1. Suitable Title - Should reflect the business and analytic goals
Examining opportunities of the participation in sequential markets and finding the optimal balance for it

2. Business Goal:
(a) Who is the stakeholder or client?
An energy company
(b) A description of the business goal. What are the business benefits of implementing this idea?
Maximizing profits from energy sales
What opportunity is it creating?
Splitting available energy between the day-ahead and intra-day auctions
What shortcoming does it address?
It would increase effectiveness of dealing with the energy market's price volatility and decrease current opportunity costs
(c) What would be considered a success?
Identifying whether participation in the intra-day auction is viable and, if it is, deriving a reccomendation for pricing auction offers

3. Analytics/Data Mining Goal:
(a) A description of the analytics objective.
Identify parameters responsible for the price difference between the day-ahead and intra-day auctions. Develop a regression model based on these parameters
(b) Is this a supervised or unsupervised task?
Supervised (price data is available)
Is it predictive or descriptive?
First part is descriptive, second part is predictive
Is it retrospective or forward-looking?
Both parts are retrospective, however the regression model can be used for prospective predictions too (assuming the utilized paremeters' data is available at the point of prediction)
(c) What is the main outcome variable(s) of interest?
One or several parameters which have a causal relationship with the prices at both auctions and the kind of their relationship
4. Data:
(a) Brief description of available data.

The volume (in MWh) and price (in €/MWh) of the energy sold at each auction sale. For the day-ahead auction the data is hourly, for the intraday auction the data is available in 15-minute increments

(b) Some guidance on the data subset that will be used and the re-processing or preparation that might be needed based on your past experience.

First of all the data needs to be scraped from the respective websites. Then the hourly data needs to be converted into 15-minute increments and merged with the intraday auction data into one DataFrame. Since there is no missing data, no respective preparations are required. A Pandas TimeStamp will be used as the primary key. An additional column will be created for the price premiums.

(c) Sample of ten rows (records) with ten columns (variables) that will be used.

Timestamp	15-minute increment	MWh (day-ahead)	€/MWh (day-ahead)	MWh (intraday)	€/MWh (intraday)
2017-01-01 00:00:00	00:00	69,20	2736,55	46,56	616,5
2017-01-01 00:15:00	00:15	69,20	2736,55	39,96	282,6
2017-01-01 00:30:00	00:30	69,20	2736,55	34,44	257,1
2017-01-01 00:45:00	00:45	69,20	2736,55	27,18	426,2
2017-01-01 01:00:00	01:00	61,07	2642,975	38,68	456,9
2017-01-01 01:15:00	01:15	61,07	2642,975	34,1	326,7
2017-01-01 01:30:00	01:30	61,07	2642,975	40,52	396,2
2017-01-01 01:45:00	01:45	61,07	2642,975	36,41	404,5
2017-01-01 02:00:00	02:00	55,47	2651,8	39,83	348,2
2017-01-01 02:15:00	02:15	55,47	2651,8	37,76	231,4
2017-01-01 02:30:00	02:30	55,47	2651,8	35,24	277,4
2017-01-01 02:45:00	02:45	55,47	2651,8	32,79	288

5. Methods:
(a) What are some data mining methods to consider?
-Supervised
-Batch learning
-Model based learning
-(Teil 1) Deskriptiv, (Teil 2) Prediktiv
(b) Which performance measures are appropriate?

For the descriptive measuring goodness of fit is appropriate. For the predictive model the prediction accuracy on the holdout set compared to the training set.

How do they map to the business goal?

The goodness of fit shows if the derived parameters are responsible for the price difference between the auctions. The prediction accuracy shows how well the causal relationship between the parameters and the price has been estimated

6. Implementation/Production:

(a) Operational requirements or constraints (for example, will the solution be run in real-time? will it require collecting new data? will it be a one-time analysis or ongoing?)

The predictive model will not be continuously run in real-time, however it certainly can be used for that. This would however require automated data collection, automated data preparation, a thorough monitoring system, and a backup system in case the prediction performance deteriorates too much