

Decoding Trump's Rhetoric through an NLP Framework

- Donald Trump's speeches during his election campaigns and presidential administration were classified by a model elaborated with Natural Language Processing (NLP) techniques and Machine Learning (ML) models
- The model identified 9 main categories in Trump's semantics: Mexico, Economy, Security, Healthcare, MAGA, Conservative Values, Elections, Administration, and Biden
- Trump's rhetoric shows that negative sentiment towards Mexico has intensified in recent months, in contrast to the previous electoral processes (2016 and 2020)

Classifying Trump's rhetoric. Donald J. Trump's speeches in his electoral campaign processes (2016, 2020, and 2024) and during his presidential administration (2017-2020) maintain an important role within the global political, social, and economic landscape. Due to the unique and controversial style that has characterized him, it is essential to understand the dynamics and evolution of its communication.

Trump's rhetoric analysis is very relevant for Mexico due to the strong economic relationship between both countries. Around 54% of Mexico's economic cycle is explained by the US. In terms of manufacturing exports, Mexico destined 84.5% of its 2023 outflows to the US. In addition, Mexico is the highest recipient of remittances from the US (surpassing India, Philippines, and China). Family remittances represent around 3.5% of Mexico's GDP, 5% of private consumption, and 26.4% of payrolls. Finally, 37.8% of foreign direct investment of 2023 came from the US.

Classifying all Trump's speeches is complex given the quantity available since 2015. In addition, any human interpretation could be subject to a certain subjectivity and polarity due to the existence of behavioral biases that limit its objective evaluation.

This essay describes a methodology that allows the extraction, classification, and interpretation of Trump's speeches published in the different types of media. These were collected from the *White House* website, the "*Certified Website of Donald J. Trump For President 2024*", the "*Right Side Broadcasting Network*" channel, and "*data.millercenter.org*". It is important to note that only the speeches that were part of his electoral campaign and during his presidential administration were selected. The rest were not considered (such as interviews) to avoid possible biases training the model.

A total of 620 *communiqués* were analyzed during the observed period (June 2015 to November 2020 and January 2023 to June 22, 2024). The following diagram shows the process applied:



Alejandro Padilla Santana
Chief Economist and Head of Research
alejandro.padilla@banorte.com



Alejandro Cervantes Llamas
Executive Director of Quantitative Analysis
alejandro.cervantes@banorte.com



José Luis García Casales
Director of Quantitative Analysis
jose.garcia.casales@banorte.com



Daniel Sebastián Sosa Aguilar
Senior Analyst, Quantitative Analysis
daniel.sosa@banorte.com

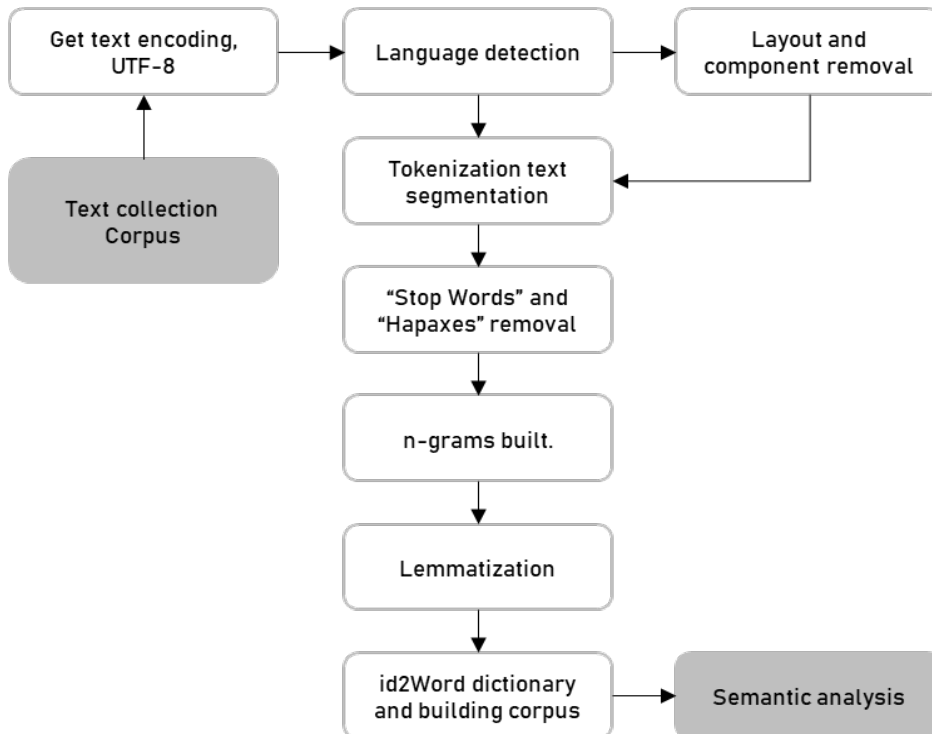


[@analisis_fundam](http://www.banorte.com/analisis-economico)

