Healthcare Analytics

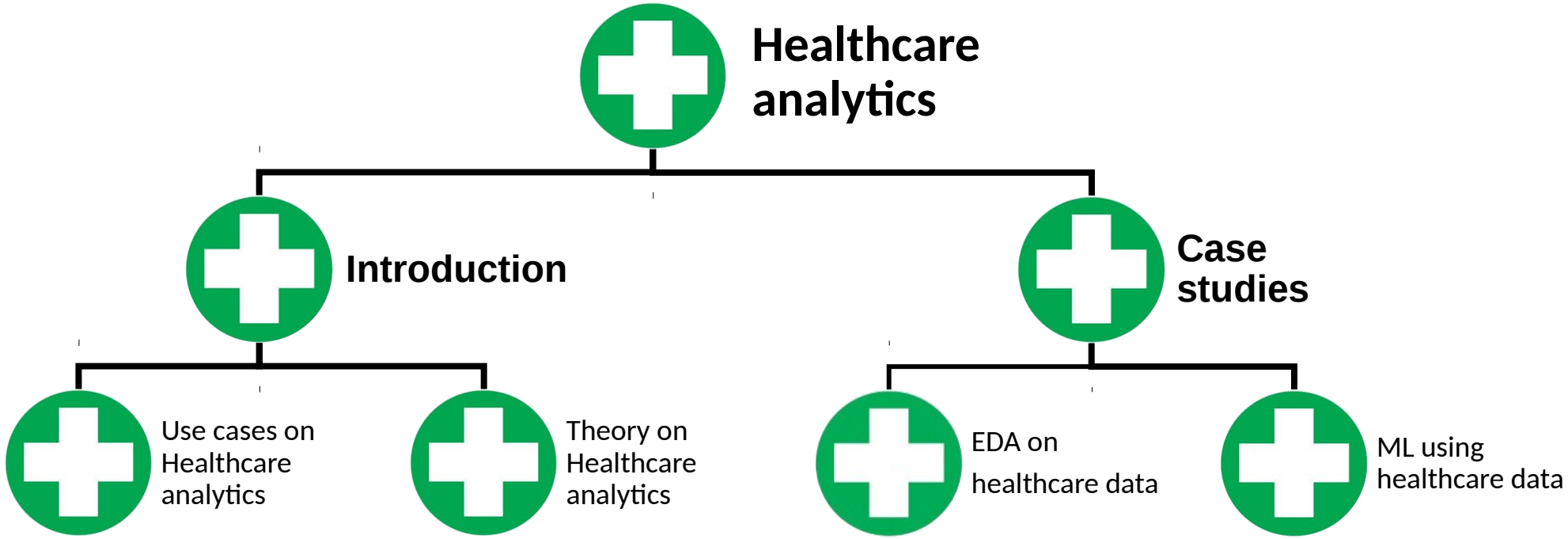
Expectations from Bootcamp

- This is an add on course
- Not meant for any kind certification per-se
- Expand your knowledge on how one can apply Analytics in Healthcare domain
- Participants expected to know/learn the algorithms and their properties
- The focus is on using and building applications for healthcare not on theory
- Expect you can code in Python and work with Tableau
- Drop mails for queries that haven’t been answered during the session

Speaker

Healthcare analytics SME

Case study structure



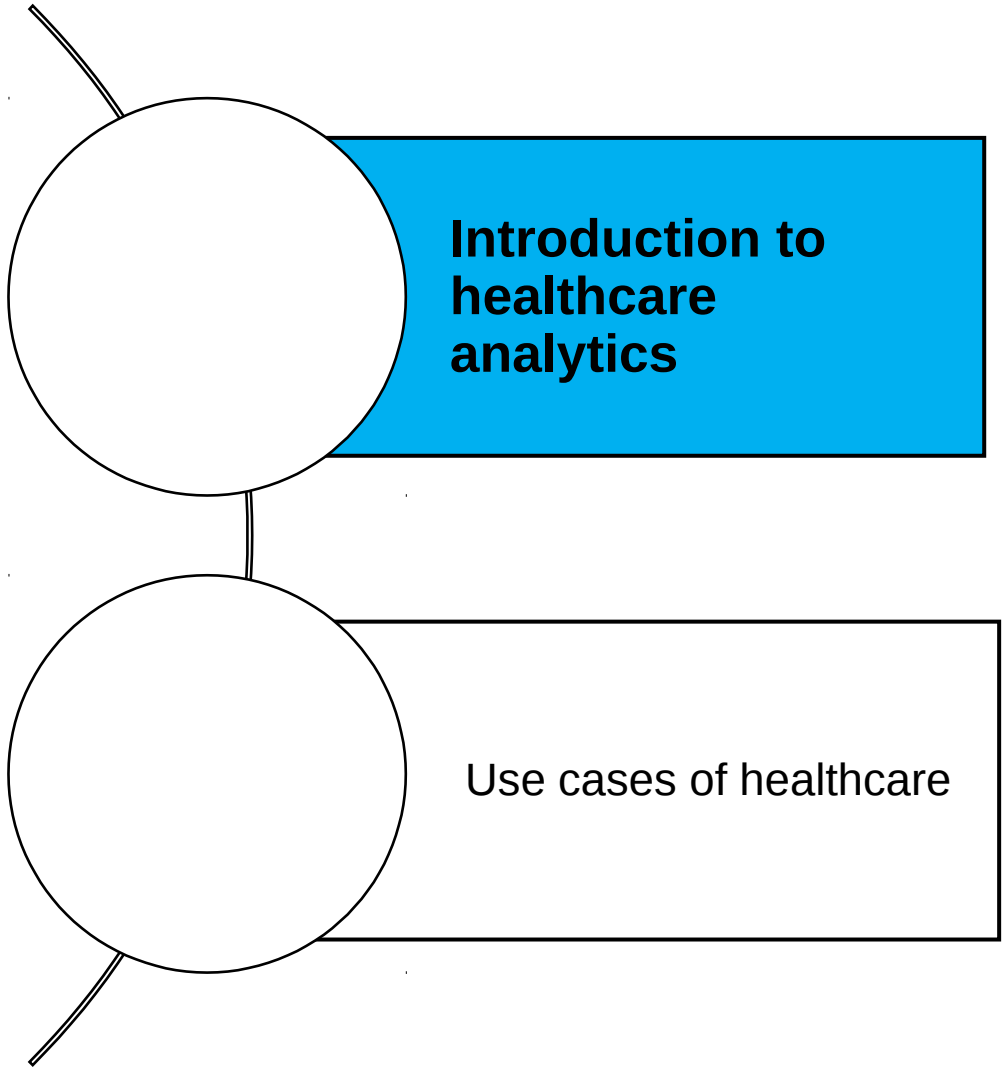
Healthcare Analytics Sessions

Session	Topic	Tools / Materials
1	Introduction to Healthcare analytics	Pre-reads
2	EDA using TB dataset	Tableau
3	Regression analysis to predict Heart disease	Python
4	Decision tree analysis to categorize high risk patients	Python

Scope of Health care analytics

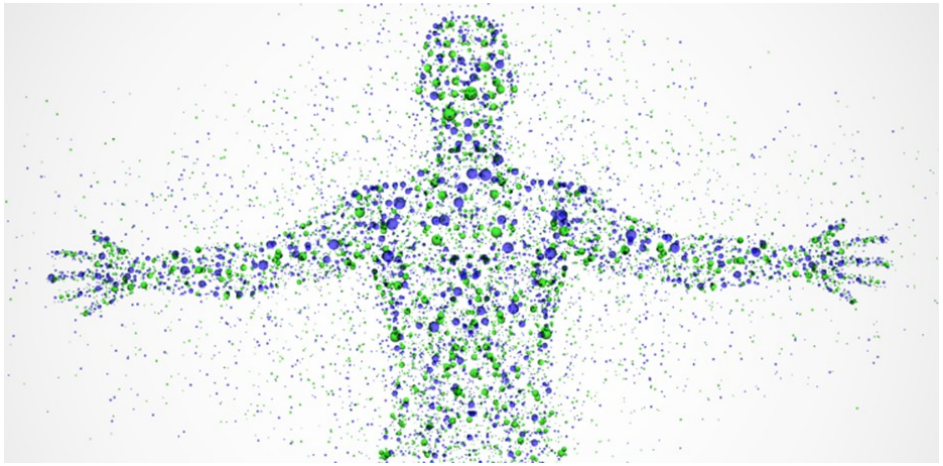
In this Bootcamp we will use healthcare data to develop Machine Learning solutions which helps us predict onset of diseases of heart disease and identify high risk patients.

Agenda



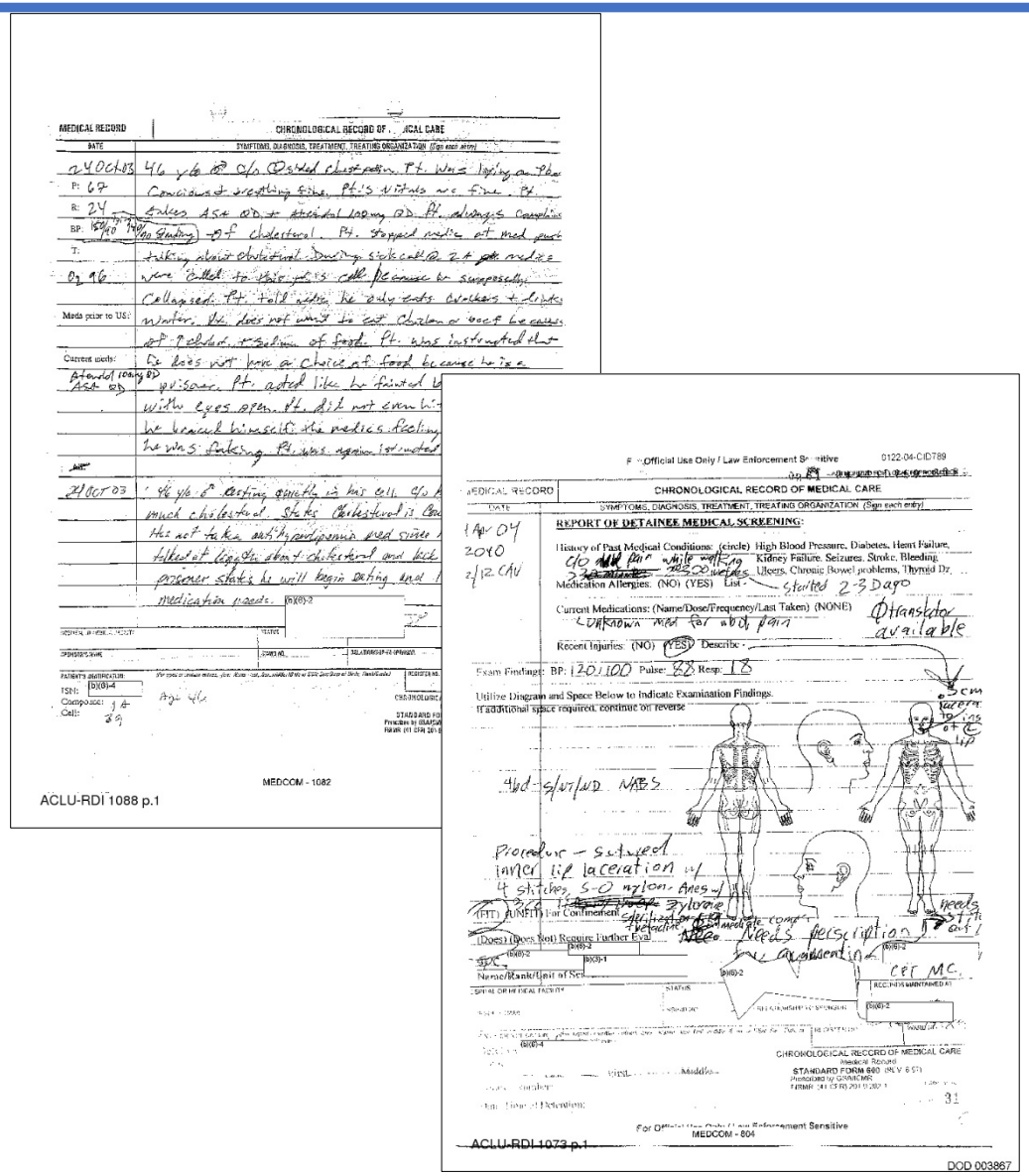
Healthcare Analytics

What is healthcare?



Healthcare allows improvement of health via the prevention, diagnosis, and treatment of disease.

