TEAM 2 – PROJECT REPORT

LivabilityX: An Analysis of Trends in Global Urban Livability

Yulin Wei; Danyang Zhang; Xingzhou Ma
weiyulin020626@163.com; {dzhang497;lyndonma}@gatech.edu

1. Introduction

Urban livability significantly impacts residents' quality of life and informs urban development and policymaking.[1] With challenges from urbanization, global warming, and socioeconomic disparities, understanding and improving livability is crucial.[2] Livability involves factors like economic stability, environmental quality, healthcare, and housing. Our project aims to develop a comprehensive Urban Livability Index to assess countries worldwide, integrating climate and socio-economic data to guide policies and urban planning. By analyzing both historical and current data, our index tracks changes in livability and assesses the impact of emerging trends.

2. Problem Definition

This project aims to create an interactive tool for evaluating global city livability using climate and socio-economic data. Current assessments are often limited in scope and lack adaptability to changing conditions. Our proposed tool will identify key factors contributing to livability and their changes over time, using machine learning to predict scores and analyze factors like climate, healthcare, cost of living, and safety. The tool will also simulate potential changes in livability, helping stakeholders improve living conditions and adapt to socio-economic and environmental challenges.

3. Literature Survey

Ahmed et al. (2019) critically review the concept of "livability," noting its broad application and inherent ambiguity in urban planning. They concluded by discussing strategies to enhance livability in urban environments, including cities, towns, and neighborhoods[3].

Current livability assessments often rely on separate indices or non-integrated measures, making it difficult to capture a city's overall livability comprehensively. For example, the study by Wang et al. (2015) compared Beijing with other global cities using the Livable Level Integrated Index (LLII) and revealed gaps, particularly in environmental quality, which underscores the need for more holistic assessments that consider multiple factors over time[4]. Kutty et al. (2022) propose a novel machine learning-based framework combining metric-distance analysis and clustering techniques to evaluate resilience and livability in 35 European smart cities[5]. Tan et al. (2016) introduced the Global Livable Cities Index (GLCI), incorporating a broader range of indicators to account for economic competitiveness and diverse populations[6], but it could not still dynamically assess climate change impacts and emerging socio-economic trends.

The study by Saeed et al. (2020) emphasized a more integrated approach by using the Analytical Hierarchical Process (AHP) to develop a composite index that accounts for disparities among cities[7]. Najafi et al. (2024) evaluate urban livability in District 1 of Tehran using GIS to analyze physical, socio-economic, and accessibility indicators at the urban block level[8]. However, they were limited to a specific region and did not fully capture global or climate-related factors.

Benita et al. (2021) proposed a Spatial Livability Index for dense urban centers, which identified patterns based on geographical correlations but still did not address the dynamic, evolving nature of urban livability under climate change[9]. Tran et al. (2021) developed transportation livability-related indicators for a green urban road rating system in Taiwan[10]. However, the study's reliance on local conditions may limit its applicability to other regions, necessitating a more flexible framework to accommodate diverse urban contexts.

Agbali et al. (2019) explore the implementation of smart city strategies through stakeholder perspectives in Manchester, Boston, and San Diego, highlighting the significance of social and technological innovations in addressing urban challenges[11]. Al-Maliki et al. (2024) propose an ICT framework for developing smart cities in Saudi Arabia, addressing the unique challenges posed by the region's religious and cultural context while identifying potential benefits and obstacles[12]. However, the proposed model may require further refinement to adapt to rapid technological change. Kalenyuk et al. (2024) analyze the impact of digital financial technologies on smart cities, categorizing innovations like payments, lending, investing, and blockchain while emphasizing the need for significant resources and collaborations to implement smart projects[13], but they caution that these advancements also introduce risks such as cyber threats and data protection challenges. Therefore, current practices are often fragmented, region-specific, or static, lacking a predictive, comprehensive, and adaptable approach that can effectively guide long-term urban planning and policy-making. A comprehensive livability statistical model will lead to effective prediction of future.

