

Job for analysts

Инструменты — Python:

- ## Облака тэгөв

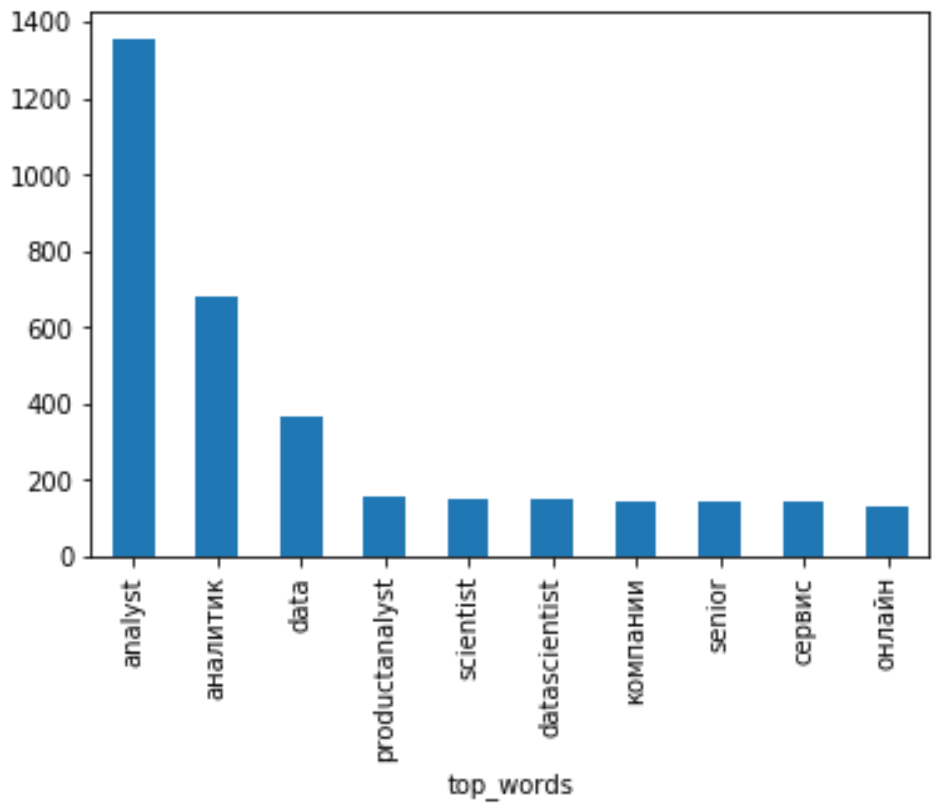
[illegible]

Данные результат стал возможен после удаления стоп слов и часто встречаемых слов не имеющих отношения к сути вакансии: 'вакансия','москва','что','это','так','вот','быть','как','в','—','к','на','за','было','требуемые','навыки','ищет','удалённо','специалиста','компания','откликаться','geekjob','описание','вакансии','ru','её','пост','санкт-петербург','санкт петербург','минск','личку','санкт','петербург','санктпетербург','оксана','пругьянова','подробнее'.

