Ich hab mich an der Kaggle-Competition "[Predicting Bike Rental Demand](https://www.kaggle.com/c/bike-sharing-demand)" Link versucht, bei der es darum geht, die Anzahl der geliehenen Stadträder in Washington D.C. vorherzusagen, abhängig von Zeit und Wetter.

Ich habe mich für dieses Projekt entschieden, weil der Datensatz und der Bewertungsmaßstab dem Bluebook in der ersten Lektion von "Machine Learning for Coders" sehr ähnlich war, sodass ich die gezeigten Arbeitsschritte gut übertragen konnte. Ich habe die Aufgabe also mit einem Random Forest Regressor gelöst.

Das ist mein allererstes Machine Learning Projekt. Im ersten habe ich das Trainingsset wie in der Lektion geteilt, weil zuerst fälschlicherweise gedacht habe, das müsste immer so gemacht werden. Erst später habe ich verstanden, dass das nur im gezeigten Beispiel so war. In Fall von "Predicting Bike Rental Demand" gab es bereits ein extra Testset, also hätte ich das gar nicht tun müssen.

Im ersten Durchlauf hatte ich also nur die Hälfte an Trainingsdaten, konnte dafür aber selbst den Score berechnen, weil ich die tatsächlichen y-Werte aus der Testhälfte als Vergleich heranziehen konnte. Ich habe damit einen RMSLE von ca. 0,35 erzielt, womit ich für das erste Mal sehr zufrieden bin.

Im zweiten Durchlauf habe ich vor allem Formfehler bereinigt und die Datensätze wie vorgesehen verwendet. Dafür bekomme ich aber kein Ergebnis mehr, weil die Competition schon geschlossen ist.

Was habe ich dabei gelernt?  
Sei hartnäckig und beiß dich durch. Sprich mit deiner Styroporente, sie ist dein bester Freund. Im Zweifelsfall, schau dir deinen Code und deine Daten an, viele Probleme lösen sich dadurch. Es ist kein Hexenwerk, sondern nur viel Puzzlearbeit. Am Anfang habe ich kaum etwas verstanden und die Lektion erschien mir wie ein unüberwindlicher Berg. Und nun bin ich fertig und es sieht nach erstaunlich wenig aus. Die wahre Arbeit und meinen Fortschritt sieht man im Endresultat nicht. Es ist winzig und sauber.