**Assignment: Software Installation and Using Open Refine**

**1.1 FileZilla Or CyberDuck**

A computer screen shot of a colorful cloud

Description automatically generated

A screenshot of a computer

Description automatically generated

**1.2 OpenRefine**

A computer screen shot of a colorful cloud

Description automatically generated

**3) MySQL (server + workbench)**

A computer screen shot of a colorful cloud

Description automatically generatedA screenshot of a computer

Description automatically generated

**4) Anaconda (Open Data Science Platform: Python )**

A screenshot of a computer program

Description automatically generated

2.a Import the data into OpenRefine and create a new project "SandyCleanup"

A screenshot of a computer

Description automatically generated

b. Remove columns where majority of the cells are empty or have "Unspecified" or “NA” values

( Do not remove the columns that are needed to complete the rest of this exercise )

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

c. Trim white spaces on all address related columns and transform addresses into title case

A screenshot of a computer

Description automatically generated

d. Convert City to title case, then Cluster and Merge the column

A screenshot of a computer

Description automatically generated

e. Clean up Descriptor Column - Cluster and Merge following text categories:

1. "Other Water problem(WZZ)", "Other Water problem(QZZ)" as "Other Water Problem"

A screenshot of a computer

Description automatically generated

2. "Commercial 421 A/B Exemptions" as "Commercial Exemption"

A screenshot of a computer

Description automatically generated

3. "Commercial Exemption" "Commercial Other Exemption" as "Commercial Exemption"

A screenshot of a computer

Description automatically generated

4. "Personal DRIE Exemption", "Personal SCHE Exemption", "Personal DHE Exemption" as "Personal

Exemption"

A screenshot of a computer

Description automatically generated

f. Clean up Location Type - Cluster and Merge following text categories:

1. "Comercial", "Commercial", "Store/Commercial" as "Commercial"

A screenshot of a computer

Description automatically generated

2. "RESIDENTIAL BUILDING", "Residential Building", "Residence" as "Residential"

A screenshot of a computer

Description automatically generated

3. "Club/Bar/Restaurant", "Bar/Restaurant", "Restaurant" as "Club/Bar/Restaurant"

A screenshot of a computer

Description automatically generated

4. "3+ Family Apt. Building", "3+ Family Apartment Building" as "3+ Family Apartment"

A screenshot of a computer

Description automatically generated

5. "Street/Sidewalk", "Street and Sidewalk" as "Street/Sidewalk"

A screenshot of a computer

Description automatically generated