

## Оглавление

1. Введение.....	1
2. Общая информация.....	2
3. Подробнее о тональности.....	6
4. Эмоджи.....	10
5. Части речи.....	14
6. Анализ твитов на брань.....	18

### 1. Введение.

В данной статье мы рассмотрели данные со службы поддержки клиентов сервиса такси Яндекс. Общее количество обращений в службу поддержки составляет 3504 твита.

Средняя длина обращения составляет 26 слова.

Тональность твитов были разделена на 3 категории:

- негативные (0.0)
- нейтральные (0.5)
- позитивные (1.0)

Количество негативных твитов – 1606, что в 3 раза больше позитивных твитов. Далее по количеству идут нейтральные твиты (1345), а затем позитивные твиты — 553.

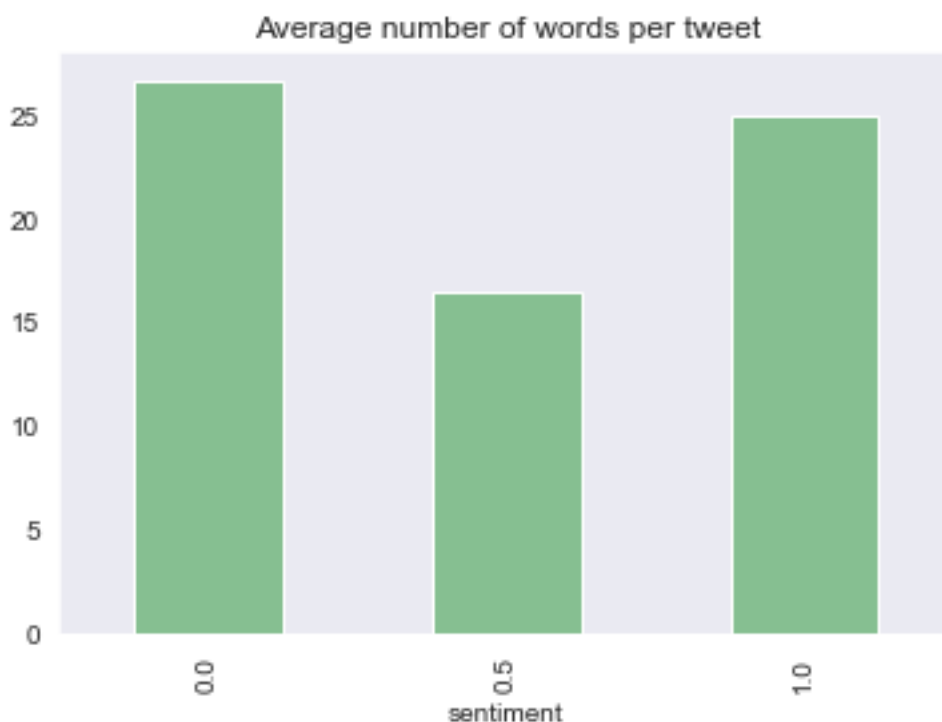
Наиболее распространенное отрицательное слово – «водитель» , «такси» и «цена», а наиболее употребляемой частью речи является существительное.



Описательная статистика слов в твитах по типу тональности:

	count	mean	std	min	25%	50%	75%	max
sentiment								
0.0	1606.0	26.615193	12.221226	1.0	17.0	26.0	37.0	53.0
0.5	1345.0	16.418587	11.034555	1.0	9.0	13.0	21.0	53.0
1.0	553.0	24.947559	12.603447	3.0	14.0	23.0	35.0	59.0

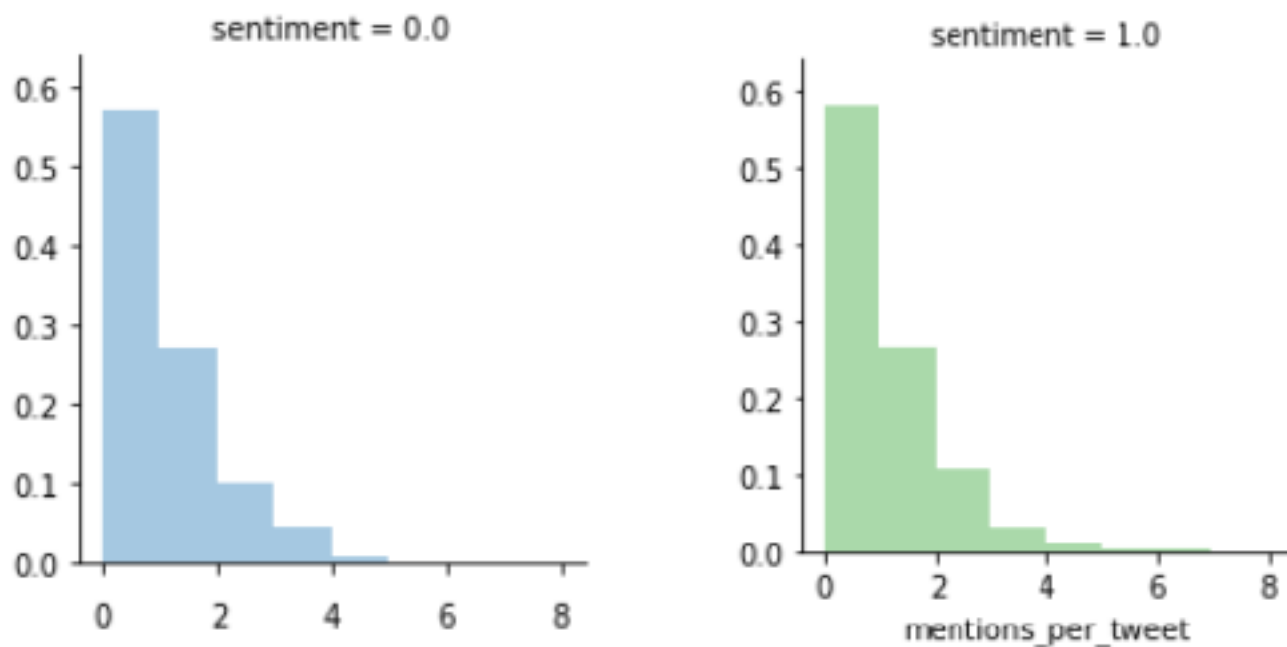
Минимальное количество слова в негативном и нейтральном обращениях - 1, максимальное - 53.



