

Оглавление

1. Введение.....	1
2. Общая информация.....	2
3. Подробнее по типу тональностей.....	5

1 Введение.

В данной статье мы рассмотрели данные со службы поддержки клиентов сервиса такси Яндекс. Общее количество обращения в службу поддержки составляет 3504 твитов. Средняя длина обращения составляет 26 слова. Тональность твитов были разделены на 3 категории:

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

* Во всех трех категориях есть очень много слов, которые не должны были попадать в те категории, в которых они сейчас находятся.

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»

Наиболее употребляемой частью речи является существительное.

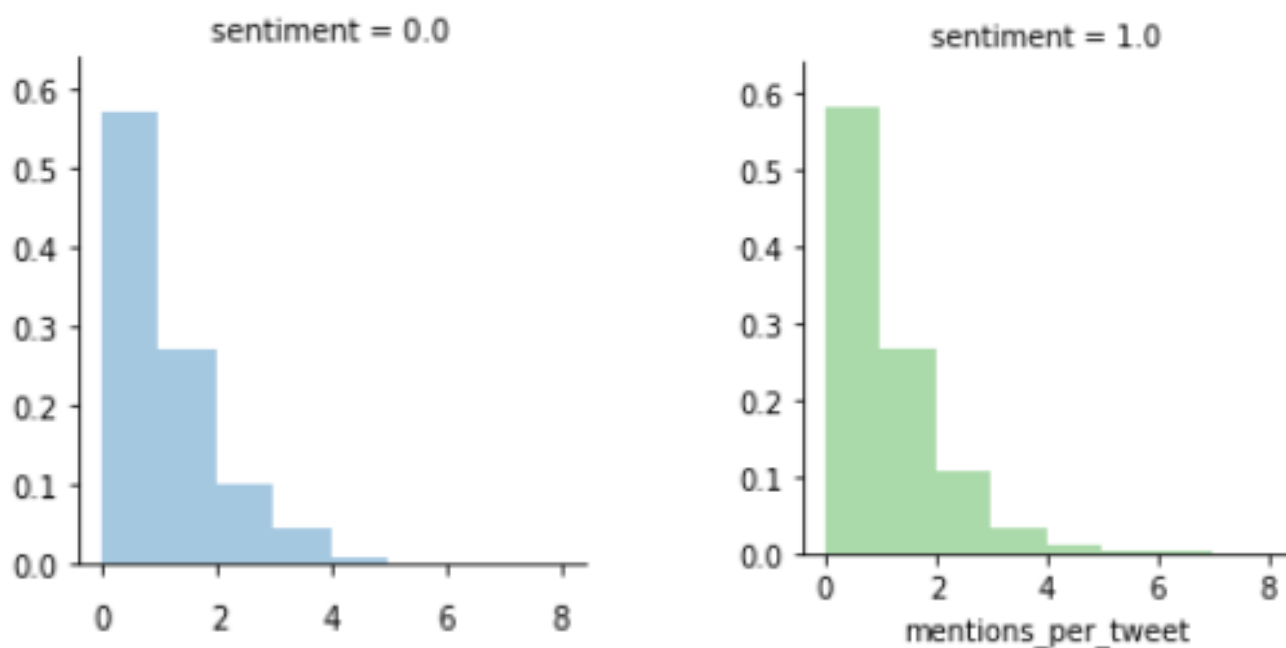
2 Общая информация.

