Navigating ARPA-H

Matt Might | might@uab.edu | @mattmight

UAB HUGH KAUL PRECISION MEDICINE INSTITUTE

ARPA-H?

Advanced Research Projects Agency for Health

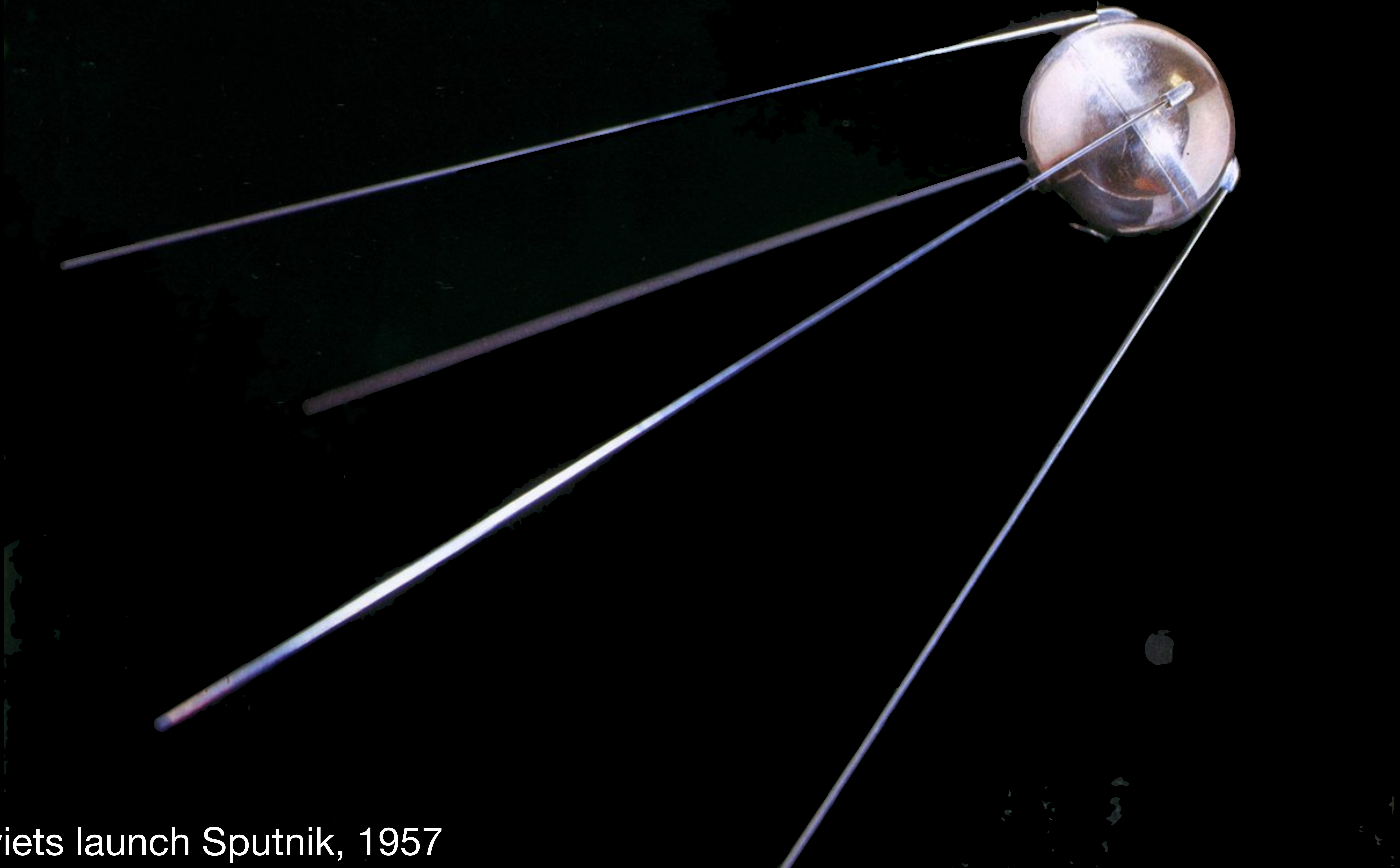
3

Where did it come from?

What are “ARPA-class” projects?

Tips for ARPA-H funding

History



Soviets launch Sputnik, 1957

Eisenhower learns of Sputnik

ARPA is born, 1958

Typical ARPA project

The Internet (originally called ARPANet)

Stealth technology

GPS

Rebranded in 1972: DARPA

And also…

1. Advanced Robotics
2. Artificial Intelligence (AI)
3. Precision Medicine and Biotechnology
4. Autonomous Vehicles and Unmanned Systems
5. Cybersecurity
6. Materials Science and Manufacturing
7. Neuroscience and Brain-Computer Interfaces
8. Hypersonics
9. Quantum Computing and Communications
10. Health and Medical Technologies
11. Space Technologies
12. Energy and Power Systems
13. Human Performance and Bioenhancement
14. Advanced Sensors and Detection Systems
15. High-Energy Laser Systems

How?

High risk, high reward

Parallel approaches

Short timelines

Fail fast philosophy

Significant deliverables

High autonomy for PMs

Limited terms for PMs

Examples: My DARPA projects

Cyberwarfare and cybersecurity

CRASH: Create unhackable OS

APAC: Malware-proof phones

STAC: Leak-proof programs

Born in 2022

Modeled on DARPA

(perhaps too closely)

Inspired by Operation Warp Speed

Two new ARPA-H Projects

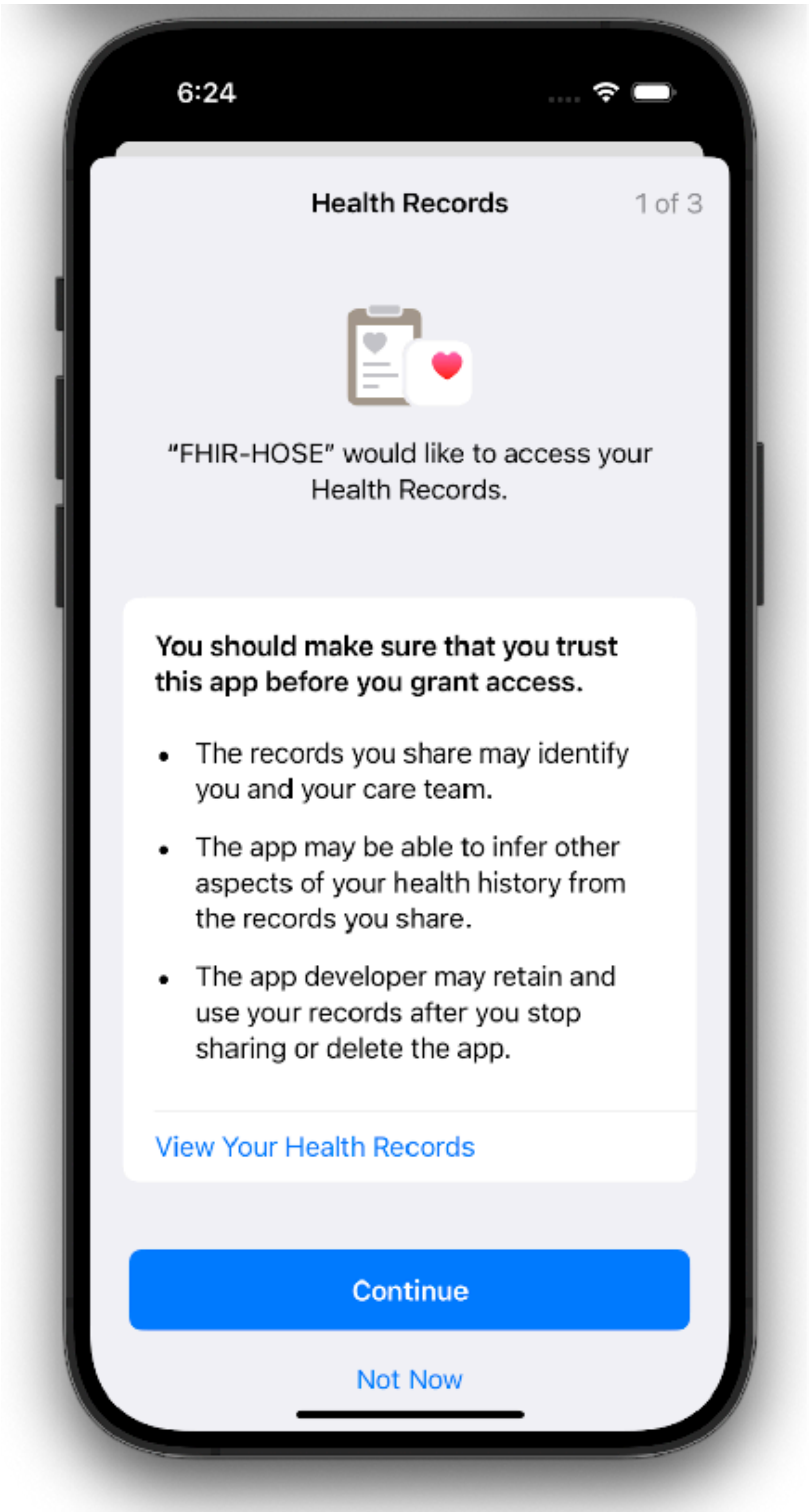
MATRIX project w/ Dave Fajgenbaum: $48 million

