

# Spis użytych technologii

sobota, 18 stycznia 2025 21:01

Do realizacji projektu zostały użyte programy:

- RStudio,
- VS Code,
- OneNote,
- Visual Studio 2022,
- Dbeaver.

A także języki programowania wraz z bibliotekami:

- Python:
  - sqlalchemy,
  - json,
  - random,
  - faker,
  - csv,
  - pandas,
- R:
  - DBI,
  - RMariaDB,
  - dplyr,
  - dbplyr,
  - ggplot2,
  - knitr,
  - lubridate,
  - RColorBrewer.

# Lista plików

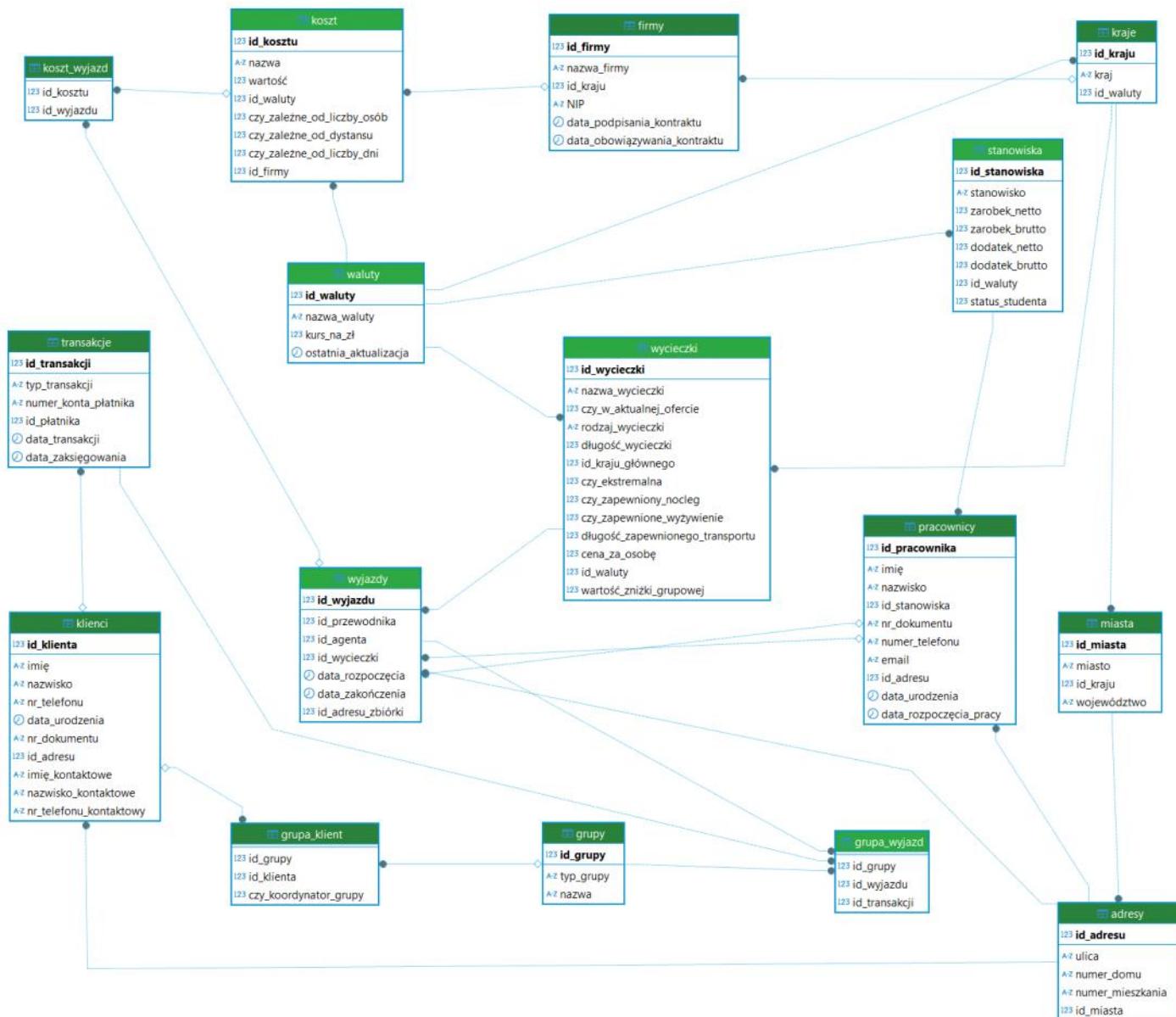
sobota, 18 stycznia 2025 21:01

Poniżej znajduje się lista plików użytych do realizacji projektu:

- `tworzenie.py`
  - Plik zawiera skrypt w języku Python generujący wszystkie puste tabele.
- `struktura.json`
  - Plik zawiera opisy kolumn napisane w SQL dostosowane do formatu JSON.
- `warunki_bazy.py`
  - W pliku znajduje się skrypt języka Python tworzący wyzwalacze dla tabel, dla których ograniczenie "CHECK" nie może zadziałać, a dla których wpisywanie lub edycja danych musi być ograniczona.
- `__init__.py`
  - Plik wspomagający czytanie funkcji z poszczególnych plików do pliku głównego tworzącego bazę.
- `czysta_baza.py`
  - W pliku znajduje się skrypt bazujący na innych plikach, który doprowadza bazę do pustego stanu gotowego na generację danych.
- `GENERACJA.py`
  - Plik zawierający skrypt generujący dane do poszczególnych tabel opierający się na poniższych plikach:
    - `aleja.txt`
    - `generał_marszałek.txt`
    - `imiona.csv`
    - `kraje_waluty.csv`
    - `miasta.txt`
    - `nazwa_wycieczki.txt`
    - `plac_skwer.txt`
    - `rodzaj_wycieczki.txt`
- `Analiza-grupa6.Rmd`
  - Plik dokonujący analizy danych zawartych w bazie oraz tworzący raport na tej podstawie.
- `Analiza-grupa6.html`
  - Plik zawierający raport z części analitycznej.

# Schemat projektu bazy danych

sobota, 18 stycznia 2025 20:51



# Lista zależności funkcyjnych

sobota, 18 stycznia 2025 18:27

Poniżej znajduje się lista zależności funkcyjnych dla poszczególnych tabel. Nazwy tabel (zmienne relacyjne) są wyszczególnione kolorem fioletowym. W każdym poszczególnym wierszu nienagłówkowym w tabelce znajdują się wypisane atrybuty, a także zbiory atrybutów, z którymi istnieją zależności funkcyjne (tzn. np. dla atrybutu id\_a i zbioru {pierwszy, drugi} istnieją zależności funkcyjne id\_a->pierwszy oraz id\_a->drugi). Wszystkie zależności funkcyjne zostały wypisane dla potencjalnej zawartości całej bazy.

- **pracownicy:**

Wszystkie={id\_pracownika, imię, nazwisko, id\_stanowiska, nr\_dokumentu, numer\_telefonu, email, id\_adresu, data\_urodzenia, data\_rozpoczęcia\_pracy}

Atrybuty	Zbiory atrybutów, z którymi są w zależności funkcyjnej	Wyjaśnienie
<b>id_pracownika</b>	wszystkie\{id_pracownika\}	atrybut id_pracownika jest kluczem głównym (jest unikatowym), zatem jego jednoelementowy zbiór jest nadkluczem
<b>nr_dokumentu</b>	wszystkie\{nr_dokumentu\}	atrybut nr_dokumentu jest unikatowy, zatem jego jednoelementowy zbiór jest nadkluczem
<b>numer_telefonu</b>	wszystkie\{numer_telefonu\}	atrybut numer_telefonu jest unikatowy, zatem jego jednoelementowy zbiór jest nadkluczem
<b>email</b>	wszystkie\{email\}	atrybut email jest unikatowy, zatem jego jednoelementowy zbiór jest nadkluczem

- **stanowiska:**

Wszystkie={id\_stanowiska, stanowisko, zarobek\_netto, zarobek\_brutto, id\_waluty, status\_studenta}

Atrybuty	Zbiory atrybutów, z którymi są w zależności funkcyjnej	Wyjaśnienie
<b>id_stanowiska</b>	wszystkie\{id_stanowiska\}	atrybut id_stanowiska jest kluczem głównym (jest unikatowym), zatem jego jednoelementowy zbiór jest nadkluczem

- **wycieczki:**

Wszystkie={id\_wycieczki, czy\_w\_aktualnej\_ofercie, rodzaj\_wycieczki, id\_kraju\_głównego, czy\_ekstremalna, czy\_zapewniony\_nocleg, czy\_zapewnione\_wyżywienie, długość\_zapewnionego\_transportu, cena\_za\_osobę, id\_waluty, wartość\_zniżki\_grupowej}

Atrybuty	Zbiory atrybutów, z którymi są w zależności funkcyjnej	Wyjaśnienie
<b>id_wycieczki</b>	wszystkie\{id_wycieczki\}	atrybut id_wycieczki jest kluczem głównym (jest unikatowym), zatem jego jednoelementowy zbiór jest nadkluczem

- **klienci:**

Wszystkie={id\_klienta, imię, nazwisko, nr\_telefonu, data\_urodzenia, nr\_dokumentu, id\_adresu, imię\_kontaktowe, nazwisko\_kontaktowe, nr\_telefonu\_kontaktowy}

Atrybuty	Zbiory atrybutów, z którymi są w zależności funkcyjnej	Wyjaśnienie
<b>id_klienta</b>	wszystkie\{id_klienta\}	atrybut id_klienta jest kluczem głównym (jest unikatowym), zatem jego jednoelementowy zbiór jest nadkluczem
<b>nr_dokumentu</b>	wszystkie\{nr_dokumentu\}	atrybut nr_dokumentu jest unikatowy, zatem jego jednoelementowy zbiór jest nadkluczem

- **grupy:**

Wszystkie={id\_grupy, typ\_grupy, nazwa}

Atrybuty	Zbiory atrybutów, z którymi są w zależności funkcyjnej	Wyjaśnienie
<b>id_grupy</b>	wszystkie\{id_grupy\}	atrybut id_grupy jest kluczem głównym (jest unikatowym), zatem jego jednoelementowy zbiór jest nadkluczem

- **grupa\_klient:**

Wszystkie={id\_grupy, id\_klienta, czy\_koordynator\_grupy}

Atrybuty	Zbiory atrybutów, z którymi są w zależności funkcyjnej	Wyjaśnienie
----------	--	-------------

Tabela nie posiada zależności funkcyjnych

- **grupa\_wyjazd:**

Wszystkie={id\_grupy, id\_wyjazdu, id\_transaksi}

Atrybuty	Zbiory atrybutów, z którymi są w zależności funkcyjnej	Wyjaśnienie
----------	--	-------------

Tabela nie posiada zależności funkcyjnych

- klienci:**

Wszystkie={id\_kraju, kraj}

Atrybuty	Zbiory atrybutów, z którymi są w zależności funkcyjnej	Wyjaśnienie
<b>id_kraju</b>	{kraj}	atrybut id_kraju jest kluczem głównym (jest unikatowym), zatem jego jednoelementowy zbiór jest nadkluczem
<b>krat</b>	{id_kraju}	atrybut krat jest unikatowy, zatem jego jednoelementowy zbiór jest nadkluczem

- miasta:**

Wszystkie={id\_miasta, miasto, id\_kraju, województwo}

Atrybuty	Zbiory atrybutów, z którymi są w zależności funkcyjnej	Wyjaśnienie
<b>id_miasta</b>	wszystkie\{id_miasta}	atrybut id_miasta jest kluczem głównym (jest unikatowym), zatem jego jednoelementowy zbiór jest nadkluczem

- adresy:**

Wszystkie={id\_adresu, ulica, numer\_domu, numer\_mieszkania, id\_miasta}

Atrybuty	Zbiory atrybutów, z którymi są w zależności funkcyjnej	Wyjaśnienie
<b>id_adresu</b>	wszystkie\{id_adresu}	atrybut id_adresu jest kluczem głównym (jest unikatowym), zatem jego jednoelementowy zbiór jest nadkluczem

- firmy:**

Wszystkie={id\_firmy, nazwa\_firmy, id\_kraju, NIP, data\_podpisania\_kontraktu, data\_obowiązywania\_kontraktu}

Atrybuty	Zbiory atrybutów, z którymi są w zależności funkcyjnej	Wyjaśnienie
<b>id_firmy</b>	wszystkie\{id_firmy}	atrybut id_firmy jest kluczem głównym (jest unikatowym), zatem jego jednoelementowy zbiór jest nadkluczem

- koszt:**

Wszystkie={id\_koszta, nazwa, wartość, id\_waluty, czy\_zależne\_od\_liczby\_osób, czy\_zależne\_od\_dystansu, czy\_zależne\_od\_liczby\_dni, id\_firmy}

Atrybuty	Zbiory atrybutów, z którymi są w zależności funkcyjnej	Wyjaśnienie
<b>id_koszta</b>	wszystkie\{id_koszta}	atrybut id_koszta jest kluczem głównym (jest unikatowym), zatem jego jednoelementowy zbiór jest nadkluczem

- koszt\_wyjazd:**

Wszystkie={id\_koszta, id\_wyjazdu}

Atrybuty	Zbiory atrybutów, z którymi są w zależności funkcyjnej	Wyjaśnienie
----------	--	-------------

Tabela nie posiada zależności funkcyjnych

- transakcje:**

Wszystkie={id\_transakcji, typ\_transakcji, numer\_konta\_płatnika, id\_płatnika, data\_transakcji, data\_zaksięgowania}

Atrybuty	Zbiory atrybutów, z którymi są w zależności funkcyjnej	Wyjaśnienie
<b>id_transakcji</b>	wszystkie\{id_transakcji}	atrybut id_transakcji jest kluczem głównym (jest unikatowym), zatem jego jednoelementowy zbiór jest nadkluczem

- wyjazdy:**

Wszystkie={id\_wyjazdu, id\_przewodnika, id\_agenta, id\_wycieczki, data\_rozpoczęcia, data\_zakończenia, id\_adresu\_zbiorki}

Atrybuty	Zbiory atrybutów, z którymi są w zależności funkcyjnej	Wyjaśnienie
<b>id_wyjazdu</b>	wszystkie\{id_wyjazdu}	atrybut id_wyjazdu jest kluczem głównym (jest unikatowym), zatem jego jednoelementowy zbiór jest nadkluczem

- waluty:**

Wszystkie={id\_waluty, nazwa\_waluty, kurs\_na\_zł, ostatnia\_aktualizacja}

Atrybuty	Zbiory atrybutów, z którymi są w zależności funkcyjnej	Wyjaśnienie
<b>id_waluty</b>	wszystkie\{id_waluty}	atrybut id_waluty jest kluczem głównym (jest unikatowym),

		zatem jego jednoelementowy zbiór jest nadkluczem
<b>nazwa_waluty</b>	wszystkie\{nazwa_waluty}	atrybut nazwa_waluty jest unikatowy, zatem jego jednoelementowy zbiór jest nadkluczem

Ponieważ wszystkie zależności w każdej tabeli zaczynają się od nadkluczy, to każda tabela jest EKNF, co oznacza, że cała baza również jest EKNF.

# Najtrudniejsze elementy projektu

sobota, 18 stycznia 2025 20:52

Naszym zdaniem najtrudniejszymi w realizacji projektu było:

- zaprojektowanie bazy tak, aby spełniała wszystkie wymagania (rozległość bazy utrudniała znormalizowanie jej, a także wydłużała czas testów),
- stworzenie programu generującego dane, tak aby zachował on przy tym spójność między tabelami, np. żeby ten sam wyjazd odbywał się na terenie tego samego kraju, w jakim obowiązuje dana waluta,
- podczas analizy dopasowanie rozmiaru wykresu do danych oraz wygenerowanie w pełni zautomatyzowanych odpowiedzi.