New DS Challenge

Our client has reached out for assistance in developing a comprehensive data analysis and modeling strategy to enhance decision-making in the oil and gas sector. They are particularly interested in leveraging data to optimize pricing strategies and gain a competitive edge.

I would like to request your assistance with the initial phase of our project. Our focus will be on the following key areas:

1. **Data Cleaning, Processing, and Consolidation:**
    - Identify and handle missing or outlier values.
    - Explore the data and document findings.
2. **Exploratory Data Analysis:**
    - Visualize and describe key variables.
    - Identify associations between related variables.
    - Explore potential trends, including short- and long-term patterns.
    - Generate additional variables that might be useful for analysis, such as monthly price volatility.
3. **Modeling:**
    - Select and describe an appropriate modeling approach.
    - Develop the model to predict `avg_preco_venda` for the months of June, July, and August 2024, and assess its performance using relevant metrics.
4. **Insights/Actionable Decisions:**
    - Propose strategies to leverage identified opportunities.

I've coordinated with the Research & Data Services team to gather relevant datasets.

Time is of the essence, so I recommend we divide the work efficiently. Your expertise and collaboration will be crucial to the success of this initiative. Thank

you all for your dedication and hard work. I look forward to achieving great results together!

Dataset Description:

This dataset is designed to support the prediction of `avg_preco_venda` (average selling price) in the oil and gas sector. Below are the key variables included:

1. **bandeira_revenda_cat**: Categorical. Represents the brand category of the reseller.
2. **data**: Date. This column contains the date of each transaction or observation.
3. **sigla_uf**: Categorical. Abbreviation of the Brazilian state.
4. **sigla_uf_nome**: Categorical. Full name of the state.
5. **id_municipio**: Categorical. Unique identifier for the municipality.
6. **id_municipio_nome**: Categorical. Municipality name.
7. **produto**: Categorical. Type of product sold.
8. **unidade_medida**: Categorical. Unit of measurement for the product.
9. **n_estabelecimentos**: Numeric. Number of reporting establishments.
10. **avg_preco_compra**: Numeric. Average purchase price.
11. **avg_preco_venda**: Numeric. Average selling price, which is the target variable for prediction.