

MOBILE BUSINESS KEYWORD SEARCH TOOL USER DOCUMENTATION

Background

The Mobile Business Keyword Spreadsheet Tool searches State of Washington Department of Revenue (DOR) business licensing data to identify potential mobile businesses that may present risk of stormwater pollution. The procedure is based on key business license information including North American Industry Classification System (NAICS) codes, keywords in the description of the business, and city endorsement codes (ETP IDs).

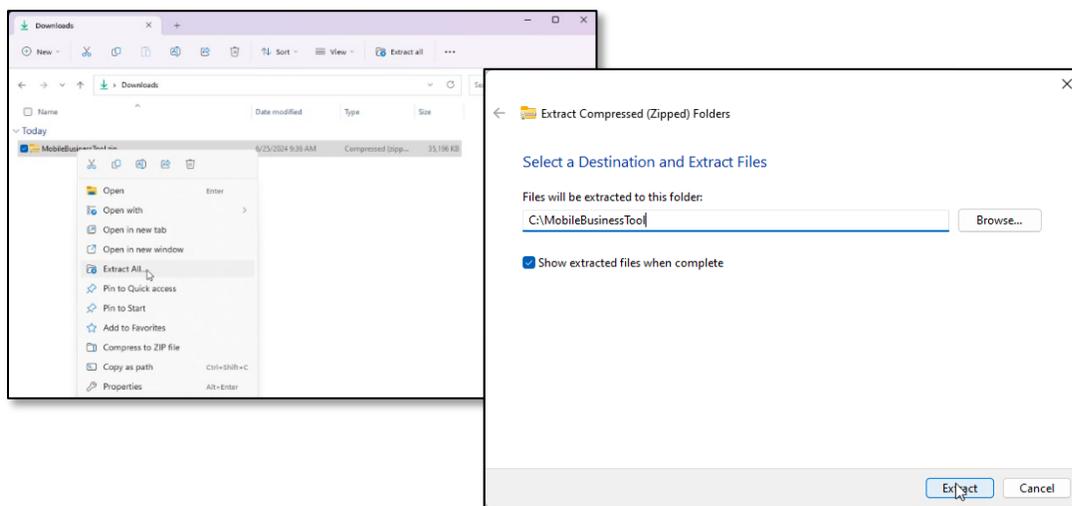
The Tool consists of a single executable file (cli.exe)¹ and a folder of comma-delimited files (.csv) that the program uses to reference, weight, and score target keywords (keywords.csv), target NAICS codes (naics.csv), and ETP IDs that relate to a business’s home jurisdiction (endorsements.csv).

The Tool was developed as part of a Stormwater Action Monitoring (SAM) Study on Mobile Business Stormwater Source Control and Coordination. This document provides instructions for using the Tool. The full description of how the Tool was developed is provided in the technical memorandum prepared for this task of the Study (Aspect Consulting, 2022). The memorandum and other deliverables from the Study can be found in the [SAM Website](#).

Setup

No installation is required.

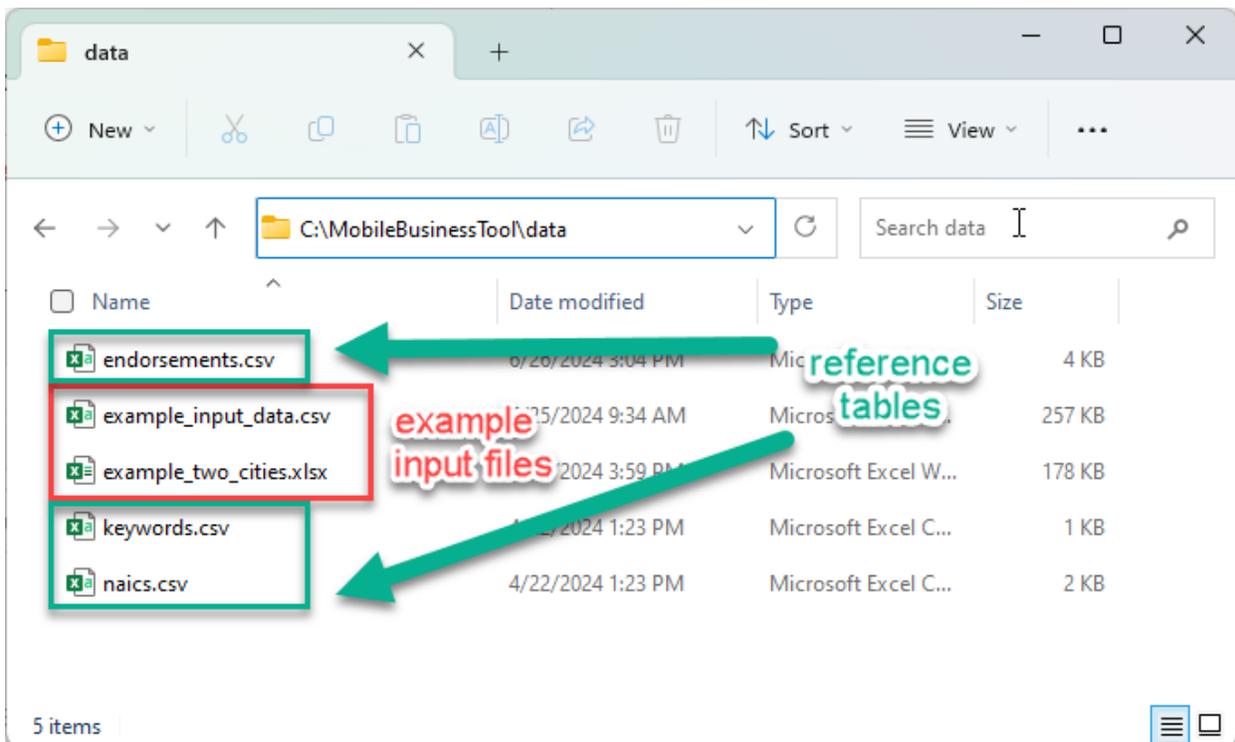
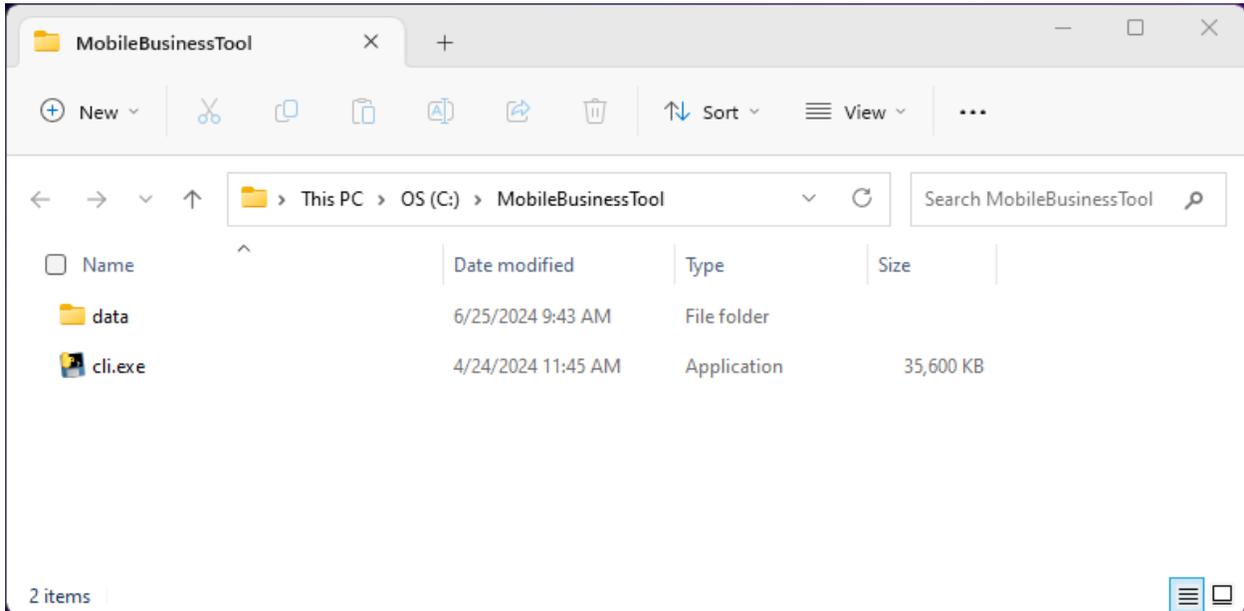
To begin, simply unzip (extract) the Mobile Business Tool zip file/folder (.zip) into a desired location on your computer:



¹“CLI” is short for “command line interface”

The unzipped folder contains two key elements:

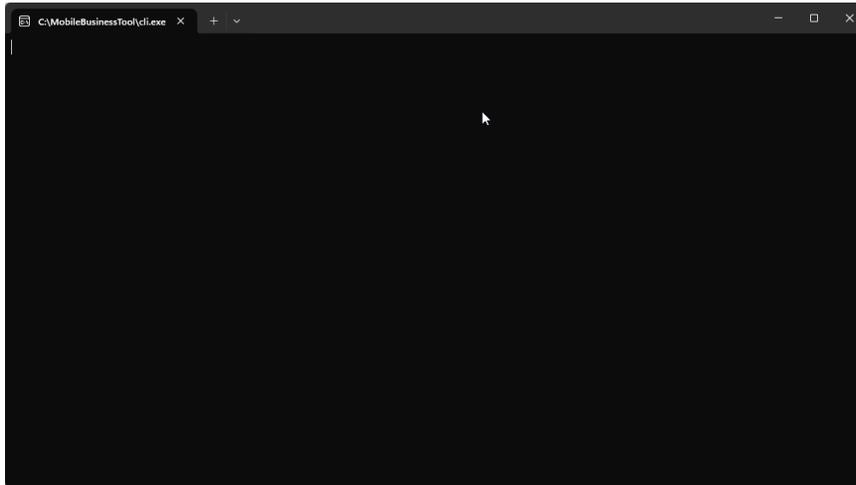
1. The executable application file itself (**cli.exe**)
2. A “**data**” folder containing three files of reference tables the application requires to run (**keywords.csv**, **naics.csv**, and **endorsements.csv**), plus two files of example input tables (**example_input_data.csv** and **example_two_cities.xlsx**).



