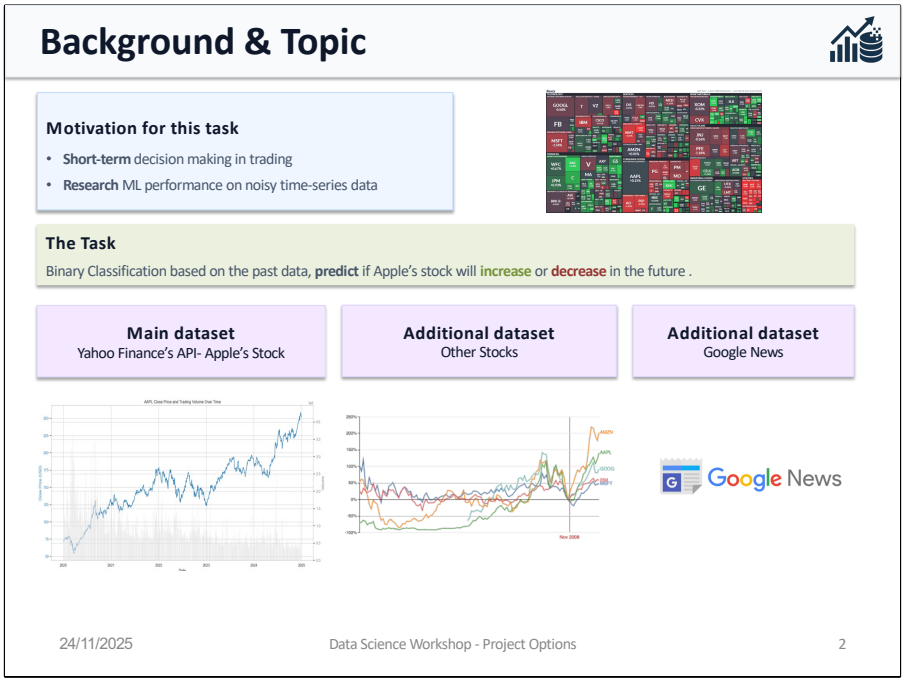
Moran

- Good afternoon. We are **Itay Mutzafi, Moran Zaks, and Shaked Schnarch**

- **Our first option** is in the financial markets: **Stocks Prediction**. The task is a Supervised Binary Classification on noisy time-series data.
- **Our second option** is in Neuroscience: **Detecting Autism Patterns in Brain Signals.**

We will now dive into the first option: Stocks Prediction.

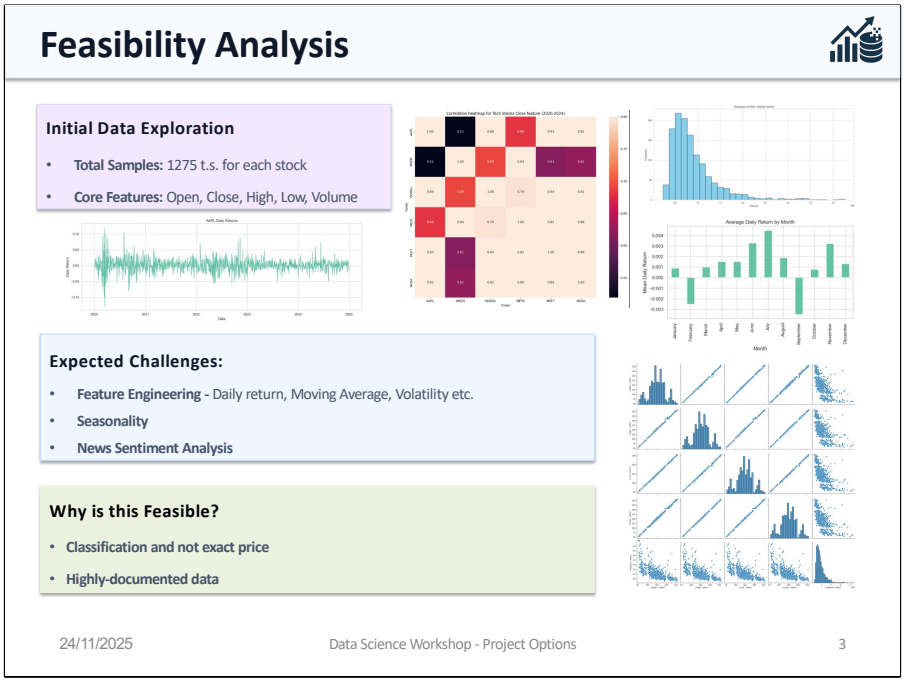


Moran

Our idea is to use public financial dataset to answer the question based on the past trading days and the new will a stock increase or decrease tommorow? We still need to dive in and check the time frames for the input and output

