# Tutorial: Junk Dimensions

In the lecture, you have learned junk dimensions (the real estate agent case study). In this tutorial, you have four major tasks. I recommend that you open both the lecture notes on junk dimensions and this tutorial sheet on the screen.

## TASK A. Exploring the operational database

1. The source operational data can be found in dw.property1. How many records are there in the property table? (Notes: you need to write down the SQL commands, which produce the answer to each question in Task A). For example, the answer to question 1 is: "`Select count(*) From dw.property1;`", and there are 50000 records.
2. How many attributes are there in the property table? (Hint: use ALL_TAB_COLUMNS)
3. What are the values in attribute ensuite? What are the values in attributes pool, spa, and aspect_facing?
4. List all attributes that have only small number of values (e.g. attributes that have a maximum of 4 different values) (Hint: use ALL_TAB_COLUMNS)
5. How many different categories of properties?
6. What are the values in the attribute num_bedrooms?
7. Investigate the values of the houseprice attribute. What is the maximum and what is the minimum?
8. Investigate the suburb. How many suburbs listed?
9. List the suburb names, together with their postcodes.
10. There are four more property tables: dw.property2, dw.property3, dw.property4, and dw.property5. Investigate some of the questions above (questions 1-9) for the other four property tables.

## TASK B. Creating a data warehouse (Option 1 – Non Junk Dimension version)

1. Read the lecture notes (Option 1, the Non Junk Dimension version). Pay attention to the star schema.
2. Create the EnsuiteDIM1 dimension table (you can copy and paste the codes from the lecture notes). Why do you think we need EnsuiteID zero?
3. Create the other dimension tables (i.e. PoolDIM1, AspectFacingDIM1, and SpaDIM1). Why don't they have ID=0, like EnsuiteID?
4. Create the TempFact1 table.

5. How many records are there in TempFact1 table? Is the number of records in TempFact1 table fewer than or more than or the same as dw.property1?
6. How many attributes are there in TempFact1? Investigate the content of Tempfact1 table. Pay attention to the values of each attribute.
7. Alter TempFact1 tables to add EnsuiteID, and then Update this attribute. How many records are affected by each update. Is the sum of records affected by the three updates equal to the total number of records in TempFact1 table?
8. Do the same for attributes PoolID, Aspect_FacingID, and SpaID.
9. Investigate the contents of TempFact1 table. Check particularly the correlation between Ensuite and EnsuiteID, for example.
10. Create PropertyFact1 table.
11. How many records are there in PropertyFact1 table? Is there any significant difference in number of records between PropertyFact1 and TempFact1? Explain why.
12. Investigate the contents of PropertyFact1 table. Compare with TempFact1 table.
13. Now, let's drop all the dimension tables, tempfact table, and fact table. And then redo creating all the four dimensions; BUT, when you create the EnsuiteDIM table, do not update EnsuiteID=0 (leave it out). Then do everything else the same as before. What I would like you to do is to check the contents of the new Fact table. See what happens when you leave out the update EnsuiteID=0 for 'null'.

## TASK C. Creating a data warehouse (Option 2 – Junk Dimension version)

1. Create a JunkDIM table, alter the table to include a JunkID, create a sequence, and update the JunkDIM table
2. Check the contents of the JunkDIM table. How many records are there? Compare the contents of the JunkDIM table and the four dimension tables initially created for Option#1 above.
3. Create TempFact2 table. How many records are there? What is the difference between TempFact2 and TempFact1 tables previously created for Option#1 above?
4. Alter TempFact2 to include a JunkID attribute
5. Show the contents of table TempFact2. Did you see that the JunkID column is still empty?
6. Now update the JunkID column, using the cursor PL/SQL. What is the / symbol after the last end;?

7. Show the contents of table TempFact2 now. Pick a record in TempFact2 table, and see whether the JunkID matches with the JunkID in JunkDIM table.
8. Create the Fact2 table, and show the contents of this file. Compare between Fact2 (Option#2) and Fact1 (Option#1) tables.

## TASK D. Exploring the data warehouse

1. Display the number of property that have pool and spa; first using Fact1 (without the junk version), and then using Fact2 (with the junk version). Compare the codes and the results.
2. Show the number of property that has north facing.
3. Show the number of property that has north facing, but no pool.