Welcome to the Machine (Learning): An NLP framework for analyzing the Fed's monetary policy statements

- **The following research note describes a natural language processing (NLP) technique that classifies and interprets the central bank's statements**
- **The model classified the semantics of each of the Fed monetary policy statements in 23 preponderant topics**
- **The most important finding was the degree of strength, given that the model can explain 9 out of 10 rate movements. The model also defines the range of variation of the terminal reference rate**
- **Analyzing today's statement, the model suggests a terminal rate between 3.5 and 4.25% (upper range)**

Fed's topics classification. The Federal Open Market Committee (FOMC) statement is the primary tool that the monetary authority uses to communicate its monetary policy. It contains the target range for the federal funds rate, transmits its medium-term perspective on the US economy, and offers insights on the outcome of future votes. However, the interpretation of the Fed's monetary policy statement could be subject to a certain subjectivity due to the existence of behavioral biases that limit an objective interpretation.

The following research note describes a natural language processing (NLP) technique that classifies and interprets the central bank's statements. In addition, it also provides a methodology to confirm whether the communication made by the Central Bank is consistent with the monetary policy implemented.

NLP and Machine Learning. In recent years, machine learning (ML) has evolved to discover common features in verbal and written communication. Quantitative research techniques have gained ground in the application of solutions for the financial industry. Following this trend, the present study uses new methodologies for the analysis and processing of unstructured data.

Text mining is a tool used for extracting a summary from a set of documents. This is commonly referred to as a corpus. Central bank communication usually includes terms or phrases that convey significant messages related to the economy's evolution and can be extracted with the appropriate text mining tools. In the sentiment analysis literature, Natural Language Processing (NLP) methods are used to build dictionaries that can be associated with sentiments.

Unstructured data, such as Fed statements, provide a large amount of information. Unlike traditional methods of macroeconomic analysis, it is necessary to use NLP techniques to extract value from the statements. As a result, the main source of information for this study are the Fed's monetary policy statements from 1978 to 2022.

July 27, 2022

www.banorte.com
@analisis_fundam

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

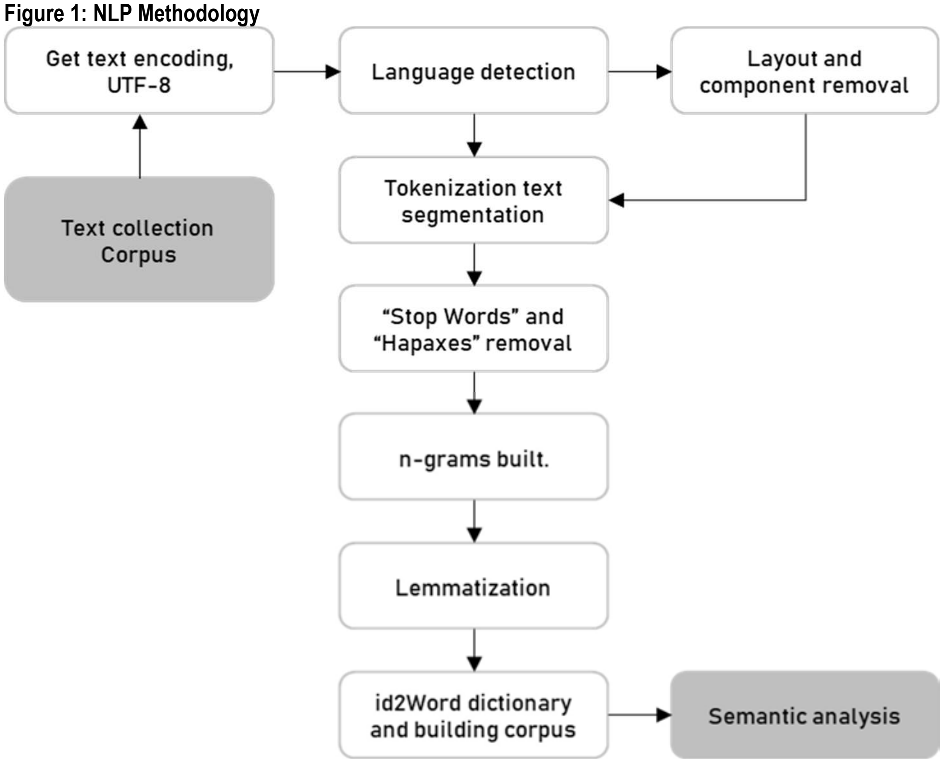
Alejandro Cervantes
Executive Director of Quantitative Research
alejandro.cervantes@banorte.com

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

Daniel Sosa
Analyst, Quantitative Research
daniel.sosa@banorte.com

The NLP model was carried out using Python, since the monetary policy statements are available in PDF and HTML formats. In total, 326 Fed statements were considered for our model (August 1978 to July 2022). The following figure shows the treatment applied to the press releases:

Figure 1: NLP Methodology



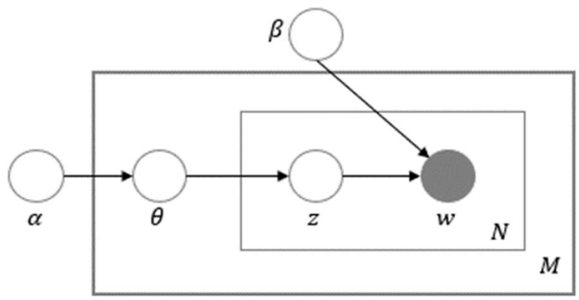
Source: Giuseppe Bruno, Text mining and sentiment extraction in Central Bank documents, IEEE International Conference on Big Data, 2016

The data engineering implemented starts-off by downloading and cleaning the text of the statements. 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: low inflation, high inflation; low unemployment, high unemployment; 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 modeling process.

The model was estimated using Latent Dirichlet Allocation (LDA). These assume that each statement can be represented by a mixture of topics, and each topic is described by a bag of words (BoW), the following figure is a general representation of the LDA model.

Figure 2: Graphic Representation of a LDA Model



For each document $\mathbf{w}$ *in a corpus* $\mathbf{D}$:

1. *Choose* $N \sim Poisson\ (\xi)$.
2. *Choose* $\theta \sim Dir\ (\alpha)$.
3. *For each of the N words* w_n:

   (a) *Choose a topic* $z_n \sim Multinomial(\theta)$.

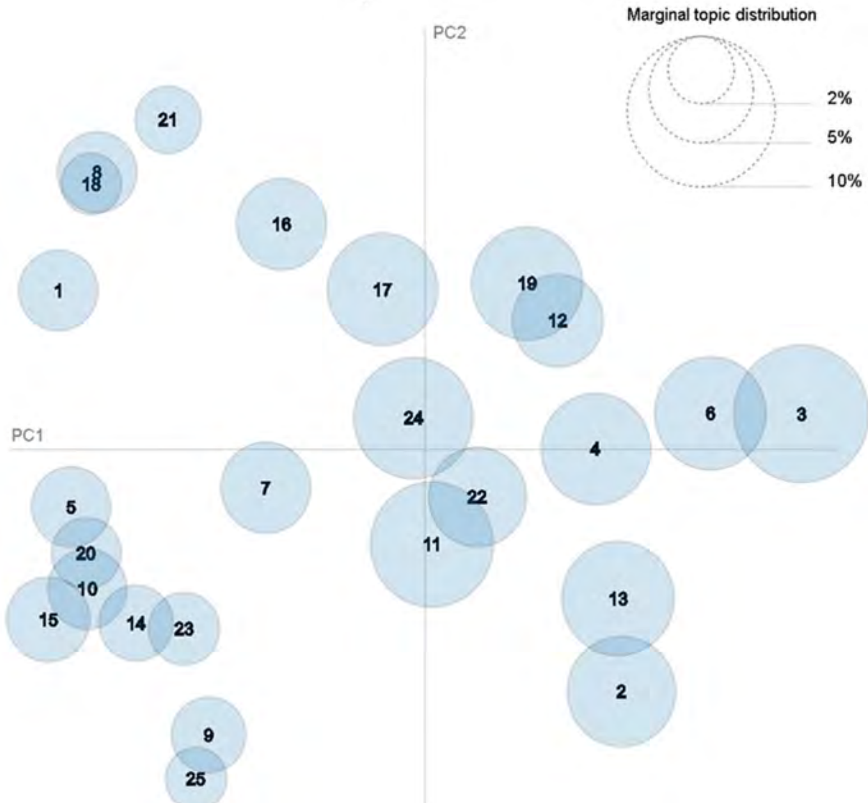
   (b) *Choose a word* w_n *from* $p(w_n|z_n, \beta)$, *a multinomial probability conditioned on the topic* z_n

Source: D.M Blei, A.Y. Ng, and M.I. Jordan. Latent dirichlet allocation. The Journal of Machine Learning Research, 3:993-1022, 2003.

