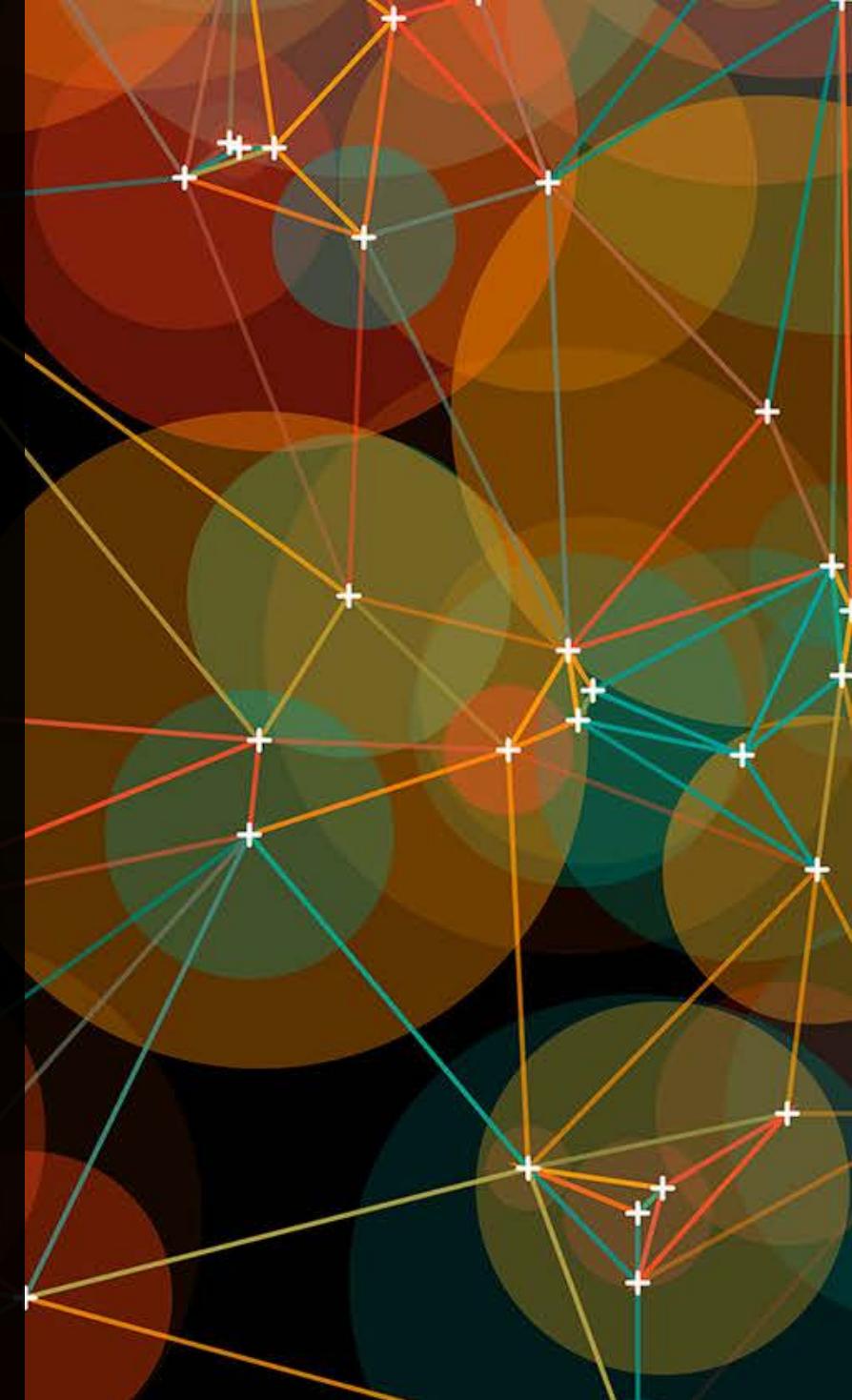




Microsoft Data Science Summit

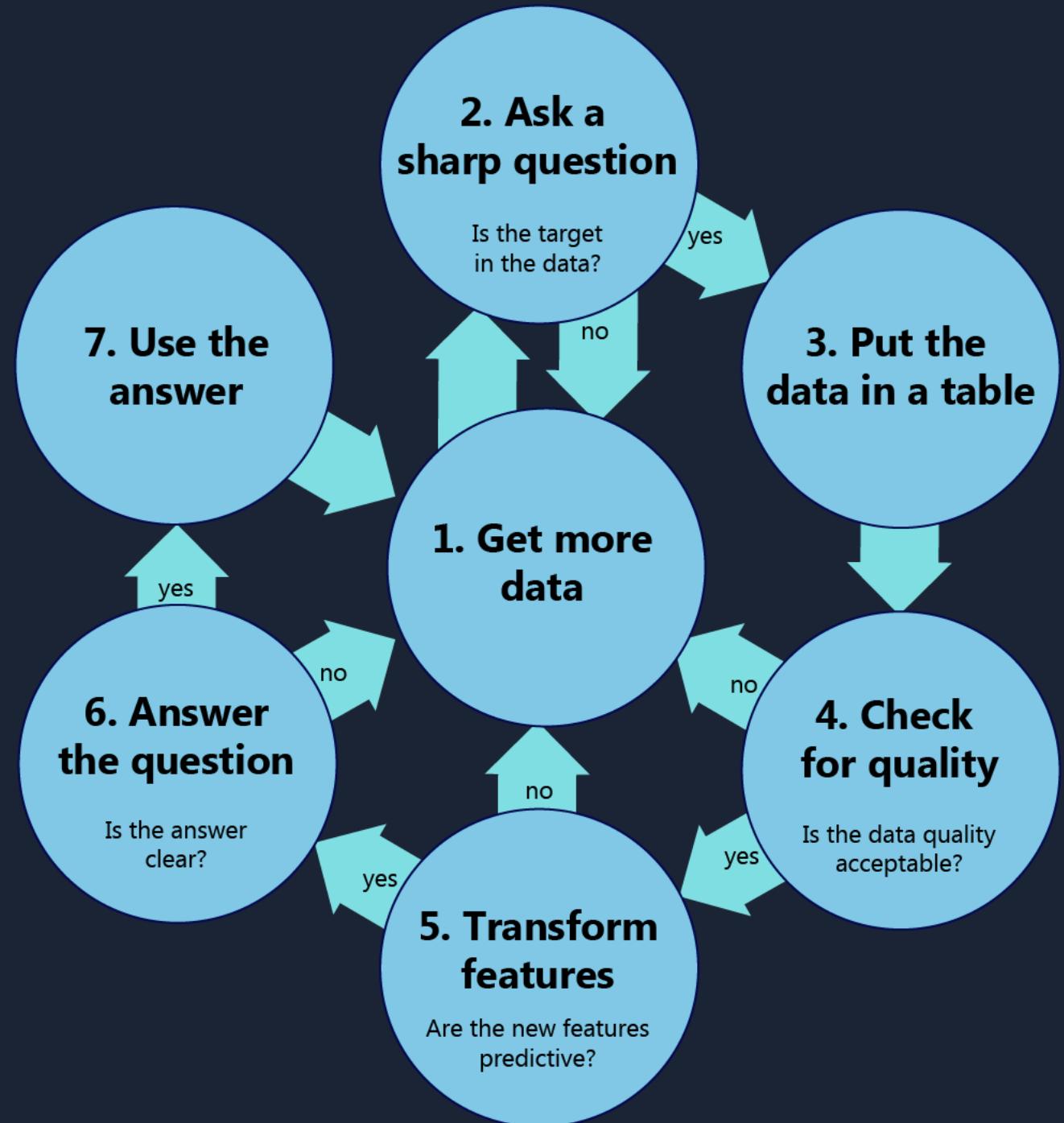
Sept 26 – 27 | Atlanta, GA

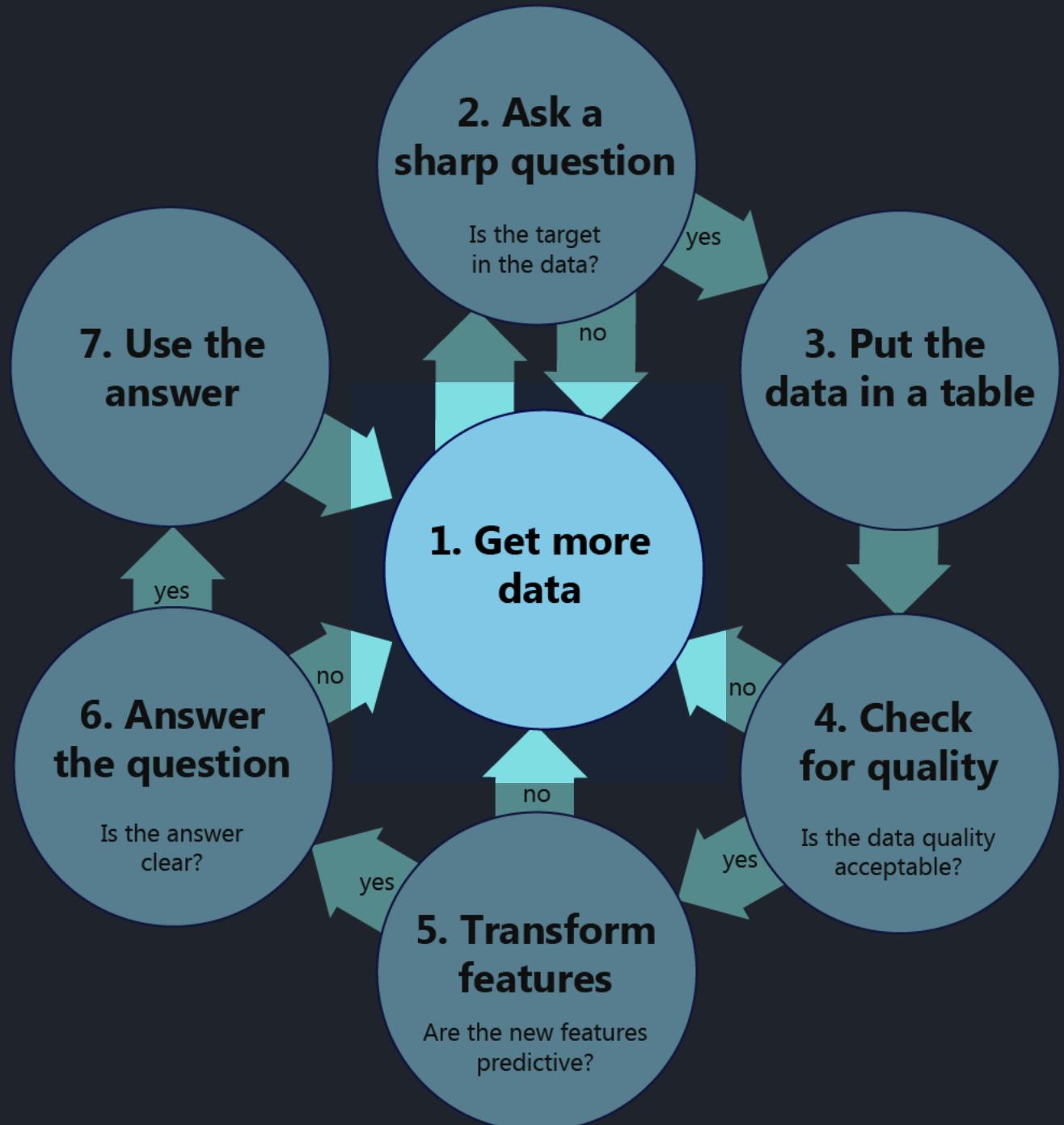


Data Science for Absolutely Everyone

Brandon Rohrer
Principal Data Scientist
Microsoft







Numbers and Names

Numbers

Amount : 38.3 degrees

Count : 39 pizzas

Money : \$1,387

Pixel brightness : 232/255

Sound intensity : .64

Names

Type : Shih Tzu

Variety : Caramel latte

ID : Air Force One

Model number : R2-D2

Category : Chocolate

Text : "Best. Show. Ever. <3"

Names that look like numbers

Phone number : 847-5609

Zip code : 90210

ID number : 007

Serial number : 100000184573

Credit card number : 5738-7539-9898-0023

Social security number : 627-42-0932

Numbers that look like names
and names that can be turned into numbers

Place : first, second, third

Size : small, medium, large

Side : left, middle, right

Time zone : Pacific, Mountain, Central, Eastern

Train stops : Kendall, Central, Harvard, Porter

Data Engineering

Measure

Collect

Store

Search

Move

Transform

Azure Event Hub

Azure Stream Analytics

Azure Data Factory

Hadoop and Spark on
Azure HDInsight

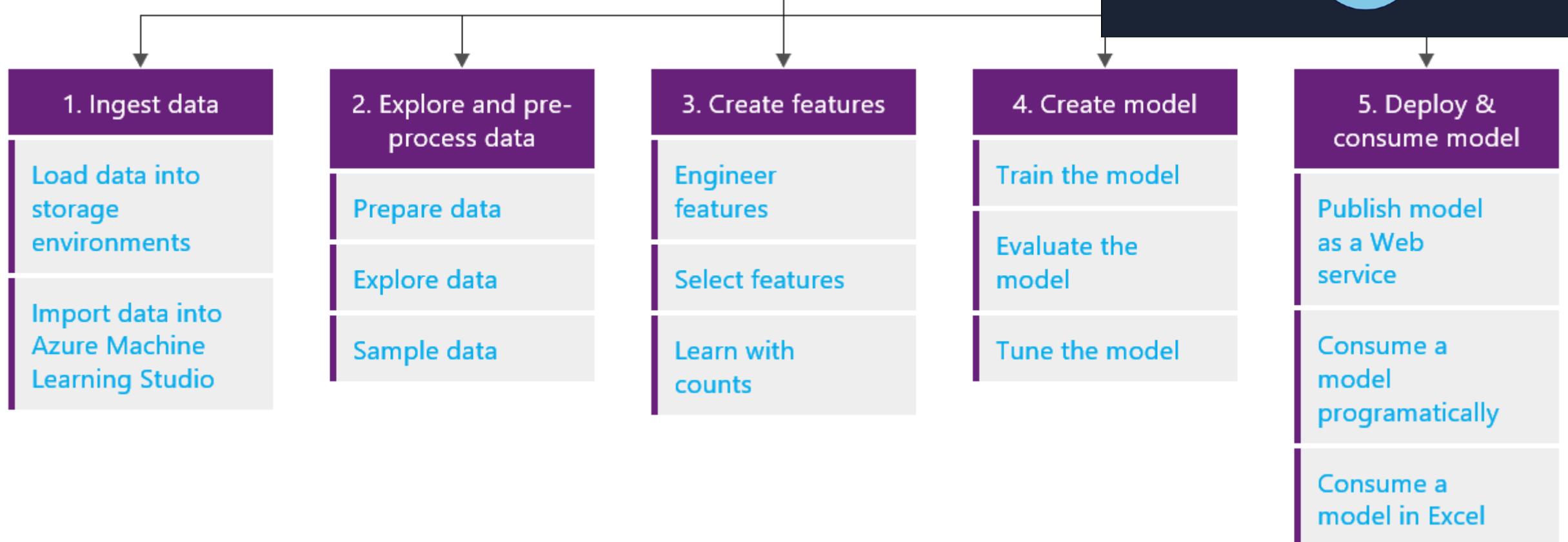
Azure Search

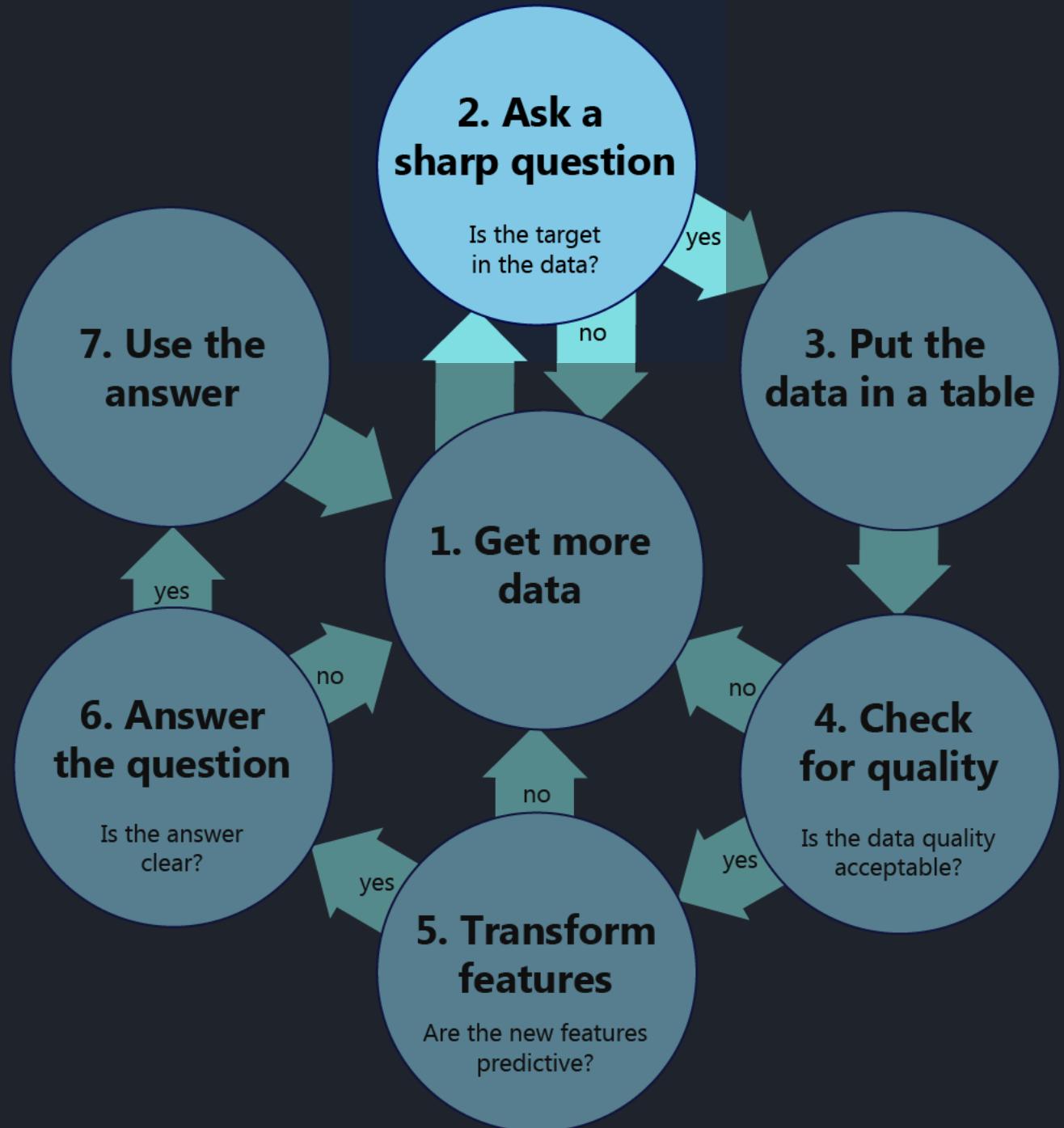
Azure DocumentDB

Azure Data Lake

Azure Data Catalog

Microsoft Data Science Process

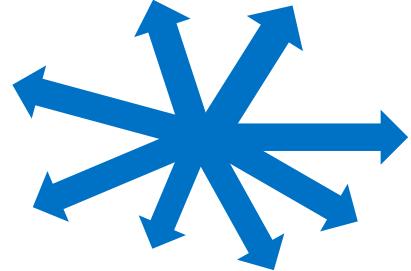




Vague questions

vs.

Sharp questions



Doesn't have to be answered
with a name or a number

What can my data tell me
about my business?

What should I do?

How can I increase my profits?



Must be answered with a
name or a number.

How many times will the
feature I built get used by a
new user?

Which route through
downtown will get me to work
the fastest?

Target

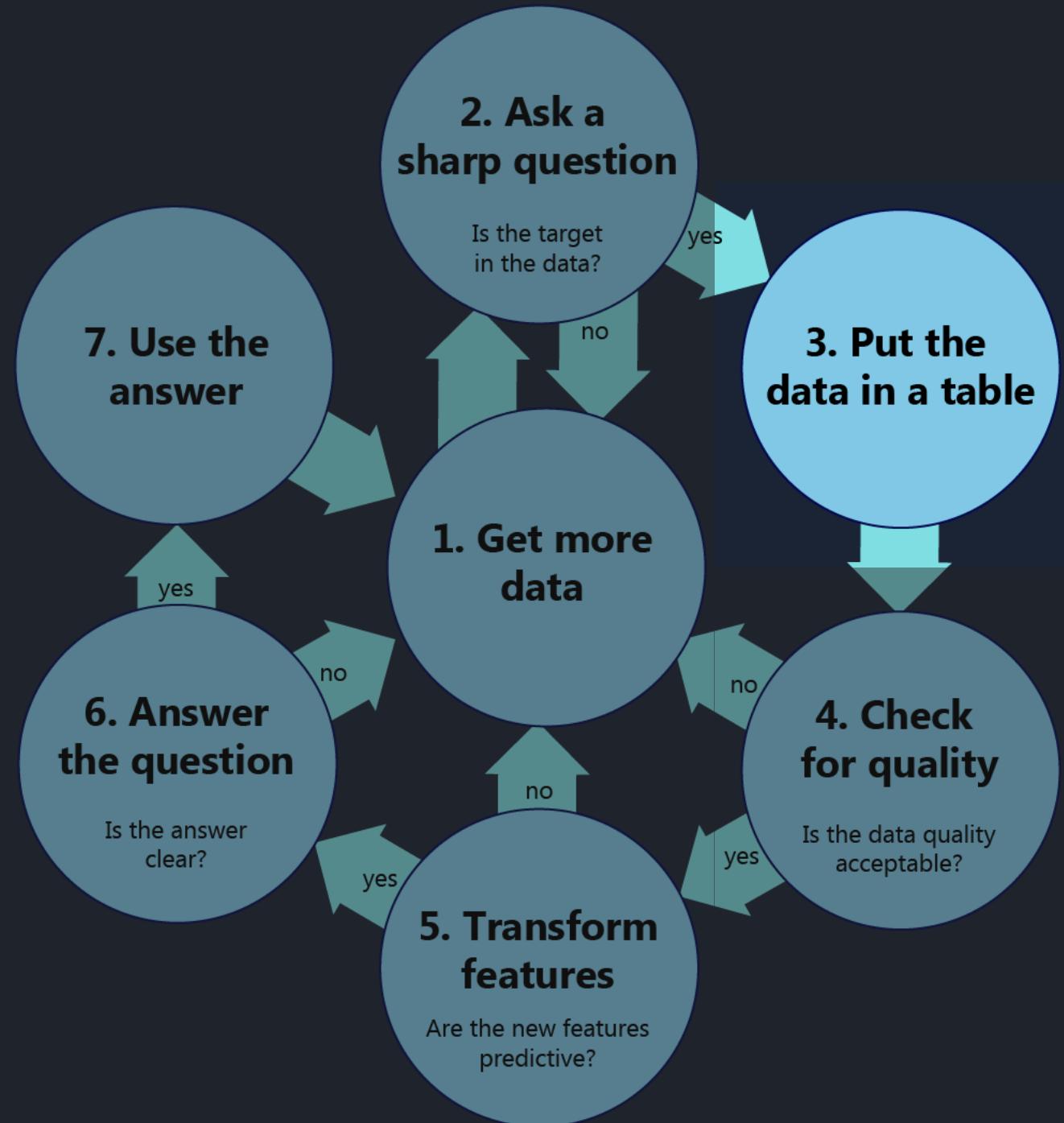
What will my stock price be next week?

