**Tutorial**

**Auction Case Study – Solutions**

**1. Data Cleaning**

After doing some data explorations, I have decided to do the following two things:

1. Delete 10 duplicate records in Auction
2. Update 2 incorrect categories

**(a) Delete 10 duplicate records in Auction**

Found 10 duplicate records in dtaniar.auction2.

Select count(\*) From dtaniar.auction2;

Count(\*)

-----------

26189

Create Auction table using distinct to remove the 10 duplicate records:

Create table auction2

As

Select distinct \*

From dtaniar.auction2;

Check if the 10 duplicate records are removed.

Select count(\*) From auction2;

Count(\*)

-----------

26179

Result shows that 10 duplicate records are removed.

OR

Use distinct when creating temp fact.

CREATE table auctiontemp as

SELECT distinct

r.r\_name || n.n\_name as LocationID,

a.a\_endtime,

i.i\_itemkey,

c.cat\_categorykey,

a.a\_auctionprice,

i.i\_retailprice,

a.a\_auctionkey

FROM item2 i,

Category2 c,

Region2 r,

Nation2 n,

Auction2 a,

Customer2 cus

WHERE a.a\_itemkey = i.i\_itemkey

AND i.i\_categorykey = c.cat\_categoryKey

AND cus.c\_nationkey = n.n\_nationkey

AND r.r\_regionkey = n.n\_regionkey

AND a.a\_username = cus.c\_username;

**(b) Update 2 incorrect categories**

First I checked the item table, and found 2 interesting records.

Select \*

From item2

Where i\_categorykey = 'cat1';

Item I1 which is a mobile phone is categorized cat1 (a null category in the category table), whereas item I2 which is a bid package is also categorized cat1.

Item I1 should have been categorized cat12 (Mobile Telephones), whereas item I2 should have been categorized as cat2 (Bid Package).

So, I updated these two item records to reflect the correct categories.

Update item2

Set i\_categorykey = 'cat12'

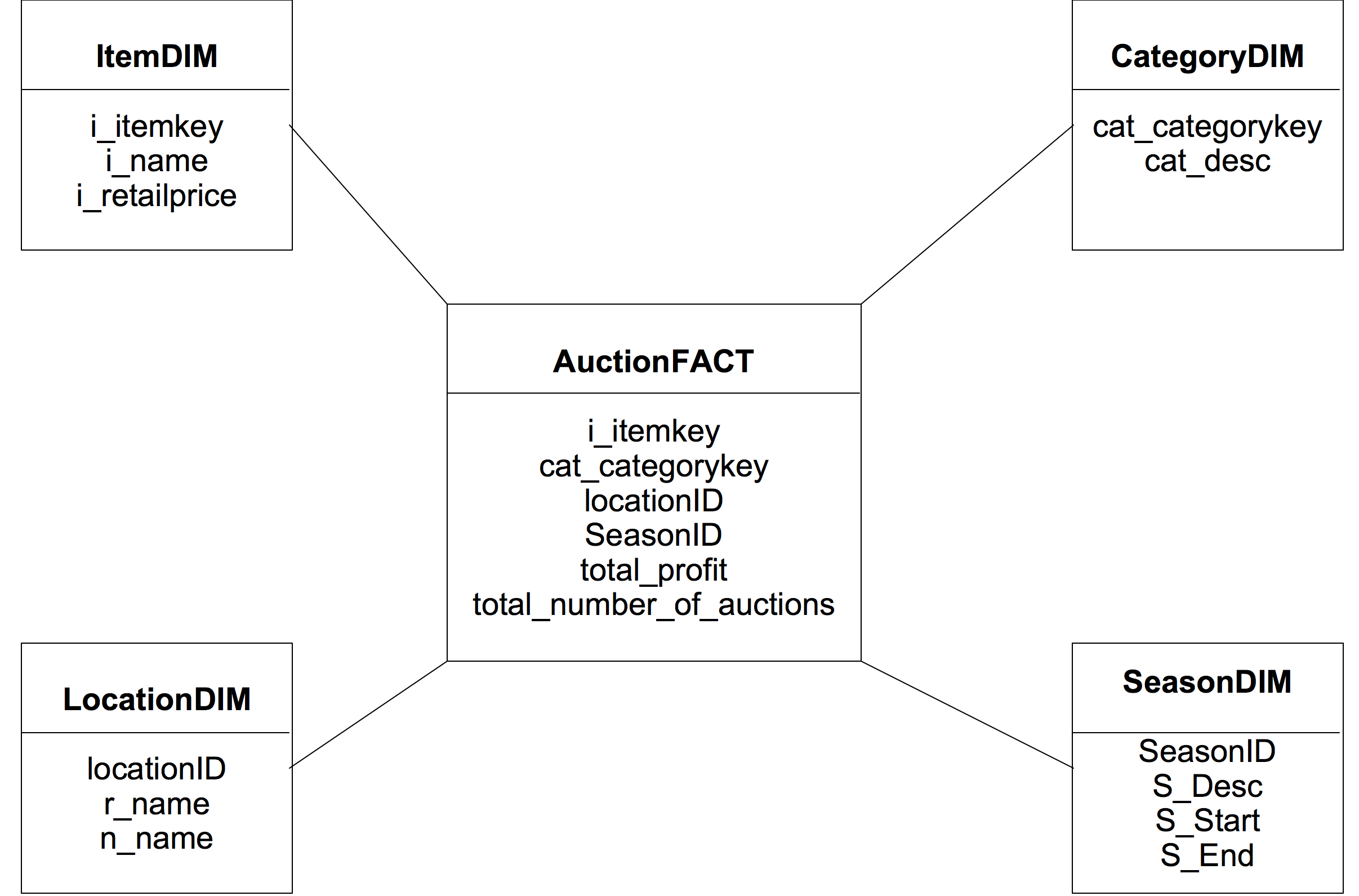
Where i\_itemkey = 'I1';