Healthcare - just 10 years back



Healthcare had limited/Restricted access to data

- Only patient data, like lab/doctor reports
- Data was recorded manually
- Realtime analysis was missing
- Treatment was based primarily on Doctors judgement

Major care providers have recognized data value

upx
Move up in life

TMC
TEXAS
MEDICAL
CENTER

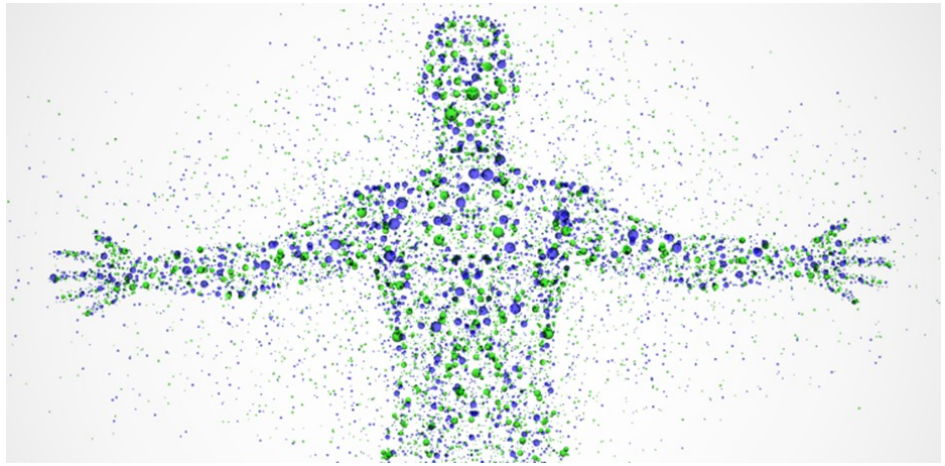
JOHNS HOPKINS
SCHOOL of MEDICINE

Mount
Sinai
Hospital

Major care providers have recognized data value

upx
Move up in life

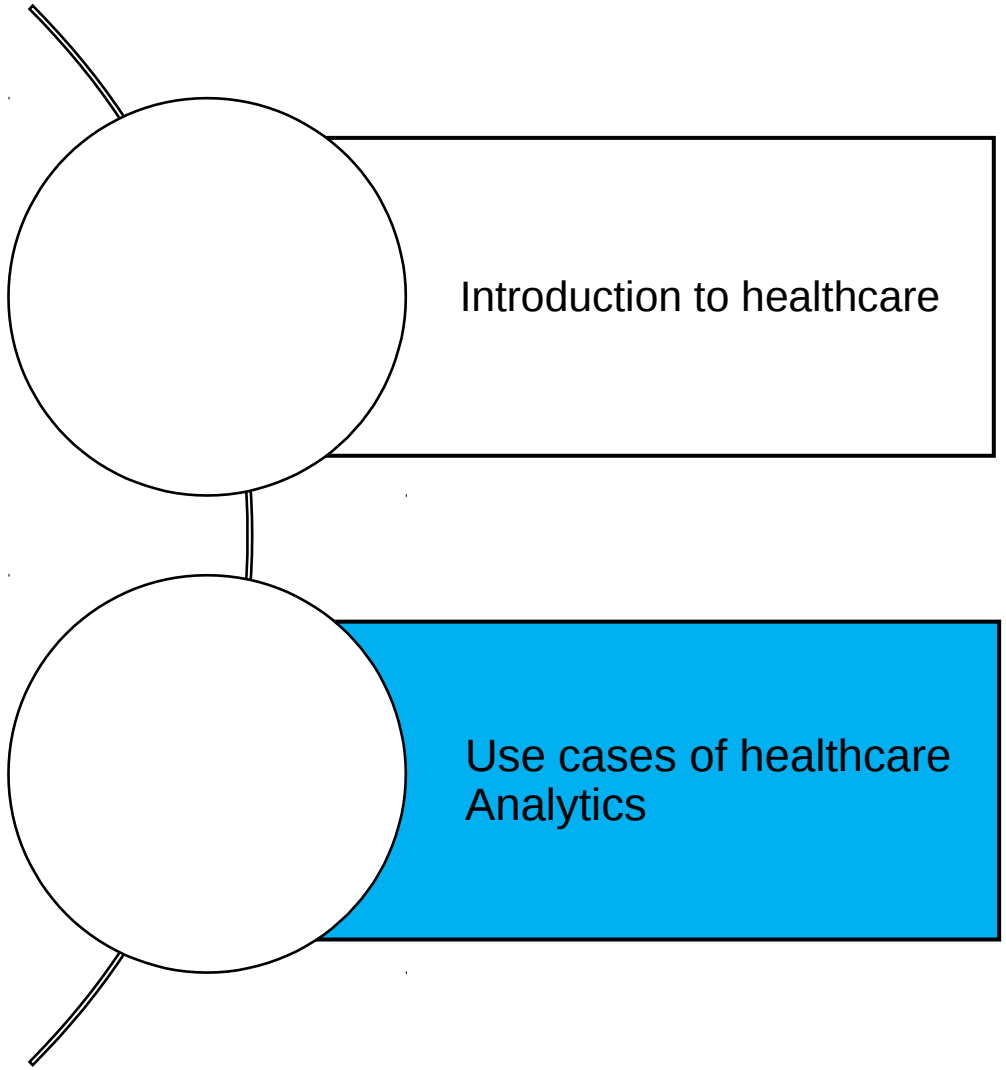
What is healthcare analytics?



Healthcare analytics now allows

- Collecting both Health, Behavioral data
- Identify best treatments for diseases
- Reduce excessive spending on health care
- Improve overall health for individual patients
- Identify preventive measures against illnesses

Agenda



Healthcare Analytics

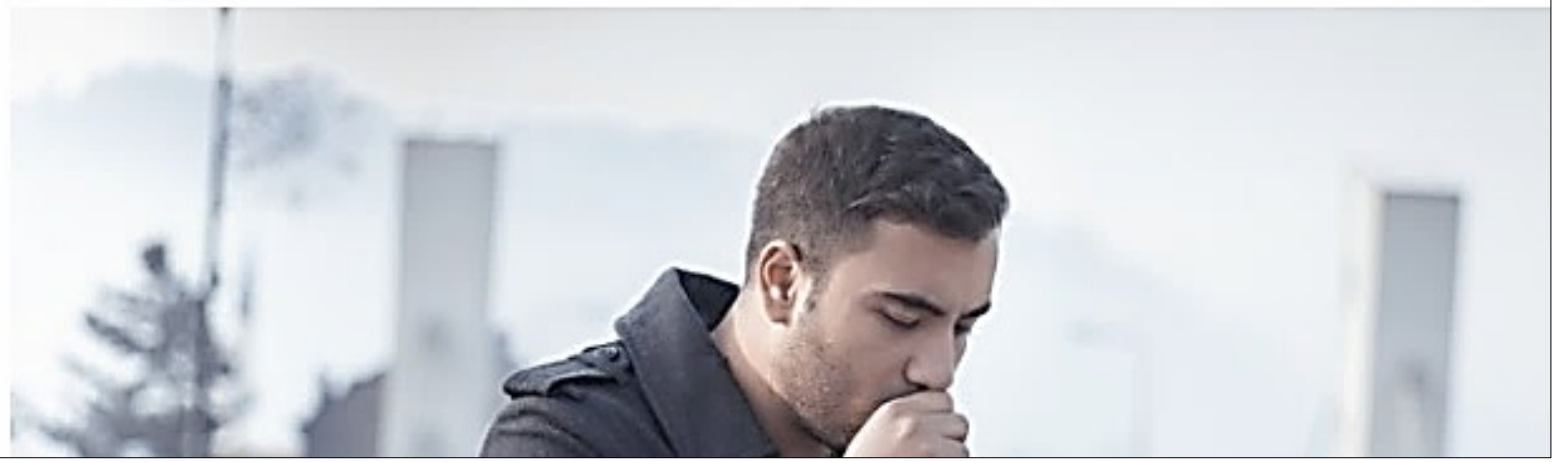
Use Case 1

Applications in personalized medicine

What's the Best Medicine for a Cough?

Ask Well

By RICHARD KLASCO, M.D. APRIL 6, 2018



The New York Times

Symptoms of a cough & cold

- ❑ Stuffy nose
- ❑ Sore throat
- ❑ Shortness of breath

- ❑ Lasts up to a week
- ❑ In severe case months

Over the counter medication

In most cases you end up buying generic cough / cold medication

Which may only alleviate some symptoms

Little evidence that these medications are effective
for all!

So, how can data and analytics can help us seek medication suitable for our personal profile for commonly occurring illnesses?

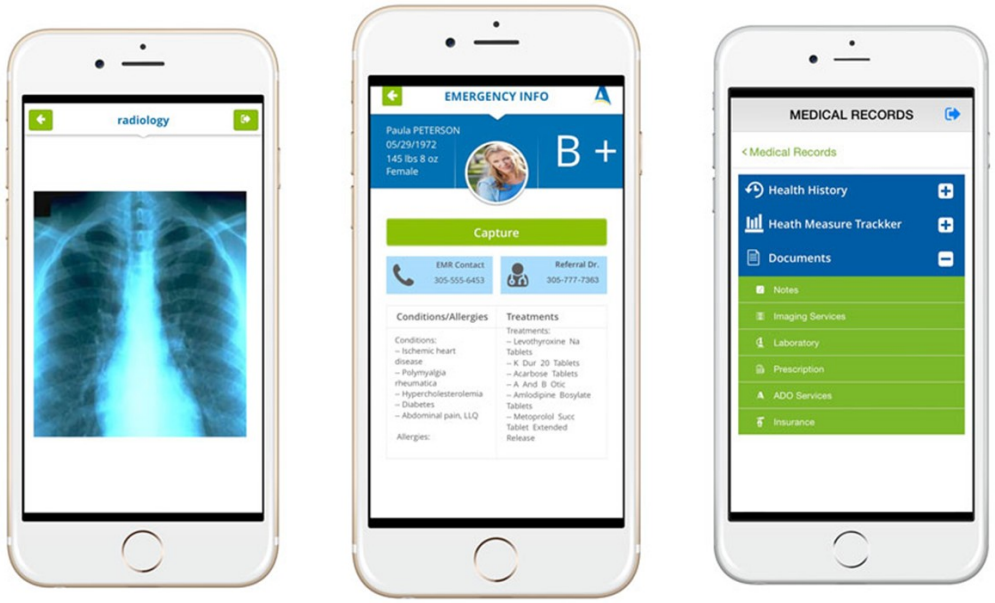
The insights depends on which **data** is available from you personal medical history, on your location, online behavior, recent climatic trends etc.

Data from your personal health records

Electronic health records give information on your Personal Health profile

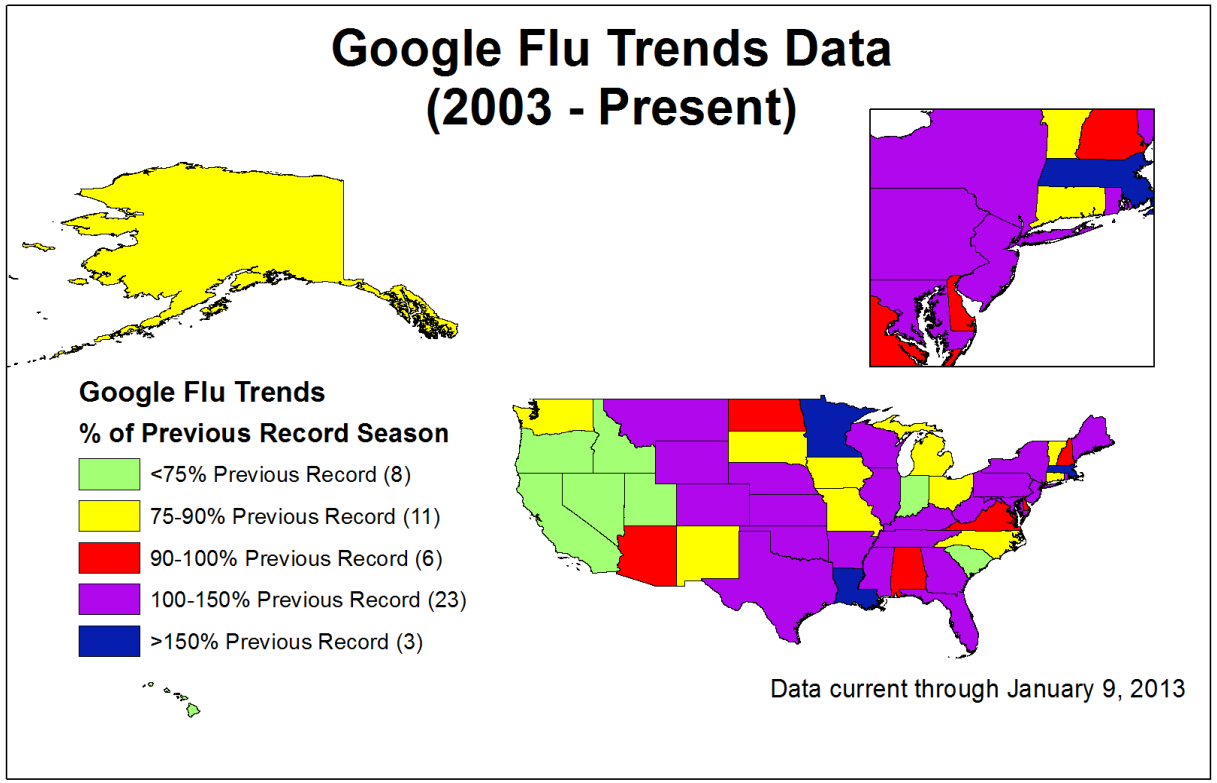
Behavioral data allows segmenting people, understanding and predicting their health care needs

Example of personal medical data



- Allergies
- Laboratory test results
- Prescription record
- Vaccinations

Behavioral data of all people can be used



- Based on user purchases of medication online
- User search analysis for medication online
- Based on **local weather conditions**
- Based on **epidemiology trends**
- Current disease spread **(e.g. Bird flu)**

Use this data healthcare applications can prescribe preventive measures

Your personal Healthcare apps could….

