***Дипломный проект - Битва районов***

***Сиденко Станислав Александрович***

***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 и обсуждение
6. Вывод

# 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).

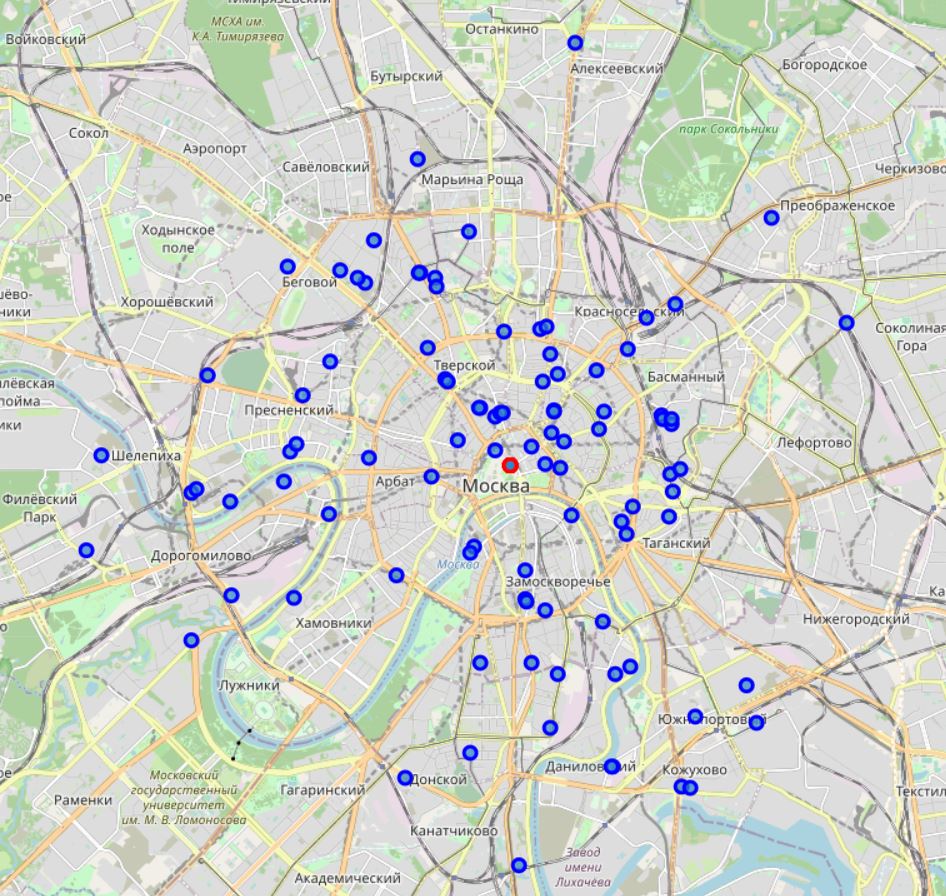


Рис. 1 Все ночные клубы Москвы.

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

**3. Методология**

В этом проекте я сосредоточил усилие для нахождения и определения территории, для наилучшего открытия ночного клуба.

Во-первых, я получил данный о местоположении всех существующих ночных клубов в пределах третьего транспортного кольца в Москве. Так же определил расстояние каждого ночного клуба до центра Москвы, за который мы взяли Красную площади и получили её координаты в начале. Для более полной картины по ночным клубам, используя Foursquare API были получены рейтинги клубов.

Во-вторых, я создам кластеры на основе полученных данных, для более точного определения будет использоваться разные свойства, а также два инструмента для кластеризации: Kmeans and DBSCAN.

В-третьих, проверю гипотезу, что, чем ближе расстояние до центра Москвы, тем выше рейтинг ночных клубов.

**4. Исследовательский Анализ**

Для начала возьмем самый распространенный метод кластеризации Kmeans, он достаточно быстрый и позволит подобрать наиболее подходящие гиперпараметры под него.

Так как мы можем предположить частный случай, что каждый ночной клуб будет относиться к отдельному кластеру, мы укажем данное допущение при поиске.

Данные по расположению ночных клубов на карте Москвы указаны в координатах, для наилучшей кластеризации переведем все координаты в радианы используя библиотеку Numpy.

Чтобы подобрать оптимальные параметры для метода кластеризации Kmeans используем GridSearchCV из библиотеки Sklearn.

Применив лучшие найденные параметры получили 75 кластера(Рис.2)

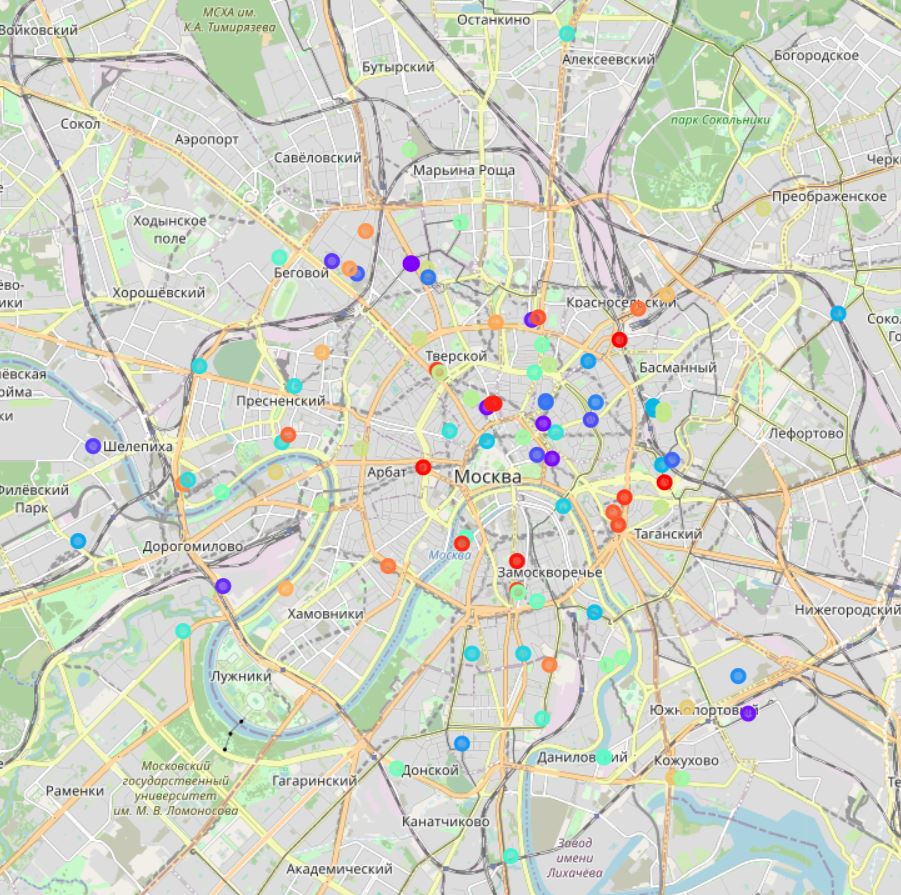


Рис.2 Разделение на кластеры методом Kmeans.

При изучении получившегося разбиения, к сожалению, никакой полезной информации выявлено не было. По этой причине был произведен повторная кластеризация методом DBSCAN (Рис.3).

