In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
rissh = pd.read_csv("train.csv")
rissh.head()
```

Out[1]:

	User_ID	Description	Browser_Used	Device_Used	Is_Response
0	id10326	The room was kind of clean but had a VERY stro...	Edge	Mobile	not happy
1	id10327	I stayed at the Crown Plaza April -- - April -...	Internet Explorer	Mobile	not happy
2	id10328	I booked this hotel through Hotwire at the low...	Mozilla	Tablet	not happy
3	id10329	Stayed here with husband and sons on the way t...	InternetExplorer	Desktop	happy
4	id10330	My girlfriends and I stayed here to celebrate ...	Edge	Tablet	not happy

In [2]:
```
rissh.shape
```

Out[2]:
```
(38932, 5)
```

In [3]:
```
rissh.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 38932 entries, 0 to 38931
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   User_ID       38932 non-null  object
 1   Description   38932 non-null  object
 2   Browser_Used  38932 non-null  object
 3   Device_Used   38932 non-null  object
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

```
 4   Is_Response   38932 non-null  object
dtypes: object(5)
memory usage: 1.5+ MB
```

In [4]:
```
rissh.describe().transpose()
```

Out[4]:

	count	unique	top	freq
User_ID	38932	38932	id16127	1
Description	38932	38932	Pros:\n-I booked the room for - nights two mon...	1
Browser_Used	38932	11	Firefox	7367
Device_Used	38932	3	Desktop	15026
Is_Response	38932	2	happy	26521

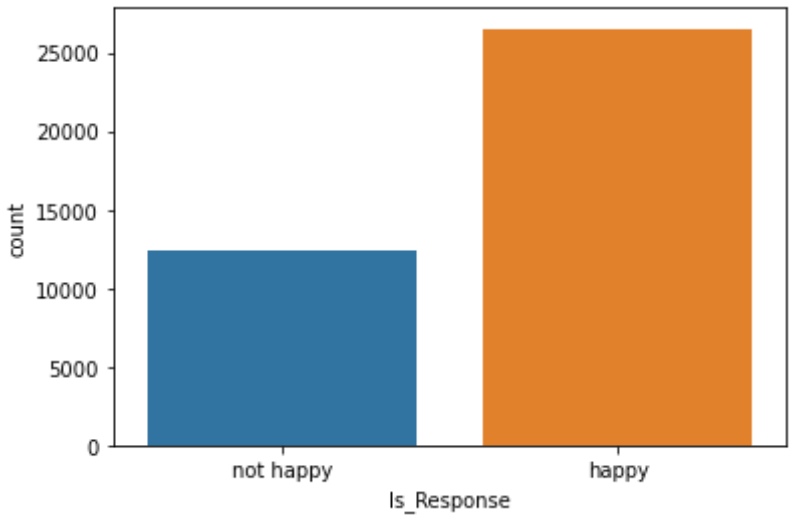
In [5]:
```
rissh.isnull().sum()      #Checking whether the dataset contains missin
g values or not
```

Out[5]:
```
User_ID          0
Description      0
Browser_Used     0
Device_Used      0
Is_Response      0
dtype: int64
```

In [6]:
```
import seaborn as sns
sns.countplot(x="Is_Response",data=rissh)
```

Out[6]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x158c67ec5e0>
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD



In [8]:
```
rissh.drop(columns=["User_ID","Browser_Used","Device_Used"],inplace=True)     # Removing useless columns
rissh.head(10)
```

Out[8]:

	Description	Is_Response
0	The room was kind of clean but had a VERY stro...	not happy
1	I stayed at the Crown Plaza April -- - April -...	not happy
2	I booked this hotel through Hotwire at the low...	not happy
3	Stayed here with husband and sons on the way t...	happy
4	My girlfriends and I stayed here to celebrate ...	not happy
5	We had - rooms. One was very nice and clearly ...	happy
6	My husband and I have stayed in this hotel a f...	not happy
7	My wife & I stayed in this glorious city a whi...	happy
8	My boyfriend and I stayed at the Fairmont on a...	happy
9	Wonderful staff, great location, but it was de...	not happy

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

In [9]:
```
import re                       #Apply first level of cleaning
import string
# Convert the data into lower-case, removes square bracketts, remove nu
mbers and punctuations
def text_clean1(text):
    text = text.lower()
    text = re.sub("\[.*?\]","",text)
    text = re.sub("[%s]" % re.escape(string.punctuation),"",text)
    text = re.sub("\w*\d\w*","",text)
    return text

clean1 = lambda x : text_clean1(x)
```

In [10]:
```
# Apply the first clean to the dataset
rissh["cleaned_Description"] = pd.DataFrame(rissh.Description.apply(cle
an1))
rissh.head(10)
```

Out[10]:

	Description	Is_Response	cleaned_Description
0	The room was kind of clean but had a VERY stro...	not happy	the room was kind of clean but had a very stro...
1	I stayed at the Crown Plaza April -- - April -...	not happy	i stayed at the crown plaza april april th...
2	I booked this hotel through Hotwire at the low...	not happy	i booked this hotel through hotwire at the low...
3	Stayed here with husband and sons on the way t...	happy	stayed here with husband and sons on the way t...
4	My girlfriends and I stayed here to celebrate ...	not happy	my girlfriends and i stayed here to celebrate ...
5	We had - rooms. One was very nice and clearly ...	happy	we had rooms one was very nice and clearly ha...
6	My husband and I have stayed in this hotel a f...	not happy	my husband and i have stayed in this hotel a f...
7	My wife & I stayed in this glorious city a whi...	happy	my wife i stayed in this glorious city a whil...
8	My boyfriend and I stayed at the Fairmont on a...	happy	my boyfriend and i stayed at the fairmont on a...

