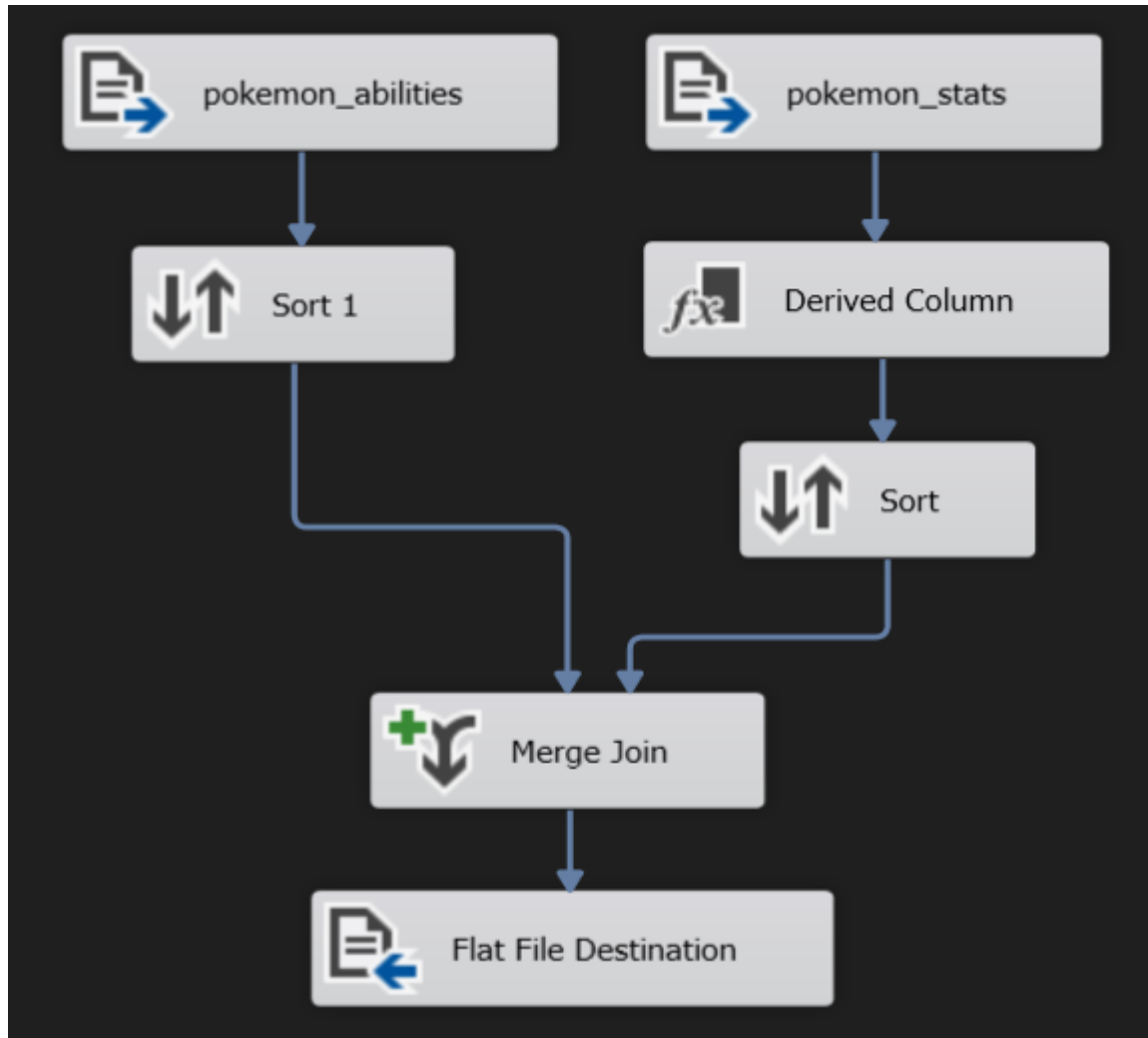


## Preprocessing Stage

### Aim of the process

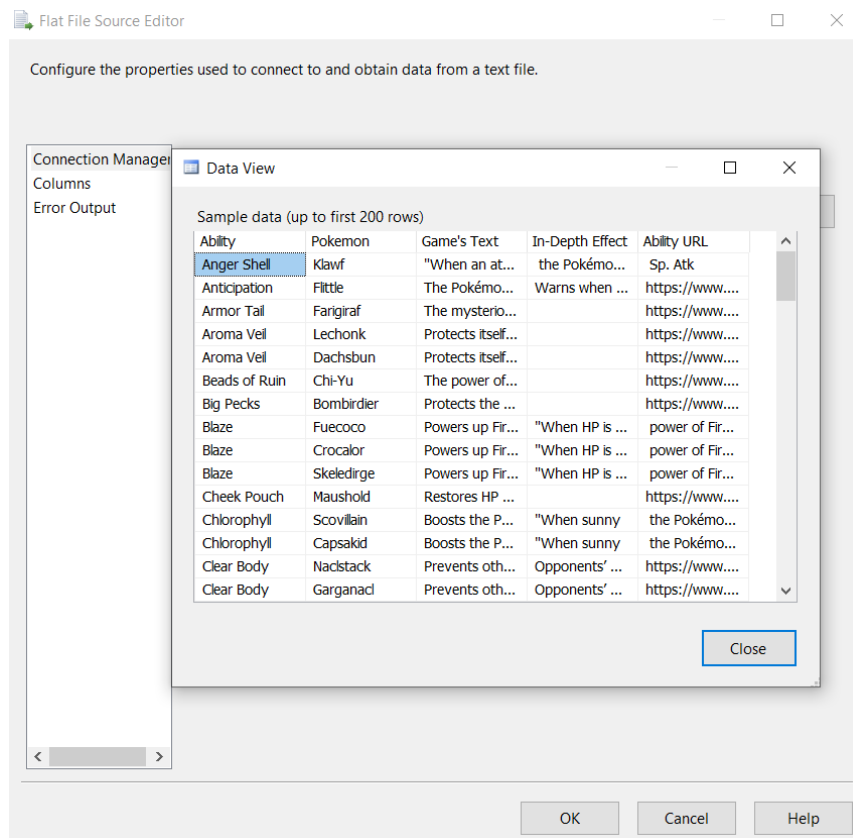
The preprocessing stage is done to obtain files ready for data cleansing done with the scripts written in python. The whole data flow process at this stage is shown below:



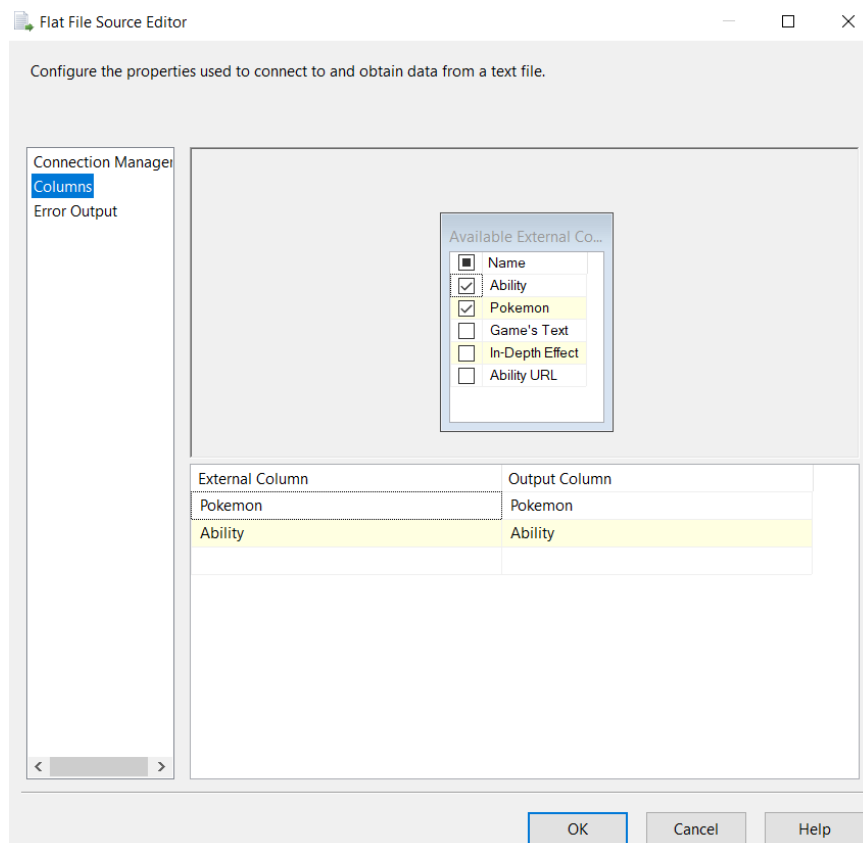
### Data sources and initial transformations

