Recipe Recommender Assignment EDA

Submitted By:

Mr. Lohith Raj M

Miss. Mitali Pensia

Miss. Rani Mondal

Agenda

01. Introduction

02. Objective of this case study

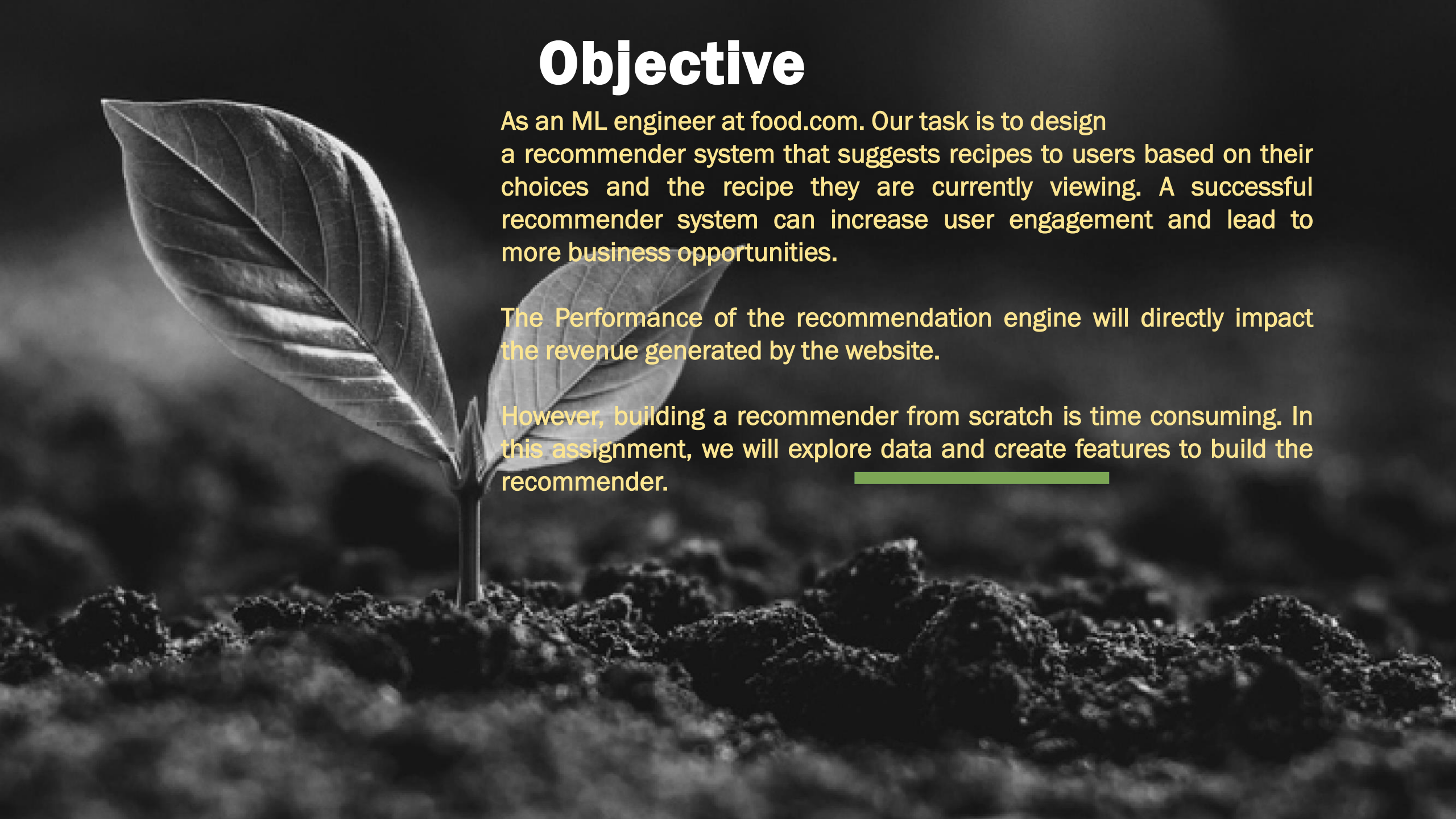
03. Steps to approach the problem:

04. Task List

05. Closing

Introduction

When a user visit a website food.com, they will notice a section called “You’ll also love”. Under this section, the website recommends recipes to the one they are looking at or based on their past rating patterns.



