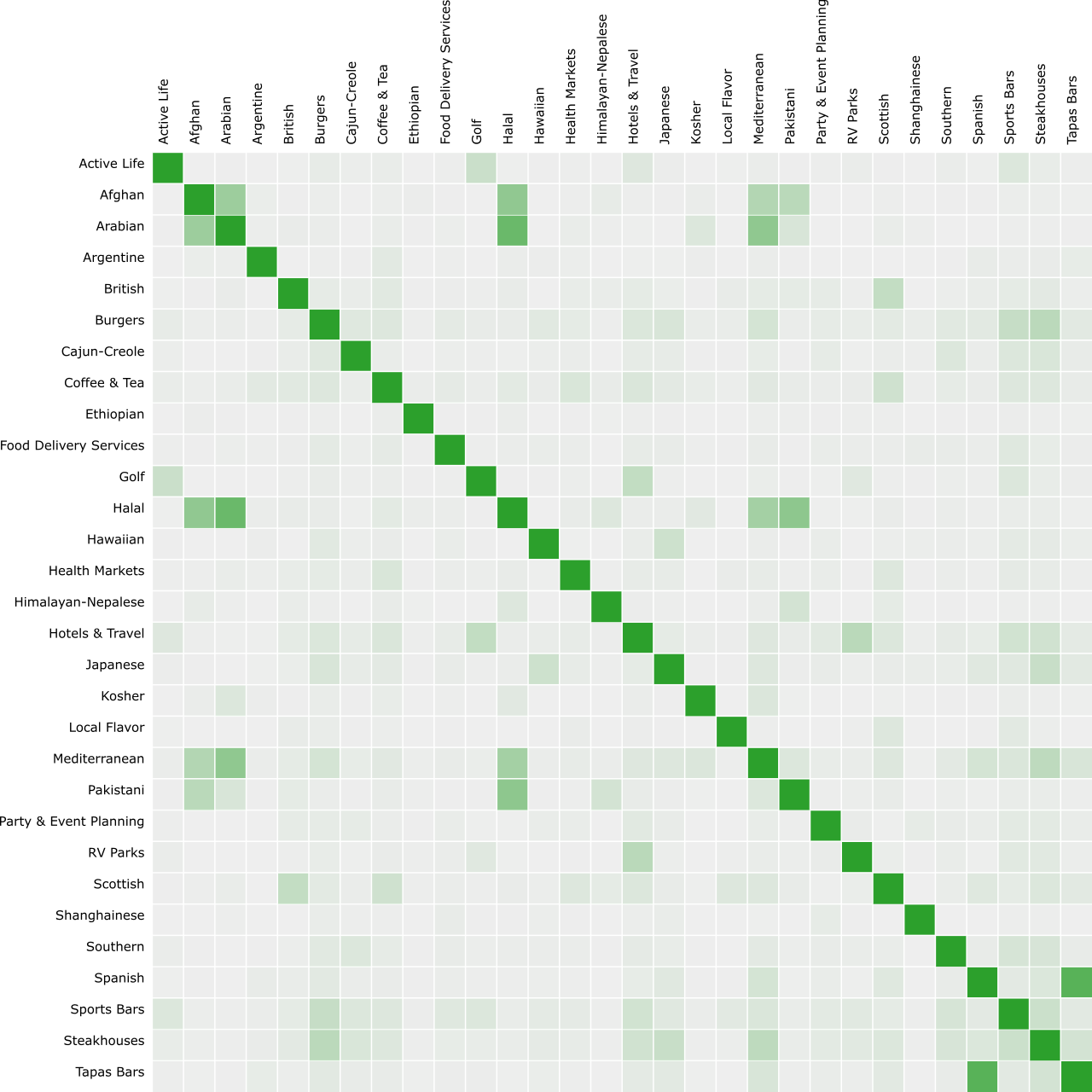
Task 2: Cuisine Clustering and Map Construction

# Task 2.1: Visualization of the Cuisine Map



## Dataset

30 cuisine reviews randomly selected from Task 2 Cuisines Sample Dataset provided.

