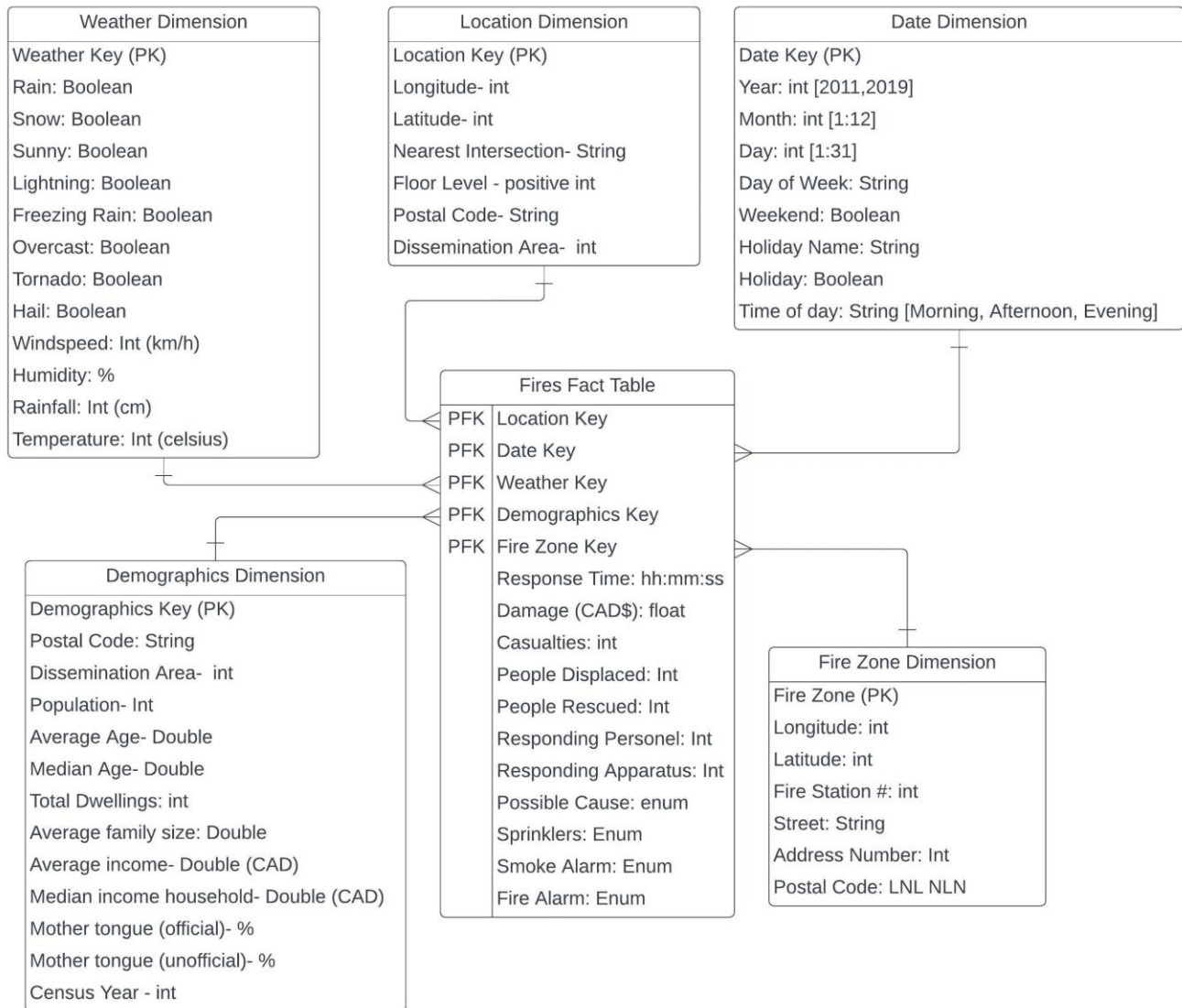


# CSI4142 Project – Conceptual Design – Group 4

Amy Leigh Dunn – 8093779 - ldunn084@uottawa.ca

Émilie Fortin - 6338858 - efort063@uottawa.ca

Jason Su – 300074626 - xsu031@uottawa.ca



## 1. Declare the grain of your data mart.

The grain of our data mart is as follows: A single fire incident happening at a specific location, at a specific time, on a specific date, in certain weather conditions, in a demographic region, with a specific cause, in a specific fire station zone. An incident includes the following measures: response time, damage (dollar amount), damage category, number of people rescued, displaced as well as casualties, number of responding personnel and apparatus, possible cause and, finally, the presence and operation of safety features such as sprinklers, smoke alarms and fire alarms.

## 2. Detail all the dimensions and dimensional attributes. You should list the domains and show sample values. (e.g., Age: integer, minimum = 0 and maximum = 130, Sample value = 35).

