**CAPM Workflow for Question Library**

**Overview**

Manages the CAPM question library by extracting raw questions, formatting them, merging into questions.json, removing duplicates, and validating the quiz application. Ensures data integrity through backups and verification.

**Metadata Requirements for Scripts**

All scripts (\*.py, \*.bat) must include:

* **File Name**: Script name (e.g., remove\_duplicates.py).
* **Owner**: Andrew.
* **Purpose**: Brief function description.
* **Version Control**: Version number (e.g., v1.0).
* **Change Log**: Creation date and updates with descriptions.

**Extraction of Data**

**Purpose**: Process raw questions from drop\_the\_format\_questions.txt to questions.json.

* **Verify Script Version**:
* type "M:\OneDrive\Documents\GitHub\excel\_to\_json\conversion\_script.py" | findstr "# Version Control:"

[Note: Replace conversion\_script.py with actual script in run\_conversion.bat.]

* **Run Conversion**:
  + Process drop\_the\_format\_questions.txt:
  + M:\OneDrive\Documents\GitHub\excel\_to\_json\run\_conversion.bat
  + Generates log in M:\OneDrive\Documents\GitHub\excel\_to\_json\logs\ (e.g., conversion\_log\_20250805\_1656xx.txt).
* **Clear Input File**:
* echo. > "M:\OneDrive\Documents\GitHub\excel\_to\_json\data\drop\_the\_format\_questions.txt"
* git add "M:\OneDrive\Documents\GitHub\excel\_to\_json\data\drop\_the\_format\_questions.txt"
* git commit -m "Cleared drop\_the\_format\_questions.txt on 2025-08-05"
* git push

**Backup**

**Purpose**: Create timestamped backups of questions.json to prevent data loss.

* **Run Backup**:
* M:\OneDrive\Documents\GitHub\excel\_to\_json\backup\_script.bat
* **Save**:
* copy "M:\OneDrive\Documents\GitHub\excel\_to\_json\data\questions.json" "M:\OneDrive\Documents\GitHub\excel\_to\_json\backup\data\_backup\_%date:~6,4%%date:~3,2%%date:~0,2%\_%time:~0,2%%time:~3,2%%time:~6,2%\questions.json"
* **Verify**:
* dir "M:\OneDrive\Documents\GitHub\excel\_to\_json\backup\data\_backup\_%date:~6,4%%date:~3,2%%date:~0,2%\_%time:~0,2%%time:~3,2%"

**Duplicate Removal**

**Purpose**: Identify and remove duplicate questions from questions.json, retaining lowest ID.

* **Check Library Size**:
* python -c "import json; print(len(json.load(open('M:\\OneDrive\\Documents\\GitHub\\excel\_to\_json\\data\\questions.json'))))"
* **Identify Duplicates**:
* python "M:\OneDrive\Documents\GitHub\excel\_to\_json\check\_duplicates.py"
* **Remove Duplicates**:
* python "M:\OneDrive\Documents\GitHub\excel\_to\_json\remove\_duplicates.py"
* **Verify** (expect 301 questions):
* python -c "import json; print(len(json.load(open('M:\\OneDrive\\Documents\\GitHub\\excel\_to\_json\\data\\questions\_updated.json'))))"
* python "M:\OneDrive\Documents\GitHub\excel\_to\_json\check\_duplicates.py" # Modify for questions\_updated.json
* **Replace**:
* copy "M:\OneDrive\Documents\GitHub\excel\_to\_json\data\questions\_updated.json" "M:\OneDrive\Documents\GitHub\excel\_to\_json\data\questions.json"
* git add "M:\OneDrive\Documents\GitHub\excel\_to\_json\data\questions.json"
* git commit -m "Removed 7 duplicates (IDs 116, 177, 165, 248, 305, 306, 307) on 2025-08-05"
* git push

**Validation**

**Purpose**: Ensure questions.json consistency and quiz application functionality.

* **Validate Consistency**:
* python -c "import json; questions = json.load(open('M:\\OneDrive\\Documents\\GitHub\\excel\_to\_json\\data\\questions.json')); print([q['id'] for q in questions if q['question'] and not q['explanation']])"
* **Test Quiz Application**: Check buttons, filtering, navigation; log issues.

**Changelog**

* 2025-08-05: Created document, added metadata requirements, sectioned into Extraction, Backup, Duplicate Removal, Validation.
* 2025-08-05: Replaced dir with Get-ChildItem -Recurse.
* 2025-08-05: Noted 'question' field for scripts.
* 2025-08-05: Added check\_duplicates.py, remove\_duplicates.py (v1.4) with detailed report.
* 2025-08-05: Enhanced backup\_script.bat with metadata, fixed log directory error.
* 2025-08-05: Removed test\_questions.xlsx and related scripts.

**Notes for Future Update**

* Clarify conversion script in run\_conversion.bat.
* Fix conversion errors (e.g., "invalid literal for int()").