LDA 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 the Federal Reserve Bank statements (refer to Chart 1). The bubble chart shows each of the topics defined in all the monetary policy statements. The size of the bubble explains the preponderance of the topic, while the dissimilarity that exists between the topics is represented by the distance between each of the bubbles; that is, the greater the distance, the greater the dissimilarity between the topics.

Chart 1: Intertopic Distance Map
Most relevant terms via multidimensional scaling



Source: Banorte

The bubble chart is a visual tool that helped find the optimal number of topics to classify all monetary policy statements. 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.

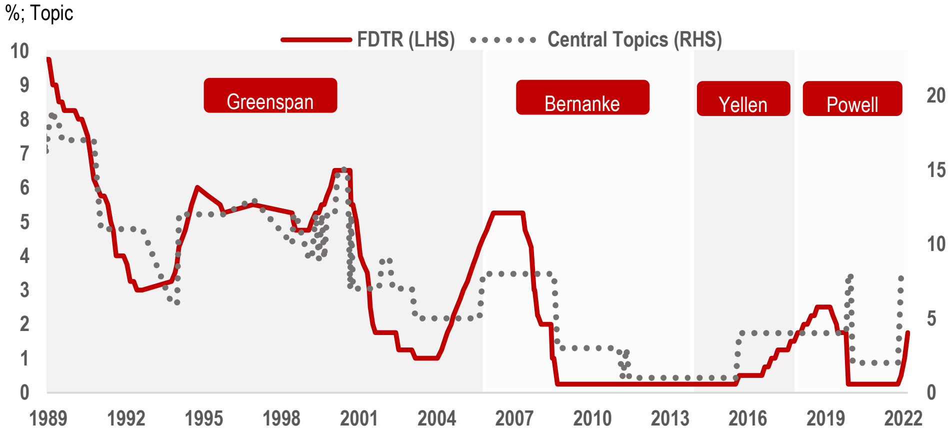
The elbow method was used to establish the optimal number of topics. An interval between 10 and 30 topics was established to measure the performance of the LDAMallet model to analyze their respective coherence. The method showed that the optimal scenario were 25 topics, with an approximate coherence of 0.43. The standard LDA model with the same scenario obtained a 0.39 coherence. Consequently, the LDAMallet model was chosen for the classification.

The LDAMallet model classified each monetary policy statement within the 25 topics. For this classification, the model establishes the weight that each topic has and assigns the one with the greatest magnitude. Even though 25 topics were identified in the corpus, the allocation resulted in 23 relevant topics.

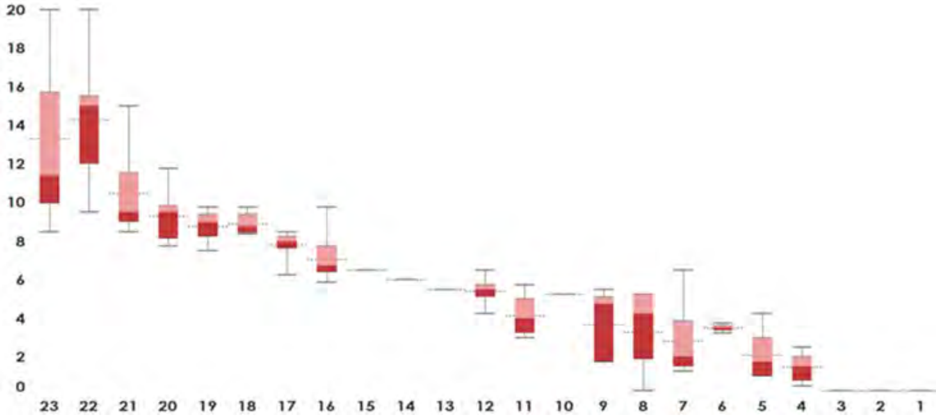
Subsequently, the relationship between the Fed funds rate and the semantic classification obtained was analyzed (refer to Chart 2). **The most important finding was the degree of strength, given that the model can explain 9 out of 10 rate movements. The model also defines the range of variation of the terminal reference rate (refer to Chart 3).**

The model suggests a terminal rate between 3.5 and 4.25% (upper range). After today's monetary policy announcement, the model classified this statement with a topic that anticipates the terminal rate between 3.5% and 4.25% The model also confirms the Fed's congruence in its communication.

