



Analytical Dataset – a school based guide to using the dataset

The Analytical Dataset is the data that sit below Insight. Whereas Insight is a tool for examining cohorts of pupils, the dataset allows analysis at pupil level, organised by Scottish Candidate Number.

The Dataset is offered to all Local Authorities shortly after the publication of each Update (typically September and February). Whilst practice varies, many LAs make this available to schools. There are multiple sheets of data provided. School improvement conversations are supported primarily by the data sheet called 'Pupils'.

The dataset may be shared in a variety of formats. This Deep Dive assumes distribution of the *raw* unformatted CSV file.

Guides to exporting CSV data into Microsoft Excel (Appendix 1) and to the codes used within the dataset (Appendix 2) are available as appendices within this Deep Dive.

This Deep Dive is intended to give users ideas on how the dataset can be used to aid the analysis of their data. It is not intended to be comprehensive, but hopes to illustrate what is possible with the Dataset, and encourage deeper use within schools.

Navigation

	A	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	scn	sex	ethnicity	age_gro	lac_gro	eal_gro	fsm_gro	asn_gro	vc_asn	simd_d	simd_v	pupil_p	leaver	leaver	destina	leaver	vc_stag	vc_leav	highest
2	60325545	2	1	3	NULL	2	2	2	1	4	7	3	1	1	4	1000001	4	3	
4	60374694	1	1	3	NULL	2	2	2	1	8	15	3	1	1	4	1000001	4	3	
8	61312552	1	1	3	NULL	2	2	2	1	8	16	3	1	1	4	1000001	4	3	
9	61283714	1	1	3	NULL	2	2	2	1	10	19	3	1	1	4	1000001	4	3	
19	64245473	2	1	3	NULL	2	2	2	1	6	11	3	1	1	4	1000001	4	3	
20	64248189	2	1	3	NULL	2	2	2	1	7	14	3	1	1	4	1000001	4	3	
23	70275155	1	1	3	NULL	2	2	2	1	8	16	3	1	1	4	1000001	4	3	
26	70191903	1	1	3	NULL	2	2	2	1	8	16	3	1	1	4	1000001	4	3	
28	70202689	1	1	3	NULL	2	2	2	1	8	16	3	1	1	4	1000001	4	3	
36	70283956	1	1	3	NULL	2	2	2	1	8	16	3	1	1	4	1000001	4	3	
38	70286416	2	1	3	NULL	2	2	1	2	5	9	3	1	1	4	1000001	4	3	
40	70286424	2	1	3	NULL	2	2	2	1	8	16	3	1	1	2	1000001	4	3	
52	70930501	2	1	3	NULL	2	2	2	1	8	15	3	1	1	4	1000001	4	3	
54	70936283	1	1	3	NULL	2	2	1	2	8	15	3	1	1	4	1000001	4	3	
56	70938103	1	1	3	NULL	2	2	2	1	9	18	3	1	1	4	1000001	4	3	
58	70986930	2	2	3	NULL	1	2	2	1	5	9	2	1	1	4	1000001	4	3	
64	71132218	2	1	3	NULL	2	2	2	1	2	4	3	1	1	4	1000001	4	3	
66	71054659	2	2	3	NULL	2	2	2	1	7	13	3	1	1	4	1000001	4	3	
69	71139646	1	1	3	NULL	2	2	2	1	6	8	3	1	1	4	1000001	4	3	



Using the Dataset

Preparing the Dataset for use

If the dataset is not already set up as such, we would recommend:

- Adding an initial column (or columns) to include **pupil names** (exported from SEEMiS) alongside their SCN.
- Changing the **column headings in Row 1** to “**Filters**” to facilitate analysis.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
	Name	Attendance	scn	year	baseCentre	stage	sex	ethnicity	age_group	lac_group	eal_group	fsm_group	asn_group	vc_asn_and_m ainstream_inte gration_group	simd_decile	simd_vigintile	pupil_points_g roup	leaver_group	leaver_destina tion_group
1	Pupil Name1		191697421	2021	1000001	5	Female	White	16 -	-	-	-	-	No	9	18	Lowest 20	-	-
2	Pupil Name2		91354721	2021	1000001	4	Female	White	<16	Home	-	FSM	ASN	< 80%	6	11	Lowest 20	-	-
3	Pupil Name3		81311102	2021	1000001	5	Female	Minorit	16	Away	-	FSM	ASN	80% +	1	1	Lowest 20	Leave Non Pos	I
4	Pupil Name4		81350957	2021	1000001	5	Female	White	16 -	-	-	-	-	No	1	2	Lowest 20	Leave Positive	I
5	Pupil Name5		80164904	2021	1000001	5	Female	White	16 -	-	-	-	-	No	8	15	Lowest 20	Leave Positive	I
6	Pupil Name6		80873476	2021	1000001	5	Female	White	16 -	-	-	FSM	ASN	80% +	2	4	Lowest 20	Leave Positive	I
7	Pupil Name7		80755546	2021	1000001	5	Female	White	16 -	-	-	-	ASN	80% +	4	8	Lowest 20	Leave Positive	I
8	Pupil Name8		81160597	2021	1000001	5	Male	White	16 -	-	EAL	-	-	No	1	2	Lowest 20	Leave Positive	I
9	Pupil Name9		80964080	2021	1000001	5	Female	White	16 -	-	-	-	-	No	5	10	Middle 60	Leave Positive	I
10	Pupil Name10		80743289	2021	1000001	5	Female	White	16 -	-	-	-	-	No	2	3	Middle 60	-	-

- It may also be useful to use the Analytical Data Guide to **assign labels** (rather than the codes) to each of the data points on the spreadsheet.

	A	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	
	Name	leaver_group	leaver_destination_group	destination	leaver_centre	vc_stage_and_winter_leaver_group	vc_leaver_stage_group	highest_scqf_course_to_date	highest_scqf_level_to_date	Numeracy highest_scqf_level_to_date	annual total tariff pts	cumulative total tariff pts	complementary tariff pts	cumulative complementary tariff pts	
1	Pupil Name1	-	-	System Leave	NULL	S5	NULL	NULL	NULL	NULL	0	0	0	0	
2	Pupil Name2	-	-	System Leave	NULL	S4	NULL	NULL	NULL	NULL	0	0	0	0	
3	Pupil Name3	Leave Non Posit		Unemployed	myschool	S5 Winter	S4 and	NULL		3	NULL	0	4	0	4
4	Pupil Name4	Leave Positive		Employed	myschool	S5	S5		4	4	4	0	62	0	62
5	Pupil Name5	Leave Positive		Employed	myschool	S5 Winter	S4 and		5	5	NULL	0	65	0	65
6	Pupil Name6	Leave Positive		Further Educ	myschool	S5	S5		4	4	4	0	80	0	80
7	Pupil Name7	Leave Positive		Further Educ	myschool	S5 Winter	S4 and		5	5	4	0	203	0	191
8	Pupil Name8	Leave Positive		Employed	myschool	S5 Winter	S4 and		5	4	4	0	228	0	216
9	Pupil Name9	Leave Positive		Further Educ	myschool	S5	S5		5	5	4	0	494	0	390
10	Pupil Name10	-	-	System Leave	NULL	S5	NULL		5	5	5	0	311	0	299



This guide poses reflective questions to illustrate how the dataset can be used to supplement Insight analysis and inform school improvement at pupil level.

Measure 1 – Local Measure: Attainment Versus Destination (Using S4 & Complementary)

Reflective Questions.

