

Praca projektowa nr3

Marcin Drzewiecki & Oskar Werner

Wstęp

Celem poniższej prezentacji jest zapoznanie odbiorcy z problemami badawczymi i wynikami implementacji analizy danych, jaką przeprowadziliśmy w związku z pracą projektową nr 3.

W naszej analizie pracowaliśmy na danych udostępnionych przez NYC Bike Share, LLC dotyczących klientów korzystających z systemu rowerów miejskich oraz ich podróży z 5 miesięcy w roku 2018.

Biblioteki których używaliśmy:

- dplyr
- ggplot2

Uznaliśmy, że najlepsza struktura danych dla naszej analizy to jedna ramka która została stworzona w następujący sposób. Do każdej ramki dodaliśmy kolumnę o nazwie “miesiąc” w której znajdowała się nazwa miesiąca, a potem złączyliśmy wszystkie zmodyfikowane ramki wierszowo.

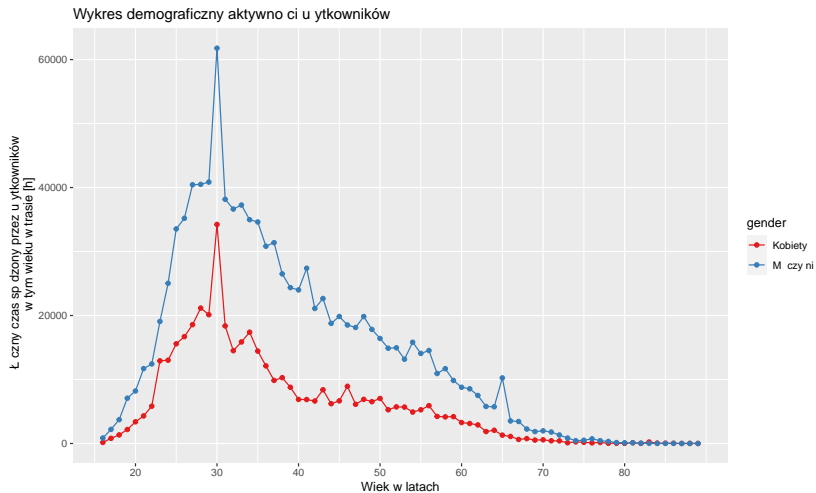
Problem 1.

Pierwszym doświadczeniem, jakie chcieliśmy przeprowadzić nad danymi z BikeShare, była analiza demograficzna pod kątem płci i wieku względem czasu, jaki użytkownicy spędzili korzystając z usług aplikacji.

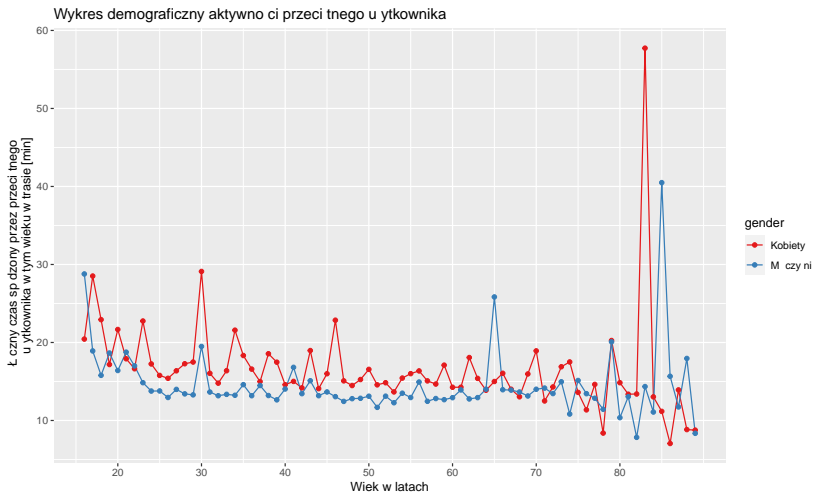
Pewne hipotezy, które sformułowaliśmy jeszcze przed rozpoczęciem pracy z danymi:

- Wraz ze wzrostem wieku aktywność w aplikacji spada w badanej grupie.
- Aktywność mężczyzn w aplikacji jest większa w każdym wieku.

Wyniki problemu 1.



Jak wyglądają dane dla przeciętnego użytkownika?



Wnioski z problemu 1.

