**Data Analytics Project**

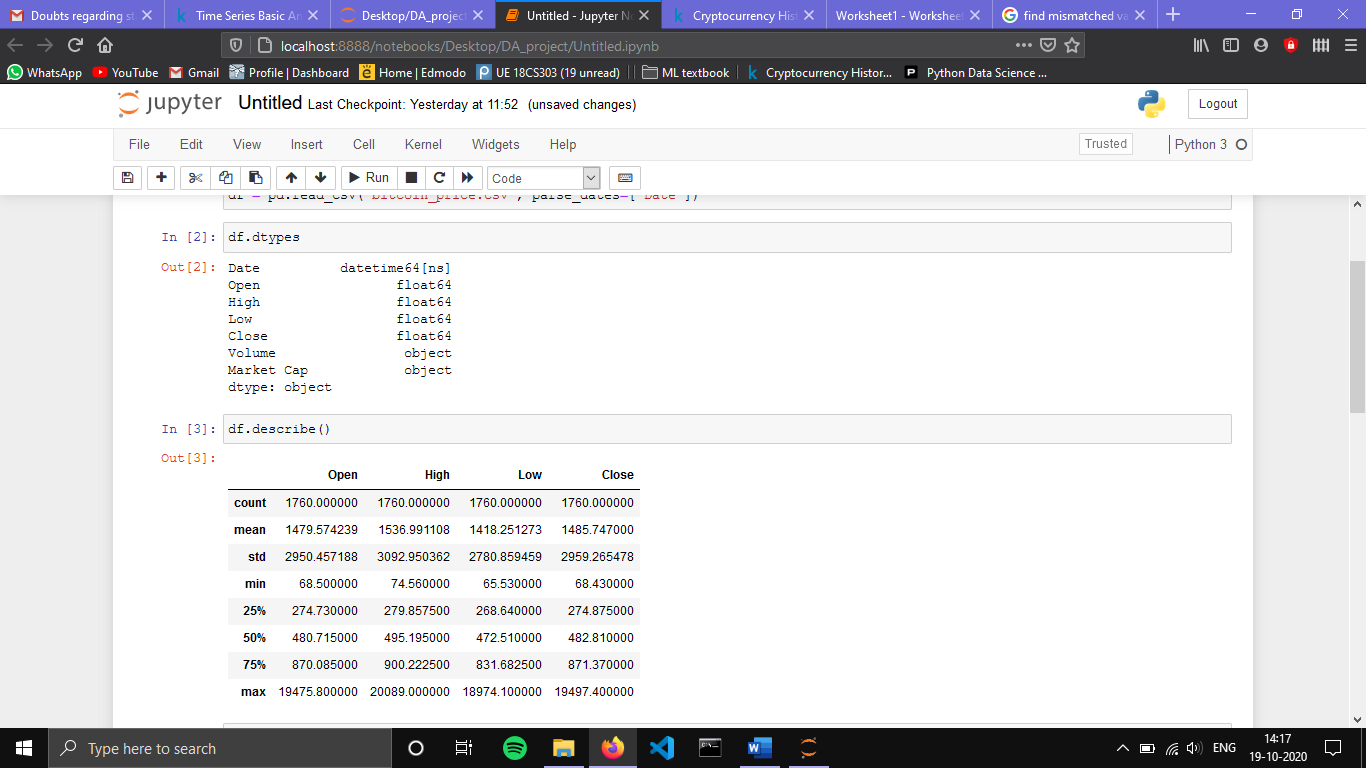
**Stock-taking of data**

**Bitcoin\_price.csv**

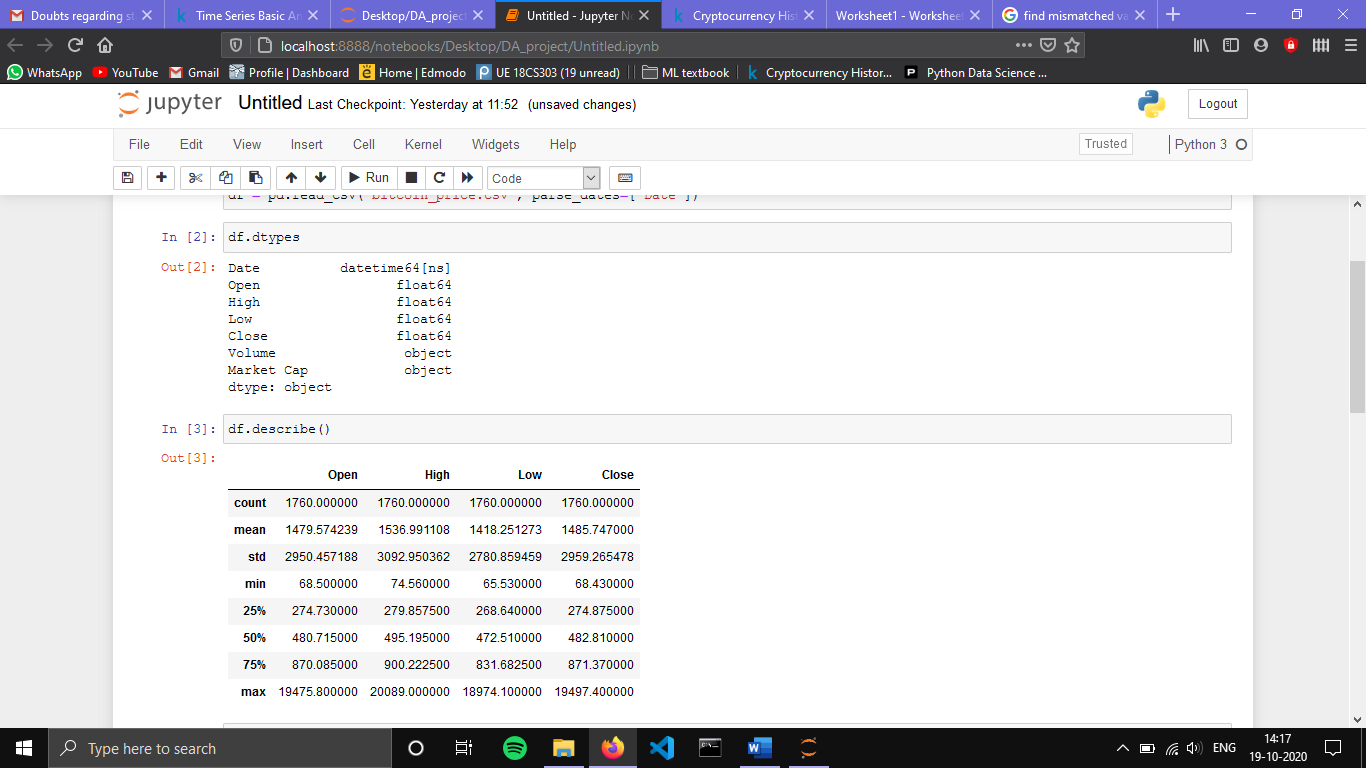
**Number of columns:** 7

**Number of rows:** 1760

**Attributes:**

****

**Summary statistics:**

****

**Attributes description:**

**Date:** date of observation

**Open:** Opening price on the given day

**High:** Highest price on the given day

**Low:** Lowest price on the given day

**Close:** Closing price on the given day

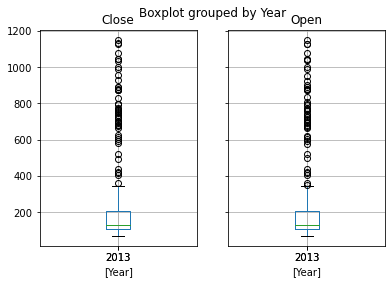
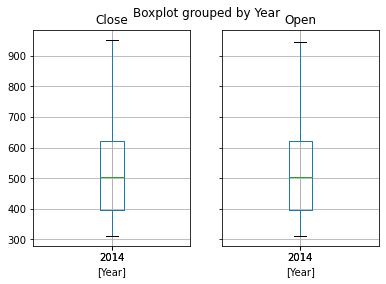
**Volume:** Volume of transactions on the given day

