Microsoft Fabric Project: DP-600 Prep with Genk Match Data

Step 1: Ingest Your Data

- Upload 'season_23_24.csv' and 'season_24_25.csv' into your Microsoft Fabric Lakehouse.

- Optionally create shortcuts or use Dataflows Gen2 to ingest from external sources like GitHub or OneLake.

Step 2: Transform Using Notebooks

- Open a Fabric Notebook.

- Read and clean both CSVs with pandas or PySpark.

- Add a 'season' column if missing.

- Combine both datasets and standardize column names.

- Save the final DataFrame as a Lakehouse table.

Step 3: Serve with SQL Endpoint

- Use the SQL Endpoint to query the Lakehouse table.

- Example: SELECT season, venue, result, COUNT(*) FROM genk_matches GROUP BY season, venue, result.

Step 4: Model the Data

- Create a semantic model.

- Define measures (e.g., Goals Per Match, Average Attendance).

- Use calculated columns (e.g., Goal Difference = GF - GA).

- Establish relationships if more tables are added.

Step 5: Visualize in Power BI

- Connect Power BI to your Lakehouse or SQL Endpoint.

- Build visuals for match trends, performance comparison, attendance patterns.

- Add slicers for filtering by season, round, opponent.

Step 6: Automate with Pipelines

- Set up a Data Pipeline to refresh the dataset automatically (e.g., simulate daily updates).

- Trigger Notebook transformations and data loads.

Step 7: Monitor and Secure

- Use Lineage View to trace data flow from ingestion to Power BI.

- Apply Row-Level Security (RLS) to restrict data by team or season.

- Set up Data Activator rules (e.g., alert if xG > 2 and result = L).

Step 8: Practice DP-600 Scenarios

- Manage workspace permissions and deployment pipelines.

- Document your solution with screenshots and lineage.

- Simulate role-based access and workspace promotion (Dev Test Prod).