

10 Academy: Artificial Intelligence Mastery

Week 10 Challenge Document

Date: 30 July - 05 Aug 2025

Change point analysis and statistical modelling
of time series data - detecting changes and
associating causes on time series data

Business objective

The main goal of this analysis is to study how important events affect Brent oil prices. This will focus on finding out how changes in oil prices are linked to big events like political decisions, conflicts in oil-producing regions, global economic sanctions, and changes in Organization of the Petroleum Exporting Countries (OPEC) policies. The aim is to provide clear insights that can help investors, analysts, and policymakers understand and react to these price changes better.

Situational Overview (Business Need)

You are a data scientist at **Birhan Energies**, a leading consultancy firm specialising in providing data-driven insights and strategic advice to stakeholders in the energy sector. With a mission to help clients navigate the complexities of the global energy market, Birhan Energies focuses on delivering actionable intelligence that supports decision-making processes for investors, policymakers, and energy companies.

You are tasked with analyzing how big political and economic events affect Brent oil prices. Understand how political decisions, conflicts in oil-producing areas, international sanctions, and OPEC policy changes affect the market.

The oil market is very unstable. This makes it hard for investors to make good decisions, manage risks, and maximize returns. Policymakers need detailed analysis to create strategies for economic stability and energy security. Energy companies need accurate price forecasts to plan operations, control costs, and secure supply chains.

As a data scientist at **Birhan Energies**, you are tasked with:

1. Finding key events that have significantly impacted Brent oil prices over the past decade.
2. Measuring how much these events affect price changes.
3. Providing clear, data-driven insights to guide investment strategies, policy development, and operational planning.

Data

The [data](#) set contains historical Brent oil prices. It includes daily prices from May 20, 1987, to September 30, 2022.

Data fields

- **Date:** Represents the date of the recorded Brent oil price. Each entry is formatted as 'day-month-year' (e.g., 20-May-87). The dataset covers daily prices from May 20, 1987, to September 30, 2022.
- **Price:** This column represents the price of Brent oil on the corresponding date. The price is recorded in USD per barrel.

Learning Outcomes:

Skills:

- Change Point Analysis & Interpretation
- Statistical Reasoning
- Using PyMC3 - a standard Bayesian modelling package in Python
- Analytical Storytelling with Data

Knowledge:

- Probability distributions and choosing the relevant one for a given task
- Bayesian inference
- Monte Carlo Markov Chain
- Model comparison
- Policy analysis

Communication:

- Reporting to government bodies

Team

Tutors:

- Mahlet
- Rediet
- Kerod

- Rehmet

Key Dates

- Discussion on the case - Wednesday 30 July 2025. Use #all-week10 to pre-ask questions.
- Interim Solution - 20:00 UTC on Sunday 01 Aug 2025.
- Final Submission - 20:00 UTC on Tuesday 05 Aug 2025

Instructions

Success in this challenge comes from making smart, strategic decisions about where to focus your effort.

Your priorities should be:

- A deep, well-explained analysis using the core Bayesian Change Point model
- Master the concepts and application of Bayesian inference and change point detection (PyMC3).
- Build a dashboard that allows you to deliver a functional, interactive product most efficiently. The quality of the insights matters more than the complexity of the tech stack.
- Clear Communication: A clear report and a simple, intuitive dashboard that tells a compelling story are the ultimate goals.

Objectives:

The global (business) objective is divided into the following sub-objectives

- Defining the data analysis workflow
- Understanding the model and data
- Extracting statistically valid insights in relation to the business objective

Task 1: Laying the Foundation for Analysis

Task 1 focuses on defining the data analysis workflow and understanding the model and data. This involves planning the analysis steps needed to achieve the project's objective and ensuring a clear understanding of the key concepts related to the project.

1. Defining the Data Analysis Workflow:
 - a. Clearly outline the steps and processes involved in analyzing the Brent oil prices data.
 - b. Research and Compile Event Data: Research major geopolitical events, OPEC decisions, and economic shocks relevant to the oil market. Compile a structured dataset (e.g., a CSV file) containing at least 10-15 key events with their approximate start dates. This will be crucial for your analysis.

