

**Company:** JPMorgan Chase & Co. - Job Sim via Forage.

**Title:** Quantitative Research Virtual Experience Program.

**Date Completed:** August, 2025.

**Duration:** 2 Weeks.

**Tools Used:** Excel, Python (Pandas, NumPy, Prophet, SKLearn, Joblib, Matplotlib)

**CASE STUDY:** Commodity Trading Market Pricing Model and Financial Predictive Model using data mining.

## Overview

This project was a two part project run by JPMorgan chase & Co. (JPMC) as a Quantitative Analysis Virtual Internship on Forage. The objective of the first part was to help a commodity trading retail client build a pricing model from existing historical gas price data. As a data analyst, my role was to download, clean and analyze the data to estimate the purchase price of gas at any date in the past and extrapolate it for one year into the future, and then write a function that is able to use the data to price the contract at any date given basic parameters like injection dates, withdrawal dates, storage cost, etc.

The second part focused on assisting the retail banking arm of JPMC, which has been experiencing higher-than-expected default rates on personal loans. The risk team wants to know if more defaults should be expected in the future and what the expected loss is likely to be. My role was to use a sample of their loan book to build a prototype predictive model that given details for any loan will predict the probability that the borrower will default (PD) and use the provided data to train the model assuming a recovery rate of 10%.

## Dataset

CSV files of historical gas price data containing 48 months of gas price data (October 2010 – September 2024) and transactional loan record data containing 10000 rows and 8 columns were provided and downloaded from forage website.

## Tools & Skills Used

**Tools:** Microsoft Excel, Python

**Skills:** Data cleaning, Data Modeling, Machine Learning, Data Visualization, Quantitative Analysis

## Process and Approach

- Imported raw data from Microsoft Excel into Python.
- Cleaned raw dataset by removing duplicates, handling missing values, and correcting inconsistencies; and then transformed and filtered the data.

- Wrote Python codes to create and train a regression model for gas contract price with one year extrapolations on the data.
- Wrote Python codes to create and train a predictive model to predict probability of defaults on the loan data.
- Created visuals to explore, analyze hidden patterns and communicate insights in the data using Matplotlib.

## **Insights & Findings**

- Gas price has been trending upward since October 2010 with peaks usually between Januarys – March and lows usually between Julys –September.
- The latest peak and low occurred at \$12.75 and \$10.90 respectively
- Gas price is expected to peak at around \$12.90 between January – April 2025 with low at around \$11.80 between May – June 2025.
- Gas prices are highly seasonal with peaks around January – April and lows around July – September every year.

## **Recommendation**

- The best time for injection is between July - September while the best time for withdrawal is between January and April.
- Gas prices are very volatile between April and June (bearish) and October and January (Bullish) and so Gas contracts should be traded with caution within this period.

## **Skills**

- Data Cleaning and transformation
- Data modeling and Analysis
- Data Visualization and Interpretation
- Data Mining and Machine Learning
- Presentation & Insights
- Charts & Graphs
- Effective Communication

## **Links & Artifacts**

- **GitHub Repo:** <https://shorturl.at/RV4pz>
- **Certificate:** <https://shorturl.at/SDw2o>

## **Reflection**

Through this project, I built on my fundamental knowledge of data mining, machine learning and data statistical modeling. Also I harnessed my programming skills in python and this further broadened my data analytics skills. Furthermore, this project introduced me into the

field of commodity trading and finance and equipped me with basic transactional principles of the two domains.