**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 where members have a shared sense of pride, commitment, and well-being is well documented, but often not integrated in more conventional assessments of individual or community life quality.

Attempts to quantify Life Satisfaction or Quality of Life for individuals and communities often consider only Income or Economic related metrics, and fail to consider broader and more subjective metrics. These subjective measures are equally important indicators of overall social and individual “well-being”, attributes that are favorably engendered by Third Places.

This analysis proposes to use Foursquare data on public and social venues (Third Places) for a sample of 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 Third Place concept has received considerable attention from academia, business, and governance since first introduced by Oldenburg.

The value of increasing our understanding of how Third Places contribute to individual and community well-being is best illustrated by the current emphasis on this topic from a variety of interests, noted in the following excerpts:

**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.*

**Workplace Environment** - *Recently, Google opened its own coffee shop branded The Coffee Lab on its campus in Mountainview, Calif. The Coffee Lab is evidence of how organizations are trying to provide employees with access to environments that offer some of the relaxed amenities of home, dubbed the “first place” by sociologist Ray Oldenburg, and the energy of a corner coffee shop—an example of a “****third place****.” Leading organizations are blending these concepts at work, the “second place.” Google’s solution provides workers with the vibe they are craving from the workplace.*

**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,” Sarah Treuhaft, senior director of PolicyLink, an organization working to advance racial and economic equity, tells Curbed. “Turning their 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****. Health care clinics and hospitals could become more community oriented—a trend that’s in its nascent stages. 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**

For this project, the primary data sets and sources are as follows:

1. Life Satisfaction Indicator metrics for New York City communities

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

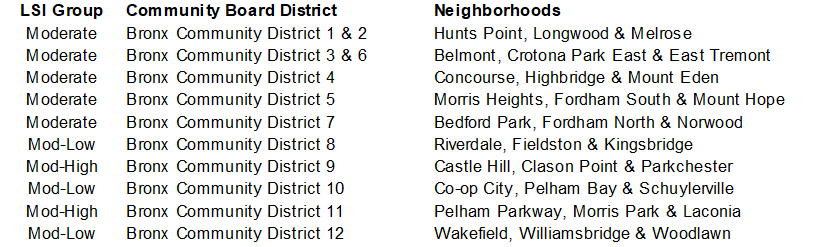
1. OECD Life Quality Index and components (as reference only)
2. Trending public venue data by category and New York neighborhoods and communities

* *Source: Foursquare*

**Data Sourcing** – API’s will be constructed or utilized if 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 indicators, such as Earnings, Unemployment, Education, Commute time to work, Household demographics, etc. PUMA data is tabulated in geographical areas designated by the Census Bureau.

For New York City, these PUMA geographical areas correspond directly to New York City Community Districts. Within each NYC Community District are multiple component neighborhoods. As an example, the following shows this data as compiled for the Bronx borough of NYC.



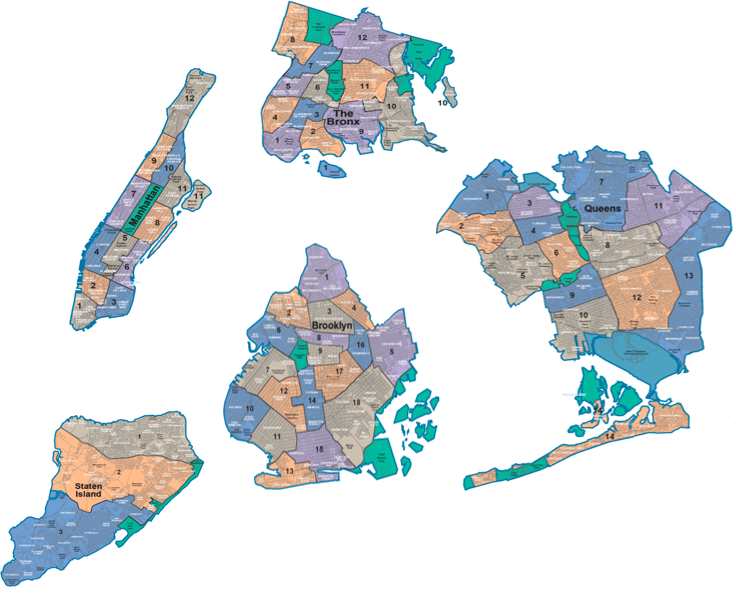
As a result of this data architecture, public venue data by Neighborhood from Foursquare can be easily combined with Census Bureau data for each Community District, forming the foundation of the analysis contained here.

Maps of the 5 New York City Boroughs are shown next. The Numbered areas represent the Community Districts within the Boroughs, that also correspond directly with U.S. Census Public Use Micro Data geographic areas.

