*Од членовите на тимот,*

**10.11.2021**

Опис на документот

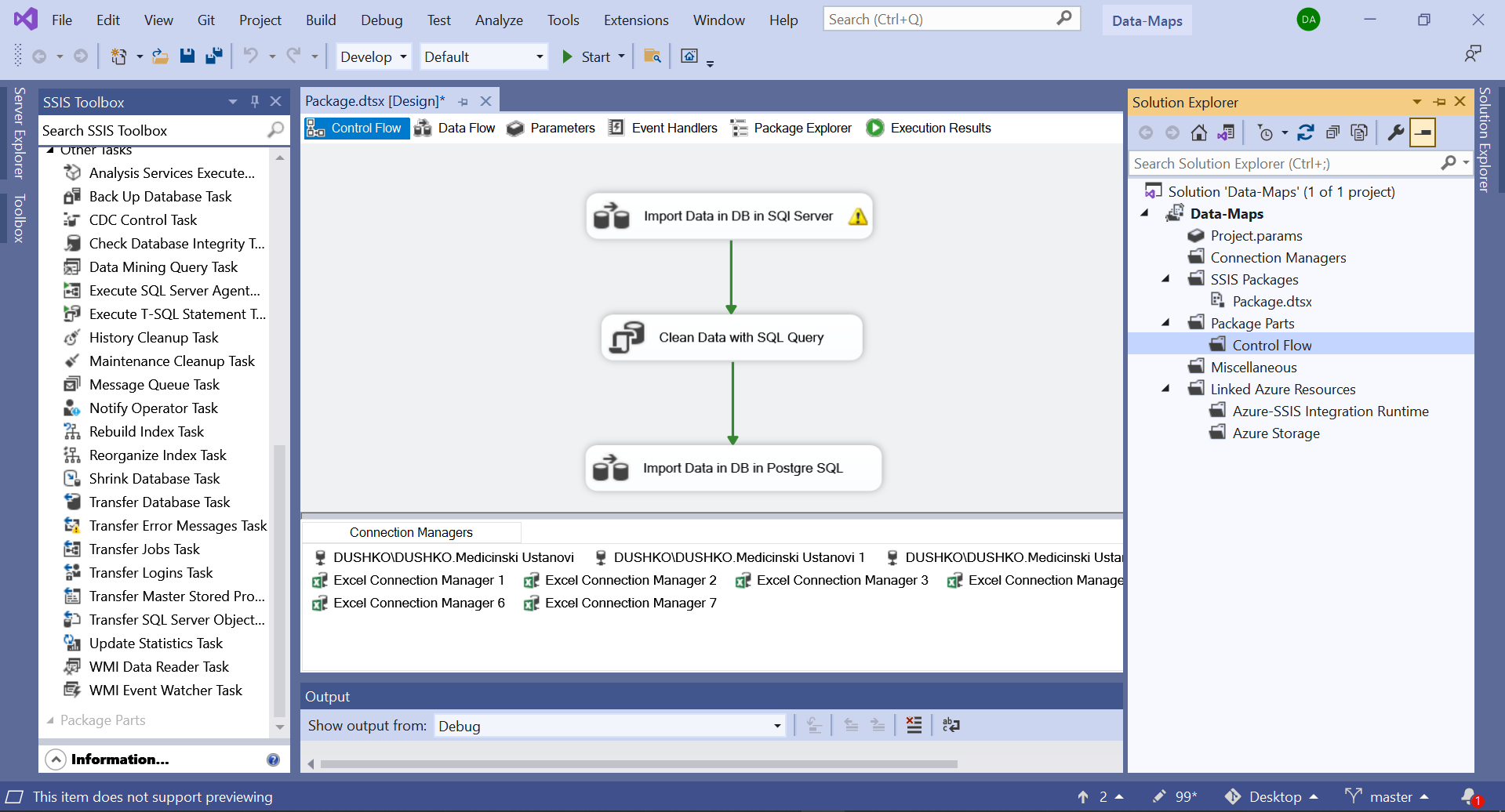
Опис на pipe - and - filter архитектурата за филтрирање и внес на податоците од изворот до дестинациската база на податоци на SQL Server

Домашна работа бр.1

*Дизајн и архитектура на софтвер*

Целиот процес на пренос и филтер на податоците од изворот (Excel) до креираната база на податоци (на SQL Server), успеавме да го постигнеме со помош на SSIS или SQL Server Integration Services, ETL алатката, во Visual Studio 2019.

Во рамки на самиот SSIS проект дефиниравме SSIS Package, кој ќе го содржи главниот Control Flow на pipe - and - filter архитектурата (слика 1).



