### **Project Ask:**

### **An Analysis of Colombia's Macroeconomic Landscape (ASK)**

This document outlines the foundational "Ask" phase for an in-depth analysis of the Colombian economy. The objective is to leverage a curated set of economic indicators from FRED to understand the interplay of key financial and economic forces that have shaped Colombia in recent decades.

#### **1. The Central Question**

Given the provided datasets, the central problem we aim to solve is:

**"How have external economic factors (like trade and currency fluctuations) and internal indicators (such as inflation, interest rates, and the housing market) collectively influenced Colombia's Gross Domestic Product (GDP) and overall economic trajectory?"**

This overarching question allows us to explore the dynamics between key drivers of economic performance and diagnose the nation's economic health.

#### **2. Stakeholder Goals and Expectations**

For this analysis, we will consider two primary stakeholder groups: **Economic Policymakers** (e.g., from the Ministry of Finance or Central Bank of Colombia) and **Foreign Investors**.

**Stakeholder Goals:**

* **Economic Policymakers:** Their primary goal is to ensure economic stability and foster sustainable growth. They would expect this analysis to reveal the primary drivers of inflation, the sensitivity of GDP to changes in interest rates and export volumes, and the overall resilience of the domestic economy.
* **Foreign Investors:** Their main goal is to assess risk and identify opportunities for return on investment. They would expect the analysis to provide clarity on currency stability (USD to COP exchange rate), the risk/return profile of government debt (10-year bond yields), and the potential growth of the domestic market, as reflected by GDP and the property market.

**Expectations for the Analysis:**

* **Clarity and Insight:** Both groups expect a clear, data-driven narrative that moves beyond simple data presentation to offer actionable insights.
* **Correlation Analysis:** An exploration of the relationships between the key variables. For example, how have changes in the USD exchange rate historically impacted the value of Colombian exports? How do interest rates correlate with residential property prices and inflation?
* **Trend Identification:** A clear visualization and explanation of the long-term trends for each economic indicator.

#### **3. Defining a Successful Outcome**

A successful outcome for this project will be a comprehensive analytical report that includes:

* **An Executive Summary:** A high-level overview of the key findings and insights tailored to both policymakers and investors.
* **A Macroeconomic Dashboard:** A series of clear, well-annotated visualizations charting the historical trends of all the key economic indicators.
* **An Analytical Narrative:** A detailed chapter explaining the observed trends and the relationships between the datasets. It will directly address the central question by telling the story of the Colombian economy as seen through the provided data.
* **Actionable Insights:** The report will conclude with specific, data-backed insights.
  + For Policymakers: Insights into which economic levers have historically been most effective in managing inflation and stimulating growth.
  + For Investors: A risk/reward summary based on historical volatility and performance of the included assets and indicators.

Success is defined not just by the completion of the analysis, but by its ability to provide a clear, compelling, and actionable economic story that empowers stakeholders to make more informed decisions.

#### **4. Required Data**

To effectively address the central question and meet stakeholder expectations, the analysis will be strictly limited to the following datasets provided:

1. **Gross Domestic Product for Colombia** (MKTGDPCOA646NWDB)
2. **Consumer Price Index Total for Colombia** (COLCPALTT01IXOBM)
3. **Consumer Price Index: Housing, Water, Electricity, Gas and Other Fuels** (COLCP040000IXNBQ)
4. **Residential Property Prices for Colombia** (QCON628BIS)
5. **Interest Rates, Long-Term Government Bond Yields: 10-Year** (COLIRLTLT01STM)
6. **Currency Conversions: US Dollar Exchange Rate for Colombia** (COLCCUSMA02STM)
7. **International Merchandise Trade Statistics: Exports** (COLXTEXVA01CXMLSAM)

### **Project Prepare:**

### **Data Sourcing, Handling, and Limitations (PREPARE)**

This document outlines the "Prepare" phase of our analysis of the Colombian economy. In this stage, we identify the specific datasets, their sources, our methods for handling them, and, most importantly, the inherent limitations of the data that will inform the subsequent analysis.

#### **1. Specific Data Required**

To address the central question defined in the "Ask" phase, we will utilize the following specific time-series datasets:

* **Gross Domestic Product:** MKTGDPCOA646NWDB
* **Total Consumer Price Index:** COLCPALTT01IXOBM
* **CPI for Housing, Water, Electricity, Gas & Other Fuels:** COLCP040000IXNBQ
* **Residential Property Prices:** QCON628BIS
* **10-Year Government Bond Yields:** COLIRLTLT01STM
* **USD to National Currency Exchange Rate:** COLCCUSMA02STM
* **Exports Commodities:** COLXTEXVA01CXMLSAM

#### **2. Data Source**

All datasets for this analysis have been sourced from the **Federal Reserve Economic Data (FRED)** database, a public repository maintained by the Federal Reserve Bank of St. Louis. FRED is a highly reputable source for national and international economic data.

