## Big data Analytics in real estate

Done By: Sukesh Pavan Bitragunta

Submitted to: Andrew Enkeboll



## Problem Statement

Big data will be very helpful when we are dealing with the large data sets. The data set is taken from the Norfolk open data source, it has 74Krows and it consist the information about the property sales in Norfolk city. By exploring data, we can find out which area has the highest sales and which domain has more sales in residential, commercial and empty land.

Data Source: https://data.norfolk.gov/Real-Estate/Property-Assessment-and-Sales-FY21/8bfx-a5g8/data



DISH	Intot
Parcel ID	float64
Extension	object
uPIN .	#loate4
tegal Description	object
Owner	object
Property Street Number	float64
Property Street Number Sufflx	object
Property Street Direction	obsect
Property Street Name	obsect
Property Street Type	object
Property Street Suffix	float64
Unit Number	obsect
Property City	obsect
Property Zip	object
Acreage	Float64
Land Square Footage	float64
Property class Description	object.
Property Use	object.
Mesidential Finished Living Area	floates
commercial Suilding Area	Float64
Improvement Year Built	Float64
Current Improvement Value	floates
current Land Value	floates
Current Total Value	Floot64
Effective Year	object
Transfer Date	object
Scantor	object.
Snantee	object
Consideration	Float64
Document Number	object

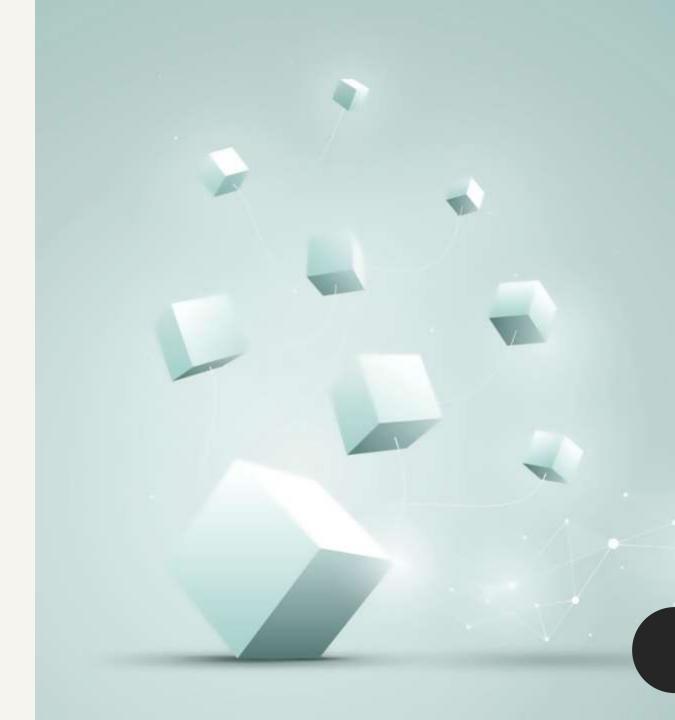
Owner	
fajardo Hernandez, Veronica S	1
s Virginia Propeties, Llc	1
10 Summer Project Llc	2
100 W21, Llc	1
1001 W 27th Street, Llc	2
Zuraw, Richard P & Patricia W	2
Zwack, Lauren P	2
Zwahlen, Norman J	1
Zwahlen, Norman J Trust	1
Zyskowski, Anita Lisa	1
Name: Property Zip, Length: 57982	, dtype: int64

The top image has the variables that are present in the data set, this will help us to know to what is the data type of the column/variable and helps us to know how to deal with them.

The bottom image is an insight drawn from the data set, that tell us how many properties does an owner has. This can further help us to evaluate the amount of money does owner kept in the property.

## Solution

This project will be done by using pyspark which is very famous for big data and perform EDA on the dataset. This EDA will help us to find the hidden insights of the dataset and let us know which sales/transactions that are beneficial to organization.



## Thank you