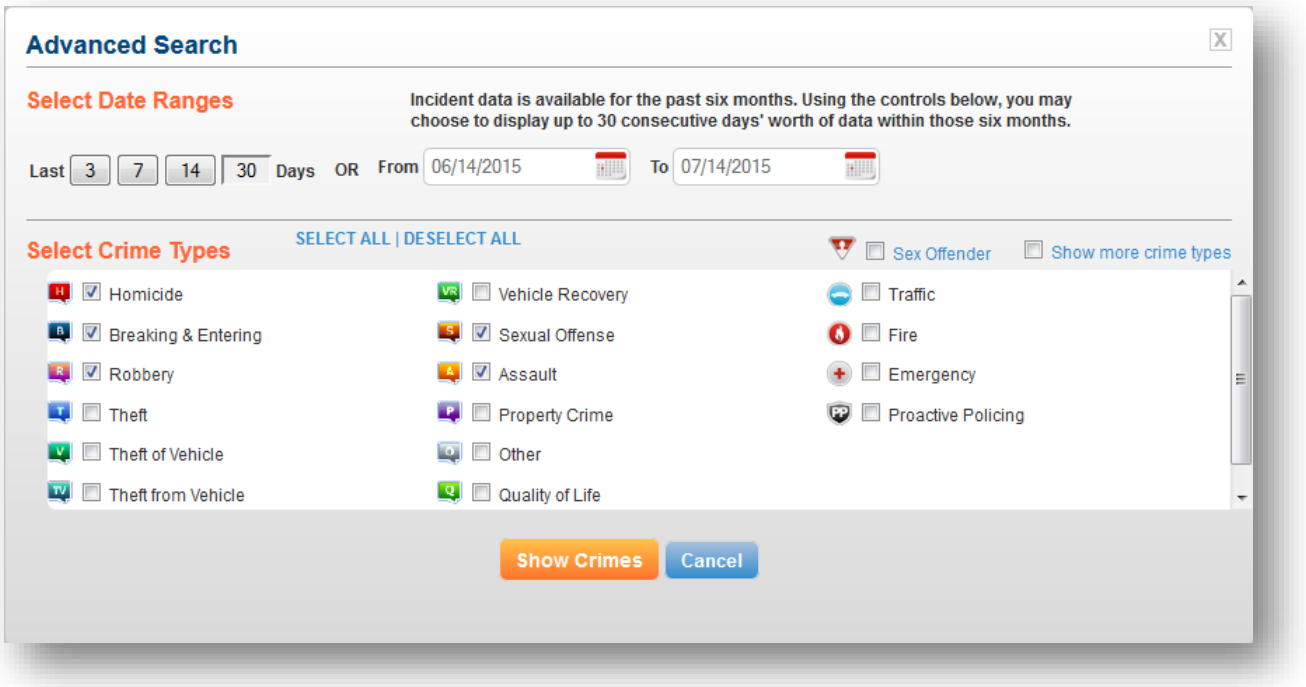
Geocoding addresses in ArcGIS

1. Downloading crime tabular data

For this exercise, we are going to download crime data for the past 15 days in San Francisco. Go to crimereports.com and in the search box type: *San Francisco, CA*.

Let's query our search to only felony crimes for the past 30 days. In the Advanced Search window, check only the following crimes: *Homicide*, *Breaking & Entering*, *Robbery*, *Sexual Offense* and *Assault* and select the las 30 days for the time frame as shown below.



Click Show Crimes to see the results on your map. You can click on any of the incidents to see the information related to each crime.

To download the data, click on the Show Details button on the top left corner, and in the new opening window on your left, click the View/Print List command at the top of the table. Select all the records in the table by clicking Control + A, then click Control + C to copy them in memory.

