*Qianhui Lu, Shenquan Wang, Yihui Zhou*

*SJSU | BUS 243*

***Database Design for CryptocurrencIES***

Table of Contents

[Business Benefits and Goals 2](#_Toc531713301)

[Business Requirements 3](#_Toc531713302)

[Key Features 5](#_Toc531713303)

[Functional Requirements 6](#_Toc531713304)

[Technical Requirement 7](#_Toc531713305)

[Infrastructure Requirement 8](#_Toc531713306)

[User Interface 8](#_Toc531713307)

[Test & Launch 12](#_Toc531713308)

[Results & Visualization 14](#_Toc531713309)

[Reference 18](#_Toc531713310)

# **Business Benefits and Goals**

Star Investment Fund is a wealth management firm with a legacy in serving wealthy individuals and families in Asia since 1980s. Recently the Star Investment Fund has opened up a branch in Palo Alto, aiming to acquire emerging clients in the technology field and to better understand leading disruptive technologies in order to explore investment opportunities for its clients in Asia.

Star Investment Fund has been passionate about incorporating advanced technologies in designing investment strategies. Since 1980s, the fund uses asset allocation framework, with fundamental research and value investing to manage assets for clients. The fund successfully avoided hazardous portfolio drawdown in Asia financial crisis in 1997, dotcom bubble in 2000, and global financial crisis in 2008 thus it has thrived till today. After financial crisis the fund has expanded into quantitative investing as solely relying on the traditional fundamental research does not meet the needs of the fund’s younger, hungrier clients any more.

Since its move to Palo Alto, the fund has been concerned about one major keyword: Cryptocurrency. The fund is struggling whether to incorporate cryptocurrency investing in its niche portfolio or not, considering a number of pros and cons. The annualized return on Bitcoin, for example, far surpasses that of any traditional asset classes, which allures many eager clients. The investment flows into cryptocurrencies through VC funds has been rising every year. Moreover, quite a number of private banks across the continents open up investment solutions on cryptocurrencies. That is evolving competition from within the industry. Meanwhile, prominent fund such as Y incubator is also looking to enter the high net worth market through its startup offerings which offers exposure to cryptocurrency investments.

Meanwhile, the extremely high volatility, unknown regulation requirements, ever evolving ICO landscape, and unstable coin exchanges have all added complexities and difficulties for Star Investment Fund to incorporate cryptocurrencies as part of their portfolio. Thus, the investment team has recently decided to perform an investment analysis on cryptocurrencies to examine risk and reward in entering the field.

The fund would like to obtain following outcomes from the database analysis project:

* Product level: narrow down the investment scope of more than 1,200 cryptocurrencies;
* Investment level: examine monthly, annual performance return and volatility;
* Trading level: examine ease of buy and sell, and volume;
* Risk level: examine several periods when there are dramatic movements in price and explore the reasons behind it; examine impactful country specific regulation risks such as in China.

The fund is aiming to base their decision on the database project and present to its board members.

# **Business Requirements**

As investment manager to many ultra-high-net-worth individuals, Star Investment Fund is only able to invest and propose investment decisions to clients that are suitable to clients and to the firm. In fact, its legacy has been built upon careful due diligence and research and very high standard in risk management. The firm adopts a risk-oriented perspective to make investment decisions and whether or not to onboard certain investment products. Thus, this project aims to solve the problem and provide advice based on the risks.

Below is a wide scope of risks the fund takes into consideration when it comes to cryptocurrency investing.

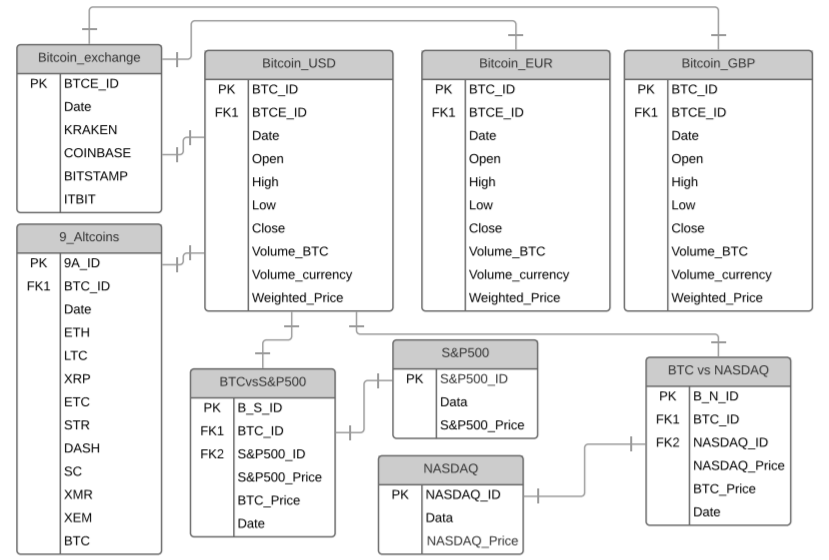
* Credit Risk: the risk that the issuer or the guarantor of a product such as a share or a bond is not able, usually for financial reasons, to repay principal or interest in relation to the product. The credit risk, to some extent, has been tackled in a large part due to the mechanism of cryptocurrencies.
* Market risk: the risk that the current value of a product falls as a result of movements in the market prices due; in particular, to changes in interest rates, foreign exchange rates, and equity and commodity prices so that the investor may not get back the money invested or may not make the returns as anticipated. This is a significant, integral risk of investing in cryptocurrencies. Though the price of cryptocurrencies is in large part decided by supply and demand; any change in regulations, any news that could remotely be linked to cryptocurrencies, could fluctuate the market easily.
* Liquidity risk: the liquidity of an investment is directly affected by the supply and demand for that investment and indirectly by other factors, including market disruptions or issues affecting the infrastructure on which the investment is traded such as a securities settlement system. The liquidity risk in cryptocurrencies links to the number of buyers and sellers in the market. This risk is not neglectable- when price of Bitcoin, for example, is high, the buyers rushed into the market as we have seen in 2017, when Bitcoin price went up 2000%; however, in 2018 when Bitcoin price has dropped more than 50% and for most of September and October the price has been lingering around $6000, the rush of buyers has died down and transaction volume has decreased.
* Currency/foreign exchange risk: the risk that currency exchange rate fluctuations may reduce gains or increase losses on foreign investments. For most Asian based clients, their wealth could in foreign currencies other than USD. For cryptocurrencies, majority of the coins traded are denominated in USD, JPY, or EUR, rarely in emerging market currencies. So, the problem of selecting an exchange and subsequently which currency to trade in for cryptocurrencies looms large for cryptocurrency investors.

