Project Objective: Creation of a platform to be able to compare each neighborhood to its perspective walk score and forecast the data

Winter Break work is highlighted!

Outline:

1. Scrape data from the internet
2. Clean data
3. Create a neural network
4. Implement the neural network with the data

Scrape Data: (Address, Zip Code and Price extracted)

Price	Address	Bathroom	Room	Furnished	Pet Allowed	Postal Code
$990.00	1438 Overdale Montréal H3G 1V3 QC Canada	1	1 1/2	Oui	Unknown	H3G 1V3
$2200.00	1045 Rue Wellington Montréal QC H3C 1V6 Canada	1	4 1/2	Oui	Non	H3C 1V6
$950.00	175 Deguire Ville St-Laurent QC H4N 1P1	1	4 1/2	Non	Non	H4N 1P1
$1550.00	Place Gennevilliers Laliberté Montréal QC H1V Canada	1	4 1/2	Oui	Unknown	
$2300.00	230 rue Dominion Montreal QC H3J 1N4	1	4 1/2	Non	Unknown	H3J 1N4
$900.00	2262 Avenue Bilaudeau Montreal H1L 4A6 QC	1	3 1/2	Oui	Non	H1L 4A6
$2600.00	635 Rue Saint Maurice Montréal QC H3C 1C7 Canada	1	3 1/2	Oui	Unknown	H3C 1C7
$790.00	2190 Rue Marie-Anne Montréal QC H2H 1M8 Canada	1	3 1/2	Oui	Non	H2H 1M8
$725.00	2710 Rue Provost Lachine QC H8S 1R4 Canada	1	3 1/2	Oui	Oui	H8S 1R4
$740.00	2990 Edouard Montpetit H4B 1Z9 Montréal QC	1	4 1/2	Oui	Unknown	H4B 1Z9
$918.00	Montréal H3T1E7 QC Canada	1	3 1/2	Non	Unknown	H3T1E7

Clean Data:

1. All data which is in the CSV file should have an address, zip code and price.
    a. If the data does not have a price or address, they must be deleted
    b. If the data has an address but no zip code, the zip code must be found
    c. If the zip code cannot be found, that house must be deleted
2. Merging all datasets together to get rid of duplicate houses.
3. Formatting or creating a new CSV file to be able to be manipulated by a Neural Network:

String	String	Latitude	Longitude	Float (Normalization)	Float (Normalization)
Full Address	Zip Code	-90 – 90	-90 – 90	Lat/90	Long/90
1438 Overdale Montréal H3G 1V3 QC Canada	H3G 1V3	45.494380	-73.572520	.50549311	-.81747244
“	“	“	“	“	“
“	“	“	“	“	“

Format the first two columns first, and save room to format the last two.

String	int	Z-score*	Float (Normalization)
Zip Code	Price	-3 – 3	Between -1 and 1
H3G 1V3	$990.00	"	"
"	"	"	"

*We will be talking about how to do this next semester with Nicolas. If you would like to get a head start, the link below explains it in detail:

https://sebastianraschka.com/Articles/2014_about_feature_scaling.html#z-score-standardization-or-min-max-scaling

4. Finally, we will be creating the master Input/output file that will be feed to the Neural Network:

Inputs (Features)	**Outputs**
Normalized float for Prices	
Normalized float for Zip Code	

Creation of Neural Network:

1. Understand what a neural Network exactly is
    a. http://www.deeplearningbook.org/
    b. http://neuralnetworksanddeeplearning.com/
    c. https://www.youtube.com/watch?v=h3l4qz76JhQ
    d. https://www.youtube.com/watch?v=aircAruvnKk
2. Learn how to implement one from Python
    a. https://pytorch.org/
    b. https://lasagne.readthedocs.io/en/latest/user/installation.html
    c. https://pythonhosted.org/nolearn/
3. Program a neural network into python that takes the above framework