e-Commerce and Machine Learning: A price predicting model for the Airbnb listings of New York City

Olsen, Lars-Peter Lützen, Luis Osorio Muntean, Bogdan

December 18, 2019

Abstract

In this paper we are analysing the benefits of Machine learning in the current e-Commerce environment. We are establishing that recommendation engines are useful both for the users as well as for the service providers. We then focus on a case study that analyzes the Airbnb listings from New York city due to its population and listings density. Based on our analysis we discover that intrinsic attributes to a listing have a greater impact than the extrinsic attributes.A price predicting model is then developed and though the results are just above average, we recognise that the model can be improved upon. Augmenting the dataset shows an increase in the performnace of the implemented models. We conclude by stating that further development is needed in order to have comparable results with the existing models implemented in e-Commerce platforms.

1 Introduction

Machine Learning is an important area for the development and creation of new models and activities that support the advancement of computing systems [1]. In recent years, such technology has been used heavily for many aspects of our modern society; from content filters for social media networks to recommendation engines on e-Commerce websites [2] and many other consumer products such as smartphones, security systems, and personal assistants [3]. IBM's Watson, Elon Musk's OpenAI and advances in the development of autonomous vehicles [4, 5] are all examples of systems that are using machine learning and that are raising awareness of such topic in public discourse.

E-Commerce is one of the areas that is using machine learning, with models and systems being built to mine large quantities of data in order to either optimize the search results that a user is seeing or to recommend similar and/or related products within a price threshold [6, 2]. Werthner et al. [7] is portraying the state of tourism and traveling before the e-Commerce boom, stating that with the advent of the internet being used by the general population, tourism will undergo a series of changes that will allow for traditional tourism-related transactions to happen in so-called smart e-Marketplaces [7].

Airbnb is the perfect embodiment of such claims, revolutionizing the meaning of traveling on a budget and shared-economy. The non-conventional business model of the company combined with e-Commerce thriving[1] has established Airbnb as a popular choice for those that are listing short term rentals and those that are traveling. Their notoriety has in part come from the recommendation engine based on machine learning that drives 99% of their bookings[2]. We believe that a solution focused on price optimization and recommendation would be beneficial for the company and its users.

Machine Learning is an approach to computer science that allows a computer program to use large amounts of data and computational techniques to make decisions without human intervention and adjust over time by 'learning'. The consensus is that machine learning algorithms become better with the quantity of data they get fed with [8]. E-Commerce is known for generating large amounts of data that is used by machine learning systems for optimization and recommendation processes [6, 2]. With tourism joining the trend, machine learning has been identified as a candidate that can improve this field [7, 9]. Similarly to how stock price prediction models work [10], the price of a fare[11] or rental can be determined as well by using a set of non-parametric data.

This paper focuses on analyzing and implementing

[1] E-commerce sales worldwide - `http://tiny.cc/0rbchz`
[2] Listing Embeddings in Search Ranking - `http://tiny.cc/fobchz`

a machine learning model that can predict the price of a listing based on various non-parametric observations from the Airbnb listings in New York City. The primary motivation for choosing New York City as our target for this project has to do both with the amount of Airbnb listings in New York City, as well as the plethora of tourist that enticed by its modern metropolis look, authentic American characteristics, distribution of tourist attractions and ethnic and racial diversity, are visiting the city every year [12].

The dataset is public and was taken from Airbnb by OpenDataSoft [3]. It contains approximately 40 thousand observations of listings from New York City each with 29 features, 8 categorical and 21 numerical.

1.1 Research Question

Inspired by the features that are present in the dataset, our research question is:

RQ: Does the price of an Airbnb listing change when different features are fluctuating?

- Q1: Is crime rate a detrimental factor in predicting the price of a listing?
- Q2: Are touristic attractions boosting the price of a listing?
- Q3: Are reviews influencing the price?
- Q4: Does the borough or neighbourhood where a listing is located have an influence on the predicted price?

1.2 Limitations

The limitations of this project come from the dataset itself. The lack of a column that specifies the amount of space that a listing has poses our main limitation. Such is the case that ML models will have difficulties predicting the price accurately for two listings situated on the same street, with the same number of amenities and similar rooms if one listing rents 50 square feet and the other 250 square feet - the price should be impacted by that as described by Limsombunchai et al. [13] when building a price recommendation engine for houses in Australia. The dataset is augmented with the top 50 most visited tourist attractions[4], but we have to keep in mind that this is heavily influenced by (a) the number of people living in a specific borough/area and (b) the amount of tourist visiting a certain part of NYC. Moreover, though the size of the dataset is decent, more observations would aid the final product.

