1) We used pandas to read the data from the csv-files into data frames. A roundtrip converter needs to be chosen in order not to loose precision.
2) We constructed a feature matrix X_train and a response vector y_train, as well as y_test from the data frames.
3) To visualize the relationship between the features and the response we used the seaborn framework. We decided to do a non-linear feature transformation on X_train and X_test using the sklearn.preprocessing module. The degree of the polynomial features was set to 3.
4) As a model we have chosen the ridge regression from module sklearn.linear_model. RidgeCV was used to choose the regularization parameter by cross-validation. We have decided to use a 10-fold cross-validation with a negative mean squared error regression loss as a scoring metric.
5) We trained our predictor using the training data.
6) We printed out the cross validation mean score (sklearn.model_selection) and the root mean squared error on the training data to be able to value the quality of the predictor.
7) We predicted the values on the test data and have written the predicted values to a csv-file using pandas

some remarks:

- The program was written in Python.
- We tested different linear models like Linear Regression or Lasso but Ridge Regression has yield the best cross validation score. We also tested Support Vector Regression with different kernels (rbf, linear, polynomial) but this has led to overfitting of the training data.
- In order to validate a lot of different regularization parameter values we used the Euler computer from ETH (ridgecv alpha = 266.83496156, cross val mean score = 0.920652272746, RMSE on training data = 7.47898298079).