|  |
| --- |
| DATA SCIENCE CAPSTONE PROJECT  ANALYSIS OF RICHMOND REAL ESTATE MARKET  IN RELATION TO  FOURSQUARE VENUES |
| KEVIN A. DILL KEVINDILL2001@GMAIL.COM |



## CAPSTONE PROJECT:

## IBM DATA SCIENCE PROFESSIONAL CERTIFICATE

### **1. INTRODUCTION**

#### 1.1 Background

Moving to an unfamiliar area in a short period of time can be a daunting task. In order to facilitate the search for housing, assessing market conditions and identifying locations of the necessary and desired amenities becomes a critical path. The constraints of the shortened time frame do not allow for extended exploratory visits. Being able to narrow the areas to conduct the search makes efficient use of time available when visiting the area.

#### 1.2 Problem

Due to the requirements of an offer of employment received by my wife, requiring a relocation to the Richmond Virginia Metropolitan Area (RVA), we found ourselves in just such a situation. This project aims to reduce the search for a suitable area to one or two zip codes that meet the requirements for amenities, commuting, and real estate pricing.

#### 1.3 Interest

As chance would have it, this situation came about at approximately the same time as the need for a final project for the IBM Professional Certificate in Data Science. Others who find themselves under similar duress may be able to use the methods here to facilitate their own move.

### **2. DATA SOURCES & ANALYSIS**

#### 2.1 Data Sources

The single requirement for data for this project is that it needs to include data from [FourSquare.com](https://foursquare.com/city-guide). Additional data sources were allowed to complete the project as necessary.

Foursquare makes data available through a webservice once an account has been established. The additional data for this project is provided by [Realtor.com](https://www.realtor.com/). Realtor.com provides a data file which is updated monthly. This file is accessible without cost and contains real estate data broken down by zip codes. Other files are available for data at the county and state level, but these were not used. These files are located [here](https://www.realtor.com/research/data/).

#### 2.2 Data cleaning and wrangling

The process began by downloading and processing the real estate data. Based on a cursory visit to the area, initial zip code areas identified for search were Richmond, Midlothian, and Chesterfield. The initial filter produced 17 zip codes for search. The data columns selected for inclusion into the search are:

*'postal\_code','zip\_name', 'median\_listing\_price', 'average\_listing\_price', 'new\_listing\_count', 'active\_listing\_count', 'total\_listing\_count', 'price\_reduced\_count', 'median\_days\_on\_market'*

Distance from the job location to the center point of each zip code area was calculated based on the geo coordinates of the respective locations. None of the initial areas selected would be excluded for distance. The decision was made to increase the number of zip code areas in an iterative process. The final list of zip code areas is: Richmond, Midlothian, and Chesterfield, Henrico, Glen Allen, Ashland, Chamberlayne, and Mechanicsville. This produced a total of 27 zip code areas. All of which still fell within the preferred 20-mile radius.

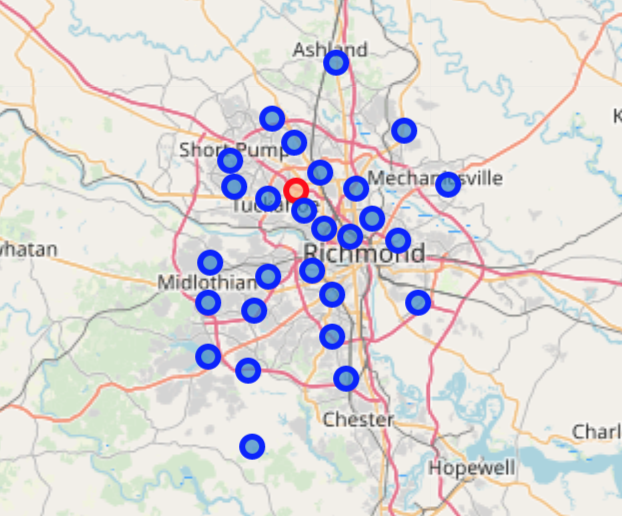
With the coordinates for each zip code, a call to the Foursquare API was made to download all venues in the zip code area within a 5-kilometer radius of the center point. This produced a data set with 2394 venues in 222 different categories. The initial list of venue categories for sorting was Grocery Stores, Veterinarian Services and Medical Services. This produced a list of 49 venues, all of which were Grocery Stores. A search of the raw data was conducted using a standard text editor for terms that may be used rather than the ones that had been chosen. Apparently Foursquare data does not contain entries for medical facilities for either humans or animals. A surrogate list of amenities was developed to sort neighborhoods on desirability, including: Grocery Store, Supermarket, Drugstore, Pharmacy, Market, Pet Store, Coffee Shop, Wine Shop, Gym/Fitness Center, Salon/Barbershop, Gym, Farmers Market, Food & Drink Shop, Performing Arts Venue, Museum, Theater, Art Museum, Dog Run, Social Club, State/Provincial Park and Lake. This produced a total of 379 different venues.

