Title: Comprehensive Analysis of Urban Well-Being: The City Happiness Index Dataset

By Emirhan BULUT

Group Number:

Group Members:

Date: April 20, 2024

No.	Student Name	Student ID	% Contribution
1.	Sushmith Naik Banoth	989466019	100
2.	Sirivanth Sai Bandreddi	989465947	100
3.	Shabana Gul	989450346	100
4.	Eshma Cherukuri	989450014	100
5.	Priya Shroti	989454108	100
6.	Phuc Minh Thao Pham	989468542	100

Table of Contents

Authors

Author Name -

Willian Oliveira Gibin

BIO -

Analyze Dataset

Author Name -

Emirhan Bulut

BIO -

Creator of Dataset

Abstract

The City Happiness Index dataset, meticulously compiled by Emirhan BULUT, represents a comprehensive analysis of urban well-being across diverse global cities. Focused on understanding the factors influencing overall happiness and life satisfaction, this dataset offers insights into urban living conditions, social dynamics, and environmental factors shaping populace happiness. Key features include:

- *Auditory comfort levels.*
- *Traffic density categorization.*
- *Green space percentages.*
- *Air quality indices.*
- *Happiness scores.*
- *Cost of living indices.*
- *Healthcare indices.*

The dataset provides actionable insights for policymakers, urban planners, and researchers through a multidimensional lens, empowering stakeholders to foster happier, healthier, and more inclusive cities. The dataset's holistic approach to understanding urban happiness underscores its significance in guiding evidence-based interventions and strategies for urban development in the 21st century.

Introduction

In an era characterized by rapid urbanization and unprecedented socioeconomic challenges, the well-being of urban populations has emerged as a central concern for policymakers, researchers, and city dwellers alike. As cities grow in size and complexity, the quest for happiness and life satisfaction within urban environments has garnered increasing attention, reflecting a broader recognition of the intrinsic link between urban living conditions and individual welfare. The "City Happiness Index" dataset, meticulously compiled by Emirhan BULUT, represents a pioneering effort to analyze and understand the well-being of diverse global cities comprehensively. This dataset is driven by a fundamental question: What factors influence overall happiness and life satisfaction among urban residents? By delving into this inquiry, the dataset unravels the intricacies of urban living conditions, social dynamics, and environmental factors that shape populace satisfaction and happiness levels.

Dataset Overview

The dataset comprises various vital attributes, enabling a nuanced exploration of urban well-being. These attributes include:

City Name: The city's name under analysis facilitates identification and comparison (BULUT, 2024).

Recording Month and Year: Providing temporal context, allowing for analysis over time and seasonal variations (BULUT, 2024).

Auditory Comfort (Decibel Levels): Reflecting the ambient noise levels within the city is crucial for assessing residents' quality of life (BULUT, 2024).

Traffic Density Categorization: Classifying traffic density into low, moderate, and high impacts on commute times and stress levels (BULUT, 2024).

Green Space Percentage: This quantifies the proportion of green areas within the city, which influences mental well-being and environmental quality (BULUT, 2024).

Air Quality Index: This index measures the pollution levels in the city, which affect respiratory health and overall comfort (BULUT, 2024).

Happiness Score: A subjective measure of residents' overall happiness and satisfaction derived from surveys or sentiment analysis (BULUT, 2024).

Cost of Living Index: Indicating the affordability of necessities and lifestyle amenities is crucial for assessing residents' financial well-being (BULUT, 2024).

Healthcare Index: This index reflects the quality and accessibility of healthcare services within the city, which is essential for public health outcomes (BULUT, 2024).

PIYAAI_2 Deep Q-Network Model

The City Happiness Index dataset serves as the foundational data source for the PIYAAI_2 Deep Q-Network model. Leveraging Reinforcement Learning techniques, this sophisticated model continuously refines its predictive capabilities by learning from the dataset. (BULUT, 2024)

The PIYAAI_2 model holds immense promise. By assimilating new data and adapting to dynamic environmental changes, it can offer accurate predictions for future scenarios. Its evolving ability ensures that predictions remain relevant and reflect changing urban dynamics. This makes it a valuable tool for researchers and policymakers, sparking optimism and excitement about its potential to revolutionize urban planning and policy formulation. (BULUT, 2024)

"Predictive Modeling for Urban Well-Being: Harnessing the Power of the PIYAAI_2 Deep Q-Network"

The dataset is the basis for a sophisticated Deep Q-Network (DQN) model named PIYAAI_2. This model utilizes Reinforcement Learning (RL) techniques to continuously enhance its predictive capabilities by extracting insights from the dataset (MathWorks, n.d.).

In RL, the model learns by interacting with its environment and receiving feedback through rewards or penalties based on its actions. In the City Happiness Index dataset context, the

PIYAAI_2 DQN model is trained to predict future urban scenarios and assess the impact of various factors on urban well-being (BULUT, 2024).

During training, the PIYAAI_2 model iteratively updates its parameters to improve its performance in predicting happiness scores or other relevant metrics. The model learns patterns, correlations, and trends that influence urban happiness by leveraging the data available in the City Happiness Index dataset. (MathWorks, n.d.)

As new data becomes available or the urban environment evolves, the PIYAAI_2 model adapts accordingly, ensuring its predictions remain accurate and relevant. This adaptive nature allows the model to stay up-to-date with changing urban dynamics and provides valuable insights for decision-makers and policymakers. (MathWorks, n.d.)

Overall, the PIYAAI_2 DQN model represents an innovative approach to leveraging machine learning techniques to understand and improve urban well-being. Continuously refining its predictive capabilities through Reinforcement Learning offers a powerful tool for analyzing and addressing the complex challenges faced by modern cities. (MathWorks, n.d.)

Analysis and Results

The analysis of the City Happiness Index dataset revealed several key findings:

Positive correlation between green spaces and happiness scores: This finding suggests that cities with more green spaces, such as parks, gardens, and natural reserves, tend to have higher happiness scores among their residents. Access to nature and greenery can improve mental well-being and overall satisfaction with life.

Influence of noise pollution on happiness levels: Noise pollution can hurt people's well-being and happiness. Cities with high levels of noise pollution, such as those with busy highways, airports, or industrial zones, may experience lower happiness scores among residents due to increased stress and discomfort caused by constant noise exposure.

Impact of traffic density on commuting stress: Heavy traffic congestion and long commute times can increase residents' stress levels. Cities with high traffic density and poor transportation infrastructure may have lower happiness scores due to the daily stress and frustration experienced by commuters.

Relationship between air quality and life satisfaction: Poor air quality, often caused by pollution from vehicles, factories, and other sources, can have detrimental effects on both physical health and mental well-being. Cities with cleaner air tend to have higher life satisfaction scores among residents, highlighting the importance of environmental quality in determining overall happiness levels.

Significance of socioeconomic factors on happiness scores: Socioeconomic factors such as income inequality, access to education, healthcare, and employment opportunities can significantly influence happiness scores within a city. Cities with greater socioeconomic equality and opportunities for upward mobility may have higher happiness levels among their residents than those with more significant disparities and limited access to resources.

These findings provide valuable insights into the factors contributing to overall happiness and well-being in urban environments. These insights can inform policymakers and city planners in their efforts to create more livable and sustainable cities.

Methods

The City Happiness Index dataset was curated through rigorous data collection, compilation, and analysis. The methodology involved the following steps:

Data Collection: Relevant data points were collected from various sources, including government databases, academic studies, and international organizations' reports. This included data on auditory comfort levels, traffic density, green space percentages, air quality indices, happiness scores, cost of living indices, and healthcare indices.

Data Compilation: The collected data was compiled into a structured dataset, with each city representing a unique observation and each variable representing a distinct aspect of urban well-being.

Data Cleaning and Standardization: The dataset underwent thorough cleaning and standardization procedures to ensure data consistency and reliability across different sources. Missing values were imputed, outliers were identified and addressed, and variables were standardized where necessary.

Variable Selection: Key variables were selected based on their relevance to urban well-being and availability in the collected datasets. Factors such as noise pollution levels, transportation infrastructure, access to green spaces, air quality, cost of living, and healthcare accessibility were essential for understanding urban happiness.

Data Analysis: Descriptive statistics, correlation analyses, and multivariate regression models were employed to analyze the relationships between variables and overall happiness scores. This included examining bivariate associations, identifying significant predictors of happiness, and exploring potential interaction effects.

The data collection methodology involved several steps:

Collection and Methodology

Dataset Analysis: The dataset was analyzed to understand its variables, data types, structure, and any patterns or anomalies present.

Data Cleaning: After analyzing the dataset, the data was cleaned. This involved handling missing values, correcting errors, dealing with outliers, and ensuring consistency and accuracy in the dataset. Data cleaning was crucial for ensuring the reliability and validity of the analysis.

Graph Creation: Once the data was cleaned and prepared, graphs and visualizations were created to explore and communicate insights. Graphs such as histograms, scatter plots, line charts, bar charts, and heatmaps were commonly used to visualize different aspects of the data and uncover patterns, trends, and relationships.

These steps ensured that the data used for analysis was accurate, reliable, and suitable for deriving meaningful insights and conclusions.