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

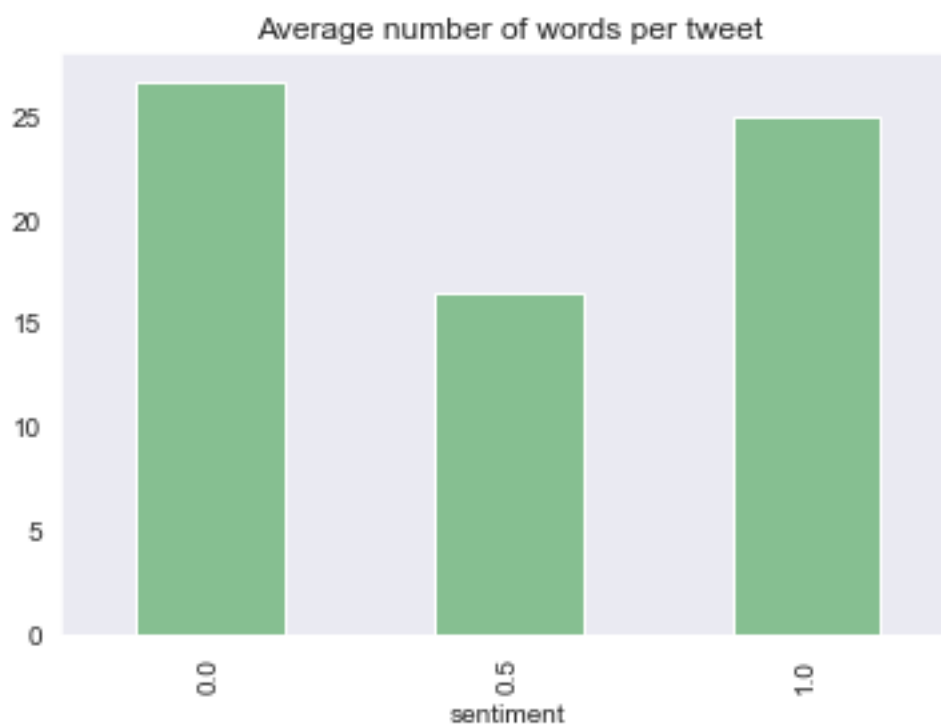
Минимальное количество слова в негативном и нейтральном обращениях - 1, максимальное - 53.

Подсчет количество упоминаний (@) в твите:



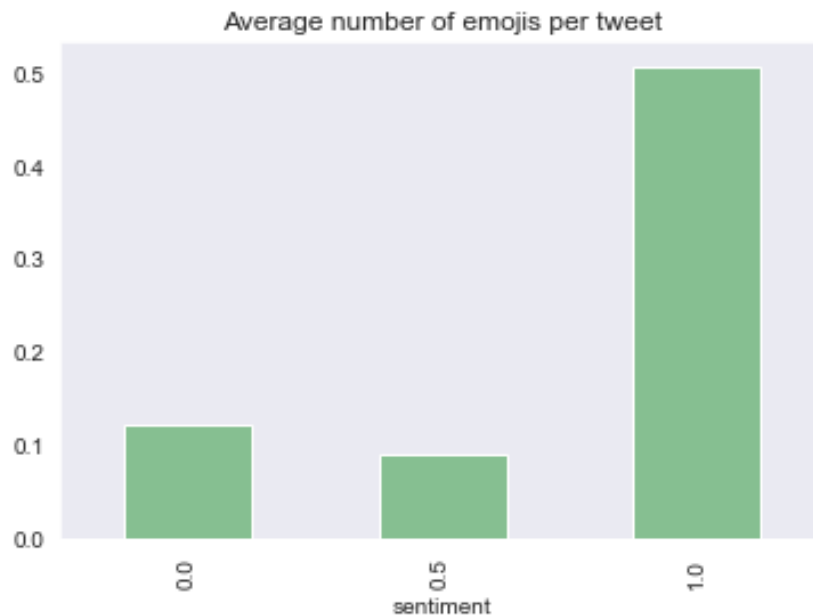
В 57% негативных и 58% во всех позитивных твитах не были отмечены аккаунты @yandex_go и/или @yandexsupport

