

# Clustering Optimization for Carsharing Fuelings

---

Бородин Матвей

Маргарита Ковалева

Клышников Михаил

Студкемп  
по математике в ИИ

Яндекс Образование

01

# Представление команды

# Команда

Студкемп  
по математике в ИИ



Михаил  
Клышников

Пару слов, если нужно



ЧатГПТ  
Опэнэйаевич

Пару слов, если нужно



Матвей  
Бородин

Пару слов, если нужно



Маргарита  
Ковалева

Руководитель

02

**Цель:** Оптимизация  
заправки  
каршерингов

# Задачи

Студкемп  
по математике в ИИ

01

Генерация синтетических данных:  
расположение и заправленность  
машин

02

Анализ данного датасета заправок  
и стоимости бензина

03

Получение глупых бейзлайнов:  
ближайшие и самые дешевые  
заправки

04

Алгоритмы кластеризации машин  
и поиск оптимальных заправок

05

Оптимизация прохода бензовоза  
по кластеру

06

Подсчет итоговых метрик и  
сравнение с бэйзлайнами

03

данные

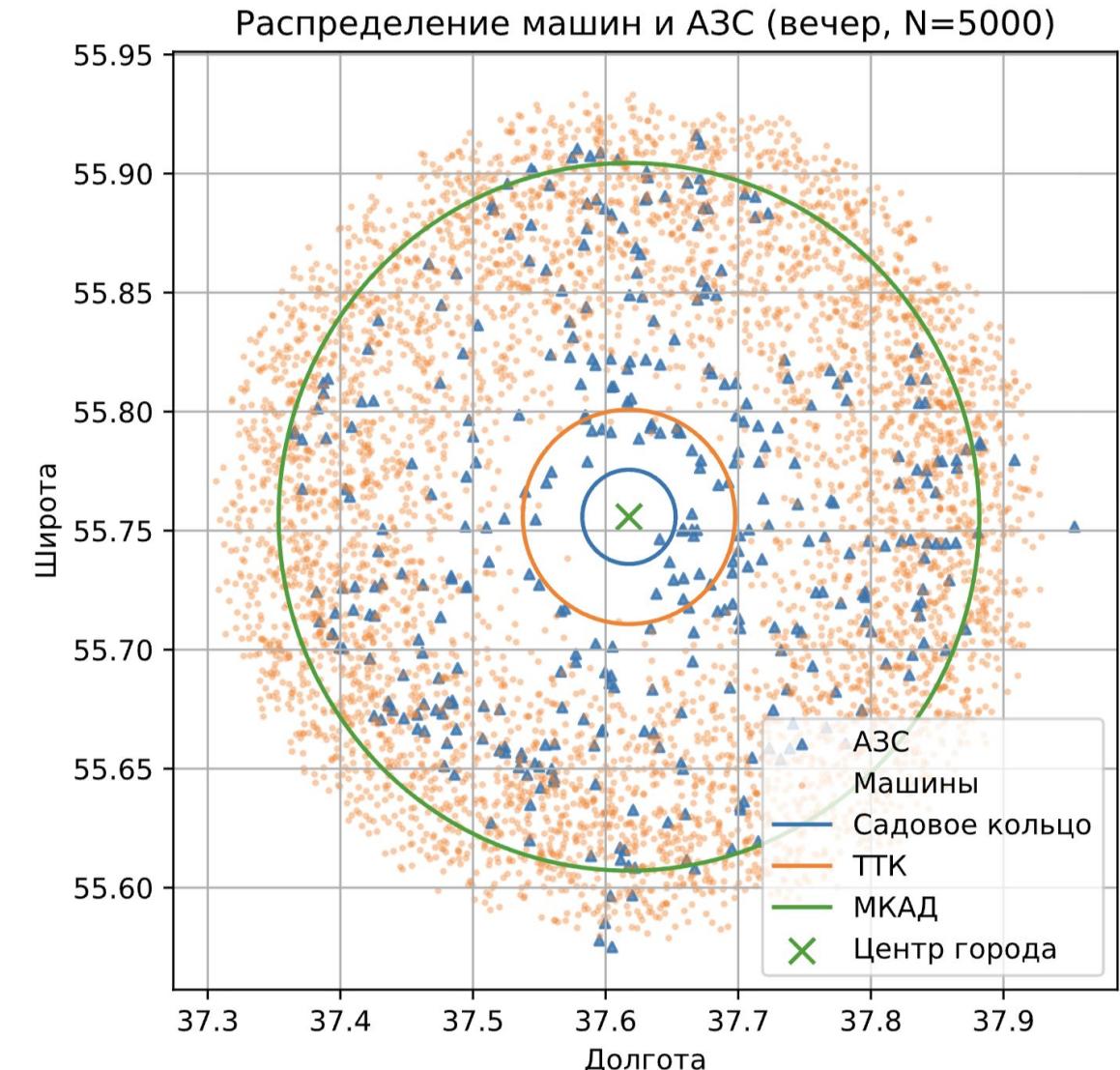
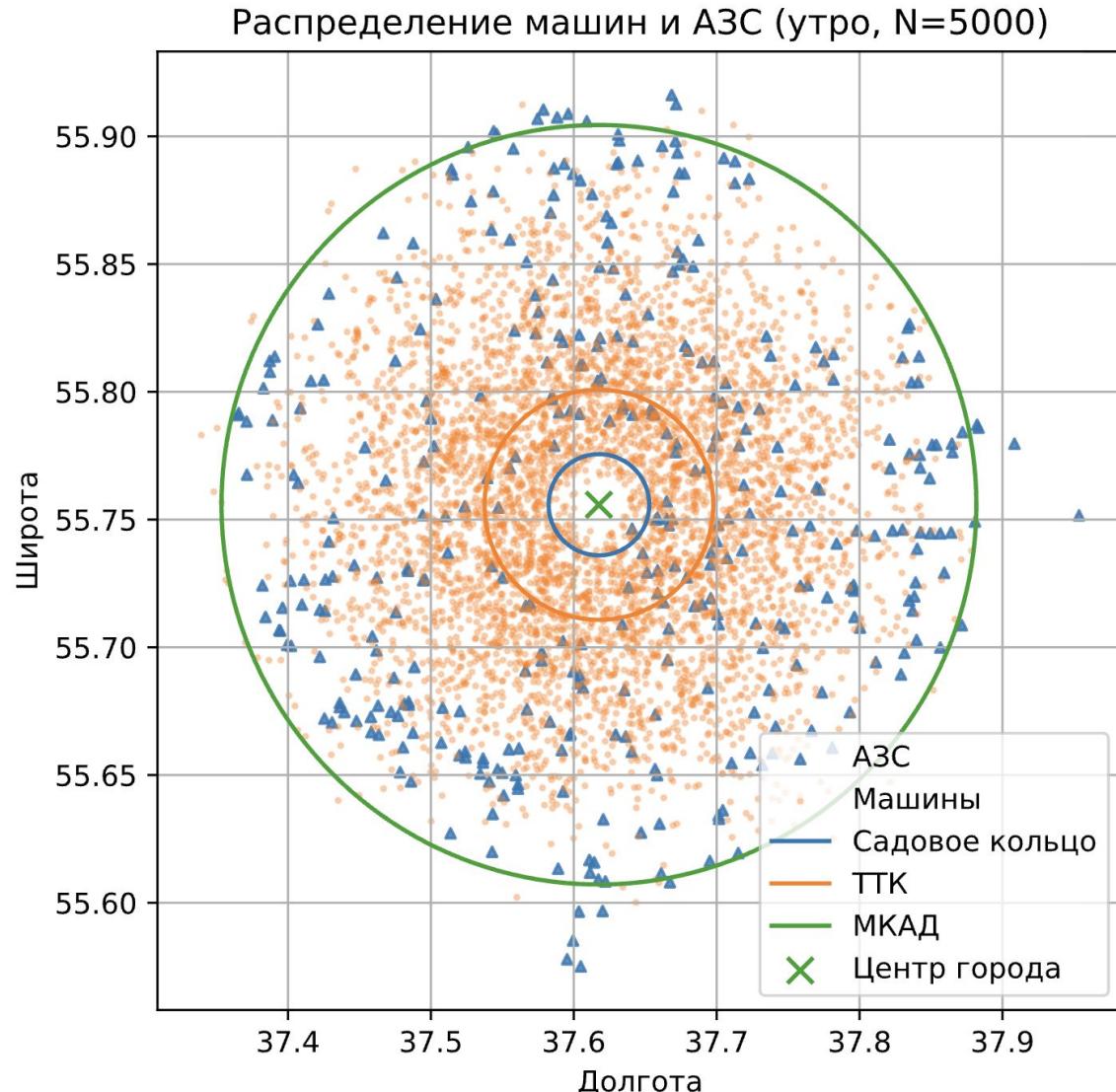
# Вопрос. Где брать данные?

