|  |  |
| --- | --- |
| **Hurtownie i Eksploracja Danych** | |
| **Temat** | Sieć restauracji |
| **Przygotowali** | Piotr Kaczmarczyk, Przemysław Postrach |

1. **Opis problemu**

W ramach projektu opracowaliśmy hurtownie danych wspomagającą prace sieci restauracji “Vege cziken w majo”. Dane pochodzą z systemu transakcji synchronizującego prace całej sieci. Zakupy mogą być dokonywane osobiście w restauracji, w specjalnych kioskach oraz online w aplikacji. Płatności są przyjmowane przez wszystkie dostępne formy min. Gotówkowo, PAYU czy kartą płatniczą. Dzięki dacie wygenerowania paragony przeprowadzana jest synchronizacja w i kolejność dodawania rekordów w bazie.

W systemie rejestrowane są również dostawy. Aplikacja wysyła informację o miejscu dostawy oraz o pracowniku odpowiedzialnym za przygotowanie i skompletowanie zamówienia. W sieci restauracji produkty są pogrupowane w kategorie i zorganizowane w różne menu. Każda restauracja posiada takie same produkty, kategorie oraz menu.

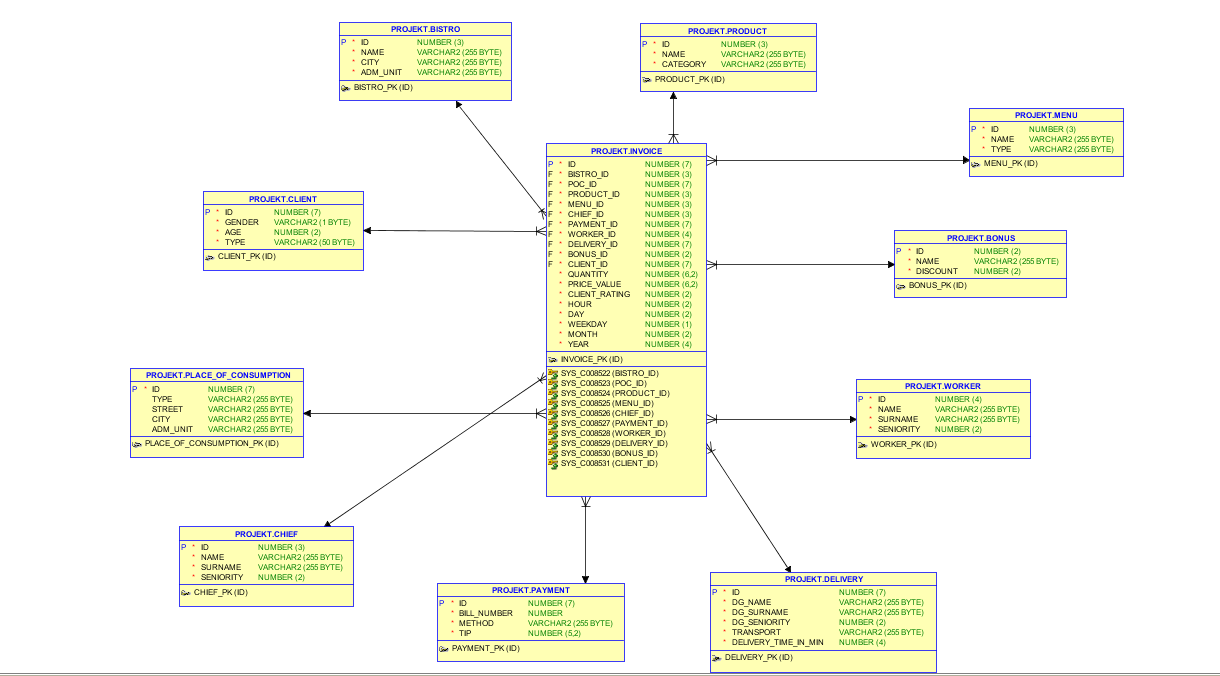
Ponadto sieć oferuje różne bonusy, które można wykorzystać przy zakupach. Dostępne dla wszystkich klientów, nie tylko posiadaczy kart stałego klienta. Z pomocą sztucznej inteligencji system może codziennie generować promocje dla klientów oraz przewidywać jakie promocje będą się najchętniej sprzedawały. Dzięki temu możemy generować popyt i podaż w zależności od stanów magazynowych czy okresów gorszej sprzedaży.

Każdy klient może ocenić każdy posiłek w skali od 1 do 10 przy kasie. Dzięki temu możemy przechowywać informacje o tym, co, kiedy i gdzie klientom podobało się bardziej lub mniej. Dane te są podstawą do analiz danych i predykcji poprzez sieci neuronowe. Najbardziej podstawowym elementem analizy jest sprawdzenie czy dana transakcja ucieszyła klienta. Pozwala to nam przewidzieć co należy poprawić i udoskonalić w naszych restauracjach.

System zarządzania siecią działa w trybie online i na bieżąco dostarcza dane do magazynu. W ramach projektu zaprojektowaliśmy go i symulowaliśmy dane wprowadzane do niego przez system na przestrzeni kilku miesięcy. Hurtowania posiada dane nie tylko o transakcji zakupionych produktach, ale też o pracownikach, klientach, dostępnym menu oraz kiedy transakcja została przeprowadzona.

1. **Hurtownia danych**

Schemat bazy danych:



Hurtowania danych realizuje klasyczny schemat gwiazdy. Wszystkie tabele posiadają połączenie jeden do wielu z tabelą “Invoice”.

Miary:

1. Quantity – waga zakupionego posiłku
2. Price\_value – cena
3. Client\_rating – ocena klienta
4. Data

Zostały wybrane ze względu na optymalizacje zapytań oraz możliwości późniejszej analizy data minerem.

Wymiary:

1. Client – informacje z karty stałego klienta
2. Place\_of\_consumption – informacje o miejscu konsumpcji
3. Bistro – informacje o restauracji
4. Menu – posiada nazwę i typ
5. Bonus – informacje o zniżce
6. Produkt – informacje o produkcie i jego kategorii
7. Worker – informacje o pracowniku
8. Chief – informacje o szefie kuchni
9. Payment – informacje o płatności
10. Delivery – informacje o dostawie

Wymiary pomagają nam zebrać bardziej szczegółowe informacje o restauracji, klientach oraz posiłkach. Dzięki temu profilowanie klientów będzie sprawniejsze. Będziemy mogli sprawdzić jakie oni mają upodobania oraz co najczęściej się sprzedaje.

1. **Instalacja**

Instalacja oraz konfiguracja bazy danych przebiegłą w skrócony sposób. Wykorzystaliśmy obraz maszyny wirtualnej opartej na WIN 10 z zainstalowanym Oracle sql server oraz sql devleoper. Została ona nam udostępniona przez prowadzącego.

Należało jedynie nadać uprawnienia do dla użytkownika “projekt” aby można było bez przeszkód wykonywać zadania na bazie danych. Zostało to zrobione poprzez egzekucje komendy:

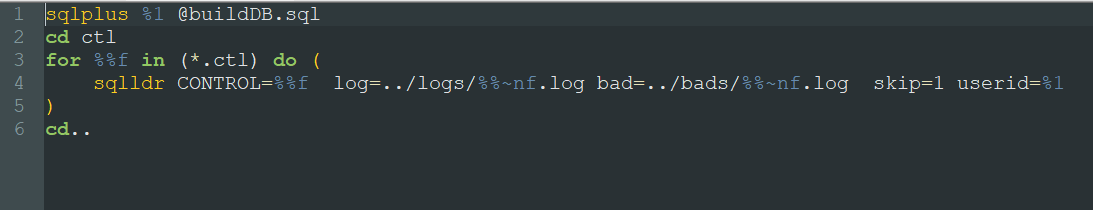
**GRANT ALL PRIVILEGES TO PROJEKT;**

Przygotowany obraz posiadał dostępny moduł “Data miner” należało go jedynie pobrać przy pierwszym uruchomieniu bazy. Często można spotkać się z błędami przy jego instalacji, dlatego na postawie miejsca oraz nazwy wykonywanego SQLa należy w odpowiednim miejscu dodać poniższą linijkę:

**alter session set "\_oracle\_script" = true;**

1. **Zasilenie hurtowni danymi**

Zasilenie hurtowni danymi odbyło się poprzez skrypt zawarty w pliku “createAndLoad.bat”.



Skrypt odpalamy w konsoli komendom:

**CreateAndLoad.bat \*użytkownik\*/\*hasło\***

W pierwszej linii zostaje uruchomiony sql plus i wykonany kod SQL usuwający bazę oraz jej zależności i tworzący tabele na nowo.

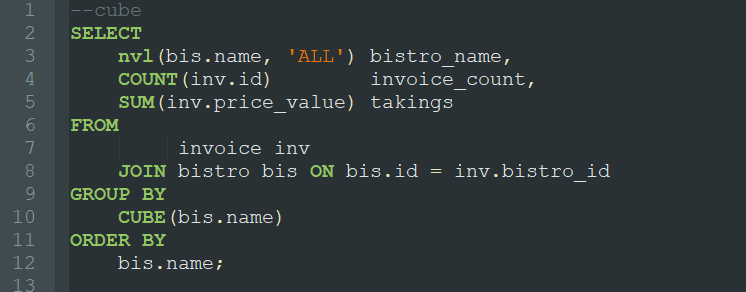
Następnie zostają wykonane skrypty ładujące dane do bazy danych. Ze względu na więzy integralności tabela “Invoice” zostaje załadowana na samym końcu. W tym celu nazwa zmieniona została na “ww\_Invoice.ctl”.

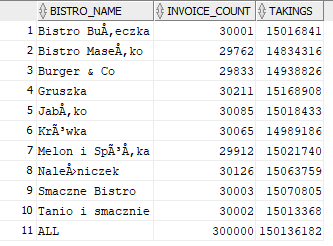
1. **Zapytania SQL**

Ustawienia bazy danych spowodowało brak obsługi polskich znaków.

