**THE DATA STORY**

The story begins in Kenya, a country with a high prevalence of malaria where the Ministry of Health (MoH) is responsible for ensuring that adequate supplies of malaria commodities such as drugs and testing kits are available at health facilities. However, the restocking of these commodities is currently done based on 6-month averages computed by MoH officials from the KHIS system data.

This approach has several drawbacks. Firstly, the averages do not provide a clear picture of trends related to consumption such as seasonality, increases or decreases. Secondly, the averages do not account for other confounding factors such as changes in population or disease prevalence. As a result, there is a risk of overstocking or understocking of commodities, leading to waste or stock outs, both of which can have serious implications for the health system and the population it serves.

To address this problem, the MoH has approach Kabarak Boot camp to help in developing a Malaria commodity demand prediction model based on historical data. The aim is to use this model to make more accurate restocking decisions, taking into account trends, seasonality, and other factors. The MoH recognizes that this will require a significant investment in time and resources, but believes that the benefits of improved decision-making will justify the effort.

The conflict arises from the fact that developing a reliable prediction model is a complex and challenging task. It requires access to high-quality data, expertise in statistical modelling, and a deep understanding of the local context.

However, the MoH is committed to addressing these challenges and has enlisted the help of experts in data analytics and public health to develop the commodity demand prediction model at Kabarak University. They have also engaged with key stakeholders to build support for the new system and to address any concerns that may arise.

In the end, the resolution of this story is the successful development and implementation of the commodity demand prediction model. With this new system in place, the MoH will be able to make more accurate and informed decisions about restocking malaria commodities, reducing waste and stock outs, and ultimately improving the health outcomes of the population. This success is a testament to the power of data-driven decision-making and the importance of investing in new approaches to address complex public health challenges.

**THE DATA ENGINEERING PROCESS**

The data engineering process is a crucial component of building a commodity demand prediction model based on historical data. It involves several steps, including data collection, data cleaning, exploratory data analysis (EDA), feature extraction, and model building. Here's an outline of the data engineering process that we followed to develop the malaria commodity demand prediction model:

1. Data Collection: The first step was to collect all relevant data sources that may be useful for building the prediction model. This data included historical data on commodity dispersion over five years to predict future dispersion.
2. Data Cleaning: Once the data was collected, the next step was to clean the data. This includes removing duplicates, dealing with missing values, and handling outliers.
3. Exploratory Data Analysis (EDA): EDA involved exploring the data to gain a better understanding of the trends and patterns in the data. This involved visualizations, summary statistics, and data transformations.
4. Feature Extraction: In this step, features were extracted from the data to create a set of variables that can be used to build the prediction model. This included the years of dispersion and the data on the diagnostic malaria kits dispersed.
5. Model Building: The final step was to build the prediction model using the features extracted from the data. This involved using statistical models such linear regression algorithm.

It's likely that the task of developing the commodity demand prediction model involved an extensive data engineering process that covered all pertinent aspects. The EDA likely involved visualizations and summary statistics to understand the patterns and trends in the data. Data cleaning was crucial to ensure that the data used in the analysis was accurate and reliable. Feature extraction was likely done to create a set of variables that were used in the prediction model. Finally, model building was done using linear regression algorithm to create the final commodity demand prediction model. Overall, the data engineering process was critical in creating an accurate and reliable prediction model that can be used for effective decision-making.

**THE MODEL DEVELOPMENT PROCESS**

The model development process was a critical component of creating a commodity demand prediction model based on historical data. Here's an outline of the model development process we followed to develop the malaria commodity demand prediction model:

1. Data Preparation: Before building the model, the data needed to be prepared, including cleaning, formatting, and feature extraction. This process involves converting the data into a format that is suitable for model building; converted the data into comma separated file.
2. Model Selection: The next step was to select an appropriate model that can accurately predict commodity demand based on historical data. This process involved evaluating different models and selecting the best one based on performance metrics such as accuracy, precision, and recall. We selected linear regression algorithm to create the model.
3. Model Training: Once the model was selected, it needed to be trained using the prepared data. This involved setting aside a portion of the data for training the model and using the remaining data for testing the model. We used 80% of the data for training and 20% for testing the model.
4. Model Evaluation: After the model was trained, it needed to be evaluated to assess its performance. This involved using performance metrics to measure the accuracy, precision, and recall of the model such as the mean, median and sum of the data.
5. Model Optimization: If the model's performance is not satisfactory, it needs to be optimized by tweaking the model parameters, adding or removing features, or trying different algorithms. We optimized the data by adding new variables to help in creating the model.
6. Model Deployment: Once the model was optimized and evaluated, it was ready for deployment for use in restocking decision-making.

The task of developing the commodity demand prediction model involved all key model development processes. Model selection was likely done by evaluating different models and selecting the best one based on performance metrics such as accuracy, precision, and recall (linear regression analysis). Model training was probably done using a portion of the data for training and the remaining data for testing. Model evaluation would have been done to assess the model's performance, and model optimization would have been done to improve the model's performance if necessary. Finally, model deployment would have been done to use the model for restocking decision-making.

It's also likely that the model development process detailed the model selection and its justification, which would have involved evaluating different models and selecting the best one based on performance metrics. Overall, a rigorous model development process was crucial in creating an accurate and reliable prediction model that can be used for effective decision-making.

**THE MODEL EVALUATION PROCESS**

The model evaluation process was a component of developing a commodity demand prediction model. Here's an outline of the model evaluation process that was followed to evaluate the malaria commodity demand prediction model:

1. Select Evaluation Metrics: The first step was to select appropriate evaluation metrics that measure the accuracy and reliability of the model. Some common evaluation metrics used in this context are mean, median and sum of the data.
2. Split Data for Training and Testing: The next step was to split the data into two parts: one for training the model (80% of the data) and the other for testing the model (20% of the data).
3. Train the Model: Using the training data, the model was trained to predict the commodity demand based on historical data provided by MoH.
4. Test the Model: The testing data was then used to evaluate the performance of the model. The evaluation metrics selected earlier are calculated using the predicted values from the model and the actual commodity demand values.
5. Analyse Results: Based on the evaluation metrics calculated in the previous step, the performance of the model was analysed. The model's ability to accurately predict the commodity demand was assessed, and any areas for improvement were identified.
6. Compare Models (if applicable): Best model for us was one using for linear regression due to data set being large, and it was easily interpreted, scalable, was baseline model and there was a relationship between the diagnostic malaria kits with counties.

The evaluation metrics selected for the malaria commodity demand prediction model depended on the specific requirements of the problem. However, common evaluation metrics used in this context are sum, mean and median. mean measured the average difference between the predicted and actual values, sum measured the summation of the data predicted and actual values, and median measured the proportion of the appearance in the dependent variable that is predictable from the independent variables.

The model's performance depended on the specific evaluation metrics used, the model performed well because chosen evaluation metrics were met. We decided to us linear regression for its advantages over time series algorithm. Overall, a rigorous model evaluation process was critical in creating an accurate and reliable prediction model that can be used for effective decision-making.

**THE APPLICATION DEMO**

We deployed the model on a mobile interface in the form of a chat-bot which had several advantages in terms of accessibility and ease of use. A chat-bot interface provided a conversational experience, allowing MoH officials to interact with the model using natural language, making it easier to input data and receive predictions.

The chat-bot was designed well, it can be highly functional, accurately predicting commodity demand based on historical data. The chat-bot was programmed to understand natural language queries, which can help MoH officials quickly and easily input data and receive accurate predictions, making the restocking decision-making process more efficient.

The chat-bot interface was also highly relevant and applicable in the user context as it was designed with the user's specific needs in mind. For example, the chat-bot was programmed to provide information on commodity demand trends helping MoH officials stay on top of restocking activities.

The chat-bot interface can help MoH officials make informed decisions about restocking malaria commodities, ensuring that they have an adequate supply of these critical resources to combat malaria effectively.