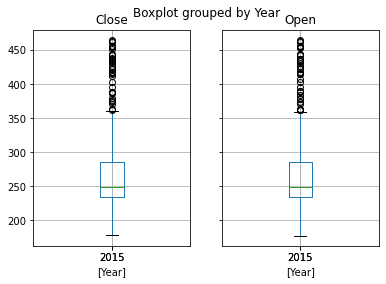
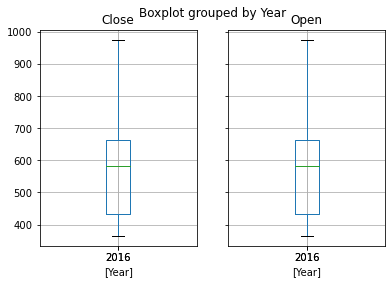
**Market\_Cap:** Market capitalization in USD

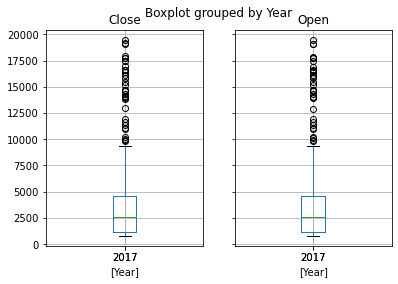
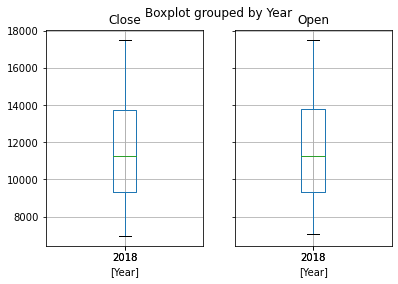
**Missing values:** There are no missing values in the dataset.

The attribute Volume has 243 (14%) mismatched values

**Boxplots for Open and Close prices grouped by Year**

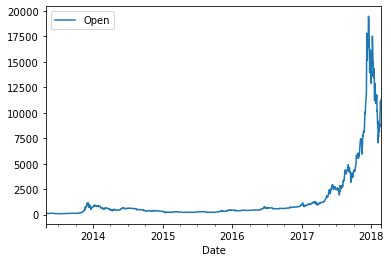
** **

** **

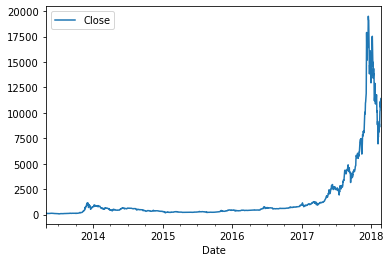
** **

From the above boxplots, we can see that there are outliers present in the Open and Close prices for the years 2013, 2015 and 2017.

**Plot of Open prices from 28/04/2013**

****

**Plot of Close prices from 28/04/2013**

****

**Data Cleaning**

The mismatched values in this dataset can be replaced by the mean value corresponding to the particular time intervals. The values of the Market\_Cap and Volume attributes must be converted to type float by removing the intermediate commas.

The values in the dataset seem to be of the same scale. There is no need for any transformation.

