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.

Introduction

Let us see how many columns are in the data set.

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

Exploratary Data Analysis

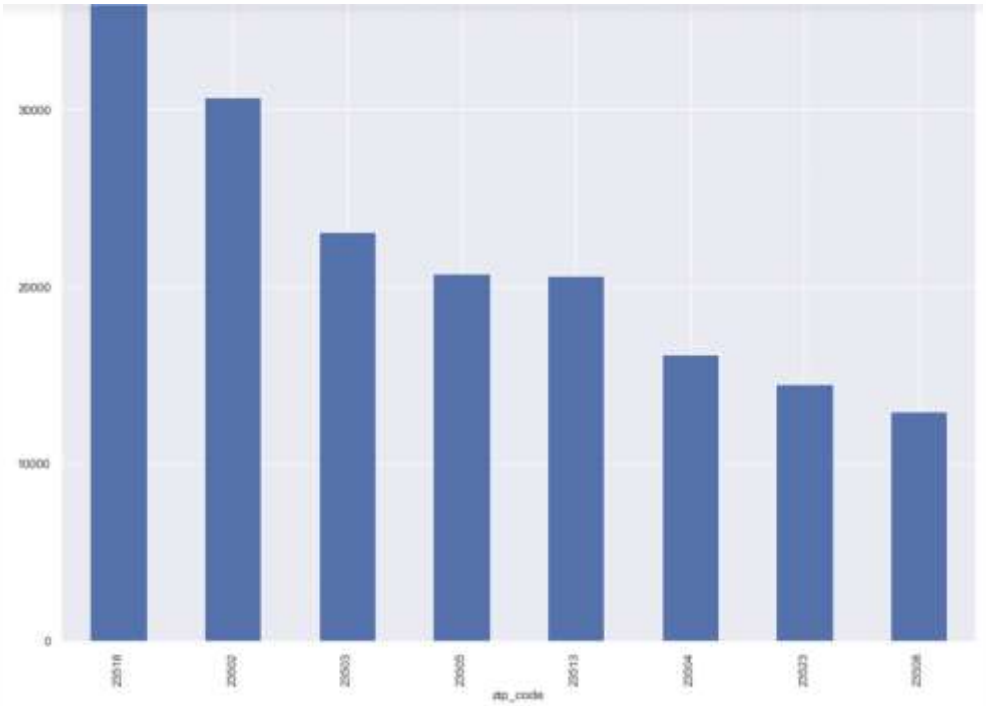
- Unnecessary Columns are removed from the data frame. Such as property street number, property suffix, Direction, Parcel ID.

- Some null values are replaced with the mean of the column and some with the mode .

Let us analyze the columns

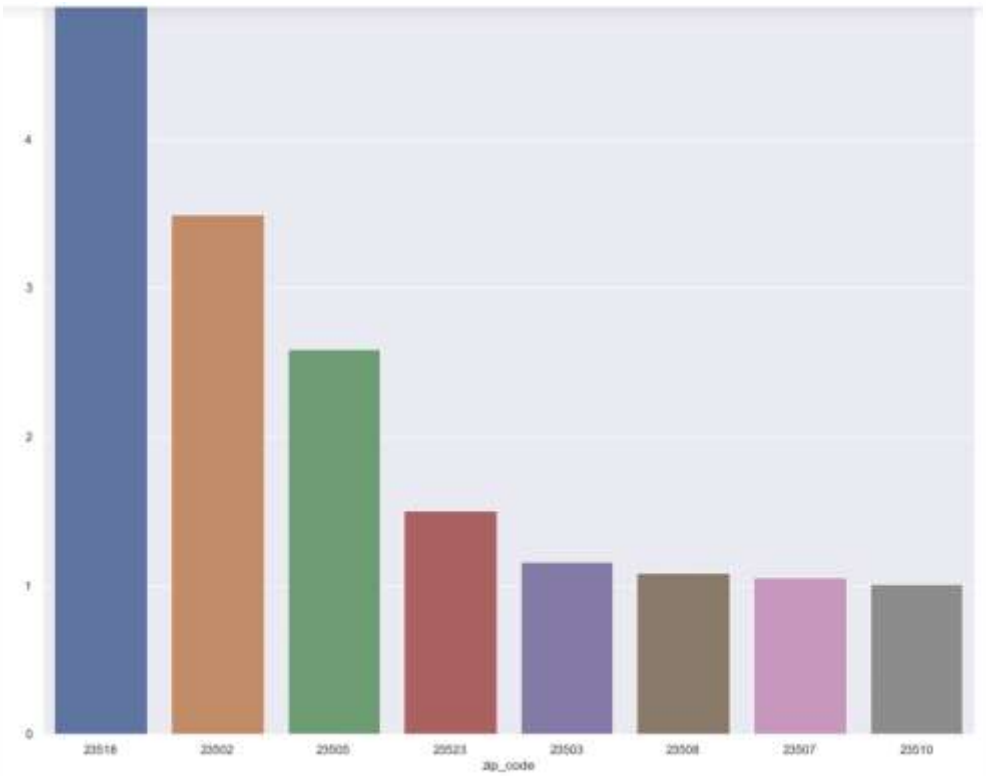
- There are total 40 plus zip codes in Norfolk city.
- Let us see the top 8 zip codes with the most acreage.

These zip codes have highest acreage compared to the other zip codes.

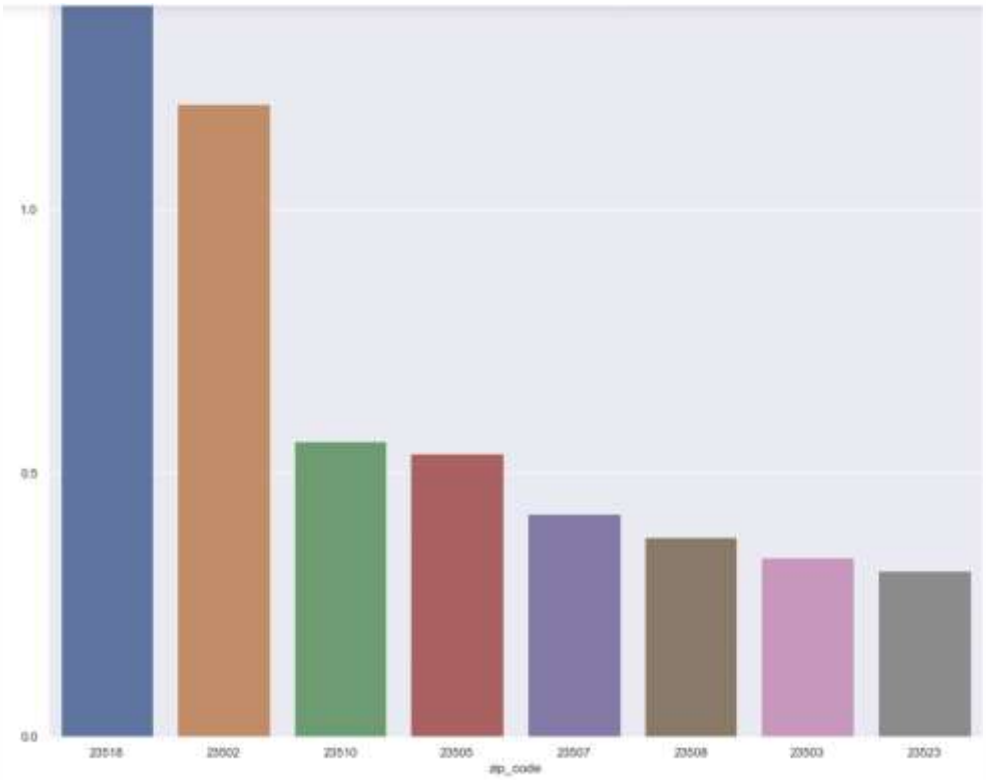


- Top 8 zip codes with the most land cost

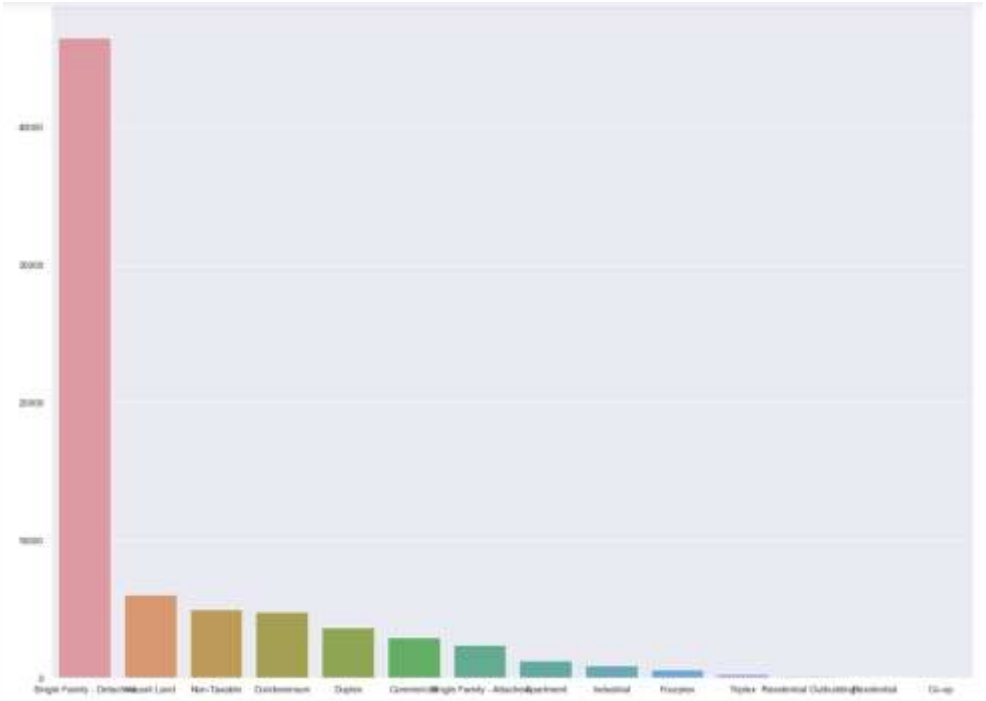
zip codes 25318, 25302, 25303 have most acreage and most land value cost.



