

IMPLEMENTING CLIMATE BUDGETING IN BUENOS AIRES

Leveraging AI Tools For Success

MAY 2026


COLUMBIA
SIPA
School of International
and Public Affairs

By

Raghad Alblwai, Alexander Bitterman, Richard
Carrión, Devan Gallagher, Ian Sun, Miranda Zanoni



Table of Contents

1	Executive Summary	↘
2	Introduction	↘
3	Problem Statement	↘
4	Context	↘
5	Stakeholder Analysis	↘
6	Data Analysis & Methodology	↘
7	AI Tool Development	↘
8	Results	↘
9	Challenges	↘
10	Recommendations & Roadmap	↘

Meet Our Team



Daniel Adam Zarrilli

Faculty Advisor



Raghad Alblwai

Travel Liaison



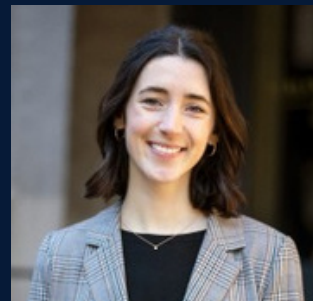
Alexander Bitterman

Project Manager



Richard Carrion

Fieldwork Coordinator



Devan Gallagher

Written Report & Presentation Editor



Ian Sun

SIPA Liaison



Miranda Zanoni

Client Liaison

Executive Summary

Climate Budgeting in Buenos Aires

LEVERAGING AI TOOLS FOR SUCCESS

This report presents our Capstone team's work for the City of Buenos Aires (CABA) to advance their implementation of climate budgeting. Our work builds on a prior Capstone engagement and represents the next phase of a multi-year effort to embed climate objectives into the city's budget governance structure.

Climate budgeting is a governance tool that helps cities map climate goals to budget decisions by tracking and reporting public expenditures that support mitigation and adaptation. For CABA, this is especially important because the city has ambitious climate commitments under its Climate Action Plan 2050 and, as a member of C40, is expected to embed climate budgeting into its regular budgeting processes by 2030. However, the city still lacks an institutionalized system for tracking how its budget allocates resources toward those goals. Our engagement focused on three workstreams:

- A comparative analysis of how peer cities approach climate budgeting
- A preliminary keyword-based tagging exercise conducted in R, a statistical programming language, against the city's FY2025 budget
- The development of an AI-assisted classification tool that applies proportional weighting to produce a defensible, reproducible climate share estimate

Our R keyword exercise was conducted across two scenarios (first 103 and then 158 keywords) and identified approximately 5.82% of the approved budget as climate-relevant. And it also revealed core limitations of any purely manual tagging approach: no proportionality logic, high false-positive rates for generic terms, and no scalable path to annual use.

Fieldwork in CABA including meetings with the Finance, Environment, and Innovation teams, confirmed the direction of AI-assisted automation, and enlightened us to richer program-level data in the Presupuestos Jurisdiccionales (Jurisdictional Budgets).

The AI tool we developed applies the OECD Rio Marker methodology, an internationally recognized framework for classifying the degree to which public expenditures support climate objectives to CABA's executed 2025 budget, classifying 45,444 budget lines across 451 unique programs and assigns proportional climate weights of 100% (Direct Action) or 40% (Enabling Co-benefit) or 0% (No Climate Action). The production methodology went through four iterative versions to eliminate sources of interpretive variance. The final result, 6.46% of total executed spending, or approximately ARS 930 billion, is a conservative lower bound designed to withstand audit, not a complete measure of the City's climate-related expenditure; the actual level of climate-related spending is likely higher than this estimate suggests. It is the first structured, reproducible climate share estimate CABA has produced for its executed budget.

Primary deliverables include a six-sheet Excel workbook with KPI summary, tagged budget lines, and a 327-line manual review queue, interactive HTML dashboard with editable weights, and a two-prompt methodology package that enables the Environment team to rerun the pipeline annually without developer support. Near-term recommendations center on completing the manual review queue, compiling an agency mandate list to support a new 10% administrative tier, and building the program-to-mandate lookup table required to reintroduce a 70% weighting tier. Long-term, migrating to a Python/API pipeline would achieve full reproducibility and enable the system to scale to annual use as a genuine governance tool rather than a retrospective reporting exercise.

Note: The figures presented in this report are estimates based on publicly available budget data available at the time of analysis and do not constitute official CABA figures. For the most accurate and up-to-date analysis of the City's climate budget, the relevant reference sources are BA Data, the Dirección General de Estadísticas, and the Dirección General de Presupuesto.

Introduction

Climate change poses an urgent and growing threat to cities worldwide. Urban areas account for roughly 70% of global greenhouse gas emissions and are simultaneously among the most vulnerable to the physical impacts of a changing climate. In this context, municipal governments play a critical role in controlling infrastructure investment decisions, land use, transport systems, and public services that collectively determine a city's emissions trajectory and its resilience to climate shocks.

Buenos Aires, the Autonomous City of Buenos Aires (CABA), has long recognized this responsibility. The city has been engaged in systematic climate action since at least 2008, producing greenhouse gas inventories, adopting a Climate Action Plan, and participating in the C40 Cities Climate Leadership Group. Its Climate Action Plan 2050, released in 2019, sets ambitious targets including a 30% reduction in emissions by 2030 and long-term goals around public transport use, vehicle electrification, and air quality.

Despite this track record, CABA currently lacks an institutionalized system for tracking how its municipal budget allocates resources toward climate goals. Climate spending is not systematically identified, tagged, or reported. This gap makes it difficult to assess whether the city's spending is aligned with its stated commitments, to demonstrate climate leadership to external partners and financiers, or to meet the emerging requirements of international networks like C40, which now requires member cities to institutionalize climate budgeting by 2030.

This report presents the work of a SIPA Capstone team engaged by CABA's Undersecretariat of Foreign Affairs to support the city's transition from climate budgeting recommendations to actual implementation. Our work builds directly on a prior Capstone engagement and represents the next phase of a multi-year effort to embed climate objectives into the city's budget governance structure.

Goals

Develop an implementation framework to embed climate budgeting into the municipal budgeting process.

Conduct a preliminary budget tagging exercise and demonstrate what percentage of the budget is climate-related.

Conduct a preliminary budget tagging exercise and demonstrate what percentage of the budget is climate-related.

Develop actionable recommendations for evolving the current budget tagging exercise into a sustainable, long-term climate budget model.

Deliverables

These deliverables are framed within the C40 Climate Budget Framework, which defines climate budgeting as the process of integrating climate goals from a city's Climate Action Plan into its existing budget process, assigning implementation responsibility across departments, tracking spending against climate targets, and covering both operating and capital budgets. Climate budgeting, in this framework, shifts from planning to implementation, creates accountability and transparency, embeds equity alongside mitigation and adaptation, and helps cities attract external climate finance.

1	A six-sheet Excel workbook containing a KPI summary, the full set of 1,232 climate tagged budget lines, a 327 line manual review queue with blank columns for reviewer decisions, the complete non-climate rows for audit completeness, and a Spanish-language methodology reference tab.
2	An interactive HTML dashboard that runs in any browser without backend infrastructure, displaying climate share KPIs and breakdowns by ministry and thematic axis, with editable weights that update the reported climate share in real time.
3	The two-prompt methodology package (Prompt 1 for tagging and Prompt 2 for output generation) which allows the environment team to replicate the full pipeline for future budget years by uploading a new file to a Claude Project with no developer involvement required.
4	A step-by-step video guide to show how to run the AI code and alter views on the dashboard.
5	A one-pager in English and Spanish for internal use, summarizing the project.

Problem Statement

The City of Buenos Aires wants a clear, operational climate budget tagging system that embeds climate goals into the city's budgeting process and governance structure. As a member of the C40 network, CABA must meet its commitments and integrate climate budgeting into municipal decision-making on policies, actions, and budget. Meeting C40's climate budgeting requirements is a primary driver for this work, and it is the reason CABA's International Relations team is centrally involved in ensuring the city's budget framework aligns with C40's standards. While the city has momentum for climate action, it faces real constraints: limited funding, staffing, data infrastructure, and specialized analytical capacity constrain its ability to implement a full climate budgeting framework. Given these constraints, the CABA seeks to advance from recommendations to implementation and expects an automated tool that will help the team tag Fiscal Year 2025's budget with the appropriate climate identifications. This tool will improve data classification and visualization, integrate financial and environmental budget data, and reduce the manual labor currently required. It will also help their team quantify climate spending across the city's ministries and operationalize the climate tagging process for subsequent years.

KEY CHALLENGES

- The city plans to shift from a retroactive (ex-post) classification exercise to embedding climate tagging proactively into the budget formulation process.
- Meeting C40's climate budgeting requirements by 2030 creates a defined timeline and external accountability mechanism.
- Institutional capacity constraints, including limited data availability at the project level, and a lack of standardized climate definitions that all ministries can apply consistently.
- Coordination between the Finance and Environment teams remains an important area to strengthen, given differences in incentive structures, workflows, and definitions of climate-related spending.
- The budget formulation period (roughly April through August) is the critical window in which climate criteria would need to be embedded to influence actual resource allocation decisions. Currently, no climate definitions or criteria are embedded in that process.
- The initial keyword-based tagging exercise was an Excel-based exercise run by the Environment team and only used keywords to match budget lines that had climate relevance.
 - This exercise also revealed the absence of any proportionality logic. A budget line tagged as climate-relevant is currently treated as 100% climate-related, even when climate may be only a secondary objective.
- In the final tagging exercise, despite obtaining more data, the program definitions and details were in separate PDF files, which made review challenging for accurate tagging.

