Property recommendations

# Introductions

Jack Bauer family is going to move to Pittsburgh. The family is recruiting a butler to help them make decisions. The tasks is:

1. A house. Jack Bauer family wants to buy a house. The requirements are:
2. The price is less than 500,000 USD.
3. It has investment potential.
4. Close to medical centers/hospitals, universities and supermarkets/malls (Target, Walmart, Whole Food, Costco, etc.).
5. Excellent traffic in surrounding areas.

To accomplish this task in this assignment, that data file containing Pittsburgh property price data that was provided accompanying the project description was utilised.

In the next section, an approach that was followed in accomplishing the recommendations is presented.

# Approach

Property data provided in the data file was analysed while taking into consecration, the requirements from the Bauer family as presented in the introduction above. In order for us to be able to build a model that could be used in analysing the data and providing proper recommendations, for the Bauer family, ratings were assigned to some of the records in the list of properties. These manually rated records were to be used as training data for our recommendation model.

Our approach in providing the rating was based on our view that, more expensive properties have a greater investment potential as it indicates that such properties attracts more interest. This therefore means that when an opportunity to sell it comes, it will not be challenging to get potential buyers.

This view led us to the top 51 expensive propertied within the $0 to $499999 bracket and assigning ratings on them. In addition to them having a greater potential as an investment, the higher prices can also serve as an indicator of a relative higher quality of the properties. These top 50 properties are initially assigned scores of 5 as they met the requirement of having an investment potential and the requirement of being below $500000. These score to be increased or decreased after analysing the distribution of the of traffic congestions and property price distribution on the map presented in <https://vietexob.shinyapps.io/traffic_real_estate> .

The maps utilised to adjust there scores shown in Figure 1 and Figure 2. Figure 1 depicts the listing price distribution while Figure 2 depicts traffic distribution measured by the average speed within a neighbourhood.

