Задача исследования отделений Почта Банк по данным из открытых источников

Аналитическая записка

Выполнила Софья Марцинович

Приложение: Jupiter notebook “Pochta Bank” (<https://github.com/smartsinovich/career-msu>)

Постановка задачи

Была поставлена задача провести исследование отделений банка по представленным данным, собранным из открытых источников. Данные включают в себя характеристики отделений и ближайших к ним окрестностей. Данные представлены в сыром виде, имеют ошибки, пересечения, неправильные записи и пр.

Описание данных

Для анализа был использован предоставленный датасет (файл c расширением csv).

Размеры датасета: (717, 33). Датасет имеет 33 категории.

Data types: "id", "post\_id", "open\_date", "street", "address", "place\_type", "shop\_centre", "dns", "lat", "lon", "response", "district", "post2", "is\_center", "price", "parking\_count", "parking\_min\_distance", "finance\_count", "shop\_count", "distance\_100", "distance\_500", "cafe\_count", "culture\_count", "job\_gov\_count", "metro\_count", "metro\_min\_distance", "transport\_count", "transport\_min\_distance", "working\_days", "work\_hours", "weekday\_availability", "office\_hours", "is\_moscow".

Пре-процессинг данных  
 В ходе анализа будет использоваться только часть исходных данных. Для дальнейшей работы создаем датасет, который состоит из наиболее важных категорий (характеристик):

data = data[[ 'post\_id', 'open\_date', 'address', 'lat', 'lon', 'district', 'post2', 'working\_days', 'work\_hours', 'weekday\_availability', 'office\_hours', 'price', 'is\_moscow']]

Первичная оценка датасета позволяет сделать следующие выводы:

1. Отделения банка распределены по округам и что разные округа имеют разное количество отделений.
2. Часть округов расположена в Москве, оставшаяся часть – за пределами Москвы.

Аназиз данных

Выбор уникальных отделений

Исходные данные содержат повторения. Выделим из исходных 717 строк уникальные.

print("Number of unique mail office:", data['address'].nunique())

Number of unique mail office: 703

Общее количество уникальных отделений 703, что меньше, чем количество записей в оригинальном датасете.

Приведем пример повторяющихся данных:

# take a peek at the address column

data['address'].head()

0 пл. Преображенская, дом 8

1 пл. Преображенская, д. 8

2 пл. Преображенская, д. 8 ТП №2

3 пл. Преображенская, дом 8

4 рп Андреевка, дом 3 Б

Name: address, dtype: object

На примере видно, что отделение по адресу пл. Преображенская, дом 8 упомянуто в датасете не менее 4 раз.

Поскольку возможна ситуация, когда одно и то же отделение присутствует в датасете под разными id, будем выявлять уникальные отделения через их географические координаты.

Для начала проверим уникальность координат:

print("Number of unique bank offices:", data['lat'].nunique())

#если берем широту (lat) - еще меньше 557

Number of unique bank offices: 557

print("Number of unique bank offices:", data['lon'].nunique())

#если берем широту (lon) -    одинаковая с lat 557

Number of unique bank offices: 557

Проверка показала, что имеем дело с 517 уникальными отделениями.

Цены на недвижимость

Изучим зависимость месторасположения отделения и стоимости недвижимости. Выделим в два датасета отделения в Москве и не в Москве. Используем категорию “is\_moscow”

df\_mos = data.loc[data['is\_moscow'] != 0] # таблица для москвы

Построим распределение цен на недвижимость для Москвы и неМосквы.

Figure 1. Цены для Москвы

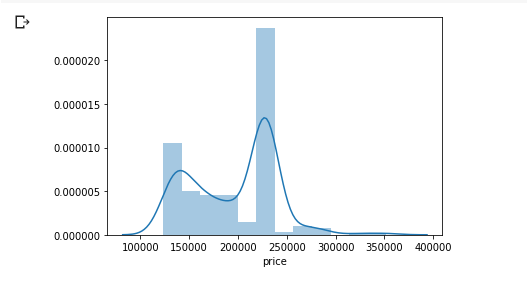
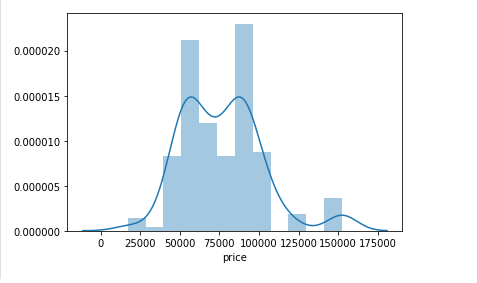


