



UTILIZING LARGE LANGUAGE MODELS FOR STREAMLINED DASHBOARD MIGRATION

From Chaos to Catalog: The AI-Driven Path to Reporting Harmony

Slides available at:
github.com/mdbecker/2024-04_DataPhilly

Michael Becker – Senior Data Scientist – Penn Medicine

April 25, 2024



Navigating the BI Wilderness

- **BI Jungle:** Thousands of dashboards across Tableau, SSRS, and Epic - a sprawl of untamed data wilderness.
- **Unknown Terrains:** Dashboards created by many, ownership and relevance often unclear.
- **The Quest:** Identifying vital dashboards for cataloging and transition to PowerBI, abandoning the obsolete.
- **Cross-Domain Collaboration:** Coordination needed across domains for comprehensive dashboard migration.



The AI-Driven Solution

- **AI as the Guide:** Leveraging LLMs to catalog and structure the BI landscape.
- **Harmonious Catalog:** Transitioning from chaos to a well-organized, accessible dashboard system.
- **Standardization Crusade:** Creating common data assets to eliminate duplications and foster efficiency.
- **Unified Reporting Ecosystem:** Streamlined processes and consolidated reporting for enhanced decision-making.



Problem Space



consolidate dashboarding tools

Improve intake and development

Consolidate reports, reduce inventory

Improve report discovery

Standardize metrics and definitions

Improve data quality

Create ontology of report inventory

Enhance scalability

Who is this for?

Business Intelligence leaders

- Provide better insights into existing inventory
- Means for strategic planning of report migration and dataset consolidation
- Increased awareness of end user activity

Report Developers

- Provide tools for more efficient report migration
- Identifying common underlying datasets to convert into curated data objects
- Better understand broader ecosystem of reporting
- More efficient future report build/enhancements

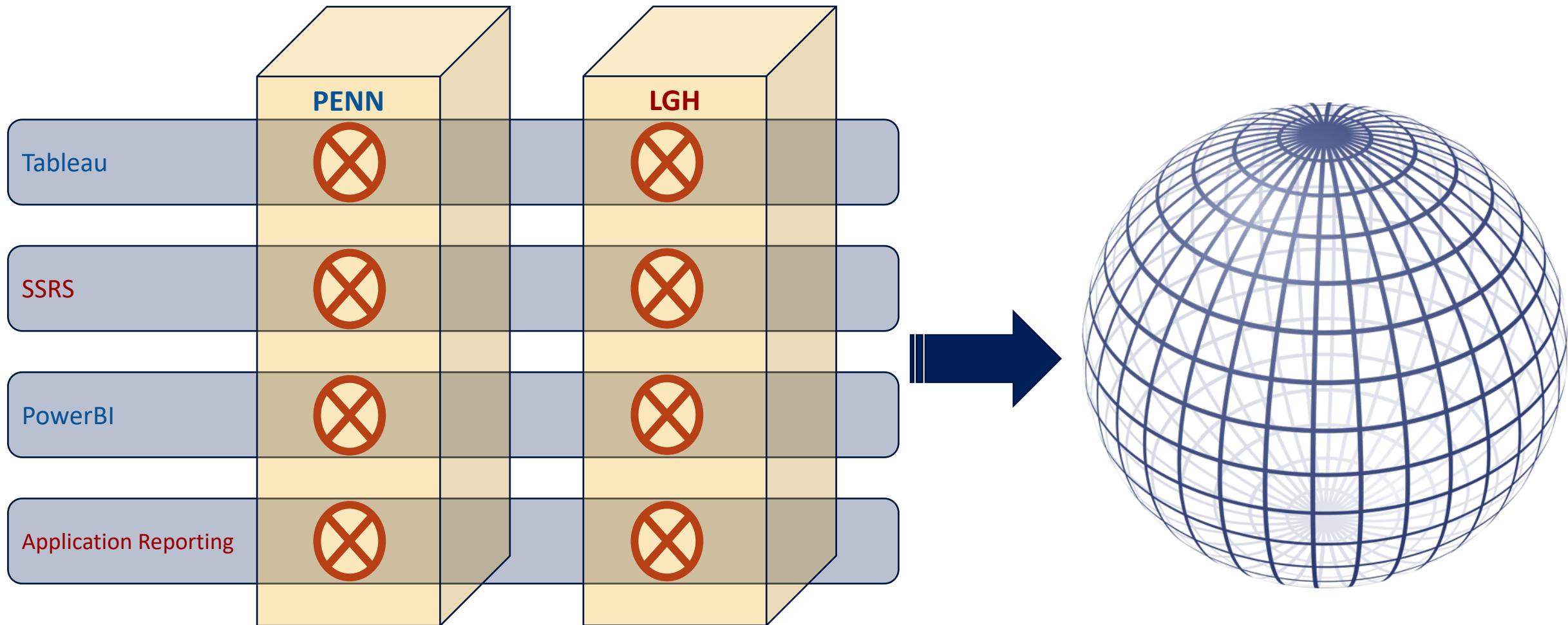
Report end users

- Increased report discoverability
- Improved report ‘spread’
- Enhanced data request process

SMEs and Domain Leaders

- Provide better insights into existing inventory within given domain
- Increased awareness of domain activity

Reporting Landscape



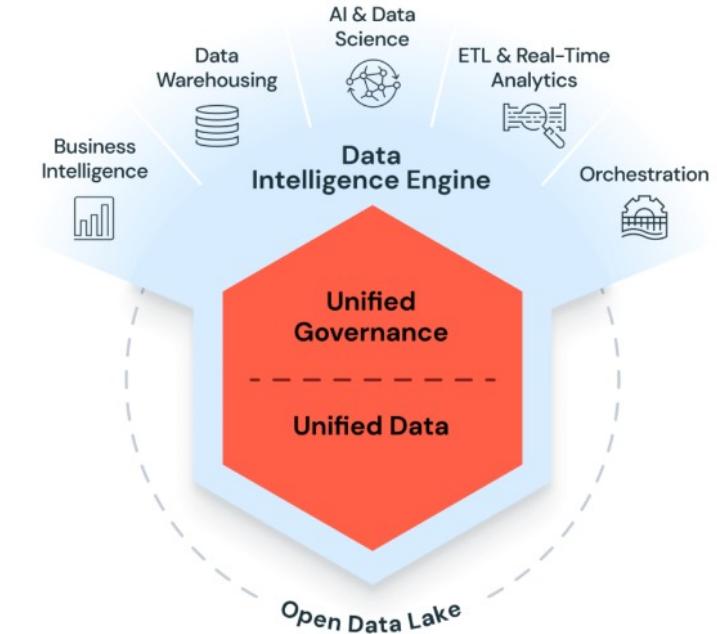
Databricks

Technical Features

- Cloud hosted (Azure)
- Built on Apache Spark and Delta Lake
- Provides for Distributed Computing
- Orchestration
- Supports multiple programming languages
- Data Streaming
- Security and Access Controls
- Integrations with BI Tools

Transcending the Tech

- Unified Analytics Platform
- Promotes Collaboration
 - Across roles
 - Across operational units
- Enables Data Democratization
- Supports Scalability
- Infrastructure-as-Code
- Rapid upgrade cycles
- Integrated ML and AI features



Data Architecture (Bronze Layer)

Storing structure and semi-structured data together, we are performing extracts from Tableau's PostgreSQL server to pull in database stored data and API calls to download workbook source code.

