

Sprawozdanie z projektu analizy danych

Jan Kowalski

1 grudnia 2025

1 Wprowadzenie

Celem projektu była analiza zbioru danych dotyczącego sprzedaży diamentów. Badaliśmy zależności między różnymi cechami diamentów, takimi jak masa karatowa, kolor i cena.

2 Opis danych

Zbiór danych zawiera następujące zmienne:

- **carat** – masa diamentu w karatach,
- **cut** – jakość szlifu,
- **color** – kolor diamentu,
- **price** – cena w dolarach.

3 Analiza statystyczna

Średnia cena diamentów została obliczona według wzoru:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad (1)$$

gdzie X_i oznacza cenę i -tego diamentu, a n liczbę obserwacji.

Korelacja pomiędzy masą a ceną obliczana była przy użyciu współczynnika Pearsona:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}. \quad (2)$$

4 Wyniki

Analiza wykazała silną dodatnią korelację między masą diamentów a ich ceną. Oznacza to, że większe diamenty mają tendencję do wyższej ceny.

Dodatkowo, regresję liniową ceny na masę można przedstawić w postaci:

$$\hat{Y} = \beta_0 + \beta_1 X, \quad (3)$$

gdzie β_0 to wyraz wolny, a β_1 współczynnik kierunkowy regresji.

5 Wnioski

Analiza danych pozwoliła stwierdzić, że masa diamentu jest kluczowym czynnikiem wpływającym na jego cenę. W przyszłych badaniach warto uwzględnić także interakcje między jakością szlifu i kolorem.