

# Анализ корреляции дополнительных услуг и рейтинга заведений Horeca (кафе)

Авторы: Карташова Екатерина, Рыжкова Валерия  
Научный руководитель: Мултанен Татьяна Генриховна



# Введение

- В России предпринимательской деятельностью занимаются примерно 3% населения, 4,2 млн человек
- Задача предпринимателя – минимизировать издержки и увеличить прибыль при открытии и ведении бизнеса
- Предпринимателям не всегда удаётся проанализировать существующие дополнительные услуги



# База данных

## Описание

Данные с сайта «Яндекс карт» об отзывах с данными об организациях Москвы и Санкт-Петербурга за 2018-2021

## Объемы

1 таблица  
105 тыс.строк в каждой

user_id	org_id	rating	ts	user_city	org_city	average_bill	rating_org	rubrics	food_delivery	breakfast	takeaway	home_terr	wi-fi	business_hours
1.7E+19	7.2E+18	2	105	msk	msk	500	4.03869	Быстрое	1	1	1	1	1	1
1.2E+19	7.2E+18	1	1072	msk	msk	500	4.03869	Быстрое	1	1	1	1	1	1
1E+19	7.2E+18	2	420	msk	msk	500	4.03869	Быстрое	1	1	1	1	1	1
3.3E+18	7.2E+18	3	643	msk	msk	500	4.03869	Быстрое	1	1	1	1	1	1
6.2E+18	7.2E+18	5	701	msk	msk	500	4.03869	Быстрое	1	1	1	1	1	1
1.6E+19	7.2E+18	5	319	msk	msk	500	4.03869	Быстрое	1	1	1	1	1	1
1.2E+19	7.2E+18	5	647	msk	msk	500	4.03869	Быстрое	1	1	1	1	1	1
6.1E+18	7.2E+18	4	597	msk	msk	500	4.03869	Быстрое	1	1	1	1	1	1
1.3E+19	7.2E+18	5	725	msk	msk	500	4.03869	Быстрое	1	1	1	1	1	1
9E+18	7.2E+18	5	345	msk	msk	500	4.03869	Быстрое	1	1	1	1	1	1
5.2E+18	7.2E+18	5	598	msk	msk	500	4.03869	Быстрое	1	1	1	1	1	1
1.7E+19	7.2E+18	5	1132	msk	msk	500	4.03869	Быстрое	1	1	1	1	1	1

**user\_id** — идентификатор пользователя

**org\_id** — идентификатор организации

**rating** — поставленная оценка в отзыве

**ts** — дата отзыва

**user\_city** — информация о городе проживания пользователя

**org\_city** — информация о городе организации

**average\_bill** — средний чек в рублях

**rating\_org** — средний рейтинг организации

**rubrics** — рубрика, указывающая на основной вид деятельности организации

**food\_delivery:handmade\_goods** — 63

последних столбца в таблице

представляют собой бинарные переменные

# Цели и гипотезы



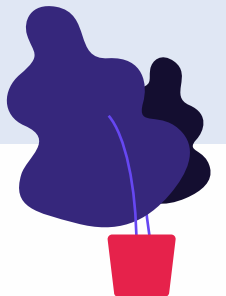
## Цель

- Выявить список тех дополнительных услуг, которые положительно влияют на рейтинг заведения и могут повысить его

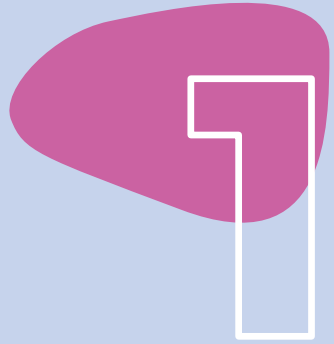


## Гипотезы

- Наш продукт будет актуален целевой аудитории
- Дополнительные услуги влияют на рейтинг заведения



# Задачи



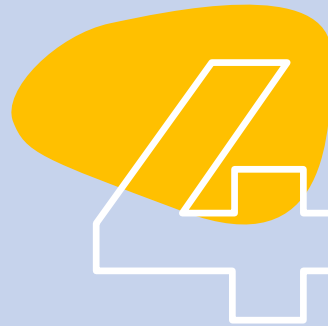
Провести  
анализ рынка  
и выявить  
потребности  
целевой  
аудитории



Сократить  
выборку и  
собрать  
недостающие  
данные



Определить  
оптимальный  
метод  
исследования  
и провести его



Выдвинуть  
рекомендации для  
предпринимателей



# Актуальность для целевой аудитории



Более 50% потребителей предварительно читают онлайн-отзывы о продукте



В 2019 году было превышение закрытий над открытиями —много непродуманных проектов заведений



95% покупателей читают отзывы, размещенные на сайтах-отзовиках



Опрошенные обращают внимание на:  
62% — интерьер,  
57% — удобство расположения,  
35% — наличие программы лояльности,  
26% — вид из окна,  
19% — фоновую музыку.



Прибыль заведения зависит от рейтинга

## Целевая аудитория

### 1) предприниматели

- имеют кафе/неопытны в ведении кафе
- планируют открыть новые точки/первое кафе

### 2) аналитики/экономисты

- избежать ошибок в отчетах
- принести прибыль компании
- понять, над какими проблемными местами предстоит работать

## Портрет

Андрей, 35 лет Санкт-Петербург

Высшее образование

Решительный, целеустремленный

Предприниматель, владеет одним кафе

-планирует расширить свой бизнес, открыть точки в центре Санкт-Петербурга

-боязнь неудачи и ошибок, потери авторитета, финансов

-неуверенность в собственных силах и опыте

-наш продукт:

анализирует его ситуацию и помогает внедрить результаты анализа, чтобы избежать прошлых ошибок и не допустить новых



