

Przetwarzanie Danych Ustrukturyzowanych - zadanie domowe 3

MACIEJ BORKOWSKI, TYMOTEUSZ KWIECIŃSKI

Analiza *wydajności* rowerzystów

PIERWSZA ANALIZA

Porównanie wydajności rowerzystów

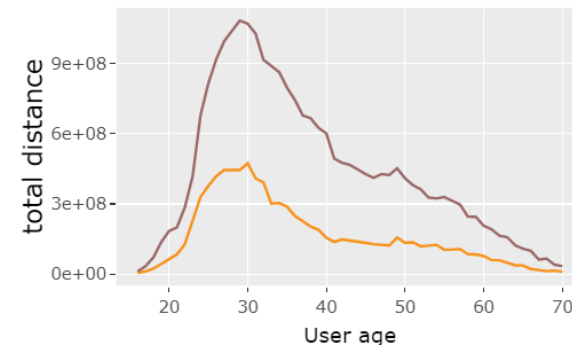
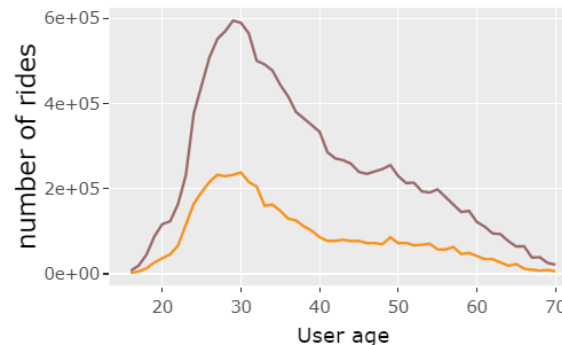
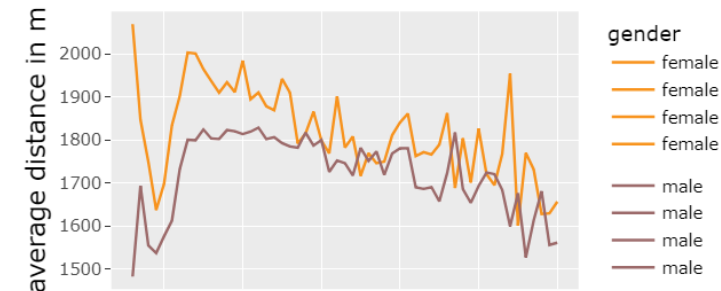
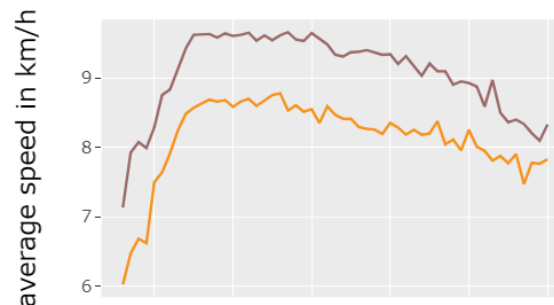
Na wykresach po prawej widnieje porównanie rowerzystów pod względem wieku oraz płci.

Jak się okazuje mężczyźni jeżdżą średnio nieznacznie szybciej od kobiet

Kobiety jeżdżą średnio dłużej niż mężczyźni.

Krzywa opisująca liczbę wypożyczeń oraz odległości pokonane przez kobiety oraz mężczyzn przypomina krzywą dzwonową

Average speed, distance and number of rides by age and gender



Jak pora roku i płeć użytkowników wpływają na czas trwania podróży?

DRUGA ANALIZA

Przygotowanie do analizy

- ❖ Do analizy wykorzystane zostały dane pochodzące z roku 2019 dla Nowego Jorku.
- ❖ Dla każdej obserwacji został przypisany rodzaj podróży:
 - ❖ Do 8 minut – bardzo krótka
 - ❖ 8-15 minut - krótka
 - ❖ 15-30 minut - średnia
 - ❖ 30-60 minut - długa
 - ❖ Ponad 60 minut – bardzo długa
- ❖ Dla każdej obserwacji została ponadto przypisana pora roku:
 - ❖ grudzień, styczeń, luty – zima
 - ❖ marzec, kwiecień, maj – wiosna
 - ❖ czerwiec, lipiec, sierpień - lato
 - ❖ wrzesień, październik, listopad - jesień

Jak pora roku wpływa na czas podróży?

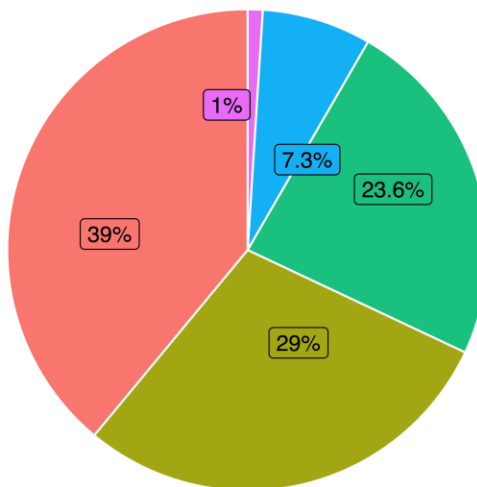
Wyniki analizy zostały przedstawione na wykresach kołowych wraz z wartością procentową.

Obserwujemy, że najczęściej bardzo krótkich podróży występuje zimą, a najmniej latem.

Ilość podróży krótkich jest porównywalna dla każdej pory roku, największe zmiany zachodzą dla podróży bardzo krótkich oraz średnich.

Proporcje ilości podróży każdego rodzaju są porównywalne dla wiosny i jesieni.