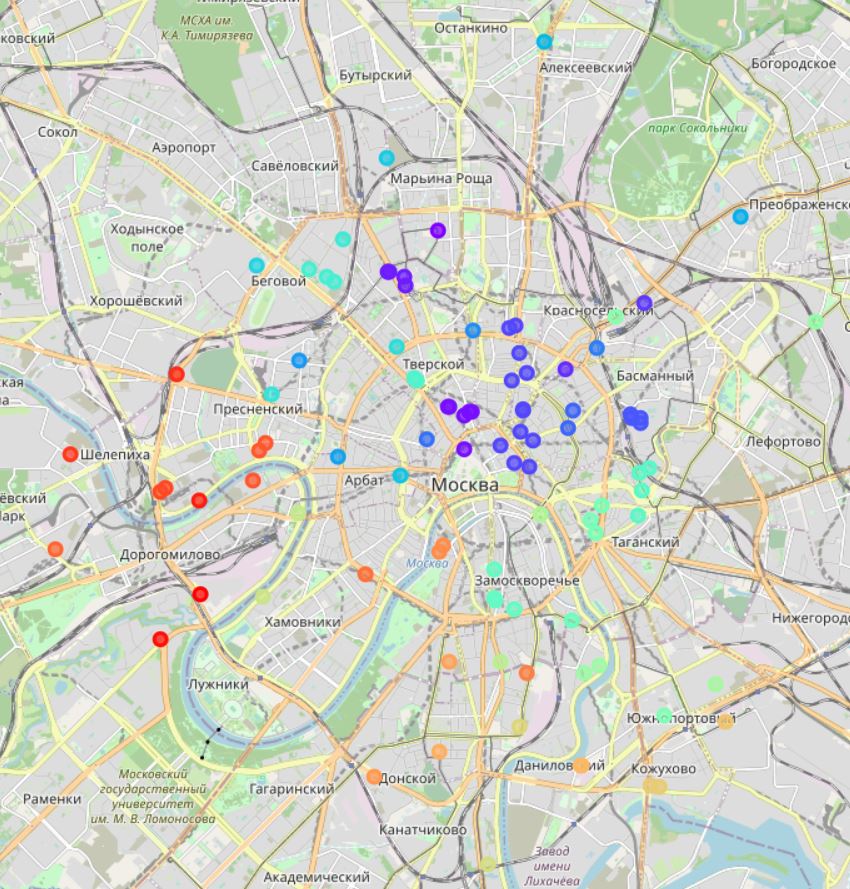


Рис.3 Разделение на кластеры методом DBSCAN.

В результате получились отчётливые группы ночных клубов, пригодных для дальнейшего анализа. Полученные 55 групп ночных клубов были внесены в таблицу.

Как видно из полученной карты, большинство ночных клубов разделены довольно большими расстояниями. Чётко прослеживаются только небольшие группы, которые более приближены к центру города.

Для более подробной информации об этих скоплениях провелась группировка и отобраны все скопления ночных клубов численностью равных или выше 4.

Эти группы были нанесены на отдельную карту(Рис.4)

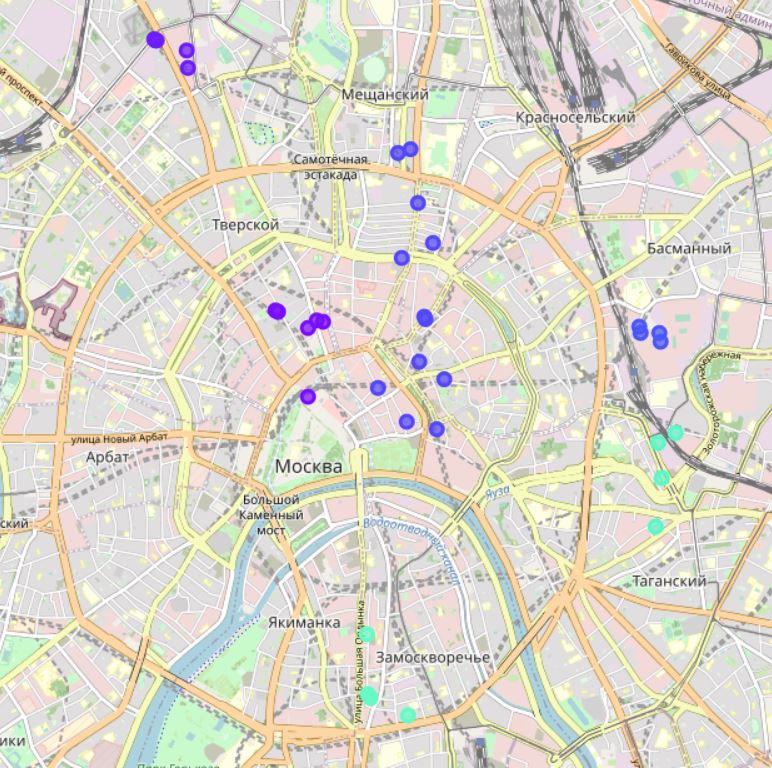


Рис.4 Отчётливые группы от 4-х клубов.

