

# University of Ottawa

## School of Electrical Engineering and Computer Science

### CSI4142 Fundamentals of Data Science

#### Project Phase 2: Physical Design and Data Staging

Student Name: Xia (Cynthia) Sheng

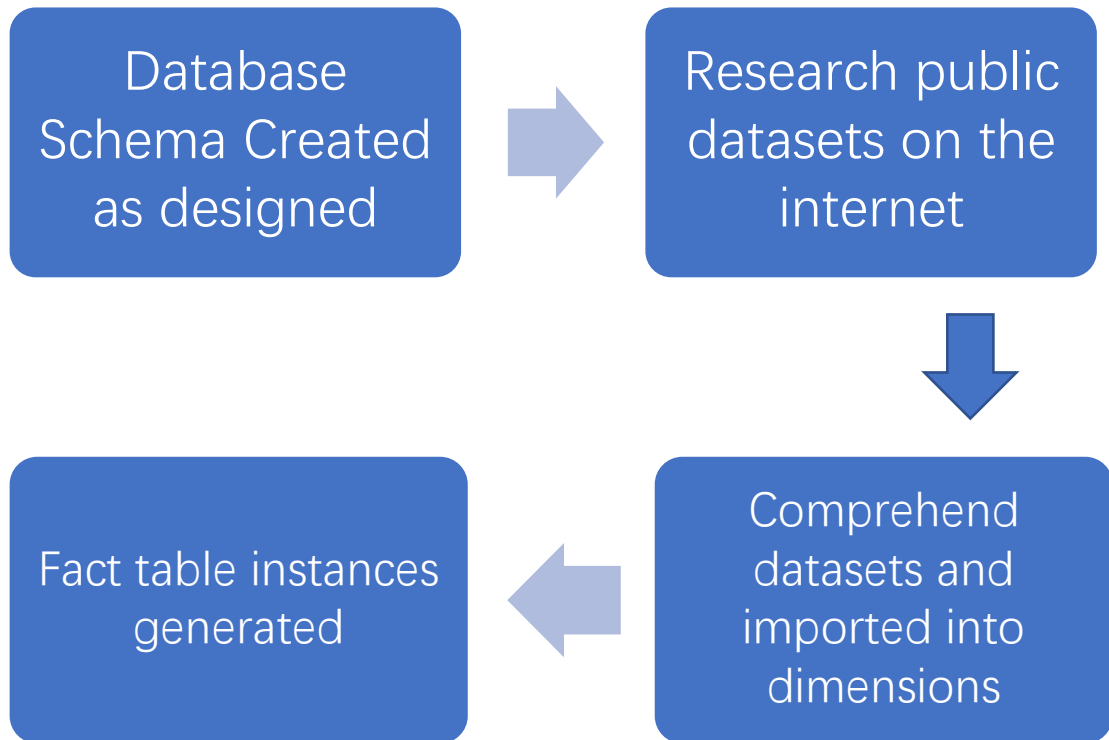
Student ID: 300091655

Student Name: Xinyu Ji

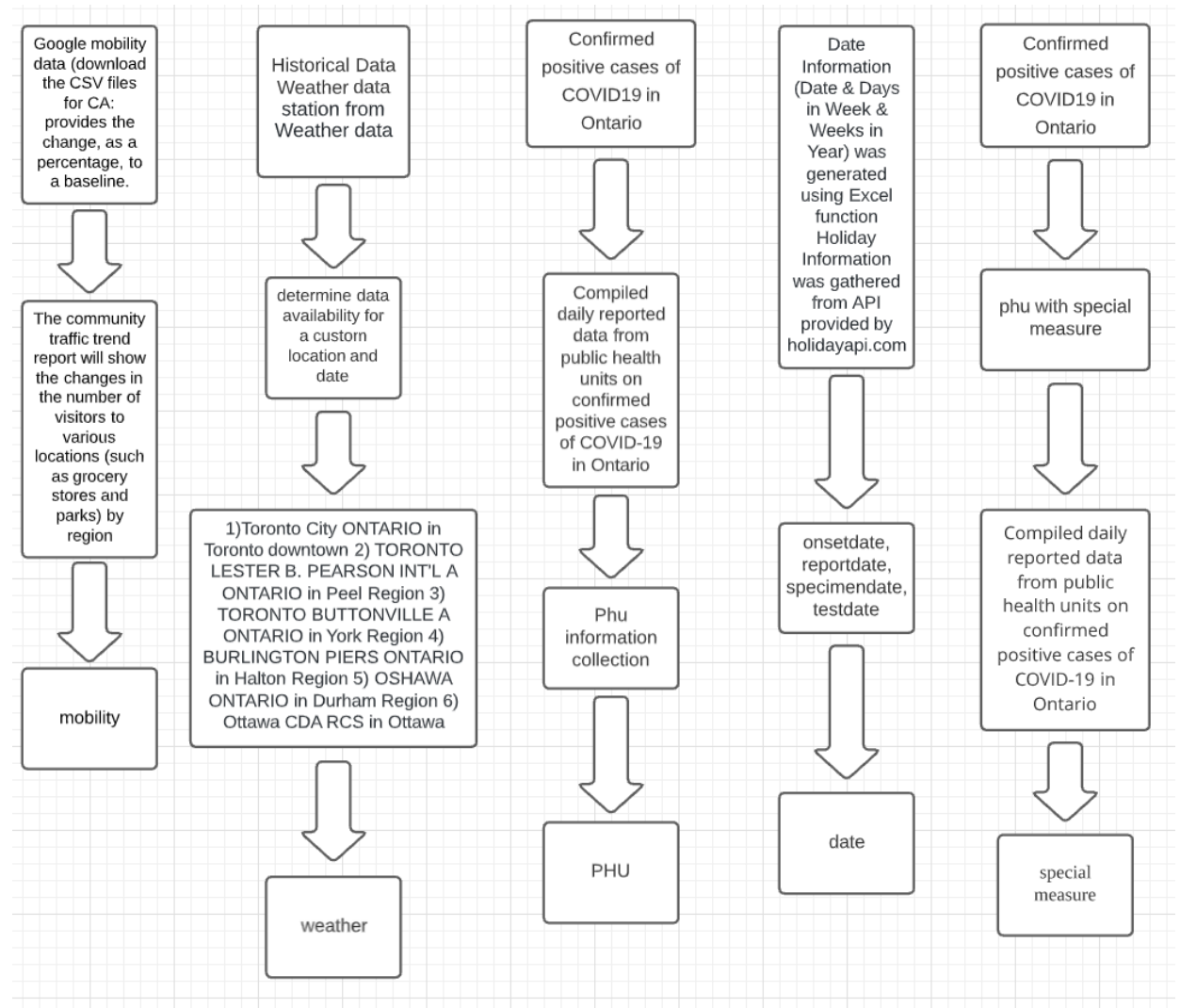
Student ID: 300061354

Student name: Heng (Scott) Zhang

Student number: 300067988



## high-level data staging plan



Data quality issues	Who solved	Solutions
Missing data in weather dimension	Xia (Cynthia) Sheng	<ul style="list-style-type: none"> <li>All missing 'snow_on_ground' and 'total_recipitation' are filled with 0.</li> <li>all missing temperature are filled with mean value of the day before and after of the missing date.</li> </ul>
Missing Data in Fact Table and Positive Case	Heng (Scott) Zhang	<p>1. Some data may not have specimen date or test date, in case of this circumstance, we use report date as their specimen data/test date</p> <p>2. Some cases have no age info. available, we use 0 during data staging</p>

		3. Gender may have 3 types of value: M for Male, F for Female, U for Unknown

