

# **CSI 4142 Fundamentals of Data Science**

## **Project Phase 1: Conceptual Design – Dimensional Model**

**University of Ottawa**

**School of Computer Science**

**Professor : Herna L Viktor**

Student Name : Jiajie Xu  
Student Number: 7881937

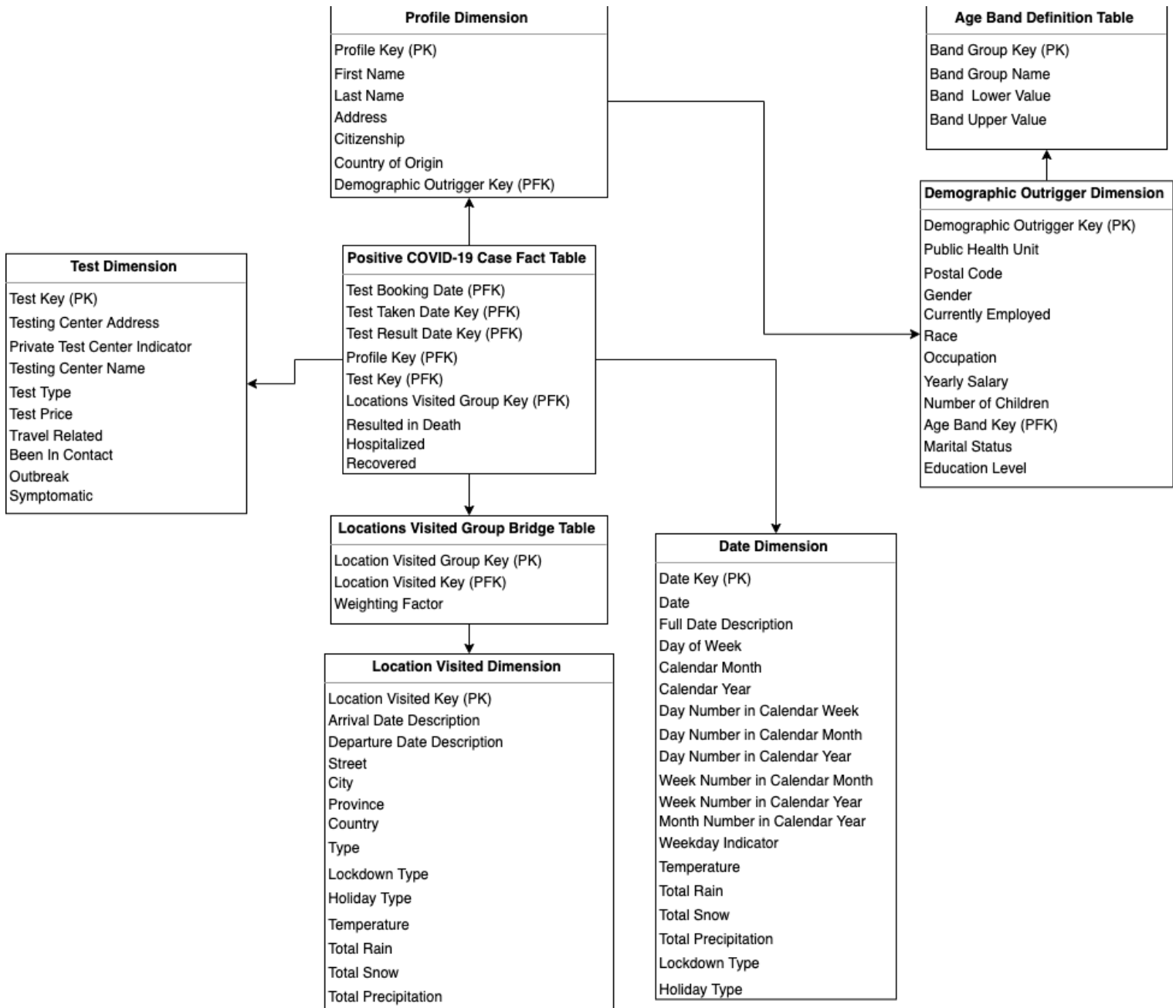
Student Name : Ali Khanafer  
Student Number: 300010614

Student Name : Vidulash Rajaratnam  
Student Number: 8190398

## Grain

Profile of an individual person who booked his COVID-19 test on a certain date, took it on another, at a certain testing center and received his results as positive on a certain day. The fact table also describes all the locations the infected person visited

## Conceptual Model



## Measures and Facts

**Table: Date Dimension**

Attribute	Type	Domain	Sample Value
Date Key	/	/	/
Date	Date	YYYY/MM/DD Where, 2019 <= YYYY, 01 <= MM <= 12 and 01 <= DD <= 31	2019/02/21
Full Date Description	String	/	"February 21st 2019"
Day of Week	String	{"Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday"}	"Monday"
Calendar Month	String	{"January", "February", "March", "April", "May", "June", "July", "August", "September", "October", "November", "December"}	"January"
Calendar Year	Number	Value >= 2019	2019
Day Number in Calendar Week	Number	1 <= Value <= 7	3
Day Number in Calendar Month	Number	1 <= Value <= 30	21
Day Number in Calendar Year	Number	1 <= Value <= 365	212
Week Number in Calendar Month	Number	1 <= Value <= 4	2
Week Number in Calendar Year	Number	1 <= Value <= 52	3
Month Number in Calendar Year	Number	1 <= Value <= 12	1
Weekday Indicator	Boolean	True/False	True

