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.	In this scenario, transfer learning is indeed a suitable approach. By using a pre-trained model like VGG-16, which has already learned general features from a large and diverse dataset, you can leverage these learned features to classify the smaller dataset of categorized photos. This approach is effective because the pre-trained model's ability to recognize patterns and features in images can be fine-tuned with your specific data, even if it's limited. This reduces the need for a large dataset and computational

	resources, making it more efficient and practical compared to generating synthetic data, which may not capture the real-world diversity and nuances present in actual photos.

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.	For the Loan Funding Prediction problem, using a synthetic data approach is more suitable. The primary issue here is the class imbalance in the dataset, which means there are not enough examples of denied loans. By generating synthetic data, you can create additional examples of the underrepresented class, thereby balancing the dataset. This will help the model learn to better identify patterns associated with loan denials, improving its performance and accuracy. Transfer learning is less effective in this context because it relies on a pre-trained model from a similar task, and it won't address the specific issue of class imbalance in your dataset.