```python
from sklearn.neighbors import KNeighborsRegressor
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler

proj = pd.read_csv("/content/2008.csv")
proj = proj.dropna(axis=0)


cat_cols = proj.select_dtypes(exclude = 'number')
num_cols = proj.select_dtypes(include = 'number')
onehot_cat_cols = pd.get_dummies(cat_cols)
onehot_cat_cols.head()
proj = pd.concat([num_cols,onehot_cat_cols],sort=True,axis=1)


y = proj["DepDelay"]
x = proj.drop("DepDelay", axis=1)


x_train, x_test, y_train, y_test = train_test_split(x,y,test_size = 0.3,random_state=0)


from sklearn.neighbors import KNeighborsRegressor
from sklearn.metrics import mean_squared_error
from math import sqrt

rmse_val = []
for K in range(20):
    K = K+1
    model = KNeighborsRegressor(n_neighbors = K)
    model.fit(x_train, y_train)  #fit the model
    pred=model.predict(x_test) #make prediction on test set
    error = sqrt(mean_squared_error(y_test,pred)) #calculate rmse
    rmse_val.append(error) #store rmse values
    print('RMSE value for k= ' , K , 'is:', error)
```

↪

```
       RMSE value for k=   1 is: 9.882884533089396
       RMSE value for k=   2 is: 9.180074288731001
       RMSE value for k=   3 is: 9.227337017261211
       RMSE value for k=   4 is: 9.438633490965023
       RMSE value for k=   5 is: 9.700683339322628
       RMSE value for k=   6 is: 9.933216135708314
       RMSE value for k=   7 is: 10.200646548449974
       RMSE value for k=   8 is: 10.496199549161291
       RMSE value for k=   9 is: 10.726966458648558
       RMSE value for k=   10 is: 10.997371915236306
       RMSE value for k=   11 is: 11.231109459698727
       RMSE value for k=   12 is: 11.476727369333288
       RMSE value for k=   13 is: 11.670937574842903
       RMSE value for k=   14 is: 11.879700264038574
       RMSE value for k=   15 is: 12.069631457585404
       RMSE value for k=   16 is: 12.2590395802624
       RMSE value for k=   17 is: 12.433982291029233
       RMSE value for k=   18 is: 12.604380518596756
       RMSE value for k=   19 is: 12.782466444951758
       RMSE value for k=   20 is: 12.936243124827326
```
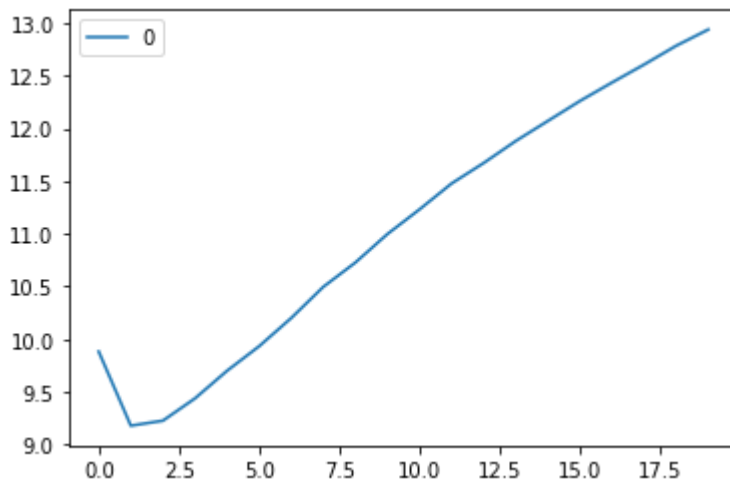
```python
import matplotlib.pyplot as plt
curve = pd.DataFrame(rmse_val)
curve.plot()
```

☐→  <matplotlib.axes._subplots.AxesSubplot at 0x7f5c2c43b240>



```python
model = KNeighborsRegressor(n_neighbors = 2)
model.fit(x_train, y_train)
pred=model.predict(x_test)
print("{} NN Score: {:.2f}%".format(2, model.score(x_test, y_test)*100))
```

☐→  2 NN Score: 93.51%

```python
from sklearn.model_selection import KFold, cross_val_score
model = KNeighborsRegressor(n_neighbors=2)
ab = KFold(10)
CV_Scores = cross_val_score(model, x, y, cv=ab).mean()
print(CV_Scores)
```

```
print(cv_scores)
```

0.8773527808368982