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
```

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	2.0	3	647	56	0
1	0	0	2.0	3	647	56	0
2	0	0	2.0	3	647	56	0
3	0	0	2.0	3	647	56	0
4	0	0	2.0	3	647	56	0