



# Introduction to Cobra

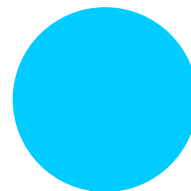
Predictive models focused on performance and interpretation

34th Data Science Leuven Meetup

Jan Beníšek, 2021/02/09



# About



**PYTHON**  
PREDICTIONS  
A TOBANIA COMPANY



**Since 2006** | **Our mission**

*Data Science with Impact*

*Being the most valuable data  
science business partner*



**Jan Benisek**

*Senior Data Scientist / Data Engineer*

[jan.benisek@pythonpredictions.com](mailto:jan.benisek@pythonpredictions.com)

[www.linkedin.com/in/janbenisek](https://www.linkedin.com/in/janbenisek)

# ARE YOU READY TO TAME THE COBRA?

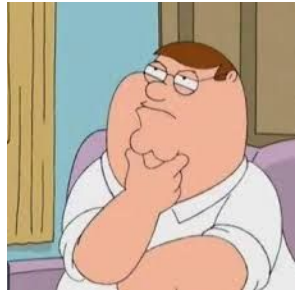


# WHAT?

Cobra allows you to **quickly** build **interpretable** models with high **predictive power**.

# Why Cobra?

*Cobra makes interpretable models equally accessible as black box models.*



## Coverage

Most projects are  
some form of binary  
classification



## Automation

Quick model building  
leaves more time for  
exploration

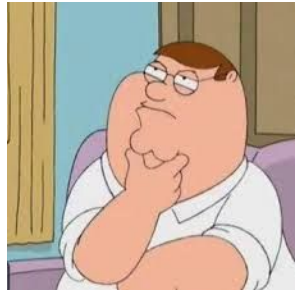


## Availability

Rapid and robust  
model building  
becomes available  
for many clients

# Why open-sourcing Cobra?

*We want to give back to the community and be transparent.*



## Community

Sharing is caring

Giving back is  
important to us

Growing together



## Transparency

Towards clients

Job candidates  
understand what we  
do

# How does Cobra work?

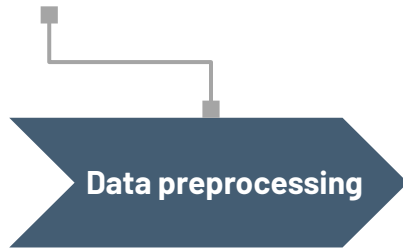
*Cobra works by combining our business expertise in model building with the latest Python Data Science libraries.*



# How does Cobra work?

*Preprocessing is the core of Cobra and its interpretability...*

- Train/selection/validation split
- Binning continuous variables into intervals
- Replace missing values
- Regroup categorical variables
- Target encoder of bins/categories

