

CITIBIKE

-

ANALIZA



Opracowali: Adrian Zareba, Zuzanna Piróg



Czym jest CitiBike?

- Citi Bike to największy system udostępniania rowerów publicznych, obsługujący nowojorskie dzielnice:
 - Bronx
 - Brooklyn
 - Manhattan
 - Queens
 - Jersey City
 - Hoboken
- Stanowi alternatywny środek transportu, umożliwiający szybkie poruszanie się po całym Nowym Jorku
- Rozpoczął działanie 27 maja 2013r.



Co oferuje CitiBike?



Annual Membership

\$15*/month

unlimited 45-minute rides on a classic bike.

[\[Join now\]](#)

*\$185 billed upfront annually

Single Ride

\$3.99/trip

one ride up to 30 minutes on a classic bike.

[Get a pass](#)

Day Pass

\$15/day

unlimited 30-minute rides in a 24-hour period on a classic bike.

[Get a pass](#)

Citi Bike oferuje dostęp do ponad 1,500 stacji rowerowych i ponad 25,000 rowerów rozmieszczonych po całym mieście.

Aby skorzystać z usług można wykupić:

- Rocznny abonament i zostać subskrybentem
- Pojedynczy przejazd lub jednodniowy dostęp i zostać klientem

Tyle możliwości...

Pojawia się wiele pytań związanych z nowojorską wypożyczalnią rowerów miejskich.

- Jakie są najpopularniejsze stacje?
- Kiedy rowerzyści są najbardziej aktywni?
- Jak pogoda wpływa na aktywność użytkowników?
- I wiele innych

Na szczęście, Citi Bike udostępnia ogólnodostępną bazę danych. Zawiera dane sięgające roku 2013 i jest na bieżąco aktualizowana



Baza danych

- Analizowana przez nas baza danych zawiera informacje z roku 2019 (z okresu 3 miesięcy wakacyjnych) i 2016
- W pliku uzyskanym ze strony Citi Bike znajdziemy dane o przejaźdżce takie jak:

- Czas trwania (sekundy)
- Godzina i data rozpoczęcia i zakończenia
- Nazwa stacji początkowej i końcowej
- Id danej stacji
- Szerokość i długość geograficzna danej stacji
- Rodzaj użytkownika (klient/subskrybent)
- Płeć użytkownika
- Rok urodzenia użytkownika

	tripduration	starttime	stoptime		
1	393	2019-08-01 00:00:01.4680	2019-08-01 00:06:35.3780		
2	627	2019-08-01 00:00:01.9290	2019-08-01 00:10:29.7840		
3	1132	2019-08-01 00:00:04.0480	2019-08-01 00:18:56.1650		
4	1780	2019-08-01 00:00:04.1630	2019-08-01 00:29:44.7940		
5	1517	2019-08-01 00:00:05.4580	2019-08-01 00:25:23.4550		
6	632	2019-08-01 00:00:11.8060	2019-08-01 00:10:43.9590		
	start.station.id	start.station.name			
1	531	Forsyth St & Broome St			
2	274	Lafayette Ave & Fort Greene Pl			
3	2000	Front St & Washington St			
4	479	9 Ave & W 45 St			
5	3312	1 Ave & E 94 St			
6	173	Broadway & W 49 St			
	start.station.latitude	start.station.longitude	end.station.id		
1	40.71894	-73.99266	408		
2	40.68692	-73.97668	3409		
3	40.70255	-73.98940	3388		
4	40.76019	-73.99126	473		
5	40.78172	-73.94594	3312		
6	40.76068	-73.98453	3707		
	end.station.name	end.station.latitude			
1	Market St & Cherry St	40.71076			
2	Bergen St & Smith St	40.68674			
3	President St & Henry St	40.68280			
4	Rivington St & Chrystie St	40.72110			
5	1 Ave & E 94 St	40.78172			
6	Lexington Ave & E 26 St	40.74146			
	end.station.longitude	bikeid	usertype	birth.year	gender
1	-73.99400	35305	Subscriber	1996	2
2	-73.99063	38822	Subscriber	1998	2
3	-73.99990	18373	Subscriber	1988	1
4	-73.99193	25002	Subscriber	1988	1
5	-73.94594	31198	Subscriber	1965	2
6	-73.98329	21628	Subscriber	1998	1

przykładowe dane

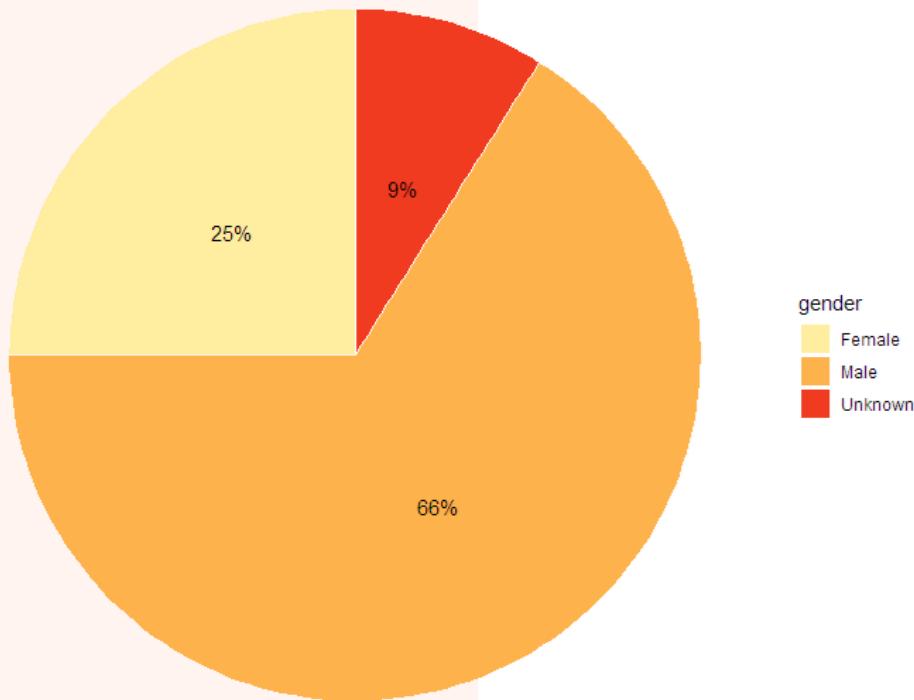




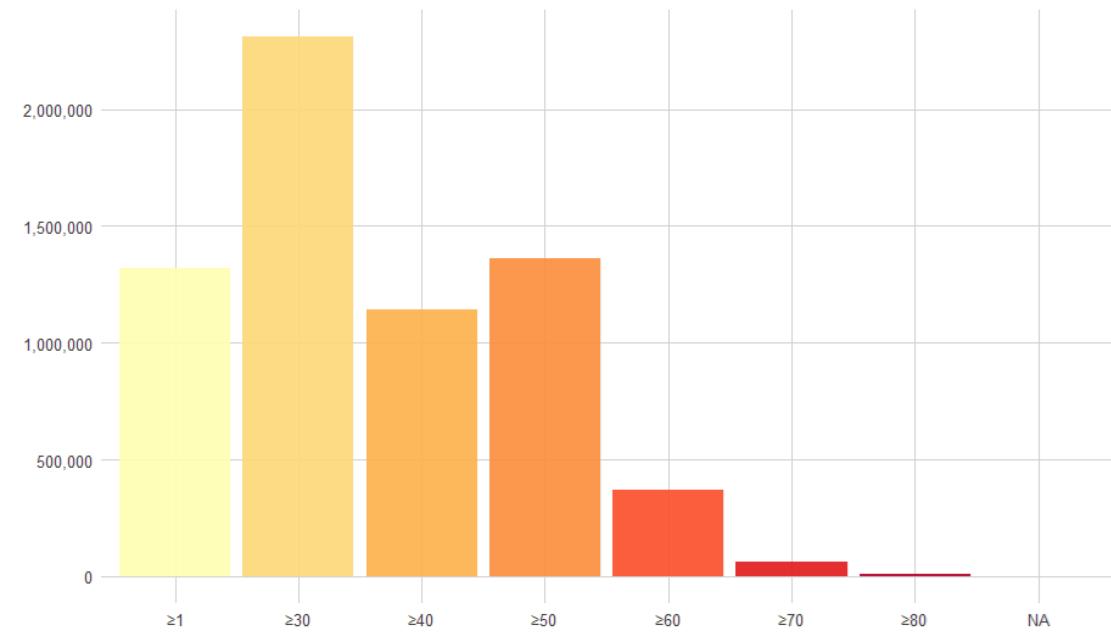
**Kto jest przeciętnym
użytkownikiem Citi Bike?**

