Analiza wyników – zestaw zadań 3

Izabela Karczewska, Natalia Kołodziejczyk, Natalia Kubańska

A screenshot of a computer

Description automatically generated

# Natężenie ruchu rowerowego -nkolo

A screenshot of a data table

Description automatically generated

## Pogoda -nkolo

A screenshot of a data table

Description automatically generated

**Średnią liczbę rowerzystów przejeżdżających przez most w poszczególnych dniach tygodnia w sumie oraz z podziałem na zachodnią i wschodnią stronę mostu**

Na moście łączącym dzielnice Fremont i Queen Anne obserwujemy największy ruch rowerzystów we wtorki, a nieco mniejszy w środy i czwartki. Początkowo może to zaskakiwać, gdyż można by przypuszczać, że większość osób wybiera się na rower rekreacyjnie w weekendy. Jednak fakt, że most jest kluczowym połączeniem dla dojeżdżających do pracy, wyjaśnia ten trend. Zauważalny jest również spadek liczby rowerzystów w poniedziałki i piątki. Istnieje prawdopodobieństwo, że w dni te ze względu na około weekendowość część osób pracuje zdalnie. Dodatkowo, większość ruchu rowerowego koncentruje się na zachodniej stronie mostu.

A screenshot of a computer

Description automatically generated

**Sumaryczna liczba rowerzystów przejeżdżających przez most w poszczególnych dniach roku w sumie oraz z podziałem na zachodnią i wschodnią stronę mostu przez pierwsze dziesięć dni stycznia 2022 roku**

A table with numbers and numbers

Description automatically generated

**Średnia liczba rowerzystów przejeżdżających przez most w poszczególnych godzinach w sumie oraz z podziałem na zachodnią i wschodnią stronę mostu, a także z podziałem na dni robocze i weekend**

A screenshot of a graph

Description automatically generated

# **Wnioski na podstawie wcześniej wklejonych ramek oraz wykresu**

Zaczynając od tego co zostało powiedziane już w poprzednim punkcie, wyróżnia się tendencja do wybierania zachodniej strony mostu. Jej przyczyny nie są jednoznacznie wyjaśnione w ogólnodostępnych źródłach, co prowadzi do różnych hipotez. Możliwe, że zachodnia część mostu jest lepiej przystosowana dla rowerzystów, na przykład poprzez szersze pasy. Inna teoria sugeruje, że usytuowanie przejść na drugą stronę mostu może wpływać na wybór tej trasy, jak również lokalizacja ważnych punktów docelowych dla użytkowników tego szlaku. Nie można wykluczyć również technicznej usterki czujnika na wschodniej stronie mostu. Aby zweryfikować te przypuszczenia, warto by było przeprowadzić analizę porównawczą dla danych z różnych lat oraz dokładniej zbadać lokalizacje, do których najczęściej zmierzają użytkownicy. Ponadto, tendencja ta jest najbardziej wyraźna w dni robocze, zwłaszcza w godzinach szczytu, co może dawać dodatkowe wskazówki co do przyczyn preferencji tej strony mostu.

Ruch rowerowy na moście charakteryzuje się rocznym rozkładem przypominającym bardzo szeroką parabolę. Liczba rowerzystów jest mniejsza na początku roku, wzrasta w kolejnych miesiącach i utrzymuje się na stabilnym poziomie aż do jesieni, kiedy to zaczyna powoli spadać. Taki trend wydaje się być naturalnym wynikiem zmian temperatur, które bezpośrednio wpływają na preferencje co do wyboru środka transportu. Występują także dwie nietypowe zmiany liczby rowerzystów. Pierwsza miała miejsce w czwartek 07.04.2022, kiedy to odnotowano znaczący wzrost, mimo braku świąt czy ważnych wydarzeń. Tego dnia obchodzony jest jednak światowy dzień zdrowia, co mogło przyczynić się do akcji promujących jazdę na rowerze. Drugą nietypową obserwację zanotowano w niedzielę 03.07.2022, jednak nie odnaleziono informacji o wydarzeniu, które mogłoby wyjaśniać ten wzrost. Analiza pogody z tych dni pokazuje, że 07.04.2022 temperatura znacznie wzrosła, co mogło zachęcić do wyboru roweru jako środka transportu, natomiast 03.07.2022 odnotowano spadek temperatury i opady deszczu, co zwykle ogranicza chęć do aktywności na zewnątrz, zwłaszcza w niedzielę.

