**BUSINESS UNDERSTANDING**

**PROBLEM STATEMENT:**

* + EABL, a leading East African brewery company, has been experiencing stock price volatility and stagnation over the past 10 years, impacting shareholder confidence and investment stability.
  + Recent regulatory actions, such as the Kenya Revenue Authority's imposition of a policy requiring EABL to file returns within 24 hours, further exacerbate challenges faced by the company.
  + The primary challenge is to devise data-driven strategies to stabilize EABL's stock performance, mitigate the impact of regulatory policies, and foster exponential growth in shareholder value.

**BUSINESS OBJECTIVES**:

* + Maintain investor confidence and prevent shareholder withdrawal amidst stock price volatility and regulatory pressures.
  + Stabilize EABL's stock performance and foster sustainable growth in stock value over time.
  + Enhance EABL's resilience to regulatory changes and policy disruptions while optimizing financial outcomes.

**CONTEXT**:

* + EABL operates in the highly competitive and dynamic East African brewery industry, facing challenges such as changing consumer preferences, regulatory scrutiny, and economic fluctuations.
  + The company's success is closely tied to its brand reputation, product quality, market positioning, and ability to adapt to evolving market conditions.
  + Regulatory interventions, such as the Kenya Revenue Authority's policy, can significantly impact EABL's operational efficiency, financial reporting processes, and investor perception.

In summary, EABL seeks data-driven solutions to address stock price volatility, regulatory challenges, and investor concerns, with the overarching goal of stabilizing stock performance and achieving sustained growth in shareholder value amidst a dynamic business environment.

**GOALS:**

1. **KEY VARIABLES AND MODEL TARGETS**:
   * Model Targets:
     + Stock Price Performance Metrics (e.g., daily closing prices, trading volume)
     + Customer Sentiment Scores
     + Regulatory Compliance Metrics
   * Success Metrics:
     + Stock Price Stability Index (measured by reduced volatility)
     + Accuracy of Sentiment Analysis Models
     + Compliance Rate with Regulatory Policies
2. **DATA SOURCES**:
   * Internal Data Sources:
     + Historical Stock Market Data (stock prices, trading volume)
     + Customer Feedback and Reviews (Sentiments)
     + Financial Reports (revenue, expenses)
     + Compliance Records (If available)
   * External Data Sources:
     + Social Media Data (customer sentiment)
     + Regulatory Policies and Updates

**HOW TO DO IT:**

1. **OBJECTIVES**:
   * formulating specific questions:
     + How can we predict and stabilize EABL's stock performance amidst regulatory pressures?
     + What is the sentiment of customers towards EABL's products, and how does it correlate with stock price movements?
     + How can we ensure compliance with regulatory policies while optimizing financial outcomes?
   * Team roles and responsibilities:
     + Data Scientists: Model development, analysis
     + Business Analysts: Requirement gathering, stakeholder liaison
     + Database Administrators: Data collection, preprocessing
   * Milestone plan:
     + Milestones include data collection, analysis, model development, and deployment.
2. **SUCCESS METRICS**:

* **Sentiment Accuracy Index (SAI)**:
  + The percentage accuracy of sentiment analysis models in correctly classifying customer sentiment towards EABL's products.
  + Formula: (Number of correctly classified sentiments / Total sentiments analyzed) \* 100
  + Target: Achieve an SAI of 90% or higher within the first six months of model deployment.
* **Volatility Reduction Ratio (VRR)**:
  + The percentage reduction in stock price volatility achieved through data-driven interventions.
  + Formula: ((Initial volatility - Final volatility) / Initial volatility) \* 100
  + Target: Achieve a VRR of at least 20% within one year of implementing recommended strategies.
* **Compliance Assurance Score (CAS)**:
  + A measure of EABL's adherence to regulatory policies, including timely filing of returns and compliance with tax regulations.
  + Formula: (Number of compliant actions / Total regulatory actions) \* 100
  + Target: Maintain a CAS of 95% or higher to mitigate regulatory risks and penalties.
* **Customer Sentiment Impact Factor (CSIF)**:
  + The quantified impact of customer sentiment on EABL's stock price movements, measured through correlation analysis.
  + Formula: Correlation coefficient between sentiment scores and stock price changes (-1 to 1)
  + Target: Achieve a minimum CSIF of 0.5, indicating a moderate to strong correlation between sentiment and stock performance.
* **Profitability Enhancement Index (PEI)**:
  + The measure of improvement in EABL's profitability attributed to data-driven strategies, such as targeted marketing and product innovation.
  + Formula: (post-intervention profitability - Pre-intervention profitability) / Pre-intervention profitability \* 100
  + Target: Attain a PEI of 15% or higher within two years, demonstrating the effectiveness of data-driven initiatives in driving financial outcomes.

These success metrics provide measurable targets for evaluating the effectiveness of this project in achieving its objectives and delivering tangible business value to EABL.

* **Model Accuracy (MA)**:
  + The percentage of correctly predicted outcomes by the predictive models, such as stock price movements or customer sentiment classifications.
  + Formula: (Number of correct predictions / Total predictions) \* 100
  + Target: Achieve a MA of 80% or higher for all predictive models.
* **Precision and Recall**:
  + Measures of the predictive model's ability to correctly identify positive (relevant) and negative (non-relevant) instances.
  + Precision: The ratio of true positive predictions to the total predicted positives.
  + Recall: The ratio of true positive predictions to the total actual positives.
  + Target: Precision and Recall scores of at least 0.8 for all critical prediction tasks.
* **Area Under the ROC Curve (AUC-ROC)**:
  + A measure of the predictive model's ability to discriminate between positive and negative outcomes across different thresholds.
  + Target: Achieve an AUC-ROC score of 0.8 or higher, indicating good model discrimination performance.
* **Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE)**:
  + Measures of the average magnitude of errors between predicted and actual values in regression tasks.
  + MAE: The average of the absolute differences between predicted and actual values.
  + RMSE: The square root of the average of the squared differences between predicted and actual values.
  + Target: Minimize MAE and RMSE scores to ensure accurate prediction of continuous variables, such as stock prices.
* **Cross-Validation Scores**:
  + Measures of model performance across multiple validation sets to assess generalizability and robustness.
  + Target: Achieve consistent and high cross-validation scores (at least above 0.7) across different folds and validation techniques.
* **Model Training Time**:
  + The time taken to train the predictive models on the available data.
  + Target: Minimize model training time to ensure efficient utilization of computational resources and timely model updates.

