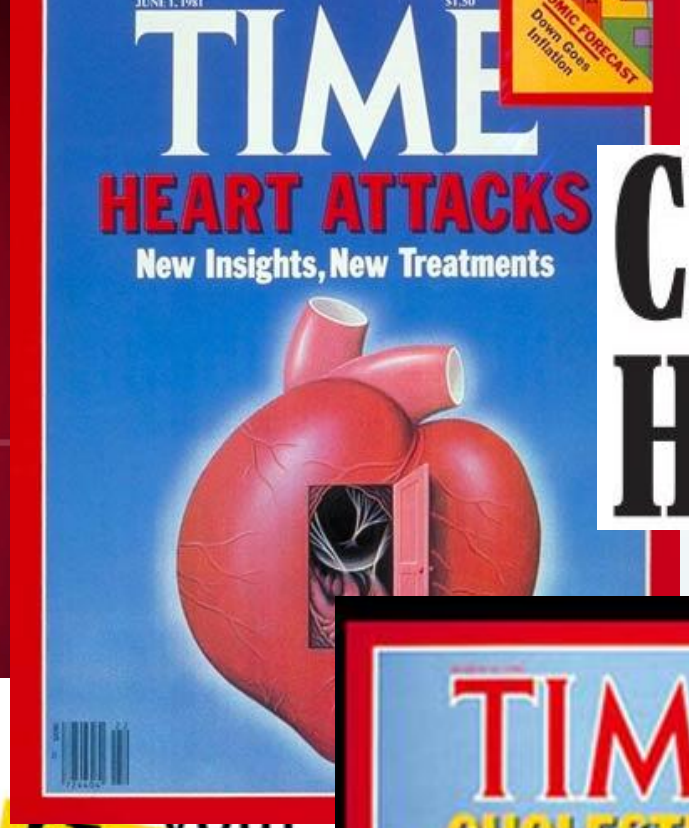
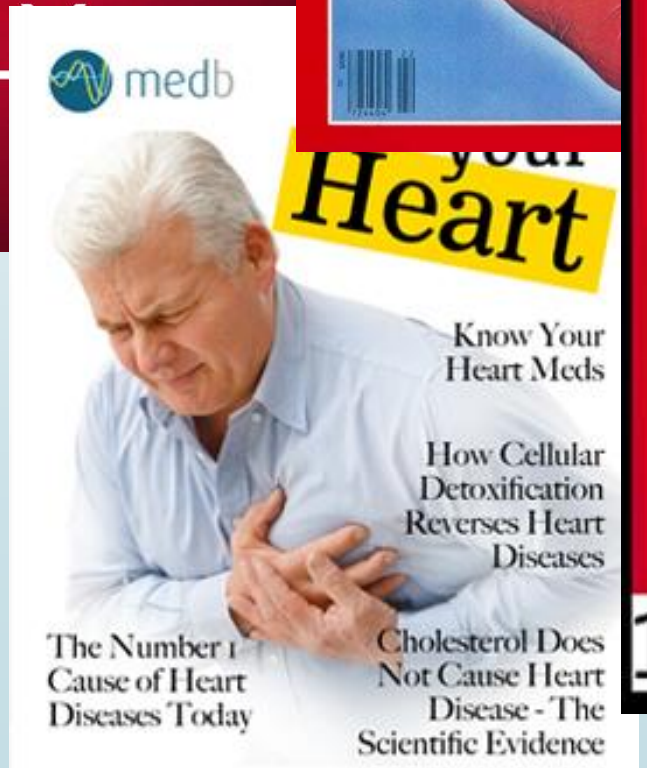

World Health Organization
REGIONAL OFFICE FOR Europe
Cardiovascular diseases are the
No. 1
cause of death
GLOBALL
#NoTobacco


TIME
HEART ATTACKS
New Insights, New Treatments


CHEESE FIGHTS
HEART DISEASE


medb
Heart
Know Your Heart Meds
How Cellular Detoxification Reverses Heart Diseases
The Number 1 Cause of Heart Diseases Today
Cholesterol Does Not Cause Heart Disease - The Scientific Evidence


TIME
CHOLESTEROL
And Now the Bad News...
1984
TIME
Eat Butter.
2014

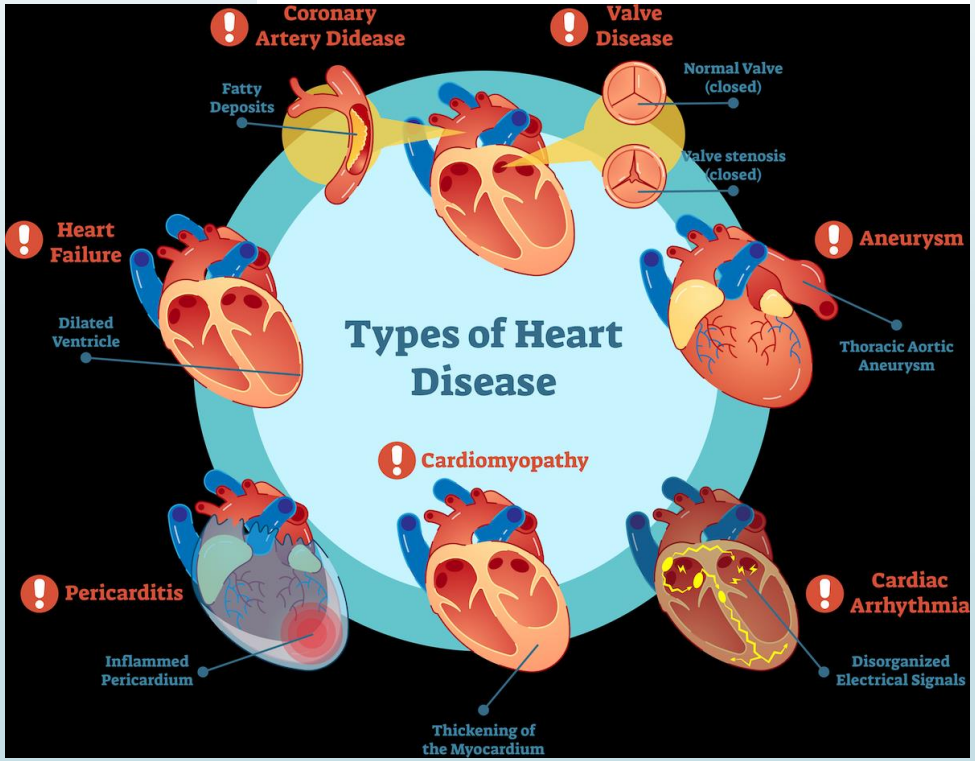
Risk factors for heart disease in different locations

by
Kalvin Dobler, Nicolas Huber, Nathalie Guttmann
Introduction to Data Science 2021

Cardiovascular disease (CVD)

- Nr. 1 worldwide for mortality and morbidity
- Mostly in low and middle-income countries
- High- socio-economic burden
- Most CVDs losses: ischemic heart disease and stroke

DALYs	by 2010	by 2020	by 2030
CVD DALYs Annual number of DALYs	153 million	169 million	187 million
Burden of CVD Percentage of all DALYs	10.4%	11.0%	11.6%
CVD rankings globally	3rd: coronary heart disease 5th: stroke	3rd: coronary heart disease 4th: stroke	3rd: coronary heart disease 4th: stroke
CVD rankings in developing countries	4th: coronary heart disease 8th: stroke	3rd: coronary heart disease 6th: stroke	3rd: coronary heart disease 5th: stroke



Research questions

- Risk factors predicting cardiovascular diseases?
- Meaningful different levels of heart diseases?
- Difference between locations?