Tendencje do jazdy na rowerze w poszczególne dni tygodnia została omówiona w poprzedniej części. Analizując wykres, można również zauważyć, że zmiany w natężeniu ruchu rowerowego pomiędzy poszczególnymi dniami są znacznie łagodniejsze po wschodniej stronie mostu.

Przyjrzenie się rozkładowi liczby rowerzystów w poszczególnych godzinach wykazało zdecydowane różnice między dniami powszednimi a weekendami. W dni robocze szczególnie dużo rowerzystów obserwuje się w godzinach porannych 7-9 oraz popołudniowych 16-18, co jest związane z typowymi godzinami dojazdu do pracy czy szkoły oraz powrotami z tych miejsc. Natomiast w weekendy brak jest wyraźnych szczytów w ruchu rowerowym. Liczba rowerzystów układa się w rozkład bliski normalnemu z medianą przypadającą na godziny południowe.­

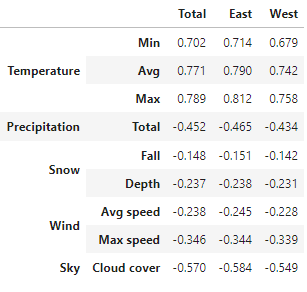
A screenshot of a graph

Description automatically generated

Statystyki miesięczne

A screenshot of a graph

Description automatically generated



# Korelacje liniowe pomiędzy parametrami pogody a liczbą rowerzystów tabela R

A screenshot of a computer

Description automatically generated

## R^2 wsp determinacji

A screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generated

# Korelacje liniowe między parametrami pogody[¶](http://localhost:8888/notebooks/OneDrive/Pulpit/Semestr6/AnalizaDanychWPythonie/Project3-DescriptiveStatisticsPart3/ZZ03_CZ03.ipynb#b)-Korelacje-liniowe-mi%C4%99dzy-parametrami-pogody) R

