

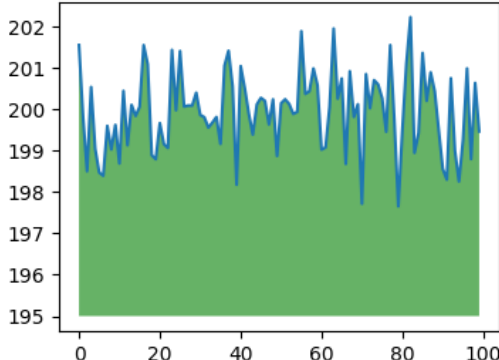
Data Collection and Preprocessing Phase

Date	15 july 2024
Team ID	740771
Project Title	Car Performance Prediction
Maximum Marks	6 Marks

Data Exploration and Preprocessing Report

Analyzed dataset for car performance prediction, handling missing values, outliers, and ensuring data normalization. Performed feature selection and engineering to enhance model accuracy. Final dataset prepared for model training.

Section	Description
	Basic statistics, dimensions, and structure of the data. s
Univariate Analysis	Exploration of individual variables (mean, median, mode, etc.).
	Relationships between two variables (correlation, scatter plots).

	<p style="text-align: center;">Sample Visualization</p> 
Multivariate Analysis	Patterns and relationships involving multiple variables. Data Overview
Outliers and Anomalies	Identification and treatment of outliers.

Data Preprocessing Code Screenshots

car performance prediction.ipynb

colab.research.google.com/drive/1V5W1dH8hMm_o9a3l4PeA1L2etQ_N3Vj#scrollTo=ijRZVzp_-Vtx

car performance prediction.ipynb

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

Connect T4

```
[ ] dataset=dataset.drop('car name',axis=1) #dropping the unwanted column.
```

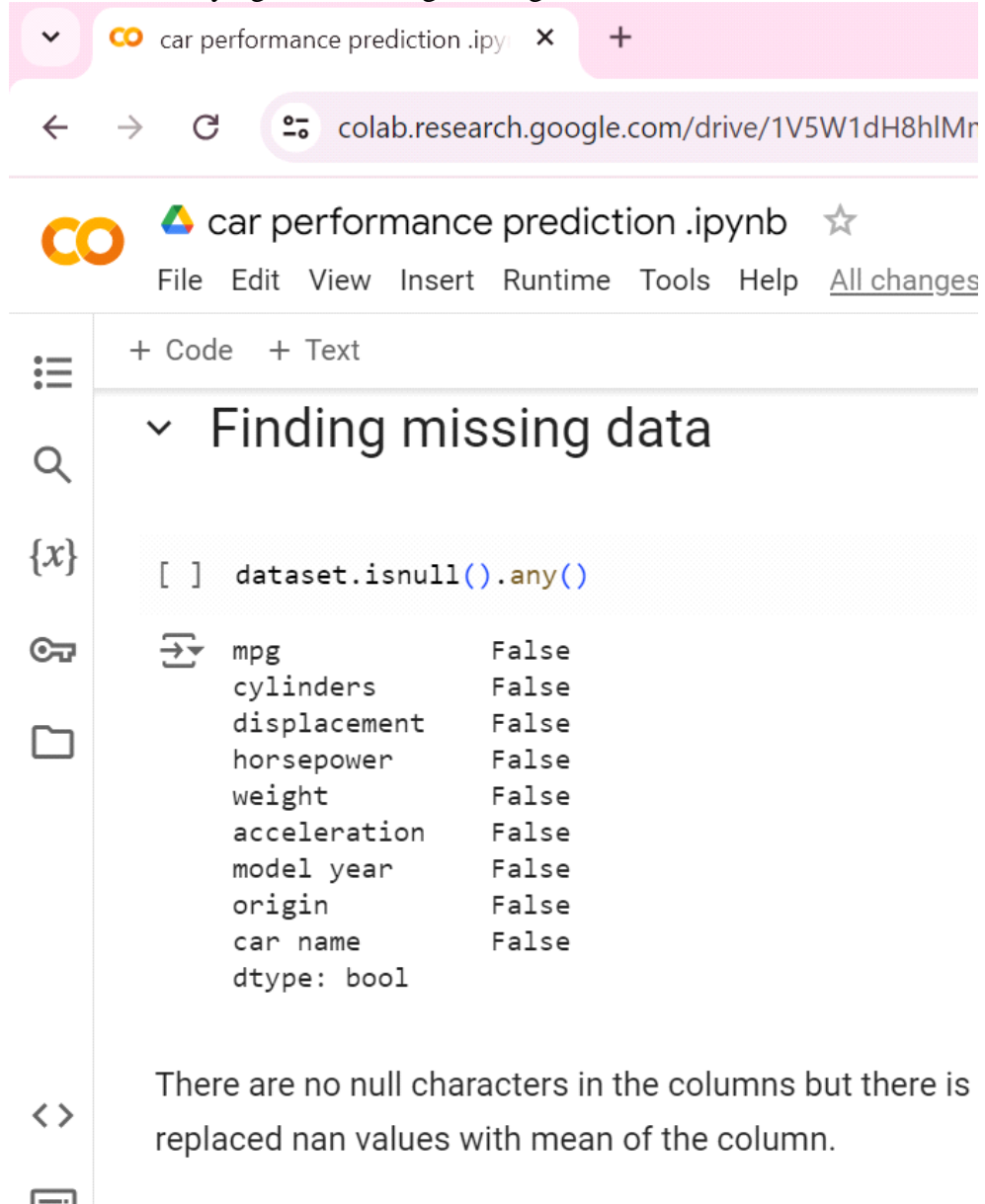
```
corr_table=dataset.corr()#Pandas dataframe.corr() is used to find the pairwise correlation of all columns in the dataframe.
corr_table
```

	mpg	cylinders	displacement	horsepower	weight	acceleration	model year	origin
mpg	1.000000	-0.775396	-0.804203	-0.771437	-0.831741	0.420289	0.579267	0.563450
cylinders	-0.775396	1.000000	0.950721	0.838939	0.896017	-0.505419	-0.348746	-0.562543
displacement	-0.804203	0.950721	1.000000	0.893646	0.932824	-0.543684	-0.370164	-0.609409
horsepower	-0.771437	0.838939	0.893646	1.000000	0.860574	-0.684259	-0.411651	-0.453669
weight	-0.831741	0.896017	0.932824	0.860574	1.000000	-0.417457	-0.306564	-0.581024
acceleration	0.420289	-0.505419	-0.543684	-0.684259	-0.417457	1.000000	0.288137	0.205873
model year	0.579267	-0.348746	-0.370164	-0.411651	-0.306564	0.288137	1.000000	0.180662
origin	0.563450	-0.562543	-0.609409	-0.453669	-0.581024	0.205873	0.180662	1.000000

Loading Data

Handling Missing Data

Code for identifying and handling missing values.



car performance prediction .ipynb

colab.research.google.com/drive/1V5W1dH8hLMr

File Edit View Insert Runtime Tools Help [All changes](#)

+ Code + Text

Finding missing data

```
[ ] dataset.isnull().any()
```

mpg	False
cylinders	False
displacement	False
horsepower	False
weight	False
acceleration	False
model year	False
origin	False
car name	False
dtype:	bool

There are no null characters in the columns but there is replaced nan values with mean of the column.

Data Transformation	Code for transforming variables (scaling, normalization).
Feature Engineering	Code for creating new features or modifying existing ones.