

PREDICTING ZILLOW'S ZESTIMATE ACCURACY

GROUP #4

GOTTA CAMILLA – IERARDI ANDREA – LAZZARA FRANCESCO – LENI THOMAS –
LEPEK ALEKSANDRA

PROJECT GOAL

The objective of the project is to verify the accuracy of the estimation in the report; moreover here there is a focus on the 8 information quality dimensions method

To reach the target, we analyze the information quality, IQ, through the JMP add-in

1- Data Evaluation and Potential Analysis

8 Information Quality Dimensions

1) DATA RESOLUTION

- 40% of data points were removed due to missing information
- Only 1,4 K valid points
- Only one state

2-3

2) DATA STRUCTURE

- Multiple sources used - besides standard sq meteres also school rankings, proximity of metro
- Data structure is satisfying, it is hetergoneous

3-4

2- Data Evaluation and Potential Analysis

8 Information Quality Dimensions

3) DATA INTEGRATION

- Data between multiple sources used a location as primary key
- There are no privacy concerns as the data is public or home owners accepted to publish it

4

4) TEMPORAL RELEVANCE

- The time span is only 3 years and due to high variation of the house prices over time the model could be easily outdated
- The data needs to be refresh periodically

3-4

3- Data Evaluation and Potential Analysis

8 Information Quality Dimensions

5) CHRONOLOGY OF DATA AND GOAL

- Goal is clear
- Retrospective
- Ex-post availability

4

6) GENERALIZABILITY

- Only applicable to Northern Virginia, because here there are more data

3

4- Data Evaluation and Potential Analysis

8 Information Quality Dimensions

7) OPERATIONALIZATION

- Not clear how this evaluation could be included in decision making process

3

8) COMMUNICATION

- Additional presentation is included which summarise on high level the findings

4

Help

This is a rating-based approach to quantifying InfoQ that scores each of the eight dimensions. This coarse grained approach rates each dimension on a 5 point scale, with 5 indicating "Very High" achievement in that dimension.

The ratings are then normalized into a desirability function for each dimension, which are then combined to produce an overall InfoQ score using the geometric mean of the individual desirabilities.

By dragging the slider handles, each dimension can be assigned a plausible range of ratings, or a specific rating.

InfoQ

Lower Bound: 0,53.00

Upper Bound: 0,64.00

Data Resolution

Low  Acceptable

Data Structure

Acceptable  High

Data Integration

High  High

Temporal Relevance

Acceptable  High

Chronology of Data and Goal

High  High

Generalizability

Acceptable  Acceptable

Operationalization

Acceptable  Acceptable

Communication

High  High

IQ= 58,5%

Table 1- InfoQ-JMP