

Wprowadzenie do analizy danych i uczenia maszynowego

Wykresy, matplotlib, seaborn

Franciszek Saliński

Koło Naukowe Data Science, Wydział MiNI PW

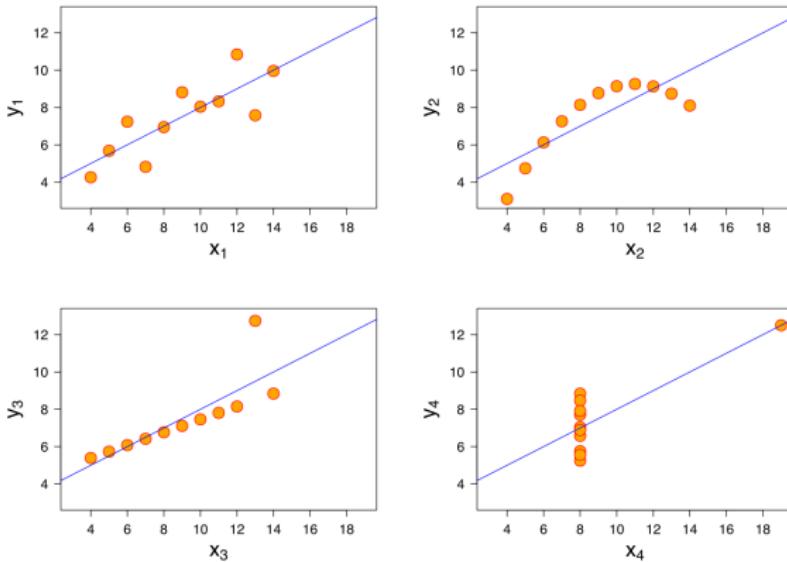
23 listopada 2025

Po co nam wizualizacja?

- Ułatwia zrozumienie danych i dostrzeżenie zależności.
- Pozwala zauważać trendy, wzorce i anomalie.
- Ujawnia aspekty niewidoczne w statystykach opisowych.
- Uatrakcyjniania sposób prezentacji wyników.
- Umożliwia jasne przedstawienie wyników analizy odbiorcom nietechnicznym.

Kwartet Anscombe'a

Każda z poniższych próbek ma identyczne lub niemal identyczne: średnią, wariancję próbkową, medianę, współczynnik korelacji, prostą regresji, współczynnik R^2 .