We will compare our resoults between training over
- only the specific stock's past
- other companies in the market's change
- news titles like 'new iphone is coming out

motivation is both trading decisions for investors etc., and for evaluating ML over noisy time-series data



Moran

By starting to look at the data we see we have basic features and will need to do feature engineering for more like daily return

daily return - predicts the percentage change in stock price from one day to the next

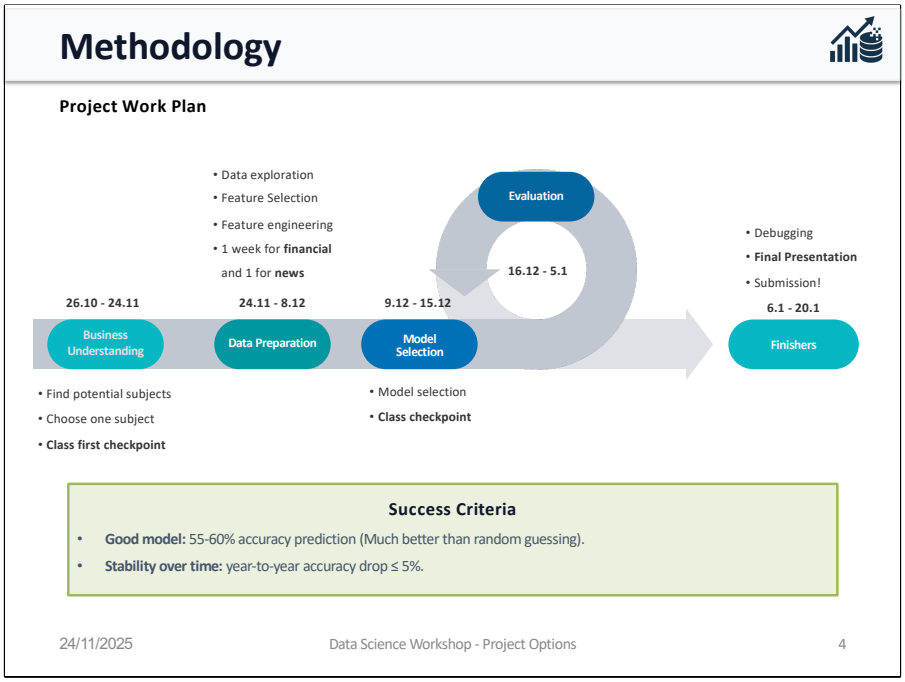
moving average is the average of the last N days of prices

volatility - תנודתיות

We will need to address the seasonality of the data (Holidays, Options expiry day)

supporting dataset will add the titles from the news. Using NLP method for sentiment analysis.

We believe this is feasible since we only predict the change and not the exact price, and we have a lot data to train on