*Слика 1. Control Flow, со сите таскови во SSIS*

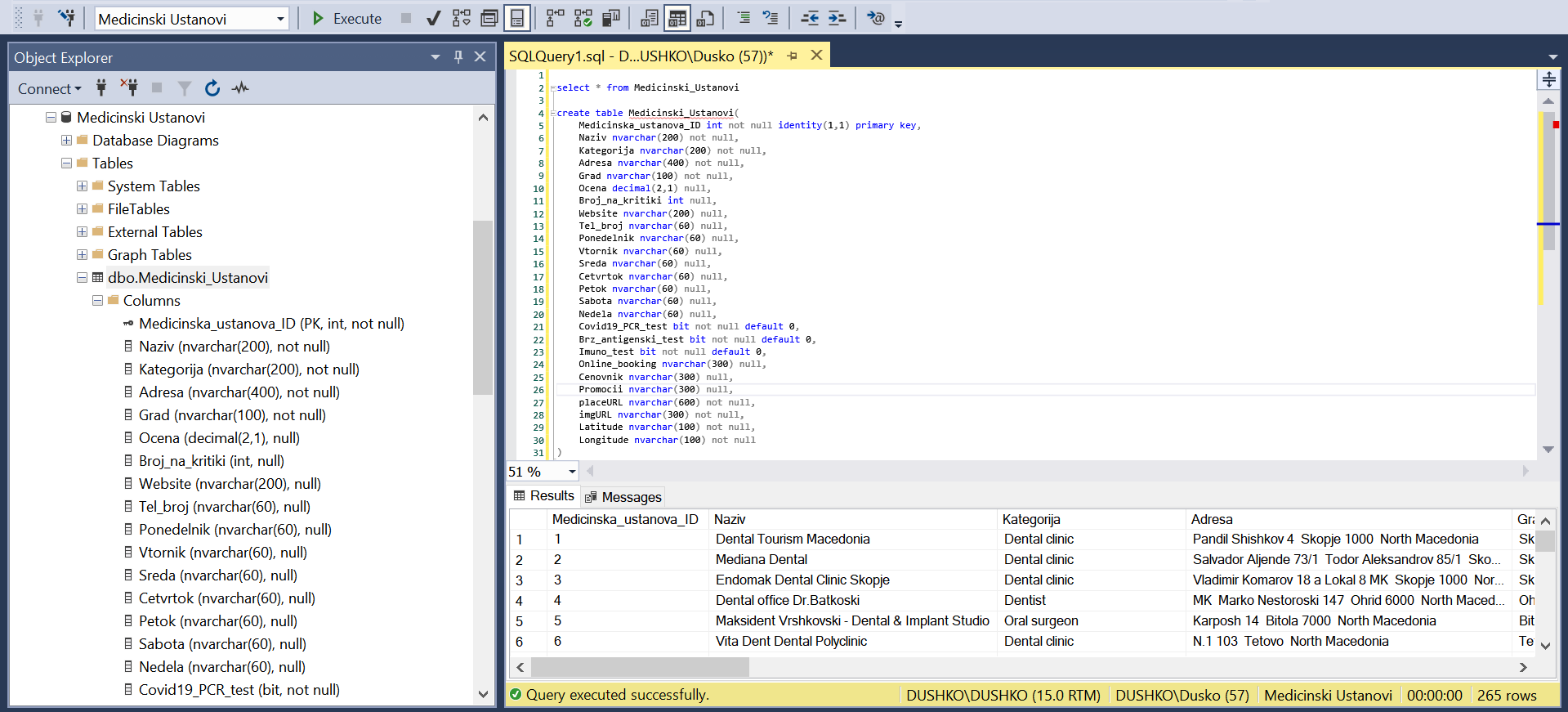
Делот од Control Flow го прикажува **редоследот** на извршување на сите таскови со цел да го добиеме бараното импортирање на податоците во база, додека **начинот** како ќе се изврши полнењето на табелите го прикажуваат самите таскови, поделени по улоги. Со секој таск, во SSIS, креираме посебен Data Flow каде ќе се наоѓаат сите филтри на податоците.

Таскот “Import Data in DB in SQL Server”, како што кажува името, ја полни креираната табела Medicinski\_Ustanovi креирана во базата Medicinski Ustanovi (слика 2).

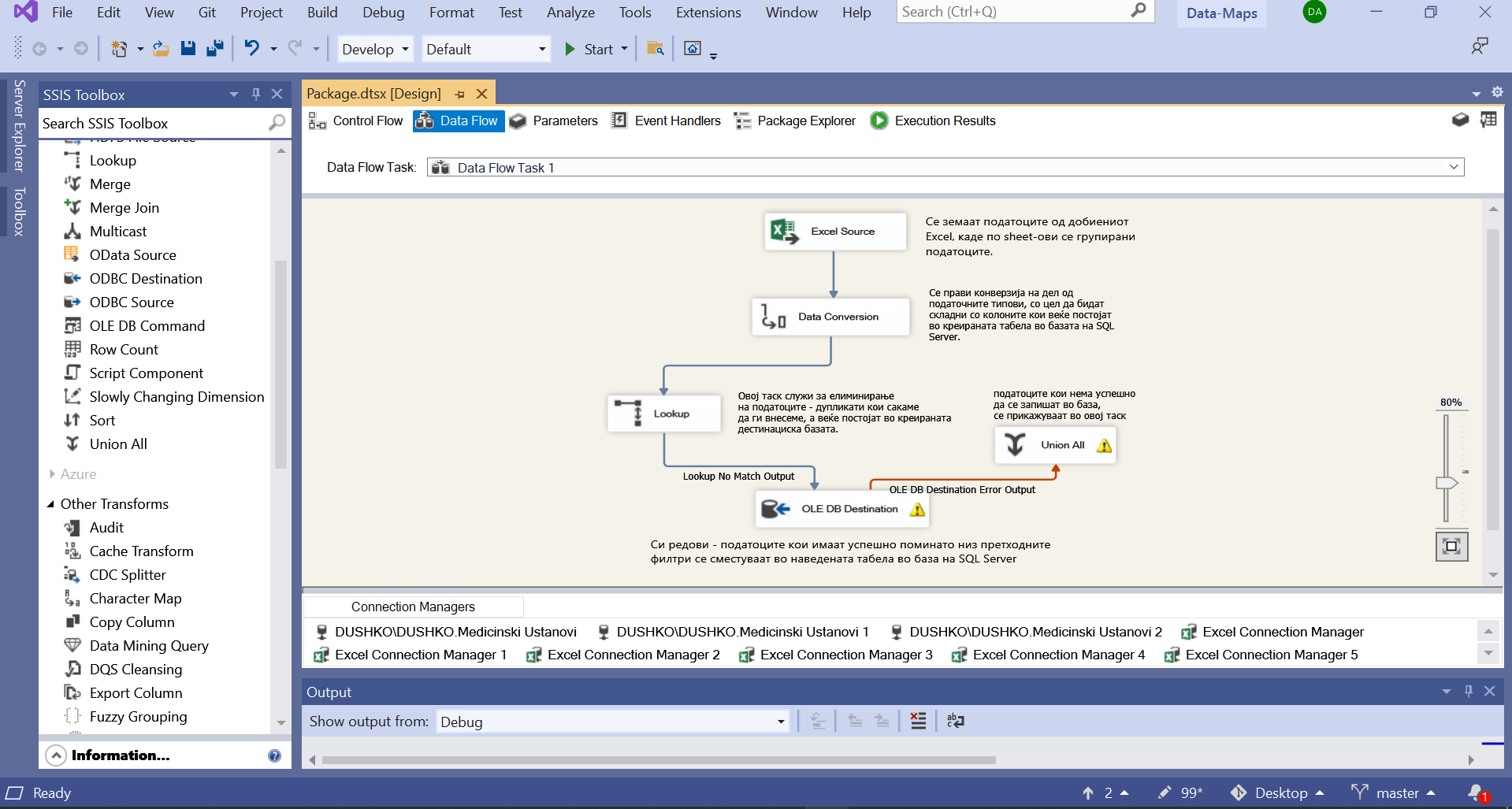
Изворот на податоците (Excel) го добивме преку Web Scraping, со помош на веб апликацијата “Phantom Buster”, каде со помош на наши филтри успеавме да добиеме сет на податоци групиран по категории (sheet - ови) кои ни беа од интерес.

