



Midjourney: coding wizard, matrix style, extreme wide portrait, black background, hacker --quality 0.25 --ar 16:9
Vary: coding wizard, matrix style, extreme wide portrait, black background, Psychedelic art --ar 16:9
Vary: Coding Wizard, extreme wide portrait, psychedelic art, black background, intense colours, cinematic lighting --ar 16:9
Zoom Out: desktop screen, black, cinematic lighting --ar 16:9
Vary region: desktop screen, data analysis, data points intensive colour, cinematic lighting --ar 16:9
Zoom Out: psychedelic digital intensive colour, cinematic lighting --ar 16:9 --zoom 2

Angewandte Generative KI in den (digitalen) Geisteswissenschaften

**Generieren.
Transformieren.
Analysieren. Visualisieren.
ChatGPT-4 Advanced
Data Analysis**

<http://chpollin.github.io/GM-DH>

Christopher Pollin

<https://chpollin.github.io>

Zentrum für Informationsmodellierung, Graz

<https://informationsmodellierung.uni-graz.at/de/>

Digital Humanities Craft OG

www.dhcraft.org

ChatGPT-4 Advanced Data Analysis is like ...

[...] the systems solve some problems that even capable students would find challenging, whereas they fail on some problems that even middling high school students would find easy.

Davis, Ernest, und Scott Aaronson. „Testing GPT-4 with Wolfram Alpha and Code Interpreter plug-ins on math and science problems“. arXiv, 14. August 2023. <https://doi.org/10.48550/arXiv.2308.05713>.

Superintelligente*r Praktikant*in, der/die Python coden kann.

Extrem motiviert - stürzt sich auf jeden Task.

Macht jedoch immer wieder Fehler und man muss alles überprüfen.

Hat eine kurze Aufmerksamkeitsspanne.

Die IT-Abteilung hat überhaupt gar kein Vertrauen, weswegen es keinen Internetzugang und nur einen Sandbox-Arbeitsplatz gibt.

... und es ist natürlich eine Maschine ...

ChatGPT-4 Advanced Data Analysis (=ADA)

(formerly known as Code Interpreter)

⚡ GPT-3.5

🤖 GPT-4

Our most capable model, great for tasks that require creativity and advanced reasoning.

Available exclusively to Plus users

GPT-4 currently has a cap of 50 messages every 3 hours.

Default

🤖 Code Interpreter Beta

Plugins Beta



“an experimental ChatGPT model that can use Python, handle uploads and downloads” <https://openai.com/blog/chatgpt-plugins>

chatGPT-4 Model + Python interpreter

Upload von bis zu 10 Files. File Download möglich

Unterstützte Formate:

- **Text** (.txt, .csv, .json, .xml etc.)
- **Image** (.jpg, .png, .gif etc.)
- **Document** (.pdf, .docx, .xlsx, .pptx etc.)
- **Code** (.py, .js, .html, .css etc.)
- **Data** (.csv, .xlsx, .tsv, .json etc.)
- **Audio** (.mp3, .wav etc.)
- **Video** (.mp4, .avi, .mov etc.)

Was kann ADA ?

Data Reading (Powerpoints, PDF, am besten strukturierte Daten wie CSV, ...)

Generieren (und bearbeiten)

Code generieren, Prototypen, “alle Bilder in einem .zip drehen”, ...

Transformieren

Data Cleaning, (un)strukturierte Daten → strukturierte Daten, XML, RDF, CSV, ...

Analysieren

Exploring Datasets, Statistische Auswertung, “über Daten und Ergebnisse reden”

Visualisieren

Data Visualization, Generating on-the-fly Charts and Maps

Wer macht was? GPT-4 vs. Python

Table 1: Tasks Best Suited for GPT-4

Task	Best Tool	Reason	Relevant Media Type
Reading Text	GPT-4	GPT-4 is proficient at understanding and interpreting text.	Text [txt, pdf, word]
Writing Text	GPT-4	GPT-4 excels at generating coherent, contextually appropriate text.	Text [txt, pdf, word]
Outlining Text	GPT-4	For extracting and summarizing the main points from the text.	Text [txt, pdf, word]
Translating Languages	GPT-4	GPT-4 is capable of translating text between different languages accurately.	Text [txt, pdf, word]
Simulating Conversations	GPT-4	GPT-4 can simulate dialogue and conversation between multiple entities.	Text [txt, pdf, word]
Answering Questions about Text	GPT-4	Able to comprehend and respond to inquiries based on provided text.	Text [txt, pdf, word]
Summarizing Text	GPT-4	GPT-4 can succinctly summarize large volumes of text.	Text [txt, pdf, word]
Drafting Emails	GPT-4	Can draft coherent and contextually relevant emails.	Text [txt, pdf, word]

Table 2: Tasks Best Suited for Python

Task	Best Tool	Reason	Relevant Media Type
Operating on Files	Python	Python can create, rename, delete, and manipulate files.	All Types
Numerical Computation	Python	Python excels at performing mathematical and numerical calculations.	Structured Data [csv, json, xml, excel]
Image Processing	Python	Python, with various libraries, can manipulate and process images.	All Types
Video Processing	Python	Python can manipulate and process video files.	All Types
Audio Processing	Python	Python is suitable for manipulating and processing audio files.	All Types
System Scripting	Python	Python can interact with the system for scripting tasks like moving files, creating directories.	All Types
Graphing & Visualization	Python	Python can create graphs and visual representations of data.	Structured Data [csv, json, xml, excel]
Data Analysis	Python	Python is proficient at analyzing structured and unstructured data.	Structured Data [csv, json, xml, excel]

Ressourcen

- **Open Interpreter.** An open-source, locally running implementation of OpenAI's Code Interpreter.
<https://github.com/KillianLucas/open-interpreter>
- **Open Interpreter**  ChatGPT Code Interpreter You Can Run LOCALLY!.
https://www.youtube.com/watch?v=SqnXUHwla3c&ab_channel=MatthewBerman
- **AI Explained.** 12 New Code Interpreter Uses (Image to 3D, Book Scans, Multiple Datasets, Error Analysis ...). https://www.youtube.com/watch?v=njf22xx8BQ&ab_channel=AIExplained
- **AI Explained.** GPT 4 Got Upgraded - Code Interpreter (ft. Image Editing, MP4s, 3D Plots, Data Analytics and more!).
https://www.youtube.com/watch?v=O8GUH0_htRM&t=311s&ab_channel=AIExplained
- **Jules White.** ChatGPT Advanced Data Analysis. Coursera.
<https://www.coursera.org/learn/chatgpt-advanced-data-analysis/lecture/sWBTrm/breaking-the-ice-how-to-get-started>

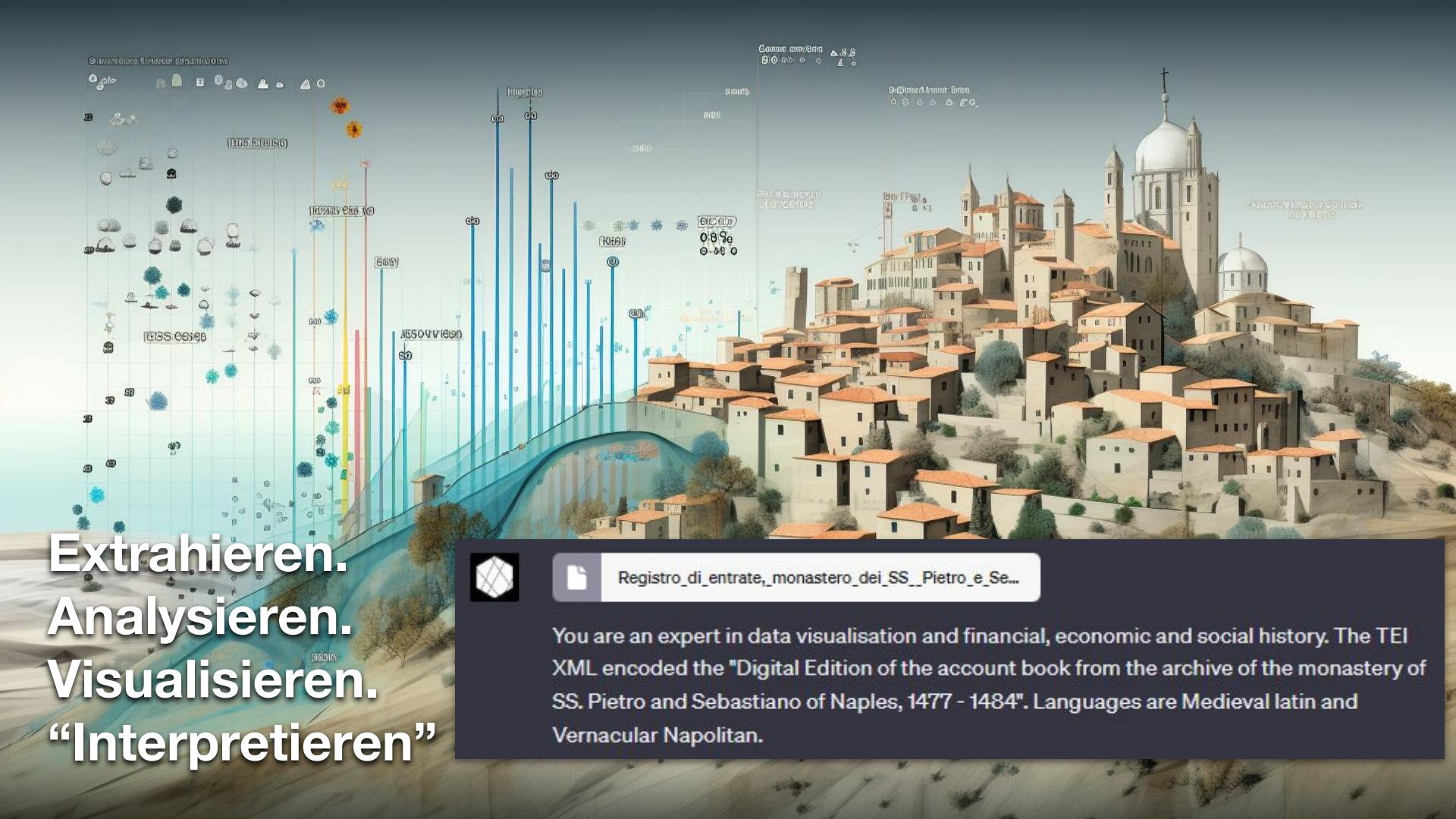
**Extrahieren.
Analysieren.
Visualisieren.
“Interpretieren”**



Registro_di_entrata,_monastero_dei_SS._Pietro_e_Se...

You are an expert in data visualisation and financial, economic and social history. The TEI XML encoded the "Digital Edition of the account book from the archive of the monastery of SS. Pietro and Sebastiano of Naples, 1477 - 1484". Languages are Medieval latin and Vernacular Napolitan.

**Extrahieren.
Analysieren.
Visualisieren.
“Interpretieren”**



TEI XML Rechnungsbuch → Datenvisualisierung 1/4

You are an expert in data visualisation and financial, economic and social history. The TEI XML encoded the "Digital Edition of the account book from the archive of the monastery of SS. Pietro and Sebastiano of Naples, 1477 - 1484". Languages are Medieval latin and Vernacular Napolitan.

You will use data visualisation to answer the following research questions:

- * Reconstruct the history of the landscape through the cultures.
- * To what extent was the monastery involved in the logic of a "regional" market, and what was the importance of wine sales and production?
- * To what extent can the monastery be understood as a microcosm in which a number of people had privileged relations with the monastery?

TEI XML Rechnungsbuch → Datenvisualisierung 2/4

Semantic of @ana:

- * bk:entry: Identifies a transaction entry.
- * bk:when: This is the date of the transaction.
- * bk:from: This identifies the economic entity from which the economic assets flow.* bk:to:
This indicates the other economic subject (the receiver) in the transaction to which
economic assets flow.
- * bk:money: This represents the monetary value involved in the transaction.
- * bk:commodity: This indicates the good or commodity involved in the transaction.
- * bk:status: This is the status of the transaction (paid, partly paid, not paid, etc.).
- * bk:debit | bk:credit: This defines the debit or credit entry of a transaction.

Visualizing Monastery's Historical Transactions. September 7, 2023. . ChatGPT August 3 Version.