AI search for treatments for all 10,000 rare disorders

AI-Enabled Biomedical Data Fabric: $9 million

Collect => Harmonize => Analyze

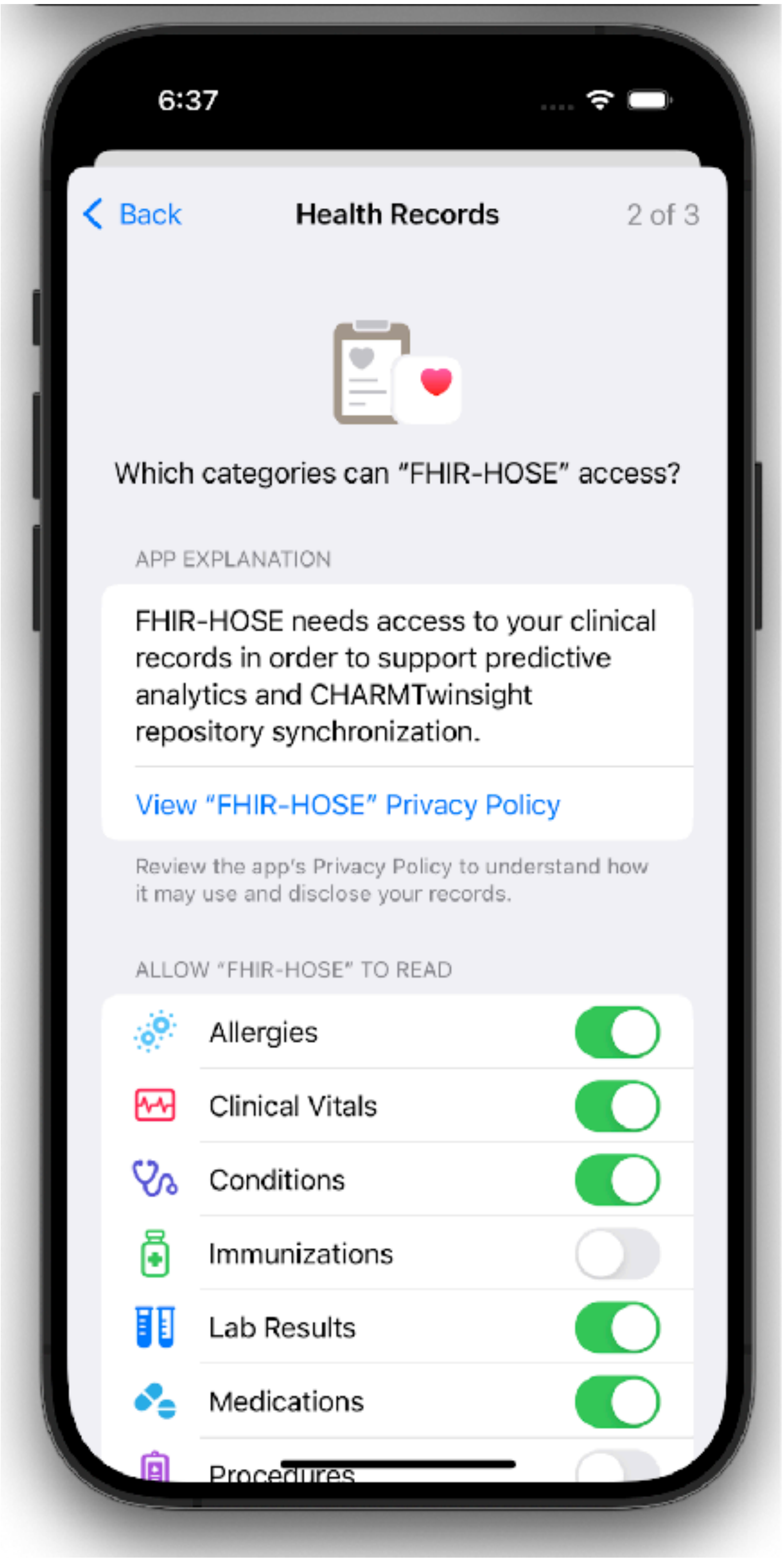
Ex: Importing patient data via phone



"FHIR-HOSE" would like to access your Health Records.

You should make sure that you trust this app before you grant access.

- The records you share may identify you and your care team.
- The app may be able to infer other aspects of your health history from the records you share.
- The app developer may retain and use your records after you stop sharing or delete the app.



Which categories can "FHIR-HOSE" access?

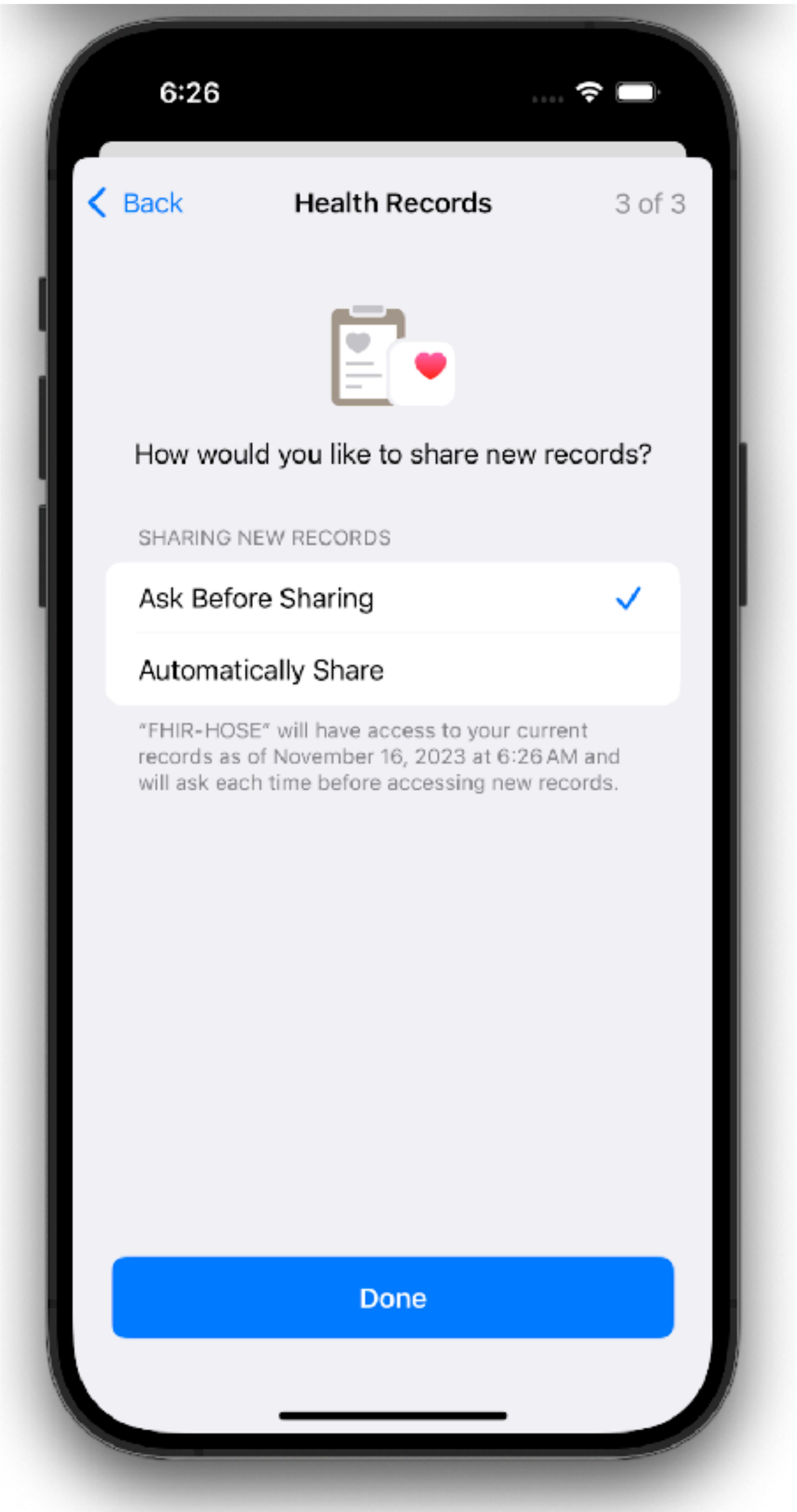
APP EXPLANATION

FHIR-HOSE needs access to your clinical records in order to support predictive analytics and CHARMTwinsight repository synchronization.