Самите филтри во “Import Data in DB in SQL Server” таскот, за полнење на табелата во дестинацијата се прикажани на слика 3.

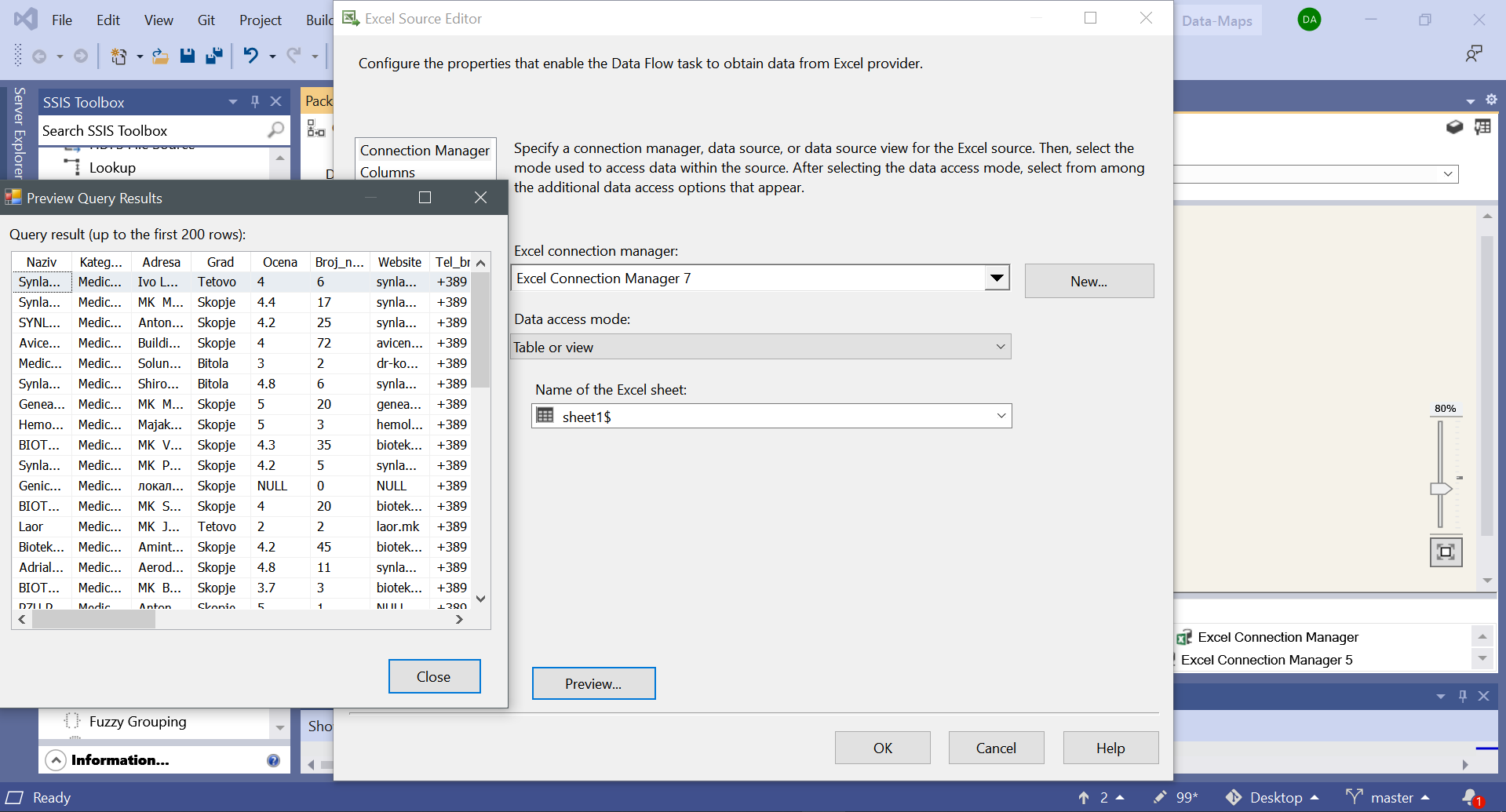
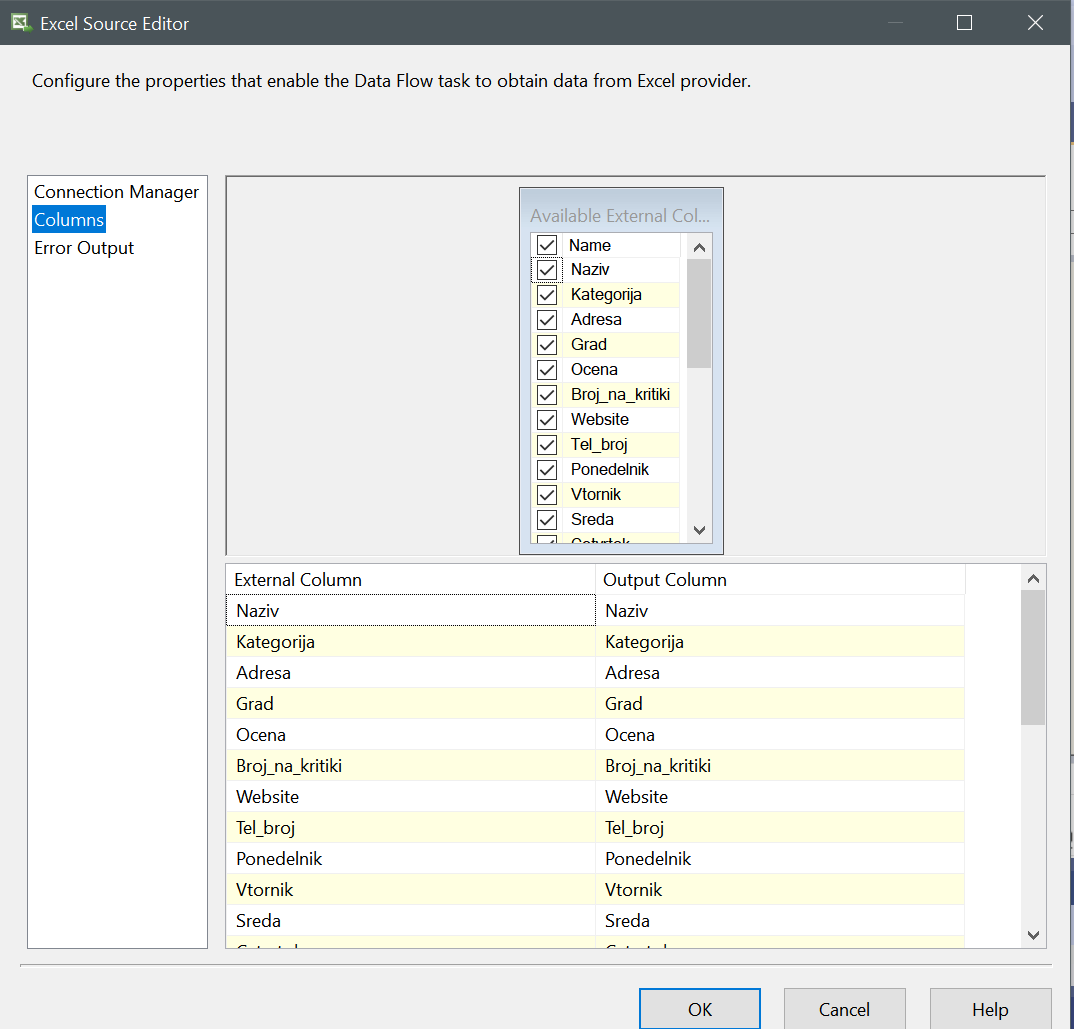
Во “Excel Source” филтерот ги поставуваме сите колони кои сакаме да ги запишеме во дестинацијата, во овој дел имаме можност за одбирање на Excel file како и селектирање на посакуваниот sheet, заедно со негов preview (слика 4).