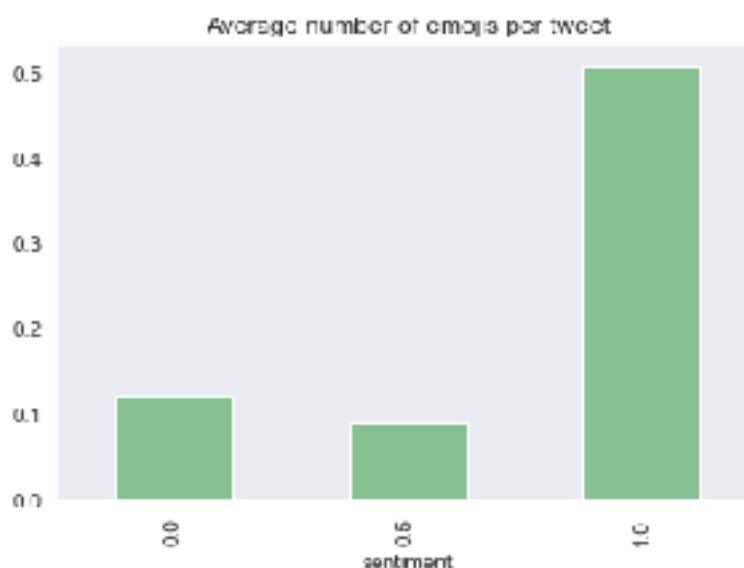
Граф. 3. Среднее количество слов в твитах, разделенных по типу тональности

Среднее количество слова в негативных 26 и 24 в позитивных твитах. Нейтральные твиты имеют в среднем 16 слов в каждом твите.

Подсчет количества упоминаний (@) в твите:



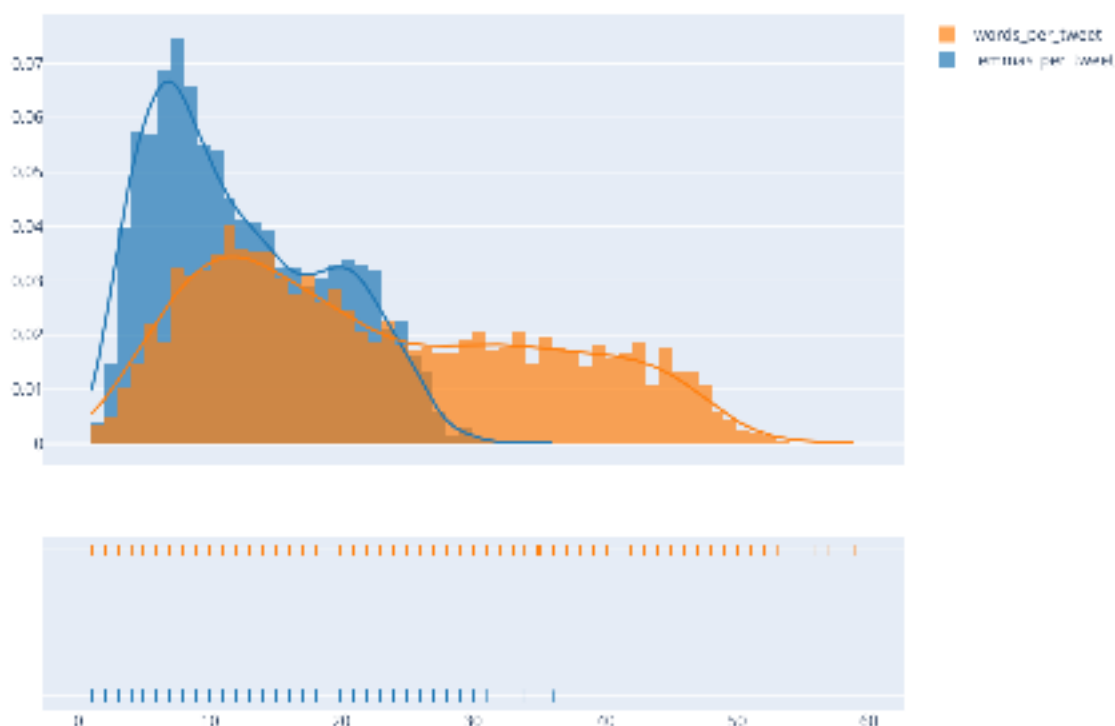
В 57% негативных и 58% во всех позитивных твитах не были отмечены аккаунты @yandex\_go и/или @yandexsupport