Context

POLITICAL LANDSCAPE

CABA operates as a quasi-federal district with its own constitution and significant decision-making authority, making it politically distinct from the surrounding Province of Buenos Aires. The city is led by Mayor Jorge Macri of the Propuesta Republicana (PRO) coalition, who has held office since 2023, and governs alongside a unicameral 60-seat legislature where no single coalition holds a decisive majority. Alliance shifts frequently, and federal politics, currently controlled by Javier Milei's La Libertad Avanza, maintain pressure on city-level decision-making.

Party	Ideology	2023 Seats Added	2025 Seats Added	Total 2026
Frente de Izquierda (FIT)	Socialist	1	1	2
Es Ahora BA	Center left	-	10	10
Unión por la Patria (UP)	Center left	10	-	10
BA Primero	Center right	-	5	5
Juntos por el Cambio (JxC)	Center right	15	-	15
Volvamos BA	Center right	-	3	3
La Libertad Avanza (LLA)	Far right	4	11	15
Total		30	30	60

Institutionally, CABA is also characterized by other degrees of decentralization. The city is divided into comunas and each is governed by locally elected committees of seven representatives. This structure encourages that governance occurs closer to neighborhoods, increasing local participation but also adding another layer of political negotiation. Legislative proposals must pass through the city legislature before being reviewed by the mayor, reinforcing a system of checks and balances but also slowing decision-making in a context where consensus is often difficult to achieve. Additionally, CABA prides itself on transparency and accountability, with a wide majority of government documents and records are publicly accessible online.

This fragmented political environment has direct implications for climate budgeting. Cross-ministry initiatives require deliberate political insulation and sustained coalition-building to survive changes in administration priorities. Public opinion further complicates the political landscape. Residents of CABA are particularly focused on issues such as public safety, economic instability, and corruption, not climate change. Economic concerns, including high, though declining, inflation remains a priority, meaning climate budgeting does not have a strong public constituency. Therefore, it is important that the government continue to follow external accountability mechanisms, like C40's 2030 deadline, in order to maintain institutional momentum.

ECONOMIC OUTLOOK

Argentina remains in a period of macroeconomic adjustment under President Javier Milei's stabilization program. Inflation has fallen significantly from the very high levels seen in 2024, with the OECD projecting it will reach 37% in 2025 and 15% in 2026, and the national fiscal position has improved through strict spending controls. However, limited borrowing options and continued uncertainty constrain new public investment at both the national and city levels.

Within this context, CABA is in a comparatively strong economic position. The city has maintained fiscal surpluses, and S&P Global Ratings affirmed the city's B- rating with a stable outlook in July 2025.

This makes climate budgeting particularly compatible with the city's financial realities. The primary value of the system is not to necessarily create new climate spending immediately, but to make existing climate-relevant expenditures visible, trackable, and defensible to internal government partners and external bodies. Over time, as economic conditions stabilize and the city's planning improves, climate budgeting could also support more deliberate prioritization across ministries, helping CABA decide which mitigation and adaptation investments should receive greater protection, coordination, and long term support. As economic conditions stabilize and planning horizons grow, a functioning climate budget system will also better position the city to attract international climate finance.



Fieldwork Image:



TECHNOLOGICAL CAPABILITIES

INSTITUTIONAL AND DATA REQUIREMENTS FOR CLIMATE BUDGETING

The implementation of climate budgeting in CABA depends not only on political support and fiscal conditions, but also on the city's ability to identify, classify, and report climate-related spending with consistency. In practice, technological capacity means having advanced software, but also incorporating clear tagging rules, structured budget data, standardized reporting, and maintaining staff capacity to apply the system across ministries. The World Resources Institute notes that climate budget tagging has become an increasingly common tool for tracking climate-related public expenditure, but that countries use very different systems depending on their institutional and technical capacity.

EXISTING FOUNDATIONS AND CURRENT LIMITATIONS IN CABA

CABA already has experience with thematic budget labeling, including gender budgeting, and has explored several methods for climate labeling, such as binary tagging, weighting, and project sheets. Our work builds on these earlier efforts and shows that climate-relevant budget lines can be identified and processed in a structured way using the city's existing budget data. At the same time, both last year's Capstone report and our analysis confirm that important limitations remain. These include weak data granularity, unresolved questions around proportional weighting, limited staffing and analytical capacity, and the absence of a fully standardized methodology that can be applied consistently across ministries. Data availability has been a persistent and significant constraint throughout this work. Our initial analysis relied on keyword matching alone and while, over the course of the project, we gained access to additional data including project descriptions, we are still dependent on publicly available data.



Fieldwork Image: The Obelisk

IMPLICATIONS FOR THE NEXT STAGE

For CABA, the next stage is no longer simply about recognizing gaps in technological capacity but also about deciding how to build a practical and institutionally feasible climate budgeting system from the tools and knowledge now in place. Our analysis shows that it is possible to build a working climate budget tagging process that is explainable, fairly reproducible, and usable for review. However, it also indicates that the city must prioritize a simpler and more reliable method that can be clearly explained and defended, rather than an advanced form of automation that may be less consistent. For CABA, this suggests that the immediate priority should be to strengthen the basic architecture of climate budgeting including clearer definitions, better project documentation, standardized ministry reporting, and a shared validation process between the Finance and Environment teams. CABA's main challenge is no longer whether climate budget tagging is possible, but how to institutionalize it in a way that is consistent, scalable, and compatible with the city's regular budgeting process.

CABA CLIMATE GOALS & CHALLENGES

CABA enacted its Climate Change Law in 2011, mandating regular updates to its climate plan. In 2019, the city released its Climate Action Plan 2050, tied to both the Sustainable Development Goals and CABA's C40 commitments, and the city is preparing to release its fourth iteration of the plan this year. The plan organizes climate action into four scopes: a prepared city, an inclusive city, an innovative and low-carbon city, and a city that prioritizes proximity. In 2020, the city added a fourth pillar focused on urban tree planting, aimed at improving carbon sequestration, rainfall capture, and urban heat island reduction, particularly important given that only about 10 percent of CABA's land area is open or green space.

CLIMATE ACTION PLAN 2050 GOALS



30% reduction in greenhouse gas emissions by 2030



100% zero-emission public buses.



80% of passengers use public or non-motorized transport



100% compliance with WHO air quality standards



30% of residences use photovoltaic energy.

KEY IMPLEMENTATION CHALLENGES

Despite this ambition, CABA faces structural barriers to climate budget implementation that directly shape the constraints our recommendations must navigate. Thus far, the city has identified spending that benefits mitigation or adaptation, but is not currently equipped to flag spending that undermines climate goals. The positive-only framing was a deliberate choice to encourage ministry participation, and aims to position climate tagging as a recognition and funding opportunity rather than a flag for where emissions are too high.

The city relies on qualitative classification criteria rather than quantitative emissions metrics, because CABA does not currently have the staffing capacity for project-level emissions modeling. While appropriate for a first-generation system, this creates a risk of inconsistency across ministries over time and makes it harder to link tagged spending to demonstrable climate outcomes.

Climate budgeting is also not currently required by city legislation, meaning progress depends more on political commitment, administrative coordination, and external accountability than on an internal binding legal mandate.

This means cross-ministry coordination remains somewhat disjointed. The Finance and Environment teams have different primary mandates, reporting structures, and workflows. Embedding climate criteria into the budget formulation process, rather than applying them retroactively requires sustained engagement and clear protocols. Urbanization pressures and climate vulnerability are concentrated in the city's informal settlements (villas), which are disproportionately located in flood-prone areas and have lower access to climate-resilient infrastructure. Any climate budgeting framework should be sensitive to these equity dimensions.

CLIMATE BUDGET PROGRESS TO DATE

When we began our engagement, climate budget implementation in CABA was effectively stalled at the formulation stage. The Environment team had attempted an initial keyword-based review of budget lines and a draft thematic taxonomy, but the process was entirely manual, difficult to explain methodologically, and not built to repeat. There was no clear proportionality logic, so every tagged budget line was treated as 100% climate-relevant regardless of how central climate actually was to the program. The data available at the time lacked project-level descriptions, which made it impossible to distinguish climate-relevant subcomponents within multi-purpose programs. And with tens of thousands of budget lines and limited staff capacity, there was no realistic path to scaling the exercise annually through the budget process. Our goal was to take that foundation and build a methodology and tool that could produce a credible, reproducible climate share figure that the city could defend publicly and could repeat in future years.

MUNICIPAL BUDGET

Understanding where climate tagging fits in CABA's budget cycle was a critical context for our recommendations. The budget process begins every year with individual ministries submitting spending requests to the Ministry of Finance, which sets overall ceilings and coordinates inter-ministry review. The Finance Ministry consolidates these into a full budget proposal, which the Mayor submits to the City Legislature in late September. After public hearings and legislative review, the budget is voted on and takes effect January 1. The Formulation period (around May through August) is the critical window: this is when climate criteria would need to be embedded to influence actual resource allocation decisions, rather than being applied retroactively after spending is already locked in.