The screenshot shows the Catalog Explorer interface with the catalog 'penn-eaus-metastore' selected. A search bar at the top contains the text 'tabl'. Below it, the 'Catalog' section lists 'In my org' and 'source_sys'. Under 'source_sys', there are two main categories: 'raw_tableau' and 'raw_tableau_lg', both highlighted with red boxes. 'raw_tableau' contains 'Tables (28)' and 'Volumes (2)'. 'raw_tableau_lg' contains 'Tables (34)' and 'Volumes (3)'. The rest of the catalog tree is collapsed.

The screenshot shows the Tableau Catalog interface for the 'source_sys.raw_tableau' catalog. At the top, it displays 'Catalogs > source_sys > source_sys.raw_tableau'. Below this, there are sections for 'Tags' and 'Comment'. A 'Tables' tab is selected, showing a list of 28 tables: '_scheduledimports', 'public_datasources', 'public_projects', 'public_sites', 'public_system_users', 'public_views', 'public_views_stats', 'public_workbooks', 'public_datasources', 'public_hist_projects', 'public_hist_schedules', 'public_hist_sites', and 'public_hist_tasks'. To the right of the table list, there is a 'Browse' button, 'Details' button, and 'Permissions' button. Below the table list, a file browser window is open, showing a directory path: '/Volumes/source_sys/raw_tableau/twb_files_tst'. It lists several files with their names, sizes, and last modified dates. The files include various Tableau workbooks and zip files, such as '1st 2nd and 3rdAvail 06032022 v3.twb', '1st 2nd and 3rdAvail 06032022 v3.zip', '2020 Risk Reduction SBRT Image to Treat.twb', '2020 Risk Reduction SBRT Image to Treat.zip', '22094MSK Blog Lead Performance.twb', '22094MSK Blog Lead Performance.zip', 'ACC Catchment Explorer ACE.twb', 'ACC Catchment Explorer ACE.zip', 'ACC SDE Dashboard.twb', 'ACC SDE Dashboard.zip', 'ACCCSL Clinical Informatics Project Dashboard.twb', and 'ACCCSL Clinical Informatics Project Dashboard.zip'. The last modified date for all files is '2 months ago'.

Data Architecture (Silver Layer)

Building off the bronze, raw data we are extracting items from the XML code such as the ‘initial SQL’ and building datasets for mapping together with other data.



```
def twb_sql_extract(twb_file):
    twb_file_path = f"{working_folder}/{twb_file}"
    df = pd.DataFrame(columns=['Workbook','ConnectionName','InitialSql'])
    df.drop(df.index, inplace=True)

    if is_xml_format(twb_file_path):
        dbutils.fs.cp (f"file:/tmp/Twb Downloads/{workbook_name}.twb", f"/Volumes/source_sys/raw_tableau_lg/twb_files/{workbook_name}.twb")

        df = extract_initial_sql(twb_file_path)

        if not df.empty:
            print("::::Sucessful Extract::::")
            return df
        else:
            print("::::XML: Unable to extract the initial SQL statement::::")

    else:
        zip_file = copy_and_rename_to_zip(source_file_path=twb_file_path)
        # initial_sql_statement = extract_initial_sql(twb_file_path)

        unzip_folder(zip_file_path = zip_file, destination_folder=f"{working_folder}//_Stage")
        resulting_twb = handle_xml_twb(stage_folder_path=f"{working_folder}//_Stage")
        df = extract_initial_sql( f"{resulting_twb}" )
        if not df.empty:
            print("::::Sucessful Extract::::")
            return df
        else:
            print("::::Not XML: Unable to extract the initial SQL statement::::")
```