## How

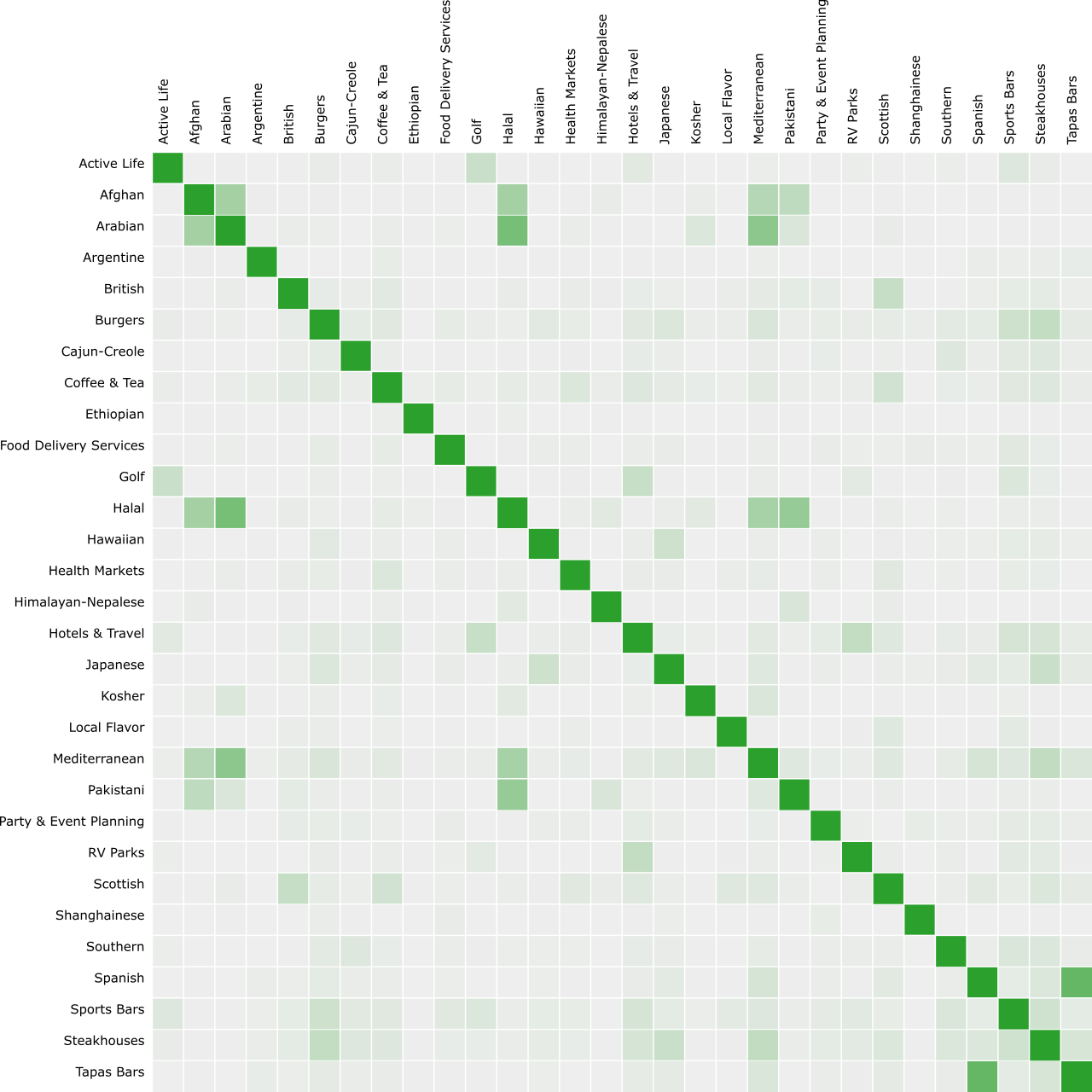
This graph is based on Word Frequency similarity only. No IFD and LDA is applied.

