# Evo Trial Job – Project summary

Following the project description, I can confirm the trial job solution has been all developed in SSIS. In order to execute the SSIS package, I decided to utilize a windows batch file which provides the most effortless way to execute the package independently on a location and working environment.

In the attached Launch.txt file you can find all the further information about how to correctly proceed with the batch file setting and execution. For your reference, below you can see the SSIS Data Flow schema.

A screenshot of a cell phone

Description automatically generated

# Evo Trial Job – Ideas for enhancement

## As already mentioned above, based on the project requirements I developed the whole solution exclusively in SSIS platform. On the other hand, if I would have a chance to utilize some other tools and approaches, I can assure you that we could achieve a way better result. Please find below some of my ideas on how the current SSIS based solution could be enhanced.

## 1. SQL DB based solution

One of the first ideas on how to enhance the current solution would be using SSIS Catalogue for storing the SSIS package and SQL DB as an intermediate for data formatting.

The catalogue provides a possibility to easily deploy new or updated packages together with very convenient SQL Agent tool for scheduling the automated jobs.

The SQL DB is the best environment for any kind of data formatting. As you can see from the SQL Command I quickly drafted for this project, it is a way more straightforward solution than developing the complex SSIS Data Flow package. Besides that, the SQL formatting can provide significantly faster results for big data volumes and it is by far more flexible for solution maintenance and customization.

DROP TABLE IF EXISTS #Temp

SELECT

[ITEM\_CODE],

[CATEGORY\_ID],

[CATEGORY\_VALUE],

DENSE\_RANK() OVER(ORDER BY [CATEGORY\_ID],[CATEGORY\_VALUE] DESC) AS [ID]

INTO #Temp

FROM (

SELECT [ITEM\_CODE], 1 AS [CATEGORY\_ID], TRIM(LOWER([CATEGORY\_MAIN])) AS [CATEGORY\_VALUE] FROM [input]

UNION

SELECT [ITEM\_CODE], 2 AS [CATEGORY\_ID], TRIM(LOWER([CATEGORY\_SECONDARY])) AS [CATEGORY\_VALUE] FROM [input]

UNION

SELECT [ITEM\_CODE], 3 AS [CATEGORY\_ID], TRIM(LOWER([CATEGORY\_COLOR])) AS [CATEGORY\_VALUE] FROM [input]

UNION

SELECT [ITEM\_CODE], 4 AS [CATEGORY\_ID], TRIM(LOWER([CATEGORY\_GENDER])) AS [CATEGORY\_VALUE] FROM [input]

) AS InputUnion

WHERE [CATEGORY\_VALUE] IS NOT NULL

SELECT DISTINCT

[ITEM\_CODE],

[ID] AS [CATEGORY\_ID]

FROM #Temp

ORDER BY [ITEM\_CODE], [ID]

SELECT DISTINCT

[ID],

[CATEGORY\_ID],

[CATEGORY\_VALUE]

FROM #Temp

ORDER BY [ID], [CATEGORY\_ID]

## 2. Powershell data manipulation

## In terms of data manipulation and process automation in the MS Windows environment, Powershell is one of the best tools that can be employed. Powershell scripts can be executed either via SSIS Process Task or separately in the command line. Using Powershell script in combination with SQL DB would provide us with a completely independent and very flexible solution. If there would be an option to utilize SQL DB with Powershell it would be probably my most favorite solution for this specific project.

## 3. Python data formatting

Python is an extremely powerful language which could be also a very helpful replacement for SSIS, Batch and PowerShell scripts. The greatest benefit of this solution would be the formatting performance and the fact that the whole solution can be placed in a single script file. On the other hand, if we would want to run the script in some other environment, we would have to first secure the Python installation. Also, the script customization and maintenance are not as easy as for SSIS or SQL.