HubLyfe PM6

Value Proposition: HubLyfe is a one-stop database to easily find all the transportation, education, salary, cost of living (rent), and lifestyle data about Boston neighborhoods for persons new to the city (a very confusing city) who are unhappy with the generic and vague Zillow and Craigslist searches which do not give you all of the necessary information to truly maximize what you get based on your means of living.

Did you fulfill your value proposition?
We fulfilled everything in our value proposition except for the transportation information. We did not implement the use of Google Maps API like we intended but still have it as a future goal. We also were not able to integrate a lot of lifestyle features into our database, but we discuss possible options to expand on this topic in future goals. Overall, we still created a one stop database with most of the information a user moving to Boston could want when moving there, or just moving to a new neighborhood in the city.

List what you initially planned to deliver (as stated in Milestone 1), and compare to what you actually delivered.

Planned:

- We plan to have a database where the user can input address/neighborhood where they want are thinking of living/working, along with the type of job they are looking for. The user would then be given relevant information about both neighborhoods, along with the average commute time between them and the average salary of the the job they are looking for, if it is a city job.
- Features:
    - Education information on public schools in the neighborhood from these data sets
    - Information on the average salary earnings for city jobs
    - Free wi-fi locations in the neighborhood from this dataset
    - Demographics (primarily average age of resident) of the neighborhood:
    - Average rent based on neighborhood and apartment type.
    - We will use Google Location API to derive average commute time between the neighborhood where the user would like to live and work.
    - We will use (scrape) Yelp to identify the highest rated restaurants in both the neighborhoods the user would like to live and work.

Delivered:

- Except providing the user the information regarding the commute time and free wifi locations, we were able to achieve the rest of our goals. User will able to input neighborhood/zip where they are thinking of living/working, along with the type of job they are looking for. The user is then given relevant information about both neighborhoods such as average rent, restaurants, public school details in the living location and available jobs and their respective salary in the working location
- Feature:
    - Education information on public schools in the neighborhood from these data sets
    - Information on the average salary earnings for city jobs

- Demographics of the neighborhood: Age data, educational attainment & ethnicity of the neighborhood
- Average rent based on neighborhood and apartment type.
- Restaurant details for any neighborhood

Your final UML. Include a description of what changed in your UML (including which milestone did you make changes and why).

We added an enumeration type of Mobile Food Walk On for Restaurant type that we missed in initial data cleaning. We changed a few data, types including AgePercentage to AgePopulation, Age to DateOfBirth, added years to all of the AgeRange enumerations, Job detail Salary and Rent price from doubles to Decimals. We made all of our changes except for age to DateOfBirth for PM3, when we actually had to bring our data in and ran into difficulties. We changed age to DateOfBirth in PM4, when we received feedback after PM3 that age would be a constantly changing data type while DatOfBirth would remain static and therefore be more convenient. All of these changes were minor, and were simply to make queries and data searching more fluent.

What went well?

Since our data was almost entirely revolving around Boston, it was very easy to find relationships between datasets since most of them included Boston specific location based information like neighborhood name or ZIP code. This also made it easy to find external sources for our ETL, As we could join our data with any other Boston specific data that included Zip code or neighborhood information. Also, our initial UML was very comprehensive and so did not need to be reworked beyond minor tweaks in data types. All our initially formed relationships maintained throughout the project.

What would you do differently?

We ran into some difficulties with the initial data sets we found. Some tables were named differently than others of the same type (i.e. Downtown Financial District vs. Financial District). This is an example of a real world problem that we did not predict, and resulted in many hours of cleaning csv data to make sure we could maintain referential integrity and key relations in future PMs. This also exposed the limitation of data that we could use, and if we could do it differently we would spend more time finding data sets with more varying information to create relations with and draw conclusions from. One example is finding more specific rent/housing data from a third party source to get more accurate pricing.

What do you plan to do next?

Up next for the HubLyfe team will be to fulfill the goals that we could not get to as stated in our value proposition. The big features that we want to implement next would be the transportation service in HubLyfe, things to do in nearby areas, and integration with Yelp. Transportation features can be handled through accessing Google Maps API, BlueBikes station locations, and traffic data. For things to do in an area, there is a TripAdvisor API to search for top rated activities, places to see, and things to do. Yelp reviews can be scraped for restaurants in a given area to complement our current data sets.