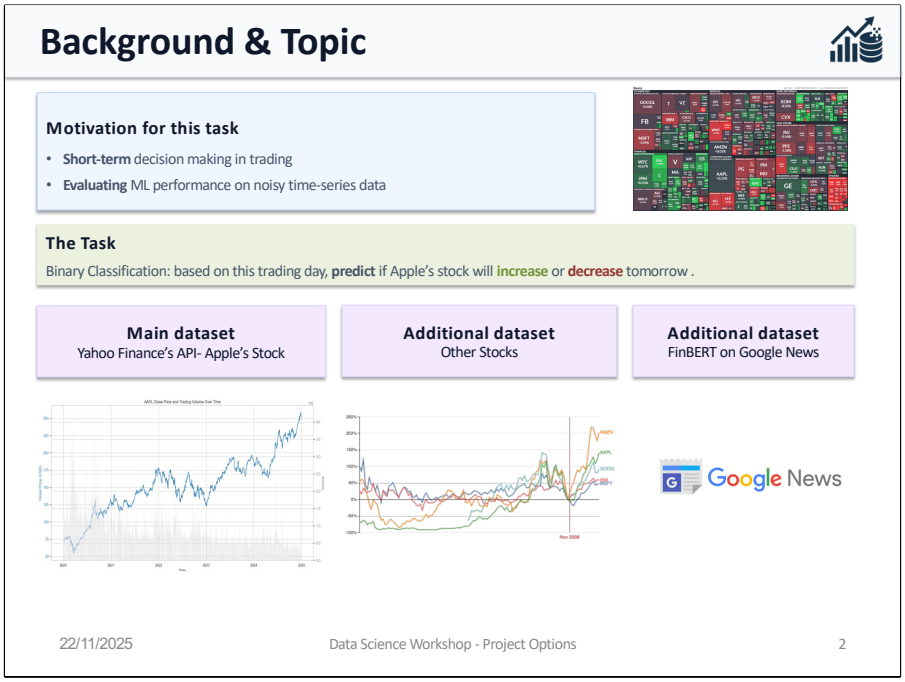
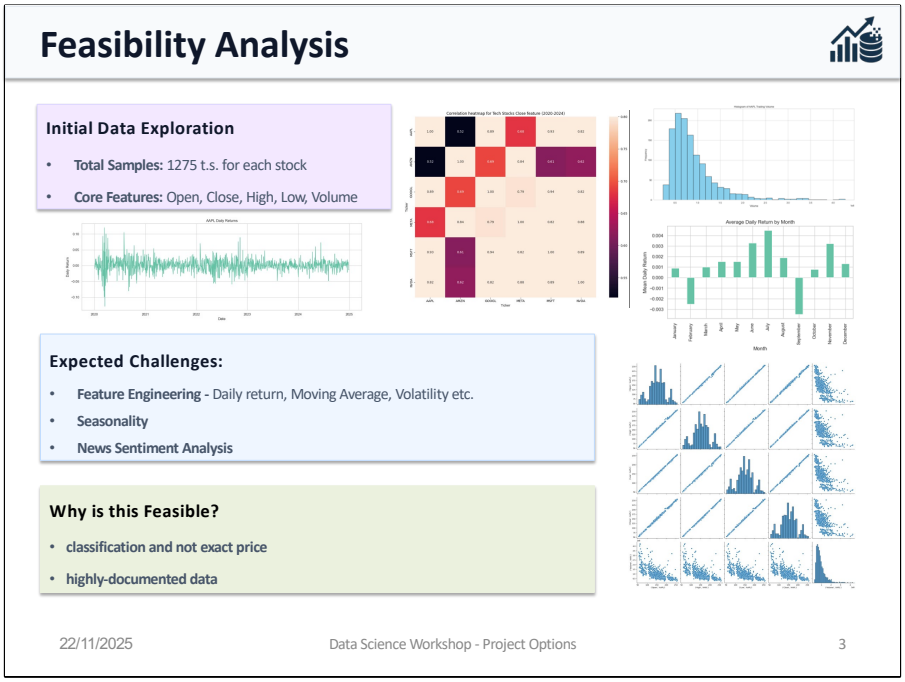
- Hi everyone.
- Our project addresses a classic Data Science challenge: Transforming raw, unstructured time-series data into structured features for classification.
- We aim to detect autism signatures in brain cells by comparing different Feature Engineering strategies, moving beyond the simple statistics currently used in research



Our idea is to use public financial dataset to answer the question 'will a stock increase or decrease tommorow?'

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



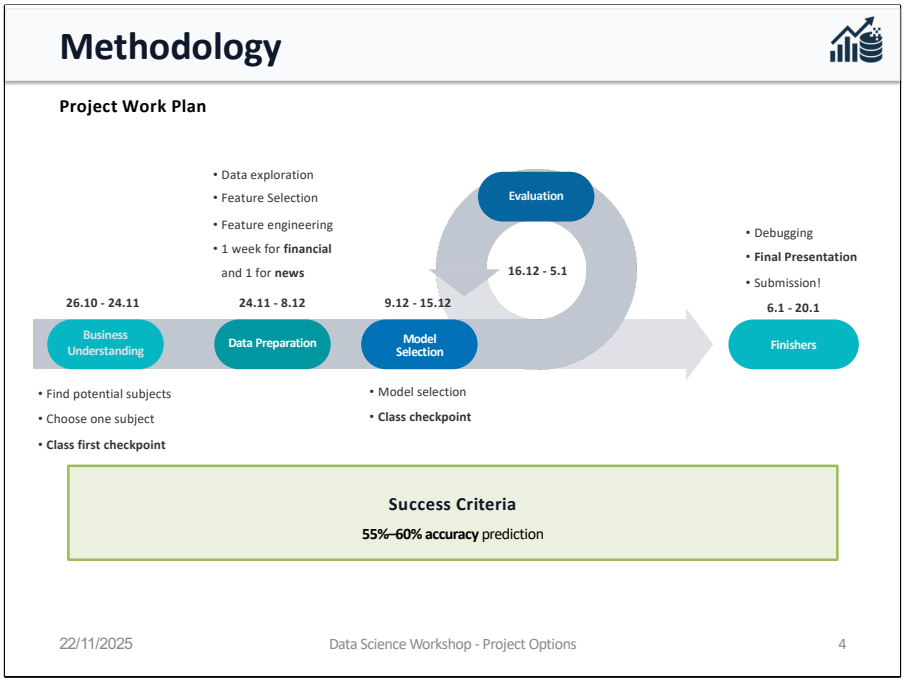
By starting to look at the data we see we have basic features and will need to do feature engineering for more like daily return

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

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

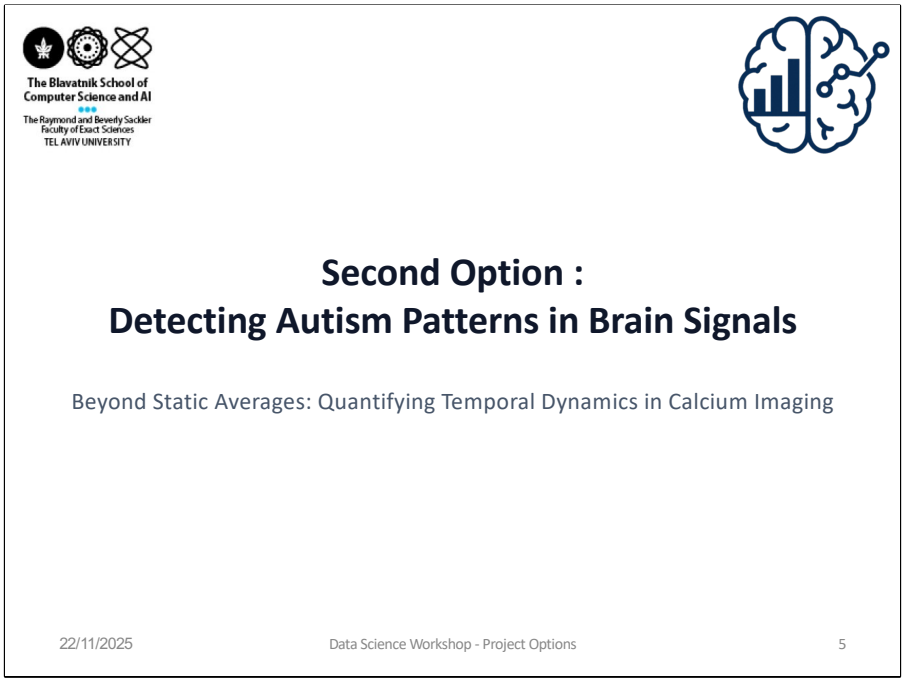
supporting dataset will add the titles from the news. Using NLP method for sentiment analysis.