- Update you on **local health trends**
- Inform on **preventive measures** on common ailments
- Inform on active exercises and diet changes
- Inform on **suitable cough** / **cold medicine** for your profile
- Alleviate the symptoms to the maximum extent possible

In severe cases physicians could….

- Validate your symptoms
- Prescribe treatment actions
- Prescribe targeted cough / cold medications

Prevention is better

Insights from health care analytics not only provide information on effective medication to allow targeted treatment, but also allow predict & prevent contracting ailments

Healthcare apps that you can examine

Prescription advice
Validating Symptoms

Medscape WebMD

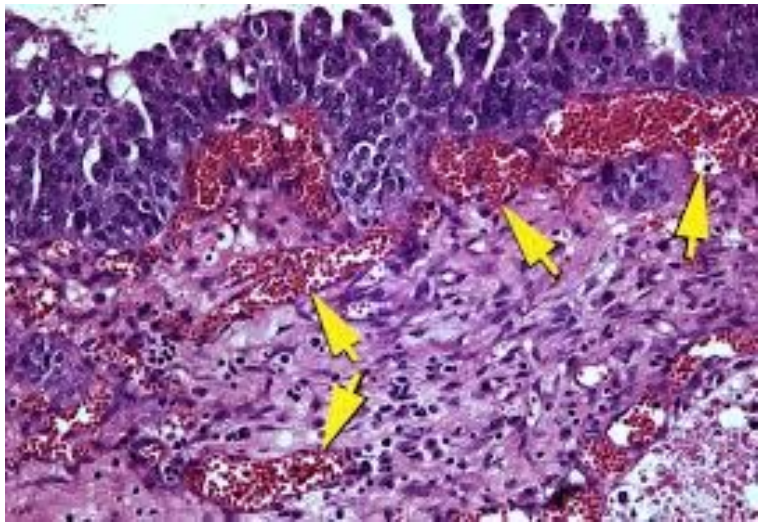
Health diet and
exercise recommendations

Mental health improvement

happify DAILY

Use Case 2

Predictive screening in healthcare



About 10 of every 100 women who develop **ovarian cancer** will have a BRCA1 or BRCA2 mutation. For women who have a BRCA1/2 mutation, the risk for early **breast cancer** and **ovarian cancer** is greatly increased. In some families **breast cancer** or **ovarian cancer** will occur due to inherited mutations in genes other than BRCA1/2. Jun 17, 2014

Predictive screening in healthcare

assessyourrisk by bright pink

1 in 8 women will develop breast cancer at some point in her lifetime. 1 in 75 will develop ovarian cancer.

Bright Pink created this tool to help you assess your personal risk level for breast and ovarian cancers. The more you know, the better prepared you are to take actions that can help reduce your risk.

Your body. Your life.
Don't leave it up to chance.

ASSESS YOUR RISK

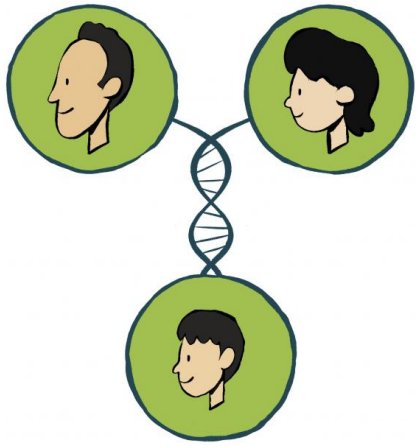
Use Case 2

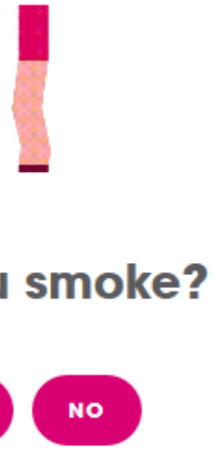
assessyourrisk.org

Healthcare and lifestyle data is used for prediction

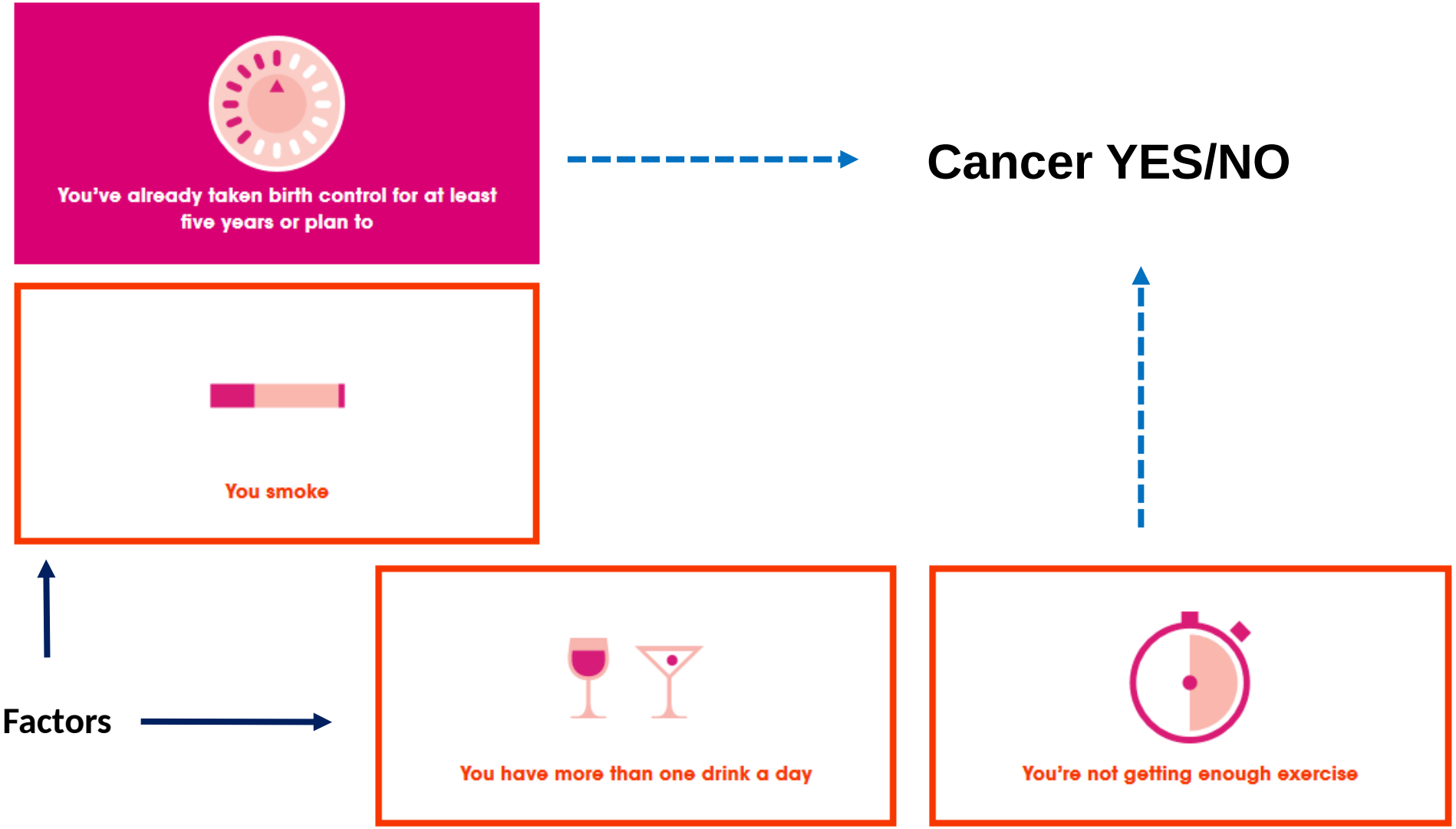
Collect data concerning

- Personal health (symptoms)
- Clinical tests results
- Family health history
- Ancestry
- Behavioral data (Drinking, Smoking habits)





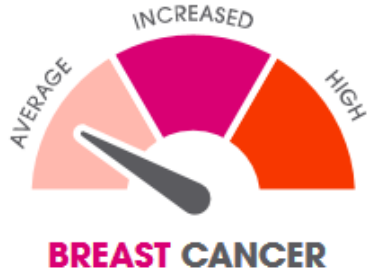
Following Cancer risk factors are analyzed



Use Case 2

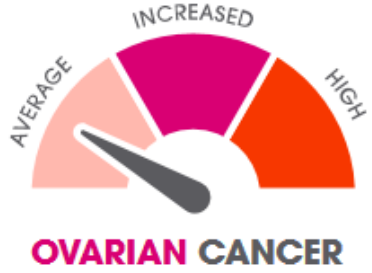
Risk is calculated by predictive tools

Baseline Risk for Breast Cancer



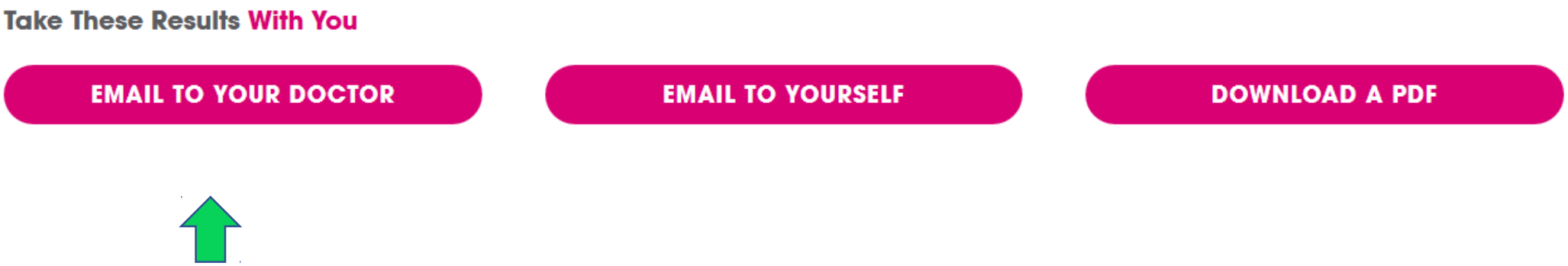
Your answers to our personal and family health history questions suggest that you are at average baseline risk for breast cancer — meaning you have up to a 12% chance of developing breast cancer in your lifetime. It's important to note, however, that 75% of breast cancers are diagnosed in average-risk women. It's great that you've taken the first step in being proactive about your health by assessing your risk: knowing the modifiable risk factors for the disease and maintaining a healthy lifestyle are still important even when your risk is average.

Baseline Risk for Ovarian Cancer



Your answers to our personal and family health history questions suggest that you are at average baseline risk for ovarian cancer — meaning you have up to a 1.3% chance of developing ovarian cancer in your lifetime. It's important to note, however, that 75% of ovarian cancers are diagnosed in average-risk women. It's great that you've taken the first step in being proactive about your health by assessing your risk: knowing the modifiable risk factors for the disease and maintaining a healthy lifestyle are still important even when your risk is average.

Allows generating actionable insights



Take immediate action!

Suggest life style changes

upx
Move up in life

YOUR RISK MANAGEMENT PLAN

Be Proactive About Your Health

Finding out your baseline risk level is the first step in being proactive. Now it's time to turn that awareness into action. Talk to your healthcare provider about these guidelines to determine which strategies are best for you.