Chart 2: Banorte's Fed Topic Index vs Fed Funds Rate
%; Topic



Source: Banorte, Federal Reserve, Bloomberg

Chart 3: Banorte's Fed Topic Index Box Plot
Central Topics


Source: Banorte, Federal Reserve, Bloomberg

References:

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

Federal Reserve Bank of San Francisco. (2021) *Taking the Fed at its Word: A New Approach to Estimating Central Bank Objectives using Text Analysis.* San Francisco: SF FED.

Giuseppe, B. (2016) *Text mining and sentiment extraction in central bank documents*. Washington D.C., IEEE.

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

Siakoulis, V. and Petropoulos, A. (2021) "Can central bank speeches predict financial market turbulence? Evidence from an adaptative NLP sentiment index analysis using XGBoost machine learning technique", *Central Bank Review*, p.p. 141-153.

Analyst Certification

We, Alejandro Padilla Santana, Juan Carlos Alderete Macal, Alejandro Cervantes Llamas, Manuel Jiménez Zaldívar, Marissa Garza Ostos, Katia Celina Goya Ostos, Francisco José Flores Serrano, José Luis García Casales, Víctor Hugo Cortes Castro, José Itzamna Espitia Hernández, Carlos Hernández García, Leslie Thalía Orozco Vélez, Hugo Armando Gómez Solís, Yazmín Selene Pérez Enríquez, Miguel Alejandro Calvo Domínguez, Daniela Olea Suárez, José De Jesús Ramírez Martínez, Gerardo Daniel Valle Trujillo, Luis Leopoldo López Salinas, Isaías Rodríguez Sobrino, Paola Soto Leal, Oscar Rodolfo Olivos Ortiz, Daniel Sebastián Sosa Aguilar and Salvador Austria Valencia 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.

GRUPO FINANCIERO BANORTE S.A.B. de C.V.