Граф 4. Среднее количество эмоджи в твитах, разделенных по типу тональности

Твиты с позитивным тоном имеют тенденцию использовать больше эмоджи в своих твитах. При ручной валидации становится понятно, что очень часто такие твиты являются саркастичными, и прямо сейчас мы работаем над решением, которое позволит отличать сарказм и иронию от позитивных твитов.

Distribution of number of words and keywords



*Граф. 5. Распределение количества слов и количества ключевых слов на твит.*

Большинство твитов имеют 5-8 ключевых слов. Диапазон слов в 30% твитах составляет от 7 до 18.

### 3.Подробнее о тональности.



*Рис 6. График распределения негативных твитов по датам*



*Рис 7. График распределения позитивных твитов по датам*

В целом на графе не наблюдается значительное увеличение или уменьшение как отрицательных, так и положительных твитов.

С 11 по 13 февраля был пик негативных обращении в службу поддержки (45 негативных). 15 января был зафиксирован максимум положительных твитов за 1 день (20).

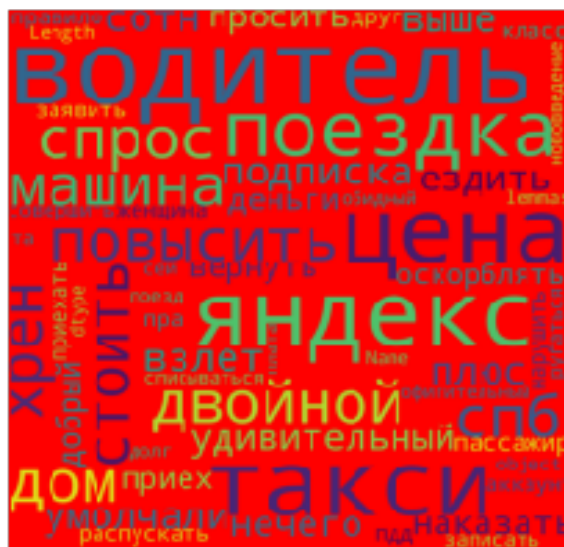


Рис 8. Визуализация наиболее употребляемых слов в негативных твитах

Возможно, слова “Водитель” и “Такси” можно объединить. Вместе слова «водитель» и «такси» встречаются в негативных твитах 1941 раз, («водитель» : 1452, «такси»: 489). Это составляет 8% от всех негативных слов.

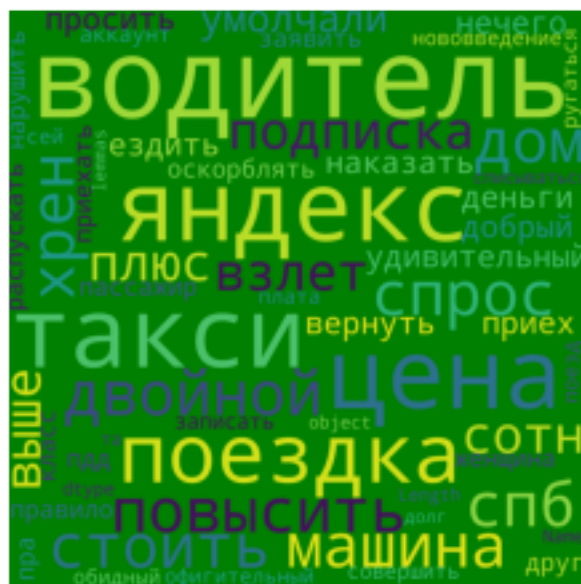


Рис 9. Визуализация наиболее употребляемых слов в позитивных твитах.

