

# US Short-term Housing Recommendations for College Students

Yuncheng Yu , Rongruo Zhou, Zhekai Dong

Department of Computer Science and Engineering, Texas A&M University, College Station, TX 77843

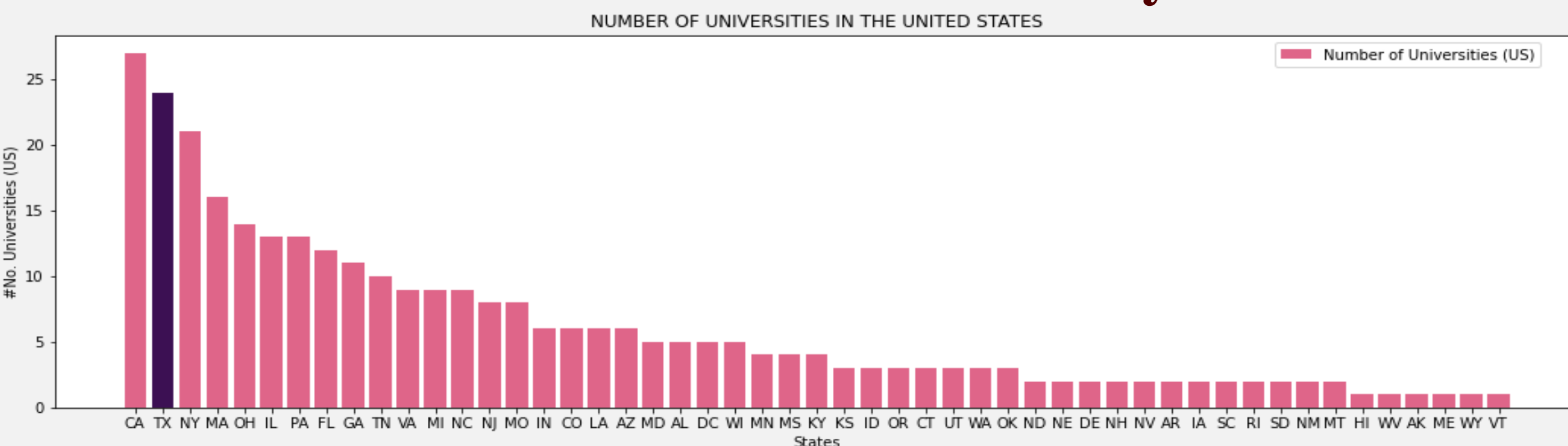
## Introduction

According to the Education Data Initiative report in August 2021, there is an average 2% annual college graduate enrollment rate increase in the United States [2]. In Texas A&M University, from 2019 to 2020 there was a 2.4% enrollment increase from a figure of 69,465 students to 71,109 [1]. As the increasing number of college students enrollment and the serious recent coronavirus pandemic situation and restrictions, short-term on campus housing for college student parents are becoming limited and therefore more off campus housing , such as Airbnb, should be considered. Thus, finding recommended short-term rental houses becomes quite a headache for the majority of U.S. college students and parents from out of state. Therefore, TeamSpirit plans to implement data mining and machine learning algorithms, such as LDA, Community Detection, Decision Tree, Random Forest, Adaboost, and Linear regression, using Austin housing as an example to establish a housing recommendation system for college families who desire to find good short-term housing in Austin.

