

TECHNICAL UNIVERSITY OF CIVIL ENGINEERING BUCHAREST
DOCTORAL SCHOOL

DATABASES AND THE COMPUTERIZATION OF THE REAL ESTATE APPRAISAL PROCESS

Research Report no. 3

Doctoral Thesis Advisor,
Prof. Univ. Eng. Dumitru Onose Phd

PhD Student,
Eng. Silvia Zeicu Chiorean

December 2016



TABLE OF CONTENTS

Chapter 1 - Introductory Notions	2
1.1. Introduction.....	2
1.2. Automated Valuation Models.....	Error! Bookmark not defined.
Chapter 2 - Sampling Procedures	3
2.1. General Notions.....	3
2.2. Probabilistic Sampling Procedures	4
2.3. Sample Volume.....	5
Chapter 3 - The Process of Developing and Automated Valuation Method.....	6
3.1. Stages of Developing and Automated Valuation Model.....	6
3.2. Identification of the subject real estate (database - bank guarantee portfolio to be valuated)	7
3.3. Establishing Assumptions in Valuation.....	10
3.4. Data Management and Data Quality Analysis	11
3.5. Establishing Model Specifications.....	15
3.6. Model Calibration.....	17
3.7. Model Testing and Quality Assurance	19
3.8. Periodic Verification of Model Accuracy.....	20
Chapter 4. Objectives and Proposals	21

Chapter 1 - Introductory Notions

1.1. Introduction

At present, specialist appraisers use various statistical methods for appraising real estate properties, a multitude of solutions provided by the mathematical statistics to be applied in this field: MRA - multiple regression analysis, time series, fuzzy logic techniques, ANN – artificial neural networks, NFS - neuro-fuzzy systems and others.

At international level, mathematical statistic-based solutions emerge from different corners of the world and from both areas of expertise, the academia (universities, research institutions) and industry (banks, appraiser companies, real estate agencies), and even more, their symbiosis with the Geographic Information Systems (GIS) or web applications.

The immediate and very palpable advantages of applying these valuation solutions are the automatization of the appraisal process and the existence of the control methods for the results achieved. The main disadvantage is the complexity of applying the suggested solutions, requiring a team of specialists from different fields - programmers, statisticians, mathematicians, appraisers, market analysts - to develop and operate systems of this nature and scale.

1.2. Automated Valuation Models

Creating an Automated Valuation Model (AVM) can be a solution for appraising real estate guarantee portfolios held by banks. AVMs are software that produce estimators of market values for the subject real estate, based on the analysis of market conditions and characteristics of comparable real estates, market information collected in advance. AVM are particularly applicable in the real estate appraisal through direct market comparison.

AVM credibility and accuracy of its results depend on the quantity and quality of the data used in the valuation, experience and training of the staff that designs and develops the model. The amount of data refers to the volume of sample data used in the valuation. As regards the quality of data in the process of developing an AVM, a special stage for the management and analysis of the data quality is designated.

Given the impossibility of using the entire data population (entire portfolio), it is necessary to extract a representative sample of the whole population. A sample is representative of the entire population if its structure is identical or very similar with the structure of the population where it has been extracted so that it is possible to extrapolate the findings to the entire population. The sample will be representative when using an objective, random sampling method, with anticipated probability and if the sample has a sufficiently large volume of data.

For the properties constituting the analysed sample, both the market values estimated by applying the assessment methods and techniques recommended by IVSs and their characteristics must be known.

Chapter 2 - Sampling Procedures

2.1. General Notions

At present, the most commonly used way of obtaining data for statistical analysis is the **statistical survey**.

The statistical survey may be probabilistic (or random) and non-probabilistic (or non-random). The differences between the two types of surveys consist in the fact that, in probabilistic procedures, the selection of the units required for sampling is random, in this respect each unit has a known p probability and different from zero to be in the sample, while the non-probabilistic procedures select judiciously the units for the sample to best present the essential characteristics of the population from which it is extracted, but it is not possible to know the probability of the unit being in that sample.

For the reasons listed above, the statistical survey using non-probabilistic methods is not recommended in appraising real estate.

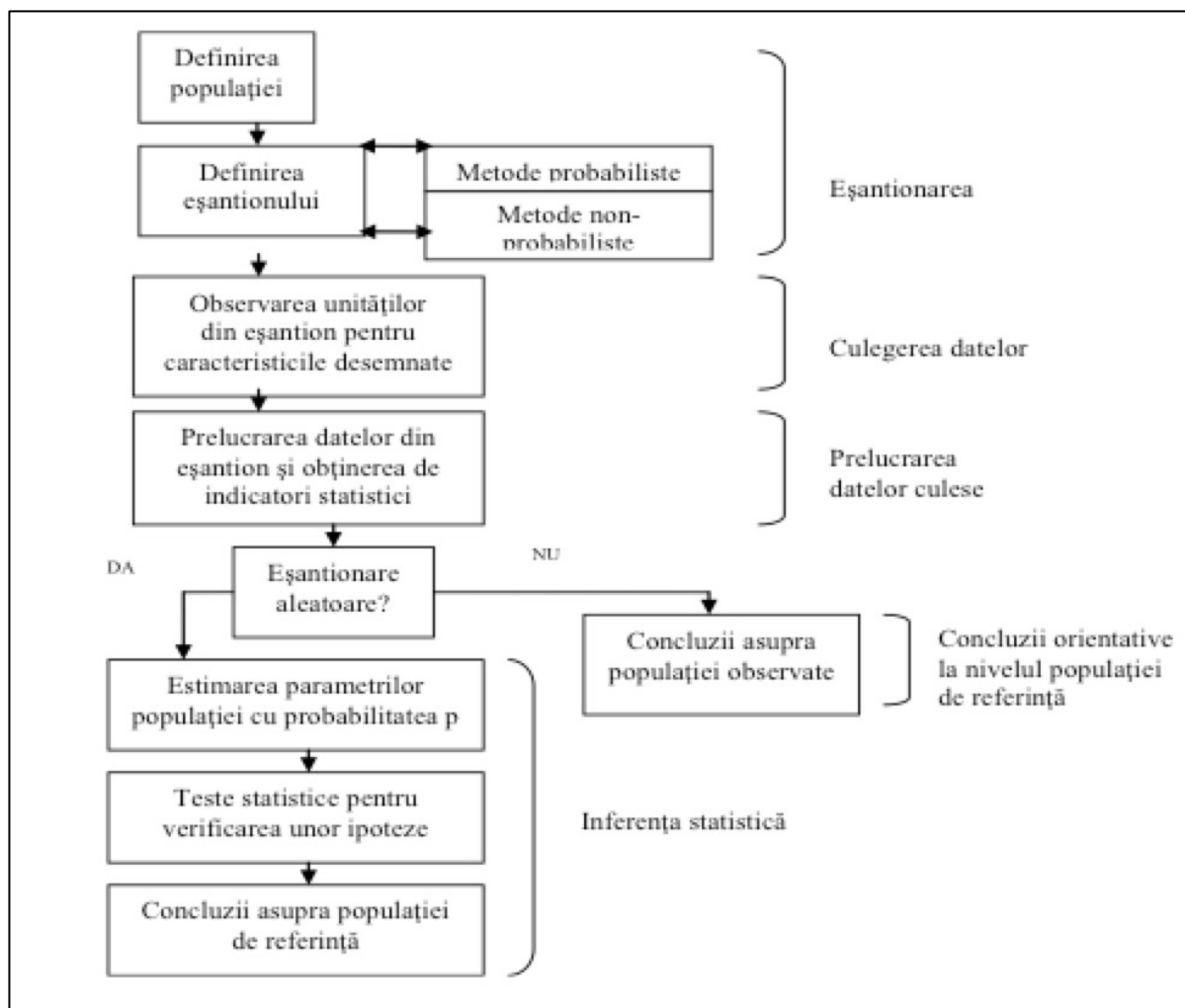


Fig. 1. Stages of research by statistical survey¹

¹ Statistică: teorie și aplicații, T. Andrei, S. Stancu, D. Traian Pele, Economică Publishing, second edition 2002;

2.2. Probabilistic Sampling Procedures

a. Simple sampling is particularly suitable for homogeneous populations consisting of simple units; The sample can be selected by applying the return process (the returning ball procedure) or the non-return process (the non-returning ball procedure). By returning, the unit selected at a given time is placed back into the population so that each population unit has the chance to be re-elected at each selection, therefore the probability is constant. By selecting without return, the selected unit is no longer returned to the population, the probability being variable, in the sense of it increasing as the sample is created.

