

```
In [1]: import pandas as pd
```

```
In [2]: df = pd.read_csv('vgsales.csv')
```

```
In [3]: df.describe()
```

Out[3]:

	Rank	Year	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
	16598.000000	16327.000000	16598.000000	16598.000000	16598.000000	16598.000000	16598.000000
	8300.605254	2006.406443	0.264667	0.146652	0.077782	0.048063	0.537441
	4791.853933	5.828981	0.816683	0.505351	0.309291	0.188588	1.555028
	1.000000	1980.000000	0.000000	0.000000	0.000000	0.000000	0.010000
	4151.250000	2003.000000	0.000000	0.000000	0.000000	0.000000	0.060000
	8300.500000	2007.000000	0.080000	0.020000	0.000000	0.010000	0.170000
	12449.750000	2010.000000	0.240000	0.110000	0.040000	0.040000	0.470000
	16600.000000	2020.000000	41.490000	29.020000	10.220000	10.570000	82.740000

```
In [7]: df = df.dropna()
```

```
In [11]: df.columns
```

Out[11]: Index(['Rank', 'Name', 'Platform', 'Year', 'Genre', 'Publisher', 'NA_Sales', 'EU_Sales', 'JP_Sales', 'Other_Sales', 'Global_Sales'], dtype='object')

```
In [13]: df.head()
```

Out[13]:

	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sal
	1	Wii Sports	Wii	2006.0	Sports	Nintendo	41.49	29.02	3.77	8.
	2	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.58	6.81	0.
	3	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.85	12.88	3.79	3.
	4	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.75	11.01	3.28	2.
	5	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	11.27	8.89	10.22	1.

```
In [14]: df.shape
```

Out[14]: (16291, 11)

```
In [23]: prices = ['NA_Sales', 'EU_Sales', 'JP_Sales', 'Other_Sales']
```

```
In [24]: #from Kaggle
def std_based(col_name,df):
    mean = df[col_name].mean()
    std = df[col_name].std()
    cut_off = std * 2
    lower, upper = mean - cut_off, mean + cut_off
    new_df = df[(df[col_name] < upper) & (df[col_name] > lower)]
    return new_df
```

```
In [25]: for col in prices:
        df = std_based(col, df)
```

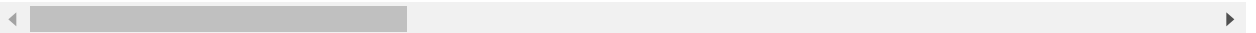
```
In [15]: df['Year'] = df['Year'].astype(int)
```

```
In [16]: # one hot encoding
one_hot = pd.get_dummies(df,columns=["Platform", "Genre", 'Publisher'])
one_hot.head()
```

Out[16]:

	Rank	Name	Year	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales	Platform_
0	1	Wii Sports	2006	41.49	29.02	3.77	8.46	82.74	
1	2	Super Mario Bros.	1985	29.08	3.58	6.81	0.77	40.24	
2	3	Mario Kart Wii	2008	15.85	12.88	3.79	3.31	35.82	
3	4	Wii Sports Resort	2009	15.75	11.01	3.28	2.96	33.00	
4	5	Pokemon Red/Pokemon Blue	1996	11.27	8.89	10.22	1.00	31.37	

5 rows × 627 columns



```
In [17]: one_hot.dtypes
```

```
Out[17]: Rank          int64
Name          object
Year          int32
NA_Sales      float64
EU_Sales      float64
...
Publisher_id Software      uint8
Publisher_imageepoch Inc.  uint8
Publisher_inXile Entertainment  uint8
Publisher_mixi, Inc        uint8
Publisher_responDESIGN     uint8
Length: 627, dtype: object
```


In [35]: x

Out[35]:

	Rank	Year	Platform_2600	Platform_3DO	Platform_3DS	Platform_DC	Platform_DS	Platform_DS	Platform_DS
0	1	2006	0	0	0	0	0	0	0
1	2	1985	0	0	0	0	0	0	0
2	3	2008	0	0	0	0	0	0	0
3	4	2009	0	0	0	0	0	0	0
4	5	1996	0	0	0	0	0	0	0
...
16593	16596	2002	0	0	0	0	0	0	0
16594	16597	2003	0	0	0	0	0	0	0
16595	16598	2008	0	0	0	0	0	0	0
16596	16599	2010	0	0	0	0	0	1	0
16597	16600	2003	0	0	0	0	0	0	0