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



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

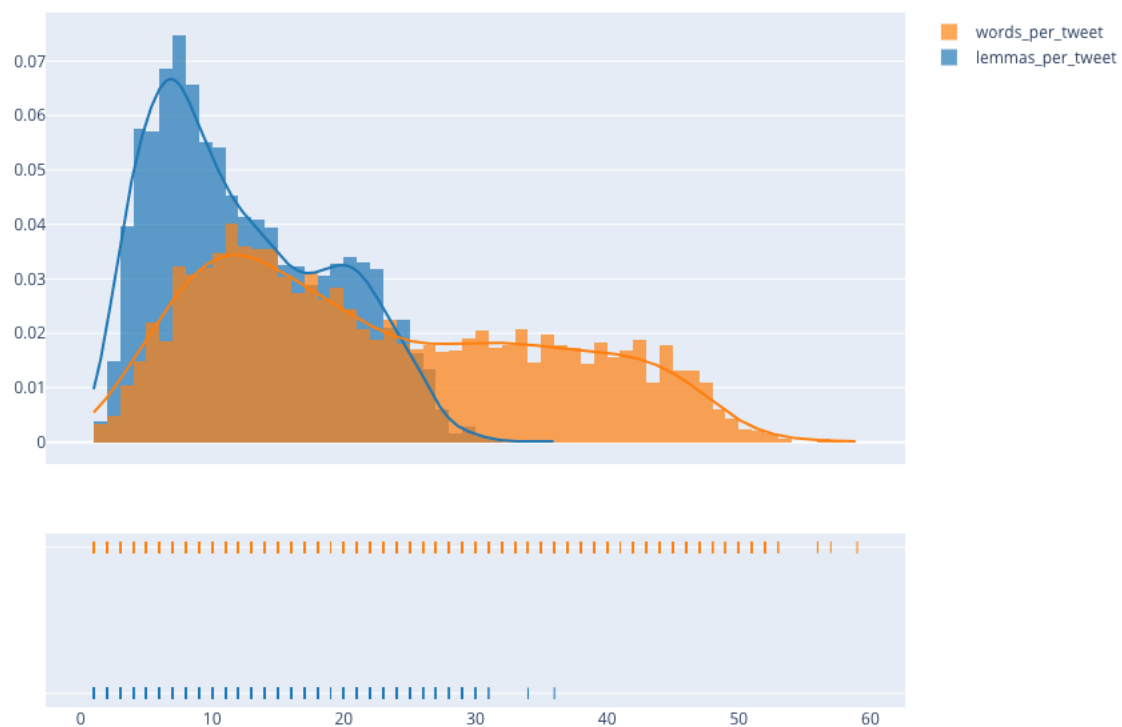
Граф 4. Среднее количество эмоджис в твитах, разделенных по типом тональности



Твиты с позитивным тоном имеют тенденцию больше использовать эмоджи в своих твитах.

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

Distribution of number of words and keywords



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

3.Подробнее по тональностям.

Рис 6. График распределение негативных твитов по датам

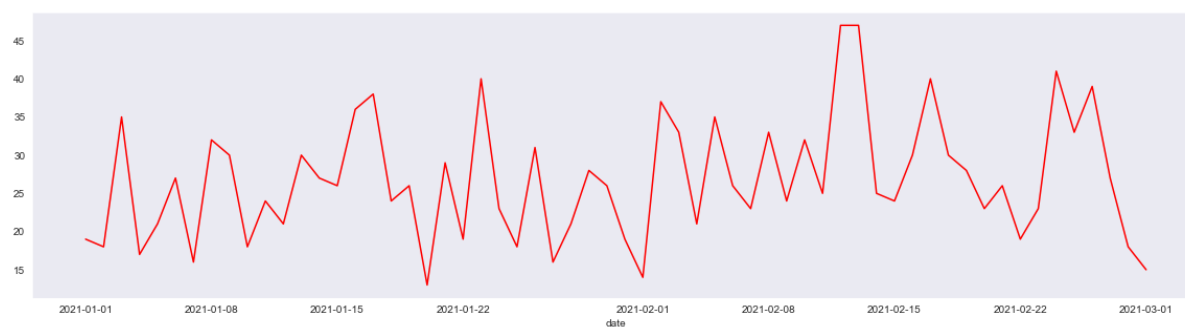
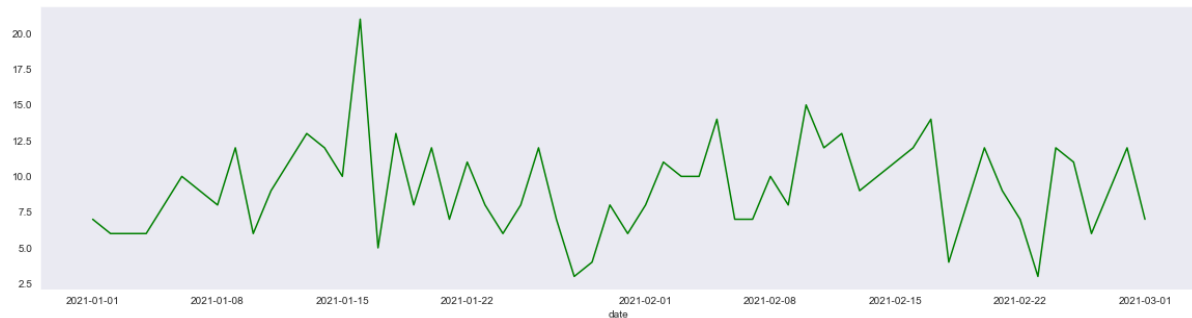


