**Applied Data Science Capstone Project**

**The Third Place**

Assessing the Impact of Community Public Venues

On Individual Life Satisfaction

1. **Introduction**

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”). Examples of Third Places shown below correspond very closely with venue categories in the Foursquare database.

Examples of Third Places:

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

The “Third Place” concept was originally proposed by sociologist Ray Oldenburg in 1991. Oldenburg proposed that Third Places were 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**

Life Satisfaction or Quality of Life measurements are typically quantitative by design, but can fail to consider broader measures which can be equally important indicators of overall social and individual “well-being” or “happiness”.

This analysis proposes to determine if data on public and social venues (Third Places) for a sample of communities can be correlated to more conventional measures of Life Quality or Satisfaction, thereby providing additional depth and insight to assessments of “well-being”.

1. **Background**

The Third Place concept has received considerable attention from academia, business, and governance since first introduced by Oldenburg. The attributes of Third Places are generally summarized as follows:

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

1. **Interest**

The value of increased understanding of how Third Places contribute to individual and community well-being is best illustrated by 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** - 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.

**Architecture** - I believe architecture can change people’s lives and change them for the better …. and also instill a sense of pride in one’s community. Civic buildings and public buildings go a long way to represent who we are. They should be commensurate with our pride of place.

**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.

Sources and accreditation for the excerpts above will be provided in the final report draft.

1. **Data**

For the project, the analysis subject will the Boroughs and Neighborhoods of New York City. The analysis process steps will be 1) extract available Quality of Life indicator data for New York City communities, 2) construct a Quality of Life or Life Satisfaction Index for each community, 3) extract trending venue data by category for each community from Four Square, and finally, 4) perform analytic processes to identify insights on how venue data (representing Third Places) on subject communities correlates to the computed Life Quality Index.

**Data Sourcing**

The primary data source for the analysis will be the U.S. Census Bureau Public Use Micro-data (PUMA). This database is an exceptionally rich source of 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, the PUMA areas correspond directly to New York City Community Districts. Each District is comprised of a group of community neighborhoods. As an example, PUMA district #4112 is NYC - Queens Community District 12, which is comprised of the neighborhoods Jamaica, Hollis, and St. Albans.

As a result of this data architecture, venue data from Foursquare can be easily correlated to PUMA data on the basis of component Neighborhoods.

The New York City Planning Commission also compiles a robust set of data by Community District Board area, from sources such as the NY Police Dept (crime rates).

Maps of the 5 New York Boroughs are shown below. The Numbered areas represent the Community Districts within the Boroughs. There are 59 Community Districts: 12 in Manhattan, 12 in the Bronx, 18 in Brooklyn, 14 in Queens, and 3 in Staten Island.

A picture containing text, map, items, air

Description automatically generated

*Accreditation for the images above will be provided in the final report draft*

**Constructing a Comparative Satisfaction Index**

A Life Quality Index will be constructed based on a selection of key indicators, extracted from the sources detail in the preceding paragraph.

Selection of Index attributes will be based on the OECD (Organization for Economic Co-operation and Development) Better Life Index which is calculated at a national level. The Index is often reported as a ranking of “National Happiness”, with each participating country ranked on its calculated Life Satisfaction and Well-Being compared to other nations. (In 2017, Norway ranked #1, the United States ranked #10).

**Key components of the OECD Index are as follows:**

|  |  |
| --- | --- |
|  |  |
| **Income and Wealth** | Household Net Income |
|  |  |
| **Jobs and Earnings** | Employment Rate |
|  |  |
| **Housing** | Rooms per person |
|  | Rent as % Income |
|  |  |
| **Health Status** | Life expectancy |
|  | Self-Reported Health |
|  |  |
| **Work-Life Balance** | Hours > 50 per week worked |
|  | Hours devoted to leisure and personal care (sleeping, eating) |
|  |  |
| **Education and Skills** | Education level attained by adult population |
|  | Student age 15+ rank on reading tests (literacy) |
|  |  |
| **Social Connections** | Frequency of engagement with family or friends |
|  | Share of people reporting a support network available in times of need |
|  |  |
| **Civic Engagement and Governance** | Number of votes cast by voting-age population |
|  |  |
| **Environmental Quality** | Broad: quality of different environmental air, water) and access to amenities |
|  | Calculated composite index of air pollution |
|  |  |
| **Personal Security** | Reported number of victims of an assault crime |
|  |  |
| **Subjective Well Being** | Reported subjective we-being |

**Statistics collected to construct a Life Quality Index for NYC Community Districts are as follows:**

|  |  |
| --- | --- |
| **Income** | Median Earnings past 12 months |
|  | Unemployment Rate |
| **Housing** | Housing Cost > 30% of Income |
| **Health** | Healthcare - % Population Uninsured |
| **Work Life Balance** | Travel To Work (minutes-average 1 way) |
| **Education and Skills** | Education - % Population Bachelor’s Degree or Above |
|  | % Households with Limited English speaking capability |
| **Social Connections** | % Households with >1 Computer |
| **Civic Engagement** | No data identified |
| **Support Network** | Count of Hospitals, clinics, treatment centers |
| **Environment** | Population density |
|  | % of Population with Green Space within walking distance |
|  | % Clean Streets as ranked by NYC Commission |
| **Personal Security** | Crime Count – Felony violent assaults reported by NYPD |
| **Subjective Well-Being** | No self-reported metric |

1. **Methodology**

A Life Satisfaction Index will be calculated for each of the 59 NYC Community Districts

The Index rankings will be segmented using K-mean cluster analysis and/or binning as follows:

1 – High Satisfaction

2 – Moderate-High Satisfaction

3 – Moderate Satisfaction

4 – Moderate to Low Satisfaction

5 – High Satisfaction

Initially, the attributes described above will not be weighted. (Analysis from the OECD suggests that attribute weighting by importance can vary dramatically between genders, age, and nationality of respondents for construction of their index).

Foursquare trending venue data will be extracted for neighborhoods in each borough. The neighborhood data will be aggregated according to the neighborhood composition of each Community District – PUMA area. The top 10 trending venue categories will be determined.

Machine Learning algorithms will be utilized to compare the Community District calculated Satisfaction Index to the respective Community venue data. The number of each category as well as the mix of categories will be used in the comparison. Classification Algorithms to be used are:

* K Nearest Neighbor
* Decision Tree
* Support Vector Machine
* Log Regression

The comparative analysis will be performed on a select sample borough and component Districts initially. If results are valid or insightful, the analysis will be extended to the full NYC Community data set.

Also noted: the Machine Learning algorithms in this course were primarily used as binary classifiers, but Python literature indicates that these algorithms can be applied to multinomial classification, as described above.