Determining the Flat Sales Prices by Flat Characteristics
Using Bayesian Network Models

Volkan Sevinç

- The need for residence is an indispensable need for people.
- Residence prices have always been a topic of interest in daily life and in various branches of science.
- Certain properties of a residence play a decisive role in the formation of it sales price.
- Thus, researchers have been trying to build various models to estimate the sales price of a residence based on its various aspects.

- Hedonic price models, which are based on the consumer theory suggested by Lancaster in 1966, have been being used widely for this purpose for a long time.

- Hedonic pricing identifies the factors and the characteristics affecting the price of an item. Thus, hedonic price models estimate the sales price of a residence considering both its internal characteristics and some external factors that are likely to affect its price.

- However, hedonic price models assume that buyers have the same amount of desire towards the same attributes of a property at the same level, which is not very possible.

- Bayesian networks however have some advantages in determining the effects of the variables on a certain feature.

- They have also some advantages over the regression models like
- hedonic pricing.

- For example, in regression models, regressors, which are used to estimate the dependent variable, are called independent variables, which are expected to be uncorrelated, and possible dependencies among them are ignored.

- Any change in the value of an independent variable affects only the dependent variable and not any other independent variables in the model.

- This way of modeling the sales price of a residence based on its various attributes may not be very realistic.

- For example, when the number of bathrooms increases in a residence, it directly causes an increase in the total area of the residence.

- Bayesian network models, on the other hand, consider the dependency relations among all the variables in the model unless some restrictions are imposed.

- As the variables exist in a network structure they all can influence each other at various strengths.

- Another disadvantage of the regression models is that they are not as flexible as Bayesian network models. That is, regression models are additive models taking the sum of the independent variables and equalizing it to the dependent variable.

- After fitting a regression model, adding some extra independent variables that are found to be significant, may dramatically change the structure and even the accuracy of the model.

- Bayesian models, however, as they mainly consider the interdependencies among the variables rather than the sum of them, are not that highly affected by new variables being added into the model.

- Bayesian networks also do not suffer from overparameterization problem very much like regression models do.

- Another advantage of Bayesian network models is that they can handle missing data more easily than regression models.

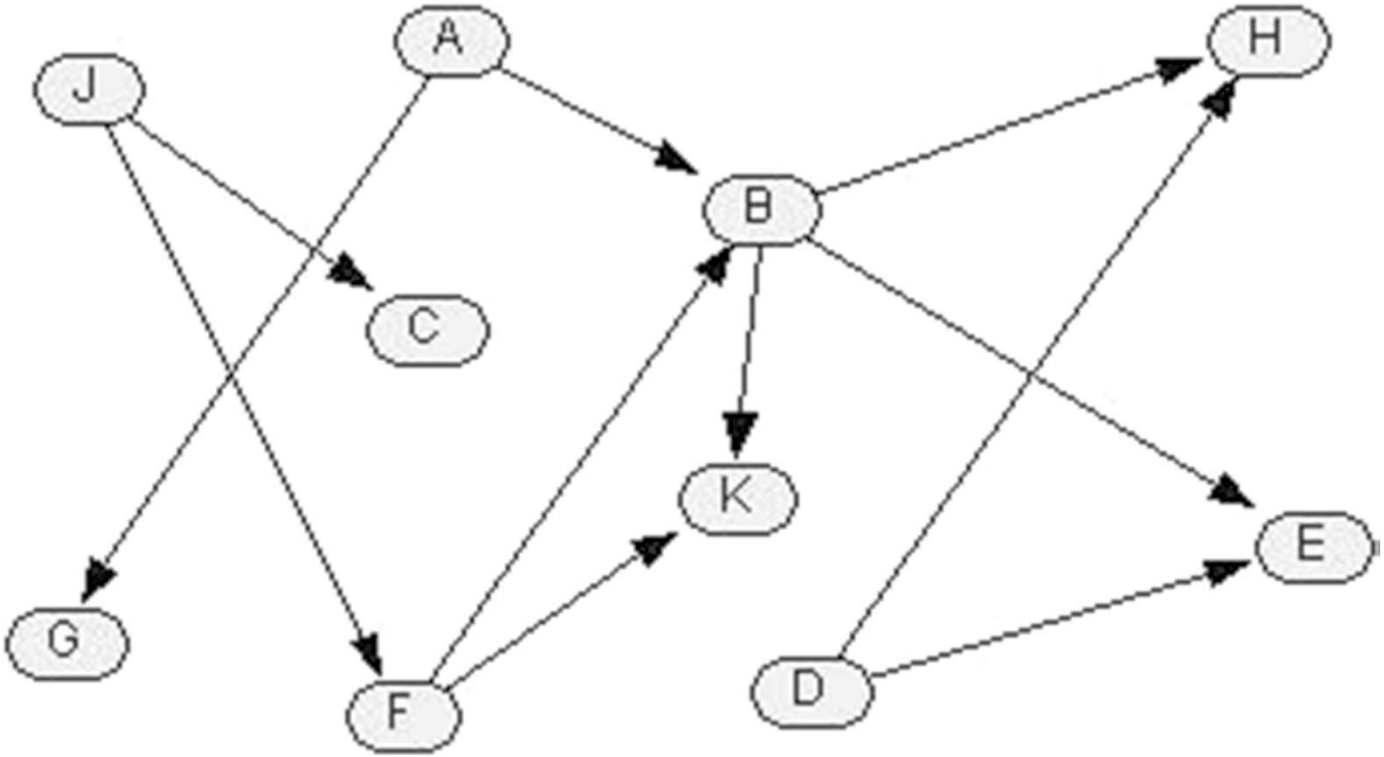
- Bayesian networks express the conditional probabilistic relations among variables as a graphical model. Bayesian networks were first proposed by Pearl in 1985.

