

MATCH 'N MERGE

USER UI GUIDANCE

Jordan Cox

Rob Holbrook

Ramon Lim

Stephanie Zhu

12/08/2017

Launching the Match N Merge Application

1. Install the Flexx package, <https://flexx.readthedocs.io/en/latest/>
2. Download the latest Match N Merge code by doing a GET on <https://github.com/rholbrook98/matchnmerge>
3. Locate Jupyter notebook, MatchNMerge.ipynb
4. Ensure that the following files do NOT exist in the current Jupyter notebook folder: results.csv, columnMapping.csv, columnMapping_user.csv, filesReady.txt, status.txt. If any of these exists, delete them on the filesystem
5. Open Jupyter notebook, MatchNMerge.ipynb
 - a. Note: Jupyter notebook must be run in Firefox, Mozilla
 - b. Note: Default download file path for Firefox must be the folder where the notebook is located
6. Select Kernel -> Restart & Run All
 - a. Javascript/HTML UI Application will be presented

Running the Match N Merge Application

Step 1: Select CSV Data Sets



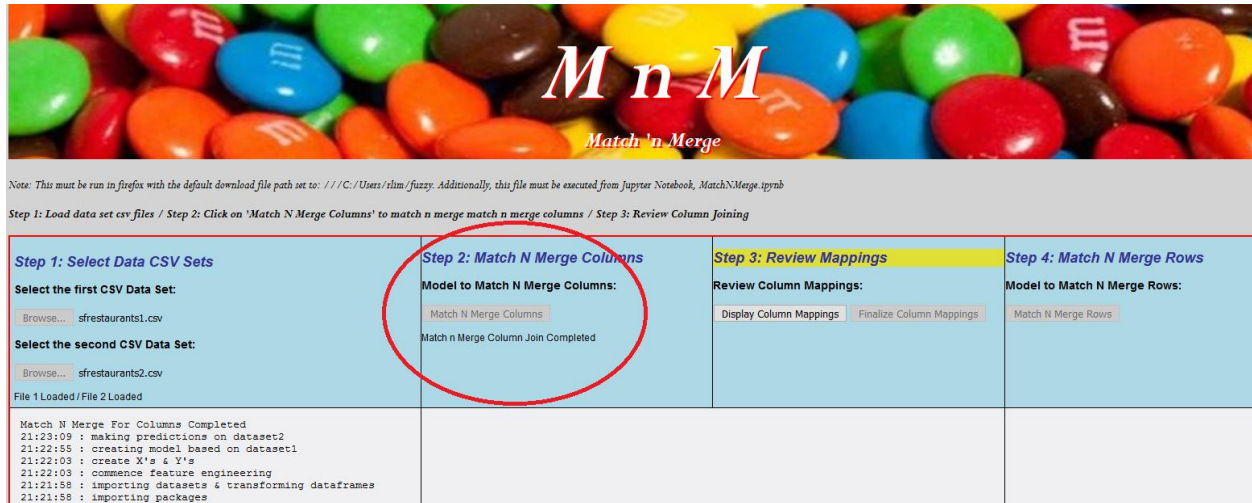
The image shows the Match 'n Merge application interface. At the top is a banner with colorful M&M's candies and the text "M n M" and "Match 'n Merge". Below the banner is a note: "Note: This must be run in firefox with the default download file path set to: //C:/Users/rlim/fuzzy. Additionally, this file must be executed from Jupyter Notebook, MatchNMerge.ipynb". Below the note is a section titled "Step 1: Load data set csv files". This section is divided into four columns: "Step 1: Select Data CSV Sets", "Step 2: Match N Merge Columns", "Step 3: Review Mappings", and "Step 4: Match N Merge Rows". The "Step 1: Select Data CSV Sets" column is circled in red. It contains two sections: "Select the first CSV Data Set:" with a "Browse..." button and "No file selected." text, and "Select the second CSV Data Set:" with a "Browse..." button and "No file selected." text. The other columns contain buttons for "Match N Merge Columns", "Display Column Mappings", "Finalize Column Mappings", and "Match N Merge Rows".

