NC STATE UNIVERSITY

Decoding Market Dynamics: A Data-Driven Approach to Real Estate Analysis

Rohan Dodda, Joshua Kotzker, Dipen Patel

Department of Computer Science and Statistics, North Carolina State University

rbdodda@ncsu.edu | jmkotzke@ncsu.edu | dpatel36@ncsu.edu

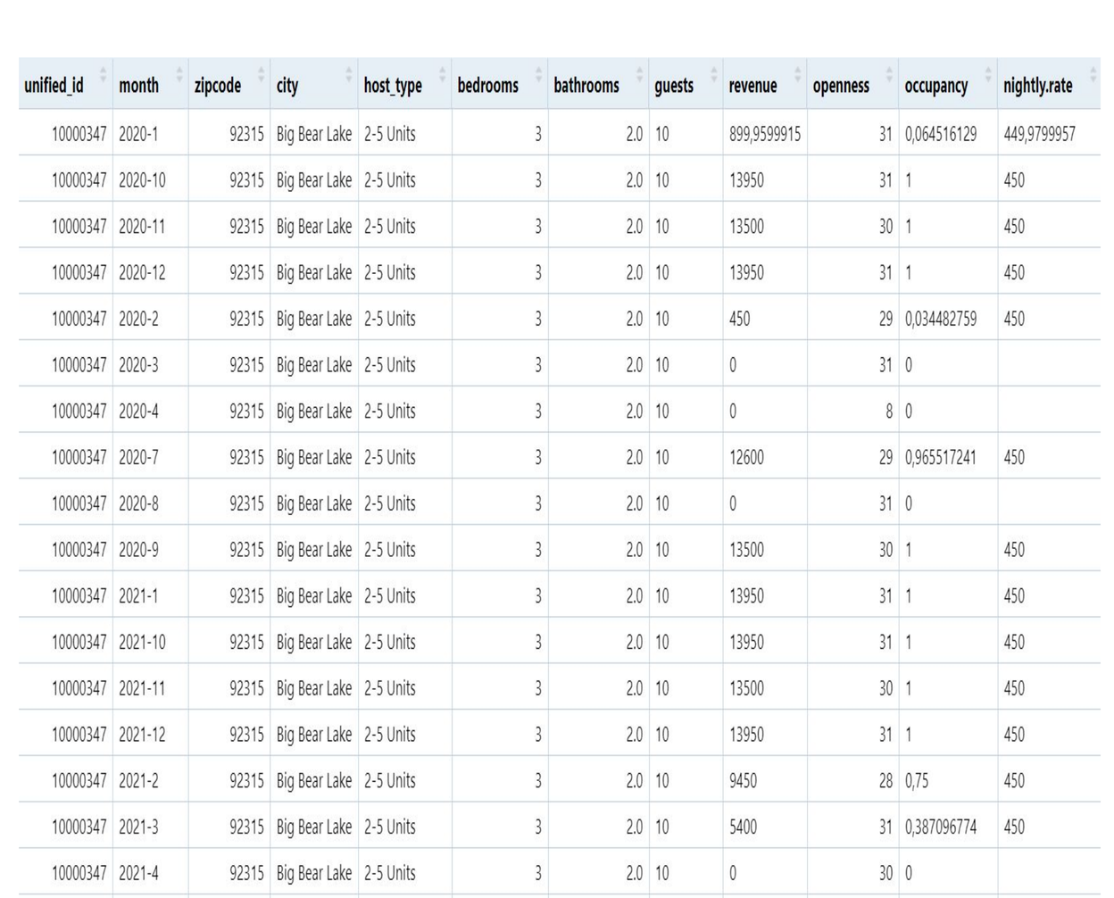
Group 13

Objective and Overview

The data we used for this report is a collection of demographic and revenue information for rentals posted on AirBnb. The properties listed are located in San Bernardino County, California and the data extends from 2019 into 2022. Our analysis leverages an extensive collection of data to attempt to reveal patterns and correlations that drive the rental real estate market. Predictive models were made to give accurate assessments on property rental pricing and revenue forecasting. We hope the information provided will be useful for those willing to supplement their income stream with rentals.

airbnb

Cleaning and Wrangling

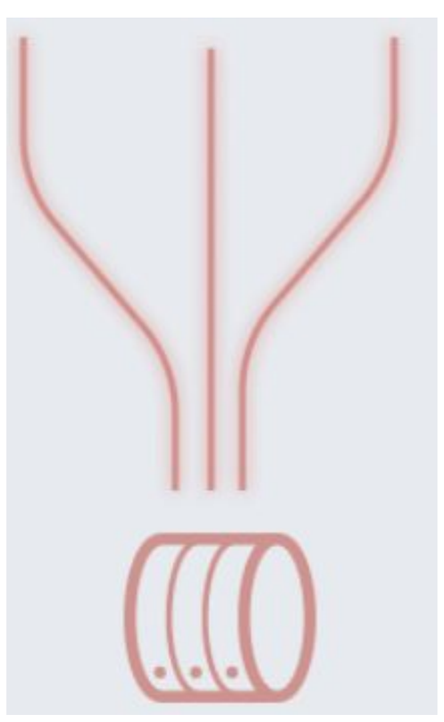


The initial phase of our analytical analysis involved data cleaning and merging. Using R and the tidyverse library, we found identifiers, corrected numerical formats and resolved inconsistencies across datasets. We transformed the yearly market analysis data, amalgamated it with amenities and geolocation information, and produced a unified, clean dataset for analysis. Character-based representations of guest counts were converted to numeric integers to streamline quantitative analysis.

Results

The results of our studies reinforced the hypotheses we had going into the creation of the models.

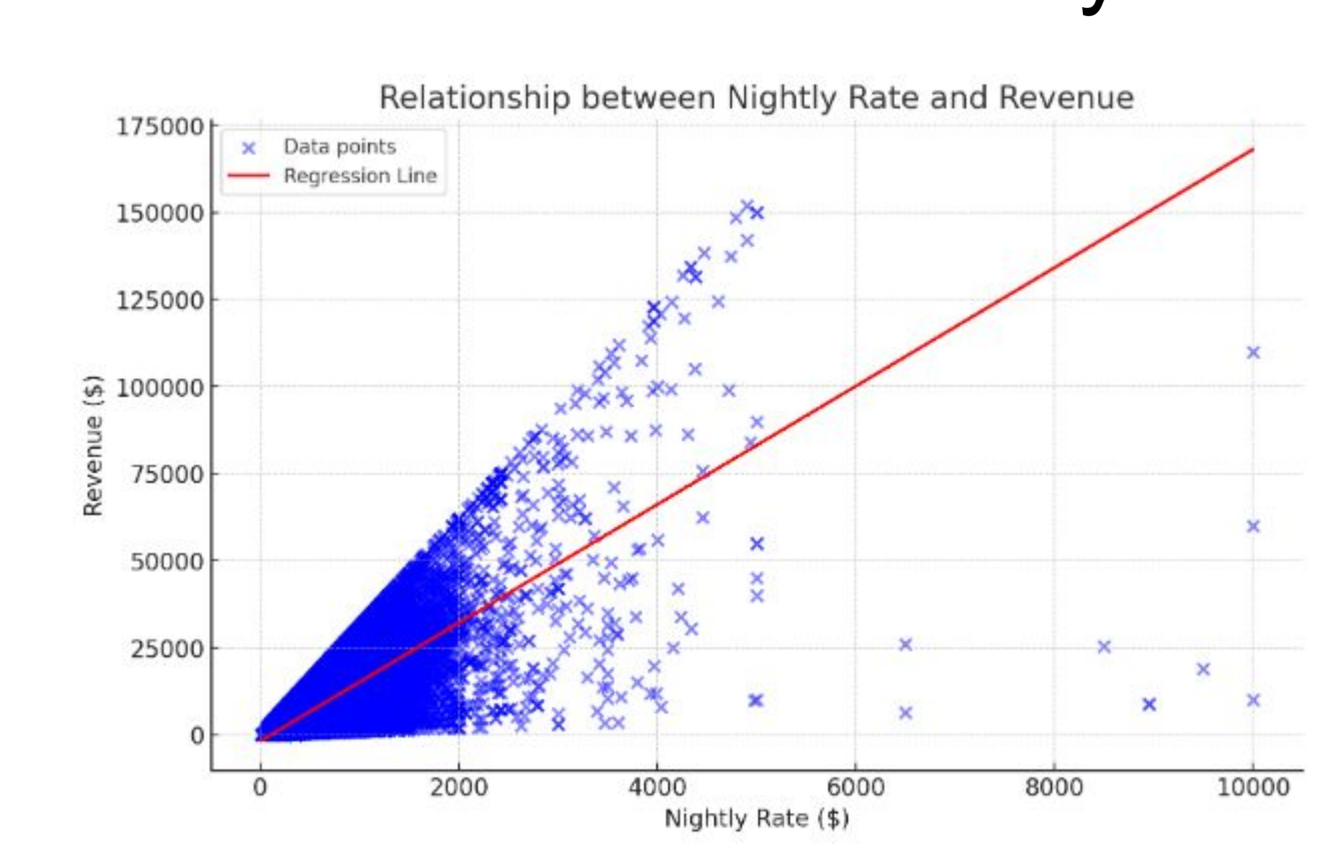
- Pricing gauges for rentals based off of their accommodations in shows units more amenities added upwards of $100s of more nightly rates.
- Definite +$1500 discrepancy between profits of individual owners and that of companies.
- Revenue of Airbnb rentals had a significant impact during COVID but has returned to stable levels in the years since



Linear Regression

(Rohan Dodda)

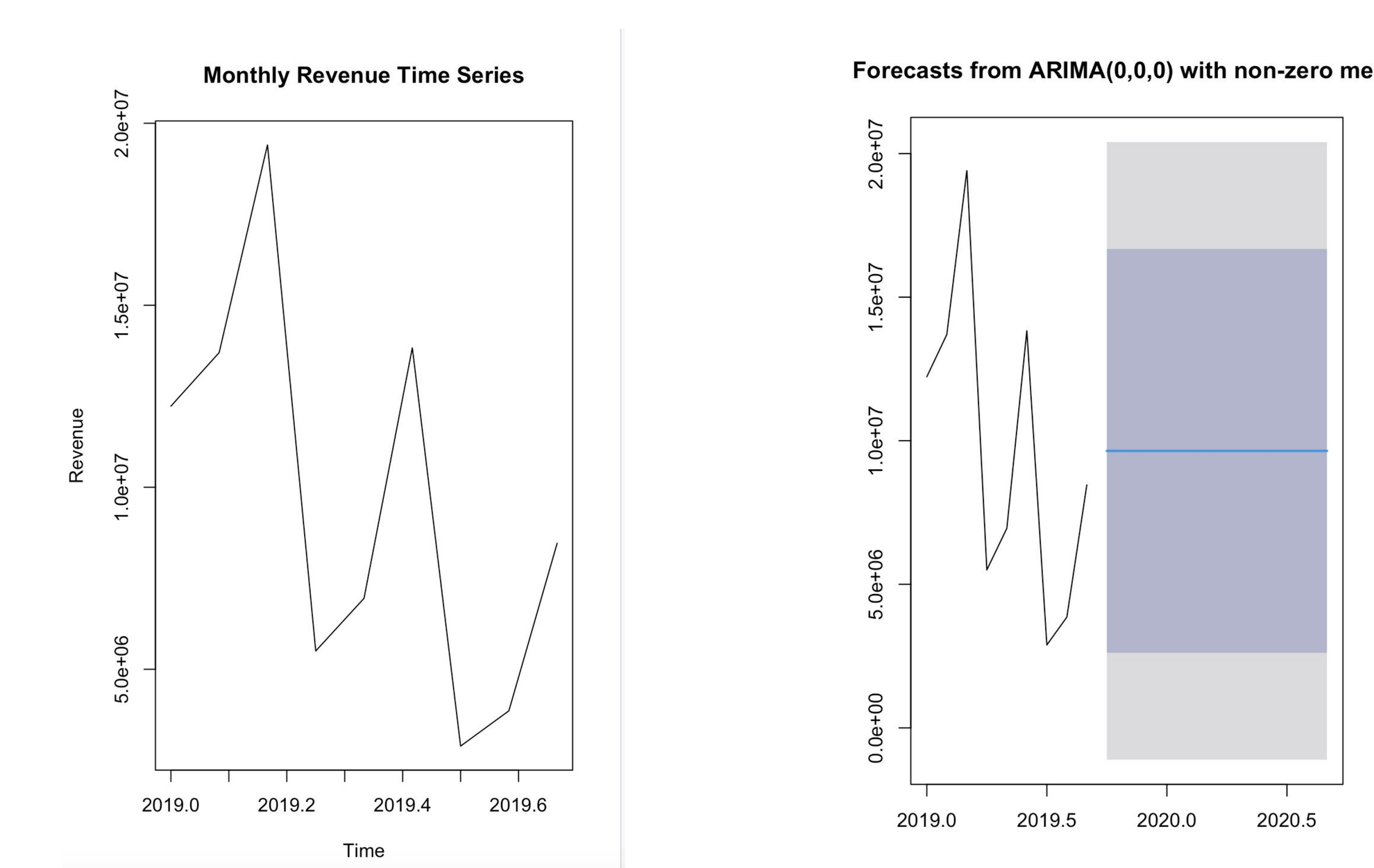
We predicted the pricing strategy for Airbnb listings, with a particular focus on **nightly rates as a determinant of overall price**. Our model analyzed the relationship between nightly rates and the listed prices of properties, considering variables such as location and properties amenities, which were significant in price formation. We carefully assessed the model's coefficients to understand the elasticity of property prices with respect to changes in nightly rates. This insight is particularly useful for hosts in strategizing their pricing to optimize earnings. We checked for multicollinearity and evaluated the model's R-squared value to gauge the explanatory power of the nightly rate on price. In doing so, our model not only aimed to predict prices based on nightly rates but also provided a quantitative foundation for understanding the pricing mechanisms within the Airbnb market in San Bernardino County.



Time Series

(Dipen Patel)

Utilizing ARIMA models, we aimed to capture and forecast future revenue trends. The choice of ARIMA(0,0,0) with a non-zero mean was based on the time series' characteristics, which suggested a model without differencing or moving average components, implying that the data did not exhibit strong autocorrelation patterns over time. This was further evidenced by the Partial Autocorrelation Function (PACF) for the original series, where no significant lags beyond the immediate period were observed, indicating that a simpler model might be appropriate for this particular time series.



Classification Tree

(Joshua Kotzker)

The goal of this study was to attempt to link the categorical variables to actionable predictions of revenue for certain conditions of AirBnb rentals. A prediction tree based off of an ANOVA model provides a clean and neat display of prediction for this project. The methodology was straightforward: we split our data into training and testing sets and developed a prediction tree without imposing restrictions on node or bucket sizes. Crucially, the model underscored the revenue discrepancy between single-owner and professional or multi-homeowner properties, aligning with our initial hypotheses.