

# SIT220/731 2022.T3: Task 2P

## Working with *numpy* Vectors (Unidimensional Data)

Last updated: 2022-10-30

### Contents

<b>1</b>	<b>Task</b>	<b>1</b>
<b>2</b>	<b>Additional Tasks for Postgraduate (SIT731) Students (*)</b>	<b>2</b>
<b>3</b>	<b>Artefacts</b>	<b>3</b>
<b>4</b>	<b>Intended Learning Outcomes</b>	<b>3</b>

### 1 Task

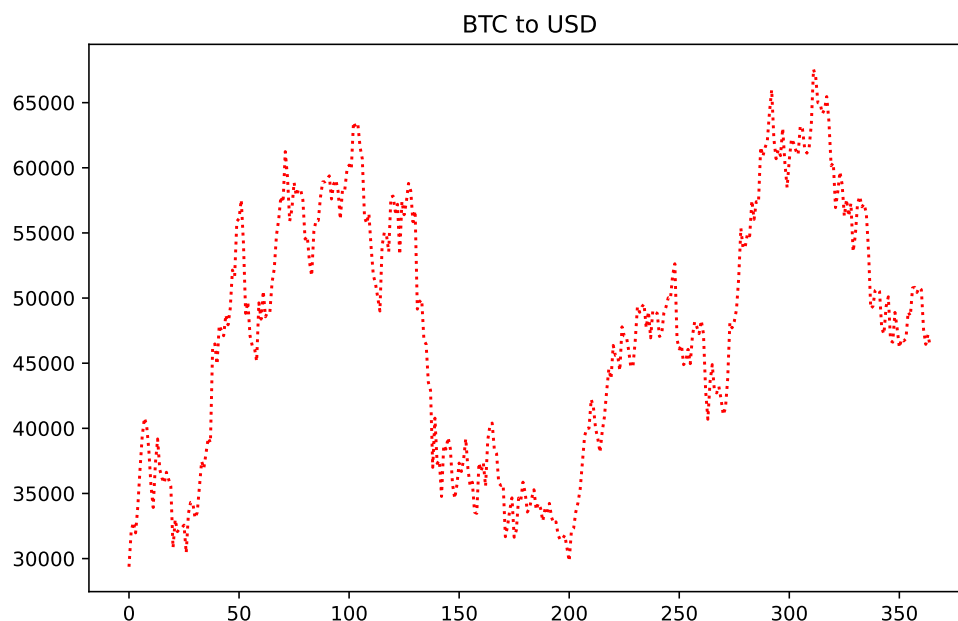
Create a single Jupyter/IPython notebook (see the *Artefacts* section below for all the requirements), where you perform what follows.

*Note that all your code must work correctly if someone decides to load an updated version of the CSV file mentioned in Step 1.*

1. Download the latest daily close BTC-to-USD data, from 2021-01-01 up to this day, available at <https://finance.yahoo.com/quote/BTC-USD> (the *Historical Data* tab).

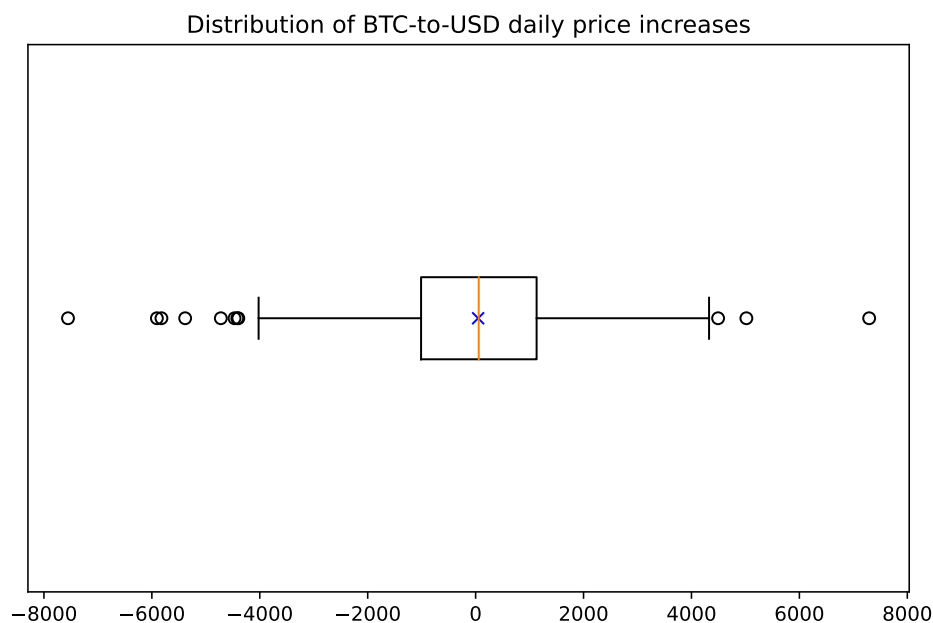
Use a spreadsheet application such as LibreOffice Calc or MS Excel to remove everything except the numeric values in the *Close* column. The column labels should also be deleted. If there are any missing data, remove them manually. Export these to a CSV file (which should only contain numbers, one per line).

2. Use `numpy.loadtxt` to read the above BTC-to-USD data as a *numpy* vector named `rates`.
3. For the 2021 data *only* (first 365 days), determine and print the following aggregates:
  - arithmetic mean,
  - median,
  - minimum,
  - maximum,
  - standard deviation,
  - interquartile range.
4. Call `matplotlib.pyplot.plot(rates, <...further_arguments...>)` to draw the data as a function of the day number (with 0 being the first day), using *red dotted line segments*. Call `matplotlib.pyplot.title` to add the plot title.



5. Using `matplotlib.pyplot.boxplot`, draw a *horizontal* box-and-whisker plot for the 2021 daily price increases/decreases as obtained by a call to `numpy.diff`. In your own words (2-3 text paragraphs), explain what can be read from such a plot.

*The blue “x” is part of the postgraduate requirements only, as explained below.*



## 2 Additional Tasks for Postgraduate (SIT731) Students (\*)

Postgraduate students, apart from the above tasks, are additionally **required** to solve/address/discuss what follows.

1. Using an additional call to `matplotlib.pyplot.plot`, mark the arithmetic mean on the box plot with a blue “x”.
2. Determine the day numbers (with 0 being the first day) with the lowest and highest observed prices in 2021.
3. Count (programmatically) how many outliers are there on the boxplot (for the definition of an outlier, consult Section 2.3. of our learning materials on the unit site). In your own words, explain what such outliers might mean in the current context.

### 3 Artefacts

The solution to the task must be included in a single Jupyter/IPython notebook (an .ipynb file) running against a Python 3 kernel.

At the start of the notebook, you need to provide: the **title** of the report (e.g., *Task 42: How Much I Love This Unit*), your **name**, **student number**, **email address**, and whether you are an **undergraduate (SIT220)** or **postgraduate (SIT731)** student.

Make sure that your notebook has a **readable structure**; in particular, that it is divided into sections. Use rich Markdown formatting.

Imagine it is a report that you would like to show to your manager or clients — you certainly want to make a good impression. Check your spelling and grammar. Also, use formal language.

Before each code chunk, briefly **explain** what its purpose is. After each code chunk, **summarise and discuss the obtained results** (in a few sentences).

---

Submit one file via OnTrack:

1. the version of the Jupyter/IPython notebook converted to a PDF file (e.g., via *File* → *Export Notebook As* → *PDF* or convert to HTML and from that to PDF with your web browser; any method will do).

You do not need to submit the .ipynb file via OnTrack, but you must store it for further reference – a marking tutor might ask for it later, e.g., at the end of the trimester.

### 4 Intended Learning Outcomes

ULO	Is Related?
ULO1 (Data Processing/Wrangling)	YES
ULO2 (Data Discovery/Extraction)	YES
ULO3 (Requirement Analysis/Data Sources)	YES
ULO4 (Exploratory Data Analysis)	YES
ULO5 (Data Privacy and Ethics)	YES