**Technical interview**

**Technical Interview – AI Product Analysis**

The purpose of this technical interview is to analyze the output of an AI product that forecasts sales at the country-brand level.

Each month, the AI model forecasts the **volume** for the next 3 years using historical monthly data. Once the forecast is generated, the volume is converted into **sales ($)** using prices manually provided by stakeholders in the system.

Each execution of the AI model is versioned using the format R\_MONTH\_YEAR (e.g., R\_JUNE\_2022, R\_JULY\_2022). To ensure everything has run correctly, the outputs of each new version are compared to those of the previous month to identify any unexpected changes or issues in the process.

For this exercise, you will compare the results of R\_JULY\_2022 with the previous run, R\_JUNE\_2022.

You are provided with **three .parquet files**:

**1. Actuals.parquet**

* Contains historical actual **volume data** at monthly level for each country-brand combination (identified by the id column).
* The run column refers to the version of the execution.
* Sales last year ($M) shows the total actual sales (in millions of dollars) for each id over the past 12 months.

**2. Prices.parquet**

* Contains the **forecasted prices** for future dates at the id-monthly level.
* Columns R\_JUNE\_2022 and R\_JULY\_2022 represent the prices provided by stakeholders for each execution version.

**3. Forecast.parquet**

* Includes the **forecasted volume** (volume\_fcst) and corresponding **sales in $** (sales\_fcst) for each id and date, computed using the stakeholder prices.
* The run column indicates the version of the forecast.

**Objective**

Your goal is to:

* **Compare** the July 2022 (R\_JULY\_2022) output against the June 2022 (R\_JUNE\_2022) version.
* **Identify** any inconsistencies, data quality issues, or changes that could indicate a problem in the July run.
* Be **proactive** in highlighting any key discrepancies, reporting important cases, and proposing **potential data validation checks** that could help avoid similar issues in future runs.

You are expected to:

* Share a **notebook or Python script** used for the analysis.
* Present your findings during the live interview in the format of your choice (e.g., plots, tables, commentary).