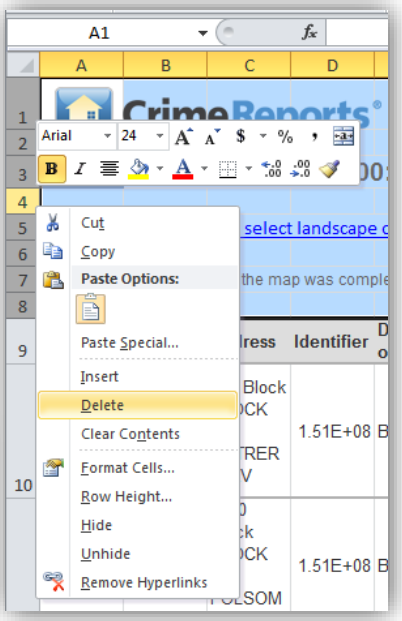
Open a new Excel spreadsheet and click Control + V to paste your table into Excel.

Notice that everything on the pop-up window got copied.

To delete the first 8 rows, select those on the left hand side, right-click and click Delete.

Select the CrimeReports logo and the print button overlaying your first row. Hit the Delete key to remove them.

Your table should look similar to the one below:



	A	B	C	D	E	F
1	Crime Type	Date/Time	Address	Identifier	Description	Agency
2	Assault	6/14/2015 0:05	500 Block BLOCK OF POTRERO AV	150524607	BATTERY	San Francisco Police Department
3	Assault	6/14/2015 1:20	1000 Block BLOCK OF FOLSOM ST	150528295	BATTERY	San Francisco Police Department

2. Preparing Your Data

Before you can import tabular data into ArcGIS, you must make sure it is in a file type and a format that ArcGIS can recognize and your field names are clean and matching geocoding categories. First we will format our data fields, and then we will save it correctly.

2.1 Formatting Tabular Files

Addresses must be correctly formatted in order to display correctly in ArcMap and fields cannot start with a number or contain special characters except for underscore (_).

- Divide your data into the following columns: Address, City, State, Zip. Enter the corresponding data into each column.
- If you are missing data from one of these categories, you do not need to include that header (zipcode in this case)
- You will need to add City and State since the same address can be repeated in several cities and states. Add two columns: City, State and fill them with the appropriate city name (San Francisco) and state name (California) for each record.*
- Read carefully the rules for field names:

  1. When creating spreadsheets, make sure **fields are fewer than 255 characters**.
  ArcGIS reads the first 255 field characters. Fields with more than 255 characters are converted to BLOB fields and are not readable. Abbreviate, manually truncate, or split any fields longer than 255

characters.

2. Check the numeric field type **before and after importing Excel data.**
ArcGIS typically converts spreadsheet numeric fields to double precision (Double), which may not meet your needs. If necessary, create new fields of the desired type and calculate values into them.

3. Check the **format for date** fields.
ArcGIS uses the Lotus date/time format. In this format, the calendar date is represented by a whole number value that represents the number of days since January 1, 1900, plus one day (due to a bug in Lotus 123 and carried over to Excel). Time is represented as the decimal portion of a 24-hour day. If date/time data is important, format the input spreadsheet using a standard Excel date/time format.

4. Follow **ArcGIS field naming rules** when creating Excel column names.
The first row of an Excel worksheet sets the name for each column. Column names become field names when an Excel worksheet is imported into ArcGIS. Always follow these naming rules:

- Column/Field names **must begin with a letter**.
- Column/Field names **must contain only letters, numbers, and the underscore character**.
- Column/Field names must be no more than 64 characters. If a name is longer than 64 characters, ArcGIS retains the first 63 characters.
- Column/Field names may not consist solely of reserved words (**date, value, name, text,** and **year**). Do not use these words in field names. See the list of reserved words. ArcGIS typically adds a trailing underscore to reserved word field names added by copying and pasting from other sources.

o Make sure the field names you have chosen are not too long and do not have spaces or other problematic characters (eg: *, &, !, #, etc). Change Crime Type to Crime_Type and Date/Time to Date_Time.

o Enter your data carefully – mistakes and typos can lead to errors in finding the address!

