***Дипломный проект - Битва районов***

***Сиденко Станислав***

***20.01.2021***

Содержание:

1. Введение:
   1. Предпосылки
   2. Проблема
   3. Интерес
2. Данные
   1. Источники данных
   2. Сбор и очистка данных
   3. Выбор характеристик
3. Методология
4. Исследовательский Анализ
5. Результат

# 1.Введение

# 1.1. Предпосылки

В мире постоянно открываются новые увеселительные заведения, такие как ночные клубы. Это отличное место для отдыха и развлечения. В них можно не только отдохнуть, а также завести интересные знакомства. Такие заведения приносят хорошую прибыль. Всегда есть большие возможности в реализации любой идеи для оформления и выбора тематики для ночного клуба.

**1.2. Проблема**

На рынке очень большая конкуренция. При всех выгодах, много заведений закрываются, делая одну из главных ошибок на старте - выбор неправильного месторасположения. В этом проекте мы выберем оптимальное месторасположение ночного клуба в Москве.

**1.3. Интерес**

Данный проект будет интересен для предпринимателей, решивших открыть ночной клуб в Москве. Так же он будет интересен арендодателям, так как можно найти клиентов под ночной клуб и, обоснованно, установить цену и выстроить долгосрочные отношения с арендатором.

**2. Данные**

**2.1 Источники данных**

Определим факторы, влияющие на успешный выбор локации для размещения ночного клуба:

* Количество ночных клубов, построенных рядом
* Расстояние до центра
* Рейтинг клубов по соседству

Источники данных для проекта:

1. Для определения координат используется библиотека geopy;
2. Все ночные клубы, их местоположения и оценки получены при помощи Foursquare API;
3. С помощью библиотеки Pandas преобразуем json информацию в табличную форму;
4. Дистанции по координатам будут определяться с помощью библиотеки haversine.

Поиск места ограничим третьим транспортным кольцом, что примерно составляет 7 км в радиусе.

**2.2. Сбор и очистка данных**

Для начала я определил координаты центра Москвы, за который была принята Красная Площадь.

Далее нужно было получить все ночные клубы в радиусе 7 км от Красной Площади. Для этого нужно использовать Foursquare API. Здесь же возникла проблема ограничения количества получаемых ночных клубов за один запрос у Foursquare API.

Для решения этой проблемы нужно разделить рассматриваемую территорию на несколько зон, чтобы получить все расположенные на ней ночные клубы.  Я поставил 4 точки, равноудаленные от центра и проведем поиск по радиусу, пересекающие друг друга. Тут же возникает вторая проблема расчёта расстояния по широте и долготе.

Так как расстояния небольшие применим упрощение: для широты - 1 градус = 111.1 км, 1 градус долготы приблизительно равен 64 км. Это позволило избавиться от длинных формул по расчету расстояния по координатам.

Для перекрытия полученных нами зон берем радиус поиска на 500 метров больше (от 7 км выбранной нами зоны). На сайте Foursquare API находим категорию ночного клуба для поиска только нужных нам ночных клубов. От Foursquare API получаем информацию в формате Json, для удобства, с помощью библиотеки Pandas, преобразуем полученные данные в табличный формат. Так же, так как зоны выборки пересекаются, убираем дубликаты клубов и соединяем получившуюся информацию в один датасет.

**2.3. Выбор характеристик**

Получилась таблица с 97 ночными клубами и 19 характеристиками.

Первым делом нужно выбрать из полученной информации только те характеристики, которые помогут в анализе. Так же изменим названия характеристик, сделав их более понятными.

Отбросим все ненужные данные оставив основное. Полученные ночные клубы, отобразим на карте(Рис.1).

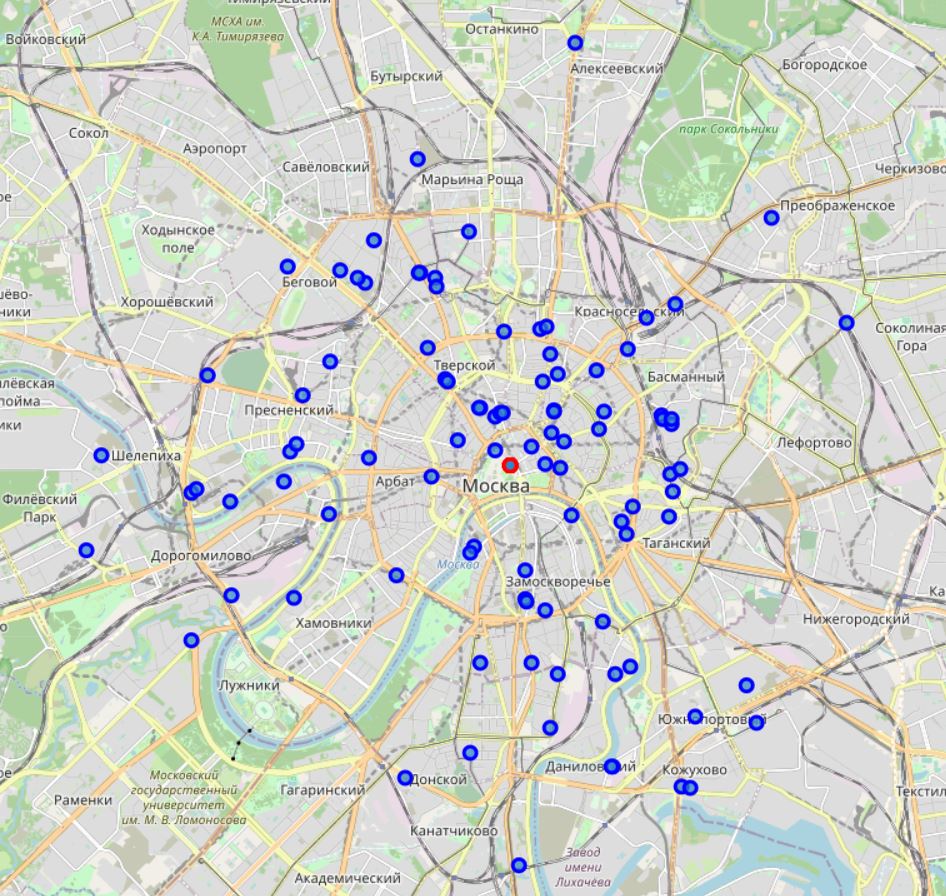


Рис. 1 Все ночные клубы Москвы.

Понадобятся данные дистанции от каждого взятого клуба до центра Москвы (Красной Площади), для этого воспользуемся библиотекой haversine. Полученные расстояния добавим в таблицу.