Project phase 1

Yiyun Ding 1004705214, Jianhong Guo 1004436862

1. Domain:

The data is mainly about courses of recreational programs offered by the city of toronto. Information such as course name, location, restrictions and facilities is included.

2. Dataset

Link:
Courses: https://open.toronto.ca/dataset/recreational-courses-historical-data/
Wards: https://open.toronto.ca/dataset/ward-profiles-2014-2018-wards/
Information relevant:
Almost all the information in the dataset is relevant except the “session_year”, “start year” , “start minute’, “start quarter”, and “end year” since we know that all of the courses in the table start and end in the same year.
learning that I have to do:
Ward means the political boundary that divides the City of Toronto into 44 areas. If I want to understand the course information in each political boundary, I need to learn where these boundaries are.
Any cleaning up you think you will have to do in order to use the data:
All data in the both tables is clearly stated and in required format. There is no need to remove any rows. The irrelevant columns should be removed to avoid redundancy.

3. Questions:

1. Which skating course lasts the least number of courses in Scarborough Southwest in summer?
2. What courses in York West are allowed for elderly people older than 80?
3. Which area’s course density is the highest?
4. Which swimming course is the most popular？

4. Schema:

Relations

Course(Course_Barcode, Name, Location, PopularityID, TimeID, Program, Course_Type, Number_of_classes).
Location(Location, Postal_Code, Ward).
Ward (Ward, Name, Area)
Facility(Location, Facility, Facility_Type, Facility_District)
Popularity(PopularityID, LimitID, Course_Waitlist, Course_Reg, Visits)

Time(TimeID, Reg_Session, Days_of_the_Week, Start_month, Start_day, Start_hour, Num_of_weeks, Course_Hours)
Limits(LimitID, Min_age, Max_age, Min_Reg, Max_Reg)
Program(Program, Section, Subsection)

Integrity Constraints
Ward ∈ {1,2,3...44}
Reg_Session ∈ {spring, summer, autumn, winter}
Course_Type ∈ {drop in, regular}
sub_section ∈ {early child, child, child/youth, youth, adult, order adult, all ages}
Popularity[PopularityID] ⊆Course[PopularityID]
Time[TimeID] ⊆Course[TimeID]
Program[Program] ⊆Course[Program]
Limits[LimitID] ⊆ Popularity[LimitID]
Facility[location] ⊆Location[location]
Course[location] ⊆Location[location]
Ward[Ward]⊆Location[Ward]