[**Advantages (+)**](#_dh0qtil7kk7t) **1**

[Data Warehouse](#_rwo2j3bevurw) 1

[Machine Learning:](#_3046g1z602nv) 1

[Power BI](#_8hgafu30nsj1) 2

[**Limitations (-)**](#_iwsoxrsyol4f) **2**

[Data Warehouse](#_94s46ipy8cc6) 2

[Power BI](#_gu6jctw6vr8k) 2

[Machine Learning](#_uxqlo5sstbfl) 3

[**Future Suggestions**](#_lg4u5417tknm) **3**

[Data Warehouse](#_m7ykpfxcqx58) 3

[Power BI](#_b4lqaqhpv27x) 3

[Machine Learning](#_92ap9k8bo3ev) 3

D11 – Project Evaluation

# Advantages (+)

## Data Warehouse

1. A Data Warehouse was created with the following attributes:
   1. Subject oriented,
   2. Integrated,
   3. Time variant,
   4. Non-volatile,
   5. Summarized
2. Two DataMarts focused on Reviews and Calendar.
   1. Improves the response time of business users by allowing users to access the particular form of data they need.
   2. A condensed, more oriented version of a warehouse of data.
   3. Business Users do not interact with the main Data Warehouse (security)

## Machine Learning:

1. We trained and evaluated the price prediction performance of six different models based on a property’s attributes:
   1. Lasso Regression,
   2. Ridge Regression,
   3. Linear Regression,
   4. Elastic Regression,
   5. Random Forest,
   6. Decision Tree: Furthermore, a different approach has been used for Random Forest and Decision Tree. A grid search has been used. **T**he data scientist sets up a **grid** of hyperparameter values and for each combination, trains a model and scores on the testing data.

## Power BI

1. Using Power BI to visualize the data from machine learning and Data Warehouse to be easier for the users to interact with.
   1. Easy filtering of data with filters and slicers
   2. Leaderboards for easy identification of the target property or host for questions the business users are looking to answer
   3. Maps for locating those leaders
2. Direct connection of the visualization tool with our SQL server
3. Power BI has helped for a sanity check on our database to confirm the correct flow of the data.

# Limitations (-)

## Data Warehouse

1. Could not create all Data Marts as there aren’t any explicit business departments available.
2. We haven’t really used the created Data Marts anywhere in our system
3. We didn’t use an OLAP system that could summarize data by eg property, time-period, or other dimension. An OLAP Cube is a data structure that allows fast analysis of data according to the multiple Dimensions that define a business problem.
4. We have used Dimension type 0 for the Dimension tables of our project. This means that values in a Table cannot change (non-volatile DW).
5. Reviewers comments have been cut when the size is over the maximum.

## Machine Learning

The machine learning algorithms that have been used were regression models that could predict the price of a property. The limitation was that we could not use classification models such as k-nearest neighbours (K-NN), Logistic Regression and Support Vector Machine. With a result, the metrics we used as evaluation of the model was only the MAE and RMSE without using confusion matrix (accuracy, precision, recall, f1-score).

# Future Suggestions

## Data Warehouse

1. Identify the different departments and create specific Data Marts for every department.
2. Using a dimension type 1, 2 or 6 for our data warehouse, instead of type 0, we can permit values to change in the data warehouse eg. when an attribute is updated. An example is if a property has some updated attributes and this has to be reflected in our data warehouse.

## Power BI

More analytical all around tool for Representations especially for locating overvalued, undervalued and hosts to notify.

## Machine Learning

1. Clustering prices so we have more evaluation metrics like (accuracy, precision, recall, F1-score).
2. After clustering use of :
   1. k-nearest neighbours (K-NN)
   2. Support Vector Machine
3. Sentiment Analysis according to the comments provided by reviewers.