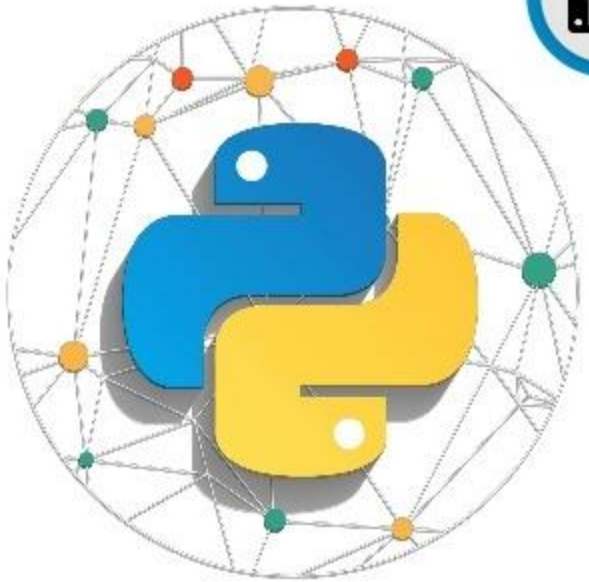
Logistic Regression

edureka!

https://www.edureka.co/ python

Topics Covered in Today's Training



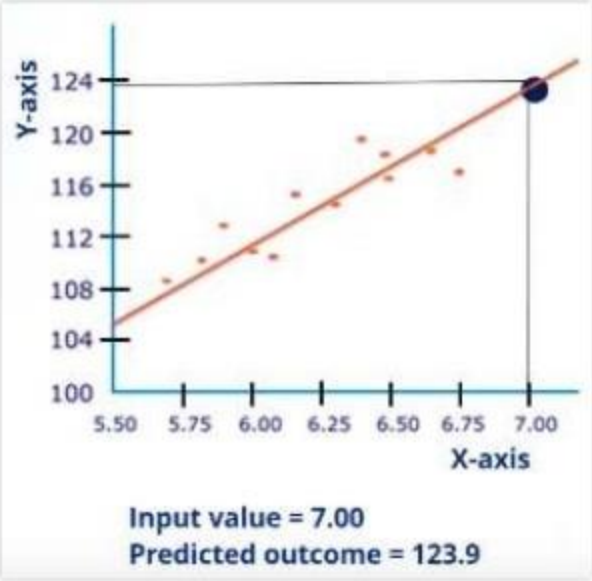
- What is Regression?
- Logistic regression: What and Why?
- Linear Vs Logistic Regression
- Use- Cases
- Demo

What Is Regression

What is Regression?

Regression Analysis is a predictive modelling technique

It estimates the relationship between a dependent (target) and an independent variable(predictor)

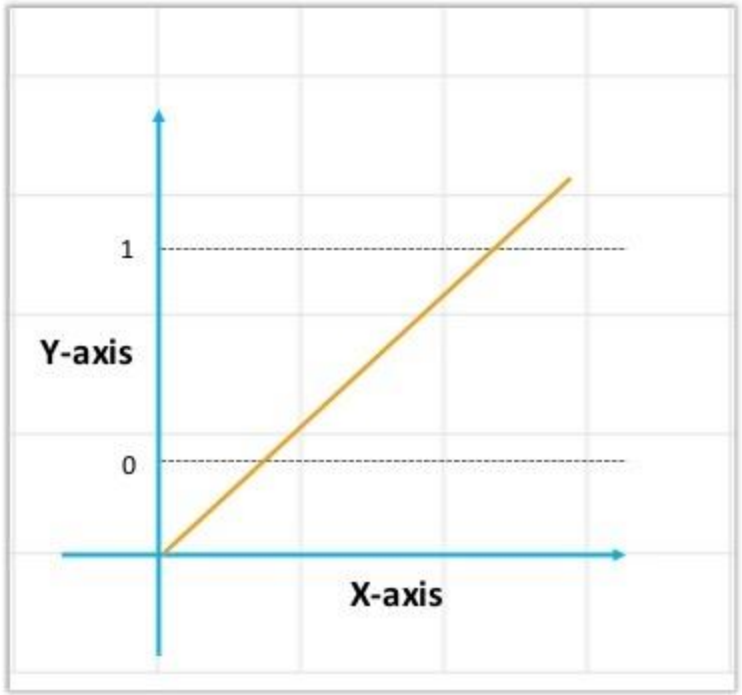


Logistic Regression: What And Why

Logistic Regression: What And Why?

