Data Engineering Challenge

You are working on a project to help a restaurant by deriving business intelligence and operational insight from food ordering data. Given menu items and orders, extract insights from the given data and document the process of getting there. Derive business metrics that capture key information, and provide some aggregations and visualizations of this information.

Use whatever tools you think are appropriate to create a deliverable that shows the insights you choose to share. Include any source code or scripts that you use to generate the data. Any visualizations can be interactive or statically-produced content.

Details about the system

You are given a menu of items that can be prepared by a restaurant, how long it takes to prepare each item, and each item's price. You're also given a list of orders placed by consumers — these orders come from different food ordering apps. Assume that every order in the list was cooked in parallel based on its item cooking time.

Menu Items are defined as follows. Cook time is expressed in seconds. Price per unit is in USD cents.

```
[
  {
    "name": "Arborio Rice and White Bean Soup",
    "cook_time": 420,
    "price_per_unit": 450
  },
  {
    "name": "Au Gratin Peas and Potatoes",
    "cook_time": 480,
    "price_per_unit": 299
  },
  ...
]
```

Orders are defined as follows. Paid per unit is in USD cents. Time is UTC.

```
[
  {
    "ordered_at": "2020-01-20T16:01:00",
    "service": "SuperEats",
    "name": "Ryan G",
    "items": [
      {
        "name": "Cassoulet for Today",
        "paid_per_unit": 262,
        "quantity": 1
      },
```

```
      {
        "name": "Rigatoni with Sausage & Peas",
        "paid_per_unit": 500,
        "quantity": 1
      }
    ]
  },
  {
    ...
  }
]
```

Extra Credit

Assume that your restaurant can only make N items in parallel, and the rest are queued. How would the N be determined? What data is needed to support this decision, what form would it take, and how would it be used?

Deliverables

Please take your time to deliver a quality solution that shows the best of your abilities. Include:

- Business metrics and other insights produced from the input data, in the form of:
  - Static PDF or HTML with output, *or*
  - Dockerized runtime with an interactive report, or equivalent.
- Supporting code that:
  - Is readable and something that would pass a thorough code review.
  - Clearly defines any intermediate data schemas used for processing or aggregating the raw data.
  - Shows any data cleansing or transformation steps.
  - *Note that you do not need to make the code runnable if it depends on frameworks or reporting tools not included with your challenge.*
- A README file that contains:.
  - A list of assumptions made in producing the solution.
  - A description of the supporting code, and the process used to generate the business metrics.
  - A description of the data pipeline that you believe could exist to power a similar system in production.
  - A summary of the business metrics or insights produced.

What we look for

This challenge is meant to help us see your best code and abilities, and to showcase your judgment. When we evaluate the challenge, we look at how focused you were on meeting the requirements, at the simplicity of your approach, at your use of appropriate data design patterns, your choices of schemas, and your ability to extract and present useful insights from

the data. Assume that you’re building something that would scale, last a while, and that you will want to be proud of.