The venue data was further processed to determine the number of each venue per zip code, as well as the most common venue for the desired venues. These values were compared to the most common venue across all venues for each area examined.

The real estate data was examined for potential relationships to price. Among the relationships explored were distance, venue type, listing count and days on market. The results yielded little in terms of strong correlations.

#### 2.3 Results and Analysis

In looking at the distribution of zip codes that were chosen a basic map was created and displayed below. The red dot is the location of the potential job and the surrounding blue points are the centers of the various zip code areas. The area of exploration is centered in RVA, not the job location itself.



**Venue Counts From The Final List Of Selected Venues For All Zip Code Areas Combined**

|  |  |
| --- | --- |
| **VENUE TYPE** | **COUNT** |
| Art Museum | 3 |
| Coffee Shop | 80 |
| Dog Run | 4 |
| Farmers Market | 10 |
| Food & Drink Shop | 3 |
| Gym | 73 |
| Market | 5 |
| Museum | 4 |
| Performing Arts Venue | 6 |
| Pet Store | 14 |
| Pharmacy | 45 |
| Salon / Barbershop | 20 |
| Social Club | 1 |
| State / Provincial Park | 1 |
| Supermarket | 95 |
| Theater | 3 |
| Wine Shop | 16 |
| Art Museum | 3 |
| Coffee Shop | 80 |
| Dog Run | 4 |

As part of the wrangling process, I merged Grocery Stores and Supermarkets, as well as Drugstores and Pharmacies, and Gyms and fitness centers. Looking at the venue counts, the limitations of the data set start to become apparent. In a metropolitan area of over 1.2 million people one would expect to find more than 95 grocery stores, 76 coffee shops, and 16 wine shops. A quick manual search lead to an [article](https://www.richmond.com/food-drink/survey-richmond-ranks-near-top-in-number-of-grocery-stores/article_4dca25ea-7283-11e5-af9f-573ee0494952.html) from 2015, where it was stated that Richmond had 122.56 grocery retailers per 100,000 residents. Further, the lack of data on medical and veterinarian facilities is troublesome; I would have expected to at least find urgent care centers.

