B

LEVERAGING LABOR MARKET DATA AND NAME ANALYSIS TO MAP SCIENTIFIC DIASPORAS

Planning Department Research Division

CITY *of* **BOSTON**

Planning Department

The **RESEARCH DIVISION** is a division of the City of Boston's Planning Department, which Mayor Wu established in 2024 to restore planning as a core function of city government. The Division compiles and analyzes current, historical, and comparative data on the City's population and economy to produce quality research that informs policy.

AUTHOR: City of Boston Planning Department, Research Division.

DATE: January 2025.

INFORMATION: Please visit **bostonplans.org/research** for more information.

REQUESTS: Please contact us at **bostonplans.org/research/research-inquiries**.

RESEARCH DIVISION

DIRECTOR OF RESEARCH
Alvaro Lima

DEPUTY DIRECTOR
Christina Kim

SENIOR DEMOGRAPHER
Phil Granberry

SENIOR ECONOMISTS
Michael Endale
Fatima Aqeel

SENIOR RESEARCH ASSOCIATES
Kevin Kang
Jake Schneider

RESEARCH ASSOCIATE
Emily Korest

RESEARCH COORDINATOR
Cecilia Takacs

RESEARCH ASSISTANTS
Sam Klugherz
Grace Julien
Na Young Moon
Tyler Yandt

INTERN
Samara Shankar

TABLE OF CONTENTS

ABSTRACT

This study leverages labor market data, name recognition, and analysis to map and understand scientific diasporas. A case study of the Brazilian scientific diaspora in the Greater Boston region illustrates the methodology developed by the Research Division of the City of Boston Planning Department in partnership with Claro Analytics and Namsor.[1] The case study's findings present Brazilians' educational and professional characteristics in Boston's scientific sectors. The case study's methodology provides a scalable framework to analyze ethnic diasporas across the United States and globally.

The primary motivation for the study was to test a methodology designed to identify existing scientific diasporas from secondary data sets to provide a way for governments in sending and receiving countries to coordinate mutually leveraging diasporas' knowledge and skills. Demographic research on immigrant communities, in general, focuses its attention on immigrants living at or below the poverty line, as it should. This is an essential means of informing social policy decisions to make them more responsive to the needs of these immigrant communities. However, less attention is generally paid to higher-skilled immigrants who make up "scientific diasporas" and are the focus of our study. While they are not the beneficiaries of social programs designed to alleviate poverty and, therefore, not necessarily the subject of studies to enhance social policies, the immigrants who make up scientific diasporas can and should be considered important subjects of studies and policy design.

---

[1]   See **APPENDIX B** and **APPENDIX C** for a description of these organizations and capabilities.

01

INTRODUCTION

The term "scientific diaspora" refers to highly educated individuals who have emigrated from their home countries and contribute to research, innovation, and development in their host countries. The term diaspora was first used to refer to communities who were forced to migrate due to political, religious, or ethno-racial persecution while continuing to preserve their identity, culture, traditions, and a reference to an ancestral homeland despite their geographic dispersion.[2] In the 1980s, the term began to be used more broadly and later included highly qualified individuals, giving rise to the term "scientific diaspora."[3] The term "scientific diaspora" appeared in the 1990s almost at the same time as the term "brain circulation" is contrasted to the conception of "brain drain,"[4] which carries a negative connotation regarding the migration of highly educated professionals.[5] Migration is no longer considered a permanent loss for the country of origin, provided that an enabling environment is created and engagement strategies designed, going beyond their physical return to their home countries.[6]

Research hubs like Boston benefit significantly from these communities and could foster cross-border links to drive scientific and economic progress. If effectively engaged, these scientists can contribute to their countries of origin by creating knowledge exchanges to increase collaboration between scientific communities in sending and receiving countries.[7] Strategies of cooperation and integration between research centers can improve the exchange of knowledge and resources to advance science and innovation across countries. However, it is not enough to build knowledge bridges. It is necessary to construct domestic policies and structures of support to favor "brain circulation."

2   The term Diaspora is derived from the Greek word meaning "a scattering or sowing of seeds" and refers to the spread of a given ethnic group outside of their country of origin. The term diaspora was traditionally used to describe the Jewish population exiled from Judea in biblical times.

3   COHEN, 2008.

4   The term "brain drain" was first used at the end of 1950s by the British Royal Society to refer to the massive exit of scientists and technologists from the United Kingdom to the United States, considered as a permanent loss for the country.

5   Another term, not at all unimportant, is the "brain waste" which refers to highly-skilled immigrants who are either unemployed or working in jobs that do not make appropriate use of their knowledge and skills representing a significant social and economic cost for both the individuals affected and society as a whole.

6   Many highly qualified Brazilians want to return to Brazil but few can find favorable employment conditions, adequate means to develop their research, etc. (PIRES and GOMIDE, 2014.) The most dramatic cases are those Brazilians ex-bolsistas who live and work in the United States, are not able to return and find themselves trapped in perennial debt. Relatório da Conferência Livre de Boston - II COMIGRAR.

7   Such Exchanges and Networks have been essential for the development of the tech sector in Taiwan, Israel, India, and China, as well as for the creation of business hubs in Ireland and Scotland.

Globalization and innovations in information and communication technologies have created new possibilities for harnessing and leveraging human talent across international borders.[8] In a knowledge-based economy, firms and nation-states know that to gain global competitiveness, they must attract, recruit, and retain human talent wherever it is available. Cross-border knowledge exchanges have become essential to national systems of science, technology, and innovation.

The Brazilian government's policies for the internationalization of science were responsible for the insertion of thousands of Brazilian researchers abroad. While many of these scientists have returned to Brazil after their studies, many have established themselves in universities, research institutions, and the private sector abroad. This phenomenon has generated an intense debate about "brain drain." More recently, Brazilian researchers in this area have forced a paradigm shift by introducing the concept of "brain circulation," arguing that highly skilled Brazilian professionals can be perceived not as a permanent loss (brain drain) but as a potential resource if identified, organized, and supported by public, private and philanthropic organizations that seek to leverage their knowledge.