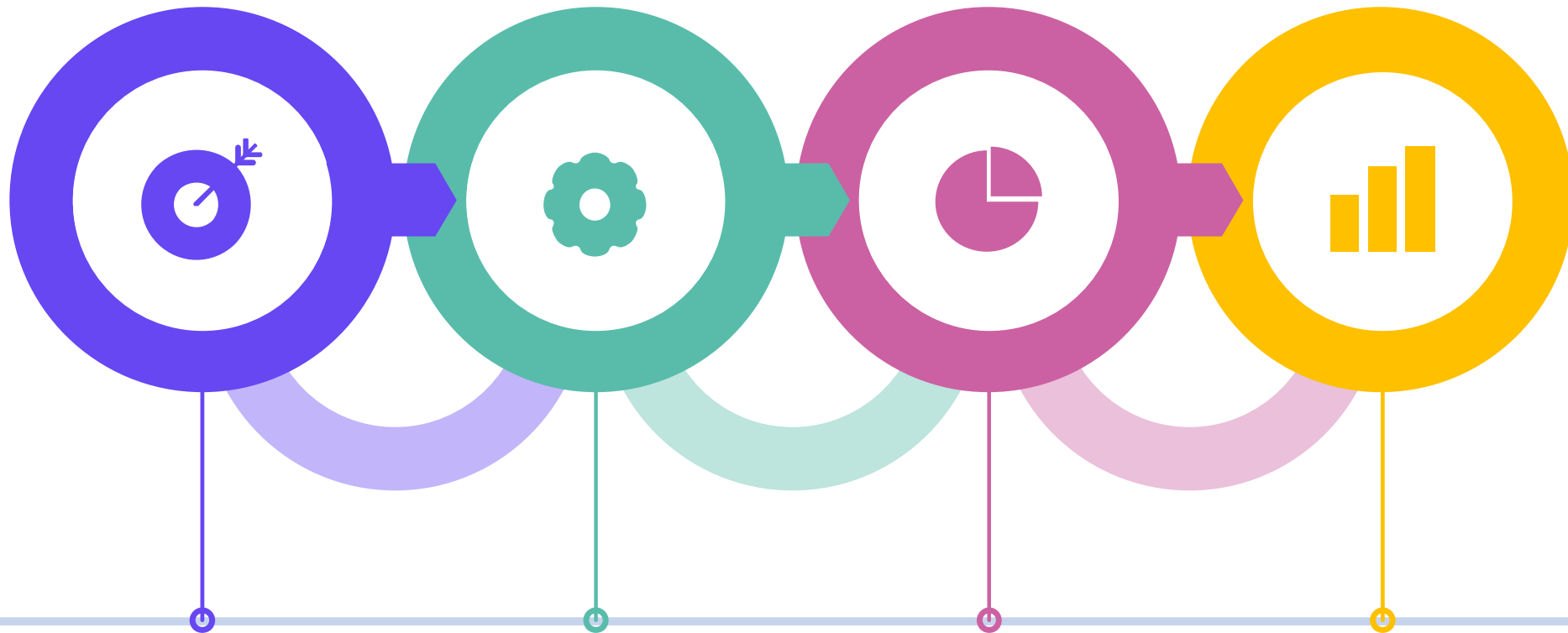
# Конкуренты

	Стоимость	Удаленная работа	Определение концепции заведения	Анализ работы заведения	Аудит кухни	Предоставляют ли конкретные доп.услуги?	Анализ целевой аудитории
Исследовательская компания «Restteam»	88,000 руб.	✗	✓	✓	✓	✗	✓
Исследовательская компания «Mozg.rest»	8,000-10,000 руб.	✓	✗	✓	✓	✗	✓
Исследовательская компания «trade-drive»	17,000-25,000 руб.	✓	✓	✓	✗	✓	✓

Вывод: на рынке нет недорогого проекта, выполняющего схожие функции



# Swot-анализ



## Сильные стороны

- инновационность в использовании данных с отзывов посетителей
- легко контролировать процесс и управлять изменениями

## Слабые стороны

- отсутствие опыта в работе
- погрешности
- неиндивидуальный подход

## Возможности

- нет продукта, выполняющего такие же функции, на рынке
- коллаборации с известными аналитическими компаниями

## Риски

- незаинтересованность ЦА
- отсутствие доверия клиентов
- появление новых игроков на рынке

# Какой рейтинг считается хорошим?

Исследования крупных компаний:



## Aliexpress

Отрицательная оценка — ★/★★

Нейтральная оценка — 3 звезды - ★★★

Положительная оценка — 4-5 звезд. ★★★★★/★★★★★



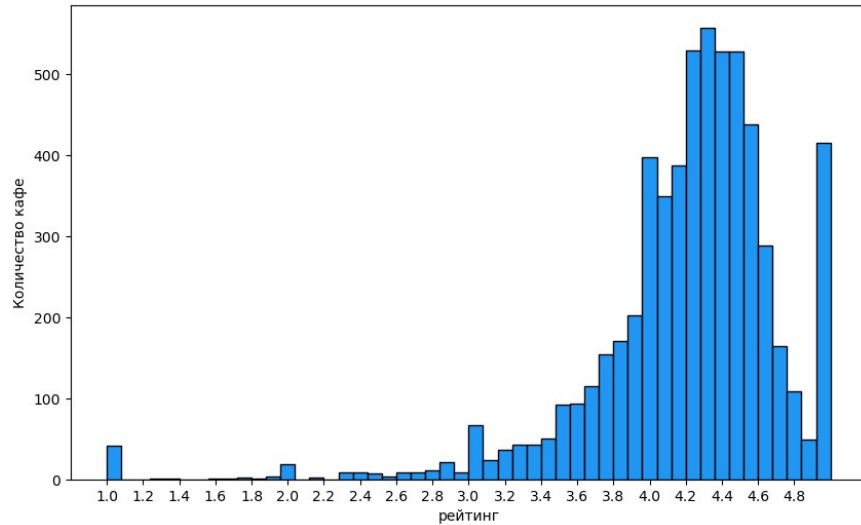
## Совкомбанк

4,5 – 4,7 баллов - рейтинг, поднимающий продажи

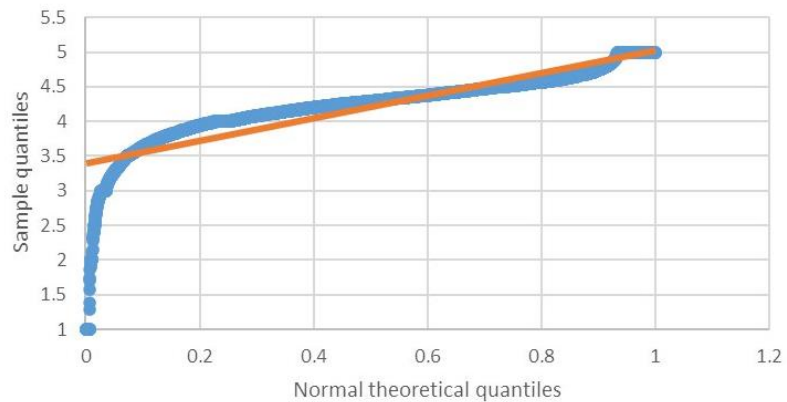
\*Максимально возможный рейтинг (рейтинг 5.0) не вызывает доверие у людей

# Проверка на нормальность

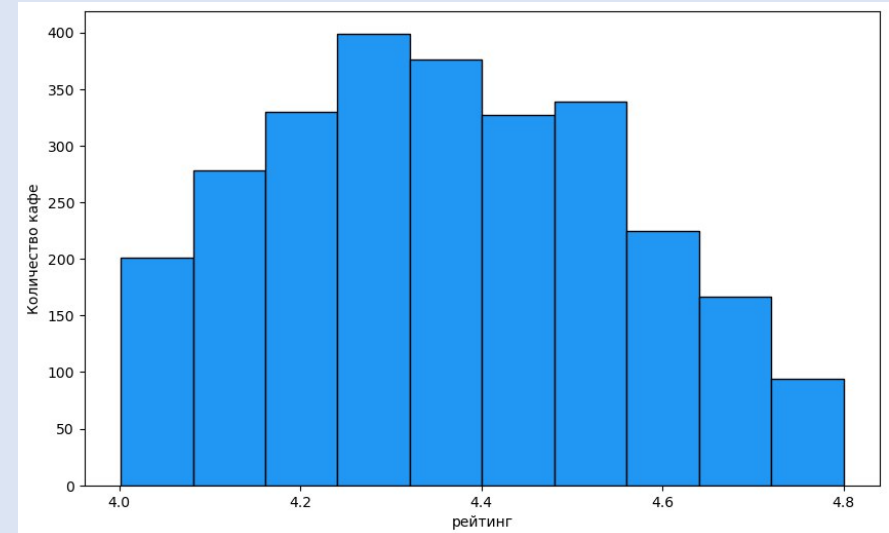
## Исходные данные



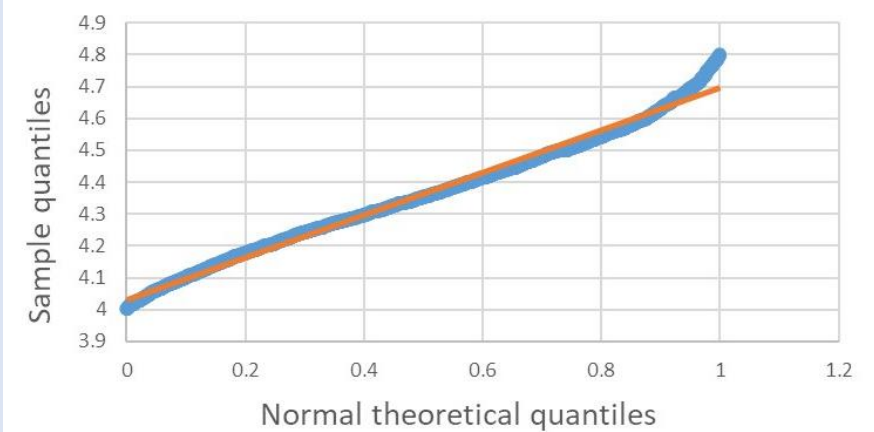
Q-Q plot



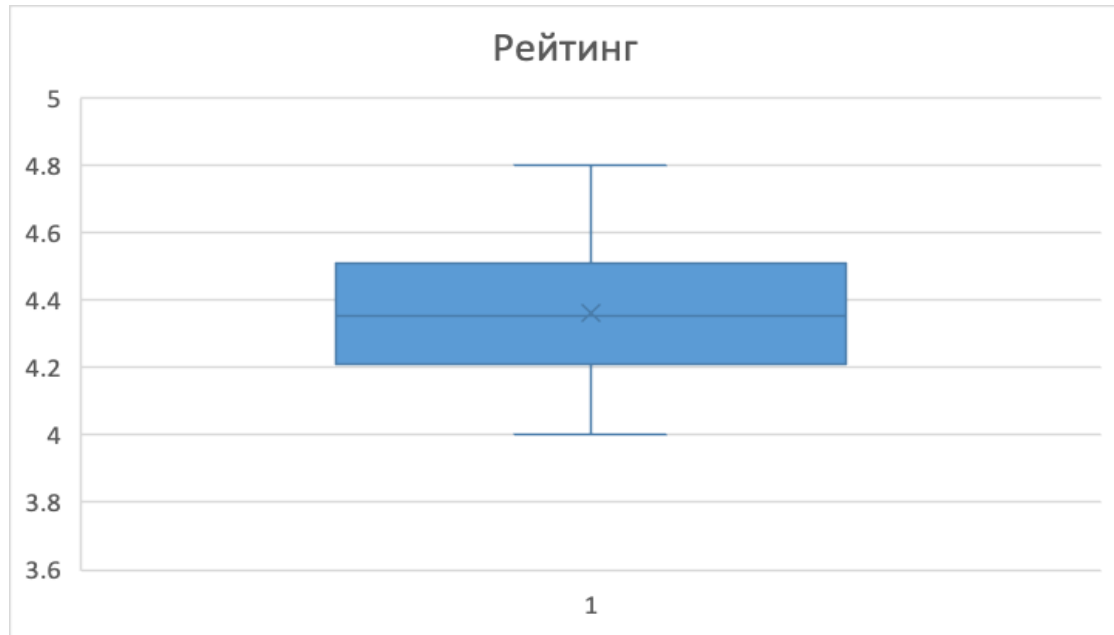
## Итоговые данные



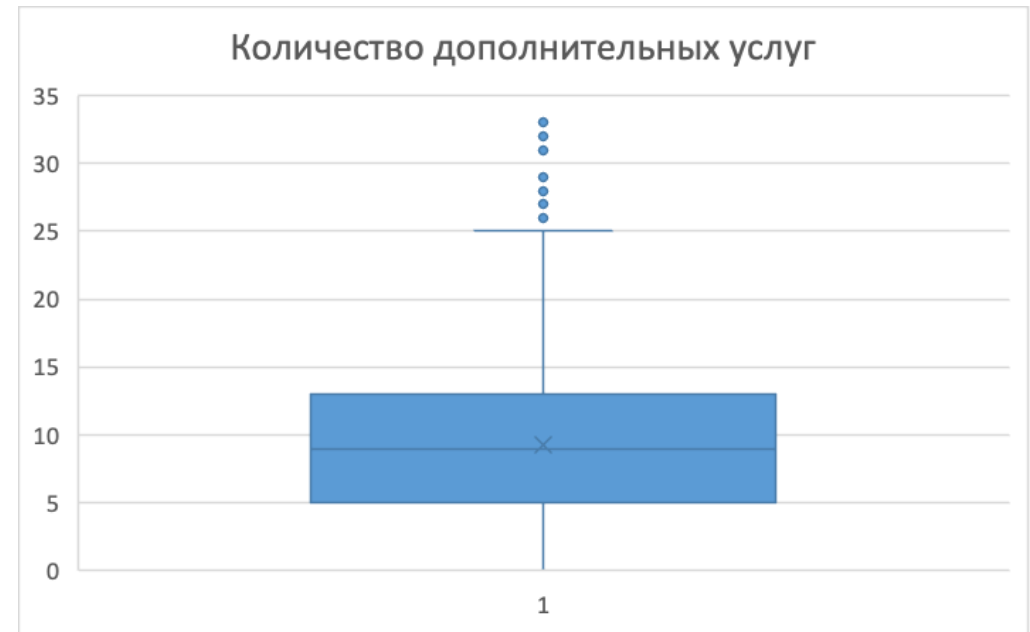
Q-Q plot



# Описательная статистика



Медианное значение = 4,35  
Размах=0,9

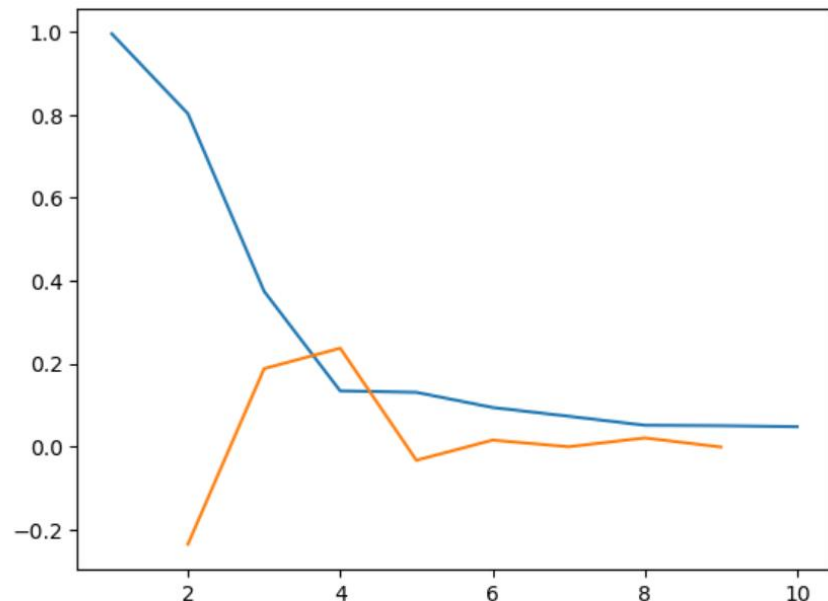


Медианное значение = 9  
Размах=25



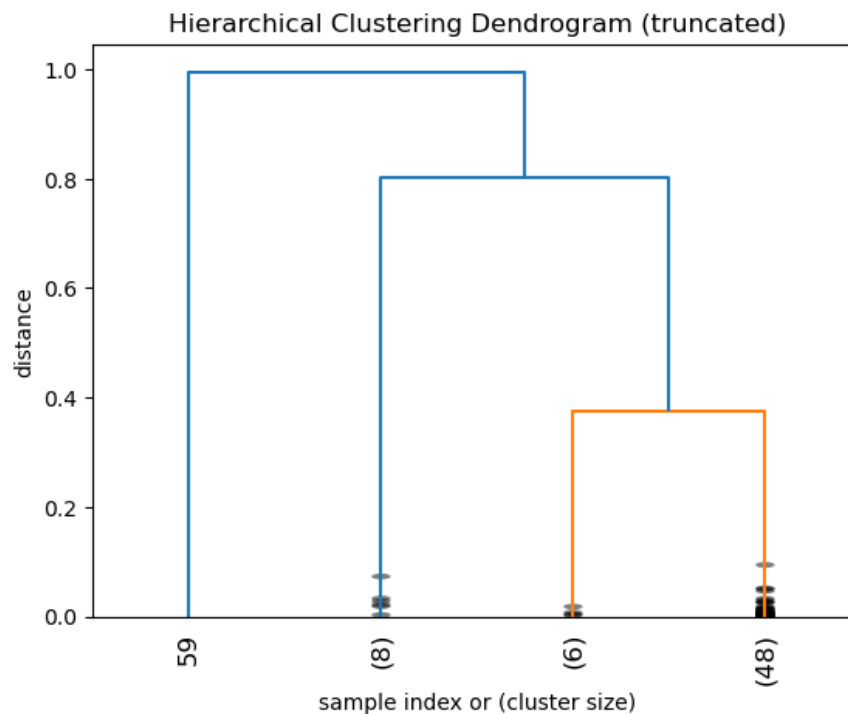