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

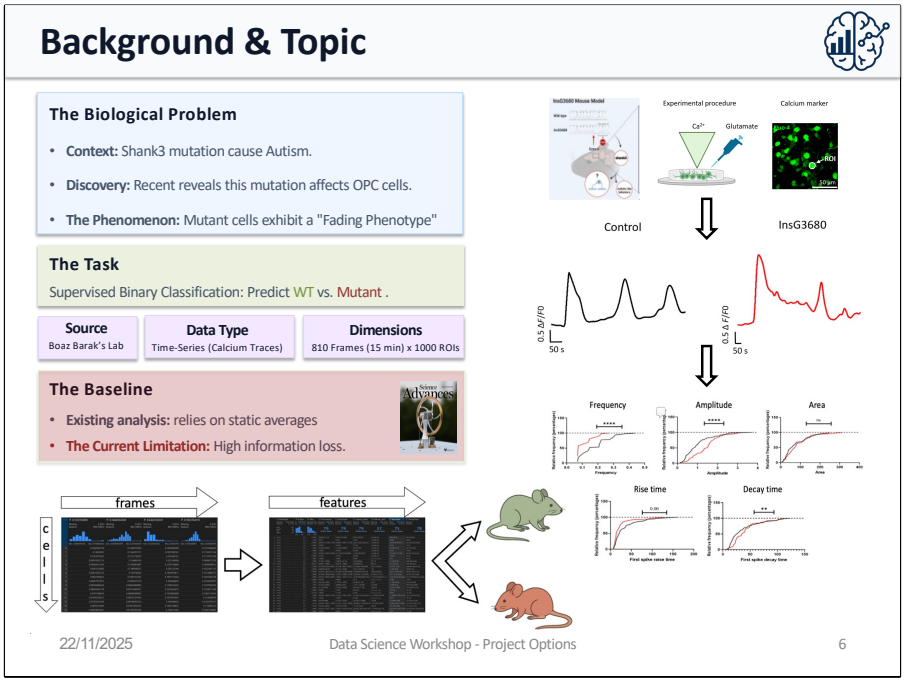


We will use the pipeline explained in class –

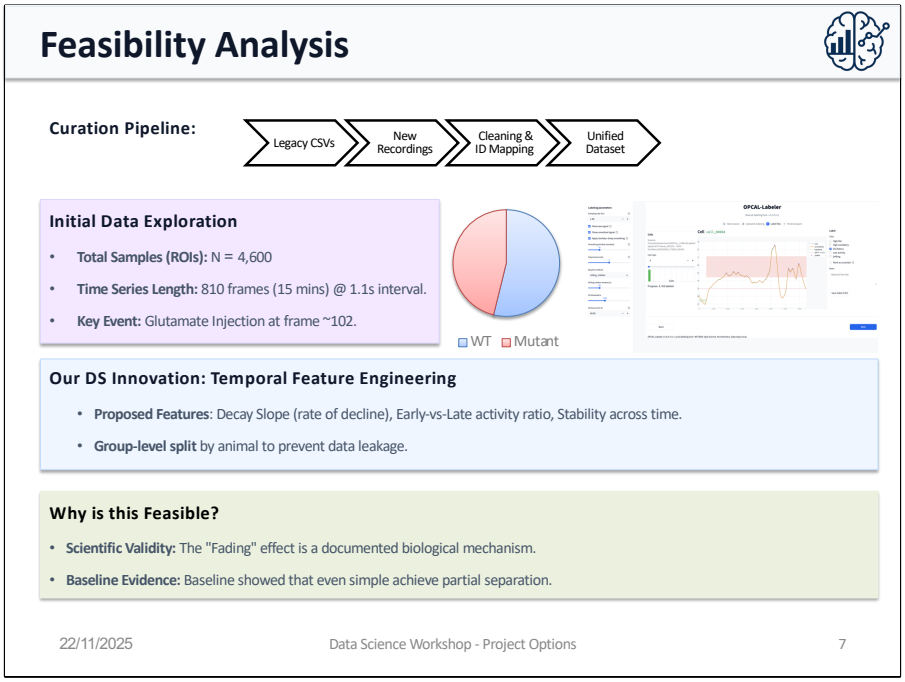
first having a clean datasets to feed the model – exploration, feature engineering
Secondly the model selection
And in the end evaluation the accuracy – for noisy data we think 55-60% of accuracy will consider a success, better then tossing a coin



