**Hackathon Prompt: Build a GenAI System Integration Hub for Barclays Using OpenAI and Streamlit**

**Objective**: Create a user-friendly platform called the **GenAI System Integration Hub** that leverages OpenAI’s GPT-4 and Streamlit to streamline data access and insights across Barclays’ data ecosystem. This platform should allow users to ask natural language questions about the data and receive relevant data summaries, statistics, visualizations, and actionable insights, all while respecting access controls.

**Key Project Components**

**1. Core Functionalities:**

• **Natural Language Querying**: Users should be able to ask questions in plain English, and the system should interpret their intent and identify the relevant datasets.

• **Intelligent Data Selection**: The AI should automatically select and connect to the required datasets via ODBC or API connectors based on the user’s query.

• **Data Analysis and Summarization**: The platform should calculate statistics, create summaries, and suggest visualizations (e.g., bar charts) as appropriate.

• **Access Control and Permissions**: Access to different data sources is controlled by Barclays’ Active Directory permissions, ensuring that only users with the correct permissions can access specific data.

• **Insightful Responses and Follow-Up Questions**: The system should provide responses similar to a data analyst’s insights, allowing users to ask follow-up questions for deeper analysis without starting over.

**2. Technical Architecture:**

• **AI Engine**:

• Utilize GPT-4 via OpenAI’s API to interpret user questions, determine required datasets, and provide insightful analysis.

• Use a data dictionary (data\_dictionary.json) to map dataset names and columns, allowing GPT-4 to dynamically understand the data structure.

• **Data Loading and Processing**:

• Design a query\_handler.py file that uses the data dictionary to dynamically load and combine data files based on user queries.

• Support various data formats like CSV (stored in a data folder), with an option to connect to live databases in the future.

• Manage dataset joins automatically by linking keys (e.g., account\_id, customer\_id) based on user query requirements.

• **Visualization and Summaries**:

• Provide visual summaries (e.g., bar charts) based on query requirements. For example, if a user asks for the “average transaction amount,” generate an appropriate bar chart.

• **Frontend Interface with Streamlit**:

• Build a user interface that allows users to type questions, select roles, and submit queries.

• Display data previews, summary statistics, visualizations, and conversation history in an intuitive layout.

• Style the interface with Barclays’ brand colors and logo, creating a professional and visually appealing design.

• **Governance and Access Requests**:

• For users without required access permissions, notify them of the datasets they need and provide instructions on how to request access.

**Dataset Integration Scope**:

The GenAI System Integration Hub connects and leverages **30 key datasets** across Barclays, providing a holistic view of critical information. These datasets span multiple departments and functions, including customer demographics, transaction details, risk assessments, credit scores, and account mappings. By integrating these datasets, the platform empowers users to conduct cross-functional analyses and gain a comprehensive understanding without needing to know the technical details of each dataset.

Each dataset is documented within a **data dictionary** that the AI engine references to dynamically identify, load, and merge relevant information as needed. This architecture allows the system to autonomously decide the appropriate datasets to answer any given query, ensuring efficient data retrieval and minimizing user effort.

**3. Step-by-Step Setup Guide:**

• **Repository Structure**:

• Organize the repository as follows:

GenAI-System-Integration-Hub/

├── src/

│ ├── app.py # Main Streamlit app

│ ├── query\_handler.py # Handles query interpretation and data processing

├── data/

│ ├── credit\_transactions.csv

│ ├── debit\_transactions.csv

│ ├── customer\_risk\_markers.csv

│ ├── account\_customer\_mapping.csv

├── data\_dictionary.json # JSON file listing datasets and columns

├── requirements.txt # Required Python packages

└── README.md # Project documentation

• **Installation Instructions**:

• Clone the repository.

• Install required packages: pip install -r requirements.txt.

• Add your OpenAI API key in query\_handler.py.

• **Running the App**:

• In the terminal, navigate to the src folder.

• Run the Streamlit app: streamlit run app.py.

**4. Barclays Core Values in the System Design**

• **Respect**: Ensures a secure, user-centric experience by adapting responses to user roles and respecting data access limitations.

• **Integrity**: Offers full transparency with traceable data sources and an auditable process, instilling trust in the insights provided.

• **Service**: Improves internal efficiency, allowing Barclays employees to focus on higher-impact tasks.

• **Excellence**: Leverages cutting-edge AI to deliver quick and accurate insights.

**5. Sample User Stories**

• As a **risk analyst**, I want to ask for “customer risk profiles by region” and receive an insightful summary along with a regional distribution chart.

• As a **manager**, I need to analyze “total transaction amounts for high-risk customers” to support compliance reporting.

**6. Example Queries and Expected Responses**

• **Query**: “Calculate the average transaction amount for high-risk customers.”

• **Expected Output**: Summary statistic (average amount), table preview, and a bar chart showing transaction distribution.

• **Query**: “Show the distribution of customer types by risk marker.”

• **Expected Output**: A summary of customer counts by risk marker, and a bar chart showing the distribution.

This prompt should serve as a clear blueprint for the hackathon project, with guidance on objectives, technical setup, and expected functionalities. Let me know if you need further clarification or additional sections for this pitch.