A Comparative Evaluation of IP Geolocation Services for Brazilian Data Enrichment

Pedro G. B. Vieira, Lucas G. L. Costa, Gisele L. Pappa

[1]Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte, MG – Brasil

{pedrobarruetavena, lucas.lage, glpappa}@dcc.ufmg.br

Abstract. *IP geolocation information is fundamental for enhancing analytical datasets, supporting applications such as recommendation systems, investigation, service personalization, among others digital fields. However, there are several geolocation discovery free online services as database enrichment tools. This work evaluate several of those services, aiming to define the most reliable service for the Brazilian network context, and compares their performance across IPv4 and IPv6. The ground truth set is extracted from Ripe Atlas's brazilians devices. The experiment shows that BigData Cloud is the best service and have better results with IPv6 results. Despite the inherent limitation in obtaining reliable IP-based geolocation data, these findings help practitioners select more reliable services for geolocation enrichment of Brazilian IP datasets.*

1. Introduction

The increasing integration of digital infrastructure into everyday activities has reinforced the importance of IP geolocation data for applications such as recommendation systems, cybersecurity, fraud detection, digital investigations, service personalization, and network analysis. In Brazil, accurate geolocation information is particularly relevant due to the country's territorial scale and regional diversity. However, since the Internet Protocol (IP) was not designed to provide geographic information, IP addresses do not inherently contain reliable location metadata. As a result, analytical datasets often require enrichment through external IP geolocation services capable of estimating attributes such as country, state, city, and coordinates.

Currently, several free services provide IP geolocation APIs and databases, but selecting the most suitable provider remains challenging, especially for Brazilian IP addresses, whose geolocation quality may vary considerably across services. Additionally, the growing adoption of IPv6 raises questions about potential differences in geolocation accuracy between IPv4 and IPv6 addresses.

In this work, we investigate the effectiveness of free IP geolocation services for Brazilian IP addresses. Our study is guided by the following research questions, focusing on the evaluation of IP geolocation services as data enrichment tools for improving the quality and completeness of analytical datasets:

RQ1 - Best Service for Brazilian IP Geolocation. Which free IP geolocation service provides the highest accuracy for Brazilian IP addresses? Answering this question supports the enrichment of Brazilian IP datasets by helping practitioners identify the most reliable service for this context.

RQ2 - Geolocation Differences Between Brazilian IPv4 and IPv6 Addresses. Are there significant differences in geolocation accuracy between Brazilian IPv4 and IPv6 addresses? Answering this question helps practitioners assess the impact of the IP version on geolocation reliability when enriching Brazilian IP datasets.

To answer these questions, we conducted an empirical experiment using geolocation data collected from Brazilian devices available in the RIPE Atlas platform as a ground-truth dataset. We queried multiple IP geolocation services using the collected IP addresses and compared the returned locations against the ground-truth coordinates to evaluate and compare the accuracy of each service.

Related works. Geospatial information has become a strategic resource for several applications, including IoT, analytical systems, and personalization mechanisms [Silva and Holanda 2025, Leite et al. 2024]. However, many datasets still lack explicit geographic attributes, creating the need for IP geolocation services that enrich non-geographic data with spatial context. Although several studies have evaluated the accuracy and consistency of popular IP geolocation databases and services [Nur 2023, Xu et al. 2018, Callejo et al. 2022], comparative analyses remain limited, particularly those focused on the Brazilian context and the many available geolocation services.

2. Data

Obtaining reliable ground-truth data for IP geolocation is a challenging task. The Internet Protocol do not inherently provide accurate location information, and publicly available geolocation data is often limited or imprecise. To address this challenge, we used data from RIPE Atlas[1], a large-scale global measurement network composed of thousands of probes distributed worldwide. These probes are maintained by volunteers, who provide metadata about their devices, including geographic location information. RIPE Atlas exposes APIs that allow researchers to retrieve the most recent IP addresses and metadata associated with the probes.

