**Julio Pochet**  
**CSD370 – Module 9.2 Assignment:**   
**Test Data Generation**

**🔹 1. Synthetic Test Data Generation**

**Method:**  
Synthetic test data is artificially created to imitate the structure and patterns of real-world data, without containing any actual sensitive information. It can be generated manually or using tools that automate the process based on schemas or data models.

**Scenario:**  
This method is ideal for testing applications that handle private data (for example: finance, healthcare) where using actual user data could violate privacy regulations. It's also helpful when real data is unavailable or incomplete during early development.

**Tools:**

* [Mockaroo](https://mockaroo.com/)
* [Tonic.ai](https://www.tonic.ai/)

**Source:**  
Tonic.ai – Guide to Synthetic Test Data Generation  
🔗 <https://www.tonic.ai/guides/guide-to-synthetic-test-data-generation>

**🔹 2. Randomized Data Generation**

**Method:**  
This method generates randomized values—such as names, phone numbers, emails, or usernames—within certain constraints. It’s useful for testing how systems handle diverse and unpredictable inputs.

**Scenario:**  
Commonly used for form validation, API testing, or fuzz testing login pages, where developers want to see how the system reacts to both typical and edge-case inputs.

**Tools:**

* [Faker (Python)](https://faker.readthedocs.io/en/master/)
* [JavaFaker](https://javadoc.io/doc/com.github.javafaker/javafaker/latest/com/github/javafaker/Faker.html)
* [RandomUser.me API](https://randomuser.me/)

**Source:**  
Faker – Python Documentation  
🔗 <https://faker.readthedocs.io/en/master/>

**🔹 3. Production Data Masking (Anonymization)**

**Method:**  
This technique takes real production data and masks sensitive elements like names, credit cards, and SSNs. The goal is to preserve data relationships and structure while protecting user privacy. It's often used for secure, high-fidelity test environments.

**Scenario:**  
Ideal for enterprises that need to test using realistic data while remaining compliant with GDPR, HIPAA, or other data protection laws. For example, a healthcare provider testing EHR systems can use masked patient records to simulate real-world scenarios safely.

**Tools:**

* [DataVeil](https://www.dataveil.com/)
* Microsoft SQL Server Dynamic Data Masking
* Delphix

**Source:**  
DataVeil – Data Masking Software  
🔗 <https://www.dataveil.com/>