Przeciętna płeć i wiek

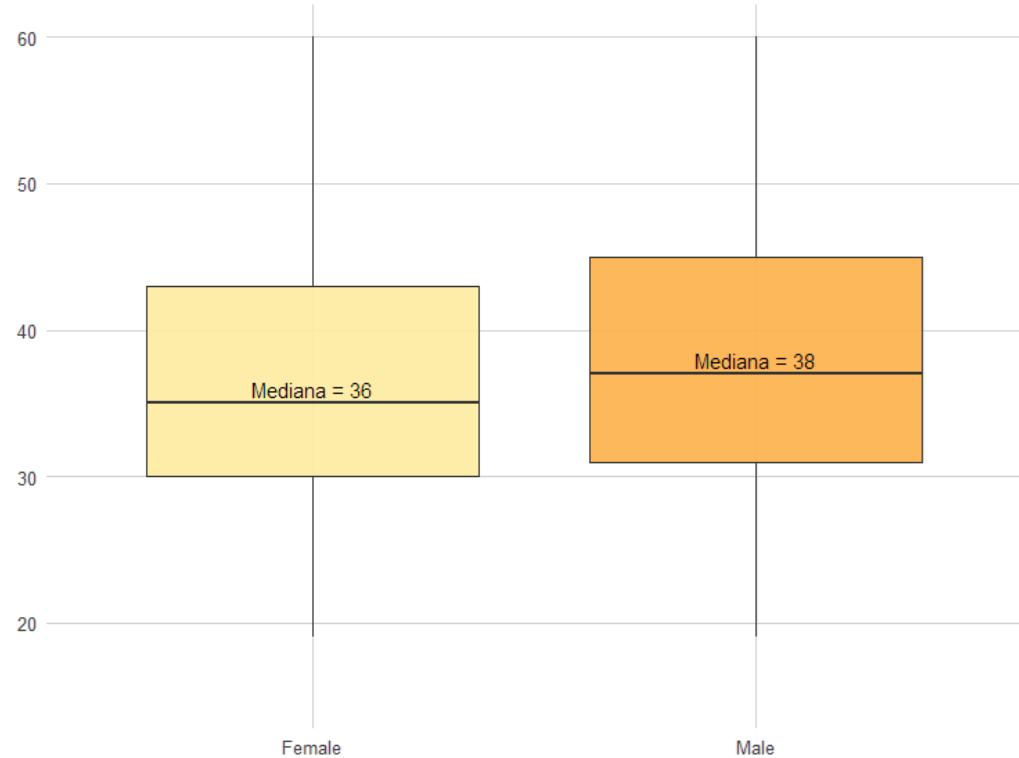
Rozłożenie płci



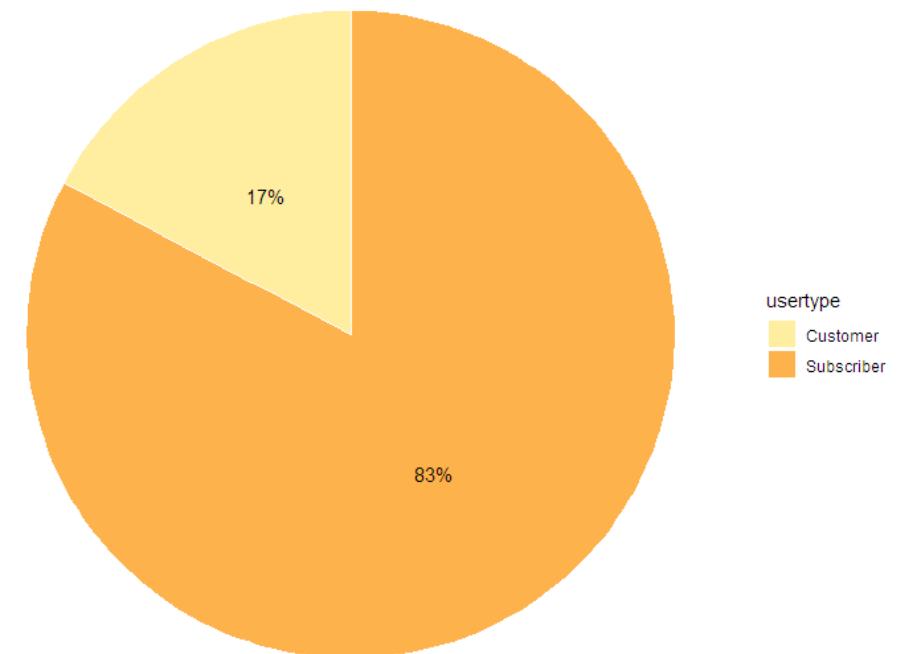
Rozkład grup wiekowych



Rozkład wieku w zależności od płci



Subskrybent a Klient jednorazowy



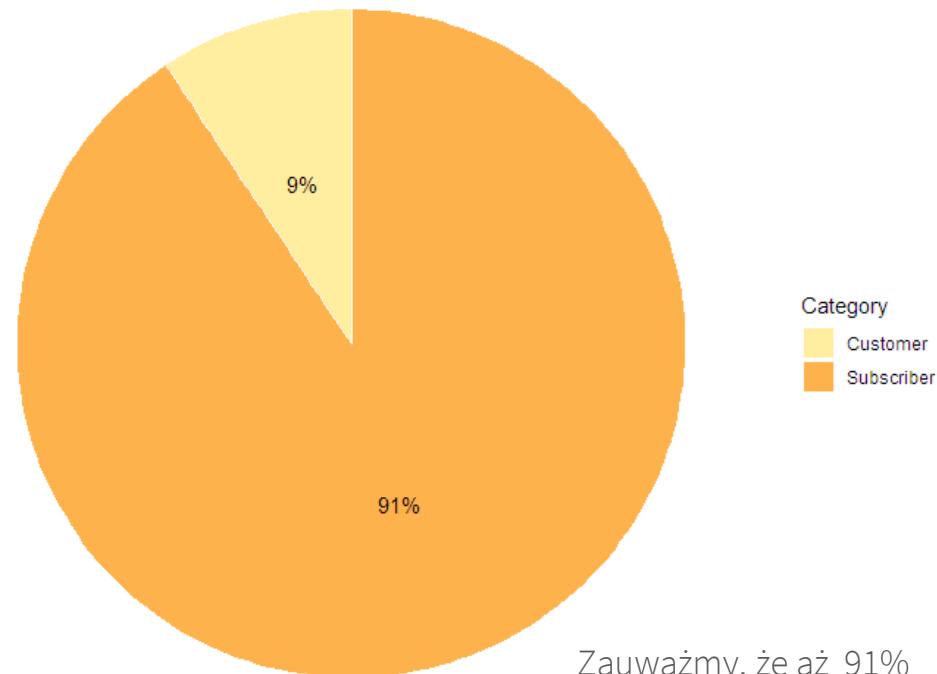
Jakie możemy wyciągnąć wnioski?

Przeciętnym użytkownikiem Citi Bike jest mężczyzna w wieku około 38 lat, który ma wykupioną roczną subskrypcję. Możemy więc założyć, że jest mieszkańców Nowego Jorku, który codziennie korzysta z usług wypożyczalni



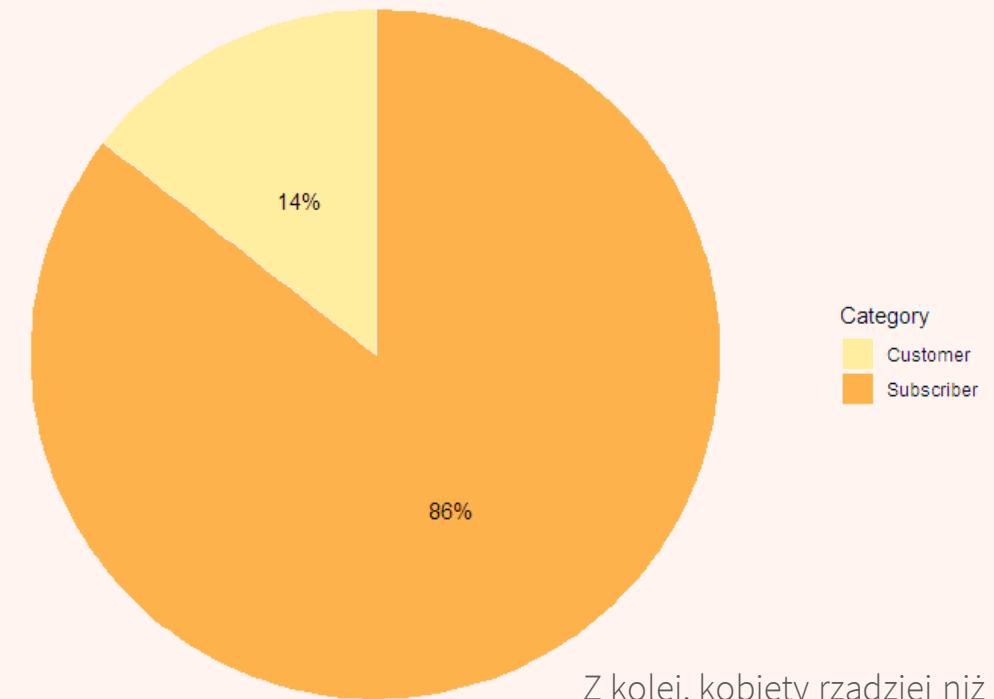
Subskrybent a płeć

Rozkład subskrybentów dla mężczyzn

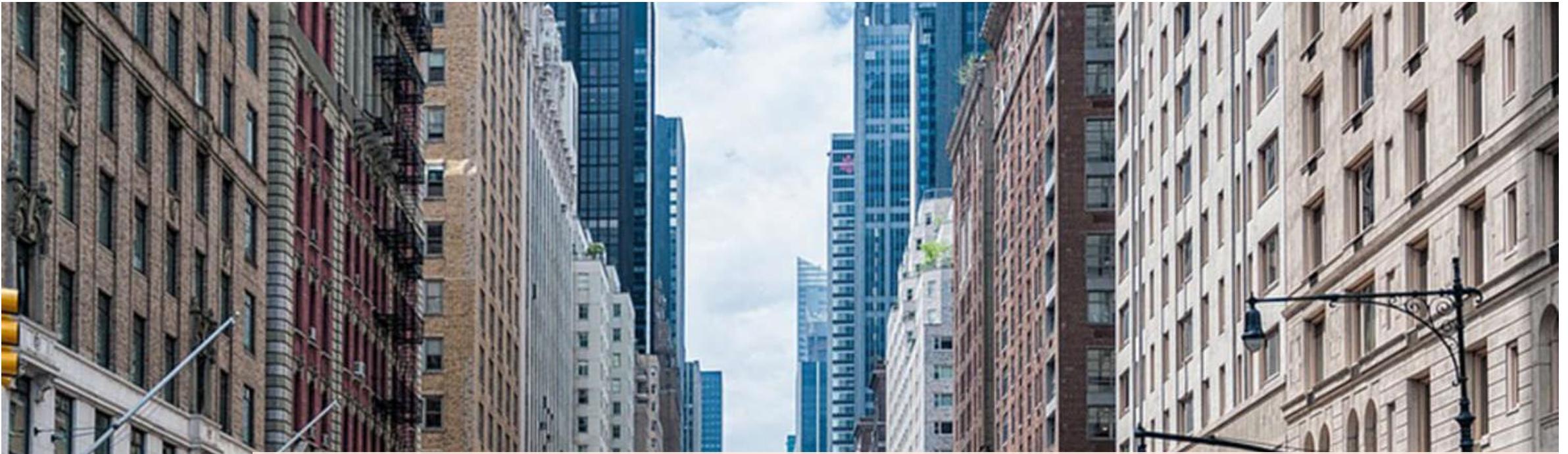


Zauważmy, że aż 91% procent mężczyzn jest subskrybentami, zaledwie 9% to jednorazowi klienci

Rozkład subskrybentów dla kobiet



Z kolei, kobiety rzadziej niż mężczyźni zostają subskrybentkami i częściej skłonne są do jednorazowych przejazdów



Jakie są najpopularniejsze stacje?



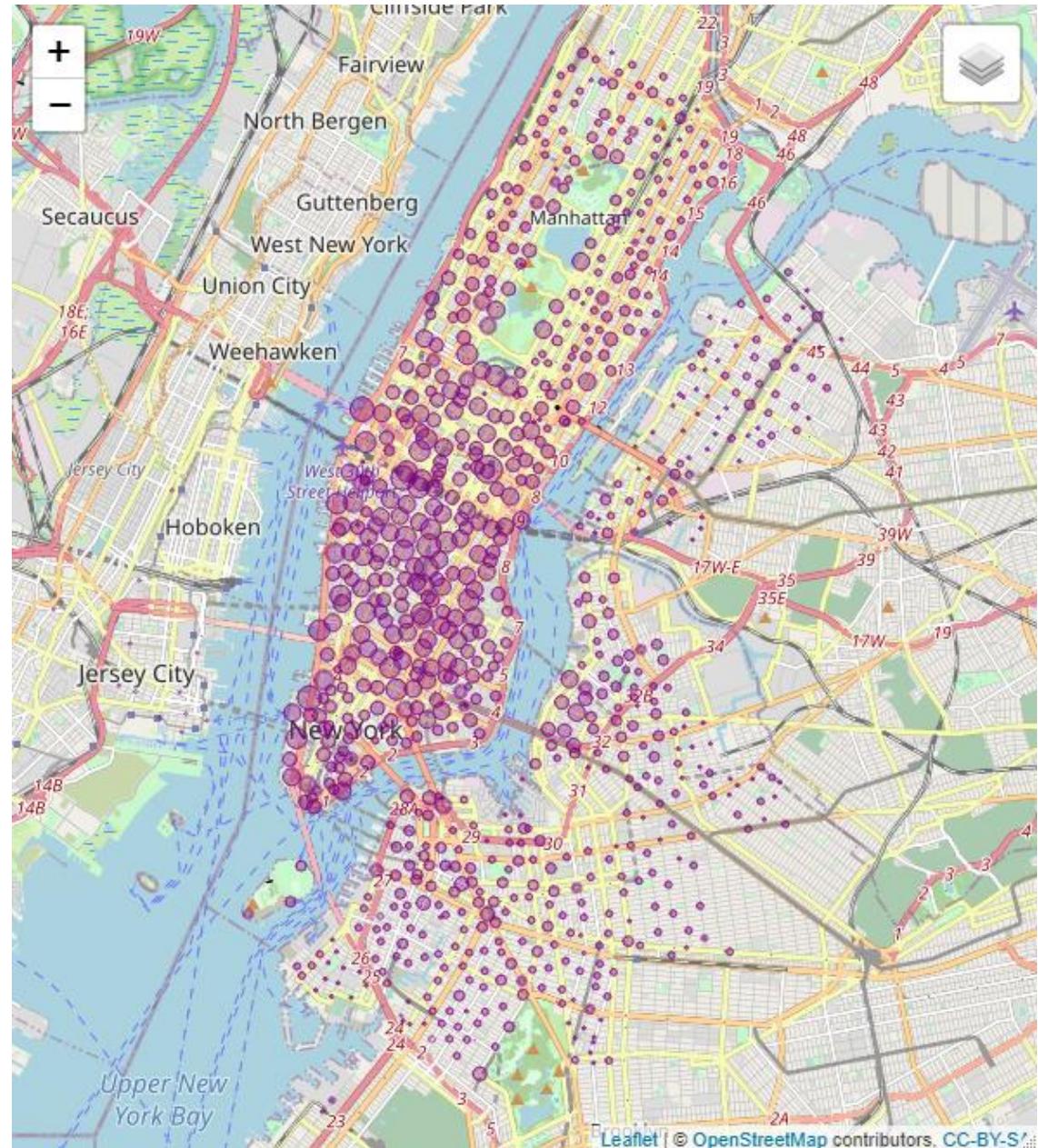
Stacje

Citi Bike oferuje możliwość korzystania z aż 1717 stacji.

Rozsiane są po całym mieście z czego najczęściej odwiedzane znajdują się w okolicy centrum.



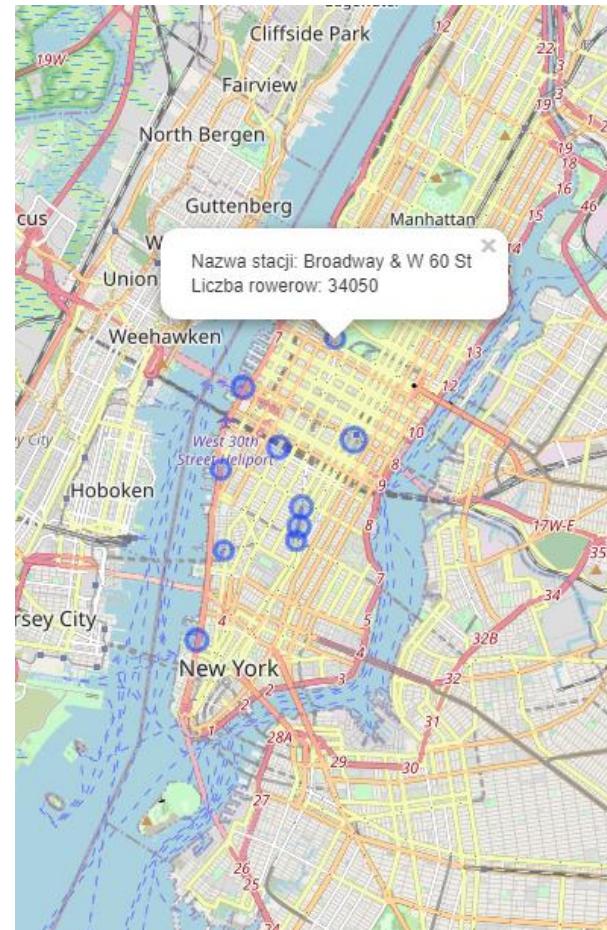
[mapaInteraktywna.html](#)



Stacje początkowe

- Najpopularniejsza stacja początkowa to Pershing Square North z której wyjechało aż 47,435 rowerzystów w przeciągu zaledwie 3 miesięcy

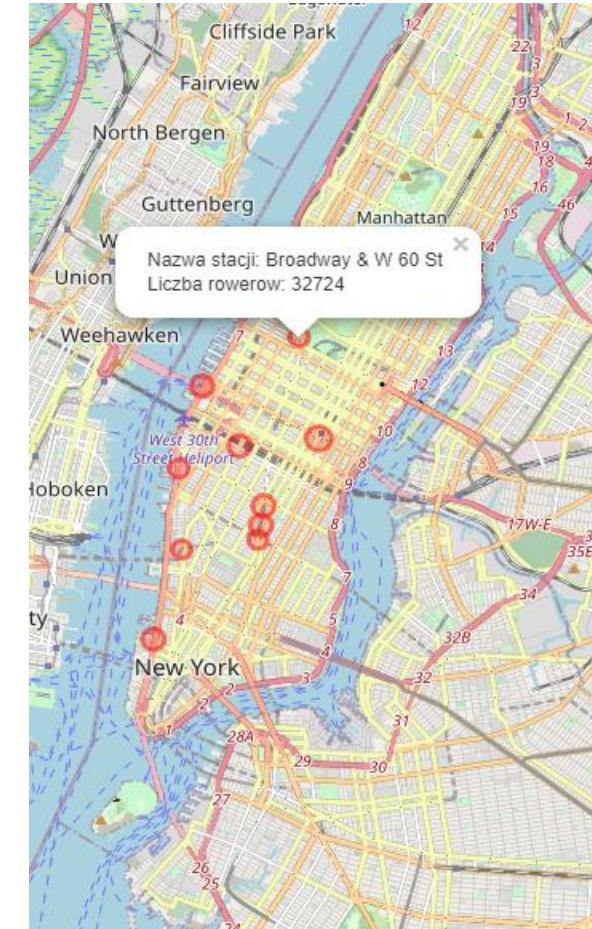
start.station.name	liczba
1 Pershing Square North	47435
2 West St & Chambers St	39603
3 12 Ave & W 40 St	39242
4 E 17 St & Broadway	37289
5 8 Ave & W 31 St	37137
6 Broadway & E 22 St	35445
7 Christopher St & Greenwich St	34864
8 Broadway & W 60 St	34050
9 Broadway & E 14 St	33521
10 W 20 St & 11 Ave	31824