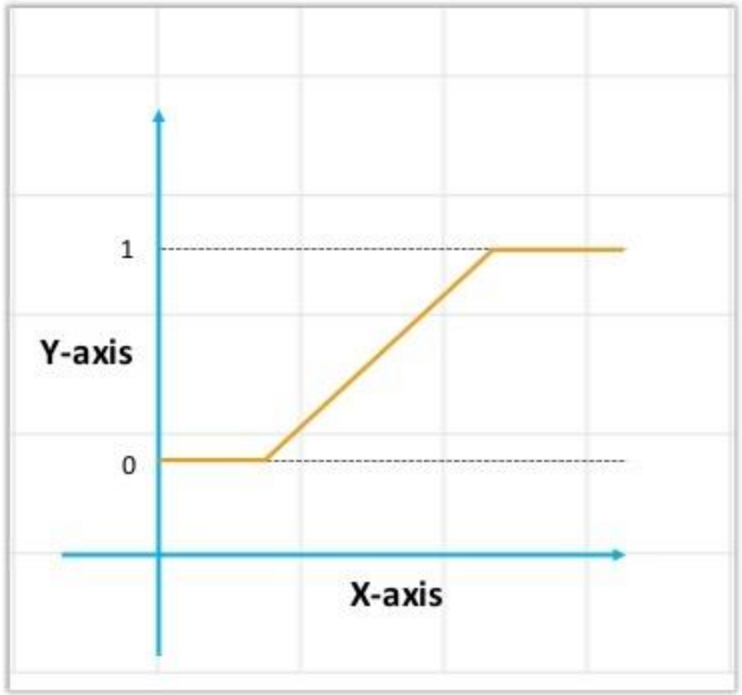
Logistic Regression produces results in a binary format which is used to predict the outcome of a categorical dependent variable. So the outcome should be discrete/ categorical such as:

Why Not Linear Regression?



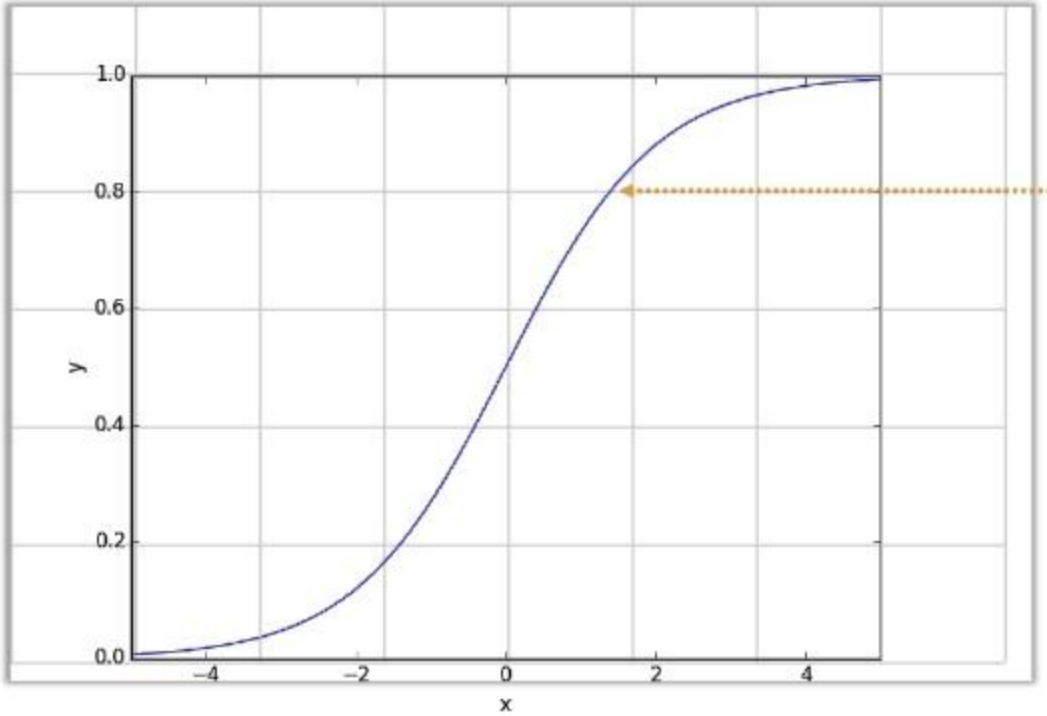
Since our value of Y will be between 0 and 1, the linear line has to be clipped at 0 and 1.

