IUM Projekt – Etap 1 (zadanie 2)

Wojciech Nowicki, nr indeksu 304088

Gustaw Daczkowski, nr indeksu 304031

1. Definicja problemu biznesowego

“Są osoby, które wchodzą na naszą stronę i nie mogą się zdecydować, którym produktom przyjrzeć się nieco lepiej. Może dałoby się im coś polecić?"

Problemem biznesowym jest stworzenie systemu, który dla użytkownika odwiedzającego sklep internetowy będzie w stanie wygenerować rekomendacje produktów, którymi to potencjalny kupujący byłby zainteresowany. Ma to na celu zwiększenie dziennej liczby wyświetleń produktów, co może zwiększyć sprzedaż przekładając się na wzrost przychodów sklepu.

2. Zadanie modelowania oraz kryteria sukcesu

Zadaniem modelowania jest przygotowanie modelu rekomendacyjnego, który na podstawie dostarczonych danych (sesje użytkowników, aktualny katalog produktów, informacje o użytkownikach) będzie w stanie wybrać takie produkty z aktualnego katalogu, które użytkownik będzie skłonny zakupić. Analityczny kryterium sukcesu będzie współczynnik:

$$\frac{rekomendacje\ kliknięte\ przez\ użytkownika}{wszystkie\ rekomendacje}$$

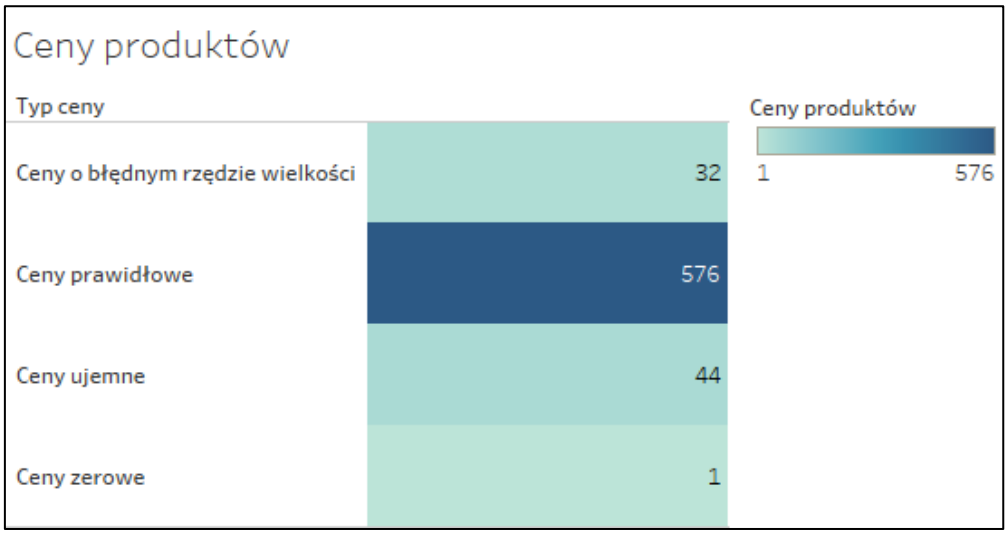
Wartość tego współczynnika należy ustalić z klientem tak aby odnieść go w sposób realny do obecnego ruchu w sklepie.

W ramach jednego zestawu rekomendacji będzie pojawiać się co najwyżej 5 przedmiotów sugerowanych dla danego użytkownika.