Winners of the awards as the best economic forecasters in Mexico by LSEG and Focus Economics in 2023



Figure 1: NLP Methodology



Source: Giuseppe Bruno, Text mining and sentiment extraction in Central Bank documents, IEEE International Conference on Big Data, 2016

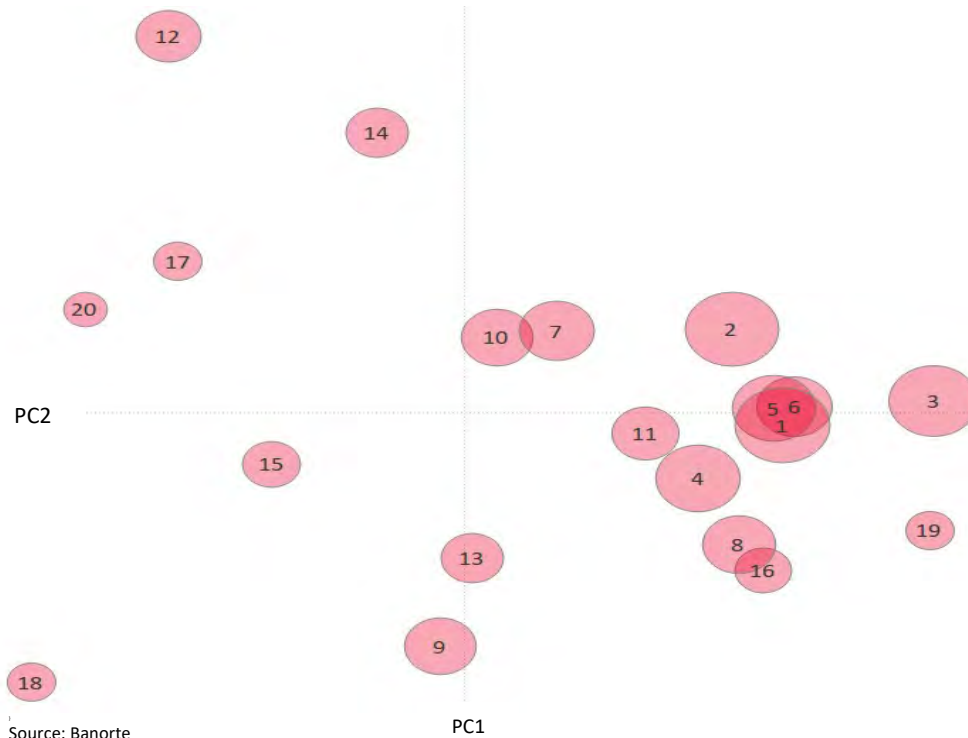
The data engineering implemented starts-off by downloading the speeches and applying a cleaning process in Python. Then, NLP techniques were applied to extract the most valuable information contained in the documents. Next, words that didn't added value (stop words), that appeared only once within the corpus (hapaxes), and that were present in 100% of the documents were eliminated. In addition, n-grams composed of 2 or more tokens (minimum processed text unit) were constructed, given that they could have high relevance in the interpretation of the document, for example: migration crisis, national security, economic growth, etc. Finally, lemmatization techniques were used to reduce the words to their common root, to condense the size of the text and eliminate the unnecessary noise. As a result, a clean corpus was obtained.

It is important to highlight that the above-mentioned cleaning process is extremely important to build the ML classification model. Prior to the model, a bag of words (BoW) was built as the main input for defining a dictionary used during the training process.

The model was estimated using Latent Dirichlet Allocation (LDA). This is a generative model that describes how the documents within a dataset were created. In this context, a dataset is a collection of documents, and a document is a collection of words. Hence, the generative model describes how each document is represented by these words. Initially, it is assumed that there are K topic distributions within the set of documents, meaning that each K of the K multinomial distributions contains V elements, where V is the number of terms in our corpus. The classification model identified the most relevant topics in the corpus built from Trump's communication (refer to Chart 1).