- Including the both cost of the
- land and cost of the building these zip codes have highest total value among the others.

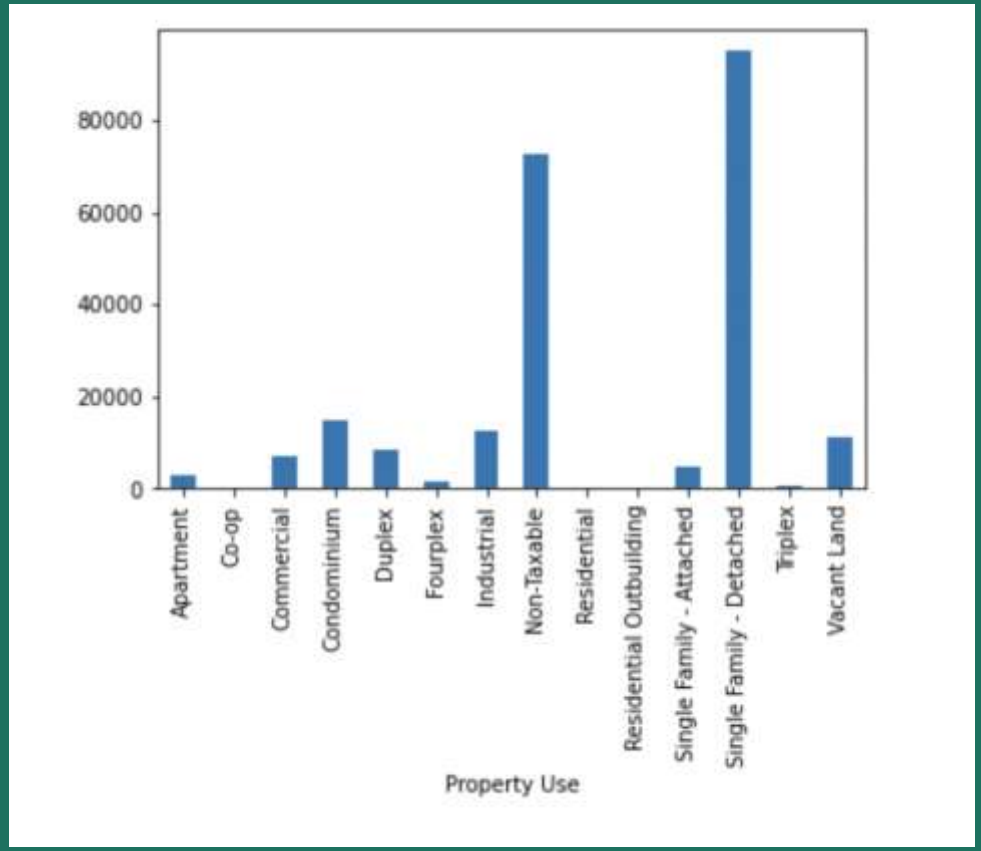


- There are different types of poperties such as single house, single house detached, Apartment (Single, Doble).
- In the given data set we have total 14 types of properties and what is the count of each property type is also calculated.



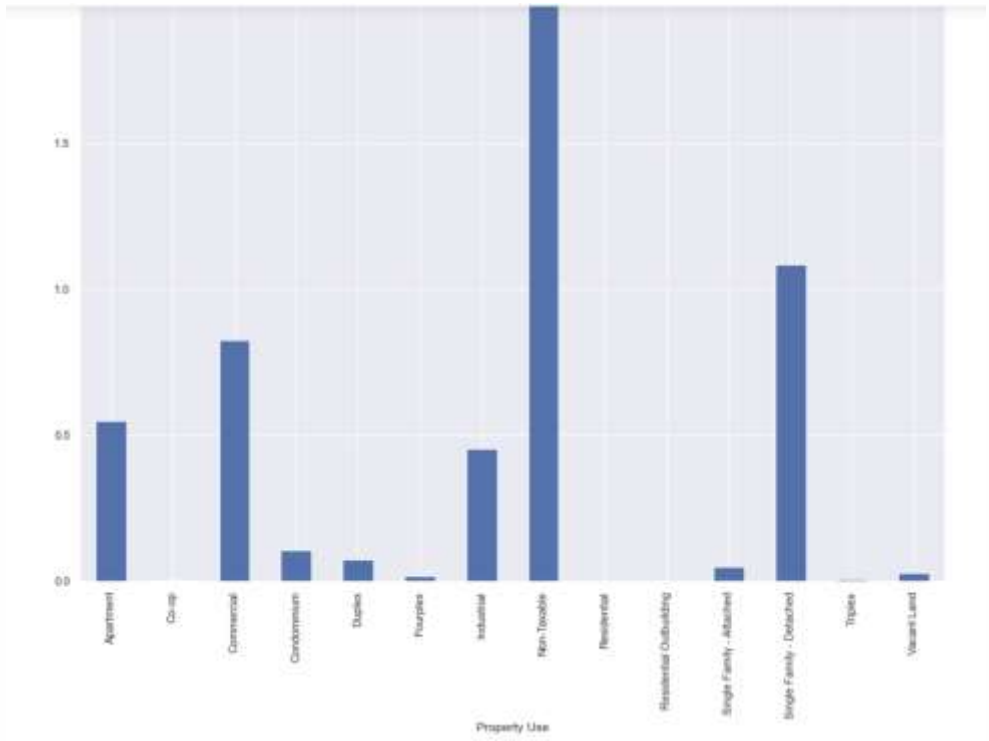
- This bar graph shows us how much acreage does each type of the property have and how their total valuation combing.

- Single family detached had highest acreage and while co-op has least compared to all others.



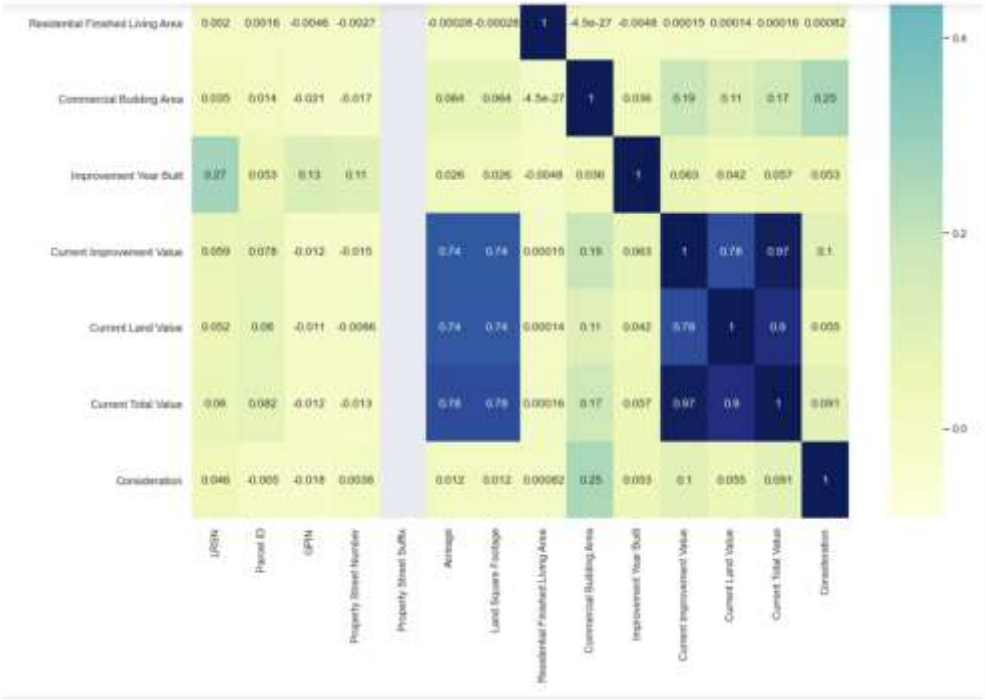
- But coming to total value, non-taxable land has more value compared to single detached.

- Condium and vacant land are almost equal but the value for vacant land is more compared to condium.



- Let us the correlation between the variable and see which land effects the total cost of the property.

- The total value has a high correlation with acreage, Land value and Improvement value.



THANK YOU