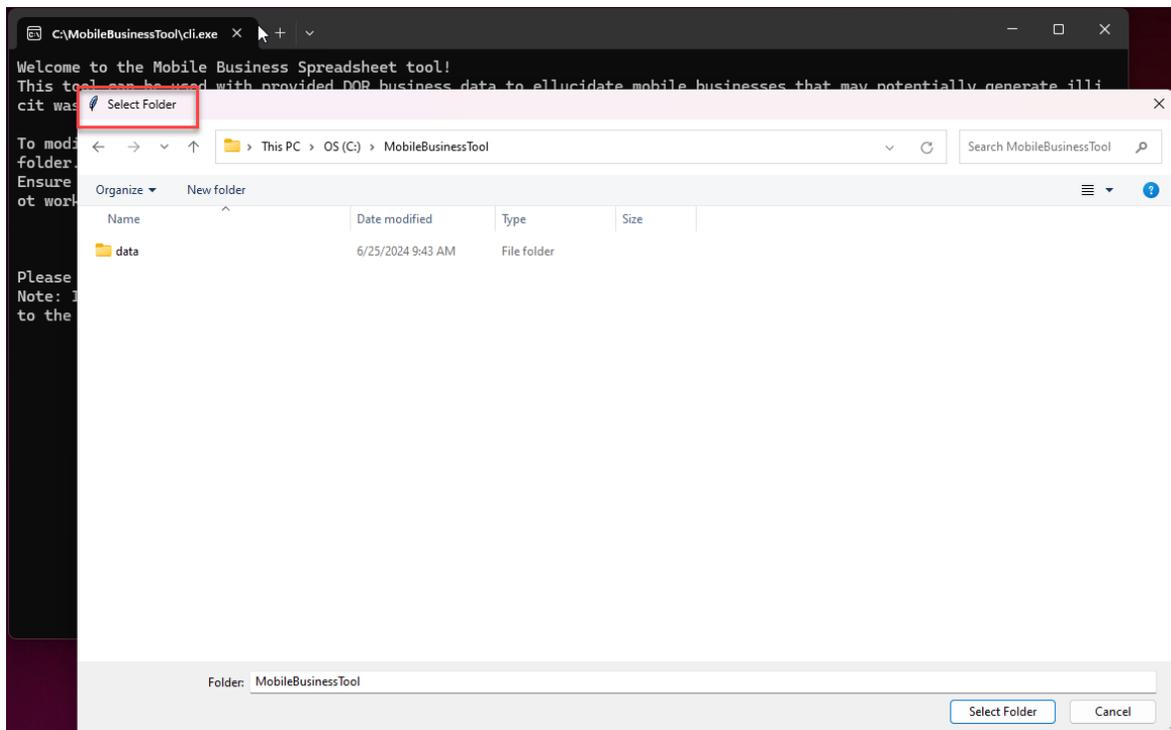
Running the Application

To run the application, double click “cli.exe”.

When the application first opens and starts to load, you will see a blank Windows Terminal (command prompt) window appear:



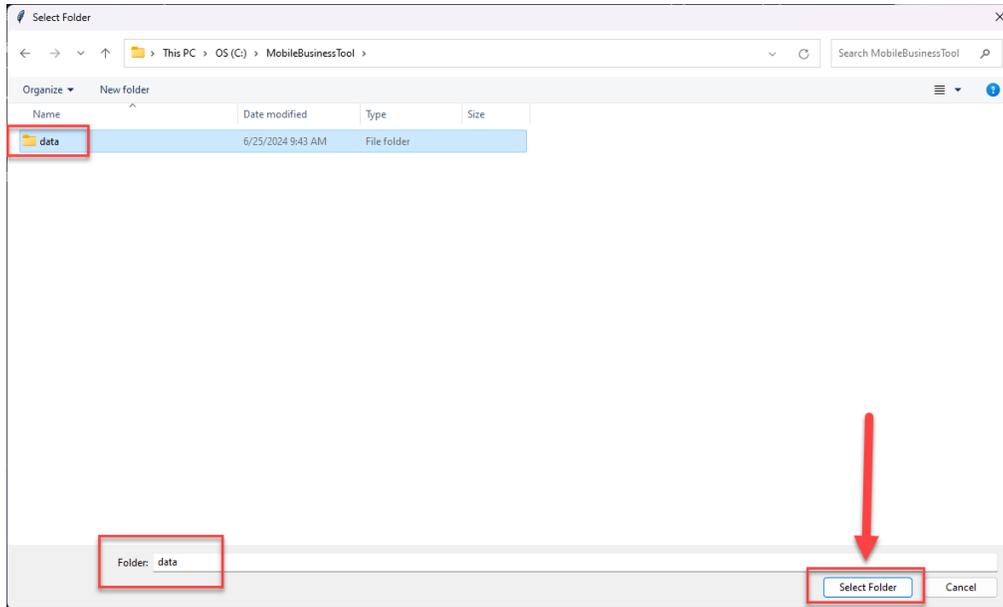
Shortly after, instructional text is printed in the terminal window and a Windows dialog box appears prompting the user to “Select Folder”:



This is asking the user to point to the folder than contains the reference tables: keywords.csv, naics.csv, and endorsements.csv (the “data” folder in the original application distribution). If any of

these files cannot be found or there is an error in processing them, the software will default to the initial pilot program values that have been hardcoded into the application.

Navigate to the appropriate folder and then click “Select Folder”.