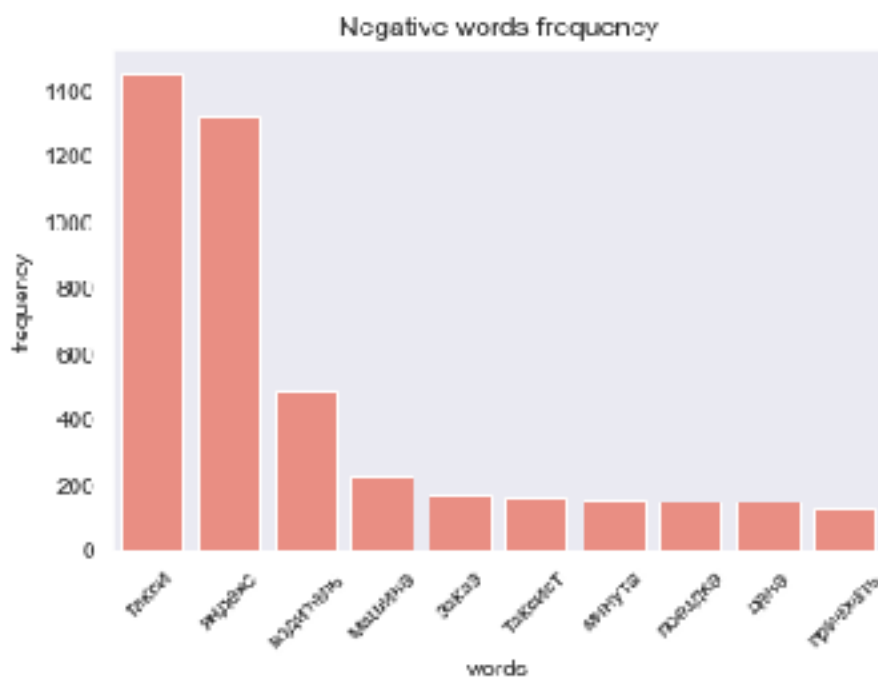


Рис 10. Частота наиболее распространенных отрицательных слов.

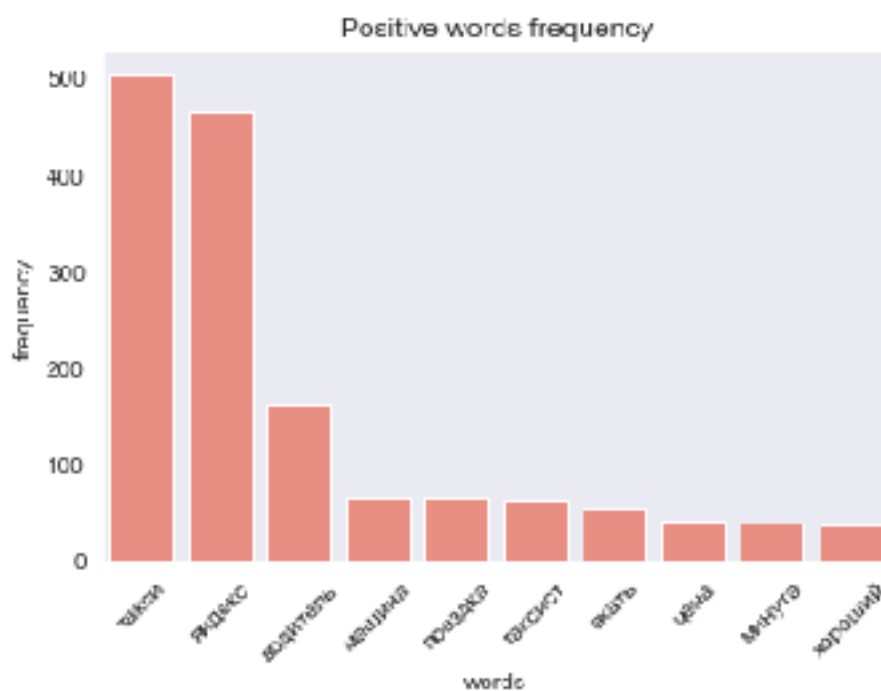


Рис 11. Частота наиболее распространенных позитивных слов



Как мы видим, как в графе отрицательных, так и положительных слов, есть много одинаковые слова. Ниже приведен список слов, которые входят в список 20 наиболее распространенных положительных слов, но не входят в список 20 наиболее распространенных отрицательных слов.

[хороший, еда, рейтинг, знать, музыка, ездить]

#### 4. Эмоджи.

Как мы уже отметили ранее, часть твитов ошибочно была отнесена к положительным. Есть вероятность что это связано с эмоджи которые выражают иронию. Для того чтобы подтвердить это, был проведён анализ эмоджи как положительных, так и отрицательных твитов.

### 10 самых распространенных эмоджи из негативных твитов с соответствующими расшифрованными ярлыками.

{ '😭': 'loudly crying face' }	6
{ '😂': 'face with tears of joy' }	5
{ '😬': 'pouting face' }	5
{ '😮': 'face with symbols on mouth' }	5
{ '🤡': 'clown face' }	3
{ '😍': 'smiling face with hearts' }	3
{ '👍': 'thumbs up' }	3
{ '😆': 'rolling on the floor laughing' }	3
{ '🙄': 'woman facepalming: light skin tone' }	3
{ '⚡': 'high voltage' }	2

Ниже приведены 2 эмоджи, которые могут вызвать сомнение относительно их присутствия в отрицательных твитах.

