

# Coursera: Applied Data Science Capstone

## Capstone Project - The Battle of Neighborhoods (Week 1)

# **1. Overview of Data and Data Sources**

## **1.1 Introduction**

The intention of this project is to predict store success based on various information on the surrounding area.

Step 1: Identify the influence of various amenities in the surrounding area for existing stores; data on store success, geospatial data and amenities in surrounding area is needed

Step 2: Predict store success for new locations; geospatial data (location of new stores) and amenities in surrounding area is needed

While existing and new store addresses and data regarding success of existing stores are only available in company internal databases, is geospatial data and data referring venues in the surrounding area available from cloud resources.

## **1.2 Business Data**

**Extracts from the company database will be used:**

- Addresses existing stores
- Addresses planned store locations
- Revenue data for existing stores, Year 2018 (data standardized) by category of products

Data will be provided as “.csv” files

## **1.3 Google Maps Geocoding API**

To convert addresses to geospatial information, Google Maps Geocoding API will be used.

Data will be accessed by Python using a geocoder library.

## **1.4 Foursquare location data**

To determine the proximity of various amenities (e.g. riding stables, riding school, riding club, competition stores), Foursquare location data is used. Data will be accessed by Foursquare API calls.