**Week 1 Assignment - Data Acquisition**

Group: Group 6 (Michael Adriel Darmawan & James Kumala)

For the week 1 assigment, we would be using the Yelp database schema from the blackboard that can be seen as below.  
Diagram

Description automatically generated

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. So, for the conceptual data warehouse design, first we will create a category dimension table. This dimension table will be used to store the category information, as well as the business\_id as foreign key. The second dimension table is business table, which stores the information about the city, as well as the business name. The last dimension table is date dimension table, which stores the date time information for each review. The fact table in this solution is review table. This will contain the stars field (used to determine the review), useful field, cool field, and funny field (used to determine the useful, funny, cool assessment). This table will also contain the foreign keys for business table and date table. In the end, this resulted in a snowflake schema.

Diagram

Description automatically generated  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Which businesses has the most reviews with the least users in Edinburgh?  
     
   In question number 2, we identify 4 points, which is the businesses, the reviews, the users, and the city. We decided to create a data warehouse schema with 1 dimension table (business table) and 1 fact table (review table). For the business table, we will use it to store the name and city information. Whereas the review table will be used to store the business\_id foreign key, review count, and review\_user\_count to know the number of users for that business.

Chart, box and whisker chart

Description automatically generated\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. What kind of weather (first selected type) was it on the top 10 days of most reviews in Las Vegas?  
     
   In this question, we can identify the weather, the time (top 10 days), the reviews, and the city. First, we need the date dimension tables. This date dimension table will be used to get the top 10 days where the most reviews took place. Then we have the review fact table to store the weather id and date id as foreign keys, as well as the review count. For the weather dimension table, we will need it to know the weather type and the city.

Diagram

Description automatically generated  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. What is the current balance of each business of the top 10 most tip counts (in 2012) in Edinburgh?  
     
   In this question, we can identify balance, business, tip, and city. Therefore, we will need finance and tip tables. For the finance table, we will use it to store the balance of a business, which is linked through the finance id foreign key in business table. In tip fact table, it will be used to store the business id and date id foreign keys, as well as the tip counts. Last but not least, we also need the date dimension table.

Diagram

Description automatically generated