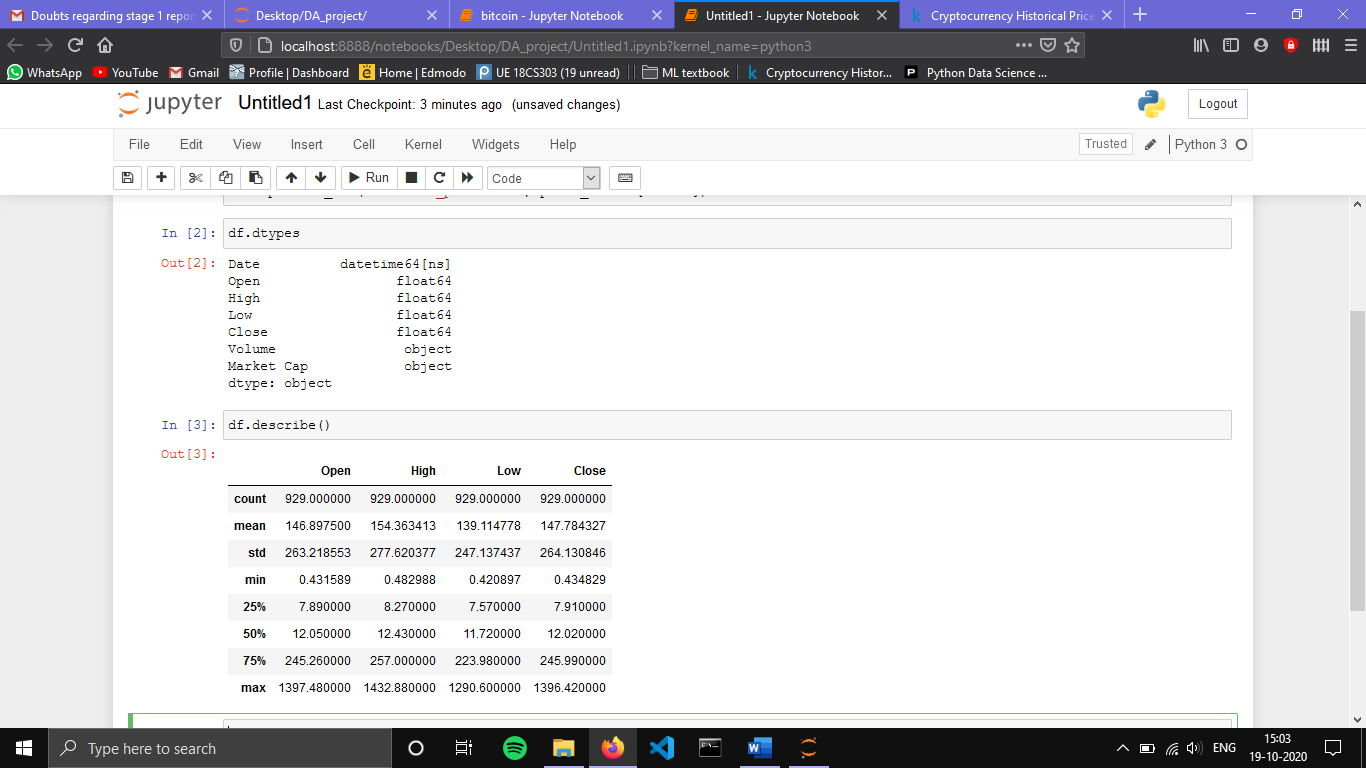
All the attributes are relevant for the analysis. There seems to be no redundant attributes. Hence there is no need for dimensionality reduction.