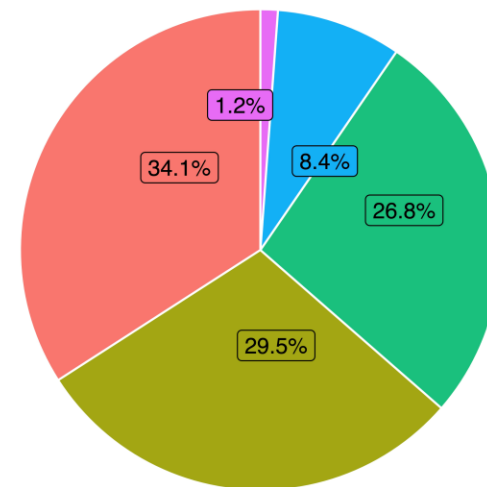
Spring



TypeOfTrip

- 1-UltraShort
- 2-Short
- 3-Medium
- 4-Long
- 5-UltraLong

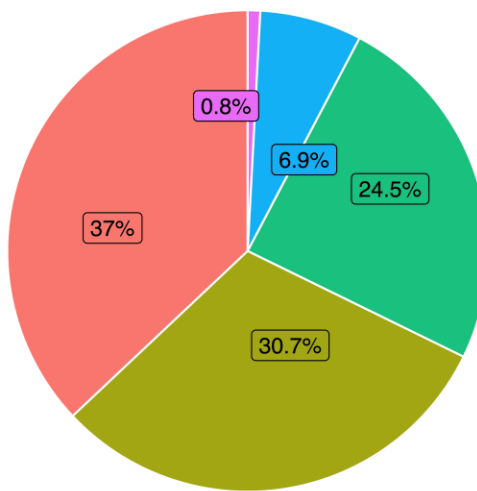
Summer



TypeOfTrip

- 1-UltraShort
- 2-Short
- 3-Medium
- 4-Long
- 5-UltraLong

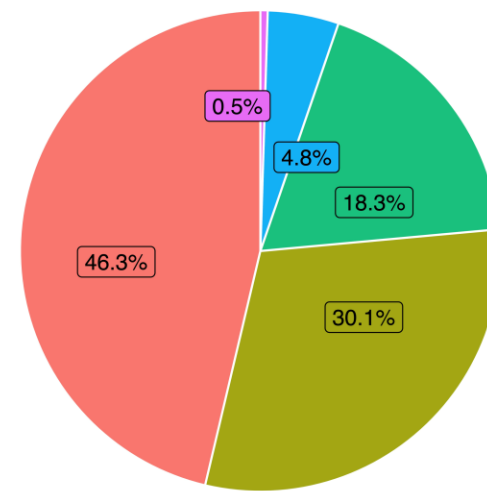
Autumn



TypeOfTrip

- 1-UltraShort
- 2-Short
- 3-Medium
- 4-Long
- 5-UltraLong

Winter



TypeOfTrip

- 1-UltraShort
- 2-Short
- 3-Medium
- 4-Long
- 5-UltraLong

Jak płeć użytkowników wpływa na czas podróży?

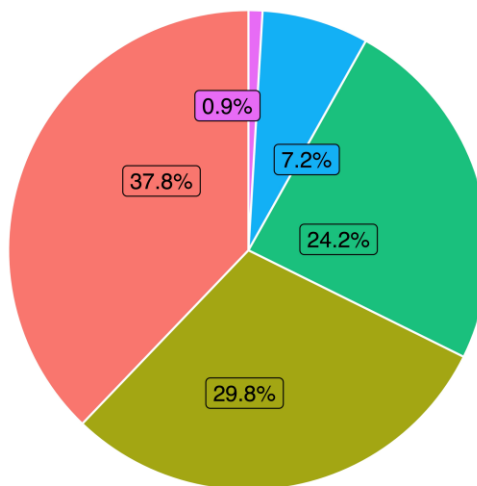
Wyniki zostały przedstawione na wykresach kołowych wraz z wartością procentową.

Rezultaty tego badania mogą być bardziej zaskakujące.

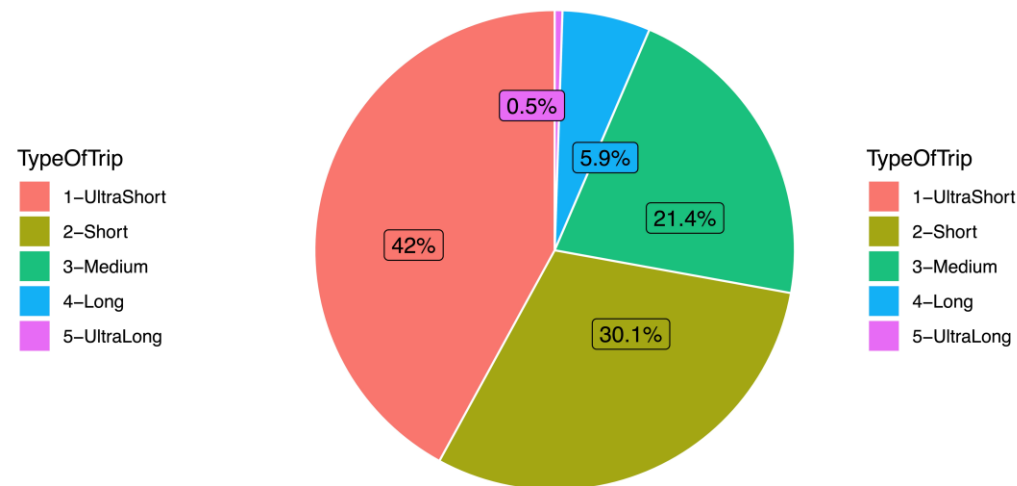
Zdecydowanie więcej bardzo krótkich podróży odbywają mężczyźni, kosztem podróży średnich i długich, które w głównej mierze odbywają kobiety.

Skłaniać do przemyśleń może natomiast zestawienie czasu trwania podróży użytkowników, którzy nie podali swojej płci. Widzimy że zdecydowana większość to podróże średnie, ponadto dużą część stanowią podróże długie i bardzo długie.

