**COMS E6111-Advanced Database Systems, Spring 2014**

**Project 3**

Abhinav Bajaj (ab3900), Rafica Abdul Rahim (ra2688)

# List of Files for submission -

apriori.py - File with the implementation code of A-priori algorithm.

INTEGRATED-DATASET.csv – generated data set file for our project

README.pdf - This file

Makefile - File containing the steps to compile and run the code on CLIC machines

example-run.txt- File containing output of the interesting sample run

# Description of Data-Set

## Data-Set Selection

We used “Restaurant Inspection Results” data from NYC Open Data data set(s).

The data set is available [here](https://data.cityofnewyork.us/Health/Restaurant-Inspection-Results/4vkw-7nck?)

## INTEGRATED-DATASET Extraction

## “INTERESTING” about Data-Set

We considered many data-sets before finalizing this data-set. Few of them are the “School Demographics and Accountability”, “NYS Math Test Results”, “ELA Test Results” and more. The other most interesting data we could think of was combining the “Demographic Statistics By Zip Code” and “Natural Gas Consumption by ZIP Code”.

# Running the program -

The project is in Python.

Run Command Options -

python apriori.py INTEGRATED-DATASET.csv <min\_sup> <min\_conf>

For example -

python apriori.py INTEGRATED-DATASET.csv 0.3 0.5

# Design Description

The major work flow or design of the program is presented in the flow diagram below

# Interesting Sample Run

## Command line specification

## Explanation of “interestingness”