

Assignment -2 Report

My Dataset is about Melbourne Housing Market, this data set

Source Of my dataset- <https://www.kaggle.com/anthonypino/melbourne-housing-market>

Intend source of my data- With this data we can predict and analyze the cost, wealth, field value of a particular area, land value , as well in business perspective we can estimate the value and shops covering the region. Based on this data analysis we can make evidence based theorem in construction area over Melbourne.

Agencies working with my data - This data was scraped from publicly available results posted every week from Domain.com.au

Variable with the dataset

Suburb: Suburb

Address: Address

Rooms: Number of rooms

Price: Price in Australian dollars

Method: S - property sold; SP - property sold prior; PI - property passed in; PN - sold prior not disclosed; SN - sold not disclosed; NB - no bid; VB - vendor bid; W - withdrawn prior to auction; SA - sold after auction; SS - sold after auction price not disclosed. N/A - price or highest bid not available.

Type: br - bedroom(s); h - house,cottage,villa, semi,terrace; u - unit, duplex; t - townhouse; dev site - development site; o res - other residential.

SellerG: Real Estate Agent

Date: Date sold

Distance: Distance from CBD in Kilometres

Regionname: General Region (West, North West, North, North east ...etc)

Propertycount: Number of properties that exist in the suburb.

Bedroom2 : Scraped # of Bedrooms (from different source)

