

Degree Programme and Code of the degree programme:

**Banking & Finance - 0229**

**Bachelor Thesis**

To obtain the academic degree:

**Bachelor of Arts in Business (BA)**

**Streamlit for Data Science for Economic Applications**

Subject: Banking

Submitted by: Diana Mirzaeva

Student ID: 1910229081

Supervisor: Alois Strobl

Providing body: University of Applied Sciences BFI Vienna

Wohlmutstraße 22

1020 Vienna

Vienna, 27.04.2022



**Statutory Declaration**

I hereby declare

that I have abided by the rules of academic work, in particular that I have authored the bachelor thesis independently and have not availed myself of any sources other than the ones specified in the appended bibliography. All passages taken from other publications word for word or in substance are identified as such. I further declare that I have not yet submitted this bachelor thesis as academic work in any form, neither in Austria nor abroad.

I am aware that also after positive assessment of the bachelor thesis, discovery of a violation of the rules of academic work (especially in case of plagiarism) will result in the initiation of proceedings to declare both the assessment as well as the academic degree granted null and void.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Place, Date |  | Author’s signature |



Non-disclosure of the bachelor thesis

Upon submitting an academic paper, the author is entitled to request that access to the submitted work be withheld for a maximum of five years from the date of submission. The request is to be granted by the programme director if the student shows probable cause that important legal or economic interests of the student are endangered.

The author requests non-disclosure until DD.MM.YYYY for the bachelor thesis at hand.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Place, Date |  | Author’s signature |

After consultation with the seminar supervisor, the requested non-disclosure of the bachelor thesis is granted by the programme director.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Place, Date |  | Programme director’s signature |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| UAS stamp |  |  |

Note: Any use of this bachelor thesis disregarding its non-disclosure notice is punishable under criminal law!



**Declaration of Consent**

With my signature I grant UAS BFI Vienna the campus-wide (UAS locations) right of use, in line with Copyright Act Sec. 18a, to make my bachelor thesis available on an intranet platform and to reproduce it for teaching and research purposes.

The author’s copyright and all other rights of use related to it shall of course remain with the author of the thesis (see § 10 Abs 1 UrhG).

I also consent to the UAS’ right to submit the thesis to an award event, e.g. the “Best Paper Award”, after they have consulted me and any co-authors of the thesis.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Place, Date |  | Author’s signature |

Table of Contents

[**List of Appendices V**](#_Toc101481937)

[List of Figures VI](#_Toc101481938)

[List of Tables VII](#_Toc101481939)

[List of Abbreviations VIII](#_Toc101481940)

[Abstract IX](#_Toc101481941)

[1 Introduction 1](#_Toc101481942)

[1.1 Research Question 1](#_Toc101481943)

[1.2 Methodology 1](#_Toc101481944)

[1.3 Structure 2](#_Toc101481945)

[2 Data Science 1](#_Toc101481946)

[2.1 Data. General Concept. 1](#_Toc101481947)

[2.2 Big Data 2](#_Toc101481948)

[2.3 Applications of Data Science 2](#_Toc101481949)

[2.3.1 Footnotes 2](#_Toc101481950)

[2.3.2 Figures and tables 2](#_Toc101481951)

[3 Python 3](#_Toc101481952)

[3.1 Introduction to Python 3](#_Toc101481953)

[3.2 Importance of Python in Economics and Finance 3](#_Toc101481954)

[3.3 Python libraries and packages 3](#_Toc101481955)

[4 Streamlit. 3](#_Toc101481956)

[4.1 Introduction to Streamlit 3](#_Toc101481957)

[4.2 Streamlit and Machine Learning 3](#_Toc101481958)

[5 Streamlit usage. Practical example. 3](#_Toc101481959)

[5.1 Data preprocessing. 3](#_Toc101481960)

[5.2 Data Analysis 3](#_Toc101481961)

[5.3 Results and STreamlit sharing. 3](#_Toc101481962)

[6 The Future of Data Science. ML and Streamlit Chances. 3](#_Toc101481963)

[7 Conclusion 3](#_Toc101481964)

[8 References 5](#_Toc101481965)

[Appendices 6](#_Toc101481966)

[Appendix A: Sample 6](#_Toc101481967)

[Appendix B: Sample 7](#_Toc101481968)

List of Appendices[[1]](#footnote-1)

**Number Caption Page**

Appendix  [A: Example 8](#_Toc52362733)

Appendix  [B: Example 9](#_Toc52362734)

List of Figures[[2]](#footnote-2)

**Number Caption Page**

[**Figure 1:** Catalogue of format templates 1](#_Toc52365462)

[**Figure 2:** Insert captions 3](#_Toc52365463)

[**Figure 3:** Add captions 4](#_Toc52365464)

List of Tables[[3]](#footnote-3)

**Number Caption Page**

[**Table 1:** Overview of format templates to be used 2](#_Toc52365493)

[**Table 2:** Sample table 4](#_Toc52365494)

List of Abbreviations[[4]](#footnote-4)

BSC Balanced Score Card

CSR Corporate Social Responsibility

MA Master of Arts in Business

POS Point of Sale

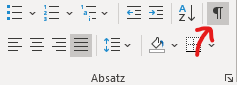
Abstract

(in English)

***Attention:*** *The section change after this paragraph should not be deleted. It allows for the change from Roman to Arabic page numbering****.***

**

*To make invisible control characters and thus section changes visible, press the key combination CTRL+SHIFT+PLUS or click on this character:*



# Introduction

We can't argue with the fact that data has become an integral part of our lives. People encounter data everywhere, in many fields. As organizations in all industries generate huge amounts of data, it is important to be able to analyze and interpret it in order to draw the right conclusions and make better decisions. Using this tool is an effective way to improve process efficiency, reduce costs, etc.

The data, datasets and its analysis are gaining its importance in our lives. It is extremely important for government, financial institutions, investors, traders and other organizations to be able to collect, analyze and interpret the data and to be prepared for demonstration and usage it for a problem solving. As the World is changing, there is a huge amount of new information every day. The value of information for companies is growing. Those who own data are better at analyzing and solving business problems, predicting market trends and understanding customer needs.

By working with different programming language and frameworks, people can analyze the data faster and also demonstrate it in easier way for a human. That is why in this paper we are going to cover Streamlit and its application for data science and economics.

## Research Question

This leads me to my research question: How can Streamlit be used for Data Science?

## Methodology

In this bachelor thesis, we would like to use both qualitative and quantitative methods as research support. The analysis of this topic will be carried out by using qualified literature, news articles and market information to see what is Streamlit and its applications, what are benefits of its usage.

This bachelor thesis starts with an introduction to data science and big data in general. Continuing the theoretical part, the author will provide an overview of its usage.

Following the theoretical introduction to data science, some information about Python and its libraries will be given. After that we will discuss different frameworks and we will focus ourselves on Streamlit. The real-world example will be given in order to give a full picture of Streamlit application to datasets. For that we are going to use “Economic Indicators” datasets coming from the different sources. After the relevant data is preprocessed, it will be “transferred” to Streamlit. After that we will be able to easily interpret our data and see the pros and cons of Streamlit.

## Structure

After the introduction, the second chapter will provide an overview of data science and big data and its impact on our lives. Some information on Python and its libraries will be presented and discussed in the third chapter. The fourth chapter will give an overview to Streamlit and its possible application to finance. In the fifth chapter well present the Streamlit example with chosen dataset and its analysis. At the end the future of data analytics will be discussed with a regard to artificial intelligence. Conclusion chapter serves as a summary. We will summarize the work.

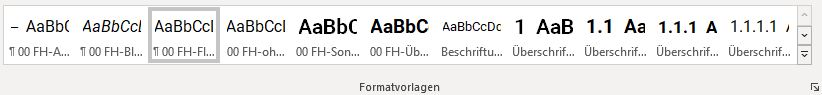
# Data Science

For the formatting of your thesis, the manual provides plenty of information in chapter 6.[[5]](#footnote-5)

## Data. General Concept.

The use of format templates is essential for uniform formatting and the direct transfer of headings to the lists. For detailed information, please refer to the manual[[6]](#footnote-6).

The most important templates for the design of your thesis are contained in the format template catalogue and labelled “00 FH-…“.



**Figure 1:** Format template catalogue

Here you can find a list of format templates:

**Table 1:** Overview of the format templates to be used

For the running text, please use the format template "00 FH-Fließtext" (Roboto, 12pt, justified, 1.5 line spacing). Avoid blank lines before and after a paragraph - these are already included in the format template.

Sample running text. Sample running text. Sample running text. Sample running text. Sample running text. Sample running text. Sample running text. Sample running text. Sample running text. Sample running text. Sample running text. Sample running text. Sample running text.

## Big Data

* Recommendations on the design of headings and levels: Start your thesis with chapter “1 Introduction”.
* Use format templates “Heading 1” to “Heading 5” to structure the chapters and subchapters.

Attention: Only by using the format templates will the headings be automatically transferred to the table of contents!

* Close your thesis with a chapter entitled “Conclusion”[[7]](#footnote-7).

Recommendation: In the tab "View", set a check mark at "Display" for "Navigation area". This will display the structure of the document on the left side of the screen and you can quickly switch to the individual chapters. This can be particularly helpful in long documents.

## Applications of Data Science

### Footnotes

In the footnotes, you will find a general note on footnotes.[[8]](#footnote-8)

### Figures and tables

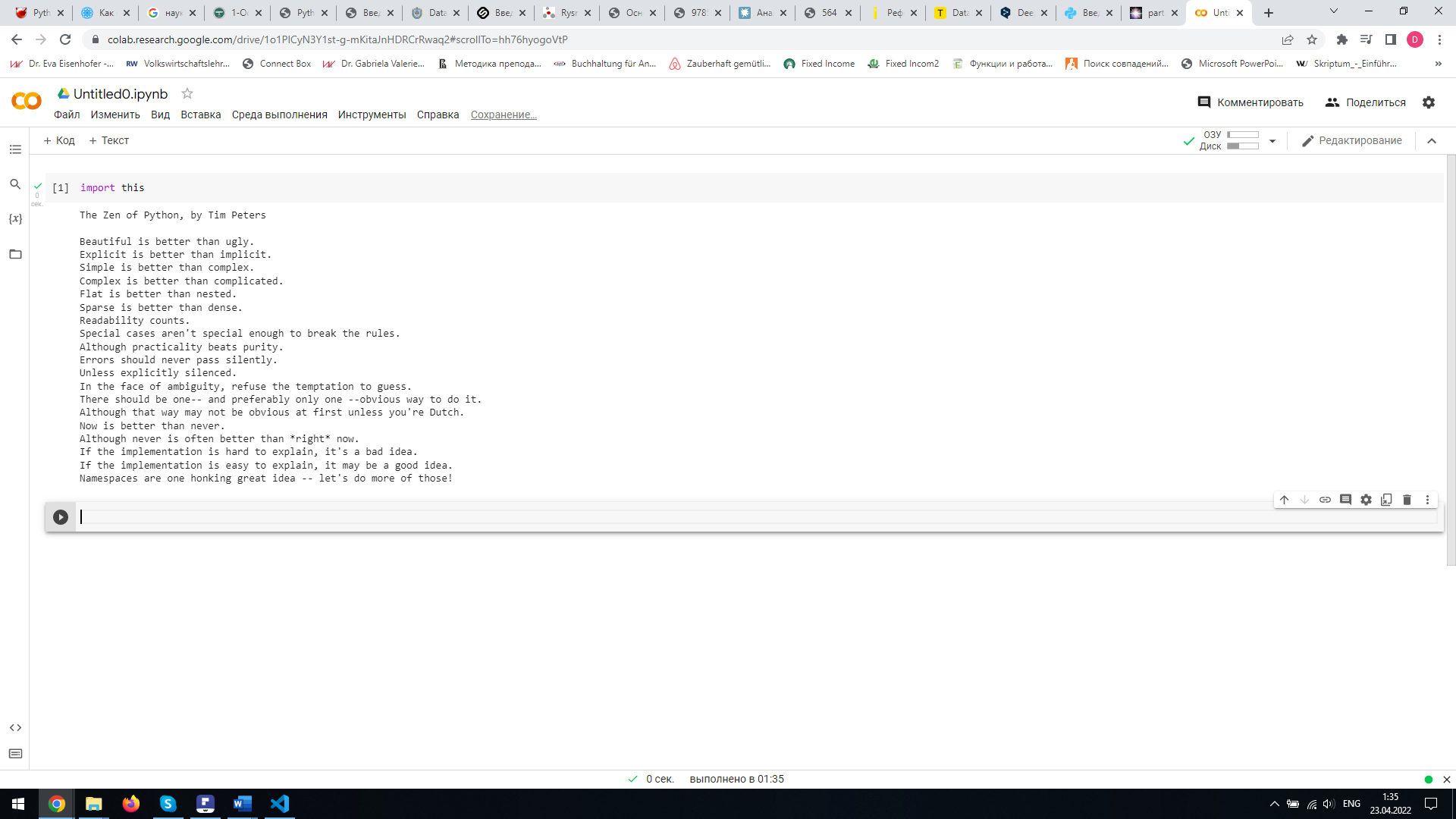
To add captions to a figure or table and thus enable automatic transfer to the

# Python

Nowadays there is a huge number of different programming languages both similar to each other and different from each other. All languages have some orientation, i.e. are able to solve only certain problems. Therefore, most often they are classified according to the type of tasks to be solved.

Python today can be confidently called the best programming language. Python has been pushing other programming languages to the limit in recent years and is gaining record popularity around the world. Python has a large community of people around the world who speak it, so if you have a problem or make a mistake, you can ask for help. It is almost a combination of incongruities: Python is suitable for web development, big data analysis, and machine learning, but it is also intuitive, easy, and beginner-friendly.

## Introduction to Python

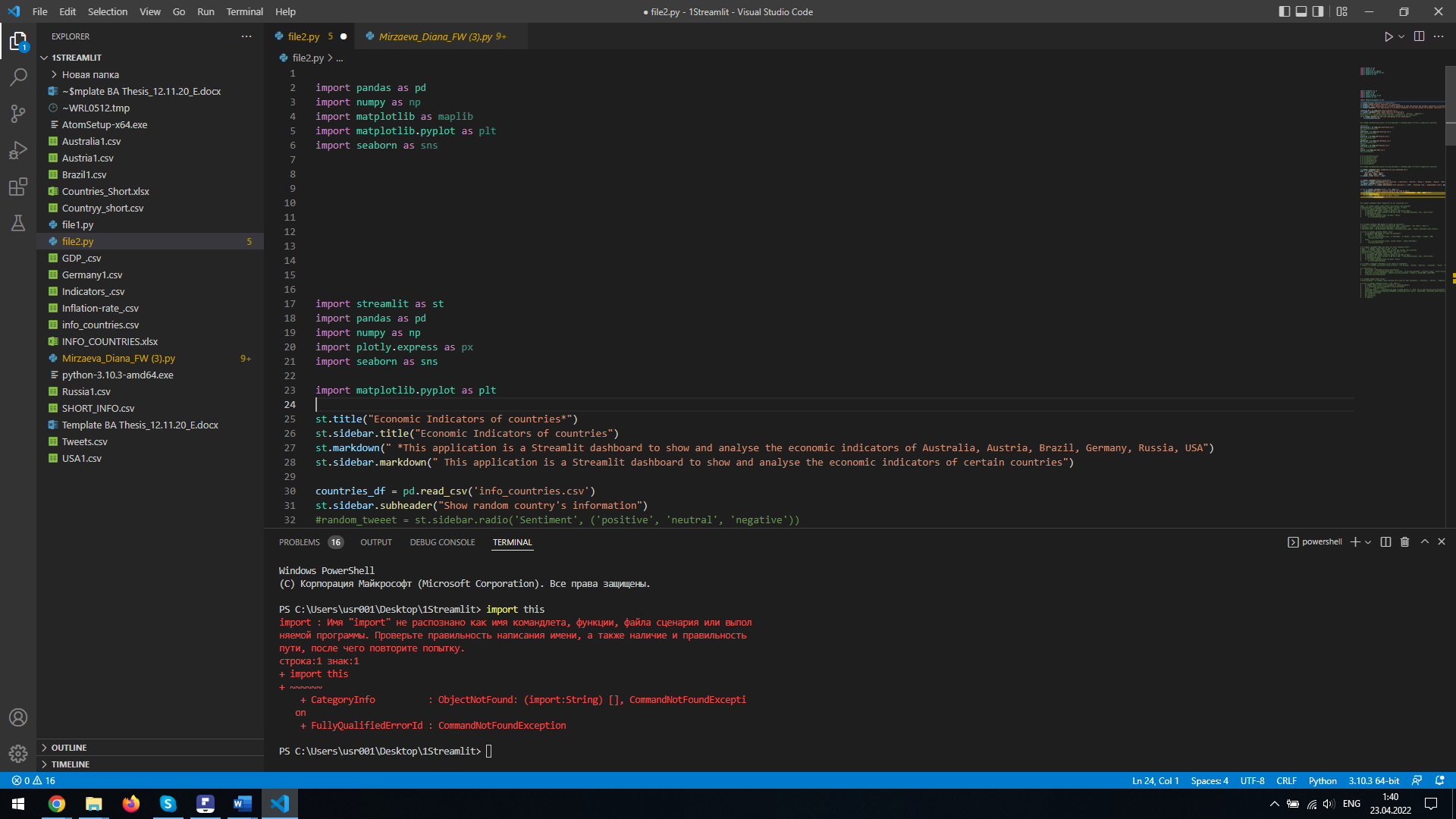
Python is a high-level general-purpose programming language with a large standard library containing many useful features and supporting several programming paradigms. Python developers adhere to a particular philosophy authored by Tim Peters. Its contents can be easily seen by running the interpreter and entering the command "import this" into it: 

In short, the developers advocate beauty, explicitness and simplicity. This philosophy also contains many important rules, such as: "Mistakes should never be glossed over." "If an implementation is hard to explain, you have a bad idea." "If the implementation is easy to explain, you probably have a good idea."

This clear but simple structuring of the code is the reason for its popularity. This code is easy to understand both for the novice programmer as well as for the expert who, for example, has changed the programming language.

## Python libraries and modules

Python is not only a programming language, but also an entire ecosystem of libraries and tools. These need to be imported as needed (such as the diagramming library) or run as a separate

The import operation makes the package available in an easily accessible way. The import operation makes the package available in the current namespace and in the current interpreter process.

The base Python distribution already includes a large set of packages and modules that extend the interpreter's functionality. General mathematical calculations do not require any modules, while specialized mathematical functions are imported through special module.

Pandas

NumPy based library for processing and Time series and table data analysis. Tightly integrated with matplotlib (data visualization) and PyTables (storage and loading of data).

NumPy

Library providing support for multidimensional arrays designed to store homogeneous and heterogeneous data.

Includes optimized functions/methods for working with such arrays

Matplotlib

One of the most popular packages for data visualization and construction both 2D and 3D diagrams in Python.

Seaborn

Scikit-learn

A popular machine learning package containing unified implementations of many algorithms, such as classification and clustering.

There are various Python environments like Jupyter notebook, GoogleColaboratory, Kaggle Kernel, Atom. Visual Studio Code was implemented in the practical exercise of this paper.

<http://www.williamspublishing.com/PDF/978-5-907203-03-7/part.pdf>

## Importance of Python in Economics and Finance

1. Введение в программирование. Основные концепции. Самые частые ошибки и работа с ними.
2. Основные библиотеки: numpy, matplotlib, sympy, pandas.
3. Python в работе с рыночными рисками: расчет волатильности, доходности, построение линейных моделей, оценка рисков методами VaR и drawdown.
4. Python в инвестиционной деятельности: построение инвестиционных портфелей, CPPI, прогнозирование доходности акций.
5. Python в корпоративных финансах: моделирование NPV методом Monte Carlo, симуляционные модели.
6. Работа с альтернативными источниками данных: анализ текстовой информаци

Advantages of using Python instead of R and Microsoft Excel

# Streamlit.

Dashboard Every facet on its own only shows a part of the data. It is when the facets connect, through a comprehensive visualization, a holistic picture of the dataset is revealed. This was the primary reason for choosing the dashboard as the final visualization. A secondary reason was also the interactivity that the dashboard offers. 36 The dashboard was made using a python library called Streamlit. The library allows for making easy interactive dashboards and more importantly allows seamless hosting capabilities using Github. Another advantage of Streamlit is the supported mapping library, pydeck.gl. Pydeck.gl is a python library based on the Javascript deck.gl library which makes creating high level spatial visualizations simple. There were two separate dataframes created out of the original raw data. One was used for the Altair (VanderPlas et al. 2018) typicality chart under the Topical subheader. The other was modified and used for all other visualizations in the entire dashboard. There were also three images used in the dashboard: the image of refugees fleeing Moria and the two timelines. The goal was to make the dashboard into a narrative visualization while also allowing for data exploration and the fullest interactivity. Incorporating elements of narrative visualization and allowing for interactivity were achieved using various methods within the Streamlit library. Figure 16 shows the principles of abstraction and elaboration implemented using the expanders. Figure 17 shows the inclusion of information from the topical and temporal facet within the spatial facet using tooltips. Figure 18 illustrates the use of selection boxes to filter the data and display results according to the user’s choice.

