

Uniwersytet im. Adama Mickiewicza w Poznaniu
Wydział Matematyki i Informatyki



Projekt bazy danych

Sklep internetowy

Piotr Chlebicki
456516

Prowadzący zajęcia:
prof. UAM dr hab. Marka Wiśła

Poznań, czerwiec 2023

1 Opis bazy danych

Celem bazy danych jest obsługa danych tworzonych przez użytkowników sklepu internetowego oraz przez sam sklep a także pomaga oceniać pewne statystyczne aspekty prowadzenia takiego sklepu.

Przyjęte założenia:

- Strona rejestruje tylko wizyty zalogowanych użytkowników.
- Każdego dnia o północy następuje krótki reset serwera w celach konserwacyjnych. Niemożliwe jest zalogowanie się na stronę jednego dnia i wylogowanie następnego.
- Nazwa użytkownika jest unikalna i nie zawiera znaków specjalnych.
- Strona rejestruje czas zalogowania/wylogowania użytkownika oraz ilość wyświetlonych ofert w trakcie danej wizyty.

Baza danych składa się z pięciu tabel:

- **users** – będąca rejestrem użytkowników,
- **sales** – będąca rejestrem zamówionych przedmiotów,
- **products** – będąca spisem oferowanych przedmiotów,
- **predictions** – służąca do zapisywania wyników prognoz dotyczących długości pobytu danego użytkownika na stronie,
- **visits** (tabela dynamiczna) – zawiera informacje o wizytach użytkowników na stronie.

Między tabelami występują następujące zależności:

- Każda prognoza powiązana jest z wizytą dla której czas jest prognozowany za pomocą pola **id_visit** obecnego w tabelach **visits** i **predictions**,
- Każda wizyta powiązana jest z dokładnie jednym użytkownikiem, pole **user_name** z tabeli **users** i **visits**,
- Każda sprzedaż produktu z tabeli **sales** powiązana jest z użytkownikiem, przy pomocy pola **user_name** z tabeli **users** i **sales**, oraz ze sprzedawanym produktem, pole **product_name** z tabeli **sales** i **products**.

Na podstawie powyższej struktury oraz wprowadzonych danych utworzono następujące raporty:

- **week_day** – wyświetla dni tygodnia od największej do najmniejszej liczby unikalnych wizyt:

Weekdays from highest to lowest ammount of unique visits:

day_name	unique_visits
-----	-----
Saturday	835
Friday	781
Wednesday	770
Sunday	751
Thursday	708
Monday	693
Tuesday	663

- **best_selling_items** – wymienia najczęściej kupowane przedmioty:

Items from most to least frequently bought:

product_name	sales
-----	-----
Soap	124
Computer keyboard	114
Fidget spinner	99
Tortilla	62
Frying pan	45
Crips	32

- **frequent_visits** (prodecura) – wyświetla najczęściej wizytujących użytkowników w wyspecyfikowanym czasie:

Most frequent visitors from 15th of may to today 2023:

user	visits
-----	-----
Marisa Hofstetter	97
Mirella Durtschi	94
Lotta Hautzig	94
Photios Mastrotis	85
Nolan Schawalder	82
Lara Duque	82
Eleftheria Condeli	81
Torben Johannsen	81
Liam Krauser	80
Nela Itschner	80

Xene Franga	79
Soren Jensen	79
Borja Cutrina	78
Grethe Due	78
Albino Hernandez	77
Christian Chistau	76
Cristiana Tavares	76
Yanni Apostolilis	76
Lidia Almirall	75
Sam Kahler	73
Thorvald Lauritzen	72
Gustav Messerschmidt	72
Fernando Valdovinos	71
Frobertus	71
Laila Schweizer	71
Elisabeth Wulff	70
Tassilo	70
Hannelore Ungar	69
Giannis Panilis	68
Wilberga	65

- **buy_orders** (procedura) – wyświetla ilość kupionych przedmiotów przez użytkownika oraz wydana ilość pieniędzy w trakcie podanej wizyty:

```

items bought money spent
-----
2                600,00

```

- **regression_metrics** (procedura) – wylicza najczęściej stosowane metryki oceniające jakość predykcji (dla konkretnego algorytmu uczenia):

```

bias MSE      RMSE MAE  conf int coverage
-----
1665 9060010 3010 2334 0.12

```

(tutaj wyniki zaokrąglono a nazwy kolumn zostały skrócone).