	A	B	C	D	E	F	G	H
1	Crime_Type	Date_Time	Address	Identifier	Description	Agency	City	State
2	Assault	6/14/2015 0:05	500 Block BLOCK OF POTRERO AV	150524607	BATTERY	San Francisco Police Department	San Francisco	California
3	Assault	6/14/2015 1:20	1000 Block BLOCK OF FOLSOM ST	150528295	BATTERY	San Francisco Police Department	San Francisco	California

** Tip: To copy an entire column with the same value, use control + C, <-, Shift + control + down, ->, Shift + control + up, control + V.*

2.2 Useable File Extensions

Once you have formatted your data, you will save it using a file type that ArcMap can recognize. The following file types can be used in ArcMap. All of these file types can be read by Microsoft Excel:

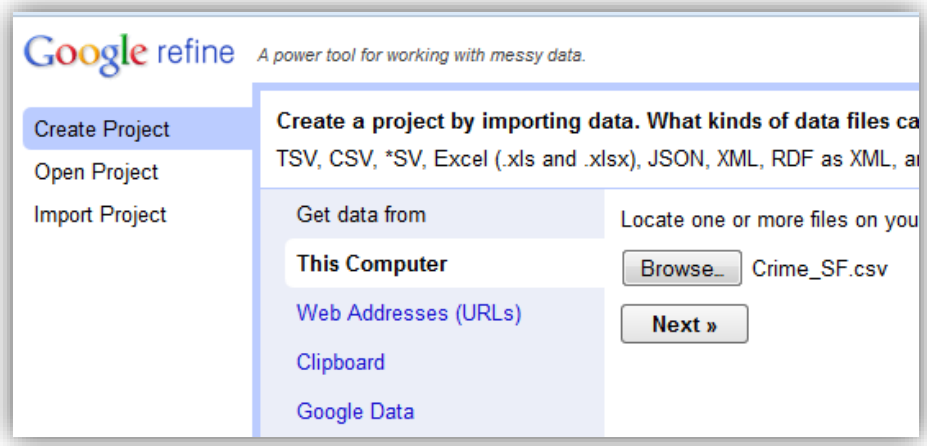
- .csv
- .txt
- .xls

Click on File, Save As, Name: *Crime_SF*. Change Save as type and select CSV (Comma delimited). Click Save. In the following windows, click OK and Yes to finish saving your file.

2.3 Cleaning records with Open Refine

Take a look at the address records in your table. Notice how most of them have the word Block or BLOCK. This will not allow ArcGIS to geocode your crime incidents.
One very powerful and free tool to clean your data is Open Refine. This easy-to-use program allows you to explore, filter, and clean your messy data with very little effort.

In your computer, go to Programs and click on google-refine to start the program.

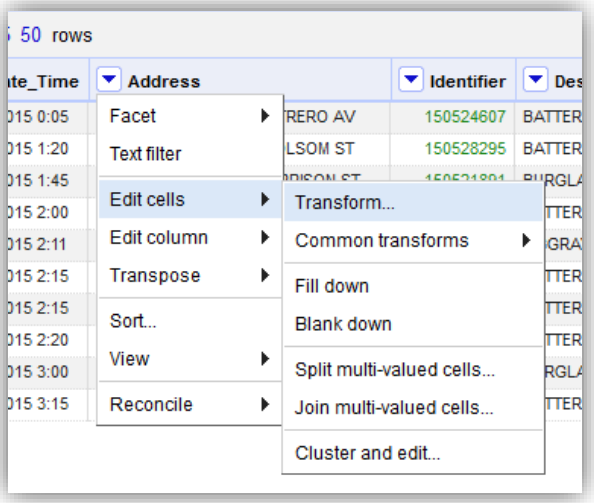


It will open a browser window, however, your data will be not be shared online so you can work with private data without any issues.

Select Get data from This Computer and click Browse. Navigate to the folder where you saved your *Crime_SF* file and click Open. Click Next.

Take a look at the columns and records and make sure everything looks correctly. On the top right corner, click Create Project. Now you can start exploring and cleaning your data.

On the *Description* field, click on the inversed arrow and select Facet – Text Facet. This allows you to see all the different categories under Description and how many records you have under each category. Close the Facet window.

