Impact of Property Investment on Land Value in the city of Vancouver (2020-2024)

Kevin Tang, Mingyang Gao, and Zhipeng Cheng

Abstract—This report investigates the relationship between property investment and land value in the city of Vancouver from 2020 to 2024. The analysis utilizes datasets on property tax and building permits to explore trends and correlations in both residential and non-residential sectors. The findings reveal a generally weak and inconsistent relationship, influenced significantly by the economic disruptions caused by the COVID-19 pandemic.

However, a moderate to strong positive correlation is observed in 2024, particularly in the non-residential property sector, indicating signs of recovery and stabilization in the post-pandemic period. These insights aim to provide a better understanding of how investments influence land value and the factors contributing to these trends.

Index Terms—Building permits, land value, property investment, Vancouver.

I. Problem Definition

THE problem being addressed is understanding the impact of property investment on land value in Vancouver over the period of 2020-2024. This involves examining the correlation between investment activities, as reflected in building permits, and the changes in land value, as indicated by property tax data. The primary challenge lies in the inconsistent and often weak relationship between these variables, influenced by external factors such as economic fluctuations and government policies. The COVID-19 pandemic introduced significant volatility in the real estate market, complicating the analysis further by disrupting both residential and non-residential property trends. Identifying clear patterns amidst this uncertainty and understanding the factors driving any observed correlations present key challenges in this analysis.

II. Methodology

A. AWS S3 For Data Warehouse

AWS S3 was utilized as the primary data storage solution for this project. S3 provided a secure, scalable, and cost-effective way to store the two Parquet and CSV raw datasets sourced from Vancouver's open data portal. These datasets were stored in S3 to facilitate seamless data access and processing. S3's compatibility with AWS Glue and Redshift ensured smooth data movement and integration across different components of the data pipeline.

B. AWS Glue For Extract, Transform, Load

AWS Glue was used for the ETL (Extract, Transform, Load) process to extract data from various sources, transform it into a consistent format, and load it into a data warehouse. This tool was selected for its serverless nature, which minimizes infrastructure management and enables automated data processing. AWS Glue's compatibility with PySpark made it an ideal choice for handling large-scale data transformation and integration tasks efficiently.

C. AWS Redshift For Data Aggregation

AWS Redshift was employed for data aggregation and SQL-based processing. This data warehouse solution was chosen for its high performance and scalability, which are essential for running complex SQL queries on large datasets. Redshift's columnar storage format and parallel processing capabilities allowed for faster data filtering, joining, and aggregation, making it well-suited for analyzing trends in building permits and property tax data. In this phase, joins of the two data sources were performed on postal codes and a mapping from postal code to city local area was used to aggregate data according to geography.

D. AWS Sagemaker For Visualization

AWS SageMaker was used to read the processed data and generate interactive visualizations. This service was chosen for its integration with Jupyter notebooks, providing an interactive environment for data exploration and analysis. SageMaker's extensive support for Python libraries, such as Pandas, NumPy, Matplotlib, and Seaborn, allowed for custom data analysis and visualization tailored to the project’s needs. During visualization, local area information from the processed data was used to associate each entry with a geographic division from the Vancouver Open Data Portal’s local area boundary dataset to achieve the 3D plots shown on notebook.