## Introduction to Streamlit

Streamlit is a tool to build interactive web applications around data and machine learning models, with Python. Like Jupyter Notebooks, Python scripts for data modeling are the only scripts required to construct a user interface; the Streamlit framework builds an interactive and user friendly interface around it. In contrast to Jupyter Notebooks, which is a web-based interactive computing program, Streamlit builds web applications using its own HTML, CSS and JavaScript, by providing a Python API. Where Jupyter notebook requires users to run blocks of Python code, Streamlit hides the script in the background and presents an interactive interface to the user. An alternative that was considered for the ITM web application, with a Python model in the back-end, was the Flask6 framework. Deploying Flask apps requires writing HTML, CSS and JavaScript, to create a user interface on top of the data model in Python. Although Flask allows more design freedom than Streamlit, it requires a longer design, development and deploy process. Streamlit offers high performance while maintaining the flexibility of rapid prototyping. Using Streamlit’s API was thus preferred for the iterative design and development process that was maintained in this project. Figure E1 shows Streamlit’s workflow. Every time the user interacts with the interface, the Streamlit Python script is run from top to bottom to update the UI. To make the application performant, only a subset of the pipeline is recomputed to display the change in the user interface, while the rest is taken from cache. All but two visualizations are directly shown when the user interface is rendered. Two interactive visualizations, presenting relationships between topics and documents in 5https://www.streamlit.io/ 6https://flask.palletsprojects.com/en/1.1.x/ 41 a two-dimensional space, are only generated on explicit user request. Generating these two visualizations requires a dimensionality reduction computation, which is kept outside of the default UI rendering process for performance reasons. All visualizations, except for these two figures, are made using Plotly7 , a Python open source graphing library. In contrast to the widely-used Matplotlib8 library, Plotly is browser-aware. Instead of creating non-interactive charts that Streamlit would display as static images, Plotly does not require Streamlit to re-render the entire interface for a interaction with a figure. For performance reasons, all visualizations are created using Plotly, which sends interactive charts to the browser. Although Streamlit (and additional visualization libraries), comes with a clean and user friendly interface design, some customization of the interface was performed. First, the default size of the left sidebar is quite narrow. Because of the design choice to put all user input in this sidebar, this sidebar is widened by manipulating the CSS file. In addition, the main color in the figures is set to turquoise (#00b4c8), which is the main color of the brand Yle9 . Although this was not a direct requirement by the stakeholders, the organization requires to use the Yle colour palette in corporate-level publications, presentations and graphics. End-users of the ITM applications might use PNG exports of the figures generated in the ITM application directly in data reports and presentations representing the company, so Yle’s colour palette is used as much as possible following this indirect requirement.

## Streamlit and Machine Learning

# Streamlit usage. Practical example.

## Data preprocessing.

## Data Analysis

## Results and STreamlit sharing.

# The Future of Data Science. ML and Streamlit Chances.

# Conclusion

How do I write a conclusion?[[9]](#footnote-9)

References

Haslehner, F. / Wala, T. / Springler, E. (2014): Bachelor- und Masterarbeiten an Fachhochschulen. Eine Kurzanleitung zur Erstellung wissenschaftlicher Arbeiten, 2. Aufl., Wien: LexisNexis

Nachbagauer, A. (2016): Jenseits der Normalarbeit. Prolegomenon zur atypischen Arbeit aus soziologischer Sicht, in: Pacic, H. (Hrsg.): Atypische Beschäftigung. Rechtsfragen jenseits der Normalarbeit, Wien: Verlag Österreich, S. 1–36

Paier, D. (2017): Was kann die hochschuldidaktische Forschung der Lehrpraxis sagen? Aktuelle Entwicklungen der hochschuldidaktischen Forschung, in: Wirtschaft und Management, Band 25, S. 7–25

Appendices

Appendix A: Sample

For example, an interview guide can be inserted here.

Appendix B: Sample

For example, a transcript can be inserted here.

1. If your paper does not include appendices, delete this list. [↑](#footnote-ref-1)
2. If your paper does not include figures, delete this list. [↑](#footnote-ref-2)
3. If your paper does not include tables, delete this list. [↑](#footnote-ref-3)
4. If your paper does not include abbreviations, delete this list. [↑](#footnote-ref-4)
5. Haslehner et al (2014) pp. 107 [↑](#footnote-ref-5)
6. Haslehner et al (2014) pp. 122 [↑](#footnote-ref-6)
7. Instructions for writing this final section can be found in Haslehner et al (2014) p. 35. [↑](#footnote-ref-7)
8. For a correct paragraph alignment of footnotes, it is crucial that you set a tab at the beginning of the footnote text. This means that the left paragraph alignment is also correct for footnotes with several lines. [↑](#footnote-ref-8)
9. Instructions for writing this final section can be found in Haslehner et al (2014) on page 35. [↑](#footnote-ref-9)