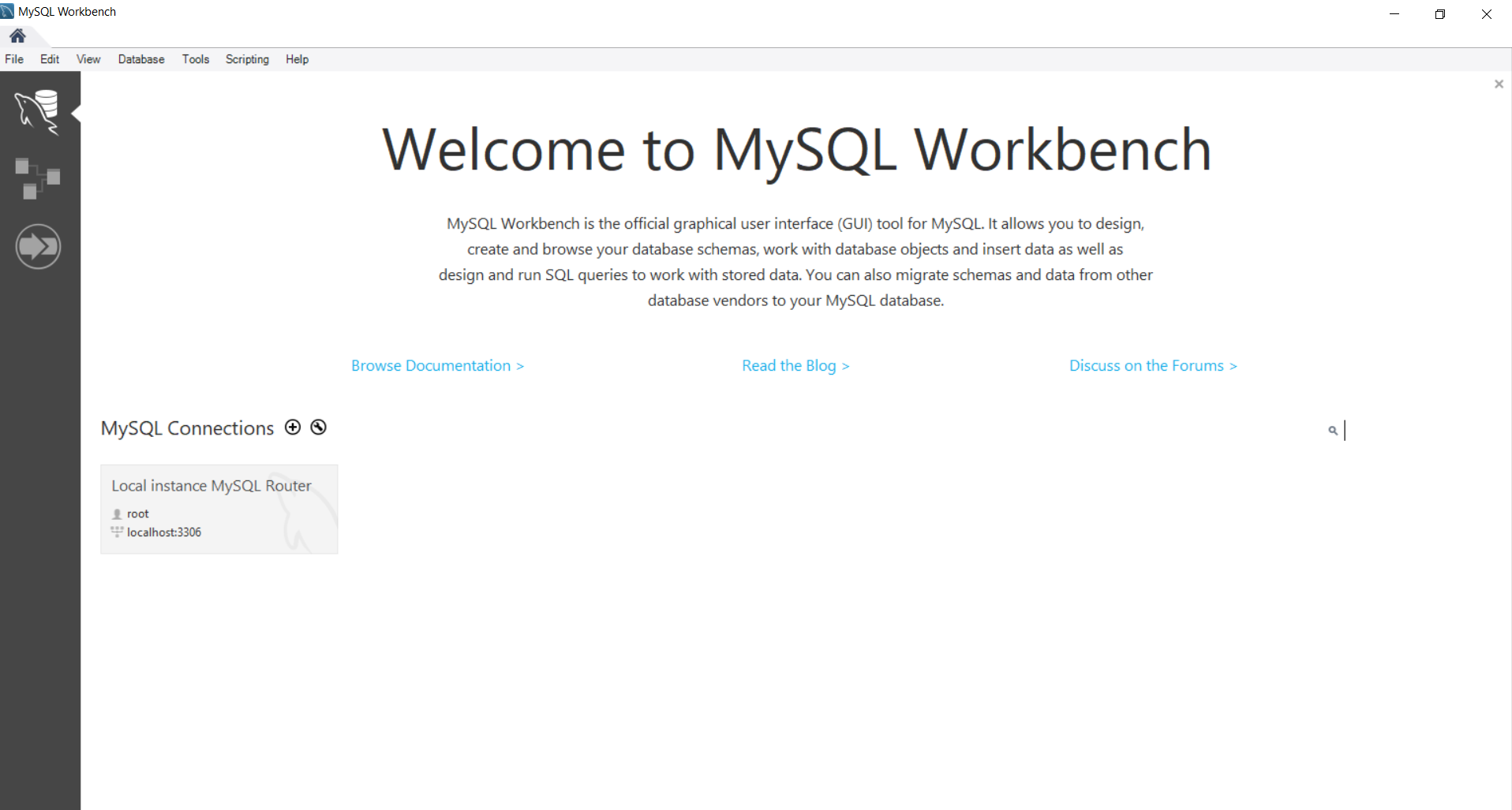
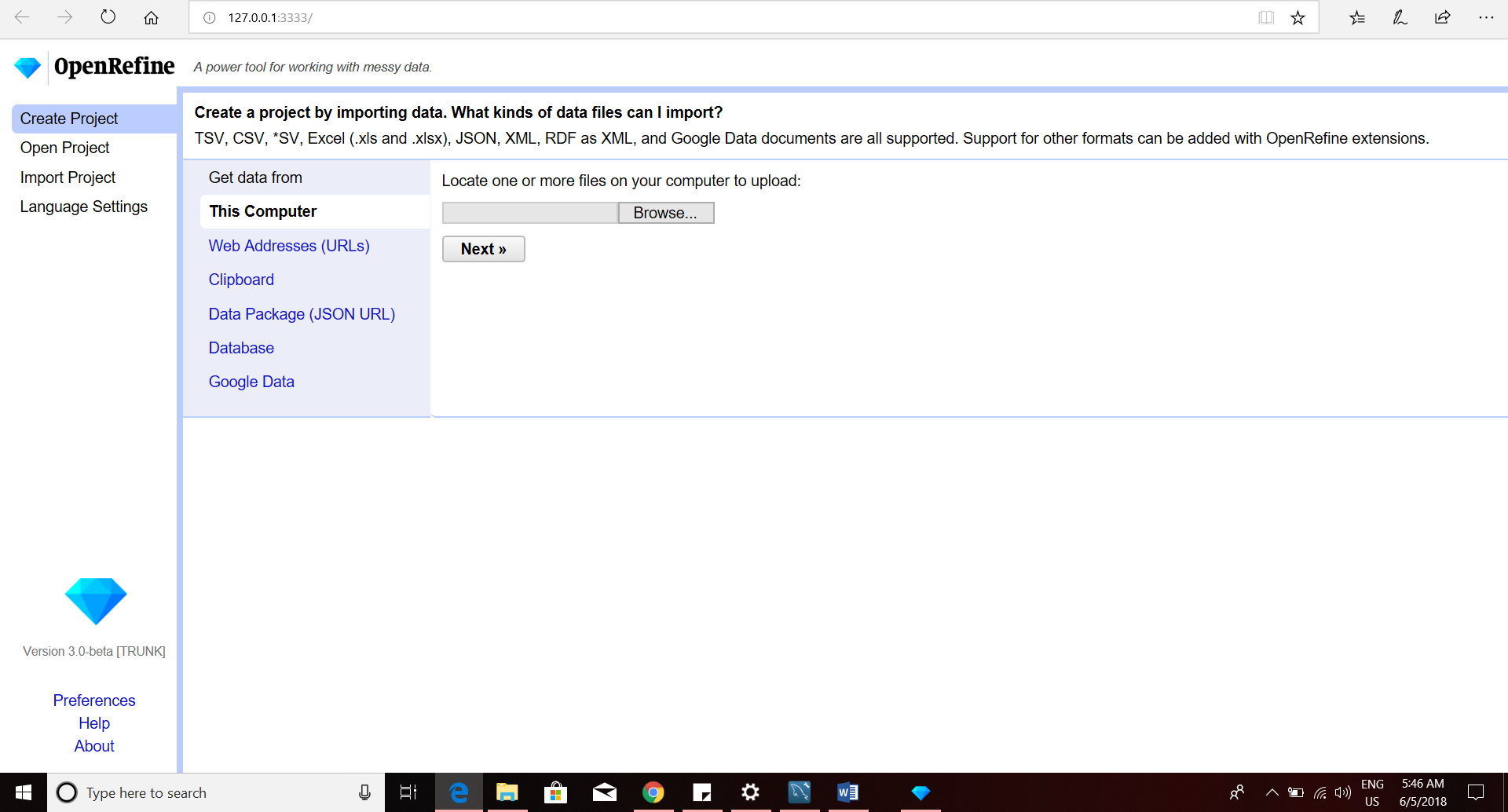