This is done by modifying the Python code provided:

1. Change the default TfidfVectorizer to CountVectorizer
2. Remove LDA model and use vector to store similarity

# Task 2.2: Improving the Cuisine Map

## IDF Applied



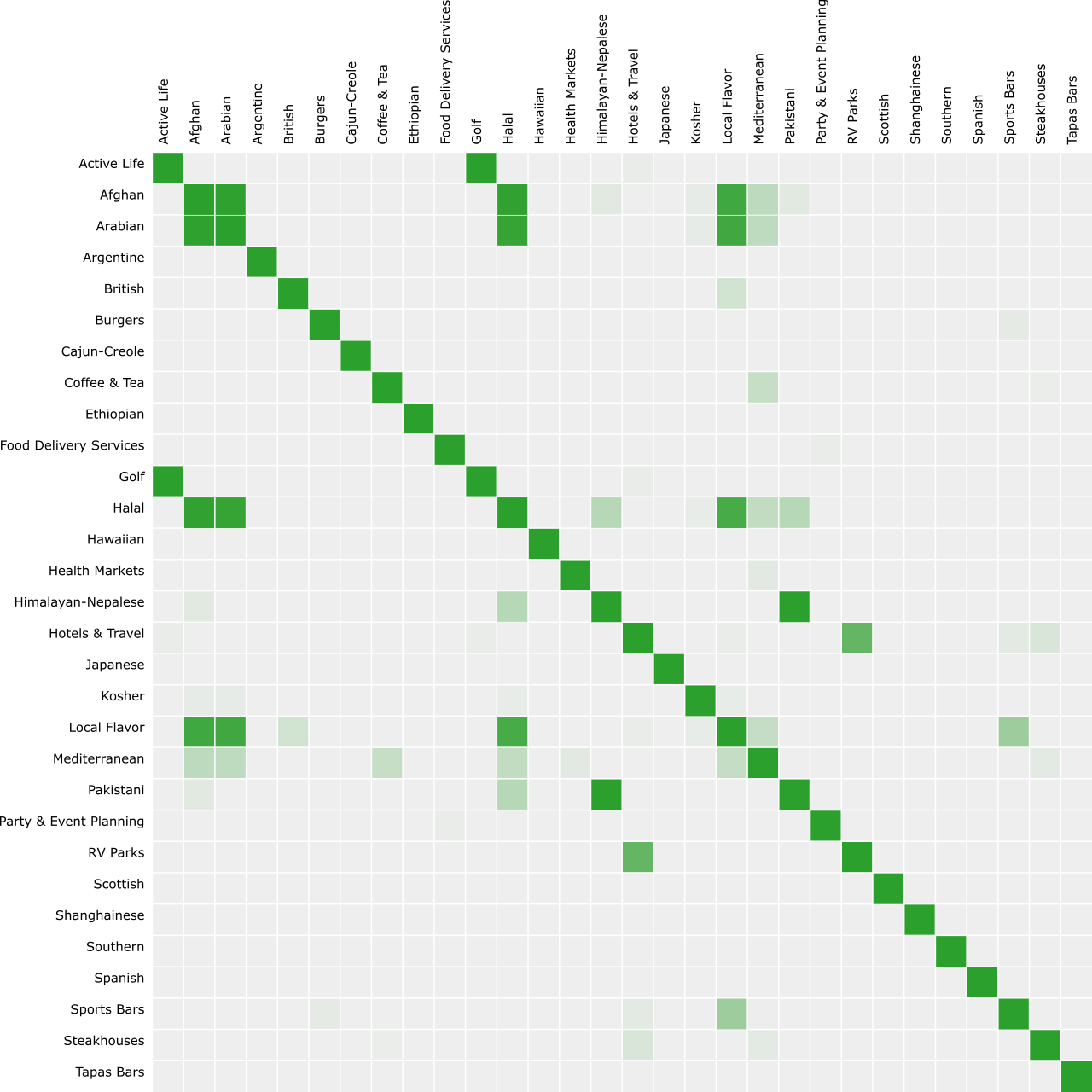
### Dataset

30 cuisine reviews randomly selected from Task 2 Cuisines Sample Dataset provided.

### How

This graph is based on cuisine similarity where TF-IDF is applied.

## LDA Applied



### Dataset

30 cuisine reviews randomly selected from Task 2 Cuisines Sample Dataset provided.