Rysunek: Kwartet Anscombe'a, źródło

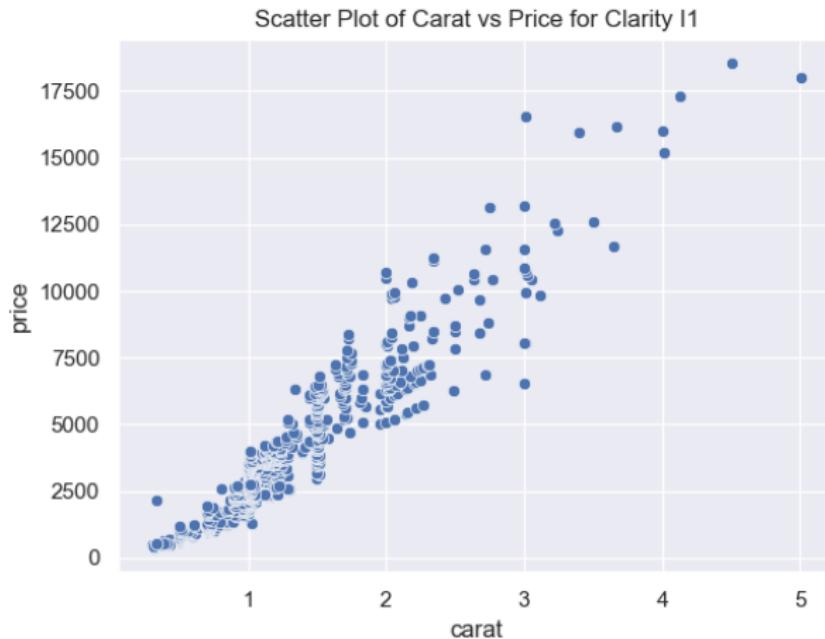
Przypomnienie podstawowych wykresów

Histogram



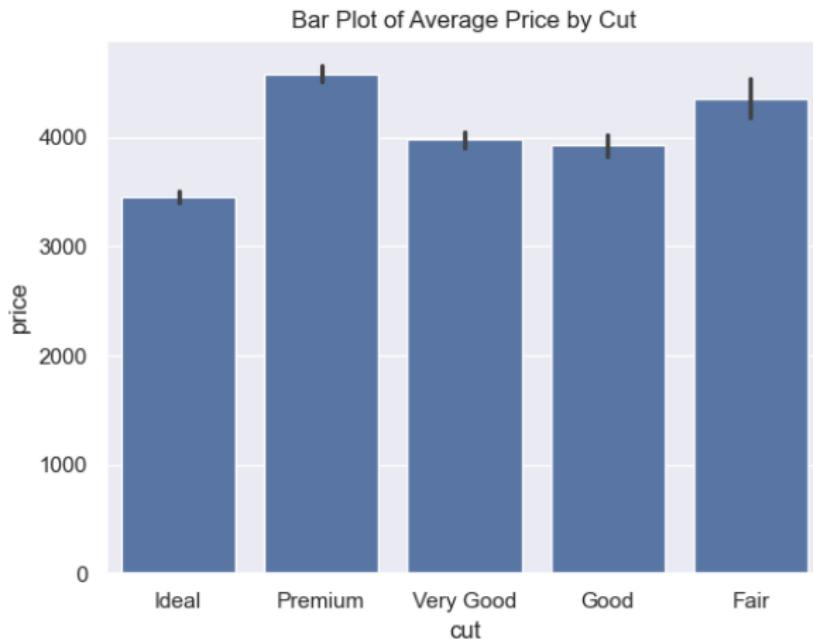
Przypomnienie podstawowych wykresów

Wykres punktowy



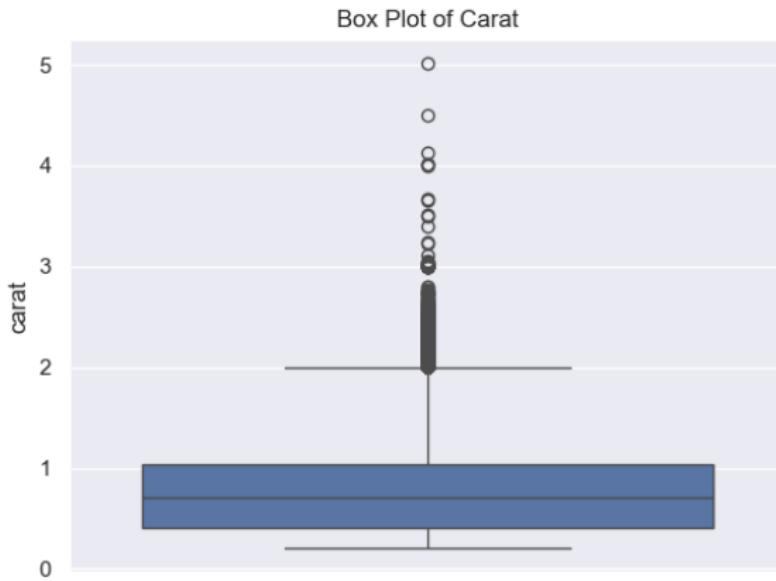
Przypomnienie podstawowych wykresów

Wykres słupkowy



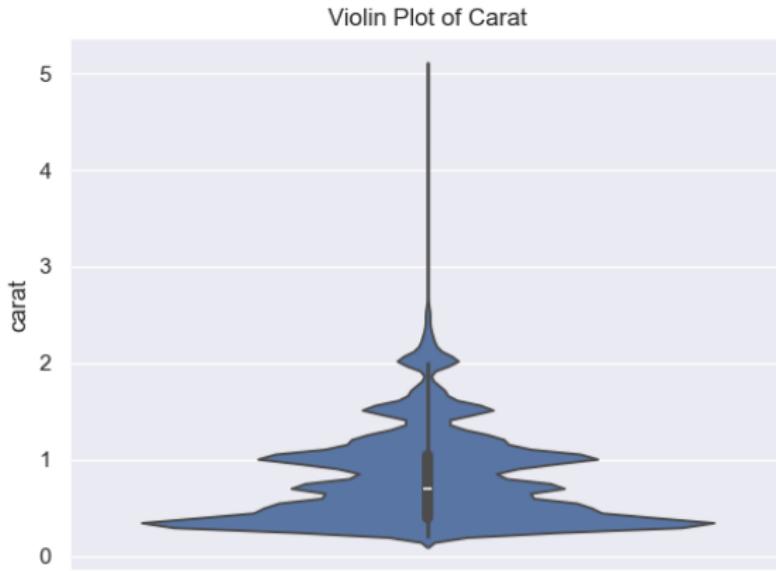
Boxplot

Wykres pudełkowy (boxplot / box-and-whiskers plot) – pozwala na wizualizację rozkładu zmiennej ciągłej, ukazując przy tym niektóre statystyki opisowane oraz wartości odstające

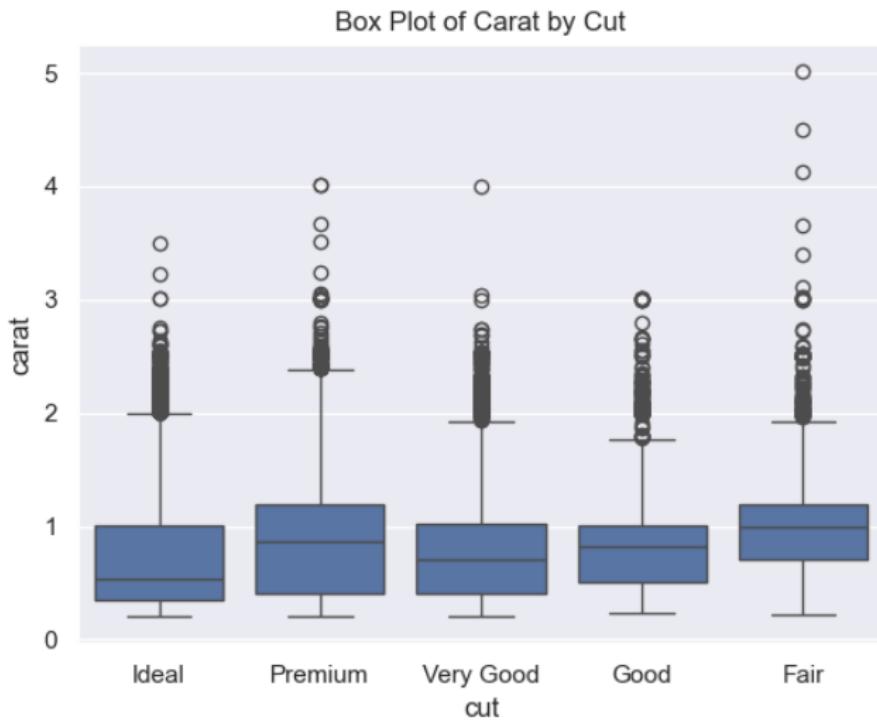


Violinplot

Wykres skrzypcowy (violinplot) – pozwala na wizualizację rozkładu zmiennej ciągłej, w tym estymowaną gęstość



Więcej wymiarów



Więcej wymiarów



Zasady poprawnej wizualizacji

- Dobieraj typ wykresu do typów i charakteru danych.
- Staraj się używać liniowej skali, zaczynającej się od zera.
- Ogranicz liczbę elementów – unikaj przeładowania wykresu.
- Stosuj czytelne kolory i oznaczenia, używaj ich tylko tam, gdzie są potrzebne.
- Używaj opisów osi, legendy i tytułu, ale tylko tam, gdzie są potrzebne.
- Zadbaj o odpowiedni kontekst: jednostki i źródła danych.