Proposed Sandbox Activities

1. **Navigating the Data Catalog**
    - **Browse and Search Data Assets:** Participants will use the search and browse functionality to locate specific datasets relevant to telecom data (e.g., "Customer Usage Data", "Network Performance Metrics", "Billing Records").
    - **Exploring Metadata Details:** Upon selecting an asset, participants will review the dataset description, tags, owners, data sources, related columns, and classifications.

Questions:

- How did browsing the data catalog and using the search function differ from the manual methods you might have used previously?
- In what ways did having a central metadata repository make it easier to understand what data assets exist?

2. **Exploring Data Lineage**
    - **View Lineage Graphs:** Participants will select a critical dataset and explore its lineage to see upstream sources (like CRM systems or network event logs) and downstream consumption points (like BI dashboards or churn prediction models).
    - **Impact Analysis Exercise:** Participants identify how a change in an upstream source column (e.g., “customer_plan_type”) could affect downstream reports or analytics. They will trace this impact visually on the lineage graph.

Questions:

- After viewing the lineage graphs, how has your understanding of data dependencies and transformations changed?
- How would the ability to visualize lineage and perform impact analysis help you make more informed decisions when changes occur in upstream systems?

3. **Data Quality Checks and Profiling**
    - **Profile a Dataset:** Participants run a profiling job on a sample customer data table to view data distribution, uniqueness, and completeness metrics.
    - **Create a Quality Rule:** They define a simple data quality rule (e.g., "Email column should not contain null values") and run it against the dataset to understand how data quality checks are configured and monitored.

Questions:

- What insights did you gain from running data profiling and quality checks that you would not have had before?
- How might the continuous monitoring of data quality influence how you trust or use certain datasets?

4. **Business Glossary and Tags**
    - **Review Glossary Terms:** Participants explore a pre-defined business glossary for the telecom domain (e.g., definitions of "ARPU - Average Revenue Per User" or "Network Node"). They will see how terms are linked to datasets and columns.
    - **Add Tags to a Dataset:** Participants apply relevant tags (e.g., "PII" or "Billing Data") to a sample dataset, learning how classification enhances discoverability and governance compliance.

Questions:

- How did having a formal business glossary and tagging capability improve clarity around key telecom metrics and terms?
- Can you give an example of how using standardized definitions could resolve misunderstandings or inconsistencies within your team?

5. **Ownership and Stewardship**
    - **Identify Data Owners:** Participants locate the designated owners and stewards for a given dataset. They can see who's responsible for ensuring data quality and accuracy.
    - **Assign Roles and Responsibilities:** In a safe sandbox environment, participants simulate assigning a data owner or steward to a dataset or a domain, understanding how roles support accountability.

Questions:

- How did identifying data owners and stewards clarify accountability and responsibility within the data ecosystem?
- If your team adopted these ownership models, how might it change the way data issues are addressed?

6. **Access Control and Policy Review**
    - **Examine a Data Policy:** Participants review a mock data access policy (e.g., restrict access to sensitive fields like customer SSN). They see how policies are defined and enforced within the tool.

    Questions:

- How did role-based access control (RBAC) and data policies help you understand the importance of regulated data access?

7. **Collaboration and Annotation**
    - **Comment on a Dataset:** Participants will leave comments, questions, or annotations on a dataset page and mention teammates (e.g., "@Alice – can you verify the correctness of the churn rate calculation?").
    - **Follow Data Assets:** They subscribe to changes on a critical dataset so they receive notifications if metadata or classifications are updated, helping understand how the tool supports ongoing governance engagement.

    Questions:

- Did leaving comments or subscribing to dataset changes improve communication among team members? How might this feature streamline issue resolution in the future?
- How do you foresee using these collaboration features in your day-to-day workflows?

8. **Monitoring and Reporting**
    - **View Data Usage Statistics:** Participants explore dashboards or reports showing which datasets are most frequently accessed, helping them understand popularity and adoption metrics.
    - **Data Quality Trend Reports:** They examine how data quality scores or alert counts have trended over time, reinforcing the importance of ongoing governance and monitoring.

    Questions:

- Comparing the “before” scenario (without governance tools) to now, which challenges do you think are most effectively addressed by OpenMetadata features?
- What aspects of the data governance tool would you prioritize implementing in your organization to achieve similar benefits?