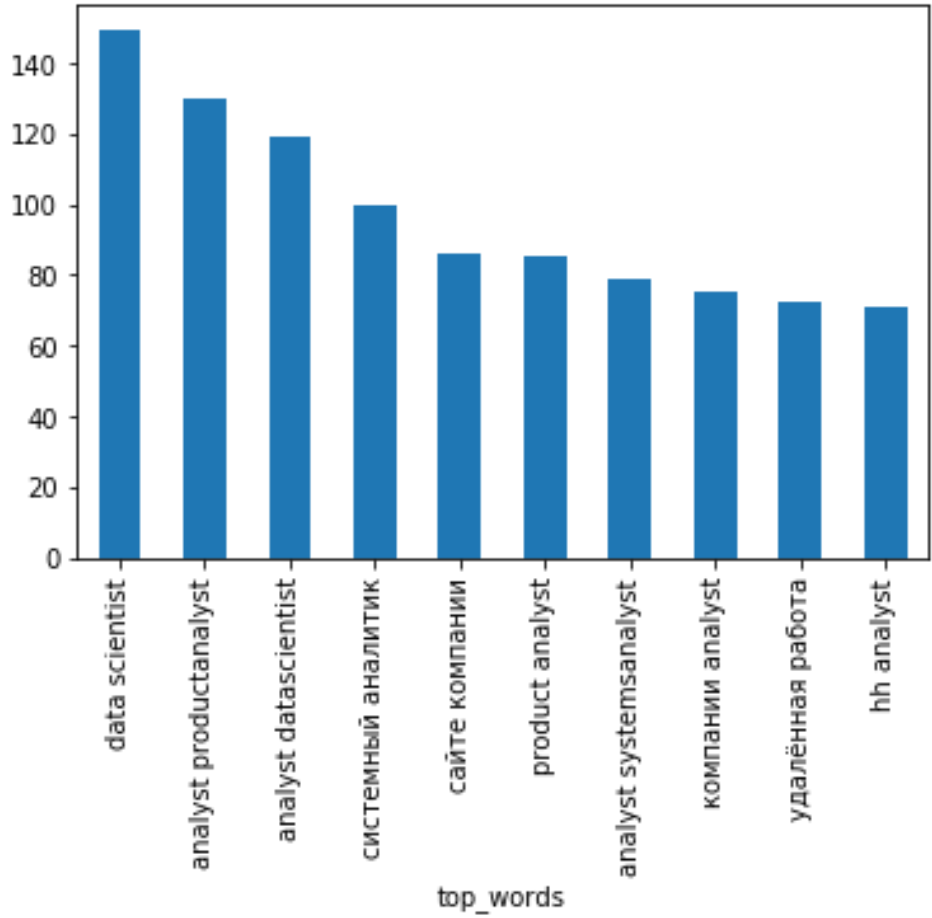
все слова переведены в нижний регистр

города исключены, что бы не вносить много шума в основную картину

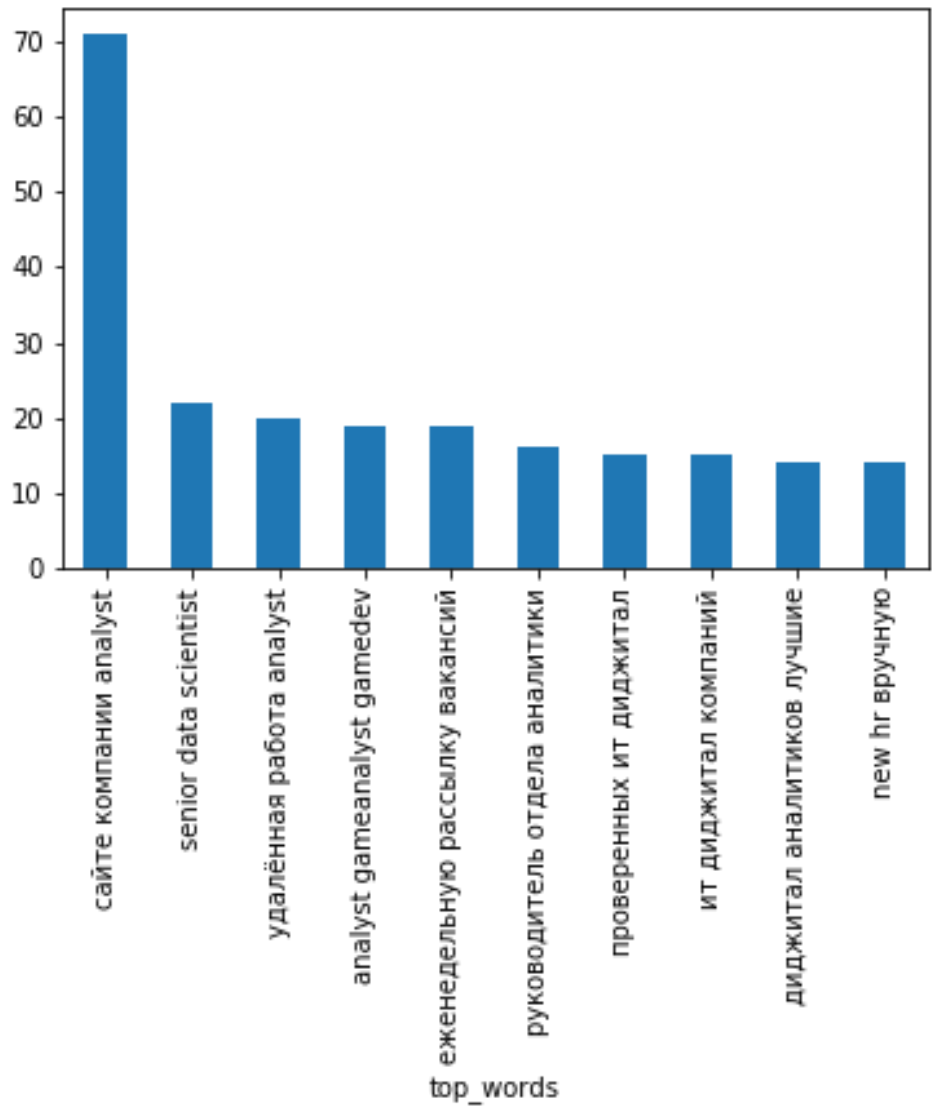