E. Others (Dekart, Binder)

Dekart was used for constructing and displaying heatmaps,

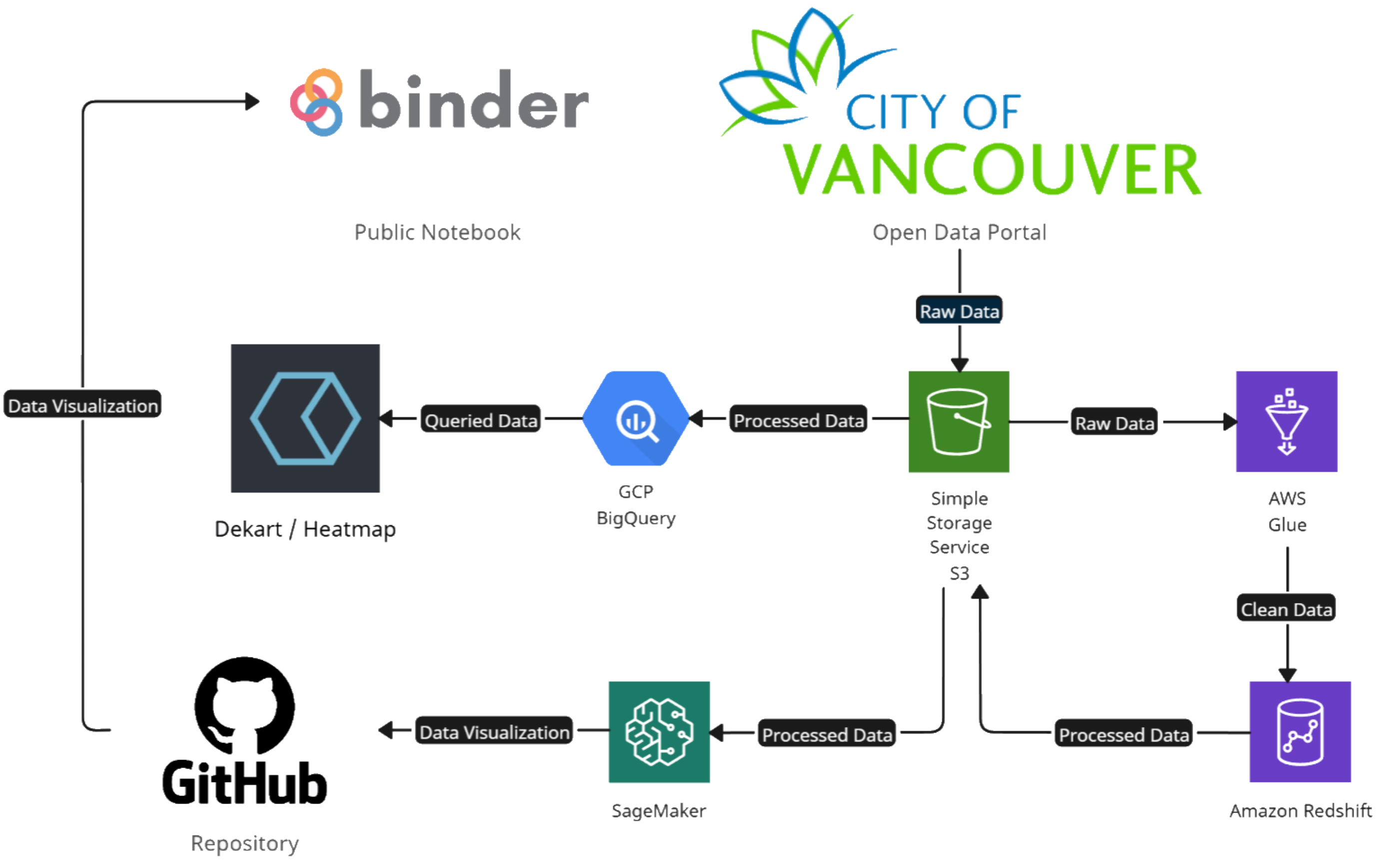

Fig. 1. Flow of data from source to visualization through the various services/technologies used in this project.

providing a visual representation of the spatial distribution of data trends. Binder was utilized to create an accessible, interactive environment for the public to explore the analysis and visualizations without the need for local environment setup.

III. Problems Encountered

A. *Geospatial Feature Incompatibility*

When attempting to read geospatial features from raw data in AWS Glue, a significant problem arose with the handling of Parquet files. Glue automatically recognized these files but insisted on reading them as binary data that cannot be read by any method supported inside the AWS Glue environment, rendering the geospatial features unusable. This issue meant that the data could not be parsed correctly for transfer to other services like Redshift or S3, stalling the ETL process. After extensive troubleshooting and testing different configurations, I decided to switch to reading in CSV files instead of Parquet. While this approach introduced inefficiencies in terms of file size and processing time, it provided the necessary clarity for parsing geospatial features effectively, allowing me to progress with the pipeline.

B. *Interservice Communication*

Moving cleaned data from Glue to Redshift presented major difficulties due to the complexity of AWS's inter-service communication. Setting up service roles, authenticated connections, and correct API configurations proved to be particularly challenging. These steps require precise understanding of permissions, resource policies, and Glue-specific nuances in how Redshift accepts data transfers. Initially, trial-and-error attempts only resulted in repeated failures. I overcame this hurdle by delving into AWS's documentation on write_dynamic_frame.from_jdbc_conf and carefully validating each parameter and configuration in my use case. This reexamination ultimately resolved the connection issues and allowed data transfer to proceed.