- c. Identify and state any assumptions and limitations of the analysis. Crucially, this must include a discussion on the difference between identifying a statistical correlation in time and proving causal impact.
Determine the main media channels and formats for communicating results to stakeholders.
2. Understanding the Model and Data:
 - a. Read the main references related to the project to grasp the key concepts and models being used.
 - b. Analyze Time Series Properties: Before modeling, investigate the Brent oil price data for key properties like trend and stationarity. Briefly discuss how these properties inform your modeling choices.
 - c. Explain the purpose of change point models in the context of analyzing price fluctuations and how they help identify structural breaks in the data.
 - d. Describe the expected outputs of a change point analysis (e.g., dates of changes, new parameter values) and its limitations.

Task 2: Change Point Modeling and Insight Generation

Part 2.1: Core Analysis (Mandatory)

1. Implement the Change Point Model: Apply a Bayesian Change Point detection model using PyMC3 to identify statistically significant structural breaks in the Brent oil price series.
2. Identify Change Points: Interpret the model's output to determine the most probable dates of significant changes in the price behavior (e.g., changes in mean price or volatility).
3. Associate Changes with Causes: Compare the detected change point dates with your researched list of key events from Task 1. Formulate hypotheses about which events likely triggered the detected shifts.
4. Quantify the Impact: For each major change point you associate with an event, describe the impact quantitatively. For example: "Following the OPEC production cut announcement around [Date], the model detects a change point, with the average daily price shifting from \$X to \$Y, an increase of Z%."

Part 2.2: Advanced Extensions (Optional)

For a more advanced analysis, or to include in a "Future Work" section of your report, you can discuss how you might:

- **Explore Other Potential Factors:** Briefly describe how you would incorporate other data sources (e.g., GDP, inflation rates, exchange rates) to build a more comprehensive explanatory model.
- **Consider Advanced Models:** Mention how other models could provide different insights. For example, a VAR (Vector Autoregression) model could analyze the dynamic relationship between oil prices and macroeconomic variables, or a Markov-Switching model could explicitly define 'calm' vs. 'volatile' market regimes.

Suggested Approach for Task 2

This workflow outlines the key steps to successfully complete the core analysis in Task 2.

1. Data Preparation and EDA:

- Load the data and ensure the Date column is converted to a datetime format.
- Plot the raw Price series over time to visually identify major trends, shocks, and periods of high volatility. This will help you form initial hypotheses.
- **Consider analyzing log returns:** Price series are typically non-stationary. To model them more effectively, consider converting prices to log returns: $\log(\text{price}_t) - \log(\text{price}_{t-1})$. This transformation often results in a stationary series, which is easier to model, and is excellent for analyzing changes in volatility.
- Plot the log returns to observe volatility clustering (periods of high volatility followed by more high volatility).

2. Building the Bayesian Change Point Model (in PyMC3)

- **Define the Switch Point (tau):** This is the unknown point in time where the data's behavior changes. Define it as a discrete uniform prior over all possible days in your dataset.
- **Define the "Before" and "After" Parameters:** Define the parameters that describe the data's behavior. For a simple model changing its mean, you would define two means.
- **Use a Switch Function:** Use `pm.math.switch` to select the correct parameter (`mu_1` or `mu_2`) based on whether the time index is before or after the switch point `tau`.
- **Define the Likelihood:** Connect the model to the data. This is typically a `pm.Normal` distribution, where the `mu` (mean) is determined by the switch function. This tells the model how likely the observed data is given the parameters.
- **Run the Sampler:** Use `pm.sample()` to run the MCMC simulation to find the posterior distributions of your parameters (`tau`, `mu_1`, `mu_2`, etc.).

3. Interpreting the Model Output:

- **Check for Convergence:** Before interpreting results, always check that the model has converged. Use `pm.summary()` and look for `r_hat` values close to 1.0. Examine the trace plots using `pm.plot_trace()`.
- **Identify the Change Point:** Plot the posterior distribution of your switch point `tau`. A sharp, narrow peak indicates high certainty about when the change occurred.
- **Quantify the Impact:** Plot the posterior distributions for the "before" and "after" parameters (e.g., `mu_1` vs `mu_2`). Comparing these distributions allows you to make probabilistic statements like, "There is a 98% probability that the mean price after the change point was higher than before."

Task 3: Developing an Interactive Dashboard for Data Analysis Results

Build a Dashboard Application using Flask (backend) and React (frontend) to visualize the results of the analysis, helping stakeholders explore how various events affect Brent oil prices.