Throughout the year, the Finance Ministry allows adjustments at each of the four quarterly reviews, and conducts an audit of the prior year's execution with current-year spending.



Stakeholder Analysis

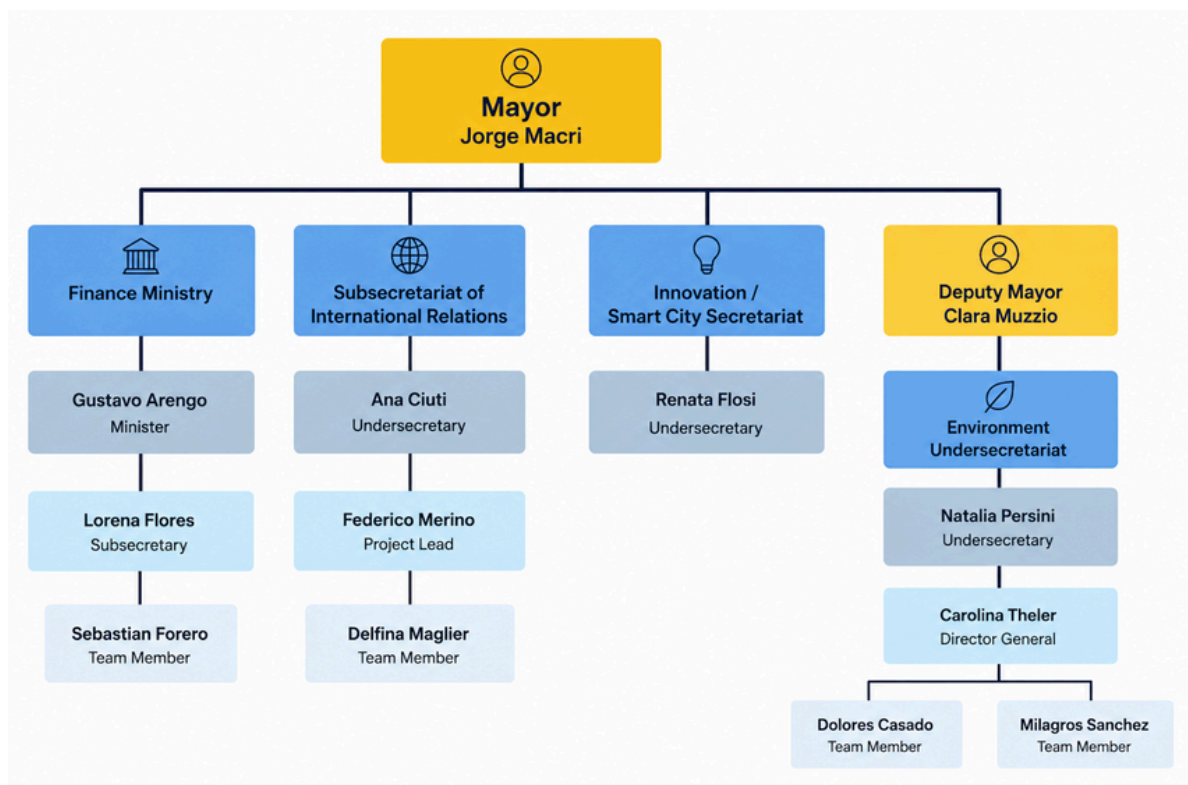
A central component of our methodology has been in-depth stakeholder interviews with CABA government officials and external experts. These conversations validated our understanding of the city's budget process and constraints, identified priorities and gaps, and grounded our recommendations in what is feasible within CABA's institutional context.

BUENOS AIRES STAKEHOLDERS

Across the city government, we engaged teams spanning foreign affairs, finance, environment, innovation, and infrastructure. While the teams we met with had buy-in to develop a more robust climate budgeting system, they flagged that there is limited institutional appetite to invest in this process. This attitude required us to be hyper aware of the operational constraints that will shape what is actually adoptable.

The International Relations team, led by Undersecretary Ana Ciuti, framed climate budgeting primarily through the lens of international positioning. Their priority is ensuring the system produces public-facing outputs that meet the standards expected by C40 and other partners, reinforcing that external credibility is a key driver of the initiative.

The Finance team was among the most consequential conversations for shaping the project's direction. They introduced us to the "Presupuestos Jurisdiccionales," a set of standardized, public-facing program description documents published alongside the city's budget. Rather than one document per budget line, these files are organized by jurisdiction and cover each program within that ministry, providing a more robust explanation of what each program is, what it does, and why it is funded. Each entry follows a consistent structure across ministries, typically including program objectives, activity descriptions, and budget justifications. This standardization was useful because it meant we could apply a consistent reading approach across the full budget rather than adapting to ministry-specific formats.



The team also confirmed the “Formulation” period as the critical window for embedding climate criteria and consistently emphasized that any proposed changes must fit within existing workflows without creating additional reporting burden. Their partnership will be essential to building a shared understanding of what climate budgeting means and why it matters, and is a prerequisite for the system to function over time.

The Environment team is the primary internal champion of the initiative. They developed the initial climate framework and keyword taxonomy underpinning the current tagging exercise, and they are eager to move from the pilot to a more institutionalized system. Their feedback confirmed that our preliminary tagging work was a useful starting point and provided specific direction on refining the classification methodology. The Environment Undersecretary Natalia Persini also highlighted the tagging process itself as a vehicle for building awareness and buy-in across city agencies. Even imperfect tagging, done transparently and iteratively, creates a shared vocabulary for discussing climate alignment and opens the door to more sophisticated analysis over time.

The Innovation team, represented by Undersecretary for Smart City Pedro Alessandri, validated the AI tool development direction we are pursuing. He noted that a semantic reasoning or embedding model trained on the city's budget data could produce climate classifications more accurately and consistently than simple keyword matching.



Fieldwork Image: CABA Government Offices

EXTERNAL STAKEHOLDERS

Our external interviews provided perspective on how peer cities and global networks are approaching climate budgeting, and what lessons are most transferable to CABA's context.

Lia Cairone, Director of Sustainability and Resiliency at the New York City Office of Management and Budget, offered a window into what a more mature climate budgeting system looks like in practice. New York City's approach focuses on advancing climate policies, funding strategic investments, and aligning spending with climate goals broadly consistent with CABA's aspirations, though operating at a much larger scale with a bigger budget office team and more granular project-level data. The transferable lessons were practical: having someone in the budget office with a cross-functional climate mandate is important for ensuring climate considerations are integrated into decisions rather than added after the fact, and any automated system requires genuine buy-in from both the finance office and line agencies. New York's focus on high-level strategic alignment rather than granular emissions accounting is a pragmatic model that may be particularly relevant for CABA given its current analytical capacity.

Sara da Silva Freitas from the C40 Network reinforced that climate budgeting is fundamentally a governance reform, not a financial modeling exercise. She stressed that political commitment and administrative ownership are prerequisites, and that data fragmentation is the most common technical barrier she has observed across cities including Oslo, Mumbai, Rio, and Sao Paulo. Her perspective was a useful reminder that CABA's challenges are not unique, and that C40's frameworks and peer learning resources are available to draw on.

Darryl Sng from Accenture provided grounding on the sequencing of AI adoption. His core message was that the city should first establish a clear taxonomy, a reliable classification framework, and a semi-manual tagging process before moving toward LLM-assisted or more advanced automation. He identified CABA's open-data infrastructure, including Presupuesto Abierto and BA Data, as a genuine comparative advantage that provides a stronger foundation for AI application than many peer cities. He also offered a useful evaluative lens: AI tools should be judged by whether they improve the city's capacity to classify spending at scale and generate results early enough to actually influence budget decisions.

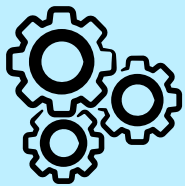


Pedestrian Plaza

KEY TAKEAWAYS FROM INTERVIEWS



Process matters as much as output: The value of climate budgeting is not just in producing a headline figure but in the institutional learning, cross-agency coordination, and accountability mechanisms that the process creates. Recommendations should be designed to strengthen this process over time.



Feasibility within existing systems: Any recommendations must work within CABA's existing budget systems, workflows, and staffing constraints. Adding new platforms or reporting requirements not integrated into how the Finance Ministry already works is unlikely to be adopted.



AI as a scaling tool, not a replacement for judgment: All stakeholders who discussed AI were clear that automated tools should support and accelerate human judgment, not replace it. Ministry staff will need to validate classifications for major budget items.



C40 alignment as both constraint and opportunity: The C40 requirement to institutionalize climate budgeting by 2030 creates a real deadline and accountability mechanism. It also means Buenos Aires can draw on frameworks, tools, and peer-learning resources that C40 has developed, rather than building a system from scratch.

Data Analysis

R EXERCISE: INITIAL BUDGETING TAGGING

We did an initial exploratory exercise to implement climate budgeting in Buenos Aires using 2025 budget data and a keyword list provided by the environmental team. We conducted an automated tagging process in R to identify potentially climate-related budget lines.

The methodology follows the programmatic budget structure and applies a two-level tagging approach. Level 1 tags a budget line when a match is found in either activity-level variables or program-level variables. Level 2 is more restrictive, tagging only when matches are found in both levels simultaneously, increasing confidence in the classification.

