Tyler Echols

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

df = pd.read_csv('weatherAUS.csv')
df.head()

to_drop = ['Location', 'Date']
df.drop(to_drop, inplace=True, axis=1)

df['RainToday'] = df['RainToday'].fillna('No')
df['RainTomorrow'] = df['RainTomorrow'].fillna('No')
missing_values_numeric_features  = [col for col in df.columns if (df.isnull().sum()[c
def impute_means(df, missing_values_columns):
    data = df.copy()
    '''Filling missing values with mean'''
    for col in missing_values_columns:
        data[col] = data[col].fillna(data[col].mean())

    return data
df = impute_means(df,missing_values_numeric_features)
df.isnull().sum()

print(df)

df.plot(x="Rainfall", y="Humidity9am", kind="scatter")
```

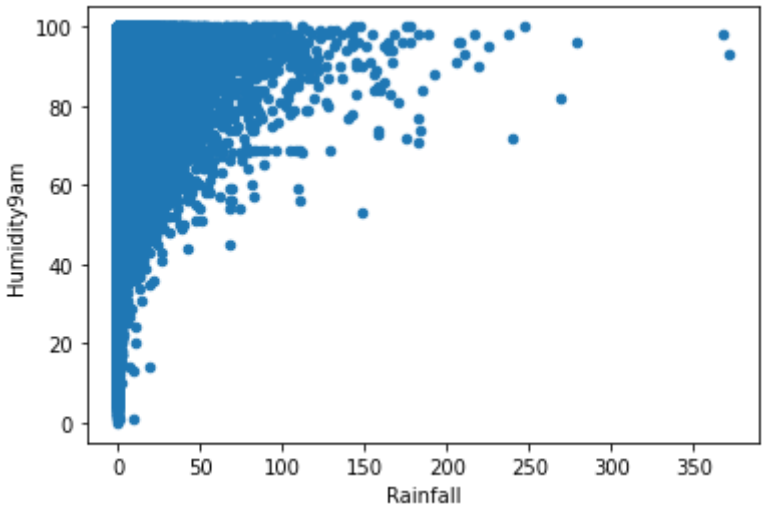
```
        MinTemp    MaxTemp  Rainfall  Evaporation  Sunshine WindGustDir  \
0          13.4  22.900000       0.6     5.468232  7.611178           W
1           7.4  25.100000       0.0     5.468232  7.611178         WNW
2          12.9  25.700000       0.0     5.468232  7.611178         WSW
3           9.2  28.000000       0.0     5.468232  7.611178          NE
4          17.5  32.300000       1.0     5.468232  7.611178           W
...         ...        ...       ...          ...       ...         ...
145455      2.8  23.400000       0.0     5.468232  7.611178           E
145456      3.6  25.300000       0.0     5.468232  7.611178         NNW
145457      5.4  26.900000       0.0     5.468232  7.611178           N
145458      7.8  27.000000       0.0     5.468232  7.611178          SE
145459     14.9  23.221348       0.0     5.468232  7.611178         NaN

        WindGustSpeed WindDir9am WindDir3pm  WindSpeed9am  ...  Humidity9am  \
0            44.00000          W        WNW          20.0  ...         71.0
1            44.00000        NNW        WSW           4.0  ...         44.0
2            46.00000          W        WSW          19.0  ...         38.0
3            24.00000         SE          E          11.0  ...         45.0
4            41.00000        ENE         NW           7.0  ...         82.0
...               ...        ...        ...           ...  ...          ...
145455       31.00000         SE        ENE          13.0  ...         51.0
145456       22.00000         SE          N          13.0  ...         56.0
145457       37.00000         SE        WNW           9.0  ...         53.0
145458       28.00000        SSE          N          13.0  ...         51.0
145459       40.03523        ESE        ESE          17.0  ...         62.0