<https://chat.openai.com/share/01b19be6-c1ad-4f92-868c-f4edc9882d8c>

TEI XML Rechnungsbuch → Datenvisualisierung 3/4

Information (index for person, organisation and places, classification of commodities, units) can be found in standOff listPerson, listOrg, listPlace, unitDecl, classDecl/taxonomy

Here is a TEI XML snippet:

```
...
<p facs="#facsc_11_TextRegion_1678960405191_649" ana="bk:entry">
    <date ana="bk:when" when-custom="13 indizione" when="1480-01-31">Die ultimo mensis ianuarii, XIII
    indictione,</date>
    avemo avuto da <persName ana="bk:from" ref="#p7">Fantino</persName> per <term ana="commodity">foglie</term>
    che ave venduto dall'orto delo nostro
    monasterio consignata a la rota per
    mano de <persName ana="bk:by" ref="#p6">dopno Iacobo de Larino</persName><measureGrp><measure ana="bk:money"
    quantity="18" unitRef="#tr.">tr.18,</measure><measure ana="bk:money" quantity="10" unitRef="#gr.">gr.10</measure></measureGrp>
    Et per <term ana="commodity">foglia</term> che ave avute lo
    dicto <orgName ana="bk:to" ref="#o1">monastero</orgName><measureGrp><measure ana="bk:money" quantity="2"
    unitRef="#tr.">tr.2</measure>,<measure ana="bk:money" quantity="13" unitRef="#gr.">gr.13</measure></measureGrp>
</p>
...

```

TEI XML Rechnungsbuch → Datenvisualisierung 4/4

Carry out the task as follows:

- * First analyse the TEI XML. Focus only on the <body> and extract the encoded transactions in the @ana semantic annotation.
- * Take the text of the historical source (text nodes in the TEI XML) and create user stories in relation to the research questions.
- * Discuss data visualisations based on the user stories that are useful for answering the research questions.
- * Implement the data visualisation using all the transactions and their associated phenomena and named entities.



Alright, let's proceed step by step:

1. **Analyse the TEI XML:** I'll parse the XML file, focusing on the `<body>` section and extracting the encoded transactions based on the `@ana` semantic annotation.
2. **Create User Stories:** Based on the extracted text of the historical source, I'll generate user stories in relation to the provided research questions.
3. **Discuss Data Visualisations:** I'll suggest potential data visualizations that can help in answering the research questions.
4. **Implement Data Visualisation:** I'll then attempt to implement the discussed visualizations using the extracted transactional data and associated phenomena.

Let's begin with the first step.

Überführung der Forschungsfragen (1., 2., 3.) in User Stories

Visualizing Monastery's Historical Transactions. September 7, 2023. . ChatGPT August 3 Version.

<https://chat.openai.com/share/01b19be6-c1ad-4f92-868c-f4edc9882d8c>

Step 2: Create User Stories

Based on the research questions and the extracted data, we can create the following user stories:

1. **Reconstruct the history of the landscape through the cultures.**
 - As a researcher, I want to visualize the variety and frequency of commodities (like wine, grains, fruits) being traded over time to understand the agricultural landscape.
 - As a researcher, I want to see the geographical distribution of trading places, organizations, and individuals to understand regional agricultural hubs.
2. **To what extent was the monastery involved in the logic of a "regional" market, and what was the importance of wine sales and production?**
 - As a researcher, I want to visualize the frequency of transactions involving the monastery as a "to" or "from" entity to gauge its involvement in the regional market.
 - As a researcher, I want to see a breakdown of commodities traded by the monastery, with a focus on wine, to understand its importance in wine sales and production.
3. **To what extent can the monastery be understood as a microcosm in which a number of people had privileged relations with the monastery?**
 - As a researcher, I want to visualize the most frequent individuals and organizations that had transactions with the monastery to understand privileged relations.
 - As a researcher, I want to see the status (paid, partly paid, not paid) of transactions involving these frequent entities to gauge the nature of their relationship with the monastery.

Auswahl potenzieller Visualisierungsarten in Abhangigkeit der User Stories

Step 3: Discuss Data Visualizations

Based on the user stories, the following data visualizations can be useful:

1. **Time Series Chart:** Plotting commodities over time will give a sense of the changing agricultural landscape.
2. **Geographical Map:** Plotting the places where transactions are happening can help understand regional agricultural hubs.
3. **Pie or Bar Chart:** Breaking down commodities traded by the monastery, especially focusing on wine.
4. **Histogram:** Displaying the frequency of individuals and organizations transacting with the monastery.
5. **Stacked Bar Chart:** Displaying transaction status (paid, partly paid, not paid) for frequent entities to understand their relationship dynamics with the monastery.

Now, let's move to the fourth step:

Visualizing Monastery's Historical Transactions. September 7, 2023. . ChatGPT August 3 Version.

<https://chat.openai.com/share/01b19be6-c1ad-4f92-868c-f4edc9882d8c>

Implementieren

Visualizing Monastery's Historical Transactions. September 7, 2023. . ChatGPT August 3 Version.

<https://chat.openai.com/share/01b19be6-c1ad-4f92-868c-f4edc9882d8c>

Step 4: Implement Data Visualisation

To implement the data visualizations, I'll first extract the relevant transactional data, including date, commodity, monetary values, entities involved, and transaction status.

Finished working

Show work ▾

The extracted transaction data includes relevant fields like date, entities involved (from and to), commodity involved, monetary values, and the status of the transaction. Here are the first few transactions for verification:

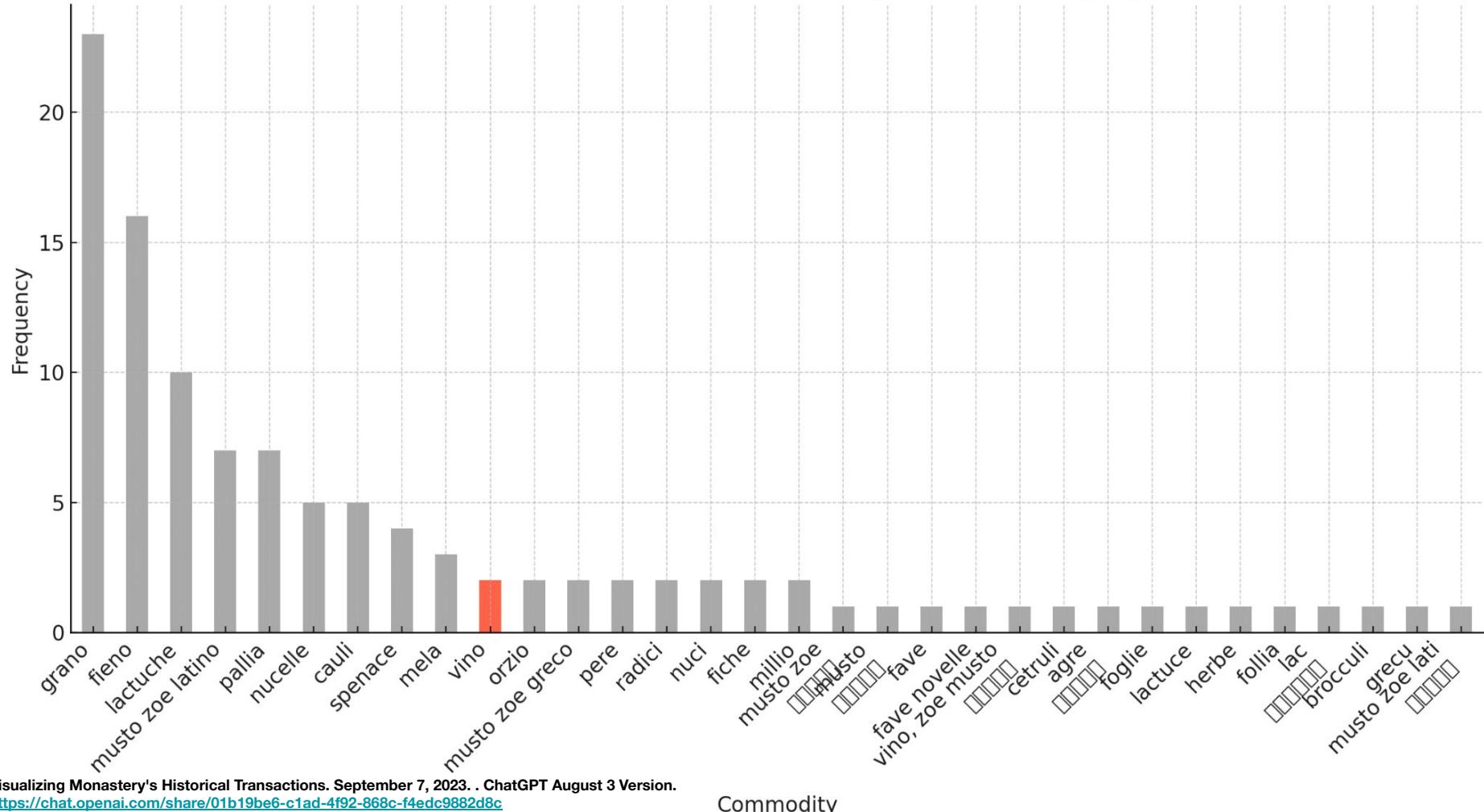
1.

- Date: 1477-06-09
- From: orto
- To: monastero
- Commodity: fave novelle
- Money: 6 tr., 10 gr.

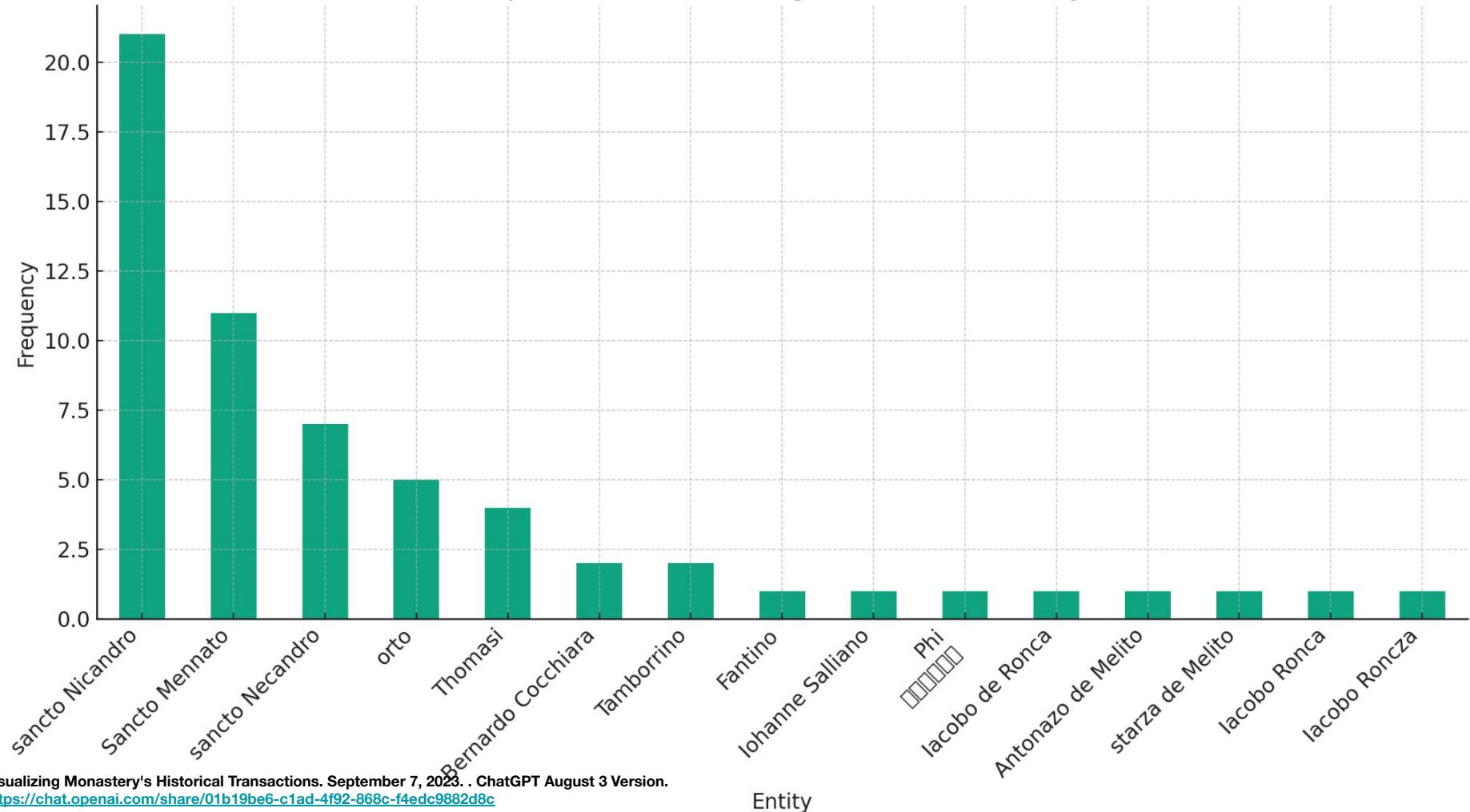
... and so on.

