Data Mart Grain

The grain of this data mart is the number of recreational facilities in Canada, along with the location in Ontario, the type of facility, the locations ethnicity statistics, the locations age statistics.

Dimensions & Dimensional Attributes

Facility Type

Facility Key

Type

- Example: Pool
- Possible values: Pool, Parks, Playgrounds, Community Centre, Aquatic & Fitness Centre, Arena, Gym, Splash Park, Picnic Site

Location

Location key

Street Address Key

City

- Example: Ottawa
- Possible values: All cities in Canada

Province/Territory

- Example: Ontario
- Possible values: All provinces/Territories in Canada

Street Address

Street Address Key

Unit number

- Example: Null
- Possible values: null, or 0 to infinity

Street number

- Example: 75
- Possible values: null, or 0 to infinity

Street name

- Example: Laurier
- Possible values: string length > 1

Street type

- Example: Ave E
- Possible values: string length > 1

Postal Code

- Example: K1N 6N5
- Possible values: string length == 6

Location Ethnicity

Location Ethnicity Key

Location Key

South Asian

Chinese

Black

Filipino

Arab

Latin American

Southeast Asian

West Asian

Korean

Japanese

Other(visible minority but selected other)

Multiple Visible Minorities

Not a visible minority

For all of the ethnicities listed above, the minimum value is 0, and the maximum value is infinity. An example of the South Asian attribute could be 33 000.

Location Age

Location Ethnicity Key

Location Age Key

0-15

16-24

25-35

36-50

51+

Sample value 0-15: 11 000. Minimum value is 0, maximum value is infinity.

Sex Location

Location Ethnicity Key

Location Sex Key

Measures/Facts
Location Key
Facility Key
Location Ethnicity
Location Age
Facility Count

Assumptions
We are assuming that the data from all datasets are from the same years, although they are all different years.

Moreover, we will assume some facility types are the same as others for our data mart. For example, a rink is the same type as an arena, outdoor swimming pools, pools, and indoor swimming pools are the same, etc.

Design Mistake	How we avoided it
Facility type was a text attribute in the fact table	We made a new table for it, making its own dimension
Ethnicity was normalized.	We avoided the normalization for the ethnicity dimension by rethinking what attributes actually needed to be stored instead of just taking everything related to it.
Location included too many attributes	We used outrigging to decrease the number of attributes we needed and also to help with analysis for the different geographical areas.
Used operational keys as primary keys for the location age and location ethnicity dimensions	Realized our mistake and created primary keys for the dimensions.

Team Summary
For our team plan, we started by brainstorming and discussing what the grain of our model was. Then, we started looking into other external datasets. After obtaining a good grain, we analyzed the datasets together and discussed the possible dimensions, then split up to create the dimension attributes. Afterwards, we discussed and made changes to the dimensional attributes. Finally, we worked together and discussed design mistakes that we made with our current model and made changes.

We worked on this deliverable together on Saturday, February 3rd, and Wednesday, February 7th.

Aside from the mandatory TA meetings, we did not have any additional meetings with the TA.

Additional References

Original Dataset: https://www.statcan.gc.ca/en/lode/databases/odrsf

Ethnicity Statistics:
https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=9810030301&pickMembers%5B0%5D=1.1&pickMembers%5B1%5D=2.2&pickMembers%5B2%5D=3.1&pickMembers%5B3%5D=4.1&pickMembers%5B4%5D=5.1

Age statistics from Province:
https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000501

Sex statistics based on city
https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/dv-vd/pyramid/index-en.htm