Of the two, the sampling without return² is more appropriate as it provides the guarantee that all the units of the sample will be different, as it is necessary for valuations, and cannot determine the market value of a property using a sample of the same comparable property, but both have a major deficiency, namely that the selection is conditioned by the ability by which elements of the population are mixed and, usually, is not applied³ in the valuation.

b. Sampling with random number tables uses Yates or Kendell tables. Initially, all the elements of the population are numbered and then the sample is extracted from the tables.

c. Systematic (mechanical) sampling consists in ordering elements by a characteristic, forming a list, dividing the population into k unit groups (equal volumes). k represents the counting step and is $\frac{N}{n}$, and the starting point is $j \leq k$ (N – population volume, n – sample volume).

d. Layered sampling. The data population is divided into layers with common features, such as: residential properties - apartments, villas, houses and others; Industrial properties; Commercial properties - office spaces, shops and others; Agricultural properties - arable lands, orchards, agro-zootechnical farms and others; Specialised properties. At each layer level, simple random sampling is carried out to construct a representative sample of the population.

This type of sampling is best suited to real estate and is already used by appraisers to manually sample market data needed for valuation, for example, depending on the age of the buildings, the size of the property or the location.

The stratification ensures higher degree of comparability of comparable properties and systematic errors or inadequate check of properties can be eliminated. On the other hand, excessive stratification of data may lead to the impossibility of applying statistical data analysis.

It is performed in several stages:

1. choosing the control variable (variables) (types of real estate properties and, possibly, different characteristics thereof) used to perform the stratification,
2. defining the layers in which the population is divided,
3. choosing the number of layers according to the size and nature of the population, but not more than 7-8 per characteristic / classification criterion⁴,

² Business Statistics in Practice, Bruce L. Bowerman, Richard T. O'Connell, McGraw Hill, fourth edition, 2007- pp. 5.

³ Statistică aplicată pentru modelarea evaluării imobiliare, Iuliu Pașca, Politehnica Publishing, 2008 pp. 16

⁴ Statistica. Teorie și aplicații, Tudorel Andrei, Stelian Stancu, Daniel Traian Pele, Economică Publishing, second edition, 2002 – pp. 209

4. allocating the sample volume over the previously defined layers:

- a. either proportionally to the number of units in the same population (number of units / sample layer must be proportional to the number of units / same population layer):

$$p_i = \frac{N_i}{N} \text{ therefore, } n_i = n \cdot p_i$$

where:

N_i – number of units in the reference population, from layer i

n - sample volume

N - population volume

n_i - number of units in the sample, in layer i .

- b) either optimally (disproportionally), taking into account both the share of each layer in the total population volume, and the variance of each layer:

$$n_i = n \cdot \frac{N_i \cdot \sigma_i}{\sum_{i=1}^k N_i \cdot \sigma_i}, \text{ with } i=1, k$$

where:

σ_i – the standard deviation of the variable used for the stratification, estimated for layer i

5. The units to be included in the sample are randomly generated for each layer.

A relative sampling error, below 5%, allows us to appreciate that the sample is representative:

$$\frac{|\bar{x} - m|}{m} \cdot 100 \leq 5\%$$

where:

$\bar{x}$ - mean of the analysed sample at the sample level (market value),

m - mean of the analysed variable at the population level

2.3. Sample Volume

The automated valuation model involves performing analysis on a sample of real estate properties and generalising the results obtained (by statistical inference) for the whole studied population (the entire portfolio of bank guarantees). To this end, it is necessary to determine the appropriate size of the sample on the basis of which the analysis and the overall assessment of the bank guarantee portfolio will be performed.

Chapter 3 - The Process of Developing an Automated Valuation Model

3.1. Stages of Developing an Automated Valuation Model

This process is presented in the *Standard on Automated Valuation Models*⁵ developed by the *International Association of Assessing Officers*, a complex process that requires close collaboration between appraisers, real estate market analysts, statisticians and software developers.

Starting from the recommendations of this standard, we propose the following sequence of steps needed to build the automated valuation model:

1. identifying the subject real estate (portfolio of bank guarantees to be valued);
2. establishing the extraordinary assumptions, hypothetical and limiting conditions;
3. data management and data quality analysis;
4. sample stratification (if it was not built by stratified sampling);
5. establishing the model specifications;
6. model calibration;
7. model testing and quality assurance;
8. model validation;
9. application of the model;
10. regular verification of the model accuracy

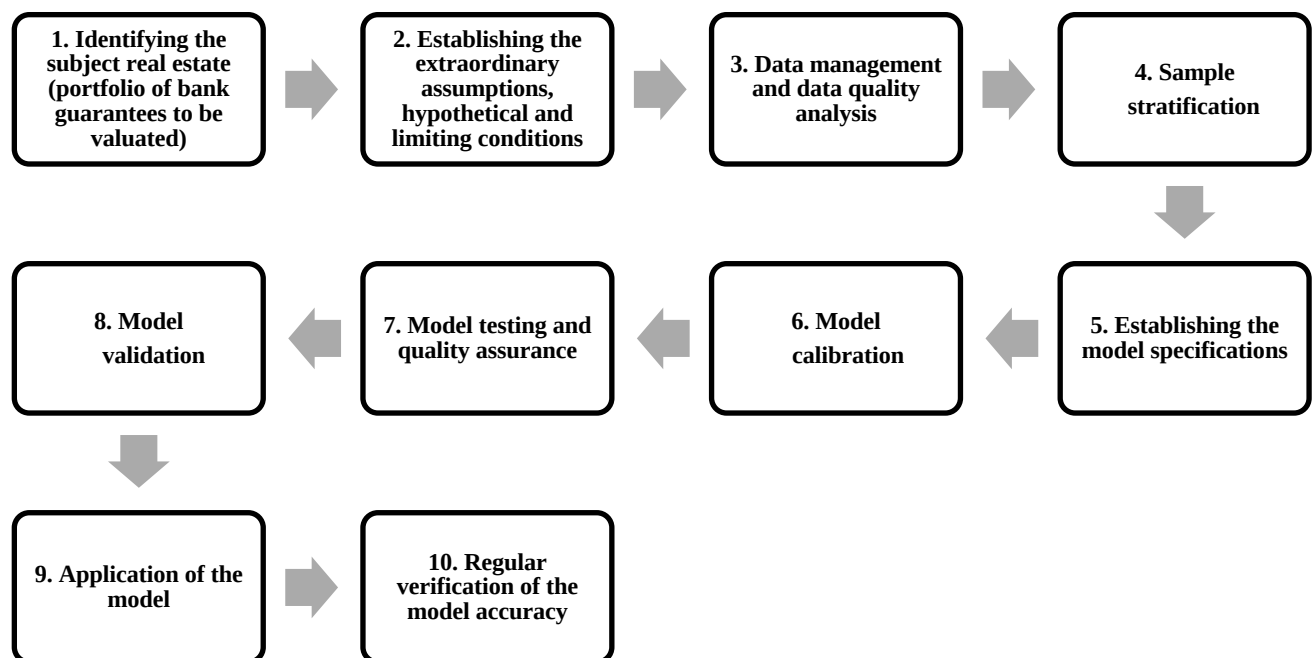


Fig. 2. Standard for automated valuation models developed by the International Association of Assessing Officers (IAAO)

⁵ Standard on Automated Valuation Models (AVMs), IAAO, 2003

3.2. Identification of the subject real estate (database - bank guarantee portfolio to be valued)

The database contains a total of 116 evaluation reports conducted in Cluj-Napoca, Cluj County (Marasti, Manastur, Gheorgheni, Grigorescu and Zorilor Districts).