Студкемп  
по математике в ИИ



# Распределение по времени суток

Студкемп  
по математике в ИИ

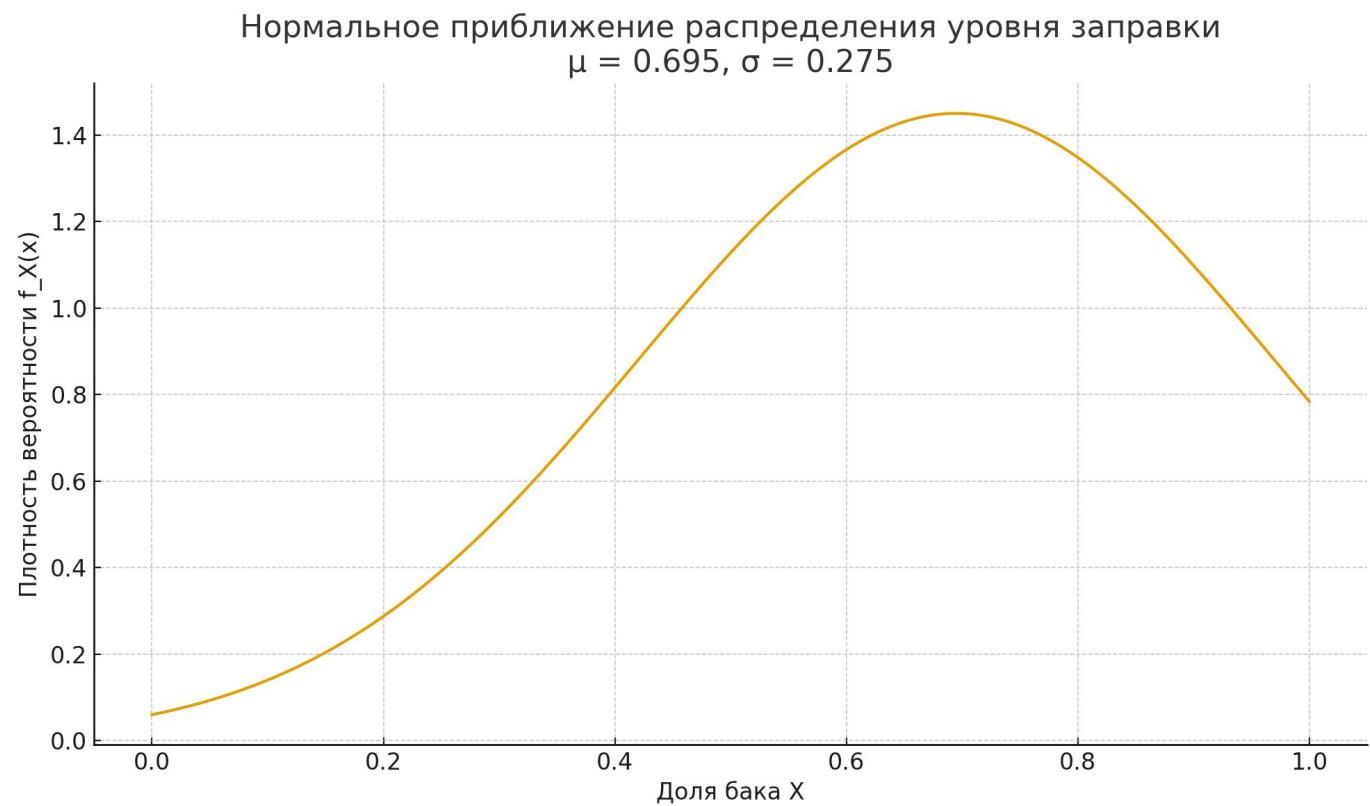


# Распределение заправленности

Студкемп  
по математике в ИИ

- Менее 10% бака - 1% парка
- 10-20% бака - 3% парка
- 20-50% бака - 20% парка
- Остальные заправки не требуют

$E[\text{затраты на топливо}] \approx 2,45$  млн руб. в идеальных условиях



# Минимальные затраты и метрика

Студкемп  
по математике в ИИ

**2,45млн**

Необходимая минимальную  
сумма для заправки машин  
при нормальном распределении  
в идеальных условиях

**±100к**

Дополнительные затраты на  
объезд бензовозов и простой  
машин

04

Бейзлайны

# Первый глупый бейзлайн: относить машины к самым ближайшим заправкам

Давайте отнесем каждую машину к самой  
ближайшей заправке, тем самым  
минимизируем объезд бензовоза.  
Посчитаем итоговую метрику ранее  
описанным способом.

2'752'812

Итоговая средняя стоимость заправить автопарк из  
5000 машин (в рублях)

# Второй глупый бейзлайн: выбирать только самые дешевые заправки

Выберем заранее набор самых дешевых заправок, и будем использовать только их. Основная проблема - большой суммарный пробег бензовозов.

2'899'304

Итоговая средняя стоимость заправить автопарк из 5000 машин (в рублях)

05

Алгоритмы

# Алгоритм на основе метрики Utility + Байесовская оптимизация

$$\text{score}_{sj} = \frac{D_j}{P_s} f_{\text{dist}}(d_{sj}) \quad \text{Cost}_{sj} = P_s \cdot D_j + c_{\text{dist}} \cdot d_{sj}$$

$$\text{Utility}_{sj} = \tilde{w}_{\text{score}} \cdot \text{score\_norm}_{sj} + \tilde{w}_{\text{cost}} \cdot (1 - \text{cost\_norm}_{sj})$$

Составляем матрицу машин-заправок со скорами и выбираем заправку с наибольшим Utility.