The remainder of this paper is organized as follows: We are describing the used methods in Section (2). We then are exploring, visualizing and augmenting the dataset as well as training and optimizing various ML models in Section (4). Section (5) presents our findings and Section (6) concludes the paper.

2 Methods

This project is a pilot study that aims to prove the ability of a machine learning model to predict the price associated with an observation based on categorical and numerical features. Without a doubt, a larger dataset and further investigation will yield better results which may or may not be comparable to the current state-of-the-art price recommendation engines [14].

This paper implements a mixed-method approach where we are combining empirical findings with desk research. In Blake[15] research is defined as a 'systematic, intensive study directed towards a fuller scientific knowledge of the subject'. This project is no exception wherein the scientific knowledge gained from the desk research is applied to nurture our empirical findings. Such findings are based on a dataset that contains all the active listings on Airbnb from New York City with the intent to build on other work in this field. [10, 14, 6].

Nunamaker et al. [16] describes research as an integrated process divided into four main stages. Such classification will represent the backbone of this paper and regardless of the chosen method, will at all times be followed. These phases - *Theory building, Experimentation, Observation, and Systems Development* - are supporting the definition that Blake [15] is giving to research as a whole. Furthermore, in Nunamaker et al. [16] five research methods are addressed and argued for. Out of these five, we have identified the *Exploratory and Developmental Research* as being the best fit for our given scenario.

With this model, we will be searching (or synthesizing) for instructions that will yield a better course of action. Hence, this project will be carried out in time-boxed iterations wherein we will process the data, create visualizations to gain insight, implement machine learning models, measure performance, opti-

[3] Website of OpenDataSoft - `https://www.opendatasoft.com/`

[4] Most visited places in NYC - `http://tiny.cc/7lhehz`

mize these models and draw conclusions for the next iteration; as described by Goldstein[5]. The findings of each iteration contribute to our body of knowledge so that better results can be achieved. These iterations will be furthered in Section (4). Several recurring themes are present in this research paper such as the impact of crime rate, reviews and tourist attractions on a listing, as well as the intrinsic features of an observation.

3 State of Machine Learning in e-Commerce

In recent times, machine learning has unlocked the potential for us to solve problems that would otherwise be impossible to solve without large amounts of resources and time, by using ML algorithms on large datasets [8, 1]. Leshno et al. [17] prove that it is possible, with the help of a multi-layer neural network with a non-polynomial activation function, to approximate any function - if enough data is provided. Such models are 'learning' from sets of data that are labeled in a process called 'supervised learning' [1, 18]. With time and enough data, the model will adjust to correctly predict the value based on variations of the same attributes[6].

Jordan et al. [6] describes how such systems will revolutionize the public sector, especially the healthcare system, commerce and science. Such called recommendation engines have been implemented by various companies that are leveraging e-Commerce [2]. One such example is Amazon [7] which states that 35% of their revenue is generated by their recommendation engine. Airbnb is using a similar item-to-item and collaborative filtering model[8] and claims that 99% of their bookings are being generated by the recommendations that the engine gives to its users. Thus far we have localized the purpose that machine learning serves in e-Commerce and more specifically in recommendation engines. With tourism being one of the early adopters of e-Commerce capabilities [7, 2]. Their feasibility is undeniable, yet we have to understand how neural networks can help us when building a price predicting model.

Price predicting models that are leveraging machine learning and neural networks already exist. One of the earliest iterations is described in Schoneburg et al. [10], where a price predicting model for three stock exchanges is described. The authors are using a combination of discrete values and show the benefits of prognosis on semi-chaotic time series [19]. Later, Limsombunchai et al. [13] develops a neural network that is taking advantage of discrete values, both intrinsic and extrinsic to build a hedonic price model on house prices in Australia. Encouraged by our findings and supported by Guggilla et al. [14] that developed the price recommendation model for Airbnb, we now understand that to create a comparable model to the existing one, a combination of discrete values will have to be explored, such as location features, amenities, review scores and rented surface.