We tested two scenarios. Scenario 1 used the original list of 103 keywords and resulted in 6,865 lines tagged, representing 23.9% of the approved budget. We analyzed the result and modified the keyword list, removing generic terms producing high false-positive rates (footnote w examples?) and added more specific terms and named programs. Scenario 2 used this refined list of 158 keywords and produced 1,391 tagged lines, representing 5.82% of the budget, reflecting a more targeted but narrower classification.

Overall, this exercise allowed us to test how available data and keyword-based methods can support climate budget classification, while highlighting the importance of refining inputs and validating results before scaling the approach

How it works

- 1. Normalize text**
Lowercase, remove accents, trim whitespace — applied to both budget text and keywords
- 2. Search 5 columns**
Match keywords against Desc_Act, Desc_Obra, Desc_Prog, Desc_Sprog, Desc_Proj
- 3. Word-boundary matching**
Prevents partial matches — 'led' cannot match 'delegaciones'
- 4. Two classification levels**
Mínimo (Act/Obra) and Mayor (Prog/Sprog/Proy)
- 5. Flag generic terms**
22 ambiguous keywords flagged separately for manual review

Budget columns searched(*)

MÍNIMO NIVEL

Activity/Works level — most broad, minimum level

Desc_Act

Desc_Obra

MAYOR NIVEL

Program/Project level — more granular

Desc_Prog

Desc_Sprog

Desc_Proj

Levels of tagging

- 1 First level of tagging** | Matched at least one level
- 2 Second level of tagging** | Matched at BOTH levels — highest confidence



Generic terms flagged

22 ambiguous keywords flagged separately for manual review

SCENARIO 1

103 keywords across 10 thematic areas

103 Keywords Original list	6,865 Lines tagged 17.1% of all lines	89.2% Generic-only rate 6,127 of 6,865 tagged	738 Confident matches 1.8% of all lines
---	--	--	--

⚠ Problem: 91.9% of tagged lines matched only generic terms

Terms like 'educacion' (4,655 false positives), 'asistencia' (1,472), and 'limpieza' (96) drove the vast majority of matches, but not necessarily climate spending.

Financial results

	Budget lines	ARS (billions)	% of approved budget
All tagged	6,865	\$3,670B	23.9%
— Generic-only	6,127	\$3,307B	21.5%
— Confident	738	\$363B	2.37%

SCENARIO 2

158 keywords — same matching process, targeted improvements

158 Keywords +55 net additions	1,391 Lines tagged 3.5% of all lines	22.9% Generic-only rate vs. 91.9% in Pass 1	1,072 Confident matches 5.06% of approved budget
---	---	--	---

What changed

3 removed	High false positive rate (educacion, asistencia, limpieza)
34 removed	Generic phrases, duplicates; replaced some words
65 kept	Retained 65 original keywords
93 added	Identified some from review of budget descriptions, added program names, added variations, replaced generic terms
22 flagged	Generic list for manual review

Financial results

	Lines	ARS (B)	% budget
All tagged	1,391	\$893.7B	5.82%
Generic-only	319	\$117.4B	0.77%
Confident	1,072	\$776.4B	5.06%
— <i>Mitigation</i>	359	\$559.2B	3.03%
— <i>Adaptation</i>	722	\$429.2B	2.80%
Both levels of tagging	489	\$484.3B	3.16%

*Note: confident Vigente is high relative to line count — climate-tagged lines are significantly larger than average (\$724M vs \$372M mean Vigente per line)
Approved budget (Vigente) = original sanctioned budget plus mid-year modifications.

THEMATIC BREAKDOWN

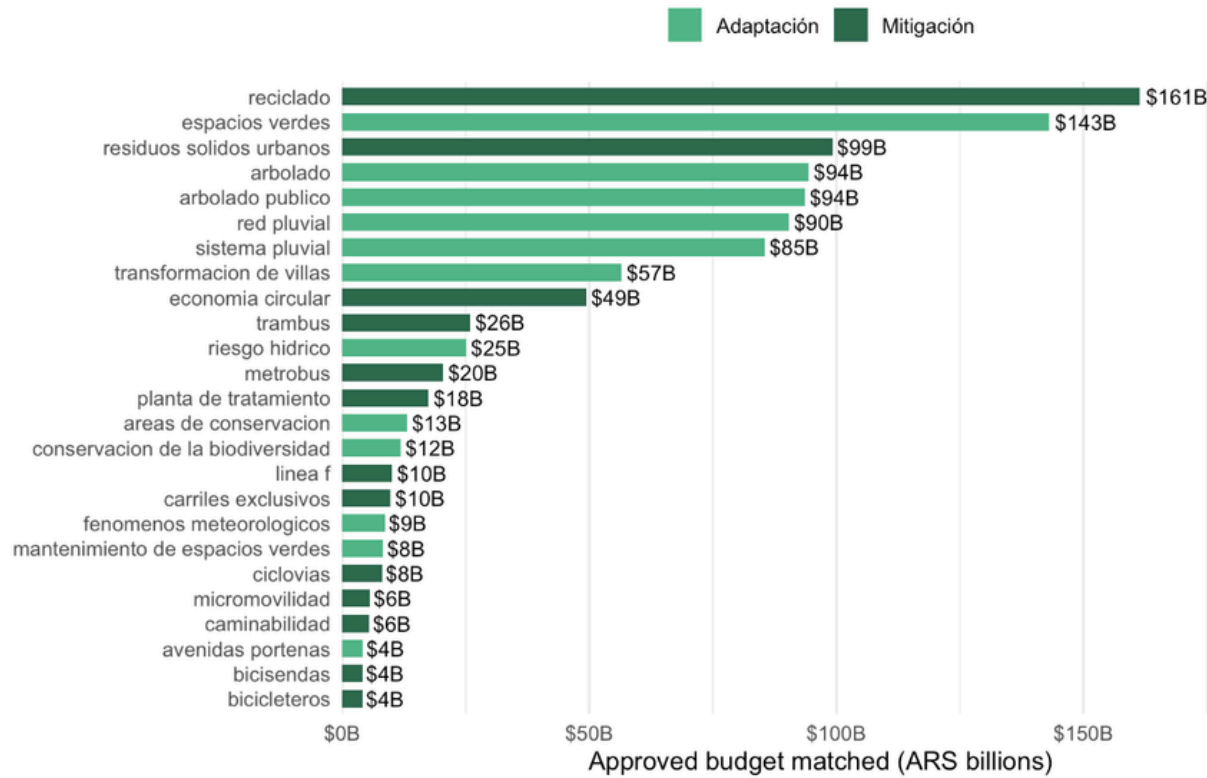
Pass 2 — confident matches, % of total approved budget (Vigente)

Thematic area	Category	% of budget	n
Waste Management	Mitigation	2.64%	507
Urban Green Infrastructure	Adaptation	2.46%	699
Hydro/water Risk Mgmt	Adaptation	1.39%	229
Low-Emission Transport	Mitigation	0.62%	176
Informal Settlement Integration	Adaptation	0.37%	52
Circular Economy	Mitigation	0.33%	35
Health & Climate Change	Adaptation	0.052%	32
Energy & Construction	Mitigation	0.007%	81
Environmental Education	Adaptation	0.001%	20
Sustainable Food	Adaptation	0.0002%	3

⚠ Energy & Construction captures very little — likely underrepresented due to generic budget coding.

TOP MATCHED KEYWORDS

Top 25 keywords by approved budget matched
ARS billions | confident matches only



This exercise was useful as a baseline but it had significant limitations. The budget data available at that stage lacked program-level descriptions, which made it impossible to distinguish climate-relevant subcomponents within multi-purpose programs or to apply proportionality weighting. During our fieldwork in CABA, the Finance team provided us with the “Presupuestos Jurisdiccionales.” These documents were a key missing part of our analysis thus far because they gave us substantive programmatic context that the budget line data alone could not provide. Where a raw budget line might only contain a brief label like "Conducción" or "Mantenimiento," the corresponding Presupuesto Jurisdiccional entry would explain the program's underlying purpose, making it considerably easier to assess whether climate-relevant keywords were genuinely applicable or simply incidental. This additional context meaningfully improved classification accuracy when combined with the budget line data.

That said, the “Presupuestos Jurisdiccionales” have real limitations. While the descriptions are more informative than the budget line fields alone, they remain relatively general. Program entries do not always contain enough specificity to distinguish between climate-relevant and non-climate subcomponents within a single multi-purpose program, and they do not include quantitative data on emissions, energy use, or environmental outputs. As a result, they improved our ability to classify programs directionally, but could not substitute for the kind of project-level detail that a more mature climate budgeting system would require.

The Innovation team also validated our goal of using an AI-based approach to replace the manual, difficult-to-scale keyword matching process. Our conversations on the ground in CABA shaped the second stage of our work.



Fieldwork Image: Recycling Center



Fieldwork Image: Across from Plaza de Mayo

Methodology

The second stage of our work was to build an AI-assisted tagging tool based on the refined keyword taxonomy from the R exercise. Our goal was to address the core limitations of the manual approach: inconsistent classification across budget lines, no proportionality logic, and a process that could not realistically scale to annual use by the Environment team. The Environment team provided several potential methodologies for us to follow and we conducted in-depth research to inform which method would work best for CABA.

