Project Part 1:

Small Data Problem Analysis Report

Complete this document and submit it with your project.

Match the scenario with the most appropriate solution and explain your choice

Scenario #1: Travel Planner Problem

A travel planning company asks customers to share pictures of past vacations/holidays so their staff can identify what kind of trips they enjoy. The company offers three basic categories of trips:

- Exploring in the Forest
- Adventure in the Desert
- Relaxing on the Beach

As part of a new online trip planning software, the company is creating an AI bot that will automatically figure out from the uploaded photos which category is likely to be most appealing to the customer. The challenge is the company has fewer than 500 photos that are categorized, and they feel it will be difficult to train a model using such little data.

Scenario #1: Travel Planner Problem Should you use transfer learning or a synthetic data approach to solve this problem? Please explain your answer in a short paragraph containing 3-5 sentences.	**Recommended Approach: Transfer Learning** Transfer learning is the most appropriate solution because the company has fewer than 500 labelled images, which is insufficient to train a reliable image classification model from scratch. Pre-trained image models already understand general visual features such as landscapes, textures, and environments that are relevant to forests, deserts, and beaches. By fine-tuning an existing model on this small dataset, the company can achieve good performance without requiring large amounts of labelled data. Synthetic data generation for images is complex and may not accurately capture real-world travel scenarios.

Scenario #2 Loan Funding Prediction Problem

A loan company has a fairly large dataset that they want to use to train a model that predicts whether or not a loan should be funded. The problem they face is the dataset they are using has a large class imbalance... they don't have enough examples of loans that were denied. This is creating a model that doesn't perform well, particularly for loans that probably should be denied.

Scenario #2: Loan Funding Prediction Problem Should you use transfer learning or a synthetic data approach to solve this problem? Please explain your answer in a short paragraph containing 3-5 sentences.	**Recommended Approach: Synthetic Data** A synthetic data approach is more suitable in this scenario because the dataset suffers from a significant class imbalance, with very few examples of denied loans. Generating synthetic samples of the minority class can help balance the dataset and allow the model to better learn patterns associated with loan denial. Unlike transfer learning, which is less effective for structured tabular data, synthetic data directly addresses the imbalance problem. This approach improves model fairness and predictive performance for underrepresented outcomes.