

**PROJEKTOVANJE SKLADIŠTA PODATAKA**

NHL GAME DATA

**Enio Kurteši IT27/2019**

**Novi Sad, 2023.**

**SADRŽAJ**

[ZADATAK I CILJEVI PROJEKTA 2](#_bookmark0)

[OPIS POSTUPKA KREIRANJA DW SISTEMA 3](#_bookmark1)

[SPECIFIKACIJA ZAHTEVA KORISNIKA 4](#_bookmark2)

[SPECIFIKACIJA MODELA 5](#_bookmark3)

[OPIS ETL PROCESA 12](#_bookmark4)

[PRIKAZ IZVEŠTAJA KAO ODGOVOR NA ZAHTEVE KOJI SU POSTAVLJENI OD STRANE](#_bookmark5) [KORISNIKA 26](#_bookmark5)

[ZAKLJUČAK 29](#_bookmark6)

# ZADATAK I CILJEVI PROJEKTA

Osnovni zadatak projekta je kreiranje Data warehouse sistema namenjenog za pohranjivanje podataka koji se tiču odigranih utakmica u američkoj hokejaškoj ligi(NHL).Kasnije će se na osnovu napravljenog skladišta odgovoriti na korisničke zahteve. Izveštaji će biti interaktivni i omogućiće uvid u praćenje izabranih statističkih parametara na nivou jedne utakmice.

Oblast odabranog projekta obuhvata veoma detaljne podatke vezane za odigrane utakmice izmedju timova u NHL ligi. Pasionirani ljubitelji sportova vole da prate podatke vezano za performanse njihovih timova kao što su: broj golova, šuteva, timska odbrana, načinjeni faulovi... Uz pomoć algoritama mašinskog učenja i istorijskih podataka o utakmicama određenih timova, možemo pokušati i da previdimo rezultate nekih narednih utakmica.

U nastavku dokumentacije biće detaljno objašnjen postupak kreiranja skladišta podataka,počevši od samog izbora dataset-a,kreiranja odgovarajuće OLAP šeme na osnovu koje će kasnije i biti implementirano skladište podataka,kao i sam ETL postupak koji za cilj ima transformaciju i pročišćavanje izvornih podataka i njihovo punjenje u kreirano skladište podataka. Na kraju,biće dat grafički prikaz i objašenjenje kreiranih izveštaja,koji predstavljaju odgovore na specificirane korisničke zahteve.

# OPIS POSTUPKA KREIRANJA DW SISTEMA

Postupak projektovanja DW sistema započinje pretraživanjem i pronalaženjem odgovarajućeg seta podataka na Internetu. Dataset koji je analiziran je javno dostupan na sajtu kaggle.com .

Nakon projektovanja OLAP šeme, gde se definišu tabele dimenzija, činjenična tabela, veze između tih tabela i atributi tabela, kreiraju se DDL naredbe za tabele dimenzija i za činjeničnu tabelu, kao i sekvence za potrebe primarnih ključeva, i kreiraju se veze između tabela. Dimenzije koje su identifikovane na osnovu izvornog seta podataka su:Home team,Away team,Venue,Official,Season,Date dimension koja predstavlja obaveznu dimenziju u svakom skladištu podataka. Takodje,kreira se skripta za vremensku dimenziju, koja odgovara vremenskoj dimenziji u OLAP šemi. Nakon toga kreira se ETL proces, u kom se vrše određene transformacije nad podacima,te se nakon toga kreirane tabele popunjavaju transformisanim,prečišćenim podacima. Na kraju, kreiraju se izveštaji na osnovu kreiranog skladišta podataka i korisničkih zahteva, čime se dobija uvid u podatke na interaktivan način putem izveštaja.

# SPECIFIKACIJA ZAHTEVA KORISNIKA

Zahtevi korisnika predstavljeni su u obliku pitanja, a na te zahteve, odnosno pitanja biće odgovoreno kasnije,nakon detaljne analize, putem izveštaja koji će biti kreirani. Treba napomenuti da izveštaji naravno ne daju isključivo jedan odgovor kojim će biti odgovoreno na konkretna postavljena pitanja,već se korišćenjem kreiranih izveštaja može odgovoriti na mnoga pitanja.

Spisak zahteva:

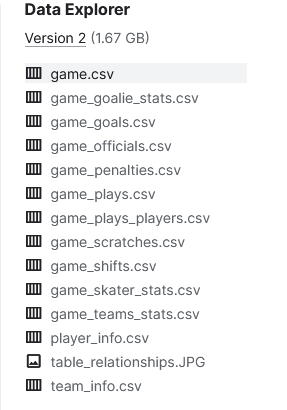
1. Koji timovi su u top 5 po broju postignutih golova kod kuće u prvoj deceniji 21. veka?
2. Koji timovi su u top 5 po broju postignutih golova u gostima u prvoj deceniji 21. veka?
3. Kako su se utakmice završavale izmedju 2008. i 2012.? Koliki je odnos golova i šuteva?
4. Koji timovi imaju najgoru timsku odbranu, grupisano po konferencijama kojima pripadaju?
5. Koji su bili rezultati u prvom kolu u datoj sezoni?
6. Koliki je broj postignutih golova domaćeg tima,broj šuteva, faulova, izgubljenih i oduzetih lopti u sezoni, kao i prosečne vrednosti ovih parametara za timove grupisane po konferencijama?
7. Koje su maksimalne vrednosti parametara na jednom meču, posmatrajući domaći tim u sezoni 2012?
8. Koji sudija je dosudio najviše faulova u poslednje 3 sezone?

**SPECIFIKACIJA MODELA**

1. **Specifikacija izvora podataka**

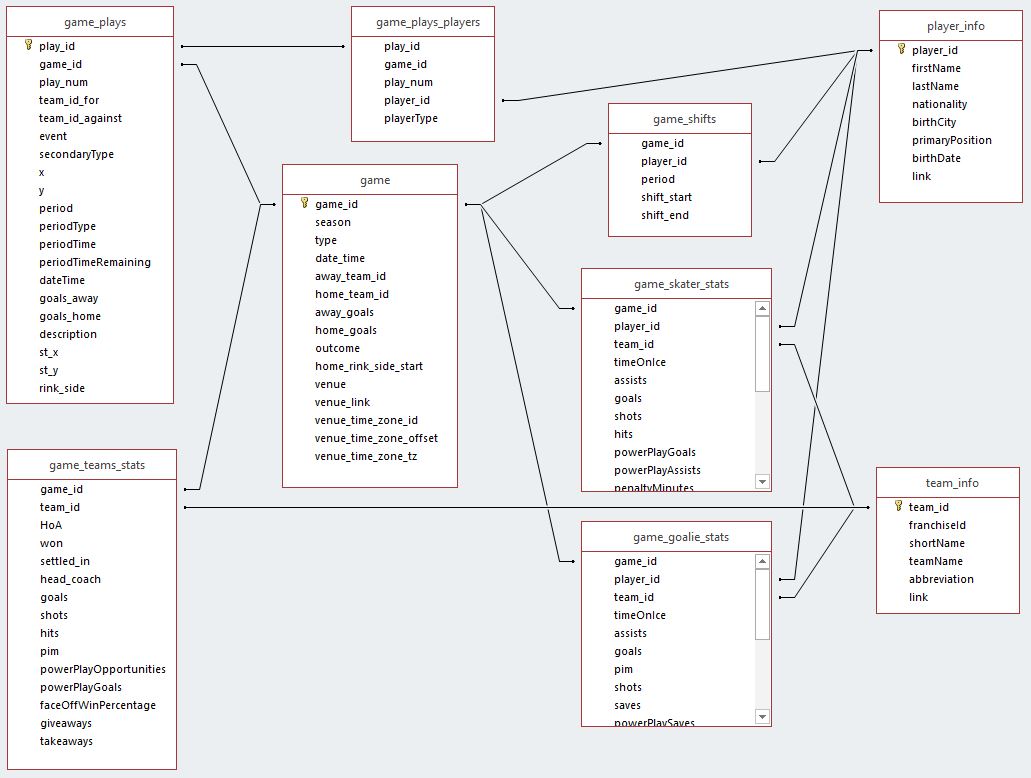
Primarni izvor podataka predstavlja set podataka pod nazivom NHL Game Data, koji se nalazi na sajtu Kaggle. Link ka izvoru podataka: https://www.kaggle.com/datasets/martinellis/nhl-game-data

Izvorni dataset sadrži tabele u csv. formatu. Reč je o izuzetno glomaznom dataset-u koji se sastoji od 13 tabela. Bilo je neophodno upoznati se sa samim dataset-om, brojnim atributima koji mi nisu bili poznati, kako bi se na ispravan način odlučilo šta treba izostaviti iz analize, a šta je pogodno za kreiranje skladišta podataka. Iz ovog dataset-a moguće je izdvojiti nekoliko Data Mart-ova. Odlučeno je da fokus bude na analizi utakmica, a da se tabele vezane za igrače ignorišu.