The **location dimension**'s domain contains the following:

A unique private key - integer, minimum: 0, maximum: size of the data set (ex. 15320)

Longitude - integer, min: -180, maximum: 180 (ex. -74)

Latitude - integer, min: -90, maximum: 90 (ex. 35)

Nearest intersection: text, sample value: Dixon Rd / 427 N Dixon Ramp

Level of origin - An integer or N/A if the fire didn't occur in a building (ex. 3)

Postal Code - A String representing the postal code in the format LNL NLN (where L represents letter and N represents number (ex. M4C 2J4)

Dissemination area [Smallest subdivision used by the canadian census] - int {range information unfindable - will determine when examining the raw data}

The **date dimension**'s domain contains the following:

A unique private key - integer, minimum: 0, maximum: size of the data set (ex. 13023)

Year - int, min: 2011, max: 2019 (ex. 2018)

A time stamp - in the following format "yyyy-mm-dd hh:mm" [hours in 24h format]  
(ex. 2010-03-14 12:30)

A time of day - text, in {midnight if h=00, night: if 00<h<06, dawn if h=6, morning if 6<h<12, noon if h=12, afternoon if 12<h<18, dusk if h=18, evening if 18<h} (ex. evening)

Month - enum [January, February, March, April, May, June, July, August, September, October, November, December] (ex. April)

Day - int, min:1, max: 31 (ex. 13)

Day of the week: enum [Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday]  
(ex. Thursday)

Holiday - boolean [true, false] (ex. false)

Holiday type - enum [Family Day, Halloween, Christmas, Hanukkah, New Years,...] Can have multiple tags (ex. Halloween)

Weekend - boolean [true, false] (ex. true)

The **weather dimension**'s domain contains the following:

A unique private key - integer, minimum: 0, maximum: size of the data set (ex. 12031)

rain : boolean [true, false] (ex. true)

Freezing rain: boolean [true, false] (ex. false)

Snow: boolean [true, false] (ex. true)

Sunny: boolean [true, false] (ex. false)

Overcast: boolean [true, false] (ex. false)

Tornado: boolean [true, false] (ex. true)

Hail: boolean [true, false] (ex. true)

Lightning/thunder: boolean [true, false] (ex. false)

Windspeed (km/h) : int, min: 0, max: 380 (ex. 230)

Humidity (%): int, min:0, max: 100 (ex. 23)

Rainfall (cm): int, min:0, max: 160 (ex. 14)

Temperature (deg. celsius): int, min: -90, max: 60 (ex. 30)

The **demographics dimension**'s domain contains the following:

A unique private key - integer, minimum: 0, maximum: size of the data set (ex. 14029)

Census year - int, min: 2011, max: 2021, sample data: 2016

Postal Code - A String representing the postal code in the format LNL NLN (where L represents letter and N represents number (ex. M4C 2J4)

Dissemination area [Smallest subdivision used by the canadian census] - int {range information unfindable - will determine when examining the raw data}

Population density per square kilometer, float - min: 0, max: unknown (ex. 2000)

Average Age of the population - float, min: 0, max: 100 (ex. 24)

Median Age of the population - float, min:0, max: 100 (ex. 15)

Total number of dwellings - int min: 0, max: unknown (ex. 2000)

Average household size - float, min: 0, max: 100 (ex. 3.5)

Median total income in 2020 among recipients (\$) - float, min:0, max: 10 000 000 (ex. 10,000.23)

Median total income of household in 2020 (\$) - float , min:0, max: 10 000 000 (ex. 10,000.23)

Mother tongue statistics (%) - official languages - int, min:0, max: 100 (ex. 65%)

Mother tongue statistics (%) - non official languages - int, min:0, max: 100 (ex. 35%)

The **Fire zone dimension**'s domain contains the following:

A unique private key - integer, minimum: 1, maximum: 84 (ex. 34)

A longitude for the station - integer, min: -180, maximum: 180 (ex. 89)

A latitude for the station - integer, min: -90, maximum: 90 (ex. 35)

Fire station street number - number min:1 max: unknown (ex. 28)

Fire station street address - text (ex. St. Clair St.)

Fire station city - text, ex: [toronto, north york, ...]

Fire station Postal Code: A String representing the postal code in the format LNL NLN (where L represents letter and N represents number (ex. M2J 4L1)

3. Detail all the measures/facts. You should list the domains and sample values. (e.g., Age: integer, minimum = 0 and maximum = 130, Sample value = 35).

