

SQL dla analityków danych : opanuj możliwości SQL-a, aby wydobywać informacje z danych / Jun Shan, Matt Goldwasser, Upom Malik, Benjamin Johnston. – Gliwice, copyright © 2023

Spis treści

Wprowadzenie	11
ROZDZIAŁ 1	
Poznanie i opisywanie danych	37
Wprowadzenie	37
Analityka danych i statystyka	38
Zadanie 1.01 — klasyfikowanie nowego zbioru danych	39
Rodzaje statystyki	40
Metody z obszaru statystyki opisowej	41
Analiza jednoczynnikowa	41
Ćwiczenie 1.01 — tworzenie histogramu	42
Ćwiczenie 1.02 — obliczanie kwartyli dla sprzedaży dodatków	50
Ćwiczenie 1.03 — obliczanie miar tendencji centralnej dla sprzedaży dodatków	54
Ćwiczenie 1.04 — obliczanie dyspersji dla sprzedaży dodatków	57
Analiza dwuczynnikowa	58
Ćwiczenie 1.05 — obliczanie współczynnika korelacji Pearsona dla dwóch zmiennych	64
Interpretowanie i analizowanie współczynnika korelacji	66
Zadanie 1.02 — eksplorowanie danych sprzedażowych z salonu samochodowego	70
Praca z niepełnymi danymi	70
Testy istotności statystycznej	71
Często używane testy istotności statystycznej	72
SQL i analityka	73
Podsumowanie	73
ROZDZIAŁ 2	
Wprowadzenie do SQL-a dla analityków	75
Wprowadzenie	75
Świat danych	76
Rodzaje danych	77
Relacyjne bazy danych i SQL	78
Wady i zalety baz SQL-owych	79
System zarządzania relacyjnymi bazami danych PostgreSQL	81
Ćwiczenie 2.01 — uruchamianie pierwszej kwerendy SELECT	83
Instrukcja SELECT	86
Klauzula WHERE	90
Klauzule AND i OR	91

Klauzule IN i NOT IN	93
Klauzula ORDER BY	94
Klauzula LIMIT	97
Klauzule IS NULL i IS NOT NULL	98
Ćwiczenie 2.02 — kwerenda SELECT z podstawowymi słowami kluczowymi dotycząca tabeli salespeople	99
Zadanie 2.01 — kwerenda SELECT z podstawowymi słowami kluczowymi dotycząca tabeli customers	102
Tworzenie tabel	103
Tworzenie pustych tabel	103
Podstawowe typy danych w SQL-u	104
Typy liczbowe	104
Typy znakowe	104
Typ logiczny	105
Daty i godziny	106
Struktury danych — format JSON i tablice	106
Ograniczenia kolumn	107
Prosta instrukcja CREATE	107
Ćwiczenie 2.03 — tworzenie tabeli w SQL-u	108
Tworzenie tabel za pomocą kwerendy SELECT	109
Aktualizowanie tabel	111
Dodawanie i usuwanie kolumn	111
Dodawanie nowych danych	112
Aktualizowanie istniejących wierszy	114
Ćwiczenie 2.04 — aktualizowanie tabeli w celu podniesienia ceny pojazdu	115
Usuwanie danych i tabel	116
Usuwanie wartości z wiersza	116
Usuwanie wierszy z tabeli	117
Usuwanie tabel	117
Ćwiczenie 2.05 — usuwanie niepotrzebnej tabeli pomocniczej	119
Zadanie 2.02 — tworzenie i modyfikowanie tabel na potrzeby działań marketingowych	120
SQL i analityka	121
Podsumowanie	121
ROZDZIAŁ 3	
Przygotowywanie danych za pomocą SQL-a	123
Wprowadzenie	123
Łączenie danych	124
Łączenie tabel za pomocą słowa kluczowego JOIN	124
Rodzaje złączeń	127
Ćwiczenie 3.01 — używanie złączeń do analizy sprzedaży w salonach	137
Podkwerendy	139
Sumy	140
Ćwiczenie 3.02 — generowanie listy gości na przyjęcie dla klientów VIP za pomocą klauzuli UNION	142