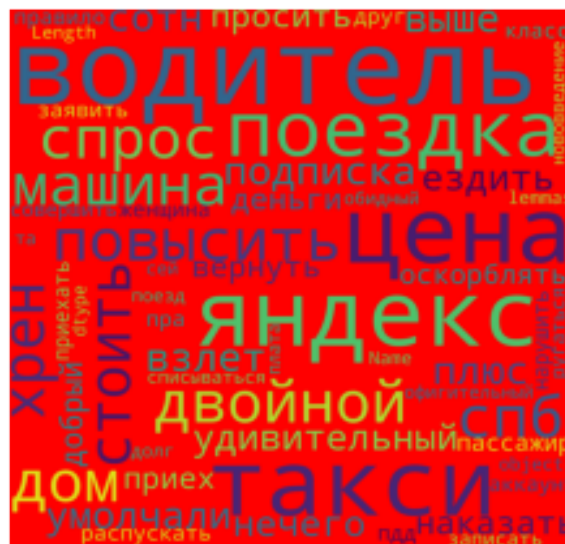
Рис 7. График распределение позитивных твитов по датам



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

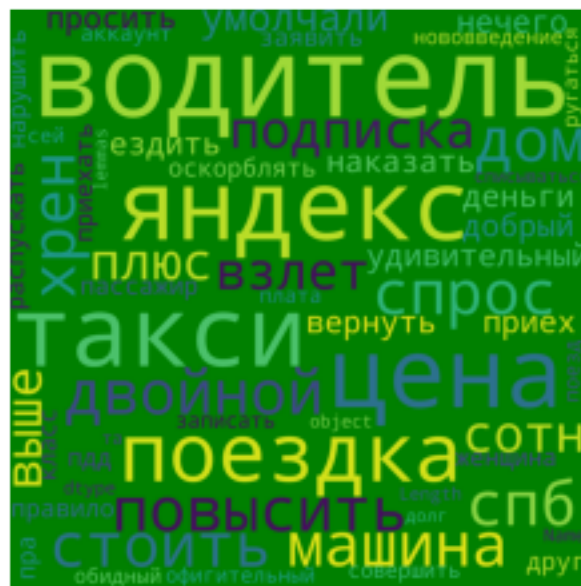


Рис 9. Частота наиболее распространенных отрицательных слов