        Humidity3pm  Pressure9am  Pressure3pm  Cloud9am  Cloud3pm  Temp9am  \
0              22.0       1007.7       1007.1  8.000000   4.50993     16.9
1              25.0       1010.6       1007.8  4.447461   4.50993     17.2
2              30.0       1007.6       1008.7  4.447461   2.00000     21.0
3              16.0       1017.6       1012.8  4.447461   4.50993     18.1
4              33.0       1010.8       1006.0  7.000000   8.00000     17.8
...             ...          ...          ...       ...       ...      ...
145455         24.0       1024.6       1020.3  4.447461   4.50993     10.1
145456         21.0       1023.5       1019.1  4.447461   4.50993     10.9
145457         24.0       1021.0       1016.8  4.447461   4.50993     12.5
145458         24.0       1019.4       1016.5  3.000000   2.00000     15.1
145459         36.0       1020.2       1017.9  8.000000   8.00000     15.0

        Temp3pm  RainToday RainTomorrow
0          21.8         No          No
1          24.3         No          No
2          23.2         No          No
3          26.5         No          No
4          29.7         No          No
...         ...        ...         ...
145455     22.4         No          No
145456     24.5         No          No
145457     26.1         No          No
145458     26.0         No          No
145459     20.9         No          No

[145460 rows x 21 columns]
<matplotlib.axes._subplots.AxesSubplot at 0x7f36ef580410>
```