The 'pokemon\_abilities' (gen9\_pokemon\_abilities.csv) data source:

## Data View:



## Choosing appropriate data columns:



Sorting the data:

Sort Transformation Editor

Specify the columns to sort, and set their sort type and their sort order. All nonselected columns are copied unchanged.

Available Input Columns

<input type="checkbox"/>	Name	Pass Thr...
<input checked="" type="checkbox"/>	Pokemon	<input type="checkbox"/>
<input type="checkbox"/>	Ability	<input checked="" type="checkbox"/>

Input Column	Output Alias	Sort Type	Sort Order	Con
Pokemon	Pokemon	ascending	1	

☐ Remove rows with duplicate sort values

OK

Cancel

Help

The 'pokemon\_stats' (gen9\_pokemon\_stats.csv) data source:

Data View:

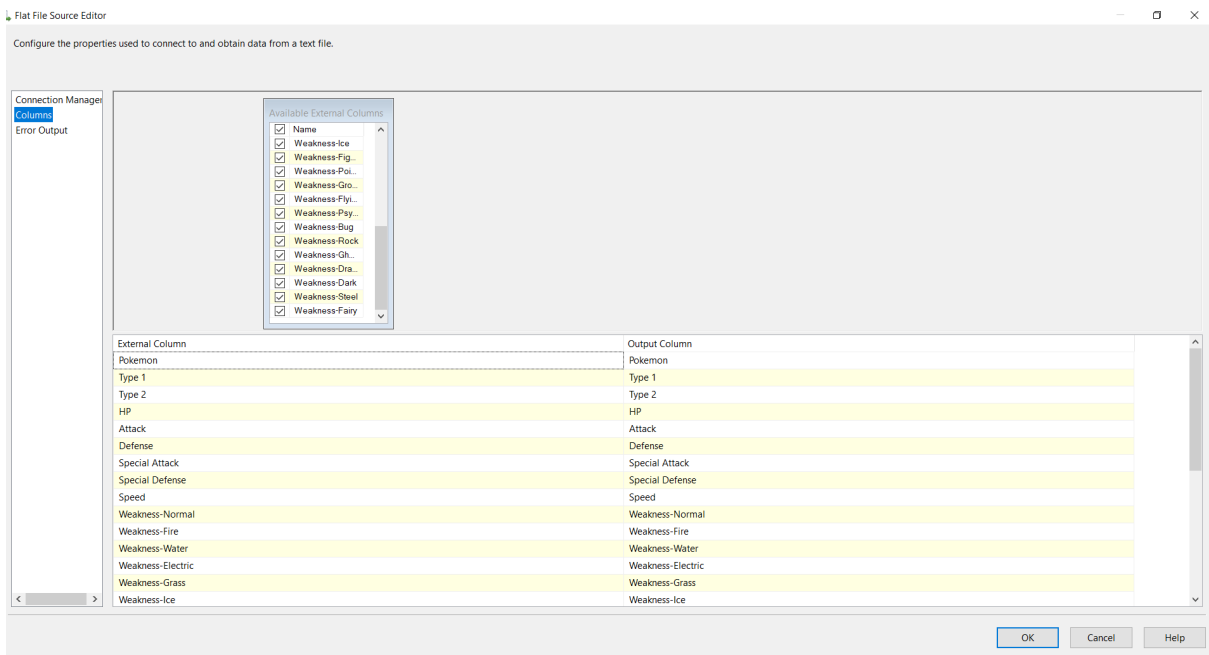
Data View

Sample data (up to first 200 rows)

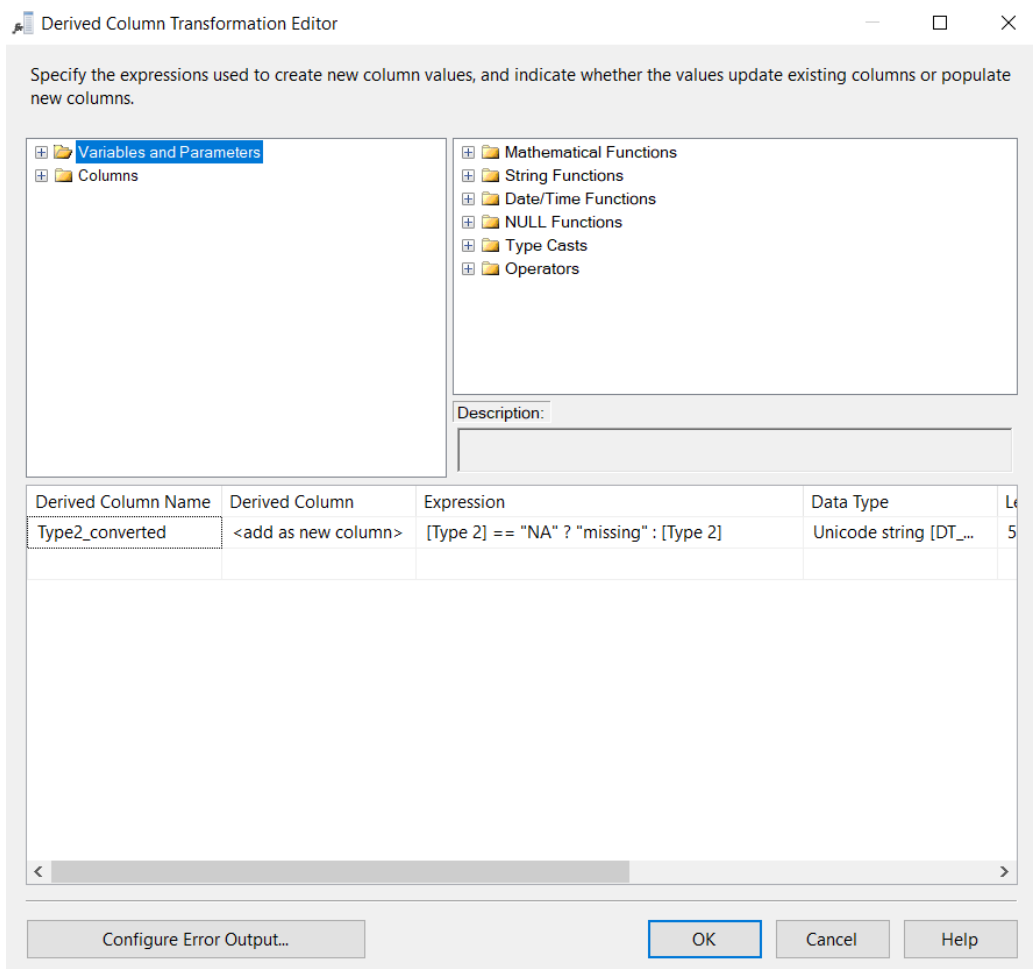
Pokemon	Type 1	Type 2	HP	Attack	Defense	Special Attack	Special Defense	Speed	Weakness-No...	Weakness-Fire	Weakness-W...	Weakness-Ele...	Weakness-Po...
Armarouge	fire	psychic	85	60	100	125	80	75	1	0.5	2	1	0.0
Baxcalbur	ice	dragon	115	145	92	75	86	87	1	1	0.5	0.5	0.0
Bellbolt	electric	NA	109	64	91	103	83	45	1	1	1	0.5	1
Bombirdier	flying	dark	70	103	85	60	85	82	1	1	1	2	0.0
Brambleghast	grass	ghost	55	115	70	80	70	90	0	2	0.5	0.5	0.0
Bramblin	grass	ghost	40	65	30	45	35	60	0	2	0.5	0.5	0.0
Brute Bonnet	dark	grass	111	127	99	79	99	55	1	2	0.5	0.5	0.0
Capsakid	grass	NA	50	62	40	62	40	50	1	2	0.5	0.5	0.0
Ceruleledge	fire	ghost	75	125	80	60	100	85	0	0.5	2	1	0.0
Cetitan	ice	NA	170	113	65	45	55	73	1	2	1	1	1
Cetoddle	ice	NA	108	68	45	30	40	43	1	2	1	1	1
Charcadet	fire	NA	40	50	40	50	40	35	1	0.5	2	1	0.0
Chi-Yu	dark	fire	55	80	80	135	120	100	1	0.5	2	1	0.0
Chien-Pao	dark	ice	80	120	80	90	65	135	1	2	1	1	1
Clodsire	ground	poison	130	75	60	45	100	20	1	1	2	0	1

