Exploratory Data Analysis Report

# Executive Summary

- Dataset has 398 rows and 9 columns.

- Numeric column with max variance: weight.

- Most frequent in horsepower: 150.

# Dataset Overview

|  |  |  |
| --- | --- | --- |
| Column | Type | Missing |
| origin | int64 | 0 |
| cylinders | int64 | 0 |
| displacement | float64 | 0 |
| horsepower | object | 0 |
| weight | int64 | 0 |
| acceleration | float64 | 0 |
| year | int64 | 0 |
| name | object | 0 |
| Kilometer\_per\_liter | float64 | 0 |

# Numeric Summary

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| index | count | mean | std | min | 25% | 50% | 75% | max |
| origin | 398.0 | 1.57 | 0.8 | 1.0 | 1.0 | 1.0 | 2.0 | 3.0 |
| cylinders | 398.0 | 5.45 | 1.7 | 3.0 | 4.0 | 4.0 | 8.0 | 8.0 |
| displacement | 398.0 | 193.43 | 104.27 | 68.0 | 104.25 | 148.5 | 262.0 | 455.0 |
| weight | 398.0 | 2970.42 | 846.84 | 1613.0 | 2223.75 | 2803.5 | 3608.0 | 5140.0 |
| acceleration | 398.0 | 15.57 | 2.76 | 8.0 | 13.82 | 15.5 | 17.18 | 24.8 |
| year | 398.0 | 1976.01 | 3.7 | 1970.0 | 1973.0 | 1976.0 | 1979.0 | 1982.0 |
| Kilometer\_per\_liter | 398.0 | 10.0 | 3.32 | 3.83 | 7.44 | 9.78 | 12.33 | 19.81 |

# Categorical Summary

## horsepower

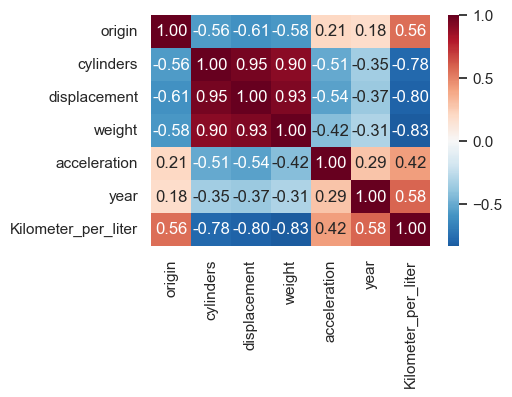
|  |  |
| --- | --- |
| Value | Count |
| 150 | 22 |
| 90 | 20 |
| 88 | 19 |
| 110 | 18 |
| 100 | 17 |

## name

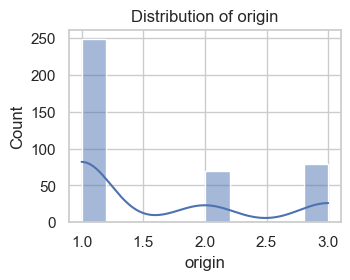
|  |  |
| --- | --- |
| Value | Count |
| ford pinto | 6 |
| amc matador | 5 |
| ford maverick | 5 |
| toyota corolla | 5 |
| chevrolet impala | 4 |

# Correlation Heatmap

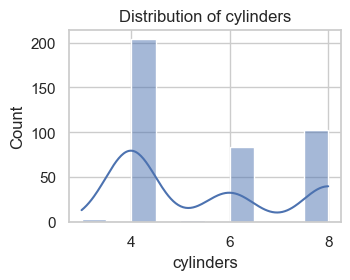
Strongest correlation: \*\*displacement\*\* vs \*\*cylinders\*\* (r=0.95).



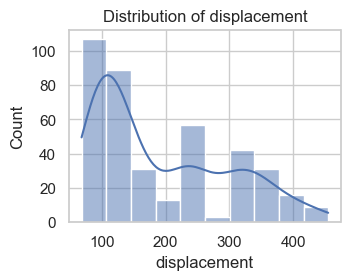
# Distributions



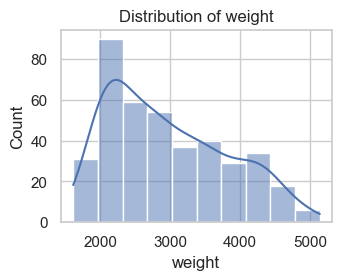
Distribution of \*\*origin\*\* is right-skewed, mean=1.57, median=1.00.



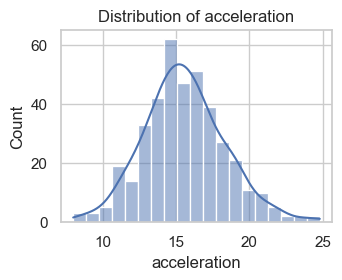
Distribution of \*\*cylinders\*\* is right-skewed, mean=5.45, median=4.00.



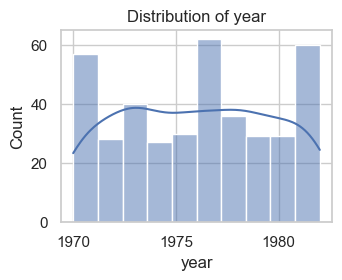
Distribution of \*\*displacement\*\* is right-skewed, mean=193.43, median=148.50.



Distribution of \*\*weight\*\* is right-skewed, mean=2970.42, median=2803.50.

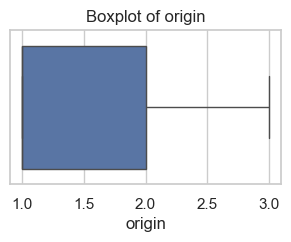


Distribution of \*\*acceleration\*\* is symmetric, mean=15.57, median=15.50.

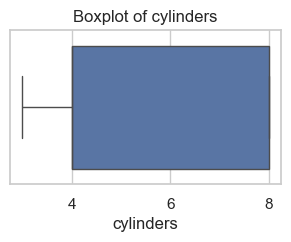


Distribution of \*\*year\*\* is symmetric, mean=1976.01, median=1976.00.

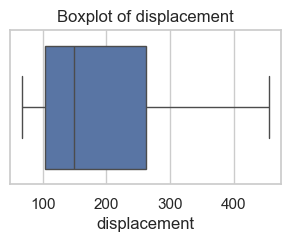
# Boxplots



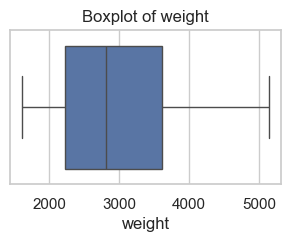
Boxplot of \*\*origin\*\*: IQR=1.00, ~0 outliers.



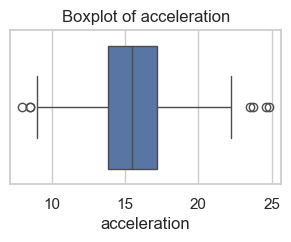
Boxplot of \*\*cylinders\*\*: IQR=4.00, ~0 outliers.



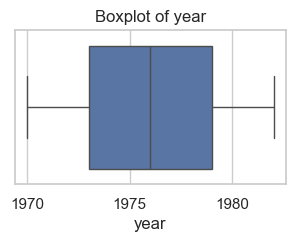
Boxplot of \*\*displacement\*\*: IQR=157.75, ~0 outliers.



Boxplot of \*\*weight\*\*: IQR=1384.25, ~0 outliers.

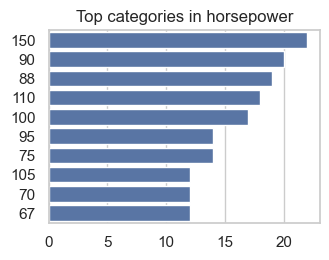


Boxplot of \*\*acceleration\*\*: IQR=3.35, ~7 outliers.

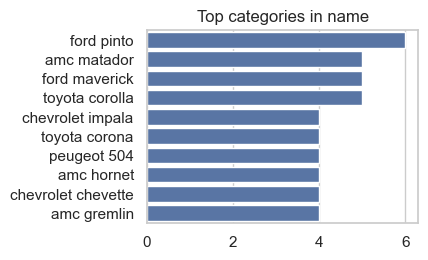


Boxplot of \*\*year\*\*: IQR=6.00, ~0 outliers.

# Top Categories



Top in \*\*horsepower\*\*: 150 (22), 90 (20), 88 (19)



Top in \*\*name\*\*: ford pinto (6), amc matador (5), ford maverick (5)

# Conclusion

This report summarizes dataset structure, missing values, distributions, correlations, and categories. Further analysis (feature engineering, statistical tests, modeling) is recommended for deeper insights.