4. Methodology

Dataset Collection

We've chosen 54 representative countries from all continents worldwide and collected their various aspects of data structured as below:

Criteria Layer	Factor Layer	ID	Source
Safety	Crime Rate	S1	WORLD POPULATION REVIEW[14] (WPR for short at below)
	Clearance Rate	S2	
	Homicide Rates	S3	
	Incidence of Sexual Violence	S4	
	Robbery Rate	S5	
	Police Officers	S6	NationMaster[15]
Health Care	Global Health Security Index	H1	
	Life expectancy at birth	H2	
	Birth rate	H3	
	Physicians > Per 1,000 people	H4	
	Hospital beds > Per 1,000 people	H5	
	Per capita health expenditure	H6	
	Nurses and midwives > per 1000	H7	
	Doctors > per 1000	H8	
	Universal Health Coverage Index	H9	WHO[16]
Purchasing Power	Average monthly disposable salary > After tax	P1	NationMaster
	Cost of Living	P2	
Cost of Living	Clothing	L1	
	Markets	L2	
	Rent	L3	
	Restaurants	L4	
	Basic utilities	L5	
	Groceries	L6	
	Sports and Leisure	L7	
	Telecommunication	L8	
	Transportation	L9	
Property Price to Income Ratio	Average monthly disposable salary > After tax	R1	
	Average Property Price	R2	Statistia[17]
Pollution	CO2 Emissions per 1000	U1	NationMaster
	PM10 > Country level > Micrograms per cubic meter	U2	
	NOx emissions per populated area	U3	
	Total renewable water resources	U4	
	Drinking Water Quality	U5	WPR
	Garbage Disposal	U6	
	Noise & Light Pollution	U7	
Traffic	Traffic Commute Time	T	Tomtom[18]
Climate	Average Temperature	C1	WorldEconomics[19]
	Average Dewpoint	C2	UNData[20]
	Natural Disaster Risk	C3	WPR

Database Building

Upon collecting all datasets, the next step was to clean and merge them into a cohesive and searchable database. The dataset was processed using Python and Pandas, where transformations were performed to standardize the various indicators.

Since data in categories of safety, healthcare and pollution have different scales and units, normalization is necessary. This project uses the data in NYC as baseline 100 to calculate the normalized X data as Z_X:

$$Z_X = \frac{X}{X_{NYC}} \times 100$$

The Index for single factor could be calculated as:

$$I_n = \sum w_X Z_X$$

Detailed w_X is given below, it can easily be seen that I_n of NYC will maintain to be 100.

	1	2	3	4	5	6	7	8	9
S	-0.3125	-0.25	-0.25	-0.1875	0.4	0.6			
H	0.15	0.2	0.05	0.15	0.1	0.15	0.05	0.1	0.05
U	-0.3333	-0.3333	-0.1667	-0.1667	0.375	0.375	0.25		

For data in categories of purchasing power, cost of living, property to income ratio and traffic, they have consistency. So the calculation method is more straightforward.

$$Z_P = \frac{P1}{P2} = \frac{P1}{L} \qquad Z_L = \sum_n Ln \qquad Z_R = \frac{R2}{R1} \qquad I_{P,L,R,T} = \frac{Z_{P,L,R,T}}{(Z_{NYC})_{P,L,R,T}} \times 100$$

For the category of climate, considering the specialty of the data, we refer to the formula given by Numbeo[21].

$$I_c = C1 + 0.555\left(6.1e^{5417.75\left(\frac{1}{273.16}-\frac{1}{C2+273.15}\right)} - 10\right) - 0.4C3$$

The overall formula for calculating the Livability Index is shown below.

$$I_{main} = 0.4I_P - I_R - 0.1I_L + 0.5I_S + 0.4I_H - 0.5I_T - 1.5^{-1}I_U + 3^{-1}I_C$$

After cleaning, all fragmented datasets were combined into a single searchable SQLite database which was integrated into a Python Flask backend, enabling efficient access for downstream analytics and visualization purposes. We also precomputed weighted metrics using custom Python scripts to allow for rapid querying. By combining all raw datasets into one structured and normalized database, we ensured the data was ready for machine learning analysis, predictive modeling, and insightful visualization.

Multi-model comparative analysis

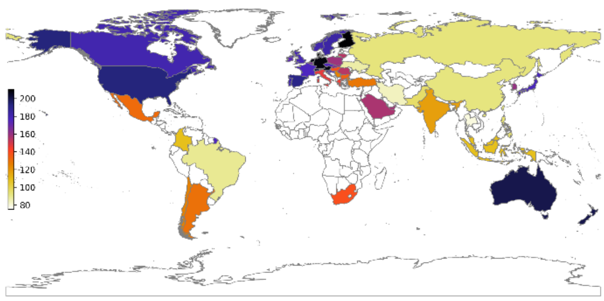
We utilized several machine learning algorithms to predict and evaluate the city livability index. The original dataset used in the prediction algorithm was calculated and analyzed using the aforementioned method, with a total of 16 sub-indices for 54 countries from 2017 to 2024, calculated every six months. For the practical significance of the data and the reliability of the model, we adopted the strategy of using Criteria Layer data, without choosing to use the final total index or raw data for prediction.

