Report Contents

[Report Contents 1](#__RefHeading___Toc58_777593657)

[1 Introduction 2](#__RefHeading___Toc592_1769055570)

[1.2 Background 2](#__RefHeading___Toc600_1769055570)

[1.3 Problem 2](#__RefHeading___Toc372_777593657)

[1.4 Stakeholders 2](#__RefHeading___Toc374_777593657)

[2 Data Sourcing, Understanding and Wrangling 2](#__RefHeading___Toc594_1769055570)

[2.2 Data API 2](#__RefHeading___Toc374_7775936571)

[2.3 Problem & Data Understanding 2](#__RefHeading___Toc374_77759365711)

[2.4 Exploratory Data Analysis 3](#__RefHeading___Toc374_77759365712)

1. Introduction

In this capstone project we will explore the idea of where to build new sports facilities.

## **Background**

To set a background for the discussion, the local government wants to build more sports facilities in the selected neighbourhood. The sports facilities are of 2 main categories – 1. Athletics and Sports 2. Gymnasium / Fitness Centre. They want to ensure that they achieve a balance between the types of facilities they build. So they want to ensure they increase the number of gyms in areas where there are fewer gyms etc.

This is our problem domain.

## **Problem**

The problem we will solve is deciding where we should build sports facility in a neighbourhood and of what type it should be. This approach is similar to anomaly detection where we find outliers i.e. areas having less sports facilities.

## **Stakeholders**

The stakeholders of this project would be local sports departments OR people interested in opening some type of sports facility business in the neighbourhood. They want to decide what type of facility and also where it is best to build it. In order to benefit the new business they will want to select areas with no facility or over-crowded facilities.

1. Data Sourcing, Understanding and Wrangling

The primary data source for this project is the location data provided by the provider FourSquare. We will access the FourSquare REST APIs to get venue and other details. Following sections look at the APIs we will use and understand the data they return.

## **Data API**

The important FourSquare API for this project are listed below:

* Search For Venues: GET: XXX/v2/venues/search
* Get Venue Categories: GET: XXX/v2/venues/categories
* Get Details of a Venue: GET: XXX/v2/venues/VENUE\_ID

## **Problem & Data Understanding**

The problem requires that we apply the model that we come up with on a certain area or neighbourhood and we should be able to run it again on other neighbourhoods. So it should except new parameters in terms of area/neighbourhood and also venue categories we want to differentiate between.

We will work with Venue Dataretrieved from FourSquare. Each Venue belongs to a list of categories wherein there is a parent category. We will be interested in venues of following parent categories for our fitness project :

* **Athletics & Sports (4f4528bc4b90abdf24c9de85) (exclusively)**
* **Gym / Fitness Center (4bf58dd8d48988d175941735)**

We will assume the user has a location in mind around which they want to check availability of fitness facilities. Hence we will assume a target geographical location given by user as input data also.

## **Exploratory Data Analysis**

<Week 2 onwards ......>

# 