

COMP 693 Industry Project Weekly Journal:

Project Title: Coastal Proximity, Environmental Risks, and Property Valuation: A Comprehensive GLM Analysis in Christchurch

Week 1 - 3: 15 July - 4 August

Work Completed and Time Spent	Weeks 1-2: <ul style="list-style-type: none"> Discussed study direction and finalized project scope with mentor – 3 hours. Gathered property transaction data for Christchurch – 10 hours. Drafted and completed project proposal – 40 hours. Established weekly meeting schedule and planned timeline – 2 hours. Conducted literature review on relevant studies – 10 hours.
Plans/Goals for Next Week	Week 3: <ul style="list-style-type: none"> Started cleaning Christchurch dataset – 15 hours. Began setting up GIS software for spatial analysis – 5 hours. Conducted research on Christchurch's coastal areas – 10 hours.
Reflections	<ul style="list-style-type: none"> Complete data cleaning for Christchurch. Finalize GIS setup and begin spatial analysis. Apply preliminary Hedonic Pricing Model analysis to Christchurch dataset.
Project Health	The initial weeks have been productive, with clear project direction and good progress in data acquisition, proposal writing, and setup for analysis.
Issues Arising	Green
	none

Week 4: 5 August - 11 August

Work Completed and Time Spent	<ul style="list-style-type: none"> Discussion on Coastal Areas: Reviewed and selected key coastal areas in Christchurch (Scarborough, Brighton, Sumner, etc.) for focused analysis – 3 hours. Mentor Guidance: Received additional guidance on the application of the Hedonic Pricing Model (HPM) and Repeat-Sales Method (RSM) – 2 hours. Data Collection: Acquired latitude and longitude data for properties in the selected coastal areas – 5 hours.
-------------------------------	---

<ul style="list-style-type: none"> Plans/Goals for Next Week 	<ul style="list-style-type: none"> Data Preparation: Prepared the transaction data for import into GIS by integrating the geocoded coordinates – 4 hours. Import the geocoded property data into QGIS for spatial analysis. Begin performing spatial analysis on the selected coastal areas to understand the impact of coastal proximity on property values. Continue refining the application of HPM and RSM with mentor's input.
Reflections	This week's progress has been significant in narrowing down the focus areas for our analysis. The acquisition of latitude and longitude data has been crucial in enabling the next steps of spatial analysis. The additional guidance on the models has provided clearer direction for the analytical approach, ensuring we can derive meaningful insights from the data.
Project Health	Green
Issues Arising	none

Week 5: 12 August - 18 August

Work Completed and Time Spent	<ul style="list-style-type: none"> Data Import: Successfully imported the geocoded property data into QGIS, mapping properties across selected coastal areas – 5 hours. Spatial Analysis Setup: Configured the GIS environment for spatial analysis, including setting up base maps and defining key layers – 3 hours. Initial Analysis: Began preliminary spatial analysis by calculating distances from properties to the coastline and generating initial visualizations – 4 hours. Model Refinement: Worked on refining the Hedonic Pricing Model (HPM) parameters in consultation with the mentor to better capture the effects of coastal proximity – 3 hours.
<ul style="list-style-type: none"> Plans/Goals for Next Week 	<ul style="list-style-type: none"> Complete the spatial analysis by performing buffer analysis and further distance calculations. Begin integrating the results of the spatial analysis with the property transaction data to start applying the Hedonic Pricing Model. Prepare initial findings and visualizations for discussion with the mentor.
Reflections	This week marked the transition from data preparation to actual analysis. Importing the data into QGIS and beginning the spatial analysis has provided a solid foundation for understanding the spatial relationships between property values and coastal proximity. The ongoing refinement of the models ensures that we are well-positioned to derive accurate and meaningful insights in the coming weeks.
Project Health	Green
Issues Arising	none

Week 6: 19 August - 25 August

Work Completed and Time Spent	<ul style="list-style-type: none"> • Data Filtering and Matching: Refined the property transaction dataset by filtering out non-coastal areas and matching coastal properties with their respective latitude, longitude, and suburb information using legal descriptions as the unique key – 6 hours. • Integration of View Data: Added the 'view' and 'scope of view' attributes to the dataset, ensuring that properties with specific view characteristics (water, city, or landscape views) are accurately represented – 3 hours. • Data Restructuring: Rearranged the dataset to organize property transactions by unique properties, consolidating multiple sales records into single entries and tracking changes over time – 4 hours. • Coastal Proximity Measurement: Imported a coastline file into QGIS and used spatial analysis tools to measure the distance from each property to the nearest coastline. This data was then matched back to the transaction dataset – 4 hours.
Plans/Goals for Next Week	<ul style="list-style-type: none"> • Finalize the HPM model specification, ensuring all relevant variables and interaction terms are correctly incorporated. • Begin running the HPM analysis on the prepared dataset, focusing on the impact of coastal proximity and views on property values. • Validate the model through cross-validation techniques and start interpreting the preliminary results.
Reflections	This week was pivotal in preparing the dataset for advanced analysis. By integrating additional attributes such as view and scope of view, and accurately measuring the proximity to the coast, the dataset is now well-prepared for the Hedonic Pricing Model analysis. The discussions and reviews of the model's structure have provided a clear direction for the next steps, ensuring that the upcoming analysis will be robust and meaningful.
Project Health	Green
Issues Arising	none

Week 7: 26 August - 1 September

Work Completed and Time Spent	<ul style="list-style-type: none"> • Refinement of Methodology: Re-evaluated the approach for estimating property values due to challenges in acquiring historical housing price index data. After thorough consideration, I decided to use the Repeat-Sales Method to estimate the 2024 property values based on the two most recent sales – 3 hours. • Data Processing: Applied the Repeat-Sales Method across the dataset. This involved calculating the Compound Annual Growth Rate (CAGR) for both Total Value and Land Value using the most recent and second most recent sale prices. Excel formulas were carefully constructed and applied across the dataset to estimate the 2024 values – 6 hours. • Validation of Results: Carefully reviewed the calculated CAGR values and the resulting 2024 estimates to ensure accuracy. Verified the calculations against known trends and adjusted where necessary to
-------------------------------	--

<ul style="list-style-type: none"> Plans/Goals for Next Week 	<p>ensure the method was applied consistently across all properties – 4 hours.</p> <ul style="list-style-type: none"> Preparation for Analysis: With the 2024 values estimated, I prepared the dataset for the upcoming application of the Hedonic Pricing Model (HPM). This included organizing and cleaning the data to ensure it is ready for regression analysis – 2 hours. Begin applying the Hedonic Pricing Model (HPM) to analyze the impact of coastal proximity, property views, and other factors on the estimated 2024 property values. Generate initial findings and visualizations that illustrate the relationship between these variables and property values. Discuss the preliminary results with my mentor to ensure the approach is robust and aligned with the project's goals.
Reflections	<p>This week marked significant progress in preparing the data for final analysis. The decision to switch to the Repeat-Sales Method for estimating 2024 property values was crucial in overcoming data limitations, and it has provided a solid foundation for the next stage of the project. The calculations were carefully validated, ensuring that the results are reliable. I feel well-prepared to begin the detailed analysis using the Hedonic Pricing Model in the coming week.</p>
Project Health	Green
Issues Arising	none

Week 8: 2 September - 8 September

<p>Work Completed and Time Spent</p>	<ul style="list-style-type: none"> Exploration of Additional Variables: After reviewing the dataset further, I identified more independent variables to include in the Hedonic Pricing Model (HPM). In addition to proximity to coast, view, and land/floor area, I discovered that land value in 2024 could be a valuable addition. After discussions and further analysis, I evaluated the pros and cons of including land value as an independent variable – 3 hours. Model Revision: Based on the new variables, I adjusted the structure of the Hedonic Pricing Model. This included modifying the regression equation to incorporate the land value and testing different specifications, such as adding interaction terms and checking for multicollinearity. Several preliminary models were run to evaluate their performance – 5 hours. Multicollinearity Testing: Given the potential for multicollinearity between land value and other variables (like land area), I performed Variance Inflation Factor (VIF) testing to assess the severity of multicollinearity in the revised model. Some variables were adjusted, and tests indicated that multicollinearity was manageable – 3 hours.
<ul style="list-style-type: none"> Plans/Goals for Next Week 	<ul style="list-style-type: none"> Fine-tune the Hedonic Pricing Model by refining the interaction terms and potentially excluding variables that contribute to multicollinearity or offer limited explanatory power. Conduct deeper validation of the model results and compare different versions of the model (e.g., with and without land value). Discuss these findings with my mentor to ensure alignment with project objectives and gain additional insights for further refinement.

Reflections	This week marked a critical phase in finalizing the structure of the Hedonic Pricing Model. I was able to make significant progress in enhancing the model by including additional variables like land value. Testing for multicollinearity was essential, and I feel more confident now that the model is becoming robust.
Project Health	Green
Issues Arising	none

Week 9: 9 September - 15 September

Work Completed and Time Spent	<ul style="list-style-type: none"> Project Scope Update: Finalized the decision to focus the project entirely on Christchurch, excluding Dunedin from the analysis. This adjustment was made to ensure that we have sufficient time for an in-depth study of Christchurch coastal property values – 2 hours. Data Preparation and Adjustments: Completed final adjustments to the dataset, focusing on properties in Christchurch's coastal suburbs. The dataset now includes log-transformed variables for property distance to the coast, land area, and floor area, as well as dummy variables for water view, flooding zone, and suburbs – 4 hours. Model Specification: Designed the final Hedonic Pricing Model (HPM) to analyse the property values in Christchurch. The model includes both categorical and continuous variables, and we defined how dummy variables would be handled in the model – 3 hours. R Analysis Setup: Began implementing the regression analysis using R. The model structure was reviewed, and initial runs of the regression were conducted to validate the setup. The process involved checking the data for any inconsistencies and ensuring that the dummy variables were correctly implemented – 4 hours. Documentation Update: Updated all relevant project documents, including the project proposal and timeline, to reflect the exclusion of Dunedin from the scope. This ensured that all future work would focus solely on Christchurch – 1 hour.
Plans/Goals for Next Week	<ul style="list-style-type: none"> Finalize the regression analysis in R, including interpreting the coefficients for the various independent variables. Perform a thorough evaluation of the results to identify key drivers of property values in Christchurch's coastal areas. Begin writing the findings and insights section based on the regression output.
Reflections	The decision to focus only on Christchurch has provided more clarity and allowed us to allocate resources and time more effectively. With the final model in place, the analysis is now progressing smoothly. Initial regression results seem promising, and the next steps will involve a deeper dive into the interpretation of the data.
Project Health	Green
Issues Arising	none

Week 10: 16 September - 22 September

Work Completed and Time Spent	<ul style="list-style-type: none"> • Removed Influential Points from Model – 8 hours Conducted an in-depth analysis of influential data points using Cook's Distance. Carefully reviewed and removed these points to improve the overall model performance. • Re-ran and Optimized Hedonic Pricing Model (HPM) on Cleaned Data – 6 hours After removing influential points, re-ran the regression model. Conducted a detailed examination of the new model summary, making necessary adjustments to optimize model accuracy and interpret results. • Extensive Diagnostic Checks and Refinement – 7 hours Performed comprehensive diagnostic checks, including residual analysis, heteroscedasticity tests, and autocorrelation testing using the Durbin-Watson test. Plotted Cook's Distance for the cleaned model, analysed results, and ensured no significant influential points remained. • Prepared for Reporting and Further Analysis – 5 hours Began structuring the analysis section for the project report, including summarizing key findings and preparing visual aids. Reviewed remaining data for any additional cleaning or necessary reanalysis.
• Plans/Goals for Next Week	<ul style="list-style-type: none"> • Complete the final iteration of the Hedonic Pricing Model and solidify findings for the report. • Draft the analysis section of the project report with a focus on clear and detailed presentation of results. • Meet with the mentor to discuss the progress and refine the remaining project steps, ensuring alignment with project goals.
Reflections	This week was intensive but crucial in refining the model and ensuring the integrity of the analysis. Removing and reassessing influential points greatly enhanced the model's reliability. The extensive diagnostic checks confirmed that the model assumptions were more closely met, resulting in a more accurate and dependable analysis of property values in Christchurch.
Project Health	Green
Issues Arising	none

Week 11: 23 September - 29 September

Work Completed and Time Spent	<ul style="list-style-type: none"> • Model Implementation (GLM): Transitioned the analysis model from the log-linear model to a Generalized Linear Model (GLM) with a Gamma family and log link, which is more appropriate given the characteristics of our data—specifically, the non-negative property values. This phase included adjusting the model specification to reflect the unique nature of the GLM structure – 5 hours.
-------------------------------	---

	<ul style="list-style-type: none"> • Assumption Checks & Diagnostic Analysis: Performed extensive diagnostic checks for the GLM, including tests for multicollinearity (VIF), residuals analysis, and heteroskedasticity (Breusch-Pagan test). These checks ensured that the model met the assumptions necessary for accurate and reliable inference – 3 hours. • Model Validation: Conducted model validation using cross-validation and deviance residuals. Additional evaluation metrics, including AIC and BIC, were compared with other models (log-linear, polynomial, interaction, and GAM). This allowed for a robust analysis of model performance – 4 hours. • Interpretation of Coefficients & Goodness-of-Fit: Interpreted the coefficients for the GLM, discussing the impact of each independent variable (e.g., distance to coast, land size, floor area, suburb, etc.) on property prices. Calculated McFadden's pseudo-R² for overall model fit. Despite slightly higher AIC/BIC values compared to other models, the GLM remains the preferred approach due to its ability to handle heteroskedasticity and the nature of the dataset – 3 hours.
<ul style="list-style-type: none"> • Plans/Goals for Next Week 	<ul style="list-style-type: none"> • Finalize GLM Reporting: Consolidate the results of the GLM analysis, focusing on model fit, interpretation, and how proximity to the coast affects property values in Christchurch. • Compare with Other Models: Although the GLM is the preferred model, I will continue comparing it with alternative models using AIC, BIC, and residual diagnostics to ensure it remains the best approach. • Prepare the Final Report: Begin writing the comprehensive analysis report, integrating the findings, interpretation of results, and the overall conclusions based on the statistical analysis.
Reflections	<p>This week marked a significant milestone with the successful implementation of the Generalized Linear Model. The diagnostic tests and assumption checks were crucial to validating the model's robustness, and the cross-validation process provided further confidence in the model's predictive capabilities. While alternative models (e.g., GAM, polynomial) performed well based on AIC/BIC, the GLM's ability to manage the dataset's variance and its flexibility made it the best option for the project. The transition from a log-linear model to a GLM has been justified, as this approach aligns well with the data's characteristics and the project's goals.</p>
Project Health	Green
Issues Arising	none

Week 12: 30 September - 6 October

Work Completed and Time Spent	<ul style="list-style-type: none"> • GLM Model Finalisation (6 hours): Finalised the Generalised Linear Model (GLM) using a Gamma distribution, ensuring the inclusion of all significant predictors. This involved reviewing model diagnostics and ensuring that the model structure accurately captured the relationships in the Christchurch coastal property dataset. Special attention was given to validating the robustness of the model and addressing any potential issues from earlier assumption checks. • Model Diagnostics and Validation (4 hours): Conducted a comprehensive set of assumption checks and diagnostic tests for the
-------------------------------	---

<ul style="list-style-type: none"> Plans/Goals for Next Week 	<p>GLM. This included analysing residuals, checking for multicollinearity using VIF, performing heteroscedasticity tests, and running cross-validation. These steps were vital in ensuring that the model was statistically sound and ready for integration into the final analysis.</p> <ul style="list-style-type: none"> Final Report Preparation - Outline and Structure (8 hours): The main focus this week was organising the content outline for the final report. A detailed structure was created based on the marker's provided template, which was adapted to fit the specific needs of this research project. This involved planning the flow of information, determining the most effective way to present the statistical findings, and incorporating the insights gained from the analysis. Discussions were held with the mentor to refine the structure, ensuring it aligned with both academic standards and project objectives. Discussion with Mentor and Report Design (2 hours): Engaged in a detailed discussion with the mentor regarding the overall report design, the most relevant findings to highlight, and how best to present the GLM model results. Based on the feedback, adjustments were made to the report structure, ensuring that the statistical insights would be presented in a clear and meaningful way to a broader audience, including those without deep statistical knowledge.
Reflections	<p>This week marked significant progress towards wrapping up the project, with a focus shifting from detailed model testing to preparation for the final report. The report design discussions were particularly useful in clarifying how to present the complex statistical findings in a way that will be accessible to various audience groups. The workload was heavier than usual, as organising and structuring the report required careful thought and planning to ensure that the final submission meets expectations.</p>
Project Health	Green
Issues Arising	none

Week 13: 7 October - 13 October

<p>Work Completed and Time Spent</p>	<ul style="list-style-type: none"> Final Report Writing: This week focused heavily on finalising the project report. The structure, content, and statistical analyses have been integrated into a cohesive document. Key sections such as the background, literature review, and detailed analysis of the Generalised Linear Model (GLM) have been completed – 12 hours. GLM Model Findings & Interpretation: Revisited and refined the GLM model findings, ensuring all coefficient estimates, assumptions, and model validations are accurately discussed. The interpretation of the
--------------------------------------	--

	<p>statistical results has been crafted to provide meaningful insights for both technical and non-technical audiences – 5 hours.</p> <ul style="list-style-type: none"> • Proofreading & Review: Started the process of proofreading the entire document and getting feedback from the project mentor on the quality and clarity of the report, especially on the statistical analysis sections – 3 hours.
<ul style="list-style-type: none"> • Plans/Goals for Next Week 	<ul style="list-style-type: none"> • Complete Final Report Review: Complete proofreading and final revisions based on the feedback provided. Ensure all sections are aligned with project goals and requirements. • Submit Final Report: Prepare and submit the final version of the report for review and grading.
Reflections	This week marked significant progress in consolidating the findings of the project into a structured, comprehensive report. Most of the writing has been completed, and the remaining focus will be on refining and ensuring accuracy.
Project Health	Green
Issues Arising	none

Week 14: 14 October - 20 October

Work Completed and Time Spent	<ul style="list-style-type: none"> • Final Report Completion and Review: This week marked the successful completion of the final project report, including all sections, appendices, and the integration of statistical analysis results. The final proofread and review of the report ensured that all components, including the R code, GIS data, and supporting materials, were well documented and clearly explained – 10 hours. • Creation of Comprehensive R Script Document: I documented all the R code used for data analysis, model testing, assumption checking, and results visualization into a well-organized script. The script ensures reproducibility and serves as a reference for future research – 6 hours. • Project Submission Preparation: Organized all project outputs (cleaned data, structured Excel file, QGIS project, and R script) into a GitHub repository for final submission. This ensures all artefacts are accessible and well-presented for submission – 3 hours. • Repository Finalization: Uploaded all relevant documents, code, data, and project files to GitHub and reviewed the content to ensure everything is properly documented and linked – 2 hours.
<ul style="list-style-type: none"> • Plans/Goals for Next Week 	<ul style="list-style-type: none"> • Project Presentation Preparation: Prepare for the final project presentation by organizing key findings, methodology, and visual elements into a cohesive presentation format. Focus will be on simplifying key concepts for the audience and ensuring clarity in data visualization.
Reflections	This week has been highly productive with the completion of the final report and the successful organization of all project artefacts. With the R script fully documented and the project files ready for submission, the focus now shifts to preparing a clear and engaging project presentation that effectively communicates the findings and methodologies to a broader audience.
Project Health	Green

Issues Arising	none
----------------	------