Figure 2. Цены для не Москвы

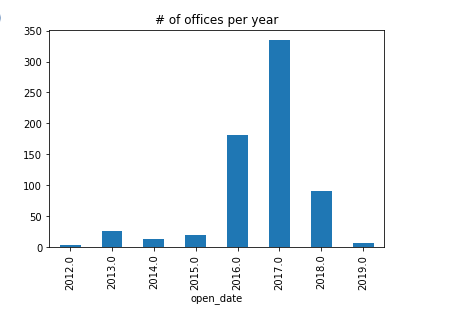
****

Цены в подмосковье более приближены к нормальному распределению.

Развитие сети отделений

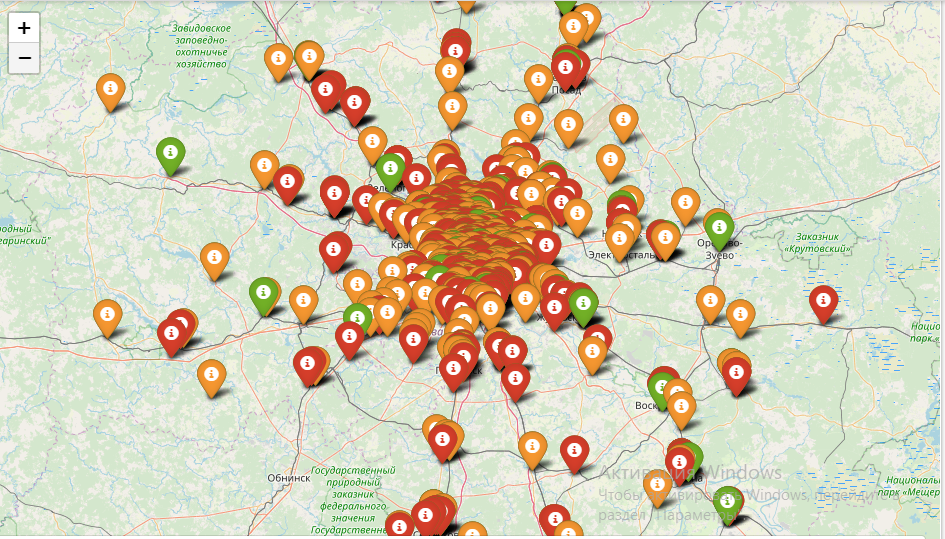
Исходный набор данных предоставляет информацию о дате открытия каждого отделения. Построим график на основе этой информации. Видно, что основное развитие сети отделений происходило в 2016, 2017 и 2018 годах. Пик приходится на 2017 год, когда было открыто большинство новых отделений.

Figure 3. Количество открытых отделений по годам



В настоящее время новые отделения практически не открываются.

Figure 5.Цветом выделены отделения, открытые в период интенсивного роста сети.



Зеленый - до 2016

Оранжевый - 2017

Красный -  2018

Деление офисов на московские и немосковские

Поделим округа на московские и немосковские. В результате получаем следущие цифры: в Москве 418 отделений, за пределами Москвы — 298 отделений.

#print("Number of unique bank office in Moscow:", data['is\_moscow'].nunique())

# москва vs немосква.

by\_join\_moscow = data[['is\_moscow']].groupby(by='is\_moscow').size().rename('number\_of\_moscow\_offices')

by\_join\_moscow

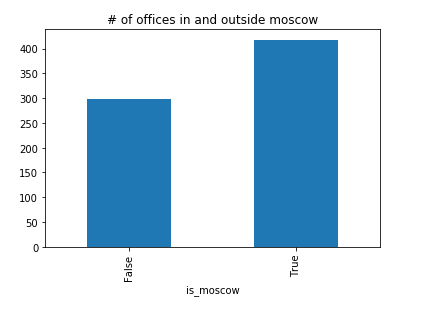
is\_moscow

False 298

True 418

Name: number\_of\_moscow\_offices, dtype: int64

Figure 6. Кол-во отделений внутри Москвы и вне Москвы



Распределение отделений внутри округов (районов)

Отделения как внутри Москвы, так и вне ее распределены по округам. Определение «округ» в нашем случае нестрогое, иногда это официальное название округа, например, «Центральный административный округ», иногда это название района, например «Речинский».