Date	Americas sales	Europe and Africa sales	Asia sales
	Competitor	Product	Market share
Product	First month users	First quarter users	First year users
	Date	Dow Jones	Nikkei

Target

What will my stock price be next week?

Date	Americas sales	Europe and Africa sales		
Product	First month users	First quarter users	Competitor	Date
2023-01-01	1000	800	500	2023-01-01
2023-01-02	1050	850	550	2023-01-02
2023-01-03	1100	900	600	2023-01-03
2023-01-04	1150	950	650	2023-01-04
2023-01-05	1200	1000	700	2023-01-05
2023-01-06	1250	1050	750	2023-01-06
2023-01-07	1300	1100	800	2023-01-07
2023-01-08	1350	1150	850	2023-01-08
2023-01-09	1400	1200	900	2023-01-09
2023-01-10	1450	1250	950	2023-01-10
2023-01-11	1500	1300	1000	2023-01-11
2023-01-12	1550	1350	1050	2023-01-12
2023-01-13	1600	1400	1100	2023-01-13
2023-01-14	1650	1450	1150	2023-01-14
2023-01-15	1700	1500	1200	2023-01-15
2023-01-16	1750	1550	1250	2023-01-16
2023-01-17	1800	1600	1300	2023-01-17
2023-01-18	1850	1650	1350	2023-01-18
2023-01-19	1900	1700	1400	2023-01-19
2023-01-20	1950	1750	1450	2023-01-20
2023-01-21	2000	1800	1500	2023-01-21
2023-01-22	2050	1850	1550	2023-01-22
2023-01-23	2100	1900	1600	2023-01-23
2023-01-24	2150	1950	1650	2023-01-24
2023-01-25	2200	2000	1700	2023-01-25
2023-01-26	2250	2050	1750	2023-01-26
2023-01-27	2300	2100	1800	2023-01-27
2023-01-28	2350	2150	1850	2023-01-28
2023-01-29	2400	2200	1900	2023-01-29
2023-01-30	2450	2250	1950	2023-01-30
2023-01-31	2500	2300	2000	2023-01-31
2023-02-01	2550	2350	2050	2023-02-01
2023-02-02	2600	2400	2100	2023-02-02
2023-02-03	2650	2450	2150	2023-02-03
2023-02-04	2700	2500	2200	2023-02-04
2023-02-05	2750	2550	2250	2023-02-05
2023-02-06	2800	2600	2300	2023-02-06
2023-02-07	2850	2650	2350	2023-02-07
2023-02-08	2900	2700	2400	2023-02-08
2023-02-09	2950	2750	2450	2023-02-09
2023-02-10	3000	2800	2500	2023-02-10
2023-02-11	3050	2850	2550	2023-02-11
2023-02-12	3100	2900	2600	2023-02-12
2023-02-13	3150	2950	2650	2023-02-13
2023-02-14	3200	3000	2700	2023-02-14
2023-02-15	3250	3050	2750	2023-02-15
2023-02-16	3300	3100	2800	2023-02-16
2023-02-17	3350	3150	2850	2023-02-17
2023-02-18	3400	3200	2900	2023-02-18
2023-02-19	3450	3250	2950	2023-02-19
2023-02-20	3500	3300	3000	2023-02-20
2023-02-21	3550	3350	3050	2023-02-21
2023-02-22	3600	3400	3100	2023-02-22
2023-02-23	3650	3450	3150	2023-02-23
2023-02-24	3700	3500	3200	2023-02-24
2023-02-25	3750	3550	3250	2023-02-25
2023-02-26	3800	3600	3300	2023-02-26
2023-02-27	3850	3650	3350	2023-02-27
2023-02-28	3900	3700	3400	2023-02-28
2023-02-29	3950	3750	3450	2023-02-29
2023-03-01	4000	3800	3500	2023-03-01
2023-03-02	4050	3850	3550	2023-03-02
2023-03-03	4100	3900	3600	2023-03-03
2023-03-04	4150	3950	3650	2023-03-04
2023-03-05	4200	4000	3700	2023-03-05
2023-03-06	4250	4050	3750	2023-03-06
2023-03-07	4300	4100	3800	2023-03-07
2023-03-08	4350	4150	3850	2023-03-08
2023-03-09	4400	4200	3900	2023-03-09
2023-03-10	4450	4250	3950	2023-03-10
2023-03-11	4500	4300	4000	2023-03-11
2023-03-12	4550	4350	4050	2023-03-12
2023-03-13	4600	4400	4100	2023-03-13
2023-03-14	4650	4450	4150	2023-03-14
2023-03-15	4700	4500	4200	2023-03-15
2023-03-16	4750	4550	4250	2023-03-16
2023-03-17	4800	4600	4300	2023-03-17
2023-03-18	4850	4650	4350	2023-03-18
2023-03-19	4900	4700	4400	2023-03-19
2023-03-20	4950	4750	4450	2023-03-20
2023-03-21	5000	4800	4500	2023-03-21
2023-03-22	5050	4850	4550	2023-03-22
2023-03-23	5100	4900	4600	2023-03-23
2023-03-24	5150	4950	4650	2023-03-24
2023-03-25	5200	5000	4700	2023-03-25
2023-03-26	5250	5050	4750	2023-03-26
2023-03-27	5300	5100	4800	2023-03-27
2023-03-28	5350	5150	4850	2023-03-28
2023-03-29	5400	5200	4900	2023-03-29
2023-03-30	5450	5250	4950	2023-03-30
2023-03-31	5500	5300	5000	2023-03-31
2023-04-01	5550	5350	5050	2023-04-01
2023-04-02	5600	5400	5100	2023-04-02
2023-04-03	5650	5450	5150	2023-04-03
2023-04-04	5700	5500	5200	2023-04-04
2023-04-05	5750	5550	5250	2023-04-05
2023-04-06	5800	5600	5300	2023-04-06
2023-04-07	5850	5650	5350	2023-04-07
2023-04-08	5900	5700	5400	2023-04-08
2023-04-09	5950	5750	5450	2023-04-09
2023-04-10	6000	5800	5500	2023-04-10
2023-04-11	6050	5850	5550	2023-04-11
2023-04-12	6100	5900	5600	2023-04-12
2023-04-13	6150	5950	5650	2023-04-13
2023-04-14	6200	6000	5700	2023-04-14
2023-04-15	6250	6050	5750	2023-04-15
2023-04-16	6300	6100	5800	2023-04-16
2023-04-17	6350	6150	5850	2023-04-17
2023-04-18	6400	6200	5900	2023-04-18
2023-04-19	6450	6250	5950	2023-04-19
2023-04-20	6500	6300	6000	2023-04-20
2023-04-21	6550	6350	6050	2023-04-21
2023-04-22	6600	6400	6100	2023-04-22
2023-04-23	6650	6450	6150	2023-04-23
2023-04-24	6700	6500	6200	2023-04-24
2023-04-25	6750	6550	6250	2023-04-25
2023-04-26	6800	6600	6300	2023-04-26
2023-04-27	6850	6650	6350	2023-04-27
2023-04-28	6900	6700	6400	2023-04-28
2023-04-29	6950	6750	6450	2023-04-29
2023-04-30	7000	6800	6500	2023-04-30
2023-05-01	7050	6850	6550	2023-05-01
2023-05-02	7100	6900	6600	2023-05-02
2023-05-03	7150	6950	6650	2023-05-03
2023-05-04	7200	7000	6700	2023-05-04
2023-05-05	7250	7050	6750	2023-05-05
2023-05-06	7300	7100	6800	2023-05-06
2023-05-07	7350	7150	6850	2023-05-07
2023-05-08	7400	7200	6900	2023-05-08
2023-05-09	7450	7250	6950	2023-05-09
2023-05-10	7500	7300	7000	2023-05-10
2023-05-11	7550	7350	7050	2023-05-11
2023-05-12	7600	7400	7100	2023-05-12
2023-05-13	7650	7450	7150	2023-05-13
2023-05-14	7700	7500	7200	2023-05-14
2023-05-15	7750	7550	7250	2023-05-15
2023-05-16	7800	7600	7300	2023-05-16
2023-05-17	7850	7650	7350	2023-05-17
2023-05-18	7900	7700	7400	2023-05-18
2023-05-19	7950	7750	7450	2023-05-19
2023-05-20	8000	7800	7500	2023-05-20
2023-05-21	8050	7850	7550	2023-05-21
2023-05-22	8100	7900	7600	2023-05-22
2023-05-23	8150	7950	7650	2023-05-23
2023-05-24	8200	8000	7700	2023-05-24
2023-05-25	8250	8050	7750	2023-05-25
2023-05-26	8300	8100	7800	2023-05-26
2023-05-27	8350	8150	7850	2023-05-27
2023-05-28	8400	8200	7900	2023-05-28
2023-05-29	8450	8250	7950	2023-05-29
2023-05-30	8500	8300	8000	2023-05-30
2023-05-31	8550	8350	8050	2023-05-31
2023-06-01	8600	8400	8100	2023-06-01
2023-06-02	8650	8450	8150	2023-06-02
2023-06-03	8700	8500	8200	2023-06-03
2023-06-04	8750	8550	8250	2023-06-04
2023-06-05	8800	8600	8300	2023-06-05
2023-06-06	8850	8650	8350	2023-06-06
2023-06-07	8900	8700	8400	2023-06-07
2023-06-08	8950	8750	8450	2023-06-08
2023-06-09	9000	8800	8500	2023-06-09
2023-06-10	9050	8850	8550	2023-06-10
2023-06-11	9100	8900	8600	2023-06-11
2023-06-12	9150	8950	8650	2023-06-12
2023-06-13	9200	9000	8700	2023-06-13
2023-06-14	9250	9050	8750	2023-06-14
2023-06-15	9300	9100	8800	2023-06-15
2023-06-16	9350	9150	8850	2023-06-16
2023-06-17	9400	9200	8900	2023-06-17
2023-06-18	9450	9250	8950	2023-06-18
2023-06-19	9500	9300	9000	2023-06-19
2023-06-20	9550	9350	9050	2023-06-20
2023-06-21	9600	9400	9100	2023-06-21
2023-06-22	9650	9450	9150	2023-06-22
2023-06-23	9700	9500	9200	2023-06-23
2023-06-24	9750	9550	9250	2023-06-24
2023-06-25	9800	9600	9300	2023-06-25
2023-06-26	9850	9650	9350	2023-06-26
2023-06-27	9900	9700	9400	2023-06-27
2023-06-28	9950	9750	9450	2023-06-28
2023-06-29	10000	9800	9500	2023-06-29
2023-06-30	10050	9850	9550	2023-06-30
2023-07-01	10100	9900	9600	2023-07-01
2023-07-02	10150	9950	9650	2023-07-02
2023-07-03	10200	10000	9700	2023-07-03
2023-07-04	10250	10050	9750	2023-07-04
2023-07-05	10300	10100	9800	2023-07-05
2023-07-06	10350	10150	9850	2023-07-06
2023-07-07	10400	10200	9900	2023-07-07
2023-07-08	10450	10250	9950	2023-07-08
2023-07-09	10500	10300	10000	2023-07-09
2023-07-10	10550	10350	10050	2023-07-10
2023-07-11	10600	10400	10100	2023-07-11
2023-07-12	10650	10450	10150	2023-07-12
2023-07-13	10700	10500	10200	2023-07-13
2023-07-14	10750	10550	10250	2023-07-14
2023-07-15	10800	10600	10300	



One target per row

Stock price	Date	Day of week	Dow Jones	Last month sales	Last quarter sales	Market share	New users last month	New users last quarter	Days since press release	Days since product release	Total users
57.3	5/21	Tue	17,245	68.8M	211.2M	23.1%	63,522	195,322	3	96	2.49M
58.8	5/22	Wed	17,289	68.8M	211.2M	23.1%	63,522	195,322	4	97	2.49M
56.9	5/23	Thu	17,115	68.8M	211.2M	23.1%	63,522	195,322	5	98	2.49M
57.4	5/24	Fri	17,278	68.8M	211.2M	23.1%	63,522	195,322	6	99	2.49M

One target per row

Aggregate

User name	Date joined
little_lil	Jan 27, 2014
popoverGuy	Jan 27, 2014
Red_Red	Jan 28, 2014
David_G_53	Jan 30, 2014
randll	Jan 30, 2014
...	...

Stock price	Date	Day of week	Dow Jones	Last month sales	Last quarter sales	Market share	New users last month	New users last quarter	Days since press release	Days since product release	Total users
57.3	5/21	Tue	17,245	68.8M	211.2M	23.1%	63,522	195,322	3	96	2.49M
58.8	5/22	Wed	17,289	68.8M	211.2M	23.1%	63,522	195,322	4	97	2.49M
56.9	5/23	Thu	17,115	68.8M	211.2M	23.1%	63,522	195,322	5	98	2.49M
57.4	5/24	Fri	17,278	68.8M	211.2M	23.1%	63,522	195,322	6	99	2.49M

One target per row

Aggregate Distribute

Quarter	Total sales
2015Q4	119.2M
Month	Total sales
2016/01	43.0M
2016/02	60.1M
2016/03	55.5M
2016/04	41.7M
2016/05	68.8M
...	...

Stock price	Date	Day of week	Dow Jones	Last month sales	Last quarter sales	Market share	New users last month	New users last quarter	Days since press release	Days since product release	Total users
57.3	5/21	Tue	17,245	68.8M	211.2M	23.1%	63,522	195,322	3	96	2.493M
58.8	5/22	Wed	17,289	68.8M	211.2M	23.1%	63,522	195,322	4	97	2.494M
56.9	5/23	Thu	17,115	68.8M	211.2M	23.1%	63,522	195,322	5	98	2.494M
57.4	5/24	Fri	17,278	68.8M	211.2M	23.1%	63,522	195,322	6	99	2.495M

One target per row

Aggregate
Distribute
Compute

Press release date	Subject
2016/03/24	Mega amazing whizbang
2016/05/03	Super widget upgrade
2016/05/18	New gizmos on the flimflam
...	...

Stock price	Date	Day of week	Dow Jones	Last month sales	Last quarter sales	Market share	New users last month	New users last quarter	Days since press release	Days since product release	Total users
57.3	5/21	Tue	17,245	68.8M	211.2M	23.1%	63,522	195,322	3	96	2.49M
58.8	5/22	Wed	17,289	68.8M	211.2M	23.1%	63,522	195,322	4	97	2.49M
56.9	5/23	Thu	17,115	68.8M	211.2M	23.1%	63,522	195,322	5	98	2.49M
57.4	5/24	Fri	17,278	68.8M	211.2M	23.1%	63,522	195,322	6	99	2.49M

One target per row

Aggregate

Measure

Distribute

Compute

Stock price	Date	Day of week	Dow Jones	Last month sales	Last quarter sales	Market share	New users last month	New users last quarter	Days since press release	Days since product release	Total users
57.3	5/21	Tue	17,245	68.8M	211.2M	23.1%	63,522	195,322	3	96	2.49M
58.8	5/22	Wed	17,289	68.8M	211.2M	23.1%	63,522	195,322	4	97	2.49M
56.9	5/23	Thu	17,115	68.8M	211.2M	23.1%	63,522	195,322	5	98	2.49M
57.4	5/24	Fri	17,278	68.8M	211.2M	23.1%	63,522	195,322	6	99	2.49M

One target per row

Aggregate
Distribute
Compute

Measure
Estimate

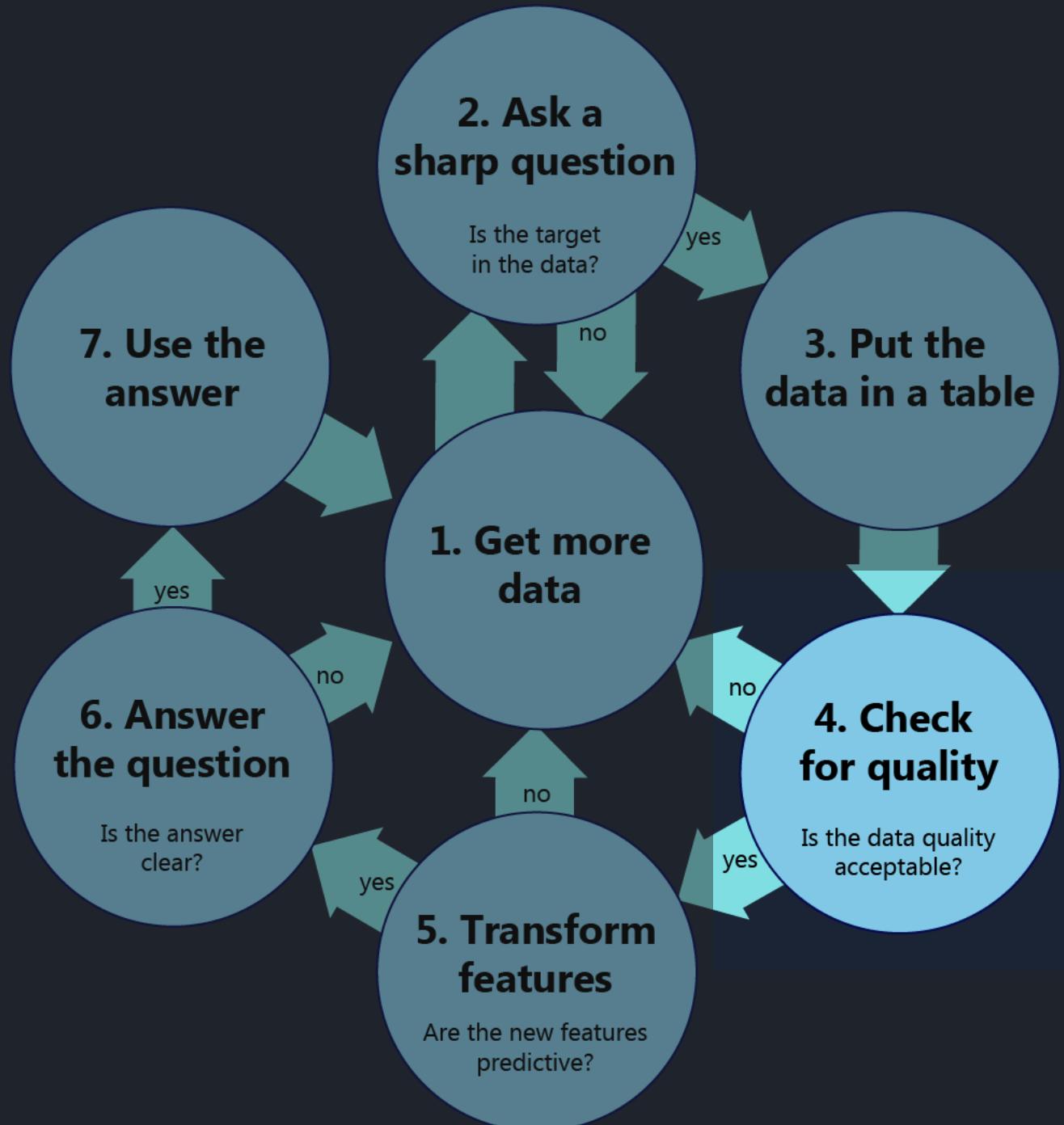
Stock price	Date	Day of week	Dow Jones	Last month sales	Last quarter sales	Market share	New users last month	New users last quarter	Days since press release	Days since product release	Total users
57.3	5/21	Tue	17,245	68.8M	211.2M	23.1%	63,522	195,322	3	96	2.49M
58.8	5/22	Wed	17,289	68.8M	211.2M	23.1%	63,522	195,322	4	97	2.49M
56.9	5/23	Thu	17,115	68.8M	211.2M	23.1%	63,522	195,322	5	98	2.49M
57.4	5/24	Fri	17,278	68.8M	211.2M	23.1%	63,522	195,322	6	99	2.49M

One target per row

Aggregate
Distribute
Compute

Measure
Estimate
Leave blanks

Stock price	Date	Day of week	Dow Jones	Last month sales	Last quarter sales	Market share	New users last month	New users last quarter	Days since press release	Days since product release	Total users
57.3	5/21	Tue	17,245	68.8M	211.2M	23.1%	63,522	195,322	3	96	2.49M
58.8	5/22	Wed	17,289	68.8M	211.2M	23.1%	63,522	195,322	4	97	2.49M
56.9	5/23	Thu	17,115	68.8M	211.2M	23.1%	63,522	195,322	5	98	2.49M
57.4	5/24	Fri	17,278	68.8M	211.2M	23.1%	63,522	195,322	6	99	2.49M