Use the Dataset to:

- *Identify the SCN of the learners with zero points*
 - Consider the narrative that lead to learners achieving zero points
 - Are there improvement messages for future learners?
- *Identify pupils with less than 100 points – roughly equivalent to three N4 passes*
 - How many learners are there?
 - Is this a significant part of the year roll?
 - Consider the narrative that lead to learners achieving these points
 - Is there a pattern to this attainment (possibly consider different demographics or groups of learners based on, for example, the sex of the learners, SIMD, FSM, EAL, LAC, ASN needs)
 - Are there improvement messages for future learners?
- *Investigate ASN patterns*
 - Are there pupils with no ASN but low attainment – what led to this?
 - Note there are two columns categorising ASN; (i) ASN Group and (ii) ASN and mainstream integration group
- *Pupils with maximum attainment*
 - Five N5 A grades would equate to 5x84 points, 420 points – how well distributed are high attainers across learner demographics?
- *Level 6 presentation in S4*
 - Are any learners *above* 420 points?
 - Those learners gaining level 6 awards in S4, what was their attainment in other subjects?
- *Link to SEEMiS SQA data*
 - Can the attainment data in the data set be linked to SEEMiS attainment records to identify patterns in subjects?
 - How well does each subject area contribute to
 - (i) the strongest five pieces of attainment (for Complementary)
 - (ii) for the highest attaining pupils, are the upper N5 grades appropriately distributed across all subject areas?
 - (iii) for the lower attaining pupils, how well do all subjects ensure learners achieve within their subject areas?



Measure 2 – Local Measure: Attainment for All (S4 & Complementary)

Reflective Questions (note some Measure 1 questions are also appropriate here)

- Did all S4 learners have at least 5 pieces of attainment?
- How well do all learners do in getting an award in each subject they study; are there patterns? *This information can be found for searching by SCN in the attainment tab of the dataset or through SEEMiS.*
- How many learners were able to gain 420 complementary tariff points?
- Did any learners gain more than 420 through level 6 presentation? Note this has the potential to skew data and hide underachievement elsewhere.
- How well do we sustain 5 x National 5 A attainment; at what rate does attainment “fall away” with 420 points becoming becomes 410, 400 etc?
- What does the “shape” of attainment within the middle 60% look like? Is there a gradual or more marked “fall off” in points?
- Curriculum offer and progression pathways can have a marked impact upon Attainment for All; what opportunities (perhaps NPAs, SfW, SQA Awards, non-SQA awards etc) are available to learners?
- In each of the above consider different demographics and/or groups of learners

In addition,

- Identify in S4 the learners with the highest 20%, middle 60%, lowest 20% - track them into S5 and S6.
- How well does S5 individual attainment build on those individuals S4 attainment? Consider points gained in S4, how many more points have been added for each candidate in S5?
- Are any pupils moving significantly between the highest 20%, middle 60%, lowest 20% (either up or down?) as they move from S4 into S5? Why did this happen?
- Were there any S5 pupils who achieved no, or very low points in S5 annual data? Ditto for S6 – what opportunities exist for all learners to be successful?
- Were there subject patterns in above? Do all subjects add value as expected as learners progress through Senior Phase? Do all subjects areas provide appropriate progression pathways?

Measure 3 – Local Measure: Literacy and Numeracy

Reflective Questions

- Which learners failed to achieve an award in one or both literacy and numeracy? What did their other attainment look like?
- Are there learners who have achieved a package of attainment elsewhere but failed to attain Lit and/or Num?
- Are there pupils who have achieved one or more level 5 Awards (such as NPAs, SfW and N5s), but whose Lit/Num is at level 4 or below?
- Similarly, are there pupils who have achieved one or more level 6 Awards (or even level 7 Awards), but whose Lit/Num is at level 5 or below?



- In each of the above consider different demographics and/or groups of learners.

Measure 4 – National Measure: Positive Destinations

Reflective Questions

- Were there any learners who achieved at Level 7, but moved to a level 6 destination?
Why was this the case?
- Likewise for learners who achieved at least one Level 6 and moved to a Level 5 or 6 destination; did any of these learners have more than 1 level 6 and therefore may have the entry requirements for Level 7 courses such as HNC and HND?
- In each of the above consider different demographics and/or groups of learners.