2 Opis tabel

Kolumna	Opis	Typ	Pusta Wartość	Standardowa wartość	Powiązania
id_visit	Standardowy kod wizyty	INT	NIE	Automatycznie zwiększany licznik	Primary Key
user_name	Nazwa użytkownika	VARCHAR(20)	NIE		Foreign Key tabeli users
date_enter	Data zalogowania	DATETIME	NIE	GETDATE()	
date_exit	Data wylogowania NULL gdy pobyt nadal trwa	DATETIME	TAK	NULL	
view_count	Ilość objerzanych ofert w trakcie wizyty	INT	NIE		

Tabela **visits**

Kolumna	Opis	Typ	Pusta Wartość	Standardowa wartość	Powiązania
user_name	Nazwa użytkownika	VARCHAR(20)	NIE		Primary Key
gender	Płeć	VARCHAR(2)	NIE		
delivery_adress	Adres dostawy	VARCHAR(100)	NIE		
nationality	Skrót ALPHA-3 nazwy kraju pochodzenia użytkownika	VARCHAR(3)	NIE		
education	Poziom edukacji	VARCHAR(10)	NIE		

Tabela **users**

Kolumna	Opis	Typ	Pusta Wartość	Standardowa wartość	Powiązania
product_name	Nazwa produktu	VARCHAR(20)	NIE		Primary Key, Foreign Key tabeli products
country_of_origin	Pochodzenie zlecenia kupna kod ALPHA-3	VARCHAR(3)	TAK		
quantity_bought	Zamówiona ilość produktu	INT	NIE		
id_visit	Wizyta podczas której złożono zamówienie	INT	NIE		Foreign Key tabeli visits

Tabela **sales**

Kolumna	Opis	Typ	Pusta Wartość	Standardowa wartość	Powiązania
id_reg	Standardowy kod predykcji	INT	NIE	Automatycznie zwiększany licznik	Primary Key
alg_name	Skrócona nazwy algorytmu uczącego	VARCHAR(8)	NIE		
id_visit	Wizyta dla której prognozowany jest czas pobytu	INT	NIE		Foreign Key tabeli visits
predicted_time	Prognozowany czas przebywania na stronie	TIME	NIE		
low_pred	Dolna wartość przedziału ufności dla prognozy	TIME	NIE		
high_pred	Górna wartość przedziału ufności dla prognozy	TIME	NIE		
var_num	Ilość wykorzystanych zmiennych objaśniających	SMALLINT	TAK		

Tabela predictions

Kolumna	Opis	Typ	Pusta Wartość	Standardowa wartość	Powiązania
product_name	Nazwa produktu	VARCHAR(20)	NIE		Primary Key
price	Cena w PLN	MONEY	NIE		
category	Kod kategorii produktu	VARCHAR(4)	TAK		
stored_quantity	Sztuki produktu w magazynach	INT	NIE		
description	Opcjonalny krótki opis produktu	VARCHAR(400)	TAK		

Tabela products

3 Raporty, opis widoków oraz procedur

Dni tygodnia od największej do najmniejszej liczby unikalnych wizyt

Do tego raportu utworzono widok `week_day`. Raport ten służy do podsumowania, w który z dni tygodnia jest najwięcej unikalnych wizyt co jest istotne jeżeli zastanawiamy się nad ewaluacją lub ewentualną zmianą serwera internetowego dla sklepu (czy nie potrzeba aby większej przepustowości). Jest to także istotne z punktu widzenia biznesowego (planowanie różnych zdarzeń/przecen i tym podobnych).

Najczęściej kupowane przedmioty

Do tego raportu utworzono widok `best_selling_items`. Raport ten służy do znalezienia które z przedmiotów oferowanych w sklepie są najczęściej zamawiane przez użytkowników. Takie informacje w sposób istotny służą w informowaniu decyzji biznesowych na przykład o tym:

- jakich przedmiotów należy zamówić lub wyprodukować więcej,
- które przedmioty kosztują za mało/za dużo w relacji do powodzenia u użytkowników,
- jakim typem przedmiotu mogą być zainteresowani użytkownicy.

Najczęściej wizytujący użytkownicy

Do tego raportu utworzono procedurę `frequent_visits` zwracającą listę użytkowników posortowanych względem częstości wizyt na stronie internetowej. Procedura przyjmuje 2 parametry `@begin` i `@end`, oba typu `DATE`, które określają w jakim przedziale czasowym zliczane powinny być wizyty, domyślną wartością dla `@end` jest `GETDATE()`. Z punktu widzenia biznesu taka informacja jest ważna, ponieważ pomaga na przykład w tworzeniu spersonalizowanych rekomendacji dla odpowiednich użytkowników czy tworzeniu spersonalizowanych ofert typu indywidualne zniżki.

Kupione przedmioty przez użytkownika oraz wydana ilość pieniędzy w trakcie wizyty

Do tego raportu utworzono procedurę `buy_orders` zwracającą ilość kupionych przedmiotów i wydanych pieniędzy w trakcie podanej wizyty (parametr, typu `INT`) co może być przydatne dla samego użytkownika strony oraz w celach opracowań statystycznych takich jak na przykład testowanie zależności pomiędzy wydaną kwotą podczas danej wizyty a dniem tygodnia.

Metryki oceniające jakość regresji

Do tego raportu utworzono procedurę `regression_metrics` wyliczającą najpopularniejsze wielkości używane w ocenie algorytmów prognozujących dla danego algorytmu uczącego (parametr). Informacja ta jest istotna dla ponieważ znacznie ułatwia dostęp do istotnych informacji dla osób zaangażowanych w opracowanie tych metod, które same z siebie są istotne z perspektywy biznesowej (aktywność użytkowników wpływa na ich decyzje).

4 Diagram ERD