Bathroom: Number of Bathrooms

Car: Number of carspots

Landsize: Land Size in Metres

BuildingArea: Building Size in Metres

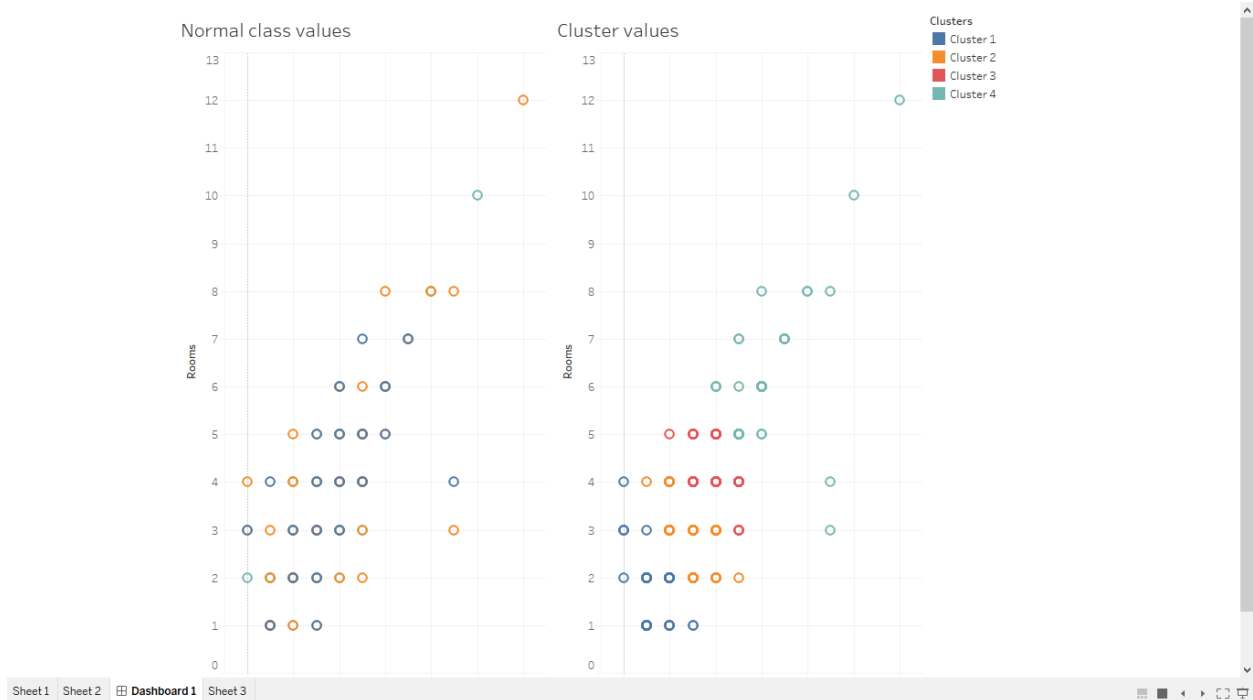
YearBuilt: Year the house was built

CouncilArea: Governing council for the area

Latitude: Self explanatory

Longitude: Self explanatory

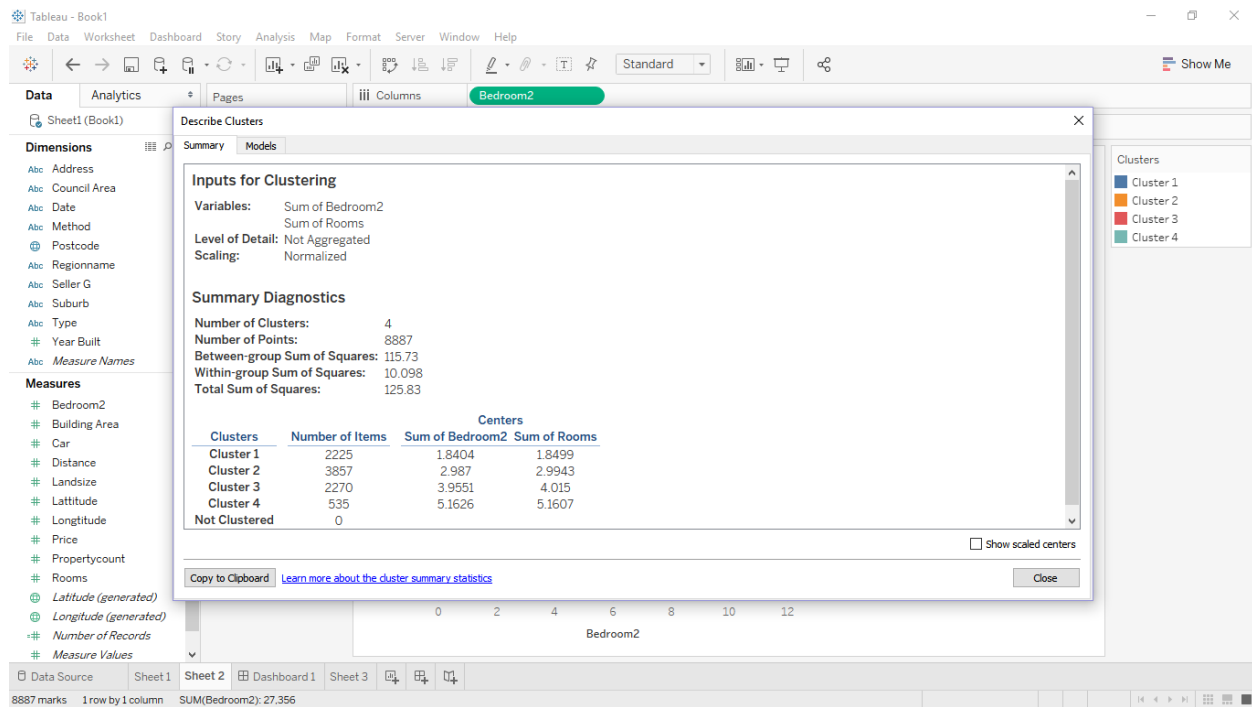
Ex-1 : Variables of Room And BedRoom2 taken for this clustering.