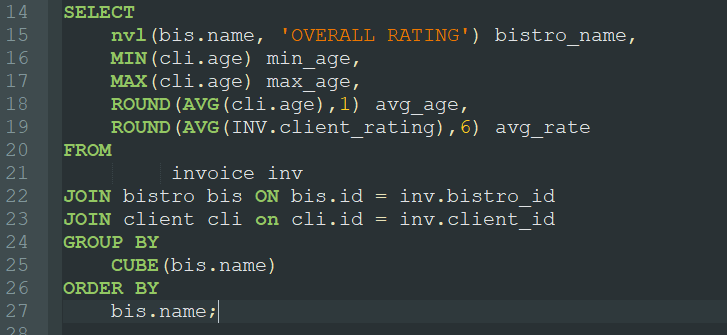
Cube:

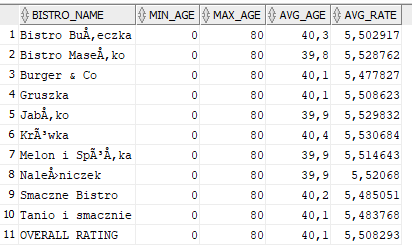
1. Zapytanie pokazuje która restauracja jest najbardziej rentowna



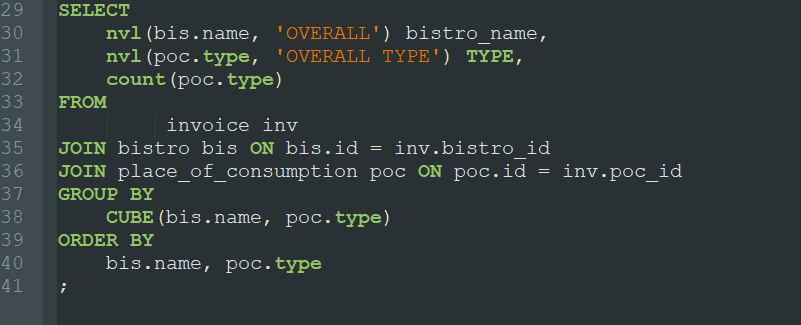


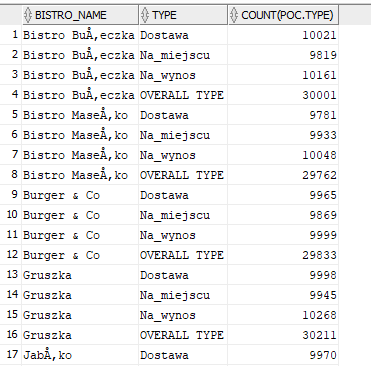
1. Zapytanie pokazuje minimalny, maksymalny wiek klienta oraz jaką ocenę wystawił.





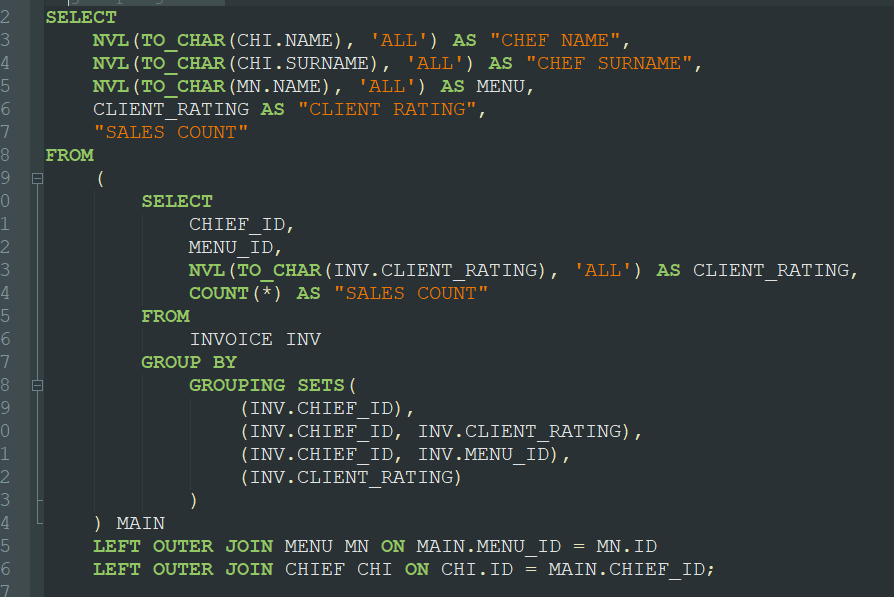
1. Zapytanie pokazuje jakie dostawy zostały wykonane dla danej restrauracji.

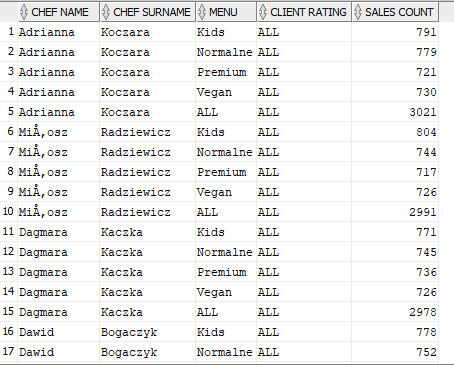




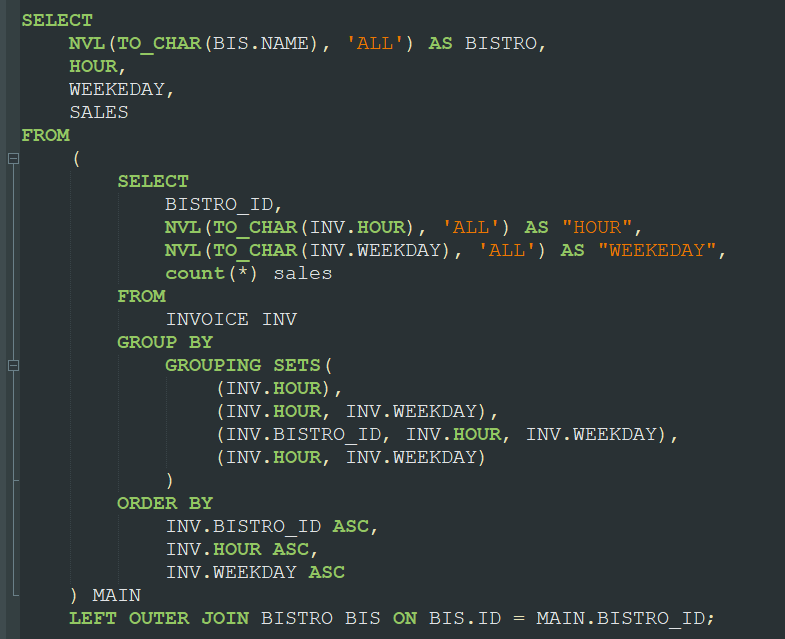
Grouping sets:

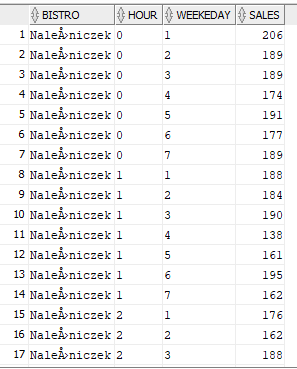
1. Zapytanie pokazuje kucharza, wykonany posiłek, ocenę klienta oraz ilość przygotowanych posiłków.



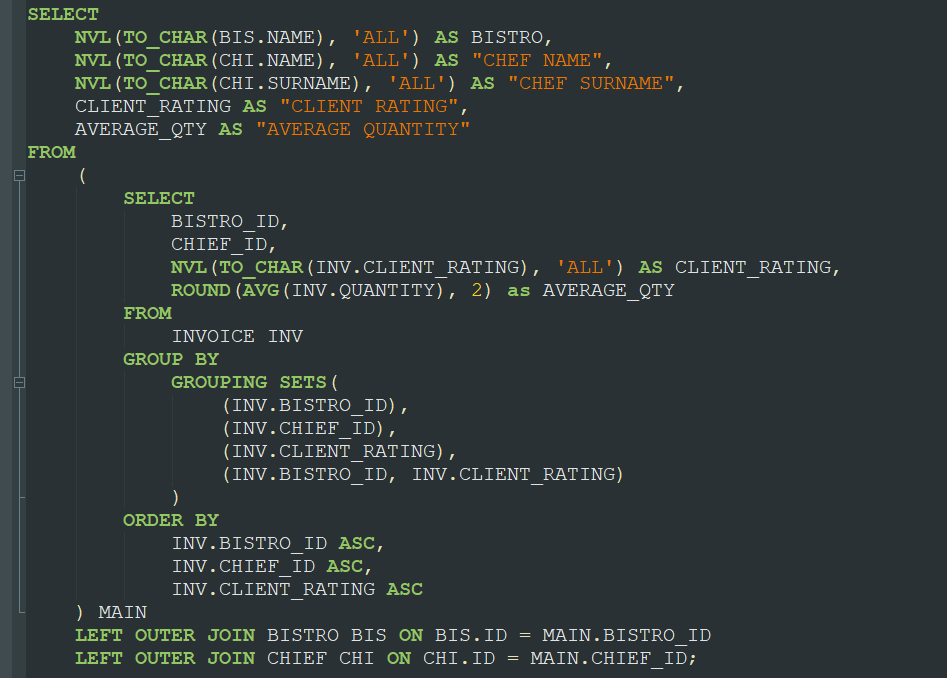


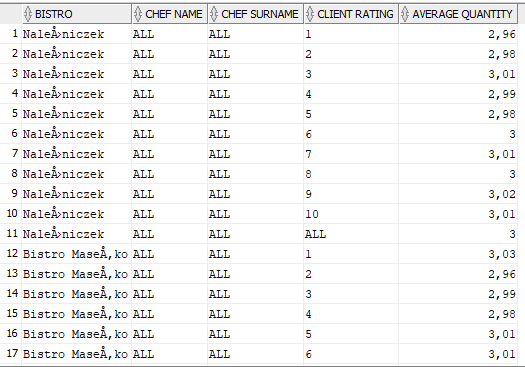
1. Zapytanie pokazuje sprzedaż z podziałem na dni tygodnia i godziny.