Топ - 10 слов



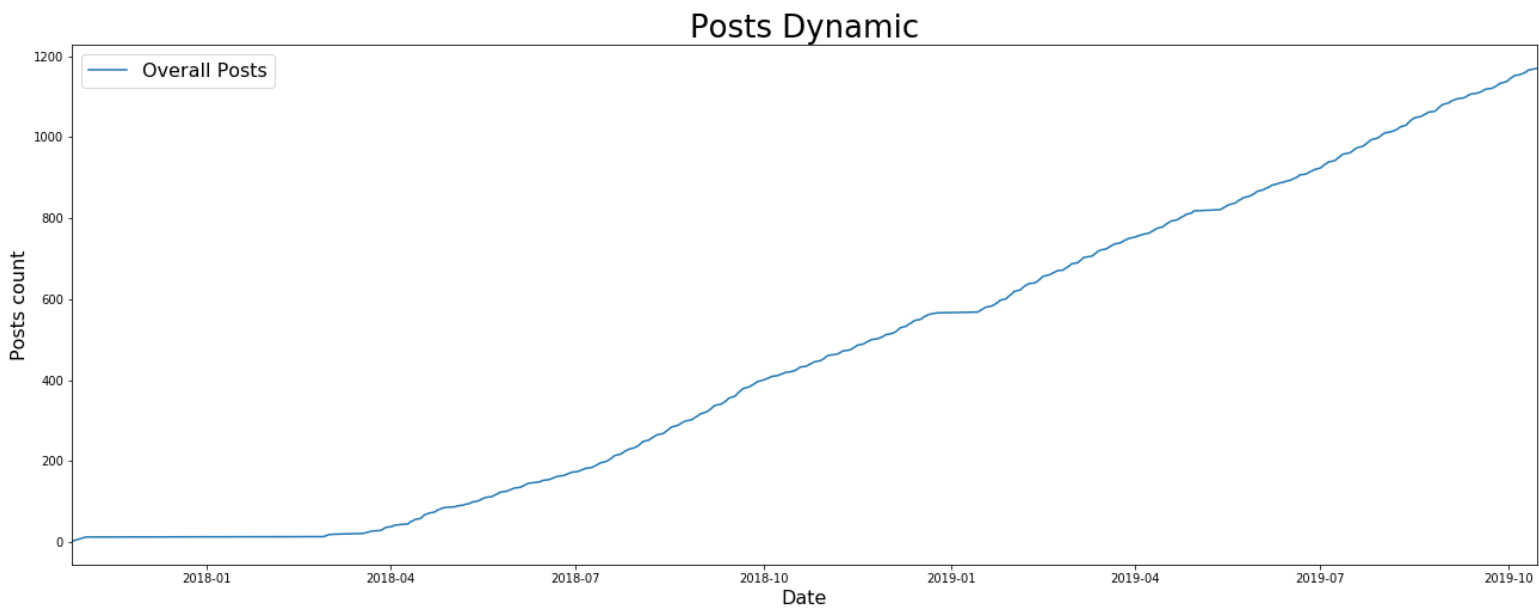
Топ - 10 пар слов



Топ - 10 троек слов

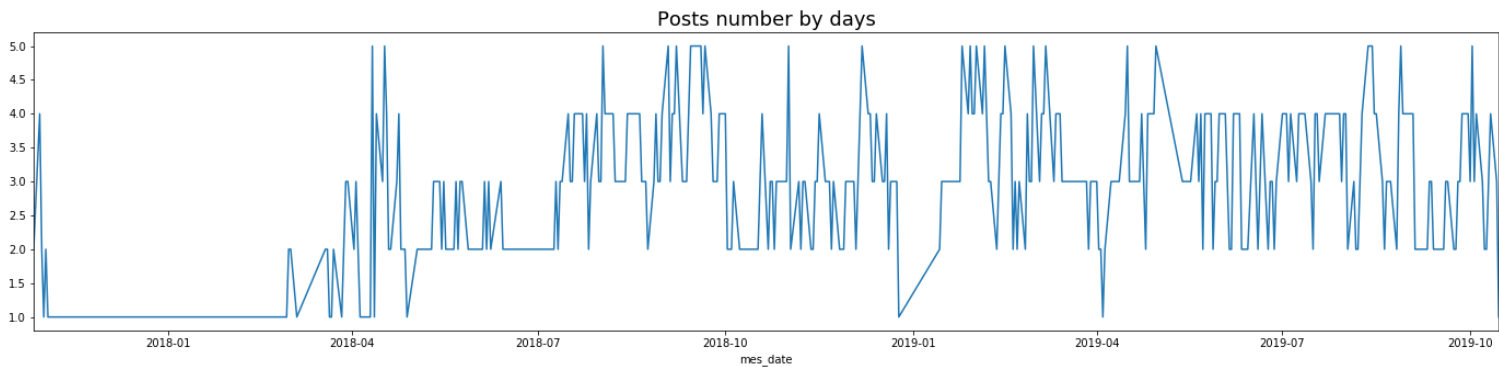


