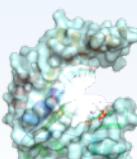


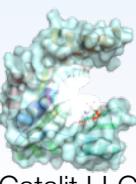


WELCOME



LOGISTICS

- doors
- water
- restrooms
- breaks and lunch
- theory / practice ratio
- computer setup
- wifi
- skill level



Catalit LLC

WHY ARE WE HERE ?



AGENDA

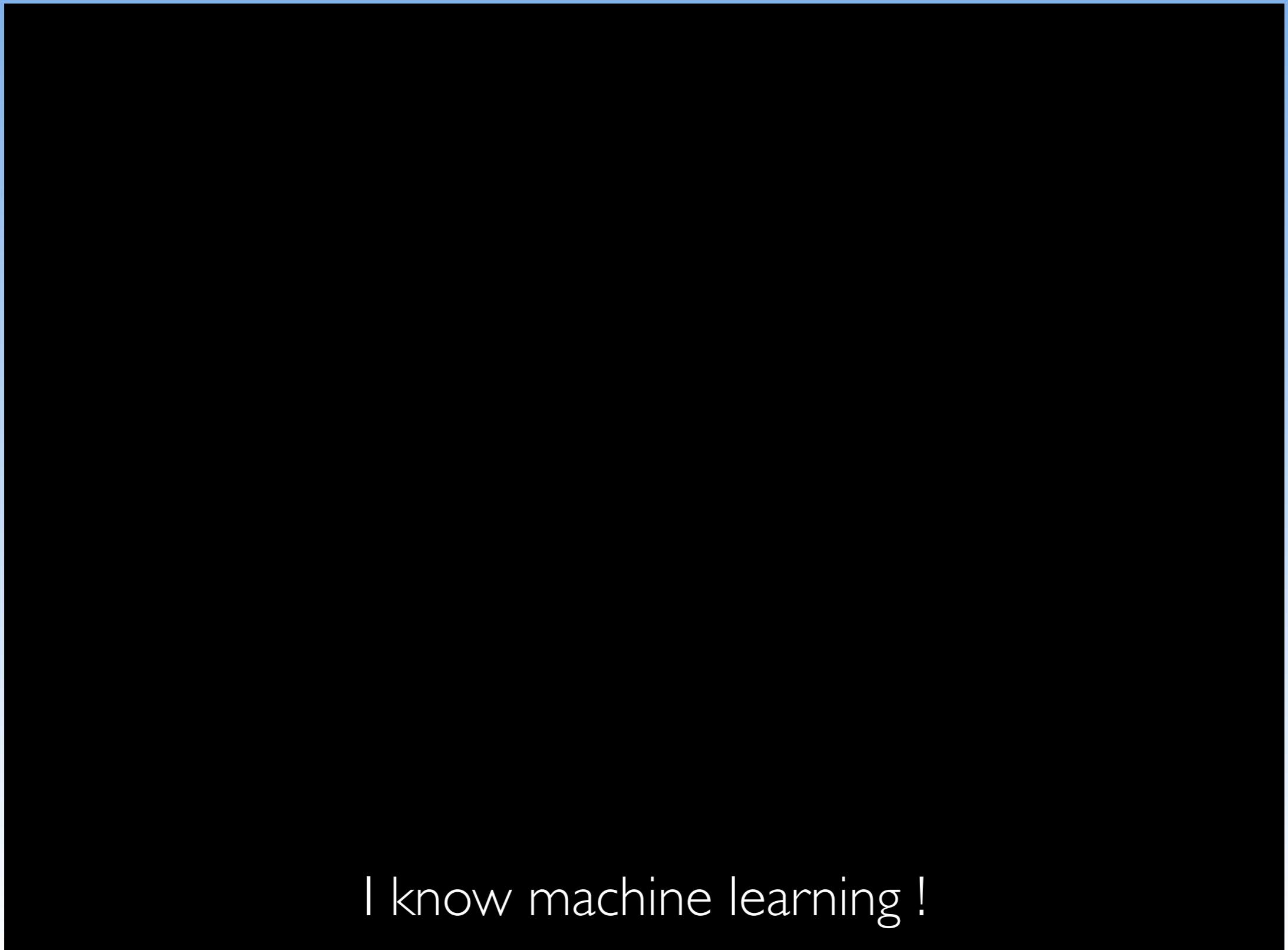
- **Saturday**
 - 09:00 am - 09:30 am Registration
 - **09:30 am - 11:20 am Exploration**
 - 11:20 am - 11:30 am Break
 - **11:30 am - 01:00 pm Regression**
 - 01:00 pm - 02:00 pm Lunch
 - **02:00 pm - 03:50 pm Classification**
 - 03:50 pm - 04:00 pm Break
 - **04:00 pm - 05:15 pm Clustering**
 - 05:15 pm - 05:30 pm Wrap up
- **Sunday**
 -
 - **09:30 am - 11:20 am Exploration/Validation**
 - 11:20 am - 11:30 am Break
 - **11:30 am - 01:00 pm Improve accuracy**
 - 01:00 pm - 02:00 pm Lunch
 - **02:00 pm - 03:50 pm Final Project**
 - 03:50 pm - 04:00 pm Break
 - **04:00 pm - 05:15 pm Final Project**
 - 05:15 pm - 05:30 pm Wrap up

GOALS

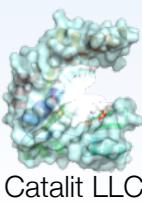
- Learn some cool ML
- Meet fellow students
- Project of your own
- Have fun



WHAT TO EXPECT



I know machine learning !