1. “smiling face with hearts” (был использован 3 раза)

## 2. “thumbs up” (был использован 3 раза)

*Полный текст твитов где были использованы эти 2 эмоджи:*

'Сегодня мне попался оочень добрый таксист. Вы встречали такого таксиста, чтобы он Вам привез тушь, которая выпала из косметички??? Я даже не заметила, что выпала. Прodelать такой путь обратно, потратить бензин ради косметоса. Желая чтобы это добро ему вернулось. #Яндекстакси 🥰'

'бля как мне нравится что яндекс повысил цену на такси в три нахуй раза 🥰'

'Ребят, у вас есть истории про неадекватных водителей такси? Хочется послушать. Просто у меня только такие и попадаются, по-моему. Например, сейчас на меня наорали из-за (внимание!) плохой погоды и стали угрожать, если я плохо оценю поездку. ЯндексТакси, соси жопу 🥰'

'Яндекс такси..... Вы совсем афигели. (Рядом стоит машина и не принимает), а что бы вы понимали я живу в не большем городе и на морозе стоять ждать водителя который в 6 км от тебя это класс 👍'

'заказала яндекс такси приехал подождал 2 минуты и уехал а я ему по случайности ещё и 5 звёзд поставила, потом 7 минут ждала другую машину хотя по карте он стоял на месте и никуда не ехал👍 а потом нажал что на месте хотя ещё ехал меня вообще бесит когда так делают'

'Прикольно, первый раз увидел что в яндекс такси указан что водитель плохо слышит. У него ещё и батарейка на слуховом аппарате села. Главное довез быстро и точно по адресу 👍'

Как мы видим, не всегда смайлы с хорошим смыслом могут быть использованы в качестве положительного комментария. 4 из 6 комментариев являются тому подтверждением.

### 10 самых распространенных эмоджи из позитивных твитов с соответствующими расшифрованными ярлыками.

{': 'face with tears of joy'}	10
{': 'smiling face with sunglasses'}	8
{': 'smiling face with hearts'}	6
{': 'slightly smiling face'}	5
{': 'grinning face with sweat'}	5
{': 'grinning face'}	5
{': 'clown face'}	5
{': 'beaming face with smiling eyes'}	4
{': 'face with rolling eyes'}	4
{': 'loudly crying face'}	4
{': 'woman shrugging'}	4
{': 'smiling face with smiling eyes'}	3

Отмеченные ранее 2 эмоджи также могут вызвать сомнение относительно их присутствия в положительных твитах.

1. “Clown face” (был использован 5 раз)
2. “Loudly crying face” (был использован 4 раза)

'@tvoitubik После домогательства водителя Яндекс.такси и открытого бездействия Яндекса — я лучше буду на заднем ездить или вообще пешком ходить 🤡'

'Итак, вызвал такси для поездки на 5 минут за 160₽, во время поездки изменил адрес на соседний дом. Знаете сколько @citymobiltaxi насчитал дополнительно? Почти 300 рублей)))) 🤡 Яндекс при таких изменениях вообще не меняет цену <https://t.co/DtGIzgDSKd>'

'обожаю яндекс такси за: приедет через 13 минут \*через минуту\* вас ожидает 🤡'

'@we\_are\_vkusvill ваша служба поддержки до сих пор нигде не ответила 🤡 Даже яндекс такси прибежали в реплаи, хотя я их не тегала Я очень люблю ваш магазин и с доставкой никогда не было проблем, но именно работа службы поддержки у вас очень ухудшилась, а отклики на вакансии не смотрите'

Смайлик “clown face” в позитивных твитах - это яркий сигнал о том, что произошла ошибка с обозначением тональности.

'почему мы уже второй раз заказываем яндекс такси а к нам приезжает машина с наклейками убера 🤡'

'я забыла сумку у такси и мы минут 20 ходили по супермаркету и только после я заметила отсутствие сумки позвонила водителю а он как оказалось все это время оставался рядом с реми и ввсматривал нас чтобы отдать сумку 🤡 спасибо яндекс такси и этому водителю'

'мой рейтинг пользователя в яндекс такси 4,98/5. я в шоке. почему так низко?😭'

Некоторые твиты были ошибочно отмечены как положительные, но исходя от анализа выше, можно сказать что эмоджи не является основной причиной данного события, так как количество ошибочно обозначенных твитов где присутствует эмоджи очень мало.

Всего лишь 8 твитов в которых имеются эмоджи выражающий негатив из всех положительных (553) твитов были ошибочно отмечены как положительные.

