

Maged Own
Coursera
March 14, 2021

Coursera project report

The identification of venues in Calgary

The purpose of this business challenge is to identify venues near random addresses in Calgary and segment and cluster them accordingly.

Calgary is a city in North America with multiple communities identified as community names and in each community there are addresses that can be explored through Foursquare API to return the venues of popularity helping to determine which area can be selected for future establishments.

The dataset was taken from the city of Calgary website and then it was processed to identify what has to be done to it to make it more useful for this project.

The link to the data is <https://data.calgary.ca/resource/ejve-em7f.json>. After the data has been parsed from the link it should look like:



	roll_number	roll_year	comm_name	year_of_construction	assessment_class_description	address	comm_code	property_type	assessment_class	unique_key	re_assessed_value	assessed_value	land_size_sf	sub_property_use
0	560981565	2021	RESIDUAL WARD 9 - SUB AREA 90	1981	Residential	105 1101 84 ST NE	09O	IO	RE	2021560981565	39500	39500	NaN	NaN
1	201641974	2021	PANORAMA HILLS	NaN	Residential	100V 175 PANATELLA HL NW	PAN	LO	RE	2021201641974	10000	10000	151	TPU
2	560985301	2021	RESIDUAL WARD 9 - SUB AREA 90	1999	Residential	805 1101 84 ST NE	09O	IO	RE	2021560985301	125000	125000	NaN	NaN
3	201208238	2021	MCKENZIE TOWNE	NaN	Residential	48V 115 PRESTWICK VI SE	MCT	LO	RE	2021201208238	10000	10000	149	TPU
4	202640405	2021	DOWNTOWN EAST VILLAGE	2015	Residential	292V 650 5 ST SE	DNE	LI	RE	2021202640405	34000	34000	0	TPU

The above will still need processing and cleaning to make it more useful. And this is what will be done in this project.