

CASE STUDY ANALYSIS

INF20016 BIG DATA MANAGEMENT

CHONNIKARN YUTHCHANA
STUDENT ID: 103801015

Table of Contents

1. Executive Summary.....	3
2. Introduction (321 words).....	4
2.1 Background	4
2.2 Aim & Objective	4
3. Data Quality Assessment (876 words)	5
3.1 Data Integrity and Completeness	5
3.2 Duplication Management.....	8
3.3 Data Correctness and Consistency	9
4.Victoria Market Overview (Insight 1).....	11
5. Victoria Real Estate Market Bubble Analysis(Insight2).....	12
6. Suburb Rental Price Comparison (Insight 3)	14
7. Real Estate Agency Performance and Property Type Specialization in Hawthorn: 2023-2024 (Insight 4).....	16
8. Conclusions	18
9. References	19
Appendices	20
Appendix A – Data Wrangling and Data Cleaning.....	20

Infographic

MELBOURNE PROPERTY MARKET OVERVIEW

An overview of rental affordability, property distribution, and agency specialization in Melbourne's diverse real estate market.

TOP 3



most affordable rentals suburbs in 2023

The Western and Northern Victoria regions offer the most affordable rental options, with Laverton, Melton West, and Kings Park showing the lowest median rental prices. These areas may attract budget-conscious renters and yield-driven investors.

Properties in demand within



6-16km.



2-14km.
from CBD

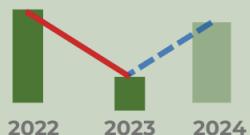


6-12km.

Distance Closer to CBD = Higher prices

Houses remain dominant within 10 km of the CBD, attracting premium buyers, while townhouses and units have higher transaction volumes in mid-range suburbs (6-12 km), targeting affordability-driven buyers.

Transaction Volume Fluctuations



Transaction volumes peaked in 2022 and dropped significantly in 2023, indicating a cooling market likely due to economic factors. A rise in listings is forecasted for 2024.

27

Solds

Jellis Craig

20

Solds

Marshall White

9

Solds

Kay

Hawthorn

Top-performing
agencies
2022-2024

In suburbs like Hawthorn, top agencies such as Jellis Craig specialize in house sales, while agencies like Marshall White focus on diverse property types, including townhouses, aligning with varied buyer needs.

1. Executive Summary

This report provides a comprehensive analysis of Melbourne's real estate market, focusing on rental affordability, property type distribution, and agency performance to assist investors in making data-driven decisions. Utilizing data from sources like Consumer Affairs Victoria and Domain.com.au, the study examines market stability, rental yields, and location-based investment strategies. A rigorous data quality assessment ensured insights were reliable by resolving duplications and handling missing values across key fields. The key objectives include understanding market stability, comparing rental yields across suburbs, examining agency specialization, and offering strategic recommendations for optimal investment based on data-driven insights.

1. **Rental Affordability:** Premium suburbs near the CBD, like Hawthorn and Hughesdale, show high rental prices due to strong demand, while outer suburbs such as Laverton and Melton West offer lower rents, presenting budget-friendly choices for renters and high-yield opportunities for investors.
2. **Location and Property Type Trends:** Proximity to the CBD significantly affects property value and demand. Houses within 10 km are highly valued, while units and townhouses, prevalent between 6-14 km, attract renters and investors seeking affordable options with central accessibility. This segmentation offers investors strategic choices between high-value investments near the CBD and rental yields in mid-range suburbs.
3. **Agency Specialization:** In suburbs like Hawthorn, agencies display specialization in certain property types. Jellis Craig leads in house sales, while Marshall White and Kay & Burton cater to townhouse sales. This insight enables investors to align with agencies specializing in their target property type.
4. **Regional Rental Trends:** Affordable rentals cluster in Western and Northern suburbs, ideal for budget-focused investments. Higher rentals in Southern and Eastern suburbs indicate central demand, guiding investors based on proximity and rental potential.

This report provides strategic guidance, helping investors balance portfolios by leveraging insights on Melbourne's real estate trends, optimizing property selection, and maximizing returns across diverse locations and property types.

2. Introduction (321 words)

2.1 Background

Melbourne Property Insights (MPI) stands at the forefront of real estate consultancy, offering a meticulous, data-driven analysis of the Melbourne property market. Our mission is to empower both investors and homeowners by utilizing a curated dataset drawn from reputable sources, including Consumer Affairs Victoria, Domain.com.au, and various government databases. In a market characterized by complexity and rapid change, making informed decisions is essential for maximizing returns while effectively managing risk. At MPI, we excel in identifying high-yield, low-risk investment opportunities and pinpointing regions poised for future value appreciation. The Melbourne property market has become a focal point of speculation, with predictions oscillating between sustained growth and potential downturns. Investors, particularly those who have amassed capital during the pandemic, are grappling with the challenges posed by this uncertainty. They require comprehensive, data-driven insights to navigate risks and optimize their investments. A deep understanding of market stability, location strategies, rental trends, and potential fluctuations is important for making sound decisions in today's unstable environment.

2.2 Aim & Objective

This analysis aims to equip MPI's investors with actionable insights, enabling them to make informed property investment decisions in the Melbourne real estate market. Our primary objectives include:

- 1. Market Stability Assessment:** Evaluating the current state of the market to determine its stability and resilience against economic shifts.
- 2. Identifying Promising Investment Locations:** Highlighting suburbs and areas that show potential for high yields and future appreciation, considering factors such as demographics and local development plans.

3. Analysis of Key Influencing Factors: Investigating elements that shape property trends, including rental yields, median house prices, and broader economic indicators, to provide a well-rounded market perspective.

4. Uncovering Hidden Trends: Employing advanced data analytics to identify emerging patterns and trends that may not be immediately apparent, guiding investors toward successful acquisitions. By addressing these focus areas, this analysis will provide crucial answers regarding market fluctuations, location advantages, and investment risks, ultimately leading to informed, strategic property decisions for our investors.

3. Data Quality Assessment (876 words)

A comprehensive data quality assessment was conducted to ensure the Melbourne property dataset addressing all critical elements across four primary areas: data integrity, duplication, completeness, and correctness. This rigorous process aimed to resolve data integrity issues and record and address anomalies systematically, ultimately creating a robust and reliable dataset for analysis as outlined below.

3.1 Data Integrity and Completeness

The dataset includes five main tables (Victoria Property 2020-2024, RE Agents, Vic Regions-Suburbs, Suburb House Price, and Suburb Rental), each providing essential information for understanding Melbourne's real estate landscape. To validate data integrity across these tables, an initial review was performed using a data dictionary to understand fields, data types, and value ranges. This revealed fields with fixed values, such as "*Property Status*" with 12 standard categories, and keywords in "*Property Type*" (e.g., 'h' for house and 'u' for units). By consistently applying these standards helped prevent data discrepancies.

In the Suburb Rental and Suburb House Price tables, several NULL values were identified across key fields (e.g., median prices for specific years). NULL values were addressed by imputing values based on calculated market median prices for each year, ensuring the dataset remained comprehensive and representative of overall trends.

However, alternative approaches were explored, such as incorporating data from reliable external sources used by MPI, including Domain and government property datasets. A comparison using Hawthorn suburb's median rent price as a reference was conducted to examine data consistency across different trusted providers. As shown in **Figure 1**, the Melbourne Property dataset indicated Hawthorn's 2023 median rent price as \$495 per week. In contrast, **Figure 2** showed data from RealEstate.com.au which recorded the 2023 rental price at \$850 per week (Real Estate, 2024), while **Figure 3** from Domain.com.au displayed a median rental price of \$750 per week in December 2023 (Domain, 2023). These external sources were cited from reputable platforms, including Consumer Affairs Victoria (Consumer Affairs Victoria, 2021).

These inconsistencies raised concerns about the reliability of external data sources. Given the substantial variation in rental prices from multiple sources, MPI opted to use internally calculated median prices to replace the missing values rather than introducing potentially conflicting external data. This decision minimized data discrepancies and provided a more consistent, reliable dataset for investors.

Suburbs	Dec 2021 Median	Dec 2022 Median	Dec 2023 Median
Hawthorn	375	400	495

Fig 1 – 2023 Rental Median Price in Hawthorn in current dataset

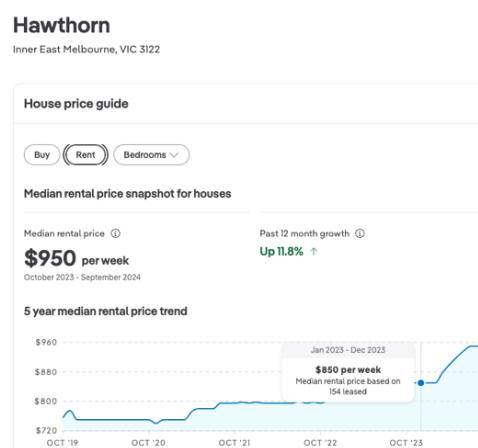


Fig 2 – 2023 Rental Median Price in Hawthorn from Realestate.com website

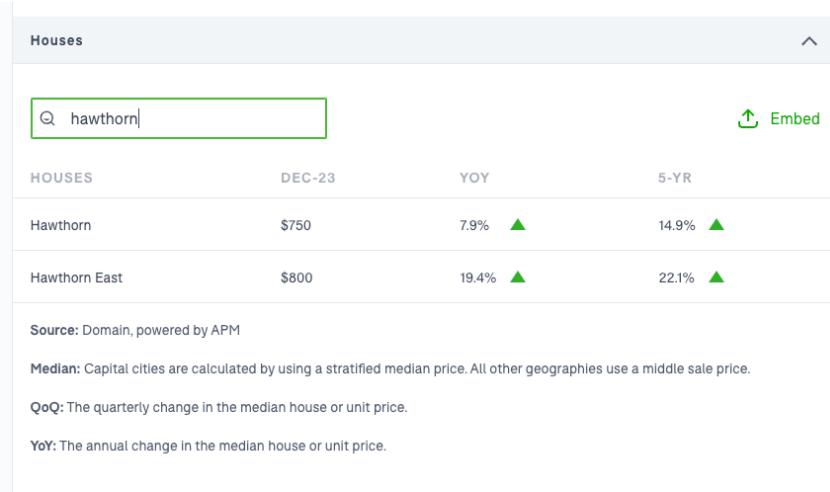


Fig 3 – 2023 Rental Median Price in Hawthorn from Domain.com website

Additionally, a data integrity issue was identified after joining the Vic Regions-Suburbs table with the Suburb Rental and Suburb House Price tables. The suburbs of Melbourne and Bellfield lacked records in both the rental and house price tables, resulting in NULL values for key fields like rental and median house prices. **Figure 4** shows the NULL values for these suburbs in the joined tables. To resolve this, records for Melbourne and Bellfield were manually added, using each year's calculated market median prices. This approach ensured consistency across the dataset without introducing external biases and filled data gaps, maintaining trend accuracy for all suburbs. And those missing value suburbs will be noted to the audience.

