

AI In Healthcare, ML/DL Tutorial

Evan Jones, UT ID: ej8387

Assignment source at:

https://github.com/etjones/aihc_hw7_llm_ej8387

Diabetic Response to Medications

Can we get an LLM to identify which medications improve outcome for diabetic patients?

Here's the specific issue I used in generating further prompts and analysis:

Identify correlations between specific medications and improved health outcomes for diabetic patients in the Siyeh Synthetic data set.

Dataset: Siyeh Synthetic Medical Data

I used Brenda Griffith's Siyeh Synthetic Medical Data dataset.

The screenshot shows a web browser window with the URL `https://data.world/siyeh/synthetic-medical-data`. The page is for the 'Synthetic Medical Data' dataset on the data.world platform, created by Brenda Griffith. The dataset has 737 files and a 'Launch workspace' button. The 'Overview' tab is selected, showing a description and summary of the dataset. The description states it is synthetic medical record data for 'Introduction to Biomedical Data Science'. The summary explains that the data was generated using Synthea, a synthetic patient generator, and is free from cost, privacy, and security restrictions. It also mentions that the data is de-identified and suitable for research. At the bottom, there is a list of files, with 'allergies.csv' highlighted, and a 'View' button next to it.

DESCRIPTION

Synthetic medical record data for Introduction to Biomedical Data Science.

SUMMARY

This dataset contains the core data to be used in projects for the textbook **Introduction to Biomedical Data Science** edited by Robert Hoyt MD FACP ABPM-CI, and Robert Muenchen MS PSAT (2019).

Data was generated using Synthea, a synthetic patient generator that models the medical history of synthetic patients. Their mission is to output high-quality synthetic, realistic but not real, patient data and associated health records covering every aspect of healthcare. The resulting data is free from cost, privacy, and security restrictions, enabling research with Health IT data that is otherwise legally or practically unavailable. De-identified real data still presents a challenge in the medical field because there are people who excel at re-identification of these data. For that reason the average medical center, etc. will not share their patient data. Most governmental data is at the hospital level. NHANES data is an exception.

You can read Synthea's first academic paper [here](#).

