

TEAM ID: PNT2022TMID32368

PROJECT NAME: DemandEst - AI powered Food Demand Forecaster

Team Leader

The image displays two screenshots of a Jupyter Notebook interface, showing the process of splitting a dataset into independent and dependent variables.

**Top Screenshot:**

### Splitting The Dataset Into Dependent And Independent Variable

In machine learning, the concept of dependent variable (y) and independent variables(x) is important to understand. Here, Dependent variable is nothing but output in dataset and independent variable is all inputs in the dataset.

With this in mind, we need to split our dataset into the matrix of independent variables and the vector or dependent variable. Mathematically, Vector is defined as a matrix that has just one column.

Let's split our dataset into independent and dependent variables.

1. The independent variable in the dataset would be considered as 'x' and the 'homepage\_featured', 'emailer\_for\_promotion', 'op\_area', 'cuisine', 'city\_code', 'region\_code', 'category' columns would be considered as independent variable.
2. The dependent variable in the dataset would be considered as 'y' and the 'num\_orders' column is considered as dependent variable.

```
In [122]: features = columns.drop(['num_orders'])
trainFinal1 = trainFinal[features]
x = trainFinal1.values
y = trainFinal['num_orders'].values
```

```
In [123]: trainFinal1.head()
```

	homepage_featured	emailer_for_promotion	op_area	cuisine	city_code	region_code	category
0	0	0	2.0	3	047	56	0
1	0	0	2.0	3	047	56	0
2	0	0	2.0	3	047	56	0

**Bottom Screenshot:**

output in dataset and independent variable is all inputs in the dataset.

With this in mind, we need to split our dataset into the matrix of independent variables and the vector or dependent variable. Mathematically, Vector is defined as a matrix that has just one column.

Let's split our dataset into independent and dependent variables.

1. The independent variable in the dataset would be considered as 'x' and the 'homepage\_featured', 'emailer\_for\_promotion', 'op\_area', 'cuisine', 'city\_code', 'region\_code', 'category' columns would be considered as independent variable.
2. The dependent variable in the dataset would be considered as 'y' and the 'num\_orders' column is considered as dependent variable.

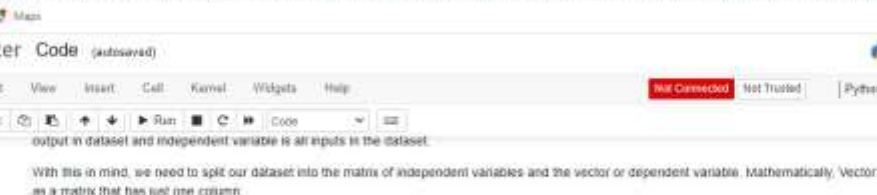
```
In [122]: features = columns.drop(['num_orders'])
trainFinal1 = trainFinal[features]
x = trainFinal1.values
y = trainFinal['num_orders'].values
```

```
In [123]: trainFinal1.head()
```

	homepage_featured	emailer_for_promotion	op_area	cuisine	city_code	region_code	category
0	0	0	2.0	3	047	56	0
1	0	0	2.0	3	047	56	0
2	0	0	2.0	3	047	56	0
3	0	0	2.0	3	047	56	0
4	0	0	2.0	3	047	56	0

