Report on Data & Task Abstraction Streamlit

# Dataset Type:

This dataset is in Tabular format where each attribute is represented by Columns and item is represented by rows. It can also be represented in Geometry type.

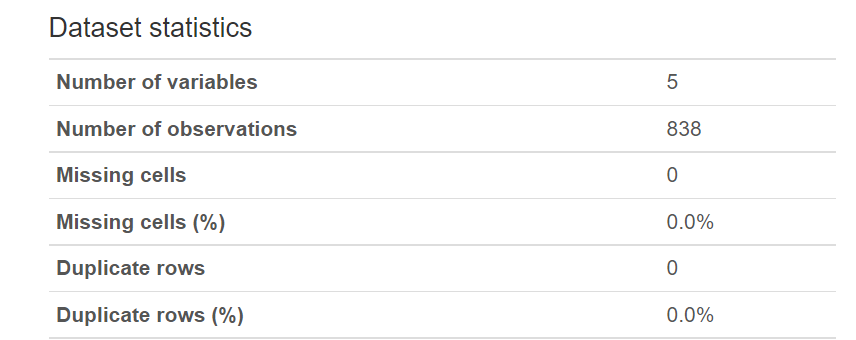
1. Number of items and fields/attributes:

There are 838 items in the Dataset and 5 attributes.

Please find below number of unique values per attribute –

|  |  |
| --- | --- |
| Attribute Name | Number of Unique Values |
| State\_Code | 50 |
| State\_Abv | 50 |
| Damage\_Code | 33 |
| Damage\_Descp | 33 |
| Amount | 837 |

Each of the 5 attributes have 838 values with duplicates



1. The types of all the attributes:

|  |  |
| --- | --- |
| Categorical | 4 |
| Quantitative | 1 |

* 1. Categorical Attributes:
     + State\_Abv
     + Damage\_Desc
     + Damage\_Code
     + State\_Code
  2. Quantitative Attributes:
     + Amount

1. Description for each attribute

* State\_Code

This attribute represents a unique numerical value for each unique state that is represented by State\_Abv

* State\_Abv

This attribute represents abbreviation for different states. It is corelated with State\_Code

* Damage\_Code

This attribute represents a unique numerical value for each unique damage description represented by attribute Damage\_Descp

* Damage\_Descp

This attribute gives a brief description of how the damage occurred

* Amount

This attribute represents the amount that was incurred by the insurance because of the damage caused

1. For each attribute determine its cardinality/range

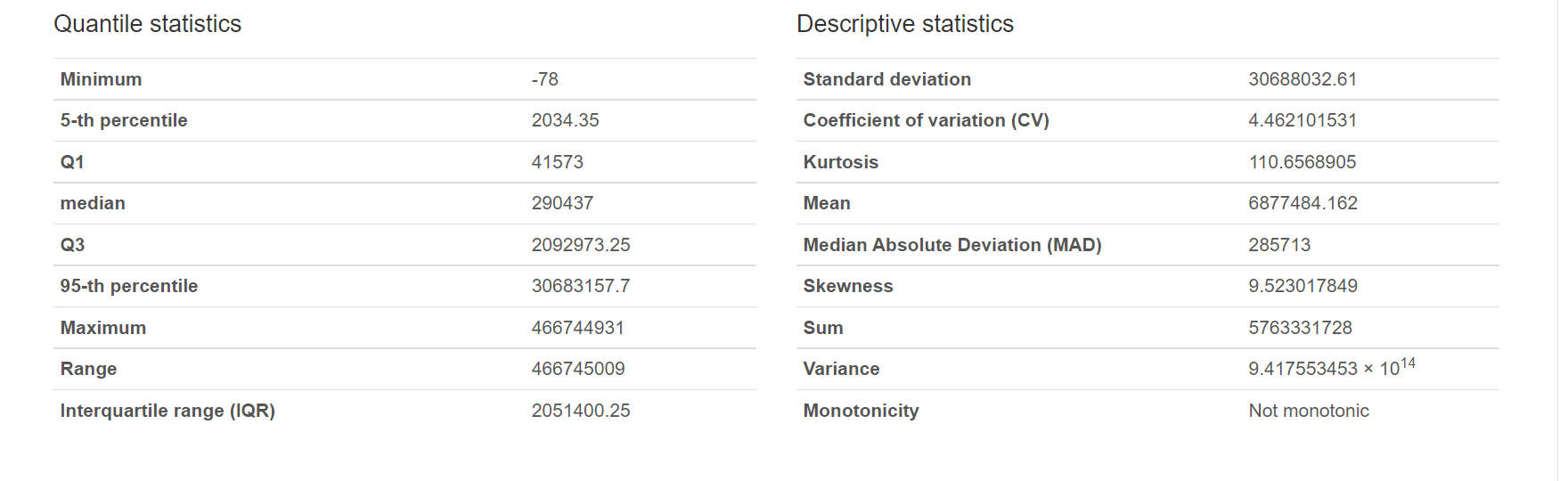
* For categorical attributes, write down the number of unique levels

