**Applied Data Science Capstone Project**

**The Third Place**

Assessing the Impact of Community Public Venues

On Individual Life Satisfaction

1. **Introduction and Background**

**The “Third Place” is a term referring to social environments that are distinct from an individual’s home (“First Place”) and work environment (“Second Place”). Third places are where people congregate and interact other than work or home**.

It has been noted that England has pubs, France has cafés, and Austria has coffee houses. In the past in the U.S., common third places included country stores, post offices, barber shops, hair salons, soda shops, and taverns.

Examples of Third Places in the U.S. today are:

* Cafes, restaurants, pubs, coffee houses - Churches
* Parks and other outdoor recreational areas - Schools, Libraries
* Indoor recreational areas such as arcades - Social and community centers

But not all public venues are automatically considered a true Third Place. The “Third Place” concept was originally proposed by sociologist Ray Oldenburg in 1991. Characteristics of a Third Place are:

* Neutral space – no economic or social status prerequisites for participation
* Accessibility – are inclusive, and readily assessable
* Accommodating – provide for the needs of participants
* Regulars – regular occupants or participants who set the mood and characteristics of the venue
* Low Profile – not extravagant or pretentious, but accepting and inclusive
* Conversation – is the primary activity, although not necessarily the only activity
* Atmosphere – typically exhibit home-like warmth, acceptance and belonging

Sociologies propose that Third Places are critical for active civic engagement and effective democratic governance, and function as anchors of community life by facilitating positive social interactions and a “sense of belonging”.

The Third Place concept has been widely embraced and developed since it’s introduction by business, local government, urban planners, architects and designers, and social policy advocates.

1. **Problem Statement**

**The value of Third Places as an integral component of strong and vibrant communities are often not integrated in more conventional assessments of individual or community life quality**.

Attempts to quantify Life Satisfaction or Quality of Life often consider only Income or Economic related metrics, and fail to consider broader and more subjective values, such as those provided by Third Places.

**This analysis proposes to use Foursquare data on public and social venues (Third Places) for a NYC communities to determine any correlation to more conventional measures of Life Quality or Satisfaction.**

Insights developed could provide additional depth to assessments of “well-being”, and identify opportunities for businesses, communities, and policy makers in their efforts to enhance community quality of life.

1. **Interest**

**The value of increasing our understanding of how Third Places contribute to individual and community well-being extends to a variety of interests, as noted in the following quotes -**

**Business** - *Long-time CEO Howard Schultz used the concept (****Third Place****) as a cornerstone in the creation of Starbucks' image of a coffee chain where people are willing to pay a bit more for coffee because of the cache and potential connection that comes with the purchase.*

**Urban Planning** – “ … *by investing in space that welcomes all people and fosters socioeconomic mixing—just like Oldenburg’s* ***third place*** *philosophy—cities can reverse social and economic fragmentation, nurture more trust and inclusion among residents, and increase environmental sustainability”.*

**Social Policy** - *“Racial equity and inclusion is the key to building successful, thriving cities and the private sector plays a critical role - turning businesses into welcoming ‘****third places****’ for all people, not only paying customers, is a positive step forward.”*

**Governance** – “*Community centers, libraries, neighborhood parks and playgrounds, and arts centers are just a sampling of the potential for public* ***third places****. Free spaces can facilitate activism and democratic participation so residents and community-based organizations can engage in community planning, development, organizing, and policy-change efforts”.*

1. **Data**

**Data by Source** - the primary data sets and sources are as follows:

1. **Trending public venue data by category and New York neighborhoods and communities**

* *Source: Foursquare*

1. **Life Satisfaction Indicator metrics for New York City communities**

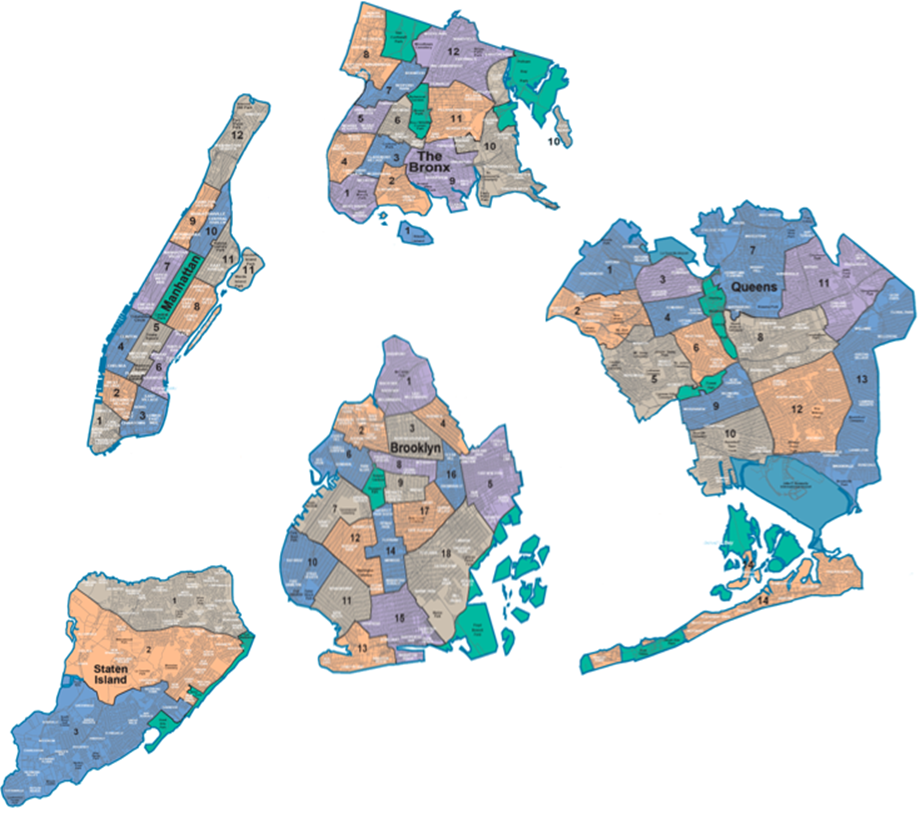
* *Sources: U.S. Census Public Use Micro Data (PUMA) Community Survey, and NYC Community Planning Boards*