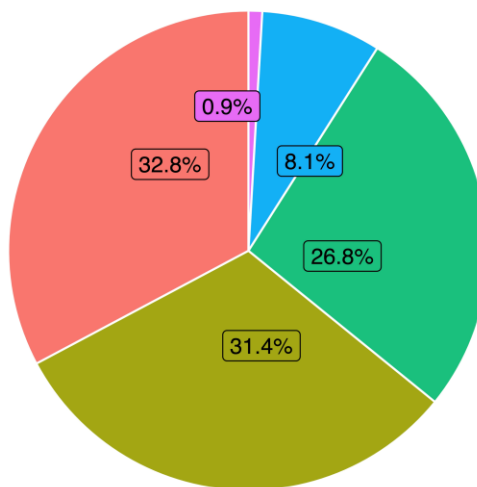
All genders



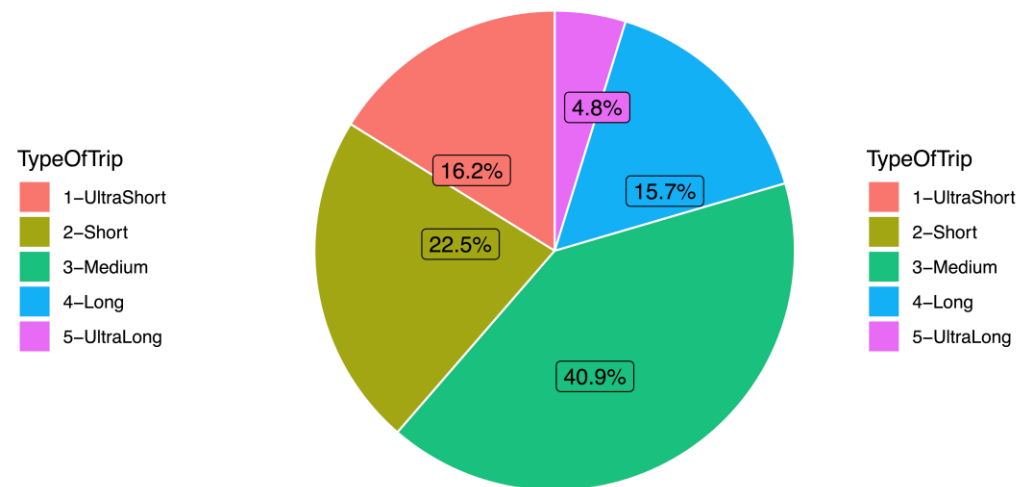
Men

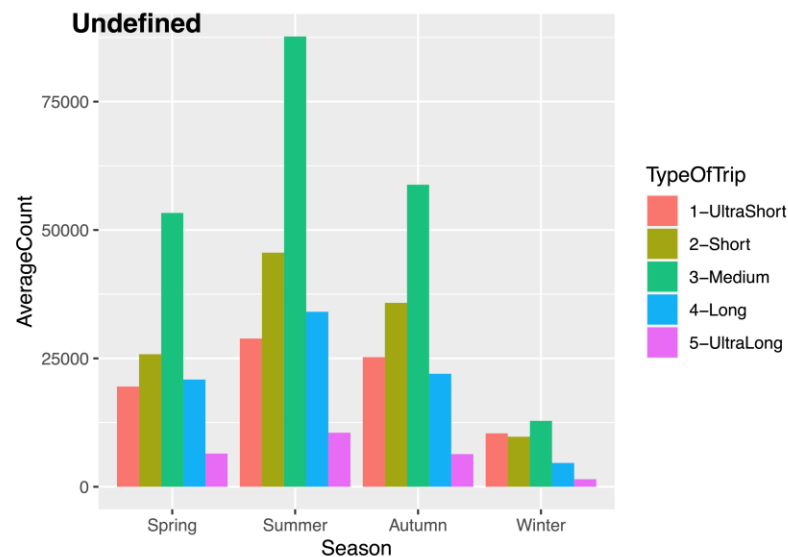
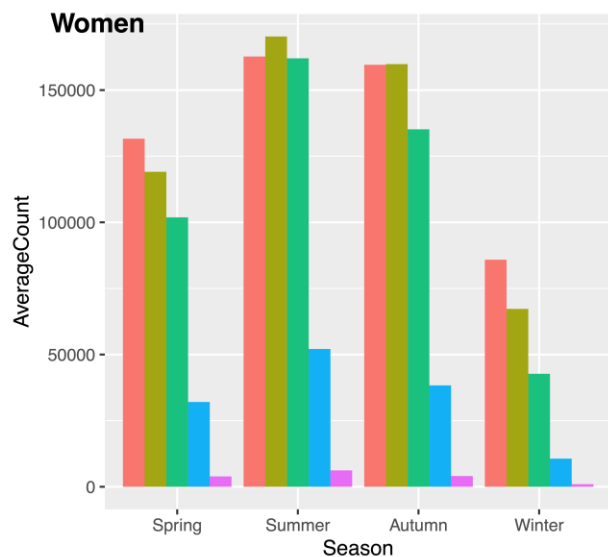
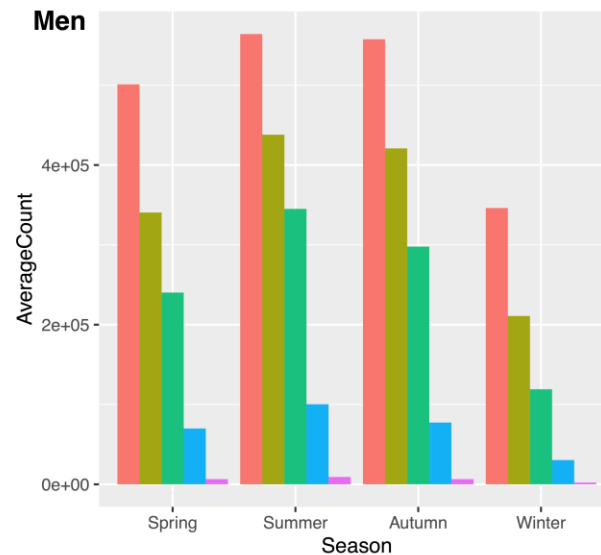
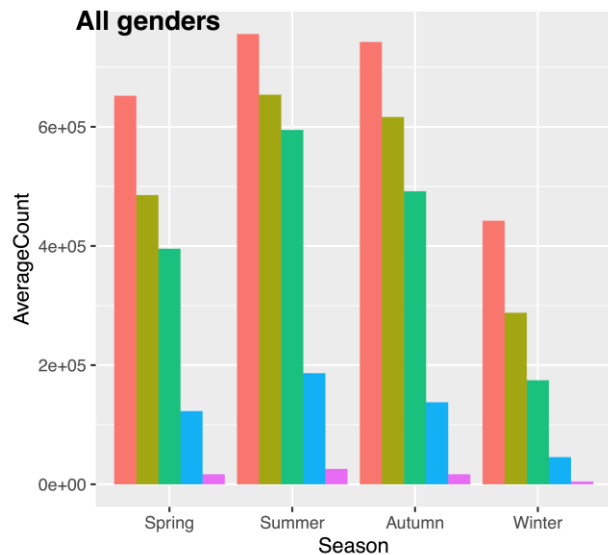


Women



Undefined





Pokusiliśmy się o zestawienie obu wspomnianych analiz na jednym wykresie, tym razem kolumnowym.