1. **IDENTIFYI NG DATA SOURCES**:
   * Relevant data sources include historical stock market data, customer feedback, financial reports, social media data, and regulatory policies.
   * Ensure data accuracy and completeness to address the project goals effectively.
   * Consider collecting additional data if existing sources are insufficient to address specific aspects of the problem.

**Impact on Business:**

* The project will enable EABL to:
  + Predict and stabilize stock performance amidst regulatory pressures, fostering investor confidence.
  + Gain insights into customer sentiment and its impact on stock prices, facilitating targeted marketing strategies.
  + Ensure compliance with regulatory policies, minimizing operational risks and financial penalties.
* Ultimately, the project will contribute to EABL's ability to maintain a competitive edge in the market, drive sustainable growth, and maximize shareholder value.

1. **PRODUCT**

After deployment, the end-user (EABL) will be able to access and interact with the product through a user-friendly interface or platform. Here's how they can access it:

1. **Dashboard or Web Application**:
   * Through a dashboard or web application accessible through a web browser.
   * EABL stakeholders can log in using their credentials to access the platform securely.
2. **Interactive Visualizations**:
   * An interactive visualizations feature presenting key insights and metrics derived from real-time data analysis.
   * Stakeholders can explore trends, patterns, and correlations in the data through interactive charts, graphs, and maps.
3. **Data Exploration and Filtering**:
   * Users can filter and drill down into the data based on various parameters such as time period, product category, customer segment, and sentiment score.
   * This allows stakeholders to customize their analysis and gain deeper insights into specific aspects of EABL's operations and market performance.
4. **Alerts and Notifications**:
   * An alerting mechanism to notify stakeholders about significant events or anomalies detected in the data.
   * For example, stakeholders can receive alerts for sudden fluctuations in stock prices, significant changes in customer sentiment, or non-compliance with regulatory policies.
5. **Model Outputs and Recommendations**:
   * Generation of automated reports or recommendations based on predictive models and analysis results.
   * Stakeholders can review these outputs to make informed decisions regarding marketing strategies, product development, regulatory compliance, and financial planning.
6. **Feedback Mechanism**:
   * A feedback mechanism for users to provide input on the usefulness and effectiveness of the product.
   * EABL can use this feedback to continuously improve the platform and tailor it to the evolving needs of the business.

Overall, the deployed data science product will empower EABL stakeholders to access, explore, and interact with data-driven insights in a user-friendly and intuitive manner, facilitating informed decision-making and driving business outcomes.

1. **PROJECT PLAN**

To address the challenge faced by EABL, a data-driven solution leveraging product sentiment analysis and time series data can be formulated

**Data Collection and Preprocessing:**

* + Gather historical stock market data for EABL, including daily closing prices, trading volume, and any relevant financial indicators.
  + Acquire sentiment data related to EABL's products from social media, customer reviews, and other relevant sources. Utilize natural language processing (NLP) techniques to extract sentiment scores.
  + Organize and preprocess the collected data to ensure consistency and accuracy for analysis.

**Time Series Analysis:**

* + Conduct time series analysis on EABL's stock performance to identify patterns, trends, and factors influencing stock fluctuations over the past 5-10 years.
  + Use statistical methods such as moving averages, trend analysis, and seasonal decomposition to understand the underlying dynamics of the stock market behavior.

**Sentiment Analysis:**

* + Perform sentiment analysis on product-related data to gauge customer perception, satisfaction, and sentiment towards EABL's offerings.
  + Classify sentiments into positive, negative, or neutral categories and quantify the intensity of sentiment shifts over time.

**Correlation Analysis:**

* + Explore the relationship between product sentiment scores and stock market performance using statistical techniques.
  + Identify correlations and causal relationships between changes in sentiment and fluctuations in stock prices.

**Predictive Modeling:**

* + Develop predictive models using machine learning algorithms such as regression, time series forecasting, or sentiment-based models.
  + Predict future stock price movements based on historical trends, sentiment analysis results, and other relevant factors.

**Policy Impact Assessment:**

* + Evaluate the potential impact of Kenya Revenue Authority's policy on EABL's operations and financial performance.
  + Incorporate policy-related variables into the analysis framework to assess their influence on stock market dynamics.

**Recommendations and Strategies:**

* + Generate actionable insights and recommendations for EABL's stakeholders based on data-driven analysis.
  + Propose strategies to mitigate the negative effects of stock tanking and regulatory challenges.
  + Suggest proactive measures to enhance product quality, customer satisfaction, and brand reputation to stabilize and grow the company's stock value.

**Continuous Monitoring and Adaptation:**

* + Establish a framework for ongoing monitoring of stock market trends, sentiment dynamics, and policy developments.
  + Continuously refine and adapt the data-driven solution based on real-time data updates and feedback from stakeholders.

By adopting a comprehensive data-driven approach encompassing time series analysis, sentiment analysis, predictive modeling, and strategic recommendations, EABL will be able to navigate challenges, stabilize stock performance, and foster exponential growth in shareholder value.