

# Cryptocurrency Data Visualization

## Business Case 5



Marjorie Kinney *m20210647*

Bruno Mendes *m20210627*

Lucas Neves *m20211020*

Farina Pontejos *m20210649*

Business Cases for Data Science

NOVA Information Management School

May 2022

[Code](#) · [Dashboard](#)

## **Table of Contents**

<b>Introduction</b>	<b>1</b>
<b>Business Understanding</b>	<b>1</b>
<b>Data Understanding &amp; Preparation</b>	<b>1</b>
<b>Design &amp; Deployment</b>	<b>2</b>
<b>Conclusion</b>	<b>4</b>

## Introduction

Data dashboards allow users to monitor multiple key performance indicators at once through the use of visual graphics. This holistic view can assist users in making informed decisions through the use of data-driven measures, rather than through intuition alone. A dashboard is typically accessed through a web browser and consists of multiple graphics and tables. In the case of a financial asset dashboard, an analytic approach is required. This type of dashboard should include comparisons and historic data, as well as the ability to drill down into underlying details.

## Business Understanding

Investments4Some is a long-standing Portuguese, privately-held hedge funds management firm. Previous work with this company involved creating predictive models for the closing value of ten cryptocurrencies to be used with their clients.

Partners of Investments4Some have now requested a flexible dashboard to be used to inform investment decisions for both financial analysts from their internal financial team and for external stakeholders. Although information about the cryptocurrencies is important, they would also like to be able to access information about as many financial assets as possible.

Because information about assets such as stocks and cryptocurrencies change over time, the use of a dashboard can be extremely useful for monitoring these changes. Being able to see if an asset is increasing or decreasing in volume, volatility, value, or momentum can help inform decisions about when to buy or sell these assets. Current and historical data must be available at the same time to help anticipate changes in the market. Predictions for future asset values, based on machine learning algorithms, can be included to assist in these decision making processes as well.

## Data Understanding & Preparation

As requested by the client, we are retrieving the information about the assets and making forecasts in real-time. The main source of market data is the Yahoo Finance API. We are able to retrieve the *Open*, *High*, *Low*, *Close*, *Adj Close*, and *Volume* of any given financial asset between any two specified dates, in this case between 100 days previous, and the current date.

All of these columns are given as float values. In addition, we are also able to retrieve the Ticker information from the same API. Specifically, we download the given stock symbol's sector and industry labels if available, or its description otherwise. We also retrieve the URL to the stock symbol's logo if it is available.

In addition to data about the selected asset, we also use the Twitter API to retrieve the most recent tweets that mention this keyword. We download the contents of the tweets, and perform some text processing operations on it to enable us to analyze the general sentiment relating to the asset.

Unlike cryptocurrencies, traditional financial markets are not always open. As such, weekends will reflect null values for these traditional assets. Instead of dropping these rows, we replace the null values under the Volume column with zeros, and use the last known Close value for the rest of the columns. We drop the Adj Close column as it is highly correlated, and most of the time equal, to the Close column.

Some stock symbols do not have values for sector or industry specified. In these cases, we instead check if the description is available and display the contents of this field. We retrieve and display the logos of those symbols where this is available.

Before we are able to perform sentiment analysis on the tweets we download, we first need to perform some text processing operations. Firstly we discard those tweets tweeted by authors with “bot” in their names, since these are automatically generated tweets made by software, and do not reflect human sentiment. After filtering these out, we remove stop words and symbols, as well as any mentions of other Twitter usernames. We also remove any extraneous whitespaces.

Finally, we use the *SentimentIntensityAnalyzer* package to return the polarity scores of each tweet. These results are composed of three parts: the neutral, positive, and negative scores. We average all these values across all the tweets retrieved to obtain a general polarity score, which can be considered a measure of how positive or negative the emotions of the tweets are.

## Design & Deployment

Based on our previous work for the client, we were able to train several models for each cryptocurrency coin specified in order to obtain the best parameters to use for each. This process however takes quite a long time to run. As the client wished to be able to view these predictions in real-time, and across different assets, it is not practical to perform a similar model-fitting process.

As such, we selected one of the models that consistently performed well across different assets. Specifically, we used the *GradientBoostingRegressor*, fitted on the seven days prior to the day being predicted.

In creating the dashboard, several design best practices were followed.

**Visualization types:** the data visualization types were selected based on how to relay the information as clearly and simply as possible. Line charts were chosen to represent asset information over time, and a tree map was chosen to represent market volume. A simple graphic was selected to summarize sentiment.

**Layout:** the layout was optimized to show the most relevant information on the top left, which is the area the eye is drawn to first in Western cultures, whereas the supplemental information was placed on a sidebar on the right.

**Clarity:** the number of visualizations was limited in order to avoid clutter and overcrowding.

**Color:** visualizations were kept as color-neutral as possible, unless encoding specific information, such as the green-yellow-red encoding for positive-neutral-negative sentiment.

**Interaction:** visualizations were created with embedded callbacks so that users can interact with the available data as needed. Hovering over various areas allow the user to view additional details encoded in the chart, and filters can be applied by the user to the second set of visualizations to select which technical analysis indicators to view.

The dashboard layout consists of two main areas. The main body occupies 75% of the available width, and the rest is occupied by a sidebar section.

The main body contains the focus of the dashboard, which are the financial plots. The first one that the user sees is the time series plot of the selected stock symbol. By default this displays the last 100 days of data. On the same panel, the user can click on another tab to show the candlestick plot of the asset.

Below the main time series plot are the various Technical Analysis indicators. We plot the five different indicators which we found to be the most relevant in our previous analysis: the *Bollinger Bands*, the *Moving Average Convergence Divergence*, the *Momentum Relative Strength Index*, the *Volatility*, and the *On Balance Volume*.

The last panel contains some information about the market in general. A treemap visualization showing the market capitalization of the constituents of the S&P500 is presented, with the assets grouped by their *Sector*. The last 30 days of *Close* prices of the top 10 assets by market capitalization is shown as sparklines next to it. While it would be ideal to retrieve this information in real-time, it is not feasible within the limitations of the API we were using. Each constituent's ticker information, which contains its market capitalization, took on average a minute each to be retrieved. We considered it to be detrimental to the user experience if one had to wait as long as it would take considering the large number of assets involved.

Some additional information is displayed in the sidebar area. The first element on the sidebar is the symbol selector. This is where the user would select and change the symbol to display. Changing this selection fetches the relevant symbol from the Yahoo Finance API and recalculates the predictions as well as regenerates the plots displayed.

Below the symbol selector is a short profile of the selected asset retrieved using the Yahoo Finance API. It displays the full name of the asset, its logo, its sector, and the industry to which it belongs, if available. If these details are not available, it instead displays a short text description.

Following this section, we display the result of the sentiment analysis of the 100 most recent tweets relating to the asset, retrieved using the Twitter API. This small number is a result of the rate limitations set by the free tier of the service. Three smiley face icons are displayed, colored red, yellow, and green, to represent negative, neutral, and positive sentiments, respectively. The opacity of each icon is determined by its polarity score, capped at a minimum of 20% for aesthetic purposes.

We have also included links to the sources of the data we used, as well as the link to the repository containing the code that generates the dashboard.

We used the Dash package to build our interactive dashboard. While the web app itself is written in Python, it is rendered in the browser, making it easy for the clients to access and

use. A Dockerfile is provided, as well as a Docker compose file, to enable Investments4Some to deploy a containerized version of the app to any server of their choice. Additionally, we deployed the dashboard to the Google Cloud Platform, making it accessible online.

## **Conclusion**

The dashboard facilitates analysis by displaying historic prices and relevant technical indicators for various assets, along with general sentiment derived from twitter posts. The forecasting can also be valuable, but it is not a substitute for informed expert opinion since these types of algorithms won't predict major market shifts in a timely manner.

In the future, we could use paid features like an unlimited number of tweets and better cloud service. The asset predictions associated with this dashboard should be monitored for performance. Periodically testing multiple models and parameters and selecting the best combination for the majority of the assets will ensure the best possible dashboard performance.