**Most Common Venues of Interest**

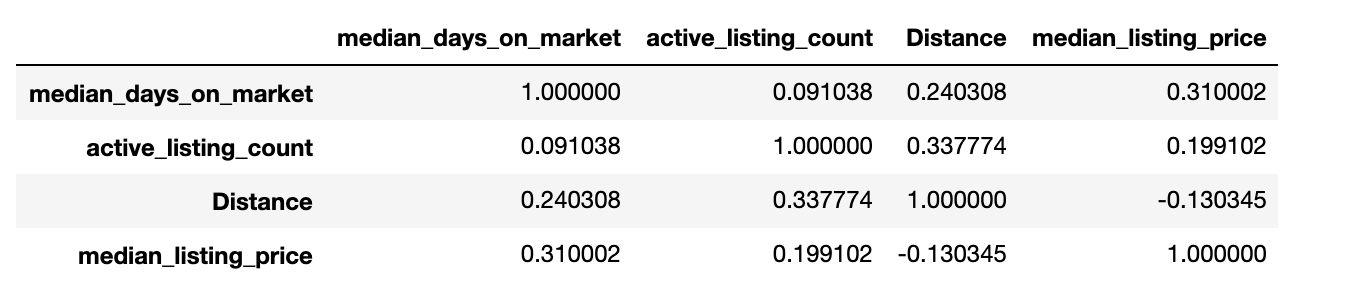
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ZipCode** | **1st Most Common Venue** | **2nd Most Common Venue** | **3rd Most Common Venue** | **4th Most Common Venue** | **5th Most Common Venue** |
| 23005 | Coffee Shop | Pharmacy | Supermarket | Wine Shop | Salon / Barbershop |
| 23059 | Supermarket | Coffee Shop | Gym | Pharmacy | Salon / Barbershop |
| 23060 | Supermarket | Coffee Shop | Wine Shop | Salon / Barbershop | Pharmacy |
| 23111 | Supermarket | Gym | Coffee Shop | Salon / Barbershop | Pharmacy |
| 23112 | Gym | Coffee Shop | Supermarket | Pet Store | Pharmacy |
| 23113 | Coffee Shop | Gym | Supermarket | Salon / Barbershop | Pharmacy |
| 23114 | Coffee Shop | Pharmacy | Supermarket | Gym | Salon / Barbershop |
| 23116 | Gym | Supermarket | Coffee Shop | Salon / Barbershop | Pharmacy |
| 23220 | Coffee Shop | Gym | Performing Arts Venue | Theater | Art Museum |
| 23221 | Coffee Shop | Gym | Supermarket | Museum | Art Museum |
| 23222 | Coffee Shop | Gym | Performing Arts Venue | Theater | Food & Drink Shop |
| 23223 | Gym | Performing Arts Venue | Museum | Coffee Shop | Pharmacy |
| 23224 | Pharmacy | Supermarket | Coffee Shop | Farmers Market | Gym |
| 23225 | Coffee Shop | Supermarket | Farmers Market | Wine Shop | Art Museum |
| 23226 | Coffee Shop | Supermarket | Gym | Wine Shop | Salon / Barbershop |
| 23227 | Gym | Coffee Shop | Pharmacy | Supermarket | Theater |
| 23228 | Supermarket | Coffee Shop | Wine Shop | Gym | Pet Store |
| 23229 | Supermarket | Wine Shop | Coffee Shop | Salon / Barbershop | Pharmacy |
| 23231 | Coffee Shop | Pharmacy | Supermarket | Gym | Salon / Barbershop |
| 23233 | Supermarket | Coffee Shop | Gym | Wine Shop | Farmers Market |
| 23234 | Supermarket | Pharmacy | Gym | Theater | Art Museum |
| 23235 | Coffee Shop | Supermarket | Gym | Pharmacy | Wine Shop |
| 23236 | Gym | Supermarket | Coffee Shop | Wine Shop | Salon / Barbershop |
| 23237 | Supermarket | Pharmacy | Gym | Salon / Barbershop | Theater |
| 23238 | Supermarket | Pharmacy | Gym | Wine Shop | Coffee Shop |

**Most Common Venues – All**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Zip Code** | **1st Most Common Venue** | **2nd Most Common Venue** | **3rd Most Common Venue** | **4th Most Common Venue** | **5th Most Common Venue** |
| 23005 | Fast Food Restaurant | Gas Station | Pizza Place | Sandwich Place | Hotel |
| 23059 | American Restaurant | Hotel | Mexican Restaurant | Supermarket | Convenience Store |
| 23060 | Supermarket | Pizza Place | Mexican Restaurant | Hotel | Sushi Restaurant |
| 23111 | Mexican Restaurant | Italian Restaurant | Sandwich Place | American Restaurant | Gym |
| 23112 | Italian Restaurant | Mexican Restaurant | American Restaurant | Gym | Coffee Shop |
| 23113 | Gym | Coffee Shop | Supermarket | Mexican Restaurant | Fast Food Restaurant |
| 23114 | Coffee Shop | Clothing Store | Supermarket | Pharmacy | Mexican Restaurant |
| 23116 | Gym | Supermarket | Italian Restaurant | Café | Convenience Store |
| 23220 | American Restaurant | Coffee Shop | Breakfast Spot | Sandwich Place | Pizza Place |
| 23221 | Brewery | American Restaurant | Coffee Shop | Greek Restaurant | Pizza Place |
| 23222 | Coffee Shop | Gym | Brewery | American Restaurant | BBQ Joint |
| 23223 | Seafood Restaurant | Bakery | Sandwich Place | Southern / Soul Food Restaurant | Gym |
| 23224 | Discount Store | Convenience Store | Pharmacy | Sandwich Place | Supermarket |
| 23225 | Coffee Shop | Park | Supermarket | American Restaurant | Mexican Restaurant |
| 23226 | Coffee Shop | American Restaurant | Supermarket | Brewery | Mexican Restaurant |
| 23227 | Gym | Convenience Store | Brewery | Pizza Place | Discount Store |
| 23228 | Vietnamese Restaurant | Mexican Restaurant | Sandwich Place | Convenience Store | Supermarket |
| 23229 | American Restaurant | Supermarket | Italian Restaurant | Mexican Restaurant | Coffee Shop |
| 23231 | Rental Car Location | Discount Store | Sandwich Place | American Restaurant | Gas Station |
| 23233 | Supermarket | American Restaurant | Department Store | Sandwich Place | Cosmetics Shop |
| 23234 | Discount Store | Convenience Store | Fast Food Restaurant | Supermarket | Pharmacy |
| 23235 | Mexican Restaurant | Coffee Shop | Supermarket | Gym | Fast Food Restaurant |
| 23236 | Mexican Restaurant | Department Store | Gym | Clothing Store | Pizza Place |
| 23237 | Discount Store | Convenience Store | Supermarket | Fast Food Restaurant | Pharmacy |
| 23238 | Supermarket | Mexican Restaurant | Sandwich Place | American Restaurant | Pizza Place |
| 23832 | Pizza Place | Convenience Store | American Restaurant | Gym | Supermarket |

