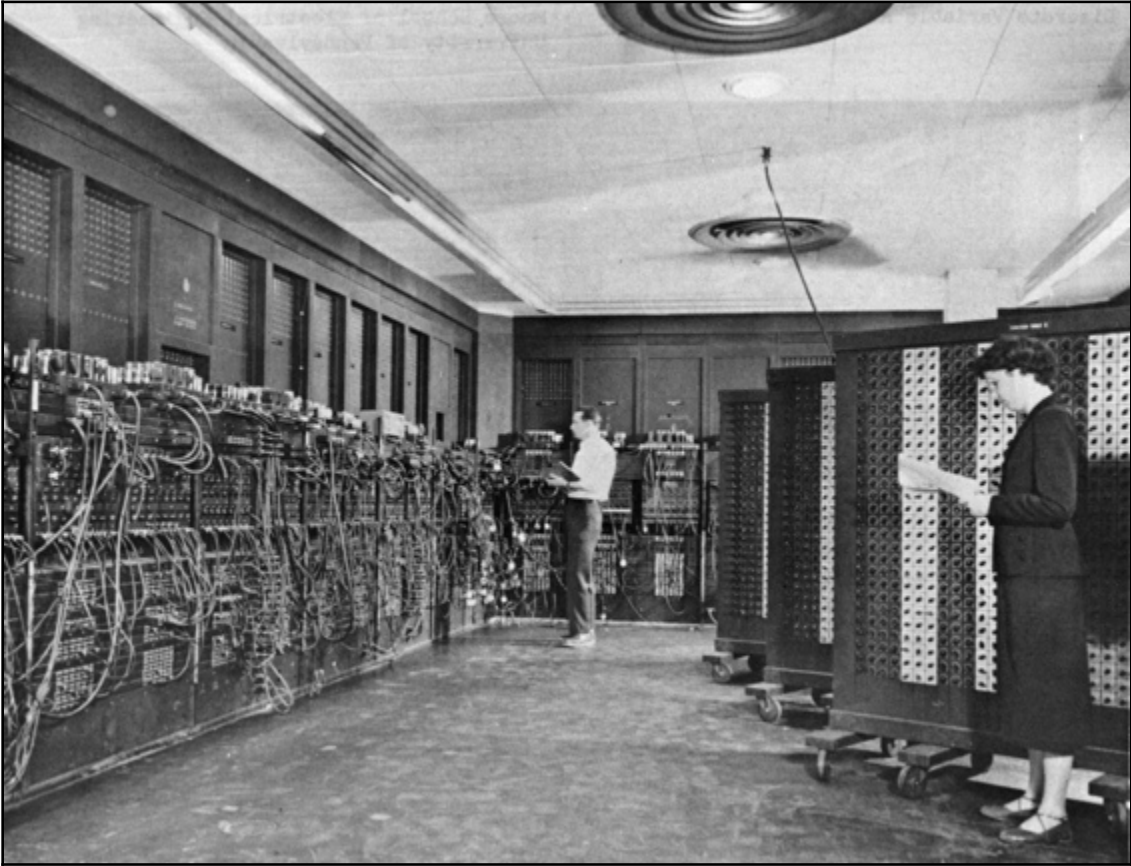
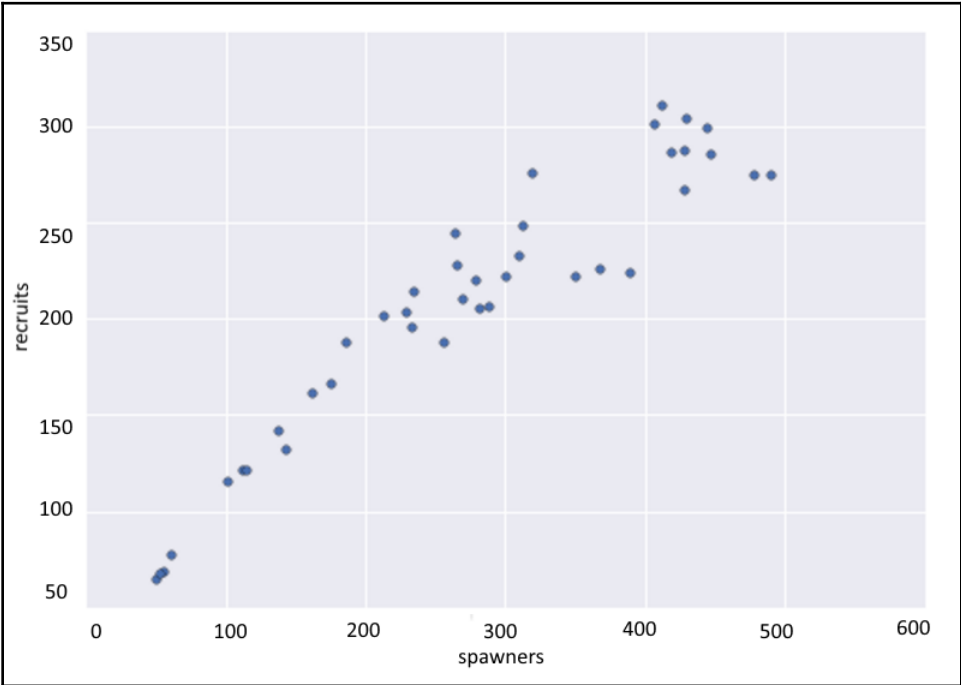
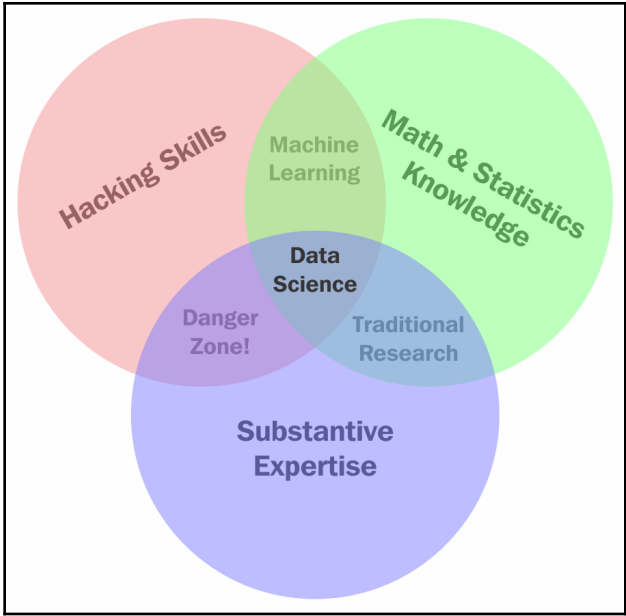
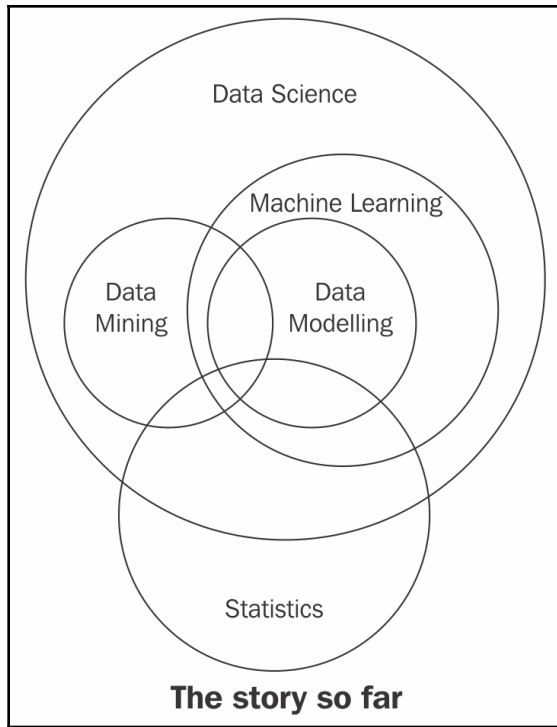


Chapter 01: How to Sound Like a Data Scientist







B. To be completed by the claimant

PLEASE PRINT

Please Answer the Following Questions:

(1) Have you been treated or examined by a doctor (other than a doctor at a hospital) since the above date? _____ ▶

Yes No

(If yes, please list the names, addresses and telephone numbers of doctors who have treated or examined you since the above date. Also list the dates of treatment or examination. If possible, send updated reports from these doctors to the Administrative Law Judge before the date of your hearing.)

DOCTORS NAME(S)	ADDRESS(ES) & TELEPHONE NO.(S)	DATE(S)

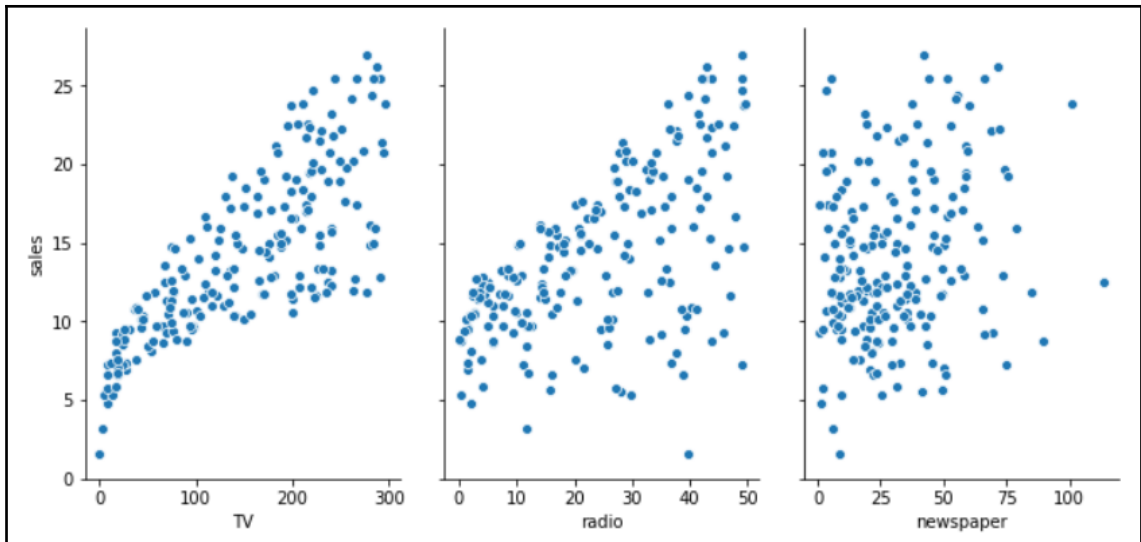
(2) What have these doctors told you about your condition?

(3) Have you been hospitalized since the above date? _____ ▶

Yes No

(If yes, please list the name and address of the hospital. Also, explain why you were hospitalized and what treatment you received.)

	TV	Radio	Newspaper	Sales
1	230.1	37.8	69.2	22.1
2	44.5	39.3	45.1	10.4
3	17.2	45.9	69.3	9.3
4	151.5	41.3	58.5	18.5
5	180.8	10.8	58.4	12.9



Machine Learning Quantitative Analyst

Bloomberg - ★★★★★☆ 282 reviews - New York, NY

The Machine Learning Quantitative Analyst will work in Bloomberg's Enterprise Solutions area and work collaboratively to build a liquidity tool for banks,...

8 days ago - [email](#)

Sponsored

Save lives with machine learning

Blue Owl - San Francisco, CA

Requirements for all data scientists. Expert in Python and core libraries used by data scientists (Numpy, Scipy, Pandas, Scikit-learn, Matplotlib/Seaborn, etc.)...

30+ days ago - [email](#)

Sponsored

Data Scientist

Indeed - ★★★★★☆ 132 reviews - Austin, TX

How a Data Scientist works. As a Data Scientist at Indeed your role is to follow the data. We are looking for a mixture between a statistician, scientist,...

Easily apply

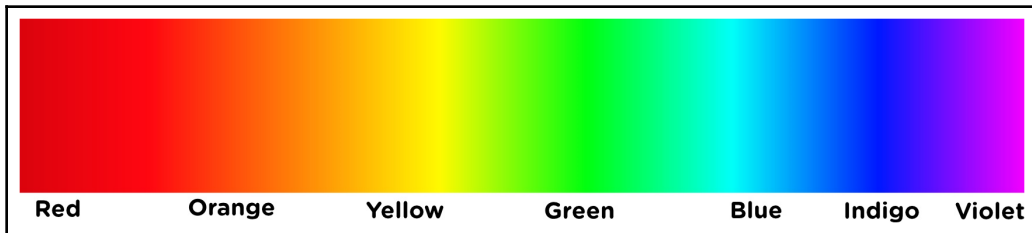
30+ days ago - [email](#)

Sponsored

10857
data 2641
analytics 857
learning 570
machine 545
machine learning 529
science 414
business 398
experience 387
marketing 313
scientist 296
data science 293
data scientist 287
marketing analytics 265
scientists 258
statistics 241
data scientists 240
algorithms 238
data analytics 229
experience data 225
applied 215
using 207
relevant 203
learning algorithms 202
analytics data 197
best 186

Chapter 02: Types of Data

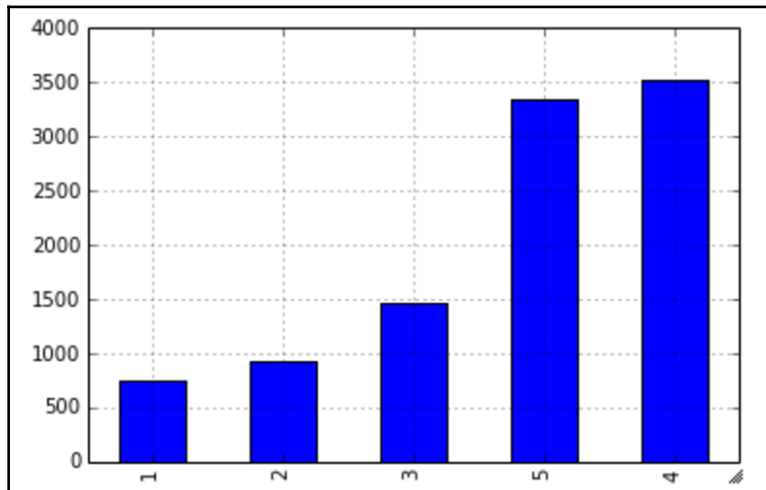
	country	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcohol	continent
0	Afghanistan	0	0	0	0.0	AS
1	Albania	89	132	54	4.9	EU
2	Algeria	25	0	14	0.7	AF
3	Andorra	245	138	312	12.4	EU
4	Angola	217	57	45	5.9	AF



Chapter 03: The Five Steps of Data Science

	business_id	date	review_id	stars	text	type	user_id	cool	useful	funny
0	9yKzy9PApeiPPOUJEtrnvk	2011-01-26	fWKvX83p0-ka4JS3dc6E5A	5	My wife took me here on my birthday for breakf...	review	rLtI8ZkDX5vH5nAx9C3q5Q	2	5	0
1	ZRJwVLyzEjQ1VAihDhYiow	2011-07-27	IjZ33sJrzXqU-0X6U8NwyA	5	I have no idea why some people give bad review...	review	0a2KyEL0d3Yb1V6aivbluQ	0	0	0
2	6oRAC4uyJCSjI1X0WZpVSA	2012-06-14	IESLBzqUCLdSzSqm0eCSxQ	4	love the gyro plate. Rice is so good and I als...	review	0hT2KtflLiobPvh6cDC8JQg	0	1	0
3	_1QQZuf4zZOyFCvXc0o6Vg	2010-05-27	G-WvGalSbqqaMHINnByoda	5	Rosie, Dakota, and I LOVE Chaparral Dog Park!!...	review	uZetI9T0NcROGOyFfughhg	1	2	0
4	6ozycU1RpktNG2-1BroVtw	2012-01-05	1uJFq2r5QfJG_6ExMRCaGw	5	General Manager Scott Petello is a good egg!!!...	review	vYmM4KtSc8ZfQBg-j5MWkw	0	0	0

	business_id	date	review_id	stars	text	type	user_id	cool	useful	funny
4372	jvvh4Q00Hq2XylcfmAAT2A	2012-06-16	ivGRamFF3KurE9bjkl6uMw	2	This review is for the chain in general. The l...	review	KLekdmo4FdNnP0huUhzZNw	0	0	0
9680	rIOnUa02zMz_ki8eF-Adug	2012-06-16	mutQE6UfjLlpJ8Wozpq5UA	2	This review is for the chain in general. The l...	review	KLekdmo4FdNnP0huUhzZNw	0	0	0



	Survived	Pclass	Name	Sex	Age
0	0	3	Braund, Mr. Owen Harris	male	22
1	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38
2	1	3	Heikkinen, Miss. Laina	female	26
3	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35
4	0	3	Allen, Mr. William Henry	male	35

	Survived	Pclass	Sex	Age
count	891.000000	891.000000	891.000000	714.000000
mean	0.383838	2.308642	0.352413	29.699118
std	0.486592	0.836071	0.477990	14.526497
min	0.000000	1.000000	0.000000	0.420000
25%	0.000000	2.000000	0.000000	20.125000
50%	0.000000	3.000000	0.000000	28.000000
75%	1.000000	3.000000	1.000000	38.000000
max	1.000000	3.000000	1.000000	80.000000

	Survived	Pclass	Name	Sex	Age
0	0	3	Braund, Mr. Owen Harris	0	22
1	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	1	38
2	1	3	Heikkinen, Miss. Laina	1	26
3	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	1	35
4	0	3	Allen, Mr. William Henry	0	35

