**Eaze Data Scientist Technical Exercise**

You are taking on the role of a Data Scientist at Eaze. The following is a short exercise (~90 min) that should help familiarize you with some of the problems we tackle and demonstrate some of your skills. This is designed to be open-ended, so while there are specific prompts feel free to think outside the box. You may use any programming language/platform you are comfortable with.

**Data**

Two files of randomized data are attached that are similar to some of the data in our system:

* delivery\_geography
  + order\_id: unique identifier per order (int)
  + latitude: delivery location (degrees)
  + longitude: delivery location (degrees)
  + estimateddeliverytime: an estimate of how long a delivery took (seconds)
* delivery\_prices
  + order\_id: unique identifier per order (int)
  + deliveredat: time delivered (timestamp)
  + totalprice: total product price (USD)

**Tasks**

1. Join these two tables in a meaningful way.
2. Find at least one source of external information that you think would augment this table well and add it into the joined table.
3. Explain (at a high level) how you would perform this join on a live data pipeline in a performant way.
4. Use any clustering/classification technique to divide the data into at least 3 and at most 20 groups. Annotate your joined table with this classification.
5. Create any visualization of the data that you think would display meaningful insight.
6. Explain (at a high level) how you could use any insights derived above to advance the company financially.

**Deliverables**

* Commented code used to solve these tasks
* A single output table with the annotated classification
* One or a series of visualizations
* Text responses to each question (total length 1-2 pages)
* Anything else you feel is relevant