Start-Tech Academy

DATA DICTIONARY

Next step should be to understand the data. You should know variable definition and distribution along with table's unique identifiers and foreign keys

Data Dictionary

A Comprehensive Data Dictionary should include

1. Definition of predictors

2. Unique identifier of each table (or Primary Keys)

3. Foreign keys or matching keys between tables

   https://youtu.be/76Y6Tg1glrQ

4. Explanation of values in case of Categorical variables

Start-Tech
ACADEMY

Examples

**Data Dictionary
House Pricing Dataset**

The data set contains 506 observations of house prices from different towns. Corresponding to each house price, data of 18 other variables is available on which price is suspected to depend

price	Value of the house
crime_rate	Crime rate in that neighborhood
resid_area	Proportion of residential area in the town
air_qual	Quality of air in that neighborhood
room_num	Average number of rooms in houses of that locality
age	How old is the house construction in years
dist1	Distance from employment hub 1
dist2	Distance from employment hub 2
dist3	Distance from employment hub 3
dist4	Distance from employment hub 4
teachers	Number of teachers per thousand population in the town

Examples

Data Dictionary
House Pricing Dataset

The data set contains 506 observations of house prices from different towns. Corresponding to each house price, data of 18 other variables is available on which price is suspected to depend

poor_prop	Proportion of poor population in the town
airport	Is there an airport in the city? (Yes/No)
n_hos_beds	Number of hospital beds per 1000 population in the town
n_hot_rooms	Number of hotel rooms per 1000 population in the town
waterbody	What type of natural fresh water source is there in the city (lake/ river/ both/ none)
rainfall	The yearly average rainfall in centimeters
bus_ter	Is there a bus terminal in the city? (Yes/No)
parks	Proportion of land assigned as parks and green areas in the town

Start-Tech
ACADEMY