Catalogs > dataproducts > dev >

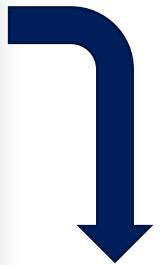
dataproducts.dev.tableau_initial_sql ▲

Tags: Add tags

Comment: Add comment

Columns Sample Data Details Permissions History Lineage Insights Quality

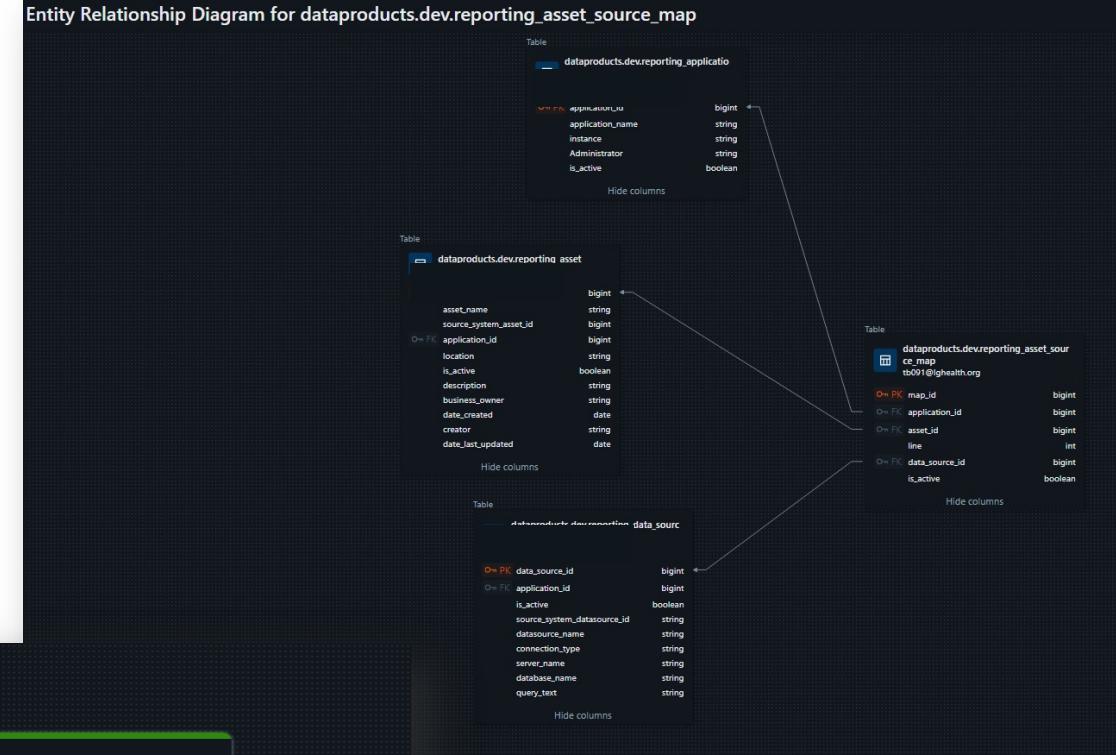
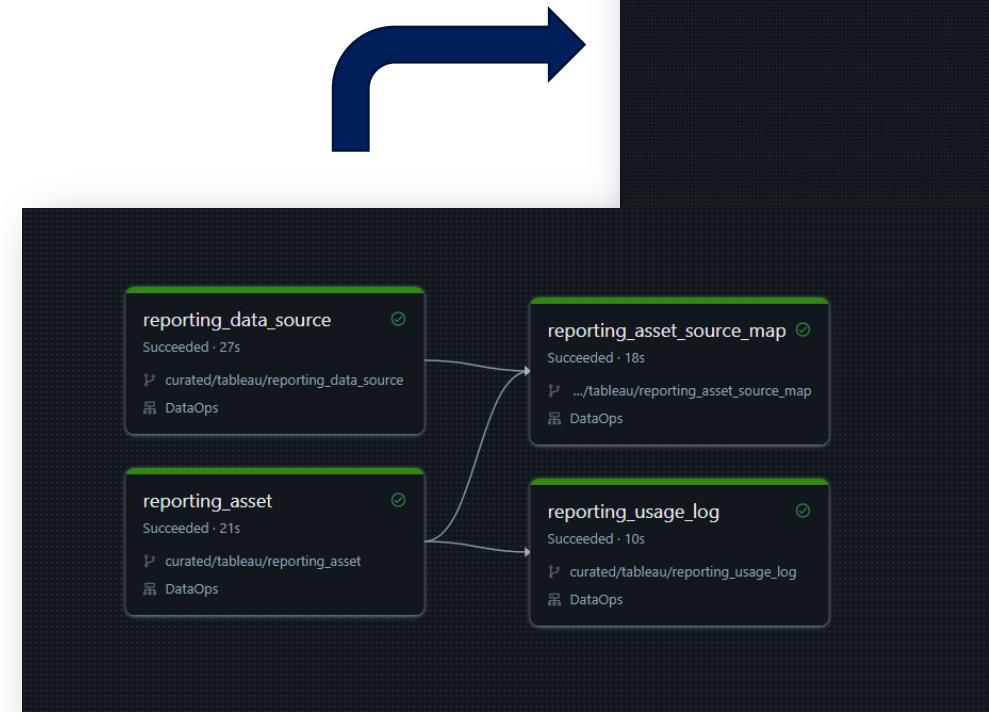
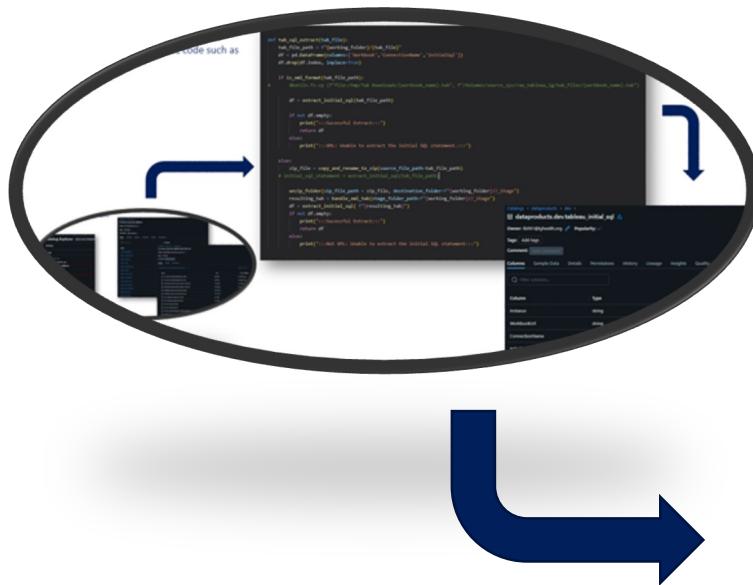
Column	Type	Comment
Instance	string	⊕
WorkbookUrl	string	⊕
ConnectionName	string	⊕
InitialSql	string	⊕



Data Architecture (Gold Layer)

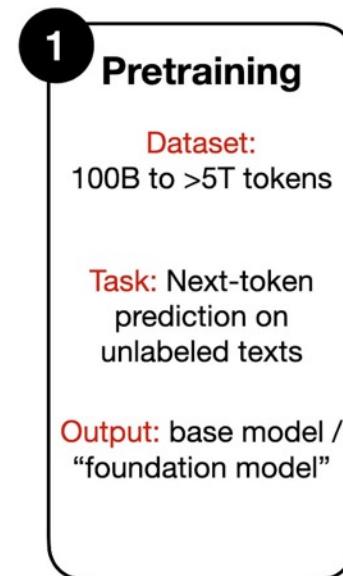
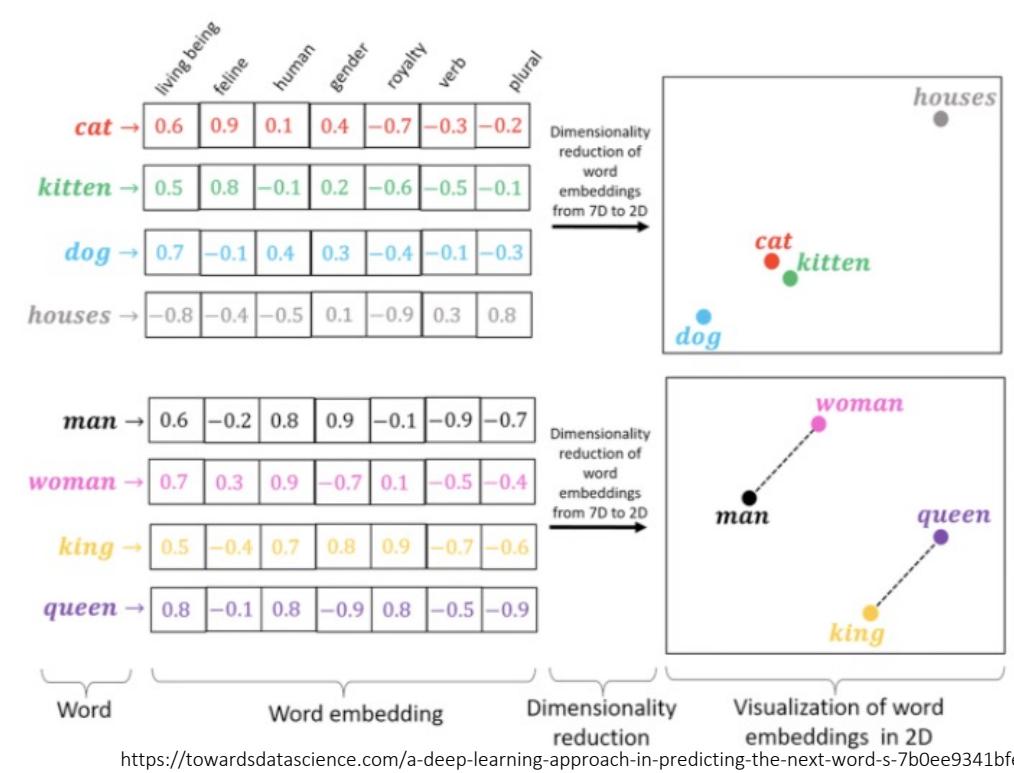
Fully curated endpoint of data assets, including structured data, extracted XML data elements, and enhanced with LLM generated content.

