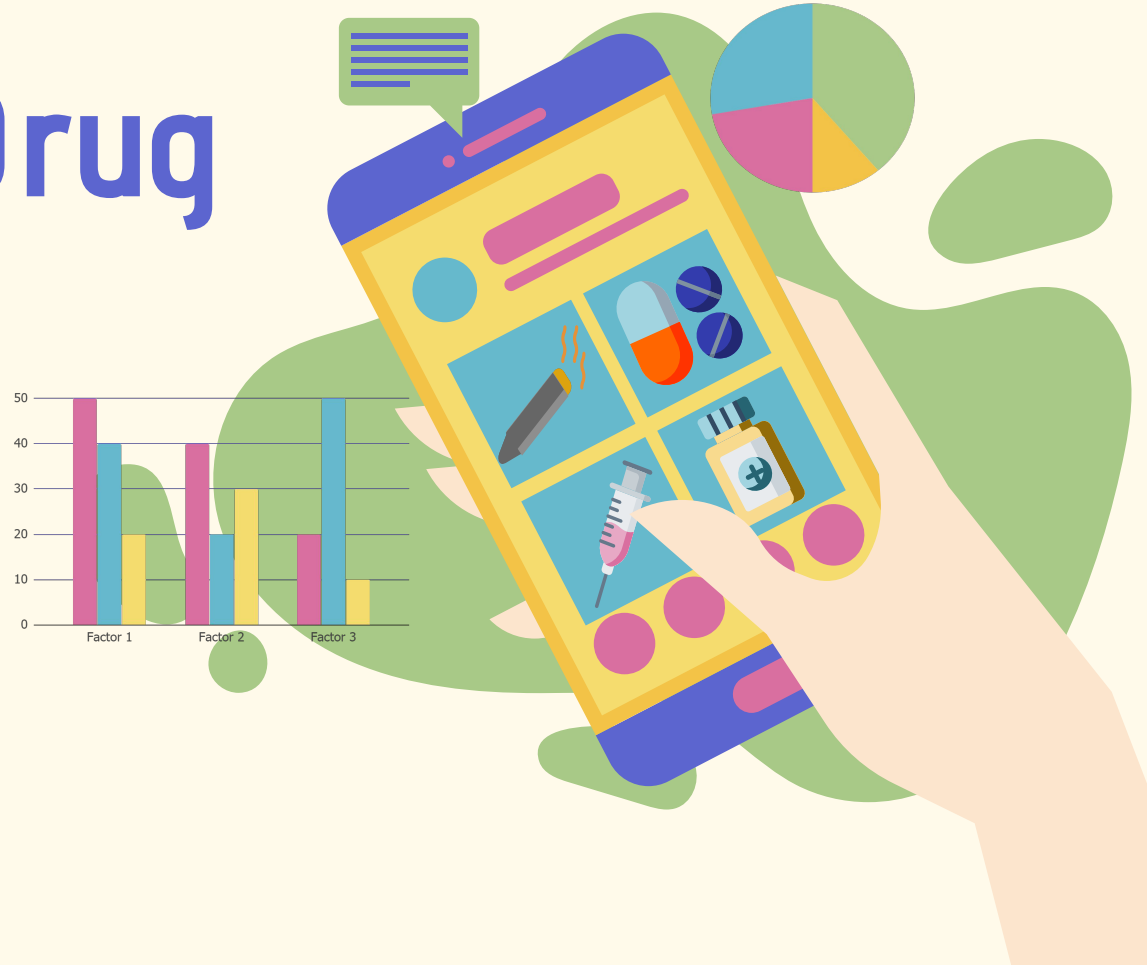
AI-driven Drug Prevention App

Frauke Albrecht,
Bendix Haß



Agenda

Business Case

target audience, app behavior, monetization

Methodology

metric, model selection, model tuning

Data

targets, features

Future Work

possibilities for further improvement

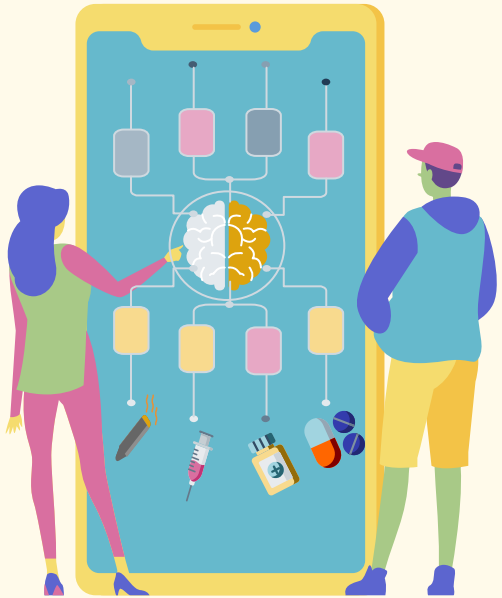
Business Case

Target Audience

- everybody who is willing to self-assess his/her risk for **drug use**

Impact

- help prevent drug use



App Behavior

1. enter demographic data
2. make personality tests
3. get personal results

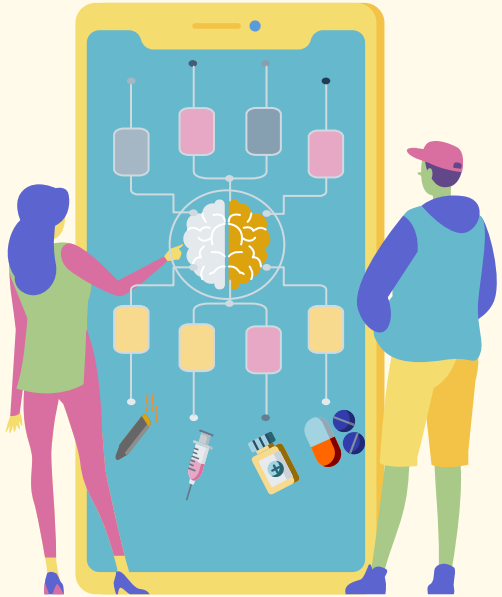
Monetization

- free basic results
- advanced premium results
- user data

Definition of Targets

User Scale