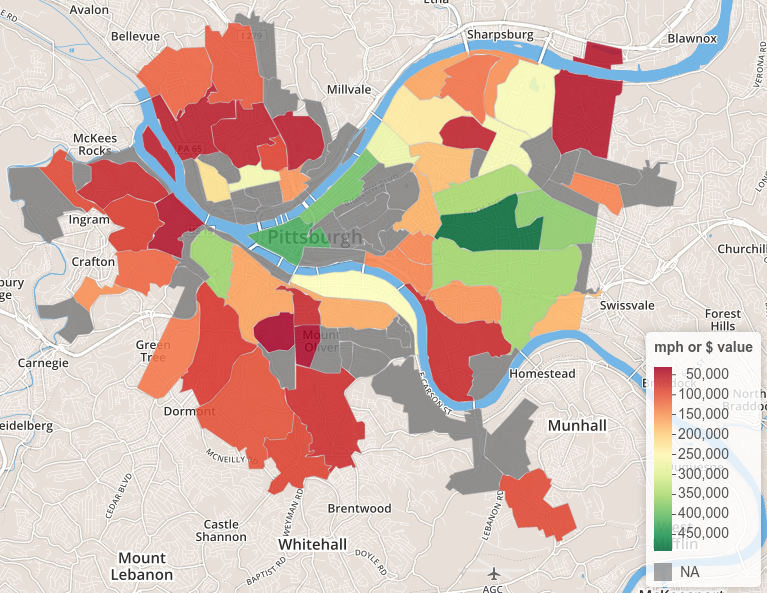


Figure : Median listing price

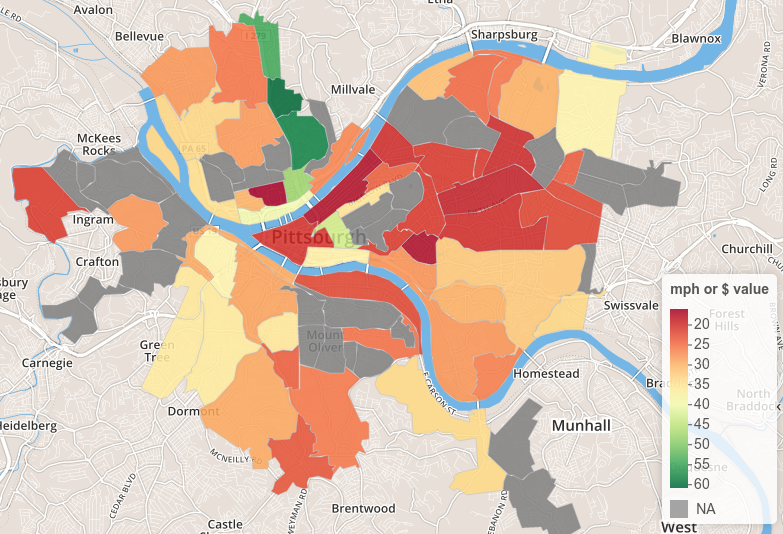


Figure : Average road speed distribution

From the score of 5 for each of the 51 properties, the score decreased or increased depending on whether the property is located in traffic congested neighborhood or not and also depending on the property is based in a neighborhood with lower median price listing or not. The 51 properties that were manually rated are presented in Table 1.

Table : Manually rated properties

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| location | street | neighborhood | price | rating |
| 371 S GRAHAM ST PITTSBURGH | S Graham St | Bloomfield | 475000 | 3 |
| 374 S PACIFIC AVE PITTSBURGH | S Pacific Ave | Bloomfield | 485750 | 2 |
| 1016 5TH AVE PITTSBURGH | 5th Ave | Bluff | 470000 | 5 |
| 1044 SAW MILL RUN BLVD PITTSBURGH | Saw Mill Run Blvd | Brookline | 450000 | 5 |
| 3421 BATES ST PITTSBURGH | Bates St | Central Oakland | 440000 | 7 |
| 806 PENN AVE UNIT 5 PITTSBURGH | Penn Ave | Cultural District | 449000 | 4 |
| 941 PENN AVE UNIT 902 PITTSBURGH | Penn Ave | Cultural District | 475000 | 6 |
| 467 WYOLA ST PITTSBURGH | Wyola St | Duquesne Heights | 440000 | 5 |
| 157 SWEETBRIAR VILLIAGE TRL PITTSBURGH | Sweetbriar Village Trl | Duquesne Heights | 445000 | 5 |
| 157 SWEETBRIAR VILLIAGE TRL PITTSBURGH | Sweetbriar Village Trl | Duquesne Heights | 445000 | 5 |
| 301 5TH AVE PITTSBURGH | 5th Ave | Fifth and Forbes Corridor | 450000 | 3 |
| 301 5TH AVE PITTSBURGH | 5th Ave | Fifth and Forbes Corridor | 460000 | 5 |
| 300 4TH AVE UNIT 2102 PITTSBURGH | 4th Ave | Fifth and Forbes Corridor | 475000 | 0 |
| 5401 GLENWOOD AVE PITTSBURGH | Glenwood Ave | Hazelwood | 475000 | 5 |
| 6126 CALLERY ST PITTSBURGH | Callery St | Highland Park | 450000 | 5 |
| 6333 JACKSON ST PITTSBURGH | Jackson St | Highland Park | 476500 | 6 |
| 943 WELLESLEY RD PITTSBURGH | Wellesley Rd | Highland Park | 479000 | 7 |
| 5710 CALLOWHILL ST PITTSBURGH | Callowhill St | Highland Park | 490000 | 8 |
| 1000 FARRAGUT ST PITTSBURGH | Farragut St | Highland Park | 490000 | 8 |
| 1220 W NORTH AVE PITTSBURGH | W North Ave | Manchester | 462500 | 5 |
| 307 S DITHRIDGE ST UNIT 802 PITTSBURGH | S Dithridge St | North Oakland | 466700 | 6 |
