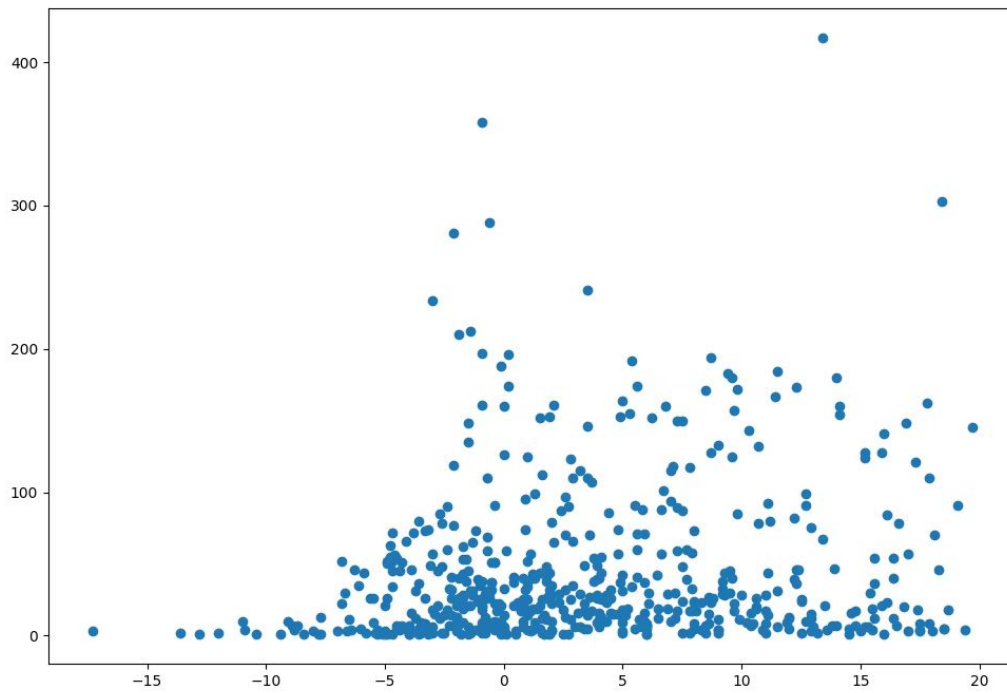
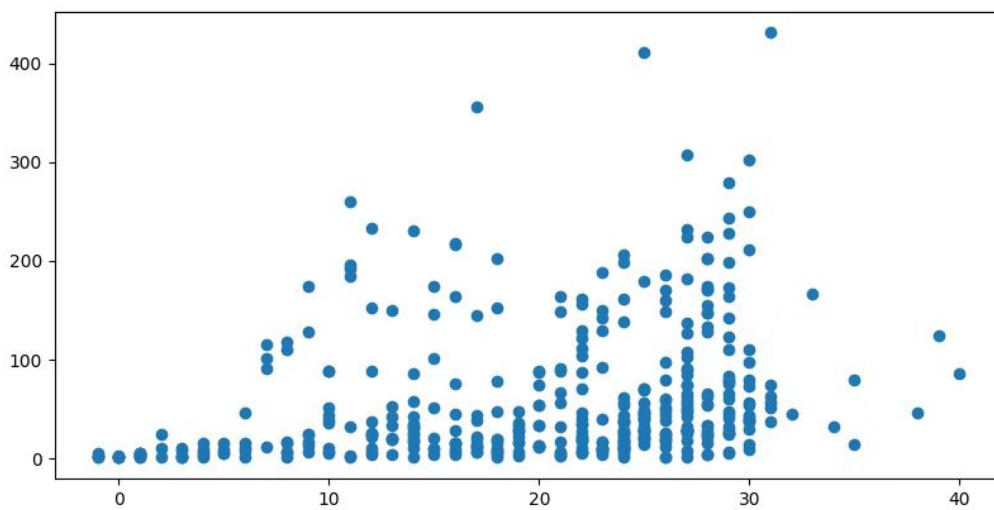