<b>Temperature</b>	Number	-40 <= Value <= 40	25
<b>Total Rain:</b>	Double	Value >=0	14
<b>Total Snow:</b>	Double	Value >=0	0
<b>Total Precipitation</b>	Number	0 <= Value <= 30	0
<b>Lockdown_type</b>	String	{“Green”, “Yellow”, “Orange”, “Red”, “Grey”}	“Red”
<b>Holiday_type</b>	String	{“Thanksgiving”, “Christmas”, “New Year’s Day”, “Canada Day”, “Labour Day”, “Good Friday”, “Victoria Day”, “Remembrance Day”, NULL}	“New Year’s Day”

**Table: Test Dimension**

<b>Attribute</b>	<b>Type</b>	<b>Domain</b>	<b>Sample Value</b>
<b>Test Key</b>	/	/	/
<b>Test Center Address</b>	String	/	“151 Brewer Way, Ottawa”
<b>Private Test Center</b>	Boolean	True/False	True
<b>Test Price</b>	Double	Value >=0	100.0
<b>Test Type</b>	String	{“PCR”, “Serology”}	“PCR”
<b>Test Center Name</b>	String	/	“Ottawa test center”
<b>Travel_related</b>	Boolean	True/False	True

<b>Been in Contact</b>	Boolean	True/False	False
<b>Outbreak</b>	Boolean	True/False	True
<b>Symptomatic</b>	Boolean	True/False	False

**Table: Location Visited Dimension**

<b>Attribute</b>	<b>Type</b>	<b>Domain</b>	<b>Sample Value</b>
<b>Location_key</b>	/	/	/
<b>Arrival Date Description</b>	String	/	"February 2nd 2019"
<b>Departure Date Description</b>	String	Value >= Arrival Date Description	"February 3rd 2019"
<b>Street</b>	String	/	"Bank Street"
<b>City</b>	String	/	"Ottawa"
<b>Province</b>	String	/	"Ontario"
<b>Country</b>	String	/	"Canada"
<b>Type</b>	String	{ "Retail and Recreation", "Grocery and Pharmacy", "Parks", "Transit Stations", "Workplaces", "Residential" }	"Parks"
<b>Holiday_type</b>	String	{ "Thanksgiving", "Christmas", "New Year's Day", "Canada Day", "Labour Day", "Good Friday", "Victoria Day", "Remembrance Day", NULL }	"New Year's Day"
<b>Lockdown_type</b>	String	{ "Green", "Yellow", "Orange", "Red", "Grey" }	"Red"
<b>Temperature</b>	Number	-40 <= Value <= 40	25

<b>Total Rain:</b>	Double	Value $\geq 0$	14
<b>Total Snow:</b>	Double	Value $\geq 0$	0
<b>Total Precipitation:</b>	Double	Value $\geq 0$	10

**Table: Profile Dimension**

Attribute	Type	Domain	Sample Value
<b>Profile_key</b>	/	/	/
<b>First Name</b>	String	/	“John”
<b>Last Name</b>	String	/	“Doe”
<b>Address</b>	String	/	“75 Laurier Ave. E”
<b>Citizenship</b>	String	/	“Canadian”
<b>Demographic Outrigger Key</b>	<b>Link to Demographic Outrigger dimension</b>		
<b>Country_of_origin</b>	String	/	“Japan”

**Table: Age-band dimension**

Attribute	Type	Domain	Sample Value
<b>Age_band_key</b>	/	/	/
<b>Age_group</b>	String	{“Children”, “Youth”, “Adults”, “Seniors”}	“Adult”
<b>Minimum_age</b>	Integer	$0 \leq \text{Value} \leq 130$	5
<b>Maximum_age</b>	Integer	$0 \leq \text{Value} \leq 130$	80

**Table: Demographic Outrigger Dimension**

Attribute	Type	Domain	Sample Value
<b>Demographic Outrigger Key</b>	/	/	/
<b>Public Health Unit</b>	String	{As Defined by Ontario's Ministry of Health}	"Erie St. Clair"
<b>Postal Code</b>	String	Value Length = 6	"K1N 0K8"
<b>Gender</b>	String	{"Male", "Female", "Other", NULL}	Female
<b>Currently Employed</b>	Boolean	True/False	True
<b>Race</b>	String	{"Aboriginal", "Asian", "Middle Eastern", "Black or African American", "Hispanic or Latino", "White", "Other"}	"Other"
<b>Occupation</b>	String	/	"Student"
<b>Yearly Salary</b>	Number	$0 \leq \text{Value} \leq 10000000$	130000
<b>Number of Children</b>	Number	$0 \leq \text{Value} \leq 12$	2
<b>Marital Status</b>	String	{"Married", "Single", "Widowed", "Divorced", "Other"}	"Single"
<b>Education Level</b>	String	{"High school", "Bachelor", "Masters", "Doctorate", "No degree"}	"College"
<b>Age Band Key</b>	<b>Link to Age Band dimension</b>		