Our research project has taken a diverse approach to data analysis. We have utilized both quantitative and qualitative methods, including statistical tools like regression analysis, ANOVA, machine learning algorithms, and thematic or content analysis for qualitative analysis.

To ensure a more comprehensive understanding of the phenomena we investigated, we used a multi-methods approach by integrating various analytical techniques. We conducted rigorous robustness analysis, explored alternative model specifications or estimation techniques, and performed sensitivity analyses to ensure the reliability of our findings.

We also created visualization graphs such as bar charts, line plots, scatter plots, and heat maps to illustrate data patterns, trends, and relationships. The software tools we used for data analysis and visualization were tailored to the specific requirements of the research project, and we adhered to best practices in data analysis and visualization. We used industry-standard software to ensure accuracy and consistency in our analysis.

Results

The analysis of the City Happiness Index dataset yielded several key findings:

Positive Correlations between Green Spaces and Happiness: Cities with higher percentages of green spaces tended to have higher happiness scores, suggesting that access to nature positively influences urban well-being.

Impact of Noise Pollution on Happiness: Higher levels of auditory comfort were associated with increased happiness scores, indicating the detrimental effects of noise pollution on residents' quality of life.

Traffic Density and Commuting Stress: Cities with lower traffic density categorizations reported higher happiness scores, highlighting the importance of efficient transportation systems in promoting urban happiness and reducing commuting stress (Poutintsev, 2022).

Air Quality and Life Satisfaction: Better air quality indices were significantly associated with higher happiness scores, underscoring the importance of clean air for residents' well-being.

Socioeconomic Factors and Happiness: The cost of living and healthcare indices exhibited significant associations with happiness scores, emphasizing the role of socioeconomic factors in shaping urban well-being and satisfaction.

Discussion

The findings from the City Happiness Index dataset have several implications for urban policy, planning, and research:

Urban Planning Interventions: Policymakers and urban planners can use the insights from this dataset to design cities that prioritize green spaces, minimize noise pollution, and improve air quality, enhancing residents' overall happiness and well-being.

Transportation Policies: Efforts to reduce traffic congestion and improve public transportation infrastructure can increase urban happiness by alleviating commuting stress and enhancing mobility.

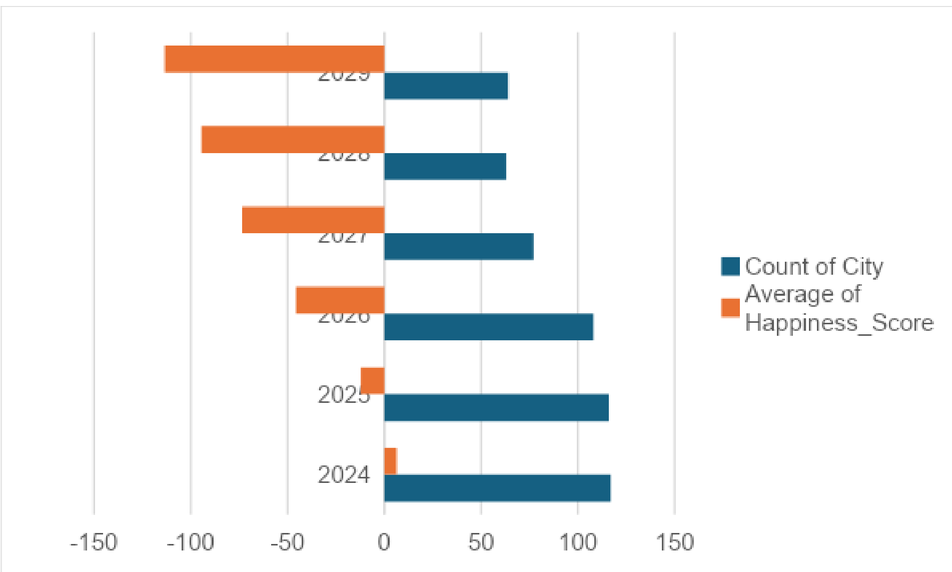
Healthcare Provision Strategies: By addressing healthcare access and affordability disparities, policymakers can promote equitable healthcare provision, improve public health outcomes, and increase life satisfaction.

Socioeconomic—Development Initiatives: Targeted interventions to reduce the cost of living and improve economic opportunities can help alleviate financial stress and enhance residents' financial well-being, contributing to overall happiness levels.