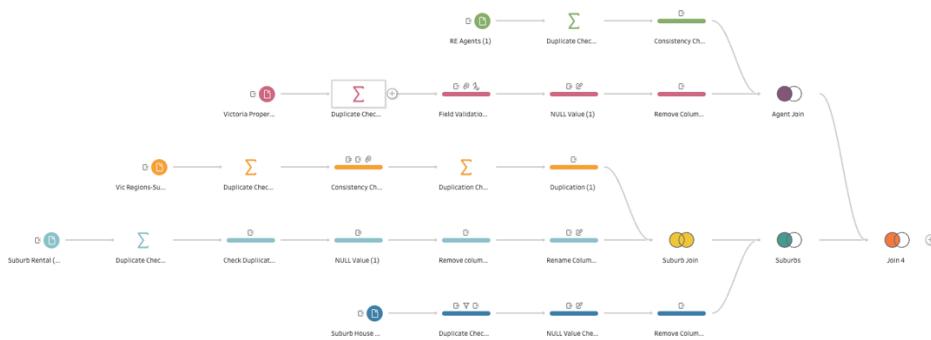


Fig 4- Joining Process

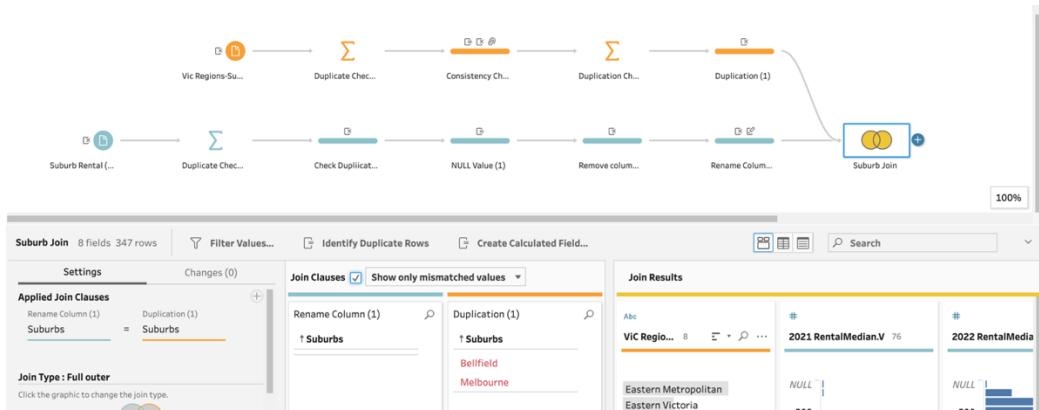


Fig 5- NULL Values for Bellfield and Melbourne in the Joined Suburb Rental and Suburb House Price Tables.

Thus, the approach of replacing the NULL value with median prices for specific years has been employed rather than using a data from another source with inconsistency data source issue. Moreover, suburbs with missing data should be flagged for further investigation to conduct a root-cause analysis into data sourcing issues, enhancing data completeness and reliability.

3.2 Duplication Management

Duplication was systematically identified and resolved across all tables. For instance, in the Suburb Rental table, duplicate entries for suburbs like Camberwell were removed after validation, ensuring only unique entries remained. The Aggregation function was used to detect and remove duplicate records in tables like Victoria Property, where suburbs such as Nunawading had multiple identical listings. By retaining only unique records, we ensured a streamlined dataset, facilitating accurate analysis and reducing the risk of bias from duplicated entries.

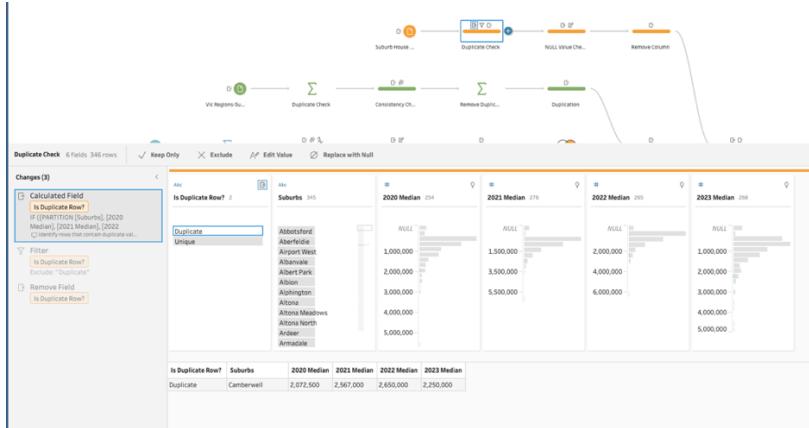


Fig 6 – Duplicate Rows identify in Suburb House Price table

3.3 Data Correctness and Consistency

Data correctness was another key focus. Inconsistencies such as regional misassignments and misspellings were corrected; for example, Camberwell's incorrect assignment to both "Eastern Metropolitan" and "Southern Metropolitan" as in Figure 6 was resolved by assigning it to the correct region based on authoritative sources. Additionally, misspellings in suburb names, such as 'Melbrne' instead of 'Melbourne,' were corrected as in Figure 7, enhancing the clarity and consistency of the data. Property types were standardized, grouping values like 'h,' 'home,' and 'unit' into consistent categories (house, townhouse, and unit) follows by Figure 8. Where data inconsistencies existed with external sources, such as rental prices differing from online real estate platforms, median prices were selected as the imputation strategy to preserve data consistency.

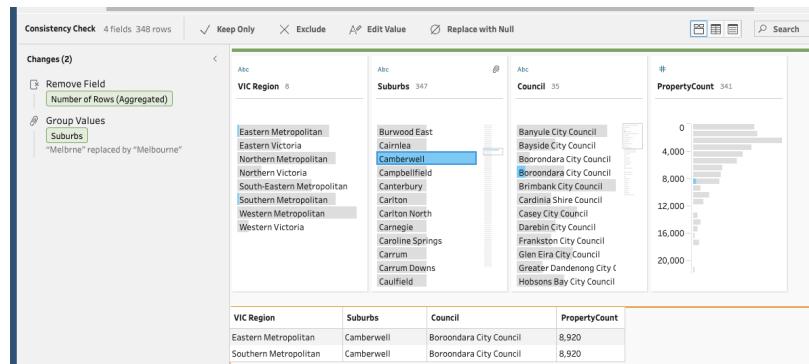


Fig 7- A wrong region assigned to Camberwell Suburb

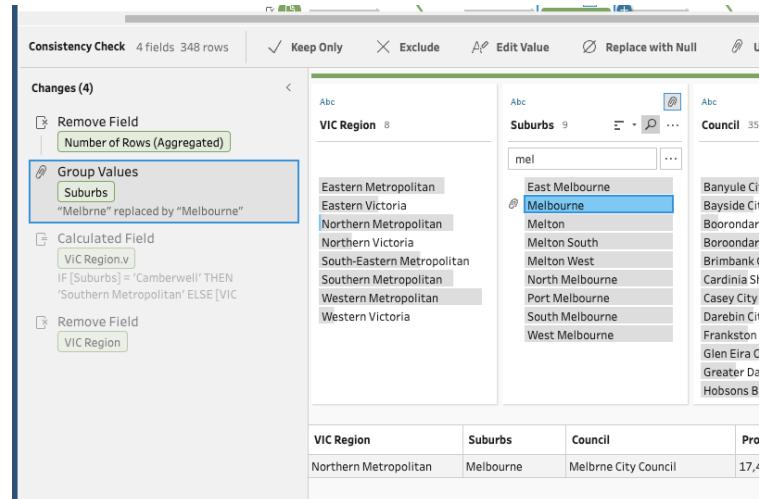


Fig 8 – Steps of resolving suburb misspelling issue.

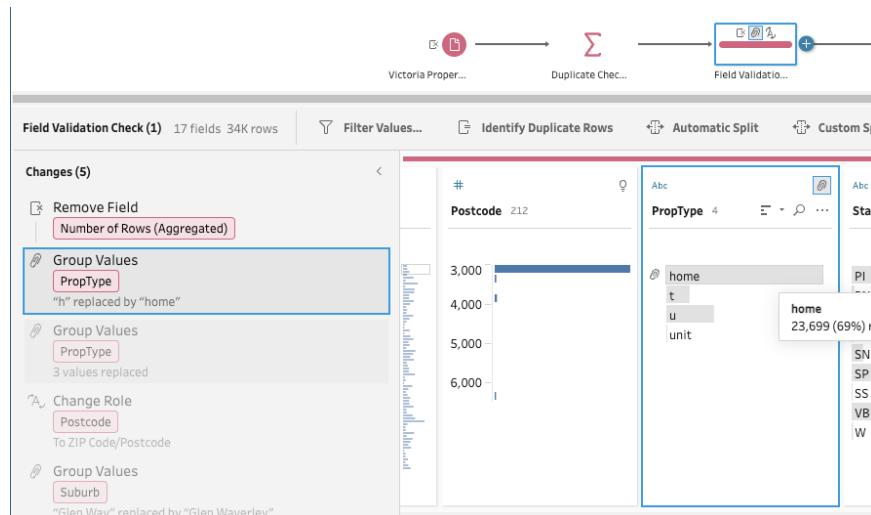


Fig 9 – Properties types Classification after Grouping

Moreover, during analysis, the postcode correctness data has been identified that there was 6262 postcode which is a western Australia suburb and deep investigate this record was from Nunawading Record so to resolve this issues the record has been correctly assigned to 3131 for Nunawading

The steps taken to address these data quality issues have resulted in a clean and reliable dataset, ready for advanced analysis. Further details on the data cleaning methods, including additional steps and justifications, are provided in the [Appendix A – Data Wrangling and Data Cleaning](#) for reference.