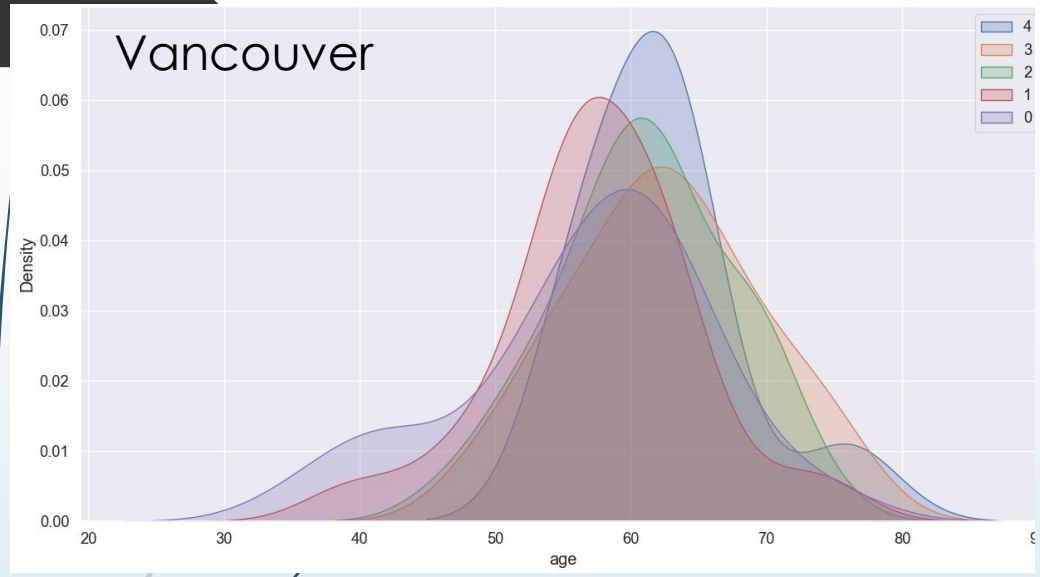
Dataset

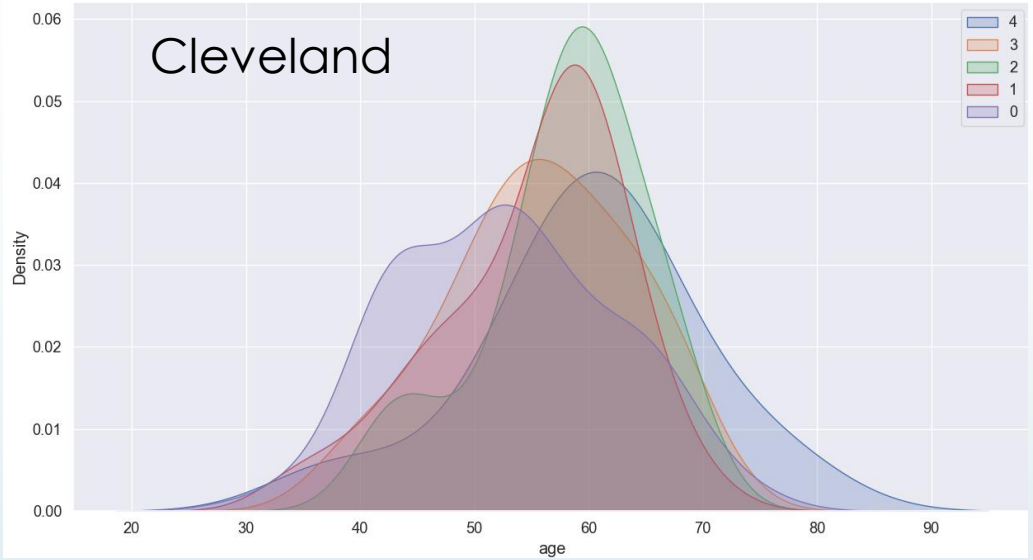
- Name: Heart Disease Data Set
- Source: UCI Machine Learning Repository
- Time: 1998
- Location: Hungary, Switzerland (incl. Zürich and Basel), Cleveland, Long Beach
- Disease: 0 : healthy, 1-4: different levels of heart disease
- Attributes (75): sex, age, chest pain location, smoking….

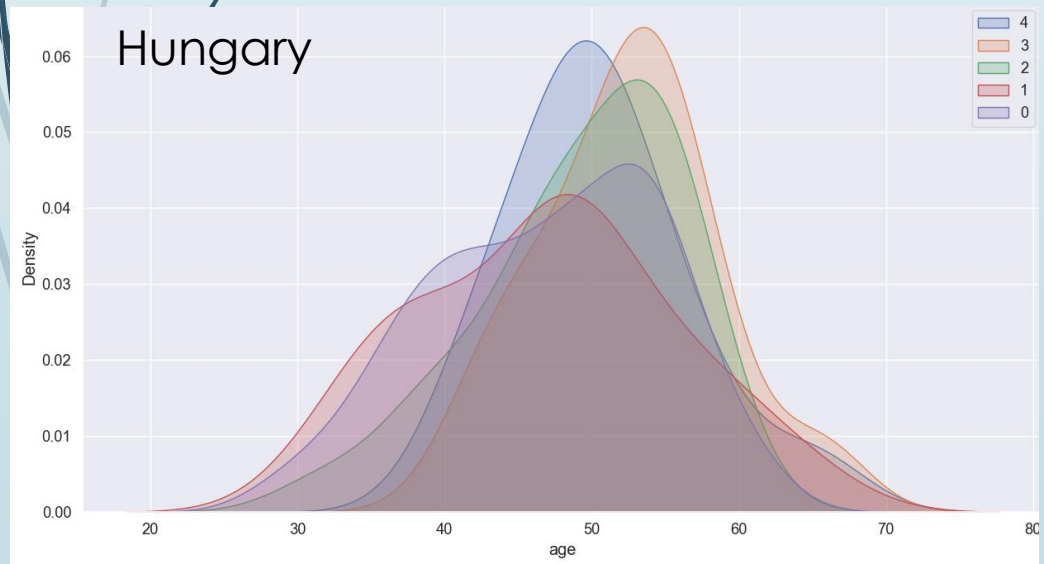
Workflow

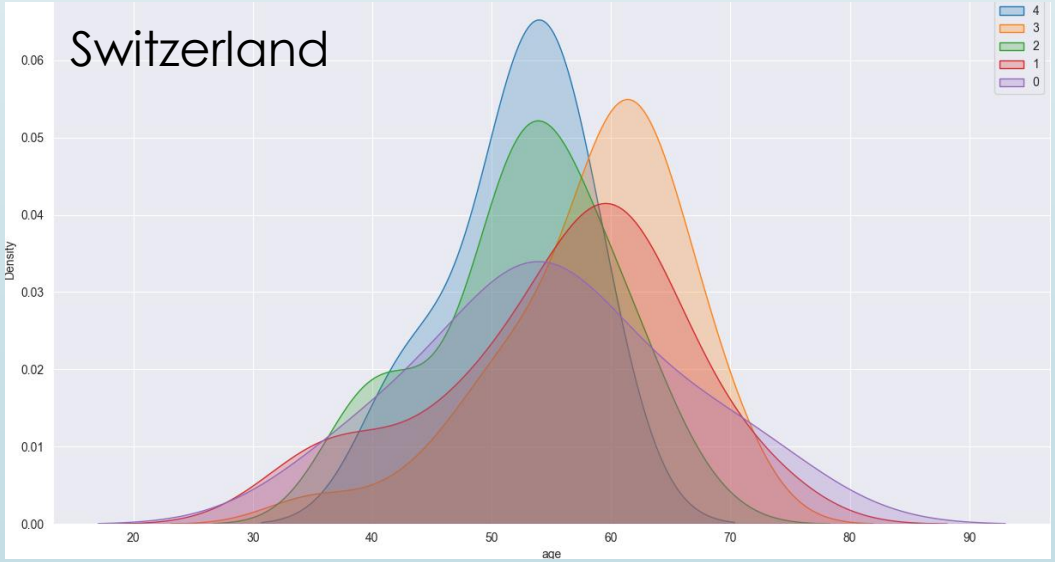
- Pre-processing: removing missing values
- Data visualization: correlation between features
- Feature selection: Random Forrests
- Dimensionality reduction: t-SNE, UMAP, Autoencoders
- Classification: logistic regression, multinomial logistic regression, naive Bayes, SVM, KNN, neural networks
- Performance measurement: Confusion matrix, ROC curve (incl. AUC)