```
import pandas as pd
```

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report, confusion_matrix

df = pd.read_csv('weatherAUS.csv')
df.head()

to_drop = ['Location', 'Date']
df.drop(to_drop, inplace=True, axis=1)

df['RainToday'] = df['RainToday'].fillna('No')
df['RainTomorrow'] = df['RainTomorrow'].fillna('No')
missing_values_numeric_features  = [col for col in df.columns if (df.isnull().sum()[c
def impute_means(df, missing_values_columns):
    data = df.copy()
    '''Filling missing values with mean'''
    for col in missing_values_columns:
        data[col] = data[col].fillna(data[col].mean())

    return data
df = impute_means(df,missing_values_numeric_features)
df.isnull().sum()

print(df)

df.plot(x="Rainfall", y="Humidity9am", kind="scatter")
df.head()
df.info()


x = np.arange(10).reshape(-1, 1)
y = np.array([0, 0, 0, 0, 1, 1, 1, 1, 1, 1])

df = LogisticRegression(solver= 'liblinear', random_state=0).fit(x,y)
df.fit(x, y)

df.classes_
df.intercept_
df.coef_
df.predict_proba(x)
df.predict(x)
df.score(x, y)
confusion_matrix(y, df.predict(x))
```

```
        MinTemp    MaxTemp  Rainfall  Evaporation  Sunshine WindGustDir  \
0          13.4  22.900000       0.6     5.468232  7.611178           W
1           7.4  25.100000       0.0     5.468232  7.611178         WNW
2          12.9  25.700000       0.0     5.468232  7.611178         WSW
3           9.2  28.000000       0.0     5.468232  7.611178          NE
4          17.5  32.300000       1.0     5.468232  7.611178           W
...         ...        ...       ...          ...       ...         ...
145455      2.8  23.400000       0.0     5.468232  7.611178           E
145456      3.6  25.300000       0.0     5.468232  7.611178         NNW
145457      5.4  26.900000       0.0     5.468232  7.611178           N
145458      7.8  27.000000       0.0     5.468232  7.611178          SE
145459     14.9  23.221348       0.0     5.468232  7.611178         NaN

        WindGustSpeed WindDir9am WindDir3pm  WindSpeed9am  ...  Humidity9am  \
0            44.00000          W        WNW          20.0  ...         71.0
1            44.00000        NNW        WSW           4.0  ...         44.0
2            46.00000          W        WSW          19.0  ...         38.0
3            24.00000         SE          E          11.0  ...         45.0
4            41.00000        ENE         NW           7.0  ...         82.0
...               ...        ...        ...           ...  ...          ...
145455       31.00000         SE        ENE          13.0  ...         51.0
145456       22.00000         SE          N          13.0  ...         56.0
145457       37.00000         SE        WNW           9.0  ...         53.0
145458       28.00000        SSE          N          13.0  ...         51.0
145459       40.03523        ESE        ESE          17.0  ...         62.0

        Humidity3pm  Pressure9am  Pressure3pm  Cloud9am  Cloud3pm  Temp9am  \
0              22.0       1007.7       1007.1  8.000000   4.50993     16.9
1              25.0       1010.6       1007.8  4.447461   4.50993     17.2
2              30.0       1007.6       1008.7  4.447461   2.00000     21.0
3              16.0       1017.6       1012.8  4.447461   4.50993     18.1
4              33.0       1010.8       1006.0  7.000000   8.00000     17.8
...             ...          ...          ...       ...       ...      ...
145455         24.0       1024.6       1020.3  4.447461   4.50993     10.1
145456         21.0       1023.5       1019.1  4.447461   4.50993     10.9
145457         24.0       1021.0       1016.8  4.447461   4.50993     12.5
145458         24.0       1019.4       1016.5  3.000000   2.00000     15.1
145459         36.0       1020.2       1017.9  8.000000   8.00000     15.0

        Temp3pm  RainToday RainTomorrow
0          21.8         No          No
1          24.3         No          No
2          23.2         No          No
3          26.5         No          No
4          29.7         No          No
...         ...        ...         ...
145455     22.4         No          No
145456     24.5         No          No
145457     26.1         No          No
145458     26.0         No          No
145459     20.9         No          No

[145460 rows x 21 columns]
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 145460 entries, 0 to 145459
Data columns (total 21 columns):
```

```
 #   Column         Non-Null Count   Dtype  
---  ------         --------------   -----  
 0   MinTemp        145460 non-null  float64
 1   MaxTemp        145460 non-null  float64
 2   Rainfall       145460 non-null  float64
 3   Evaporation    145460 non-null  float64
 4   Sunshine       145460 non-null  float64
 5   WindGustDir    135134 non-null  object 
 6   WindGustSpeed  145460 non-null  float64
 7   WindDir9am     134894 non-null  object 
 8   WindDir3pm     141232 non-null  object 
 9   WindSpeed9am   145460 non-null  float64
 10  WindSpeed3pm   145460 non-null  float64
 11  Humidity9am    145460 non-null  float64
 12  Humidity3pm    145460 non-null  float64
 13  Pressure9am    145460 non-null  float64
 14  Pressure3pm    145460 non-null  float64
 15  Cloud9am       145460 non-null  float64
 16  Cloud3pm       145460 non-null  float64
 17  Temp9am        145460 non-null  float64
 18  Temp3pm        145460 non-null  float64
 19  RainToday      145460 non-null  object 
 20  RainTomorrow   145460 non-null  object 
dtypes: float64(16), object(5)
memory usage: 23.3+ MB
array([[3, 1],
       [0, 6]])
```




```
import pandas as pd
import numpy as np
import seaborn as sb
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn import tree
import pydotplus
from sklearn.tree import DecisionTreeRegressor
import matplotlib.pyplot as plt
import matplotlib.image as pltimg
import matplotlib.pyplot as plt
from sklearn.neighbors import KNeighborsRegressor
```

```
df = pd.read_csv('weatherAUS.csv')
print('\Data Frame \n', df)

# Data Exploration
df.head()

df.info()

df.describe()

df.shape

df.columns


# Data Cleaning
df = df.drop(columns=['Location', 'Date', 'WindGustDir', 'WindDir9am', 'WindDir3pm'])
print(df)


# Graphs
df.plot(x="Humidity9am", y="Rainfall", kind="scatter")
df.plot(x= "Humidity3pm", y="Rainfall", kind = "scatter")




# Linear Regression

print(df.head())
print('\n Data Frame Dimensions', df.shape)


X = df.iloc[:, 0:12]

y = df.iloc[:, 13]

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state

print('train size:', X_train.shape)
print('test size:', X_test.shape)



# Decision Tree
# we are going to try and create a decision tree
d = {'WindGustDir': 1}

df['WindGustDir'] =  df['WindGustDir'].map(d)

d = {'YES': 1, 'NO': 0}
```

```
df['Go'] = df['Go'].map(d)

print(df)

features = ['Rainfall', 'Presure3pm', 'Pressure9am', 'WindDir3pm', 'WindDir9am', 'Win

X = df[features]
y = df['Go']

print(X)
print(y)

dtree = DecisionTreeClassifier()
dtree = dtree.fit(X, y)

data = tree.export_graphviz(dtree, out_file=None, feature_names=features)
graph = pydotplus.graph_from_dot_data(data)


# kNN
regressor = KNeighborsRegressorn(n_Rainfall=3)
regressor.fit(X_train, y_train)

X_train_scaled = scaler.transform(X_train)
X_test = scaler.transform(X_test)



#Analysis
# Based on the results used for this project I have came to the conclusion that there
# Some data can be read and calculated properly but if not defined properly it will h
# it is possiable to get whatever results needed. Python seems to run at a decent eno
# from a lack of uncorprative data, or hard to translate. They both should work with



# My Thought's
# this Language is just as frustrating to work in as R. Though there are some things
# However if it is to be belevied that Python is better than R for machine learning t
# and conversions need to calculate data and display data. Even though most of my val
```

```
\Data Frame
                Date Location  MinTemp  MaxTemp  Rainfall  Evaporation  \
0         2008-12-01   Albury     13.4     22.9       0.6          NaN
1         2008-12-02   Albury      7.4     25.1       0.0          NaN
2         2008-12-03   Albury     12.9     25.7       0.0          NaN
3         2008-12-04   Albury      9.2     28.0       0.0          NaN
4         2008-12-05   Albury     17.5     32.3       1.0          NaN
...              ...      ...      ...      ...       ...          ...
145455    2017-06-21    Uluru      2.8     23.4       0.0          NaN
145456    2017-06-22    Uluru      3.6     25.3       0.0          NaN
145457    2017-06-23    Uluru      5.4     26.9       0.0          NaN
145458    2017-06-24    Uluru      7.8     27.0       0.0          NaN
145459    2017-06-25    Uluru     14.9      NaN       0.0          NaN

        Sunshine WindGustDir  WindGustSpeed WindDir9am  ... Humidity9am  \
0            NaN           W           44.0          W  ...        71.0
1            NaN         WNW           44.0        NNW  ...        44.0
2            NaN         WSW           46.0          W  ...        38.0
3            NaN          NE           24.0         SE  ...        45.0
4            NaN           W           41.0        ENE  ...        82.0
...          ...         ...            ...        ...  ...         ...
145455       NaN           E           31.0         SE  ...        51.0
145456       NaN         NNW           22.0         SE  ...        56.0
145457       NaN           N           37.0         SE  ...        53.0
145458       NaN          SE           28.0        SSE  ...        51.0
145459       NaN         NaN            NaN        ESE  ...        62.0

        Humidity3pm  Pressure9am  Pressure3pm  Cloud9am  Cloud3pm  Temp9am  \
0              22.0       1007.7       1007.1       8.0       NaN     16.9
1              25.0       1010.6       1007.8       NaN       NaN     17.2
2              30.0       1007.6       1008.7       NaN       2.0     21.0
3              16.0       1017.6       1012.8       NaN       NaN     18.1
4              33.0       1010.8       1006.0       7.0       8.0     17.8
...             ...          ...          ...       ...       ...      ...
145455         24.0       1024.6       1020.3       NaN       NaN     10.1
145456         21.0       1023.5       1019.1       NaN       NaN     10.9
145457         24.0       1021.0       1016.8       NaN       NaN     12.5
145458         24.0       1019.4       1016.5       3.0       2.0     15.1
145459         36.0       1020.2       1017.9       8.0       8.0     15.0

        Temp3pm  RainToday  RainTomorrow
0          21.8         No            No
1          24.3         No            No
2          23.2         No            No
3          26.5         No            No
4          29.7         No            No
...         ...        ...           ...
145455     22.4         No            No
145456     24.5         No            No
145457     26.1         No            No
145458     26.0         No            No
145459     20.9         No           NaN

[145460 rows x 23 columns]
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 145460 entries, 0 to 145459
```

