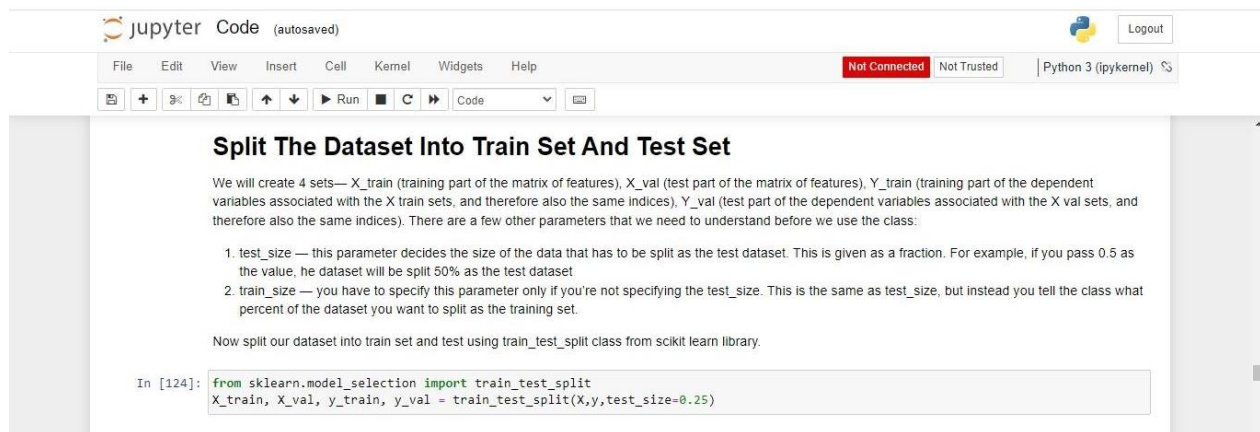


TEAM ID: PNT2022TMID52731

PROJECT NAME: DemandEst - AI powered Food Demand Forecaster

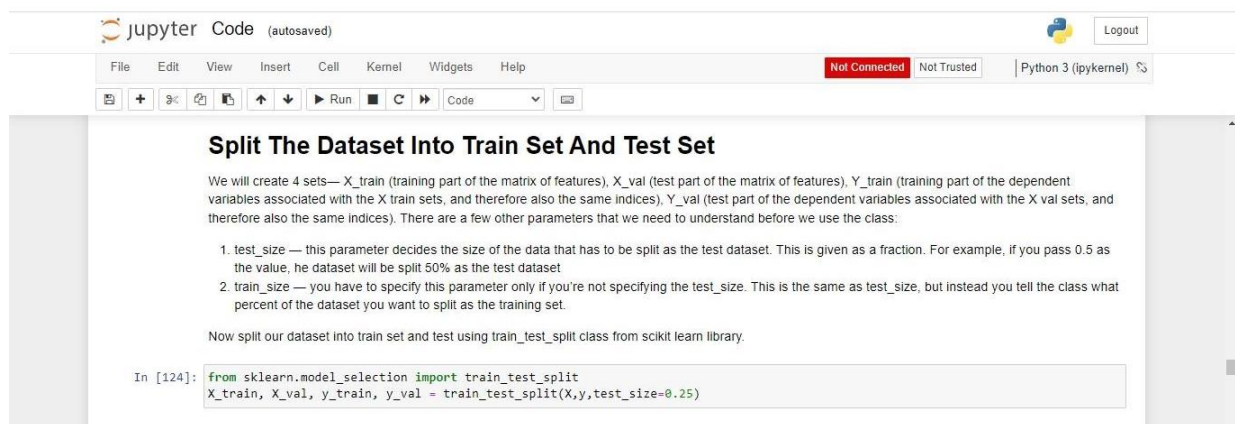
Team Leader



The screenshot shows a Jupyter Notebook interface. At the top, there's a header with the Jupyter logo, the text 'jupyter Code (autosaved)', and a 'Logout' button. Below this is a menu bar with 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', 'Widgets', and 'Help'. To the right of the menu bar, there's a status bar showing 'Not Connected', 'Not Trusted', and 'Python 3 (ipykernel)'. Below the menu bar is a toolbar with various icons for file operations, editing, and running code. The main content area has a title 'Split The Dataset Into Train Set And Test Set'. Below the title, there's a paragraph explaining the goal: '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:'. This is followed by a numbered list of two items: 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 the list, there's a paragraph: 'Now split our dataset into train set and test using train\_test\_split class from scikit learn library.' At the bottom, there's a code cell with the following 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)
```

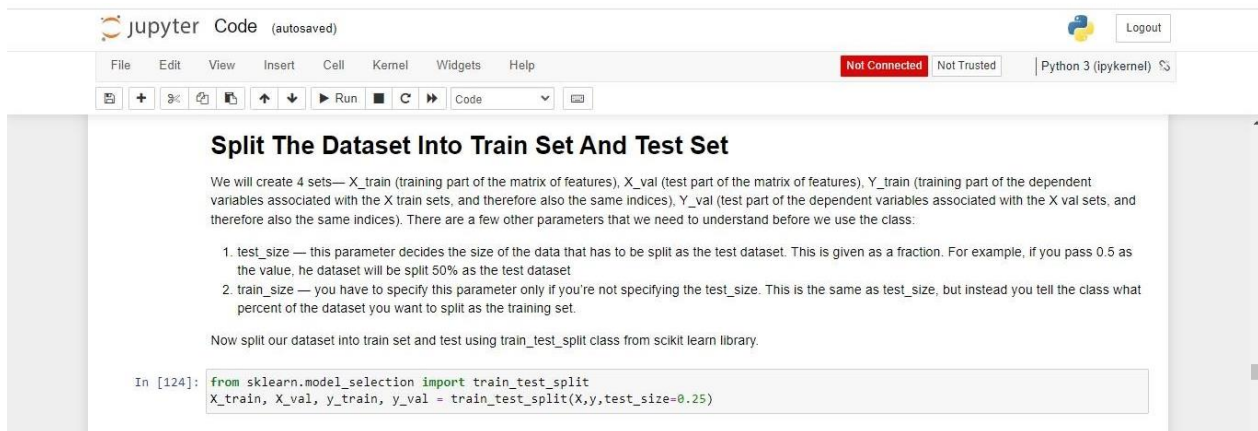
Team Member 1



The screenshot shows a Jupyter Notebook interface. At the top, there's a header with the Jupyter logo, the text 'jupyter Code (autosaved)', and a 'Logout' button. Below this is a menu bar with 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', 'Widgets', and 'Help'. To the right of the menu bar, there's a status bar showing 'Not Connected', 'Not Trusted', and 'Python 3 (ipykernel)'. Below the menu bar is a toolbar with various icons for file operations, editing, and running code. The main content area has a title 'Split The Dataset Into Train Set And Test Set'. Below the title, there's a paragraph explaining the goal: '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:'. This is followed by a numbered list of two items: 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 the list, there's a paragraph: 'Now split our dataset into train set and test using train\_test\_split class from scikit learn library.' At the bottom, there's a code cell with the following 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)