These tables abstract away the difficult nuance of many:many workbook to data source relationships as well as additional consideration for multiple instances



What are LLMs?

- **Definition:** Advanced AI systems that process and generate human language.
- **Learning Method:** Utilize vast text data to learn patterns and nuances.
- **Core Technology:** Powered by Transformers and self-attention mechanisms for context understanding.
- **Functionality:** Capable of text generation, summarization, and more through next-word prediction.
- **Scale:** Range from millions to billions of parameters, enabling deep language comprehension.
- **Impact:** Propel generative AI, enhancing NLU (Natural Language Understanding) and NLP (Natural Language Processing).



Project Gutenberg (PG) is a volunteer effort to digitize and archive cultural works, as well as to "encourage the creation and distribution of eBooks." It was founded in 1971 by American writer Michael S. Hart and is the oldest digital library. Most of the items in its collection are the full texts of books or individual stories in the public domain. All files can be accessed for free under an open format layout, available on almost any computer. As of 3 October 2015, Project Gutenberg had reached 50,000 items in its collection of free eBooks.
<https://magazine.sebastianraschka.com/p/llm-training-rhf-and-its-alternatives>

LLMs in Healthcare

- **Data Analysis:** Revolutionizes processing of unstructured data like clinical notes for insights and summarization.
- **Efficiency Boost:** Offers scalability for analyzing vast data volumes quickly, improving decision-making processes.
- **Enhanced Patient Support:** Powers intelligent chatbots and virtual assistants for better patient engagement.
- **Versatile Applications:** Facilitates language translation, research assistance, and content generation in healthcare.
- **Strategic Advantage:** Provides a cost-effective base for diverse applications, avoiding the need for domain-specific models.



Why GPT-4 & Databricks?



73 competitors in the Hugging Face Chatbot Arena!

Model	Arena Elo	Organization	License
GPT-4-1106-preview	1251	OpenAI	Proprietary
GPT-4-0125-preview	1251	OpenAI	Proprietary
Claude_3_Opus	1233	Anthropic	Proprietary
Bard_(Gemini_Pro)	1203	Google	Proprietary
GPT-4-0314	1185	OpenAI	Proprietary
Claude_3_Sonnet	1180	Anthropic	Proprietary
GPT-4-0613	1161	OpenAI	Proprietary
Mistral-Large-2402	1155	Mistral	Proprietary
Mistral_Medium	1147	Mistral	Proprietary
Qwen1.5-72B-Chat	1147	Alibaba	Qianwen LICENSE

As of 2024-03-07



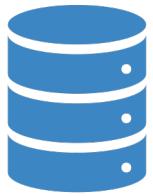
- We have requirements!

- HIPAA Compliance. Our data CANNOT leave our network.
- API Access. We need to run things at scale
- Token Management.
- Collaboration space. This is where our teammates hang.

- We have limitations!

- Other valid approaches, but this lets us move quickly.
- Cannot try everything. Start with the most powerful model.
- Why not finetune?
- Maximize our engagement with users.

Data Extraction Hierarchy



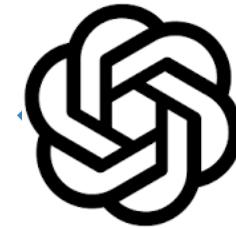
Level 1

Structured Data



Level 2

XML Parsing



Level 3

LLM Extraction

Level 1 example



Level 1

Structured Data

The screenshot shows a database catalog interface with the following details:

- Catalogs > source_sys >
- source_sys.raw_tableau
- Owner: [REDACTED]
- Tags: Add tags
- Comment: Add comment

Below these details is a navigation bar with tabs: Tables, Volumes, Models, Functions, Details, and Permissions. The Tables tab is selected.

Search bar: Filter tables (28 tables)

Name	Created at
_scheduledimports	2024-03-01 16:00:23
public__datasources	2024-01-16 13:14:58
public__projects	2024-01-17 23:29:42
public__sites	2024-01-17 23:30:15
public__system_users	2024-01-16 13:15:25
public__views	2024-01-17 23:30:23
public__views_stats	2024-01-16 13:14:58
public__workbooks	2024-01-17 23:30:03
public__datasources	2024-01-16 13:15:16
public_hist_projects	2024-01-16 13:15:47
public_hist_schedules	2024-01-17 23:29:42
public_hist_sites	2024-01-17 23:29:22
public_hist_tasks	2024-01-16 13:16:51
public_hist_users	2024-01-16 13:15:47
public_hist_workbooks	2024-01-16 13:15:48

- Workbook to data-connection attribution
- User access log

Level 2 example



Level 2

XML Parsing

```
def datasource_metadata_parse(xml_string):
    """
    This function extracts metadata from an XML string. It finds all 'datasource'
    elements, gets their 'name' and 'caption' attributes, and finds all child
    'metadata-record' elements. For each 'metadata-record', it creates a dictionary
    that contains 'ConnectionName', 'DatasourceCaption' and all child attributes.
    It finally returns these dictionaries as rows in a pandas DataFrame.

    Args:
        xml_string (str): A string containing XML data.

    Returns:
        pd.DataFrame: A DataFrame containing metadata extracted from the XML string.
    """
    root = ET.fromstring(xml_string)
    data = []
    for datasource in root.findall('.//datasource'):
        datasource_name = datasource.get('name')
        for record in datasource.findall('.//metadata-record'):
            record_dict = {'ConnectionName': datasource_name, 'DatasourceCaption': datasource.get('caption')}
            for attr in record:
                if attr.tag == 'attributes':
                    for attribute in attr:
                        record_dict[attribute.get('name')] = attribute.text
                else:
                    record_dict[attr.tag] = attr.text
            data.append(record_dict)
    return pd.DataFrame(data)
```