Distribution of age for the type of disease

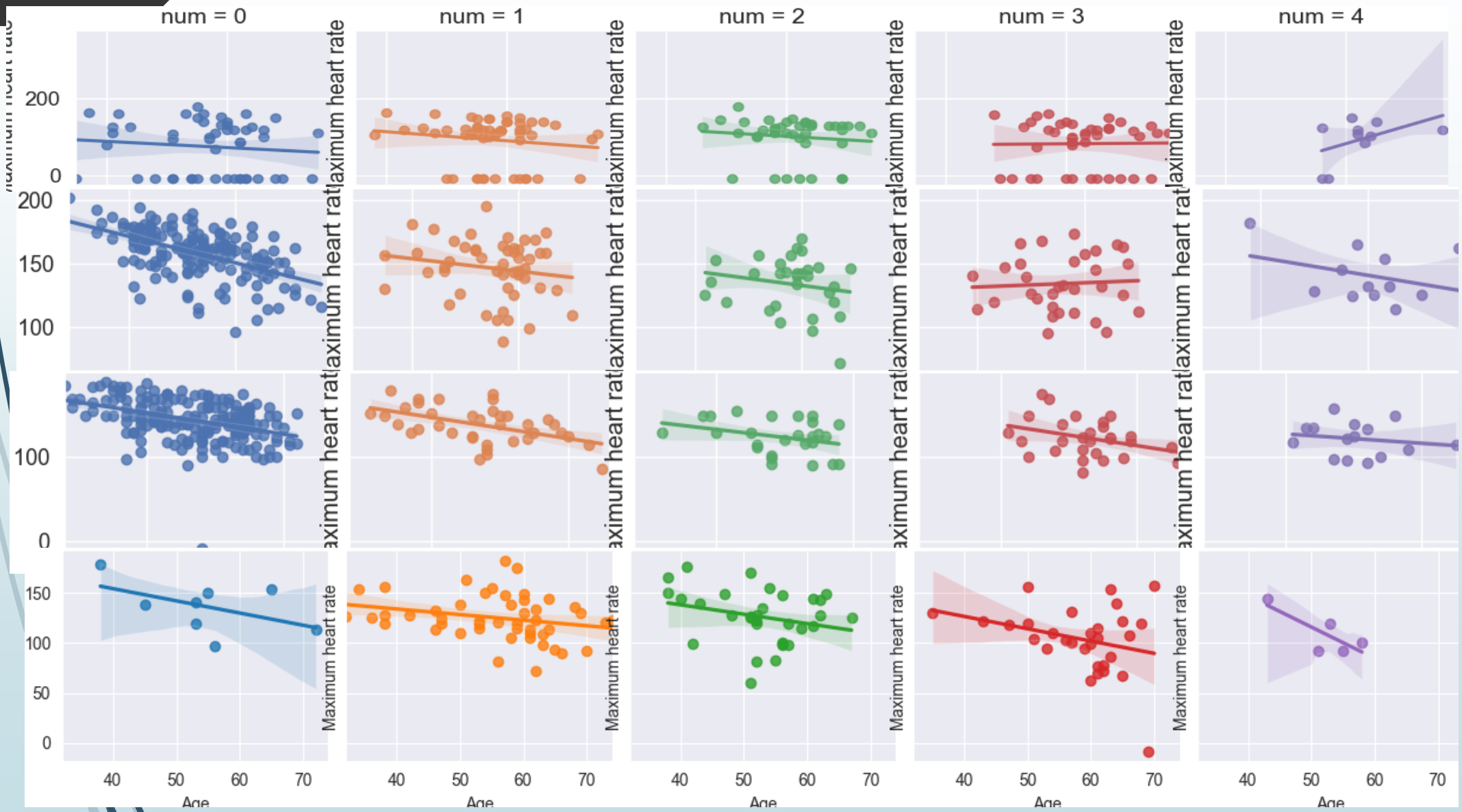








Effect of age on max. heart rate



Vancouver

Cleveland

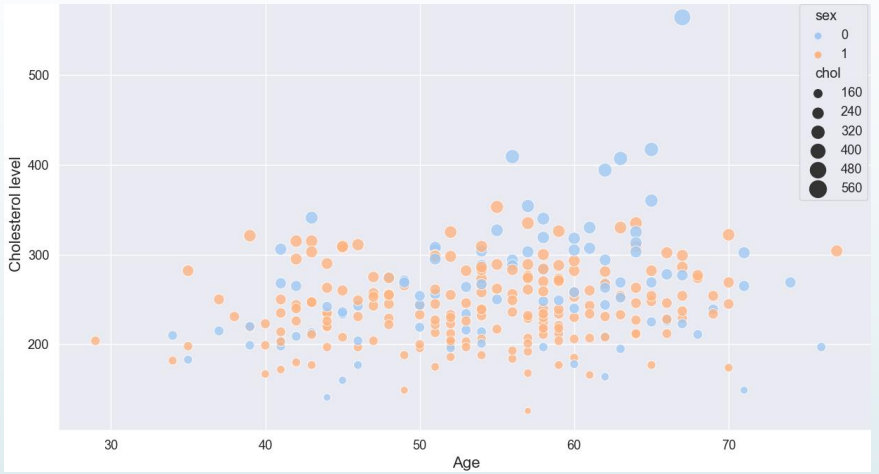
Hungary

Switzerland

Cholesterol levels

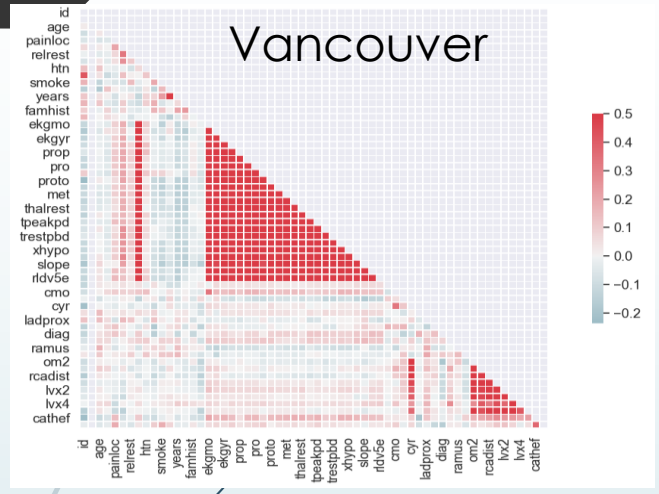
Vancouver

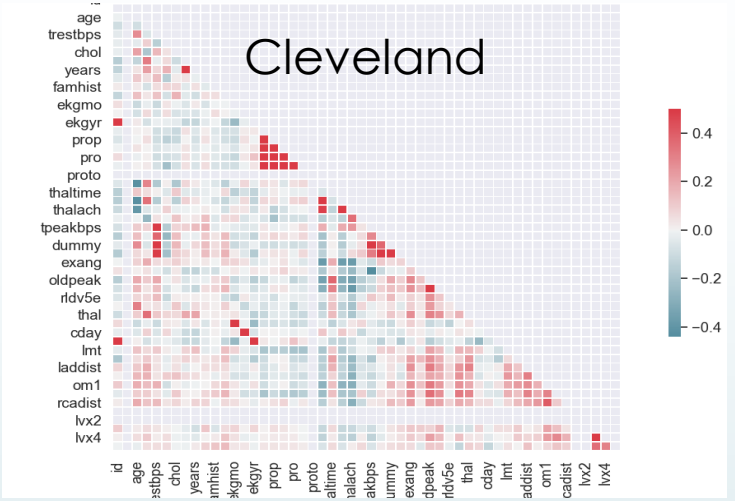
