FACTOR ANALYSIS



**Component Matrix**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 |
| price | .665 | .411 | .322 | -.068 | .214 |
| bedrooms | .720 | .266 | -.349 | .204 | -.336 |
| floors | .563 | -.444 | .283 | .161 | .451 |
| waterfront | .126 | .290 | .695 | -.501 | -.258 |
| condition | -.178 | .666 | -.332 | -.210 | .499 |
| zipcode | -.303 | .251 | .469 | .739 | .025 |
| new\_bathrooms | .910 | .128 | -.153 | .148 | -.124 |
| yr\_built (Binned) | .515 | -.686 | -.006 | -.178 | .082 |
| sqft\_living (Binned) | .883 | .204 | .017 | .017 | .051 |

Extraction Method: Principal Component Analysis.

a 5 components extracted.

We highlights in our new factors the most important components for each one. For factors number 5 we don’t have any very important component then the others.

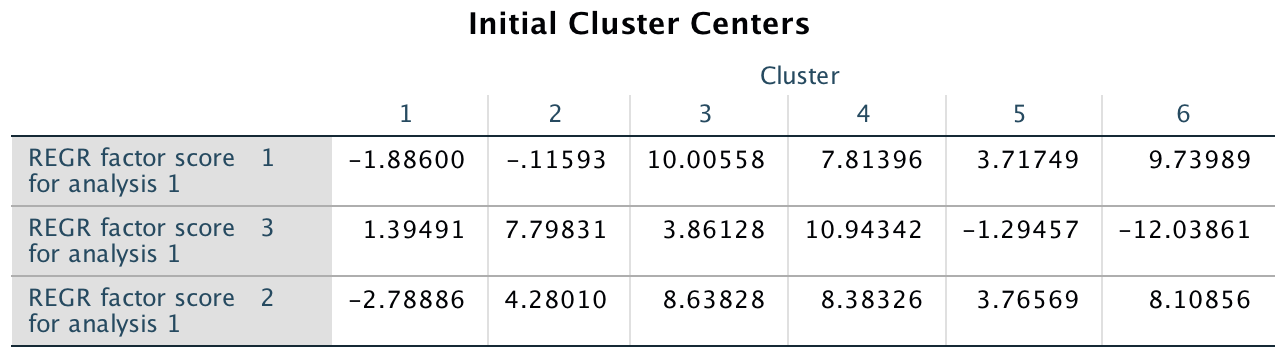
2STEP CLUSTER ANALYSIS

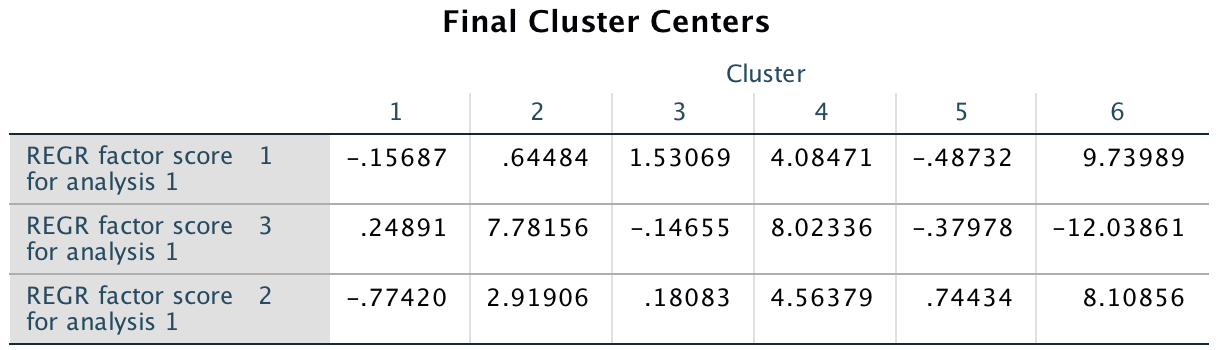
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Cluster** | **percent(values)** | **percent(values)** | **V4** | **V5** |
| 1 | 0.819 | 177 | 1 | 0.8189515569333272 |
| 2 | 13.3253 | 2,880 | 2 | 13.325313468745662 |
| 3 | 18.4657 | 3,991 | 3 | 18.465738213112477 |
| 4 | 27.2197 | 5,883 | 4 | 27.21972886688567 |
| 5 | 22.3292 | 4,826 | 5 | 22.329153750057834 |
| 6 | 17.8411 | 3,856 | 6 | 17.841114144265024 |



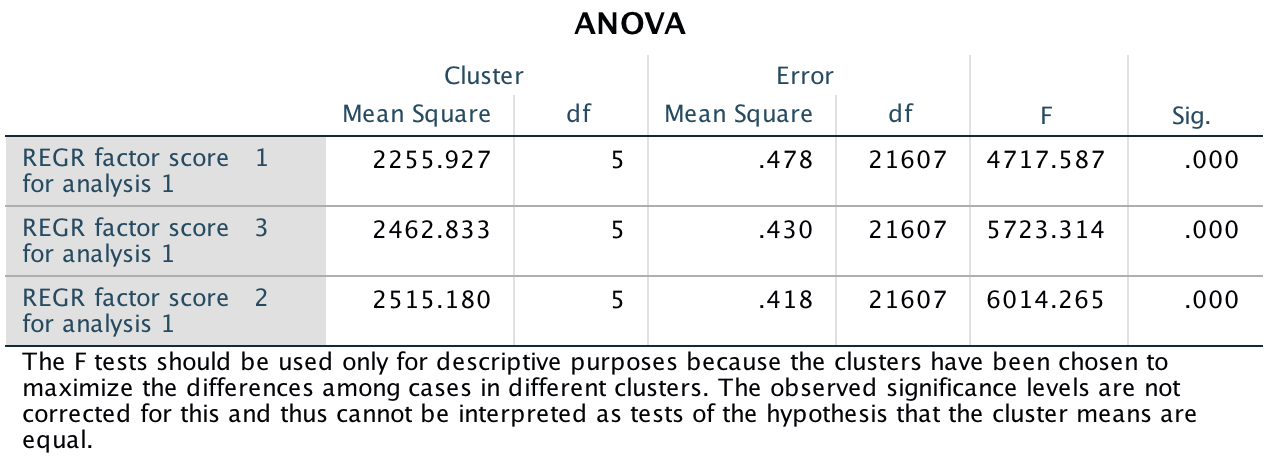
With the 2Step analysis we try find the best number of clusters. The algorithm give us as the best choice K = 6. Analysing the Silouhette graph we can see that the quality of the clusters is not very good, but this is the best result we could get from this dataset.

K-MEANS

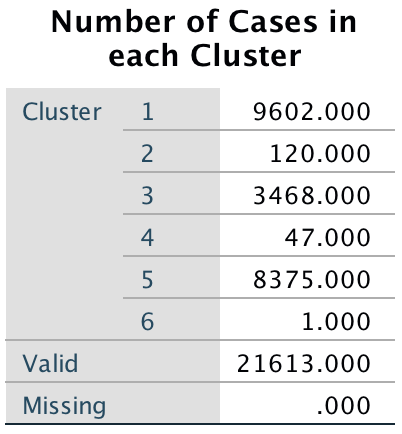




do we have to explain the cluster? We think that for example Cluster 4 and 6 has the huge amount of bathroom bedrooms and maybe they are big houses, because in the factor analysis we found that factors 1 is more correlated to bedrooms, bathrooms and sqft\_living.



All the 3 factors are significant.



In particular, for CLUSTER 6 the observation is the n° 15871. We present now the value for this one

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Price | Bedroom | Sqft\_liv | Floors | Waterfront | Condition | Year | Zipcode | Bathrooms |
| 6.4E+005 | 33 | 1620 (group n°7) | 1.0 | 0 | 5 | 1947  (group n° 4) | 98103 | 58.00 |

|  |  |  |
| --- | --- | --- |
| Factor 1 | Factor 2 | Factor 3 |
| 9.73989 | 8.10856 | -12.03861 |

We think this observation is the only one in the cluster number 6 because for Factor n° 1 (found with the factor analysis) we have that the data are explained most with bedrooms ( 72%) bathrooms (91%) and this house has a very high number of this two things.