

A description of the problem and a discussion of the background

Problem: Identifying the possibility to establish a financial advisory institution in San Jose, by analysing the economic potential of the population and exploring the venues in each neighbourhood.

Background: San Jose is one of the famous cities of California, USA. It is part of the Silicon Valley which hosts many information technology companies. These companies employ many IT professionals who are quite busy with the technological innovation and development of new software products. These professionals are fully occupied with their daily hectic routine which consumes a significant portion of their time. Consequently, it leaves them with less time to fully concentrate on their financial matters especially study different opportunities and make suitable investment related decisions. Therefore, an institution which could manage their finances and provide investment advice would be of great support to them. From this perspective, San Jose offers a huge potential for establishing a financial institution to manage the finances and advise on investments for the people working in this city.

Analysis of economic indicators and venues of San Jose: Before such an institution can be set up, it is important to study the different parameters across different locations in San Jose, which could indicate the investment potential. Accordingly, I have taken up the analysis of socio economic indicators of people residing in different neighbourhoods of San Jose and the venues in each neighbourhood of San Jose. The analysis will be very helpful for institutions to potentially setup a shop and provide customized financial investment services for individuals working in San Jose along with their families.

A description of the data and how it will be used to solve the problem.

The data is retrieved from the Spatial Data Repository of NYU [2] (U.S. Neighbourhoods greenness measures and social variables). The data includes many attributes and the following attributes would be mainly used for analysis:

- Population density
- Average high income
- Percentage owning houses
- Percentage renting
- Median age
- Median high income
- City parks.

In addition, the Neighbourhood name and the coordinates (latitude and longitude) are the attributes.

The .json file from the website contains data related to many other cities in the USA, in addition to San Jose. Therefore, data needs to be pre-processed and filtered to retain only the San Jose relevant indicators.

In order to obtain the data related to venues, I would be using Four Square to explore and analyse each Neighbourhood of San Jose to identify the venues in the neighbourhood. Both the income related data and venue related data would be merged to provide a comprehensive view of San Jose and the neighbourhoods. I would be using the K-means algorithm to cluster the neighbourhoods.

On the one hand the income levels provide an indicator for economic potential of the population to setup an advisory financial institution. On the other hand, the data related to venues in different neighbourhoods of San Jose provides an indicator of economic activity within the neighbourhoods. These two when merged together would provide the basis to potential investors and enable them to decide and setup an investment advisory institution.