- grouped drugs by feature correlation
- resulting in 2 groups:
  1. **Ecstasy**, Amphetamines, Cannabis, Cocaine, Ketamine, Legal highs, LSD, Magic mushrooms
  2. **Benzodiacepines**, Amphetamines, Cocaine, Heroin, Methadone



Time Scale

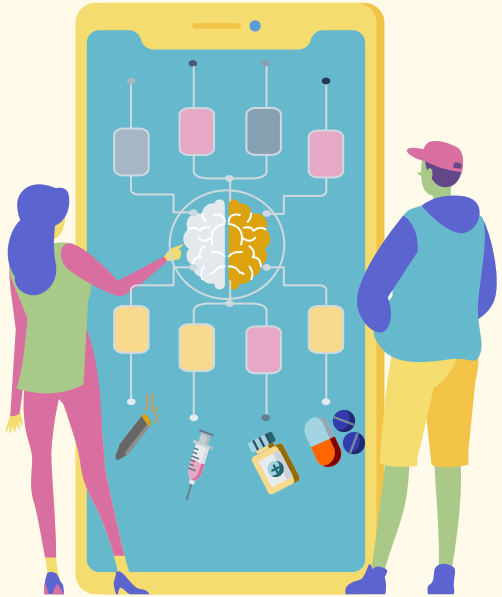
1. last used within last month?
2. last used within last year?

4 target permutations □ 4 independent models

Definition of Targets

User Scale

- grouped drugs by feature relation
- resulting in 2 groups:
  1. **Ecstasy**, Amphetamines, Cannabis, Cocaine, Ketamine, Legal highs, LSD, Magic mushrooms
  2. **Benzodiacepines**, Amphetamines, Cocaine, Heroin, Methadone



Time Scale

1. last used within last month?
2. last used within last year?