## Team Member 1

[illegible]



```
output in dataset and independent variable is all inputs in the dataset.

With this in mind, we need to split our dataset into the matrix of independent variables and the vector or dependent variable. Mathematically, Vector is defined as a matrix that has just one column.

Let's split our dataset into independent and dependent variables.

1. The independent variable in the dataset would be considered as 'x' and the 'homepage_featured', 'emailer_for_promotion', 'cp_area', 'cuisine', 'city_code', 'region_code', 'category' columns would be considered as independent variable.

2. The dependent variable in the dataset would be considered as 'y' and the 'num_orders' column is considered as dependent variable.

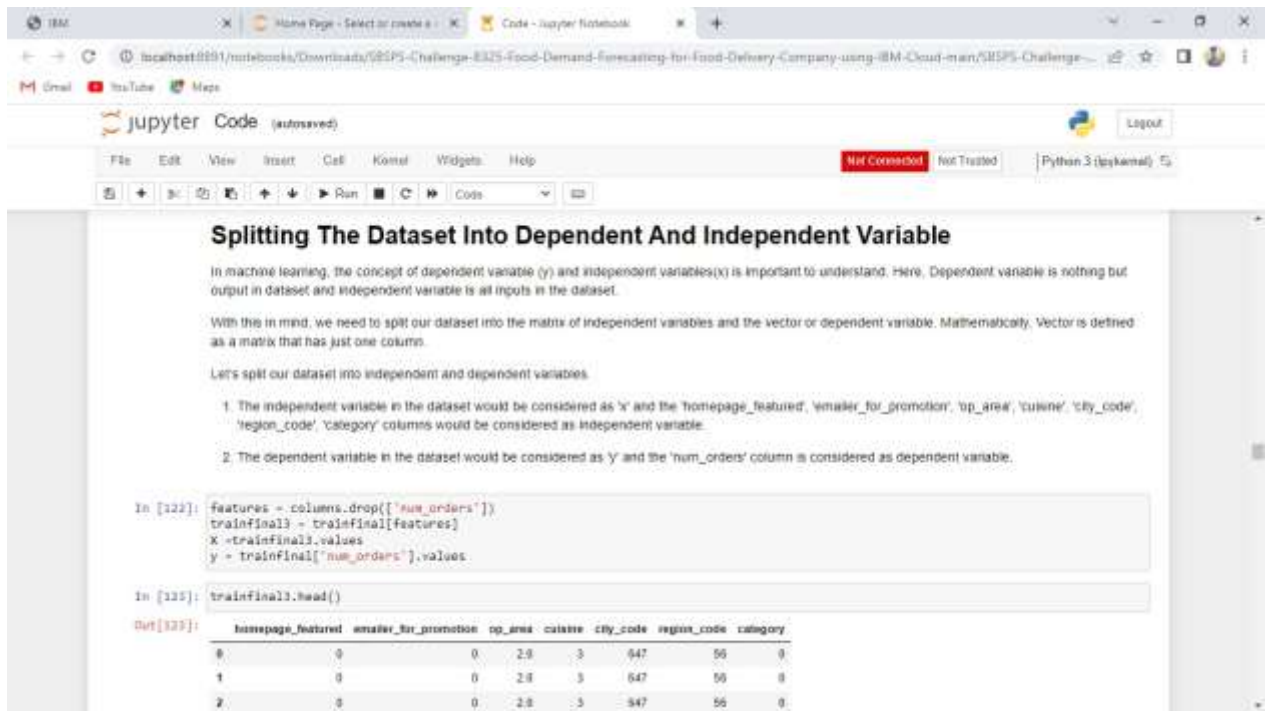
In [122]: features = columns.drop(['num_orders'])
trainfinal3 = trainfinal[features]
X = trainfinal3.values
y = trainfinal['num_orders'].values

In [123]: trainfinal3.head()

Out[123]:
```

	homepage_featured	emailer_for_promotion	cp_area	cuisine	city_code	region_code	category
0	0	0	2.0	3	047	56	0
1	0	0	2.0	3	047	56	0
2	0	0	2.0	3	047	56	0
3	0	0	2.0	3	047	56	0
4	0	0	2.0	3	047	56	0

## Team Member 2



The screenshot shows a Jupyter Notebook interface with a title bar indicating it's running on IBM Cloud. The notebook is titled "Splitting The Dataset Into Dependent And Independent Variable". The content includes a brief explanation of dependent and independent variables in machine learning, followed by a list of two points: 1. The independent variable 'x' consists of 'homepage\_featured', 'emailer\_for\_promotion', 'op\_area', 'cuisine', 'city\_code', 'region\_code', and 'category'. 2. The dependent variable 'y' is 'num\_orders'. Below this, there are two code cells. The first cell defines 'features' as all columns except 'num\_orders', splits the data into 'trainfinal3' and 'testfinal3', and extracts the 'num\_orders' values into 'X' and 'y'. The second cell displays the first three rows of 'trainfinal3'.

### Splitting The Dataset Into Dependent And Independent Variable

In machine learning, the concept of dependent variable (y) and independent variables(x) is important to understand. Here, Dependent variable is nothing but output in dataset and independent variable is all inputs in the dataset.

With this in mind, we need to split our dataset into the matrix of independent variables and the vector or dependent variable. Mathematically, Vector is defined as a matrix that has just one column.

