Projekt zaliczeniowy

Proces ETL

Grupa projektowa: 26

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Imię | Nazwisko | Numer albumu | Grupa dziekańska | Wkład w prace nad projektem[[1]](#footnote-1) | Udział procentowy |
| Marek | Kopeć | 186912 | 1111 | Pisanie kodu aplikacji, modyfikowanie kodu aplikacji, tworzenie bazy danych, stworzenie dokumentacji technicznej oraz instrukcji instalacji i obsługi aplikacji | 33,3% |
| Mateusz | Sajler | 194599 | 1111 | Pisanie kodu aplikacji, modyfikowanie kodu aplikacji, tworzenie bazy danych, stworzenie dokumentacji technicznej oraz instrukcji instalacji i obsługi aplikacji | 33,3% |
| Arek | Hadała | 192980 | 1111 | Pisanie kodu aplikacji, modyfikowanie kodu aplikacji, tworzenie bazy danych, stworzenie dokumentacji technicznej oraz instrukcji instalacji i obsługi aplikacji | 33,3% |

**Dokumentacja Techniczna:**

1.Aplikacja została napisana w technologii programowania  języka R z wykorzystaniem takich bibliotek jak : rvest, dplyr, stringr, odbc, dbi. System zarządzania baza danych został stworzony w technologii MySQL.

2.Aby aplikacja poprawnie działała zalecamy uruchomienie aplikacji na sprzęcie o minimalnych wymaganiach sprzętowych jak niżej podane:

System : Windows Vista/7/8/10

Procesor : 2 core procesor

Ram : 4GB

Karta graficzna : zintegrowana

Dysk: 50GB

Aby aplikacja poprawnie działała musimy również zainstalować dodatkowe biblioteki takie jak:

-library(rvest)

-library(dplyr)

-library(stringr)

-library(odbc)

-library(dbi)

-library(rstudioapi)

Należy użyć polecenia install.packages()

3.Linki do oprogramowania które tworzą to środowisko.

<https://www.rstudio.com/>

<https://www.mysql.com/>

4.Instrukcja instalacji aplikacji:

Instalacja aplikacji jak i cały proces działania odbywa się w programie RStudio, gdzie instalowane są wszystkie biblioteki potrzebne do kompilacji i uruchomienia programu oraz wywoływany jest sam kod programu.

5.Model danych użyty w projekcie.

Strukturalnie:

|  |  |  |
| --- | --- | --- |
| nazwa zmiennej | opis zmiennej | zwracana wartość/przyjmowane parametry |
| oferty\_df | Data frame zawierający wartości pobierane ze scrappowanej strony internetowej. |  |
| choices | Zmienna służąca do komunikacji z użytkownikiem. | Zwracana wartość to województwo dla którego przeprowadzone będzie wyszukiwanie. |
| otodom | Zmienna przechowująca link do strony z ogłoszeniami | Link do strony jest “hard coded”, poszerzany jest w pętli o ilość podstron |
| page | Zmienna page przechowuje link do strony, dodatkowo poszerzone o numer podstrony |  |
| nazwy | zmienna przechowująca nazwy ogłoszeń |  |
| metraż | zmienna przechowująca metraż mieszkania |  |
| ceny | zmienna przechowująca ceny mieszkania |  |
| link | zmienna przechowująca link do podstrony |  |
| piętro | zmienna przechowująca piętro na którym znajduje się mieszkanie |  |
| liczba\_pokoi | zmienna przechowująca ilość pokoi znajdujących się w mieszkaniu |  |
| dodatkowo | zmienna pomocnicza używana do procesu transformacji |  |
| parametry | zmienna wykorzystywana do pobrania reszty informacji o mieszkaniu |  |
| rok\_budowy | zmienna przechowująca rok w którym zakończono budowę mieszkania |  |
| czynsz\_dodatkowy | zmienna przechowująca dodatkowy czynsz |  |
| kaucja | zmienna przechowująca dodatkowy czynsz |  |
| rodzaj\_zabudowy | zmienna przechowująca dodatkowy czynsz |  |
| parametr | pomocnicza zmienna zastosowana podczas transformacji danych pobieranych ze strony |  |
| conn | zmienna zawierająca ustawienia oraz dane logowania i połączenia z bazą dancyh | parametry: driver - baza danych, “server” - nazwa serwera, “Database” - nazwa baz danych, “UID” - username, “PWD” - password, port - wykorzystywany port do połączenia |
| write.csv | metoda dzięki której pobrane dane przesyłane są do pliku CSV | parametry: nazw pliku, rozszerzenie |

**Instrukcja obsługi aplikacji:**

1.Instrukcja uruchomienia aplikacji

Aplikacja uruchamiana jest poprzez RStudio. Aby aplikacja poprawnie działała musimy otworzyć RStudio, zainstalować wszystkie wymagane biblioteki oraz uruchomić kod programu.

2.Opis funkcjonalności aplikacji

Podstawową funkcjonalnością aplikacji jest pobieranie danych ze strony otodom dla wszystkich ogłoszeń znajdujących się na stronie. Dzięki aplikacji możemy również transformować dane (usuwać niepotrzebne znaki, przekształcać do odpowiedniej formy).  
Dzięki aplikacji jesteśmy w stanie połączyć się z uprzednio skonfigurowaną bazą danych. Z poziomu aplikacji nawiązane jest połączenie dzięki zainstalowanej bibliotece odbc. Aplikacja generuje plik csv co umożliwia szybszą analizę danych jeżeli jest taka konieczność.

3.Opis scenariuszy użycia aplikacji

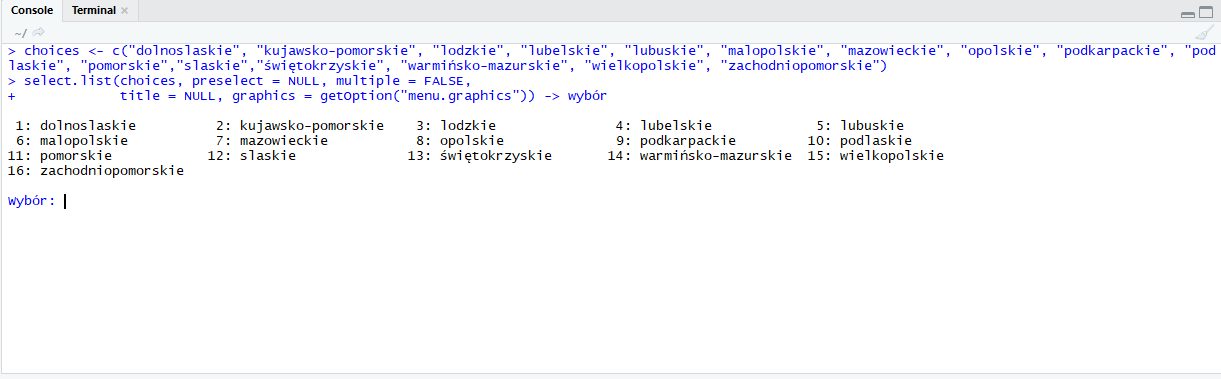
W zależności od przyjętych parametrów aplikacja może zwracać różne wartości. Liczba możliwych scenariuszy działania aplikacji jest równa liczbie możliwych parametrów czyli w naszym wypadku liczby województw. Dzięki aplikacji możemy wybrać z którego województwa oferty mieszkań będa pobierane do naszej bazy danych. Mamy do wyboru 16 województw co prowadzi do 16 możliwych, różnych scenariuszy.