Tip! Use individual models

4 target permutations □ 4 independent models

Definition of Features

dropped features crossed out

Personality Tests

- Neuroticism
- Extraversion
- Openness to Experience
- Agreeableness
- Conscientiousness
- Impulsivity
- ~~Sensation Seeking~~

Demographic Data

- Age
- Gender
- Education
- Country
- ~~Ethnicity~~

Legal Drug Data

- ~~Caffeine~~
- ~~Chocolate~~
- Nicotine
- ~~Alcohol~~

Definition of Features

Dropped features crossed out

Tip! Drop those from the user survey

Personality Tests

- Neuroticism
- Extraversion
- Openness to Experience
- Agreeableness
- Conscientiousness
- Impulsivity
- ~~Sensation Seeking~~

Demographic Data

- Age
- Gender
- Education
- Country
- ~~Ethnicity~~

Legal Drug Data

- ~~Caffeine~~
- ~~Chocolate~~
- Nicotine
- ~~Alcohol~~

Methodology

Metric

Precision:

- reduce False Positives (non-user is predicted user)

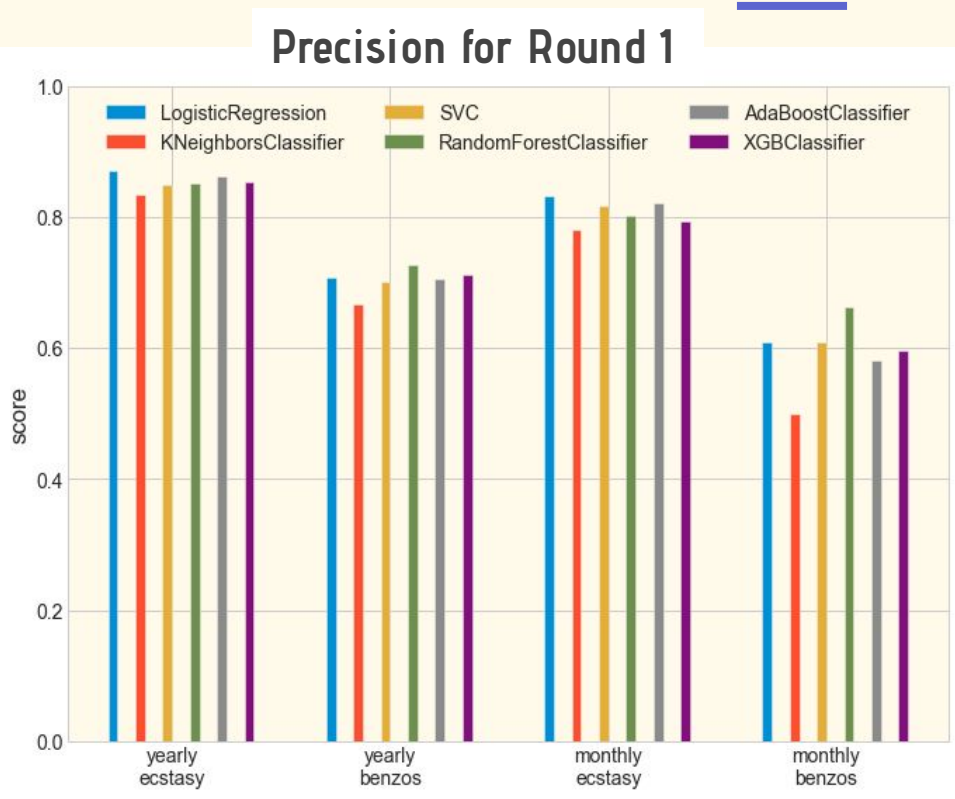
Recall:

- reduce False Negatives (user is predicted non-user)

Model Types

1. Logistic Classifier
2. K-nearest-neighbors
3. Support-vector-machine
4. Random Forest
5. AdaBoost
6. XGBoost

Round 1 & 2: Model Selection



Rules for round 1:

- 6 models per target
- 4 targets
- default parameters
- 5-fold cross-validation
- metric: precision

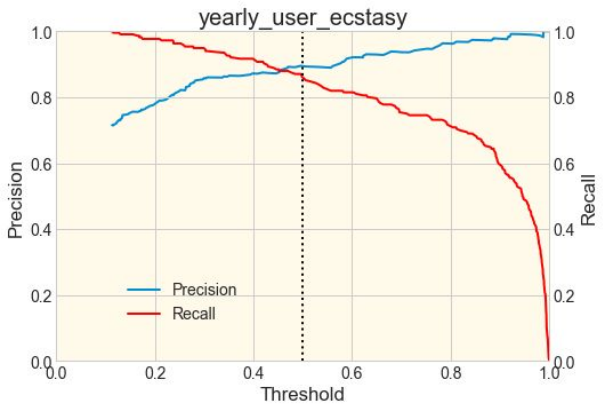
Rules for round 2:

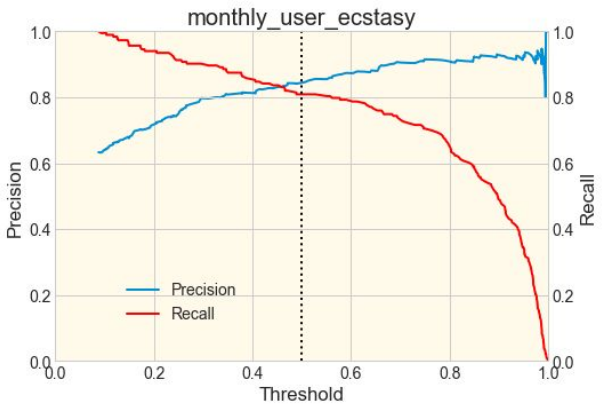
- 3 models per target
- 4 targets
- GridSearch
- 5-fold cross-validation
- metric: precision

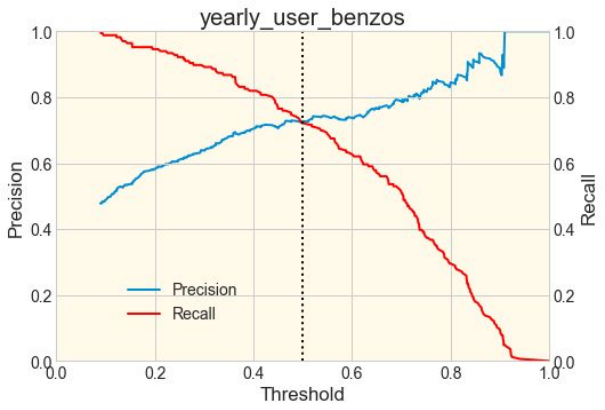
Winner:

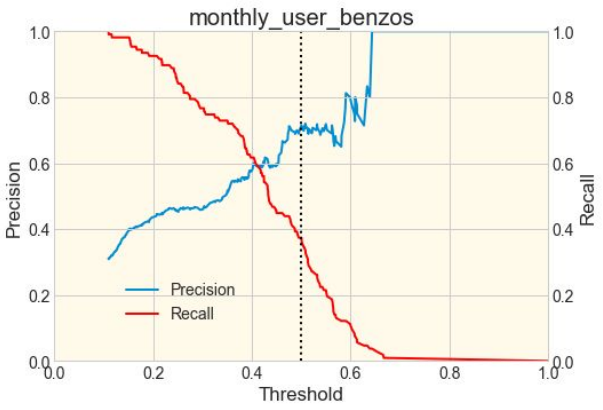
- LogisticRegression
- Precision: 72..89%