*Слика 2. Креираната табела Medicinski\_Ustanovi на SQL Server во SSMS*



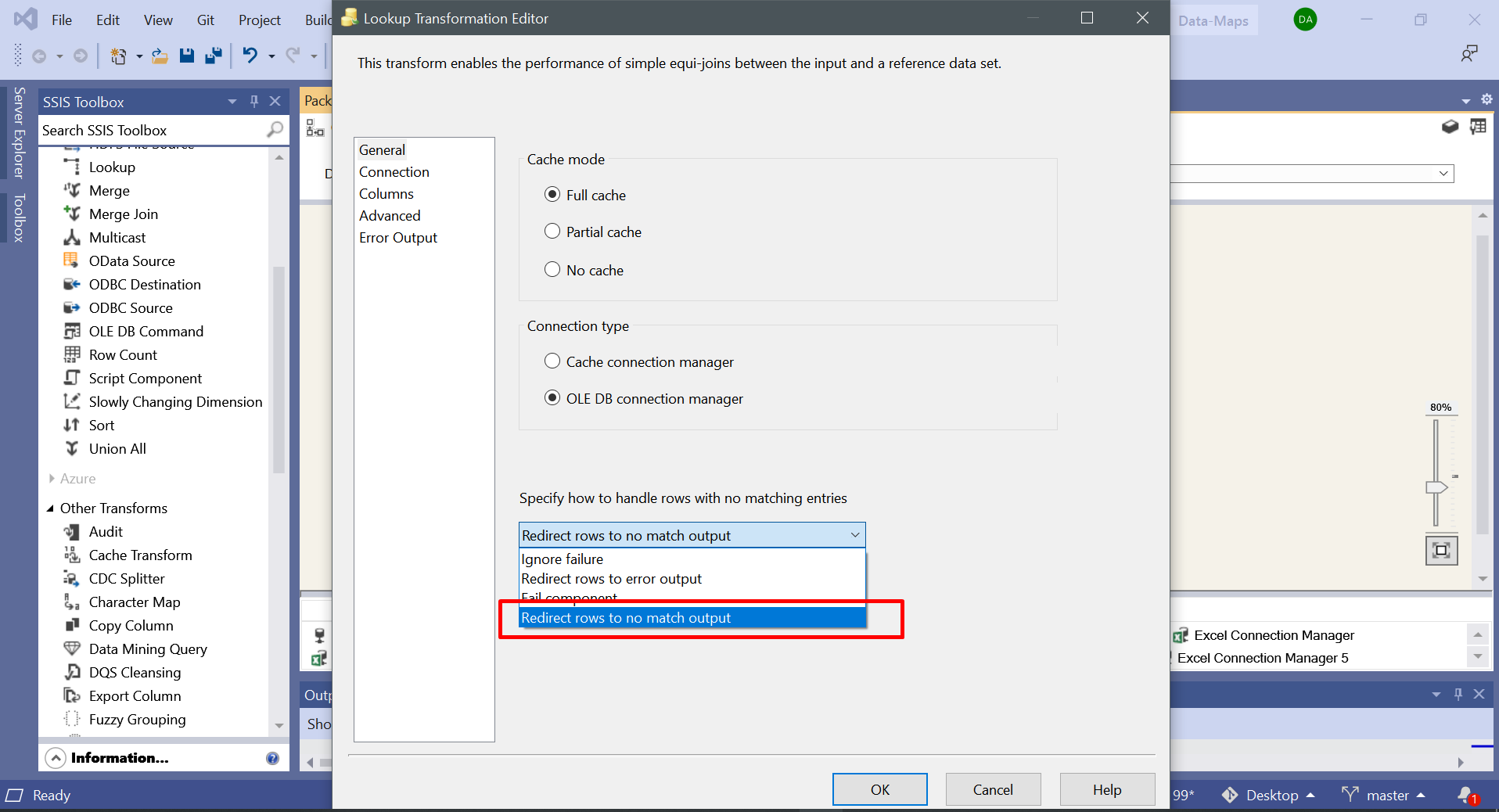
*Слика 3. Филтри на трансофрмација на податоците од изворот до дестинацијата*