*Slika 1: Izvorni set podataka*

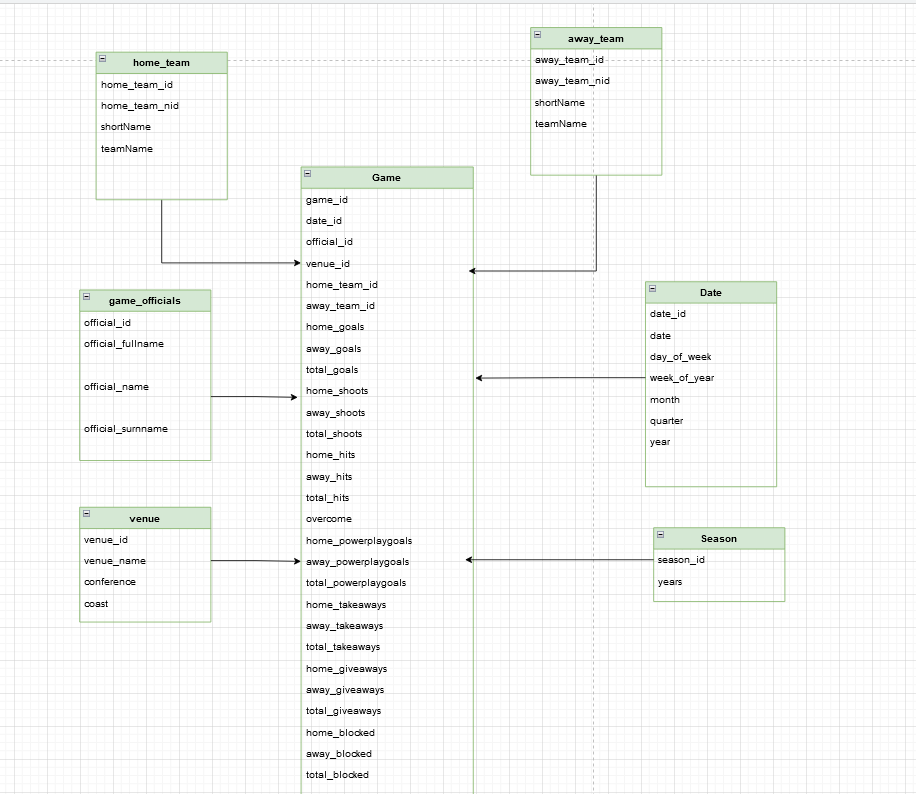
Prikaz dijagrama klasa,odnosno OLTP šeme:



*Slika 2:OLTP šema*

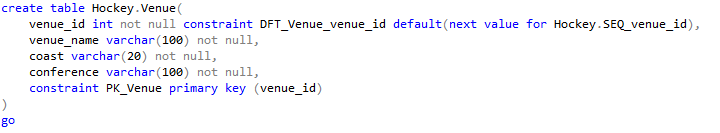
1. **Specifikacija ciljnog DataWarehouse sistema**

Na osnovu OLTP šeme podataka, korisničkih zahteva i procesa poslovanja, kreirana je OLAP šema. Atributi vezani za samu odigranu utakmicu definisani su kao mere u okviru tabele činjenica, dok su tabele koje sadrže podatke koji se ponavljaju, i vezani su za veći broj utakmica, izdvojene kao tabele dimenzija. Na slici 3 prikazana je zvezda šema koja je kreirana na osnovu identifikovanih tabela dimenzija i tabele činjenice.



*Slika 3: OLAP šema*

* 1. **Specifikacija zahtevanih dimenzija**



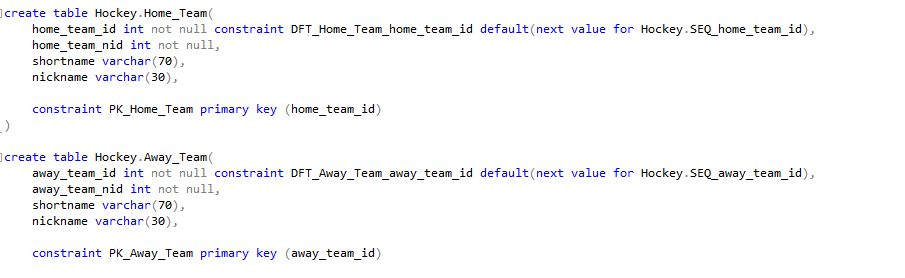
Dimenzija Venue sastoji se od četiri obeležja:

1. venue\_id – veštački ključ koji se kreira preko sekvence

2. venue\_name – ime dvorane u kojoj se igraju mečevi

3. coast – deo SAD gde se nalazi dvorana

4. conference – država, odnosno grad u kom se nalazi dvorana



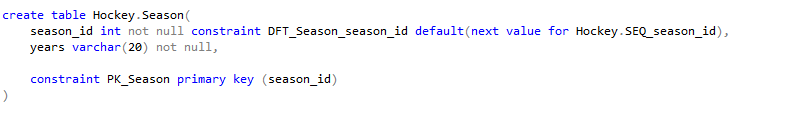
Dimenzije Home team i Away team imaju ista obeležja:

1. home\_team\_id – veštački ključ kreiran preko sekvence

2. home\_team\_nid – prirodni ključ koji dolazi iz OLTP

3.shortname – ime tima

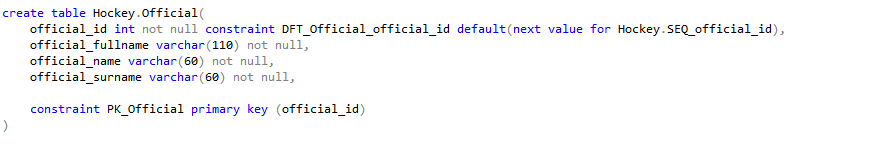
4. nadimak tima



Dimenzija Season ima dva obeležja:

1. season\_id – veštački ključ kreiran preko sekvence

2. years-sezone u godinama u formatu 2016/2017



Dimenzija Official ima četiri obeležja:

1.official\_id – veštački ključ kreiran preko sekvence

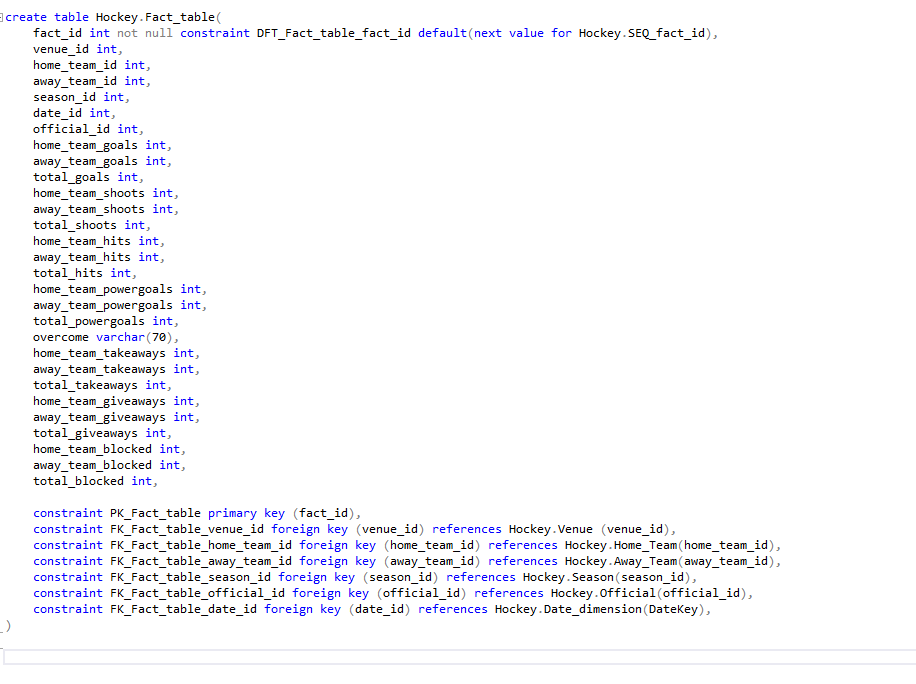
2. official\_fullname – ime i prezime sudije

3. official\_name – ime sudije

4. official\_surname – prezime sudije

Dimenzija Date je preuzeta iz javno dostupnog skripta za kreiranje datuma i ona se nalazi u DDL fajlu koji će biti predat u prilogu.

* 1. **Specifikacija zahtevanih mera**



Činjeničnu tabelu predstavlja izvorna tabela Game koja je uz određene modifikacije pretvorena u tabelu činjenica. Činjenična tabela sadrži mere relevantne za praćenje jedne odigrane utakmice. U njoj se nalaze veštački ključevi tabela dimenzija koji je okružuju kao i mere koje su interesantne za praćenje. Mere od interesa su sledeće:

1. home\_team\_goals – broj postignutih golova domaćina na meču

2. away\_team\_goals – broj postignutih golova gosta na meču

3. total\_goals – ukupan broj golova na meču

4. home\_team\_shoots – broj šuteva na meču od strane domaćeg tima

5. away\_team\_shoots – broj šuteva na meču od strane gostujućeg tima

6. total\_shoots – ukupan broj šuteva na meču

7. home\_hits – broj faulova domaćeg tima na meču

8. away\_hits – broj faulova gostujućeg tima na meču

9. total\_hits – ukupan broj faulova na meču

10.home\_team\_powergoals – broj golova domaćeg tima kad ima prednost u broju igrača u polju

11. away\_team\_powergoals – broj golova gostujućeg tima kad ima prednost u broju igrača u polju

12.total\_powerplaygoals – ukupan broj takvih golova

13. overcome – način na koji se završila utakmica. Opcije: pobeda domaćina regularno, pobeda domaćina u produžecima, pobeda gosta regularno, pobeda gosta u produžecima. U hokeju nerešen rezultat ne postoji.

14. home\_team\_takeaways – broj izgubljenih lopti na nivou domaćeg tima

15. away\_team\_takeaways – broj izgubljenih lopti na nivou domaćeg tima

16. total\_takeaways – ukupan broj izgubljenih lopti na meču

17.home\_team\_giveaways – broj ukradenih lopti od strane domaćeg tima

18. away\_team\_giveaways – broj ukradenih lopti od strane gostujućeg tima

19. total\_giveaways – ukupan broj ukradenih lopti

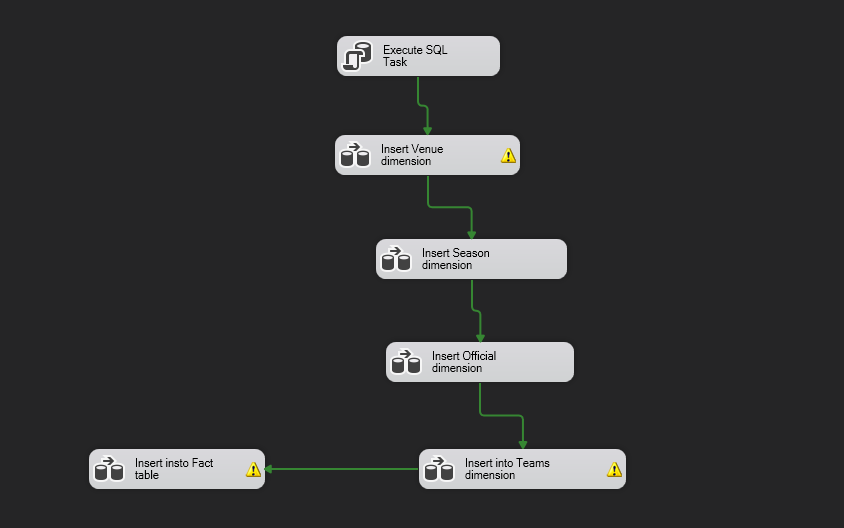
20. home\_team\_blocked – blokirani udarci domaćeg tima

