# Number Oracle: Big Mart Sales Predictive Analysis

## Milestone 1: Project Initialization and Planning Phase

The "Project Initialization and Planning Phase" marks the project's outset, defining goals, scope, and stakeholders. This crucial phase establishes project parameters, identifies key team members, allocates resources, and outlines a realistic timeline. It also involves risk assessment and mitigation planning. Successful initiation sets the foundation for a well-organized and efficiently executed machine learning project, ensuring clarity, alignment, and proactive measures for potential challenges.

### Activity 1: Define Problem Statement

Predict the sales of products across various outlets for Big Mart to assist in inventory management and sales strategy. The goal is to analyze historical sales data, understand the key drivers affecting sales, and build a predictive model to forecast future sales.

**Ref. template:** Click Here

**Number Oracle Problem Statement Report:** Click Here

### Activity 2: Project Proposal (Proposed Solution)

**Proposed Project:** The "Number Oracle" project aims to develop a machine learning model for predicting product sales in Big Mart outlets. Using a comprehensive dataset that includes product attributes, store information, and historical sales data, the project seeks to create a predictive model to optimize stock levels, enhance sales strategies, and improve overall inventory management.

**Ref. template:** Click Here

**Number Oracle Project Proposal Report:** Click Here

### Activity 3: Initial Project Planning

Outline key objectives, define the scope, and identify stakeholders for the Big Mart sales prediction system. This includes setting timelines, allocating resources, and determining

the overall project strategy. Establish a clear understanding of the dataset, formulate goals for analysis, and plan the workflow for data processing.

**Ref. template:** Click Here

**Number Oracle Project Planning Report:** Click Here

**Milestone 2: Data Collection and Preprocessing Phase**

**Activity 1: Data Collection Plan, Raw Data Sources Identified, Data Quality Report**
**Raw Data Sources Identified:** Gather relevant Big Mart sales data from various sources. Ensure data quality through verification and address missing values. Identify raw data sources such as internal sales records, product attributes, and store information.

**Ref. template:** Click Here

**Number Oracle Data Collection Report:** Click Here

**Activity 2: Data Quality Report**
Ensure the dataset's quality by verifying the data, addressing missing values, and adhering to ethical guidelines. Establish a reliable foundation for predictive modeling. Data quality is ensured through thorough verification, addressing missing values, and maintaining adherence to ethical guidelines, establishing a reliable foundation for predictive modelinge stablishing a reliable foundation for predictive modeling.

**Ref. template:** Click Here

**Number Oracle Data Quality Report:** Click Here

**Activity**

**3: Data Exploration and Preprocessing**

**Data Exploration:** Clean, encode, and organize the dataset for subsequent exploratory analysis and machine learning model development. Tasks include handling missing values, encoding categorical variables, and normalizing numerical features**.**

**Ref. template:** Click Here

**Number Oracle Data Exploration and Preprocessing Report:** Click Here

## Milestone 3: Model Development Phase

The Great! In the Model Development Phase, we will focus on building, training, and evaluating the predictive models for the Big Mart Sales dataset. Here's a step-by-step guide to help you through this phase:

Activity 1: Feature Selection Report

The Feature selection is a crucial step in the model development phase. It involves identifying and retaining the most relevant features in the dataset that contribute to the predictive power of the model while removing redundant or irrelevant features. Effective feature selection can enhance model performance, reduce overfitting, and improve interpretability.

**Ref. template:** Click Here

**Number Oracle Feature Selection Report:** Click Here

**Activity 2: Model Selection Report**

The Model selection is the process of choosing the best model(s) for predicting the target variable, Outlet_Sales, from a given set of candidate models. This involves training multiple models, tuning their hyperparameters, and evaluating their performance using various metrics.

**Ref. template:** Click Here

**Number Oracle Model Selection Report:** Click Here

### Activity 3: Initial Model Training Code, Model Validation and Evaluation Report

In this section, we'll provide the code for initial model training, validation, and evaluation. The chosen model for detailed implementation is CatBoost, based on its superior performance in the previous model selection report.

**Ref. template:** Click Here

**Number Oracle Model Development Phase Template:** Click Here

## Milestone 4: Model Optimization and Tuning Phase

The goal of this phase is to improve the performance of the model by fine-tuning its hyperparameters and exploring additional techniques like feature selection and ensembling.

### Activity 1: Hyperparameter Tuning Documentation

The Hyperparameter tuning is a critical step in optimizing machine learning models. It involves selecting the best set of parameters that provide the highest model performance. This documentation provides details on the hyperparameter tuning process for the CatBoost model in the Big Mart Sales Predictive Analysis project.

### Activity 2: Performance Metrics Comparison Report

The Performance Metrics Comparison Report This report provides a comparison of the performance metrics for various models tested during the model selection and tuning phases. The goal is to identify which model performs best based on evaluation metrics such as Mean Squared Error (MSE) and R-squared ($R^2$).

### Activity 3: Final Model Selection Justification

The Final Model Selection This section provides a detailed justification for selecting the CatBoost model, potentially in an ensemble setup, as the final model for predicting Big Mart Sales. The justification is based on a thorough comparison of different models, hyperparameter tuning, and evaluation of performance metrics.

. **Ref. template: Click Here**

**Number Oracle  Model Optimization and Tuning Phase Report: Click Here**

**Milestone 5: Project Files Submission and Documentation**

For project file submission in Github, Kindly click the link and refer to the flow. **Click Here**

For the documentation, Kindly refer to the link. Click Here

**Milestone 6: Project Demonstration**

In the upcoming module called Project Demonstration, individuals will be required to record a video by sharing their screens. They will need to explain their project and demonstrate its execution during the presentation.