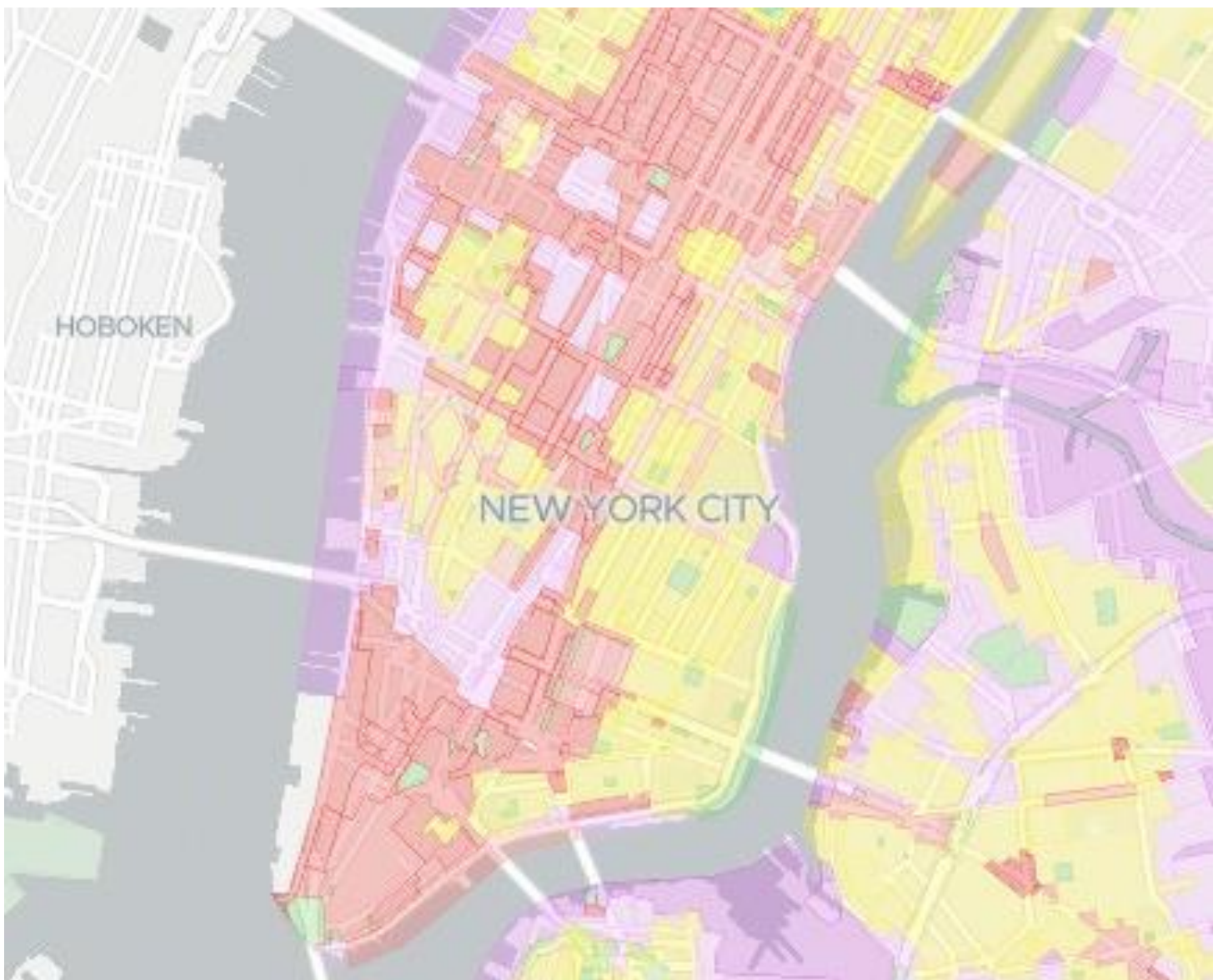
Dane dotyczące ilości rodzajów podróży dla każdej pory roku są średnią z odpowiadających im trzech miesięcy.

Dodatkowe obserwacje, jakie możemy poczynić, są związane z różnicami w proporcjach pomiędzy rodzajami podróży dla różnych pór roku.

Przykładowo dla kobiet widzimy, że latem oraz jesienią ilość podróży bardzo krótkich, krótkich oraz średnich była porównywalna, natomiast dla niezdefiniowanej płci najwięcej było podróży o średnim czasie trwania latem.

Sprawdzanie typu dzielnic

TRZECIA ANALIZA



Podział dzielnic na mieszkaniowe i pracownicze

Okazuje się, że w Nowym Jorku są wydzielone dystrykty mieszkaniowe i biznesowe.

Do dzielnic mieszkaniowych należą między innymi: East Village, Williamsburg, Upper West Side.

Do dzielnic biznesowych należą: Midtown, Financial District.

Przygotowanie do badania

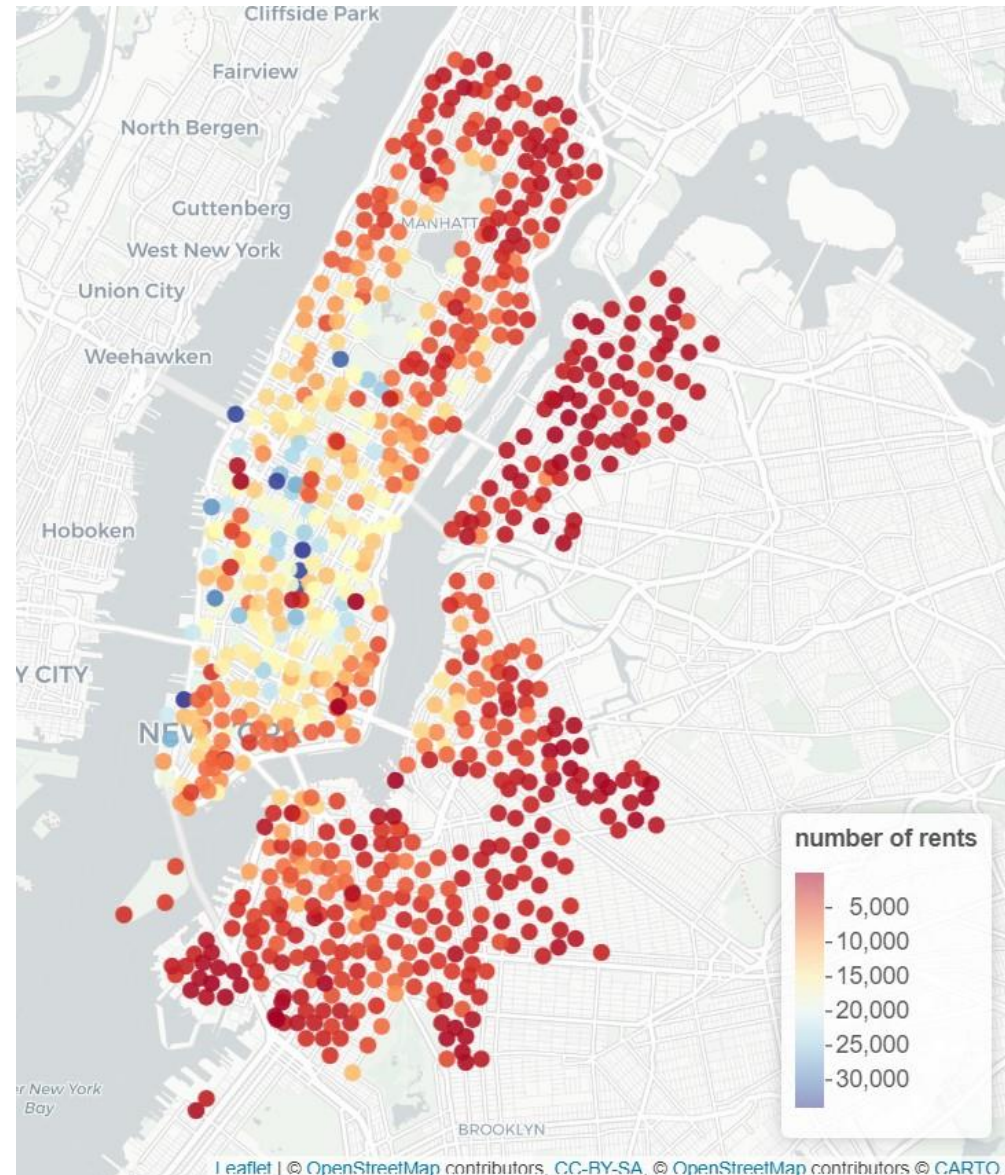
- ❖ Wykorzystałem ramki danych z 04.2019 – 06.2019
- ❖ Przyjąłem, że nowojorczyacy zaczynają pracę w godzinach 7:30 – 9:30 oraz, że kończą ją 15:30 – 17:30.