ID	First name	Last name	Birth year	Height	Birthplace	Identity is secret	Can fly	Alignment	Wears cape
7435	Bruce	Wayne	1969*	6' 2"	Gotham	Y	3	anti-villain	black
0958	Ororo	Munroe	--1979--	5' 11"	Manhattan		9	good	long
9471	Diana	Trevor	1618	5' 8"	Paradise Island	Y	Jet	truth	rarely
9483	Janet	Van Dyne	19.42	5' 4"	Cresskill		tiny	Good	Not really
0696	Peter	Parker	1111983	5' 10"	Queens	Y	Fall	right	never
5531	Harleen	Quinzell	1981	5' 2"	Gotham	Y	-	evil	no
4734	Erik	Lehnsherr	1-9-3-2	6' 0"	Hamburg		Lev.	mutants	Absolutely
7757	Natasha	Romanova	1983	5' 7"	St. Petersburg		jet	depends	No way
0323	Jean	Grey	"1977"	5' 6"	Annandale		No	good	Mostly not
3980	Clark	Kent	"1954"	6' 4"	Krypton	Y	12	Truth	always
3057	Victor	Von Doom	"1943"	6' 2"	Latveria		1	Bad	yes
0573	Stephen	Strange	1968	6' 2"	Philadelphia		not	light	Y
7452	Thor	Odinson	2287 BC	6' 6"	Norway		10	Good	Of course
1437	Selina	Kyle	1998	5' 7"	Gotham	Y	NA	Neutral	It clashes
1883	Raven	Darkholme	..1911..	5' 10"	unknown	Y	no	mostly bad	Not really
5830	Kara	Zor-el	1961	5' 7"	Krypton	Y	fast	G	Yes

ID	First name	Last name	Birth year	Height	Birthplace	Identity is secret	Can fly	Alignment	Wears cape
7435	Bruce	Wayne	1969*	6' 2"	Gotham	Y	3	anti-villain	black
0958	Ororo	Munroe	--1979--	5' 11"	Manhattan		9	good	long
9471	Diana	Trevor	1618	5' 8"	Paradise Island	Y	Jet	truth	rarely
9483	Janet	Van Dyne	19.42	5' 4"	Cresskill		tiny	Good	Not really
0696	Peter	Parker	1111983	5' 10"	Queens	Y	Fall	right	never
5531	Harleen	Quinzell	1981	5' 2"	Gotham	Y	-	evil	no
4734	Erik	Lehnsherr	1-9-3-2	6' 0"	Hamburg		Lev.	mutants	Absolutely
7757	Natasha	Romanova	1983	5' 7"	St. Petersburg		jet	depends	No way
0323	Jean	Grey	"1977"	5' 6"	Annandale		No	good	Mostly not
3980	Clark	Kent	"1954"	6' 4"	Krypton	Y	12	Truth	always
3057	Victor	Von Doom	"1943"	6' 2"	Latveria		1	Bad	yes
0573	Stephen	Strange	1968	6' 2"	Philadelphia		not	light	Y
7452	Thor	Odinson	2287 BC	6' 6"	Norway		10	Good	Of course
1437	Selina	Kyle	1998	5' 7"	Gotham	Y	NA	Neutral	It clashes
1883	Raven	Darkholme	..1911..	5' 10"	unknown	Y	no	mostly bad	Not really
5830	Kara	Zor-el	1961	5' 7"	Krypton	Y	fast	G	Yes

ID	First name	Last name	Birth year	Height	Birthplace	Identity is secret	Can fly	Alignment	Wears cape
7435	Bruce	Wayne	1969	6' 2"	Gotham	Y	3	anti-villain	black
0958	Ororo	Munroe	1979	5' 11"	Manhattan		9	good	long
9471	Diana	Trevor	1618	5' 8"	Paradise Island	Y	Jet	truth	rarely
9483	Janet	Van Dyne	1942	5' 4"	Cresskill		tiny	Good	Not really
0696	Peter	Parker	1983	5' 10"	Queens	Y	Fall	right	never
5531	Harleen	Quinzell	1981	5' 2"	Gotham	Y	-	evil	no
4734	Erik	Lehnsherr	1932	6' 0"	Hamburg		Lev.	mutants	Absolutely
7757	Natasha	Romanova	1983	5' 7"	St. Petersburg		jet	depends	No way
0323	Jean	Grey	1977	5' 6"	Annandale		No	good	Mostly not
3980	Clark	Kent	1954	6' 4"	Krypton	Y	12	Truth	always
3057	Victor	Von Doom	1943	6' 2"	Latveria		1	Bad	yes
0573	Stephen	Strange	1968	6' 2"	Philadelphia		not	light	Y
7452	Thor	Odinson	-2287	6' 6"	Norway		10	Good	Of course
1437	Selina	Kyle	1998	5' 7"	Gotham	Y	NA	Neutral	It clashes
1883	Raven	Darkholme	1911	5' 10"	unknown	Y	no	mostly bad	Not really
5830	Kara	Zor-el	1961	5' 7"	Krypton	Y	fast	G	Yes

ID	First name	Last name	Birth year	Height	Birthplace	Identity is secret	Can fly	Alignment	Wears cape
7435	Bruce	Wayne	1969	6' 2"	Gotham	Y	3	anti-villain	black
0958	Ororo	Munroe	1979	5' 11"	Manhattan		9	good	long
9471	Diana	Trevor	1618	5' 8"	Paradise Island	Y	Jet	truth	rarely
9483	Janet	Van Dyne	1942	5' 4"	Cresskill		tiny	Good	Not really
0696	Peter	Parker	1983	5' 10"	Queens	Y	Fall	right	never
5531	Harleen	Quinzell	1981	5' 2"	Gotham	Y	-	evil	no
4734	Erik	Lehnsherr	1932	6' 0"	Hamburg		Lev.	mutants	Absolutely
7757	Natasha	Romanova	1983	5' 7"	St. Petersburg		jet	depends	No way
0323	Jean	Grey	1977	5' 6"	Annandale		No	good	Mostly not
3980	Clark	Kent	1954	6' 4"	Krypton	Y	12	Truth	always
3057	Victor	Von Doom	1943	6' 2"	Latveria		1	Bad	yes
0573	Stephen	Strange	1968	6' 2"	Philadelphia		not	light	Y
7452	Thor	Odinson	-2287	6' 6"	Norway		10	Good	Of course
1437	Selina	Kyle	1998	5' 7"	Gotham	Y	NA	Neutral	It clashes
1883	Raven	Darkholme	1911	5' 10"	unknown	Y	no	mostly bad	Not really
5830	Kara	Zor-el	1961	5' 7"	Krypton	Y	fast	G	Yes

ID	First name	Last name	Birth year	Height	Birthplace	Identity is secret	Can fly	Alignment	Wears cape
7435	Bruce	Wayne	1969	74	Gotham	Y	3	anti-villain	black
0958	Ororo	Munroe	1979	71	Manhattan		9	good	long
9471	Diana	Trevor	1618	68	Paradise Island	Y	Jet	truth	rarely
9483	Janet	Van Dyne	1942	64	Cresskill		tiny	Good	Not really
0696	Peter	Parker	1983	70	Queens	Y	Fall	right	never
5531	Harleen	Quinzell	1981	62	Gotham	Y	-	evil	no
4734	Erik	Lehnsherr	1932	72	Hamburg		Lev.	mutants	Absolutely
7757	Natasha	Romanova	1983	67	St. Petersburg		jet	depends	No way
0323	Jean	Grey	1977	66	Annandale		No	good	Mostly not
3980	Clark	Kent	1954	76	Krypton	Y	12	Truth	always
3057	Victor	Von Doom	1943	74	Latveria		1	Bad	yes
0573	Stephen	Strange	1968	74	Philadelphia		not	light	Y
7452	Thor	Odinson	-2287	78	Norway		10	Good	Of course
1437	Selina	Kyle	1998	67	Gotham	Y	NA	Neutral	It clashes
1883	Raven	Darkholme	1911	70	unknown	Y	no	mostly bad	Not really
5830	Kara	Zor-el	1961	67	Krypton	Y	fast	G	Yes

ID	First name	Last name	Birth year	Height	Birthplace	Identity is secret	Can fly	Alignment	Wears cape
7435	Bruce	Wayne	1969	74	Gotham	Y	3	anti-villain	black
0958	Ororo	Munroe	1979	71	Manhattan	NA	9	good	long
9471	Diana	Trevor	1618	68	Paradise Island	Y	Jet	truth	rarely
9483	Janet	Van Dyne	1942	64	Cresskill		tiny	Good	Not really
0696	Peter	Parker	1983	70	Queens	Y	Fall	right	never
5531	Harleen	Quinzell	1981	62	Gotham	Y	-	evil	no
4734	Erik	Lehnsherr	1932	72	Hamburg	NA	Lev.	mutants	Absolutely
7757	Natasha	Romanova	1983	67	St. Petersburg	NA	jet	depends	No way
0323	Jean	Grey	1977	66	Annandale		No	good	Mostly not
3980	Clark	Kent	1954	76	Krypton	Y	12	Truth	always
3057	Victor	Von Doom	1943	74	Latveria	Missing	1	Bad	yes
0573	Stephen	Strange	1968	74	Philadelphia		not	light	Y
7452	Thor	Odinson	-2287	78	Norway		10	Good	Of course
1437	Selina	Kyle	1998	67	Gotham	Y	NA	Neutral	It clashes
1883	Raven	Darkholme	1911	70	unknown	Y	no	mostly bad	Not really
5830	Kara	Zor-el	1961	67	Krypton	Y	fast	G	Yes

ID	First name	Last name	Birth year	Height	Birthplace	Identity is secret	Can fly	Alignment	Wears cape
7435	Bruce	Wayne	1969	74	Gotham	Y	3	anti-villain	black
0958	Ororo	Munroe	1979	71	Manhattan	N	9	good	long
9471	Diana	Trevor	1618	68	Paradise Island	Y	Jet	truth	rarely
9483	Janet	Van Dyne	1942	64	Cresskill	N	tiny	Good	Not really
0696	Peter	Parker	1983	70	Queens	Y	Fall	right	never
5531	Harleen	Quinzell	1981	62	Gotham	Y	-	evil	no
4734	Erik	Lehnsherr	1932	72	Hamburg	N	Lev.	mutants	Absolutely
7757	Natasha	Romanova	1983	67	St. Petersburg	N	jet	depends	No way
0323	Jean	Grey	1977	66	Annandale	N	No	good	Mostly not
3980	Clark	Kent	1954	76	Krypton	Y	12	Truth	always
3057	Victor	Von Doom	1943	74	Latveria	N	1	Bad	yes
0573	Stephen	Strange	1968	74	Philadelphia	N	not	light	Y
7452	Thor	Odinson	-2287	78	Norway	N	10	Good	Of course
1437	Selina	Kyle	1998	67	Gotham	Y	NA	Neutral	It clashes
1883	Raven	Darkholme	1911	70	unknown	Y	no	mostly bad	Not really
5830	Kara	Zor-el	1961	67	Krypton	Y	fast	G	Yes

ID	First name	Last name	Birth year	Height	Birthplace	Identity is secret	Can fly	Alignment	Wears cape
7435	Bruce	Wayne	1969	74	Gotham	Y	3	anti-villain	black
0958	Ororo	Munroe	1979	71	Manhattan	N	9	good	long
9471	Diana	Trevor	1618	68	Paradise Island	Y	Jet	truth	rarely
9483	Janet	Van Dyne	1942	64	Cresskill	N	tiny	Good	Not really
0696	Peter	Parker	1983	70	Queens	Y	Fall	right	never
5531	Harleen	Quinzell	1981	62	Gotham	Y	-	evil	no
4734	Erik	Lehnsherr	1932	72	Hamburg	N	Lev.	mutants	Absolutely
7757	Natasha	Romanova	1983	67	St. Petersburg	N	jet	depends	No way
0323	Jean	Grey	1977	66	Annandale	N	No	good	Mostly not
3980	Clark	Kent	1954	76	Krypton	Y	12	Truth	always
3057	Victor	Von Doom	1943	74	Latveria	N	1	Bad	yes
0573	Stephen	Strange	1968	74	Philadelphia	N	not	light	Y
7452	Thor	Odinson	-2287	78	Norway	N	10	Good	Of course
1437	Selina	Kyle	1998	67	Gotham	Y	NA	Neutral	It clashes
1883	Raven	Darkholme	1911	70	unknown	Y	no	mostly bad	Not really
5830	Kara	Zor-el	1961	67	Krypton	Y	fast	G	Yes

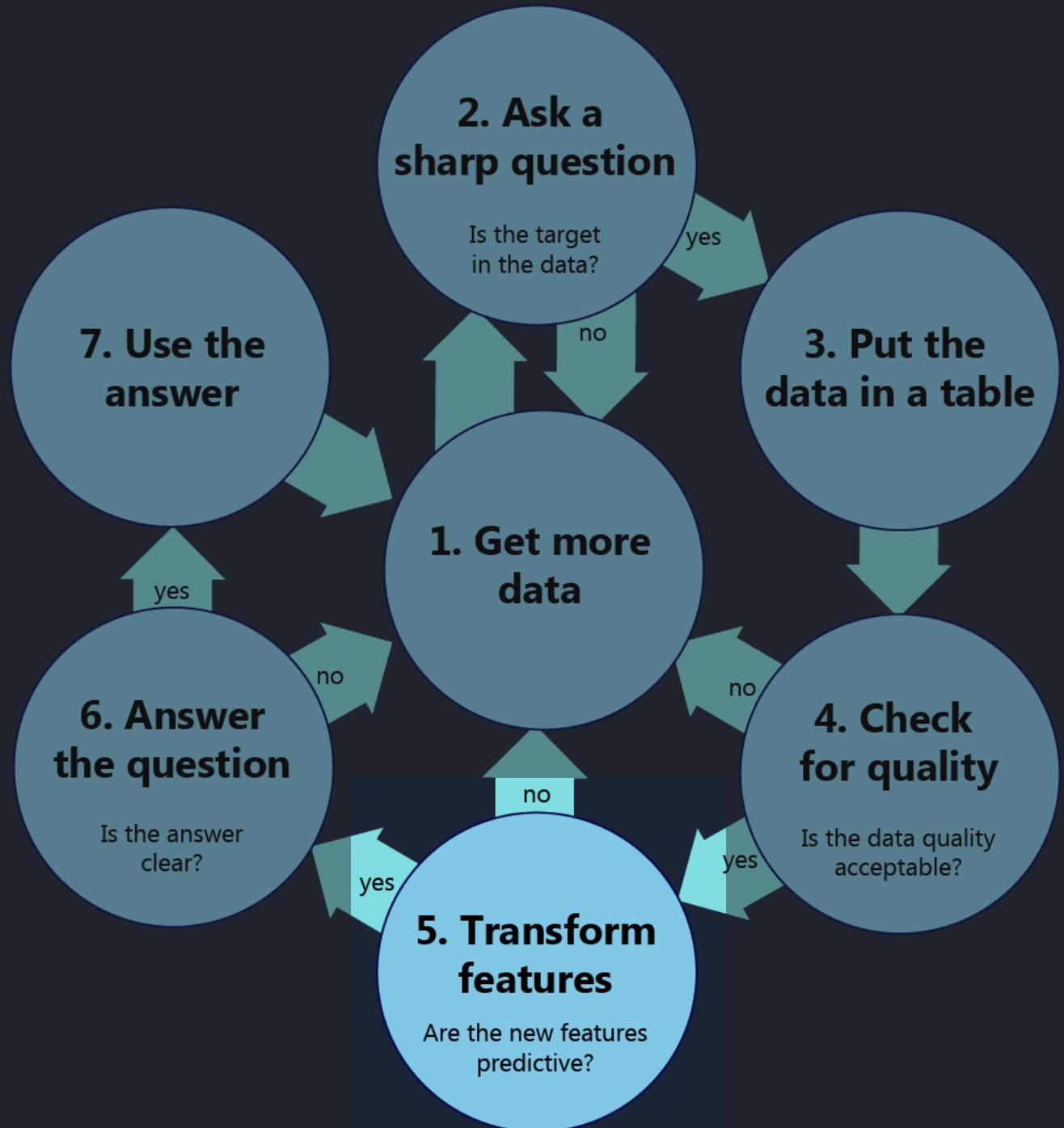
ID	First name	Last name	Birth year	Height	Birthplace	Identity is secret	Can fly	Alignment	Wears cape
7435	Bruce	Wayne	1969	74	Gotham	Y	N	anti-villain	black
0958	Ororo	Munroe	1979	71	Manhattan	N	Y	good	long
9471	Diana	Trevor	1618	68	Paradise Island	Y	N	truth	rarely
9483	Janet	Van Dyne	1942	64	Cresskill	N	Y	Good	Not really
0696	Peter	Parker	1983	70	Queens	Y	N	right	never
5531	Harleen	Quinzell	1981	62	Gotham	Y	N	evil	no
4734	Erik	Lehnsherr	1932	72	Hamburg	N	N	mutants	Absolutely
7757	Natasha	Romanova	1983	67	St. Petersburg	N	N	depends	No way
0323	Jean	Grey	1977	66	Annandale	N	N	good	Mostly not
3980	Clark	Kent	1954	76	Krypton	Y	Y	Truth	always
3057	Victor	Von Doom	1943	74	Latveria	N	N	Bad	yes
0573	Stephen	Strange	1968	74	Philadelphia	N	N	light	Y
7452	Thor	Odinson	-2287	78	Norway	N	Y	Good	Of course
1437	Selina	Kyle	1998	67	Gotham	Y	N	Neutral	It clashes
1883	Raven	Darkholme	1911	70	unknown	Y	N	mostly bad	Not really
5830	Kara	Zor-el	1961	67	Krypton	Y	Y	G	Yes

ID	First name	Last name	Birth year	Height	Birthplace	Identity is secret	Can fly	Alignment	Wears cape
7435	Bruce	Wayne	1969	74	Gotham	Y	N	anti-villain	black
0958	Ororo	Munroe	1979	71	Manhattan	N	Y	good	long
9471	Diana	Trevor	1618	68	Paradise Island	Y	N	truth	rarely
9483	Janet	Van Dyne	1942	64	Cresskill	N	Y	Good	Not really
0696	Peter	Parker	1983	70	Queens	Y	N	right	never
5531	Harleen	Quinzell	1981	62	Gotham	Y	N	evil	no
4734	Erik	Lehnsherr	1932	72	Hamburg	N	N	mutants	Absolutely
7757	Natasha	Romanova	1983	67	St. Petersburg	N	N	depends	No way
0323	Jean	Grey	1977	66	Annandale	N	N	good	Mostly not
3980	Clark	Kent	1954	76	Krypton	Y	Y	Truth	always
3057	Victor	Von Doom	1943	74	Latveria	N	N	Bad	yes
0573	Stephen	Strange	1968	74	Philadelphia	N	N	light	Y
7452	Thor	Odinson	-2287	78	Norway	N	Y	Good	Of course
1437	Selina	Kyle	1998	67	Gotham	Y	N	Neutral	It clashes
1883	Raven	Darkholme	1911	70	unknown	Y	N	mostly bad	Not really
5830	Kara	Zor-el	1961	67	Krypton	Y	Y	G	Yes

ID	First name	Last name	Birth year	Height	Birthplace	Identity is secret	Can fly	Alignment	Wears cape
7435	Bruce	Wayne	1969	74	Gotham	Y	N	Good	black
0958	Ororo	Munroe	1979	71	Manhattan	N	Y	Good	long
9471	Diana	Trevor	1618	68	Paradise Island	Y	N	Good	rarely
9483	Janet	Van Dyne	1942	64	Cresskill	N	Y	Good	Not really
0696	Peter	Parker	1983	70	Queens	Y	N	Good	never
5531	Harleen	Quinzell	1981	62	Gotham	Y	N	Bad	no
4734	Erik	Lehnsherr	1932	72	Hamburg	N	N	Bad	Absolutely
7757	Natasha	Romanova	1983	67	St. Petersburg	N	N	Good	No way
0323	Jean	Grey	1977	66	Annandale	N	N	Good	Mostly not
3980	Clark	Kent	1954	76	Krypton	Y	Y	Good	always
3057	Victor	Von Doom	1943	74	Latveria	N	N	Bad	yes
0573	Stephen	Strange	1968	74	Philadelphia	N	N	Good	Y
7452	Thor	Odinson	-2287	78	Norway	N	Y	Good	Of course
1437	Selina	Kyle	1998	67	Gotham	Y	N	Neutral	It clashes
1883	Raven	Darkholme	1911	70	unknown	Y	N	Bad	Not really
5830	Kara	Zor-el	1961	67	Krypton	Y	Y	Good	Yes

ID	First name	Last name	Birth year	Height	Birthplace	Identity is secret	Can fly	Alignment	Wears cape
7435	Bruce	Wayne	1969	74	Gotham	Y	N	Good	black
0958	Ororo	Munroe	1979	71	Manhattan	N	Y	Good	long
9471	Diana	Trevor	1618	68	Paradise Island	Y	N	Good	rarely
9483	Janet	Van Dyne	1942	64	Cresskill	N	Y	Good	Not really
0696	Peter	Parker	1983	70	Queens	Y	N	Good	never
5531	Harleen	Quinzell	1981	62	Gotham	Y	N	Bad	no
4734	Erik	Lehnsherr	1932	72	Hamburg	N	N	Bad	Absolutely
7757	Natasha	Romanova	1983	67	St. Petersburg	N	N	Good	No way
0323	Jean	Grey	1977	66	Annandale	N	N	Good	Mostly not
3980	Clark	Kent	1954	76	Krypton	Y	Y	Good	always
3057	Victor	Von Doom	1943	74	Latveria	N	N	Bad	yes
0573	Stephen	Strange	1968	74	Philadelphia	N	N	Good	Y
7452	Thor	Odinson	-2287	78	Norway	N	Y	Good	Of course
1437	Selina	Kyle	1998	67	Gotham	Y	N	Neutral	It clashes
1883	Raven	Darkholme	1911	70	unknown	Y	N	Bad	Not really
5830	Kara	Zor-el	1961	67	Krypton	Y	Y	Good	Yes