Response time, time stamp, hh:mm:ss, min: 00:00:01, max: 24:59:59, sample value: 00:21:05

Total Casualties - int , min:0, max:15, sample value: 0

Estimated dollar loss (CAD\$) - float (2 decimals precision), min: 0.00, max: 50000000, sample value: 100

Estimated number of people displaced - int, min: 0, max: 999, sample value: 2

Number of persons rescued - int, min:0, max: 86, sample value:7

Number of responding personnel - int, min:0, max:1275, sample value:5

Number of responding Apparatus - int, min:0, max:436, sample value: 2

Possible Cause - enum, value among: [intentional or unintentional]

Sprinkler System Presence and operation [OFM code] - enum, value among: [No sprinkler system; Sprinkler system present and operated; sprinkler system present, did not operate ; Sprinkler system present, operation undetermined; Sprinkler system presence undetermined]

Smoke Alarm Presence and operation [OFM code] - enum, value among: [No smoke alarm; Smoke alarm present and operated; Smoke alarm present, did not operate ; Smoke alarm present, operation undetermined; Smoke alarm presence undetermined]

Fire Alarm System Operation and presence [OFM code] - enum, value among [No fire alarm; Fire alarm present and operated; Fire alarm present, did not operate ; Fire alarm present, operation undetermined; Fire alarm presence undetermined]

#### 4. Remember to detail all your assumptions.

We assumed that the time of the fire being declared to the station corresponds to the time of the fire, therefore we are assuming that the blaze was noticed immediately and that emergency services were contacted immediately, which may not be accurate.

For the demographics data, we made sure to only include categories that used - 100% data and not ones that used 25% sample data, for increased data quality.

For real world phenomena, such as temperature or wind speed, the minimum and maximum values are slightly larger or smaller than the largest values ever recorded on earth, to ensure that the range would be sufficiently large.

Since the census data is only gathered every 5 years, it does not exactly reflect the demographics data found for each dissemination area, at the exact date of the fire. We are assuming that this demographic data is accurate enough by rounding the date to the nearest census date.

Whenever a measure is followed by [OFM code], it indicates that this is an official classification by the Ontario Fire Marshal's office. These codes are represented by an enum describing what each code represents will be found in detail in the descriptions and metadata.

5. Make a checklist (Use Tables) of the “10 design mistakes” mentioned at the end of Module 1, Part C and show how you avoided/handled those mistakes.

Mistake	How we avoided/handled the mistakes
1) Place Text Attributes in the Fact table	All text attributes are changed to enum values.
2) Limit verbose descriptions to save space	All relevant details will be included regardless of space requirements.
3) Normalise to save space	We ensured that no dimension was normalized and we will follow a star configuration.
4) Ignore the need to track changes	All data involved is historical, therefore no changes will need to be made over the course of this project.
5) Add new hardware to solve all query performance issues	We ensured that our database was not normalized to prevent slowing down of queries due to joins.
6) Use operational keys as primary keys	We created a unique Private key for each dimension to replace operational keys. If necessary for the dimension, we will use a table to link the unique private key to the current operational key.
7) Neglect to declare (and comply with) the grain	We declared our grain and we will ensure that we comply with it by comparing it to the datamart at different stages.
8) Neglect a detailed design	We have created a detailed conceptual model. In the future, we will create detailed models of the data cleaning and transformation.
9) Expect users to query normalised data (repeat of 3)	See point 3.
10) Fail to conform Facts and Dimensions	Most facts and dimensions are conformed already, but we have chosen specific formats and parameters to ensure consistency.

6. A summary of your team's work plan, including the times and dates you met, how you divided the work, and how you often meet with the TA.

Dates we met up:

- Friday, February 3rd 2023
- Monday, February 6th 2023
- Wednesday, February 8th 2023
- Friday, February 10th 2023
- Sunday, February 12th 2023

Dates we met with the TA:

- Monday, February 6th 2023

Name	Contributions
Amy Dunn	Contacted the creator of the data set to resolve ambiguity about certain categories, researched techniques for building the data mart, completed the design decisions checklist
Émilie Fortin	Worked on detailing our assumptions, detailing each of the dimensions and detailing the measures
Jason Su	Created the diagram (1st - 4th draft), declaring the grain, aided with detailing each of the dimensions and detailing the measures

7. Add a list of additional references you used when creating your model

<https://open.toronto.ca/dataset/fire-incidents/>  
<https://open.toronto.ca/dataset/fire-station-locations/> <https://open.toronto.ca/dataset/toronto-fire-services-run-areas/> [https://climate.weather.gc.ca/historical\\_data/search\\_historic\\_data\\_e.html](https://climate.weather.gc.ca/historical_data/search_historic_data_e.html) <https://www.geocoder.ca/?locate=43.6865581769999%2C+-79.599419244&geoit=GeoCode> <https://www.ontario.ca/page/office-fire-marshal>  
[Census Profile \(statcan.gc.ca\)](#)  
[Census Profile, 2016 Census \(statcan.gc.ca\)](#)  
<https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/index.cfm?Lang=E>  
[Fire Station Locations – City of Toronto](#)  
[Intelligent Diagramming | Lucidchart](#)  
<https://stackoverflow.com/questions/37926684/entering-value-into-search-bar-and-downloading-output-from-webpage>