

ABInBev

DATA ENGINEERING

RECRUITMENT USE CASE

DISCLAIMER

This document is property of AB InBev and usage is limited to recruitment purposes. This should not be shared externally by ABI employees or external candidates.

The tasks and scenarios contained in the exercises are fictional and exclusive for technology proficiency evaluation during recruitment processes. The candidates will not be asked to submit or share solutions with ABI employees. Similarly, the results will not be used by ABI for business purposes.

This evaluation does not represent any form the guarantee of a consequent employment offer or limited time work relation.

Candidates are will not obligated to perform the exercise and are free to present results, or not, during the recruitment process.

MAIN INSTRUCTIONS

WHAT

This test is composed by two exercises related to data & analytics. The main objective is to evaluate candidate ability to respond to data problems and build solutions using data engineering techniques.

HOW

Execution can be performed using any infrastructure or platform of candidate choice and availability.

Suggested environments: Azure, AWS, Google Cloud or other cloud providers can be a good option. Stand alone tech stacks on personal notebook/PC are also accepted, such as MySQL, Postgree, Jupyter, Spark/Scala terminal.

WHEN

At candidate's choice and time availability. The exercises shouldn't take more than 2 or 3 hours to be executed. Results will be presented in a 45 minutes session to be agreed between candidate and recruiters.

Presenting the results

- 45 minutes session → 30 min presentation for architecture and solution (code & results) + 15 minutes for Q&A.
- Schedule to be agreed between candidate and recruiters.
- Presentation decks or PowerPoint material are not required. The focus here is to demonstrate coding and solution skills =)

Evaluation criteria

1. **Coding Skills** – Ability to solve problems using proposed technology: usage of libraries, functions, tech functionalities, performance approach, code reutilization and coding standards (indentation, documentation).
2. **Solution Architecture** – Data structures designed to answer proposed problems. Aspects evaluated: methodologies applied, cardinality, standards, naming conventions, data types, relationships.
3. **Problem Solving Skills** – What was the extension and breath on the proposed solution? How much the proposed solutions is aligned with the business usage and purpose? Is it scalable? Does it support big volumes of data, if not what are the next steps?
4. **Tech presentation Skills** – Candidate ability to explain the “though process”, key aspects, tradeoffs or even specific bottlenecks of proposed solution.

SCENARIO

Consider you are a data engineering for a fictional beverage startup:

- After the sprint planning, two tasks were assigned to you by the project squad.
 - Along with the data solution, they also expect you to use your own data architecture and infrastructure, that later can become the official data platform.
-

TECH STACK

- **Programming Language** → The company is working to create a data platform. The only requirement from the data architect is to stay under three data engineering frameworks. No other language can be used.
 1. SQL – basic data analysis framework available
 2. Python – most used framework to build data engineering pipelines.
 3. Pyspark/Scala – Top notch framework used by productionized and performant pipelines.
- You can use any combination of those frameworks to create your solution. If you have knowledge in more than one, prefer to use the highest # in the scale. For instance, if you are experienced in SQL and Python, prefer Python. If you Know Python and Pyspark, use Spark.
- **Interface** → At your choice. The most important for the MVP is to prove that you can resolve the data problem using the proposed languages. For instance, you can implement using Jupyter notebooks, Ambari queries, SQL queries, Data Bricks notebooks, Zeppelin Notebooks or any other data implementation environments.
- **Infrastructure** → You are free to introduce your own databases and infrastructure to design your solution. You can propose data lake/blob storage solutions, Hadoop based data lakes, high performance SQL based databases such as Big Query, Synapse, Redshift or even local stand alone instances connecting directly to the .CSV files. You are welcome to demonstrate the top of your knowledge. Remember that as much reliable and robust your solution can be, better accepted it will be by the fictional project squad.

BUSINESS CASE 1 – Beverage Sales

Briefing

Your mission is to build an MVP for a Beverage Sales analytics platform.

Data Resources

1. Beverage Sales → abi_bus_case1_beverage_sales.csv
2. Beverage Channel Features → abi_bus_case1_beverage_channel_group.csv

Requirements

1. Build a data pipeline to merge the beverage channel features available in the interface #2 with the transactional sales from interface #1.
Expected result: Data from both .CSV are merged correctly and can be grouped or transformed.
2. Implement a data ingestion pipeline for the beverage sales data using dimensional data modeling. The solution must contain dimensions (at least two), fact (at least one) and especially summary tables based on the data provided. Your approach should be able to answer most of the business questions based on the KPIs provided. Implement and perform all the transformations and aggregations you find important. Make sure you can understand and explain your design decisions for the fictional team members represented by the recruiters.
Expected result: Physical data model implemented with 100% of data provided ingested.
3. What are the next steps, enhancements and future features that could be added to your solution after the MVP?
Expected result: List of items with enhancements and features expected in the future to improve proposed MVP.

BUSINESS CASE 1 – Beverage Sales

Briefing

Your mission is to build an MVP for a Beverage Sales analytics platform.

Data Resources

1. Beverage Sales → abi_bus_case1_beverage_sales.csv
2. Beverage Channel Features → abi_bus_case1_beverage_channel_group.csv

Requirements

4. Create queries/scripts that can answer following business questions using the data structure created:

Expected result: Data processes and results.

- 4.1 What are the Top 3 Trade Groups (TRADE_GROUP_DESC) for each Region (Btlr_Org_LVL_C_Desc) in sales (\$ Volume)?
- 4.2 How much sales (\$ Volume) each brand (BRAND_NM) achieved per month?
- 4.3 Which are the lowest brand (BRAND_NM) in sales (\$ Volume) for each region (Btlr_Org_LVL_C_Desc)?

Important: The queries/scripts need to be based on your MVP created in item #2, not the source files.