

**Analysis of Denver Airbnb Listings in June 2024: Distinguishing Characteristics and Variations  
by Superhost Status**

Briana Palencia, Edda Phillips, Jigna Chaudhary, Pratik More

CU Denver Business School

BANA 6620: Computing for Business Analytics

Dr. Ziyi Wang

December 11, 2024

## Abstract

This analysis examines characteristics that contribute to the success of Airbnb listings in Denver during the month of June 2024, with a focus on differences between superhosts and non-superhosts. Using exploratory data analysis (EDA), k-means clustering, linear regression, and random forest, we explored relationships between listing performance metrics, host characteristics and neighborhood, and how they differ with superhost status.

Our findings show that both superhosts (88%) and non-superhosts (82%) most commonly offer 'entire place' listings. Superhosts charge an average of \$526 per night, compared to \$197 for non-superhosts, and maintain a higher average review score (4.8 vs. 4.7). Superhosts also cancel 12.3% fewer bookings and respond faster to inquiries (1.54 hours vs. 5.34 hours). Five Points, Denver's most-listed neighborhood, has an even mix of superhosts and non-superhosts, while other top neighborhoods show slight variations. Clustering analysis reveals that popular neighborhoods show only small differences in average review ratings between high and low-rated clusters. Additionally, each year of hosting correlates with a 0.009-point increase in ratings. An OpenAI-powered chat model was then developed to provide hosts with strategies to improve their listings and achieve superhost status.

In conclusion, the analysis reveals that superhost status is influenced by responsiveness, guest satisfaction, and reliability, rather than location or listing type. Superhosts maintain higher ratings, faster response times, and higher overall review ratings. Our predictive modeling from random forest and through OpenAI highlights the importance of these factors, providing insights for hosts to improve performance and increase their chances of becoming superhosts in Denver's competitive market.

## Introduction

Airbnb has become a major player in the accommodation sector, competing alongside traditional lodging options. In 2023, 448 million Airbnb nights were booked on the platform (Statista, 2023). Meanwhile, Denver has become a popular travel destination, seeing a 3% rise in visitors in 2023 (Visit Denver, 2023). Currently, high-performing Airbnb lists are crucial for meeting the demands of travelers in Denver. As travel continues to increase after the COVID pandemic, understanding what factors make an Airbnb successful amongst guests is essential. One key factor is a host's superhost status, awarded to hosts who maintain a rating of more than 4.8, respond to more than 90% of inquiries within 24 hours, and complete at least 10 stays or 3 long-term bookings per year (Airbnb, 2024). This analysis seeks to identify the characteristics that distinguish high-performing Airbnb listings in Denver, and explore how these factors vary by superhost status, providing insight for hosts looking to improve their listings' performance.

## Methods

Our analysis was conducted on a dataset of 5,016 Airbnb listings in Denver for June 2024 from Inside Airbnb (2024), containing information about listing characteristics, host metrics, neighborhood, availability and review scores.

**Data Cleaning:** To prepare the dataset for analysis, we addressed missing values, corrected data inconsistencies, and expanded categorical variables (e.g., room types).

**Exploratory Data Analysis (EDA):** Descriptive statistics and data visualizations (e.g., pie charts, bar charts, box plots, scatter plots) were used to summarize key variables and identify trends and patterns. Correlation matrices identified significant variables for further analysis.

**Clustering:** To identify trends, k-means clustering was applied to group popular neighborhoods (based on the number of listings) by average overall review ratings.

**Regression Analysis:** A linear regression model was developed to examine the relationship between a host's experience (measured in years) and their listing's overall review rating.

**Random Forest:** A random forest model was developed to further enhance the predictive accuracy of superhost status. This method identified the key variables influencing superhost status.

**OpenAI Chatbot (LangChain):** LangChain was used to integrate the random forest model with OpenAI, allowing predictions on superhost status and providing tailored recommendations based on listing data.

## Results

- The most popular room type for super and non-superhosts are listings offering the 'entire place'. Of all the 2024 listings, 88% of superhosts and 82% of non-superhosts offered entire place listings. This includes entire homes or entire sections of a home. (Exhibit 1)
- On average, superhosts charge \$526 per night, while non-superhosts charge \$197 per night. However, superhosts have a much wider range of prices. The prices of superhost listings vary more, with a typical difference of about \$5,208, while non-superhosts' prices are more consistent, with a difference of around \$330. (Exhibit 2)
- To qualify for superhost status, a host must achieve an overall guest rating of 4.8 or higher. However, the average guest rating for non-super hosts was 4.7, indicating only a slight difference in review scores based on status. There was a marked difference, however, in a host's acceptance rate, or the rate of cancellations by the host. Non-superhosts canceled guest bookings 12.3% more than superhosts. (Exhibit 3)
- Superhost status does not necessarily align with neighborhood popularity. For example, in Denver's most-listed neighborhood, Five Points, superhosts and non-superhosts each represent about 8.5% of the listings. In Highland, the second most-listed area, non-superhosts outnumber superhosts by 2.5%, while in the third most popular neighborhood, superhosts exceed non-superhosts by just 1.3%. (Exhibit 4)
- Of the top ten neighborhoods by listing count, all but Union Station and Gateway-Green Valley Ranch belong to Cluster 1, which has average review ratings ranging from 4.86 to 4.91. Cluster 0, which includes these two neighborhoods, has average ratings of 4.80 and 4.78. With a small difference of 0.13, neighborhood appeal does not appear to significantly impact listing performance, as these lower-rated neighborhoods are not at the bottom of the top 10 and the difference in ratings is minimal. (Exhibit 5)
- Superhosts respond to guest inquiries much more quickly and more often than non-superhosts. The average response time for superhosts is 1.54 hours, compared to 5.34 hours for

non-superhosts, which is nearly 3.5 times faster. Superhosts responded 99% of the time, whereas non-superhosts responded 94%. (Exhibit 3, Exhibit 6)

- As hosts gain more experience (years as host), their overall Airbnb rating improves. Starting with an average review rating of 4.78, each additional year of experience a host has leads to an increase of 0.009 points in their overall listing's rating. (Exhibit 7)
- A machine learning model (Random Forest Classifier) was trained on the dataset to predict superhost status, giving 80% accuracy. Features were scaled using StandardScaler, and the model was optimized with class weights to handle slight imbalances in the target variable. (Exhibit 8)
- The Airbnb Superhost Classification Chatbot was developed to assist hosts in assessing their potential to achieve superhost status based on manually defined criteria derived from Airbnb's hosting standards. The chatbot serves as an interactive and user-friendly tool that provides practical feedback to improve hosting metrics. (Exhibit 9)

## Conclusion

This analysis reveals that hosts who achieve superhost status consistently demonstrate higher performance metrics across several dimensions. Superhosts are more likely to offer 'entire-place' accommodations, command higher (yet more variable) nightly rates, and maintain stronger guest satisfaction, as reflected in their response times, acceptance rates, and overall review ratings. While differences in guest rating averages between superhosts and non-superhosts are subtle, the reliability and timeliness of superhosts—exemplified by faster response times and fewer cancellations—stand out as key factors that distinguish top-performing hosts.

Neighborhood popularity does not directly correlate with superhost status. Superhosts are distributed broadly across neighborhoods, suggesting that the quality of host practices plays a more significant role in enhancing performance than neighborhood appeal or location. Moreover, as hosts gain experience, their ability to meet or exceed guest expectations improves over time.

The predictive modeling results confirm the importance of these host attributes, with the random forest model demonstrating an 80% accuracy in classifying superhost status. Integrating this model into an interactive chatbot provides a valuable resource for hosts seeking recommendations. Taken together, these findings highlight that achieving and maintaining superhost status depends less on any single characteristic—such as location or listing type—and more on guest satisfaction, responsiveness, and trustworthiness. These insights can guide new and existing hosts in prioritizing the key elements needed to enhance their performance and thrive in Denver's increasingly competitive Airbnb market.

## Documentation

Exhibit 1: Pie charts showing the percentage of room types in listings for both super and non-superhosts

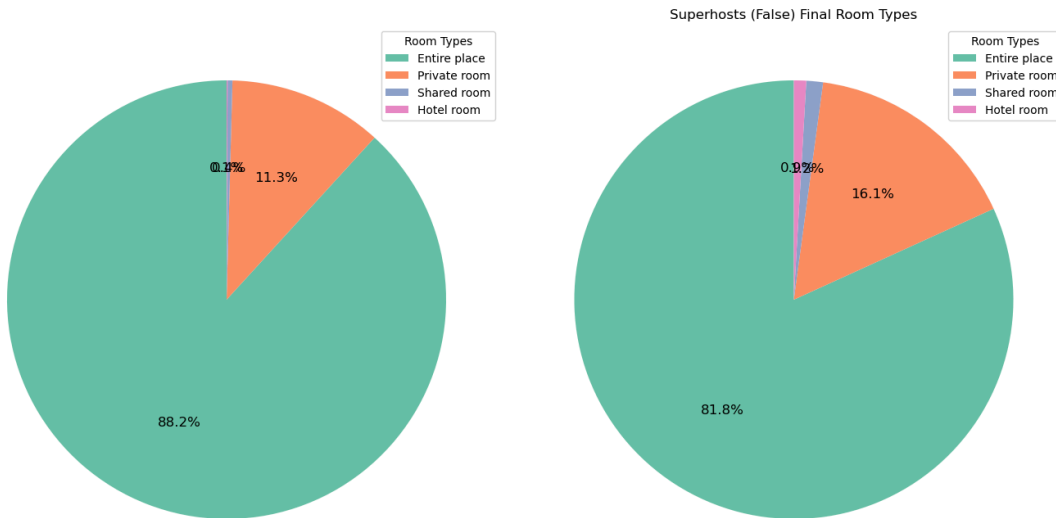


Exhibit 2: Boxplots showing the price spread of super and non-superhosts comparatively, excluding outliers

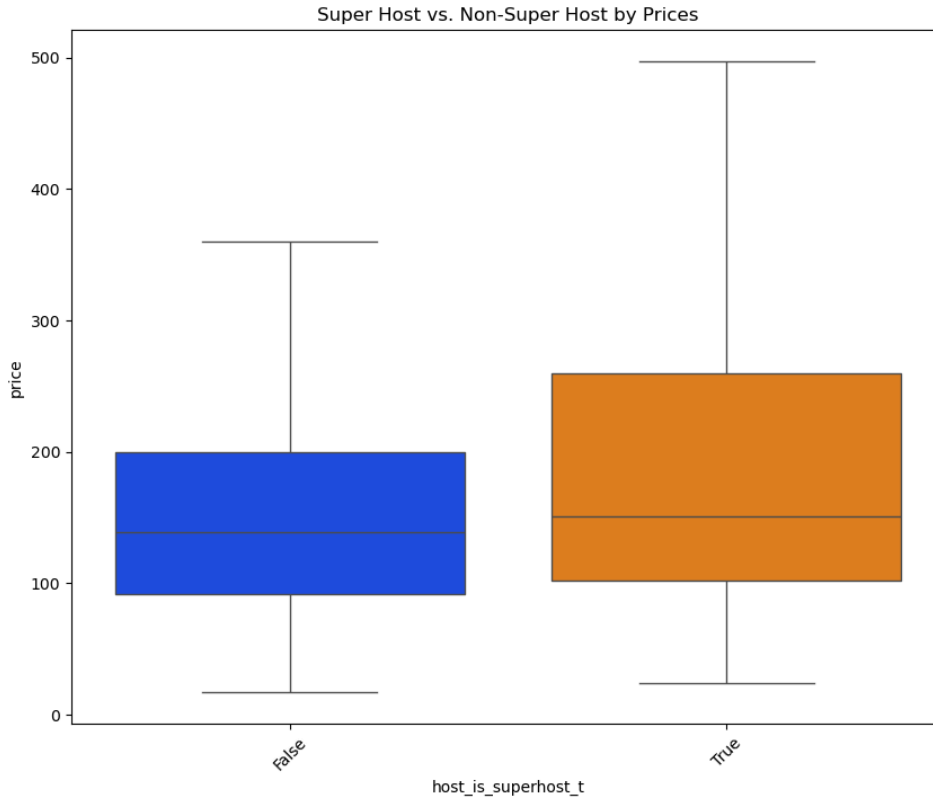


Exhibit 3: Side-by-side comparison of the host's response rate and acceptance rates

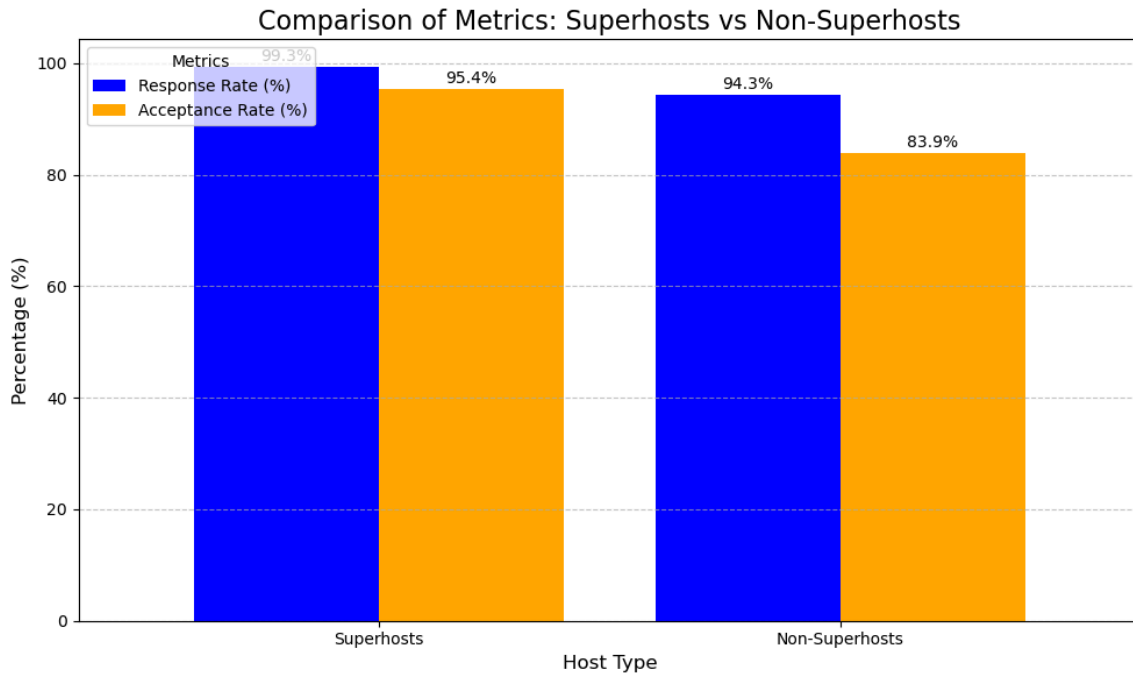


Exhibit 4: Proportion distribution (%) of super and non-superhosts in the top 10 neighborhoods based on listings count

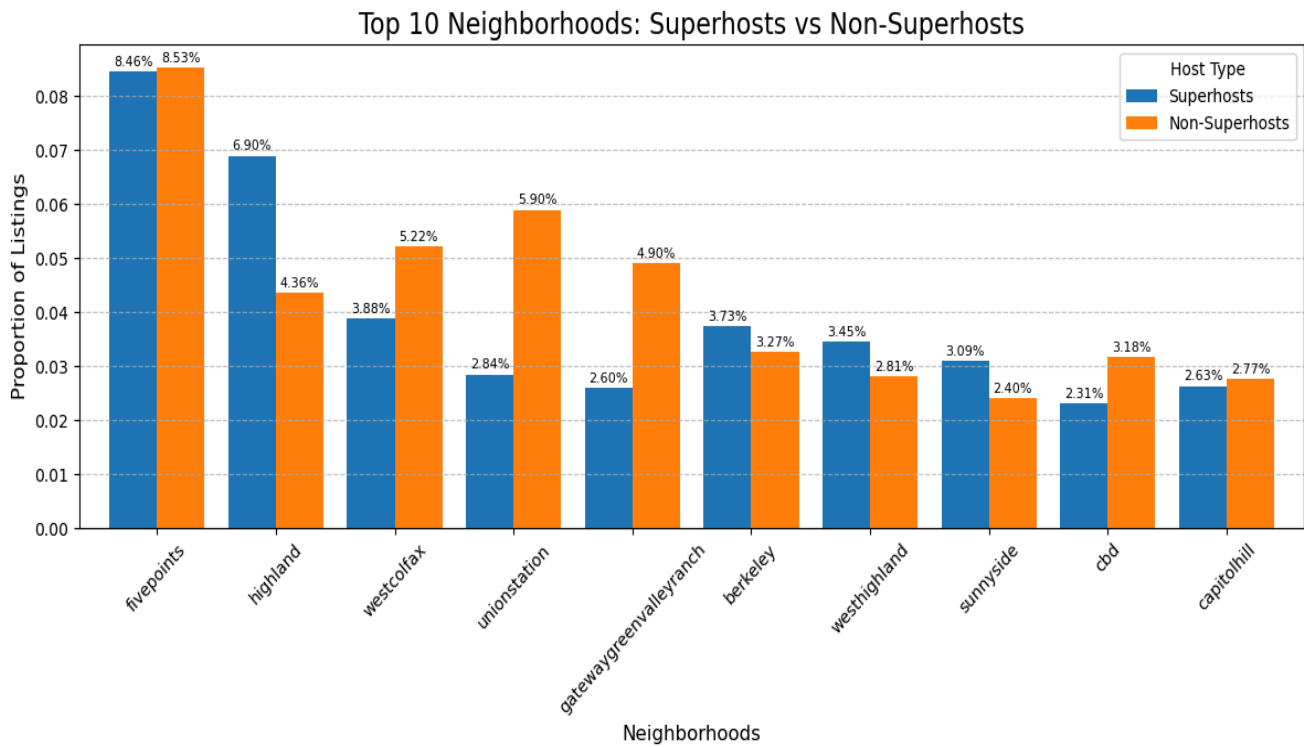


Exhibit 5: K-means cluster analysis of top 10 Neighborhoods (by listing count) grouped into two clusters based on average overall review ratings

K-means Clustering of the Top 10 Neighborhoods (by listing count) on  
on Average Overall Review Ratings:

	Neighborhood	Cluster	review_scores_rating
0	neighborhood_fivepoints	1	4.874956
1	neighborhood_highland	1	4.891810
2	neighborhood_westcolfax	1	4.869168
3	neighborhood_unionstation	0	4.801168
4	neighborhood_gatewaygreenvaleyranch	0	4.781186
5	neighborhood_berkeley	1	4.893810
6	neighborhood_westhighland	1	4.886630
7	neighborhood_sunnyside	1	4.911670
8	neighborhood_baker	1	4.874535
9	neighborhood_capitolhill	1	4.862678

Exhibit 6: Line plots comparing superhosts vs non-superhosts with their response rate and time

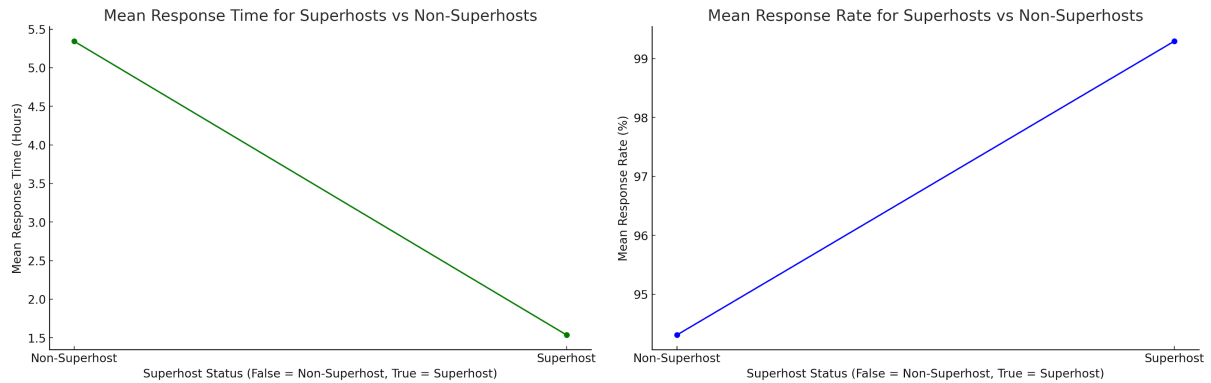
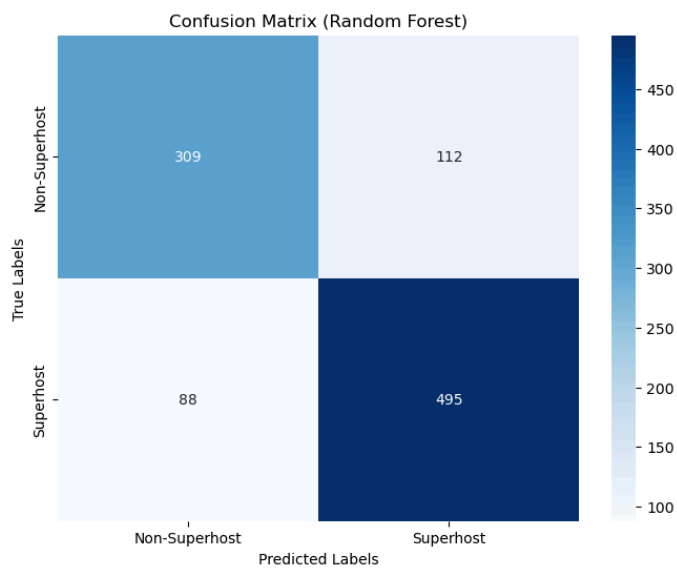


Exhibit 7: Linear regression model showing the baseline review score and how many points the overall rating improves each year as a host

OLS Regression Results						
Dep. Variable:	review_scores_rating	R-squared:	0.015			
Model:	OLS	Adj. R-squared:	0.015			
Method:	Least Squares	F-statistic:	77.54			
Date:	Sat, 07 Dec 2024	Prob (F-statistic):	1.76e-18			
Time:	08:09:24	Log-Likelihood:	25.827			
No. Observations:	5016	AIC:	-47.65			
Df Residuals:	5014	BIC:	-34.61			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	4.7893	0.009	543.438	0.000	4.772	4.807
years_as_host	0.0093	0.001	8.806	0.000	0.007	0.011
Omnibus:	6279.333	Durbin-Watson:	1.873			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1198888.775			
Skew:	-6.742	Prob(JB):	0.00			
Kurtosis:	77.529	Cond. No.	22.0			
Notes:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.						

Exhibit 8: Confusion matrix and random tree classification report showing the accuracy of predicting superhost status





## Random Forest Classification Report:

	precision	recall	f1-score	support
False	0.78	0.73	0.76	421
True	0.82	0.85	0.83	583
accuracy			0.80	1004
macro avg	0.80	0.79	0.79	1004
weighted avg	0.80	0.80	0.80	1004

## Exhibit 9: Illustrative output of Airbnb superhost chatbot

---

```
.. Welcome to the Airbnb Superhost Chatbot!
   Provide the following inputs about your hosting profile:
```

```
--- Prediction ---
```

```
Prediction: Superhost
```

```
--- Explanation ---
```

```
Congratulations! Based on the provided criteria, you qualify as a Superhost!
```

---

```
✓ if __name__ == "__main__": ...
```

---

```
.. Welcome to the Airbnb Superhost Chatbot!
   Provide the following inputs about your hosting profile:
```

```
--- Prediction ---
```

```
Prediction: Not a Superhost
```

```
--- Explanation ---
```

```
Based on the criteria, you do not currently qualify as a Superhost. Consider the following:
```

- Improve your host response rate to at least 90%.
- Improve your host acceptance rate to at least 90%.
- Increase your availability to at least 100 days in the next year.
- Strive for more frequent bookings to increase reviews per month.
- Aim to improve guest satisfaction to achieve a review score of 4.5 or higher.

## References

- Airbnb. (2024). *Superhost*. Airbnb. Retrieved from <https://www.airbnb.com/superhost>
- Inside Airbnb. (2024). *Denver data*. Retrieved from <https://insideairbnb.com/denver/>
- OpenAI. (2024). *ChatGPT (Dec 15 version) [Large language model]*. OpenAI. Retrieved from <https://openai.com/chatgpt>
- Statista. (2023). *Airbnb - statistics & facts*. Retrieved from <https://www.statista.com/topics/2273/airbnb/#topicOverview>
- Visit Denver. (2023). *Denver tourism continues strong growth in 2023*. Retrieved from <https://www.denver.org/articles/post/denver-tourism-continues-strong-growth-in-2023/>

## Acknowledgements

AI was utilized to support the development of code for exploratory data analysis (EDA) and predictive modeling. It provided guidance and suggestions for specific coding challenges encountered during our final project.