Review the app's Privacy Policy to understand how it may use and disclose your records.

ALLOW "FHIR-HOSE" TO READ

Allergies ●
Clinical Vitals ●
Conditions ●
Immunizations ○
Lab Results ●
Medications ●
Procedures ○

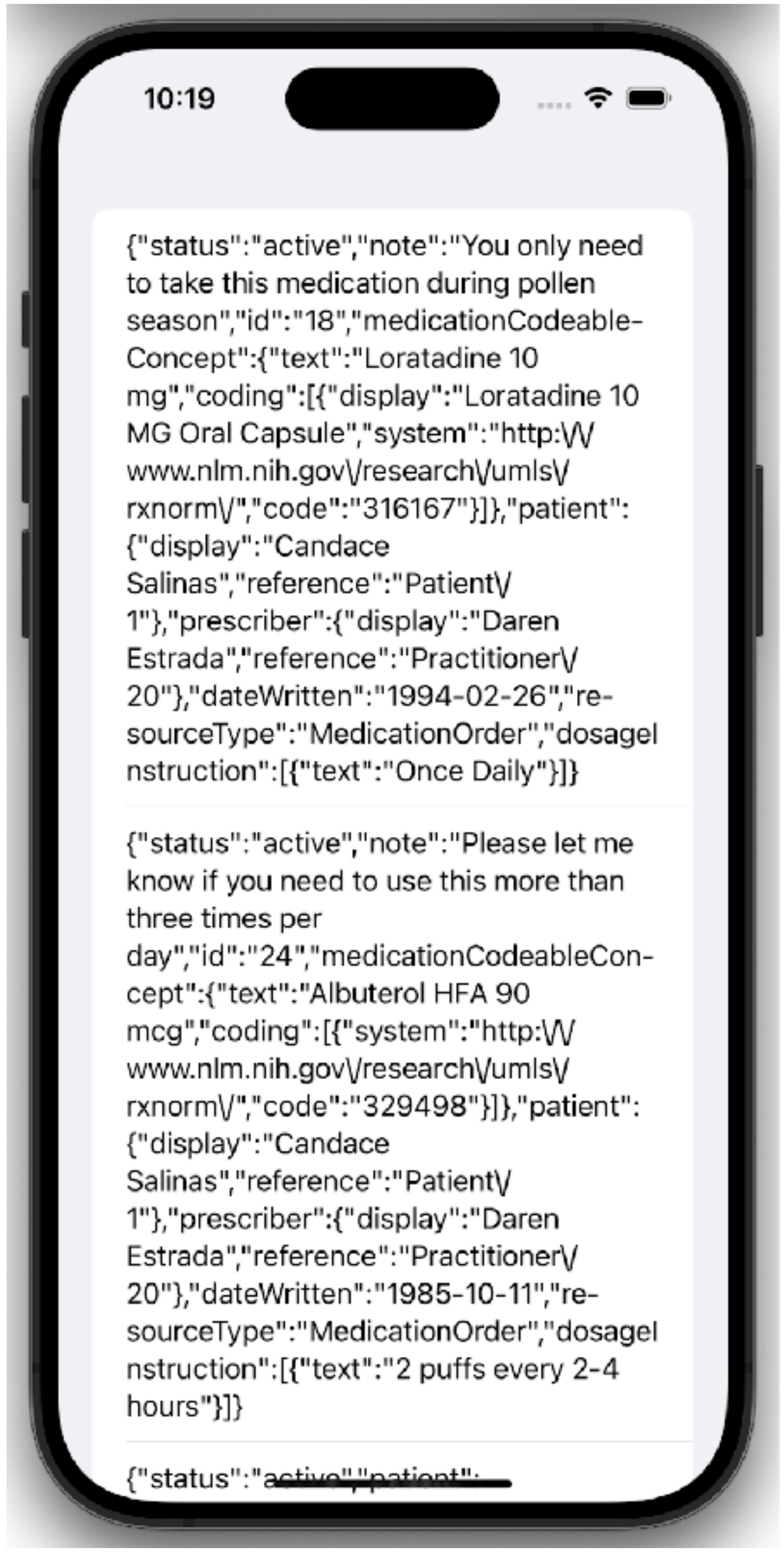


How would you like to share new records?

SHARING NEW RECORDS

Ask Before Sharing ✓
Automatically Share

"FHIR-HOSE" will have access to your current records as of November 16, 2023 at 6:26 AM and will ask each time before accessing new records.



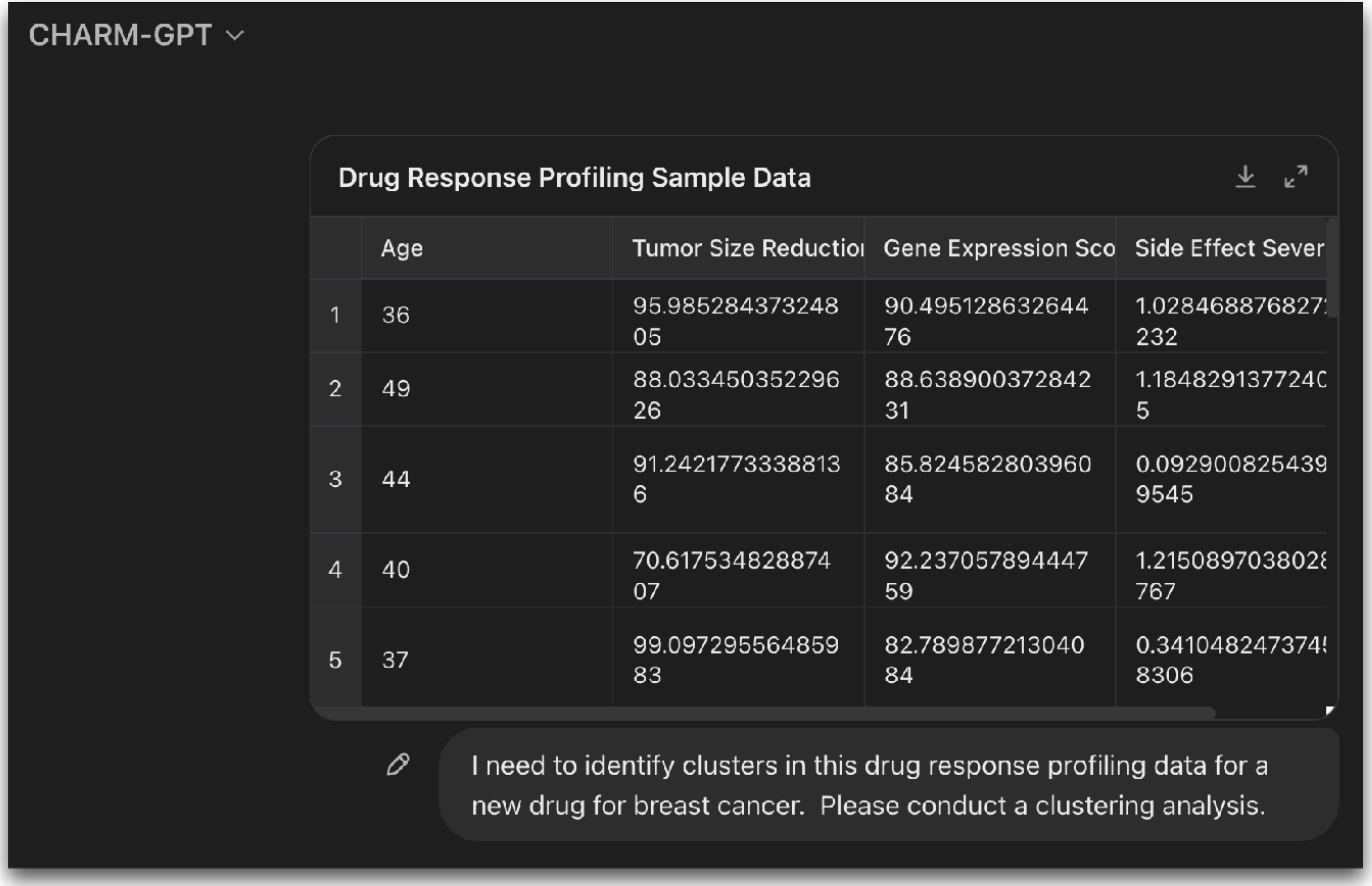
{"status":"active","note":"You only need to take this medication during pollen season","id":"18","medicationCodeableConcept":{"text":"Loratadine 10 mg","coding":[{"display":"Loratadine 10 MG Oral Capsule","system":"http:\/\/www.nlm.nih.gov\/research\/umls\/rxnorm\/","code":"316167"}]},"patient":{"display":"Candace Salinas","reference":"Patient\/1"},"prescriber":{"display":"Daren Estrada","reference":"Practitioner\/20"},"dateWritten":"1994-02-26","resourceType":"MedicationOrder","dosageInstruction":[{"text":"Once Daily"}]}