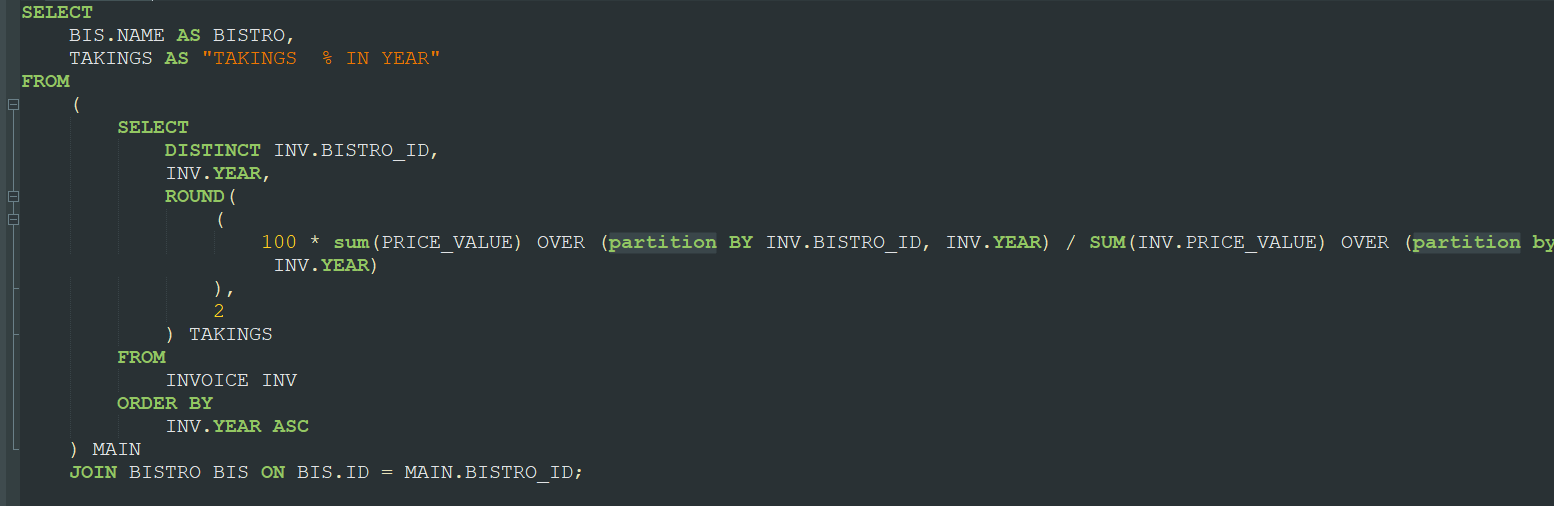
1. Zapytanie pokazuje jaki szef przygotował jaką ilość jedzenia, dodatkowo pokazana została średnia ocena klienta.

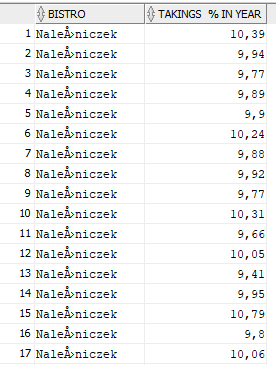




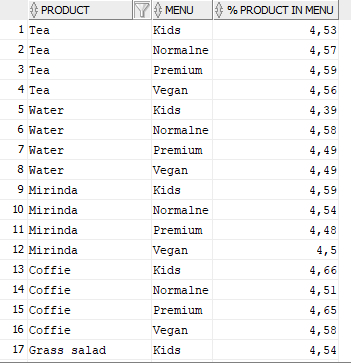
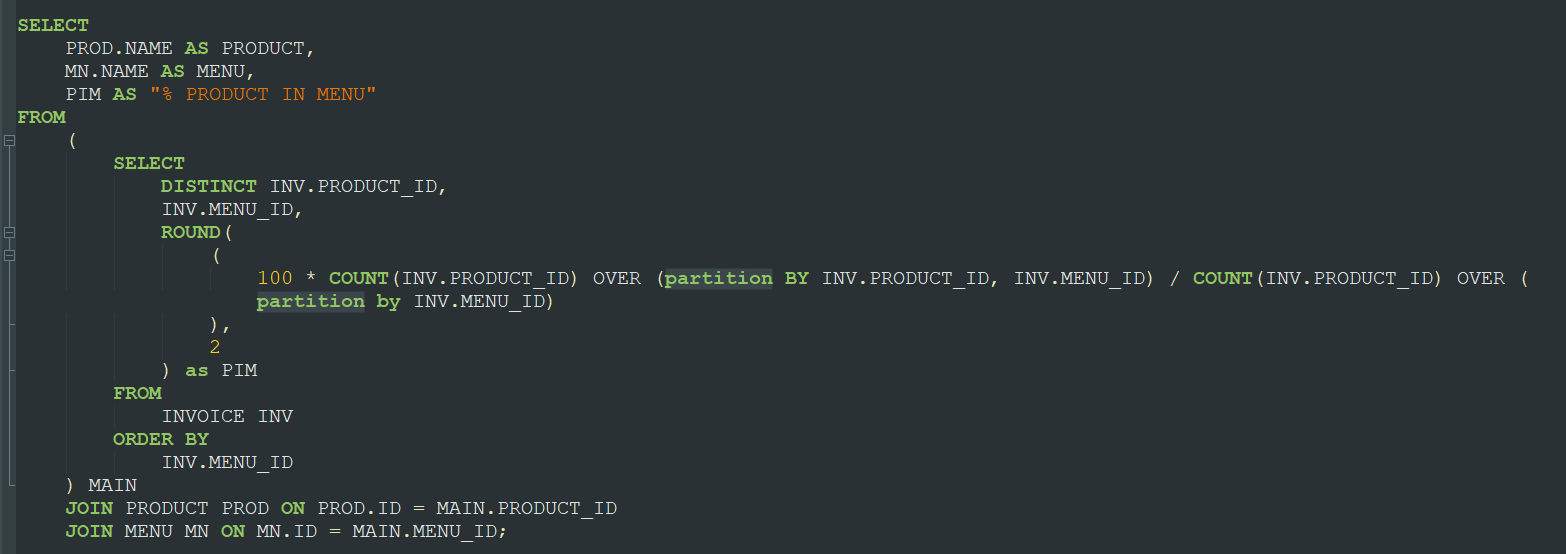
Partition:

1. Zapytanie pokazuje % utargu danej restauracji.

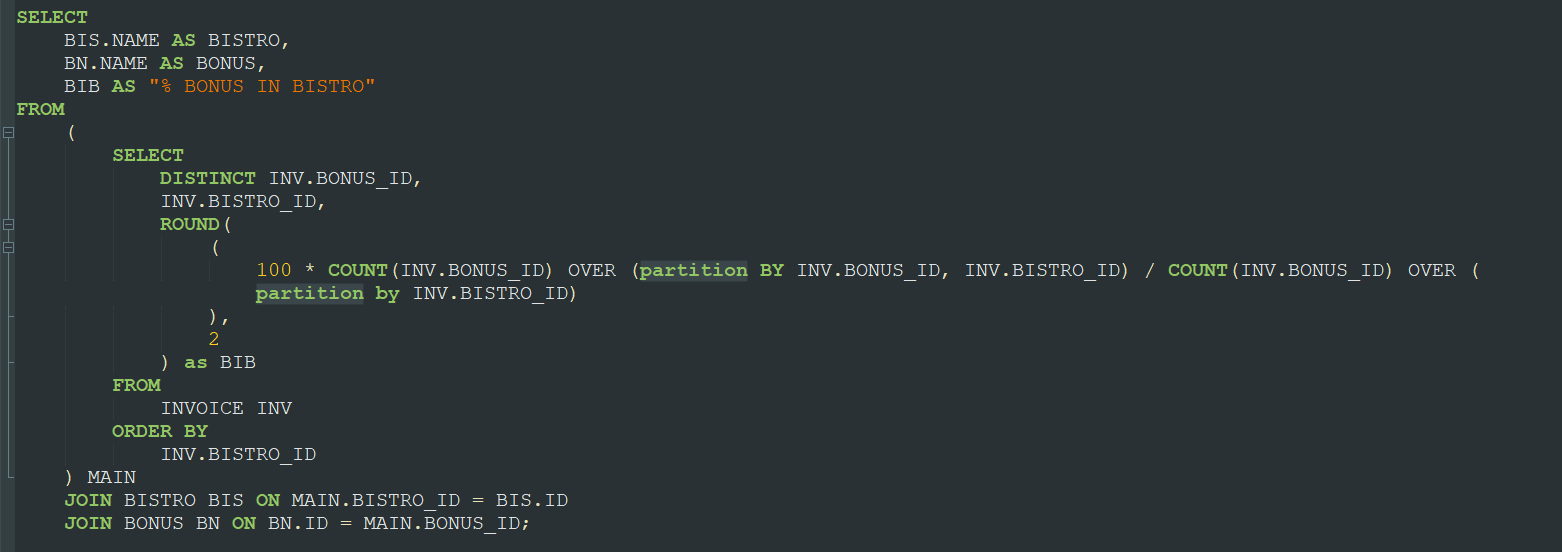


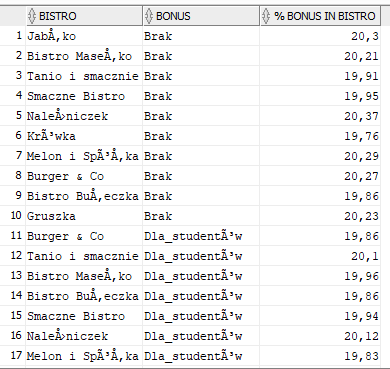


1. Zapytanie pokazuje najczęściej kupowany produkt w menu



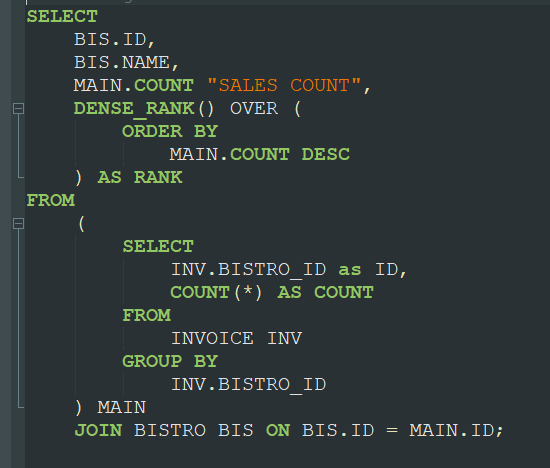
1. Zapytanie pokazuje rabaty w danym bistro.

