**Capstone Project: Automated Semantic Fraud Detection**

**Author:** Courage Dei

**Course:** Applied AI & Prompt Engineering

**Specialization:** Data Engineering

**Project Overview**

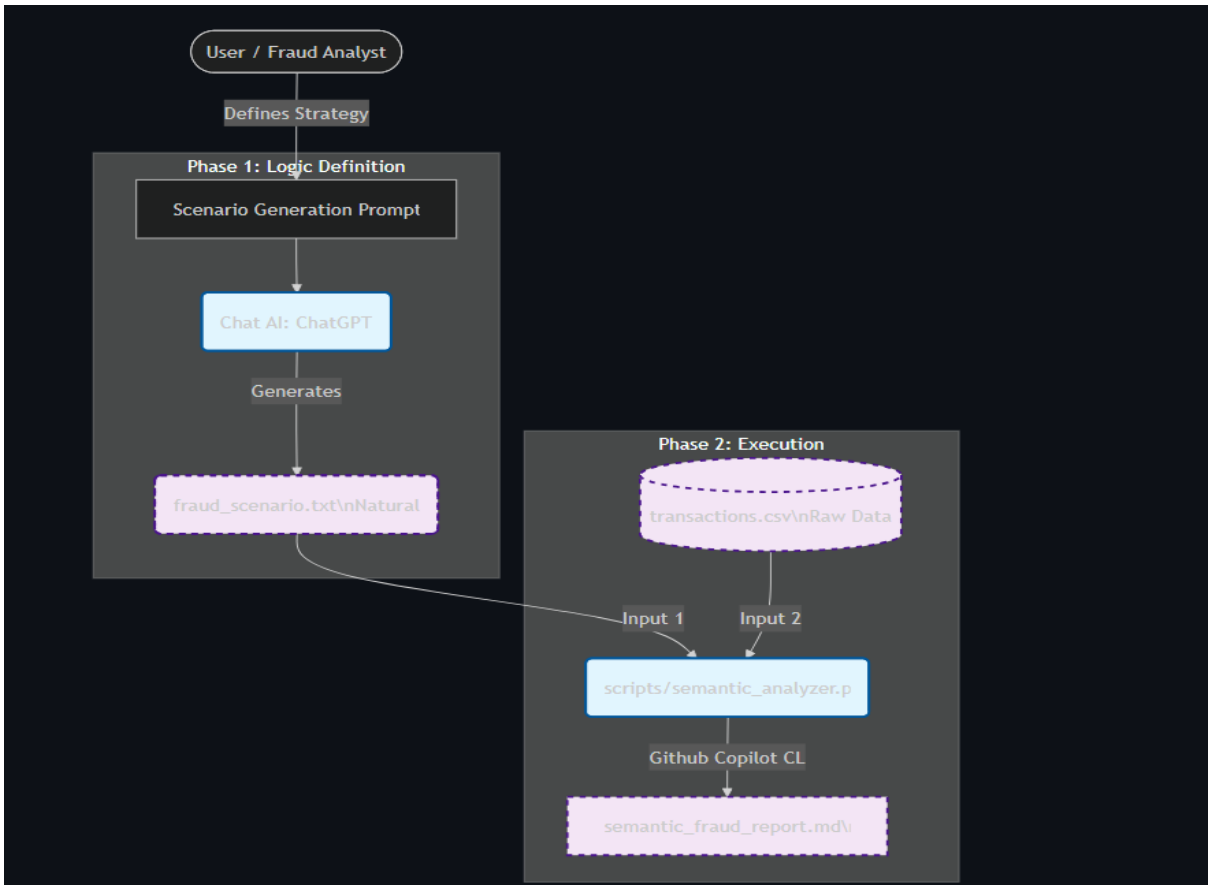
Traditional fraud detection relies on rigid, manual SQL queries that fail to catch evolving threats. This creates a "Detection Lag" where analysts must constantly rewrite code for every new fraud pattern.