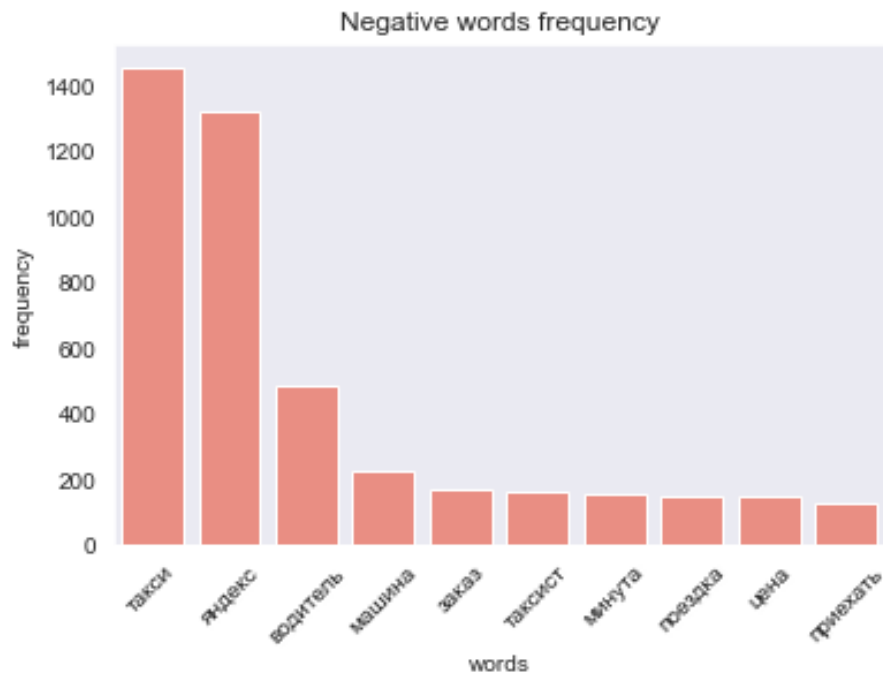


Рис 10. Частота наиболее распространенных позитивных слов

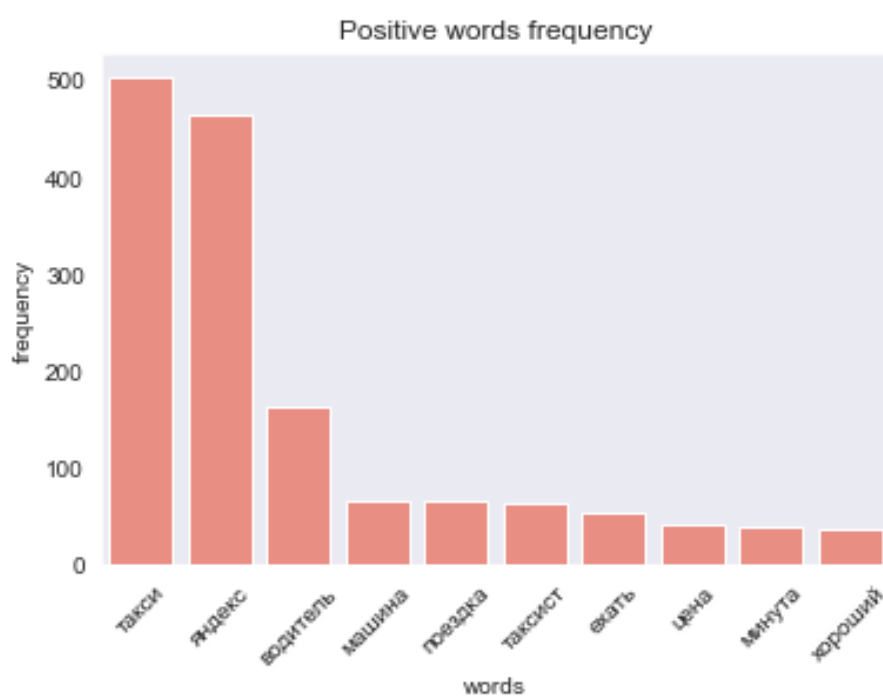


Рис 11. Части речи в часто встречаемых словах негативных твитов

