**BANA7038 PROJECT REPORT**

**TEAM MEMBERS:** Saurav Nandi, Reshma Punukollu, Pavan Motagi

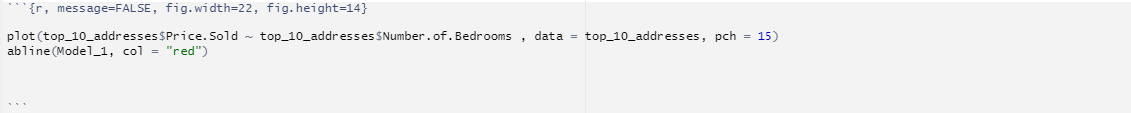
**ROLES:**

* Saurav Nandi – Responsible for analysis in R
* Reshma Punukollu – Responsible for analysis in Python
* Pavan Motagi – Responsible for analysis in Tableau
* All – Responsible equally for the presentation and the report.

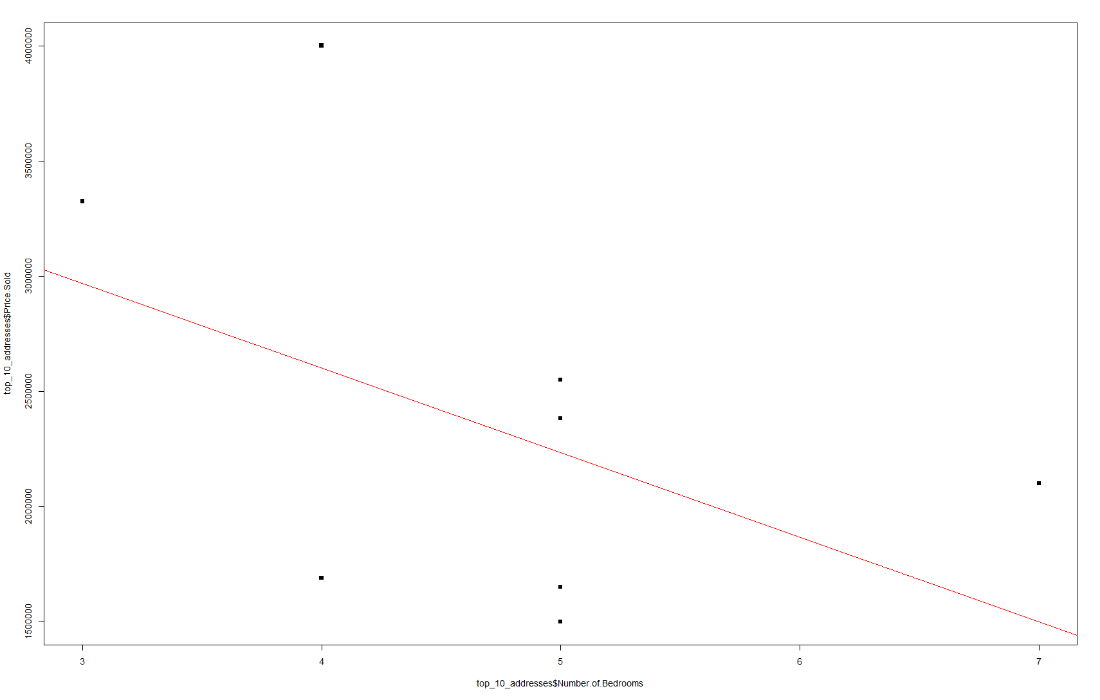
**Primary Goal of the project:**

* To analyze the housing data gathered for Cincinnati, OH, and gain an insight into the purchasing trends and preferences of people to help people moving into the city to make a sound decision.
* To achieve this, we have performed analysis into the housing data in R, Python, and Tableau.

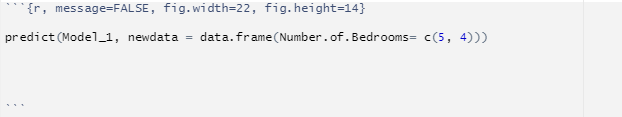
1. **Data Analysis in R**



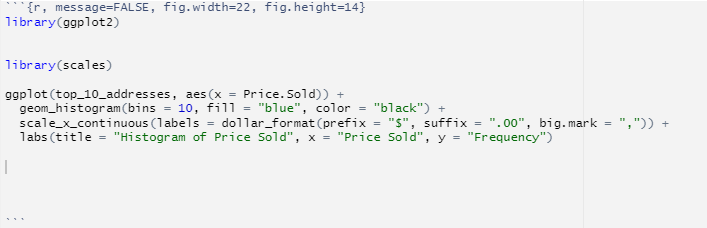
The above code creates a linear regression model that takes the number of bedrooms as the independent variable and its Price sold field as the dependent variable respectively. The resultant linear regression model is shown in the figure given below:



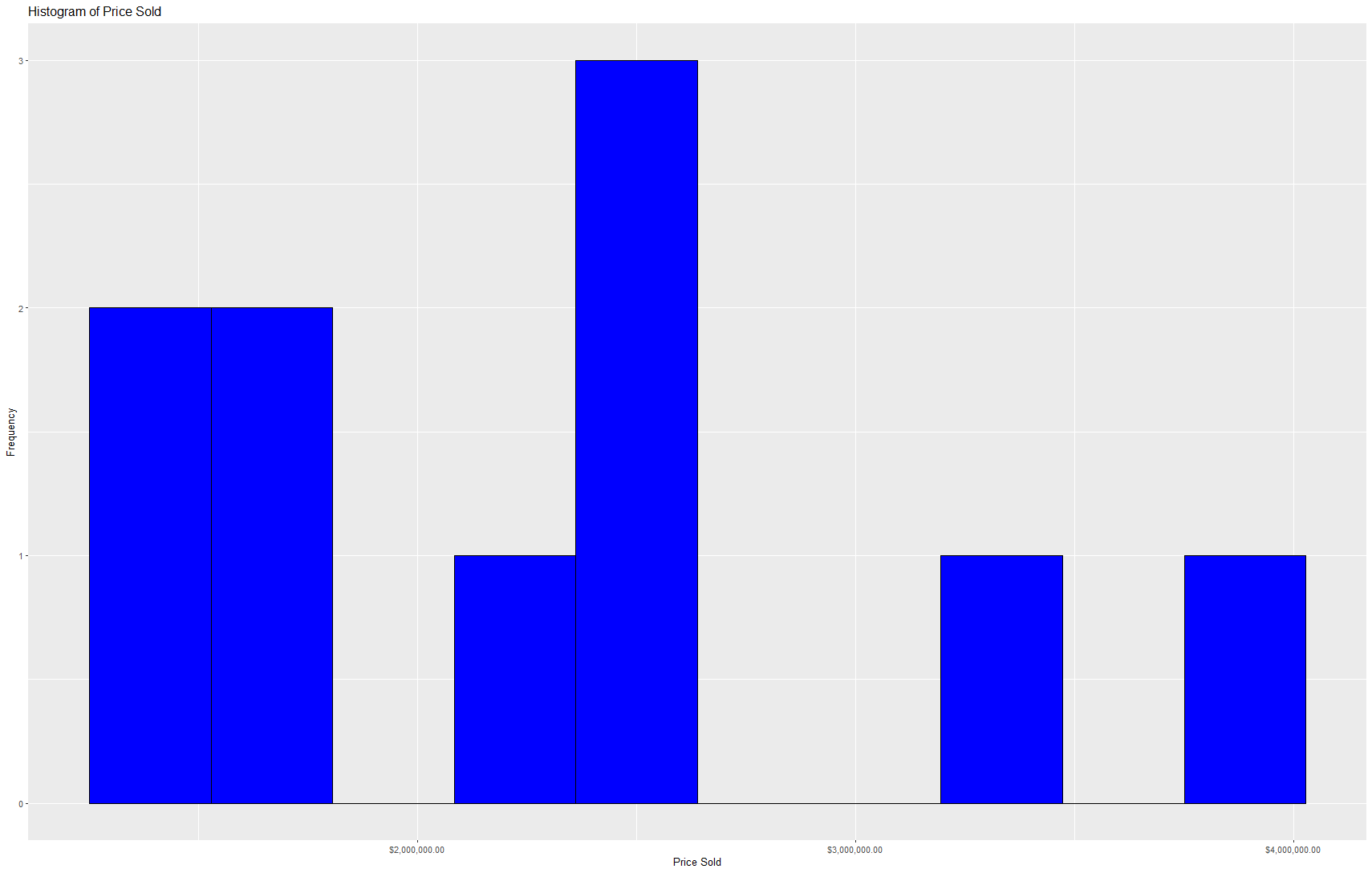
From the above linear regression analysis, we can infer that the slope formed because of linear regression line is negative. We have also predicted the price of bedrooms, using the linear regression model used above.



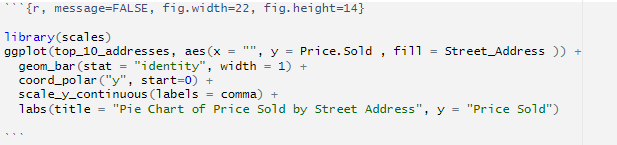
This code predicts the price value of the number of bedrooms, which are 4 and 5, respectively. The predicted price values of 5 and 4 bedrooms are respectively 2234741 and 2601793. Thus, real estate managers are required to manage the number of bedrooms in their estates, so that they can sell their estates at an affordable price to a lot of customers to increase their profits in the long run.



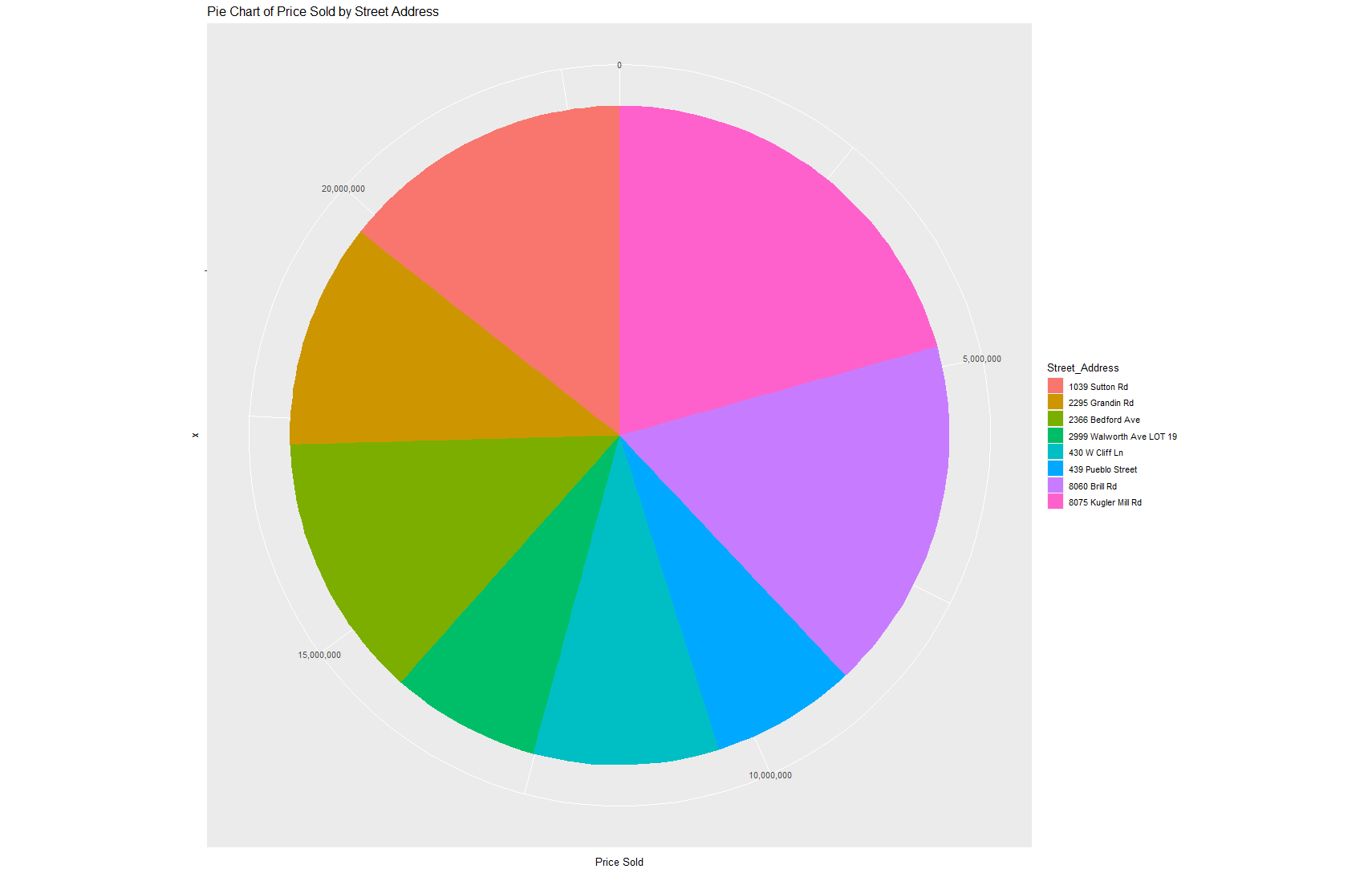
The above code will give us the histogram where frequency of price sold of different real estates will be provided.



From the above histogram analysis, it is important for real estate brokers across different regions of Ohio to sell their estates and houses between the ranges of 2 million to 3 million dollars to increase their profits.

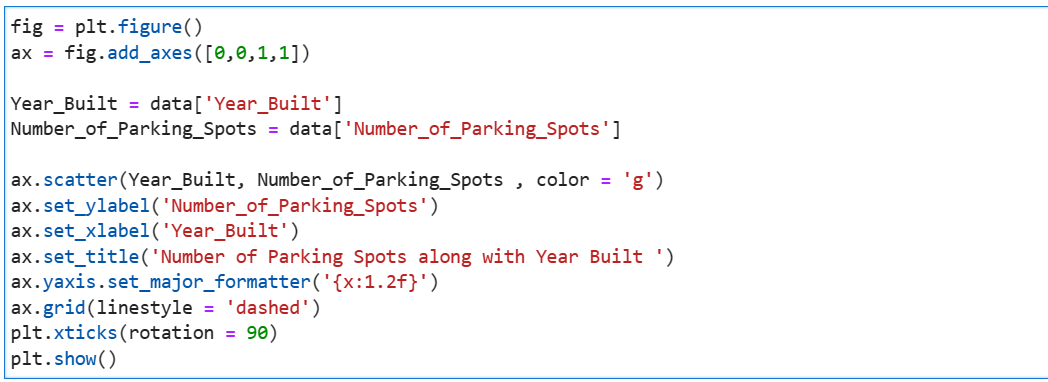


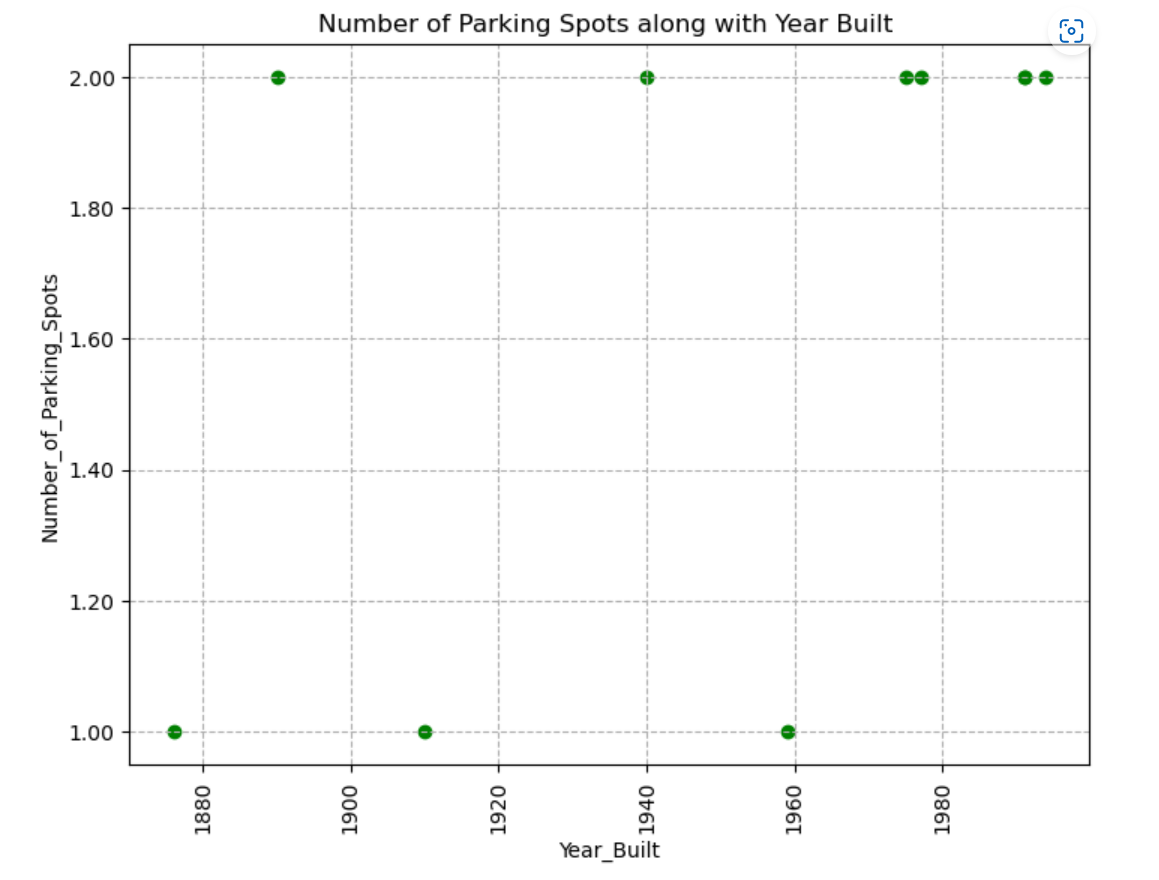
The above code gives an output of pie chart representation where Price sold with respect to different street addresses will be shown. The given result is shown below:



So, from the above pie chart analysis, 8075 Kugler Mill Road has sold estates at the highest prices, so real estate brokers and managers should invest heavily in building estates and houses in this street to boost their sales and profits respectively.

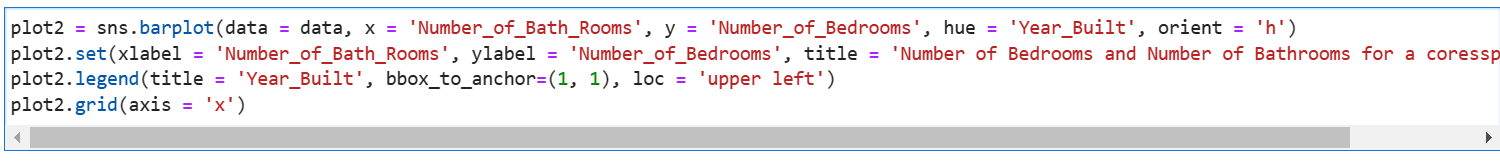
1. **Data Analysis in Python**

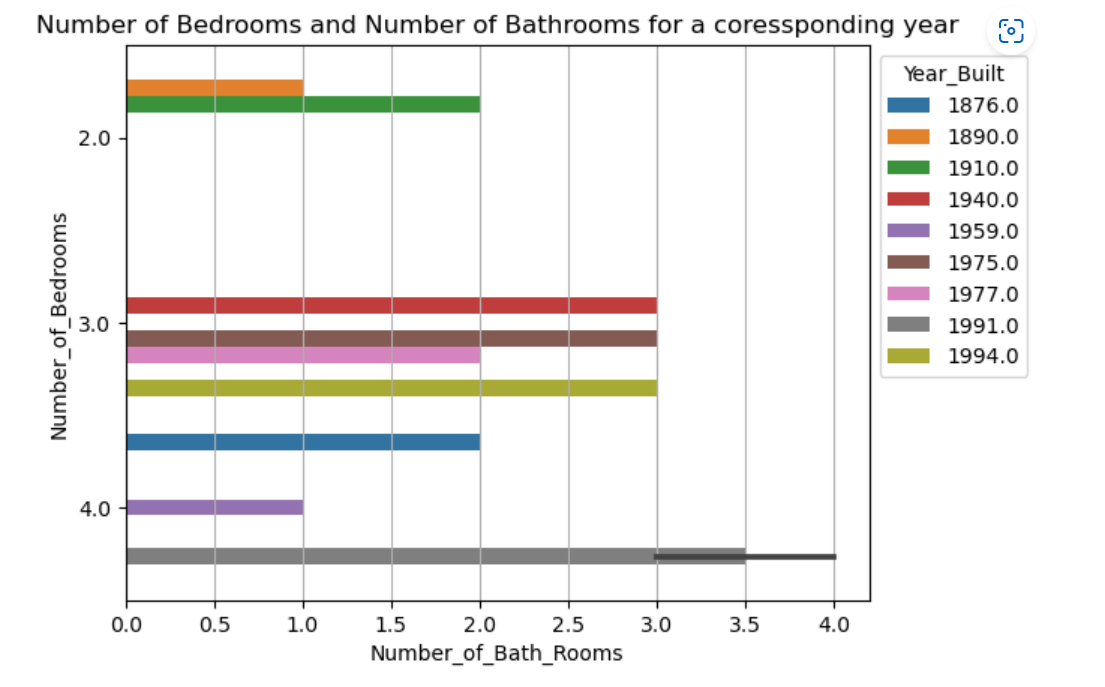




By plotting the trend of parking spaces built in a year, you were able to see the number of parking lots allocated for each house each year. This visualization could be useful for understanding how the demand for parking spaces has changed over time in the Cincinnati area, as well as identifying any trends or patterns that may exist in the data.

Overall, this kind of data analysis can be helpful for a variety of purposes, such as understanding real estate trends, identifying potential areas for improvement or investment, and more. Using Python to conduct data analysis is a powerful tool that allows you to efficiently and effectively process large amounts of information and create clear and concise visualizations to help communicate your findings.

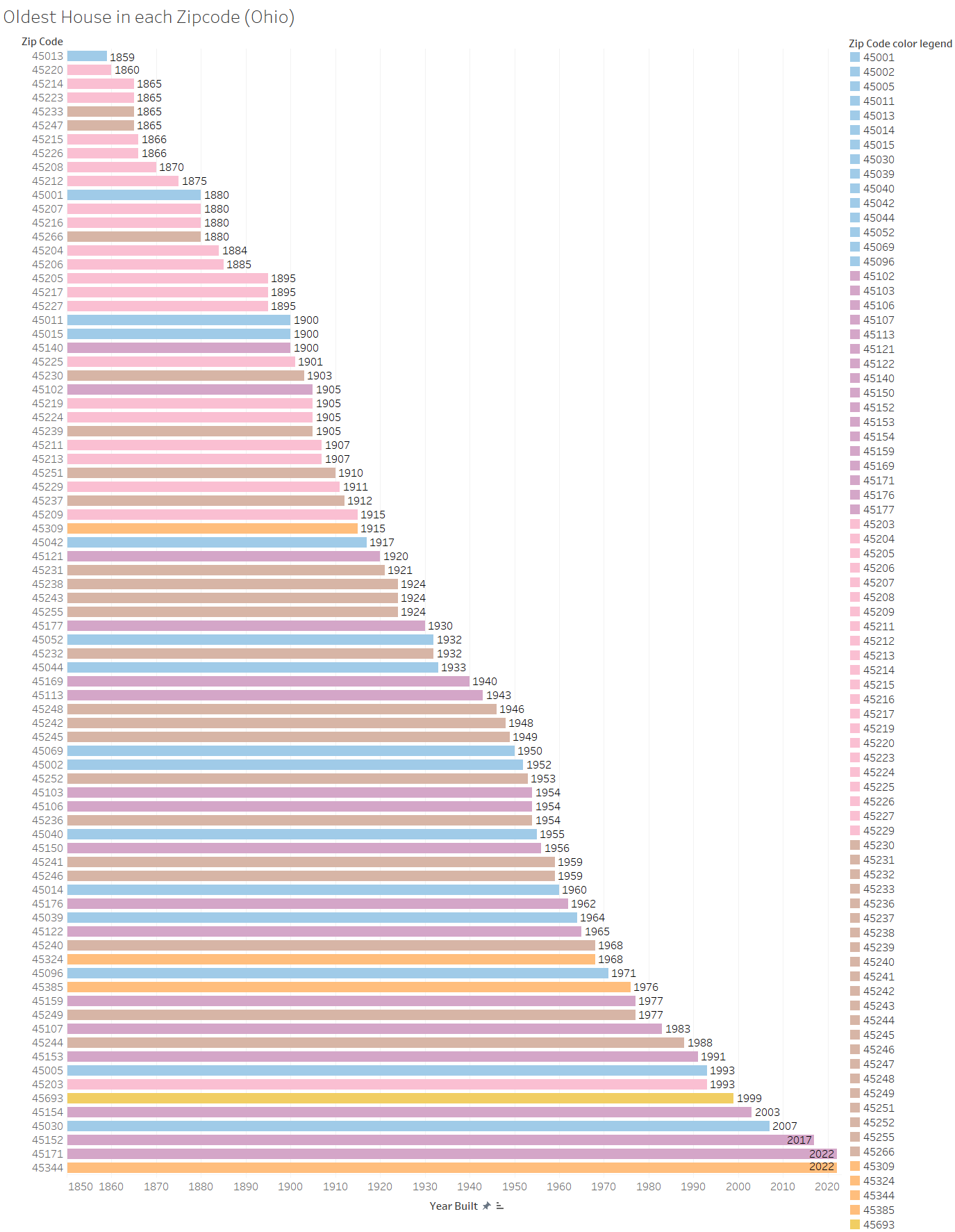




We conducted an analysis of the relationship between the number of bedrooms and number of bathrooms in the greater Cincinnati area for a given year, using data analytics techniques. By plotting this relationship, we were able to identify the number of bathrooms that were built for each bedroom in the past, as well as the factors that contributed to this trend.

This analysis can be useful for gaining insights into the housing market in Cincinnati, such as understanding the demand for different types of homes and how the number of bathrooms has changed over time in relation to the number of bedrooms. Identifying these trends can help inform future decision-making in the real estate industry, such as determining which types of properties are likely to be in high demand and therefore more valuable.

1. **Data Analysis in Tableau**



The first plot shown above represents the oldest house in each of the zip codes in the Greater Cincinnati area. And as seen from the visualization, we can see that the oldest house is in 45013 which comes under the Butler County.

The second plot shown below contains the analysis of the average house prices in the various Cincinnati zip codes. From the analysis and visualization, we can observe that the highest average price of houses is 45243 located in Hamilton County.

Chart, table, treemap chart

Description automatically generated