16291 rows × 621 columns

```

In [32]: x_train_na, x_test_na, y_train_na, y_test_na = train_test_split(x, y_na, test_size=0.2,
x_train_eu, x_test_eu, y_train_eu, y_test_eu = train_test_split(x, y_eu, test_size=0.2,
x_train_jp, x_test_jp, y_train_jp, y_test_jp = train_test_split(x, y_jp, test_size=0.2,
x_train_other, x_test_other, y_train_other, y_test_other = train_test_split(x, y_other, test_size=0.2,

```

In [34]: model = LogisticRegression()

```
In [36]: scores_na = cross_val_score(model, x, y_na, cv=10)
print("CV scores of North America = {}".format(scores_na))
print("mean CV score of North America= {}".format(scores_na.mean()))
```

C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model_logistic.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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
CV scores of North America = [0.1993865  0.37507673 0.78514426 0.83855126 0.726
82627 0.78575813
0.72559853 0.7170043  0.56721915 0.4014733 ]
mean CV score of North America= 0.6122038436769143
```



```
In [37]: scores_eu = cross_val_score(model, x, y_eu, cv=10)
print("CV scores of Europe = {}".format(scores_eu))
print("mean CV score of Europe = {}".format(scores_eu.mean()))
```

C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model_logistic.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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:7
62: 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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
```

```
CV scores of Europe = [0.11472393 0.72559853 0.76058932 0.80417434 0.82197667
0.8004911
0.76058932 0.74769797 0.67403315 0.61325967]
mean CV score of Europe = 0.6823133993906457
```

```
C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:7
62: 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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
```



```
In [38]: scores_jp = cross_val_score(model, x, y_jp, cv=10)
print("CV scores of Japan = {}".format(scores_jp))
print("mean CV score of Japan = {}".format(scores_jp.mean()))
```

C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model_logistic.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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:7
62: 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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:7
62: 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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:7
62: 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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
```

```
CV scores of Japan = [0.52822086 0.7980356 0.82750153 0.82259055 0.80662983 0.
76918355
```

```
0.80110497 0.76979742 0.74585635 0.75015347]
```

```
mean CV score of Japan = 0.7619074143119156
```

```
C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.p
y: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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
```

```
In [39]: scores_other = cross_val_score(model, x, y_other, cv=10)
print("CV scores of Other = {}".format(scores_other))
print("mean CV score of Other = {}".format(scores_other.mean()))
```

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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
```

C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model_logistic.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> (<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

```
In [40]: params = {'C': [0.1, 1.0, 10.0]}
lr_grid_na = GridSearchCV(model, param_grid = params, scoring='accuracy', n_jobs=
lr_grid_eu = GridSearchCV(model, param_grid = params, scoring='accuracy', n_jobs=
lr_grid_jp = GridSearchCV(model, param_grid = params, scoring='accuracy', n_jobs=
lr_grid_other = GridSearchCV(model, param_grid = params, scoring='accuracy', n_jo
```

```
In [41]: lr_grid_na.fit(x_train_na, y_train_na)
print(lr_grid_na.score(x_test_na, y_test_na))
print(lr_grid_na.best_params_)
```

```
0.625038355323719
{'C': 1.0}
```

```
In [42]: lr_grid_eu.fit(x_train_eu, y_train_eu)
print(lr_grid_eu.score(x_test_eu, y_test_eu))
print(lr_grid_eu.best_params_)
```

```
0.7134090211721387
{'C': 10.0}
```

```
In [43]: lr_grid_jp.fit(x_train_jp, y_train_jp)
print(lr_grid_jp.score(x_test_jp, y_test_jp))
print(lr_grid_jp.best_params_)
```

```
0.7898128260202516
{'C': 0.1}
```

C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model_logistic.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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)
n_iter_i = _check_optimize_result(

```
In [44]: lr_grid_other.fit(x_train_other, y_train_other)
print(lr_grid_other.score(x_test_other, y_test_other))
print(lr_grid_other.best_params_)
```

```
0.8637618901503529
{'C': 0.1}
```

C:\Users\Sajjad\anaconda3\lib\site-packages\sklearn\linear_model_logistic.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> (<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 (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)
n_iter_i = _check_optimize_result(

```
In [ ]:
```