Update item2

Set i\_categorykey = 'cat2'

Where i\_itemkey = 'I2';

**2. Star Schema Diagram**



**3. SQL commands to create dimension and fact tables**

Drop Table itemDIM;

Drop Table CategoryDIM;

Drop Table LocationDIM;

Drop Table SeasonDIM;

Drop Table AuctionTemp;

Drop Table AuctionFact;

CREATE table itemDim as

SELECT distinct i\_itemkey, i\_name, i\_retailprice

FROM item2;

CREATE table categorydim as

SELECT distinct \*

FROM category2;

CREATE table LocationDim as

SELECT distinct

r.r\_name || n.n\_name as LocationID,

r.r\_name,

n.n\_name

FROM nation2 n, region2 r

WHERE r.r\_regionkey = n.n\_regionkey;

CREATE table SeasonDim

(SeasonID number,

S\_Desc varchar2(15),

S\_Start varchar2(15),

S\_End varchar2(15));

Insert into SeasonDim values (1, 'Spring', 'September', 'November');

Insert into SeasonDim values (2, 'Summer', 'December', 'February');

Insert into SeasonDim values (3, 'Autumn', 'March', 'May');

Insert into SeasonDim values (4, 'Winter', 'June', 'August');

CREATE table auctiontemp as

SELECT

r.r\_name || n.n\_name as LocationID,

a.a\_endtime,

i.i\_itemkey,

c.cat\_categorykey,

a.a\_auctionprice,

i.i\_retailprice,

a.a\_auctionkey

FROM item2 i,

Category2 c,

Region2 r,

Nation2 n,

Auction2 a,

Customer2 cus

WHERE a.a\_itemkey = i.i\_itemkey

AND i.i\_categorykey = c.cat\_categoryKey

AND cus.c\_nationkey = n.n\_nationkey

AND r.r\_regionkey = n.n\_regionkey

AND a.a\_username = cus.c\_username;

ALTER table auctiontemp

ADD(SeasonID number);

Update auctiontemp

Set SeasonID = 1

Where to\_char(a\_endtime, 'MM')>= '09' and to\_char(a\_endtime, 'MM') <='11';

Update auctiontemp

Set SeasonID = 2

Where to\_char(a\_endtime, 'MM')>= '12' or to\_char(a\_endtime, 'MM') <='02';

Update auctiontemp

Set SeasonID = 3

Where to\_char(a\_endtime, 'MM')>= '03' and to\_char(a\_endtime, 'MM') <='05';