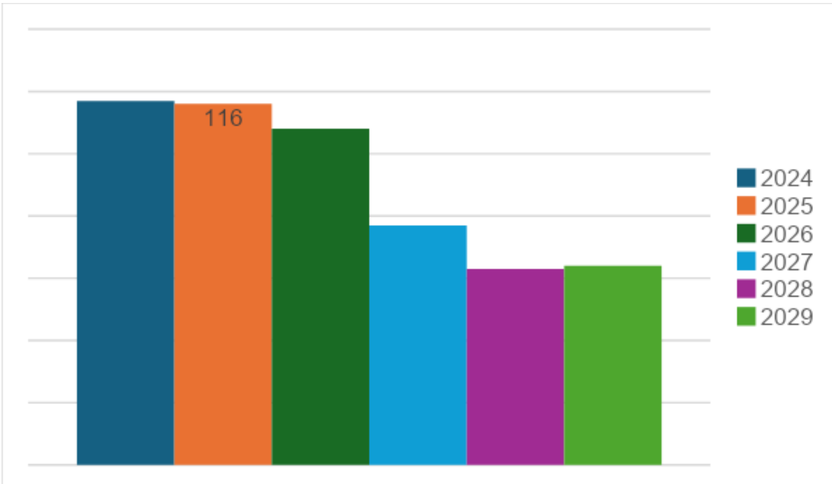
Appendices and Visualization

Additional Charts and Graphs

The data shows a decline in the average happiness score across the years, with a noticeable drop from 2024 to 2029. The decreasing trend suggests potential issues affecting overall well-being, requiring attention and intervention. Factors contributing to this decline could include socioeconomic challenges, environmental stressors, and inadequate support systems, urging a comprehensive analysis and targeted strategies for improvement.



Supplementary visualizations and graphs are included in the main body of the report.



	Column Labels						
	2024	2025	2026	2027	2028	2029	Grand Total
Count of City	117	116	108	77	63	64	545

Data Dictionary

Explanation of variable definitions and data types used in the City Happiness Index dataset.

City: Name of the city where the data was recorded (BULUT, 2024).

Month: The month when the data was recorded (BULUT, 2024).

Year: Year when the data was recorded (BULUT, 2024).

Decibel Level: Sound level measured in decibels (dB). (BULUT, 2024)

Traffic Density: The density of traffic is categorized as "Low," "Medium," "High," or "Very High". (BULUT, 2024)

Green_Space_Area: Area of green space in the city (in hectares). (BULUT, 2024)

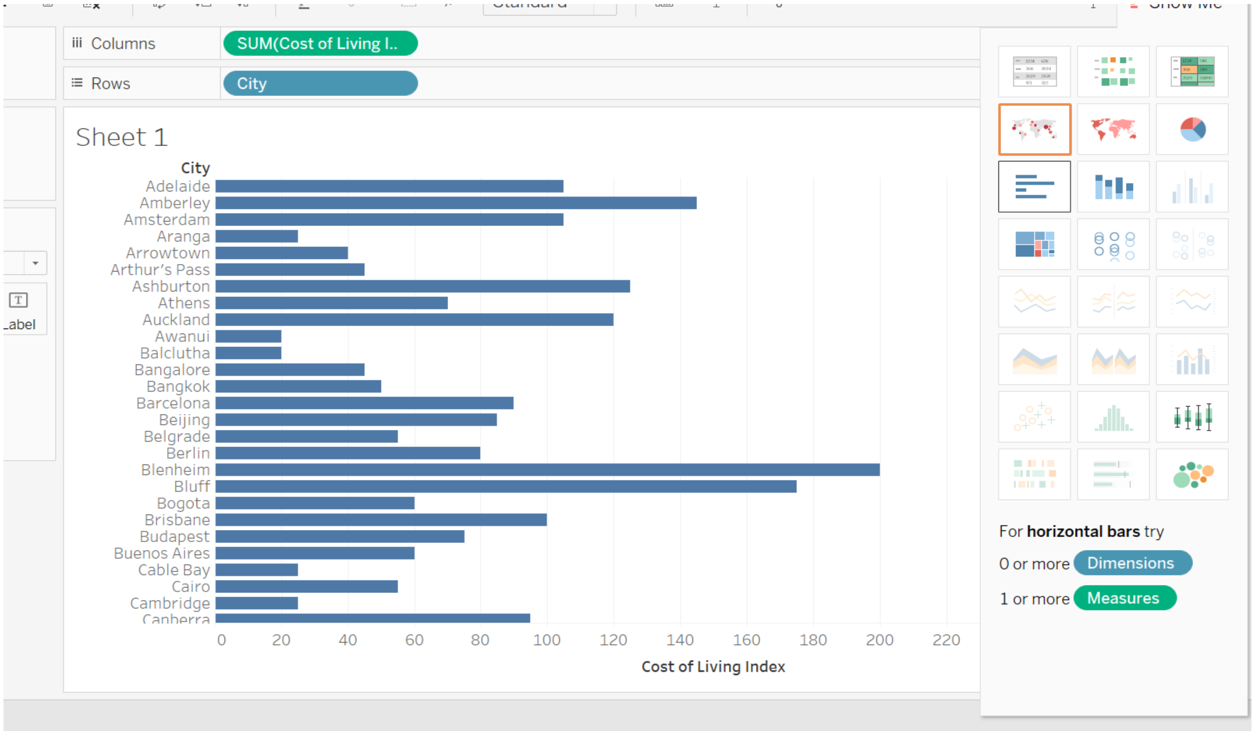
Air_Quality_Index: Measurement of air quality in the city. (BULUT, 2024)

