Lifehack: Find the best city for you!

Group 14
Paul Girdler, Hyungkyu Lim, Venkata Sai Abhishekh Sarvepalli

Abstract

We surveyed students to define the problem statement and find the specific factors that were most important to them in finding a city to relocate to upon graduation. The students wanted to be empowered to make the right decision and the majority of students were concerned about trade-offs between opportunity and lifestyle factors. We scraped and aggregated various free datasets to gather the required information, and analyzed it in a number of ways. Our most interesting findings were:

- Big metro areas are generally high in opportunities, fun and vibrant, and surprisingly low in crime, but costs are high.
- There are trade-offs between OPPORTUNITY & LIFESTYLE. But there are hidden gems.

Data

We gathered the data from a variety of places. The majority of the data was scraped from the web using python. Although this was tedious all the complete datasets required payment to access.

Table – Datasets

Dataset(s)	Description
https://www.kaggle.com/sl6149/data-scientist-job-market-in-the-us https://www.nerdwallet.com/blog/studies/best-places-stem-grads-2016/	Jobs and opportunities by US cities
www.zillow.com/data	Real-estate and economic data about US zip codes
http://www.statsamerica.org/USCP/about.aspx https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population https://gist.github.com/Miserlou/c5cd8364bf9b2420bb29 https://en.wikipedia.org/wiki/Cities_and_metropolitan_areas_of_the_United_States	General statistics of US cities • Population • Location • City/Metro relationship
https://www.data.gov/	Various government datasets
https://www.bls.gov/data/	Various governments datasets related to labor market activity, working conditions, and price changes in the economy
https://en.wikipedia.org/wiki/List_of_cities_by_sunshine_duration https://en.wikipedia.org/wiki/List_of_cities_by_temperature https://www.nerdwallet.com/blog/sunniest-cities/ https://www.currentresults.com/Weather/US/annual-snowfall-by-city.php https://wallethub.com/edu/cities-with-the-best-worst-weather/5043/ https://www.usclimatedata.com/ https://www.ggweather.com/normals/ https://www.currentresults.com/Weather/US/average-city-weather.php https://www.nerdwallet.com/blog/sunniest-cities/	Each US cities climate normal • Ave min max temp • Ave annual snow days • Ave annual sunshine days • Ave monthly rainfall • Ave rain days

https://simplemaps.com/data/us-cities https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population	US cities population data based on census
https://www.nerdwallet.com/blog/credit-cards/best-cities-for-quality-of-life/ https://wallethub.com/edu/best-worst-cities-for-recreation/5144/	Quality of Life for US cities
https://www.numbeo.com/cost-of-living/region_rankings.jsp?title=2018&region=019 https://www.numbeo.com/cost-of-living/cpi_explained.jsp https://www.nerdwallet.com/blog/mortgages/home-search/most-affordable-cities-united-states/	Cost of Living Index for US cities
https://en.wikipedia.org/wiki/List_of_United_States_cities_by_crime_rate	Crime Statistics for US cities.

Analysis and Results

We wanted to empower students with the data to help them narrow down which US city will be most suitable to live in based on their individual requirements. Based on our survey of the students in our data-science class, we tried to identified a problem statement and scope.

The six broad factors we would examine:

1. Weather
2. Job opportunities in area
3. Housing
4. Low crime rate
5. Vibrant Night Life & Culture
6. Affordability of the city.

Please note our final analysis was condensed for impact, and some plots and observations were removed.

Firstly, we analyzed in terms of job opportunities. Here are our main observations:
1. Big metro areas are generally full of opportunities.
2. There are some other metro areas that are hidden gems based on out plot.
3. Searchable map is the best way to visualize opportunity clustering around the country.

Secondly, we analyzed weather. Here are our observations

- Weather radials are an engaging way to show seasonal temperature variation
- Unfortunately, Buffalo scored consistently poor in terms of weather.

Thirdly, we analyzed the trade-off between job opportunities vs life style factors.
We mostly used a quadrant plot as this is very effective at highlighting trade-offs (it is also favoured by the biggest consulting firms such as PwC). All our plots are annotated and labeled to give an intuitive understanding of the trade-offs.

Next, we analyzed cities in terms of crime. Here are our observations:

- Violent crime varies by almost 10 X from highest to lowest.
- The plot generated shows a power distribution.
- Surprisingly, the violent crime is not a big city problem.

Lastly, we analyzed in terms of

- STEM job opportunity vs cost living trade off
- STEM job opportunities vs recreation trade off.

Our main observations here are that there are some hidden gems that have high cost and living and low cost, like Austin, Texas. Otherwise, as you would expect cities that are vibrant and full of opportunity are expensive.

Conclusion

Our most interesting findings were:

- Big metro areas are generally high in opportunities, fun and vibrant, and surprisingly low in crime, but costs are high.

There are trade-offs between OPPORTUNITY & LIFESTYLE. But there are hidden gems.

Why we deserve bonus marks!

1. **We did a real problem and had to work to a set scope. As such we could not change our problem and scope to suit what we felt like doing.**
2. **We had to deal with and overcome real life issues, such as data integrity, data availability, with solutions such as web-scraping, data-cleaning and data-wrangling.**
   For example
   Standardizing Data:
   New York, New York
   New York-Newark-Jersey City, NY-NJ-PA MSA
   New York, NY MA
   New York Metro
   New York, NY
   Cleaning Data:
   '23^{o}C': remove non-numeric convert to number.
   '1,000': remove non-numeric convert to number.
   Geolocation Data convert to useable format.