The main models employed include Random Forest Regression, Gradient Boosting Regression, Supported Vector Regression and k-Nearest Neighbors Regression.

After conducting a comprehensive qualitative and quantitative analysis and comparison of the livability prediction data generated by various algorithms for countries around the world in the coming years, we believe that the prediction results of the grading boost regression model are the most suitable for display.

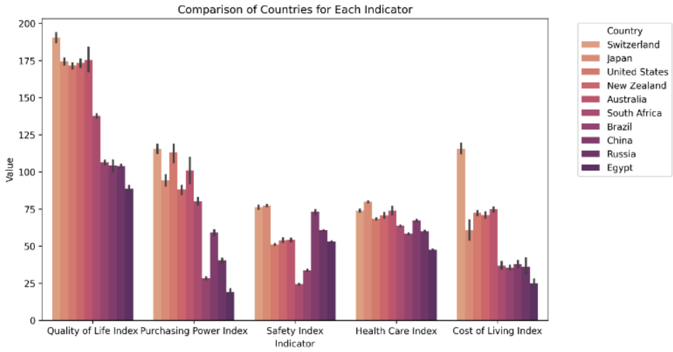
Visualization

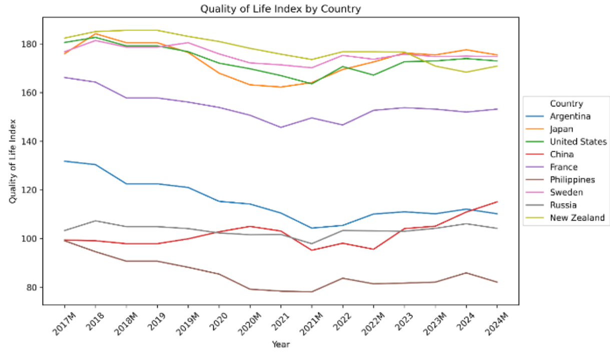
1. Index World Maps: The main index and criteria layer indexes will be presented on world maps. We visualized world livability data using libraries of geopandas and matplotlib in Python. For example, here is the 2017 Quality of Life Index visualization. The whole set of maps including prediction has presented on team's poster and demo.



2. Current Analysis: Illustrated a structured

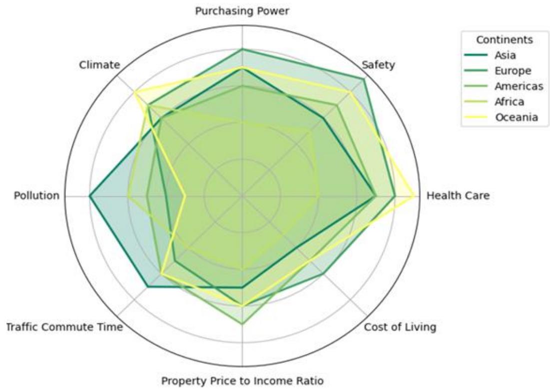
process for assessing livability, integrating various indicators into an indexed dataset with 9 chosen representative countries for each continent.





3. Trend Analysis: The line graph depicts the Quality-of-Life Index for 9 chosen representative countries from 2019 to 2024, showcasing fluctuations and trends over time.

4. Analysis by Continent: Visualization of Livability indicators corresponding to different continents. The average of each continent is used as a reference.



5. Experiment and Evaluation

Testbed Construction

The testbed for this experiment is designed to analyze the impact of multiple independent variables on the Quality-of-Life Index using machine learning models. To ensure the experiment's generalizability and reproducibility, all experiments are conducted with the same set of features and target variables, and evaluated using various models. The testbed construction involves the following aspects:

Independent Variables: These include a range of socio-economic and quality-of-life indicators, such as purchasing power, safety, cost of living, healthcare, etc.

Dependent Variable: The overall Quality-of-Life Index, representing a country's overall livability.

Model Selection: Several machine learning models (e.g., Linear Regression, Support Vector Machine, Random Forest) are employed for fitting the data and predicting the index.