# Кластерный анализ

## Метод «локтя» (кластеризация)



Рекомендованное количество кластеров: 4

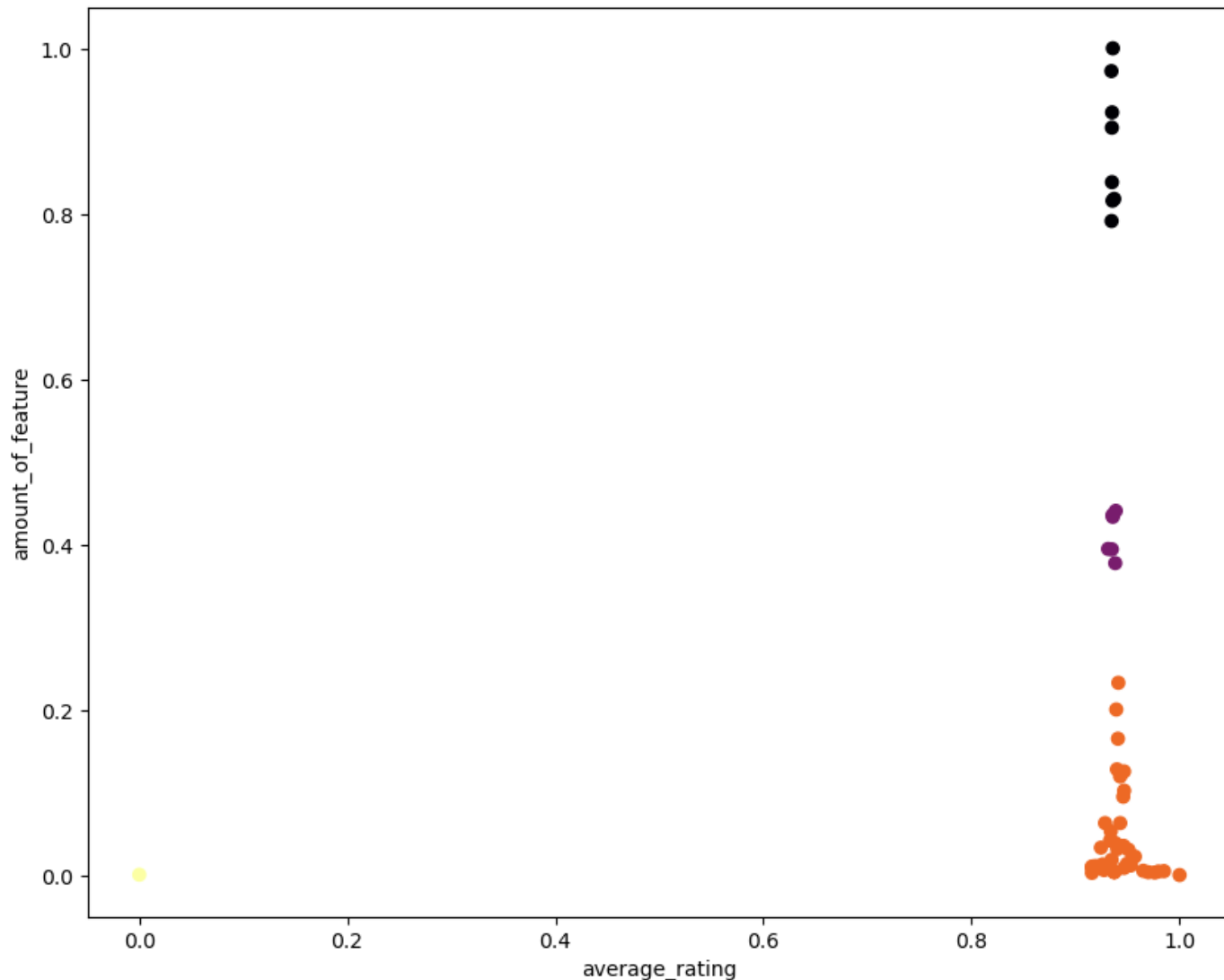
## Дендограмма



## Алгоритм работы:

1. Присвоение порядкового номера каждой доп. услуге
2. Расчет среднего рейтинга заведений, содержащих определенную доп услугу
3. Нормирование данных
4. Вычисление евклидова расстояния
5. Определение оптимального количества кластеров методом “локтя”
6. Построение дендограммы, диаграммы распределения кластеров

# Диаграмма распределения кластеров



- Плотность – величина взаимосвязи объектов
- Самую большую плотность имеет оранжевый цвет
- Желтым цветом обозначен едва ли заметный выброс

\*Каждая точка на графике – отдельная дополнительная услуга

# Вывод из кластерного анализа

Выявили 4 кластера, первый из которых является искомым набором доп. услуг чаще всего встречающихся у кафе с хорошим и отличным рейтингом

1 кластер - 8 дополнительных услуг

2 кластер - 6 дополнительных услуг

3 кластер - 48 дополнительных услуг

4 кластер - 1 дополнительная услуга – выброс

	average_rating	amount_of_feature	amount_of_features_in_klaster
KMeans			
1	4.358876	2049.625000	8
2	4.360673	957.833333	6
3	4.387619	89.062500	48
4	0.000000	0.000000	0



# Вывод из кластерного анализа

Набор дополнительных услуг, который стоит внедрять в кафе, чтобы достичь хорошего рейтинга:

- |                           |                             |
|---------------------------|-----------------------------|
| 1.food_delivery           | - доставка                  |
| 2.breakfast               | - завтрак                   |
| 3.takeaway                | - еда на вынос              |
| 4.summer_terrace          | - летняя терраса            |
| 5.wi-fi                   | - вай-фай доступ            |
| 6.business_lunch          | - комплексный обед          |
| 7. payment_by_credit_card | - возможность оплаты картой |
| 8.coffee_to_go            | - кофе на вынос             |

Набор дополнительных услуг, который стоит внедрять в кафе, чтобы достичь отличного рейтинга (выше 4.5):

- |                          |  |
|--------------------------|--|
| 9. closed_for_quarantine | - меры безопасности в случае карантина           |
| 10. online_takeaway      | - возможность заказа на вынос онлайн             |
| 11. karaoke              | - караоке  |
| 12. special_menu         | - специальное меню                               |
| 13. sports_broadcasts    | - спортивные передачи                            |
| 18. wheelchair_access    | - доступ для людей с ограниченными возможностями |



# Результаты

1

Мы провели анализ рынка, целевой аудитории и выяснили, что продукт будет полезен потребителям и будет актуален для предпринимателей

2

Вывели список дополнительных услуг, которые должны быть в каждом кафе, если они хотят иметь хороший рейтинг.

Вывели список дополнительных услуг, которые потенциально могут повысить рейтинг заведению.

# Перспективы



создать интерфейс взаимодействия с пользователем (landing-page)



научиться самостоятельно парсить карты и собирать данные



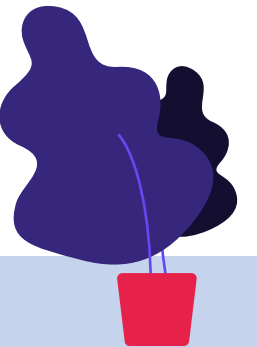
персонализировать проект под каждого пользователя



