

Spis treści

Wprowadzenie	9
Rozdział 1. Wprowadzenie do SQL-a dla analityków	33
Wprowadzenie	33
Świat danych	34
Rodzaje danych	34
Analityka danych i statystyka	35
Rodzaje statystyki	36
Zadanie 1.01 — klasyfikowanie nowego zbioru danych	37
Metody z obszaru statystyki opisowej	37
Rozkład danych	38
Ćwiczenie 1.01 — tworzenie histogramu	38
Ćwiczenie 1.02 — obliczanie kwartyli dla sprzedaży dodatków	43
Tendencja centralna	46
Ćwiczenie 1.03 — obliczanie miar tendencji centralnej dla sprzedaży dodatków	47
Dyspersja	48
Ćwiczenie 1.04 — obliczanie dyspersji dla sprzedaży dodatków	49
Analiza dwuczynnikowa	50
Wykresy punktowe	50
Ćwiczenie 1.05 — obliczanie współczynnika korelacji Pearsona dla dwóch zmiennych	56
Zadanie 1.02 — eksplorowanie danych sprzedażowych z salonu samochodowego	61
Praca z niepełnymi danymi	61
Testy istotności statystycznej	62
Często używane testy istotności statystycznej	63
Relacyjne bazy danych i SQL	64
Wady i zalety baz SQL-owych	64
Podstawowe typy danych w SQL-u	66
Typy liczbowe	66
Typy znakowe	66
Typ logiczny	67
Daty i godziny	67
Struktury danych — format JSON i tablice	68

Wczytywanie tabel — kwerenda SELECT	68
Podstawowa budowa i działanie kwerendy SELECT	68
Podstawowe słowa kluczowe w kwerendach SELECT	69
Ćwiczenie 1.06 — kwerenda SELECT z podstawowymi słowami kluczowymi dotycząca tabeli salespeople	76
Zadanie 1.03 — kwerenda SELECT z podstawowymi słowami kluczowymi dotycząca tabeli customers	77
Tworzenie tabel	78
Tworzenie pustych tabel	78
Ćwiczenie 1.07 — tworzenie tabeli w SQL-u	79
Tworzenie tabel za pomocą kwerendy SELECT	80
Aktualizowanie tabel	81
Dodawanie i usuwanie kolumn	81
Dodawanie nowych danych	81
Aktualizowanie istniejących wierszy	83
Ćwiczenie 1.08 — aktualizowanie tabeli w celu podniesienia ceny pojazdu	84
Usuwanie danych i tabel	85
Usuwanie wartości z wiersza	85
Usuwanie wierszy z tabeli	86
Usuwanie tabel	86
Ćwiczenie 1.09 — usuwanie niepotrzebnej tabeli	87
Zadanie 1.04 — tworzenie i modyfikowanie tabel na potrzeby działań marketingowych	87
SQL i analityka	88
Podsumowanie	89
Rozdział 2. Przygotowywanie danych za pomocą SQL-a	90
Wprowadzenie	90
Łączenie danych	91
Łączenie tabel za pomocą słowa kluczowego JOIN	91
Rodzaje złączeń	93
Ćwiczenie 2.01 — używanie złączeń do analizy sprzedaży w salonach	102
Podkwerendy	103
Sumy	104
Ćwiczenie 2.02 — generowanie listy gości na przyjęcie dla klientów VIP za pomocą klauzuli UNION	105
Wyrażenia WITH	107
Przekształcanie danych	108
Funkcja CASE WHEN	108
Ćwiczenie 2.03 — używanie funkcji CASE WHEN do pobierania list klientów z danego regionu	109
Funkcja COALESCE	111
Funkcja NULLIF	112
Funkcje LEAST i GREATEST	113
Funkcja CASTING	114
Funkcje DISTINCT i DISTINCT ON	115
Zadanie 2.01 — używanie SQL-a do tworzenia modelu wspomagającego sprzedaż	118
Podsumowanie	119