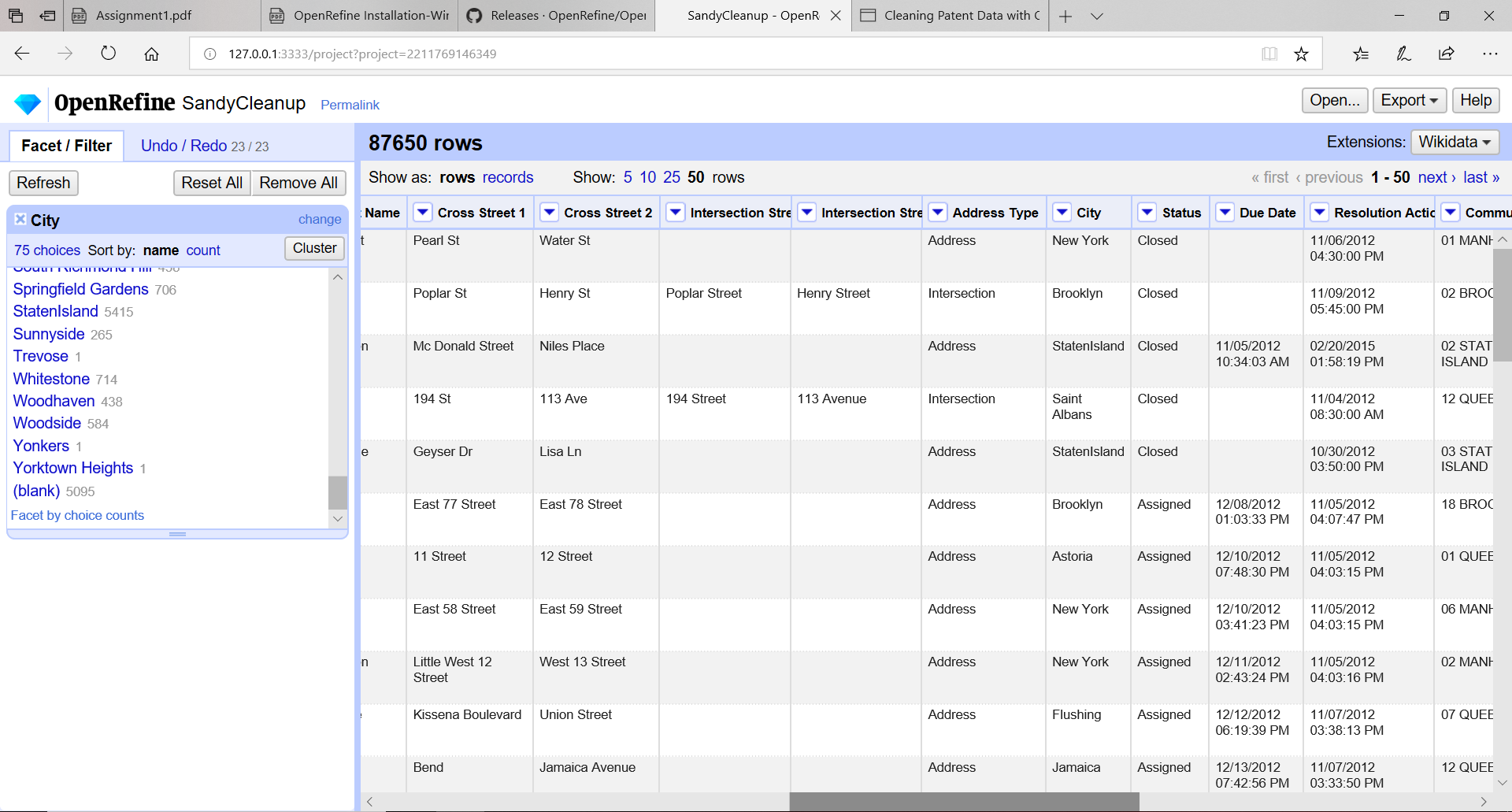
**Installation: My SQL workbench**