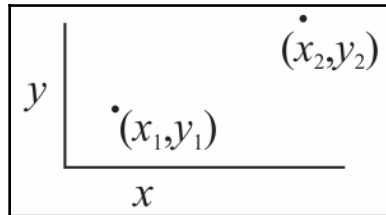
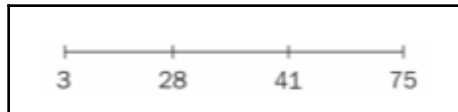
Chapter 04: Basic Mathematics

	Office 1	Office 2	Office 3
HR	57	67	65
Engineering	89	87	98
Management	94	84	60

Customer: M_1

$$\begin{pmatrix} 5 \\ 1 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} 4 \\ 5 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \cdot 4 \\ 1 \cdot 5 \\ 3 \cdot 1 \end{pmatrix} = \begin{pmatrix} 20 \\ 5 \\ 3 \end{pmatrix}$$

(5.4) → user loves comedies and this move is funny
 +
 (1.5) → user hates romance but this move is romantic
 +
 (3.1) → user doesn't mind action and the move is not action packed
 28



$$2^4 = 2 \cdot 2 \cdot 2 \cdot 2 = 16$$

exponent (pointing to 4)
 base (pointing to 2)

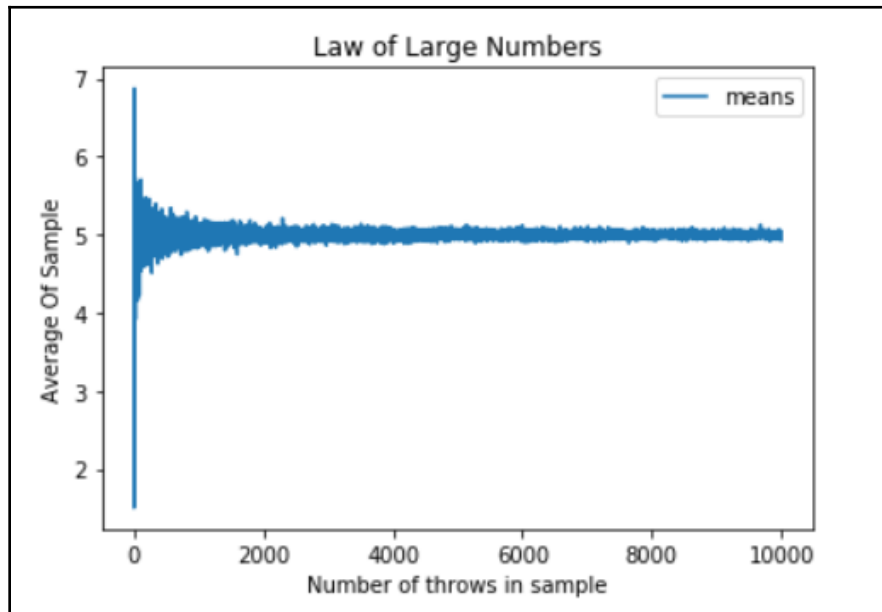
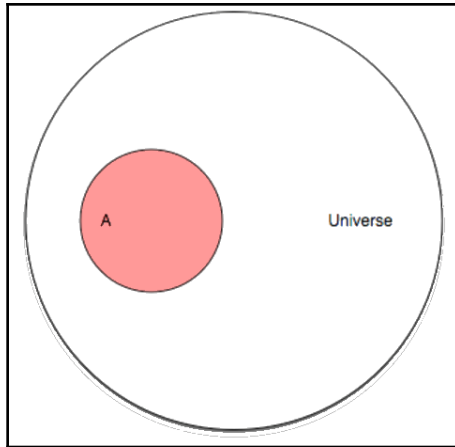
$$\log_2(16) = 4$$

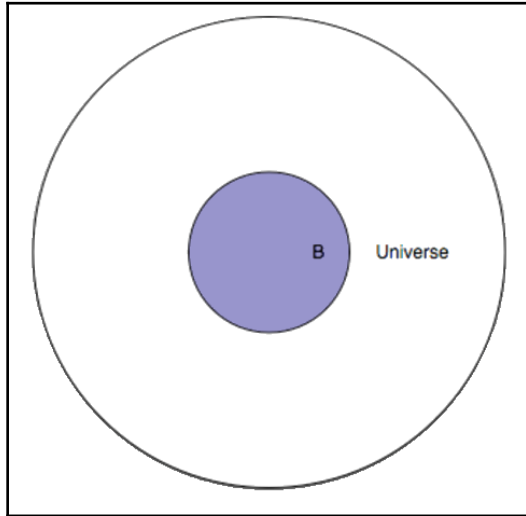
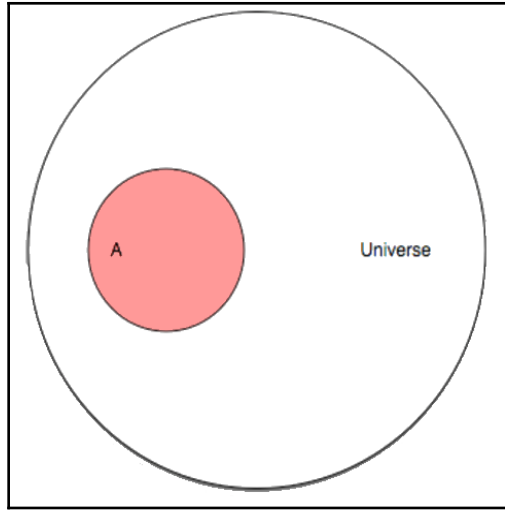
base logarithm

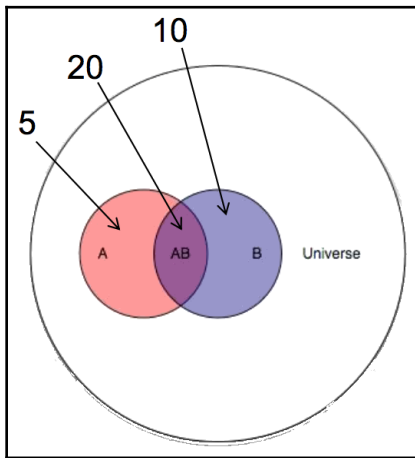
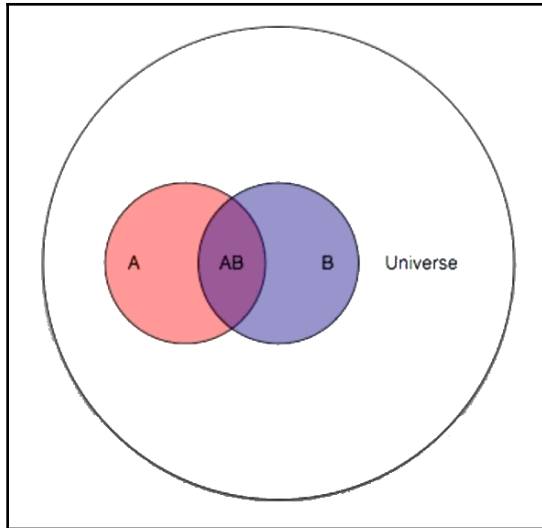
$$\log_2(16) = 4 \leftrightarrow 2^4 = 16$$

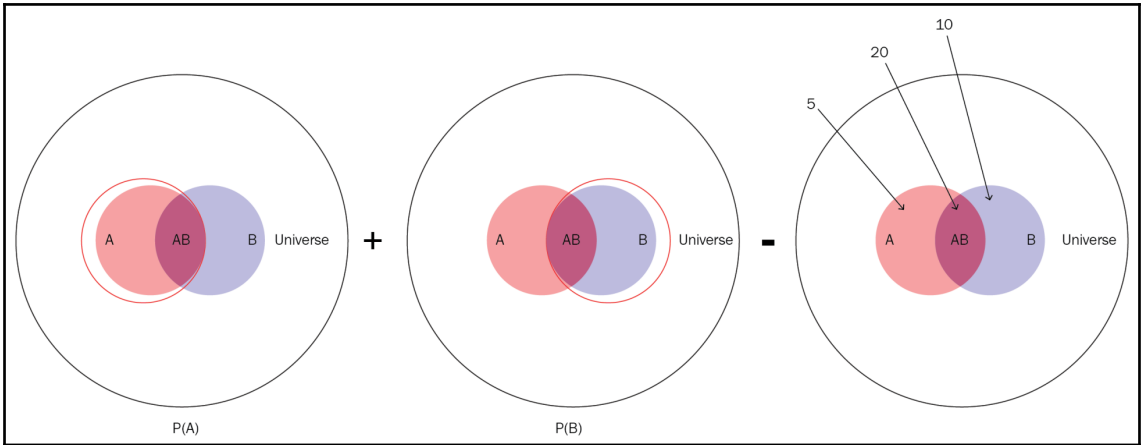
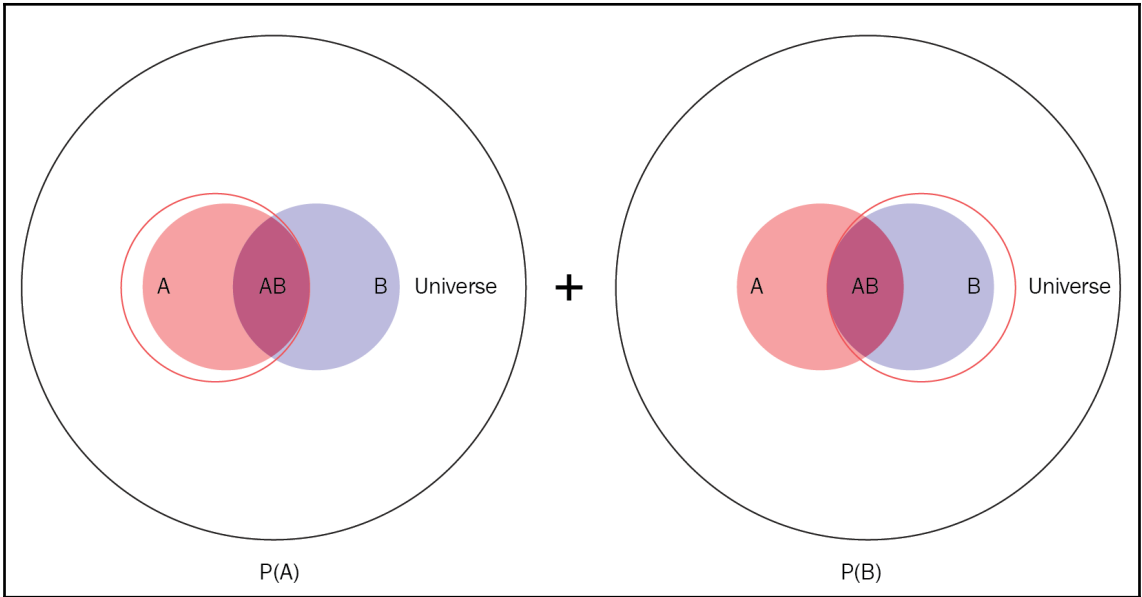
$$2 = e^{.03t} \leftrightarrow \log_e(2) = .03t$$

Chapter 05: Impossible or Improbable – A Gentle Introduction to Probability



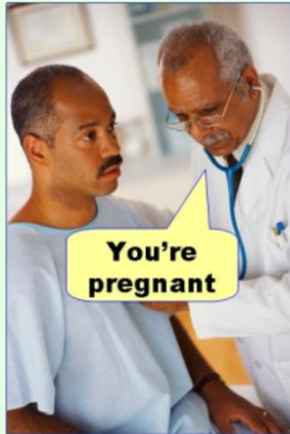






n=165	Predicted:	
	NO	YES
Actual: NO	50	10
Actual: YES	5	100

Type I error
(false positive)



Type II error
(false negative)

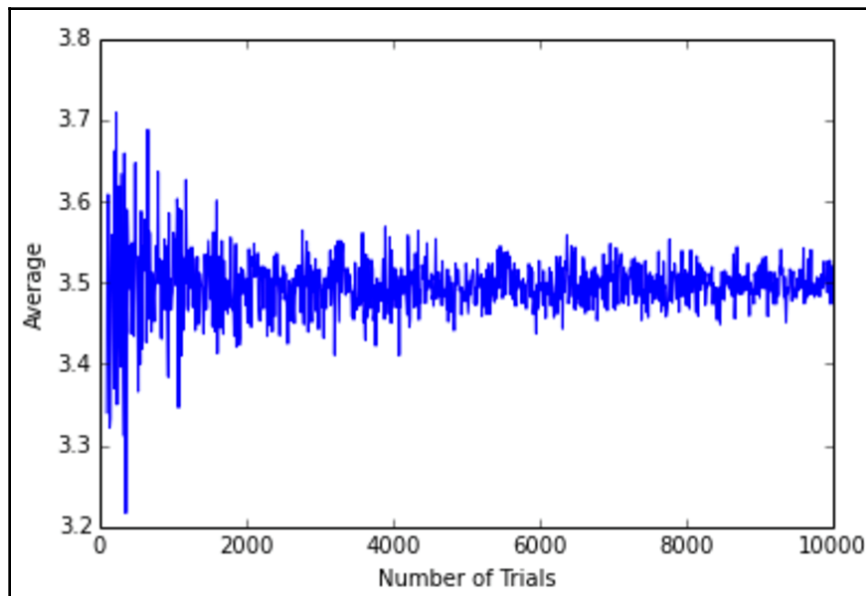


Chapter 06: Advanced Probability

	Sex	Survived
0	male	no
1	female	yes
2	female	yes
3	female	yes
4	male	no

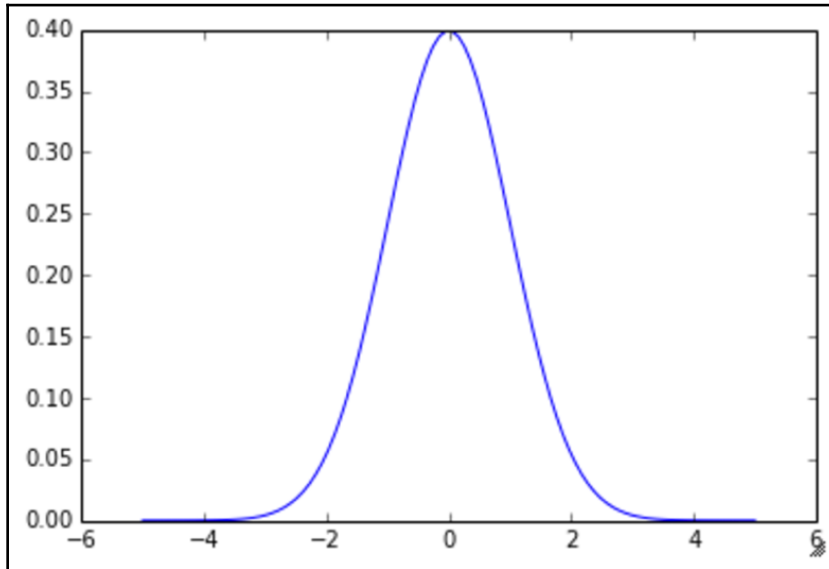
X = the outcome of a single dice roll

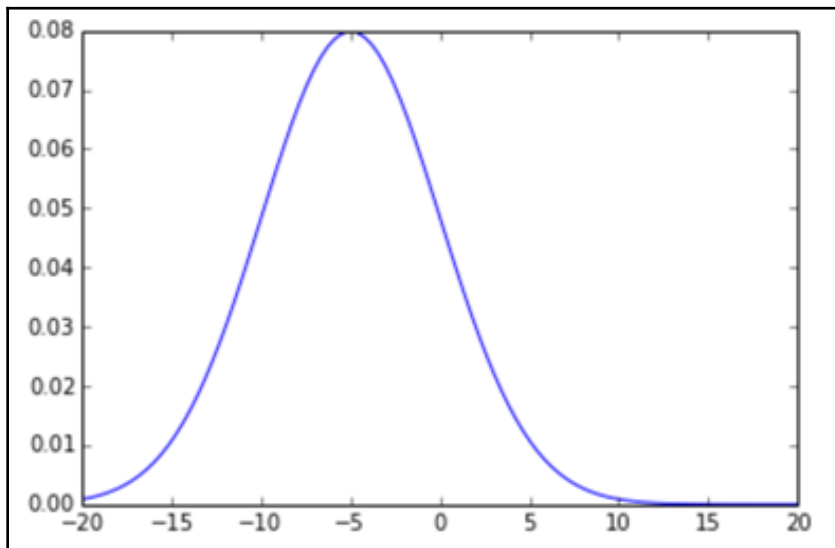
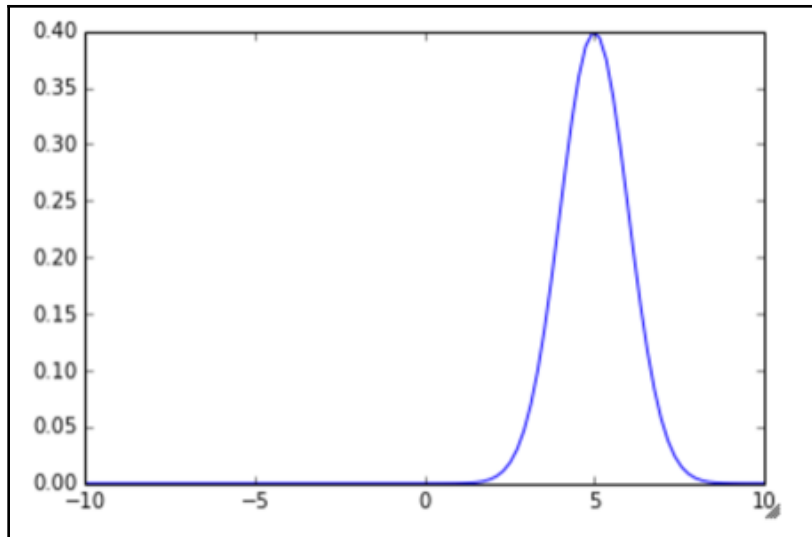
Value	$X = 1$	$X = 2$	$X = 3$	$X = 4$	$X = 5$	$X = 6$
Probability	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

