CLUSTERING GEOGRAPHIC POINTS FOR WAREHOUSES LOCATION

BUSINESS PROBLEM

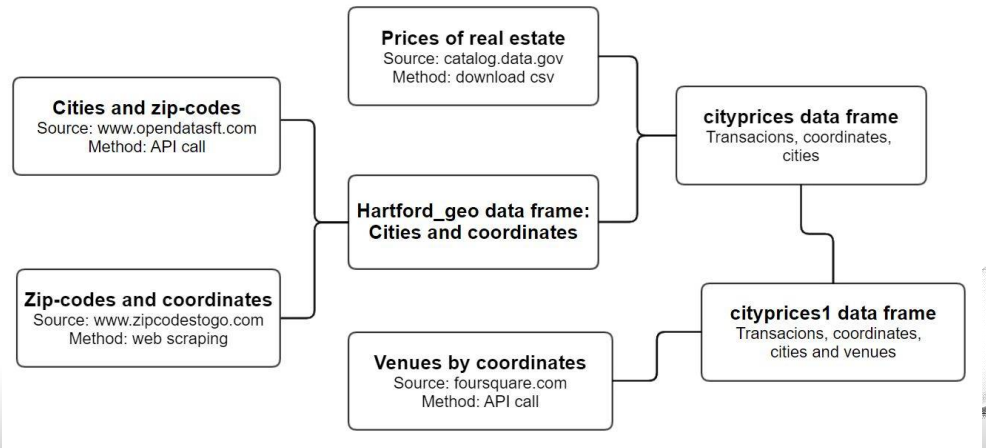
BUSINESS PROBLEM:

- WHOLESALE DISTRIBUTOR WITH INTENTION OF EXPANDING ITS OPERATIONS TO NEW GEOGRAPHIC AREA
- HOW TO EVALUATE WHICH GEOGRAPHIC LOCATION IS BETTER FOR ACQUIRE A PROPERTY TO SET THE DISTRIBUTION WAREHOUSES UP?

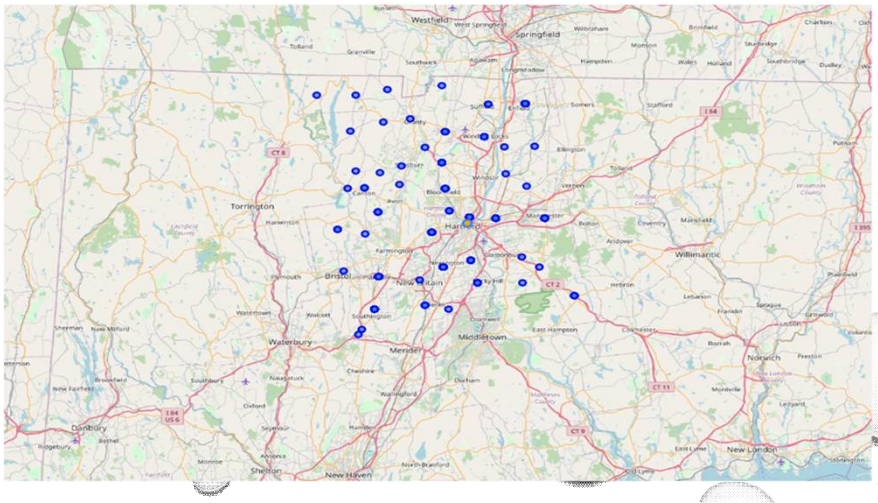
SOURCES

- ALL CITIES WITH ITS ZIP CODES, SCRAPED FROM HTTPS://WWW.ZIPCODESTOGO.COM.
- ALL CITIES WITH ITS GEOGRAPHIC COORDINATES, IN JSON FORMAT THROUGH AN API CALL TO HTTPS://PUBLIC.OPENDATASOFT.COM/API/
- PRICES OF REAL ESTATE PROPERTY IN THE AREA OF INTEREST, IN CSV FORMAT DOWNLOADED FROM CATALOG.DATA.GOV
- VENUES TO WHOM OUR ITEMS WILL BE DISTRIBUTED, EXTRACTED IN JSON FORMAT FROM FOURSQUARE.COM, THROUGH AN API CALL

