Task 2.1

This is a similarity matrix between different cuisines from IDF

	Amusement Parks	Argentine	Asian Fusion	Bakeries	Breweries	Cafeteria	Cajun-Creole	Casinos	Comfort Food	Diners	Ethiopian	Event Planning & Services	Filipino	Fondue	Fruits & Veggies	Grocery	Internet Cafes	Live-Raw Food	Mediterranean	Mongolian	Personal Shopping	Polish	Pretzels	Ramen	Shopping	Sushi Bars	Tea Rooms	Tex-Mex	Thai	Tour
Amusement Parks	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Argentine	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Asian Fusion	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
Bakeries	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Breweries	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cafeteria	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cajun-Creole	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Casinos	0	0	0	0	0	0	0	1	0	0	0	0.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comfort Food	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Diners	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Ethiopian	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Event Planning & Services	0	0	0	0	0	0	0	0.4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Filipino	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fondue	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fruits & Veggies	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Grocery	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Internet Cafes	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Live-Raw Food	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Mediterranean	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Mongolian	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
Personal Shopping	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Polish	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Pretzels	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Ramen	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0.9	0	0	0	0	0	0
Shopping	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Sushi Bars	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Tea Rooms	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Tex-Mex	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Thai	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Tours	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Where number is normalized similarity between two cuisines.

Here it indicates that There are barely any similarities between topics. This is because IDF is too sparse and it is difficult to identify similarity in such a sparse vector.

I also used LDA model to generate topics model and compare similarity between different cuisines based on topics and its corresponding weights, which is already provided in sample code.

	Adult Entertainment	Amusement Parks	Arabian	Brazilian	Cuban	Fondue	German	Malaysian	Mongolian	Pakistani	Ramen	Street Vendors	Turkish	Seafood	Breakfast & Brunch	Casinos	Cheesesteaks	Cinema	Comfort Food	Convenience Stores	Event Planning & Services	Food Delivery Services	Ice Cream & Frozen Yogurt	Internet Cafes	Italian	Latin American	Pubs	Tex-Mex	Thai	Tour
Adult Entertainment	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0
Amusement Parks	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0
Arabian	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0
Brazilian	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0.2	0.1	0	0	0	0	0	0
Cuban	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0
Fondue	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0
German	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0.1	0	0	0	0	0	0	0
Malaysian	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0
Mongolian	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0
Pakistani	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0.3	0	0	0	0	0	0	0
Ramen	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0
Street Vendors	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0
Turkish	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0
Seafood	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0.1	0	0	1	0	0	0	0
Breakfast & Brunch	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0.1	0	0	0	0	0	0	0
Casinos	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0.1	1	0	0	0	0	0.2	0
Cheesesteaks	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0.1	0	0	0	0	0	0	0
Cinema	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0.1	0	0	0	0	0	0	0
Comfort Food	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0.1	0	0	0	0	0	0	0
Convenience Stores	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0.1	0	0	0	0	0	0	0
Event Planning & Services	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0.1	0	0.3	0	0	0	0	0
Food Delivery Services	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0.1	0	0	0	0	0	0	0
Ice Cream & Frozen Yogurt	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.3	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1	0.1	0.2	0.1	0.1	0.1	0.2	0.1
Internet Cafes	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0.1	1	0	0	0	0	0.2	0
Italian	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.3	0	0.2	0	1	0	0	0	0.9	0
Latin American	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0.1	0	0	1	0	0	0	0
Pubs	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	1	1	0	0
Tex-Mex	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	1	1	0	0
Thai	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2	0	0	0	0	0	0	0.2	0.2	0.9	0	0	0	1	0
Tours	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	1

Here it indicates that more similarity is discovered using LDA model. This is mainly because topic models are more condense vectors, and more significant topics will provide higher weight when comparing similarities.

But the result by no mine is meaningful, because there are many similarities doesn't make sense. For example, "Ice Cream and Frozen Yogurt" cuisine has almost identical similarity to all other cuisines which is not intuitive. "Turkish" and "Ramen" cuisines are also similar to each other which is not a meaningful result.