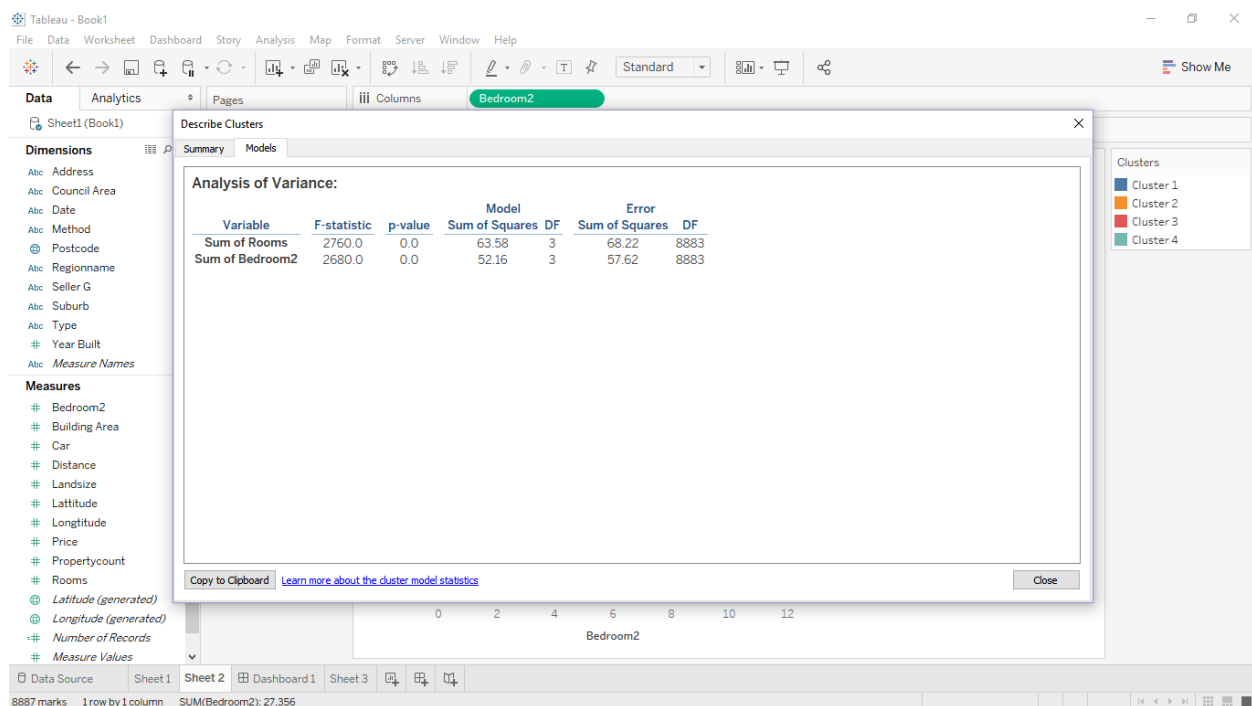
In theory, data points that are in the same group should have similar properties and/or features, while data points in different groups should have highly dissimilar properties and/or features. Clustering is a method of unsupervised learning and is a common technique for statistical data analysis used in many fields. In Clustering grouping of similar objects or same type of data can be formed together and perform process. we can use clustering analysis to gain some valuable insights from our data by seeing what groups the data points fall into when we apply a clustering algorithm.

With the above example two variable take to prove or gain the cluster. In this room and bathroom2 having a great co-relation with one another this clustering is formed.

With operation 4 cluster is formed this explains that data points are properly shifted or assigned with one another.



The above pic represent the nature of the points discussed in clustering.

