

Laboratory 4

Auction Case Study – Solutions

1. Data Cleaning

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

- a. Delete 10 duplicate records in Auction
- b. 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.

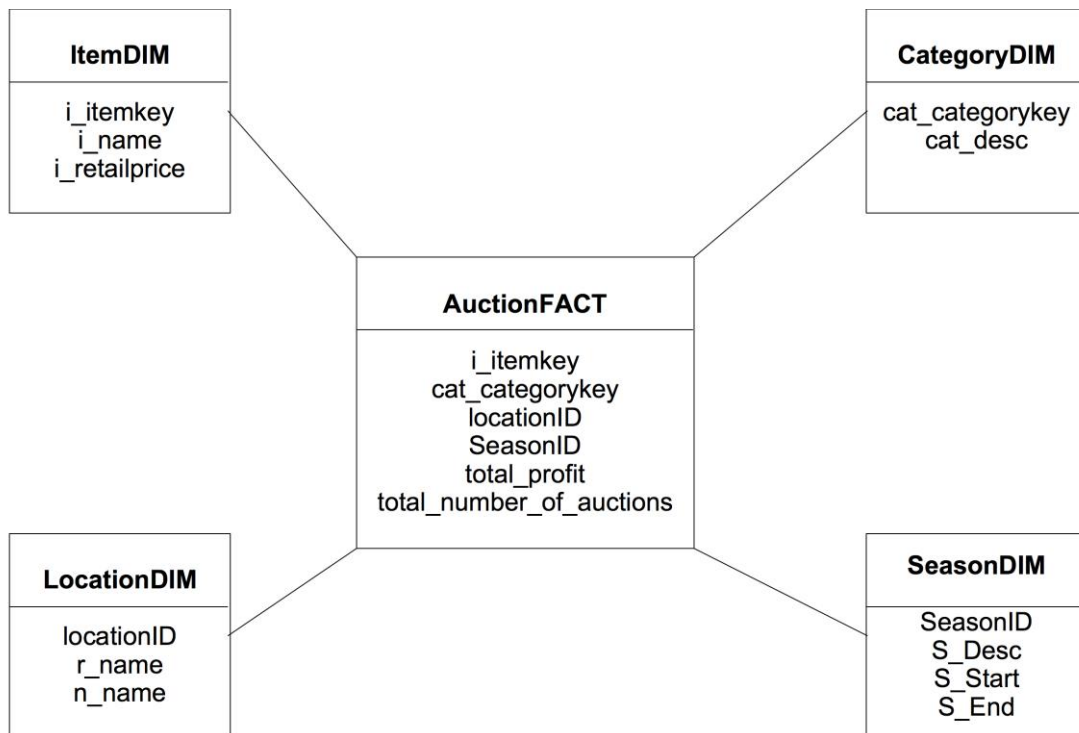
```

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, Autumn, 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