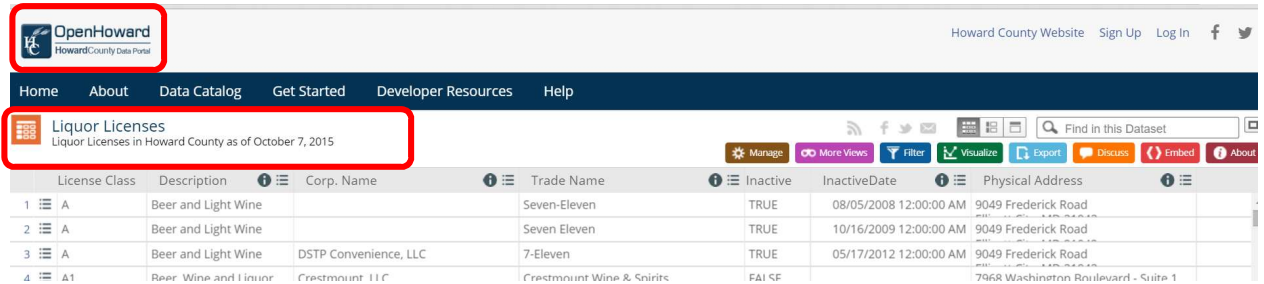


Workflow Documentation

1. Go to the source of focal dataset: <https://opendata.howardcountymd.gov/Licenses-and-Permits/Liquor-Licenses/tk3t-mn7e>



OpenHoward
Howard County Data Portal

Howard County Website Sign Up Log In

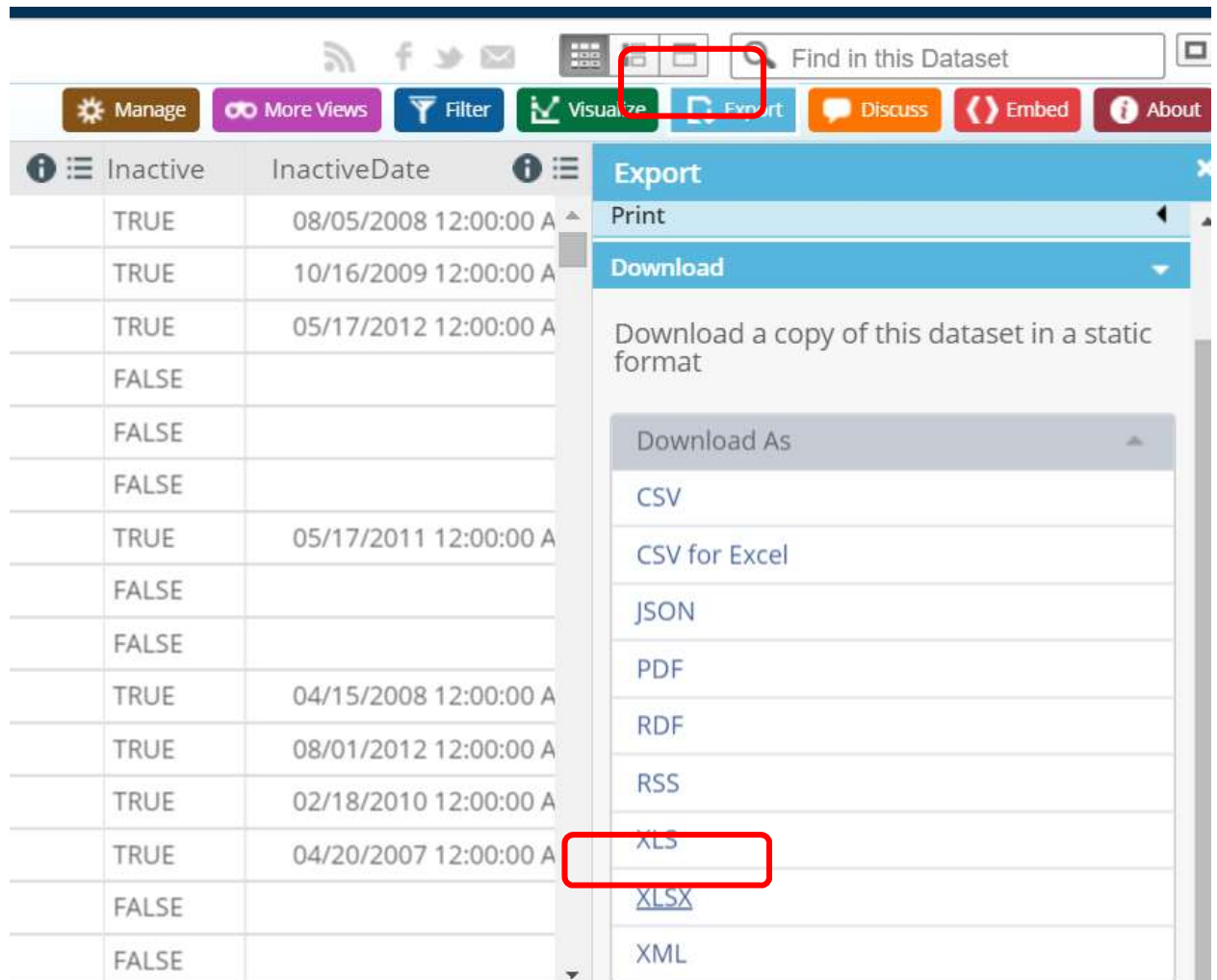
Home About Data Catalog Get Started Developer Resources Help