Show more ▼

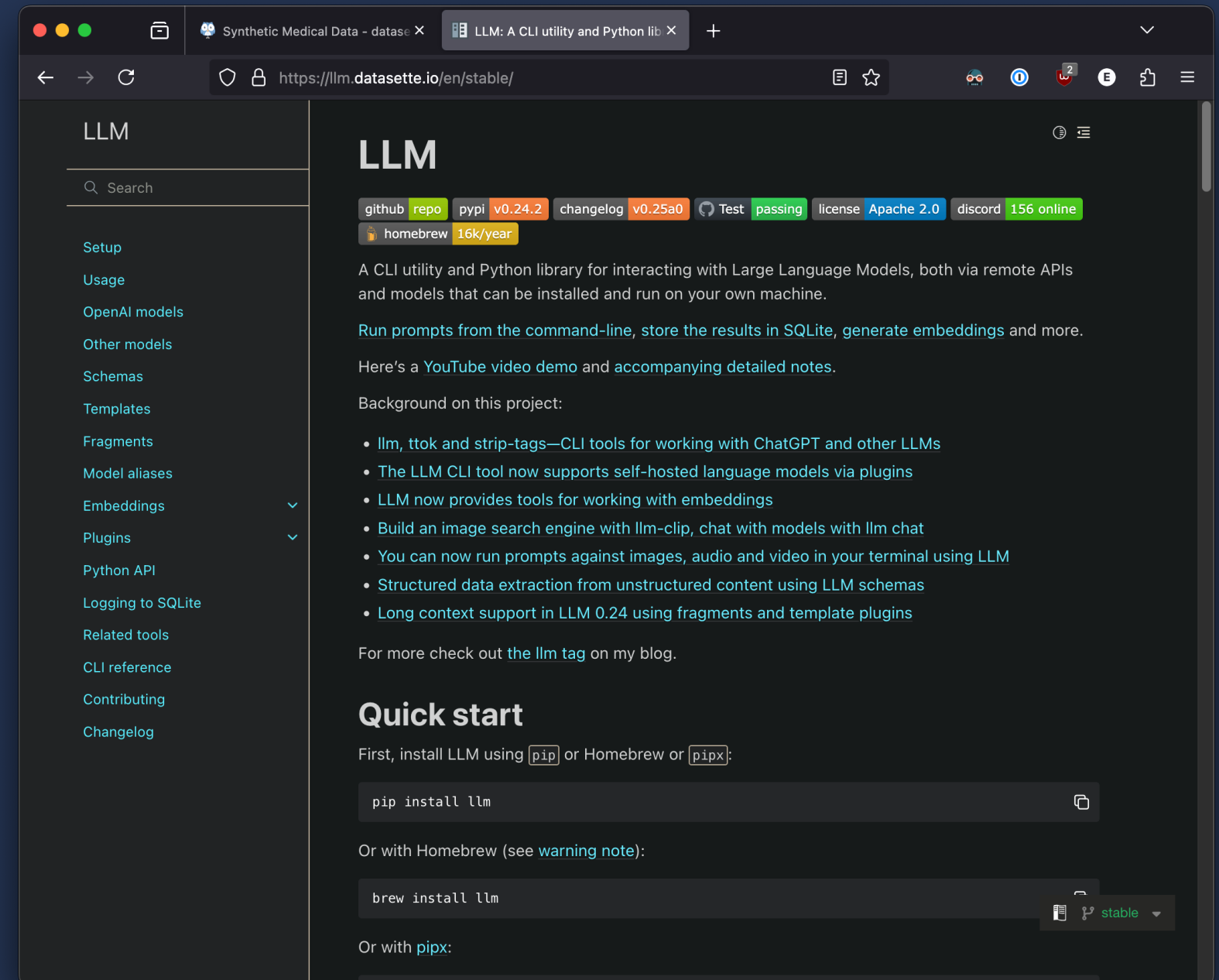
11 files

allergies.csv
Request more info

View

Toolset

- Simon Willison's brilliant llm Python package allows programmatic or command-line interactions with many LLMs
- pandas for data loading and filtering
- Google's Gemini 2.0 Flash LLM has a generous 1M token context window, great for data analysis
- Codeium's Windsurf AI-enabled IDE *radically* smooths friction for developing small-scale tools
- Deckset for Markdown -> Slide deck generation



Data Filtering: 22Mb -> 776kb

- The Siyeh dataset is 22Mb, too large to fit in LLM context windows.
- **BUT**, we don't need all that data; just the parts that describe diabetic patients
- Using Windsurf, I wrote a data-filtering script, [data_preparation.py](#) to extract only Diabetes-relevant data from the original dataset.
- [llm_data_loader.py](#) digests that data and concatenates it into a single file formatted for LLMs

```
def format_data_for_llm(
    dataset: Dict[str, pd.DataFrame], max_patients: int = 20, max_observations: int = 5
) -> str:
    """Format the processed data for LLM input.

    Args:
        dataset: Dictionary of processed dataframes
        max_patients: Maximum number of patients to include
        max_observations: Maximum number of observations per patient to include

    Returns:
        Formatted data as a string
    """
    # Start with dataset summary
    if "summary" in dataset:
        formatted_data = "# Dataset Summary\n\n"
        formatted_data += json.dumps(dataset["summary"], indent=2) + "\n\n"
    else:
        formatted_data = ""

    # Add medication effectiveness summary
    formatted_data += "# Medication Effectiveness Summary\n\n"

    outcomes_df = dataset["medication_outcomes"]

    # Focus on HbA1c outcomes
    hba1c_outcomes = outcomes_df[outcomes_df["obs_code"] == "4548-4"]
    > if not hba1c_outcomes.empty: ...

    # Add blood glucose outcomes
    glucose_outcomes = outcomes_df[outcomes_df["obs_code"].isin(["2339-0", "2345-7"])]
    > if not glucose_outcomes.empty: ...

    # Select a subset of patients for detailed examples
    patient_outcomes = outcomes_df.groupby("patient").size().reset_index(name="count")
    patient_outcomes = patient_outcomes.sort_values("count", ascending=False)
    selected_patients = patient_outcomes.head(max_patients)["patient"].tolist()

    # Add patient examples
    formatted_data += "# Patient Examples\n\n"

    > for patient_id in selected_patients: ...

    return formatted_data
```