**Ethereum\_price.csv**

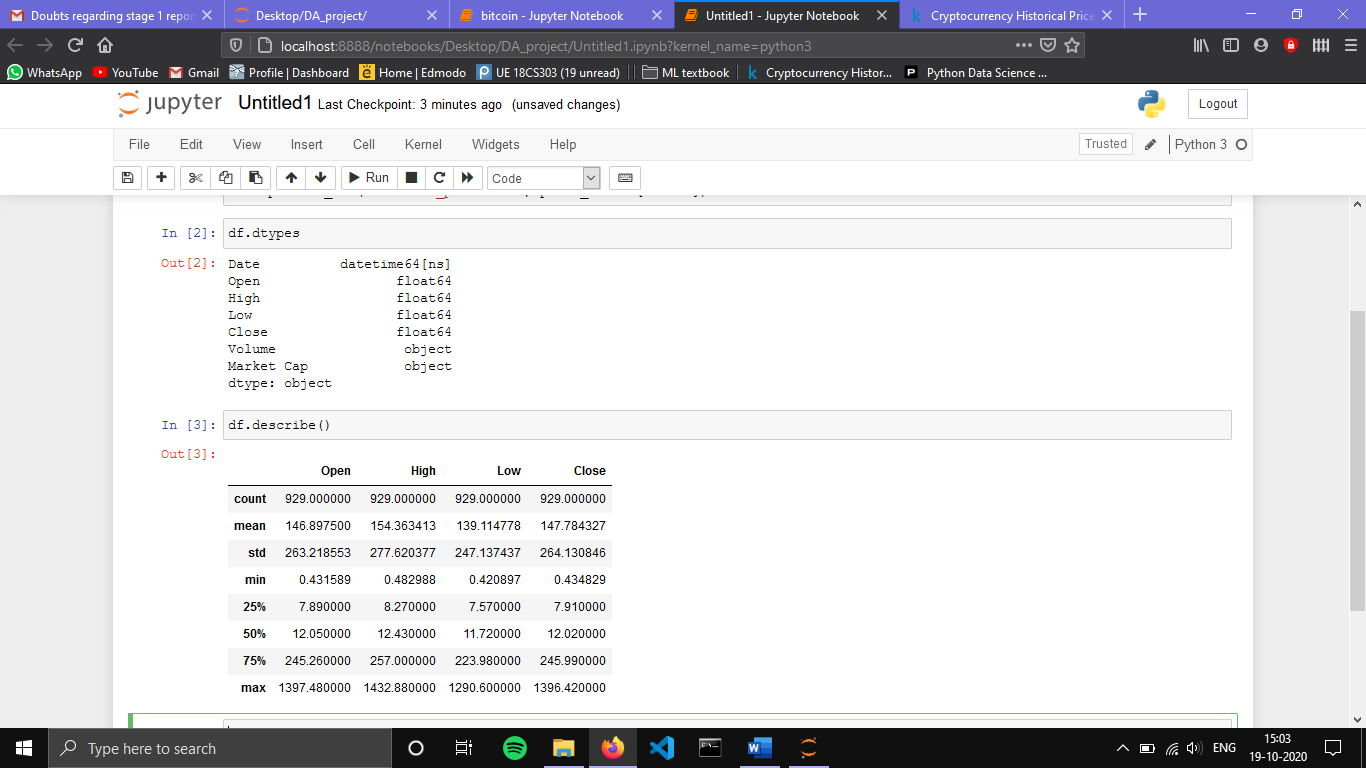
**Number of columns:** 7

**Number of rows:** 929

**Attributes:**

****

**Summary statistics:**

****

**Attributes description:**

**Date:** date of observation

**Open:** Opening price on the given day

**High:** Highest price on the given day

**Low:** Lowest price on the given day

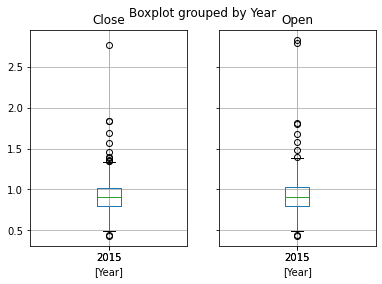
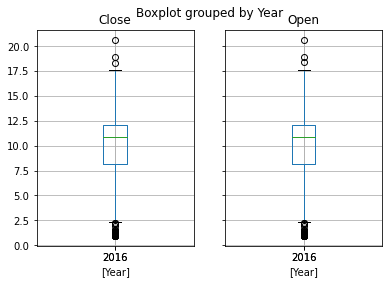
**Close:** Closing price on the given day

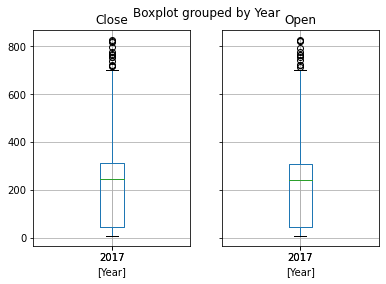
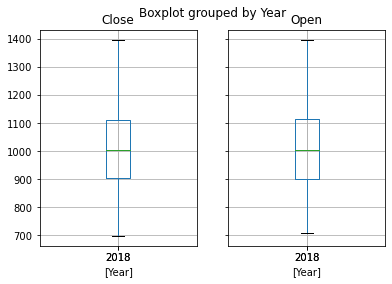
**Volume:** Volume of transactions on the given day

**Market\_Cap:** Market capitalization in USD

**Missing values:** There are no missing or mismatched values in the dataset.

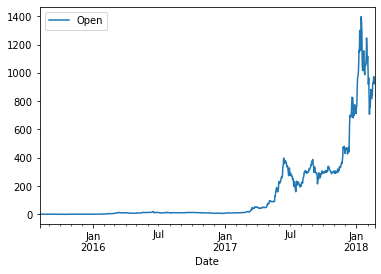
**Boxplots for Open and Close prices grouped by Year**