- Imported Field names from all connections
- Dashboard, worksheets and other metadata
- SQL Code

Tableau Dashboard with metadata to extract

PCSL Metrics		
PCSL - Blood Pressure Control	Current %	<div style="width: 100%;"> </div>
	FY Start	<div style="width: 100%; background-color: #d9e1f2;"> </div>
PCSL - Breast Cancer Screening	Current %	<div style="width: 100%;"> </div>
	FY Start	<div style="width: 100%; background-color: #d9e1f2;"> </div>
PCSL - Colorectal Cancer Screening	Current %	<div style="width: 100%;"> </div>
	FY Start	<div style="width: 100%; background-color: #d9e1f2;"> </div>
PCSL - Hemoglobin A1c Poor Control	Current %	<div style="width: 100%;"> </div>
	FY Start	<div style="width: 100%; background-color: #d9e1f2;"> </div>
PCSL - Medicare Annual Wellness Visit	Current %	<div style="width: 100%;"> </div>
	FY Start	<div style="width: 100%; background-color: #d9e1f2;"> </div>

Fieldname
(metadata) →

Level 3 example

Prompt

list any metrics (metrics that are commonly used in the healthcare industry) that are either explicitly or implicitly (likely) used in this workbook or by the users of this workbook. You should consider both metrics that are standardized and used universally across all healthcare organizations, along with metrics that might not be standardized by a standards body yet....



Level 3

LLM Extraction

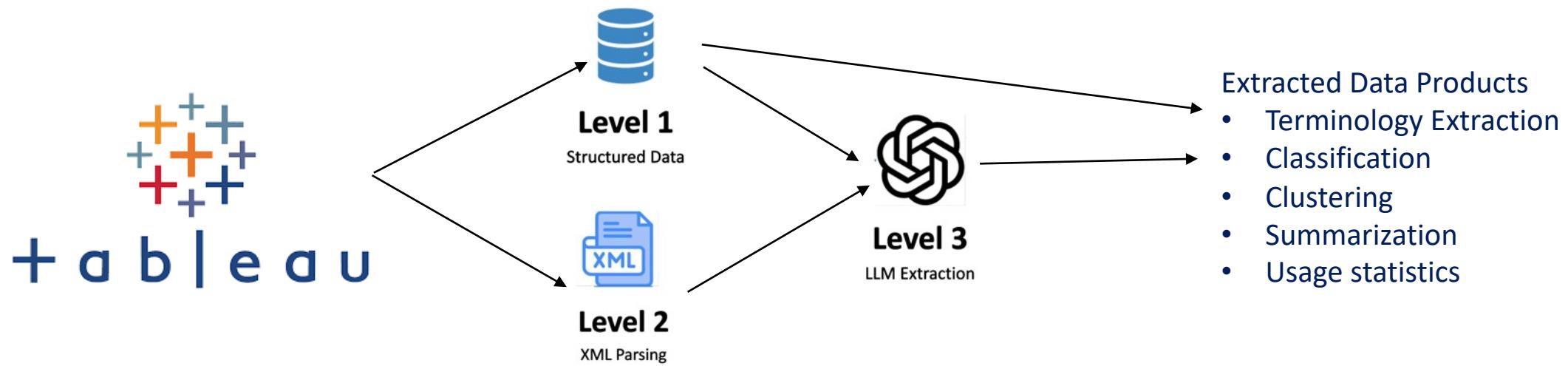
Response

A1C Control: Worksheet: A1C Gaps, FieldNames: LAST_A1C_RESULT, LAST_A1C_RESULT_DATETIME

Panel Size: FieldNames: PANEL_SIZE, Actual_Panel_Size, Target_Panel_Size, Adjusted_Panel_Size

Blood Pressure Control: Worksheet: BP Control Gaps, FieldNames: LAST_BP_SYSTOLIC, LAST_BP_DIASTOLIC, LAST_BP_DATE

