Geolocation Extraction From Reddit Text Data

Mila Stillman*, Anna Kruspe

Technische Hochschule Nürnberg, Fakultät Informatik, Hohfelderstraße. 40, Nürnberg
Hochschule München, Lothstraße 64, München

Abstract

Reddit has been an important source of news and information exchange in the past two decades. This social media platform is composed of communities forming spaces where users with common interests can share their experiences and opinions. Many of these online communities, called 'subreddits', are highly active, with some subreddits having tens of millions of followers. The extraction of geographic information from social media text data has been an increasing topic of research in the last years. While most of the work has been done on Twitter data, recent restrictions to its Application Programming Interface (API) have limited the access for academic research. Reddit.com could be a good alternative source due to its depth of discussions and communal interactions. Additionally, it allows more anonymity to users, thereby rendering it more ethically acceptable as it reduces users' traceability. However, this characteristic complicates the process of objective verification and evaluation. Nevertheless, we believe that Reddit is a rich source of information that could be used for training models for geolocation extraction. This paper presents a few first such experiments of extracting geographic data from location-based communities on Reddit and classifying posts on a city-level using BERT for text classification.

Keywords

Geospatial Text Mining, Social Media, BERT Classification, Reddit

1. Introduction

Geolocation extraction from text, also called geoparsing, is a subfield of geospatial data extraction and analysis, and refers to the determination of the geographic location of unstructured textual data including social media data. Geoparsing can be beneficial to the discovery of common interests, opinions or issues in specific geographic locations on different scales. The authors in [1] identified seven main application domains of geoparsing, namely geographic information retrieval, support during disaster events, crime management, disease spread management, traffic management, tourism management and spatial humanities. Geolocated social media posts were also found to be useful in remote sensing applications [2]. For example, social media data from Twitter was used to classify buildings in urban regions in combination with remote sensing images [3], to quickly obtain situational awareness in developing crises [4, 5], and to support the detection of misinformation spread during the Russo-Ukrainian war [6, 7, 8].

Reddit is a well established platform for online discourse with 52 million daily active users, 303.4 million posts and two billion comments in 2020 [9]. According to statista.com there are

GeoExT 2024: Second International Workshop on Geographic Information Extraction from Texts at ECIR 2024, March 24, 2024, Glasgow, Scotland

*Research conducted while at Technische Hochschule Nürnberg. Supported by Hochschule München.
mila.stillman@hm.edu (M. Stillman); anna.kruspe@hm.edu (A. Kruspe)
0009-0008-8976-8855 (M. Stillman); 0000-0002-2041-9453 (A. Kruspe)

around two billion monthly visits to Reddit, and more than 130,000 active communities, or subreddits [10]. Subreddits are sections on the website where discussions are organized in a forum-like fashion. Moderators ensure that posts are relevant and adhere to community rules, with automation playing an increasing role in the process [9]. Redditors are predominately male (around 65%), and the majority of web traffic comes from English speaking countries: USA (around 48.5%), Canada and UK (each approximately 7%) and Australia (4.63%). Germany is the sole non-English speaking country within the top five countries active on Reddit (around 3%) [11]. The range of topics of subreddits covers almost any conceivable topic. Thus, topics related to locations are included, i.e., subreddits exist for many large cities worldwide, as well as for smaller cities with local communities. In this paper, we conduct first experiments on Reddit posts from location-related subreddits using a purely textual approach. The premise underlying our study is that posts and comments from location-focused subreddits with ongoing moderation, are likely to contain textual cues enabling automated methods to determine the geographic origin directly from the content.

2. Related work

The topic of geographic information extraction from text is challenging due to considerable linguistic variability inherent in such data and its complex relations within a geospatial framework. A large part of the research was focused on Named Entity Recognition (NER), which is the categorization of important entities within unstructured text, specifically location related entities [1]. To-date, research using social media data was mainly focused on Twitter data due to the size of the platform and the previously available API, which included precise geolocation in the form of GPS-based coordinates and place location with variable granularity. Research on extracting geolocation based on the textual content, as well as based on user-location profiling or a combination of both has been conducted, e.g., in the shared task on geolocation prediction at the Workshop on Noisy User-generated Text (WNUT) in 2016 [12].