- Bayesian networks have two main structures as graphical and probabilistic. The graphical structure of Bayesian networks is based on a relationship structure called Directed Acyclic Graph (DAG).

- A DAG consists of nodes and directed arrows that connect them to each other.

- Variables in the model are represented by nodes, and arrows indicate the conditional probabilistic relations among the variables.

A Bayesian network



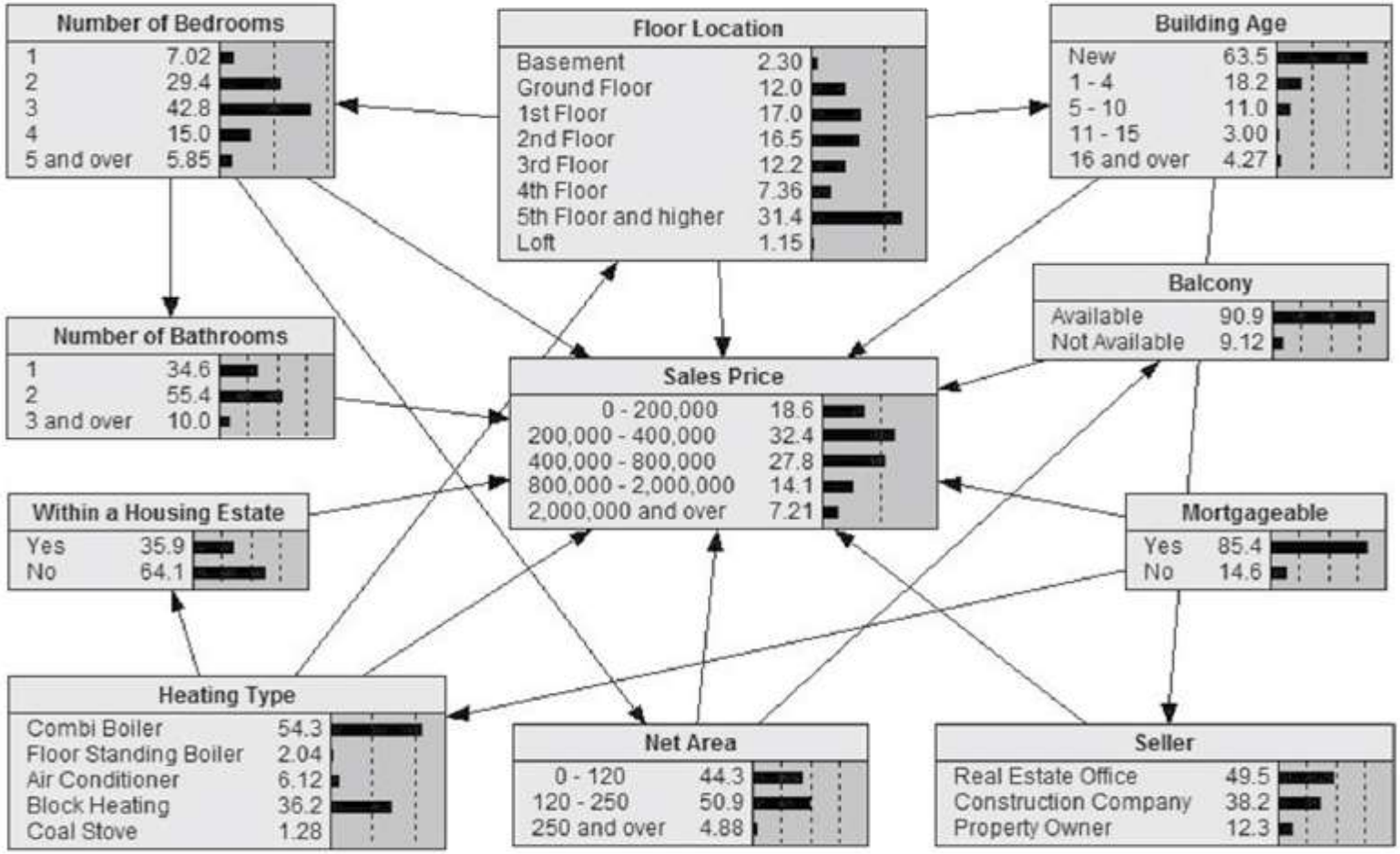
- Flats are the most popular types of residences across the European Union. In fact, Eurostat reports that, 46% of people in the EU countries lives in flat type of residences, 18.6% in semi-detached houses and 34.7% in detached houses.

