|  |
| --- |
| Inland Revenue |
| **Property Technical Requirements**  **Inferred Sales Estimates**  Document Version: 1.5 |
| **IN CONFIDENCE** |

Table of Contents

[1 Purpose 4](#_Toc34818162)

[2 Business Requirements 4](#_Toc34818163)

[3 Solution Overview 5](#_Toc34818164)

[3.1 Solution Diagram 5](#_Toc34818165)

[4 Functional Requirements 6](#_Toc34818166)

[5 Non-Functional Requirement 12](#_Toc34818167)

[6 APPENDIX ONE – Estimation Decision Trace 14](#_Toc34818168)

[7 APPENDIX TWO – Estimation Source Trace 15](#_Toc34818169)

[8 Terms 17](#_Toc34818170)

[8.1 Definition of a Dealer 17](#_Toc34818171)

[8.2 Definition of a Builder 17](#_Toc34818172)

[8.3 Definition of a Developer 17](#_Toc34818173)

Document Control

|  |  |  |
| --- | --- | --- |
| **Author** | **Version** | **Change** |
| Mark Openstein | 1.0 | Initial draft |
| Mark Openstein | 1.1 | Refinements to FR-06 logic based on SQL build |
| Mark Openstein | 1.2 | Minor formatting updates |
| Mark Openstein | 1.3 | More scenarios added to FR06; scenario 6, 7 and added appendix 1 |
| Mark Openstein | 1.4 | Added NFR-06 |
| Mark Openstein | 1.5 | Updated numbering and added FR-07, aggregate marker requirements updated, added appendix two, general formatting, NFR-06 expanded to level 5, appendix 2 added details |

# Purpose

This document will outline the requirements for a solution to determine a property sale value estimate. The estimation will be used when a sales value is unavailable for the property sales transaction.

# Business Requirements

The main requirements here are for auditing purposes by the business to find properties where follow up is required for tax elated purposes.

|  |  |  |
| --- | --- | --- |
| **Requirement** | **Priority** | **Description** |
| BR-01 | must | Sales without values will need an estimation to be calculated. The sales values for all properties are needed regardless if there is a sales value available, this is required for improved compliance reporting of sales within an area by a seller for tax audits.[[1]](#footnote-1) |
| BR-02 | should | Sales values should be reviewed and adjusted to correctly reflect the correct sale value per property transaction. The sales values for properties are applied across titles that formed part of the same agreement, causing misleading sales figures on property records for reporting purposes.[[2]](#footnote-2) |

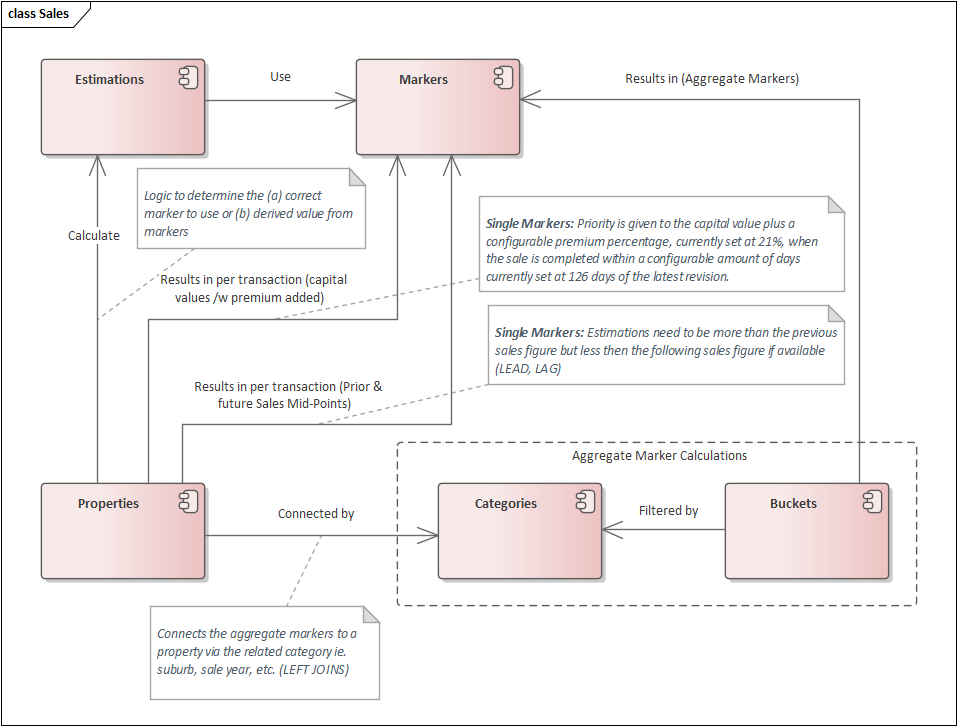
# Solution Overview

Properties that don’t have a sales figure provided for within the data have a mechanism that will generate estimates using the below concepts;