Catalit LLC

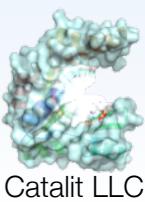
WHAT TO EXPECT

- Accelerated learning
- Interactive
- Fun / Discovery



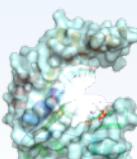
Catalit LLC

MACHINE LEARNING

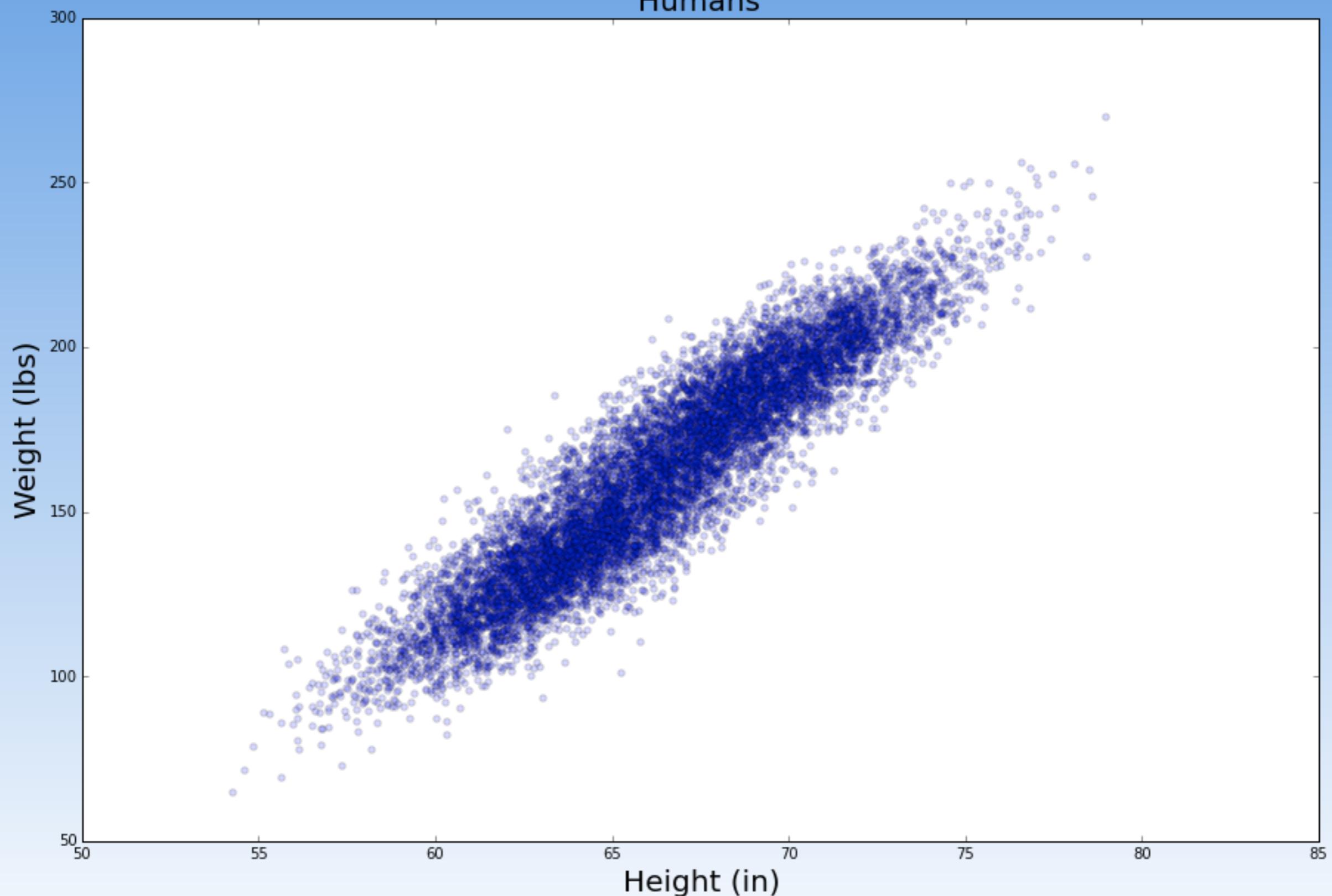


Catalit LLC

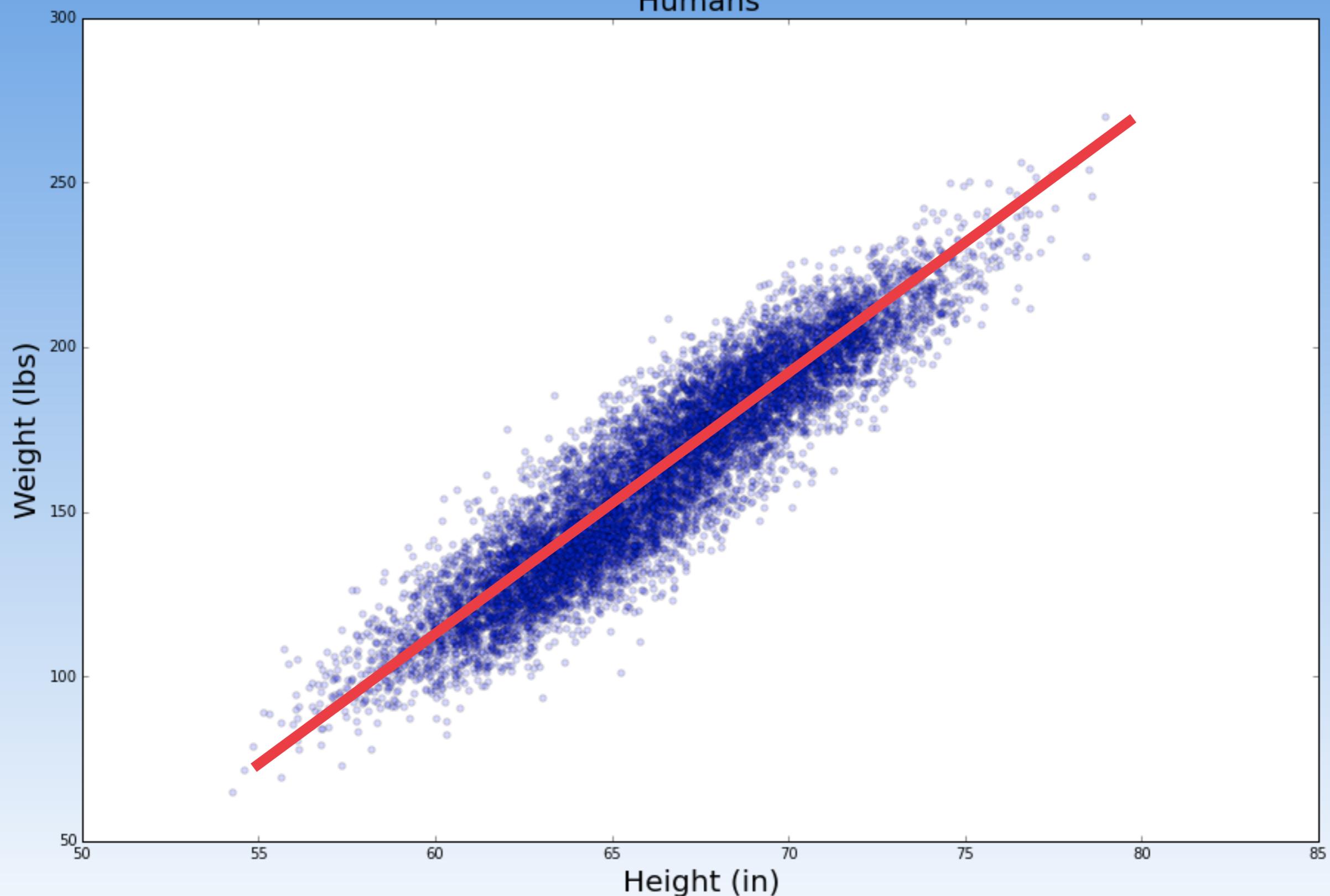
YOU ARE ALREADY
ABLE TO PERFORM MACHINE
LEARNING WITH YOUR BRAIN....



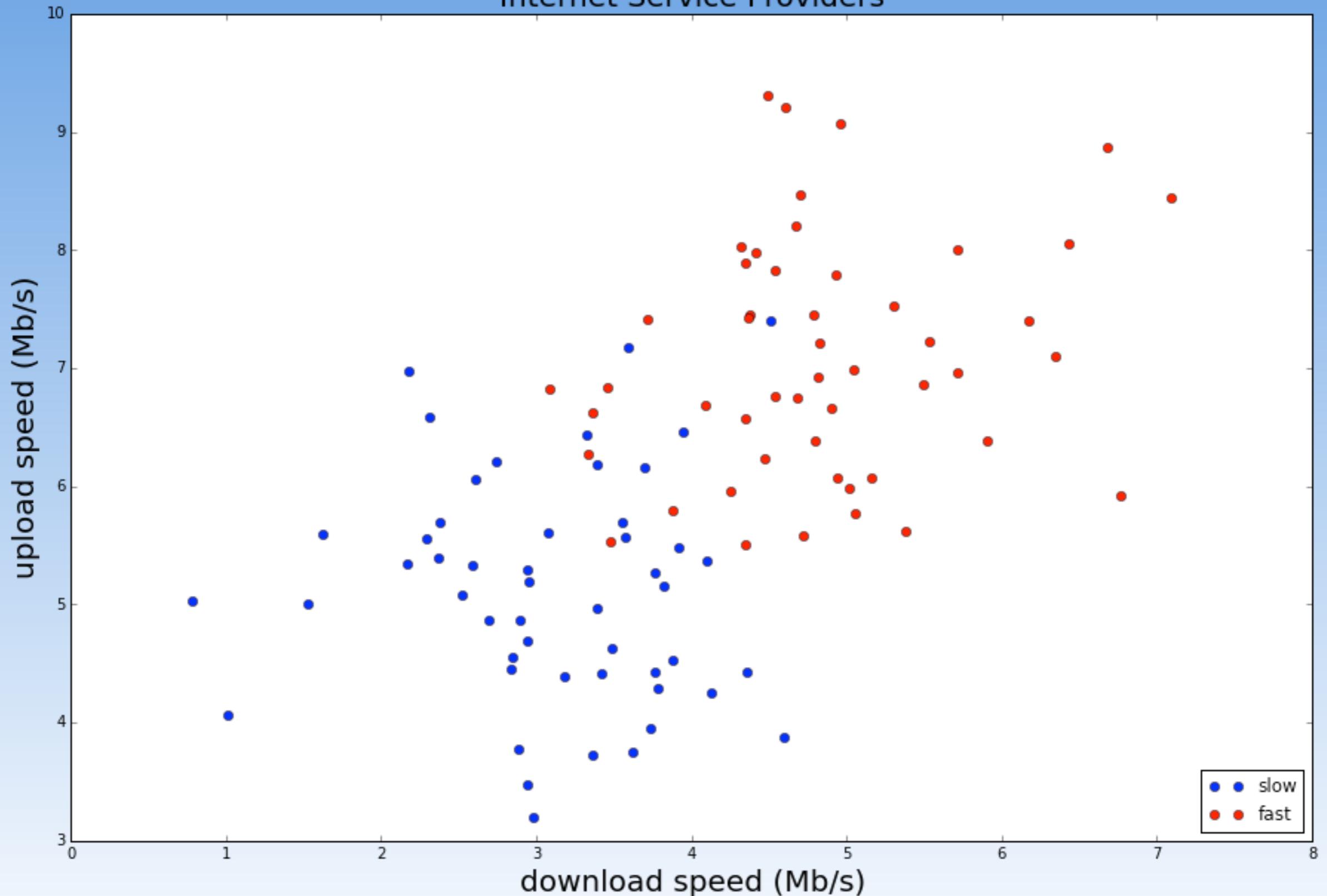
Humans



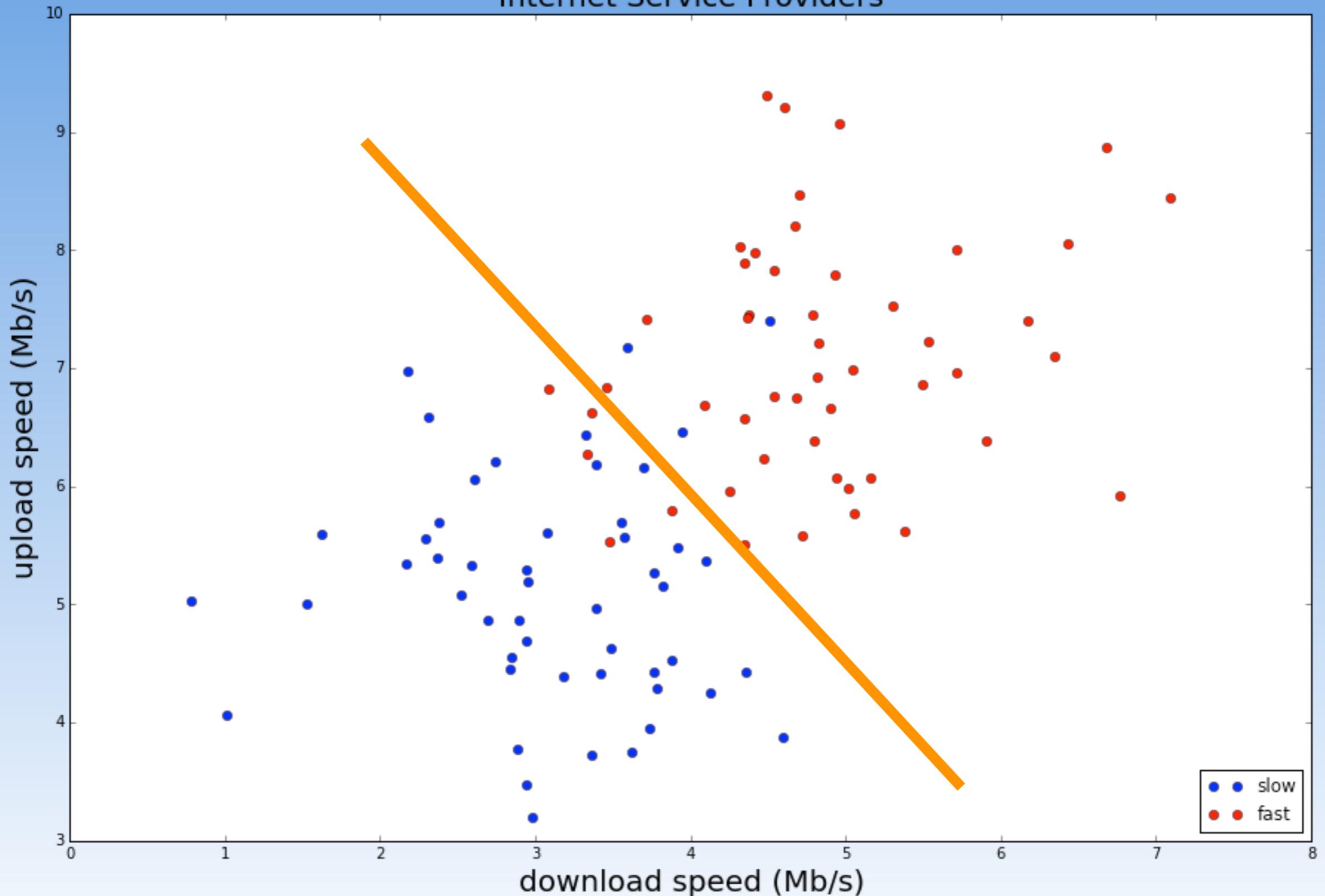
Humans



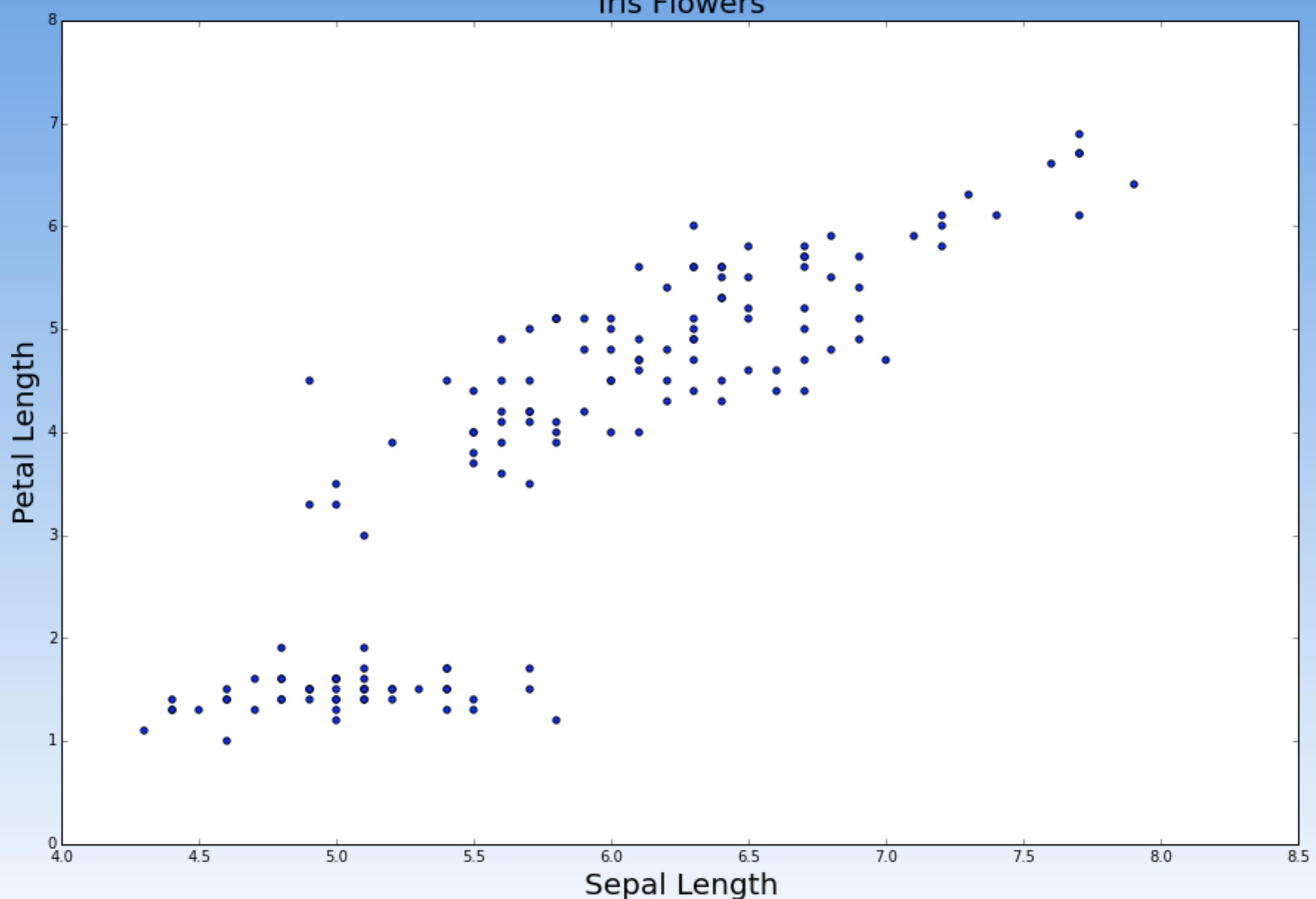
Internet Service Providers



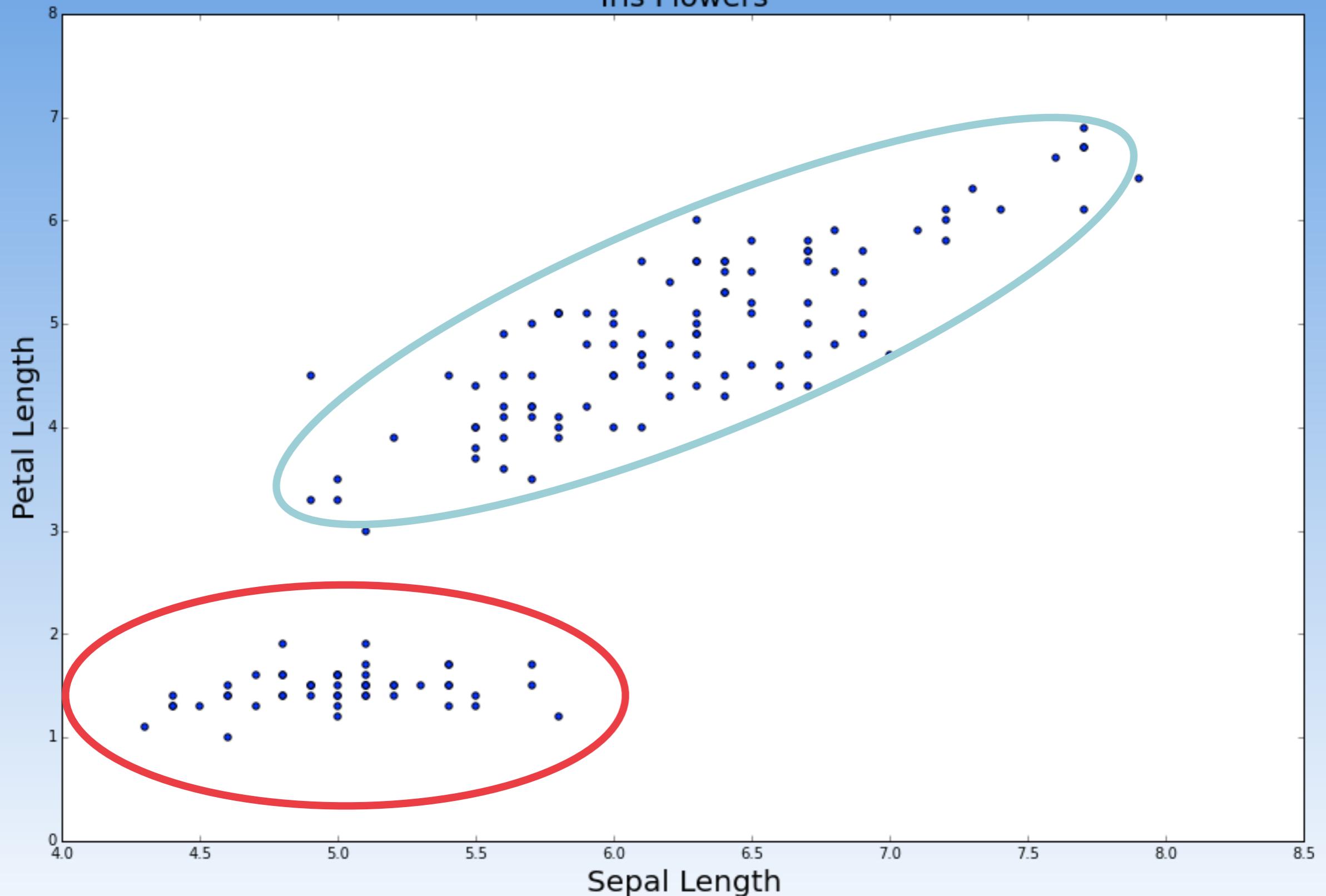
Internet Service Providers



Iris Flowers



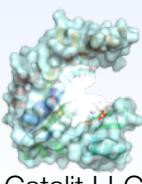
Iris Flowers



REGRESSION

CLASSIFICATION

CLUSTERING

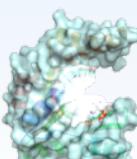


MACHINE LEARNING

"A field of study that gives computers the ability to learn without being explicitly programmed." (1959)



Arthur Samuel, AI pioneer
Source: Stanford

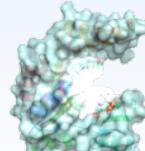
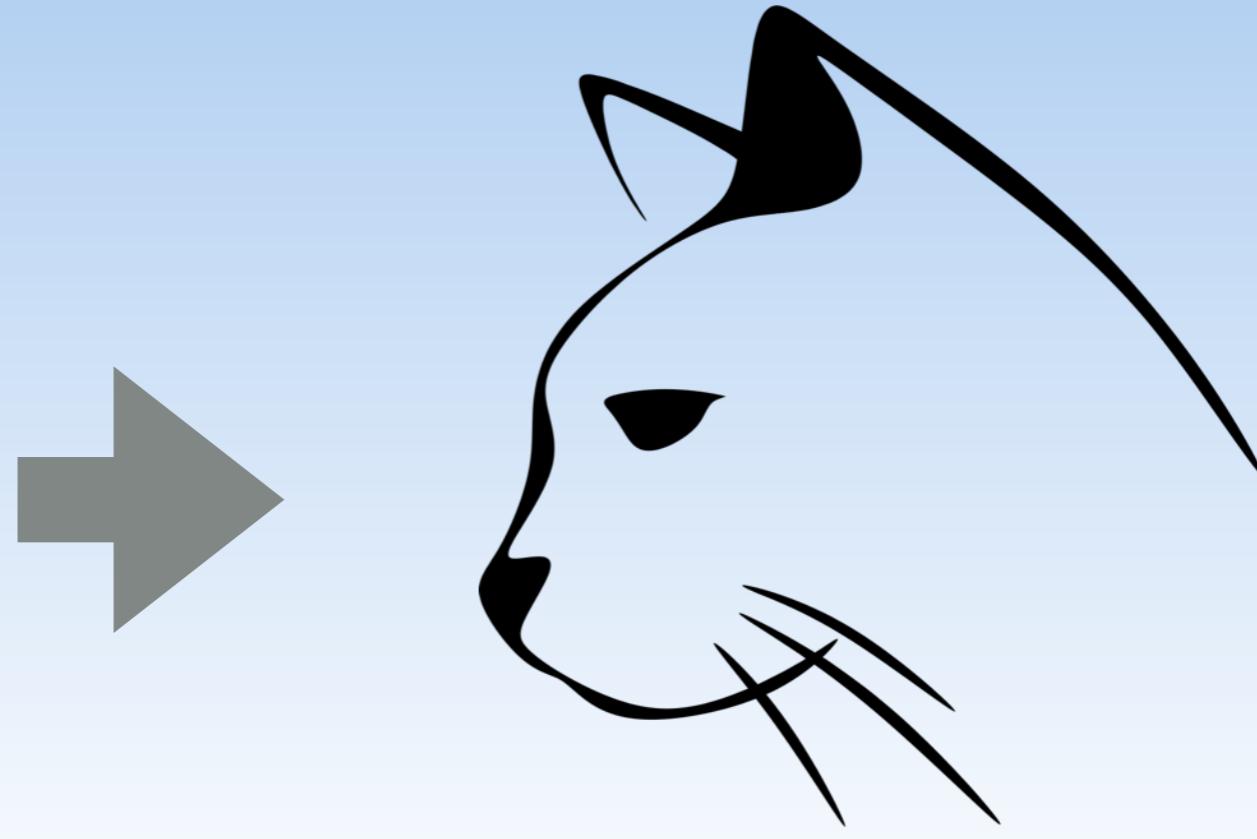


Catalit LLC

LEARNING



Learning is not memorizing. It is
generalizing the conclusions
to previously unseen examples



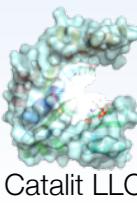
Catalit LLC

SUPERVISED LEARNING



©2014 Creative Safety Supply

<http://www.realsafety.org/wp-content/uploads/2014/11/safety-supervisors-interaction.png>



Catalit LLC

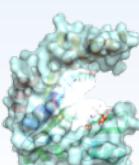
SUPERVISED LEARNING

	Age	Gender	Annual Salary	Months in residence	Months in job	Current Debt	Paid off credit
	23	M	\$30,000	36	12	\$5,000	Yes
Client 2	30	F	\$45,000	12	12	\$1,000	Yes
Client 3	19	M	\$15,000	3	1	\$10,000	No
Client 4	25	M	\$25,000	12	27	\$15,000	?

SUPERVISED LEARNING

	Age	Gender	Annual Salary	Months in residence	Months in job	Current Debt	Paid off credit
Client 1	23	M	\$30,000	36	12	\$5,000	Yes
Client 2	30	F	\$45,000	12	12	\$1,000	Yes
Client 3	19	M	\$15,000	3	1	\$10,000	No
Client 4	25	M	\$25,000	12	27	\$15,000	?

Record OR Data Point



Catalit LLC

SUPERVISED LEARNING

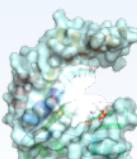
	Age	Gender	Annual Salary	Months in residence	Months in job	Current Debt	Paid off credit
Client 1	23	M	\$30,000	36	12	\$5,000	Yes
Client 2	30	F	\$45,000	12	12	\$1,000	Yes
Client 3	19	M	\$15,000	3	1	\$10,000	No
Client 4	25	M	\$25,000	12	27	\$15,000	?