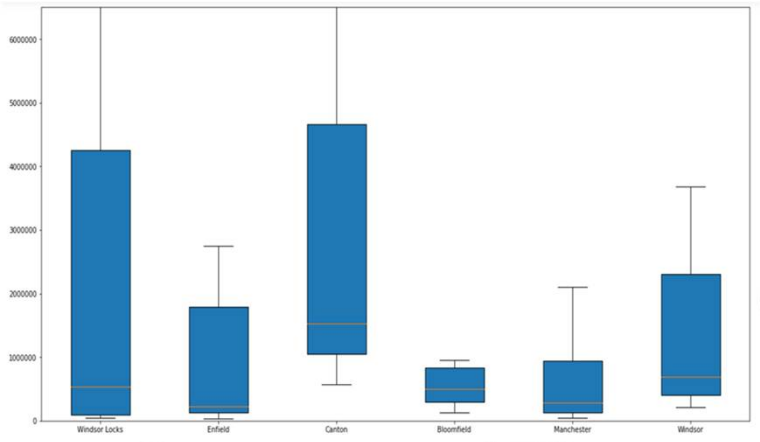
DATA PREPARATION



VISUAL ASSESSMENT OF GEOGRAPHIC DISPERSION

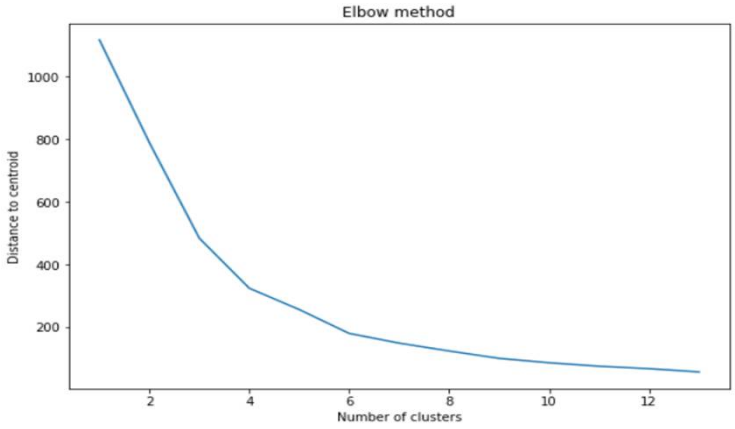


DIFFERENCES IN PRICES



MACHINE LEARNING TOOLS APPLICATION – K-MEANS CLUSTERING

- ELBOW METHOD USED
- CHOSEN K = 6



RESULTS: MEAN VALUES

Labels	longitude	latitude	SaleAmount	StatDistance	Venue Counts
0	-72.807124	41.713834	1,079,346.1	65.125573	15.072464
1	-72.521871	41.771109	1,093,011.2	64.840289	6.260000
2	-72.604996	41.918199	52,714,285.7	64.923615	11.000000
3	-72.679790	41.804220	1,081,363.0	64.998237	7.153846
4	-72.879071	41.666106	1,011,284.9	65.197640	6.447368
5	-72.563927	41.976495	846,760.4	64.882684	15.038462

RESULTS: DISCUSSION

CLUSTERS 1, 3 AND 4: REASONABLE PRICES, THEIR MEAN DISTANCE RELATIVELY LOW VENUES OF INTEREST IN THEIR PROXIMITY

CLUSTER 2: PROPERTIES WITH LEVELS OF PRICES THAT MAKES IT IMPOSSIBLE TO BE CONSIDERED

CLUSTER 0 : REASONABLE PRICES, SOMEWHAT FAR FROM THE STATION, BUT A GOOD CANDIDATE FOR NUMBER OF VENUES NEAR.

CLUSTER 5: BEST COMBINATIONS OF FEATURES ACCORDING TO OUR CRITERIA