

Візуальний аналіз даних популярності статті на сайті habr.com

Підготовчий етап

1. Імпортуйте необхідні бібліотеки.
2. Завантажимо надані у файли з вихідними даними. Ознайомтеся з ними.
3. Виведіть дані так як показано на зображенні.

	0	1	2
url	https://habrahabr.ru/post/18284/	https://habrahabr.ru/post/18285/	https://habrahabr.ru/post/18286/
domain	habrahabr.ru	habrahabr.ru	habrahabr.ru
post_id	18284	18285	18286
published	2008-01-01 18:19:00	2008-01-01 18:30:00	2008-01-01 18:34:00
author	@Tapac	@DezmASter	@DezmASter
flow	develop	design	design
polling	False	False	False
content_len	4305	7344	8431
title	Новогодний подарок блогерам — WordPress 2.3.2	Сумасшедшие яйца, или сервис для отслеживания ...	Сумасшедшие яйца, или сервис для отслеживания ...
comments	0	1	47
favs	0	1	72
views	236	353	1200
votes_plus	0	1	45
votes_minus	0	0	5
views_lognorm	-0.792687	-0.485214	0.44609
favs_lognorm	-1.34407	-0.831946	1.79402
comments_lognorm	-2.43687	-1.87319	0.574656

4. Видаліть данні, що завершуються на _lognorm
5. Застосуйте для виводу статистичної інформації функцію DataFrame.describe()

	count	mean	std	min	25%	50%	75%	max
post_id	134137.0	181307.054265	81766.350702	18284.0	115565.0	182666.0	257401.0	314088.0
content_len	134137.0	7516.929699	8724.771640	1.0	2179.0	4949.0	9858.0	352471.0
comments	134137.0	39.625994	62.398958	0.0	7.0	19.0	48.0	2266.0
favs	134137.0	71.999866	145.854135	0.0	8.0	27.0	78.0	8718.0
views	134137.0	16631.013084	31479.819691	3.0	2100.0	7600.0	18700.0	1730000.0
votes_plus	133566.0	35.536888	42.461073	0.0	11.0	22.0	45.0	1466.0
votes_minus	133566.0	8.050035	11.398282	0.0	2.0	5.0	10.0	445.0

6. Стовець published (час публікації) містить рядки. Щоб ми могли працювати з цими даними як з датою/ часом публікації, перетворіть їх до типу datetime:
7. Створіть кілька стовпців на основі даних про час публікації:

```
df['year'] = [d.year for d in df.published]
df['month'] = [d.month for d in df.published]
df['dayofweek'] = [d.isoweekday() for d in df.published]
df['hour'] = [d.hour for d in df.published]
```

Завдання:

У кожному пункті пропонується побудувати графік і з його допомогою відповісти на питання.

1. В якому місяці (і якого року) було найбільше публікацій?
 - березень 2016
 - березень 2015
 - квітень 2015
 - квітень 2016
2. Проаналізуйте публікації в місяці з попереднього питання. Виберіть один або кілька варіантів:
 - Один або кілька днів сильно виділяються із загальної картини
 - На Хабре завжди більше статей, ніж на гіктаймсе
 - По суботах на гіктаймс і на Хабрахабр публікують приблизно однакове число статей

Підказки: побудуйте графік залежності числа публікацій від дня; використовуйте параметр `hue`;

3. Коли найкраще опублікувати статтю?
 - Найбільше переглядів набирають статті, опубліковані о 12 годині дня
 - У опублікованих в 10 ранку постів найбільше коментарів
 - Найбільше переглядів набирають статті, опубліковані в 6 годині ранку
 - Максимальна кількість коментарів на гіктаймсе набрала стаття, опублікована о 9 годині вечора
 - На Хабре денні статті коментують частіше, ніж вечірні.
4. Кого з топ-20 авторів найчастіше мінусують?
 - @Mordatyj
 - @Mithgol
 - @alizar
 - @ ilya42

5. Порівняйте суботи і понеділки

Чи правда, що по суботах автори пишуть в основному вдень, а по понеділках - в основному ввечері?