Quant Team Inductions

25 Batch

Round 1

Dataset

Here's the dataset you'll require for this task.

This dataset presents a synthetic, daily record of financial market activities pertaining to major firms engaged in artificial intelligence, specifically OpenAI, Google, and Meta.

It incorporates both quantitative financial indicators and qualitative event markers that are likely to influence firm-level stock performance, including the public release of models such as GPT (OpenAI), Gemini (Google), and Llama (Meta).

In addition, the dataset records expenditures on research and development (R&D) in AI-related products and services as well as revenue streams attributable to these activities.

The time horizon of the dataset extends from 1 January 2015 to 31 December 2024.

Tasks

Basics

1. Import & clean the dataset (handle missing/ null values if any).
2. Do basic statistical analysis - such as average R&D spending, average AI revenue, average stock impact per company. Also identify which company spent the most on R&D overall, and which company grew AI revenue the fastest.

Correlations

Use df.corr() to obtain correlations between the following three sets of parameters:

1. AI Revenue vs R&D Spending
2. AI Revenue vs Stock Impact
3. R&D Spending vs Stock Impact

Visualization & Plots

1. Plot the correlations using heatmaps.
2. Plot R&D spending with AI Revenue for each company over time.

   Identify which top 3 companies that have the most ROI (Revenue divided by R&D).

3. Use a scatter-plot to observe the realtion between R&D revenue and Stock Impact.
4. Name the top 2 events where maximum stock impact was observed overall (2 each for negative and positive).
   Which company reacts most strongly to events?

Takeaway & Analysis

Write a short write-up of less than 120 words about your takeaway and insights from this analysis.
What company has the best strategy?
What questions does this visualization or data answer?
etc.
You are encouraged to be opinionated and personalized as this is an insight for us into how **you** think. The use of technical terminology is encouraged but not mandatory, and its absence will not result in any penalty.

Submission Format

Submit a Jupyter Notebook (.ipynb) via the Google Form provided.

Ensure your code is clean, well-commented, and outputs (plots, tables, insights) are clearly visible.

Deadline: 30th August, 2025

Please Note: Any submissions generated with AI will be disqualified, and students found using AI will be ***blacklisted from all future inductions****.*