**Brainstorm ideas**

* Divorce rate
* Community mobility (Google)
* Music preferences
* Students’ grades

**Introduction**

**Questions**

* What are the mobility trends in Canada?
  + How do mobility trends change as time progresses?
  + How do mobility trends differ across provinces and territories in Canada?
  + Which provinces were the most/least active in social distancing?
* What factors drove these trends?
  + Number of cases in Canada
  + Number of cases around the world
  + Temperature
  + Precipitation
  + Weather
  + Number of days into social-distancing
  + Population density
  + Urbanness
  + Economy
* Can we cross reference the mobility trends with Google Trends?
* Apple Mobility Trends Report?
* Do these trends reflect the number of cases increases in Canada?
  + Social-distancing score

**Data Collection**

* Google Community Mobility Reports
* WHO COVID-19 map data
* Government of Canada COVID-19 data
* Government of Canada Ontario Toronto City station weather data

**Data Cleaning**

* Lesson learned: always check for null values from the start

**Data Exploration**

**Modeling**

* Linear Regression (fast)
* Decision Tree (fast)
* Random Forest (accurate but slow)
* Gradient Boosting Tree (accurate but slow)

**Evaluation**

**Deployment**

**To-do**

* What should we explore? What plots should we make?
  + What places did people visit the most? Bar plot
  + Cross reference most/least popular places with Google Trends. Line plot
  + How do the mobility trends vary from day to day across different categories in Canada/provinces and territories? Line plot
* Devise an evaluation metric to measure how well did people do in social distancing – Social Distancing score
  + Visualize social distancing in map. Grace? Geospatial data? Google map API?
  + How does social distancing score vary from day to day across categories in Canada/provinces and territories? Line plot
* Build a statistical model to investigate what factors contribute to the social distancing score
  + Multiple Linear Regression
    - P-value
  + Random Forest Regression
    - Permutation Feature Importance

**Problems encountered**

* How to analyze weather (i.e. temperature, precipitation, etc.) of a country?
* We couldn’t collect weather data for a country. Should we narrow down our model from predicting the social-distancing score of a country to the social-distancing score of a province (e.g. Ontario)?
* WHO COVID-19 data and Government of Canada Ontario Toronto City station weather data does not have an active link for download.
* How to create a valid social-distancing score?
* What is the best way to visualize data?
* Is there a more efficient way to clean data?