Revenue Prediction Mini-Project

Project Overview

In this mini-project, you will work with open banking transaction data to build a revenue prediction pipeline. The dataset includes various transaction types for multiple companies over time. This project is designed to assess your ability to explore data, build a machine learning pipeline, and handle real-world data issues. We are interested in seeing how you approach this problem, how you break it down and your thought process, rather than seeing a fully fledged, production ready code - this task should take you around a couple of hours to complete for this purpose!

Your main tasks:

```
Task 1: Explore the dataset in a Jupyter notebook and document your findings.

Task 2: Build a Python script that implements a machine learning pipeline to predict revenue.
```

Instructions

Task 1: Data Exploration (Jupyter Notebook)

- Load the provided dataset (open_banking_data.csv) into a Jupyter notebook.
- Perform Exploratory Data Analysis (EDA) to understand the dataset:
- Use visualizations or summary statistics to present your findings.
- Highlight any observations or issues in the data.

Task 2: Build a Machine Learning Pipeline (Python Script)

- Create a Python script to implement a machine learning pipeline:
- Algorithm: You can use any machine learning algorithm to build the model.
- Expectations: The focus is not on achieving high performance but on demonstrating how you structure your pipeline and approach the problem.

Deliverables

1. Jupyter Notebook (Task 1): Includes your data exploration and observations.
2. Python Script (Task 2): Implements the machine learning pipeline.
3. Summary: Write a short paragraph (not more than one page) explaining your approach, decisions made, and any key findings.

Additional Notes

- Create a repository and add PeterHaugh, camerongj1 and nima-liberis as collaborators.
- Use the dataset provided (open_banking_data.csv).
- Focus on clarity, organization, and the logical progression of your work.