Note: The reference tables in this folder can be modified by the user to customize the list of keywords searched, the NAICS codes, and the endorsements. However, **care should be taken to preserve the names of the files as well as the structure of the data (columns, column names)**. If file names or column names are changed, the application will fail to load the tables and default to values hardcoded into the software. See the “**Reference Tables**” section in this document for more information. After the data folder is identified, the application reads the values of the reference tables into the system:

```
C:\MobileBusinessTool\cli.exe x + v - □ x
Welcome to the Mobile Business Spreadsheet tool!
This tool can be used with provided DOR business data to elucidate mobile businesses that may potentially generate il
licit waste.

To modify target keywords, city ETP-IDs, or priority NAICS codes, please modify the corresponding csv files in the dat
a folder.
Ensure the edited files follow the same format as the defaults (as described in the documentation), or the program may
not work as intended.

Please select the data folder containing the keywords, naics, and endorsement csv files.
Note: If the city_summary files cannot be found, or there is an issue with the file or schema, the program will defaul
t to the default values.

Loading data...

Loading CSV file...

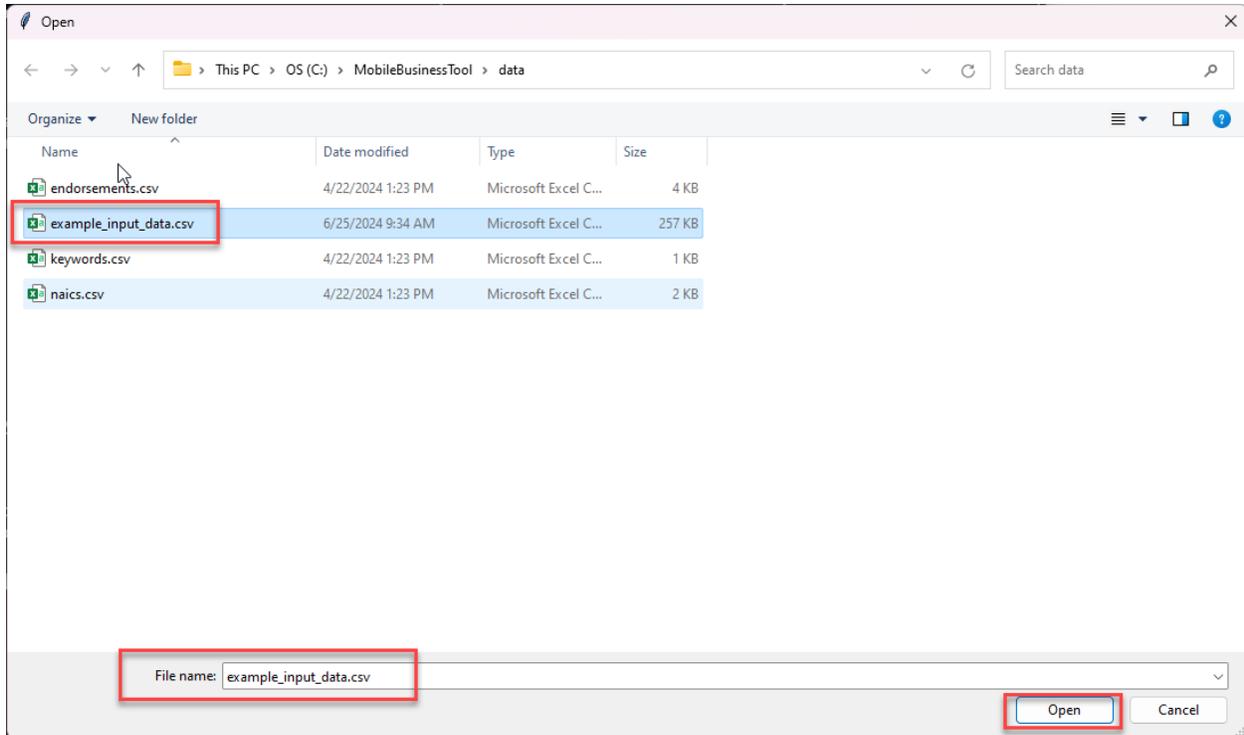
Loading CSV file...

Loading CSV file...

Successfully loaded keywords, NAICS, and endorsement data!

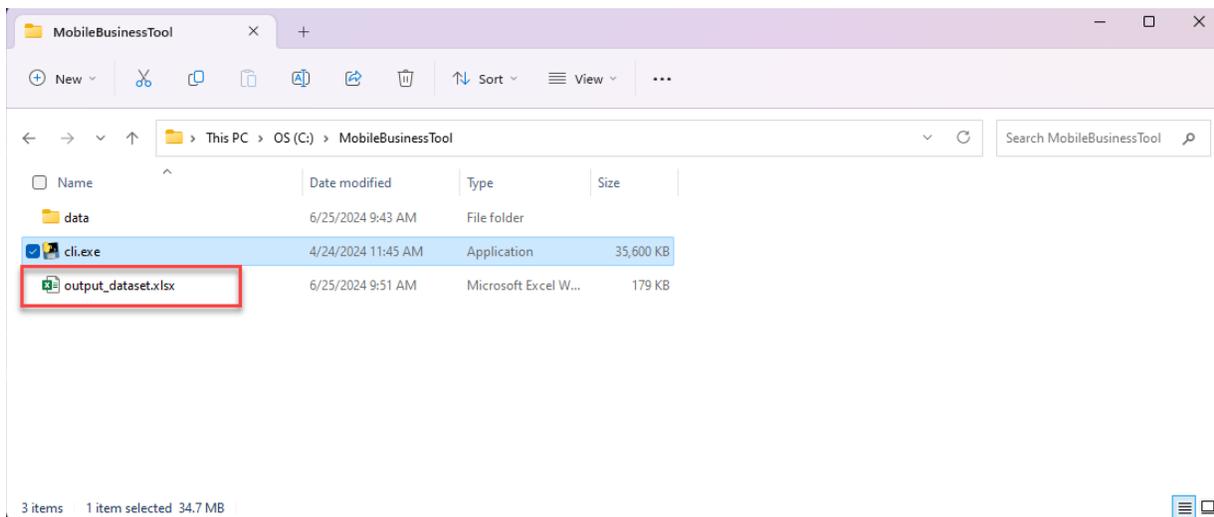
Please select the business data file.
Ensure this file follows the schema/format as described in the documentation.
```

In the next window that appears, the user is prompted to point to the “business data file”. This is the file provided by the user that contains the business license information, including business name, NAICS Codes, business description, UBI number, ETP IDs, etc. The Data folder includes an example data set that can be used for this step.



This file can be stored anywhere on your computer in either csv or xlsx format.

Assuming the input business data file contains all the required fields with the expected names (see following section), the application will run and generate an output file called “**output_dataset.xlsx**” in the same directory as where the cli.exe is located. This file can be used to review potential mobile businesses identified by the Tool.



Note: The application will always write out the results to output_dataset.xlsx in the same directory as the cli.exe **and will overwrite any existing copy of output_dataset.xlsx**. Previous versions of output_dataset.xlsx should be renamed or moved to another directory if you do not want them to be overwritten.

Business Data File Requirements

The application prompts the user to select a table of business information extracted from the Department of Revenue Business Licensing Service Portal (BLS Portal). These can either be provided as .csv or .xlsx.

Multiple cities (separated by sheets in an .xlsx file) or a single city's data (.csv or .xlsx) can be provided.

The final section of this document describes the process of downloading data from the BLS Portal.

FILE AND SHEET NAMES

To successfully look up and reference the endorsement (ETP) codes for mobile businesses in endorsements.csv, **the input file name and/or sheet name(s) should be identical to the city name(s) (in all caps) as it appears in endorements.csv**.

For CSVs, the file name should be the city name in all caps, just as it appears in endoresments.csv (e.g., "SEATAC.csv" or "REDMOND.csv")

For XLSX files, the sheet name(s) should correspond to the (all caps) city names as they appear in endorements.csv. Excel files can contain multiple sheets with data for different cities.

Note: in any case, the name of the file/sheet(s) MUST be less than 30 characters or the application will fail.

REQUIRED FIELDS

It is recommended that the user's input file matches the format in **example_input_data.csv**.

However, the application will run if the following five fields are provided (along with any other columns/fields):

Column/Field	Expected Field Name
Business Description	"ProdDescription":
Business Unified Business Identifier Number (UBI)	"bus-ubi"
(Legal) Name of the Business	"LE-name"
Primary business NAICS codes	"DOR-NAICS-Primary"
Secondary business NAICS codes	"DOR-NAICS-Secondary"

If the input data file contains the five fields above and all are named using the **Expected Field Names** above, then the application should run without additional user input.