With this data structure in place, we can now proceed to implement the visualizations.

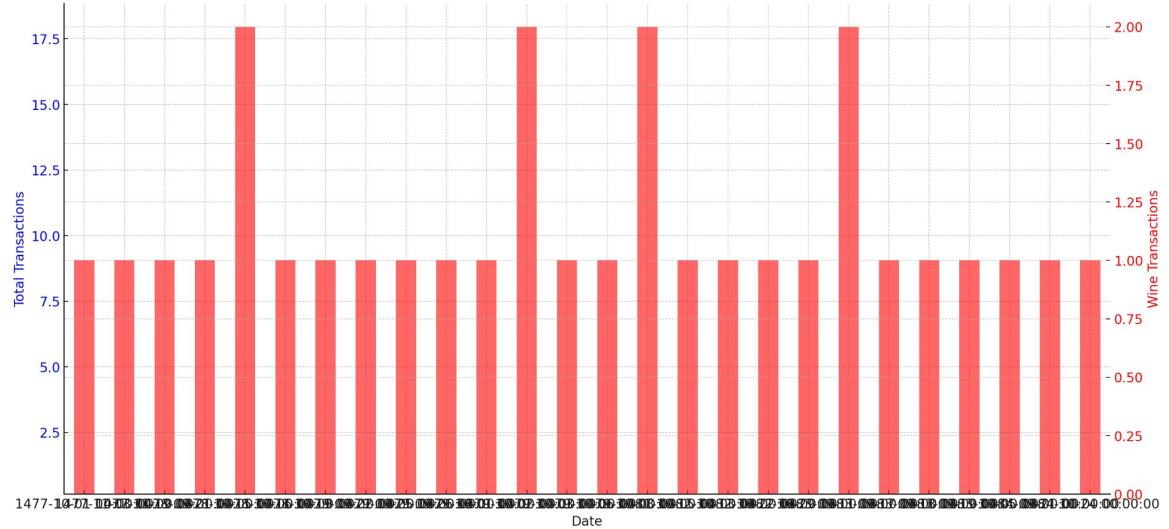
Commodities Traded by the Monastery with Wine Highlighted



Top Entities Transacting with the Monastery



Total and Wine Transactions Over Time

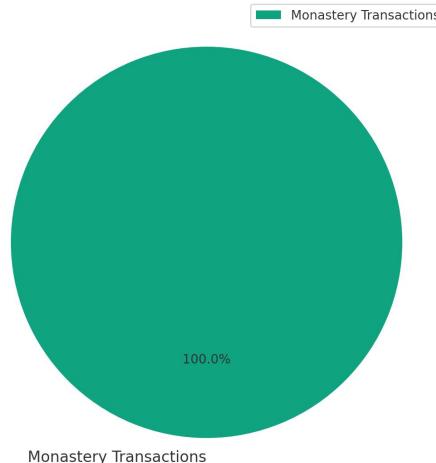


Visualizing Monastery's Historical Transactions. September 7, 2023 . ChatGPT August 3 Version.

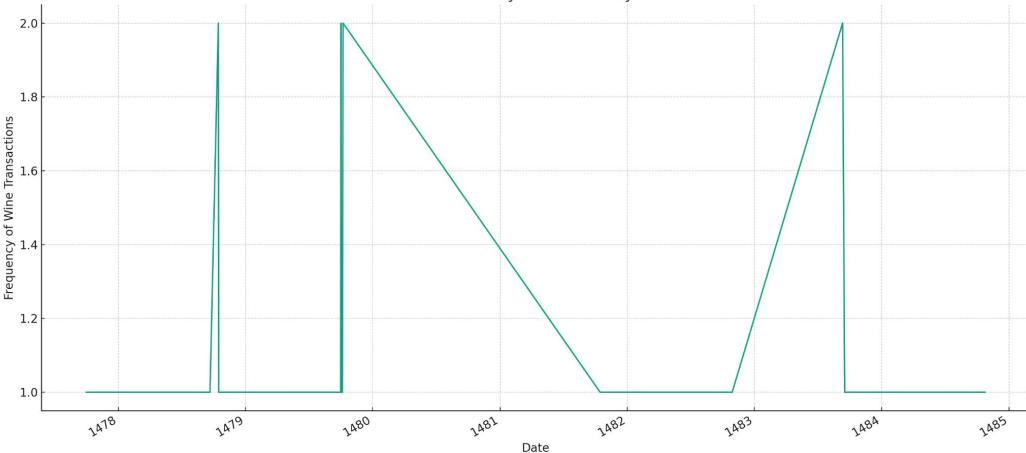
<https://chat.openai.com/share/01b19be6-c1ad-4f92-868c-f4edc9882d8c>

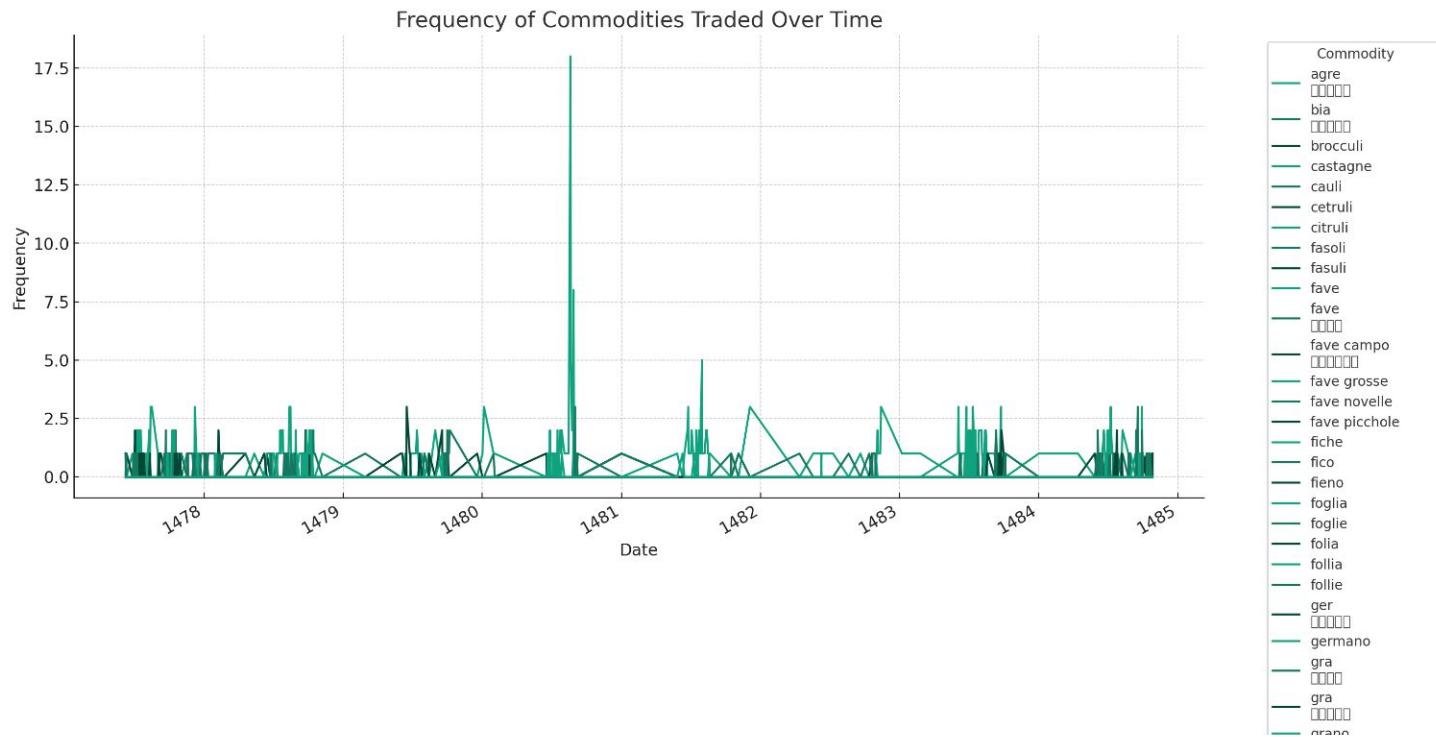
“Schlechte” Visualisierungen

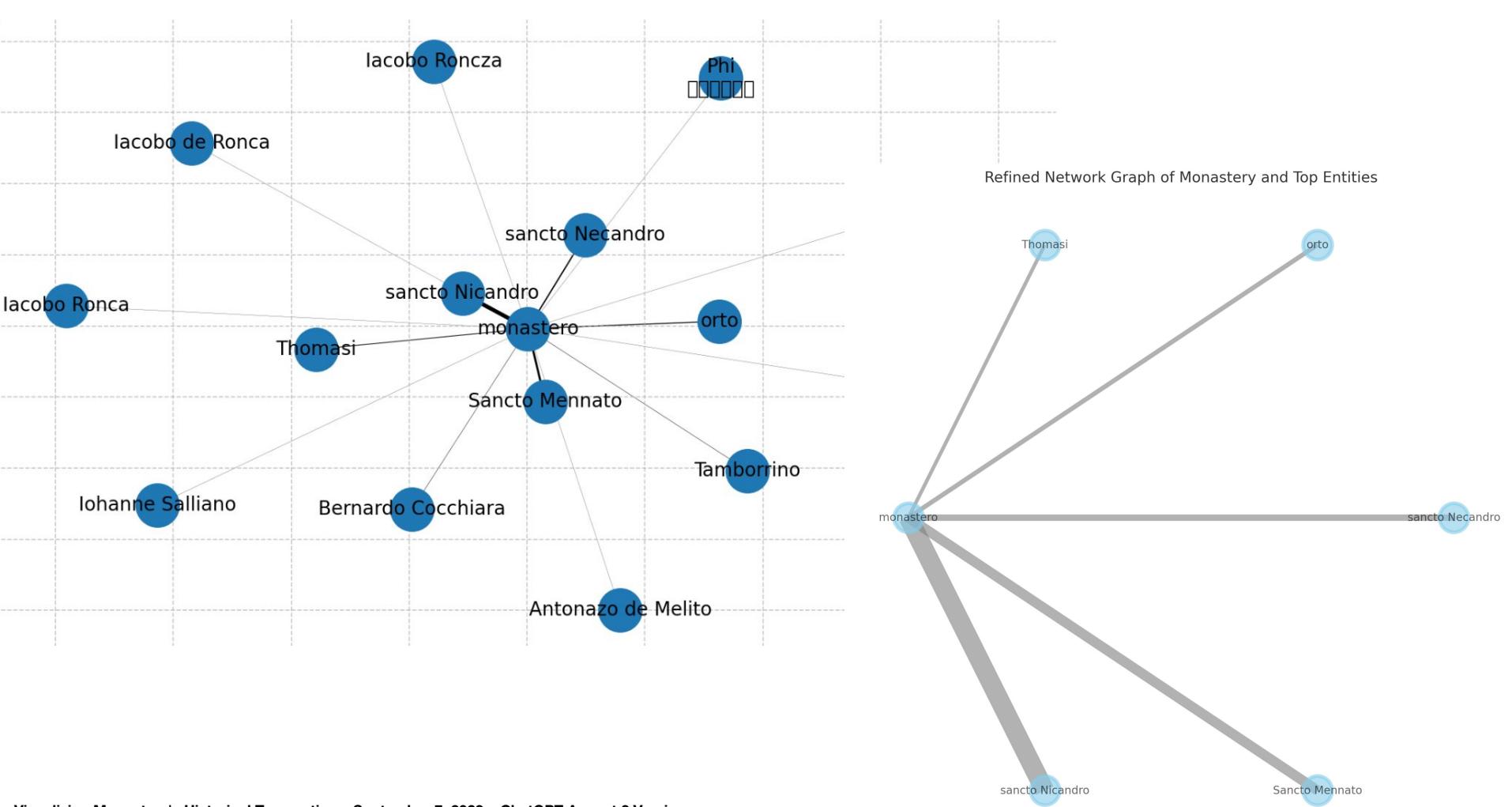
Proportion of Transactions Involving the Monastery



Wine Transactions by the Monastery Over Time

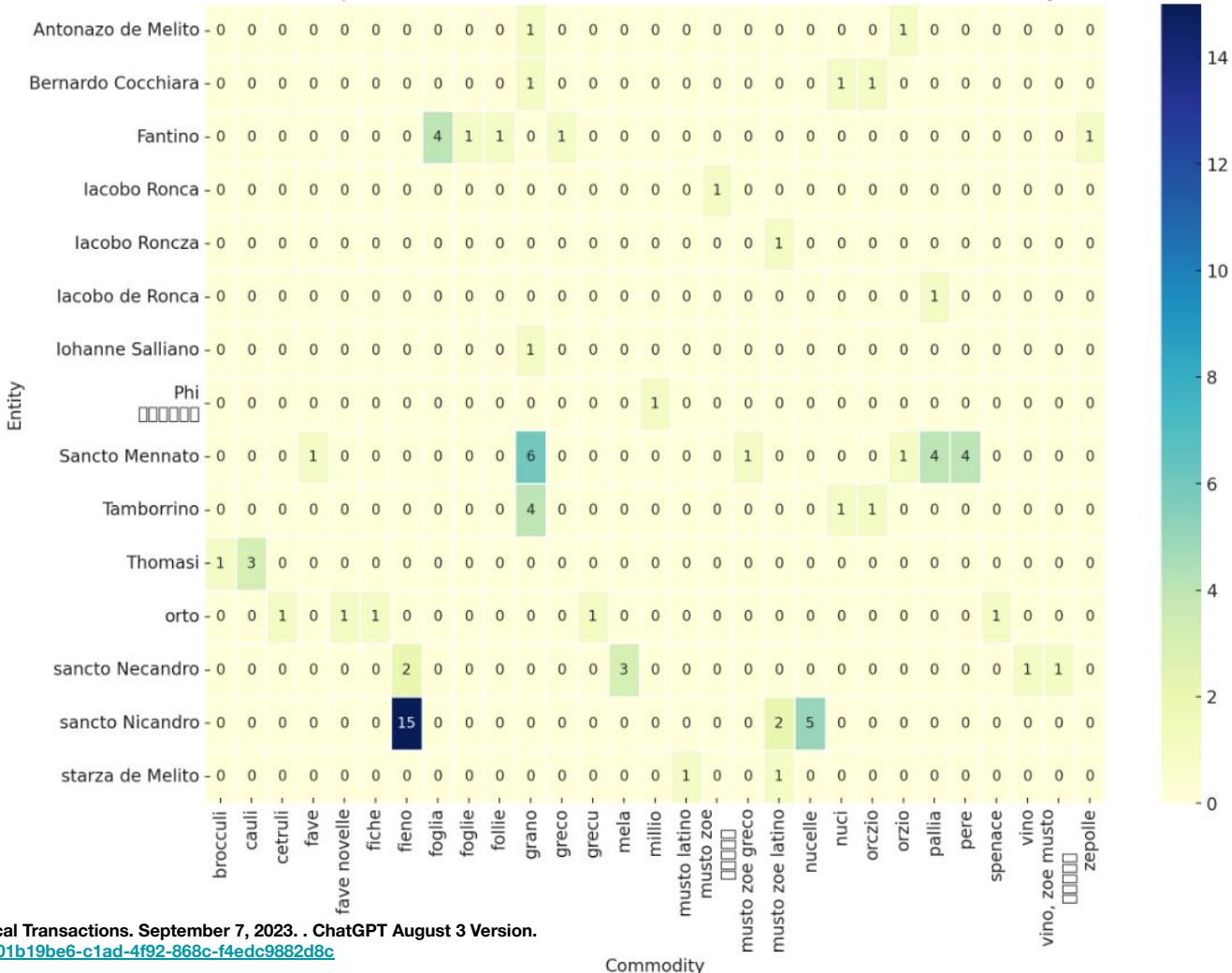






Refined Network Graph of Monastery and Top Entities

Heatmap of Entities vs. Commodities Transactions with the Monastery



“Interpretieren” - und weitere Gedanken



give answers to the following research questions:

- * Reconstruct the history of the landscape through the cultures.
- * To what extent was the monastery involved in the logic of a "regional" market, and what was the importance of wine sales and production?
- * To what extent can the monastery be understood as a microcosm in which a number of people had privileged relations with the monastery?]

Klingt irgendwie gut, wird jedoch vermutlich nicht Bestand haben bzw. ist eher eine “offensichtliche” Erkenntnis.

Aber wo stehen wir in 2-3 Jahren?

Wie sieht es aus, wenn ich 1000 solcher “Interpretationen” mache und KI-Agenten die Ergebnisse gegeneinander diskutieren lasse?

Es spricht nichts dagegen, Sekundärliteratur in die Diskussion einzubringen.

1. Reconstruct the history of the landscape through the cultures:

From the transaction records of the monastery, we observed that several commodities were traded frequently, giving insights into the agricultural practices and preferences of the period. The prominence of commodities such as wine, grain, oil, and fruit suggests that these were primary agricultural products cultivated in the region. The continuous trading of these commodities over the observed years indicates stable agricultural practices and possibly a consistent demand for these products. Thus, the landscape during this period can be characterized by vineyards, grain fields, olive orchards, and other fruit-bearing trees.

2. To what extent was the monastery involved in the logic of a "regional" market, and what was the importance of wine sales and production?:

The monastery's frequent and varied transactions highlight its active role in the regional market. It wasn't just a religious institution but a significant economic entity. The consistent trading of a wide range of commodities indicates the monastery's integration into the regional market dynamics. Among the traded commodities, wine stood out both in terms of transaction frequency and potential economic value. The regularity of wine transactions and its emphasis in the records underscores the importance of wine sales and production for the monastery. This suggests that the monastery might have had its vineyards or had close ties with local wine producers, making wine a cornerstone of its economic activities.

3. To what extent can the monastery be understood as a microcosm in which a number of people had privileged relations with the monastery?:

The monastery's transaction records reveal repeated interactions with specific entities, suggesting the existence of preferred or privileged partners. Some entities had frequent and high-value transactions, implying a deeper economic relationship with the monastery. The network graph further illustrated these relationships, with certain entities emerging as central players in the monastery's economic network. These consistent interactions indicate that the monastery wasn't just an isolated religious institution; it was an integral part of a larger economic ecosystem. Within this ecosystem, certain entities (or individuals) held privileged positions, possibly due to trust, long-standing partnerships, or mutual economic benefits.

Extrahieren. Analysieren. Visualisieren.

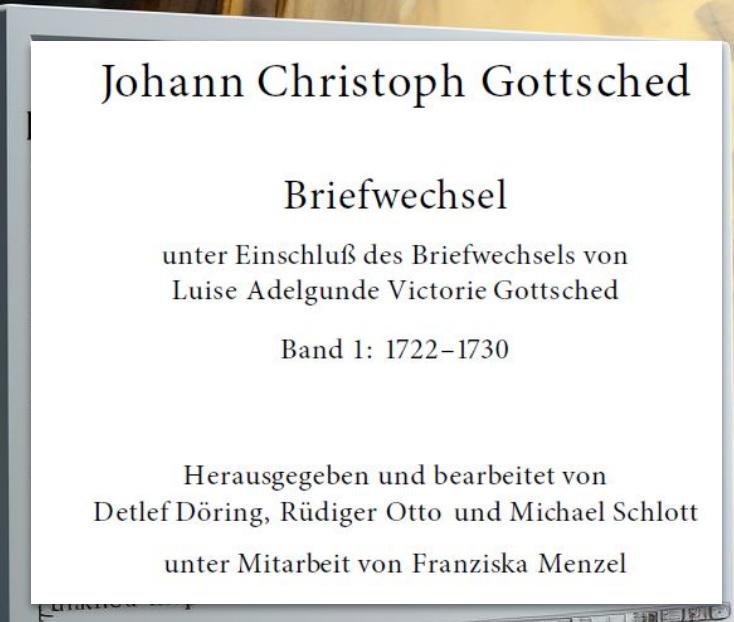
Johann Christoph Gottsched

Briefwechsel

unter Einschluß des Briefwechsels von
Luise Adelgunde Victorie Gottsched

Band 1: 1722–1730

Herausgegeben und bearbeitet von
Detlef Döring, Rüdiger Otto und Michael Schlott
unter Mitarbeit von Franziska Menzel



Midjourney: <https://s.mj.run/32MoH4uIA> full body, 18th century man watching modern computer, Baroque style, vortex of pdf documents, dreamy --quality 0.25 --ar 16:9 --weird 10
Vary: pdf icon, computer screen, Baroque style, --weird 10 --ar 16:9
Vary: <https://s.mj.run/tJ5NmFx60II> pdf viewer with open document --ar 16:9

PDF → Datenvisualisierung

PDF mit 636 Seiten. | Auf Seite 71-481 sind Briefe | Print PDF Edition. **Kein triviales Beispiel!!!**

Döring, Detlef, Otto, Rüdiger and Schlott, Michael. Band 1 1722–1730, Berlin, Boston: De Gruyter, 2007. <https://doi.org/10.1515/9783110924572>

3. GOTTLIEB SIEGFRIED BAYER AN GOTTSCHED
Königsberg April 1724 [6]

Überlieferung

Abschrift: Dresden, SLUB, M 166 I, Nr. 4, S. 12 (unsere E
Druck: Uhl, S. 344.

Bayer hat nie unterlassen, an Gottsched zu denken, und er glau
ihn vergessen hat. Was er über Gottsches Person an Mencke
hat, werde er von diesen erfahren haben. Den Verleger Sch
Pietschs und Bayers Namen grüßen.

1. LUDWIG PHILIPP THÜMMIG AN GOTTSCHED,
Halle 4. Mai 1722

Überlieferung

Original: Leipzig, UB, 0342 I, Bl. 1. 1 S.
Abschrift: Dresden, SLUB, M 166 I, Nr. 1, S. 3.
Druck: Uhl, S. 365–366.

Thümmig bedauert, daß Gottsches Brief nicht die erw
gewesen sei. Er kann sich daher nicht mit Gottsches Einw
wünscht, es möge ihm recht bald eine Abschrift zukom
längst eine Untersuchung über die Unsterblichkeit der
wird Johann Jakob Meckelburg diese und andere Abhandl
Gottsched möge bitte den Freund Christoph Friedrich B:

ⁱⁱ wieviel ändert Bearb. aus wie

¹¹ Christian Bernhard Thamm (Korrespondent), ein Kommil
am 22. September 1711 in Königsberg immatrikuliert; vg
Briefe, S. 16. Er verfaßte ein Gedicht zu Ehren von Gottsch
vgl. Reicke, S. 70. Als Adressat kommt indes auch der kr
meister Johann Thamm (1663–1727) in Frage. Gottsched
sertation *Genuinam omnipraesentiae divinae notionem* (Mit

VIRO NOBILISSIMO, DOCTISSIMO Que/ Jol
GOTTSCHEDIO/ s[alutem] d[icit] p[lurimam]

Vehementer doleo, Literas Tuas dissertationem,¹
non habuisse comitem. Solæ enim in manus mea
siderio Tuo satisfacere, neque objectiones rimar

5 Pietsch an Gottsched – März 1724

2. JOHANN VALENTIN PIETSCH AN GOTTSCHED,
Königsberg 16. März 1724

Überlieferung

Original: Leipzig, UB, 0342 I, Bl. 2–4. 5 S.
Abschrift: Dresden, SLUB, M 166 I, Nr. 2, S. 4–7.
Druck: Waniek, S. 47–48 (Teildruck).

