CAS Applied Data Science, University of Bern, S. Haug

**Assignment Work Instructions for Module 1**

The assignment work for Module 1 is a Conceptual Design Report for a Data Science project you would like to perform. Ideally during the CAS with some dataset you find interesting. Please use or orient yourself according to the template below. Remove this page for the submission version. Key information is the following.

**Language:** English or German

**Deadline:** To be defined in class

**Deliverables M1:** Conceptual Design Report, (this document), GitHub Repository, including also Jupyter Notebook and poster from Module 2.

**Expected effort and length:** About 30 hours, minimum 5 pages

**Further formal quality requirements:**

* All references to be listed in corresponding section and cited with number in text
* All tables and figures to have numbered legends with short explanations (tables above, figures below) and be referenced in text (Figure 1: blablan, Table 1: blblbl).
* Figures to be as self explanatory as possible, e.g. plots with at least axis labeling including units.

**Your names etc**

Your addresses

Your emails

**Data Science Project**

**Project Name**

**Conceptual Design Report**

**16th October 2020**

# Abstract

# 

# Table of Contents

[**Abstract**](#_buc6q0k08dmn) **1**

[**Table of Contents**](#_z0ssh6k3nrnr) **2**

[**1 Project Objectives**](#_5bx4zkckkkd) **3**

[**2 Methods**](#_q672ig652t9r) **3**

[**3 Data**](#_wauol8kayjkz) **3**

[**4 Metadata**](#_h4s0m7kg1q1p) **3**

[**5 Data Quality**](#_c5rpsdy8g2ak) **3**

[**6 Data Flow**](#_x5u0l8hx0kbh) **4**

[**7 Data Model**](#_mtk27zhjxb66) **4**

[**8 Risks**](#_vf6vykh0xvv7) **4**

[**9 Preliminary Studies**](#_r9xh3qn8v2wb) **4**

[**10 Conclusions**](#_4hg0yma181gd) **4**

[**Acknowledgements**](#_7gj90am4irap) **4**

[**Appendix X**](#_7djimjyxr32l) **4**

[**References and Bibliography**](#_g859w1rzj1iv) **5**

# 

# 1 Project Objectives

If we take the market valuation of a company to reflect public sentiment about the future of that company (its discounted expected future earnings) then it is reasonable to assume that the leading stock market indices for an economy, comprising the stock price movements of that country’s highest valued companies, are good indicators for the public sentiment about the future of that whole economy. With this project we try to explore the impact that the rise and fall in corona cases and deaths has had on the public sentiment of the worlds 10 biggest economies (based on their GDP in 2019 <https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?year_high_desc=true>). We therefore try to investigate whether there are indicators that the general public believes that the coronavirus has a long-lasting effect on economic prosperity.

We will use the following stock indices for each country:

|  |  |  |
| --- | --- | --- |
| **Country** | **Stock index** | **RIC** |
| United States of America | S&P 500 | ^SPX |
| China | SSE Composite | ^SSEC |
| Japan | Nikkei 225 | ^N225 |
| Germany | DAX | ^GDAXI |
| India | BSE Sensex | ^BSESN |
| United Kingdom | FTSE 100 | ^FTSE |
| France | CAC 40 | ^FCHI |
| Italy | FTSE MIB | ^FTMIB |
| Brazil | IBOVESPA | ^BVSP |
| Canada | S&P/TSX Composite | ^GSPTSE |

# Major world market indices: <https://www.investing.com/indices/major-indices>

# 2 Methods

0.5-1.0 page

Which infrastructure, tools, software libraries, statistical methods etc do you intend to use. It is clear that you may not know this at this stage, but try to make yourself some plan, even if it is going to change during the CAS.

# 3 Data

* Coronavirus data will come from the European Center for Disease Prevention and Control (<https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide>), for the space from 01.10.19-01.10.20.
* Financial markets data will come from yahoo finance: <https://finance.yahoo.com/>
* Data for leading economic indicators comes from the OECD: <https://www.oecd.org/sdd/fin-stats/>
* ETFs for sector/industry-analysis: <https://www.justetf.com/de/etf-lists.html>

# 4 Metadata

What metadata is required for reproducing your analysis?

Where do you store the metadata, how can people access it?

# 5 Data Quality

What are the quality requirements you have to reach your analysis goal (precision ...)?

Are they met? If not, do you expect a significant impact on your results,

Any measures to improve the quality?

# 6 Data Flow

Explain with a figure and words how the data flow of your project will be, from the data source to the final plots and numbers and.

# 7 Data Model

Draw and explain your data model at the conceptual level, the logical level and the physical level. .

Conceptual

Logical (with dataframes and with databases)

Physical (infrastructure needs)

# 8 Risks

What can go wrong?

When this and that goes wrong, what counter measures do you have?

What will be the impact on the quality of the aimed output, project time schedule, project cost ?

# 9 Preliminary Studies

Plots and numbers from Module 2.

# 10 Conclusions

….

# Acknowledgements

Acknowledge persons or institutions that helped you with the CDR here.

# Appendix X

If you have something to attach to your report, do it here.

# References and Bibliography

Please number any information source you used in the report with corresponding links here [1]:

[1] S. Haug et al., How to make a CDR, own brain, 2020 (put a weblink or DOI here)

[2]