Model Tuning: Selecting the threshold



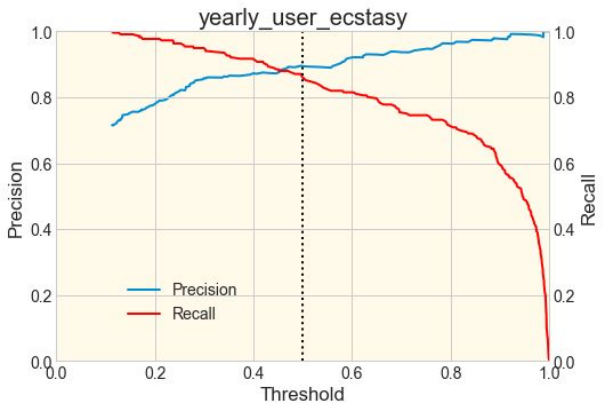


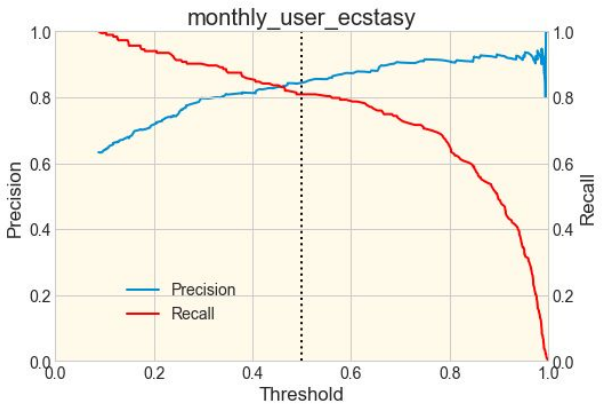


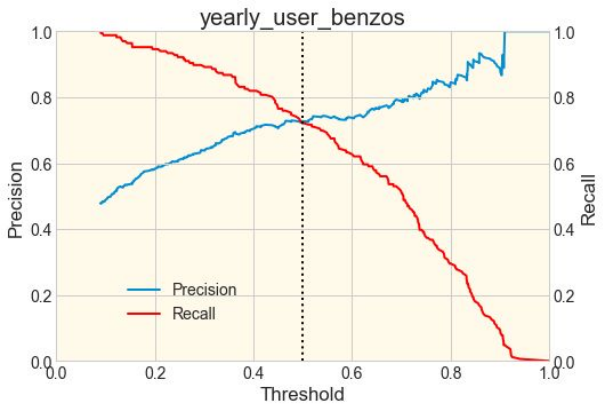


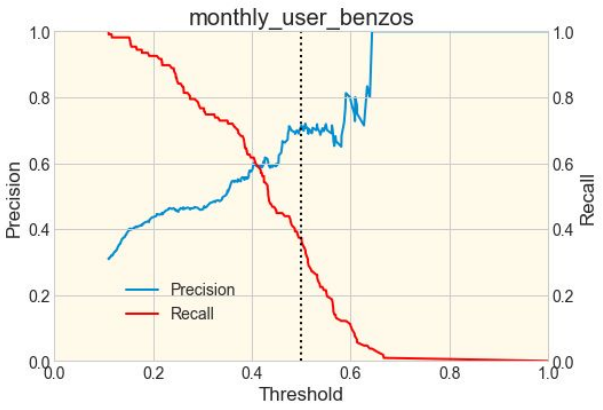
- default threshold of .5 delivers reasonable outcome
- threshold can be adapted at will

Model Tuning: Selecting the threshold









- default threshold of .5 delivers reasonable outcome
- threshold can be adapted at will

Tip! adapt the threshold to differ user experience

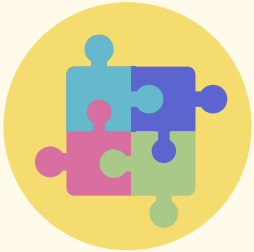
Future Work

More Data

model is biased for ethnicity “white”

