

**BIDV PROJECT**

**Department of Computer Science and Engineering**

**The NorthCap University, Gurugram**

**Name- Shivraj and Gaurav**

**Roll No. – 20csu349 and 20csu342**

**Class- DS-D (7)**

**ABOUT DATA SET**

* **This dataset contain price and are where the AirBNB house are present in Vienna.**
* **The data was taken from the site of github and was updated till the last date.**
* **The data set was contained in the form of the csv file.**

**INDEX**

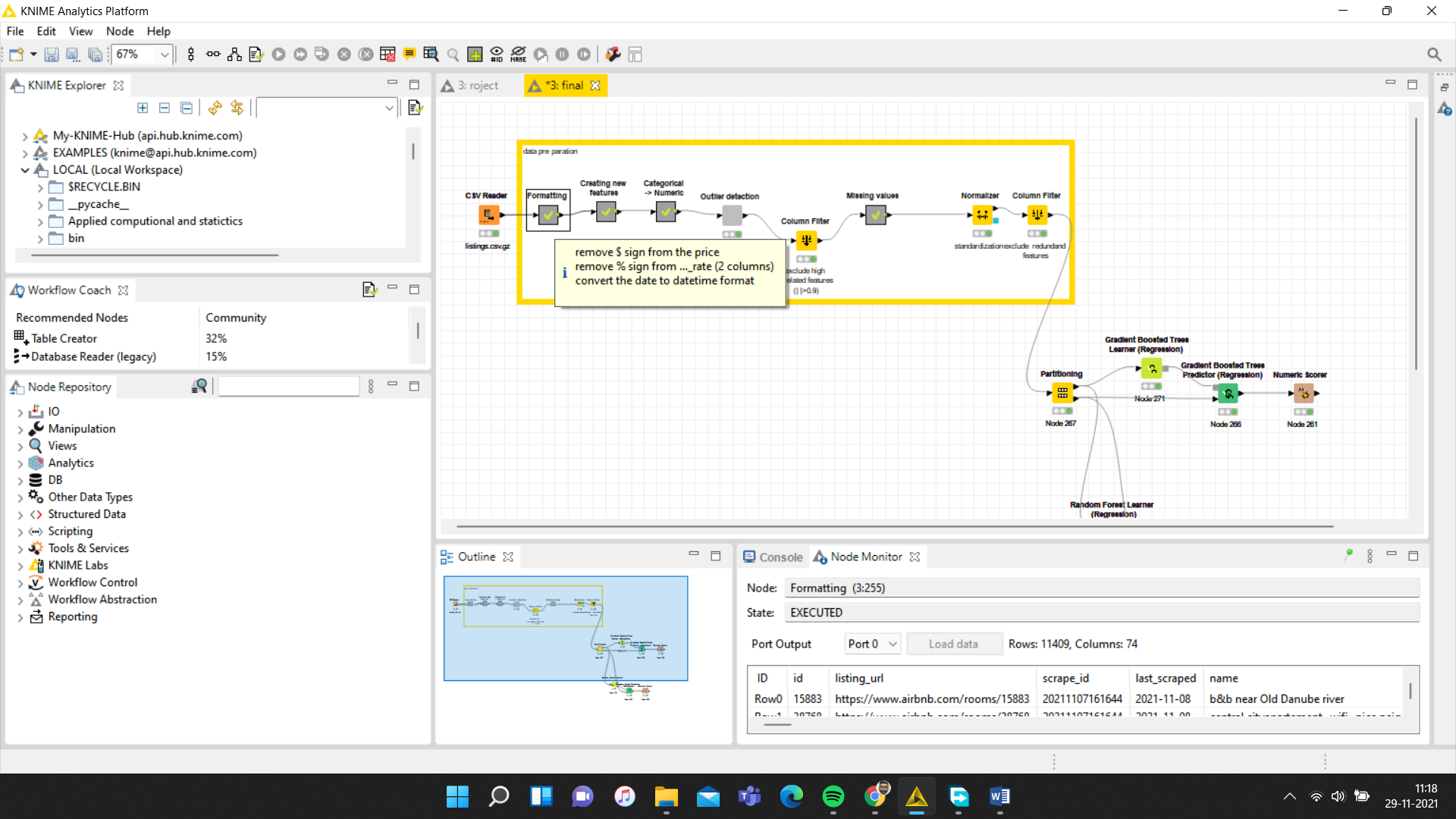
|  |  |
| --- | --- |
| **S.no** | **Experiment** |
| **1.** | **Intro** |
| **2.** | **About data** |
| **3.** | **Knime** |
| **4.** | **Conclusion** |

**The Price prediction of AirBNB**

* First we will do the data pre-processing on the file and delete or exclude all the unwanted data from the project.

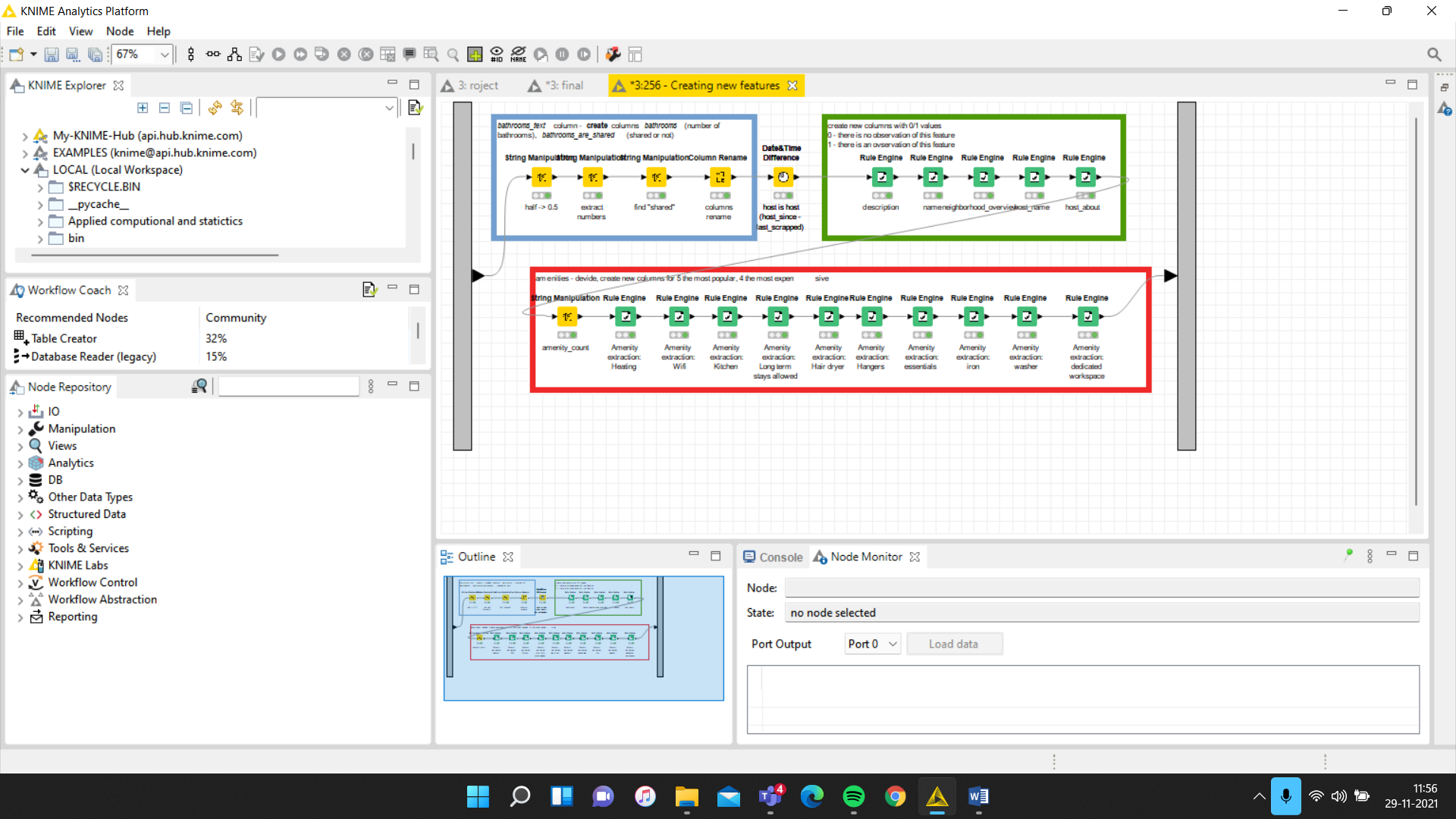
1. Formatting metanode

Clear all the special characters from the data .

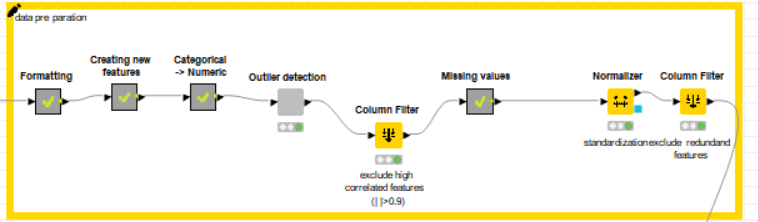


1. Create features

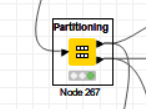
Add new columns in data.



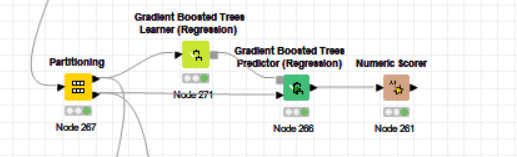
1. Add new data, delete missing value standardization and exclude the redundancy from the data.



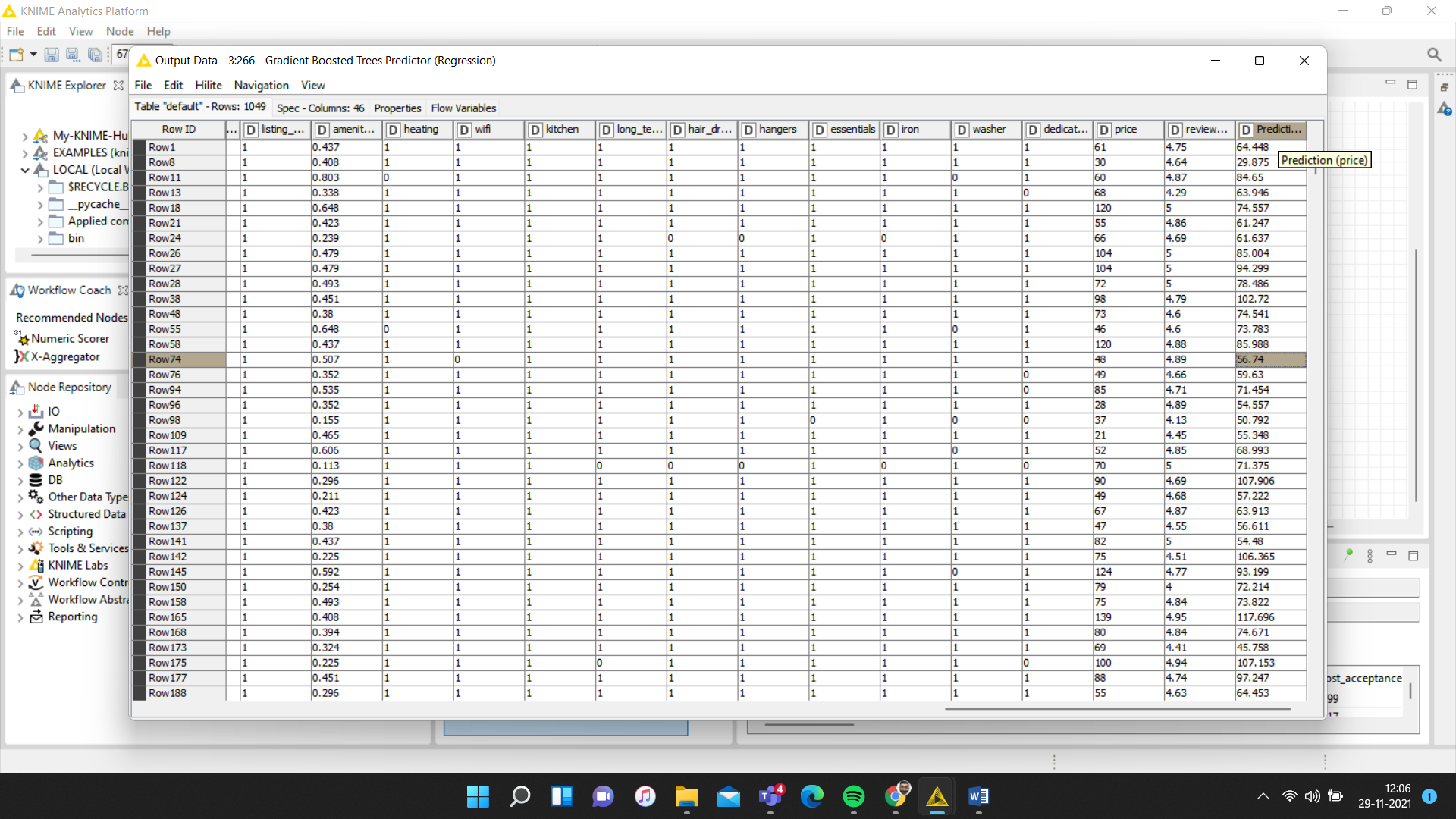
* Partition of data for test and train data.

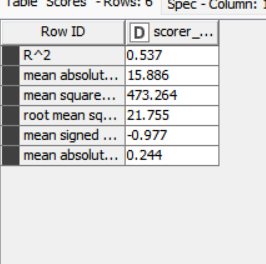


* Gradient Boosting for regression

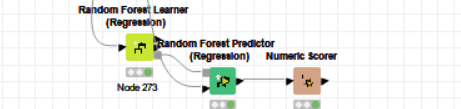


We find the predicted price of the property and then we check the r square value.

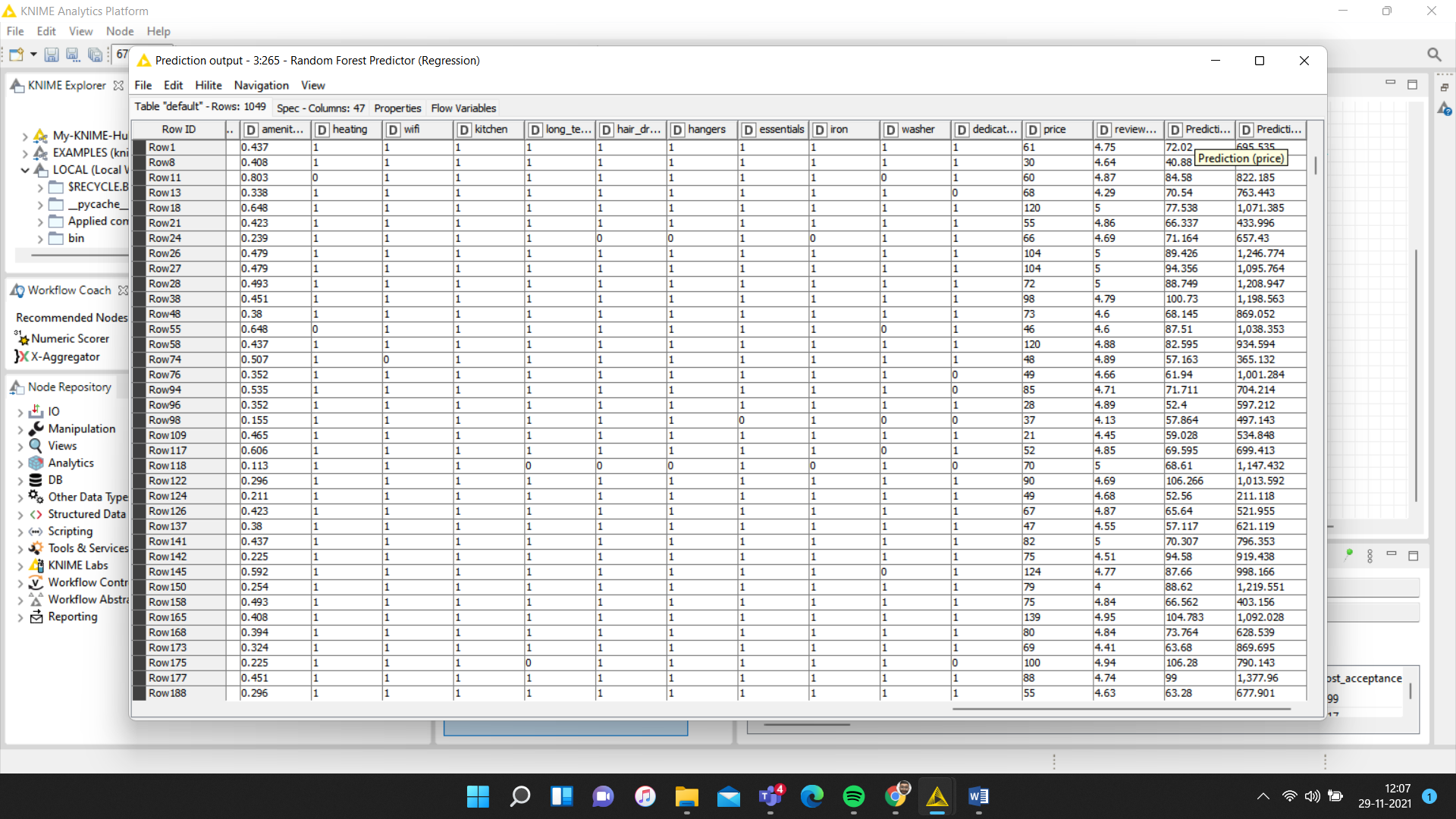


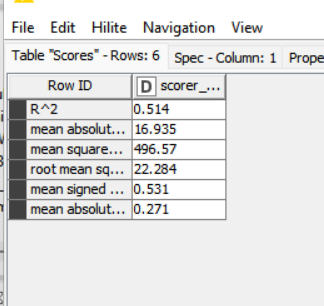


* Random forest leaner



We find the predicted price and variance of the property and then we check the r square value.





**Conclusion**

In the project, we analyse the Airbnb properties in Vienna, extract meaningful information from the data, and establish a forecasting model that would answer the price-oriented questions we generated. As a first step of data preparation, normalization and format changes were performed on the necessary columns. In the next stage missing data treatment and outlier extraction were conducted in order to clean the dataset. As a result, all features were present, no missing values and/or outliers remained. After familiarizing ourselves with all the data, we formulated an important and overarching question to guide our research, which is: How do house prices and incomes differ in different districts and/or neighbourhoods in Vienna? We also had two sub-questions to get more insight into the matter: Which features have the biggest impact on price? Also, what are the most popular and most expensive amenities? We created models to answer these questions and tested these models for score prediction as well. We compared our results in this report. As a result, we achieved an r^2 value of 57% using Gradient Boosting Machine Learner for price prediction. Ultimately, our goal was to be able to make price predictions as a way to encourage future hosts and potential users to make price adjustments based on certain features in a particular property. This project has succeeded in establishing a model that can guide users in terms of pricing.