In 2018, bitcoin price has undergone a sharp drop of more than 78% until this report writing, which is a sharp drop from close to $20,000 to now $3,750. Apart from the late cycle economy effect and Sino-US trade war tension, bitcoin has encountered a few fundamental challenges as well. One is that, Chicago Mercantile Exchange (CME) Group has opened Bitcoin futures trading this year giving investor means to trade from the other angle. Investors now have a powerful weapon to short bitcoins. The second one, is the creation and quick rise of Bitcoin Cash, which now ranks top 5 in the top traded cryptocurrencies. Furthermore, there is further division in the Bitcoin Cash into Bitcoin S.V. and Bitcoin ABC to see which one takes up 51% market cap first.

# **Key Features**

There are several key features for this project, including unique database design, easy using user interface and special centralized bitcoin analysis.

The database was designed base on the entity relation diagram (ERD) with the raw data from the open source. The ERD for this project is shown in *Figure 1*. By designing the ERD, the data is normalized, data duplication and redundancy were minimized. In order to keep the database safe, it is fully independent from the open source database. Once the API processing finished, the new local database was generated. Only admin user will have the privilege to perform such update. The unique database design continued with the abilities of data backup and recovery. To achieve those abilities, three different databases were created on three different team’s machines with same data. Furthermore, those machines crossed both Windows 10 and Mac OS X for extra safety.



*Figure 1. ERD for Database Design*

The user-friendly user interface provides straight-forward experience for users. Star Investment Fund could obtain specific data through the user interface by simple requirements selection and mouse clicks. After acquiring data, users could also save the data to their document folders.

Since the database was generated by the ERD, which designed to have entity of Bitcoin\_USD to be the center of other entities. The whole database was centralized around Bitcoin. In other words, the client could easily achieve special centralized Bitcoin analysis. For example, Star Investment Fund could compare price of different cryptocurrency with Bitcoin like Ethereum, Litecoin, Ripple with same date. Furthermore, the client would also be able to see the investment trend by discuss the cryptocurrency’s advantages over traditional financial tools like S&P500 or NASDAQ and create virilizations with the data that this project provided.

# **Functional Requirements**

Due to high standard of Star Investment Fund, this project had set specific goals about the accomplishments for clients at beginning. Those goals included ability of easy data inquiry, wide range of data selection, visualizations for different data comparisons and useful investment advices.

The main idea behind the database design was trying to accomplish easy data inquiry for clients. As database designer, we constantly remind ourselves of these priorities when developing ERD, sourcing data, spinning up the database and designing user interface. As a result, this project successfully delivered unique database and easy using user interface. Those features ensured robust yet user-friendly data inquiry process for our users.

Because investors are interested in diversifying the investment portfolio, the varieties of cryptocurrencies shall be analyzed. Thus, the database contained total of nine alternative cryptocurrencies in a table to be compared with Bitcoin. Meanwhile, investors are also interested in benchmarking traditional investment asset classes with cryptocurrencies. So, the database also included the data for S&P500 and NASDAQ. The wide coverage of benchmarking data ensures a comprehensive analysis of Bitcoin and cryptocurrency performance by our users and the investors from Star Investment Fund.

Various data visualization tools were employed to help our user grasp the trends and performance of the various assets in comparison with Bitcoin. Given the natural of our time-series data, line charts would show clear continuous data trends instead of data points from a MySQL table. Hence, various line charts were created to illustrate the price fluctuation. In this way, clients would have a better understanding for those specific item value in comparison to Bitcoin. Same theory was also applied for showing the correlations across the top alternative cryptocurrencies to give clients a clear idea of those correlations.

At the end of this project, the mission was to provide investment advice for clients on investment strategy for Bitcoin and cryptocurrency asset class. Even though the advice itself is simple, the process of analyzing the data and arriving at the conclusion is a meandering process. As expected, the final advice was generated with needed avidness and discussion.

# **Technical Requirement**

As mentioned earlier, we initialized our database design by creating an ERD. Lucidchart is used to create the ERD. To develop this database, we then determined our project environment. We decided to use Python environment given its flexibility to work with MySQL, the availability of python packages to work with data, for data visualization, and its strong community.

Next step in our database design is sourcing our data. As discussed earlier, we are looking to incorporate various data into our database including: Bitcoin price in other currencies, Bitcoin trading volume in other currencies, Bitcoin price daily fluctuation, alternative coins performance, benchmarking traditional stock market performance. Due to lack of central management entity, the price of cryptocurrency also varies slightly based on the exchanges it is traded on. Thanks to the robust cryptocurrency open source community, we can get cryptocurrency prices from four of the major cryptocurrency exchanges from one source, Quandl. They provide user-friendly API for us to call for the data in our Python Jupyter Notebook environment. After acquiring the prices there, we took the average of prices from these exchanges. For alternative coins performance, we used the API from another cryptocurrency exchange, Poloniex. For data on stock market, we got our data from Yahoo! Finance. After we obtained the data, we created the database in MySQL Workbench. Then we went back to Python to compose SQL query, produce visualizations and develop graphical user interface.