COMPARATIVE CITY EXAMPLES

City / Jurisdiction	Climate Budget Share / Scale	Methodology Type	Weighting Approach	Classification System	Budget Integration
 Mumbai	~37.8% of capital budget climate-aligned	Climate budget tagging aligned with climate action plan	Full attribution (100% of tagged expenditures)	Action-based classification (climate action plan sectors)	Fully integrated into budget formulation (circulars, templates, departmental inputs)
 New York City	Large-scale capital investment (multi-billion annually)	Climate expenditure tagging within capital planning framework	Binary attribution (climate / non-climate)	Program- and agency-based classification	Partially integrated (strong in capital planning, limited in operating budget)
 Montreal	~20–30% estimated climate-relevant spending	Climate lens applied during budget review	Qualitative attribution (non-standardized partial tagging)	Thematic classification (e.g., mitigation, resilience)	Partially integrated (applied during budget review stage)
 Paris	~20%+ of expenditures climate tagged	Climate budget tagging framework	Tiered attribution (positive / neutral / negative impact)	Functional classification based on climate impact	Fully integrated into budget formulation and approval
 London	Climate spending aligned with net-zero targets	Program-based climate budgeting approach	Binary attribution (program-level classification)	Sector-based classification (e.g., transport, energy)	Integrated into strategic planning and budgeting processes
 Colombia (Rio marker approach)	Applied across total public expenditure	Coefficient-based climate budget tracking	Weighted attribution (e.g., 100% principal, 40% significant)	Rio marker classification (principal vs. significant objectives)	Primarily applied during reporting and tracking
 Buenos Aires (Proposed)	To be established (baseline assessment required)	Hybrid approach: tagging aligned with climate action plan + coefficient-based weighting	Tiered and weighted attribution (100% principal, 40% significant, 0% non-climate)	Dual classification: climate action plan sectors + Rio marker categories	Phased integration: pilot → integration into budget formulation → citywide expansion

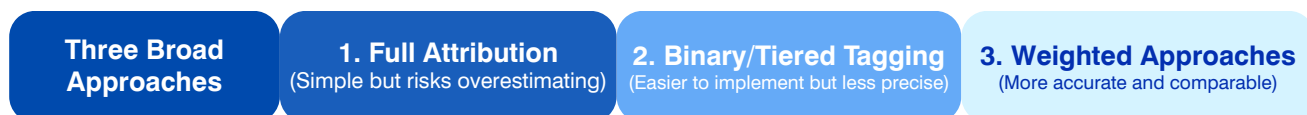
The table above compares how different cities and frameworks structure climate budgeting, with a focus on three key dimensions: how climate expenditures are identified, how they are weighted, and how they are integrated into the budget process. While cities differ in their institutional arrangements and policy contexts, a central methodological question cuts across all cases: how to determine the proportion of public spending that can be meaningfully attributed to climate action.

Across the cases we reviewed, three broad approaches emerge, each with different implications for accuracy, usability, and policy relevance.

A first group of cities, including Mumbai and London, relies primarily on full attribution linked to climate action plans, where expenditures aligned with climate priorities are counted as fully climate-relevant. This approach is effective for embedding climate considerations into budget processes and ensuring strong alignment with policy goals. However, it tends to overestimate climate spending, as it does not account for the fact that many public investments, particularly in urban infrastructure, serve multiple objectives simultaneously.

A second group, including New York City and Paris, uses tagging systems that classify expenditures as climate-related or not, sometimes with tiered categories reflecting positive or negative impacts. These approaches are relatively straightforward to implement and can be integrated into existing budget frameworks with limited disruption. However, they are limited in their ability to capture the degree of climate relevance within multi-purpose programs, as they typically rely on binary or categorical classifications without proportional weighting.

A third approach, used by Colombia and reflected in the Rio Marker framework, introduces coefficient-based weighting, whereby expenditures are assigned a percentage that reflects their contribution to climate objectives. This method directly addresses the limitations of both full attribution and binary tagging by recognizing that climate relevance exists along a spectrum. As a result, it provides a more accurate and comparable estimate of climate spending across sectors and programs, and has become a widely used standard in international climate finance reporting.



This comparison highlights a clear trade-off. Simpler systems, such as full attribution or binary tagging, are easier to implement but risk overstating or mischaracterizing climate spending. More advanced systems, such as weighted approaches, require greater methodological clarity and institutional capacity, but produce more precise and policy-relevant results.

In practice, many governments are moving toward hybrid models that combine policy alignment, tagging, and weighting within a single framework. These systems retain the usability of tagging approaches while incorporating the analytical rigor of proportional weighting. This evolution reflects a broader shift in climate budgeting from descriptive reporting toward more accurate measurement and decision-support tools.

Within this context, weighting emerges as a critical foundation for robust climate budgeting. By quantifying the extent to which each expenditure contributes to climate objectives, it enables governments to move beyond simple classification and toward a more nuanced understanding of how public resources are allocated in practice.

IMPLICATIONS FOR BUENOS AIRES

Our comparative analysis demonstrates that CABA is well-positioned to adopt a weighting-based climate budgeting system aligned with the OECD Rio Marker methodology, while extending it through the use of automated classification and city-specific policy frameworks. This approach is consistent with international practices and allowed us to proceed to the next stage of our AI tool development.

THE THREE TIERS (METHODOLOGY A)

Rio Markers classify activities into three tiers, each corresponding to a weight applied to the activity's financial value when calculating climate expenditure:

Score	Label	Definition
 2 (100%)	Principal Objective	Climate is the fundamental purpose of the activity, the program would not exist without its climate goal.
 1 (40%)	Significant Objective	Climate is an explicit but secondary consideration, the program would exist regardless, but includes climate-relevant components.
 0 (0%)	Not Targeted	The activity has no explicit climate objective.

INITIAL FIVE-TIER PROPOSAL (METHODOLOGY B)

We initially explored a five-tier framework that added two intermediate weights 70% (Mandated Integration) and 10% (Administrative Support), between the standard OECD tiers. This extension was inspired by practices at specific institutions (New Zealand's Disaster Risk Reduction approach, Germany's Environment Ministry weighting for mandated programs) and aimed to capture nuance that the three-tier OECD structure misses.



Both methodologies, the standard three-tier OECD framework (Methodology A) and the extended five-tier framework (Methodology B), were implemented throughout the exploratory phase so that results could be compared. As we will describe in Section 7, we elected not to carry the five-tier framework forward into the production delivery, but it remains part of our analytical record.

AI Tool Development

Our first approach was a single large prompt that instructed Claude to perform the entire workflow in one pass. We directed Claude to read the budget, look up keywords in the taxonomy, apply the weighting methodology, classify each line, calculate climate expenditure, and generate both Excel and HTML deliverables. The prompt was approximately 2,000 words long and included methodology definitions, keyword references, decision rules, and output specifications.

This design followed the intuitive assumption that more instruction produces better outputs. In practice, it produced subsequent problems that informed much of our project. Informed by our research, we applied the OECD Rio Marker methodology to the executed 2025 budget:

- 1 Read 45,444 budget line items from the city's executed 2025 budget
- 2 Classify each budget program as: Mitigation, Adaptation, Both, or None
- 3 Assign a climate weight (0%, 40%, or 100%) per program
- 4 Calculate the actual climate spending in Argentine pesos (Devengado)
- 5 Produce a structured output ready for team review and future dashboard development

CLAUDE PROJECT RATIONALE

We chose to implement the pipeline inside a Claude Project rather than as standalone chats. A Project in the Claude interface allows:

- Custom instructions that persist across every new chat in the project
- Shared knowledge files that every chat can access
- Clean separation between stable configuration (the methodology) and variable inputs (the budget file)
- Easy team sharing by granting access to the project rather than redistributing files

This architecture matched our constraints well: it requires no infrastructure, no developer involvement, and can be used by non-technical team members by simply opening a new chat and uploading the budget file.

THE 11 THEMATIC AXES

Our classification uses CABA's own climate framework with 11 thematic axes split between Mitigation (reducing emissions) and Adaptation (adjusting to climate impacts):

TYPE	THEMATIC AXIS
Mitigation	Eje 1 — Low-emission transport (cycling, electric vehicles, public transit)
Mitigation	Eje 2 — Waste management (recycling, composting, biogas)
Mitigation	Eje 3 — Energy and sustainable construction (LED, solar, insulation)
Mitigation	Eje 4 — Circular economy (reuse, green certification)
Adaptation	Eje 5 — Urban green infrastructure and nature-based solutions
Adaptation	Eje 6 — Heat wave preparedness (climate shelters, early warning)
Adaptation	Eje 7 — Hydrometeorological risk (flood defenses, drainage)
Adaptation	Eje 8 — Urban integration of informal settlements
Adaptation	Eje 9 — Environmental education
Adaptation	Eje 10 — Sustainable food systems (urban agriculture)
Adaptation	Eje 11 — Health and climate change (heat health, dengue control)

KEY CHALLENGES DURING DEVELOPMENT

We went through several iterations throughout development, discovering and resolving four main challenges.

1

INCONSISTENT RESULTS ACROSS RUNS

A recurring problem throughout development was that the same prompt, run in a different chat session or by a different team member, produced different outputs. Climate share percentages shifted between runs, the count of tagged lines changed, and justifications for identical budget lines were worded differently each time. This was not a minor inconvenience. A climate expenditure figure that changes between runs cannot be defended to auditors, and a methodology that produces different results for different users cannot be published. Addressing this became the central challenge of the project.