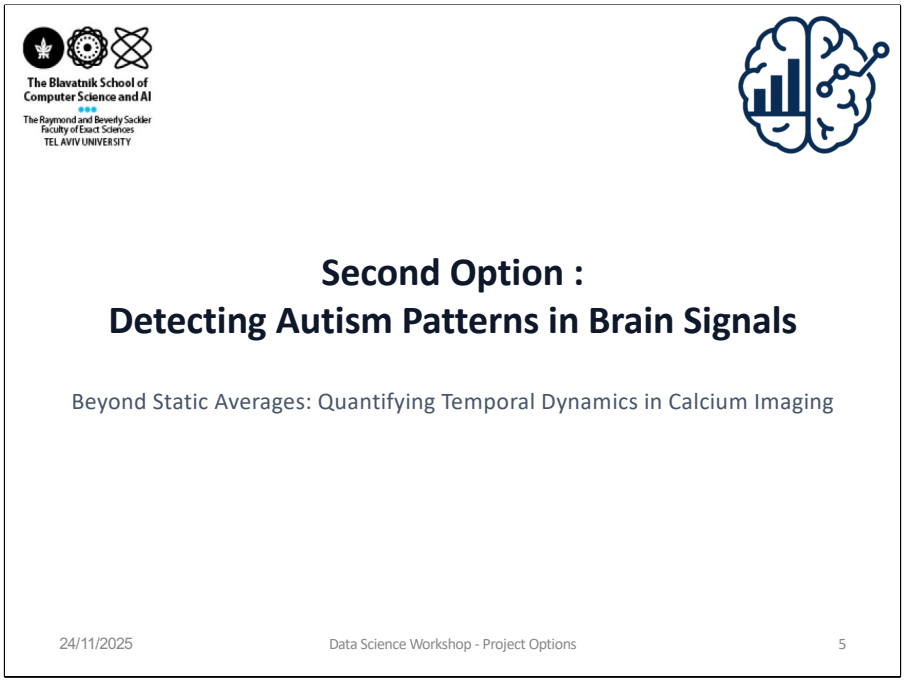


Itay

We will use the pipeline explained in class -Here is our methodology for the Stocks Prediction project.
We begin with *Business Understanding* — defining the task and reviewing the financial data from Yahoo Finance.

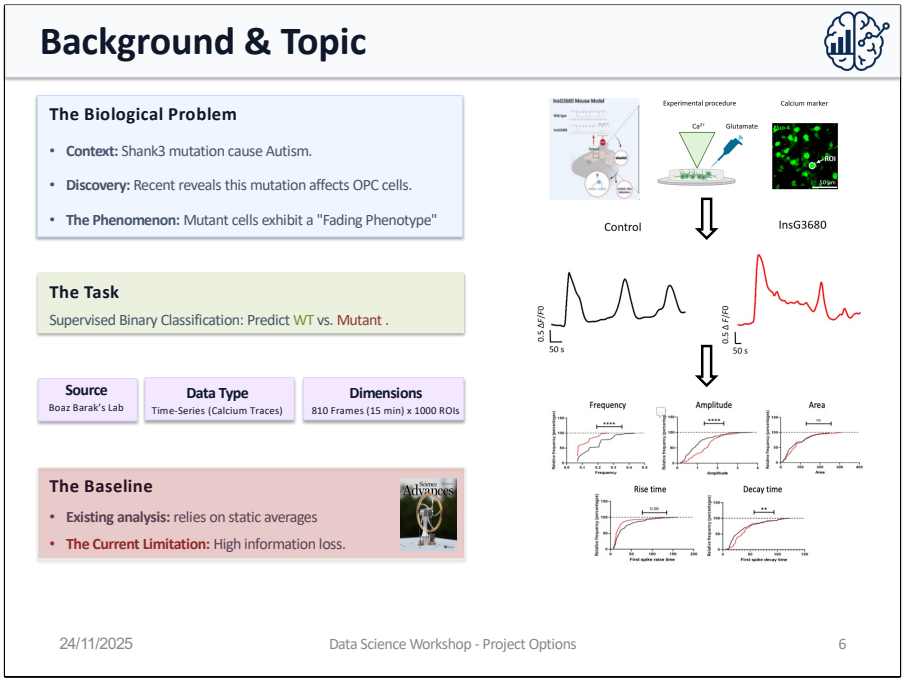
Data Preparation: cleaning the dataset, merging financial data with the news sentiment features, and scaling everything.