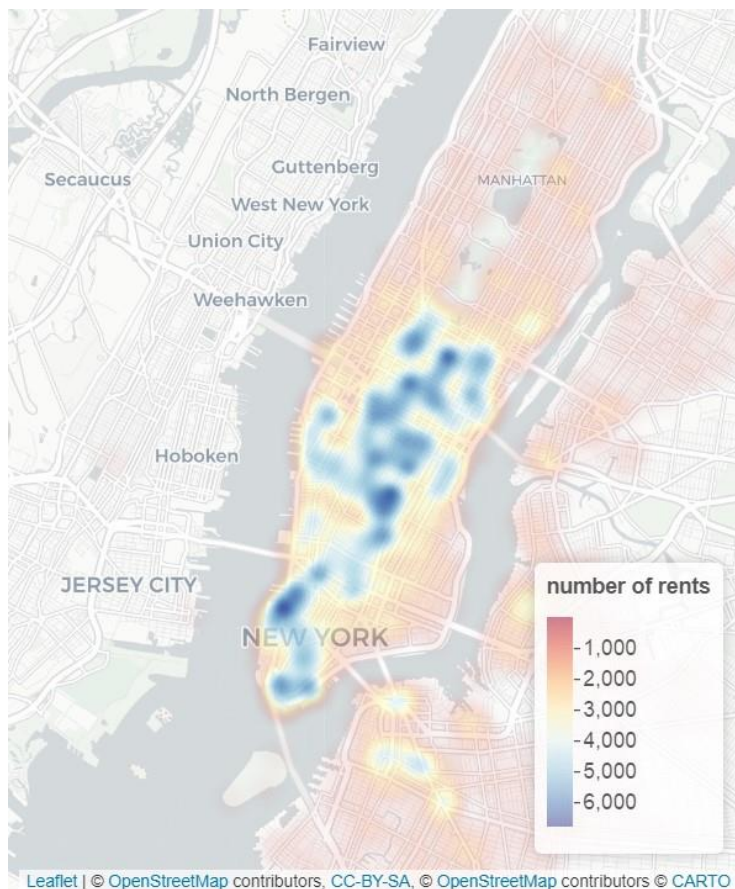
Mapa stacji

Pierwsza wizualizacja danych –
naniesienie stacji na mapę Nowego
Jorku.

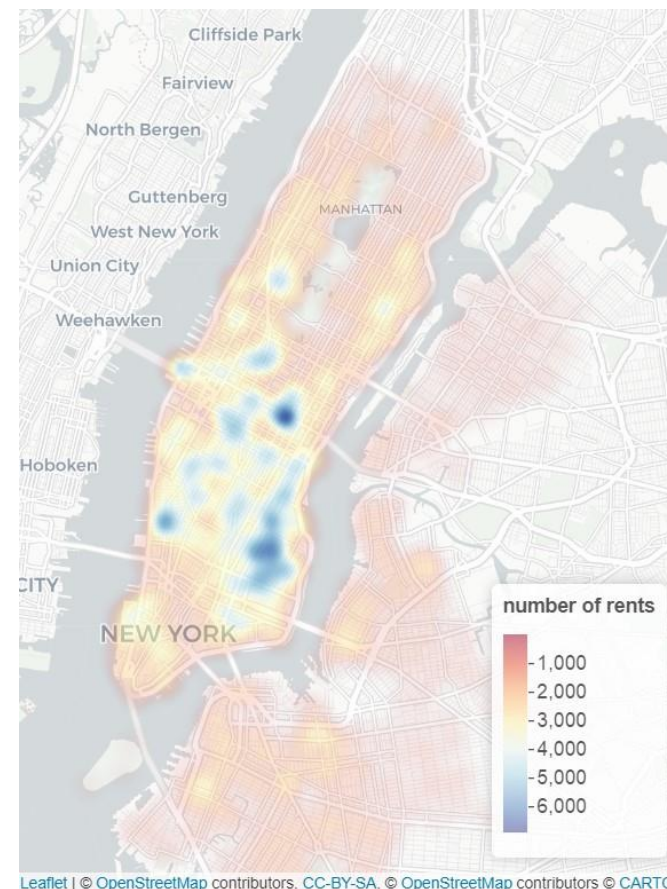
Każda stacja została pokolorowana na
kolor reprezentujący liczbę
wypożyczeń rowerów na stacji

Jednak przedstawiona wizualizacja nie
pozwala na wyciągnięcie
odpowiednich wniosków