|  |  |
| --- | --- |
| **Concept** | **Overview** |
| Properties | This is the sales property record that has no sales value provided or an incorrect value, that we are inferring a value for. |
| Estimations | The estimation values are chosen by looking at the different markets generated and selecting the correct marker value or a calculation of one or several markers are made and then chosen. |
| Markers | Markers are generated by buckets and the property sales record (Single Marker). Buckets generate Aggregate Markers that can be medians, sum, averages, etc. that are grouped by categories within a bucket. |
| Buckets | Buckets are the data we wish to use to generate a marker value, they output markers and are grouped by categories. |
| Categories | Used to determine the marker values for the bucket and used connect back onto the property record i.e. sale year, suburb, building age, seller type, etc… |

## Solution Diagram

The diagram shows how the concepts interact, to derive an estimation.



**Figure 1 – solution conceptual diagram**

# Functional Requirements

These requirements set out to specify in detail the logical components of the what will be built to address the business requirements previously specified.

|  |  |  |  |
| --- | --- | --- | --- |
| on hold | started | done | not started |

|  |  |  |  |
| --- | --- | --- | --- |
| **Requirement** | **Priority** | **Description** | **Status** |
| **General Requirements** | | | |
| FR-01[[3]](#footnote-3) | could | Property sales for land that with high capital value prices e.g. 50 million and no sales values, could have the capital values equally divided against all related sale properties i.e. 100 properties each with a capital value of 50 Million will be divided equally for every property with the same QPID value[[4]](#footnote-4) | on hold |
| FR-02[[5]](#footnote-5) | could | Property sales for land that would otherwise result in high sales prices e.g. 50 million, *could* be equally divided against all related properties i.e. 100 properties each with a sale value of 50 Million will be divided equally for every property with the same QPID value[[6]](#footnote-6) (these high sales values if not corrected would impact marker calculations unless explicitly ignored.) | on hold |
| FR-03 | should | Markers should be used in combination with one another to select the correct final inferred value (estimations) i.e. averages and median markers will result in a final marker that is the average of both markers, several marker values resulting from a bucket have no value the next highest marker will be selected or selecting of | done |
| FR-04 | must | When the final inferred value (estimations) returns a NULL value calculated from selecting one of the markers, the next available marker value will be supplied, in an order of an agreed precedence e.g. capital values /w 200 000 default | done |
| FR-05 | must | Inferred property sales values (estimations) will be derived from one or more **Single or Aggregate Markers**[[7]](#footnote-7), specifically;   * Averages * Medians * Combined Aggregate Averages * Property Capital Values plus % premium * Property Capital Values with 200K default * Prior & Future Sales Mid-Point * Purchase Cost *Plus+* Consent Costs (TBD)[[8]](#footnote-8) * Latitude and longitude Marker groups (TBD) | started |
| **Single Marker Requirements** | | | |
| FR-06[[9]](#footnote-9) [[10]](#footnote-10) [[11]](#footnote-11) | must | As a final estimation check, when the calculated aggregate marker estimate returns a value that is higher than the next available sales value for the property or lower than the previous sales value (go back twice or forward for scenario 2), then a mid-point value will be calculated between the previous sale and future sale if available, with consideration for capital values *without* premiums provided; *lower letters are CV, upper are sale values*  When the **middle sale** or **middle sales+** calculated estimates are too high compared to latest sale then;  **scenario 1: middle sale**  This scenario applies to sales without a value that have a previous sale and forward sale value that can be used to determine the mid sale value  (A=$100K, x=$80K), (B=$0, y=$320K), (C=$500K, z=$440K), therefore, B = if (A+C/2) > y then (A+C/2) else y  **scenario 2: middle sales+** [[12]](#footnote-12)  This scenario applies to sales without a value that have a sales value two sales previously and one sales forward or have a sales value two sales forward and one previous sale that can be used to determine the mid sale value  (A=$100K, x=$80K), (B=$0, y=$140K), (C=$0, z=$440K), (D=$560K, v=$490K),  therefore, C = if (A+D/2) > z then (A+D/2) else z  therefore, CC=C  therefore, B = if (A+CC/2) > y then (A+CC/2) else y  **scenario 3: final sale** (no results)  This scenario applies to sales without a value for the final sale and cannot provide an estimate  (A=$100K, x=$80K), (B=$0, y=$140K), therefore, keep current estimate  **scenario 4: single sale** (no results)  This scenario applies to a single sale without a value and cannot provide an estimate  (A=$0, x=$80K), therefore, keep current estimate  **scenario 5: second sale** [[13]](#footnote-13)  