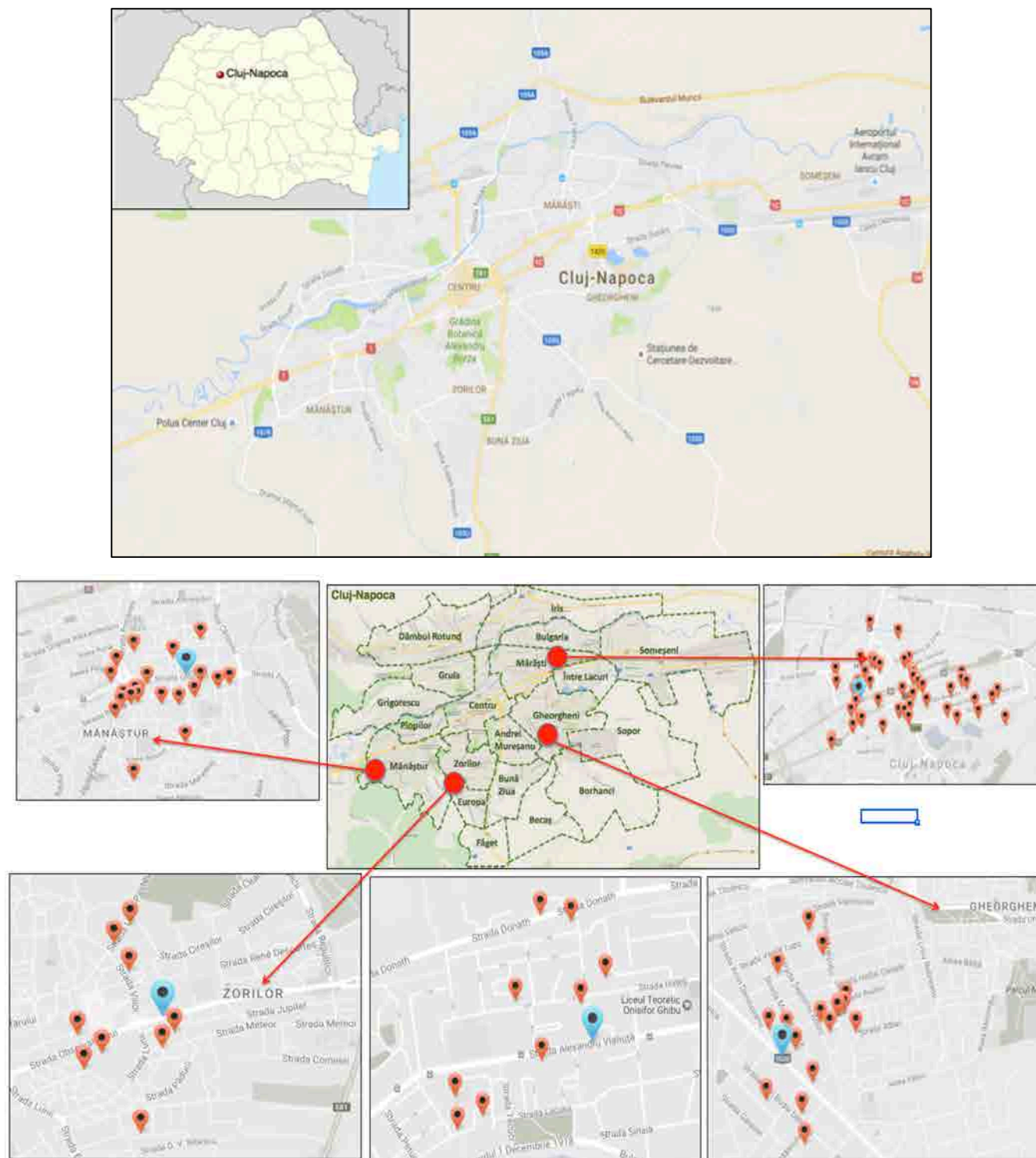


Fig. 3. Location of subject real estate - Cluj-Napoca City

*Tab. 1 Guarantees portfolio to be valuated - database
(Sampling by location)*

No.	ID	Type of real estate	Location	Area of real estate	Number of rooms	Useful floor area [sq m]	Floor (Ground/intermediate/las floor)	Finishing (low/average/high)	Transaction price (Eur)	Unit transaction price (Eur/sq m)
1	46	Apartment	Cluj-Napoca	Marasti	2	40.00	Intermediate	High	60,000	1,500.00
2	38	Apartment	Cluj-Napoca	Marasti	2	48.00	Intermediate	Average	65,000	1,354.17
3	45	Apartment	Cluj-Napoca	Marasti	2	48.00	Intermediate	Average	69,900	1,456.25
4	19	Apartment	Cluj-Napoca	Marasti	2	50.00	Intermediate	Average	68,500	1,370.00
5	30	Apartment	Cluj-Napoca	Marasti	2	50.00	Intermediate	Average	66,990	1,339.80
6	33	Apartment	Cluj-Napoca	Marasti	2	50.00	Intermediate	Average	69,000	1,380.00
7	36	Apartment	Cluj-Napoca	Marasti	2	50.00	Intermediate	Average	62,000	1,240.00
8	37	Apartment	Cluj-Napoca	Marasti	2	50.00	Intermediate	High	68,000	1,360.00
9	42	Apartment	Cluj-Napoca	Marasti	2	50.00	Last	Average	57,000	1,140.00
10	13	Apartment	Cluj-Napoca	Marasti	2	52.00	Intermediate	High	72,800	1,400.00
11	18	Apartment	Cluj-Napoca	Marasti	2	52.00	Ground	Average	71,500	1,375.00
12	26	Apartment	Cluj-Napoca	Marasti	2	52.00	Intermediate	Average	69,000	1,326.92
13	27	Apartment	Cluj-Napoca	Marasti	2	52.00	Intermediate	High	68,000	1,307.69
14	28	Apartment	Cluj-Napoca	Marasti	2	52.00	Ground	Average	71,500	1,375.00
15	31	Apartment	Cluj-Napoca	Marasti	2	52.00	Intermediate	Average	76,000	1,461.54
16	34	Apartment	Cluj-Napoca	Marasti	2	52.00	Ground	Average	66,000	1,269.23
17	39	Apartment	Cluj-Napoca	Marasti	2	52.00	Last	Average	62,000	1,192.31
18	40	Apartment	Cluj-Napoca	Marasti	2	52.00	Last	Average	61,000	1,173.08
19	41	Apartment	Cluj-Napoca	Marasti	2	52.00	Last	Average	60,000	1,153.85
20	43	Apartment	Cluj-Napoca	Marasti	2	52.00	Last	Average	63,000	1,211.54
21	44	Apartment	Cluj-Napoca	Marasti	2	52.00	Last	Low	57,000	1,096.15
22	2	Apartment	Cluj-Napoca	Marasti	2	53.00	Intermediate	Average	62,000	1,169.81
23	5	Apartment	Cluj-Napoca	Marasti	2	53.00	Intermediate	Average	61,000	1,150.94
24	8	Apartment	Cluj-Napoca	Marasti	2	53.00	Last	Average	58,000	1,094.34
25	11	Apartment	Cluj-Napoca	Marasti	2	53.00	Intermediate	High	69,900	1,318.87
26	16	Apartment	Cluj-Napoca	Marasti	2	53.00	Ground	Average	66,000	1,245.28
27	17	Apartment	Cluj-Napoca	Marasti	2	53.00	Ground	Average	68,000	1,283.02
28	20	Apartment	Cluj-Napoca	Marasti	2	53.00	Ground	Average	63,000	1,188.68
29	22	Apartment	Cluj-Napoca	Marasti	2	53.00	Ground	Average	63,000	1,188.68
30	24	Apartment	Cluj-Napoca	Marasti	2	53.00	Ground	Average	63,000	1,188.68
31	25	Apartment	Cluj-Napoca	Marasti	2	53.00	Ground	Average	65,000	1,226.42
32	1	Apartment	Cluj-Napoca	Marasti	2	54.00	Last	Average	63,000	1,166.67
33	29	Apartment	Cluj-Napoca	Marasti	2	54.00	Last	Average	65,000	1,203.70
34	32	Apartment	Cluj-Napoca	Marasti	2	54.00	Last	Average	63,000	1,166.67
35	35	Apartment	Cluj-Napoca	Marasti	2	54.00	Last	Average	65,000	1,203.70
36	3	Apartment	Cluj-Napoca	Marasti	2	56.00	Intermediate	Average	66,000	1,178.57
37	4	Apartment	Cluj-Napoca	Marasti	2	58.00	Last	Average	70,000	1,206.90
38	6	Apartment	Cluj-Napoca	Marasti	2	58.00	Intermediate	High	71,000	1,224.14
39	9	Apartment	Cluj-Napoca	Marasti	2	58.00	Last	Average	70,000	1,206.90
40	12	Apartment	Cluj-Napoca	Marasti	2	58.00	Intermediate	High	82,000	1,413.79
41	15	Apartment	Cluj-Napoca	Marasti	2	58.00	Last	Average	70,000	1,206.90
42	7	Apartment	Cluj-Napoca	Marasti	2	60.00	Ground	High	75,000	1,250.00
43	10	Apartment	Cluj-Napoca	Marasti	2	63.00	Ground	Low	75,000	1,190.48
44	21	Apartment	Cluj-Napoca	Marasti	2	65.00	Ground	Average	79,000	1,215.38
45	23	Apartment	Cluj-Napoca	Marasti	2	65.00	Ground	Average	79,000	1,215.38
45 Total apartments Cluj-Napoca, Marasti District										
46	1	Apartment	Cluj-Napoca	Manastur	2	52.00	Intermediate	High	69,500	1,336.54
47	2	Apartment	Cluj-Napoca	Manastur	2	50.00	Intermediate	High	61,500	1,230.00
48	3	Apartment	Cluj-Napoca	Manastur	2	50.00	Intermediate	Average	51,500	1,030.00
49	4	Apartment	Cluj-Napoca	Manastur	2	52.00	Ground	Average	62,000	1,192.31
50	5	Apartment	Cluj-Napoca	Manastur	2	52.00	Ground	Average	57,900	1,113.46
51	6	Apartment	Cluj-Napoca	Manastur	2	52.00	Ground	High	65,000	1,250.00
52	7	Apartment	Cluj-Napoca	Manastur	2	37.00	Intermediate	High	48,000	1,297.30
53	8	Apartment	Cluj-Napoca	Manastur	2	43.00	Intermediate	Average	48,000	1,116.28
54	9	Apartment	Cluj-Napoca	Manastur	2	39.00	Intermediate	Average	48,000	1,230.77
55	10	Apartment	Cluj-Napoca	Manastur	2	49.00	Intermediate	Average	61,800	1,261.22
56	11	Apartment	Cluj-Napoca	Manastur	2	42.00	Intermediate	Average	52,000	1,238.10
57	12	Apartment	Cluj-Napoca	Manastur	2	45.00	Ground	Average	53,300	1,184.44
58	13	Apartment	Cluj-Napoca	Manastur	2	43.00	Intermediate	Low	46,000	1,069.77