- Save the CSV file to the same folder as the Jupyter notebook (you can set the default file download path in the Firefox browser to download files to that folder)
- Select the first data set (CSV file)
- Select the second data set (CSV file)

Step 2: Match N Merge Columns

Match 'N Merge – IEOR 290 – Cox, Holbrook, Lim, Zhu

Click on the **“Match N Merge Columns”** button to start the machine learning modeling process to match columns across the two data sets



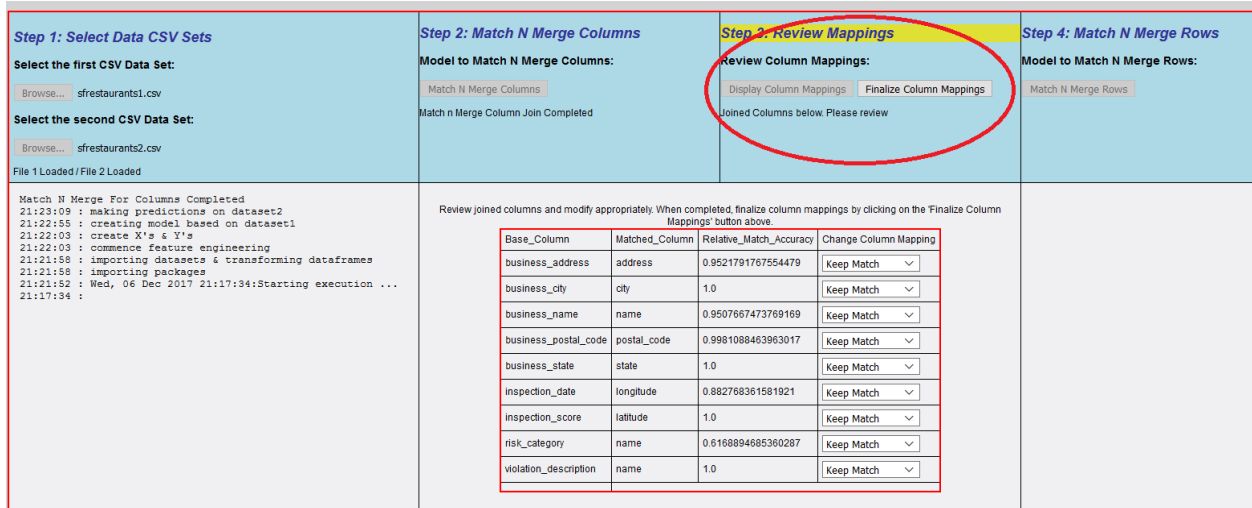
The screenshot shows the Match 'n Merge application interface. At the top is a banner with colorful M&M's candies and the text "M n M Match 'n Merge". Below the banner is a note about running the application in Firefox. The main interface is divided into four steps: Step 1: Select Data CSV Sets, Step 2: Match N Merge Columns, Step 3: Review Mappings, and Step 4: Match N Merge Rows. Step 2 is currently active and highlighted with a red circle. It contains a button labeled "Match N Merge Columns" and a status message "Match n Merge Column Join Completed". Step 1 shows two CSV files loaded: "sfrestaurants1.csv" and "sfrestaurants2.csv". Step 3 shows a button labeled "Display Column Mappings". Step 4 shows a button labeled "Match N Merge Rows". A status box at the bottom left displays the following log:

```
Match N Merge For Columns Completed
21:23:09 : making predictions on dataset2
21:22:55 : creating model based on dataset1
21:22:03 : create X's & Y's
21:22:03 : commence feature engineering
21:21:58 : importing datasets & transforming dataframes
21:21:58 : importing packages
```

- Wait for column modeling background python function to complete
- Status box on the bottom left-hand section of the UI informs you when this has completed
- Once completed, the **“Display Column Mappings”** button is enabled

Step 3: Review Mappings

Click on the **“Display Column Mappings”** button to view the best column matches based on the results of the predictive modeling.