One of the main challenges in designing policies to support the exchange of knowledge (i.e., brain "circulation" as opposed to "drain") is the lack of available data on the Brazilian scientific diaspora. Most national censuses only provide a counting or estimates of the number of Brazilians with higher educational attainment living in the country or other geographic unit.[9] This report seeks to provide a methodology for identifying, mapping, and analyzing scientific diasporas across multiple geographic scales using the Brazilian scientific diaspora in the Greater Boston area as a case study. Boston, one of the cities that receives the most grant funding from the National Institutes of Health, is a global hub for scientific research, making for an excellent geography to test this methodology. The methodology described here does not preclude using alternative quantitative and qualitative methods to complement the results provided.

---

8   LIMA, 2010.

9   For context, **APPENDIX A** provides a brief profile of Brazilians in Massachusetts.

The following section details the methodology of this case study, including the primary data providers, study area definition, filters, and queries used to identify, map, and analyze the Brazilian scientific diaspora. The final section explores the results of this study in Greater Boston. It demonstrates how this case study can be scaled to other metropolitan areas and larger geographies, such as the United States or the world.[10]

---

10   For a preliminary example of these extensions, see Powerpoint Presentation for details.

02

CASE STUDY METHODOLOGY

DATA PROVIDERS

To identify scientific diasporas, it is essential to use tools that facilitate the identification of a diaspora or ethnic group living within a particular region, as well as analyze their educational and professional backgrounds. Two data providers contributed to this analysis: (1) Claro Analytics, a labor market intelligence platform that aggregates professional profile data from LinkedIn and job boards to provide real-time insights into employment trends, and (2) Namsor, which is a name analysis company that uses algorithms to infer diaspora membership, country of origin, and gender from personal names.

Claro collects and aggregates professional profile data across labor market platforms and allows advanced searches based on job titles, industries, degrees, skills, and location.[11]

Namsor has developed two name recognition models that are integrated directly within Claro. The diaspora model matches names to global ethnic groups based on linguistic and cultural patterns, and the country of origin model predicts a name's likely country of origin using database names from around the world.

This case study demonstrates how Claro's database can be combined with Namsor's name recognition models to create targeted queries to analyze the Brazilian scientific diaspora at multiple geographic levels, focusing on the Greater Boston labor market.

STUDY AREA DEFINITION

The study area chosen for the Case Study is the Greater Boston area to capture the entire Boston labor market. This region includes the City of Boston and surrounding municipalities within 25 miles of Boston, collectively contributing to Boston's expansive labor market. By focusing on Greater Boston, we can also capture professionals who might otherwise be underrepresented if only the City of Boston was used as the study area. This is because many LinkedIn users prefer

---

11   **APPENDIX C** provides more information about Claro's data collection and categorization methodology.

to list their location as their metropolitan areas rather than the precise cities or towns they live in.

FIGURE 1
Study Area Map
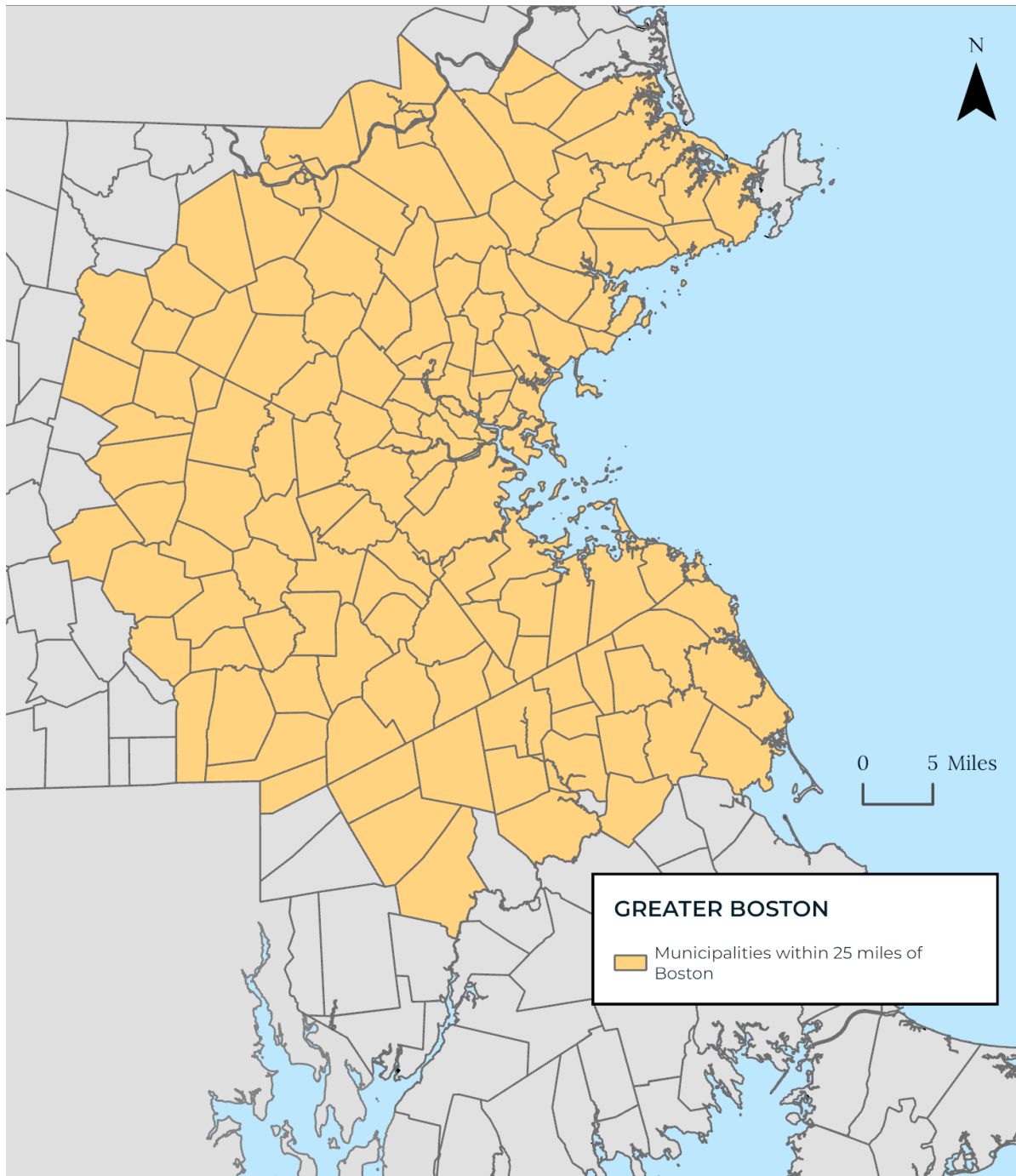
IDENTIFYING BRAZILIAN PROFILES

The study employs a robust filtering methodology in Claro Analytics to identify Brazilian profiles in Greater Boston accurately.[12] These filters are designed to maximize the inclusion of Brazilian professionals while minimizing misclassification from other Lusophone or non-Brazilian groups. The key filters used are described below.

NAME ANALYSIS FOR BRAZILIAN ORIGIN

Leveraging the integrated Namsor name analysis tool, the study identifies profiles with Portuguese names of Brazilian origin. Namsor's algorithm analyzes linguistic patterns and cultural markers in names to differentiate Brazilian names from those of other Lusophone groups. This step ensures that individuals with high-probability Brazilian origins are included in the analysis.

KEYWORD FILTERS FOR BRAZILIAN CITIES AND UNIVERSITIES

Additional keyword filters refine the search to capture profiles explicitly linked to Brazil. This step broadens the dataset to include profiles that might not be identified solely by name analysis. These filters look for:

• Mentions of major Brazilian cities (e.g., São Paulo, Rio de Janeiro, Belo Horizonte) in location fields.

• Academic histories include references to Brazilian educational institutions (e.g., Universidade de São Paulo, Universidade Federal de Minas Gerais).

EXCLUSION OF PROFILES FROM OTHER LUSOPHONE COUNTRIES

To reduce false positives, the methodology excludes profiles likely to belong to professionals from other Portuguese-speaking countries, such as Portugal, Cabo Verde, Mozambique, or Angola. By applying this exclusion criterion, the study ensures that the dataset accurately represents individuals connected to Brazil. Profiles are flagged for exclusion if:

• Names match patterns typical of non-Brazilian Lusophone origins.

• Location or education history references non-Brazilian Lusophone cities or institutions.

---

12   See Powerpoint Presentation for details.

IDENTIFYING "SCIENTIFIC" PROFILES

Specific filters are applied to identify individuals with advanced academic qualifications and professional engagement in scientific fields to refine the dataset further and focus on the Brazilian scientific diaspora. These filters ensure the study targets the most relevant subset of the Brazilian community contributing to research and innovation in Greater Boston. Two samples differing in scope are used to capture the "scientific" aspect of the Brazilian scientific diaspora:[13]

- **High-Constrained Sample**: To understand the most advanced professionals precisely, the study applies a strict filter to individuals explicitly listed as PhD or MD holders.

- **Wider-Net Sample**: To capture the broader scientific diaspora, a filter includes profiles associated with scientific industries, regardless of degree type, to account for professionals contributing to science-driven sectors.

HIGH-CONSTRAINED SAMPLE: DOCTORATE DEGREES (PHD, MD)

- **Educational Attainment**: The dataset includes only individuals with doctorate-level qualifications such as PhDs or MDs. This filter ensures the focus is on highly educated professionals likely to be engaged in advanced research, academia, or specialized industries.

- **Degree Fields**: Additional refinement considers degrees in fields typically associated with scientific research, including but not limited to life sciences, engineering, medicine, and technology.

WIDER-NET SAMPLE: EMPLOYMENT IN SCIENTIFIC INDUSTRIES

- **Industry Categories:** Claro provides industry classification data, allowing the study to include professionals working in ten scientific sectors:
    1. Hospitals & Health Care
    2. Biotechnology
    3. Pharmaceuticals
    4. Higher Education
    5. Medical Devices
    6. Environmental Services
    7. Civil Engineering
    8. Medical Practice
    9. Research
    10. Computer Software

13　　　The definition of "scientific" can be further restricted (e.g. to include only PhDs) or relaxed (to include , for example, College+) depending on the objective of the researcher.

CHALLENGES AND CONSIDERATIONS

While the methodology outlined for identifying and analyzing the Brazilian scientific diaspora is robust, several challenges and considerations must be addressed to ensure the accuracy and comprehensiveness of the study.

ACCURACY OF NAME ANALYSIS

Namsor's name recognition tool provides a powerful means to infer Brazilian origins. Namsor's diaspora model identifies Lusophone names with 85%-95% accuracy, including Brazilian, Portuguese, Luso-Asian, and Luso-African origins. The country of origin model differentiates Lusophone diasporic groups with ~75% accuracy, distinguishing between Portuguese, Brazilian, Angolan, Mozambican, and Cape Verdean origin. Distinguishing Brazilian names from other Lusophone groups can be challenging, especially for individuals with shared naming conventions.[14]

REPRESENTATION BIAS

The dataset's reliance on LinkedIn and other professional platforms introduces inherent biases. First, not all professionals, particularly those in less digitally connected industries, actively maintain LinkedIn profiles, potentially underrepresenting certain groups. Second, Claro Analytics only scrapes public LinkedIn profiles, which introduces bias. Finally, the accuracy of profile information (e.g., education and employment history) depends on the user's input, which may be incomplete or outdated.

REGIONAL FOCUS AND SCALABILITY

While the study focuses on Greater Boston, expanding the analysis to other regions or global scales requires additional considerations. In areas with lower internet penetration, professional data may be sparse or inconsistent, affecting the scalability of the methodology.

---

14   More information about the Namsor name origin algorithm is available in **APPENDIX C**.

03

CASE STUDY RESULTS

FOCUS ON GREATER BOSTON

PROFILE ANALYSIS

The study identified over 13,000 Brazilian professional profiles in the Greater Boston area using Claro. Of these:

1. More than 300 individuals (2.4% of all Brazilian diaspora profiles in Greater Boston) hold doctorate degrees (PhDs or MDs), representing a highly educated subset of the diaspora.

2. Over 1,200 individuals (8.9% of all Brazilian diaspora profiles in Greater Boston) are employed in scientific industries. The scientific industries used to identify this are described in the previous section.