Why Not Linear Regression?



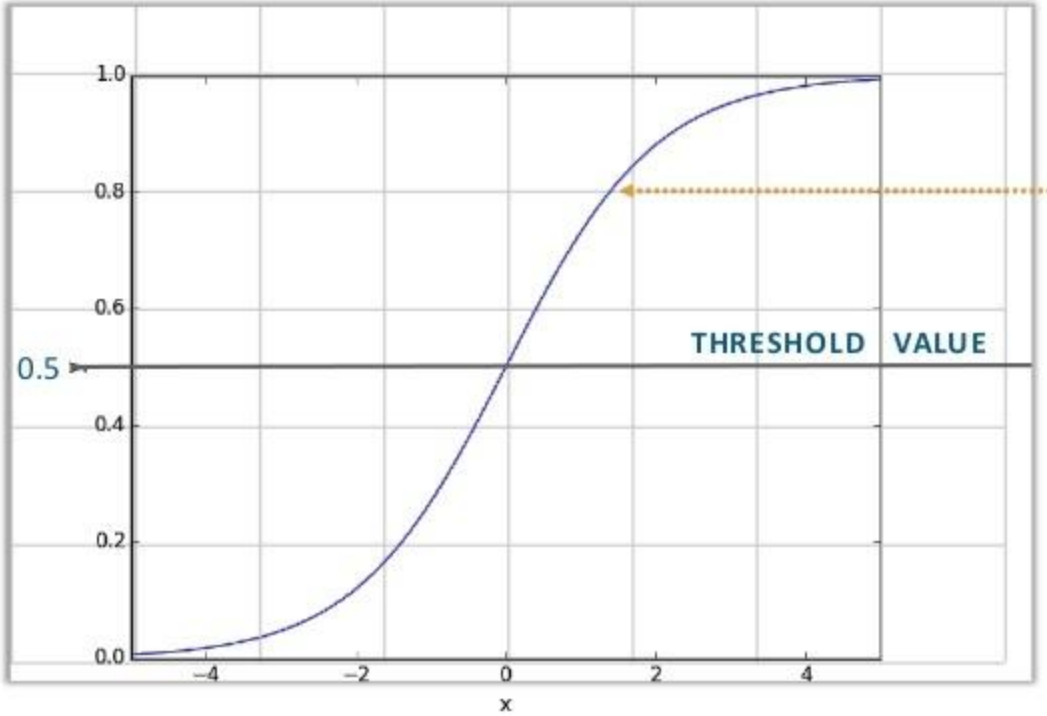
With this, our resulting curve cannot be formulated into a single formula. Hence we came up with **Logistic**!

Logistic Regression Curve



The Sigmoid “S” Curve

Logistic Regression Curve



The Sigmoid "S" Curve

With this, the threshold value indicates the probability of winning or losing

Logistic Regression Equation

The Logistic Regression Equation is derived from the Straight Line Equation

Equation of a straight line

$Y = C + B1X1 + B2X2 + \ldots$ ⟶ Range is from –(infinity) to (infinity)

Logistic Regression Equation

The Logistic Regression Equation is derived from the Straight Line Equation

Equation of a straight line

$Y = C + BIX1 + B2X2 + \ldots.$ ⟶ Range is from –(infinity) to (infinity)