** **

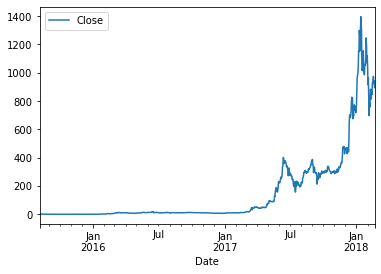
** **

From the above boxplots, we can see that there are outliers present in the Open and Close prices for the years 2015, 2016 and 2017.

**Plot of Open prices from 07/08/2015**

****

**Plot of Close prices from 07/08/2015**

****

**Data Cleaning**

The values of the Market\_Cap and Volume attributes must be converted to type float by removing the intermediate commas.

The values in the dataset seem to be of the same scale. There is no need for any transformation.

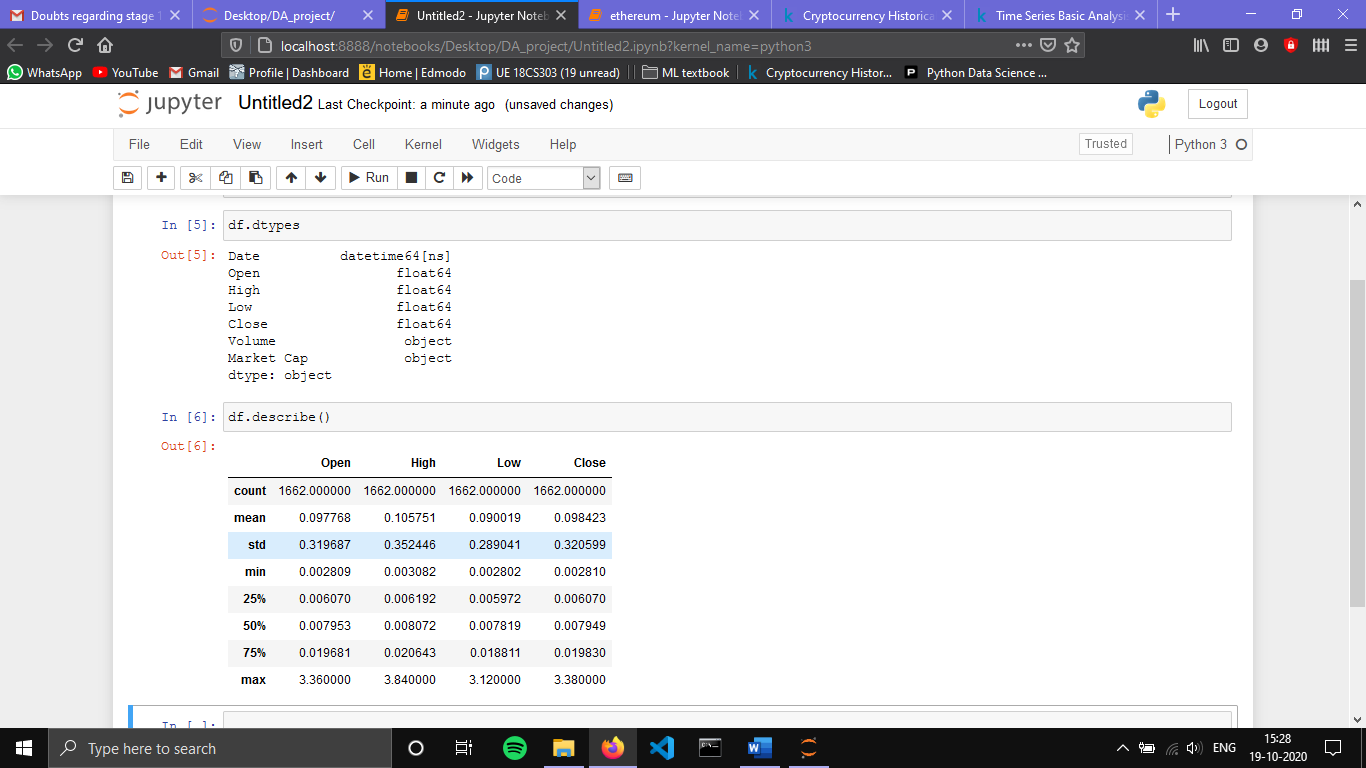
All the attributes are relevant for the analysis. There seems to be no redundant attributes. Hence there is no need for dimensionality reduction.