# **Infrastructure Requirement**

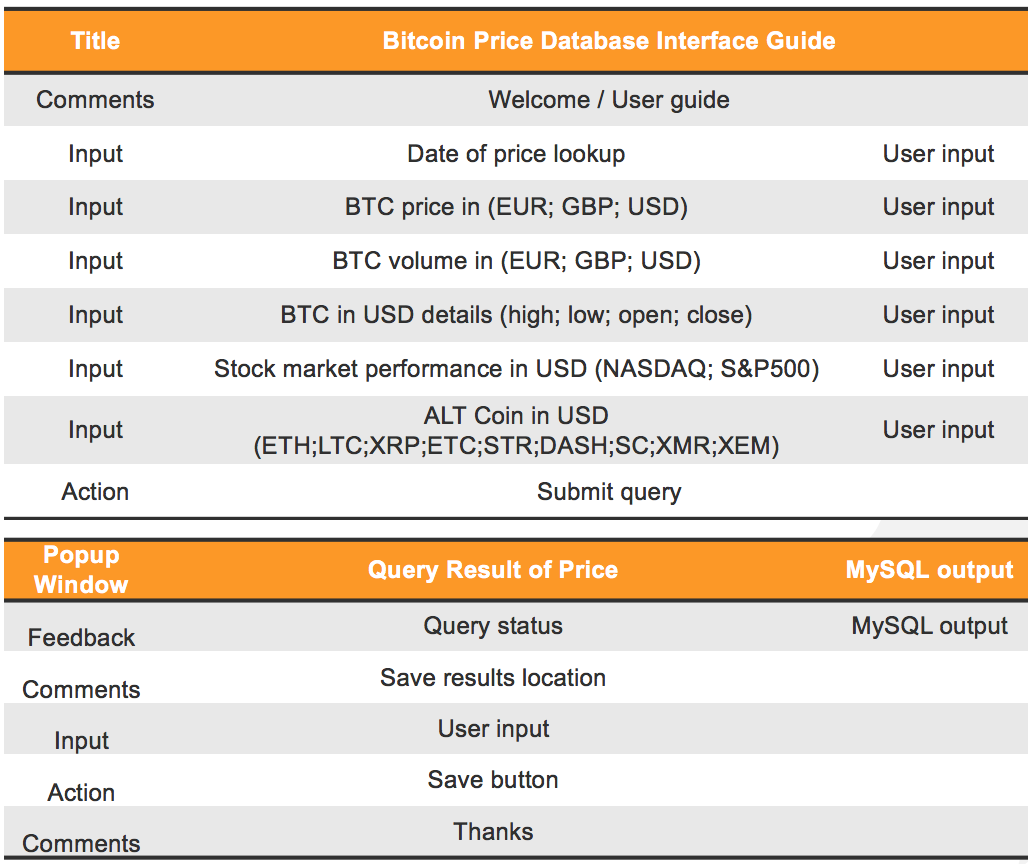
Python is our main working environment. The Python packages we used includes: Numpy, Pandas, Pickle, Quandl, Plotly, Sqlalchemy, Pymysql, Tkinter, Pandas, Time, Messagebox, Datetime, Re, and Os. We used MySQL Workbench to create schemas, designate users and their respective privileges, create table views and test the SQL queries.

We also explored tools like HTML with PHP for our graphical user interface. Given the fact that PHP is a server-side language, for every query we execute, we would have to refresh the site to make sure the query pass through the PHP server of choice again. As mentioned earlier, our goal is to create an easy to use interface for our users. Hence, we decided to not to use PHP and strictly use Python from data process to data visualization to UI.

As mentioned earlier, we used computers with Windows 10 OS and Mac OS as hardware. Python program can be run standalone with no need to set up external servers.