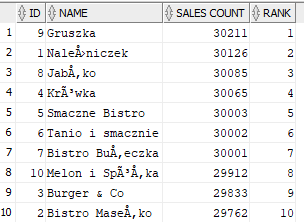




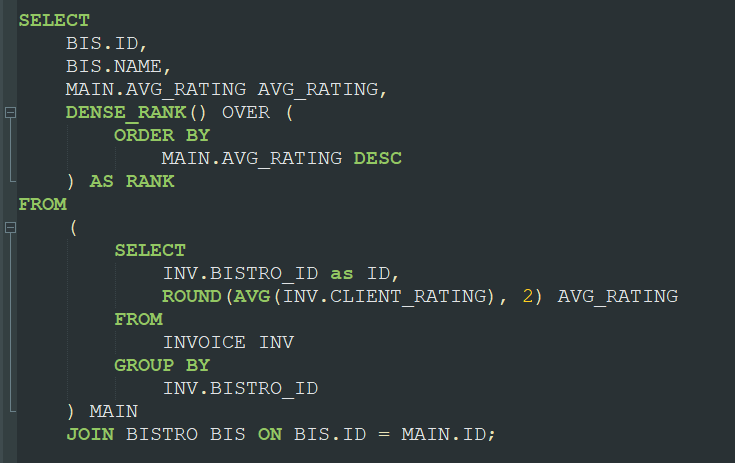
Rank:

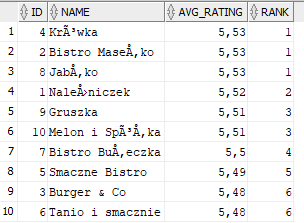
1. Ranking restauracji ze względu na sprzedaż



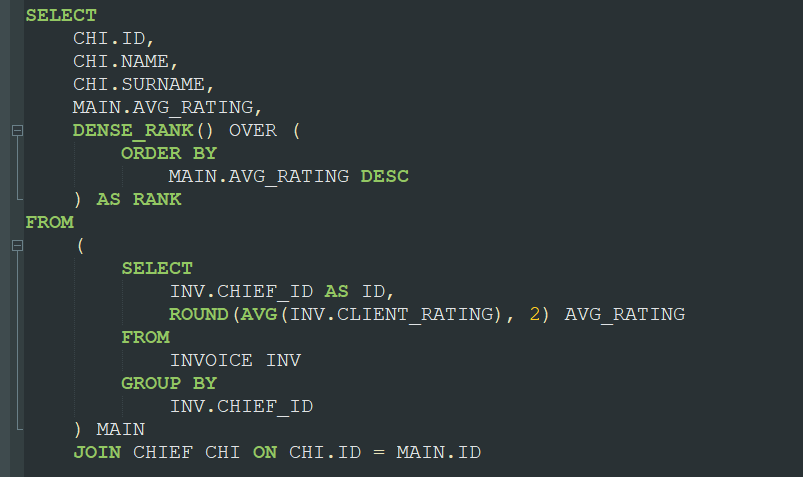


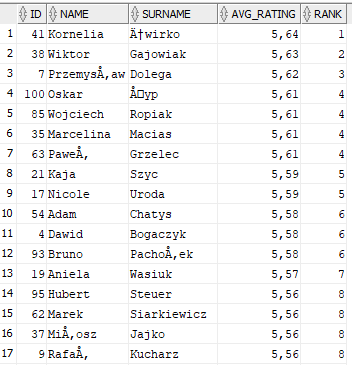
1. Ranking ze względu na oceny klientów





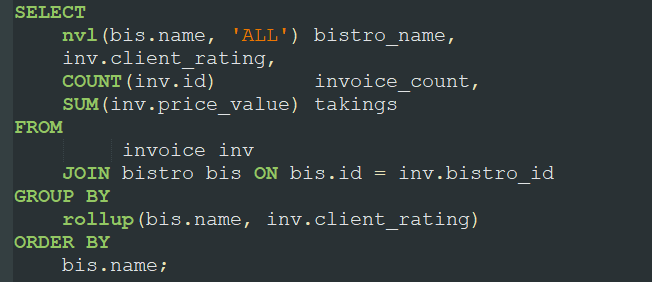
1. Ranking szefów kuchni

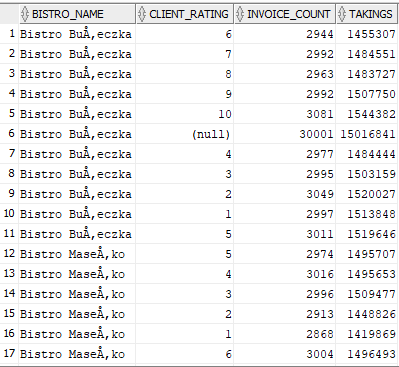




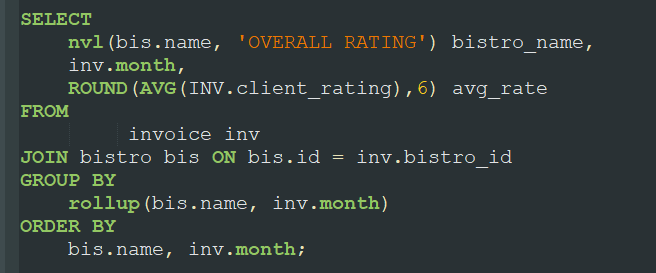
Rollup:

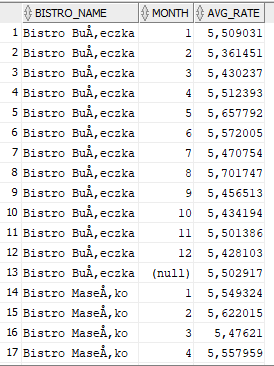
1. Zapytanie pokazuje utarg w danych miesiącach i latach.



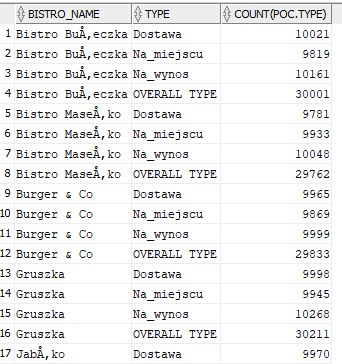
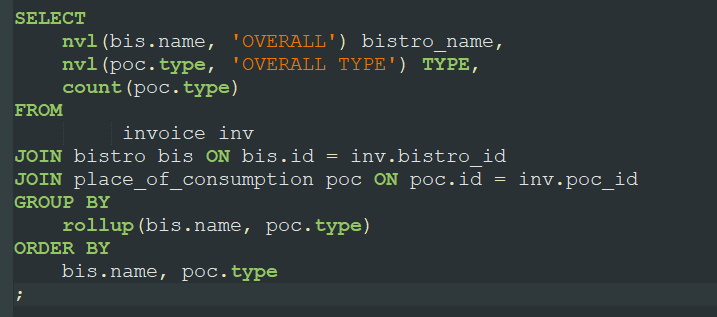


1. Zapytanie zwraca liczbę sprzedaży w danym bistro w danym miesiącu i roku danego produktu.



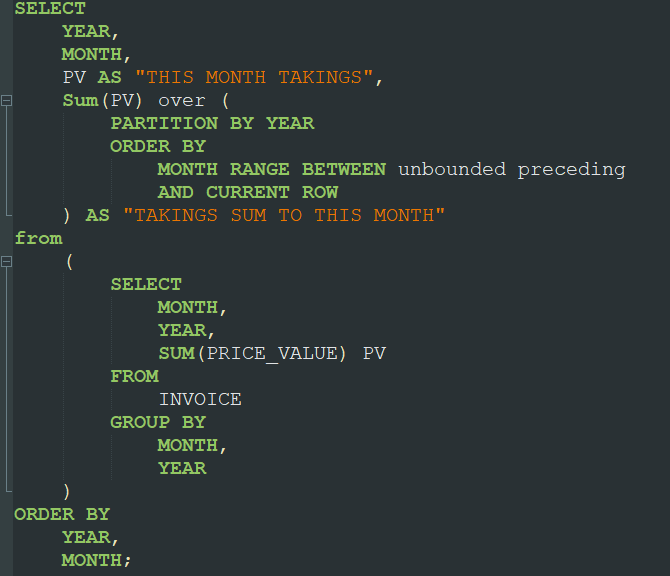


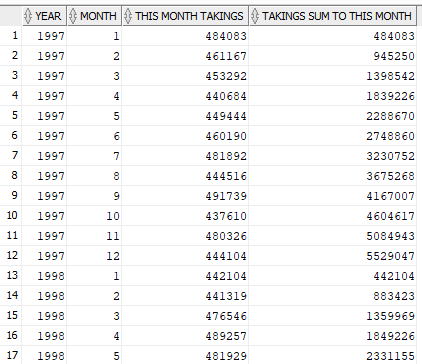
1. Zapytanie zwraca liczbę zamówień danego menu w restauracji



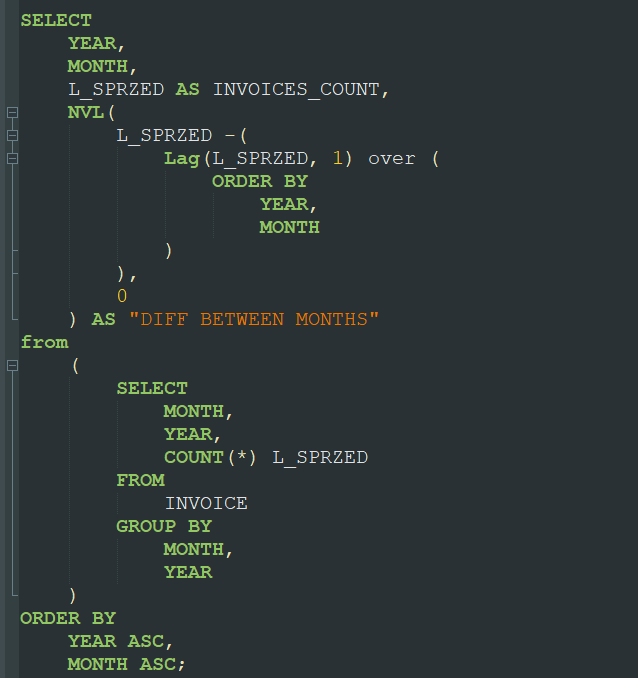
Window:

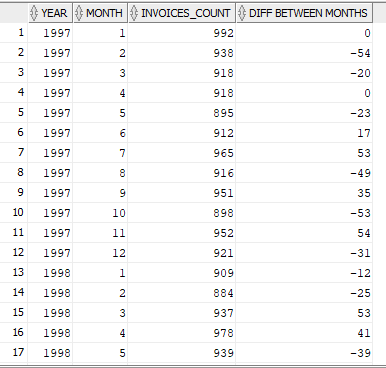
1. Zapytanie zwraca przyrosty dochodu z miesiąca na miesiąc



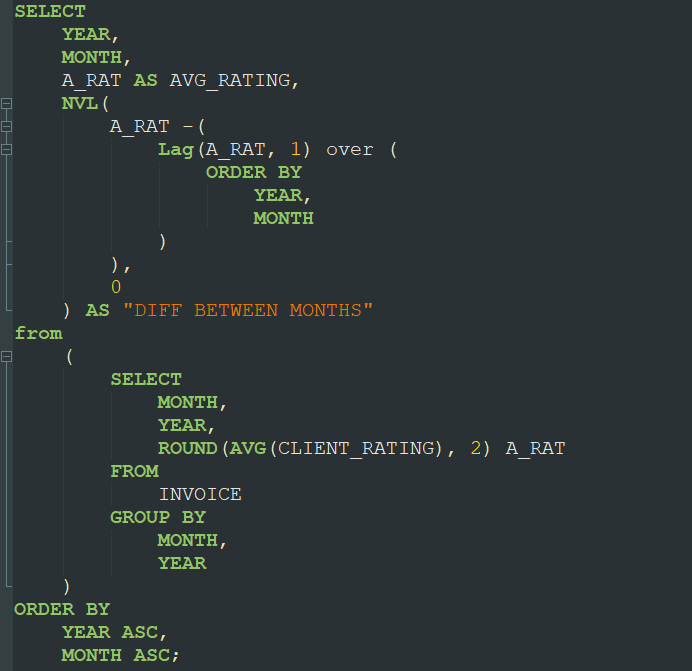


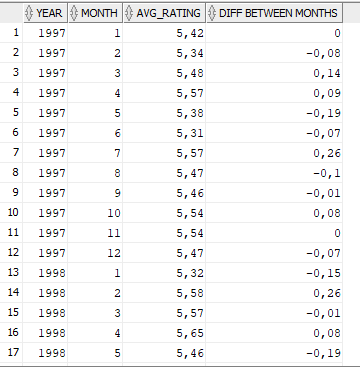
1. Zapytanie zwraca różnicę w ilości sprzedaży między bieżącym miesiącem a poprzednim.



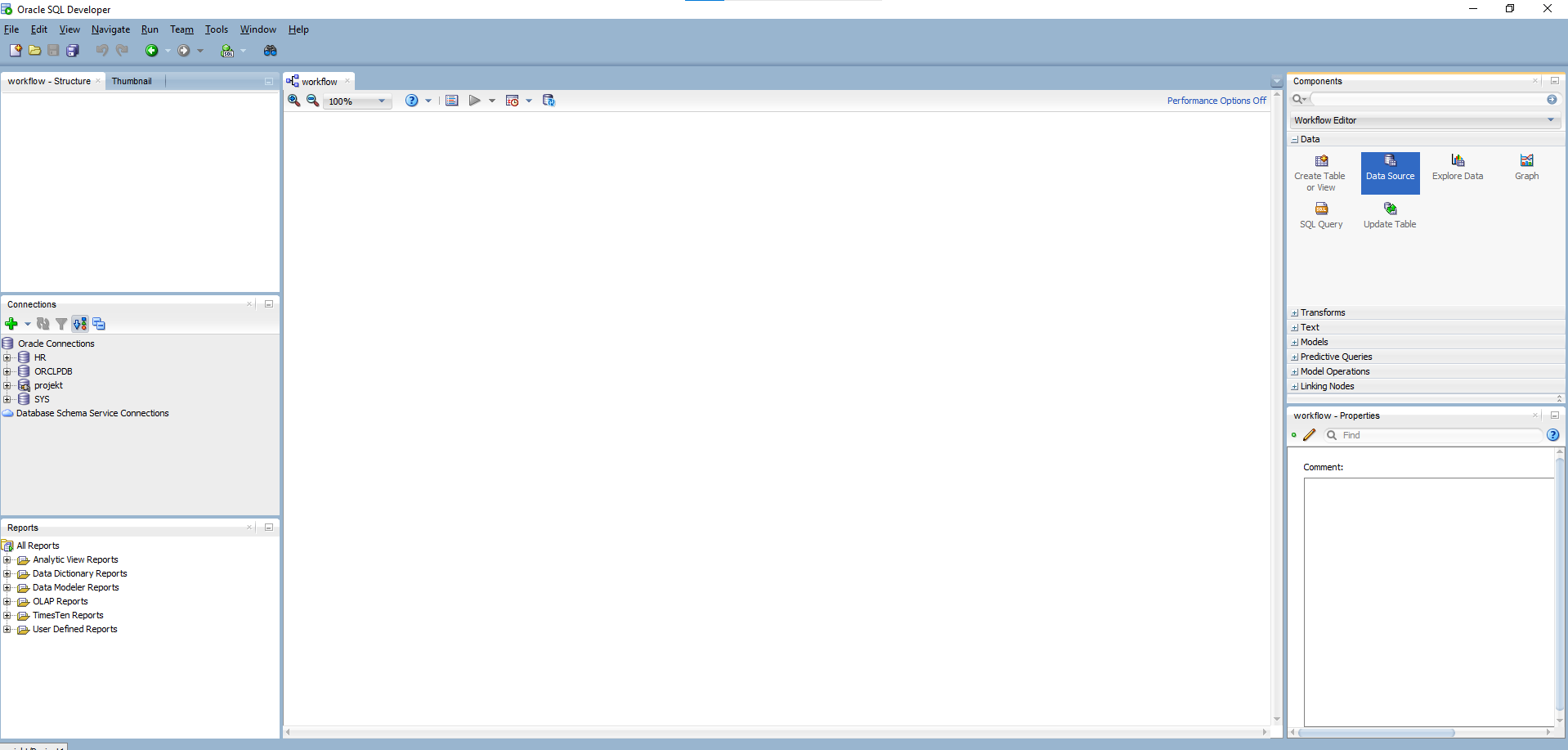


1. Zapytanie zwraca, jak się zmieniała średnia ocen klientów.



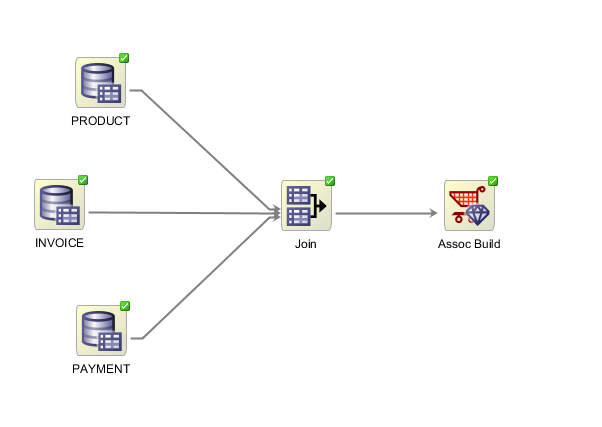


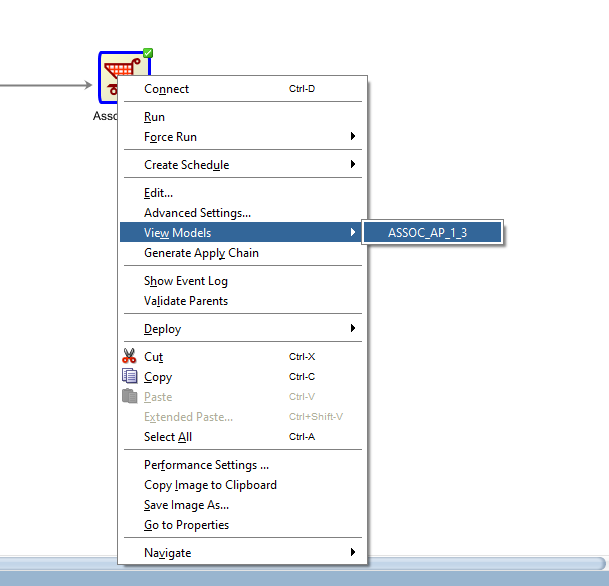
1. **Eksploracja danych**

Po uruchomieniu data minera należy stworzyć nowy projekt. Dla każdego podejścia eksploracji danych został stworzony osobny workflow. Do zbudowania schematu zostało wykorzystane menu po prawej stronie, z którego możemy tworzyć schematy.

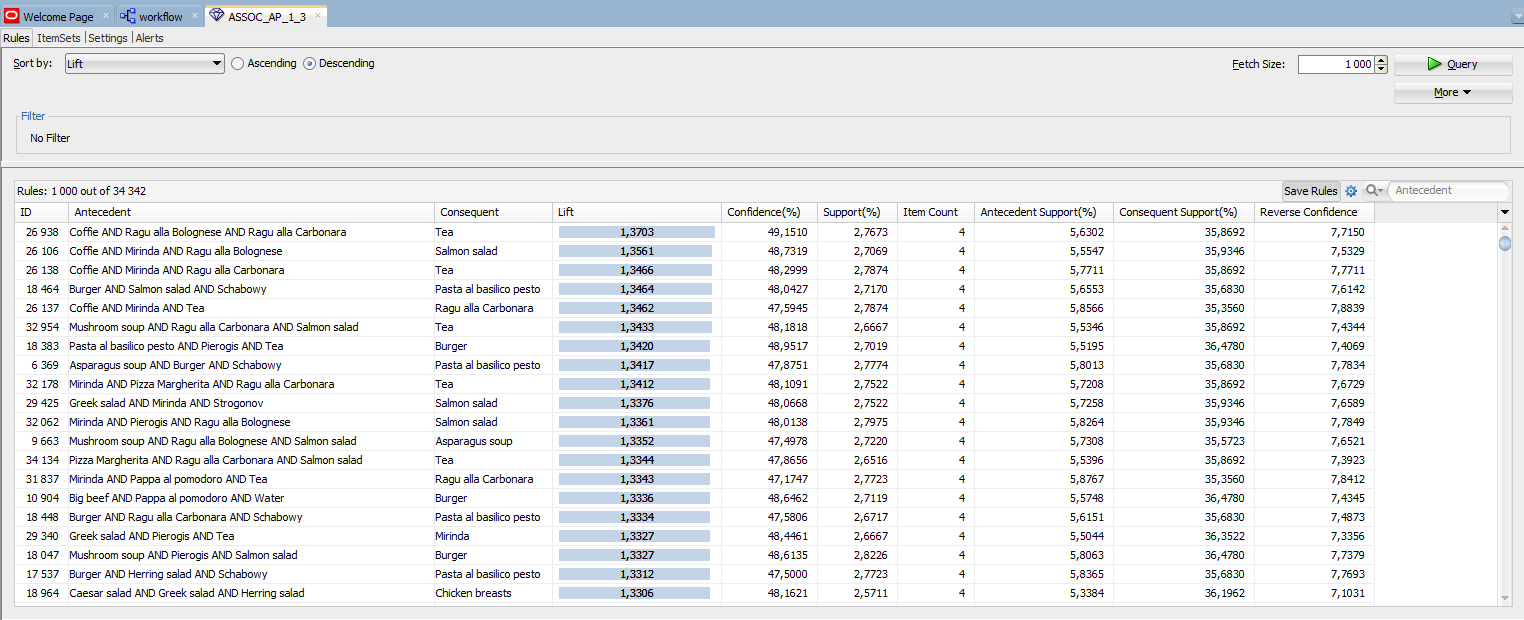
1. Reguły asocjacyjne

Schemat:



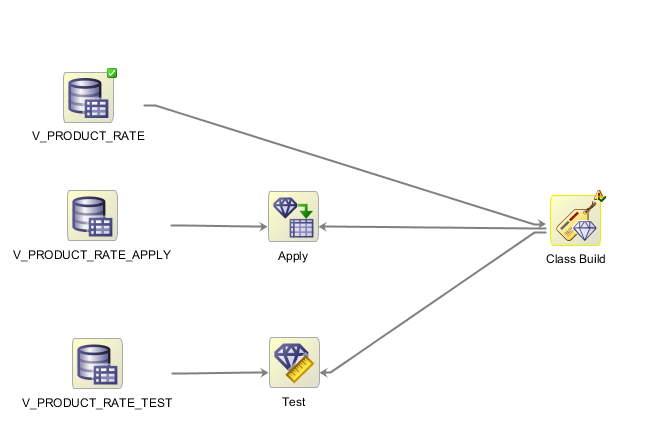
Po uruchomieniu modelu możemy podejrzeć wynik. Jest on dostępny po kliknięciu prawym klawiszem na ostatni moduł. 

Otrzymany wynik:

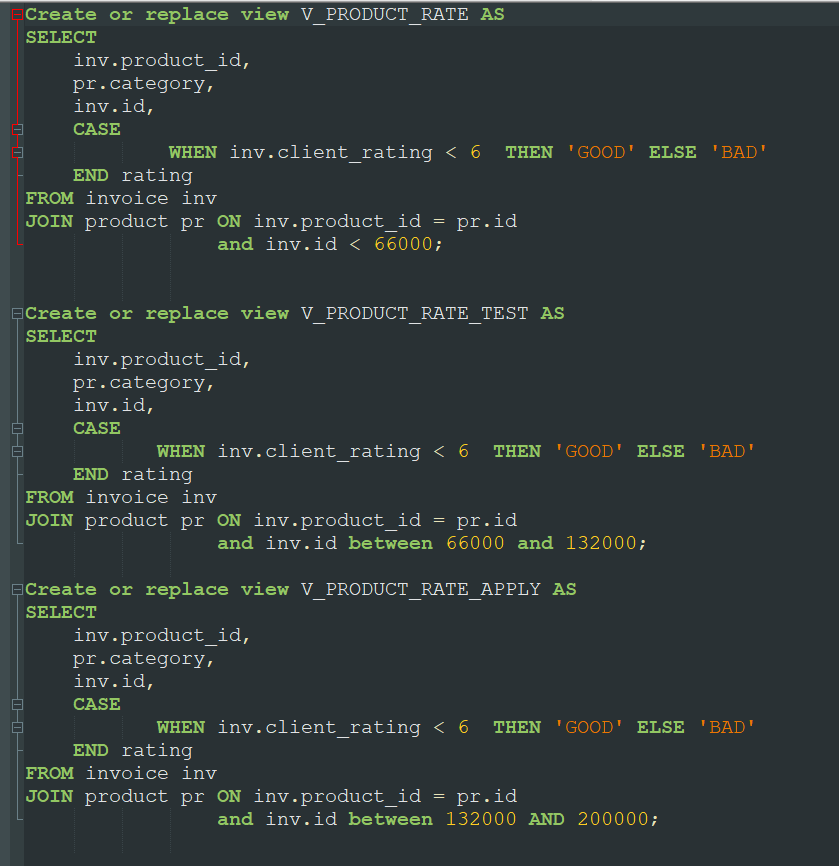


1. Klasyfikacja

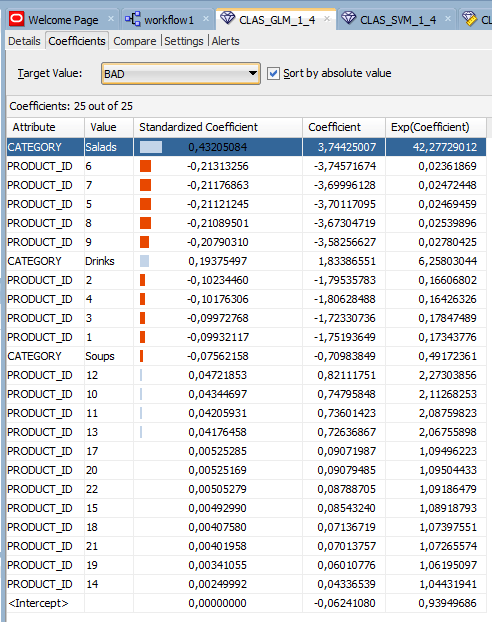
Schemat został zbudowany w taki sam sposób jak poprzedni, ale z wykorzystaniem innych bloczków.

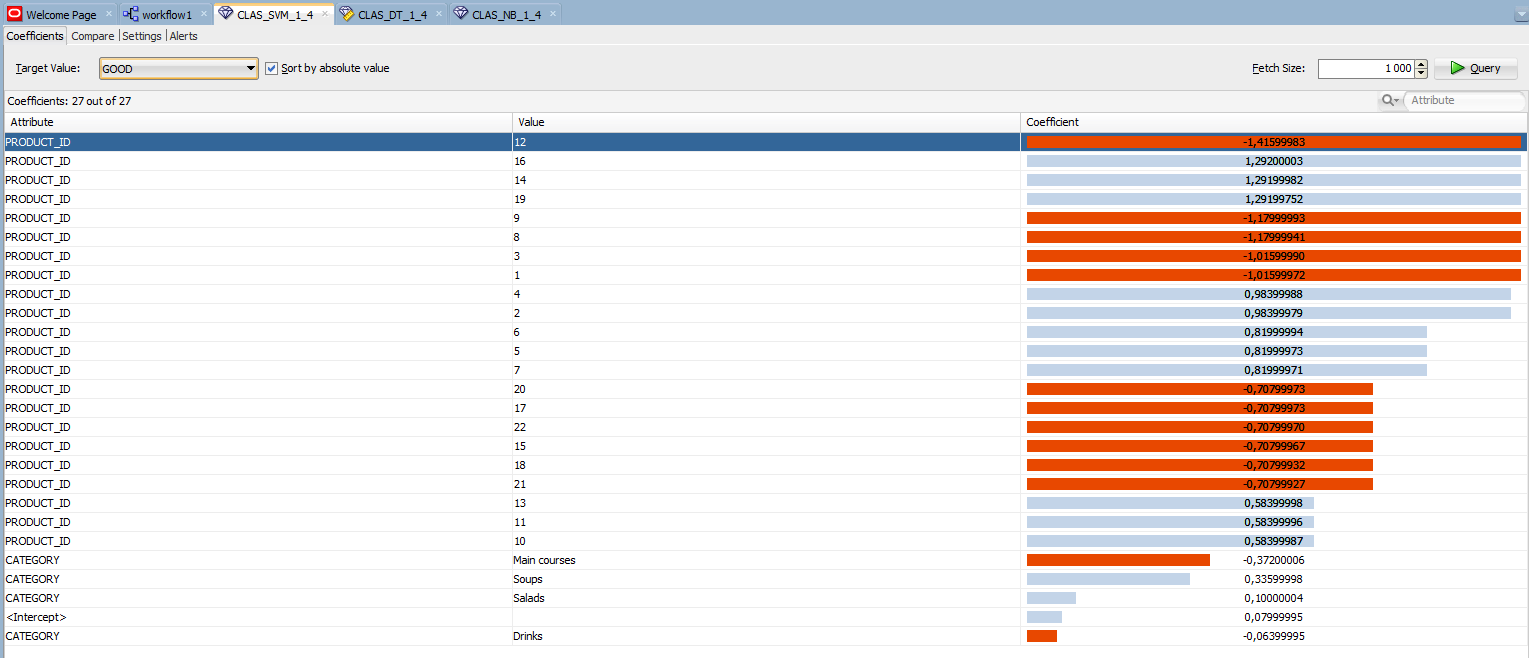
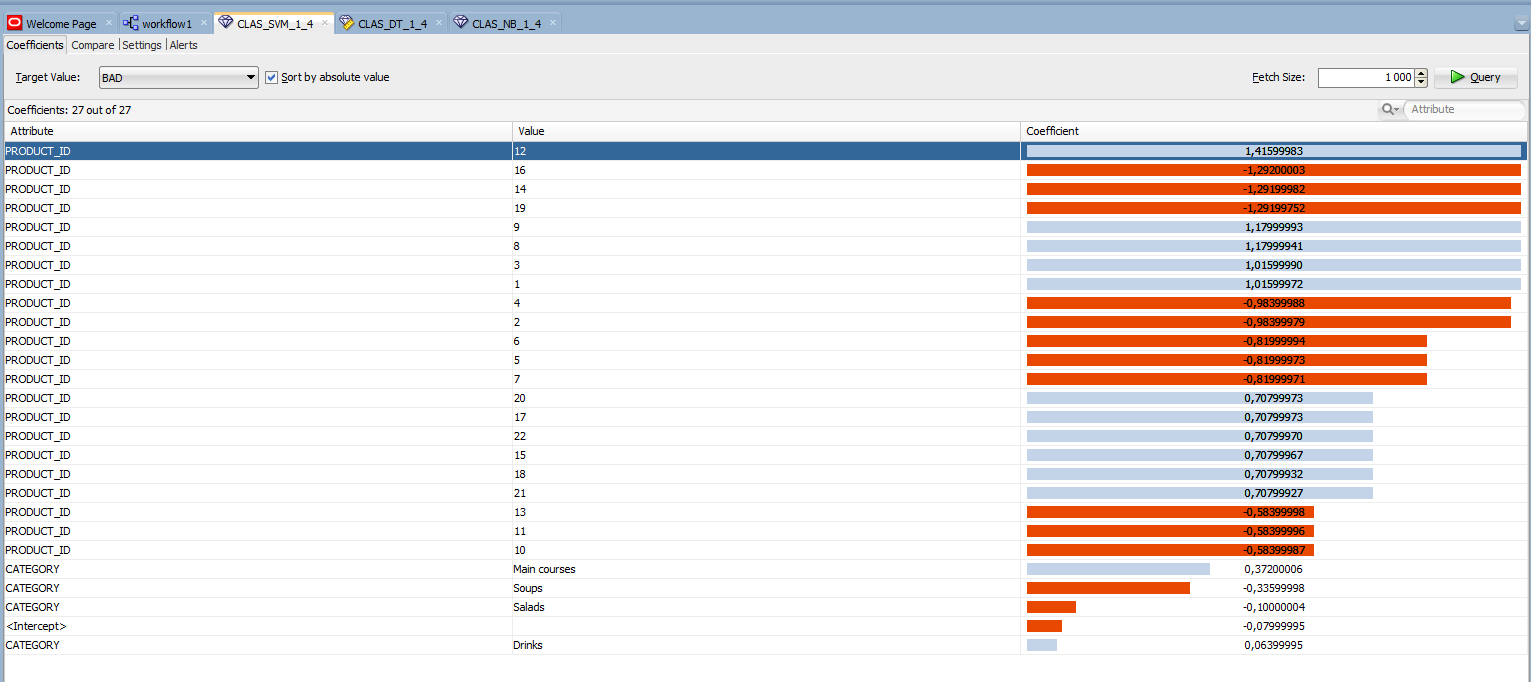
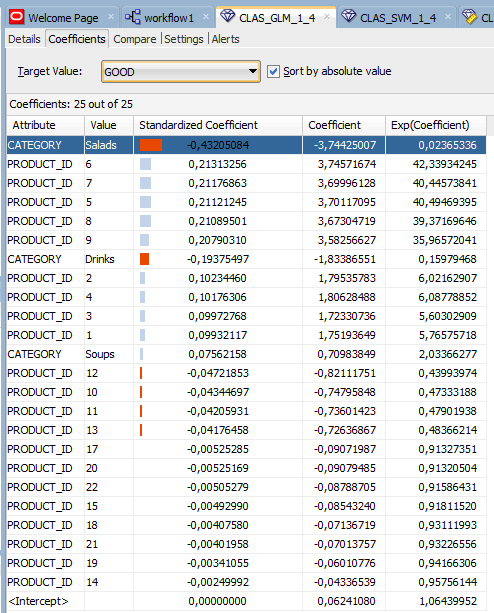