Tool Usage

- You know who does a better job writing specific, nuanced LLM prompts than people do?

LLMs.

All prompts generated by Claude 3.7 Sonnet.

- llm has a very handy --fragments flag that lets us send entire files to an LLM, while efficiently storing all queries and responses in a local database. It's magic!

```
llm -m gemini-2.0-flash -f data/llm_input_data.txt -f prompts/zero_shot_prompt.md > responses/zero_shot_response_1.md
```

```
llm -m gemini-2.0-flash -f data/llm_input_data.txt -f prompts/few_shot_prompt.md > responses/few_shot_response_1.md
```

```
llm -m gemini-2.0-flash -f data/llm_input_data.txt -f prompts/chain_of_thought_prompt.md > responses/chain_of_thought_response_1.md
```

Zero-shot prompt

Analyze the provided Siyeh Synthetic medical dataset in llm_input_data.txt to identify correlations between specific medications and improved health outcomes for diabetic patients.

The dataset contains:

- A summary of the dataset with statistics on diabetic patients, conditions, medications, and observations
- Medication effectiveness summaries for HbA1c and blood glucose outcomes
- Detailed patient examples with demographics, medication histories, and health measurements

Focus on:

1. Analyzing the medication effectiveness summaries to identify which medications show the strongest correlations with improved outcomes
2. Examining the patient examples to understand individual responses to medications
3. Identifying patterns in health metric changes (HbA1c, blood glucose) following medication use
4. Determining which medications appear most effective across multiple patients
5. Considering potential confounding factors like patient demographics and comorbidities

Provide a detailed analysis of which medications appear most effective for diabetic patients in this dataset and explain your reasoning.

Zero-shot results

(Complete response [here](#) (100 lines))

Summary:

1. Insulin formulations and metformin demonstrate the strongest evidence for glycemic control.

2. Semaglutide shows promising results but has limited sample size.

3. Effectiveness varies by patient demographics and diabetes severity.

Okay, I will analyze the provided Siyeh Synthetic medical dataset to identify medications that show the strongest correlations with improved health outcomes for diabetic patients, considering the specified factors.

1. Analysis of Medication Effectiveness Summaries:

- **HbA1c Outcomes:** The table reveals medications with the largest *mean change* reduction in HbA1c. Here are some notable ones:
 - **insulin** human isophane **70** UNT/ML / Regular Insulin Human **30** UNT/ML Injectable Suspension [**Humulin**]: Shows a significant mean change of -2.98 and a percentage change of -38.40%. This makes sense, as insulin is a direct intervention for lowering blood sugar and HbA1c.
 - **24** HR Metformin hydrochloride **500** MG Extended Release Oral Tablet: Shows a substantial mean change of -1.35 (-15.31%). Metformin is a first-line oral medication for type 2 diabetes.
 - **3** ML liraglutide **6** MG/ML **Pen** Injector: Shows a mean change of -1.06 with a % change of -9.98%. Liraglutide is a GLP-1 receptor agonist, known for its HbA1c-lowering effects.
 - canagliflozin 100 MG Oral Tablet shows a decent Mean Change (-1.23) and Mean % Change (-13.01%).
- **Blood Glucose Outcomes:**
 - **Insulin** Lispro **100** UNT/ML Injectable Solution [**Humalog**]: shows a good Mean Change (-47.00) with a % change of -35.07%
 - **Atorvastatin** **80** MG Oral Tablet and **Captopril** **25** MG Oral Tablet show a significant mean change of -26.00, however, there is only one observation so the results may not be very accurate or representative.
 - **Donepezil** hydrochloride **10** MG / Memantine hydrochloride **28** MG [**Namzaric**]: shows a substantial mean change of -18.00 (-16.66%). Donepezil/ Memantine are used to treat Alzheimer's, these results may be spurious or the patient has other issues that influence blood sugar and HbA1c

It's important to note some medications show a *positive* mean change in glucose, which could be undesirable for diabetic patients. These should be investigated further.