HochwohlEdler/ Insonders geehrtester Herr Magister

Euere HochwohlEdl: erweisen mir eine rechte Freundschaft daß Sie mich
Ihres geneigten Andenkens zu versichern belieben wollen, und fehlet mein
Vergnügen recht vollkommen zu machen nichts mehr als eine kleine Nach
richt von Ihrem glücklichen Zustand, denn ich glaube nicht daß es einer
Person von Ihren Qualitäten, unglücklich gehen kann, es sey denn daß

Briefe aus PDF extrahieren

I am uploading a PDF of "Johann Christoph Gottsched's Correspondence. Historical-critical edition". Concentrate only on **pages 71 to 481**. These pages contain correspondence. Correspondences can be 1 page or more. They always start with a pattern such as "**199. Luise Adelgunde Victorie Kulmus an Gottsched, Danzig 27. Oktober 1730 [188]**"

There are **212 letters in total**. The last one is "**212. Georg Christian Wolff to Gottsched, Strasbourg 29.December 1730 [205]**" begins on page 478 and ends on page 481.

Here is the first correspondence, on page 71. It starts at page 71 and ends on page 72. It contains footnotes, numbered lines (5, 10, 15, ...), a heading with the number and title of the correspondence, date, summary and the text itself.

...
{FULLTEXT OF A LETTER WITH FOOTNOTES}
...

Extract correspondence information from the provided PDF, focusing on capturing the complete and accurate details of each letter, including the number, participants, date, start page, and end page. Employ advanced text extraction and processing techniques to handle variability and complexity in text structures and formatting.

Deep Dive into Conversation

Advanced Data Analysis - PDF to
DataVis. September 22, 2023. ChatGPT
August 3 Version.

<https://chat.openai.com/share/7870d8c2-bccd-4580-9f79-cf0159b10781>

Finished working Hide work ^

python Copy code

```
from PyPDF2 import PdfFileReader
import re

# Path to the uploaded PDF file
file_path = "/mnt/data/Gottsched-Edition.pdf"

# Define the range of pages to extract text from
start_page = 71
end_page = 481

# Initialize a PdfFileReader object
with open(file_path, 'rb') as file:
    pdf_reader = PdfFileReader(file)

# Check if the end_page is within the total number of pages in the PDF
if end_page > pdf_reader.numPages:
    raise ValueError(f"The PDF only has {pdf_reader.numPages} pages, but you requested page {end_page}.")

# Extract text from the specified range of pages
extracted_text = {}
for page_number in range(start_page - 1, end_page): # Page numbering is 0-based
    page = pdf_reader.getPage(page_number)
    extracted_text[page_number + 1] = page.extractText()

# Check the extraction for the first few pages
extracted_text_preview = {page: text for page, text in list(extracted_text.items())[:5]}
```

RESULT

```
{71: 'Thümmig an Gottsched - Mai 1722\nn3\nn1.\nLudwig Philipp Thümmig an Go',  
72: '4\nThümmig an Gottsched - Mai 1722\nnCum Te rer um metaphysicarum ama',  
73: 'Pietsch an Gottsched - März 1724\nn5\nn2.\nJohann Valentin Pietsch an'}
```

It seems like the text extraction from the PDF was successful. Next, we will parse the extracted text to identify the correspondences, extracting their number, participants, date, start page, and end page. The pattern indicated for the start of each correspondence is a number followed by a period, then names, a location, and a date.

For each identified correspondence:

- We will take the number preceding the period as the correspondence number.
- The text between the number and the location will be taken as the participants.
- The text after the location will be taken as the date.
- The page number where this pattern is found will be the start page of the correspondence.
- The end page will be the page number just before the next correspondence starts.

Let's begin by identifying and extracting the details of each correspondence.

Finished working

Show work ▾

It seems that the regular expression pattern did not match any correspondence headers in the extracted text. This could be due to variations in the text format or the presence of special characters or line breaks in the text.

I will try a more flexible approach, searching for possible matches and refining the pattern. Specifically, I will look for lines starting with a number and a period to identify potential starts of correspondences, and then adjust the regular expression pattern accordingly.

GPT-4 erkennt Position von Personen, Ort, Datum etc.

GPT-4 Strategie sind RegEx

Now that we have extracted a few lines surrounding the potential matches, we can observe the structure around the correspondence start and refine our regular expression pattern to accurately capture the correspondence details such as number, participants, and date.

From the extracted surrounding lines, it appears that the correspondence details are structured as follows:

- The correspondence number is followed by a period.
- The participants are listed after the period and before the comma.
- The date is listed after the location and the last comma.

Let's refine our regular expression pattern based on these observations and reapply it to the extracted text to accurately capture the correspondence details.

Finished working

Show work ▾

It appears that the refined pattern was able to identify the start of correspondences, but there are still inaccuracies in the captured details. The participants and date fields are not captured correctly, with some extraneous text included in the participants field and the date field capturing incorrect information.

GPT-4 erkennt, dass die eigenen RegEx noch nicht zum gewünschten Ergebnis führen und versucht immer wieder neue RegEx.

To address this, I will loosen the pattern slightly, especially around the date capturing group, and attempt the extraction again. Additionally, I will try to capture some lines around potential starts of correspondences to refine the pattern more accurately.

Finished working Hide work ^

python

Copy code

```
# Extract several lines around the potential starts of correspondences to refine the pattern
surrounding_lines_dict = {}

# Extract 2 lines before and after the potential match to have more context
for page, matches in potential_matches.items():
    text = extracted_text[page]
    surrounding_lines_dict[page] = extract_surrounding_lines(text, [int(match['start']) - 2, int(match['start']) + 2])

# Check the extracted surrounding lines for the first few potential matches
surrounding_lines_refinement_preview = {page: lines for page, lines in list(surrounding_lines_refinement.items())[:5]}
surrounding_lines_refinement_preview
```

RESULT

```
{71: {1: 'Thümmig an Gottsched - Mai 1722\n3\n1.\nLudwig Philipp Thümmig a',  
73: {2: 'Pietsch an Gottsched - März 1724\n5\n2.\nJohann Valentin Pietsch',  
75: {2: 'Pietsch an Gottsched - März 1724\n7\nAlbdenn stehet i mein Carl'
```

Wir können uns jederzeit den Code und auch die Ergebnisse des ausgeführten Python-Codes ansehen.

To refine the pattern further, let's attempt to capture only text that appears to be part of a date (i.e., text containing digits, month names, and possibly some punctuation like periods or commas) after the last comma in the line. Specifically, we'll try to capture text that follows the last comma and that precedes any newline character, excluding any non-date-like text.

Let's adjust the pattern and attempt the extraction again.

Finished working

Show work ▾



continue



Finished working

Show work ▾

It appears that the final refined pattern still didn't capture the correspondences accurately, and we have incorrect and repeated entries in the extracted correspondences.

Given the variability and complexity in text structures and formatting, refining the regular expression further might not be sufficient to accurately capture all correspondence details.

Wenn ChatGPT-4 stoppt und uns nicht direkt fragt, wie es weitergehen soll, dann einfach “continue” drücken.



List exactly all the rules you use to apply the regular expression and the usual procedures to identify the beginning of a correspondence. Also list any challenges. Take a deep breath and think step by step about how to solve the problem.



Certainly! Here is a step-by-step breakdown of the current approach and rules applied, along with challenges identified.

Approach to Identify the Beginning of Correspondence:

1. Identify Potential Starts:

- Look for lines starting with a number followed by a period, as it is the common pattern observed at the start of each correspondence.

2. Extract Participants:

- From the identified line, extract the text between the number (at the start of the line) and the first comma as participants.
- If participants are not found in the identified line, extend the search to subsequent lines until participants are found or until a stopping condition is met.

3. Extract Date:

- From the identified line, extract the text after the last comma as the date.
- If the extracted date does not look like a valid date, extend the search to subsequent lines until a valid date is found or until a stopping condition is met.

4. Handle Variability and Exceptions:

Scorer	Optimizer / Source	Instruction position	Top instruction	Acc
Baselines				
PaLM 2-L	(Kojima et al., 2022)	A_begin	Let's think step by step.	71.8
PaLM 2-L	(Zhou et al., 2022b)	A_begin	Let's work this out in a step by step way to be sure we have the right answer.	58.8
PaLM 2-L		A_begin	Let's solve the problem.	60.8
PaLM 2-L		A_begin	(empty string)	34.0
text-bison	(Kojima et al., 2022)	Q_begin	Let's think step by step.	64.4
text-bison	(Zhou et al., 2022b)	Q_begin	Let's work this out in a step by step way to be sure we have the right answer.	65.6
text-bison		Q_begin	Let's solve the problem.	59.1
text-bison		Q_begin	(empty string)	56.8
Ours				
PaLM 2-L	PaLM 2-L-IT	A_begin	Take a deep breath and work on this problem step-by-step.	80.2
PaLM 2-L	PaLM 2-L	A_begin	Break this down.	79.9
PaLM 2-L	gpt-3.5-turbo	A_begin	A little bit of arithmetic and a logical approach will help us quickly arrive at the solution to this problem.	78.5
PaLM 2-L	gpt-4	A_begin	Let's combine our numerical command and clear thinking to quickly and accurately decipher the answer.	74.5
text-bison	PaLM 2-L-IT	Q_begin	Let's work together to solve math word problems! First, we will read and discuss the problem together to make sure we understand it. Then, we will work together to find the solution. I will give you hints and help you work through the problem if you get stuck.	64.4
text-bison	text-bison	Q_end	Let's work through this problem step-by-step:	68.5
text-bison	gpt-3.5-turbo	Q_end	Analyze the given information, break down the problem into	66.5

Yang, Chengrun, Xuezhi Wang, Yifeng Lu, Hanxiao Liu, Quoc V. Le, Denny Zhou, und Xinyun Chen. „Large Language Models as Optimizers“. arXiv, 6. September 2023.
<https://doi.org/10.48550/arXiv.2309.03409>.

“Take a deep breath and think step by step”

Ich habe den Überblick verloren. Sag mir bitte, was du da versuchst zu matchen.

4. Handle Variability and Exceptions:

- Manually inspect and adjust the conditions and string manipulations to handle the variability and exceptions in the text format more accurately.
- Specifically, ensure that participants are captured only till the last occurrence of a name before any comma.
- Ensure that the date is captured after the last occurrence of a location or a name and a comma, and only if it looks like a valid date.

Challenges:

1. Variability in Text Format:

- The text format varies greatly, with different structures, punctuations, and newline characters, making it challenging to create a one-size-fits-all solution.

2. Inaccurate Extraction of Participants and Date:

- The heuristic approach still didn't accurately extract the participants and the date. The date field is especially challenging, often capturing incorrect information or not capturing at all.

3. Incorrect Entries:

- The heuristic approach sometimes creates incorrect entries, capturing lines that are not the start of correspondences.

4. Complexity in Text Structures:

- The complexity in text structures, with multiple commas in a line and important information spanning multiple lines, complicates the extraction process.

Potential Solution:

1. Refine Heuristic Approach:

- Continue to refine the conditions and string manipulations based on manual inspection of the text, to handle variability and exceptions more accurately.
- Specifically, focus on accurately capturing the date by adjusting the conditions to ensure

Die Aufgabe ist nicht einfach mit regulären Ausdrücken zu lösen, aber ich konnte einen "Refine, verify, repeat"-Workflow einrichten.

Next Steps:

1. Further Refinement:

- More rigorously validate the extracted participants and dates to exclude any invalid or irrelevant text.
- Handle the edge cases where incorrect entries are being created, ensuring only valid correspondences are captured.

2. Validation and Verification:

- Validate each extracted correspondence against known patterns and formats to ensure the accuracy of the extraction.
- Cross-verify the extracted details with the original text to ensure that no valid correspondence is missed and no incorrect entry is included.

Let's proceed with these further refinements and validations.

Finished working Show work ▾

Despite the more rigorous refined heuristic approach, the extraction still contains the incorrect entry where the participants' field captured 'änder t Bearb. aus', and the date '16. August 17/17'.

Next Steps:

1. Addressing Incorrect Entries:

- Examine the specific conditions or patterns that lead to the creation of incorrect entries and address them.
- If necessary, manually correct or remove such incorrect entries.