The root cause is that large language models generate outputs probabilistically. Even with identical inputs, each run samples from a slightly different distribution, which introduces randomness that compounds when the prompt requires judgment rather than rule-following. Our original prompt had several such instructions, including an AI inference fallback, free-form justification text, and weight distinctions that asked the model to assess whether climate was a program's principal or significant objective. Each of these gave the model room to interpret, and that room produced variance.

2

THE PRESUPUESTOS JURISDICCIONALES DID NOT MAP CLEANLY TO THE BUDGET STRUCTURE

When the Finance team shared the Presupuestos Jurisdiccionales, the expectation was that their program descriptions could be matched directly to rows in the executed budget file, giving the classification model richer context for each line. In practice, the match was much harder to execute than anticipated. The PDFs describe programs at a relatively high level, while the Excel budget contains hundreds of sub-lines under each program, many of them carrying identical labels like "Conducción" or "Mantenimiento" that give no indication of the program they belong to. Attempts to match individual rows to PDF paragraphs through fuzzy matching produced poor results, including cases where lines were matched to the wrong ministry entirely.

The consequence was that the OCR extraction and matching pipeline we initially built, which took over 20 minutes to run and still produced a 51% accuracy rate, had to be abandoned. We moved instead to program-level classification, where the model classifies each of the 451 unique programs once and that classification is then applied to all sub-lines beneath it. This was a more reliable approach, but it meant that nuance within multi-purpose programs was lost.

3

THE METHODOLOGY REQUIRED DATA WE DID NOT HAVE

The initial five-tier framework included a 70% weight for programs whose design is legally required to incorporate climate criteria, covering programs subject to ACUMAR environmental mandates, IVC sustainability standards, or the 2019 Buenos Aires Building Code. Applying this tier correctly requires a definitive lookup connecting each program to its governing legal mandate. That information does not exist in the budget fields, the keyword taxonomy, or any structured data source we had access to. It lives in legal documents and in the institutional knowledge of the Environment team.

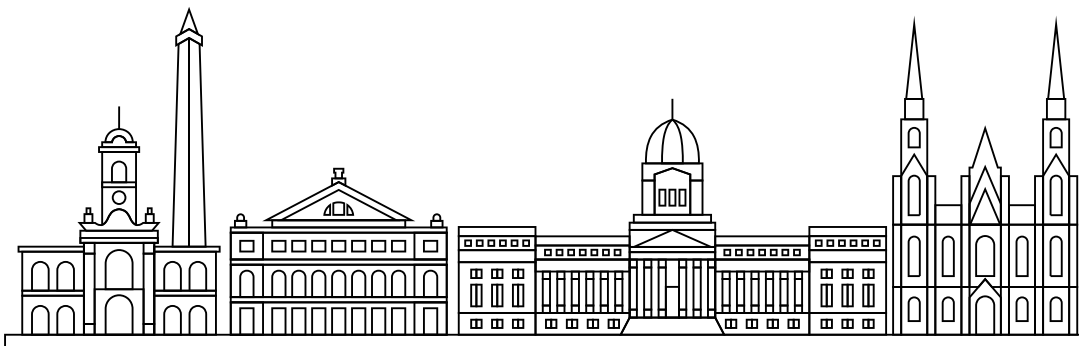
Without that lookup table, applying the 70% tier meant asking the model to infer mandate applicability from program names alone, which reintroduced exactly the kind of judgment-based reasoning that was causing our reproducibility problem. The 2x2 study made this concrete: five different design configurations produced five different counts of lines at the 70% tier, ranging from 0 to 1,230, for the same underlying budget. The tier was dropped from the production methodology as a result, though the path to reintroducing it is clear once the lookup table is built.

4

KEYWORD AMBIGUITY

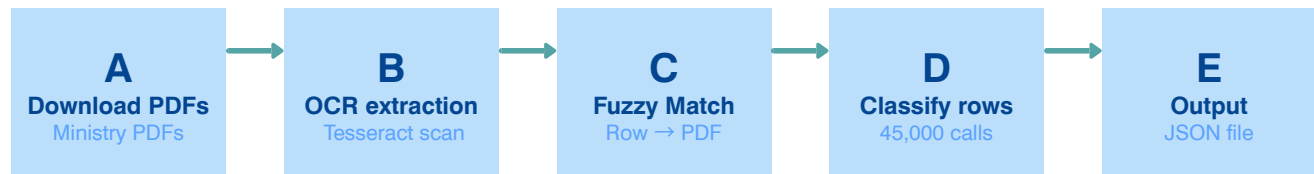
The keyword taxonomy included several terms that carry different meanings depending on context. The clearest example is "aislamiento," which refers to thermal insulation in building and energy contexts but to quarantine or social isolation in health and social service contexts. The taxonomy flagged this with bracketed qualifiers noting it should only be applied in construction or energy contexts, but the first version of the prompt stripped those qualifiers and treated every match as climate-relevant. The result was false positive tags on programs with no connection to climate, artificially inflating the reported climate share.

Fixing this required restructuring how the prompt handles qualified keywords entirely, moving from a simple match to a slot-filled output format that forces the model to record which field the keyword was found in before a tag is applied. This change also had the side effect of reducing the human review queue from 7,411 lines to 327, since many of the earlier flags had been triggered by qualified keywords firing incorrectly on non-climate content.



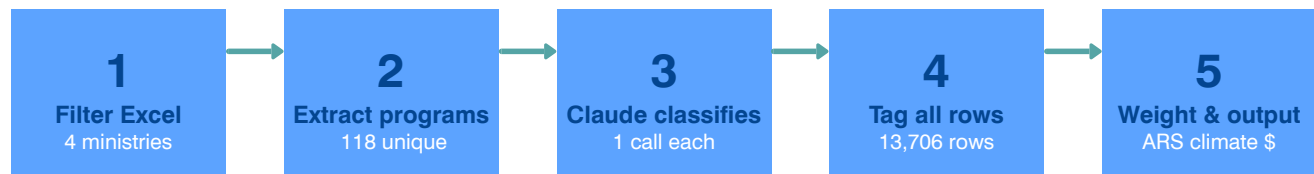
WHAT WE BUILT: STEP BY STEP

ORIGINAL PIPELINE (REPLACED)



20+ min OCR | Poor match quality | 18 hour runtime

FINAL PIPELINE (IN USE)



No PDFs | 118 API calls | Under 5 minutes

THE WEIGHTING SCALE

Each program receives one of three weights, based on how central climate is to its purpose:

Weight	Category	When applied
100%	Direct Action	Climate IS the primary objective. The program would not exist without it. Example: waste recycling, flood drainage, urban parks.
40%	Enabling Action	The program has a meaningful climate co-benefit alongside another primary goal. Example: public transport subsidies, LED street lighting.
0%	Administrative	No climate connection. General management and support. Example: Conducción, legal services, communications.

CLIMATE BUDGET FORMULA:

$$\text{Devengado (executed spend)} \times \text{Climate Weight \%} = \text{Presupuesto Climático}$$

We use Devengado, money actually spent, not the approved or adjusted budget, because it reflects real investment.

Design Configuration

Before settling on the final methodology and weighting scale, we ran a systematic comparison of the four design configurations. We varied whether the AI was allowed to use inference beyond keyword matching and whether the PAC2050 Climate Action Plan was supplied as a reference document.

Configuration	Method	Share	Climate (ARS)	Tag=1	@100	@70	@40
Without AI + With CA	A	9.90%	1,427 B	6,162	1,375	—	4,787
Without AI + With CA	B	10.41%	1,500 B	6,162	1,375	1,230	3,557
Without AI + Without CA	A	14.74%	2,125 B	9,223	3,694	—	5,529
Without AI + Without CA	B	14.92%	2,151 B	9,223	3,694	437	5,092
With AI + With CA	A	9.40%	1,355 B	5,486	1,482	—	4,004
With AI + With CA	B	9.56%	1,378 B	5,486	1,482	193	3,811
With AI + Without CA	A	7.85%	1,131 B	4,881	1,082	—	3,799
With AI + Without CA	B	8.08%	1,165 B	4,881	1,082	566	3,233

Note: In the table, "CA" denotes the PAC2050 Climate Action Plan reference. Climate spending is expressed in billions of Argentine Pesos. Tag=1 is the count of lines classified as climate-relevant. @100 / @70 / @40 are the row counts at each weight tier. All eight results were produced from the same 45,444-row CABA 2025 executed budget and the same keyword taxonomy.

We found the same budget, same taxonomy, and same underlying methodology were producing different results. Climate share estimates ranged from 7.85% to 14.94% of the total executed spending, a spread of roughly 1 trillion ARS. This was driven entirely by how the classification prompts were structured, not by any difference in the data itself. This finding shaped every subsequent design decision: it demonstrated that without a principled and consistent methodology, any single climate share estimate could be challenged.