* *Within the 5 Boroughs, are 59 Community Districts: 12 in Manhattan, 12 in the Bronx, 18 in Brooklyn, 14 in Queens, and 3 in Staten Island.*
* *The 59 Community Districts are comprised of 331 total Neighborhoods*

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

A complete list of Neighborhoods by Community District and Borough is shown in the Appendix**.**



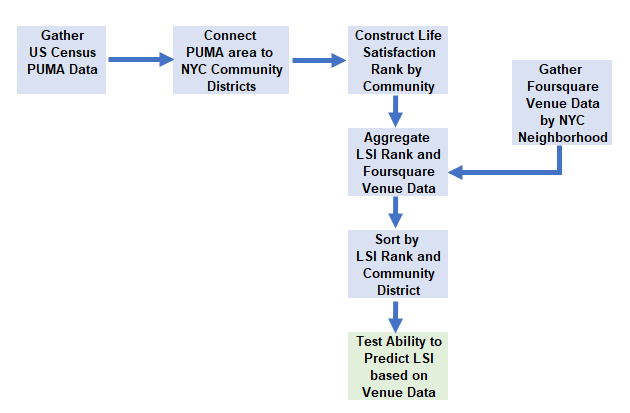
Source: Wikipedia

1. **Methodology**

To gain insights into how potential Third Place venues might correlate to Communities with different Life Satisfaction indicators the following process was followed;

* *Data on trending Public Venues was extracted from Four Square for all NYC neighborhoods.*
* *U.S. Census Bureau and NYC Planning Commission data comprising key Life Satisfaction Indicators (LSI) was extracted via API’s.*
* *Using cluster analysis methods, the LSI indicator data was grouped and used to rank NYC Community Districts from Low to High for apparent Life Satisfaction.*
* *The LSI data and Foursquare Trending Venue data was merged for each NYC Community District.*
* *Machine Learning algorithms were used to predict each Community District’s assigned LSI Rank based on their Venue composition and mix.*

These steps are summarized as follows:



**Constructing the comparative Life Satisfaction Indicator Ranking**

A Life Satisfaction Indicator (LSI) data set was constructed on a selection of key indicators, extracted from the sources detailed in the preceding paragraphs, for all 59 NYC Community Districts.

The selection of attributes was based on the OECD (Organization for Economic Co-operation and Development) Better Life Index components. The OECD Index is calculated at a national level bi-annually, and is widely reported and referenced. The LSI attributes were not weighted. (OECD analysis suggests that attribute weighting by importance can vary dramatically between genders, age, and nationality of respondents).

Cluster analysis was then used to group and rank the Indicators as follows:

**1 – High Satisfaction**

**2 – Moderate to High Satisfaction**

**3 – Moderate Satisfaction**

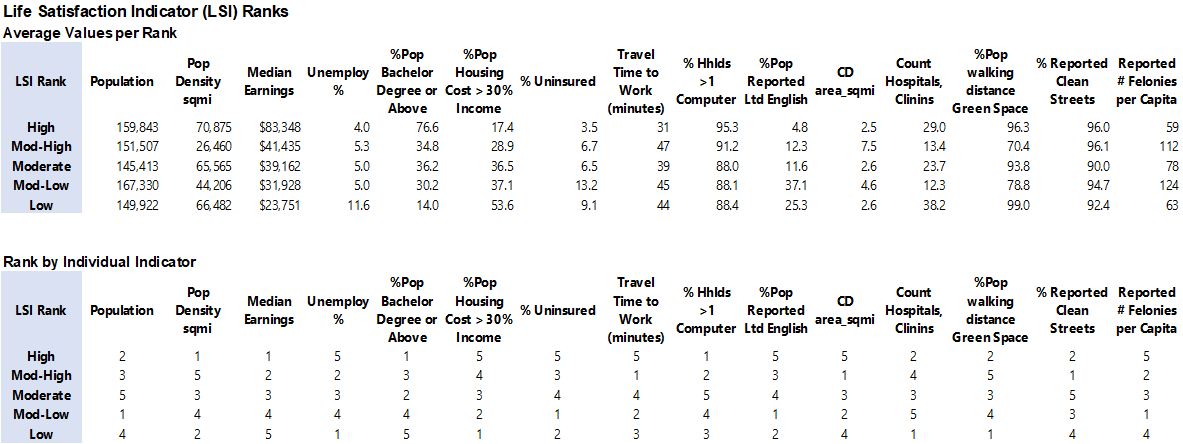
**4 – Moderate to Low Satisfaction**

**5 – Low Satisfaction**

Indicators used in the OECD Index and this Capstone Project analysis are shown below:



The following shows the average values determined for each indicator in the LSI Rank Groups as produced by K-means cluster analysis.

****

**Aggregation**

Shown below is a sample of LSI Ranked Community Districts and Neighborhoods, with the mean frequency of occurrence for several Trending Foursquare Venue categories. This aggregation was the basis for the Machine Learning models.



**Classification Analysis using Machine Learning**

Once the Foursquare and LSI attribute data were segmented by NYC Community District and LSI Rank, Machine Learning classification algorithms were utilized.

Machine Learning methods were used to determine if the mix and composition of the Trending Venue data within Ranked Community Districts was differentiated enough to allow accurate prediction of the LSI Rank for each.