2. Validation and Verification:

- Further validate each extracted correspondence against known patterns and formats to ensure the accuracy of the extraction.
- Cross-verify the extracted details with the original text to ensure that no valid



Give me a txt file in which you write in each line all the correspondences that you can filter out with the current code.

I will look at it as a human being and help you.



"50. Johann Georg Bock an Gottsched, Königsberg 6. April 1728 [32.84]"

is always followed "Überlieferung" some lines later.

Original: Leipzig, UB, 0342 I, Bl. 55–58. 7 S.

Abschrift: Dresden, SLUB, M 166 I, Nr. 41, S. 79–85.

1. Participants: Ludwig Philipp Thümmig an Got ts
2. Participants: Johann Valentin Pietsch an Got t
3. Participants: änder t Bearb. aus, Date: 16. Au
4. Participants: Daniel Heinrich Arnoldt an Got t
5. Participants: Daniel Heinrich Arnoldt an Got t
6. Participants: Got tlieb Siegfried Bayer an Got
7. Participants: Heinrich von Sanden an Got tsche
8. Participants: Got tlieb Siegfried Bayer an Got
9. Participants: Daniel Heinrich Arnoldt an Got t
11. Participants: Johann Heinrich Kreuschner an G
12. Participants: Johann Christian Clodius an Got
13. Participants: John Lesingham an Got tsched, D
14. Participants: Got tsched an Johann Ernst Philippi, Date: 14. August 1726, Start Page: 109, End Page: 110
15. Participants: Johann Friedrich Schreiber an Got tsched, Date: 1. November 1726, Start Page: 111, End Page: 117
17. Participants: Leipzig, Date: 15. März 1727, Start Page: 118, End Page: 119
18. Participants: Leipzig, Date: 29. März 1727, Start Page: 120, End Page: 120
19. Participants: Got tsched an Johann Ulrich König, Date: 3. Mai 1727, Start Page: 121, End Page: 123
20. Participants: Kur fürst von Sachsen, König in Polen, Date: 25. Mai 1727, Start Page: 124, End Page: 124
21. Participants: Christian Franz Schwenschuch an Got tsched, Date: 18. Juni 1727, Start Page: 125, End Page: 127
22. Participants: John Lesingham an Got tsched, Date: 19. Juli 1727, Start Page: 128, End Page: 131
23. Participants: Johann Friedrich Schreiber an Got tsched, Date: 30. Juli 1727, Start Page: 132, End Page: 134
24. Participants: Got tsched an Johann Ernst Philippi, Date: 6. August 1727, Start Page: 135, End Page: 135
25. Participants: Johann Got tlieb Krause an Got tsched, Date: 27. August 1727, Start Page: 136, End Page: 137
26. Participants: Johann Georg Bock an Got tsched, Date: 21. September 1727, Start Page: 138, End Page: 141
27. Participants: Heinrich Graf von Bünau an Got tsched, Date: 16. Oktober 1727, Start Page: 142, End Page: 141
28. Participants: Johann Got tlieb Krause an Got tsched, Date: 22. Oktober 1727, Start Page: 142, End Page: 143
29. Participants: Bayreuth, Date: 25. Oktober 1727, Start Page: 144, End Page: 155
30. Participants: Johann Christian Clodius an Got tsched, Date: 1725–1727, Start Page: 156, End Page: 162

Ich schaue mir die Ergebnisse an und gebe Feedback, was nicht funktioniert hat: "X ist so, sollte aber so sein".

136 wurde nicht erfasst. 137 stimmt wieder ...

136. JOHANN THEODOR JABLONSKI AN GOTTSCHED,
Berlin 7. März 1730 [154]

Überlieferung

Original: Leipzig, UB, 0342 I, Bl. 202–203. 3 S.
Abschrift: Dresden, SLUB, M 166 I, Nr. 117, S. 226–227.

HochEdler, hochgelahrter/ Hochgeehrter Herr Professor

Dero geehrte beide vom 18 Jan. und 5 Febr. das erste an mich, das zweite
an das Directorium der Societät der Wissenschaften gerichtet, sind zu recht
eingelaufen, und habe ich die schuldigkeit auf jenes zu beobachten, nur
darum aufgeschoben, damit wegen des letzteren wozu mir in dem vorigen

von Racknitz an Gottsched – 1730

321

innern, daß dieselbe von einem der größten Redner und Dichter unserer
Zeiten aufgesetzt sey, der sich seiner Freiheit bedient. Aber die mittel-
mässigen und ungeübten Köpfe könnten vielleicht das in mir suchen, was
die Kraft der Worte in der ordentlichen Rede mit sich bringet. Helffen
doch E. HochEdelGeb. diesen Leuten, daß sie in diesen Fehler nicht ver-
fallen und nicht höhere Gedanken von mir haben, als ich verdiene. Ein
Poet verleiht nichts, wenn er bekennet, er habe als ein Poet geschrieben.
Ich habe die Ehre mit der reinesten hochachtung zu sey

E. HochEdelGebohrnen/ Ergebenster Diener/ J. L. Mosheim.

Helmstedt/ d. 1. Martii/ 1730.

10

Unter vielen Unruhen zu einer Zeit, da ich einen Sohn¹⁰ beerdige.
Ich bitte gehorsamst den Einschluß nicht ungütig zu deuten.

135. GALLUS MAXIMILIAN FREIHERR VON RACKNITZ
AN GOTTSCHED,
Wittenberg 1. März 1730

15

Überlieferung

Original: Leipzig, UB, 0342 I, Bl. 200–201. 3 S.
Abschrift: Dresden, SLUB, M 166 I, Nr. 116, S. 223–225.

WohlEdler hochgeschätzter/ Insonders hochgeertester Herr Professor./ Gro-
ßer Freund.

20

Diu tacui. Es ist wahr und hab ich die Grenzen der hofflichkeit ziemlich
überschritten. Allein entweder hochst-notwendige Verrichtungen, oder
ein kranker Leib haben mich von meiner Schuldigkeit so lange abgeholt-
ten. Es haben Sich unterdeßen viele Veränderungen zugetragen und Ew:
WohlEdlen besitzen rühmlichst jetzo die Stelle eines Professoris¹ worzu ich

25

¹⁰ Georg Christoph Mosheim (* 1729).

¹ Vgl. Einleitung.

```
{  
    "Number": 18,  
    "Participants": "Leipzig",  
    "Date": "",  
    "Start Page": 120,  
    "End Page": 120,  
    "Content": "Montag 31. Juli, 1809, S. 724.\nDen Auszug aus einem Brief Gottscheds an Meister veröffentlichte dessen Neffe, der\nSchriftsteller und Theologe Leonhard Meister",  
    "Footnotes": "",  
    "Überlieferung Metadata": "Druck: Morgenblatt für gebildete Stände Nr. 181, Montag"  
},
```

Die extrahierten Daten als JSON erzeugen.

Funktioniert nicht ideal, aber irgendwie so,
dass es doch funktionieren könnte!

52

Gottsched an Meister – März 1727

18. GOTTSCHED AN JOHANN HEINRICH MEISTER (LEMAITRE),
Leipzig 29. März 1727 [17.29]

Überlieferung

Druck: Morgenblatt für gebildete Stände Nr. 181, Montag 31. Juli, 1809, S. 724.

5 Den Auszug aus einem Brief Gottscheds an Meister veröffentlichte dessen Neffe, der Schriftsteller und Theologe Leonhard Meister (1741–1811) unter dem Titel: Gottsched's Geringsschätzung seiner Poesie und der deutschen Sprache. Aus einem Schreiben Gottsched's an meinen sel. Oheim, Heinrich Meister, der weil. als Hofprediger in Bückeburg lebte, und hernach als Dekan und Pfarrer zu Küsnach am Zürcher=See starb.

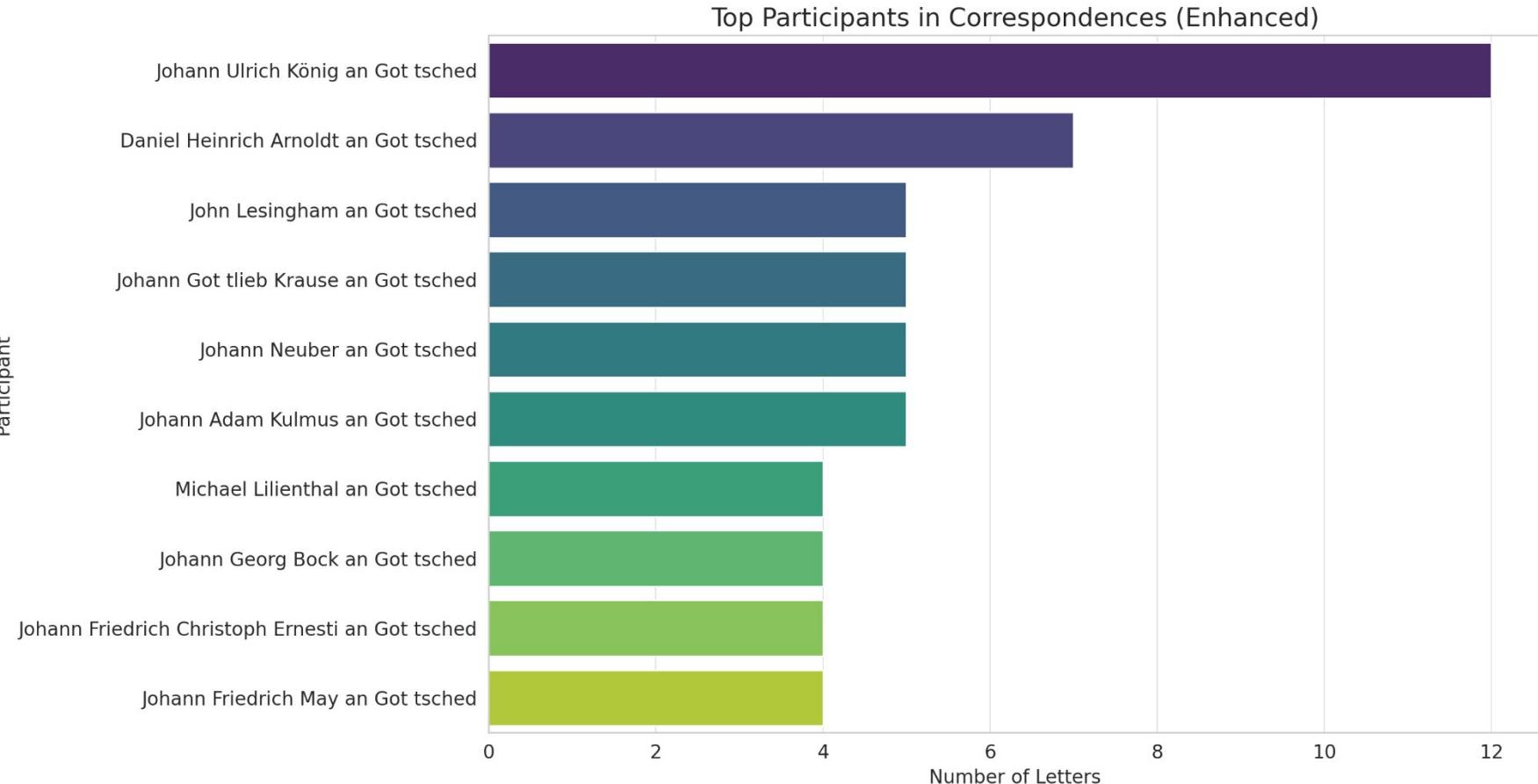
10 Leipzig, den 29 März 1727. Durch E. E. bin ich izt vollkommen überführt, daß ich die Uebergabe meiner Ode¹ an Se. Hochfürstl. Durchlaucht, den Herrn Marggrafen (von Bayreuth),² keinen bessern Händen hätte anvertrauen können, als wirklich von mir geschehen. Ungeachtet ich mich diesmal nicht darnach habe richten können, so werde mir doch in's Künftige die Zeilen:

D'un poète flatteur l'ame basse et servile
Est toujours pour les Grands en louange fertile,³