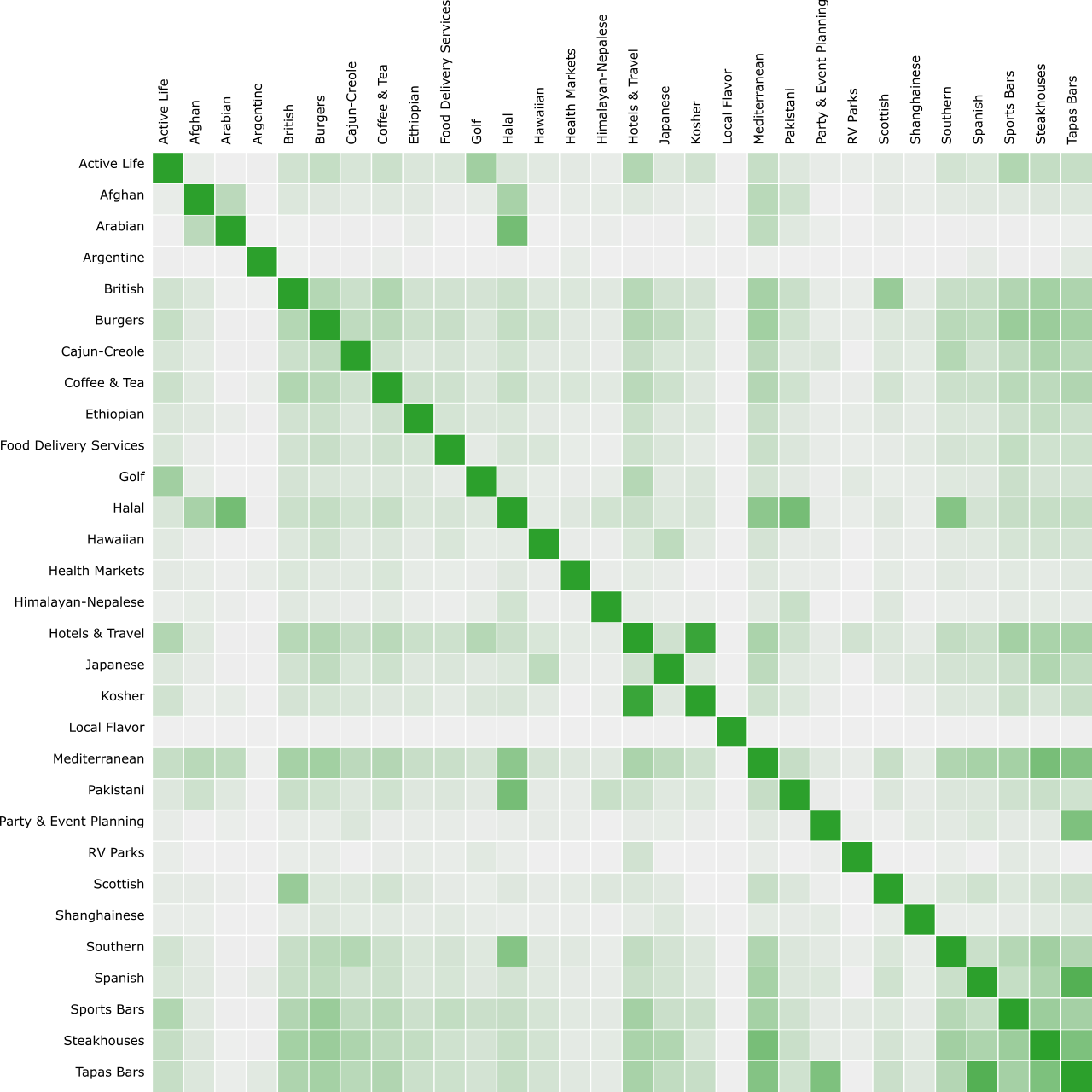
### How

This graph is based on Word Frequency similarity and LDA is applied.

This is done by modifying the Python code provided:

1. CountVectorizer is used
2. LDA model is used

## IDF + LDA Applied



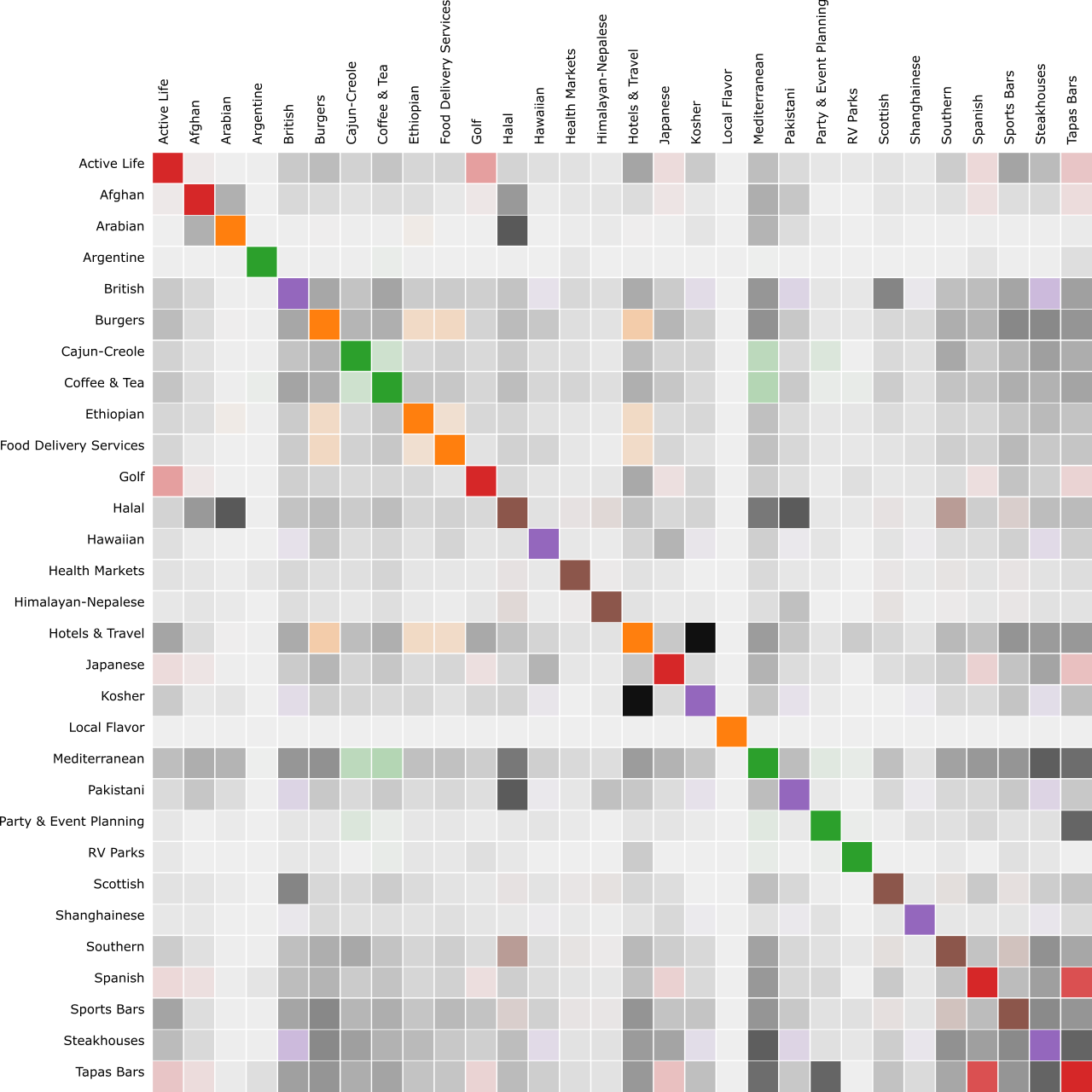
### Dataset

30 cuisine reviews randomly selected from Task 2 Cuisines Sample Dataset provided.

### How

This graph is based on cuisine similarity when TF-IDF and LDA are applied. This is part of function that the Python tool provided.

# Task 2.3: Incorporating Clustering in Cuisine Map



### Dataset

30 cuisine reviews randomly selected from Task 2 Cuisines Sample Dataset provided.

### How

K-Means is applied to the IDF+LDA similarity dataset. K is set to 5.