Executive Summary - Happipad Capstone Project

Happipad exists to provide alternative avenues for access to housing opportunities for both renters and landlords, with a focus on affordability and accessibility. Our goal for this project was to utilize their data in two ways: to employ data modelling techniques to provide additional insights into their business, and use dashboarding to display geographical demand and temporal changes in rental activities on the Happipad platform.

We developed two dashboards: a Renter Overview and a Property Overview. To assist Happipad in understanding the housing supply, the Property Overview displays the geographical distribution of properties across Canada, and other important features such as the distribution of prices and property types and features in different locations. The Renter Overview helps give insight into housing demand and its changes over time by displaying the distribution of renters across Canada and their budget information.

We also employed machine learning modelling techniques in an attempt to predict the price of rental contracts based on the available data of the homes, such as the city, province, type of bed, and more factors. Happipad can use this information to understand the most important features of homes or listings on their prices, and to notify users if their listings have a price far from the predicted price in order to assist hosts in getting their room rented. Based on the success in previous research of the XGBoost, LightGBM, and random forest models of predicting home prices, we decided to use these models. The best performer was the random forest, with an RMSE of 133 (compared to the average rental price of $835 overall).

To assist the predictive modeling and explore the data further, we applied Ollama for natural language processing (NLP) for the free text property titles and descriptions provided in the dataset. A Mistral model was selected because of its light weight and relatively high efficiency. Highly detailed instructions, in combination with JSON-based outputs, and allowing for high flexibility with analysis of open-text responses led to better extraction of the data. Although this information did not improve model performance, this is likely due to the small sample size, and it has the potential to improve predictions as there is more data collected.

Taken together, we have provided Happipad with an effective pipeline to input their future data to get visual business insights through dashboards, as well as better understand the price distributions of their rental contracts.