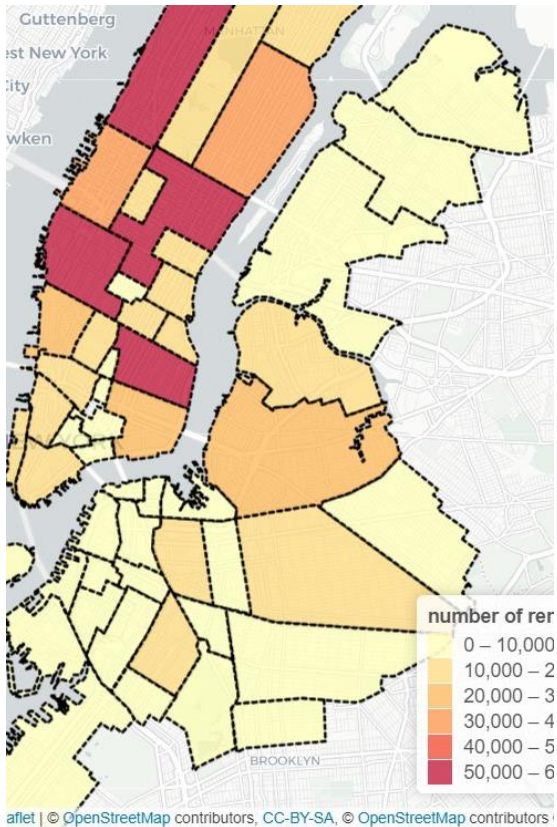


Przyjazdy nad ranem (7:30 – 9:30)

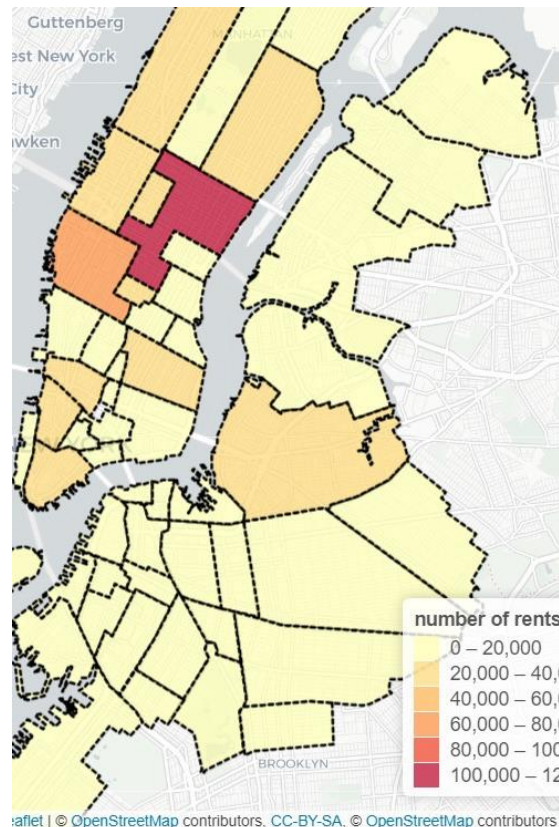


Odjazdy nad ranem (7:00 – 9:00)

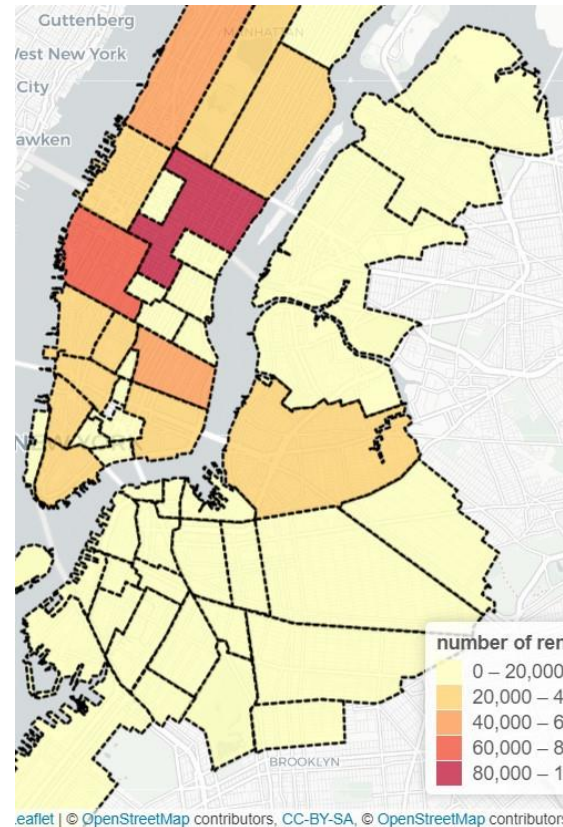
Mapa ciepła



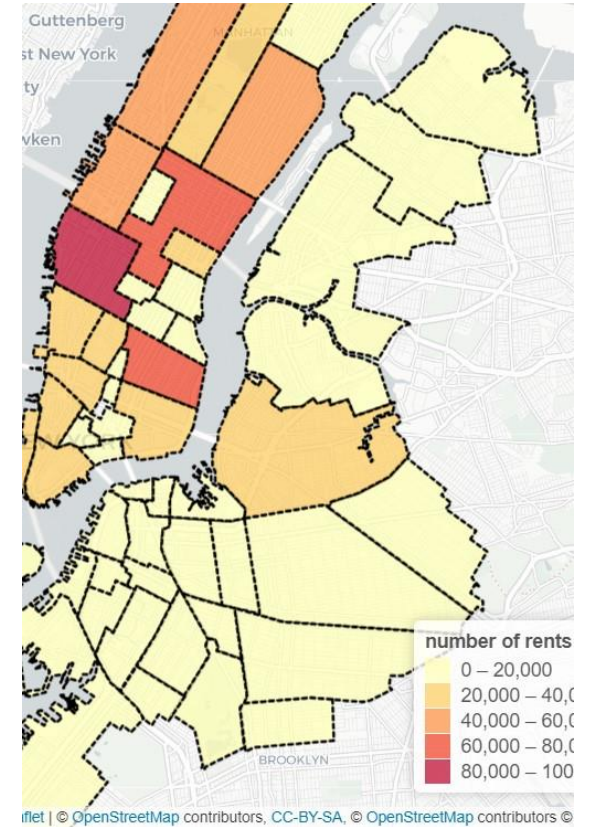
Odjazdy nad ranem (7:00 – 9:00)



Przyjazdy nad ranem (7:30 – 9:30)



Odjazdy po południu (15:30 – 17:30)