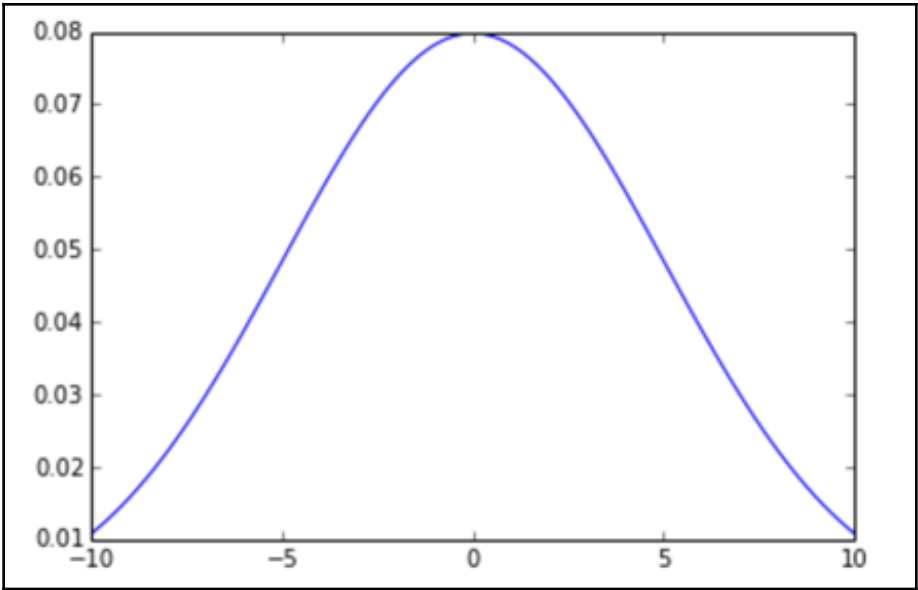


Value	$X = 0$	$X = 1$	$X = 2$	$X = 3$	$X = 4$
Probability	0.02	0.07	0.25	0.4	0.26

value x_i	0	1	2	3	4	5
Probability	0.23730	0.39551	0.26367	0.08789	0.01465	0.00098



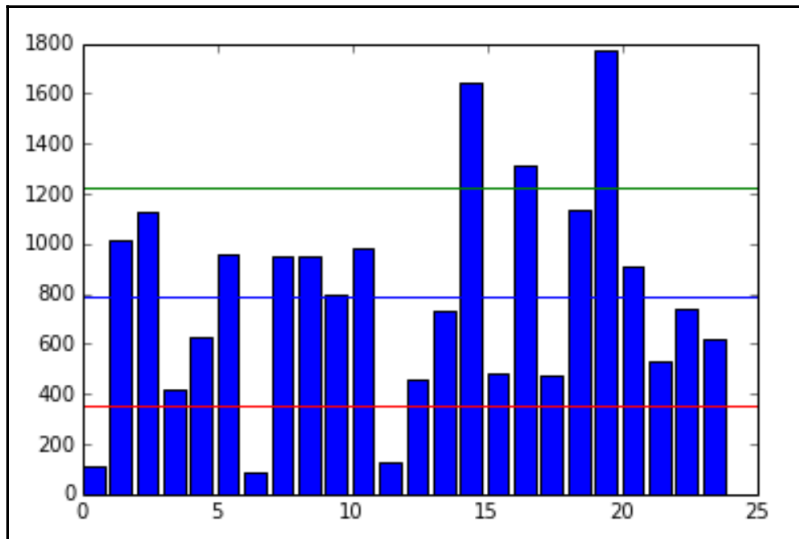




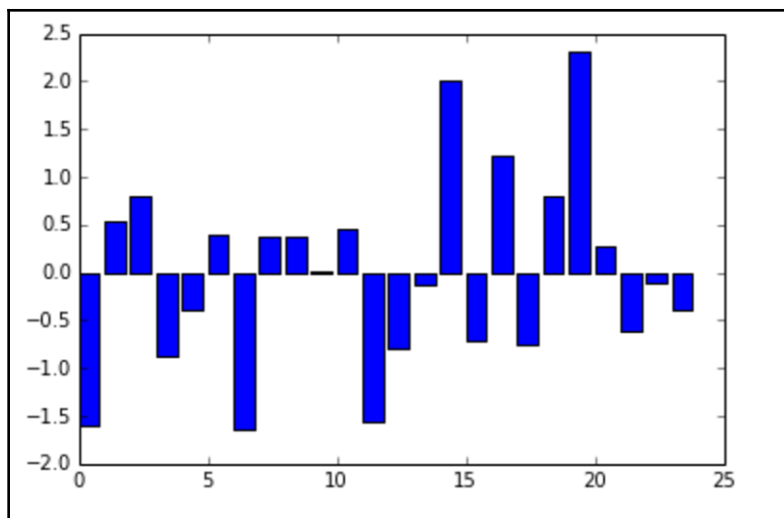
Chapter 07: Basic Statistics

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n}}$$

$$s = \sqrt{\frac{(109 - 789)^2 + (1017 - 789)^2 + \dots + (621 - 789)^2}{24}}$$



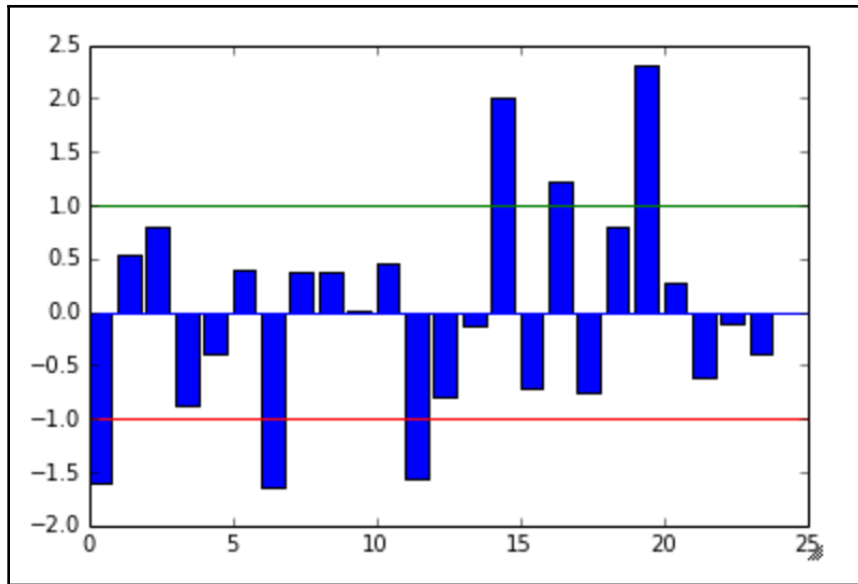
Department	Mean Salary	SD	CoV
Mailroom	\$25,000	\$2,000	8.0%
Human Resources	\$52,000	\$7,000	13.5%
Executive	\$124,000	\$42,000	33.9%



$$\text{Z-score of } (\bar{x}) = \frac{x - \bar{x}}{s} = \frac{0}{s} = 0$$

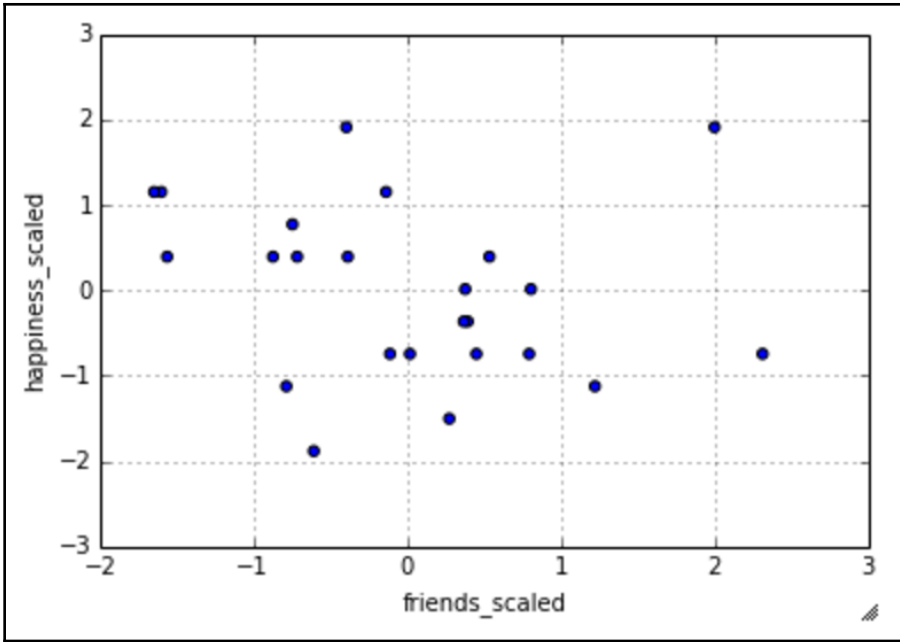
$$\text{Z-score of } (x + s) = \frac{(\bar{x} + s) - \bar{x}}{s} = \frac{s}{s} = 1$$

$$\text{Z-score of } (x - s) = \frac{(\bar{x} - s) - \bar{x}}{s} = \frac{-s}{s} = -1$$

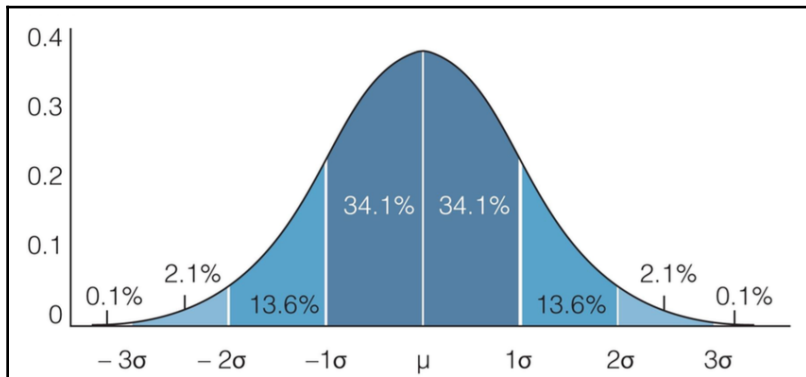


	friends	happiness
0	109	0.8
1	1017	0.6
2	1127	0.3
3	418	0.6
4	625	0.6

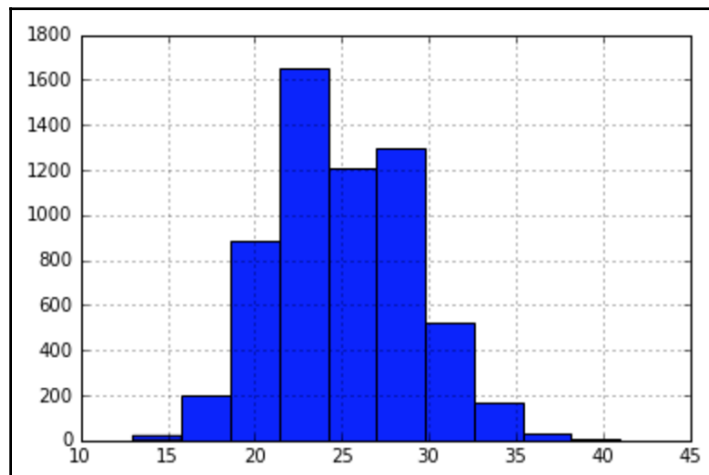
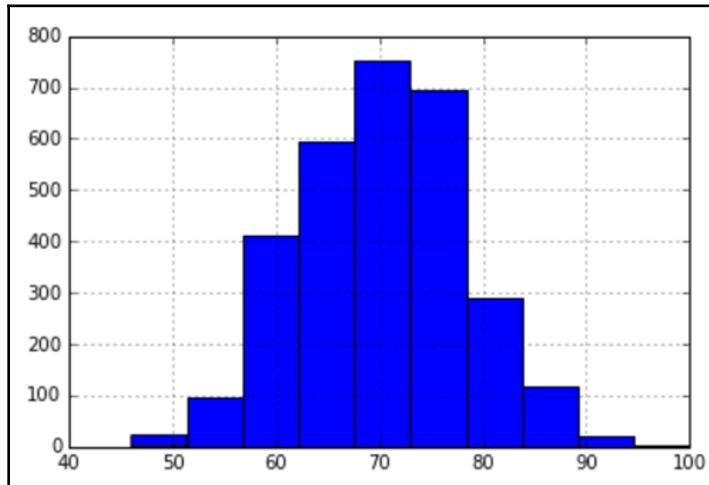
	friends_scaled	happiness_scaled
0	-1.599495	1.153223
1	0.536040	0.394939
2	0.794750	-0.742486
3	-0.872755	0.394939
4	-0.385909	0.394939

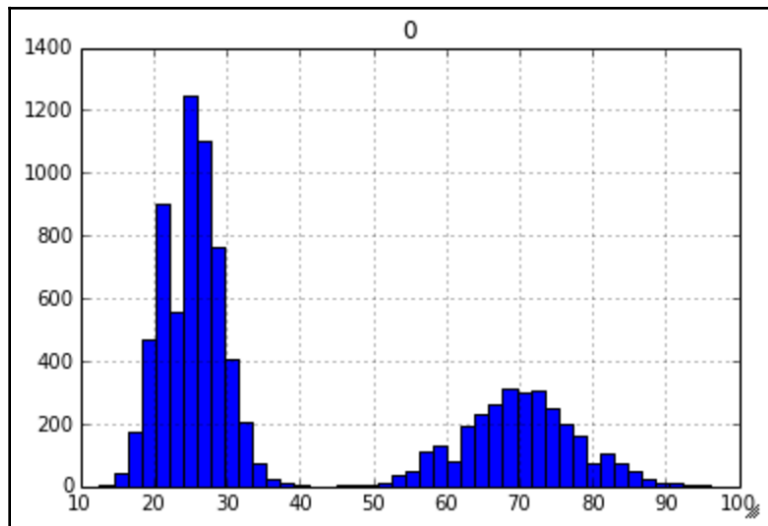
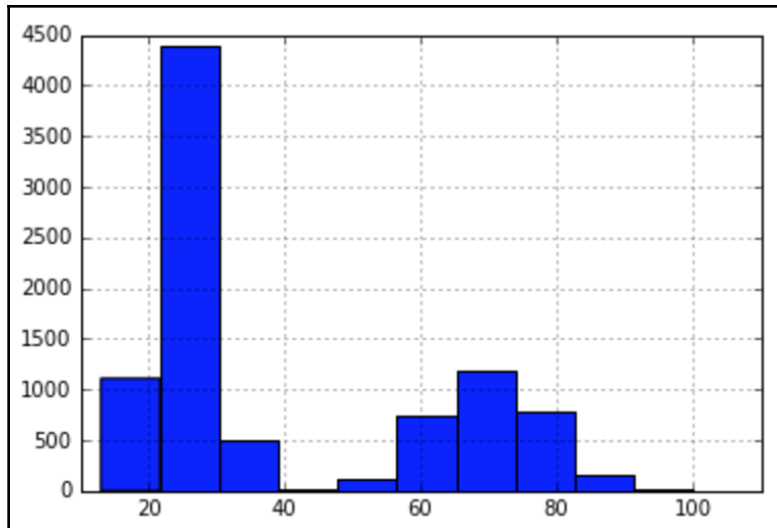


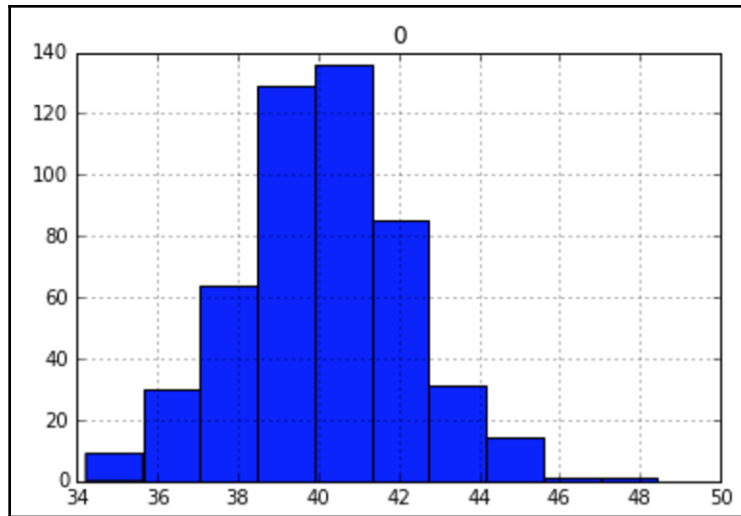
	friends	happiness
friends	1.000000	-0.216199
happiness	-0.216199	1.000000