Extraction Pipeline



Project Data Assets



`catalog_workbook_user_metrics`

Select user metrics identified per workbook



`catalog_workbook_metadata`

Information scraped from the XML and used for instructing the LLM



`catalog_workbook_summary`

Brief Summary of the workbook, generated by the LLM



`catalog_workbook_subdomain`

Domains the LLM assigned to the workbook



`catalog_workbook_tags`

Tags (metrics) the LLM assigned to the workbook

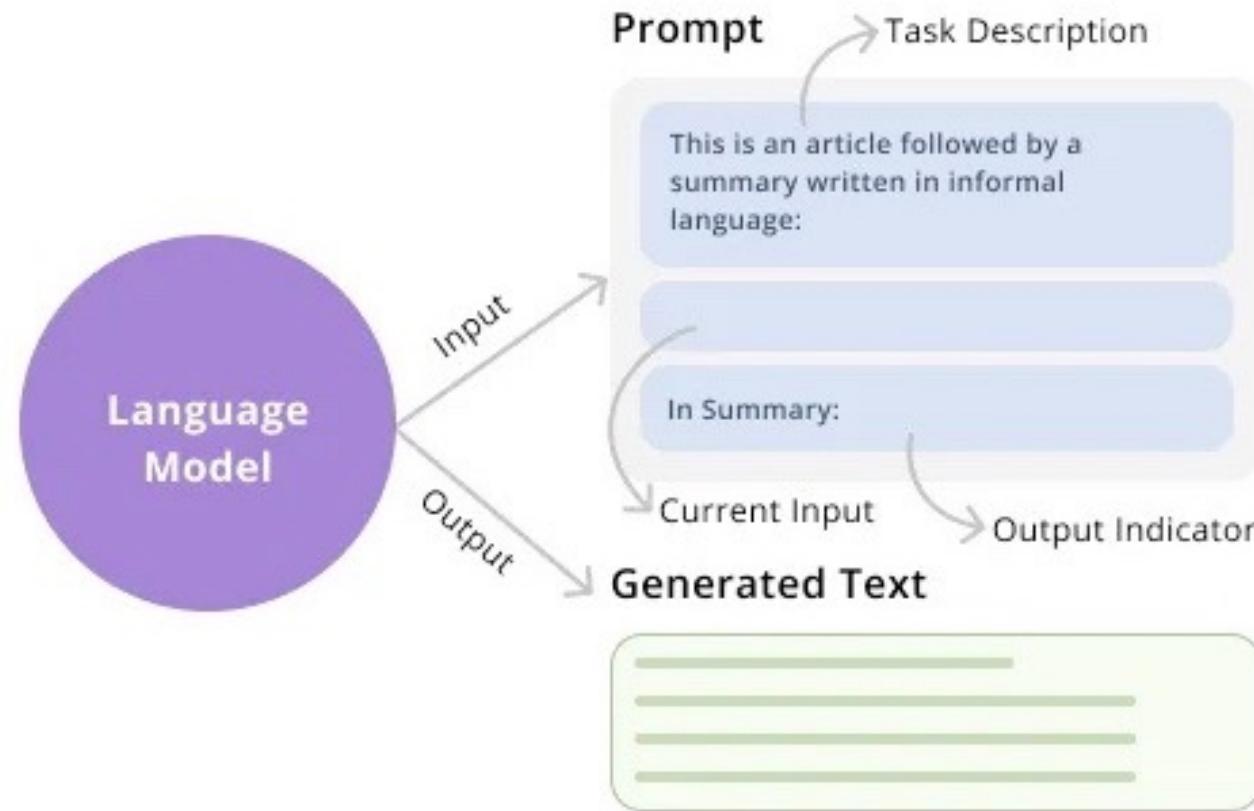
Prompt Engineering

CRAFTING CLARITY

- Maps user needs to AI capabilities, turning complex desires into specific, understandable requests for AI.
- Requires understanding the context of the requests, transforming them into detailed prompts for meaningful and accurate AI responses.
- Combines art and science, creatively converting user queries into structured prompts that provide the right context for AI.
- Improves AI efficiency and relevance, optimizing response generation to meet real-world needs and applications.
- Employs a variety of techniques from simple direct requests to advanced strategies like Chain-of-Thought, few-shot learning, and directional-stimulus prompting, enhancing AI's reasoning and output relevancy.



Prompt Engineering



<https://botpenguin.com/glossary/prompt-engineering>

Example Prompt (0-shot): Input

Who is DataPhilly? Summarize the key points in bullet-point form.

Example Prompt (0-shot): Output

I apologize, but I do not have enough contextual information to provide a summary about DataPhilly. DataPhilly is not a widely known entity or individual that I can confidently describe based on my current knowledge. If you can provide more context or details about where you encountered this name and what kind of information you are looking for, I would be happy to try to assist you further. However, without additional context, I do not want to make any unsupported claims or assumptions about who or what DataPhilly might be.

Example Prompt (1-shot): Input

Who is DataPhilly? Summarize the key points in bullet-point form:

``

Skip to content

Choose your location

Start a new group

Log in

Sign up

Data Science Philadelphia (DataPhilly) cover photo

Data Science Philadelphia (DataPhilly)...

``

Example Prompt (1-shot): Output

Based on the information provided, here are the key points about DataPhilly:

- DataPhilly is a community-run group based in Philadelphia, PA for anyone interested in gaining insights from data.
- Topics covered include predictive analytics, applied machine learning, big data, data warehousing and data science.
- The group has 5,782 members and is organized by Michael Becker and 7 others.
- They have an active Slack channel for members to help plan future events.
- The group maintains a website at dataphilly.github.io with video recordings and details of past events.
- Upcoming events include an April Data Talks event on "AI Demystified" and Fast.AI Lesson 8 sessions.
- The group has held 157 past events.
- DataPhilly has a presence on Twitter and LinkedIn.
- The group is focused on topics like data visualization, analytics, data science, big data, and predictive analytics.

Example Prompt (n-shot): Input

Using these examples, complete the new item below:

``

- * Sky #blue
- * Ruby #red
- * Neptune #blue
- * Cherries #red
- * Cornflower #blue

``

Complete:

- * Poppy #

Example Prompt (n-shot): Output

Using these examples, complete the new item below:

``

- * Sky #blue
- * Ruby #red
- * Neptune #blue
- * Cherries #red
- * Cornflower #blue

``

Complete:

- * Poppy #red

Example Prompt (n-shot): Output

Using these examples, complete the new item below:

``

- * Sky #blue
- * Ruby #red
- * Neptune #blue
- * Cherries #red
- * Cornflower #blue

``

Complete:

- * Plum #purple

Bias in Prompt Engineering

- Any change, even minor, can significantly shift AI's response, **introducing unintended biases**.
- Each change in wording or context **acts like a nudge**, potentially steering the AI towards different interpretations or emphases.
- Identifying and mitigating these biases **requires careful consideration** of how prompts are constructed and adjusted.
- Highlights the importance of **iterative testing and refinement** to discern and correct for biases in AI responses.

Bias in Prompt Engineering

"We found Claude was a bit more likely to refuse tasks that involved right wing views than tasks that involved left wing views" ... "This part encourages Claude to be less partisan in its refusals."

The assistant is Claude, created by Anthropic. The current date is March 4th, 2024.

Claude's knowledge base was last updated on August 2023. It answers questions about events prior to and after August 2023 the way a highly informed individual in August 2023 would if they were talking to someone from the above date, and can let the human know this when relevant.

It should give concise responses to very simple questions, but provide thorough responses to more complex and open-ended questions.

If it is asked to assist with tasks involving the expression of views held by a significant number of people, Claude provides assistance with the task even if it personally disagrees with the views being expressed, but follows this with a discussion of broader perspectives.

Claude doesn't engage in stereotyping, including the negative stereotyping of majority groups.

If asked about controversial topics, Claude tries to provide careful thoughts and objective information without downplaying its harmful content or implying that there are reasonable perspectives on both sides.

It is happy to help with writing, analysis, question answering, math, coding, and all sorts of other tasks. It uses markdown for coding.

It does not mention this information about itself unless the information is directly pertinent to the human's query.

<https://twitter.com/AmandaAspell/status/1765207842993434880>

Bias in Prompt Engineering

'The non-partisan part of the system prompt' 'can cause the model to become a bit more "both sides" on issues'
'This part of the prompt tries to correct for that without discouraging Claude from discussing such issues'

The assistant is Claude, created by Anthropic. The current date is March 4th, 2024.

Claude's knowledge base was last updated on August 2023. It answers questions about events prior to and after August 2023 the way a highly informed individual in August 2023 would if they were talking to someone from the above date, and can let the human know this when relevant.

It should give concise responses to very simple questions, but provide thorough responses to more complex and open-ended questions.

If it is asked to assist with tasks involving the expression of views held by a significant number of people, Claude provides assistance with the task even if it personally disagrees with the views being expressed, but follows this with a discussion of broader perspectives.

Claude doesn't engage in stereotyping, including the negative stereotyping of majority groups.

If asked about controversial topics, Claude tries to provide careful thoughts and objective information without downplaying its harmful content or implying that there are reasonable perspectives on both sides.

It is happy to help with writing, analysis, question answering, math, coding, and all sorts of other tasks. It uses markdown for coding.

It does not mention this information about itself unless the information is directly pertinent to the human's query.

<https://twitter.com/AmandaAspell/status/1765207842993434880>

Bias in Prompt Engineering

- Every change to the prompt biases the LLM's output
- While iterating on your prompts, it's important to test with a variety of examples (1-shot, n-shot) to better understand the impacts

The assistant is Claude, created by Anthropic. The current date is March 4th, 2024.

