

Experiment Overview

Introduction

The current global Corona crisis not only has a huge impact on health systems around the world, but also on the economic situation. Gold is traditionally known to be a go-to asset in times of crisis and uncertainty and its price therefore usually increases when stock indices collapse. However, the Austrian newspaper Kurier recently published an article ¹ claiming that current market activity during the Corona crisis contradicts this theory. They also state that the gold price will stabilize again in the medium term as this also happened on other occasions.

This experiment aims at a data-driven approach to make validations of these assumptions easier. The development of the number of Corona cases and deaths in each country over time is compared to the corresponding gold prices. Since we have not yet overcome the current situation, validating assumptions about medium-term consequences on the gold price cannot be made now, but might be with future data.

Architecture

The tool acts in two phases. In a first step, data about Corona infections and the gold prices have to be imported from separate CSV files.

The Corona file has to have at least the following column definitions: `day`, `month`, `year`, `cases`, `deaths`, `countriesAndTerritories`. `Day`, `month` and `year` obviously describe the date of the record, whereas `cases` and `deaths` show the number of new infections/deaths on this day in the country, `countriesAndTerritories` contains.

The gold price file must at least consist of the two columns `Datum` and `Schlusskurs` (German for date and closing price) which describe the price of a troy ounce gold in US-Dollar for that date.

The data import is done in the `preprocess.py` Python script. This script also handles all the data preparation steps needed for later visualization. Firstly, the number of Corona cases and deaths has to be accumulated over time grouped by country because the input data only contains the newly confirmed cases/deaths per day. Additionally, only records between user-supplied start- and end-dates are considered. However, days without cases (after being accumulated) are ignored to only focus on the timespan with confirmed cases. Some preparation is also necessary for the gold price data. Gold is not traded on weekends which results in lack of data for those days. To be able to draw a continuous curve, the gold price on weekends is set to the then last known value (Friday).

Then, both data are merged into a single temporary CSV file with columns `countriesAndTerritories`, `date`, `cases`, `deaths`, `price` along the date.

The `visualize.py` script then reads the temporary file, creates directories for all countries present and splits the file into one file (`data.csv`) for each country within the corresponding directory. It also draws a diagram (`diagram.png`) for each country containing curves for the number of cases and deaths and the gold price. The overall architecture of the tool can be seen in Fig. 1.

Fig. 2 shows the diagram for Austria as an output example.

¹<https://kurier.at/wirtschaft/coronavirus-verliert-gold-seinen-ruf-als-krisenwaehrung/400787762>

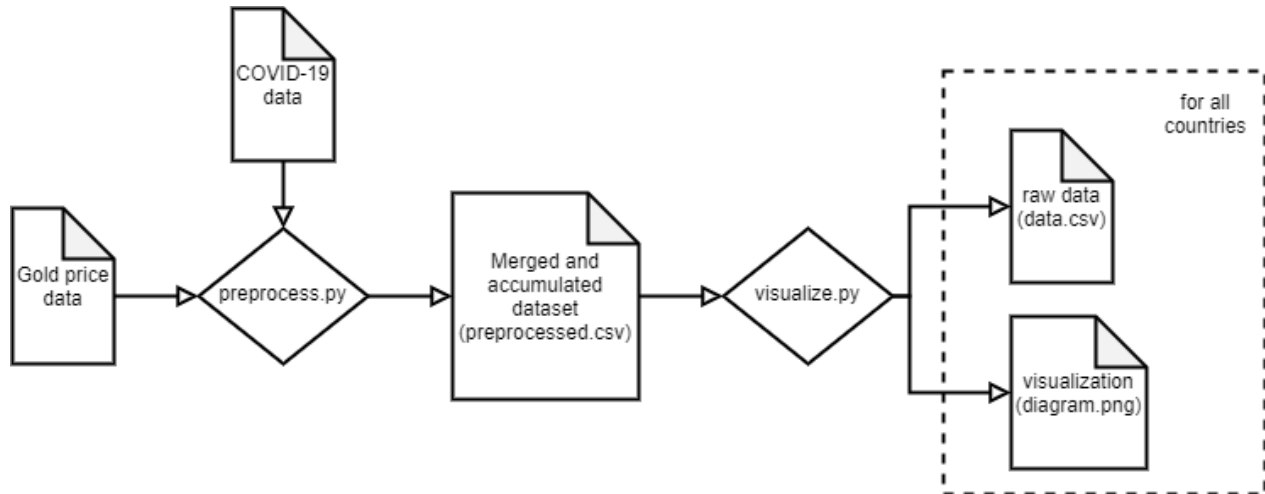


Figure 1: Logical architecture of the data analysis tool

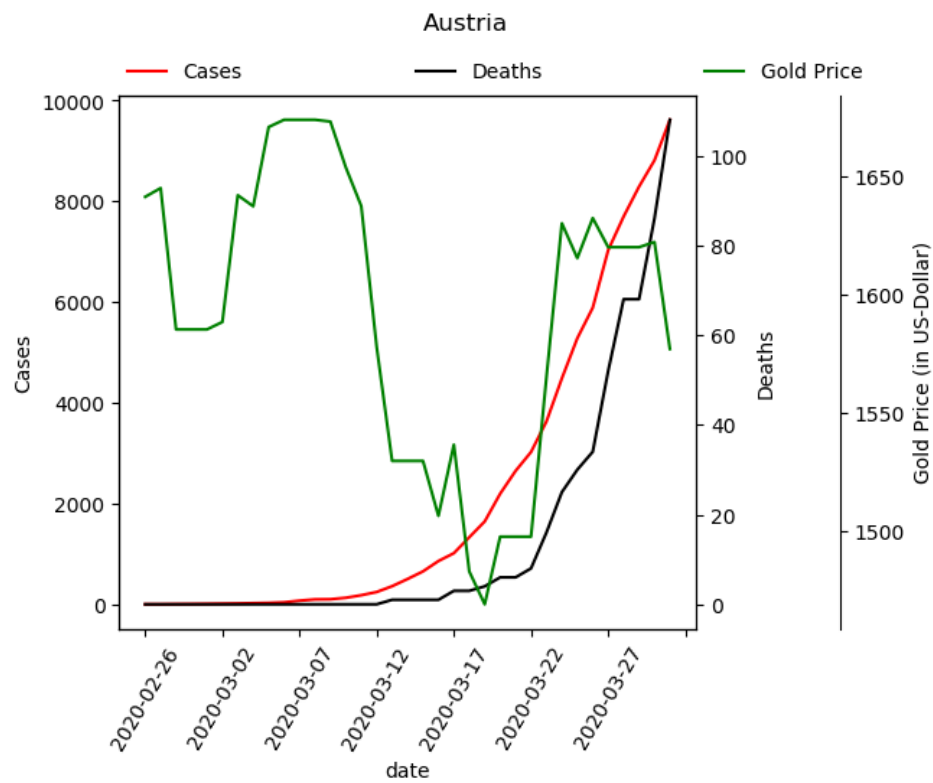


Figure 2: Output diagram for Austria