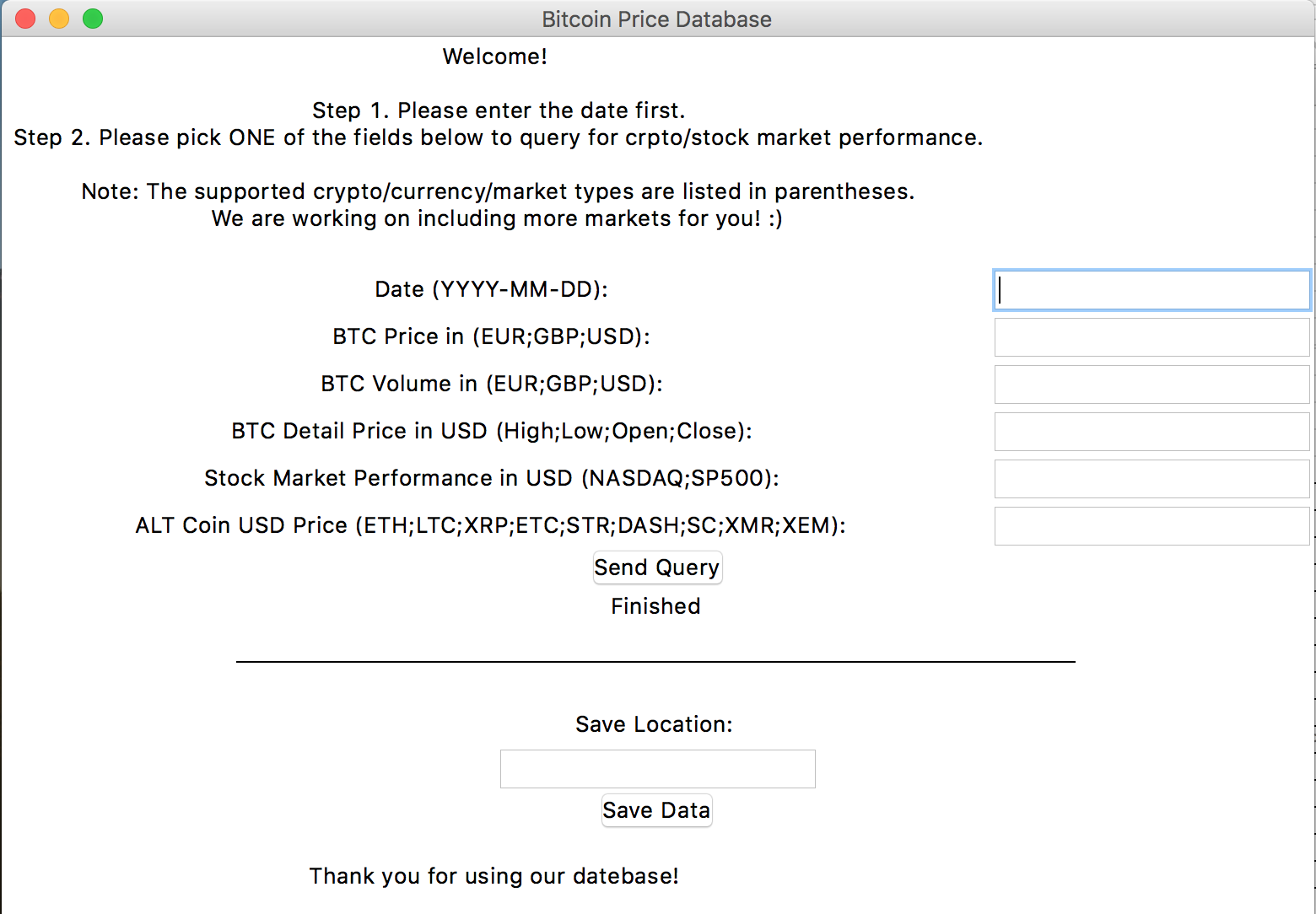
# **User Interface**

Before we started building the graphical user interface, we first sketched out the schema for the UI, which is shown in *Figure 2.* We will have some comments to welcome the users, and to provide some brief instructions for them. Then we’ll have the inputs sections. After which, we will have a submit query action button. A popup window will then show up to provide users with the query results or any additional feedback or error messages. Then the query status will be displayed, and the user will have the option to input a local directory path if they’d like to save the data in .csv format. It is illustrated in the *Figure 2*.



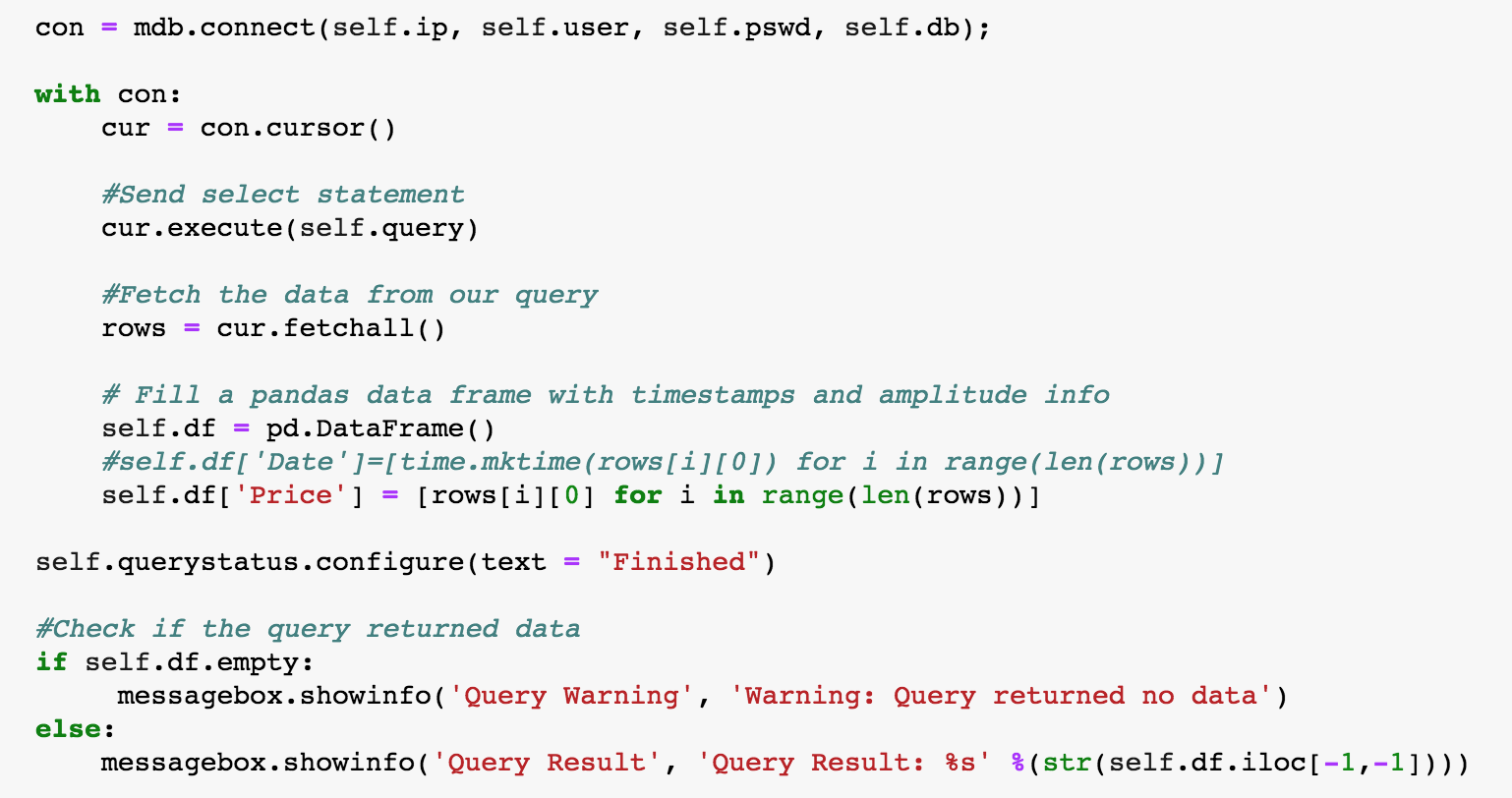
*2. Graphic User Interface Schematic*

After designing the schematics, we moved on to actually coding the graphic user interface. We again evaluated the available technologies like HTML, PHP and decided to continue with Python to develop a simple but robust UI for our users. Python’s de facto GUI package, tkinter is want we ended up heavily reply on for the UI. Seeing *Figure 3* for what our actual UI looks like.

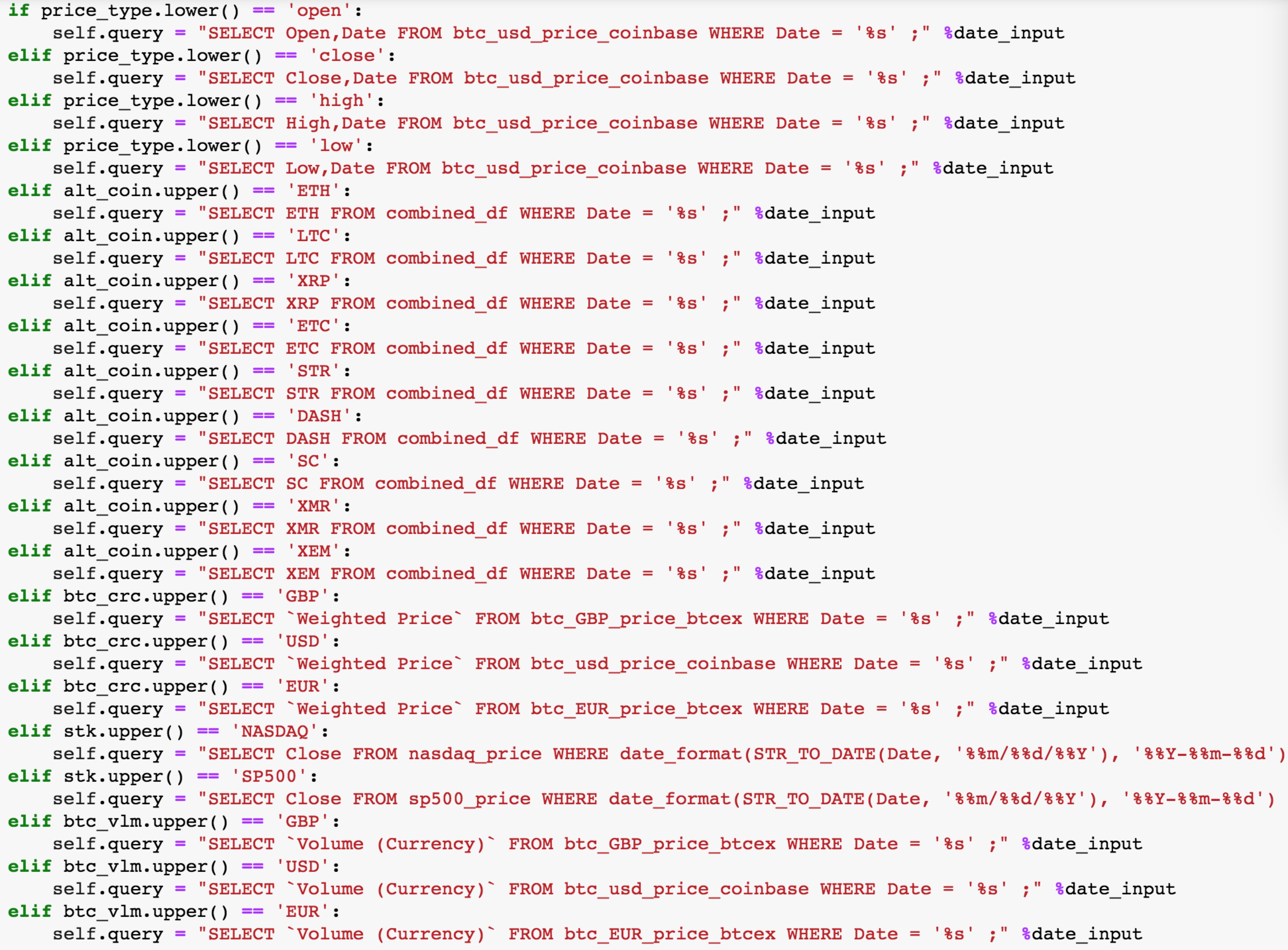


*Figure 3. Completed GUI (Initial Window)*

When developing the GUI, the panels were designed first. The panels can be separated into query panel and saving panel and a line is used between the two to delineate. Then we wrote two functions for sending the query and saving the file respectively. In the function of sending the query, we first parsed the user inputs to validate the user input. We created error handling variable to confirm the input format is okay, otherwise, an error message will be returned to the users. Then we set the queries and established the connections to MySQL from python, which is shown in *Figure 4.* Then we compose the file-saving function. Similarly, we also will prompt an error message if the user directory inputted is invalid.



*Figure 4. Connection to MySQL to fetch query*



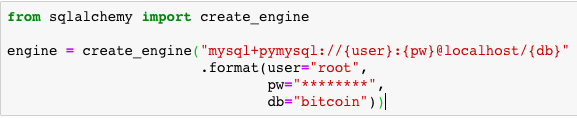
*Figure 5. SQL Query written in Python as part of the query function*

We learned an important and interesting lesson that while writing SQL in Python, we need to be extra careful when using writing the time format like %m/%d/%Y because the code will go through Python compiler first and will be read as python’s callable object. Yet %m is not a legal callable object. Hence it would return error message to us. To avoid such error, we need to use %% so it escapes the % and one % will be left for SQL to compile and read, which is shown in *Figure 5.*

# **Test & Launch**

We chose to work with Jupyter notebook in conjunction with Python. A few python packages are used in the project including Pymysql, Sqlalchemy to transport data from notebook to database.

Sample coding are attached in *Figure 6* and *Figure 7*.

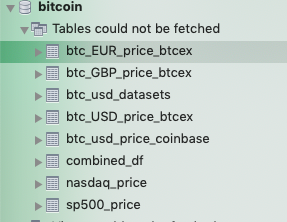


*Figure 6. To connect Jupyter notebook to MySQL to load data*



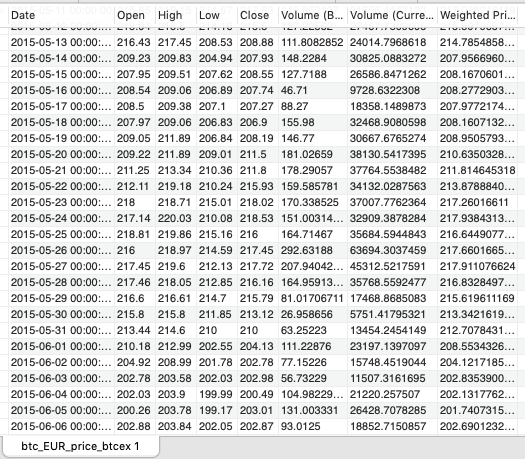
*Figure 7. To create scheme in database and transport data*

After transporting all data from Jupyter notebook to database in created schema “Bitcoin”, we get the database ready, which are shown in *Figure 8*, *Figure 9* and *Figure 10*.

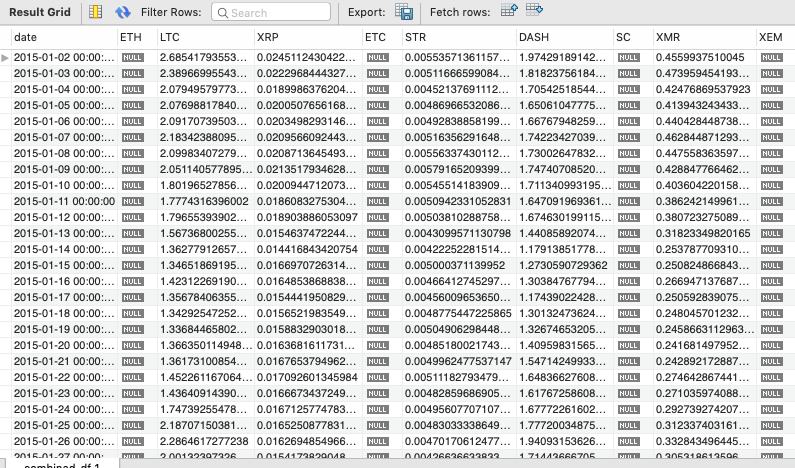


*Figure 8. Tables in bitcoin schema*

Below is a screen shot for the databases built in MySQL Workbench.



*Figure 9. Table view for BTC to EUR*

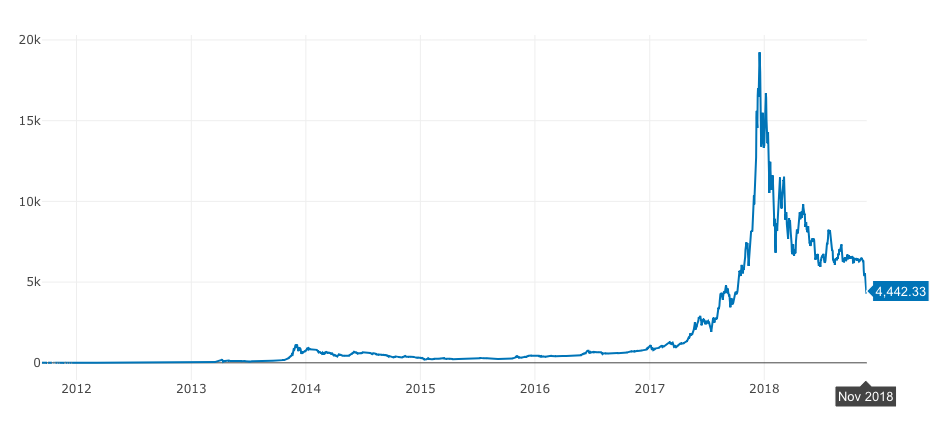


*Figure 10. Table view for alternative coins*

# **Results & Visualization**

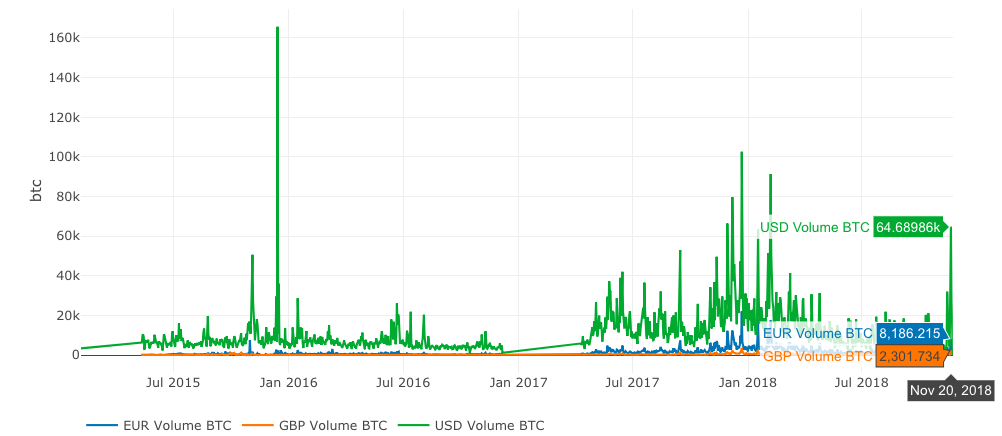
We have worked on 5 parts of visualization based on the database, query and research.