| 307 S DITHRIDGE ST UNIT 803 PITTSBURGH | S Dithridge St | North Oakland | 466700 | 6 |
| 703 GETTYSBURG ST PITTSBURGH | Gettysburg St | Point Breeze | 470300 | 5 |
| 245 CONOVER RD PITTSBURGH | Conover Rd | Point Breeze | 476545 | 5 |
| 6952 CLARIDGE PL PITTSBURGH | Claridge Pl | Point Breeze | 480000 | 5 |
| 131 YORKSHIRE DR PITTSBURGH | Yorkshire Dr | Point Breeze | 499000 | 5 |
| 4770 WALLINGFORD ST PITTSBURGH | Wallingford St | Shadyside | 447300 | 5 |
| 715 S AIKEN AVE PITTSBURGH | S Aiken Ave | Shadyside | 450000 | 6 |
| 242 EMERSON ST PITTSBURGH | Emerson St | Shadyside | 455000 | 9 |
| 246 EMERSON ST PITTSBURGH | Emerson St | Shadyside | 455000 | 6 |
| 586 MOORHEAD PL PITTSBURGH | Moorhead Pl | Shadyside | 480000 | 5 |
| 426 COVENTRY RD PITTSBURGH | Coventry Rd | Shadyside | 490000 | 4 |
| 19 S 18TH ST PITTSBURGH | S 18th St | Southside Flats | 445000 | 5 |
| 101 MERRIMAN MEWS PITTSBURGH | Merriman Mews | Southside Flats | 475000 | 5 |
| 101 MERRIMAN MEWS PITTSBURGH | Merriman Mews | Southside Flats | 475000 | 5 |
| 1113 E CARSON ST PITTSBURGH | E Carson St | Southside Flats | 480000 | 4 |
| 58 S 17TH ST PITTSBURGH | S 17th St | Southside Flats | 480000 | 4 |
| 2212 E CARSON ST PITTSBURGH | E Carson St | Southside Flats | 480000 | 4 |
| 105 MERRIMAN MEWS PITTSBURGH | Merriman Mews | Southside Flats | 482000 | 3 |
| 1123 WIGHTMAN ST PITTSBURGH | Wightman St | Squirrel Hill North | 443500 | 9 |
| 1660 MURRAY AVE UNIT 26 PITTSBURGH | Murray Ave | Squirrel Hill North | 450000 | 5 |
| 1536 ASBURY PL PITTSBURGH | Asbury Pl | Squirrel Hill North | 451000 | 5 |
| 1029 MURRAYHILL AVE PITTSBURGH | Murrayhill Ave | Squirrel Hill North | 479000 | 4 |
| 1031 MURRAYHILL AVE PITTSBURGH | Murrayhill Ave | Squirrel Hill North | 485000 | 5 |
| 1853 SHAW AVE PITTSBURGH | Shaw Ave | Squirrel Hill South | 443000 | 10 |
| 6740 FOREST GLEN RD PITTSBURGH | Forest Glen Rd | Squirrel Hill South | 460000 | 5 |
| 6429 BARTLETT ST PITTSBURGH | Bartlett St | Squirrel Hill South | 464000 | 5 |
| 6347 PHILLIPS AVE PITTSBURGH | Phillips Ave | Squirrel Hill South | 465000 | 5 |
| 2734 FERNWALD RD PITTSBURGH | Fernwald Rd | Squirrel Hill South | 489000 | 4 |
| 6219 NICHOLSON ST PITTSBURGH | Nicholson St | Squirrel Hill South | 497000 | 3 |
| 0 W NORTH AVE PITTSBURGH | North Ave | Unknown | 462500 | 5 |