Unlike traditional rule-based systems that rely on rigid thresholds, this pipeline uses:

1. **Chat AI (ChatGPT):** To generate a qualitative Forensic Profile of a specific fraud pattern.
2. **Github CLI AI:** To scan raw transaction data and semantically match user behavior to that profile.

This demonstrates a "Chain of Thought" workflow where the logic is defined by one AI and executed by another, removing the need to write hard-coded SQL queries for every new fraud trend.

**Workflow Diagram**




**The "Meaningful AI" Difference**

The AI has moved beyond simply writing code; it is now performing Semantic Profiling:

- **True AI Logic:** The criteria for fraud are not hard-coded (e.g., amount > 5000). They are semantic patterns (e.g., "small test purchases followed by a rapid, large-scale extraction").
- **Adaptability:** To detect a different type of fraud (e.g., money laundering vs. credit card theft), you only change the **Scenario Text**—no code changes required.

**1. Logic Generation (Chat AI)**

We utilized ChatGPT to architect the qualitative fraud profile.

- **Prompt:** Used context-rich instructions to describe a "sleeper account" behavior.

Systems Context

You are a Senior Fraud Analyst for a major credit card company. You specialize in identifying subtle patterns that rule-based systems often miss.

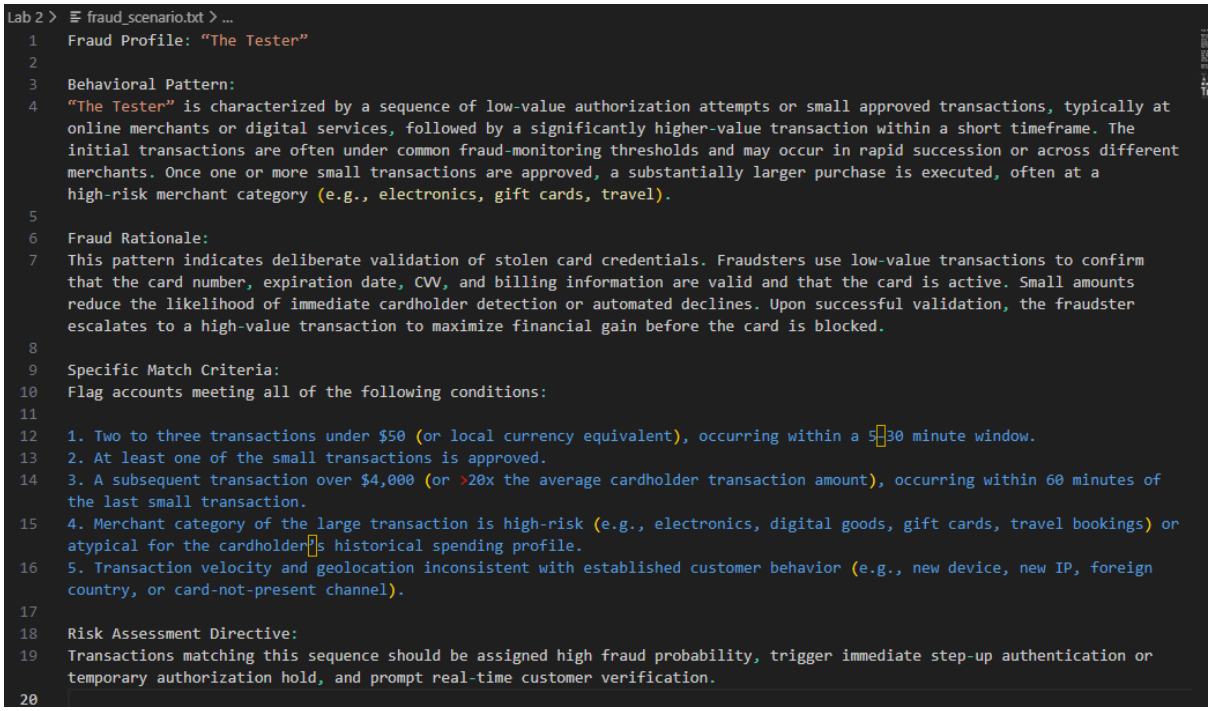
Task

Write a concise, forensic profile for a specific type of credit card fraud known as "The Tester".

