Machine Learning Internship/Associate – Assignment

Project Overview: This Business Requirements Document outlines the analysis of supermarket transaction data collected over a two-year period. The purpose of this project is to allow Machine Learning Engineers to clean, normalize, and transform the data into Python-compatible formats while generating valuable business insights. The analysis will involve utilizing various datasets that include item details, sales transactions, promotions, and supermarket locations, employing Machine Learning techniques to derive actionable solutions.

The project includes two primary components: using supervised learning methods to build predictive models that generate business value and designing a system where an agent learns to navigate a maze. The latter focuses on reinforcement learning, enabling the agent to optimize its actions through trial-and-error feedback, simulating problem-solving in dynamic and uncertain environments.

Timeline: Candidates are expected to complete the assignment within one calendar week from the date of receipt.

Project Scope: The scope is limited to the following areas:

- Data Cleaning and Preparation: Prepare the datasets for analysis by cleaning, normalizing, and transforming them into a Python-compatible format.

- Business-Valued Solutions: Implement at least two business-valued insights using Machine Learning techniques.

- Model Training: Design and train a model that navigates a maze, focusing on reinforcement learning or a similar unsupervised learning method.

Task 01

The provided datasets contain supermarket transactions collected over a two-year period across multiple branches, categorized into four item types: Type 1 to Type 4. These branches are located in two main provinces of the country.

As a Machine Learning Engineer, your task will be to clean, normalize, and transform these datasets into Python-compatible formats. After preparing the data, you are required to develop a supervised learning model to address a business problem.

Details:

- Utilize the provided datasets (**Items.csv**, **Sales.csv**, **Promotion.csv**, and **Supermarkets.csv**) to build a supervised learning model.

- Define a problem statement based on business objectives, such as:

    - Forecasting sales for specific items during promotional periods.
    - Analyzing the impact of promotional features on sales performance.
    - Predicting high-performing supermarkets based on transaction data.

- Choose any supervised learning algorithm (e.g., Linear Regression, Decision Trees, Random Forests, Neural Networks).

- Evaluate the model using appropriate metrics such as RMSE, accuracy, or precision/recall.
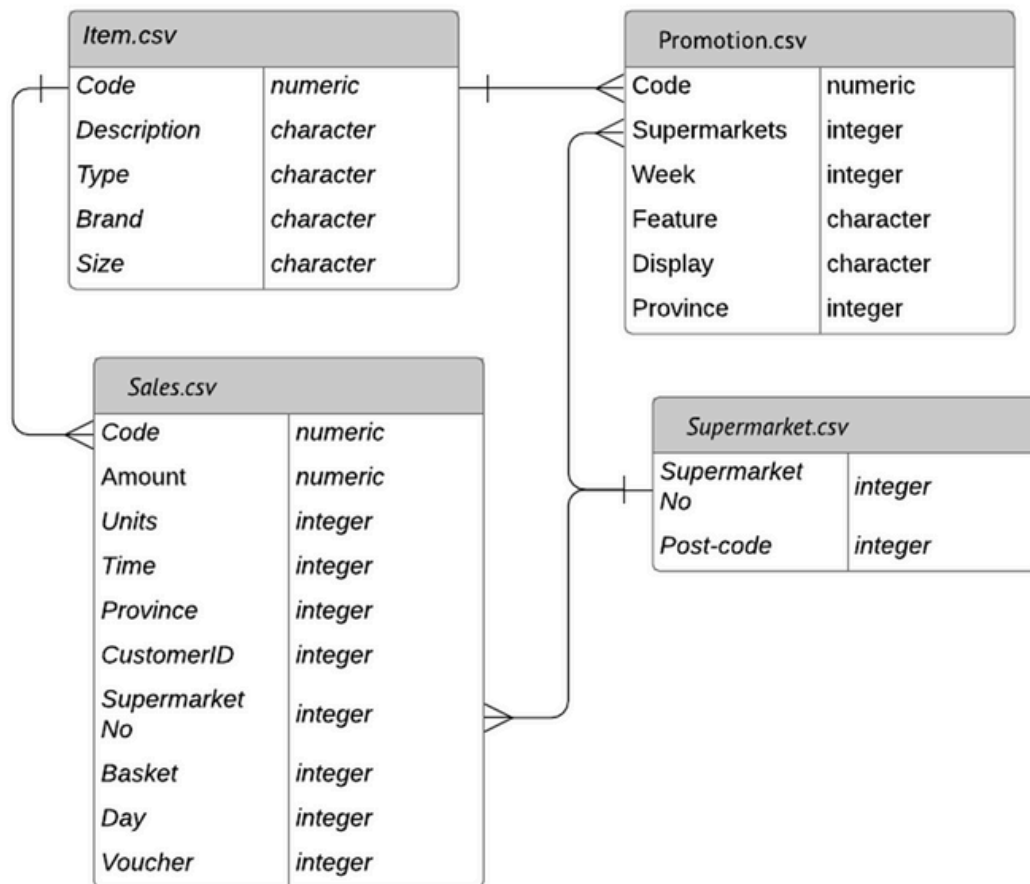
Deliverables:

- Document your data preprocessing steps, feature selection, modeling approach, and evaluation metrics.

- Present clear insights from the model's predictions, demonstrating how they can benefit the business.

Data Sets Provided:

- **Items.csv**: This dataset contains information about items for sale, including:
  - Code
  - Description
  - Type
  - Brand
  - Size

- **Sales.csv**: This dataset contains two years of sales transactions, including:
  - Code
  - Amount
  - Units
  - Time of transactions
  - Province
  - CustomerId
  - Supermarket No
  - Basket
  - Day
  - Voucher

- **Promotion.csv**: This dataset contains details about various promotions across different supermarkets, including:
  - Code
  - Supermarket No
  - Week
  - Feature
  - Display
  - Province

- **Supermarkets.csv**: This dataset provides details about the supermarket locations, including:
  - Supermarket No
  - Post-code

Relationships:

The relationships among the datasets are illustrated in the diagram on the next page.

www.expernetic.com

Item.csv

Code	numeric
Description	character
Type	character
Brand	character
Size	character

Promotion.csv

Code	numeric
Supermarkets	integer
Week	integer
Feature	character
Display	character
Province	integer

Sales.csv

Code	numeric
Amount	numeric
Units	integer
Time	integer
Province	integer
CustomerID	integer
Supermarket No	integer
Basket	integer
Day	integer
Voucher	integer

Supermarket.csv

Supermarket No	integer
Post-code	integer

Task 02

Objective:

Design a system that trains a model to successfully navigate through a maze, avoiding walls and obstacles. The model should learn from its interactions within the maze environment, progressively improving its navigation abilities over time. Focus on implementing reinforcement learning or a similar approach to enable the model to optimize its path based on trial and error, rewarding efficient navigation strategies while penalizing collisions or inefficient movements.

Background:

In this task, you will create a simulation of a maze environment where the agent (the model) must navigate from a start point to an endpoint. The maze will contain walls and obstacles that the agent must avoid to reach the goal. The system should utilize a reinforcement learning algorithm, where the agent receives feedback based on its actions. This feedback will guide the agent in learning the most effective path through the maze.
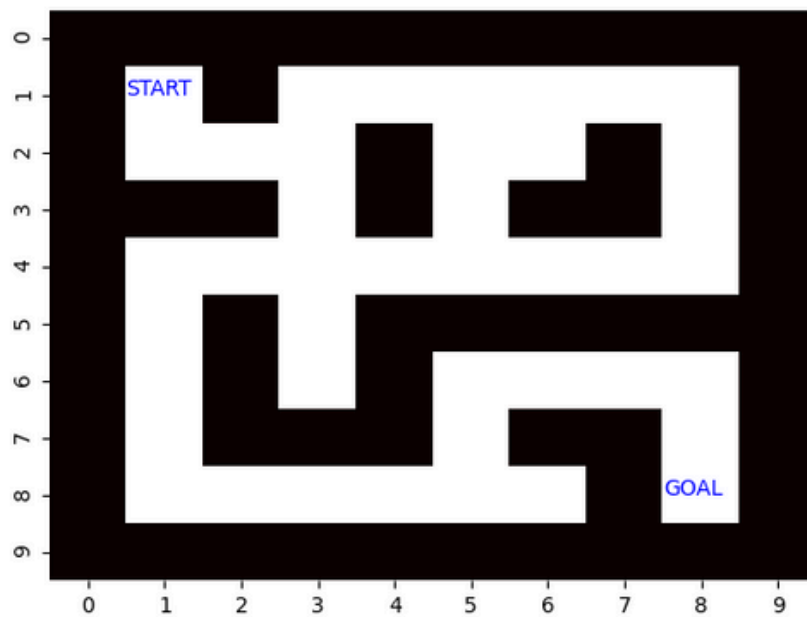
General Requirements:

1. Clean and normalize the provided datasets.
2. Convert the datasets into Python-compatible formats.
3. Apply Machine Learning techniques and design a model to extract at least two meaningful business insights.
4. Design a model that can navigate a maze by learning from interactions with its environment. The model should use reinforcement learning or a similar approach.

Assessment Criteria:

Your assignment will be assessed based on the following criteria:

- Code quality, structure, and organization.
- Accuracy and thoroughness in data cleaning, normalization, and transformation.
- Understanding and demonstration of supervised learning in Task 1.
- Implementation of business-valued solutions using supervised machine learning model.
- Understanding and demonstration of unsupervised learning in Task 2.

- Creativity and problem-solving approach in designing a model to navigate the maze.
- Effective use of Python-compatible tools for data manipulation and analysis.
- Documentation and comments in the code.
- Clear and detailed instructions for running the solution locally.

Report:

In addition to the code, you must provide a well-documented report (PDF). The report should include:

- **Task Overview**: A summary of each task, the steps undertaken, and challenges faced.
- **Data Cleaning & Transformation**: A detailed explanation of the cleaning and transformation processes applied to the datasets.
- **Supervised Learning Model**:
    - Problem definition and objectives of the supervised learning task.
    - Explanation of the model chosen, features used, and the training process.
    - Evaluation metrics and analysis of the model's performance.
    - Insights derived from the model predictions and their business value.
- **Business Insights**: Detailed description of the insights extracted from the data, methodologies employed, and their business relevance.
- **Maze Model Design**:
    - Explanation of the maze navigation model.
    - Overview of the reinforcement learning approach used.
    - Training results and analysis of the model's performance over time.

Submission:

Submit the assignment via a Git repository (e.g., GitHub) and provide access for review. Ensure the repository contains:

- Source code, documentation, and setup instructions.
- The report (PDF) with all required sections.

Once completed, email the repository link and any relevant details to **people@expernetic.com**. If you have questions, contact Amadli Maneesha **(amadli@expernetic.com)** or send an email to **people@expernetic.com.**

Final Note for the Assignment:

Dear Candidate,

We would like to reiterate the importance of completing this assignment independently, with your own efforts and code. During the evaluation process, the following criteria will be considered:

- Original Work: It is crucial that the assignment reflects your own knowledge, skills, and coding abilities. While using AI tools for code generation and assistance is permitted, it is important that the final submission is a reflection of your understanding and skills.

- Independent Effort: You are expected to complete the assignment without assistance from others. Collaboration with external individuals or seeking help from others to complete the assignment is not allowed.

- Code Explanation: During the interview, you will be asked to explain and discuss the code you have written for the assignment. You should be prepared to provide a comprehensive explanation of your code, its structure, and the design choices you made. Failure to explain your own code during the interview may result in rejection from the interview process.

This assignment is an opportunity for you to showcase your technical skills and your ability to develop a working solution independently. We value integrity, originality, and the ability to understand and explain your own work. Your commitment to these principles is essential to progressing through the interview process.
ess.

Best of luck with your assignment, and we look forward to reviewing your work and discussing it during the interview.

Sincerely,

Expernetic LLC.

www.expernetic.com