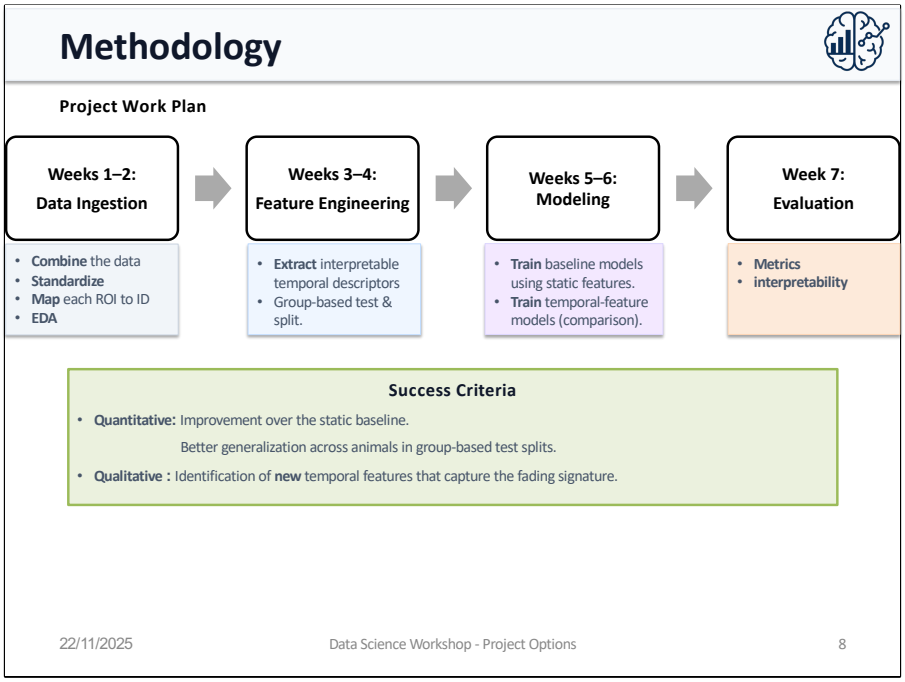
- Let's move on to our second project option.
- While the first option dealt with financial markets, this project focuses on **Neuroscience**.
- Here, we aim to use Time-Series Classification to distinguish between healthy brain cells and those carrying an Autism-related mutation.



- Let's start with the context. We are analyzing Calcium signals from specific brain cells called **OPCs**.
- Please look at the graphs on the right. The **Control cell** (in black) shows stable activity over time. The **Mutant cell** (in red) starts strong but then collapses. We call this the **'Fading Phenotype'**.
- **Here is the Data Science gap:** Currently, researchers analyze these signals by taking the **average** amplitude. But mathematically, a signal that 'crashes' and a signal that is 'stable' can have the exact same average. This leads to **Information Loss**.
- Our goal is to solve this by building a model that captures the **temporal shape**—the actual dynamics of the signal—rather than just its static average.
- **Our Task** is Supervised Binary Classification. We need to predict if a cell is 'Wild Type' (Healthy) or 'Mutant'.



- **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.



- Our methodology follows a standard and realistic DS pipeline.

- Success for us means three things:
  Quantitatively, 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 temporal features capture.

Stocks vs. Brain Signals

Stocks

- ✓ Excellent documentation and reliable API
- ✓ Include data integration between two data sources
- ✓ Feasible and intuitive
- ✗ Hard to achieve strong results
- ✗ Complex feature engineering
- ✗ Train-test split

Brain Signals

- ✓ Clear and well-justified scientific value
- ✓ Highly feasible classification task
- ✓ Chance of clear and strong results
- ✗ Domain-specific challenge
- ✗ Hard to visualize for general audience
- ✗ Train-test split

Current Decision: focus on stocks