This scenario applies only when there is a second sale where the first sale had no value provided  (A=$0, x=$80K), (B=$185K, y=$120K), therefore,  A= (x+B/2) >x then (x+B/2) else x  **scenario 6: forward sales[[14]](#footnote-14)**  This scenario applies to sales without a value for the current sale and for at least 2 previous sales but has a sales value for at least the first forward sale *(the scenario uses the similar logic as scenario 5 but the net is wider to include more sales)*  (U=$0, n=$80K), (T=$0, v=$80K), (A=$0, x=$80K), (B=$185K, y=$120K), therefore,  A= (x+B/2) >x then (x+B/2) else x  **scenario 7: previous sales**  This scenario applies to sales without a value for the current sale and for at least 2 forward sales but has a sales value for at least the first previous sale *(the scenario uses the similar logic as scenario 5, but the net is wider to include more sales)*  (B=$185K, y=$50K), (A=$0, x=$80K), (U=$0, n=$180K), (T=$0, v=$280K), therefore,  A= (x+B/2) >x then (x+B/2) else x  **scenario unknown** (no results)  This scenario applies to sales that have no value for multiple sales and don’t fall into other scenarios therefore a sales value cannot be determined as there is too many zero values for the property  **scenario out of Scope** (no results)  This scenario applies to all record that have a sales value already | done |
| FR-07 | must | Generate two different alternatives to the base revision capital value in the sale as;   * Capital Value /w [x] premium (21%) * Capital value /w [$200 000] as a default when there is a multi-sale CV (generally these are large CV’s) | done |
| FR-08[[15]](#footnote-15) | must | Prioritise the use of the CV value of the sale with a **[21%]** increase, when the sale is done within **[126 days]** of the revision of the CV for the property and the CV value is below 3 500 000 million and greater than 0 *(when determining the use of the aggregate markers)* | done |
| FR-09 | could | A supplemental marker could be created that looks at the estimate based on aggregate markers as a high value and compares this to the capital value as a low value and uses similar logic to FR06 | on hold |
| **Bucket Requirements** | | | |
| FR-10 | must | All buckets calculating their markers will exclude property records with a sale value of 0 | done |
| FR-11**\*\***[[16]](#footnote-16) | must | All buckets calculating their markers will exclude sales and capital values greater than 3 500 000 Million | done |
| FR-12[[17]](#footnote-17) | must | All buckets calculating their markers will exclude multi sale records with the sales code 'M11','M12','M13','M21','M22','M23','M31','M32','M33','M43' | done |
| FR-13 | must | All buckets calculating their markers will exclude sales above 3 500 000 million [[18]](#footnote-18) | done |
| **Bucket 1.1 and 1.2: Aggregate Markers based on property only within any sales period** | | | |
| FR-14 | must | The solution will infer estimation values calculated by markers for any type of sale within the area for properties sold in the sale year | done |
| FR-15 | must | Markers will be determined by buckets and categorised and filtered by;   * **Connection categories:** sale year, suburb, area, land size, property type (residential/non-residential) and building age | done |
| FR-16 | must | Markers will result in the following;   * All Property Sales Averages * All Property Sales Median * All Property Average of Above Combined | done |
| **Bucket 1.3 and 1.4: Aggregate Markers based on all property and seller type** | | | |
| FR-17 | should | The solution will infer estimation values for any type of sale by looking at the seller type[[19]](#footnote-19) and their sales within the area for properties sold in the sale year i.e. builder, developer, dealer *(sellers not identified will not be included)* | not started |
| FR-18 | should | Markers will be determined by buckets and categorised and filtered by;   * **Connection categories:** sale year, suburb, area, land size, property type (residential/non-residential), seller type (developer, builder dealer), seller entity and building age[[20]](#footnote-20) | not started |
| FR-19 | should | Markers will result in the following;   * All Property Seller Sales Averages * All Property Seller Sales Median * All Property Seller Average of Above Combined | not started |
| **Bucket 1.5 and 1.6: Aggregate Markers based on residential property for Individual seller types[[21]](#footnote-21)** | | | |
| FR-20 | must | The solution will infer estimation values for residential properties sales by looking at the seller type not being a developer, builder or developer within the area for properties sold in the sale year | started |
| FR-21 | must | Markers will be determined by buckets and categorised and filtered by;   * **Connection categories:** sale year, suburb, area, land size, seller type (other), property type (residential only) and building age | started |
| FR-22 | must | Markers will result in the following;   * Residential Property Individual Bright-line Averages * Residential Property Individual Bright-line Median * Residential Property Individual Bright-line Average of Above Combined | started |