{"status":"active","note":"Please let me know if you need to use this more than three times per day","id":"24","medicationCodeableConcept":{"text":"Albuterol HFA 90 mcg","coding":[{"system":"http:\/\/www.nlm.nih.gov\/research\/umls\/rxnorm\/","code":"329498"}]},"patient":{"display":"Candace Salinas","reference":"Patient\/1"},"prescriber":{"display":"Daren Estrada","reference":"Practitioner\/20"},"dateWritten":"1985-10-11","resourceType":"MedicationOrder","dosageInstruction":[{"text":"2 puffs every 2-4 hours"}]}

{"status":"active","patient":

My medical record in an open standards format: FHIR

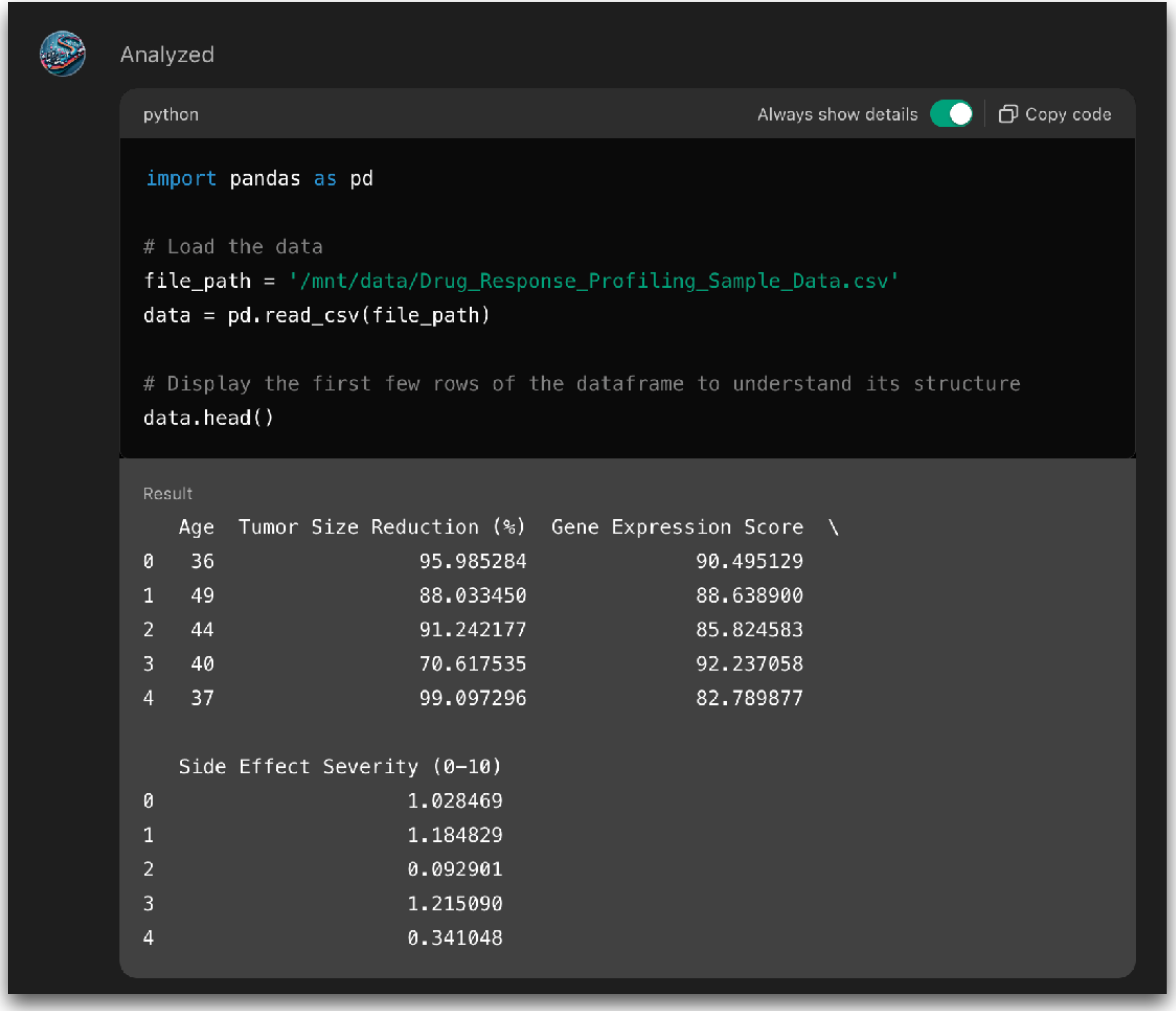
Automated analysis: CHARM-GPT



Drug Response Profiling Sample Data

	Age	Tumor Size Reductio	Gene Expression Sco	Side Effect Sever
1	36	95.98528437324805	90.49512863264476	1.028468876827 232
2	49	88.03345035229626	88.63890037284231	1.184829137724 5
3	44	91.2421773338136	85.82458280396084	0.0929008254399545
4	40	70.61753482887407	92.23705789444759	1.215089703802 767
5	37	99.09729556485983	82.78987721304084	0.341048247374 8306

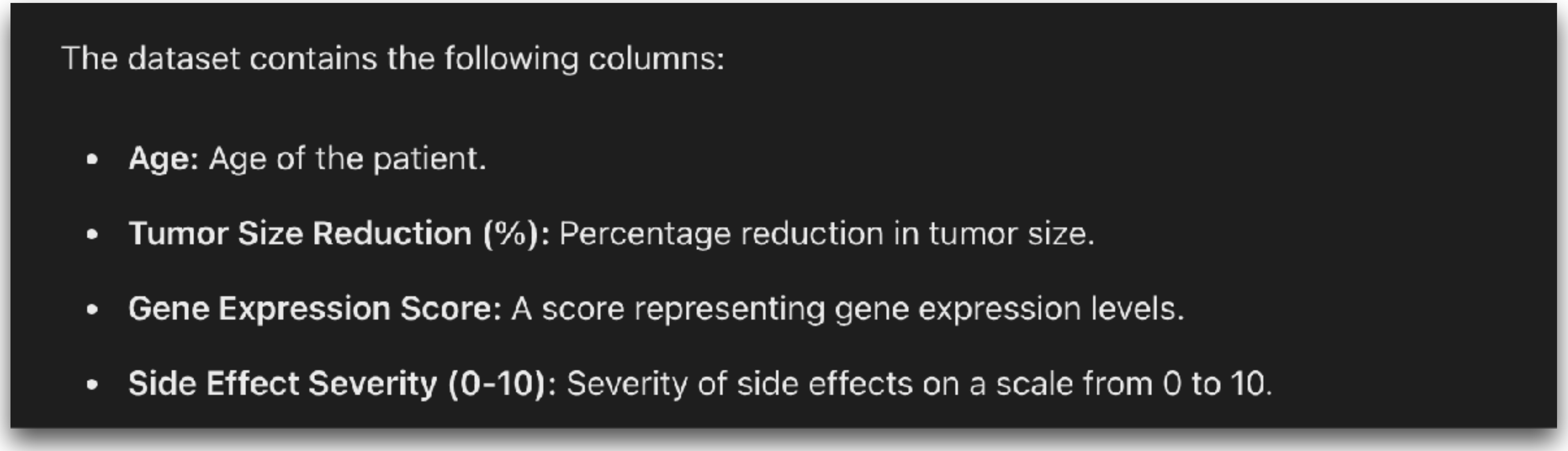
I need to identify clusters in this drug response profiling data for a new drug for breast cancer. Please conduct a clustering analysis.



```python
import pandas as pd

# Load the data
file_path = '/mnt/data/Drug_Response_Profiling_Sample_Data.csv'
data = pd.read_csv(file_path)

# Display the first few rows of the dataframe to understand its structure
data.head()
```