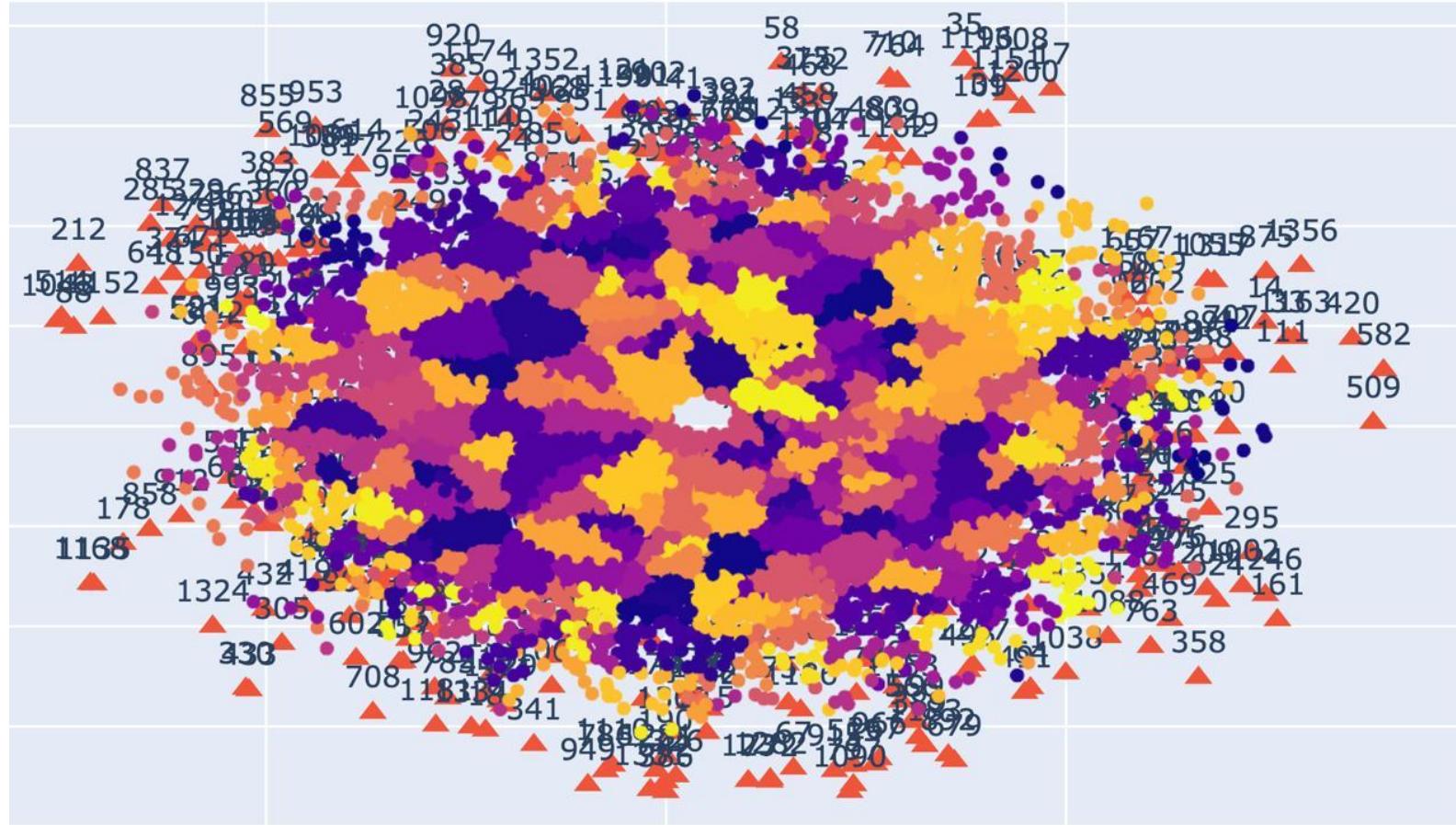
Для подбора гиперпараметров используется Байессовская оптимизация.

2'699'227

Итоговая средняя стоимость заправить автопарк из 5000 машин (в рублях)

