

The Battle of Neighborhoods: Exploring Soccer Field Business Opportunity

Coursera Applied Data Science Capstone Project

Week 4 Submission – Introduction and Data

I-Hsin Lin

1 July 2021

Contents

1. Introduction	1
1.1. Background Information	1
1.2. Problem Statement	1
2. Data	2

1. Introduction

1.1. Background Information

The Government of Indonesia has announced that the nation's capital will be moved to the East Kalimantan region. The location of the new capital city is located in part of North Penajam Paser Regency and partly in Kutai Kertanegara Regency, East Kalimantan. The government considered the location due to several factors such as infrastructure, supporting facilities, and being separated from the business center. Some time ago, Kompas.com summarized that in East Kalimantan there are three international standard football stadiums. Thus, it can also be said that the East Kalimantan region also has an attraction in the field of sports¹.

1.2. Problem Statement

With the aforementioned prospect, the neighborhoods around the soccer fields may come as the priority for developing. Various stakeholders may be interested to explore business opportunities in East Kalimantan.

This data science project is thus carried out to help them answer the following question: Which of the East Kalimantan regions (around soccer fields) are strategic for investment?

The project will be of interest to fellow entrepreneurs.

1. <https://bola.kompas.com/read/2019/08/26/19490918/ada-3-stadion-bertaraf-internasional-di-kawasan-kalimantan-timur?page=all>

2. Data

In order to explore potential answer to the problems, the following data are required:

1. **The names of administrative regions in East Kalimantan and their corresponding postal codes.** The regions include three levels of subdivision: city, district, subdistrict. The information was scraped from a directory on indonesiapostcode.com. The region names are useful to perform analysis across different sub-regions. The postal codes are needed to obtain coordinates of each subdistricts.
2. **Geographical coordinates** of East Kalimantan and its subdistricts, which will in turn be needed to utilize Foursquare API in the subsequent step. Coordinates are obtained using GeoPy libraries, with [Nominatim](#) API.
3. **Information about venues in East Kalimantan regions:** the names, category, venue latitudes, venue longitudes. These are obtained using [Foursquare](#) API. The subdistricts of East Kalimantan will be clustered based on their surrounding venues to find the best location candidates.