Bitcoin Price Analysis

Objectives:

- Extract BTC prices from Yahoo Finance and key words frequency from Google Trends
- Manipulate, transform, and merge datasets to prepare variables
- Run simple regression against the key words 'Bitcoin'
- Analyze regression statistics and conlcude the significance

In [17]:
```
import pytrends
import numpy as np
import pandas as pd
import pandas_datareader.data as pdr
from datetime import datetime, timedelta

from scipy import stats
import statsmodels.api as sm

import matplotlib.pyplot as plt
%matplotlib inline
```

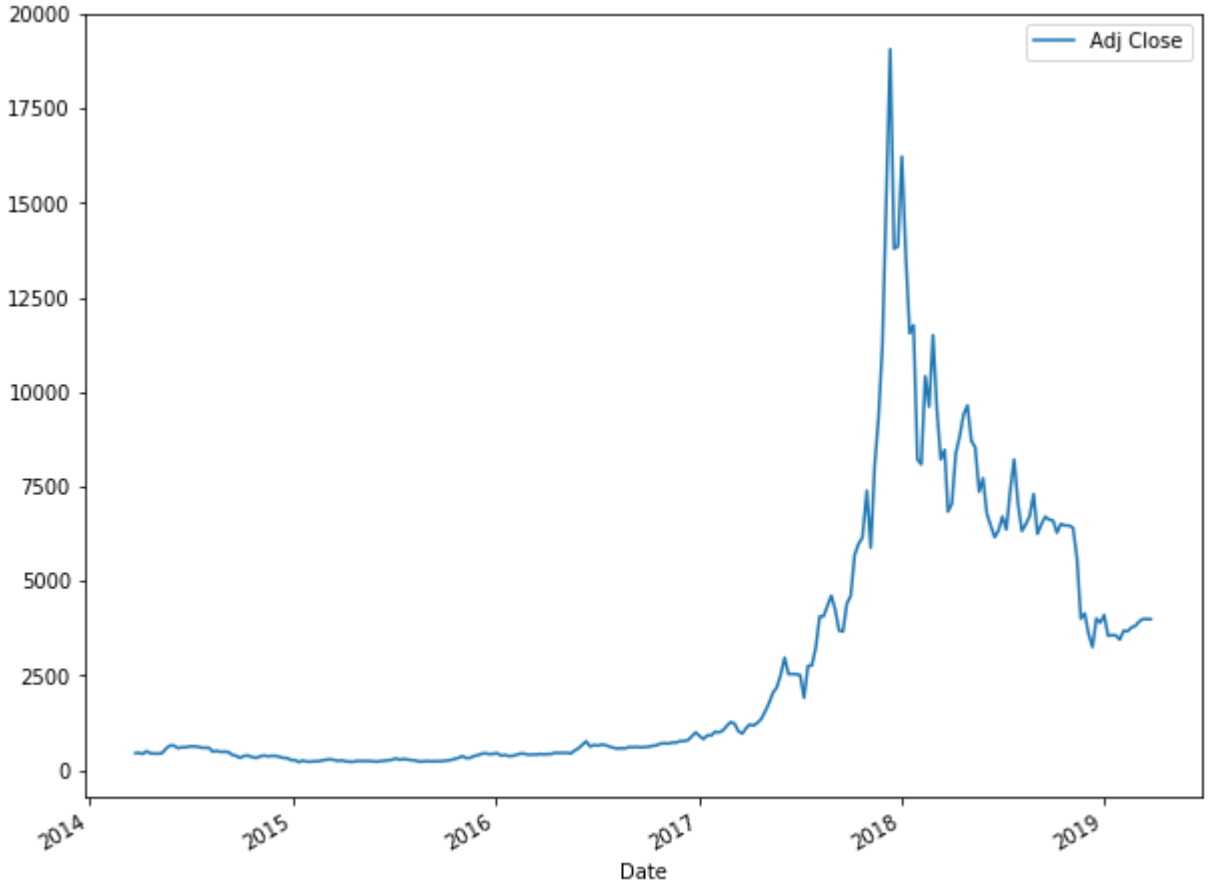
In [18]:
```
#Setting the end date to today
end = datetime.today()

#Start date set to one year back
start = datetime(end.year-5,end.month,end.day)

#using yahoo finance to grab cryptocurrency data
btc = pdr.get_data_yahoo('BTC-USD',start,end,interval='w')
# ETH = pdr.DataReader('ETH-USD','yahoo',start,end)
# LTC = pdr.DataReader('LTC-USD','yahoo',start,end)
```

In [3]:
```
#Set the figure sizes
plt.rcParams['figure.figsize'] = (10,8)
```

In [4]:
```
#Plot the Bitcoin price movements over the past 5 years
btc['Adj Close'].plot(legend = True);
```

