Weekly Summary Report

# Week 1 and Week 2: Data Preparation, Synthetic Data Generation, and Predictive Analysis Planning

## 1. Overview of Tasks Completed

### Week 1: Data Collection and Preparation

Objective: The primary objective during week 1 was to create a foundational dataset that includes country and state information, specifically for the USA, UK, and India.  
Actions Taken:  
- A DataFrame was created to list all the states of the USA, UK, and India, which served as the basis for subsequent tasks.  
- The dataset was further expanded to include synthetic data representing sales, total users, revenue, and profit for two products: Apple AirTags and Jio Tags.

### Week 2: Enhancing the Dataset and Planning Predictive Analysis

Objective: Enhance the dataset by adding new features and prepare for predictive analysis tasks.  
Actions Taken:  
- Additional columns (Made\_in\_India and Cheaper\_Price) were added to simulate the scenario where the products are manufactured in India at a reduced cost.  
- The concept of Cheaper\_Price was based on a hypothetical 25% reduction in price, assuming the cost-effectiveness of manufacturing in India.  
- A synthetic dataset was generated for various states across different countries, including calculated fields for sales, total users, total revenue, and profit.  
- Plans were laid out to perform visualization tasks using different models and to predict the success rate and accuracy for products manufactured in India at a lower cost.

## 2. Key Learnings

### Data Preparation and Cleaning:

- Learned the importance of thorough data preparation, which is critical for ensuring the quality and accuracy of the dataset used in further analysis.  
- Experienced the process of structuring a DataFrame with real-world geographical data (countries and states) and integrating synthetic data for sales and user statistics.

### Feature Engineering:

- Developed an understanding of feature engineering by adding hypothetical columns (Made\_in\_India, Cheaper\_Price) to simulate real-world scenarios.  
- Learned to apply domain knowledge to enhance the dataset for predictive analysis.

### Synthetic Data Generation:

- Gained experience in generating synthetic data to simulate real-world conditions, which is essential when working with limited or no real data.  
- Learned to create additional features, such as total revenue and profit, using logical assumptions and calculations.

### Predictive Analysis Planning:

- Recognized the importance of planning predictive analysis tasks in advance, including defining the scope (e.g., predicting success rate and accuracy for products manufactured in India).  
- Identified the need for creating visualization charts to better understand data distribution and trends before diving into model training.

## 3. Challenges Faced

### Data Synthesis Logic:

- One of the primary challenges was defining the logic for generating synthetic data that accurately reflects real-world conditions, particularly in calculating the number of users and ensuring realistic revenue and profit figures.

### Feature Engineering Complexity:

- Creating features like Made\_in\_India and Cheaper\_Price required careful consideration of the implications on downstream predictive analysis and ensuring these features were meaningful in the context of the analysis.

## 4. Conclusion

The tasks completed in weeks 1 and 2 laid a strong foundation for the subsequent predictive analysis. Through careful data preparation, synthetic data generation, and thoughtful feature engineering, we have created a robust dataset that will allow us to explore predictive scenarios effectively. The upcoming weeks will focus on visualizing this data and building models to predict the success of strategic decisions, such as manufacturing products in India at a lower cost.