4. Victoria Market Overview (Insight 1)

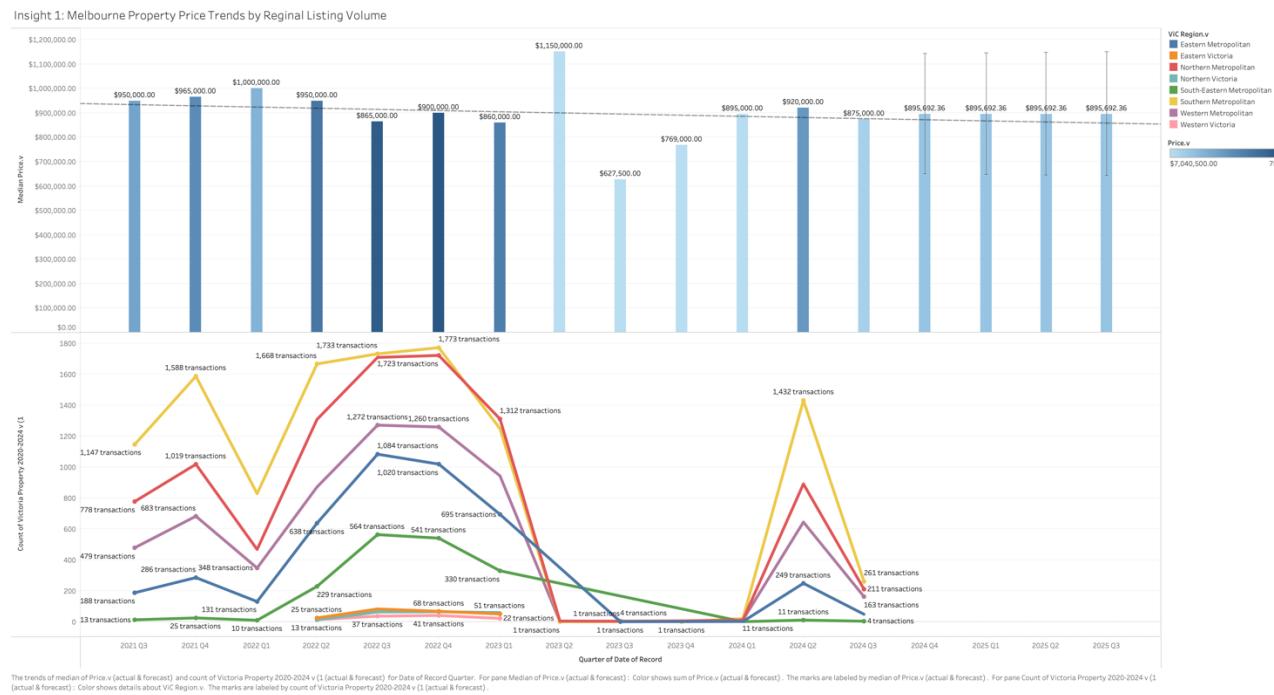


Fig 10- Insight 1 Melbourne Property Price Trends by Regional Listing Volume

This visualization provides a nuanced view of Melbourne's property market, showing a relatively stable median price trend with notable fluctuations in transaction volumes across regions and quarters from 2021 Q1 to 2024 Q3. The forecasted median price between **2024 Q4 and 2025 Q3** indicates a continuation of the stable trend around \$895,000, suggesting market resilience amid economic fluctuations. Regions with consistently high transaction counts, **Southern Metropolitan, Northern Metropolitan, and Western Metropolitan**, demonstrate robust demand, aligning with Gyourko, Mayer, and Sinai's (2013) findings that strong regional demand influences price stability in real estate markets. The significant transaction volumes in these areas reflect sustained interest, potentially due to favourable facilities or economic opportunities that attract both buyers and investors. Lower transaction volumes, especially in **Northern and Western Victoria**, suggest less competitive environments, which might appeal to buyers seeking quieter or undervalued areas. As Case and Shiller (2003) emphasize, understanding these dynamics is crucial for investors, as high-transaction regions with stable prices often present lower-risk investment opportunities.

5. Victoria Real Estate Market Bubble Analysis(Insight2)

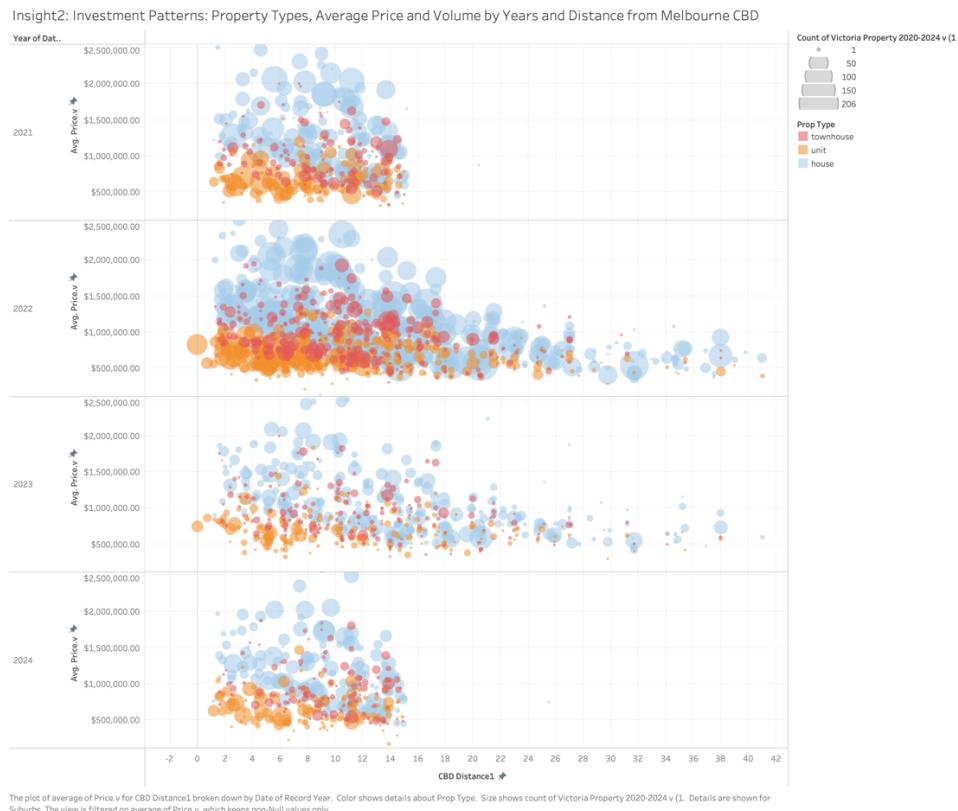


Fig 11- Insight 2 Investment Patterns: Property Types, Average Price and Volume by Years and Distance from Melbourne CBD

The bubble analysis of Melbourne property types reveals distinct investment patterns based on proximity to the CBD. Properties closer to the CBD, particularly within 10 km, command higher prices, with houses in high demand due to their larger size and central location. This aligns with research by Manning, Rodriguez, and Ghosh (1999), which highlights that properties in central areas often offer revenue-enhancing advantages due to accessibility and proximity to high-demand amenities. Additionally, Saberi et al. (2017) underscore the impact of accessibility, noting that proximity to transport and amenities significantly increases property demand and value in metropolitan areas.

Units are concentrated within 14 km of the CBD, typically priced below \$1 million, attracting investors seeking affordable entry points and stable rental yields.

Townhouses, commonly found between 6-12 km, offer moderate pricing, appealing to investors aiming for balanced capital growth and rental returns. Houses are more broadly distributed, with significant volumes within 6-16 km, making them ideal for long-term investments due to their potential for high capital appreciation.

Sales volume trends show a peak in transactions within 20 km of the CBD during 2022, followed by a decline in 2023. By 2024, sales were primarily concentrated within 16 km, with fewer properties sold in outer suburbs. This change reveals a growing investor preference for central locations, likely driven by accessibility and stable demand in inner areas. This distribution of property types by distance allows investors to align their strategies with financial goals, targeting central areas for high-value investments while opting for townhouses and units in suburban areas for rental income and moderate growth.