Stacje końcowe

- Ponownie najpopularniejsza stacja, tym razem końcowa, to Pershing Square North w której skończyło przejażdżkę 47,049 rowerzystów.

	end.station.name	liczba
1	Pershing Square North	47049
2	West St & Chambers St	41955
3	12 Ave & W 40 St	40354
4	E 17 St & Broadway	38405
5	Broadway & E 22 St	37496
6	8 Ave & W 31 St	37352
7	Christopher St & Greenwich St	34940
8	Broadway & E 14 St	33495
9	W 20 St & 11 Ave	32763
10	Broadway & W 60 St	32724



- Zauważmy, że najpopularniejsze stacje początkowe i końcowe pokrywają się, co oznacza to, że to właśnie w okolicach tych stacji możemy najczęściej spotkać rowerzystów Citi Bike

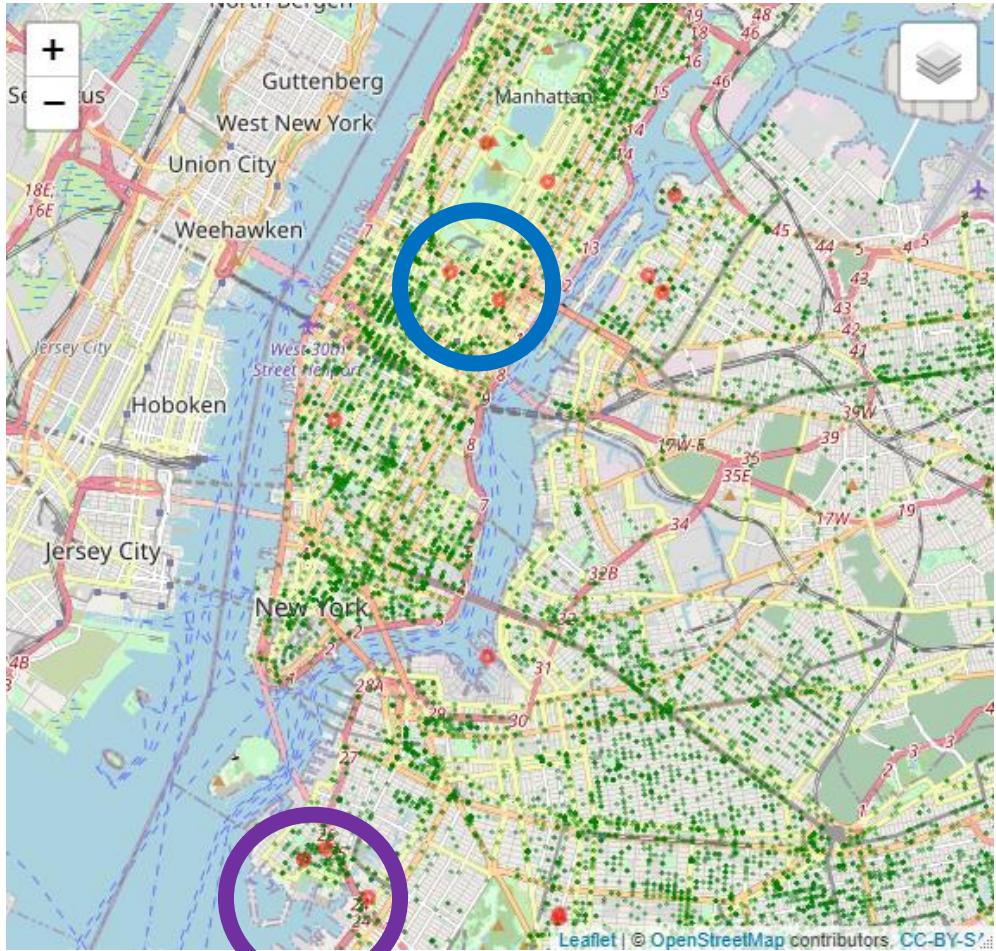
A co z najmniej popularnymi stacjami?

Aby odpowiedzieć na to pytanie skorzystamy z bazy danych przestępstw które miały miejsce w Nowym Jorku w 2019r. w czasie wakacji. Dowiemy się czy duża ilość przestępstw wpływa na zmniejszoną aktywność na stacji

