

Wina zadania

Dane: **wina.csv**

Dane zawierają informacje z prawie 130 tysięcy win takie jak ocena, cena wina, nazwisko i nick recenzenta, treść recenzji oraz nazwa i pochodzenie wina.

Oceny przyznawane są winom w testach wg. następującej skali:

- 98–100 najwyższy poziom jakości
- 94–97 znakomite
- 90–93 doskonałe, bardzo zalecane
- 87–89 bardzo dobre, polecane
- 83–86 dobre, nadaje się do codziennego spożycia
- 80–82 akceptowalne

ZADANIA

1. Ile jest zmiennych, ile jest obserwacji? Jakiego typu są dane?
2. Wyznaczyć wina, które otrzymały ocenę co najmniej 94 oraz kosztują mniej niż \$25.
3. Wylosować 1% wszystkich obserwacji.
4. Określić 3 wina, które zdobyły najwięcej punktów.
5. Określić 100 najtańszych win.
6. Utworzyć listę win od największej do najmniejszej liczby punktów
7. Wybrać cztery zmienne (cechy) mówiące o pochodzeniu wina: country, oraz po kolejno zmienne od province do region_2.
8. Zamienić nazwę zmiennej (cechy) points na score.
9. Dodać zmienną “price_pln”, która będzie ceną wina w złotych. Przymijmy kurs dolara do złotego na poziomie 3,94.
10. Ile średnio kosztuje butelka ocenianego wina?
Dla ceny butelki wina wyznaczyć następujące kwantyle: 0, 0.1, 0.25, 0.50, 0.75, 0.9, 1
11. Wyznaczyć medianę dla ceny butelki wina.
12. Które wina mają najlepszy stosunek ceny do jakości?
13. Z otrzymanych wyników wyznaczyć tylko te obserwacje, które uzyskały 90 lub więcej punktów.
14. Który kraj produkuje najtańsze wina?
15. Dodać średnią ocenę dla win oraz liczbę win z danego kraju.