

Statystyka w biznesie - Projekt 2

Semestr letni 2024/25

Jako analityk biznesowy otrzymuje Pani/Pan zbiór danych pochodzący z systemu transakcyjnego przedsiębiorstwa. W zbiorze danych zawarte są informacje ilościowe o produktach (SKU) zakupionych przez klientów (PSD – punkt sprzedaży detalicznej) w analizowanym okresie oraz wartościach tych zakupów. Wykorzystując załączone dane należy przygotować raport opisujący strukturę sprzedaży przedsiębiorstwa oraz przedstawić propozycję segmentacji klientów (na podstawie podobieństwa ich zakupów). Wszystkie obliczenia powinny zostać wykonane w środowisku statystycznym R.

Raport nie powinien przekraczać 12 stron. Oczekiwana struktura i zawartość raportu przedstawia się następująco:

1. Wstęp – krótkie wprowadzenie do tematu analizy;
2. Opis zbioru danych – informacje o liczbie obserwacji, liczbie i rodzajach zmiennych, informacje o przygotowaniu zbioru do analizy (w szczególności, czy i dlaczego usunięto pewne obserwacje / rekordy z analizy); przedstawienie podstawowych statystyk opisowych dot. zmiennych w zbiorze, w tym analiza Pareto dla SKU i klientów;
3. Opis metodologii segmentacji – opisanie wykorzystanych metod segmentacji (w tym uzasadnienie wyboru danej metody) i miar wykorzystanych do określenia jakości przeprowadzonej segmentacji (optymalności liczby klastrów);
4. Wyniki segmentacji – przedstawienie i omówienie wyników segmentacji oraz profilowania (scharakteryzowania) segmentów (w tym zaprezentowanie wybranych tabel z wynikami i wykresów);
5. Podsumowanie.

Raport może być przygotowany w parach (dwie osoby, inna para niż przy projekcie 1). Termin na oddanie raportu to **15.06 godz. 23:59**. Raport powinien zostać zawieszony na Niezbędniku w sekcji zadania. Raport proszę przestać w formacie **docx** (tak, żebym mogła dodawać komentarze) oraz **pdf** (tak, żebym mogła wrócić w każdej chwili do wersji oryginalnej). Proszę załączyć także kod **R** (plik w formacie .R).

Raport jest obowiązkowy i wchodzi z wagą 25% w skład oceny końcowej z przedmiotu. Sam raport oceniany będzie na skali 0-25 pkt., w tym:

- 0-15 pkt – poprawność doboru metod analizy danych, poprawność opisu przeprowadzonych analiz i ich wyników
- 0-5 pkt – redakcja raportu, poprawność językowa
- 0-5 pkt – ocena kodu R (przejrzystość i poprawność kodu, w tym opatrzenie kodu komentarzami).

Po zakończeniu oceniania raportów otrzymają Państwo ode mnie indywidualne maile z ogólnymi uwagami, komentarzami w tekście raportu oraz punktacją.

W razie pytań proszę o kontakt.

Powodzenia!