Let's split our dataset into independent and dependent variables.

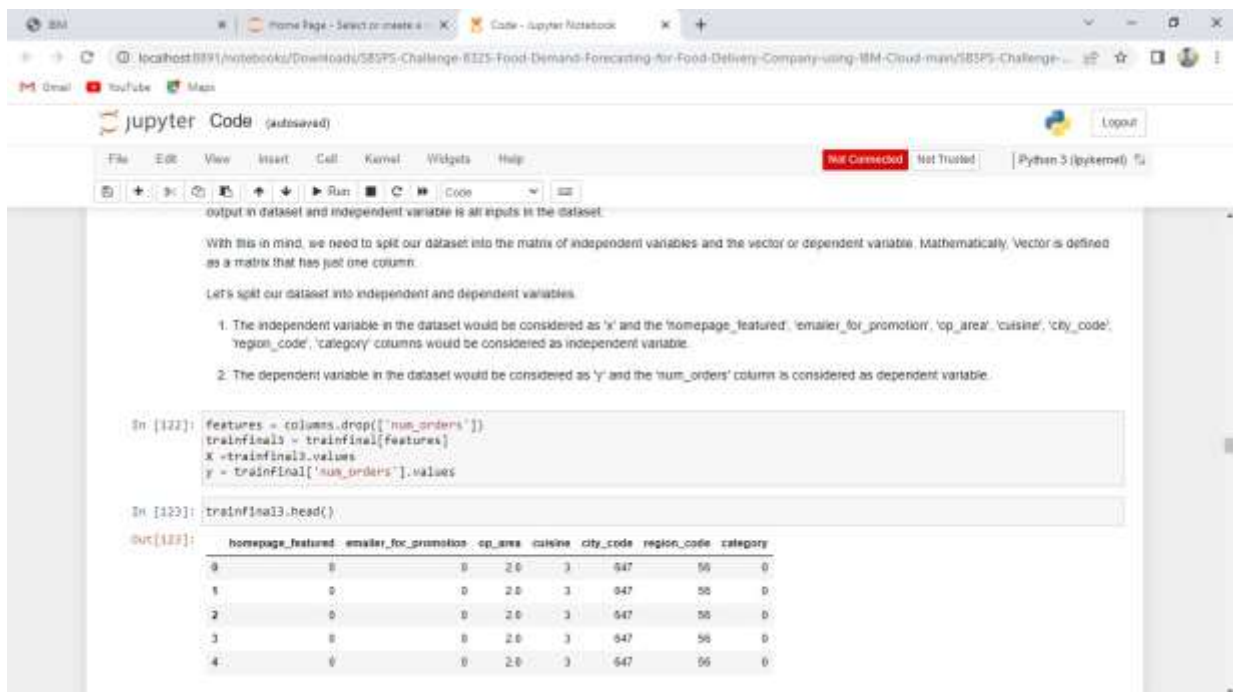
1. The independent variable in the dataset would be considered as 'x' and the 'homepage\_featured', 'emailer\_for\_promotion', 'op\_area', 'cuisine', 'city\_code', 'region\_code', 'category' columns would be considered as independent variable.
2. The dependent variable in the dataset would be considered as 'y' and the 'num\_orders' column is considered as dependent variable.

```
In [122]: features = columns.drop(['num_orders'])
          trainfinal3 = trainfinal[features]
          X = trainfinal3.values
          y = trainfinal['num_orders'].values
```

```
In [123]: trainfinal3.head()
```

```
Out[123]:
```

	homepage_featured	emailer_for_promotion	op_area	cuisine	city_code	region_code	category
0	0	0	0	2.0	3	647	56
1	0	0	0	2.0	3	647	56
2	0	0	0	2.0	3	647	56



This screenshot shows the same Jupyter Notebook interface, but the output of the 'trainfinal3.head()' command now displays five rows of data (indices 0 to 4).