Разделим округа на московские и немосковские.

df\_mos['district'].value\_counts() #в москве

Центральный административный округ 19

Южный административный округ 13

Марьино 12

Москва 11

Алексеевский 10

..

Аэропорт 1

улица Димитрова 1

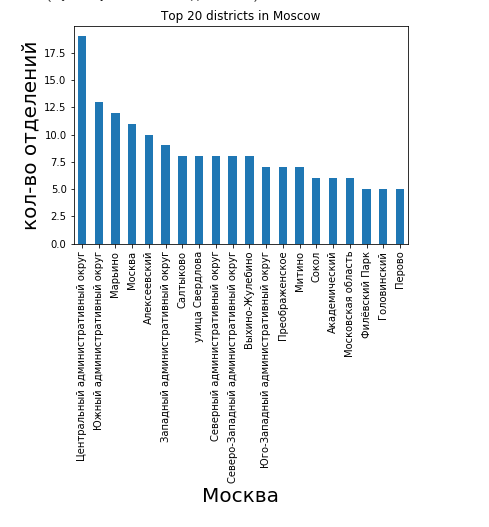
Проспект Вернадского 1

Квартал №532 1

Южное Медведково 1

Name: district, Length: 152, dtype: int64

Figure 7. Кол-во отделений по округам



count\_by\_district = df\_mos['district'].value\_counts()

count\_by\_district.describe()

count 152.000000

mean 2.750000

std 2.675001

min 1.000000

25% 1.000000

50% 2.000000

75% 3.000000

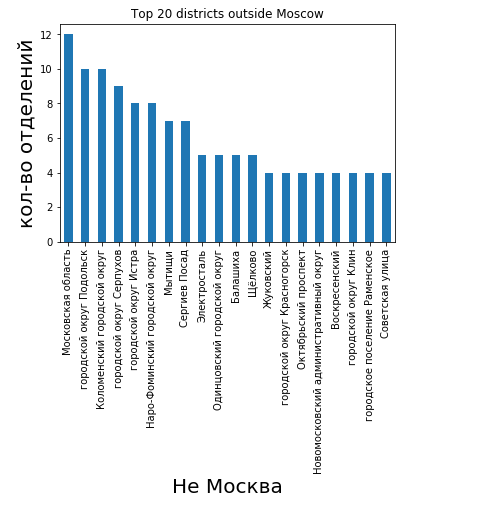
max 19.000000

Name: district, dtype: float64

В Москве большинство округов имеют всего 2-3 отделения, mean number = 2.75.

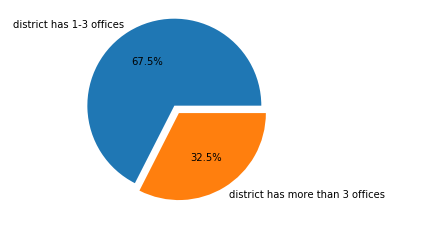
Самое большое количество отделений (19) в Центральном административном округе.