C. *GeoJSON Parsing Corruption*

A severe issue occurred during the unloading of processed data from Redshift to S3, where hidden rows of incorrectly parsed GeoJSON data, generated by ST_GeomFromGeoJSON operations, corrupted the entire table. These hidden errors were challenging to detect, and their presence meant that the unloading process failed repeatedly. To address this, I abandoned the GeoJSON-to-geometry conversion process entirely and instead opted for a simpler, more controlled approach. By extracting latitude and longitude directly from the original column through string parsing, I bypassed the problematic GeoJSON processing. This adjustment ensured table integrity and allowed the unloading operation to complete.

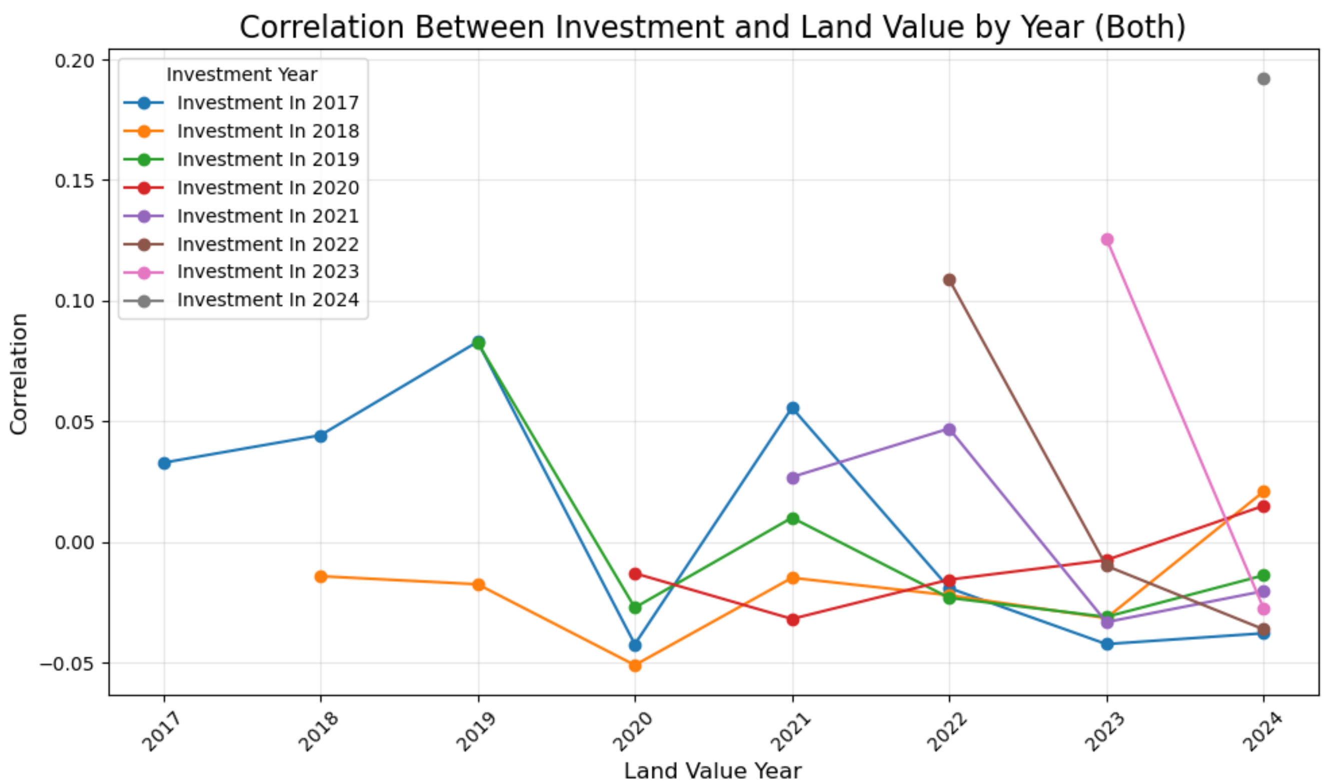

Fig. 2. Correlation between property investment made in specific year and land value in all subsequent years for both residential and non-residential property use.