Features

SUPERVISED LEARNING

	Age	Gender	Annual Salary	Months in residence	Months in job	Current Debt	Paid off credit
	23	M	\$30,000	36	12	\$5,000	Yes
Client 2	30	F	\$45,000	12	12	\$1,000	Yes
Client 3	19	M	\$15,000	3	1	\$10,000	No
Client 4	25	M	\$25,000	12	27	\$15,000	?

Labels



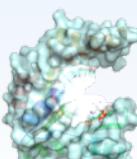
Catalit LLC

SUPERVISED LEARNING

	Age	Gender	Annual Salary	Months in residence	Months in job	Current Debt	Paid off credit
	23	M	\$30,000	36	12	\$5,000	Yes
Client 2	30	F	\$45,000	12	12	\$1,000	Yes
Client 3	19	M	\$15,000	3	1	\$10,000	No
Client 4	25	M	\$25,000	12	27	\$15,000	?

Goal: generalize to new data

SUPERVISED LEARNING



Catalit LLC

UNSUPERVISED LEARNING

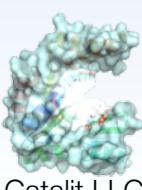


UNSUPERVISED LEARNING

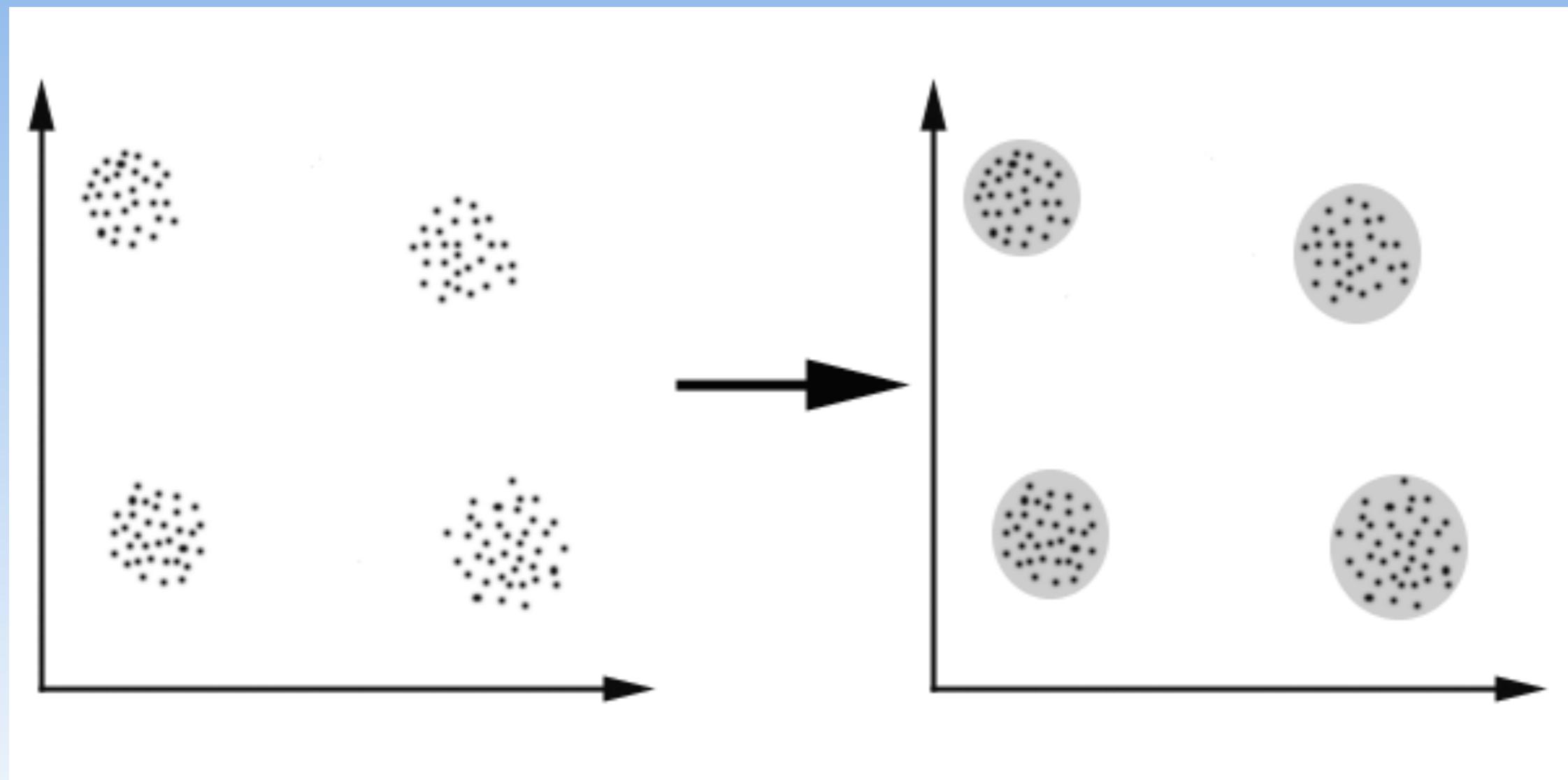
no Labels

	Age	Gender	Annual Salary	Months in residence	Months in job	Current Debt	
Client 1	23	M	\$30,000	36	12	\$5,000	!!!
Client 2	30	F	\$45,000	12	12	\$1,000	!!!
Client 3	19	M	\$15,000	3	1	\$10,000	!!!
Client 4	25	M	\$25,000	12	27	\$15,000	!!!

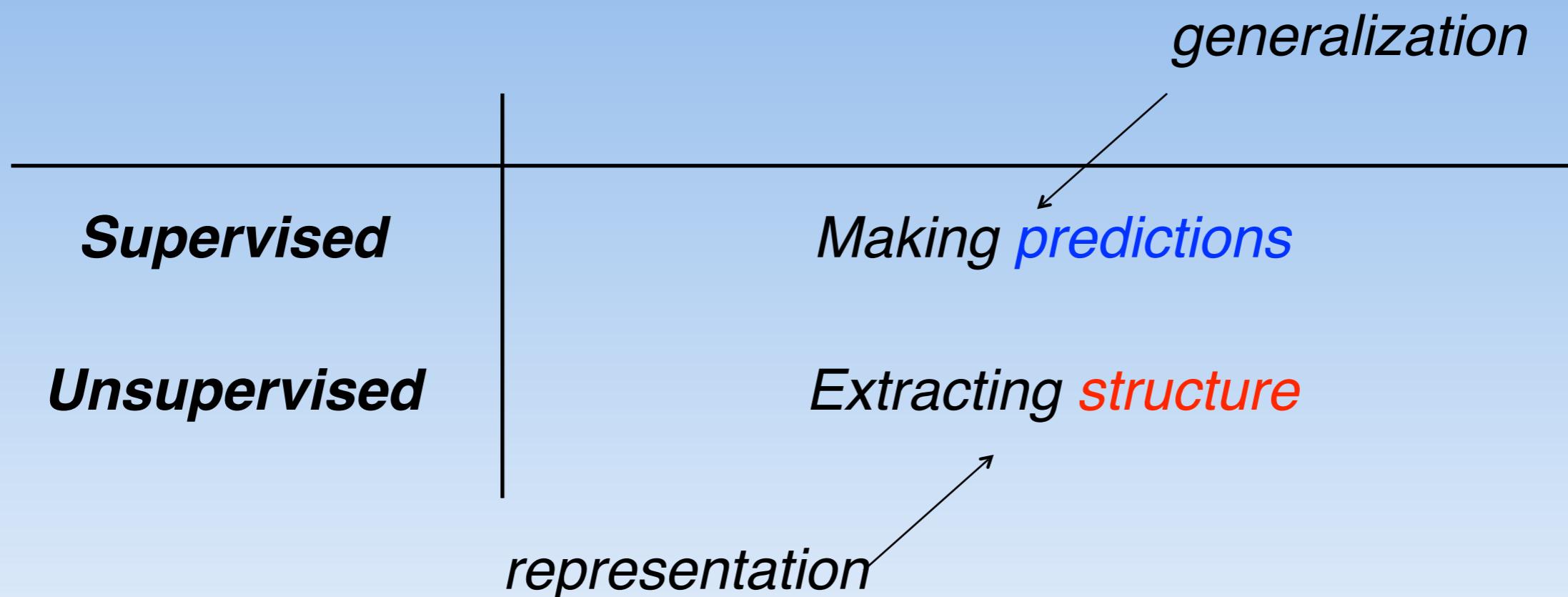
Goal: find patterns



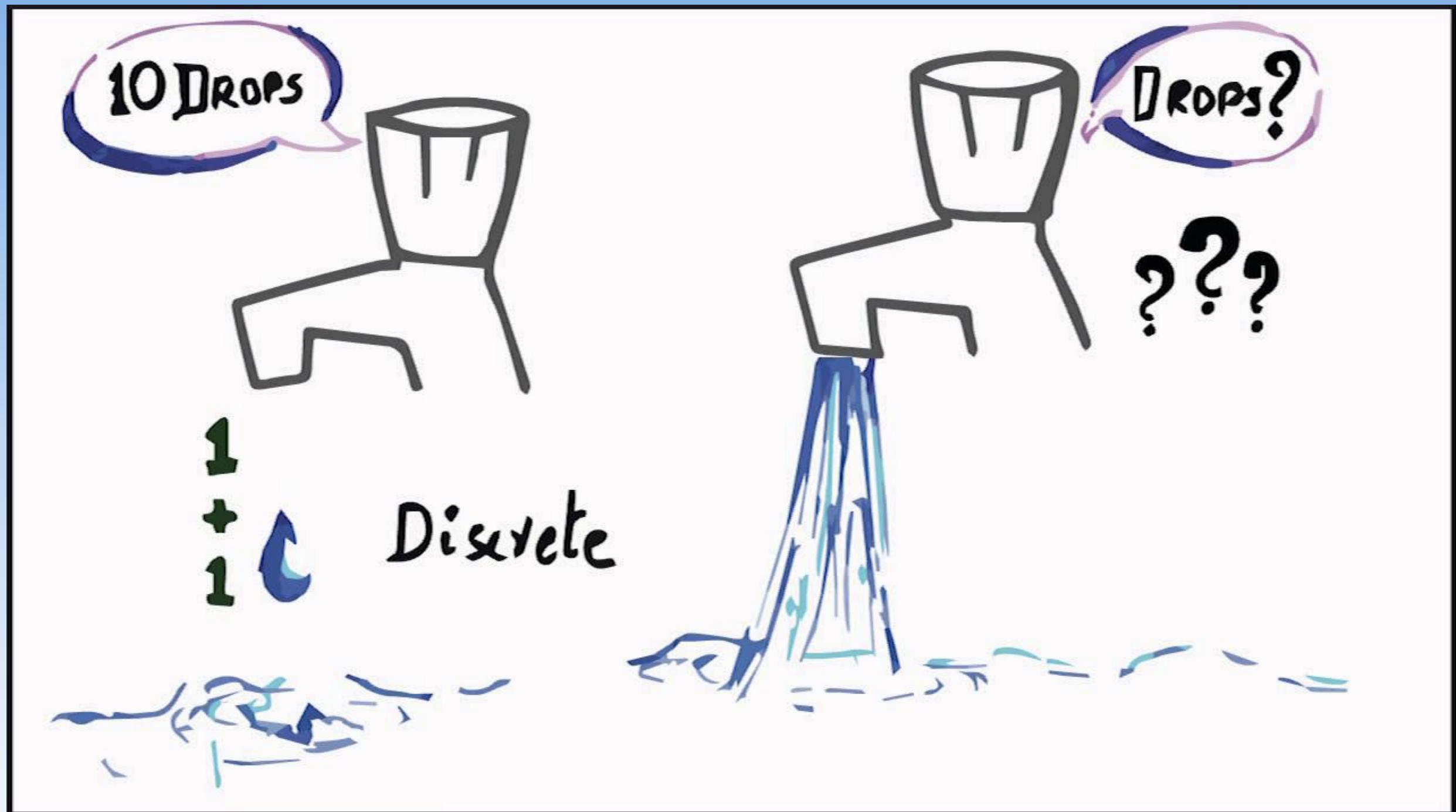
UNSUPERVISED LEARNING



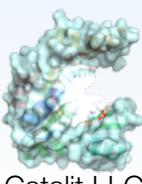
MACHINE LEARNING



PREDICTION TYPE



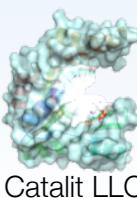
PREDICTION TYPE



Catalit LLC

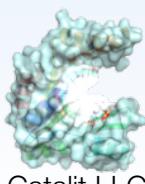
PREDICTION TYPE

CONTINUOUS	CATEGORICAL
Height of children	Eye colors
Weight of cars	Courses at university
Speed of the train	Highest degree
Temperature	Gender
Stock price	Is email spam or not



