Week 1 Assignment  
Group: Group 6 (Michael Adriel Darmawan & James Kumala)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. What are the best (review, useful, funny & cool) 10 restaurants of the last 3 years (2014, 2015 & 2016) in Last Vegas?  
     
   For this question, we identified 4 points. They are the reviews, the restaurants, the time, and the city. Therefore, we created in total 3 dimension tables (Business Types, Cities, Businesses, and Users) and 1 fact table (Reviews). The Business Types table will define the type of business, whether it is restaurant, theatre, bookstore, and etc. The Cities table will be used to know which city a particular business is from. These two tables then will help to identify the Business tables. For the Reviews table, which is a fact table, it will be an “amalgamation” of Business and Users table, where a user can give a review on a given business, then the data is stored using the scheme of Reviews table. This resulted in Snowflake scheme.  
   Diagram

   Description automatically generated  
   \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. Which businesses has the most reviews with the least users in Edinburgh?  
     
   In question number 2, we also see the same 4 points we identified previously in question number 1. This time, it asks for a business which has the most review, meaning that we need to have the dimension table of businesses. This also includes the cities dimension table. The difference here is that we do not need the business types dimension table, because we are talking about business in general, whereas question number 1 was talking about restaurant type of businesses specifically. We also need the reviews fact table, since it is asking about the most review. The question then follows by giving a condition that It should be with the least users in the city of Edinburgh, therefore we need users dimension table as well. So, this has a snowflake scheme as well.   
     
   A picture containing table

   Description automatically generated\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
3. What kind of weather (first selected type) was it on the top 10 days of most reviews in Las Vegas?  
     
   In this question, we are asked to find the weather on top 10 days where the most reviews are generated in the city of Las Vegas. First, we identify that we need to have a dimension table of weathers to store each day’s weather data for given city. Then, to know the top 10 days of most reviews we can group by the reviews using the date of the reviews, then we count the number of the reviews in each day before finally sorting it descending order. This results in a star scheme.  
     
   Chart, diagram, box and whisker chart

   Description automatically generated  
   \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
4. What is the current balance of each business of the top 10 most tip counts (in 2012) in Edinburgh?  
     
   For this question, we identify 4 points that could be made into 1 fact table and 2 dimension tables. First is the business dimension table. Each business has a record of its own balance. Then for the tip counts, we will have an extra tips fact table, containing how much the tip is, and the time the tip was given to a particular business. For the cities table, we can actually combine it with the business table, creating a new city field for businesses dimension table, but it can be used as an enhancement later on. Therefore, it creates a snowflake scheme.  
     
   Diagram

   Description automatically generated