APPENDIX 1 – Importing the CSV Dataset file into Excel

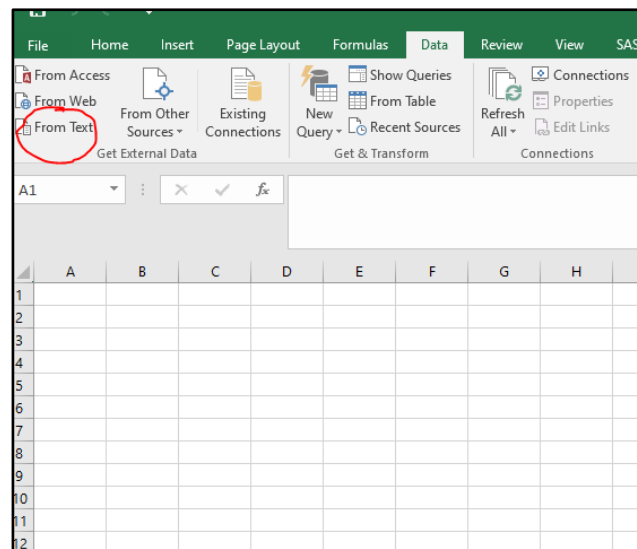
Converting text data into Microsoft Excel data

Instructions for users of the Data Set to import the Data Set into Microsoft Excel from the original CSV file.

Ensure you have unzipped text documents. If they are in a zip folder you can simply right click on the folder and select “Extract all” or copy and paste the files out of the zip folder.