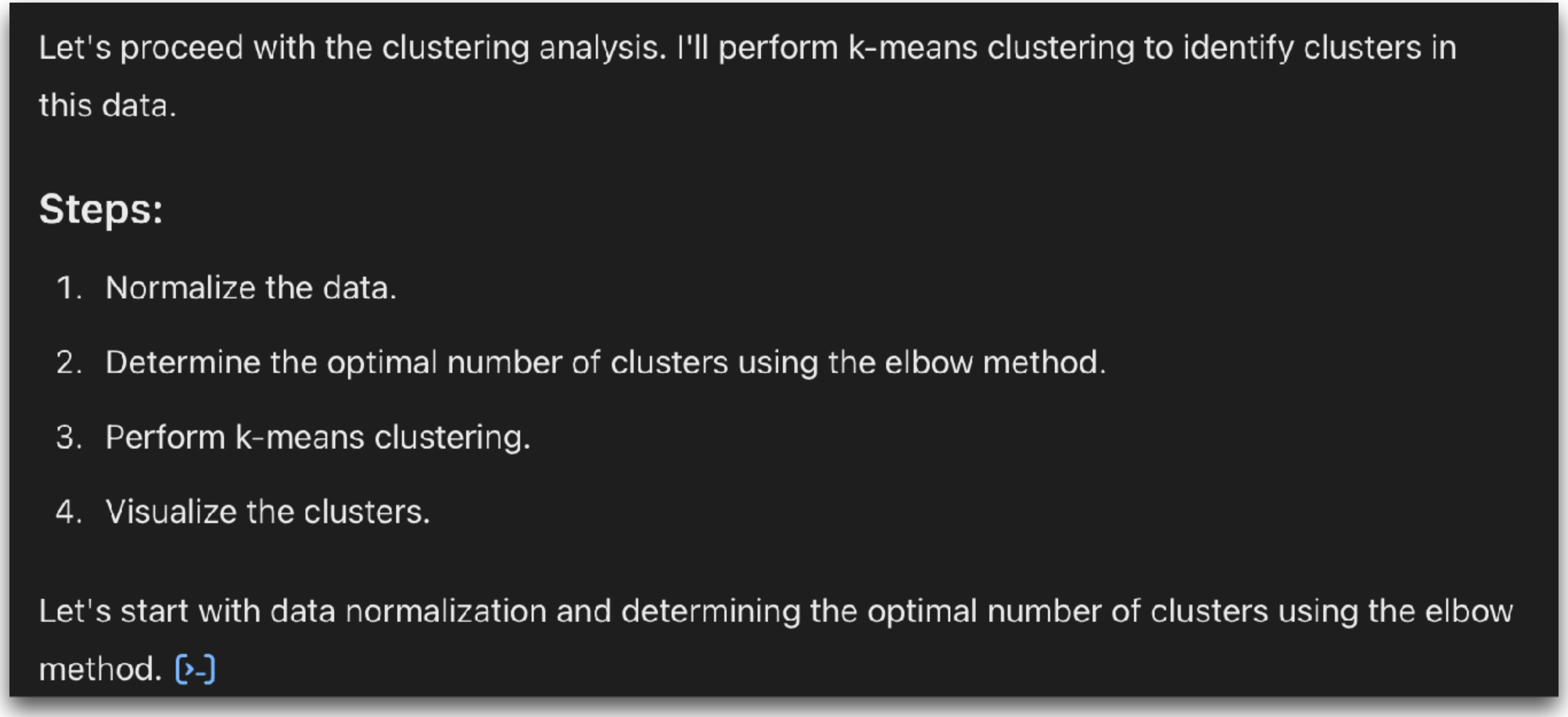
```
   Age  Tumor Size Reduction (%)  Gene Expression Score  \
0   36                 95.985284              90.495129
1   49                 88.033450              88.638900
2   44                 91.242177              85.824583
3   40                 70.617535              92.237058
4   37                 99.097296              82.789877

   Side Effect Severity (0-10)
0                     1.028469
1                     1.184829
2                     0.092901
3                     1.215090
4                     0.341048
```

CHARM-GPT writes code to look at the data and understand its format/structure.



The dataset contains the following columns:

- **Age:** Age of the patient.
- **Tumor Size Reduction (%):** Percentage reduction in tumor size.
- **Gene Expression Score:** A score representing gene expression levels.
- **Side Effect Severity (0-10):** Severity of side effects on a scale from 0 to 10.



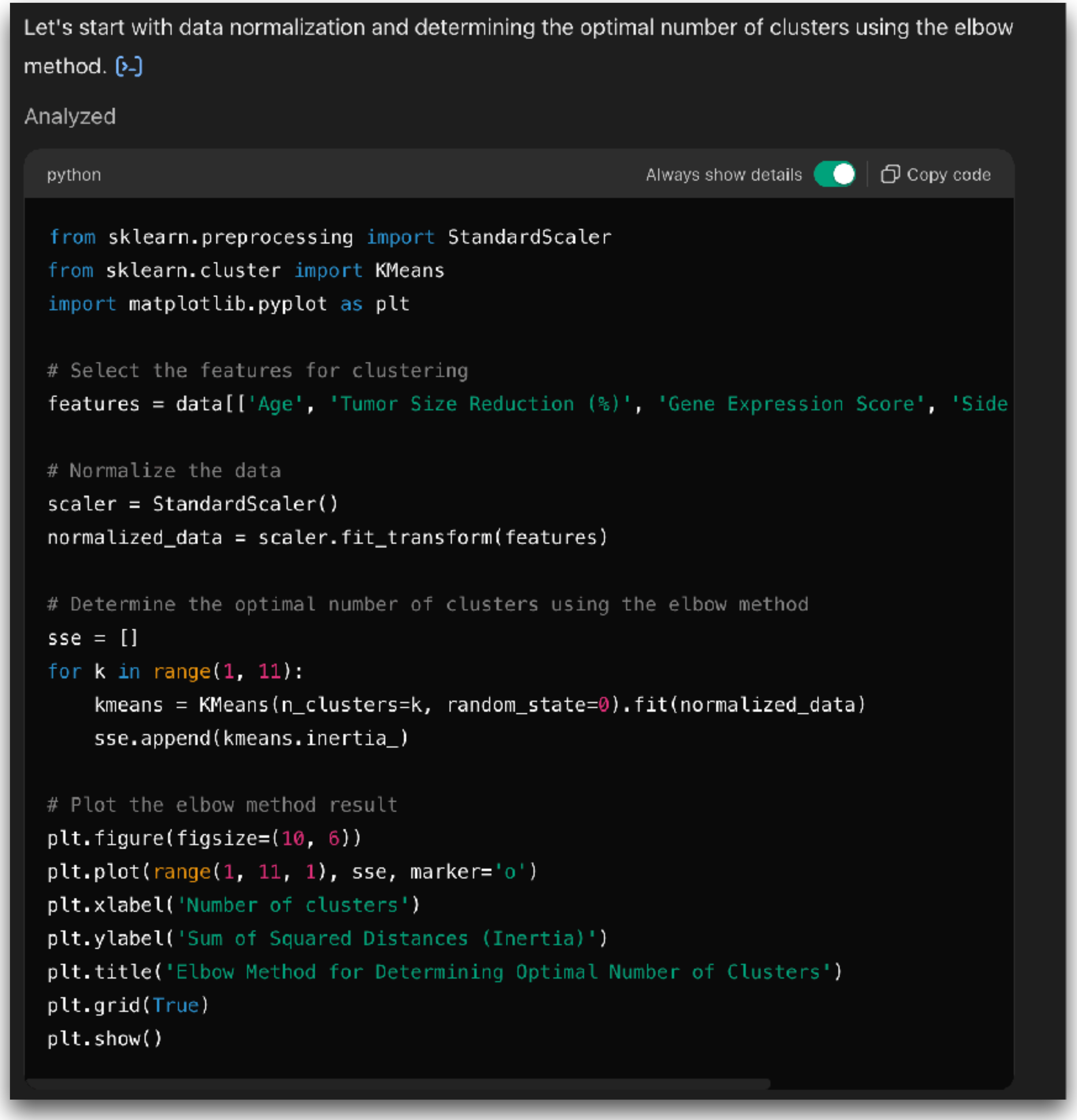
Let's proceed with the clustering analysis. I'll perform k-means clustering to identify clusters in this data.