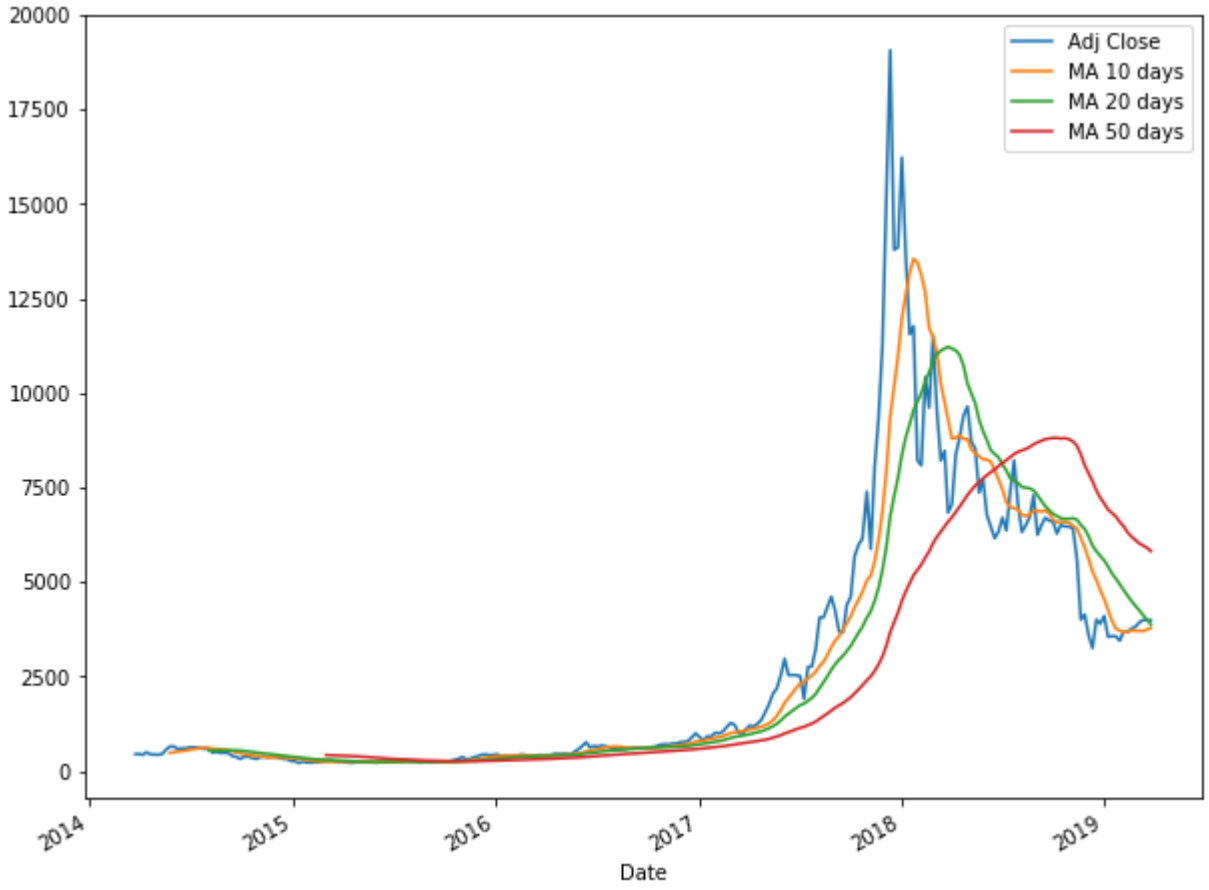


In [5]:
```
#Plot 10,20,and 50 days moving average with daily Bitcoin prices

ma_days = [10,20,50]

for ma in ma_days:
     column_name = "MA %s days"%(str(ma))
     btc[column_name] = btc['Adj Close'].rolling(window=ma,center=False).mean()

btc[['Adj Close','MA 10 days','MA 20 days','MA 50 days']].plot(legend=True);
```



Use Pytrend package to extract weekly google searches data

In [6]:
```
from pytrends.request import TrendReq
pytrends = TrendReq(hl='en-US', tz=360)
kw_list = ["Bitcoin","Ethereum"]
```

In [7]:
```
pytrends.build_payload(kw_list, cat=0, timeframe='today 5-y', geo='', gprop='')

btc_searches = pytrends.interest_over_time()
```