# Non-Functional Requirement

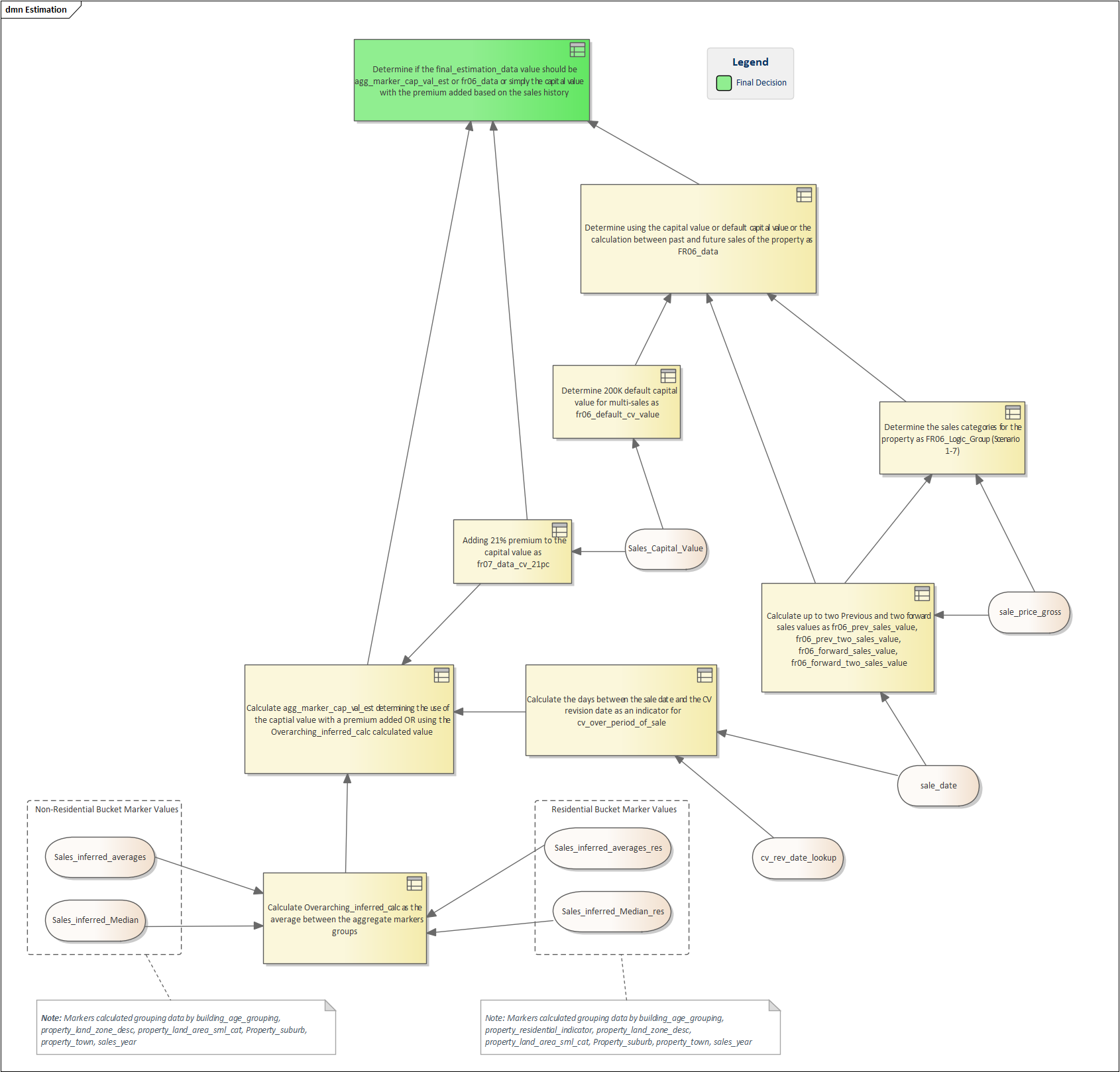
These requirements are not specific to the functionality of the final solution; however, they will ensure a robust, scalable solution that ensures data quality is considered.