- Wraz ze wzrostem wieku aktywność w aplikacji badanej grupy badawczej maleje
- Aktywność mężczyzn w aplikacji jest większa w każdym wieku.
- Proporcja aktywności mężczyzn i kobiet w aplikacji w różnych grupach wiekowych jest zmienna.
- najwyższą aktywnością wykazują się użytkownicy w wieku ok. 30 lat (a nie jak przypuszczaliśmy najmłodsi użytkownicy).

Przypuszczenia

Przyczyną tego może być zarówno potencjalna większa potrzeba korzystania z usług aplikacji u tej starszej grupy wiekowej, tudzież mniejsza popularność aplikacji wśród młodych osób.

Problem 2

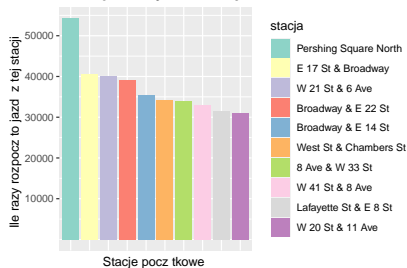
Kolejnym zagadnieniem, jakie chcieliśmy poruszyć jest częstość odwiedzania danych stacji przez użytkowników aplikacji.

Nasze hipotezy:

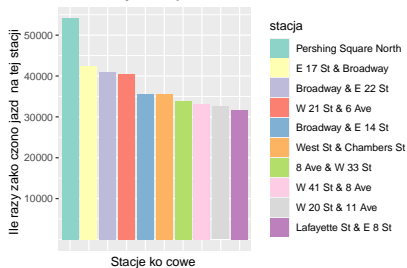
- Stacje, na których najczęściej rozpoczynają się trasy, równie często są stacjami, na których trasy się kończą.
- Wśród najczęściej odwiedzanych stacji są te w okolicach Broadwayu czy Central Parku

Wyniki

Top 10 stacji z których rozpoczyna się jazd



Top 10 stacji na których ko czy si jazd



Wnioski

- Największego ruchu rowerowego możemy się spodziewać w centrum miasta.
- Najczęściej odwiedzana stacja - t.j. Pershing Square North, znajduje się obok Central parku i dworca głównego.

Przypuszczenia

Może to świadczyć o tym, że użytkownikami BikeShare bywają często osoby korzystające z kolei.

Problem 3

Z jaką średnią prędkością przemieszczają się rowerami użytkownicy w różnym wieku?

Nasze hipotezy:

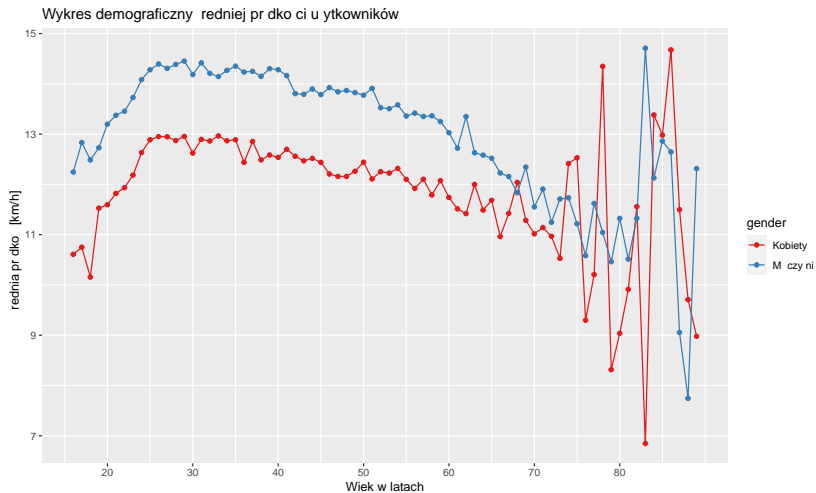
- Mężczyźni poruszają się szybciej od kobiet.
- Ze wzrostem wieku średnia prędkość użytkowników maleje.

Implementacja kodu

Dystans liczony po współrzędnych wg metryki taksówkowej.

Przyjęte 112km - 1 stopień geograficzny.

Wyniki



Wnioski

- Mężczyźni poruszają się szybciej od kobiet
- Największe wahania w średnich występują u najstarszych grup wiekowych

Przypuszczenia

Wahania te mogą wynikać z szeroko zróżnicowanej aktywności czy stanu zdrowia użytkowników w tym wieku, bądź z fałszywości danych użytkowników.