Let's try to reduce the Logistic Regression Equation from Straight Line Equation

$Y = C + BIX1 + B2X2 + \ldots.$ ⟶ In Logistic equation Y can be only from 0 to 1

Logistic Regression Equation

The Logistic Regression Equation is derived from the Straight Line Equation

Equation of a straight line

$Y = C + BIX1 + B2X2 + \ldots.$ ⟶ Range is from –(infinity) to (infinity)

Let's try to reduce the Logistic Regression Equation from Straight Line Equation

$Y = C + BIX1 + B2X2 + \ldots.$ ⟶ In Logistic equation Y can be only from 0 to 1

Now , to get the range of Y between 0 and infinity, let's transform Y

$\frac{Y}{1-Y}$ } Y= 0 then 0, Y= 1 then infinity ⟶ Now, the range is between 0 to infinity

Logistic Regression Equation

The Logistic Regression Equation is derived from the Straight Line Equation

Equation of a straight line

$Y = C + BIX1 + B2X2 + \ldots.$ ⟶ Range is from –(infinity) to (infinity)

Let's try to reduce the Logistic Regression Equation from Straight Line Equation

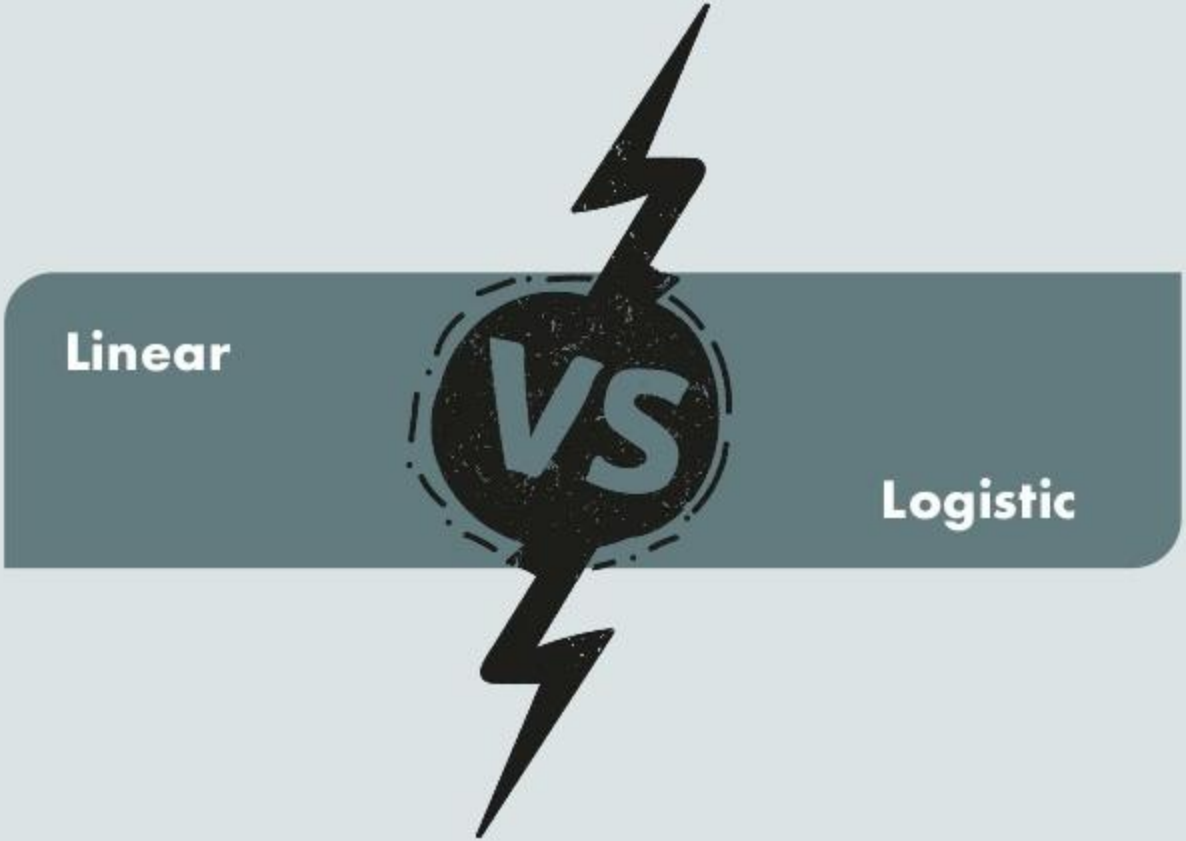
$Y = C + BIX1 + B2X2 + \ldots.$ ⟶ In Logistic equation Y can be only from 0 to 1