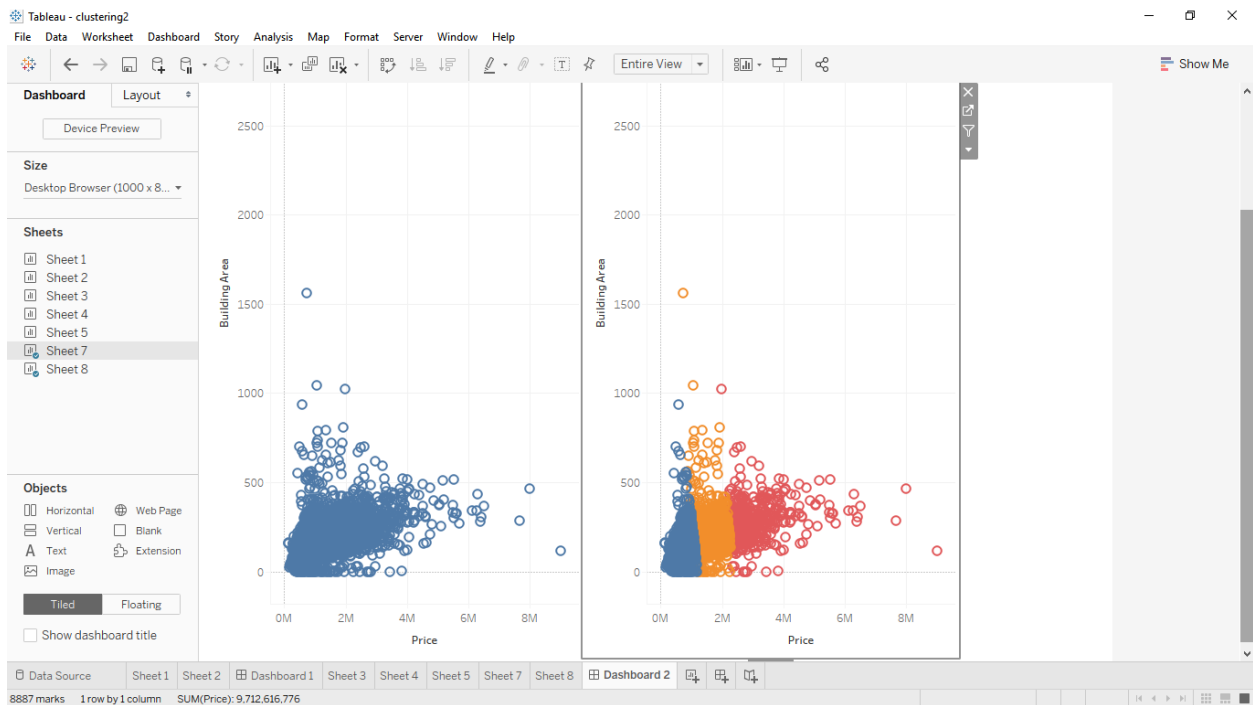


This explains the finalized report of the cluster.

Ex – 2

This Represent the factor of Clustering with another set of variable here the performance of the cluster is much efficient than the previous one.