Causality of Personality

check assumption: personality is independent of drug use

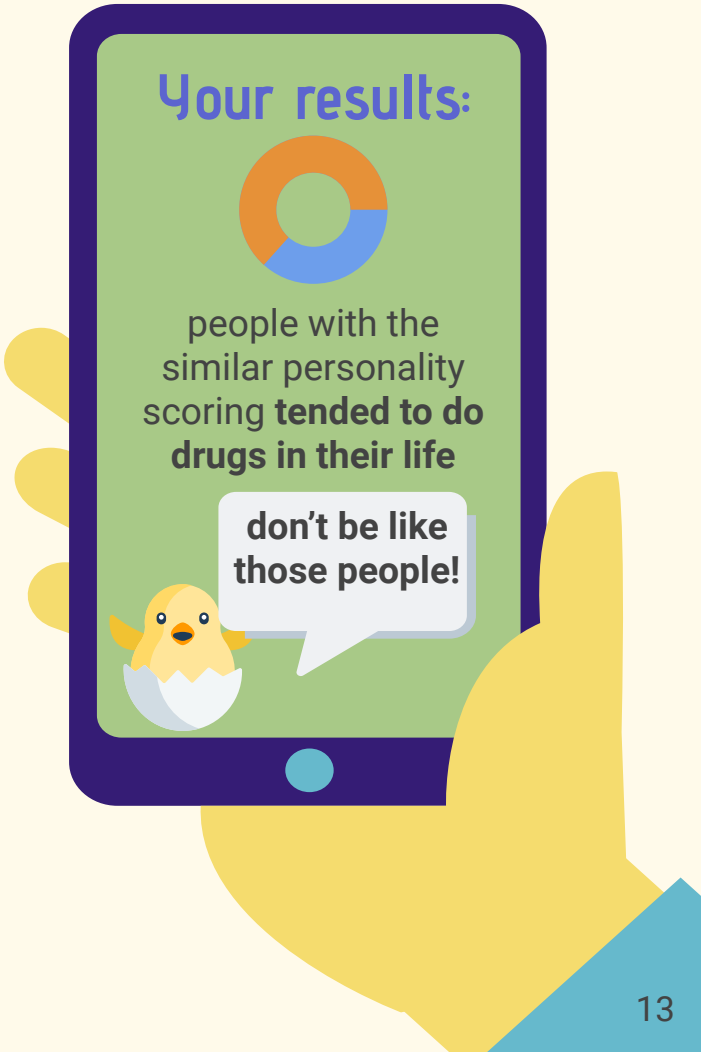


Model Tweaking

metric is focused on **not unsettling people:** double check that by A/B testing

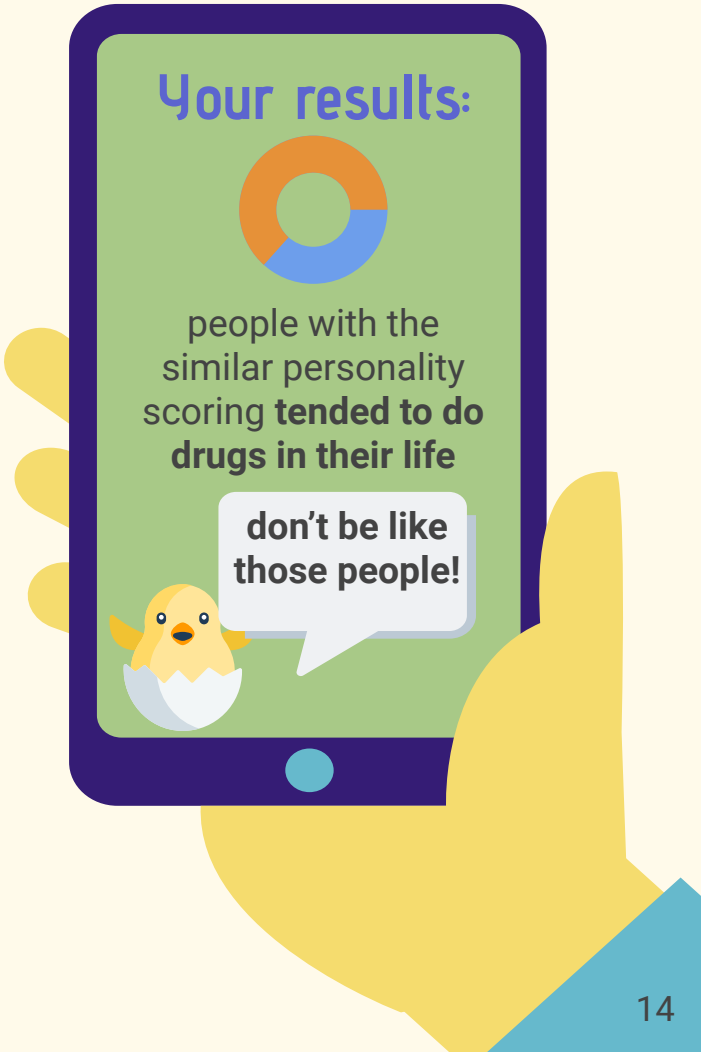
Basic function

based on the his demographic inputs and the results of the personality tests the user gets the feedback whether he is biased to do drugs in the future



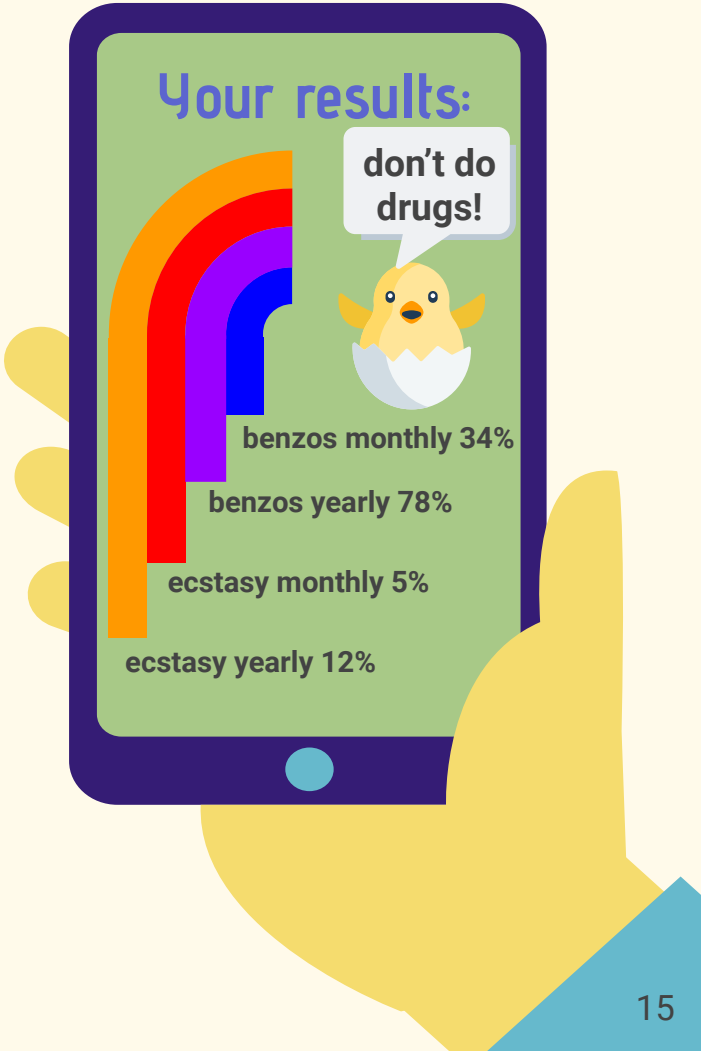
Premium function

based on the his demographic inputs and the results of the personality tests the user gets the feedback which drugs and at which time scale the user is biased to do drugs.



Premium function

based on the his demographic inputs and the results of the personality tests the user gets the feedback which drugs and at which time scale the user is biased to do drugs.



Thanks!

Do you have any questions?

Fun Fact: The highest correlation exists between: **LSD** and **magic mushrooms**