Previous work by [Izhikevich et al. 2024] indicates that RIPE Atlas is a reliable source of geolocation ground truth, reporting inconsistencies in only 1.5% to 3.96% of the analyzed probes under their evaluation methodology. This suggests that the vast majority of probe owners provide correct and trustworthy location information.

The dataset was collected on April 20, 2026, using the most recent probe information available at the time of extraction. In total, the dataset contains 109 probes with valid IPv4 addresses, of which 64 also provided valid IPv6 addresses. The number of available IPs is inherently dynamic and may vary over time due to probe connectivity and address reassignment. Furthermore, Figure 1, which presents the distribution of probe uptime in days, shows that most probes were recently activated or reconnected, increasing confidence that their geolocation metadata had been updated recently.

2.1. Location Inference Method

To determine the city, state, and country associated with each probe based on its geolocation coordinates, we performed spatial analyses using GeoPandas[2] and geographic shapefiles containing administrative boundaries. The geographic coordinates provided by RIPE Atlas were spatially matched against these boundary datasets to infer the corresponding municipality, state, and country for each probe.

[1]RIPE Atlas `https://atlas.ripe.net/`
[2]GeoPandas `https://geopandas.org/en/stable/`

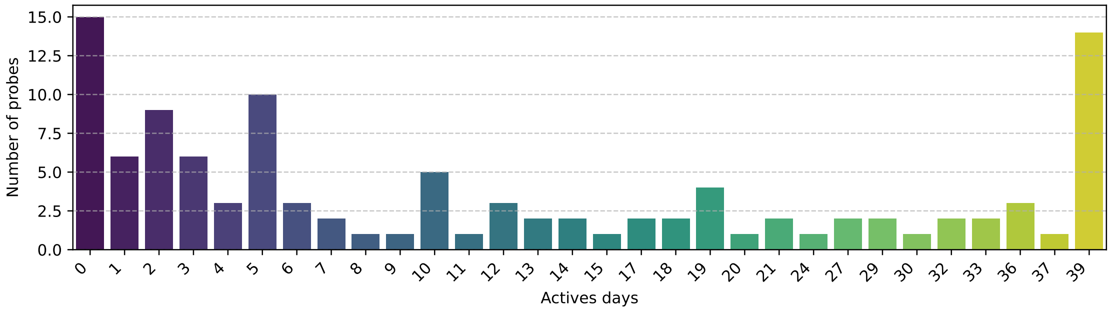


Figure 1. Distribution of active days for the probes.

To identify whether a probe was located within Brazilian territory, we used shapefiles from the Natural Earth Project[3], a public repository of geographic datasets. For municipality-level inference, we employed the official Brazilian municipal mesh provided by IBGE[4], which contains the administrative boundaries of Brazilian cities. This dataset enabled the assignment of each probe coordinate to its corresponding municipality.

The resulting dataset contains IP addresses distributed across 52 Brazilian cities. São Paulo has the most probes (24), followed by Osasco and Porto Alegre, with 9 and 7 probes, respectively. Additionally, 40 cities are represented by a single IP address, reflecting a geographically diverse but unevenly distributed dataset.

3. Experimental Evaluation

The evaluation process is organized into four parts: the adopted methodology (Section 3.1), the selection of the analyzed services (Section 3.2), the experimental setup (Section 3.3), and the discussion of the obtained results (Section 3.4).

3.1. Methodology

The experimental methodology was structured into three complementary steps to ensure a consistent and reliable evaluation process. First, we collected RIPE Atlas probes and applied the location inference method described in Section 2.1 to identify Brazilian probes and determine their corresponding municipalities, states, and countries, establishing the ground-truth dataset. Second, we searched for publicly available IP geolocation APIs and selected services according to predefined technical criteria, prioritizing widely accessible free solutions. Third, we queried all selected services using the collected IPv4 and IPv6 addresses and stored the returned geolocation information for subsequent analysis. Based on these results, we conducted two evaluations to answer RQ1 and RQ2.

3.2. Data Enrichment Service Selection

The selection of IP geolocation services was based on a search for prominent providers that satisfied the technical requirements of the experimental setup. To ensure feasibility and consistency, the selected services were required to: (i) provide geographic coordinates via free-tier APIs; (ii) support both IPv4 and IPv6 addresses; and (iii) offer sufficient rate limits to execute the experiment without interruption. The most relevant services satisfying

[3]Natural Earth Data `https://www.naturalearthdata.com/`
[4]IBGE Municipal Mesh 2025 `https://www.ibge.gov.br/`

these criteria were selected, resulting in a total of 19 IP geolocation services evaluated in this study, as shown in Table 1.

3.3. Experimental Setup

Among the 109 collected RIPE Atlas probes, all probes provided valid IPv4 addresses, while 64 also provided valid IPv6 addresses. Based on this dataset, two evaluations were conducted: (i) An evaluation using all 109 IPv4 addresses together with the 64 available IPv6 addresses, aiming to answer RQ1. (ii) A second evaluation restricted to the 64 probes containing both IPv4 and IPv6 addresses, aiming to answer RQ2 by enabling a direct comparison between the two protocols under the same conditions.

In both evaluations, all 19 selected services were queried, and their geolocation responses were stored for later processing. After retrieving the coordinates returned by each service, we applied the location inference method described in Section 2.1 to determine the corresponding city, state, and country associated with each response.

The evaluation considered the geographic error, measured in kilometers, between the ground-truth coordinates and those returned by each service. Based on these distances, we computed descriptive statistics, including mean error, standard deviation, median error, and the 75th percentile. Additionally, we conducted pairwise statistical significance tests comparing the best-performing service against all others. Finally, we compared the inferred city, state, and country against the ground truth to assess categorical geolocation accuracy at different administrative levels.

3.4. Results

This section presents and discusses the results obtained in the experimental evaluation, aiming to answer the two proposed research questions.

RQ1 - Best Service for Brazilian IP Geolocation. Table 1 presents the results of the first evaluation, including all analyzed services, metrics, and statistical test results. The results reveal substantial variation in geolocation performance across providers, likely due to differences in their underlying databases and inference methodologies.

Overall, the mean geolocation error remained relatively high, indicating limited precision for Brazilian IP addresses. Likewise, city-level accuracy results show that most services still struggle to provide reliable geolocation at the municipal scale. The results also reveal high standard deviation values and substantial differences between mean and median geolocation errors for most services. This behavior suggests the presence of significant outliers, in which a small number of predictions with extremely large errors inflate the mean distance error, while the median remains comparatively lower.

Among all evaluated services, BigData Cloud achieved the best overall performance across most metrics. In the IPv6 evaluation, BigData Cloud obtained the best results for all considered metrics. For IPv4, the service also dominated most metrics, with the exception of mean error and state-level accuracy, where IP2Location achieved slightly better results. However, the median and 75th percentile metrics indicate that the lower mean error observed for IP2Location is strongly influenced by outlier behavior.

In particular, BigData Cloud achieved median geolocation errors of only 0.3 km for both IPv4 and IPv6, while maintaining the 75th percentile below 8 km in both cases.