The authors in [13] used statistical machine learning methods such as K-Nearest Neighbor and Support Vector Machines to extract geolocation from Tweets without an indicated location. Others used deep neural networks [14] for this task. The authors in [15] used CNNs and user metadata, achieving an accuracy of 52.8% and 92.1% on city and country level predictions respectively. Research has employed Knowledge Graphs (KG) generated from Gazeteers and other geographic resources to link entities mentioned in Tweets with relevant geographic information [16, 17]. Wing and Baldridge [18] used a hierarchical approach via logistic regression models on nodes in geodesic grids. Huang and Carley [19] worked on hierarchical approaches using initial classification at a country-level, and subsequently refining it to city-level. Recently, Transformer-based language models, such as Bidirectional Encoder Representations from Transformers (BERT) [20] have also been used extensively due to their success in other NLP tasks [21, 22, 23, 24, 25]. E.g., [21] achieved an F1-score of 0.77 for location detection in Indonesia. Hybrid approaches have also been utilized, e.g., the authors in [25] used geographic data from OpenStreetMap [26] fused with a BERT model, achieving an F1-score of up to 0.88 on certain Twitter datasets. More detailed information about previous work on geoparsing using Twitter data can be found in the survey by Hu et al. [1].

Since 2023, Twitter API is not affordable for academic research. In addition, Twitter reduced Tweets with precise geolocation to less than 0.2% in 2021 [27] and many of the user-set place names and geolocations on Twitter were proven unreliable [28]. Therefore, research needs to be extended beyond those labels and platform. Reddit is one of the strong alternative candidates with a freely available API. To our knowledge, research on geoparsing using Reddit data was limited due to the lack of ground-truth data. Harrigian [29] attempted to classify Reddit user geolocation using subreddit metadata and extracting localized data via queries for questions such as 'where do you live?'. The results showed that models trained on Reddit data significantly outperformed transferred models trained on Tweets.

3. Methodology and Results

3.1. Data collection and model

We used a subset of the Reddit Data Dump dataset from Academic Torrents[1], which was collected using the Pushshift API [30]. Namely we used posts from the last quarter of 2022 for the training and validation, and tested the model on data from February 2023 to avoid immediate temporal influences. Subreddits of large cities by population in the USA and in Germany were chosen manually for the classification task (the full list is available under https://github.com/Milast/GeoExt_Reddit). Overall there were 181 subreddits from 126 cities that included the names of cities such as '/rBerlin'[2], as well as other location-related subreddits such as '/rChicagoFoods', '/rportlandmusic', etc., which were aggregated under the relevant city name. To train the model, only cities with over 100 posts during the mentioned time frame were included in the experiments, resulting in 57 cities being selected.

A pre-trained BERT model from Hugging Face [31] was used for the text classification. We chose the multilingual `bert-base-multilingual-cased` model, due to the ability of the case sensitive model to deal with the capitalization of nouns in the German language. The title and content of Reddit posts were combined and used for the classification. The code from the tutorial of Winastwan [32] was modified to fit the task. The model comprises a BERT tokenizer with a maximal text length of 512, which generated the input into the BERT transformer model with 12 layers of transformer encoder and hidden size of 768. A classification layer was used to discriminate between the cities. The fine-tuned model was then evaluated on the test data.

3.2. Experiments and results

First, a country-level classification was performed and achieved 98.2% accuracy (see AppendixC). This result was somewhat expected due to the nature of BERT embeddings, providing higher similarity between languages than semantic similarity [33]. Since the focus was on two countries with a large language difference (US and Germany), we then conducted a non-hierarchical city-level classification, assuming that country level differences would be captured automatically. After the city-level classification of the original text, we used the SpaCy library in Python [34] for the NER to keep posts that included Named Entities (NEs) that could imply, or infer, a

[1]www.academictorrents.com

[2]The notation '/r' indicates a subreddit on the platform.

