24 Hour Python API Challenge / Project

GOAL

The goal of this test is to create a consumable RESTful API for importing, storing and retrieving sales insights.

OVERVIEW

This test is built to help us assess three aspects:

1. How you manage to import a large amount of data from a CSV file
2. How you design / build a database schema to store the imported CSV data
3. How you approach designing a RESTful API. We are interested in how you organize code. While this application is trivial, we'd like to see it structured as if it would become an enterprise application.
4. Your familiarity with the current Python ecosystem.
5. Your use of Python in modern codebases.

PREREQUISITES

Use this public domain Kaggle dataset for the CSV import, database and analysis.

1. A year's worth of sales from a fictitious pizza place, including the date and time of each order and the pizzas served, with additional details on the type, size, quantity, price, and ingredients.
    a. https://www.kaggle.com/datasets/mysarahmadbhat/pizza-place-sales

REQUIREMENTS

You are allowed to utilize the following libraries:

1. Python
2. Any Python library you see fit e.g. for API development or data science

You may use supporting libraries to implement the features required in this test e.g. Django or FastAPI. You may also opt NOT to use any libraries.

Use Case

1. Create a public github repository to contain your work.
    a. Follow what you understand to be best git development practices
    b. Share this repository with us upon completion of the exercise
2. Build a simple backend application that supports the following oversimplified use-cases
    a. Create an import mechanism for the CSV file

        i. This import mechanism should allow for regular updates i.e. users can upload new CSV files to add and update the existing database.
        ii. You may choose to do the CSV import via a API endpoint or via another mechanism e.g. uploading files into a dedicated folder and running a job to synchronize the database.
    b. Data is stored in a structured way in a new database schema which is suitable for the CSV file
        i. Please also explain why you have chosen this schema and how it is suitable for the imported data
    c. Process the data upon import or build a way to do it regularly e.g. scheduled job to create statistics about the dataset
        i. This should yield the following insights
            1. Ranking of the most popular pizzas
            2. Most popular days and most popular pizzas on these days
            3. A ranking of the most popular ingredients i.e. based on how many pizzas have sold and their ingredients compile which ingredients / pizza toppings are most popular.
    d. Create a simple REST API to consume the insights. You are free to design your own schema for the REST API.
        i. Have the API return JSON data
    e. Bonus: This is optional, but we do value your ideas for additional insights into the sales data. What statistics would you like to produce which might be important for the business. More extra points for implementing your ideas and making these insights available via API endpoints.
3. Create a video recorded (ex. via zoom) demonstration of the application. This should be a walkthrough with voiceover discussing the application and explaining each of the decisions, considerations and tradeoffs that were made in your code. All of the use cases, as defined above, should be demonstrated across each of the targeted platforms: web browser
    a. Walkthrough video with voiceover of you running the web application and showing that all features work
    b. Walkthrough video with voiceover of you showing your code and explaining each of the decisions, considerations and tradeoffs that were made in your code.