

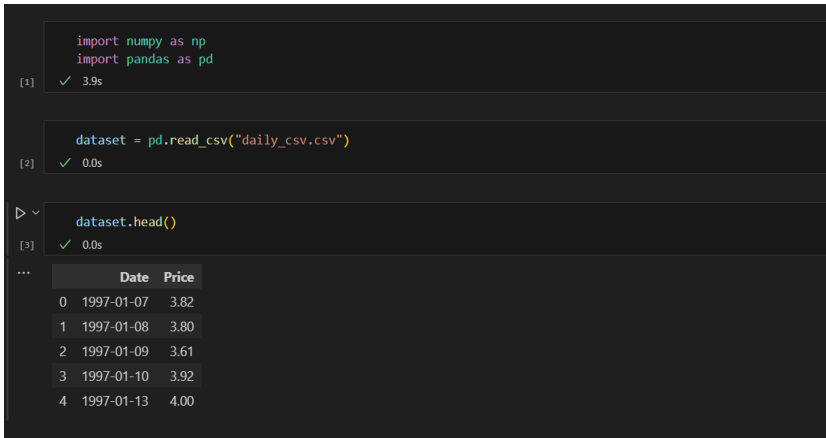
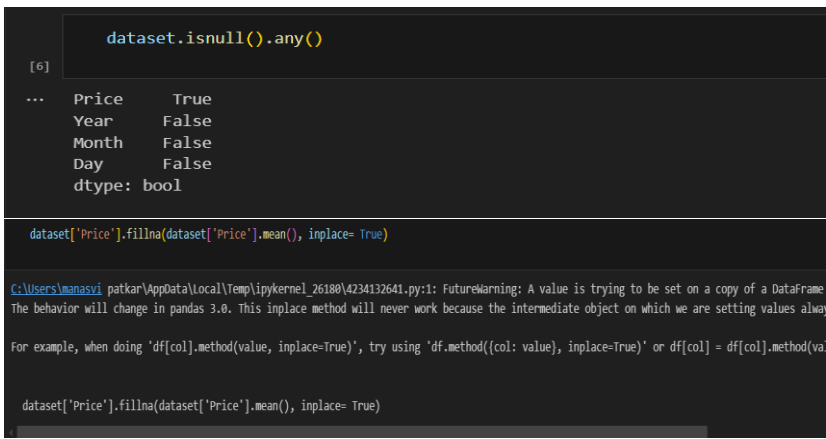
Data Collection and Preprocessing Phase

Date	17 July 2024
Project Title	Machine Learning Approach For Predicting The Price Of Natural Gas
Maximum Marks	6 Marks

Data Exploration and Preprocessing Template

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	Description																																				
Data Overview	Dimension: 5938 rows x 2 columns Description statistics:																																				
	<table><tr><th></th><th>A</th><th>B</th></tr><tr><th>1</th><th>Date</th><th>Price</th></tr><tr><td>2</td><td>1997-01-07</td><td>3.82</td></tr><tr><td>3</td><td>1997-01-08</td><td>3.8</td></tr><tr><td>4</td><td>1997-01-09</td><td>3.61</td></tr><tr><td>5</td><td>1997-01-10</td><td>3.92</td></tr><tr><td>6</td><td>1997-01-13</td><td>4</td></tr><tr><td>7</td><td>1997-01-14</td><td>4.01</td></tr><tr><td>8</td><td>1997-01-15</td><td>4.34</td></tr><tr><td>9</td><td>1997-01-16</td><td>4.71</td></tr><tr><td>10</td><td>1997-01-17</td><td>3.91</td></tr><tr><td>11</td><td>1997-01-20</td><td>3.26</td></tr></table>		A	B	1	Date	Price	2	1997-01-07	3.82	3	1997-01-08	3.8	4	1997-01-09	3.61	5	1997-01-10	3.92	6	1997-01-13	4	7	1997-01-14	4.01	8	1997-01-15	4.34	9	1997-01-16	4.71	10	1997-01-17	3.91	11	1997-01-20	3.26
		A	B																																		
	1	Date	Price																																		
	2	1997-01-07	3.82																																		
	3	1997-01-08	3.8																																		
	4	1997-01-09	3.61																																		
	5	1997-01-10	3.92																																		
	6	1997-01-13	4																																		
	7	1997-01-14	4.01																																		
	8	1997-01-15	4.34																																		
	9	1997-01-16	4.71																																		
	10	1997-01-17	3.91																																		
11	1997-01-20	3.26																																			

Univariate Analysis	-
Bivariate Analysis	-
Multivariate Analysis	-
Outliers and Anomalies	-
Data Preprocessing Code Screenshots	
Loading Data	 <pre> import numpy as np import pandas as pd [1] ✓ 3.9s dataset = pd.read_csv("daily_csv.csv") [2] ✓ 0.0s dataset.head() [3] ✓ 0.0s ... Date Price 0 1997-01-07 3.82 1 1997-01-08 3.80 2 1997-01-09 3.61 3 1997-01-10 3.92 4 1997-01-13 4.00 </pre>
Handling Missing Data	 <pre> dataset.isnull().any() [6] ... Price True Year False Month False Day False dtype: bool dataset['Price'].fillna(dataset['Price'].mean(), inplace=True) C:\Users\manasvi\patkar\AppData\Local\Temp\ipykernel_26180\4234132641.py:1: FutureWarning: A value is trying to be set on a copy of a DataFrame The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always exists. For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(va dataset['Price'].fillna(dataset['Price'].mean(), inplace=True) </pre>
Data Transformation	-

Feature Engineering	-
Save Processed Data	-