```

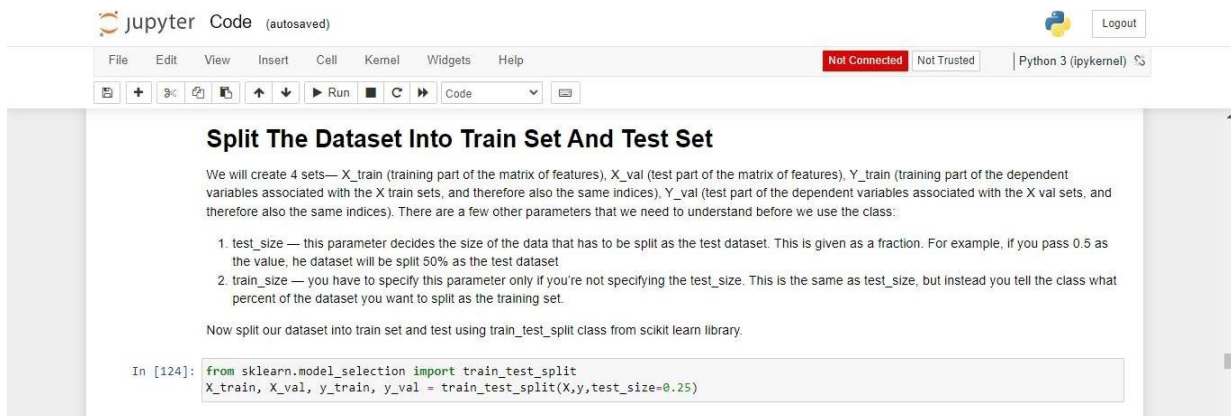
## Team Member 2



The screenshot shows a Jupyter Notebook interface. At the top, the title bar reads "jupyter Code (autosaved)". The menu bar includes File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. On the right, there are status indicators: "Not Connected", "Not Trusted", and "Python 3 (ipykernel)". Below the menu bar is a toolbar with icons for file operations, running, and code execution. The main content area has a title "Split The Dataset Into Train Set And Test Set". The text explains the creation of four 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), and  $Y_{val}$  (test part of the dependent variables associated with the  $X_{val}$  sets, and therefore also the same indices). It mentions that there are a few other parameters that need to be understood before using the class. Two parameters are listed: 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, it says "Now split our dataset into train set and test using `train_test_split` class from scikit learn library." A code cell is shown with the following 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 3



The screenshot shows a Jupyter Notebook interface. At the top, the title bar reads "jupyter Code (autosaved)". The menu bar includes File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. On the right, there are status indicators: "Not Connected", "Not Trusted", and "Python 3 (ipykernel)". Below the menu bar is a toolbar with icons for file operations, running, and code execution. The main content area has a title "Split The Dataset Into Train Set And Test Set". The text explains the creation of four 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), and  $Y_{val}$  (test part of the dependent variables associated with the  $X_{val}$  sets, and therefore also the same indices). It mentions that there are a few other parameters that need to be understood before using the class. Two parameters are listed: 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, it says "Now split our dataset into train set and test using `train_test_split` class from scikit learn library." A code cell is shown with the following 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)
```