21. away\_team\_blocked – blokirani udarci gostujućeg tima

22. total\_blocked – ukupan broj blokiranih udaraca

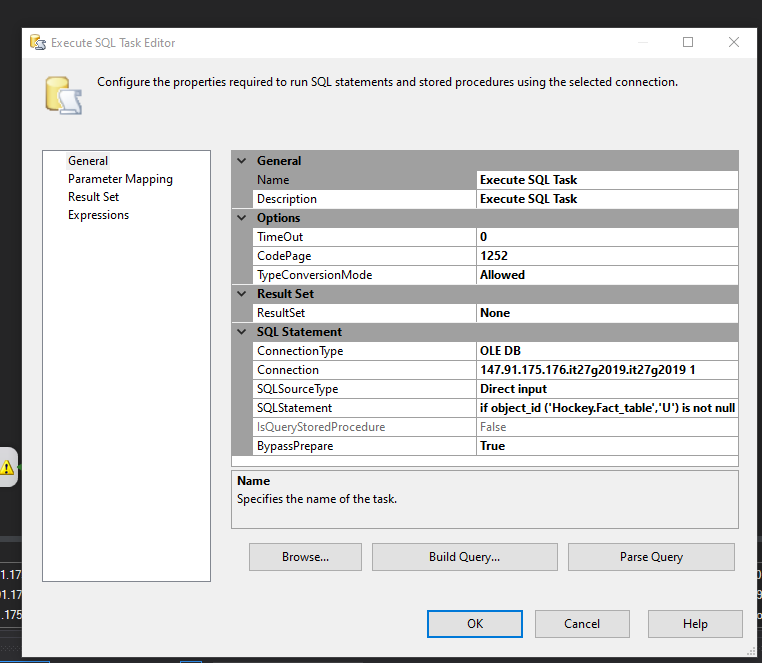
# OPIS ETL PROCESA

Za realizaciju ETL procesa korišćen je Visual Studio 2019, Microsoft SQL Server Managment Studio kao i poseban paket pod nazivom Integration Services Project.



*Slika 4: Etl proces*

Prvi korak u kreiranju ETL procesa jeste kreiranje tabela dimenzija koristeći DDL naredbe, konkretno pomoću create table naredbe, i kreiranje sekvenci za svaku od kreiranih tabela korišćenjem create sequence naredbe. Korišćenjem komponente Execute SQL Task, podešena je konekcija na server u kome je smeštena studentska baza. Nakon toga, u kartici SQL statement kopiran je SQL kod za kreiranje svih neophodnih tabela. Kao što se može videti na slici 4, dimenzije i tabela činjenica se pune jedna nakon druge. Prvo su izabrane dimenzije koje nisu bile kompleksne za izvršavanje neophodnih transformacija pre punjenja u željenu destinaciju.

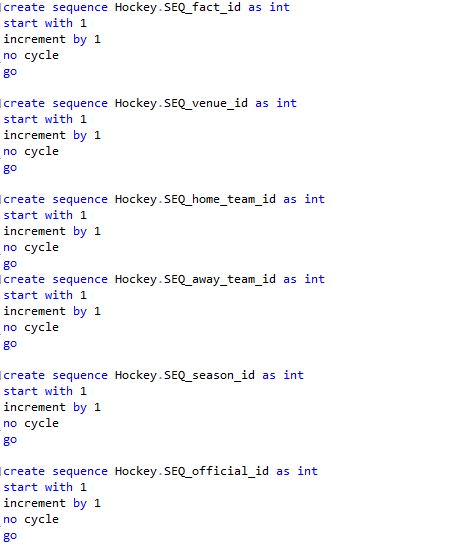


*Slika 5: Execute SQL Task*

Na slici 6 prikazano je uslovno brisanje. Svaki put kada se pokrene Execute SQL Task izvršiće se brisanje postojećih tabela i sekvenci i kreiranje novih.

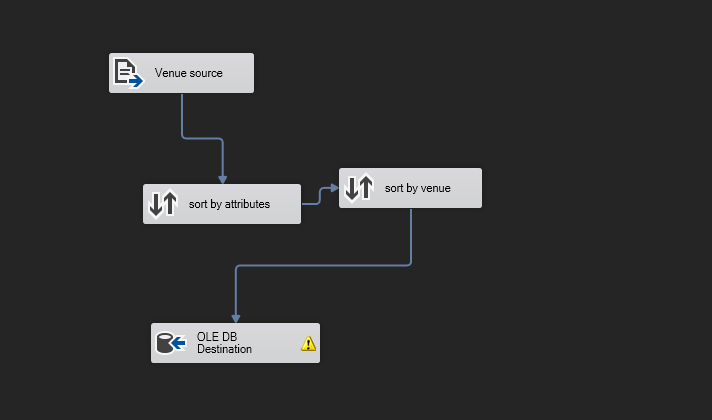


*Slika 6: Uslovno brisanje objekata*

**

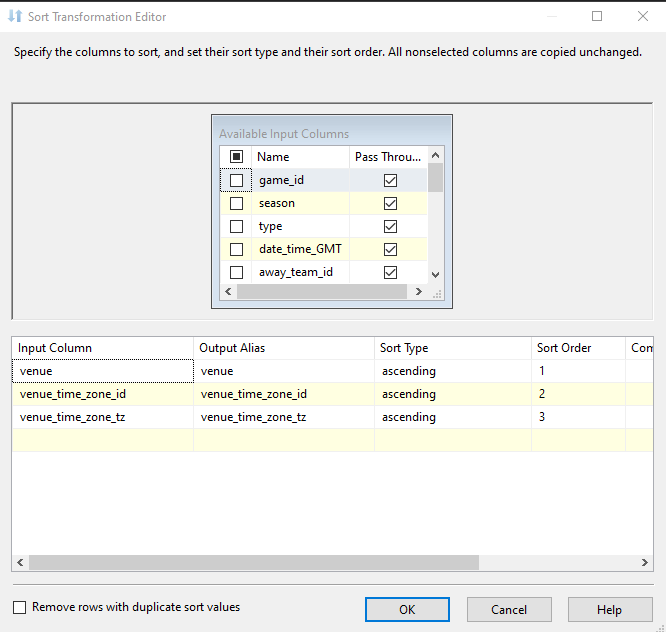
*Slika 7: Kreiranje sekvenci*

Data flow se sastoji iz nekoliko komponenti. Imamo izvor podataka, koji se može nalaziti u različitim formatima poput Flat file source, Ole DB source, XML source... Nakon što učitamo podatke, uz pomoć različitih alata za transformaciju možemo modifikikovati, prečistiti i izdvojiti željene podatke koji se obično na kraju upisuju na krajnju destinaciju. Destinacija može biti tekstualni fajl, tabela u bazi podataka, excel fajl...



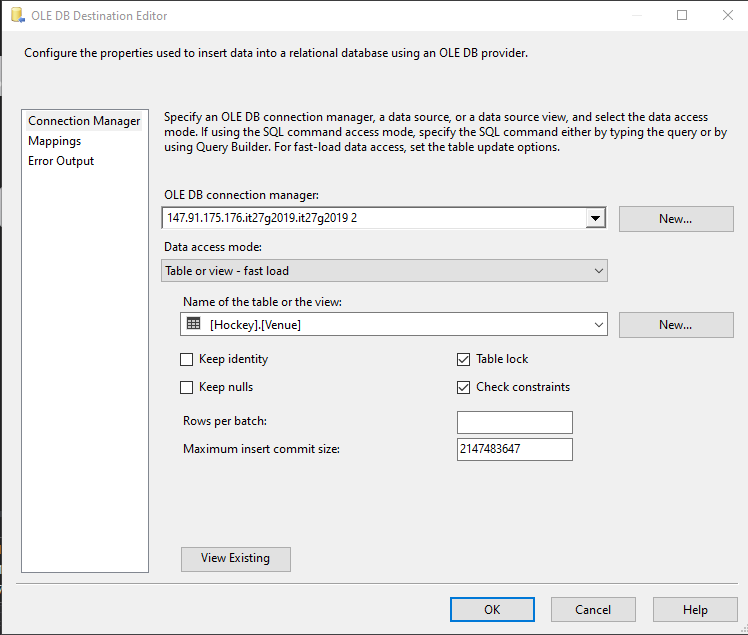
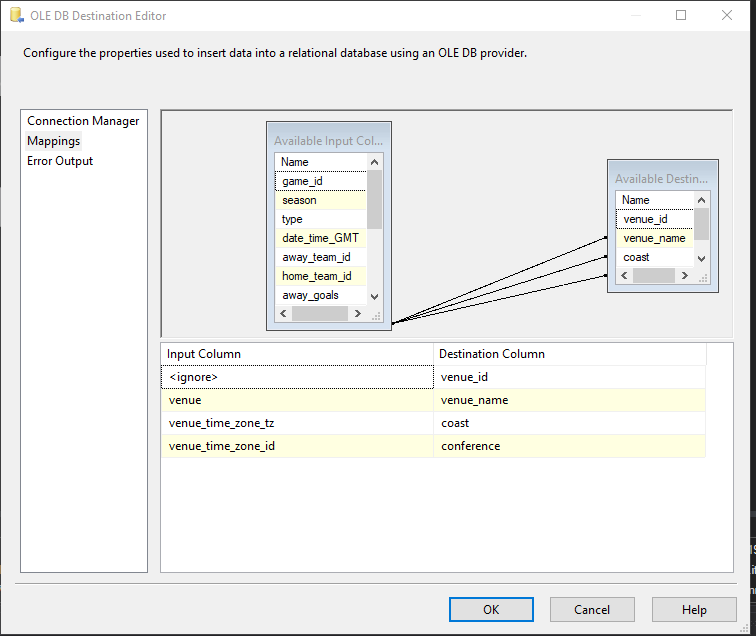
*Slika 8: Data flow za tabelu Venue*

Kao što se može videti, data flow za dimenziju Venue nije kompleksan. Sastoji se od flat file source-a koji se učitava. Podaci za venue su se nalazili u izvornoj Game tabeli i bilo je neophodno uz pomoć sort komponente izdvojiti obeležja koja su relevantna samo za Venue.