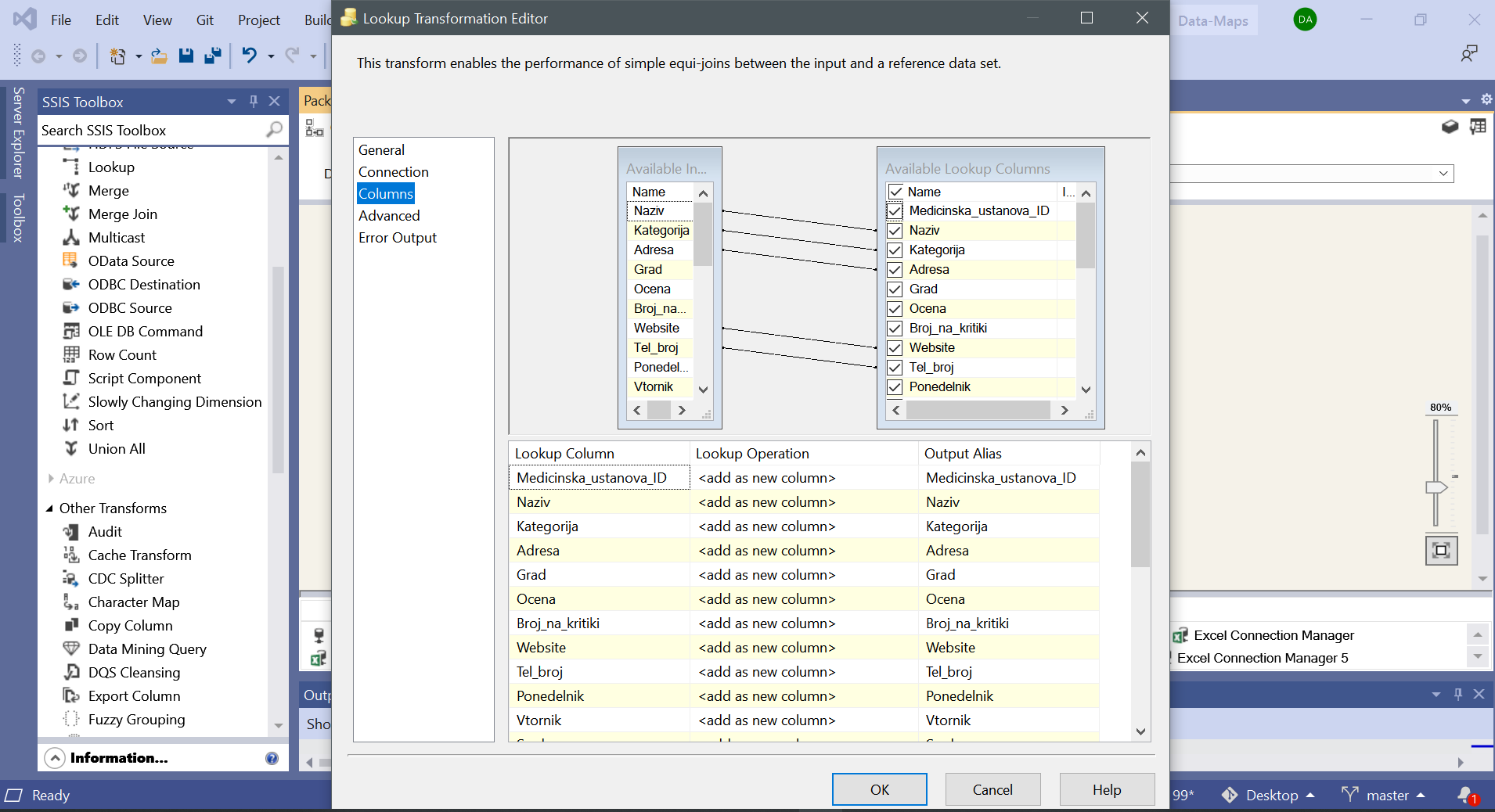


*Слика 4. Excel Source филтер*

Сите филтрирани податоци продолжуваат во следниот “Data Conversion” филтер, каде се прави конверзија на дел од податочните типови од изворот, со цел да бидат складни со колоните кои веќе постојат во креираната табела - дестинација.

Во “Lookup” филтерот, успеавме да постигнеме ефективен начин на импорт на податоците во дестинацијата, со тоа што поставуваме “Lookup No Match Output” за елиминирање на сите дупликати - редови од изворот на податоците. Во овој дел ги поврзуваме колоните од изворот со колоните од дестинацијата по кои сакаме да правиме елеминирање на дупликат податоци (слика 5).