COMBINED

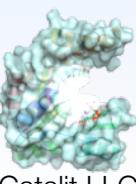
	CONTINUOUS	CATEGORICAL
SUPERVISED	REGRESSION	CLASSIFICATION
UNSUPERVISED	DIMENSION REDUCTION	CLUSTERING



COMBINED

	CONTINUOUS	CATEGORICAL
SUPERVISED	REGRESSION	CLASSIFICATION
UNSUPERVISED	DIMENSION REDUCTION	CLUSTERING

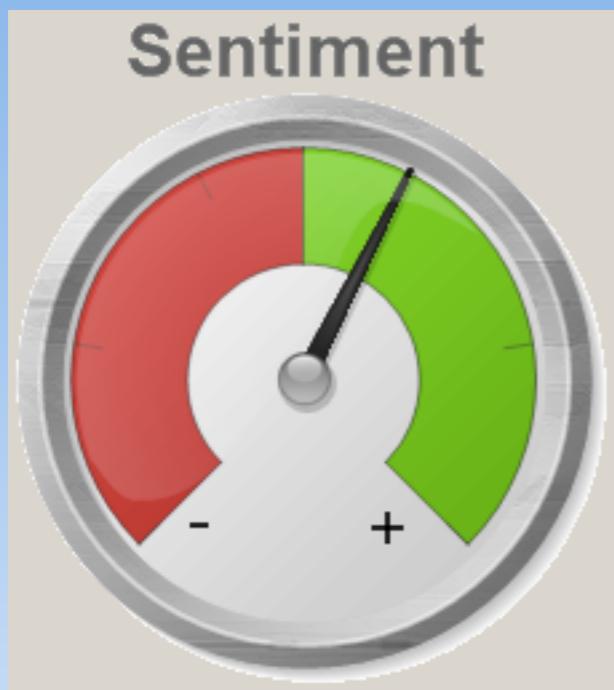
APPLICATIONS



Catalit LLC

APPLICATIONS

Sentiment Analysis



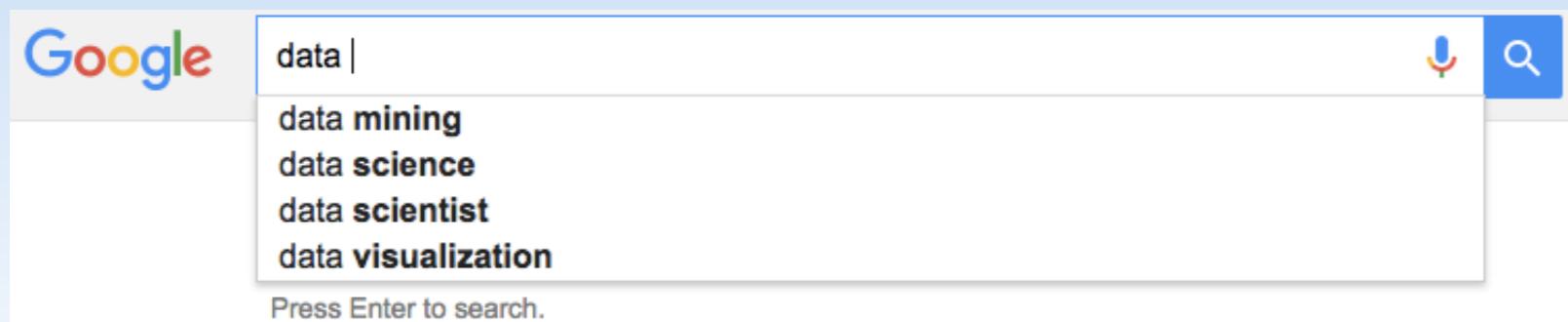
Genome prediction



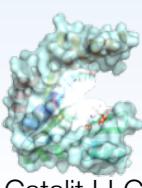
Heart Monitoring



Keyword prediction



A screenshot of a Google search bar. The user has typed "data |" into the search field. Below the search field, a dropdown menu lists suggestions: "data mining", "data science", "data scientist", and "data visualization". There is also a link to "Press Enter to search." and a microphone icon for voice search.



Catalit LLC

APPLICATIONS

Book recommendation

Recommendations for You in Books

Cracking the Coding Interview: 150...
➤ Gayle Laakmann McDowell
Paperback
★★★★★ (166)
\$39.95 \$23.22
Why recommended?

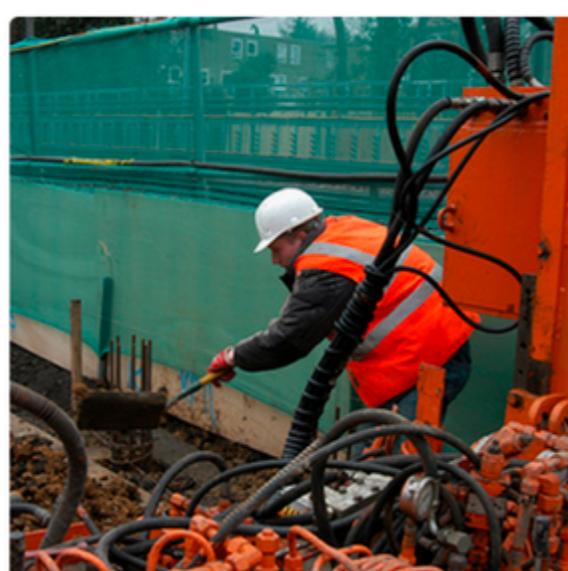
Introduction to Algorithms
Thomas H. Cormen, Charles E.
Hardcover
★★★★★ (85)
\$92.00 \$80.00
Why recommended?

Data Mining: Practical Machine...
➤ Ian H. Witten, Eibe Frank,
Mark A. Hall
Paperback
★★★★★ (27)
\$69.95 \$42.09
Why recommended?

Caption generation



"man in black shirt is playing guitar."



"construction worker in orange safety vest is working on road."

Movie recommendation

NETFLIX

Watch Instantly | Browse DVDs | Your Queue | Movies You'll ❤️

Congratulations! Movies we think You will ❤️

Add movies to your Queue, or Rate ones you've seen for even better suggestions.

Spider-Man 3	300	The Rundown	Bad Boys II
Add	Add	Add	Add
★★★★★ (Not Interested)	★★★★★ (Not Interested)	★★★★★ (Not Interested)	★★★★★ (Not Interested)

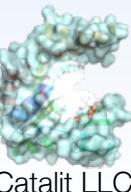
Las Vegas: Season 2 (6-Disc Series)	The Last Samurai	Star Wars: Episode III	Robot Chicken: Season 3 (2-Disc Series)
Add	Add	Add	Add
★★★★★ (Not Interested)	★★★★★ (Not Interested)	★★★★★ (Not Interested)	★★★★★ (Not Interested)

Human recognition

I'm not a robot

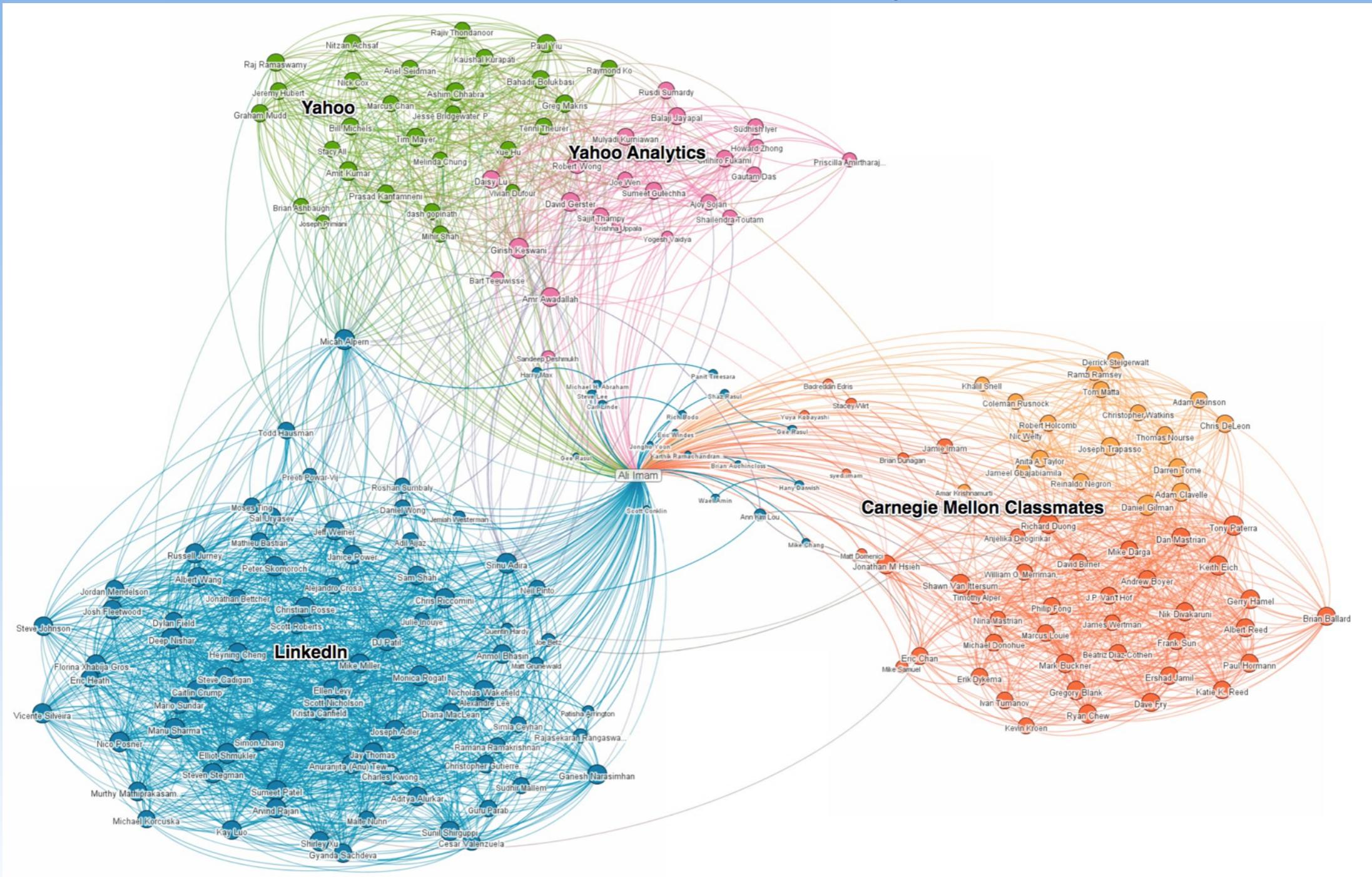
reCAPTCHA

[Privacy - Terms](#)



APPLICATIONS

Social Network Analysis

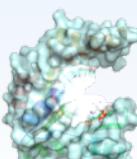


APPLICATIONS

House price prediction



Document classification



Catalit LLC



Search Facebook



Francesco

Home



Francesco Mosconi

[Edit Profile](#)

FAVORITES

[News Feed](#)[Messages](#)[Events](#)[Saved](#)[Update Status](#)[Add Photos/Video](#)[Create Photo Album](#)

What's on your mind?