No.	ID	Type of real estate	Location	Area of real estate	Number of rooms	Useful floor area [sq m]	Floor (Ground/intermediate/las floor)	Finishing (low/average/high)	Transaction price (Eur)	Unit transaction price (Eur/sq m)
59	14	Apartment	Cluj-Napoca	Manastur	2	49.00	Intermediate	Low	50,000	1,020.41
60	15	Apartment	Cluj-Napoca	Manastur	2	45.00	Intermediate	Average	54,500	1,211.11
61	16	Apartment	Cluj-Napoca	Manastur	2	44.00	Intermediate	Average	54,000	1,227.27
62	17	Apartment	Cluj-Napoca	Manastur	2	45.00	Intermediate	Average	54,500	1,211.11
63	18	Apartment	Cluj-Napoca	Manastur	2	46.00	Last	Average	56,300	1,223.91
64	19	Apartment	Cluj-Napoca	Manastur	2	47.00	Intermediate	Average	55,000	1,170.21
65	20	Apartment	Cluj-Napoca	Manastur	2	47.00	Ground	Average	53,000	1,127.66
66	21	Apartment	Cluj-Napoca	Manastur	2	36.00	Last	Average	44,000	1,222.22
67	22	Apartment	Cluj-Napoca	Manastur	2	51.47	Last	Average	54,000	1,049.15
66	Total apartments Cluj-Napoca, Manastur District									
68	1	Apartment	Cluj-Napoca	Gheorgheni	2	40.00	Intermediate	Average	52,900	1,322.50
69	2	Apartment	Cluj-Napoca	Gheorgheni	2	38.00	Last	Average	51,700	1,360.53
70	3	Apartment	Cluj-Napoca	Gheorgheni	2	38.00	Last	Average	48,900	1,286.84
71	4	Apartment	Cluj-Napoca	Gheorgheni	2	57.00	Intermediate	High	81,000	1,421.05
72	5	Apartment	Cluj-Napoca	Gheorgheni	2	63.00	Intermediate	Average	77,000	1,222.22
73	6	Apartment	Cluj-Napoca	Gheorgheni	2	63.00	Intermediate	Average	70,000	1,111.11
74	7	Apartment	Cluj-Napoca	Gheorgheni	2	50.00	Last	Average	62,000	1,240.00
75	8	Apartment	Cluj-Napoca	Gheorgheni	2	38.00	Intermediate	High	53,000	1,394.74
76	9	Apartment	Cluj-Napoca	Gheorgheni	2	30.00	Intermediate	Average	42,000	1,400.00
77	10	Apartment	Cluj-Napoca	Gheorgheni	2	40.00	Last	High	54,000	1,350.00
78	11	Apartment	Cluj-Napoca	Gheorgheni	2	48.00	Ground	Average	71,000	1,479.17
79	12	Apartment	Cluj-Napoca	Gheorgheni	2	30.00	Intermediate	Average	40,000	1,333.33
80	13	Apartment	Cluj-Napoca	Gheorgheni	2	30.00	Intermediate	Average	40,800	1,360.00
81	14	Apartment	Cluj-Napoca	Gheorgheni	2	29.00	Intermediate	Average	38,000	1,310.34
82	15	Apartment	Cluj-Napoca	Gheorgheni	2	30.00	Intermediate	Average	43,000	1,433.33
83	16	Apartment	Cluj-Napoca	Gheorgheni	2	33.00	Intermediate	High	49,000	1,484.85
84	17	Apartment	Cluj-Napoca	Gheorgheni	2	31.00	Intermediate	Average	45,000	1,451.61
85	18	Apartment	Cluj-Napoca	Gheorgheni	2	32.00	Intermediate	High	50,500	1,578.13
86	19	Apartment	Cluj-Napoca	Gheorgheni	2	32.00	Last	Average	36,000	1,125.00
87	20	Apartment	Cluj-Napoca	Gheorgheni	2	32.00	Last	Average	36,000	1,125.00
88	21	Apartment	Cluj-Napoca	Gheorgheni	2	32.00	Ground	Average	40,900	1,278.13
88	Total apartments Cluj-Napoca, Gheorgheni District									
89	1	Apartment	Cluj-Napoca	Grigorescu	2	50.00	Intermediate	Average	65,000	1,300.00
90	2	Apartment	Cluj-Napoca	Grigorescu	2	52.00	Intermediate	Average	67,900	1,305.77
91	3	Apartment	Cluj-Napoca	Grigorescu	2	52.00	Intermediate	Average	67,900	1,305.77
92	4	Apartment	Cluj-Napoca	Grigorescu	2	43.00	Last	Average	54,300	1,262.79
93	5	Apartment	Cluj-Napoca	Grigorescu	2	44.00	Last	High	61,400	1,395.45
94	6	Apartment	Cluj-Napoca	Grigorescu	2	45.00	Last	Average	54,000	1,200.00
95	7	Apartment	Cluj-Napoca	Grigorescu	2	48.00	Last	Average	65,000	1,354.17
96	8	Apartment	Cluj-Napoca	Grigorescu	2	47.00	Last	Average	52,000	1,106.38
97	9	Apartment	Cluj-Napoca	Grigorescu	2	44.00	Intermediate	Average	57,500	1,306.82
98	10	Apartment	Cluj-Napoca	Grigorescu	2	53.00	Intermediate	Average	66,500	1,254.72
99	11	Apartment	Cluj-Napoca	Grigorescu	2	52.00	Last	Low	54,000	1,038.46
100	12	Apartment	Cluj-Napoca	Grigorescu	2	50.00	Ground	Average	57,000	1,140.00
101	13	Apartment	Cluj-Napoca	Grigorescu	2	50.00	Ground	Average	55,000	1,100.00
102	14	Apartment	Cluj-Napoca	Grigorescu	2	52.00	Intermediate	Average	61,000	1,173.08
103	15	Apartment	Cluj-Napoca	Grigorescu	2	50.00	Intermediate	Average	63,000	1,260.00
104	16	Apartment	Cluj-Napoca	Grigorescu	2	44.00	Intermediate	Average	53,500	1,215.91
105	17	Apartment	Cluj-Napoca	Grigorescu	2	46.00	Intermediate	Average	64,000	1,391.30
106	18	Apartment	Cluj-Napoca	Grigorescu	2	45.00	Intermediate	Average	60,000	1,333.33
107	19	Apartment	Cluj-Napoca	Grigorescu	2	48.00	Ground	Average	61,500	1,281.25
107	Total apartments Cluj-Napoca, Grigorescu District									
108	1	Apartment	Cluj-Napoca	Zorilor	2	54.00	Intermediate	High	76,000	1,407.41
109	2	Apartment	Cluj-Napoca	Zorilor	2	54.00	Intermediate	Average	66,000	1,222.22
110	3	Apartment	Cluj-Napoca	Zorilor	2	50.00	Intermediate	Average	66,000	1,320.00
111	4	Apartment	Cluj-Napoca	Zorilor	2	60.00	Intermediate	Average	86,700	1,445.00
112	5	Apartment	Cluj-Napoca	Zorilor	2	50.00	Last	Average	66,000	1,320.00
113	6	Apartment	Cluj-Napoca	Zorilor	2	50.00	Ground	Average	71,000	1,420.00
114	7	Apartment	Cluj-Napoca	Zorilor	2	48.00	Intermediate	Average	65,000	1,354.17
115	8	Apartment	Cluj-Napoca	Zorilor	2	48.00	Intermediate	Average	67,000	1,395.83
116	9	Apartment	Cluj-Napoca	Zorilor	2	44.00	Ground	High	65,000	1,477.27
116	Total apartments Cluj-Napoca, Zorilor District									

3.3. Establishing Assumptions in Valuation

The database consists of valuation reports with the purpose of assessing the loan guarantee during July-November of the current year, and comprises 2-room apartments with areas ranging between 40 and 65 sq m.

For the realization of the database it was taken into account that the dependent variable is the market value of the apartments, expressed in Eur and independent variables, the useful floor area of the real estate, the number of rooms, the floor and their finishes, according to the table below:

Tab. 1 Short description of database variables

Tip proprietate imobiliara	Apartment
Zona proprietatii imobiliare	Cluj-Napoca, Jud. Cluj
Numar camere	2
Suprafata utila [mp]	40-65mp
Etaj	parter/intermediar/ultimul
Finisaje	inferioare/medii/superioare
Pret de tranzactie	Exprimat in Eur
Pret de tranzactie unitar	Exprimat in Eur/mp

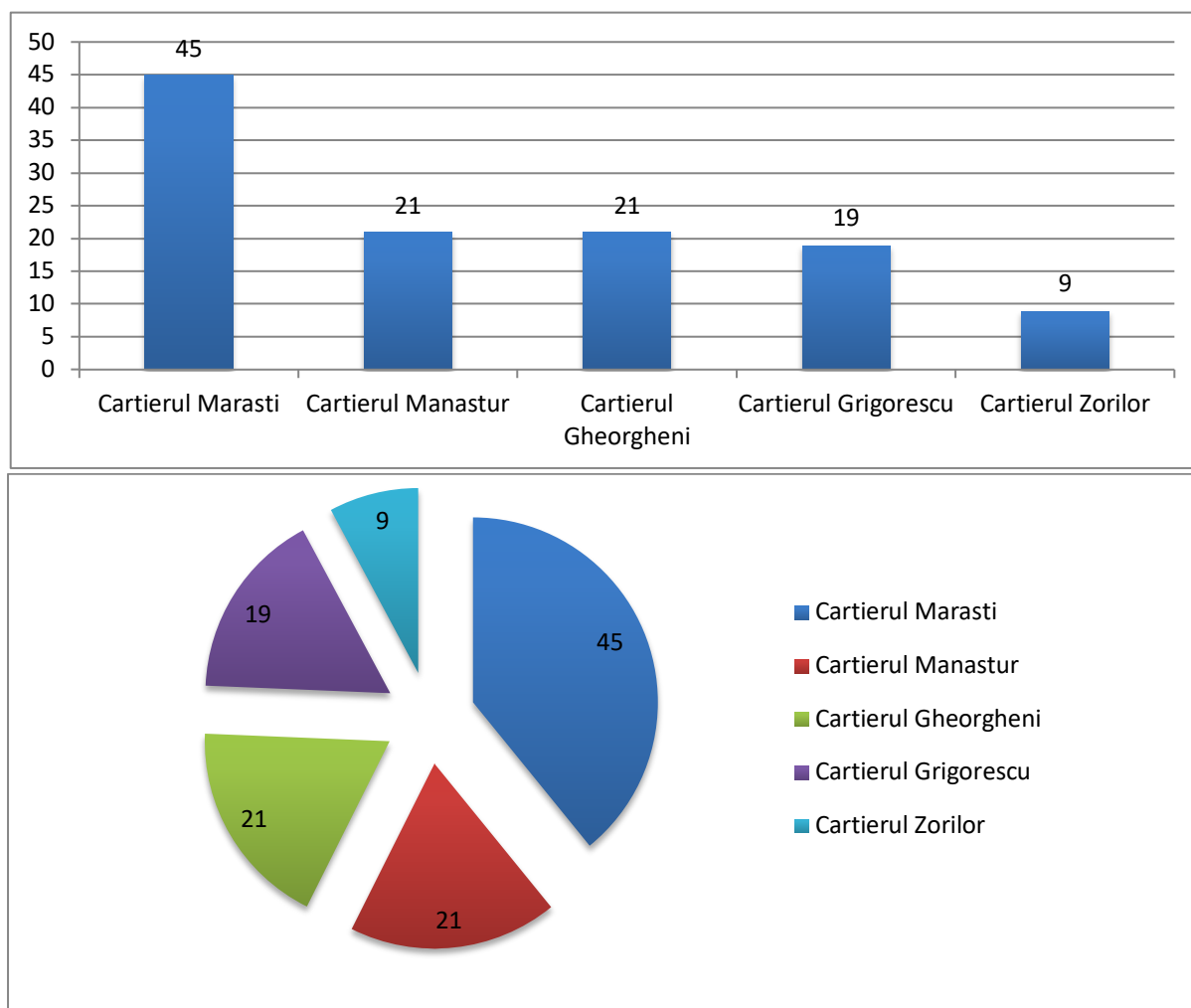


Fig. 4. Distribution of subject real estate properties by location

3.4. Data Management and Data Quality Analysis

The market analyst must use statistical tools to systematically and iteratively "scan" market data to determine the shortcomings, errors, inconsistencies, extreme values (aberrant values) in the data series analysed (data quality analysis).

In support of this stage, statistical analysis software provides graphical techniques and quantitative techniques to explore the data series. These techniques can be used to establish the representativeness of the sample to identify the frequency distribution function (normal, uniform, χ^2 , lognormal, Student, Fisher, Poisson, etc.) and its characteristics, or to detect the aberrant values of the analysed series.

Statistical Indicators

Three statistical indicators are the **mean, median, and mode**. These indicators are used to identify the characteristic variable in a population or sample. Measures that refer to a population are called parameters.

The *mean*, which is commonly called simple arithmetic mean, is by far the most used parameter. It is obtained by dividing the sum of all variables in a population by the number of variables.

An indicator of the key trend is the *median*, it divides the values of the observations forming a sample of data in equal intervals. To identify the median, the values of the observations are arranged in ascending order, as in the example below. If the total number of observations is odd, the median is represented by the value of the central position (in the middle of the data series). If the total number of observations is par, as in the example below, the median is the arithmetic mean of the value of the two observations in the middle of the data series (median).

The *mode* is the value of the observations most frequently appearing in the analysed data sample.

Dispersion Sizes

Dispersion sizes show which variation occurs in a variable. These are useful as they can be compared to the characteristics of a known distribution - such as normal distribution - in order to determine if a particular set of inferential statistics can be used⁶ on the parameters.⁷ Moreover, it facilitates the comparison of two sets of data to determine which one varies more.

Standard Deviation and Variance

The two fundamental sizes describing the dispersion - the standard deviation (or the mean square deviation) and the variance (or dispersion) - take into account the type of data distribution. In addition, standard deviations can further run statistical analyses, allowing inference and the deduction of conclusions about the degree of uncertainty associated with that inference. For this reason, standard deviation is a statistic of the frequently calculated and reported sample.

⁶ Deductible;

⁷ For example, if the data set has a distribution that is sufficiently close to the normal distribution, then statistical methods based on the normal distribution can be used to make inferences about the parameters of the population to which the sample belongs.

To determine the degree of dispersion of sample data, the mean square deviation can be calculated. By reporting the mean square deviation to the sample mean, the coefficient of variation characterising the data series is calculated.

Its level may indicate the degree of difficulty of the forecast to be made on the basis of the data in the sample. Thus, if the coefficient of variation is:

- less than 0.5 it can be considered that the economic indicator analysed is easy to predict;
- ranging from 0.5 to 1, it can be considered that the forecast of the analysed economic indicator has an average degree of difficulty;
- higher than 1, the series of data characterising the economic indicator analysed has high variability, which implies a high degree of difficulty of the forecast. A high coefficient of variation represents, in some cases the signal of the existence of extreme values that can be recorded by some observations in the sample. In these situations, it is necessary to identify them and the causes that could cause these extreme values. Given that these observations are not similar (in terms of the determinants of their value) to the other in the sample, it is advisable to remove them from the sample.