Table 1
Average and macro testing results for the classification task using BERT for the three experiments

	Precision	Recall	F1-score	Precision (avg)	Recall (avg)	F1-score (avg)
Original text	0.64	0.44	0.50	0.54	0.47	0.48
Text filtered for location-inferring NER	0.68	0.54	0.58	0.61	0.57	0.57
Text filtered for location specific NER	0.77	0.65	0.69	0.72	0.68	0.69

location, such as organizations, events, etc. The NER tags used and their meanings can be found in the appendix A. After seeing an improvement of 0.08 in the overall F1-score, we pursued to further filtering the posts to keep those with NEs related specifically to a geographic location, namely 'GPE' (Countries, cities, states) and 'LOC' (Non-GPE locations, mountain ranges, bodies of water). The results are summarized in table 1. There was a significant improvement in the classification after filtering the posts for NEs that could potentially infer a location, and another improvement after filtering for location specific NEs. It is important to mention that the size of the dataset changed after filtering, which might have had an impact on the results. The size of the training and testing data, and the full results per city for the three experiments were added in appendixB and appendixC respectively. Some communities had more accurate classification results in all of three experiments, namely large German cities such as Munich, Berlin and Hamburg, as well as Las Vegas and Kanzas City achieving up to 70-80% accuracy while other cities only around 50%. Interestingly, both Las Vegas and Kanzas City included an aggregated subreddit '<city_name>_r4r', which stands for Redditors for Redditors and facilitates connections between users. These subreddits improved the classification results to a larger extent than other topics such as food or music, and could be studied further in the future.

4. Conclusion and future work

In this paper, we presented the initial results for geoparsing experiments on recent Reddit data. The analysis was done in a statistical manner without attempting to locate specific users, adhering to ethical recommendations. We performed these experiments using BERT classification on country and city levels. For countries, the classification worked near-perfectly due to the influence of language. For cities, we also included experiments where NER was used to detect posts that contain either location-inferring NEs, or location-specific NEs, showing an improving classification accuracy. There are many influences that should be analyzed next such as linguistic ambiguities, influences of specific NER tags and aggregated subreddits, and extension to other countries. Geoparsing using Large Language Models (LLMs) could also be compared to supervised classification, as well as utilizing KGs constructed from Gazeteers and other sources, with or without the assistance of LLMs. Other future work could be on hierarchical classification, using additional training data and different BERT models, and examining user networks in the localized communities. Although working with Reddit data is challenging due to the anonymous nature of the platform and the lack of explicit geolocation information, our experiments demonstrate an interesting basis for further research on geoparsing.

References

[1] X. Hu, Z. Zhou, H. Li, Y. Hu, F. Gu, J. Kersten, H. Fan, F. Klan, Location reference recognition from texts: A survey and comparison, ACM Computing Surveys 56 (2023) 1–37.

[2] X. X. Zhu, Y. Wang, M. Kochupillai, M. Werner, M. Häberle, E. J. Hoffmann, H. Taubenböck, D. Tuia, A. Levering, N. Jacobs, et al., Geoinformation harvesting from social media data: A community remote sensing approach, IEEE Geoscience and Remote Sensing Magazine 10 (2022) 150–180.

[3] M. Häberle, M. Werner, X. X. Zhu, Geo-spatial text-mining from twitter–a feature space analysis with a view toward building classification in urban regions, European journal of remote sensing 52 (2019) 2–11.

[4] A. Kruspe, Few-shot tweet detection in emerging disaster events, in: AI+HADR Workshop @ NeurIPS, 2019.

[5] A. Kruspe, J. Kersten, F. Klan, Review article: Detection of actionable tweets in crisis events, Natural Hazards and Earth System Sciences (NHESS) Special Issue: Groundbreaking technologies, big data, and innovation for disaster risk modelling and reduction (2020).