**Data Acquisition** – API’s were constructed or utilized where offered by sources to extract data sets

**Data Architecture** - U.S. Census Bureau Public Use Micro-data Areas (PUMA) database contains a wide variety of economic data. For NYC, the PUMA geographical areas correspond directly to NYC Community Board Districts, allowing Four Square Neighborhood Venue data to be accurately combined with the Census Bureau data.

**Data Set** - 59 NYC Community Board Districts comprised of 331 neighborhoods, each with LSI ranks of 1 through 5, combined with the 23,500 Four Square venues, with 259 unique categories.

**New York City Boroughs and Community Districts (numbered areas on maps)**



Source: Wikipedia

1. **Methodology**

**For insights into how Third Place venues correlate to Communities with different Life Satisfaction indicators the following process was followed;**

* + 1. ***Using Cluster Analysis, a Life Satisfaction rank was constructed for each NYC Community based on U.S. Census data***
    2. ***Machine Learning algorithms were used to predict each Community District’s assigned Life Satisfaction average Rank using their respective sets of Public Venues.***

**Life Satisfaction Indicator Ranks applied with Cluster Analysis-**

1 **–** High Satisfaction

2 – Moderate to High Satisfaction

3 – Moderate Satisfaction

4 – Moderate to Low Satisfaction

5 – Low Satisfaction

**Average Life Satisfaction Values by Ranked categories-**



**Classification Analysis using Machine Learning**

Machine Learning models were used to attempt prediction of LSI Rank for Communities based on their Public Venue data. The Machine Learning algorithms utilized were:

- K Nearest Neighbor - Log Regression

- Decision Tree - Naïve Bayes

- Support Vector Machine - Random Forest

All data was normalized. Additionally, all venues with average frequency of occurrence less than 2% were excluded. The Machine Learning algorithm parameters were optimized.

1. **Results**

**Machine Learning Classification Models were able to successfully predict Life Satisfaction Ranks for NYC Communities using their respective sets of Public Venues from Four Square.**

Results by algorithm type are shown below, including Accuracy and Precision scores, and the number of prediction errors (misclassifications of neighborhood LSI ranks).



1. **Conclusion**

**The high accuracy rates of the Machine Learning model predictions indicates that the composition of Public Venues are differentiated between Communities with varying ranges of Life Satisfaction indicators.**

**This initial finding can have valuable insights for the role that public venues and Third Places can play in developing successful business strategies, urban planning, and community development policies and governance.**

**Given the premise that Public Venues can promote active civic engagement and positive social interaction, these differences in Venue composition must be better understood:**

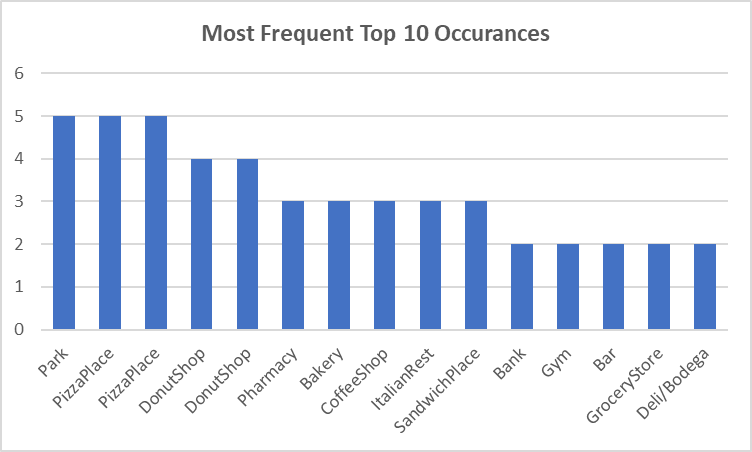
* *Do these differences represent a more advantaged set of venues for some Communities verses a more disadvantaged set for other Communities, or*
* *Do they indicate that different venue sets for Communities with varying Life Satisfaction indicators more effectively facilitate positive social interactions, stronger civic engagement and “sense of belonging*”?

**VIII. Appendix**

**Additional Detail from Analysis**

Below are summaries of Top 10 Trending venues aggregated for Communities within each LSI group.





**Observations on the Top Venues grouped by LSI Rank:**

Across all Communities, only 2 Venues were in the Top 10 for all 5 LSI Rank Groups:

* *Pizza places, and Parks each appeared in all 5 groups*

Only 1 Venue was in Top 10 Trending Venues for 4 of 5 LSI Groups:–

* *Donut Shops*

Of all 50 Top 10 Trending Venues, 32 were Retail Food related and distributed across all LSI Groups:

* *Restaurants, Bakeries, etc.*
* *Bakeries, Coffee Shops, Sandwich Places, and Italian Restaurants each appeared in 3 groups*

Of the 50 Top 10 Trending Venues, 13 were non-Retail food Related