Динамика постов по месяцам

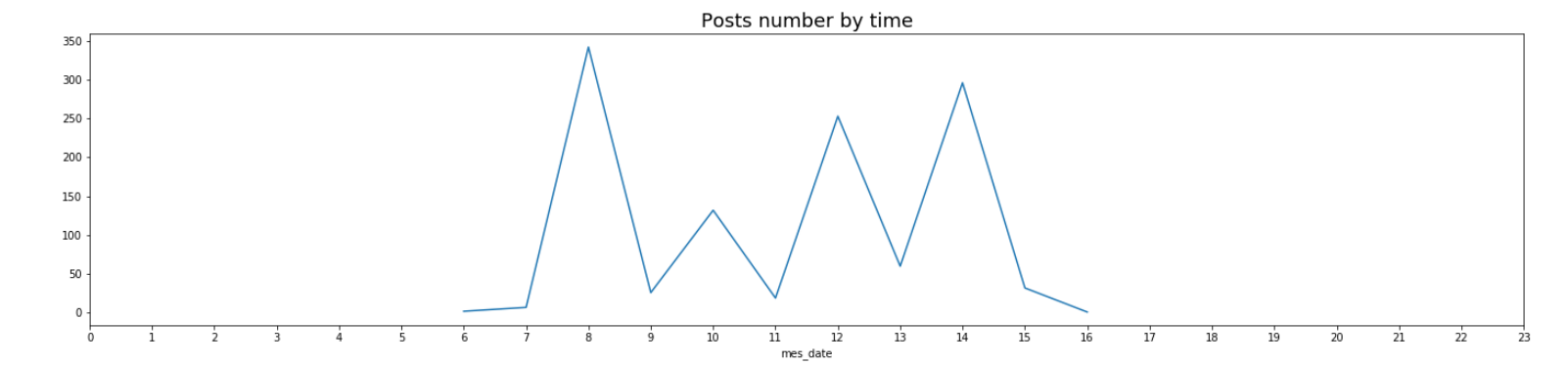


Динамика с накопление, поэтому видим достаточно быстрый подъем с 07/2018 и быстрое увелечение количесва ~ в 6 раз