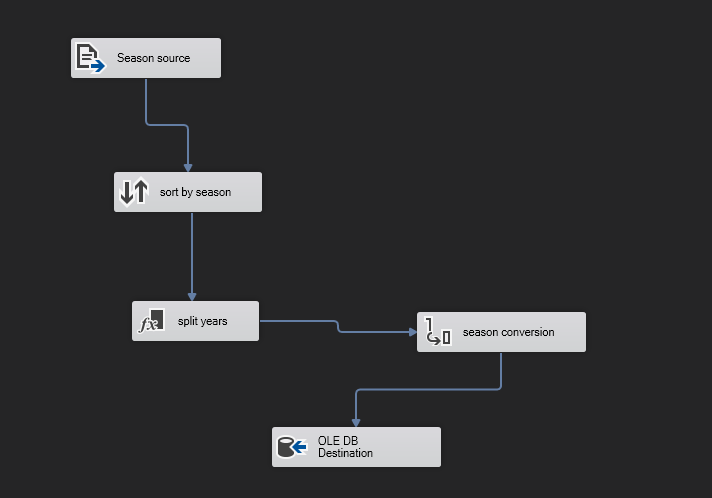
*Slika 9: Uzimanje relevatnih obeležja za dimenziju Venue*

Nakon uzetih obeležja, vrši se upisivanje na željenu destinaciju u tabeli Venue koja je prethodno uz pomoć Execute SQL Task-a kreirana u bazi. Neophodno je odabrati željenu tabelu, a nakon toga izvršiti mapiranja obeležja, da bi sistem znao gde koje obeležje treba da se upiše.

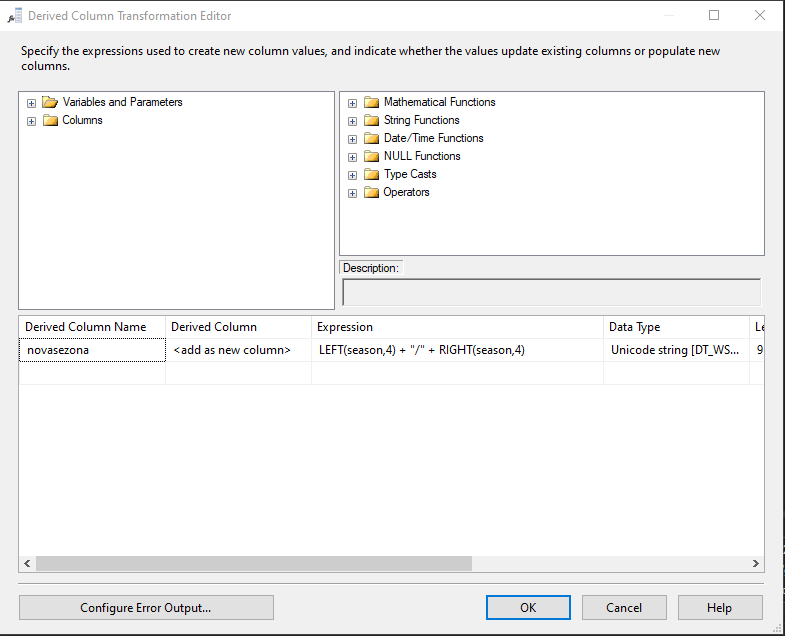
 

*Slika 10: Upis u tabelu Hockey.Venue*

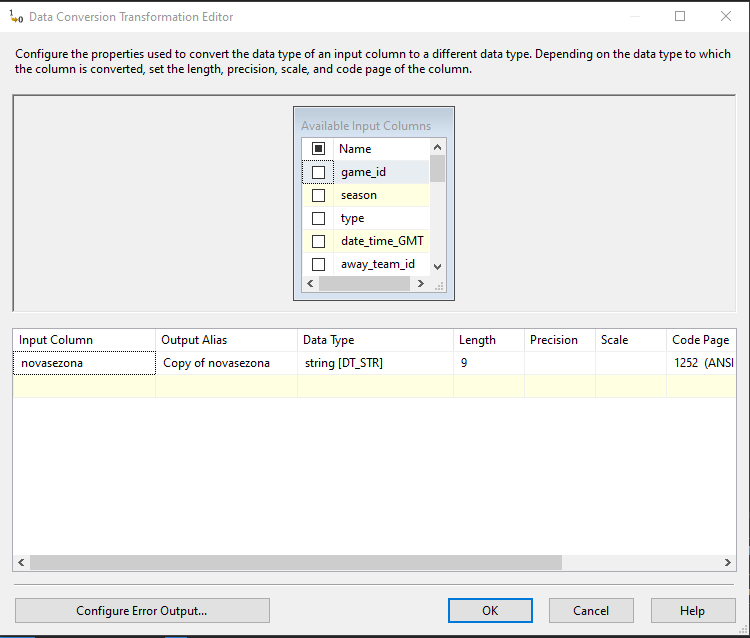
Nakon dimenzije Venue, sledi kreiranje data flow-a za dimenziju Season. Dimenzija Season se takođe nalazila izvorno u tabeli Game i za nju je bilo potrebno izvući relevantno obeležje i odraditi dodatne transformacije preko komponente Derived Column, gde je bio cilj da izvornu sezonu koja je bila u obliku 20152016 pretvorimo u oblik 2015/2016. Funkcija koja je korišćenja je : LEFT(season,4) + "/" + RIGHT(season,4). Left komponenta će u ovom slučaju uzeti prva 4 broja, na nju će se dodati / i kasnije uz pomoć right funkcije uzimamo 4 broja gledajući poslednje. Nakon uspešno izvršene transformacije, sledeći korak je konverzija nove kolone u odgovarajući tip podatka, da bi moglo da se izvrši mapiranje i upis u tabelu u bazi podataka.



*Slika 11: Data flow tabele Season*

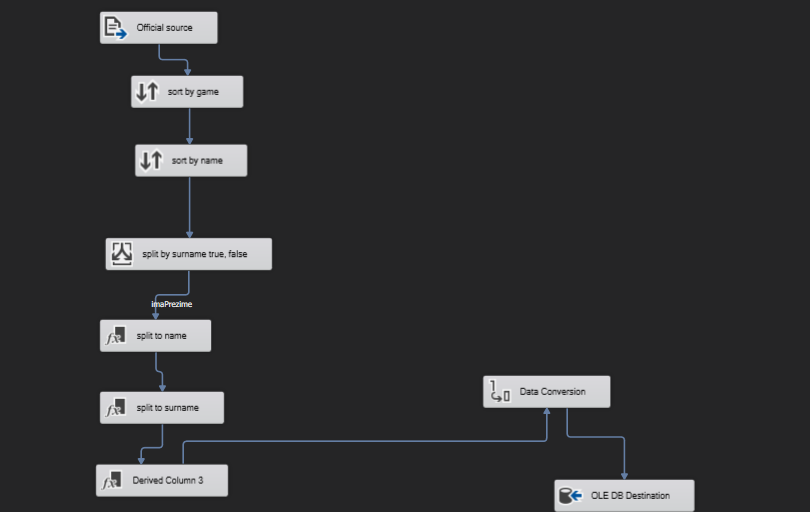


*Slika 12: Derived column u tabeli Season*



*Slika 13: Podešavanje Data conversion komponente*

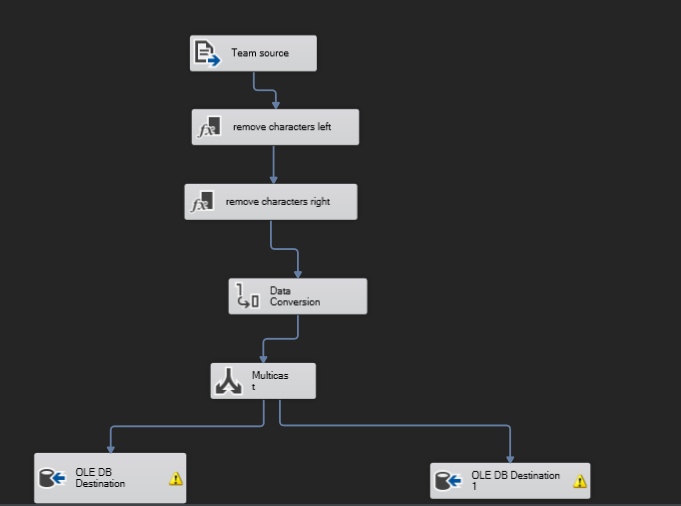
Naredna dimenzija koja se kreira jeste Official. U njoj se nalaze podaci o sudijama koji su sudili u poslednjih 19 sezona. U izvornom setu podataka, za svaku utakmicu je bilo prikazano 4 sudije sa tog meča. Dve sudije u polju i dve linijske sudije. Zbog praktičnosti, odlučeno je, za potrebe izrade ovog skladišta, da posmatramo utakmicu kao da je na njoj jedan sudija. Iz tog razloga, bilo je neophodno izvršiti prvo sortiranje po game\_id i uključiti opciju da se sklone redovi koji imaju duplikat vrednosti po tom obeležju. Na taj način, za svaku utakmicu ostao je samo jedan sudija. Nakon toga, primećeno je da se u izvornoj tabeli nalaze sudije čije koji nisu imali prezime već samo ime. Korišćenjem komponente Conditional Split podelili smo sudije na one koji imaju prezime i one koji nemaju. Funkcija koja je korišćena prilikom ove provere je: TOKENCOUNT(official\_name," ") >= 2. Funkcija tokencount broji koliko imamo elemenata u obeležju, ovde smo ih splitovali po praznom polju. Zaključujemo da ukoliko je broj stringova u official\_name >= 2, to znači da imamo ime i prezime i da taj sudija može da uđe u analizu. Sudije koje ne zadoljavaju ovaj uslov nisu uzete u razmatranje prilikom rada na ovom projektu. Nakon toga, vršimo split punog imena sudije na ime i prezime. Funkcija koja je korišćena za razdavajanje imena je : LEFT(official\_name,FINDSTRING(official\_name," ",1) - 1), a funkcija koja je korišćena za razdvajanje prezimena je : RIGHT(official\_name,LEN(official\_name) - FINDSTRING(official\_name," ",1)). Nakon toga, vrši se data conversion imena i prezimena i upis u bazu podataka. Slika data flow-a dimenzije Official je prikazana na slici 14.