Steps:

1. Normalize the data.
2. Determine the optimal number of clusters using the elbow method.
3. Perform k-means clustering.
4. Visualize the clusters.

Let's start with data normalization and determining the optimal number of clusters using the elbow method.

CHARM-GPT decides on a strategy for analyzing the data.

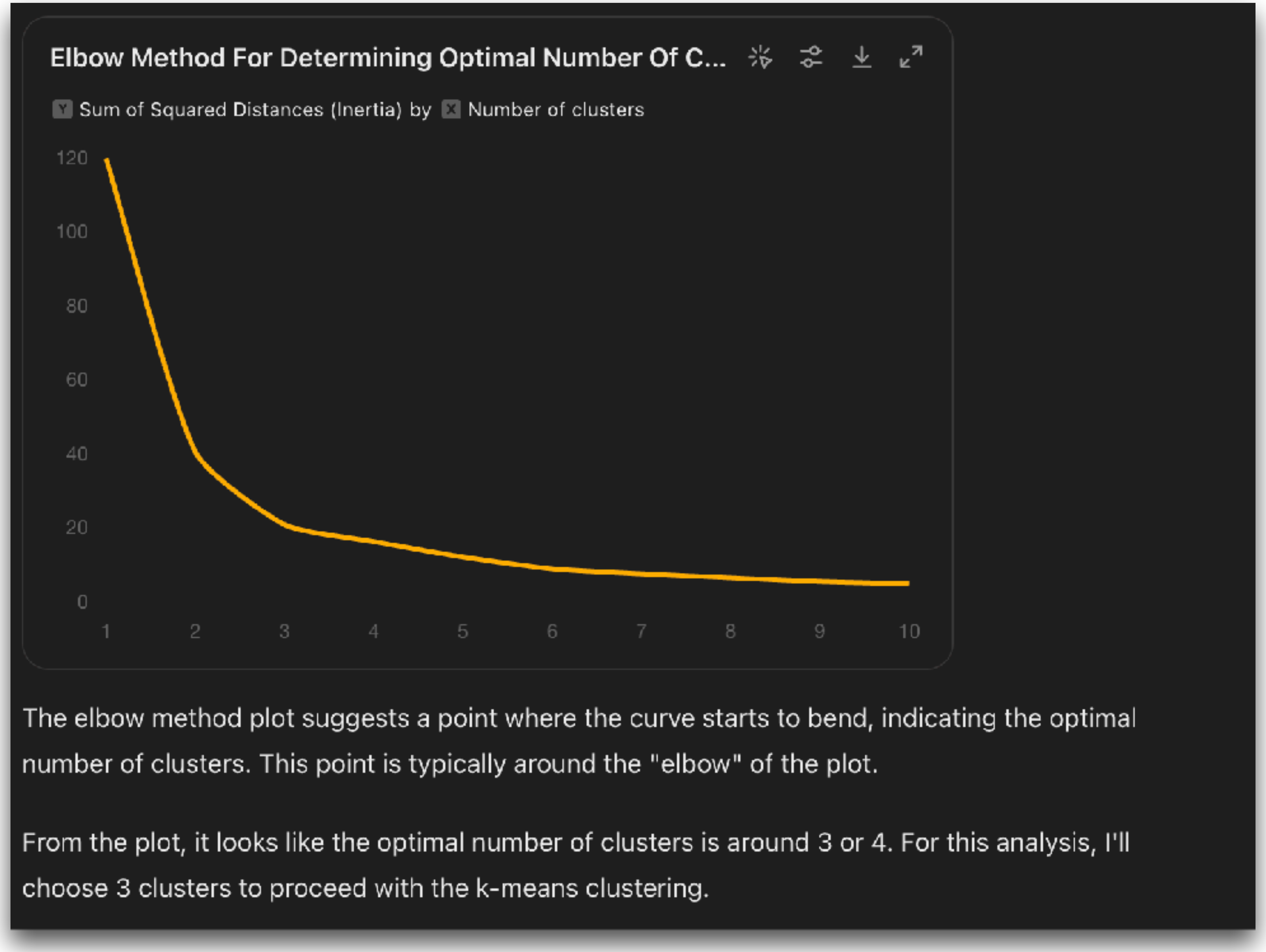


Let's start with data normalization and determining the optimal number of clusters using the elbow method.

Analyzed

```python
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

# Select the features for clustering
features = data[['Age', 'Tumor Size Reduction (%)', 'Gene Expression Score', 'Side

# Normalize the data
scaler = StandardScaler()
normalized_data = scaler.fit_transform(features)

# Determine the optimal number of clusters using the elbow method
sse = []
for k in range(1, 11):
    kmeans = KMeans(n_clusters=k, random_state=0).fit(normalized_data)
    sse.append(kmeans.inertia_)

# Plot the elbow method result
plt.figure(figsize=(10, 6))
plt.plot(range(1, 11, 1), sse, marker='o')
plt.xlabel('Number of clusters')
plt.ylabel('Sum of Squared Distances (Inertia)')
plt.title('Elbow Method for Determining Optimal Number of Clusters')
plt.grid(True)
plt.show()
```

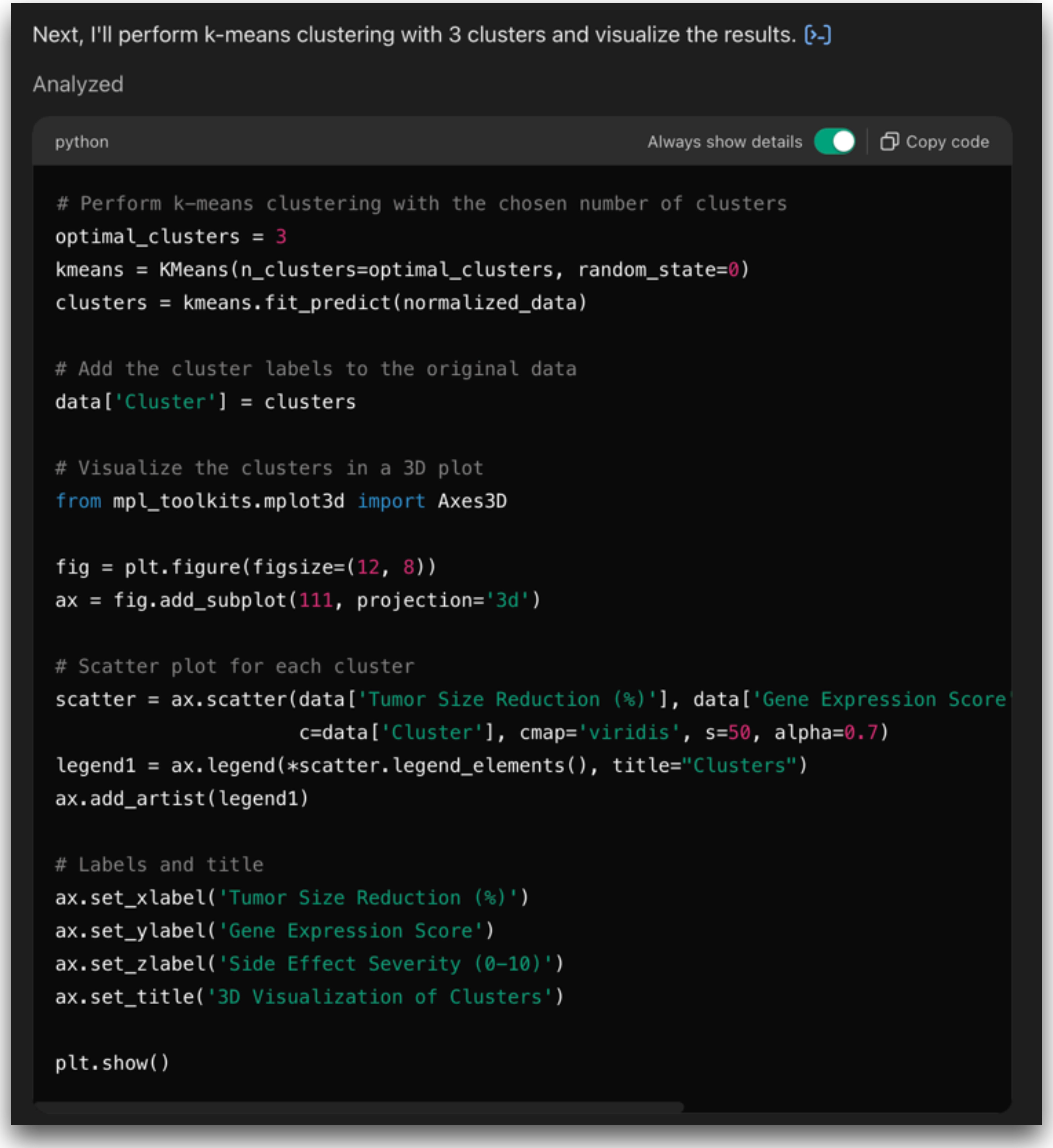
CHARM-GPT writes code to use the elbow method to estimate the number of clusters.