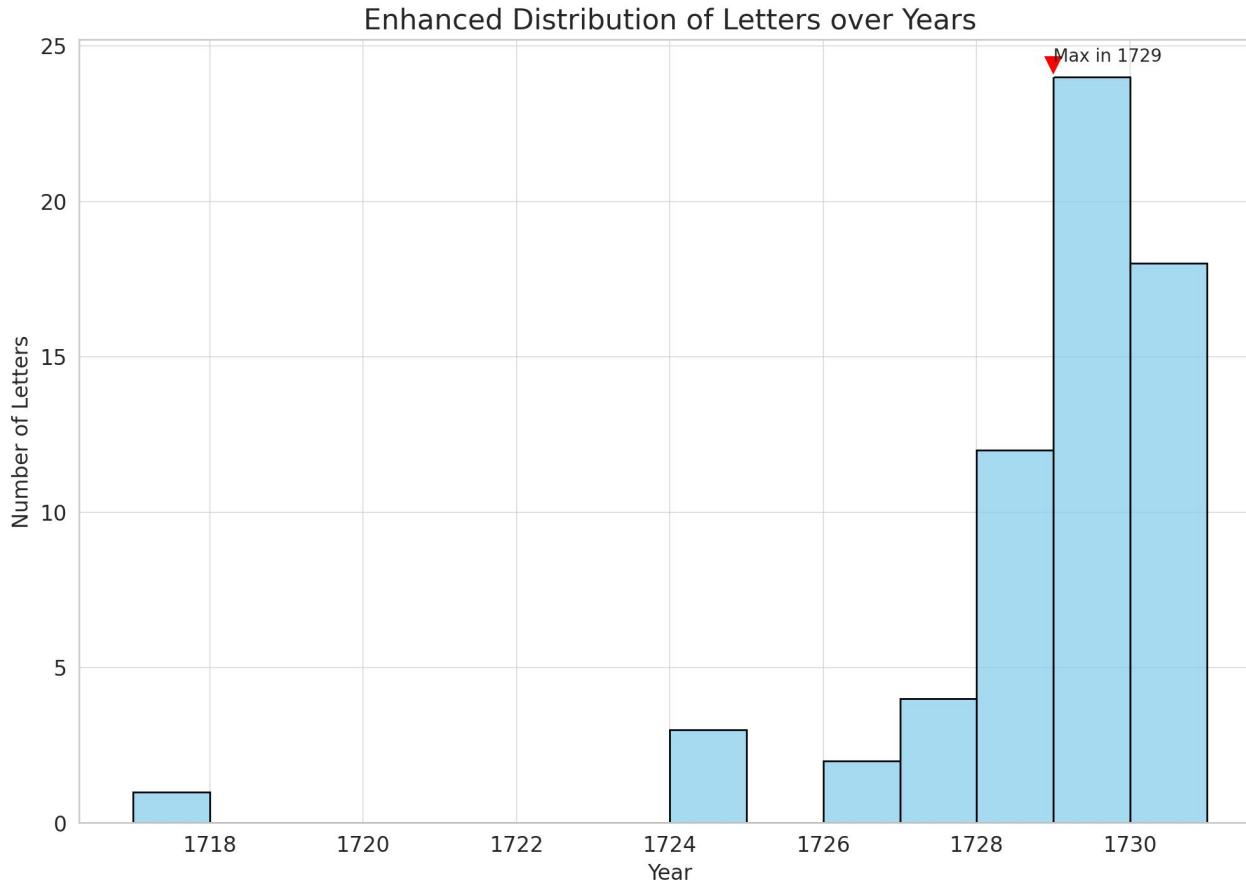
zur Lehre dienen lassen. ---- Ich wär' es wohl zufrieden, daß die mittelmäßigen Poeten aus der Republik verbannt würden,⁴ wenn ich nicht selbst

20 mit darunter begriffen wäre. Mein Horatius hat mir schon oft durch sein: Mediocribus esse poetis,⁵ eine Schamröthe abgenöthigt, und ich gestehe es, daß ich niemals ohne ein Widerstreben des Gewissens ein Blatt drucken lasse. Bey uns Deutschen aber heißt es: Inter coecos regnat luscus.⁶ Wir sind noch weit von derjenigen Richtigkeit der Gedanken und Ausdrücke entfernt, die in den schlechtesten französischen Poesien herrscht. Ich

PDF → Datenvisualisierung: Ergebnisse



PDF → Datenvisualisierung: Ergebnisse



PDF → Datenvisualisierung

- Einige Briefe sind richtig erfasst, aber einige sind nicht korrekt erfasst.
- In der Regel keine triviale Aufgabe.

- Text aus dem PDF muss passen: "Got tsched"
- Es gibt Informationen wie Zeilennummer,
Seitenüberschriften, Seitenzahl, die man nicht haben will.
- Inhalte treten optional auf: Datum, Namen, Orte, ...
- Es gibt ähnliche Strukturen, auf die gematcht wird.

- **Fazit:**

- Auch wenn es nicht fehlerfrei funktioniert hat, ist es ein spannendes Beispiel.
- Man müsste ausprobieren ganz viele Iterationen (halb)automatisiert
durchlaufen zu lassen. Nach dem Motto: "Schau hier ist Brief X: da hast du Y
falsch gemacht. Finde eine bessere Regel, um die Struktur programmatisch
zu erfassen" - Mit diesen "Feedback Iterations" müsste noch einiges
Herauszuholen: Experimentdauer: 2-3 Stunden.

Dann habe ich mich nochmal hingesetzt und mein Prompting verbessert

<https://chat.openai.com/share/ac753b84-1091-4d22-a378-55026be87353>

```
C: > Users > chris > Downloads > extracted_letters_refined.txt
1 Number: 1.
2 Sender: Ludwig Philipp Thümmig
3 Recipient: Gottsched
4 Place: Halle
5 Date: 4. Mai 1722
6 Start Page: 3
7 End Page: 4
8
9 Number: 2.
10 Sender: Johann Valentin Pietsch
11 Recipient: Gottsched
12 Place: Königsberg 1
13 Date: 6. März 1724
14 Start Page: 5
15 End Page: 9
16
17 Number: 4.
18 Sender: Daniel Heinrich Arnoldt
19 Recipient: Gottsched
20 Place: Königsberg 1
21 Date: 7. Mai 1724
22 Start Page: 10
23 End Page: 4
24
25 Number: 5.
26 Sender: Daniel Heinrich Arnoldt
27 Recipient: Gottsched
28 Place: Königsberg 1
29 Date: 1. August 1724
30 Start Page: 5
31 End Page: 18
32
```

Use pdfminer to extract the text. We need a good quality of the text.

Ignore the patterns on the first line of each page:

...

{new page}

Thümmig an Gottsched - Mai 1722

3

{new page}

4

Thümmig an Gottsched - Mai 1722

{new page}

Pietsch an Gottsched - März 1724

5

{new page}

6

Pietsch an Gottsched - März 1724

{new page}

Pietsch an Gottsched - März 1724

7

...

Here is a matching rule to identify the beginning of a letter:

...

NAME AN NAME,
PLACE DATE

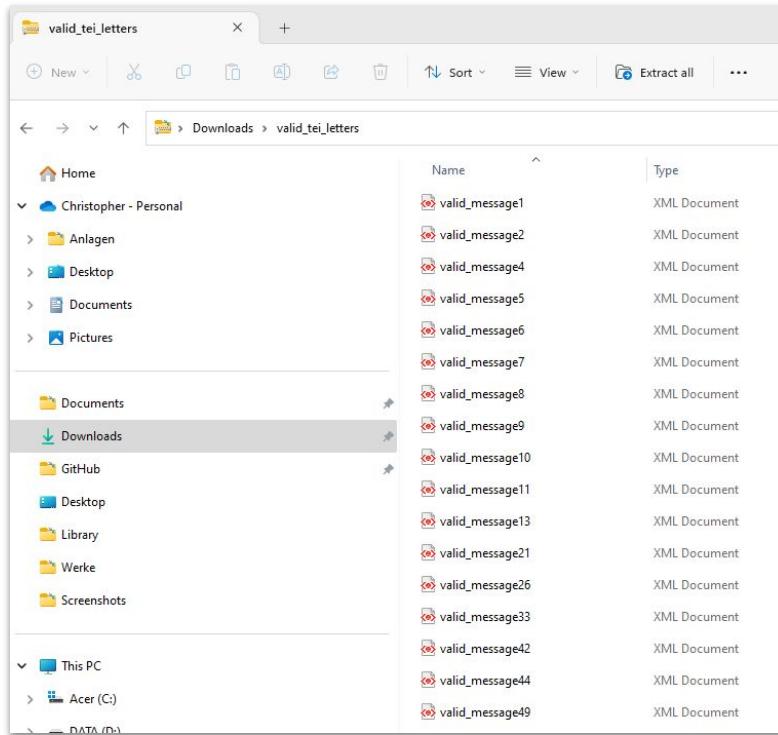
Überlieferung

...

PDF → TEI

ZIP mit 116/212

<correspDesc> TEI XML



```
1 <?xml version="1.0" ?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0" version="3.3.0">
3   <teiHeader>
4     <fileDesc>
5       <titleStmt>
6         <title>Letter Correspondence</title>
7       </titleStmt>
8       <publicationStmt>
9         <p>Letter correspondence extracted from PDF.</p>
10      </publicationStmt>
11      <sourceDesc>
12        <p>Extracted from provided PDF.</p>
13      </sourceDesc>
14    </fileDesc>
15    <correspDesc xml:id="message158">
16      <correspAction type="sent">
17        <persName>Christian von Loß</persName>
18        <placeName>Dresden </placeName>
19        <date when=" 1730-05-09"/>
20      </correspAction>
21      <correspAction type="received">
22        <persName>Gottsched</persName>
23      </correspAction>
24    </correspDesc>
25  </teiHeader>
26  <text>
27    <body>
28      <p>Full text of the letter goes here.</p>
29    </body>
30  </text>
31 </TEI>
```

Key findings

Allgemeine Arbeitsweise

- **Schritt-für-Schritt-Arbeitsweise:** Komplexere Aufgaben in Teilaufgaben zerlegen.
- **Klare Information und Bereitstellung von Kontext:** Es ist wichtig, alle erforderlichen Informationen und den Gesamtzusammenhang einer Aufgabe zu verstehen und zu kommunizieren.
- **Feedback und Verbesserung:** Detailliertes Feedback und Verbesserungsvorschläge sind notwendig, um qualitative Ergebnisse zu erzielen.
- **Perfektion und Fehlerfreundlichkeit:** Es sollte nicht erwartet werden, dass Arbeiten oder Ergebnisse von Anfang an perfekt und fehlerfrei sind.
- **Überprüfung der Korrektheit:** Alle Arbeiten und Ergebnisse sollten immer auf ihre Korrektheit überprüft werden.
- **“Document as Knowledge vs. Document as an Object”:** Der Umgang mit Dokumenten, die man als Kontext nutzen möchte, ist anders als der Umgang mit Objekten, die man "erarbeiten" möchte.

Key findings

Umgang mit Code und Technologie

- **Klarheit in der Code-Beschreibung:** Die Anforderungen und Einschränkungen des Codes müssen klar und präzise formuliert werden, um Missverständnisse und Fehler zu vermeiden.
- **Wahrung des Code-Execution-State:** Da der Code-Execution-State schnell "abstürzen" kann, ist es notwendig, kontinuierlich am Code zu arbeiten und sicherzustellen, dass keine Daten verloren gehen.

Forschungs- und Entwicklungsprozess

- **Schärfen von Prototyping und Forschungsfragen:** Eine klare Definition und Formulierung der Ziele sind während der Entwicklungsphase entscheidend.
- **Relevanz von User Stories:** Das Extrahieren von User Stories hilft dabei, Kontext zu generieren: Übertragung von Forschungsfragen auf Implementierung.

Anhang



ufbas-tei.xml



ufbas-categories.xml

Analyse des TEI XML des “Urfehdenbuch X der Stadt Basel”

This TEI XML represents the Urfehdenbuch X of the city of Basel (1563-1569). In this TEI XML a historical text has been annotated.

Encoding guidelines:

...