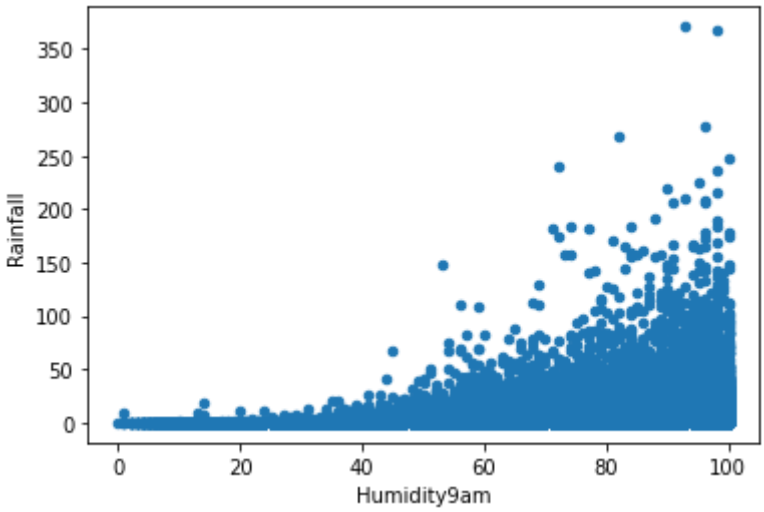
```
Data columns (total 23 columns):
 #   Column         Non-Null Count   Dtype  
---  ------         --------------   -----  
 0   Date           145460 non-null  object 
 1   Location       145460 non-null  object 
 2   MinTemp        143975 non-null  float64
 3   MaxTemp        144199 non-null  float64
 4   Rainfall       142199 non-null  float64
 5   Evaporation    82670 non-null   float64
 6   Sunshine       75625 non-null   float64
 7   WindGustDir    135134 non-null  object 
 8   WindGustSpeed  135197 non-null  float64
 9   WindDir9am     134894 non-null  object 
 10  WindDir3pm     141232 non-null  object 
 11  WindSpeed9am   143693 non-null  float64
 12  WindSpeed3pm   142398 non-null  float64
 13  Humidity9am    142806 non-null  float64
 14  Humidity3pm    140953 non-null  float64
 15  Pressure9am    130395 non-null  float64
 16  Pressure3pm    130432 non-null  float64
 17  Cloud9am       89572 non-null   float64
 18  Cloud3pm       86102 non-null   float64
 19  Temp9am        143693 non-null  float64
 20  Temp3pm        141851 non-null  float64
 21  RainToday      142199 non-null  object 
 22  RainTomorrow   142193 non-null  object 
dtypes: float64(16), object(7)
memory usage: 25.5+ MB
        MinTemp  MaxTemp  Rainfall  Evaporation  Sunshine  WindGustSpeed  \
0          13.4     22.9       0.6          NaN       NaN           44.0   
1           7.4     25.1       0.0          NaN       NaN           44.0   
2          12.9     25.7       0.0          NaN       NaN           46.0   
3           9.2     28.0       0.0          NaN       NaN           24.0   
4          17.5     32.3       1.0          NaN       NaN           41.0   
...         ...      ...       ...          ...       ...            ...   
145455      2.8     23.4       0.0          NaN       NaN           31.0   
145456      3.6     25.3       0.0          NaN       NaN           22.0   
145457      5.4     26.9       0.0          NaN       NaN           37.0   
145458      7.8     27.0       0.0          NaN       NaN           28.0   
145459     14.9      NaN       0.0          NaN       NaN            NaN   

        WindSpeed9am  WindSpeed3pm  Humidity9am  Humidity3pm  Pressure9am  \
0               20.0          24.0         71.0         22.0       1007.7   
1                4.0          22.0         44.0         25.0       1010.6   
2               19.0          26.0         38.0         30.0       1007.6   
3               11.0           9.0         45.0         16.0       1017.6   
4                7.0          20.0         82.0         33.0       1010.8   
...              ...           ...          ...          ...          ...   
145455          13.0          11.0         51.0         24.0       1024.6   
145456          13.0           9.0         56.0         21.0       1023.5   
145457           9.0           9.0         53.0         24.0       1021.0   
145458          13.0           7.0         51.0         24.0       1019.4   
145459          17.0          17.0         62.0         36.0       1020.2   

        Pressure3pm  Cloud9am  Cloud3pm  Temp9am  Temp3pm RainToday  \
0            1007.1       8.0       NaN     16.9     21.8        No   
1            1007.8       NaN       NaN     17.2     24.3        No   
```

