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.	For the Travel Planner Problem, **transfer learning** is the best approach. • The dataset of fewer than 500 labeled photos is too small to train a deep learning model effectively from scratch. • Using a **pre-trained image classification model** that has already learned general visual features, the company can fine-tune it to recognize specific vacation categories like forests, deserts, and beaches.

	• This approach saves time, reduces the need for large amounts of data, and improves accuracy

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.	Use a **synthetic data** approach combined with resampling and cost-sensitive training. • The dataset is large but severely imbalanced , generating additional realistic minority-class examples with methods like SMOTE or a tabular GAN helps the model learn the denied-class decision boundary without discarding majority data. • Complement synthetic augmentation with class weighting or focal loss and evaluate with recall to ensure improved detection of denials. • Transfer learning is less appropriate here because pretrained image/text models don’t directly apply to tabular loan data.