assignTags Documentation

Function
assignTags efficiently assigns tags to merchants, or any entity, according to the merchant's category and subcategory. New tags to be assigned to merchants are merged with the merchant's existing tags. Tags assigned to merchants are unique, in lower-case, and have had accents removed.

Implementation
Step 1:
- Create tags.xlsx, a legend of search terms to be applied to merchants (entities) according to their category and subcategory labels. It has columns:
  - *Category*: an entity category.
  - *Subcategory*: a subcategory of a *Category*. Leave cell empty for tags that apply to any entity in the category.
  - *Tags*: a comma-separated list of new tags (e.g. ["tag1", "tag2"]) to be assigned to entities that have been labelled with subcategory in the other sheet, merchants.xlsx.

Step 2:
- Create merchants.xlsx, containing tagging information for merchants (entities). It has columns:
  - *Brand_Name*: an entity name.
  - *Category*: as previously defined.
  - *Subcategories*: a comma-separated list of *Subcategory* subcategories that apply to the entity.
  - *Existing_Tags*: tags that have already been assigned to an entity.

Step 3:
- Run the sheets through the program.
- The output file called "assigned tags.csv" contains the columns:
  - *Name*: as previously defined.
  - *Category*: as previously defined.
  - *Subcategories*: as previously defined.
  - *Tags*: a generated column that contains the new tags assigned to the entity based on the subcategories the user has labelled the entity with.

Possible Additions
- Handling user data entry errors.
- Could link directly to a database.

Example

Step 1: create or retrieve legend for tag assignment (tags.xlsx)

Category	Subcategory	Assignment Conditions	Tags
ART		category catch-all	['art']
ART	SUPPLIES	sells art supplies	['art supplies', 'paint', 'brush', 'sketch', 'draw', 'color', 'canvas', 'fabric']
ART	GALLERY	is an art gallery	['gallery', 'art gallery', 'photography']
BOOKS		category catch-all	['book', 'chapters', 'indigo', 'coles', 'bookstore', 'library']
BOOKS	COMICS	is a comic shop	['comic', 'collectable']
BOOKS	USED	is a used book store	['used books', 'used bookstore']
BUTCHER		category catch-all	['butcher', 'meat', 'steak', 'sausage', 'grass fed', 'beef']
CAFE		category catch-all	['cafe', 'coffee']
CAFE	COFFEE	is a coffee shop	['coffee roaster', 'coffee beans', 'espresso', 'cappiccino', 'latte', 'americano']
CAFE	TEA	is a tea shop	['tea']
CAFE	BUBBLE_TEA	is a bubble tea shop	['bubble tea', 'boba']
FAST_FOOD		catch-all	['fast food', 'food', 'sandwich', 'lunch', 'drive thru']
FAST_FOOD	CLASSIC	sells classic American fast food	['burger', 'fries', 'hot dog', 'chicken nuggets', 'milk shake']
FAST_FOOD	PIZZA	is a pizza shop	['pizza', 'slice']
FAST_FOOD	HEALTHY	sells health fast food	['salad', 'bowl', 'poke', 'healthy food']
GAS_STATION		catch-all	['gas station', 'gas', 'convenience store', 'cigarettes', 'snacks', 'drinks']
TRAVEL		catch-all	['travel', 'tour', 'sightseeing', 'landmarks', 'vacation']
YOGA		catch-all	['yoga', 'yoga mat', 'hot yoga', 'fitness']

Step 2: create or retrieve table containing merchant data (merchants.xlsx)

Brand_Name	Category	Subcategories	Existing_Tags
Art Gallery of Ontario	ART	GALLERY	["AGO", "Art Gallery of Ontario", "Gallery", "Art"]
Jane's Art Supplies	ART	SUPPLIES	["art supplier", "paint"]
Second Hand Comics	BOOKS	COMICS, USED	["Marvel", "comics"]
The Blue Butcher	BUTCHER		["butcher", "meat"]
Green Tea House	CAFE	TEA, BUBBLE_TEA	["Green Tea House"]
Black Coffee Roasters	CAFE	COFFEE	["coffee", "pastries"]
Bubba's Burgers	FAST_FOOD	CLASSIC	[]
Pizza Pizza	FAST_FOOD	PIZZA	["pizza", "wings", "delivery"]
The Poke Bowl	FAST_FOOD	HEALTHY	[]
Petro Canada	GAS_STATION		[]
NYC Sightseeing Tours	TRAVEL		["Tourism"]
House of Yoga	YOGA		["yoga"]

Step 3: run assignTags.py to obtain the following output (assigned tags.csv)

	Brand_Name	Category	Subcategories	Existing_Tags	Tags
0	Art Gallery of Ontario	ART	['GALLERY']	['AGO', 'Art Gallery of Ontario', 'Gallery', 'Art']	['ago', 'art', 'gallery', 'art gallery', 'art gallery of ontario', 'photography']
1	Jane's Art Supplies	ART	['SUPPLIES']	['art supplier', 'paint']	['fabric', 'paint', 'sketch', 'brush', 'art supplies', 'color', 'art', 'art supplier', 'canvas', 'draw']
2	Second Hand Comics	BOOKS	['COMICS', 'USED']	['Marvel', 'comics']	['book', 'marvel', 'library', 'comics', 'used books', 'indigo', 'comic', 'collectable', 'used bookstore', 'bookstore', 'chapters', 'coles']
3	The Blue Butcher	BUTCHER		['butcher', 'meat']	['beef', 'steak', 'butcher', 'sausage', 'meat', 'grass fed']
4	Green Tea House	CAFE	['TEA', 'BUBBLE_TEA']	['Green Tea House']	['cafe', 'coffee', 'green tea house', 'bubble tea', 'tea', 'boba']
5	Black Coffee Roasters	CAFE	['COFFEE']	['coffee', 'pastries']	['cafe', 'cappiccino', 'americano', 'latte', 'coffee roaster', 'pastries', 'coffee', 'espresso', 'coffee beans']
6	Bubba's Burgers	FAST_FOOD	['CLASSIC']	[]	['sandwich', 'hot dog', 'chicken nuggets', 'milk shake', 'food', 'fries', 'lunch', 'burger', 'fast food', 'drive thru']
7	Pizza Pizza	FAST_FOOD	['PIZZA']	['pizza', 'wings', 'delivery']	['sandwich', 'slice', 'food', 'pizza', 'wings', 'delivery', 'lunch', 'fast food', 'drive thru']
8	The Poke Bowl	FAST_FOOD	['HEALTHY']	[]	['sandwich', 'healthy food', 'salad', 'poke', 'food', 'bowl', 'lunch', 'fast food', 'drive thru']
9	Petro Canada	GAS_STATION		[]	['cigarettes', 'gas', 'drinks', 'convenience store', 'gas station', 'snacks']
10	NYC Sightseeing Tours	TRAVEL		['Tourism']	['travel', 'vacation', 'tourism', 'tour', 'sightseeing', 'landmarks']
11	House of Yoga	YOGA		['yoga']	['yoga mat', 'hot yoga', 'fitness', 'yoga']