6. Suburb Rental Price Comparison (Insight 3)

Insight 3: 2023 Top 3 Most Affordable Suburbs by Median Rental Price and Region

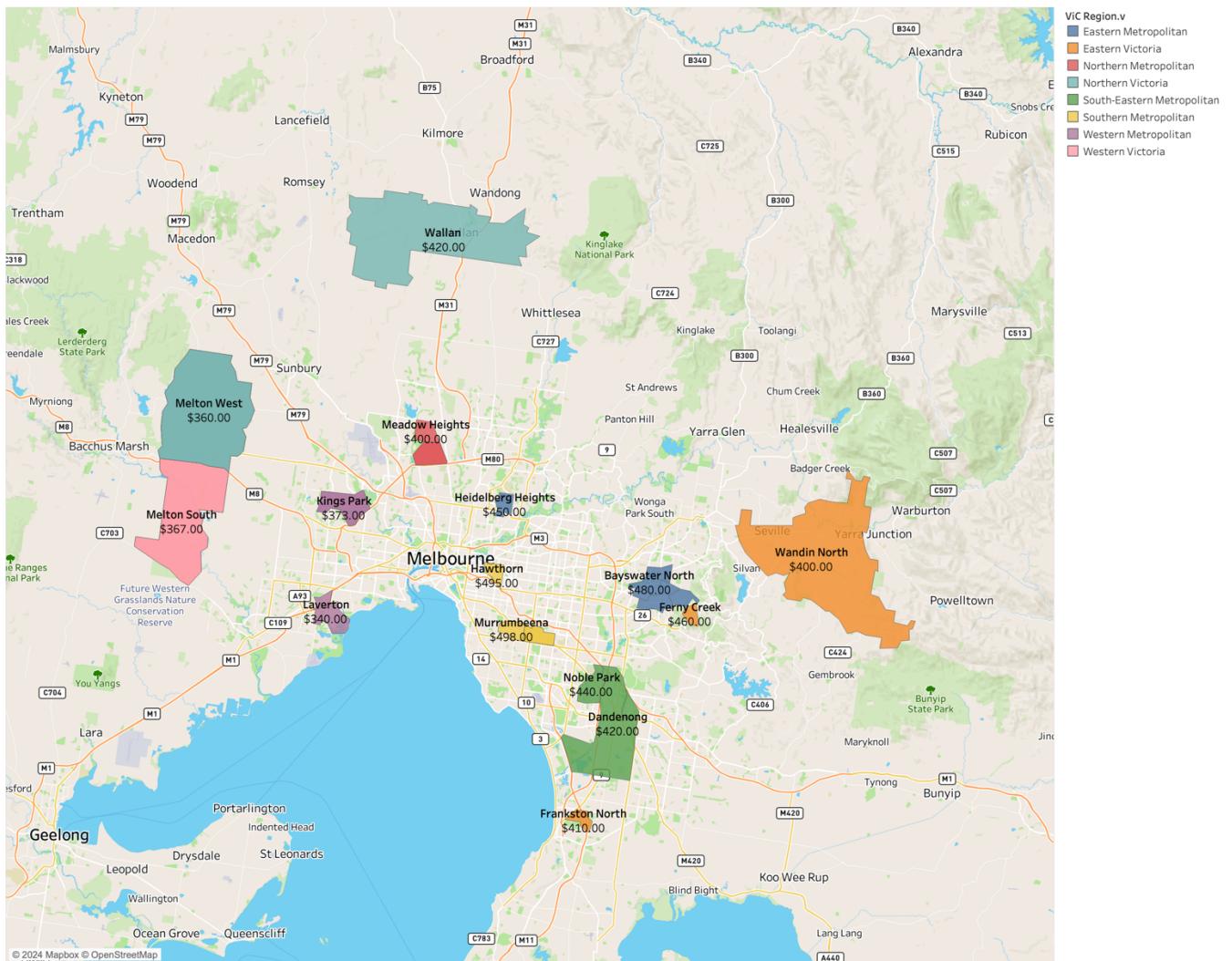


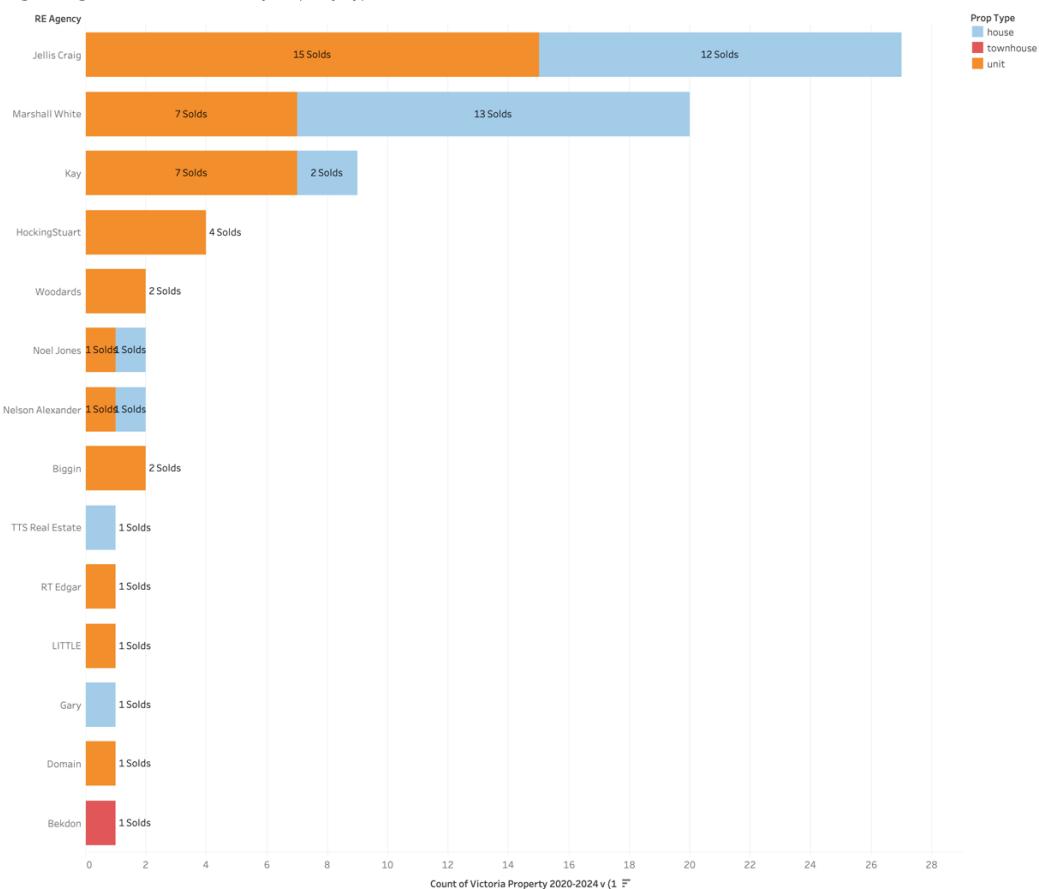
Fig12- Insight 3: 2023 Top 3 Most Affordable Suburbs by Median Rental Price and Region.

The 2023 rental price map highlights Melbourne's most affordable suburbs, primarily located in Western Victoria and Northern Metropolitan areas. Suburbs like Laverton (\$340), Melton West (\$360), and Kings Park (\$373) offer low rental costs, making them attractive for budget-conscious renters and investors seeking steady demand.

In contrast, Southern and Eastern Metropolitan areas command higher rental prices, such as Hughesdale (\$498) and Murrumbeena (\$498), likely due to their nearness to transport links and central amenities. Research shows that accessibility to public transport and amenities can significantly drive-up rental values in metropolitan areas, attracting professionals seeking suburban living with central convenience (Saberi et al., 2017). Furthermore, urban regeneration efforts often lead to increased demand and rental prices in well-connected regions, as renters balance affordability with access to infrastructure and services (Sage Journals, 2020). This insight helps renters and investors align their choices with budget considerations, illustrating how rental affordability shifts with regional proximity and demand dynamics. Renters prioritizing cost savings might target Western and Northern suburbs, while investors seeking premium returns can focus on Eastern and Southern regions with higher rental values.

7. Real Estate Agency Performance and Property Type Specialization in Hawthorn: 2023-2024 (Insight 4)

Insight 4: Agent Sold Performance by Property Types in Hawthorn on 2023 & 2024



Count of Victoria Property 2020-2024 v (1 for each RE Agency. Color shows details about Prop Type. The marks are labeled by count of Victoria Property 2020-2024 v (1. The data is filtered on Status, Date of Record Year and Suburb. The Status filter keeps Sold, Sold After Auction, Sold not disclosed, Sold Prior and Sold after auction price not disclosed. The Date of Record Year filter keeps 2023 and 2024. The Suburb filter keeps Hawthorn.

Fig13- Agent Sold Performance by Property Types in Hawthorn for 2023 & 2024

This visualization showcases the performance of real estate agencies in Hawthorn across property types (house, townhouse, and unit) in 2023 and 2024. Jellis Craig remains the top-performing agency with 27 sales, primarily focusing on houses, followed by Marshall White with 20 sales across houses and townhouses. Kay & Burton also shows significant activity, particularly in townhouse sales. The concentration of house sales among top agencies reflects strong demand for high-value properties in Hawthorn. According to research (Turnbull & Waller, 2018), agency specialization can enhance market performance by aligning with client preferences and building a strong

reputation. Agencies like Jellis Craig and Marshall White have leveraged their expertise in house sales, capturing a competitive advantage by meeting the high demand in this category. Meanwhile, Marshall White and Kay & Burton's diversification into townhouse sales may be a strategic move to cater to a growing segment of buyers looking for more affordable, centrally located options. For sellers and investors, this insight underscores the benefits of choosing agencies with a proven track record in specific property types, aligning with the strategic advantage specialization provides in competitive markets.

8. Conclusions

This analysis concludes that Melbourne's real estate market offers diverse investment opportunities across various regions and property types. While the CBD-adjacent suburbs remain premium markets due to accessibility and high demand, outer regions present affordable options with steady rental yields. Specialized agencies demonstrate performance advantages in specific property types, particularly in competitive suburbs like Hawthorn. Addressing data integrity challenges strengthened the analysis, underscoring the importance of accurate data in high-stakes investment contexts. Overall, the findings support strategic, data-driven decision-making for investors aiming to balance affordability, growth potential, and risk.

9. References

- Consumer Affairs Victoria, V. G. (2021, May 4). *Property data*.
Www.consumer.vic.gov.au. <https://www.consumer.vic.gov.au/housing/buying-and-selling-property/property-data>
- Domain. (2023, December). *Domain Rental Report - December 2023 | Domain*.
Domain. <https://www.domain.com.au/research/rental-report/december-2023/>
- Manning, C., Rodriguez, M., & Ghosh, C. (1999). Devising a Corporate Facility Location Strategy to Maximize Shareholder Wealth. *Journal of Real Estate Research*, 17(3), 321–340. <https://doi.org/10.1080/10835547.1999.12090981>
- Palm, M., Raynor, K. E., & Warren-Myers, G. (2020). Examining building age, rental housing and price filtering for affordability in Melbourne, Australia. *Urban Studies*, 58(4), 809–825. <https://doi.org/10.1177/0042098020927839>
- Real Estate. (2024). *Real Estate, Property & Homes for Sale - realestate.com.au*.
Realestate.com.au. <https://www.realestate.com.au/>
- Saberi, M., Wu, H., Amoh-Gyimah, R., Smith, J., & Arunachalam, D. (2017). Measuring housing and transportation affordability: A case study of Melbourne, Australia. *Journal of Transport Geography*, 65, 134–146. <https://doi.org/10.1016/j.jtrangeo.2017.10.007>
- Turnbull, G. K., & Waller, B. D. (2018). (What) do top performing real estate agents deliver for their clients? *Journal of Housing Economics*, 41, 142–152.
<https://doi.org/10.1016/j.jhe.2018.06.005>
- Victorian Electoral Commission. (n.d.). *Southern Metropolitan Region*.
Www.vec.vic.gov.au. <https://www.vec.vic.gov.au/electoral-boundaries/state-regions/southern-metropolitan-region>
- Your Investment property. (n.d.). *Hawthorn, VIC 3122: Suburb Profile & Property Report | YIP*. Yourinvestmentpropertymag.com.au.
<https://www.yourinvestmentpropertymag.com.au/top-suburbs/vic/3122-hawthorn>

Appendices

Appendix A – Data Wrangling and Data Cleaning

1. Review Data Dictionary

Firstly, begin by reviewing the Data Dictionary that attached in Melbourne Property Insight 2020 -2023 to ensure a clear understanding of the fields, data types and value ranges for each dataset.

By investigating the data dictionary sheet, there are 5 tables includes Victoria Property 2020-2024, RE Agents, Vic Regions-Suburbs, Suburb House Price and Suburb Rental. As a result, some fields have been stated that there are fixed value like Property Status which includes 12 value and Property Type that uses keywords like h for house and u for units.

2. Suburb Rental Table

- Using Aggregate function to check the duplication: as a result, there are duplicated rows in Camberwell Suburbs.
- Data Completeness issues: NULL Values in Dec 2021 Median, Dec 2022 Median and Dec 2023 Median. With the 16 suburbs that the value of median price are missing,

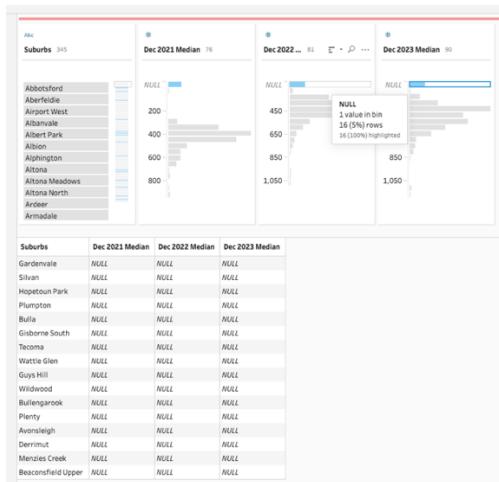


Fig x – Data Completeness issues in Suburb Rental Table

To resolve the problems, each year market median price has been introduced by calculating from all suburbs' median price and Replacing the NULL value in each year by the market median price value.