[Friends](#)[Post](#)**Boris Marcone** shared Daniele Caluri's post.

12 hrs ·

**Paola Cinquina** and 2 others

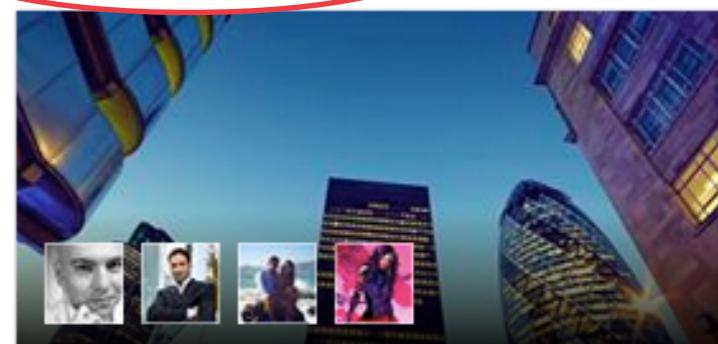
PEOPLE YOU MAY KNOW

**Maciej Szczepanek**

2 mutual friends

[Add Friend](#)[See All](#)

SUGGESTED GROUPS

**The Fintech Society**

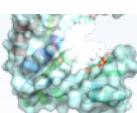
4 friends · 474 members

[+ Join](#)**Shape Europe**

2 friends · 416 members

[+ Join](#)[See All](#)

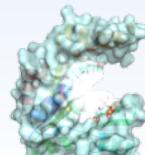
SPONSORED

[Create Ad](#)

Catalit LLC

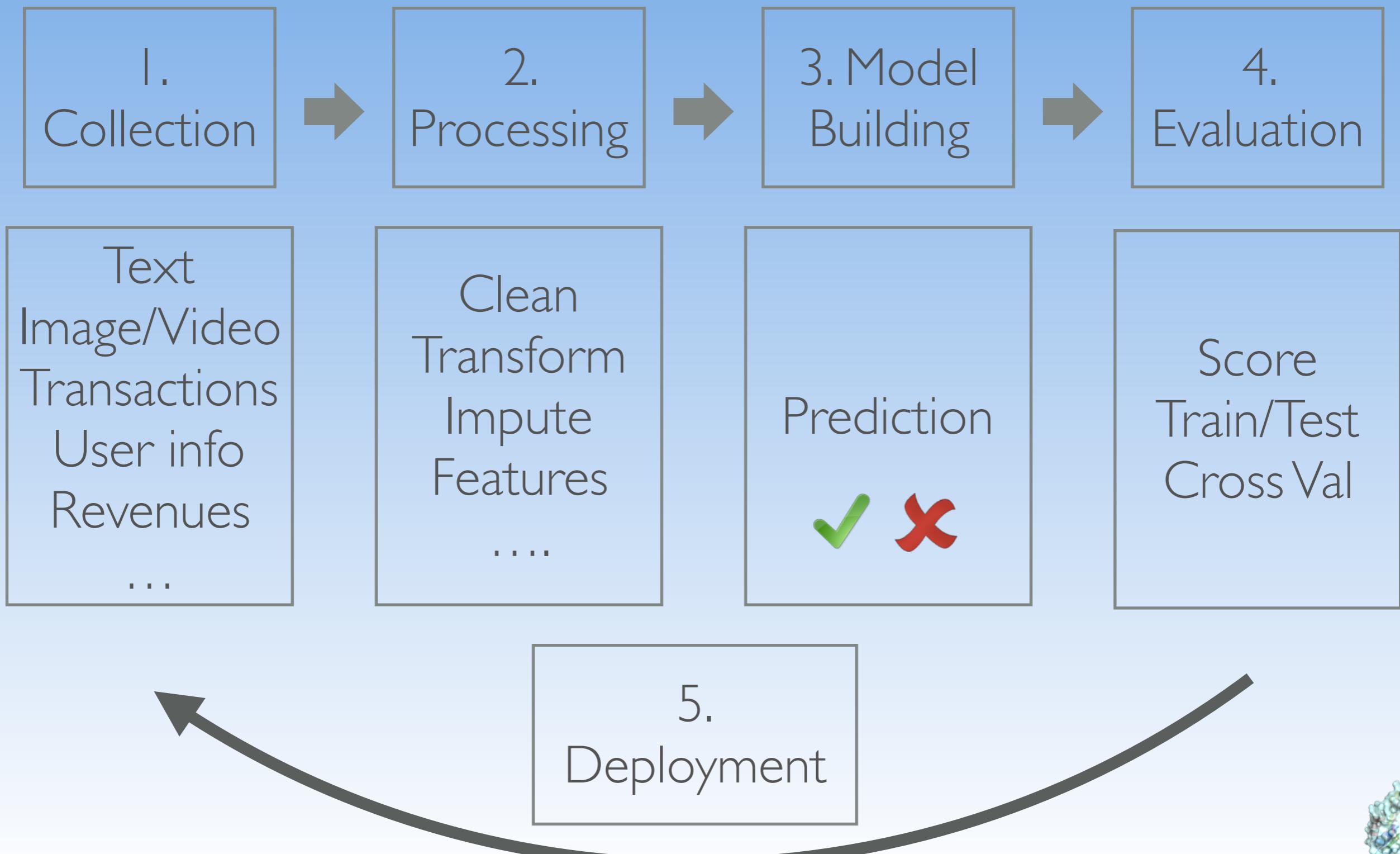
APPLICATIONS

Stock market prediction



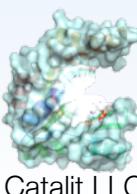
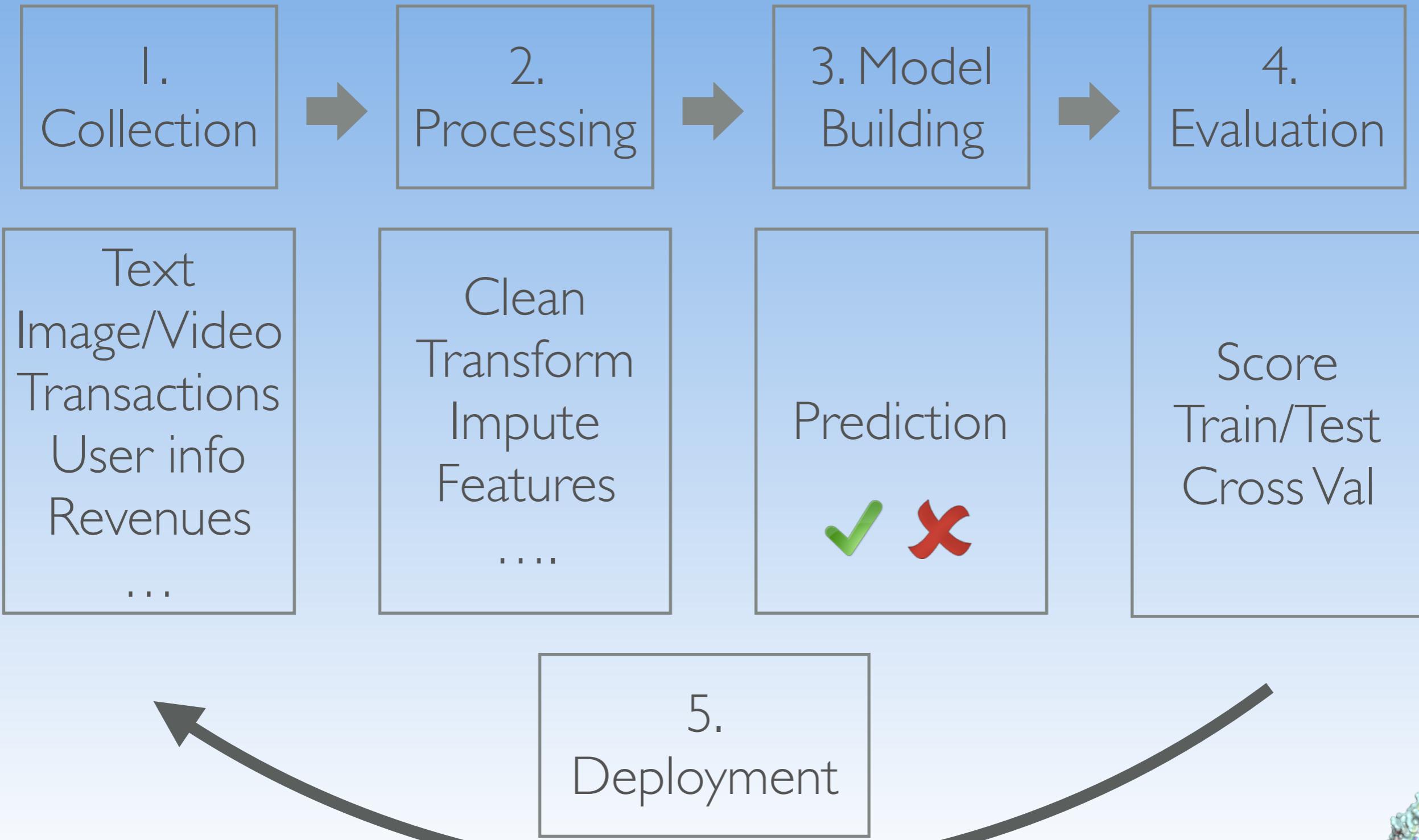
Catalit LLC