About Genetic Counseling and Testing

If you want to confirm that your baseline risk truly is average, consider talking with your healthcare provider or a genetic counselor about whether genetic testing is right for you. If you need help finding a genetic counselor to talk to in person or on the phone, visit the National Society of Genetic Counselors for more information.

About Genetic Counseling and Testing

Keep Up with Well-Woman Exams

These annual exams, which are covered in all insurance policies, should include:

- Clinical breast and pelvic exams starting at age 20
- Pap smear — though it's important to know that these check for cervical cancer, not for ovarian cancer, and you may not get one every year
- Mammograms yearly beginning at age 40

More about well-woman exams

Know Your Normal

Knowing what's normal for your body is key to early detection, since it will help you notice when something out of the ordinary is happening. To start with, get to know the typical look and feel of your breasts -- everybody's breasts are different! While you do this, also make sure to learn the signs and symptoms of breast cancer so you can recognize when something that's not normal for you might be problematic. The same is true for knowing how your body typically reacts with menstruation and what sort of digestive fluctuations are normal for you, along with the signs and symptoms of ovarian cancer. If you do notice an atypical change in your breasts or in your digestion or menstruation and it persists for 2–3 weeks, be sure to bring it to your healthcare provider's attention.

More about breast and ovarian self-awareness

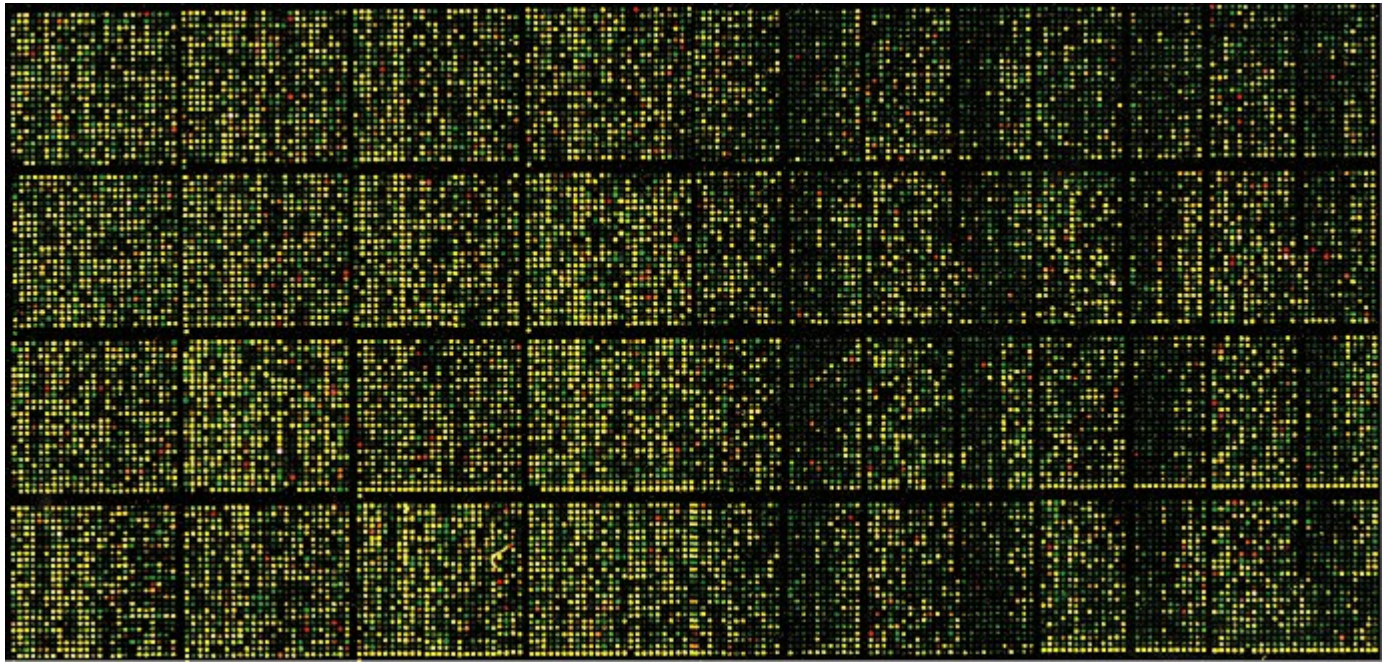
Risk Reduction Options

Your best risk reduction options center on maintaining a healthy lifestyle. Improving your overall health can improve your overall risk level. Review the "How Life Affects Risk" section above to learn ways you can make a difference.

More about lifestyle risk reduction

Which other fields apply healthcare analytics?

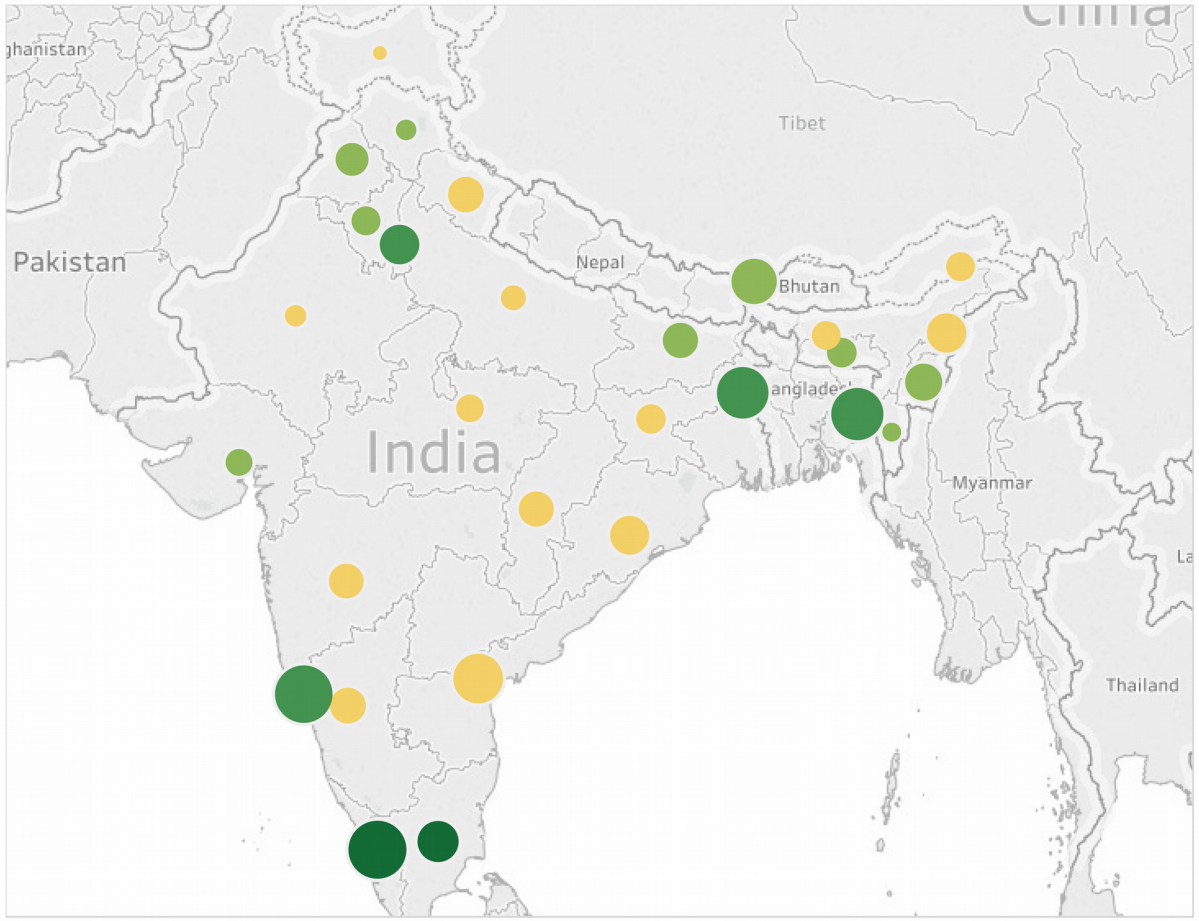
In which fields healthcare analytics is applied?



Medical Screening

Identifying risk of cancer
or heart disease in you family

In which fields healthcare analytics is applied?

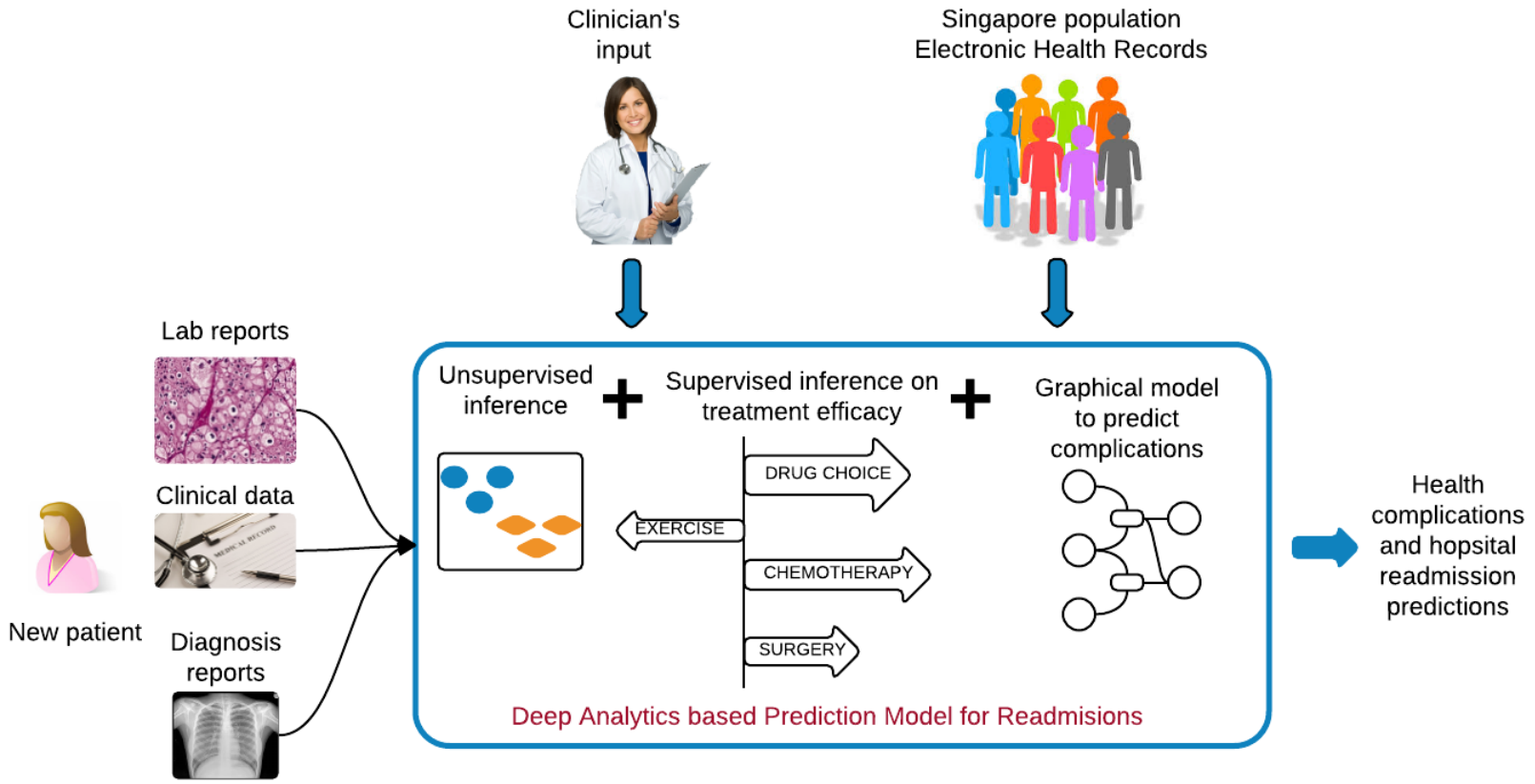


Map based on Longitude (generated) and Latitude (generated). Color shows NFHS III(2005-06)-Number of women in age group 15-49 per 100-000 reported Diabetes-Total. Size shows NFHS III(2005-06)-Number of men in age group 15-49 per 100-000 reported Diabetes-Total. Details are shown for India/States/UTs.

Epidemiology

Understanding why more Indians are becoming diabetic

In which fields healthcare analytics is applied?