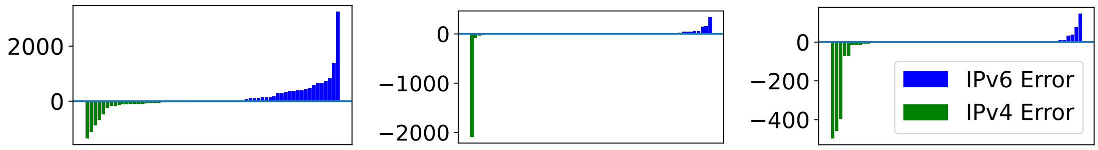
	IPV4							IPV6						
					Accuracy (%)							Accuracy (%)		
Service	$\bar{E}$	M_d	P_{75}	**Wilcoxon**	c	s	n	$\bar{E}$	M_d	P_{75}	**Wilcoxon**	c	s	n
BigData Cloud [BigData 2026]	99.3 ± 268.0	**0.3**	**7.9**	–	**76**	90	**100**	**43.4 ± 160.6**	**0.3**	**0.3**	–	**86**	**97**	**100**
IP2Location [IP2Location 2026]	**92.4 ± 325.5**	5.7	15.7	$W = 1,629$ $p < 0.05$	62	**93**	**100**	65.0 ± 196.4	6.9	16.2	$W = 311$ $p < 0.05$	60	94	**100**
IP Geolocation [IPGeolocation 2026]	130.7 ± 396.2	6.4	32.0	$W = 1,652$ $p < 0.05$	59	91	100	413.2 ± 1557.9	8.7	100.1	$W = 199$ $p < 0.05$	59	84	97
IP-API [IP-API 2026]	220.0 ± 1083.1	5.5	29.5	$W = 1,821$ $p < 0.05$	64	92	96	110.6 ± 352.0	9.4	33.4	$W = 287$ $p < 0.05$	63	92	97
API Techniknews [Techniknews 2026]	220.0 ± 1083.1	5.5	29.5	$W = 1,821$ $p < 0.05$	64	92	96	110.6 ± 352.0	9.4	33.4	$W = 287$ $p < 0.05$	63	92	97
ipwhois.io [Ipwhois.io 2026]	228.3 ± 981.1	7.0	48.4	$W = 1,508$ $p < 0.05$	62	91	99	407.5 ± 1,630.1	8.3	73.4	$W = 275$ $p < 0.05$	59	89	97
FindIP [FindIP 2026]	266.4 ± 993.6	7.3	79.1	$W = 1,328$ $p < 0.05$	53	87	99	389.4 ± 1,651.6	6.4	49.3	$W = 254$ $p < 0.05$	59	90	97
Geoapify [Geoapify 2026]	266.4 ± 993.6	7.3	79.1	$W = 1,328$ $p < 0.05$	53	87	99	389.4 ± 1,651.6	6.4	49.3	$W = 254$ $p < 0.05$	59	90	97
Free IP API [FreeIPAPI 2026]	322.8 ± 1,239.7	7.3	78.2	$W = 1,351$ $p < 0.05$	52	87	98	387.1 ± 1,651.9	6.4	45.3	$W = 257$ $p < 0.05$	60	90	97
IPWho [IPWho 2026]	492.0 ± 2,259.6	6.5	26.0	$W = 1,727$ $p < 0.05$	62	84	90	309.6 ± 1,168.0	9.8	62.3	$W = 337$ $p < 0.05$	57	83	87
IP2 Info [Info 2026]	516.4 ± 2,314.4	5.3	31.0	$W = 1,711$ $p < 0.05$	64	84	88	441.8 ± 1,204.2	15.8	370.7	$W = 181$ $p < 0.05$	46	67	73
apiip [Apiip 2026]	532.1 ± 2,316.5	5.5	46.7	$W = 1,599$ $p < 0.05$	63	83	88	441.9 ± 1,204.1	15.8	370.7	$W = 181$ $p < 0.05$	46	67	73
ipdata [Ipdata 2026]	551.8 ± 2,316.1	5.8	85.2	$W = 1,631$ $p < 0.05$	59	82	87	464.3 ± 1,205.5	17.2	372.9	$W = 172$ $p < 0.05$	44	63	71
Geolocated.io [Geolocated.io 2026]	607.9 ± 1,991.1	12.6	286.3	$W = 1,196$ $p < 0.05$	50	80	95	679.5 ± 2,182.2	7.1	80.1	$W = 213$ $p < 0.05$	57	86	94
IPDetails [IPDetails 2026]	620.5 ± 2,343.6	6.0	134.4	$W = 1,313$ $p < 0.05$	56	79	86	263.6 ± 536.6	15.0	366.6	$W = 225$ $p < 0.05$	48	70	76
IPstack [IPstack 2026]	689.6 ± 2,028.8	14.5	155.0	$W = 1,325$ $p < 0.05$	49	80	93	1,550.2 ± 2,938.0	23.3	813.9	$W = 118$ $p < 0.05$	46	70	83
ipapi [Ipapi 2026]	689.6 ± 2,028.8	14.5	155.0	$W = 1,325$ $p < 0.05$	49	80	93	1,550.2 ± 2,938.0	23.3	813.9	$W = 118$ $p < 0.05$	46	70	83
The IP API [TheIPAPI 2026]	1,517.3 ± 3,663.5	27.5	763.5	$W = 793$ $p < 0.05$	31	68	87	934.7 ± 2,183.8	18.7	348.0	$W = 155$ $p < 0.05$	43	75	89
Abstract API [Abstract 2026]	2,353.2 ± 3,548.9	935.6	1,754.1	$W = 234$ $p < 0.05$	14	21	80	441.8 ± 1,204.2	15.8	370.7	$W = 181$ $p < 0.05$	46	67	73

