Comparison of two cities: Houston and Chicago

Yuan Zhang

1. **Introduction**
   1. **Background**  
      My husband and I lived in Houston for almost 7 years. My husband has recently got a job offer in Chicago, and we plan to move to Chicago by the end of September. We are familiar with the climate, food, museums, sports and all the fun in Houston, which can be regarded as our second hometown in the world. On the other hand, we have never been to Chicago and the only impression for Chicago is global, multicultural, and cosmopolitan city.
   2. **Problem**  
      My goal is to perform a comparison of the two cities to see how similar or dissimilar they are. Such techniques allow users to identify similar neighborhoods among cities based on amenities or services being offered locally, and thus can help in understanding the local area activities.
   3. **Interest**  
      There will be other job seekers like us would be interested in the difference between two cities, this will help them to make a better decision when they move to a new city. There will be some company would like to expand their business, they would benefit from such analysis. Any people who live in these two cities can also refresh their ideas by checking one city analysis. Such analysis tool is very extensible to other cities comparison, which can give us a basic idea when we want to compare two cities in the world.
2. **Data Preparation**
   1. **City neighborhoods data from Wikipedia**  
      City data was extracted from the respective Wikipedia pages using Request and Beautifulsoup libraries in Python. There is only neighborhoods data in Wikipedia page, we need to first extract the neighborhoods data and get the coordinate from geopy.
   2. **Weather and climate data from** [**https://www.timeanddate.com/weather/usa/**](https://www.timeanddate.com/weather/usa/)  
      We can search for different city crime rate and types from this website, and use Beautifulsoup to scrape the data. The research here is limited to the most recent whole year.
   3. **City crime data is from** [**https://www.neighborhoodscout.com/il/**](https://www.neighborhoodscout.com/il/)  
      We can search for different city crime rate and types from this website, and use Beautifulsoup to scrape the data. The research here is limited to the most recent whole year.
   4. **Job data is form** [**https://www.indeed.com/**](https://www.indeed.com/)  
      I tried to scrape the job listing from indeed. I used “data scientist” as a target position. I am sure most of the young people that plan to move these two cities will be interested in this kind of information.
   5. **Foursquare location data**  
      [Foursquare](https://foursquare.com/) is a social location service that allows users to explore the world around them. We took advantage of Foursquare data to extract venue based information for all of the neighborhoods under within two cities. The call to the API returns a JSON file and we need to turn that into pandas data-frame.