|  |  |  |  |
| --- | --- | --- | --- |
| on hold | started | done | not started |

|  |  |  |  |
| --- | --- | --- | --- |
| **Requirement** | **Priority** | **Description** | **Status** |
| NFR-01[[22]](#footnote-22) | should | The solution should be tested by an accuracy score to ensure continual improvements | done |
| NFR-02 | should | Inferred property sales logic (estimation logic) should have accuracy of up to 70% to 130% of actual sales values when looked at those sales transactions | done |
| NFR-03 | could | The accuracy score could be configurable for easily changing metrics to get the best score;   * configure range * configure % score * configure sale/cv days * configure price caps | done |
| NFR-04 | should | The solution should be able to (in the future) link and be superseded by property valuations data from external systems i.e. CoreLogic[[23]](#footnote-23) | on hold |
| NFR-05 | should | The solution built[[24]](#footnote-24) should be able to be incorporated into the current code base used to drive the property dashboard | on hold |
| NFR-06 | should | The solution should provide for a quality indicator for the final estimation (Marker/CV calculations) used: *(The below data values are ranked by the likely hood of their accuracy)*   |  |  |  | | --- | --- | --- | | **level** | **data** | **quality** | | Quality Level -1 | undetermined | unknown | | Quality Level 1 | fr06 calc data | very high | | Quality Level 2[[25]](#footnote-25) | aggregate marker within 30%[[26]](#footnote-26) of fr06 calc data | high | | Quality Level 3 | aggregate marker data | medium | | Quality Level 4 | capital value /w premium | medium | | Quality Level 5 | base capital value /w 200K multi-sale default | low |   *Example data (exported 10/3/2020)* | done |

# APPENDIX ONE – Estimation Decision Trace

The below diagram illustrates how the final estimation is determined by choosing one of the single or aggregate markers.