TABLE 1
Brazilian Diaspora Profiles by Geographic Level

	Total Brazilian Diaspora Profiles	Doctorate (PhD, MD)	Doctorate (PhD, MD) %	Scientific Industry	Scientific Industry %
Greater Boston	13,862	319	2.3%	1,176	8.5%
United States	467,223	6,448	1.4%	21,963	4.7%
World *(excluding Brazil)*	1,770,994	19,810	1.1%	79,125	4.5%

Source: Claro Analytics, Namsor, City of Boston Planning Deparrtment Research Division.

EDUCATIONAL BACKGROUND

Brazilians in Greater Boston with advanced degrees and those in scientific sectors often receive their education from a mix of top Brazilian and local Boston institutions.

1. *Brazilian Institutions*. The University of São Paulo (USP) is the most represented institution in the list of alma maters, followed by other notable institutions such as the Universidade Estadual de Campinas (UNICAMP) and the Universidade Federal de Minas Gerais (UFMG).

2. *Boston-Based Institutions*. Many Brazilian professionals pursue further studies at Boston's renowned universities, such as Northeastern University, Boston University, and the Massachusetts Institute of Technology (MIT).

Top Schools Among Brazilians in Greater Boston with a PhD or MD

School	Count
Universidade De Sao Paulo	80
Universidade Estadual De Campinas	33
Massachusetts Institute Of Technology	29
Universidade Federal De Minas Gerais	27
Harvard Medical School	24
Universidade Federal Do Rio De Janeiro	21
Universidade De Brasilia	20
Harvard University	19
Boston University	16
Northeastern University	16

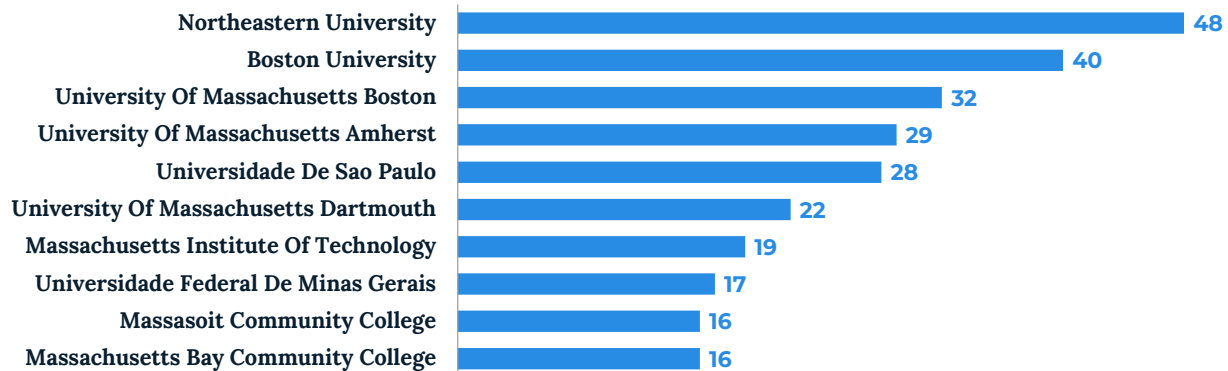*Source: Claro Analytics, Namsor, City of Boston Planning Deparrtment Research Division.*

Top Schools Among Brazilians in Greater Boston in a Scientific Industry

School	Count
Northeastern University	48
Boston University	40
University Of Massachusetts Boston	32
University Of Massachusetts Amherst	29
Universidade De Sao Paulo	28
University Of Massachusetts Dartmouth	22
Massachusetts Institute Of Technology	19
Universidade Federal De Minas Gerais	17
Massasoit Community College	16
Massachusetts Bay Community College	16

Source: Claro Analytics, Namsor, City of Boston Planning Deparrtment Research Division.

KEY EMPLOYERS

Several institutions in Greater Boston play a pivotal role in employing Brazilian professionals.

1. *Harvard University*: Employs the highest number of Brazilians with PhDs/MDs (27 professionals) and the second-highest number of Brazilians among those in scientific sectors, reflecting the university's strong connection to global talent.
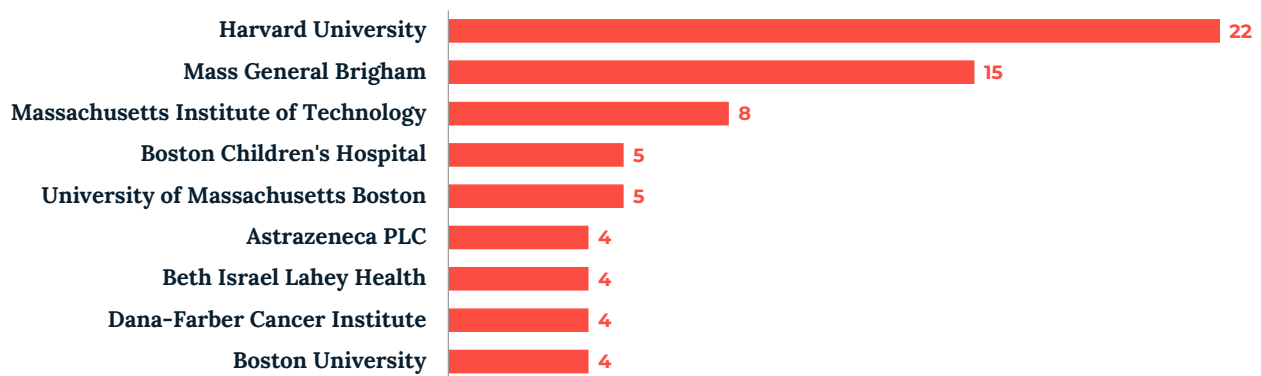
2. *Mass General Brigham*: A leading employer in the healthcare sector, hiring doctorate holders and scientific professionals across multiple disciplines. Mass General Brigham is the top employer among Brazilians in scientific industries, with 27 professionals.

3. *MIT and Boston University*: These premier universities offer significant opportunities in academia, research, and innovation, further cementing Greater Boston as a hub for the scientific diaspora.