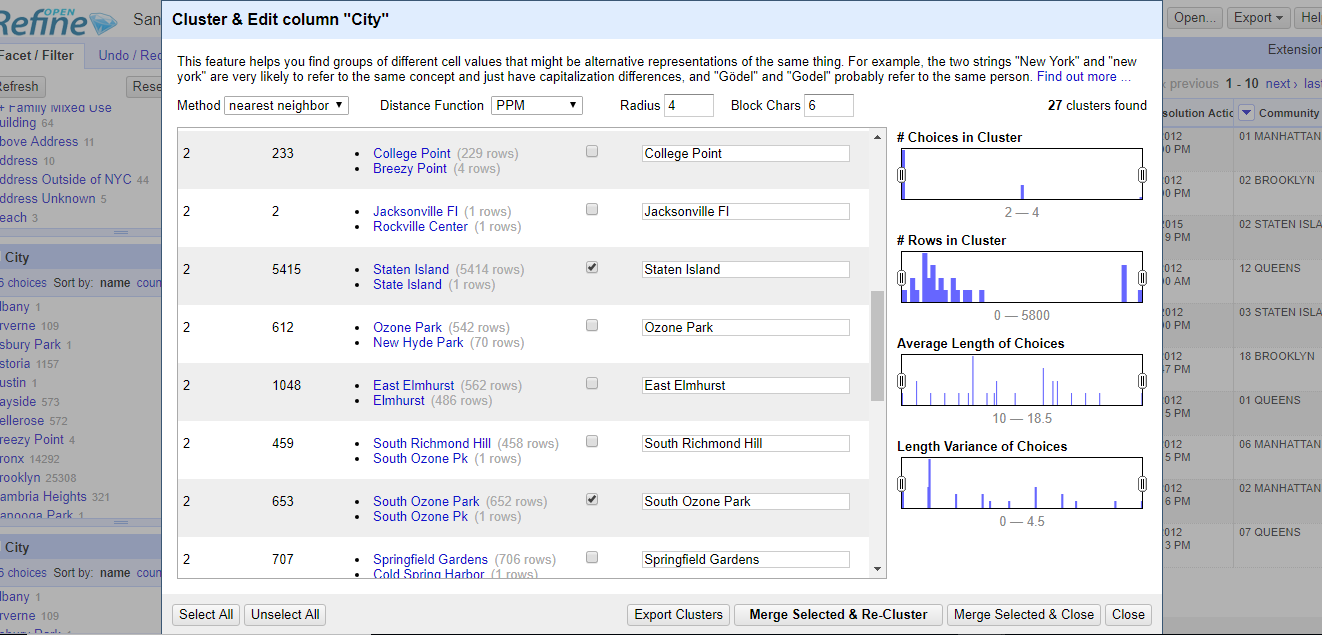


**Installation: Open Refine version 3.0 Beta**



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

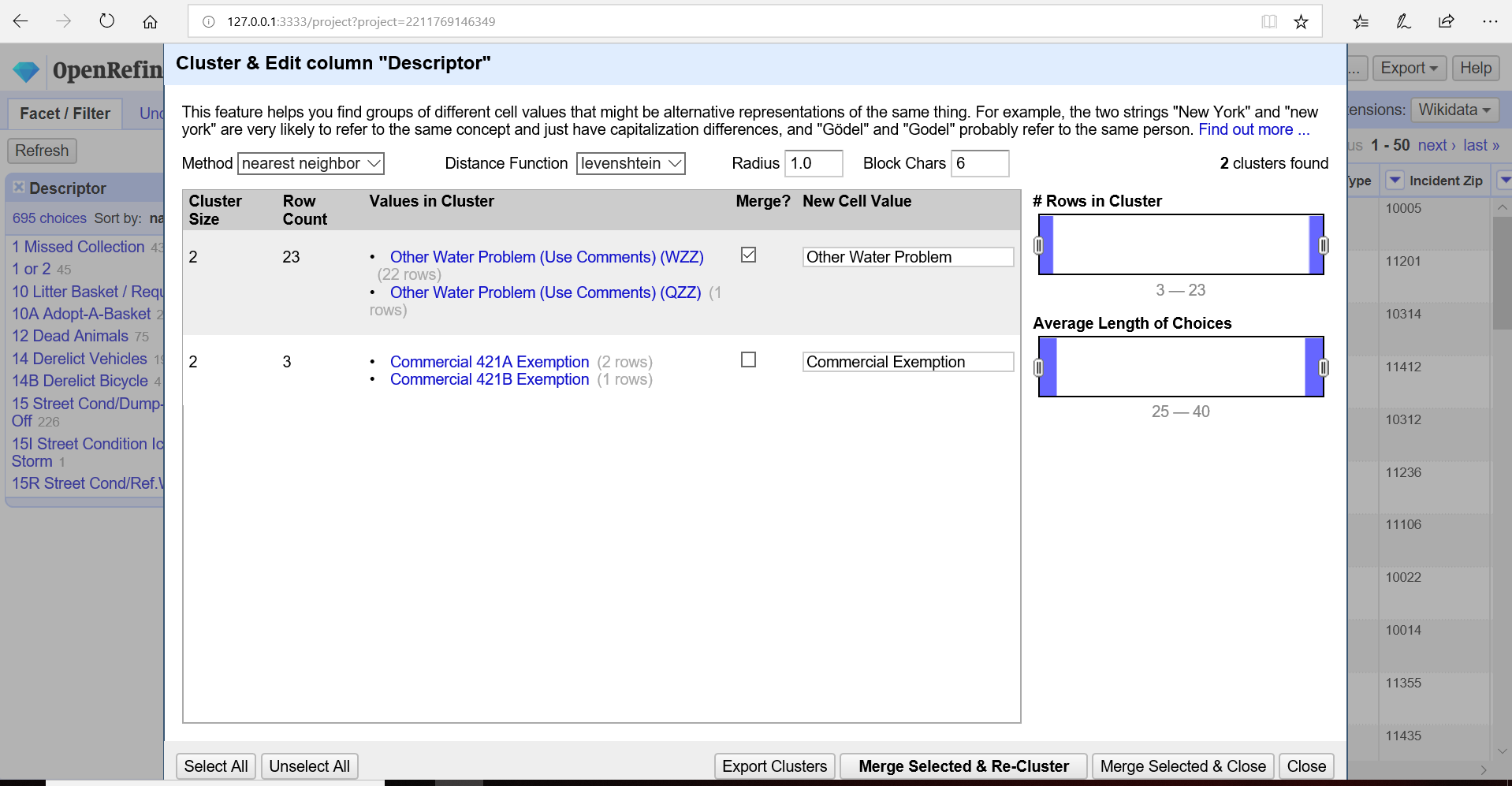