*Slika 14: Podešavanje konekcije u OLE DB destination komponenti*

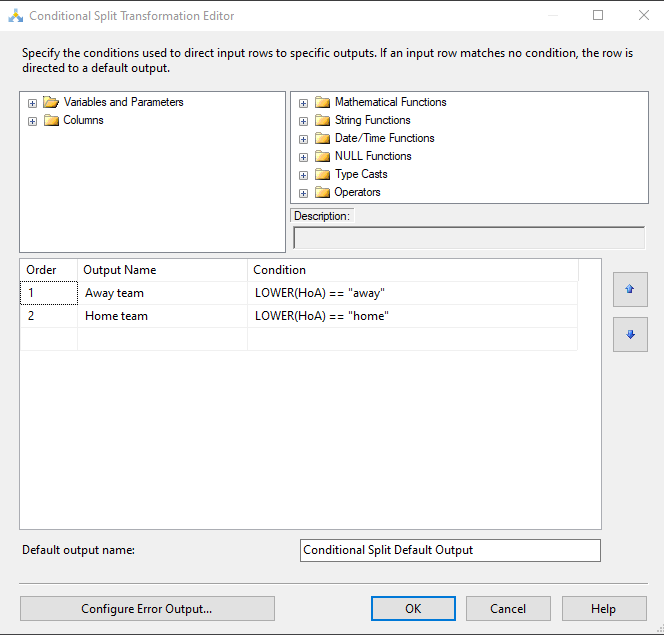
Sledeći će biti prikazan data flow za učitavanje podataka u tabele Home team i Away team, one u sebi sadrže iste podatke. Prvi zadatak koji se našao na putu jeste čišćenje team\_id obeležja. On se u izvornoj tabeli nalazio u obliku “2” i to nije bilo odgovarajuće za rad. Bilo je neophodno ukloniti navodnike i ostaviti čist broj.To se uradilo uz pomoć funkcije REPLACE.

Primer: REPLACE(["team\_id"],LEFT(["team\_id"],1),""). Funkcija prvi karakter sa leve strane zamenjuje praznim stringom. Analogno tome, uklanjanje navodnika sa desne strane se sprovodi funkcijom RIGHT. Primer: REPLACE(["team\_id"],RIGHT(["team\_id"],1),""). Nakon ovog postupka, vrši se data conversion novog team\_id-a. I onda se uz pomoć komponente Multicast, ovi podaci razlivaju na dva Ole DB destinationa,, jer je iste podatke neophodno upisati u tabelu Home Team i u tabelu Away team.

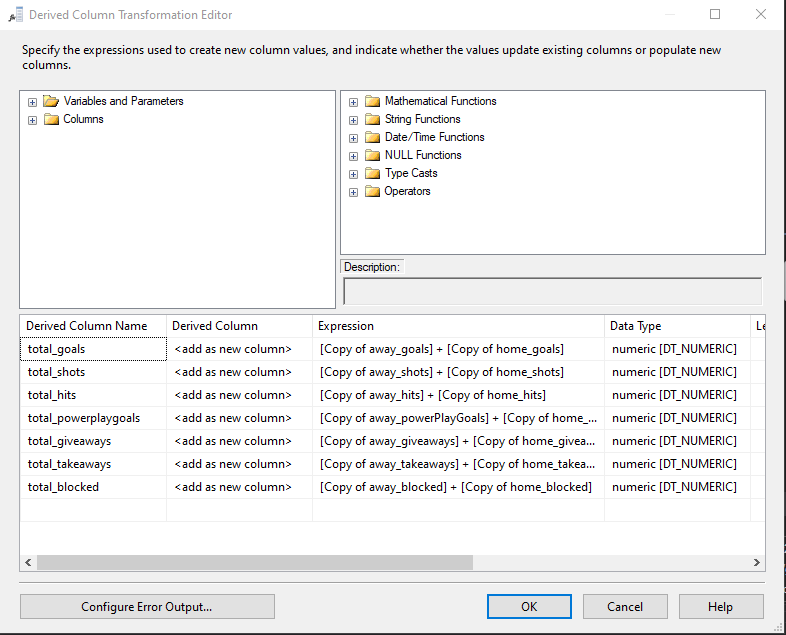


*Slika 15: Data flow za tabele Home team i Away team*

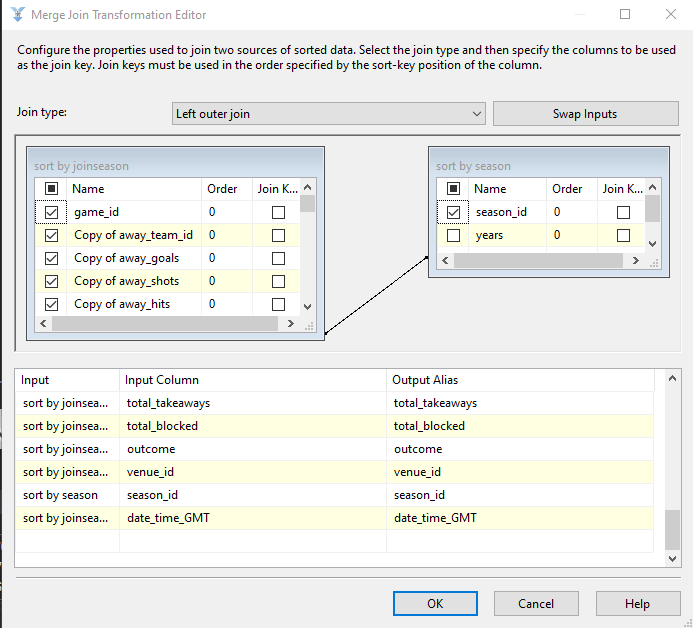
Nakon kreiranja svih neophodnih dimenzija, bilo je vreme za popunjavanje čijenične tabele. Ona je sama po sebi veoma kompleksna i sada će biti objašnjen celokupni postupak. Imali smo tabelu Team Stats gde su za svaku utakmicu postojale dve torke koja su se ticale domaćeg i gostujućeg tima. Ideja je bila da se u jednoj torci činjenične tabele nalaze svi ti podaci. Rešenje koje se nametnulo kao logično jeste izvršenje conditional split-a, gde bi se iz izvorne tabele izdvojile torke koje kao obeležje HoA(home or away) imaju vrednost ili home ili away.Nakon toga se izvršava sortiranje po game\_id koji je zajedničko obeležje i vrši se spajanje ove dve faktički nove tabele. Neophodni data conversion je odrađen i onda se pristupa komponenti derived column gde se izračunavaju mere od interesa u tabeli činjenica poput: total\_goals, total\_shoots, total\_hits, total\_blocked, total\_giveaways, total\_takeaways...Sledeći korak je učitavanje flat file source Game odakle se izvlači venue, season, datum, overcome...Nakon toga se učitava iz ole db source Venue tabela, uzimamo naziv dvorane i vršimo spajanje sa dotadašnjim podacima koji se provlače otpočetka dataflow-a. Učitavanje Season je zanimljivo jer se kreira nova kolona samo za join uz pomoć ove funkcije: LEFT(years,4). Izvorni podaci su u obliku 20152016 a u bazi su upisani kao 2015/2016, na ovaj način elegantno rešavamo problem nekompatibilnosti podataka i uspešno izvršavamo spajanje.Sledeću stvar predstavlja učitavanje Official tabele. Dosadašnji dataflow spajamo sa Official tabelom uz pomoć fullname-a sudije. Neophodno je napomenuti da je od interesa prilikom spajanja ovih dimenzija da se u merge join-u preuzme njihov veštački ključ. Spajanje sa Date dimenzijom se vrši konverzijom datuma iz OLTP-a pomoću funkcije LEFT(date\_time\_GMT,10), jer je analizom utvrđeno da takav način omogućava spajanje sa formatom koji podržava SQL server managment studio. Poslednja stvar jeste učitavanje Home team i Away team tabele, sortiranje po njihovom obeležjima i spajanje tako da se na kraju celokupnog procesa u fact tabeli nalaze sva obeležja koja su potrebna za punjenje.



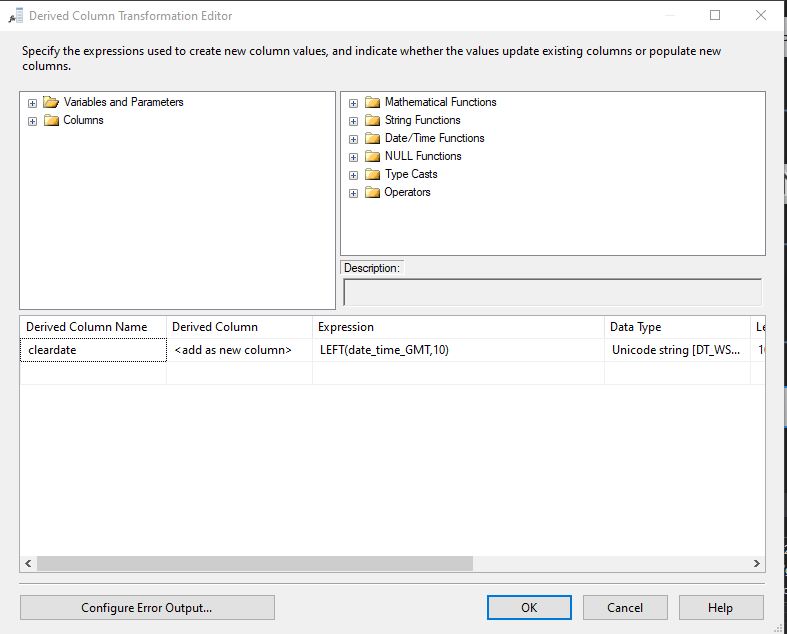
*Slika 17: Podela na domaće i gostujuće, pre njihovog spajanja u jednu torku*