There are many approaches to resolving the NULL values in table includes finding a reference median price of the suburb from the trustable source such as RealEstaste website, Domain website however from the research show that the value can take from the website also but the dataset will have inconsistency issues for example in dataset, Hawthorn has a median rental price in 2023 as 495\$ however in the realestate website the median rental price in 2023 is 850\$ per week and in Domain Website is 750\$ per week however in dataset the median rental price in 2023 is 495\$ per week. Which make this approach drop due to the inconsistency that happened. ned. So the solution is replacing NULL value with the median price of a market in each year and the root cause analysis need to be examine to further deep into which source the data came from to enhance a data consistency and completeness

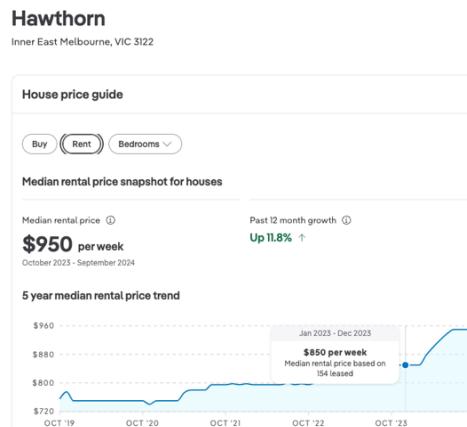


Fig x – 2023 Rental Median Price in Hawthorn from realestate website

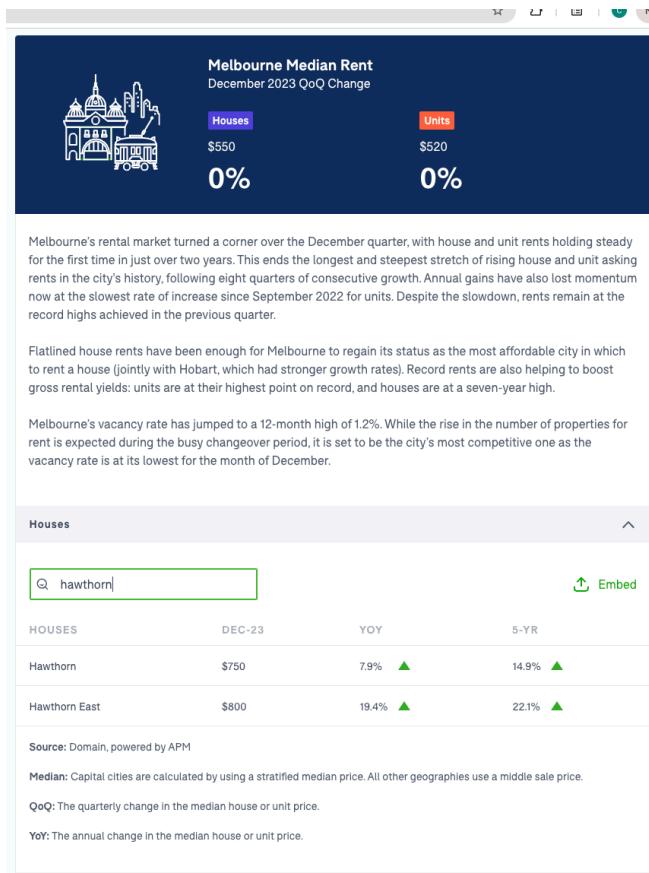


Fig x – 2023 Rental Median Price in Hawthorn from Domain website

Suburbs	Dec 2021 Median	Dec 2022 Median	Dec 2023 Median
Hawthorn	375	400	495

Fig x – 2023 Rental Median Price in Hawthorn in current dataset

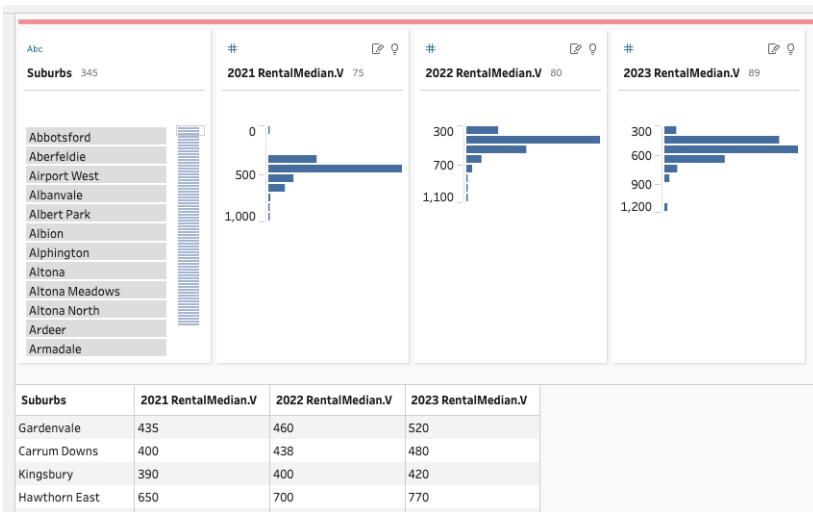


Fig x – The Completeness of Data after Replacing NULL Value

2. Suburb House Price Table

1. Data Duplication Issue, by identifying the duplication value using 'Identify Duplicate Rows' and have been remove the duplicated rows which is Camberwell.

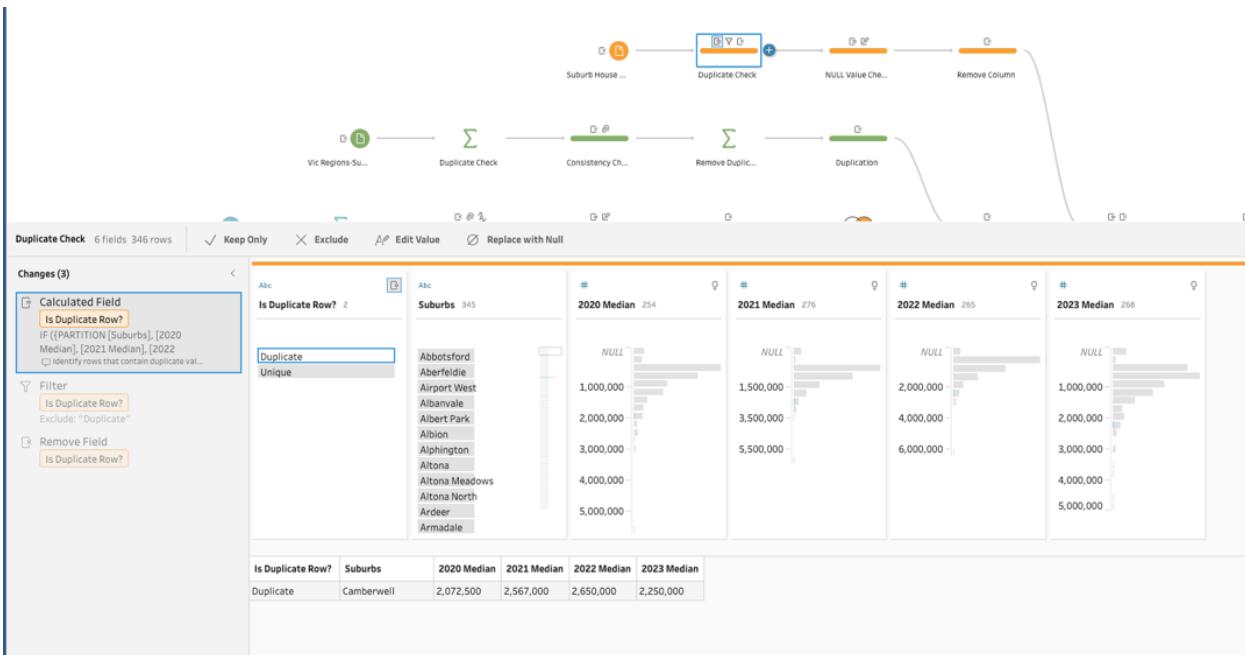


Fig x – Duplicate Rows identify in Suburb House Price table

2. Data Completeness Issues

From identifying the data completeness in the table from 2020 to 2022 Median Price has a NULL values in 13 suburbs includes Dockland, Hopetoun park, Silvan, Kooyong, Guys Hill, Bulla, Wildwood, Bullengarook, Gisborne South, Plumpton, Southbank, Caulfield East and Hillside. And 1 suburb, Caufield has been missing 2023 Median Price.

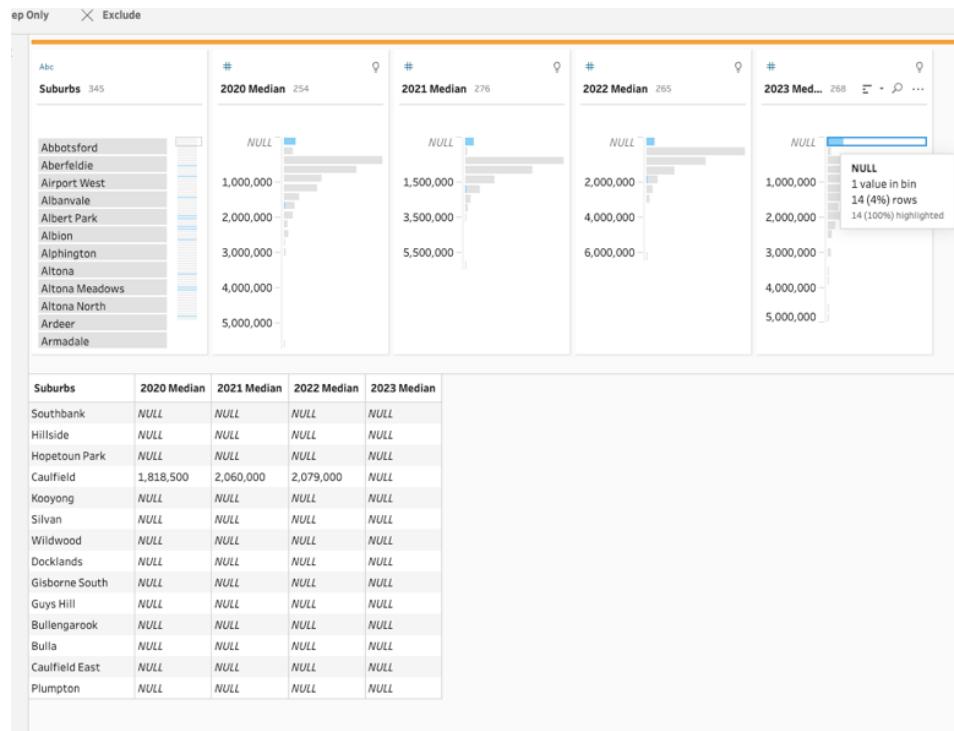


Fig x- NULL Values in 2023 median Price in Suburb House Price table

The same approaches as in Suburb Rental Price have been introduced by calculating market median price for each year and replacing the NULL value. Additionally, if burbs like Caufield have a record of median prices in recent years from 2020 to 2022, the median price that missing in 2023 is replaced by the average of median price from 2020-2022. As in the photo below, the market median price of each year has been calculated in Market Median Field and the each year Median Price.v is a verified row.

NULL Value Check		13 fields	345 rows	Keep Only	Exclude	1 Recommendation	Print	Copy	Search										
Changes (9)																			
	Suburbs	345	2020 Median Price.V	254	2020 Median	254	2020 Market Median	1	2021 Median	276	2021 Market Median	1	2021 Median Price.V	133	2022 Median	265	2022 Market Median	1	2022 Median Price.V
Calculated Field	Abbotsford		0	NULL	0	NULL	0	NULL	0	NULL	0	NULL	0	NULL	0	NULL	0	NULL	
Calculated Field	Aberfeldie		3,000,000	NULL	6,000,000	NULL	1,000,000	NULL	1,500,000	NULL	3,000,000	NULL	3,000,000	NULL	6,000,000	NULL	2,000,000	NULL	
Calculated Field	Albert Park		2,000,000	NULL	3,000,000	NULL	3,500,000	NULL	5,500,000	NULL	4,000,000	NULL	6,000,000	NULL	7,000,000	NULL	7,000,000	NULL	
Calculated Field	Altmore		4,000,000	NULL	4,000,000	NULL	4,000,000	NULL	4,000,000	NULL	4,000,000	NULL	4,000,000	NULL	4,000,000	NULL	4,000,000	NULL	
Calculated Field	Altmore Meadows		5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	
Calculated Field	Altmore North		5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	
Calculated Field	Ardeer		5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	
Calculated Field	Armadale		5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	5,000,000	NULL	
Renamed Field	Decklands	892,500	2020 Median Price.V	892,500	2020 Median	892,500	2020 Market Median	1,050,000	2021 Median	1,050,000	2021 Market Median	1,050,000	2022 Median	1,050,000	2022 Market Median	1,050,000	2023 Median	1,050,000	2023 Market Median
Renamed Field	Hopetoun Park	892,500	2020 Median Price.V	892,500	2020 Median	892,500	2020 Market Median	1,050,000	2021 Median	1,050,000	2021 Market Median	1,050,000	2022 Median	1,050,000	2022 Market Median	1,050,000	2023 Median	1,050,000	2023 Market Median
Renamed Field	Silvan	892,500	2020 Median Price.V	892,500	2020 Median	892,500	2020 Market Median	1,050,000	2021 Median	1,050,000	2021 Market Median	1,050,000	2022 Median	1,050,000	2022 Market Median	1,050,000	2023 Median	1,050,000	2023 Market Median
Calculated Field	Kooyong	892,500	2020 Median Price.V	892,500	2020 Median	892,500	2020 Market Median	1,050,000	2021 Median	1,050,000	2021 Market Median	1,050,000	2022 Median	1,050,000	2022 Market Median	1,050,000	2023 Median	1,050,000	2023 Market Median
Calculated Field	Guy's Hill	892,500	2020 Median Price.V	892,500	2020 Median	892,500	2020 Market Median	1,050,000	2021 Median	1,050,000	2021 Market Median	1,050,000	2022 Median	1,050,000	2022 Market Median	1,050,000	2023 Median	1,050,000	2023 Market Median
Calculated Field	Bulla	892,500	2020 Median Price.V	892,500	2020 Median	892,500	2020 Market Median	1,050,000	2021 Median	1,050,000	2021 Market Median	1,050,000	2022 Median	1,050,000	2022 Market Median	1,050,000	2023 Median	1,050,000	2023 Market Median
Calculated Field	Wildwood	892,500	2020 Median Price.V	892,500	2020 Median	892,500	2020 Market Median	1,050,000	2021 Median	1,050,000	2021 Market Median	1,050,000	2022 Median	1,050,000	2022 Market Median	1,050,000	2023 Median	1,050,000	2023 Market Median
Calculated Field	Bulleen Park	892,500	2020 Median Price.V	892,500	2020 Median	892,500	2020 Market Median	1,050,000	2021 Median	1,050,000	2021 Market Median	1,050,000	2022 Median	1,050,000	2022 Market Median	1,050,000	2023 Median	1,050,000	2023 Market Median
Calculated Field	Glenmore South	892,500	2020 Median Price.V	892,500	2020 Median	892,500	2020 Market Median	1,050,000	2021 Median	1,050,000	2021 Market Median	1,050,000	2022 Median	1,050,000	2022 Market Median	1,050,000	2023 Median	1,050,000	2023 Market Median
Calculated Field	Pumpherston	892,500	2020 Median Price.V	892,500	2020 Median	892,500	2020 Market Median	1,050,000	2021 Median	1,050,000	2021 Market Median	1,050,000	2022 Median	1,050,000	2022 Market Median	1,050,000	2023 Median	1,050,000	2023 Market Median
Calculated Field	Southbank	892,500	2020 Median Price.V	892,500	2020 Median	892,500	2020 Market Median	1,050,000	2021 Median	1,050,000	2021 Market Median	1,050,000	2022 Median	1,050,000	2022 Market Median	1,050,000	2023 Median	1,050,000	2023 Market Median
Calculated Field	Caulfield East	892,500	2020 Median Price.V	892,500	2020 Median	892,500	2020 Market Median	1,050,000	2021 Median	1,050,000	2021 Market Median	1,050,000	2022 Median	1,050,000	2022 Market Median	1,050,000	2023 Median	1,050,000	2023 Market Median
Calculated Field	Hillside	892,500	2020 Median Price.V	892,500	2020 Median	892,500	2020 Market Median	1,050,000	2021 Median	1,050,000	2021 Market Median	1,050,000	2022 Median	1,050,000	2022 Market Median	1,050,000	2023 Median	1,050,000	2023 Market Median

Fig x – The Table after Replacing NULL value with the market median price

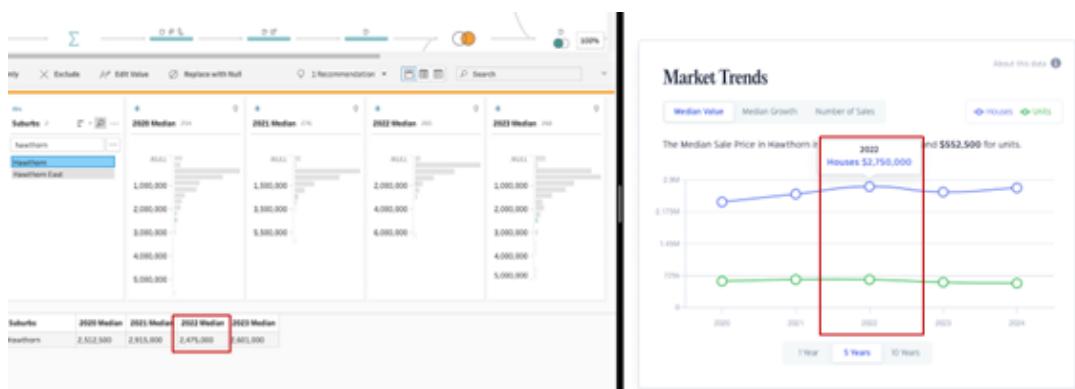


Fig x – compare the value of 2022 Median Price in Hawthorn in database and source from YourInvestmentProperty

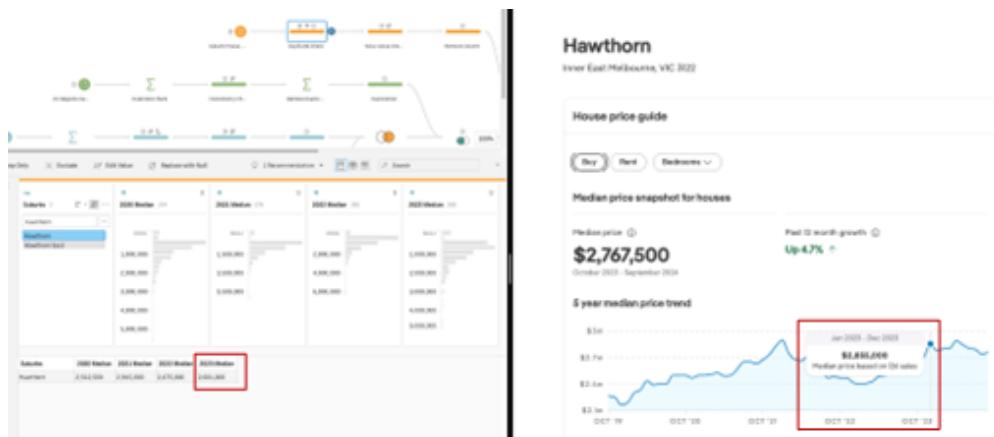


Fig x – compare the value of 2023 Median Price in Hawthorn in database and source from RealEstate

From two picture above, show that the data completeness also can be replaced by the information from website as well but the data source of current dataset and the source from internet is not consistent and accurate so replacing with the median value has been chosen.

3. Vic Region Suburbs Table

This table includes the overview of each suburb like the Victoria region where the suburb is located, council that take care and the Property Count.

- Firstly, investigating into **data duplication**, as a result, the table has a unique rows of 347 suburbs.

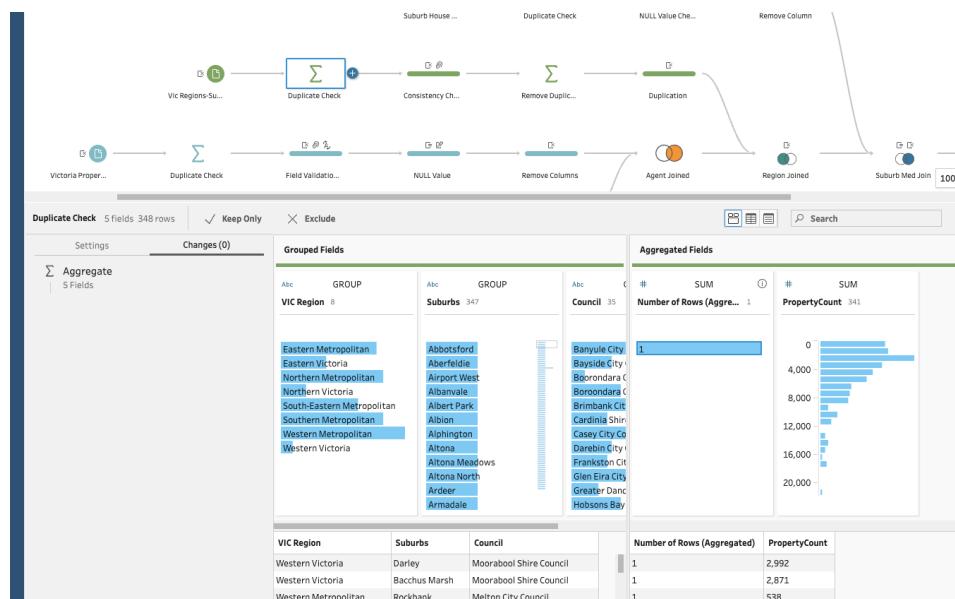


Fig x – Results of Identifying Duplicate Rows

- The **data correctness** has been investigated that Camberwell has been assigned to both regions Eastern Metropolitan and Southern Metropolitan which correctly Camberwell should be assigned to Southern Metropolitan (Victorian Electoral Commission, n.d.).

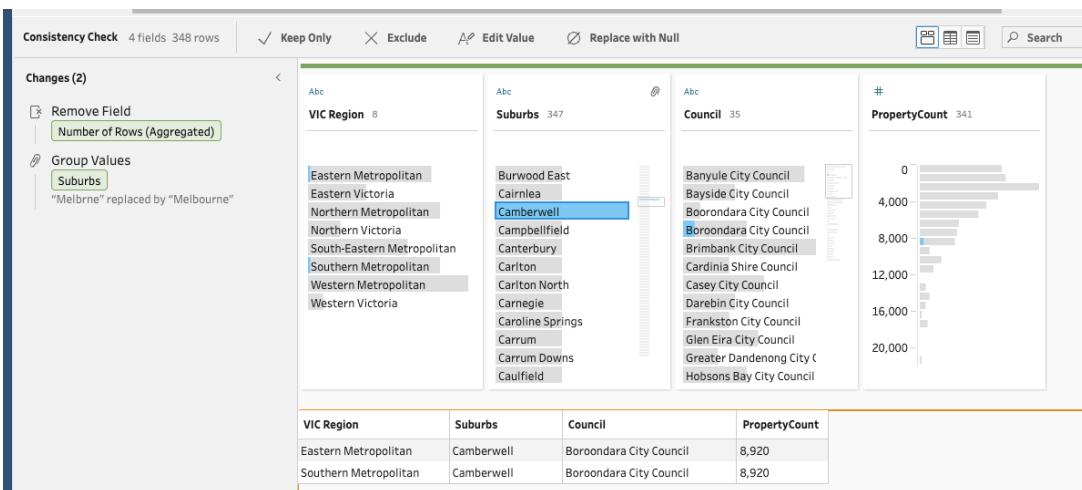


Fig x- A wrong region assigned to Camberwell Suburb

To resolve the issue, another calculation field which contains a function to check if the suburb is Camberwell then change the Region to 'Southern Metropolitan'. After that the wrong assigned Region has been changed from Eastern to Southern Metropolitan. Moreover, Aggregate function for identifying the duplicate rows of Camberwell suburbs has been identified as well.

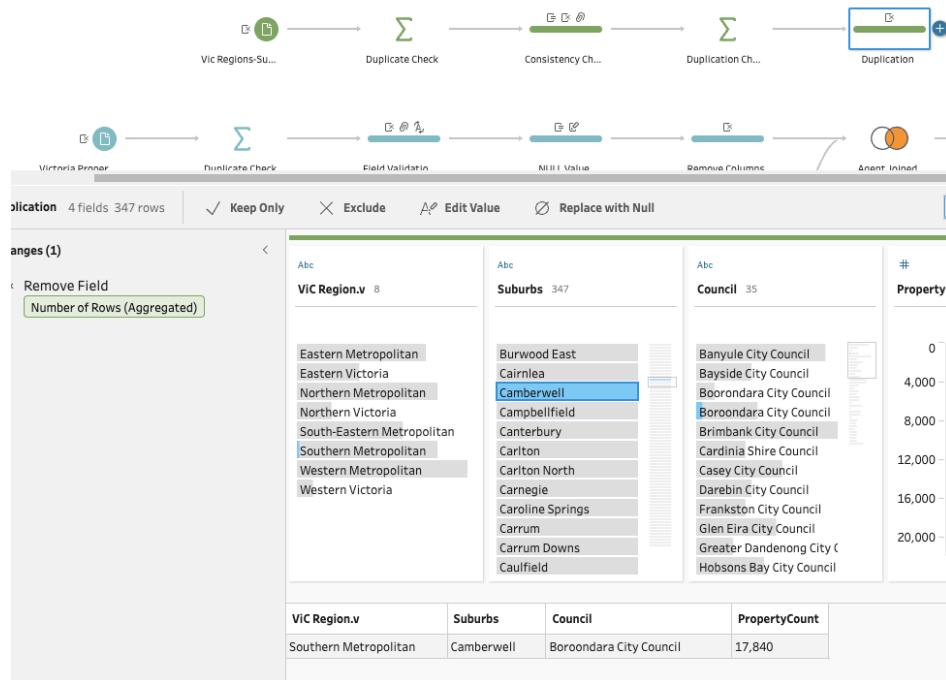


Fig x – Camberwell has been assigned to the right region.

c. Data Misspelling

In the Region Column, from the investigating, Melbourne suburb has been misspelling from 'Melbrne'. Then this value has been change to the correct spelling 'Melbourne'.

The screenshot shows a data consistency check interface with the following details:

- Changes (4)**:
 - Remove Field: Number of Rows (Aggregated)
 - Group Values: Suburbs, "Melbrne" replaced by "Melbourne"
 - Calculated Field: VIC Region.v
IF [Suburbs] = 'Camberwell' THEN 'Southern Metropolitan' ELSE [VIC Region]
 - Remove Field: VIC Region
- VIC Region**: A list of regions: Eastern Metropolitan, Eastern Victoria, Northern Metropolitan, Northern Victoria, South-Eastern Metropolitan, Southern Metropolitan, Western Metropolitan, Western Victoria.
- Suburbs**: A list of suburbs starting with 'mel': East Melbourne, Melbourne, Melton, Melton South, Melton West, North Melbourne, Port Melbourne, South Melbourne, West Melbourne.
- Council**: A list of councils: Banyule City, Bayside City, Boorondara, Boroondara, Brimbank City, Cardinia Shire, Casey City, Darebin City, Frankston, Glen Eira City, Greater Dandenong, Hobsons Bay.
- Table View**: A table showing the resolution:

VIC Region	Suburbs	Council	Pro
Northern Metropolitan	Melbourne	Melbrne City Council	17,4

Fig x – Steps of resolving suburb misspelling issue.

4. Victoria Property 2021-2024 table

From this table contains the data of listing transaction which includeed the agent, address and the overview insight of the listing realestate. The data quality has been investigated as belows:

a. Identifying Duplication Data

The table has been identify the duplicated rows in Nunawading suburbs and already removed inside the Aggregate Step.

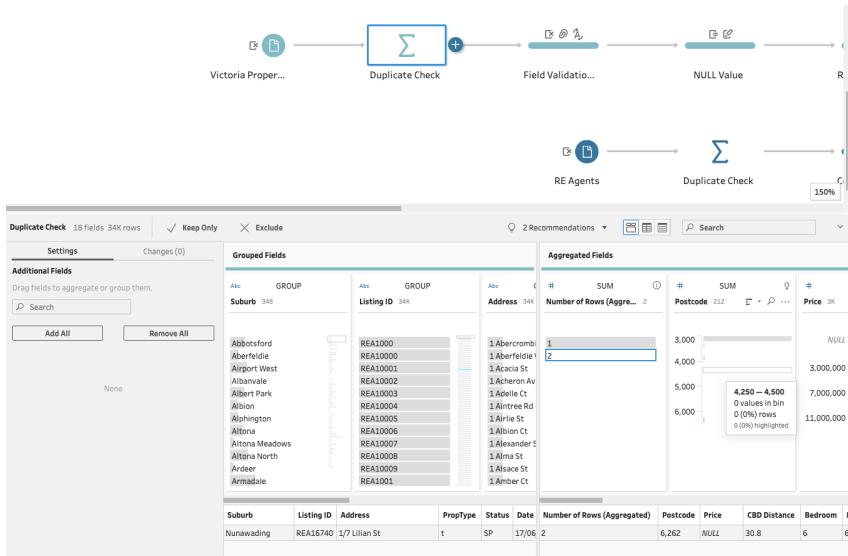


Fig x – Identifying the duplicate value in the table

b. Identifying Data Validation in Property Type Column

From the Data Dictionary, the property types have been described in like h-house, t-townhouse, u-unit/flat/apartment which in the dataset the value is not classified properly which includes h, home, t, u, and unit. To reslove this problem, the same property types has been group as 3 classification which are 'home, townhouse, and unit'

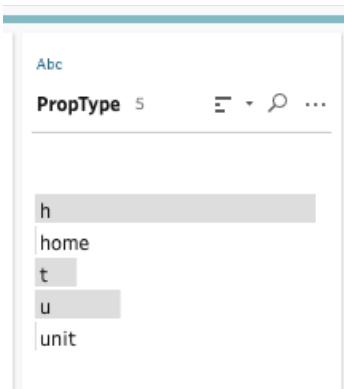


Fig x – Properties types before data cleaning

Suburb	Listing ID	Address	PropType
Fitzroy	REA12849	29 Bell St	townhouse
Broadmeadows	REA23633	3/61 Cuthbert St	house
Newport	REA9777	9 Blackshaws Rd	house
Sunshine	REA2433	6 robinson St	house
Airport West	REA3805	2/137 McNamara Av	house
Richmond	REA30180	6/4 McGrath Ct	unit

Fig x – Properties types Classification after data cleaning

c. Identifying Data Validation in Suburbs Column

The Results of the investigation show that there are a misspelling of the Glen Waverley suburb which has 2 values include 'Glen Wav' and 'Glen Waverley' so the solution is grouping both value together and rename to the correct spelling.