протестировать вторую модель решения и выявить достоинства/недостатки



доказать достоверность двух методов решения



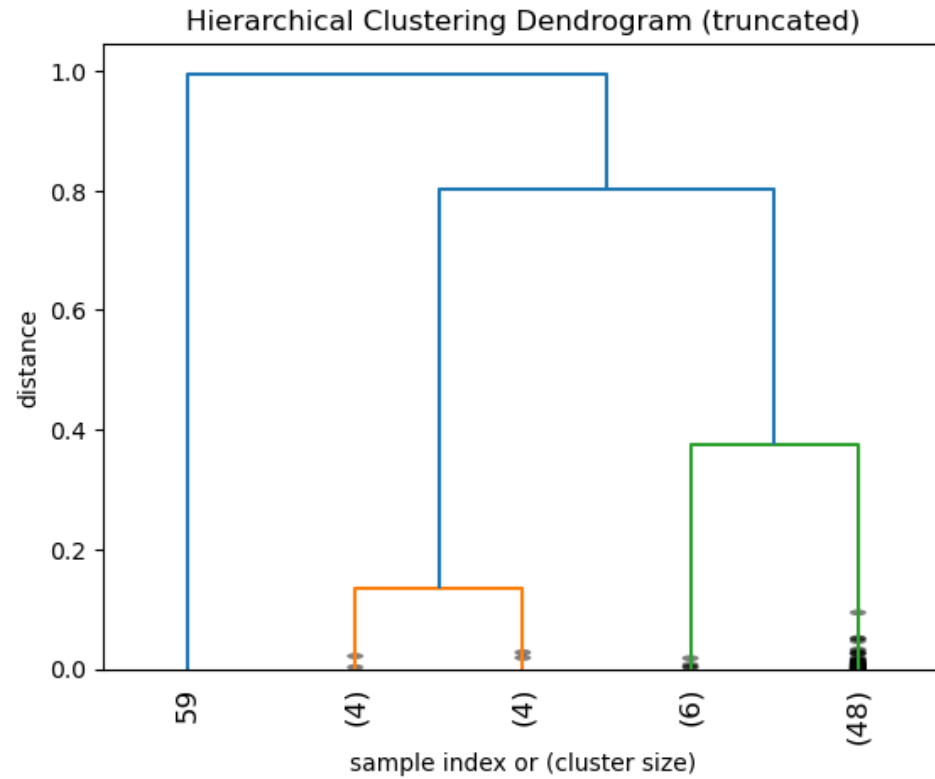
# Приложение

## Использованная литература

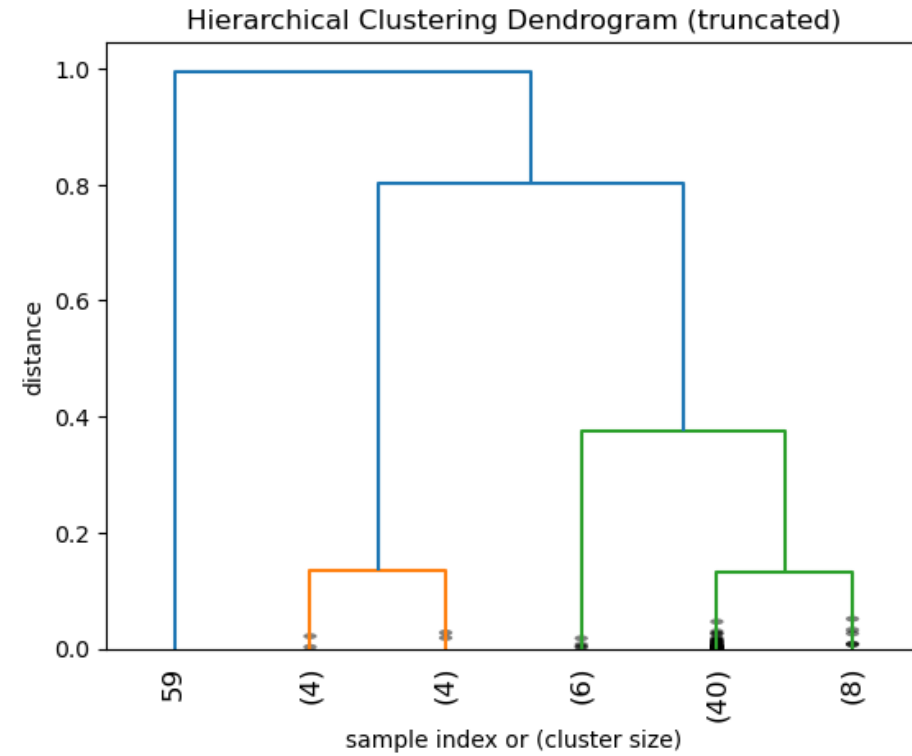
<a href="https://pntr.io/27facts">https://pntr.io/27facts</a>	сайт с исследованиями о значимости отзывов
<a href="https://thebell.io/amp/v-rossii-3-predprinimatelej-eto-namnogo-menshechem-v-mire">https://thebell.io/amp/v-rossii-3-predprinimatelej-eto-namnogo-menshechem-v-mire</a>	статистика предпринимателей в России
<a href="https://vc.ru/marketing/221088-kak-polozhitelnye-otzyvy-ubivayut-vashi-prodazhi">https://vc.ru/marketing/221088-kak-polozhitelnye-otzyvy-ubivayut-vashi-prodazhi</a>	прибыль зависит от рейтинга от рейтинга
<a href="https://www.rbc.ru/spb_sz/25/11/2020/5fbe22cb9a794707aa1d09c1">https://www.rbc.ru/spb_sz/25/11/2020/5fbe22cb9a794707aa1d09c1</a>	информация о упадочном бизнесе
<a href="https://iom.anketolog.ru/2021/02/15/kafe-restorany">https://iom.anketolog.ru/2021/02/15/kafe-restorany</a>	на что обращают внимание посетители при выборе заведения приема пищи
<a href="https://restteam.ru/consulting/audit-restorana?ysclid=ld0qibp98c561647791">https://restteam.ru/consulting/audit-restorana?ysclid=ld0qibp98c561647791</a>	один из игроков на рынке
<a href="https://info.mozg.rest">https://info.mozg.rest</a>	один из игроков на рынке
<a href="https://trade-drive.ru/automatization/rest/?ysclid=ldf39ja5jk85641997">https://trade-drive.ru/automatization/rest/?ysclid=ldf39ja5jk85641997</a>	один из игроков на рынке
<a href="https://yandex.ru/support/business-priority/manage/rating.html">https://yandex.ru/support/business-priority/manage/rating.html</a>	как формируются отзывы
<a href="https://sovcombank.ru/blog/biznesu/kak-rabotayut-otzivi-trendi-i-rekomendatsii/amp?utm_referrer=https%3A%2F%2Fwww.google.com%2F">https://sovcombank.ru/blog/biznesu/kak-rabotayut-otzivi-trendi-i-rekomendatsii/amp?utm_referrer=https%3A%2F%2Fwww.google.com%2F</a>	анализ понятия “хороший рейтинг”
<a href="https://business.aliexpress.ru/help/article/otzivi-pokupatelei">https://business.aliexpress.ru/help/article/otzivi-pokupatelei</a>	анализ понятия “хороший рейтинг”
<a href="https://habr.com/ru/post/578754/">https://habr.com/ru/post/578754/</a>	нормальность
<a href="http://statistica.ru/theory/klasterizatsiya-metod-k-srednikh/">http://statistica.ru/theory/klasterizatsiya-metod-k-srednikh/</a>	кластерный анализ

# Приложение

Для 5 кластеров



Для 6 кластеров



# Приложение

## Использованные формулы

### Формула Евклидова расстояния

$$\rho(x, y) = \|x - y\| = \sqrt{\sum_{p=1}^n (x_p - y_p)^2}, \text{ где } x, y \in R^n$$

### Метод k-средних

Ряд наблюдений  $(x^{(1)}, x^{(2)}, \dots, x^{(m)})$ ,  $x^{(j)} \in R^n$ .

Метод k-средних разделяет  $m$  наблюдений на  $k$  групп (или кластеров) ( $k \leq m$ )  $S = \{S_1, S_2, \dots, S_k\}$ , чтобы минимизировать суммарное квадратичное отклонение точек кластеров от центроидов этих кластеров:  $\min \left[ \sum_{i=1}^k \sum_{x \in S_i} \|x^{(j)} - \mu_i\|^2 \right]$ , где  $x^{(j)} \in R^n$ ,  $\mu_i \in R^n$ ,  $\mu_i$  - центроид для кластера  $S_i$ .