#### **3. Data Collection and Storage**

* **Collection:** The data has been pre-collected and made available for this project in the form of individual .csv files for each time series.
* **Storage and Security:** The data is currently stored locally. As this is public, non-confidential data, there are no stringent data privacy or security concerns (e.g., Personally Identifiable Information).
* **Ethical Considerations:** The primary ethical obligation is to use the data transparently. This involves acknowledging the source (FRED), being clear about any transformations applied to the data, and honestly reporting the data's limitations to ensure the analysis is not misleading.

#### **4. Data Limitations**

A thorough initial analysis of the provided data reveals several critical limitations that must be addressed before and during the analysis phase.

* **Mismatched Frequencies:** The datasets are recorded at different time intervals.
  + **Annual:** Gross Domestic Product (MKTGDPCOA646NWDB)
  + **Quarterly:** CPI for Housing (COLCP040000IXNBQ), Residential Property Prices (QCON628BIS)
  + **Monthly:** Total CPI (COLCPALTT01IXOBM), Exchange Rate (COLCCUSMA02STM), Interest Rates (COLIRLTLT01STM), Exports (COLXTEXVA01CXMLSAM)

This mismatch is the most significant analytical challenge. To compare these series, we must standardize them to a single frequency (likely quarterly or annually). This process, known as resampling, can involve averaging or imputation, which can smooth out volatile data or introduce artifacts. The chosen method will be clearly documented.

* **Varying Time Periods:** The datasets cover different historical time spans. For instance, the exchange rate data begins in 1957, while the CPI for Housing data only starts in 2009. A comprehensive, multi-variable analysis will be restricted to the overlapping time frame, which begins in **Q1 2009**. Any analysis of longer-term trends for individual variables must be done in isolation.
* **Absence of Missing Data:** A positive finding is that none of the datasets contain missing or null values within their reported time frames. This simplifies the cleaning process considerably.
* **Data Definitions:** The analysis is constrained by the precise definitions used by the sourcing institution for each indicator. For example, "Residential Property Prices" is a nominal index, and understanding its composition is key to interpreting its movements correctly. Our analysis will rely on the standardized FRED definitions for each series.

***Answering Your Key Questions (SQL Version) (Process)***

***1. What errors, inaccuracies, or missing values are present in the data?*** The initial analysis reveals that the datasets are generally clean and do not suffer from incorrect entries or random missing values within their reported timelines. The primary challenge is not one of errors, but of structure:

* **Inconsistent Frequencies:** The data is recorded at different intervals: Gross Domestic Product is annual, some housing and property indices are quarterly, and the remaining indicators are monthly.
* **Divergent Timelines:** Each economic indicator's data begins and ends in different years. A direct comparison is impossible without first aligning these timelines.

***2. How can the data be cleaned and transformed to make it usable?*** The transformation strategy in SQL is to standardize all seven datasets to a single, common frequency. The chosen frequency would be **quarterly**. This involves:

* **Aggregation for Monthly Data:** For indicators like interest rates, exchange rates, and the total CPI, the three monthly records within each quarter would be grouped together and mathematically averaged. This provides a single, representative value for the quarter. For a cumulative value like exports, the three monthly figures would be summed to create a quarterly total.
* **Propagation for Annual Data:** For the annual GDP data, the single value for a given year would be assigned to each of the four quarters within that year. This allows the annual figure to be compared alongside the quarterly indicators.
* **Alignment of Quarterly Data:** The datasets that are already quarterly are simply aligned to a standard quarter-start date (e.g., January 1st, April 1st) to ensure they can be joined perfectly with the newly transformed datasets.

***3. What tools and techniques will be applied for data cleaning?*** The entire process would be conducted within a **SQL Database** (such as SQL Server, PostgreSQL, etc.). The specific techniques are:

* **Structured Table Creation:** First, defining a rigid table structure for each dataset. This enforces the correct data types (e.g., ensuring dates are stored as DATE and values as FLOAT or DECIMAL), which prevents data-type errors.
* **Date and Grouping Functions:** Using built-in SQL functions to extract the YEAR and QUARTER from each observation date. These are used in a GROUP BY clause, which is the core technique for collecting all the monthly records of a quarter together for aggregation.
* **Aggregate Functions:** Employing mathematical functions like AVG() (average) and SUM() to perform the calculations on the groups created by the GROUP BY clause.
* **Common Table Expressions (CTEs):** To keep the process organized and readable, the transformation for each dataset would be performed inside its own CTE. This technique breaks the complex problem into a series of logical, sequential steps.
* **Join Operations:** Using JOIN commands to combine the results from all the individual CTEs into a final, master table. The join is performed by matching the standardized quarterly dates, uniting all the different economic indicators onto a single timeline.