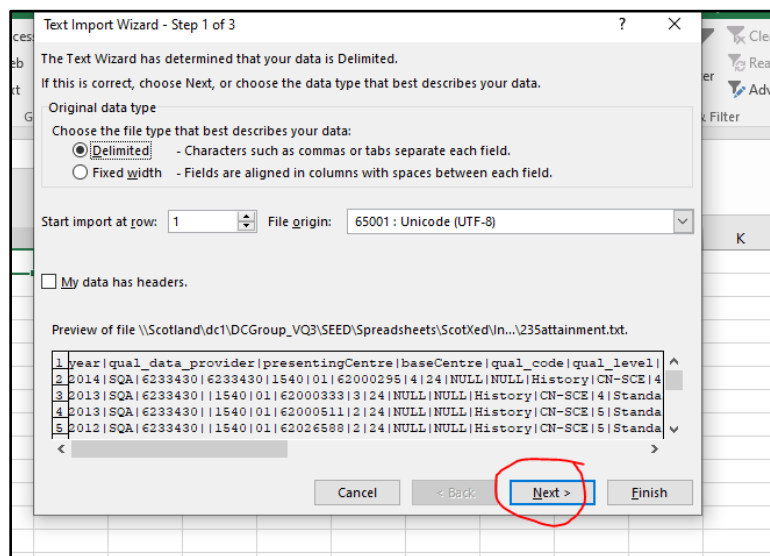
Open Microsoft Excel

- Select the **Data** menu
- Select “**From Text**” option



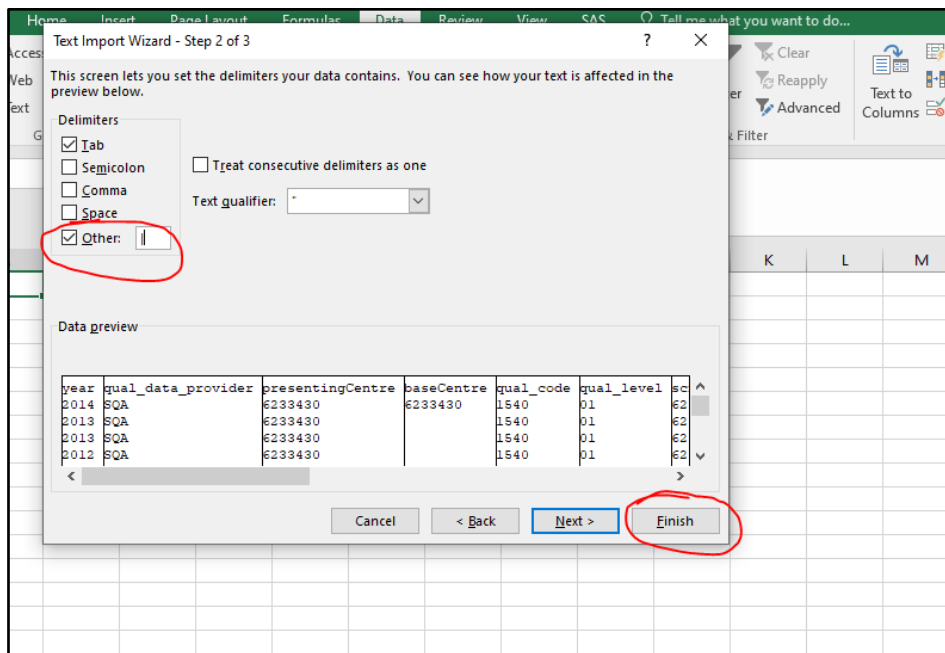
- Now select the data you want to use from the file explorer (text file).

- The dialogue box (opposite) should open, press “**Next**”.

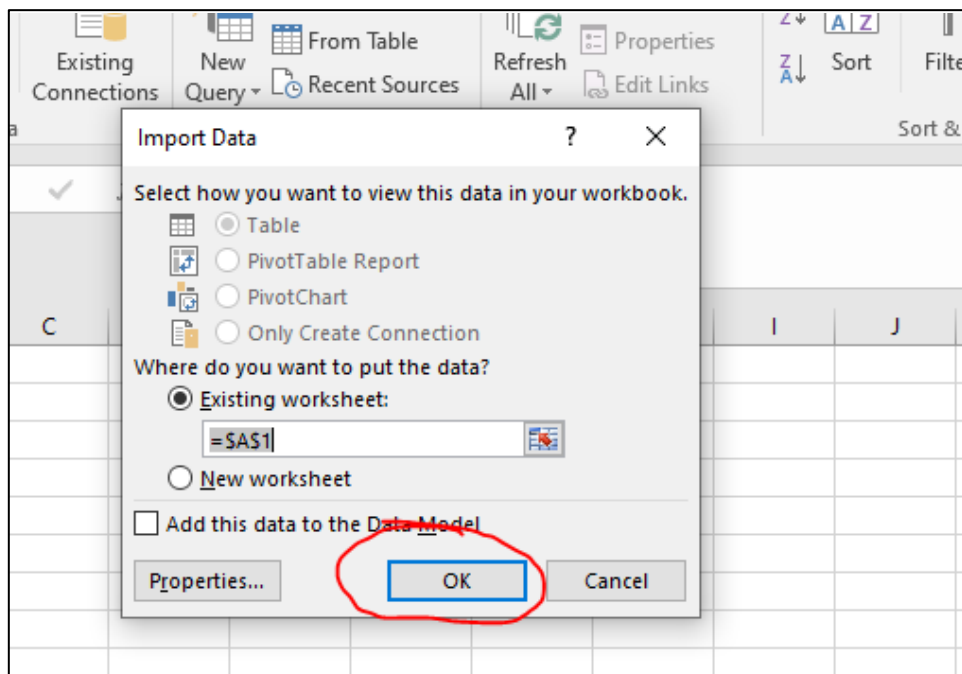




→ Now select “Other” and type in a “|” symbol. Then press “Next”.



→ The below dialogue box should open. Now press “OK”.



→ Now repeat the process for each of the files in a new tab each time or in a new document depending on your preference.