2. Examining Patient Examples:

Patient: 6e09f851

- This patient's history shows they are on Amlodipine, Clopidogrel, Simvastatin, Nitroglycerin, Alendronic acid, Acetaminophen/Hydrocodone, and Naproxen.
- The table shows an increase in glucose after the use of Amlodipine, Simvastatin and Alendronic acid.

Few-shot prompt

Task: Analyze correlations between medications and improved health outcomes for diabetic patients in the Siyeh Synthetic dataset provided in llm_input_data.txt.

Example 1:

Input: Patient with Type 2 Diabetes prescribed Metformin 500mg

- Before medication: HbA1c 8.2%, Fasting glucose 165 mg/dL
- After 6 months: HbA1c 7.1%, Fasting glucose 130 mg/dL

Analysis: This patient showed significant improvement in glycemic control after Metformin initiation, with HbA1c reduction of 1.1% and fasting glucose reduction of 35 mg/dL.

Example 2:

Input: Patient with Type 1 Diabetes prescribed Insulin Glargine

- Before medication: HbA1c 9.3%, Fasting glucose 210 mg/dL
- After 6 months: HbA1c 7.4%, Fasting glucose 140 mg/dL

Analysis: This patient demonstrated substantial improvement with Insulin Glargine, achieving a 1.9% reduction in HbA1c and 70 mg/dL reduction in fasting glucose.

Example 3:

Input: Patient with Type 2 Diabetes prescribed Sitagliptin

- Before medication: HbA1c 7.8%, Fasting glucose 145 mg/dL
- After 6 months: HbA1c 7.6%, Fasting glucose 140 mg/dL

Analysis: This patient showed minimal improvement with Sitagliptin, suggesting it may be less effective for this particular case.

Now, analyze the Siyeh Synthetic dataset in llm_input_data.txt to identify which medications correlate with the best outcomes for diabetic patients. The file contains:

- Medication effectiveness summaries showing mean changes in HbA1c and blood glucose
- Patient examples with detailed medication histories and health outcomes
- Timelines showing health metrics before and after medication administration

Consider factors like:

1. Changes in HbA1c and blood glucose levels as shown in the medication effectiveness summaries
2. Consistency of results across multiple patients
3. Magnitude of improvements in health metrics
4. Time between medication start and observed improvements
5. Patient demographics that might influence outcomes

Provide a comprehensive analysis of medication effectiveness for diabetic patients in this dataset.

Few-shot results

See [few_shot_response_1.md](#) for complete response (62 lines) .

Summary:

1. Insulin formulations most effectively reduce HbA1c, while metformin shows strong, consistent glycemic control.

2. Combination therapies appear more effective than monotherapy for poorly controlled diabetes.

3. Medication effectiveness varies based on patient characteristics and diabetes severity.

Okay, I will analyze the provided Siyeh Synthetic dataset to identify medications that correlate with improved health outcomes for diabetic patients. My analysis will consider the factors you outlined, focusing on HbA1c and blood glucose level changes, consistency, magnitude, timing, and potential demographic influences.