## 5. Части речи.

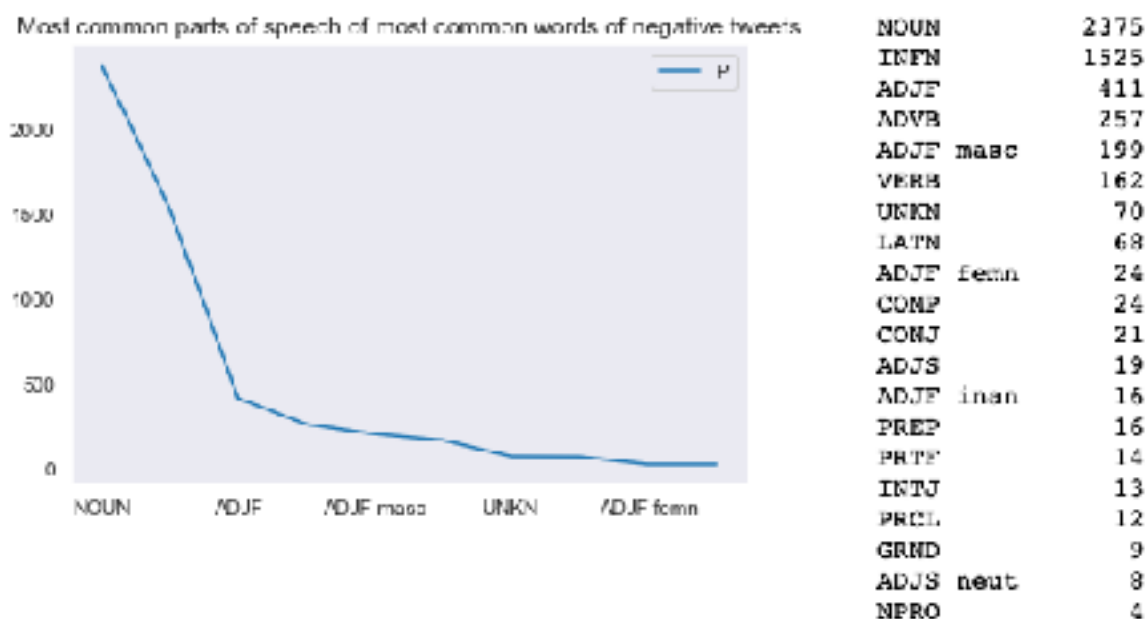


Рис 11. Части речи в часто встречаемых словах негативных твитов

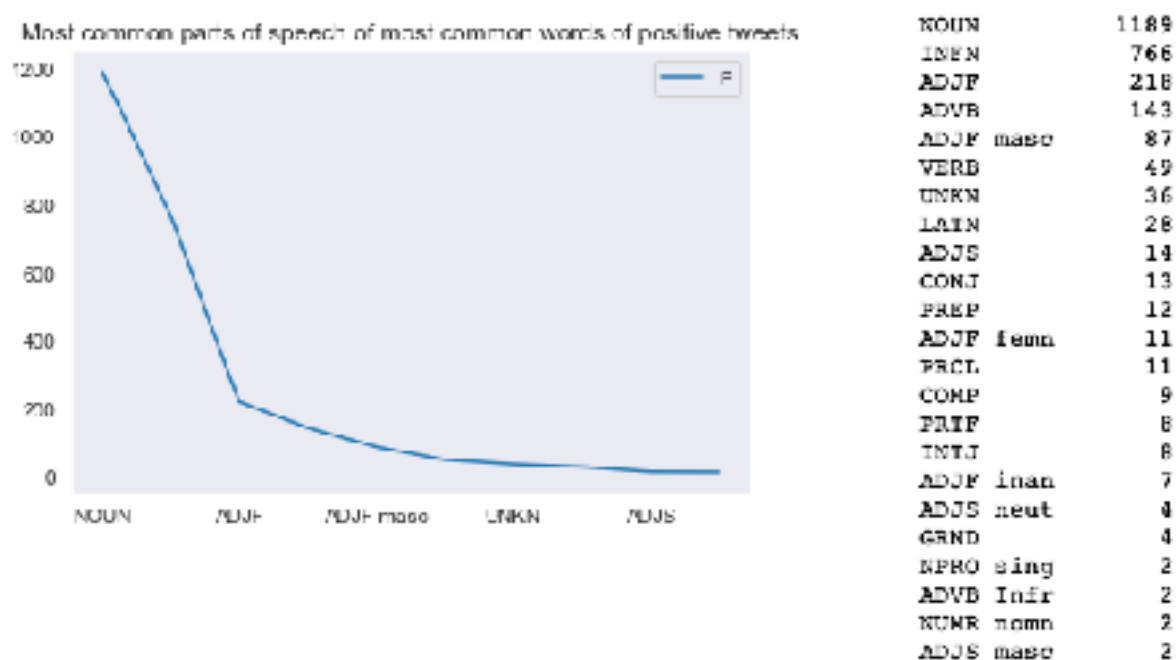


Рис 12. Части речи в наиболее часто встречаемых словах позитивных твитов.

Количество существительных (2375 в негативных и 1189 в позитивных твитах) сильно превышает остальные части речи.

Существительные составляют 45% от всех ключевых слов в негативном контексте (5276) и также от всех ключевых слов в позитивном контексте (2635)

Около 10% слов относятся к зарубежному языку или вовсе не являются словами.

Ниже приведено визуальное подтверждение о полученных данных, из чего можно сделать вывод о том, что большинство обращений в службу поддержки связаны с водителями, таймингом и стоимостью.