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

	Description	Is_Response	cleaned_Description
9	Wonderful staff, great location, but it was de...	not happy	wonderful staff great location but it was defi...

In [12]:
```
# Apply second round of cleaning for removing the " " and new line
def text_clean2(text):
    text = re.sub("['""....]",'',text)
    text = re.sub("\n","",text)
    return text

clean2 = lambda x : text_clean2(x)
```

In [15]:
```
# See the updated text
rissh["Updated_Description"] = pd.DataFrame(rissh["cleaned_Description"
].apply(clean2))
rissh.head(10)
```

Out[15]:

	Description	Is_Response	cleaned_Description	Updated Description	Updated_Description
0	The room was kind of clean but had a VERY stro...	not happy	the room was kind of clean but had a very stro...	the room was kind of clean but had a very stro...	the room was kind of clean but had a very stro...
1	I stayed at the Crown Plaza April -- - April -...	not happy	i stayed at the crown plaza april april th...	i stayed at the crown plaza april april th...	i stayed at the crown plaza april april th...
2	I booked this hotel through Hotwire at the low...	not happy	i booked this hotel through hotwire at the low...	i booked this hotel through hotwire at the low...	i booked this hotel through hotwire at the low...
3	Stayed here with husband and sons on the way t...	happy	stayed here with husband and sons on the way t...	stayed here with husband and sons on the way t...	stayed here with husband and sons on the way t...
4	My girlfriends and I stayed here to celebrate ...	not happy	my girlfriends and i stayed here to celebrate ...	my girlfriends and i stayed here to celebrate ...	my girlfriends and i stayed here to celebrate ...

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

	Description	Is_Response	cleaned_Description	Updated Description	Updated_Description
5	We had - rooms. One was very nice and clearly ...	happy	we had rooms one was very nice and clearly ha...	we had rooms one was very nice and clearly ha...	we had rooms one was very nice and clearly ha...
6	My husband and I have stayed in this hotel a f...	not happy	my husband and i have stayed in this hotel a f...	my husband and i have stayed in this hotel a f...	my husband and i have stayed in this hotel a f...
7	My wife & I stayed in this glorious city a whi...	happy	my wife i stayed in this glorious city a whil...	my wife i stayed in this glorious city a whil...	my wife i stayed in this glorious city a whil...
8	My boyfriend and I stayed at the Fairmont on a...	happy	my boyfriend and i stayed at the fairmont on a...	my boyfriend and i stayed at the fairmont on a...	my boyfriend and i stayed at the fairmont on a...
9	Wonderful staff, great location, but it was de...	not happy	wonderful staff great location but it was defi...	wonderful staff great location but it was defi...	wonderful staff great location but it was defi...

In [19]:
```
from sklearn.model_selection import train_test_split
Independent_Var = rissh.Updated_Description
Dependent_Var = rissh.Is_Response
IV_train, IV_test, DV_train, DV_test = train_test_split(Independent_Var
,Dependent_Var,test_size = 0.1,random_state=225)
print("IV_train :" , len(IV_train))
print("IV_test :",len(IV_test))
print("DV_train:",len(DV_train))
print("DV_test:",len(DV_test))
```

```
IV_train : 35038
IV_test : 3894
DV_train: 35038
DV_test: 3894
```

In [20]:
```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

```
tvec = TfidfVectorizer()
clf2 = LogisticRegression(solver = "lbfgs")

from sklearn.pipeline import Pipeline
```

In [23]:
```
model = Pipeline([("Vectorizer",tvec),("Classifier",clf2)])
model.fit(IV_train,DV_train)

from sklearn.metrics import confusion_matrix    # To check the accuracy
 of the model
predictions = model.predict(IV_test)
print(confusion_matrix(predictions,DV_test))
```

```
C:\ProgramData\Anaconda3\lib\site-packages\sklearn\linear_model\_logist
ic.py:762: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown
in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression
  n_iter_i = _check_optimize_result(
```

```
[[2417  304]
 [ 154 1019]]
```

Model Predictions

In [29]:
```
from sklearn.metrics import accuracy_score,precision_score,recall_score

print("The accuracy is :", accuracy_score(predictions, DV_test))
print("The precisioon score is :",precision_score(predictions, DV_test,
 average="weighted"))
print("The Recall score is :",recall_score(predictions,DV_test,average=
"weighted"))
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

```
The accuracy is : 0.8823831535695943
The precisioon score is : 0.8889271415963718
The Recall score is : 0.8823831535695943
```

Predictions on New Reviews

In [38]:
```
customer_1 = ["It is good product"]
result = model.predict(customer_1)
print(result)
```

```
['happy']
```

In [39]:
```
customer_2 = ["I am frustated"]
result = model.predict(customer_2)
print(result)
```

```
['not happy']
```

In [ ]:
```

```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD