***STATYSTYCZNA ANALIZA DANYCH – PROJEKT KOŃCOWY***

*Temat projektu: Zbiór danych zawiera liczbę rowerów publicznych wypożyczonych na godzinę w systemie udostępniania rowerów w Seulu, wraz z odpowiednimi danymi pogodowymi i informacjami o wakacjach.*

Obraz zawierający zrzut ekranu

Opis wygenerowany automatycznie

Poniższe badania mają na celu zbadanie wpływu liczby wypożyczeń rowerów w zależności od czynników atmosferycznych oraz sezonowych. Zakładamy, w cieplejsze pory będą miały pozytywny wpływ na tę wartość (będzie więcej wypożyczeni w miesiącach o wzmożonej aktywności słonecznje, większej temperaturze oraz w trakcie

Zmienna objaśniana:

* Liczba wypożyczeń

Zmienne objaśniające:

Ilościowe:

* Temperatura
* Ilość słońca

Jakościowe:

* Godzina
* Miesiąc
* Pora roku

Dla zmiennej objaśnianej wykonano statystyki opisowe oraz miary położenia i rozproszenia.

*Mediana = 542 –* wartość środkowa posortowanego zbioru danych

*Średnia = 729 –* średnia liczba wypożyczeń rowerów w danym miesiącu

*Moda = 122 –* najczęstsza liczba wypożyczeń

*Odchylenie standardowe = 642.35 –* miara odstawania wartości od średniej. 642 jest dużą wartością, co oznacza, że mamy do czynienia z

*Wariancja = 412615.02*

*1. kwantyl = 214*

*3. kwantyl = 1084*

*Kurtoza = 0.82*

Przedstawiono je na wykresie ramka-wąsy oraz histogramie:

Obraz zawierający zrzut ekranu, tekst, diagram, Prostokąt

Opis wygenerowany automatycznie

Mamy większość wartości odstających w maksimum, a minimum zamyka się w wykresie.

Obraz zawierający Wykres, zrzut ekranu, linia, diagram

Opis wygenerowany automatycznie

Każdy słupek reprezentuje średnią liczbę wypożyczonych rowerów na dany miesiąc, gdzie wartościami jest stan wypożyczeń na dany dzień. Wykres jest prawoskośny, co oznacza, że dane skoncentrowane są po lewej stronie (przez większość czasu liczba wypożyczeni była skoncentrowana przy niższych wartościach, przy medianie)

Dla skategoryzowanej zmiennej objaśniającej (miesiąc) wykonano histogramy, by przedstawić rozkład pojedynczych składowych.

W zależności od miesiąca:

styczeń: luty:

Obraz zawierający diagram, tekst, zrzut ekranu, Wykres

Opis wygenerowany automatycznie Obraz zawierający diagram, Wykres, linia, zrzut ekranu

Opis wygenerowany automatycznie

marzec: kwiecień:

Obraz zawierający diagram, Wykres, linia, zrzut ekranu

Opis wygenerowany automatycznie Obraz zawierający diagram, Wykres, linia

Opis wygenerowany automatycznie

maj: czerwiec:

Obraz zawierający diagram, Wykres, linia, zrzut ekranu

Opis wygenerowany automatycznie Obraz zawierający diagram, Wykres, linia, design

Opis wygenerowany automatycznie

lipiec: sierpień:

Obraz zawierający diagram, Wykres, linia, zrzut ekranu

Opis wygenerowany automatycznie Obraz zawierający diagram, Wykres, zrzut ekranu, linia

Opis wygenerowany automatycznie

wrzesień: październik:

Obraz zawierający diagram, Wykres, zrzut ekranu, linia

Opis wygenerowany automatycznie Obraz zawierający diagram, Wykres, design

Opis wygenerowany automatycznie

listopad: grudzień:

Obraz zawierający diagram, Wykres, zrzut ekranu, design

Opis wygenerowany automatycznie Obraz zawierający diagram, Wykres, Prostokąt, linia

Opis wygenerowany automatycznie

Widać, że najbardziej „spłaszczony” wykres był na wiosnę i zimę, a więc od marca do sierpnia, zahaczając o wrzesień.