Now , to get the range of Y between 0 and infinity, let's transform Y

Y Y= 0 then 0
$1-Y$ Y= 1 then infinity ⟶ Now, the range is between 0 to infinity

Let us transform it further, to get range between –(infinity) and (infinity)

$\log \left[\frac{Y}{1-Y}\right] \Rightarrow Y = C + BIX1 + B2X2 + \ldots.$ ⟶ **Final Logistic Regression Equation**


Linear
VS
Logistic

Linear Vs Logistic Regression

Linear Regression

1. Continuous variables
2. Solves Regression Problems
3. Straight line

Logistic Regression

1. Categorical variables
2. Solves Classification Problems
3. S-Curve

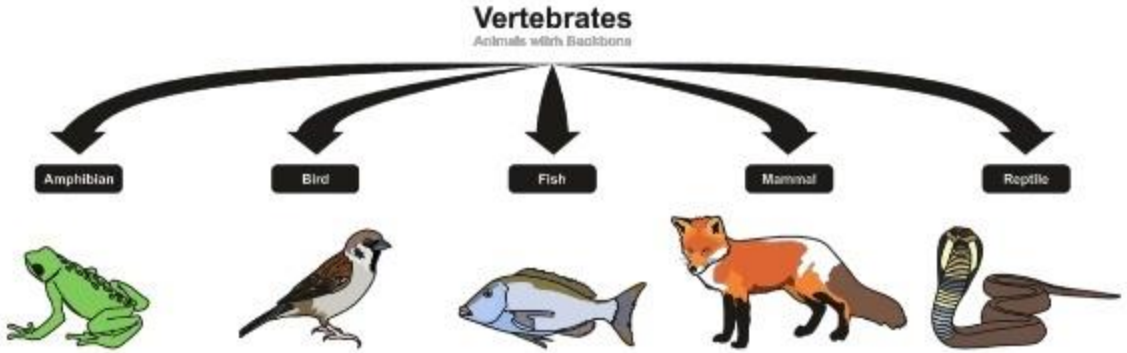
Logistic Regression: Use - Cases

Logistic Regression: Use - Cases

Weather Predictions

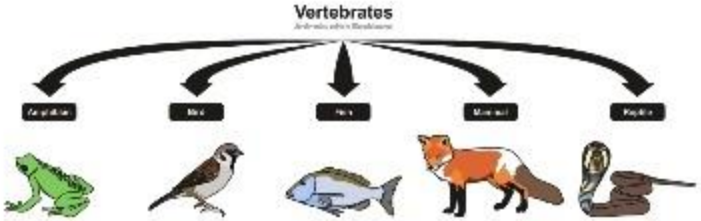
Logistic Regression: Use - Cases

Classification Problems



Logistic Regression: Use - Cases

Determines Illness

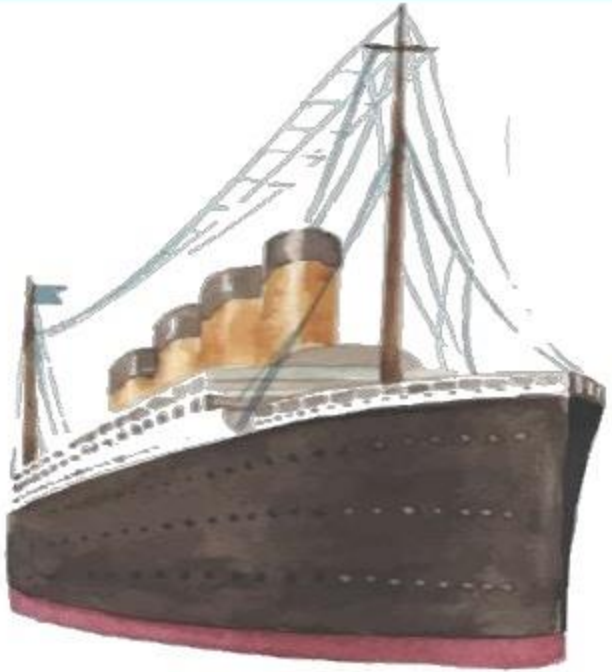


Demo

Titanic Data Analysis

1

Explore titanic dataset and explore about the people, both those who survived and those who did not. With today's technology, answering questions through data analysis is now easier than ever.