Database Analysis - Analysis of the Subject Property Portfolio

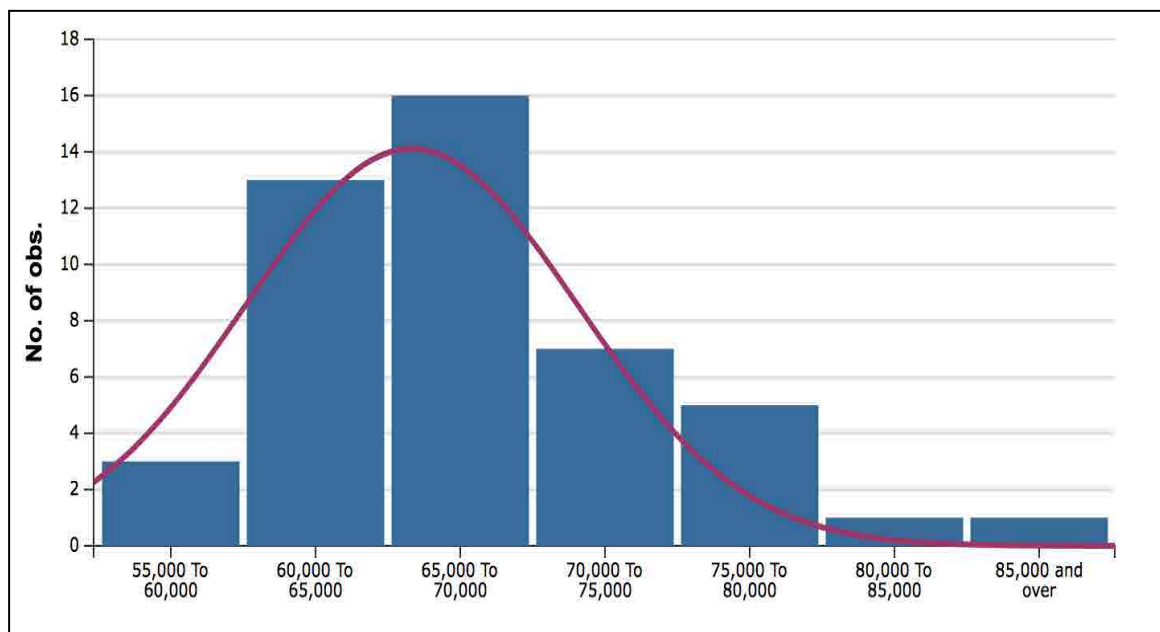


Fig. 5. Frequency of property distribution based on the estimated market value

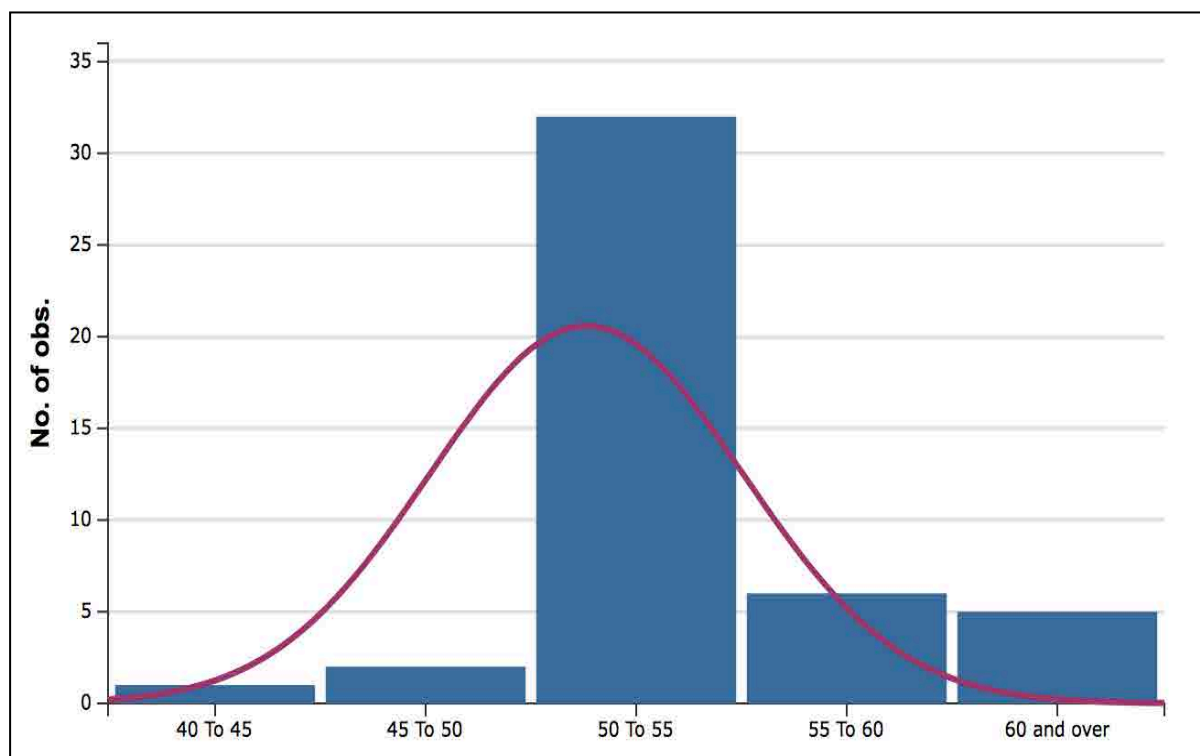


Fig. 6. Frequency of property distribution depending on the surface of the subject properties

Count	46	Mean Deviation	5,044.06427
Mean	67,458.47826	Second Moment	41,316,104.20605
Mean LCL	65,528.57658	Third Moment	2.1561E+11
Mean UCL	69,388.37995	Fourth Moment	6.2E+15
Variance	42,234,239.85507		
Standard Deviation	6,498.78757	Sum	3,103,090
Mean Standard Error	958.19395	Sum Standard Error	44,076.92178
Coefficient of Variation	0.09634	Total Sum Squares	2.1E+11
		Adjusted Sum Squares	1,900,540,793.47826
Minimum	57,000		
Maximum	87,000	Geometric Mean	67,164.20574
Range	30,000	Harmonic Mean	66,880.93296
		Mode	63,000
Median	66,495		
Median Error	177.06570	Skewness	0.81187
Percentile 25% (Q1)	63,000	Skewness Standard Error	0.34240
Percentile 75% (Q3)	70,000	Kurtosis	3.64642
IQR	7,000	Kurtosis Standard Error	0.64247
MAD (Median Absolute Deviation)	8,000	Skewness (Fisher's)	0.83950
Coefficient of Dispersion (COD)	0.07555	Kurtosis (Fisher's)	0.86531

Fig. 7. Statistical data quality analysis - 'market value' dependent variable

Count	46	Mean Deviation	3.09546
Mean	53.58696	Second Moment	19.41635
Mean LCL	52.26396	Third Moment	34.67334
Mean UCL	54.90995	Fourth Moment	1,830.36039
Variance	19.84783		
Standard Deviation	4.45509	Sum	2,465
Mean Standard Error	0.65687	Sum Standard Error	30.21589
Coefficient of Variation	0.08314	Total Sum Squares	132.985
		Adjusted Sum Squares	893.15217
Minimum	40		
Maximum	65	Geometric Mean	53.40705
Range	25	Harmonic Mean	53.22664
		Mode	52
Median	53		
Median Error	0.12138	Skewness	0.40527
Percentile 25% (Q1)	52	Skewness Standard Error	0.34240
Percentile 75% (Q3)	54	Kurtosis	4.85514
IQR	2	Kurtosis Standard Error	0.64247
MAD (Median Absolute Deviation)	6	Skewness (Fisher's)	0.41906
Coefficient of Dispersion (COD)	0.05455	Kurtosis (Fisher's)	2.21650

Fig. 8.

Statistical analysis of data quality - the independent variable 'usable floor area of apartments'