Clinical analytics

Providing better treatment

Summary of use cases

- Healthcare analytics is a discipline which uses patient data to develop solutions for betterment of health
- Various products like mobile apps to web applications use past data to provide healthcare insights (**prescriptive and predictive insights**)
- The data collected can be coming from individual medical records and from behavioral data

Session 1 – part 2

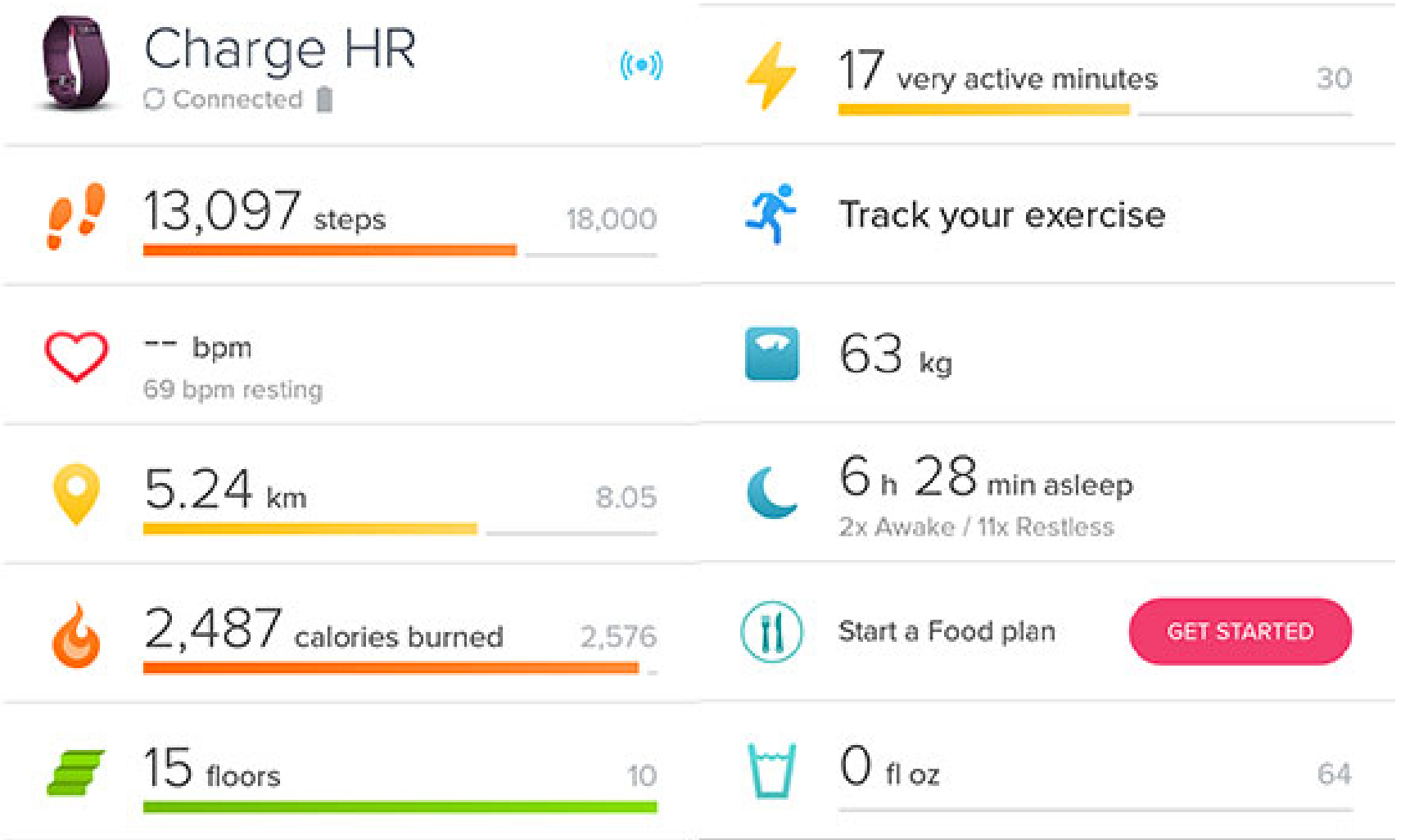
Agenda

upx
Move up in life

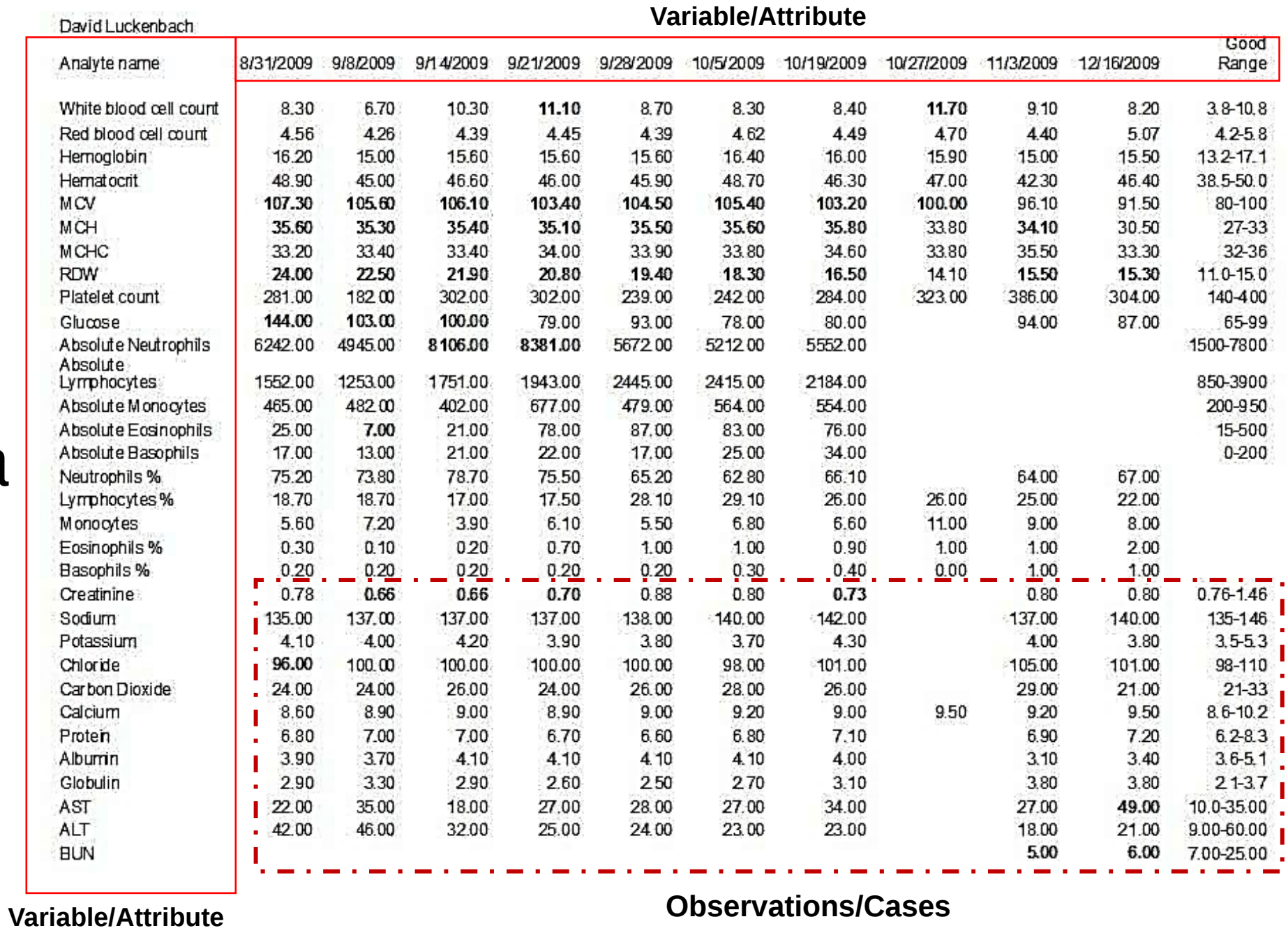
Healthcare Analytics

So what is the nature of data we are dealing in health care analytics scenario?

Structured data



Structured data



David Luckenbach

Analyte name	8/31/2009	9/8/2009	9/14/2009	9/21/2009	9/28/2009	10/5/2009	10/19/2009	10/27/2009	11/3/2009	12/16/2009	Good Range
White blood cell count	8.30	6.70	10.30	**11.10**	8.70	8.30	8.40	**11.70**	9.10	8.20	3.8-10.8
Red blood cell count	4.56	4.26	4.39	4.45	4.39	4.62	4.49	4.70	4.40	5.07	4.2-5.8
Hemoglobin	16.20	15.00	15.60	15.60	15.60	16.40	16.00	15.90	15.00	15.50	13.2-17.1
Hematocrit	48.90	45.00	46.60	46.00	45.90	48.70	46.30	47.00	42.30	46.40	38.5-50.0
MCV	**107.30**	**105.60**	**106.10**	**103.40**	**104.50**	**105.40**	**103.20**	**100.00**	96.10	91.50	80-100
MCH	**35.60**	**35.30**	**35.40**	**35.10**	**35.50**	**35.60**	**35.80**	33.80	**34.10**	30.50	27-33
MCHC	33.20	33.40	33.40	34.00	33.90	33.80	34.60	33.80	35.50	33.30	32-36
RDW	**24.00**	**22.50**	**21.90**	**20.80**	**19.40**	**18.30**	**16.50**	14.10	**15.50**	**15.30**	11.0-15.0
Platelet count	281.00	182.00	302.00	302.00	239.00	242.00	284.00	323.00	386.00	304.00	140-400
Glucose	**144.00**	**103.00**	**100.00**	79.00	93.00	78.00	80.00		94.00	87.00	65-99
Absolute Neutrophils	6242.00	4945.00	**8106.00**	**8381.00**	5672.00	5212.00	5552.00				1500-7800
Absolute Lymphocytes	1552.00	1253.00	1751.00	1943.00	2445.00	2415.00	2184.00				850-3900
Absolute Monocytes	465.00	482.00	402.00	677.00	479.00	564.00	554.00				200-950
Absolute Eosinophils	25.00	**7.00**	21.00	78.00	87.00	83.00	76.00				15-500
Absolute Basophils	17.00	13.00	21.00	22.00	17.00	25.00	34.00				0-200
Neutrophils %	75.20	73.80	78.70	75.50	65.20	62.80	66.10		64.00	67.00	
Lymphocytes %	18.70	18.70	17.00	17.50	28.10	29.10	26.00	26.00	25.00	22.00	
Monocytes	5.60	7.20	3.90	6.10	5.50	6.80	6.60	11.00	9.00	8.00	
Eosinophils %	0.30	0.10	0.20	0.70	1.00	1.00	0.90	1.00	1.00	2.00	
Basophils %	0.20	0.20	0.20	0.20	0.20	0.30	0.40	0.00	1.00	1.00	
Creatinine	0.78	**0.66**	**0.66**	**0.70**	0.88	0.80	**0.73**		0.80	0.80	0.76-1.46
Sodium	135.00	137.00	137.00	137.00	138.00	140.00	142.00		137.00	140.00	135-146
Potassium	4.10	4.00	4.20	3.90	3.80	3.70	4.30		4.00	3.80	3.5-5.3
Chloride	**96.00**	100.00	100.00	100.00	100.00	98.00	101.00		105.00	101.00	98-110
Carbon Dioxide	24.00	24.00	26.00	24.00	26.00	28.00	26.00		29.00	21.00	21-33
Calcium	8.60	8.90	9.00	8.90	9.00	9.20	9.00	9.50	9.20	9.50	8.6-10.2
Protein	6.80	7.00	7.00	6.70	6.60	6.80	7.10		6.90	7.20	6.2-8.3
Albumin	3.90	3.70	4.10	4.10	4.10	4.10	4.00		3.10	3.40	3.6-5.1
Globulin	2.90	3.30	2.90	2.60	2.50	2.70	3.10		3.80	3.80	2.1-3.7
AST	22.00	35.00	18.00	27.00	28.00	27.00	34.00		27.00	**49.00**	10.0-35.00
ALT	42.00	46.00	32.00	25.00	24.00	23.00	23.00		18.00	21.00	9.00-60.00
BUN									**5.00**	**6.00**	7.00-25.00