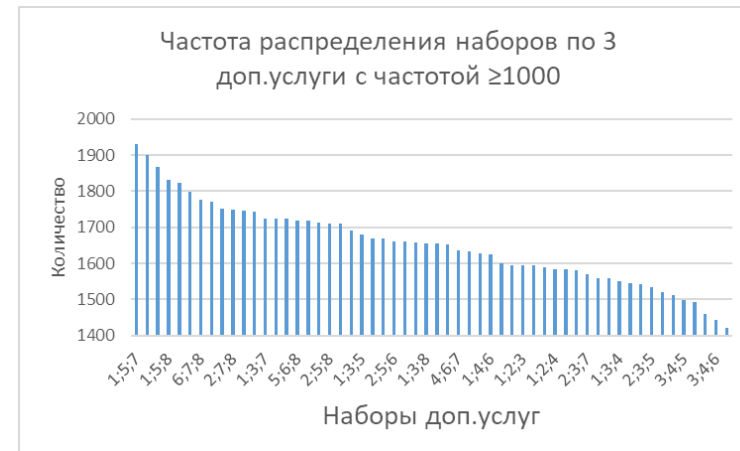
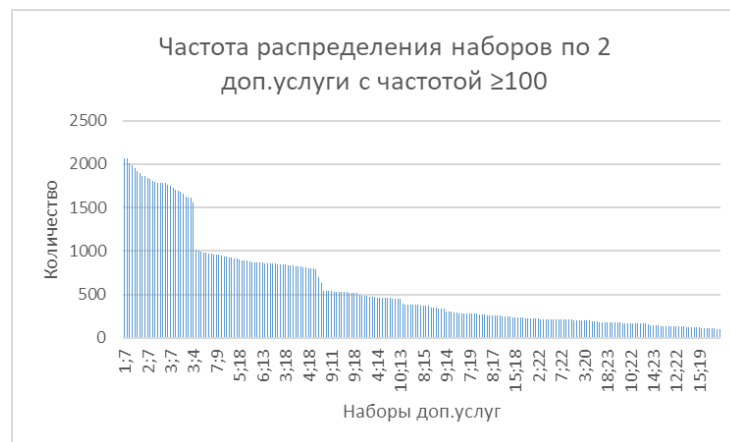
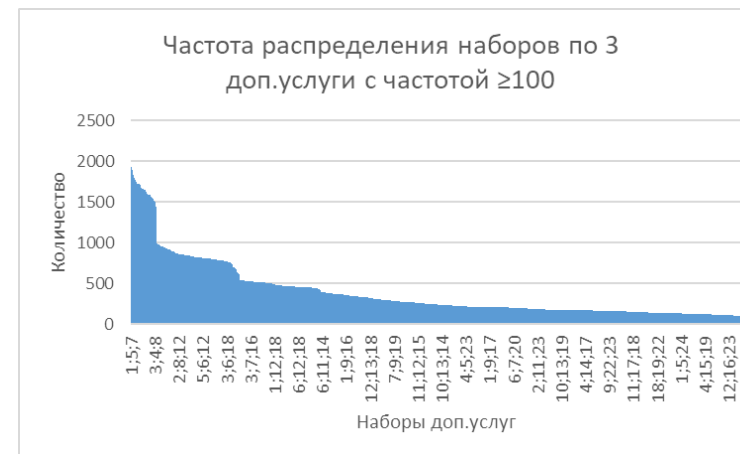
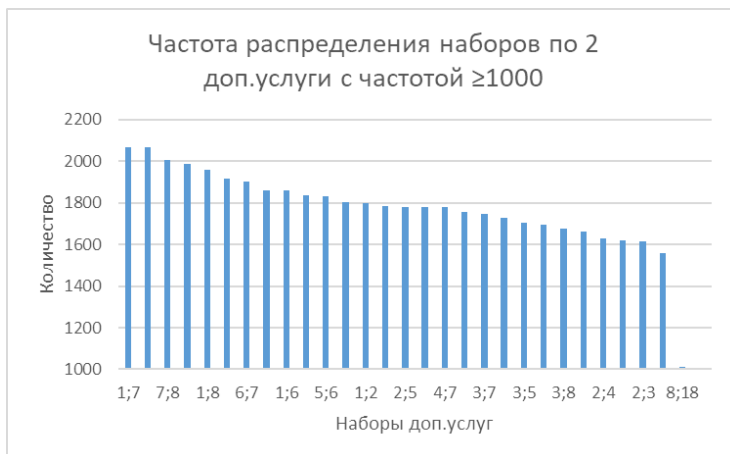
### Метод локтя (кластеризация)

Метод состоит в построении графика объясненной вариации в зависимости от количества кластеров и выборе локтя кривой в качестве количества кластеров для использования.

# Приложение

## Анализ другим методом

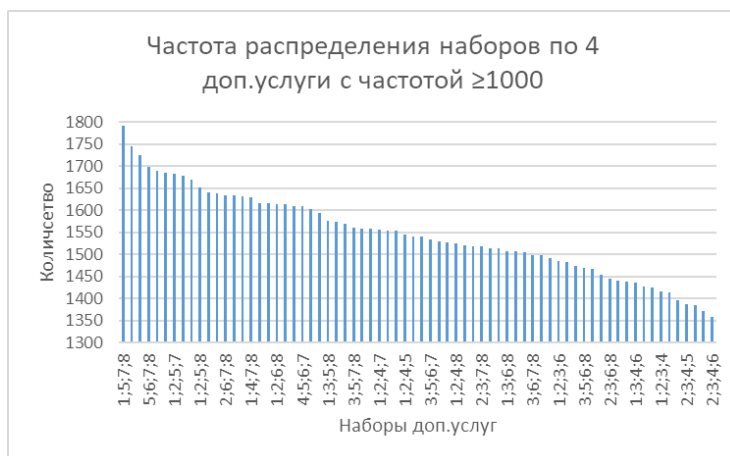
Заведения с хорошим и отличным рейтингом (4.0 и выше)



# Приложение

## Анализ другим методом

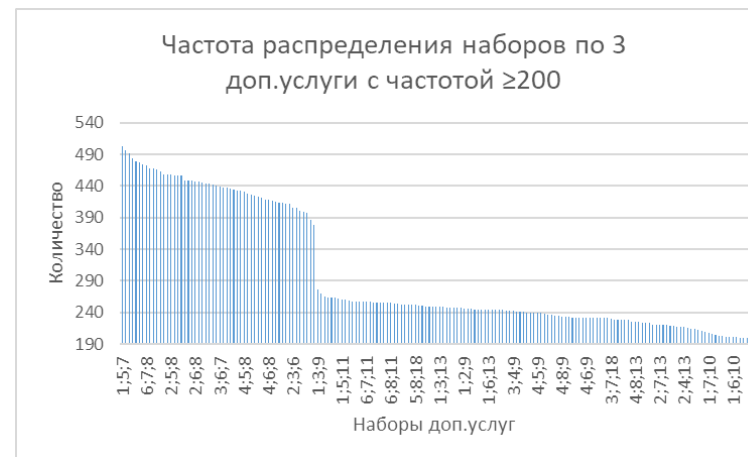
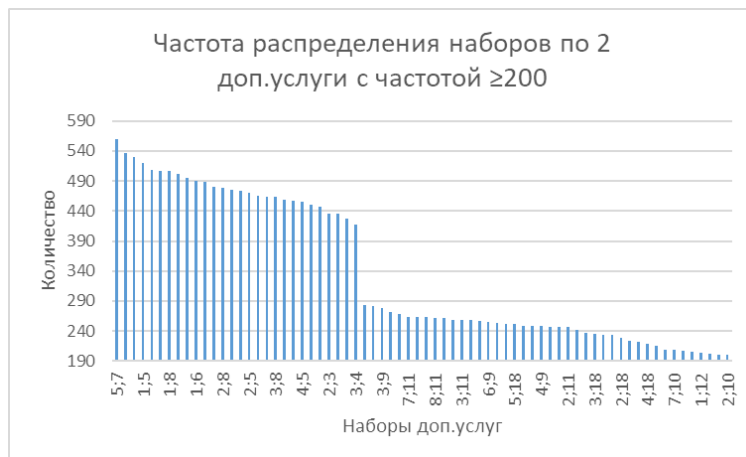
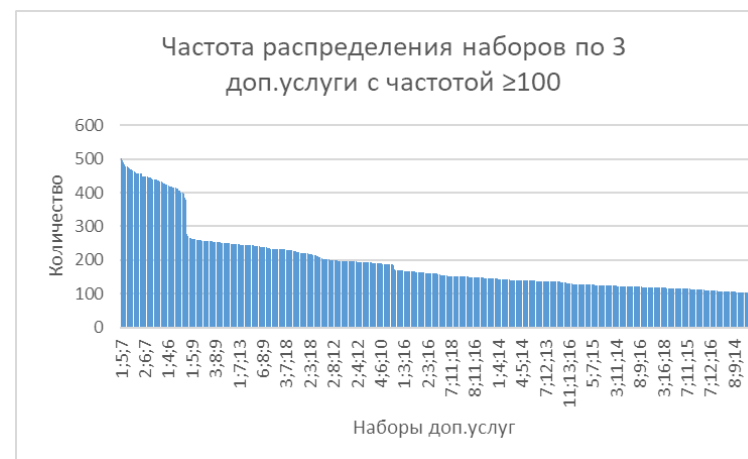
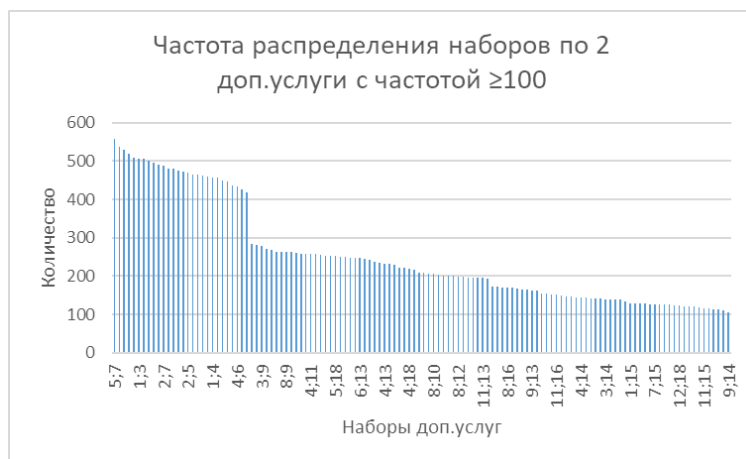
Заведения с хорошим и отличным рейтингом (4.0 и выше)