- Key Components of the Dashboard:
 - Backend (Flask):
 - Develop APIs to serve data from the analysis results, making it accessible for the React frontend.
 - Handle requests for different datasets, model outputs, and performance metrics.
 - Integrate data sources for real-time updates (optional, if necessary).
 - Frontend (React):
 - Create an intuitive and user-friendly interface to display the analysis results.
 - Design interactive visualizations to show how different events correlate with changes in oil prices.
 - Include features like filters, date ranges, and comparisons, allowing users to explore data around specific events or time periods.
 - Ensure responsiveness for various devices (desktop, tablet, mobile).
 - React tools for charts([Recharts](#), [React Chart.js 2](#), [D3.js](#))
- Key Features:
 - Present historical trends, forecasts, and correlations with events (e.g., political decisions, conflicts, economic sanctions).
 - Allow users to see how specific events influenced Brent oil prices, with features like "event highlight" to visualize spikes or drops in prices.

- Enable users to filter data, select date ranges, and drill down into details for deeper insights.
- Display key indicators like volatility, average price changes around key events.

Tutorials Schedule

In the following, the colour **purple** indicates morning sessions, and **blue** indicates afternoon sessions.

Wednesday: Data Analysis Techniques

- [Introduction to the challenge \(Mahlet\)](#)
- [Data Science Workflow \(Rediet\)](#)

Thursday: Change Point Analysis

- [Change Point Analysis \(Kerod\)](#)

Friday: Bayesian inference

- [Bayesian inference and Monte Carlo Markov Chain \(Rediet\)](#)
- [How to build react dashboards \(Rehmet\)](#)

Interim Submission

- Interim report - Covering task - 1
 - A 1-2 page document outlining your planned steps.
 - A structured Tabdata with dates and event descriptions.
- Link to your GitHub code.

Feedback

You may not receive detailed comments on your interim submission but will receive a grade.

Final Submission

- A blog post entry (which you can submit for example to Medium publishing) or a pdf report.
- Link to your Github code, and make sure to screenshots demonstrating anything else you have done.

Feedback


You will receive comments/feedback in addition to a grade.

References

Data Science Workflow

1. <https://www.datascience-pm.com/data-science-workflow/>
2. <https://towardsdatascience.com/mastering-the-data-science-workflow-2a47d8b613c4>
3. <https://www.knowledgehut.com/blog/data-science/data-science-workflow>
4. <https://medium.com/codex/a-comprehensive-guide-to-master-the-data-science-workflow-739295117d67>

Change Point Analysis

1. <https://forecastegy.com/posts/change-point-detection-time-series-python/>
2. <https://eecs.wsu.edu/~cook/pubs/kais16.2.pdf>
3. <https://jagota-arun.medium.com/change-point-detection-in-time-series-bcf01409010e>
4. <https://www.iese.fraunhofer.de/blog/change-point-detection/>
5.  Change Point Detection in Time Series

Bayesian Change point detection

1. [Bayesian Changepoint Detection with PyMC3](#)
2. [Model Comparison using PyMC3](#)

Bayesian inference and Monte Carlo Markov Chain

1. [As a introduction to Bayesian statistics and the python package \(PyMC3\)](#)
2. https://warwick.ac.uk/fac/sci/statistics/staff/academic-research/steel/steels_homepage/bayesiantsrev.pdf
3. <https://www.embecosm.com/2021/12/18/forget-arima-going-bayesian-with-time-series-analysis/>
4. https://www.lancaster.ac.uk/~romano/post/change_point_online_intro/
5. https://www.pymc.io/blog/chris_F_pydata2022.html
6. <https://machinelearningmastery.com/markov-chain-monte-carlo-for-probability/>
7. <https://towardsdatascience.com/monte-carlo-markov-chain-mcmc-explained-94e3a6c8de11>

Alternative Data Source

1. [Petroleum Trade: Imports From OPEC Countries](#)
2. [Energy Consumption Transportation Sector](#)

React dashboard templates

1. <https://github.com/flatlogic/react-dashboard>

2. <https://github.com/creativetimofficial/light-bootstrap-dashboard-react?tab=readme-ov-file>
3. <https://github.com/flatlogic/react-dashboard>
4. <https://github.com/marmelab/react-admin>
5. <https://github.com/coreui/coreui-free-react-admin-template>