Project Phase 4: Data Mining

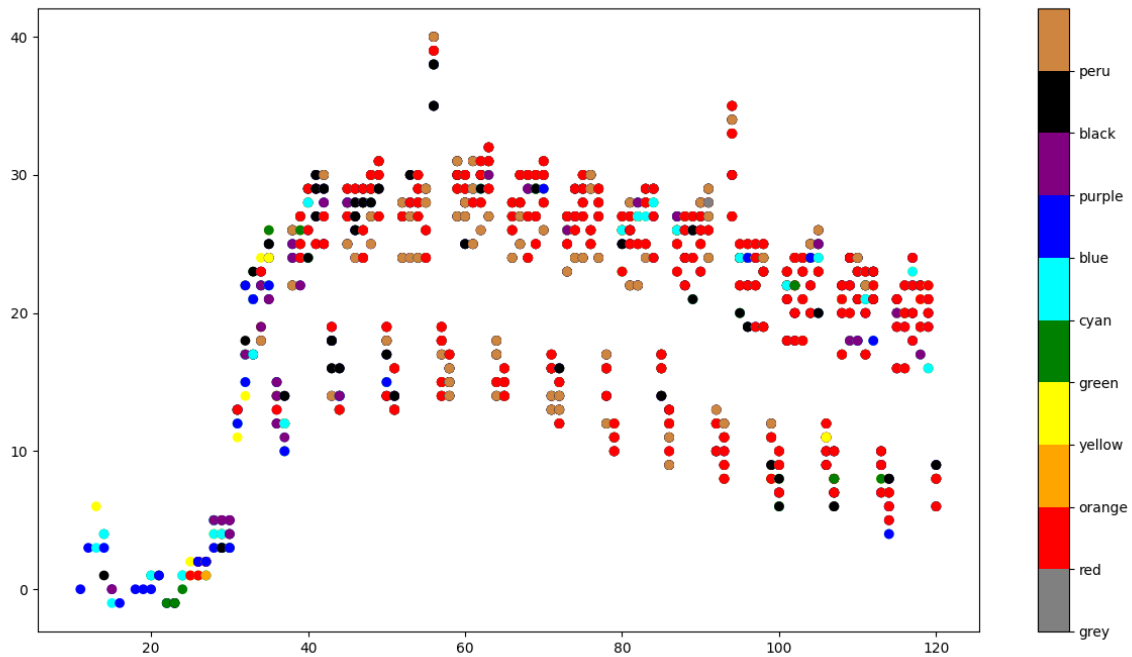
Part A



This is the graph of reported cases for each temperature level



This is the number of cases by percent increase of residential



This is the graph of cases to mobility with color denoting age group

We don't have any missing value in the dataset due to the data staging phase and we don't have any categorical attributes. However, we did some normalisation of numeric attributes. For the residential data we have the maximum number of 40 and minimum number of -4 so we rescaled it using $(x - (-4))/40$. For the low temperature the maximum is 19.7 and the minimum is -21.1 so we rescaled it using $(x - (-21.1))/40.8$. So we have both the data with minimum value of 0 and maximum value of 1.

Part B

These were the results from using daily low temperature and residential mobility to try and predict age group.

Decision Tree Results:

Time to complete: 0.007001399993896484

	precision	recall	f1-score	support
0	0.00	1.00	0.00	0
1	0.00	0.00	0.00	87
2	0.00	0.01	0.00	160
3	0.02	0.02	0.02	168
4	0.01	0.01	0.01	152
5	0.02	0.01	0.01	171
6	0.06	0.01	0.02	149
7	0.00	0.00	0.00	138
8	0.00	0.00	0.00	128
9	1.00	0.00	0.00	97
accuracy			0.01	1250
macro avg	0.11	0.10	0.01	1250
weighted avg	0.09	0.01	0.01	1250

Gradient Boosting Results:

Time to complete: 0.5291285514831543

	precision	recall	f1-score	support
1	0.00	0.00	0.00	87
2	0.05	0.02	0.03	160
3	0.20	0.01	0.02	168
4	0.11	0.65	0.19	152
5	0.11	0.14	0.12	171
6	0.05	0.02	0.03	149
7	0.00	0.00	0.00	138
8	0.00	0.00	0.00	128
9	1.00	0.00	0.00	97
accuracy			0.10	1250
macro avg	0.17	0.09	0.04	1250
weighted avg	0.15	0.10	0.05	1250

Random Forest Results:

Time to complete: 0.033007144927978516

	precision	recall	f1-score	support
1	0.00	0.00	0.00	87
2	0.02	0.02	0.02	160
3	0.00	0.00	0.00	168
4	0.00	0.00	0.00	152
5	0.04	0.04	0.04	171
6	0.01	0.01	0.01	149
7	0.00	0.00	0.00	138
8	0.00	0.00	0.00	128
9	0.00	0.00	0.00	97
accuracy			0.01	1250
macro avg	0.01	0.01	0.01	1250
weighted avg	0.01	0.01	0.01	1250

Overall, these results were not terribly good as far as precision and recall go. It is perhaps noteworthy that the results that tended to give the most precision and recall were in the 40s and 50s age groups. This could be attributed to the fact that they make up the largest age groups in Canada. For some reason, the 90s age group sometimes gave 100% precision with 0 percent recall.