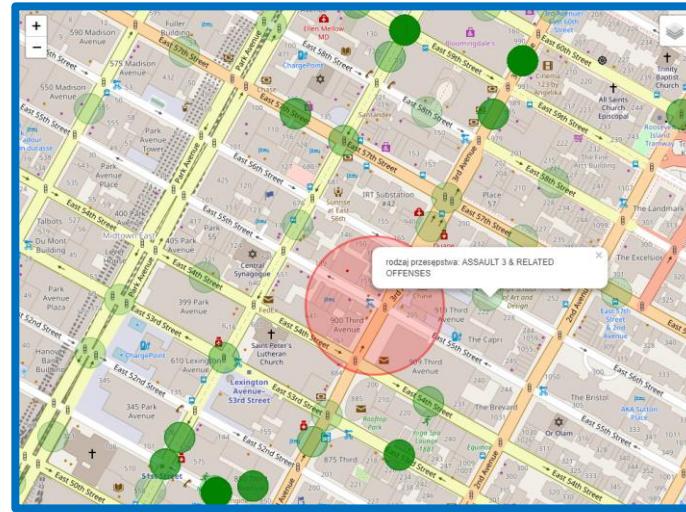


Miejsca w których dokonano aresztowań

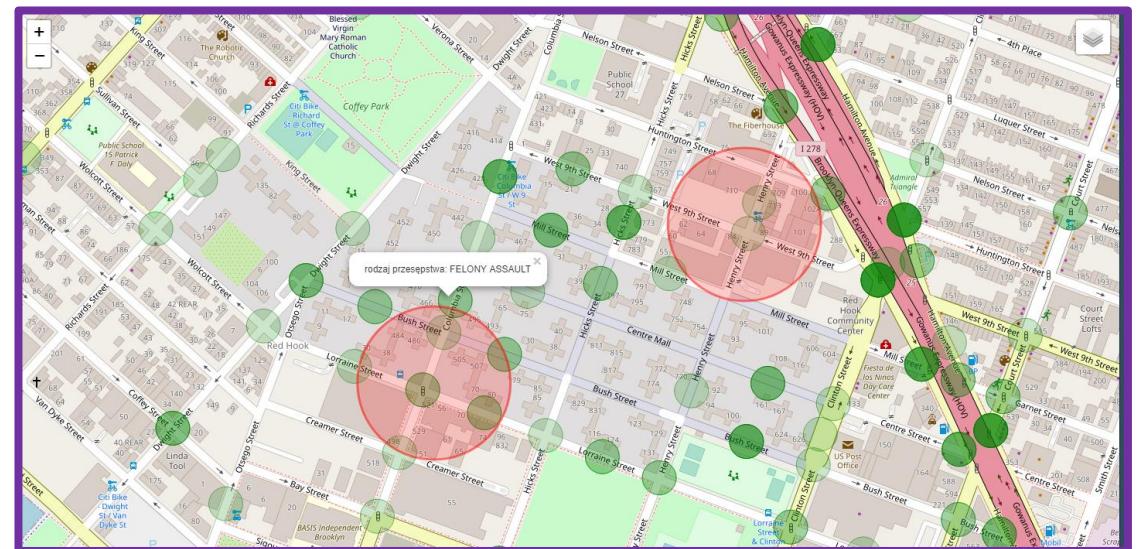
Stacje początkowe



Najrzadziej uczęszczane stacje początkowe

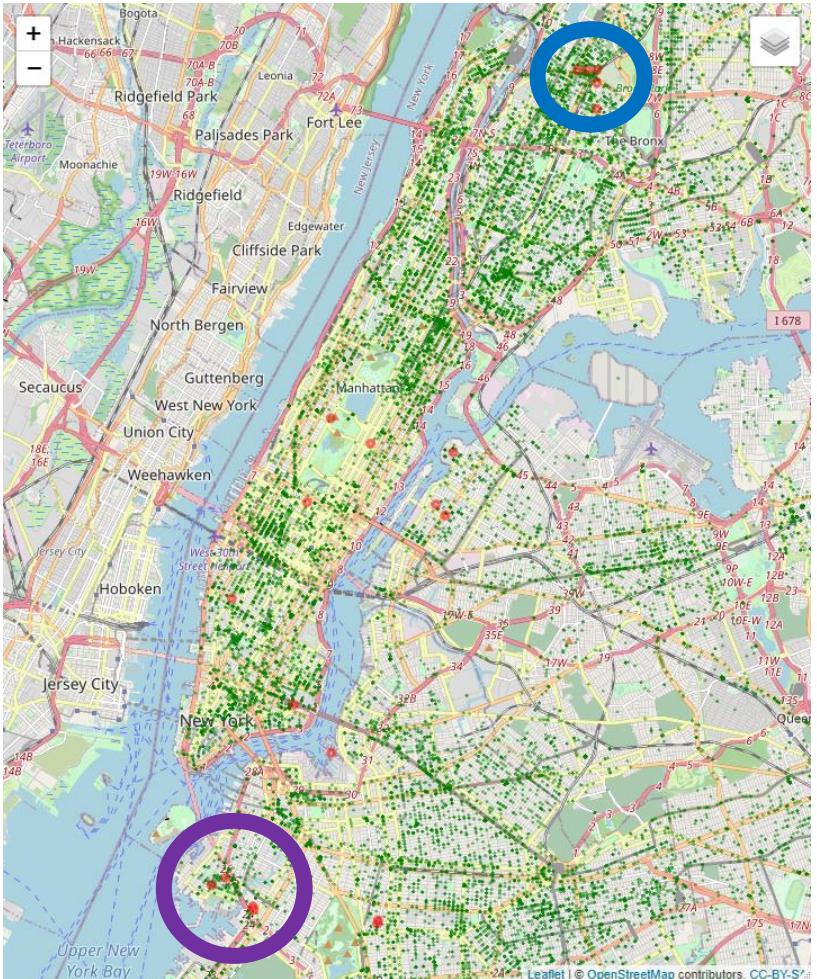


Powód Aresztowania: Napaść

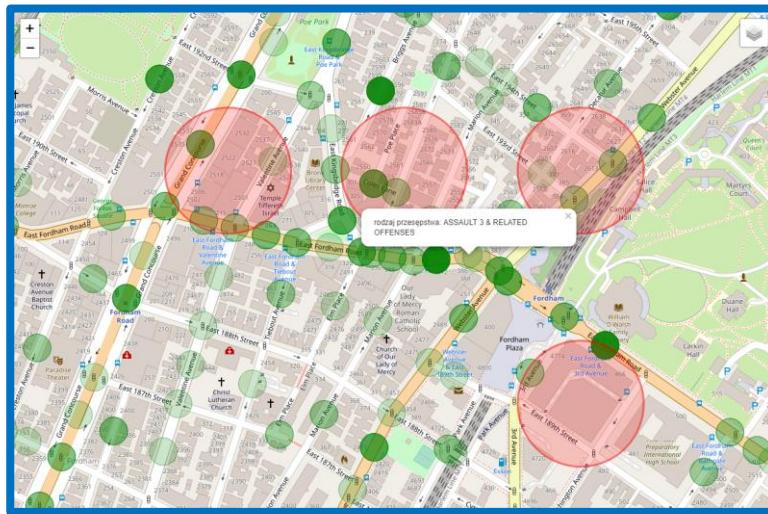


Powód Aresztowania: Napaść

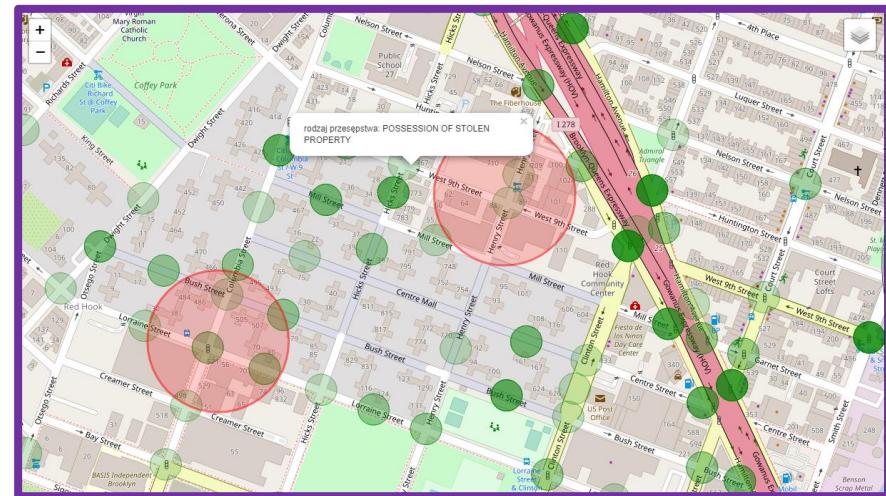
Stacje końcowe



Najrzadziej uczęszczane stacje końcowe



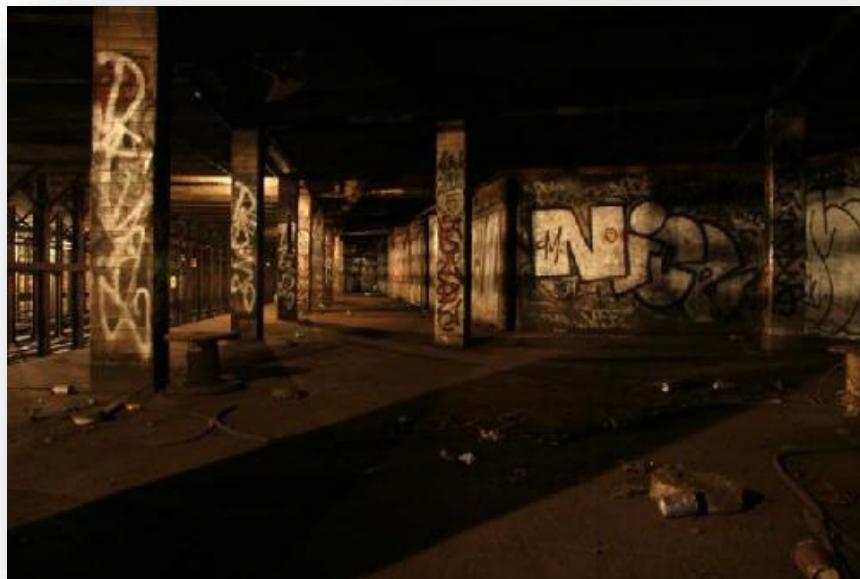
Powód Aresztowania: Napaść



Powód Aresztowania: Kradzież

Wnioski

Na podstawie dołączonych map, możemy zauważyc, że znaczna część stacji z najmniejszą aktywnością znajduje się w miejscach, w których jest zwiększoną liczba aresztowań. Jest to więc uzasadnione, że właśnie w tych rejonach najrzadziej spotkamy rowerzystę Citi Bike, który będzie kończył tam swoją podróż.

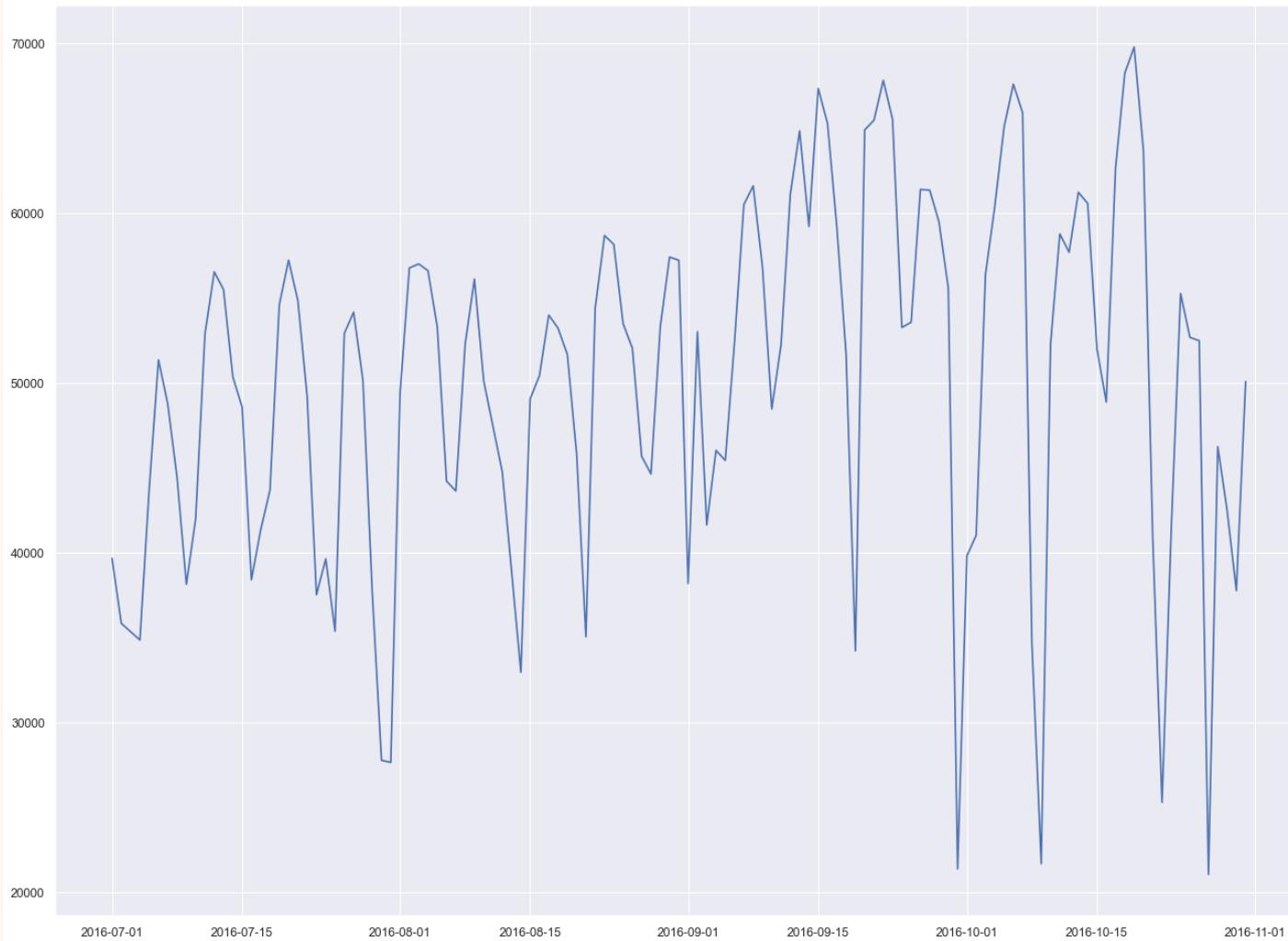


Zobaczmy teraz jak wygląda sprawa z osią czasu

Mając dane rozłożone w czasie można wyciągnąć wnioski na temat ich tendencji do alteracji. Jeżeli dane wykresowe, gdzie osią x jest czas, są stałe (lub zbliżają się do funkcji stałej) nie za dużo z nich wynika. W naszym przypadku jednak sytuacja jest odwrotna.

