

Name	Project name	Data sets	Goals
Pallavi	Santander Product Recommendation	Kaggle	Learn and implement recommendation engine
Greg	Text predictor and features extraction	Twitter	Provided a started sentence, provide the next word, and try to guess the gender, location and age of the person typing
Ana	Compare ML methods for text classification	Public (my webpage)	Show how different ML methods compare on text classification tasks. Apply new libraries to a well known task. Be able to reproduce results in scientific papers.
Colin	If the shoe fits..	Web Scraping of various websites.	The first objective is: To predict the price of mens shoes based upon a variety of categorical properties with an accuracy of over 20% above the baseline. The second objective is: Given a sample photo of a shoe I would like to be able to recommend a similar shoe from an online retailer with a 70% accuracy.
Matthew	Ja, logo, Ja!	Public (scraping)	Discover correlations between logo features and popularity / commercial success. Exact target variables TBD depending on EDA findings
Daniel	Munchy - Molecular Gastronomy According to Cultural backgrounds	- Yelp (will scrape but against ToS so will see) or use challenger data	Conduct exploratory data analysis, create hypothesis and use ML after data has been cleaned to see if there are any patterns for what could make a food store/restaurant more successful
		- Mayor of London data (public)	
		- Cookpad (scrape)	
		- Google maps data (scrape)	
Andras	Untitled Capstone Project #51	an activity logging platform such as Strava, Garmin, Movescount, etc.	Predict Marathon finishing times based on training data and calculate how you do compared to others with similar invested effort
Raj	Project Ratatouille	Yummly API, ImageNet	Provide a health score for the meal based on an image of the meal / gradients
Dina	VC portfolio analysis	Crunchbase and Angellist API	Analyze features that help predict startup succuss and best combination of investments for venture capital portfolio
Ian	Project Youtube	Youtube	Predict how various features influence the ranking and / or views.
Chloe	Brain tumours MRI scans analysis	The cancer imaging archive - Rembrandt study	Predict tumour type based on MRI scans analysis
Dorian	Location suggestions generator for opening restaurant	TripAdvisor, Justeat,DataGov	In function of the entry (Type of cuisine, postcode, average price?,... return a likelihood of success of a place(1-10) + propose the best areas.
Yulia	Breath sound classification and prediction	Audio signals	Identify and classify a signal into breath or not breath. Predict breath pattern and identify a change in pattern.
Martin	Can Changes in Non-OPEC production level affect oil prices?	BP, EIA	To be able to predict changes in oil prices based on Non-OPEC oil production changes.

Vincent	Predicting Premier League match outcomes using weather data.	Kaggle, wunderground.com	Use different classification methods to beat the baseline accuracy of football outcome predictions
Mark	Investment sense without the expense	Morningstar performance tables	To determine which markets are efficient enough to reduce the dependence on active fund managers
Ignacio	Airbnb prices	Airbnb dataset	to recommend the optimal price to optimize revenue
Revathy	Magic Sound of Cocount Somthing fishy	Audio signals collected by me Kaggle	Classify the type of coconut using sound Classify the type of fish species using image recognition
Jaye	Deforestation	Global Forest Watch	Determine social, economic and environmental features which are potential predictors of the rate in which trees / forests are being created / destroyed.
Wally	The Counted [- a fork]	"The Counted" 's database, The Washington Post's database	i) Performing unsupervised learning analysis on the data ii) Integrating crime level data into a model predicting frequency of killings by state or police department.