**e. Clean up Descriptor Column - Cluster and Merge following text categories:**

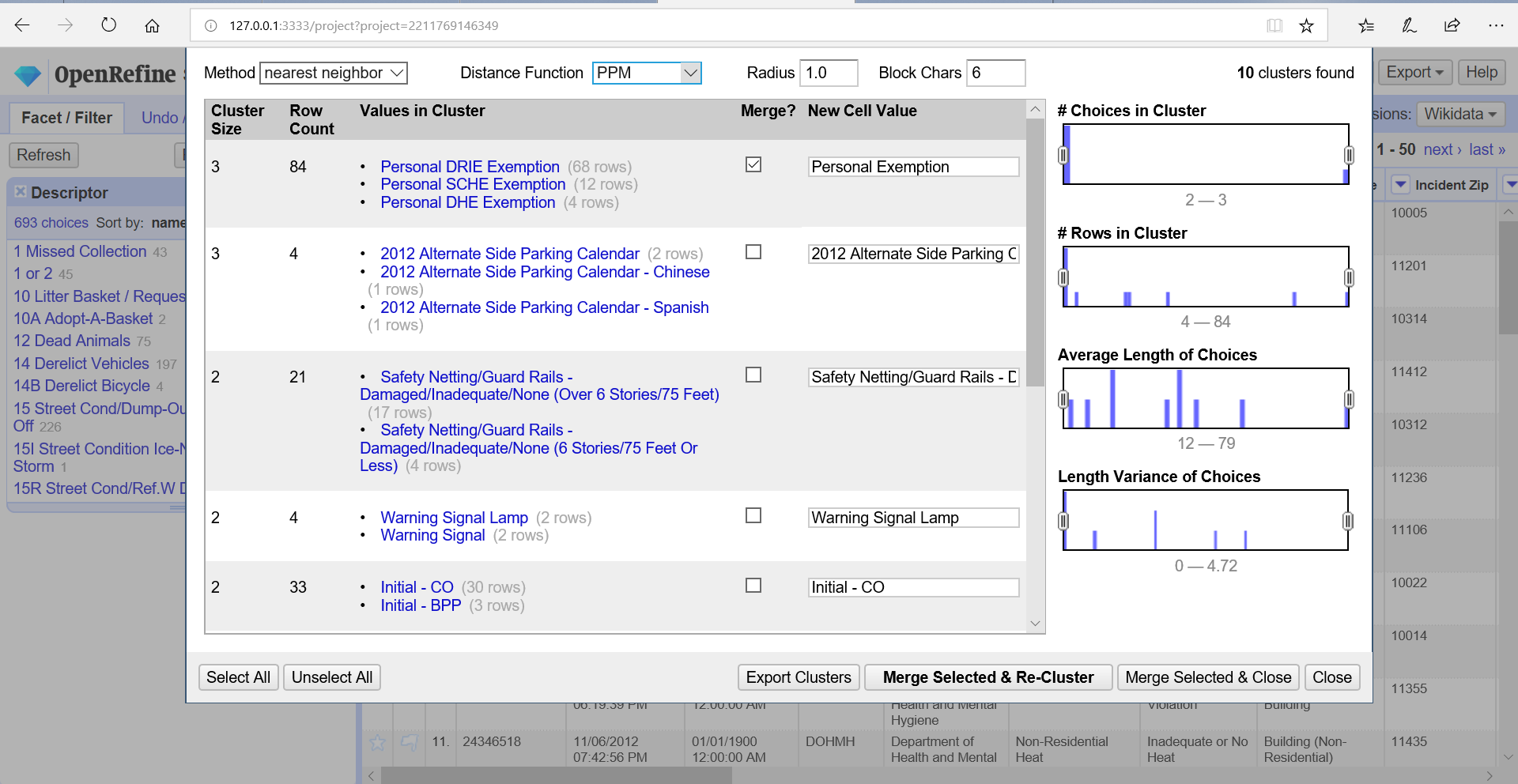
**1. "Other Water problem(WZZ)", "Other Water problem(QZZ)" as "Other Water Problem"**