However, if the five required fields exist but have different names, the application will prompt the user to identify them, one-by-one:

```
C:\MobileBusinessTool\cli.exe x + v
Processing example_input_data...
Could not find city example_input_data in endorsements! Not filtering by ETP-ID.
columns = ['Run-Date', 'Name', 'Firm', 'Description', 'UBI', 'NAICS-Primary', 'NAICS-Secondary', 'ETP', 'Mail-City', 'Mail-State', 'Mail-Zip', 'Home-Operated-Business', 'Phys-City', 'Phys-State', 'Phys-Zip', 'City Limits']
Column ProdDescription not found in dataframe.
Please type in the name of the column that contains the Business Description, exactly as it appears in the provided business data. Detected column names are listed above.
Description
Column bus-ubi not found in dataframe.
Please type in the name of the column that contains the Business Unified Business Identifier Number, exactly as it appears in the provided business data. Detected column names are listed above.
UBI
Column LE-name not found in dataframe.
Please type in the name of the column that contains the Legal Name of the Business, exactly as it appears in the provided business data. Detected column names are listed above.
Name
Column DOR-NAICS-Primary not found in dataframe.
Please type in the name of the column that contains the The primary business NAICS codes, exactly as it appears in the provided business data. Detected column names are listed above.
NAICS-Primary
```

a list of column names in the provided input table, for reference

User-typed identification of alternate field name(s) for each required field

Reference Tables

The reference values stored in **keywords.csv**, **naics.csv**, and **endorsements.csv** impact how potential mobile business are sorted, weighted, scored, filtered, or otherwise flagged by the software.

Users of this software can adjust the lists of reference values stored in **keywords.csv**, **naics.csv**, and **endorsements.csv**. However, **only the contents of the each should be adjusted. Any changes to the number of columns, the columns names, or the file names/types would break the functionality to the tool.**

KEYWORDS

The file **keywords.csv** has two columns: “**Keyword**” and “**Weight**”. The values in the keyword column are strings of characters that the Tool searches for in the business description that are *potentially* indicative of mobile business—or explicitly indicative of a business that isn’t of interest. The “**Weight**” is a numerical value between 0 and 1 that represents a degree of relevance, with “1” being highly relevant/indicative and 0 being the least relevant/indicative. It is suggested to use values of 0.25, 0.5, 0.75, and 1 to have four tiers of relevance for various keywords.

Entries with a value of “-1” are there to explicitly identify and exclude keywords/descriptions that are not of interest for stormwater. For example, a business might have the word “mobile” in its description, which, by default, has a weight of 0.75. However, if the description also has the word “wireless” or is part of the term “mobile home” then the weight of “-1” for “wireless” or “mobile home” tells the Tool that that instance of the word “mobile” isn’t of interest in that case.

The keywords.csv file can also contain common misspellings or typos (e.g. “service” in addition to “service”).

Some examples from **keywords.csv** are below:

Keyword	Weight	Keyword	Weight	Keyword	Weight
arborist	0.5	mobile auto	1	pressure wash	1
carpet clean	0.75	mobile clean	1	repair	0.25
food truck	1	mobile service	1	service	0.25
food vendor	0.25	oil	0.5	service	0.25
grease	0.5	on site	0.5	vactor	1
hvac	0.25	paint	0.25	vent hood	1
installation	0.25	pesticide	0.5	window wash	1
landscape	0.75	pet groom	0.25	mobile + modular	-1
maintenance	0.5	pipe clean	0.5	mobile home	-1
mobile	0.75	pool	0.25	wireless	-1

NAICS CODES

The file **naics.csv** contains three columns: “**HIGHER_PRIORITY_NAICS**”, “**MEDIUM_PRIORITY_NAICS**”, and “**LOWER_PRIORITY_NAICS**”. Each of these columns contains lists of NAICS codes that are ranked high, medium, or low priority for the purpose of identifying mobile businesses. The NAICS ranking was done by the Study team based on perceived risk of stormwater pollution (see technical memorandum for full explanation, Aspect 2022). NAICS codes can be moved from one column to another to adjust their priority, additional NAICS codes can be added to any of the columns, and NAICS codes can also be removed from the table. However, blank cells should be avoided between entries in each column.

ENDORSEMENTS/ETP CODES

An Endorsement Type Identifier (ETP-ID) specific to “non-resident” businesses indicates that a business is applying for a license in a city in which their business does not reside—which is an indicator that a business is located outside the licensed jurisdiction to do work. While a non-resident endorsement alone does not indicate a mobile business, it would be a likely attribute of a mobile business.

The file **endorsements.csv** has two columns: “City”, and “ETP_ID”. The values in the “City” column need to match the file/sheet name (as described in the previous section) and the value in “ETP_ID” is the non-resident ETP ID code used by that city.

If the system finds a match in endorsements.csv for the given city name (as provided in the file/sheet name of the input data), **the output of the tool will be filtered for only those businesses that have that endorsement code.**

If the system does not find a match in endorsements.csv for the names of the file/sheet, the system will simply warn the user and not perform any filtering or sorting based on the ETP-ID.

