

Module 8: Capstone

Lab

Applying Safe Enterprise AI Pipeline Blueprints (15:15 - 17:00)

End-to-End Ingestion · Programmatic Safeguards · Executive Summarization



The Capstone Mission

Deploying programmatic boundaries over an unpolished, raw online customer dataset



Execution Milestones

1. Sandbox Pipeline

Establish local data controls to download, clean, and process the dataset safely. Compute metrics programmatically to maintain complete accuracy.

- ✓ Programmatic Ingestion
- ✓ Explicit Math Layer

2. Guarded Summarization

Activate the local privacy filter to mask target user IDs before generating prompts. Leverage AI strictly for stylistic linguistic narrative generation.

- 🛡️ PII Column Elimination
- ✂️ Executive Storytelling

The Target Dataset

Live IBM Telco Dataset

A standard 7,043 record telecom directory stored on a public GitHub server. Ideal for verifying secure, production-grade cloud workflows.



Format: Raw CSV



Public Repository Link

Audited Hazards

Contains dangerous text representations inside numeric zones (e.g., missing spaces in billing indices) and sensitive, clear-text identifiers.



Object-typed Numeric Fields



Raw Customer Tracking IDs

1. Raw Data Ingestion

Pandas Connection & Audit

Load the live online CSV dataset directly via its raw URL using Pandas.

Data Issue Resolved:

The 'TotalCharges' column contains raw string spaces. We replace them and convert the column to float.

```
import pandas as pd

# Raw Online
url = "https://raw.githubusercontent.com/IBM/telco-customer-churn-on-icp4d/master/data/Telco-Customer-Churn.csv"
df = pd.read_csv(url)

# Fix unpolished numeric
df['TotalCharges'] = pd.to_numeric(
    df['TotalCharges'].replace(" ", None)
)
df.dropna(subset=['TotalCharges'], inplace=True)
```

2. Activating the Privacy Shield

Excluding Direct PII

Prior to framing prompts for external AI workspace APIs, we mask sensitive customer IDs and drop identifiable demographics.

 Unique Token Mapping

 Demographic Scrim

```
def apply_privacy_shield (dataframe):  
    safe_df = dataframe.copy()  
  
    # Mask direct tracking  
    safe_df['customerID'] = [  
        f"ANON-ID-  
        {i:04d}" for i in range(len(safe_df))  
    ]  
  
    # Eliminate tracking  
    safe_df = safe_df.drop(  
        columns=['gender', 'Partner']  
    )  
    return safe_df  
  
shielded_df = apply_privacy_shield(df)
```

3. Deterministic Pre-Calculations

Isolating Analytical Facts

Calculate segment summaries locally. The output payload will serve as the prompt context, locking the LLM to true arithmetic results.

- ✓ Math computed in Python
- ⚠ Zero AI calculation drift

```
# Extract key segment
total_records = len(shielded_df)
metrics
churn_rate = (shielded_df['Churn'] == 'Yes').mean() * 10
avg_monthly = shielded_df['MonthlyCharges'].mean() 0

# Format local calculations context
raw_metrics = f"""
SEGMENT DETAILS: - Account Count:
{total_records} - Overall Churn Rate:
{churn_rate:.2f}% - Average Billing Charge:
${avg_monthly:.2f} """

print(raw_metrics)
```

4. Prompt Context Construction

Linguistic Narrative Output

Inject the pre-calculated numerical context into a highly structured system prompt structure.

Core Rule Set:

The AI is forbidden from computing metrics. It strictly serves to translates dry calculations into business strategy narratives.

```
prompt = f'''You are an executive strategy advisor. Convert these clean  
Python metrics into an update. Context Metrics: {raw_metrics}  
Requirements: 1. Divide narrative into 'Operational Impact' and  
'Strategic Intervention Priorities'. 2. Do not explain variables;  
focus on actionable executive summaries. '''
```

```
print(prompt)
```

Lab Wrap-Up & Next Steps

Troubleshooting Syntax

Common syntax errors involve type-casting on masked data columns or string split actions. Ensure Pandas datatypes are fully mapped prior to formatting LLM context vectors.

- ❓ Check `.info()` for Object categories
- ❓ Verify PII column exclusions

Resource Distribution

Secure your dynamic templates, prompt libraries, and database connectors locally. These pipeline modules serve as an elite starting point to drive secure AI architectures throughout your career journey.

- ✅ Export notebook blueprints
- ✅ Launch your local agent ecosystem

Grand Finale Q&A

Solving remaining syntax questions & toolkit distribution

Troubleshooting Breakouts · Modular Blueprints · Your Enterprise Strategy Journey