Chart 1: Intertopic Distance Map

Most relevant terms via multidimensional scaling



The bubble chart shows each of the topics identified. The size of the bubble explains the preponderance of the topic, while the dissimilarity that exists in the topics is represented by the distance between each of the bubbles; that is, the greater the distance, the greater the dissimilarity in the topics.

The previous chart is a visual tool that helped to find the optimal number of topics to classify all the speeches. It is important to highlight that the ML model was trained through iterative processes that allowed to identify the optimal hyperparameters of the model to categorize the latent topics within the corpus.

To evaluate the efficiency of the model, two intrinsic metrics were used:

- 1) **Perplexity** allows us to measure how well the model predicts a sample. In the NLP context, it tells us the level of uncertainty that the model has by assigning probabilities to the text; that is, how much entropy the text has (the higher the entropy, the higher the level of uncertainty of the model). The closer this measure is to zero (including negative values), the more relevant the words are. However, the optimization of this measure does not necessarily reflect a better interpretation of the topics, since it is possible to have high perplexity and a null business sense relation.
- 2) **Coherence**. This metric helps us measure the degree of similarity in the semantics of the most relevant words within the topics. There are different coherence measures. In this case we use the “C_v” measure, based on subarrays, segmentation of top words, cosine distances and the use of NMPI (Normalized Mutual Point Information). The higher coherence, the more similar the semantics.

The optimal scenario is low perplexity combined with high coherence. However, a higher level of coherence is more relevant than low perplexity. In the search for the highest coherence, we use a variant of the LDA application, the “LDAMallet” model. The latter uses Gibbs sampling, unlike the standard LDA model which uses variational Bayesian methods. It is important to emphasize that both models require a priori the number of topics in which they will classify our corpus.

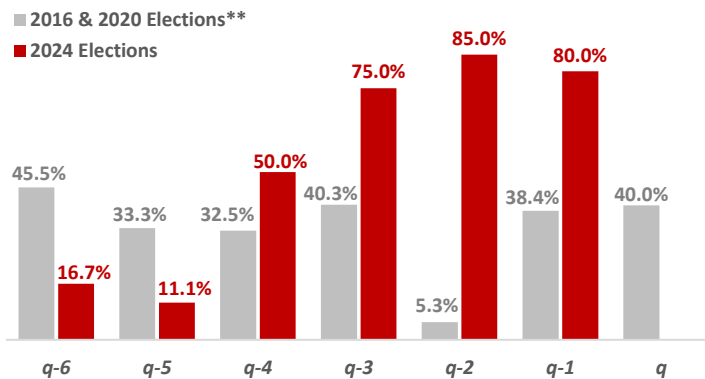
The elbow method was used to establish the optimal number of topics. An interval between 10 and 30 topics was established to measure the performance of the LDAMallet model to analyze their respective coherence. The method showed that the optimal scenario were 20 topics, with an approximate coherence of 0.4. The standard LDA model with the same scenario obtained a 0.36 coherence. Consequently, the LDAMallet model was chosen for the classification.

LDAMallet model identified 20 topics, then establishes the weight that each topic has in the speech and assigns the one with the greatest magnitude. Then, after analyzed the keywords of the topics, the speeches were classified into 9 categories: Mexico, Economy, Security, Public Health, MAGA (“Make America Great Again”), Conservative Values, Elections, Administration and Biden.

Negative rhetoric towards Mexico has intensified. The topics classification for each of the *communiqués* shows that Trump's negative sentiment towards Mexico has strongly intensified between December 2023 and June 2024 (expressed in *q-3*, *q-2*, and *q-1* bars in Chart 2). In contrast, the average negative bias in the 2016 and 2020 electoral processes was substantially lower than the current process (28% of total speeches in 2016 and 2020 vs. 80% in 2024). In addition, the word cloud built on the communication that allude to a negative bias towards Mexico since the beginning of the current electoral campaign supports the negative connotation in the semantics if we consider the relevance of words such as: “border”, “crime”, “cartel”, and “security” (refer to Chart 3).

Chart 2: Trump’s speeches with negative bias about Mexico

% of total speeches made in the quarters previous to the last three electoral processes*.