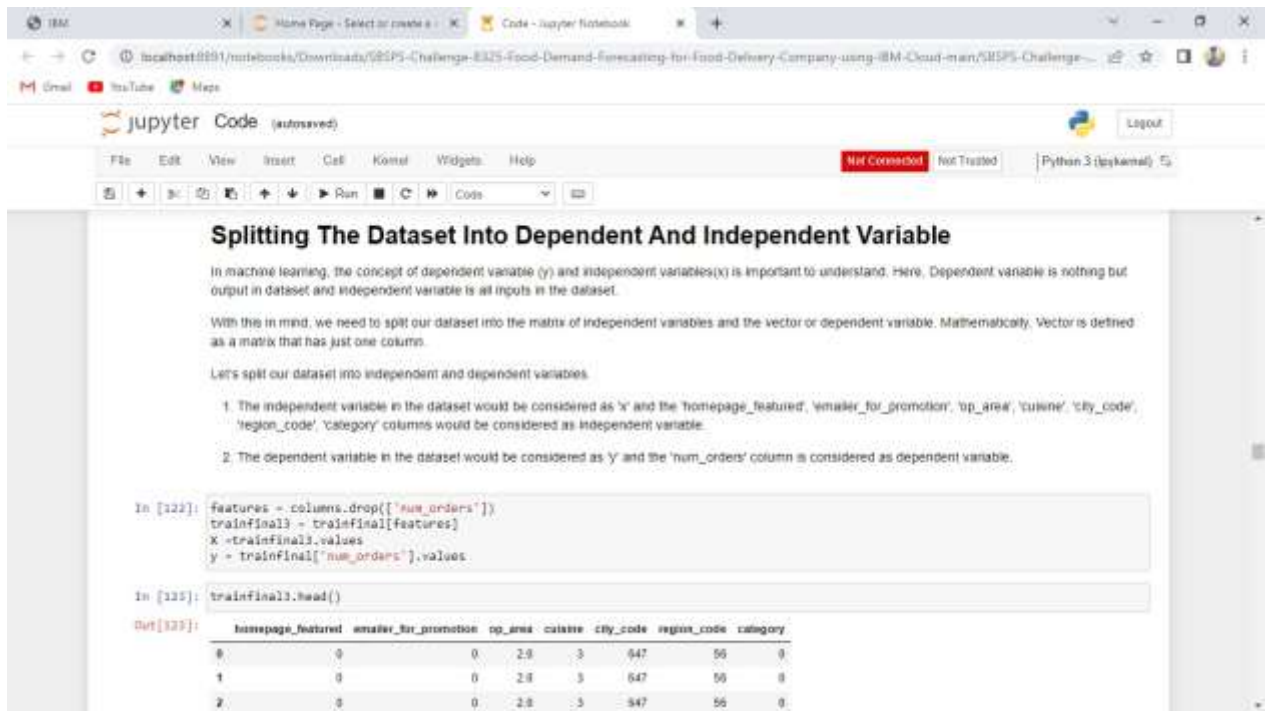
```
In [122]: features = columns.drop(['num_orders'])
          trainfinal3 = trainfinal[features]
          X = trainfinal3.values
          y = trainfinal['num_orders'].values
```

```
In [123]: trainfinal3.head()
```

```
Out[123]:
```

	homepage_featured	emailer_for_promotion	op_area	cuisine	city_code	region_code	category
0	0	0	0	2.0	3	647	56
1	0	0	0	2.0	3	647	56
2	0	0	0	2.0	3	647	56
3	0	0	0	2.0	3	647	56
4	0	0	0	2.0	3	647	56

## Team Member 3



The screenshot shows a Jupyter Notebook interface with the following content:

### Splitting The Dataset Into Dependent And Independent Variable

In machine learning, the concept of dependent variable (y) and independent variables(x) is important to understand. Here, Dependent variable is nothing but output in dataset and independent variable is all inputs in the dataset.

With this in mind, we need to split our dataset into the matrix of independent variables and the vector or dependent variable. Mathematically, Vector is defined as a matrix that has just one column.

Let's split our dataset into independent and dependent variables.