ID	First name	Last name	Birth year	Height	Birthplace	Identity is secret	Can fly	Alignment	Wears cape
7435	Bruce	Wayne	1969	74	Gotham	Y	N	Good	Y
0958	Ororo	Munroe	1979	71	Manhattan	N	Y	Good	Y
9471	Diana	Trevor	1618	68	Paradise Island	Y	N	Good	N
9483	Janet	Van Dyne	1942	64	Cresskill	N	Y	Good	N
0696	Peter	Parker	1983	70	Queens	Y	N	Good	N
5531	Harleen	Quinzell	1981	62	Gotham	Y	N	Bad	N
4734	Erik	Lehnsherr	1932	72	Hamburg	N	N	Bad	Y
7757	Natasha	Romanova	1983	67	St. Petersburg	N	N	Good	N
0323	Jean	Grey	1977	66	Annandale	N	N	Good	N
3980	Clark	Kent	1954	76	Krypton	Y	Y	Good	Y
3057	Victor	Von Doom	1943	74	Latveria	N	N	Bad	Y
0573	Stephen	Strange	1968	74	Philadelphia	N	N	Good	Y
7452	Thor	Odinson	-2287	78	Norway	N	Y	Good	Y
1437	Selina	Kyle	1998	67	Gotham	Y	N	Neutral	N
1883	Raven	Darkholme	1911	70	unknown	Y	N	Bad	N
5830	Kara	Zor-el	1961	67	Krypton	Y	Y	Good	Y



rows

columns

65670

3

0

1

2

view as



	0	1	2
5.107477	5.135881	60.479023	
5.113939	5.141432	61.419001	
5.117143	5.13772	82.774271	
5.118805	5.145063	62.552338	
5.119299	5.144294	66.799533	
5.11949	5.140815	77.870507	
5.120502	5.147892	64.326006	
5.121868	5.14889	61.743756	
5.121949	5.149292	64.493967	
5.123392	5.148504	69.140338	
5.124216	5.148921	69.449809	

Feature engineering

Sometimes you have to massage the data before it becomes useful in answering your question.

rows

65670

columns

3

view as



0 1 2

5.107477 5.135881 60.479023

5.113939 5.141432 61.419001

5.117143 5.13772 82.774271

5.118805 5.145063 62.552338

5.119299 5.144294 66.799533

5.11949 5.140815 77.870507

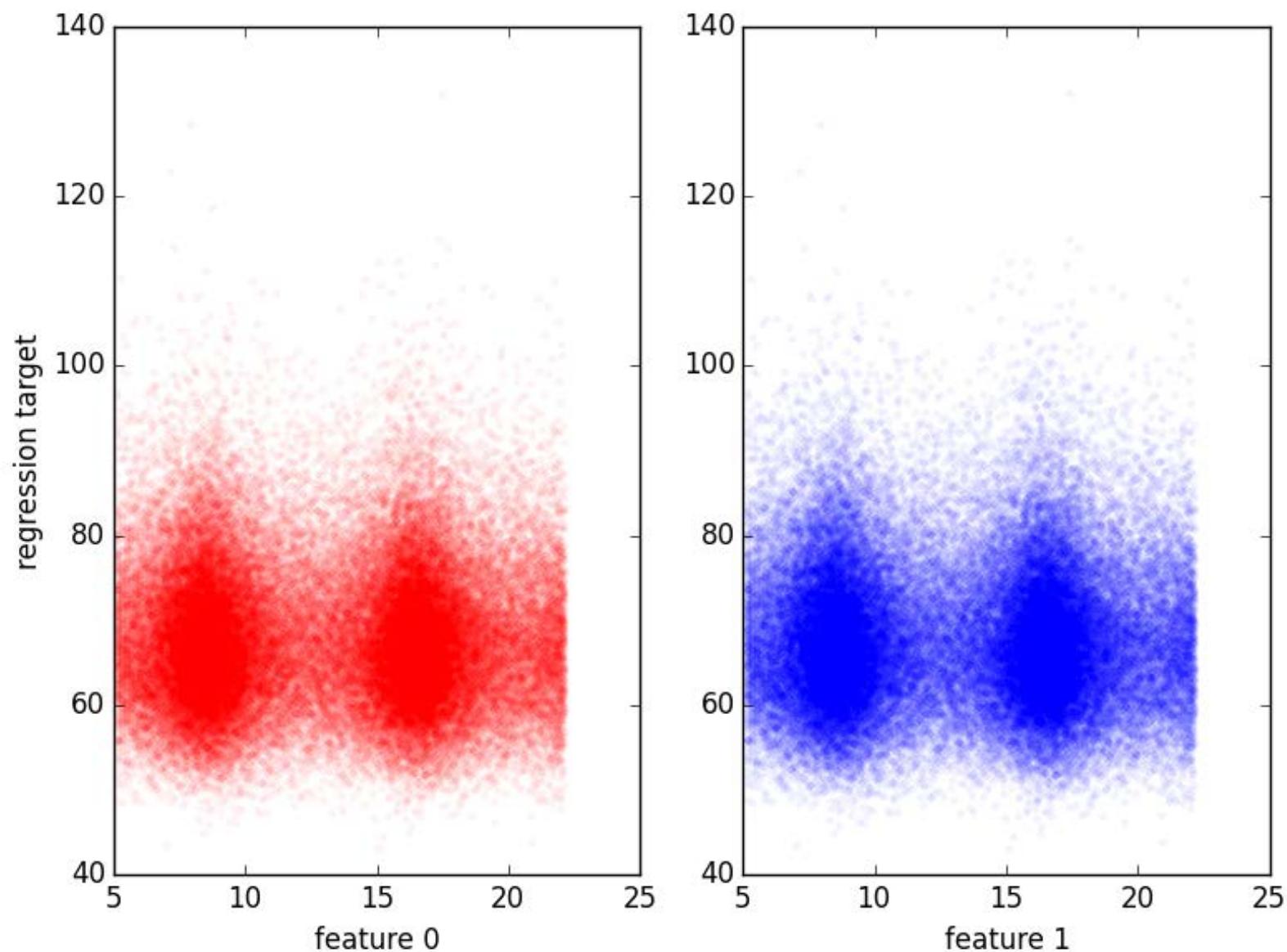
5.120502 5.147892 64.326006

5.121868 5.14889 61.743756

5.121949 5.149292 64.493967

5.123392 5.148504 69.140338

5.124216 5.148921 69.449809



rows

65670

columns

3

view as



0 1 2

5.107477 5.135881 60.479023

5.113939 5.141432 61.419001

5.117143 5.13772 82.774271

5.118805 5.145063 62.552338

5.119299 5.144294 66.799533

5.11949 5.140815 77.870507

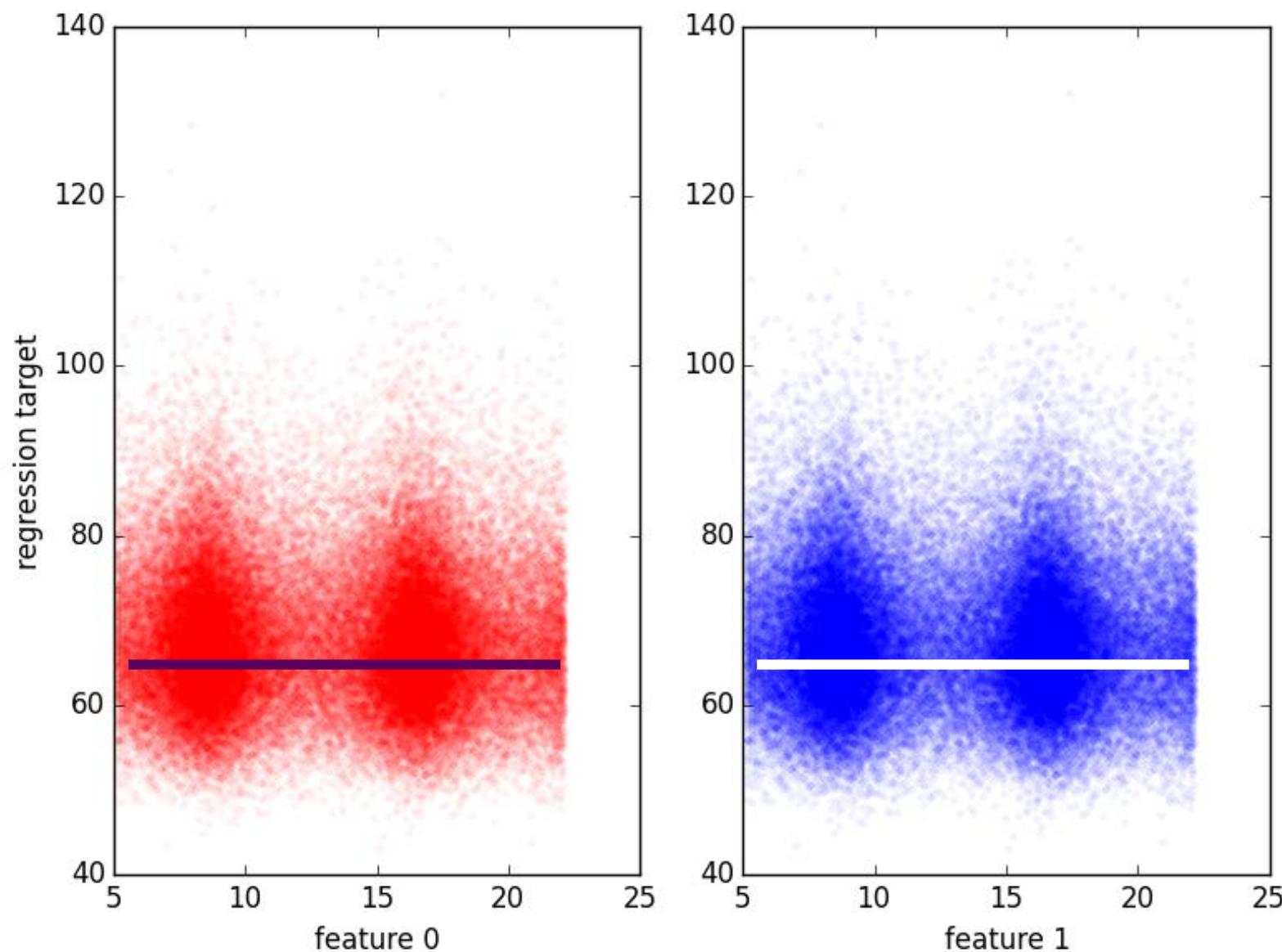
5.120502 5.147892 64.326006

5.121868 5.14889 61.743756

5.121949 5.149292 64.493967

5.123392 5.148504 69.140338

5.124216 5.148921 69.449809



rows

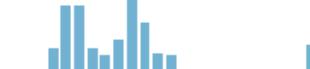
columns

65670

3

0 1 2

view as



5.107477 5.135881 60.479023

5.113939 5.141432 61.419001

5.117143 5.13772 82.774271

5.118805 5.145063 62.552338

5.119299 5.144294 66.799533

5.11949 5.140815 77.870507

5.120502 5.147892 64.326006

5.121868 5.14889 61.743756

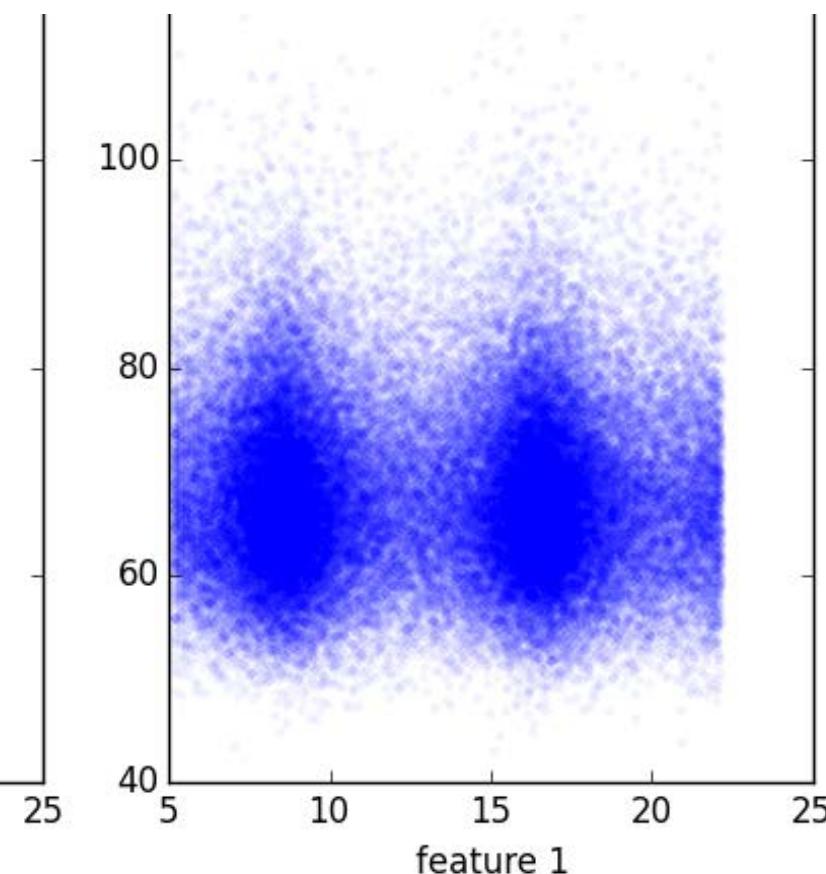
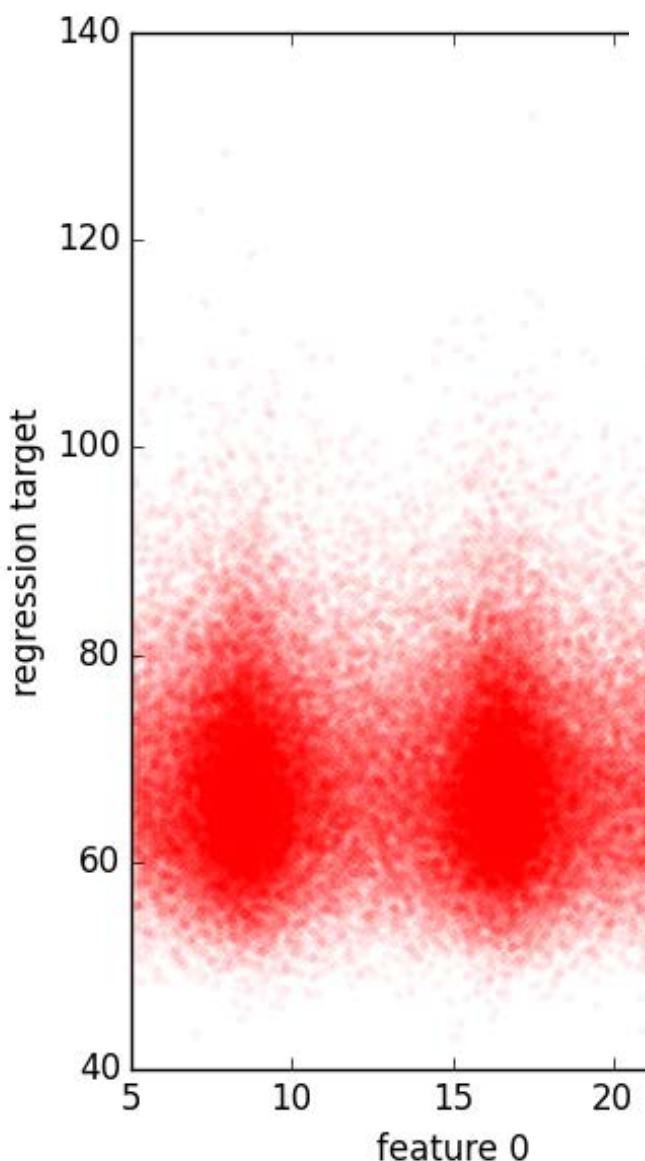
5.121949 5.149292 64.493967

5.123392 5.148504 69.140338

5.124216 5.148921 69.449809

Metrics

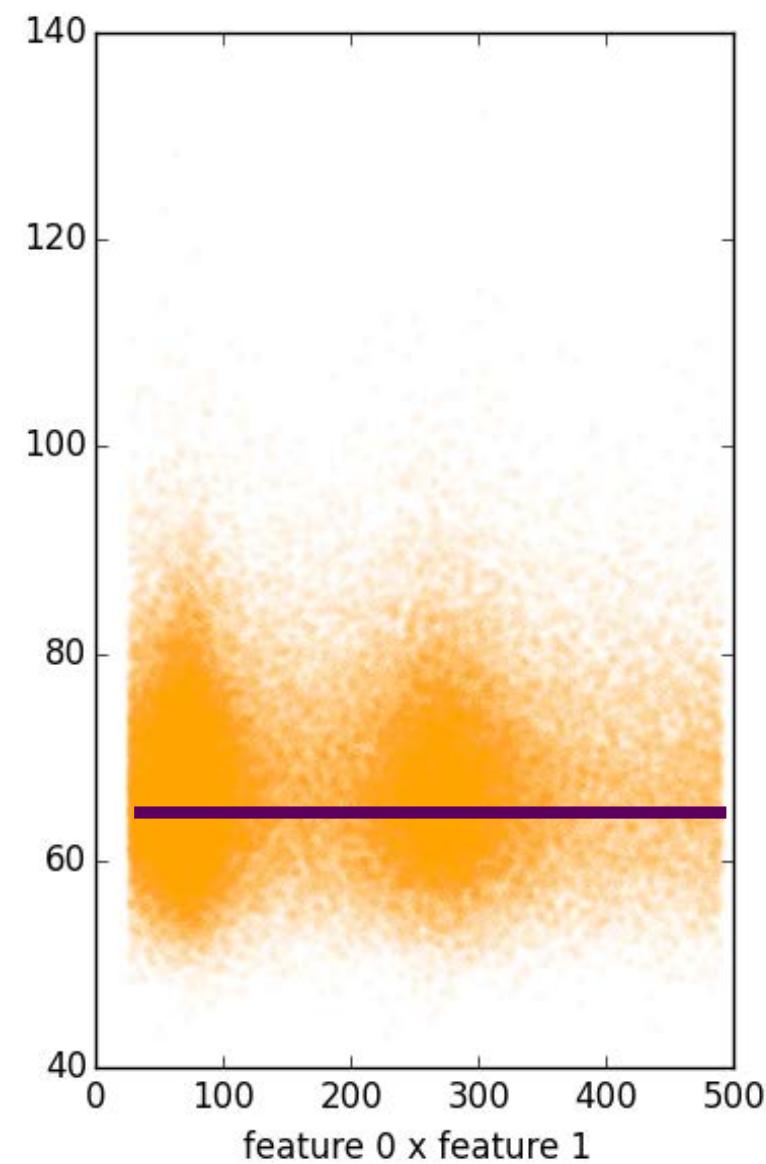
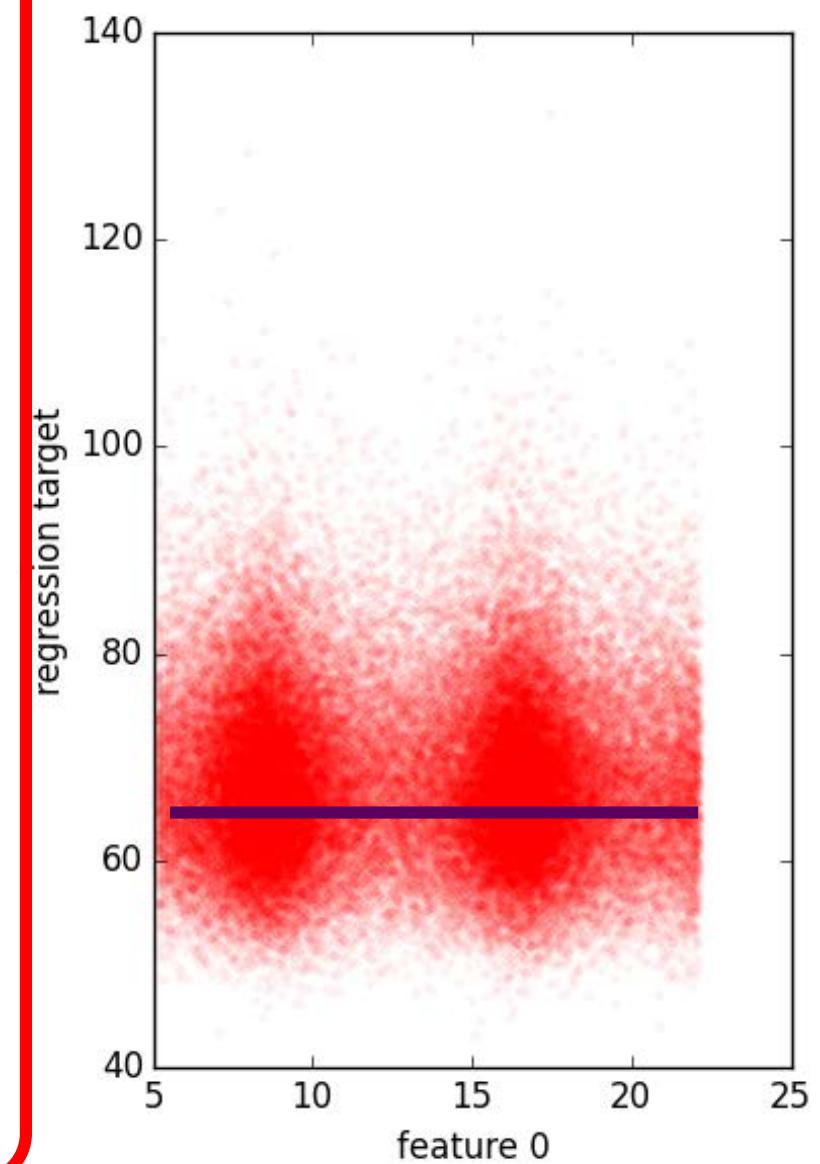
Mean Absolute Error	6.485434
Root Mean Squared Error	8.280206
Relative Absolute Error	0.991422
Relative Squared Error	0.983903
Coefficient of Determination	0.016097



columns

4

0	1	2	Multiply(1_0)
5.107477	5.135881	60.479023	26.231395
5.113939	5.141432	61.419001	26.292971
5.117143	5.13772	82.774271	26.290449
5.118805	5.145063	62.552338	26.336574
5.119299	5.144294	66.799533	26.335178
5.11949	5.140815	77.870507	26.318351
5.120502	5.147892	64.326006	26.359789
5.121868	5.14889	61.743756	26.371937
5.121949	5.149292	64.493967	26.374413
5.123392	5.148504	69.140338	26.3778
5.124216	5.148921	69.449809	26.384186
5.126409	5.154655	62.028089	26.42487