- According to Turkish Statistical Institute (TurkStat), similar to the EU, flat type residences also are the most popular residence units in Turkey.

- Therefore, in this study, we aim to estimate the sales prices of flats depending on their various characteristics by using a model other than the classic hedonic pricing models, such as a Bayesian network model.

- The data used in the study were drawn from a property sales website sahibinden.com in October 2019 during various connection sessions to the website.
- The data belong to 24 randomly selected cities in Turkey.
- These cities are Adana, Ankara, Antalya, Aydın, Balıkesir, Bursa, Denizli, Diyarbakır, Düzce, Erzurum, Eskişehir, Gaziantep, Kayseri, Kocaeli, Konya, İstanbul, İzmir, Mersin, Muğla, Ordu, Sakarya, Samsun, Tekirdağ and Tabzon.
- The data consist of 5000 observations corresponding to 11 variables indicatingvarious features of a flat.
- The variables were determined based on the results of Lancaster who suggested consumer theory which led to the development of hedonic price models.

Node		Levels	Definition
1	Number of Bedrooms	1 2 3 4 5 and over	Indicates the number of bedrooms in the flat in addition to 1 living room.
2	Building Age (years)	New 1 – 4 5 – 10 11 – 15 16 and over	Indicates the age of the building which the flat is a part of.
3	Balcony	Available Not Available	Indicates whether or not the flat has a balcony.
4	Net Area (m^2)	0 – 120 120 – 250 250 and over	Specifies the amount of net area of the flat in meter squares (m^2).
5	Heating Type	Combi Boiler Floor Standing Boiler Air Conditioner Block Heating Coal Stove	Specifies the heating type used in the flat.
6	Sales Price (TL)	0 – 200000 200000 – 400000 400000 – 800000 800000 – 2000000 2000000 and over	Indicates the sales price of the flat in Turkish Lira (TL).
7	Mortgageable	Yes No	Indicates the eligibility of the flat to get mortgage credit from banks.
8	Number of Bathrooms	1 2 3 and over	Specifies the number of bathrooms in the flat.
9	Seller	Real Estate Office Construction Company Property Owner	Specifies the type of the seller as a real estate office, a construction company or a property owner
10	Within a Housing Estate	Yes No	Indicates whether the flat is located in a housing estate area.
11	Floor Location	Basement Ground Floor 1st floor 2nd Floor 3rd Floor 4th Floor 5th Floor and Higher Loft	Specifies the floor on which the flat is located.