Close

Choosing appropriate data columns:



Transforming 'Type 2' column:



## Sorting the data:

Sort Transformation Editor

Specify the columns to sort, and set their sort type and their sort order. All nonselected columns are copied unchanged.

Available Input Columns

<input type="checkbox"/> Name	Pass Thr...
<input checked="" type="checkbox"/> Pokemon	<input type="checkbox"/>
<input type="checkbox"/> Type 1	<input checked="" type="checkbox"/>
<input type="checkbox"/> Type 2	<input type="checkbox"/>
<input type="checkbox"/> HP	<input type="checkbox"/>
<input type="checkbox"/> Attack	<input type="checkbox"/>
<input type="checkbox"/> Defense	<input type="checkbox"/>
<input type="checkbox"/> Special ...	<input type="checkbox"/>
<input type="checkbox"/> Special ...	<input type="checkbox"/>
<input type="checkbox"/> Speed	<input type="checkbox"/>
<input type="checkbox"/> Weakne...	<input type="checkbox"/>
<input type="checkbox"/> Weakne...	<input type="checkbox"/>
<input type="checkbox"/> Weakne...	<input type="checkbox"/>
<input type="checkbox"/> Weakne...	<input type="checkbox"/>

Input Column	Output Alias	Sort Type	Sort Order	Comparison Flag
Pokemon	Pokemon	ascending	1	

☐ Remove rows with duplicate sort values

OK Cancel Help

## Merge join

Merge Join Transformation Editor

Configure the properties used to join two sources of sorted data. Select the join type and then specify the columns to be used as the join key. Join keys must be used in the order specified by the sort-key position of the column.

Join type: Inner join Swap Inputs

Sort 1

<input checked="" type="checkbox"/> Name	Ord...	Join...
<input checked="" type="checkbox"/> Pokemon	1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> Ability	0	<input type="checkbox"/>

Sort

<input checked="" type="checkbox"/> Name	Ord...	Join...
<input type="checkbox"/> Pokemon	1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> Type 1	0	<input type="checkbox"/>
<input checked="" type="checkbox"/> Type2_converted	0	<input type="checkbox"/>

Input	Input Column	Output Alias
Sort 1	Pokemon	Pokemon
Sort 1	Ability	Ability
Sort	Type 1	Type1_form
Sort	Type2_converted	Type2_form

OK Cancel Help

## Destination

Flat File Destination Editor

Configure the properties used to connect to and insert data into a text file.

Connection Manager:  
Mappings

Flat File connection manager:  
Flat File Connection Manager 6 New...

☒ Overwrite data in the file Update...

Header:  
Pokemon,Ability,Type1\_form,Type2\_form

Preview...

OK Cancel Help

Flat File Destination Editor

Configure the properties used to connect to and insert data into a text file.

Connection Manager:  
Mappings

Available Input Columns:

Name
Pokemon
Ability
Type1_form
Type2_form

Available Destination Columns:

Name
Pokemon
Ability
Type1_form
Type2_form

Input Column Mapping:

Input Column	Destination Column
Pokemon	Pokemon
Ability	Ability
Type1_form	Type1_form
Type2_form	Type2_form

OK Cancel Help