Vancouver Property Tax Report 2023

Talk With Data: Tableau, IBM SPSS Statistics And PowerBI

Submitted To:
Prof Fatemeh Abkenari (Parissa), Faculty Lecturer

Submitted By:

Abhilash Dikshit

Shamim Sherafati

Business Questions

1. Understanding Property Distribution:

- What is the distribution of Property Identification Numbers (PIDs) based on legal types?
- How does this distribution impact the overall property tax landscape?

2. Analysing Land Value Trends:

- How has the average current and previous land value varied across different legal types over the years?
- Are there any notable trends or disparities in land value appreciation?

3. Exploring Tax Dynamics:

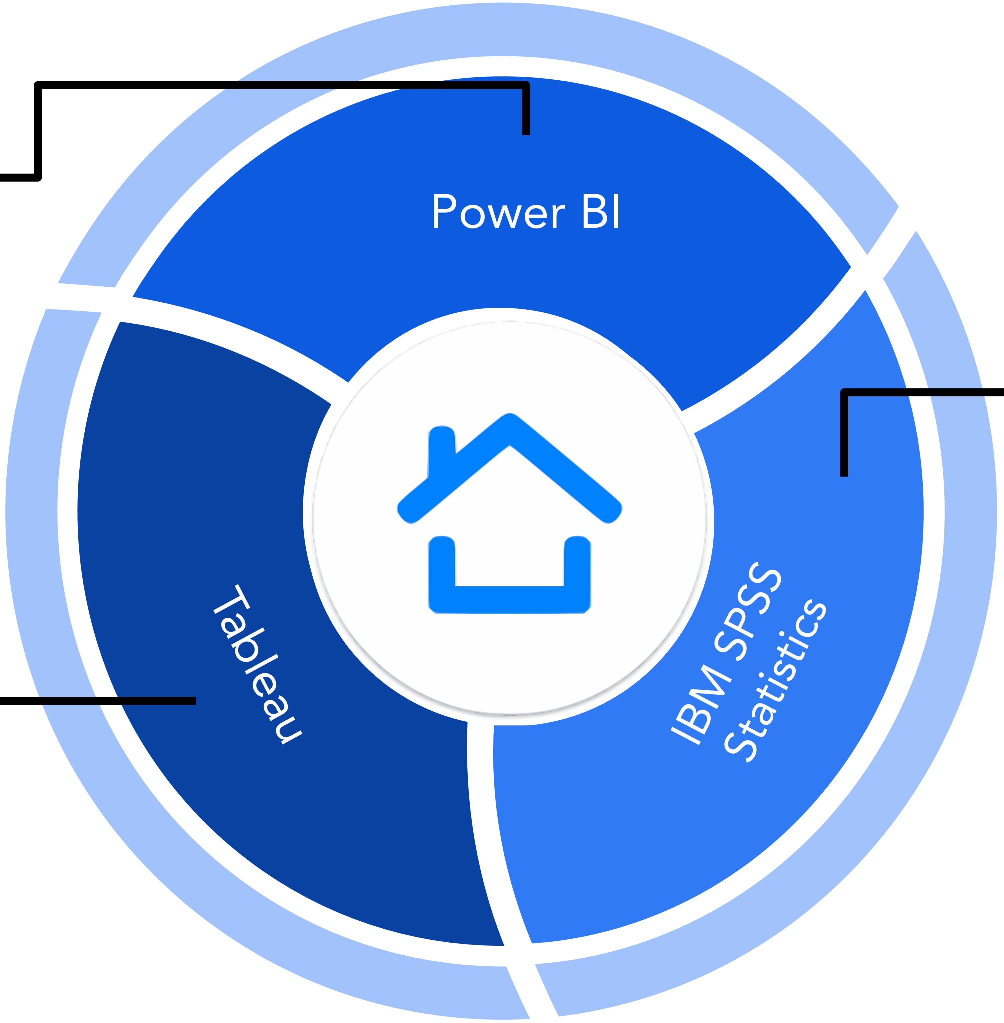
- What insights can be derived from the average tax levy in different legal types and zoning classifications?
- How do tax patterns differ among various property categories, and what factors contribute to these differences?

Implementation Roadmap

Data Visualization and Analysis

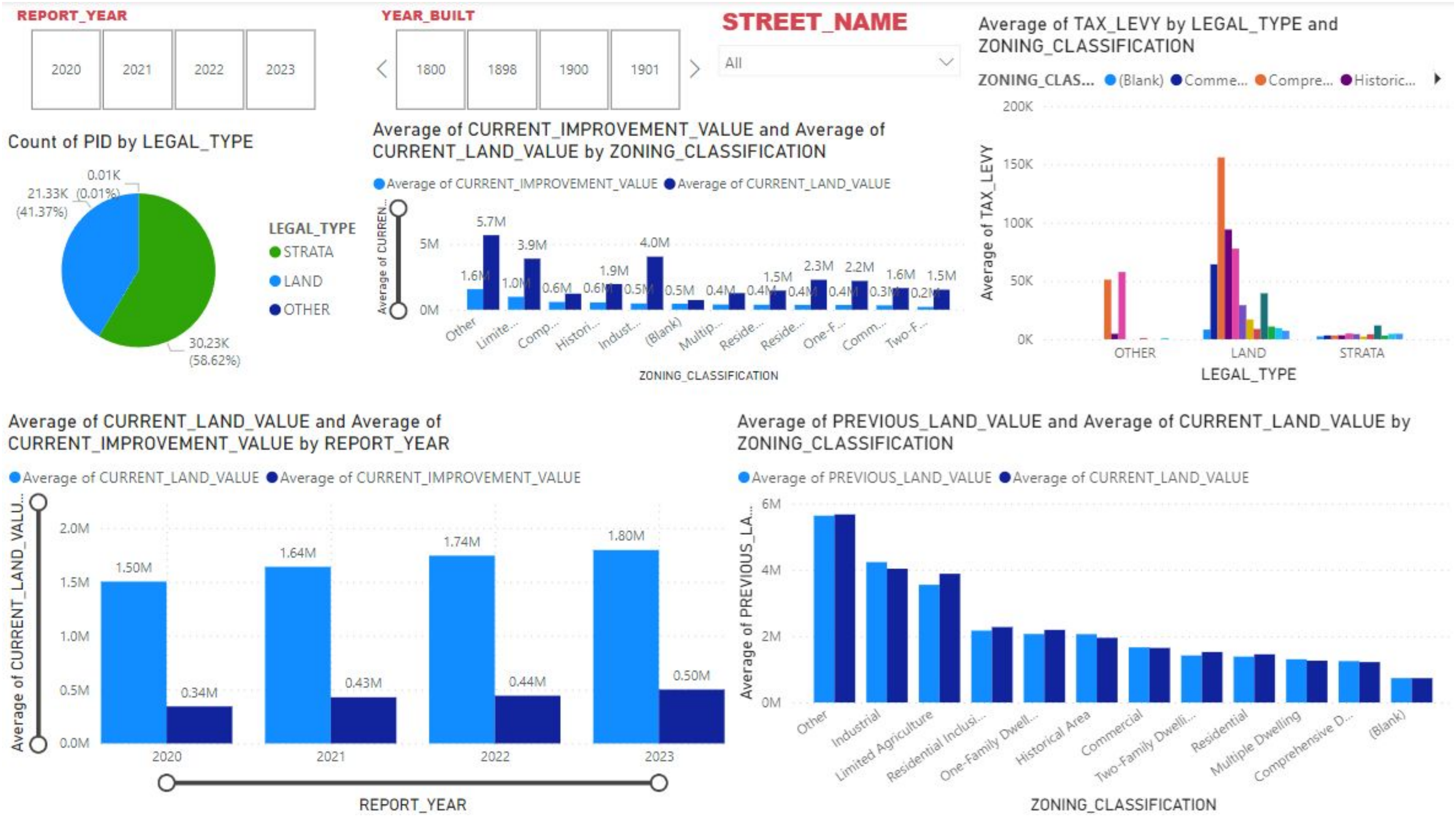
Statistical Analysis

Data Storytelling



Power BI

Data Analysis For Property Tax In Vancouver From 2020-2023



This report about property taxes gives us a detailed look at how real estate works in a specific area. It shows things like Property Identification Numbers (PIDs) to see how many properties are "STRATA" (like condos or apartments) and how many are just LAND.

It also looks at how the values of LAND and STRATA have stayed steady over the years.

All these charts and graphs help us understand property taxes better, and we can use this information to make smart decisions in planning and developing policies.

IBM SPSS Statistics

Anova And Coefficients

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	14379111830024946000.000	3	4793037276674981900.000	63743.295	<.001[b]
	Residual	633503157328157900000.000	842505	75192806847218.470		
	Total	777294275628407400000.000	842508			

a. Dependent Variable: CURRENT_LAND_VALUE
b. Predictors: (Constant), TAX_LEVY, REPORT_YEAR, YEAR_BUILT

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-77172245.320	17086722.747		-4.517	<.001
	YEAR_BUILT	-15000.537	318.946	-.046	-47.032	<.001
	REPORT_YEAR	53466.419	8456.504	.006	6.323	<.001
	TAX_LEVY	63.409	.147	.426	432.468	<.001

a. Dependent Variable: CURRENT_LAND_VALUE

The regression model shows a moderate relationship (R = 0.430) between the predictors (TAX_LEVY, REPORT_YEAR, YEAR_BUILT) and the dependent variable (CURRENT_LAND_VALUE).

This high F value suggests that there's a significant difference among the groups in the regression model. In simpler terms, it indicates that the variables in the model are contributing significantly to explaining the variance in the dependent variable. This strong statistical significance can provide confidence in the reliability of the regression model.

These coefficients provide a standardized measure of the variable's effect on the dependent variable.

In this case, TAX_LEVY has the largest standardized coefficient (Beta = 0.426), indicating that it has the strongest impact on the CURRENT_LAND_VALUE among the variables considered in this model.

Total Variance Explained: Principal Component Analysis

Total Variance Explained:

Component	Initial Eigenvalues: Total	Initial Eigenvalues: % of Variance	Initial Eigenvalues: Cumulative %	Extraction Sums of Squared Loadings: Total	Extraction Sums of Squared Loadings: % of Variance	Extraction Sums of Squared Loadings: Cumulative %
1	3.283	46.893	46.893	3.283	46.893	46.893
2	1.784	25.490	72.383	1.784	25.490	72.383
3	1.306	18.652	91.035	1.306	18.652	91.035
4	.355	5.070	96.105			
5	.239	3.419	99.524			
6	.026	.377	99.901			
7	.007	.099	100.000			

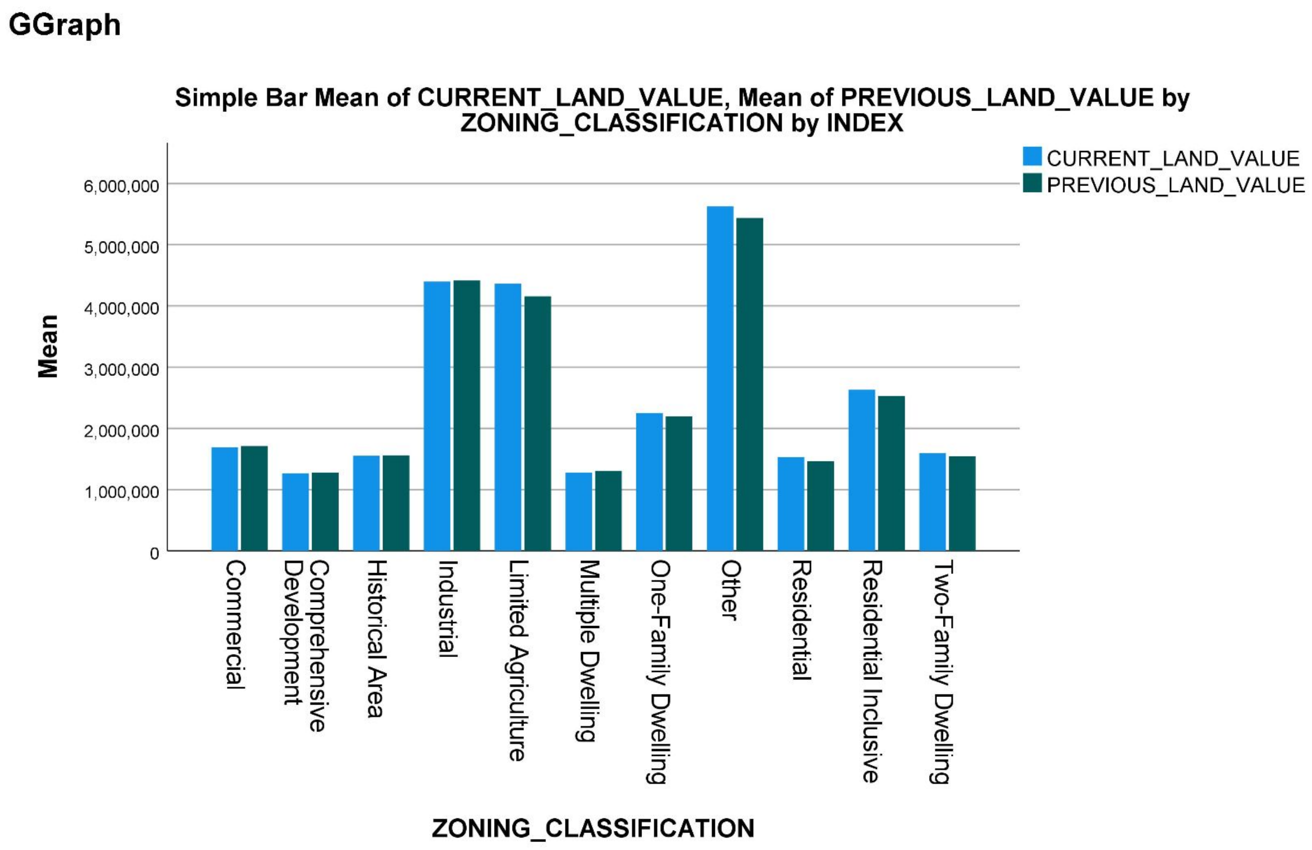
Extraction Method: Principal Component Analysis.

The Eigenvalues represent the amount of variance each principal component holds. The first component explains the highest amount of variance, followed by the second, and so on. The % of Variance (Extraction) represents the proportion of total variance explained by each component. For example:

Component 1 explains 46.893% of the total variance.
The first three components cumulatively explain 91.035% of the total variance.

Components 4, 5, 6, and 7 have eigenvalues below 1, suggesting that they explain less variance than would be expected from a single variable and might not be considered substantial components.

GGraph

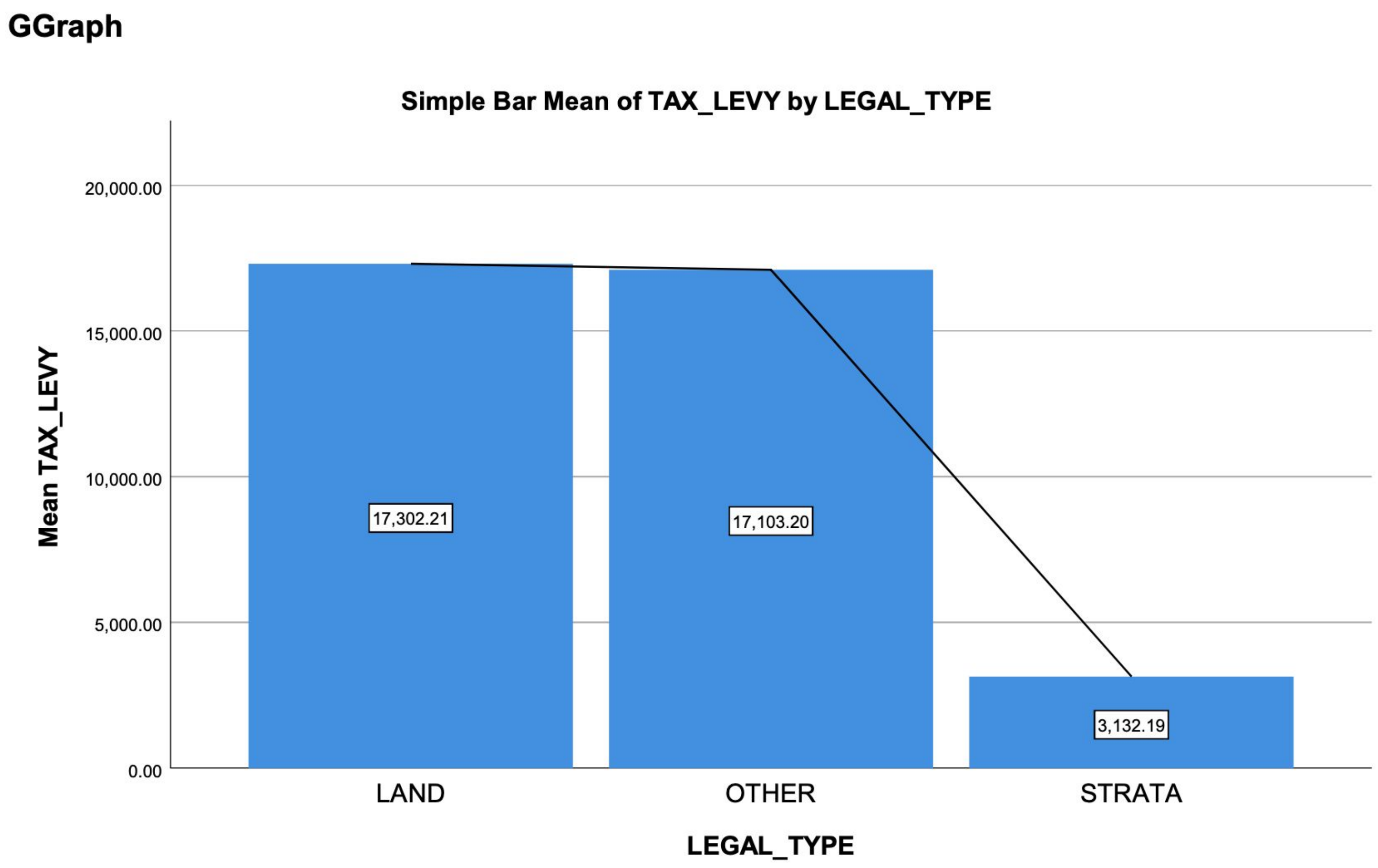


Examining the trends in property taxes reveals interesting changes in the average present values of land from 2020 to 2023.

The bar chart indicates a consistent stability in land and strata values, in contrast to notable fluctuations observed in other property categories.

These fluctuations, rising from 22.19 million in 2020 to a peak of 29.59 million in 2023, warrant a more thorough investigation into the underlying trends and their effects on the urban environment.

The Clustered Column Chart



The clustered column chart offers a glimpse into the distribution of tax obligations among different legal classifications and zoning categories within the land sector.

Remarkably, comprehensive development shows the highest tax responsibility at 149k, trailed by historical and industrial zones at 83k and 82k, respectively.

In contrast, residential properties consistently carry lower tax burdens, prompting inquiries into the factors influencing their contributions.

KMeans in SPSS

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
FOLIO	873484	2.E+10	8.E+11	4.99E+11	2.496E+11
LAND_COORDINATE	873484	1963206	84531342	49897988.80	24958687.1
FROM_CIVIC_NUMBER	427023	0	6705	866.30	875.560
TO_CIVIC_NUMBER	871240	1	31888	2389.22	1994.596
CURRENT_LAND_VALUE	860914	0	3568531000	1749660.86	10057195.3
CURRENT_IMPROVEMENT_VALUE	860914	0	876401000	451709.80	4766582.444
TAX_ASSESSMENT_YEAR	860914	2020	2023	2021.51	1.118
PREVIOUS_LAND_VALUE	850976	0	3488433000	1736914.94	10018672.3
PREVIOUS_IMPROVEMENT_VALUE	850976	0	652775000	424229.23	4279501.536
YEAR_BUILT	847604	1800	2022	1984.36	29.752
BIG_IMPROVEMENT_YEAR	847604	200	2022	1991.84	19.664
TAX_LEVY	861605	.00	9760300.00	8964.5822	64805.61572
NEIGHBOURHOOD_CODE	873484	1	30	16.55	8.943
REPORT_YEAR	873484	2020	2023	2021.51	1.118
Valid N (listwise)	408964				

Iteration History[a]

	Change in Cluster Centers				
Iteration	1	2	3	4	5
1	2.124E+10	4.795E+10	6.586E+10	5.855E+10	2.249E+10
2	89527432.5	1054130033	4105377168	1.096E+10	5140786280
3	2705046934	74276335.2	326012571	1.253E+10	3561126.772
4	4416061941	.000	.000	8217883326	.000
5	3731941572	.000	.000	4851437123	3561126.772
6	5234178953	.000	.000	4544508626	3457819.269
7	349452198	.000	.000	332836349	13412611.5
8	.000	.000	.000	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 8. The minimum distance between initial centers is 147957741241.038.

As the iterations progress, there is a trend towards decreased changes in cluster centers, with some clusters stabilizing and displaying minimal alterations. Ultimately, convergence is achieved as Z-scores stabilize at 0.000 for all clusters, signifying the establishment of relatively stable cluster compositions. This iterative process underscores the evolution and eventual stabilization of distinct clusters, enabling the discernment of unique groupings within the dataset based on inherent similarities and patterns.

K-Means in SPSS

Initial Cluster Centers

	Cluster 1	2	3	4	5
FOLIO	2.E+11	6.E+11	8.E+11	2.E+10	3.E+11
LAND_COORDINATE	17259506	58027529	84212495	2463732	32071895
FROM_CIVIC_NUMBER	3505	1	101	3	3606
TO_CIVIC_NUMBER	688	2727	1520	1980	5665
CURRENT_IMPROVEMENT_VALUE	273000	93000	140000	171000	199000
TAX_ASSESSMENT_YEAR	2021	2023	2022	2022	2021
PREVIOUS_LAND_VALUE	739000	1459000	942000	3412000	306000
PREVIOUS_IMPROVEMENT_VALUE	270000	92200	144000	162000	195000
YEAR_BUILT	2007	1913	1987	1978	2016
BIG_IMPROVEMENT_YEAR	2007	1975	1987	1978	2016
TAX_LEVY	2922.50	6693.41	11324.60	11729.90	1572.78
CURRENT_LAND_VALUE	727000	1606000	1077000	3607000	338000
NEIGHBOURHOOD_CODE	26	20	18	1	23
REPORT_YEAR	2021	2023	2022	2022	2021

The initial cluster centers for the five clusters in the dataset display distinct property characteristics across various parameters. Cluster 1 is delineated by FOLIO (172595060452), LAND_COORDINATE (17259506), and encompasses properties with FROM_CIVIC_NUMBER (3505) to TO_CIVIC_NUMBER (688).

Cluster 2 showcases different property attributes, featuring higher values in various aspects such as LAND_COORDINATE and FROM_CIVIC_NUMBER, accompanied by varying tax assessments and property values.

Clusters 3, 4, and 5 exhibit distinct profiles in terms of property characteristics, tax levies, land values, and neighborhood codes, highlighting diverse property traits within the dataset across different clusters and their respective features in the context of assessment years and neighborhood codes.

Final Cluster Centers

	Cluster 1	2	3	4	5
FOLIO	2.E+11	6.E+11	8.E+11	1.E+11	3.E+11
LAND_COORDINATE	16787907	62710156	77183255	12461173	29310298
FROM_CIVIC_NUMBER	1055	739	569	1207	680
TO_CIVIC_NUMBER	1815	1293	2982	1689	4433
CURRENT_IMPROVEMENT_VALUE	248653	225896	215230	242435	182428
TAX_ASSESSMENT_YEAR	2022	2022	2022	2022	2022
PREVIOUS_LAND_VALUE	613324	697615	496153	625964	390772
PREVIOUS_IMPROVEMENT_VALUE	247852	226423	211900	242538	181309
YEAR_BUILT	2002	1996	2002	2000	2004
BIG_IMPROVEMENT_YEAR	2003	1997	2002	2001	2004
TAX_LEVY	2566.78	2934.09	2191.88	2733.80	1747.98
CURRENT_LAND_VALUE	619315	703252	513888	627172	409741
NEIGHBOURHOOD_CODE	22	16	17	24	20
REPORT_YEAR	2022	2022	2022	2022	2022

Cluster 1, identified by FOLIO and LAND_COORDINATE, comprises properties with a range from civic numbers 1055 to 1815. These properties, predominantly built around 2002, have shown significant improvements by 2003.

Cluster 2 exhibits distinct property attributes with higher values in various aspects like LAND_COORDINATE and civic numbers, indicating diverse property characteristics.

Clusters 3, 4, and 5 also display unique property profiles, showcasing varying property values, tax assessments, and neighborhood codes. These final clusters help identify and differentiate property traits within the dataset, offering insights into diverse property types and their respective attributes across different clusters.

KMeans in SPSS

ANOVA

	Cluster Mean Square	Cluster df	Error Mean Square	Error df	F	Sig.
FOLIO	6.181E+27	4	7.588E+20	408959	8145482.197	<.001
LAND_COORDINATE	6.181E+19	4	7.588E+12	408959	8145482.195	<.001
FROM_CIVIC_NUMBER	4667901246	4	707414.064	408959	6598.542	<.001
TO_CIVIC_NUMBER	6.128E+10	4	2334745.486	408959	26246.378	<.001
CURRENT_IMPROVEMENT_VALUE	2.333E+13	4	7.927E+10	408959	294.326	<.001
TAX_ASSESSMENT_YEAR	64.659	4	1.254	408959	51.560	<.001
PREVIOUS_LAND_VALUE	6.992E+14	4	6.498E+11	408959	1075.946	<.001
PREVIOUS_IMPROVEMENT_VALUE	2.430E+13	4	1.012E+11	408959	240.029	<.001
YEAR_BUILT	909928.194	4	213.091	408959	4270.145	<.001
BIG_IMPROVEMENT_YEAR	928955.540	4	148.312	408959	6263.536	<.001
TAX_LEVY	1.021E+10	4	50969006.5	408959	200.337	<.001
CURRENT_LAND_VALUE	6.359E+14	4	6.085E+11	408959	1045.030	<.001
NEIGHBOURHOOD_CODE	1035520.183	4	79.282	408959	13061.288	<.001
REPORT_YEAR	64.659	4	1.254	408959	51.560	<.001

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Number of Cases in each Cluster

Cluster	1	75272.000
	2	200394.000
	3	42842.000
	4	71274.000
	5	19182.000
Valid		408964.000
Missing		464520.000

These differences indicate varied property characteristics within different clusters. However, it's important to note that these results are for descriptive purposes only, as the clusters were intentionally selected to maximize differences among cases. Therefore, the observed significance levels do not confirm the equality of cluster means and should be interpreted cautiously.

Multiplayer Perception

Multilayer Perceptron

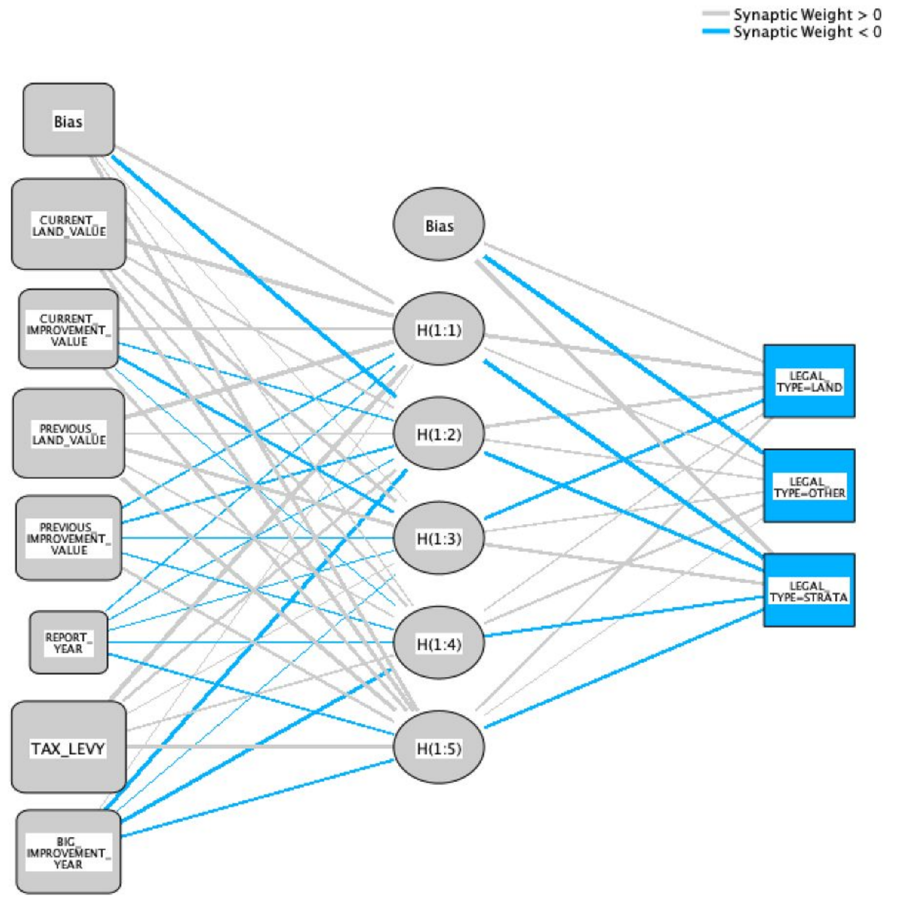
Case Processing Summary

		N	Percent
Sample	Training	671924	80.0%
	Testing	167518	20.0%
Valid		839442	100.0%
Excluded		34042	
Total		873484	

Network Information

Input Layer	Covariates	1	CURRENT_LAND_VALUE
		2	CURRENT_IMPROVEMENT_VALUE
		3	PREVIOUS_LAND_VALUE
		4	PREVIOUS_IMPROVEMENT_VALUE
		5	REPORT_YEAR
		6	TAX_LEVY
		7	BIG_IMPROVEMENT_YEAR
	Number of Units[a]		7
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		5
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	LEGAL_TYPE
	Number of Units		3
	Activation Function		Softmax
	Error Function		Cross-entropy

a. Excluding the bias unit



Hidden layer activation function: Hyperbolic tangent

Output layer activation function: Softmax

The Case Processing Summary indicates that 80% of the dataset was allocated for training purposes, consisting of 671,924 cases, while the remaining 20% (167,518 cases) formed the testing subset. A total of 839,442 cases were considered valid, with 34,042 cases excluded from analysis, resulting in a dataset totaling 873,484 cases.

These covariates underwent standardization. The model design excludes the bias unit in its computations.

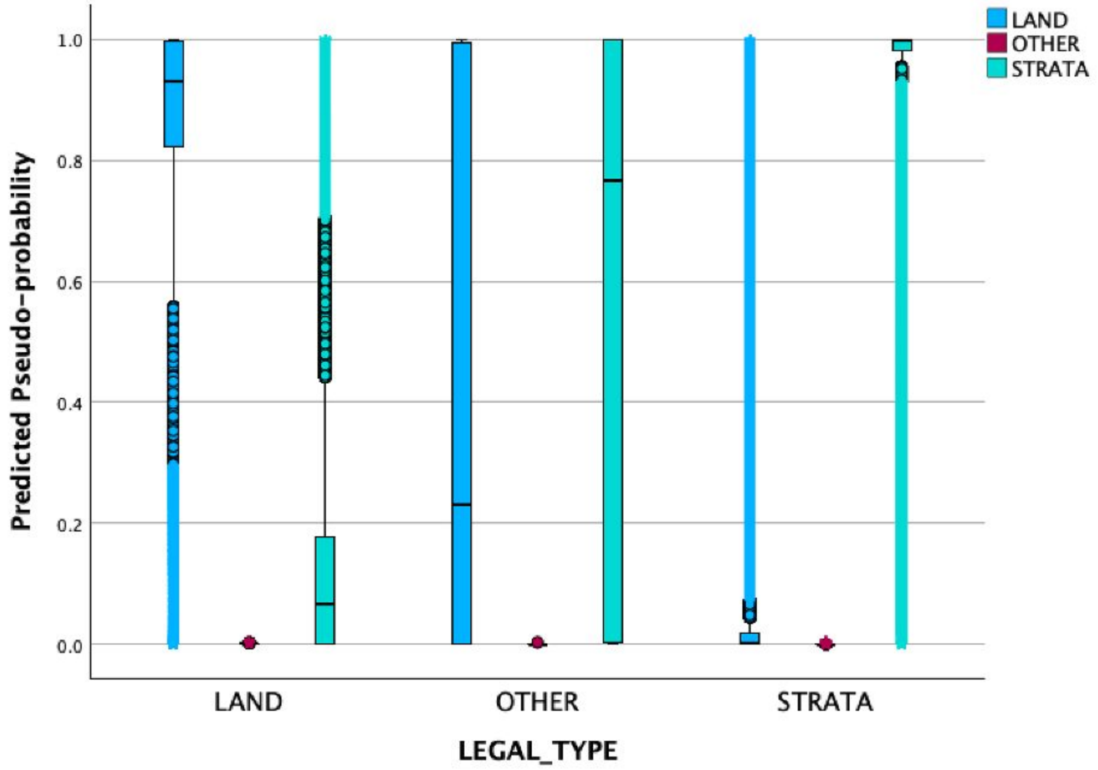
Multiplayer Perception

Classification

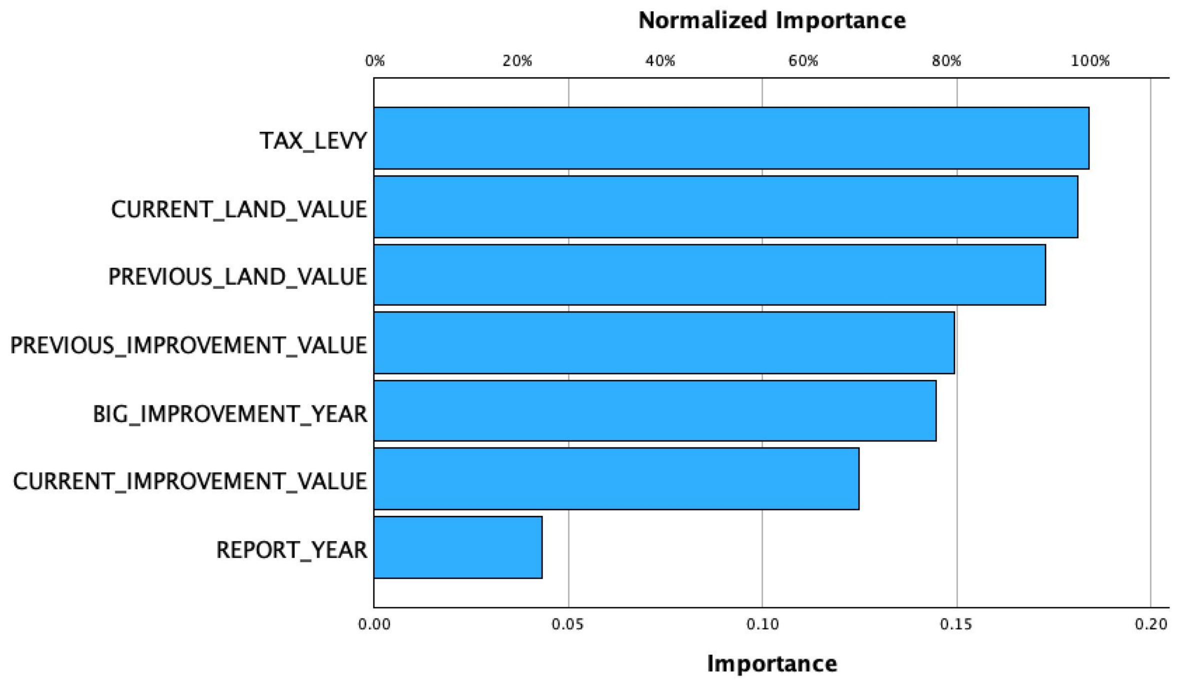
Sample	Observed	Predicted LAND	Predicted OTHER	Predicted STRATA	Percent Correct
Training	LAND	267585	0	10846	96.1%
	OTHER	150	0	159	0.0%
	STRATA	32830	0	360354	91.7%
	Overall Percent	44.7%	0.0%	55.3%	93.5%
Testing	LAND	66913	0	2650	96.2%
	OTHER	34	0	47	0.0%
	STRATA	8154	0	89720	91.7%
	Overall Percent	44.8%	0.0%	55.2%	93.5%

Dependent Variable: LEGAL_TYPE

The Classification section presents the observed and predicted outcomes for different categories—LAND, OTHER, and STRATA—across both training and testing datasets. In the training set, the model accurately predicted LAND in 96.1% of cases, STRATA in 91.7%, while it couldn't correctly predict any OTHER category instances. Similarly, in the testing set, the model achieved high accuracy in predicting LAND (96.2%) and STRATA (91.7%) but couldn't predict any instances of the OTHER category. The overall accuracy for the prediction of LEGAL_TYPE across all categories stands at 93.5%.



Normalized Importance



Independent Variable Importance

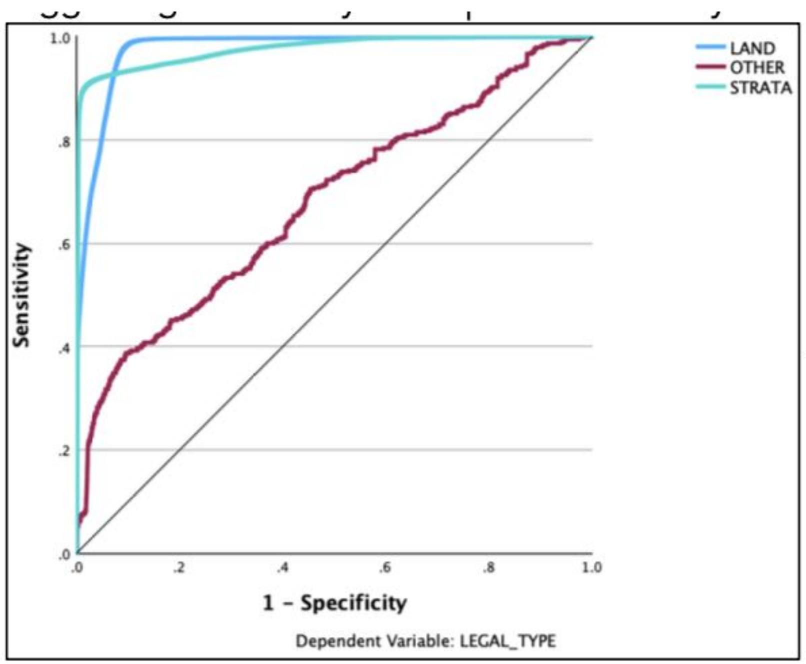
	Importance	Normalized Importance
CURRENT_LAND_VALUE	.181	98.4%
CURRENT_IMPROVEMENT_VALUE	.125	67.8%
PREVIOUS_LAND_VALUE	.173	93.8%
PREVIOUS_IMPROVEMENT_VALUE	.149	81.2%
REPORT_YEAR	.043	23.4%
TAX_LEVY	.184	100.0%
BIG_IMPROVEMENT_YEAR	.145	78.6%

Normalized Importance provides insights into the contribution of variables to the model.

Notably, BIG_IMPROVEMENT_YEAR holds the highest importance at 100%, followed by PREV_IMPROVEMENT_VALUE at 75.1%, and CURRENT_LAND_VALUE at 78.2%.

These findings underscore the significance of these variables in predicting legal types based on property characteristics.

Area Under the Curve (AUC)



Area Under the Curve

		Area
LEGAL_TYPE	LAND	.976
	OTHER	.679
	STRATA	.976

AUC values indicate strong predictive performance, particularly for LAND and STRATA with scores of 0.976 each. The OTHER category exhibits a slightly lower AUC of 0.679, suggesting a relatively lower predictive ability.

Tableau

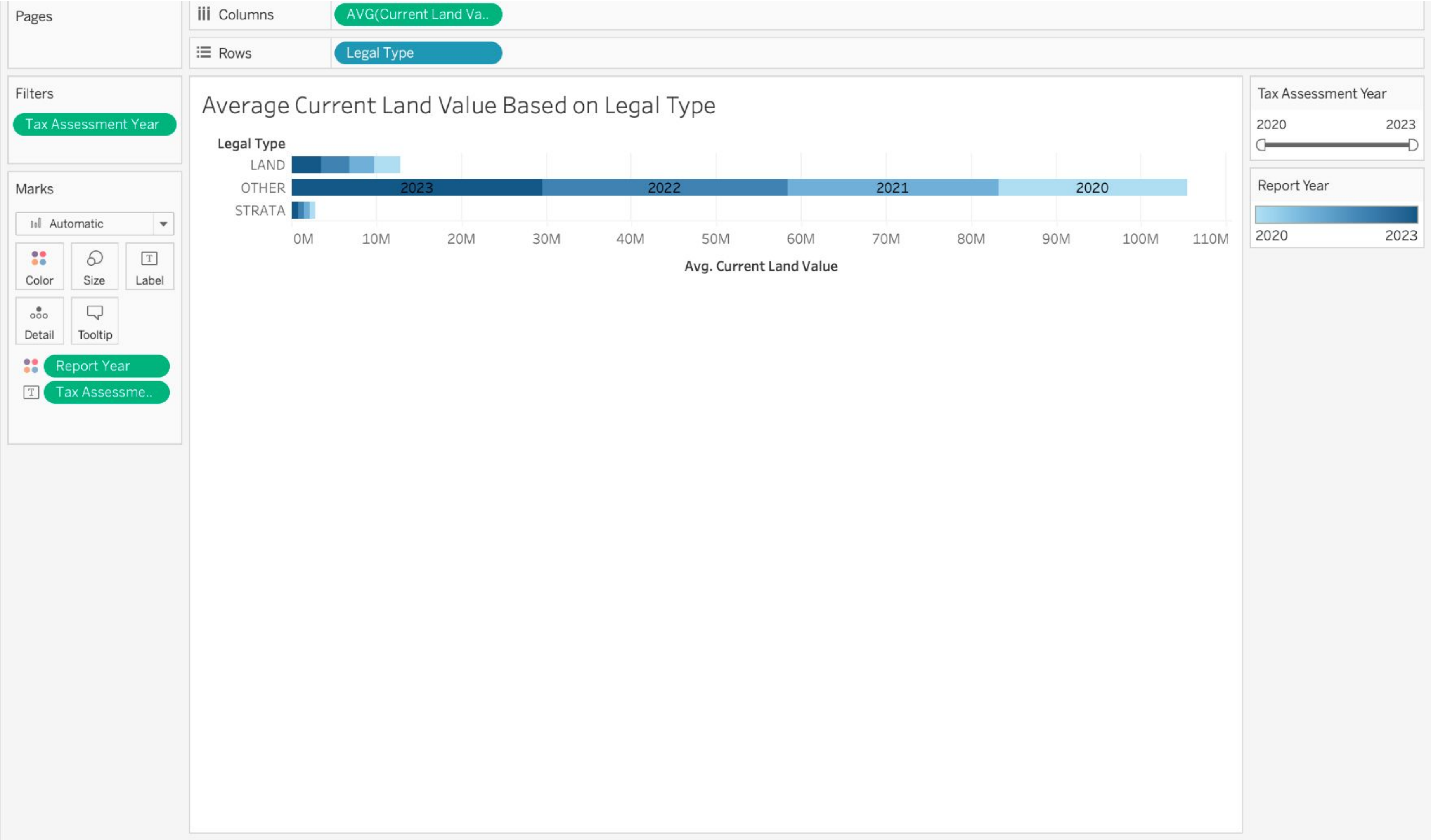
Average Previous and Current Land Value Based On Zone Classification

Looking at the stacked area chart that shows historical data, we see how the years when properties were built relate to how much the land was worth on average.

Some important things we noticed were a big drop from 1800 to 1890, a big increase around 1900, and then changes after 1900.

These changes make us want to investigate what things like money, society, and how things were built affected how much land was worth at different times.

Average Current Land Value Based On Legal Type

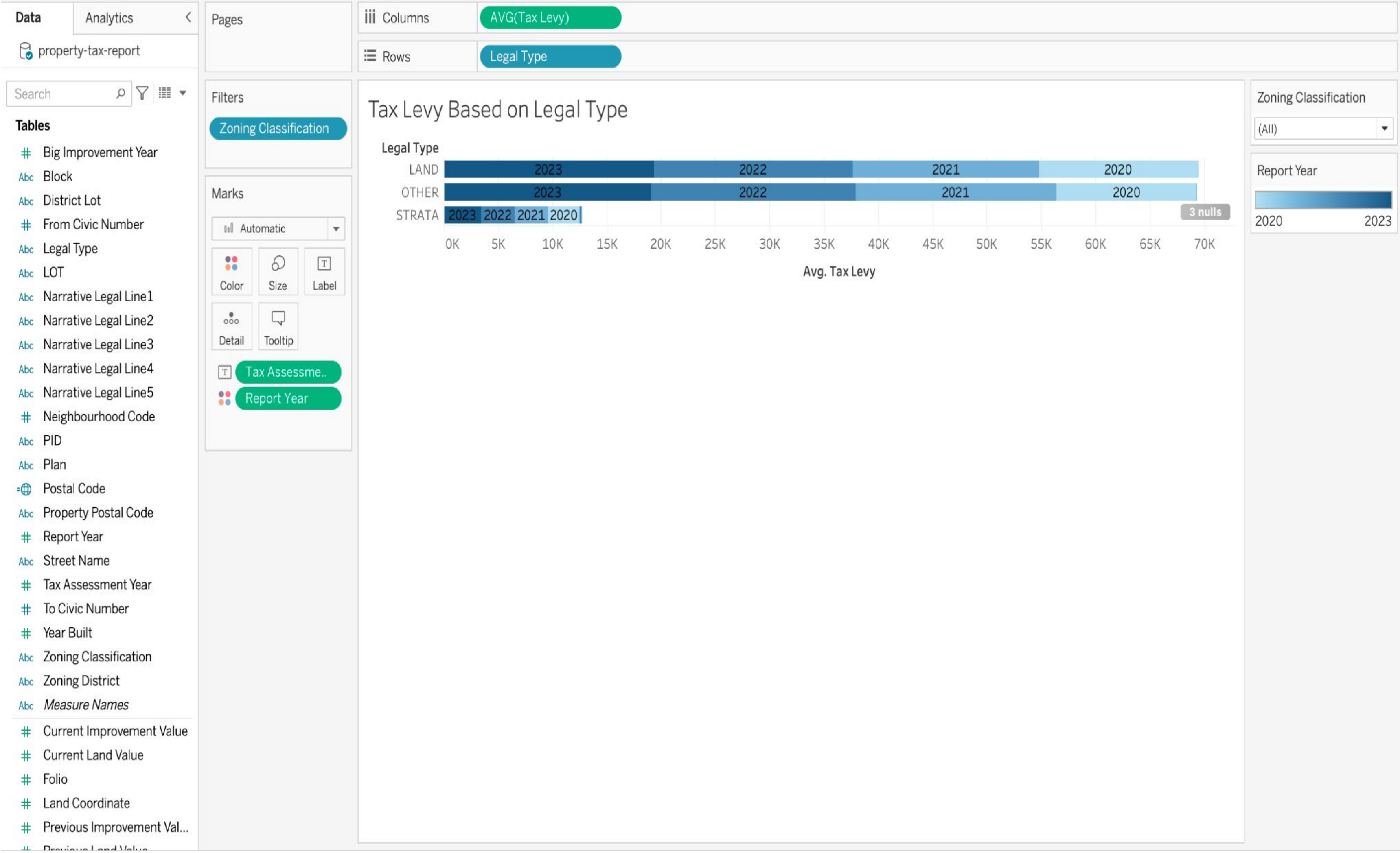


Looking at how property taxes changed over time, we noticed interesting changes in how much land was worth between 2020 and 2023.

The chart showed that land and strata values stayed pretty much the same, but there were big ups and downs in the value of 'other' types of properties.

These values went up a lot from 22.19M in 2020 to a high of 29.59M in 2023. This makes us want to find out more about why these changes happened and how they affected the city.

Tax Levy Based On Legal Type

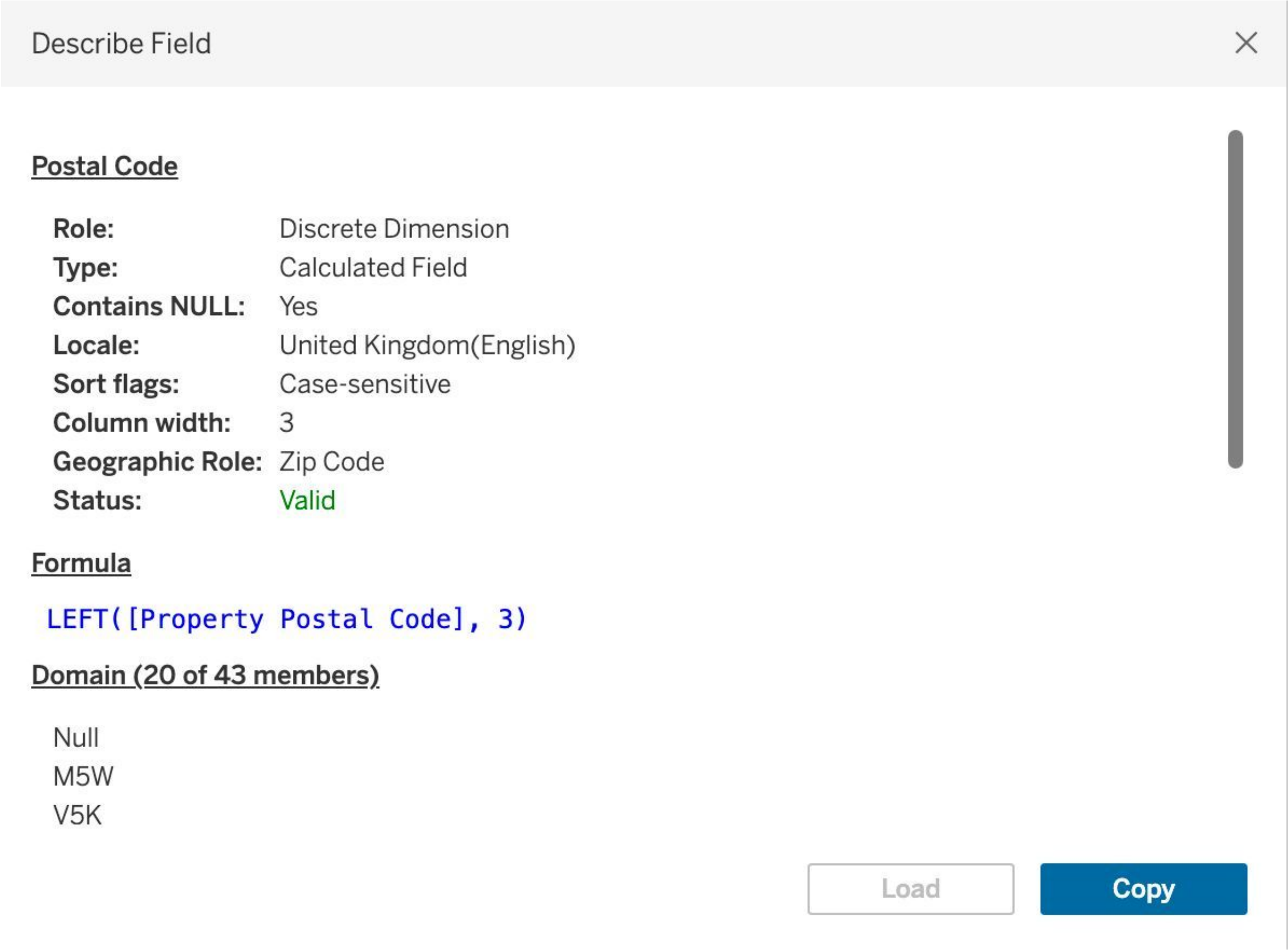


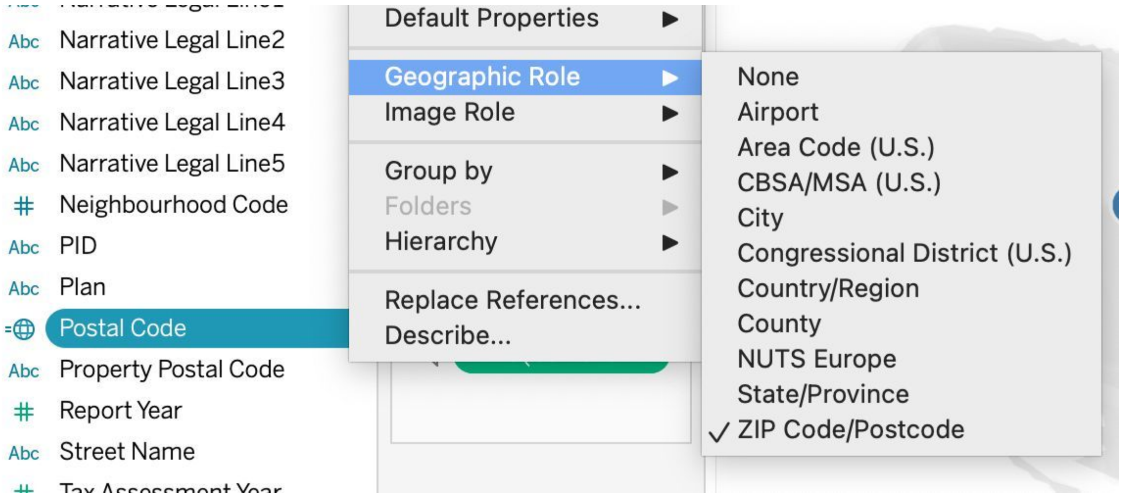
The chart shows how much tax different places and types of buildings pay.

We saw that places like comprehensive development pay the most taxes, about 149k. Historical and industrial zones pay around 83k and 82k.

But places where people live (residential areas) always pay less tax compared to the other places. This makes us curious about why homes pay lower taxes than other places.

DAX Implementation For Map View: Average Current Land Value Based On Vancouver Postal Code