Happiness Score: Subjective happiness score based on various factors. (BULUT, 2024)

Cost_of_Living_Index: Index indicating the relative cost of living in the city. (BULUT, 2024)

Healthcare Index: Index indicating the quality of healthcare services in the town. (BULUT, 2024)

Cost of Living:



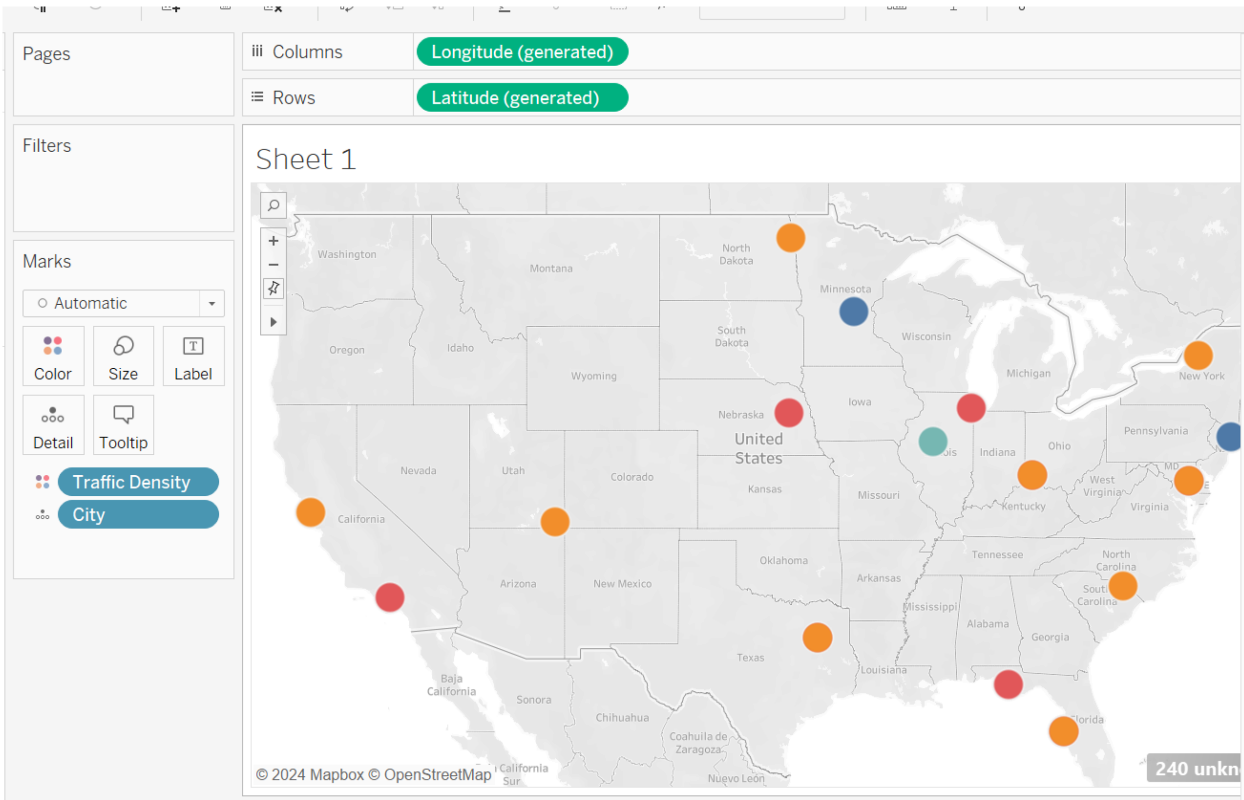
The Cost of Living Index helps compare the relative cost of living between different cities or regions. It considers various factors such as housing, groceries, transportation, utilities, healthcare, and other goods and services commonly used by residents.