Elbow Method For Determining Optimal Number Of C...
Y Sum of Squared Distances (Inertia) by X Number of clusters
120
100
80
60
40
20
0
1
2
3
4
5
6
7
8
9
10

The elbow method plot suggests a point where the curve starts to bend, indicating the optimal number of clusters. This point is typically around the "elbow" of the plot.

From the plot, it looks like the optimal number of clusters is around 3 or 4. For this analysis, I'll choose 3 clusters to proceed with the k-means clustering.

CHARM-GPT estimates 3 clusters from the elbow method.

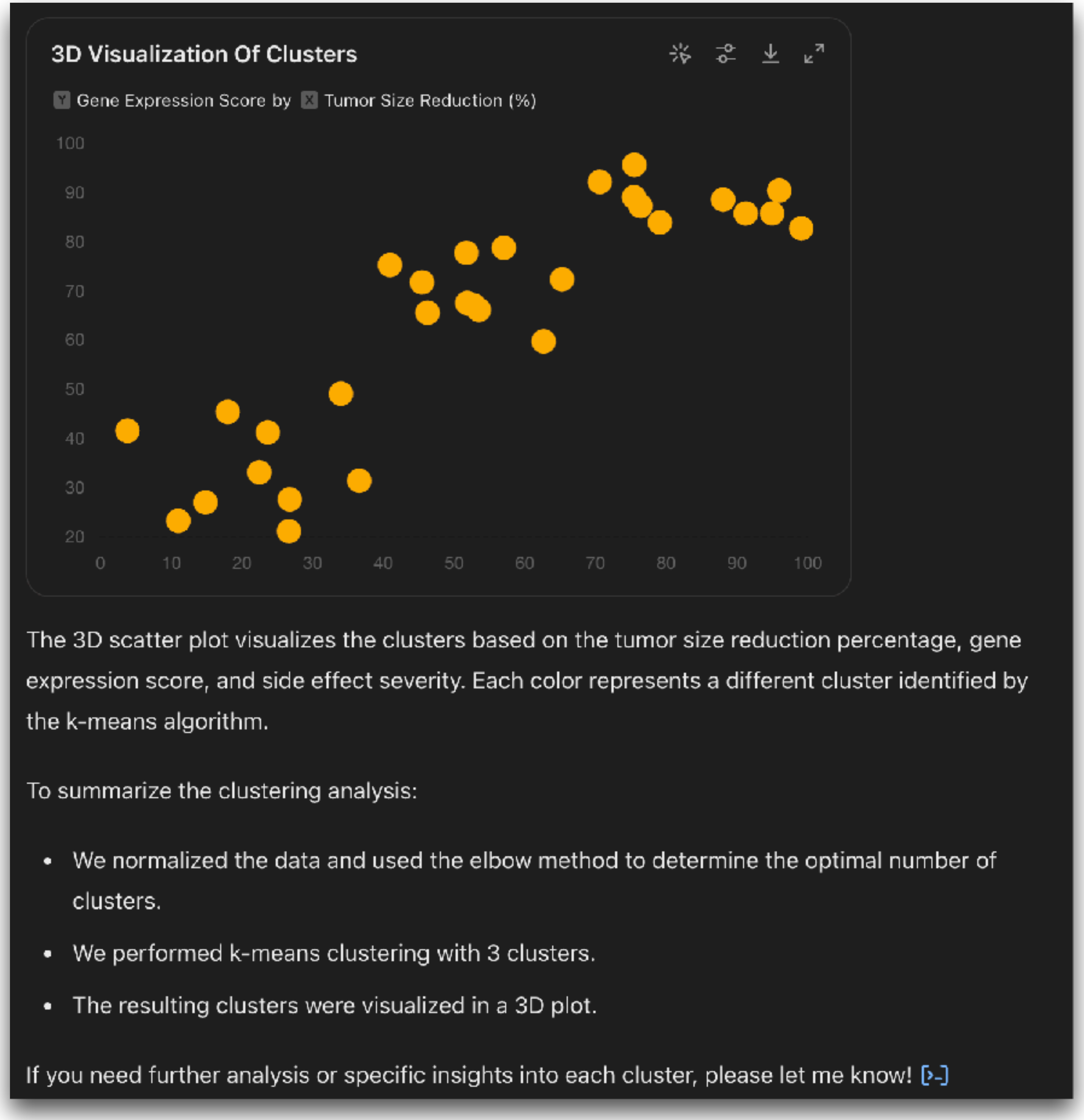


Next, I'll perform k-means clustering with 3 clusters and visualize the results.

Analyzed

```python
# Perform k-means clustering with the chosen number of clusters
optimal_clusters = 3
kmeans = KMeans(n_clusters=optimal_clusters, random_state=0)
clusters = kmeans.fit_predict(normalized_data)

# Add the cluster labels to the original data
data['Cluster'] = clusters

# Visualize the clusters in a 3D plot
from mpl_toolkits.mplot3d import Axes3D

fig = plt.figure(figsize=(12, 8))
ax = fig.add_subplot(111, projection='3d')

# Scatter plot for each cluster
scatter = ax.scatter(data['Tumor Size Reduction (%)'], data['Gene Expression Score
                     c=data['Cluster'], cmap='viridis', s=50, alpha=0.7)
legend1 = ax.legend(*scatter.legend_elements(), title="Clusters")
ax.add_artist(legend1)

# Labels and title
ax.set_xlabel('Tumor Size Reduction (%)')
ax.set_ylabel('Gene Expression Score')
ax.set_zlabel('Side Effect Severity (0-10)')
ax.set_title('3D Visualization of Clusters')

plt.show()
```

CHARM-GPT writes code to run k-means clustering with 3 clusters.


3D Visualization Of Clusters
Y Gene Expression Score by X Tumor Size Reduction (%)
100
90
80
70
60
50
40
30
20
0
10
20
30
40
50
60
70
80
90
100

