

Scouting Insight Officer Pre-Task

You will be provided with:

- 2 CSV files (Players.csv, ReportingInsight.csv).

What is expected throughout the tasks:

- A presentation slide deck summarising your findings and answers from **all the tasks**. This deck will be used at the interview stage.
- You can explore and manipulate the data using any program or coding method you're most comfortable with. The purpose of these tasks is to help us understand your thought process.

Task 1.1: Player Data Audit & Cleaning

Scenario

You have been given a raw export of Player data (**Players.csv**). This dataset is known to contain issues with duplicates, missing values and inconsistencies. Your job is to **clean** the data to create a reliable and accurate player table.

Your job is to:

- Use any method to explore the dataset and identify the problems.
- Dedupe the data to create a reliable player table.
- Produce a summary that outlines the state of the data before and after your cleaning.

What's Expected

- **Document your approach:** provide explanations of what you did and why.
- **Identify and fix issues:** detect and correct duplicates, inconsistent naming, and other data problems.
- **Show impact:** include metrics showing before and after improvements (e.g. number of duplicates removed).
- **Final Table:** a "cleaned" version of the players table.

Task 1.2: Handling PlayerID Updates Across Related Tables

Scenario

In a real-world environment, player data would typically live within a structured database containing multiple related tables. For example, Players, Reports, Contracts, and Appearances. These tables are linked by shared fields such as PlayerID.

During your audit, you notice that several players appear more than once in the dataset, each with a different PlayerID. Once you've identified which PlayerID should be retained for each player, you'll need to consider how these changes would cascade through the wider database to maintain consistency and data integrity.

Outline how you would approach this ensuring that all related tables remain accurate when PlayerIDs are merged or updated.

Your response should outline:

- How you would identify which tables or datasets are affected by PlayerID changes.
- How you would ensure that relationships between tables remain correct after updates or merges.
- Steps you would take to make this process consistent and repeatable for future data updates.

Task 2: Scouting Insights Visualisation

Scenario

You are given a cleaned dataset, ReportingInsight.csv, containing scouting and reporting information.

Your job is to:

Explore the dataset and create visuals that surface meaningful and interesting insights. You decide what to investigate and how to present it.

What's Expected

- **Tool choice:** Use Tableau, Power BI or any visualisation tool you prefer.
- **Exploration:** Analyse the dataset to find patterns, trends, outliers or stories you think matter.
- **Explanation:** Include a written summary that highlights your key insights and why you chose those visuals.
- **Screenshots:** Screenshots of your visuals (from Tableau, Power BI or similar).