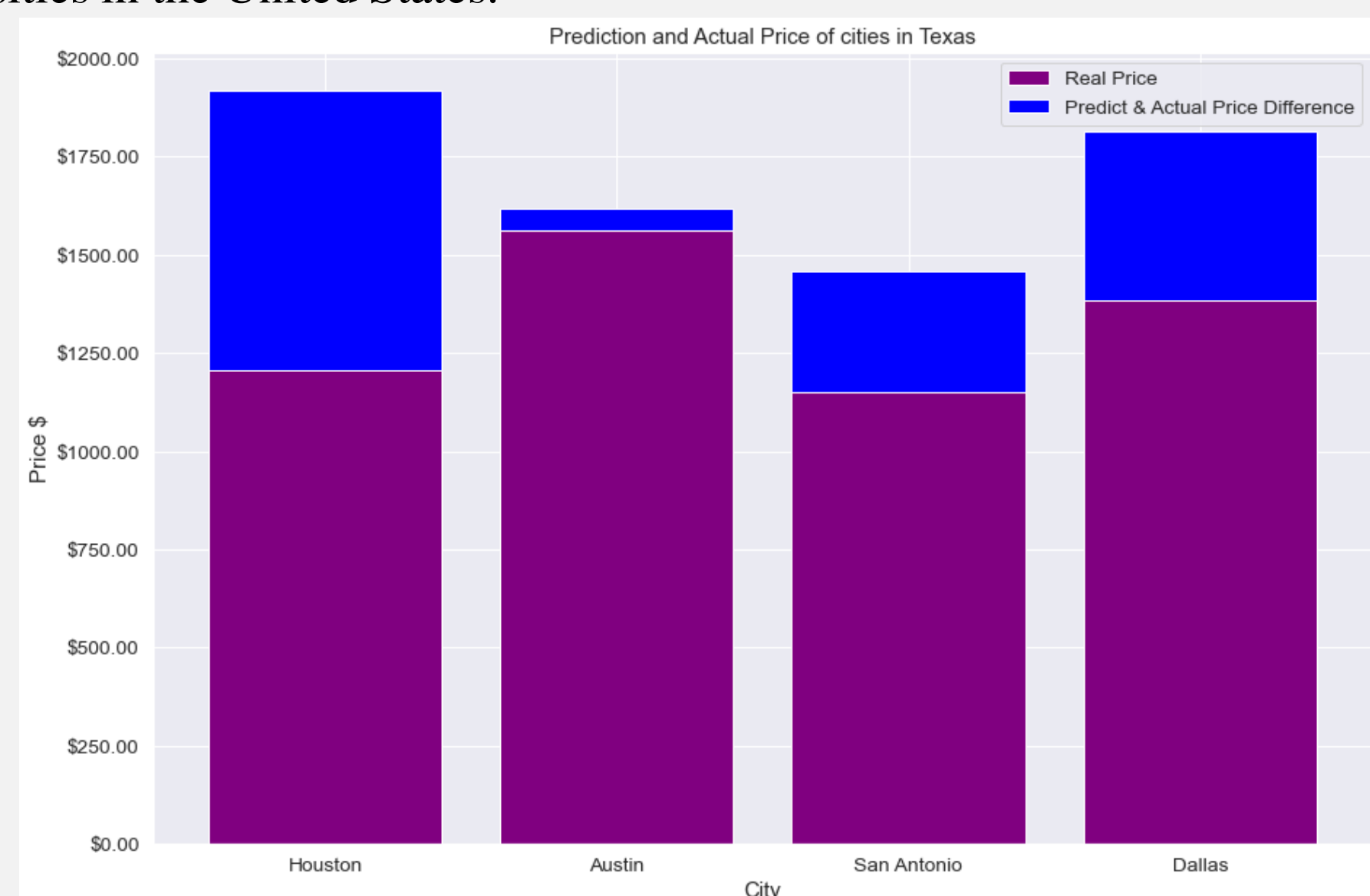
## Current Related Work

Most current related work can be categorized into two common approaches, which are either using some economic factors [3 ], such as GDP, or population density [4 ] to predict the rental price tendency or evaluate potential market rental price. Those analytic features are mainly based from statistics results to filter and sort housing options for users. However, in this project, the TeamSpirit provides a more accurate and user friendly approach based on users’ feedback and review to make house recommendations for users.