4.Opis menu i widoków okna aplikacji

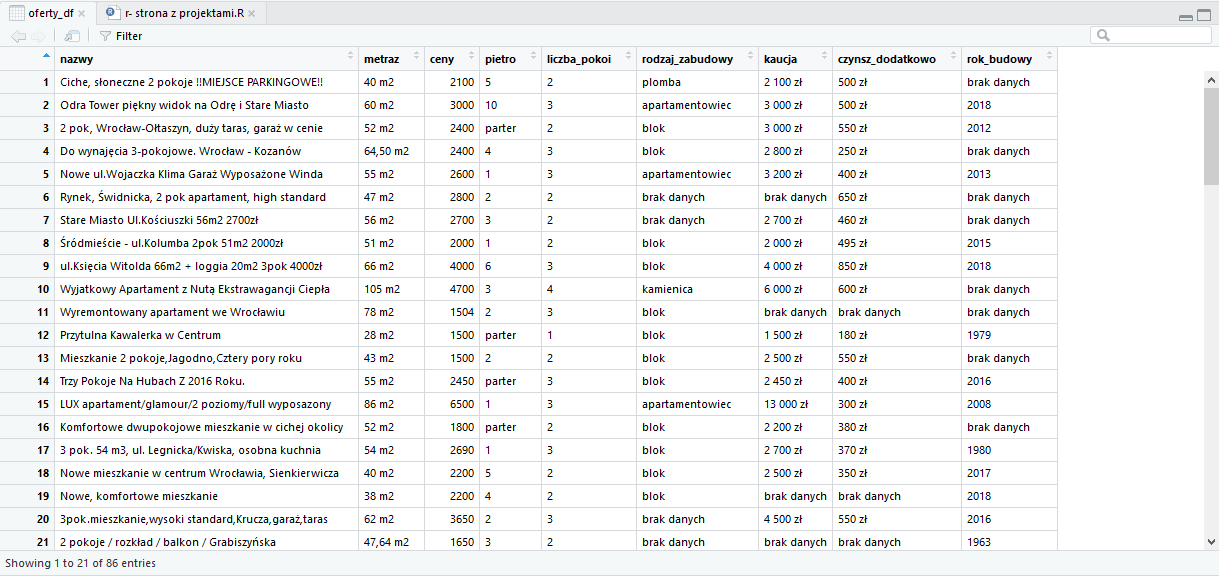
Menu aplikacji pozwala na wybór województwa z którego aplikacja będzie pobierać dane na temat ofert mieszkań. Jest to jedyny parametr który wybieramy w menu aplikacji. Po uruchomieniu aplikacji i wyborze parametru otrzymujemy dane w bazie danych z wybranymi wartościami.

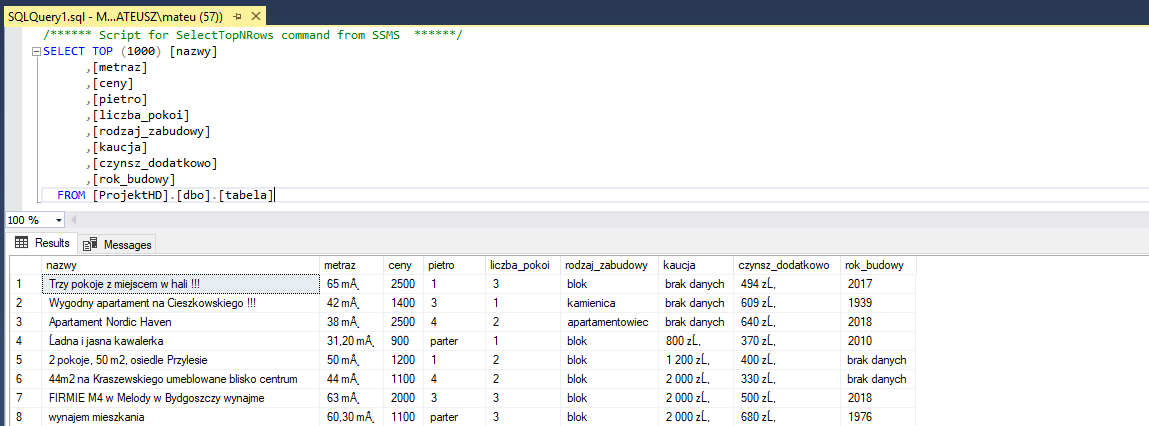
Screeny działania aplikacji:

**APLIKACJA - WYBÓR WOJEWÓDZTWA**

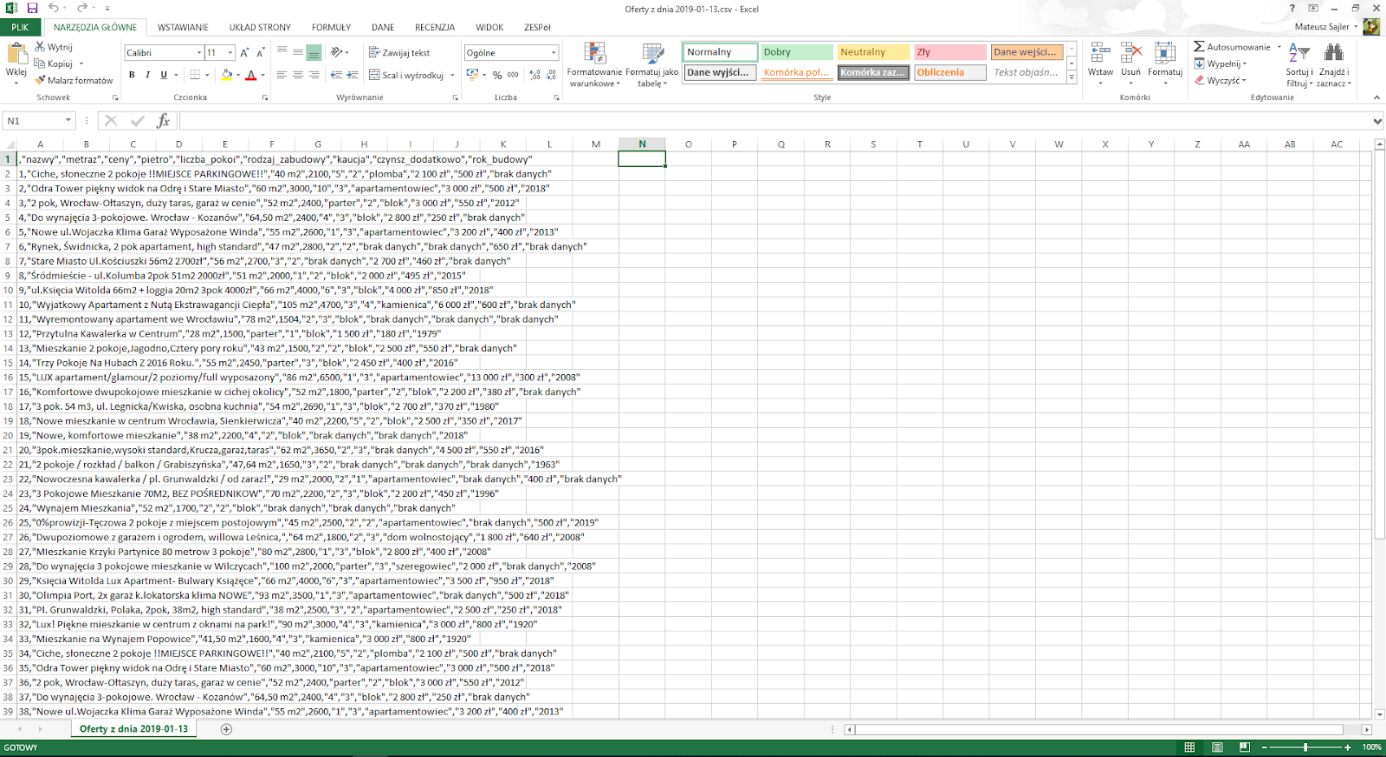


**APLIKACJA - OUTPUT**



**BAZA DANYCH**  

**PLIK CSV**



1. |  |
   | --- |
   | \_\_/70 pkt |

   [↑](#footnote-ref-1)