First chart shows the historical price of bitcoin since 2012 when trading started in multiple exchanges. Bitcoin price has returned close to 20million times since the lowest price to highest price in end of 2017. In 2018, the speed and magnitude at which bitcoin price drops is astonishing- this year alone, bitcoin price has vanished 78%, which is shown in *Figure 11.*



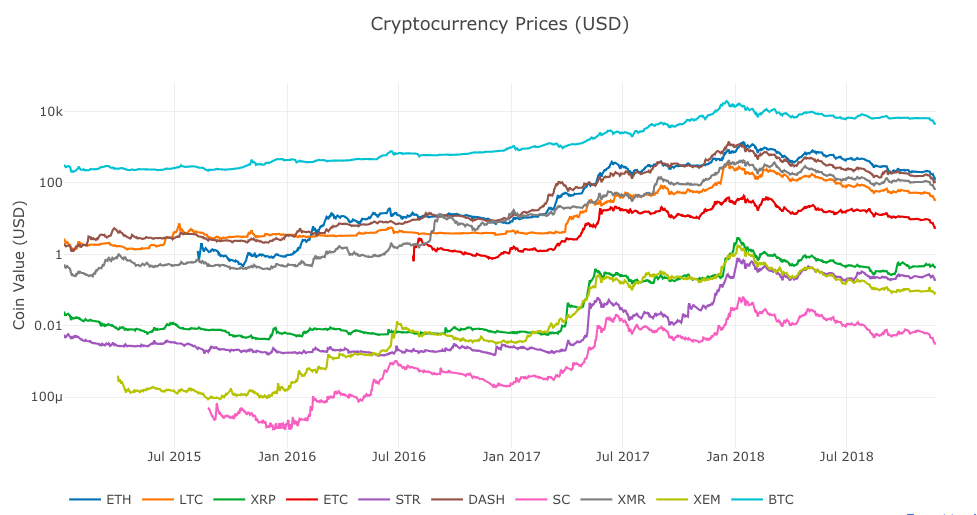
*Figure 11. Bitcoin Price*

*Figure 12* shows the trading volume with different denomination currencies, USD, EUR, and GBP. It is quite clear from the chart that USD is the dominant currency bitcoins are traded in.



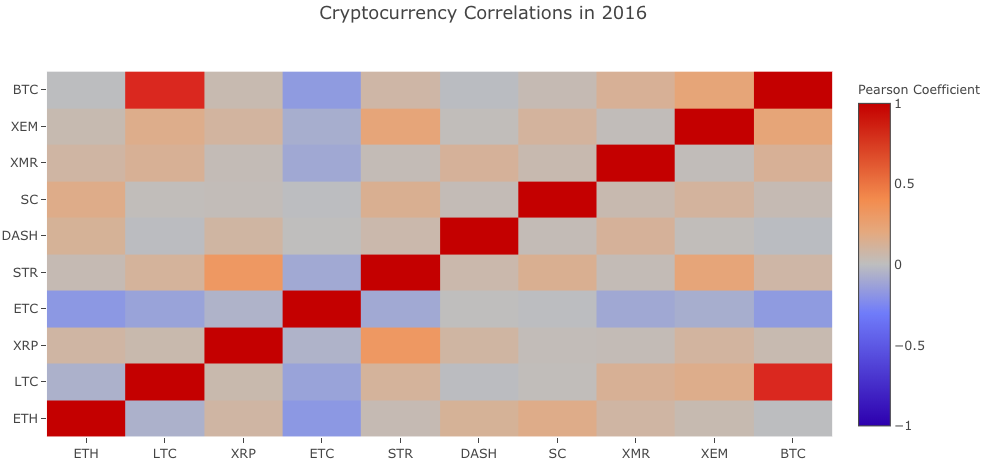
*Figure 12. Bitcoin Trading Volume in different currencies*

*Figure 13* shows historical price changes between 10 cryptocurrencies including Bitcoin.