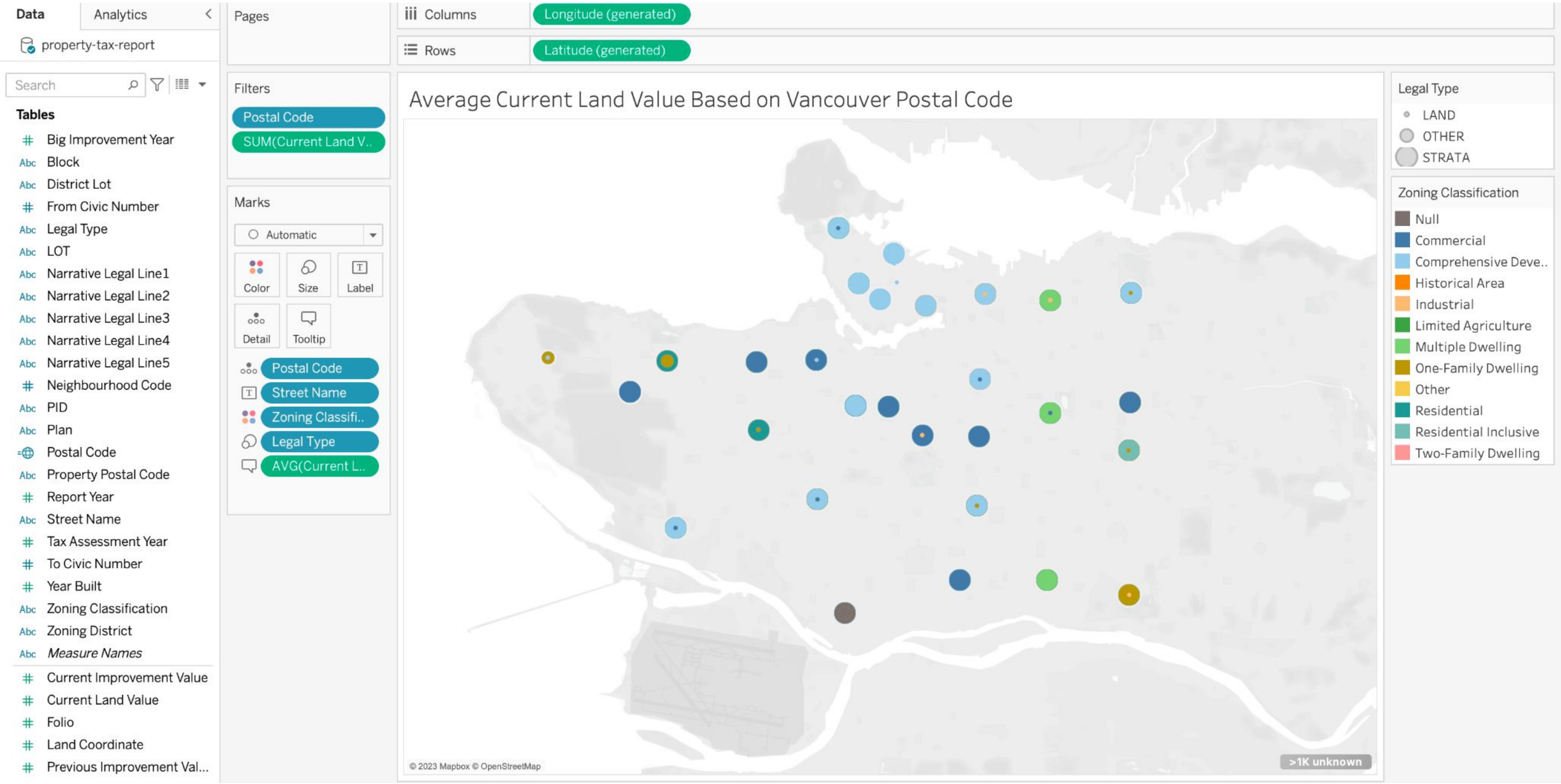




Step 1: Calculated Field “Postal Code” based on Property Postal Code column.

Step 2: Set the Geographic Role as ZIP Code/Postcode for Canadian Postal code

Map View: Average Current Land Value Based On Vancouver Postal Code

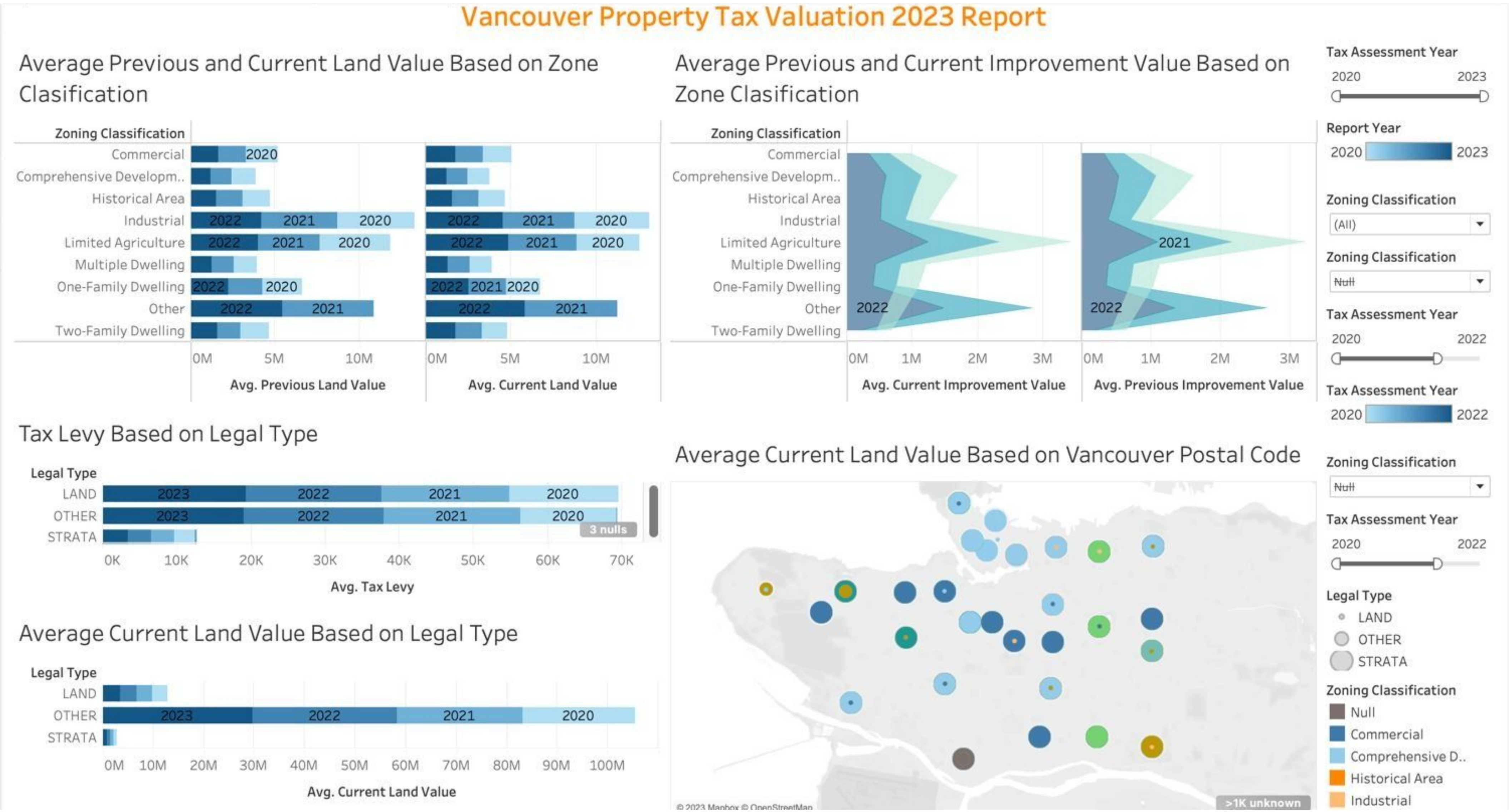


The map shows how different postal codes and zoning areas are spread across Vancouver.

We can see clusters of comprehensive development in downtown Vancouver, separate zones for businesses, and groups of zones in the south near the airport.

This shows how the city plans where things like buildings and businesses should be located, making the city diverse and well-planned.

Dashboard Analysis For Vancouver Property Tax Valuation From 2020-2023



Conclusion

Conclusion: Business Questions

1. Distribution of Property IDs by Legal Types: Property IDs are divided into categories like homes, businesses, industries, and others. Visuals show how many of each type there are in the dataset.

2. Impact of Distribution on Property Taxes: Different property types affect taxes differently. For instance, even if there are many homes, they might pay less tax compared to other types, affecting how much money the city gets from taxes.

3. Changes in Land Values Over Time for Different Types: We saw how property values changed from 2020 to 2023. Some types stayed steady, while others changed a lot. For example, 'other' types of properties jumped from 22.19M to 29.59M.

4. Notable Trends in Property Value Changes: Some types of properties had big ups and downs in their values. This needs a closer look to understand why these changes happened.

5. Insights from Tax Amounts in Different Property Types: Different property types pay different taxes. For example, businesses might pay more than homes. Understanding why this happens can tell us about how properties are used and where they are.

6. Differences in Tax Patterns Among Property Types: Taxes vary based on what the property is used for and where it's located. For instance, certain areas or types of properties pay more taxes due to rules about what can be done there.

References

References

Property tax report. (n.d.). https://opendata.vancouver.ca/explore/dataset/propertytax-report/table/?sort=-tax_assessment_year

Wikipedia contributors. (2023, November 26). SPSS. Wikipedia. https://en.wikipedia.org/wiki/SPSS

Minewiskan. (2023, January 20). Learn DAX basics in Power BI Desktop - Power BI. Microsoft Learn. https://learn.microsoft.com/en-us/power-bi/transform-model/desktop-quickstart-learn-dax-basics