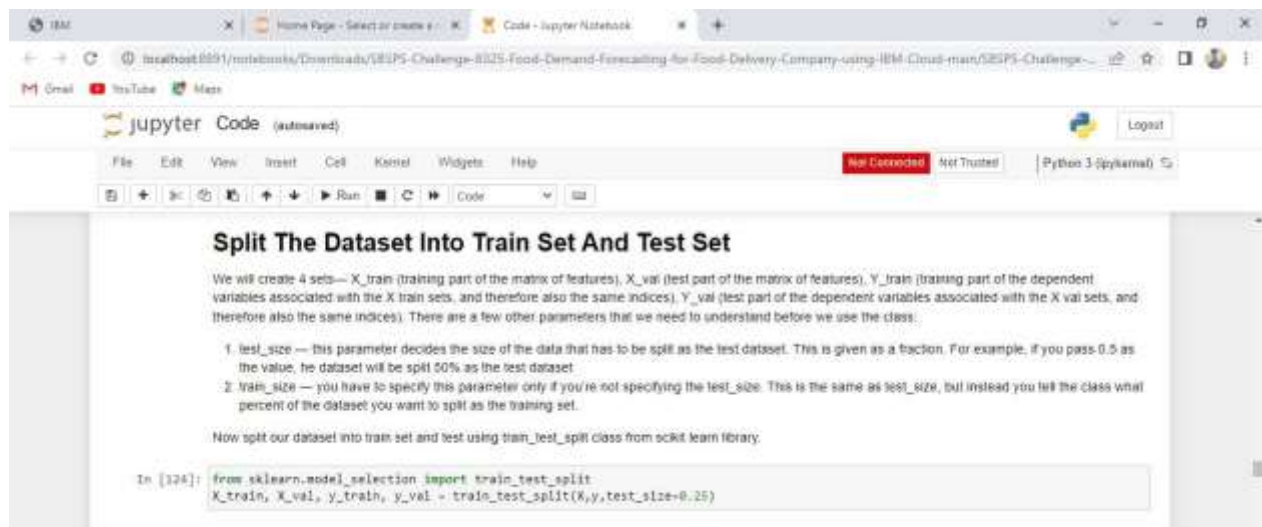


TEAM ID: PNT2022TMID32368

PROJECT NAME: DemandEst - AI powered Food Demand Forecaster

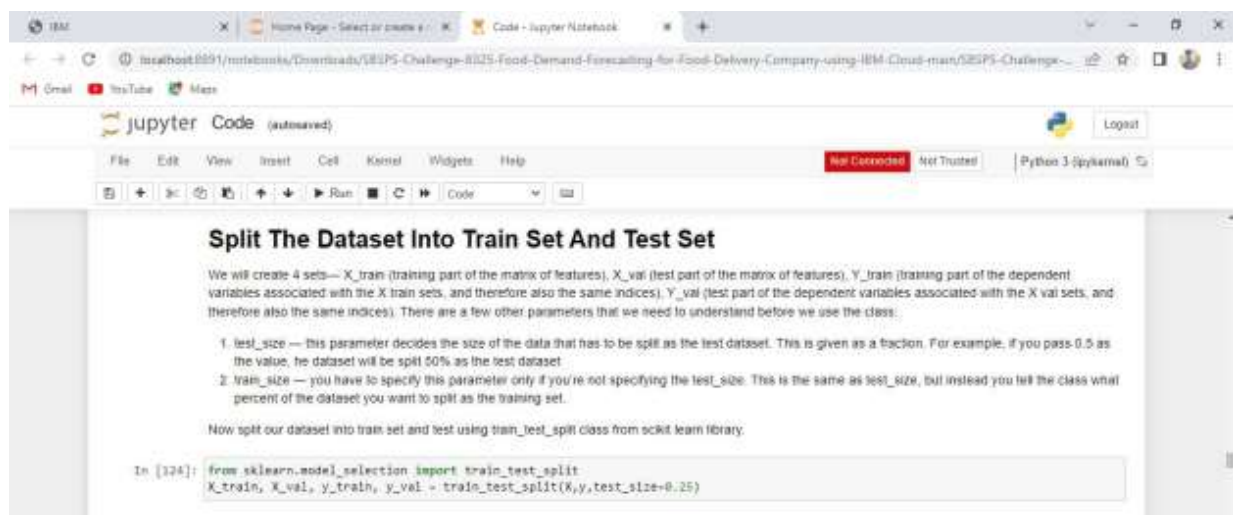
Team Leader



The screenshot shows a Jupyter Notebook interface with a title bar indicating it's running on IBM Cloud. The notebook is titled "Split The Dataset Into Train Set And Test Set". The content includes a paragraph explaining the need to split data into training and testing sets, followed by two numbered points: 1. `test_size` — this parameter decides the size of the data that has to be split as the test dataset. This is given as a fraction. For example, if you pass 0.5 as the value, the dataset will be split 50% as the test dataset. 2. `train_size` — you have to specify this parameter only if you're not specifying the `test_size`. This is the same as `test_size`, but instead you tell the class what percent of the dataset you want to split as the training set. Below this, a note states: "Now split our dataset into train set and test using `train_test_split` class from `scikit learn` library." The code cell shows the following Python code:

```
In [124]: from sklearn.model_selection import train_test_split
X_train, X_val, y_train, y_val = train_test_split(X, y, test_size=0.25)
```

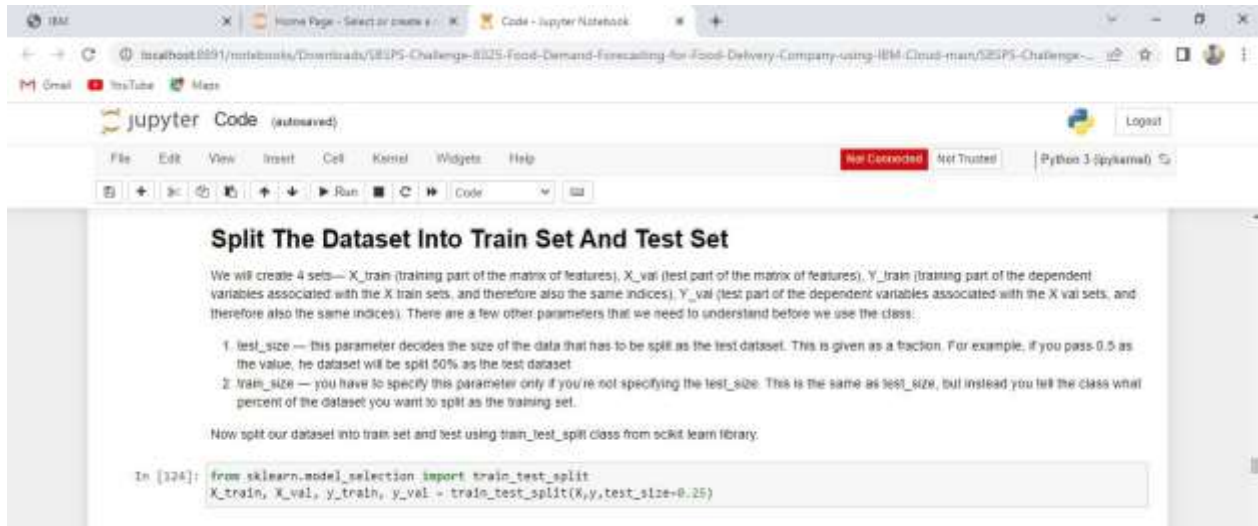
Team Member 1



This screenshot is identical to the one above, showing the same Jupyter Notebook interface with the title "Split The Dataset Into Train Set And Test Set". It contains the same explanatory text and code for splitting the dataset using `train_test_split` from the `scikit learn` library. The code cell shows:

```
In [124]: from sklearn.model_selection import train_test_split
X_train, X_val, y_train, y_val = train_test_split(X, y, test_size=0.25)
```

Team Member 2`



The screenshot shows a Jupyter Notebook titled "Split The Dataset Into Train Set And Test Set". The notebook is running on a local host (localhost:8891) and is using the Python 3 (ipykernel) environment. The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help) and a toolbar with icons for file operations, cell execution, and output viewing. The main content area contains the following text:

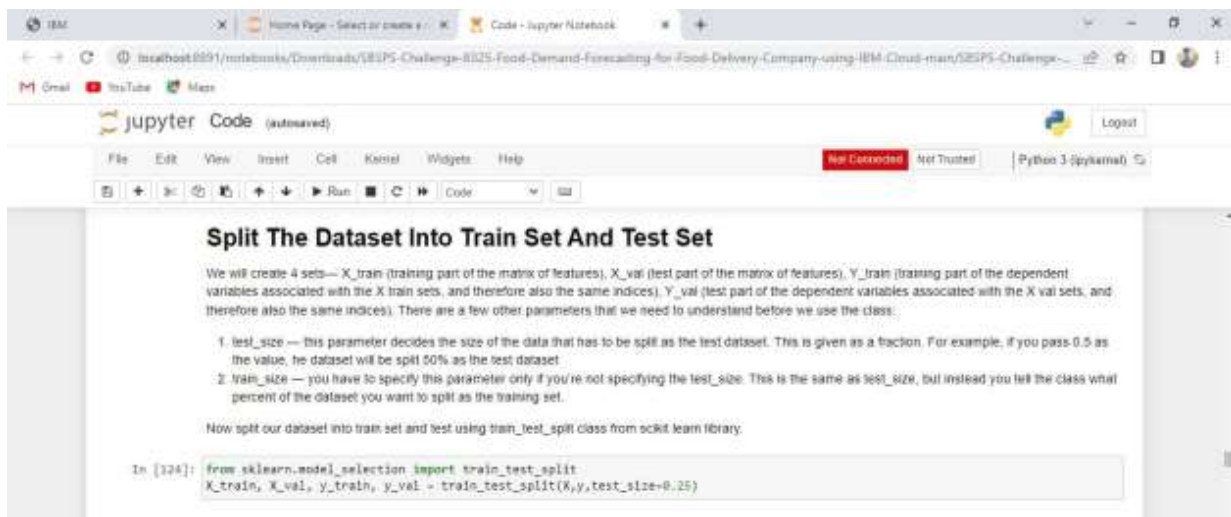
We will create 4 sets— X_{train} (training part of the matrix of features), X_{val} (test part of the matrix of features), Y_{train} (training part of the dependent variables associated with the X_{train} sets, and therefore also the same indices), Y_{val} (test part of the dependent variables associated with the X_{val} sets, and therefore also the same indices). There are a few other parameters that we need to understand before we use the class:

1. `test_size` — this parameter decides the size of the data that has to be split as the test dataset. This is given as a fraction. For example, if you pass 0.5 as the value, the dataset will be split 50% as the test dataset.
2. `train_size` — you have to specify this parameter only if you're not specifying the `test_size`. This is the same as `test_size`, but instead you tell the class what percent of the dataset you want to split as the training set.

Now split our dataset into train set and test using `train_test_split` class from `skit learn` library.

```
In [124]: from sklearn.model_selection import train_test_split
X_train, X_val, y_train, y_val = train_test_split(X, y, test_size=0.25)
```

Team Member 3



The screenshot shows a Jupyter Notebook titled "Split The Dataset Into Train Set And Test Set". The notebook is running on a local host (localhost:8891) and is using the Python 3 (ipykernel) environment. The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help) and a toolbar with icons for file operations, cell execution, and output viewing. The main content area contains the following text:

We will create 4 sets— X_{train} (training part of the matrix of features), X_{val} (test part of the matrix of features), Y_{train} (training part of the dependent variables associated with the X_{train} sets, and therefore also the same indices), Y_{val} (test part of the dependent variables associated with the X_{val} sets, and therefore also the same indices). There are a few other parameters that we need to understand before we use the class:

1. `test_size` — this parameter decides the size of the data that has to be split as the test dataset. This is given as a fraction. For example, if you pass 0.5 as the value, the dataset will be split 50% as the test dataset.
2. `train_size` — you have to specify this parameter only if you're not specifying the `test_size`. This is the same as `test_size`, but instead you tell the class what percent of the dataset you want to split as the training set.

Now split our dataset into train set and test using `train_test_split` class from `skit learn` library.

```
In [124]: from sklearn.model_selection import train_test_split
X_train, X_val, y_train, y_val = train_test_split(X, y, test_size=0.25)
```