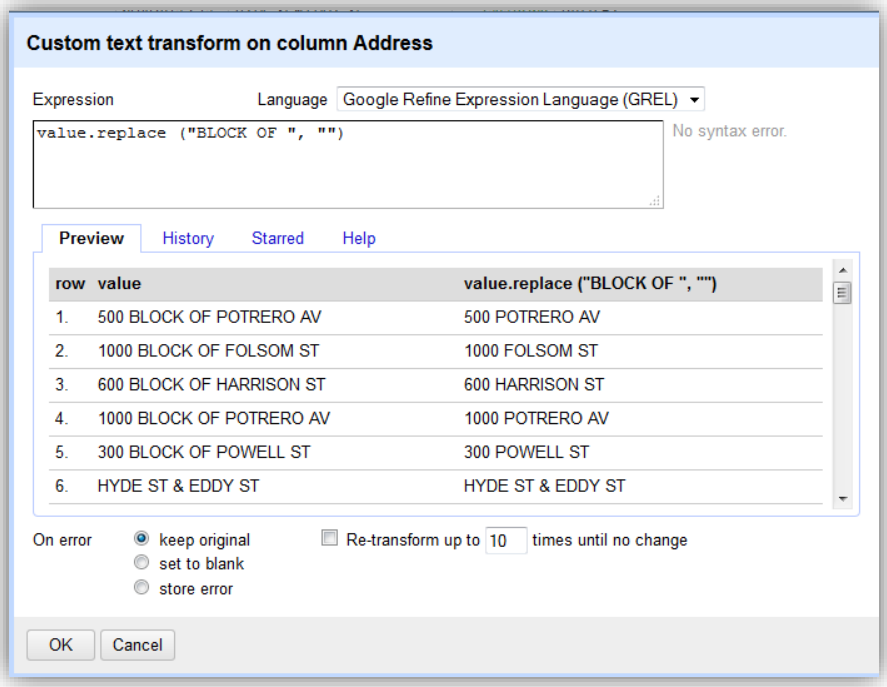


To clean the *Address* column, click on the inversed arrow and select Edit cells – Transform. This will open a new window where you can see how your changes affect the field. In the *Expression* box, type:

value.replace (“Block ”, “”)

This will replace the word Block and its trailing space with an empty space. Take a look at the transformation on the new field and click OK.

Repeat the same steps to remove the text “BLOCK OF ”, as shown below. Click OK to save your edits. Confirm your Address column is clean and ready to be geocoded in ArcMap.



To save your edits in a new file, click the Export button on the top right corner, select comma-separated value and select the Open with Microsoft Excel option. Save your new file as a CSV file and name it *Crime_SF_clean*.

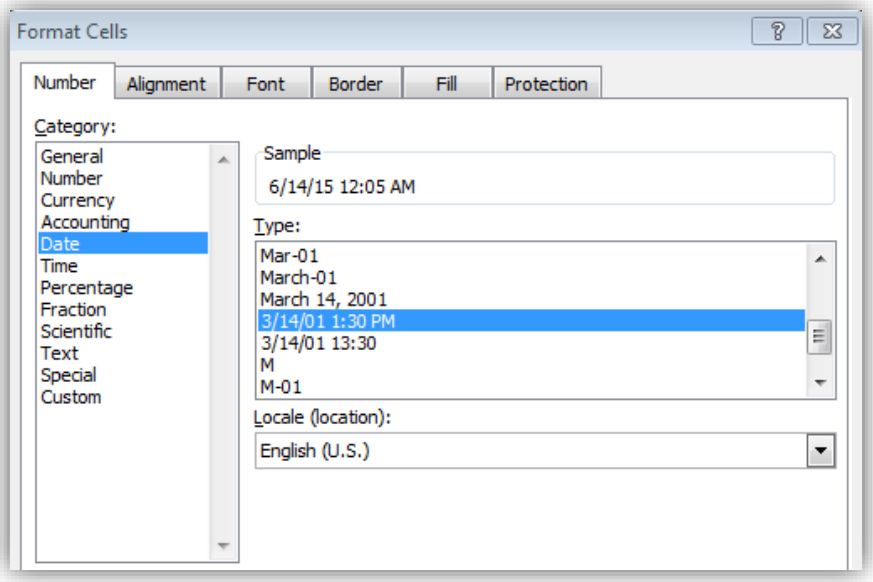
3. Importing Your Data in ArcGIS

Before you attempt to geocode or geolocate your tabular data, make sure that ArcGIS can read all of your columns without any errors.