Динамика постов по дням

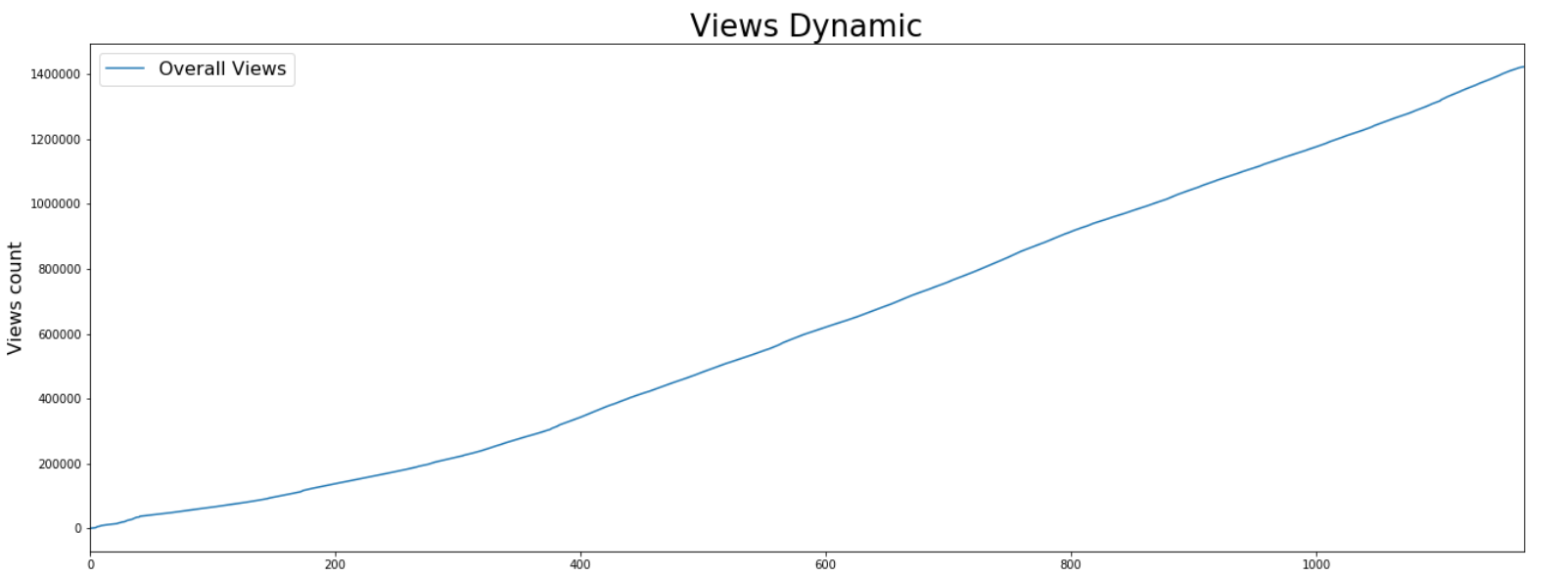


Динамика постов по часам



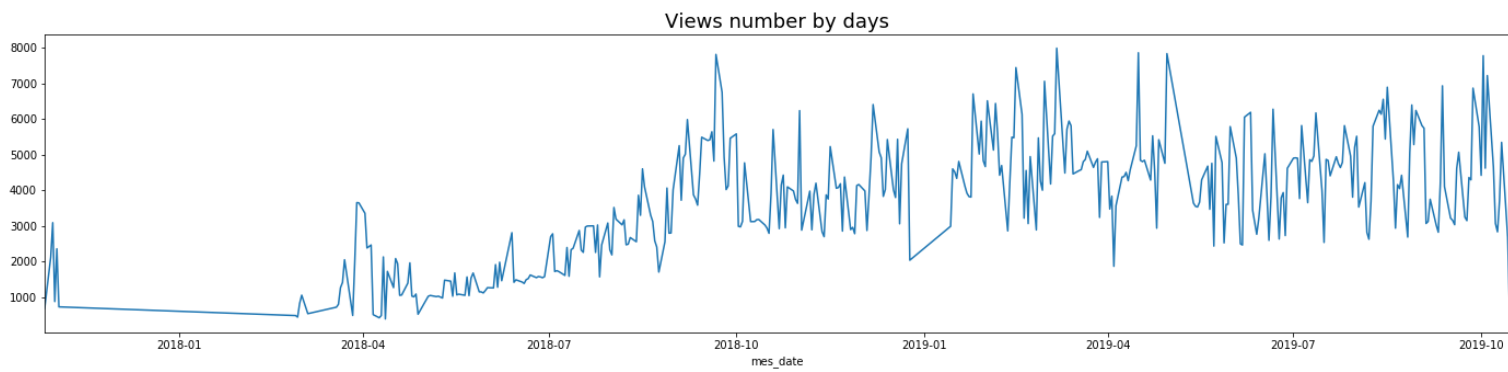
Пики размещения вакансий 8 утра, 12 и 14 часов.

Динамика просмотров



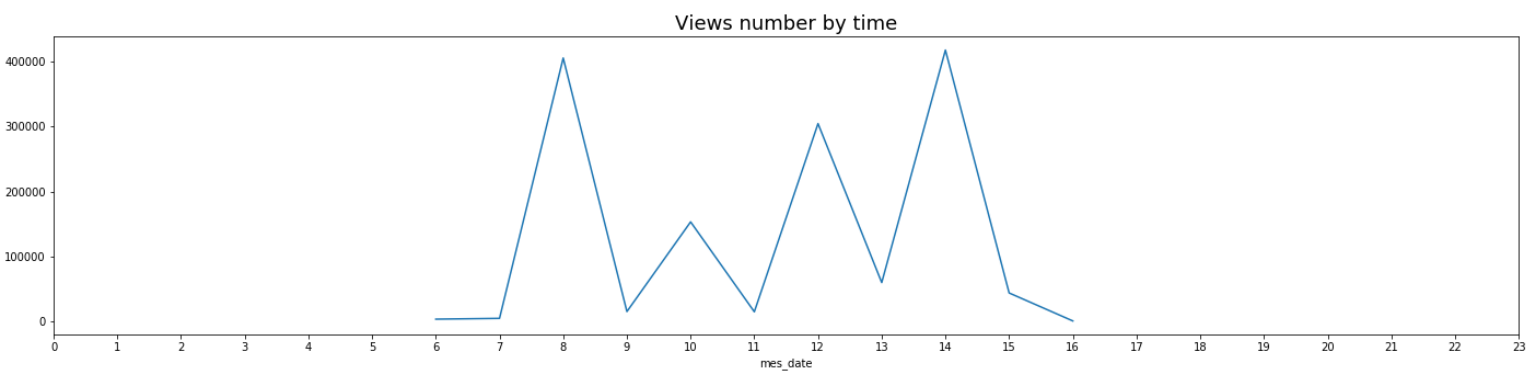
Судя по графику просмотры достаточно линейно, но точно оценить динамику сложно. Посмотрим по дням.

