
pd.DataFrame({'vif': vif[1:]}, index=df.columns[1:]).T
```

	population_list	atm_cnt	bank_list	stop_list
vif	1.396799	1.372072	1.904529	1.958901

Генетический алгоритм



это эвристический алгоритм поиска, используемый для решения задач оптимизации и моделирования путём случайного подбора, комбинирования и вариации искоемых параметров с использованием механизмов, аналогичных естественному отбору в природе.

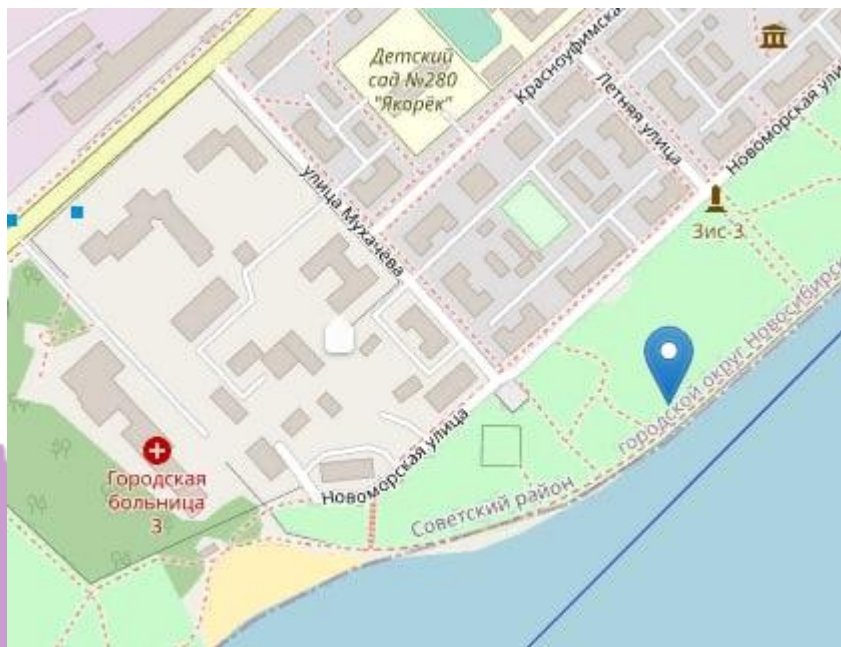
Результаты вычислений

```
category_id=['category1', 'category2','category3','category4']  
coordinat=[]  
for i in category_id:  
    coordinat.append(location(i,'Новосибирск'))  
for i in range(4):  
    print("Координаты банкомата " , category_id[i] , " – ",coordinat[i])
```

Координаты банкомата	category1	–	(54.84847843465986, 82.97390668520336)
Координаты банкомата	category2	–	(54.8171302314008, 83.09541006946871)
Координаты банкомата	category3	–	(54.86157327583376, 82.97621963139348)
Координаты банкомата	category4	–	(54.84112661283599, 83.10007774967546)

Результаты вычислений

category1



category2



Результаты вычислений

category3



category4