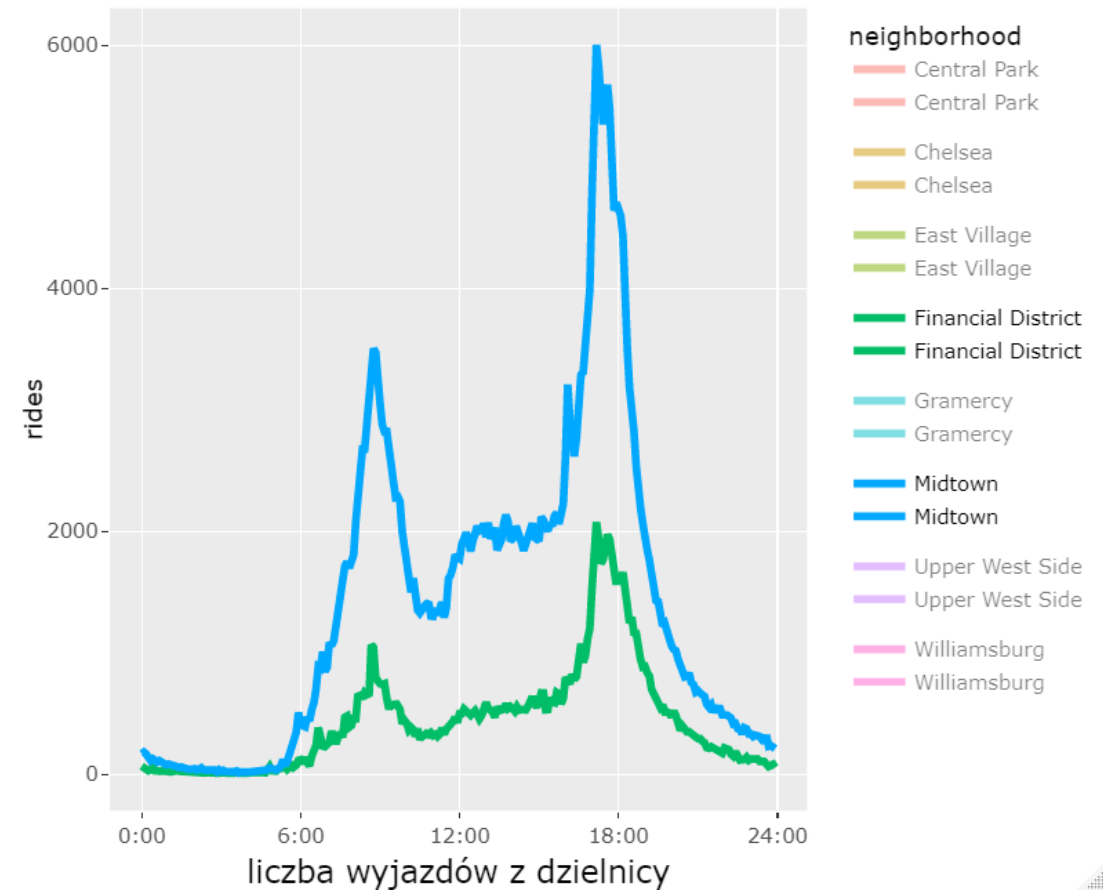
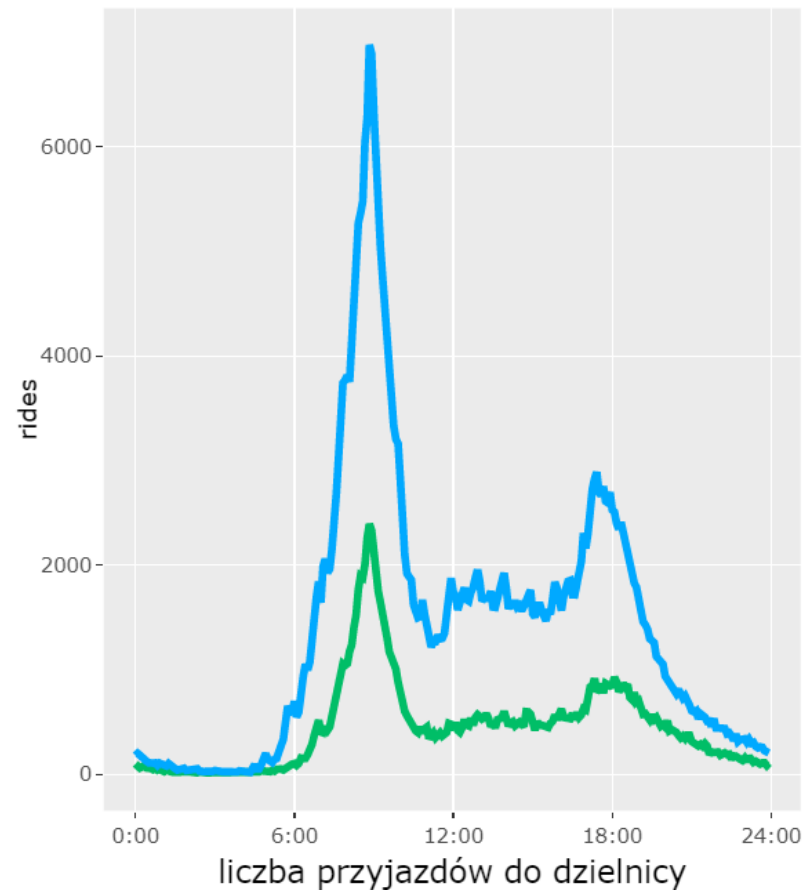


Przyjazdy po południu (16:00 – 18:00)

