

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 74K rows and it consists 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>



LRN	int64
Parcel ID	Float64
Extension	object
APN	Float64
Legal Description	object
Owner	object
Property Street Number	Float64
Property Street Number Suffix	object
Property Street Direction	object
Property Street Name	object
Property Street Type	object
Property Street Suffix	Float64
Unit Number	object
Property City	object
Property Zip	object
Acreage	Float64
Land Square Footage	Float64
Property Class Description	object
Property Use	object
Residential Finished Living Area	Float64
Commercial Building Area	Float64
Improvement Year Built	Float64
Current Improvement Value	Float64
Current Land Value	Float64
Current Total Value	Float64
Effective Year	object
Transfer Date	object
Grantor	object
Grantee	object
Consideration	Float64
Document Number	object

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.

```
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
```

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