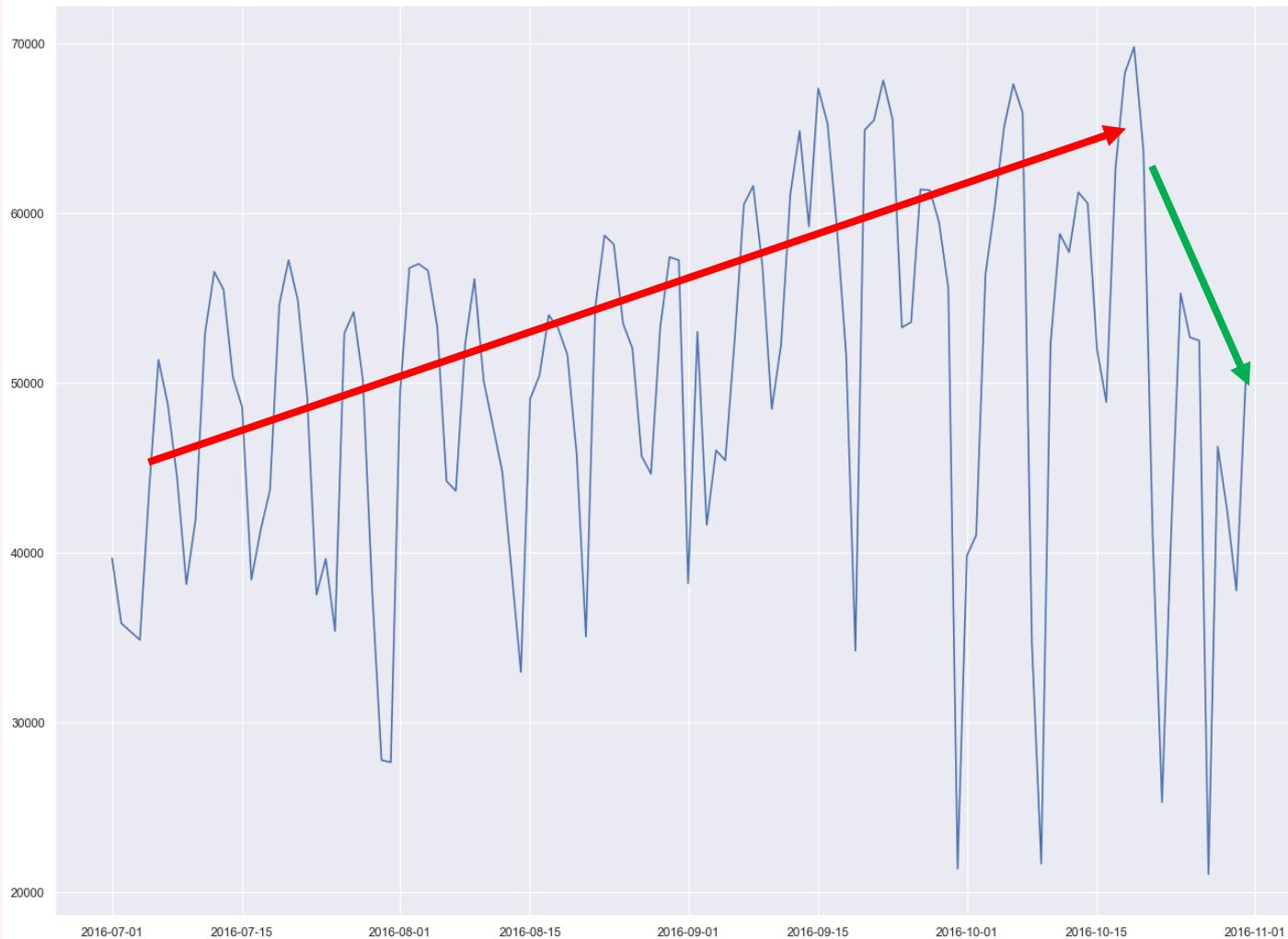


Zacznijmy od danych na przestrzeni dni z okresu 2016.07 - 2016.10 w NYC



Na osi y jest podana ilość użytkowników korzystających z usług citybike na poszczególny dzień z osi x.

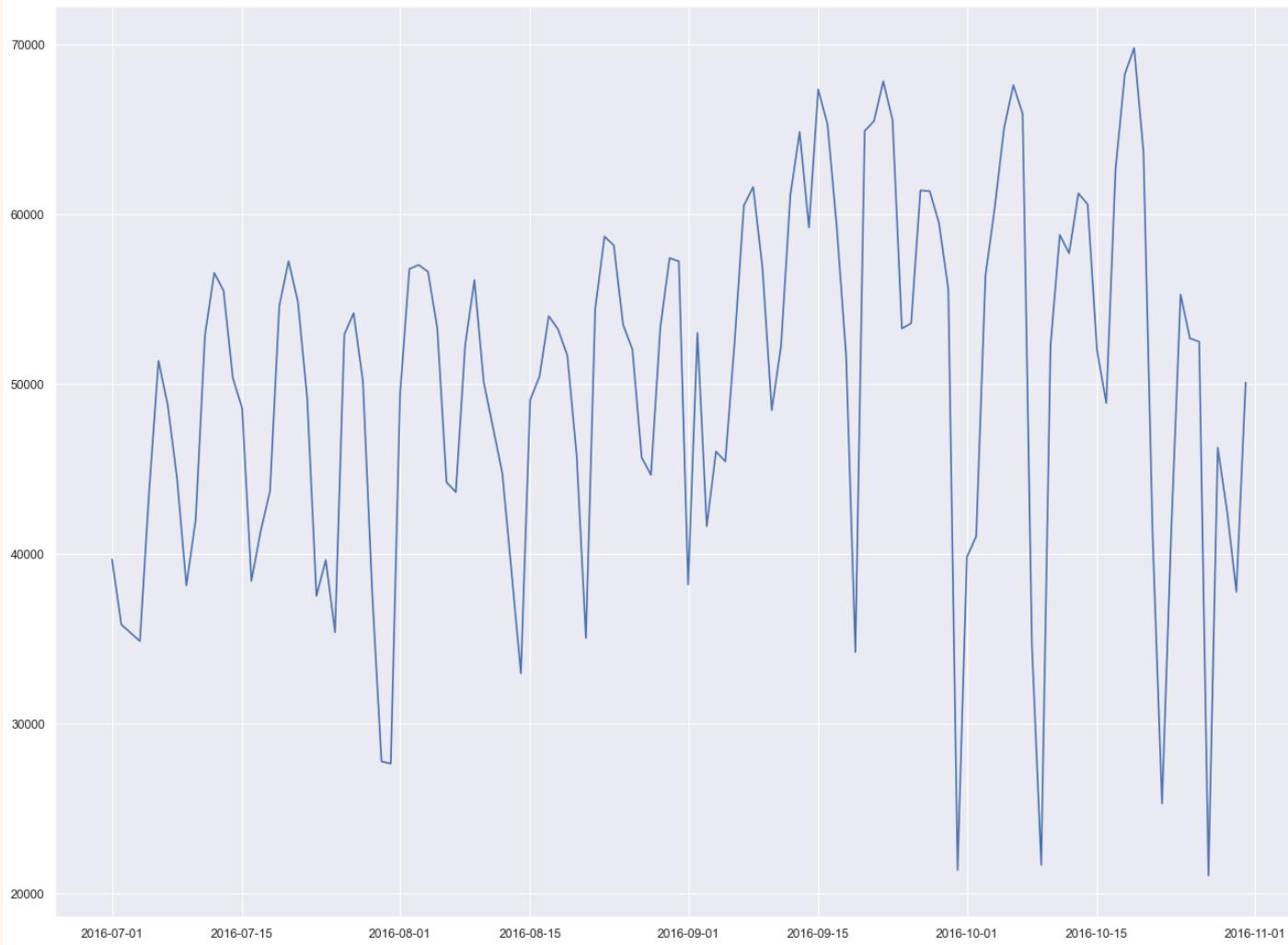
Zacznijmy od danych na przestrzeni dni z okresu 2016.07 - 2016.10 w NYC



Jak widać z wykresu mamy do czynienia z rosnącą zależnością opisaną prostą koloru czerwonego. Jest ona oczywista, gdyż wynika ze wzrostu liczby oferowanych rowerów przez citybike.