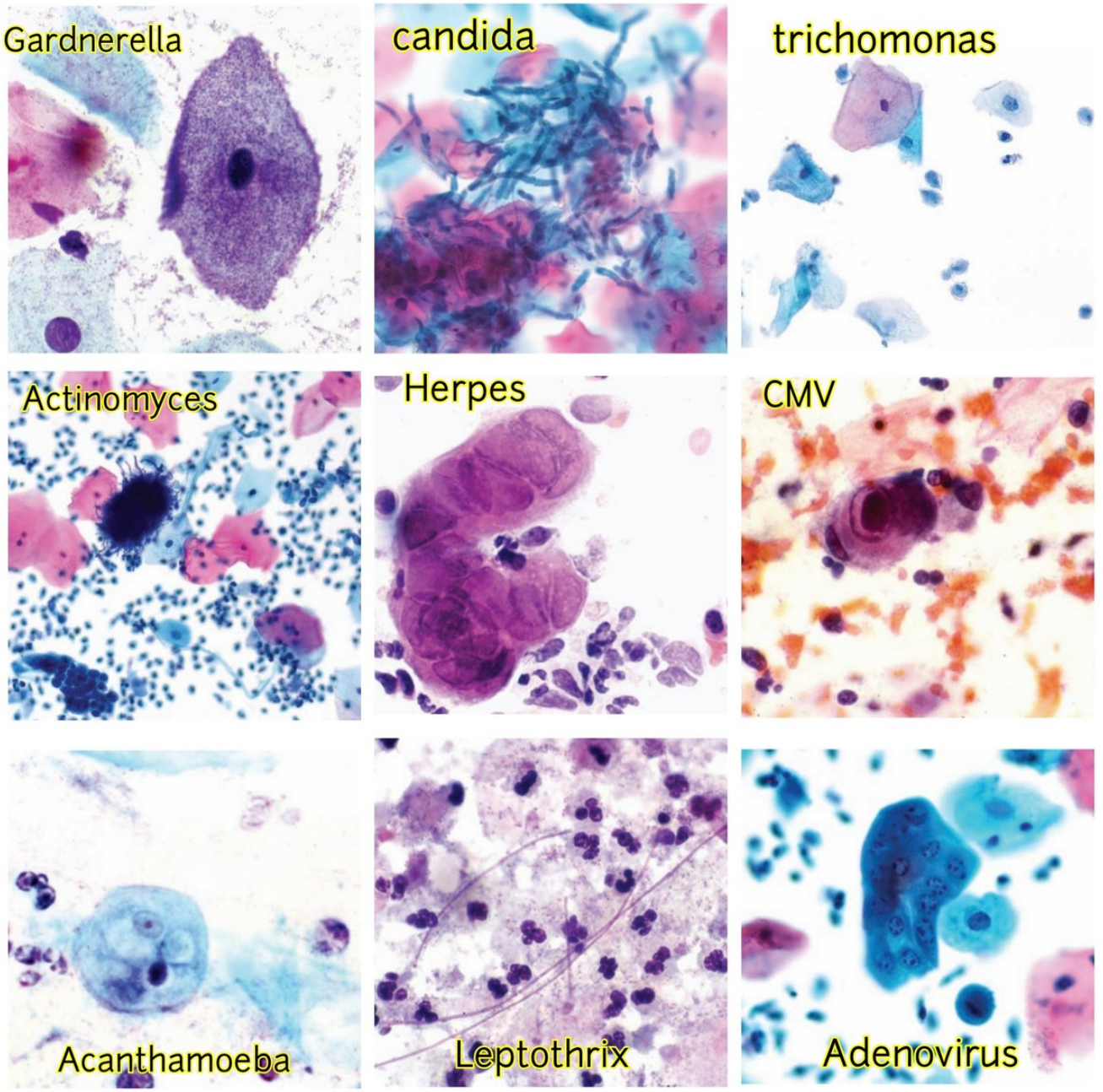
80% of all clinical data is unstructured

Unstructured data

(radiology images)

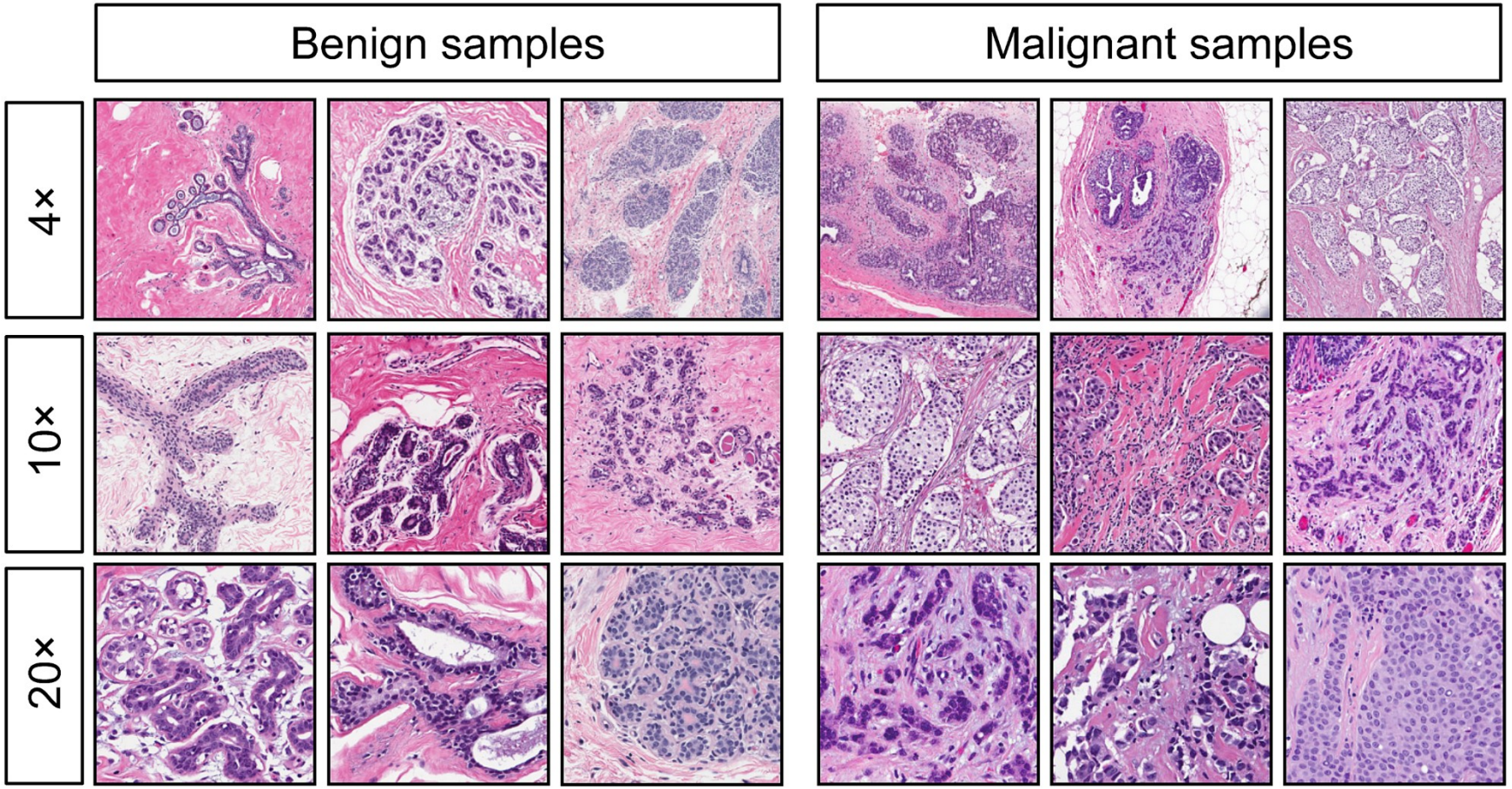
Unstructured data

(microscopic image data)

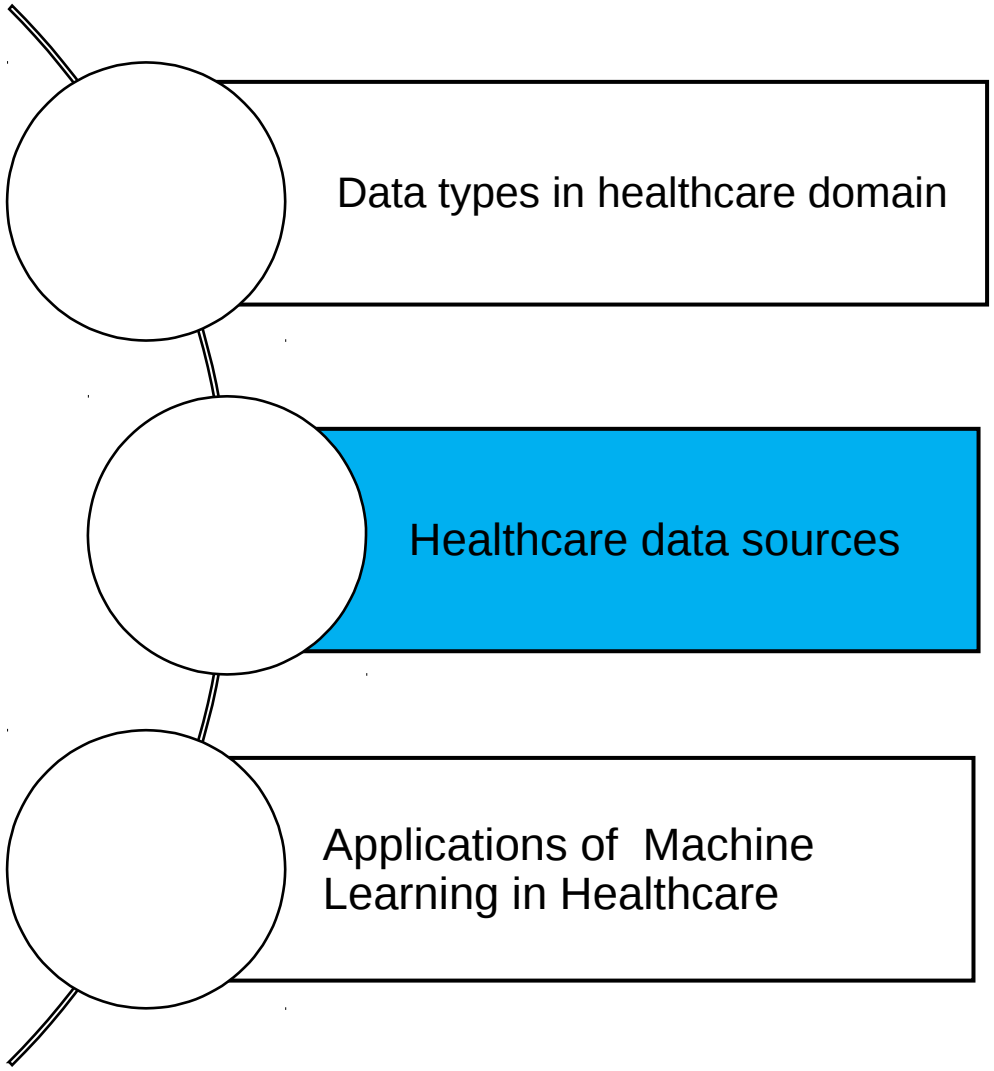


Unstructured data

(pathology/histology)



Agenda



Healthcare Analytics

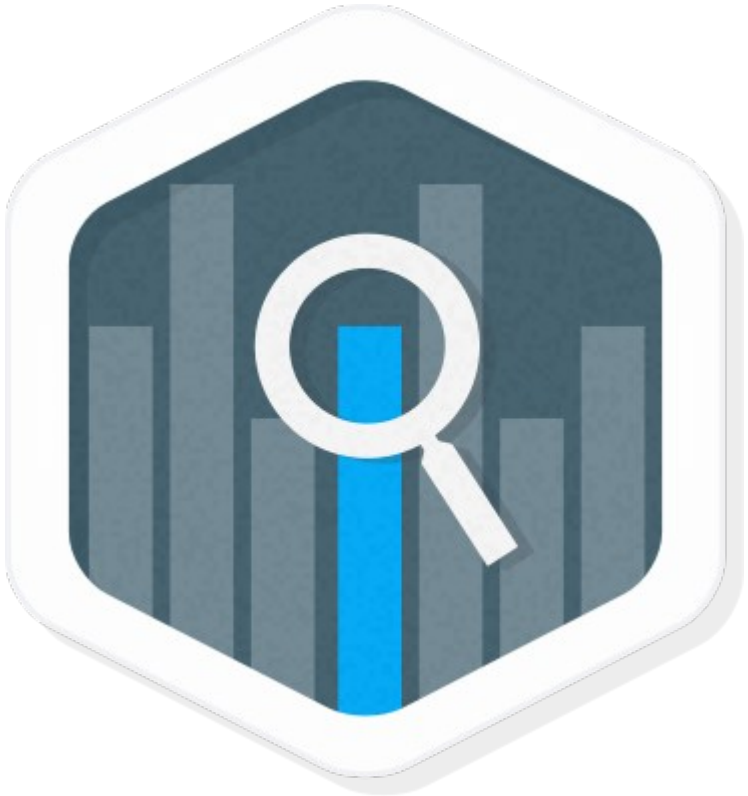
What are the sources of healthcare data?