Task 2.2

To Improve similarity discovery, I decreased the topics generated from LDA model, and new similarity matrix is as following:

	Adult Entertainment	Arabian	Barbeque	Beer	Breweries	British	Cajun-Creole	Caribbean	Cheesesteaks	Cinema	Convenience Stores	Cuban	Ethiopian	Event Planning & Services	Fondue	Food Delivery Services	Ice Cream & Frozen Yogurt	Internet Cafes	Karaoke	Malaysian	Mediterranean	Mongolian	Pakistani	Persian-Iranian	Polish	Seafood	Soul Food	Street Vendors	Tapas Bars	Turk
Adult Entertainment	1	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0.1	0	0	0	0.4	0	0.1	0	0	
Arabian	0	1	0	0	0	0	0	0	0	0	0	0	0.1	0	0.1	0	0	0	0	0	0	0.1	0	0	1	0.4	0	0.1	0	0
Barbeque	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0.3	0	1	0	0
Beer	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.3	0	0	0	0
Breweries	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0.4	1	0	0	0
British	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0.3	0	0	0	0
Cajun-Creole	0	0	0	0	0	0	1	0	0	0	1	0	0.1	0	0.1	0	1	0	1	0	0	0.1	0	0	0	0.4	0	0.1	0	0
Caribbean	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.4	0	0	0	0
Cheesesteaks	0	0	1	0	0	0	0	0	1	0	0	0	0.1	0	0.1	0	0	0	0	0	1	0.1	0	0	0	0.4	0	1	0	0
Cinema	0	0	0	1	0	0	0	1	0	1	0	0	0.1	0	0	0	0	0	0	0	0	0.1	0	0	0	0.4	0	0	0	0
Convenience Stores	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0.1	0	0	0	0.4	0	0	0	0
Cuban	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0.1	1	0	1	0	0	0	0.1	0	0	0	0.4	0	0.1	0	0.1
Ethiopian	0.1	0.1	0	0	0	1	0.1	0	0.1	0.1	0	1	1	0	0.1	1	0	1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.4	0.1	0.1	0.1	0.1
Event Planning & Services	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0.3	0	1	1	0	0.3	0	0	1	0
Fondue	0	0.1	0	0	1	0	0.1	0	0.1	0	0	0.1	0.1	0	1	0	0	0	0	0.1	0.1	1	0	0.1	0.1	0.4	1	0.1	0	0.1
Food Delivery Services	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	1	0	1	0	0	0	0.1	0	0	0	0.4	0	0	0	0
Ice Cream & Frozen Yogurt	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0.4	0	0	0	0
Internet Cafes	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0.3	0	0	0	0
Karaoke	0	0	0	0	0	0	1	0	0	0	1	0	0.1	0	0	0	1	0	1	0	0	0.1	0	0	0	0.4	0	0	0	0
Malaysian	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0.1	0	0	0	0	1	0.1	0.1	0	0.1	0.1	0.4	0	0.1	0	0.1
Mediterranean	0	0	1	0	0	0	0	0	1	0	0	0	0.1	0.3	0.1	0	0	0	0	0	1	0.1	0.3	0.3	0.1	0.4	0	1	0.3	0.1
Mongolian	0.1	0.1	0	0	1	0	0.1	0	0.1	0.1	0.1	0.1	0.1	0	1	0.1	0	0	0.1	0.1	0.1	1	0.1	0.1	0.1	0.4	1	0.1	0.1	0.1
Pakistani	0	0	0	0	0	0	0	0	0	0	0	0	0.1	1	0	0	0	0	0	0	0.3	0.1	1	1	0	0.4	0	0.1	1	0
Persian-Iranian	0	0	0	0	0	0	0	0	0	0	0	0	0.1	1	0.1	0	0	0	0	0.1	0.3	0.1	1	1	0.1	0.4	0	0.1	1	0.1
Polish	0	1	0	0	0	0	0	0	0	0	0	0	0.1	0	0.1	0	0	0	0	0.1	0.1	0.1	0	0.1	1	0.4	0	0.1	0	0.1
Seafood	0.4	0.4	0.3	0.3	0.4	0.3	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.3	0.4	0.4	0.4	0.3	0.4	0.4	0.4	0.4	0.4	0.4	0.4	1	0.4	0.4	0.4	0.4
Soul Food	0	0	0	0	1	0	0	0	0	0	0	0	0.1	0	1	0	0	0	0	0	0	1	0	0	0	0.4	1	0	0	0
Street Vendors	0.1	0.1	1	0	0	0	0.1	0	1	0	0	0.1	0.1	0	0.1	0	0	0	0	0.1	1	0.1	0.1	0.1	0.1	0.4	0	1	0	0.1
Tapas Bars	0	0	0	0	0	0	0	0	0	0	0	0	0.1	1	0	0	0	0	0	0	0.3	0.1	1	1	0	0.4	0	0	1	0
Turkish	0	0	0	0	0	0	0	0	0	0	0	0.1	0.1	0	0.1	0	0	0	0	0.1	0.1	0.1	0	0.1	0.1	0.4	0	0.1	0	1

As seen, there are more similarities discovered in this topic model setting.

But similar topics are not grouped together in this confusion matrix. This is improved in Task 2.3

Task 2.3

I used K-means to cluster different cuisines. Feature of teach cuisine is the normalized sum of weights in different topics.