Table 1. All results: $\bar{E}$, M_d, and P_{75} denote the mean ± standard deviation, median, and 75th percentile of the geolocation error (km), respectively. Wilcoxon reports pairwise comparisons against BigData Cloud based on geolocation error (km). c, s, and n denote city, state, and country accuracy.

In contrast, IP2Location presented median errors above 5 km and 75th percentile values above 15 km for both IP versions. These results indicate that BigData Cloud provides the most consistent and accurate geolocation estimates overall.

To statistically validate these findings, we performed 18 pairwise comparisons using the Wilcoxon signed-rank test, comparing BigData Cloud against each of the other services in terms of geolocation error (km), as reported in Table 1. All tests indicated statistically significant differences ($p < 0.05$), with BigData Cloud consistently achieving lower error across all comparisons. These results remain robust when considering multiple comparisons across the 18 tests.

RQ2 - Performance Differences Between Brazilian IPv4 and IPv6 Addresses. Figure 2 presents the results of the second experiment, comparing the geolocation performance between IPv4 and IPv6 addresses. Figures 2a and 2b illustrate the differences in mean and median geolocation error, respectively, across all evaluated services. Additionally, Figure 2c highlights the comparison between IPv4 and IPv6 errors specifically for BigData Cloud, the best-performing service identified in the first evaluation.



(a) Mean Error. (b) Median Error. (c) BigData Cloud Error.

Figure 2. X-axis: all probes; Y-axis: difference between IPv6 and IPv4 geolocation errors (km). (a) Mean error across services, (b) median error across services, and (c) Error for the BigData Cloud service.

The results suggest that, when considering all services together, IPv4 generally achieves lower mean and median geolocation errors than IPv6, indicating better overall geolocation precision. The discrepancy between mean and median values once again reinforces the existence of outliers affecting the error distribution.

However, this behavior differs for BigData Cloud. Unlike the overall trend observed across services, BigData Cloud achieved better performance for IPv6 addresses, producing lower geolocation errors than IPv4. This result suggests that, besides being the most accurate service overall for Brazilian IP geolocation, BigData Cloud also demonstrates superior capability in handling IPv6 geolocation data.

4. Conclusion

This study evaluated several free IP geolocation services as data enrichment tools for improving the quality and completeness of analytical datasets, focusing on their geographic precision and administrative-level accuracy for Brazilian IP addresses. To conduct the evaluation, a ground-truth dataset was built from Brazilian RIPE Atlas probes containing both IPv4 and IPv6 addresses.

The results show that BigData Cloud achieved the best overall geolocation performance among the evaluated services. Additionally, although IPv4 generally presented lower geolocation errors across the full set of services, BigData Cloud obtained better results with IPv6 addresses, making it the most effective option for the proposed task. These findings help practitioners select more reliable services for geolocation enrichment of Brazilian IP datasets.