***4. How can the integrity and quality of the cleaned data be maintained?*** The SQL approach ensures data integrity and produces a high-quality result through:

* **Reproducibility:** The entire transformation is contained within a single SQL script. This script acts as perfect documentation and can be re-run at any time to produce the exact same result, ensuring the process is transparent and repeatable.
* **Structural Integrity:** By first loading the data into strongly-typed tables, the database engine itself prevents the introduction of improperly formatted data.
* **Honest Handling of Missing Data:** By using a LEFT JOIN strategy to build the final table, we ensure that the complete timeline is preserved, even for periods where some indicators don't have data. Instead of dropping these periods, SQL will place a NULL value in the cell, which is the correct and honest way to represent a known missing observation. This prevents data loss and provides a clear picture of the data's actual limitations.

**What patterns or trends are evident in the data?**

There are two main stories happening at the same time:

* **A Steady Climb in Costs:** Three key indicators have consistently increased over the last decade: the general cost of living (inflation), the price of housing, and the exchange rate (meaning it takes more Colombian Pesos to buy one US Dollar).
* **Unstable Economic Growth:** The nation's total economic output (GDP) and its income from selling goods to other countries (Exports) are very closely linked. They rose to a peak around 2014 but have been much more volatile and have struggled to regain those levels since.

**How do these findings relate to the problem at hand?**

These findings show a clear connection between internal and external economic forces:

* The Colombian economy's health is strongly tied to its **export performance**. When exports rise, the national economy grows.
* At the same time, the **weakening currency** (a higher exchange rate) is directly linked to a higher cost of living and more expensive housing, which presents a major challenge for the financial stability of households and businesses.

**What statistical methods could be used to test the hypotheses?**

To be more certain about these connections, the next step would be to use more advanced calculations. For example, we could use a method called **regression analysis** to measure exactly how much a change in exports impacts GDP. Other tests could help determine if a change in one area, like interest rates, can reliably predict a future change in another, like inflation.

**What new questions emerge from the analysis?**

This analysis leads to several important new questions:

1. What specific event caused the sharp downturn in both exports and the overall economy around 2014?
2. If a weaker currency is often thought to help exporters, why hasn't it led to stronger, more sustained economic growth in recent years?
3. How has the steep rise in the cost of living and housing truly affected the financial well-being of the average Colombian family?

**Advanced Analysis: Regression and Causality**

The goal here is to go beyond just observing trends and correlations. We want to **measure the impact** of different economic factors on GDP and **test if one factor can predict another**. We will use two powerful statistical methods for this.

**Part 1: Regression Analysis - Measuring What Drives GDP**

A regression model helps us measure the specific impact of a set of independent variables (like exports and interest rates) on a single dependent variable (GDP).

**The Goal:** To answer the question: "By how much does a change in exports, interest rates, or the exchange rate affect Colombia's GDP?"

**How to Understand the Results:**

The output table gives us three key pieces of information:

1. **R-squared (0.792):** This is the most important measure of the model's overall success. It means that **79.2% of the changes in Colombia's GDP can be explained** by the changes in exports, interest rates, and the exchange rate combined. This is a very strong fit.
2. **Coefficients (coef):** This tells us the specific impact of each variable.
   * **Exports (15.1807):** For every **$1 increase in exports**, the model predicts that **GDP will increase by about $15.18**. This shows the powerful multiplying effect of exports on the economy.
   * **Interest Rates (-7.339e+09):** For every **1-point increase in the 10-year bond yield**, GDP is predicted to **decrease by about $7.34 billion**. This suggests higher interest rates tend to slow down the economy.
   * **Exchange Rate (1.529e+07):** For every **1-peso increase in the USD/COP exchange rate** (a weaker peso), GDP is predicted to **increase by about $15.3 million**.
3. **P-values (P>|t|):** This tells us if the results are statistically significant. A value less than 0.05 is considered significant. Here, all three variables have P-values of 0.000, meaning their impact on GDP is **highly statistically significant**.

**Part 2: Granger Causality - Testing for Predictability**

This test helps us check if the past values of one variable are useful for predicting the future values of another. It's not "causation" in the everyday sense, but rather a test of predictive power.