Model Selection, which works as a **loop**. We train a model, evaluate it, and **if it reaches our accuracy goal of 55–60%, we accept it and move on**. If not, we loop back — try another model, tune parameters, or improve the features. Only once a model meets the accuracy threshold do we proceed to *Finishers*: debugging, pipeline polishing, and preparing the final report and presentation.
Success Criteria: 1. Meaningful: the engineered features must improve accuracy and add **unique information**
2. Stability: Model performance remains stable when tested on **different years of data.**



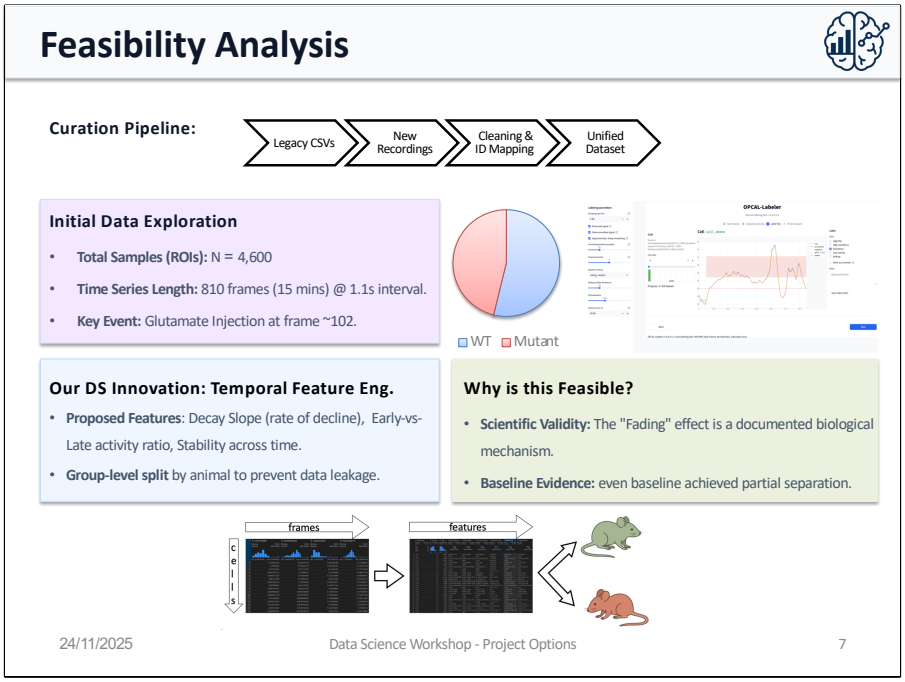
Shaked

- Now, we will move on to our second project option.
- While the first option dealt with financial markets, this project focuses on **Neuroscience**.
- Here, we aim to use Time-Series Classification to distinguish between healthy brain cells and those carrying an Autism-related mutation.



Shaked

- The Biological Problem is that a genetic mutation called Shank3 causes specific brain cells (OPCs) to exhibit what we call the 'Fading Phenotype.'
- If you look at the graphs on the right, the Control cell (black) shows stable activity, but the Mutant cell (red) starts strong and then collapses.
- Our Task is Supervised Binary Classification: we need to predict if a cell is 'Wild Type' (Healthy) or 'Mutant'.
- Here is the Data Science Gap: The existing analysis, the baseline, relies only on static averages. This is a major limitation because a 'crashed' signal and a 'stable' signal can have the exact same average, leading to significant **Information Loss.**
- We aim to capture the temporal shape—the actual dynamics of the signal—rather than just its static average.