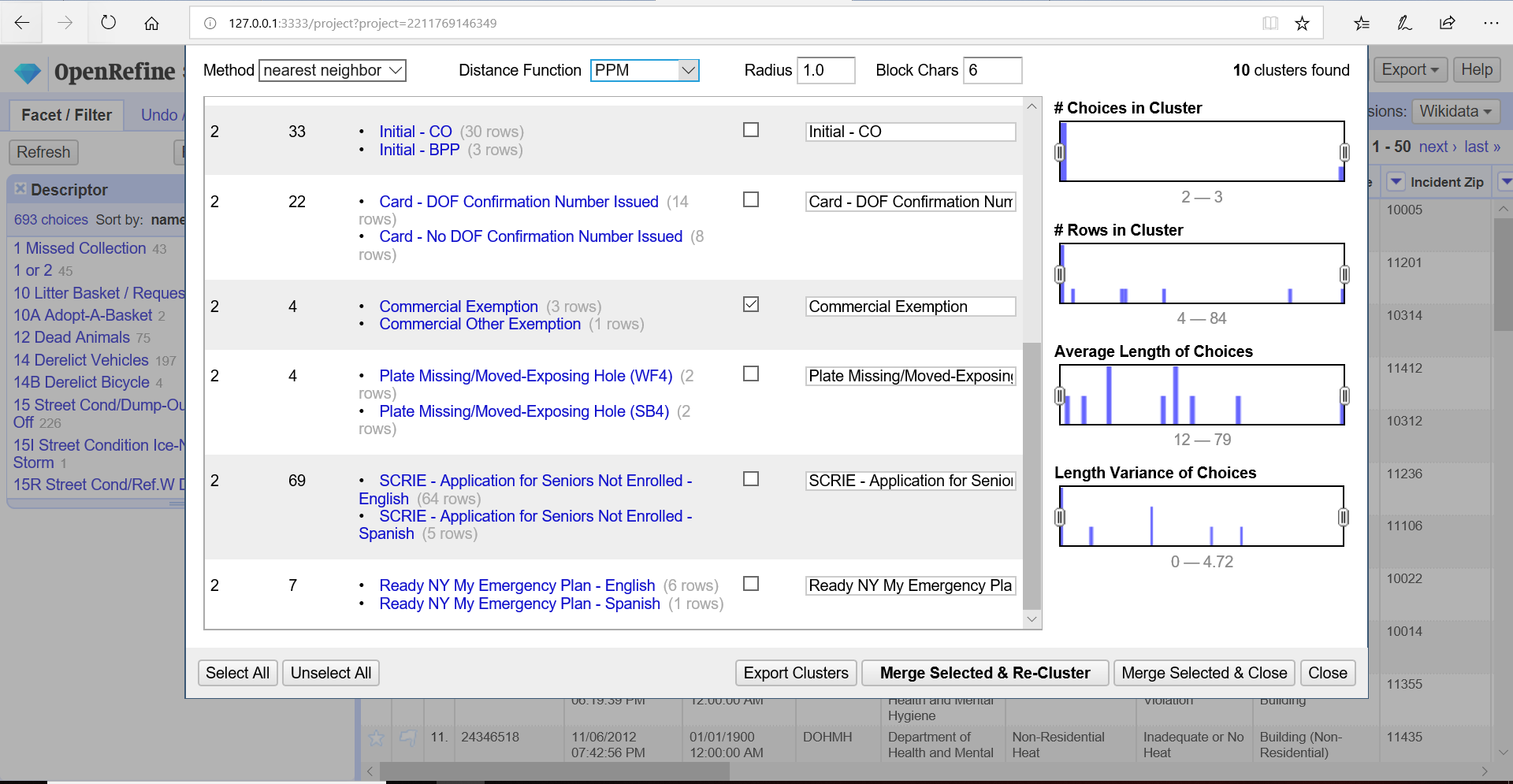
**2. "Commercial 421 A/B Exemptions" as "Commercial Exemption"**



**e.3. "Commercial Exemption" "Commercial Other Exemption" as "Commercial Exemption"**

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





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

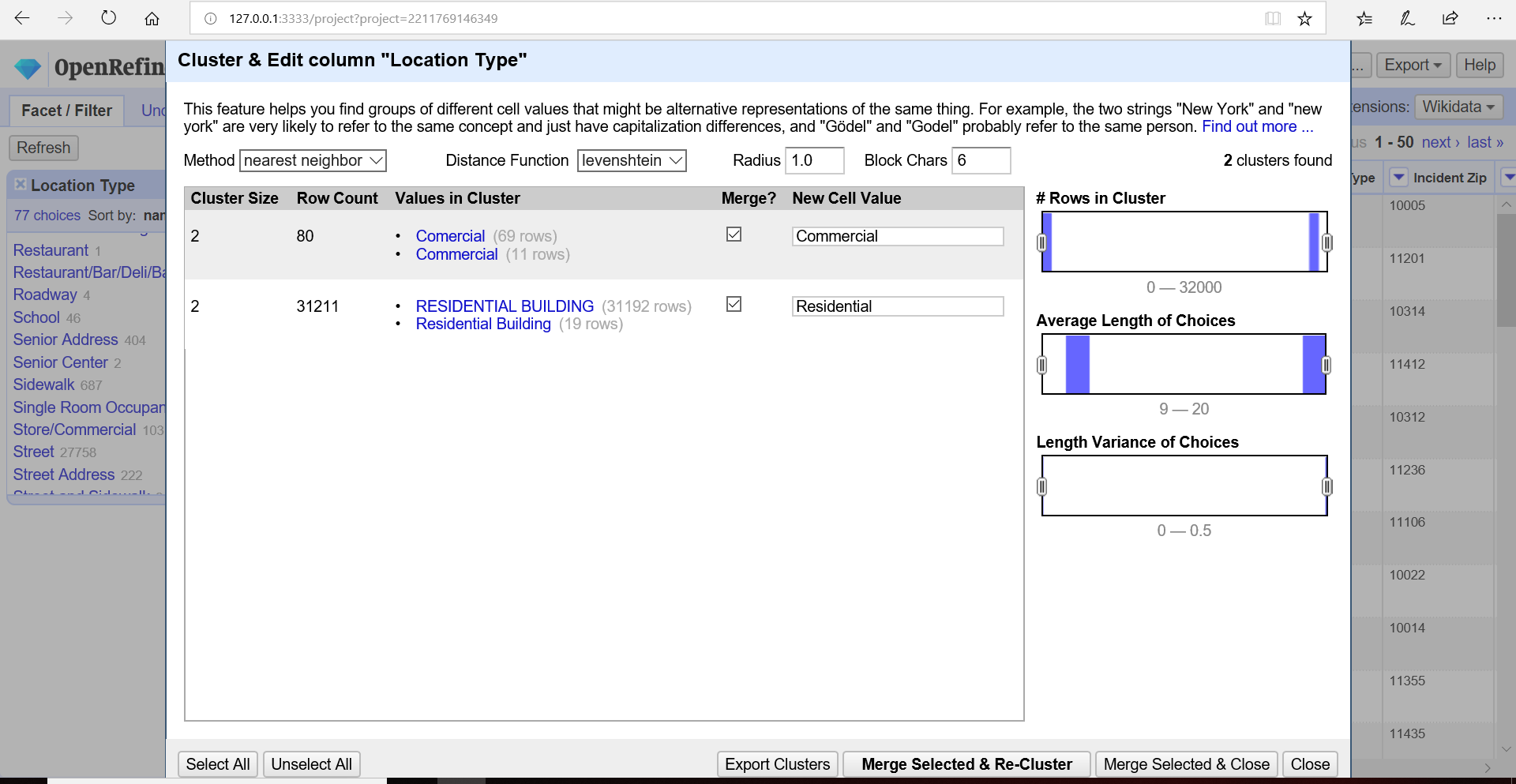
**1. "Comercial", "Commercial", "Store/Commercial" as "Commercial"**