Chapter 08: Advanced Statistics







	Under/Normal	Over	Obesity	Extreme Obesity	Total
Observed	102	178	186	34	500

	Under/Normal	Over	Obesity	Extreme Obesity	Total
Observed	102	178	186	34	500
Expected	156	165.5	147	31.5	500

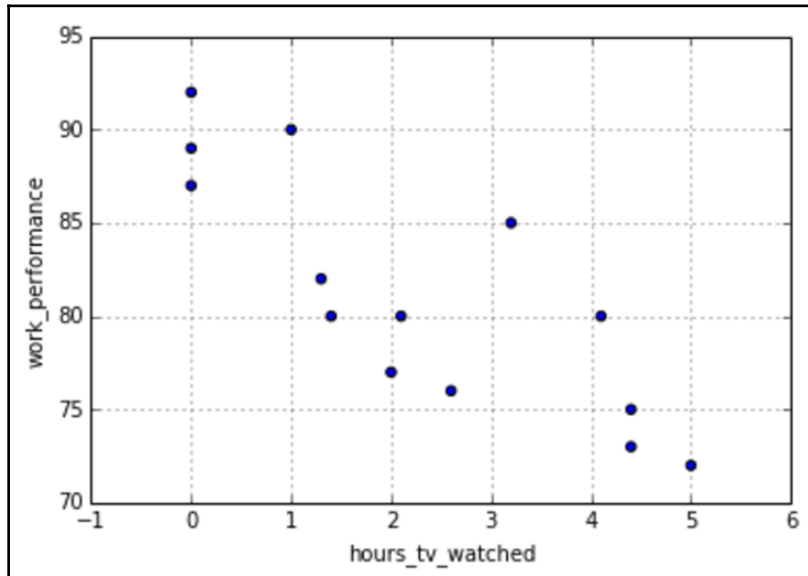
Test Statistic: $\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected}$ for $df = 3$

$$= \frac{(102 - 156)^2}{156} + \frac{(178 - 165.5)^2}{165.5} + \frac{(186 - 147)^2}{147} + \frac{(34 - 31.5)^2}{31.5} = 30.18$$

Test Statistic: $\chi^2 = \sum \frac{(Observed_{r,c} - Expected_{r,c})^2}{Expected_{r,c}}$
 over r rows and c columns

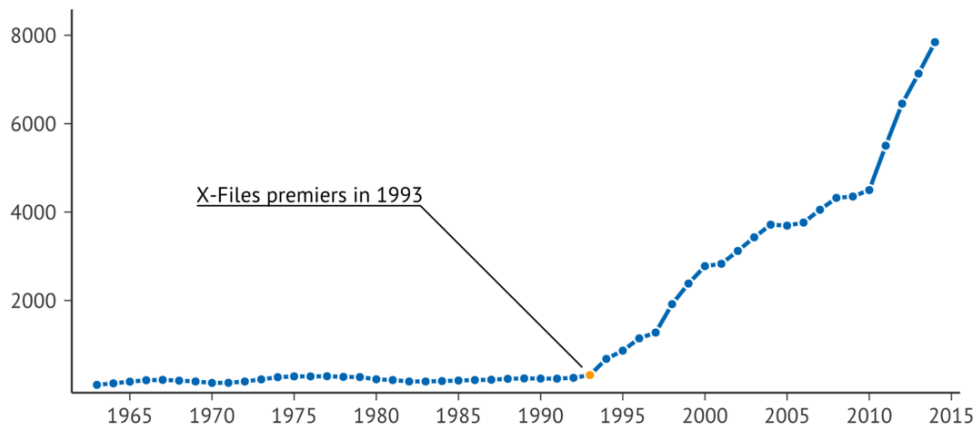
Degrees of Freedom = $(r - 1) \cdot (c - 1)$

Chapter 09: Communicating Data



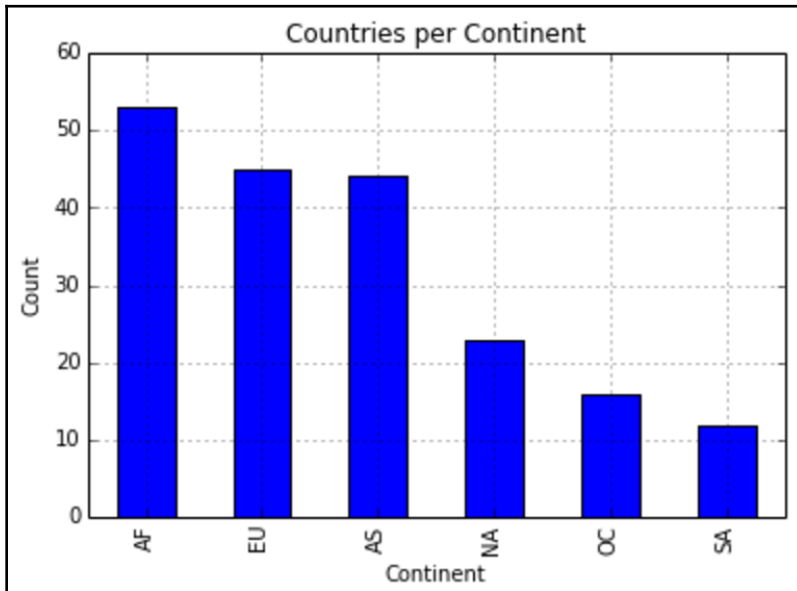
Aliens Watch the X-Files

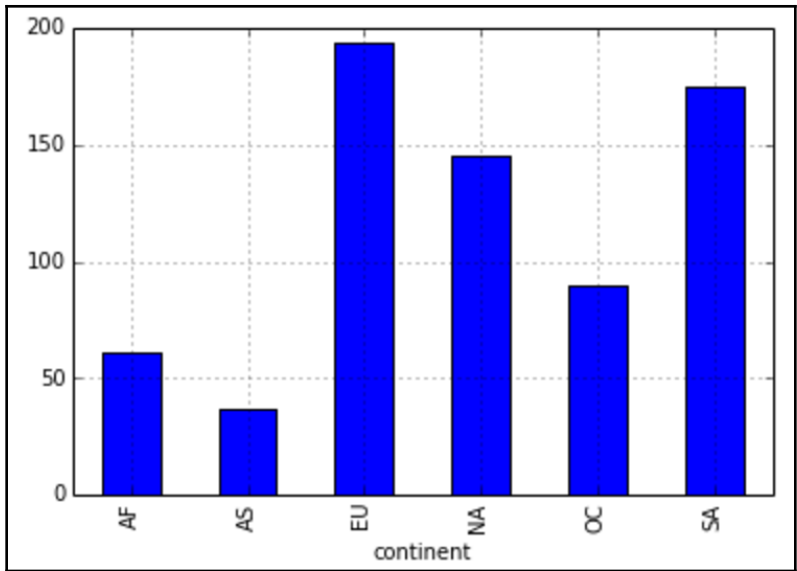
Total reported UFO sightings per year since 1963.



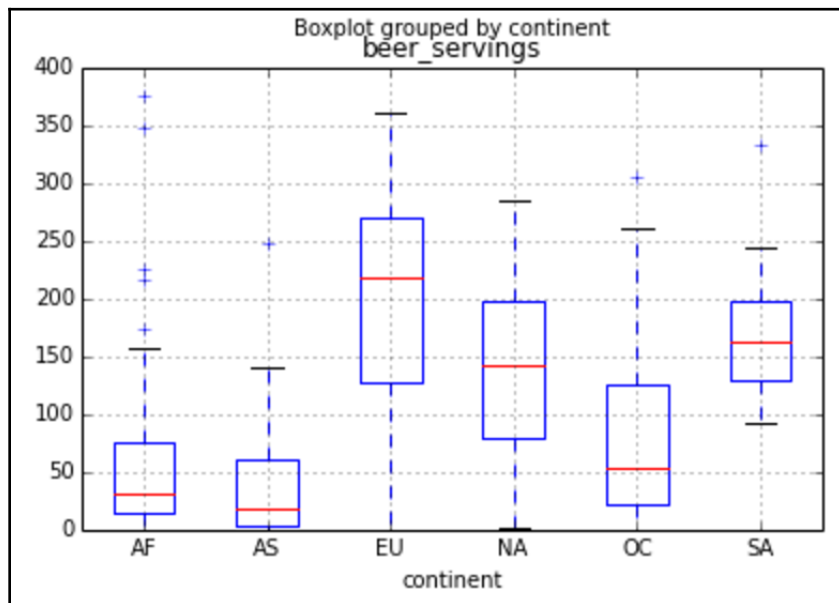
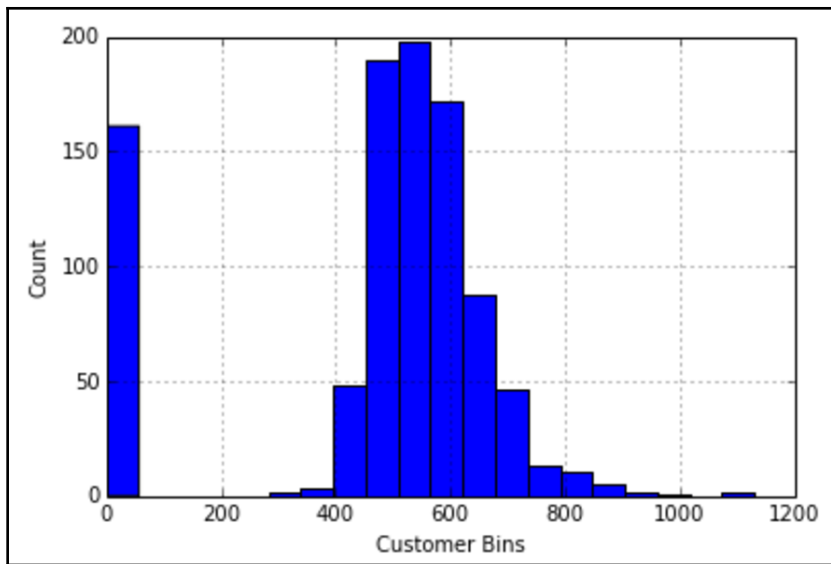
QE

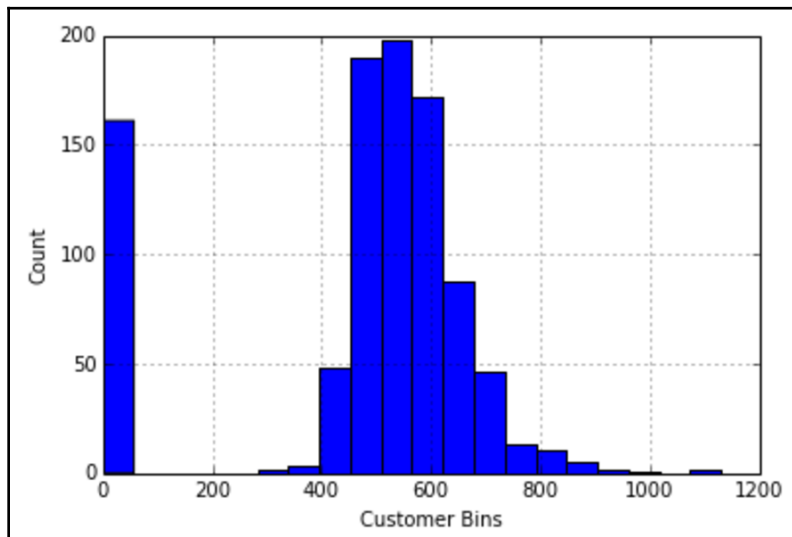
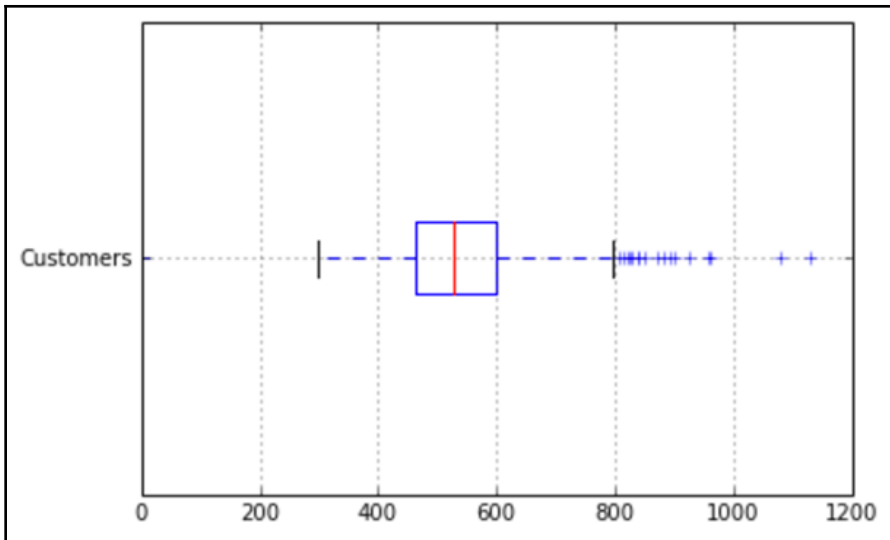
Source: NUFORC

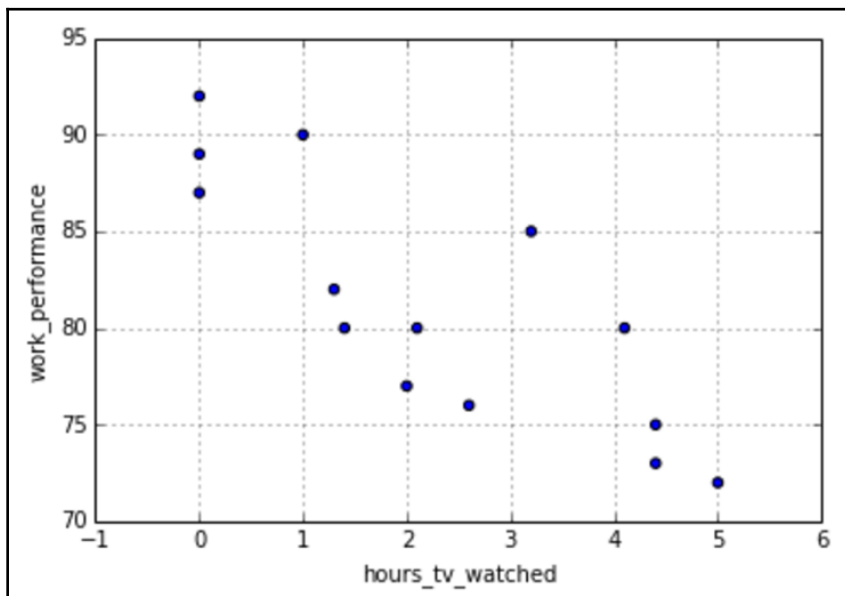
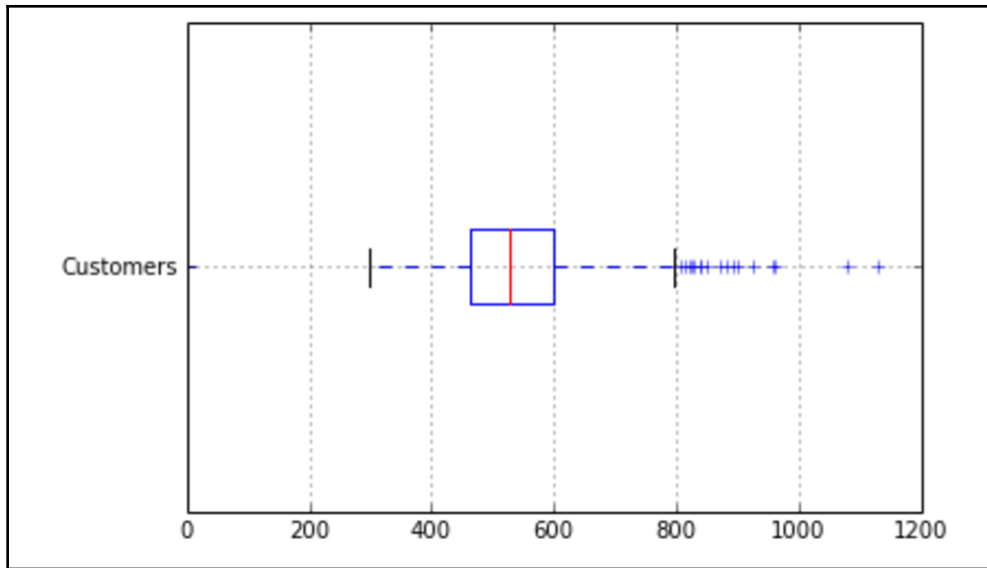


