# Dokumentacja: Projekt Modelu Analitycznego

1. Informacje o projekcie

- Nazwa projektu: Hurtownia Danych BikeStores\_Analytics

- Cel projektu: Stworzenie modelu analitycznego oraz procesów zasilania danych w celu analizy sprzedaży w BikeStores.

2. Zakres dokumentacji

- Model analityczny: Projekt schematu danych hurtowni.

- Proces ETL: Szczegóły dotyczące ekstrakcji, transformacji i ładowania danych.

- Technologie: Narzędzia i technologie użyte w projekcie.

- Optymalizacje: Metody optymalizacji bazy analitycznej.

3. Analiza bazy relacyjnej pobranej z pliku `BikeStores.bak`

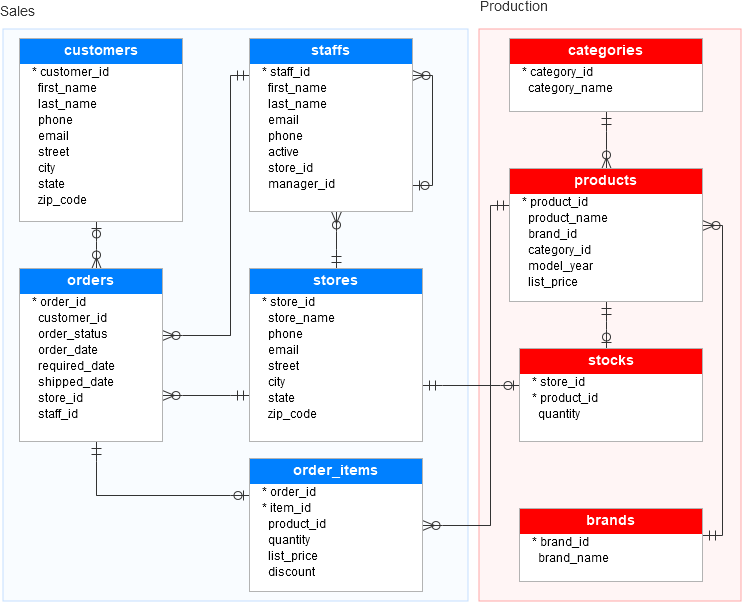
3.1. Informacje o bazie relacyjnej

- Plik ` BikeStores.bak`: Pełna kopia zapasowa bazy relacyjnej BikeStores, zaimportowana w celu zrozumienia struktury danych

- Cel analizy: Zbadanie struktury, zależności oraz danych w celu zaprojektowania hurtowni danych.

- Typ danych: Dane sprzedaży, produktów, klientów, zamówień oraz sklepów.

3.2. Struktura bazy relacyjnej



3.2.1. Tabele i ich opis

Tabela **sales.orders**

Opis: Przechowuje informacje o zamówieniach.

Klucz główny: order\_id

Kolumny kluczowe:

customer\_id: Identyfikator klienta.

store\_id: Identyfikator sklepu.

order\_date: Data zamówienia.

Tabela **sales.order\_items**

Opis: Szczegóły pozycji zamówień.

Klucz główny: order\_id, item\_id.

Kolumny kluczowe:

order\_id: Identyfikator zamówienia.

product\_id: Identyfikator produktu.

quantity: Ilość zamówiona.

list\_price: Cena jednostkowa.

discount: Rabat.

Tabela **production.products**

Opis: Dane o produktach.

Klucz główny: product\_id.

Kolumny kluczowe:

product\_name: Nazwa produktu.

brand\_id: Identyfikator marki.

category\_id: Identyfikator kategorii.

Tabela **production.categories**

Opis: Kategorie produktów.

Klucz główny: category\_id.

Kolumny kluczowe:

category\_name: Nazwa kategorii.

Tabela **sales.customers**

Opis: Dane klientów.

Klucz główny: customer\_id.

Kolumny kluczowe:

first\_name, last\_name: Imię i nazwisko klienta.

email: Adres e-mail.

phone: Numer telefonu.

Tabela sales.stores

Opis: Informacje o sklepach.

Klucz główny: store\_id.

Kolumny kluczowe:

store\_name: Nazwa sklepu.

city, state: Lokalizacja sklepu.

Tabela **production.brands**

Opis: Informacje o markach produktów.

Klucz główny: brand\_id.

Kolumny kluczowe:

brand\_name: Nazwa marki.

Tabela **production.stocks**

Opis: Przechowuje informacje o dostępnych zapasach produktów w sklepach.

Klucz główny: Brak (tabela łączy produkty i sklepy).

Kolumny kluczowe:

store\_id: Identyfikator sklepu.

product\_id: Identyfikator produktu.

quantity: Ilość dostępnych produktów w sklepie.

Tabela **sales.staffs**

Opis: Informacje o pracownikach.

Klucz główny: staff\_id.