4. *Other Notable Employers*: Institutions like Dana-Farber Cancer Institute, AstraZeneca, and Boston Children's Hospital also contribute to integrating Brazilian expertise into Greater Boston's labor market.
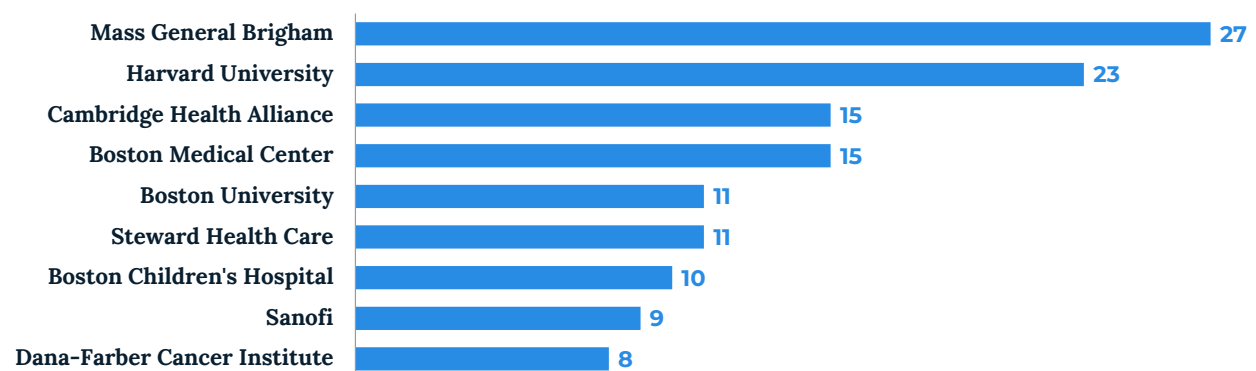
Top Employers Among Brazilians in Greater Boston with a PhD or MD

Employer	Count
Harvard University	22
Mass General Brigham	15
Massachusetts Institute of Technology	8
Boston Children's Hospital	5
University of Massachusetts Boston	5
Astrazeneca PLC	4
Beth Israel Lahey Health	4
Dana-Farber Cancer Institute	4
Boston University	4

Source: Claro Analytics, Namsor, City of Boston Planning Deparrtment Research Division.

Top Employers Among Brazilians in Greater Boston in a Scientific Industry

Employer	Count
Mass General Brigham	27
Harvard University	23
Cambridge Health Alliance	15
Boston Medical Center	15
Boston University	11
Steward Health Care	11
Boston Children's Hospital	10
Sanofi	9
Dana-Farber Cancer Institute	8

Source: Claro Analytics, Namsor, City of Boston Planning Deparrtment Research Division.

OCCUPATION TRENDS

The Brazilian scientific diaspora in Greater Boston holds a diverse range of occupations, reflecting their high levels of education and specialization in various scientific fields. The most common roles include:

1. *Researchers and Professors*

   • Teachers and Professors are the largest occupational group among Brazilians with doctorates in Greater Boston, with over 42 professionals in these roles. This is also the largest occupational group among Brazilians in scientific sectors, with 53 professionals.

   • Their contributions span academia and research institutions, driving knowledge creation and dissemination in engineering, biology, and social sciences.

2. *Biologists*

   • 23 professionals with advanced degrees work as biologists, representing a strong presence in life sciences.

   • Their roles often involve conducting molecular biology, genetics, and environmental science research, contributing to Boston's leadership in biotech and health research.

3. *Medical and Health Services Managers*

   • Many Brazilians with PhDs or MDs occupy leadership positions in the healthcare sector, coordinating and improving medical services at hospitals and clinics. 16 Brazilian professionals with PhDs or MDs work in these roles, and 53 Brazilian professionals in scientific industries work in these roles.

4. *Physicians and Nurses*

   • 8 professionals with PhDs/MDs work as physicians in Greater Boston, and 35 professionals in the scientific sector work as registered nurses. These providers enhance the region's medical expertise and support its renowned healthcare infrastructure and services.

5. *Natural Sciences Managers and Laboratory Technicians*

   • Many Brazilians contribute to scientific and technical operations, managing
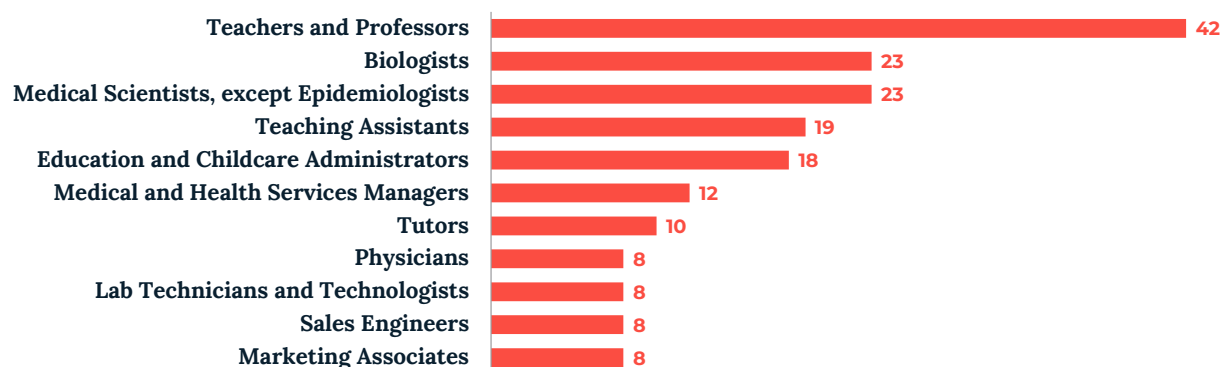
research projects, or supporting experiments in academic and industry labs.

6. *Sales Engineers and Technology Specialists*

- Some professionals with scientific backgrounds apply their expertise in the private sector, particularly in roles bridging engineering and business, such as sales engineers in tech and medical device companies.