Example of Primary data sources

Clinical	Pharmaceutical	Insurance	Behavior
• EMR • Clinical notes • Lab reports • XRAYS • Medical tests	• Clinical trials • Patient drug history	• Patient Medical claims	• Health questionnaire • Exercise data • Patient responses • Social media data

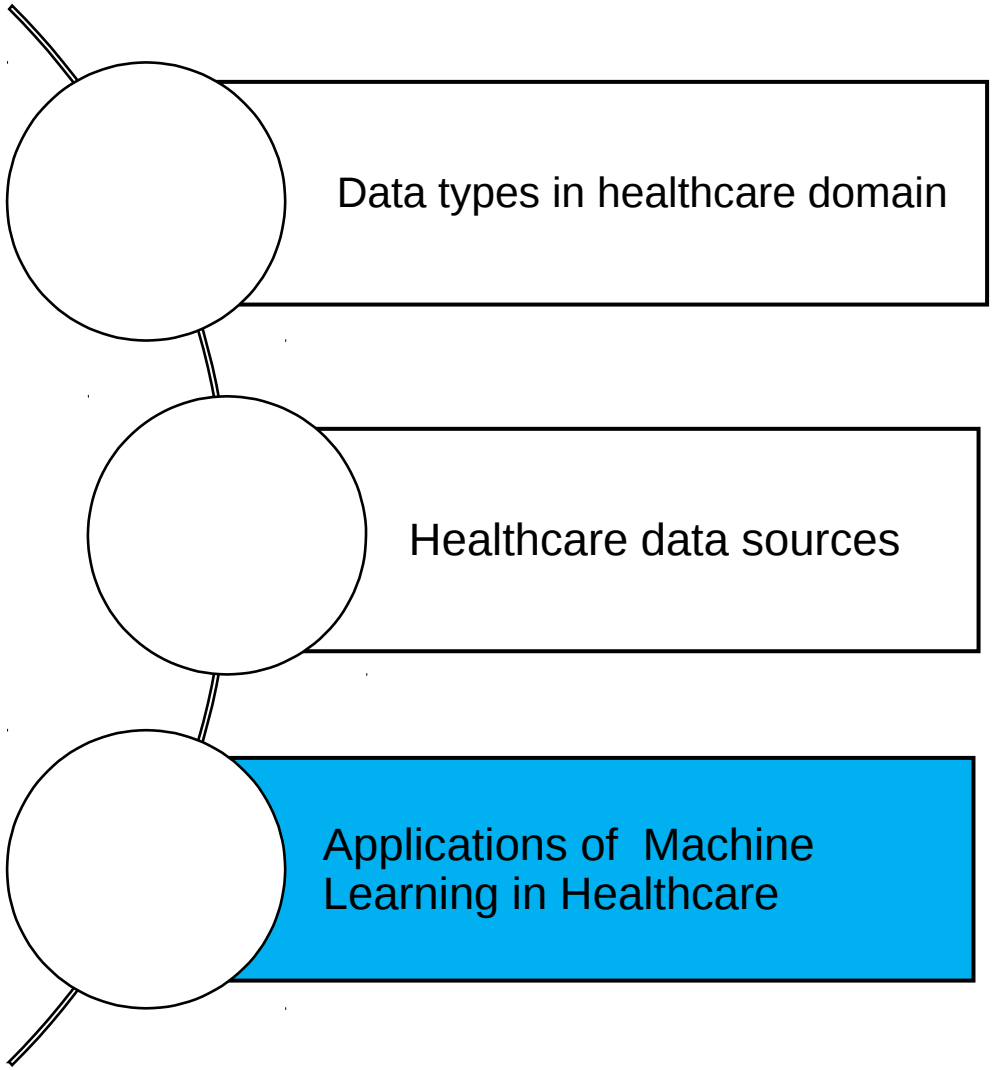
What secondary sources of data do we have?

Secondary data sources



- Hospital operations data
- Insurance claim data (blinded)
- External market research data
- Population studies data

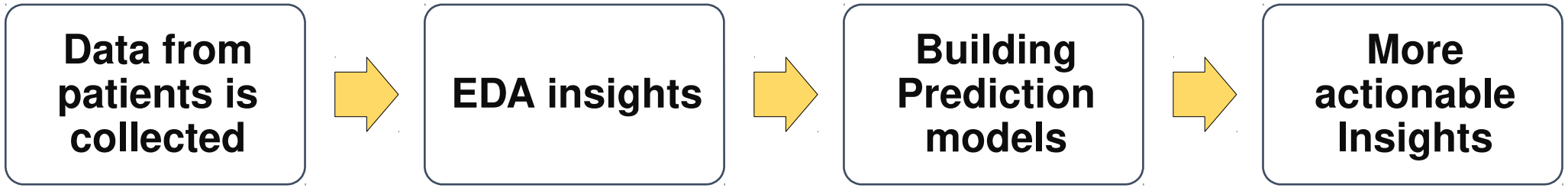
Agenda



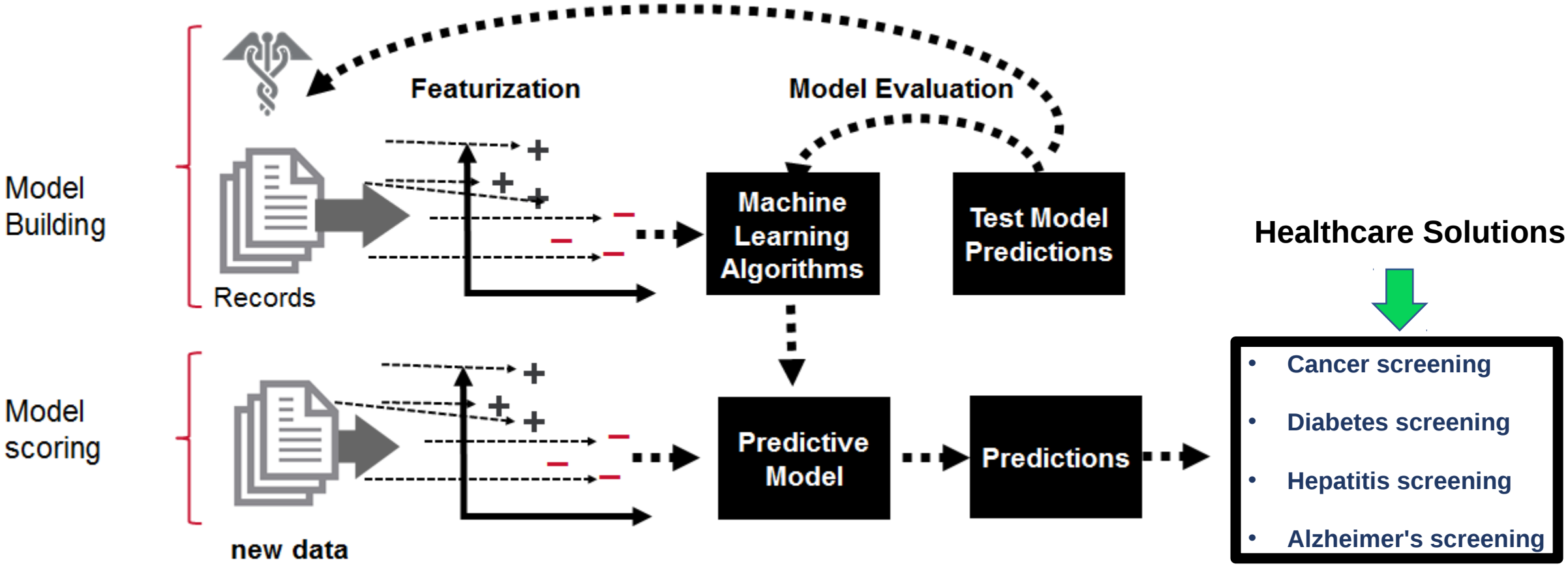
Healthcare Analytics

How do we develop ML solutions using health-care data?

How does ML help with more insights?



Developing Healthcare Analytics products



Summarizing

Data types

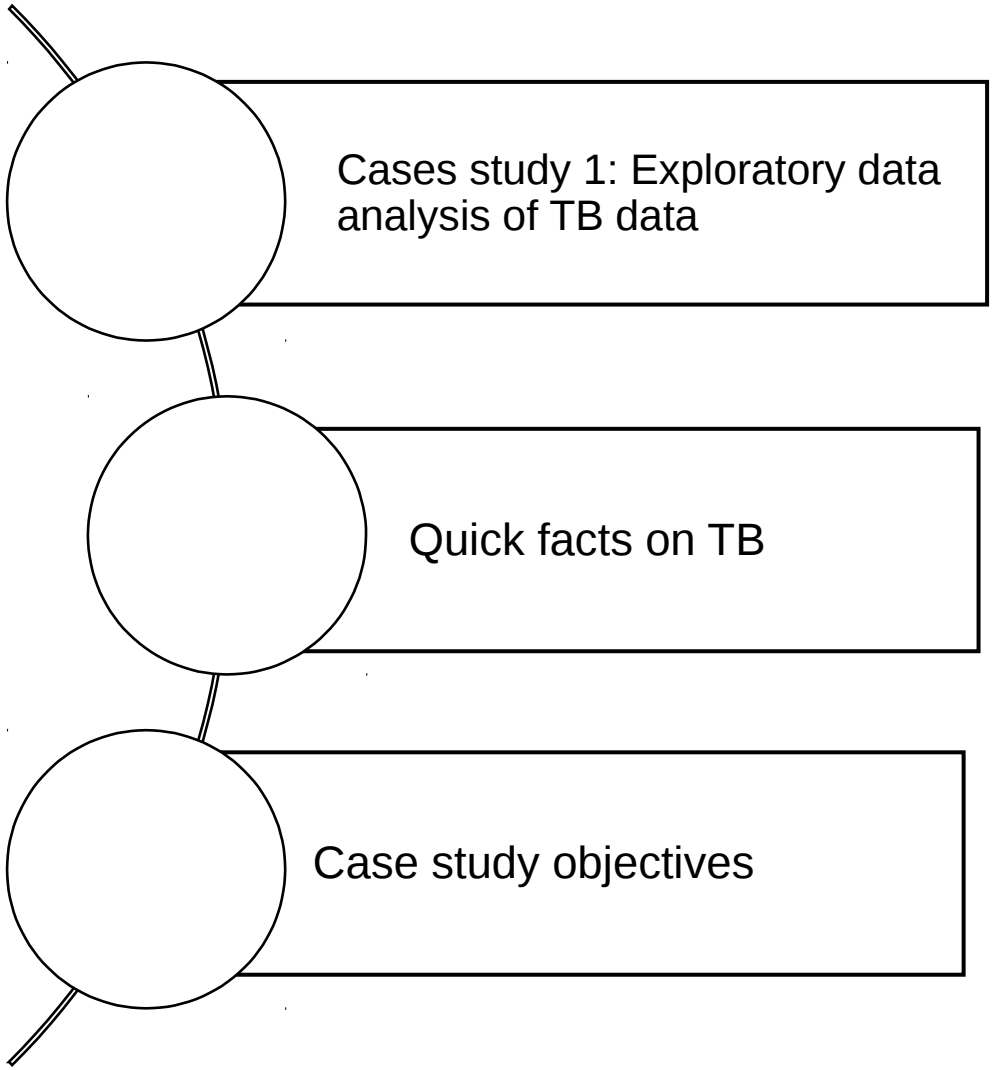
- Health care data is both structured & unstructured

