A list of data quality issues you encountered and how you handled them.

1. There are some missing values in the source data, we decided to set them to null. We use the mean value of a specific class or group to fill the missing values if necessary. We also use the most probable value to fill the missing values.
2. There are some values that are not in the correct formats, we decided to set a standard unit of measure.
3. Different regions have different values of mobility.
4. As for the special measure, there may be different states in a time interval and the latter state will overwrite the previous state. For example, in a certain period, the area is first controlled and then it is locked down. Then its state will not only have the value of lockdown, but also has the value of control. Also, different areas may have different time zone for special measure.
5. PHU reports based on the Peel district of the location, which is a clinic in the Peel district.
6. As for handling the date quality issue, we filtered out the values of non-compliant areas, which are areas outside of Ottawa and the Greater Toronto Area.

CSI4142 Physical Design and Data Staging - Project Planning

Group 26

Deliverable	Team member(s) responsible	Expected completion date	Actual completion date
Create database instance and tables	Xingpeng Wang	2021/3/10	2021/3/9
Staging of Date dimension	Xingpeng Wang	2021/3/10	2021/3/8
Staging of Patient dimension	Yifei Wang	2021/3/10	2021/3/9
Staging of PHU dimension	Xingpeng Wang	2021/3/10	2021/3/9
Staging of Mobility dimension	Yifei Wang	2021/3/10	2021/3/8
Staging of Weather dimension	Tianqi Chen	2021/3/10	2021/3/8
Map PHU, Mobility and Weather dimensions	Xingpeng Wang	2021/3/10	2021/3/9
Staging of Special Measures dimension	Yifei Wang	2021/3/10	2021/3/8
Surrogate key pipeline – including role-playing dates	Xingpeng Wang	2021/3/10	2021/3/9
Staging of fact table – including FKs and measures	Yifei Wang	2021/3/10	2021/3/8
Data quality handling and reporting	Tianqi Chen	2021/3/10	2021/3/11
Others - if any			

Estimate time (hours) to complete	Actual time (hours) to complete	Notes (if any)
5 hours	3 hours	
5 hours	2 hours	
5 hours	1 hour a half	
5 hours	3 hours	
5 hours	2 hours	
5 hours	3 hours	
5 hours	1 hour a half	
5 hours	3 hours	
5 hours	2 hours	
5 hours	3 hours	
5 hours	1 hour a half	