Liquor Licenses
Liquor Licenses in Howard County as of October 7, 2015

Manage More Views Filter Visualize Export Discuss Embed About

	License Class	Description	Corp. Name	Trade Name	Inactive	InactiveDate	Physical Address
1	A	Beer and Light Wine		Seven-Eleven	TRUE	08/05/2008 12:00:00 AM	9049 Frederick Road
2	A	Beer and Light Wine		Seven Eleven	TRUE	10/16/2009 12:00:00 AM	9049 Frederick Road
3	A	Beer and Light Wine	DSTP Convenience, LLC	7-Eleven	TRUE	05/17/2012 12:00:00 AM	9049 Frederick Road
4	A1	Beer, Wine and Liquor	Crestmount LLC	Crestmount Wine & Spirits	FALSE		7968 Washington Boulevard - Suite 1

2. Download a copy of the dataset by selecting Export option. In 'Download As' option, select 'XLSX'. Name the downloaded file (*Dataset 1*)– *Focal Dataset -Liquor_Licenses.xlsx*



Find in this Dataset

Manage More Views Filter Visualize Export Discuss Embed About

Export

Print

Download

Download a copy of this dataset in a static format

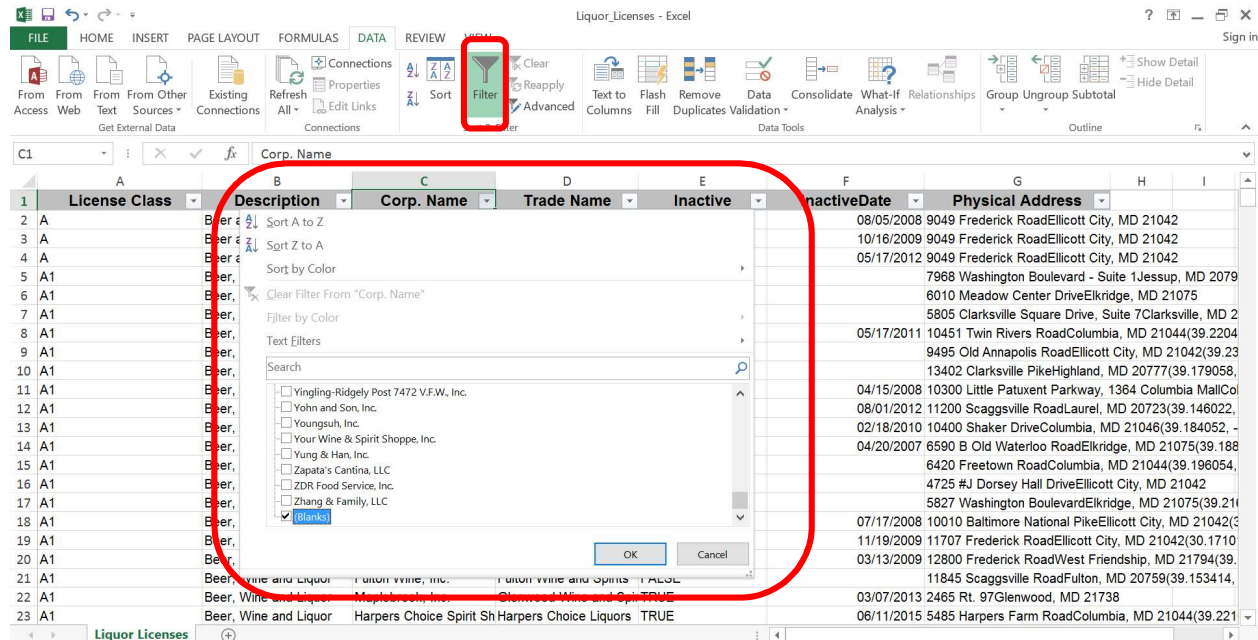
Download As

- CSV
- CSV for Excel
- JSON
- PDF
- RDF
- RSS
- XLS**
- XLSX**
- XML

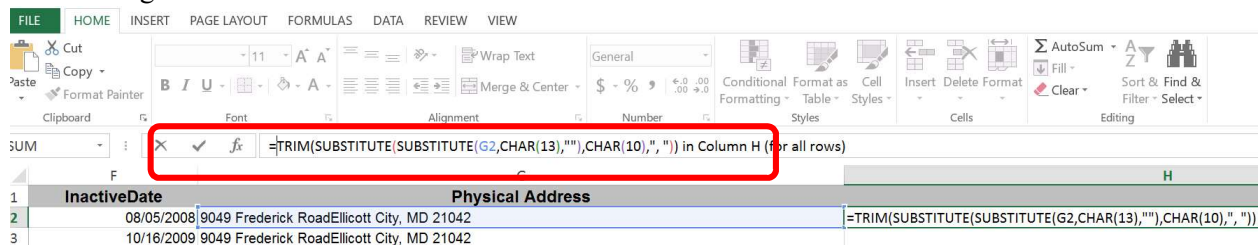
Inactive	InactiveDate
TRUE	08/05/2008 12:00:00 AM
TRUE	10/16/2009 12:00:00 AM
TRUE	05/17/2012 12:00:00 AM
FALSE	
FALSE	
FALSE	
TRUE	05/17/2011 12:00:00 AM
FALSE	
FALSE	
TRUE	04/15/2008 12:00:00 AM
TRUE	08/01/2012 12:00:00 AM
TRUE	02/18/2010 12:00:00 AM
TRUE	04/20/2007 12:00:00 AM
FALSE	
FALSE	

3. The next step is cleansing the data extracted. Following activities are carried out in this step

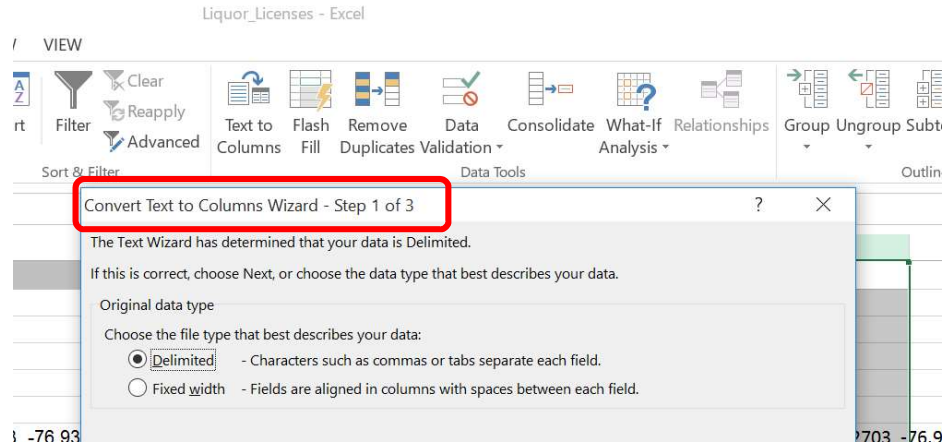
- a. Remove Null values: Use 'Filter' option in Excel on each Column of data set to determine 'Blanks'. Delete the blank rows identified.