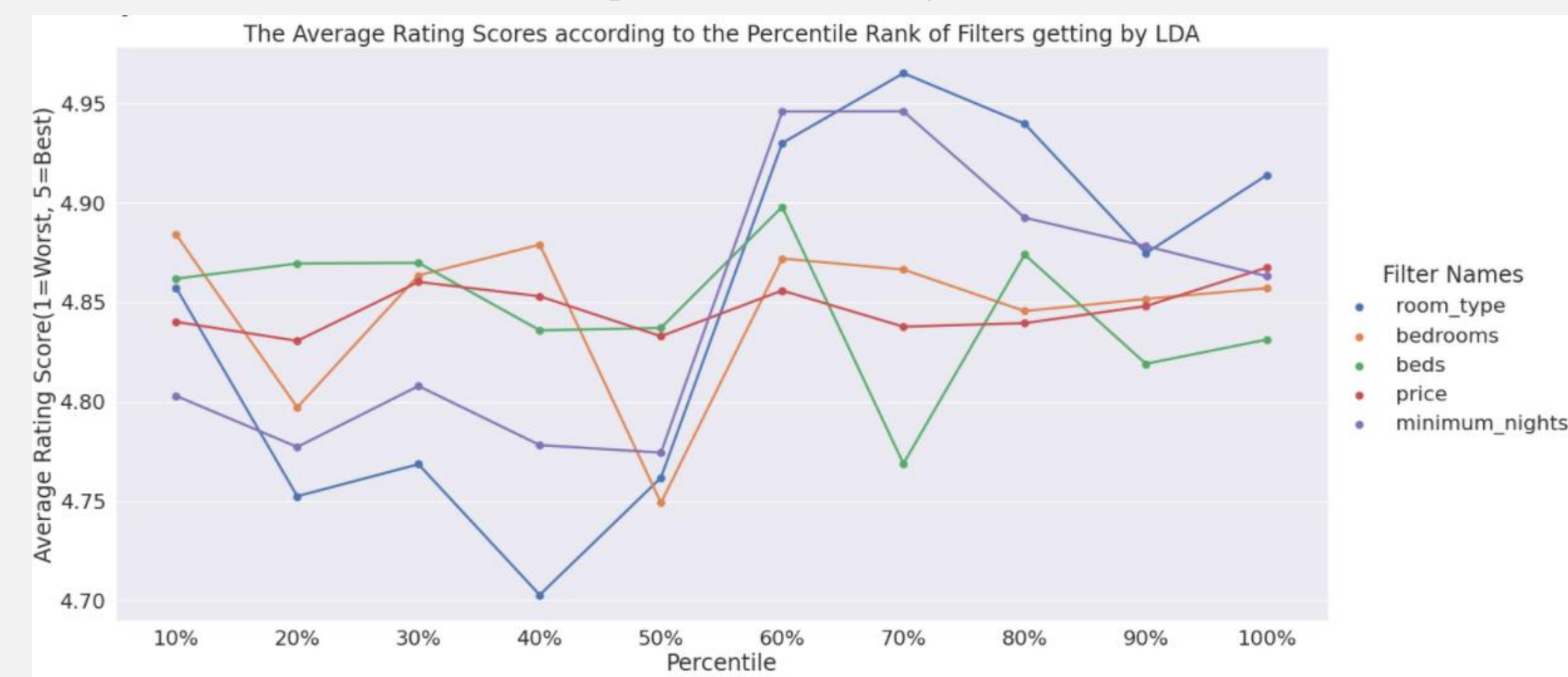
## Results Evaluation and Analysis



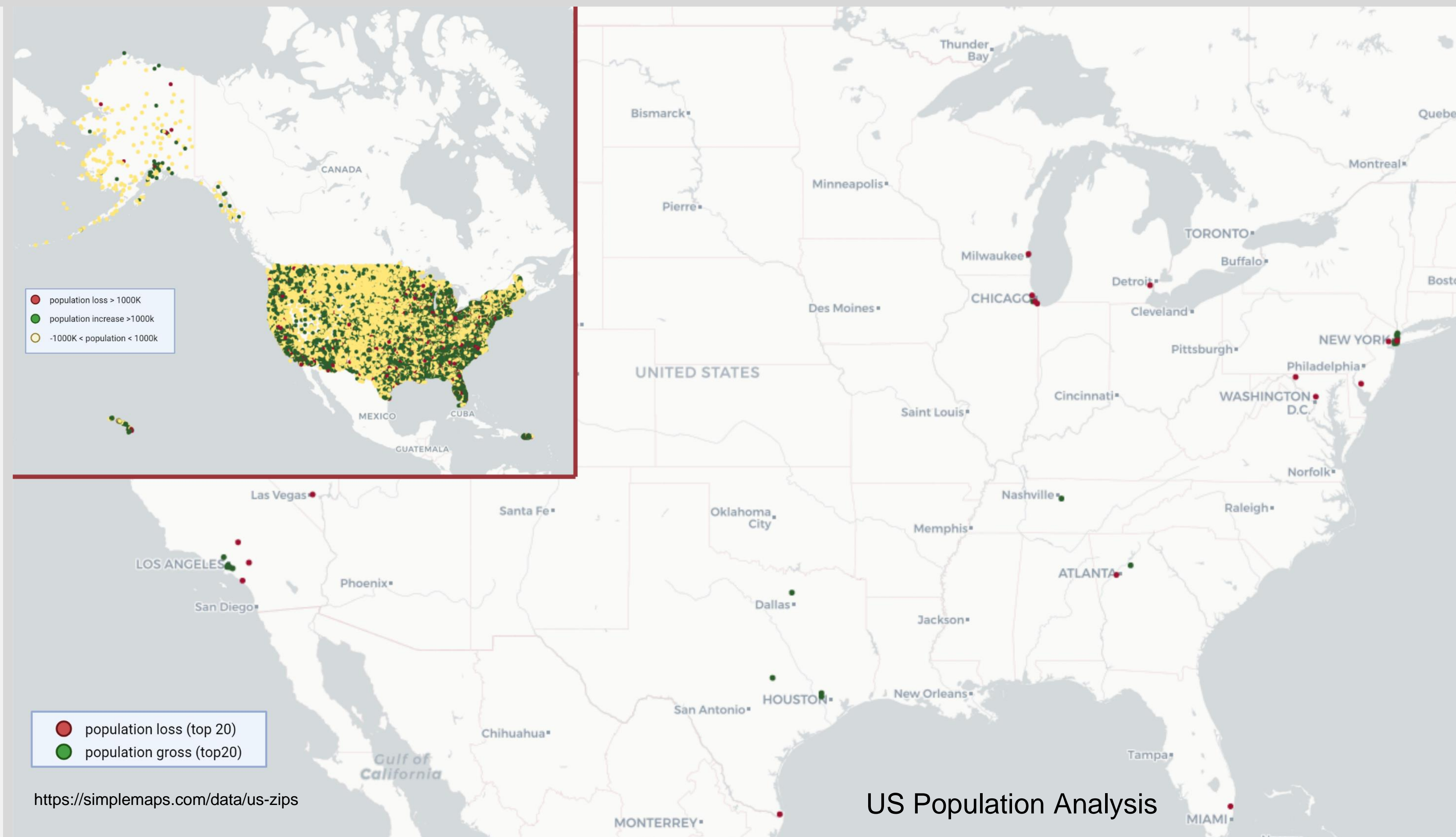
Using the recent university statistic data and comparing the US population from 2010 to 2021, Texas has the top population gross within the nation and second most amount of universities in the United States.



This training process provides an average error of estimation for Decision Tree is 21, for Random Forest is 16, for Adaboost is 23, for Linear Regression is 14. After the combination of all results, the average error of estimation is 19. Compared with the real data, the accuracy is 98%. With this training model, the result of difference between actual and predict rent price are 715, -58, 307, 432. Austin has the smallest price difference among all of these four cities in Texas, which leads the conclusion that the Austin has the most accurate and predictable housing market.



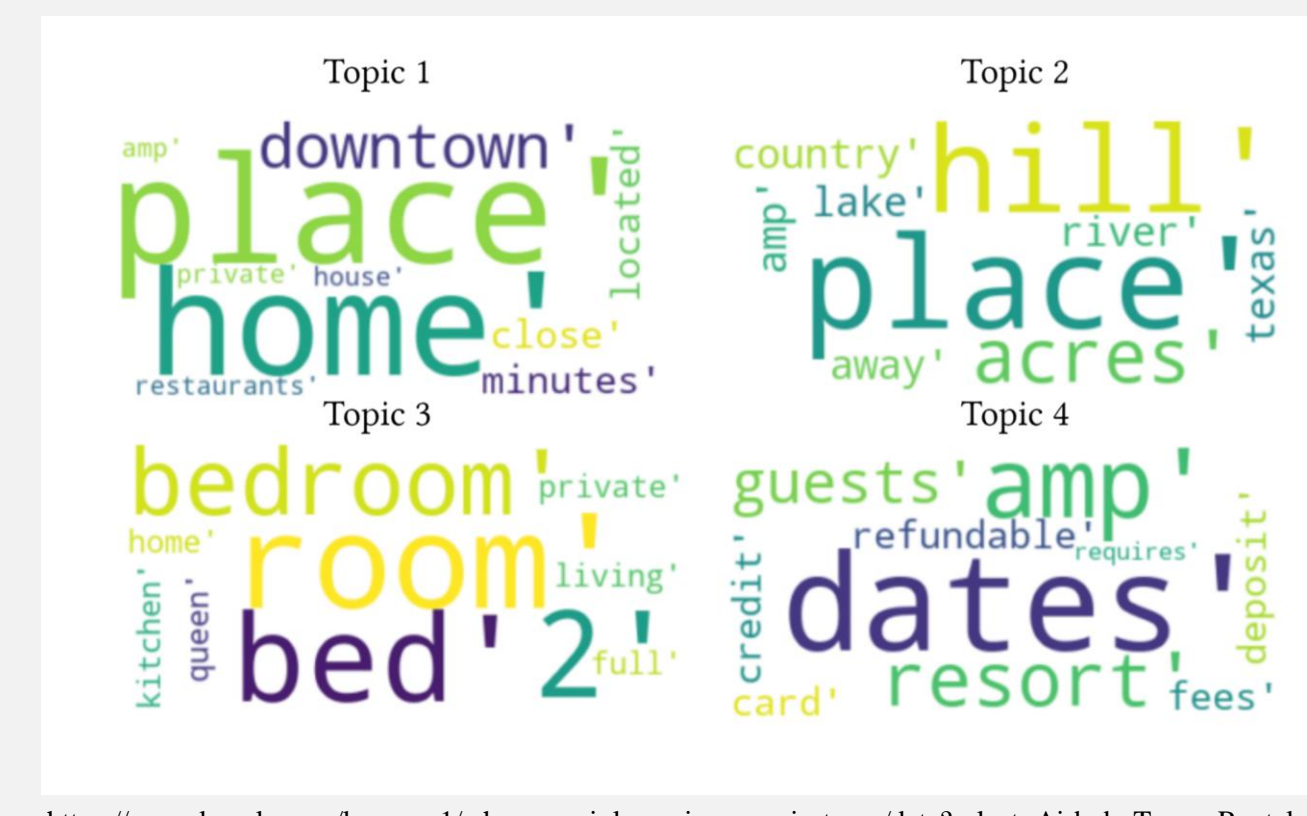
The above plot analyzes how the value of each filtering domain correlates to the general rating scores. For example, as shown in the plot, the price does not affect the rating score a lot since the houses with the lowest and highest prices have similar rating scores. But a higher minimum\_nights has a significant positive influence on the rating score. The above plot represents how the rating score changes based on the value increasing of each feature from low to high. In short, by observing the plot, it’s possible to find a ‘sweet point’ in the middle of the plot. To be specific, the user is most likely to find a high evaluated housing when the price is around \$150, the bedrooms and beds are both 2, the minimum nights is 2, and the room type is entire home or apartment.



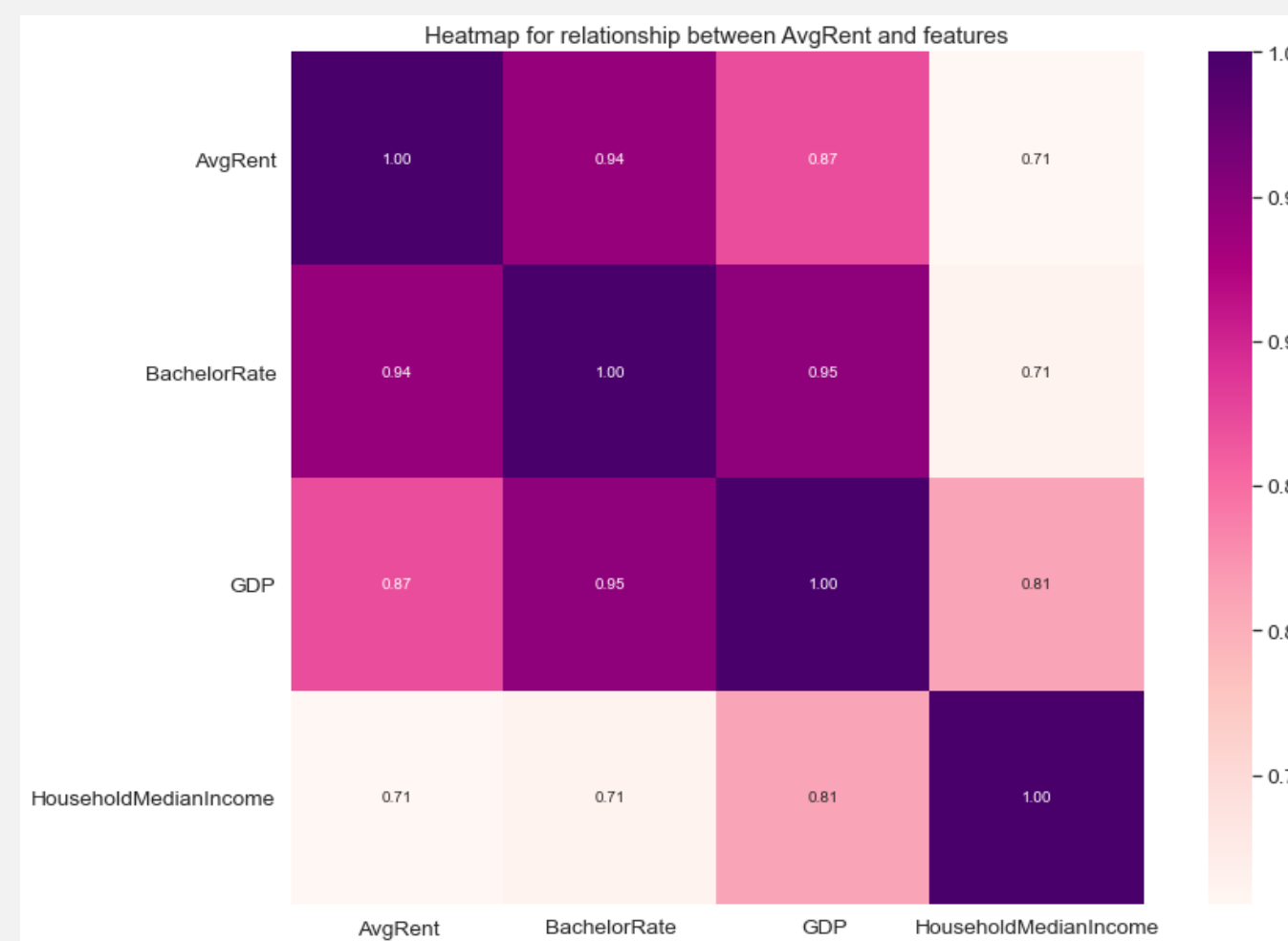
## Methods & Solutions

By analyzing the most recent Texas Airbnb Rentals dataset from Kaggle and using the text analysis **LDA** data mining Algorithm provided by the Spark Machine Learning (ML) library, four most important topics based on the rental property description of the dataset can be discovered to show what people care more about while seeking rental houses.

Utilizing the Pearson correlation coefficient with all of these features, the top 3 features in this dataset which are most correlated to the target label are “Bachelor Rate”, “GDP”, and “HouseholdMedianIncome”. The graph below shows the heatmap for the relation between target label and each features. From the previous part, there are lots of commonly mentioned key words in the housing community collected by LDA. The one which is easier to collect is “Commute Time”. The graph below represents seaborn.heatmap. Then, implement the training process by implementing the k-cross validation method and use machine learning algorithms such as Decision Tree, Random Forest, Adaboost, and Linear Regression to train and predict the rent price. Once all four model are trained, combine them and then divided by 4 to obtain the final predict result. Next, for four cities in Texas which are “Houston”, “Dallas”, “Austin”, and “San Antonio”, collecting the dataset and using the trained models above, obtain the predicting rent price.



https://www.kaggle.com/kamcaal/why-gun-violence-increase-in-texas/data?select=Airbnb\_Texas\_Rentals.csv



Using the Inside Airbnb public database of the 2021 Austin Texas Airbnb reviews and listings data, the TeamSpirit applied the **Community Detection** Algorithm by implementing the Trawling algorithm based on frequent itemset mining using FPGrowth to find potential user communities. Four user communities which are housing cost performance, location, cleanliness, and communication can be determined. By exploring the houses chosen by the users in each community and sorting them according to communities’ general and the specific scores related to the domain of each community. After that, top four recommended housings from each community as shown in the table below.

listing_url	name	description	host_location	room_type	bedrooms	beds	price	minimum_nights	review_scores_rating
<a href="https://www.airbnb.com/rooms/48449085">www.airbnb.com/rooms/48449085</a>	Charming Austin...	Just a 7-minute...	Austin, Texas	Entire home/apt	2	3	\$204	2	5
<a href="https://www.airbnb.com/rooms/4224760">www.airbnb.com/rooms/4224760</a>	Very Cozy 1BD w...	Close to downto...	Austin, Texas	Private room	1	1	\$90	1	5
<a href="https://www.airbnb.com/rooms/4508376">www.airbnb.com/rooms/4508376</a>	Camp Congress	Looking for the p...	Austin, Texas	Entire home/apt	7	12	\$1,500	2	5
<a href="https://www.airbnb.com/rooms/39558606">www.airbnb.com/rooms/39558606</a>	New + Modern 3...	This charming h...	Austin, Texas	Entire home/apt	3	4	\$439	2	5

http://insideairbnb.com/get-the-data.html



The previous LDA Algorithm results of four important topics’ keywords can be used as a filter and combine our dataframe to more accurately filter the results of recommendations for users.

## Conclusions and Discussions

Based on the datasets obtained from U. S. Census Bureau and university statistics, Texas has the second amount of universities and top population gross and should be a good target to study. Then by using the Machine learning algorithms trained model, Austin has the most predictable housing market with the least difference between the predict and actual price, which lead to the analysis and rent recommendation in the next step and comes to the key portion of the project, the recommendation system utilizing of Community Detection Algorithm toward housing reviews data is a reliable way to find suitable housing choices with excellent ratings. Deploying the filtering function so that the user can set the filtering options to find the perfect matching of the houses they want, like setting price range, number of beds, number of rooms. And that would be the complete version of our project. The main issue is always looking for usable datasets during the project. As the returned results seem to be fantastic housing choices by manually looking at some of them, the project goal is achieved. A short come of our project is only making short-term Airbnb recommendations instead of long-term choices such as apartments or houses. That’s generally due to the lacking of data. It would be great to have more detailed data in the future.

1. References
1. <https://today.tamu.edu/2020/09/18/texas-am-fall-enrollment-increases-2-4-percent/>
2. <https://educationdata.org/number-of-college-graduates/>
3. <https://www.noradarealestate.com/blog/housing-market-predictions/>
4. [https://www.jchs.harvard.edu/sites/default/files/Harvard\\_JCHS\\_Americas\\_Rental\\_Housing\\_2020.pdf](https://www.jchs.harvard.edu/sites/default/files/Harvard_JCHS_Americas_Rental_Housing_2020.pdf)