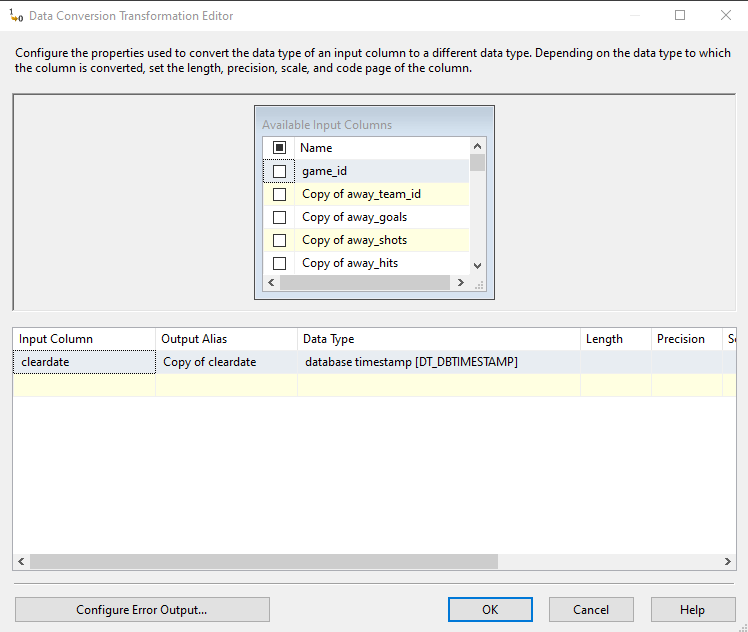
*Slika 18: Računanje mera*



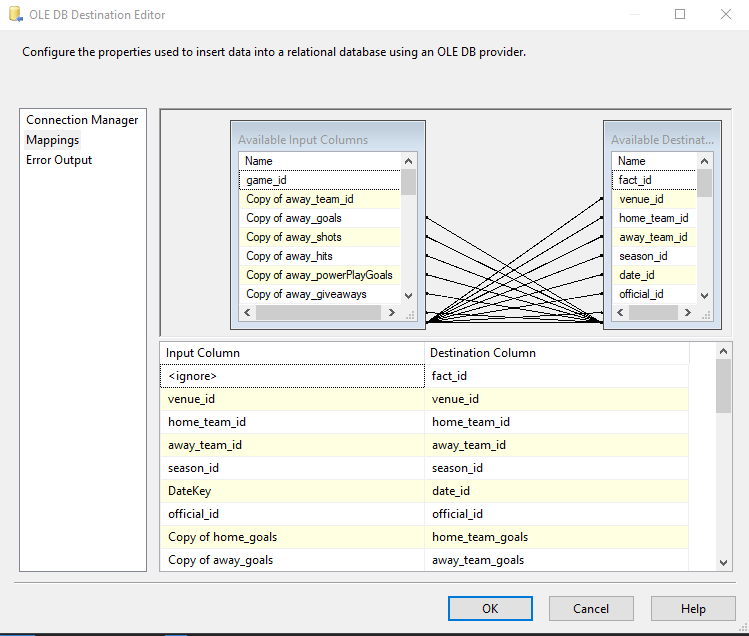
*Slika 19: Stanje podataka nakon spajanja po sezoni*



*Slika 20: Čišćenje datuma*



*Slika 21:Konverzija datuma pre spajanja sa datumom iz baze*



*Slika 22:Konačno mapiranje*

# PRIKAZ IZVEŠTAJA KAO ODGOVOR NA ZAHTEVE KOJI SU POSTAVLJENI OD STRANE KORISNIKA

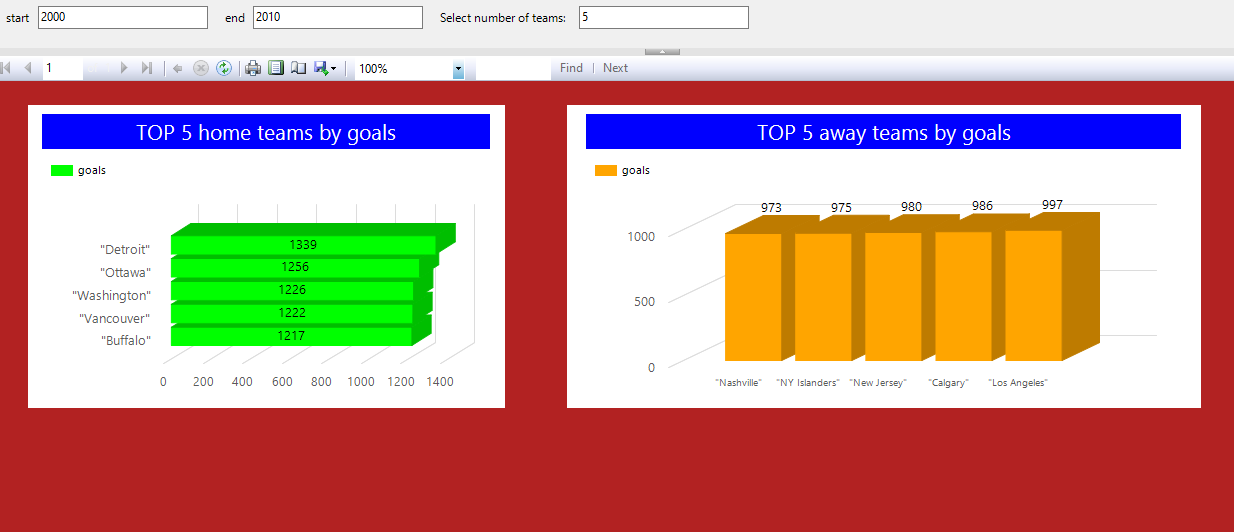
Sledeći izveštaji kreiraju se u okviru Visual Studio-a korišćenjem tipa projekta koji se naziva Report Server Project. Izveštaji se kreiraju tako što se nakon kreiranja projekta, kreira Shared Data Source u kom se definiše konekcija ka malopre kreiranom skladištu podataka. Na taj način, svi izveštaji će koristiti te podatke. Nakon kreiranja Data source-a, kreiraju se izveštaji koji imaju zadatak da daju odgovore na svaki korisnički zahtev. Koristeći Query dizajner, biraju se tabele koje su potrebne za kreiranje izveštaja i ukoliko je potrebno vrše se filtracija i sortiranje. Nakon učitavanja podataka, bira se određeni način prikaza izveštaja, i u nastavku će biti prikazane tabele, chart-ovi kao što su 3D Funnel, 3D Clustered Column,Stacked Bar...Izveštaji su interaktivni, korisnik može da unosi željene parametre na osnovu kojih će se i sam prikaz podataka menjati.

1. **Prvi izveštaj**

Prvi kreirani izveštaj ima za cilj da odgovori na sledeći korisnički zahtev:

* + Koji timovi su bili najefikasniji kod kuće, a koji timovi u gostima u prvoj deceniji 21. veka?

Izveštaj omogućava korisniku da sam izabere koliko najboljih timova u odnosu na date golove u u gostima i kod kuće će mu se prikazati,a po defaultu se prikazuje top 5 timova.



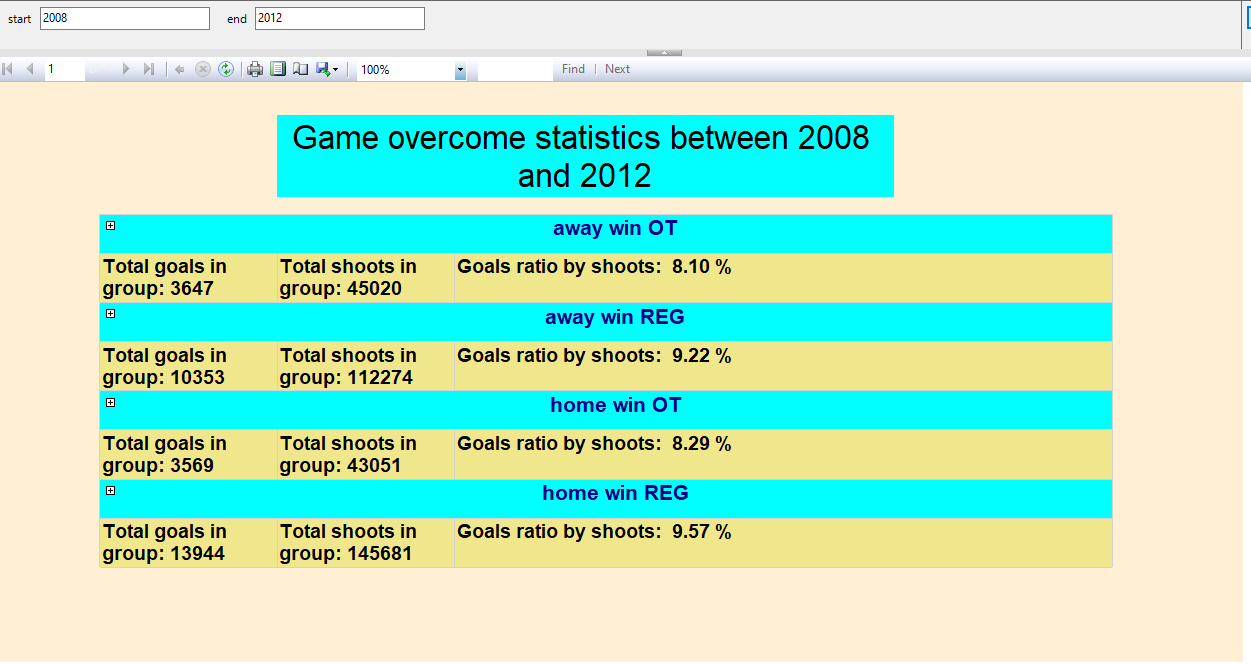
*Slika 27: Prvi izveštaj*

1. **Drugi izveštaj**

Drugi kreirani izveštaj ima za cilj da odgovori na sledeće korisničke zahteve:

* Kako su se najčešće završavale utakmice u periodu izmedju 2008. i 2012. godine?
* Kakav je odnos postignutih golova u odnosu na broj šuteva u tom periodu, grupisano po načinu završetka?