80%

of Trump’s speeches in June allude to a negative bias towards Mexico

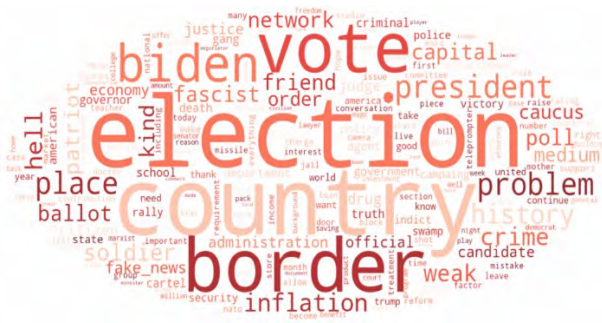
Source: Banorte

* q-x alludes to the previous quarters to the electoral processes. For example, q refers to the months of September, October, and November of the electoral year; q-1 refers to the months of June, July, and August of the same year, and so on.

** average between both electoral processes

Chart 3: Word cloud of the most relevant topics

Most relevant words in speeches made since 2023



Source: Banorte

In addition, it is important to highlight that in the current rhetoric, security issues are associated with a negative bias toward Mexico. Our model identified 4 topics related to the security problem and two of them are associated with Mexico. These has been used frequently in recent speeches. However, security topics that exclude Mexico have not appeared since June 2023. If we consider the strong dependence of the Mexican economy on the U.S. economy, we can consider Trump's greater negative bias towards Mexico as a risk factor.

On March 26, the judge Juan Merchan imposed a gag order that bars Trump from making public statements about the fraud case for which he was convicted. However, Trump has repeatedly violated this order. Despite these restrictions, the negative rhetoric against Mexico has remained constant, given that 80% of Trump's speeches in June continue to have a negative bias towards Mexico as their main topic.

Trump's rhetoric could represent a risk factor for the bilateral economic relationship.

While it is currently impossible to assess who could be the winner of the electoral process that will take place on the 5th November, Trump's negative bias towards Mexico is a risk factor worth considering for the bilateral economic relationship. To analyze the foreign policy that could be implemented and the challenges that could arise ahead in the scenario that Donald Trump is elected to another term, we will continue to monitor the proportion of Trump's speeches that contain a negative connotation towards Mexico in the coming months.

REFERENCES:

Blei, D.M., Ng, A.Y., and Jordan, M.I. (2003) *Latent Dirichlet Allocation*. The Journal of Machine Learning Research, 3:993-1022.

Liu, D., & Lei, L. (2018) The appeal to political sentiment: An analysis of Donald Trump's and Hillary Clinton's speech themes and discourse strategies in the 2016 US presidential election. *Discourse, Context & Media*.

Savoy, J. (2017) Trump's and Clinton's Style and Rhetoric during the 2016 Presidential Election. *Journal of Quantitative Linguistics*, 25(2), 168-189.

Reed, C. (2012) *Latent Dirichlet Allocation: Towards a Deeper Understanding*, Colorado Reed Tutorial, p. 2.

Analyst Certification.

We, Alejandro Padilla Santana, Juan Carlos Alderete Macal, Alejandro Cervantes Llamas, Marissa Garza Ostos, Katia Celina Goya Ostos, Francisco José Flores Serrano, José Luis García Casales, Santiago Leal Singer, Víctor Hugo Cortes Castro, José Itzamna Espitia Hernández, Leslie Thalía Orozco Vélez, Hugo Armando Gómez Solís, Carlos Hernández García, Yazmín Selene Pérez Enríquez, Cintia Gisela Nava Roa, Miguel Alejandro Calvo Domínguez, José De Jesús Ramírez Martínez, Daniel Sebastián Sosa Aguilar, Gerardo Daniel Valle Trujillo, Luis Leopoldo López Salinas, Marcos Saúl García Hernández, Juan Carlos Mercado Garduño, Ana Gabriela Martínez Mosqueda, Jazmín Daniela Cautencos Mora, Andrea Muñoz Sánchez and Paula Lozoya Valadez, certify that the points of view expressed in this document are a faithful reflection of our personal opinion on the company (s) or firm (s) within this report, along with its affiliates and/or securities issued. Moreover, we also state that we have not received, nor receive, or will receive compensation other than that of Grupo Financiero Banorte S.A.B. of C.V for the provision of our services.

Relevant statements.