APPENDIX 2 – Guide to the codes used within the Dataset

Overview

Insight provides an intuitive, flexible, powerful web based front end to allow users to explore data and carry out analyses. The front end also allows users to download data for further analysis.

Where users wish to carry out further analysis that requires them to link the Insight data to other datasets that they hold, an *Insight Analytical Dataset* is available on request.

This document provides a brief overview of the analytical dataset.

Background

Format

The data is provided as a series of vbar (pipe) delimited text files. These files are:

- Schools
- Pupils
- Attainment
- Subjects

Scope

The data is provided at individual level and no additional disclosure control is applied. An education authority or school may request data *only* for pupils based at their centre(s). The definition of base centre used is the one adopted by Insight.

Governance

Formal data sharing agreements should be put in place to ensure both parties to the share understand the basis, limitations of use and obligations of each party.

The Files

Schools

The schools file contains:

Field Name	Description
SeedCode	The SEED (Scottish Educational Establishment Database) unique centre identifier. A 7 digit number.
SchoolName	The School Name held on the SEED database.



Pupils

The pupil data contains characteristics of the pupil, as classified by Insight and drawing upon relevant pupil census, destinations and attainment data.

The pupil file contains:

Field Name	Description
scn	The SCN (scottish candidate number) of the pupil. A 9 digit number.
year	The year upon which the classification is based. Note that this represents the attainment year. A 4 digit number.
centre	The SEED code of the centre at which the pupil was based. A 7 digit number.
stage	The classified stage: <ul style="list-style-type: none"> • 4 – S4 • 5 – S5 • 6 – S6
sex	The classified sex <ul style="list-style-type: none"> • 1 – Male • 2 – Female
ethnicity	The classified ethnicity <ul style="list-style-type: none"> • 1 – White • 2 – Minority ethnic • 3 – Not known
age_group	The classified age <ul style="list-style-type: none"> • 1 – Under 16 • 2 – 16 • 3 – 17 • 4 – 18 • 5 – Over 18
lac_group	The classified looked after child status <ul style="list-style-type: none"> • 1 – Looked after at home • 2 – Looked after away from home • NULL – Not looked after or status not known
eal_group	The classified English as Additional Language status <ul style="list-style-type: none"> • 1 – EAL • 2 – Other
fsm_group	The classified Free School Meals status <ul style="list-style-type: none"> • 1 – FSM • 2 – Not FSM
asn_group	The classified Additional Support Needs status <ul style="list-style-type: none"> • 1 – ASN • 2 – Other
vc_asn_and_mainstream_integration_group	The classification of ASN used for Virtual Comparator construction <ul style="list-style-type: none"> • 1 - No • 2 - Yes, 80% or more mainstream integration • 3 - Yes, less than 80% mainstream integration
simd_decile	The Scottish Index of Multiple Deprivation (SIMD) decile of the pupil derived from their home postcode.
simd_vigintile	The SIMD vigintile of the pupil derived from their home postcode.
pupil_points_group	The cohort assignment of the pupil based on tariff points