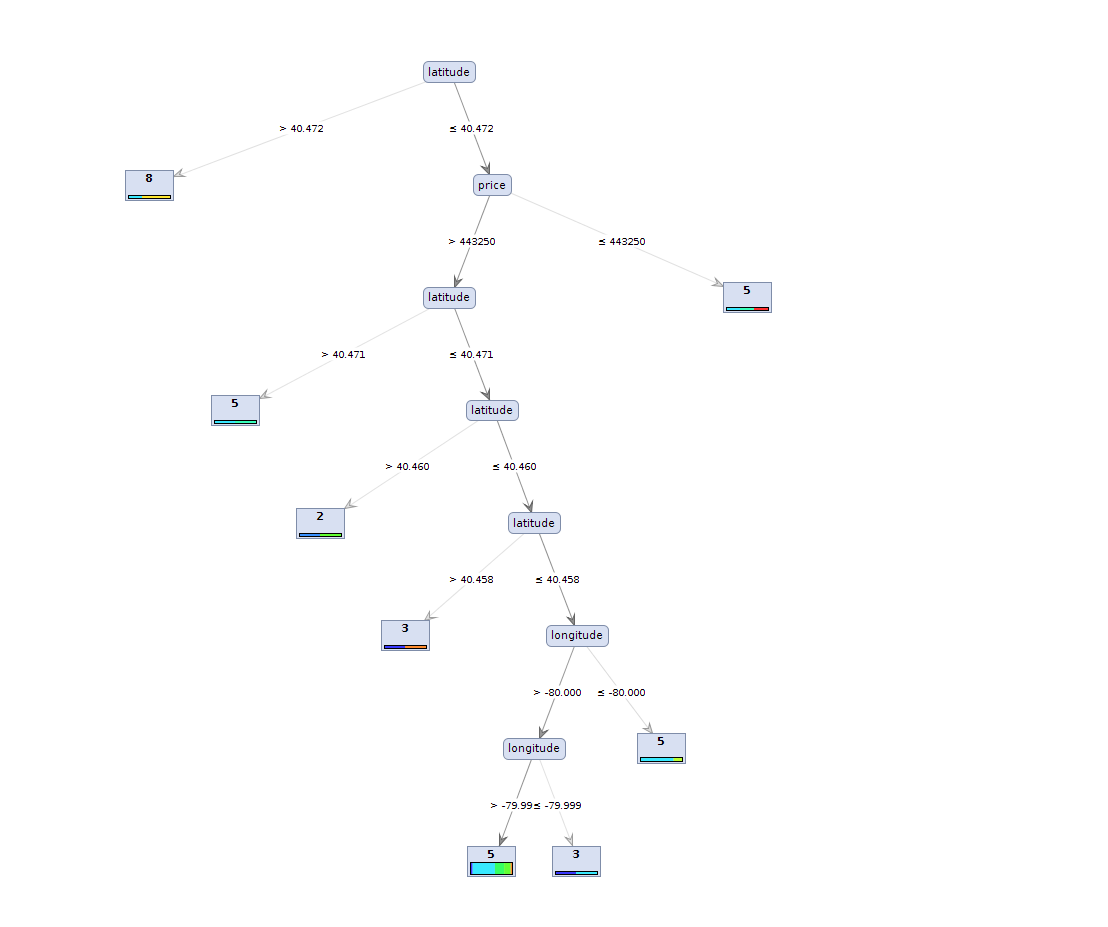


Figure : Decision Tree based on the training data

The decision tree can be read as follows: is a property is geographically placed below 40.472 on latitude and it is priced $443250, it is generally rated 5. This was obtained by running a process on RapidMiner with the data in Table 1. The developed model was found to have 54$ in accuracy. Since properties that would be recommended to the Bauer family are of the highest rating, Figure 4 shows predicted ratings with the highest one being the 8.

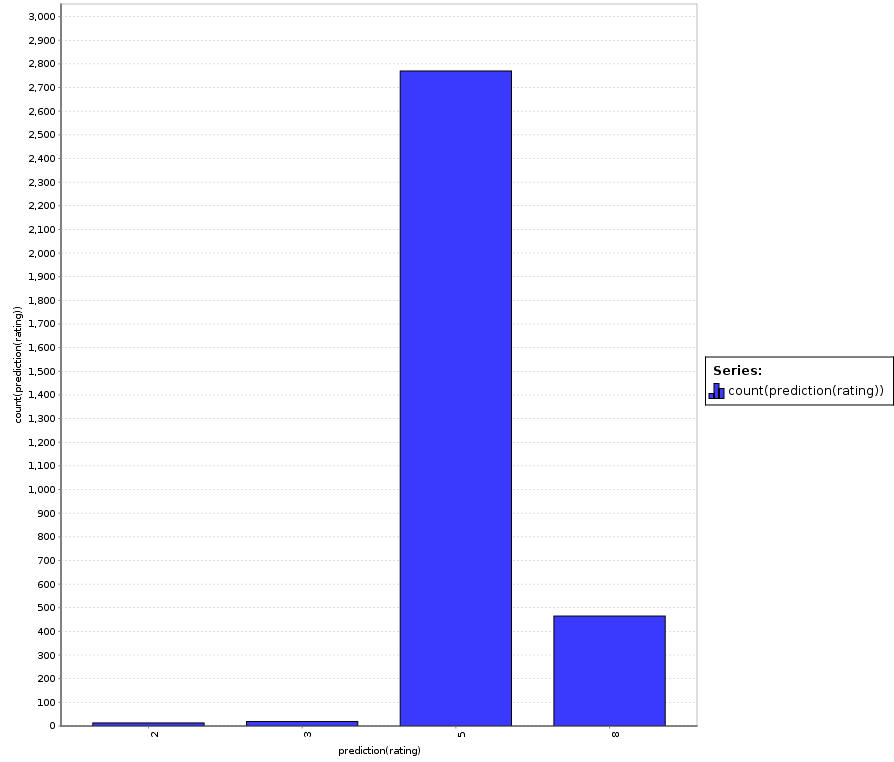


Figure : Count per predicted rating

Now that the rating that the rating that to be used for recommendation is 8, properties with this rating in Pittsburgh were found to be located in Upper Lawrenceville, Summer Hill, Stanton Heights, Perry North, Marshall-Shadeland, Highlands Park and Bloemfield neighborhoods as seen in Figure 5.

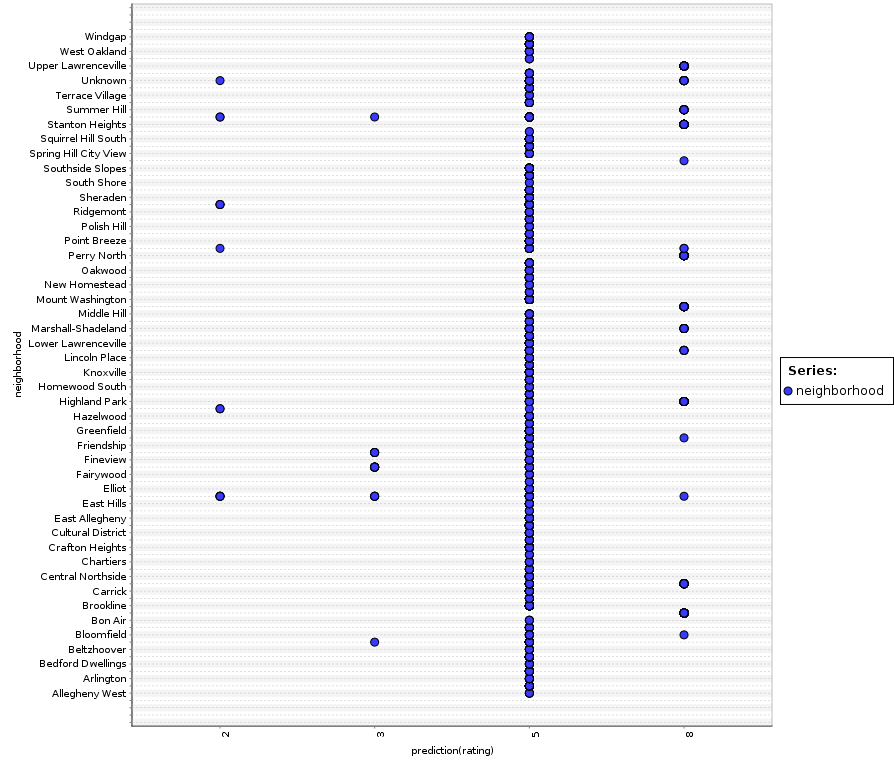


Figure : Predicted rating per neighborhood

# Conclusion and recommendations.

In this project conducted research and analysis of property data from Pittsburgh with the aim to recommend properties for the Bauer family based on the requirements that they had provided which are:

a) The price is less than 500,000 USD.

b) It has investment potential.

c) Close to medical centers/hospitals, universities and supermarkets/malls .

d) Excellent traffic in surrounding areas.

These requirements were taken into consideration which developing the prediction model. The prediction model was tested and was found to have 54% accuracy which means its predictions can be trusted.

The model therefore predicted properties from Upper Lawrenceville, Summer Hill, Stanton Heights, Perry North, Marshall-Shadeland, Highlands Park and Bloemfield neighborhoods. These predictions can be compared with the price listings and traffic congestions as depicted in Figure 1 and Figure 2 respectively. Summer Hill for example is one of the best neighborhoods in terms of traffic congestion.