In accordance with current laws and internal procedures manuals, analysts are allowed to hold long or short positions in shares or securities issued by companies that are listed on the Mexican Stock Exchange and may be the subject of this report; nonetheless, equity analysts have to adhere to certain rules that regulate their participation in the market in order to prevent, among other things, the use of private information for their benefit and to avoid conflicts of interest. Analysts shall refrain from investing and holding transactions with securities or derivative instruments directly or through an intermediary person, with Securities subject to research reports, from 30 calendar days prior to the issuance date of the report in question, and up to 10 calendar days after its distribution date.

Compensation of Analysts.

Analysts' compensation is based on activities and services that are aimed at benefiting the investment clients of Casa de Bolsa Banorte and its subsidiaries. Such compensation is determined based on the general profitability of the Brokerage House and the Financial Group and on the individual performance of each analyst. However, investors should note that analysts do not receive direct payment or compensation for any specific transaction in investment banking or in other business areas.

Last-twelve-month activities of the business areas.

Grupo Financiero Banorte S.A.B. de C.V., through its business areas, provides services that include, among others, those corresponding to investment banking and corporate banking, to a large number of companies in Mexico and abroad. It may have provided, is providing or, in the future, will provide a service such as those mentioned to the companies or firms that are the subject of this report. Casa de Bolsa Banorte or its affiliates receive compensation from such corporations in consideration of the aforementioned services.

Over the course of the last twelve months, Grupo Financiero Banorte S.A.B. C.V., has not obtained compensation for services rendered by the investment bank or by any of its other business areas of the following companies or their subsidiaries, some of which could be analyzed within this report.

Activities of the business areas during the next three months.

Casa de Bolsa Banorte, Grupo Financiero Banorte or its subsidiaries expect to receive or intend to obtain revenue from the services provided by investment banking or any other of its business areas, by issuers or their subsidiaries, some of which could be analyzed in this report.

Securities holdings and other disclosures.

As of the end of last quarter, Grupo Financiero Banorte S.A.B. of C.V. has not held investments, directly or indirectly, in securities or derivative financial instruments, whose underlying securities are the subject of recommendations, representing 1% or more of its investment portfolio of outstanding securities or 1 % of the issuance or underlying of the securities issued.

None of the members of the Board of Grupo Financiero Banorte and Casa de Bolsa Banorte, along general managers and executives of an immediately below level, have any charges in the issuers that may be analyzed in this document.

The Analysts of Grupo Financiero Banorte S.A.B. of C.V. do not maintain direct investments or through an intermediary person, in the securities or derivative instruments object of this analysis report.

Guide for investment recommendations.

| | Reference |
|------|---|
| BUY | When the share expected performance is greater than the MEXBOL estimated performance. |
| HOLD | When the share expected performance is similar to the MEXBOL estimated performance. |
| SELL | When the share expected performance is lower than the MEXBOL estimated performance. |

Even though this document offers a general criterion of investment, we urge readers to seek advice from their own Consultants or Financial Advisors, in order to consider whether any of the values mentioned in this report are in line with their investment goals, risk and financial position.

Determination of Target Prices

For the calculation of estimated target prices for securities, analysts use a combination of methodologies generally accepted among financial analysts, including, but not limited to, multiples analysis, discounted cash flows, sum-of-the-parts or any other method that could be applicable in each specific case according to the current regulation. No guarantee can be given that the target prices calculated for the securities will be achieved by the analysts of Grupo Financiero Banorte S.A.B. C.V, since this depends on a large number of various endogenous and exogenous factors that affect the performance of the issuing company, the environment in which it performs, along with the influence of trends of the stock market, in which it is listed. Moreover, the investor must consider that the price of the securities or instruments can fluctuate against their interest and cause the partial and even total loss of the invested capital.

The information contained hereby has been obtained from sources that we consider to be reliable, but we make no representation as to its accuracy or completeness. The information, estimations and recommendations included in this document are valid as of the issue date, but are subject to modifications and changes without prior notice; Grupo Financiero Banorte S.A.B. of C.V. does not commit to communicate the changes and also to keep the content of this document updated. Grupo Financiero Banorte S.A.B. of C.V. takes no responsibility for any loss arising from the use of this report or its content. This document may not be photocopied, quoted, disclosed, used, or reproduced in whole or in part without prior written authorization from Grupo Financiero Banorte S.A.B. of C.V.