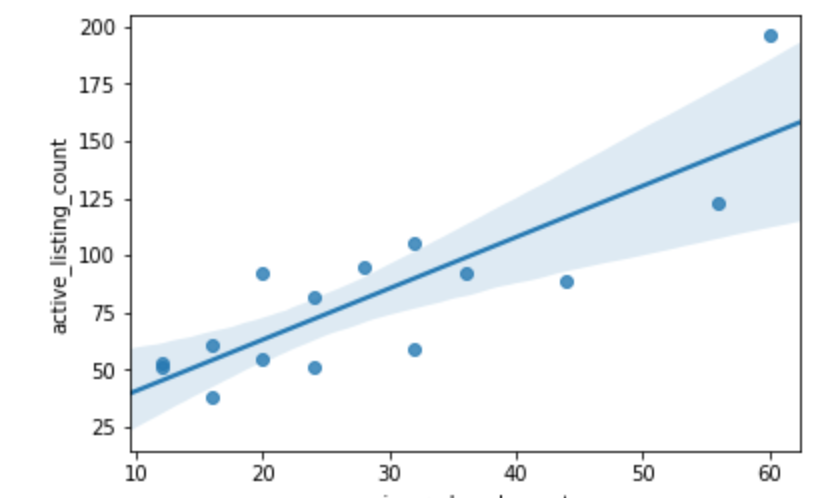
In comparing the frequency of the desired items to all items it would seem that Richmond, like many other cities, has a strong affinity for coffee. Other items such as fast food and convenience stores are expected to be frequent. Of note, however, is the listing of some eating places as “joints.”

**Correlation Table for Selected Data**



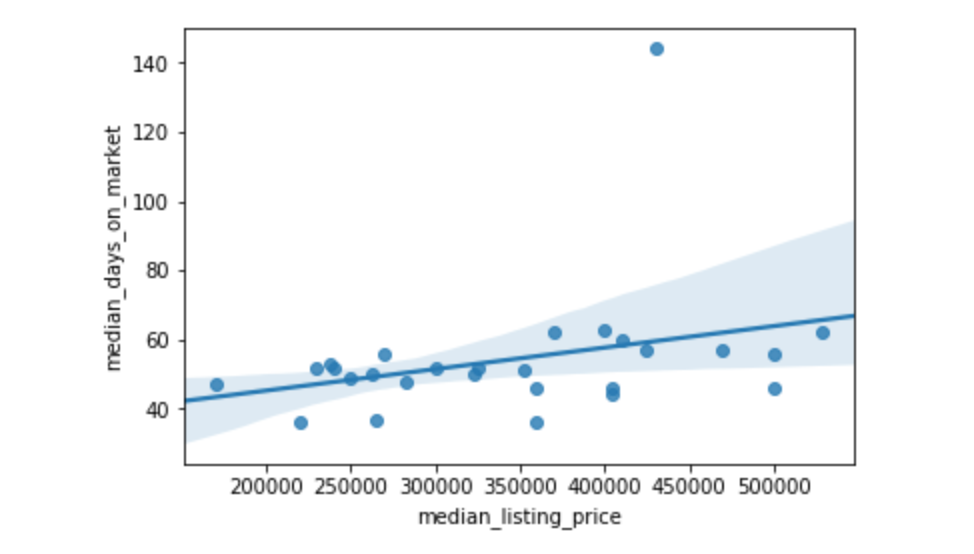
In trying to find a strong correlation between various attributes and either price or distance, I ran across a fairly homogenous data set. It is difficult to tell if this is an artifact of the limited data on the Richmond area in Foursquare, or if there is indeed no relationship. In limiting the data to areas with a median list price of less than $360,000, a correlation between active listing and reduced prices did appear as shown below.