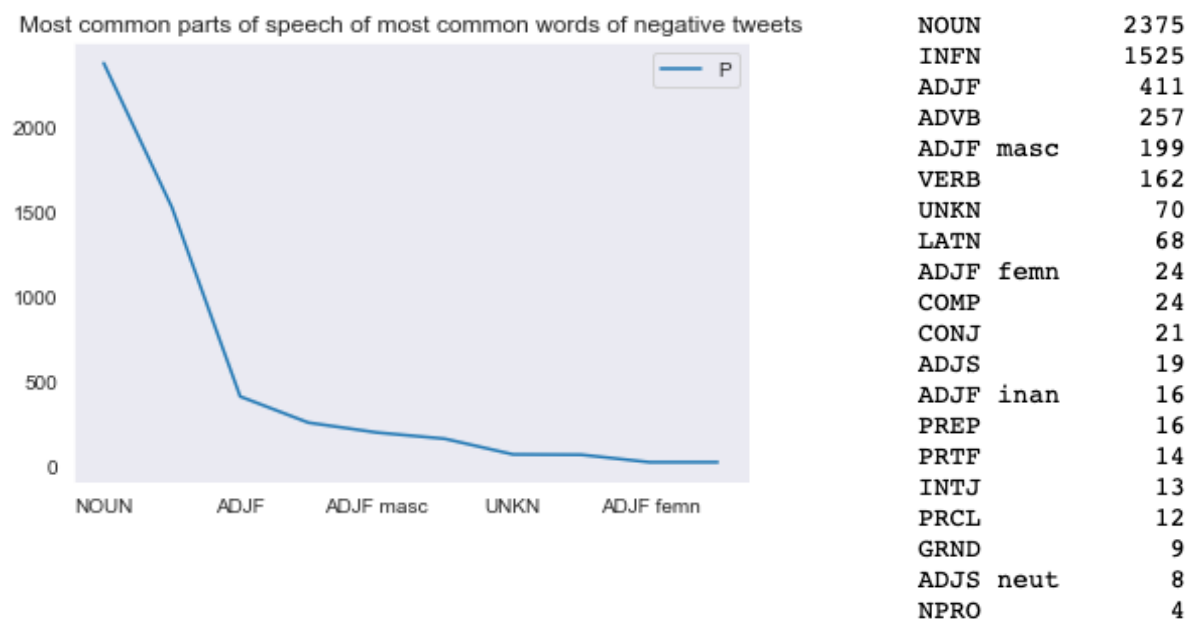
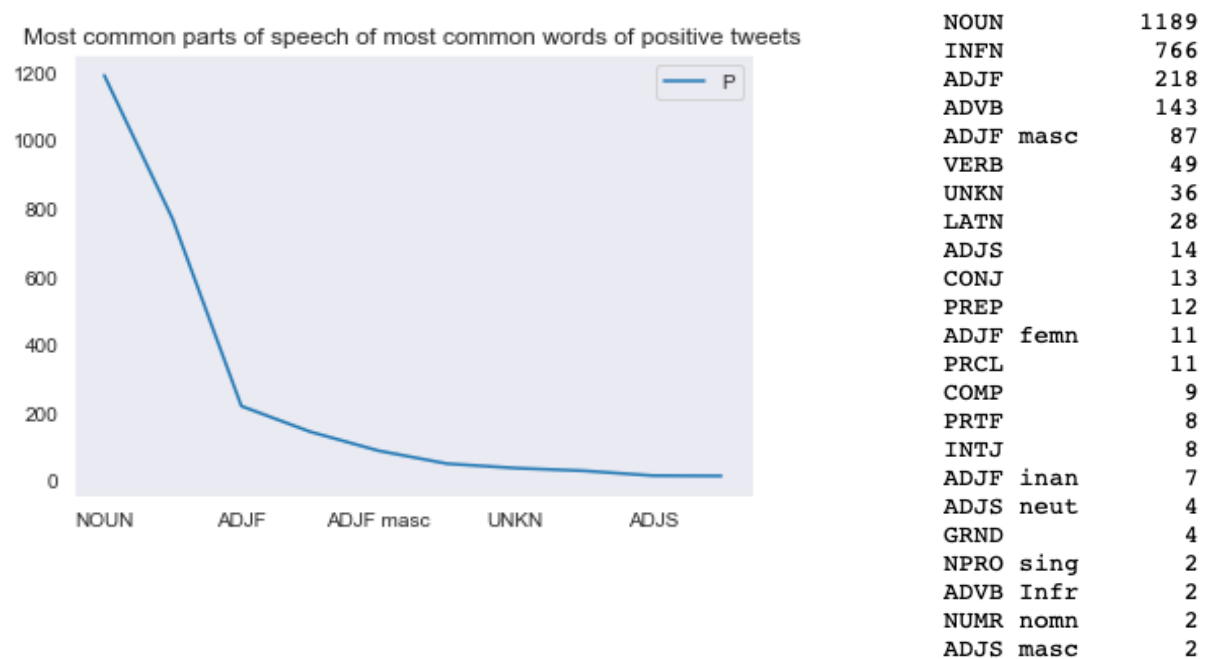


Рис 12. Части речи в часто встречаемых словах позитивных твитов.



Количества 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.