Research and Strategy			
Alejandro Padilla Santana	Chief Economist and Head of Research	alejandro.padilla@banorte.com	(55) 1103 - 4043
Raquel Vázquez Godinez	Assistant	raquel.vazquez@banorte.com	(55) 1670 - 2967
Itzel Martínez Rojas	Analyst	itzel.martinez.rojas@banorte.com	(55) 1670 - 2251
Lourdes Calvo Fernández	Analyst (Edition)	lourdes.calvo@banorte.com	(55) 1103 - 4000 x 2611
María Fernanda Vargas Santoyo	Analyst	maria.vargas.santoyo@banorte.com	(55) 1103 - 4000
Economic Research			
Juan Carlos Alderete Macal, CFA	Executive Director of Economic Research and Financial Markets Strategy	juan.alderete.macal@banorte.com	(55) 1103 - 4046
Francisco José Flores Serrano	Director of Economic Research, Mexico	francisco.flores.serrano@banorte.com	(55) 1670 - 2957
Katia Celina Goya Ostos	Director of Economic Research, Global	katia.goya@banorte.com	(55) 1670 - 1821
Yazmín Selene Pérez Enríquez	Senior Economist, Mexico	yazmin.perez.enriquez@banorte.com	(55) 5268 - 1694
Luis Leopoldo López Salinas	Manager Global Economist	luis.lopez.salinas@banorte.com	(55) 1103 - 4000 x 2707
Market Strategy			
Manuel Jiménez Zaldívar	Director of Market Strategy	manuel.jimenez@banorte.com	(55) 5268 - 1671
Fixed income and FX Strategy			
Leslie Thalía Orozco Vélez	Senior Strategist, Fixed Income and FX	leslie.orozco.velez@banorte.com	(55) 5268 - 1698
Isaías Rodríguez Sobrino	Strategist, Fixed Income, FX and Commodities	isaias.rodriguez.sobrino@banorte.com	(55) 1670 - 2144
Equity Strategy			
Marissa Garza Ostos	Director of Equity Strategy	marissa.garza@banorte.com	(55) 1670 - 1719
José Itzamna Espitia Hernández	Senior Strategist, Equity	jose.espitia@banorte.com	(55) 1670 - 2249
Carlos Hernández García	Senior Strategist, Equity	carlos.hernandez.garcia@banorte.com	(55) 1670 - 2250
Víctor Hugo Cortes Castro	Senior Strategist, Technical	victorh.cortes@banorte.com	(55) 1670 - 1800
Paola Soto Leal	Analyst	paola.soto.leal@banorte.com	(55) 1103 - 4000 x 1746
Oscar Rodolfo Olivos Ortiz	Analyst	oscar.olivos@banorte.com	(55) 1103 - 4000
Corporate Debt			
Hugo Armando Gómez Solís	Senior Analyst, Corporate Debt	hugoa.gomez@banorte.com	(55) 1670 - 2247
Gerardo Daniel Valle Trujillo	Analyst, Corporate Debt	gerardo.valle.trujillo@banorte.com	(55) 1670 - 2248
Quantitative Research			
Alejandro Cervantes Llamas	Executive Director of Quantitative Research	alejandro.cervantes@banorte.com	(55) 1670 - 2972
José Luis García Casales	Director of Quantitative Research	jose.garcia.casales@banorte.com	(55) 8510 - 4608
Daniela Olea Suárez	Senior Analyst, Quantitative Research	daniela.olea.suarez@banorte.com	55) 1103 - 4000
Miguel Alejandro Calvo Domínguez	Senior Analyst, Quantitative Research	miguel.calvo@banorte.com	(55) 1670 - 2220
José De Jesús Ramírez Martínez	Senior Analyst, Quantitative Research	jose.ramirez.martinez@banorte.com	(55) 1103 - 4000
Daniel Sebastián Sosa Aguilar	Analyst, Quantitative Research	daniel.sosa@banorte.com	(55) 1103 - 4000
Salvador Austria Valencia	Analyst, Quantitative Research	salvador.austria.valencia@banorte.com	(55) 1103 - 4000
Wholesale Banking			
Armando Rodal Espinosa	Head of Wholesale Banking	armando.rodal@banorte.com	(55) 1670 - 1889
Alejandro Aguilar Ceballos	Head of Asset Management	alejandro.aguilar.ceballos@banorte.com	(55) 5004 - 1282
Alejandro Eric Faesi Puente	Head of Global Markets and Institutional Sales	alejandro.faesi@banorte.com	(55) 5268 - 1640
Alejandro Frigolet Vázquez Vela	Head of Sólida Banorte	alejandro.frigolet.vazquezvela@banorte.com	(55) 5268 - 1656
Arturo Monroy Ballesteros	Head of Investment Banking and Structured Finance	arturo.monroy.ballesteros@banorte.com	(55) 5004 - 5140
Carlos Alberto Arciniega Navarro	Head of Treasury Services	carlos.arciniega@banorte.com	(81) 1103 - 4091
Gerardo Zamora Nanez	Head of Transactional Banking, Leasing and Factoring	gerardo.zamora@banorte.com	(81) 8173 - 9127
Jorge de la Vega Grajales	Head of Government Banking	jorge.delavega@banorte.com	(55) 5004 - 5121
Luis Pietrini Sheridan	Head of Private Banking	luis.pietrini@banorte.com	(55) 5249 - 6423
Lizza Velarde Torres	Executive Director of Wholesale Banking	lizza.velarde@banorte.com	(55) 4433 - 4676
Osvaldo Brondo Menchaca	Head of Specialized Banking Services	osvaldo.brondo@banorte.com	(55) 5004 - 1423
Raúl Alejandro Arauzo Romero	Head of Transactional Banking	alejandro.arauzo@banorte.com	(55) 5261 - 4910
René Gerardo Pimentel Ibarrola	Head of Corporate Banking	pimentelr@banorte.com	(55) 5004 - 1051
Ricardo Velázquez Rodríguez	Head of International Banking	rvelazquez@banorte.com	(55) 5004 - 5279
Víctor Antonio Roldan Ferrer	Head of Commercial Banking	victor.roldan.ferrer@banorte.com	(55) 1670 - 1899