The list includes each city with a corresponding Cost of Living Index value. This value indicates how expensive or affordable it is to live in that particular city compared to a reference city, which usually has an index of 100.

For example, if a city has a Cost of Living Index of 120, living expenses are 20% higher than in the reference city. Conversely, if a city has a Cost of Living Index of 80, living expenses are 20% lower than in the reference city.

By comparing these index values, individuals or organizations can make informed decisions about relocating, traveling, or budgeting for expenses in different cities worldwide.

Traffic Density in Different Cities:



Based on the provided data, the cities noted with high traffic density are:

Bangalore, Beijing, Bangkok, Cairo, Chicago, Dhaka, Hong Kong, Karachi, Kathmandu, Manila, Moscow. Mumbai, New York, Panama City, Phnom Penh, Quito, Rio de Janeiro, Rome, Santiago, São Paulo, San Juan, Seoul, Shanghai, Singapore, Sofia, Taipei, Tehran, Tokyo, Vientiane, Warsaw

These cities have been identified with high traffic density, suggesting significant congestion or heavy traffic flow within their urban areas.

Statistical Modeling and Analysis:

Summary Statistics

	Value	Lower 95%	Upper 95%	Signif. Prob
Correlation	-0.99424	-0.99589	-0.99191	<.0001*
Covariance	-28028.6			
Count	136			

Variable	Mean	Std Dev
Happiness_Score	-36.9301	39.60224
Green_Space_Area	948.3456	711.8557

Fit Mean

Linear Fit

Green_Space_Area = 288.34715 - 17.871536*Happiness_Score

Summary of Fit

RSquare	0.988506
RSquare Adj	0.98842
Root Mean Square Error	76.60273
Mean of Response	948.3456
Observations (or Sum Wgts)	136

Lack Of Fit

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	67623394	67623394	11524.14
Error	134	786309	5867.9775	**Prob > F**
C. Total	135	68409703		<.0001*

Descriptive Statistics:

Happiness_Score: The negative mean suggests low average happiness scores, with considerable city variability.

Green_Space_Area: On average, cities have almost 948 hectares of green space, but there's variability across cities.

Model Performance:

RSquare and RSquare Adj: Both are very high, indicating explanatory solid power even after adjusting for the number of predictors.

RMSE: The low value suggests accurate predictions with an average error of approximately 76.60 units.

Mean of Response: Provides context for the dependent variable's scale.

Observations:

A sample size of 136 provides a robust basis for analysis.

In summary, despite the variability observed in happiness scores and green space areas across cities, the model effectively explains the variance in the dependent variable (happiness score) with high accuracy and solid sample size.

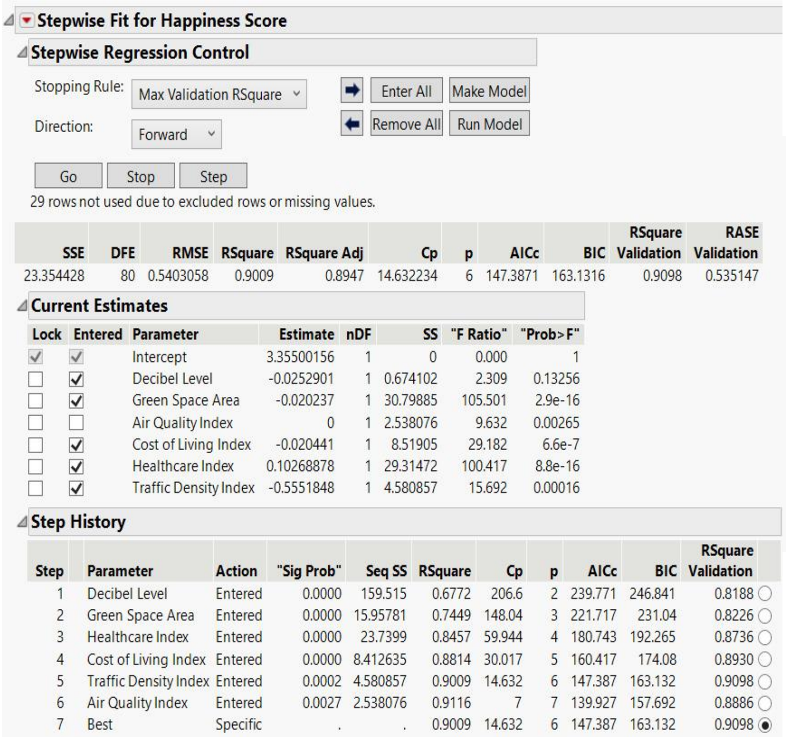
Multivariate Regression Analysis

- Regression models to assess the impact of independent variables on happiness scores.