Limitations and Future Work. This work is limited by the inherent difficulty of obtaining reliable IP geolocation ground truth, which restricts the volume of available data. Future work includes expanding the evaluation to additional regions while maintaining a focus on Brazil, as well as investigating correlations between geolocation errors and geographic or network-related factors.

Acknowledgments. This work was supported by MPMG, CNPq, CAPES, and FAPEMIG.

Declaration of Generative AI use. Only for linguistic and grammatical refinement.

References

Abstract (2026). Abstract API. `https://www.abstractapi.com/api/ip-geolocation-api`. Accessed: 2026-05-03.

Apiip (2026). apiip. `https://apiip.net/`. Accessed: 2026-05-03.

BigData (2026). Big Data Cloud. `https://www.bigdatacloud.com`. Accessed: 2026-04-30.

Callejo, P., Gramaglia, M., Cuevas, R., and Cuevas, A. (2022). A deep dive into the accuracy of ip geolocation databases and its impact on online advertising. *IEEE Transactions on Mobile Computing*, 22(8):4359–4373.

FindIP (2026). FindIP. `https://findip.net/`. Accessed: 2026-05-03.

FreeIPAPI (2026). Free IP API. `https://freeipapi.com/`. Accessed: 2026-05-03.

Geoapify (2026). Geoapify. `https://www.geoapify.com`. Accessed: 2026-05-03.

Geolocated.io (2026). Geolocated.io. `https://app.geolocated.io/`. Accessed: 2026-05-03.

Info, I. (2026). Ip2 info. `https://ip2info.com/pt`. Accessed: 2026-05-03.

IP-API (2026). IP-API. `https://ip-api.com/`. Accessed: 2026-05-03.

IP2Location (2026). IP2Location. `https://www.ip2location.com/`. Accessed: 2026-05-03.

Ipapi (2026). ipapi. `https://ipapi.com/`. Accessed: 2026-05-03.

Ipdata (2026). ipdata. `https://ipdata.co/`. Accessed: 2026-05-03.

IPDetails (2026). IPDetails. `https://ipdetails.io/`. Accessed: 2026-05-03.

IPGeolocation (2026). IP Geolocation. `https://ipgeolocation.io`. Accessed: 2026-05-03.

IPstack (2026). IPstack. `https://ipstack.com/`. Accessed: 2026-05-03.

IPWho (2026). IPWho. `https://www.ipwho.org/`. Accessed: 2026-05-03.

Ipwhois.io (2026). ipwhois.io. `https://ipwhois.io/`. Accessed: 2026-05-03.

Izhikevich, K., Du, B., Rao, S., Ukani, A., and Izhikevich, L. (2024). Trust, but verify, operator-reported geolocation.

Leite, N., Campelo, C. E. C., and Silva, S. D. (2024). Leveraging geographic feature embeddings for enhanced location-based recommendation systems. In *Anais do XXXIX SBBD*, pages 354–366, Porto Alegre, RS, Brasil. SBC.

Nur, A. Y. (2023). Accuracy and coverage analysis of ip geolocation databases. In *2023 International Balkan Conference on Communications and Networking*, pages 1–6.

Silva, D. and Holanda, M. (2025). Uma arquitetura para big data geoespacial de dispositivos iot com níveis heterogêneos de mobilidade. In *Anais do XL Simpósio Brasileiro de Bancos de Dados*, pages 42–55, Porto Alegre, RS, Brasil. SBC.

Techniknews (2026). API Techniknews. `https://api.techniknews.net/`. Accessed: 2026-05-03.

TheIPAPI (2026). The ip api. `https://theipapi.com/`. Accessed: 2026-05-03.

Xu, W., Tao, Y., and Guan, X. (2018). Experimental comparison of free ip geolocation services. In *International Conference on Security with Intelligent Computing and Big-data Services*, pages 198–208. Springer.