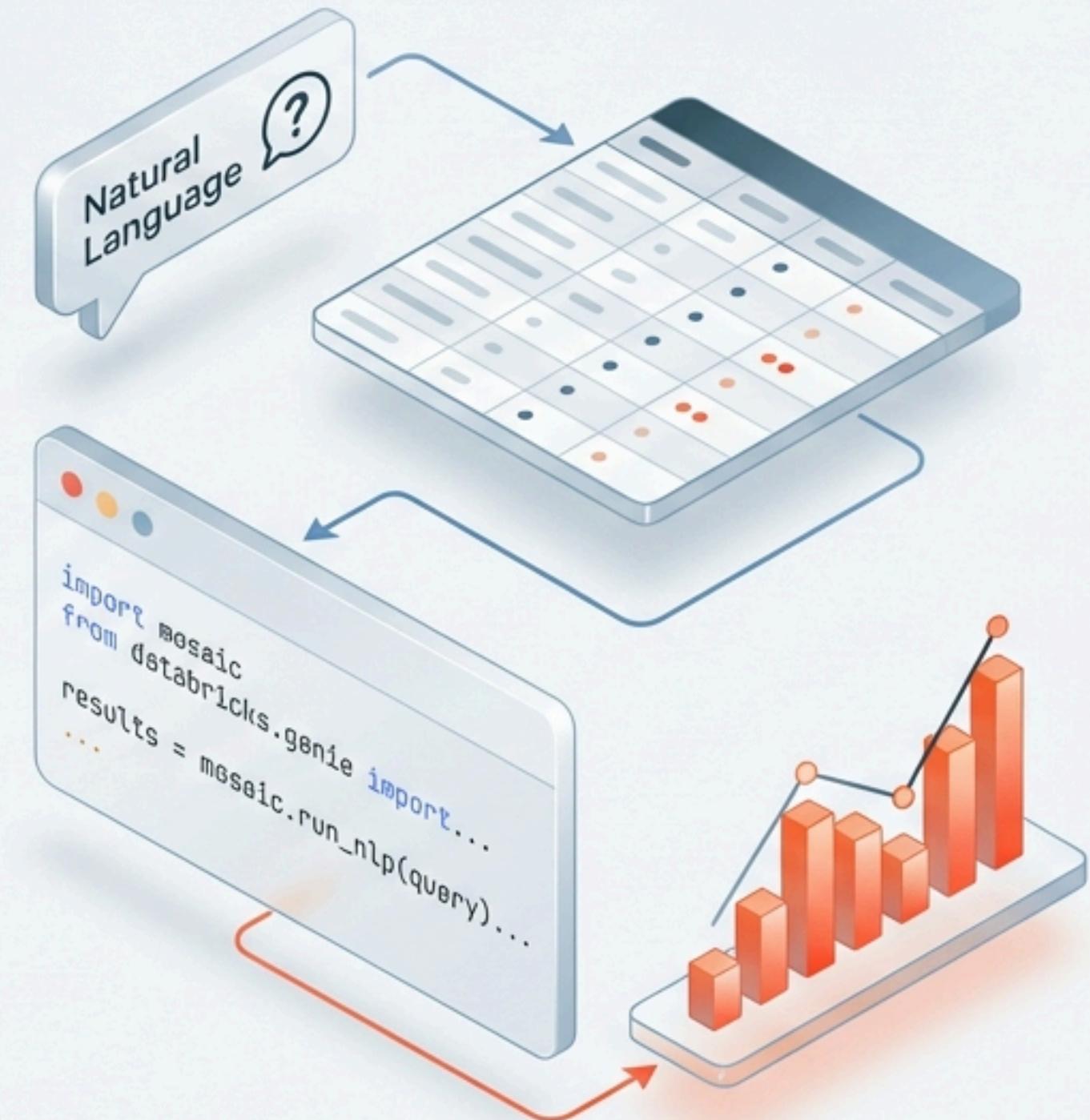




Day 14: AI-Powered Analytics

From Natural Language Questions
to Deployable NLP Pipelines

- ✓ 1. Querying with Genie
- ✓ 2. Exploring Mosaic AI
- ✓ 3. Engineering NLP Tasks
- ✓ 4. Synthesizing Insights



The Workspace: Preparing the Data Environment

The analyst enters a 'New Space' in Databricks Genie, loading three primary tables for investigation: ecommerce data for Oct/Nov 2019 and gold products.

The screenshot shows the Databricks Genie interface with a dark theme. On the left is a sidebar with navigation links: Home, Workspace, Recents, Catalog, SQL Editor, Genie (which is selected), Library, Recent, Coreslats, Tusds, Interop, Compilation, and Data windows. The main area is titled 'New Space' and contains a 'Data' section with the instruction 'Link existing or upload new data to start discovering insights.' Below this is a table view:

Table	Type
ecommerce_2019_oct	Table
ecommerce_2019_nov	Table
gold_products	Table

A red box highlights the 'gold_products' table. To the right of the table is an 'Assistant' panel with a message: 'How ana, their secure your assistans?' and a 'Query Queue' section with a dropdown set to 'Query queue'. At the bottom right is a message input field with placeholder 'Type a message...' and a row of small icons.

