# How to get (students) happy

A data scientific investigation of factors of happiness in young people aged 15 to 30

Goal: We want to investigate important factors in the lives of young people, which are demographically close to the student population of maastricht university. With help of this analysis and visualizations we want to help in guiding ways how students' happiness could be improved.

Dataset: https://www.kaggle.com/miroslavsabo/young-people-survey
Health and habit data of more than a thousand young people between 15-30 years acquired by the Comenius University in Bratislava. They surveyed 1010 teenagers and young adults in regards to 150 topics, falling under 8 main themes: Music preferences (19 items), Movie preferences (12 items), Hobbies & interests (32 items), Phobias (10 items), Health habits (3 items), Personality traits, views on life, & opinions (57 items), Spending habits (7 items), Demographics (10 items)

We spent most time investigating *Personality traits,views on life and opinions* as well as *Demographics*.

Findings:  Young people's happiness is highly dependent on their inclusiveness within a social structure. People who feel lonely are a lot more likely to also feel unhappy (Pearson correlation coefficient: -0.43, p-value: 7.68e-48). The same (although less pronounced) is observed for people with a lower number of friends (correlation: 0.32, p-value: 1.56e-25). Furthermore we observed a significant linear relationship between how positive a person evaluated his own personality and how positive his outlook on life was(correlation: 0.30, p-value: 8.37e-23). Demographic variables like gender, whether a person grew up in a city or the countryside, the financial background etc. did not make a significant difference. This suggests that mainly soft factors, like confidence, openness, etc. are important in experiencing happiness. Any organisation aiming to set optimal conditions for staff and students should therefore try to help build on such. Furthermore, organisations should value social interactions a lot more. In the light of these findings adding additional social components to the curriculum could significantly benefit students and staff's happiness. With the gradient boost learning algorithm xgboost we were able to predict the exact class of happiness (on a scale from 1 to 5) given the person's personality and demographic information as described by the author of the dataset to an accuracy of 60%(see appendix 1 for more information) and faced with a binary classification of happy/unhappy (where people with a happiness score below 3 are considered unhappy) we achieved an accuracy of 73%(see appendix 2). When analysing the most important features of the classification using the game theoretical based SHAP[1] interpretation we found that Loneliness, Energy level, Number of friends and Positivity of personality were the most influential features for the model. We also looked into interaction between those features so we observed that when energy levels are already low the impact of the loneliness variable on the model is relatively low, however when the energy level is high the impact of the feature on the model performance can be very big. Source code:https://github.com/blazejdolicki/data-madness

---

[1] Lundberg, S. M., Erion, G. G., & Lee, S. I. (2018). Consistent individualized feature attribution for tree ensembles.

# APPENDIX:

*Additional note on extended movie time:* As the matter of happiness and social factors in general is a difficult one our movie shows slight overtime of about 10 seconds(excluding credits). This was als done to still be able to convey a funny and hopefully relatable story to make the experience for any viewer a bit more exciting and less dull. We hope you enjoy it.

1 - Happiness classification evaluation(highly unbalanced data):

|  | precision | recall | f1-score | support |
|---|---|---|---|---|
| 1.0 | 0.00 | 0.00 | 0.00 | 1 |
| 2.0 | 0.00 | 0.00 | 0.00 | 6 |
| 3.0 | 0.47 | 0.28 | 0.35 | 29 |
| 4.0 | 0.55 | 0.86 | 0.67 | 51 |
| 5.0 | 1.00 | 0.14 | 0.25 | 14 |
| accuracy |  |  | 0.53 | 101 |
| macro avg | 0.40 | 0.26 | 0.25 | 101 |
| weighted avg | 0.55 | 0.53 | 0.47 | 101 |

2 - Binary happiness evaluation:

|  | precision | recall | f1-score | support |
|---|---|---|---|---|
| False | 0.67 | 0.44 | 0.53 | 36 |
| True | 0.74 | 0.88 | 0.80 | 65 |
| accuracy |  |  | 0.72 | 101 |
| macro avg | 0.70 | 0.66 | 0.67 | 101 |
| weighted avg | 0.71 | 0.72 | 0.71 | 101 |

3-Absolute feature importance and interaction example approximated according to SHAP:
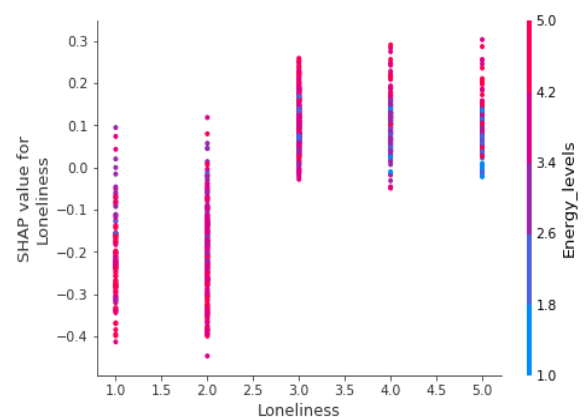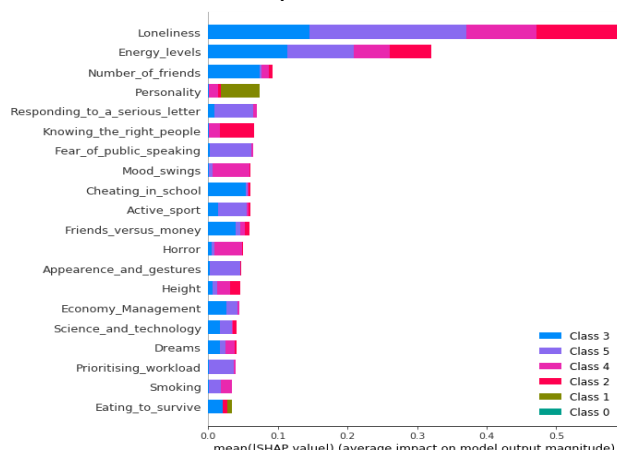


Left: Absolute feature importance for each happiness level. Right: Interaction of loneliness and Energy level and the effect on the feature importance.