

Data Profiling

Please find the below summary for the data profiling performed on Iowa_Liquor_Sales data set.

Iowa Liquor Products Data Set:

The given text represents the analysis of several fields in a dataset. Each field has a certain percentage of missing values, a certain number of unique values, and certain statistics such as shortest and longest values, minimum and maximum value counts, and remarks.

- Some fields have a small number of value counts, and it is suggested that if appropriate, some value levels could be combined together. For example, fields like 'Name', 'SCC', 'State Bottle Retail', 'Item Number', 'Bottle Volume (ml)', 'State Case Cost', 'State Bottle Cost', 'Inner Pack', 'Age', 'Vendor', 'Pack', 'Category Name', 'Vendor Name', and 'Item Description' have some value levels with low counts.
- One field 'UPC' has 0.1% missing values, and all other fields have 0% missing values. The 'Report Date' field has only one unique value, and it appears to be a date.
- Overall, the analysis suggests that there may be some data cleaning and preprocessing required in the dataset, such as handling missing values, combining value levels, and potentially removing outliers or errors.

Iowa Liquor Sales By Year And County Data Set:

The dataset contains information on various attributes such as "Sale (Dollars)", "Volume Sold (Gallons)", "County", and "Date".

- There are several fields such as "Name", "SCC", "State Bottle Retail", etc. which have 0% missing values. However, the field "County" has 1% missing values.
- The field "Name" has the maximum unique values of 6,350, while the field "County" has 105 unique values.
- The field "Sale (Dollars)" has the highest value of 19,538,877.03 dollars, whereas "Volume Sold (Gallons)" has the highest value of 1,034,588.02 gallons.
- The field "County" has a minimum value count of 1, and a maximum value count of 11.
- The field "Date" has 11 unique values, and all the values belong to the year 2015.

The remarks given for each field suggest that some values in the fields have a small number of value counts, and it may be appropriate to consider combining some value levels together.

Iowa Liquor Sales By Year Data Set:

There are only 11 records in this data set, with the same single value for "Sale (Dollars)" and "Volume Sold (Gallons)." This suggests that this data set may not be representative of the entire sales data for a larger population, but rather a specific subset or sample.

Iowa Liquor Stores Data Set:

The dataset contains information about liquor sales in the state of Iowa.

- There are 2,277 unique stores in the dataset.
- The dataset contains information about sales made on October 4th, 2022.
- The "Zip Code" field contains 491 unique values, and some of the values have a small number of value counts.

- The "Address" field contains 2,198 unique values, and some of the values have a small number of value counts.
- The "Store Address" field has 5.4% missing values, and contains 1,950 unique values. It is represented as a point using longitude and latitude coordinates.
- The "Name" field contains 2,237 unique values, and some of the values have a small number of value counts.
- The "City" field contains 458 unique values, and some of the values have a small number of value counts.
- The "Store" field contains 2,277 unique values, and some of the values have a small number of value counts.
- The "Store Status" field contains only 2 unique values: "A" and "B". The majority of stores have a status of "A".
- The "Sale (Dollars)" field contains only 11 unique values, and all the values are the same: 255,203,313.57.
- The "Volume Sold (Gallons)" field contains only 11 unique values, and all the values are the same: 4,954,851.80.
- The "Date" field contains only 1 unique value: 2012.

Total City Population By Year Data Set:

The data set seems to be related to US cities and their populations.

- The FIPS code is a unique identifier for each city, with 938 unique values present in the data set.
- The population column ranges from 8 to 126,132 with 5,637 unique values.
- The data set contains data for 31 years from 1990 to 2021.
- There are 933 unique primary point locations, which could be useful for spatial analysis.
- There are 950 unique cities in the data set, with some cities having very few entries (as indicated by the small number of minimum value counts for each city).

Iowa Liquor Sales Data Set:

The dataset contains information about liquor sales from 1,988 stores in Iowa.

- The County_Number column has the highest percentage of missing values at 29.7%. It is recommended to consider imputing these missing values.
- The Store_Location column has 9.9% missing values and contains information about the geographical location of stores in the form of POINT coordinates.
- There are 57 unique values in the Category column.
- The Bottles_Sold column has 353 unique values, with a minimum value count of 1 and a maximum value count of 498,101.
- The Item_Description column has 3,707 unique values, with a minimum value count of 1 and a maximum value count of 50,288.
- Some columns have a small number of value counts, and it is recommended to consider combining some value levels together if appropriate.