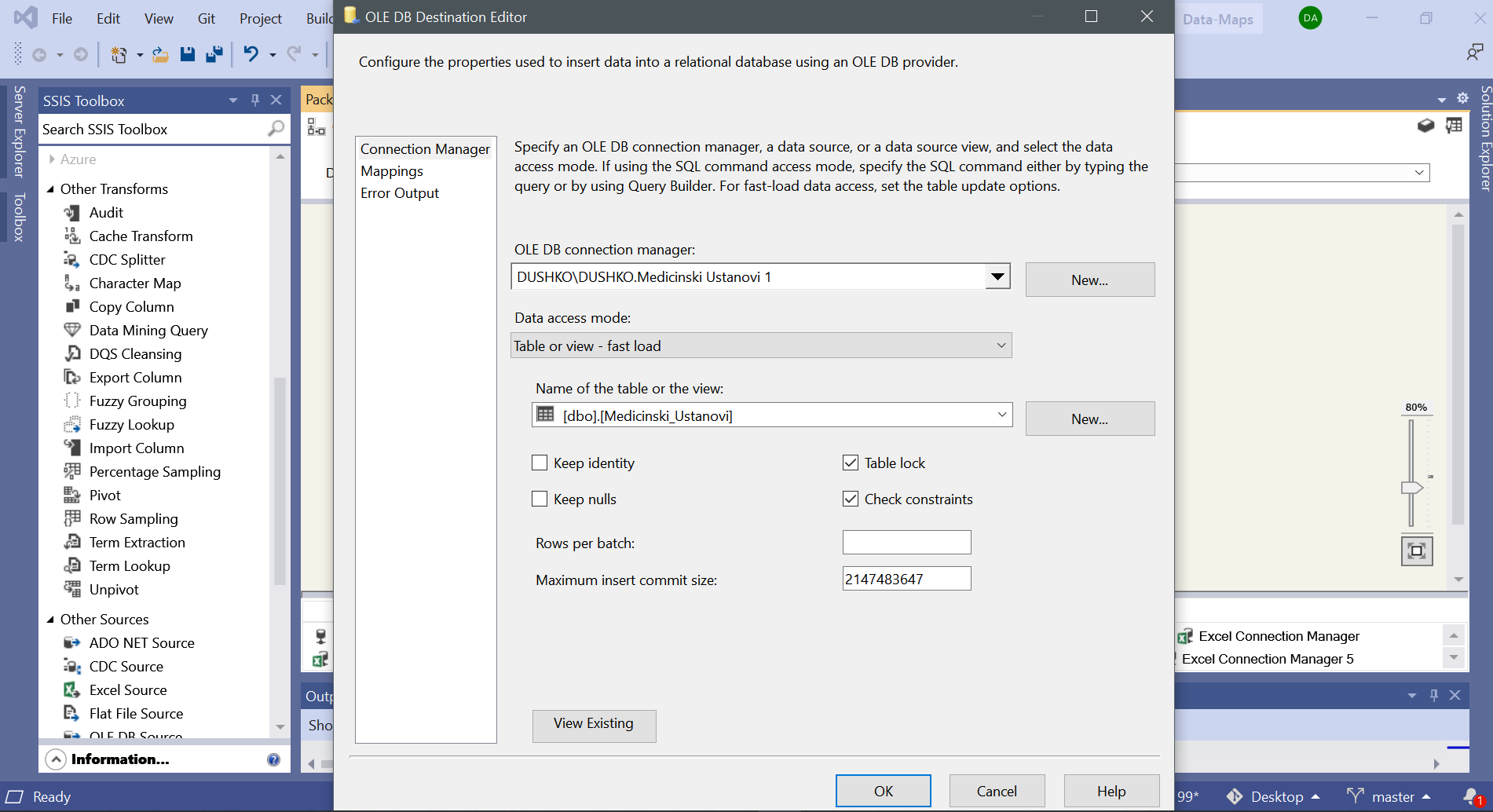




*Слика 5. Селектирање на колони за филтер*

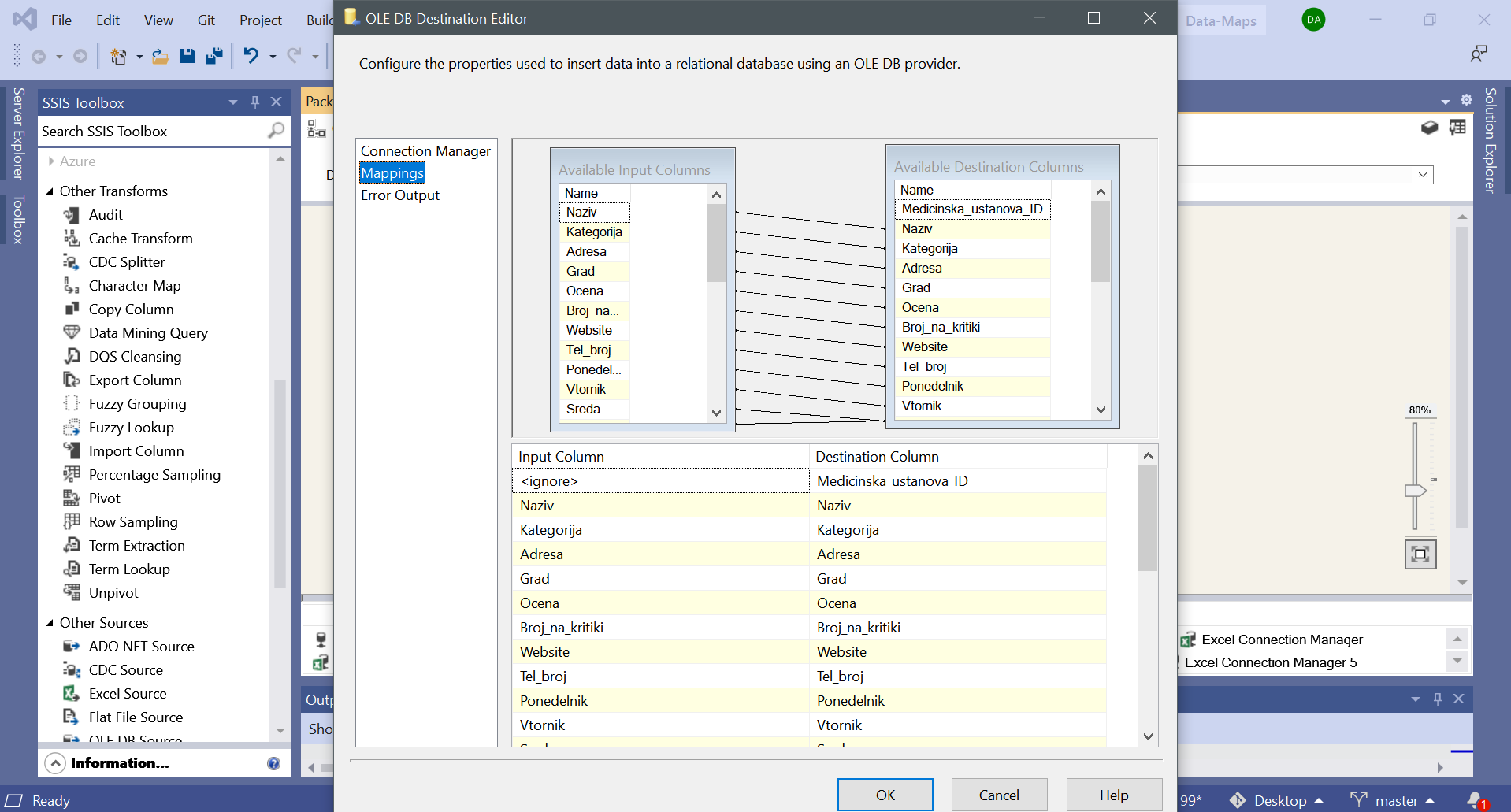
Покрај селектирањето на колони, во овој дел мора да поставиме и конекција до SQL Server-от со цел да ја добиеме посакуваната база како и дестинациската табелата во неа.