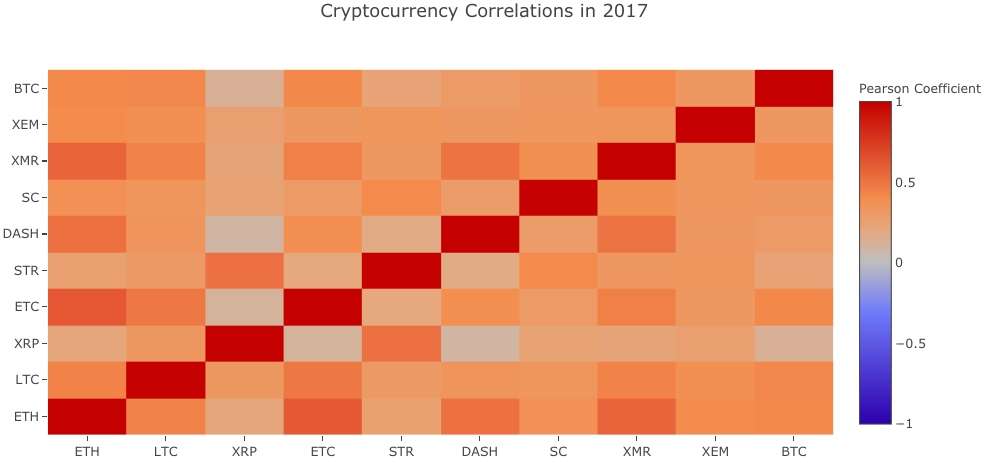


*Figure 13. Alternative Coin performance*

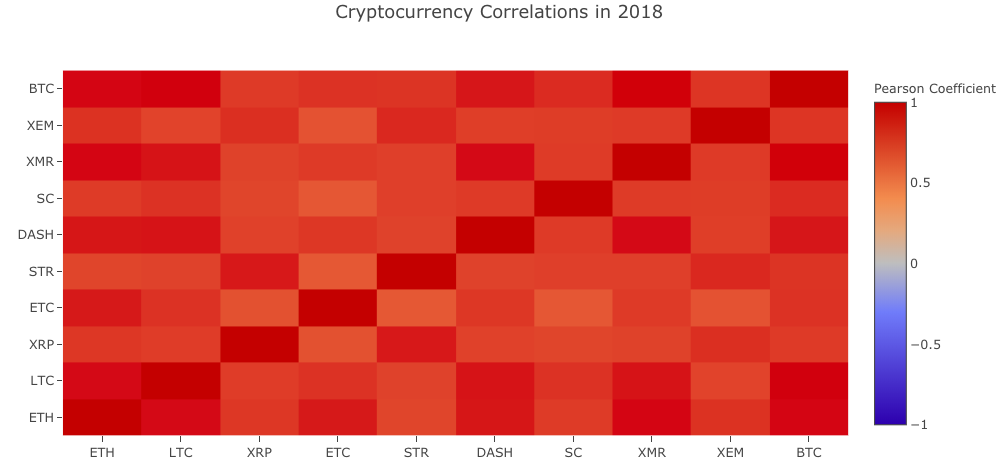
Fourth part, we took a further look at the Pearson Correlations across three years (2016,2017, and 2018) between 10 cryptocurrencies. It is very clear that in 2016, cryptocurrencies are less correlated, with in fact many of them showing negative correlation, which is shown in *Figure 14.* In 2017, the correlations turn positive for many of the cryptocurrency pairs, which is shown in *Figure 15.* In 2018, all cryptocurrencies are highly-correlated and they went down together, which is shown in *Figure 16.* In the investment world, high correlation between assets is scarier than falling prices.



*Figure 14. Alternative Coin performance correlation in 2016*

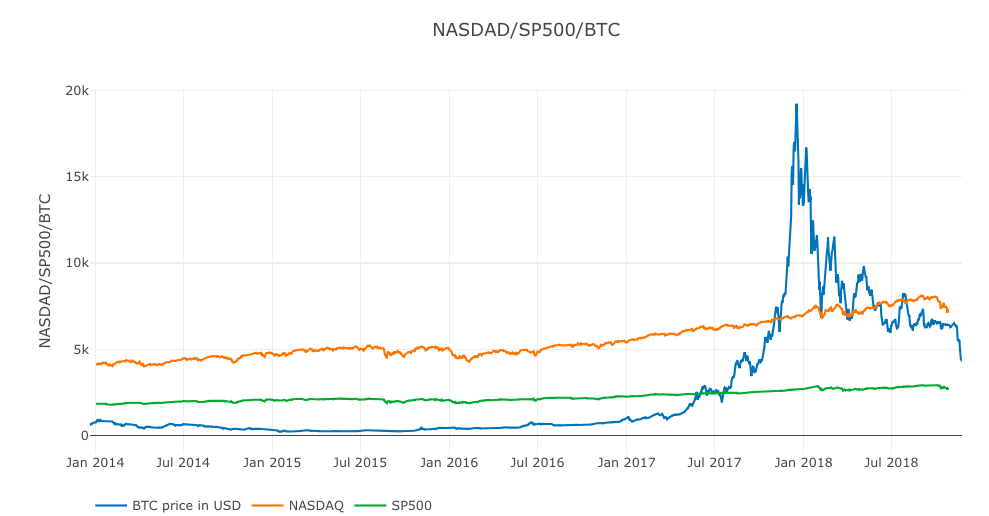


*Figure 15. Alternative Coin performance correlation in 2017*



*Figure 16. Alternative Coin performance correlation in 2018*

*Figure 17* compares Bitcoin prices with traditional asset prices such as S&P500 and NASDAQ.



*Figure 17. Bitcoin performance benchmarking with stock market*

# **Reference**

Sarah, P. (2017). Analyzing Ethereum, Bitcoin, and 1200+ other Cryptocurrencies using PostgreSQL. Retrieved from Medium | Timescale, <https://blog.timescale.com/analyzing-ethereum-bitcoin-and-1200-cryptocurrencies-using-postgresql-3958b3662e51>

Patrick,T.(2017) A Data-driven approach to cryptocurrency speculation, Retrieved from <https://blog.patricktriest.com/analyzing-cryptocurrencies-python/>

Silota, (2018) Using SQL to analyze Bitcoin, Ethereum & Cryptocurrency Performance

Retrieved from Silota, <http://www.silota.com/docs/recipes/sql-bitcoin-ethereum-cryptocurrency-analysis.html>

Quandl (2018), Quandl API for Bitcoin Data, Retrieved from QuandlBlog, <https://blog.quandl.com/api-for-bitcoin-data>

Poloniex API (2018), API documentation, Retrieved from Poloniex, <https://poloniex.com/support/api/>

Dealing Data (2016) Making a MySQL GUI with Tkinter, Retrieved from Dealing Data, <http://www.dealingdata.net/2016/08/21/Python-MySQL-GUI/>