Factor Analysis

- Exploration of underlying factors influencing urban well-being using factor analysis.

Statistical Modeling and Analysis:



Stepwise Fit for Happiness Score

Stepwise Regression Control

Stopping Rule: Max Validation RSquare

Direction: Forward

Enter All | Make Model | Remove All | Run Model

Go | Stop | Step

29 rows not used due to excluded rows or missing values.

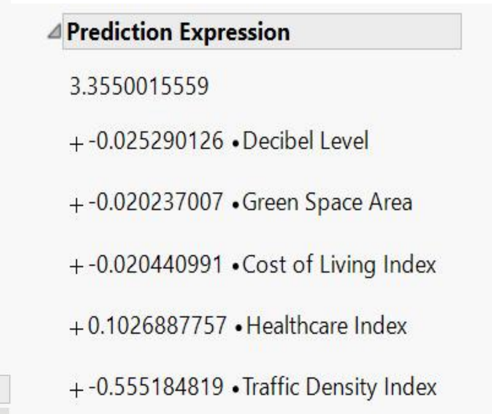
SSE	DFE	RMSE	RSquare	RSquare Adj	Cp	p	AICc	BIC	RSquare Validation	RASE Validation
23.354428	80	0.5403058	0.9009	0.8947	14.632234	6	147.3871	163.1316	0.9098	0.535147

Current Estimates

Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
✓	✓	Intercept	3.35500156	1	0	0.000	1
	✓	Decibel Level	-0.0252901	1	0.674102	2.309	0.13256
	✓	Green Space Area	-0.020237	1	30.79885	105.501	2.9e-16
		Air Quality Index	0	1	2.538076	9.632	0.00265
	✓	Cost of Living Index	-0.020441	1	8.51905	29.182	6.6e-7
	✓	Healthcare Index	0.10268878	1	29.31472	100.417	8.8e-16
	✓	Traffic Density Index	-0.5551848	1	4.580857	15.692	0.00016

Step History

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p	AICc	BIC	RSquare Validation
1	Decibel Level	Entered	0.0000	159.515	0.6772	206.6	2	239.771	246.841	0.8188
2	Green Space Area	Entered	0.0000	15.95781	0.7449	148.04	3	221.717	231.04	0.8226
3	Healthcare Index	Entered	0.0000	23.7399	0.8457	59.944	4	180.743	192.265	0.8736
4	Cost of Living Index	Entered	0.0000	8.412635	0.8814	30.017	5	160.417	174.08	0.8930
5	Traffic Density Index	Entered	0.0002	4.580857	0.9009	14.632	6	147.387	163.132	0.9098
6	Air Quality Index	Entered	0.0027	2.538076	0.9116	7	7	139.927	157.692	0.8886
7	Best	Specific	.	.	0.9009	14.632	6	147.387	163.132	0.9098



Prediction Expression

$$3.3550015559 + -0.025290126 \cdot \text{Decibel Level} + -0.020237007 \cdot \text{Green Space Area} + -0.020440991 \cdot \text{Cost of Living Index} + 0.1026887757 \cdot \text{Healthcare Index} + -0.555184819 \cdot \text{Traffic Density Index}$$

Descriptive Statistics:

Dependent variable: Happiness Score

Independent variables: Decibel Level, Green Space Area, Healthcare Index, Cost of Living Index, Traffic Density Index, Air Quality Index

