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.	Transfer learning would probably be the best since the data consists of pictures containing common objects (presumably there are lots of similar pictures that have been taken). This assumes that there is an appropriate existing model for landscape identification. Since many of the images submitted may have nothing to do with forests, deserts, or beaches, a classifier with classes of “forest”, “desert”, “beach”, or “other” may be the best option initially if the existing data is labeled that way. Trying to figure out the effect of images unrelated to the offered trip types may not be possible with so little data.

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.	Synthetic data could be used to correct the class imbalance (generate examples for the denied class). Transfer learning would not be useful for correcting the class imbalance. The resulting model would quantify the expertise of the loan officers. However, it would also be useful to look at how profitable each loan was (based on whether the customer paid on time, defaulted, etc.) if possible to identify any loans that should have been denied and potentially improve upon the expertise of the loan officers.