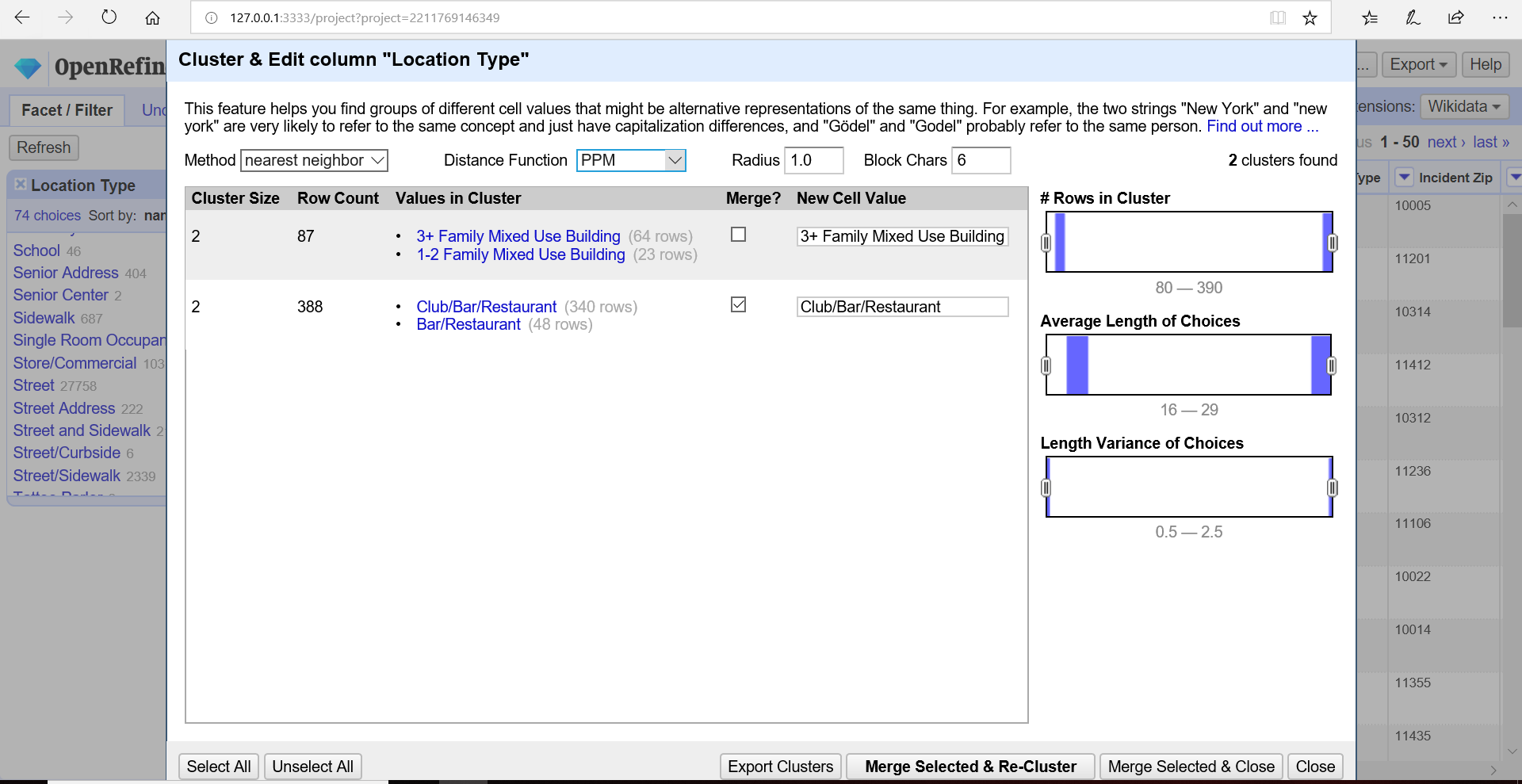
**2. "RESIDENTIAL BUILDING", "Residential Building", "Residence" as "Residential"**