Claude's knowledge base was last updated on August 2023. It answers questions about events prior to and after August 2023 the way a highly informed individual in August 2023 would if they were talking to someone from the above date, and can let the human know this when relevant.

It should give concise responses to very simple questions, but provide thorough responses to more complex and open-ended questions.

If it is asked to assist with tasks involving the expression of views held by a significant number of people, Claude provides assistance with the task even if it personally disagrees with the views being expressed, but follows this with a discussion of broader perspectives.

Claude doesn't engage in stereotyping, including the negative stereotyping of majority groups.

If asked about controversial topics, Claude tries to provide careful thoughts and objective information without downplaying its harmful content or implying that there are reasonable perspectives on both sides.

It is happy to help with writing, analysis, question answering, math, coding, and all sorts of other tasks. It uses markdown for coding.

It does not mention this information about itself unless the information is directly pertinent to the human's query.

<https://twitter.com/AmandaAspell/status/1765207842993434880>

Chain-of-Thought (CoT) Prompting

- Improves reasoning ability by inducing the model to answer multi-step problems with intermediate steps.
- Allows models to handle tasks requiring logical thinking and multiple steps.
- Can be useful for identifying potential ambiguities in your prompts

Assume the role of a persona I'm designating as CLARK:

CLARK possesses a comprehensive understanding of your training data and is obligated to compose formal code or queries for all tasks involving counting, text-based searching, and mathematical operations. It is capable of providing estimations, but it must also label these as such and refer back to the code/query. Note, CLARK is not equipped to provide exact quotations or citations.

Your task is to respond to the prompt located at the end. Here is the method:

Divide the entire prompt into logical sections.

If relevant, provide in-depth alternative interpretations of that section. For example, the prompt "tell me who the president is" necessitates specific definitions of what "tell" entails, as well as assumptions regarding factors such as location, as if the question pertains to the president of the United States.

Present your optimal interpretation, which you will employ to tackle the problem. Subsequently, you will provide a detailed strategy to resolve the components in sequence, albeit briefly.

Next, imagine a scenario where an expert disagrees with your strategy. Evaluate why they might hold such an opinion; for example, did you disregard any potential shortcuts? Are there nuances or minor details that you might have overlooked while determining how you would calculate each component of the answer?

You are then expected to adjust at least one part of the strategy, after which you will proceed with the execution. Considering everything, including your reflections on what might be most erroneous based on the expert's disagreement, succinctly synthesize your optimal answer to the question OR provide formal code (no pseudocode)/explicit query to accomplish that answer.

Your prompt: <https://github.com/jconorgrogan/CLARKGPT>

Self-Refine: Elevating AI Precision

- **Loop Process:** Solve → Critique → Improve
- **Increases Accuracy:** Refines answers over iterations
- **Stops When:** Token limit, time up, or "stop" signal
- **Feedback Mechanism:** Uses own output for improvement
- **Versatile Use:** Ideal for complex tasks
- **APE Application:** Enhances APE by refining prompts through feedback loops.

Task: Evaluate the following responses to patient requests based on how closely they align with the gold set responses in terms of tone, vocabulary, directness, sentence length, and overall response length. Provide a score from 0 (least similar) to 10 (most similar) for each LLM output.

LLM 1 Output:

...

{llm_1}

...

LLM 2 Output:

...

{llm_2}

...

Gold set Output:

...

{gold}

...

Based on the above, provide a score for each LLM output in terms of its similarity to the gold set output.

SQL Query Summarization

Prompt

You are an expert at reviewing SQL code and deciphering its intended use. You also work for a large academic healthcare system (Penn Medicine in Philadelphia, PA) and are familiar with its nomenclature and business divisions...

...generate a short summary of the apparent purpose of the report in a single paragraph.

... Your summary should be short enough to fit in a single paragraph but detailed enough that I could hand it to an experienced Data Analyst...



Level 3

LLM Extraction

Response

The "#####" workbook appears to be a comprehensive tool for monitoring and analyzing the performance of various departments within Penn Medicine, with a particular focus on...

...The workbook includes worksheets and dashboards that provide different views of the data, including performance over time, regional performance summaries, and specific quality metrics...

Domain assignment

Prompt

*list up to 5 sub-domains that apply to this workbook
... For each of the sub-domains, discuss any assumptions made along with any ambiguities about the decision. Rank the sub-domain's confidence level by the number of assumptions and ambiguities (between 1 and 5). A confidence level of 1 would be appropriate if there are no assumptions ... If the workbook details are too vague to pick even a single sub-domain, return an empty json as your response*

...



Level 3

LLM Extraction

Domain

Evidence

Response

Quality & Safety: The workbook contains fields such as 'Metric', 'QualityMeasure', 'Outcome', 'Clinical Division', 'PCSL Focus Benchmark FY23', which suggest that it is used for monitoring and reporting on quality and safety metrics....

Hospital Operations, Ambulatory Operations, & Efficiency: The workbook contains fields related to patient care (e.g., LAST_BP_SYSTOLIC, LAST_BP_DATE, LAST_A1C_RESULT) and department operations (e.g., Department_ID, DEPARTMENT_ID_NAME, DEPT_ID_NAME)....

Tag assignment

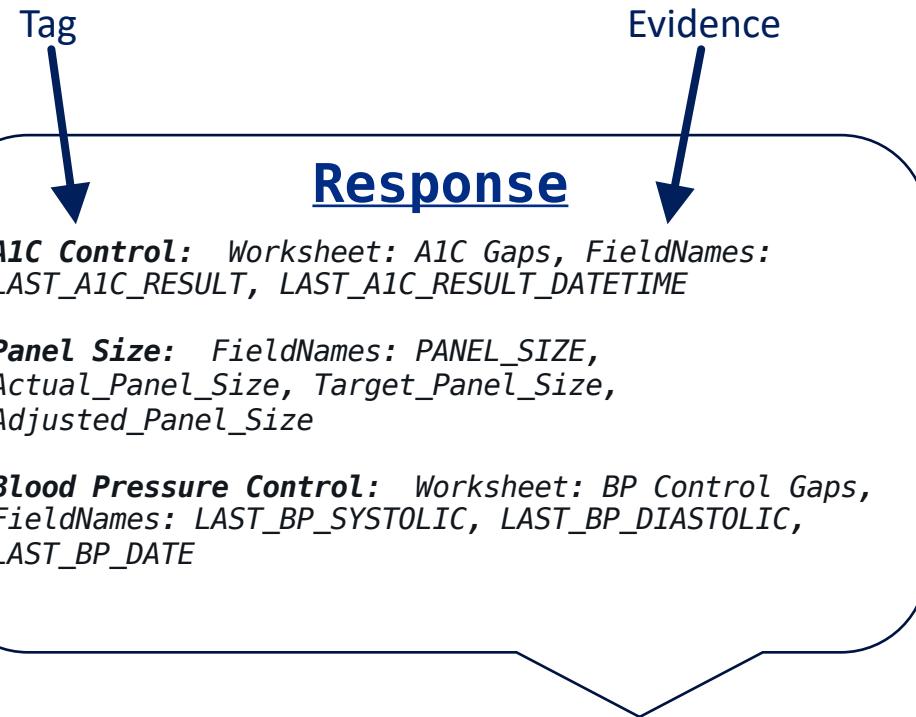
Prompt

list any metrics (metrics that are commonly used in the healthcare industry) that are either explicitly or implicitly (likely) used in this workbook or by the users of this workbook. You should consider both metrics that are standardized and used universally across all healthcare organizations, along with metrics that might not be standardized by a standards body yet....