# Приложение

## Анализ другим методом

### Заведения с отличным рейтингом (4.5 и выше)

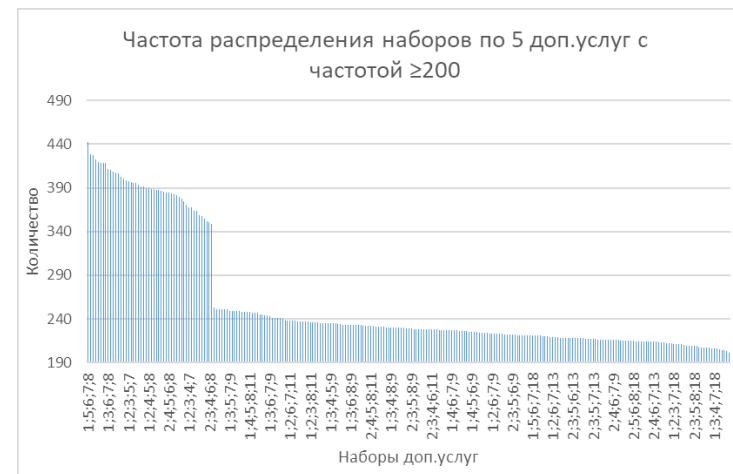
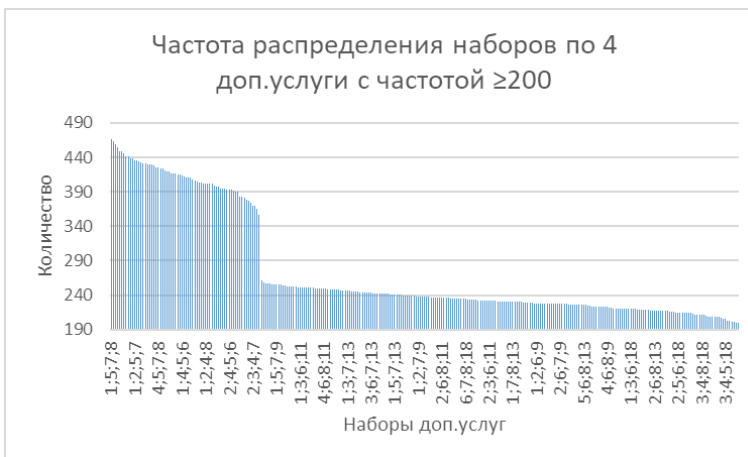
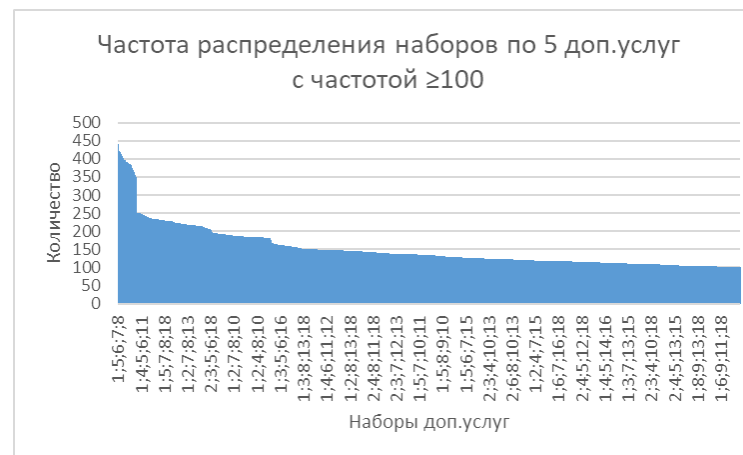




# Приложение

## Анализ другим методом

### Заведения с отличным рейтингом (4.5 и выше)



# Приложение

## Результаты опроса

вопросы/люди	Тимофей	Гриша	Денис	Саша Сухорукова	Варя	Саша Кудряшова	Лена	Саша Рояла
1 На какие параметры вы смотрите при выборе кафе?	на меню на улице, на красивую вывеску, наличие кафе на местности	отзывы, меню, расположение	отзывы, меню, средний чек	отзывы, мен, наличие определенных видов блюд	оценка близких, отзывы	цена, отзывы друзей	отзывы в интернете	отзывы
2 Какими доп. Услугами вы пользуетесь внутри кафе?	бизнес-ланч, парковка	завтраки, wi-fi, туалет	wi-fi	wi-fi, безналичная оплата	акции	бизнес-ланчи	бизнес-ланчи	безналичная оплата, безналичные чаевые
3 Важно ли для вас наличие неиспользуемых доп. Услуг?	нет	да	нет	нет	да, поднимает уровень кафе в глазах, это хорошо, даже если не нужно	нет	да, мешает звук	нет
4 На что вы обращаете внимание при посещении кафе?	вывеска	дизайн, контингент, персонал, частота	атмосфера, музыка, контингент	дизайн, количество мест, национальная кухня	атмосфера, контингент	атмосфера, громкая музыка не нравится	дизайн, атмосфера	контингент, вкус, качество подачи, дизайн
5 Почему вы выбрали именно кафе, а не другие заведения?	шаговая доступность	расположение, быстрое принятие пищи, меньше времени на ожидание	дешевизна, расположение	ожидание меньше	чаще рестораны, дешевизна, неформальная обстановка	дешевизна, вкусно	дешевизна, национальные блюда, единообразие кухни	доступность
6 Будете ли вы рекомендовать кафе друзьям, если вам там понравилось?	да	да	да	да	да	да	да	да
Оставьте ли вы отзыв, если вам все понравилось?	да	нет	нет	возможно	нет	нет	нет	нет
Оставьте ли вы отрицательный отзыв, если вам не понравилось?	только в крайнем случае	да	да	да	нет, но пожалуйсь окружению	нет	нет	нет