ML STEPS

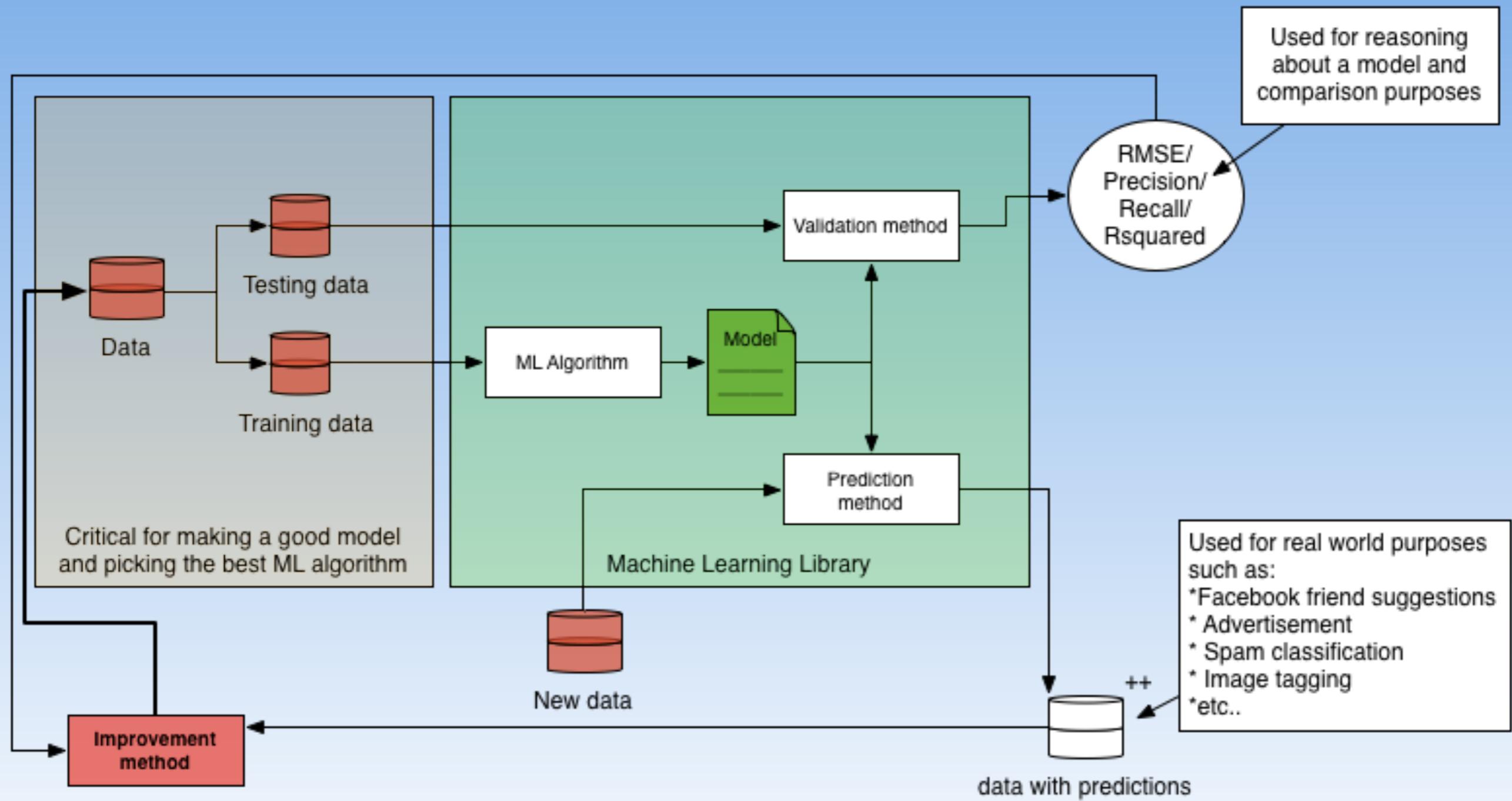


Exploration

ML STEPS

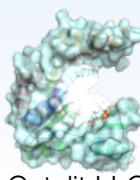


ML PIPELINE



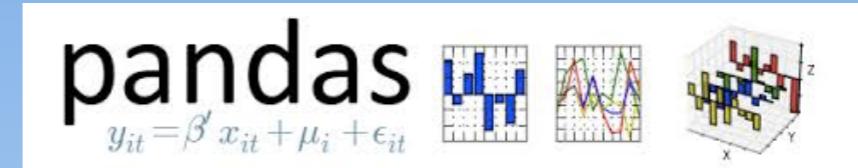
DATA SCIENCE WITH PYTHON

OPEN SOURCE



Catalit LLC

PYDATA ECHOSYSTEM



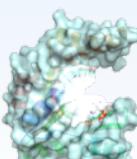
Bokeh



Catalit LLC

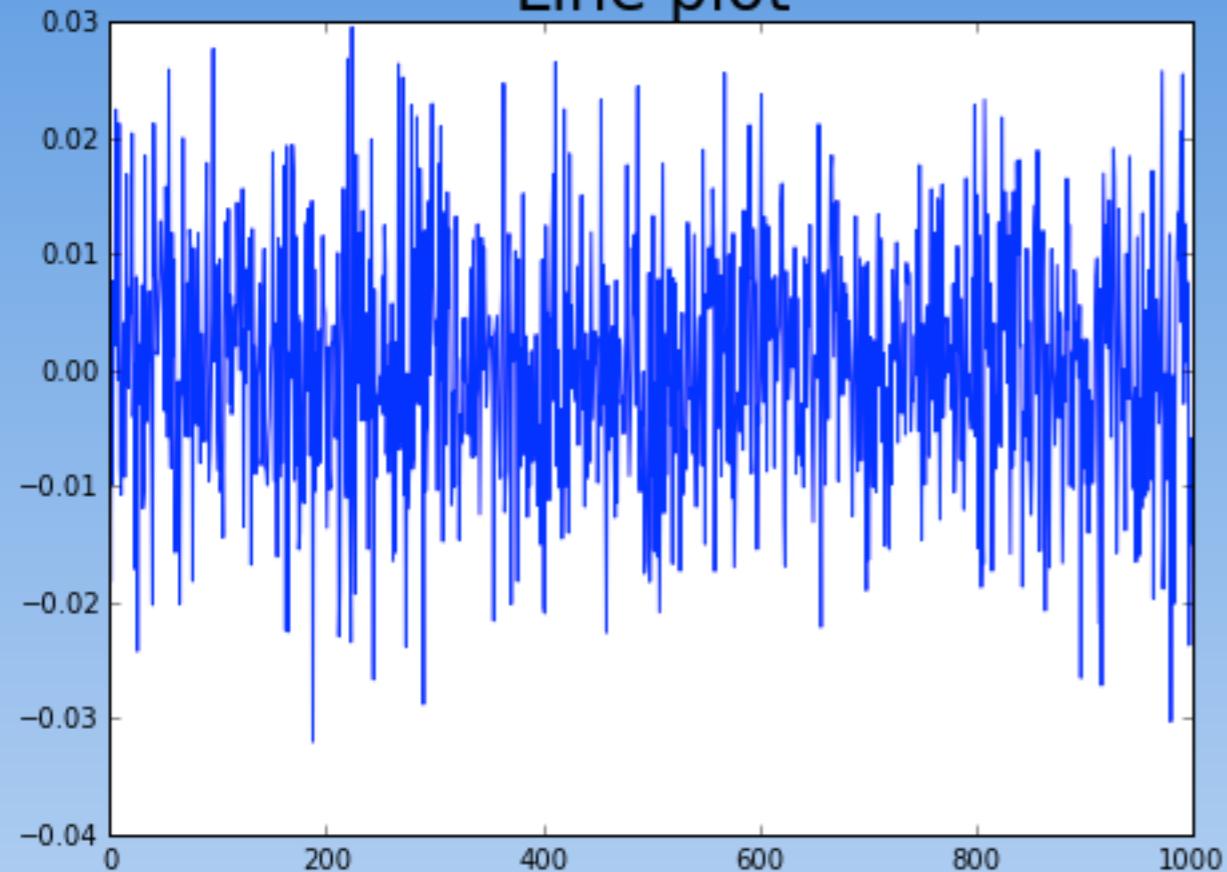
DATA VISUALIZATION

EXPLORATION

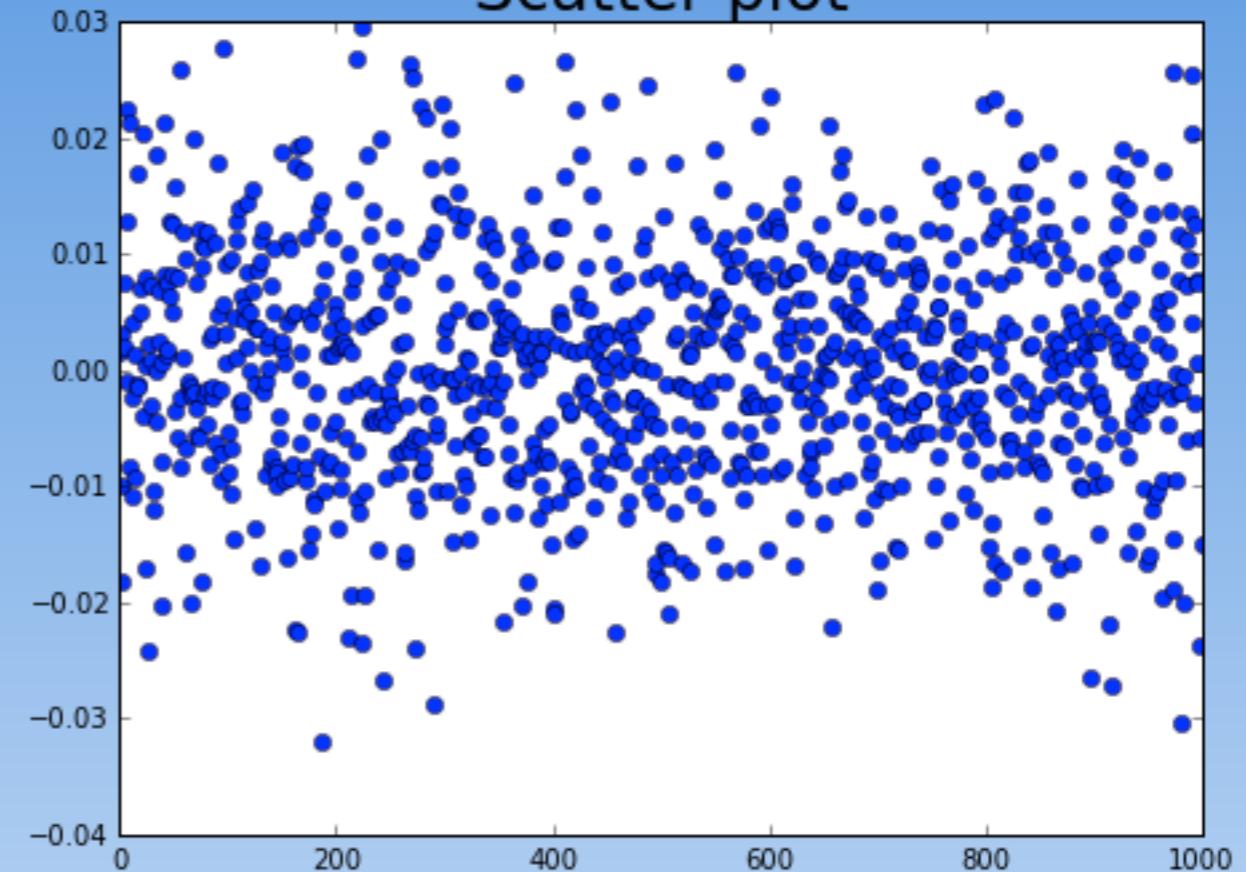


Catalit LLC

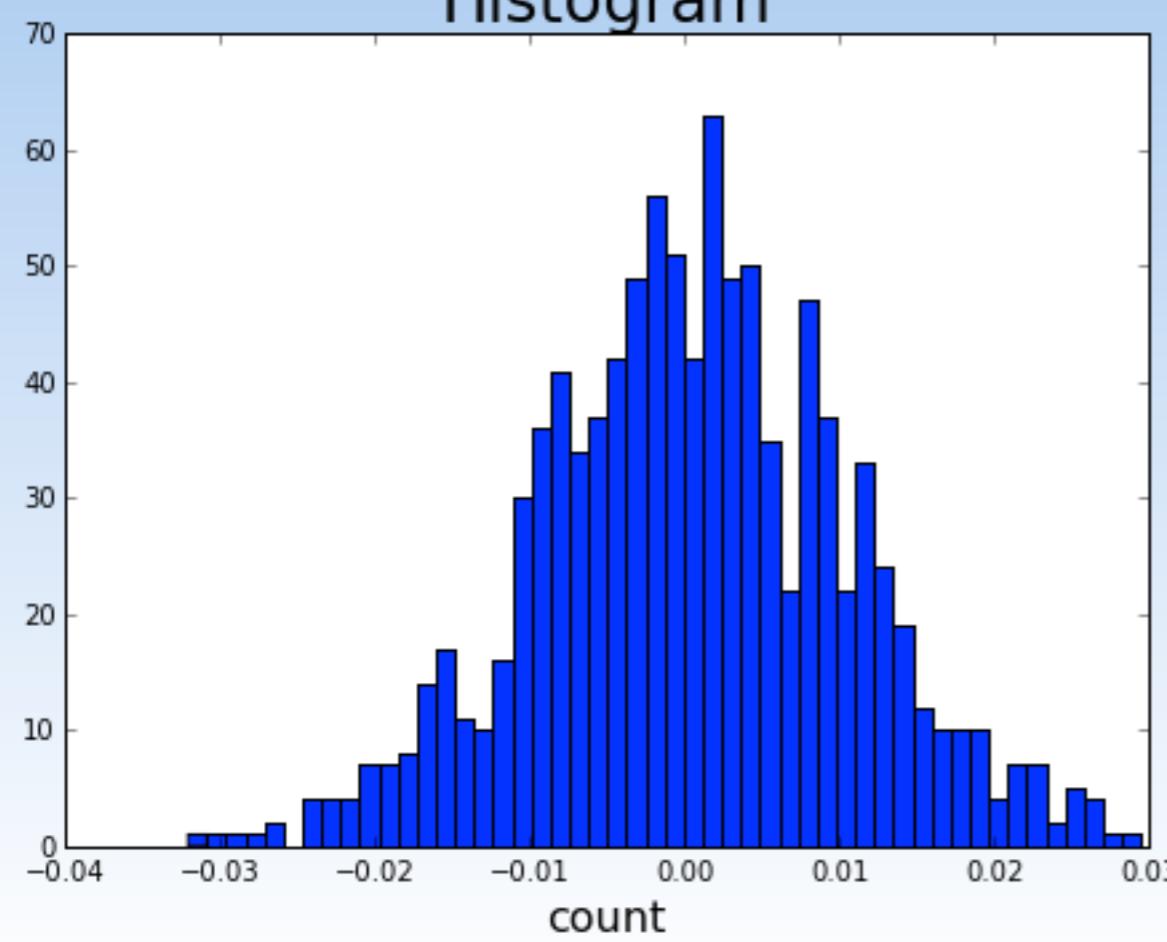
Line plot



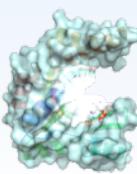
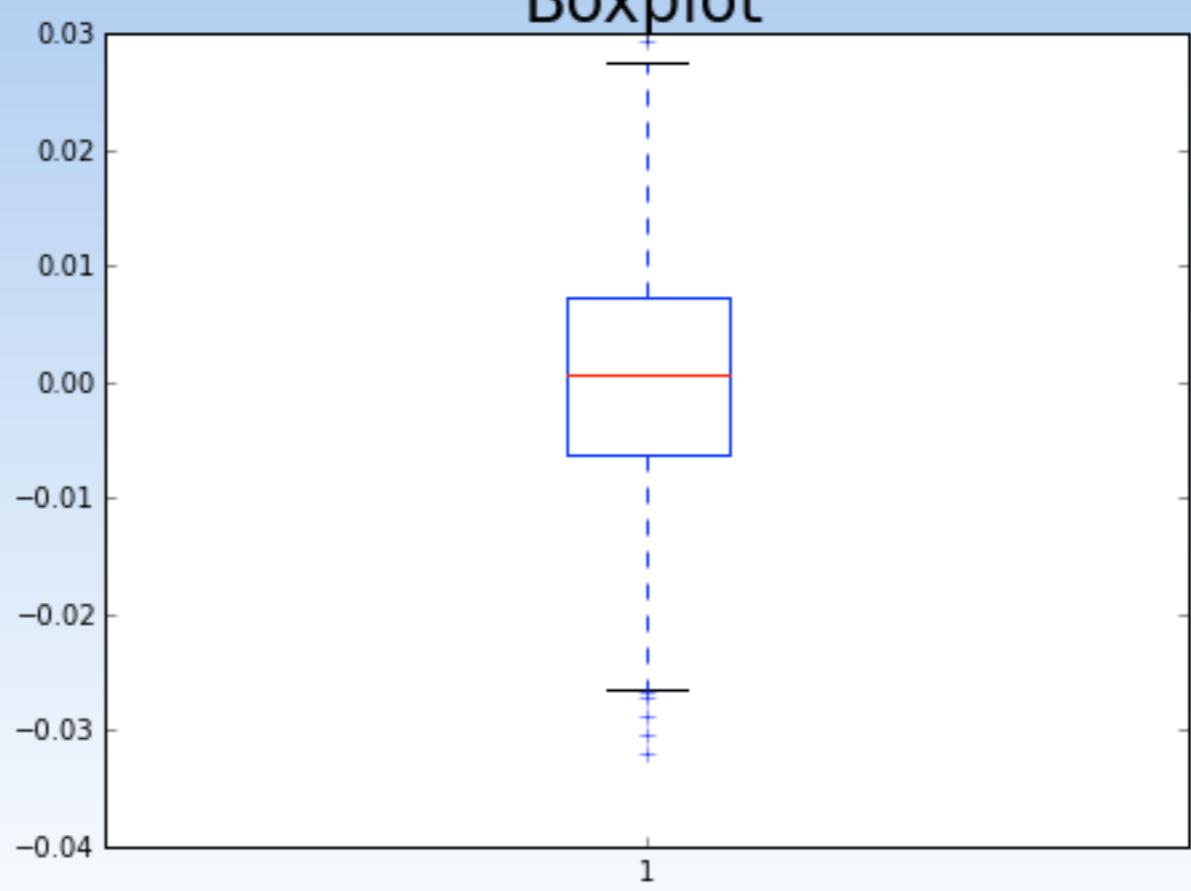
Scatter plot



Histogram

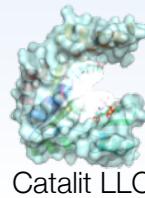
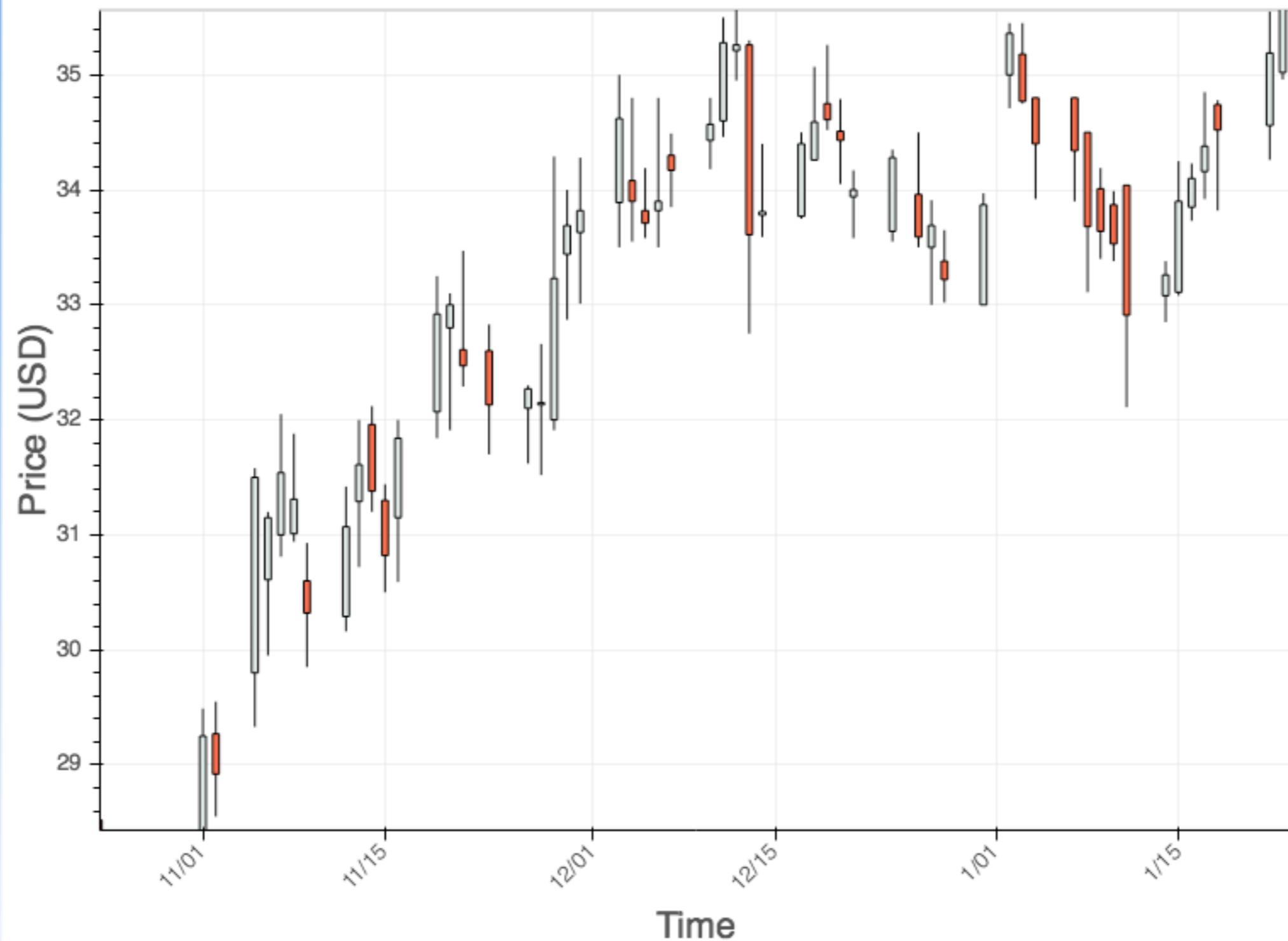


Boxplot



Catalit LLC

TSLA Candlestick



Catalit LLC

goo.gl/QvWD2Y

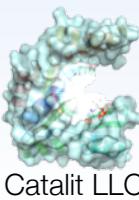
Name	Date Modified
00_prerequisites_check.py	Today, 2:36 PM
01a_multiple_plots.py	Today, 2:36 PM
01b_plot_csv.py	Today, 2:36 PM
02_regression.py	Today, 2:36 PM
03_classification.py	Today, 2:36 PM
04_clustering.py	Today, 2:36 PM
05_cleaning.py	Today, 2:36 PM
06a_improving_accuracy.py	Today, 2:36 PM
06b_feature_importance.py	Today, 2:36 PM
advanced	Today, 2:36 PM
data	Today, 2:36 PM
environment.yml	Today, 2:36 PM
finalproject	Today, 2:36 PM
README.md	Today, 2:36 PM
slides	Today, 2:38 PM



Catalit LLC

VISUALIZATION LAB

- cd <folder>
- conda env create
- source activate dataweekends
- python 00_prerequisites_check.py
- Houston, we are go! :)



Catalit LLC

VISUALIZATION LAB

- `python 01a_multiple_plots.py`
- `python 01b_plot_csv.py`
- Visualize a datasets
- Tell a story to the person next to you
- ask clarifying questions on the story
- switch roles