The objective was to gain insights on how Third Place public venues might differ between various Communities with highly varied Life Satisfaction indicators, thus potentially providing valuable reference data for business and policy makers. The Machine Learning algorithms utilized were:

- K Nearest Neighbor - Log Regression

- Decision Tree - Naïve Bayes

- Support Vector Machine - Random Forest

The analysis was performed on the 59 NYC Community Districts comprised of 331 neighborhoods, with LSI ranks of 1 through 5, combined with the respective mix of public venues totaling 259 unique venues.

All data was normalized. Additionally, all venues with average frequency of occurrence less than 2% were excluded. The Machine Learning algorithms were used for multiclass classification.

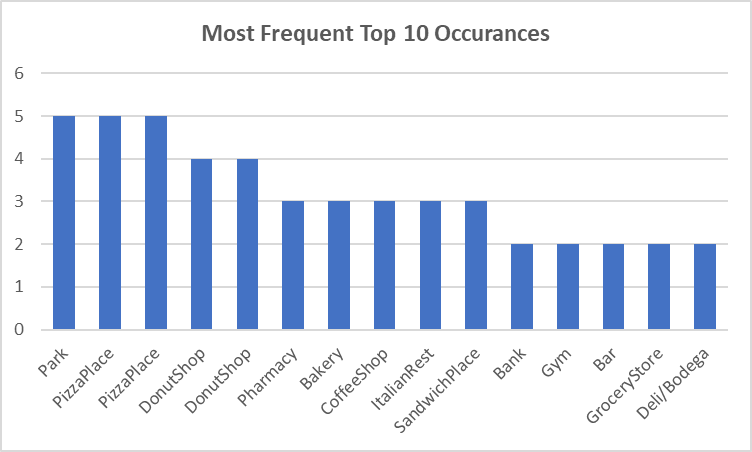
1. **Results**

**Results of Classification Analysis**: the analysis indicated that the mix and composition of Venue data sets are relatively differentiated by Neighborhood and Community, when segmented within LSI Ranking groups. Differentiation is adequate enough to allow prediction of the respective LSI rank with a relatively high accuracy rate. Results by algorithm type are shown below, including Accuracy and Precision scores, and the number of prediction errors (misclassifications) of neighborhood LSI ranks.



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





Observations on the Top Venues grouped by LSI Rank:

Across all Communities, only 2 Venues were in the Top 10 for all 5 Ranked 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*

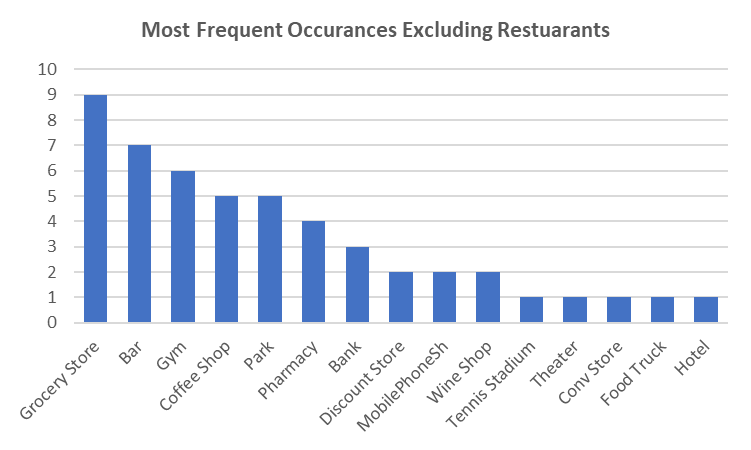
Of all 50 Top 10 Trending Venues, 32 were Retail Food related and distributed across all LSI Groups:

* *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

Top venues Excluding Restaurants were also aggregated for Communities by LSI Rank, as follows:





Observations on the Top Venues Excluding Restaurants, grouped by LSI Rank:

Of the 50 Top 10 Trending Venues excluding Restaurants, 4 venues appeared in all 5 LSI Groups at least once:

* *Grocery/Supermarket – 9 appearances, Bars – 7 appearances, Coffee Shops and Parks – 5 each*

Venues appearing in 4 of 5 LSI groups:

* *Gym/Fitness Centers – 6 appearances, Pharmacy – 4 appearances*

4 Venues appeared only within the High Satisfaction LSI Group:

* *Tennis stadium, Theater, Convenience Store, and Food truck*

1. **Conclusion**

The classification analysis indicates the mix of Public Venues are differentiated between Communities with wide ranges of Life Satisfaction indicators. The degree of differentiation is sufficient to allow accurate prediction of a Community’s LSI range based on their respective public venue mix.

Promoting the concept of the Third Place is key to developing strong and vibrant communities where members have a shared sense of pride, commitment, and well-being.

Identifying the relationship between public venues as potential Third Places and other Life Satisfaction Indicators is an important foundational step.

1. **Further Analysis**

Next steps to utilize this data would be to develop additional insights on key drivers of differentiation of public venues in different Community Districts. Important questions would be:

What are the primary differences in Venue mix: different Venues, frequency of occurrence, or mix?

Does the Venue mix for different LSI indicate desired or undesirable states?

What are the implications or opportunities for Communities and Policy makers?