The first trial is to cluster all cuisines to 3 different groups. And this is the similarity matrix

	Adult Entertainment	Amusement Parks	Arabian	Brazilian	Cuban	Fondue	German	Malaysian	Mongolian	Pakistani	Ramen	Street Vendors	Turkish	Seafood	Breakfast & Brunch	Casinos	Cheesesteaks	Cinema	Comfort Food	Convenience Stores	Event Planning & Services	Food Delivery Services	Ice Cream & Frozen Yogurt	Internet Cafes	Italian	Latin American	Pubs	Tex-Mex	Thai	Tours
Adult Entertainment	1	0	0	0	0	0	0	1	0.1	0	0.1	0.1	0.1	0.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Amusement Parks	0	1	0.1	0	0	0.1	0	0.1	0.1	0	0.1	0.1	0.1	0.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Arabian	0	0.1	1	0.1	0	0.1	1	0.1	0.1	0	0.1	0.1	0.1	0.4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Brazilian	0	0	0.1	1	0	1	0	0	0.1	0	0.1	0.1	0.1	0.4	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
Cuban	0	0	0	0	1	0	0	0	0.1	0	0.1	0.1	1	0.4	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0
Fondue	0	0.1	0.1	1	0	1	0.1	0.1	0.1	0	0.1	0.1	0.1	0.4	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
German	0	0	1	0	0	0.1	1	0	0.1	0	0.1	0.1	0.1	0.4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Malaysian	1	0.1	0.1	0	0	0.1	0	1	0.1	0	0.1	0.1	0.1	0.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Mongolian	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1	0.1	0.1	0.1	0.1	0.4	0	1	0.1	0.1	0.1	0	0	0.1	1	0	0	0	0	0.1	1	1
Pakistani	0	0	0	0	0	0	0	0	0.1	1	0.1	0.1	0.1	0.4	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
Ramen	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1	1	0.1	0.4	0	0.1	0.1	0.1	0.1	0	0	0.1	0.1	0.1	0	0.1	0	0.1	0.1	0.1
Street Vendors	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1	1	0.1	0.4	0	0	0.1	0.1	0.1	0	0	0.1	0	0	0	0	0	0.1	0	0.1
Turkish	0.1	0.1	0.1	0.1	1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1	0.4	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0.1
Seafood	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	1	0.3	0.3	0.4	0.4	0.4	0.3	0.3	0.4	0.4	0.3	0.3	0.4	0.3	0.4	0.3	0.4
Breakfast & Brunch	0	0	0	1	0	1	0	0	0	0	0	0	0	0.3	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
Casinos	0	0	0	0	0	0	0	0	1	0	0.1	0	0	0.3	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	1
Cheesesteaks	0	0	0	0	0	0	0	0	0.1	0	0.1	0.1	0	0.4	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0
Cinema	0	0	1	0	0	0	1	0	0.1	0	0.1	0.1	0	0.4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Comfort Food	0	0	0	0	1	0	0	0	0.1	0	0.1	0.1	1	0.4	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0
Convenience Stores	0	0	0	0	0	0	0	0	0	1	0	0	0	0.3	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
Event Planning & Services	0	0	0	0	1	0	0	0	0	0	0	0	1	0.3	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0
Food Delivery Services	0	0	0	1	0	1	0	0	0.1	0	0.1	0.1	0	0.4	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
Ice Cream & Frozen Yogurt	0	0	0	0	0	0	0	0	1	0	0.1	0	0	0.4	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	1
Internet Cafes	0	0	0	1	0	1	0	0	0	0	0.1	0	0	0.3	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
Italian	0	0	0	0	0	0	0	0	0	0	0	0	0	0.3	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0
Latin American	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0.4	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0
Pubs	0	0	0	0	1	0	0	0	0	0	0	0	1	0.3	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0
Tex-Mex	0	0	0	0	0	0	0	0.1	1	0.1	0.1	0	0.4	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
Thai	0	0	0	0	0	0	0	1	0	0.1	0	0	0.3	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1
Tours	0	0	0	0	0	0	0	1	0	0.1	0.1	0.1	0.4	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	1	

The second trial is to cluster all cuisines to 4 different groups. And this is the similarity matrix

