1. Data cleaning / Feature engineering. In order
   1. Weather table: change the event column to two columns:
      1. Change to rain 1/ 0 (based on whether ‘Rain’ or ‘Thunderstorm’ is in it) and fog 1 / 0 (based on whether ‘Fog’ is in event column). If it is empty or doesn’t have it, make it 0.
   2. Weather: change the precipitationsuminches column:
      1. Currently, it is a string. Some of the entries are ‘T’ for “Trace”. Change Trace to something more meaningful, and make the whole column a float.
   3. For the b-cycle table, there are missing checkout\_kiosk\_id’s. Can we add in the checkout\_kiosk\_id from the name of the checkout station (there’s a table containing all the checkout station names and kiosks)
   4. Features to delete
      1. B-cycle table
         1. Trip id
         2. Bicycle id
         3. Yearmonth
      2. Weather table
         1. Dewpoint
         2. Remove everything but average humidity
         3. Sea level pressure
         4. Average visibility
         5. Only keep average wind speed
   5. B-cycle row removal
      1. Trim rides that last longer and shorter (from the same station) than a certain amount.
      2. Trim membership types and things like that?
      3. Possibly trim year?
   6. B-cycles: remove rows with empty values
      1. Truncate dates
   7. Join weather and B-cycle tables on date?
2. Data mining preparation
   1. Distances between stations
      1. Could we do this with google maps? Could we do this with latitude longitude
   2. Distance in dates could be modulo.
   3. Standardize / normalize
3. Data mining ideas
   1. Unsupervised learning: do we get any interesting clusters?
   2. Supervised learning
      1. Picking a station & the time and such, can we predict where they’re going / how long the trip will be.