Part 3; progress, challenges & lessons

Steve Asten

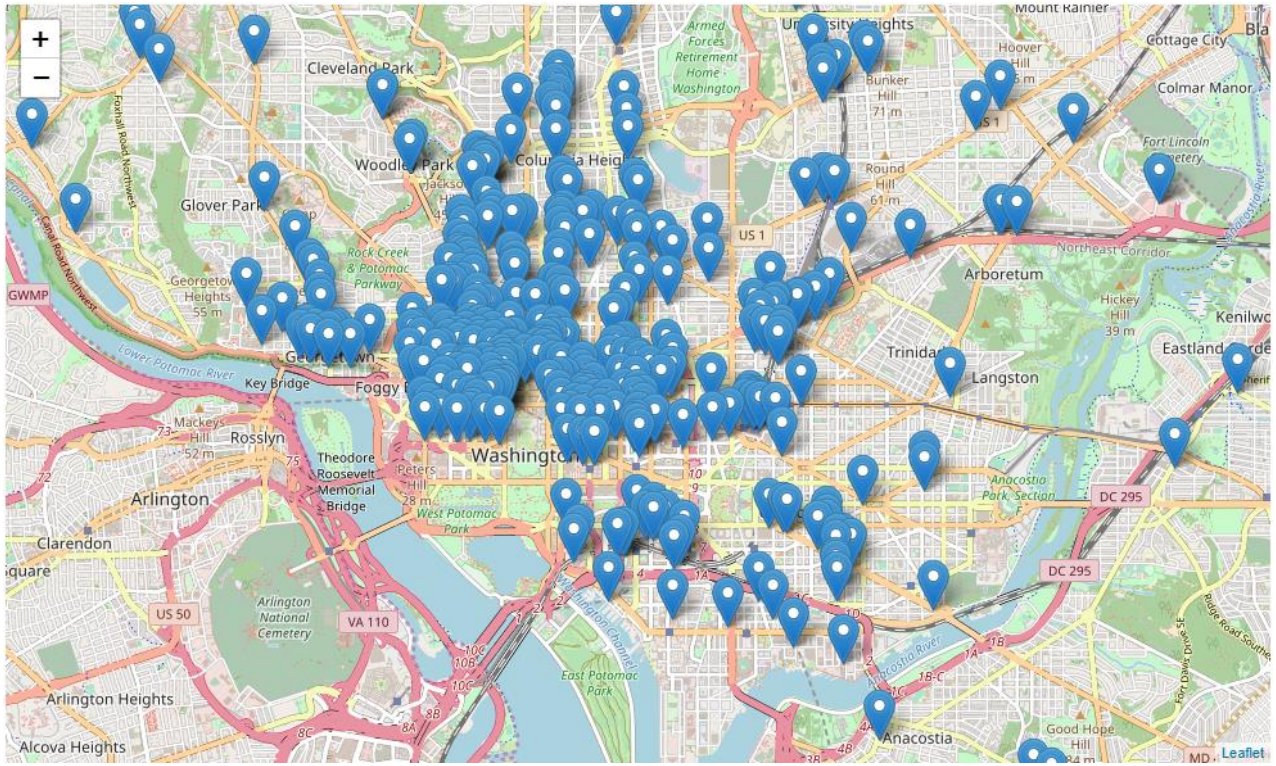
Progress 1: population is a bad indicator

- Lots of tourist attractions in DC; some zip codes have very low population but very high traffic.

zip_code	count	es	mate_total_2015	stimate_total_2016	op_change_rate
20037	18		15687	15285	-0.025626
20007	20		26687	26415	-0.010192
20010	8		32544	32421	-0.003779
20009	41		51499	51508	0.000175
20016	15		35101	35192	0.002593
20008	8		28141	28261	0.004264
20006	31		2905	2920	0.005164
20005	33		12947	13063	0.008960

Progress 2: Found all current DC businesses (including coffee shops) using Yelp API

- Currently I only have coffee shops plotted on the map. There's a lot of them, making it impossible to decide on a location purely from visualization.



Progress 3: Discovered Foursquare dataset

- I’m planning to use this to find where people are coming from when they go to a coffee shop; whether it’s a restaurant, movie theater, office, etc. This could help me predict the best location for my new coffee shop, based on current establishment locations.

Challenges/lessons

1. As we go through machine learning classes and learn new models, I'm still not entirely sure how I will apply them to my specific project.
2. Working with the foursquare dataset is very frustrating. It included a Cities table, but no foreign key in the check-ins table and thus no way to filter check-ins by city. Still trying to figure out how to efficiently geocode 250k sets of coordinates.
3. EDA for me has been difficult and unpredictable, and I should not have underestimated it when planning how much time I would put into it.
4. Yelp dataset was incomplete, and Yelp API only returns 50 results per search; need to use offsets & loops, which results in quickly using up daily-limit.