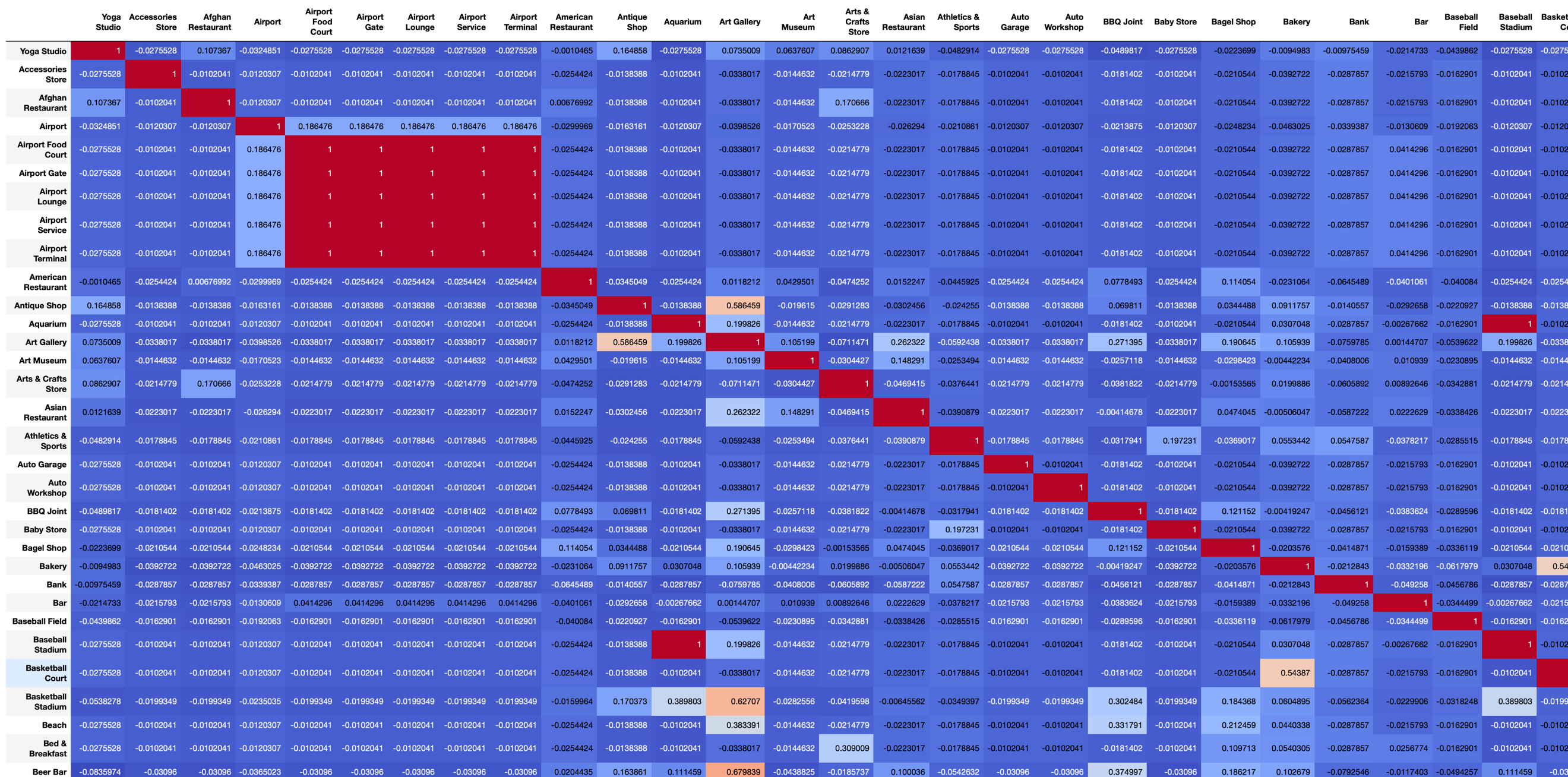
1. **Introduction**

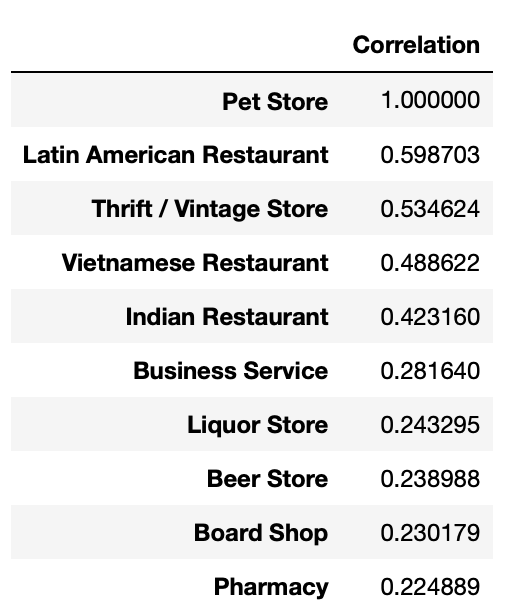
An entrepreneur is interested in opening a pet store in Toronto. They would like to determine what is the best neighborhood for the pet store. The entrepreneur does not want to simply open up the pet store by looking at neighborhoods that currently have a high number of existing pet stores as the supply in those existing neighborhoods may exceed current demands. Therefore, the goal is to identify neighborhoods that have characteristics that are suitable for pet stores but currently is undersupplied. They believe that the best indicators for identifying the location are the type of venues that are situated in a particular neighborhood. In particular, it is assumed that the types of venues in a particular neighborhood act as a proxy to represent the local demographic, including their needs and interests. In this project, we will determine which type of venues are often established in conjunction with pet stores. Once we understand what the top type of venues are, we will proceed to locate neighborhoods that have these venues as potential targets for the new pet store.

1. **Data** 
   1. We will use data from Wikipedia to define the neighborhoods in Toronto.
   2. We will use foursquare data to build a list of the venues in each neighborhood in Toronto.
2. **Methodology**
   1. In the first stage we will attempt to determine the relationship between Pet Stores and other types of venues. This is accomplished by first counting the number of venues for each type of venue in each neighborhood in Toronto. We will then calculate the correlation between each type of venue with Pet Stores. The top 5 venues with the highest correlation to the instances of Pet Stores will therefore be identified as the top most related venues, in other words, these types of venues represent the demographic who are likely suitable customers of our new pet store.
   2. In the second stage, we will identify the neighborhoods that fit the characteristics above best. This is accomplished by the counting the instances of the top venue types as determined above, and applying a weighting to that count. The weighting is the correlation calculated above. The weighted count for relevant venues is stored as the “weighted score” for each neighborhood in Toronto. The weighted score will be sorted to identify the neighborhoods that best fits our targeted characteristics.
   3. In the final stage, we will examine the top neighborhoods and eliminate the neighborhoods with a high number of existing pet stores. The remaining neighborhoods will then be the suitable neighborhoods for the new pet store.
3. **Results**
   1. In the first stage, the correlation between all types of venues were determined.

The correlation matrix is presented below:



The venues with the highest correlation to instances of Pet Stores are:



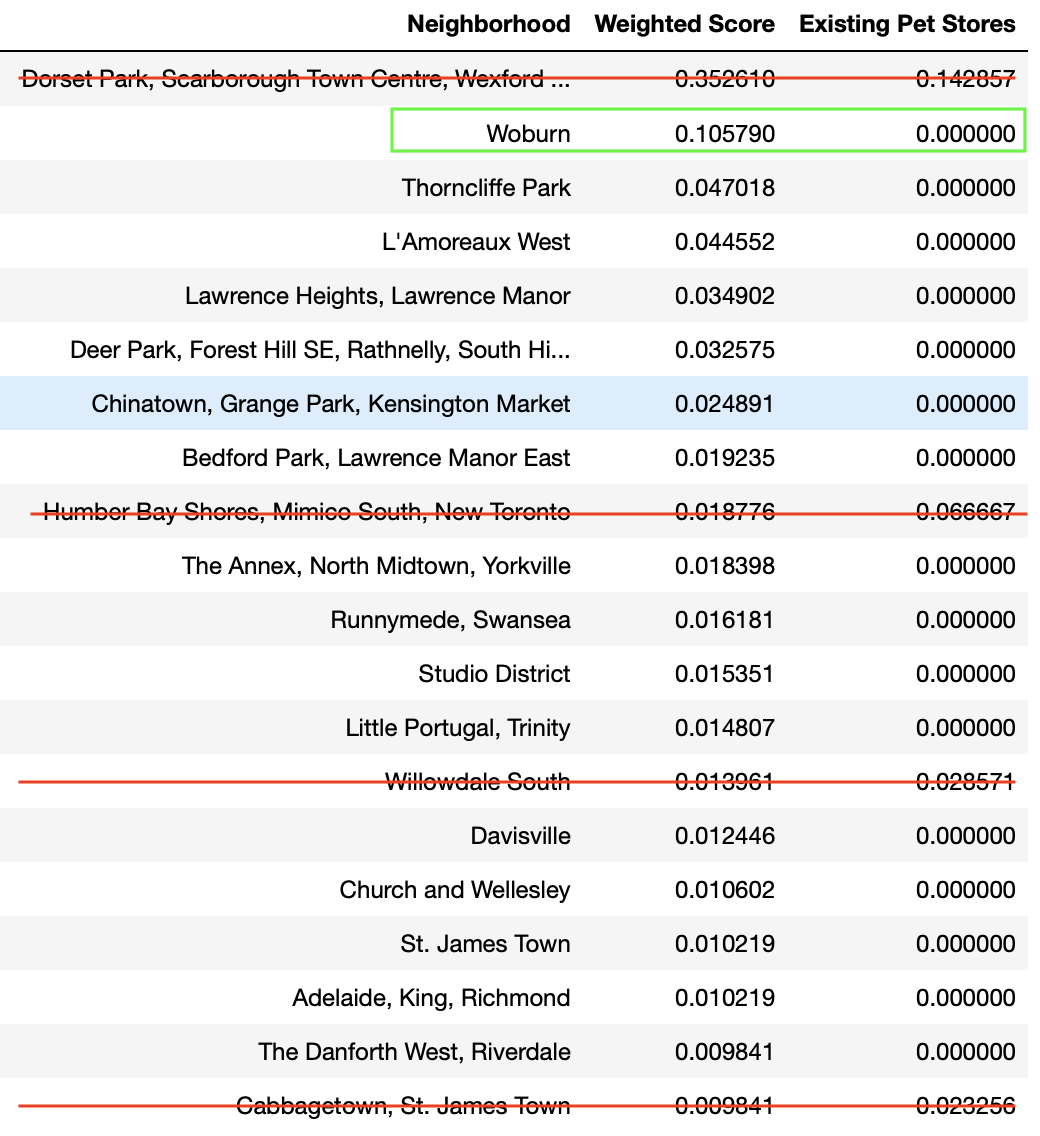
We use these results as a representation for the interests of the demographic that will be suitable for a pet store.

* 1. In the second stage, the neighborhoods with the highest weighted score are:



The weighted score takes into consideration the top 5 venue types that are related to the instances of pet stores.

* 1. In the final stage, we examine the neighborhoods with the best fit the target characteristics for the new pet stores and eliminate the neighborhoods that currently already have a high number of pet stores. The first neighborhood on our list, “Dorset Park Scarborough Townscentre, Wexford …”, already has a high number of current pet stores. Therefore, we move onto the second neighborhood on our list, “Woburn”, and determine that this is the second most suitable neighborhood and the ideal location for a new store as it has a count of current pet stores.



1. **Discussion**

The top result from 4.3, “Dorset Park Scarborough Townscentre, Wexford …”, has a high weighted score and also exhibits a high count for existing pet stores, this is expected as neighborhoods that current exhibit characteristics suitable for pet stores should have many pet stores. This result supports the assumptions and basis behind our methodology.

In this project, we utilized instances of venues as a proxy for the demographic in a particular neighborhood. This approach incorporates a number of major assumptions such as these venues are successful and that each instance of a venue will support the same number of patrons. Ofcourse these assumptions are not necessarily true. To improve the model further we can only include venues that have existed for longer duration as a way to filter out venues that are not successful and hence incorrectly located. To gain further accuracy, we can also adjust the instances of difference types of venues by incorporating the number of patrons each venue type can service. For example, we may attached a lower weighting to smaller venues and higher weightings to venues that support a higher number of customers.

Moving beyond venue data, we can also incorporate demographic data such as the population age, income, etc. Other data points may serve a stronger proxy for the type of demographic that is most well suited to a Pet Store.