На полученной карте основное скопление ночных клубов, при детальном рассмотрении, можно привязать:

* самое большое количество клубов открыто в пределах садового кольца;
* все клубы располагаются рядом с театрами, музеями, бизнес парками.

Далее проверим гипотезу о связи удаленности от центра города и рейтинге клуба.

Для этого выделим все клубы, которые имеют рейтинг выше 7.0. Эти клубы отобразим на карте (Рис.5).

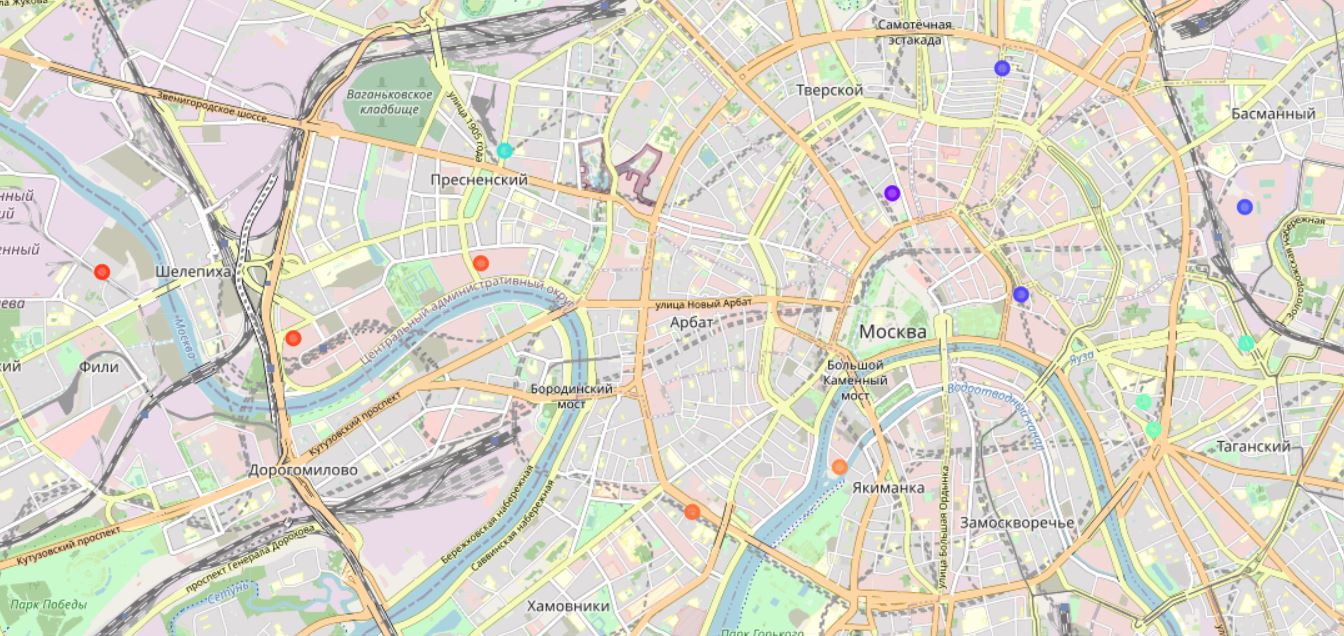


Рис.5 Клубы с рейтингом 7.0 и выше.

Посчитаем отношение общего количества клубов к клубам с рейтингом выше 7.0. Получилось, что всего 14,5% всех клубов имеют высокую оценку. В наших данных есть большое количество клубов с рейтингом 0, эти клубы можно считать новыми. Количество этих клубов ко всем составляет около 54,5%.