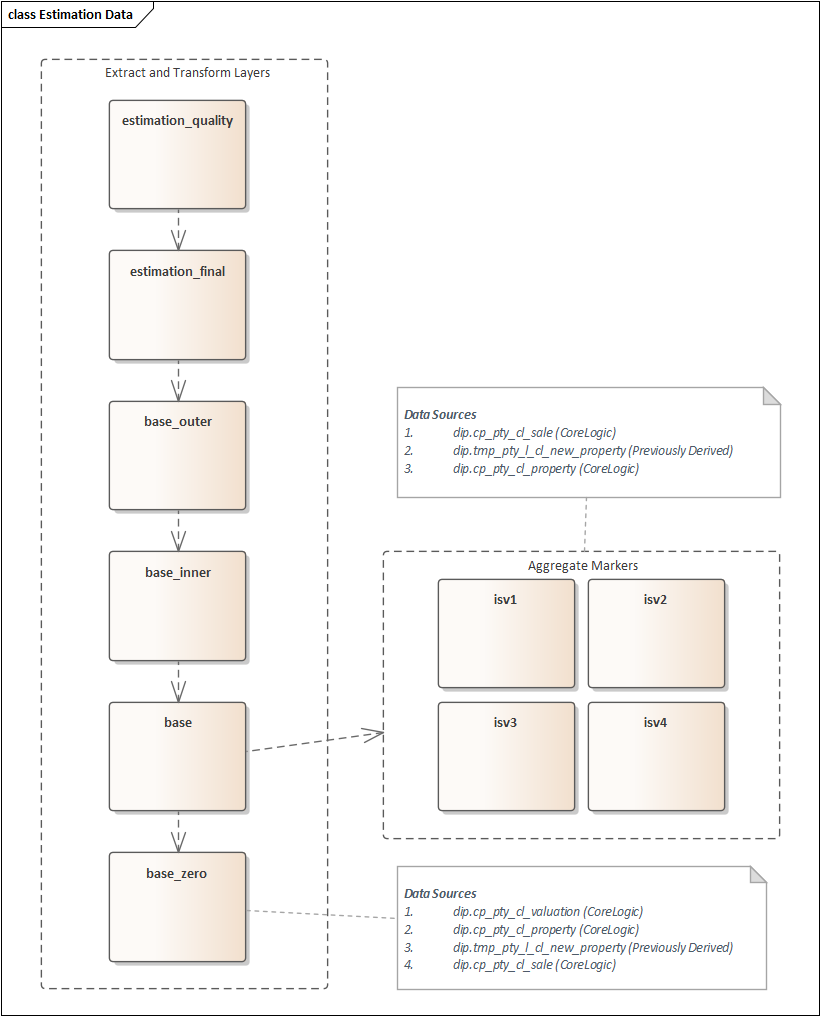


**Figure 2 – estimation decision trace**

# APPENDIX TWO – Estimation Source Trace

The below diagram identifies the source data and illustrates how the estimation logic layers its transformations of the data.

|  |  |
| --- | --- |
| **Extract Layer** | **description** |
| Estimation\_quality | This data determines what quality level the estimate is based on the final value used |
| Estimation\_final | The data determines using the FR6 calculated data or the averaged marker data or the capital value with a premium added |
| Base\_outer | This data calculates and determines using the FR6 calculated data or the capital value with a 200K default for multi-sales |
| Base\_inner | This data determines using the averaged combined marker values or the capital value with a premium added. Identifies the FR6 categories |
| Base | This data averages markers different values and identifies days between a CV and a sale value |
| Base\_zero | This data gives us our base de-duped property record with previous and forward sales values |
| Isv1 | This data bucket results in the marker: All Property Sales Averages |
| Isv2 | This data bucket results in the marker: All Property Sales Median |
| Isv3[[27]](#footnote-27) | This data bucket results in the marker: Residential Property Sales Averages |
| Isv4 | This data bucket results in the marker: Residential Property Sales Median |



**Figure 3 – estimation source trace**

# Terms

## Definition of a Dealer [[28]](#footnote-28)

<https://teams.microsoft.com/l/file/CE1CD858-14BF-4857-B8CD-E447B9DB1D32?tenantId=fb39e3e9-23a9-404e-93a2-b42a87d94f35&fileType=docx&objectUrl=https%3A%2F%2Firnz.sharepoint.com%2Fsites%2FProperty550%2FShared%20Documents%2FGeneral%2FIncrements%20and%20Sprints%2FPI%203%2FSprint%202%2FBasic%20Definition%20of%20Dealers.docx&baseUrl=https%3A%2F%2Firnz.sharepoint.com%2Fsites%2FProperty550&serviceName=teams&threadId=19:27b9d86a8e5c4eb5ba91ca9251a87483@thread.skype&groupId=f138f4bc-e166-4ce6-ba72-9c2bc378e971>

## Definition of a Builder

<https://teams.microsoft.com/l/file/6EABE2FD-A9BA-481C-979B-C14338EA76E5?tenantId=fb39e3e9-23a9-404e-93a2-b42a87d94f35&fileType=docx&objectUrl=https%3A%2F%2Firnz.sharepoint.com%2Fsites%2FProperty550%2FShared%20Documents%2FGeneral%2FIncrements%20and%20Sprints%2FPI%203%2FSprint%202%2FBasic%20Definition%20of%20Builders.docx&baseUrl=https%3A%2F%2Firnz.sharepoint.com%2Fsites%2FProperty550&serviceName=teams&threadId=19:27b9d86a8e5c4eb5ba91ca9251a87483@thread.skype&groupId=f138f4bc-e166-4ce6-ba72-9c2bc378e971>