Kartogramy prezentujące wypożyczenia rowerów w różnych porach dnia

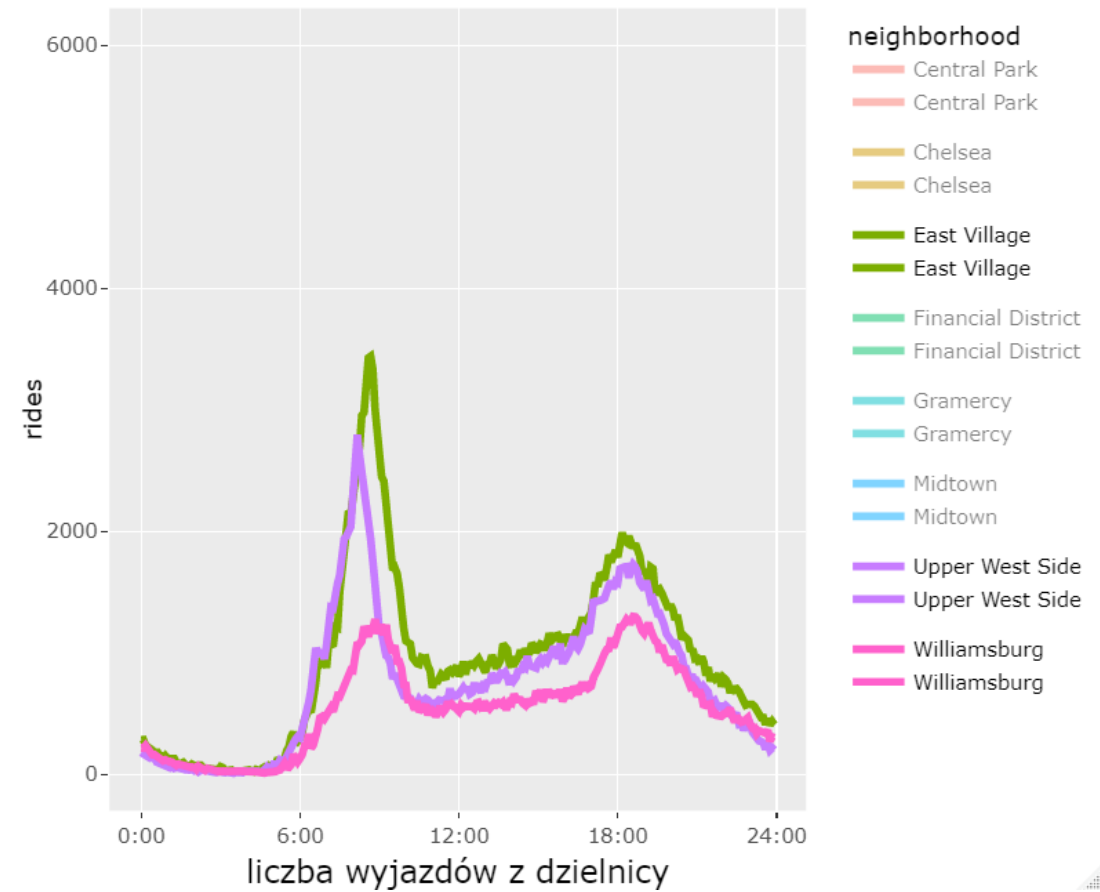
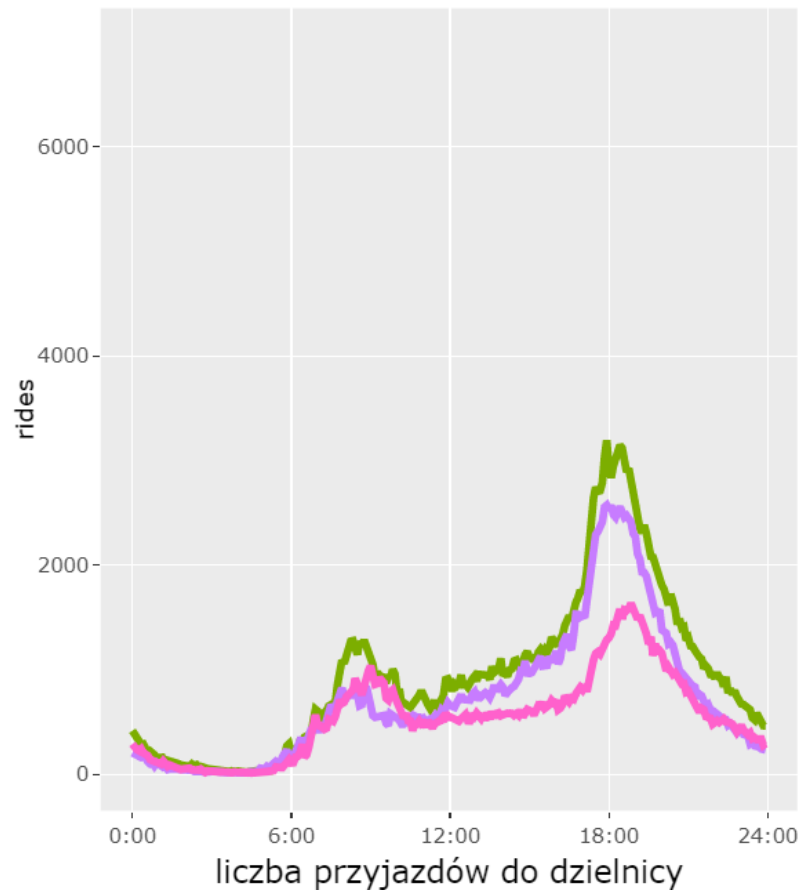
Na podstawie powyższych map można już wyciągnąć pewne wnioski – dzielnice **East Village** oraz **Upper West Side** mają intensywniejsze kolory w czasie charakterystycznym dla dzielnic mieszkaniowych.

Liczba odjazdów oraz przyjazdów do dzielnic



Ostateczne wykresy – dzielnice biznesowe

Liczba odjazdów oraz przyjazdów do dzielnic



Ostateczne wykresy – dzielnice mieszkaniowe

Podsumowanie

- ❖ Dla dzielnic o największym ruchu, takich jak: **East Village**, **Upper West Side** możemy z łatwością stwierdzić, że są dzielnicami mieszkaniowymi, zaś **Financial District** oraz **Midtown** są biznesowe – widać znaczną różnicę w liczbie wypożyczeń rowerów o danych godzinach
- ❖ Niestety z otrzymanych wizualizacji ciężko jest stwierdzić, jakiego typu są inne dzielnice. W dzielnicy mieszkaniowej **Williamsburg** można zaobserwować jedynie nieznaczne zwiększenie liczby przyjazdów rowerami do dzielnicy.
- ❖ W przypadku przeważająco biznesowych dzielnic **Gramercy** oraz **Chelsea** różnice wypożyczeń były nieznaczne, ale opisywany trend wciąż był obecny

Wniosek

Z danych opisujących użytkowanie rowerów miejskich można wysnuć wniosek, że:

Jesteśmy w stanie (z pewną dokładnością) określić charakter dzielnicy – mieszkaniowy lub biznesowy, na podstawie analizy liczby przyjazdów oraz odjazdów na stacje w godzinach rozpoczęcia i zakończenia pracy.



Do przetworzenia i wizualizacji danych korzystaliśmy z pakietów takich jak *data.table*, *dplyr*, *ggplot2* czy *plotly*.



Do wizualizacji danych przedstawionych na mapach użyliśmy ponadto pakietu *Shiny*, a także biblioteki *Leaflet* do obsługi map.



Realizując projekt korzystaliśmy z repozytorium
GitHub: <https://github.com/Fersoil/pdu3.git>.

Dodatki

Dziękujemy za uwagę