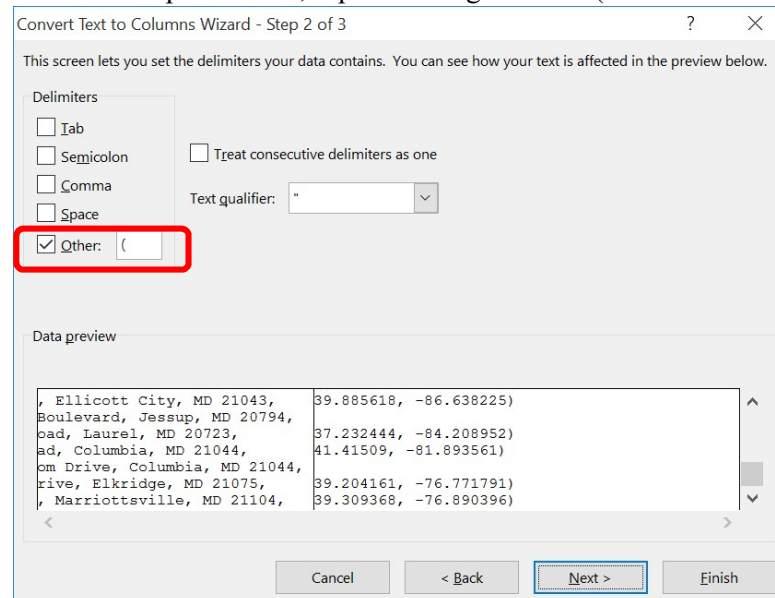
- b. Cleansing Physical Address column:
Add function - =TRIM(SUBSTITUTE(SUBSTITUTE(G2,CHAR(13),""),CHAR(10),"")) in Column H (for all rows). With this formula the line breaks present in cells in Column G get removed.



- c. Separating Latitude and Longitude from Physical Address (Column G): Select Column H
-> Go to Data -> Text to Column
In Convert text to column wizard, select 'Delimited'; Click Next



In second step of wizard, separate using 'Other - ('. Click Next-> Finish



- d. Repeat Step c, to further cleanse the column H and separate Address, Latitude and Longitude in different columns.
 - e. Once the columns are separated, delete original column G for physical address
 - f. The new formatted sheet is represented as
4. Next, extract the additional data set from : <https://opendata.howardcountymd.gov/Public-Safety/Howard-County-Police-Department-Call-For-Service-2/qccx-65fg> using Step 1 and Step 2.
 5. Name the downloaded data set as *Additional Dataset (Dataset 2) – Police CAD Event.xlsx*

6. Merging the two data sets:

- a. Create subset of Dataset 1 – Liquor License - Columns for State and Zip code can be deleted as we will merge on Address line 1
- b. Create subset of Dataset 2 – Police Department call for service – Columns for ‘Beat’ can be deleted. ‘Beat’ data is not relevant to the analysis
- c. In Dataset 2, Computer_Aided_Dispatch_Event_Type column can be filtered to remove incidents not related to criminal activity like ‘Child Abuse’, ‘911 Hangup’ etc
- d. In Dataset 2, Rename Column ‘Location’ to ‘Address Line 1’.
- e. Within each excel file create a sheet ‘Subset’ to store the subsets of raw data created above.
- f. Create new excel sheet : *Merged - Liquor Store and Criminal Incidents.xlsx*
- g. Merge the two subsets created in point 5a and 5b – Copy data from Dataset 1 and then from Dataset 2. Keep Column ‘Address Line 1’ common.

Limitations of Merging:

Address formats can vary in the two datasets which makes merging via address inconsistent. Latitude and longitude can be used effectively to merge however the second data set does not have latitude and longitude associated with it. A Perl script, can be easily used to find latitude and longitude on inputting the address. However, given the scope of assignment, the implementation of Perl script has not been performed.