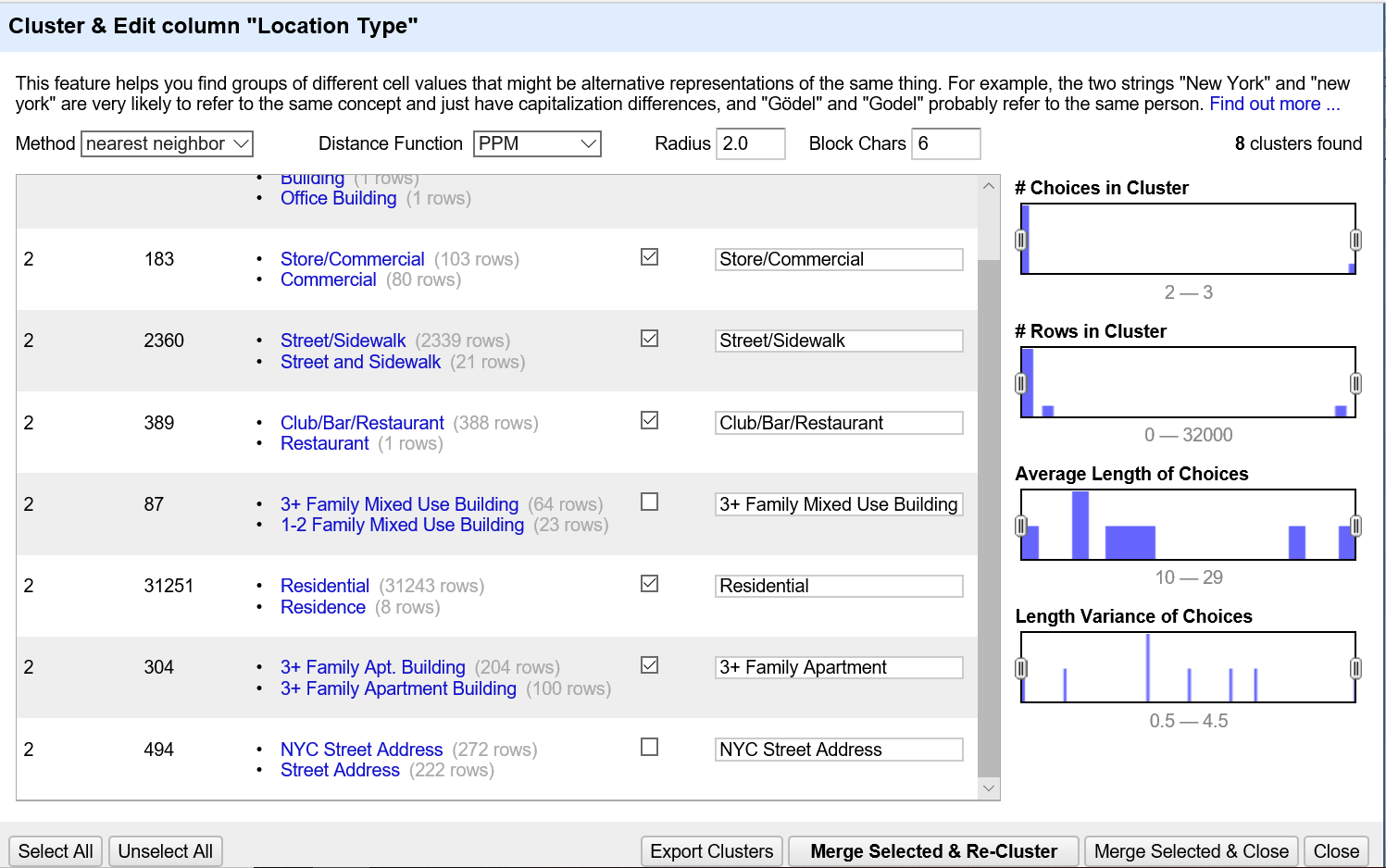
**3. "Club/Bar/Restaurant", "Bar/Restaurant", "Restaurant" as "Club/Bar/Restaurant"**

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

**5. "Street/Sidewalk", "Street and Sidewalk" as "Street/Sidewalk “by changing radius=2 in nearest neighbor function:**







* 1. **g. Online web services such as the following can be used to fetch the address given a geocode:**

**Google Reverse Geocoding for a Latitude/Longitude**

**Web Service API Example:**

**https://maps.googleapis.com/maps/api/geocode/json?latlng=40.714224,-73.961452**

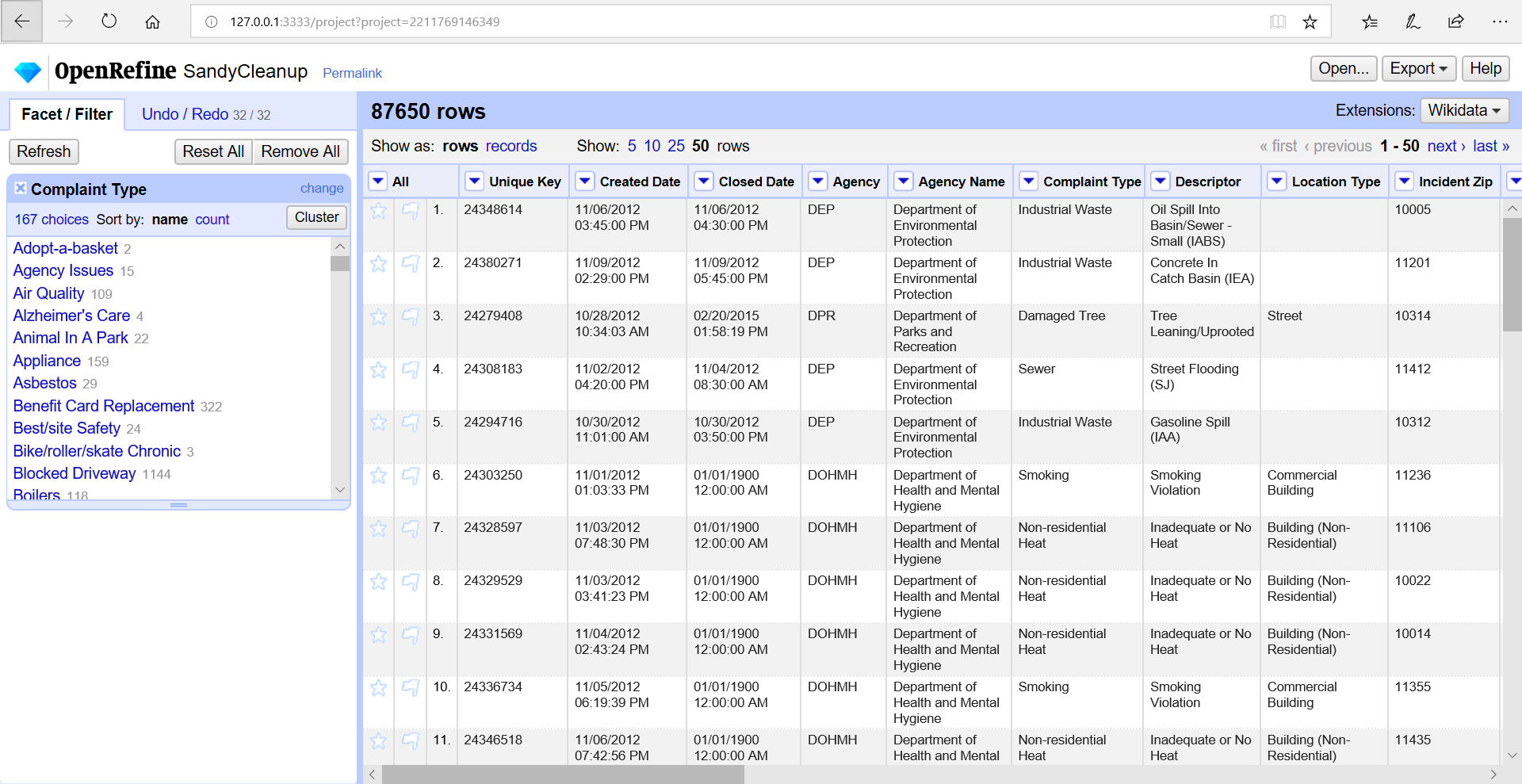
**Formulate the URL expression in OpenRefine that would fetch the complete JSON results from this web service API (You do not need to invoke the API or download the results of the web service call):**