Cleveland

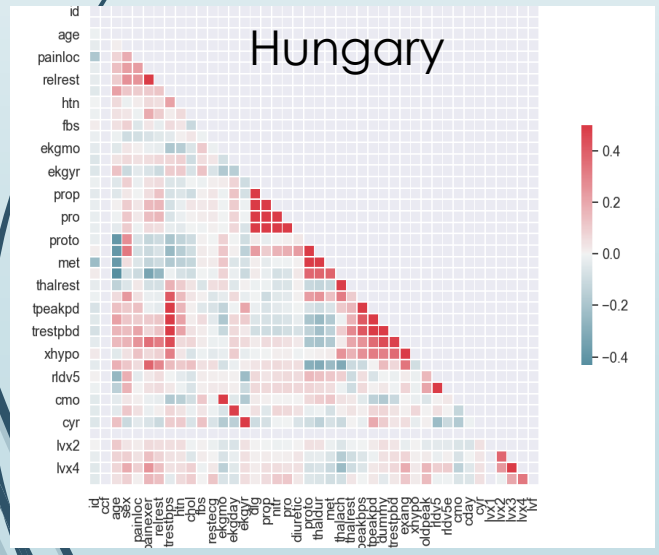


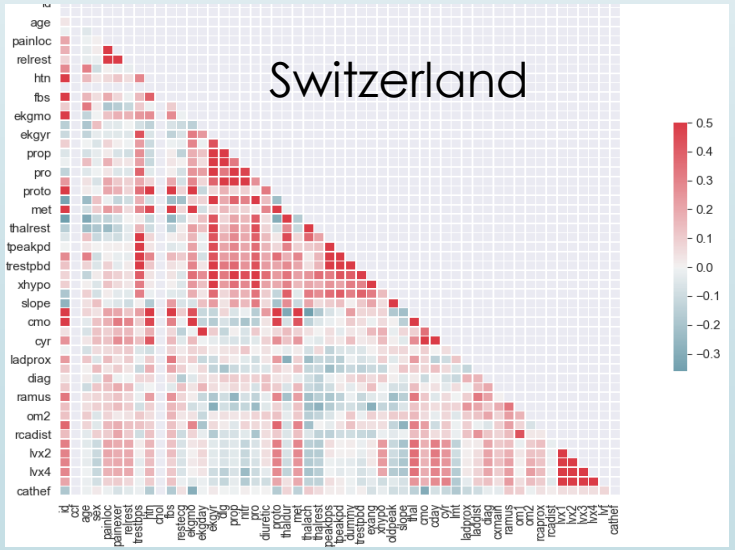
Hungary

Correlation between features



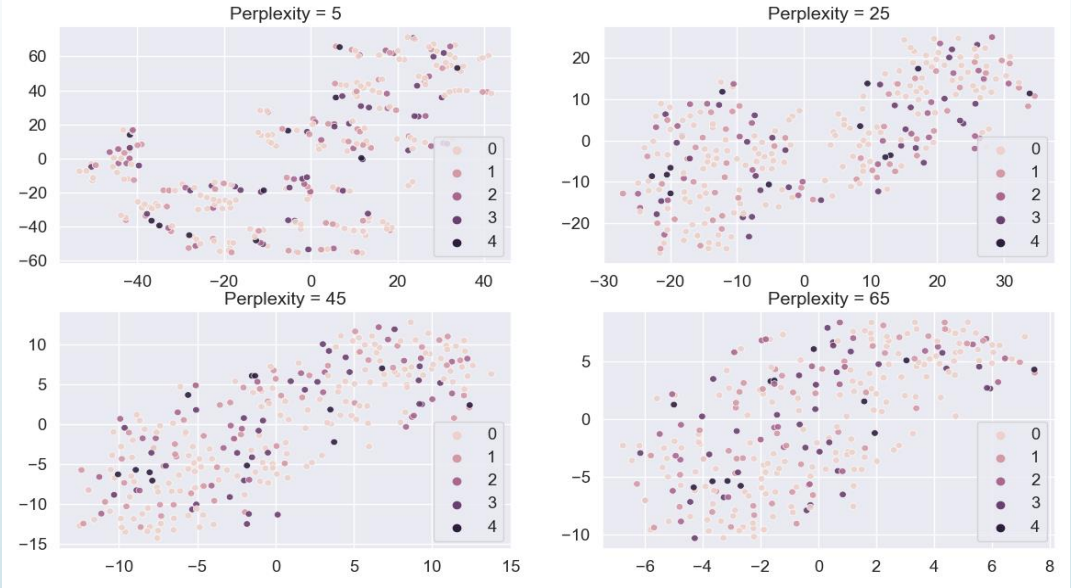






t-SNE

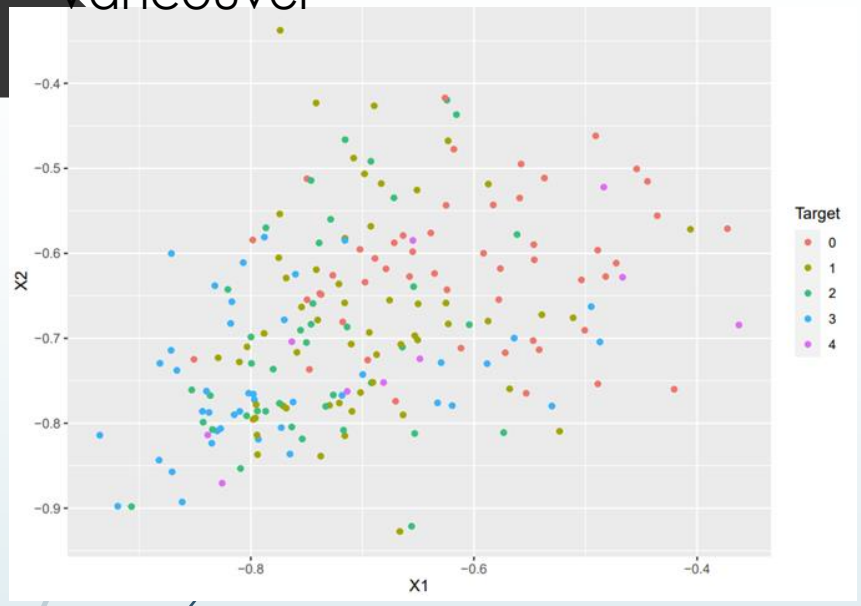
Cleveland



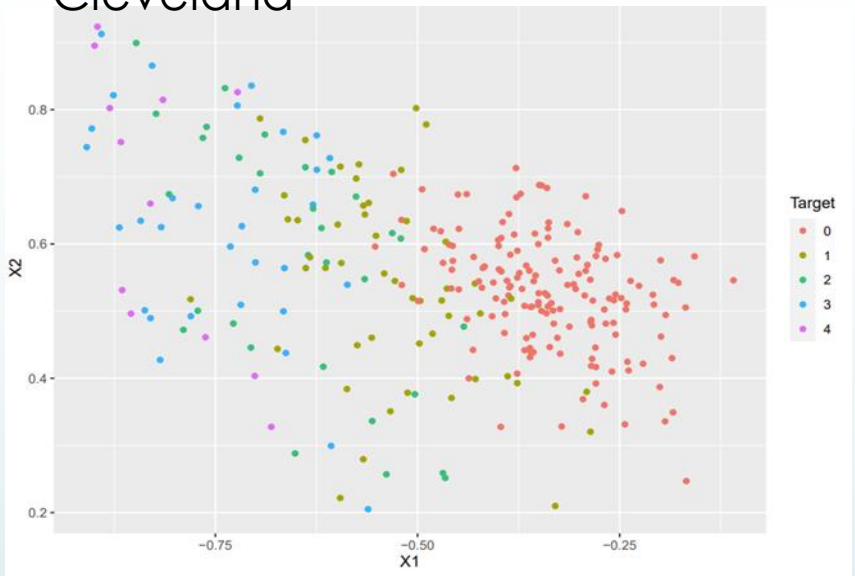
No pattern for all locations-> UMAP and autoencoders

Autoencoders

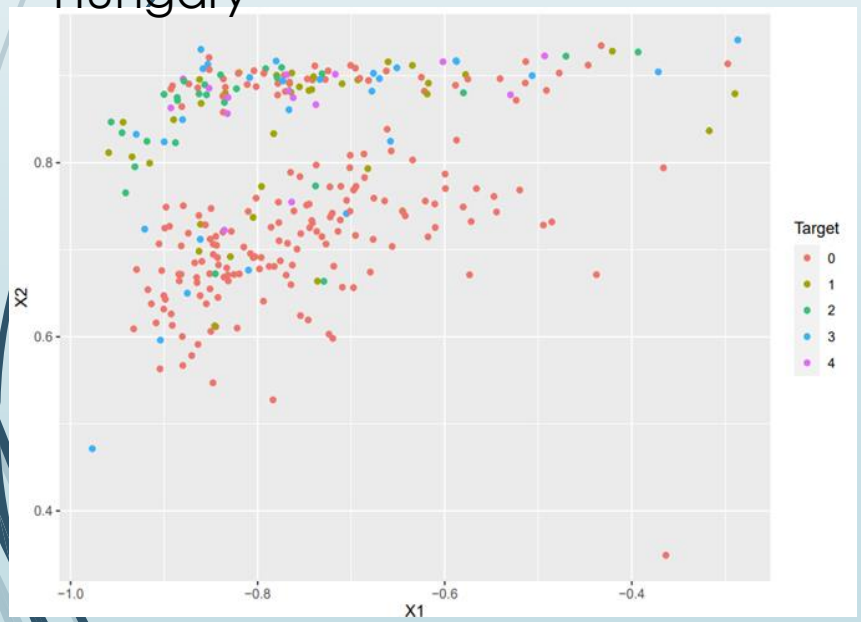
Vancouver



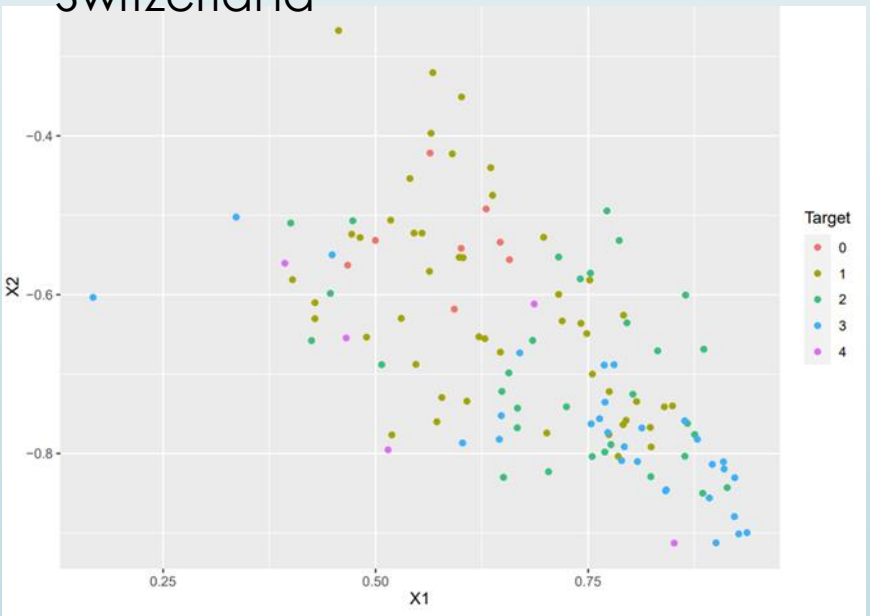
Cleveland



Hungary



Switzerland

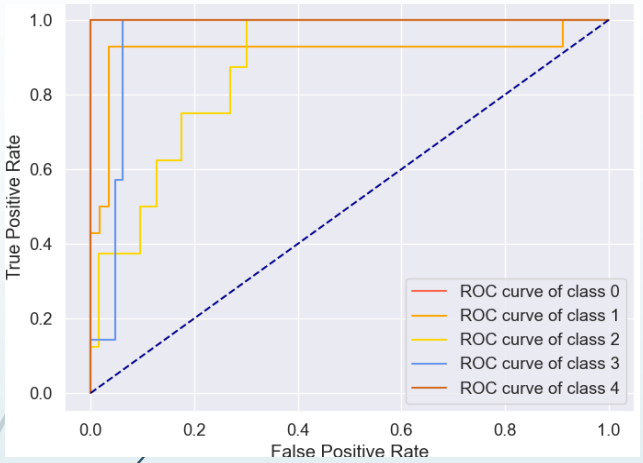


Top 5 most important risk factors

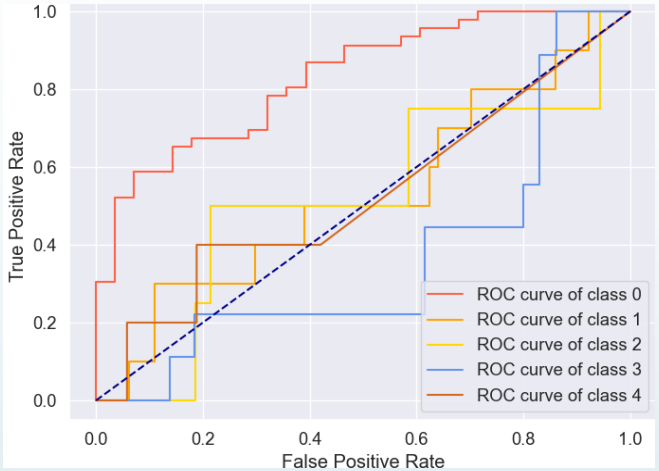
Cleveland	Vancouver	Switzerland	Hungary
Distal left anterior descending artery	Proximal right coronary artery	Circumflex	Chest pain
Thallium scintigraphy	Proximal left anterior descending arte	ID	Pain provoked by exertion
First obtuse marginal branch	ID	Maximum heart rate acheived	Induce ST depression relative to rest
Number of major vessel	Circumflex	Peak exercise systolic blood pressure	Lvx4
Proximal right coronary artery	Day of cardiac catheterization	Age of patients	Exercise-induced angina

