**Schema of Project Community Recommendation**

The overall purpose of this project is to design a website/APP that could help people move from other cities to choose the optimal community that he could live in when he gives some demographical information to the system. Also, this product could be helpful to urban planner to develop the environment of city to make people live better.

**The information we need as input:**

1. Disposable income

This part gives the priority of entertainment equipment.

We need to know the disposable income the user has and to determine which range of rental amount would be suitable.

1. Buy or rent an apartment

This part decides the choice of rental or sales price.

Buying an apartment and renting one are different for prices.

1. Age

The density of hospital would be more important for people older than 45 years old than younger people.

1. Tendency of social networking or leisure activities and frequency

This part decides the priority of leisure activities.

For people who is outgoing the choice of community would contain the consideration of entertainment equipment.

1. Education level
2. Have kids or not, and the number of kids

For people have kids the priority of education resources would be higher. And for different levels of education we would recommend respectively.

1. The education situation of kids (if they have kids)
2. Shop online or offline (for grocery)

The density of grocery would be less important if people shop them online via Amazon Prime or other applications.

1. Have cars or not and the purpose of having a car

If people have cars, we would provide the information of parking fees and the average price of gasolines.

1. Work at home or need to commute?

If people work at home, then the traffic situation would be less important for them when at weekdays.

**Data in system**

We divide Washington.D.C into 22 communities by zip code. For each community we need to gather the following data:

1. The average price of sale or rental

We need to have access to these data and to update them dynamically. The prices are different for the number of beds the apartment has.

1. The safety indexes

We need the frequency of different types of crimes and the density of police offices.

The types of crime and the frequency of happening would be a factor to consider whether the user have kids or not. But for people have kids this priority would be higher.

1. The density of public traffic

The density of public transportation like subway and bus would be important for people commute and don’t have car.

1. The density of different levels of grocery

We would rank different grocery first, then use the density as a factor to rank the community.

1. The education resources (for users who have kids)

We need the information of elementary school, middle school and high school to rank the community. Our data are all public schools, but we could provide the information of private schools via links from third party. This part could include the density of library.

1. The average price of gas (for car owners)

For car owners, the average price of gas impact their daily life so we need to get the price information.

1. The parking fees (for car owners)

Get the average parking fee for the community.

1. The appreciation of home (for buyers)

For home buyers, the appreciation of the building for past 3 to 5 years would make sense for them to consider the value of investment.

1. The delivery situation of online shopping

The delivery of online shopping would influence the life quality if people shop online often and it could be more important than live nearby a grocery.

1. The density of hospital (for user older than 45 or have special situation)

For people who has the need to see doctors often, the quality and density of hospital in the community would be important. And the average waiting time of hospital would be another consideration to rank.

1. Traffic situation

We need to analyze the traffic situation of the community such as the average traffic jam time.

1. Leisure

Parks, fitness center, concerts, cinema, sporting events. It depends on how often people go to these places and their disposable income.

For all these features we need to get a ranking for them compared to the average, the maximum and the minimum. The grade will be ranked as 0 to 5 stars and we would present a heatmap of the whole district.

The methodology of ranking is the following: calculate the percentile for each features, if the percentile is higher than 80%, give it 5, 60% to 80%, give it 4, 40%-60%, give it 3, 20% to 40%, give it 2, and less than 20% is 1.

We need to grasp data online dynamically to update the rankings as frequent as possible.

**The outcome**

Overall, we want to select 3 to 5 communities for the user. The first choice is the place that most meet the demand of the user. For people have kids, the education resource is the highest priority. For people has health issue, the hospital resource is the highest priority, etc. There would be an algorithm to choose the priority as we recommend.