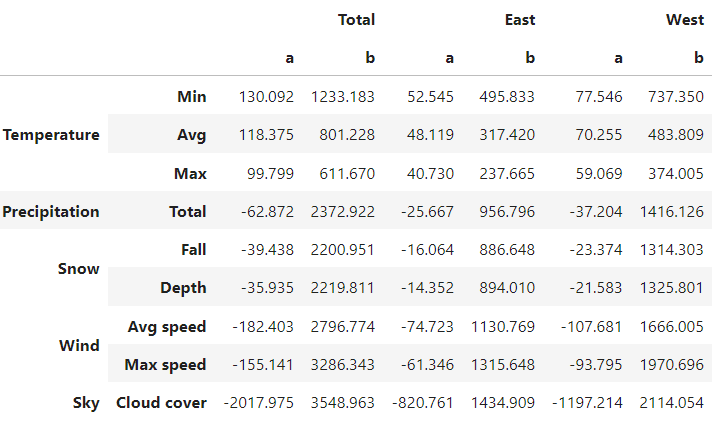
A screenshot of a graph

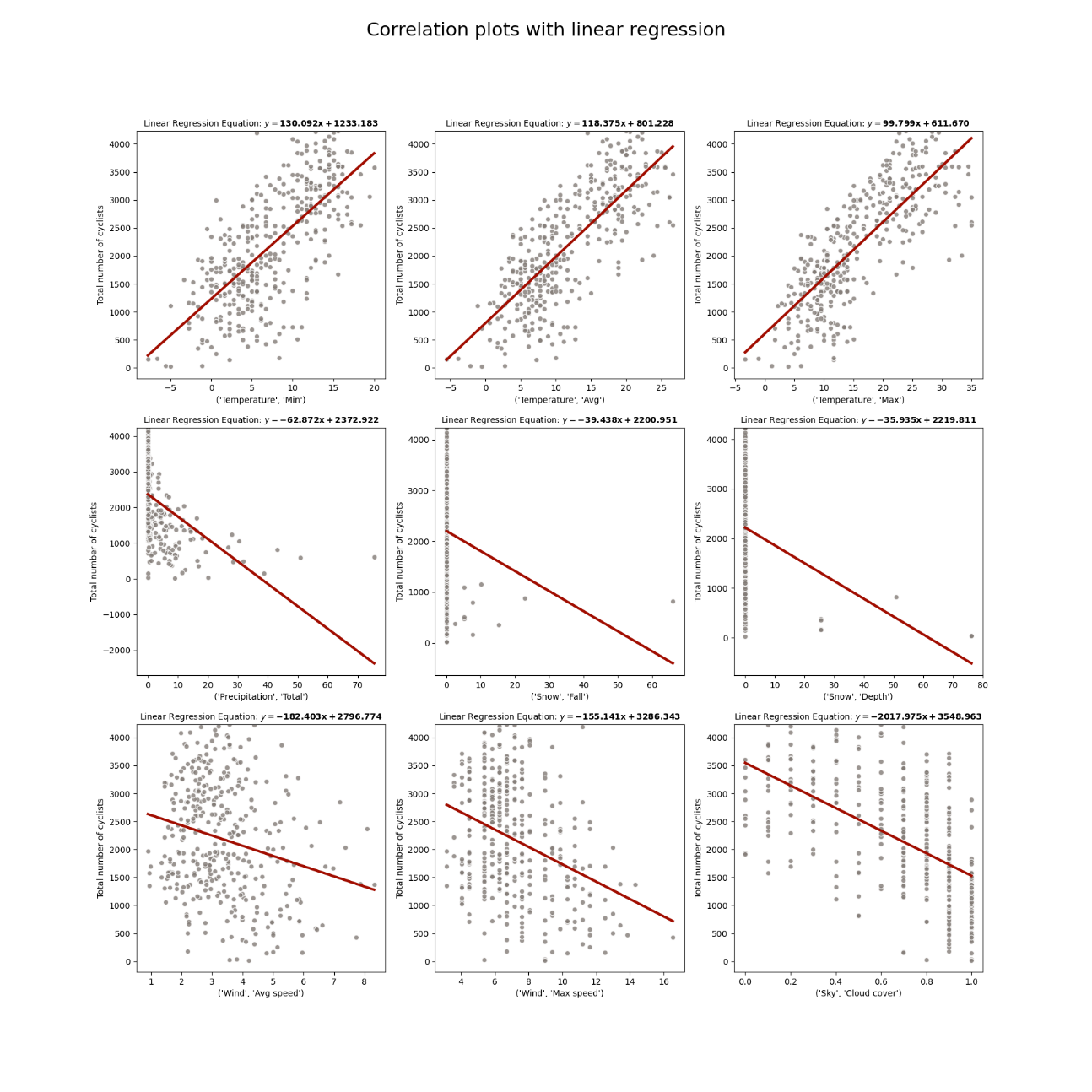
Description automatically generated

## R2

A screenshot of a color chart

Description automatically generated





**Regresja Liniowa**

Współczynnik zostały dopasowane zarówno do liczby rowerzystów po wschodniej, zachodniej i obu stronach mostu. Widać między nimi tendencję - dla łącznej liczby rowerzystów są największe, a potem kolejno dla zachodniej i wschodniej. Jest to poprawny objaw, ponieważ wcześniej omówiliśmy tendencje co do większej liczby obserwacji po zachodniej stronie. W dalszej części szerzej zostanie omówione dopasowanie regresji dla łącznej liczby rowerzystów po obu stronach mostu. Wnioski są analogiczne, a wykresy, które zostały wygenerowane tylko dla całkowitej liczby rowerzystów dobrze to obrazują.