The screenshot shows the Match 'n Merge application interface with Step 3: Review Mappings highlighted by a red circle. The interface is divided into four steps: Step 1: Select Data CSV Sets, Step 2: Match N Merge Columns, Step 3: Review Mappings, and Step 4: Match N Merge Rows. Step 3 is currently active and highlighted. It contains a button labeled "Display Column Mappings" and a status message "Joined Columns below: Please review". Step 1 shows two CSV files loaded: "sfrestaurants1.csv" and "sfrestaurants2.csv". Step 2 shows a button labeled "Match N Merge Columns" and a status message "Match n Merge Column Join Completed". Step 4 shows a button labeled "Match N Merge Rows". A status box at the bottom left displays the following log:

```
Match N Merge For Columns Completed
21:23:09 : making predictions on dataset2
21:22:55 : creating model based on dataset1
21:22:03 : create X's & Y's
21:22:03 : commence feature engineering
21:21:58 : importing datasets & transforming dataframes
21:21:58 : importing packages
21:21:52 : Wed, 06 Dec 2017 21:17:34:Starting execution ...
21:17:34 :
```

Review joined columns and modify appropriately. When completed, finalize column mappings by clicking on the 'Finalize Column Mappings' button above.

Base_Column	Matched_Column	Relative_Match_Accuracy	Change Column Mapping
business_address	address	0.9521791767554479	Keep Match
business_city	city	1.0	Keep Match
business_name	name	0.9507667473769169	Keep Match
business_postal_code	postal_code	0.9981088463963017	Keep Match
business_state	state	1.0	Keep Match
inspection_date	longitude	0.882768361581921	Keep Match
inspection_score	latitude	1.0	Keep Match
risk_category	name	0.6168894685360287	Keep Match
violation_description	name	1.0	Keep Match

Mapping table presented to the user:

Match 'N Merge – IEOR 290 – Cox, Holbrook, Lim, Zhu

Base_Column	Matched_Column	Relative_Match_Accuracy	Change Column Mapping
business_address	address	0.9521791767554479	Keep Match ▾
business_city	city	1.0	Keep Match ▾
business_name	name	0.9507667473769169	Keep Match ▾
business_postal_code	postal_code	0.9981088463963017	Keep Match ▾
business_state	state	1.0	Keep Match ▾
inspection_date	longitude	0.882768361581921	Keep Match ▾
inspection_score	latitude	1.0	Keep Match ▾
risk_category	name	0.6168894685360287	Keep Match ▾
violation_description	name	1.0	Keep Match ▾

Column Mapping Table:

Base_Column: Columns from data set 2

Match_Column: Best column match to data set 1

Relative_Match_Accuracy: Relative measure of column matching modeling performance

Change Column Mapping: Legend:

Keep Match -- keep matching of data set 2 column to data set 1 column

Do Not Match -- do not match these two columns (will be excluded from row match)

Data Set 1 Column Name - match specific data set 1 column to data set 2 column

Click on the “**Finalize Column Mappings**” button to finalize changes.

Step 4: Match N Merge Rows

Click on the “**Match N Merge Rows**” button to start the process of matching rows across the two data sets based on the column matches