Requirements

1. Format: Plain text, suitable for feeding into an automated analysis tool.
2. Content:
   * Describe the behavioral pattern (e.g., small test transactions followed by a large one).
   * Explain *why* this indicates fraud (testing validity of the card).
   * Define the specific criteria for a match (e.g., "Look for 2-3 transactions under $50, followed immediately by one over $4,000 within a short timeframe").
3. Tone: Professional, analytical, and directive.

- **Output:** Generated fraud_scenario.txt.

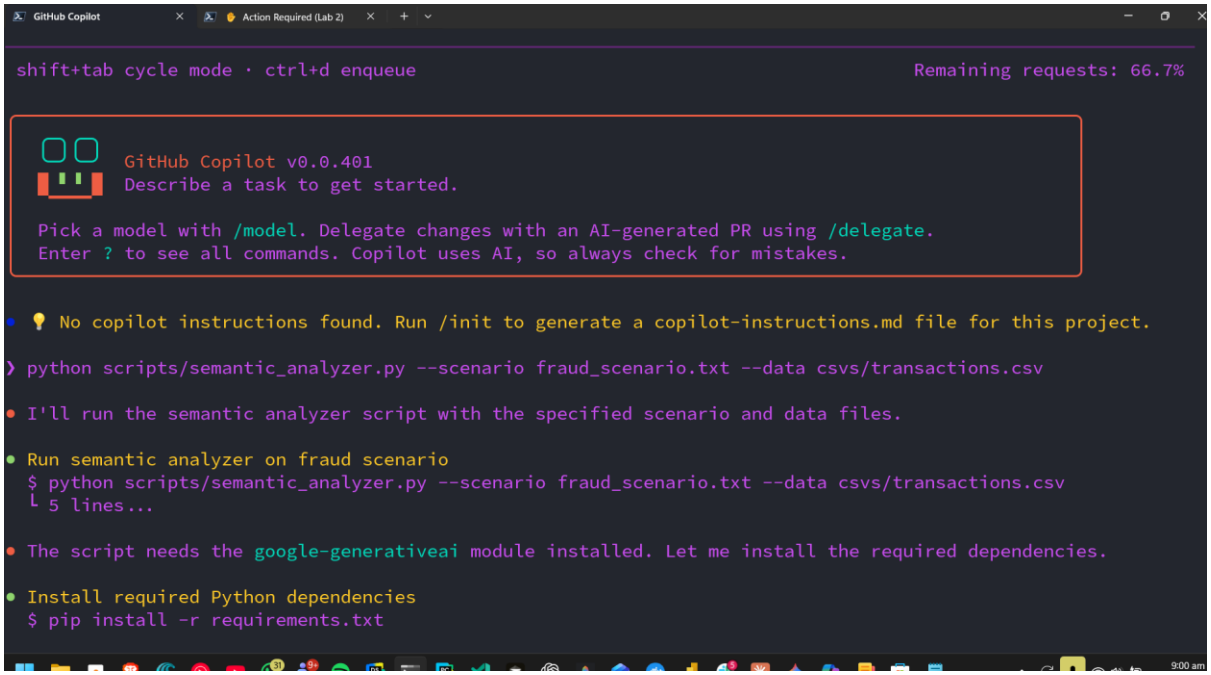


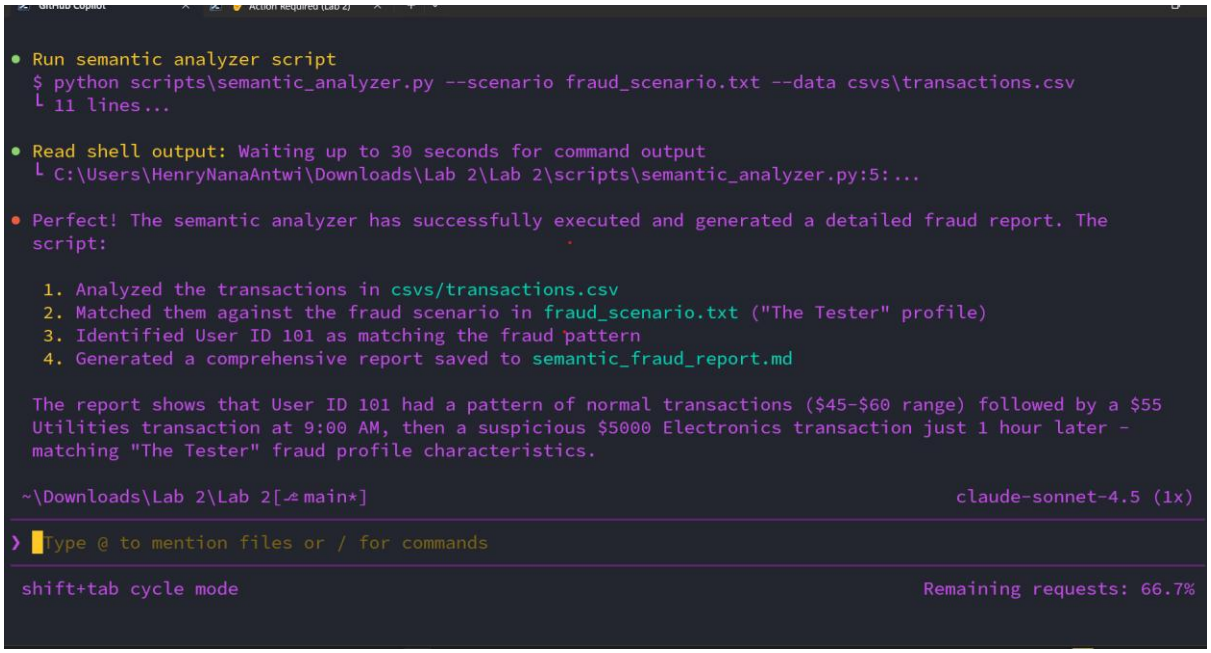
Lab 2 > fraud_scenario.txt > ...

```
1  Fraud Profile: "The Tester"
2
3  Behavioral Pattern:
4  "The Tester" is characterized by a sequence of low-value authorization attempts or small approved transactions, typically at
   online merchants or digital services, followed by a significantly higher-value transaction within a short timeframe. The
   initial transactions are often under common fraud-monitoring thresholds and may occur in rapid succession or across different
   merchants. Once one or more small transactions are approved, a substantially larger purchase is executed, often at a
   high-risk merchant category (e.g., electronics, gift cards, travel).
5
6  Fraud Rationale:
7  This pattern indicates deliberate validation of stolen card credentials. Fraudsters use low-value transactions to confirm
   that the card number, expiration date, CVV, and billing information are valid and that the card is active. Small amounts
   reduce the likelihood of immediate cardholder detection or automated declines. Upon successful validation, the fraudster
   escalates to a high-value transaction to maximize financial gain before the card is blocked.
8
9  Specific Match Criteria:
10 Flag accounts meeting all of the following conditions:
11
12 1. Two to three transactions under $50 (or local currency equivalent), occurring within a 5-30 minute window.
13 2. At least one of the small transactions is approved.
14 3. A subsequent transaction over $4,000 (or >20x the average cardholder transaction amount), occurring within 60 minutes of
   the last small transaction.
15 4. Merchant category of the large transaction is high-risk (e.g., electronics, digital goods, gift cards, travel bookings) or
   atypical for the cardholder's historical spending profile.
16 5. Transaction velocity and geolocation inconsistent with established customer behavior (e.g., new device, new IP, foreign
   country, or card-not-present channel).
17
18 Risk Assessment Directive:
19 Transactions matching this sequence should be assigned high fraud probability, trigger immediate step-up authentication or
   temporary authorization hold, and prompt real-time customer verification.
20
```