Отобразим на точечной диаграмме отношение дистанции до центра и рейтинга клубов (Рис.6)

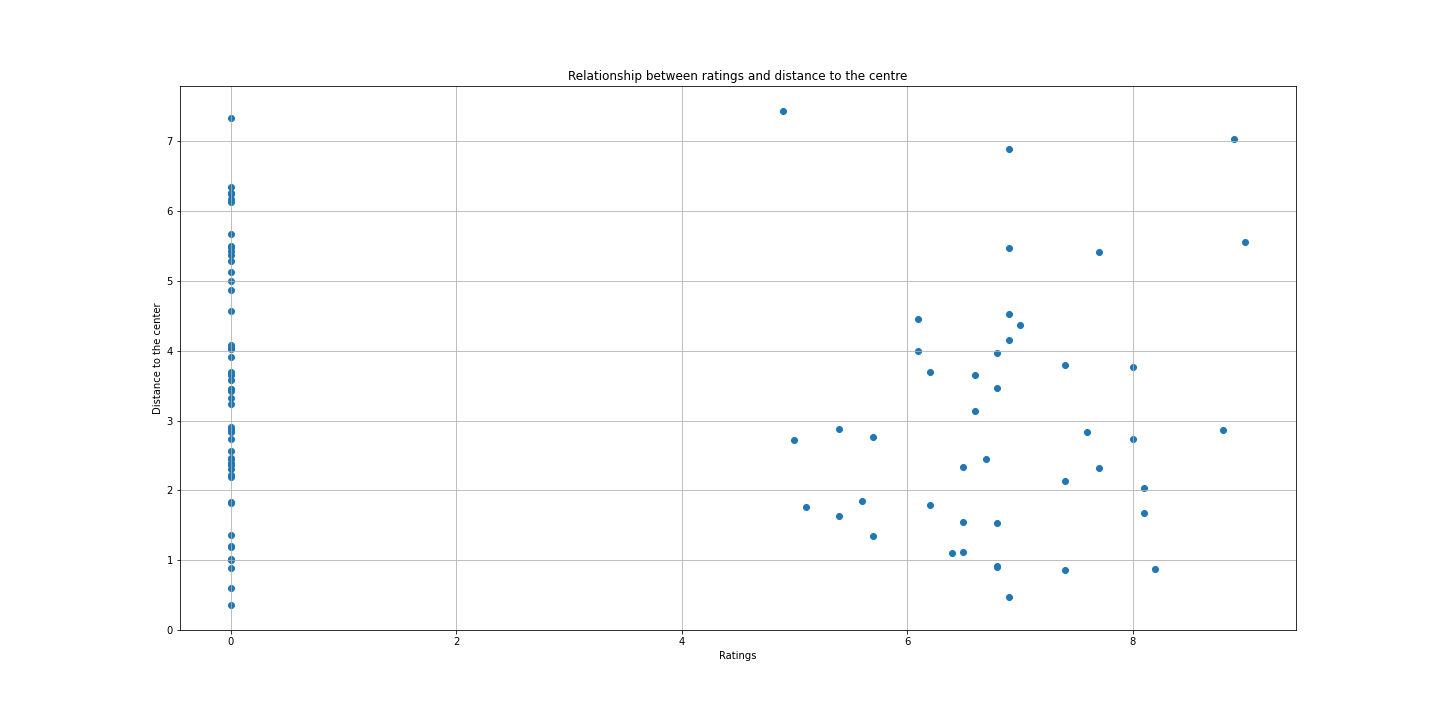


Рис.6 Отношение рейтинга клуба к удаленности от центра.

После рассмотрения карты и точечной диаграммы, можно сделать вывод, что гипотеза об отношении рейтинга к дистанции от центра не подтвердилась.

1. **Результат и обсуждение**

В этом исследовании были рассмотрены ночные клубы Москвы для выявления наилучших зон для открытия новых ночных клубов.

Как видно из исследования, конкурентоспособными являются приблизительно 14.5% клубов. Это те клубы, которые поддерживают свой рейтинг выше 7.0. Чёткого распределения заведений с высоким рейтингом не прослеживается, они разбросаны по исследуемой территории, что дает простор для конкурентной среды и это дает широкие возможности для открытия новых, что как раз и наблюдается в виде 54.5%. Очень большое количество новых ночных клубов дает возможность раскрутиться в этой сфере, используя профессиональный сервис и клиентоориентированность, это подтверждается очень маленьким процентом клубов с высоким рейтингом.