**Ripple\_price.csv**

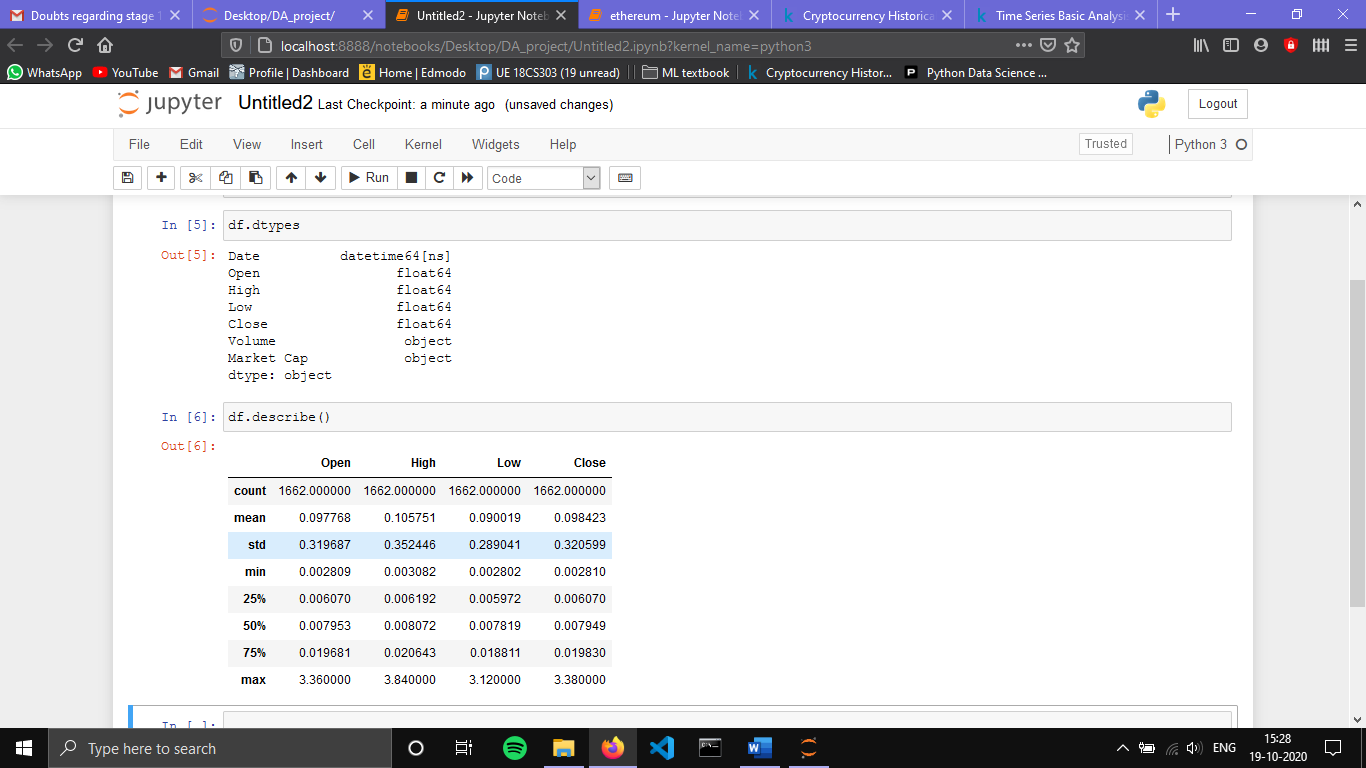
**Number of columns:** 7

**Number of rows:** 1662

**Attributes:**

****

**Summary statistics:**

****

**Attributes description:**

**Date:** date of observation

**Open:** Opening price on the given day

**High:** Highest price on the given day

**Low:** Lowest price on the given day

**Close:** Closing price on the given day

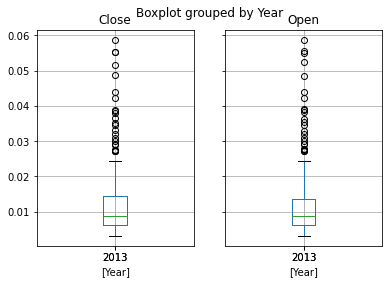
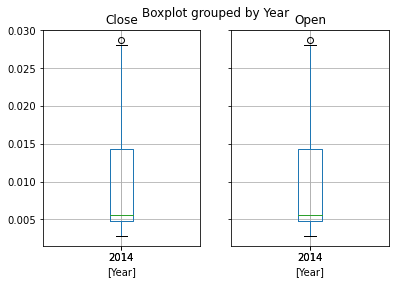
**Volume:** Volume of transactions on the given day

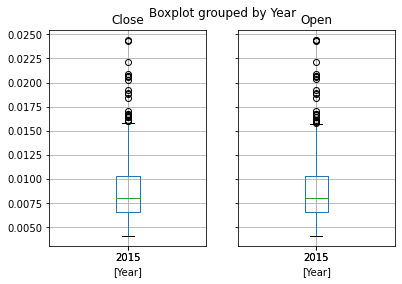
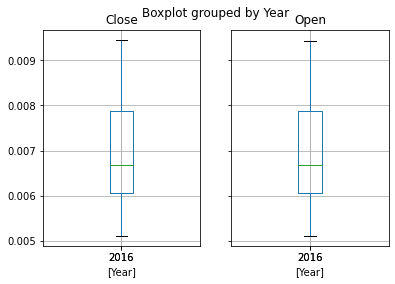
**Market\_Cap:** Market capitalization in USD

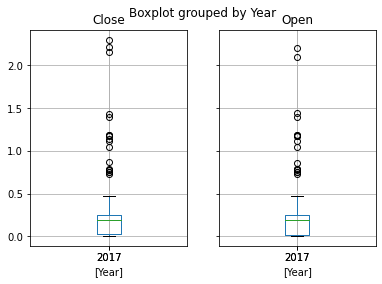
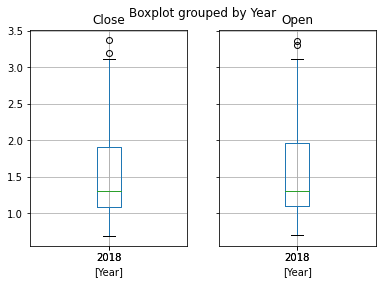
**Missing values:** There are no missing values in the dataset.

The attribute Volume has 145 (9%) mismatched values.

