VanReaper – Vancouver Real Estate Analysis & Predictions

SFU SIMON FRASER UNIVERSITY

Chirag Ahuja, Ekramul Hoque, Pavan Kosaraju, Rohith Sooram

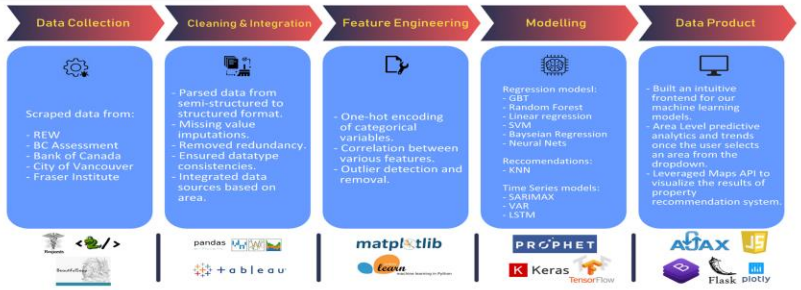
Introduction

Buying a property is financially the biggest decision you make in life. Given the enormity of the purchase and complexity of the factors that go into the decision, the situation easily leads to decision paralysis. Our data product aims to simplify this problem by helping people make informed decisions by encapsulating our machine learning models inside a intuitive and easy to use web frontend. The web frontend provides property and rental recommendations to people based on their preferences thereby eliminating the need for visiting multiple websites to get property information.

For people who are looking for places in a specific locality, our product provides area level predictive analytics of the real estate once the user selects name of the area from the dropdown. We have leveraged the power of Google Maps to visualize both - the results of recommendation systems and area level predictive analytics provided by our backend machine learning models.

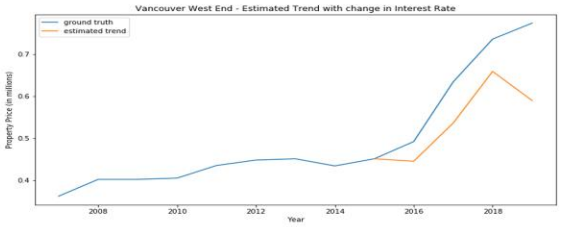
In addition to that, we have discovered intriguing correlations between the real estate price and other features like school rating, crime rate in the area and the lending interest rates prevalent at the time.

Data Science Pipeline

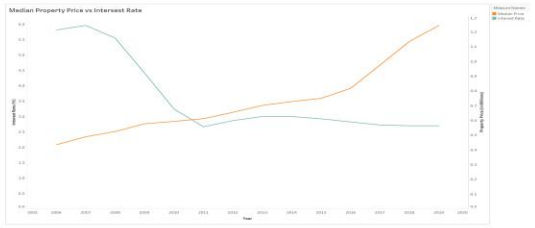


Time Series Analysis

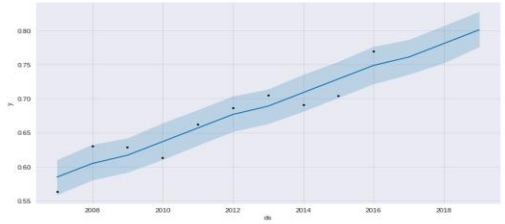
Trend estimations using LSTM.



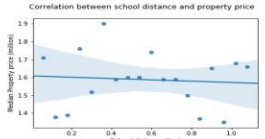
Inverse correlation between interest rates and property prices.

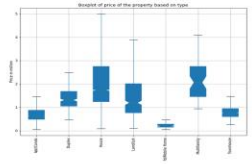


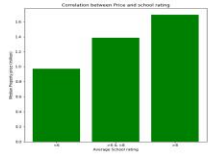
Time series trend analysis and predictions done using Facebook Prophet library on Vancouver Housing Prediction.