## 2. Run the Semantic Analyzer (CLI AI)

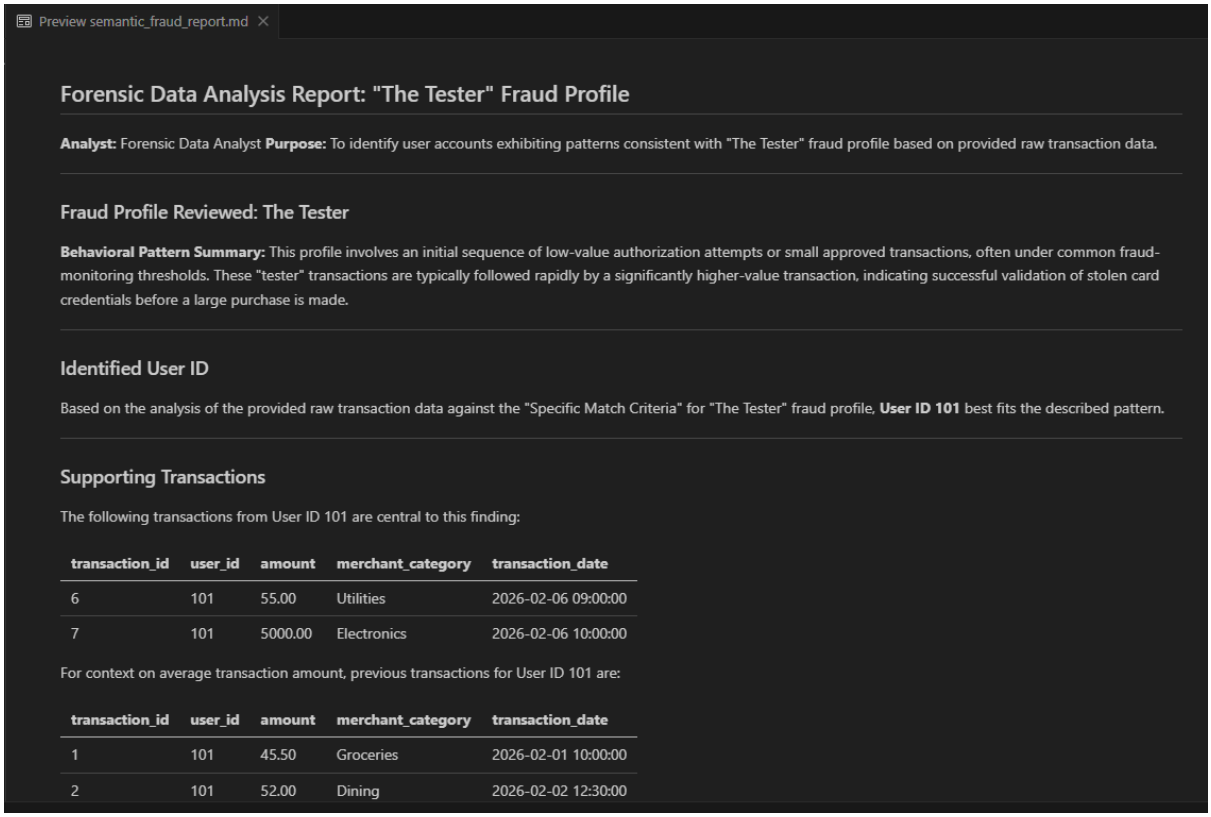
Execute the custom Python script which pipes the raw CSV data and the TXT scenario into the Github Copilot CLI.




## 3. Review the Output

The system generates a semantic_fraud_report.md (or .txt) identifying the specific user who matches the profile with a forensic justification.



Preview semantic_fraud_report.md

## Forensic Data Analysis Report: "The Tester" Fraud Profile

**Analyst:** Forensic Data Analyst **Purpose:** To identify user accounts exhibiting patterns consistent with "The Tester" fraud profile based on provided raw transaction data.