	<ul style="list-style-type: none"> 1 - Lowest 20% 2 - Middle 60% 3 – Highest 20% <p>Note: The pupil points group that is assigned to each pupil is based on the top 20%, middle 60% and bottom 20% attainment across all pupils at a national level. In Insight the pupils in the top 20%, middle 60% and bottom 20% attainment it is recalculated based only on achievement in the cohort you are looking at, for example your school or LA. This means that using the pupil points variable will give you slightly different values as some of the bottom 20% of S5 pupils in your establishment may actually be in the middle 60% of pupils nationally. In order to replicate the figures from Insight you would need to order the pupils in your cohort by cumulative total tariff points and then work out the means for the top, middle, bottom groups.</p>
leaver_group	<p>The classification of the pupil as a leaver</p> <ul style="list-style-type: none"> 0 – Is not a leaver 1 – Is a leaver
leaver_destination_group	<p>The pupil's leaver destination group</p> <ul style="list-style-type: none"> 1 – Positive 2 – Non Positive NULL – Not defined (not a leaver or not positive)
destination	<p>The detailed initial destination category assigned from SLDR and pupil census</p> <ul style="list-style-type: none"> 1 - Activity Agreement (not used from February 2020 update onwards) 2 - Employed 3 - Further Education 4 - Higher Education 5 - Training 6 - Voluntary Work 7 - Stayed on at school 8 - Any positive destination 9 - Not known 11 – Excluded 12 – Unemployed Seeking 13 – Unemployed Not Seeking 14 - System leaver (a pupil who appears in the pupil census one year but does not appear the next and who is not classified as a school leaver) 15 – Personal Skills Development (February 2020 update onwards – used when a leaver is undertaking a specific Personal Skills Development activity after leaving school)
leaver_centre	<p>The SEED code of the pupils leaving centre</p>
vc_stage_and_winter_leaver_group	<p>Stage group used for VC matching, stage based VCs</p> <ul style="list-style-type: none"> 1 – S4 2 – S5 winter leavers 3 – S5 4 – S6
vc_leaver_stage_group	<p>Stage group used for VC matching, leaver based VCs</p> <ul style="list-style-type: none"> 1 – S4 and S5 winter leavers 2 – S5



	<ul style="list-style-type: none"> • 3 – S6 • NULL – Not defined
highest_scqf_course_to_date	Highest SCQF level of courses attained by this pupil to date
highest_lit_scqf_level_to_date	Highest SCQF level of courses defined as <i>literacy</i> attained by this pupil to date
highest_num_scqf_level_to_date	Highest SCQF level of courses defined as <i>numeracy</i> attained by this pupil to date
ann_total_tariff_pts	Total Insight tariff points for the academic year
cum_total_tariff_pts	Cumulative Insight tariff points for all eligible attainment in the senior phase
ann_comp_tariff_pts	Complementary Insight tariff points for the academic year
cum_comp_tariff_pts	Cumulative Insight tariff points for all eligible attainment in the senior phase



Attainment

The attainment data contains details of course and unit attainment, linked to pupils by their SCN.

The attainment file contains:

Field Name	Description
year	Attainment year
qual_data_provider	Award provider, e.g. <ul style="list-style-type: none"> SQA – Scottish Qualifications Authority YS – Youth Scotland
centre	The SEED code of the presenting centre
qual_code	The <i>Qual Provider</i> qualification code
qual_level	The <i>Qual Provider</i> level code. Note that this may not be the same as SCQF level and in many cases will not be.
scn	The pupils SCN
qual_result	The <i>qual provider</i> result code
scqf_credit_points	The number of SCQF credit points associated with the qualification
is_literacy	Indicates whether it meets the <i>Insight</i> definition of literacy: <ul style="list-style-type: none"> 0 – Not literacy 1 – Literacy NULL – Not defined (not literacy)
is_numeracy	Indicates whether it meets the <i>Insight</i> definition of numeracy: <ul style="list-style-type: none"> 0 – Not numeracy 1 – numeracy NULL – Not defined (not numeracy)
qual_name	The qualification name
qual_product_type	The qualification product type, as provided by the <i>qual provider</i>
scqf_level	The SCQF level of the qualification
qualificationType	The description of the qualification used by the <i>qual provider</i> and based on product type and level



Subjects

The subject points data contains details of deduplicated attainment by subject, linked to pupils by their SCN. Insight subject definitions are used.

The attainment file contains:

Field Name	Description
scn	The SCN of the pupil
year	Attainment year
presenting_centre	The SEED code of the presenting centre
name	The name of the subject
scqf_level	The SCQF level of the attainment
tariff_points	The Insight tariff points of attainment for that pupil, at that presenting centre, in that subject and at that SCQF level