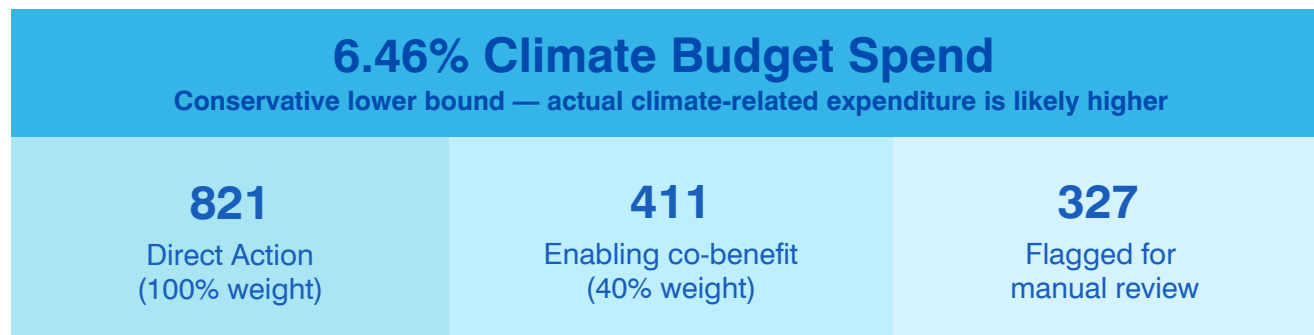
The exploratory phase also revealed a subtler problem. Even when explicit AI inference was turned off, the model still exercised judgment through the free-form justification text it was asked to produce. Because the AI was required to explain each classification decision in writing, it would interpret the result and that interpretation introduced variances upstream, even when the tagging rules themselves were meant to be strict. Removing this source of variability required us to restructure the outputs entirely and replace the written justifications with a fixed template of five slot-filled fields that the AI populates from matched data rather than writing freely.

The final methodology went through four iterations before reaching a stable, defensible result. The key changes across versions were:

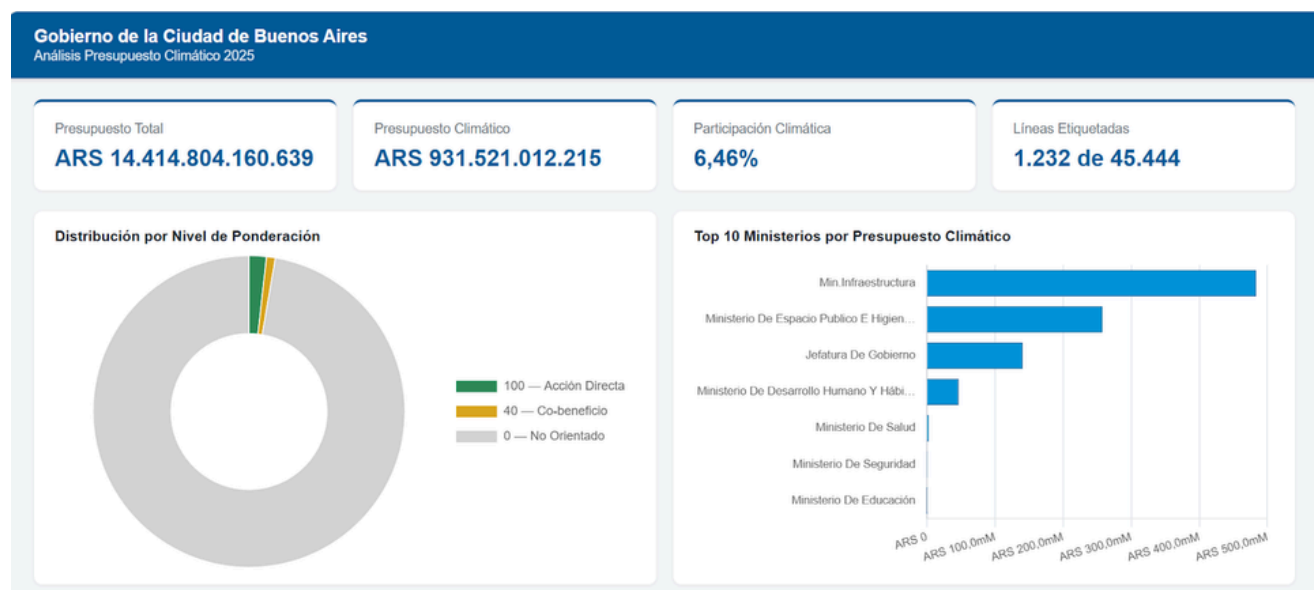
- **Removing AI inference entirely and replacing it with strict keyword matching**
- **Introducing a mechanical rule that assigns 100% weight when a climate keyword appears in the program name (Desc_Prog) and 40% when it appears only in sub-activity or project fields**
- **Restricting the human review flag to operate on the eight core budget fields rather than program description text that contained generic language and was generating thousands of irrelevant flags**

Results

The production version (V4) classifies 821 budget lines as Direct Action (100% weight), 411 as Enabling Co-benefit (40% weight), and flags 327 lines for human review, down from an initial review queue of 7,411 lines in the first deterministic draft. Total climate spending comes to approximately ARS 930 billion, representing 6.46% of the city's 2025 executed budget.



This figure sits below all eight exploratory configurations, which is by design: it reflects only what the data directly supports, excluding any classification that requires interpretive judgment. It should be understood as a conservative lower bound. As the methodology matures, with a validated mandate lookup table, an expanded keyword taxonomy, and manual review of the flagged lines, the reported share will increase, but on a more defensible evidentiary basis.



Note: The figures presented in this report are estimates based on publicly available budget data available at the time of analysis and do not constitute official CABA figures. For the most accurate and up-to-date analysis of the City's climate budget, the relevant reference sources are BA Data, the Dirección General de Estadísticas, and the Dirección General de Presupuesto.

This method is the strongest one for the Environment team's use case, because:

Among the eight exploratory configurations, the lowest climate share (7.85%, With AI + Without Climate Action) is still higher than the production figure (6.46%). This is because the production prompt goes further than removing AI inference, it also restructures the justifications, enforces bracketed qualifier logic, narrows the soft-signal check, and expands the blocklist. Each of these removes a source of overtagging that the exploratory configurations still contained.

The highest climate share (14.92%, Without AI + Without Climate Action + Method B) is the least defensible result. It looks strict because "Without AI" was chosen, but in reality the justification-driven interpretation was doing the work of the inference fallback without being named, an undocumented source of variance that auditors would find hard to accept.

Methodologies B produces consistently higher totals than their A counterparts because the 70% tier captures additional lines. Given the concerns documented in Section 7 about the data required to support the 70% tier, Methodology A is the more defensible choice in the current data environment.

The production approach sits at the conservative end of the possible results. It is the result most likely to hold up under audit challenge, because it excludes every interpretation the data does not strictly support. If the Environment team wants to report a higher figure, they can do so, but only by adding the specific data (legal mandate lookup, expanded taxonomy) needed to justify the difference.

In short, the 2×2 study establishes that any single climate share figure can be contested on methodological grounds. The production approach is the most defensible answer given the data currently available. Every step that produced a higher figure in the exploratory configurations can be traced to a prompt structure that invited interpretation. The production approach removes those invitations.

Note: *Given that the methodology is at an early stage of implementation and validation for this jurisdiction, the results may present margins of error or interpretive inconsistencies. The estimates should not be considered definitive or conclusive values, but rather technical approximations subject to further review and methodological refinement.*

Deliverables



Prompt 1: Tagging

Custom instructions for a Claude Project - upload any CABA budget, get tagged_budget.csv + review_queue.csv.
Zero infrastructure needed.



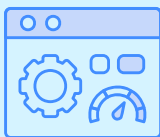
Prompt 2: Deliverables

Takes tagged CSV and produces the formatted Excel workbook + self-contained HTML dashboard in one run.



6-Sheet Excel Workbook

- | | |
|---------------------------------------|--------------------|
| 1) Resumen (KP/s + charts) | 4) No Climáticos |
| 2) Detalle Clima (1,232 tagged lines) | 5) Datos Completos |
| 3) Revisión Manual (327 review lines) | 6) Metodología |



Dashboard

BA-branded; Editable weights that update KPIs live; Filters by ministry, axis, weight; Virtual scroll for 45K rows; Excel/CSV download.



Methodology Documentation

Full audit trail: this report + README + OECD Rio Markers reference + decisions log template for the environment team.



Step-By-Step Video Explainer

Guided video of how to use AI tool/dashboard.