Evaluation Metrics: Model performance is evaluated using R^2 and Mean Squared Error (MSE) to measure the accuracy of predictions[22].

Experiment Design

Before starting the experiment, the data will be cleaned to remove any missing or outlier values, ensuring the integrity and accuracy of the dataset. Next, relevant independent variables related to the quality-of-life index will be selected, including purchasing power, cost of living, safety, healthcare, climate, and others. Finally, the variables will be standardized using methods such as Z-score normalization to ensure they are on the same scale.

Several machine learning models will be used for fitting the data. The models selected are as follows:

Linear Regression: A basic regression model to assess the linear relationship between the independent and dependent variables.

Random Forest: An ensemble learning method that uses multiple decision trees to make predictions, suitable for capturing complex nonlinear relationships.

Support Vector Machine: A regression model that finds the best hyperplane to fit the data, capable of handling high-dimensional features and nonlinear relationships.

K-Nearest Neighbors (KNN): A model that predicts values based on the distances between sample points, useful for capturing nonlinear relationships.

Evaluation Metrics

The following metrics will be used to evaluate the performance of the models:

R^2 (Coefficient of Determination): Measures the proportion of variance in the dependent variable explained by the model. A value close to 1 indicates a better fit.

Mean Squared Error (MSE): Measures the difference between predicted and actual values. Lower MSE values indicate better model performance.
Sensitivity Analysis: Analyzes how different features affect the dependent variable (quality of life index), evaluating the contribution of each feature, and visualizing the sensitivity analysis through graphs.

Experimental Process

The dataset will be split into a training set and a test set, typically using 70% of the data for training and 30% for testing. The training set will be used to train the models, while the test set will be used to evaluate model performance. Each model will be trained on the training set, and evaluated on the test set by comparing R^2 values and MSE. The model with the best performance will be selected for further sensitivity analysis. One or more independent variables (e.g., purchasing power) will be adjusted within a specified range to observe changes in the quality-of-life index and the resulting sensitivity analysis curve will be plotted.

Model Performance Evaluation

In this experiment, we applied four machine learning models to fit the Quality-of-Life Index: Linear Regression, Random Forest, Support Vector Machine Regression, and K-Nearest Neighbors Regression. The performance of each model was evaluated using R^2 (Coefficient of Determination) and Mean Squared Error (MSE). The results are summarized in the following table:

Model	R^2 Value	Mean Squared Error (MSE)
Linear Regression	0.875	0.002
Random Forest	0.857	0.005
Support Vector Machine	0.96	0.007
K-Nearest Neighbors	0.94	0.010

From the table, it is evident that the Linear Regression model achieved the highest R^2 value (0.99) and the lowest MSE, indicating it performed best in this experiment. While other models also performed reasonably well, their R^2 values and MSE were slightly worse. These results suggest that the relationship between the Quality-of-Life Index and the predictor variables is relatively linear, making Linear Regression the most suitable model for this dataset.

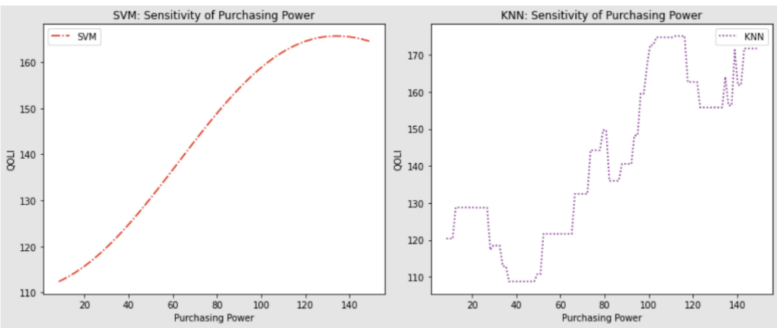
Sensitivity Analysis: Relationship Between Purchasing Power and Quality of Life Index

To further explore the impact of independent variables on the index, we selected "Purchasing Power" as the primary feature for analysis. Four different models were used to fit the relationship between Purchasing Power and Quality-of-Life Index. Sensitivity analysis curves were plotted to observe the effects of Purchasing Power on Quality-of-Life Index under each model. The results are as follows.