Figure 8. Кол-во отделений по округам вне Москвы



Большинство округов вне Москвы имеют 2 отделения, mean number = 2.22

Figure 9. Округа с разным кол-вом отделений



Выводы и рекомендации

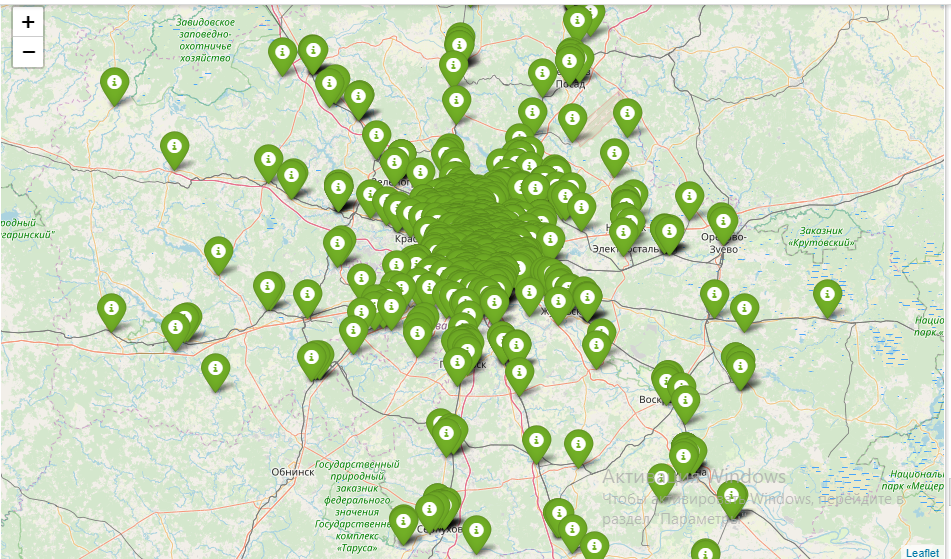
Из полученных данных следует, что строительство сети отделений в целом завершено. Наши выводы носят не рекомендательный, а скорее оценочный характер.

Согласно заявлениям первых лиц банка:

(*с) «Мы придуманы как система, позволяющая обычному человеку, вне зависимости от того, где он проживает – в большом городе, в маленьком или в селе, получить доступ к базовому набору банковских услуг, необходимому на сегодняшнем этапе развития цивилизации для нормальной жизни. Поэтому у нас есть все возрасты. В стратегии на первые годы прописано, что мы фокусируемся на двух типах клиентов, для которых мы делаем целевое предложение: это пенсионеры и активные посетители почты.»[[1]](#endnote-1)*

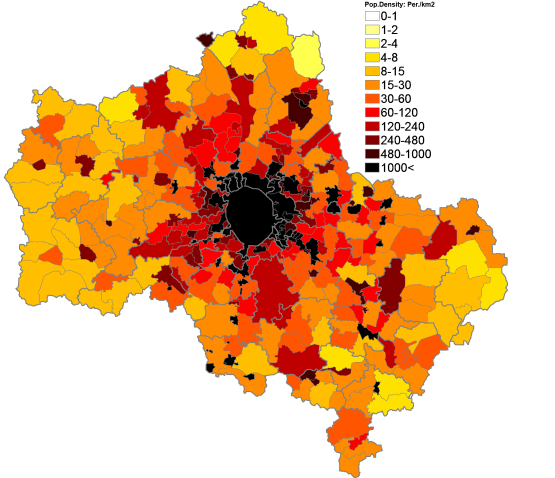
Иными словами, сеть отделений нужно было развивать там, где есть клиенты.

Figure 10. Расположение отделений на карте Москвы и МО



Видно, что отделения в основном открывались в центральных районах Москвы не смотря на чрезвычайно высокие цены, а не в быстро растущих новых районах подмосковья.

Figure 11. Плотность населения для МО



Возможно это связано с тем, что согласно сложившейся практике отделения банка вне столицы открываются в помещениях отделений связи почты России.

Такая практика представляется вполне разумной, если учитывать цены на недвижимость, а также многолетние попытки почты России получить особые условия по аренде.

*(с) «В перечень преференций, предоставляемых назначенному оператору почтовой связи, целесообразно дополнительно включить следующие:  
• ставки по аренде помещений и земельных участков для размещения объектов почтовой связи на уровне арендных ставок для иных социальных объектов.»[[2]](#endnote-2)*

Также банком пробуются новые форматы помещений. Так согласно Википедии: *«В январе 2019 г. в рамках пилотного проекта три первых отделения Почта Банка открылись в магазинах «Пятёрочка» в городах Санкт-Петербург и Калуга. Открытие таких мини-отделений планируется в более чем 50 магазинах «Пятёрочка» по всей стране».*

1. Ссылка на источник: <https://www.vedomosti.ru/finance/characters/2018/01/23/748740-pochta-bank> [↑](#endnote-ref-1)
2. Ссылка на источник:<https://www.pochta.ru/documents/10231/726548998> [↑](#endnote-ref-2)