ROC curve

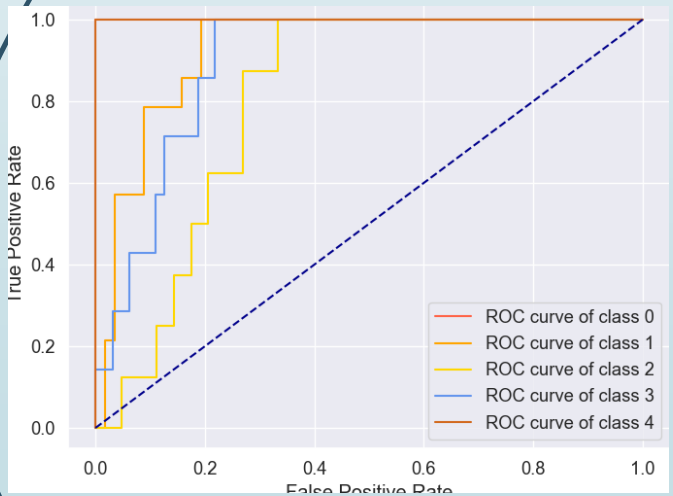
Cleveland: Logistic Regression



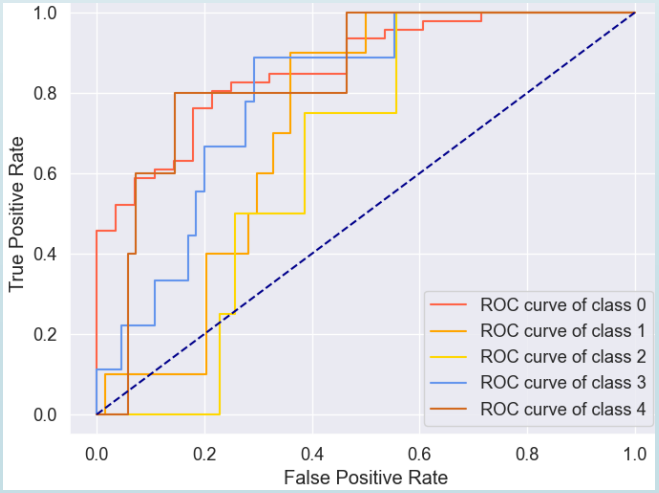
Hungary: Logistic Regression



Cleveland: naive Bayes



Hungary: naive Bayes



Performance measurement

Methods	Cleveland	Hungarian	Vancouver	Switzerland
Log regression accuracy	84%	59%	74%	65%
Log regression AUC-Score	95%	75%	85%	-
Naive Bayes accuracy	77%	46%	54%	-
Naive Bayes AUC-Score	93%	55%	88%	-
SVC linear accuray	86%	58%	84%	-
SVC multi (deg 3) accuracy	68%	62%	38%	-
SVC kernel (rbf) accuracy	58%	62%	20%	-
KNN accuracy	73%	57%	46%	42%
Neural network accuracy	48%	42%	8%	23%

Conclusion

- Variation in accuracy between methods for prediction of heart disease
- For some dataset: prediction not good

- Difficulty to differentiate the different targets
  - Merge all levels for heart disease or more patients data for each targets

- Minor differences between locations

Limitations and Outlook

- Old dataset
- Switzerland dataset unbalanced
- Missing description of features
- Missing (important) features (e.g cholesterol in CH)
- Some selection of features senseless

- Fine tune models generators
- Use of more features
- Find updated datasets

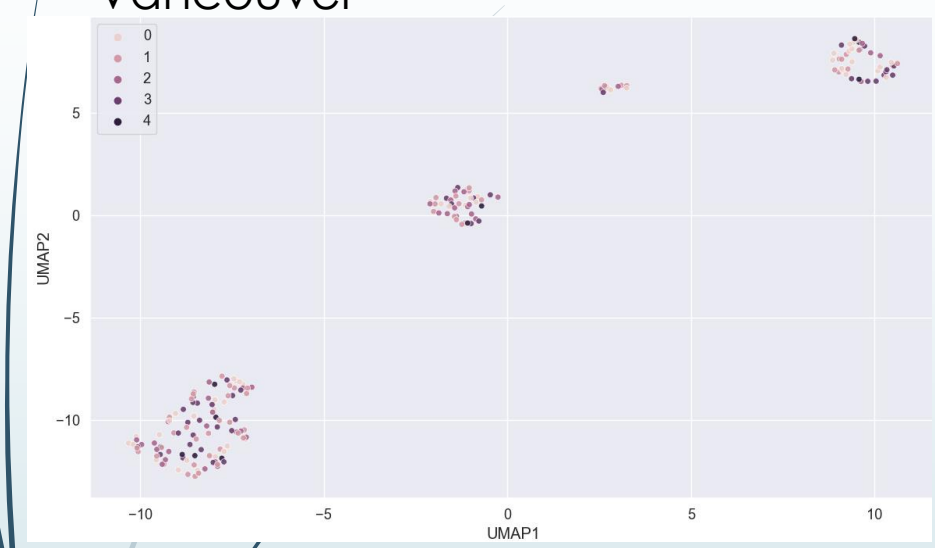
Thank you for your attention

References

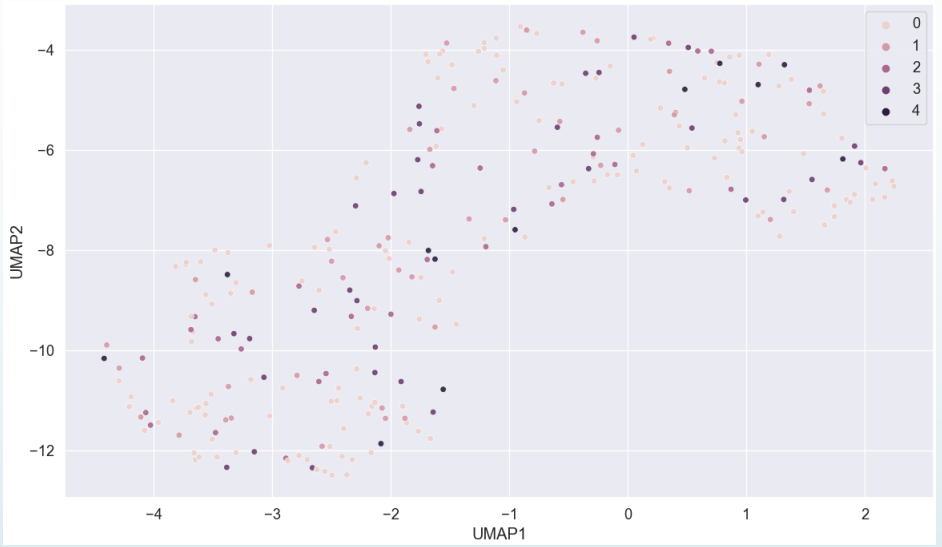
- D'Agostino R. B., Vasan R. S., Pencina M. J., Wolf P. A., Cobain M., & Massaro J. M., Kannel W. B. (2008). General Cardiovascular Risk Profile for Use in Primary Care. *Circulation*, *117*(6), 743-753. doi:10.1161/CIRCULATIONAHA.107.699579
- Dalen, J. E., Alpert, J. S., Goldberg, R. J., & Weinstein, R. S. (2014). The epidemic of the 20(th) century: coronary heart disease. *The American journal of medicine*, *127*(9), 807-812. doi:10.1016/j.amjmed.2014.04.015
- Dinu, M., Pagliai, G., & Sofi, F. (2017). A Heart-Healthy Diet: Recent Insights and Practical Recommendations. *Current cardiology reports*, *19*(10), 95. doi:10.1007/s11886-017-0908-0
- Kreatsoulas, C., & Anand, S. S. (2010). The impact of social determinants on cardiovascular disease. *The Canadian journal of cardiology*, *26 Suppl C*(Suppl C), 8C-13C. doi:10.1016/s0828-282x(10)71075-8
- Monteiro, C. A., Moubarac, J.-C., Cannon, G., Ng, S. W., & Popkin, B. (2013), Ultra-processed products are becoming dominant in the global food system. *obesity reviews, 14*(S2), 21-28. doi:10.1111/obr.12107
- WHO. (2010). *Global status report on noncommunicable diseases 2010* (). Retrieved April 02, 2021 from
- https://apps.who.int/iris/bitstream/handle/10665/44579/9789240686458_eng.pdf;jsessionid=842AF06136BEC4C69DEFE189288DFED6?sequence=1
- WHO. (2017). *Cardiovascular diseases (CVDs)*. Retrieved April 02, 2021 from https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
- WHO. (2020). *The top 10 causes of death*. Retrieved April 02, 2021 from https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death
- https://twitter.com/who_europe/status/1153277944531488768
- http://content.time.com/time/covers/0,16641,19810601,00.html
- https://feedabrain.com/change-of-heart-on-fat/
- https://twitter.com/skynews/status/936351789464551426
- https://www.berelianimd.com/blog/i-cant-believe-its-not-butter-thats-been-clogging-my-arteries-for-the-last-60-years

