Harvard Business School

DRAFT
JULY 29, 2025

IAVOR BOJINOV

Cloud Cost Optimization: A Hands-on Generative AI Exercise

Introduction

You've just stepped in as the new Chief Technology Officer (CTO) at FlavorFusion Foods, a consumer-packaged-goods company that has grown at a dizzying pace. Unfortunately, its cloud infrastructure costs have grown even faster, skyrocketing from $380,000 to a staggering $2.4 million per month in just six months. The CFO is alarmed, and the board is demanding answers.

Your mission is to use generative AI as your trusted analyst to investigate this crisis. You will need to analyze six months of cloud cost data, identify the key drivers of the cost explosion, and build an actionable recovery plan to present to the executive team. To help you, you can leverage a generative AI tool with data analysis capabilities.

Company Context

- **FlavorFusion Foods:** A rapidly growing CPG company that scaled from a regional startup to a national brand in 18 months.
- **Infrastructure:** The company operates a complex environment including a high-traffic e-commerce platform, a sprawling supply chain system, multiple data analytics platforms, and core corporate systems (ERP, marketing automation, etc.).
- **The Challenge:** Cloud spend has reached an unsustainable $2.4 million per month, with non-production environments shockingly accounting for 62% of the total.

Task 1: Understand the Crisis with Exploratory Data Analysis

First, you need to get a handle on the data and confirm the scale of the problem. Exploratory Data Analysis (EDA) is like being a detective with a new case file. You start by summarizing the evidence, looking for initial clues, and getting a feel for the situation before you identify your main suspects.

1. Begin by loading the provided `cloud_costs_daily.csv` data file into your AI assistant.
2. Ask your AI assistant to perform an initial analysis to understand the data's structure and check for any quality issues.
3. Work with your assistant to visualize the spending trends to see the problem for yourself.

Professor Iavor Bojinov prepared this exercise as the basis for class discussion rather than to illustrate either effective or ineffective handling of an administrative situation.

- **Example prompt:** Can you first check it for any data quality issues like missing values? Then, create two charts: one showing the total daily cloud cost over the six-month period, and another showing the total monthly cost trend. Be sure to generate visualizations and emphasize the key takeaways."

4. Dig one level deeper to see what's driving the trend.

    - **Example prompt:** "That confirms the cost explosion. Now, can you break it down further? Please create new charts showing the monthly spending trend separated by both Resource Type and by Department."

Task 2: Find the Big Money with Cost Concentration Analysis

Now that you've confirmed the crisis, it's time to find out where the money is really going. In any cost problem, a few areas typically account for the majority of the spend. This is where you'll find your biggest opportunities for immediate savings.

1. Focus on the most expensive individual resources.

    - **Example prompt:** "Help me find where we're hemorrhaging money. First, show me a table of the top 20 most expensive resources by their average daily cost. Include their average utilization percentage in the table. Plot these over the last 6 months."

2. Analyze the massive non-production environment spending.

    - **Example prompt:** "Can you give me a breakdown of the total non-production spend by environment (e.g., Development, Test, Staging) and by resource type?"

3. Identify specific, high-cost, low-use resources that are clear targets for cuts.

    - **Example prompt:** "Drill down into that non-production waste. List all specific resources in the Dev, Test, or Staging environments that cost more than $5,000 per month but have an average utilization of less than 30%."

Task 3: Uncover Systemic Waste with Clustering Analysis

Individual wasteful resources are one thing, but you need to find systemic patterns of inefficiency. Clustering is a machine learning technique that automatically groups similar things together. Think of it as an algorithm that can sort a messy closet for you. We'll use it to find groups of wasteful resources that share common characteristics.

1. Ask your AI assistant to run a clustering analysis to group resources by their cost and usage patterns.

    - **Example prompt:** "Let's use clustering to find patterns of waste. For each resource, calculate its average daily cost and average utilization over the last 30 days. Then, use a k-means algorithm to create 5 clusters based on those two features. Plot the results."

2. Interpret the clusters to find the most wasteful groups.

    - **Example prompt:** "Now, can you analyze those clusters? For each one, show me the average cost and average utilization. Based on that, give each cluster a descriptive label, like 'High-Cost, Low-Use' or 'Efficient & Optimized'."