Przy wnioskowaniu na temat zależności liczby wypożyczeni od temperatury nie może obejść się bez analizy tej zależności.

Obraz zawierający tekst, zrzut ekranu, diagram, linia

Opis wygenerowany automatycznieObraz zawierający diagram, Wykres, tekst, zrzut ekranu

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, diagram, linia

Opis wygenerowany automatycznie Obraz zawierający tekst, diagram, zrzut ekranu, Wykres

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, diagram, linia

Opis wygenerowany automatycznie Obraz zawierający diagram, Wykres, linia, zrzut ekranu

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, diagram, linia

Opis wygenerowany automatycznieObraz zawierający diagram, tekst, Wykres, zrzut ekranu

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, linia, diagram

Opis wygenerowany automatycznieObraz zawierający diagram, Wykres, zrzut ekranu, tekst

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, diagram, linia

Opis wygenerowany automatycznie Obraz zawierający diagram, tekst, Wykres, linia

Opis wygenerowany automatycznie

W pierwszym przypadku ((-20) - (-10) stopni) trzeci kwartyl jest bardzo wąskim obszarem. Skala osiąga górną granicę wartości rzędu mniejszego od tysiąca, a powiększa się dopiero, gdy temperatura rośnie. Znaczny przyrost następuje przy przejściu z 0-10 do 10-20, gdzie trzeci kwartyl przybiera o wiele większy obszar przy wyższej wartości górnej. Prawa strona histogramu (wyższe wartości) jest bardziej spłaszczona, co oznacza, że więcej osób decyduje się na wybranie rowera jako główny środek transportu przy bardzo wysokiej temperaturze niż przy bardzo niskiej.

Idąc za ciosem, za zmienną objaśniającą obrano ilość słońca w MJ/m2

Obraz zawierający tekst, zrzut ekranu, diagram, linia

Opis wygenerowany automatycznie Obraz zawierający tekst, diagram, Wykres, zrzut ekranu

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, diagram, linia

Opis wygenerowany automatycznie Obraz zawierający diagram, Wykres, linia, zrzut ekranu

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, diagram, linia

Opis wygenerowany automatycznie Obraz zawierający diagram, tekst, Wykres, zrzut ekranu

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, diagram, linia

Opis wygenerowany automatycznieObraz zawierający diagram, Wykres, linia, zrzut ekranu

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, diagram, Prostokąt

Opis wygenerowany automatycznie Obraz zawierający diagram, Wykres, linia, zrzut ekranu

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, diagram, linia

Opis wygenerowany automatycznie Obraz zawierający diagram, Wykres, linia, zrzut ekranu

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, diagram, linia

Opis wygenerowany automatycznie Obraz zawierający diagram, Wykres, linia, design

Opis wygenerowany automatycznie

Przeprowadzono weryfikację hipotezy o zgodności rozkładu zmiennej zależnej z rozkładem normalnym. W tym celu wykorzystano test Shapiro-Wilka.

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

Wartość p jest mniejsza od wartości W, co oznacza, że nie mamy rozkładu normalnego w naszym zbiorze.

Zbadano związek korelacyjny pomiędzy badanymi zmiennymi:

Obraz zawierający tekst, zrzut ekranu, kwadrat, diagram

Opis wygenerowany automatycznie

Między badanymi zmiennymi za każdym razem występuje korelacja większa bądź równa korelacji słabej. Biorąc pod uwagę liczbę wypożyczeń rowerów, najsłabsza jest w przypadku miesięcy, ponieważ wiele czynników poza porą roku wpływa mocniej na decyzję ludzi w poszczególnych miesiącach. Najsilniejsza korelacja jest w zależności od temperatury i godziny w trakcie dnia.

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Oprogramowanie multimedialne

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Oprogramowanie multimedialne

Opis wygenerowany automatycznie

Obraz zawierający tekst, oprogramowanie, Oprogramowanie multimedialne, wyświetlacz

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, Oprogramowanie multimedialne

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, oprogramowanie, komputer

Opis wygenerowany automatycznie