Aby można było podzielić dane na testowe, uczące oraz weryfikujące zostały stworzone 3 widoki dzielące je na równe części.



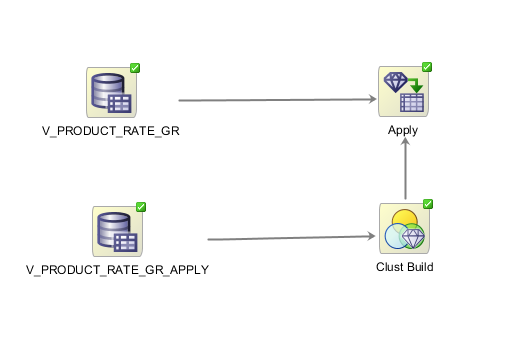
Po uruchomieniu otrzymujemy wyniki:



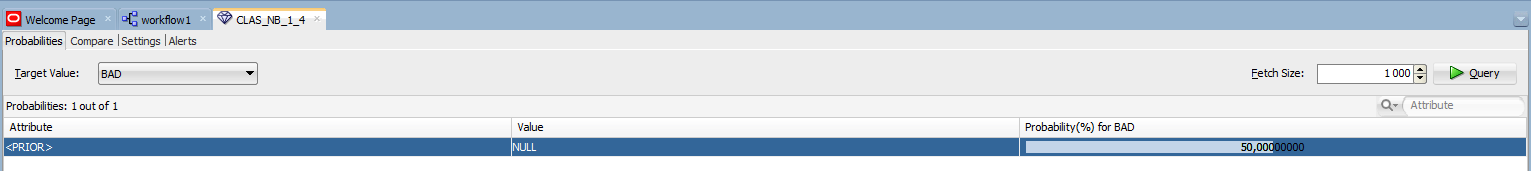


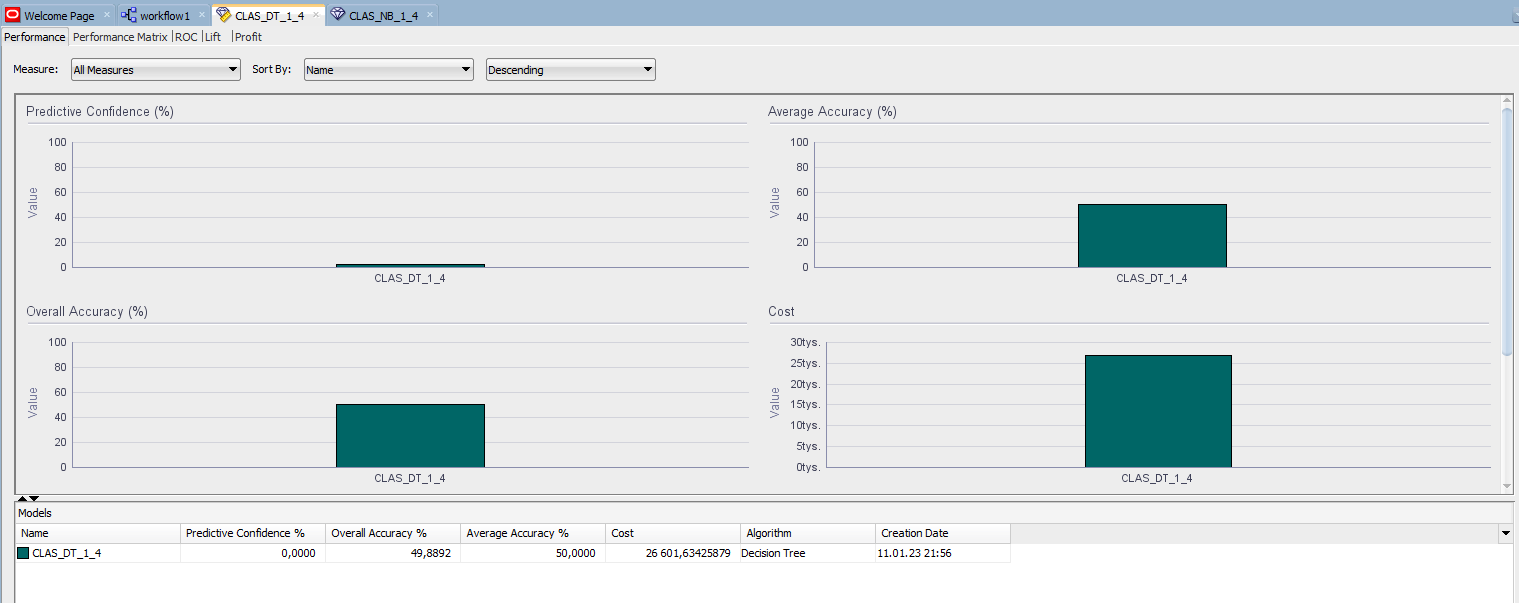
1. Grupowanie

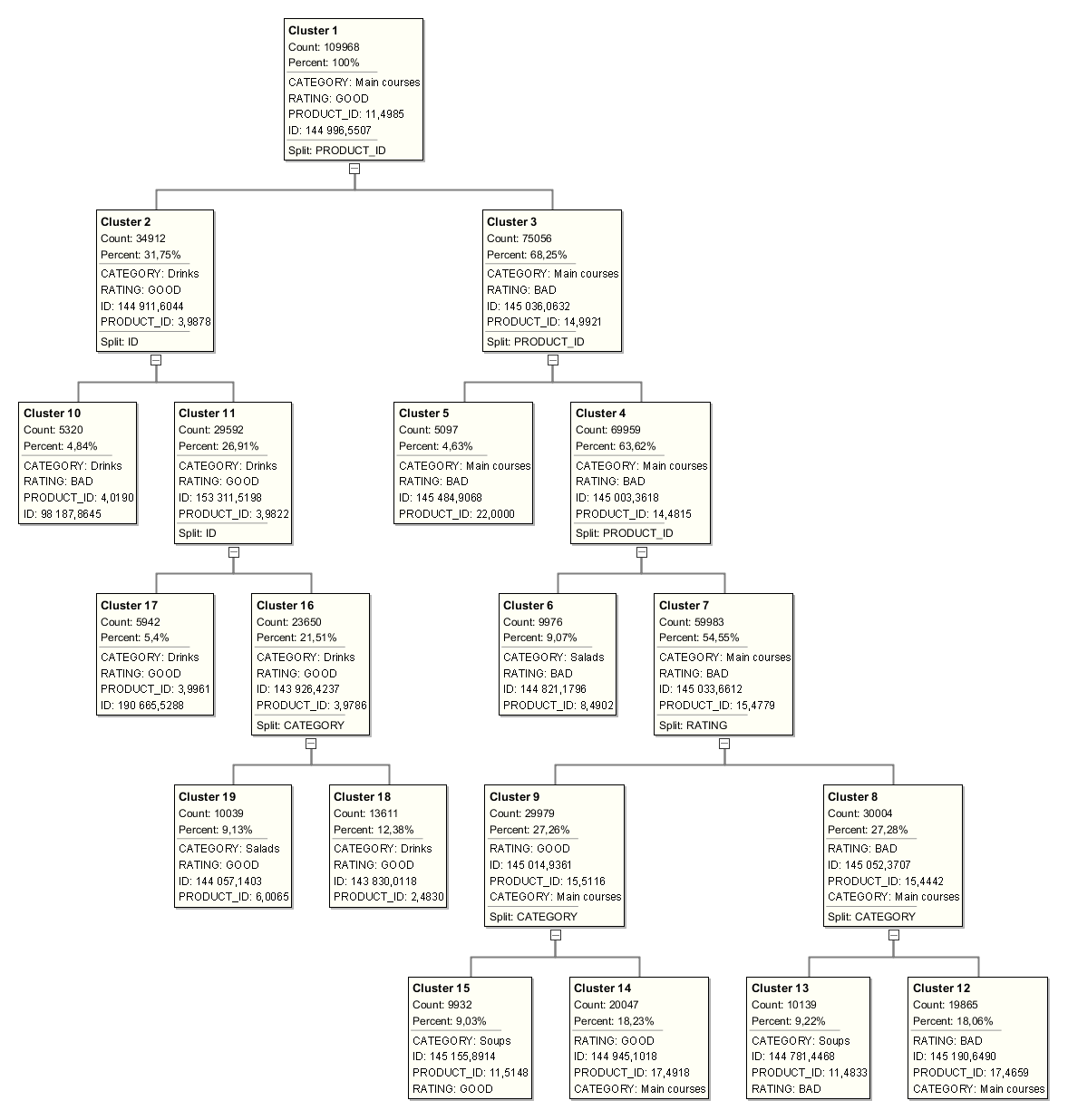
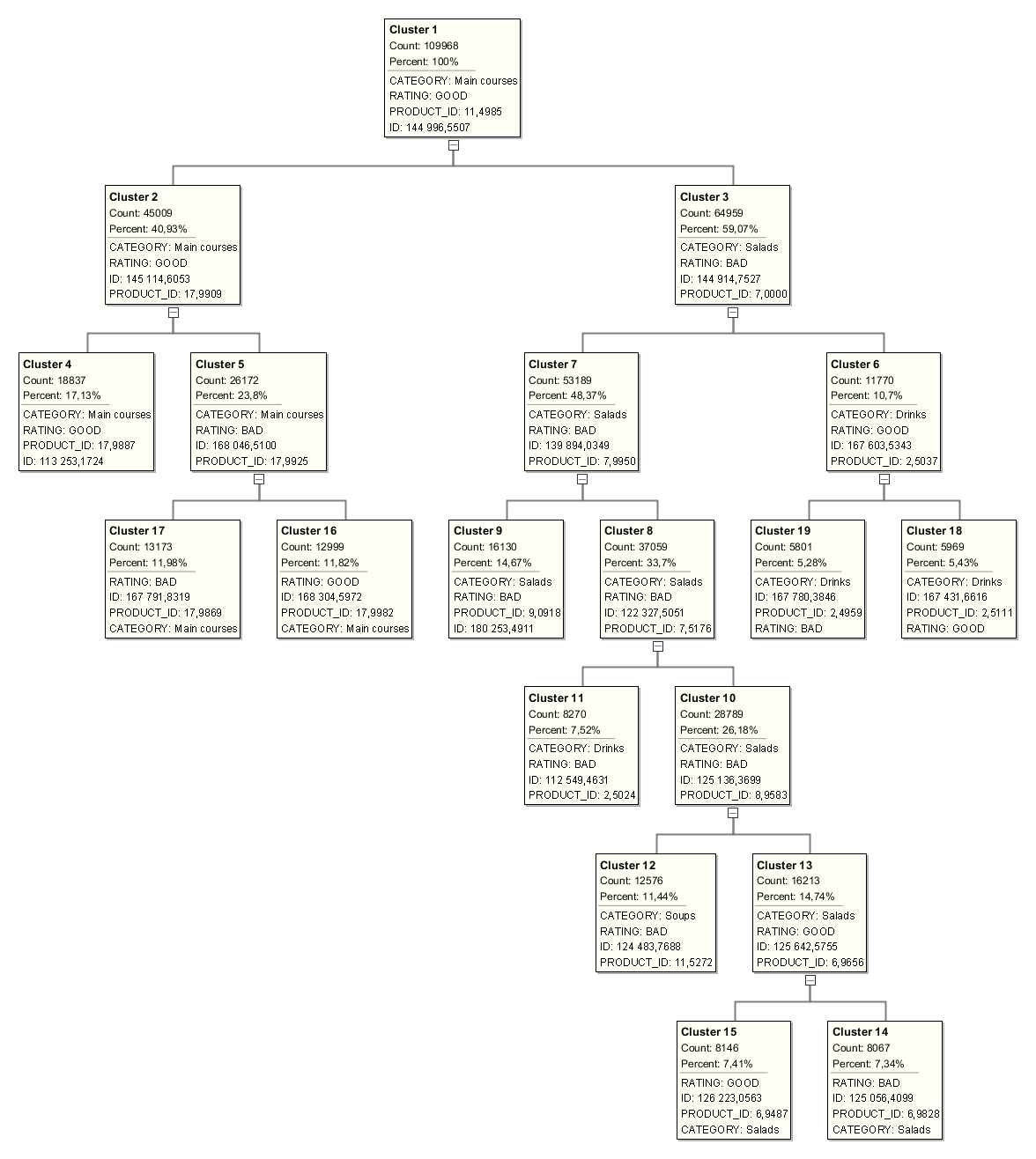
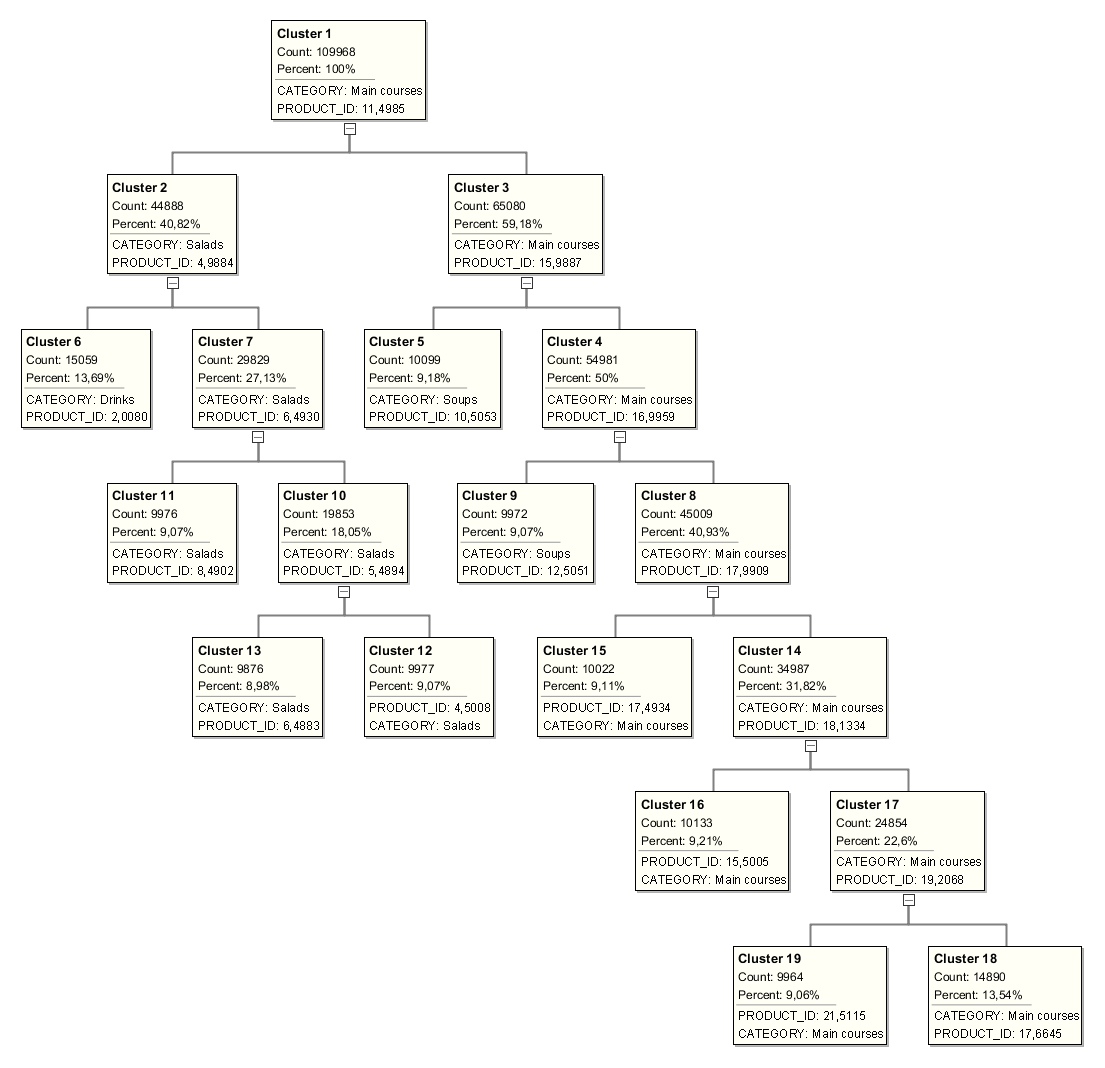
Zostało wykonane podobnie do dwóch poprzednich schematów jednak z innymi bloczkami.



Otrzymane wyniki:







1. **Wnioski**

Projekt pozwolił zbudować hurtowanie danych oraz zapoznać się z pracą z nimi. Mogliśmy poznać jak optymalnie stworzyć hurtownię aby gromadzone dane były przydatne dla biznesu.

Stworzone zapytania pozwoliły nam się zaznajomić z zaawansowanymi konceptami jakie możemy spotkać w bazach danych Oracle SQL. Pokazuje nam to jakie możliwości raportowania możemy osiągnąć samym kodem bez zewnętrznych aplikacji. Ze względu na losowość danych nie niosą one zbyt wiele informacji.

Poznany moduł data miner pozwolił nam spojrzeć na to jak dane są przetwarzane oraz jakie zależności można wyłapać. Dzięki temu wiemy, jak tworzyć takie modele oraz używać ich w praktyce.