# Кластеризация

Студкемп  
по математике в ИИ

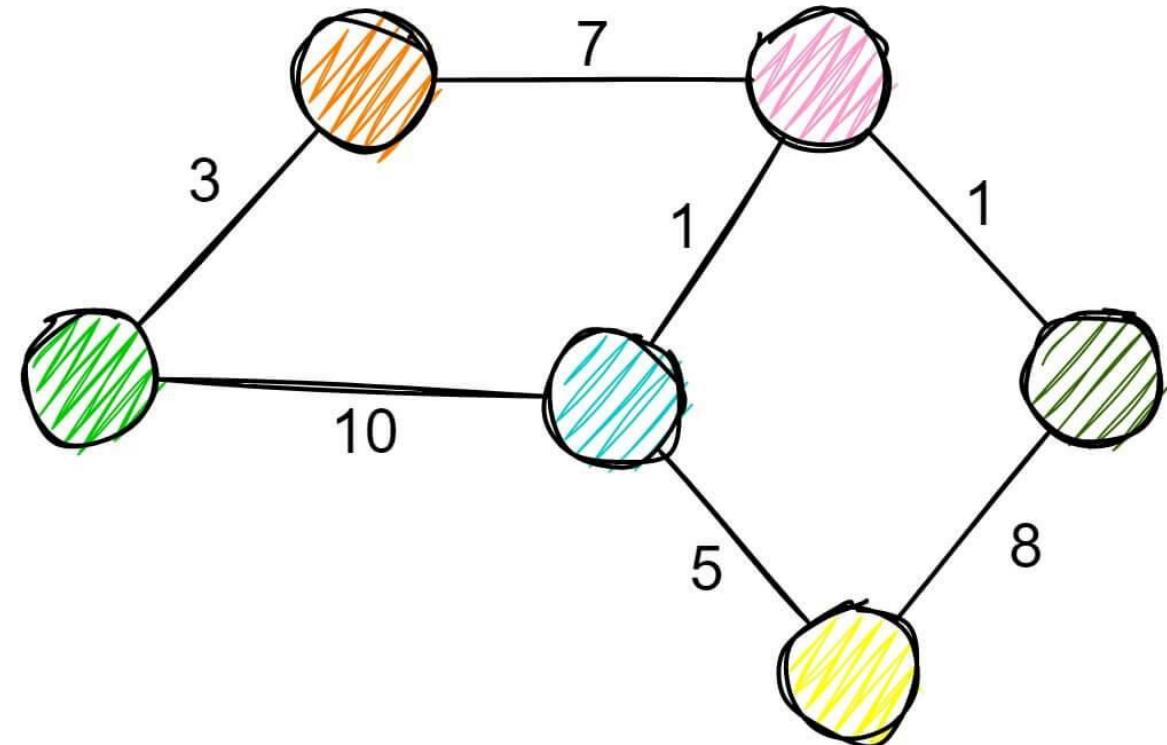


# Алгоритм обхода машин

При малом кол-ве данных будем использовать алгоритм Хелда–Карпа для поиска самого оптимального маршрута обезода не заправленных машин. ( $n < 12$ ).

В остальном случае будем обезжать жадно: сначала заправим самые “пустые” машины, затем заправим соседние, вернемся обратно и так далее.

Важное замечание: расстояние между 2 точками - манхеттенское.



# Кластеризация машин + соотнесение кластеров по заправкам

В алгоритме используется приближение:  
считается, что у всех машин одинаковая  
недозаправленность.

Алгоритм состоит из 2 этапов:

- Кластеризация машин. Каждый кластер  
это один обезд бензовоза.
- Определение самой оптимальной  
заправки для кластера.

Метрика между кластером и заправкой:

$$f(\text{cluster}, \text{station}) = P * \Sigma(D) + c * \text{dist}(\text{cluster}, \text{station})$$

