

# Requirements TSADF

## Features to implement

### **Input**

- Data file upload, requirements:
  - it should be in the supported .csv format
  - there must be time stamps (with a recognized format) present in the first column of this file
  - the file should be no larger than a certain size
- Time-series frequency:
  - automatically test whether can be derived from time stamps
  - otherwise calculate time-series frequency by input time span and interval
  - if time stamps not equally spaced, message that they should fix the input file
- Acceptable value range [0-1000]
- Possible seasonal patterns [hourly, daily, weekly, monthly]
  - determine seasonality by showing first 100 datapoints
  - limited to provided choices
  - optional: compare seasonalities including no seasonality
- Interactive threshold selection (*log2 range* times) [automatic, interactive]

### **Visualisation**

Based on the required inputs above, there should be five clear distinct sections in the application:

1. The first section should be dedicated to selecting the input file which the program will use to run. There should be clear instructions regarding the requirements of the file. After selecting the file, these requirements will be checked and displayed with the help of a pass or fail symbol per requirement. Accordingly, the application will proceed or display a message with the errors (and re-prompt for an input file).
2. The second section should be dedicated to determining the time-series frequency (TSF). The program should display a loading wheel indicating that the program is trying to derive the TSF from the time stamps. If this succeeds, a passing symbol should be displayed and the program should proceed. If the TSF can not be derived automatically, the user is prompted for the time span and the interval in [seconds, minutes, hours, days]. There should also be an option where the user can let the program know that the time stamps are not equally spaced, after which the program will display a message that the user should fix the intervals or select another input file.
3. The third section is simple. Here the user is displayed two input fields which both require a number, together representing the acceptable value range.
4. In the fourth section the user is asked to select the seasonality of the input file by entering a number followed by a time unit from a dropdown menu. This number represents the length of one seasonal pattern in the time series data. If the user is not sure, the user can click on a help button, and the first 100 data points are shown to assist the user in making his decision.
5. Optional: the fifth section. In case the user has selected interactive threshold selection, an  $n$  number of patterns should be displayed after which the user is presented the question: "Do you think most of the red points are outliers?" [yes, no]. This process repeats until all of the patterns have been processed.

After this the result of the program should be displayed with the option to be saved to a (log.txt) file with an image of the detected anomalies.

## User stories

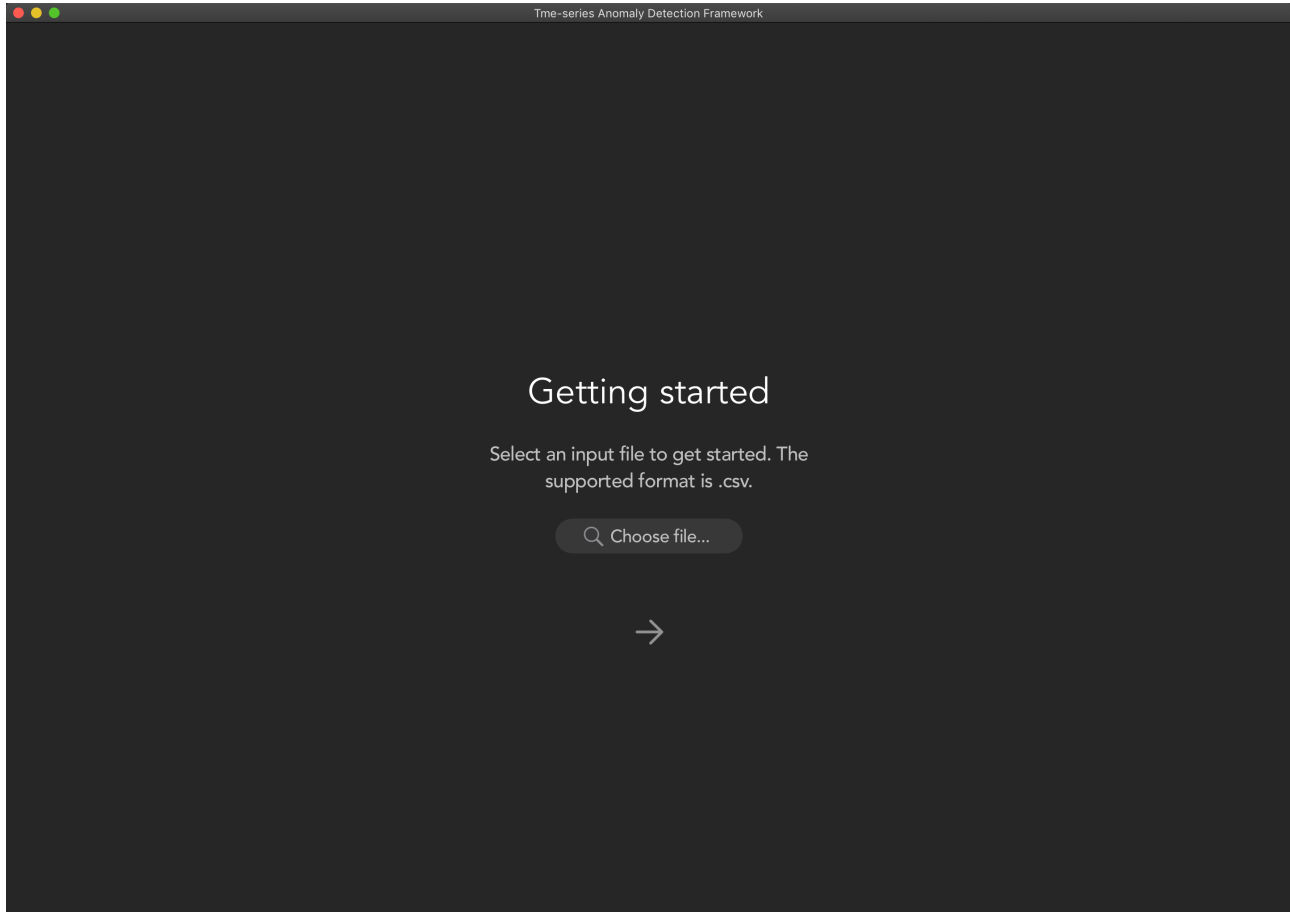
1. As a user I can upload a .csv file, so that I can run the program with input
2. As a user I can see if the input file format is OK, so that I can fix it if it is not
3. As a user I can see if there is a time-series frequency detected, so that I can fix the
4. As a user I can set the acceptable value range, so that I can filter out extreme outliers
5. As a user I can select a seasonal pattern so that the program can detect outliers based on this seasonality
6. As a user I can compare seasonalities to see what seasonality fits my input best. (optional)
7. As a user I can interactively determine the threshold so that the patterns can be reviewed individually. (optional)
8. As a user I can save a report with the detected anomalies shown in a scatterplot so that I can review the results.

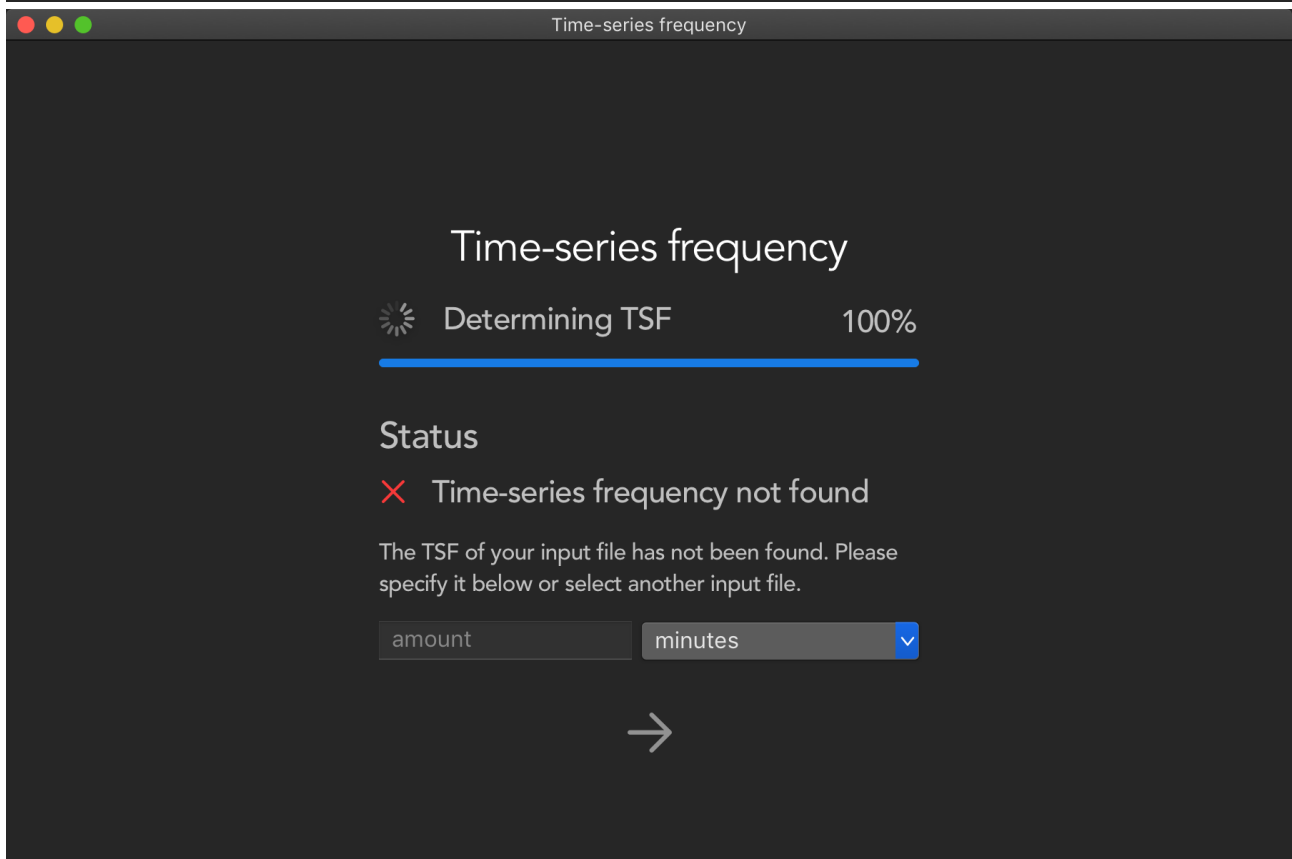
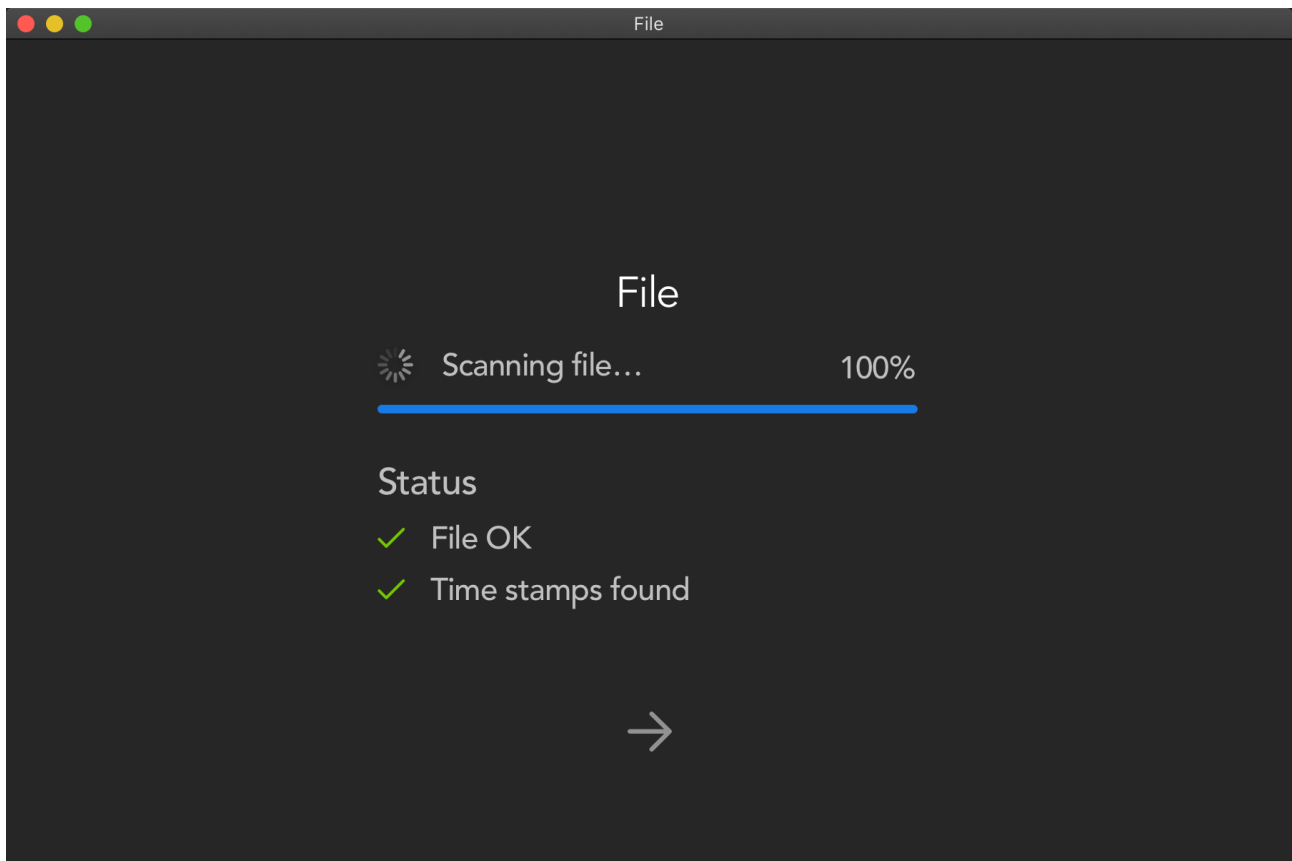
### Additional functionalities

- The program should backend check if the TSF can be derived automatically.

### Mock-up

The following images are potential designs of the user interface. Most of the functionalities discussed above are present. This design purposes as a guideline for design choices along the way.





Acceptable value range

Acceptable value range

Select an acceptable value range for the data points in your input file. Values out of this range will be ignored.

Values

Lower boundary:Upper boundary:

startend

→

Seasonality

Seasonality

Select the seasonality of the data in your input file below. Please also specify the amount of time units.

amountweeks

By continuing, the program will start detecting the anomalies in your data. Please wait. This may take a while.

→