[6] S. Rode-Hasinger, M. Häberle, D. Racek, A. Kruspe, X. X. Zhu, TweEvent: A dataset of Twitter messages about events in the Ukraine conflict, in: Information Systems for Crisis Response and Management (ISCRAM), 2023.

[7] S. Rode-Hasinger, A. Kruspe, X. X. Zhu, True or False? Detecting False Information on Social Media Using Graph Neural Networks, in: Proceedings of the Eighth Workshop on Noisy User-generated Text (W-NUT 2022), Association for Computational Linguistics, Gyeongju, Republic of Korea, 2022, pp. 222–229. URL: https://aclanthology.org/2022.wnut-1.24.

[8] J. Niu, M. Stillman, P. Seeberger, A. Kruspe, A dataset of Open Source Intelligence (OSINT) Tweets about the Russo-Ukrainian war, in: Information Systems for Crisis Response and Management (ISCRAM), 2024.

[9] Reddit.com, Reddit's 2020 year in review, 2020. URL: https://www.redditinc.com/blog/reddits-2020-year-in-review.

[10] statista, Statista.com, 2024. URL: https://www.statista.com/statistics/443332/reddit-monthly-visitors/imilarweb.com/website/reddit.com/#overview.

[11] SimilarWeb, Reddit.com, 2024. URL: https://www.similarweb.com/website/reddit.com/#overview.

[12] B. Han, A. Rahimi, L. Derczynski, T. Baldwin, Twitter geolocation prediction shared task of the 2016 workshop on noisy user-generated text, in: Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT), 2016, pp. 213–217.

[13] T. Julie, M. Sadouanouan, T. Yaya, A geolocation approach for tweets not explicitly georeferenced based on machine learning, in: International Symposium on Distributed Computing and Artificial Intelligence, Springer, 2023, pp. 223–231.

[14] I. Lourentzou, A. Morales, C. Zhai, Text-based geolocation prediction of social media users with neural networks, in: 2017 IEEE International Conference on Big Data (Big Data), IEEE, 2017, pp. 696–705.

[15] B. Huang, K. M. Carley, On predicting geolocation of tweets using convolutional neural networks, in: Social, Cultural, and Behavioral Modeling: 10th International Conference,

SBP-BRiMS 2017, Washington, DC, USA, July 5-8, 2017, Proceedings 10, Springer, 2017, pp. 281–291.
[16] T. Miyazaki, A. Rahimi, T. Cohn, T. Baldwin, Twitter geolocation using knowledge-based methods, in: Proceedings of the 2018 EMNLP Workshop W-NUT: The 4th Workshop on Noisy User-generated Text, 2018, pp. 7–16.
[17] F. Lovera, Y. Cardinale, D. Buscaldi, T. Charnois, A knowledge graph-based method for the geolocation of tweets, in: Workshop Proceedings of the 19th International Conference on Intelligent Environments (IE2023), IOS Press, 2023, pp. 53–62.
[18] B. Wing, J. Baldridge, Hierarchical discriminative classification for text-based geolocation, in: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 336–348.
[19] B. Huang, K. M. Carley, A hierarchical location prediction neural network for twitter user geolocation, arXiv preprint arXiv:1910.12941 (2019).
[20] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).
[21] L. F. Simanjuntak, R. Mahendra, E. Yulianti, We know you are living in bali: Location prediction of twitter users using bert language model, Big Data and Cognitive Computing 6 (2022) 77.
[22] I. Salogni, Salogni at geolingit: Geolocalization by fine-tuning bert, in: Proceedings of the Eighth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2023), CEUR. org, Parma, Italy, 2023.
[23] K. Lutsai, C. H. Lampert, Geolocation predicting of tweets using bert-based models, arXiv preprint arXiv:2303.07865 (2023).
[24] Y. Scherrer, N. Ljubešić, Helju@ vardial 2020: Social media variety geolocation with bert models, in: Proceedings of the 7th Workshop on NLP for Similar Languages, Varieties and Dialects, 2020, pp. 202–211.
[25] X. Hu, Z. Zhou, J. Kersten, M. Wiegmann, F. Klan, Gazpne2: A general and annotation-free place name extractor for microblogs fusing gazetteers and transformer models (2021).
[26] OpenStreetMap contributors, Planet dump retrieved from https://planet.osm.org , https://www.openstreetmap.org, 2017.
[27] A. Kruspe, M. Häberle, E. J. Hoffmann, S. Rode-Hasinger, K. Abdulahhad, X. X. Zhu, Changes in twitter geolocations: Insights and suggestions for future usage, arXiv preprint arXiv:2108.12251 (2021).
[28] A. Kumar, J. P. Singh, N. P. Rana, Authenticity of geo-location and place name in tweets (2017).
[29] K. Harrigian, Geocoding without geotags: A text-based approach for reddit. arxiv, 2018.
[30] J. Baumgartner, S. Zannettou, B. Keegan, M. Squire, J. Blackburn, The pushshift reddit dataset, in: Proceedings of the international AAAI conference on web and social media, volume 14, 2020, pp. 830–839.
[31] H. Face, huggingface.co, 2024. URL: https://huggingface.co/bert-base-multilingual-cased.
[32] R. Winastwan, Text classification with bert in pytorch, 2021. URL: https://towardsdatascience.com/text-classification-with-bert-in-pytorch-887965e5820f.
[33] J. Libovickỳ, R. Rosa, A. Fraser, How language-neutral is multilingual bert?, arXiv preprint arXiv:1911.03310 (2019).

[34] M. Honnibal, I. Montani, spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing, 2017. To appear.

A. Named Entity Recognition (NER) tags

Table 2
NER tags used during the city-level classification experiments.

NER Tag	Description
GPE	Countries, cities, states.
LOC	Non-GPE locations, mountain ranges, bodies of water.
FAC	Buildings, airports, highways, bridges, etc.
ORG	Companies, agencies, institutions, etc.
MONEY	Monetary values, including unit.
NORP	Nationalities or religious or political groups.
EVENT	Named hurricanes, battles, wars, sports events, etc.
PRODUCT	Objects, vehicles, foods, etc. (Not services.)
WORK_OF_ART	Titles of books, songs, etc.

B. Data volume

Table 3
Data volume for the training and testing data for the three experiments.

	Training data	Testing data
Original text (and Country-level)	114,530	45,365
Text filtered for location inferable NER	83,288	33,174
Text filtered for location specific NER	52,996	21,385

C. Extended results for the country-level and city-level experiments

Table 4
Results for the country-level classification on the testing data.

	precision	recall	f1-score	support
USA	0.99	0.99	0.99	43240
Germany	0.79	0.79	0.79	2125
Weighted Avg Accuracy	0.98	0.98	0.98	45365

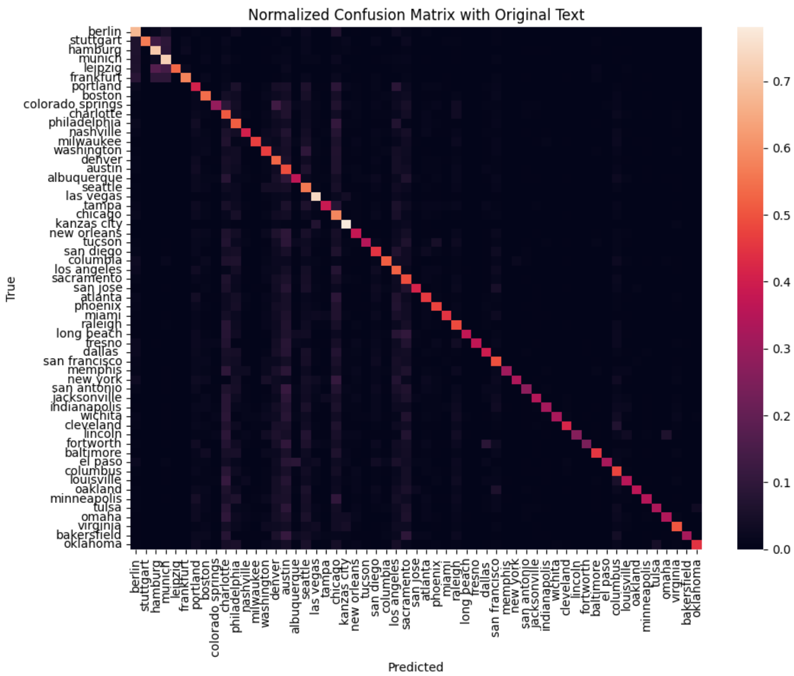

Figure 1: Normalized confusion matrix on original text of Reddit submissions.

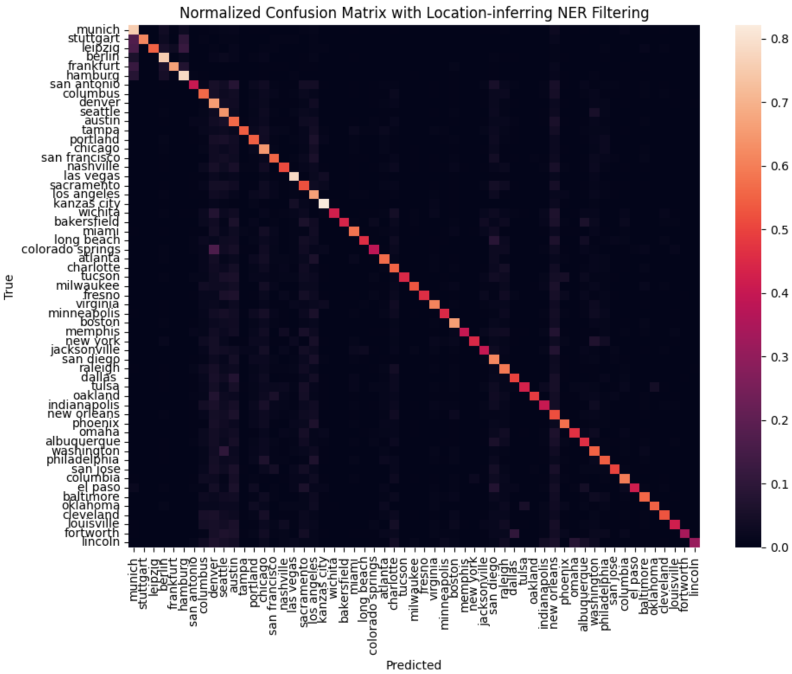

Figure 2: Normalized confusion matrix on Reddit submissions after NER filtering for location-inferring named entities.

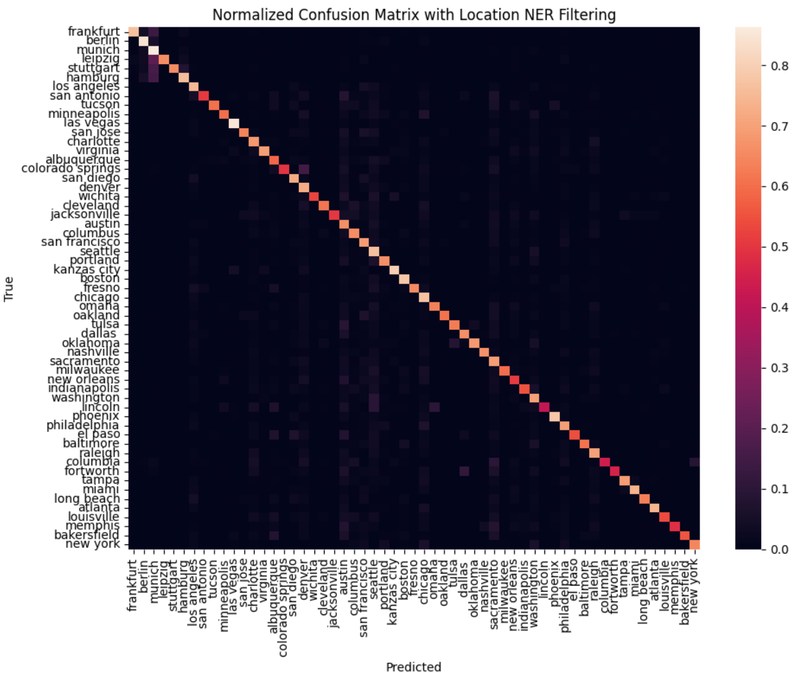

Figure 3: Normalized confusion matrix after NER filtering for location-specific named entities.

Table 5
Results by city for the three experiments including precision, recall, F1-score and number of supporting test samples.