D. *Confounding Factors in Data Source*

In my first attempt to correlate property investment with land value, I used adjacent data representing property investment through the number of employees in newly issued business licenses. However, this approach introduced numerous confounding factors, such as varying industry impacts and seasonal hiring trends, which muddied the results and produced irrelevant correlations. After discussing the problem and re-evaluating the available datasets, I was convinced to switch to one with more directly relevant metrics—the building permits dataset, which provided project values in CAD. This change drastically improved the correlation analysis, aligning the dataset more closely with the problem's requirements.

IV. Results

A. *Correlation between Property Investment and Land Value*

The analysis of the data reveals a variable relationship between property investment and land value across the years from 2017 to 2024. The correlation coefficients fluctuate significantly, indicating periods of weak, negative, or moderate positive relationships depending on the economic conditions and external factors.

For overall properties, the correlation between investment and land value remained mostly low or negative between 2017 and 2023, with the impact of economic growth becoming less directly impactful over time. Notably, the years 2020 to 2022 exhibited negative correlations, largely attributed to the disruptions caused by the COVID-19 pandemic. By 2024, a moderate positive correlation of 0.192159 emerged, suggesting signs of stabilization and recovery in market conditions as the post-pandemic phase allowed investments to exert a more consistent influence on land values.

B. *Crisis Effects*

The COVID-19 pandemic had a profound effect on the relationship between property investment and land value. The period between 2020 and 2022 was marked by significant economic uncertainty and volatility, with disrupted investment trends and market behaviors leading to generally weak or negative correlations. Residential properties experienced particularly weak correlations during this time, reflecting shifts in housing demand, suburban migration trends, and fluctuating interest rates. Non-residential properties also suffered, as the demand for commercial spaces such as offices and retail venues declined sharply. However, by 2023 and 2024, positive correlations began to reemerge, particularly in the non-residential sector, signaling a recovery as businesses adapted to new market dynamics.

C. *Residential vs. Non-Residential*

The data highlights distinct differences in how investment impacted residential and non-residential properties over the studied period. Residential properties showed consistently weak or negative correlations, with only occasional positive values, such as a mild improvement in 2019. The pandemic

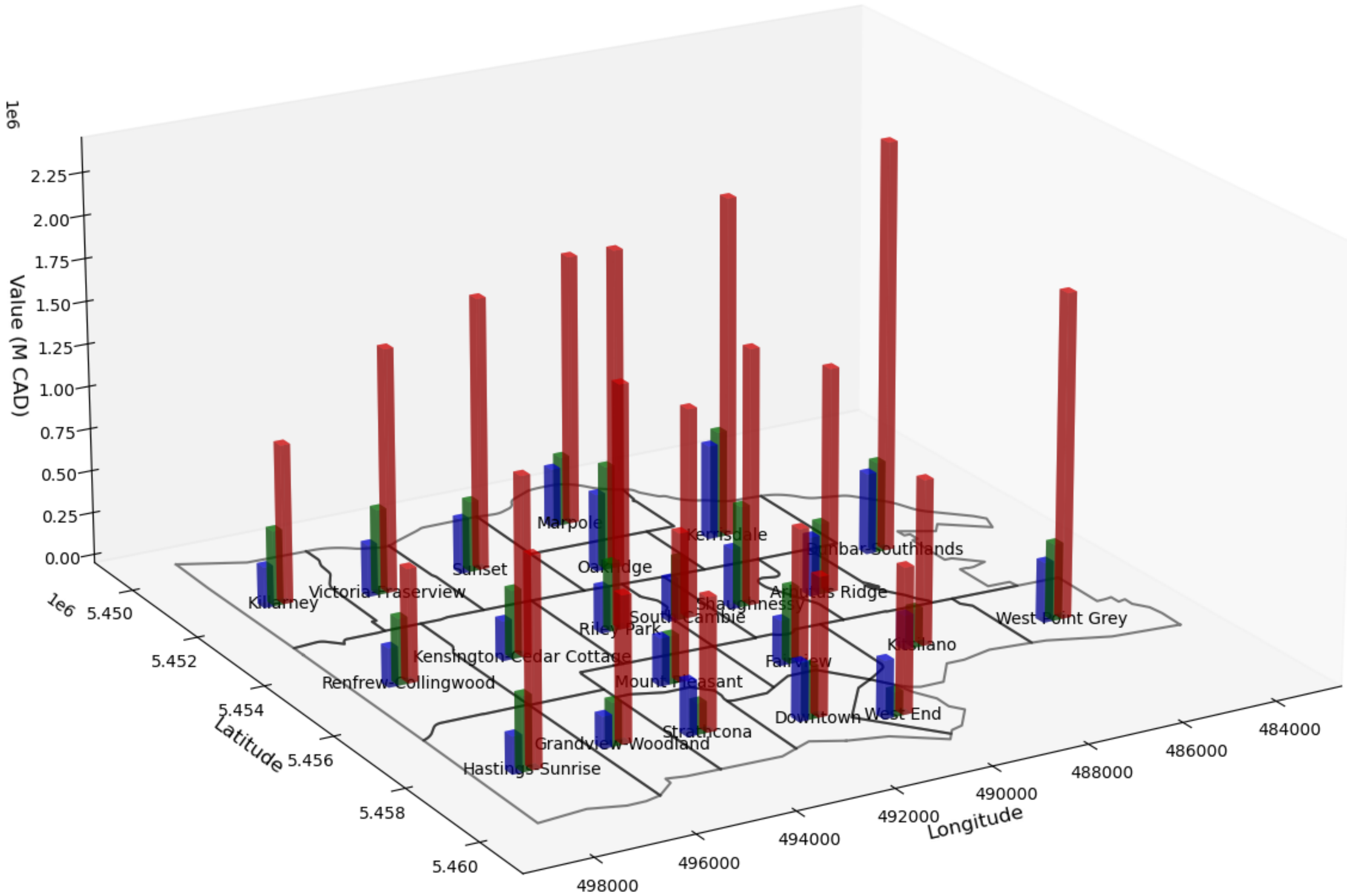

Fig. 3. Value of land, new property projects, and improvements by local area in the city of Vancouver for year 2022.

years amplified these inconsistencies due to changes in consumer behavior and economic conditions, resulting in weak relationships between investment and land value.

Non-residential properties demonstrated greater fluctuations, with periods of both strong positive and weak negative correlations. A particularly strong positive correlation was recorded in 2019, reflecting favorable economic conditions and significant returns in the sector. The pandemic significantly disrupted this trend, with negative correlations in 2020 and weak or near-zero correlations in the following two years. By 2024, however, non-residential properties exhibited a strong positive correlation of 0.193192, the highest since 2019, reflecting the sector's recovery and adaptation to post-pandemic realities.

D. Property Investment and Land Value by Local Area

The relationship between property investment and land value varies significantly by local area, influenced by neighborhood characteristics and investment patterns.

High-value neighborhoods such as Dunbar-Southlands, West Point Grey, and Kerrisdale consistently exhibit some of the highest land and improvement values, underscoring their attractiveness and strong demand. These areas also demonstrate significant growth in project values, indicating that property investments are actively driving development and land value increases. For instance, Kerrisdale saw land values surpassing $2 million per lot by 2023, reflecting the area's robust investment activity.

Emerging neighborhoods like Renfrew-Collingwood and Strathcona exhibit lower land values but show moderate growth in improvement and project values, suggesting that investment is contributing to gradual revitalization. These areas hold potential for future appreciation as development efforts continue. Similarly, Hastings-Sunrise and Victoria-Fraserview are experiencing rapid increases in land and project values, driven by gentrification trends and ongoing investment activities.

In contrast, areas like Oakridge and Riley Park show more stable trends with slight declines in property use, while areas such as Kerrisdale and Killarney have seen a noticeable reduction in residential property development. Downtown remains a standout for both residential and non-residential developments, maintaining high demand across both categories.

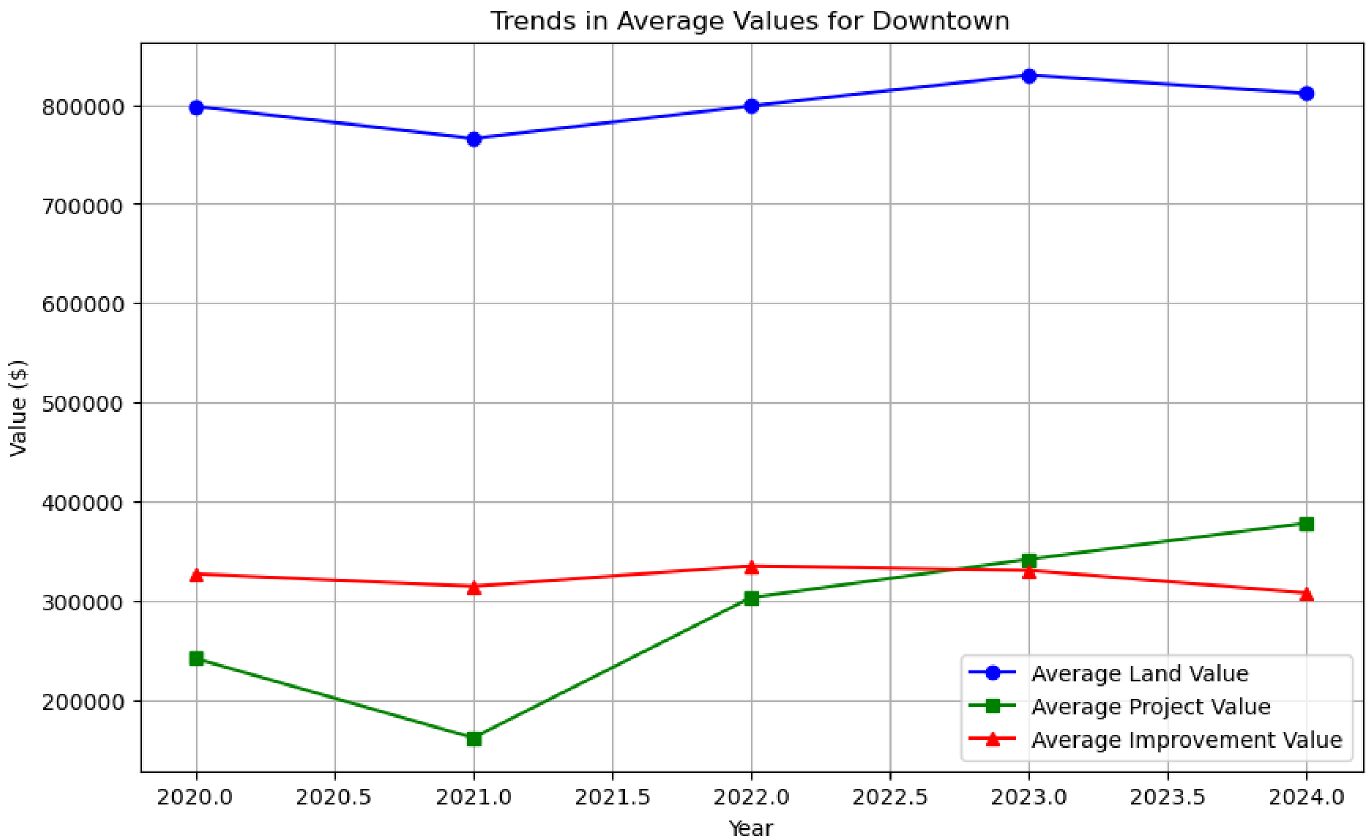

Fig. 4. Trend in average values of three categories for Downtown Vancouver from 2020 to 2024.

V. Lessons from Implementation

The process of implementation in this project revealed that geospatial data is highly specialized and requires distinct methods for effective handling. Traditional data processing tools may not fully support the complexity of geospatial formats and result in challenges in parsing and integration.

It was also observed that data cleanliness is crucial for producing reliable analysis. Ensuring data accuracy and consistency is essential for maintaining the integrity of the results, as errors or inconsistencies, such as hidden issues in data parsing, can disrupt processes and affect outcomes. This highlights the importance of thorough data validation and cleaning at each stage to ensure the reliability of the analysis.

Confounding factors were found to significantly impact the ability to draw meaningful conclusions. External influences, such as industry trends, seasonal patterns, or pandemics, can obscure relationships between data points and lead to misleading results. It highlights the need to carefully select relevant data sources and consider their context to ensure that the analysis accurately reflects the intended variables and relationships.

VI. Summary For Grading Guidance

Getting the data – 1
ETL – 2
Problem – 1
Algorithmic Work – 1
Bigness/parallelization – 2
UI – 4
Visualization – 5
Technologies – 4

References / Data Used

[1] "Property tax report," Vancouver.ca, 2020. https://opendata.vancouver.ca/explore/dataset/property-tax-report (accessed Dec. 06, 2024).

[2] "Issued building permits," Vancouver.ca, 2017. https://opendata.vancouver.ca/explore/dataset/issued-building-permits (accessed Dec. 06, 2024).

[3] "Local area boundary," Vancouver.ca, 2019. https://opendata.vancouver.ca/explore/dataset/local-area-boundary (accessed Dec. 06, 2024).