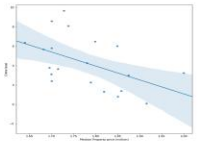


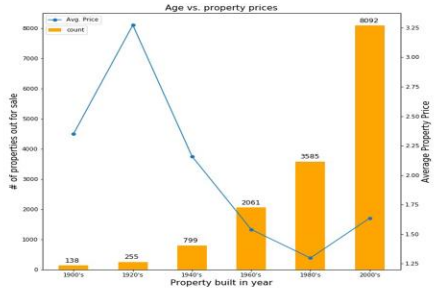
Exploratory Data Analysis

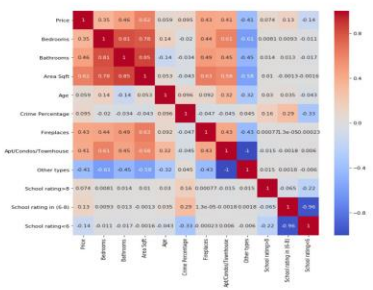


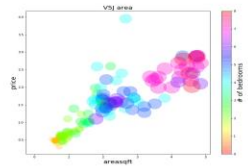


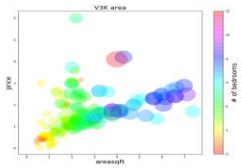


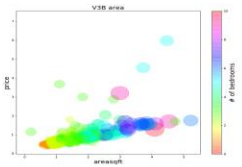






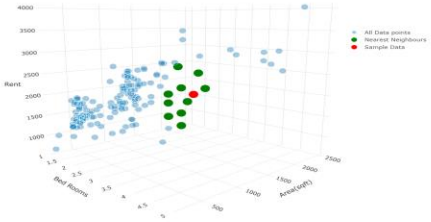




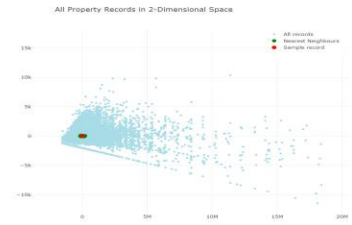


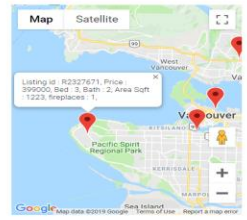
Recommendation Systems

10 nearest neighbours used in the Property Recommendation System in 3D space.

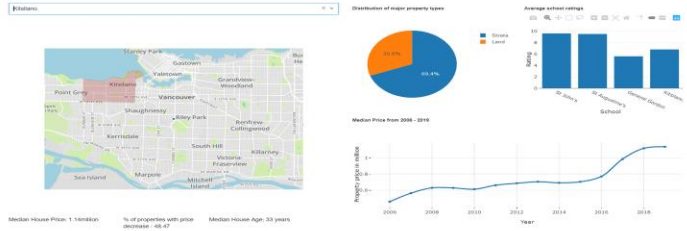


2D plot of 30 nearest neighbours used in the Rent Recommendation System. On the right, these neighbours have been pinned on Google Maps.





Summary



We have scraped data from multiple sources, performed exploratory data analysis and feature engineering to get the data in proper shape for feeding it into predictive models. Regressions models like Gradient Boosting Trees, Random Forrest, and Time Series models like ARIMA, LSTMs and Vector Auto Regressor were used to predict future prices of the properties. Property Recommendation System was implemented using k-nearest neighbors algorithm in Sci-kit learn. Finally all the statistical models were serialized, persisted and deployed using an interactive Flask Web Application.

Lessons learnt

- The importance of collecting more data: Sometimes in the real world, improving model accuracy is more about collecting more labelled data than applying advanced algorithms on limited data.
- Feature engineering: Real world data is messy and the scraped data did not have the features in the form that can directly be ingested by the models. We had to perform feature engineering to transform the features into a format that can be fed into the models.

Future work

After significantly more amount of data has been collected, we would like to do a further dive into different state of the art deep learning models to predict housing prices in Vancouver. We will Explore advanced econometric regression techniques like hedonic regression and splines to predict property rates.