Rozdział 3. Agregacja i funkcje okna	120
Wprowadzenie	120
Funkcje agregujące	120
Ćwiczenie 3.01 — używanie funkcji agregujących do analizowania danych	123
Funkcje agregujące z klauzulą GROUP BY	124
Klauzula GROUP BY	124
Klauzula GROUP BY dla kilku kolumn	129
Ćwiczenie 3.02 — obliczanie cen dla typów produktów za pomocą klauzuli GROUP BY	130
Klauzula GROUPING SETS	131
Funkcje agregujące dla zbiorów uporządkowanych	132
Klauzula HAVING	133
Ćwiczenie 3.03 — obliczanie wyników i wyświetlanie danych z użyciem klauzuli HAVING	134
Stosowanie funkcji agregujących do oczyszczania danych i sprawdzania ich jakości	135
Znajdowanie brakujących wartości za pomocą klauzuli GROUP BY	135
Pomiar jakości danych za pomocą funkcji agregujących	137
Zadanie 3.01 — analizowanie danych sprzedażowych z użyciem funkcji agregujących	138
Funkcje okna	139
Podstawy funkcji okna	140
Ćwiczenie 3.04 — analizowanie zmian współczynnika podawania danych przez klientów w czasie	144
Słowo kluczowe WINDOW	146
Obliczanie statystyk z użyciem funkcji okna	147
Ćwiczenie 3.05 — określanie pozycji na podstawie daty zatrudnienia	148
Ramka okna	149
Ćwiczenie 3.06 — motywowanie pracowników lunchem	151
Zadanie 3.02 — analizowanie sprzedaży z wykorzystaniem ramek okna i funkcji okna	153
Podsumowanie	154
Rozdział 4. Importowanie i eksportowanie danych	155
Wprowadzenie	155
Polecenie COPY	156
Kopiowanie danych za pomocą narzędzia psql	157
Konfigurowanie poleceń COPY i \copy	159
Użycie poleceń COPY i \copy do masowego wczytywania danych do bazy	160
Ćwiczenie 4.01 — eksportowanie danych do pliku w celu dalszego przetwarzania ich w Excelu	161
Zastosowanie języka R do bazy danych	165
Po co korzystać z języka R?	165
Wprowadzenie do języka R	165
Zastosowanie języka Python do bazy danych	168
Po co korzystać z języka Python?	168
Wprowadzenie do języka Python	168
Ułatwianie dostępu do baz PostgreSQL w Pythonie za pomocą narzędzi SQLAlchemy i pandas	171
Czym jest SQLAlchemy?	172