Zero-Code Querying: Identifying High-Value Products

Genie translates natural language into structured SQL results instantly.

Cause

Which products have the highest conversion rate?

Effect

product_id	conversion_rate
22500269	100
26401013	100
2702353	100
8000369	100
28721305	100

← Automatic sorting without
SQL 'ORDER BY'.

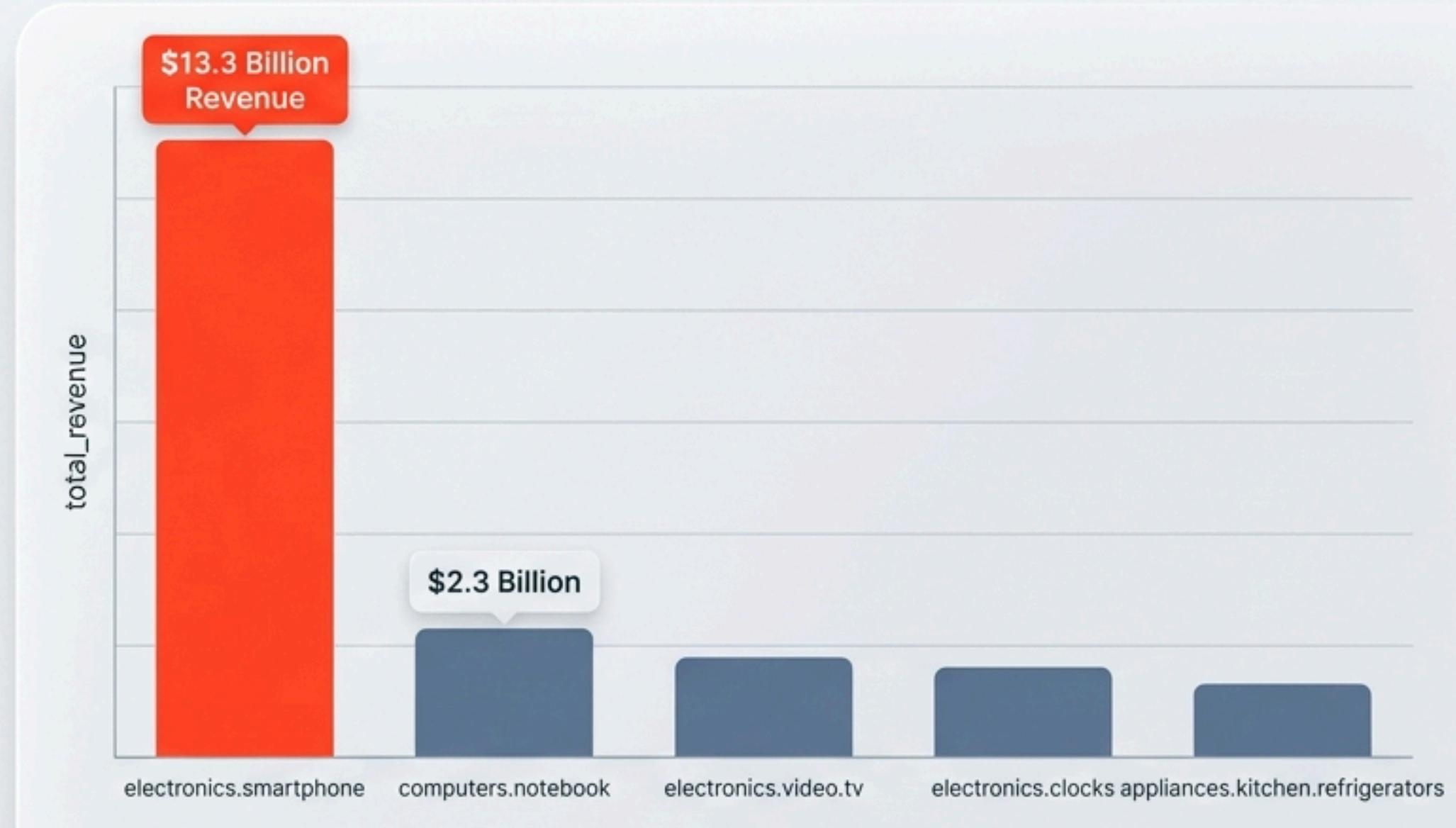
Visualizing Volatility: Automatic Trend Detection



Category Dominance: Aggregating Revenue at Scale

Complex joins and aggregations performed via a single question.

