ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Airbnb Pricing Model

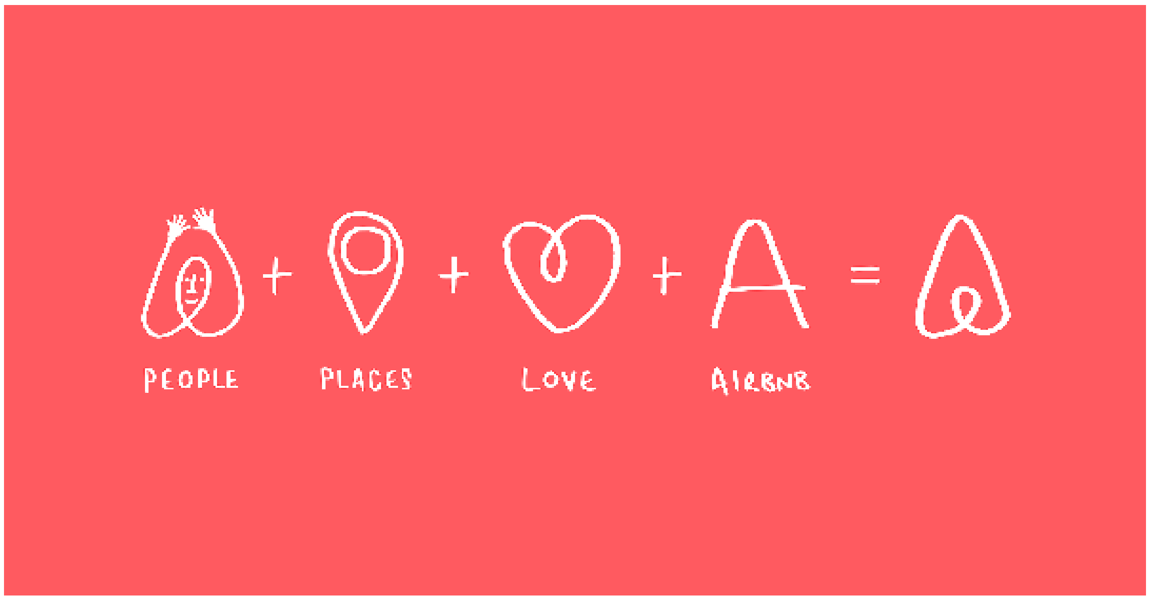
Machine Learning and Content Analytics

Final Project

Students: Ioannis Dekoulakos, Anna Makropoulou**,** Anastasios Moraitis Anastasios Safras

Academic Year 2021 - 2022

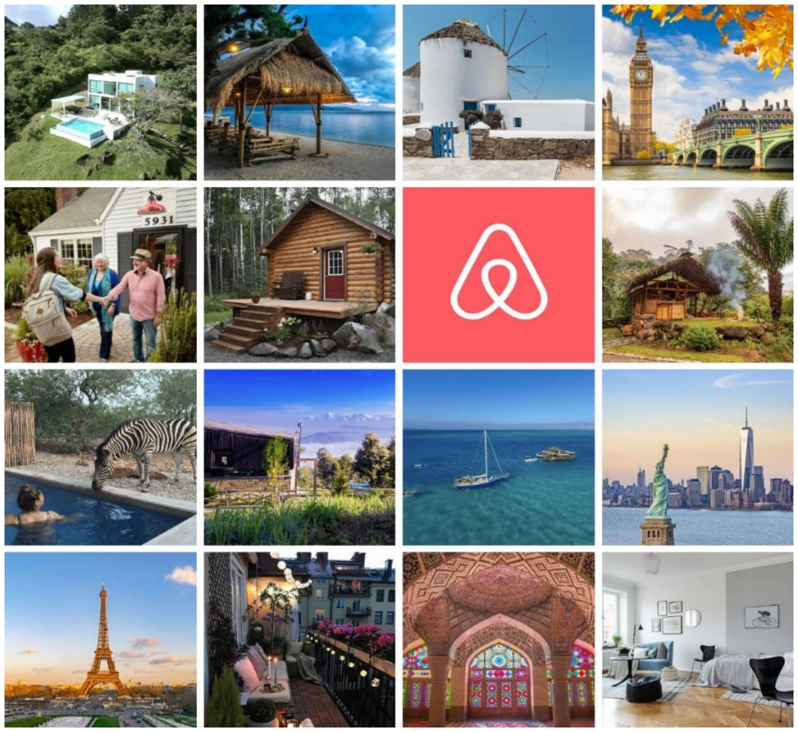
Part-Time Program



Contents

Abstract

Nowadays, many Airbnb hosts underestimate their properties and do not know how to maximize their profits. Airbnb is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. Based in San Francisco, California, the platform is accessible via website and mobile app. Airbnb does not own any of the listed properties; instead, it profits by receiving commission from each booking. The company was founded in 2008 by Brian Chesky, Nathan Blecharczyk, and Joe Gebbia. Airbnb is a shortened version of its original name, AirBedandBreakfast.com.



Airbnb is a platform where anyone can list a property for short-term rentals and at the same time individuals from across the globe, can book a property. It operates in more than 200 countries and has attracted so many fans and loyal recurrent customers that in many cases has overtaken the respective hotel industry, with red-hot growth. So, it is obvious that the competition has increased, and It has become a necessity for hosts to follow a more professional and smart management, especially in the pricing part. By observing this tendency in the market, the aim of this project is to build a service that will help Airbnb hosts get an estimate for the price of their property so that they would be able to maximize their profits or alternatively help novice hosts decide if they will eventually convert their property to an Airbnb listing, and at what price. The project focuses on the Greek market as the data collected are related solely to it.

1. Introduction

Background

Airbnb is a short-term rental service that allows everyone to list and book accommodations. Airbnb was founded in 2008 and the initial idea was that anyone that wanted extra income, could rent a room in his house to a traveler at a more competitive price than a hotel room. Later on, most users were able to choose from renting an entire place including flats, villas, beach houses etc. Within a decade, Airbnb managed to grow rapidly and now it is considered as the largest platform of accommodation around the world. In fact, Airbnb operates as of June 30 2021, in more than 220 countries around the world.
All in all, Airbnb is a web platform accessible via site and mobile application which brought huge changes in the way people can travel worldwide. It introduced a revolutionary peer-to-peer platform of accommodation, in which anyone can host guests or book a place to stay instantly. The opportunity that travelers have, to experience a stay like locals do, the direct communication with the owner and the reduced amount of money required for lodging larger spaces are only a few of the reasons why Airbnb has become a success story in the industry.

Setting the Scene

Since most of the Airbnb hosts are users with no prior expertise in the travel and real estate industry, one of the important challenges that face, is how they will ultimately decide the listing price of their rental or if their initial assumptions as far as the pricing scheme decide, is the optimal. Even though Airbnb suggests a base-line price for the property, based on some basic criteria which among them are the number of bedrooms, bathrooms and location, the final price suggested is nothing more than a low average price of other properties with similar features nearby (location-wise). As a result, the final booking price of the rental may not be the most competitive one or in many cases the most realistic one, as only a few factors are taken into consideration, with them being mostly quantifiable features like the number of bedrooms or bathrooms as mentioned above, without focusing equally on rich attributes like the view, high ceilings or the existence of fireplace etc. and text attributes like the name and description of the listing. Therefore, the need for creating a pricing-scheme leveraging the power of the text and visual attributes offered along with the rest information accompanying each rental, could:

- Improve the final price suggestions,
- upgrade the user experience
- and in many cases boost SEO for home listings and travel websites.

Our Mission

The project aims to build an easy-to-use platform leveraging our custom machine learning models built with data from the Greek Airbnb market. The purpose of this service is to offer the end-user a range of suggested pricing segments based on the criteria set as input. Such criteria are the name, description, location, the number of bedrooms, photos of the rental and other features. Consequently, any user expressing an interest to upload a listing on Airbnb can be assisted with a pricing model that will consider more criteria than the basic Airbnb's pricing model or other models in the market in order to improve or verify their initial assumptions regarding the rental's price. Moreover, the users will be benefited from this service in terms of having a clear pricing view and an estimator for the investment of their property, thus they can take into consideration more data that will help them decide whether they want to proceed with the Airbnb idea. In addition, the user would be able to choose among different pricing strategies for their listing, each one of them will result in a different forecast for a potential income. The pricing segments offered from our model *verify or suggest* an update on the price the owner sets as input. For example, for an initial price which leads to characterizing a property as expensive the model could suggest an update to the average pricing segment. Thus, the user should reconsider changing the initial assumed price to a more appropriate one. As a result, the pricing segments offered that match or need to be updated are the following:

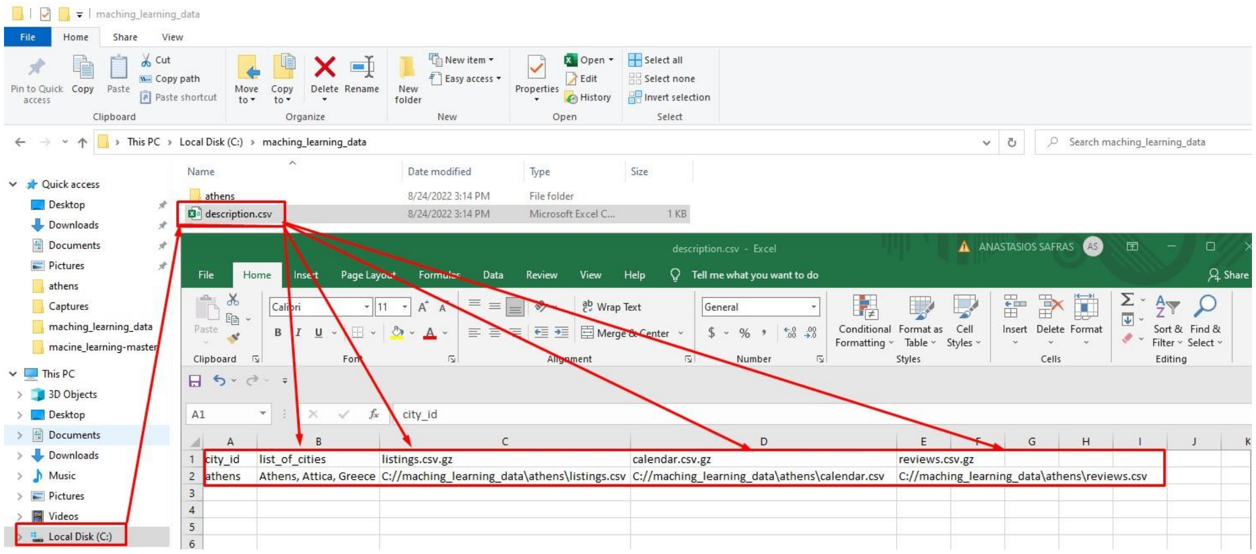
- Expensive
- Average
- Cheap

2. Data Collection

Initially the data collection was performed. The data were retrieved from the following url: http://insideairbnb.com/get-the-data.html. We downloaded data only for Athens, Greece and the columns kept refer to each property's attributes. For example, some attributes that were considered are property id, name, description, bedrooms, latitude, longitude etc. Also, the null values were defined as "unknown". The entire code defining for which attributes of each property id and for which city, data should be downloaded, can be found in the constants.py file.

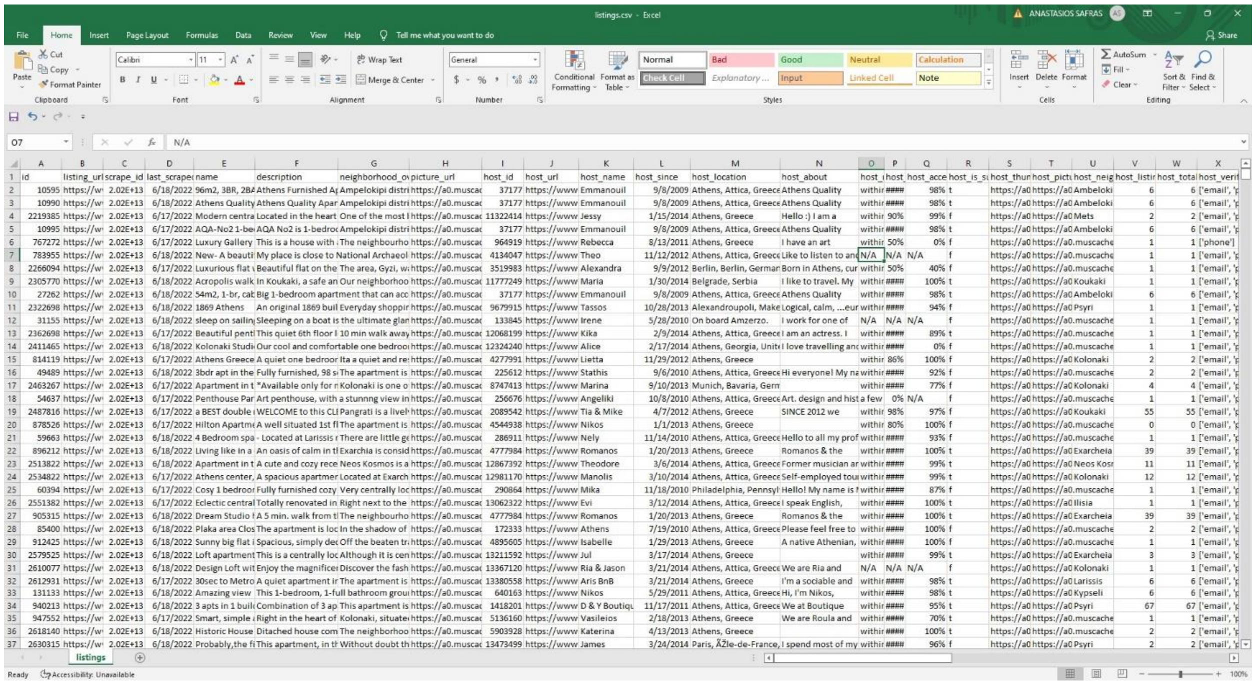
After defining in the constants.py file from were and what data should be downloaded, the data acquisition took place, in order for them to be stored in the directory contained in the constants.py file. Furthermore, when the data acquisition process is completed a csv file is created, named descriptions.csv.

The description.csv contains the columns city_id, list_of_cities, listings.csv.gz, calendar.csv.gz, reviews.csv.gz as shown below:



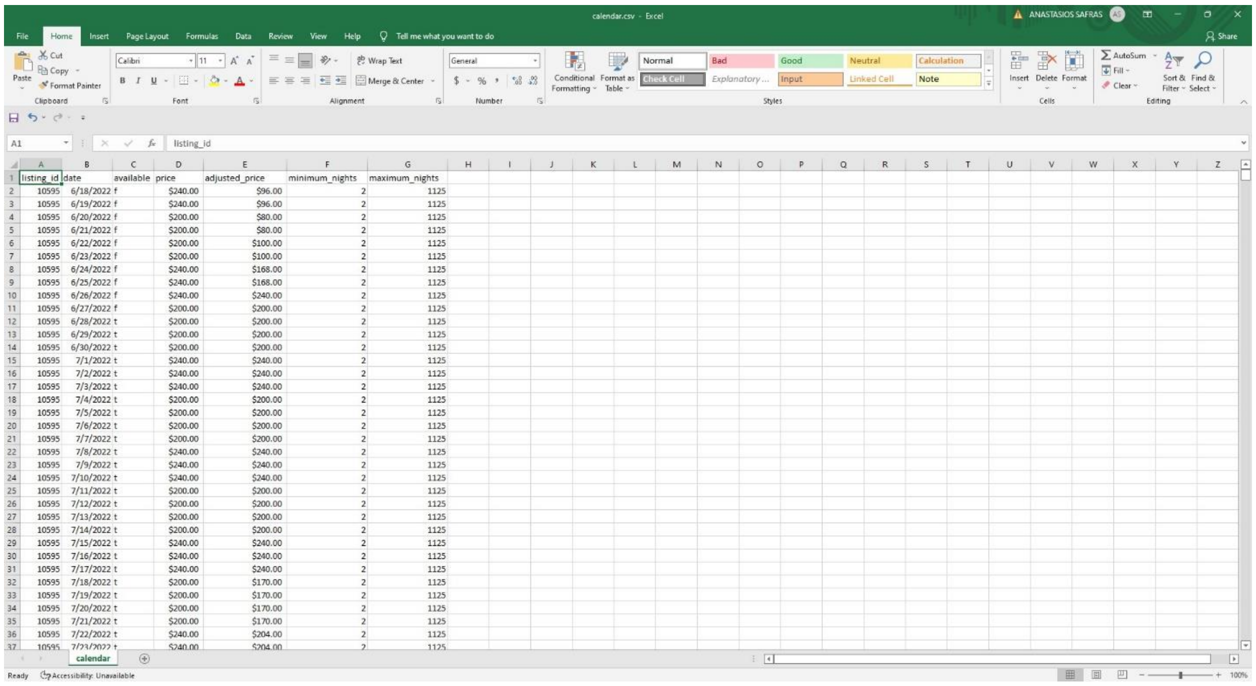
	A	B	C	D	E
1	city_id	list_of_cities	listings.csv.gz	calendar.csv.gz	reviews.csv.gz
2	athens	Athens, Attica, Greece	C://maching_learning_data\athens\listings.csv	C://maching_learning_data\athens\calendar.csv	C://maching_learning_data\athens\reviews.csv

