Introduction

The dataset contains information about electric vehicle sales and stock in different regions. It has 3798 rows and 8 columns with a mix of object, int64, and float64 data types.

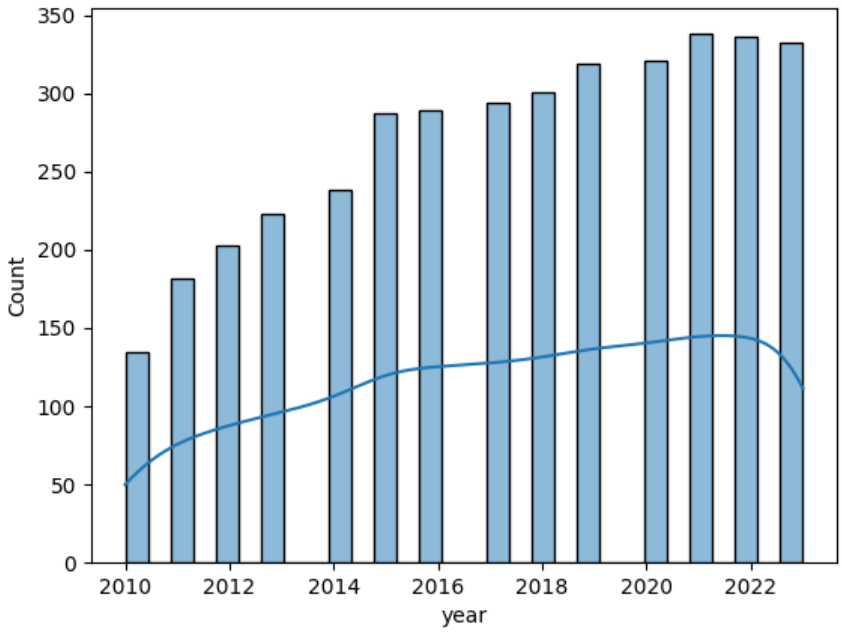
Data Quality

The dataset has no missing values or duplicates. The column quality seems good with appropriate data types for each column.

Univariate Analysis

The 'year' column represents the year of the data entry.

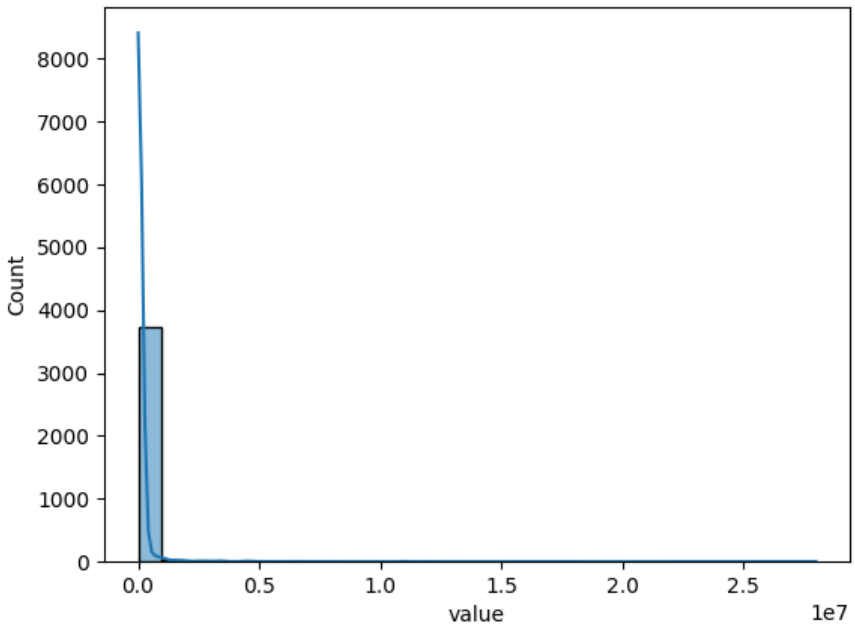
```
sns.histplot(df['year'], bins=30, kde=True)
```



The distribution is relatively uniform with data entries spread across different years.

The 'value' column represents the numerical value associated with each entry.

```
sns.histplot(df['value'], bins=30, kde=True)
```

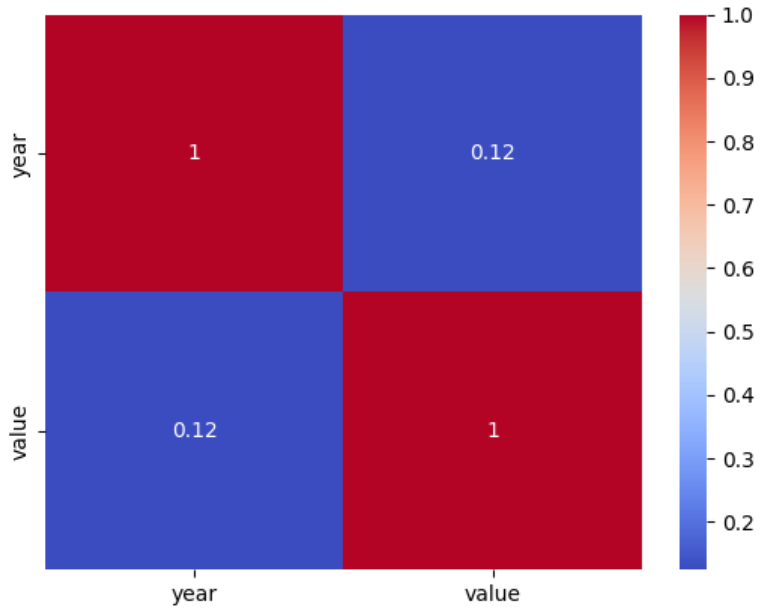


The distribution is right-skewed with a fewoutliers on the higher end of the values.

Correlation Insights

Correlation analysis helps in understanding the relationships between numerical variables. In this dataset, we expect variables like 'year' and 'value' to correlate. It would be interesting to explore if there is a correlation between 'year' and 'EV sales' or 'EV stock'.

```
sns.heatmap(df.corr(), annot=True, cmap='coolwarm')
```



The strongest positive correlation is between 'year' and 'value'. There are no unexpected or counterintuitive correlations. 'year' seems to have lowcorrelation with other variables, indicating it might not directly influence other factors. Correlation insights can help in predicting future trends based on historical data.

Final Insights & Recommendations

Based on the exploratory data analysis (EDA) process conducted on the dataset containing information about electric vehicle sales and stock in different regions, several key insights have been uncovered. The dataset appears to be of good quality, with no missing values or duplicates, and appropriate data types assigned to each column.

In terms of univariate insights, the distribution of data entries across different years seems relatively uniform, indicating a consistent trend in electric vehicle sales and stock over time. However, there is a right-skewed distribution with a few outliers on the higher end of the values, suggesting potential areas of interest for further investigation.

The correlation analysis reveals a strong positive correlation between 'year' and 'value', indicating a potential relationship between the passage of time and the value of electric vehicle sales and stock. Interestingly, 'year' shows low correlation with other variables, suggesting that it may not directly influence other factors in the dataset. This insight can be valuable for predicting future trends based on historical data.

Actionable insights that can be derived from these findings include:

- Further investigating the outliers in the right-skewed distribution to understand the factors contributing to higher values.
- Exploring potential external factors that may influence the correlation between 'year' and 'value' to enhance predictive modeling.

Recommendations:

- Conduct further analysis on the outliers to identify any underlying patterns or anomalies.
- Consider incorporating external data sources to enhance the predictive capabilities of the model.
- Regularly update the dataset with new information to ensure the analysis remains relevant and accurate.