Combine Two Pandas Dataframes

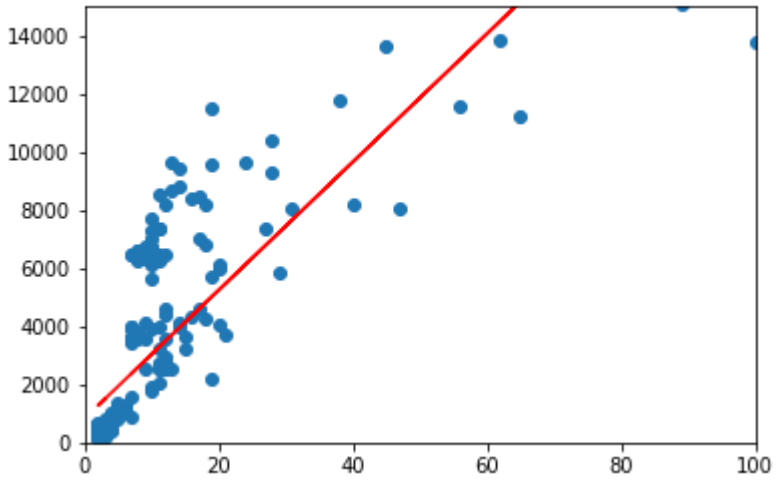
In [8]:
```
btc_prices = btc['Adj Close'].iloc[1:-1]
```

In [9]:
```
df = pd.concat([btc_prices.reset_index(drop=True),btc_searches.reset_index(drop=False)], axis=1)
```

In [19]:
```
#df = df.set_index('date')
```

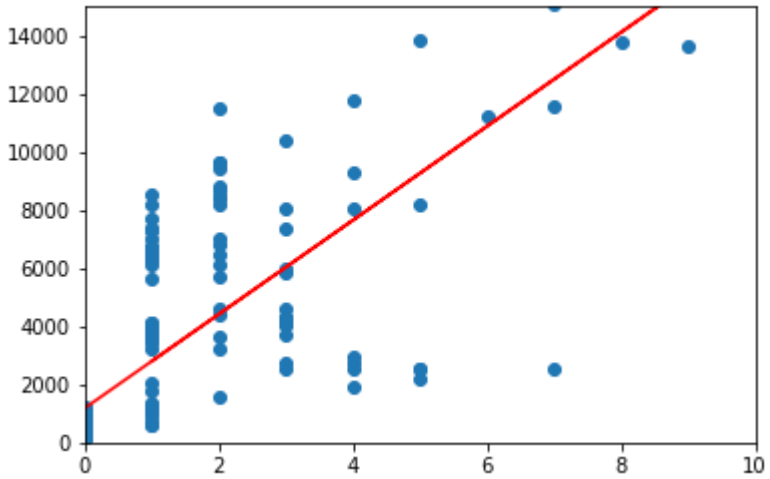
In [20]:
```
#Draw the scatter plot with Bitcoin searches as an independent variable,
#and Bitcoin average weekly prices as a dependent variable
X1 = df['Bitcoin']
Y1 = df['Adj Close']
plt.scatter(X1,Y1)
plt.axis([0,100,0,15000])

#Draw the trend line
z = np.polyfit(X1,Y1,1)
p = np.poly1d(z)
plt.plot(X1,p(X1),"r")
plt.show()
```



In [21]:
```
#Draw the scatter plot with Bitcoin searches as an independent variable,
#and Bitcoin average weekly prices as a dependent variable
X2 = df['Ethereum']
Y2 = df['Adj Close']
plt.scatter(X2,Y2)
plt.axis([0,10,0,15000])

#Draw the trend line
z = np.polyfit(X2,Y2,1)
p = np.poly1d(z)
plt.plot(X2,p(X2),"r")
plt.show()
```



Simple Linear Regression Statistics

In [22]:
```
#Apply statsmodel to run regression and conclude detailed stats
x_hat = sm.add_constant(X1)
reg = sm.OLS(Y1, x_hat).fit()
```

In [23]:
```
reg.summary()
```

Out[23]:

OLS Regression Results

Dep. Variable:	Adj Close	**R-squared:**	0.681
Model:	OLS	**Adj. R-squared:**	0.680
Method:	Least Squares	**F-statistic:**	550.4
Date:	Sun, 24 Mar 2019	**Prob (F-statistic):**	6.19e-66
Time:	21:45:04	**Log-Likelihood:**	-2343.6
No. Observations:	260	**AIC:**	4691.
Df Residuals:	258	**BIC:**	4698.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	830.5498	147.943	5.614	0.000	539.221	1121.878
Bitcoin	220.5914	9.402	23.461	0.000	202.076	239.107

Omnibus:	47.274	**Durbin-Watson:**	0.413
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	98.681
Skew:	0.904	**Prob(JB):**	3.73e-22
Kurtosis:	5.416	**Cond. No.**	18.8

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [24]:
```
#Or use scipy to run linear regression
slope, intercept, r_value, p_value, std_err = stats.linregress(X1,Y1)
stats.linregress(X1,Y1)
```

Out[24]:
```
LinregressResult(slope=220.59141247124444, intercept=830.5497519919704, rvalue=0.8251438681435742, pvalue=6.1
88115567238563e-66, stderr=9.402385580504816)
```

Insights:

1. This univariate regression explains that 64.8% of the Bitcoin prices is explained by 'Bitcoin' frequencies in Google searches
2. p-values for the slope and intercept are both smaller than 0.01. The model has decent prediction power for future Bitcoin prices.