4 Analysis

The dataset contains a little over 40000 observations with a mix of 29 categorical and numerical values. It was extracted from a larger dataset with 400000 observations and 70+ columns [9]. We are focusing on the Airbnb listings from New York City as the listing number is relatively high and the scope is limited to the development of a machine learning model that predicts prices in one city where similar conditions are present. Additionally, New York City is one of the most visited metropolia in the world [12] as well as because our results can be compared with the solution developed by SAP [14] which addresses a similar problem.

Based on the findings from Section (3), we carefully extracted the features that were intrinsic or extrinsic to a listing. Null values will be replaced depending on the respective feature and several columns such as room type, borough or neighborhood will be one-hot-encoded [10]. The dataset will be augmented with features such as crime rate per borough, income per borough, the average size of a room/apartment and tourist attractions in the proximity of 750 meters.

The Pearson correlation is used to quantify the degree of correlation between different data points. Adler et al. are proving that the Pearson Correlation is superior to the Mander's overlap coefficient [20]. While the two are similar, they differ in the use of either the absolute intensities (MOC) and the deviation from the mean (PCC), hence we are making use

[5]Deconstructing Data Science - `http://tiny.cc/9uiehz`
[6]Regression Definition - `http://tiny.cc/5m2dhz`
[7]Amazon's ML model - `http://tiny.cc/lcldhz`
[8]Recommendation engines - `http://tiny.cc/xe2dhz`

[9]Dataset can be found here - `http://tiny.cc/9d5dhz`
[10]One-Hot-Encoding - `http://tiny.cc/ta5dhz`

of it to understand if there is a correlation between the aforementioned themes.

4.1 Assumptions

Going further, we have made several assumptions such as that the hosts have been faithful when creating their listing and that the information is correct. Another assumption when augmenting our dataset with crime rate and population density was that these features are the same throughout a borough, no matter the distance between them. We are also assuming the surface that is rented based on the property type and room type column - As lofts can be 2000 sqft. while the average apartment does not exceed 1000 sqft; assumptions made based on researching the average size of the most prevolent living spaces in New York City. Lastly, when augmenting our dataset with crime rate information, we are considering all crime as being equal in gravity, regardless of its nature - ergo all the observations of crime will be used to compute the crime rate per borough.

4.2 Data processing and visualization

When processing the data, our first step was to check for null values and handle them. The missing names of listings have been replaced with an empty string, security deposit and cleaning fee features were replaced with 0 where the values were null and the review scores were replaced with 0 as it was clear that such listings did not get any review whatsoever. The review scores were using mixed scales therefore we have converted them to percentages.

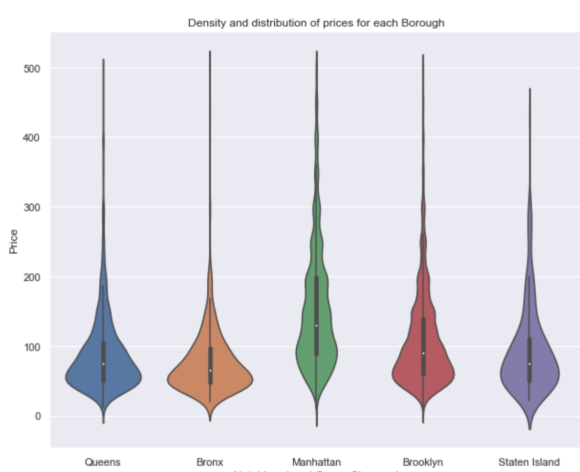

Figure 1: Distribution of prices per borough

Using descriptive statistics we can visualize the price distribution per borough (Fig.1) and what is the listing density throughout the five boroughs of NYC. Manhattan and Brooklyn have the most listings with 18970 and 16691 listings respectively, whereas Queens, Bronx and Staten Island count a total of 4676 listings.

Room types fall into three categories, either entire home/apartment, private room and shared room, the latter having only 2.9%of the listings whereas the other two being evenly distributed at around 48.5%. Fig.2 shows the areas where most listings are present as well as how expensive they are. The property type field, as per Airbnb, falls into two categories: Normal and Unique experiences. As the latter has a total of around 100 listings and 10 categories that a listing can fall into, we have decided to remove such entries mainly because most of them are boats, tents and RVs which can be moved easily and therefore skew our results (with location-related features becoming irrelevant).