**Answer: After removing all brackets and injecting comma between longitude and Longitude value.**

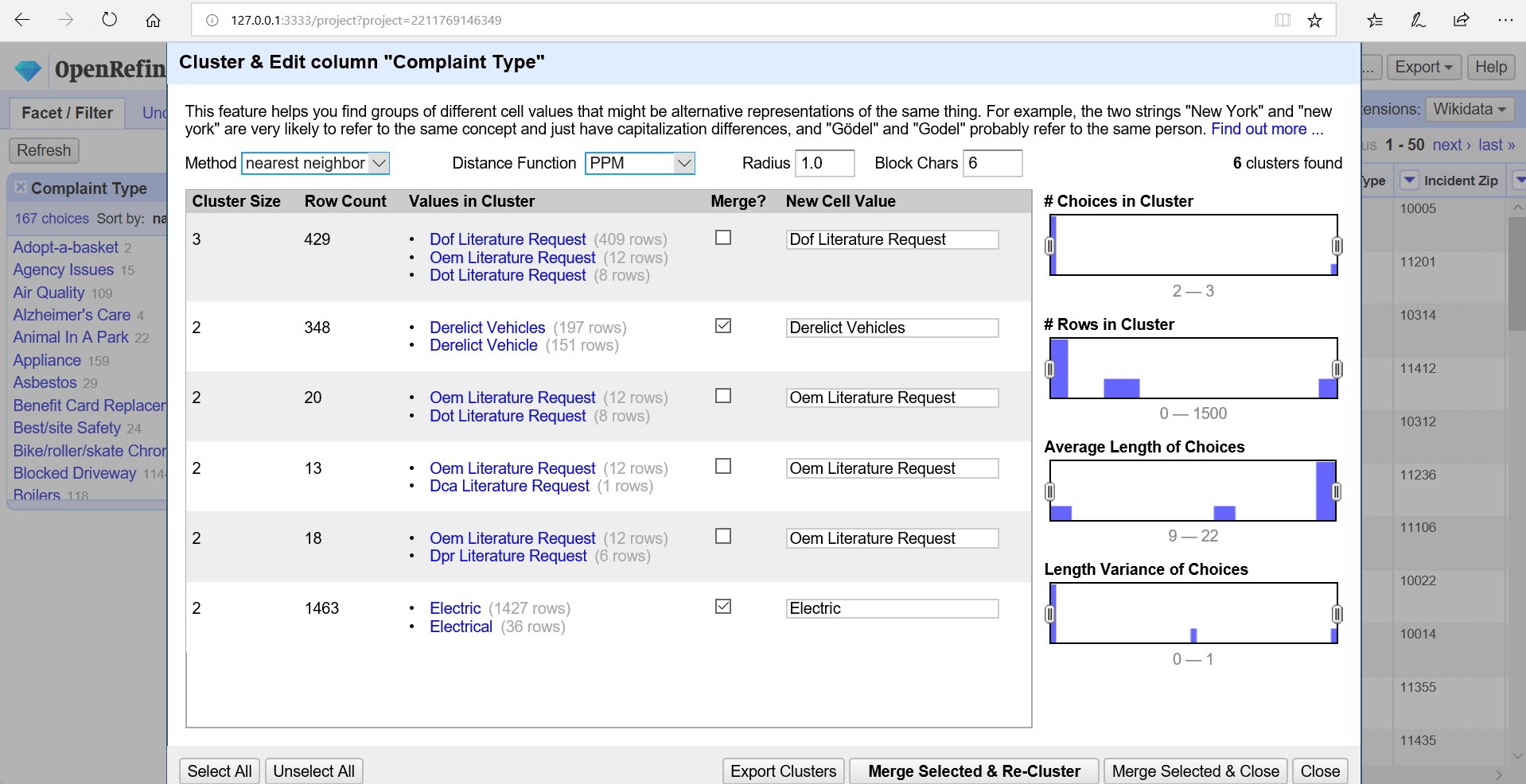
'https://maps.googleapis.com/maps/api/geocode/json?latlng='+ value.replace("(","").replace(" ","").replace(")","")

**h. Look for at least two other clean up opportunities and execute using Open Refine:**

**column complaint type to title case:**



**Cluster column values: “Derelict Vehicles” and “Derelict Vehicle” as Derelict Vehicles:**



**Descriptor column: Cluster**

**1. “Other (Explain Below)” and “Other - Explain Below” to Other - Explain Below**

**2. “Blocked Sidewalk” and “Sidewalk Blocked” to Blocked Sidewalk**