What are the top 5 brands by number of purchases over time?



Cause

Effect

The AI Analyst: Automated Contextual Insights

Genie does not just visualize data; it analyzes it, offering textual summaries that highlight anomalies and dominance.



Market Dominance

Revenue for smartphones is more than five times higher than the next highest category, indicating a strong dominance in sales.



Anomaly Detection

The most significant increase occurred on November 17, 2019, with 185,195 purchases.



Shifting Gears: Engineering the NLP Pipeline

Transitioning from Genie's analysis to Mosaic AI's engineering environment.

▷ Run

```
# Simple sentiment analysis or text classification
!pip install torch
!pip install transformers
from transformers import pipeline

sentiment_analyzer = pipeline("sentiment-analysis")
result = sentiment_analyzer("Databricks notebooks are very useful for data science.")
display(result)
```

The Obstacle: Handling Dependency Conflicts

▷ Run

```
# Simple sentiment analysis or text classification
!pip install torch
!pip install transformers
from transformers import pipeline

sentiment_analyzer = pipeline("sentiment-analysis")
result = sentiment_analyzer("Databricks notebooks are very useful for data science.")
display(result)
```

NameError: name 'transformers' is not defined

The environment throws a compilation error during pipeline execution.

The Fix: AI-Assisted Troubleshooting

Databricks Assistant diagnoses and patches the environment automatically.

A screenshot of a Databricks Notebook interface demonstrating AI-assisted troubleshooting. On the left, a code cell contains Python code for sentiment analysis:

```
# Simple sentiment analysis or text classification
+ %pip install torch
!pip install transformers
from transformers import pipeline

sentiment_analyzer = pipeline("sentiment-analysis")
result = sentiment_analyzer("Databricks notebooks are very
    useful for data science.")

display(result)
```

The code cell includes a dashed red arrow pointing from the error message in the Assistant window back to the line where `%pip install torch` is present. A "Run" button is visible at the top of the code cell.

On the right, an "Assistant" window provides diagnostic information and a fix:

Assistant

The root cause of your error is that the transformers pipeline requires the torch package, but torch is not installed...

I have fixed your code by adding `%pip install torch`.

Fix ✓

Pipeline Executed: Generating Sentiment Scores

```
Downloading model.safetensors: 100% [===== >] 268M/268M
```

```
Downloading tokenizer_config.json: 100% [===== >]
```

label	score
POSITIVE	0.87848...

Model artifacts downloaded and inference complete in < 2 minutes.

Production Standards: Integrating MLflow

```
# Log to MLflow  
import mlflow
```

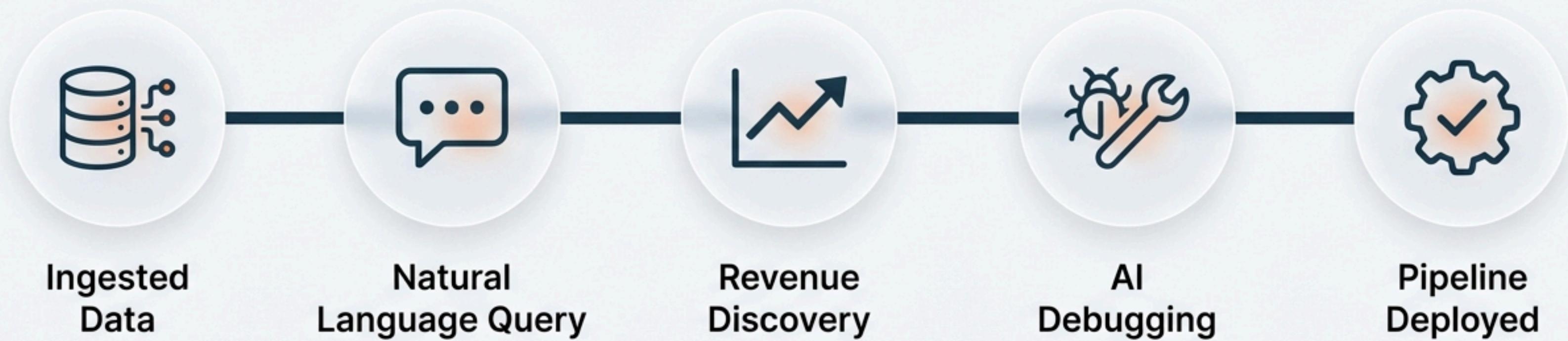
Assistant

I updated your code to add
'import mlflow' to resolve the
NameError.

```
with mlflow.start_run(run_name="sentiment_model"):  
    mlflow.log_param("model", "distilbert-sentiment")  
    mlflow.log_metric("accuracy", 0.95)
```

The code is now production-ready with automated tracking and metrics logging.

Mission Accomplished: The Day 14 Summary



Complex data science workflows reduced to conversational queries and AI-guided debugging.