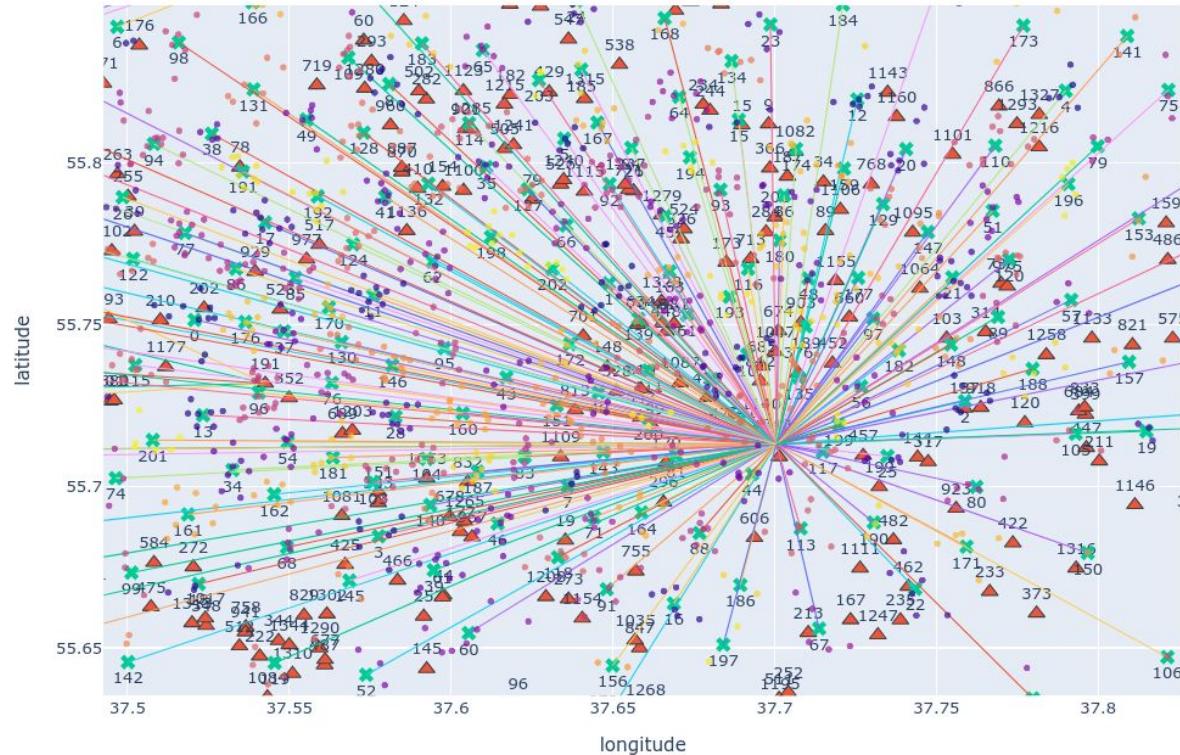
2'608'866

Итоговая средняя стоимость заправить автопарк из  
5000 машин (в рублях)

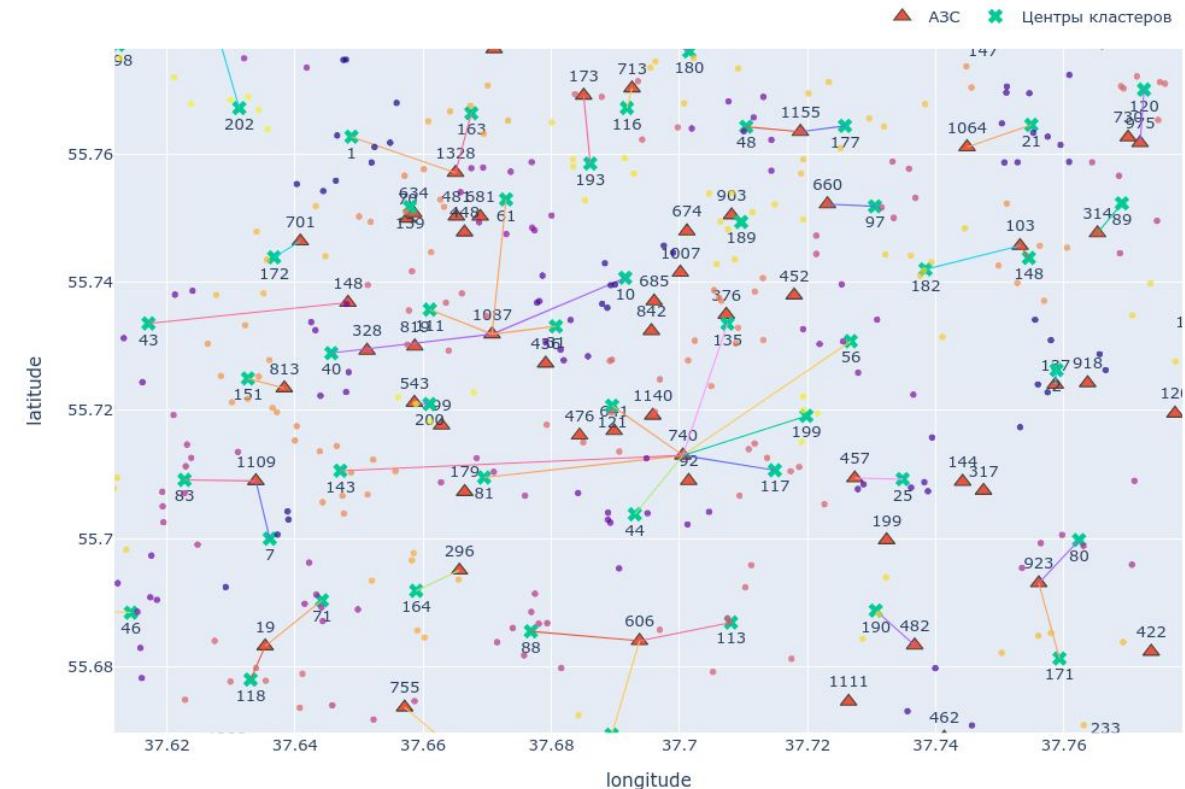
# Изменение гиперпараметра - стоимость проезда километра

Студкемп  
по математике в ИИ

C\_DIST\_PER\_KM=9



C\_DIST\_PER\_KM=700



# Распределение кластеров по заправкам

Студкемп  
по математике в ИИ

Посмотрим на распределение кластеров по заправкам. Всего вышло 143 различных станции - в нашем алгоритме это равносильно 143 бензовозам (столько нет). Цель на будущее - использовать меньше бензовозов, что приближено к реальной жизни, и распараллелить их работу.