Просмотры по дням



Уже более отчетливо прослеживается тренд, хотя рост не слишком большой и явно замедлился после 07/2019

Просмотры по часам



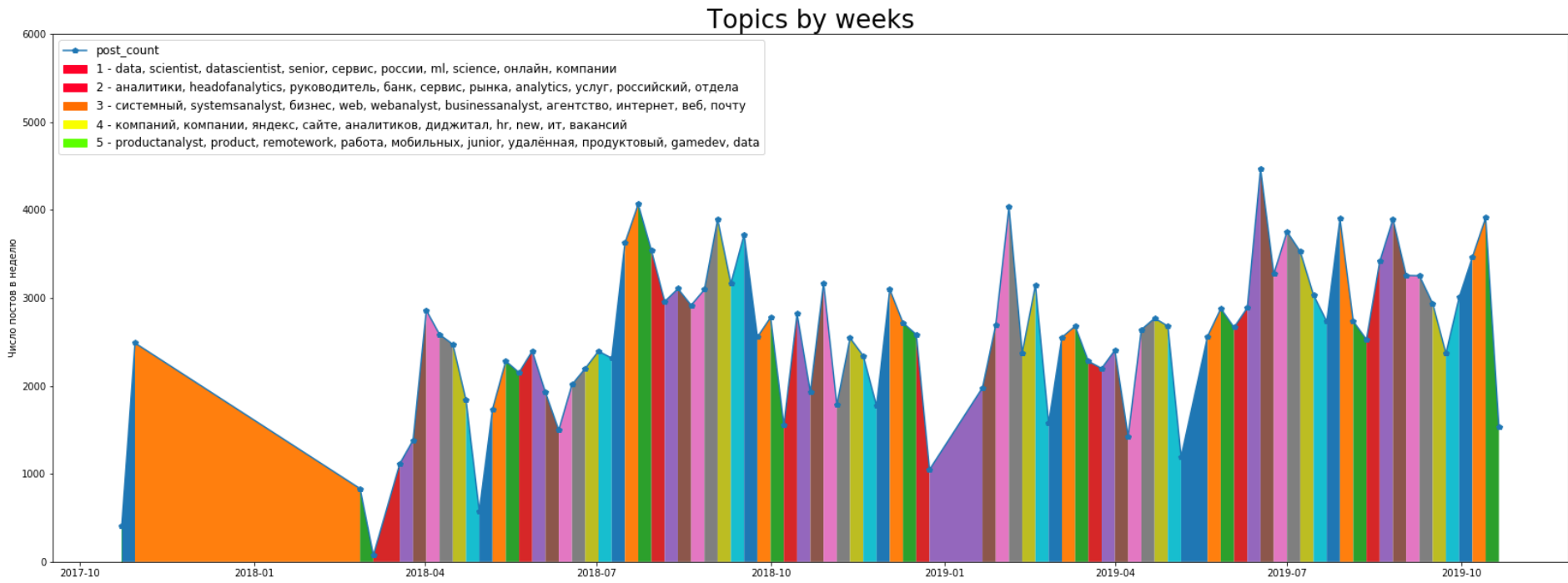
Пики просмотров совпадают с пиковым временем размещения сообщения. Это говорит о том, что подписчики сразу после публикации смотрят пост.

Тематическое моделирования

Построив Латентное размещение Дирихле (LDA) можно увидеть выделение 5 групп, которые я сгруппировал следующим образом: 'data scientist', 'руководители', 'system analyst', 'диджитал', 'product analyst'.

Тема	Наиболее часто встречающиеся слова (14 слов)
0 topic : data scientist	data, scientist, datascientist, senior, сервис, россии, ml, science, онлайн, компании, откликнуться, крупнейший, проект, relocation
1 topic : руководители	аналитики, headofanalytics, руководитель, банк, сервис, рынка, analytics, услуг, российский, отдела, данных, задачи, проект, head
2 topic : system analyst	системный, systemsanalyst, бизнес, web, webanalyst, businessanalyst, агентство, интернет, веб, почту, решений, digital, разработчик, области
3 topic : диджитал	компаний, компании, яндекс, сайте, аналитиков, диджитал, hr, new, ит, вакансий, mail, marketinganalyst, сервис, group
4 topic : product analyst	productanalyst, product, remotework, работа, мобильных, junior, удалённая, продуктовый, gamedev, data, разработчик, онлайн, engineer, игр

Распределение тем по неделям



Интересный момент, в самом начале было много вакансия системных аналитиков, а потом доли по тематикам стали стабилизироваться