Open ArcCatalog and navigate to your Crime_SF_Clean.csv file. Click on the Preview tab and notice that the Date_time is not displaying. If you read the rules for date formats in section 2.1, ArcGIS needs a standard date/time format in Excel in order to read the data.

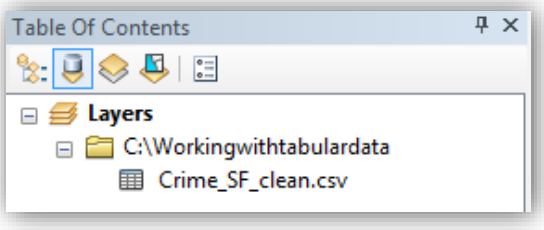
Close ArcCatalog. Open the Crime_SF_Clean.csv file in Excel and select the entire column except for the column name (use Control + Shift + Down to do this). Right-click and select Format Cells.

In the Number tab, select the Date category and as type scroll down to choose the 3/14/01 1:30PM format as display here.



Click OK and Save your changes. Check that all the fields now are readable in ArcCatalog. Close ArcCatalog and open ArcMap.

In ArcMap, open a new map document. In the ArcCatalog side window navigate to your file and drag it to the Table of Contents. Notice that the view changes from List by Drawing Order to List by Source (Note: this is the only way in ArcMap that you are able to view tables).



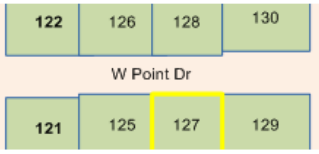
Open the attribute table and confirm once again that all

fields display correctly. Close the attribute table view.

4. Geocoding addresses in ArcGIS

Geocoding is the process of transforming a description of a location, such an address or a name of a place, to a location on the earth's surface. The resulting locations are output as geographic features with attributes, which can be used for mapping or spatial analysis.

In order to geocode addresses, we need an address reference dataset and an address locator. The reference dataset contains a database with the location of addresses for a particular region or locality. The address locator is the entity that specifies the method to interpret a particular type of address input, relate it with the reference dataset and deliver a matching option back to the user interface. Here is an example of how the process works:

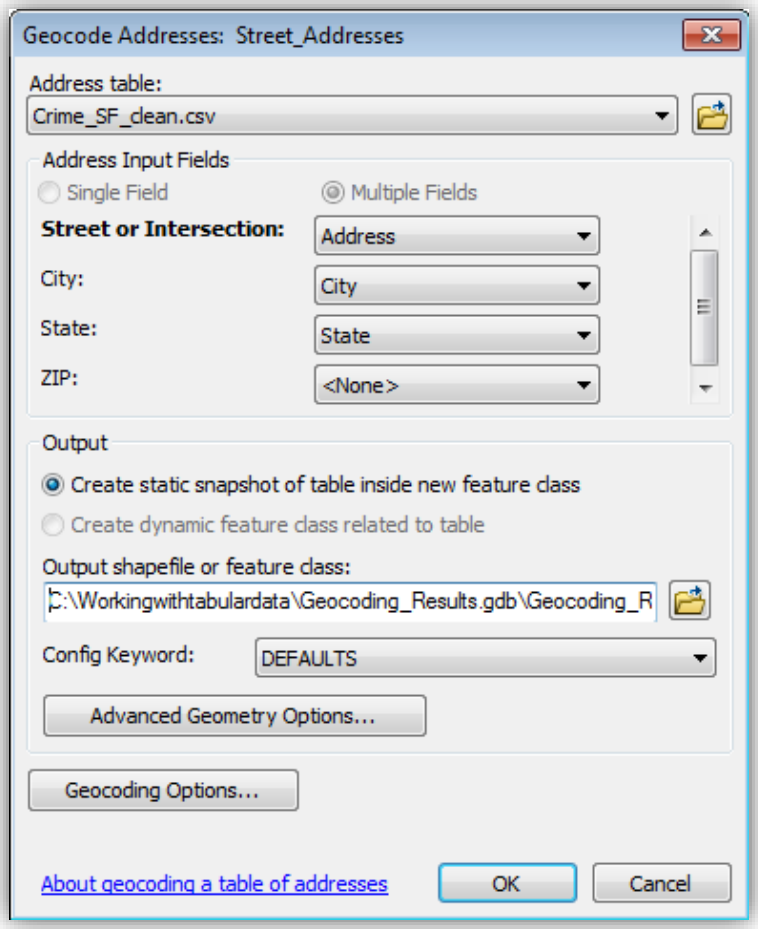
<table>
<tr><th>Step</th><th>Conceptual example</th></tr>
<tr><td>1. Original address entered</td><td>127 West Point Drive, Olympia, WA 98501</td></tr>
<tr><td>2. Address Parsed</td><td>127 | West | Point | Drive | Olympia | WA | 98501
127 | West Point | Drive | Olympia | WA | 98501</td></tr>
<tr><td>3. Multiple representations of the address created</td><td>
<table>
<tr><td>House number</td><td>127</td><td></td><td></td></tr>
<tr><td>Direction</td><td>west</td><td>w</td><td></td></tr>
<tr><td>Street Name</td><td>point</td><td>west point</td><td></td></tr>
<tr><td>Street type</td><td>drive</td><td>dr</td><td>drv</td></tr>
<tr><td>City</td><td>Olympia</td><td></td><td></td></tr>
<tr><td>State</td><td>wa</td><td>washington</td><td>wash</td></tr>
<tr><td>ZIP</td><td>98501</td><td></td><td></td></tr>
</table>
</td></tr>
<tr><td>4. Address locator searched</td><td> 
</td></tr>
<tr><td>5. Score of each potential match established</td><td>
<table>
<tr><th>Street</th><th>Number</th><th>Direction</th><th>Match Score</th></tr>
<tr><td>Point</td><td>127</td><td>W</td><td>100</td></tr>
<tr><td>Point</td><td>129</td><td>W</td><td>85</td></tr>
<tr><td>West Point</td><td>121</td><td></td><td>80</td></tr>
<tr><td>West Point</td><td>137</td><td></td><td>80</td></tr>
<tr><td>Pointe</td><td>138</td><td>W</td><td>70</td></tr>
</table>
</td></tr>
<tr><td>6. List of candidates filtered</td><td>
<table>
<tr><th>Street</th><th>Number</th><th>Direction</th><th>Match Score</th></tr>
<tr><td>Point</td><td>127</td><td>W</td><td>100</td></tr>
<tr><td>Point</td><td>124</td><td>W</td><td>85</td></tr>
</table>
</td></tr>
<tr><td>7. Best candidate matched</td><td>127 W Point Dr, Olympia, WA 98501</td></tr>
<tr><td>8. Matched feature indicated</td><td> 
</td></tr>
</table>

We will use an address locator already created for us that contains addresses for the entire US. In the ArcCatalog window, create a connection with GIS (\\libstorage.clemson.edu). Navigate to Geocoding_data_2014 – Geocoding data. In this folder, you will see different types of Address locators that will match your data depending on your attributes (zipcodes, CityState, etc)

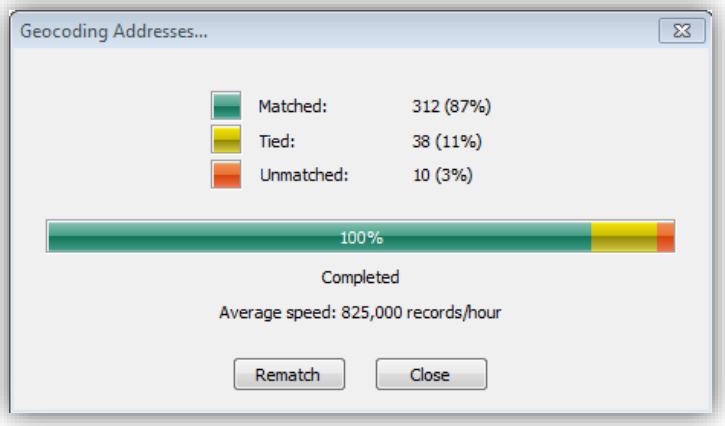
In ArcMap, go to Customize, Toolbars, and check the Geocoding toolbar. Drag the Address_Points locator to your Table Of Contents. Notice how this is the default geocoder in your Geocoding toolbar.

Click the Geocode Addresses button Select the Street_Addresses as your Address Locator. In the following window, select your Crime_SF_clean table and make sure the fields in your table match the input fields required for the locator to work. Save your results in your working folder. Make sure to create a new file geodatabase (Geocoding_Results) and save your new file as: Geocoding_SF_Crime_Street_Addresses.

Click the Geocoding Options button and change the Spelling sensitivity to 20 and the Minimum candidate and match scores to 20 as well.



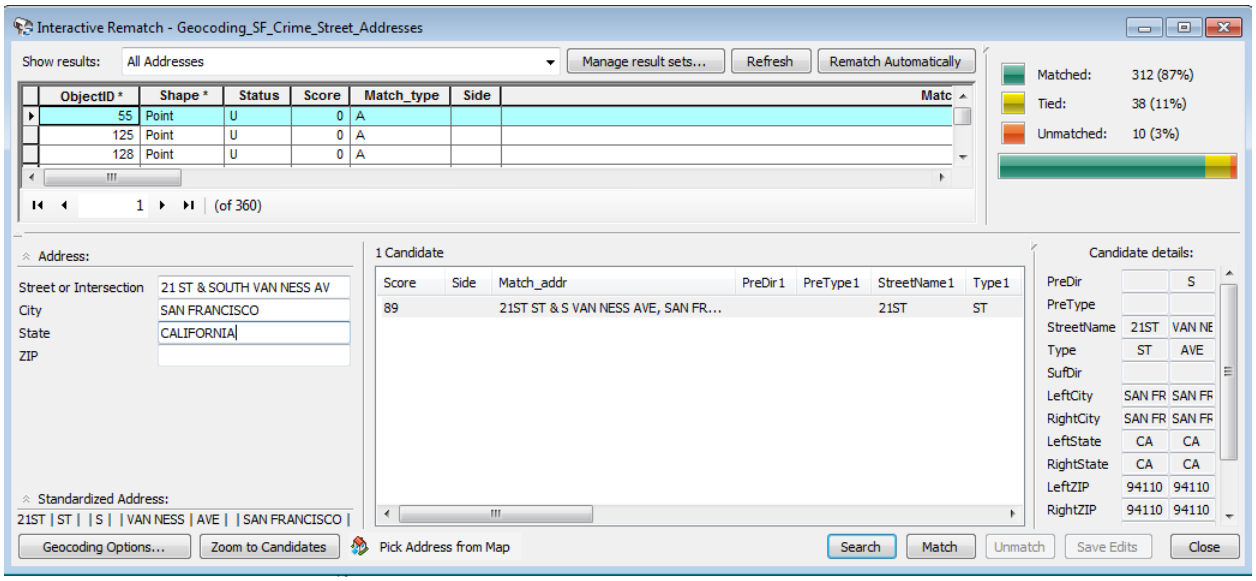
Click OK to start the geocoding process. Once it is done, you should see a results window that looks similar to this:



Click the Rematch button to manually inspect the addresses that didn't match. In the Status column, right-click to Sort descending to see all the records with a U (for unmatched) at the top of the column.

Select the first one: *1 st & south van ness av*. Let's inspect why this address didn't match. Open Google maps and type this address. Notice that South Van Ness Avenue never intersects with 1st street.

Assume a data error on this entry and change the Street or Intersection field to 21st & south van ness av. Click Search and you will notice a candidate with a score of 89.



Click Match. Notice that your unmatched records now are only 9 instead of 10. You can repeat the same process with the rest of the unmatched addresses. Close the Interactive Rematch window.

Open the attribute table of your new geocoded layer and review the matching fields. Congratulations, you are a **tiger of a geocoder**!