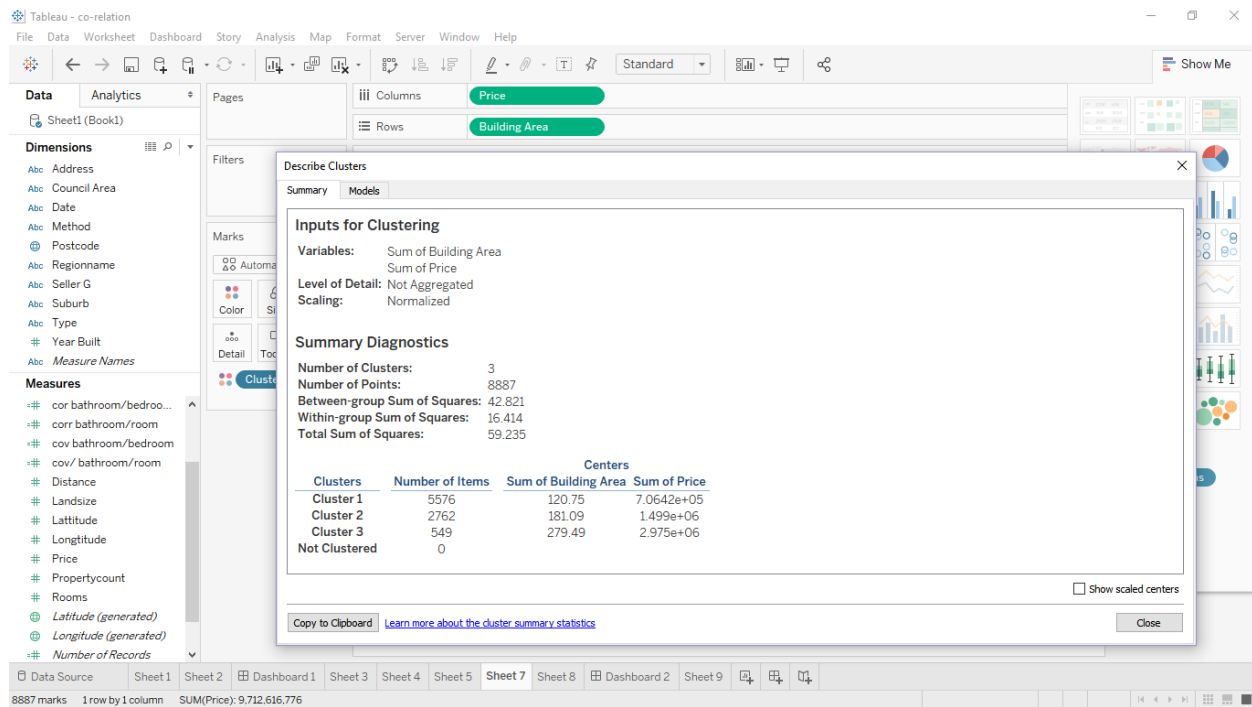
Variable of Building area with price



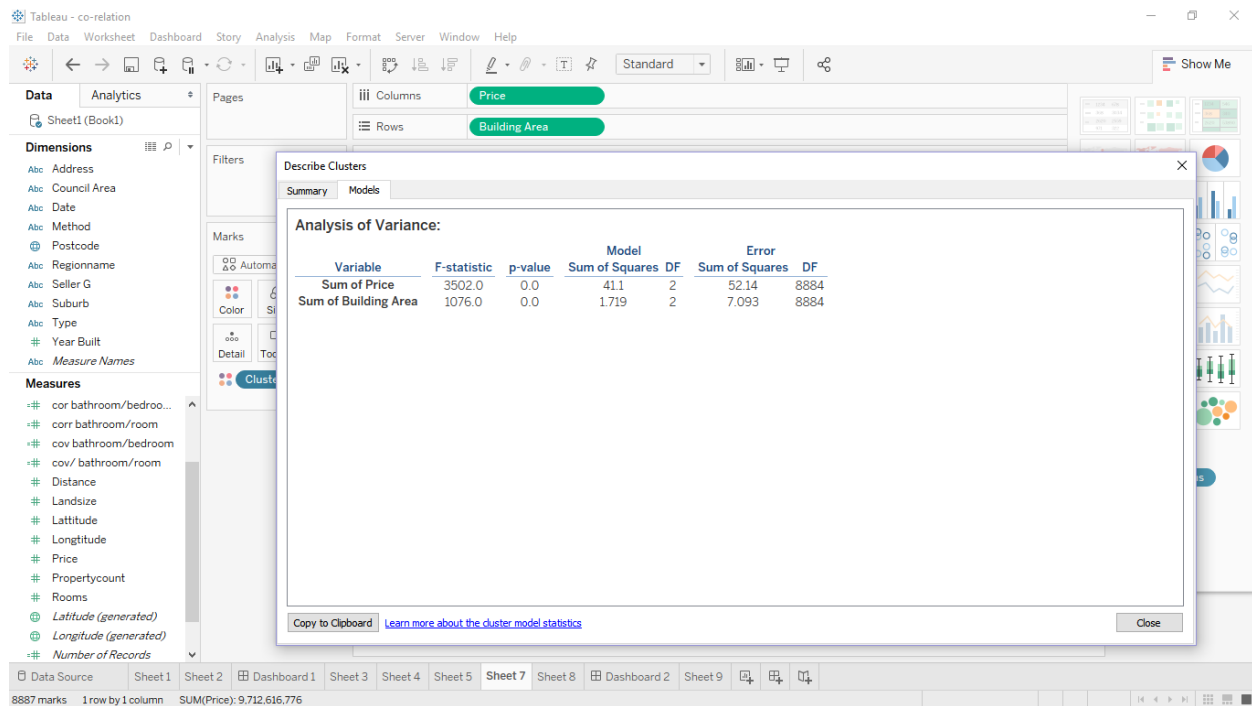
In this cluster this assigned more properly than the previous one. Here we don't have any No cluster objects but some point turns as border point Which is also associated with the cluster.

Noise region are less in this cluster while comparing with the previous one.