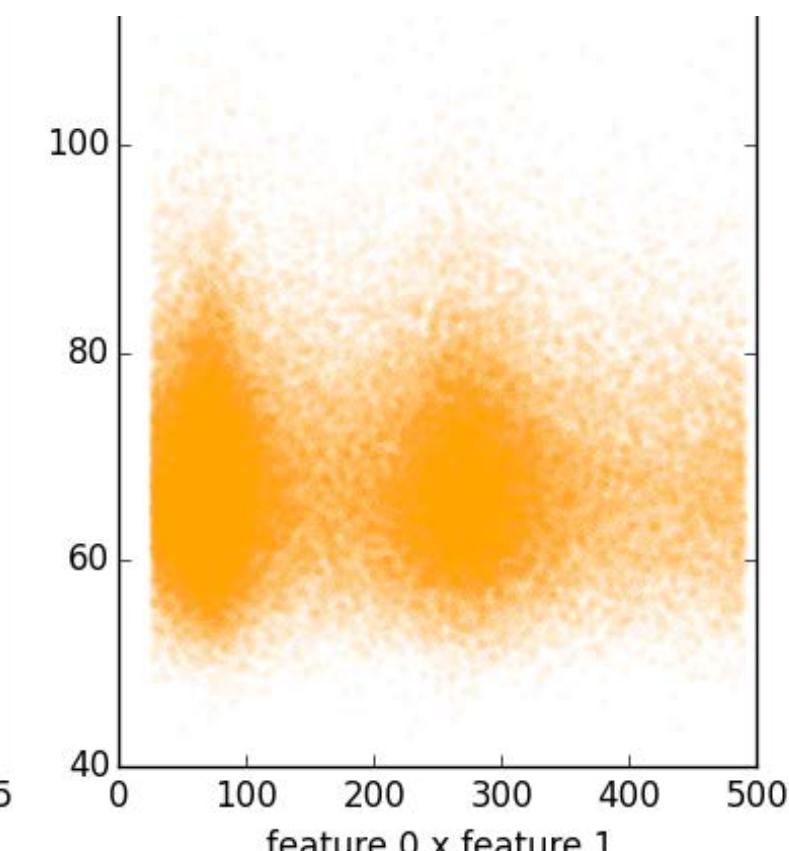
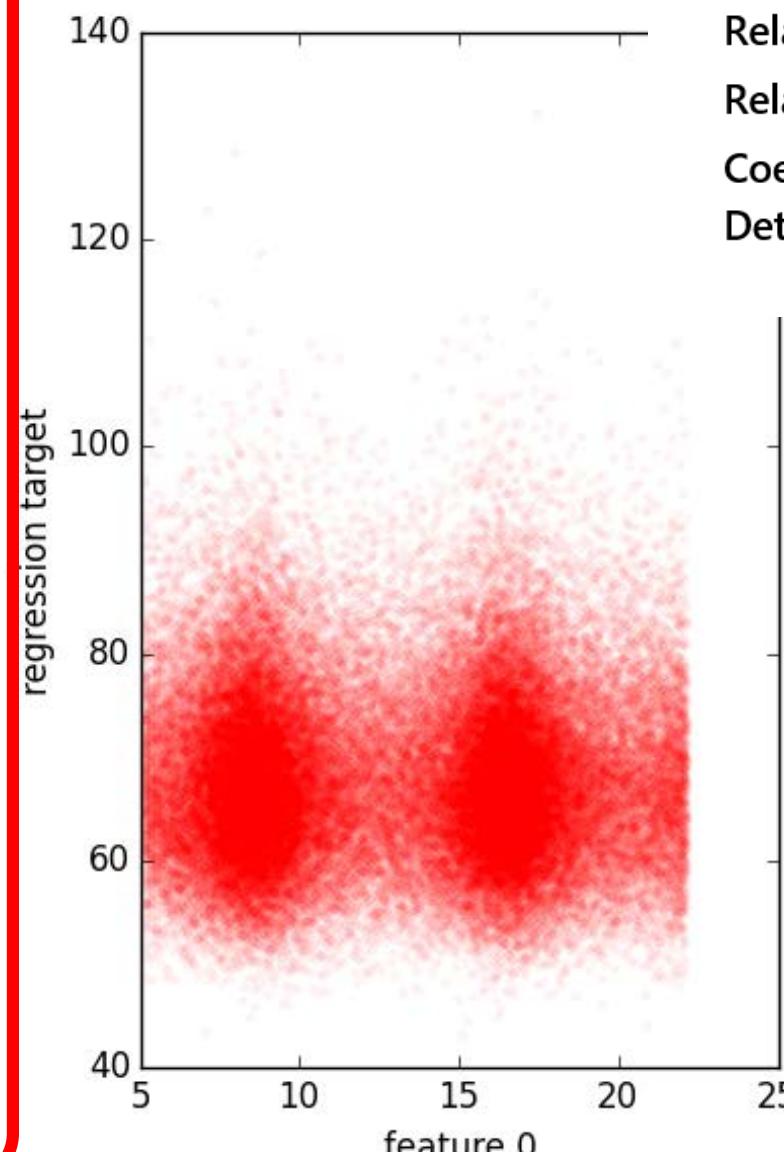
columns

4

0	1	2	Multiply(1_0)
5.107477	5.135881	60.479023	26.231395
5.113939	5.141432	61.419001	26.292971
5.117143	5.13772	82.774271	26.290449
5.118805	5.145063	62.552338	26.336574
5.119299	5.144294	66.799533	26.335178
5.11949	5.140815	77.870507	26.318351
5.120502	5.147892	64.326006	26.359789
5.121868	5.14889	61.743756	26.371937
5.121949	5.149292	64.493967	26.374413
5.123392	5.148504	69.140338	26.3778
5.124216	5.148921	69.449809	26.384186
5.126409	5.154655	62.028089	26.42487

Metrics

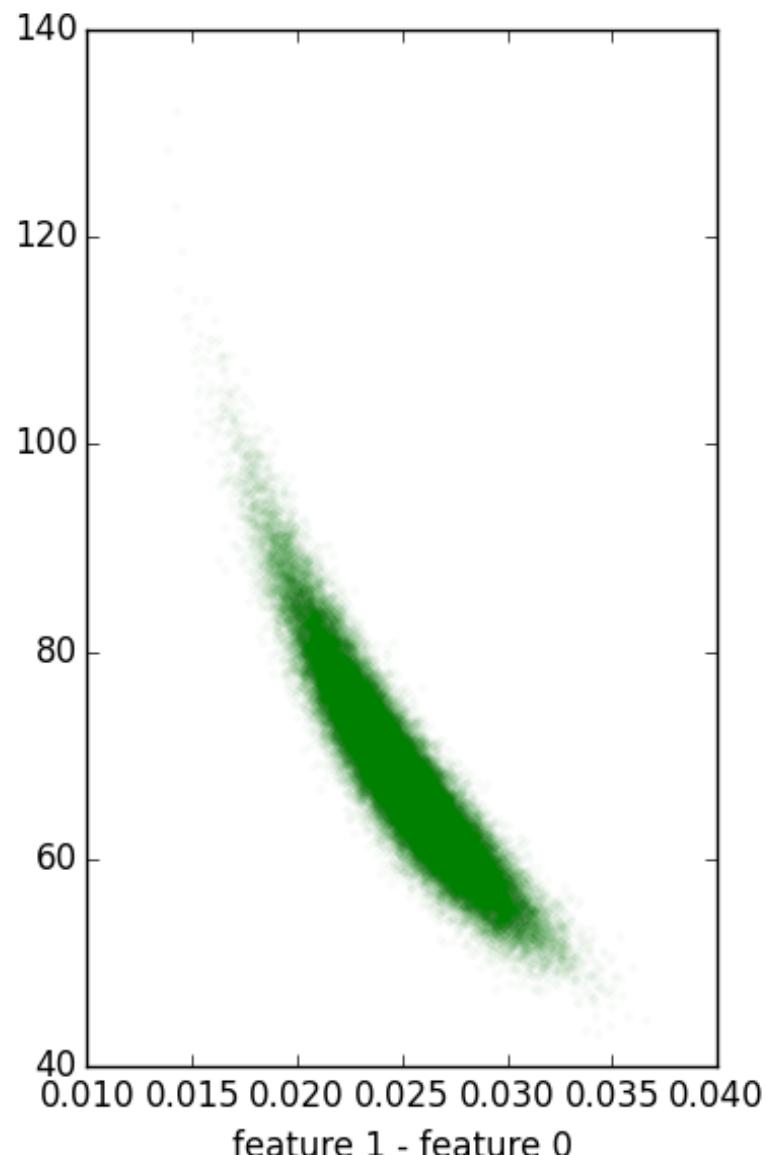
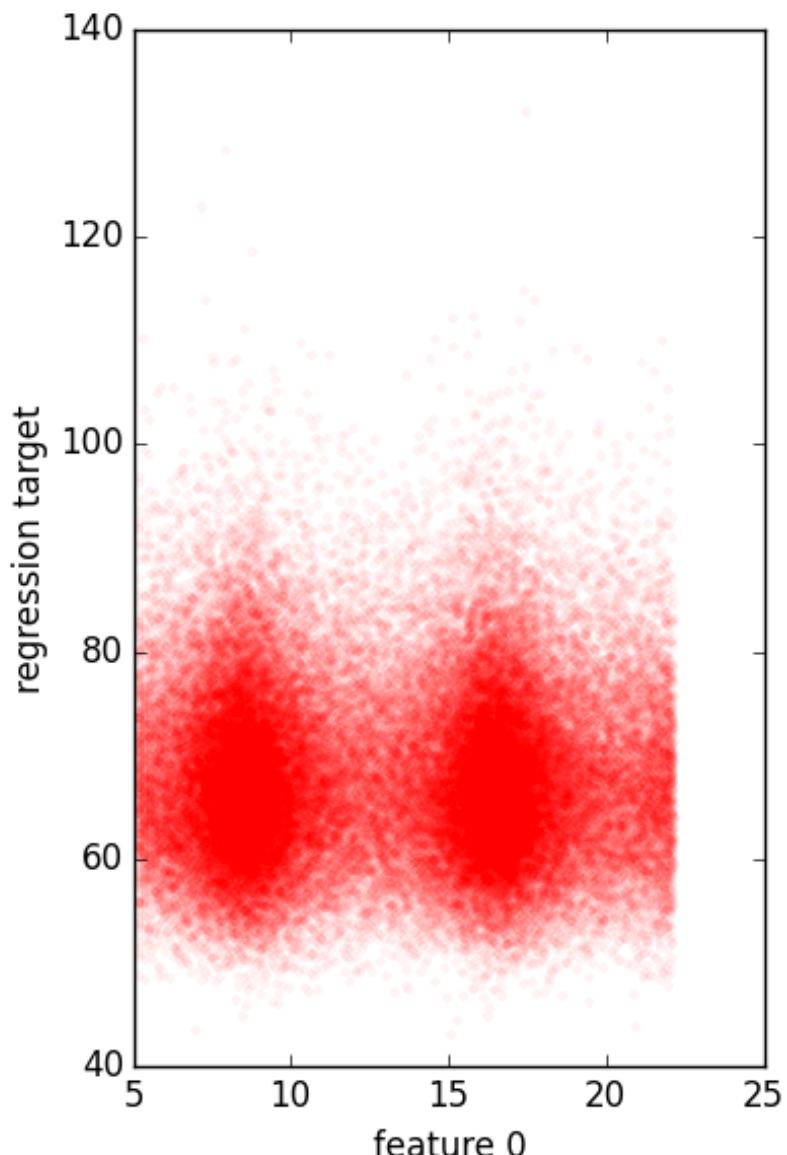
Mean Absolute Error	6.491614
Root Mean Squared Error	8.285875
Relative Absolute Error	0.992366
Relative Squared Error	0.98525
Coefficient of Determination	0.01475



columns

4

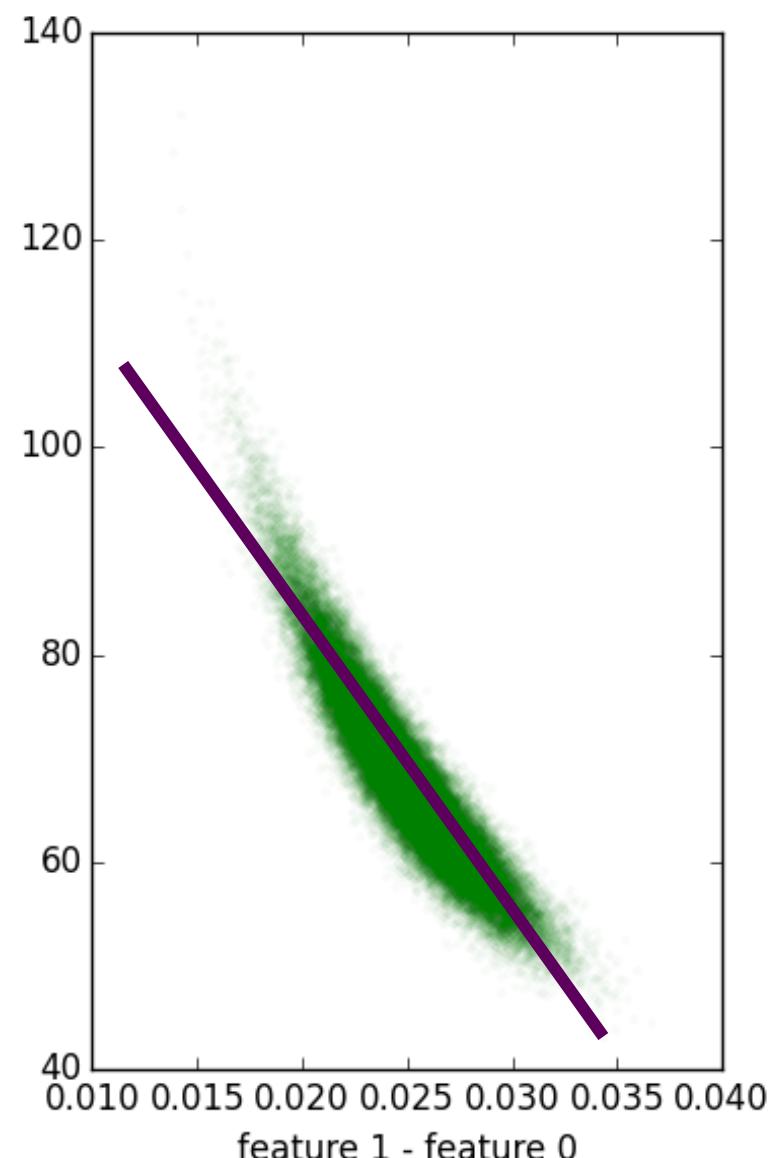
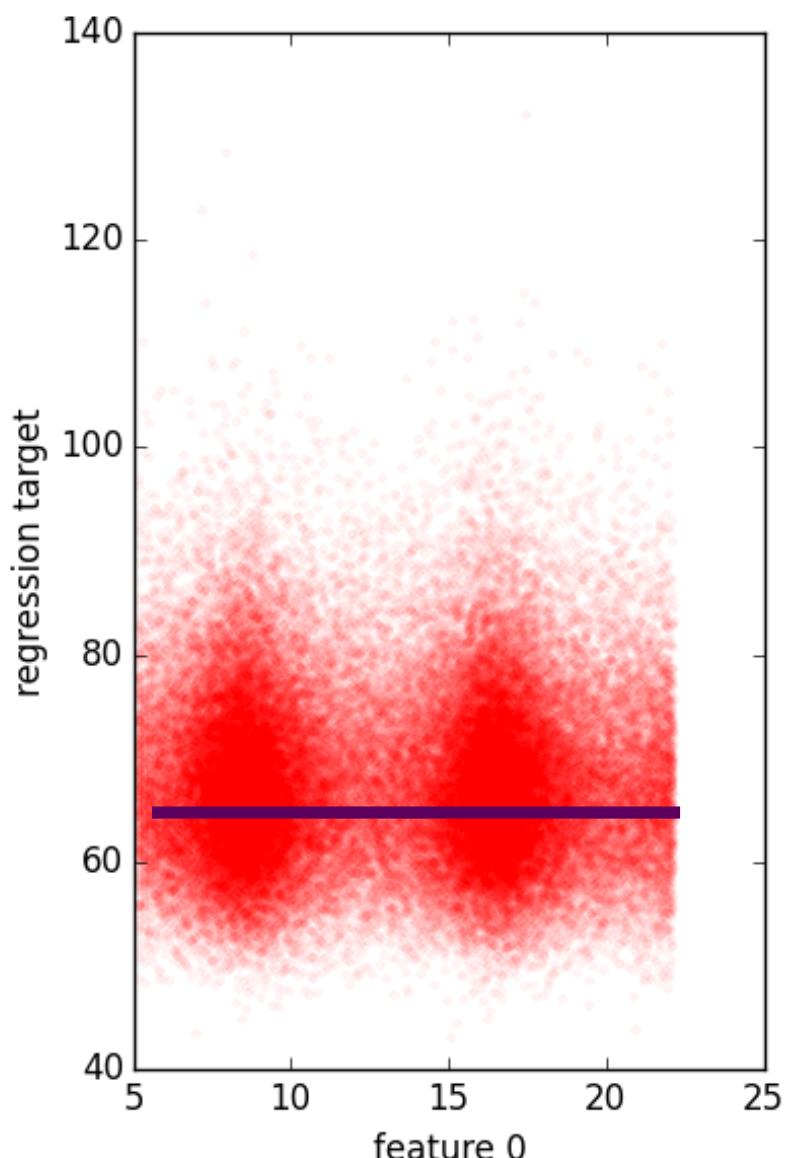
0	1	2	Subtract(1_0)
5.107477	5.135881	60.479023	0.028404
5.113939	5.141432	61.419001	0.027493
5.117143	5.13772	82.774271	0.020578
5.118805	5.145063	62.552338	0.026258
5.119299	5.144294	66.799533	0.024995
5.11949	5.140815	77.870507	0.021325
5.120502	5.147892	64.326006	0.02739
5.121868	5.14889	61.743756	0.027022
5.121949	5.149292	64.493967	0.027343
5.123392	5.148504	69.140338	0.025112
5.124216	5.148921	69.449809	0.024705
5.126409	5.154655	62.028089	0.028246



columns

4

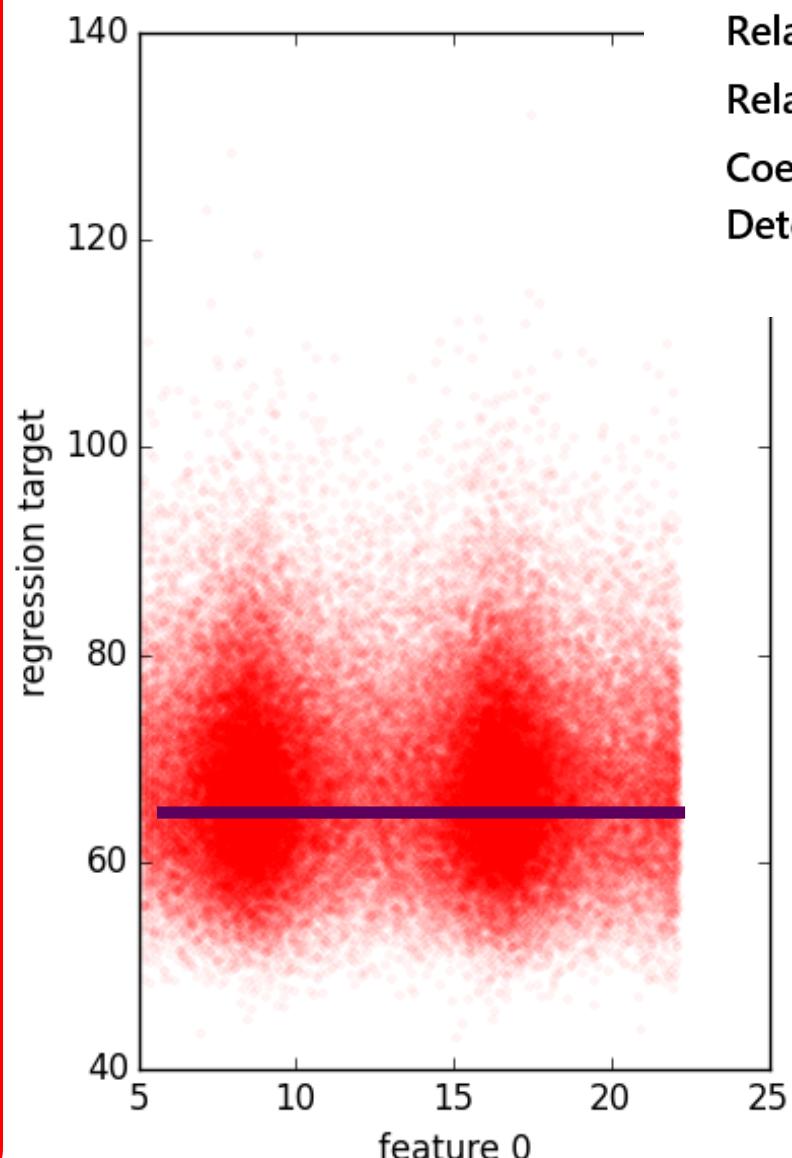
0	1	2	Subtract(1_0)
5.107477	5.135881	60.479023	0.028404
5.113939	5.141432	61.419001	0.027493
5.117143	5.13772	82.774271	0.020578
5.118805	5.145063	62.552338	0.026258
5.119299	5.144294	66.799533	0.024995
5.11949	5.140815	77.870507	0.021325
5.120502	5.147892	64.326006	0.02739
5.121868	5.14889	61.743756	0.027022
5.121949	5.149292	64.493967	0.027343
5.123392	5.148504	69.140338	0.025112
5.124216	5.148921	69.449809	0.024705
5.126409	5.154655	62.028089	0.028246



