# Determining Which Hotel in Times Square is Best for Tourists

Bradley Betts

June 26, 2020

## 1. Introduction

### 1.1 Background

New York City is a very popular location for both tourists visiting the United States, and American citizens that wish to explore another state. In fact, New York City boasts about one third of the tourists in North America. A vast majority of these visitors will explore and/or reside in the Times Square area. It's likely that when someone is going to another location that isn't their home, they choose to stay in a hotel. In doing so, they might also use programs like Trivago, to inform them of the prices of staying at these different hotels. Certain services like this will also inform the user how far the hotels are from the closest major city. On top of that, people can simply google sites of interest in their designated city of choice. However, most people choose which hotel to stay at not only because of their cost or only because of what's nearby; it is usually a mixture of both variables.

### 1.2 Problem

This project seeks to determine which hotel within Times Square is the best choice based on cost per night, and number of nearby venues. The chosen hotel will alter based on a limitation to how much the user would spend and what venues are within a specific distance.

### 1.3 Interest

Tourists venturing to New York City would for sure be interested in this project because it will make travelling to this bumbling city much easier to plan. Hotel executives would also find this information useful, as it would show them the cost of competition nearby, with the addition of which venues they could possibly market to their customers to increase their likelihood of staying in their hotel.

## 2. Data Acquisition and Cleaning

### 2.1 Data sources

The location of all the venues in New York City can be found using the Foursquare API, and the location of specific hotels will be gathered from an open-source CSV file on Kaggle shown here.

### 2.2 Data cleaning

The main problem that came with gathering the data was finding the cost of each hotel, as there was not a dataset of each of these hotel costs to be found. This is partially since many hotels have

different kinds of rooms, each with a different cost. However, simply researching each hotel and their respective cost solves this issue, as there are only 42 hotels in the Times Square hotel database. The price of each hotel was determined by finding how many kinds of rooms there are and calculating the mean cost. The results were concatenated to the CSV file.

### 2.3 Feature selection

There were some features in the CSV file that proved to be unnecessary, such as address, phone number, website and borough. Therefore, the only features that remained were hotel name, latitude, longitude and the price that was gathered from the data scraping.