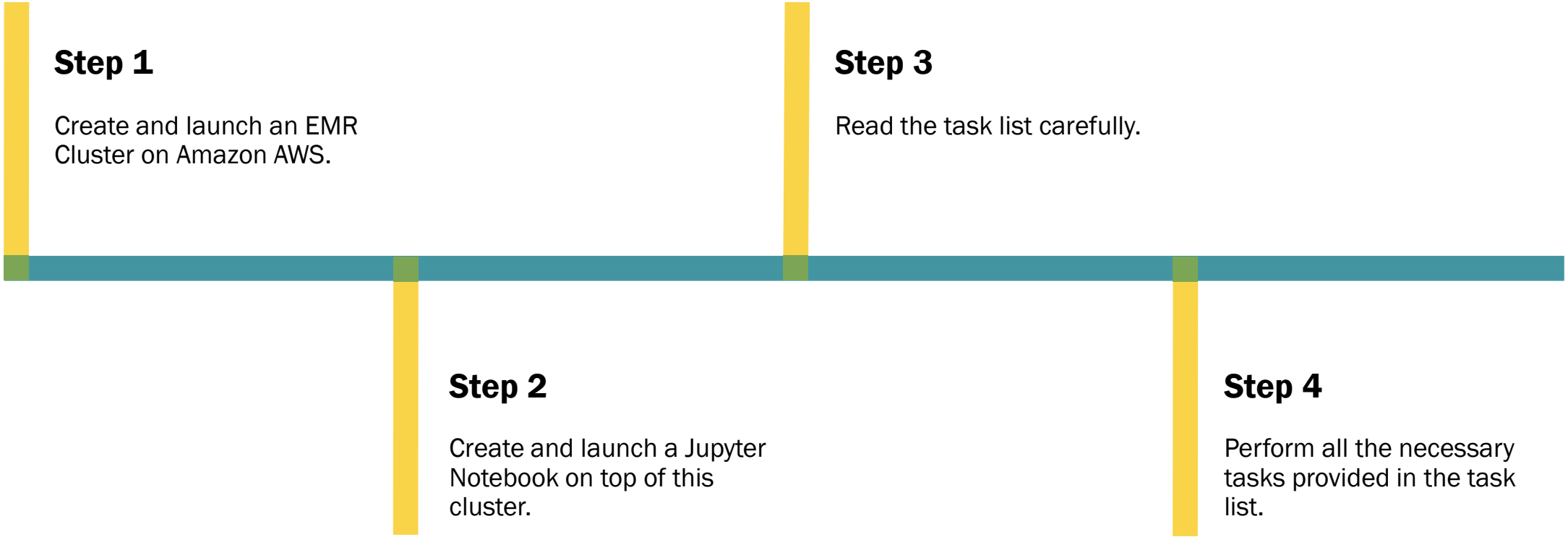
Objective

As an ML engineer at food.com. Our task is to design a recommender system that suggests recipes to users based on their choices and the recipe they are currently viewing. A successful recommender system can increase user engagement and lead to more business opportunities.

The Performance of the recommendation engine will directly impact the revenue generated by the website.

However, building a recommender from scratch is time consuming. In this assignment, we will explore data and create features to build the recommender.

Steps to approach the Problem



Task List

Task 1

- **Read the data:**
- Read ‘RAW_recipes_cleaned.csv’ from S3 bucket.
- Ensuring each field has the correct data type.

Task 2

- **Extracting individual features from the ‘nutrition’ column:**
- Separating the array into seven individual columns to create new columns named ‘calories’, ‘total_fat_PDV’, ‘sugar_PDV’, ‘sodium_PDV’, ‘protein_PDV’, ‘saturated_fat_PDV’, and ‘carbohydrates_PDV’.

Task List

Task 3

- **Standardizing the nutrition values:**
- Converting the nutrition values from absolute to relative terms.
- The nutritional columns are converted to nutrition-per-100 calorie columns.

Task 4

- **Convert the ‘tags’ column from a string to an array of strings:**
- The ‘tags’ column is a String Type column but holds an array of strings
- So ‘tags’ column is converted from strings to an array of strings:

Task List

Task 5

- **Read the second data file:**
- Read the ‘RAW_interactions_cleaned.csv’ from S3 bucket.
- Joining this ‘raw_recipes_df’ with the ‘raw_ratings_df‘ and creating ‘interaction_level_df’.
- The resulting data frame have all the rows of ‘raw_ratings_df’.

Task 6

- **Creating time-based features:**
- Time-based features are created using the ‘review_date’ and the ‘submitted’ from ‘interaction_level_df’.
- Creating features that capture the time passed between review date and the date on which the recipe was submitted.

Thank you