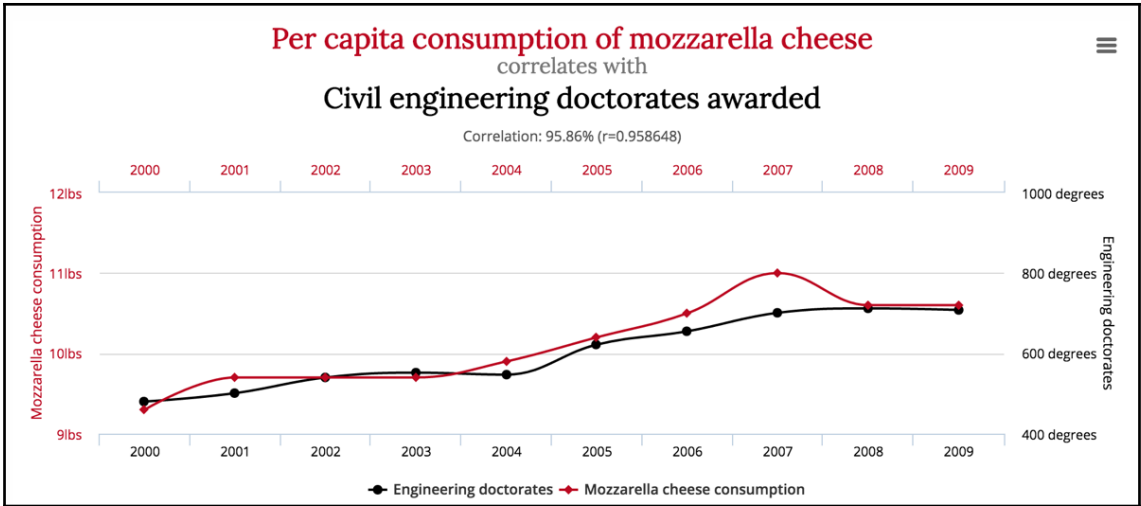


	Store	DayOfWeek	Date	Sales	Customers	Open	Promo	StateHoliday	SchoolHoliday
0	1	5	2015-07-31	5263	555	1	1	0	1
1	2	5	2015-07-31	6064	625	1	1	0	1
2	3	5	2015-07-31	8314	821	1	1	0	1
3	4	5	2015-07-31	13995	1498	1	1	0	1
4	5	5	2015-07-31	4822	559	1	1	0	1

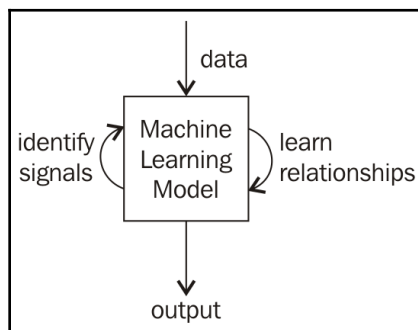
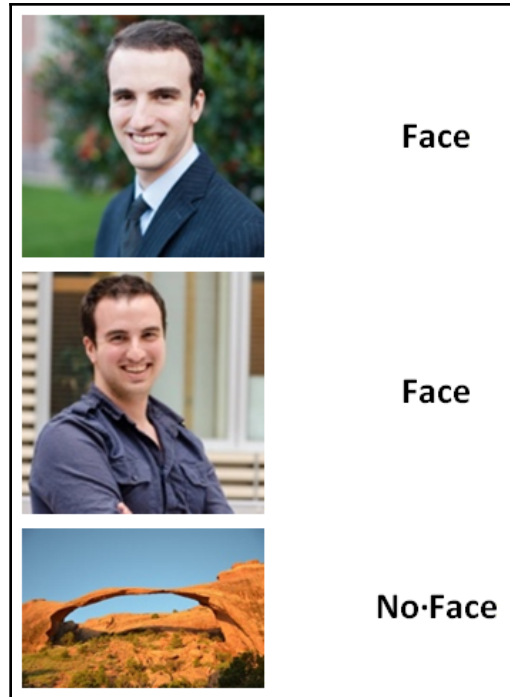


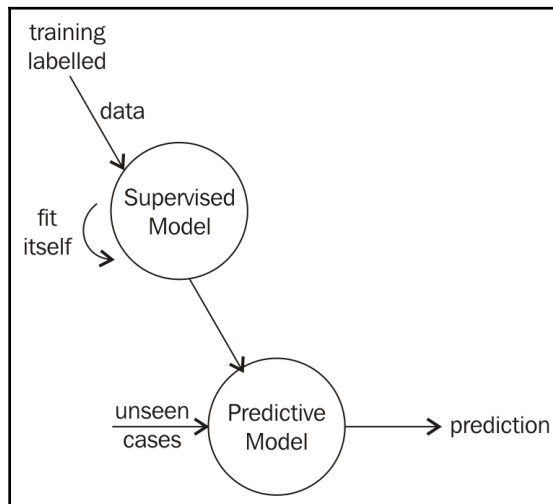
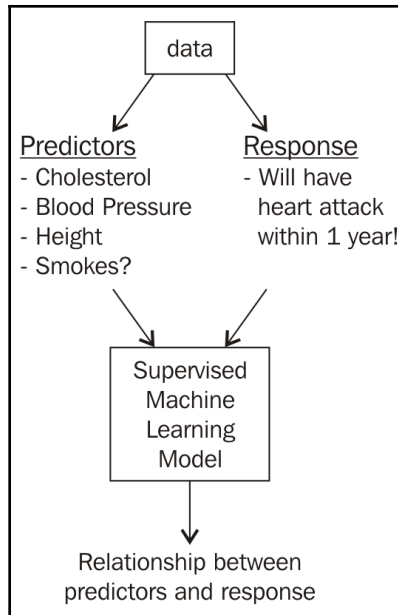


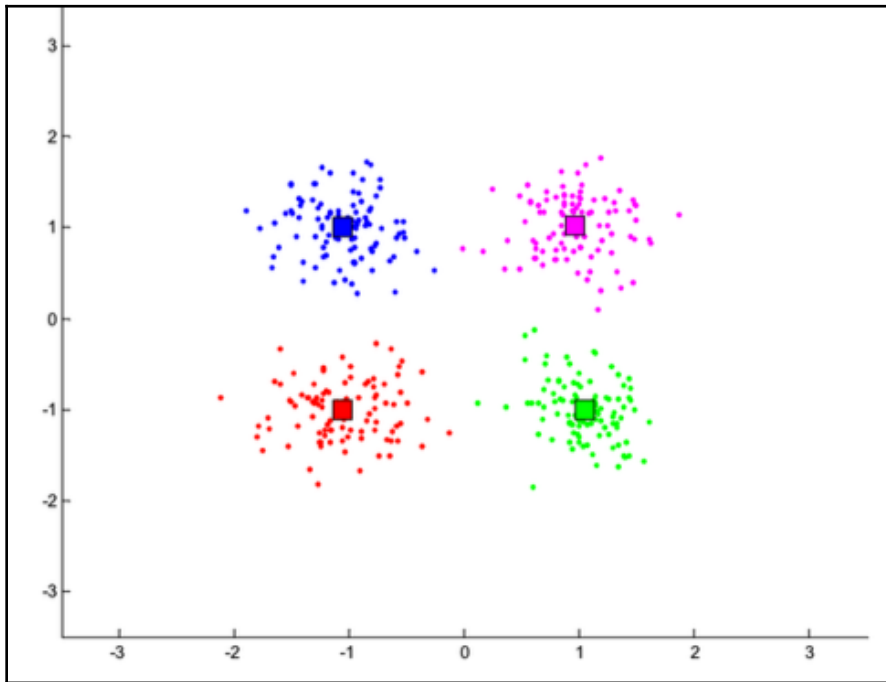
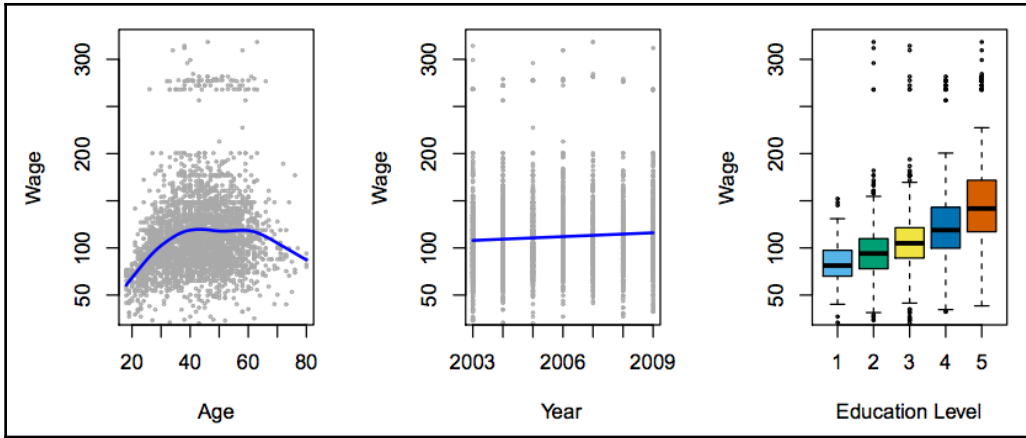


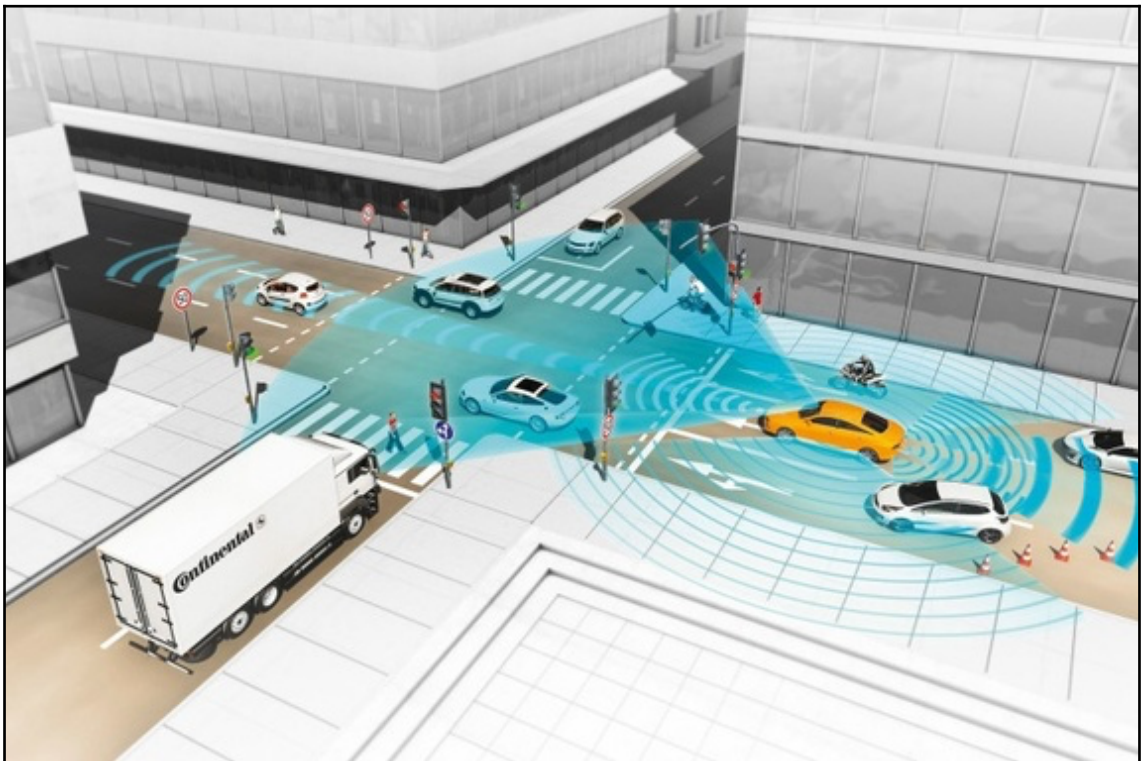
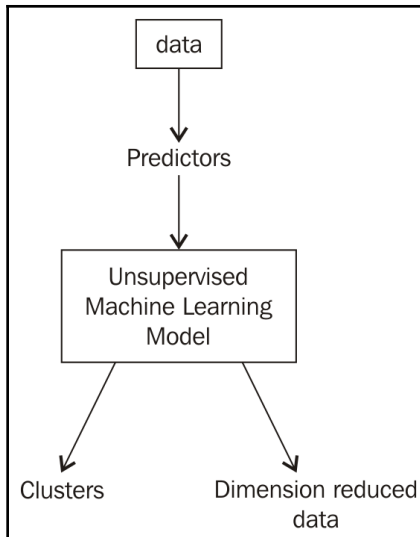


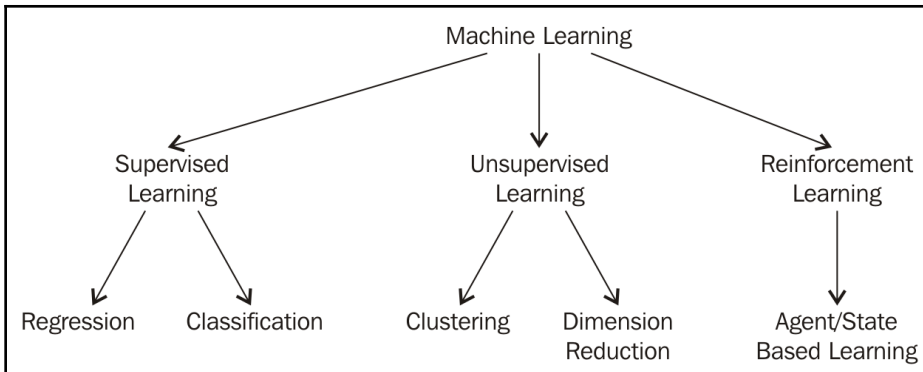
Chapter 10: How to Tell If Your Toaster Is Learning — Machine Learning Essentials



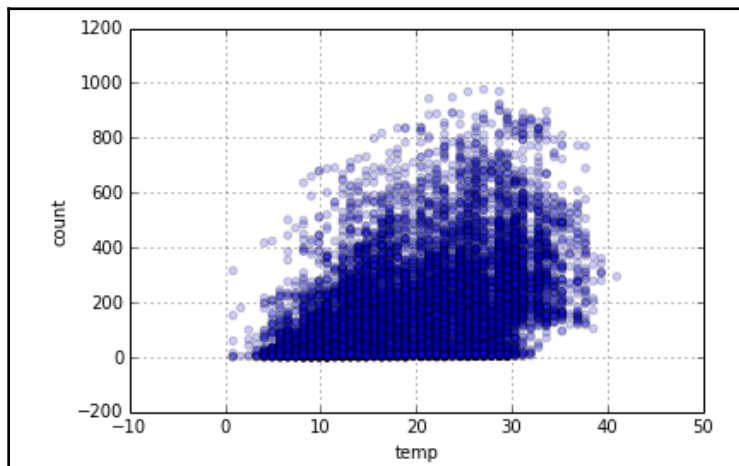


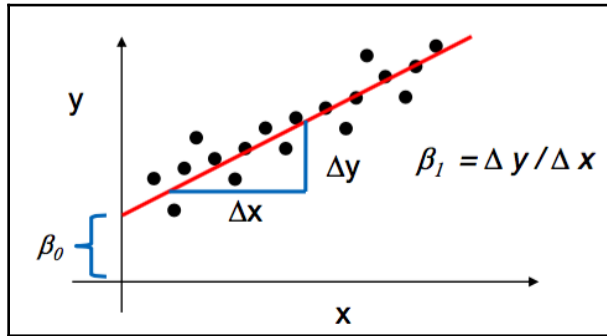
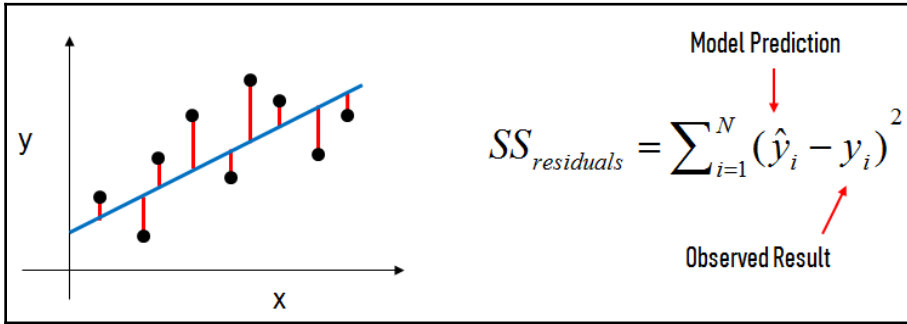
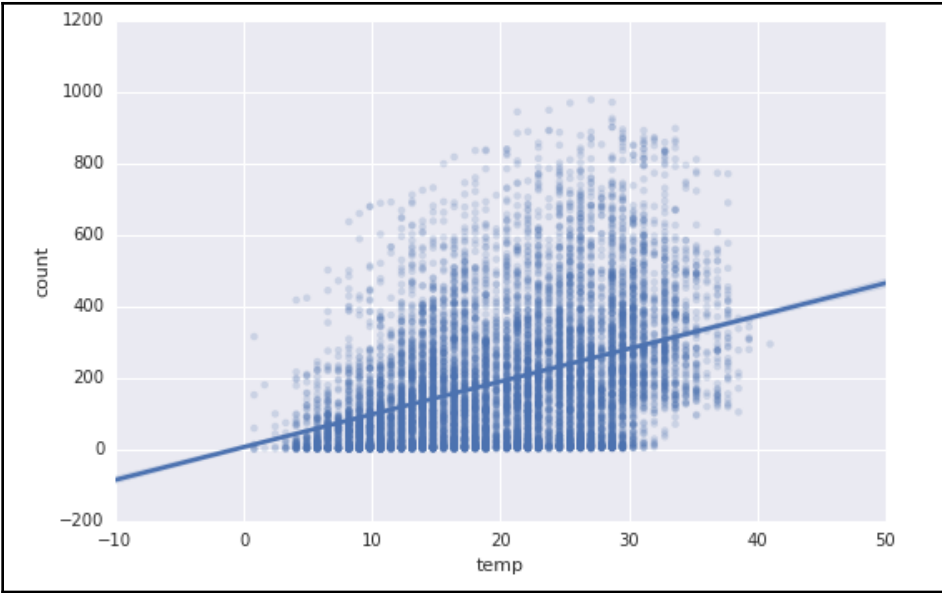