Sensitivity Analysis Plot

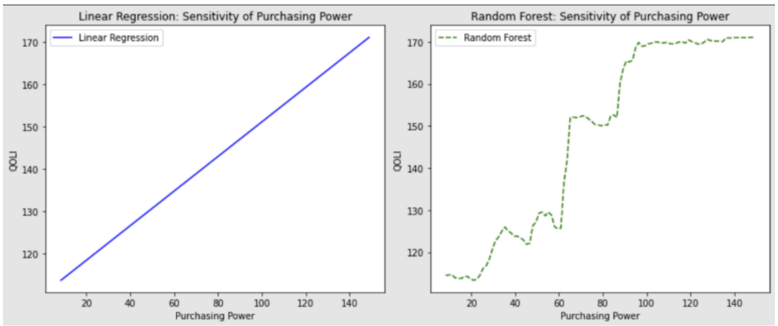
We first calculated the predicted values under the four models based on the range of values for "Purchasing Power," and plotted the corresponding sensitivity analysis curves. Each model's sensitivity plot illustrates the impact of Purchasing Power on Quality-of-Life Index at different levels.

Observations and Results



Linear Model: For the Linear Regression Model, the relationship between Purchasing Power and Quality-of-Life Index shows a relatively smooth linear trend. The model fits well and has a high R^2 value, indicating a strong linear association between the two variables.
Decision Tree: The Decision Tree Regressor reveals a more complex relationship between Purchasing Power and Quality-of-Life Index. The sensitivity plot shows some non-linear patterns, suggesting the model captures more intricate changes. However, the model also risks overfitting in certain ranges.
Random Forest: The Random Forest Regressor demonstrates robust performance in modeling the relationship[23]. The sensitivity plot is relatively smooth, and the model manages noise effectively, providing stable predictions across different values of Purchasing Power.

SVM: The Support Vector Machine Regressor (SVM) shows strong fitting ability in the higher ranges of Purchasing Power, while in lower ranges, the sensitivity is lower[24]. This suggests that the model may not fully capture the variations in the data at lower levels of Purchasing Power.



Results Analysis and Discussion

The models we used to predict the Total Quality-of-Life Index showed high accuracy across different algorithms. The strong performance of the models, with R^2 values close to 1, indicates that the Total Quality-of-Life Index is a reliable representation of the living conditions in the countries studied. The robustness of the model outputs, even when incorporating noise or variations in the data, further reinforces the reliability of the index.

In particular, the sub-indices that contribute to the Total index—such as Purchasing Power, Safety, and Health Care—demonstrated consistent relationships with the overall index. This highlights that these factors, individually and in combination, are crucial in determining the overall quality of life in different countries. The reliability of these indices offers a solid foundation for further studies on the impact of various factors on living conditions globally. Our results suggest that these indices could be widely applicable for comparative studies and for making informed decisions in policy-making and economic planning.

6. Conclusions and discussion

Our findings indicate that the livability index is a robust tool for measuring and understanding the quality of life, with strong model performance across multiple algorithms. The results demonstrate that, as the world recovers from the COVID-19 pandemic, overall livability has shown significant improvements.

In line with previous research, the livability of regions such as Western and Northern Europe, along with Northern America, remains consistently high, largely driven by their economic stability, health care systems, and low pollution levels. These regions are likely to maintain their position as the most livable areas in the foreseeable future.

In addition, while livability in China has historically lagged behind some developed regions, our study indicates that it is poised for significant improvements in the coming years. Thanks to two decades of achievements in various areas, China now exceeds many developed countries in certain indicators like safety and purchasing power. The data also shows that China has seen the most significant growth in livability after COVID-19. The prediction results which indicate that China will reach the livability levels of elementary developed countries can be given trustworthy.

While this study presents promising insights into global livability trends, it is essential to recognize that livability is a complex and multifaceted concept. Further research is needed to explore additional factors such as political stability, social equality, and environmental sustainability, which may also influence the quality of life in different regions.

Overall, the findings from this study contribute valuable data for policymakers and global organizations, providing a framework to assess and improve livability across nations. As global challenges, such as climate change and economic inequality, continue to evolve, the livability index we developed will be a useful tool for future analysis and decision-making.

7. Team Acknowledgement

Despite the major changes in the team members due to unexpected events such as the reorganization of the school district during the project, we still tried our best to overcome a series of difficulties and successfully made the planned progress, in which each of the existing team members equally shared the workload more than the original one, and we sincerely thank each of the team members for their unselfish contributions as well as the teacher, Dr. Liu, for her timely and crucial help.

Members of D. Zhang, Y. Wei, and X. Ma have contributed a similar amount of effort to all aspects of this project, including code, poster and reports.