Uzależnienie liczby rowerzystów od temperatury wydaje się trafnym wyborem – ma pozytywny wpływ na liczbę rowerzystów. Funkcje zarówno dla średniej, maksymalnej i minimalnej temperatury wydają się być dopasowane dobrze.

Między opadami a liczbą rowerzystów widać ujemną relację. Jednak stosowanie regresji liniowej do modelowania tej zależności może być nieoptymalne, co sugeruje potrzebę zastosowania bardziej złożonych metod statystycznych, które lepiej odzwierciedlą nieliniową naturę tego związku.

Objaśnianie ilości opadów śniegu oraz jego głębokości wydaje się być ciężkie do wymodelowania za pomocą regresji. Istnieje ujemna zależność jednak nie jest ona liniowa. Jest to bardziej tendencyjność. Wydaje się, że dane nie są wystarczająco dobrze zbalansowane pod tym kątem (w Seattle nie pada aż tak często śnieg) aby tworzyć model regresji liniowej na ich podstawie.

Średnia prędkość wiatru oraz maksymalna prędkość wiatru mają także negatywny wpływ na liczbę rowerzystów. Dane wskazują, że silniejszy wiatr może zniechęcać rowerzystów, jednak duże zróżnicowanie wartości sugeruje, że prosta regresja liniowa może być tu zbyt uproszczonym podejściem.

Większe zachmurzone wpływa na obniżenie liczby rowerzystów. Analiza zachmurzenia jest mierzona w skali co 10%. Pełniejsze wykorzystanie skali 0-100% mogłoby pozwolić na lepsze zrozumienie tej zależności.

A screenshot of a calculator

Description automatically generated

A graph of a curve

Description automatically generated with medium confidence

**Regresja Krzywoliniowa**

Użycie regresji krzywoliniowej do modelowania zależności między liczbą rowerzystów, a zachmurzeniem i opadami jest lepszym rozwiązaniem w szczególności dla tego drugiego. Funkcja wpasowała się w dane dobrze i odzwierciedla naturę tej zależności, w której od pewnej ilości opadów liczba rowerzystów utrzymuje się na zbliżonym poziomie. W przypadku zachmurzenia, choć zastosowanie modelu krzywoliniowego nie przynosi tak spektakularnych rezultatów, jak w przypadku opadów, nadal jest to bardziej adekwatne podejście niż prosta regresja liniowa.

**Uwagi do analizy**

* do lepszego zrozumienia tendencji przydatne wydaje się również przeanalizowanie danych z innych lat, które mogłyby utwierdzić nas w pewnych założeniach i dałaby bardziej szerokie spojrzenie
* Nie jest to problem bezpośrednio związany z samą analizą, a bardziej rzeczywistość, z którą będziemy musieli się godzić zapewne nie raz w pracy z danymi, ale nasuwa się ciekawość czym była spowodowana tendencja w preferencji zachodniej strony mostu. Ze względu na jego lokalizację ciężko jest mieć jakieś intuicyjnie przypuszczenia na ten temat nie będąc tam nigdy. Przykładowo, gdyby była mowa o moście Dębnickim i okazałoby się, że więcej osób przejeżdża rowerem po lewej stronie (idąc od AGH) nasuwałoby się, że jest to bliżej Wawelu i zazwyczaj więcej osób spędza czas właśnie po tej stronie
* Przydatne mogłoby okazać się stworzeniu również innych modeli regresji, przykładowo wielorakiej, aby zaobserwować łączny wpływ większej liczby czynników