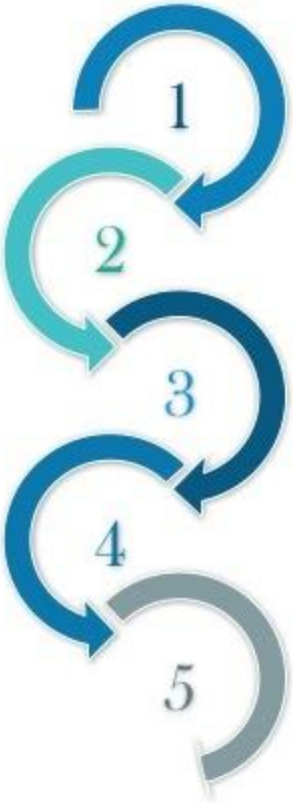


What factors made people more likely to survive the sinking of the Titanic?

Titanic Dataset

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN	S
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	C
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7000	G6	S
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500	C103	S
12	13	0	3	Saundercock, Mr. William Henry	male	20.0	0	0	A/5. 2151	8.0500	NaN	S
13	14	0	3	Andersson, Mr. Anders Johan	male	39.0	1	5	347082	31.2750	NaN	S
14	15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14.0	0	0	350406	7.8542	NaN	S

Implement Logistic Regression



Collecting Data

Analyzing Data

Data Wrangling

Train & Test

Accuracy Check

Implement Logistic Regression

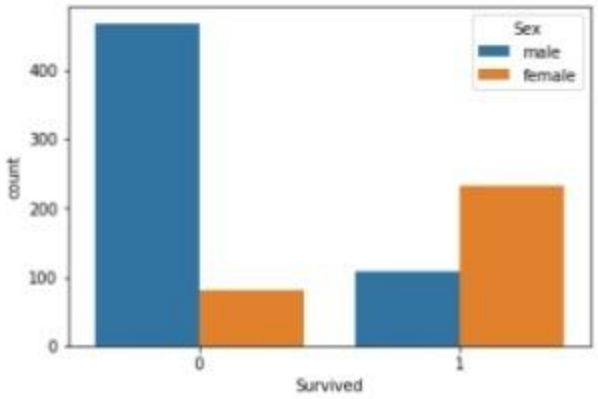
Collect Data: Import Libraries

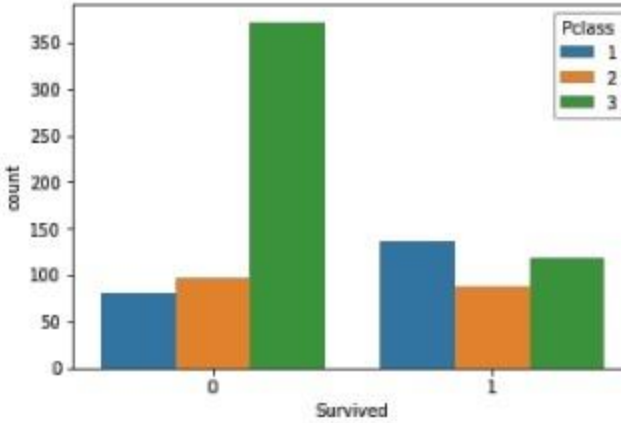
```
In [33]: import pandas as pd
         import numpy as np
         import seaborn as sns
         import matplotlib.pyplot as plt
         import math
         %matplotlib inline

         titanic_data= pd.read_csv('Titanic.csv')
```