Kolumny kluczowe:

first\_name, last\_name: Imię i nazwisko pracownika.

email: Adres e-mail pracownika.

phone: Numer telefonu pracownika.

store\_id: Sklep, w którym pracuje pracownik.

manager\_id: Identyfikator menedżera pracownika.

3.2.2. Relacje w bazie relacyjnej

- Klucze obce:

- `sales.orders.customer\_id → sales.customers.customer\_id`

- `sales.orders.store\_id → sales.stores.store\_id`

- `sales.staffs.store\_id → sales.stores.store\_id`

- `sales.order\_items.order\_id → sales.orders.order\_id`

- `sales.order\_items.product\_id → production.products.product\_id`

- `production.products.category\_id → production.categories.category\_id`

- `production.products.brand\_id → production.brands.brand\_id`

- `production.stocks.store\_id → sales.stores.store\_id`

- `production.stocks.product\_id → production.products.product\_id`

- `sales.staffs.manager\_id → sales.staffs.staff\_id`

4. Architektura i model analityczny



4.1. Struktura hurtowni danych

- Schemat gwiazdy:

Hurtownia oparta jest na modelu gwiazdy, składającym się z:

- Tabela faktów:

- `Fact\_Sales`: Przechowuje szczegóły transakcji, takie jak ilość, rabaty, wartość sprzedaży, oraz identyfikatory wymiarów (np. `product\_id`, `customer\_id`, `date\_id`).

- Tabele wymiarów:

- `Dim\_Products`: Dane o produktach (nazwa, kategoria, marka).

- `Dim\_Customers`: Dane klientów (imię, nazwisko, lokalizacja).

- `Dim\_Stores`: Informacje o sklepach (nazwa, lokalizacja).

- `Dim\_Staffs`: Dane o pracownikach

- `Dim\_Categories`: Dane o kategoriach produktów, identyfikatory i nazwy kategorii

- `Dim\_Time`: Dane kalendarzowe

4.2. Relacje między tabelami

Tabela faktów łączy się z tabelami wymiarów za pomocą kluczy obcych:

Fact\_Sales.product\_id → Dim\_Products.product\_id

Fact\_Sales.customer\_id → Dim\_Customers.customer\_id

Fact\_Sales.store\_id → Dim\_Stores.store\_id

Fact\_Sales.date\_id → Dim\_Time.date\_id

Fact\_Sales.category\_id → Dim\_Categories.category\_id  
Fact\_Sales.staff\_id → Dim\_Staffs.staff\_id

5. Proces ETL (Ekstract, Transform,Load)

5.1. Ekstrakcja danych

Źródła danych:

Relacyjna baza danych `BikeStores` została zaimportowana z pliku *BikeStores.bak*

5.2. Transformacja danych

Stworzono nową strukturę danych wg schematu gwiazdy (*star schema*).

proces jest zawarty w pliku *baza\_analityczna.sql*

W wyniku transformacji powstały tabele;

**Tabele wymiarów** (Dim\_Stores, Dim\_Products, Dim\_Customers, Dim\_Time, Dim\_Categories, Dim\_Staffs) zawierają atrybuty opisujące sklepy, produkty, klientów, czas i pracowników, z kluczami głównymi.

**Tabela faktów** (Fact\_Sales) przechowuje dane sprzedaży i łączy się z tabelami wymiarów za pomocą kluczy obcych.

Transformacja ogranicza się do zdefiniowania struktury tabel i ich relacji.

5.3. Ładowanie danych

Ładowanie do tabel faktów i wymiarów:

Tabele wymiarów są ładowane jako pierwsze, aby zagwarantować poprawne klucze obce.

Tabela faktów jest ładowana z użyciem kluczy obcych.

Proces ładowania jest zawarty w pliku *baza\_analityczna.sql*

6. Szczegóły modelu analitycznego

6.1. Schemat tabel

**Fact\_Sales**: Dane o transakcjach sprzedaży  
**Dim\_Products**: Dane o produktach

**Dim\_Customers**: Dane o klientach

**Dim\_Time**: Dane o czasie

**Dim\_Categories**: Dane o kategoriach produktów

**Dim\_Stores**: Dane o sklepach

**Dim\_Staffs**: Dane o pracownikach

6.2. Sprawdzenie poprawności procesu

W celu sprawdzenia poprawności procesu ETL stworzono 7 zapytań SQL

do bazy relacyjnej *(baza\_relacyjna.sql).* Te same zapytania SQL utworzono dla bazy analitycznej *(baza\_analityczna.sql)*

7. Optymalizacje

CREATE INDEX idx\_fact\_sales\_product\_id ON Fact\_Sales (product\_id);

CREATE INDEX idx\_fact\_sales\_date\_id ON Fact\_Sales (date\_id);

Indeksy zostały utworzone w celu przyspieszenia wyszukiwania i łączenia danych w tabeli Fact\_Sales na podstawie kluczowych kolumn product\_id oraz date\_id. Dzięki temu operacje analityczne, takie jak filtrowanie po produktach czy agregacje danych według dat, są bardziej wydajne.