columns

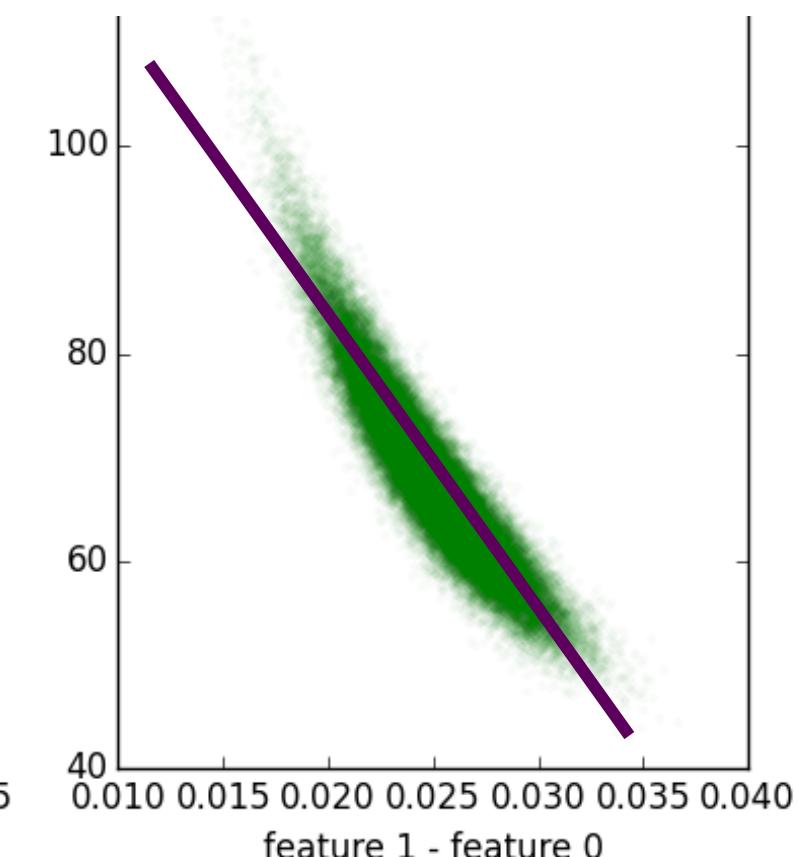
4

0	1	2	Subtract(1_0)
5.107477	5.135881	60.479023	0.028404
5.113939	5.141432	61.419001	0.027493
5.117143	5.13772	82.774271	0.020578
5.118805	5.145063	62.552338	0.026258
5.119299	5.144294	66.799533	0.024995
5.11949	5.140815	77.870507	0.021325
5.120502	5.147892	64.326006	0.02739
5.121868	5.14889	61.743756	0.027022
5.121949	5.149292	64.493967	0.027343
5.123392	5.148504	69.140338	0.025112
5.124216	5.148921	69.449809	0.024705
5.126409	5.154655	62.028089	0.028246



Metrics

Mean Absolute Error	2.243981
Root Mean Squared Error	2.834526
Relative Absolute Error	0.343035
Relative Squared Error	0.1153
Coefficient of Determination	0.8847



Other feature engineering tricks

Data-specific

- Images (SIFT)

- Text (TF-IDF)

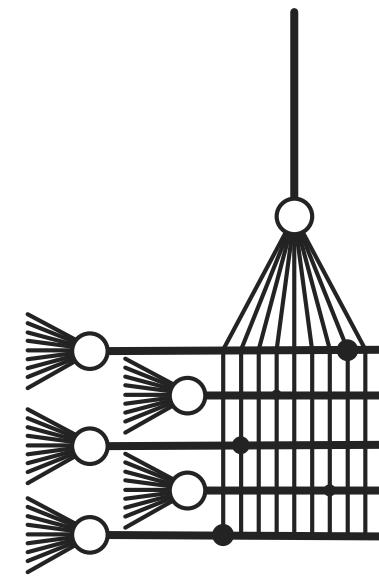
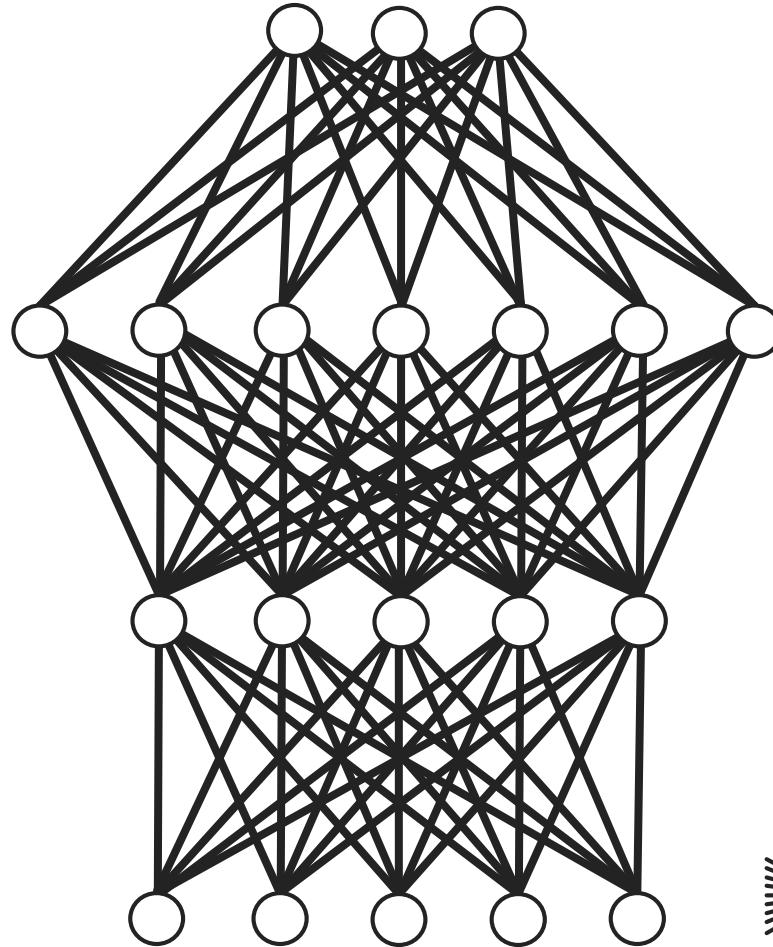
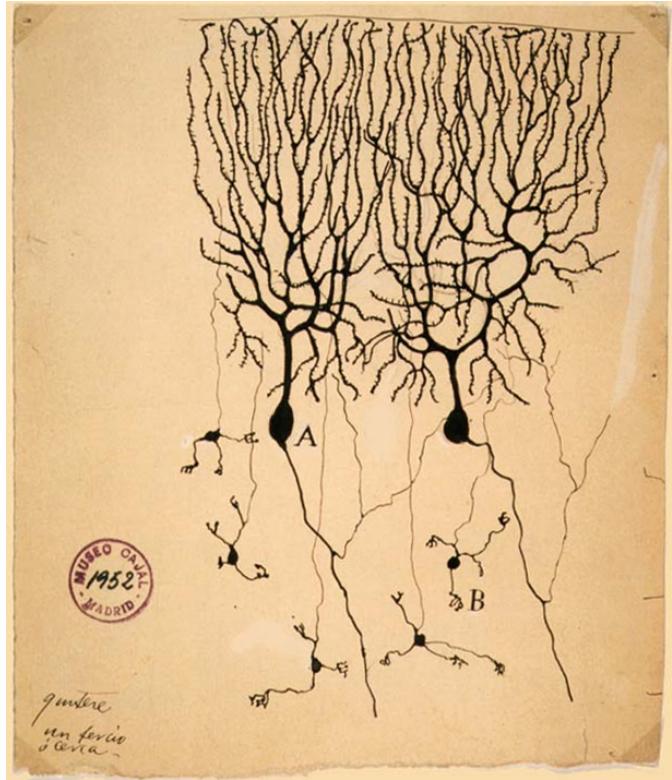
Domain specific

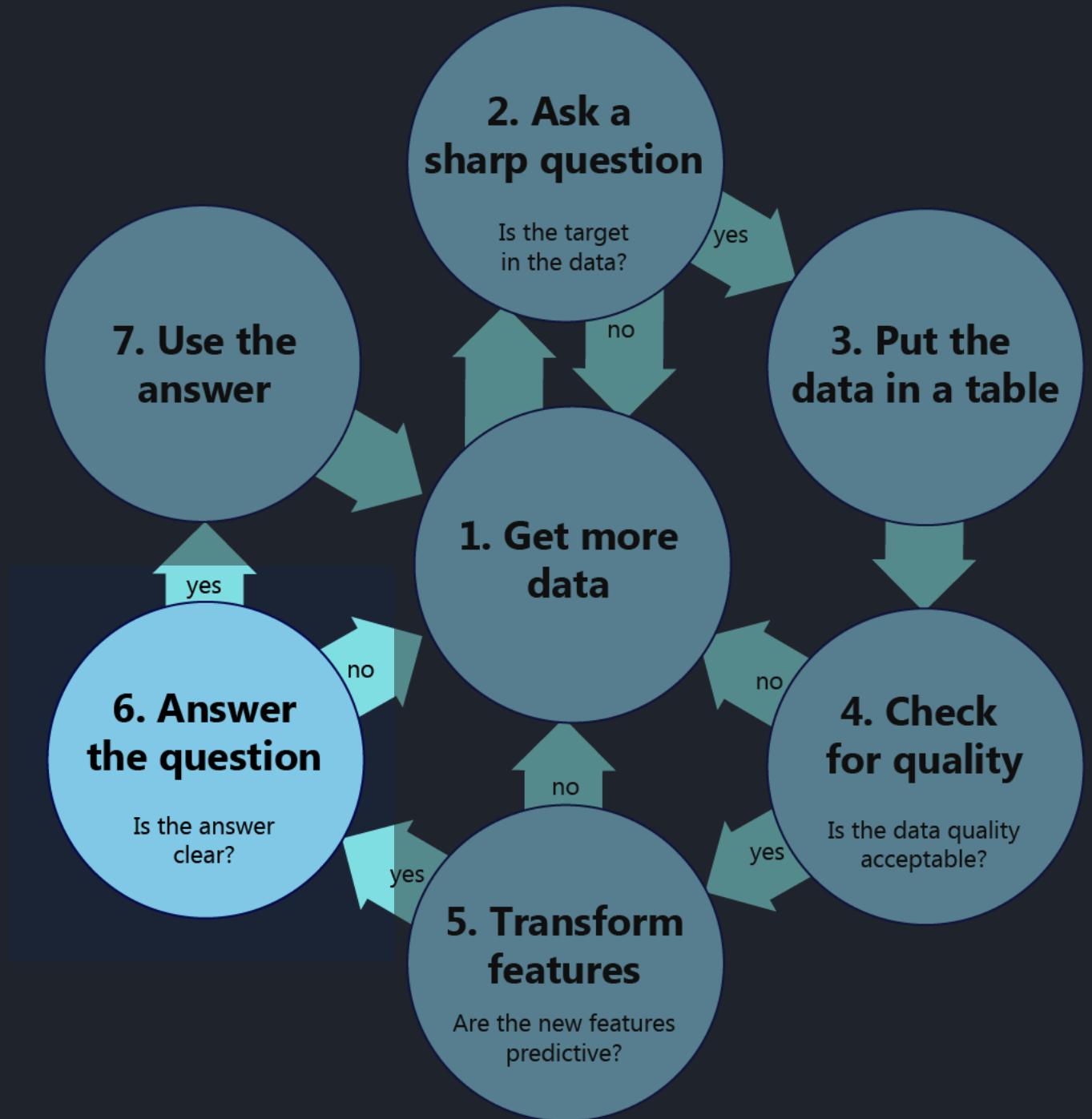
- Econometric, agricultural, sociological, ...

Deep learning

- Images, text, audio

Deep Learning Demystified





1. How much / how many?
2. Which category?
3. Which groups?
4. Is it weird?
5. Which action?

[algorithm]



How much / how many?

What will the temperature
be next Tuesday?

What will my fourth quarter
sales in Portugal be?

How many new followers
will I get next week?



Which category?

Is this an image of a cat or a dog?

Which aircraft is causing this radar signature?

What is the topic of this news article?

[classification]



Which groups?

Which shoppers have similar tastes in produce?

Which viewers like the same kind of movies?

What is a natural way to break these documents into five topic groups?



Is this weird?

Is this pressure reading unusual?

Is this internet message typical?

Is this combination of purchases
very different from what this
customer has made in the past?



Which action?

Should I raise or lower the temperature?

Should I vacuum the living room again or stay plugged in to my charging station?

Should I brake or accelerate in response to that yellow light?



Diamonds

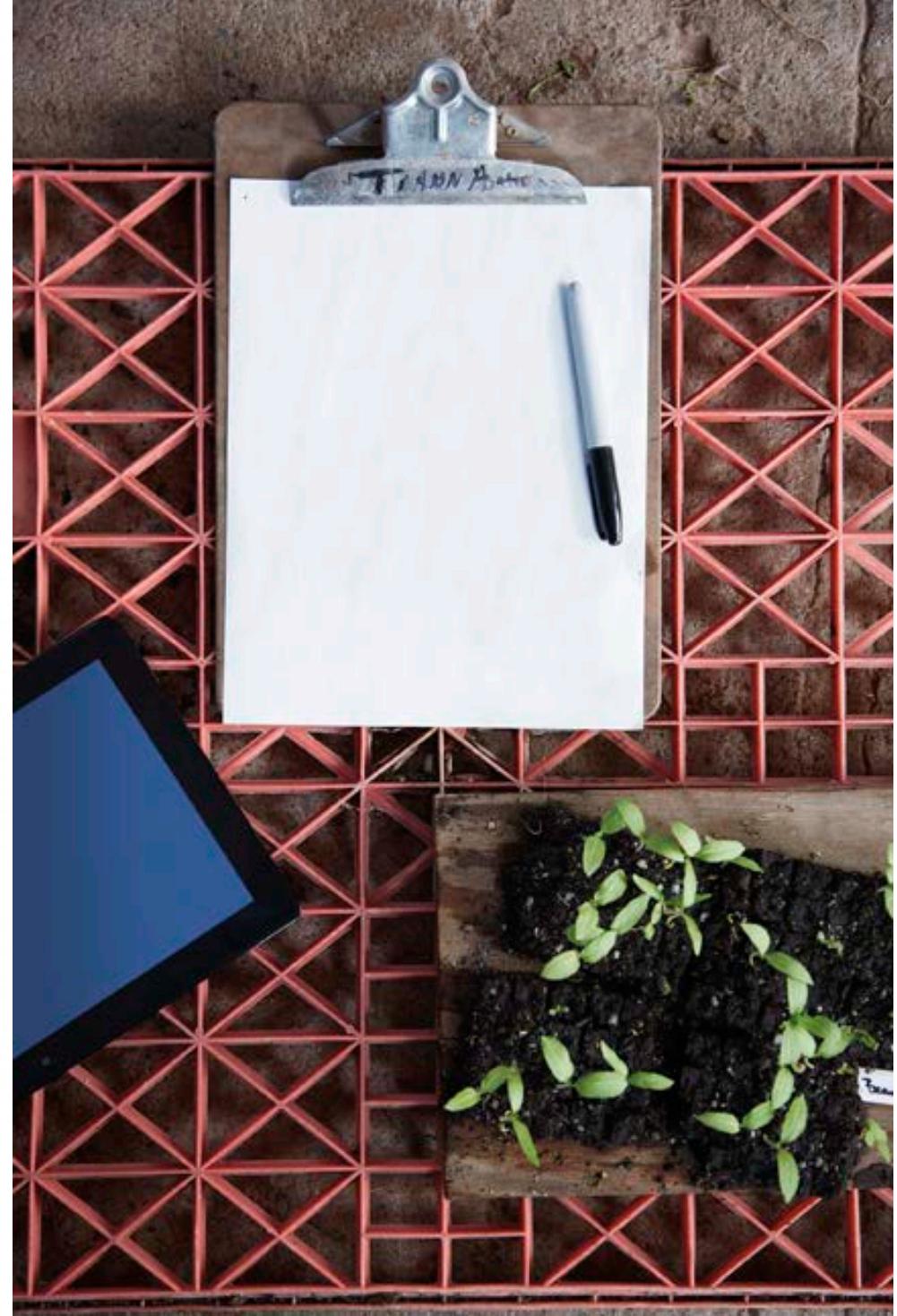


carats

1.01
.49
.31
1.51
.37
.73
1.53
.56
.41
.74
.63
.6
2.06
1.1
1.32
2.02
.34

price

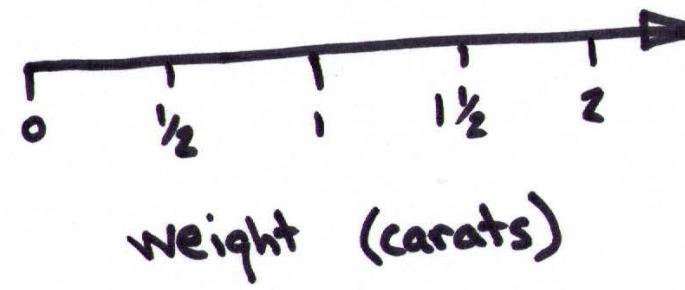
\$7,366
985
544
9,140
493
3,011
11,413
1,814
876
2,690
1,991
4,172
11,764
4,682
6,171
15,996
695



Diamonds



<u>carats</u>	<u>price</u>
1.01	\$7,366
.49	985
.31	544
1.51	9,140
.37	493
.73	3,011
1.53	11,413
.56	1,814
.41	876
.74	2,690
.63	1,991
.6	4,172
2.06	11,764
1.1	4,682
1.32	6,171
2.02	15,996
.34	695



Diamonds



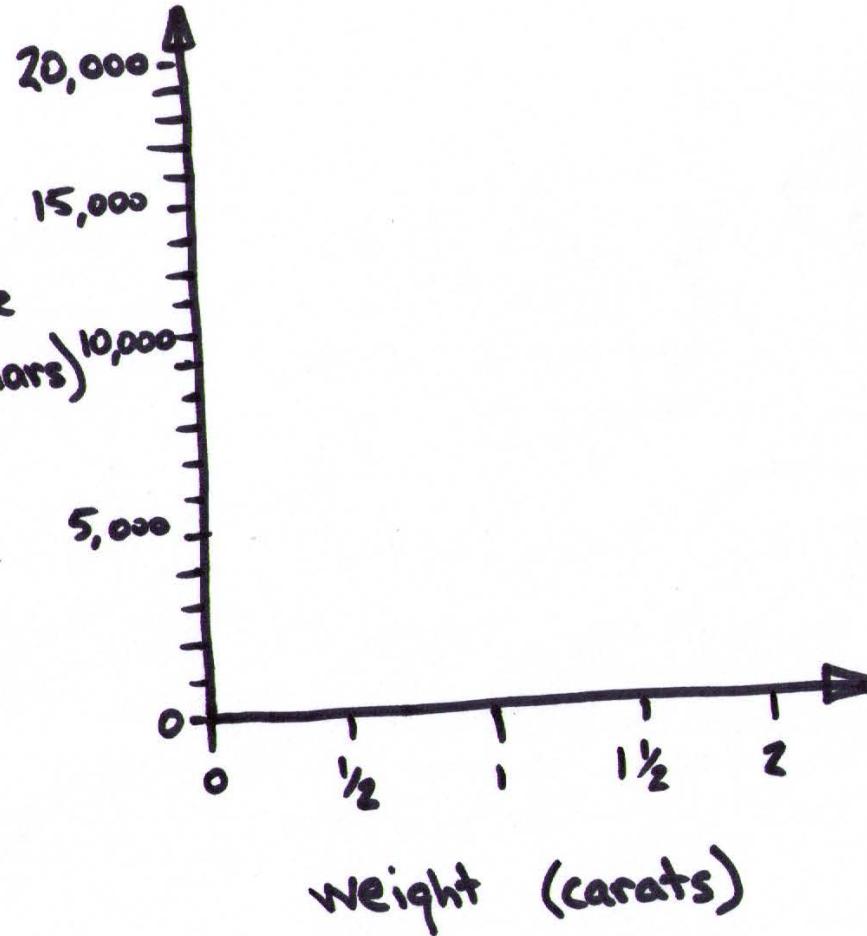
carats

1.01
.49
.31
1.51
.37
.73
1.53
.56
.41
.74
.63
.6
2.06
1.1
1.32
2.02
.34

price

\$7,366
985
544
9,140
493
3,011
11,413
1,814
876
2,690
1,991
4,172
11,764
4,682
6,171
15,996
695

price
(dollars)