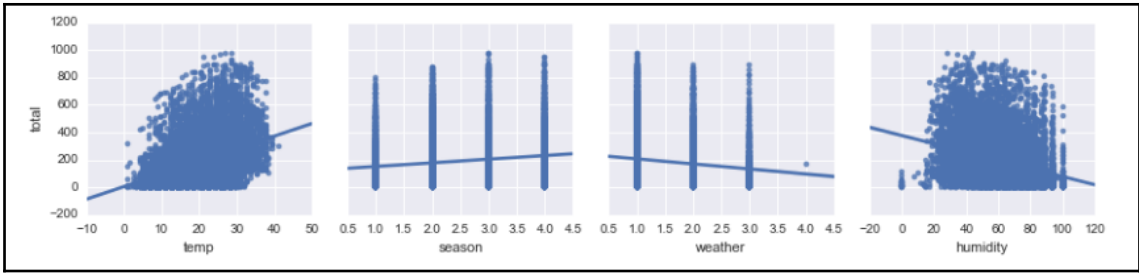




	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	casual	registered	count
0	2011-01-01 00:00:00	1	0	0	1	9.84	14.395	81	0	3	13	16
1	2011-01-01 01:00:00	1	0	0	1	9.02	13.635	80	0	8	32	40
2	2011-01-01 02:00:00	1	0	0	1	9.02	13.635	80	0	5	27	32
3	2011-01-01 03:00:00	1	0	0	1	9.84	14.395	75	0	3	10	13
4	2011-01-01 04:00:00	1	0	0	1	9.84	14.395	75	0	0	1	1







Mean Absolute Error (MAE) is the mean of the absolute value of the errors:

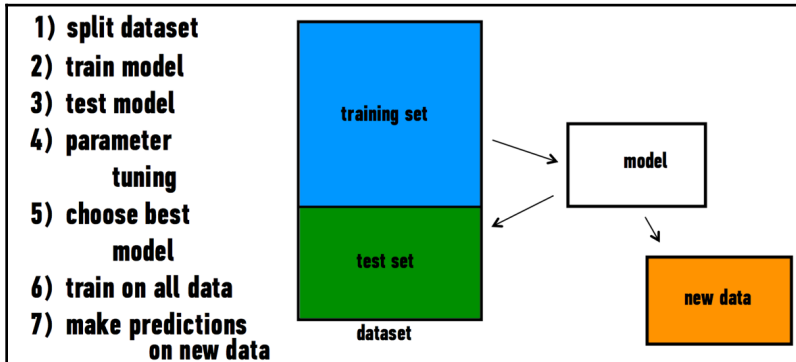
$$\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Mean Squared Error (MSE) is the mean of the squared errors:

$$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

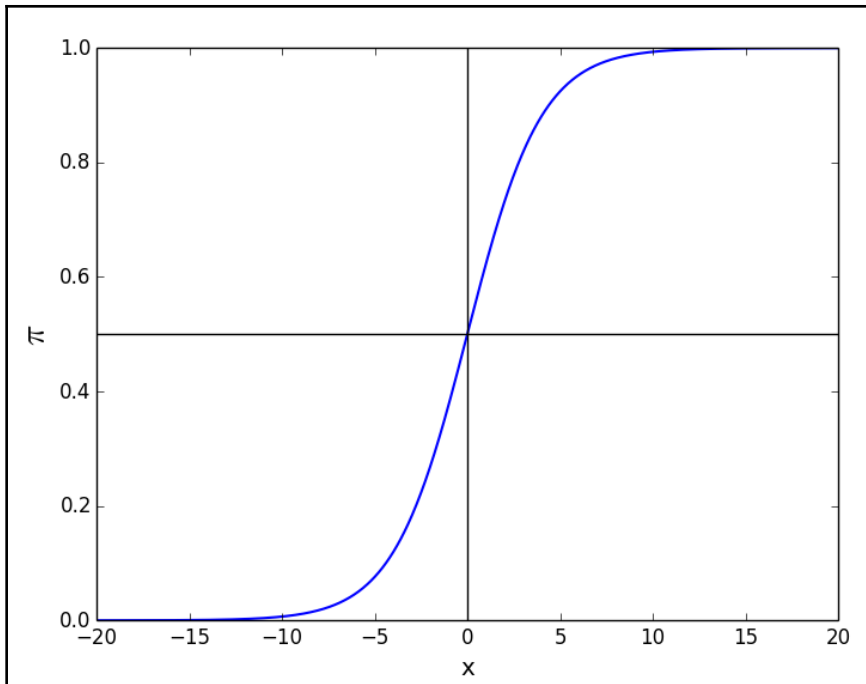
Root Mean Squared Error (RMSE) is the square root of the mean of the squared errors:

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$



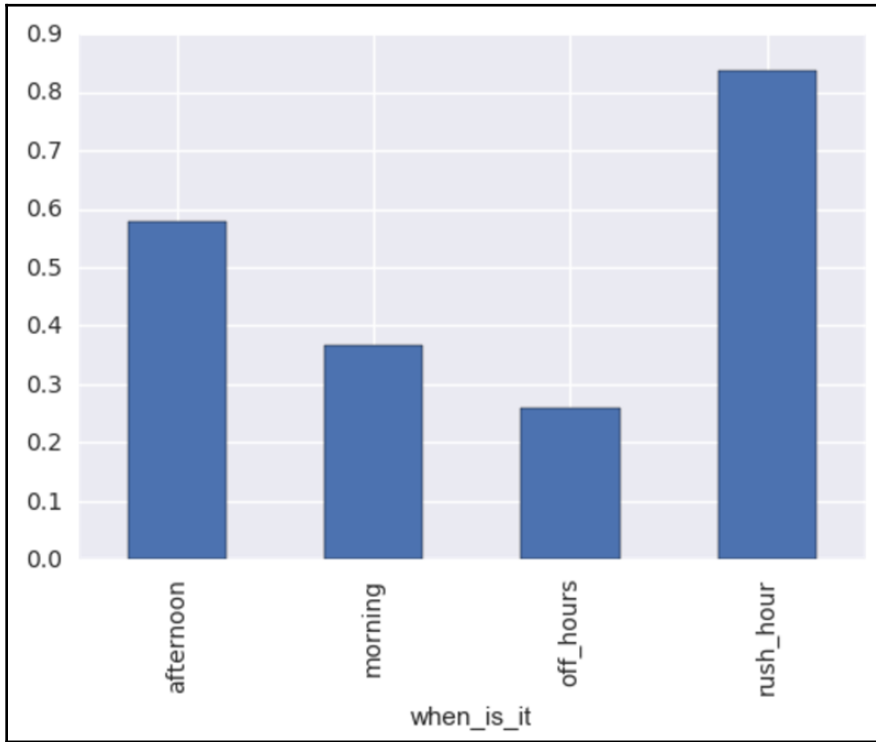
	probability	odds
0	0.10	0.111111
1	0.20	0.250000
2	0.25	0.333333
3	0.50	1.000000
4	0.60	1.500000
5	0.80	4.000000
6	0.90	9.000000

	probability	odds	logodds
0	0.10	0.111111	-2.197225
1	0.20	0.250000	-1.386294
2	0.25	0.333333	-1.098612
3	0.50	1.000000	0.000000
4	0.60	1.500000	0.405465
5	0.80	4.000000	1.386294
6	0.90	9.000000	2.197225



Major (k=4)	Engineering	Business	Literature
Computer Science	0	0	0
Engineering	1	0	0
Business	0	1	0
Literature	0	0	1
Business	0	1	0
Engineering	1	0	0

	when_is_it	above_average
0	off_hours	False
1	off_hours	False
2	off_hours	False
3	off_hours	False
4	off_hours	False



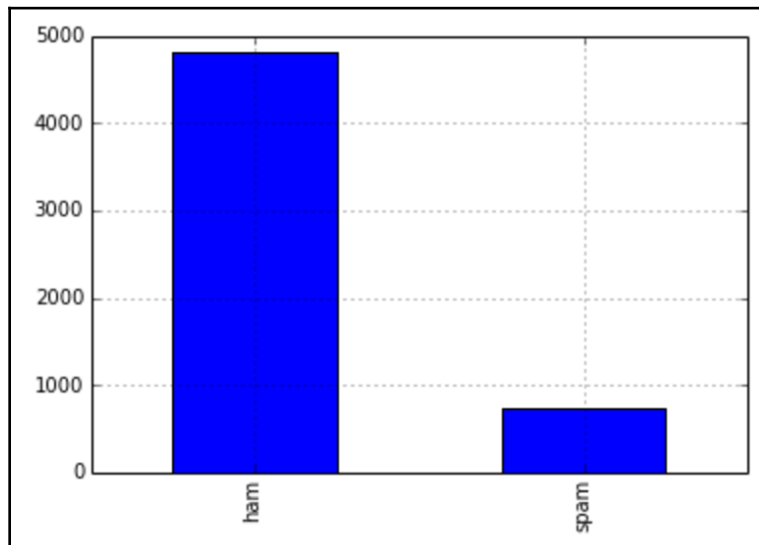
	when__afternoon	when__morning	when__off_hours	when__rush_hour
0	0.0	0.0	1.0	0.0
1	0.0	0.0	1.0	0.0
2	0.0	0.0	1.0	0.0
3	0.0	0.0	1.0	0.0
4	0.0	0.0	1.0	0.0

	when__morning	when__off_hours	when__rush_hour
0	0.0	1.0	0.0
1	0.0	1.0	0.0
2	0.0	1.0	0.0
3	0.0	1.0	0.0
4	0.0	1.0	0.0

Chapter 11: Predictions Don't Grow on Trees — or Do They?

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

	label	msg
0	ham	Go until jurong point, crazy.. Available only ...
1	ham	Ok lar... Joking wif u oni...
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...
3	ham	U dun say so early hor... U c already then say...
4	ham	Nah I don't think he goes to usf, he lives aro...
5	spam	FreeMsg Hey there darling it's been 3 week's n...
6	ham	Even my brother is not like to speak with me. ...
7	ham	As per your request 'Melle Melle (Oru Minnamin...



$$P(\text{spam} | \text{send cash now}) = P(\text{send cash now} | \text{spam}) * P(\text{spam}) / P(\text{send cash now})$$

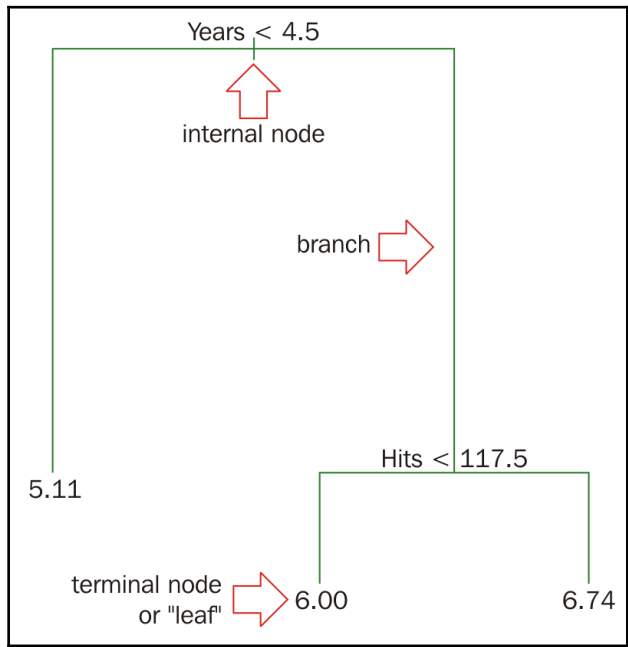
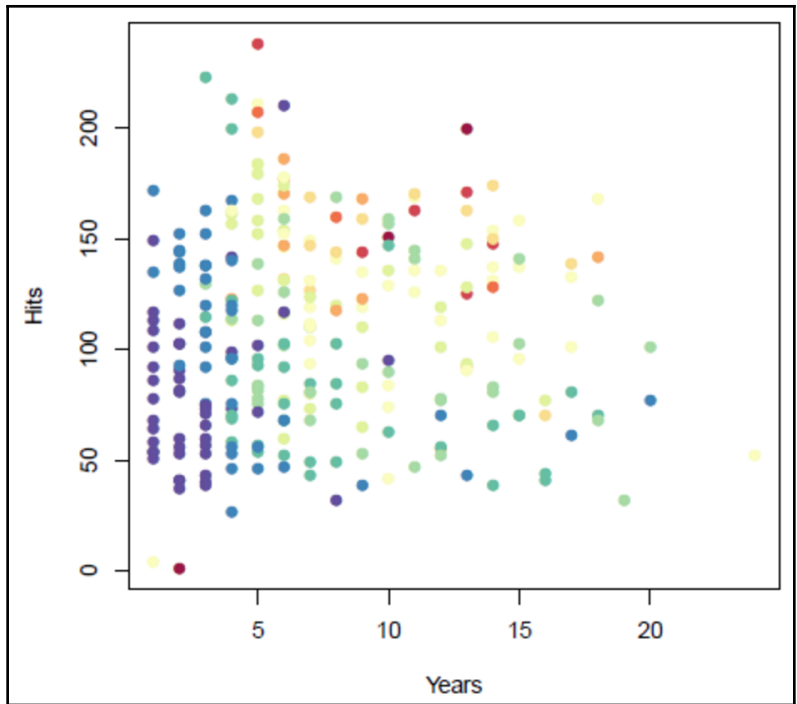
$$P(\text{ham} | \text{send cash now}) = P(\text{send cash now} | \text{ham}) * P(\text{ham}) / P(\text{send cash now})$$

$$P(\text{send cash now} | \text{spam}) * P(\text{spam}) \text{ VS } P(\text{send cash now} | \text{ham}) * P(\text{ham})$$

$$P(\text{send cash now} | \text{spam}) = P(\text{send} | \text{spam}) * P(\text{cash} | \text{spam}) * P(\text{now} | \text{spam})$$

	44	cab	call	me	please	tonight	you
0	0	0	1	0	0	1	1
1	0	1	1	1	0	0	0
2	1	0	1	1	2	0	0

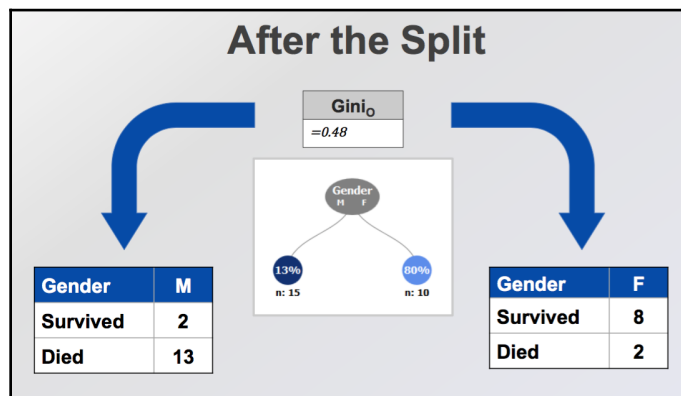
	44	cab	call	me	please	tonight	you
0	0	0	1	1	1	0	0



$$1 - \sum \left(\frac{\text{class}_i}{\text{total}} \right)^2$$

$$1 - \left(\frac{\text{survived}}{\text{total}} \right)^2 - \left(\frac{\text{died}}{\text{total}} \right)^2$$

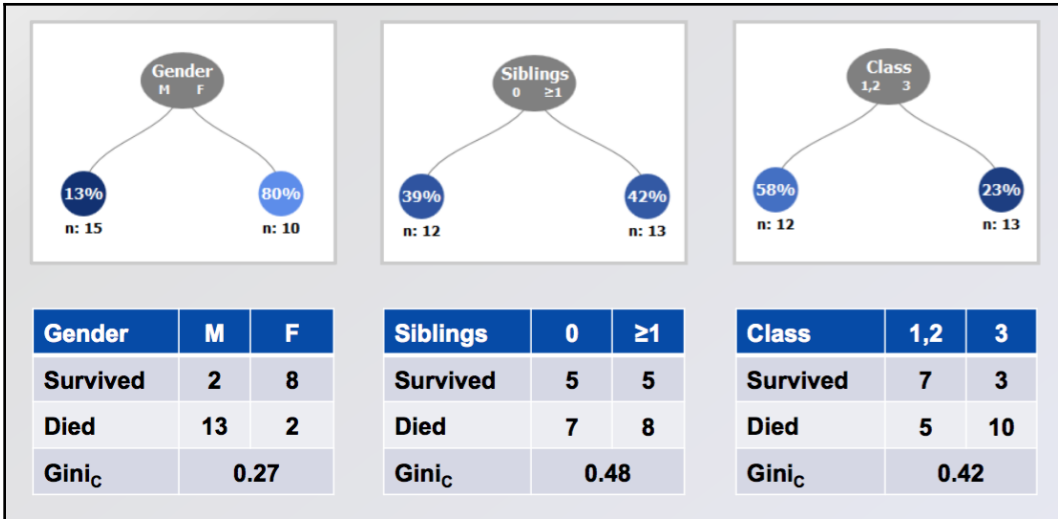
$$1 - \left(\frac{10}{25} \right)^2 - \left(\frac{15}{25} \right)^2 = 0.48$$