Наиболее часто употребляемые существительные

word	part of speech	count
такси	NOUN	1452
яндекс	NOUN	1321
водитель	NOUN	489
машина	NOUN	230
заказ	NOUN	174
таксист	NOUN	161
минута	NOUN	159
поездка	NOUN	153
цена	NOUN	149
рубль	NOUN	104

Наиболее часто употребляемые глаголы (инфинитив)

word	part of speech	count
приехать	INFN	131
ехать	INFN	127
ждать	INFN	83
написать	INFN	83
работать	INFN	83
знать	INFN	74
вызвать	INFN	71
заказать	INFN	63
пыльцать	INFN	63
стоять	INFN	63

Наиболее часто употребляемые прилагательные

word	part of speech	count
свой	ADJF	118
хороший	ADJF	52
самый	ADJF	44
нужный	ADJF	39
большой	ADJF	36
новый	ADJF	30
полный	ADJF	30
плохой	ADJF	23

Наиболее часто употребляемые наречие

word	part of speech	count
вообще	ADV	88
какой	ADV	33
вчераш	ADV	31
домой	ADV	29
просто	ADV	24
постоянно	ADV	18
бесплатно	ADV	15
точно	ADV	14

Выше приведено визуальное подтверждение о полученных данных, из чего можно сделать вывод о том, что большинство обращений в службу поддержки связаны с водителями,

таймингом и стоимостью. Наречия показывают экспрессивность твитов.

## 6. Анализ твитов на брань.

- Количество нецензурных слов в негативных твитах - 407.
- Количество уникальных слов - 77
- На каждый третий негативный твит приходится одно нецензурное слово
- Наиболее распространенное слово - “блять”
- Частота наиболее распространенного слово - 58
- Самая распространенная часть речи - существительное
- Максимальное количество нецензурных слов в день — 14

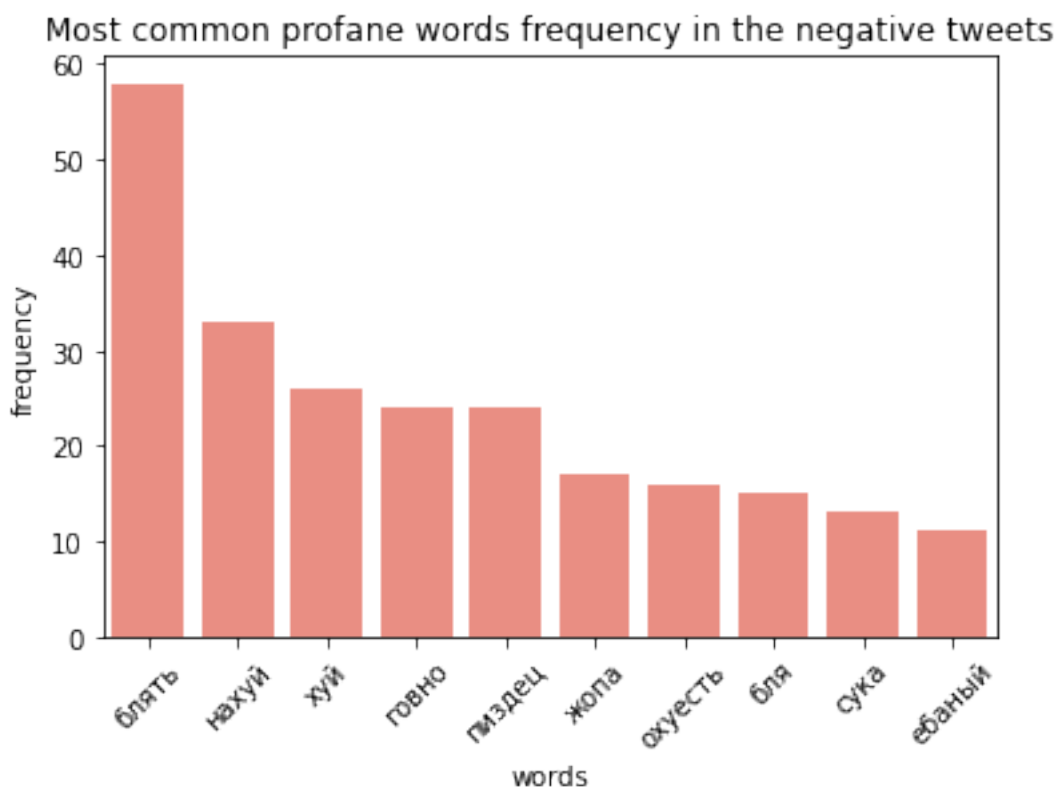
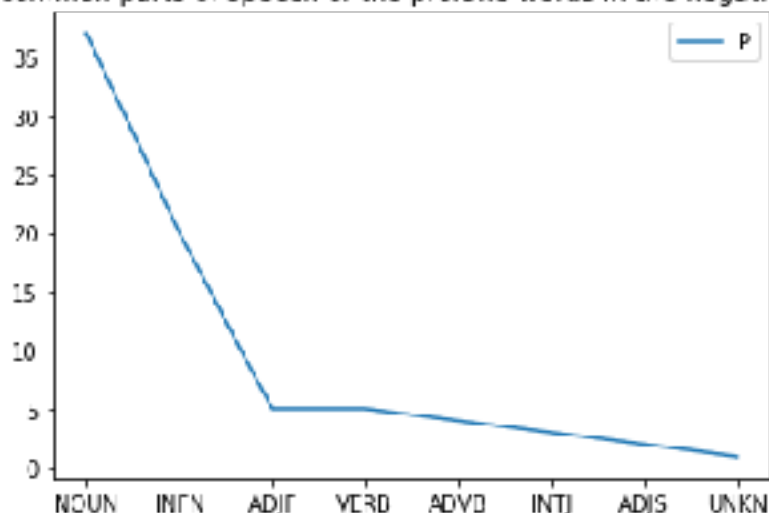