Это увеличит простоя и совсем немного увеличит сумму проезда.

Топ станций по числу кластеров, которые к ним привязаны:

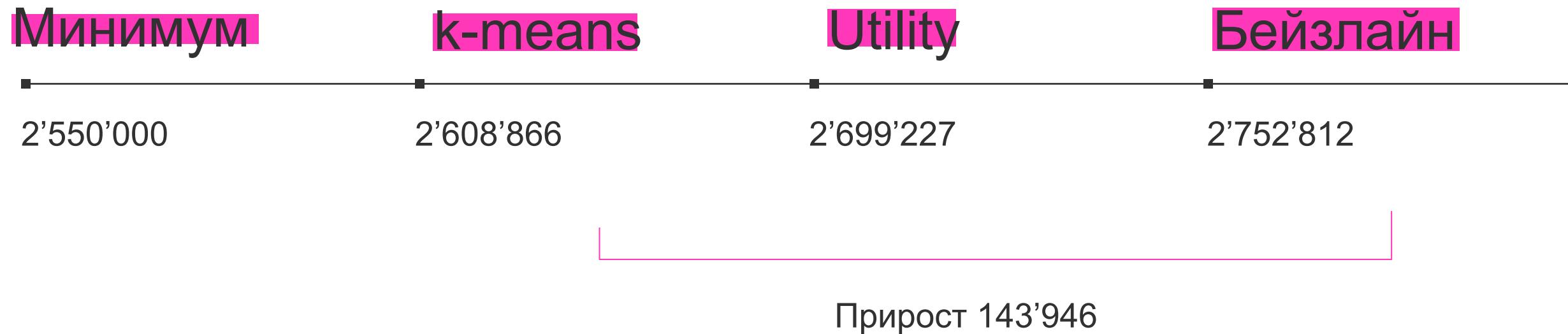
best_station_idx	num_clusters	price	lat	lon
23	191	9	56.30	55.731731
76	740	5	55.29	55.712989
33	263	4	60.95	55.796460
19	173	3	60.59	55.769146
10	79	3	60.75	55.788616
16	148	3	59.50	55.736821
60	606	3	60.81	55.684050
80	813	3	61.39	55.723477
101	977	3	60.69	55.770124
119	1143	3	60.95	55.821725
118	1136	3	60.02	55.778922
11	103	2	62.20	55.745700
9	78	2	60.80	55.798580
1	34	2	60.95	55.794012
0	19	2	60.79	55.683219
136	1290	2	58.58	55.649909
138	1315	2	60.95	55.819760
115	1109	2	62.69	55.708996
132	1241	2	60.73	55.805480
121	1155	2	60.86	55.763489
130	1237	2	61.00	55.793280
129	1216	2	63.90	55.804829
127	1203	2	60.70	55.717203
86	847	2	61.45	55.649931
7	60	2	60.63	55.837810
71	701	2	60.95	55.746401
57	584	2	61.45	55.676329
39	373	2	60.86	55.660674
70	692	2	60.89	55.765280
27	234	2	60.54	55.817997

06

Результаты

# Сравнение метрик

Студкемп  
по математике в ИИ



07

Дальнейшие  
планы на проект

# Дальнейшие планы на проект

Студкемп  
по математике в ИИ

- 
- Учитывать реальное кол-во бензовозов.  
Придумать алгоритм, при котором  
бензовозы могут обслуживать машины с  
нескольких заправок.
  - Решить задачу для других городов (Санкт-  
Петербург, Екатеринбург, и т.д.)
  - Использовать реальные данные вместо  
констант
  - Учитывать простоя в формировании  
маршрута бензовозов.

# Спасибо!

---

Бородин Матвей

Маргарита Ковалева

Клышников Михаил

Студкемп  
по математике в ИИ

Яндекс Образование