Истото конектирање е потребно да се направи и во “OLE DB Destination”. Во овој дел ги селектираме и полињата за филтер и елиминирање на NULL вредности (слика 6).



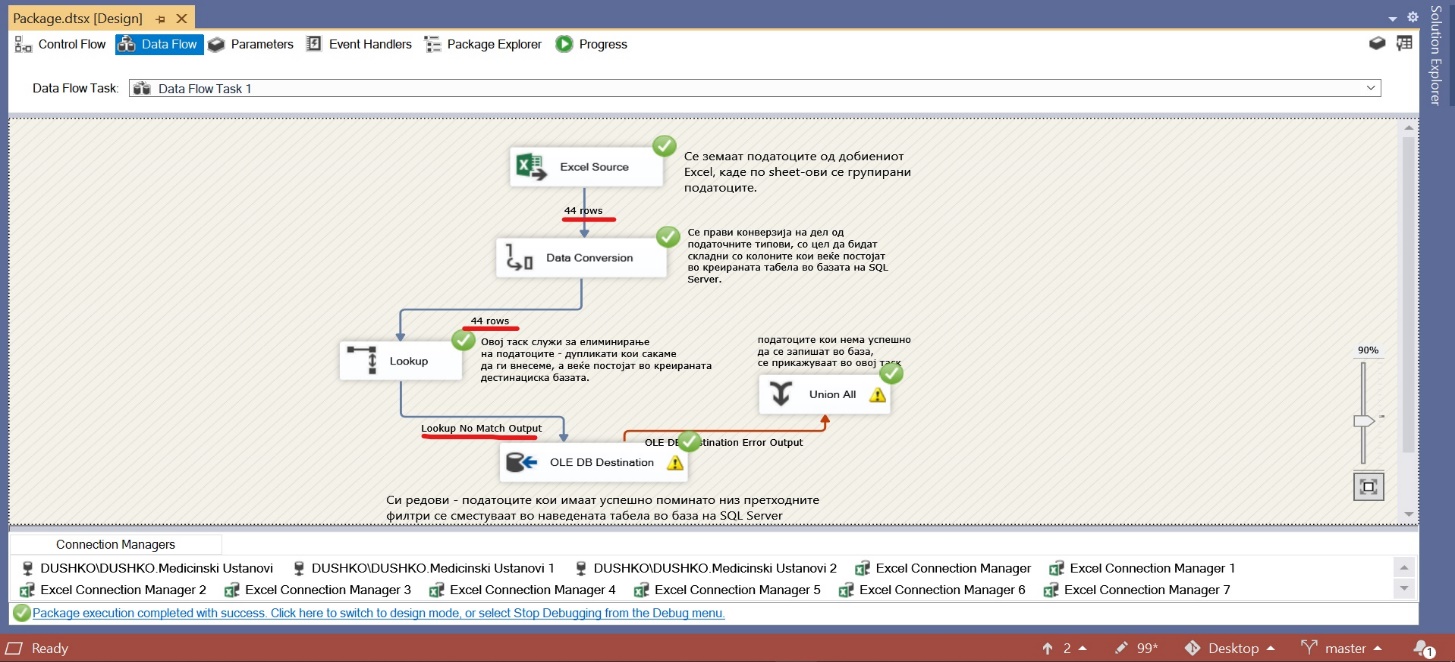
*Слика 6. Кокекција до база на сервер и елиминирање на NULL вредности*

Со цел SSIS да препознае кои колони од изворот сакаме да ги мапираме во кои колони од дестинациската табела во “OLE DB Destination” ги поврзуваме изворните колони со дестинациските, каде мора изворната колона да има ист податолен тип како и онаа во дестинациската табела (слика 7).



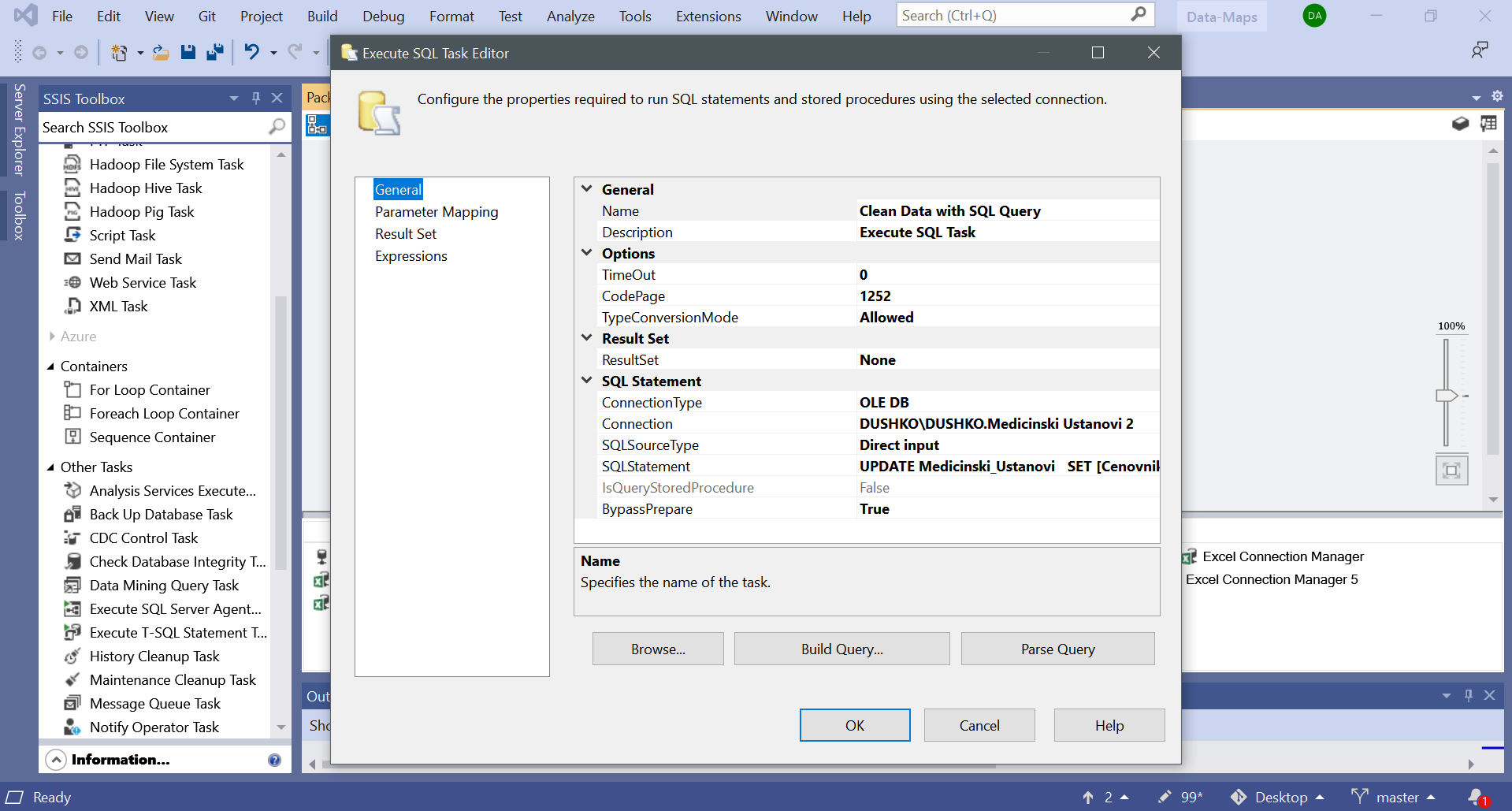
*Слика 7. Мапирање на извор со дестинација*