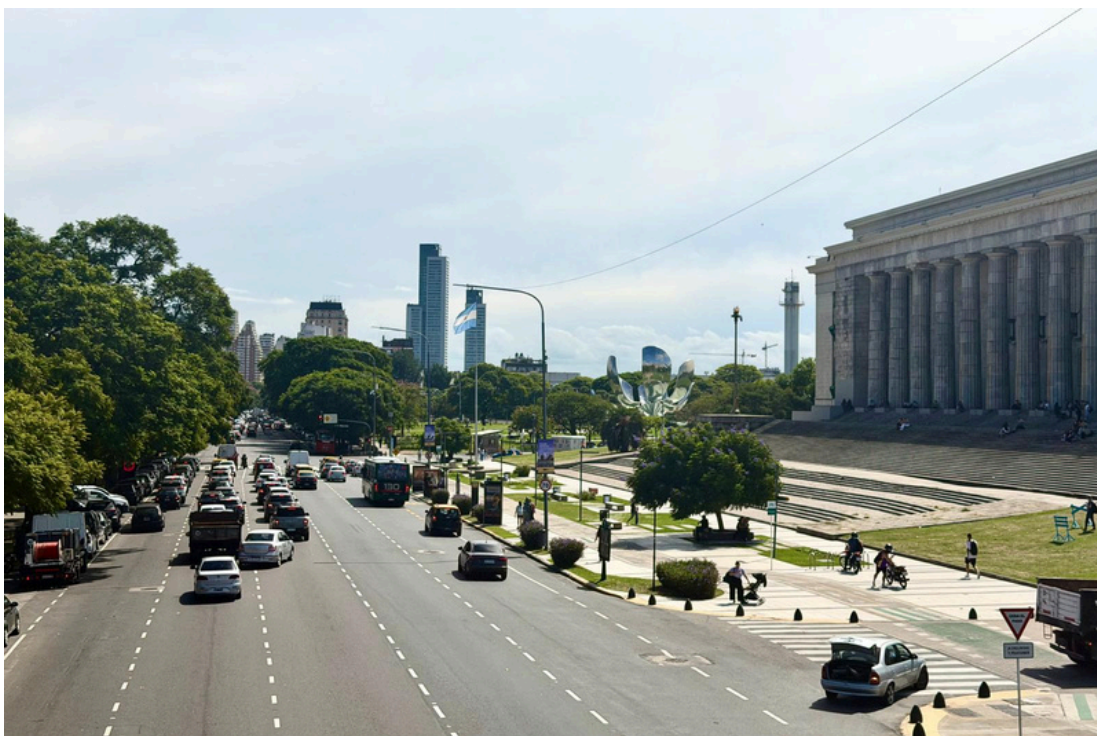
Challenges

The current system has several limitations that the city's team should understand before using the results for public reporting.

The 6.46% figure is a lower bound, not a complete count. By removing AI inference and accepting only keyword-matched classifications, the methodology will overlook climate-relevant programs that lack exact keyword matches in the budget fields. Higher figures are possible and may be more accurate, but they require additional data to justify, specifically, an expanded keyword taxonomy and a validated program-to-mandate lookup table for the 70% tier.

The chat-based workflow cannot achieve full reproducibility. Even with deterministic rules, two runs of the same prompt on the same file may differ on a small number of genuinely ambiguous cases, because language models sample outputs probabilistically. For the current use case, a first-generation pilot exercise, this level of consistency is adequate. For official reporting, migrating the methodology to a Python pipeline using the Claude API with temperature set to zero would eliminate this residual variance entirely.

The methodology has not yet been validated against an independent manual classification. Given that this is the first time CABA has produced a climate share figure for its executed budget, there is no prior baseline to compare against. A small stratified sample (roughly 100 lines tagged independently by a reviewer) would provide an empirical accuracy check and should be the first post-delivery activity.



Recommendations

VALIDATE AND REVIEW REMAINING QUEUE

In the near term, the most valuable action is to work through the 327-line review queue.

This serves two purposes: it refines the reported climate share, and it generates the decisions needed to expand the keyword taxonomy over time. Alongside this, the environment team should compile a short list of agencies whose mandate is exclusively environmental (candidates include APRA, Dirección General Sistema Pluvial, and Dirección General Reciclado). This list would enable a 10% administrative support tier to be added to the methodology quickly and without sacrificing reproducibility.

STRENGTHEN DATA FOUNDATIONS

Over the medium term, the priority should be building the program-to-mandate lookup table that would support the 70% tier.

This requires the environment team, in coordination with the Finance Ministry, to map specific programs to the legal mandates that require climate criteria, ACUMAR, IVC sustainability standards, the 2019 Building Code, with citations. Once this table exists, the 70% tier can be reintroduced into the methodology without reintroducing the ambiguity that made it unreliable in the exploratory phase. Progress on this front is also directly relevant to CABA's C40 commitments: the network requires member cities to institutionalize climate budgeting by 2030, and a methodology that can be defended, reproduced, and audited annually is a prerequisite for meeting that standard. The near- and medium-term steps described here are not just internal improvements; they are the building blocks of a system that can satisfy external scrutiny.

INTEGRATE MINISTRIES IN THE PROCESS

A related area for medium/long-term consideration is broadening ministry participation in the data collection process.

The current methodology relies entirely on publicly available budget data, which was a practical and appropriate choice for a first-generation pilot. However, this also means the system is limited by what the public record contains. If other ministries hold program-level information, project documentation, and planning data, that could meaningfully improve classification accuracy and support more proactive tagging during budget formulation rather than after the fact. A future engagement could focus on developing the protocols and inter-ministerial agreements needed to bring that data into the pipeline, particularly during the formulation window between May and August when climate criteria could actually influence resource allocation decisions rather than simply describe them after spending is locked in.

MAP OUT LONG-TERM TECHNICAL DEVELOPMENT

Long-term, migrating the pipeline from a chat-based workflow to a Python script using the Claude API would be the single most impactful improvement. This would bring reproducibility close to 100%, remove the current context window constraints, and allow the system to scale to future budget years without manual handling of CSV files. With a developer resource and a few weeks of work, the methodology could also be wrapped as a simple web interface, allowing the environment team to upload a budget file and receive a tagged output without running any prompts directly. That architecture would also support automatic drift detection, flagging if a new run produces a materially different climate share than the prior year without a corresponding change in the budget or taxonomy, which would serve as an ongoing quality check.

TRANSFORM INTO A DECISION-MAKING TOOL

For future capstone teams, the most important unfinished work is expanding this tool from a retrospective analysis instrument into a prospective decision-support tool. The current system answers the question: how much of last year's budget was climate-relevant? The more useful question for actual climate governance is: as next year's budget is being formulated, which programs should be prioritized, expanded, or redirected to improve the city's climate performance? Embedding the tagging methodology into the budget formulation period, rather than applying it after the fact to executed spending, would make it a genuine governance tool rather than a reporting exercise.

Appendix

REFERENCES

1. C40 Cities. (2024). C40 Climate Budget Framework. C40 Cities Climate Leadership Group. <https://www.c40.org/>
2. Ciudad Autónoma de Buenos Aires. (2019). Plan de Acción Climática 2050. Ministerio de Ambiente y Espacio Público, GCBA. https://buenosaires.gob.ar/gcaba_historico/agenciaambiental/cambioclimatico/planes-de-accion-climatica-de-la-ciudad
3. Merino, F. (2026, March). Senior Advisor, Undersecretariat for Foreign Affairs, City of Buenos Aires. In-person interview, Buenos Aires [Fieldwork].
4. Finance Team, City of Buenos Aires. (2026, March). Ministry of Finance (Ministerio de Hacienda), GCBA. In-person meeting, Buenos Aires [Fieldwork].
5. Government of the Autonomous City of Buenos Aires. (2025). Political and Institutional Structure of CABA. GCBA. [Finance Team deck]
6. S&P Global Ratings. (2025, July 1). Research update: City of Buenos Aires 'B-' rating. <https://www.spglobal.com/ratings/en/regulatory/article/-/view/sourcelid/101632928>
7. World Resources Institute. (2023, September). Monitoring Climate Finance. <https://files.wri.org/d8/s3fs-public/2023-09/pfp-monitoring-climate-finance.pdf>
8. SIPA Capstone Team (2024–25). Climate Budgeting in Buenos Aires: Recommendations Report [Prior Capstone]. Columbia University School of International and Public Affairs.
9. Theler, C., & Casado, D. (2026, March). General Director of Environmental Policy and Strategy / Technical Assistant for Policy and Environmental Strategy, Undersecretariat of Environment, City of Buenos Aires. In-person interview, Buenos Aires [Fieldwork].
10. Climate Budget Report: FY 2024-25. Mumbai Climate Action Plan. https://mcap.mcgm.gov.in/wp-content/uploads/2024/06/English_Climate-Budget-Report-2024-25_compressed.pdf
11. Mayor's Office of Management and Budget. New York City Climate Budgeting. Executive Budget. (2025). <https://www.nyc.gov/assets/omb/downloads/pdf/exec24-nyccb.pdf>
12. Climate Budget. Montreal (2026). <https://mtl.ged.montreal.ca/constellio/?collection=mtlca&portal=REPDVCVDM&lg=en#!displayDocument/00000133437>
13. Greater London Authority, The Mayor's Budget Guidance 2026-27. (2025). <https://www.london.gov.uk/sites/default/files/2025-07/Mayor%27s%20Budget%20Guidance%202026-27%20FINAL%20%28edit%29.pdf>
14. Paris Climate Action Plan 2024-2030. (June 2025). <https://cdn.paris.fr/paris/2025/06/25/plan-climat-en-9E8Q.pdf>
15. OECD Development Co-operation Directorate and OECD Development Assistance Committee, Results of the Survey on the Coefficients Applied to Climate Change Rio Marker Data When Reporting to the UNFCCC, DAC Working Party on Development Finance Statistics. (2024, October). [https://one.oecd.org/document/DCD/DAC/STAT\(2024\)28/REV1/en/pdf](https://one.oecd.org/document/DCD/DAC/STAT(2024)28/REV1/en/pdf)
16. AI-Automated Climate Budget Tagging: A Capstone Project Report Designing a Reproducible AI-Based Pipeline for the Buenos Aires City Government Executed Budget 2025. (2025, May). https://drive.google.com/file/d/1_GctfrM0g5tRFx1ReX5leURU06xAuf2J/view?usp=sharing