Match 'N Merge – IEOB 290 – Cox, Holbrook, Lim, Zhu

Step 1: Select Data CSV Sets	Step 2: Match N Merge Columns	Step 3: Review Mappings	Step 4: Match N Merge Rows																												
Select the first CSV Data Set: <input type="button" value="Browse..."/> sfrestaurants1.csv Select the second CSV Data Set: <input type="button" value="Browse..."/> sfrestaurants2.csv File 1 Loaded / File 2 Loaded	Model to Match N Merge Columns: <input type="button" value="Match N Merge Columns"/> Match n Merge Column Join Completed	Review Column Mappings: <input type="button" value="Display Column Mappings"/> <input type="button" value="Finalize Column Mappings"/> Column Mappings Finalized	Model to Match N Merge Rows: <input type="button" value="Match N Merge Rows"/> Match n Merge Row Join Completed																												
Match N Merge For Rows Completed 21:38:17 : Creating results output file 21:38:17 : Aligning rows 21:36:33 : Starting row merge 21:36:33 : Match N Merge For Columns Completed 21:23:09 : making predictions on dataset2 21:22:55 : creating model based on dataset1 21:22:03 : create X's & Y's 21:22:03 : commence feature engineering 21:21:58 : importing datasets & transforming dataframes 21:21:58 : importing packages 21:21:52 : Wed, 06 Dec 2017 21:17:34:Starting execution ... 21:17:34 :	Review joined columns and modify appropriately. When completed, finalize column mappings by clicking on the 'Finalize Column Mappings' button above. <table border="1"> <thead> <tr> <th>Base_Column</th> <th>Matched_Column</th> <th>Relative_Match_Accuracy</th> <th>Change Column Mapping</th> </tr> </thead> <tbody> <tr> <td>business_address</td> <td>address</td> <td>0.9521791767554479</td> <td><input type="button" value="Keep Match"/> <input type="button" value="Change"/></td> </tr> <tr> <td>business_city</td> <td>city</td> <td>1.0</td> <td><input type="button" value="Keep Match"/> <input type="button" value="Change"/></td> </tr> <tr> <td>business_name</td> <td>name</td> <td>0.9507667473769169</td> <td><input type="button" value="Keep Match"/> <input type="button" value="Change"/></td> </tr> <tr> <td>business_postal_code</td> <td>postal_code</td> <td>0.9981088463963017</td> <td><input type="button" value="Keep Match"/> <input type="button" value="Change"/></td> </tr> <tr> <td>business_state</td> <td>state</td> <td>1.0</td> <td><input type="button" value="Keep Match"/> <input type="button" value="Change"/></td> </tr> <tr> <td>inspection_date</td> <td>longitude</td> <td>0.882768361581921</td> <td><input type="button" value="Keep Match"/> <input type="button" value="Change"/></td> </tr> </tbody> </table>		Base_Column	Matched_Column	Relative_Match_Accuracy	Change Column Mapping	business_address	address	0.9521791767554479	<input type="button" value="Keep Match"/> <input type="button" value="Change"/>	business_city	city	1.0	<input type="button" value="Keep Match"/> <input type="button" value="Change"/>	business_name	name	0.9507667473769169	<input type="button" value="Keep Match"/> <input type="button" value="Change"/>	business_postal_code	postal_code	0.9981088463963017	<input type="button" value="Keep Match"/> <input type="button" value="Change"/>	business_state	state	1.0	<input type="button" value="Keep Match"/> <input type="button" value="Change"/>	inspection_date	longitude	0.882768361581921	<input type="button" value="Keep Match"/> <input type="button" value="Change"/>	Download results.csv file
Base_Column	Matched_Column	Relative_Match_Accuracy	Change Column Mapping																												
business_address	address	0.9521791767554479	<input type="button" value="Keep Match"/> <input type="button" value="Change"/>																												
business_city	city	1.0	<input type="button" value="Keep Match"/> <input type="button" value="Change"/>																												
business_name	name	0.9507667473769169	<input type="button" value="Keep Match"/> <input type="button" value="Change"/>																												
business_postal_code	postal_code	0.9981088463963017	<input type="button" value="Keep Match"/> <input type="button" value="Change"/>																												
business_state	state	1.0	<input type="button" value="Keep Match"/> <input type="button" value="Change"/>																												
inspection_date	longitude	0.882768361581921	<input type="button" value="Keep Match"/> <input type="button" value="Change"/>																												

When process has completed, click on the 'Download results.csv file' to view results.