UMAP

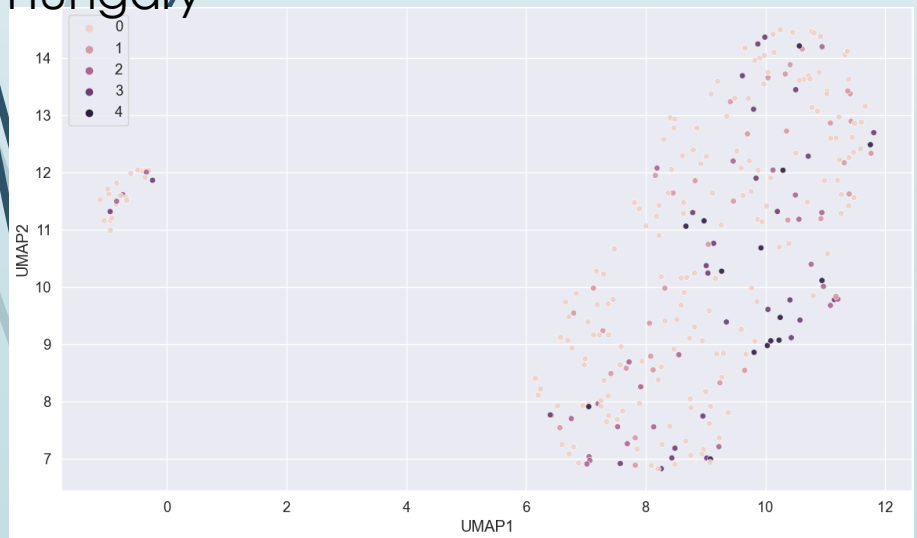
Vancouver



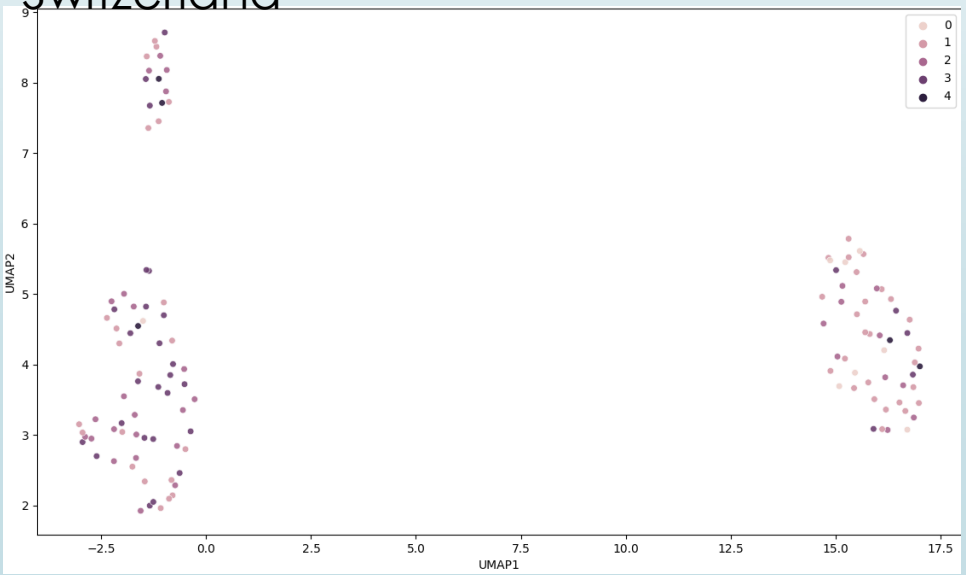
Cleveland



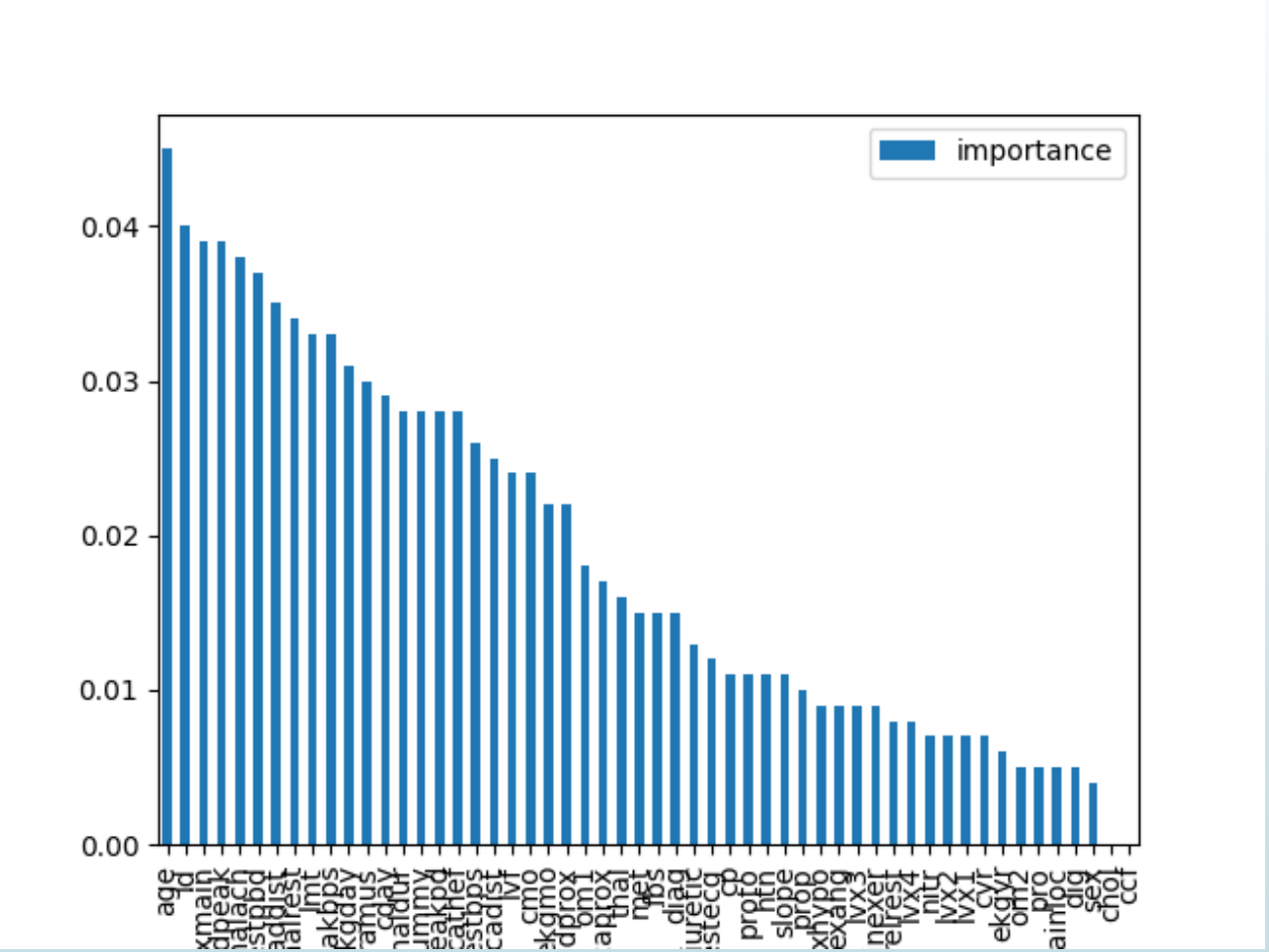
Hungary



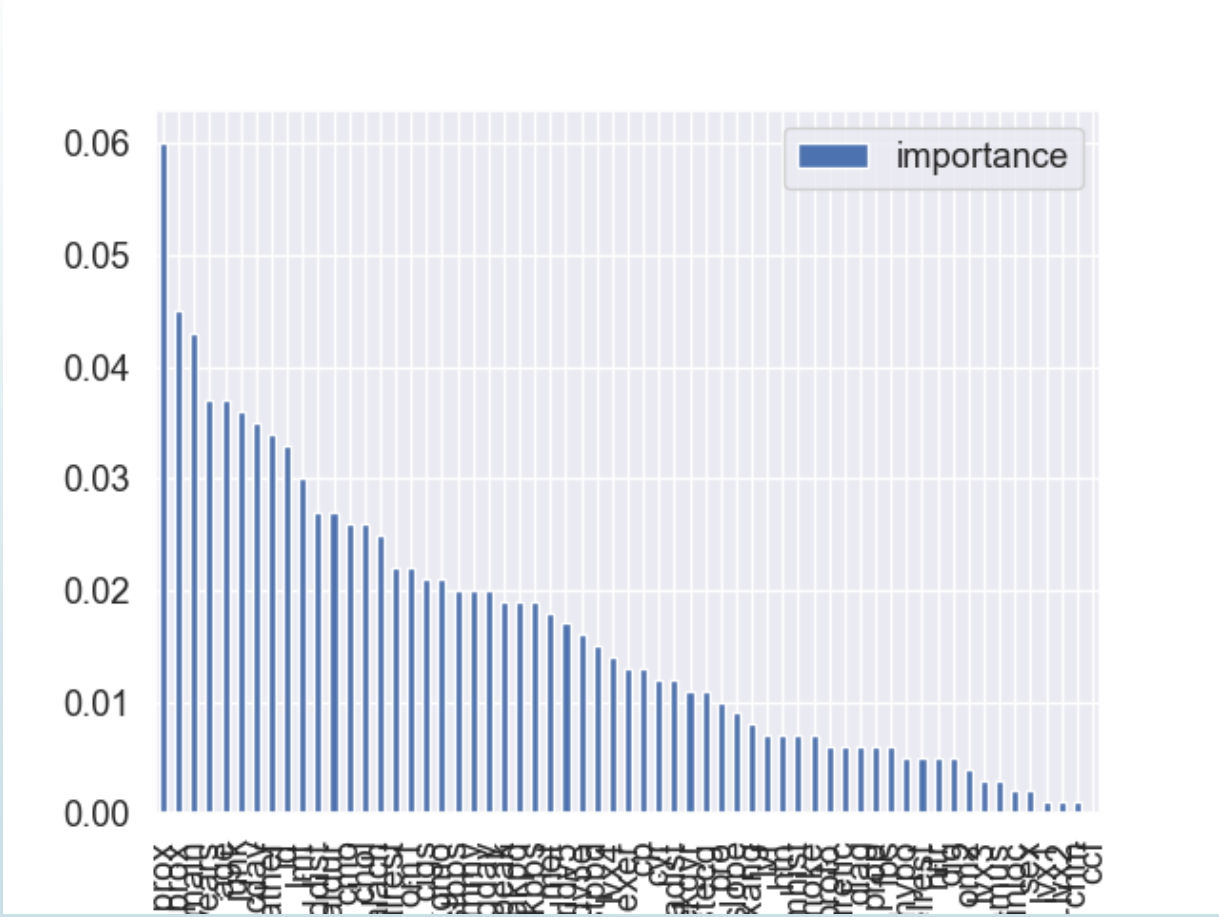
Switzerland



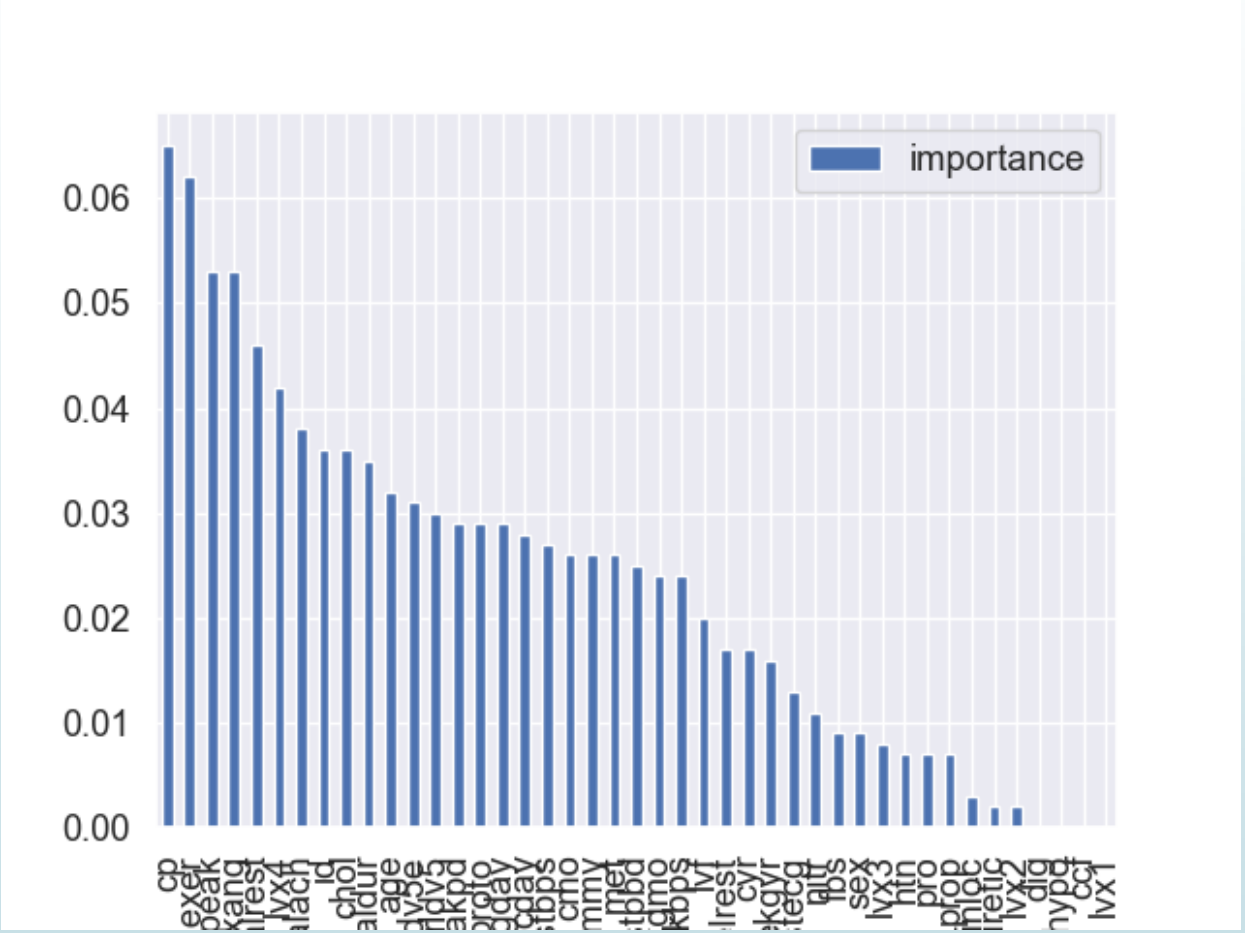
Feature importance Switzerland



Feature importance Vancouver



Feature importance Hungary



Feature importance Cleveland