Nagły spadek oznaczony kolorem zielonym jest związany z obniżoną temperaturą, o której zależności przedstawiamy w następnych slajdach.

Tendencja dnia tygodnia



Odchyły na wykresie wyglądają na okresowo zmienne o pewnym okresie, co na przestrzeni dni miesiąca może symbolizować tendencję dni tygodnia.

Kiedy najczęściej?

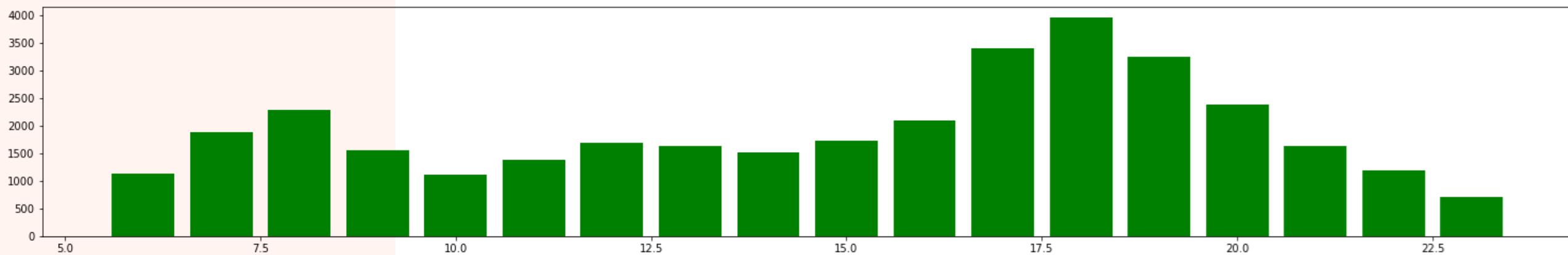
Z poprzedniego wykresu z NYC można odczytać, że ludzie najczęściej korzystają z usług citybike w środku tygodnia.

Przyczyną może być transport do pracy użytkowników. Jak to sprawdzić?



Wykres zależności ilości użytkowników od pory dnia

Jak widać z zamieszczonego wykresu godziny 7:30 oraz 17:30 cieszą się największym zainteresowaniem, wtedy właśnie ludzie transportują się między domem a miejscem pracy. (Jest to wykres danych z Jersey City)

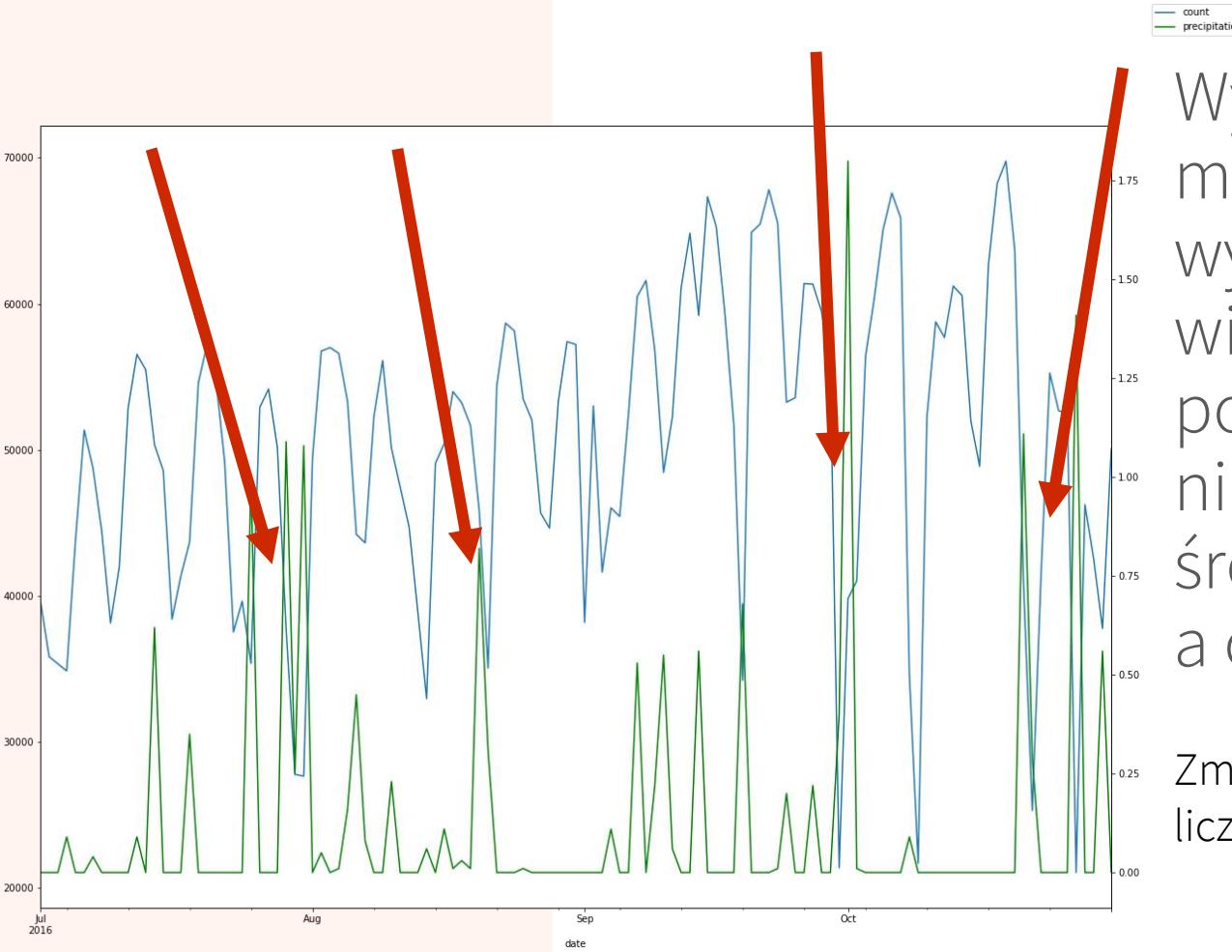


Odchyły od normy

Z poprzednich slajdów wywnioskowaliśmy, że citybike jest głównie wykorzystywane w transporcie do pracy. Co jednak znaczą i z czego wynikają widoczne na wykresie odchyły?



Wpływ pogody



Wystarczy nałożyć panującą w danym miejscu w danym dniu pogodę aby wyciągnąć proste wnioski. Odchyły widoczne na wykresie związane są z pogodą. Podczas burzy ludzie niechętnie korzystają z roweru jako środka transportu. Sprawa oczywista... a oto dowód:

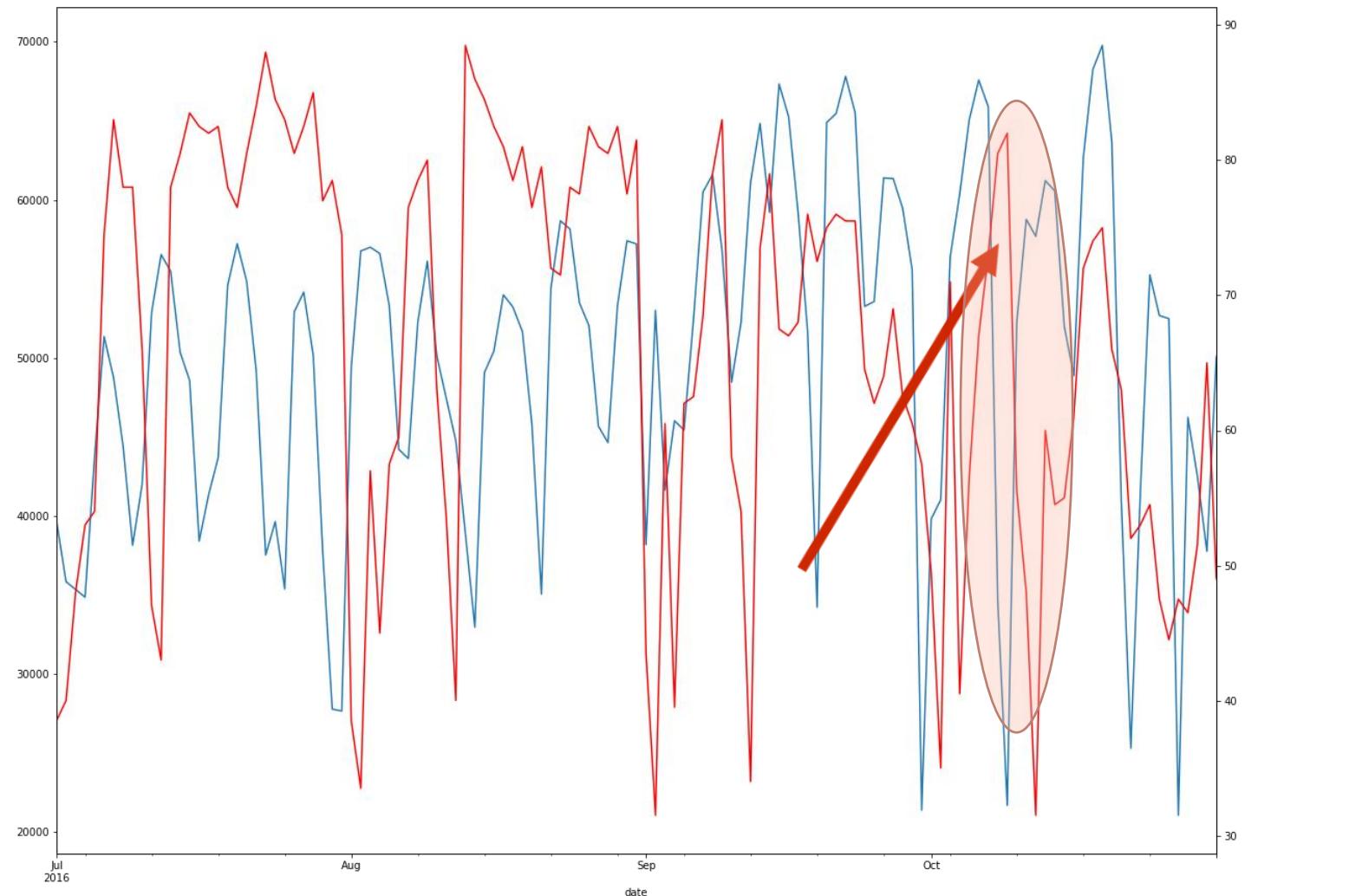
Zmienna koloru zielonego to ilość dziennych opadów liczona w centymetrach po prawej stronie wykresu

Wpływ pogody

Kiedy jest zbyt gorąco lub zbyt zimno zainteresowanie citibike spada.