Izveštaj omogućava korisniku da unese početnu i krajnju godinu,te se u zavinosti od prosledjenih paramatera od strane korisnika i sam izveštaj menja,odnosno prikazuje podatke za različite periode. Ovaj izveštaj prikazuje statistiku postignutih golova, broj šuteva i odnos između ova dva parametra. Klikom na dugme +, otvaraju se sve utakmice koje su u tom periodu završene na taj način. Možemo videti da je najveći broj golova postignut kada domaći tim pobeđuje u regularnom toku meča što nam daje indiciju da je ovakav ishod mečeva najčešći i da u takvom tipu utakmica nema nekih kalkulacija već utakmice obiluju golovima i šansama.

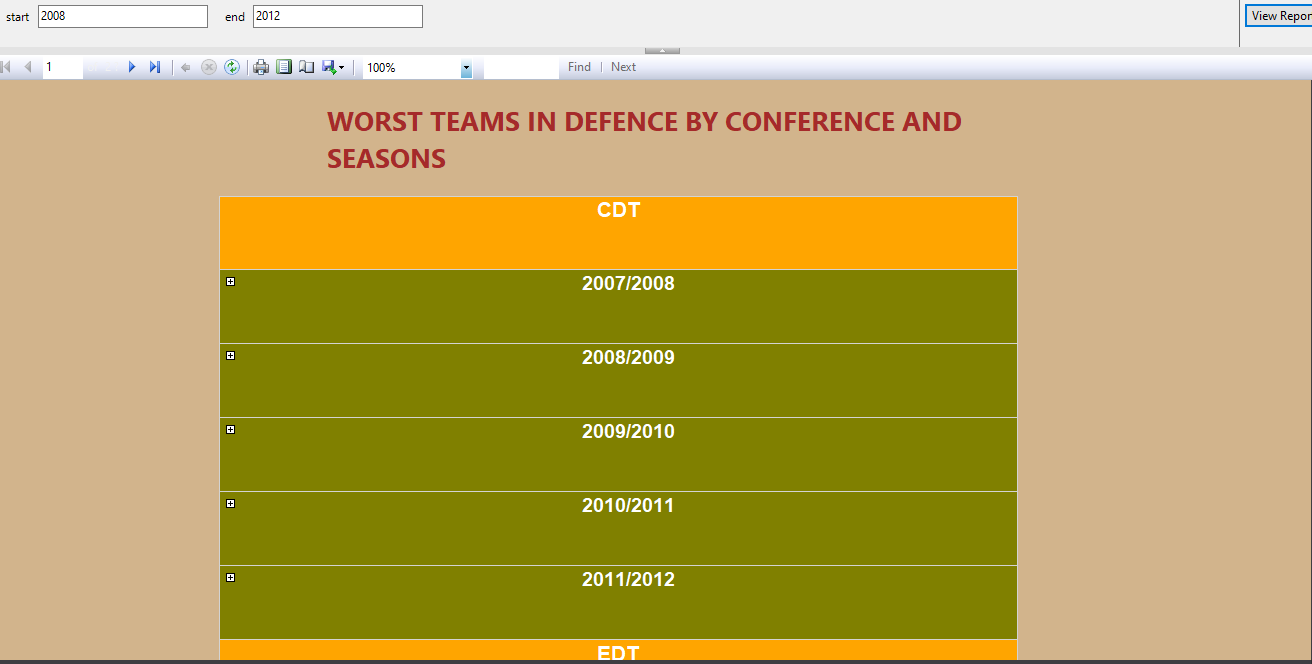


*Slika 28: Drugi izveštaj*

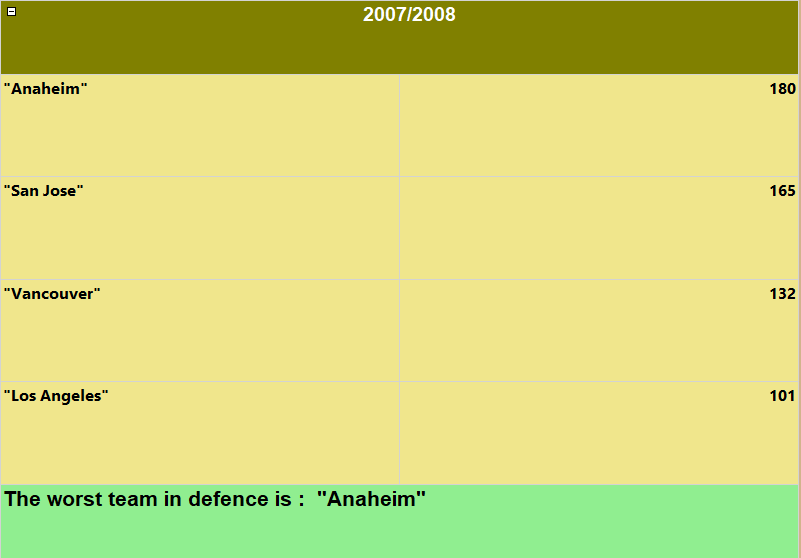
1. **Treći izveštaj**

Treći kreirani izveštaj ima za cilj da odgovori na sledeći korisnički zahtev:

* Koji timovi su imali najgoru timsku odbranu gledajući sezone i konferencije u kojima se nalaze? Ovaj izveštaj omogućava izbor sezona od interesa i na osnovu njih možemo videti koliko su timovi imali izgubljenih lopti u toj sezoni, što daje indiciju da im je timska odbrana i komunikacija bila na nezavidnom nivou.



*Slika 31: Treći izveštaj*



*Slika 32: Treći izveštaj,nastavak*

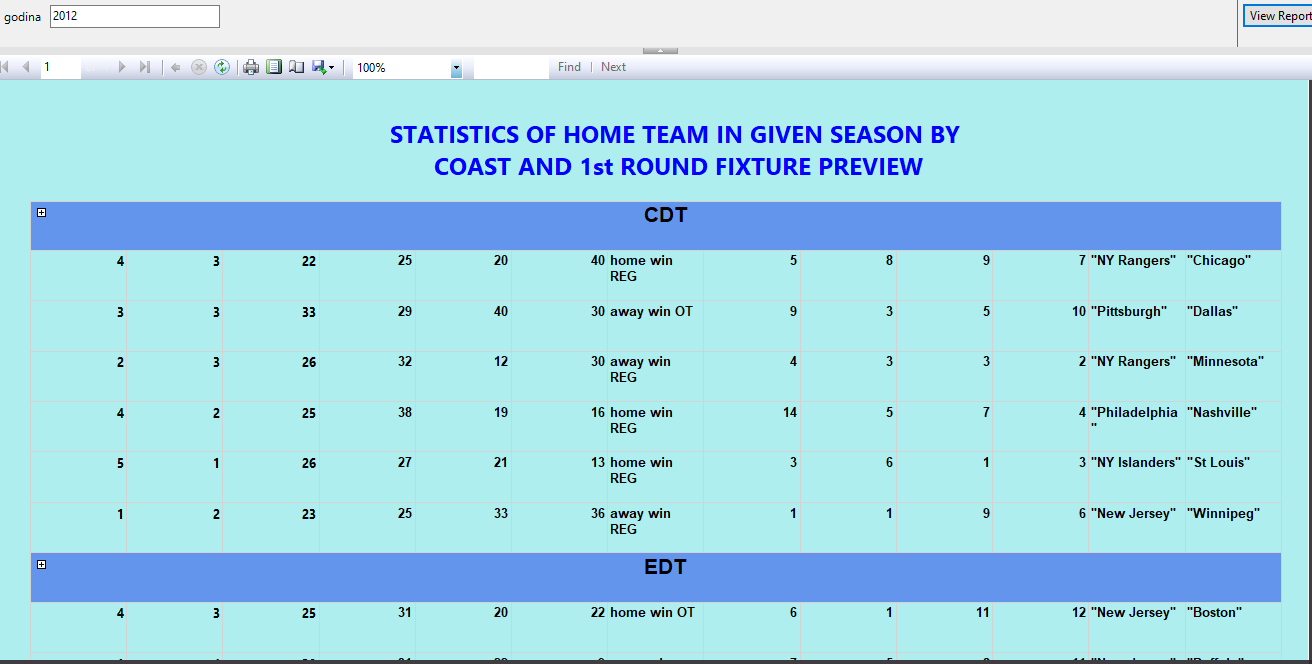
Kao što se može videti na slici 32. najgoru timsku odbranu u konferenciji PDT za sezonu 2007/2008 je imao tim Anaheim. Rezultati su sortirani opadajuće pa možemo zaključiti da je drugi najgori tim u ovoj klasi ekipa San Jose.

1. **Četvrti izveštaj**

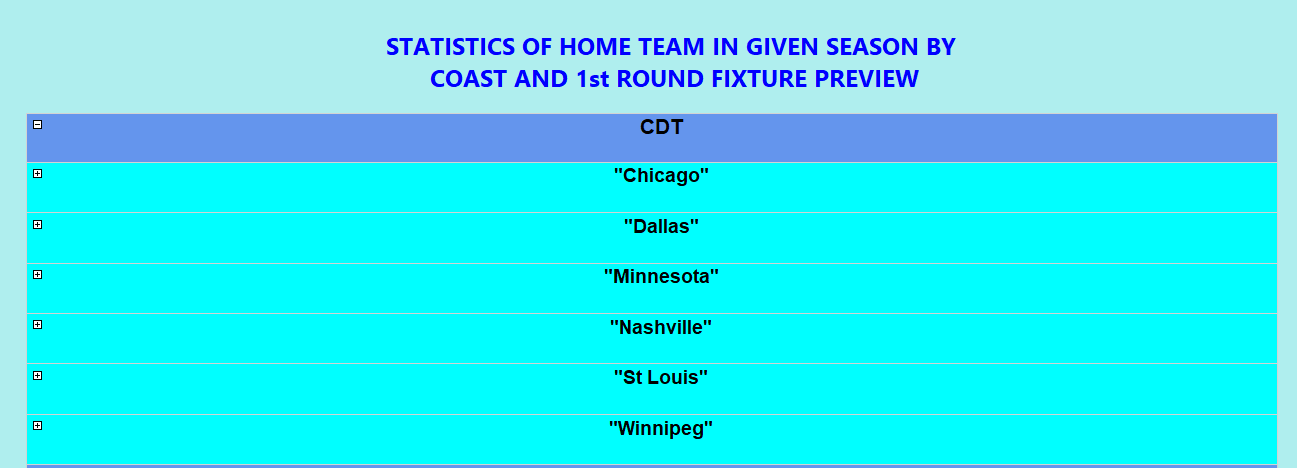
Četvrti kreirani izveštaj ima za cilj da odgovori na sledeće korisničke zahteve:

* Koji su bili rezultati prvog kola po konferencijama u datoj sezoni?
* Koliki je ukupan broj golova,šuteva, faulova,ukradenih i izgubljenih lopti za domaći tim u toj sezoni? Koliki je prosečan broj gore navedenih parametara?
* Koje su maksimalne vrednosti ovih parametara koje je tim postigao u toku te sezone?