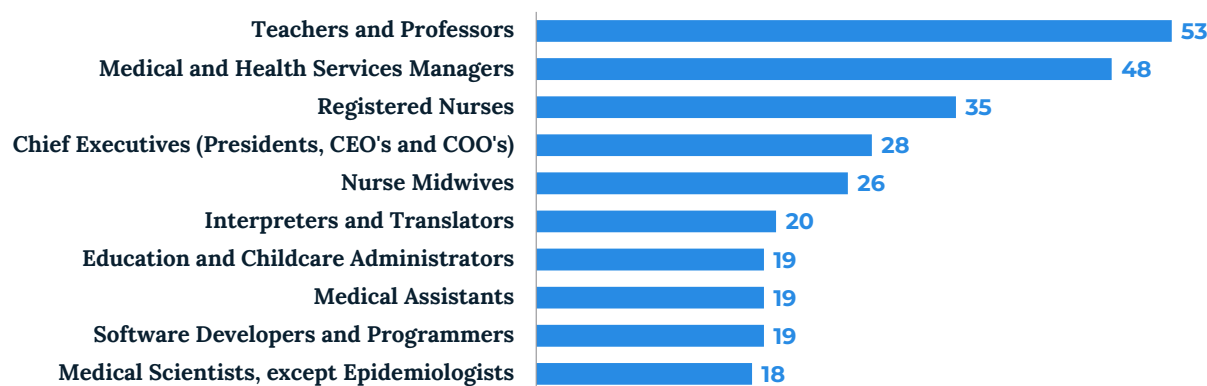### Top Occupations Among Brazilians in Greater Boston with a PhD or MD

Occupation	Value
Teachers and Professors	42
Biologists	23
Medical Scientists, except Epidemiologists	23
Teaching Assistants	19
Education and Childcare Administrators	18
Medical and Health Services Managers	12
Tutors	10
Physicians	8
Lab Technicians and Technologists	8
Sales Engineers	8
Marketing Associates	8

Source: Claro Analytics, Namsor, City of Boston Planning Deparrtment Research Division.

Top Occupations Among Brazilians in Greater Boston in a Scientific Industry

Occupation	Value
Teachers and Professors	53
Medical and Health Services Managers	48
Registered Nurses	35
Chief Executives (Presidents, CEO's and COO's)	28
Nurse Midwives	26
Interpreters and Translators	20
Education and Childcare Administrators	19
Medical Assistants	19
Software Developers and Programmers	19
Medical Scientists, except Epidemiologists	18

Source: Claro Analytics, Namsor, City of Boston Planning Deparrtment Research Division.

UNITED STATES & GLOBAL BRAZILIAN SCIENTIFIC DIASPORA[15]

UNITED STATES

The Brazilian scientific diaspora in the US mirrors patterns seen in Greater Boston, with 6,448 doctorate holders and nearly 22,000 professionals in scientific industries nationwide. Graduates of Brazil's premier institutions (USP, UNICAMP, UFMG) predominate, with most employed at leading research universities like University of California (74) and UNC (40). These professionals concentrate in research (560), higher education (503), and healthcare (262) sectors, primarily as professors (796) and researchers. Key metropolitan areas — New York, Miami, Los Angeles, and San Francisco — serve as critical hubs due to their research infrastructure and established Brazilian communities.

GLOBAL PRESENCE

Internationally, nearly 20,000 Brazilian PhDs and 79,000 scientific professionals contribute to innovation worldwide. While maintaining strong educational ties to Brazilian institutions (USP: 2,059), they have established significant presence in multinational corporations like Johnson & Johnson (200) and Sanofi (154). The global diaspora works predominantly in healthcare (16,360) and higher education (14,541) across Europe, Asia, and Oceania, applying their expertise to address complex scientific and technological challenges through international collaboration networks.

---

15   For a preliminary example of these extensions, Powerpoint Presentation for details.

04

CONCLUSION

This report demonstrates a robust methodology for identifying, mapping, and analyzing scientific diasporas using the Brazilian scientific diaspora in Greater Boston as a case study. By leveraging labor market intelligence from Claro Analytics and name analysis powered by Namsor, the study highlights Brazilians' educational and professional contributions to Boston's scientific sectors. The findings emphasize the strategic importance of these diasporas in fostering innovation, knowledge exchange, and international collaboration.

The study reveals that Greater Boston is a critical hub for Brazilian professionals. Over 13,000 profiles have been identified, including more than 300 with advanced degrees (PhDs and MDs). These individuals are concentrated in key industries such as higher education, health care, biotechnology, and pharmaceuticals. Leading employers such as Harvard University and Mass General Brigham illustrate this community's depth of integration and impact on Boston's economy.

The case study highlights the potential for this methodology to scale beyond Greater Boston, providing a replicable framework for analyzing diasporic communities across various geographic scales. Policymakers, academic institutions, and employers can use this methodology to understand better and engage scientific diasporas, fostering cross-border collaboration and leveraging the unique expertise of these communities. In doing so, host regions and countries of origin can benefit from the phenomenon of "brain circulation," turning a potential "brain drain" into an opportunity for global scientific and economic progress.

This case study underscores the potential of innovative data-driven tools to map and harness the capabilities of diaspora communities. It provides a template for future research and policymaking.

BIBLIOGRAPHY

ARAÚJO, Galber Rodrigues; CALDEIRA B., S. LIMA, Melissa. Mapeamento da Diáspora Brasileira de Ciência, Tecnologia e Inovação na Áustria, Eslováquia e Eslovênia. Embaixada do Brasil em Viena, 2021.

BALBACHEVSKY, Elizabeth; COUTO SILVA, Eduardo. A Diáspora Científica Brasileira: Perspectivas para sua Articulação em Favor da Ciência Brasileira. Parcerias Estratégicas, v. 16, n. 33, p. 163-176, 2011.

BEZERRA, Fernanda Mendes; SILVEIRA NETO, R. M. Existe 'Fuga de Cérebros' no Brasil? Evidências a Partir dos Censos Demográficos de 1991 e 2000. Economia, v.9, n.3, p. 435-456, 2008.

BRINKERHOFF, J. M. Creating an Enabling Environment for Diasporas' Participation in Homeland Development. International Migration. Geneva, v. 50, n.1, p. 75-95.

CARNEIRO, A. M.; GIMENEZ, A. M. N.; GRANJA, C. D.; BALBACHEVSKY, E.; CONSONI, F.; ANDRETTA, V. F. Diáspora Brasileira de Ciência, Tecnologia e Inovação: Panorama, Iniciativas Auto-organizadas e Políticas de Engajamento. Ideias, Campinas, SP, v.11, 1-29, e 020010, 2020.

