**Data Quality Form**

**👉 Business Question:  How well do Second-Hand Legos retain their value over time?**

For each item below, give a score ranging from 1 to 5.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **4** | **5** |
| Poor | Below Average | Average | Good | Excellent |

https://data.mendeley.com/datasets/v9hhs66vm3

1. **Relevance**
   * **DESCRIPTION**: How relevant is this data to the business question you are trying to answer? Are all of the fields useful as a part of the analysis?
   * Data is directly relevant to understanding our question, and all the fields are useful
   * **SCORE**: 5
2. **Timeliness**
   * **DESCRIPTION**: How up to date is the data? Is it recent enough and covering a long enough range of time to be useful?
   * Data only goes up to 2019 from 1987. While this is an excellent spread, including data across the Covid-19 pandemic would be valuable in understanding how Legos respond to market shocks.
   * **SCORE**: 5
3. **Historical Usage**
   * **DESCRIPTION**:  Has this data been utilized in other analyses?  If there is not much usage of the data source, is there a reason for this?  Could other users be avoiding this data set for a particular reason? <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3999078> Data is coming from a specific analysis attempting to analyze change in Lego prices over time.
   * See also the follow up study, that suggests Lego is a positive and stable investment choice: https://www.sciencedirect.com/science/article/abs/pii/S0275531921001604
   * **SCORE**: **5**
4. **Cleanliness**
   * **DESCRIPTION**: How clean is the data set? Are there a large amount of null values, duplicated rows, or inconsistent formatting of the data? Will it take a lot of work to get it in usable form?
   * Data is already in a highly functional form, and includes key data points like number of minifigures that are critical in understanding this kind of data
   * **SCORE**: **4**
5. **Credibility**
   * **DESCRIPTION**: What source is this data set coming from?  Is this considered a trustworthy source?  Are there any potential biases that this data source may hold that could skew their presentation of the data?
   * This data source is coming from a scholarly article and analysis that ultimately pulled from eBay pricing. The web scraping can be considered reliable given this is a peer reviewed study, and its authors are in good standing. **SCORE**: **5**

**TOTAL SCORE: 5+4+5+4+5 = 23**

**Data Quality Form**

**👉 Business Question:   How well do various Crypto-Currencies do in retaining their value?**

For each item below, give a score ranging from 1 to 5.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **4** | **5** |
| Poor | Below Average | Average | Good | Excellent |

https://data.world/chasewillden/cryptocurrency-price-by-date-2013-february-2018

1. **Relevance**
   * **DESCRIPTION**: How relevant is this data to the business question you are trying to answer? Are all of the fields useful as a part of the analysis?
   * Data is completely relevant
   * **SCORE**: 5
2. **Timeliness**
   * **DESCRIPTION**: How up to date is the data? Is it recent enough and covering a long enough range of time to be useful?
   * Data covers 2013-2018, which is the furthest back data regarding cryptocurrency goes. It would be better if this data went up to 2023 and has some cleaning that is needed.
   * **SCORE**: 3
3. **Historical Usage**
   * **DESCRIPTION**:  Has this data been utilized in other analyses?  If there is not much usage of the data source, is there a reason for this?  Could other users be avoiding this data set for a particular reason?
   * This data has not been utilized in other analyses, though this appears to be due to its lack of inclusion of data up to 2023.
   * **SCORE**: 3
4. **Cleanliness**
   * **DESCRIPTION**: How clean is the data set? Are there a large amount of null values, duplicated rows, or inconsistent formatting of the data? Will it take a lot of work to get it in usable form?
   * Data is solid, but some of the older data has null or inconsistent values. It will take some work to bring this into usable data.
   * **SCORE**: 2
5. **Credibility**
   * **DESCRIPTION**: What source is this data set coming from?  Is this considered a trustworthy source?  Are there any potential biases that this data source may hold that could skew their presentation of the data?
   * Dataset is coming from data.world dataset that was scraped by another user. This data has been used by several others for small scale analyses. The author of this data-set appears to be a proponent of cryptocurrency, which could lead to some bias.
   * **SCORE**: 4

**TOTAL SCORE: 5+3+3+2+4 = 17**