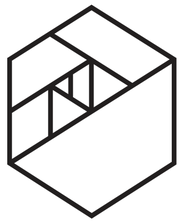
# Heart Disease Indicators app

## Data Engineering

project for Metis EDA Bootcamp

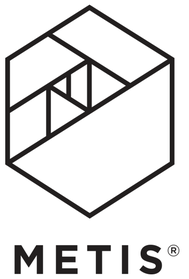
by Krystian Krystkowiak, 2022

# Introduction




- **Early heart disease identification** may be desired by medical institutions
- Also: insurances, medical apps, fitness or nutritionist, **individuals conscious about health**
- GOAL: **Web app** that can help **increase awareness** and raise red flag during **initial questionnaire**.

# Methodology - Data




- Behavioural Risk Factor Surveillance System, conducts **annual telephone surveys** to gather data on the public **health status**
- Recent (February 15, 2022) includes data from 2020.
- Answers of **319k U.S. residents**
- Features: **HeartDisease (target)**, BMI, Smoking, AlcoholDrinking, Stroke, PhysicalHealth, MentalHealth, DiffWalking, Sex, AgeCategory, Race, Diabetic, PhysicalActivity, GenHealth, SleepTime, Asthma, KidneyDisease, SkinCancer

# Methodology - Pipeline

METIS®

## DATA INGESTION

- Behavioural Risk Factor Surveillance System questionnaire data
- download from Kaggle
- 300k+ rows of data

kaggle

## TESTING/ROBUSTNESS

- Manual, annual data update.

## DATA STORAGE

- Store data as a relational database with SQLAlchemy and SQL commands

SQL

## PROCESSING

- Clean data using Python, NumPy and pandas
- Aggregate data with pandas
- Build predicting model

## DEPLOYMENT

- Visualize the data and create a web application
- Deploy application to the web with Github and Streamlit

GitHub

Streamlit

# Results

Share

Check if you should visit the doctor!

Sex

Female

Age Category

18-24

Race

White

Height? (kg)

160.00

50.00 230.00

Weight? (cm)

65.00

35.00 200.00

General Health

Excellent

Phisical Activity

Yes

## Heart Disease Indicators

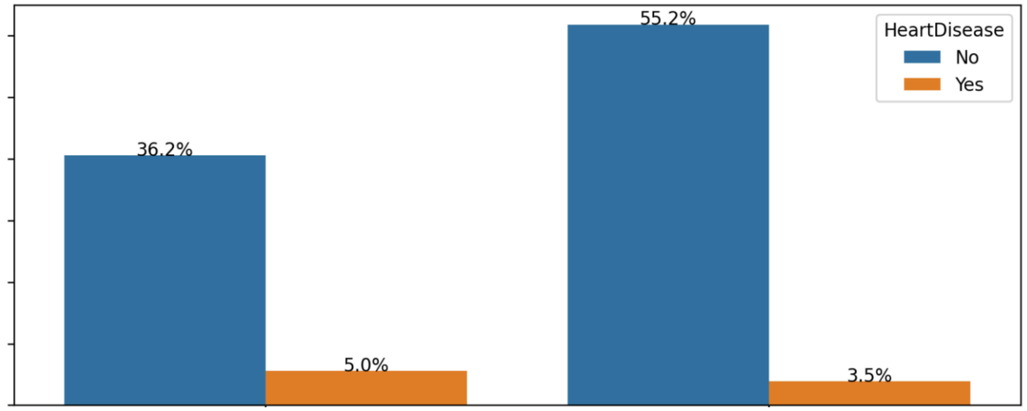
From Behavioural Risk Factor Surveillance System dataset.

Show Raw Data & Target Distribution

### Heart Disease vs Diferent Features

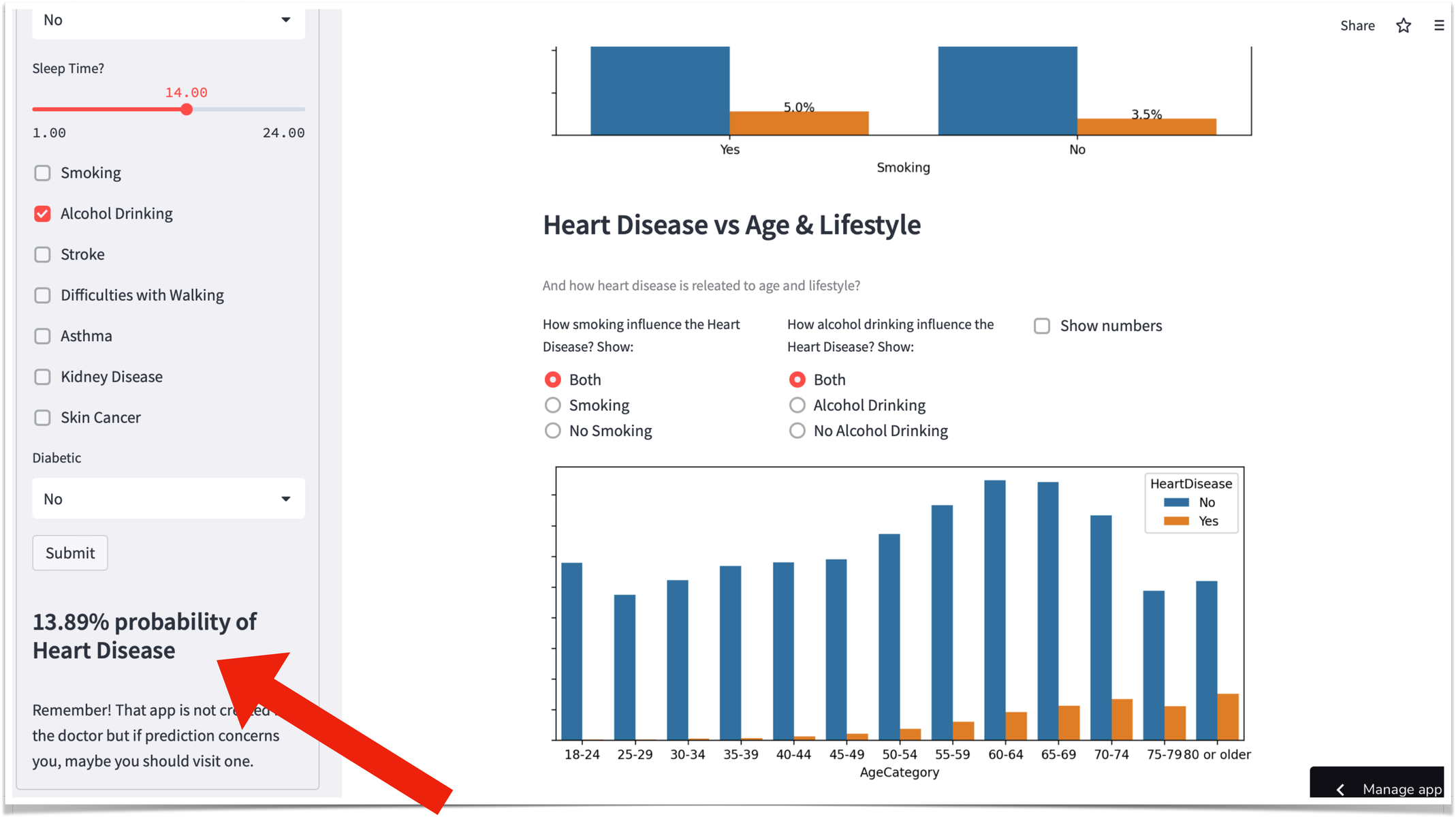
How heart disease is releated to deferent features from dataset?

Smoking

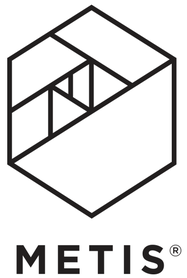



Manage app

# Results


No
Sleep Time?
14.00
1.00
24.00
Smoking
Alcohol Drinking
Stroke
Difficulties with Walking
Asthma
Kidney Disease
Skin Cancer
Diabetic
No
Submit
13.89% probability of Heart Disease
Remember! That app is not cr ed
the doctor but if prediction concerns
you, maybe you should visit one.
Share
5.0%
3.5%
Yes
No
Smoking
Heart Disease vs Age & Lifestyle
And how heart disease is releated to age and lifestyle?
How smoking influence the Heart Disease? Show:
Both
Smoking
No Smoking
How alcohol drinking influence the Heart Disease? Show:
Both
Alcohol Drinking
No Alcohol Drinking
Show numbers
HeartDisease
No
Yes
18-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80 or older
AgeCategory
Manage app

# Future Work&Conclusions




- Data ingestion as **annual automatic scrape** from BRFSS CRON
- **Cloud SQL** aws
- **Model** improvement

# Future Work&Conclusions

METIS®

- Data **consistency**!
- Constructed pipeline is good **base** for similar apps
- **Healthy** life style
  regular checkups

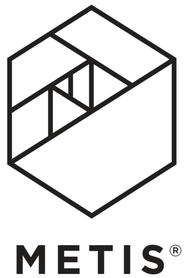
# Thank you!

## Questions?

project for Metis EDA Bootcamp

by Krystian Krystkowiak, 2022

# Appendix




- Original dataset: https://www.cdc.gov/brfss/
- Dataset: www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease (initially cleaned by Kamil Pytlak at Kaggle)
- App: https://krystkowiakk-metis-project-7--streamlit-appstreamlit-app-xuz5iy.streamlit.app/