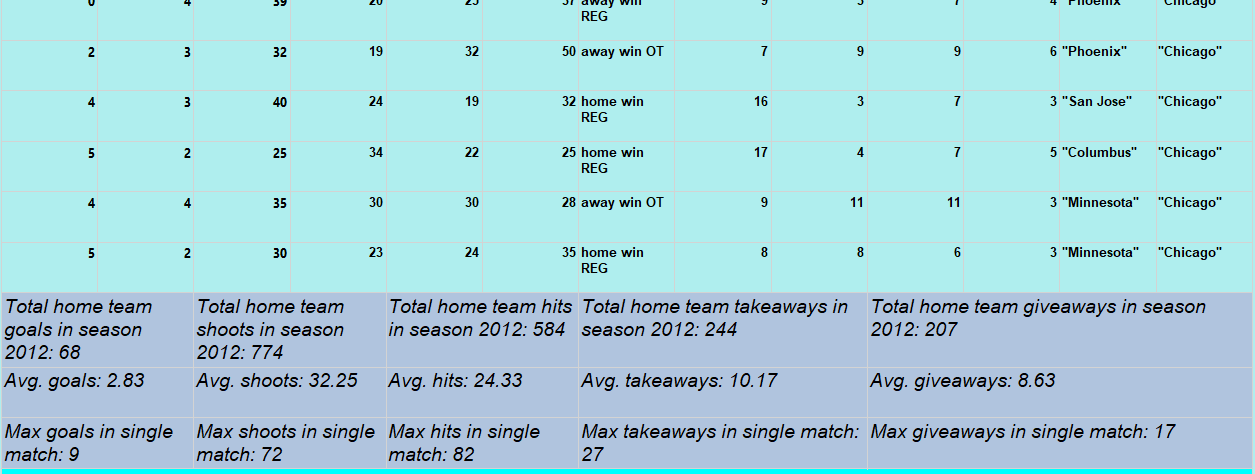
Izveštaj omogućava korisniku unese željenu sezonu koja mu omogućava da vidi rezultate prvog kola u toj sezoni, proširenjem konferencije, otvaraju se imena domaćih timova Pored toga,izveštaj prikazuje i opširnu statistiku koja se tiče dešavanja na svakoj pojedinačnoj utakmici.



*Slika 33: Četvrti izveštaj*

**

*Slika 34: Četvrti izveštaj*

**

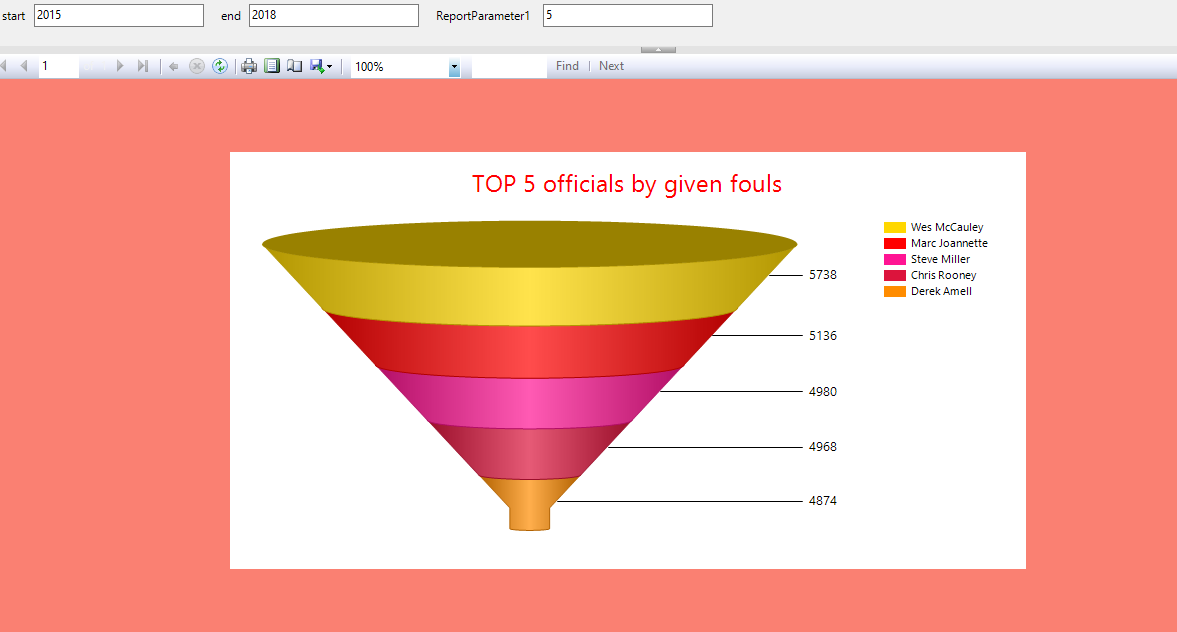
*Slika 34: Četvrti izveštaj, statistika za Chicago*

1. **Peti izveštaj**

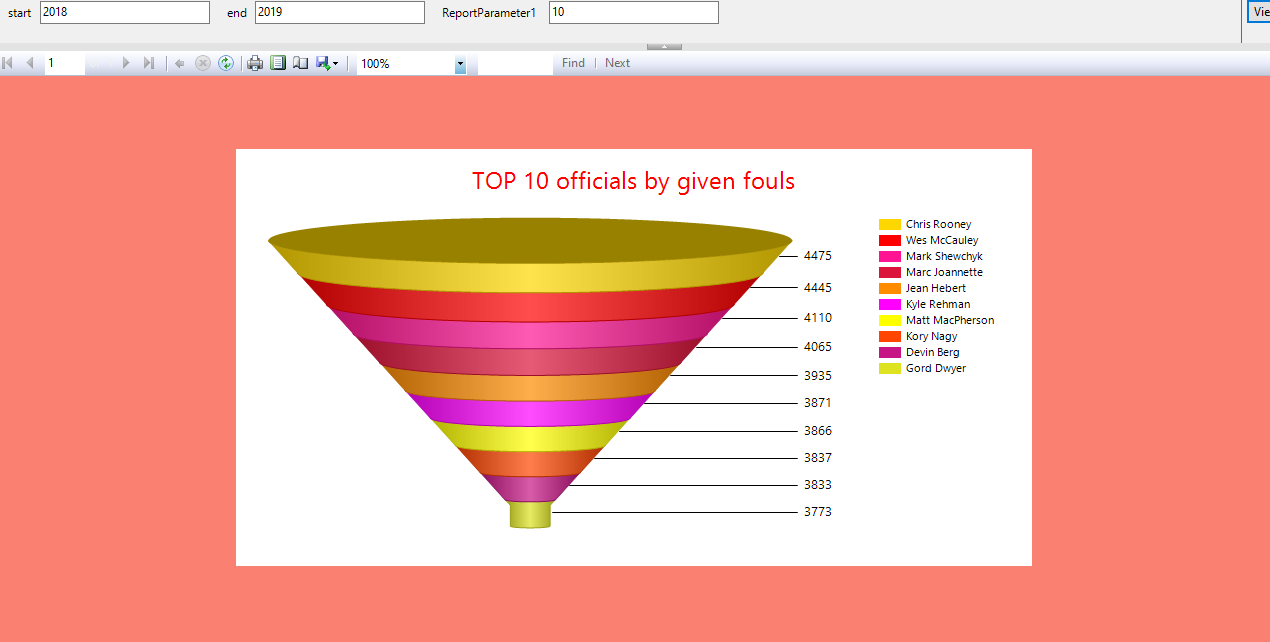
Peti kreirani izveštaj ima cilj da odgovori na sledeći korisnički zahtev:

* Koje sudije imaju mali prag tolerancije kada je reč o faulovima? Odnosno, koje sudije najčešće sviraju prekršaje u poslednje 3 sezone?

Ovaj izveštaj je prikazan putem 3D Funnel-a. Interesuje nas koje sudije često sviraju prekršaje i isključenja. Pre utakmice bitno je znati ko je glavni arbitar na meču, kako bi se tim pripremio i znao koliki stepen tolerancije sudija ima za grubu igru. Često se dešava da igrači budu isključeni iz utakmice i onda se mogu desiti powerplaygoals koji je spomenut na početku dokumentacije, odnosno, postižu se golovi dok protivnički tim ima manje igrača u polju.



*Slika 35: Peti izveštaj*



*Slika 36: Peti izveštaj*

# ZAKLJUČAK

U okviru ove dokumentacije prikazan je i objašnjen proces kreiranja DW sistema na osnovu izvornog seta podatala koji sadrži podatke o odigranim utakmicama u NHL ligi.

Postupak je podrazumevao razumevanje podataka, kreiranje OLAP šeme, implementaciju iste, zatim kreiranje ETL procesa i popunjavanje skladišta podacima, i na kraju odgovaranje na korisničke zahteve putem izveštaja.

Dalji razvoj i ispitivanje ovog dataseta može uključivati kreiranje novog Data Mart-a gde bi se temeljno analizirale performanse igrača po timovima. Bliži se zimski prelazni rok i klubovima i agentima bi sigurno bilo interesantno da imaju uvid u to ko je u dobroj formi i koji transfer bi bilo korisno da se realizuje u narednom periodu.