In [20]:

```
import pandas as pd
import numpy as np
from sklearn.linear_model import Perceptron
from sklearn import metrics
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt
%matplotlib inline
```

In [22]:

```
data = pd.read_csv('files/weather.csv', parse_dates=True, index_col=0)
data.head()
```

Out[22]:

	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am	WindDir3pm	WindSpeed9am	...	Humi
Date												
2008-02-01	19.5	22.4	15.6	6.2	0.0	NaN	NaN	S	SSW	17.0	...	
2008-02-02	19.5	25.6	6.0	3.4	2.7	NaN	NaN	W	E	9.0	...	
2008-02-03	21.6	24.5	6.6	2.4	0.1	NaN	NaN	ESE	ESE	17.0	...	
2008-02-04	20.2	22.8	18.8	2.2	0.0	NaN	NaN	NNE	E	22.0	...	
2008-02-05	19.7	25.7	77.4	NaN	0.0	NaN	NaN	NNE	W	11.0	...	

5 rows × 22 columns

In [23]:

```
#Look for missing data points
data.isna().sum()
```

Out[23]:

```
MinTemp             3
MaxTemp             2
Rainfall            6
Evaporation        51
Sunshine           16
WindGustDir      1036
WindGustSpeed    1036
WindDir9am         56
WindDir3pm         33
WindSpeed9am       26
WindSpeed3pm       25
Humidity9am        14
Humidity3pm        13
Pressure9am        20
Pressure3pm        19
Cloud9am          566
Cloud3pm          561
Temp9am             4
Temp3pm             4
RainToday           6
RISK_MM             0
RainTomorrow        0
dtype: int64
```

In [25]:

```
#Remove 'dirty' columns
dataset = data.drop(['WindGustDir', 'WindGustSpeed', 'Cloud9am', 'Cloud3pm', 'WindDir9am', 'WindDir3pm', 'RainTod
dataset.head()
```

Out[25]:

	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindSpeed9am	WindSpeed3pm	Humidity9am	Humidity3pm	Pressure9am	Pressu
Date											
2008-02-01	19.5	22.4	15.6	6.2	0.0	17.0	20.0	92.0	84.0	1017.6	1
2008-02-02	19.5	25.6	6.0	3.4	2.7	9.0	13.0	83.0	73.0	1017.9	1
2008-02-03	21.6	24.5	6.6	2.4	0.1	17.0	2.0	88.0	86.0	1016.7	1
2008-02-04	20.2	22.8	18.8	2.2	0.0	22.0	20.0	83.0	90.0	1014.2	1
2008-02-05	19.7	25.7	77.4	NaN	0.0	11.0	6.0	88.0	74.0	1008.3	1

In [26]:

```
dataset_clean = dataset.dropna()
len(dataset), len(dataset_clean)
```

Out[26]:

```
(3337, 3187)
```

In [27]:

```
#Create training and test datasets
X = dataset_clean[dataset_clean.columns[:-1]]
y = dataset_clean['RainTomorrow']
y = np.array([0 if value == 'No' else 1 for value in y])
```

In [28]:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=0)
```

In [29]:

```
#Train and test the model
clf = Perceptron(random_state=0)
clf.fit(X_train, y_train)
y_pred = clf.predict(X_test)
metrics.accuracy_score(y_test, y_pred)
```

Out[29]:

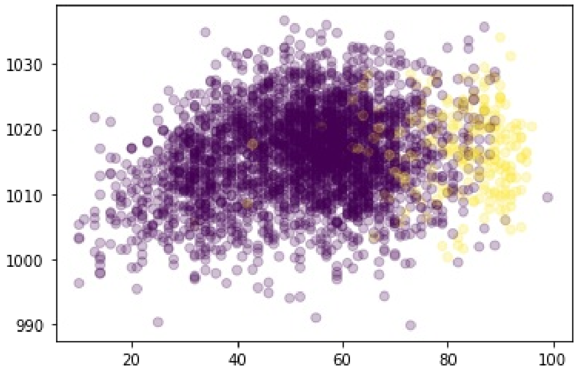
```
0.835633626097867
```

In [30]:

```
#Plot the result
fig, ax = plt.subplots()
y_pred = clf.predict(X)
ax.scatter(x=X['Humidity3pm'], y=X['Pressure3pm'], c=y_pred, alpha=.25)
```

Out[30]:

```
<matplotlib.collections.PathCollection at 0x7fdcb2383370>
```



In [ ]: