Datasheet for the Dataset of Leading Causes of Death in Alberta*

Yiyi Yao

April 11, 2024

We made a dataset to analysis house prices in Beijing from 2011 to 2017. This helps us see price increase patterns and figure out a fitting model.

Motivation

1. *For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.*

   - The dataset was created to enable analysis of house prices in Beijing. I aim to find connection between variables of house and house prices. Hopefully fit a model to predict house prices.

2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*

   - Ruiqurm from Kaggle. Data used is from Lianjia which is a large house trading platform in China.

Composition

1. *What do the instances that comprise the dataset represent (for example, documents, photos, people, countries)? Are there multiple types of instances (for example, movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.*

   - This dataset has only one csv file.

2. *How many instances are there in total (of each type, if appropriate)?*

   - 1

---

*Code and data are available at: https://github.com/Yaoee111/house-price.git.

3. *Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (for example, geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (for example, to cover a more diverse range of instances, because instances were withheld or unavailable).*

   - No, I believe that we could collect more data from other trading platform.

4. *What data does each instance consist of? "Raw" data (for example, unprocessed text or images) or features? In either case, please provide a description.*

   - number of ladder; date; price; size

5. *Is there a label or target associated with each instance? If so, please provide a description.*

   - No

6. *Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (for example, because it was unavailable). This does not include intentionally removed information, but might include, for example, redacted text.*

Collection process

1. *How was the data associated with each instance acquired? Was the data directly observable (for example, raw text, movie ratings), reported by subjects (for example, survey responses), or indirectly inferred/derived from other data (for example, part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.*

   - This dataset is a well-organized csv file that everyone can download for free.

2. *Who was involved in the data collection process (for example, students, crowdworkers, contractors) and how were they compensated (for example, how much were crowdworkers paid)?*

   - the original data was collect by the author herself.

3. *Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (for example, recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.*

   - 2011-2017

4. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)?*

- from Kaggle

5. *Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.*

- It's a open data. anyone can download for free and without notice the author.

Uses

1. *Has the dataset been used for any tasks already? If so, please provide a description.*

- Yes. Acturally this dataset were used by many people for a long time

2. *Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.*

- https://github.com/Yaoee111/house-price.git

1 References