**Efficio Cloud Analytics Case Study**

Welcome! Thank you for taking the time to participate in the process. To start off, a few ground rules –

(1) Please remember - one of the main outputs of this study is your code base, please include anything you want us to know in your readme or comments. The easier it is to see your thinking, the better. While we might ask some questions about the decisions/tradeoffs considered in subsequent rounds, that is only as needed.
(2) Please feel free to ask questions – email your HR contact, and we will make sure to get back to you by end of day.
(3) Please use Python.

Efficio is working with a client looking to mitigate their cloud computing cost. As part of this initiative, one of our stakeholders has asked the team to set up a suite of models for improved forecasting & spend visibility. Your only data input is their cost & billing usage report for the past 18 months.

The client would like you to develop a forecast for their cloud spend for the next quarter to build up their budget. Through this process, they also want to be able to understand which spend areas are going to cause growth so they can prioritize specific AWS services for optimization efforts. They hope your model will help address both of these areas.

Historically, the client has found this very difficult to do. Due to the high volatility in some elements of their spend – they find it quite difficult to budget from year to year. It is also difficult to differentiate between unusual/unexpected events that require remediation versus normal fluctuation. As an example – storage costs will often be significantly more stable, but slow upticks over time can drive a huge budget impact. By contrast compute is quite often spiky & has bursts related to regularly scheduled jobs.

The data they have provided has detail on the AWS Service – product, line item type, line item description, usage type, usage type aws, and operation. You also have information on which portions of the client infrastructure are using the services – product (tag) and environment (tag). You can use whichever combination of features are most helpful to help shape an accurate forecast and understand the spend drivers at an actionable level.

Your deliverables include:

- Your forecast:
    - A report summarizing your analysis & modeling approach, including visualizations & key insights
    - A forecast of spend broken down by Product (Product (tag)) for the next six months - you can decide on the granularity (both time & which attributes to use) and method
    - Code and documentation for the model, including any preprocessing steps, feature engineering, and model selection.
- Based on your forecast, lay out the main drivers of spend increase for the client, and whether you believe they should be prioritized for optimization efforts. This can be in terms of the client's infrastructure components (product (tag) & environment (tag)), or the AWS service information, or both!