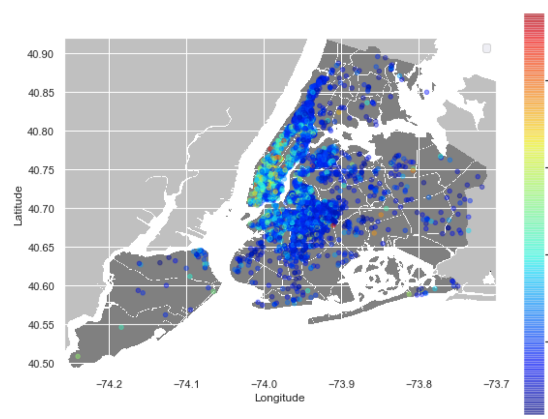

Figure 2: Map that shows where the listings are situated. Warmer colors means more expensive. The dataset has been stratified for this plot because otherwise it would be difficult to read; therefore we are using 10% of the total listings to display the price distribution and which areas have the most listings.

The neighborhoods and boroughs have been one-hot-encoded so that the extrinsic features for an Airbnb listing that are describing the location can be leveraged at a later point by the regression models. Inspired by the work of Levantis et al. [21] concerning crime and how it influences tourism, we have computed the crime rate per borough with the help of another dataset which contained the valid reportings for crime between January 2018 and January 2019 [11]. Table (1) showcases the crime rate per borough per 1000 citizens, the average income per borough and the population per borough.

Borough	Crime_RT	AVG_Wage	Population
Bronx	69.05	$35.564	1.432.132
Brooklyn	51.76	$48.758	2.582.830
Manhattan	68.63	$175.960	1.628.701
Queens	38.60	$46.829	2.278.906
Staten I.	42.74	$54.908	476.179

Table 1: Demographics per Borough in NYC. Crime rate is calculated per borough per 1000 people.

That concluded the initial processing of the dataset. During the two iterations, we will be focusing on different themes and features of the dataset to expand our body of knowledge and get better results when applying ML algorithms.

4.3 Iteration I

During the first iteration, we have focused on analyzing extrinsic features of a listing, such as tourist attractions, crime rate, borough population and longitude, and latitude. Sub questions 1, 2 and 4 will be furthered in this section. The first step was to augment the dataset with the crime rate, average income and population features described in Section (4.2). Next, we have computed in a different column the number of points of interest in a radius of 750 meters of the listing. This distance was changed multiple times mainly because while highly populated, New York City does is in the 24th place when it comes to land area[12]. We assume that a listing that has more points of interest in its vicinity will have a higher price that one that does not.

Lastly, we have applied the Pearson Correlation Coefficient and classic regression algorithms on the dataset. There does not seem to be any obvious correlation between the extrinsic values from our dataset. The best ML model was Random Forest which scored a 0.645 R2 score. In similar price predicting models such as the work of Guggilla et al. [14] and Limsombunchai et al. [13] an accurate location does facilitate a better prediction of the price. This is done with the help of boroughs, zip codes, latitude and longitude. Concerning that, we have one-hot-encoded the neighborhoods. It does seem to have better results when we are not including the neighborhoods - this could be because we do not have enough observations to balance our number of features. We will consider further refinements to this approach in future iterations of the project.

Although this iteration did not yield exceptional results, it provided us with a good understanding of what are the main features that drive the price in an Airbnb listing. Section (5) will further our conclusions for this iteration.

4.4 Iteration II

The focus of the second iteration has been the intrinsic values of a listing. Such features are the number of bedrooms, bathrooms, beds, how many guests the listing is for and the amenities that each listing has. As explained in Guggilla et al. [14] there is a correlation between the amenities that a listing has and how the price fluctuates. Initially, we have used natural language processing[13] to extract the most common amenities from the name and description of a listing - this approach did not yield results as hosts would rather use the tick box selection tool to describe the present amenities.

Following up on that, we implemented the same process to extract the amenities from the already present feature. Furthermore, these amenities were one-hot-encoded. Each observation was augmented with such columns wherein each column held the presence or absence of a particular amenity.

As in the first iteration, we have then computed Pearson's Correlation Coefficient to determine which attributes are correlated and if the amenities are correlated to the price. It seems that amenities have a rather small correlation with the price, but interestingly enough, there is a strong correlation between price and the number of people that can be accommodated, the number of beds and bedrooms. This will suffice to train the ML model.