3. Analiza danych

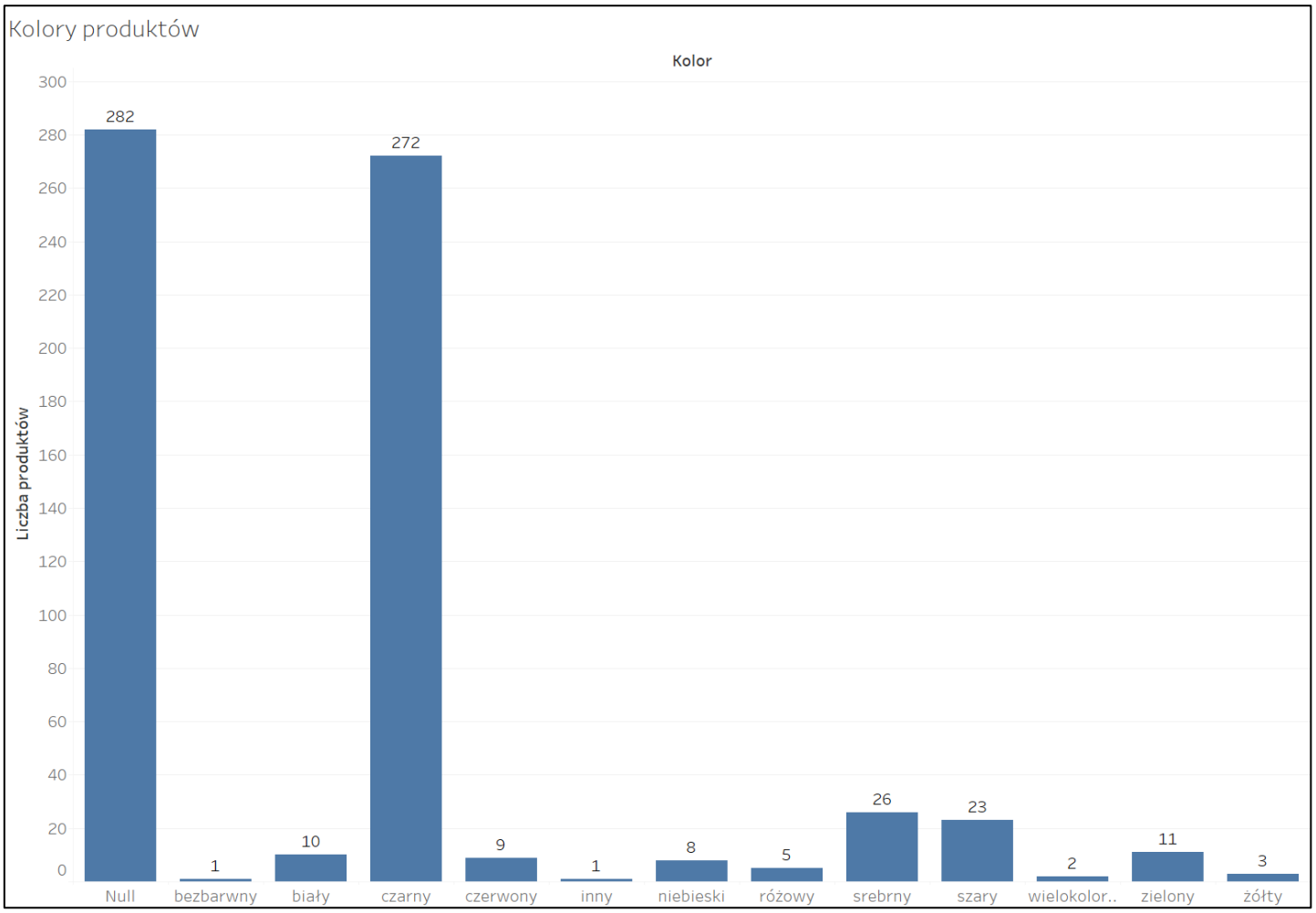
Produkty:

- Ceny produktów rozkładają się następująco:



Sugerowana poprawa danych: Do cen ujemnych użyć funkcji moduł. Produkty z cenami zerowymi usunąć z katalogu, a produkty z cenami o złym rzędzie wielkości spróbować naprawić (po konsultacji z klientem) ręcznie - o ile kontekst (czyli inne podobne produkty o właściwych cenach) na to pozwala.

