Homework 5 - Extra Credit - 75 points

Due: Friday, December 20th, 3PM EST

Note:
I will only grade this homework if you do not have an A.

DASK

1. **Same as Homework 2, Question 7, In Dask - 25 points**

   Solve:
   1. Find food service and active restaurants ("status" = "ACTIVE" **and** ""rpt_area_desc" = "Food Service") ***closest*** to the following coordinate: **of 35.994914, -78.897133,** and show the first 20
   2. With that restaurant in (a) as your center point, find the number of foreclosures within a 1 mile radius

   You can use external libraries for calculating coordinate distances.

   For Python notebooks, the ***haversine*** library is available in Jupyterhub's bigdata environment.

2. **Same as HW1 Q2.1, Language Models, in Dask – 50 points**

   Input: hw1text.zip (provided in class website)
   Solve: conditional probability of w2 given w1, P(w2|w1)

From HW1:

Usually we are not as interested in the probability of a single word, but instead on the conditional probability of, say, 'york' given that the word 'new' precedes it. Specifically, we say P(york|new).

SOLVE (**using DASK**):

Compute the top 3 most likely words following "new".