	Location NER				Location-implying NER				Original text			
City	P	R	F1	Support	P	R	F1	Support	P	R	F1	Support
berlin	0.87	0.84	0.86	474	0.73	0.76	0.74	629	0.52	0.67	0.59	867
frankfurt	0.94	0.77	0.84	142	0.92	0.67	0.77	165	0.86	0.58	0.69	192
hamburg	0.86	0.76	0.81	188	0.72	0.79	0.75	219	0.66	0.71	0.68	257
leipzig	1.00	0.66	0.79	111	1.00	0.55	0.71	133	1.00	0.53	0.69	143
munich	0.70	0.86	0.77	371	0.69	0.76	0.72	470	0.65	0.72	0.69	535
stuttgart	1.00	0.66	0.79	99	1.00	0.62	0.76	110	0.96	0.56	0.71	131
albuquerque	0.44	0.59	0.50	248	0.49	0.46	0.48	437	0.52	0.36	0.43	601
atlanta	0.81	0.74	0.77	424	0.61	0.57	0.59	672	0.50	0.46	0.48	938
austin	0.58	0.66	0.62	1225	0.51	0.56	0.53	1990	0.39	0.49	0.43	2845
bakersfield	0.98	0.56	0.71	102	0.97	0.44	0.61	152	0.97	0.32	0.48	217
baltimore	0.74	0.61	0.67	303	0.60	0.56	0.58	452	0.72	0.45	0.55	564
boston	0.81	0.78	0.79	848	0.77	0.66	0.71	1185	0.66	0.54	0.59	1460
charlotte	0.63	0.68	0.65	623	0.52	0.55	0.54	972	0.25	0.51	0.34	1289
chicago	0.60	0.77	0.67	851	0.50	0.65	0.57	1331	0.32	0.58	0.41	1897
cleveland	0.80	0.62	0.70	352	0.79	0.53	0.63	489	0.75	0.42	0.54	615
colorado springs	0.88	0.50	0.64	221	0.55	0.38	0.45	360	0.60	0.29	0.39	479
columbia	0.94	0.44	0.60	104	0.45	0.60	0.52	219	0.48	0.52	0.50	307
columbus	0.70	0.66	0.68	668	0.55	0.56	0.55	970	0.50	0.48	0.49	1245
dallas	0.82	0.66	0.73	443	0.76	0.50	0.60	697	0.59	0.39	0.47	965
denver	0.64	0.72	0.68	954	0.46	0.66	0.54	1512	0.43	0.53	0.48	2040
el paso	0.90	0.54	0.68	101	0.90	0.42	0.57	134	0.90	0.32	0.47	171
fortworth	0.90	0.46	0.61	158	0.90	0.34	0.49	249	0.86	0.24	0.38	345
fresno	0.99	0.66	0.79	122	0.99	0.46	0.63	180	0.99	0.37	0.54	234
indianapolis	0.63	0.55	0.59	245	0.56	0.40	0.47	443	0.67	0.32	0.43	586
jacksonville	0.79	0.50	0.62	222	0.44	0.40	0.42	328	0.57	0.34	0.43	445
kanzas city	0.76	0.80	0.78	218	0.71	0.82	0.76	537	0.68	0.78	0.73	993
las vegas	0.89	0.85	0.87	638	0.78	0.78	0.78	1133	0.74	0.74	0.74	1723
lincoln	0.88	0.41	0.56	109	0.81	0.31	0.45	179	0.81	0.26	0.39	219
long beach	0.84	0.63	0.72	174	0.70	0.46	0.56	275	0.82	0.37	0.51	364
los angeles	0.66	0.75	0.70	698	0.43	0.67	0.52	1042	0.33	0.52	0.40	1486
louisville	0.74	0.54	0.62	350	0.79	0.42	0.55	518	0.58	0.35	0.44	659
memphis	0.93	0.49	0.64	206	0.93	0.40	0.56	338	0.78	0.31	0.44	478
miami	0.88	0.74	0.80	395	0.66	0.58	0.62	532	0.74	0.45	0.56	725
milwaukee	0.93	0.60	0.73	310	0.81	0.53	0.64	447	0.77	0.47	0.58	541
minneapolis	0.75	0.60	0.67	290	0.73	0.45	0.56	445	0.69	0.35	0.46	628
nashville	0.88	0.67	0.76	400	0.63	0.51	0.57	606	0.56	0.40	0.47	835
new orleans	0.49	0.51	0.50	434	0.25	0.52	0.34	760	0.53	0.38	0.44	1047
new york	0.64	0.65	0.64	99	0.78	0.45	0.57	159	0.74	0.33	0.46	212
oakland	0.89	0.61	0.72	216	0.89	0.49	0.63	296	0.88	0.36	0.51	376
oklahoma	0.87	0.68	0.77	98	0.82	0.55	0.66	171	0.85	0.45	0.58	229
omaha	0.87	0.63	0.73	221	0.84	0.47	0.60	349	0.83	0.34	0.48	483
philadelphia	0.56	0.70	0.62	508	0.63	0.54	0.59	935	0.36	0.52	0.42	1473
phoenix	0.79	0.78	0.79	417	0.70	0.58	0.63	650	0.57	0.47	0.52	884
portland	0.63	0.66	0.64	426	0.51	0.55	0.53	757	0.36	0.41	0.38	1159
raleigh	0.46	0.70	0.55	344	0.43	0.59	0.50	583	0.38	0.49	0.43	746
sacramento	0.42	0.69	0.52	518	0.41	0.52	0.46	880	0.29	0.49	0.37	1236
san antonio	0.81	0.51	0.62	321	0.63	0.40	0.49	559	0.65	0.27	0.38	825
san diego	0.70	0.72	0.71	532	0.42	0.62	0.50	826	0.49	0.45	0.47	1179
san francisco	0.53	0.69	0.60	526	0.61	0.56	0.58	833	0.48	0.50	0.49	1115
san jose	0.85	0.64	0.73	271	0.78	0.50	0.61	406	0.65	0.41	0.51	527
seattle	0.58	0.76	0.66	1125	0.51	0.64	0.57	1580	0.41	0.56	0.47	2014
tampa	0.83	0.69	0.75	280	0.79	0.54	0.64	409	0.57	0.39	0.46	606
tucson	0.96	0.60	0.74	253	0.95	0.45	0.61	408	0.83	0.35	0.49	581
tulsa	0.90	0.62	0.74	194	0.87	0.43	0.58	308	0.88	0.33	0.48	404
virginia	0.76	0.69	0.72	377	0.66	0.61	0.63	525	0.67	0.51	0.58	665
washington	0.56	0.70	0.62	688	0.53	0.55	0.54	1019	0.51	0.47	0.49	1311
wichita	0.96	0.53	0.68	150	0.97	0.43	0.59	211	0.95	0.32	0.48	274