References to the data model are made by references in the global attribute @ana. In addition to references to the data model, monetary amounts are encoded using <measure>. Abbreviations are resolved using <ex>. Notes are assigned to an unspecified additional layer in the @change attribute. The @rend attribute is used to describe visual appearances that deviate from the normal text flow (position of annotations on the margin: margin, superscript: super). Editors' annotations are marked with @resp="#editor". Marks with squiggles or floral characters are coded as <metamark>. Line breaks at the beginning of headings or notes should be added implicitly, as they are encoded in the text flow of the paragraphs. Eight hands (specified in the @scribe attribute of the <handShift> encoding) are distinguishable without further identification, so the source description omits the encoding. The list of identifiable and referenced locations is captured in the <back> of the text. The RDF representation primarily uses a vocabulary developed for mapping the data model in the project from the namespace "http://gams.uni-graz.at/ufbas/#", which is explained in the category list and is available as a TEI-coded taxonomy. The mapping of these categories into a formal ontology that can be further used in historical crime research remains a research desideratum that could not be achieved with the resources available to the teaching project.⁵ The relationships between factual statements and text are expressed via the oa:hasTarget property. Place names are encoded by skos:prefLabel.

...

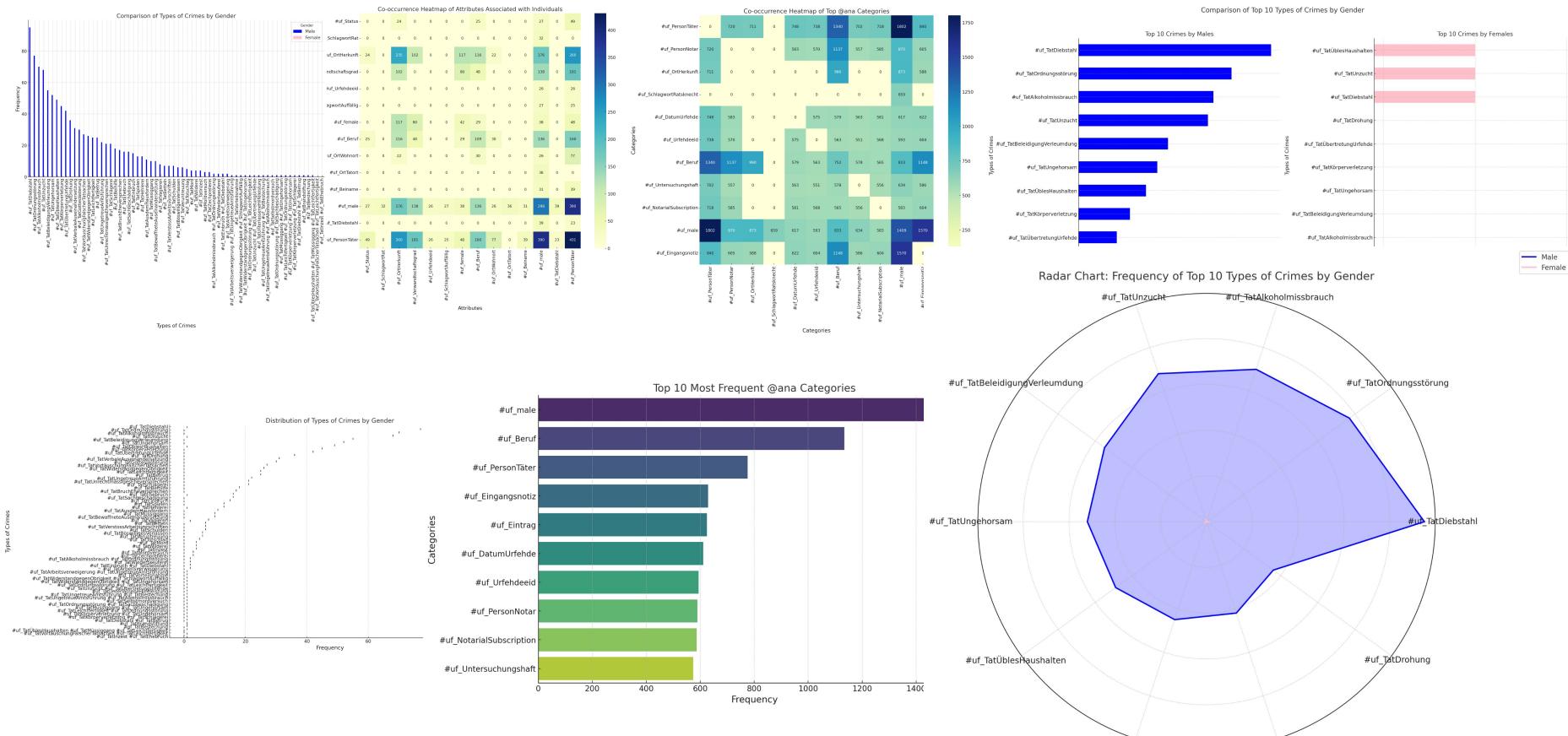
The second file (ufbas-categories.xml) contains the categories.

Just focus on the <text> and ignore the teiHeader.

You are an expert in criminal history and exploratory Data Analysis. I would like you to analyse the dataset in detail and perform several analyses to explore the dataset. Be very detailed and work step by step.

UFBAS - Analyse. 04.09.2023. ChatGPT August 3 Version.

<https://chat.openai.com/share/69a41d52-887b-4ab2-b090-2aeab759158a>





Take a deep breath and analyse the given task step by step: Convert the given <taxonomy> element in the teiHeader of the TEI XML document to SKOS.

```
<taxonomy> /t:TEI/t:teiHeader/t:encodingDesc/t:classDecl/t:taxonomy
```

Here is a small sample XML from the file

```
...  
  
<taxonomy ana="bk:Taxonomy">  
  <gloss xml:lang="en">Commodities</gloss>  
  <category xml:id="c_dairyproduct">  
    <catDesc><term ref="wd:Q185217" xml:lang="en">dairy product</term></catDesc>  
    <category xml:id="c_butter">  
      <catDesc><term ref="wd:Q34172" xml:lang="en">butter</term></catDesc>  
    </category>  
  </category>  
  <category xml:id="c_service">  
    <catDesc><term ref="wd:Q7406919" xml:lang="en">service</term></catDesc>  
    <category xml:id="c_work">  
      <catDesc><term ref="wd:Q6958747" xml:lang="en">work</term></catDesc>  
    </category>  
    <category xml:id="c_rental">  
      <catDesc><term ref="wd:Q157171" xml:lang="en">renting</term></catDesc>  
      <category xml:id="c_oxenrental">  
        <catDesc>  
          <term xml:lang="en">oxen rental</term>  
        </catDesc>  
      </category>  
      <category xml:id="c_cartrental">  
        <catDesc>  
          <term xml:lang="en">cart rental</term>  
        </catDesc>  
      </category>  
      <category xml:id="c_waggonrental">  
        <catDesc>  
          <term xml:lang="en">waggon rental</term>  
        </catDesc>  
      </category>  
      <category xml:id="c_heiferrental">  
        <catDesc>  
          <term xml:lang="en">heifer rental</term>  
        </catDesc>  
      </category>
```

Transforming TEI <taxonomy> to SKOS.

```
...  
  
<taxonomy ana="bk:Taxonomy">  
  <gloss xml:lang="en">Commodities</gloss>  
  <category xml:id="c_dairyproduct">  
    <catDesc><term ref="wd:Q185217" xml:lang="en">dairy product</term></catDesc>  
    <category xml:id="c_butter">  
      <catDesc><term ref="wd:Q34172" xml:lang="en">butter</term></catDesc>  
    </category>  
  </category>  
  <category xml:id="c_service">  
    <catDesc><term ref="wd:Q7406919" xml:lang="en">service</term></catDesc>  
    <category xml:id="c_work">  
      <catDesc><term ref="wd:Q6958747" xml:lang="en">work</term></catDesc>  
    </category>  
    <category xml:id="c_rental">  
      <catDesc><term ref="wd:Q157171" xml:lang="en">renting</term></catDesc>  
      <category xml:id="c_oxenrental">  
        <catDesc>  
          <term xml:lang="en">oxen rental</term>  
        </catDesc>  
      </category>  
      <category xml:id="c_cartrental">  
        <catDesc>  
          <term xml:lang="en">cart rental</term>  
        </catDesc>  
      </category>  
      <category xml:id="c_waggonrental">  
        <catDesc>  
          <term xml:lang="en">waggon rental</term>  
        </catDesc>  
      </category>  
      <category xml:id="c_heiferrental">  
        <catDesc>  
          <term xml:lang="en">heifer rental</term>  
        </catDesc>  
      </category>
```

```

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#">
  <skos:Concept rdf:about="https://example.com#d74fdde2944f475adc4a85e349d4ee7b">
    <skos:prefLabel>butter</skos:prefLabel>
    <skos:inScheme rdf:resource="https://example.com#Taxonomy"/>
    <skos:relatedMatch rdf:resource="wd:Q34172"/>
  </skos:Concept>
  <skos:Concept rdf:about="https://example.com#fafl131e912cbcf06e50542c3dbeb81b">
    <skos:prefLabel>dairy product</skos:prefLabel>
    <skos:inScheme rdf:resource="https://example.com#Taxonomy"/>
    <skos:relatedMatch rdf:resource="wd:Q185217"/>
    <skos:narrower rdf:resource="https://example.com#d74fdde2944f475adc4a85e349d4ee7b"/>
  </skos:Concept>
  <skos:Concept rdf:about="https://example.com#67e92c8765a9bc7fb2d335c459de9eb5">
    <skos:prefLabel>work</skos:prefLabel>
    <skos:inScheme rdf:resource="https://example.com#Taxonomy"/>
    <skos:relatedMatch rdf:resource="wd:Q6958747"/>
  </skos:Concept>
  <skos:Concept rdf:about="https://example.com#566fb919852303bc5517b6e0dcfdda7e">
    <skos:prefLabel>oxen rental</skos:prefLabel>
    <skos:inScheme rdf:resource="https://example.com#Taxonomy"/>
  </skos:Concept>
  <skos:Concept rdf:about="https://example.com#f128a7c5935262e7f3735bclb3da4684">
    <skos:prefLabel>cart rental</skos:prefLabel>
    <skos:inScheme rdf:resource="https://example.com#Taxonomy"/>
  </skos:Concept>
  <skos:Concept rdf:about="https://example.com#39dc1562e737b09b9b7439906cd14b29">
    <skos:prefLabel>waggon rental</skos:prefLabel>
    <skos:inScheme rdf:resource="https://example.com#Taxonomy"/>
  </skos:Concept>
  <skos:Concept rdf:about="https://example.com#eea0f050f1473f07661a6c4ac51d3ffc">
    <skos:prefLabel>heifer rental</skos:prefLabel>
    <skos:inScheme rdf:resource="https://example.com#Taxonomy"/>
  </skos:Concept>
  <skos:Concept rdf:about="https://example.com#ac9898e898b0be19ff47e35bcbde3e4f">
    <skos:prefLabel>renting</skos:prefLabel>
    <skos:inScheme rdf:resource="https://example.com#Taxonomy"/>
    <skos:relatedMatch rdf:resource="wd:Q157171"/>
    <skos:narrower rdf:resource="https://example.com#566fb919852303bc5517b6e0dcfdda7e"/>
    <skos:narrower rdf:resource="https://example.com#f128a7c5935262e7f3735bclb3da4684"/>
    <skos:narrower rdf:resource="https://example.com#39dc1562e737b09b9b7439906cd14b29"/>
    <skos:narrower rdf:resource="https://example.com#eea0f050f1473f07661a6c4ac51d3ffc"/>
  </skos:Concept>

```

- **renting**

BT : service
 TT : service
 NT : oxen rental
 NT : cart rental
 NT : waggon rental
 NT : heifer rental

- **rhubarb**

BT : fruit vegetable
 TT : food

- **ribbon**

BT : textile
 TT : textile

- **ribs**

BT : meat
 TT : food

- **rosin**

BT : dried fruit
 TT : food

- **rum**

BT : alcoholic
 TT : food
 NT : St. Rum
 NT : B Rum
 NT : N E Rum
 NT : best Rum

- **rye**

BT : staple food
 TT : food
 NT : spring rye

S

- **s. b. potatoes**

BT : potatoes
 TT : food

- **saleratus**

BT : unkown
 TT : unkown

- **salt beef**

BT : beef
 TT : food

<taxonomy>
→ SKOS.