Supportive evidence are presented below



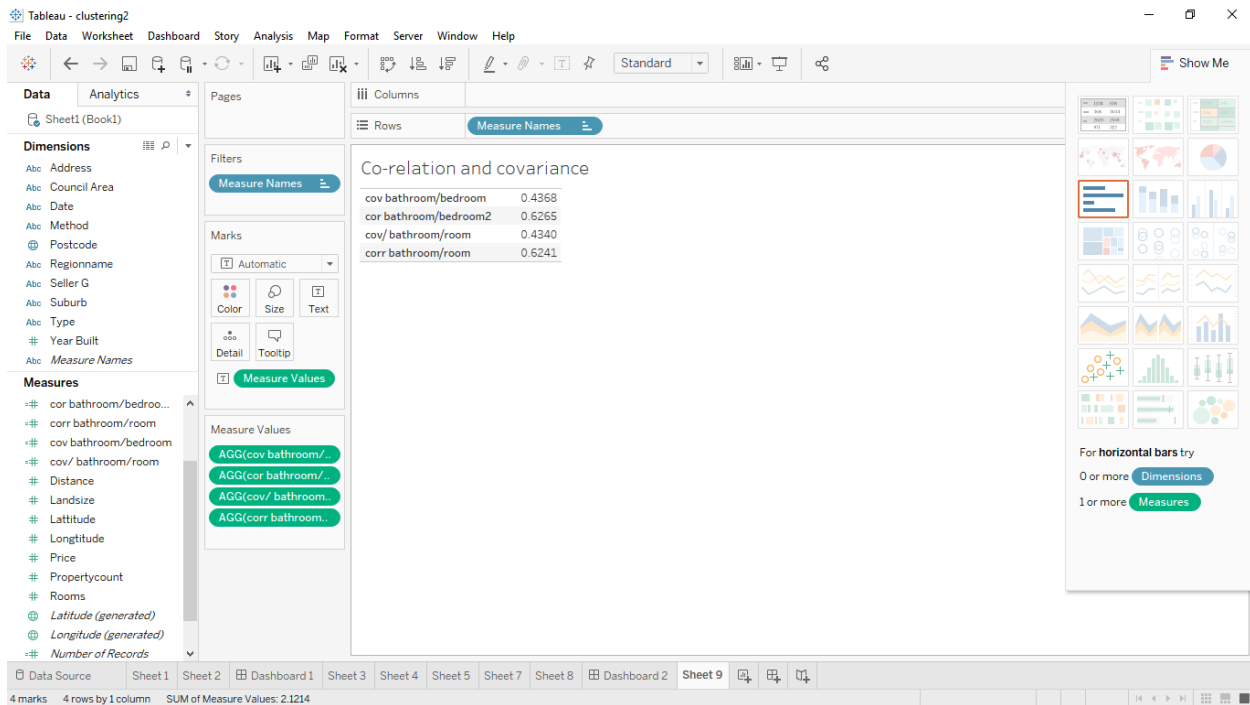
Clustering models with final outcomes



Ex - 3

Co – relation

Correlation is described as the analysis which lets us know the association or the absence of the relationship between two variables 'x' and 'y'. To find a numerical value expressing the relationship between variables. With-in this definition two variables take to find the relation between has been discussed below



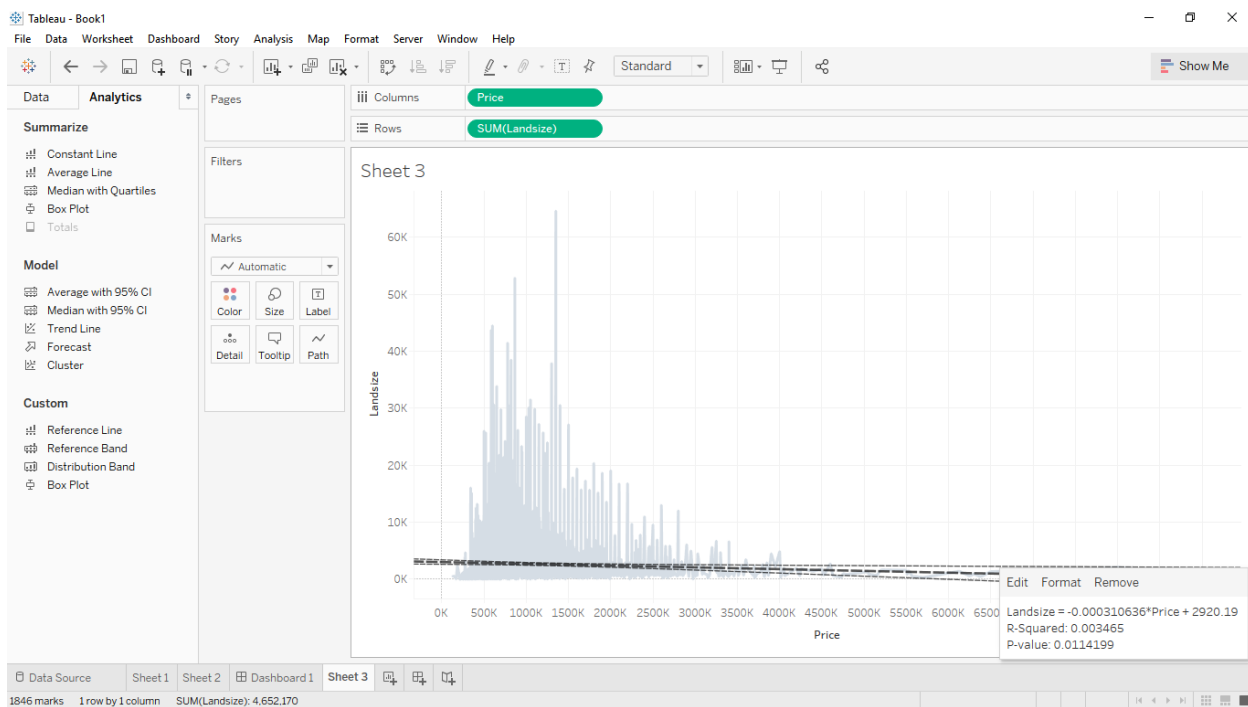
The co – relation values shows .6 and .6 in both set of values for bathroom and bedroom as well as for bathroom and room are the variable taken here. This value denotes that positive co – relation between them in two aspects. Co Variance results about the how closely this co-ordination relate to one another.