Analysis of Medication Effectiveness

Based on the provided data, here's a summary of my findings:

1. Medications with Notable HbA1c Reduction:

- **24 HR Metformin hydrochloride 500 MG Extended Release Oral Tablet:** Shows a substantial mean HbA1c reduction of -1.35, with a mean % change of -15.31%. This is a strong indicator of effectiveness in a notable sample size (n=17).
- **3 ML liraglutide 6 MG/ML Pen Injector:** Demonstrates a mean HbA1c reduction of -1.06, with a mean % change of -9.98%. This also has a decent sample size (n=10).
- **insulin human isophane 70 UNT/ML / Regular Insulin Human 30 UNT/ML Injectible Suspension [Humulin]:** Exhibits the most significant mean reduction in HbA1c and HbA1c percentage of -2.98, and -38.40% respectively. However, only four cases (n=4).

2. Medications with Blood Glucose Changes:

- **24 HR Metformin hydrochloride 500 MG Extended Release Oral Tablet:** Large average increase of 73.82 mg/dL which is about 85%, might suggest that some diabetic individuals may not respond well to the drug.
- **3 ML liraglutide 6 MG/ML Pen Injector:** Large average increase of 37.90 mg/dL which is about 53%, might suggest that some diabetic individuals may not respond well to the drug.
- **Insulin Lispro 100 UNT/ML Injectible Solution [Humalog]:** Most significant glucose reduction with -47 mg/dL but only one patient was on the treatment, meaning more data is required.
- **Atorvastatin 80 MG Oral Tablet and Captopril 25 MG Oral Tablet:** Resulted in a glucose reduction of -26 mg/dL but only one patient was on the treatment, meaning more data is required.

3. Consistency Across Patients:

- The medication effectiveness summaries provide aggregated data, but to truly assess consistency, we need to look at individual patient timelines.
- Patient **6e09f851** shows that when Atorvastatin, Captopril, and Naproxen sodium were administered on Myocardial Infarction event day, blood glucose level reduced by -26.00 mg/dL, which is consistent with summary statistics.
- Patient **9fdc58d1**, showed that Clopidogrel, Simvastatin, Amlodipine, and

Chain of Thought prompt

Analyze the Siyeh Synthetic medical dataset in llm_input_data.txt to identify correlations between specific medications and improved health outcomes for diabetic patients. Think through this problem step by step:

Step 1: First, review the dataset summary section.

- Note the total number of diabetic patients in the dataset
- Identify the key health metrics being tracked (HbA1c, blood glucose, etc.)
- Understand the scope of the medication data available

Step 2: Analyze the medication effectiveness summaries.

- Examine the HbA1c Outcomes by Medication table
- Note which medications show the largest mean reductions in HbA1c
- Consider the sample size (Count) for each medication to assess reliability
- Repeat this analysis for the Blood Glucose Outcomes by Medication table
- Identify medications that perform well across both metrics

Step 3: Examine the patient examples to understand individual responses.

- Look at the demographic information for each patient
- Review their medication histories and when each medication was started
- Analyze the health outcomes tables showing before/after values
- Study the timelines showing how metrics changed relative to medication start dates

Step 4: Look for patterns in medication effectiveness.

- Identify medications that consistently show improvements across multiple patients
- Note the magnitude of improvements for different medications
- Consider the timeframe needed to see improvements with different medications
- Look for any medications that show minimal or negative effects

Step 5: Consider potential confounding variables.

- Examine whether patient demographics correlate with medication effectiveness
- Look for patterns in comorbidities that might influence outcomes
- Consider whether medication combinations appear more effective than single medications
- Note any timeline patterns that might suggest external factors affecting outcomes

Step 6: Draw conclusions about medication effectiveness.

- Rank medications by their effectiveness in improving glycemic control
- Identify which medications show the most consistent positive outcomes
- Note any medications that appear ineffective in this dataset
- Consider which medications might be most appropriate for specific patient subgroups

Based on this step-by-step analysis, provide a detailed report on which medications correlate with the best outcomes for diabetic patients in the Siyeh Synthetic dataset, including supporting evidence from your analysis.

Chain of Thought results

(Complete response [here](#) (70 lines))

Summary:

- 1. Insulin formulations show best HbA1c reduction, while metformin offers consistent effectiveness with fewer side effects.
- 2. Patient demographics and comorbidities significantly impact medication effectiveness.
- 3. Limited sample sizes for most medications suggest caution in generalizing results.