	Pubs	Salvadoran	Vegetarian	Diners	Fruits & Veggies	Internet Cafes	Italian	Shopping	Sushi Bars	Vietnamese	Argentine	Brazilian	Bubble Tea	Malaysian	Persian-Iranian	Shanghainese	Street Vendors	Seafood	Breweries	British	Casinos	Comfort Food	Ice Cream & Frozen Yogurt	Live-Raw Food	Mediterranean	Soul Food	Tapas Bars	Tea Rooms	Thai	Ven
Pubs	1	0.1	0.4	0	0	0.6	0	0	0	0	0	0	0.6	0	0	0	0	0.4	0	0.6	0	0	0	0	0	0	0.6	0	0	0.6
Salvadoran	0.1	1	0.1	0	0.1	0.1	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1	0.1	0.1	0.5	1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1	0.1	0.1
Vegetarian	0.4	0.1	1	0.6	0.3	0.7	0	0.4	0	0.4	0.4	0	0.7	0	0	0.7	0.4	0.5	0	0.7	0	0	0	0.4	0.4	0	0.7	0	0	0.7
Diners	0	0	0.6	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0.3	0	0	0	0	0	0	0	0	0	0	0	0
Fruits & Veggies	0	0.1	0.3	0	1	0	0	1	0	1	1	0	0	0	0	0	1	0.3	0	0	0	0	0	1	1	0	0	0	0	0
Internet Cafes	0.6	0.1	0.7	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0.3	0	1	0	0	0	0	0	0	1	0	0	1
Italian	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0.3	0	0	0	0	0	0	0	0	0	0	0	0
Shopping	0	0.1	0.4	0	1	0	0	1	0	1	1	0	0	0	0	0	1	0.4	0	0	0	0	0	1	1	0	0	0	0	0
Sushi Bars	0	0.1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0.4	0	0	0	1	0	0	0	0	0	0	0	0
Vietnamese	0	0.1	0.4	0	1	0	0	1	0	1	1	0	0	0	0	0	1	0.4	0	0	0	0	0	1	1	0	0	0	0	0
Argentine	0	0.1	0.4	0	1	0	0	1	0	1	1	0.1	0.1	0.1	0.1	0.1	1	0.4	0	0	0	0	0	1	1	0	0	0	0	0
Brazilian	0	0.1	0	0	0	0	0	0	1	0	0.1	1	0.1	0.1	0.1	0.1	0.1	0.4	0	0	0	1	0	0	0	0	0	0	0	0
Bubble Tea	0.6	0.1	0.7	0	0	1	0	0	0	0	0.1	0.1	1	0.1	0.1	0.1	0.1	0.4	0	1	0	0.1	0	0.1	0.1	0.1	1	0	0	1
Malaysian	0	0.1	0	0	0	0	1	0	0	0	0.1	0.1	0.1	1	0.1	0.1	0.1	0.4	0	0	0	0	0	0	0	0	0	0	0	0
Persian-Iranian	0	1	0	0	0	0	0	0	0	0	0.1	0.1	0.1	0.1	1	0.1	0.1	0.4	1	0	0	0	0	0	0	0	0	1	0	0
Shanghainese	0	0.1	0.7	1	0	0	0	0	0	0	0.1	0.1	0.1	0.1	0.1	1	0.1	0.4	0	0.1	0.1	0.1	0	0.1	0.1	0.1	0.1	0	0	0
Street Vendors	0	0.1	0.4	0	1	0	0	1	0	1	1	0.1	0.1	0.1	0.1	0.1	1	0.4	0	0	0	0.1	0	1	1	0.1	0.1	0	0	0
Seafood	0.4	0.5	0.5	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	1	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
Breweries	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0.4	1	0	0	0	0	0	0	0	0	1	0	0
British	0.6	0.1	0.7	0	0	1	0	0	0	0	0	0	1	0	0	0.1	0	0.4	0	1	0	0	0	0	0	0	1	0	0	1
Casinos	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0.4	0	0	1	0	0	0	0	0	0	0	0	0
Comfort Food	0	0.1	0	0	0	0	0	0	1	0	0	1	0.1	0	0	0.1	0.1	0.4	0	0	0	1	0	0	0	0	0	0	0	0
Ice Cream & Frozen Yogurt	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.4	0	0	0	0	1	0	0	0	0	0	0	0
Live-Raw Food	0	0.1	0.4	0	1	0	0	1	0	1	1	0	0.1	0	0	0.1	1	0.4	0	0	0	0	0	1	1	0	0	0	0	0
Mediterranean	0	0.1	0.4	0	1	0	0	1	0	1	1	0	0.1	0	0	0.1	1	0.4	0	0	0	0	0	1	1	0	0	0	0	0
Soul Food	0	0.1	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0.1	0.1	0.4	0	0	0	0	0	0	0	1	0	0	1	0
Tapas Bars	0.6	0.1	0.7	0	0	1	0	0	0	0	0	0	1	0	0	0.1	0.1	0.4	0	1	0	0	0	0	0	0	1	0	0	1
Tea Rooms	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0.4	1	0	0	0	0	0	0	0	0	1	0	0
Thai	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.4	0	0	0	0	0	0	0	1	0	0	1	0
Venues & Event Spaces	0.6	0.1	0.7	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0.4	0	1	0	0	0	0	0	0	1	0	0	1

Both clustering results improved the quality of this confusion matrix as similar cuisines are grouped together.