**Boxplots for Open and Close prices grouped by Year**

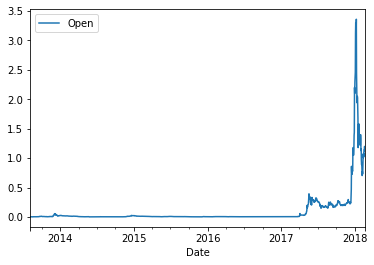
 

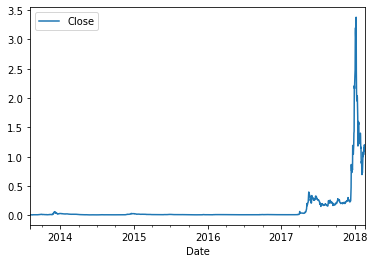
 

From the above boxplots, we can see that there are outliers present in the Open and Close prices for the years 2013, 2014 and 2015, 2016 and 2018.

**Plot of Open prices from 04/08/2013**

****

**Plot of Close prices from 04/08/2013**

****

**Data Cleaning**

The mismatched values in this dataset can be replaced by the mean value corresponding to the particular time intervals. The values of the Market\_Cap and Volume attributes must be converted to type float by removing the intermediate commas.

The values in the dataset seem to be of the same scale. There is no need for any transformation.

All the attributes are relevant for the analysis. There seems to be no redundant attributes. Hence there is no need for dimensionality reduction.