

# Olympic host cities

This is capstone project for Coursera course “Applied Data Science Capstone”

Date: 11/2018

By: Charles Fung

## Introduction

As the International Olympic Committee (IOC) stated about the host city selection process:

Hosting the Olympic Games offers manifold benefits and opportunities to a Candidate City and the host region and country. Many years of careful and precise planning are required to host successful Olympic Games, with all of the relevant organisations, authorities and stakeholders working together as one united team, to ensure that the Games leave a positive, longterm and sustainable legacy.

(<https://www.olympic.org/all-about-the-candidature-process>)

There are a lot of fiscal and financial commitments and it is a multi-year effort by the local government to host the Olympic games. There have been about 20 cities that have host the games and, unfortunately, not all of them were successful in terms of economic or popularity gains. Therefore, it is highly critical for a city to evaluate and decide if it is willing to get involved.

Hosting an Olympic Game involves many areas: sport facilities, transportation, communication, accommodation. To be successful and memorable event for the visitors, it also require support facilities like restaurants and night life etc.

In this report, we will attempt to look at the popular venues around different cities, and try to see if the number and type of venues are relevant to the determination of a city being selected to host an Olympic Game.

Target audiences for this report will include:

- 1) IOC, NOC(National Olympic Committees) involved in the election process
- 2) City governments to evaluate chance of being selected. They may also find out what venues they are lacking and make changes to improve their chance in the future.

Since the Winter Games require cooler weather conditions, the host cities are limited to those at higher latitudes. This report will limit itself to Summer Games.

# Data

## Past host cities

There is a data set from Kaggle on past Olympic athlete events, which lists for each medal the athlete, year, city, etc.

("120 years of Olympic history: athletes and results" [https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results#athlete\\_events.csv](https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results#athlete_events.csv))

We will use this data set to extract the name of the host cities.

## Best cities in the world

There are many lists on the Internet on the best cities in the world. We have picked "World's best cities" lists 100 cities (<https://www.bestcities.org/rankings/worlds-best-cities/>).

We will do a web scraping to extract the name of these cities, and filter those that have not host any Olympic Games.

## City coordinates

The cities from both list will form the base for comparison.

We will use geopy API to retrieve the coordinates for the cities.

## Venues in the cities

We will use FourSquare Explore API to retrieve 100 top-pick venues around the cities, and use the frequencies of venue categories to classify the cities.

The venue categories from FourSquare API are very fine-grained (such as Ethiopian Restaurant, Dim Sum Restaurant or Filipino Restaurant under Food) and they are too specific to be useful.

We will translate the venue categories (see below) to their top-level categories for easy comparison.

## Venue categories

FourSquare Categories API which group all the categorie under top-level categories like Arts & Entertainment, College & University, Event, etc.

We will use this API to retrieve all categories, and build a mapping dictionary for this purpose.