* Identify a station(s) of interest (the longest wait time station for-example) for lets say the three most important clusters
* Determine what that cluster’s characteristics are
* What interventions can we add for that cluster
* If we do the intervention, what is it’s impact (does it reduce the wait times, can it make the prediction of waiting over 5 min or longer), if we increase (double, 1.5x) capacity we can recompute the amount of times you have to wait longer than 5 min, and determine the percent difference)
  + Assume station gets restocked every day (make up a cost to restock station ($150)) compare it to the one time fixed cost of adding extra docks (lets say $3000) and look at monthly cost differential to implement fix
* Check correlation between wait time and inflow/outflow
* In prediction model, double capacity and determine the number of 1’s that have been decreased or increased
* Run clusters with prediction
* Add features (average trip duration per start station, average trip duration for end station) after train test split
* Make train/test split into two CSV files for everyone to work off of

Current Visuals

Clusters – All Data

Clusters – Weekly snapshots to see if stations move clusters