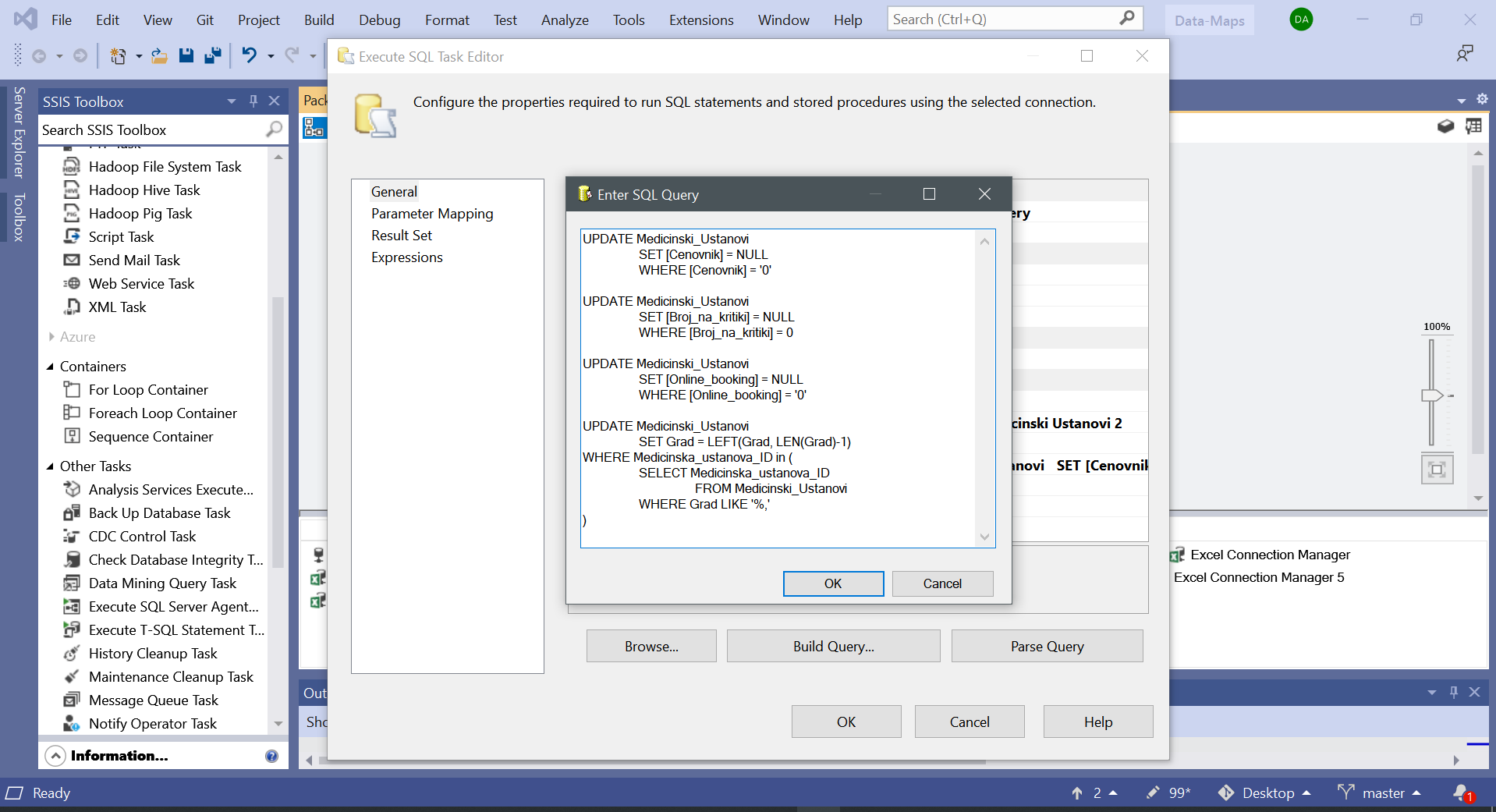
Успешното извршување на овој таск е прикажан на слика 8, но откако е веќе наполнета дестинациската табела, тука може да се види дека “Lookup” филтерот не ги “пропушта” редовите кои веќе постојат во дестинацијата.



*Слика 8. Успешно извршен таск*

Откако ќе се наполни табелата во дестинацијата останува уште да направиме филтер на внесените податоци со помош на SQL Query напишано во SSIS. Во овој дел правиме чистење на постоечките податоци во SQL Server. Таскот кој го овозможува ова е “Clean Data with SQL Query” прикажан на слика 1, додека самото query e прикажано на слика 9.





*Слика 9. Query за чистење и средување на податоците*