Update auctiontemp

Set SeasonID = 4

Where to\_char(a\_endtime, 'MM')>= '06' and to\_char(a\_endtime, 'MM') <='08';

CREATE table AuctionFact as Select

t.LocationID,

t.SeasonID,

t.i\_itemkey,

t.cat\_categorykey,

sum((t.a\_auctionPrice/0.02) \* 0.75 - t.i\_retailPrice) as total\_profit,

count(t.a\_auctionkey) as total\_number\_of\_auctions

FROM auctiontemp t

GROUP BY

t.locationID,

t.SeasonID,

t.i\_itemkey,

t.cat\_categorykey;

**4. SQL commands to answer the two queries**

--Q1 What is the total profit list based on item category and item name?

Select f.cat\_categorykey, c.cat\_desc, f.i\_itemkey, i.i\_name, sum(f.total\_profit) as Total\_Profit

From AuctionFact f, itemDim i, categorydim c

Where f.cat\_categorykey = c.cat\_categorykey

And f.i\_itemkey = i.i\_itemkey

Group By f.cat\_categorykey, c.cat\_desc, f.i\_itemkey, i.i\_name;

CAT\_CATEGORYKEY CAT\_DESC

----------- -----------------------------------------

I\_ITEMKEY

------------------------------

I\_NAME

------------------------------------------------------

TOTAL\_PROFIT

------------

cat12 Mobiles+%26+Telephones

I151

XPERIA X10 Sony Ericsson

5414.25

cat6 TV

I188

Blu-ray Home Cinema System Samsung - HTBD1252 5.1ch

86

cat7 Home+and+garden

I187

Veuve Clicquot Carte Etui Chic 75cl

-52

265 rows selected.

--Q2 What is the total number of auctions in each season (e.g. Summer,

Autum, Winter, Spring) in different regions/nations?

-- Based on Region and Nation

Select f.SeasonID, s.S\_Desc , f.locationID as Region\_Nation, sum(f.total\_number\_of\_auctions) as Total\_Auction

From AuctionFact f, SeasonDim s

Where f.SeasonID = s.SeasonID

Group By f.SeasonID, s.S\_Desc , f.locationID;

SEASONID S\_DESC REGION\_NATION TOTAL\_AUCTION

------- --------------- ------------------- -------------

2 Summer AFRICAMOZAMBIQUE 585

3 Autumn ASIACHINA 219

3 Autumn AFRICAALGERIA 218

2 Summer ASIACHINA 481

3 Autumn EUROPEGERMANY 235

2 Summer EUROPEFRANCE 479

75 rows selected.

OR

-- Based on Region only

Select f.SeasonID, s.S\_Desc, l.r\_name as Region, sum(f.total\_number\_of\_auctions) as Total\_Auction

From AuctionFact f, SeasonDim s, LocationDim l

Where f.SeasonID = s.SeasonID

And f.LocationID = l.LocationID

Group By f.SeasonID, s.S\_Desc, l.r\_name;

SEASONID S\_DESC REGION TOTAL\_AUCTION

---------- -------------- ------------------ -------------

2 Summer AMERICA 2677

3 Autumn ASIA 1110

2 Summer ASIA 2361

2 Summer AFRICA 2605

15 rows selected.

OR

-- Based on Nation only

Select f.SeasonID, s.S\_Desc, f.LocationID, l.n\_name as Nation, sum(f.total\_number\_of\_auctions) as Total\_Auction

From AuctionFact f, SeasonDim s, LocationDim l

Where f.SeasonID = s.SeasonID

And f.LocationID = l.LocationID

Group By f.SeasonID, s.S\_Desc, f.LocationID, l.n\_name;

SEASONID S\_DESC NATION TOTAL\_AUCTION

------- ------------ ------------------ -------------

3 Autumn JAPAN 236

3 Autumn CANADA 243

3 Autumn MOROCCO 186

2 Summer JAPAN 542

2 Summer CHINA 481

3 Autumn IRAQ 154

2 Summer ROMANIA 430

2 Summer BRAZIL 529

4 Winter CANADA 397

75 rows selected.

**THE END**