Używanie Pythona w narzędziu Jupyter Notebook	172
Pobieranie danych z bazy i ich zapisywanie w bazie za pomocą pakietu pandas	174
Ćwiczenie 4.02 — wczytywanie i wizualizowanie danych w Pythonie	175
Zapisywanie danych w bazie za pomocą Pythona	177
Zwiększanie szybkości zapisu w Pythonie za pomocą polecenia COPY	178
Odczyt i zapis plików CSV w Pythonie	179
Najlepsze praktyki z obszaru importowania i eksportowania danych	181
Pomijanie podawania hasła	181
Zadanie 4.01 — używanie zewnętrznego zbioru danych do wykrywania trendów sprzedażowych	182
Podsumowanie	183
Rozdział 5. Analityka z wykorzystaniem złożonych typów danych	184
Wprowadzenie	184
Wykorzystywanie typów danych z datami i czasem do analiz	185
Wprowadzenie do typu date	185
Przekształcanie typów danych	188
Przedziały	189
Ćwiczenie 5.01 — analiza danych z szeregów czasowych	191
Przeprowadzanie analiz geoprzestrzennych w PostgreSQL	192
Długość i szerokość geograficzna	193
Reprezentowanie długości i szerokości geograficznej w PostgreSQL	193
Ćwiczenie 5.02 — analizy geoprzestrzenne	195
Stosowanie tablicowych typów danych w PostgreSQL	197
Wprowadzenie do tablic	197
Ćwiczenie 5.03 — analizowanie sekwencji z użyciem tablic	200
Stosowanie formatu JSON w PostgreSQL	201
JSONB — wstępnie przetworzone dane w formacie JSON	203
Dostęp do danych z pól w formacie JSON lub JSONB	204
Stosowanie języka JSONPath do pól w formacie JSONB	206
Tworzenie i modyfikowanie danych w polu w formacie JSONB	208
Ćwiczenie 5.04 — przeszukiwanie obiektów JSONB	209
Analiza tekstu za pomocą PostgreSQL	210
Tokenizacja tekstu	211
Ćwiczenie 5.05 — analizowanie tekstu	212
Wyszukiwanie tekstu	216
Optymalizowanie wyszukiwania tekstu w PostgreSQL	218
Zadanie 5.01 — wyszukiwanie i analiza transakcji sprzedaży	220
Podsumowanie	221
Rozdział 6. Wydajny SQL	222
Wprowadzenie	222
Metody skanowania baz danych	224
Plany wykonywania kwerend	224
Skanowanie sekwencyjne i inne metody skanowania	224
Ćwiczenie 6.01 — interpretowanie działania planera kwerend	226
Zadanie 6.01 — plany wykonywania kwerendy	229
Skanowanie indeksu	230
Indeks w postaci B-drzewa	231

Ćwiczenie 6.02 — kwerenda ze skanowaniem indeksu	233
Zadanie 6.02 — skanowanie indeksu	237
Indeks z haszowaniem	238
Ćwiczenie 6.03 — tworzenie kilku indeksów z haszowaniem, aby zbadać ich wydajność	239
Zadanie 6.03 — stosowanie indeksów z haszowaniem	243
Skuteczne korzystanie z indeksów	244
Wydajne złączenia	245
Ćwiczenie 6.04 — ocenianie zastosowania złączeń wewnętrznych	246
Zadanie 6.04 — stosowanie wydajnych złączeń	252
Funkcje i wyzwalacze	253
Definicje funkcji	253
Ćwiczenie 6.05 — tworzenie funkcji, które nie przyjmują argumentów	254
Zadanie 6.05 — definiowanie funkcji zwracającej maksymalną wartość sprzedaży	257
Ćwiczenie 6.06 — tworzenie funkcji przyjmujących argumenty	258
Polecenia \df i \sf	259
Zadanie 6.06 — tworzenie funkcji przyjmujących argumenty	260
Wyzwalacze	260
Ćwiczenie 6.07 — tworzenie wyzwalaczy do aktualizowania pól	263
Zadanie 6.07 — tworzenie wyzwalacza do śledzenia średniej liczby kupionych sztuk	267
Kończenie pracy kwerend	268
Ćwiczenie 6.08 — anulowanie długo działającej kwerendy	269
Zadanie 6.08 — kończenie długo działającej kwerendy	270
Podsumowanie	271
Rozdział 7. Metoda naukowa i rozwiązywanie problemów w praktyce	272
Wprowadzenie	272
Studium przypadku	273
Metoda naukowa	273
Ćwiczenie 7.01 — wstępne zbieranie danych za pomocą technik SQL-a	274
Ćwiczenie 7.02 — pobieranie informacji sprzedażowych	276
Zadanie 7.01 — ilościowa ocena spadku sprzedaży	280
Ćwiczenie 7.03 — analiza czasu rozpoczęcia sprzedaży	281
Zadanie 7.02 — analiza hipotezy dotyczącej różnicy w cenie sprzedaży	288
Ćwiczenie 7.04 — analiza zależności wzrostu sprzedaży od współczynnika otwarć e-maili	290
Ćwiczenie 7.05 — analiza skuteczności e-mailowej kampanii marketingowej	297
Wnioski	300
Badania terenowe	300
Podsumowanie	301
Dodatek	303
Skorowidz	345