Lund University Finance Society

Advanced Python Workshops Project

VT 25

Created by: *Elliot Brunstorp*

Description

You may work individually or in groups of up to three. Choose *any* dataset that interests you (public, private, scraped, simulated, or API-driven) and pose a clear analytical question. There is **no restriction on size**, pick something that supports meaningful exploration. Your task is to

- explore and clean the data,
- build and justify an analytical or predictive model,
- evaluate your work with appropriate metrics, and
- communicate the key insights through an interactive dashboard.

There are no predefined "correct" answers; assessment is based on the rigour and clarity of your process. The project shall be built using Anaconda (Jupyter, Spyder).

Project Steps

1. **Exploratory Data Analysis** Investigate basic properties of your dataset. Compute relevant proportions, distributions, and summary statistics for variables you believe matter most.

2. **Feature Selection and Cleaning**

   2.a) Handle missing values, inconsistent types, and outliers.

   2.b) Identify variables with low predictive or explanatory power via correlation checks, statistical tests (ANOVA, χ^2, Pearson's r), or dimensionality reduction (e.g. PCA).

3. **Model Building** Create at least one suitable model. Possible choices include
   - logistic or linear regression,
   - clustering algorithms,
   - decision trees or random forests,
   - neural networks,
   - any other well-motivated method.

4. **Evaluation** Use metrics that match your objective. Depending on which model you use, you can carry out multiple tests.

5. **Dashboard and Visualisation** Build a dashboard in Plotly and/or Dash that shows for example:
   - key descriptive statistics,
   - most relevant variables or feature importances,
   - model performance visuals (for example confusion matrix, score table),
   - any additional plots that support your narrative.

Submission

Your group's submission must include:

1. all Python source files or notebooks,
2. the dashboard as `.html` (and, if built with Dash, its source code and a screenshot),
3. any data files *or* a script that retrieves them,
4. any other files or requirements necessary to run your code.

Submit a single compressed archive via Canvas by 23:59 CET on 13 June 2025.

Examples

For inspiration, see the gallery at `https://dash.gallery/Portal/` and the financial report sample at `https://github.com/plotly/dash-sample-apps/tree/main/apps/dash-financial-report`