### CSI4142 Physical Design and Data Staging - Project Planning

Deliverable	Team member(s) responsible	Expected completion date	Actual completion date	Estimate time (hours) to complete	Actual time (hours) to complete
Create database instance and tables	Heng (Scott) Zhang ,Xia (Cynthia) Sheng,	2021-02-19	2021-02-18	2	1h45min
Staging of Date dimension	Heng (Scott) Zhang	2021-02-20	2021-02-19	0.5	0.5
Staging of Patient dimension	Heng (Scott) Zhang	2021-02-20	2021-02-19	1	45min
Staging of PHU dimension	Xinyu Ji	2021-02-20	2021-02-19	0.5	0.5
Staging of Mobility dimension	Heng (Scott) Zhang	2021-02-20	2021-02-19	1	0.5
Staging of Weather dimension	Xia (Cynthia) Sheng	2021-02-21	2021-02-19	1	45min
Map PHU, Mobility and Weather dimensions	Heng (Scott) Zhang	2021-02-25	2021-02-19	2	1h45min
Staging of Special Measures dimension	Xinyu Ji	2021-02-20	2021-02-19	0.5	0.5
Surrogate key pipeline - including role-playing dates	Xia (Cynthia) Sheng	2021-02-28	2021-02-19	2	1h45min
Staging of fact table - including FKs and measures	Heng (Scott) Zhang ,Xia (Cynthia) Sheng	2021-02-28	2021-02-19	2	1h45min
Data quality handling and	Xinyu Ji	2021-03-01	2021-02-19	1	1