Ex – 4

Regression

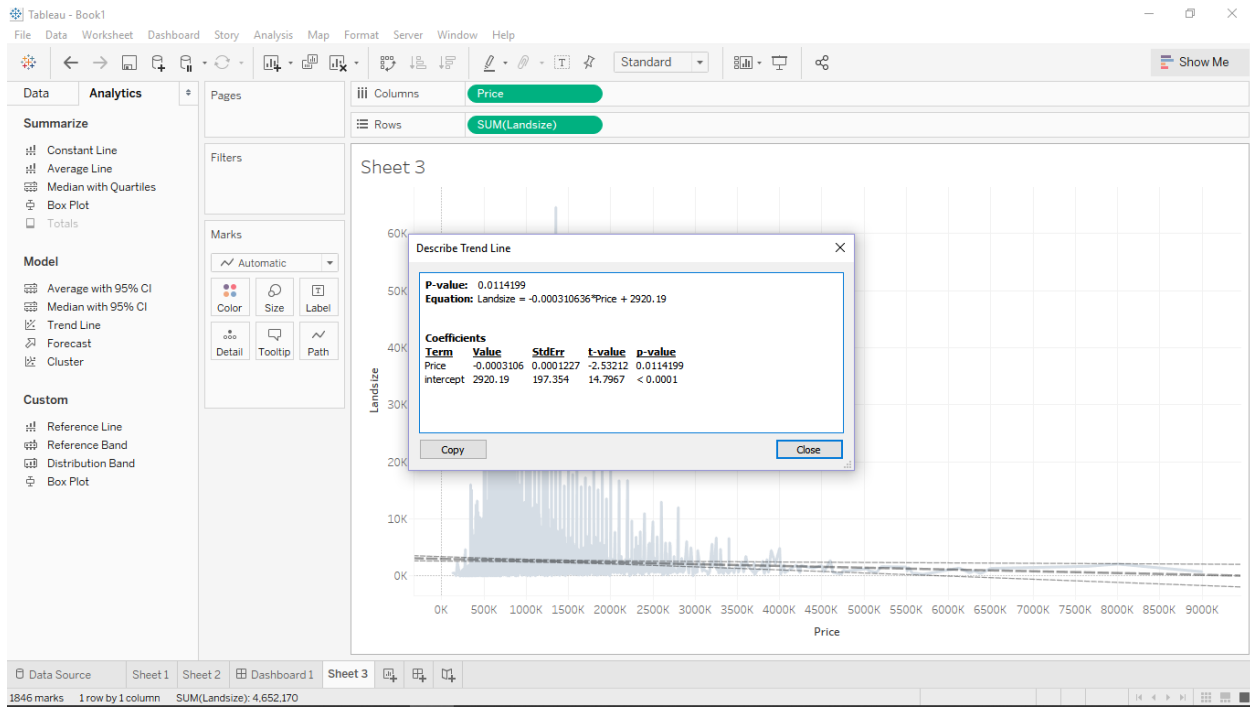
Regression describes how an independent variable is numerically related to the dependent variable. To estimate values of random variable on the basis of the values of fixed variable. This Provides the value of how this dependent variable variable makes a causation effect to the independent discussed here.

Price and Landsize are the two variable taken as x and y.



Thus it forms a regression line those values are discussed here, wheatear it fits the best fit line or not on the basis of less than 0.5 value in P

It explained and It fits the line in a negative way.



All factors of the

