I AM THE TITLE

ETL Process Plan

Lucerne, 01/03/2024

Group 8

Rodrigo Gonzalez Alonso rodrigo.gonzalez.alonso@gmail.com

Victor Anton victor.anton@stud.hslu.ch

Chris Imholz [chris.imholz@stud.hslu.ch](mailto:chris.imholz@stud.hslu.ch)

Professor

Ramón Christen ramon.christen@hslu.ch

Table of contents

[Introduction 2](#_Toc160200253)

[Architecture and procedure 2](#_Toc160200254)

[Extraction (Extract) 2](#_Toc160200255)

[Transformation (Transform) 3](#_Toc160200256)

[Laden (Load) 3](#_Toc160200257)

[Tools and technologies 3](#_Toc160200258)

[Summary 3](#_Toc160200259)

[Conclusion 3](#_Toc160200260)

[References 3](#_Toc160200261)

## Introduction

The topic / subject area you want to treat in the task

A brief explanation of your choice

The considered attributes you will use for connecting the 3 data sources and answering questions

And the research questions ...

1. main research question (I):
2. secondary research question (II):
3. secondary research question (III):

The questions should be verifiable, i.e. not as open as for example: Is there a country

GOAL:

*Development of a data anlysis by April 23, 2024 that enables institutional investors to effectively invest in Euronext listed shares (European companies from the AEX, BEL20, ISEQ Overall Index, PSI 20, CAC 40, FTSE MIB and OBX indices) based on robust ESG criteria.*

## Architecture and procedure

A complete and reasonable overview diagram (e.g. context diagram).

Ein Bild, das Text, Screenshot, Diagramm, Schrift enthält.

Automatisch generierte Beschreibung

### Extraction (Extract)

Identification and description of data sources.

Methods for data collection.

Challenges in data collection and how to overcome them.

### Transformation (Transform)

Description of the data cleansing process.

Description of the transformation steps to convert the data into a format suitable for analysis.

Application of methods for data enrichment and normalization.

### Laden (Load)

Description of how and where the transformed data is stored.

Selection of the data storage system and justification for this choice.

Security measures for data access and storage.

### Tools and technologies

List and description of the tools and technologies used in the ETL process.

Justification for the selection of specific tools.

## Summary

Summary of the key points of the ETL process.

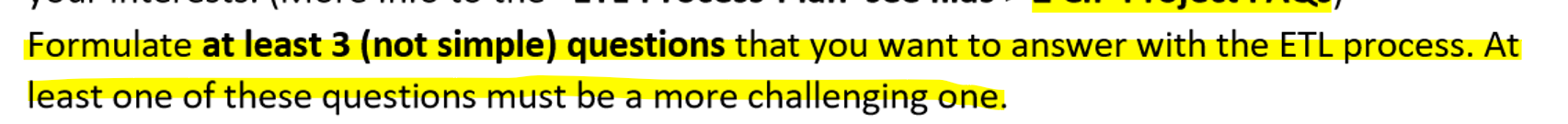
Outlook on the next steps after completion of the ETL process.

## Conclusion

## References

INFORMATION:

data set with a reasonable dimension: 20 observations (recordings) will be too less for serious data science applications. Choose a quantity that you can work with reasonably, also in terms of performance. However, a resulting data set of the ETL output of 20 observations might be reasonable for very specific topics.



Id = ISIN

Ein Bild, das Text, Schrift, Reihe, Screenshot enthält.

Automatisch generierte Beschreibung

Ein Bild, das Text, Schrift, weiß, Algebra enthält.

Automatisch generierte Beschreibung

Ein Bild, das Text, Schrift enthält.

Automatisch generierte Beschreibung

FAQ:

3.1 **PROCESS PLAN -** What belongs into the 'ETL process plan'?

**Answer:** The ETL process plan is a kind of a feasibility study of your project.  
You can define the complexity level by yourself. As lecturers, we will give you feedback on your plan with respect to the feasibility, complexity and possible risks. The process plan will be evaluated for the degree of details, transparency, comprehensibility, complexity and completeness.

* A table of contents and references
* The topic / subject area you want to treat in the task
* A brief explanation of your choice
* A complete and reasonable overview diagram (e.g. context diagram)
* Data sources that you will scrape
* The considered attributes you will use for connecting the 3 data sources and answering questions
* And the research questions ...  
  The questions should be verifiable, i.e. not as open as for example: Is there a country in the world that had fewer than 1,000 COVID-19 patients on 1 March 2021? You will NEVER collect all the data from all the countries in the world for the same point in time within the given time of the CIP project. Such an "open" (possibly also too broad) question is not "verifiable" for us lecturers.   
  Questions which are not verifiable and/or whose answers cannot be traced and validated are NOT suitable for our CIP project.