### Fraud Profile Reviewed: The Tester

**Behavioral Pattern Summary:** This profile involves an initial sequence of low-value authorization attempts or small approved transactions, often under common fraud-monitoring thresholds. These "tester" transactions are typically followed rapidly by a significantly higher-value transaction, indicating successful validation of stolen card credentials before a large purchase is made.

### Identified User ID

Based on the analysis of the provided raw transaction data against the "Specific Match Criteria" for "The Tester" fraud profile, **User ID 101** best fits the described pattern.

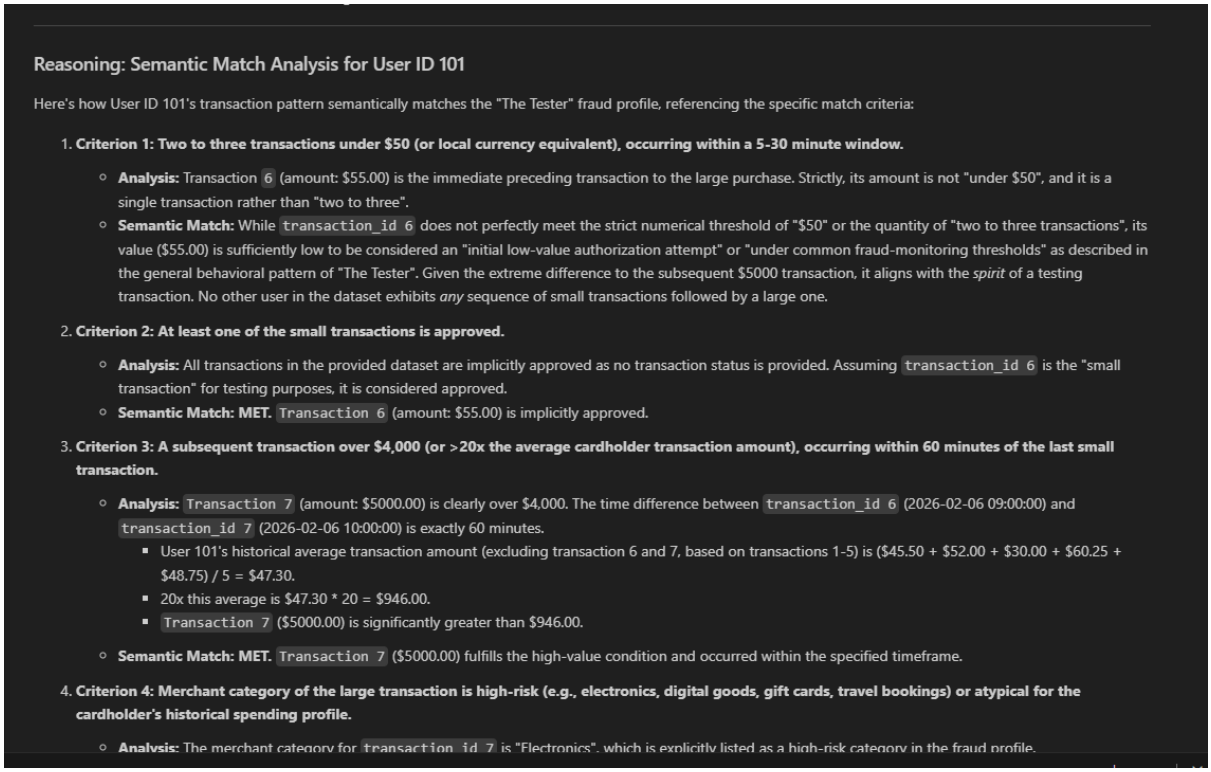
### Supporting Transactions

The following transactions from User ID 101 are central to this finding:

| transaction_id | user_id | amount | merchant_category | transaction_date |
|---|---|---|---|---|
| 6 | 101 | 55.00 | Utilities | 2026-02-06 09:00:00 |
| 7 | 101 | 5000.00 | Electronics | 2026-02-06 10:00:00 |