**The Goal:** To answer the question: "Do past changes in interest rates help us predict the direction of future inflation?"

**How to Understand the Results:**

The output shows the test results for different time lags (from 1 to 4 quarters in the past). We look at the P-Value to see if the relationship is significant.

* At a lag of 1 and 2 quarters, the P-Value is greater than 0.05, so there is **no significant** predictive relationship.
* However, at a lag of **3 and 4 quarters**, the P-Values are 0.0225 and 0.0000, which are both less than 0.05.

This means that changes in interest rates **do not immediately predict** inflation in the next quarter or two. But they **are statistically significant for predicting inflation 3 to 4 quarters (9-12 months) into the future**. This makes economic sense, as it often takes time for changes in monetary policy (like interest rates) to work their way through the economy and affect prices.

**The Story of the Analysis Phase**

The goal of the analysis phase was to take the clean, organized dataset and make it tell the story of the Colombian economy. We did this in three main steps, each building on the last.

**Step 1: First Impressions - Visualizing the Economic Trends**

First, we created charts for each economic indicator to see how they behaved over time since 2009. This visual check immediately revealed a two-part story:

* **A Story of Rising Costs:** We saw a clear, relentless upward trend in the cost of living. The lines for general inflation (CPI), housing prices, and the U.S. dollar exchange rate all moved steadily upwards, showing that life has become progressively more expensive.
* **A Story of Volatile Growth:** In contrast, the lines for the country's main growth engines—GDP and Exports—were much more unstable. They looked like a rollercoaster, rising to a high point around 2014 and then struggling in the years that followed. This immediately told us that the economy's growth has not been smooth.

This first step gave us the basic plot points. The next step was to understand how these plot points were connected.

**Step 2: Connecting the Dots - Uncovering Relationships**

Next, we created a correlation matrix, which is like a relationship map for our data. It showed us exactly how strongly each economic factor was connected to the others. The key findings were:

* **The Engine of the Economy:** We confirmed a very strong positive connection (+0.84) between **GDP and Exports**. This was a critical insight: for all practical purposes, the health of the Colombian economy is directly tied to the success of its export sector.
* **The "Cost of Living" Cluster:** We found that **inflation, housing prices, and the exchange rate** were all extremely tightly linked (correlations above +0.9). They all move together in lockstep. This means that when the peso weakens against the dollar, it's almost always accompanied by a rise in prices for goods and housing.

This step helped us understand the "why" behind our initial observations. We now knew *which* variables were moving together.

**Step 3: The Deeper Dive - Measuring and Predicting**

Finally, we used more advanced statistical models to measure the precise impact of these relationships and to test for predictive power.

* **Measuring the Drivers of GDP (Regression):** We built a model that explained **79.2%** of all the changes in GDP—a very high score. It gave us specific measurements:
  + **Exports** had a powerful multiplier effect, adding about **$15** to the economy for every **$1** earned from trade.
  + **Higher Interest Rates** acted as a brake, slowing the economy down.
  + A **weaker currency** provided a small, but statistically significant, net boost to GDP.
* **Testing for Predictability (Causality Test):** We tested if changes in interest rates could predict future inflation. The test revealed that they could, but not immediately. It showed that changes in interest rates have a significant power to predict the direction of inflation about **9 to 12 months** down the line, confirming the well-known time lag of monetary policy.

**Conclusion: The Complete Economic Narrative**

By combining these three steps, we moved from raw data to a comprehensive story. The analysis showed that the Colombian economy is best understood as being heavily reliant on a volatile export sector for its growth. This external dependency creates internal challenges, where a weakening currency—while offering a minor boost to GDP—is strongly linked to a rising cost of living for its citizens. Finally, we learned that the government's attempts to manage this through interest rates are effective, but their impact on prices is significantly delayed.

Phase 5: Share - From Data to Decisions

**Key Insights and Recommendations**

This analysis yields three critical insights that can guide strategic decisions:

1. **Insight: The Export Engine is Over-revved.** The economy's heavy dependence on exports makes it vulnerable to global shocks (e.g., commodity price crashes like the one around 2014). This is the single biggest source of economic volatility.
   * **Recommendation (Policymakers):** Prioritize strategic initiatives aimed at diversifying the economy's growth drivers. Fostering domestic innovation, technology, and service sectors can build resilience against fluctuations in global trade.
2. **Insight: The Currency-Inflation Feedback Loop.** A weakening peso is not a benign event; it is directly associated with a higher cost of living and a hotter housing market.
   * **Recommendation (Policymakers):** Develop targeted policies to shield the most vulnerable populations from the inflationary effects of currency depreciation, such as social safety nets or subsidies for essential goods.
   * **Recommendation (Investors):** Currency risk must be a primary consideration in any investment thesis. Hedging strategies are crucial. Look for companies with strong pricing power that can pass on inflation to consumers.
3. **Insight: Monetary Policy is a Long-Term Tool.** Our analysis shows that changes in interest rates can take nearly a year to meaningfully impact inflation.
   * **Recommendation (Policymakers):** Maintain a consistent, long-term monetary policy strategy. Avoid reactive, short-term changes, and clearly communicate the expected delay in policy impact to manage public and market expectations.
   * **Recommendation (Investors):** Pay close attention to the Central Bank's forward guidance. Anticipate that policy shifts will not have an immediate effect, and position portfolios for the economic environment 9-12 months in the future.

This structured presentation turns our complex data analysis into a clear, persuasive, and actionable narrative, empowering stakeholders to make informed decisions.

**Phase 6: Act - Transforming Insights into Action**

Based on our comprehensive analysis of the Colombian economy, here are the recommended actions, their anticipated impact, and a framework for measuring success.

**What specific actions are recommended based on the findings?**

The recommendations are tailored to our two key stakeholder groups:

**For Economic Policymakers:**

1. **Launch an Economic Diversification Initiative:** The analysis confirmed that Colombia's GDP is critically dependent on a volatile export sector.
   * **Action:** Create a national strategy with targeted incentives (tax credits, grants) to accelerate growth in non-traditional sectors like technology, business services, and sustainable tourism.
2. **Implement an "Inflation Shield" Program:** The data shows a direct link between currency devaluation and the rising cost of living.
   * **Action:** Design policies to cushion the impact on the population, such as temporary subsidies on essential imported goods or expanding social safety nets during periods of high inflation.
3. **Reinforce Forward-Looking Monetary Policy Communication:** Our analysis proved that interest rate changes take 9-12 months to significantly impact inflation.
   * **Action:** The Central Bank should proactively and repeatedly communicate this time lag to the public and financial markets to manage expectations and build confidence in its long-term strategy.

**For Foreign Investors:**

1. **Mandate Currency Risk Hedging:** The single greatest financial risk identified is the steady devaluation of the Colombian Peso.
   * **Action:** Incorporate currency hedging (e.g., using financial instruments like forwards or options) as a standard, non-negotiable component of any investment strategy in Colombia.
2. **Adopt a Dual-Focus Sector Analysis:** The data shows a disconnect between export growth and domestic consumer health.
   * **Action:** When evaluating opportunities, focus due diligence not only on a company's export potential but also on its vulnerability to a decline in domestic consumer purchasing power.
3. **Prioritize Central Bank Guidance:** The causality analysis showed that interest rate movements have a predictable, albeit delayed, effect.
   * **Action:** Treat the Central Bank's official communications and meeting minutes as a primary source of intelligence to anticipate future market conditions and position portfolios accordingly.

**How can stakeholders apply these insights to improve decision-making?**

* **Policymakers** can use the regression model's finding (that exports have a ~$15 multiplier effect on GDP) to justify budget allocations for trade infrastructure. They can use the 9-12 month causality result to time their policy interventions and defend their long-term approach against short-term political pressure.
* **Investors** can use the correlation data to build more resilient financial models that explicitly price in currency risk. They can shift from a purely growth-focused strategy to a more nuanced, risk-adjusted approach that balances opportunities with the clear macroeconomic challenges.

**What is the anticipated impact of these actions?**

* For **policymakers**, the anticipated impact is a **more stable and resilient economy** less prone to boom-and-bust cycles caused by global commodity prices. This would lead to more sustainable long-term GDP growth and increased public confidence.
* For **investors**, the impact will be **improved risk-adjusted returns**. By actively managing the primary risk (currency) and focusing on resilient sectors, investors can reduce portfolio volatility and increase the likelihood of success.

**How can the success of these actions be measured and monitored?**

Success must be measured with clear Key Performance Indicators (KPIs).

1. **To Measure Economic Diversification:** Track the **percentage of GDP contributed by non-traditional sectors**. Success would be a measurable increase in this KPI over a 3-5 year period.
2. **To Measure Inflation Shielding:** Monitor **household consumption data and consumer confidence indices** on a quarterly basis. Success would be seeing these indicators remain stable during periods of high currency volatility.
3. **To Measure Investment Strategy Success:** Track the **Sharpe ratio** (risk-adjusted return) of the Colombia-focused portfolio. Success would be a higher Sharpe ratio compared to an unhedged benchmark.