REFERENCES

[1] Schindler, Mirjam, and Rita Dionisio. "Planning for Urban Livability: Integrating Socio-Spatial Indicators in City-Making." *Journal of Urban Technology* (2024): 1-30.
[2] Ruth, Matthias, and Rachel S. Franklin. "Livability for all? Conceptual limits and practical implications." *Applied geography* 49 (2014): 18-23.
[3] Ahmed, Nora Osama, Amr Mostafa El-Halafawy, and Ahmed Mohamed Amin. "A critical review of urban livability." *European Journal of Sustainable Development* 8.1 (2019): 165-165.
[4] Wang, Jiao, et al. "A comparative study of Beijing and three global cities: A perspective on urban livability." *Frontiers of Earth Science* 5 (2011): 323-329.
[5] Kutty, Adeeb A., et al. "Urban resilience and livability performance of European smart cities: A novel machine learning approach." *Journal of Cleaner Production* 378 (2022): 134203.
[6] Giap, Tan Khee, Woo Wing Thye, and Grace Aw. "A new approach to measuring the liveability of cities: the Global Liveable Cities Index." *World Review of Science, Technology and Sustainable Development* 11.2 (2014): 176-196.
[7] Saeed, Urooj, et al. "An integrated approach for developing an urban livability composite index—A cities' ranking road map to achieve urban sustainability." *Sustainability* 14.14 (2022): 8755.
[8] Najafi, Ehsan, et al. "A GIS-based Evaluation of Urban Livability using Factor Analysis and a Combination of Environmental and Socio-economic Indicators." Journal of Geovisualization and Spatial Analysis 8.2 (2024): 1-20.
[9] Benita, Francisco, Vyacheslav Kalashnikov, and Bige Tunçer. "A spatial livability index for dense urban centers." Environment and Planning B: Urban Analytics and City Science 48.7 (2021): 2006-2022.
[10] Tran, Nam Hoai, et al. "Developing transportation livability-related indicators for green urban road rating system in Taiwan." Sustainability 13.24 (2021): 14016.
[11] Agbali, Mohammed, et al. "Are smart innovation ecosystems really seeking to meet citizens' needs? Insights from the stakeholders' vision on smart city strategy implementation." Smart Cities 2.2 (2019): 307-327.
[12] Al-Maliki, Saeed Q. Al-Khalidi, Ahmed Abdulmutallab Ahmed, and Osman Ali Nasr. "The Establishment of Smart Cities: Existing Challenges and Opportunities–The Case of Saudi Arabia." (2024).
[13] Kalenyuk, Iryna, et al. "FINANCIAL INNOVATIONS IN THE SMART CITY ECOSYSTEM." Financial & Credit Activity: Problems of Theory & Practice 1.54 (2024).
[14]"Crime & Safety." World Population Review, worldpopulationreview.com/categories/crime-safety. Accessed 12 Oct. 2024.
[15]"Home." NationMaster, www.nationmaster.com. Accessed 29 Oct. 2024.
[16]"WHO | Health Indicators." World Health Organization, data.who.int/indicators/i/3805B1E/9A706FD. Accessed 19 Oct. 2024.
[17]"House Price Index: Emerging and Advanced Economies Worldwide." Statista, www.statista.com/statistics/1427342/house-price-index-emerging-and-advanced-economies-worldwide/. Accessed 1 Nov. 2024.
[18]"Traffic Index." TomTom, www.tomtom.com/traffic-index/. Accessed 8 Nov. 2024.
[19]"Temperatures." World Economics, www.worldeconomics.com/Indicator-Data/ESG/Environment/Temperatures/. Accessed 21 Oct. 2024.
[20]"Climate Change Indicators." UNdata, data.un.org/Data.aspx?d=CLINO&f=ElementCode%3A05. Accessed 22 Oct. 2024.
[21] "Indices Explained." Numbeo, www.numbeo.com/climate/indices_explained.jsp. Accessed 18 Nov. 2024.
[22] Botchkarev, Alexei. "Evaluating performance of regression machine learning models using multiple error metrics in azure machine learning studio." Available at SSRN 3177507 (2018).
[23] Cootes, Tim F., et al. "Robust and accurate shape model fitting using random forest regression voting." Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part VII 12. Springer Berlin Heidelberg, 2012.
[24] Cui, Dapeng, and David Curry. "Prediction in marketing using the support vector machine." Marketing Science 24.4 (2005): 595-615.