Problem 4

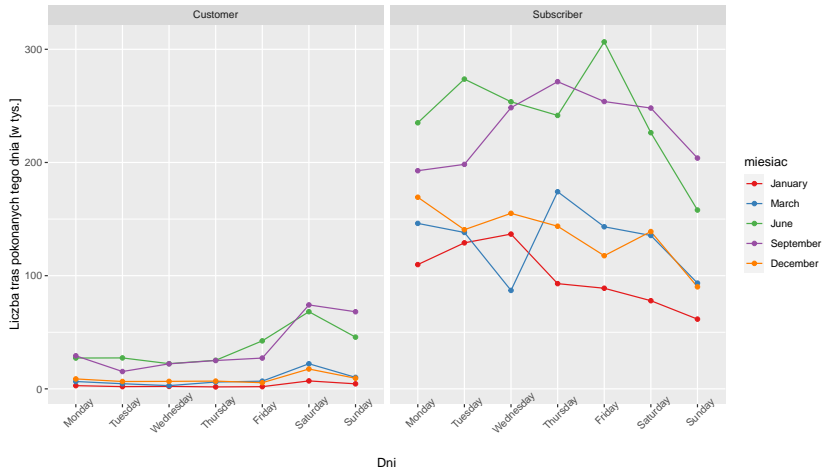
W których dniach tygodnia aktywność użytkowników w aplikacji jest największa?

Nasze hipotezy:

- Użytkownicy jednorazowi najczęściej jeżdżą rowerem w weekend.
- Stali subskrybenci jeżdżą równie często w każdym dniu tygodnia.

Wyniki

Najpopularniejsze dni



Wnioski

- Wykres może wskazywać na wypożyczanie rowerów przez jednorazowych użytkowników do przejechania weekendowej wycieczki rowerowej.
- Wykres nie wykazuje jednoznacznie popularności dni dla stałych subskrybentów.

Problem 5

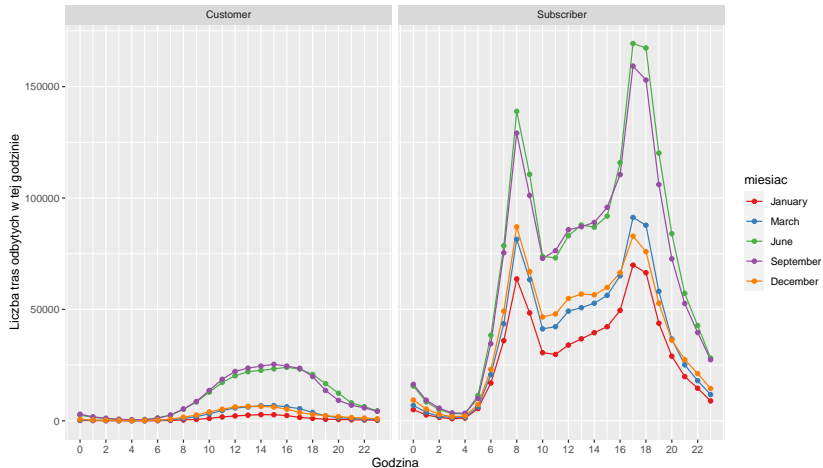
W jakich godzinach użytkownicy jednorazowi oraz stali subskrybenci są najbardziej aktywni?

Nasze hipotezy:

- Jednorazowi użytkownicy najczęściej jeżdżą na rowerze w godzinach popołudniowych.
- Aktywność stałych subskrybentów w aplikacji jest największa w godzinach szczytu.

Wyniki

Najpopularniejsze godziny



Wnioski

- Jednorazowi użytkownicy jeżdżą najczęściej w godzinach popołudniowych, co pokrywa się z wnioskiem z poprzedniego problemu badawczego, że użytkownicy rowerowi najczęściej używają rowerów do przejeżdżania wypraw rowerowych, ponieważ wyprawy rowerowe najczęściej się jeździ w weekend w popołudniowych porach.
- Aktywność subskrybentów jest największa w godzinach szczytu dwa razy w ciągu dnia, co może wskazywać na używanie roweru jako środka transportu do i z powrotem z pracy lub szkoły.

Dziękujemy za uwagę :)