In this iteration, we trained ML models, optimized them and computed different metrics to determine which one is the most appropriate for the task. We ran the classic ML Regression algorithms first - Linear Regression, SVR, XGBRegressor and Random Forest. The default configuration of the Random Forest performed the best with an R2 score of 0.62. We then optimized it by hyper tuning the parame-

[11] Crime rate dataset - http://tiny.cc/4enehz

[12] Top 150 Cities by area - https://en.wikipedia.org/wiki/List_of_United_States_cities_by_area

[13] Stemming and Lemmatization - http://tiny.cc/vgsehz

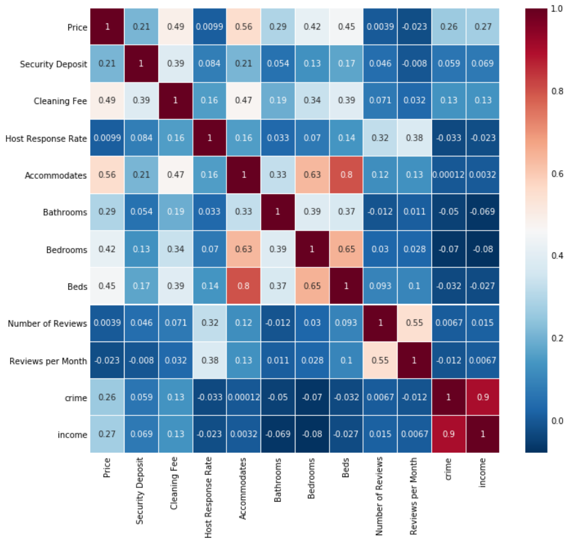

Figure 3: Pearson Correlation Matrix to show the correlation between the price and number of bedrooms, beds and how many people a listing can accommodate. This gives us a good indication that the price of a listing is not random and to some extent is predictable.

ters. The best configuration that we have reached is 1000 estimators, 5 samples split, 1 sample leaf, a max depth of 40. This showed some improvements, from 0.62 R2 Score to 0.67. Fig.4 showcases the mean of the actual values per borough versus the mean of the predicted values.

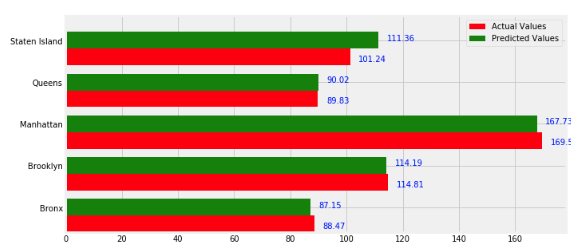

Figure 4: Mean true values compared to Mean predicted values of the optimised Random Forest per borough.

Lastly, we have run multiple configurations of an MLP Regressor - a different number of hidden layers with a different number of perceptrons. The better performing model had a 0.38 R2 score and after it was hyper tuned (1 hidden layer, 90 perceptrons respectively and 800 iterations), the R2 score was 0.592, a significant boost in performance. Other metrics will be discussed in Section (5).

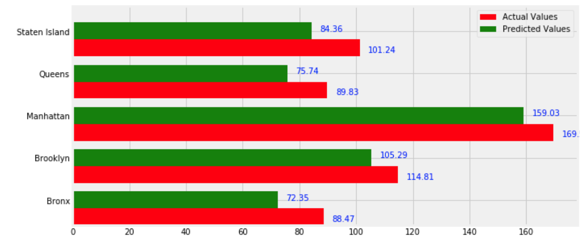

Figure 5: Mean true values compared to Mean predicted values of the optimised MLP per borough

4.5 Iteration III

In the third iteration, we have focused our attention on the biggest limitation of this project, which is the rented surface. We believe that by correlating the price of a listing with the rented surface, better results will be achieved. As the dataset did not have the surface feature well defined, we have augmented it with different surface sizes based on two existing features: property type and room type. The reason for that is simply because New York City offers a wide variety of property types that are not equal in size, value and facilities[14]. For example, a normal apartment that has 1 bedroom is around 700sqft, but a loft that has roughly the same facilities will be almost twice as big on an average. We are mentioning the fact that this is part of our assumptions, therefore it is not as accurate as of the other features from the dataset.

Similarly, we first augmented the dataset, then used the Pearson Correlation Coefficient, trained the models, optimized them and drawn conclusions based on that. The PCC did have a strong correlation with the price of a listing, 0.51, and therefore will be a contributing feature in this iteration. Table 2 better describes the metrics that we have applied and the scores that we have gotten. The models had similar results, with slight improvements. The optimized Random Forest model had an R2 score of 0.675, while the Neural Network scored 0.602 when optimized. While the R2 score is improving, we are assuming that because the dataset is not scaled that we are getting such low improvements despite having a better correlation between the price and other features.