## Definition of a Developer

<https://teams.microsoft.com/l/file/3C6AFEFC-0441-4925-92BB-80CED64664E9?tenantId=fb39e3e9-23a9-404e-93a2-b42a87d94f35&fileType=docx&objectUrl=https%3A%2F%2Firnz.sharepoint.com%2Fsites%2FProperty550%2FShared%20Documents%2FGeneral%2FIncrements%20and%20Sprints%2FPI%203%2FSprint%202%2FBasic%20Definition%20of%20Developers.docx&baseUrl=https%3A%2F%2Firnz.sharepoint.com%2Fsites%2FProperty550&serviceName=teams&threadId=19:27b9d86a8e5c4eb5ba91ca9251a87483@thread.skype&groupId=f138f4bc-e166-4ce6-ba72-9c2bc378e971>

1. Some values are not available as there is no regulatory requirement from councils to provide the information, this is done best endeavours basis [↑](#footnote-ref-1)
2. A buyer may sell a large piece of land for 113mil with many titles on and the data is incorrectly capturing 113mil for all title as the sales amount [↑](#footnote-ref-2)
3. Relates to FR-02 and a work around used for bucket requirements [↑](#footnote-ref-3)
4. Issue here is it might be a large piece of land and only 10 titles created, and the rest will be sub-divided, and properties built on them, how would you split a large sales price effectively? [↑](#footnote-ref-4)
5. The land or building was bought in a single purchase for all the related titles or land or subdivisions and the high sales values/capital values we get is repeated for all properties [↑](#footnote-ref-5)
6. Issue here is it might be a large piece of land and only 10 titles created, and the rest will be sub-divided, and properties built on them, how would you split a large sales price effectively? [↑](#footnote-ref-6)
7. Additional markers can be included at a later stage which will improve overall accuracy [↑](#footnote-ref-7)
8. Details still to be drafted and determined if this would provide a better estimate or not [↑](#footnote-ref-8)
9. Additional scenarios such as 0 sales values for multiple sales i.e. no sales vales simply won’t be addressed by FR-06 [↑](#footnote-ref-9)
10. There is a data issue with the data set, I’m currently using from the DIP, in that I’ve got a few duplicate records and the logic is picking up duplicate values – need to investigate if the logic can “skip” these [↑](#footnote-ref-10)
11. To deal with high CV values we will use a default of 200 000 for multi sales codes [↑](#footnote-ref-11)
12. Logic here still needs to be explored via the code/data – this logic doesn’t account for more than two 0 sales either previously or forward [↑](#footnote-ref-12)
13. This scenario will use x (the capital value /w premium of [21%]) [↑](#footnote-ref-13)
14. Potential miss in logic is NULL for previous and NULL for next sales with values in opposite direction (QPID: 3164461, TITLE: 336441, ADD: 123 Waihi Rd SUB: Waione) [↑](#footnote-ref-14)
15. The % increase is configurable and 21% was chosen when looking at actual sales figures to get a more precise match [↑](#footnote-ref-15)
16. Up for review when targeting the developer sales bucket [↑](#footnote-ref-16)
17. Need this as to not skew aggregate markers when there is large multi sales figure [↑](#footnote-ref-17)
18. Up for review when targeting the developer sales bucket, this number was chosen based on best results for matching [↑](#footnote-ref-18)
19. The solution assumes precedence has been applied to the seller type previously [↑](#footnote-ref-19)
20. It is assumed that the building age is grouped into categories, old, new, very old, very new, unknown [↑](#footnote-ref-20)
21. Completed, however for only the residential indicator records and not including seller information (thus far seller type has not been needed for the calculations) [↑](#footnote-ref-21)
22. NFR1, 2 and 3 are used as part of FR-08 [↑](#footnote-ref-22)
23. It is recommended to request if this valuation data is available year-on-year that we can ingest into the DIP (Inland Revenue contact up for negotiation December 2020) [↑](#footnote-ref-23)
24. Solution is currently being built stand-alone using LIN and CORELOGIC data, may be an issue for FR-15-20 [↑](#footnote-ref-24)
25. There is no underlying calculated value here its simply checking how close the aggregate value is to the FR06 value. [↑](#footnote-ref-25)
26. The percentage value is configurable and when the aggregate marker calc value is used it would fall in either QL2 or QL3 [↑](#footnote-ref-26)
27. ISV3 and ISV4 are related to requirements FR-20 - FR-23 and are only partially implemented as of 10/03/2020 [↑](#footnote-ref-27)
28. <https://www.gra.co.nz/articles-by-matthew-gilligan/are-you-a-property-dealer> [↑](#footnote-ref-28)