Diamonds



carats

1.01

.49

.31

1.51

.37

.73

1.53

.56

.41

.74

.63

.6

2.06

1.1

1.32

2.02

.34

price

\$7,366

985

544

9,140

493

3,011

11,413

1,814

876

2,690

1,991

4,172

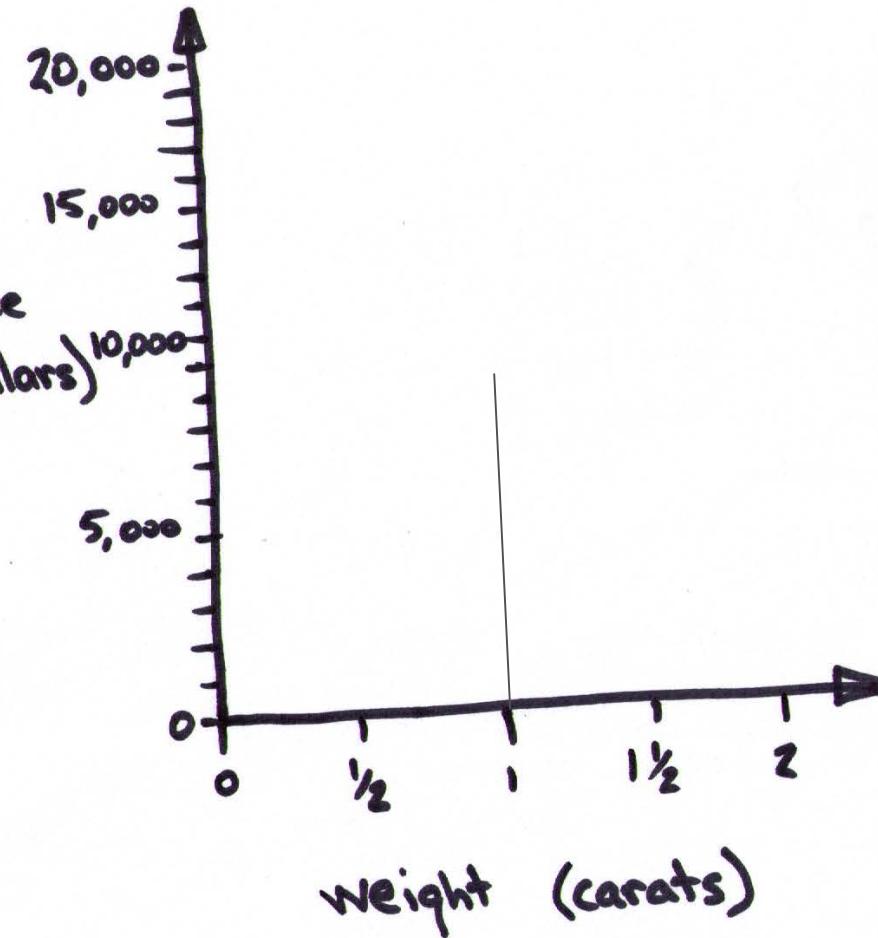
11,764

4,682

6,171

15,996

695



Diamonds



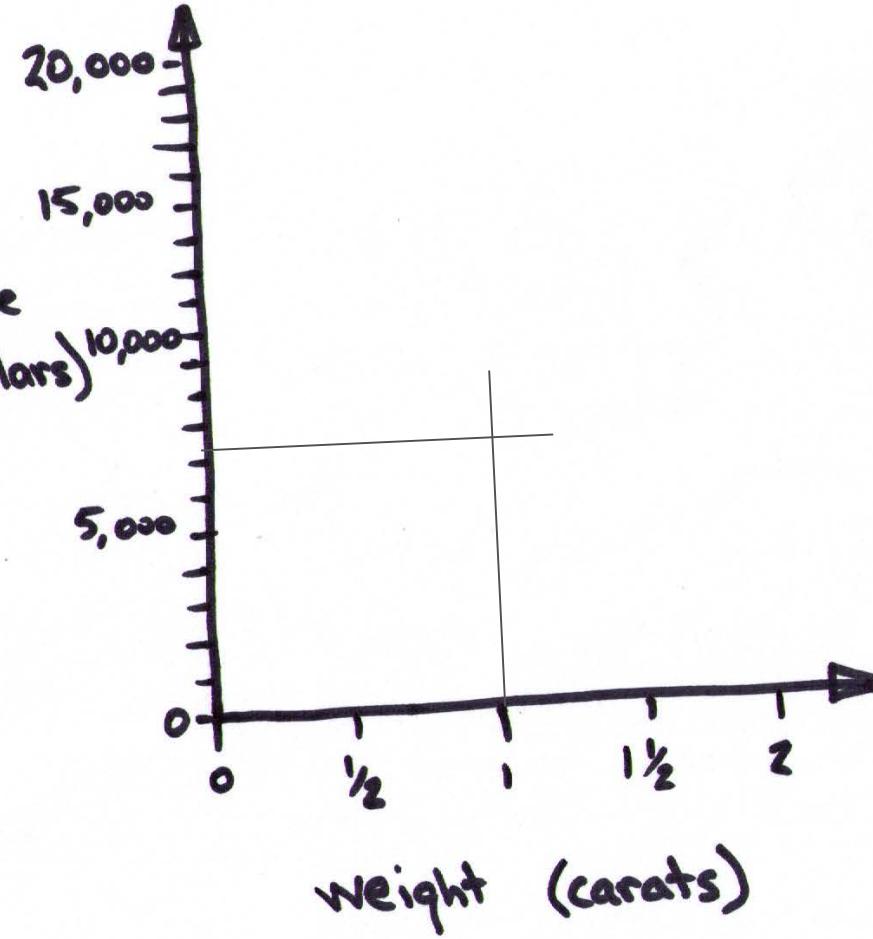
carats

1.01
.49
.31
1.51
.37
.73
1.53
.56
.41
.74
.63
.6
2.06
1.1
1.32
2.02
.34

price

\$7,366
985
544
9,140
493
3,011
11,413
1,814
876
2,690
1,991
4,172
11,764
4,682
6,171
15,996
695

price
(dollars)



Diamonds

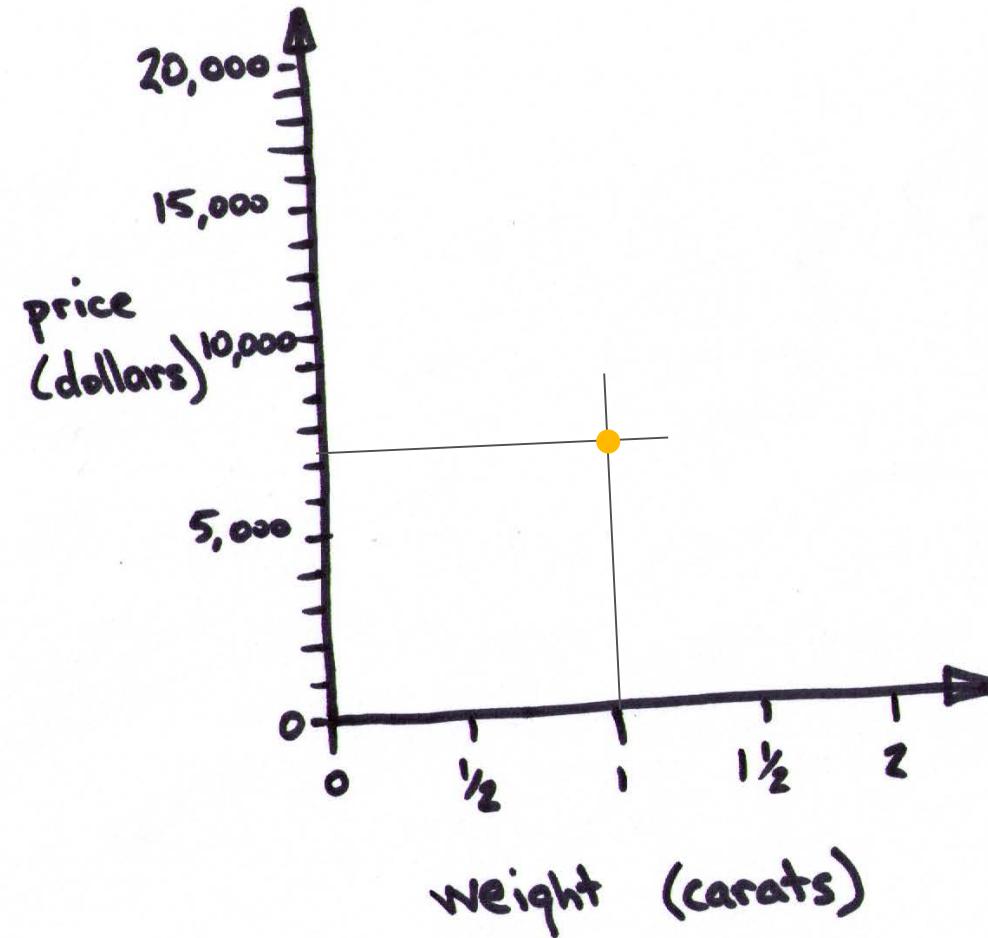


carats

1.01
.49
.31
1.51
.37
.73
1.53
.56
.41
.74
.63
.6
2.06
1.1
1.32
2.02
.34

price

\$7,366
985
544
9,140
493
3,011
11,413
1,814
876
2,690
1,991
4,172
11,764
4,682
6,171
15,996
695



Diamonds

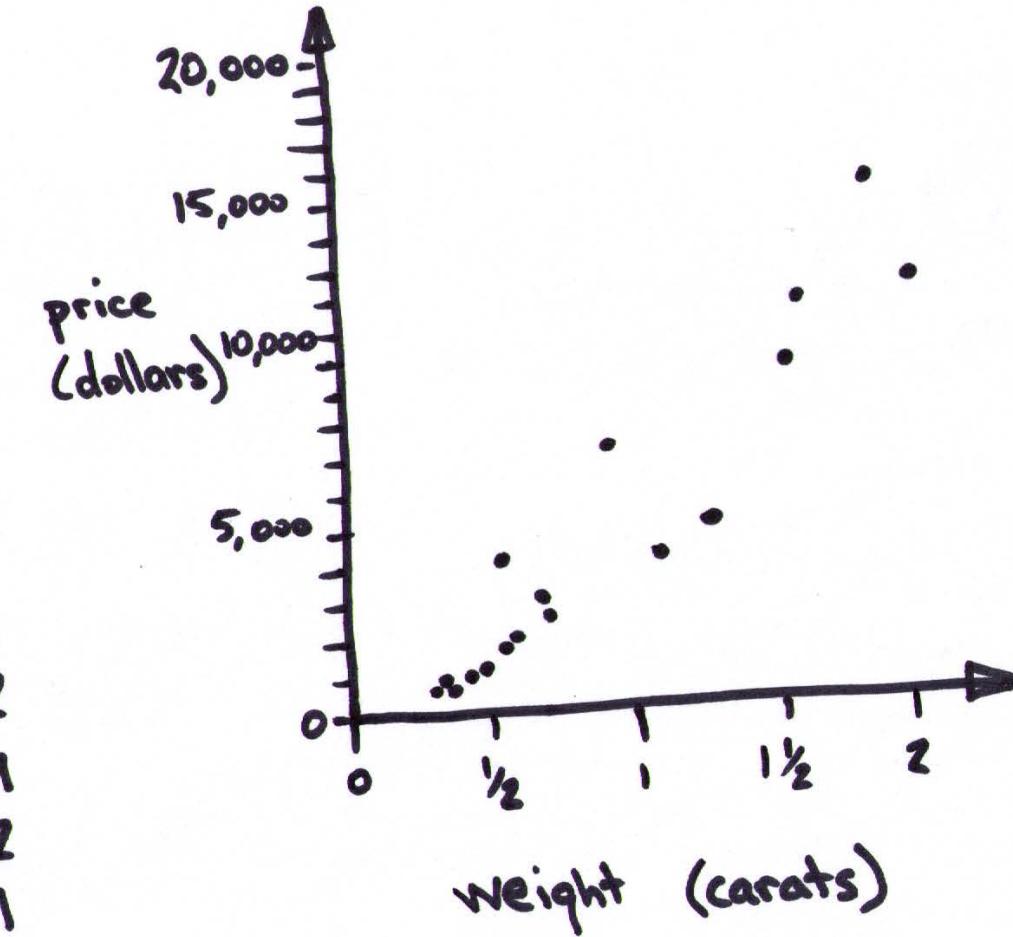


carats

1.01
.49
.31
1.51
.37
.73
1.53
.56
.41
.74
.63
.6
2.06
1.1
1.32
2.02
.34

price

\$7,366
985
544
9,140
493
3,011
11,413
1,814
876
2,690
1,991
4,172
11,764
4,682
6,171
15,996
695



Diamonds

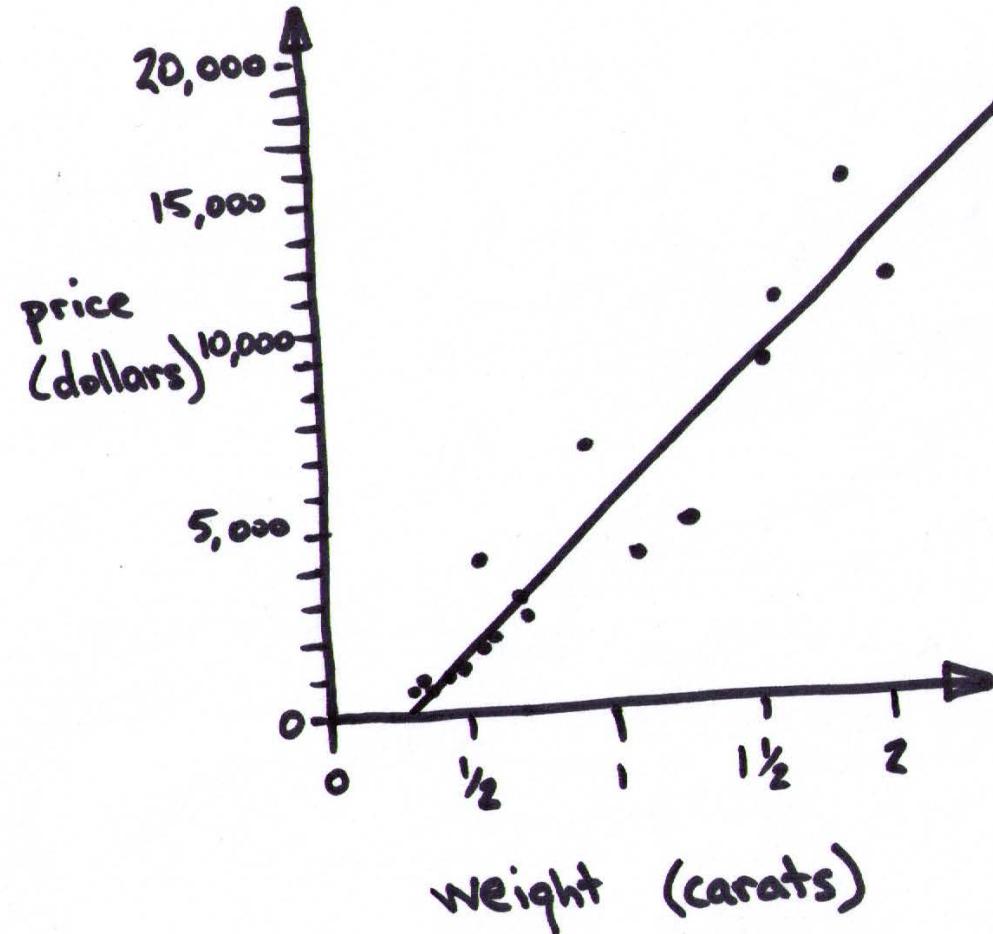


carats

1.01
.49
.31
1.51
.37
.73
1.53
.56
.41
.74
.63
.6
2.06
1.1
1.32
2.02
.34

price

\$7,366
985
544
9,140
493
3,011
11,413
1,814
876
2,690
1,991
4,172
11,764
4,682
6,171
15,996
695



Diamonds

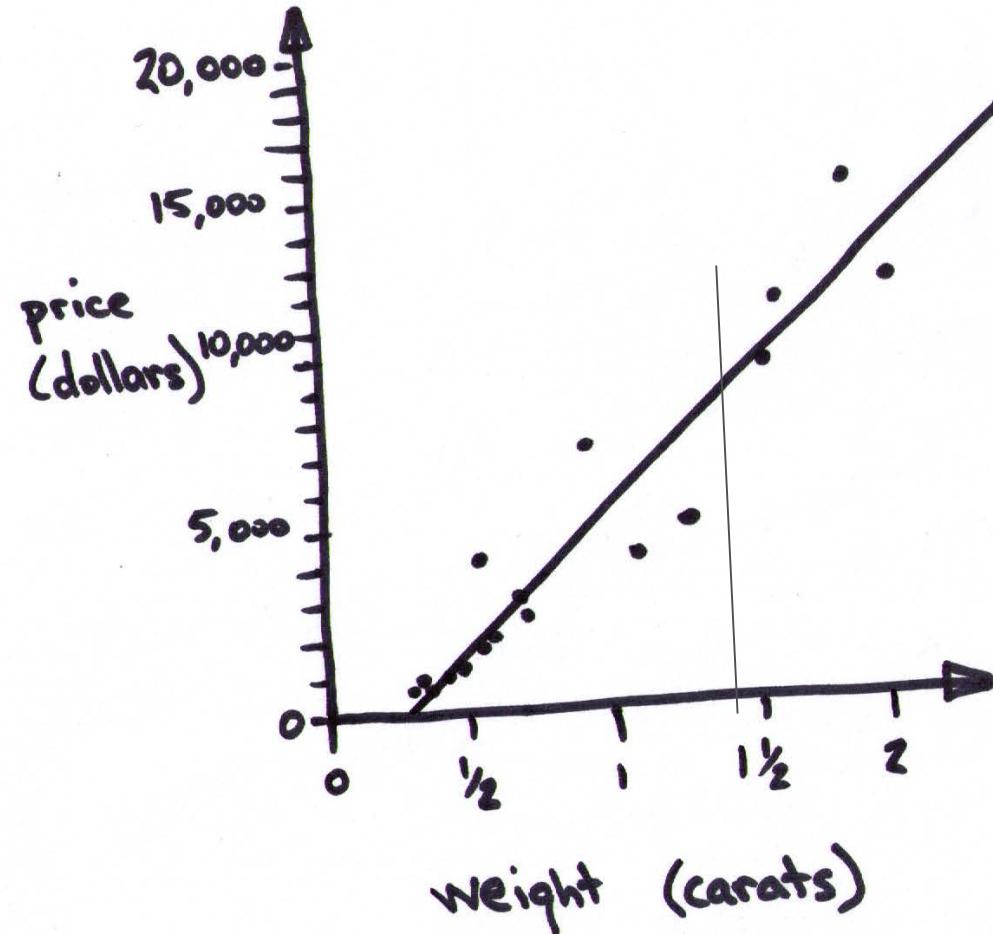


carats

1.01
.49
.31
1.51
.37
.73
1.53
.56
.41
.74
.63
.6
2.06
1.1
1.32
2.02
.34

price

\$7,366
985
544
9,140
493
3,011
11,413
1,814
876
2,690
1,991
4,172
11,764
4,682
6,171
15,996
695



Diamonds



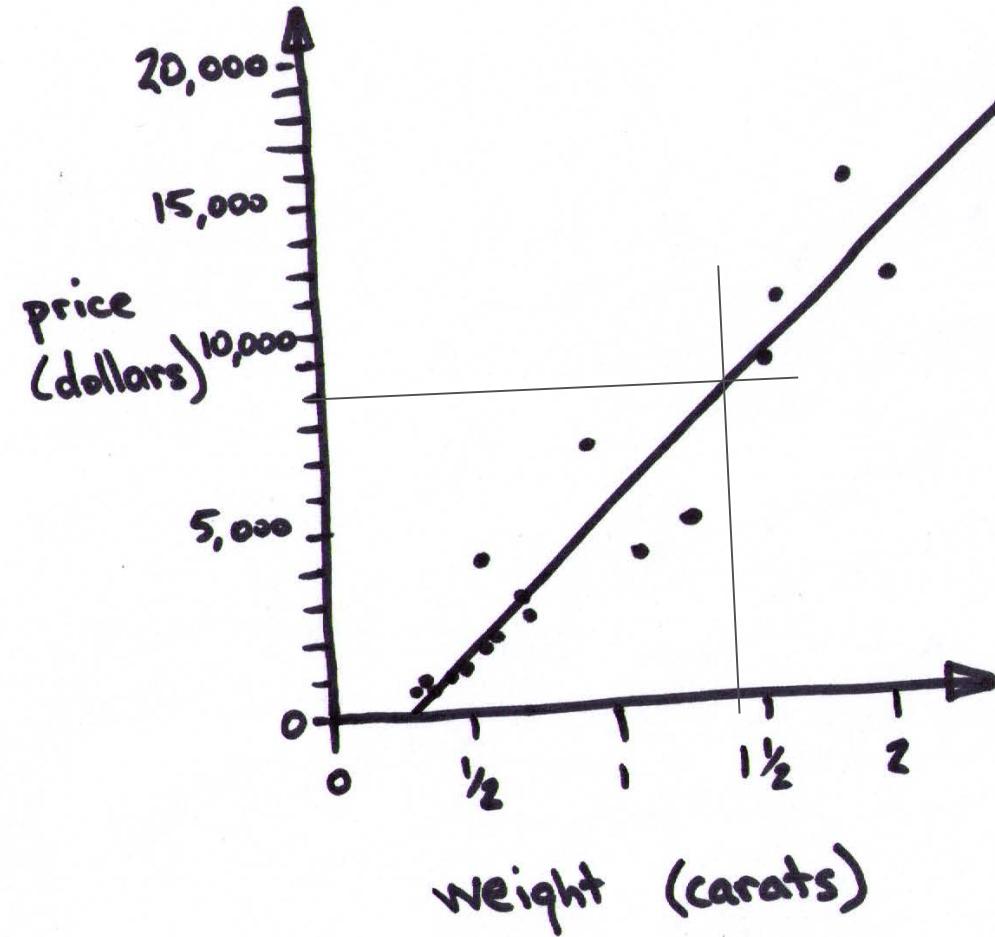
[prediction]

carats

1.01
.49
.31
1.51
.37
.73
1.53
.56
.41
.74
.63
.6
2.06
1.1
1.32
2.02
.34

price

\$7,366
985
544
9,140
493
3,011
11,413
1,814
876
2,690
1,991
4,172
11,764
4,682
6,171
15,996
695



