

PAD 05

Praca Domowa 5 – Czyszczenie danych

Zadanie 1 (5 pkt)

Wgraj zawartość pliku orders.csv do ramki danych.

Przyjrzyj się danym, wykorzystując poznane metody (describe, info, head) zanalizuj dane.

Następnie:

- Zamień order_date na format daty. Upewnij się, że w rezultacie data będzie w formacie: rok-miesiąc-dzień;
- Sprawdź ile i jakie unikalne wartości znajdują się w kolumnie T-shirt category;
- Wyczyść kolumnę tshirt_category normalizując nazwy produktów. Upewnij się, że dla każdej kategorii funkcjonuje jedna nazwa, np. zamiast white i wh jest tylko jedna wersja white. Skorzystaj z klasy string i metody *apply*;
- Rozdziel kolumnę tshirt category na 3 kolumny: tshirt_type, która będzie mówiła o tym czy jest to koszulka, bluza czy coś innego, tshirt_size, która będzie informowała o rozmiarze koszulki, i tshirt_colour, która będzie informowała o kolorze. Skorzystaj z metody *apply*.
- Znajdź zamówienia złożone między 31 grudnia 2014 a 2 sierpnia 2016.

Zadanie 2 (4 pkt)

Wgraj plik customers.csv do ramki danych. Następnie:

- Wypisz kolumny;
- Porównaj rozmiar ramki z rozmiarem ramki w zadaniu 1;
- Zmień nazwę kolumny *customerID* na *customer_id* upewniając się, że dane zmiana zostanie zapisana w pamięci dzięki jednemu z argumentów metody do zmiany nazwy kolumny;
- Połącz ramkę orders z ramką customers po kolumnie customer_id. Z której metody skorzystasz? Zostaw odpowiedź w komentarzu do zadania w kodzie (np. blok komentarzem `''' '''`).

Zadanie 3 (1 pkt)

Nową ramkę (połączenie orders i customers) zapisz do pliku csv, ale zapisz bez kolumny z indeksem.