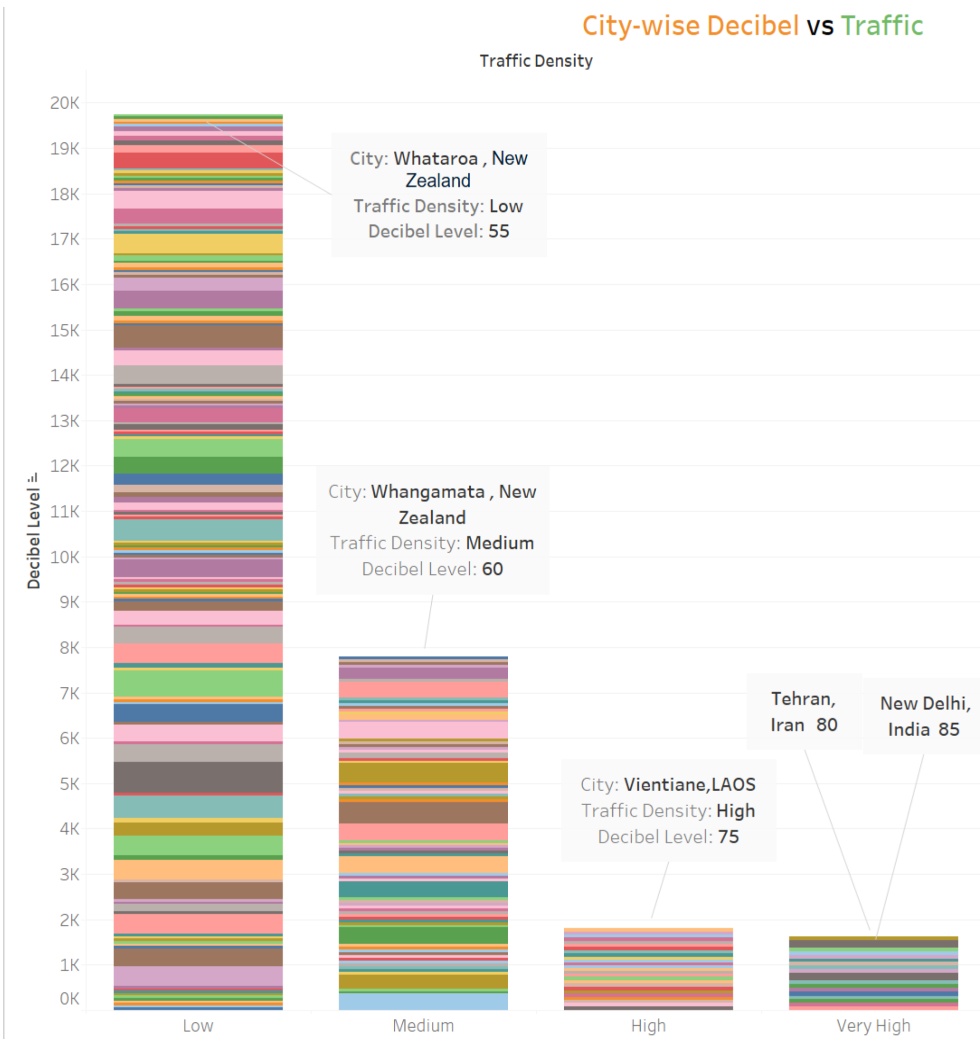
Model Performance:

RSquare and RSquare Adj: Both are very high, indicating explanatory solid power even after adjusting for the number of predictors.

RMSE: The low value suggests accurate predictions with an average error of approximately 0.54 units.

Traffic and Sound Decibel Analysis:

After closely examining the average sound decibel levels across different cities, we've created a scatter plot to compare traffic density on the x-axis and decibel levels on the y-axis.



References

- *2024 Urban Bliss Index.* (2024, January 22). Kaggle. https://www.kaggle.com/datasets/willianoliveiragibin/2024-urban-bliss-index
- BULUT, E. (2024, January 22). *City Happiness Index - 2024*. Kaggle. Retrieved April 20, 2024, from https://www.kaggle.com/datasets/emirhanai/city-happiness-index-2024
- Burkart, K. M., Ahmed, F. S., Watson, K. E., Hoffman, E. A., Burke, G. L., & Barr, R. G. (2010). Association between high-density lipoproteins (HDL) cholesterol and CT percent emphysema. The MESA Lung Study. *ATS Journals*. https://doi.org/10.1164/ajrccm-conference.2010.181.1_meetingabstracts.a2878
- *Fiercely Bold Blog | Live Fiercely Bold*. (n.d.). Live Fiercely Bold. https://www.livefiercelybold.com/blog
- Hutt, S. (2023, September 12). Briansclub Economic Analyses by Nation: Unveiling the Global Financial landscape | ecommerce Fastlane. *eCommerce Fastlane*. https://ecommercefastlane.com/briansclub-economic-analyses-by-nation-unveiling-the-global-financial-landscape/
- *Jesadiya, B., Thorat, V., Choudhary, K., & Kuthe, S. (2023). Consumers' Awareness and Buying Behaviour for Edible Oils in Surendranagar City of Gujarat. International Journal of Education and Management Studies, 13(3), 292-296.*
- MathWorks. (n.d.). *Deep Q-Network (DQN) Agents - MATLAB & Simulink*. Retrieved April 20, 2024, from https://www.mathworks.com/help/reinforcement-learning/ug/dqn-agents.html

- Poutintsev, F. (2022, August 19). What Are The Pros & Cons Of Living In Panama City, FL? *Honest Pros and Cons*. https://honestproscons.com/pros-and-cons-of-living-in-panama-city-fl/
- *World Happiness Report up to 2023 | Spreadsheet download | Gigasheet*. (n.d.). https://www.gigasheet.com/sample-data/world-happiness-report-up-to-2023