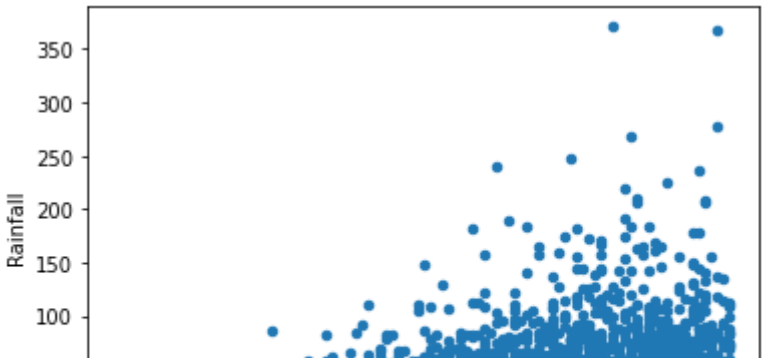
```
2            1008.7       NaN       2.0      21.0      23.2         No
3            1012.8       NaN       NaN      18.1      26.5         No
4            1006.0       7.0       8.0      17.8      29.7         No
...             ...       ...       ...       ...       ...        ...
145455       1020.3       NaN       NaN      10.1      22.4         No
145456       1019.1       NaN       NaN      10.9      24.5         No
145457       1016.8       NaN       NaN      12.5      26.1         No
145458       1016.5       3.0       2.0      15.1      26.0         No
145459       1017.9       8.0       8.0      15.0      20.9         No

       RainTomorrow
0                No
1                No
2                No
3                No
4                No
...             ...
145455           No
145456           No
145457           No
145458           No
145459          NaN

[145460 rows x 18 columns]
   MinTemp  MaxTemp  Rainfall  Evaporation  Sunshine  WindGustSpeed  \
0     13.4     22.9       0.6          NaN       NaN           44.0
1      7.4     25.1       0.0          NaN       NaN           44.0
2     12.9     25.7       0.0          NaN       NaN           46.0
3      9.2     28.0       0.0          NaN       NaN           24.0
4     17.5     32.3       1.0          NaN       NaN           41.0

   WindSpeed9am  WindSpeed3pm  Humidity9am  Humidity3pm  Pressure9am  \
0          20.0          24.0         71.0         22.0       1007.7
1           4.0          22.0         44.0         25.0       1010.6
2          19.0          26.0         38.0         30.0       1007.6
3          11.0           9.0         45.0         16.0       1017.6
4           7.0          20.0         82.0         33.0       1010.8

   Pressure3pm  Cloud9am  Cloud3pm  Temp9am  Temp3pm RainToday RainTomorrow
0       1007.1       8.0       NaN     16.9     21.8        No           No
1       1007.8       NaN       NaN     17.2     24.3        No           No
2       1008.7       NaN       2.0     21.0     23.2        No           No
3       1012.8       NaN       NaN     18.1     26.5        No           No
4       1006.0       7.0       8.0     17.8     29.7        No           No

 Data Frame Dimensions (145460, 18)
train size: (116368, 12)
test size: (29092, 12)
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
/usr/local/lib/python3.7/dist-packages/pandas/core/indexes/base.py in
get_loc(self, key, method, tolerance)
   3360             try:
-> 3361                 return self._engine.get_loc(casted_key)
   3362             except KeyError as err:
```

4 frames

```
pandas/_libs/hashtable_class_helper.pxi in
pandas._libs.hashtable.PyObjectHashTable.get_item()

pandas/_libs/hashtable_class_helper.pxi in
pandas._libs.hashtable.PyObjectHashTable.get_item()

KeyError: 'WindGustDir'

The above exception was the direct cause of the following exception:

KeyError                                  Traceback (most recent call last)
/usr/local/lib/python3.7/dist-packages/pandas/core/indexes/base.py in
get_loc(self, key, method, tolerance)
   3361                 return self._engine.get_loc(casted_key)
   3362             except KeyError as err:
-> 3363                 raise KeyError(key) from err
   3364
   3365         if is_scalar(key) and isna(key) and not self.hasnans:

KeyError: 'WindGustDir'
```

SEARCH STACK OVERFLOW





Double-click (or enter) to edit

1s completed at 10:49 PM