**Error description:**

Number of columns in training and test data differ after all preprocessing is done. This is creating an issue while making predictions using ML models later on in the code.

**Analysis:**

**train\_df.shape =**

Before handling null values - (1460,81) [includes Saleprice column]

After handling null values and before scaling and encoding - (1459,75)

After scaling and encoding - (1459,226)

**test\_df.shape =**

Before handling null values - (1459,80) [excludes Saleprice column]

Before handling null values and after scaling and encoding - (1447,74)

After scaling and encoding - (1447,210)

**Summary:**

Total categorical columns for encoding = 37 [43(total categorical columns) - 6 (dropped due to high no. of null values)]

No. of ordinal columns = 12

No. of nominal columns = 37 - 12 = 25

Feature scaling and Ordinal encoding will not affect the number of columns in the dataset. Only nominal encoding can affect the number of columns. The number of nominal columns is 25 for both training and testing data before encoding begins.Therefore, the issue is with encoding of nominal columns.

After encoding, number becomes different. 176 for training data and 160 for testing data.

**Conclusion:**

*One Hot encoder was used twice - once for training data and once for testing data. It was changed so that it is used only once for training data for fitting and transforming and then used only for transforming the test data. This resolved the error. Now, the number of columns are 226 for training data and 225 for testing data after preprocessing is complete. The difference of one is due to the SalePrice (target) column.*