3. Calculate the savings potential from your most wasteful cluster.

- **Example prompt:** "For the cluster you labeled 'High-Cost, Low-Use', list all the resources within it. Then, calculate the total monthly cost of that entire cluster. Which department owns the most resources in this group?"

Task 4: Drive Departmental Accountability

The cost problem isn't just technical; it's organizational. You need to identify which departments are driving the cost explosion and may need better governance and controls.

1. Analyze spending growth by department to see who the biggest spenders are.
    - **Example prompt:** "Now let's analyze spending by department for January 2025. Show me a table and plot of each department's total spend and their percentage growth in spending from August 2024 to January 2025."
2. Compare the operational efficiency of each department.
    - **Example prompt:** "Which departments are the least efficient? Create a chart that shows the average resource utilization percentage for each department."
3. Investigate specific departmental spending patterns that look suspicious.
    - **Example prompt:** "Marketing's spending growth seems very high. Can you show me Marketing's top 10 most expensive resources? And for the Engineering department, list all of their idle resources that still cost more than $100 per day."

Task 5: Predict and Prevent Future Waste

Great leaders don't just solve today's problems; they prevent tomorrow's. You can use machine learning to build a predictive model that acts as an early warning system for waste. We'll build a model that learns from past data to predict which resources are likely to become idle *before* they waste money.

1. Work with your AI assistant to prepare the data and build a predictive model.
    - **Example prompt:** "Let's build a model to predict future waste. First, create a new target variable called 'became_idle', marking it TRUE if a resource's utilization dropped below 10% for a sustained period. Then, create features from the data like utilization trend, resource age, environment, and department."
2. Train the model and use it to identify at-risk resources.
    - **Example prompt:** "Now, train a Random Forest classification model to predict the 'became_idle' target. Once the model is trained, use it to list the top 30 resources that are most likely to become idle in the next month."
3. Quantify the potential impact of your predictive model.
    - **Example prompt:** "If these predictions are accurate, how much money could we save next month by proactively shutting down or rightsizing these 30 resources before they become idle?"

Task 6: Build Your Recovery Plan

You've analyzed the data, identified the problems, and even predicted future issues. Now it's time to bring it all together. Based on your conversation with your AI assistant, create a clear, actionable plan to cut costs and establish long-term governance.

1. Summarize your findings and generate the core components of your plan.
    - **Example prompt:** "Based on our entire conversation, create an emergency cost reduction plan to cut at least $500,000 in monthly spend. The plan should include a list of at least 10 specific resources to terminate immediately and 10 resources to right-size, showing the potential savings for each. Also, recommend new governance policies we should implement, like auto-shutdown rules and budget limits by department."
2. Structure the findings into a clear roadmap.
    - **Example prompt:** "That's a great start. Now, organize that plan into a 30-day roadmap. Create a table with three columns: 'Timeline' (e.g., Week 1, Month 1), 'Action', and 'Estimated Monthly Savings'. This will be the document I share with the CFO."

Appendix: Data Dictionary

File: cloud_costs_daily.csv

Variable	Description	Example Values
date	Date of the measurement	"2024-08-01"
resource_id	Unique identifier	"PROD-WEB-001"
resource_name	Descriptive name	"Customer Portal Web Server 1"
resource_type	Type of cloud service	"Web Server", "Database", "Storage", "Analytics Cluster"
application	Business system using it	"E-commerce Platform", "Customer Analytics"
department	Who pays for it	"Engineering", "Marketing", "Operations", "Data Science", "Finance"
environment	Where it runs	"Production", "Development", "Test", "Staging"
daily_cost	Cost for this day ($)	1,127.00
utilization_percent	How busy it was (0-100%)	15.2
peak_utilization	Maximum utilization (%)	78.5
idle_hours	Hours with <5% usage	18

usage_pattern	When it's used	"24/7 Steady", "Business Hours", "Weekly Batch"
size_current	Resource size	"Small", "Medium", "Large", "XLarge", "2XLarge"
is_production	Production resource?	TRUE/FALSE
has_shutdown_schedule	Auto-shutdown configured?	TRUE/FALSE
business_hours_only	Should only run 9-5?	TRUE/FALSE
last_reviewed_date	Last optimization review	"2024-11-01"