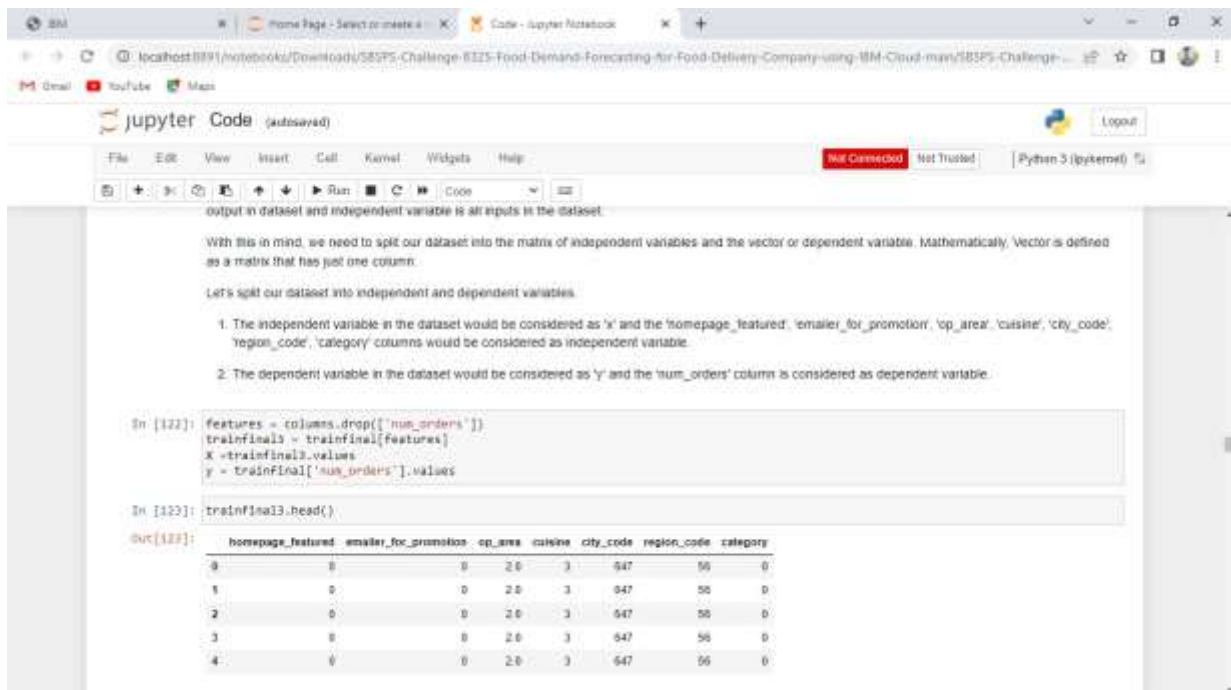
1. The independent variable in the dataset would be considered as 'x' and the 'homepage\_featured', 'emailer\_for\_promotion', 'op\_area', 'cuisine', 'city\_code', 'region\_code', 'category' columns would be considered as independent variable.
2. The dependent variable in the dataset would be considered as 'y' and the 'num\_orders' column is considered as dependent variable.

```
In [122]: features = columns.drop(['num_orders'])
          trainfinal3 = trainfinal[features]
          x = trainfinal3.values
          y = trainfinal['num_orders'].values

In [123]: trainfinal3.head()
```

Out[123]:

	homepage_featured	emailer_for_promotion	op_area	cuisine	city_code	region_code	category
0	0	0	0	2.0	3	647	56
1	0	0	0	2.0	3	647	56
2	0	0	0	2.0	3	647	56



The screenshot shows a Jupyter Notebook interface with the following content:

output in dataset and independent variable is all inputs in the dataset.

With this in mind, we need to split our dataset into the matrix of independent variables and the vector or dependent variable. Mathematically, Vector is defined as a matrix that has just one column.

Let's split our dataset into independent and dependent variables.

1. The independent variable in the dataset would be considered as 'x' and the 'homepage\_featured', 'emailer\_for\_promotion', 'op\_area', 'cuisine', 'city\_code', 'region\_code', 'category' columns would be considered as independent variable.
2. The dependent variable in the dataset would be considered as 'y' and the 'num\_orders' column is considered as dependent variable.

```
In [122]: features = columns.drop(['num_orders'])
          trainfinal3 = trainfinal[features]
          x = trainfinal3.values
          y = trainfinal['num_orders'].values

In [123]: trainfinal3.head()
```

Out[123]:

	homepage_featured	emailer_for_promotion	op_area	cuisine	city_code	region_code	category
0	0	0	0	2.0	3	647	56
1	0	0	0	2.0	3	647	56
2	0	0	0	2.0	3	647	56
3	0	0	0	2.0	3	647	56
4	0	0	0	2.0	3	647	56