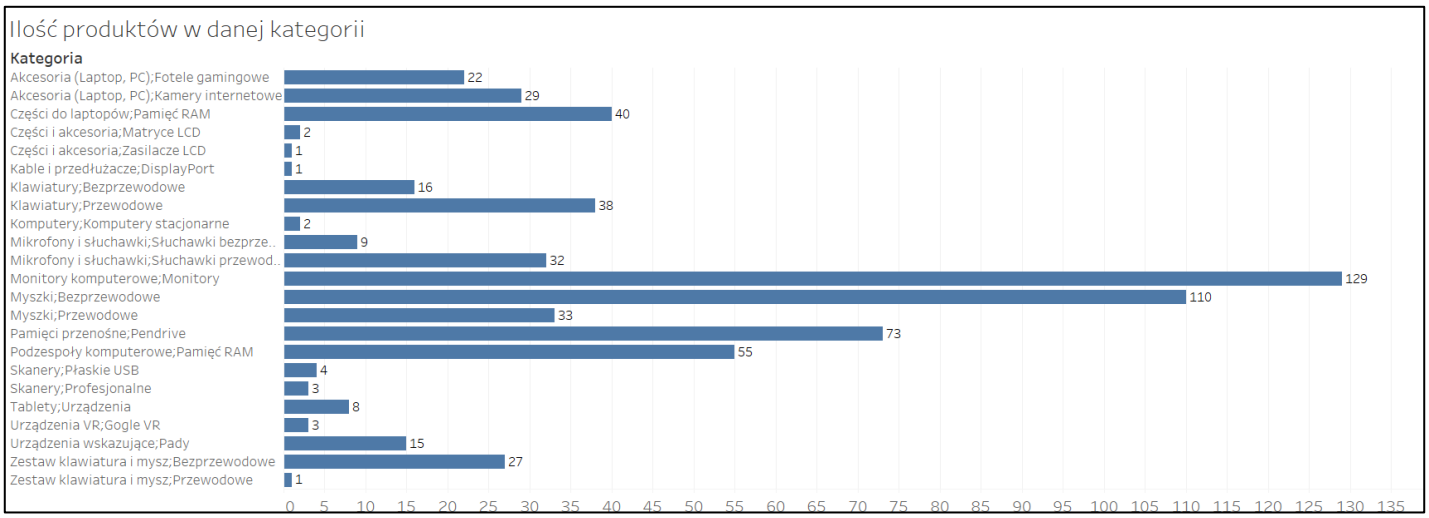
- Rozkład kolorów wśród produktów:



Dla wielu produktów (z tych, które mają wartość Null) kolor nie ma znaczenia lecz znajdują się tam też takie, dla których kolor miałby sens i realny wpływ (tutaj prośba o więcej danych). Ponadto są przypadki, dla których kolor nie znajduje się w *optional_attributes*, a widnieje w nazwie – w takich sytuacja kolor zostanie wyciągnięty z nazwy produktu i przekazany do *optional_attributes* w celu ułatwienia analizy.

- Przydatną informacją poza oceną użytkowników, byłaby także liczba tych ocen – pozwalałoby to określić rzetelność oceny produktu. (prośba o udzielenie tych danych)
- Rekordy z polskimi literami zakodowanymi według konwencji “unicode escape” zostaną przekonwertowane na odpowiedniki literowe
- Ścieżka do produktu (kategorie) zostanie rozbita – obecna forma nie jest "przyjazna" do przetwarzania.

- [opcjonalnie] Rozszerzenie oferty sklepu, gdyż w niektórych kategoriach znajduje się tylko jeden produkt. (prośba o więcej danych - być może dodanie produktów):



Użytkownicy:

- Rekordy z polskimi literami zakodowanymi według konwencji “unicode escape” zostaną przekonwertowane na odpowiedniki literowe
- Na podstawie imion zostanie dokonany podział użytkowników według płci

Sesje:

- Dane będą służyć do trenowania i walidacji modelu - najważniejsze z punktu widzenia całego zadania.

Dostawy (opcjonalnie):

- Na podstawie dat zamówienia i doręczenia można określić średni czas dostawy danego produktu – może to być przydatny atrybut przy tworzeniu modelu o ile użytkownicy w sklepie internetowym mają wgląd do takiej statystyki.

4. Ograniczenia i wnioski

- model można wytrenować jedynie na podstawie dostarczonych danych klienta, testowanie może odbywać się jedynie w trybie offline
- dane są obarczone różnymi błędami – niektóre trzeba naprawić, inne usunąć
- użytkownicy sklepu nie są skłonni podawać szczegółowych danych o sobie, w związku z tym najwięcej potrzebnych danych dla rekomendacji należy wyciągnąć z historii ich poczynań w sklepie