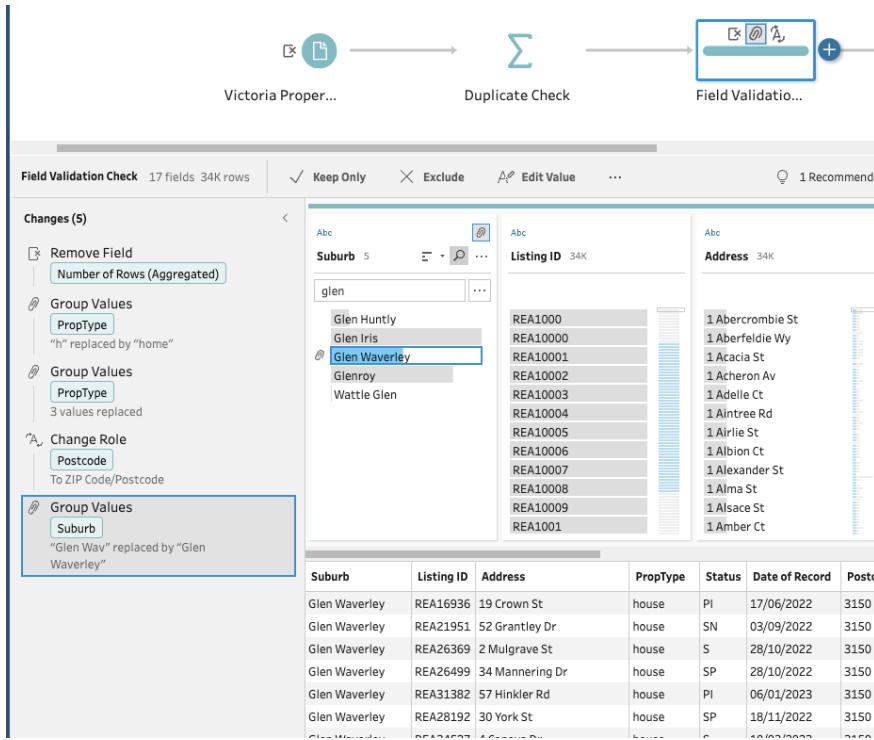


Fig x – Grouping Glen Waverley Suburb

d. Data Completeness

There are 3 column that contains a NULL value including Price, Bathroom, Car-Garage.

- Price :** The approach to resolve the issue is create a suburb median price which referencing the price of each suburb and then calculating the each suburb median price after that the price column has been updated if the null value fields has been replace with that suburb median price and if there are null value in suburb median price is replace with 'N/A' which let the stakeholders knows that this needs to be investigated further.
- Bathroom and Car Garage:** The NULL value in this column has been replaced by 'N/A' value which is better from leave the value as 'NULL' which can be the case that the data is missing during the entering the data.

5. RE Agents

This table contains the records of agent information in each listing transactions. Firstly the data duplication has been identified in Jellis Craig Rows and has been removed by Aggregate Step. Additionally, the quality of data in this table is good nothing to be changed.

6. Finally, the data integrity has been identify by joining the tables that Bellfield and Melbourne are not recorded in the Suburb Rental and Suburb House Price table.

However, joining the table just a step to identifying the data integrity issue only

Data Integrity issues

After joining table, a data integrity issue is identify as from joining Vic-Suburbs Region table with Suburb Rental table together. As a results, the mismatch values was Bellfield and Melbourne that does not have a record in a Suburb Rental Table. Also when joined with Suburb House Price table, the null value are showed that Belfield and Melbourne have a record of house price and rental price as NULL.

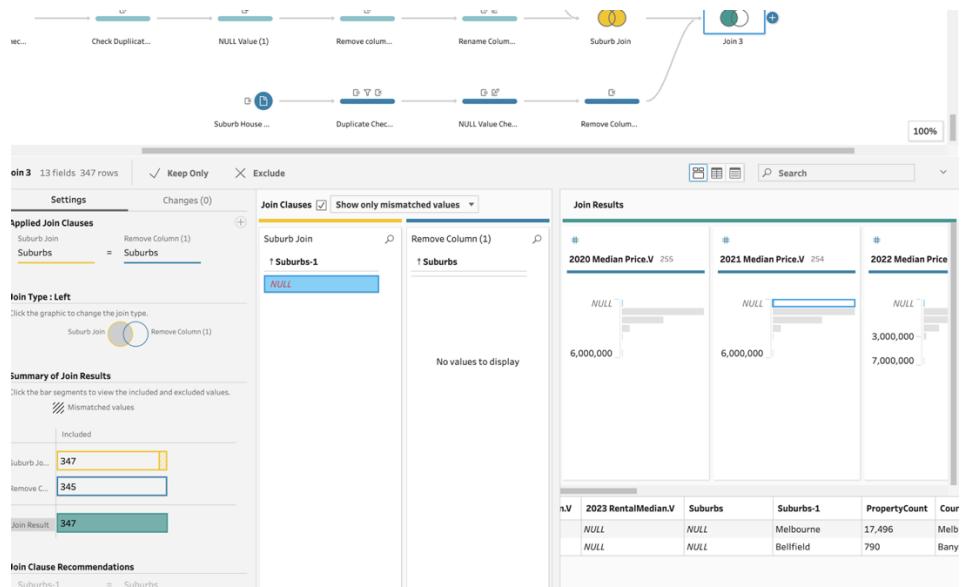
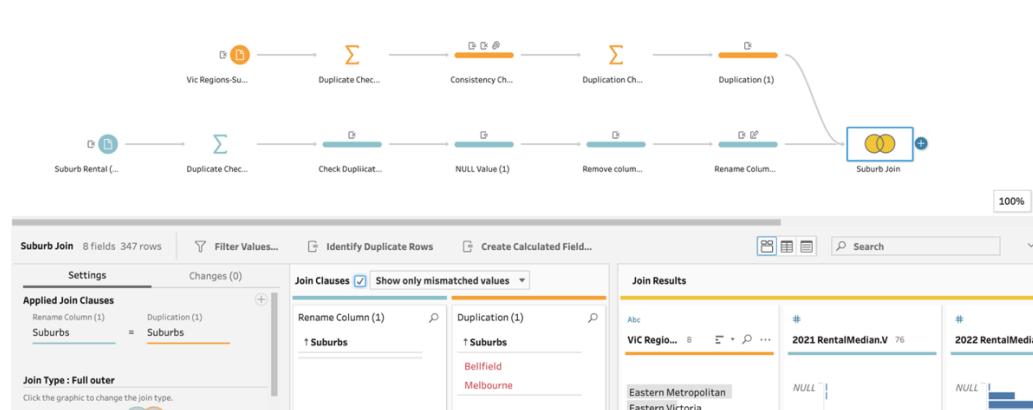


Fig- A NULL values of Bellfield and Melbourne in joined tables



To address this problems, the 2 missing suburb has been filled with the median price.

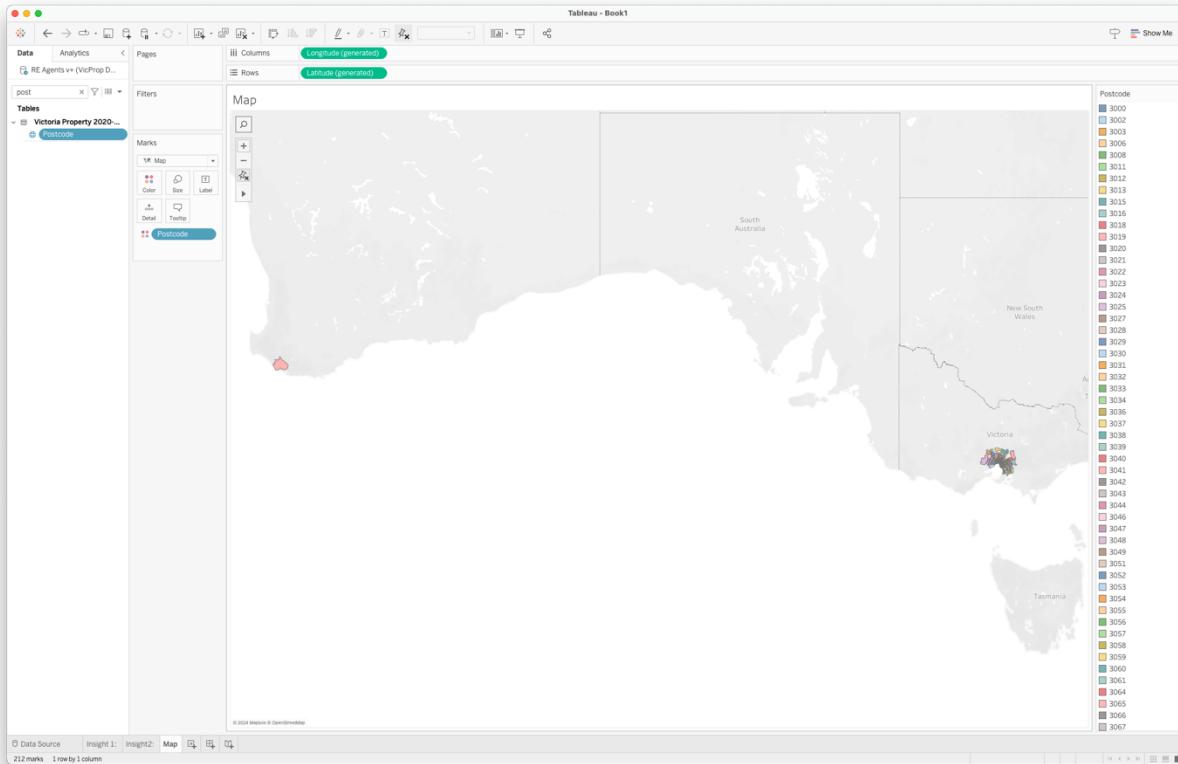
	Suburbs	2020 Median Price.V	2021 Median Price.V	2022 Median Price.V	2023 Median Price.V
343	Bentleigh	1510000	1510000	1748500	1885000
344	Croydon	785000	785000	920000	845500
345	Elsternwick	2081000	2081000	2175000	2325000
346	Darley	500000	500000	610000	640000
347	Bellfield	892500	912500	1051250	975000
348	Melbourne	892500	912500	1051250	975000
349					

	Suburbs	2021 RentalMedian.V	2022 RentalMedian.V	2023 RentalMedian.V
343	Malton West	340	350	360
344	Burnside	430	440	460
345	Cranbourne East	410	445	490
346	Flemington	380	430	500
347	Bellfield	435	460	520
348	Melbourne	435	460	520
349				

RE Agents v+ (VicProp Dataset Cleaned)



Fig- Data Integrity (Table Joining in Tableau)



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Suburb	Listing ID	Address	PropType	Status	Date of Record	Postcode	Price.v	CBD Distance	Bedroom	Bathroom.v	Car-Garage.v	Landsize	Building Area	Built Year	Latitude	Longitude
10639	Nunawading	REIA16740	1/7 Lilian St	townhouse	SP	17/6/2022	6262	960000	30.8	6	6.4		810	542.4107143	4000	-75.65356	290.33554
34451																	

This record in Vic Property Table has been wrongly assigned to 6262 post code which was a suburb in Western Australia not Victoria, so the issue was resolved by assigning the correct post code to Nunawading 3131.