**Relationship of Active Listings to Reduced Prices in Areas with Median Price Less Than $360K**

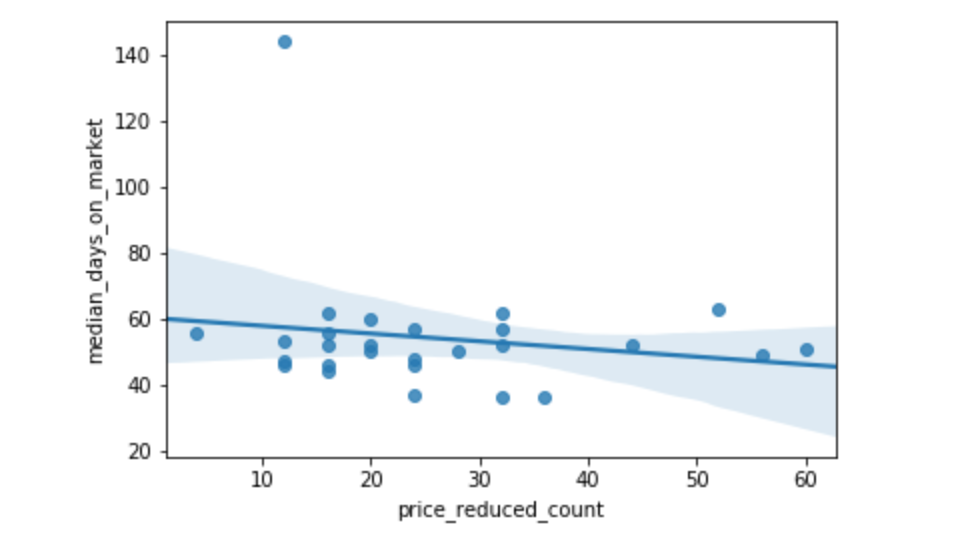


Unfortunately, this only includes 15 of the 27 areas examined. Otherwise most relationships between the data and prices was fairly tenuous at best. The selected charts below are exemplar of the data set in full.

**Relationship of Days on Market to Listing Price**



**Relationship of Days on Market to Reduced Price Count**



**General Market Snapshot**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Zip Code** | **Display Name** | **Distance(Mi)** | **Median Listing price** | **Active Listing Count** |
| 23224 | Richmond City | 8.31 | $170,050 | 53 |
| 23228 | Lakeside | 2.24 | $220,050 | 59 |
| 23222 | Richmond City | 5.99 | $230,045 | 89 |
| 23237 | Chesterfield County | 14.54 | $237,500 | 51 |
| 23234 | Chesterfield County | 11.27 | $240,000 | 61 |
| 23223 | East Highland Park | 8.52 | $249,950 | 123 |
| 23231 | Henrico County | 12.38 | $262,780 | 95 |
| 23225 | Richmond City | 6.13 | $265,050 | 51 |
| 23227 | Henrico County | 4.49 | $270,040 | 38 |
| 23235 | Bon Air | 6.84 | $283,357 | 82 |
| 23236 | Chesterfield County | 9.49 | $300,000 | 92 |
| 23114 | Chesterfield County | 10.73 | $323,000 | 55 |
| 23111 | Mechanicsville | 11.44 | $325,050 | 105 |
| 23112 | Chesterfield County | 14.22 | $352,040 | 196 |
| 23060 | Henrico County | 3.63 | $359,550 | 92 |
| 23229 | Tuckahoe | 2.17 | $360,000 | 56 |
| 23832 | Chesterfield County | 14.03 | $369,950 | 125 |
| 23116 | Mechanicsville | 9.22 | $400,000 | 145 |
| 23233 | Henrico County | 5.50 | $404,500 | 76 |
| 23220 | Richmond City | 5.34 | $405,050 | 60 |
| 23238 | Henrico County | 4.67 | $410,050 | 62 |
| 23838 | Chesterfield County | 19.59 | $425,050 | 82 |
| 23005 | Ashland | 10.06 | $430,800 | 76 |
| 23059 | Hanover County | 5.79 | $470,040 | 201 |
| 23226 | Henrico County | 1.59 | $499,950 | 38 |
| 23221 | Richmond City | 3.54 | $500,000 | 27 |
| 23113 | Chesterfield County | 8.49 | $529,050 | 123 |

### **3. Conclusions and Next Steps**

Unfortunately, the data set does not provide the insight into the area that was hoped for. In further exploration of the Foursquare data, it appears to only be useful in a small number of very large metropolitan areas, of which Richmond is not one. In order to get more useful information related to availability of goods and services, a different data provider will have to be selected. The real estate data, while useful, may not be sufficiently granular to be helpful in narrowing the search; neighborhoods within zip code areas can be very differently priced and have access to rather different venues. The fact that all the areas under consideration where within a 20-mile radius re-enforces the need for more granular data than was obtained.