Eksploracja danych

Raport nr 1, grupa A

W pliku *cereals* zawierającym dane o różnych rodzajach płatków śniadaniowych interesować nas będą następujące predyktory:

* *Manuf*  – producent (jedna z firm oznaczonych jako A, G, K, N, P, Q, R),
* *Potass* – zawartość potasu,
* *Calories* – liczba kalorii,

oraz zmienna celu:

* *Shelf –* półka, na której wykładany jest ten produkt, licząc od podłogi – im wyższa, tym lepsze jakościowo i droższe produkty.

1. Wczytaj plik *cereals*.
2. Wykonaj zwykły i znormalizowany zestawiony wykres słupkowy dla zmiennej *Manuf,* oznaczając kolorem wartości zmiennej *Shelf*.
3. Zbuduj tabelę krzyżową dla zmiennych *Manuf* (w kolumnach) i *Shelf (w wierszach).* Dane umieszczone w tabeli powinny być procentowe (procent liczony w kolumnie).
4. Zidentyfikuj obserwacje odstające ze względu na wartości zmiennej *Potass.* Ile ich jest?
5. Wykonaj zwykły i znormalizowany zestawiony histogram dla zmiennej Potass, oznaczając kolorem wartości zmiennej Shelf.
6. Na podstawie zmiennej *Calories* utwórz nową zmienną *Calories\_binned* zgodnie ze schematem:
8. ,

Wykonaj zwykły i znormalizowany zestawiony wykres słupkowy dla zmiennej *Calories\_binned,* oznaczając kolorem wartości zmiennej *Shelf*.

Raport powinien być przygotowany w estetycznej formie w postaci pliku PDF, zawierającego tylko wyniki (tabele, wykresy) określone w poleceniach zadania. Wszystkie wyniki powinny zostać zinterpretowane i krótko skomentowane (wnioski). **Praca musi być samodzielna**. Jako rozwiązanie przesyłamy plik PDF oraz (osobno) wszystkie pliki robocze w postaci **archiwum ZIP**.