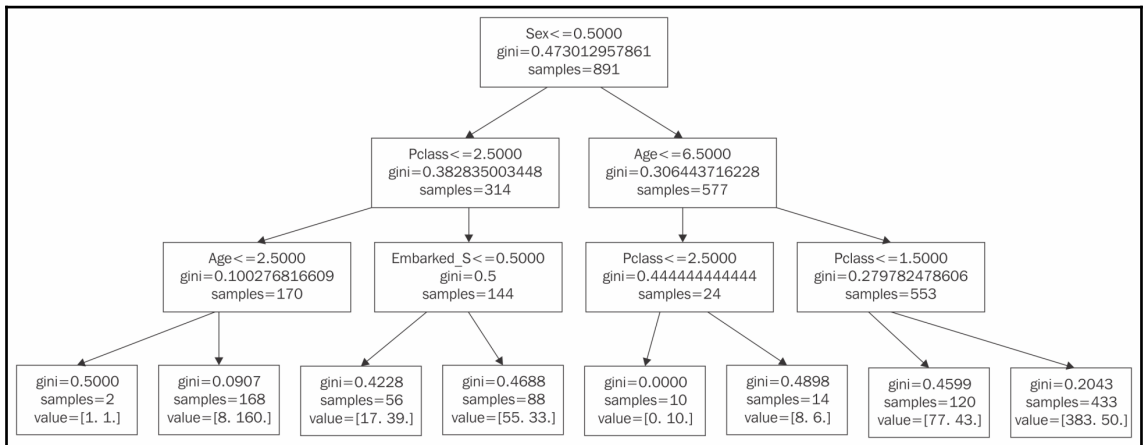
$$\text{gini}(m) = 1 - \left(\frac{2}{15} \right)^2 - \left(\frac{13}{15} \right)^2 = .23$$

$$\text{gini}(f) = 1 - \left(\frac{8}{10} \right)^2 - \left(\frac{2}{10} \right)^2 = .32$$

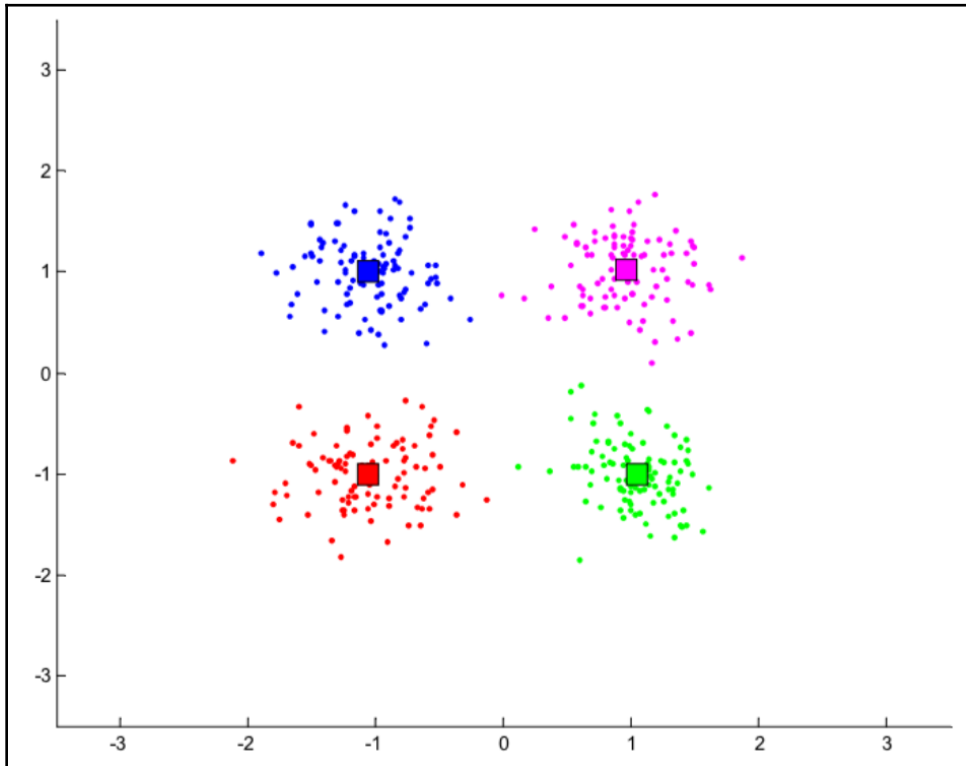
$$\text{Gini}(M)(M / M + F) + \text{Gini}(F)(F / M + F) = 0.23(15 / 10 + 15) + 0.32(10 / 10 + 15) = 0.27$$

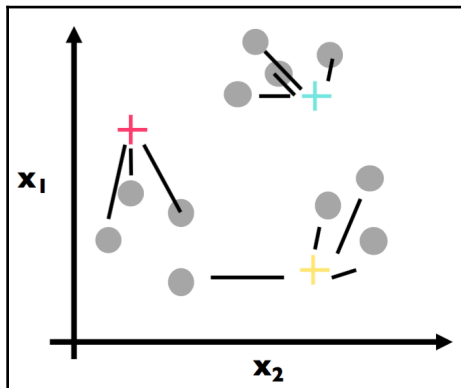
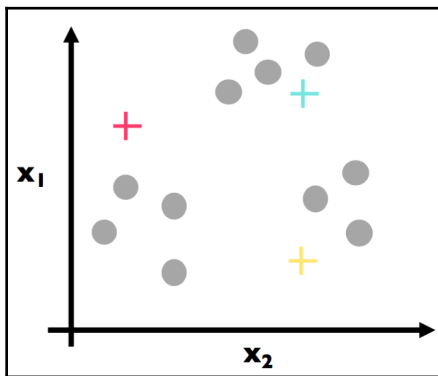
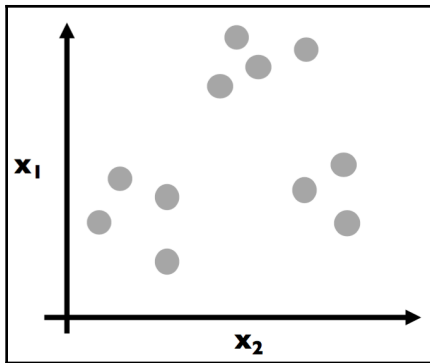


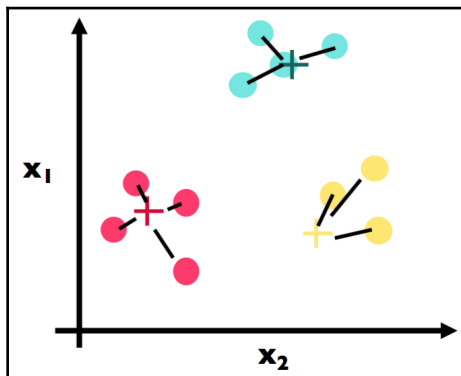
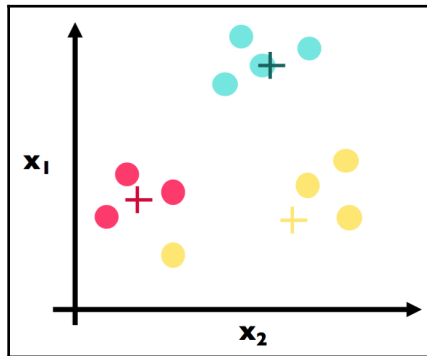
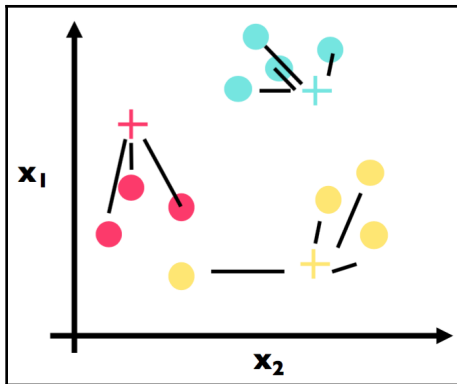
	Pclass	Sex	Age	Embarked_Q	Embarked_S
0	3	1	22.0	0.0	1.0
1	1	0	38.0	0.0	0.0
2	3	0	26.0	0.0	1.0
3	1	0	35.0	0.0	1.0
4	3	1	35.0	0.0	1.0

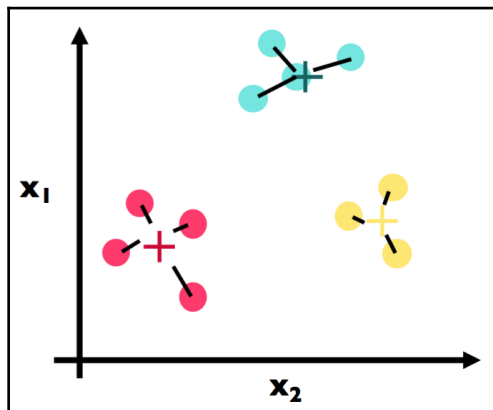
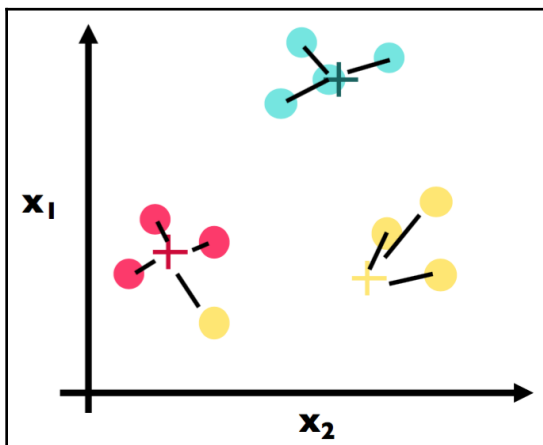
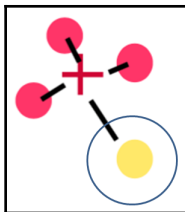


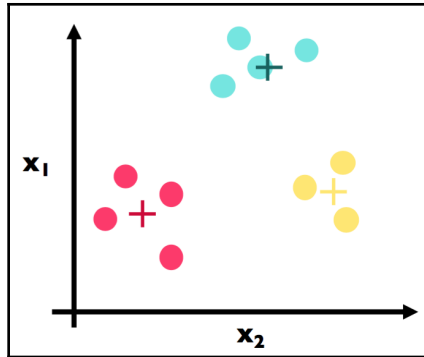
	feature	importance
0	Pclass	0.242664
1	Sex	0.655584
2	Age	0.064494
3	Embarked_Q	0.000000
4	Embarked_S	0.037258





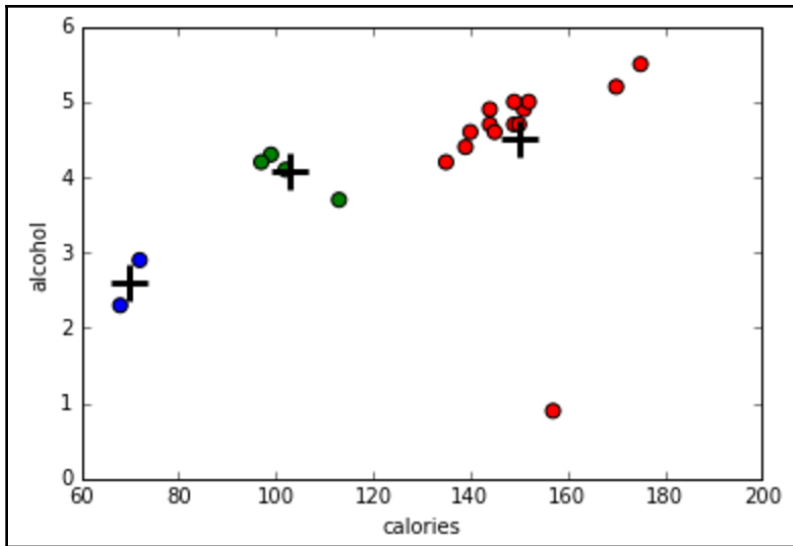






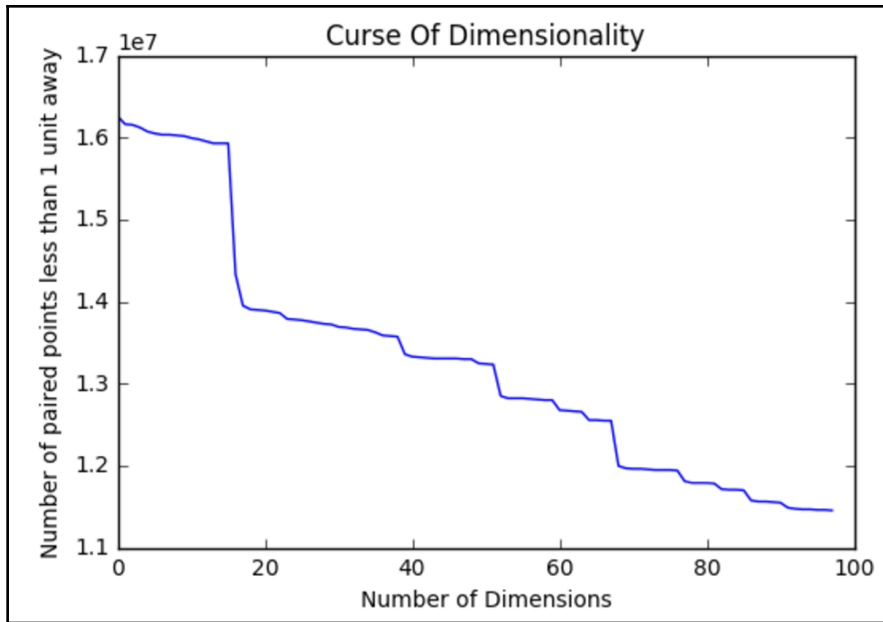
	name	calories	sodium	alcohol	cost
0	Budweiser	144	15	4.7	0.43
1	Schlitz	151	19	4.9	0.43
2	Lowenbrau	157	15	0.9	0.48
3	Kronenbourg	170	7	5.2	0.73
4	Heineken	152	11	5.0	0.77

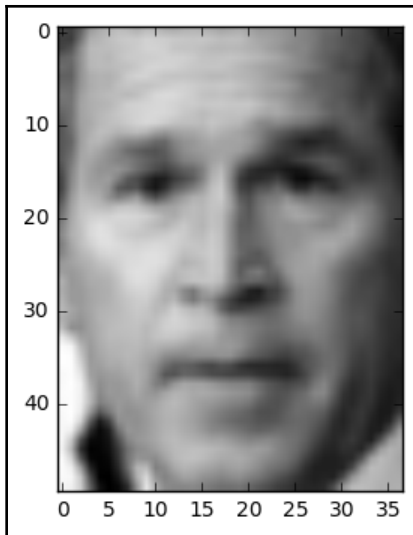
	calories	sodium	alcohol	cost
cluster				
0	150.00	17.0	4.521429	0.520714
1	102.75	10.0	4.075000	0.440000
2	70.00	10.5	2.600000	0.420000

