Module 7: AI Executive Reporting Techniques & Data Privacy

Hands-on Lab Worksheet & Conceptual Verification

Environment: Anaconda Jupyter Notebook & Python 3 (Time Parameters Omitted)

Course Objectives

This hands-on workshop walks through bridging local data workflows with public AI infrastructure responsibly. By completing this worksheet, you will master:

- **Automated Summaries:** Efficiently transferring structured math outputs from Python engines to LLMs to build clean, executive bullet-point reports.
- **Contextualizing Insights:** Translating static performance records and dry metrics into dynamic corporate operational narratives.
- **The Privacy Shield:** Constructing script boundaries to systematically mask data and filter variable attributes prior to public AI processing.

Part 1: Mock Data Setup

Execute the following script block within your local environment to construct an enterprise baseline containing financial parameters mixed with sensitive **Personally Identifiable Information (PII)**.

```
import pandas as pd
import numpy as np

np.random.seed(42)

data = {
    'customer_id': ['CUST-101', 'CUST-102', 'CUST-103', 'CUST-104', 'CUST-105'],
    'customer_name': ['Sarah Jenkins', 'Michael Chang', 'Amanda Ross', 'David Smith',
'Elena Rostova'],
    'email': ['sjenkins@corp.com', 'mchang@tech.io', 'amanda.r@services.net',
'dsmith@global.org', 'elena.r@fin.ru'],
    'region': ['North America', 'APAC', 'North America', 'EMEA', 'EMEA'],
    'q2_revenue_usd': [120000, 85000, 250000, 45000, 190000],
    'q3_revenue_usd': [115000, 92000, 210000, 48000, 205000],
    'churn_risk_score': [0.2, 0.1, 0.75, 0.3, 0.85]
}

df = pd.DataFrame(data)
print("--- RAW ENTERPRISE DATA LOADED ---")
```

Part 2: The Privacy Shield (Data Anonymization & Masking)

Exercise 1: Complete the operational workflow function below to guarantee that identification strings are completely sanitized before generating prompts for external language frameworks.

```python
def apply_privacy_shield(dataframe):
    shielded_df = dataframe.copy()

    # 1. Pseudonymize unique customer IDs globally
    shielded_df['customer_id'] = [f"ANON-ID-{i:03d}" for i in range(1, len(dataframe) +
1)]

    # 2. TASK: Add masking configuration for the 'customer_name' series
    shielded_df['customer_name'] = [f"Client_{chr(65+i)}" for i in range(len(dataframe))]

    # 3. TASK: Completely drop high-exposure variables (e.g., 'email')
    shielded_df = shielded_df.drop(columns=['email'])

    return shielded_df

safe_df = apply_privacy_shield(df)
```

Part 3: Mathematical Pre-Calculations (AI Reporting Execution Layer)

To avoid calculation hallucinations common in Large Language Models, aggregations and arithmetic transformations must always stay within native Python compilation blocks prior to linguistic hand-off.

```python
total_q2 = safe_df['q2_revenue_usd'].sum()
total_q3 = safe_df['q3_revenue_usd'].sum()
revenue_growth_pct = ((total_q3 - total_q2) / total_q2) * 100

high_risk_clients = safe_df[safe_df['churn_risk_score'] > 0.60]['customer_name'].tolist()
avg_churn_risk = safe_df['churn_risk_score'].mean() * 100

raw_calculations_context = f"METRICS CONTEXT:\n- Total Q2 Revenue: ${total_q2:,.2f}\n-
Total Q3 Revenue: ${total_q3:,.2f}\n- QoQ Revenue Growth Rate: {revenue_growth_pct:.2f}%
\n- Average Portfolio Churn Risk: {avg_churn_risk:.1f}%\n- High Churn Risk Accounts:
{high_risk_clients}"
```

Part 4: Contextualizing Insights & Automated Briefings

> **Instruction Workflow**
>
> Execute the string generation below inside your notebook. Copy the terminal print output text and prompt it into your enterprise AI assistant to evaluate the resulting professional, corporate summary asset.

```
llm_prompt = f"""
You are an elite corporate strategy advisor. Transform the following raw Python
calculations into a concise, professional executive narrative update.

Rules:
1. Do NOT reveal or seek out any actual customer personal identities (Use the provided
masked labels).
2. Group insights into two clear bulleted sections: 'Financial Vector Performance' and
'Risk Mitigation Priorities'.
3. Translate dry numbers into narrative strategy adjustments (e.g., relate negative QoQ
drop to corrective planning measures).

{raw_calculations_context}
"""
print(llm_prompt)
```

Part 5: Reflection & Conceptual Verification

Question 1

Why is it critical to compute totals, metrics variations, and statistical aggregations inside Python/Pandas environments before passing structural data profiles out to an LLM?

Question 2

When sharing a bare relational database layout structure (a Schema file without records) with cloud utilities to produce SQL code blocks, which privacy safeguards should remain active?