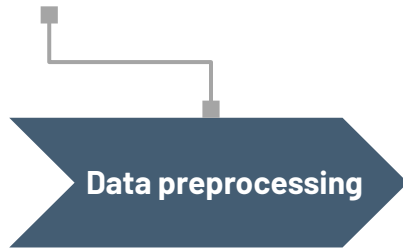




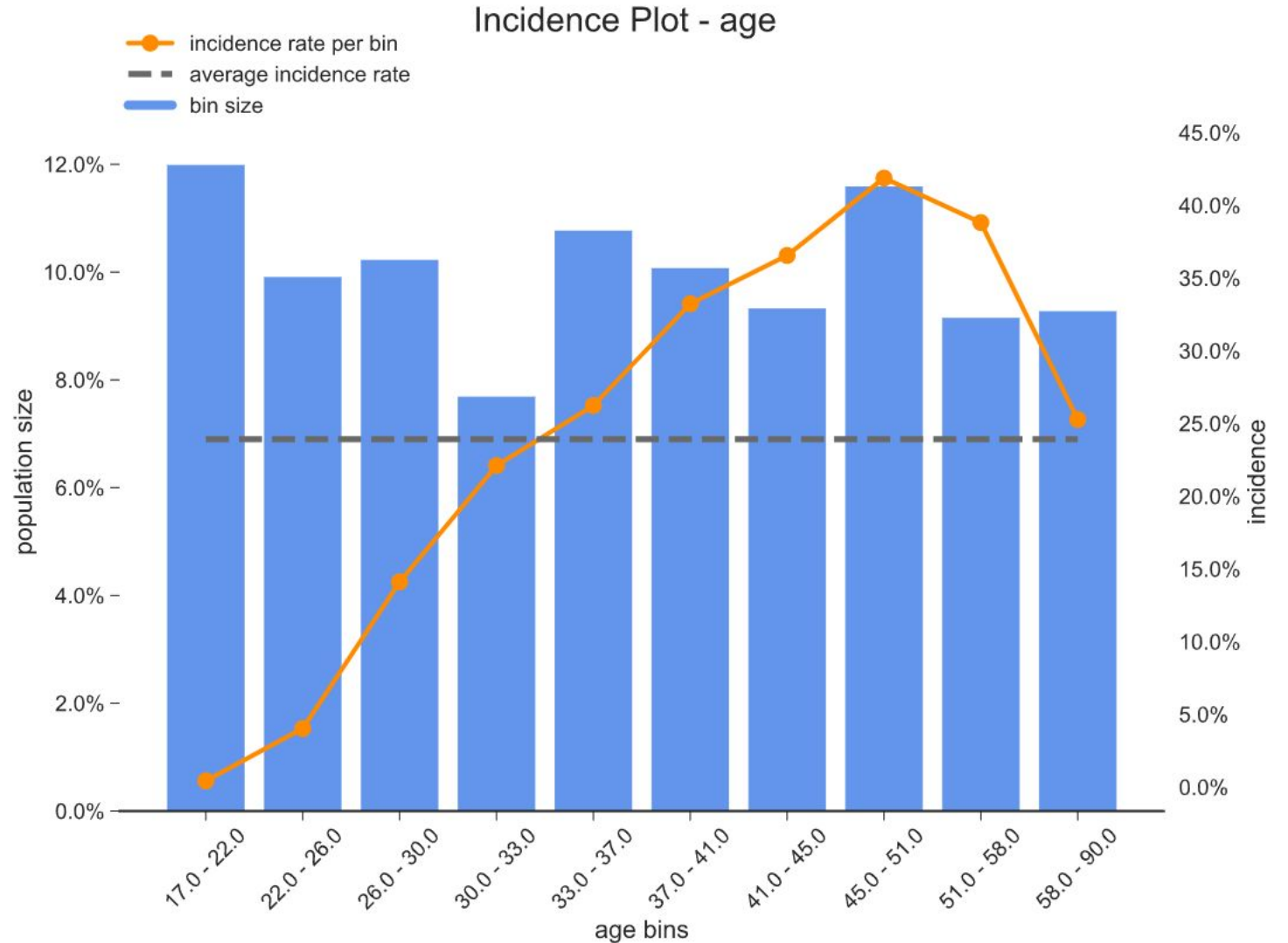
# How does Cobra work?

*...and incidence plots the main tool to open the box.*

- Train/selection/validation split
- Binning continuous variables into intervals
- Replace missing values
- Regroup categorical variables
- Target encoder of bins/categories



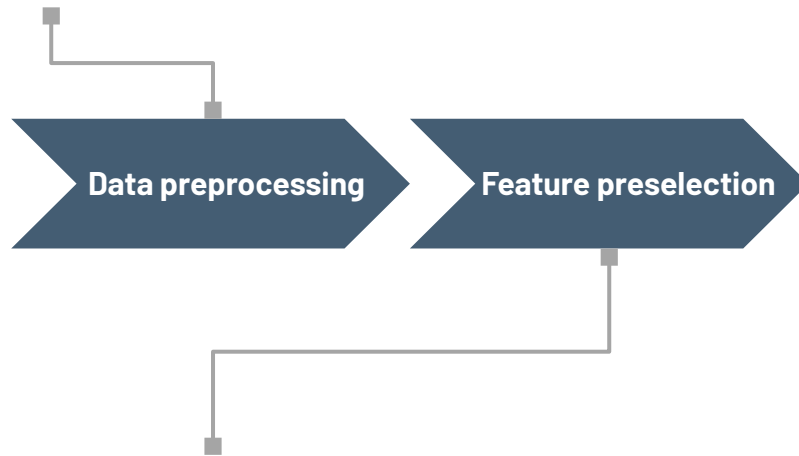
[Geert's article](#)



# How does Cobra work?

*We select only strong features and drop those which overfit.*

- Train/selection/validation split
- Binning continuous variables into intervals
- Replace missing values
- Regroup categorical variables
- Target encoder of bins/categories



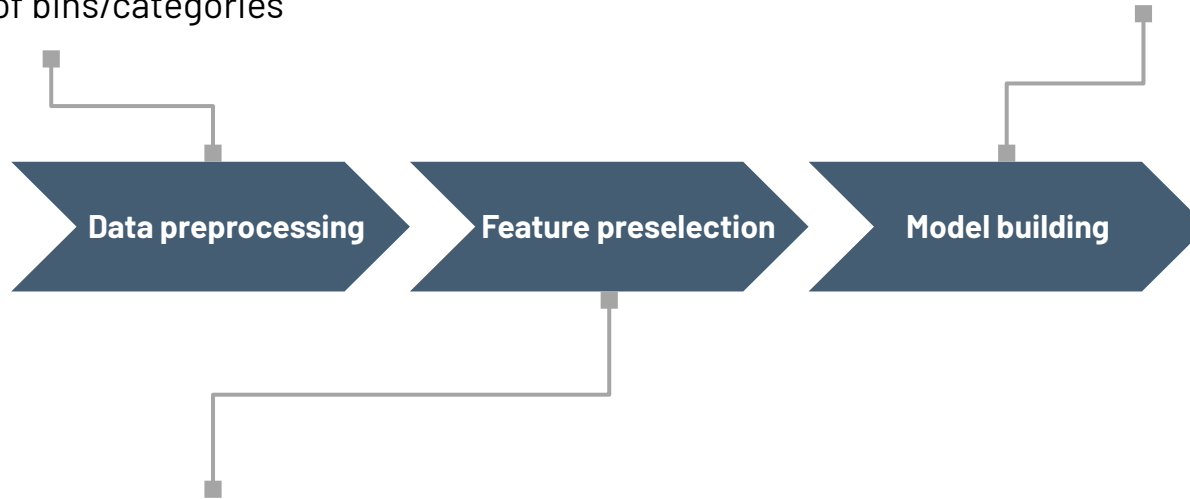
- Univariate preselection based on AUC
- Correlations & feature importance

# How does Cobra work?

*Logistic regressions with forward feature selection.*

- Train/selection/validation split
- Binning continuous variables into intervals
- Replace missing values
- Regroup categorical variables
- Target encoder of bins/categories

- Forward feature selection
- Multiple logit models are built
- Comparison of models
- Feature importance plots



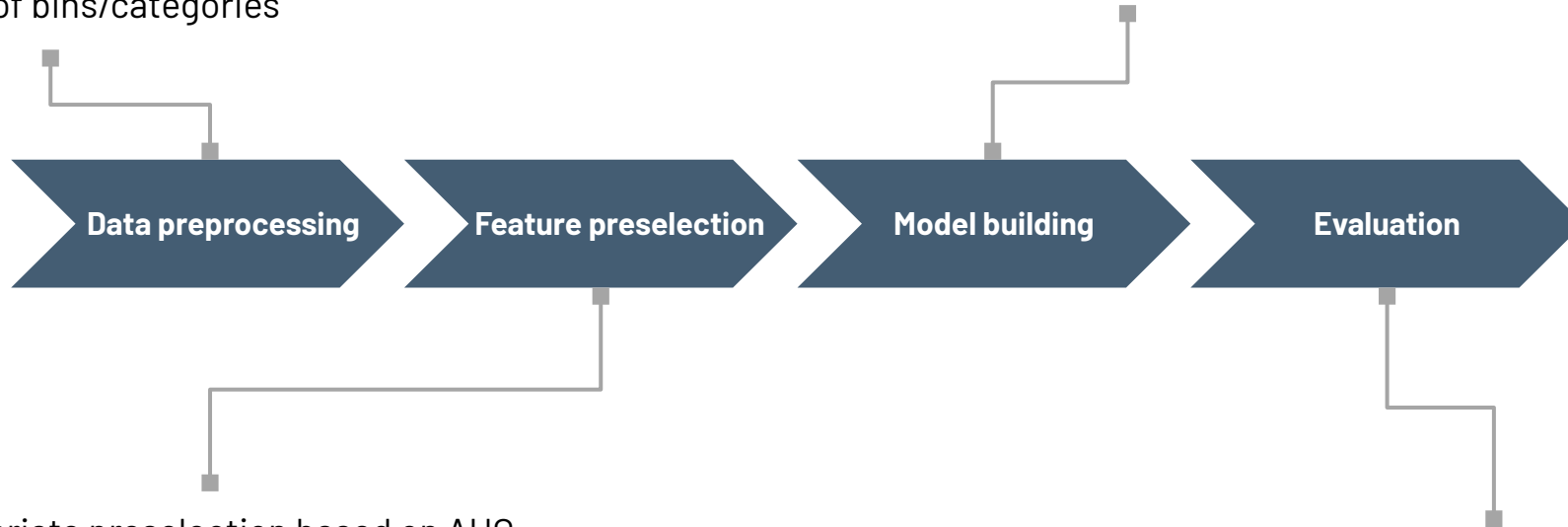
- Univariate preselection based on AUC
- Correlations & feature importance

# How does Cobra work?

*Scalar metrics and plots let us see how the model performs.*

- Train/selection/validation split
- Binning continuous variables into intervals
- Replace missing values
- Regroup categorical variables
- Target encoder of bins/categories

- Forward feature selection
- Multiple logit models are built
- Comparison of models
- Feature importance plots



- Univariate preselection based on AUC
- Correlations & feature importance

- Scalar metrics (accuracy, AUC, ...)
- Evaluation plots - ROC, Confusion matrix, lift, ...
- Predictor Incidence Graphs (PIGs) for understanding



Let's get our hands dirty



**Thank you!**



**Questions?**