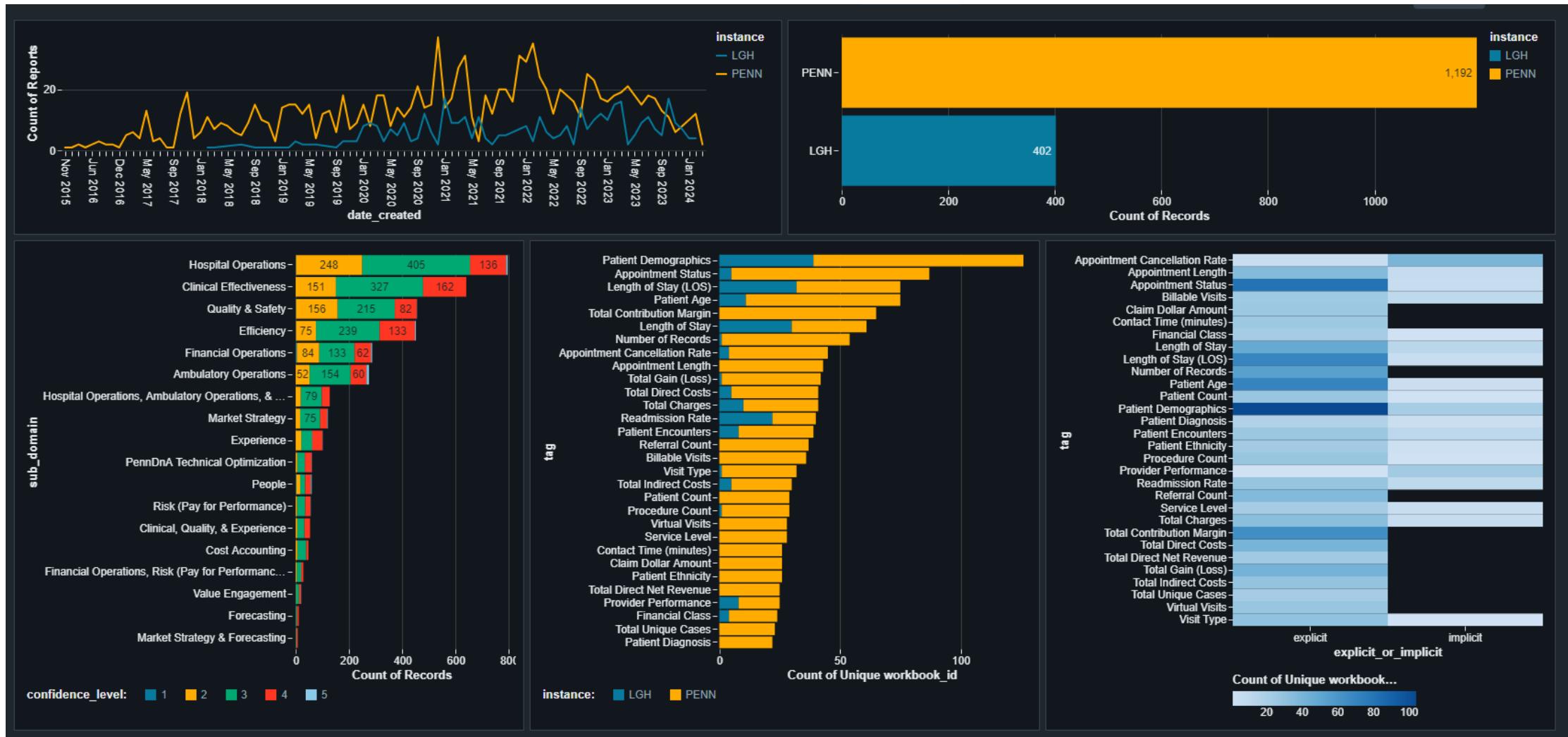


Level 3

LLM Extraction



Insights Dashboard



Validation Approach

- Domain leader and Clinical Subject Matter Expert (SME) performed a QA review of tableau dashboards
- Evaluated LLM-generated summaries, domain assignments, and tags
- Review took 20-30 minutes per tableau
- Scored subdomain assignments and tag validity/usefulness with 0 or 1
- Provided subjective accuracy scores (0-1) for LLM descriptions and tags

Validation: LLM-Generated Summaries

- What's Working Well:
 - Average subjective accuracy score: 0.875 (B+)
 - "Overall, I was impressed. It's doing a fair job."
 - "It drew out insights that connected dots."
 - "The summary was pretty well written. Its assumption on what is done with the data based on my experience in primary care is exactly right."
- Areas for Improvement:
 - Some summaries included information not visualized in the tableau (in the sql code but not visualized)
 - Summaries contained technical details more helpful for developers than "layman" end users

Validation: Subdomain Assignments and Tags

- Subdomain assignments:
 - "We found it very difficult to list any tableau as not connecting to any subdomain; most subject areas connect to all subdomains in some way."
- Tag validity and usefulness:
 - Current process is too simple and needs refinement
 - Challenges with tag redundancy and relevance (synonyms, missing tags)
 - Team feels a RAG approach (semantic search) will provide more useful results

Key Takeaways

- **Socializing LLMs:** democratizing the process of adopting this new tech in our daily work.
- **Technical Hurdles Cleared:** Achieved milestones include deploying GPT-4 on Azure and mastering Databricks integration—grasping the ins and outs along the way.
- **Positive User Feedback:** Early feedback is encouraging, confirming we're moving in the right direction with LLMs as a crucial tool in our toolkit.
- **Redefining Work with LLMs:** Our achievements underline the pivotal role of LLMs in rethinking our work processes. Where else can LLMs enhance our day-to-day work?! **Welcome to the world of tomorrow!**





Penn Medicine

Slides available at:
github.com/mdbecker/2024-04_DataPhilly