|  |  |
| --- | --- |
| Attribute Name | Number of Unique Values |
| State\_Abv | 50 |
| Damage\_Descp | 33 |
| State\_Code | 50 |
| Damage\_Code | 33 |

* For quantitative attributes, specify the range from min to max and note any other characterization that seems potentially useful (cyclic? Anything else?)

|  |  |  |  |
| --- | --- | --- | --- |
| Attribute Name | Number of Unique Values | Min | Max |
| Amount | 837 | -78 | 466744931 |

Amount:



The Ordering direction for the above attribute is Sequential.

* For ordered attributes, consider whether it would be more useful to treat them categorical or quantitative, or to preserve them as ordered.

For Ordered attribute, it will be useful to use a metric to convert it to numerical because only numerical values are used in most of the machine learning models.

1. Write two questions you would like to answer with this data set.

* Most frequent damage types

The most common damage type is with “Excess Moisture/Precip/Rain”

Damage\_Descp - Excess Moisture/Precip/Rain, Damage\_Code – 31

* State with highest number of damages

KS State\_Abv has highest number of damages

State\_Abv – KS, State\_Code – 20

* State with lowest no of damages

AK State\_Abv has lowest number of damages

State\_Abv – AK

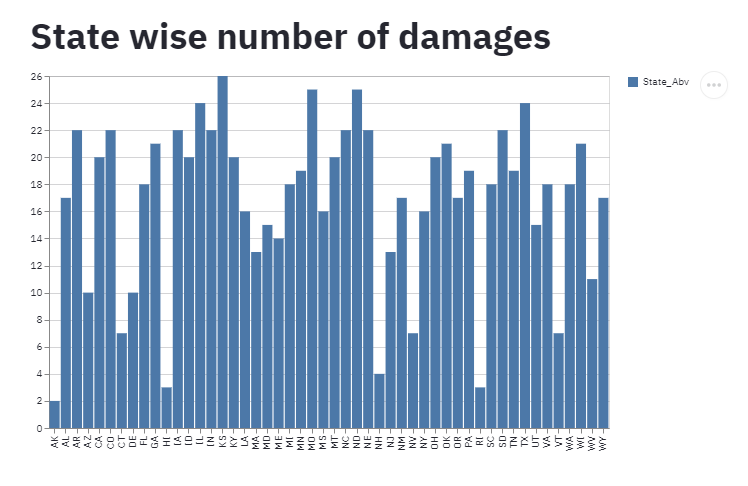
* State which has highest amount

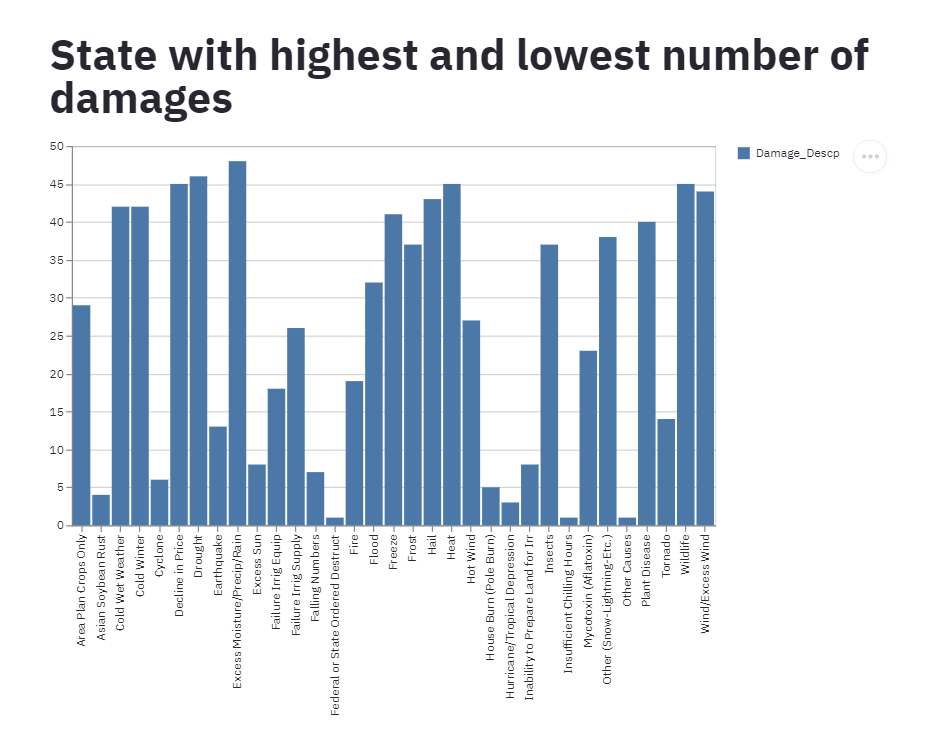