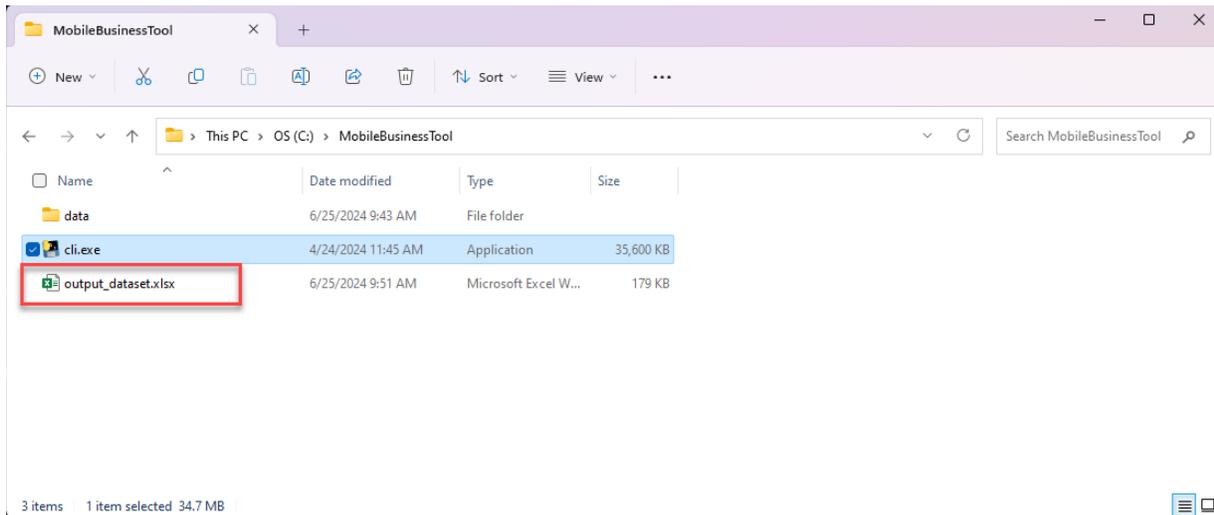
```
Processing example_input_data...
```

```
Could not find city example_input_data in endorsements! Not filtering by ETP-ID.
```

If filtering the output dataset for this non-resident ETP-ID is not desired, simply change the name of the input file/sheet to be something other than the city name(s) or remove the entry from endorsements.csv.

Output

After successfully running, the Mobile Business Spreadsheet tool creates an output file called “**output_dataset.xlsx**” in the same directory as the *cli.exe*. **The application will overwrite any existing copy of output_dataset.xlsx.** Any previous version of output_dataset.xlsx should be renamed or moved to another directory if you do not want it to be overwritten.



The output excel file contains the following sheets/tabs:

- A sheet (or multiple sheets) containing the input data, sorted, with added formatting and additional fields (see below)
- Higher Priority NAICS Summary
- Medium Priority NAICS Summary
- Lower Priority NAICS Summary
- City Summary
- NAICS Summary
- Keyword Summary

- **City Summary:** A breakdown of the minimum, maximum, average, and median weighted keyword search scores for each input city

City	Minimum Weighted Search Score	Average Weighted Search Score	Median Weighted Search Score	Maximum Weighted Search Score
CITY NAME 1	0	0.214285714	0.25	1.25
CITY NAME 2	0	0.277552204	0.25	3

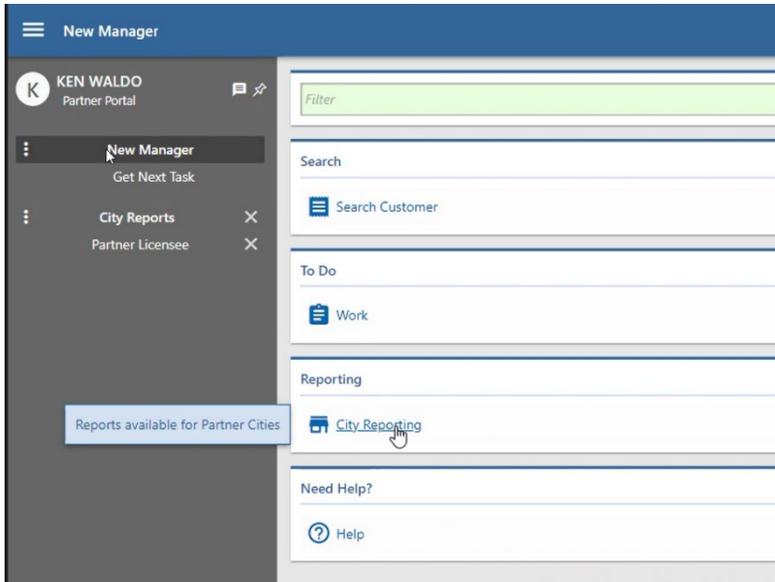
- **NAICS Summary:** A summary of the number of businesses and weighted score statistics by NAICS priority.

NAICS Priority	NAICS CODES	Business Count	Minimum Weighted Search Score	Average Weighted Search Score	Median Weighted Search Score	Max Weighted Search Score
HIGHER_PRIORITY_NAICS	['482', '485', '484', '493', '4881', '4882', '4884', '4889', '2211', '441', '447', '811192', '8112', '8113', '8114', '8111']	21	-1	0.30952381	0.25	1.25
MEDIUM_PRIORITY_NAICS	['1152', '236', '237', '238', '311', '312', '321', '3221', '3222', '325', '3241', '326', '316', '331', '332', '333', '334', '335', '336', '423140', '423930', '423110', '4233', '4237', '424930', '4244', '4246', '4247', '4248', '444', '445', '5321', '5324', '621910', '6112', '6113', '6115', '327', '4238', '6111']	541	-1	0.170979667	0.25	1.75
LOWER_PRIORITY_NAICS	['323', '722', '712']	7	0	0.142857143	0.25	0.25

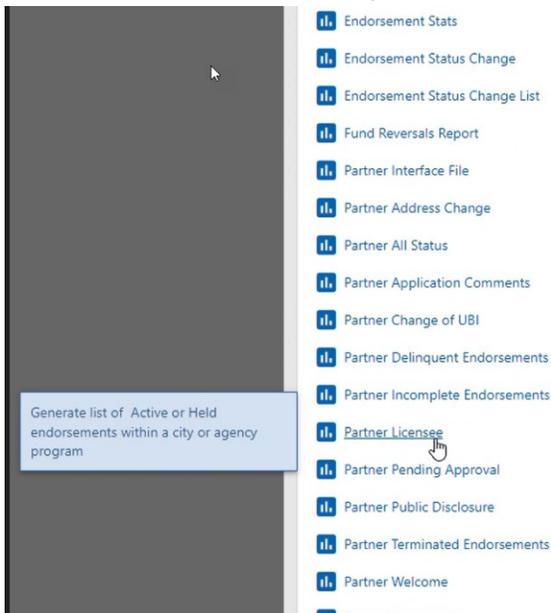
- **Keyword Summary:** A count of the number of matches for all the keywords listed in keywords.csv.

Downloading Business Licensing Data from the Department of Revenue Business License Service

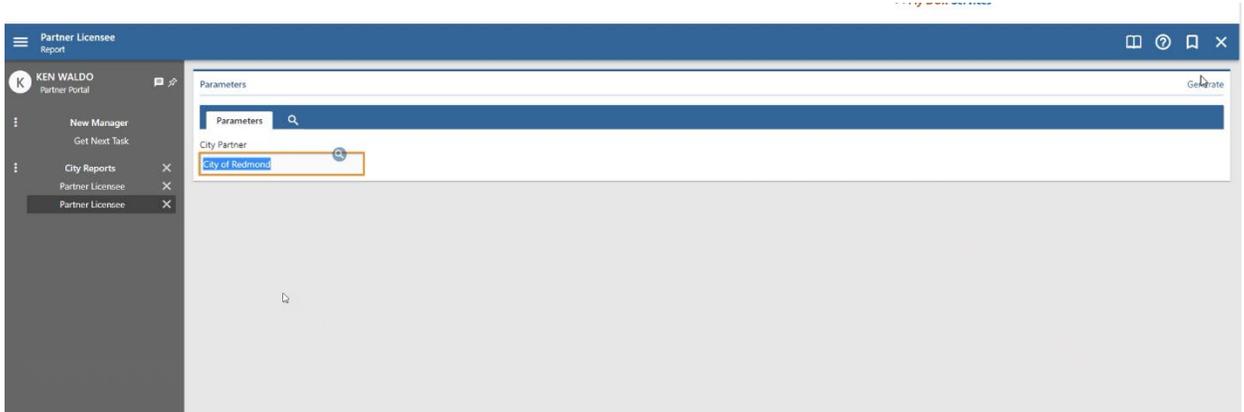
1. Log onto the BLS Portal and navigate to the Reporting tab and select City Reporting:



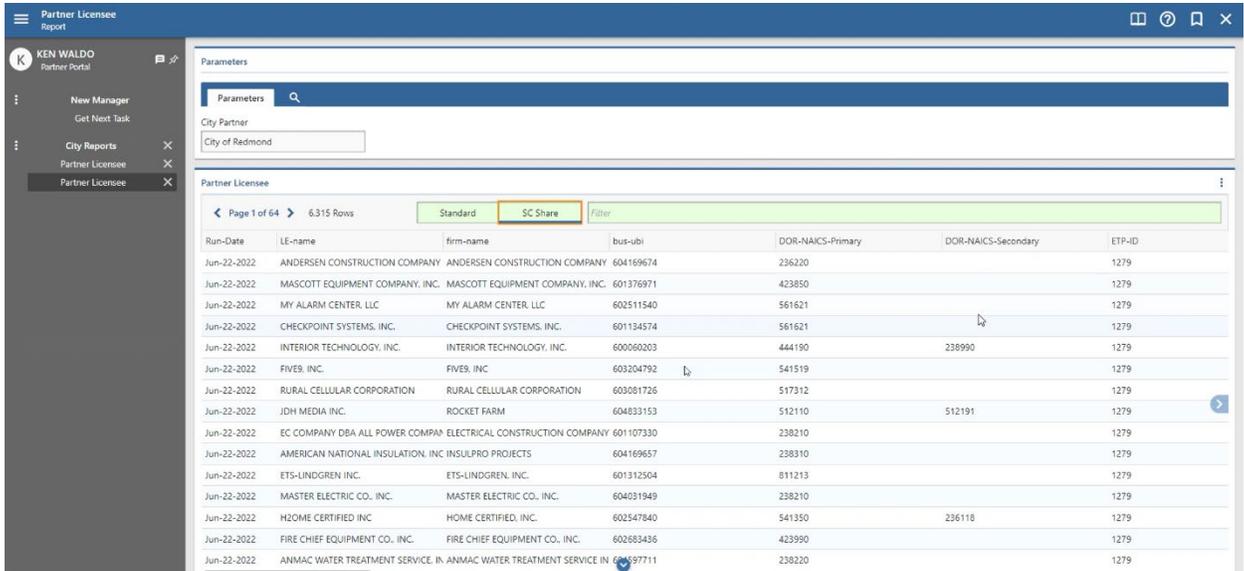
2. Select the Partner Licensee option:



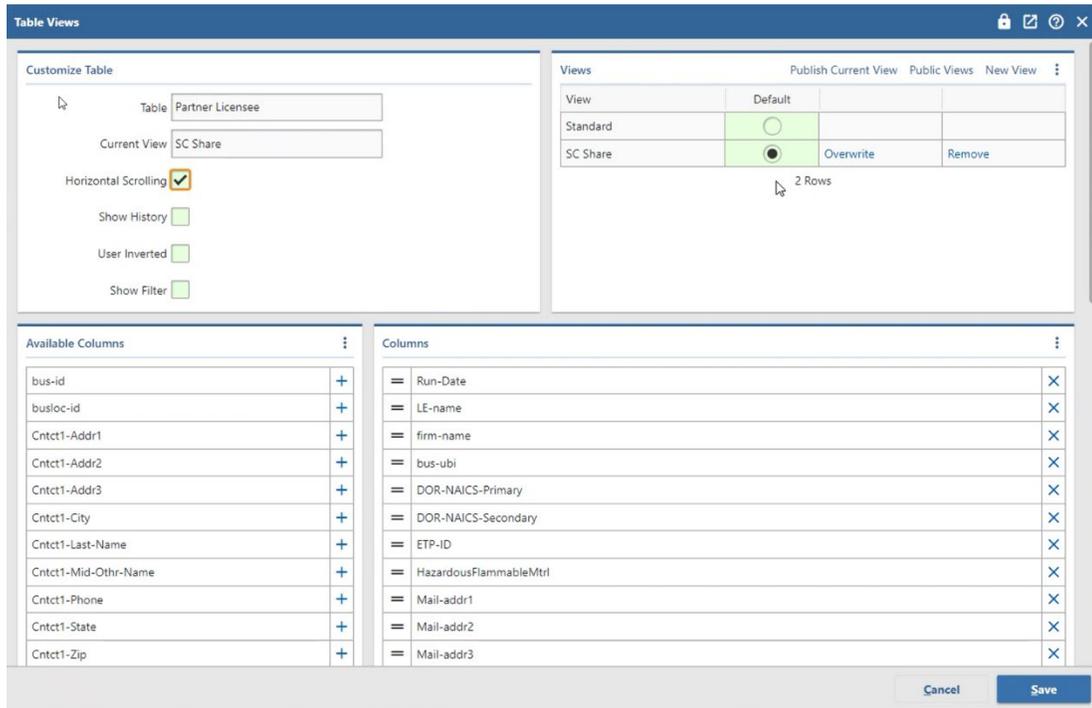
3. Enter the jurisdiction name and click **Generate**



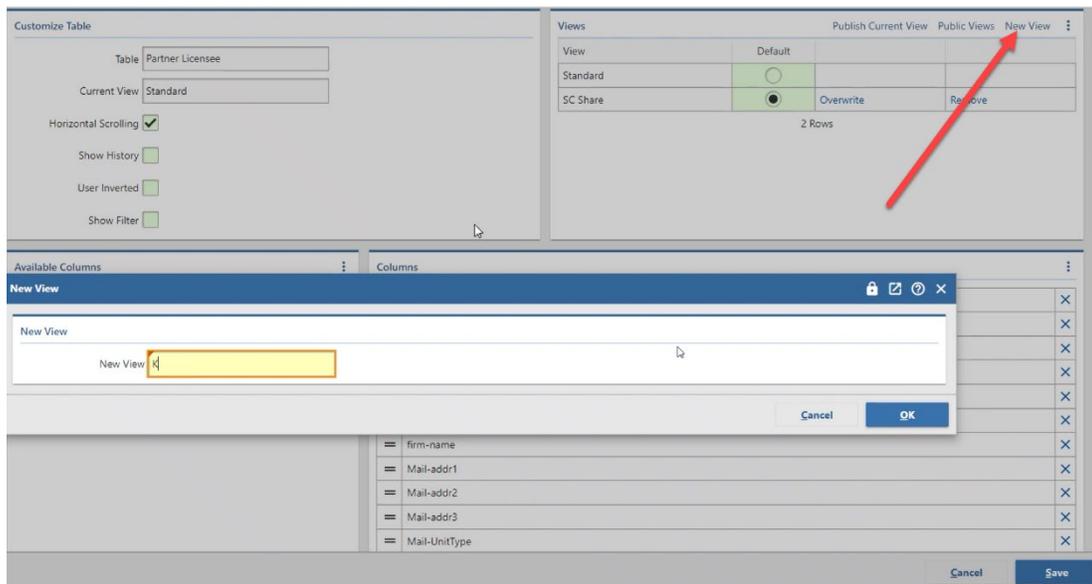
4. All current business licenses in the jurisdiction will appear in a preview window:



- Open the Table Views window to adjust the data output:



- Open a New View to customize the table and name it “BLS license download_[6-digit date]”:



7. Add (+ symbol) or remove (x symbol) columns to include only the following data fields. If necessary, rearrange the data fields to be in this order (using the parallel lines symbol on the left of the column name):

- bus-ubi, the business license number
- LE-name, legal business name
- firm-name, the “doing business as” name
- Mail-addr1, the mailing address of the business
- Mail-addr2, the mailing address of the business
- Mail-addr3, the mailing address of the business
- Mail-UnitType, the mailing address of the business
- Mail-Unit, the mailing address of the business
- Mail-City, the mailing address of the business
- Mail-State, the mailing address of the business
- Mail-Zip, the mailing address of the business
- Mail-Zip+4, the mailing address of the business
- Phys-addr1, the physical address of the business
- Phys-addr2, the physical address of the business
- Phys-addr3, the physical address of the business
- Phys-UnitType, the physical address of the business
- Phys-Unit, the physical address of the business
- Phys-City, the physical address of the business
- Phys-State, the physical address of the business
- Phys-Zip, the physical address of the business
- Phys-Zip+4, the physical address of the business
- ETP-ID, the endorsement identifiers for the business (includes “non-resident”)
- HazardousFlammableMtrl, indicates if the business uses hazardous or flammable materials
- ProdDescription, provides the business description as written by the applicant
- DOR-NAICS-Primary, indicates the primary NAICS code for the business
- DOR-NAICS-Secondary, indicates the secondary NAICS code for the business

8. Save the custom table that you created with the columns listed above

The screenshot shows the 'Table Views' application interface. The top section is titled 'Table Views' and contains two main panels: 'Customize Table' and 'Views'.

Customize Table Panel:

- Table: Partner Licensee
- Current View: SC Share
- Horizontal Scrolling:
- Show History:
- User Inverted:
- Show Filter:

Views Panel:

View	Default		
Standard	<input type="radio"/>		
SC Share	<input checked="" type="radio"/>	Overwrite	Remove

2 Rows

Available Columns Panel:

Column Name	Action
bus-id	+
busloc-id	+
Cntct1-Addr1	+
Cntct1-Addr2	+
Cntct1-Addr3	+
Cntct1-City	+
Cntct1-Last-Name	+
Cntct1-Mid-Othr-Name	+
Cntct1-Phone	+
Cntct1-State	+
Cntct1-Zip	+

Columns Panel:

Column Name	Action
Run-Date	X
LE-name	X
firm-name	X
bus-ubi	X
DOR-NAICS-Primary	X
DOR-NAICS-Secondary	X
ETP-ID	X
HazardousFlammableMtrl	X
Mail-addr1	X
Mail-addr2	X
Mail-addr3	X

At the bottom right of the interface, there are two buttons: 'Cancel' and 'Save'. A red arrow points to the 'Save' button.