# Database Systems, CSCI 4380-01
# Homework # 1(a)
# Due Thursday September 12, 2018 at 11:59:59 PM

**Homework Statement.** This homework is worth 2.5% of your total grade. If you choose to skip it, Midterm #1 will be worth 2.5% more. Remember, practice is extremely important to do well in this class. I recommend that not only you solve this homework, but also work on homeworks from past semesters. Link to those is provided in the Piazza resources page.

This homework aims to teach you how to construct complex queries using relational algebra. Please do the parts in sequence. The questions get harder and build on your knowledge of relational algebra from previous parts. Each question is equal weight.

**Database Description.** Suppose you are given the following database for AirBnB for a single city (shortened from the actual public Airbnb database):

```
hosts(host_id, host_name, host_url, host_since, host_location, host_about,
  host_response_time, host_acceptance_rate, host_is_superhost, host_identity_verified)

neighborhoods(neighbourhood_group,neighbourhood)

listings(listing_id,name,host_id,neighbourhood,latitude,longitude,
  room_type, price, minimum_nights,number_of_reviews, scores_rating, scores_accuracy,
  scores_cleanliness, scores_checkin, scores_communication,scores_location)

calendar(listing_id,date, available, price)

reviews(review_id, listing_id, date, reviewer_id, reviewer_name, comments)
```

`Hosts` are individuals who rent out their apartments.

`Neighbourhoods` contains the names of specific areas of the city (e.g. `Manhattan` (neighbourhood group) and `SoHo` (neighbourhood.

`Listings` are for specific properties being rented. A listing can be for an entire home, a room in a home or another property type (given by `room_type`). Each listing contains information about where the property is, where there is a requirement of minimum number of nights to book and the average scores (out of 5) for different aspects of the property: rating, accuracy of description, cleanliness, checkin process, location and the communication with the host. The price attribute in listings is the average price for this property.

`Calendar` contains the price of the property for each day of the year and whether it is available or booked (`available` is true or false) on that day.

Note: All date fields are formatted as `mon-day-year`, e.g. `01-31-2019`. You can assume that you can check if a date value `X` comes after another value `Y` by checking whether `X` $>$ `Y`.

Write the following queries using relational algebra (pay attention to the attributes required in the output!):

**Question 1.** The following queries only need a single SELECT ($\sigma$), followed by a PROJECT ($\pi$) and RENAMING ($\rho$) as necessary:

(a) Return the id and name of all listings in the `Murray Hill` neighbourhood for a whole house (`room_type`) with 4.5 or higher scores for accuracy, cleanliness and communication and is for less than $80.

(b) Return the id, name and URL of all superhosts hosts who have been a host since 2016 and have an acceptance rate of 100%.

**Question 2.** The following queries combine SELECT ($\sigma$), SET operations ($\cap, \cup, -$), PROJECTION ($\pi$) and RENAMING ($\rho$) as necessary:

(a) Return id of listings that are either available for at least one day in the month of November 2019 or are hosted by host with id 2845.

(b) Return the id of listings that are not available on any day in November 2019 and have at least 100 reviews.

**Question 3.** The following queries combine SELECT ($\sigma$) statements with any number of JOINS as neede ($\bowtie$, theta or natural) (or CARTESIAN PRODUCT), followed by a PROJECT ($\pi$) and RENAMING ($\rho$) as necessary:

(a) Return the id and name of all hosts who have at least one listing with score 3 or lower for cleanliness for a property that is available on date `'9/10/2019'`.

(b) Return the id, name of all listings that has a review by a reviewer (`reviews.reviewer_name`) that has the same name as the host of the listing (`hosts.host_name`).

**SUBMISSION INSTRUCTIONS.** Submit a PDF document for this homework using Gradescope. No other format and no hand written homeworks please. No late submissions will be allowed.

If the gradescope for homework submissions is not immediately available, we will announce it on Piazza when it becomes available.