

IBM Data Science Capstone Background and Data Source

By: David Tossolini

The Problem

Los Angeles is the cultural, financial, and commercial center of Southern California. It is the most populous city in California with 88 cities in Los Angeles County, California. Each city having a mayor and a city council. It's total area is 502.76 sq mi (1,302.15 km²) while its total Land area is 468.74 sq mi (1,214.03 km²). Its 2018 total population estimate was 3,990,456.

Billions in venture capital funding have poured into Los Angeles at an unprecedented rate since 2013. Talent from Silicon Valley and around the world has begun to migrate to the Los Angeles area. Unlike San Francisco, Los Angeles spans a significantly larger land area with less geographic urban centralization.

The sprawling metropolitan area presents both challenges and opportunities. The newly minted "Silicon Beach" is a sub section of this large landmass consisting of 11 "beach cities". Analyzing and understanding the large geospatial area presents a unique challenge with the opportunity for equally unique rewards.

Investors looking to profit from this change will need to answer the question of "where", a relatively difficult one given the vastness of LA. As a starting point understanding the current geographic landscape, commercial venues and demographics will be key in developing a strategy for commercial investments. By anticipating the likely change and identifying unfilled needs, savvy investors may capture significant returns through well identified bets.

The Data

To paint a picture of the Beach cities three data sources were used:

1. Population data from the Chief Executive Office of Los Angeles County
2. Resident income and age data from <https://datausa.io>
3. The Foursquare API for locating venues and their type in each city