[14] Types of Properties NYC - `https://www.eznycrealty.com/nyc-property-type`

5 Findings

Based on the analysis from Section (4), we have established that although there isn't a strong correlation between the features of the dataset and the price of an Airbnb listing from New York City, a predictive model can be trained. The best performing models were the optimized Random Forest and the optimized MLP Regressor. These scored 0.67 and 0.60 respectively. Though such scores are not encouraging at first, seeing how the performance of the MLP Regressor almost doubled when hyper tuned, we are certain that with more tuning and better processing of the data, the MLP Regressor will see improvements and outscore the Random Forest. Table .2 showcases the results of the trained models. Different metrics were used, but we have primarily focused on the R2 score which describes how close the data is to be fitted by the regression line.

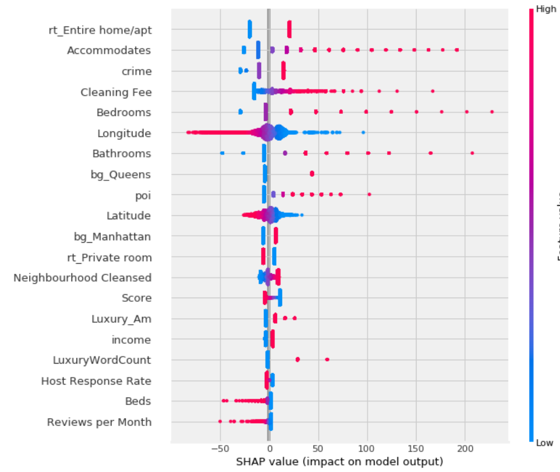

Figure 6: Impact of the attributes on the output of the Random Forest Model

Each iteration investigated the same research questions defined in Section (1.1) from different angles. This allowed us to yield better instructions and results for a machine learning price predicting model that is using as a case study the listings of Airbnb New York City.

	LR	MLP	MLPO	RF	RFO
R2	0.55	0.60	0.60	0.61	0.67
MSE	5299.6	4705.6	4651.8	4609.3	3896.5
RSE	72.8	68.6	68.2	67.9	62.4
MAE	45.6	41.63	41.6	39.6	36.3
EVS	0.55	0.62	0.61	0.61	0.67

Table 2: Metrics for the ML models that we have trained without XGBoost Regressor, and SVR. MLP means multi-layer perceptron and O stands for Optimized.

The main research question was whether the price of an Airbnb listing change when different attributes are fluctuating. We believe that the answer to that question is a definite yes; the price is mostly correlated with the number of bedrooms, bathrooms and how many people a listing can accommodate.

6 Conclusion

This paper represents a mixed-method approach to developing a price predicting machine learning model. The capabilities of a price recommendation engine in the context of the Airbnb platform are explored. With thorough research into how ML models are implemented in e-Commerce and what seems to influence a predicted price, we built a working model that has an average R2 score (0.67 for Random Forest and 0.60 for MLP Regressor).

Following up with our main research question described in Section (1.1), we have proven that the price of a listing is fluctuating based on certain features belonging to that observation. Be it through the use of PCC, or by having different iterations where a different set of features were being used to build a predictive model, the results are encouraging and further investigation can yield better results.

Q1: The crime rate is not a detrimental factor when predicting the price of a listing as there is a low correlation between the two. This claim is supported to some extent by Gilbert et al. when stating that people want to experience New York City because it embodies the American culture and diversity [12], and therefore are not concerned with how much crime an area has when booking a listing.

Q2: Touristic attractions have a negligible impact on the price and therefore ran the models without the computed points of interest and got similar results - a difference of 2-5% on the R2 score.

Q3: Zervas et al. [22] is critiquing the Airbnb review system, proving that all the listings from Airbnb that have a review, do not score lower than 3.5 stars. This alone is answering the question of whether reviews impact the price on Airbnb. Our empirical findings show that there is little to no correlation between the review score and the price of the listing, as well as the fact that there is no review score below 90% - that is for the observations that did not have a null value in the columns referring to the review score.

Q4: Fig.6 showcases how the attributes are influencing the output of the Random Forest Model. From it, the borough does not influence the predicted price. The MLP Regressor has comparable results. It does seem however that the position on the map, independent of the borough is correlated to the price and does help the price fluctuate. Further development is required to thoroughly answer this question as idealy only one field, such as zip code, would be used to determine the location of a listing.

Though there are a few obvious correlations between the attributes of the dataset and the price of an Airbnb listing, further work is needed. Section(7) describes what are the next steps for such a model to be feasible and reliable. Larger amounts of data would be needed to yield better results as well. We conclude by stating that the price of an Airbnb listing does fluctuate with the attributes of an observation.

7 Future Work

In the future, we are determined to improve the model by tuning further the parameters of the MLP Regressor. The model would be aided if a new column would be added to the dataset with the postal code, to increase the localization of a listing and increase the correlation of where a listing is situated and the actual price of it. A pipeline to process the data automatically is expected to be delivered as well. Lastly, we will be looking at other metrics as we know that a high R2 score does not denote a performant model.

Repository can be found here: `https://github.com/bgz10/DS_MS2_NYC`

References

[1] Andrew Ng. What Artificial Intelligence Can and Can't Do Right Now. Technical report, 2016.

[2] Hong Wei Yang, Zhi Geng Pan, Xi Zhao Wang, and Bing Xu. A personalized products selection assistance based on e-commerce machine learning. In *Proceedings of 2004 International Conference on Machine Learning and Cybernetics*, volume 4, pages 2629–2633, 2004.

[3] Yann Lecun, Yoshua Bengio, and Geoffrey Hinton. Deep learning, may 2015.

[4] IDA Univers. MACHINE LEARNING: BLIV KLOGERE PÅ TEKNOLOGIEN BAG KUNSTIG INTELLIGENS. *IDA Univers*, 2017.

[5] Mariusz Bojarski, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Prasoon Goyal, Lawrence D. Jackel, Mathew Monfort, Urs Muller, Jiakai Zhang, Xin Zhang, Jake Zhao, and Karol Zieba. End to End Learning for Self-Driving Cars. apr 2016.

[6] M. I. Jordan and T. M. Mitchell. Machine learning: Trends, perspectives, and prospects, jul 2015.

[7] Hannes Werthner and Francesco Ricci. E-COMMERCE AND TOURISM. Technical Report 12, 2004.

[8] Henrik Strøm, Steven Albury, and Lene Sørensen. Machine Learning Performance Metrics and Diagnostic Context in Radiology. page 6, 2018.

[9] S Cankurt and A Subasi. Developing tourism demand forecasting models using machine learning techniques with trend, seasonal, and cyclic components, 2015.

[10] E. Schöneburg. Stock price prediction using neural networks: A project report. *Neurocomputing*, 2(1):17–27, 1990.

[11] Manolis Papadakis. Predicting Airfare Prices. Technical report.

[12] David Gilbert and Claire Hancock. New York City and the transatlantic imagination: French and English tourism and the spectacle of the modern metropolis, 1893-1939. *Journal of Urban History*, 33(1):77–107, nov 2006.

[13] Visit Limsombunchai. House Price Prediction: Hedonic Price Model vs. Artificial Neural Network. Technical report.

[14] Praneeth Guggilla, Snigdha Gutha, and Goutam Chakraborty. Price Recommendation Engine for Airbnb.

[15] Stewart P. Blake. *Managing for Responsive Research and Development.* W.H.Freeman & Co Ltd, 1978.

[16] Jay Nunamaker, Minder Chen, and Titus Purdin. Systems Development in Information Systems Research.

[17] Moshe Leshno, Vladimir Ya Lin, Allan Pinkus, and Shimon Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6(6):861–867, 1993.

[18] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely Connected Convolutional Networks. Technical report, 2018.

[19] Jeffrey D. Scargle. An introduction to chaotic and random time series analysis. *International Journal of Imaging Systems and Technology*, 1(2):243–253, 1989.

[20] Jeremy Adler and Ingela Parmryd. Quantifying colocalization by correlation: The pearson correlation coefficient is superior to the Mander's overlap coefficient. *Cytometry Part A*, 77(8):733–742, aug 2010.

[21] Theodore Levantis and Azmat Gani. Tourism demand and the nuisance of crime. *International Journal of Social Economics*, 27(7-10):959–967, 2000.

[22] Georgios Zervas, Davide Proserpio, and John Byers. A First Look at Online Reputation on Airbnb, Where Every Stay is Above Average. *SSRN Electronic Journal*, apr 2018.