Implement Logistic Regression

Analyzing Data

Creating different plot to check relationship between variables

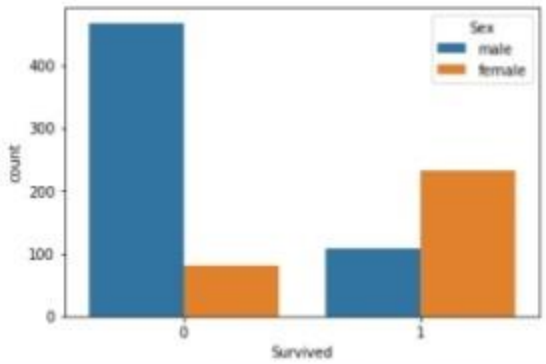


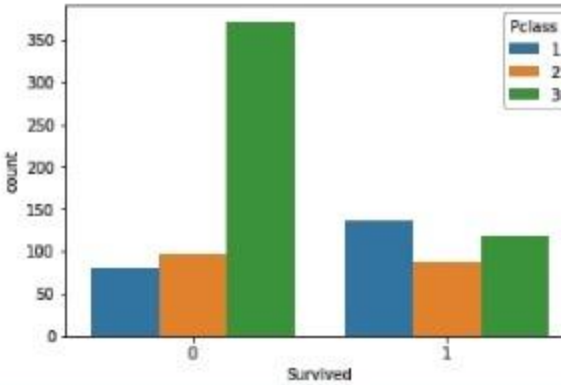


Implement Logistic Regression

Analyzing Data

Creating different plot to check relationship between variables





Implement Logistic Regression

Data Wrangling

Clean the data by removing the Nan values and unnecessary columns in the dataset

	Survived	Age	SibSp	Parch	Fare	male	Q	S	2	3
0	0	22.0	1	0	7.2500	1	0	1	0	1
1	1	38.0	1	0	71.2833	0	0	0	0	0
2	1	26.0	0	0	7.9250	0	0	1	0	1
3	1	35.0	1	0	53.1000	0	0	1	0	0
4	0	35.0	0	0	8.0500	1	0	1	0	1

Implement Logistic Regression

Train & Test Data

Build the model on the train data and predict the output on the test data

```
logistic = LogisticRegression()
logistic.fit(train_X,train_Y)
```

Implement Logistic Regression

Accuracy Check

Calculate accuracy to check how accurate your results are.

```
from sklearn.metrics import accuracy_sc
accuracy_score(y_test,predictions)*100
```

SUV Data Analysis

2 A car company has released a new SUV in the market. Using the previous data about the sales of their SUV's, they want to predict the category of people who might be interested in buying this.

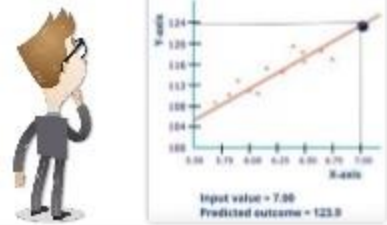
What factors made people more interested in buying SUV?

SUV Predictions

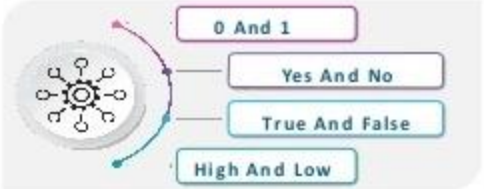
	User ID	Gender	Age	Estimated Salary	Purchased
0	15624510	Male	19	19000	0
1	15810944	Male	35	20000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0
5	15728773	Male	27	58000	0
6	15598044	Female	27	84000	0
7	15694829	Female	32	150000	1
8	15600575	Male	25	33000	0
9	15727311	Female	35	65000	0

Session in a Minute

What is regression?



Logistic: What & Why?



Linear VS Logistic

Use - Cases

Demo1: Titanic Analysis

Demo2: SUV Prediction

Don’t just learn it, **MASTER** it with

edureka!

Thank You

For more information please visit our website
www.edureka.co