Merging Data

Merge the main training dataset with its corresponding feature data.

Chapter Goals:

- Create the final dataset by merging the training and features DataFrames

A. The final dataset

For the same organizational reasons we had in merging the features and stores DataFrames, we'll now merge the training and combined features DataFrames.

Remember that the stores DataFrame contains potentially useful features listed weekly by store, and the stores DataFrame contains the type and size of each store.

```python
train_df = pd.read_csv('weekly_sales.csv')
print(train_df.columns.tolist())

# Merged and imputed stores + features
print(merged_features.columns.tolist())
```

Viewing the columns of the train_df and merged_features DataFrames.

The code above shows that the two DataFrames share the features `'Store'`, `'Date'`, and `'IsHoliday'`. Therefore, we merge the DataFrames on these three features.

While the `'Date'` feature is useful in the sense that it allows us to identify important values for a given week, like unemployment rate or CPI, it's not used directly in training a machine learning model. Therefore, we drop it from the final dataset.

```
features = ['Store', 'Date', 'IsHoliday']
final_dataset = train_df.merge(merged_features, on=features)
final_dataset = final_dataset.drop(columns=['Date'])


print(final_dataset.columns.tolist())
```

Creating the final dataset.

```
features = ['Store', 'Date', 'IsHoliday']
final_dataset = train_df.merge(merged_features, on=features)
final_dataset = final_dataset.drop(columns=['Date'])


print(final_dataset.columns.tolist())
```