Data Science Internship – 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 data scientists 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 Data Analytics and Machine Learning techniques to derive actionable solutions.

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 Data Analytics and 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 data scientist, your task will be to clean, normalize, and transform these datasets into Python-compatible formats. After preparing the data, you are required to undertake at least two business-valued solutions using Data Analytics and Machine Learning techniques.

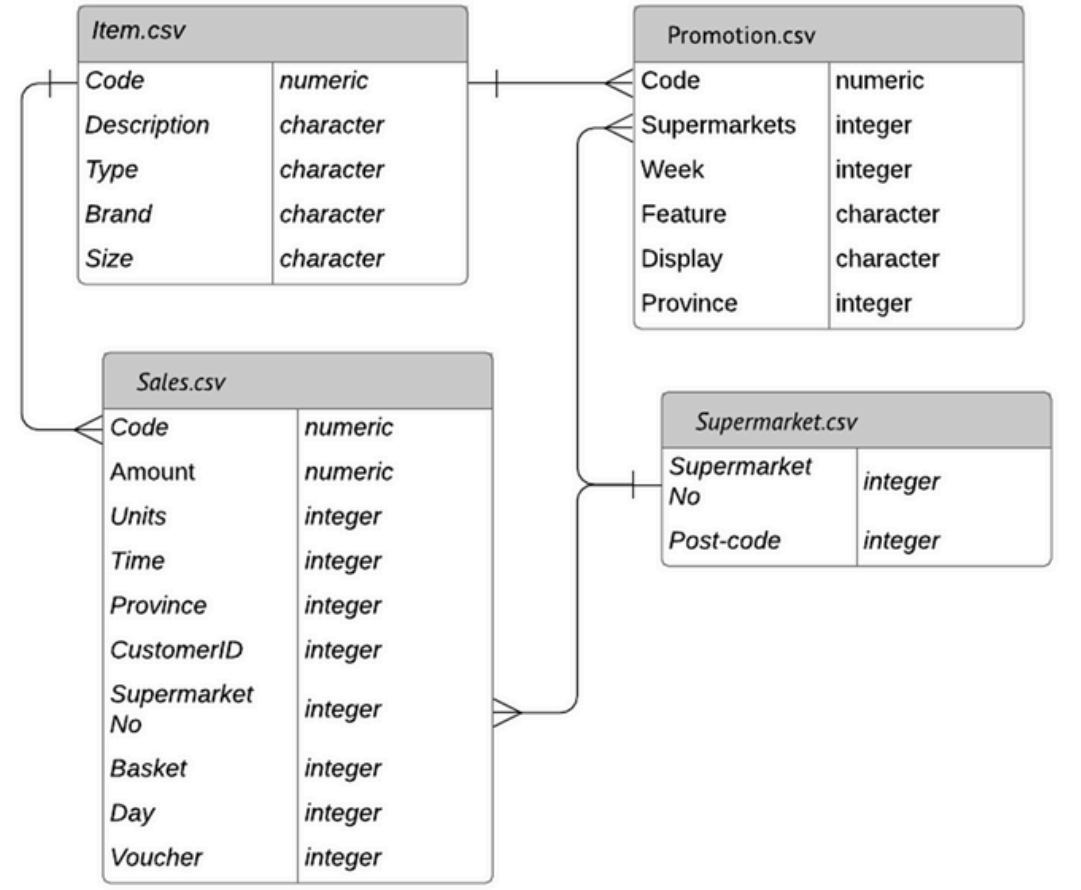
You will work with four datasets: Item.csv, Sales.csv, Promotion.csv, and Supermarkets.csv.

Data Sets Provided:

- **Sales.csv**: This dataset contains two years of sales transactions, including:
    - Code
    - Description
    - Type
    - Brand
    - Size
- **Promotion.csv**: This dataset contains details about various promotions across different supermarkets, including:
    - Code
    - Description
    - Type
    - Brand
    - Size
- **Supermarkets.csv**: This dataset provides details about the supermarket locations, including:
    - Code
    - Description
    - Type
    - Brand
    - Size

Relationships:

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



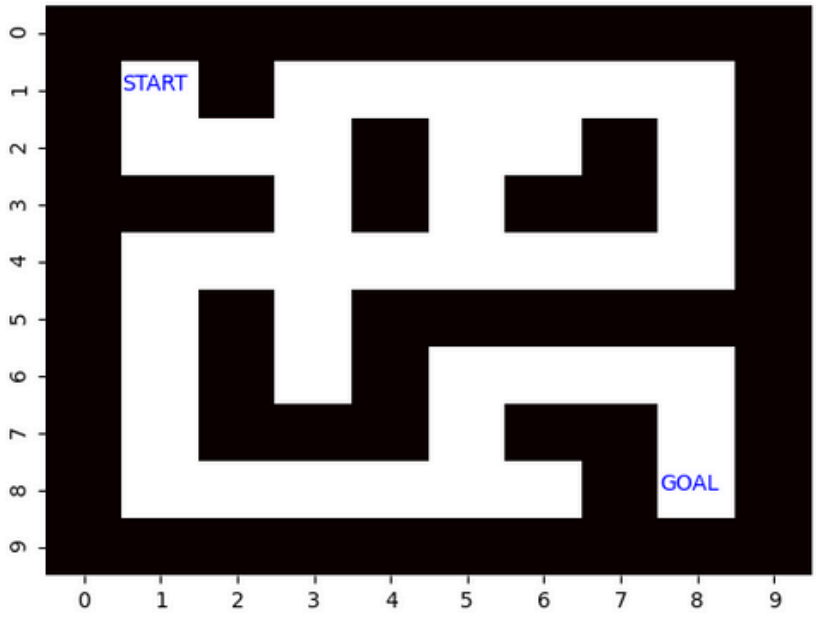
Task 02

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 and Data Analytics techniques 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.
- Implementation of business-valued solutions using Data Analytics and Machine Learning techniques.
- Understanding and demonstration of unsupervised learning in Task 2.
- Creativity and problem-solving approach in designing a model to navigate the maze.

- 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.

Submission:

Please submit your assignment via a Git repository (e.g., GitHub) and provide access to the repository for review. Ensure that your repository contains the complete source code, documentation, and any necessary setup instructions. Once you are done with the assignment, please email the Git Repo with relevant details to **people@expernetic.com**. If you have any questions, please 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.