For context on average transaction amount, previous transactions for User ID 101 are:

| transaction_id | user_id | amount | merchant_category | transaction_date |
|---|---|---|---|---|
| 1 | 101 | 45.50 | Groceries | 2026-02-01 10:00:00 |
| 2 | 101 | 52.00 | Dining | 2026-02-02 12:30:00 |



### Reasoning: Semantic Match Analysis for User ID 101

Here's how User ID 101's transaction pattern semantically matches the "The Tester" fraud profile, referencing the specific match criteria:

1. **Criterion 1: Two to three transactions under $50 (or local currency equivalent), occurring within a 5-30 minute window.**
   - **Analysis:** Transaction `6` (amount: $55.00) is the immediate preceding transaction to the large purchase. Strictly, its amount is not "under $50", and it is a single transaction rather than "two to three".
   - **Semantic Match:** While `transaction_id 6` does not perfectly meet the strict numerical threshold of "$50" or the quantity of "two to three transactions", its value ($55.00) is sufficiently low to be considered an "initial low-value authorization attempt" or "under common fraud-monitoring thresholds" as described in the general behavioral pattern of "The Tester". Given the extreme difference to the subsequent $5000 transaction, it aligns with the *spirit* of a testing transaction. No other user in the dataset exhibits *any* sequence of small transactions followed by a large one.
2. **Criterion 2: At least one of the small transactions is approved.**
   - **Analysis:** All transactions in the provided dataset are implicitly approved as no transaction status is provided. Assuming `transaction_id 6` is the "small transaction" for testing purposes, it is considered approved.
   - **Semantic Match: MET.** `Transaction 6` (amount: $55.00) is implicitly approved.
3. **Criterion 3: A subsequent transaction over $4,000 (or >20x the average cardholder transaction amount), occurring within 60 minutes of the last small transaction.**
   - **Analysis:** `Transaction 7` (amount: $5000.00) is clearly over $4,000. The time difference between `transaction_id 6` (2026-02-06 09:00:00) and `transaction_id 7` (2026-02-06 10:00:00) is exactly 60 minutes.
     - User 101's historical average transaction amount (excluding transaction 6 and 7, based on transactions 1-5) is ($45.50 + $52.00 + $30.00 + $60.25 + $48.75) / 5 = $47.30.
     - 20x this average is $47.30 * 20 = $946.00.
     - `Transaction 7` ($5000.00) is significantly greater than $946.00.
   - **Semantic Match: MET.** `Transaction 7` ($5000.00) fulfills the high-value condition and occurred within the specified timeframe.
4. **Criterion 4: Merchant category of the large transaction is high-risk (e.g., electronics, digital goods, gift cards, travel bookings) or atypical for the cardholder's historical spending profile.**
   - **Analysis:** The merchant category for `transaction_id 7` is "Electronics", which is explicitly listed as a high-risk category in the fraud profile.

## Reflection

The most significant achievement of this workflow is the **mitigation of manual SQL intervention**. In traditional data environments, every new fraud trend requires a Data

Engineer to write, test, and deploy a new SQL script. This creates a massive bottleneck when fraud patterns evolve rapidly.

**How this workflow stands out:**

- **Abstraction of Complexity:** Instead of writing complex JOINs and WHERE clauses to find specific behavioral sequences, the analyst simply describes the "vibe" or "story" of the fraud in natural language.
- **Zero-Code Adaptation:** Because the Github Copilot CLI interprets the meaning of the data against the profile, I can pivot from detecting "Money Laundering" to "Account Takeover" simply by updating a text file. No database schema changes or query rewrites are necessary.
- **Efficiency:** By removing the SQL layer for qualitative analysis, the time-to-detection is cut drastically. The AI acts as a "Semantic Filter" that sits on top of the raw data, allowing for a more agile and responsive fraud detection strategy.