Diamonds

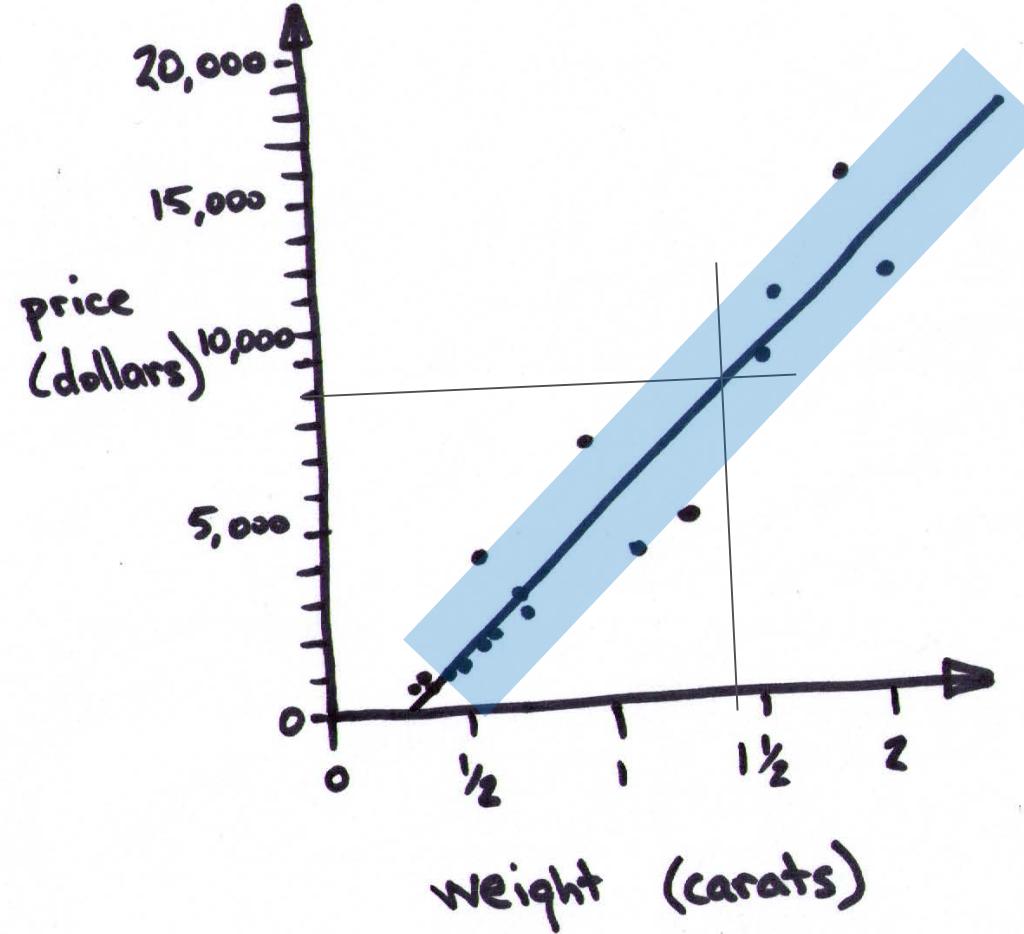


carats

1.01
.49
.31
1.51
.37
.73
1.53
.56
.41
.74
.63
.6
2.06
1.1
1.32
2.02
.34

price

\$7,366
985
544
9,140
493
3,011
11,413
1,814
876
2,690
1,991
4,172
11,764
4,682
6,171
15,996
695



Diamonds

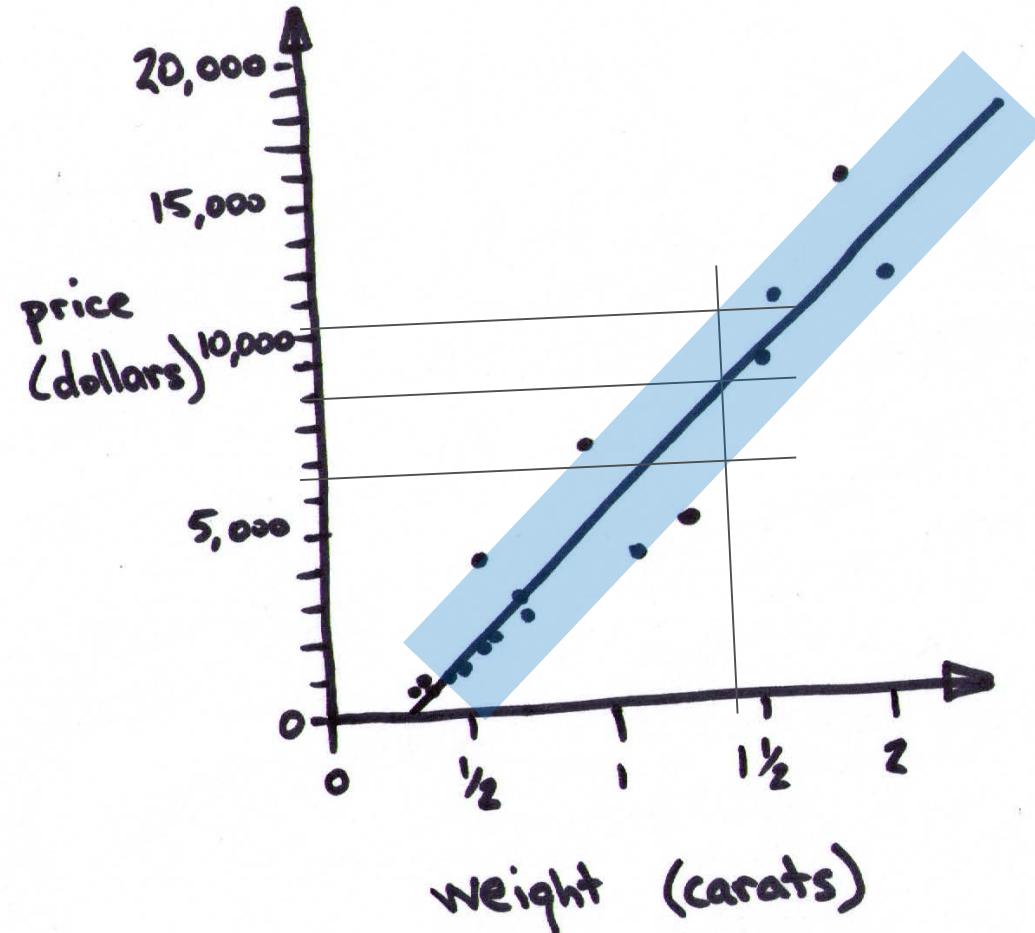


carats

1.01
.49
.31
1.51
.37
.73
1.53
.56
.41
.74
.63
.6
2.06
1.1
1.32
2.02
.34

price

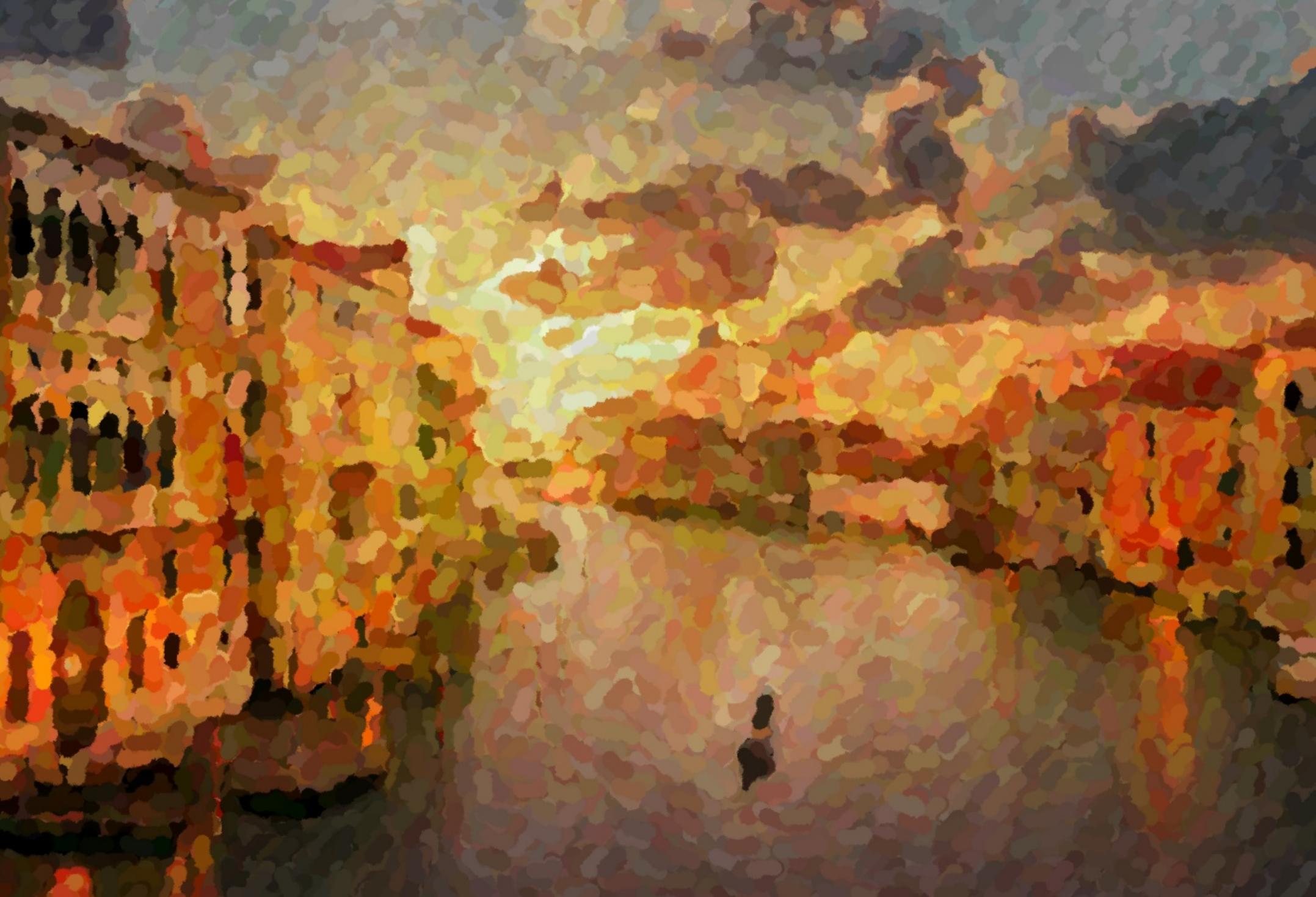
\$7,366
985
544
9,140
493
3,011
11,413
1,814
876
2,690
1,991
4,172
11,764
4,682
6,171
15,996
695



Not enough data

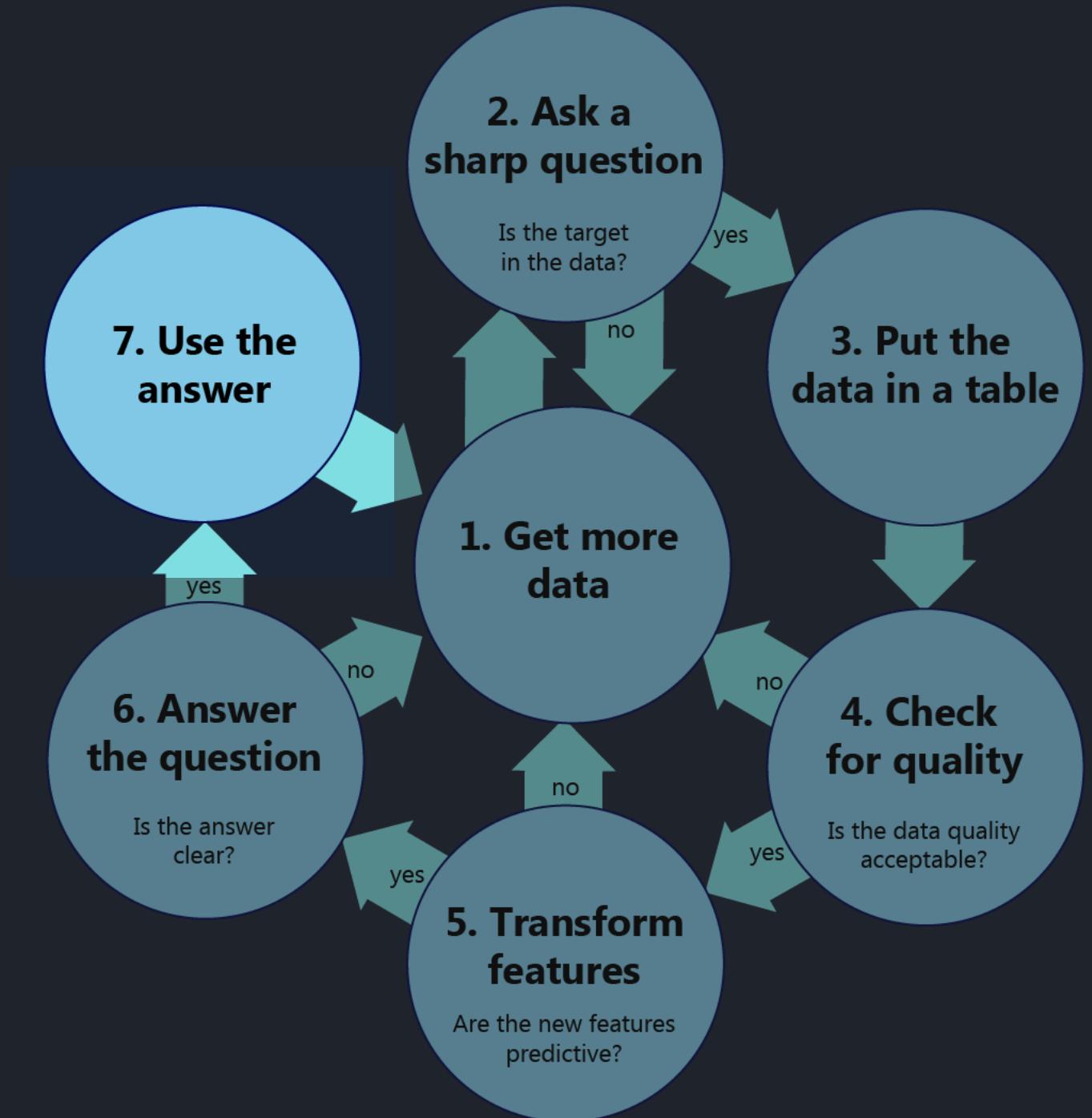


Barely enough data



Enough data





Ways to use your answer

Make a web service (Azure Machine Learning)

Make a decision

Set a price

Publish your code on GitHub

Write a PDF showing your results

Build a dash board (Power BI)

Gap 1

Nearly all machine learning algorithms assume that the world does not change.



Gap 2

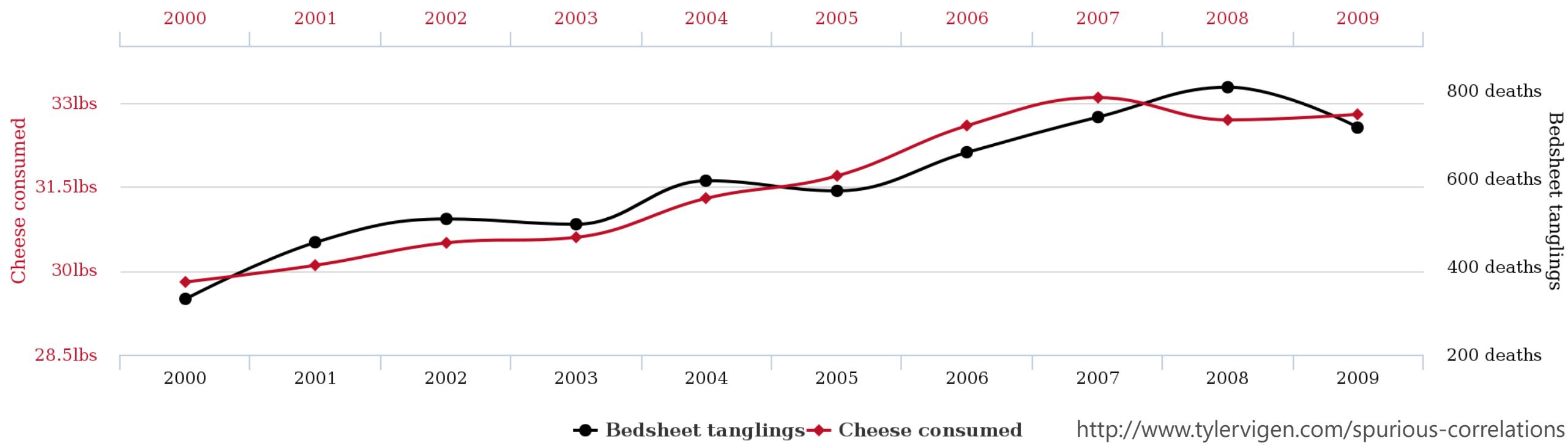
Most machine learning algorithms take a lot of examples to learn.



Gap 3

Machine learning can't tell what caused what.

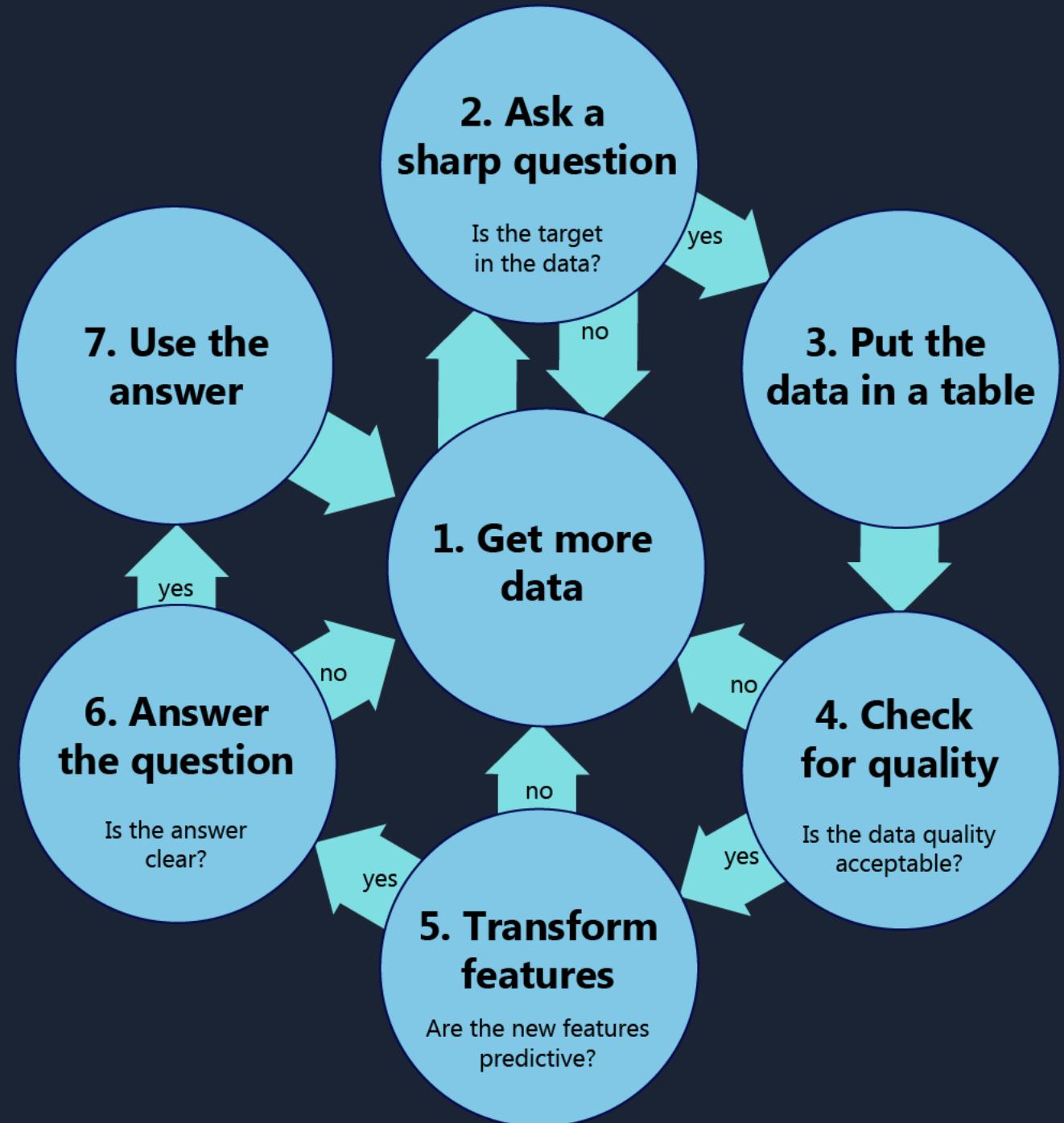
Per capita cheese consumption
correlates with
Number of people who died by becoming tangled in their bedsheets



Human insight and judgment close the gap

We're good at making reasonable guesses without enough information





Resources

1. Get more data.
2. Ask a sharp question.
3. Put the data in a table.
4. Check for quality.
5. Transform features.
6. Answer the question.

7. Use the answer.

Presentations

- [Microsoft Data Science Process](#)
[Asking a question](#)
- [Methods for handling missing values](#)
[Feature engineering example](#)
[Turn your data into a picture](#)
[Questions machine learning can answer](#)
[Algorithms for business use cases](#)
[Machine learning algorithm cheat sheet](#)
[Choosing a machine learning algorithm](#)
- [Cortana Intelligence Gallery](#)
- Data Science for Absolutely Everyone ([slides](#), [pdf](#), [video](#))
Data Science 101 ([slides](#), [pdf](#), [video](#))
The Other Stuff ([pdf](#), [video](#))
Demystifying neural networks ([slides](#), [pdf](#), [video](#))
How Convolutional Neural Networks work ([video](#), [blog](#), [slides](#), [pdf](#))

Thanks!

Questions? Want to chat about data?

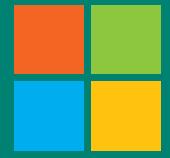
[LinkedIn](#) Brandon Rohrer

[@ brohrer](#)

[brohrer.github.io](#)

brohrer@microsoft.com

Special thanks to Diane Rohrer for image and layout design.



Microsoft