Raquel Vázquez Godínez
Assistant
raquel.vazquez@banorte.com
(55) 1670 - 2967



María Fernanda Vargas Santoyo
Analyst
maria.vargas.santoyo@banorte.com
(55) 1103 - 4000 x 2586

Economic Research



Juan Carlos Alderete Macal, CFA
Executive Director of Economic Research and Market Strategy
juan.alderete.macal@banorte.com
(55) 1103 - 4046



Yazmín Selene Pérez Enríquez
Senior Economist, Mexico
yazmin.perez.enriquez@banorte.com
(55) 5268 - 1694

Market Strategy



Santiago Leal Singer
Director of Market Strategy
santiago.leal@banorte.com
(55) 1670 - 1751



José Itzamna Espitia Hernández
Senior Strategist, Equity
jose.espitia@banorte.com
(55) 1670 - 2249



Leslie Thalía Orozco Vélez
Senior Strategist, Fixed Income and FX
leslie.orozco.velez@banorte.com
(55) 5268 - 1698



Juan Carlos Mercado Garduño
Strategist, Equity
juan.mercado.garduno@banorte.com
(55) 1103 - 4000 x 1746

Quantitative Analysis



Alejandro Cervantes Llamas
Executive Director of Quantitative Analysis
alejandro.cervantes@banorte.com
(55) 1670 - 2972



José De Jesús Ramírez Martínez
Senior Analyst, Quantitative Analysis
jose.ramirez.martinez@banorte.com
(55) 1103 - 4000



Andrea Muñoz Sánchez
Strategist, Quantitative Analysis
andrea.muñoz.sanchez@banorte.com
(55) 1105 - 1430



Alejandro Padilla Santana
Chief Economist and Head of Research
alejandro.padilla@banorte.com
(55) 1103 - 4043



Itzel Martínez Rojas
Analyst
itzel.martinez.rojas@banorte.com
(55) 1670 - 2251



Francisco José Flores Serrano
Director of Economic Research, Mexico
francisco.flores.serrano@banorte.com
(55) 1670 - 2957



Cintia Gisela Nava Roa
Senior Economist, Mexico
cintia.nava.roa@banorte.com
(55) 1105 - 1438



Marissa Garza Ostos
Director of Equity Strategy
marissa.garza@banorte.com
(55) 1670 - 1719



Carlos Hernández García
Senior Strategist, Equity
carlos.hernandez.garcia@banorte.com
(55) 1670 - 2250



Marcos Saúl García Hernández
Analyst, Fixed Income, FX and Commodities
marcos.garcia.hernandez@banorte.com
(55) 1670 - 2296



Ana Gabriela Martínez Mosqueda
Strategist, Equity
ana.martinez.mosqueda@banorte.com
(55) 5261 - 4882



José Luis García Casales
Director of Quantitative Analysis
jose.garcia.casales@banorte.com
(55) 8510 - 4608



Daniel Sebastián Sosa Aguilar
Senior Analyst, Quantitative Analysis
daniel.sosa@banorte.com
(55) 1103 - 4000 x 2124



Lourdes Calvo Fernández
Analyst (Edition)
lourdes.calvo@banorte.com
(55) 1103 - 4000 x 2611



Katia Celina Goya Ostos
Director of Economic Research, Global
katia.goya@banorte.com
(55) 1670 - 1821



Luis Leopoldo López Salinas
Economist, Global
luis.lopez.salinas@banorte.com
(55) 1103 - 4000 x 2707



Víctor Hugo Cortes Castro
Senior Strategist, Technical
victorh.cortes@banorte.com
(55) 1670 - 1800



Hugo Armando Gómez Solís
Senior Analyst, Corporate Debt
hugo.gomez@banorte.com
(55) 1670 - 2247



Gerardo Daniel Valle Trujillo
Analyst, Corporate Debt
gerardo.valle.trujillo@banorte.com
(55) 1670 - 2248



Paula Lozoya Valadez
Analyst, Equity
paula.lozoya.valadez@banorte.com
(55) 1103 - 4000 x 2060



Miguel Alejandro Calvo Domínguez
Senior Analyst, Quantitative Analysis
miguel.calvo@banorte.com
(55) 1670 - 2220



Jazmin Daniela Cuautencos Mora
Strategist, Quantitative Analysis
jazmin.cuautencos.mora@banorte.com
(55) 1670 - 2904