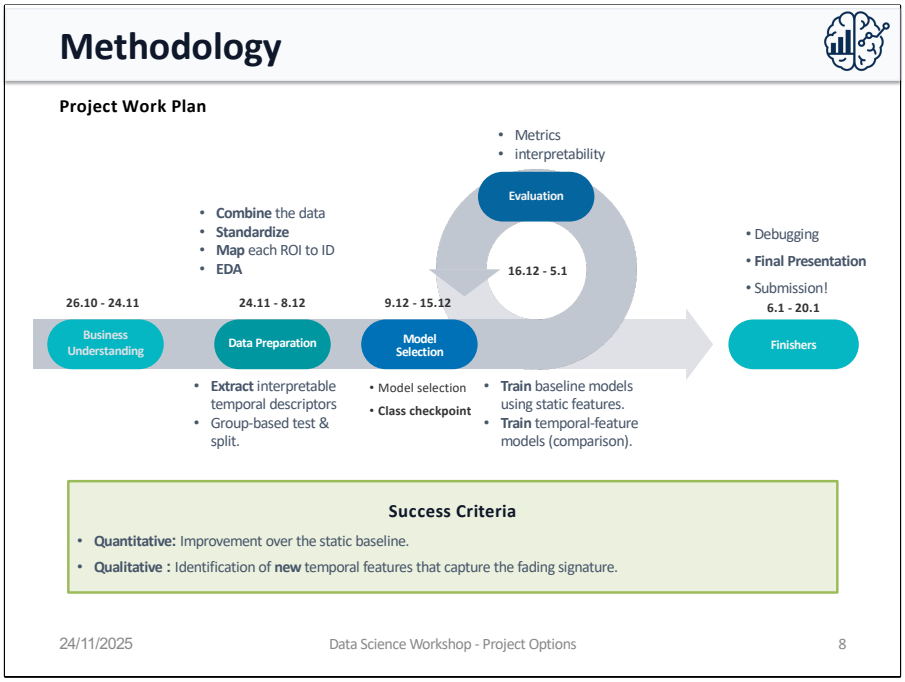


Shaked

- **Our curation pipeline** is…

- **Our Hypothesis** is that by using Machine Learning to capture the *rate of decay* and *stability over time*, we can significantly outperform the current baseline methods

- **Why is this project feasible?** First, the data is real and biologically validated. The 'Fading' effect is not random noise; it is a documented mechanism. Second, even simple manual analysis in the past showed some separation.

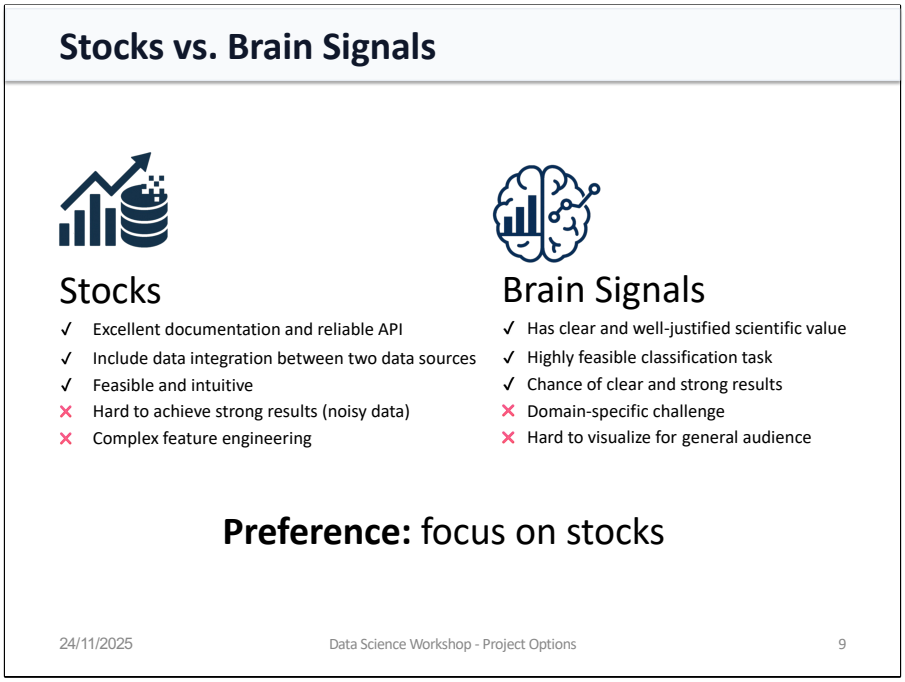


Itay

- Our methodology follows a standard and realistic DS pipeline.
- Just introduce the success criteria here, don't talk again about the steps.
- Success for us means three things: **Quantitative**, showing clear improvement over the static baseline.

  Qualitatively, identifying temporal features that reflect the fading pattern.

  Comparatively, reproducing the original analysis and demonstrating what additional information our features capture.



Itay

The **Stocks** :
+ excellent documentation, reliable APIs, and the added value of integrating **Google News sentiment**, which **enriches the feature** space and makes the prediction task more realistic.
- Noisy financial markets make strong performance difficult.

The **Brain Signals:**
+ has strong scientific value and a high chance of achieving clear results,
- harder to visualize and explain to a general audience.

Considering all factors, our **preferred option is the Stocks project**, mainly because of **its well-structured data pipeline**, the **integration of news sentiment**, strong documentation, and the Strong ML challenge.