В исследовании хорошо прослеживается тренд в расположении клубов близко к зонам туризма и театрам\музеям, так же очень популярны бизнес парки.

В связи со всеми выявленными трендами и плотностью ночных клубов в Москве можно выявить три основные рекомендованные зоны. Они находятся близко к туристическим зонам, так же в них отсутствуют скопления ночных клубов (зеленые зоны на рис.7).

Помимо этого, выявлены 2 зоны, отмечены красным, в которых не рекомендуется открытие заведений. Несмотря на то, что они удовлетворяют условиям, в этих зонах находятся большие скопления религиозных и государственных учреждений, помимо туристических.

Очень большой разброс ночных клубов по городу, даёт возможность к более большому выбору с минимальной конкуренцией.

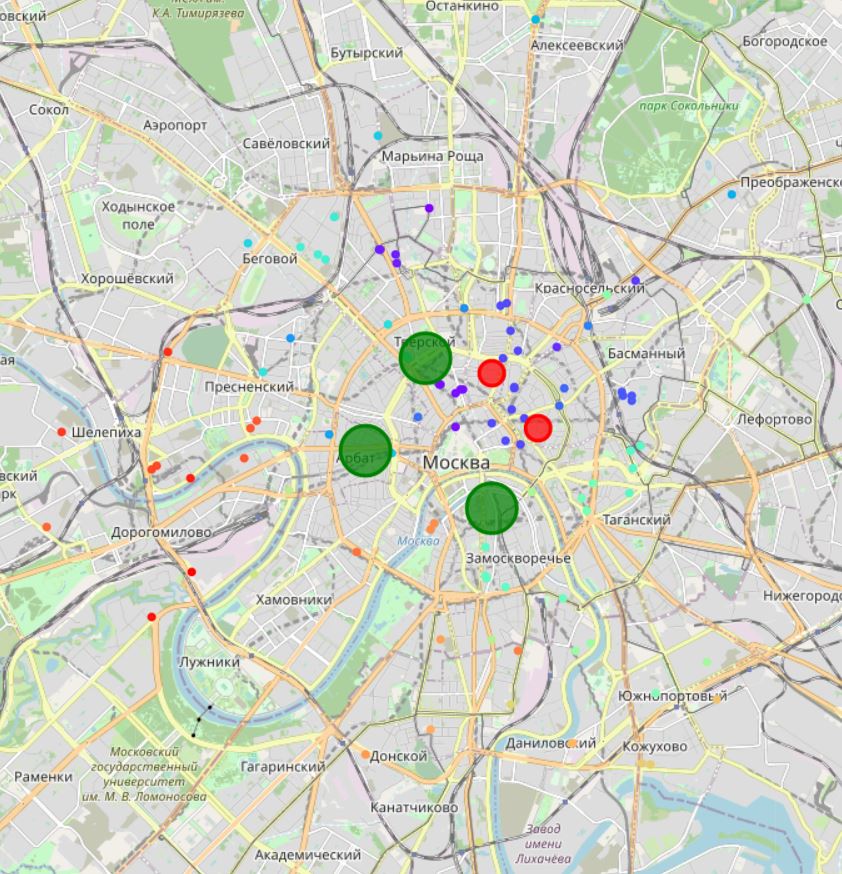


Рис 7. Наилучшие и наихудшие зоны для открытия ночных клубов.

1. **Заключение**

Целью этого проекта было определение благоприятных зон для открытия ночных клубов. Для предпринимателей это хорошее подспорье к их бизнес плану, а для арендодателей возможность заинтересовать арендаторов. Были созданы несколько карт для отображения результатов. Была опровергнута теория о связи дистанции до центра и рейтинге клубов. Очень большое количество новых ночных клубов и их разрозненность по карте может быть причиной недавней эпидемии и возможностью для открытия новых клубов при ослаблении мер, связанных с ней.