

Postęp prac – Projekt WBO

Udało się poprawnie przeprowadzić proces ETL oraz zbudować schemat hurtowni danych w postaci gwiazdy. Dużym problemem okazał się typ danych przechowywany w pierwotnych plikach CSV, stąd istniała konieczność poprawy informacji o głosowaniach. W tabeli dim_location będę czyścić wpisy, które zawierają wartości null. Dodatkowo planuję także powiązać wszystkie kody pocztowe z osiedlami we Wrocławiu (w pierwotnym źródle danych duża część kodów pocztowych nie ma przypisanych Osiedli). Schemat Hurtowni prezentuje się następująco:



Tabela faktów zawiera klucze obce z wyróżnionych wymiarów.

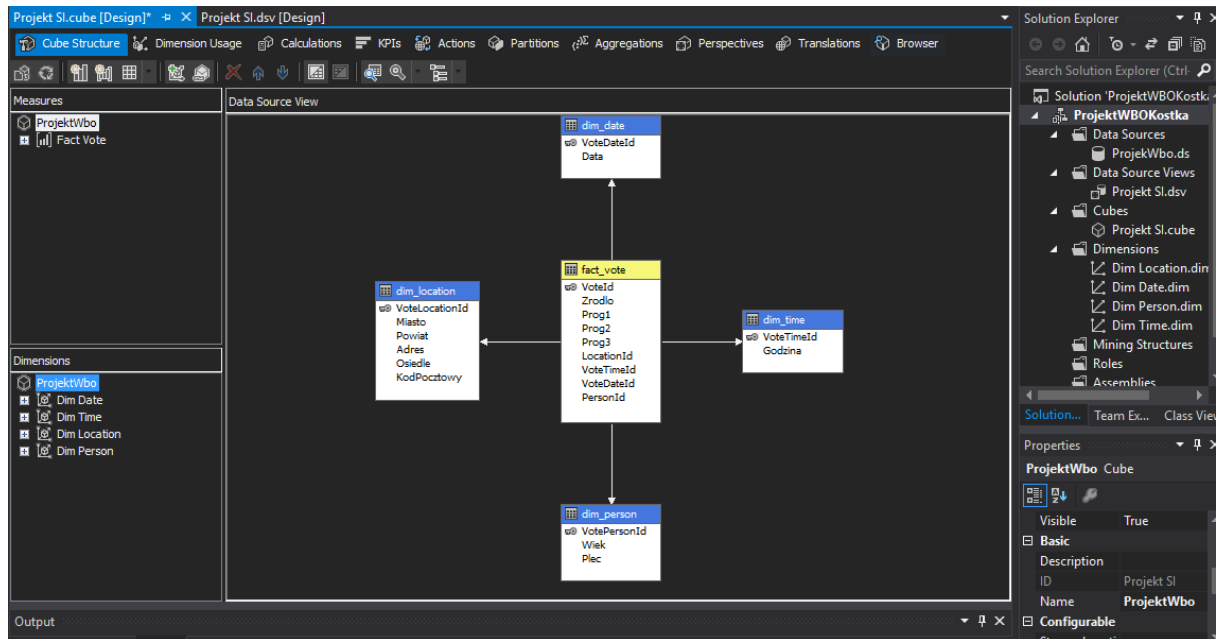
dim_person – zawiera informacje o płci oraz wieku głosującego

dim_time – zawiera informacje o godzinie głosowania

dim_date – zawiera informacje o dacie głosowania

dim_location – zawiera informacje o lokalizacji głosu

Na podstawie schematu i przy wykorzystaniu narzędzia Visual Studio, stworzyłem projekt kostki.



Posłuży ona do wyszczególnienia następujących informacji:

- jaki wiek głosujących dominuje ?
- które osiedla są najaktywniejsze ?
- jakie miasta spoza Wrocławia brały udział w WBO ?
- Które godziny sygnalizują wzmożony ruch w głosowaniu ?
- Która data i jednocześnie jaki dzień tygodnia sygnalizuje najaktywniejszy dzień głosowania ?
- Kto chętniej głosuje, kobiety czy mężczyźni ?

Proponowane tematy badań będą rozszerzał o kolejne pomysły.
Chciałbym także do tego wykorzystać informacje o projektach zwycięskich.

Podsumowując:

W przyszłym tygodniu chciałbym pokazać wstępne raporty oraz wyliczenia oraz chciałbym dopracować jeszcze dane, znajdujące się aktualnie w schemacie hurtowni.