Here's a detailed report analyzing the Siyeh Synthetic medical dataset to identify medications correlated with improved health outcomes for diabetic patients:

Step 1: Dataset Summary Review

- **Number of Diabetic Patients:** 498. This represents the population from which the data is drawn.
- **Key Health Metrics:** Glucose, Body Mass Index, Hemoglobin A1c (HbA1c), Diastolic Blood Pressure, and Systolic Blood Pressure. These are the key indicators used to measure diabetic health and medication effectiveness.
- **Medication Data:** The dataset includes a list of 65 medications analyzed and comprehensive patient medication histories.

Step 2: Analysis of Medication Effectiveness Summaries

- **HbA1c Outcomes:** The table highlights the medications showing the largest mean reductions in HbA1c.
 - **insulin** human isophane **70** UNT/ML / Regular Insulin Human **30** UNT/ML Injectable Suspension [**Humulin**] shows the greatest mean reduction of -2.98, with a mean % change of -38.40%. However, the count is only 4, making it less statistically reliable.
 - **24** HR Metformin hydrochloride **500** MG Extended Release Oral Tablet shows a significant mean reduction of -1.35 with a count of 17. Mean % change of -15.31%.
 - **3** ML liraglutide **6** MG/ML **Pen** Injector has a mean reduction of -1.06, a count of 10, and a mean % change of -9.98%.
 - **canagliflozin** **100** MG Oral Tablet shows a mean reduction of -1.23 with a count of 4. Mean % change of -13.01%.
- **Blood Glucose Outcomes:** Examining the blood glucose outcomes reveals a different picture.
 - **24** HR Metformin hydrochloride **500** MG Extended Release Oral Tablet resulted in a mean blood glucose increase of 73.82, with a mean % change of 85.48%, and count of 17. This is counterintuitive and requires further scrutiny.
 - **3** ML liraglutide **6** MG/ML **Pen** Injector resulted in a mean blood glucose increase of 37.90, with a mean % change of 53.12%, and count of 10.
 - **Insulin** Lispro **100** UNT/ML Injectable Solution [**Humalog**] shows a mean reduction of -47 with a count of 1.
- **Initial Observations:** It appears **24** HR Metformin hydrochloride **500** MG Extended Release Oral Tablet and **3** ML liraglutide **6** MG/ML **Pen** Injector effectively lower HbA1c, but increase blood glucose. Insulin seems to lower both, but has a very low count. Canagliflozin seems effective, but also with very low

Prompting Evaluation

Surprisingly, each prompt returned very similar results:

- Responses ranged from 62 to 100 lines, something like 2-4 pages.
- Each identified Insulin, Metformin, and Semaglutide as the most effective drugs for managing diabetic blood chemical markers.
- Each had similar caveats about demographics, sample size, and synthetic data quality.
- Each dealt with the pre-digested data set clearly; there was no clear benefit to adding extra examples in the few-shot prompt, or to adding multi-step instructions.

As of Spring 2025, LLMs have been progressing more by using internal Chain of Thought processes. It's not clear that explicitly providing your own chain of thought prompt will necessarily improve a model response unless you really do want to take different analysis steps than a common-sense approach.

Possible Improvements

- Each LLM response highlighted similar data quality issues of demographic representation, sample size, etc.
- These are the same issues any scientist faces when trying to make an argument with data; they're not unique to LLMs.
- To truly evaluate LLM data analysis performance, I would want to have a known "correct" interpretation which the LLM might or might not identify. As it is, each prompt identified the most popular and effective medications against diabetes. It's difficult to tell if this is insightful or simply obvious from the dataset.
- A better test might be to identify previously unknown patterns in medication data. Examples:
 - GLP-1 inhibitors were caused weight loss in diabetic patients. It took 15+ years for off-label uses to be noticed and applied.
 - Viagra was originally developed as a hypertension medication; its anti-ED effects were only noticed after the fact. Would an LLM have noticed this effect on its own?