Wykres koloru czerwonego to średnia temperatura panująca danego dnia liczona w stopniach F.

Temperatura również ma znaczenie



Typy użytkowników (Dane JC)

Ludzie mają zapisane w genach łączenie się w pewne odmienne grupy, z czego każda przejawia inne zachowanie na różnych płaszczyznach kulturowych.

Przykładem takich dywersyfikacji w społeczności są subskrybenci citibike oraz niezarejestrowani użytkownicy.

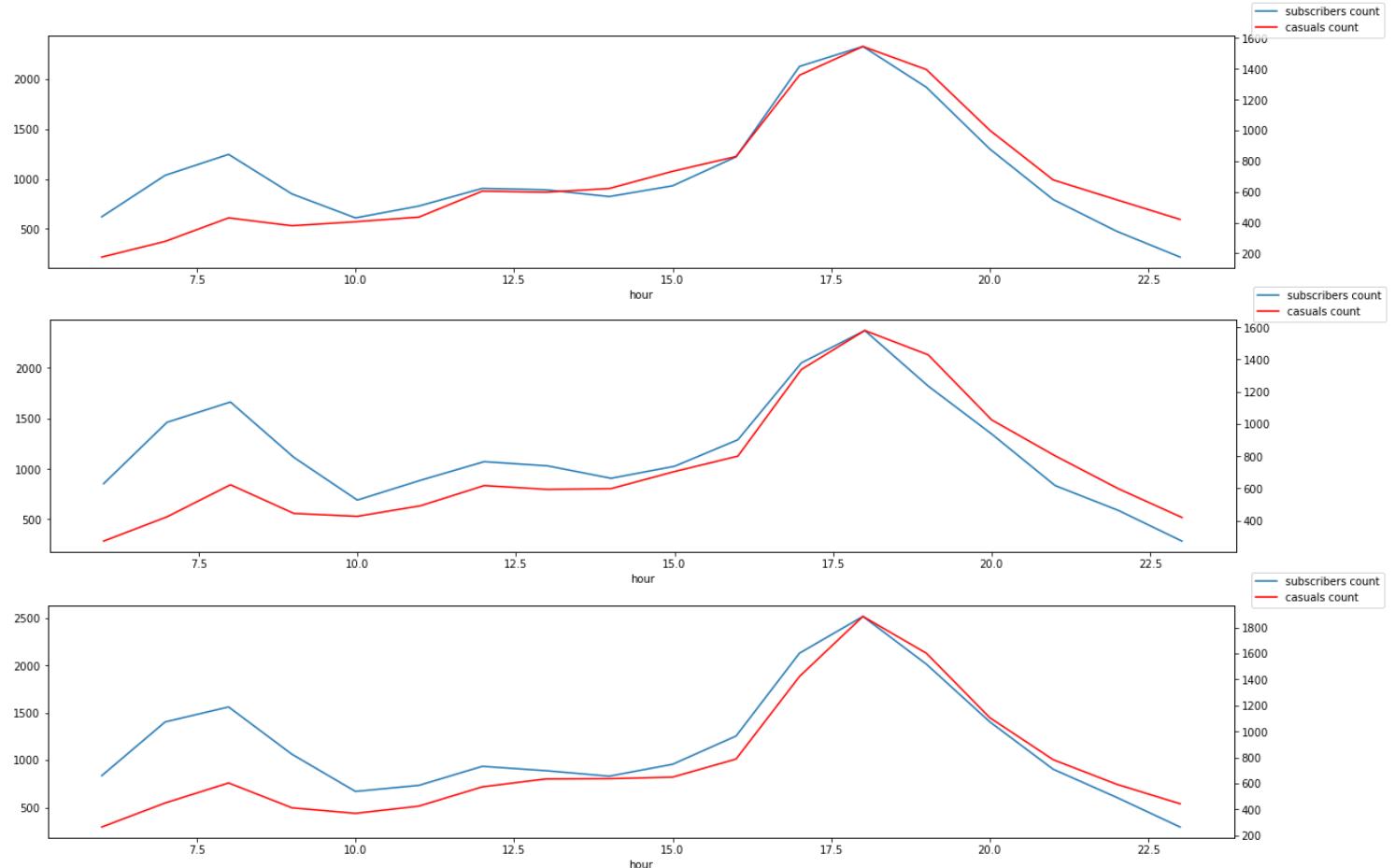


Typy użytkowników (Dane JC)

Zobaczmy na wykresy zależności subskrybentów (subscriber) od zwykłych użytkowników (casual).

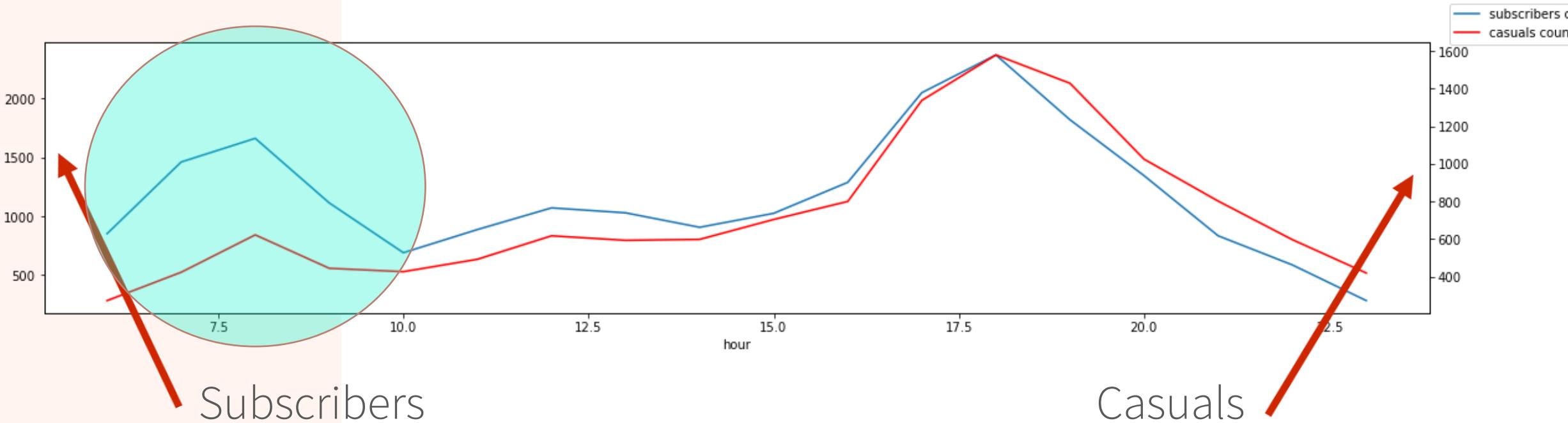
Obok zamieszczone wykresy pokazują zależność popularności użytkowania względem pory dnia (pon - śr).

Na czerwono (casual), na niebiesko (subscriber)



Typy użytkowników. (Dane JC)

Jak widać z wykresu (zwłaszcza wtorkowego) ilość subskrybentów jest większa od ilości „casuals”, a zwłaszcza w godzinach porannych (wtedy kiedy użytkownik przemieszcza się do miejsca pracy)



Typy użytkowników (Dane JC)

Wnioski są proste, subskrybenci znaczco przeważają ilością „casuals” oraz stosunkowo wracając do poprzedniego wykresu subskrybenci dużo częściej od „casuals” korzystają z usług citibike w godzinach porannych (w dojeździe do pracy).



Reasumując

Z danych pobranych z NYC oraz JC system rowerów publicznych oferowany przez citibike jest najczęściej wykorzystywany przez mężczyzn w wieku około 38 lat, którzy są subskrybentami i przemieszczają się docierając do pracy/firmy. Najbardziej znaczącym do tego czynnikiem zewnętrznym, który ma diametralny wpływ na zmianę w danych jest pogoda, a zwłaszcza ilość opadów danego dnia. Największa aktywność występuje na stacjach w centrum miasta, zaś najmniejsza w miejscach o zwiększonej aktywności przestępcej.





DZIĘKUJEMY ZA UWAGĘ!

Ramki danych:

<https://www.citibikenyc.com/system-data>

<http://w2.weather.gov/climate/xmacis.php?wfo=okx>

https://www.meteoblue.com/en/weather/archive/export/new-york_united-states_5128581

<https://www.kaggle.com/datasets/ajkarella/nyc-crime-stats>



mapalinteraktywna.html