State\_Abv- MO has highest amount

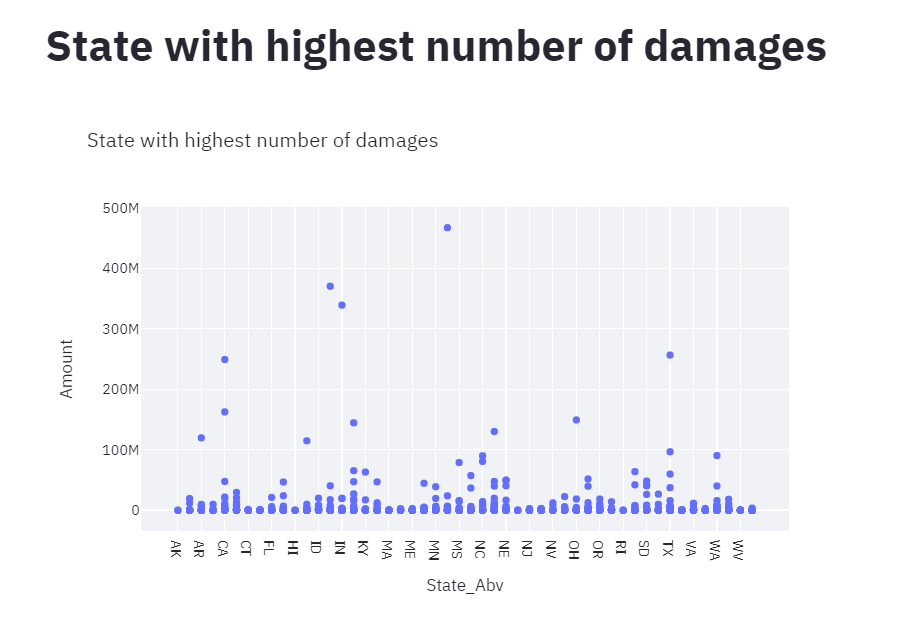
1. For each question, write the following information:
   1. Do you need a chart in order to answer this question?

Yes, having a chart will help us save the manual effort required to search in the dataset every time and also helps us Visualize the data better which gives us a better overall understanding of the dataset and correlation of the attributes.

* 1. If none of your questions require a chart, try to create a new ones that might benefit from one.







* 1. Which fields/attributes do you need to use to answer the question?

Damage\_Descp and State\_Abv were the attributes used to answer the questions 1 and 2.

State\_Abv and Amount attributes were used to answer questions 3 and 4.

* 1. Do you need to transform the data in order to answer the question? If yes, what transformations are needed?

The frequency for each unique value in the attribute had to be counted to plot the bar chart.

* 1. Do dataset type and attribute type change when you need to transform the data? If yes, how do they change?

There is no change in Dataset and attribute types

* 1. Do you have all the data you need to answer this question, or would you need additional data fields that are not provided here?

All the data to answer the questions are available.

* Pandas Profile Output File:

