

**BAHRIA UNIVERSITY**

**KARACHI CAMPUS**

**Department of Computer Science**

**Report**

**Name:** Ibrahim Deen

**Class and Section:** BS (CS)-7a

**Enrollment:** 02-134162-

**Registration:**

**Subject:** Data Mining

**Teacher:** Miss Fasiha Ikram

**Assignment:** Assignment # 2

**ACKNOWLEDGEMENT**

All due thanks to almighty, the most beneficent and merciful for blessings and light of success and devoted leadership of our teachers that made us a standalone pillar. My heartedly thanks are due to our most respectable teacher **Miss Fasiha Ikram** devoted and keen support helped us achieving this goal.

**Preface**

Bitcoin is the longest running and most well known cryptocurrency, first released as open source in 2009 by the anonymous Satoshi Nakamoto. Bitcoin serves as a decentralized medium of digital exchange, with transactions verified and recorded in a public distributed ledger (the blockchain) without the need for a trusted record keeping authority or central intermediary. Transaction blocks contain a SHA-256 cryptographic hash of previous transaction blocks, and are thus "chained" together, serving as an immutable record of all transactions that have ever occurred. As with any currency/commodity on the market, bitcoin trading and financial instruments soon followed public adoption of bitcoin and continue to grow. Included here is historical bitcoin market data at 1-min intervals for select bitcoin exchanges where trading takes place. Happy (data) mining!

***INTRODUCTION***

Bitcoin charts for the data. The various exchange APIs, for making it difficult or unintuitive enough to get OHLC and volume data at 1-min intervals that I set out on this data scraping project. Satoshi Nakamoto and the novel core concept of the blockchain, as well as its first execution via the bitcoin protocol. I'd also like to thank viewers like you! Can't wait to see what code or insights you all have to share.

***GOAL***

Goal of this kernel is to compare NN and ARIMA modelling. We will be predicting Bitcoin prices with help of Bitcoin historical data.

***CONTEXT:***

There are 4 csv files. CSV files for select bitcoin exchanges for the time period of Jan 2012 to July 2018, with minute to minute updates of OHLC (Open, High, Low, Close), Volume in BTC and indicated currency, and weighted bitcoin price. Timestamps are in Unix time. Timestamps without any trades or activity have their data fields forward filled from the last valid time period. If a timestamp is missing, or if there are jumps, this may be because the exchange (or its API) was down, the exchange (or its API) did not exist, or some other unforseen technical error in data reporting or gathering.

coincheckJPY\_1-min\_data\_2014-10-31\_to\_2018-06-27.csv

bitflyerJPY\_1-min\_data\_2017-07-04\_to\_2018-06-27.csv

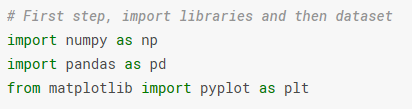
coinbaseUSD\_1-min\_data\_2014-12-01\_to\_2018-06-27.csv

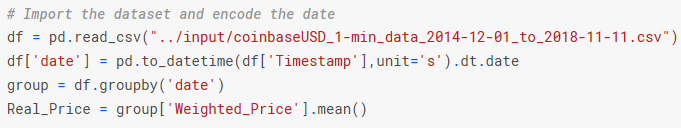
bitstampUSD\_1-min\_data\_2012-01-01\_to\_2018-06-27.csv

All from different Bitcoin exchanges

***DATA FILTRING***

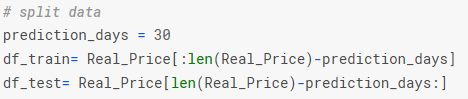
*RNN To predict bitcoin prices*





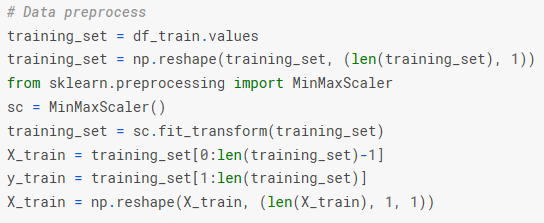
*Bitcoin predictions are going to be for a month, that is why we need to split the dataset accordingly*

***Data Splitting***

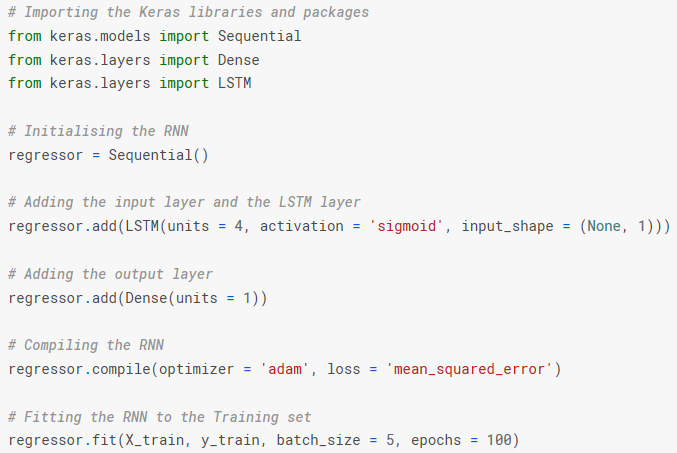


***Data Pre-Processing:***

*Some pre-processing is also necessary:*

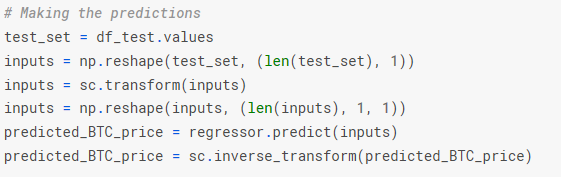


*Now keras to build the rNN, Long short-term memory!!! LSTM*

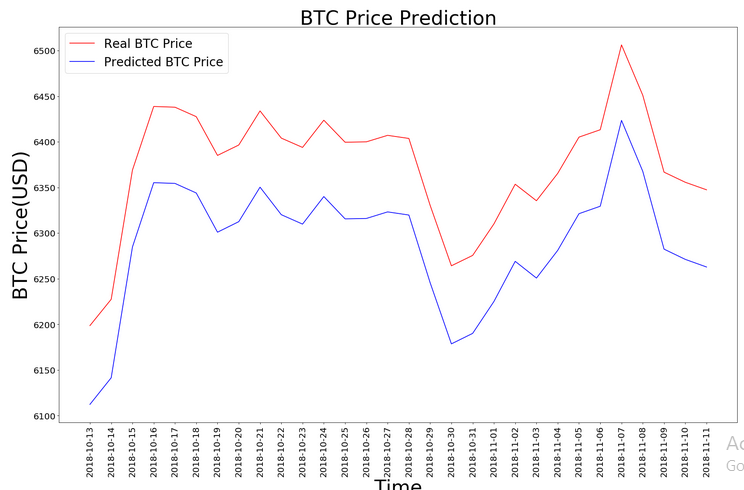


***HYPOTHISIS:***

**Predicting price:**

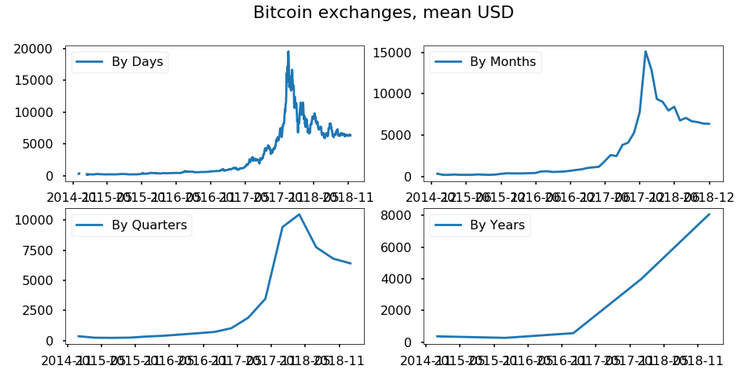






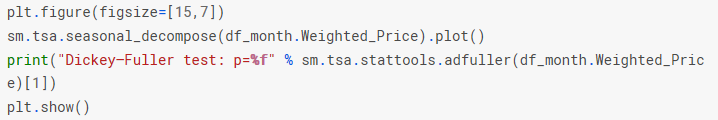
**Visualizing Trends:**

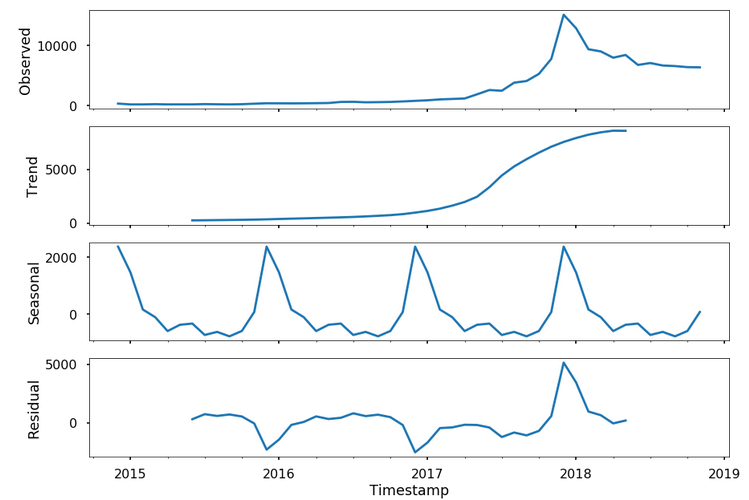




**Stationarity check and STL-decomposition of the series:**

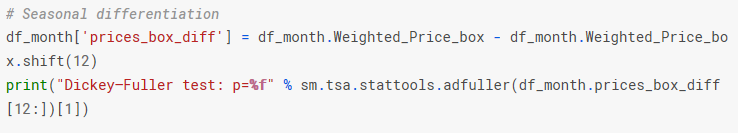
Lower the p value the better. Stationarity is our models main assumption and dickey fuller is just hypothesis test of the unit root test





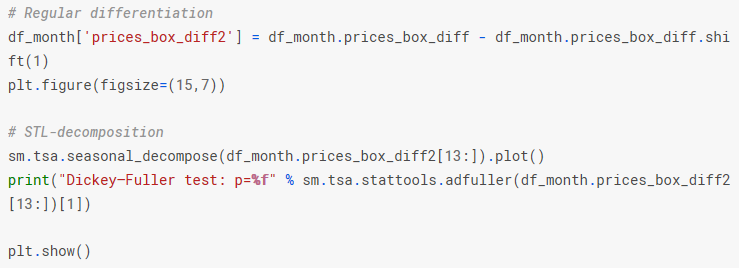
Obviously not stationary, hence we ought transform our data. First Box-cox transformation then check the test

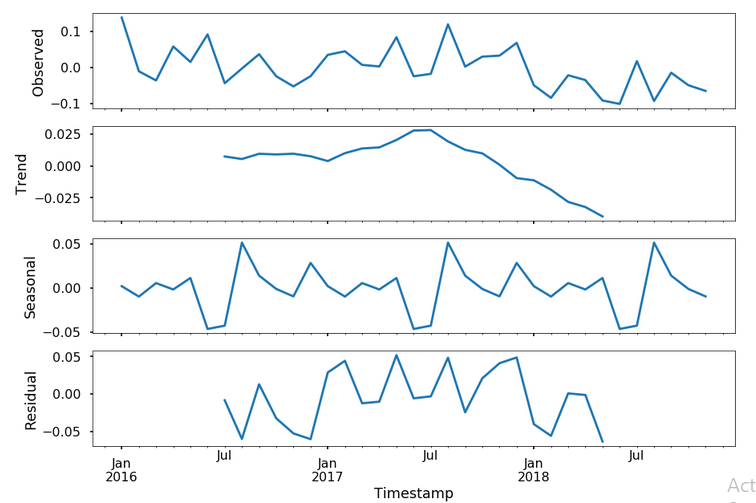
**Seasonal differentiation**



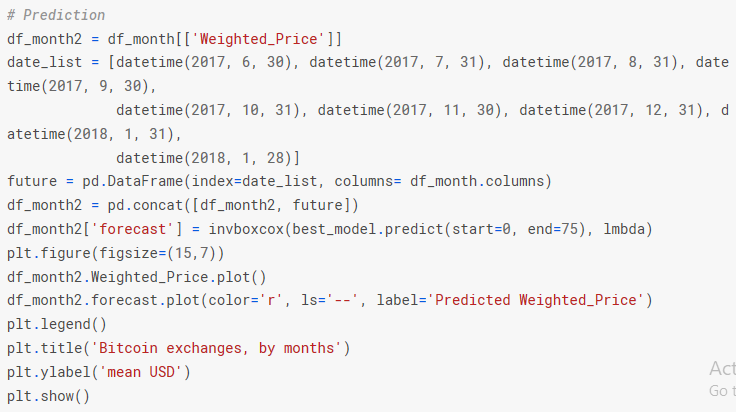
Again series is not stationary,

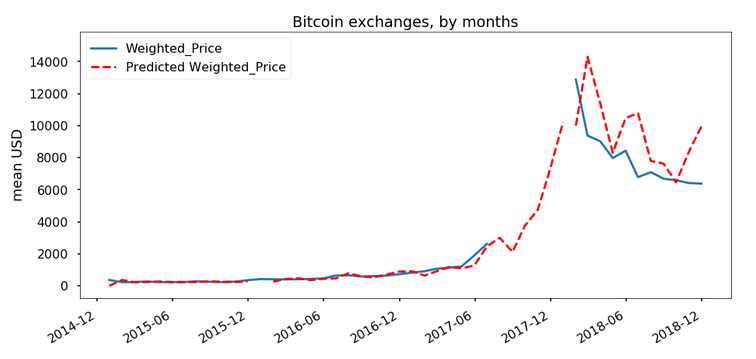
**Regular differentiation:**



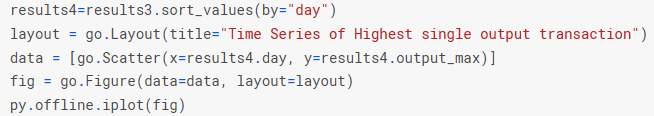


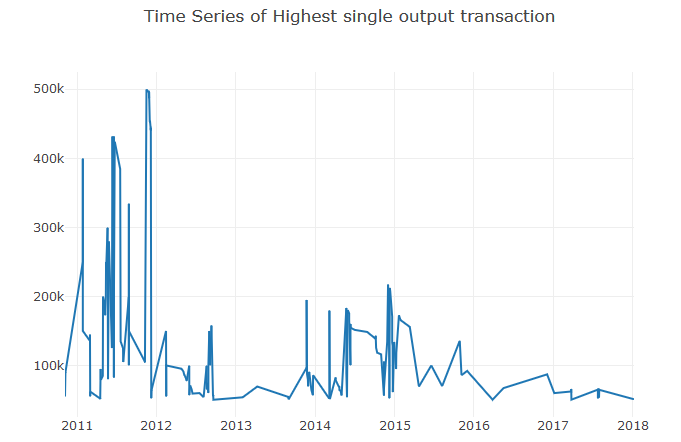
**Relation By Month:**



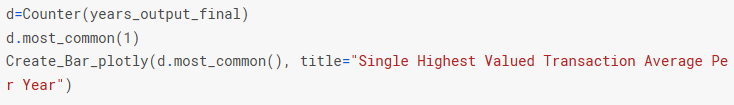


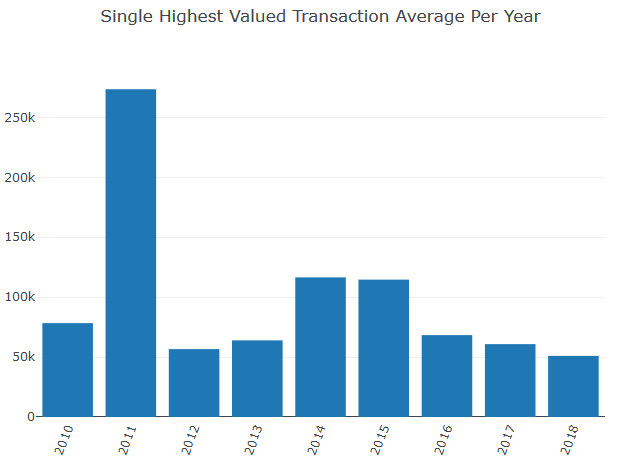
**Timeseries plot for highest valued transactions**

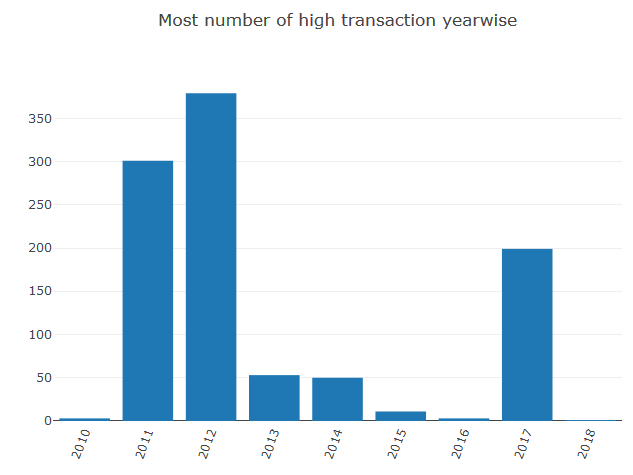




**Single highest Value Transaction average per Year**





**Which year have most number of top1000 transactions**

***CONCLUSION:***

This means that in 2016 there were just three transactions above 50343 Bitcoins !!!

Moreover in 2018 there was just one transaction above 50343,

Suprisingly in 2017, due to the high bitcoins price and its popularity there were 199 transactions that were above 50343 btc.

2012 seems to have majority of top1000 transactions, followed by 2011.

One thing to notice that high amount transaction got quite low in 2013,2014,2015 and very low in 2016.