Bayesian network for estimating the sales prices of flats depending on their characteristics

Table 10 The correct classification rates of all models employed in the study

	The Bayesian network model	The artificial neural network model	The decision tree model	The logistic model
Number of correctly classified cases	908	747	727	740
Number of incorrectly classified cases	592	753	773	760
Percentage(%)	61	49.80	48.50	49.30

Table 11 Sensitivity analysis for flat sales prices

Variable	Entropy score
Number of bathrooms	0.12302
Number of bedrooms	0.11548
Net area	0.07552
Floor location	0.03940
Heating type	0.03678
Building age	0.01949
Mortgageable	0.00656
Within a housing estate	0.00610
Seller	0.00450
Balcony	0.00304

Some Interesting Highlights:

- Heating type has less impact on flat sales prices compared to bathroom, bedroom and net area factors.
- Floor location has a low impact rate on flat prices.
- Existence of balcony did not appear to be a significant factor affecting the flat sales prices
- Mortgageability of a flat does not seem to have a significant effect on its sales price.
- There is no significant difference among the prices of the flats being sold by a real estate office, a construction company or a property owner.
- Prices are not affected by ages of buildings as long as the usage areas of the flats are large. In other words, prices of flats do not decrease dramatically in time, as long as they have large usage areas.

Table 12 Findings of the Bayesian network model and the hedonic pricing models

	The Bayesian network model	The hedonic pricing models
Number of bathrooms and sales prices		42 models were examined
	Reports a positive relation	36 models report a positive relation
		1 model reports a negative relation
		5 models report no relation
Number of bedrooms and sales prices		42 models were examined
	Reports a positive relation	23 models report a positive relation
		9 models report a negative relation
		10 models report no relation
Net area and sales prices		69 models were examined
	Reports a positive relation	62 models report a positive relation
		4 models report a negative relation
		3 models report no relation
Building age and sales prices		79 models were examined
		7 models report a positive relation
	Reports a negative relation	64 models report a negative relation
		8 models report no relation
Existence of balcony and sales prices		12 models were examined
		10 models report a positive relation
		0 model reports a negative relation
	Reports no relation	2 models report no relation
Heating type and sales prices		1 model was examined
	Reports a strong significant effect	Reports a strong significant effect
Floor location and sales prices		1 model was examined
	Reports a weak significant effect	Reports a continuous relation between them instead of linear
Mortgageability and sales prices		
	Reports no significant effect	No record
Being in a housing estate and sales prices		
	Reports no significant effect	No record
Seller type and sales prices		1 model was examined
	Reports no significant effect	Reports a significant effect


Number of Bathrooms and Sales Prices
•The result of the Bayesian network model is compatible with 86% of the hedonic pricing model results being examined
Number of Bedrooms and Sales Prices
•The result of the Bayesian network model is compatible with 55% of the hedonic pricing model results being examined
Net Area and Sales Prices
•The result of the Bayesian network model is compatible with 90% of the hedonic pricing model results being examined
Building Age and Sales Prices
•The result of the Bayesian network model is compatible with 81% of the hedonic pricing model results being examined
Existence of Balcony and Sales Prices
•The result of the Bayesian network model is compatible with 17% of the hedonic pricing model results being examined
Heating Type and Sales Prices
•The result of the Bayesian network model is 100% compatible with the hedonic pricing model result being examined
Floor Location and Sales Prices
•The result of the Bayesian network model is parallel to the result of the hedonic pricing model being examined
Mortgageability and Sales Prices
•The result of the Bayesian network model is not comparable with any hedonic pricing model results as there is no record available
Being in a Housing Estate and Sales Prices
•The result of the Bayesian network model is not comparable with any hedonic pricing model results as there is no record available
Seller Type and Sales Prices
•The result of the Bayesian network model is 100% incompatible with the hedonic pricing model result being examined