Wyrażenia WITH	144
Oczyszczanie danych	145
Funkcja CASE WHEN	145
Ćwiczenie 3.03 — używanie funkcji CASE WHEN do pobierania list klientów z danego regionu	147
Funkcja COALESCE	148
Funkcja NULLIF	149
Funkcje LEAST i GREATEST	151
Funkcja CASTING	152
Przekształcanie danych	153
Funkcje DISTINCT i DISTINCT ON	153
Zadanie 3.01 — używanie SQL-a do tworzenia modelu wspomagającego sprzedaż	156
Podsumowanie	157

ROZDZIAŁ 4

Analiza danych z wykorzystaniem funkcji agregujących	159
Wprowadzenie	159
Funkcje agregujące	160
Ćwiczenie 4.01 — używanie funkcji agregujących do analizowania danych	165
Funkcje agregujące z klauzulą GROUP BY	166
Klauzula GROUP BY	167
Klauzula GROUP BY dla kilku kolumn	173
Ćwiczenie 4.02 — obliczanie cen dla typów produktów za pomocą klauzuli GROUP BY	174
Klauzula GROUPING SETS	175
Funkcje agregujące dla zbiorów uporządkowanych	177
Funkcje agregujące z klauzulą HAVING	178
Ćwiczenie 4.03 — obliczanie wyników i wyświetlanie danych z użyciem klauzuli HAVING	180
Stosowanie funkcji agregujących do oczyszczania danych i sprawdzania ich jakości	181
Znajdowanie brakujących wartości za pomocą klauzuli GROUP BY	181
Sprawdzanie unikatowości danych za pomocą funkcji agregujących	184
Zadanie 4.01 — analizowanie danych sprzedażowych z użyciem funkcji agregujących	185
Podsumowanie	186

ROZDZIAŁ 5

Analizowanie danych za pomocą funkcji okna	188
Wprowadzenie	188
Funkcje okna	189
Podstawy funkcji okna	191
Ćwiczenie 5.01 — analizowanie zmian współczynnika podawania danych przez klientów w czasie	197
Słowo kluczowe WINDOW	200

Obliczanie statystyk z użyciem funkcji okna	202
Ćwiczenie 5.02 — określanie pozycji na podstawie daty rejestracji	203
Ramka okna	204
Ćwiczenie 5.03 — motywowanie pracowników lunchem	207
Zadanie 5.01 — analizowanie sprzedaży z wykorzystaniem ramek okna i funkcji okna	209
Podsumowanie	211

ROZDZIAŁ 6

Importowanie i eksportowanie danych	212
Wprowadzenie	212
Polecenie COPY	213
Uruchamianie polecenia psql	214
Instrukcja COPY	215
Instrukcja \COPY w narzędziu psql	218
Tworzenie tymczasowych widoków	219
Konfigurowanie poleceń COPY i \COPY	221
Użycie poleceń COPY i \COPY do masowego wczytywania danych do bazy	222
Ćwiczenie 6.01 — eksportowanie danych do pliku w celu dalszego przetwarzania ich w Excelu	223
Zastosowanie języka Python do bazy danych	228
Wprowadzenie do języka Python	228
Ułatwianie dostępu do baz PostgreSQL w Pythonie za pomocą narzędzi SQLAlchemy i pandas	233
Czym jest SQLAlchemy?	233
Stosowanie Pythona z wykorzystaniem pakietów SQLAlchemy i pandas	235
Pobieranie danych z bazy i ich zapisywanie w bazie za pomocą pakietu pandas	237
Zapisywanie danych w bazie za pomocą Pythona	238
Ćwiczenie 6.02 — wczytywanie, wizualizowanie i zapisywanie danych za pomocą Pythona	238
Zwiększanie szybkości zapisu w Pythonie za pomocą polecenia COPY	243
Odczyt i zapis plików CSV w Pythonie	245
Najlepsze praktyki z obszaru importowania i eksportowania danych	247
Pomijanie podawania hasła	247
Zadanie 6.01 — używanie zewnętrznego zbioru danych do wykrywania trendów sprzedażowych	248
Podsumowanie	250

ROZDZIAŁ 7

Analizy z wykorzystaniem złożonych typów danych	251
Wprowadzenie	251
Wykorzystywanie typów danych z datami i czasem do analiz	252
Wprowadzenie do typu date	252
Przekształcanie typów danych	255
Przedziały	257