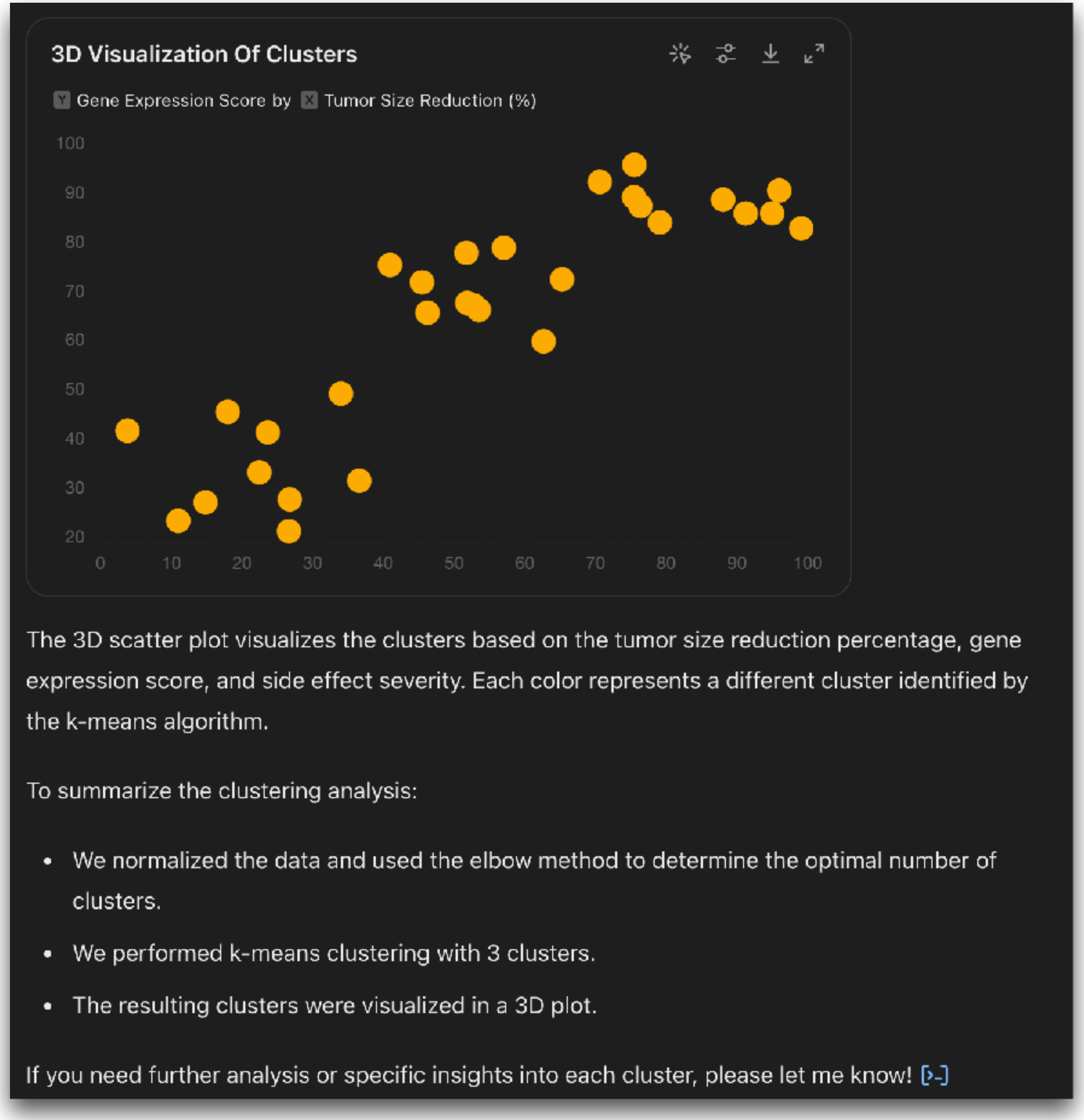
The 3D scatter plot visualizes the clusters based on the tumor size reduction percentage, gene expression score, and side effect severity. Each color represents a different cluster identified by the k-means algorithm.

To summarize the clustering analysis:

- We normalized the data and used the elbow method to determine the optimal number of clusters.
- We performed k-means clustering with 3 clusters.
- The resulting clusters were visualized in a 3D plot.

If you need further analysis or specific insights into each cluster, please let me know! [>_]

CHARM-GPT visualizes the 3 clusters.


3D Visualization Of Clusters
Y Gene Expression Score by X Tumor Size Reduction (%)
100
90
80
70
60
50
40
30
20
0
10
20
30
40
50
60
70
80
90
100

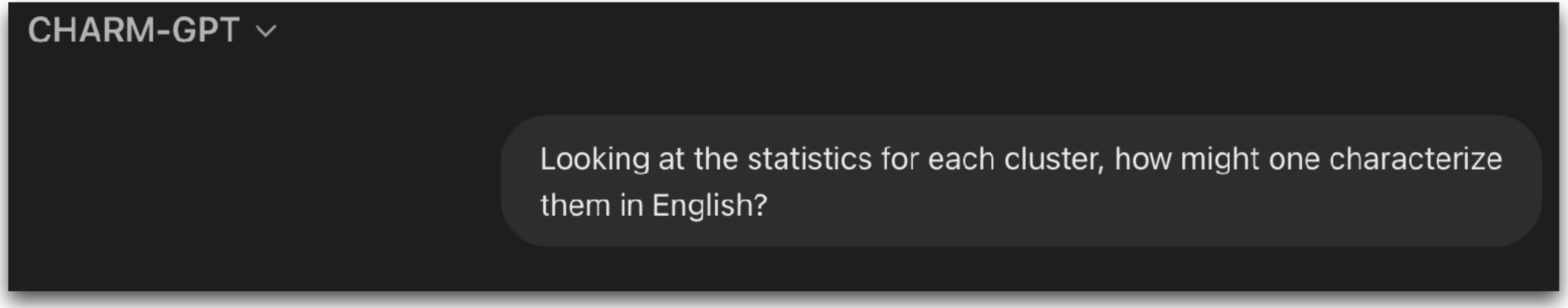
The 3D scatter plot visualizes the clusters based on the tumor size reduction percentage, gene expression score, and side effect severity. Each color represents a different cluster identified by the k-means algorithm.

To summarize the clustering analysis:

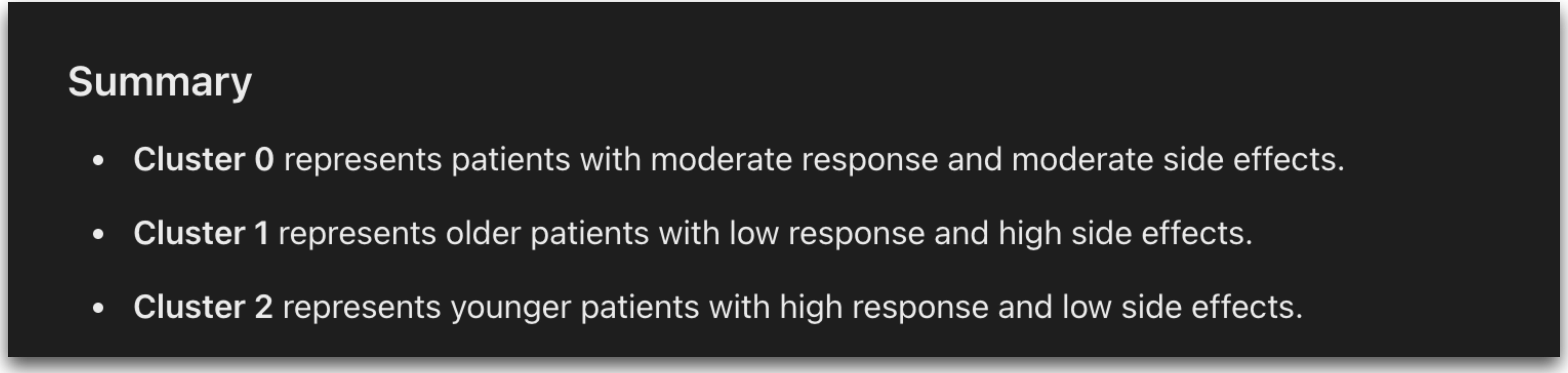
- We normalized the data and used the elbow method to determine the optimal number of clusters.
- We performed k-means clustering with 3 clusters.
- The resulting clusters were visualized in a 3D plot.

If you need further analysis or specific insights into each cluster, please let me know!

CHARM-GPT visualizes the 3 clusters.



Looking at the statistics for each cluster, how might one characterize them in English?



Summary

- **Cluster 0** represents patients with moderate response and moderate side effects.
- **Cluster 1** represents older patients with low response and high side effects.
- **Cluster 2** represents younger patients with high response and low side effects.

One more moving forward…

HealthyDocs: *Secure* medical records: $10 million

Tips on ARPA-H proposals

Highly focused calls

Short turnarounds: 3-4 weeks

~75 pages for technical section

PMs have high discretion

Needs a “secret weapon”

Makes impossible goal possible

Teaming often encouraged

Teaming can be forced post-award!

So, team or be teamed

Deliverables must be delivered

Payment *only upon delivery*

Budgets broken out by milestone

Contracting is a nightmare

Subcontracting is *awful*

Risky if deliverable spans subs

Contracting is lethargic

MATRIX: Intent to award in Nov

MATRIX: *Still* in subcontracting

BDF: Intent to award in February

BDF: *Still* in subcontracting

Use AI for RFA analysis

Use AI to make checklist

Use AI for proposal critique

Use AI for “Task Description Doc”

Also, don’t use ChatGPT

Do use Microsoft Copilot

Thank you!