COHEN, R. Global Diasporas: An Introduction. 2. Ed. Inglaterra: Routledge, 2008.

COUTO E SILVA, E.; SOUZA, G. H. A Primeira Rede da Diáspora Científica Brasileira - Brasileiros com Alta Qualificação Contribuindo para o Desenvolvimento Econômico, Científico e Tecnológico do Brasil. In: Conferências Sobre as Comunidades Brasileiras no Exterior, 1. 2008, Rio de Janeiro, RJ. Anais. Brasília, DF: Fundação Alexandre de Gusmão, 2008.

FANGMENG, T. Brain Circulation, Diaspora and Scientific Progress: A Study of the International Migration of Chinese Scientists, 1998-2006. Asian and Pacific Migration Journal, v. 25, n. 3, p. 296-319, 2016.

GAMLEN, A. Diaspora Institutions, and Diaspora Governance. International Migration Review. V. 48, p. 180-217, 2014.

GOERGEN, Pedro. A Internacionalização dos Programas de Pós-graduação. Revista Espaço Pedagógico, v. 19, n. 2, 2012.

KUZNETSOV, Y. Diaspora Networks and the International Migration of Skills: How Countries Can Draw on their Talent Abroad. WBI Development Studies. Washington, DC: World Bank. 2006.

LIMA, Alvaro. Living Here and There: How Immigrants are Creating Transnational Social Spaces that Transform Communities and Nations, Boston, Dec. 2010.

MEYER, J. B. Network Approach Versus Brain Drain: Lessons from the Diaspora. International Migration, v. 39, n.5, p. 91-110, 2001.

MEYER, J.; BROWN, M. Scientific Diasporas: A New Approach to the Brain Drain. Management of Social Transformations, n. 41, p. 1-21, 2003.
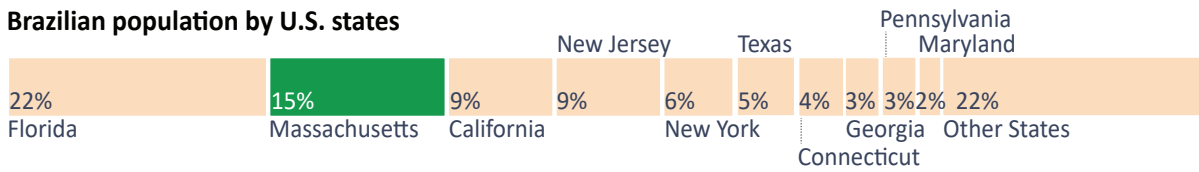
PIRES, R. R. C.; GOMIDE, A. A. Análise Comparativa: Arranjos de Implementação e Resultados de Políticas. In: GOMIDE, A. A.; PIRES, R. R. (ed.) Capacidades Estatais e Democracia: Arranjos Institucionais de Políticas Públicas. Brasília, DF: IPEA, p. 351-349, 2014.

SÉGUIN, B. et al. Scientific Diasporas as an Option for Brain Drain: Re-circulating Knowledge for Development. International Journal of Biotechnology, Switzerland, v. 8, n. 12, p. 78-90, 2006.
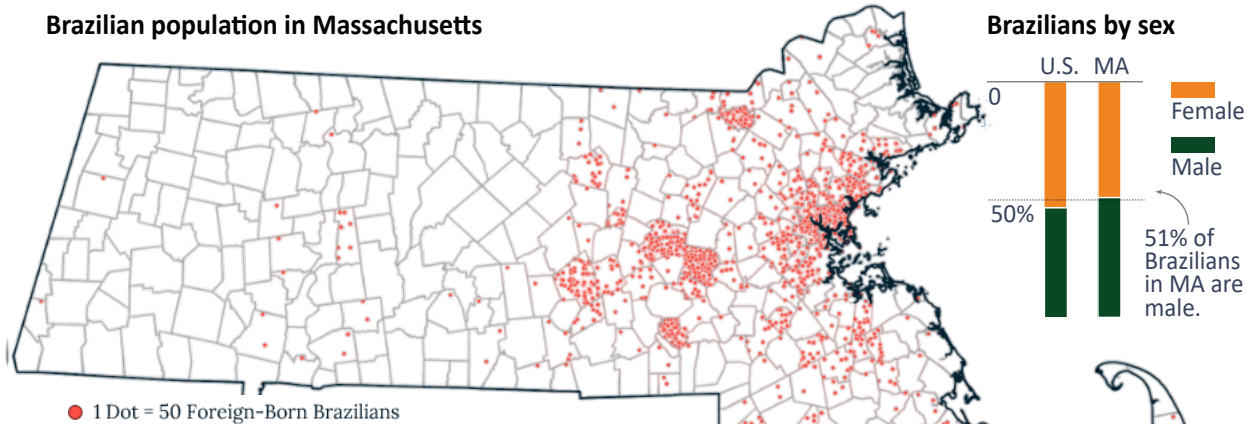
TEIXEIRA, Júlia Rocha. Diáspora Brasileira de Ciência, Tecnologia e Inovação: Perfil dos Pesquisadores no Exterior, Revista Discente Planície Científica, v. 6, n. 1, jan./jun. 2024.

PROFILE OF BRAZILIANS IN MASSACHUSETTS[16]