**Table: COVID-19 Daily Results Fact Table**

Attribute	Type	Domain	Sample Value
Test Booking Date Key	/	/	/
Test Taken Date Key	/	/	/
Test Result Date Key	/	/	/
Profile Key	/	/	/
Test Key	/	/	/
Resulted in Death	Boolean	True/False	False
Hospitalized	Boolean	True/False	True
Recovered	Boolean	True/False	True

## Assumptions

- Any entry in the data mart is of a person who has already tested positive for COVID-19.
- Since the project description says we're tracking the positive COVID-19 cases in two big Canadian cities, we chose to work with Toronto and Ottawa. We chose those two because they share the same time zone and are both in Ontario, so we can keep track of a person's public health unit
- Holidays would account for all possible nationwide holidays that would affect all provinces in Canada, particularly with a priority in Ontario. Holidays that are only observed in select provinces are not included (Ex. Islander Day in PE).
- COVID-19 test types are based on the 2 main methods used in Ontario, such as PCR test using samples collected via swabs, and serology testing which included someone's blood sample.
- Lockdown type is based on the system incorporated in Ontario, which is based on a specific color-code. Green for prevent, yellow for protect, orange for restrict, red for control and grey for full lockdown.
- Hospitalization includes individuals who have been admitted to the ICU department and those who have visited the emergency department as a result of covid related symptoms. Those who have received a positive test and have not visited the hospital are not considered in this measure.



## Work Plan

### **Division of Work:**

#### **Everyone contributed equally in:**

- Determining the grain of the conceptual design
- Determining dimensions and attributes for the tables
- Determining values and measures for the fact table

#### **Ali Khanafer:**

- Designing the conceptual model
- Updating dimension and fact tables
- Giving input on the conceptual design

#### **Jiajie Xu and Vidulash Rajaratnam:**

- Making the dimension and fact tables
- Updating dimension and fact tables
- Overall, giving input on the conceptual design

### **Meeting dates:**

- 01/30/2021 1:00-2:00 pm
  - Planned to discuss about the overall structure of conceptual design
  - Everyone got to discuss their own ideas and opinions for the project
  - We combined our ideas and came up with the basic outline for the grain and the individual dimension tables
- 02/02/2021 2:00-3:00 pm
  - Finalized our dimension and fact tables along with the grain for the model
  - Designed the conceptual model and made modifications accordingly until all group members were satisfied
- 02/03/2021 1:00-1:30 pm
  - Double-checked the grain, tables and the conceptual model
  - Added final modifications to the dimension table attributes
  - Final meeting to check if all the requirements for the deliverable were met
  - Put together the final report for submission

## References

### 1. Datasets

- a. <https://www.convertcsv.com/csv-viewer-editor.htm>
- b. <https://www.convertcsv.com/csv-viewer-editor.htm>
- c. [https://www.gstatic.com/covid19/mobility/2021-01-26\\_CA\\_Ontario\\_Mobility\\_Report\\_en.pdf](https://www.gstatic.com/covid19/mobility/2021-01-26_CA_Ontario_Mobility_Report_en.pdf)
- d. [https://climate.weather.gc.ca/climate\\_data/daily\\_data\\_e.html?hlyRange=%7C&dlyRange=1994-06-01%7C2008-12-27&mlyRange=1994-01-01%7C2006-12-01&StationID=26775&Prov=ON&urlExtension=\\_e.html&searchType=stnProv&optLimit=yearRange&StartYear=2000&EndYear=2021&selRowPerPage=25&Line=0&Month=12&Day=30&lstProvince=ON&timeframe=2&Year=2008](https://climate.weather.gc.ca/climate_data/daily_data_e.html?hlyRange=%7C&dlyRange=1994-06-01%7C2008-12-27&mlyRange=1994-01-01%7C2006-12-01&StationID=26775&Prov=ON&urlExtension=_e.html&searchType=stnProv&optLimit=yearRange&StartYear=2000&EndYear=2021&selRowPerPage=25&Line=0&Month=12&Day=30&lstProvince=ON&timeframe=2&Year=2008)
- e. <https://www.arcgis.com/home/item.html?id=26c902bf1da44d3d90b099392b544b81>
- f. <https://data.ontario.ca/dataset/status-of-covid-19-cases-in-ontario-by-public-health-unit-phu/resource/d1bfe1ad-6575-4352-8302-09ca81f7ddfc>
- g. <https://data.ontario.ca/dataset/covid-19-assessment-centre-locations/resource/c60993bb-3988-4648-9be9-398dee480514>

### 2. Dashboards

- a. <https://www.ottawapublichealth.ca/en/reports-research-and-statistics/daily-covid19-dashboard.aspx>