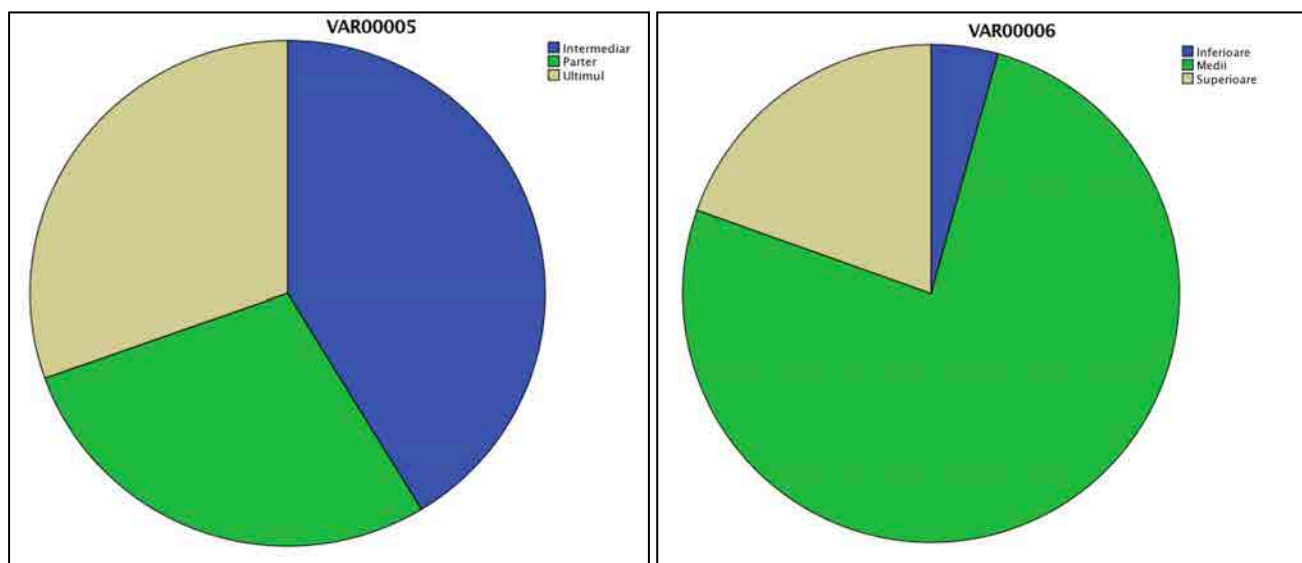


Fig. 9. Statistical analysis of real estate properties according to the qualitative variables - finishing and depending on the floor

Stratification

It is the process by which the real estate forming a heterogeneous sample are grouped by type of comparable real estate according to their physical and economic characteristics in order to minimise the differences within a layer and maximise the differences between layers.

Following this step, the mean of the market values of the properties can be estimated at the level of each layer created within the analysed sample in order to determine the index of increase / decrease of the market value of the bank guarantee portfolio. For the estimation of this index, we propose the following formula (derived from Paasche-type weighted aggregate calculation method):

$$Indice = \frac{\sum_{i=1}^m p_{i,t} \cdot \mu_{i,t}}{\sum_{i=1}^m p_{i,t} \cdot \mu_{i,t-1}} \cdot 100$$

where:

$p_{i,t}$ = share of each layer in the analysed sample (see Stratified Sampling);

μ = mean of property market values in each layer;

m = number of layers in the analysed sample;

t = current time;

$t-1$ = time of the last reassessment.

This index helps identify the possible diminution of the value of the bank guarantee portfolio, in which case it is necessary to re-evaluate it with the help of an AVM, and consequently, the following steps are undertaken.

3.5. Establishing Model Specification

This stage refers to the process of designing the optimal structure of AVM and consists in selecting the appropriate variables and defining its structure. The model is designed for each individual layer (sub-sample of comparable properties).

The appraiser may select from the multitude of variables that characterise the real estate analysed (potential explanatory variables), those which could have an economic impact on the real estate value (dependent variable). The higher the number of explanatory variables selected, the higher the risk that the information contained therein becomes redundant. On the other hand, the inclusion of a too low number of explanatory variables could cause the regression model not to explain enough the true evolution of the dependent variable, which would lead to a bias of estimators (regression coefficients).

The choice of explanatory variables must be made by the appraiser, not only in terms of size of the correlation coefficients with the dependent variable, but also by taking into consideration the economic significance that could characterise the relationship between the explanatory variable and the dependent variable.

Moreover, it examines whether there is a strong correlation between the selected explanatory variables. Where such a correlation is identified, the analysis will include only one of the explanatory variables in the interrelated group. One can opt for the strongest correlated with the

dependent variable. The more the correlation coefficient value matches 1 or -1, the stronger the variables are correlated.

The variables can be quantitative (or scaling), expressed numerically, measured on a scale and are subject to arithmetical operations, or can be qualitative (or categorical) variables that are not numerically expressed (not measured on a scale) but by attributes, but can be encoded by category, and can be of the following type:

- nominal, no internal order
- ordinal, with a certain internal non-quantifiable order, for example: the state of the building can be "very good" (coded with 0), "good" (coded 1), "satisfactory" (coded 2) or "unsatisfactory" (coded with 3).

It is very important to correctly identify and define all the model variables and then to encode (transform) the qualitative variables accordingly.

With regard to the structure of the model, it can be additive (linear), multiplicative or hybrid (nonlinear) (review *Standard on Automated Valuation Models (AVMs)*, IAAO, 2003).

In the additive structure, the contribution of the explanatory variables is summed:

$$Y_i = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \dots$$

where:

Y_i - market value of the subject property - the dependent variable;

β_0 - model constant (or the free term);

x_i - explanatory variables;

β_i - coefficients of the explanatory variables.

In the multiplicative model, the contribution of explanatory variables is multiplied:

$$Y_i = \beta_0 \cdot x_1^{\beta_1} \cdot x_2^{\beta_2} \cdot \dots$$

This type of model is more difficult to calibrate because the variables have to be converted to logarithmic format, but the advantage is that it allows corrections to be applied pro rata to the value of the subject property.

The hybrid (nonlinear) model is a combination of the additive and the multiplicative model:

$$Y_i = \pi QG \cdot (\pi QC \cdot \Sigma AC + \pi QT \cdot \Sigma AT + \Sigma A)$$

where:

πQG - product of the general qualitative variables (applies to the property as a whole);

πQC - product of the qualitative variables for the building;

ΣAC - sum of additive variables for the building;

πQT - product of the qualitative variables for the land;

ΣAT - sum of additive variables for the land;

ΣA - sum of other variables (additional variables).

It also analyses whether there is a strong correlation between the selected explanatory variables. If such a correlation is identified, only one of the explanatory variables in the group of those correlated should be retained in the analysis. The more strongly correlated with the dependent variable can be chosen. The closer the value of the correlation coefficient to the values 1 and -1, the more closely the variables are correlated.

Correlations			
		Pretvanzare	Suprafata
Pretvanzare	Pearson Correlation	1	.615**
	Sig. (2-tailed)		.000
	N	46	46
Suprafata	Pearson Correlation	.615**	1
	Sig. (2-tailed)	.000	
	N	46	46

** . Correlation is significant at the 0.01 level (2-tailed).

Fig. 10. Calculation of correlation coefficient

The correlation coefficient was calculated in order to determine the correlation between the sales price dependent variable and the usable floor area variable. If its value is close to zero – the two variables show a very weak linear dependence, if it is zero - the variables are independent, if it is close to 1 - variables have a direct linear dependence (positive) and if it is close to -1 - inverse linear dependence (negative).

3.6. Model Calibration

Among the methods of calibration, the most used are those based on statistical methods such as multiple linear regression and nonlinear regression.

The term of regression is used to describe any process in which one or more known variables (characteristics of the analysed real estate) are used to determine an unknown variable (in this case the market value of the real estate appraised) or to predict its value.

Simple Linear Regression

Simple linear regression is used to predict a dependent variable using a single explanatory variable.

The simple linear regression model has the form:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = \overline{1, n}$$

where:

β_0 – constant, i.e. the value of y when x=0;

β_1 – slope of regression line;

ε_i – the residual term (error) quantifies the incidental influence of other variables on y_i and represents the difference between the value observed (y_i) and the estimated value ($\hat{y}_i$) of the dependent variable:

$$\varepsilon_i = y_i - \hat{y}_i$$

The slope of the regression line indicates the mean value at which y changes to a change of a unit of x . If the slope is positive, y increases with increasing the value of x , and if it is negative, y decreases as the value of x increases. The coefficients of the model, β_0 and β_1 , as well as the residual terms ε_i can be determined by various methods, including the smallest square method.

For the simple linear regression analysis, the following steps are performed:

1. estimating the model parameters;
2. testing and validating the regression model;
3. performing predictions.

In order to validate the linear regression model, a series of hypotheses must be observed (*Gauss-Markov*):

- the variance of residual terms is constant;
- the residual terms ε_i have a mean of zero;
- the residual terms ε_i are not correlated with each other;
- the residual terms ε_i are not correlated with the explanatory variable x_i ;
- the residual terms ε_i normally distributed.

In this respect, it is necessary to perform the model test in order to ensure its quality (see step 7).

Multiple Linear Regression

Multiple linear regression expresses the relationship between a dependent variable and at least two explanatory variables, and its pattern is:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_m x_{mi} + \varepsilon_i, \quad i = \overline{1, n}$$

where:

- β_0 – free term;
- $\beta_1, \beta_2, \dots, \beta_m$ – multiple regression coefficients;
- ε_i – residual term.

The steps and hypotheses taken into account to achieve the multiple regression model are the same as those specified for the simple regression.

