

2. Project

Final Deliverables:	1
Things to think about:	1
1. Kickstarter Project Success	2
2. Personality vs. Risk of Drug Use	2
3. Zindi challenge	3

Final Deliverables:

1. Slide deck PDF pushed to GitHub designed for non-technical stakeholders that outline findings, recommendations, and future work (10 min presentation).
2. Jupyter notebook following PEP8 designed for data science/technical audience.
3. Python script for generating (takes .csv as argument and saves the model locally) and running your model from the terminal (takes test.csv and model as arguments and outputs accuracy and predictions as .csv) (you can still use the template repo : <https://github.com/neuefische/ds-modeling-pipeline-template/>)

Things to think about:

- Check for data imbalance!
- What would be the right performance metric - precision, recall, accuracy, F1 score, or something else? (Check TPR?)
- Try different (at least 3) machine learning algorithms to check which performs best on the problem at hand

Submission Date:

- Monday afternoon: presentations

1. Kickstarter Project Success (Medium)

In recent years, the range of funding options for projects created by individuals and small companies has expanded considerably. In addition to savings, bank loans, friends & family funding and other traditional options, crowdfunding has become a popular and readily available alternative.

Kickstarter, founded in 2009, is one particularly well-known and popular crowdfunding platform. It has an all-or-nothing funding model, whereby a project is only funded if it meets its goal amount; otherwise no money is given by backers to a project.

A huge variety of factors contribute to the success or failure of a project—in general, and also on Kickstarter. Some of these are able to be quantified or categorized, which allows for the construction of a model to attempt to predict whether a project will succeed or not. The aim of this project is to construct such a model and also to analyse Kickstarter project data more generally, in order to help potential project creators assess whether or not Kickstarter is a good funding option for them, and what their chances of success are.

2. Personality vs. Risk of Drug Use (Easy-Medium)

The abuse of tobacco, alcohol and illicit drugs is costly to our society. According to the National Institute on Drug Abuse, more than 740 billion is lost annually due to lost work productivity, crime and healthcare.

People with certain personality traits may be at increased risk for drug use problems, and studying personality may help researchers better understand and treat these problems. Attempting to link genetic disposition to substance use disorders has proven unsuccessful. There are however many studies that have shown positive correlations between risky personalities and drug use. Understanding this relation could lead to better treatment.

The dataset used for this project is referenced. There are limitations to the study since the collected sample is biased with respect to the general world population but it remains useful for risk evaluation. Each column of attributes has been normalized using T-scores and is described below.

T Score

The term t score has different meanings in different settings:

- In introductory statistics, t score is often used synonymously with [t statistic](#).
- In psychometrics, a t score is a type of standard score computed by multiplying a [z-score](#) by 10 and adding 50.
- In bone density tests, a t score compares bone mineral density to a reference mean.

Often, the meaning is clear from the context. In educational research, the psychometric definition generally applies. In papers on osteoporosis, the bone density definition is probably the intended definition.

Sometimes, though, the meaning is not clear. Some statistics texts use the term to mean "t statistic". Others use the psychometric definition.

To avoid confusion, the Stat Trek website avoids the term "t score". It uses "t statistic", instead.]

EScore: Escore (Real) is NEO-FFI-R Extraversion. Extraversion is one of the five personality traits of the Big Five personality theory. It indicates how outgoing and social a person is. A person who scores high in extraversion on a personality test is the life of the party.

Might have to figure out the other kinds of scores from the internet.

Some references:

1. Drug, Wikipedia URL: <https://en.wikipedia.org/wiki/Drug>
2. The Five Factor Model of personality Model of Personality and Evaluation of Drug Consumption risk, E.Fehrman, A.K. Muhammad, E.M. Mirkes, V. Egan, A.N Gorban. URL: <https://arxiv.org/abs/1506.06297>
3. Detecting and Assessing Alcohol and Other Drug Use. URL: <https://www.ncbi.nlm.nih.gov/books/NBK236259/>
4. UCI-Machine Learning Repository URL: <archive.ics.uci.edu/ml/datasets/Drug+consumption+%28quantified%29>

3. Zindi challenge (Medium - Advanced)

Please choose one of the following challenges on zindi. You need to follow the requirements form the challenge and also make a stakeholder presentation

- A. <https://zindi.africa/competitions/womxn-in-big-data-south-africa-female-headed-households-in-south-africa/> (Advanced, but has a starter notebook)
- B. <https://zindi.africa/competitions/xente-fraud-detection-challenge/> (medium)
- C. <https://zindi.africa/competitions/airqo-ugandan-air-quality-forecast-challenge> (Advanced, also with starter notebook)

Assignment of topics:

Kickstarter	Drug Risk	Zindi challenge A	Zindi challenge B	Zindi challenge C

Groups:

Group 1			
Group 2			
Group 3			
Group 4			
Group 5			