Ćwiczenie 7.01 — analiza danych z szeregów czasowych	259
Przeprowadzanie analiz geoprzestrzennych w PostgreSQL	261
Długość i szerokość geograficzna	262
Reprezentowanie długości i szerokości geograficznej w PostgreSQL	262
Ćwiczenie 7.02 — analizy geoprzestrzenne	265
Stosowanie tablicowych typów danych w PostgreSQL	268
Wprowadzenie do tablic	268
Ćwiczenie 7.03 — analizowanie sekwencji z użyciem tablic	271
Stosowanie formatu JSON w PostgreSQL	274
JSONB — wstępnie przetworzone dane w formacie JSON	211
Dostęp do danych z pól w formacie JSON lub JSONB	277
Stosowanie języka JSONPath do pól w formacie JSONB	280
Tworzenie i modyfikowanie danych w polu w formacie JSONB	282
Ćwiczenie 7.04 — przeszukiwanie obiektów JSONB	283
Analiza tekstu za pomocą PostgreSQL	286
Tokenizacja tekstu	286
Ćwiczenie 7.05 — analizowanie tekstu	288
Wyszukiwanie tekstu	293
Optymalizowanie wyszukiwania tekstu w PostgreSQL	296
Zadanie 7.01 — wyszukiwanie i analiza transakcji sprzedaży	298
Podsumowanie	300

ROZDZIAŁ 8

Wydajny SQL 301

Wprowadzenie	301
Znaczenie wysoce wydajnego kodu w SQL-u	302
Metody skanowania baz danych	303
Plany wykonywania kwerend	304
Ćwiczenie 8.01 — interpretowanie działania planera kwerend	305
Zadanie 8.01 — plany wykonywania kwerendy	309
Skanowanie indeksu	310
Indeks w postaci B-drzewa	311
Ćwiczenie 8.02 — tworzenie indeksu na potrzeby skanowania	312
Zadanie 8.02 — stosowanie skanowania indeksu	317
Indeks z haszowaniem	318
Ćwiczenie 8.03 — tworzenie kilku indeksów z haszowaniem, aby zbadać ich wydajność	320
Zadanie 8.03 — stosowanie indeksów z haszowaniem	323
Skuteczne korzystanie z indeksów	324
Kończenie pracy kwerend	325
Ćwiczenie 8.04 — anulowanie długo działającej kwerendy	326
Funkcje i wyzwalacze	328
Definicje funkcji	328
Ćwiczenie 8.05 — tworzenie funkcji, które nie przyjmują argumentów	329
Zadanie 8.04 — definiowanie funkcji zwracającej maksymalną wartość sprzedaży	332
Ćwiczenie 8.06 — tworzenie funkcji przyjmujących argumenty	333

Zadanie 8.05 — tworzenie funkcji przyjmujących argumenty	335
Wyzwalacze	336
Ćwiczenie 8.07 — tworzenie wyzwalaczy do aktualizowania pól	339
Zadanie 8.06 — tworzenie wyzwalacza do śledzenia średniej liczby kupionych sztuk	344
Podsumowanie	345

ROZDZIAŁ 9

Odkrywanie prawdy za pomocą SQL-a — studium przypadku 346

Wprowadzenie	346
Studium przypadku	347
Metoda naukowa	347
Ćwiczenie 9.01 — wstępne zbieranie danych za pomocą technik SQL-a	348
Ćwiczenie 9.02 — pobieranie informacji sprzedażowych	351
Zadanie 9.01 — ilościowa ocena spadku sprzedaży	356
Ćwiczenie 9.03 — analiza czasu rozpoczęcia sprzedaży	358
Zadanie 9.02 — analiza hipotezy dotyczącej różnicy w cenie sprzedaży	367
Ćwiczenie 9.04 — analiza zależności wzrostu sprzedaży od współczynnika otwarć e-maili	368
Ćwiczenie 9.05 — analiza skuteczności e-mailowej kampanii marketingowej	377
Wnioski	382
Badania terenowe	382
Podsumowanie	383

Dodatek 385

Skorowidz 429