Рис 13. Визуальная репрезентация бранных слов в дотасете



## Графическая репрезентация распространения брани по частям речи

Most common parts of speech of the profane words in the negative tweets



NOUN	37
INFN	20
ADJF	5
VERB	5
ADVB	4
INTJ	3
ADJS	2
UNKN	1

Наиболее часто употребляемые существительные

word	part of speech	count
хуй	NOUN	28
говно	NOUN	24
пиздец	NOUN	24
жопа	NOUN	17
сука	NOUN	13
сбучий	NOUN	10
хрен	NOUN	9
хуйня	NOUN	9
хер	NOUN	7

Наиболее часто употребляемые глаголы инфинитив

word	part of speech	count
ебать	INFN	11
заебать	INFN	7
ахуеть	INFN	4
пиздеть	INFN	4
охуеть	INFN	4
наебать	INFN	3
выебываться	INFN	3
охуевать	INFN	3
ебануться	INFN	2

Наиболее часто употребляемые  
существительные

word	part of speech	count
такси	NOUN	503
яндекс	NOUN	465
водитель	NOUN	164
машина	NOUN	67
поездка	NOUN	67
таксист	NOUN	65
цена	NOUN	43
минута	NOUN	40
рейтинг	NOUN	28

Наиболее часто употребляемые  
глаголы (инфинитив)

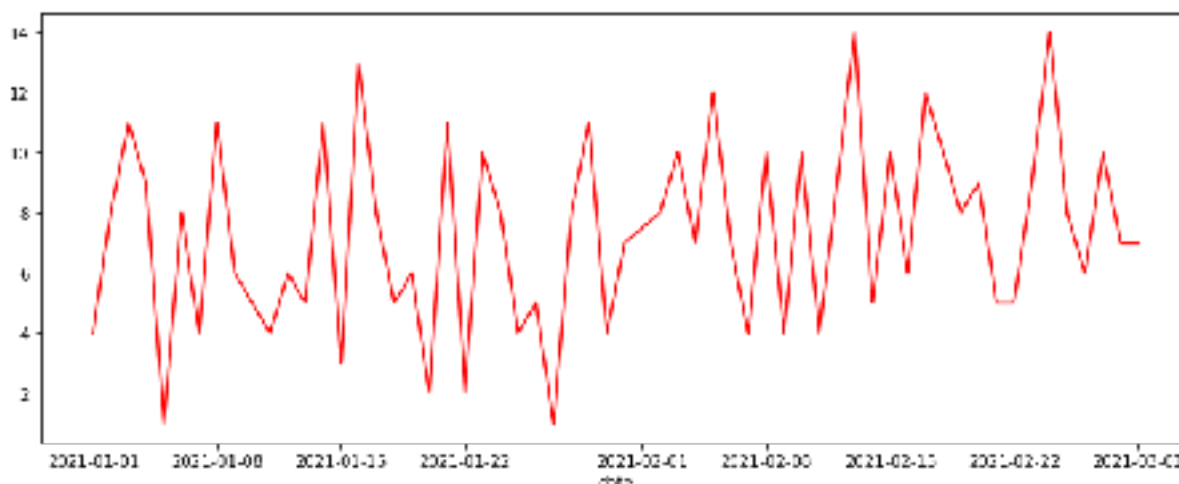
word	part of speech	count
ехать	INFN	54
приехать	INFN	32
знать	INFN	28
ездить	INFN	22
работать	INFN	22
заказать	INFN	21
вызвать	INFN	21
думать	INFN	20
любить	INFN	18

Наиболее часто употребляемые  
прилагательные

word	part of speech	count
хороший	ADJF	37
свой	ADJF	26
большой	ADJF	17
самый	ADJF	17
нужный	ADJF	15
интересный	ADJF	14
новый	ADJF	14
приятный	ADJF	11
высокий	ADJF	10
классный	ADJF	9

Наиболее часто употребляемые  
наречия

word	part of speech	count
вообще	ADVБ	24
впервые	ADVБ	8
сразу	ADVБ	7
домой	ADVБ	6
поэтому	ADVБ	6
дважды	ADVБ	5
обратно	ADVБ	5
сильно	ADVБ	5
круто	ADVБ	5
настолько	ADVБ	5



*Рис 14. График распределения негативных твитов в которых содержатся нецензурные слова по датам*

На 13 и 24 февраля приходится пик негативных твитов в которых были использованы нецензурная брань (14), тогда как 5 января и 26 января эта цифра сводится к 1.

## **Выводы**

- 1) Большинство негативных обращений содержат в себе претензию относительно таких параметров сервиса, как поведение водителя, время ожидания и изменение цены поездки.
- 2) Многие пользователи иронично используют эмоджи в негативных твитах.
- 3) Пользователи крайне экспрессивно отзываються о сервисе.
- 4) Многие пользователи используют иронию не только при помощи графических элементов, но и словесно, подразумевая под “классным сервисом” на самом деле плохой.