Application of ML solutions

- Disease risk prediction
- Targeted medical treatments

Session 2

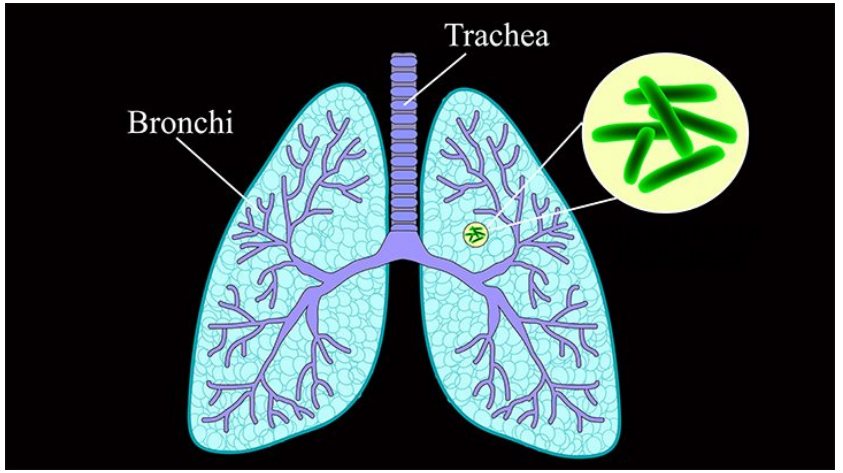
Agenda



Healthcare Analytics

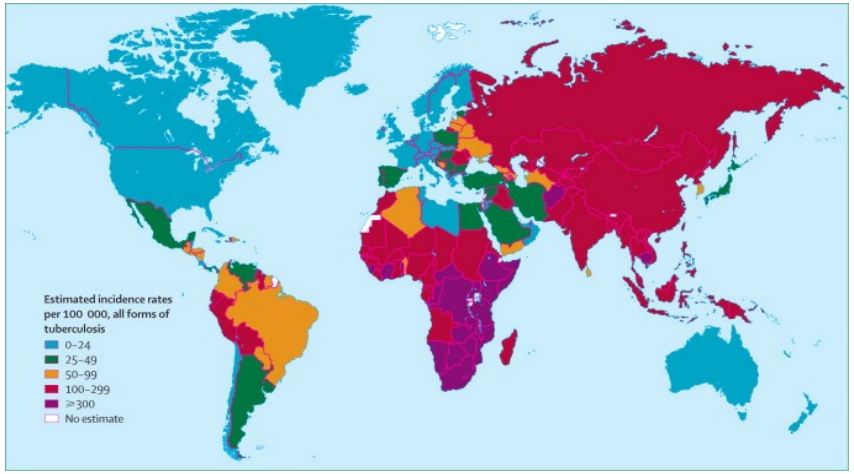
Case study 1

Perform EDA on burden of Tuberculosis globally



Quick facts about Tuberculosis

- Approx 25% of the world's population infected
- 1.4 million TB-related deaths in 2016



Economic cost is rising significantly

MDR-TB treatment could cost up to 25L in Mumbai, says study

Some forms of tuberculosis are so expensive to treat that an average Mumbai family could spend over half its annual income on a single patient.

Malathy Iyer | TNN | October 03, 2016, 06:09 IST

Exploring the TB dataset

The dataset contains **15+** variables containing
Over **5000** observations from years **1990 – 2013**
of **50+** countries

who.int/tb/en/

Objectives of the case study

- Analyse Global trends of TB burden
- Find active geographical hotspots
- Find regions where mortality rates have dropped down
- Analyse how do HIV patients fare when infected with TB

Key findings example summary

- Globally in 2013, almost _____ million people fell ill with TB and _____ million died due to TB and TB/HIV.
- In 2013, the largest number of new TB cases occurred in the _____
- TB is a leading killer of HIV-positive people causing _____one fourth of all HIV-related deaths.
- The estimated number of people falling ill with TB each year (new cases or incidence case rate) is _____
- The TB death rate dropped by _____ between 1990 and 2013
- The TB incidence rate in many countries of the World has _____

Session 3

Case study 2

Regression analysis to predict risk of heart disease

National Heart, Lung, and Blood Institute

framinghamheartstudy.org

What factors are responsible for heart risk?

The death of **US President Roosevelt** while in office initiated the studies to understand Heart diseases better.

The Framingham heart study allowed identifying **Multiple risk factors** which directly contributed for Heart disease using data from cohorts of patients over a couple of years

Following risk factors were identified

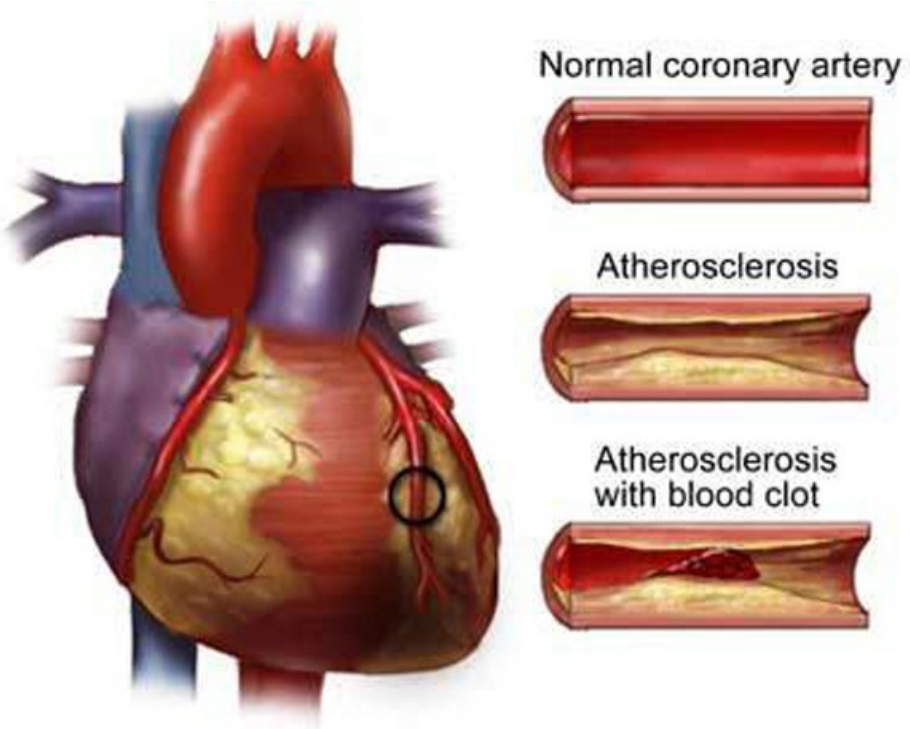
Predictors

- Male Sex
- Age
- Systolic Blood Pressure (SBP)
- Use of Antihypertensive treatment (yes/ no)
- Smoking
- Diabetes mellitus
- Total cholesterol
- HDL cholesterol
- BMI replacing lipids in a simpler model

Objectives of the case study

- Perform EDA of the heart risk data set
- Develop a classification model in Python to train and test the model
- Determine the accuracy of predictions on heart disease
- Develop a treatment plan for high risk patients

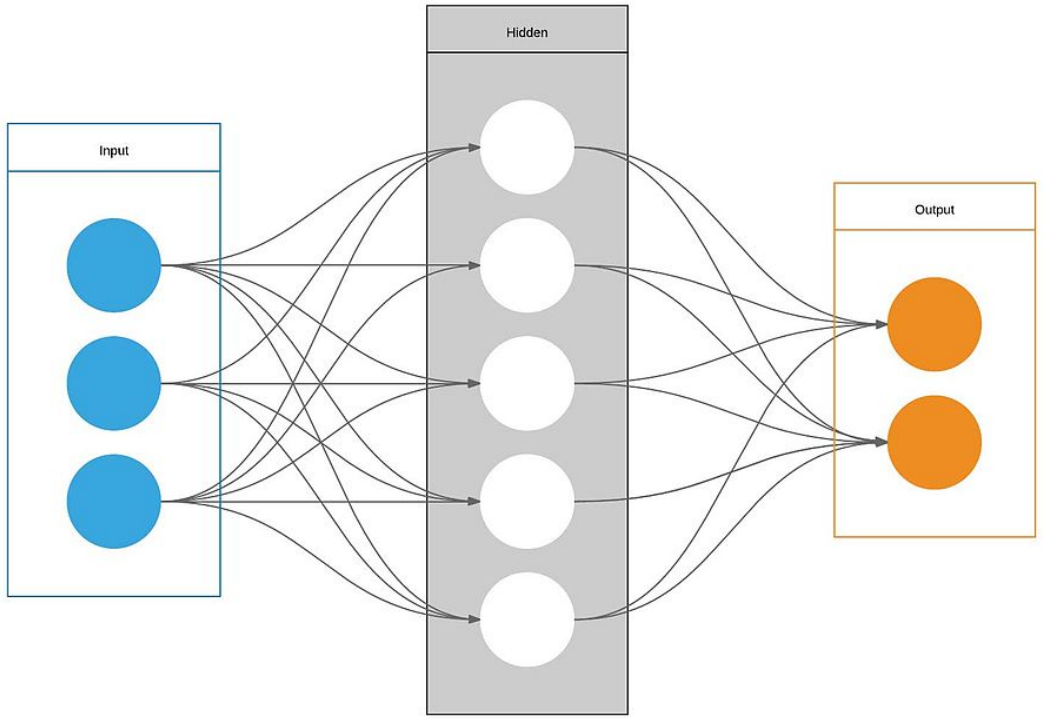
Exploring the Heart disease screening dataset



Contains 16 variables about Heart Risk factors and over 4000 observations

Machine learning model like Logistic regression can be used to predict risk of heart disease

How do we build a predictive model?



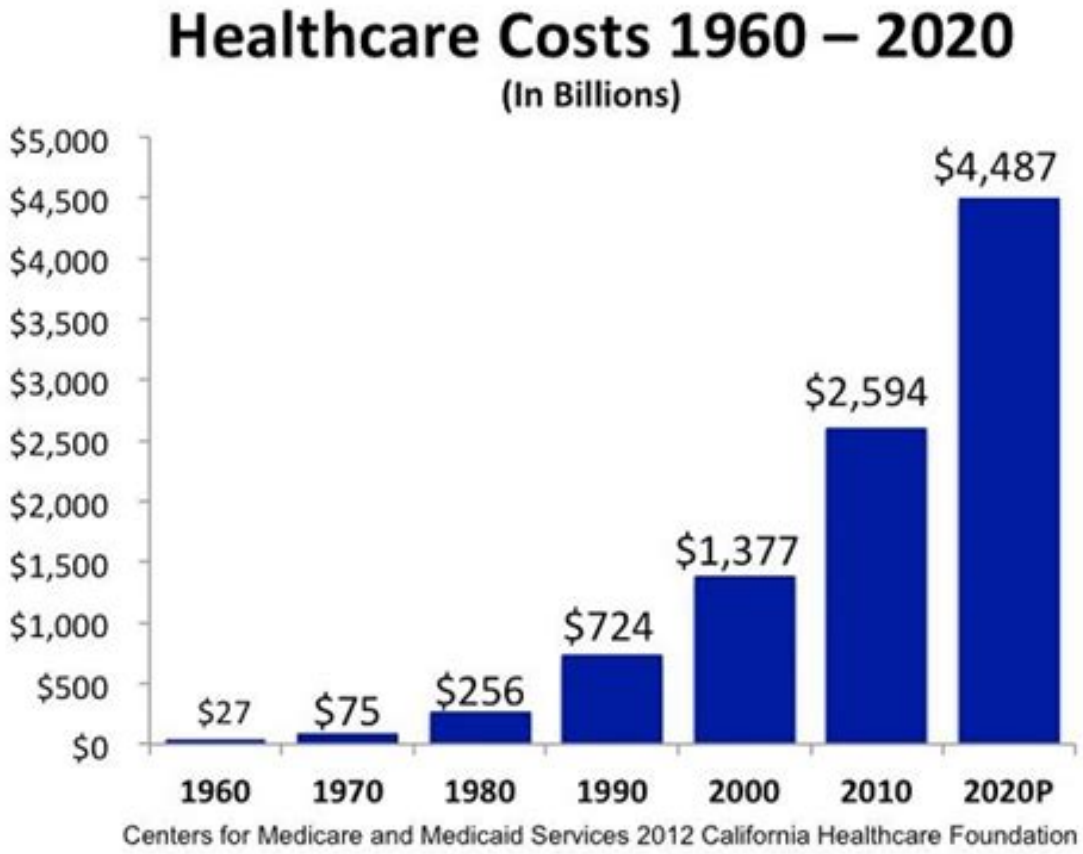
- Randomly split patients data into training and testing sets
- Use logistic regression **on training set** to predict whether or not a patient experienced CHD within 10 years of first examination
- Evaluate predictive power of model on **test set**

Data Input → **Model** → **PREDICTION**

Session 4

Reducing health care costs using Decision Trees

Objectives of the case study



Healthcare costs are a major expense for insurance providers and governments

Lets consult **Scion healthcare mgmt**.
company identify high risk patients
to lower healthcare costs

Objectives of the case study

Goal: improve the quality of cost prediction

- Predicting high-risk patients chances of developing disease
- Identify factors necessary for classification of patients into risk categories
- **Category 5** high risk **Category 1** Low risk, on a scale of 1 to 5
- Allows the insurer to serve the patients much better

Description of claims dataset

Past Claims data

Contains 16 variables and
above 45 thousand observations

Any queries ? Now or any time in future, please write to

arun@upxacademy.com

If it’s a technical issue **please attach a screenshot** of the problem description and inform the exact details

www.upxacademy.com | info@upxacademy.com | 1800-123-1260

Follow us on