Then in file "athens" the 3 separate csv files containing the listings, the calendar and the reviews can be found. The listings file contains information on each property, defined by a unique id , as presented below:



id	listing_url	scrape_id	last_scrape	name	description	neighborhood_o	picture_url	host_id	host_url	host_name	host_since	host_location	host_about	host_	host_	host_acce	host_is_su	host_thun	host_pictu	host_neig	host_listir	host_total	host_verif
10595	https://w	2.02E+13	6/18/2022	96m2, 3BR, 2BA	Athens Furnished Ap	Ampelokipi distri	https://a0.muscac	37177	https://www	Emmanouil	9/8/2009	Athens, Attica, Greece	Athens Quality	withir	####	98%	t	https://a0	https://a0	Ambeloki	6	6	['email', 'p
10990	https://w	2.02E+13	6/18/2022	Athens Quality	Athens Quality Apar	Ampelokipi distri	https://a0.muscac	37177	https://www	Emmanouil	9/8/2009	Athens, Attica, Greece	Athens Quality	withir	####	98%	t	https://a0	https://a0	Ambeloki	6	6	['email', 'p
2219385	https://w	2.02E+13	6/17/2022	Modern centra	Located in the heart	One of the most l	https://a0.muscac	11322414	https://www	Jessy	1/15/2014	Athens, Greece	Hello :) I am a	withir	90%	99%	f	https://a0	https://a0	Mets	2	2	['email', 'p
10995	https://w	2.02E+13	6/17/2022	AQA-No2 1-bed	AQA No2 is 1-bedroo	Ampelokipi distri	https://a0.muscac	37177	https://www	Emmanouil	9/8/2009	Athens, Attica, Greece	Athens Quality	withir	####	98%	t	https://a0	https://a0	Ambeloki	6	6	['email', 'p
767272	https://w	2.02E+13	6/17/2022	Luxury Gallery	This is a house with	The neighbourho	https://a0.muscac	964919	https://www	Rebecca	8/13/2011	Athens, Greece	I have an art	withir	50%	0%	f	https://a0	https://a0.muscache		1	1	['phone']
783955	https://w	2.02E+13	6/18/2022	New- A beauti	My place is close to	National Archaeol	https://a0.muscac	4134047	https://www	Theo	11/12/2012	Athens, Attica, Greece	Like to listen to and	N/A	N/A	N/A	f	https://a0	https://a0.muscache		1	1	['email', 'p
2266094	https://w	2.02E+13	6/17/2022	Luxurious flat v	Beautiful flat on the	The area, Gyzi, w	https://a0.muscac	3519983	https://www	Alexandra	9/9/2012	Berlin, Berlin, German	Born in Athens, cur	withir	50%	40%	f	https://a0	https://a0.muscache		1	1	['email', 'p
2305770	https://w	2.02E+13	6/18/2022	Acropolis walk	In Koukaki, a safe an	Our neighborhoo	https://a0.muscac	11777249	https://www	Maria	1/30/2014	Belgrade, Serbia	I like to travel. My	withir	####	100%	t	https://a0	https://a0	Koukaki	1	1	['email', 'p
27262	https://w	2.02E+13	6/18/2022	54m2, 1-br, cab	Big 1-bedroom apartment that can acc		https://a0.muscac	37177	https://www	Emmanouil	9/8/2009	Athens, Attica, Greece	Athens Quality	withir	####	98%	t	https://a0	https://a0	Ambeloki	6	6	['email', 'p
2322698	https://w	2.02E+13	6/18/2022	1869 Athens	An original 1869 buil	Everyday shoppir	https://a0.muscac	9679915	https://www	Tassos	10/28/2013	Alexandroupoli, Make	Logical, calm, ...eur	withir	####	94%	f	https://a0	https://a0	Psyri	1	1	['email', 'p
31155	https://w	2.02E+13	6/18/2022	sleep on sailin	Sleeping on a boat is the ultimate glan		https://a0.muscac	133845	https://www	Irene	5/28/2010	On board Amzerzo.	I work for one of	N/A	N/A	N/A	f	https://a0	https://a0.muscache		1	1	['email', 'p
2362698	https://w	2.02E+13	6/17/2022	Beautiful pent	This quiet 6th floor l	10 min walk away	https://a0.muscac	12068199	https://www	Kika	2/9/2014	Athens, Attica, Greece	I am an actress. I	withir	####	89%	t	https://a0	https://a0.muscache		1	1	['email', 'p
2411465	https://w	2.02E+13	6/18/2022	Kolonaki Studi	Our cool and comfortable one bedroo		https://a0.muscac	12324240	https://www	Alice	2/17/2014	Athens, Georgia, Unite	I love travelling and	withir	####	0%	f	https://a0	https://a0.muscache		1	1	['email', 'p
814119	https://w	2.02E+13	6/17/2022	Athens Greece	A quiet one bedroor	Ita a quiet and re	https://a0.muscac	4277991	https://www	Lietta	11/29/2012	Athens, Greece		withir	86%	100%	f	https://a0	https://a0	Kolonaki	2	2	['email', 'p
49489	https://w	2.02E+13	6/18/2022	3bdr apt in the	Fully furnished, 98 s	The apartment is	https://a0.muscac	225612	https://www	Stathis	9/6/2010	Athens, Attica, Greece	Hi everyone! My na	withir	####	92%	f	https://a0	https://a0.muscache		2	2	['email', 'p
2463267	https://w	2.02E+13	6/17/2022	Apartment in t	*Available only for n	Kolonaki is one o	https://a0.muscac	8747413	https://www	Marina	9/10/2013	Munich, Bavaria, Germ		withir	####	77%	f	https://a0	https://a0	Kolonaki	4	4	['email', 'p
54637	https://w	2.02E+13	6/17/2022	Penthouse Par	Art penthouse, with a stunnng view in		https://a0.muscac	256676	https://www	Angeliki	10/8/2010	Athens, Attica, Greece	Art. design and hist	a few	0%	N/A	f	https://a0	https://a0.muscache		1	1	['email', 'p
2487816	https://w	2.02E+13	6/17/2022	a BEST double	WELCOME to this CLI	Pangrati is a livel	https://a0.muscac	2089542	https://www	Tia & Mike	4/7/2012	Athens, Greece	SINCE 2012 we	withir	98%	97%	f	https://a0	https://a0	Koukaki	55	55	['email', 'p
878526	https://w	2.02E+13	6/17/2022	Hilton Apartme	A well situated 1st fl	The apartment is	https://a0.muscac	4544938	https://www	Nikos	1/1/2013	Athens, Greece		withir	80%	100%	f	https://a0	https://a0.muscache		0	0	['email', 'p
59663	https://w	2.02E+13	6/18/2022	4 Bedroom spa	- Located at Larissis r	There are little g	https://a0.muscac	286911	https://www	Nely	11/14/2010	Athens, Attica, Greece	Hello to all my prof	withir	####	93%	f	https://a0	https://a0.muscache		1	1	['email', 'p
896212	https://w	2.02E+13	6/18/2022	Living like in a	An oasis of calm in th	Exarchia is consid	https://a0.muscac	4777984	https://www	Romanos	1/20/2013	Athens, Greece	Romanos & the	withir	####	100%	t	https://a0	https://a0	Exarcheia	39	39	['email', 'p
2513822	https://w	2.02E+13	6/18/2022	Apartment in t	A cute and cozy rece	Neos Kosmos is a	https://a0.muscac	12867392	https://www	Theodore	3/6/2014	Athens, Attica, Greece	Former musician ar	withir	####	99%	t	https://a0	https://a0	Neos Kosr	11	11	['email', 'p
2534822	https://w	2.02E+13	6/17/2022	Athens center,	A spacious apartmer	Located at Exarch	https://a0.muscac	12981170	https://www	Manolis	3/10/2014	Athens, Attica, Greece	Self-employed tour	withir	####	99%	t	https://a0	https://a0	Kolonaki	12	12	['email', 'p
60394	https://w	2.02E+13	6/17/2022	Cosy 1 bedroor	Fully furnished cozy	Very centrally loc	https://a0.muscac	290864	https://www	Mika	11/18/2010	Philadelphia, Pennsyl	Hello! My name is I	withir	####	87%	f	https://a0	https://a0.muscache		1	1	['email', 'p
2551382	https://w	2.02E+13	6/17/2022	Eclectic central	Totally renovated in	Right next to the	https://a0.muscac	13062322	https://www	Evi	3/12/2014	Athens, Attica, Greece	I speak English,	withir	####	100%	t	https://a0	https://a0	Ilisia	1	1	['email', 'p
905315	https://w	2.02E+13	6/18/2022	Dream Studio I	A 5 min. walk from tl	The neighbourho	https://a0.muscac	4777984	https://www	Romanos	1/20/2013	Athens, Greece	Romanos & the	withir	####	100%	t	https://a0	https://a0	Exarcheia	39	39	['email', 'p
85400	https://w	2.02E+13	6/18/2022	Plaka area Clos	The apartment is loc	In the shadow of	https://a0.muscac	172333	https://www	Athens	7/19/2010	Athens, Attica, Greece	Please feel free to	withir	####	100%	f	https://a0	https://a0.muscache		2	2	['email', 'p
912425	https://w	2.02E+13	6/18/2022	Sunny big flat i	Spacious, simply dec	Off the beaten tr	https://a0.muscac	4895605	https://www	Isabelle	1/29/2013	Athens, Greece	A native Athenian,	withir	####	100%	f	https://a0	https://a0.muscache		1	1	['email', 'p
2579525	https://w	2.02E+13	6/18/2022	Loft apartment	This is a centrally loc	Although it is cen	https://a0.muscac	13211592	https://www	Jul	3/17/2014	Athens, Greece		withir	####	99%	t	https://a0	https://a0	Exarcheia	3	3	['email', 'p
2610077	https://w	2.02E+13	6/18/2022	Design Loft wit	Enjoy the magnificer	Discover the fash	https://a0.muscac	13367120	https://www	Ria & Jason	3/21/2014	Athens, Attica, Greece	We are Ria and	N/A	N/A	N/A	f	https://a0	https://a0	Kolonaki	1	1	['email', 'p
2612931	https://w	2.02E+13	6/17/2022	30sec to Metro	A quiet apartment ir	The apartment is	https://a0.muscac	13380558	https://www	Aris BnB	3/21/2014	Athens, Greece	I'm a sociable and	withir	####	98%	t	https://a0	https://a0	Larissis	6	6	['email', 'p
131133	https://w	2.02E+13	6/18/2022	Amazing view	This 1-bedroom, 1-full bathroom grou		https://a0.muscac	640163	https://www	Nikos	5/29/2011	Athens, Attica, Greece	Hi, I'm Nikos,	withir	####	98%	t	https://a0	https://a0	Kypseli	6	6	['email', 'p
940213	https://w	2.02E+13	6/18/2022	3 apts in 1 buil	Combination of 3 ap	This apartment is	https://a0.muscac	1418201	https://www	D & Y Boutiqu	11/17/2011	Athens, Attica, Greece	We at Boutique	withir	####	95%	t	https://a0	https://a0	Psyri	67	67	['email', 'p
947552	https://w	2.02E+13	6/17/2022	Smart, simple	Right in the heart of	Kolonaki, situate	https://a0.muscac	5136160	https://www	Vasileios	2/18/2013	Athens, Greece	We are Roula and	withir	####	70%	t	https://a0	https://a0.muscache		1	1	['email', 'p
2618140	https://w	2.02E+13	6/18/2022	Historic House	Ditached house com	The neighborhoo	https://a0.muscac	5903928	https://www	Katerina	4/13/2013	Athens, Greece		withir	####	100%	t	https://a0	https://a0.muscache		2	2	['email', 'p
2630315	https://w	2.02E+13	6/18/2022	Probably,the fi	This apartment, in th	Without doubt th	https://a0.muscac	13473499	https://www	James	3/24/2014	Paris, ÃŽle-de-France,	I spend most of my	withir	####	96%	f	https://a0	https://a0	Psyri	2	2	['email', 'p

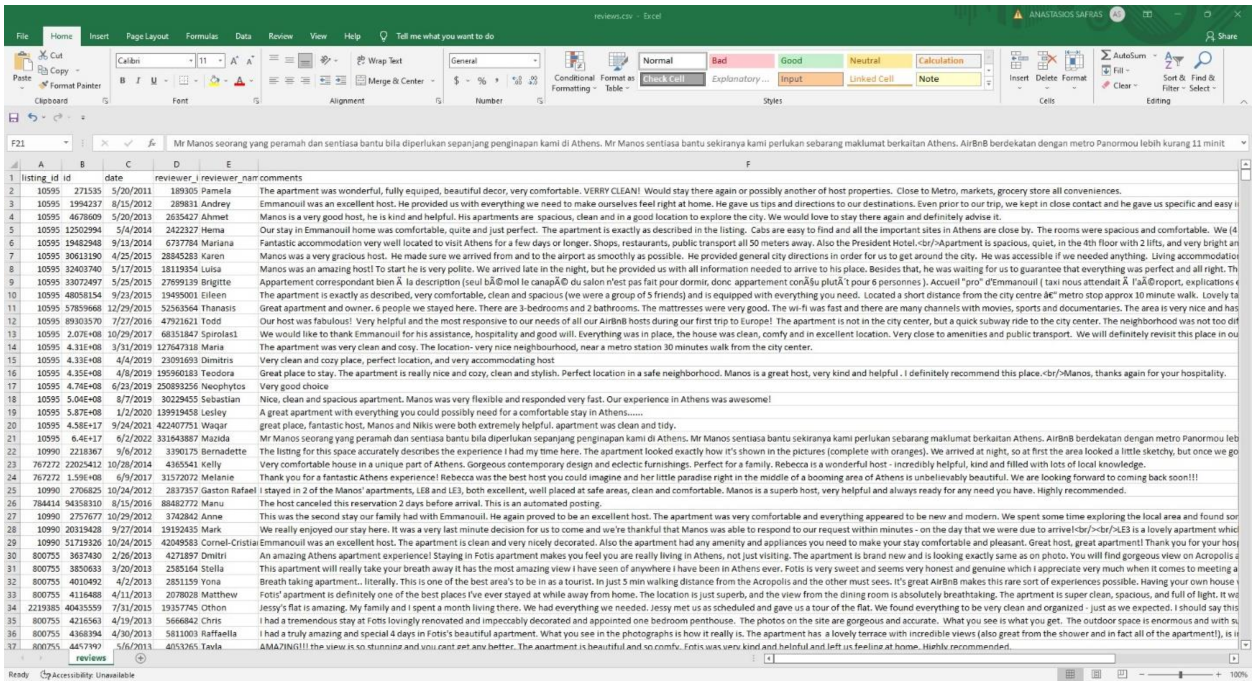
The calendar file contains daily information related to the availability of the listing, as follows:



A1 listing_id

listing_id	date	available	price	adjusted_price	minimum_nights	maximum_nights
10595	6/18/2022	f	$240.00	$96.00	2	1125
10595	6/19/2022	f	$240.00	$96.00	2	1125
10595	6/20/2022	f	$200.00	$80.00	2	1125
10595	6/21/2022	f	$200.00	$80.00	2	1125
10595	6/22/2022	f	$200.00	$100.00	2	1125
10595	6/23/2022	f	$200.00	$100.00	2	1125
10595	6/24/2022	f	$240.00	$168.00	2	1125
10595	6/25/2022	f	$240.00	$168.00	2	1125
10595	6/26/2022	f	$240.00	$240.00	2	1125
10595	6/27/2022	f	$200.00	$200.00	2	1125
10595	6/28/2022	t	$200.00	$200.00	2	1125
10595	6/29/2022	t	$200.00	$200.00	2	1125
10595	6/30/2022	t	$200.00	$200.00	2	1125
10595	7/1/2022	t	$240.00	$240.00	2	1125
10595	7/2/2022	t	$240.00	$240.00	2	1125
10595	7/3/2022	t	$240.00	$240.00	2	1125
10595	7/4/2022	t	$200.00	$200.00	2	1125
10595	7/5/2022	t	$200.00	$200.00	2	1125
10595	7/6/2022	t	$200.00	$200.00	2	1125
10595	7/7/2022	t	$200.00	$200.00	2	1125
10595	7/8/2022	t	$240.00	$240.00	2	1125
10595	7/9/2022	t	$240.00	$240.00	2	1125
10595	7/10/2022	t	$240.00	$240.00	2	1125
10595	7/11/2022	t	$200.00	$200.00	2	1125
10595	7/12/2022	t	$200.00	$200.00	2	1125
10595	7/13/2022	t	$200.00	$200.00	2	1125
10595	7/14/2022	t	$200.00	$200.00	2	1125
10595	7/15/2022	t	$240.00	$240.00	2	1125
10595	7/16/2022	t	$240.00	$240.00	2	1125
10595	7/17/2022	t	$240.00	$240.00	2	1125
10595	7/18/2022	t	$200.00	$170.00	2	1125
10595	7/19/2022	t	$200.00	$170.00	2	1125
10595	7/20/2022	t	$200.00	$170.00	2	1125
10595	7/21/2022	t	$200.00	$170.00	2	1125
10595	7/22/2022	t	$240.00	$204.00	2	1125
10595	7/23/2022	t	$240.00	$204.00	2	1125

Finally, the reviews file refers to the reviews of each listing made by past guests along with the free text of the entire review, which is presented below:



F21 Mr Manos seorang yang peramah dan sentiasa bantu bila diperlukan sepanjang penginapan kami di Athens. Mr Manos sentiasa bantu sekiranya kami perlukan sebarang maklumat berkaitan Athens. AirBnB berdekatan dengan metro Panormou lebih kurang 11 minit

listing_id	id	date	reviewer_i	reviewer_nam	comments
10595	271535	5/20/2011	189305	Pamela	The apartment was wonderful, fully equiped, beautiful decor, very comfortable. VERRY CLEAN! Would stay there again or possibly another of host properties. Close to Metro, markets, grocery store all conveniences.
10595	1994237	8/15/2012	289831	Andrey	Emmanouil was an excellent host. He provided us with everything we need to make ourselves feel right at home. He gave us tips and directions to our destinations. Even prior to our trip, we kept in close contact and he gave us specific and easy i
10595	4678609	5/20/2013	2635427	Ahmet	Manos is a very good host, he is kind and helpful. His apartments are spacious, clean and in a good location to explore the city. We would love to stay there again and definitely advise it.
10595	12502994	5/4/2014	2422327	Hema	Our stay in Emmanouil home was comfortable, quite and just perfect. The apartment is exactly as described in the listing. Cabs are easy to find and all the important sites in Athens are close by. The rooms were spacious and comfortable. We (4
10595	19482948	9/13/2014	6737784	Mariana	Fantastic accommodation very well located to visit Athens for a few days or longer. Shops, restaurants, public transport all 50 meters away. Also the President Hotel.<br/>Apartment is spacious, quiet, in the 4th floor with 2 lifts, and very bright an
10595	30613190	4/25/2015	28845283	Karen	Manos was a very gracious host. He made sure we arrived from and to the airport as smoothly as possible. He provided general city directions in order for us to get around the city. He was accessible if we needed anything. Living accommodatio
10595	32403740	5/17/2015	18119354	Luisa	Manos was an amazing host! To start he is very polite. We arrived late in the night, but he provided us with all information needed to arrive to his place. Besides that, he was waiting for us to guarantee that everything was perfect and all right. Th
10595	33072497	5/25/2015	27699139	Brigitte	Appartement correspondant bien Ã la description (seul bÃ©mol le canapÃ© du salon n'est pas fait pour dormir, donc appartement conÃ§u plutÃ´t pour 6 personnes). Accueil "pro" d'Emmanouil (taxi nous attendait Ã l'aÃ©roport, explications e
10595	48058154	9/23/2015	19495001	Eileen	The apartment is exactly as described, very comfortable, clean and spacious (we were a group of 5 friends) and is equipped with everything you need. Located a short distance from the city centre â€" metro stop approx 10 minute walk. Lovely ta
10595	57859668	12/29/2015	52563564	Thanasis	Great apartment and owner. 6 people we stayed here. There are 3-bedrooms and 2 bathrooms. The mattresses were very good. The wi-fi was fast and there are many channels with movies, sports and documentaries. The area is very nice and has
10595	89303570	7/27/2016	47921621	Todd	Our host was fabulous! Very helpful and the most responsive to our needs of all our AirBnB hosts during our first trip to Europe! The apartment is not in the city center, but a quick subway ride to the city center. The neighborhood was not too dif
10595	2.07E+08	10/29/2017	68351847	Spirolas1	We would like to thank Emmanouil for his assistance, hospitality and good will. Everything was in place, the house was clean, comfy and in excellent location. Very close to amenities and public transport. We will definitely revisit this place in ou
10595	4.31E+08	3/31/2019	127647318	Maria	The apartment was very clean and cosy. The location- very nice neighbourhood, near a metro station 30 minutes walk from the city center.
10595	4.33E+08	4/4/2019	23091693	Dimitris	Very clean and cozy place, perfect location, and very accommodating host
10595	4.35E+08	4/8/2019	195960183	Teodora	Great place to stay. The apartment is really nice and cozy, clean and stylish. Perfect location in a safe neighborhood. Manos is a great host, very kind and helpful . I definitely recommend this place.<br/>Manos, thanks again for your hospitality.
10595	4.74E+08	6/23/2019	250893256	Neophytos	Very good choice
10595	5.04E+08	8/7/2019	30229455	Sebastian	Nice, clean and spacious apartment. Manos was very flexible and responded very fast. Our experience in Athens was awesome!
10595	5.87E+08	1/2/2020	139919458	Lesley	A great apartment with everything you could possibly need for a comfortable stay in Athens......
10595	4.58E+17	9/24/2021	422407751	Waqar	great place, fantastic host, Manos and Nikis were both extremely helpful. apartment was clean and tidy.
10595	6.4E+17	6/2/2022	331643887	Mazida	Mr Manos seorang yang peramah dan sentiasa bantu bila diperlukan sepanjang penginapan kami di Athens. Mr Manos sentiasa bantu sekiranya kami perlukan sebarang maklumat berkaitan Athens. AirBnB berdekatan dengan metro Panormou leb
10990	2218367	9/6/2012	3390175	Bernadette	The listing for this space accurately describes the experience I had my time here. The apartment looked exactly how it's shown in the pictures (complete with oranges). We arrived at night, so at first the area looked a little sketchy, but once we go
767272	22025412	10/28/2014	4365541	Kelly	Very comfortable house in a unique part of Athens. Gorgeous contemporary design and eclectic furnishings. Perfect for a family. Rebecca is a wonderful host - incredibly helpful, kind and filled with lots of local knowledge.
767272	1.59E+08	6/9/2017	31572072	Melanie	Thank you for a fantastic Athens experience! Rebecca was the best host you could imagine and her little paradise right in the middle of a booming area of Athens is unbelievably beautiful. We are looking forward to coming back soon!!!
10990	2706825	10/24/2012	2837357	Gaston Rafael	I stayed in 2 of the Manos' apartments, LE8 and LE3, both excellent, well placed at safe areas, clean and comfortable. Manos is a superb host, very helpful and always ready for any need you have. Highly recommended.
784414	94358310	8/15/2016	88482772	Manu	The host canceled this reservation 2 days before arrival. This is an automated posting.
10990	2757677	10/29/2012	3742842	Anne	This was the second stay our family had with Emmanouil. He again proved to be an excellent host. The apartment was very comfortable and everything appeared to be new and modern. We spent some time exploring the local area and found sor
10990	20319428	9/27/2014	19192435	Mark	We really enjoyed our stay here. It was a very last minute decision for us to come and we're thankful that Manos was able to respond to our request within minutes - on the day that we were due to arrive!<br/><br/>LE3 is a lovely apartment whic
10990	51719326	10/24/2015	42049583	Cornel-Cristiai	Emmanouil was an excellent host. The apartment is clean and very nicely decorated. Also the apartment had any amenity and appliances you need to make your stay comfortable and pleasant. Great host, great apartment! Thank you for your hos
800755	3637430	2/26/2013	4271897	Dmitri	An amazing Athens apartment experience! Staying in Fotis apartment makes you feel you are really living in Athens, not just visiting. The apartment is brand new and is looking exactly same as on photo. You will find gorgeous view on Acropolis a
800755	3850633	3/20/2013	2585164	Stella	This apartment will really take your breath away it has the most amazing view i have seen of anywhere i have been in Athens ever. Fotis is very sweet and seems very honest and genuine which i appreciate very much when it comes to meeting a
800755	4010492	4/2/2013	2851159	Yona	Breath taking apartment.. literally. This is one of the best area's to be in as a tourist. In just 5 min walking distance from the Acropolis and the other must sees. It's great AirBnB makes this rare sort of experiences possible. Having your own house
800755	4116488	4/11/2013	2078028	Matthew	Fotis' apartment is definitely one of the best places I've ever stayed at while away from home. The location is just superb, and the view from the dining room is absolutely breathtaking. The aprtment is super clean, spacious, and full of light. It wa
2219385	40435559	7/31/2015	19357745	Othon	Jessy's flat is amazing. My family and I spent a month living there. We had everything we needed. Jessy met us as scheduled and gave us a tour of the flat. We found everything to be very clean and organized - just as we expected. I should say this
800755	4216563	4/19/2013	5666842	Chris	I had a tremendous stay at Fotis lovingly renovated and impeccably decorated and appointed one bedroom penthouse. The photos on the site are gorgeous and accurate. What you see is what you get. The outdoor space is enormous and with su
800755	4368394	4/30/2013	5811003	Raffaella	I had a truly amazing and special 4 days in Fotis's beautiful apartment. What you see in the photographs is how it really is. The apartment has a lovely terrace with incredible views (also great from the shower and in fact all of the apartment!), is i
800755	4457392	5/6/2013	4053265	Tayla	AMAZING!!! the view is so stunning and you cant get any better. The apartment is beautiful and so comfy. Fotis was very kind and helpful and left us feeling at home. Highly recommended.

3. Models

Model I: Price Prediction

Importing the attributes of Airbnb listings created a data frame which has 56 attributes and 10911 observations (columns and rows as well).

Upon importing the dataset, actions regarding changes in data types of variables, exclusion of specific columns without valuable information, as well as merging some levels of categorical variables were considered. All performed transformations will be analyzed below.

A general Overview of the Dataset, after removing the variables not taken into consideration.

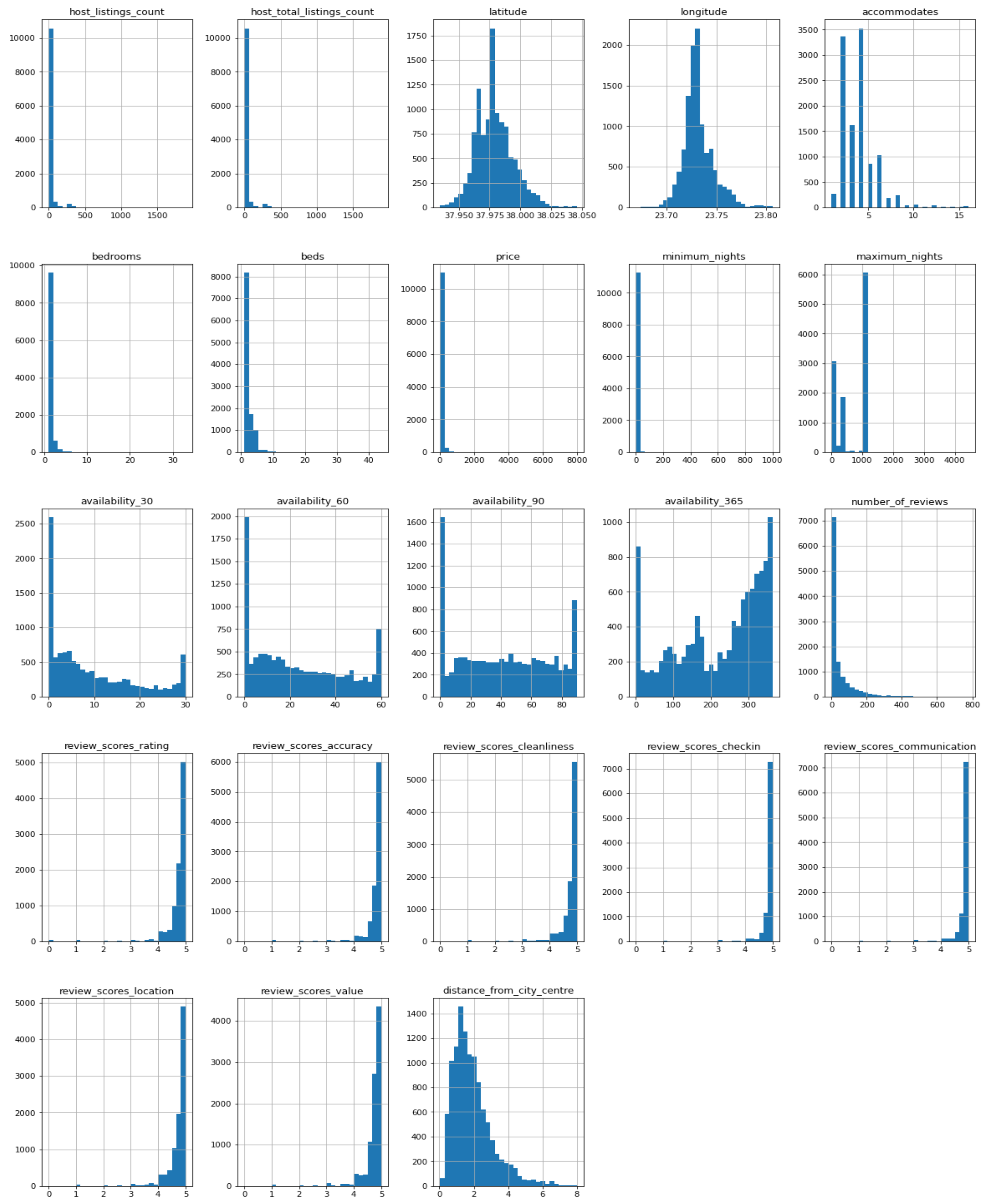


Figure 1: Histograms of all the numeric variables in the dataset

Descriptive Analysis and Data Exploration

We will keep a logical range 20-500 euros, since the extreme prices seems to be not representative for the area:

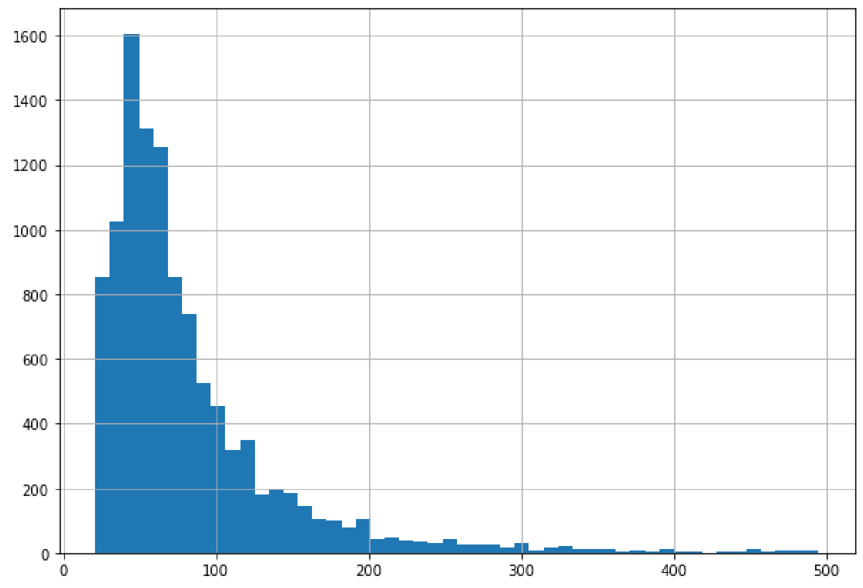


Figure 2: Histogram of Price

The Nightly advertised prices range from euros 21 to euros 495 euros. As expected, most property prices seem to be in the range 40-70 euros per day.

How have prices changed over time?

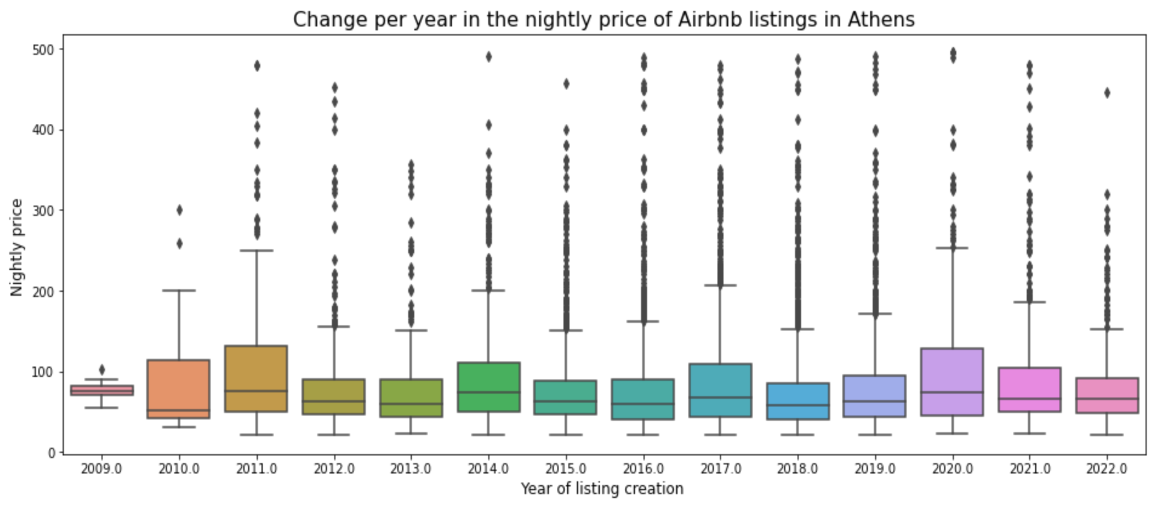


Figure 3: Box-plots for the nightly Price over the years

The average price per night for Airbnb listings in Athens seems to be higher in years 2011, 2014 & 2020. After 2020, it was expected to see a reduction in prices due to Covid-19, flights and tourism to our country decreased significantly.

In the following graphs the Average number of people accommodated are presented:

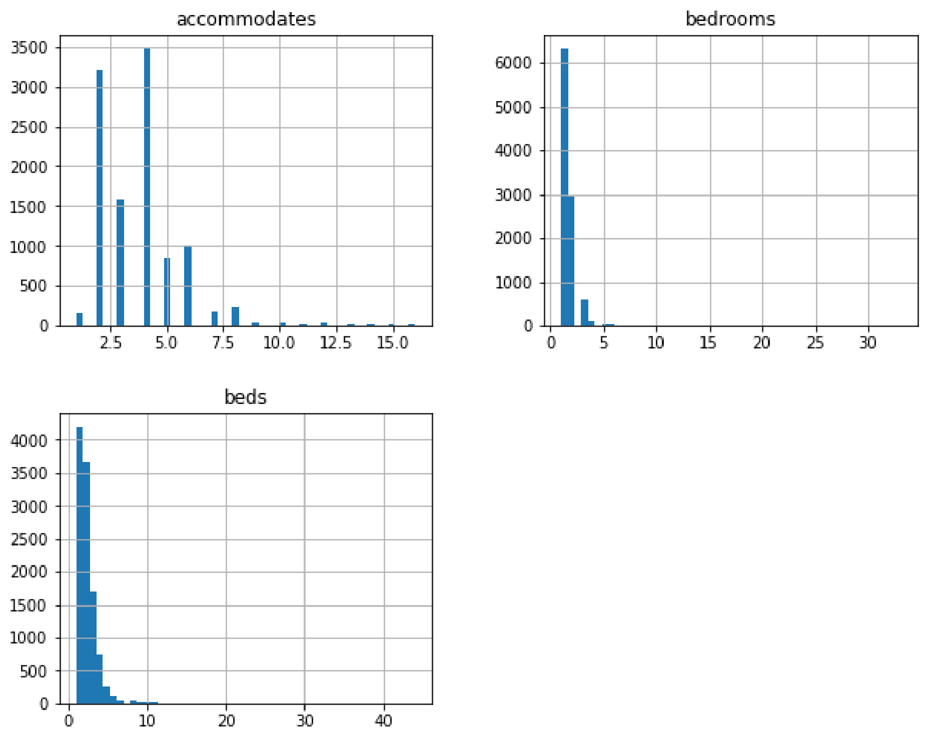


Figure 4: Bra-plots for the Average number of people accommodated

The most common property setup sleeps until 4 people and has 2 or 3 bedrooms.

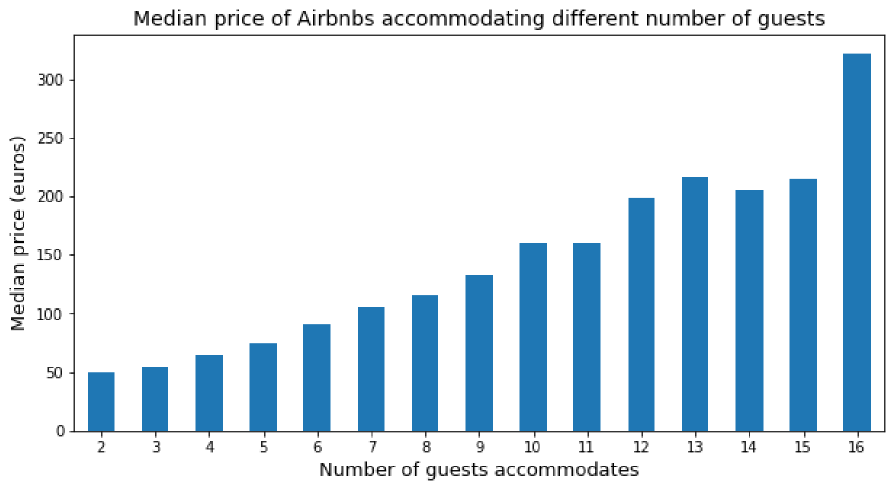


Figure 5: Median Price per number of Guests Accomodated

As expected, the more people a house accommodates, the more expensive its price is.

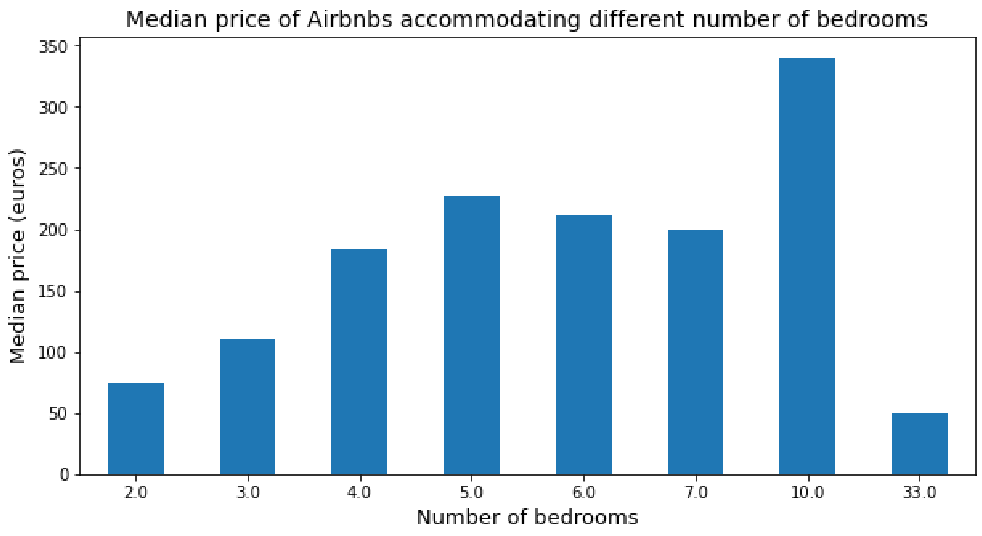


Figure 6: Median Price per Number of Bedrooms

The number of bedrooms has a significant effect on the price, it is noteworthy that a listing with 33 bedrooms has the lowest price. This makes sense, most likely this listing is a big old hotel or a hostel.

Hosts-Reviews-location

I. Hosts

How long have hosts been listing properties on Airbnb in Athens?

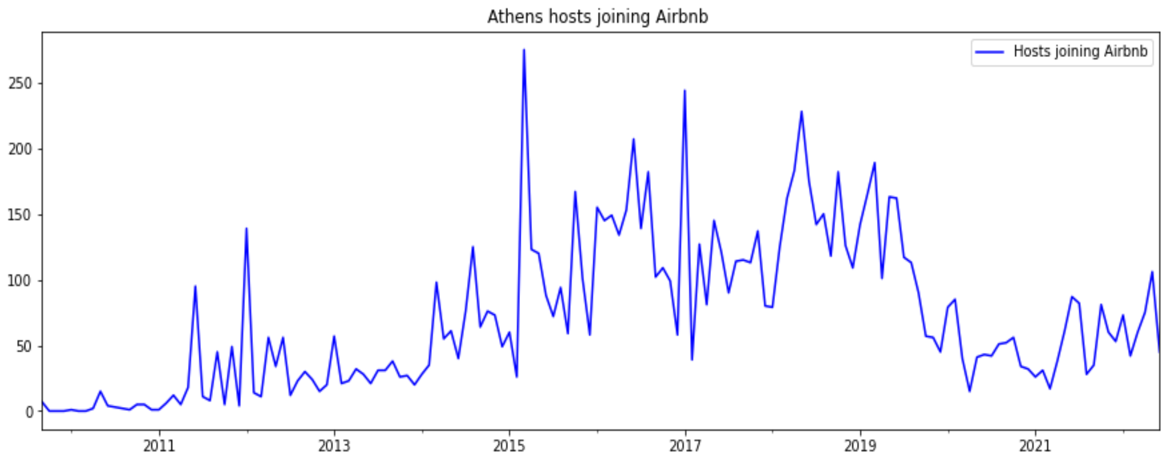


Figure 7:Yearly Evolution of Hosts Joining Airbnb

The oldest Greek listing that is currently live on Airbnb was first listed on the site around 2009. From 2011 onwards, the number of listings started increasing considerably.

We can see that there is a big peak in the number of hosts joining Airbnb in 2015. The Airbnb had become famous in Greece as well, encouraging more and more people to take advantage of their homes with a short-term rental.

II. Reviews

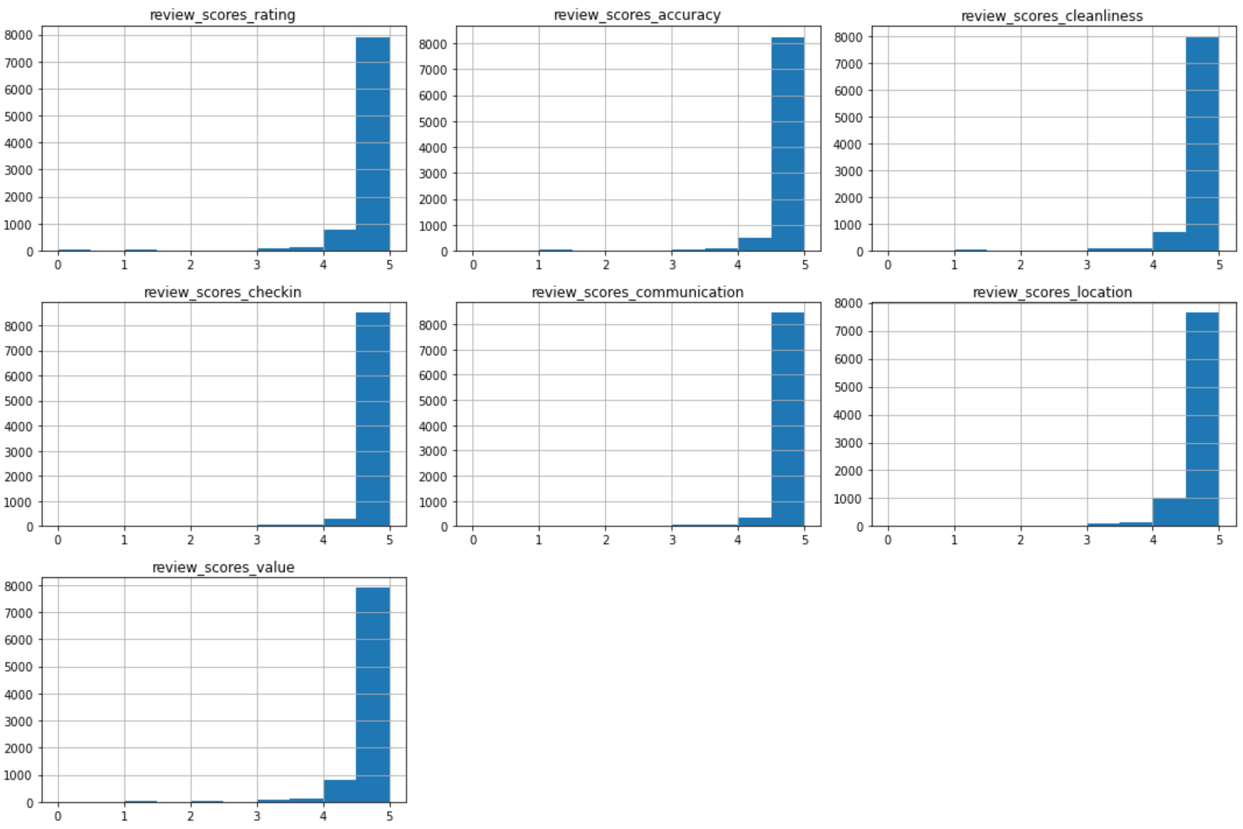


Figure 8: Reviews Score Rating

It seems that the highest percentage of scores is close to 5, something that conforms to the standards of Greek hospitality

III. Location

- the colour represents the price
- the size of the circle represents the number of reviews (so, the occupancy)

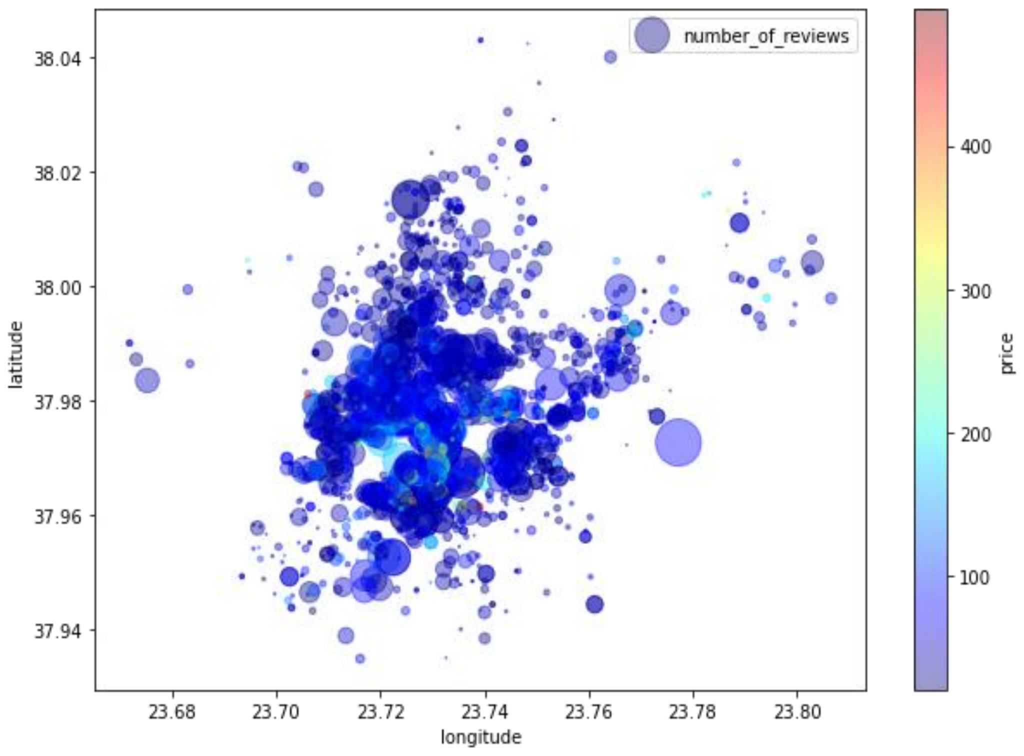


Figure 9: Volume of Reviews per Location

More central listings seem to have better occupancy at higher prices.

IV. Amenities

We kept some important variables and turned them into dummy variables:

- Breakfast
- check_in_24h
- air-conditioning
- balcony
- nature_and_views
- bed linen
- TV
- coffee_machine
- white_goods
- elevator

- parking
- outdoor_space
- hot_tub_sauna_or_pool
- internet
- pets_allowed
- Essentials
- Kitchenette
- Pool
- Self-check-in
- Stove
- *heating*

We compared the proportions of these features that are true or and the median price of each category, exploring the relationship between the category and price.

Hosts vs Super Hosts

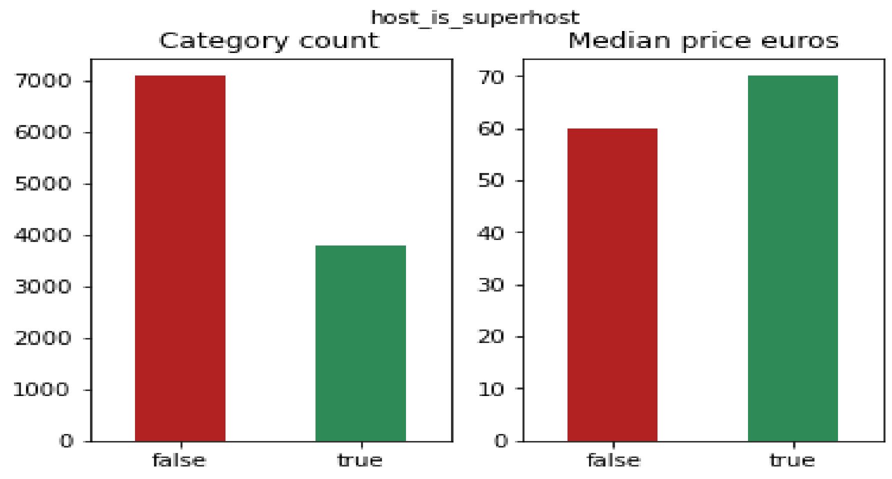


Figure 10: Hosts vs Super Hosts

About 35% of hosts are super hosts which improves the median price per night of their Airbnb listings.

Pets allowed vs Not allowed

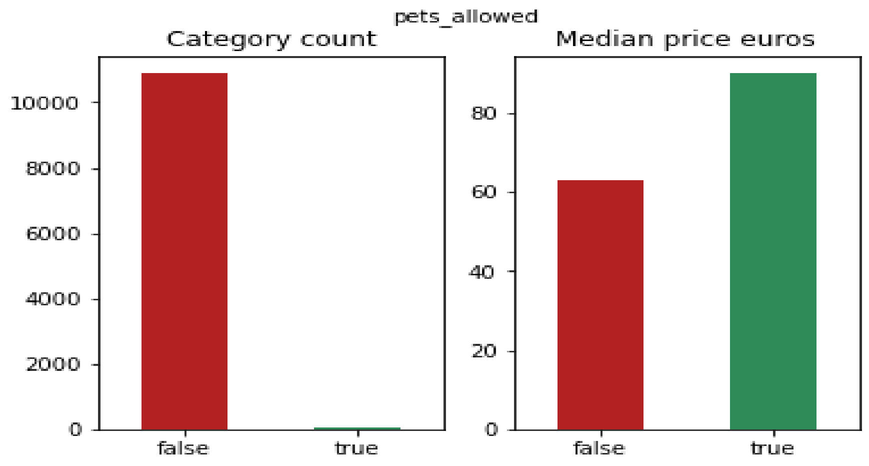


Figure 11: Pets Allowed

Only 2% of hosts allow pets, but this seems to significantly affect the price.

General Amenities

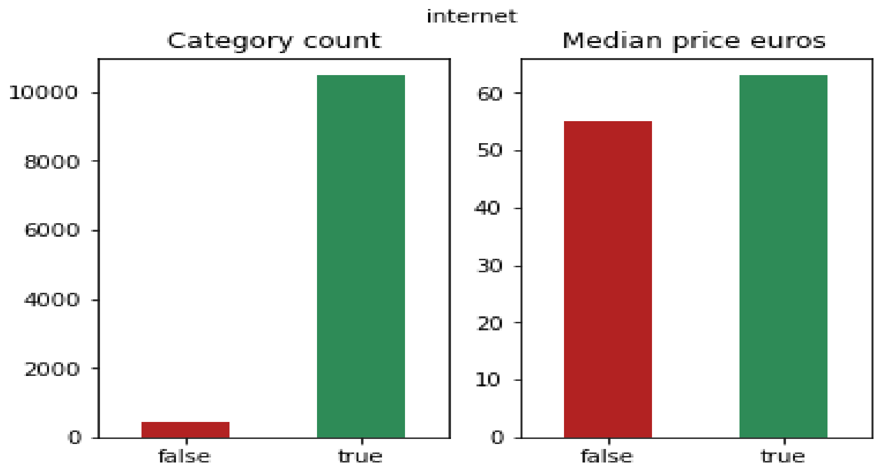


Figure 12: General Amenities

The offer of general amenities per listing does play a significant role in the variation of the median price as it does not differ notably form the median price of listings ,which do not offer general amenities.

Nature and Views

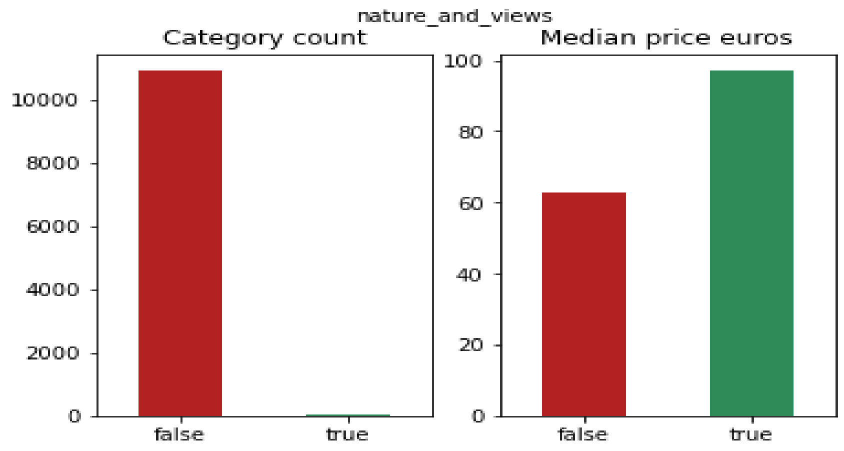


Figure 13: Nature and Views

Whether a rental offers a view or view, this does not seem to affect the median price.

Internet

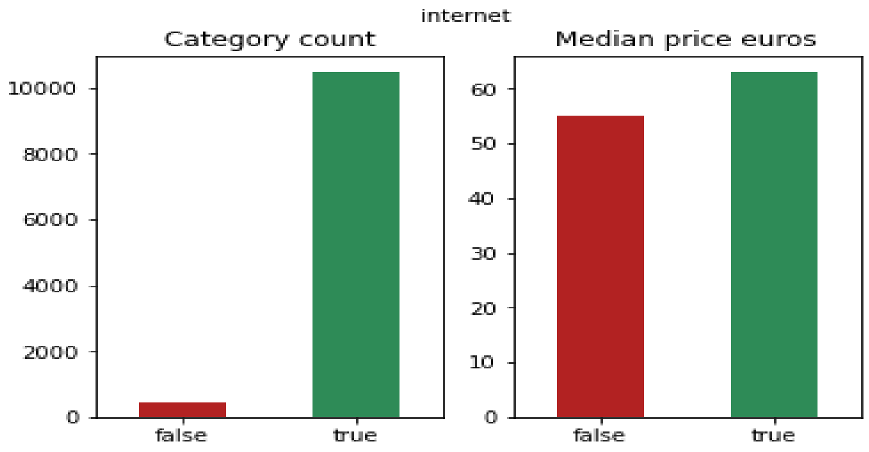


Figure 14: Internet

Only a small number of listings does include internet but similarly as the previous amenities offered this does not seem to affect the median price. The same stands for the price of rental including heating system.

Heating

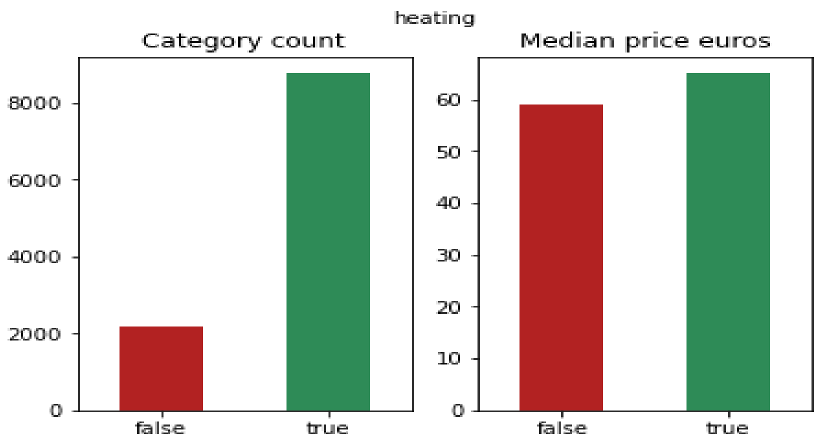


Figure 15: Heating

Methodology Set Up – Algorithm

We built three different models in order to predict the price variable.

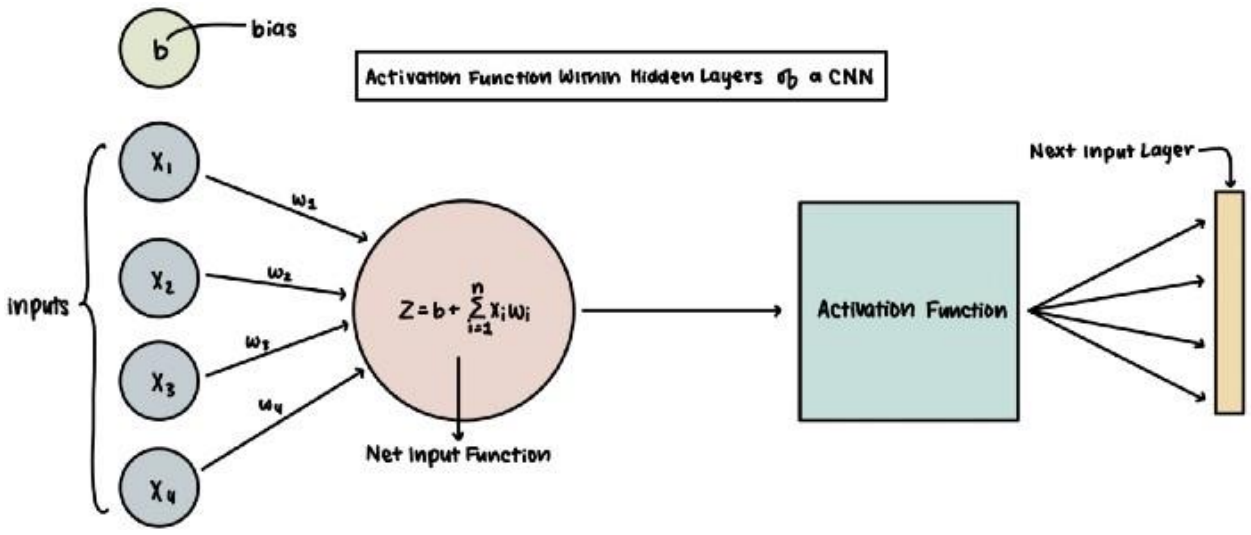
All 3 models have the same number of layers, but different hyperparameters. We distinguish those that use a different optimizer:

1. Optimizer = Adam, loss function = Mean Absolute Error, epochs = 250, batch_size = 128, Dense layers(>1 neurons) activation function: selu
2. Optimizer = SGD, loss function = Mean Absolute Error, epochs = 250, batch_size = 128
3. Optimizer = Ftrl, loss function = Mean Absolute Error, epochs = 450, batch_size = 6

Before we move forward it is important that we understand the role of activation and loss functions.

Activation Functions

Activation functions map a neuron's inputs to that of the next layer. In doing so, deep learning architectures are designed to tackle non-linear problems, which are much harder to solve than linear ones.

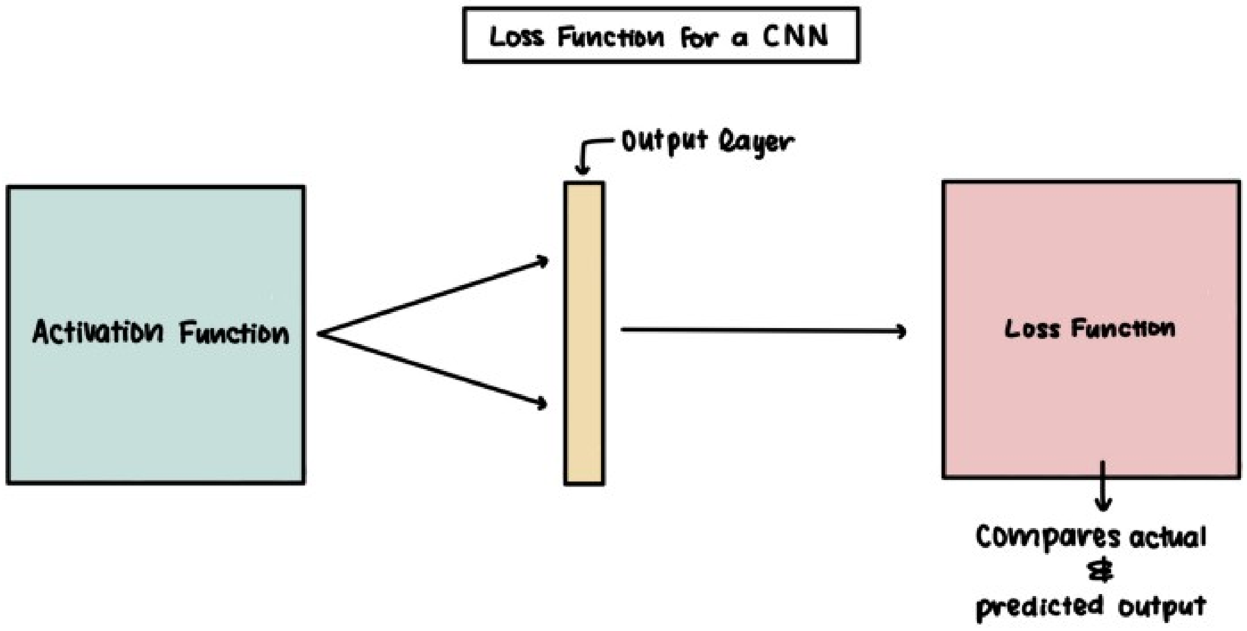


In this example, a neuron receives four inputs, each having a weight to determine its importance within the network. The net input is the summation of the input-weight products with the bias subsequently added.

Afterward comes the activation function, which will make the final call as to what gets passed onto the next input layer. While there are many different activation functions, here are some commonly used ones.

Loss Functions

Loss functions are synonymous with "cost functions" as they calculate the function's loss to determine its viability.



Above we have an activation function outputting probabilities for the output layer. By comparing the predicted values and the actual answer, we can calculate a model's accuracy and update the weights throughout our network. For example, let's say that a Convolutional Neural Network predicts that an image has a 30% chance of being a cat, 10% chance of being a frog, and 60% chance of being a horse. If the actual label for the image were a cat, then there would be a very high loss.

Choosing the best loss function is crucial, considering that it is the best indicator of whether a model is over-or-under fit.

We used MAE instead of MSE, which places a more significant penalty in comparison to MAE when the difference between the predicted and actual value increases.

1st Model Using Adam Optimizer

Adam is an optimization algorithm that can be used instead of the classical stochastic gradient descent procedure to update network weights iterative based in training data. the algorithm calculates an exponential moving average of the gradient and the squared gradient, and the parameters beta1 and beta2 control the decay rates of these moving averages.

The initial value of the moving averages and beta1 and beta2 values close to 1.0 (recommended) result in a bias of moment estimates towards zero. This bias is overcome by first calculating the biased estimates before then calculating bias-corrected estimates.

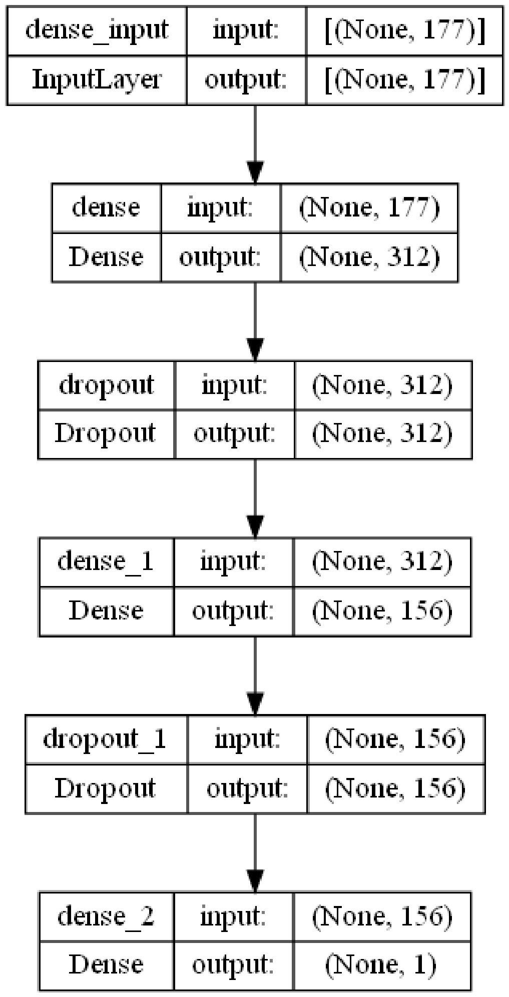
The following output of the first model contains information on the parameters, the dense layers and the dropout rates:

- **Parameters** (params): are the weights and biases that will be used for computation in all neurons of the CNN.
- **Dense Layer**: A dense layer is just a regular layer of neurons in a neural network. Each neuron receives input from all the neurons in the previous layer, thus densely connected. The layer has a weight matrix **W,** a bias vector **b,** and the activations of previous layer **a.**
- Dropout is a technique used to tackle overfitting. Dropout consists in randomly setting a fraction rate of input units to 0 at each update during training time, which helps prevent overfitting. Dropout is a regularization technique, which aims to reduce the complexity of the model with the goal to prevent overfitting.

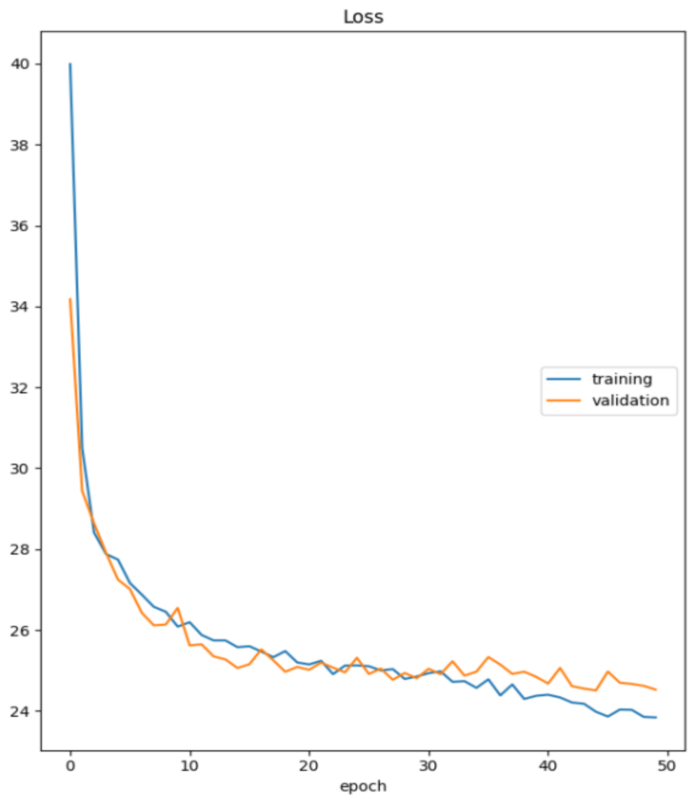
The Mean Absolute Error for the validation and test data is significantly low as shown in the output below:

```
_________________________________________________________________
Layer (type) Output Shape Param #
=================================================================
dense_17 (Dense) (None, 312) 55536

dropout_11 (Dropout) (None, 312) 0

dense_18 (Dense) (None, 156) 48828

dropout_12 (Dropout) (None, 156) 0

dense_19 (Dense) (None, 1) 157
=================================================================
Total params: 104,521

Trainable params: 104,521

Non-trainable params: 0
_________________________________________________________________
Train Data Mean Absolute Error: 22.93

Validation Data Mean Absolute Error: 24.52

Test Data Mean Absolute Error: 24.46
```

The plot of the model built using the Adam optimizer, trained for 250 epochs is the following:


dense_input | input: | [(None, 177)]
InputLayer | output: | [(None, 177)]
dense | input: | (None, 177)
Dense | output: | (None, 312)
dropout | input: | (None, 312)
Dropout | output: | (None, 312)
dense_1 | input: | (None, 312)
Dense | output: | (None, 156)
dropout_1 | input: | (None, 156)
Dropout | output: | (None, 156)
dense_2 | input: | (None, 156)
Dense | output: | (None, 1)

The loss function of the 1st model demonstrating how well it performs in the absence of overfitting problems:



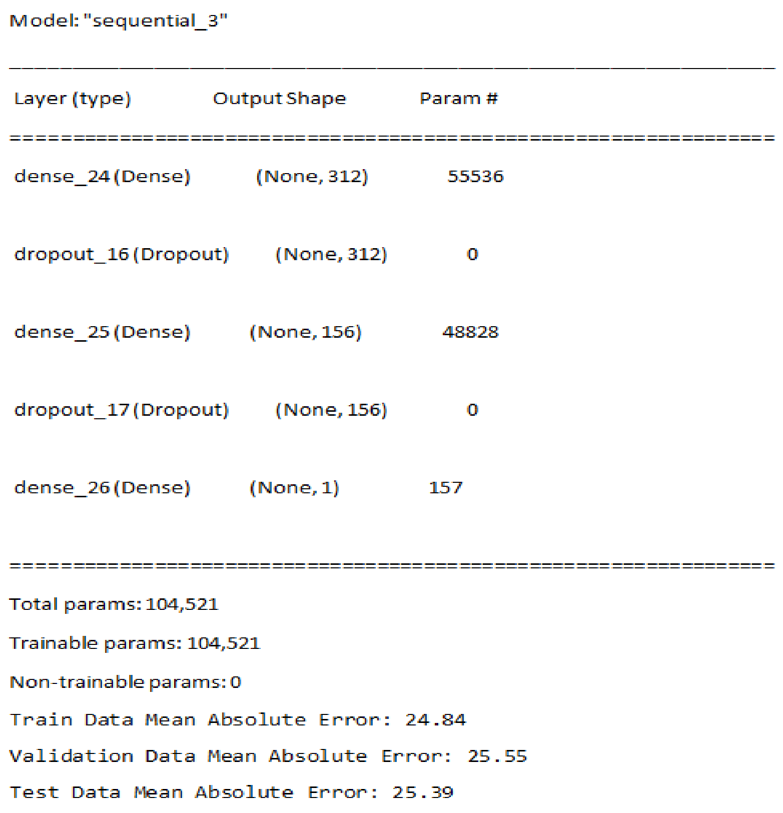
2nd Model Using SGD Optimizer

SGD algorithm is an extension of the Gradient Descent and it overcomes some of the disadvantages of the GD algorithm. Gradient Descent has a disadvantage that it requires a lot of memory to load the entire dataset of n-points at a time to compute the derivative of the loss function. In the SGD algorithm derivative is computed taking one point at a time.

The advantage of SGD is that memory requirement is less compared to the GD algorithm as the derivative is computed taking only 1 point at once. Some disadvantages are:

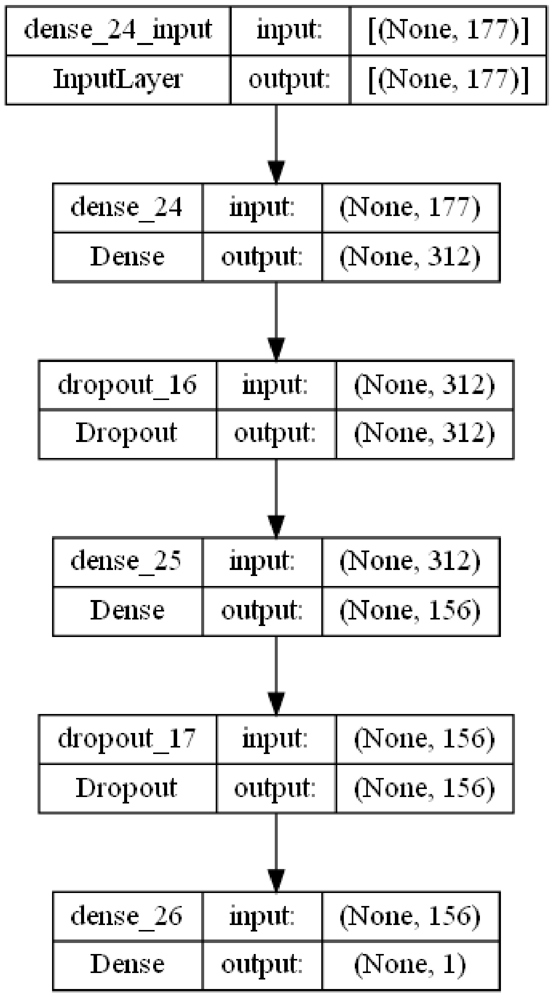
- The time required to complete 1 epoch is large compared to the GD algorithm.
- Takes a long time to converge.
- May stuck at local minima.

The model built using the SGD Optimizer performs equally well, with low MAE:

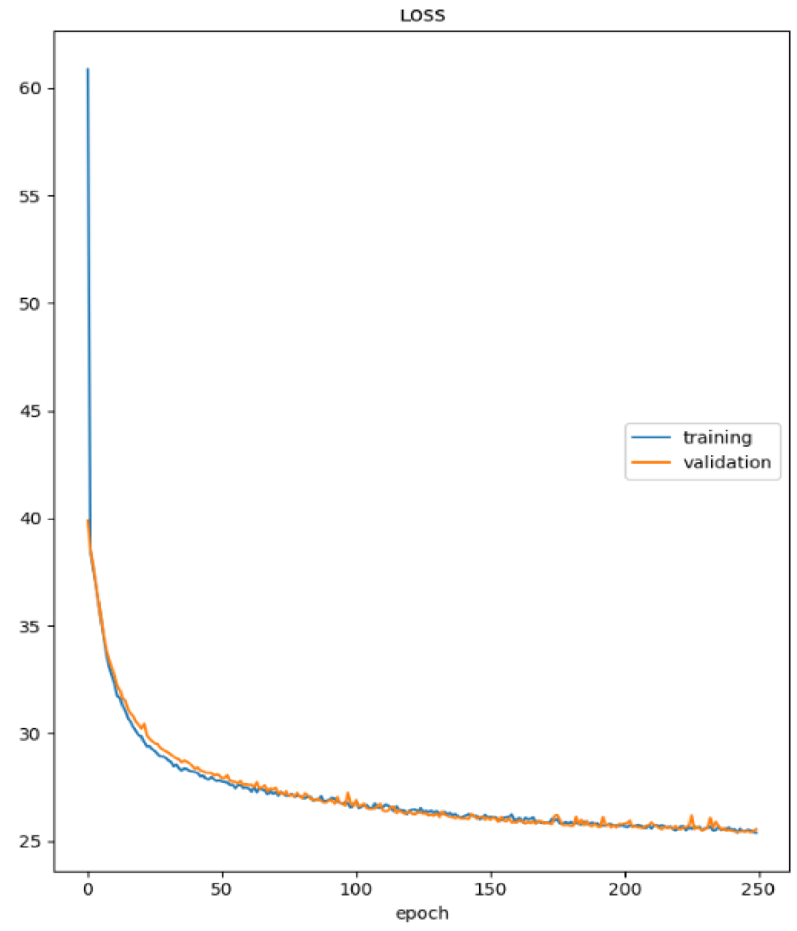


```
Model: "sequential_3"
_________________________________________________________________
 Layer (type)            Output Shape          Param #
=================================================================
 dense_24 (Dense)          (None, 312)          55536


 dropout_16 (Dropout)      (None, 312)           0


 dense_25 (Dense)         (None, 156)          48828


 dropout_17 (Dropout)      (None, 156)           0


 dense_26 (Dense)         (None, 1)           157

=================================================================
Total params: 104,521
Trainable params: 104,521
Non-trainable params: 0
Train Data Mean Absolute Error: 24.84
Validation Data Mean Absolute Error: 25.55
Test Data Mean Absolute Error: 25.39
```

The plot of the model built using the SGD optimizer, trained for 250 epochs is the following:



The loss function of the 2nd model returns very good results:



3rd Model Using FTRL Optimizer

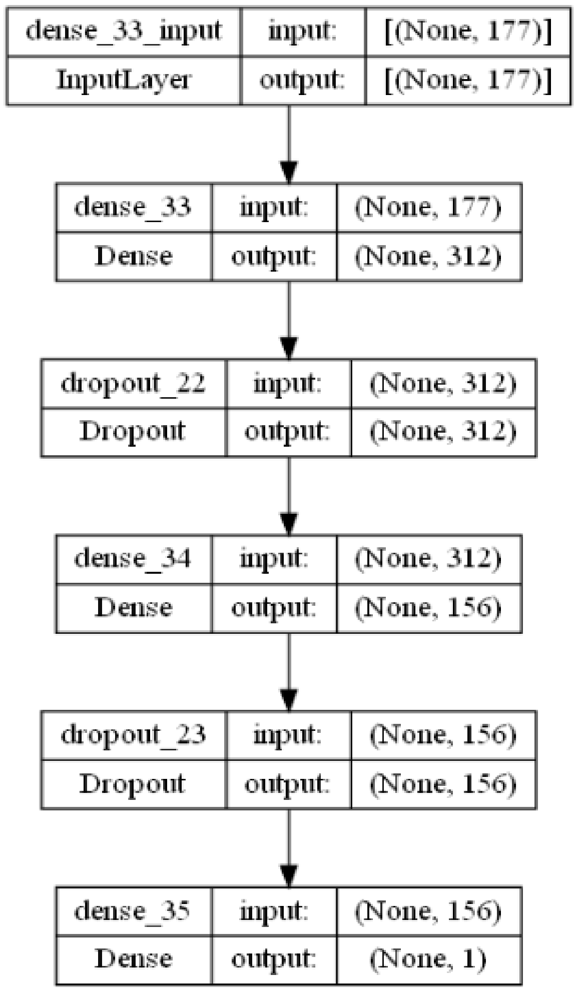
"Follow The Regularized Leader" (FTRL) is an optimization algorithm developed at Google for click-through rate prediction in the early 2010s. It is most suitable for shallow models with large and sparse feature spaces.

The model trained using the FTRL optimizer demonstrates the highest MAE, equal to 27.19 in the test dataset.

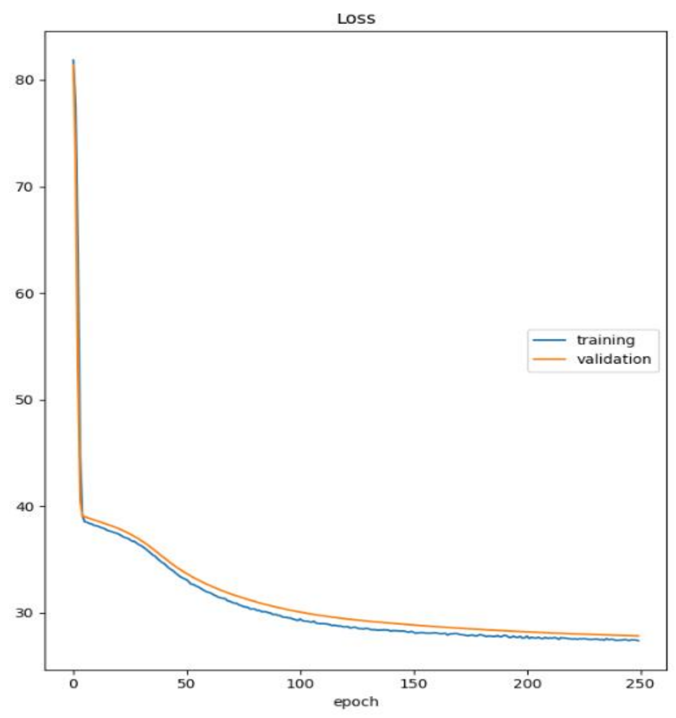
3rd model

```
Model: "sequential_11"

_________________________________________________________________

 Layer (type)              Output Shape          Param #

=================================================================

 dense_33 (Dense)          (None, 312)           55536


 dropout_22 (Dropout)      (None, 312)           0


 dense_34 (Dense)          (None, 156)           48828


 dropout_23 (Dropout)      (None, 156)           0


 dense_35 (Dense)          (None, 1)             157


=================================================================

Total params: 104,521

Trainable params: 104,521

Non-trainable params: 0

Train Data Mean Absolute Error: 27.10

Validation Data Mean Absolute Error: 27.82

Test Data Mean Absolute Error: 27.19
```

The plot of the model with FTRL:



The loss function:



Model II: Using Tensorflow to Recommend Price Adjustment

Dataset Overview

Based on the:

- name,
- description,
- neighborhood overview and
- price

attributes we built a CNN model which initially assigns the listing to a potential category derived from the quantiles of the prices. A listing based on the price set as input can be characterized as:

- Expensive
- Average
- Cheap

If the initial assumed price is "fair" then the initial pricing segment the listing belongs to is verified, since it matches the predicted pricing segment. On the contrary, in the case where the property is initially characterized as "expensive" and it is predicted as "average", the owner can consider adjusting the price based on the suggested pricing segment. Consequently, lowering the price of this particular listing could be the optimal way to proceed.

Data Pre-processing

At this step cleaning the data took place. Three core tasks were performed as far as the data cleaning is concerned:

- Firstly, the null values were handled by setting them to "unknown", as specified in the constants.py.
- Secondly, the price was fixed by replacing the "," to "." and finally normalized.
- Then NA values found in listings were removed.
- Dollar sign symbol "$" was added to define price variable.
- Also, fixed points for the range of prices were determined to determine the categories "cheap" – "average" – "expensive" and finally substitute the price column with the aforementioned categories while the predicted price retains the initial price value of the listing.

- Moreover, a new column containing the concatenation of the text attributes (name, description, neighborhood overview) was created to serve the purpose of the independent variable while price is the dependent variable.

In addition the dataset was split with a 10 to 1 ratio, to training and test.

Before proceeding with the model algorithms some further cleaning was conducted as far as the new columns with the concatenated text is concerned. More specifically, we removed:

- Emojis
- URLs
- Punctuation strings
- Translated words

After this further step of cleaning was implemented on the training and test dataset, we were able to proceed with the model's algorithms.

To conclude we have:

Initial dataset shape: 11321 x 6

Training dataset shape: 10107 x 6

Test dataset: 1125 x 6

Methodology Set Up – Algorithm

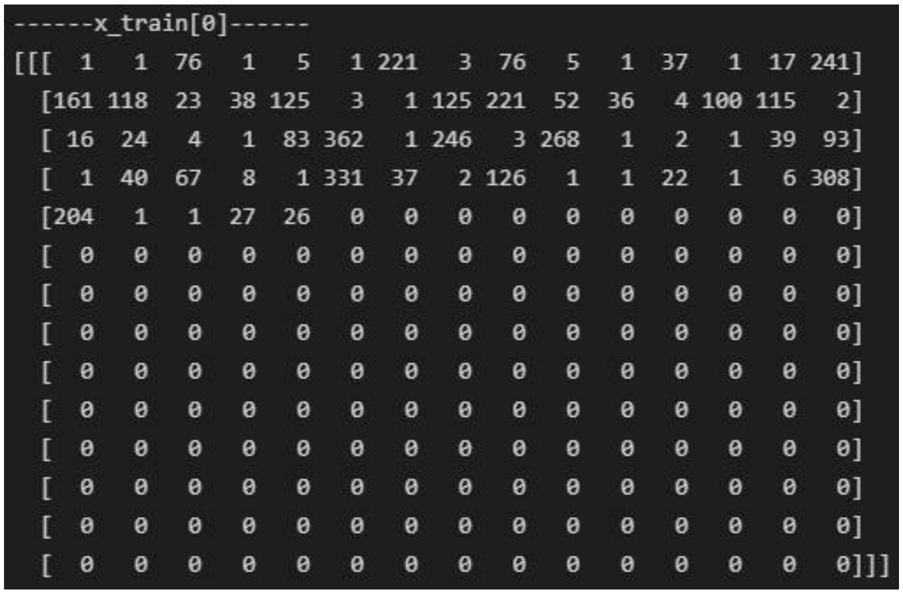
CNN is a kind of neural network; its convolutional layer differs from other neural networks. For example to perform image classification, CNN goes through every corner, vector and dimension of the pixel matrix. Convolutional layers consist of multiple features like detecting edges, corners, and multiple textures, making it a special tool for CNN to perform modeling. That layer slides across the image matrix and can detect all its features. This means each convolutional layer in the network can detect more complex features. As the feature expands, we need to expand the dimension of the convolutional layer.

We can consider text data as sequential data like data in time series. These kinds of data are represented by a one-dimensional matrix. Hence, we must work with a one-dimensional convolution layer. The idea of the model is almost the same, but the data type and dimension of convolution layers change. To work with TextCNN, we require a word embedding layer and a one-dimensional convolutional network.

Word embedding represents the density of the word vector. It is a different way to preprocess the data. This embedding can map semantically similar words. It does not consider the text as a human language but maps the structure of sets of words used in the corpus. They aim to map words into a geometric space which is called an embedding space.

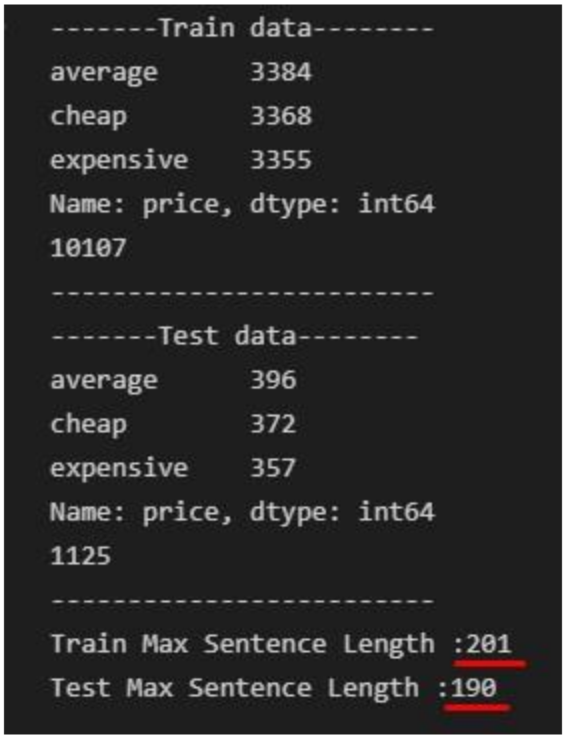
Most common words do not have a large index in our embedding space. Instead, the most uncommon words will get a higher index value which will be word count + 1 because they hold some information. Those whose occurrence is moderate will be given a moderate index value. Finally, 0 value is reserved and won't be provided to any text.

We used a tokenizer to prepare our text input for the model, after we turned our free text into a sequence we encountered a problem, each sequence has a different length of words, and to specify the length of word sequence, we need to provide a maximum length parameter and to solve this, we need to use pad_sequence(), which simply pads the sequence of words with zeros.



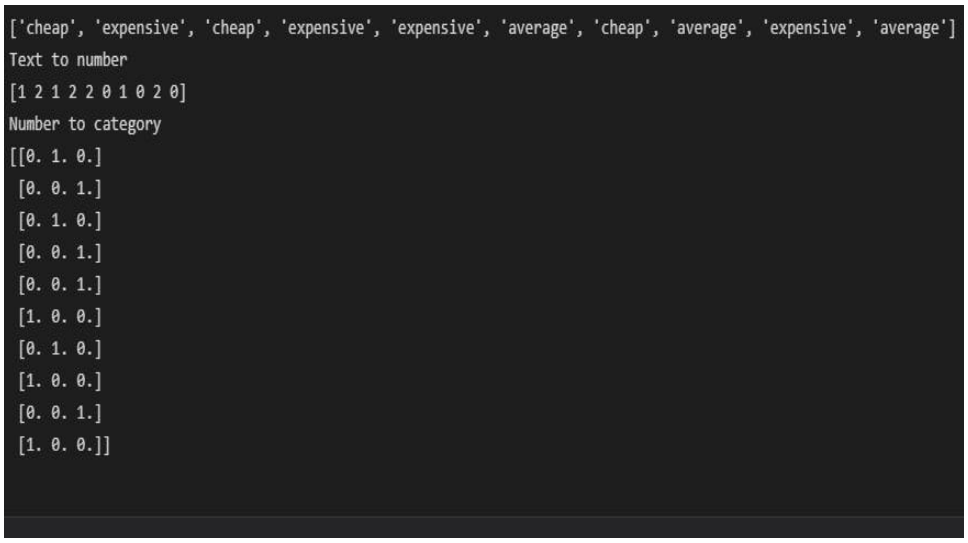
```
------x_train[0]------
[[[  1   1  76   1   5   1 221   3  76   5   1  37   1  17 241]
  [161 118  23  38 125   3   1 125 221  52  36   4 100 115   2]
  [ 16  24   4   1  83 362   1 246   3 268   1   2   1  39  93]
  [  1  40  67   8   1 331  37   2 126   1   1  22   1   6 308]
  [204   1   1  27  26   0   0   0   0   0   0   0   0   0   0]
  [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0]
  [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0]
  [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0]
  [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0]
  [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0]
  [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0]
  [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0]
  [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0]
  [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0]]]
```

As you can see the maximum length is 210, we concluded that number is safe because the maximum number of words in our text was equal to 201.



```
-------Train data--------
average      3384
cheap        3368
expensive    3355
Name: price, dtype: int64
10107
-------------------------
-------Test data--------
average      396
cheap        372
expensive    357
Name: price, dtype: int64
1125
-------------------------
Train Max Sentence Length :201
Test Max Sentence Length :190
```

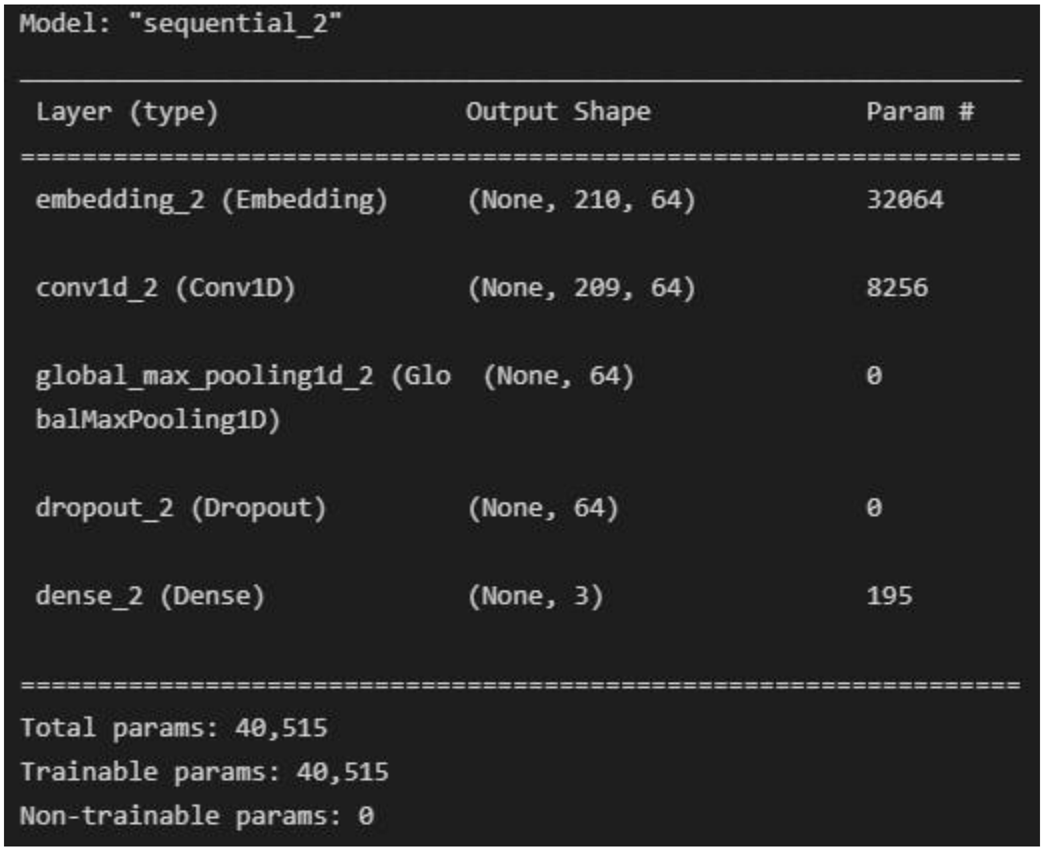
Next we transformed the text to a number and that number to a category, that way the "cheap" label was represented by the number 1 and the category was { [0],[1],[0] }, the "expensive" label with the number 2 and category { [0],[0],[1] } and last the "average" label with the number 0 and the category { [1],[0],[0] } as you can see below the first 10 rows.



```
['cheap', 'expensive', 'cheap', 'expensive', 'expensive', 'average', 'cheap', 'average', 'expensive', 'average']
Text to number
[1 2 1 2 2 0 1 0 2 0]
Number to category
[[0. 1. 0.]
 [0. 0. 1.]
 [0. 1. 0.]
 [0. 0. 1.]
 [0. 0. 1.]
 [1. 0. 0.]
 [0. 1. 0.]
 [1. 0. 0.]
 [0. 0. 1.]
 [1. 0. 0.]]
```

Keras is a high-level python deep learning API that's easier for ML beginners, as well as researchers. It is integrated as part of Tensorflow 2.0. To build the we used the Sequential model A Sequential model is appropriate for a plain stack of layers where each layer has exactly one input tensor and one output tensor (source: https://www.tensorflow.org/guide/keras/sequential_model).

The model created has an embedding dimension of 64 and the maximum sequence length is 210. For the convolution layer we use 64 filters and we looked at 2 words at the time, the convolution layers also used the ReLu activation function. For the output layer we used 3 neurons with sigmoid as activation, because of the small number of available data we encountered heavy overfitting, which we tried to reduce by adding l2 regularizers, dropouts. The model's summary is shown below:



```
Model: "sequential_2"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding_2 (Embedding)     (None, 210, 64)           32064

 conv1d_2 (Conv1D)           (None, 209, 64)           8256

 global_max_pooling1d_2 (Glo  (None, 64)               0
 balMaxPooling1D)

 dropout_2 (Dropout)         (None, 64)                0

 dense_2 (Dense)             (None, 3)                 195

=================================================================
Total params: 40,515
Trainable params: 40,515
Non-trainable params: 0
```

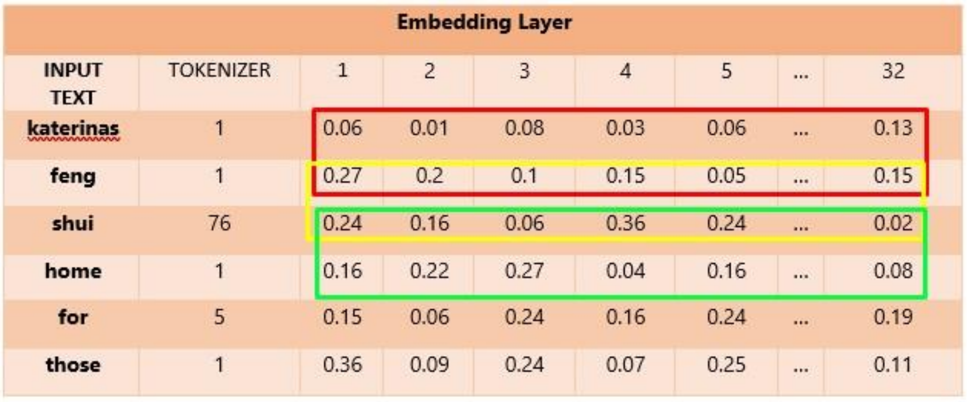
In order to better understand the shape of the output we can have a look at the following representation:

For example the first text from the training text is the following: "katerinas feng shui home for those who love…". Let's take the first 6 words.

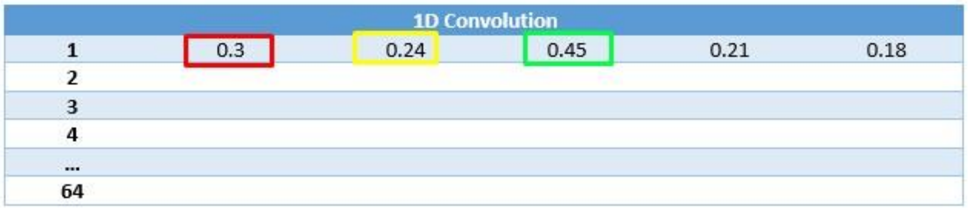
INPUT TEXT	katerinas	feng	shui	home	for	those

INPUT TEXT	**katerinas**	**feng**	**shui**	**home**	**for**	**those**
TOKENIZER	1	1	76	1	5	1

Embedding Layer								
INPUT TEXT	TOKENIZER	1	2	3	4	5	...	32
katerinas	1	0.06	0.01	0.08	0.03	0.06	...	0.13
feng	1	0.27	0.2	0.1	0.15	0.05	...	0.15
shui	76	0.24	0.16	0.06	0.36	0.24	...	0.02
home	1	0.16	0.22	0.27	0.04	0.16	...	0.08
for	5	0.15	0.06	0.24	0.16	0.24	...	0.19
those	1	0.36	0.09	0.24	0.07	0.25	...	0.11



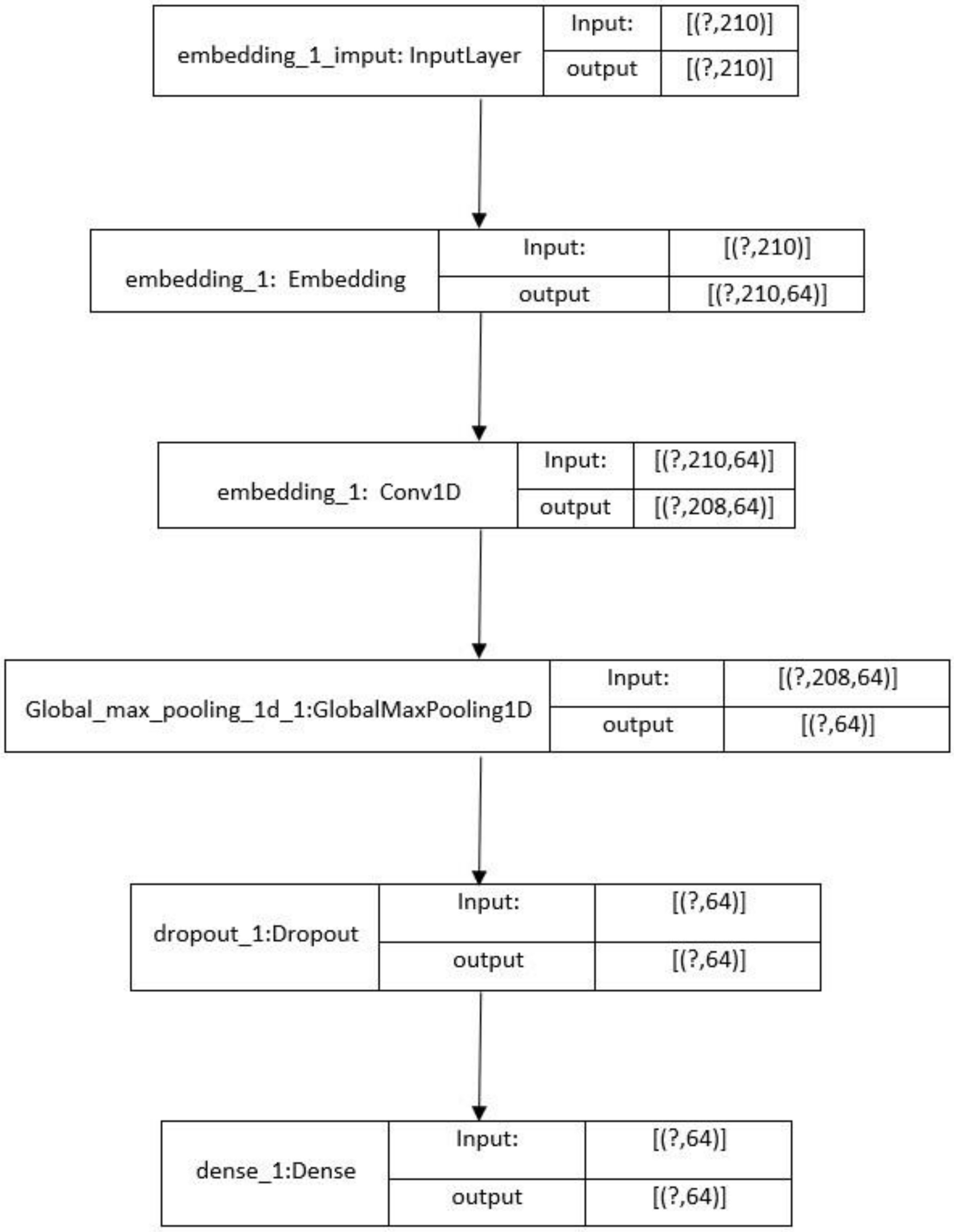
Embedding Layer								
INPUT TEXT	TOKENIZER	1	2	3	4	5	...	32
katerinas	1	0.06	0.01	0.08	0.03	0.06	...	0.13
feng	1	0.27	0.2	0.1	0.15	0.05	...	0.15
shui	76	0.24	0.16	0.06	0.36	0.24	...	0.02
home	1	0.16	0.22	0.27	0.04	0.16	...	0.08
for	5	0.15	0.06	0.24	0.16	0.24	...	0.19
those	1	0.36	0.09	0.24	0.07	0.25	...	0.11



1D Convolution					
1	0.3	0.24	0.45	0.21	0.18
2					
3					
4					
...					
64					

	1D Convolution					Global Max Pooling
1	0.3	0.24	0.35	0.21	0.18	0.35
2	0.16	0.42	0.27	0.36	0.19	0.42
3	0.22	0.28	0.24	0.16	0.45	0.45
4	0.27	0.22	0.02	0.22	0.37	0.37
…	0.24	0.55	0.08	0.21	0.38	0.55
64	0.15	0.36	0.19	0.57	0.29	0.57

The plot of the final model, where the batch size in our case was 256:

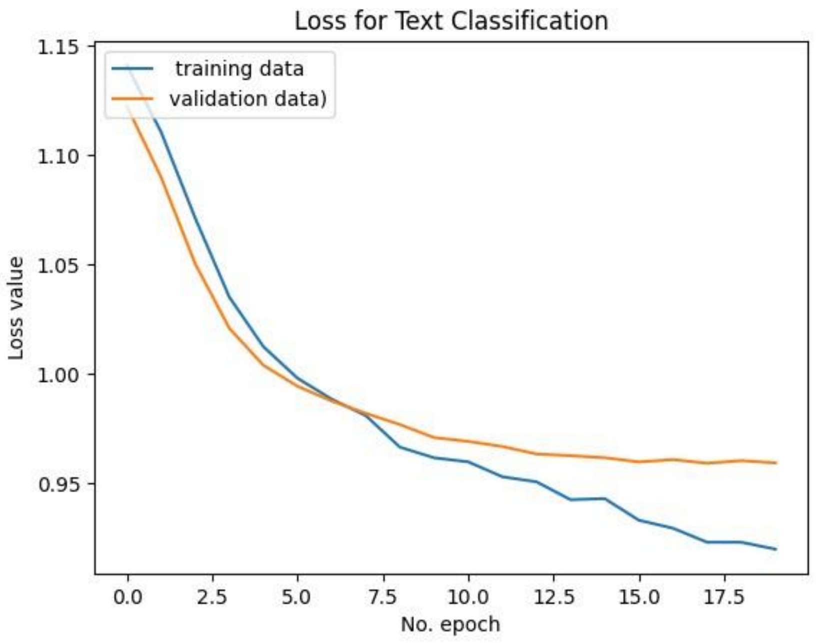

embedding_1_imput: InputLayer
Input: [(?,210)]
output [(?,210)]
embedding_1: Embedding
Input: [(?,210)]
output [(?,210,64)]
embedding_1: Conv1D
Input: [(?,210,64)]
output [(?,208,64)]
Global_max_pooling_1d_1:GlobalMaxPooling1D
Input: [(?,208,64)]
output [(?,64)]
dropout_1:Dropout
Input: [(?,64)]
output [(?,64)]
dense_1:Dense
Input: [(?,64)]
output [(?,64)]

The model was trained for 20 epochs (We run it for 100 epochs with minimum change on the results) and the results arise are encouraging if we consider the small dataset used. At first, let's explain the metrics to be evaluated:

- ***Precision***: The number of instances that are relevant, out of the total instances the model retrieved.
- ***Recall***: The number of instances which the model correctly identified as relevant out of the total relevant instances.

Categories	Precision	Recall	f1-score	Support
cheap	0.56	0.66	0.61	372
average	0.49	0.36	0.41	396
expensive	0.62	0.69	0.65	357

For example, for the expensive category we have 62% precision, meaning that out of 100 listings in total 62 were predicted correctly. Also, recall is 62% indicating that out of 100 flagged as expensive, 62 were identified correctly.
We reduced the problem with the overfitting but it was still there, at a lower scale as we can see from the graph below.



After seeing that the model still has an issue with overfitting, we decided to try using an LSTM model. Nothing changed in terms of the labels or the sequence of the text with the help of the tokenizer.

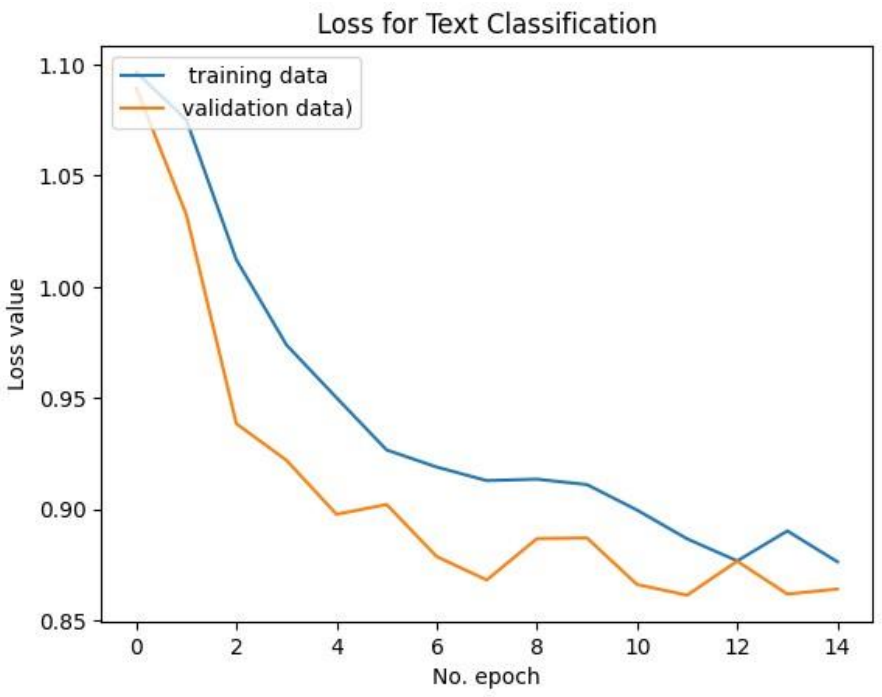
LSTM Modeling

- Vectorize listings text, by turning each text into a sequence.
- Limit the data set to the top 500 words.
- Set the max number of words in each complaint at 210.
- First layer is the embedded layer that uses 64 length vectors.
- Variational dropout is set to 0.6
- Dropout and recurrent dropout was set to 0.6
- Next is the LSTM layer with 64 memory units
- Output is the layer that gives us 3 values same as the first model.
- As an activation function, we used softmax.
- We run 15 epochs with a batch size or 256.

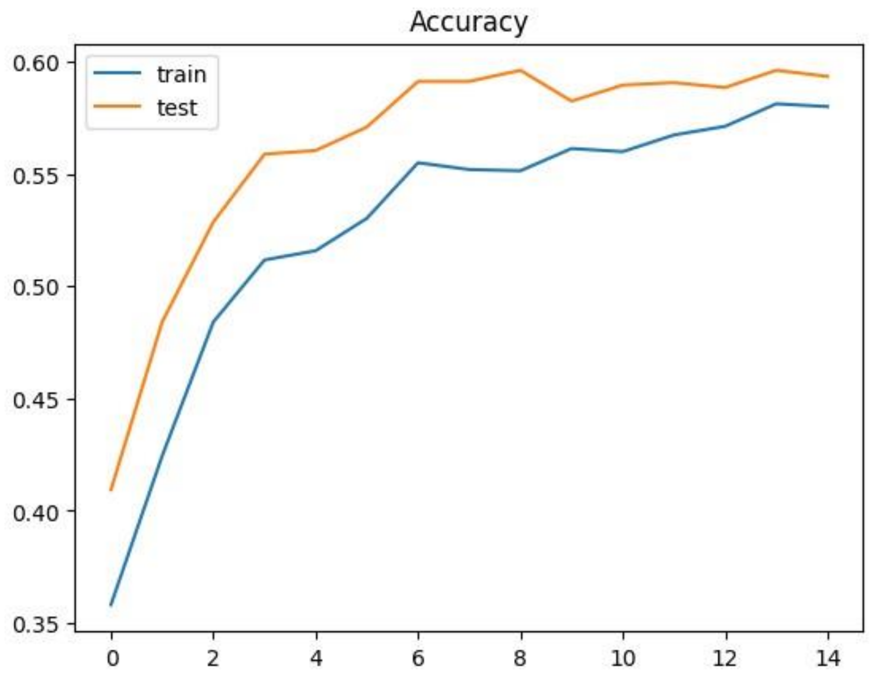
After the model trained for 15 epochs the results, we almost made overfitting go extinct in the expense of prediction. For us it was worth the tradeoff because we managed to solve our biggest problem which was the overfitting because of the small dataset we used. Let's see the new models results:

Categories	Precision	Recall	f1-score	Support
cheap	0.53	0.79	0.63	372
average	0.47	0.19	0.27	396
expensive	0.61	0.69	0.65	357

We can compare the loss values from the graph below and the previous graph and see the reduction of overfitting.



We could also check an accuracy plot for the LSTM model.



4. Business Value and Further Discussion

AI/ML solutions model human reasoning by allowing the computer to make judgement calls (“inference”) based on past positive or negative outcomes. Just as for human reasoning, the result remains an educated guess. Building more sophisticated models can increase accuracy, but the business value of that accuracy needs to be weighed against the cost of building and training the model. As a result, although we are trying to leverage the power of machine learning models to gain insights, we should always keep in mind the business value these models bring.

It is well-known that explaining an ML model to an audience with no experience or expertise in this field can be a complicated and difficult procedure. Also, there is always the risk of overinflating expectations when using terms as Machine Learning and Artificial Intelligence. Consequently, our aim is to focus on the improvements such models bring in the industry we focused.

There is no doubt that our models need improvements and enhancements in order to maximize the value brought to our potential customers, but at the same time we try to keep in mind the business aspect of the project, which is to keep our clients happy with results. Thus, when building the models our goal is not only to achieve the highest accuracy, precision or recall KPI values but spot the turning point where we can offer our client a tool that can reduce the amount of time spent to decide a price for their listing and help him/her improve their initial assessments, while highlighting the value of right decisions vs the cost of wrong ones.

In Model I section, 3 models predicting the price of the listing were developed. The business value brought by such models is great. The client can submit the rental and by adding its features, he/she can hustle-free obtain a suggested price. This might sound simple but in reality it saves a lot of time and effort from a the time-consuming process of finding listings with similar attributes and characteristics in order to decide the final price. In addition, the advantage of getting a more competitive price should be highlighted, since this is something that in the long run will maximize the profits of the client.

As far as the ModelI is concerned, without getting into details on the development of the algorithm, what we tried to achieve is quite straightforward and the possible value gained by the customer is notable. Also, it is easy to use and extra thought in understanding complicated terms or calculating formulas is requires. The customer has simply to consider adjusting the initial input price based on the suggestion made. Furthermore, it is considerably trouble-free to explain how it works. More specifically by leveraging text attributes of listings like the name, the description overview, the neighborhood overview and the price.

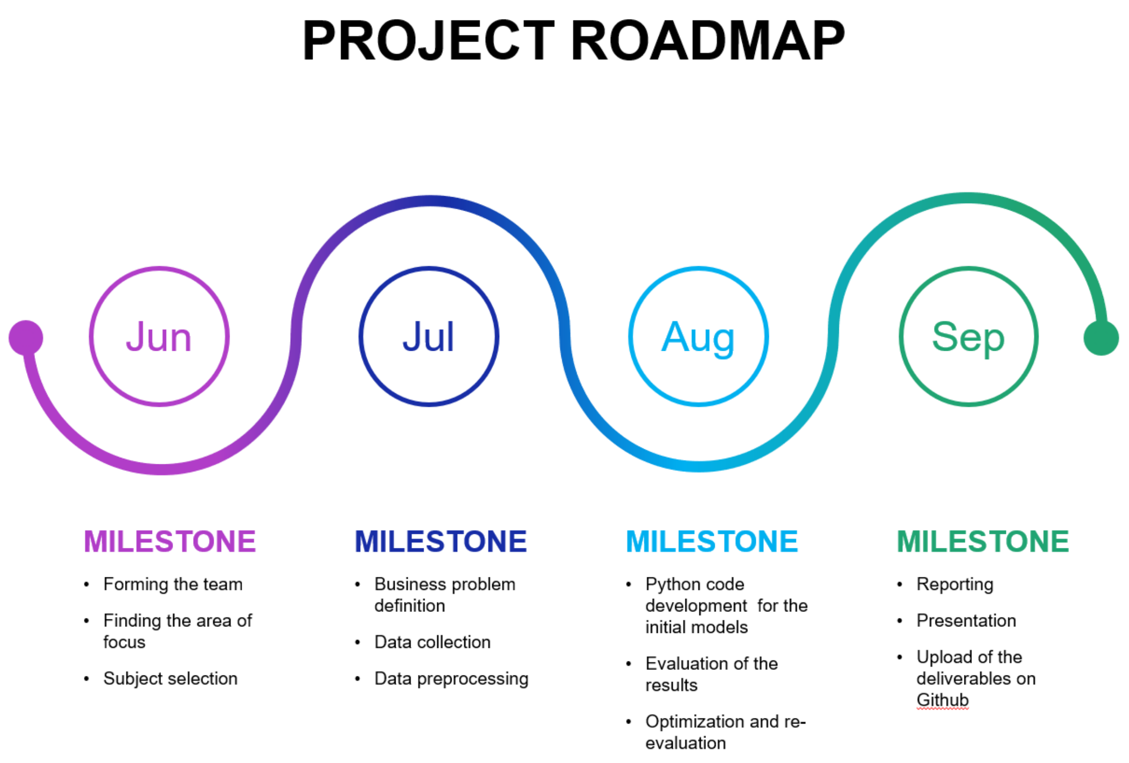
5. Members – Roles

The team is consisted of 4 members with different backgrounds and consequently the contribution and perspective of each individual was valuable on this project. The entire process was a team effort but at the same time, each member was working independently in order to bring new ideas, recommendations and improvements on the project. The core tasks performed can be summarized as:

- Brainstorming on the industry of interest where the project would focus
- Identifying the problem that needed to be answered
- Data Collection
- Data Cleaning
- Code development
- Model selection and algorithms
- Reporting
- Presentation of the results

6. Time Plan – Project Roadmap

Below is demonstrated the time-plan of the project with the core tasks performed:


PROJECT ROADMAP
Jun
Jul
Aug
Sep
MILESTONE
• Forming the team
• Finding the area of focus
• Subject selection
MILESTONE
• Business problem definition
• Data collection
• Data preprocessing
MILESTONE
• Python code development for the initial models
• Evaluation of the results
• Optimization and re-evaluation
MILESTONE
• Reporting
• Presentation
• Upload of the deliverables on Github

References

1. https://www.tensorflow.org/api_docs/python/tf/keras/optimizers/Ftrl
2. B. Jason, How to Choose Loss Functions When Training Deep Learning Neural Networks (2020), Machine Learning Mastery.
3. B. Pragati, 12 Types of Neural Network Activation Functions: How to Choose? (2022), V7Labs.
4. S. Kumar, Overview of various Optimizers in Neural Networks (2020).