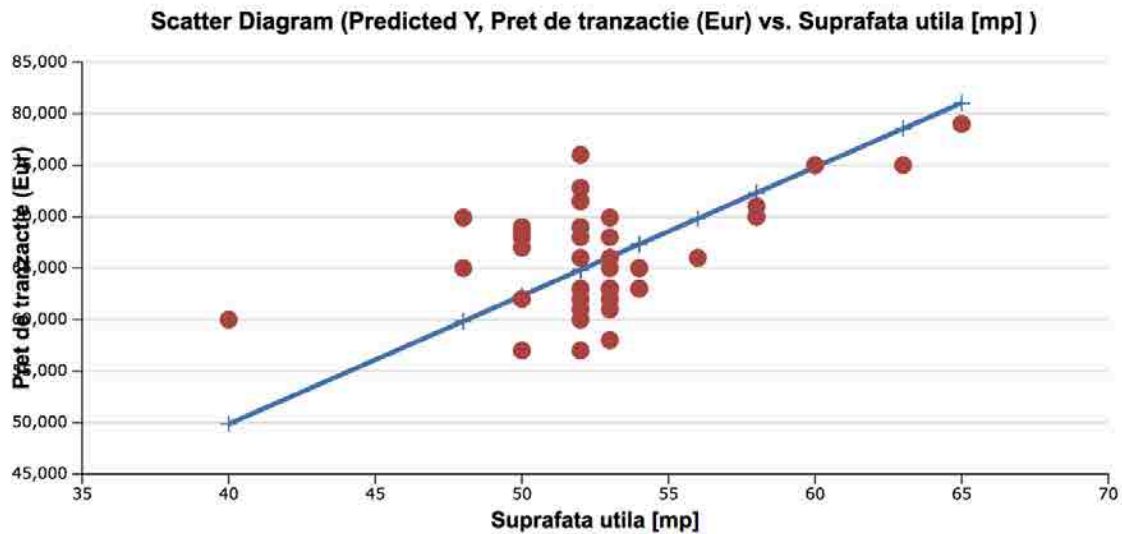


Fig. 11. Scatter diagram - price / area

Regression Statistics

<i>R</i>	0.99728
<i>R-square</i>	0.99457
<i>Adjusted R-square</i>	0.99457
<i>S</i>	4,985.34531
<i>N</i>	44

$$\text{Pret de tranzactie (Eur)} = 1,246,69698 * \text{Suprafata utila [mp]}$$

Linear regression coefficient *r* shows that there is a strong dependence between the two variables, i.e. between the transaction price and the useful floor area of the apartments. *R* square determination coefficient is 99%. The standard error indicates that the market value predicted by the model established may be different by EUR 4,985.34 against the mean of apartments sold.

3.7. Model Testing and Quality Assurance

Model testing is intended to determine whether it ensures the accuracy and integrity necessary for the estimates so achieved to be extrapolated by statistical inference to the entire population of a bank guarantee. For this purpose, a set of data on the real estate is used, for which the sale prices or market values are known, based on which the testing shall be performed. This set of data, used to test the model, was the basis for the execution and calibration of the model and represents 80% of the sample volume (*n*).

As stated above, there are a variety of statistical tests that can be applied to obtain the necessary information, some of them parametric, applicable to the sets of data with normal frequency distribution, such as Student (*t*), Fisher (*F*) Testing and other non-parametric for series that do not have normal frequency distribution, such as Chi-square, Shapiro-Wilk, Kolmogorov-Smirnov,

Man-Withney, Durbin-Watson tests.

In practice, the model specification and the calibration stage is performed iteratively: model specification - testing model specification by calibration - model specification correction - testing the specifications corrected by repeating these operations until obtaining the confidence level designed.

Model Validation

Once the model was calibrated and tested, its validation is performed based on the set of data representing the remaining 20% of the sample analysed, using statistical tests.

Applying the Model

The automated valuation model takes into account, as values of the explanatory variables, the values corresponding to the characteristics of the subject real estate and their market value is estimated as such.

3.8. Periodic Verification of the Model Accuracy

It is recommended to check the model accuracy by comparing the estimated market values to those estimated using the valuation methods and techniques recommended by IVS. The large differences existing between the two estimated values are a signal of the need to update both the analysis sample and the automated valuation model.

Chapter 4. Objectives and Proposals

Global valuation techniques for property tax purposes existed long before the development of automated valuation models (AVMs) and are currently considered as a subset thereof. Overall rating models have been developed by assessors for property tax, to improve productivity and equity in non-rural locations, where human power was insufficient to determine the function of estimating the defined value.

Additionally, in the early years of developing the global assessment techniques, tax assessors were in an advantageous position in which they could work with large amounts of data that had been converted electronically. They continue to use AVMs as a way to automate valuation and use large series of digitally encoded data from their own databases. Internet access to trusted data from tax authorities and other sources has granted assessors access to consistent data resources needed in statistical analysis.

Initial research on AVMs focused on **neural networks and expert systems** instead of regression-based models. Neuronal networks 'learn' the relationships between variables to develop and improve their internal algorithm, unknown by price estimation. They do not have a theoretical basis for algorithm development and can only be tested by comparing the estimated results with a known standard value. Because of the black box decision model, they have not developed many practical applications. Expert systems create decision models that try to mimic the behaviour of the expert (for example, the assessor).

Essentially, they automate the human process of problem solving. For example, some AVMs involve expert system components to select comparable properties or leased properties.

Expert systems are a branch of artificial intelligence that uses specialised knowledge to solve a problem at the level of a human expert. All definitions of Expert Systems highlight a specific feature, namely that they use information extracted from human experience and can, therefore, provide decision-making at the competence level corresponding to the information received and the methods of reasoning implemented. An expert system is not called a program but a system as it incorporates many different technologies such as knowledge base, inference mechanisms, explanation facilities, etc.

Bibliography

1. Appraisal Institute, 2011. *Evaluarea proprietății imobiliare*. Ediția a 13-a ed. București: s.n.
2. James T. McClave, P. G. B. T. S., 2005. *Statistics for Business and Economics*. Ed. a 9-a ed. Upper Saddle River: Prentice-Hall.
3. Paunescu V., Calin M., Manea R. Moscovici A., Salagean T. 2017. «GIS in Active Monitoring of Green Spaces.» Modern Tehnologies for the 3rd Millenium 73-78.
4. Arcgis, Survey123. 2017. <https://survey123.arcgis.com>. <https://survey123.arcgis.com>. 12 de 12. Last access: 12 de 12 de 2017. <https://survey123.arcgis.com>
5. Droj G., Droj. L. 2016. «GIS based automated valuation models – A GENUINE SOLUTION FOR REAL ESTATE VALUATION IN ROMANIA.» RevCAD “1 Decembrie 1918” University of Alba Iulia 20: 45-52.
6. Mark R. Linné, MAI, SRA with Michelle M. Thompson,. 2010. *Visual Valuation*. The Appraisal Institute.
7. Kohler, H., 2002. *Statistics for Business and Economics*. Ed. a 4-a ed. Stamford: Thomson.
8. Appraisal Institute, 2004. *Evaluarea proprietății imobiliare*. Ed a 2-a canadiana ed. Bucurști: s.n.
9. ANEVAR GEV 520 - Asociația Națională a Evaluatorilor Autorizați din România, 2015. *Standardele de evaluare a bunurilor - Evaluarea globală a portofoliilor de garanții bancare*. București: ANEVAR.
10. T. Andrei, S. S. D. T. P., 2002. *Statistică: teorie și aplicații*. Ediția a 2-a ed. s.l.:Editura Economică.
11. Bruce L. Bowerman, R. T. O. M. H., 2007. *Business Statistics in Practice*. 4th edition ed. s.l.:s.n.
12. Pasca, I., 2008. *Statistică aplicată pentru modelarea evaluării imobiliare*. s.l.:Editura Politehnică
13. Wikipedia, 2017. *Wikipedia*. [Interactiv]
Available at: <https://ro.wikipedia.org/wiki/Cluj-Napoca>
[Accesat 01 07 2017].
14. Maps, G., 2017. *Google Maps*. [Interactiv]
Available at: <https://www.google.ro/maps/place/Cluj-Napoca>
[Accesat 2017].
15. Cluj-Napoca, P., 2017. *Primăria Cluj-Napoca*. [Interactiv]
Available at: <http://www.primariaclužnapoca.ro>
[Accesat 01 07 2017].
16. EPI, A. i., 2017. *Analize imobiliare EPI*. [Interactiv]
Available at: <http://acp.imobiliare.ro/epi>
[Accesat 01 07 2017].
17. Șchiopu, C., 2007. *Inspecția proprietăților în scopul evaluării*. 3 ed. București: Iroval.
18. Badea, A.-C., 2013. *Cadastru, Bănci de date și aplicații GIS în zone urbane*. București: Conspress.