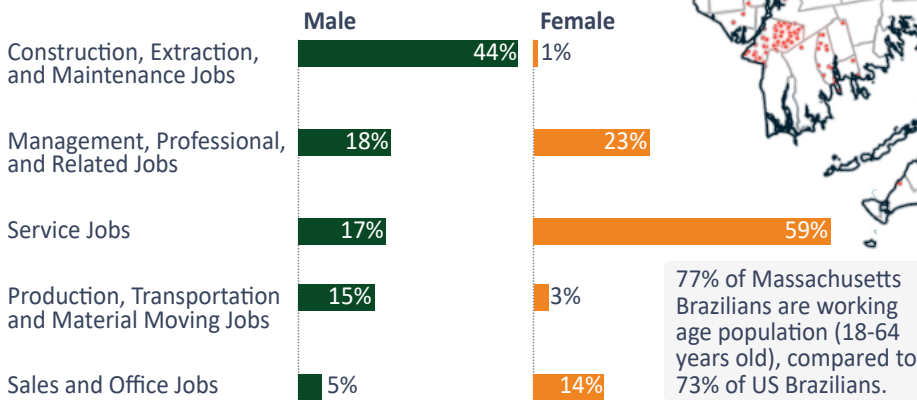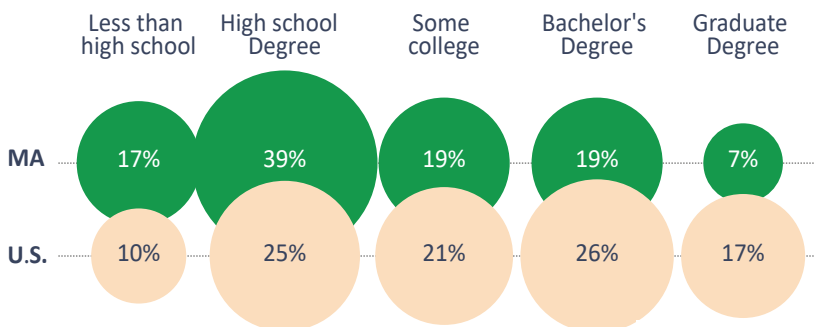## Brazilian population by U.S. states

Florida	Massachusetts	California	New Jersey	New York	Texas	Connecticut	Georgia	Pennsylvania	Maryland	Other States
22%	15%	9%	9%	6%	5%	4%	3%	3%	2%	22%

Brazilian population in Massachusetts

● 1 Dot = 50 Foreign-Born Brazilians

Brazilians by sex

U.S. MA

Female
Male

51% of Brazilians in MA are male.

Top Brazilian occupations in Massachusetts

	Male	Female
Construction, Extraction, and Maintenance Jobs	44%	1%
Management, Professional, and Related Jobs	18%	23%
Service Jobs	17%	59%
Production, Transportation and Material Moving Jobs	15%	3%
Sales and Office Jobs	5%	14%

77% of Massachusetts Brazilians are working age population (18-64 years old), compared to 73% of US Brazilians.

Brazilian population over 25 years old by educational attainment

	Less than high school	High school Degree	Some college	Bachelor's Degree	Graduate Degree
MA	17%	39%	19%	19%	7%
U.S.	10%	25%	21%	26%	17%

---

16   U.S. Census Bureau, American Community Survey (ACS, 2021).

APPENDIX B:
CLARO ANALYTICS METHODOLOGY

Claro Analytics is a labor market company whose mission is to help organizations make more innovative, strategic talent decisions through actionable labor market data. Claro has partnered with data-driven organizations for over a decade to gain a competitive advantage in their respective industries and geographies. This unwavering dedication to client success has solidified its position as the premier labor market intelligence platform.

Claro collects data about the workforce 24/7, 365. Updates for people data ("Talent Supply Insights") are pushed to the leading platform for users to view in real-time as Claro's proprietary technology detects updates. In addition, the Claro team performs data deployments approximately once every week for job posting data ("Talent Demand Insights"), spending the interim period making every reasonable effort to cleanse the data of duplicates and expired postings.

The data is entirely sourced from public sites and public profiles, so any profiles tagged as "Private" — or otherwise hidden from public view — will not be present within Claro's dataset. All profiles are tagged with a timestamp of the last time they were updated, providing clarity into a particular dataset's vintage. Profiles are refreshed on a priority queue system, where the profiles for regions most in demand by Claro's current clients are refreshed before those in more niche regions to ensure active users see the mostrecent data.

Claro currently has over 620 million profiles from across the globe. Each profile has its attributes parsed into different searchable categories such as skills, current employer, degree, location, etc. A few categories — education, company, occupation, and industry — benefit from extensive data science efforts to standardize the variations naturally present in profiles, which a human interpreter might understand to be the same. For example, the Occupation field lists a person's current occupation based on their job title and experience. These data points are fed through a machine learning algorithm trained to match key items to the Occupation from the SOC,

corresponding to the individual's occupation. The industry field follows a similar process; however, rather than analyzing job titles and experience, the current employer's name determines the industry. This allows users to resolve a common conundrum of "a data scientist who works for General Motors" or "a marketing manager who works for Microsoft." The occupation of the first would be 'Data Scientists', and the industry would be 'Automotive,' while the second occupation would be 'Marketing Managers' in the industry 'Computer Software.'

APPENDIX C:
NAMSOR MODEL

Namsor is a Big Data company that developed specialized data mining software that recognizes personal names' linguistic or cultural origin in any alphabet or language to accurately classify personal names by gender, country of origin, or ethnicity. Namsor's mission is to make sense of Big Data and help to understand international flows of money, ideas and people.

Namsor uses a natural language processing (NLP) model, embedding a representation of a word as real-valued vectors that encode the word's meaning so that the words closer in the vector space are expected to be similar. Names closer to the vector space belong to the same gender/geographic/linguistic/racial/ethnic-cultural group.

Namsor name analysis uses algorithms to infer diaspora, country of origin, and gender from personal names:

- *Diaspora*: The algorithm matches names to global ethnic groups based on linguistic and cultural patterns.
- *Country of Origin*: It predicts the likely country of origin for a name using databases of names from around the world.

These two models are integrated with Claro's leading platform.

NAMSOR ALGORITHM ACCURACY AND LIMITATIONS

The accuracy of the Namsor algorithm varies depending on the ethnic group and country of origin. For ethnic groups with distinct names, such as Russians, Nigerians, Turkish, Japanese, and Polish, the algorithm performs at a high level of accuracy. The following graph visualizes a prototype of Namsor V3 onto a 2-dimensional vector space differentiating names from these five ethnic groups.

For names belonging to the same linguistic group, such as French, Portuguese, Italian, and Spanish, as well as groups with significant cross-migration and cultural exchanges, the algorithms perform at different levels of accuracy. For example, Portuguese from Portugal and Brazilian Portuguese are hard to differentiate. The current model identifies Lusophone names (diaspora) with 85%-95% accuracy, including Brazilian, Portuguese, Luso-Asian, and Luso-African origins, and differentiates Lusophone diaspora groups (country of origin) with ~75% accuracy (e.g., Portugal, Brazil, Angola, Mozambique, Cape Verde).