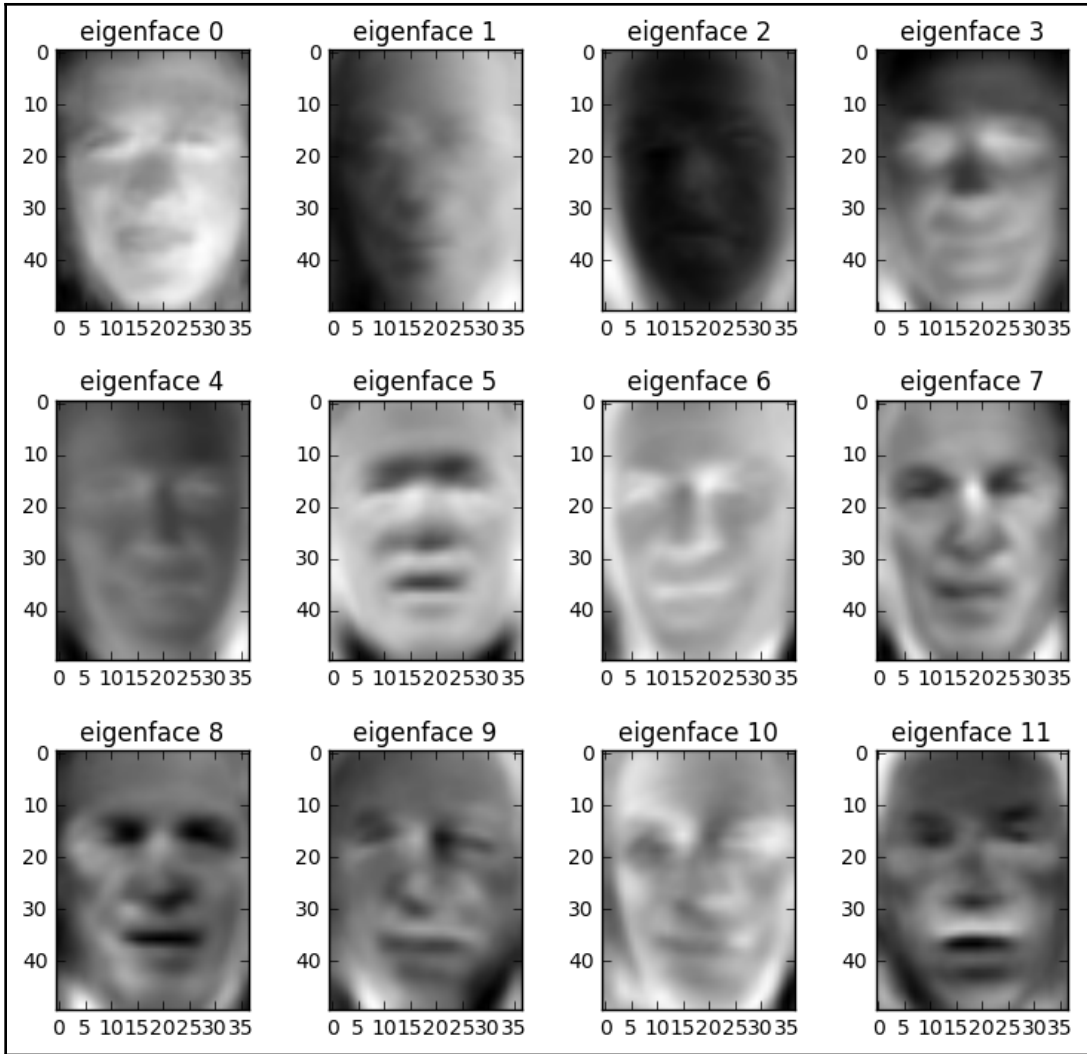


$$SC = \frac{b-a}{\max(a,b)}$$



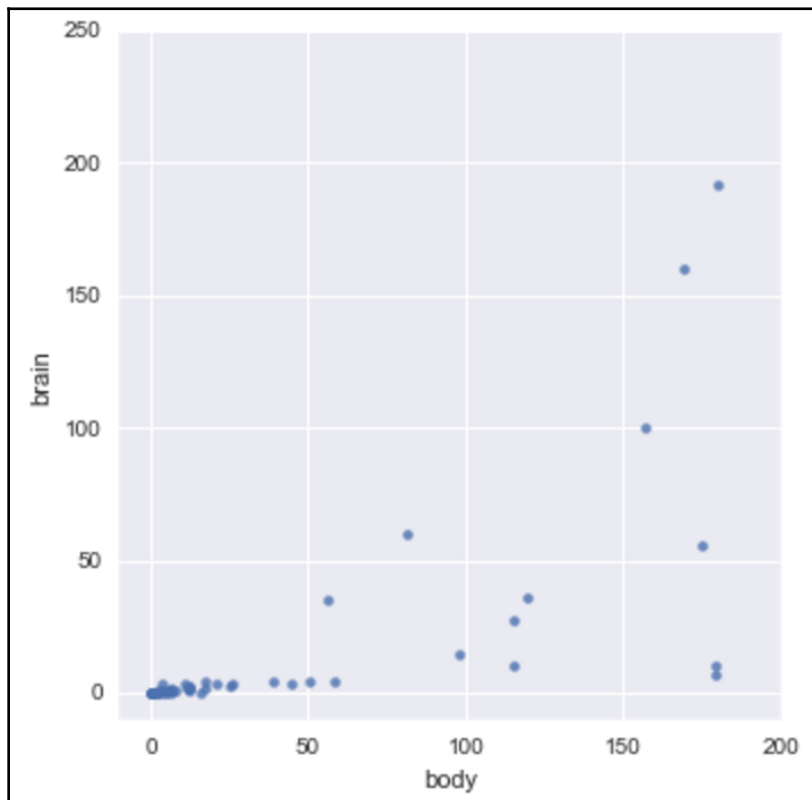


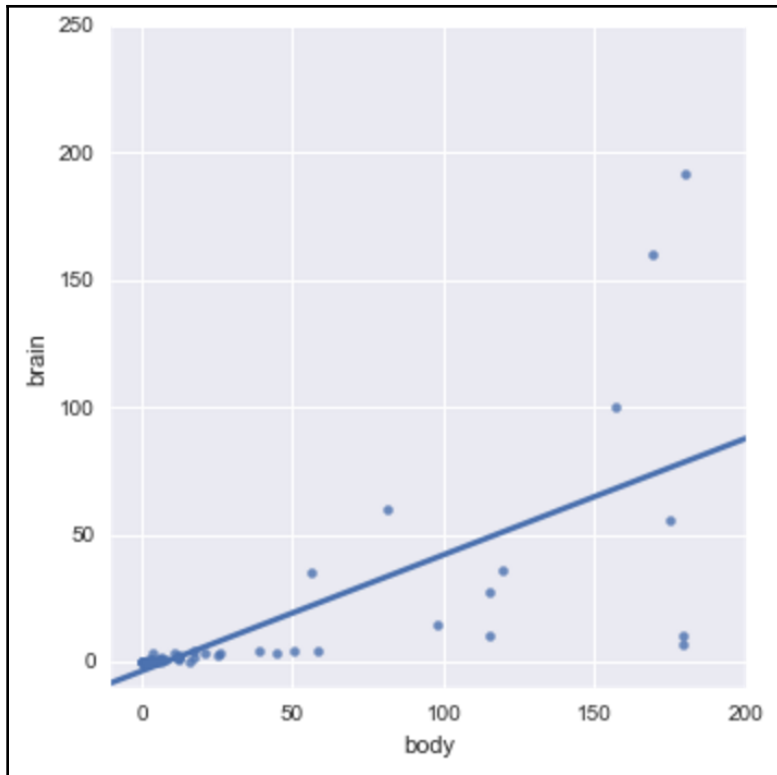




Chapter 12: Beyond the Essentials

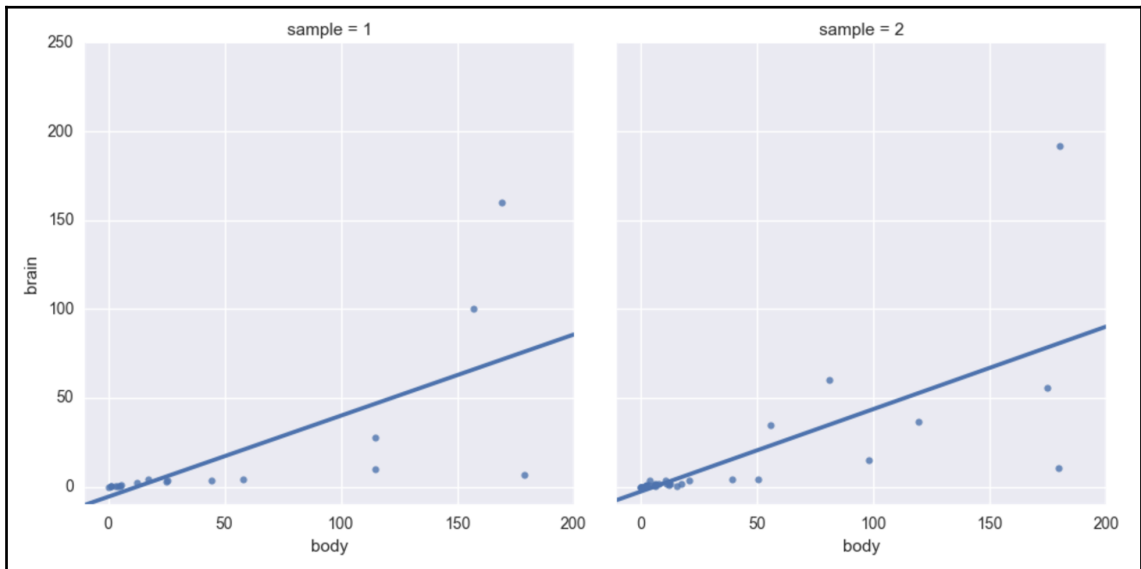
	brain	body
id		
1	3.385	44.5
2	0.480	15.5
3	1.350	8.1
4	465.000	423.0
5	36.330	119.5

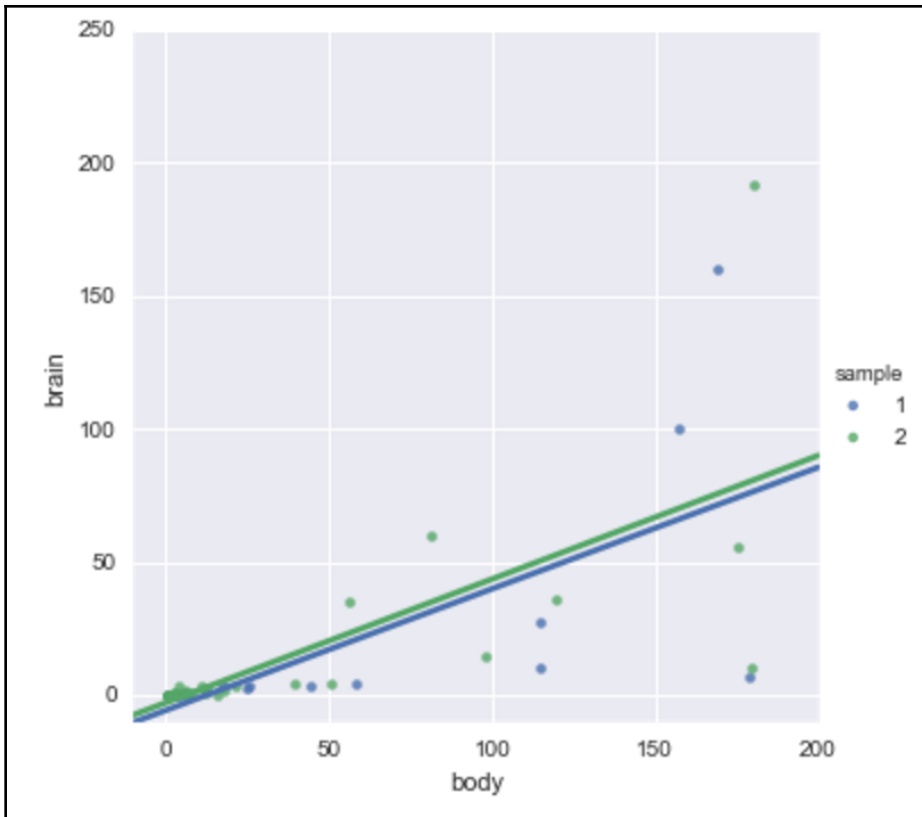


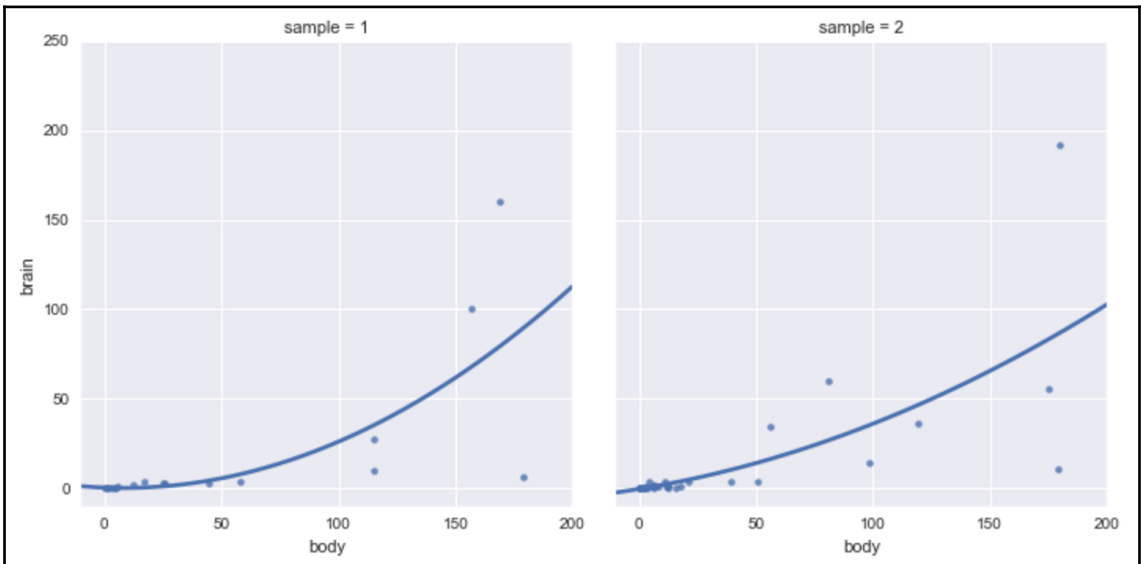
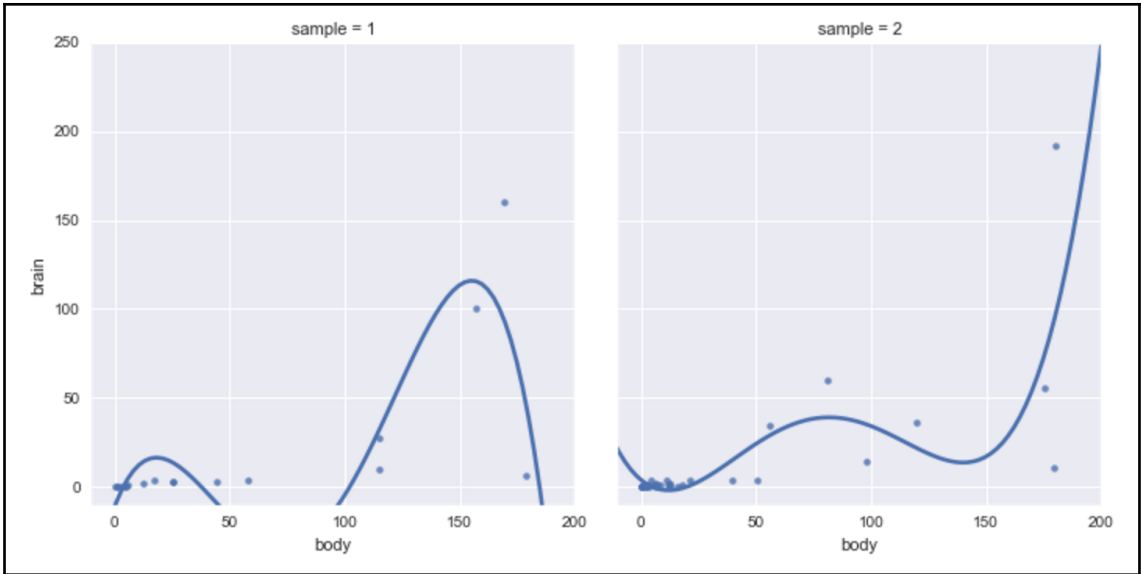


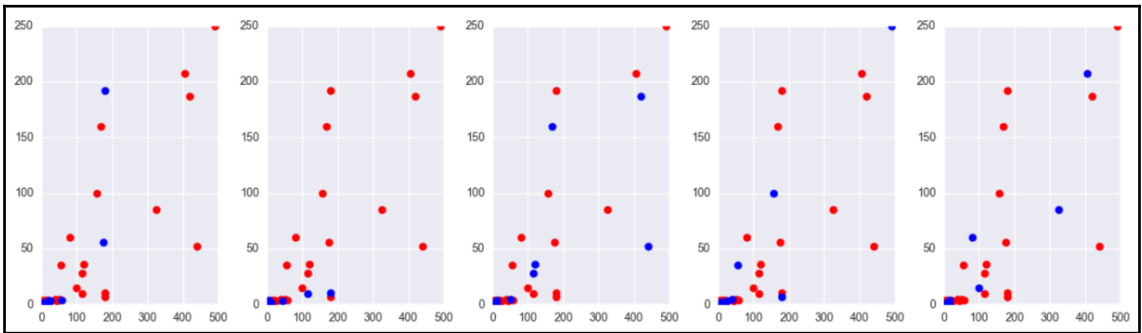
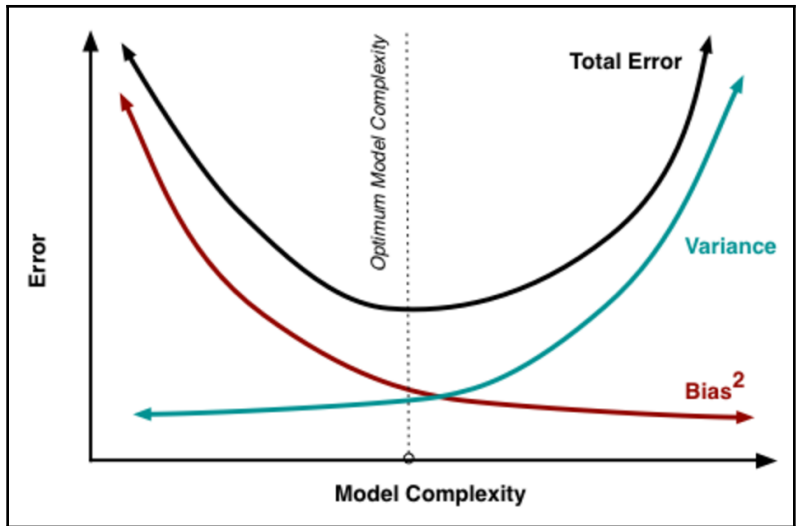
	brain	body	sample
id			
1	3.385	44.5	1
2	0.480	15.5	2
3	1.350	8.1	2
5	36.330	119.5	2
6	27.660	115.0	1

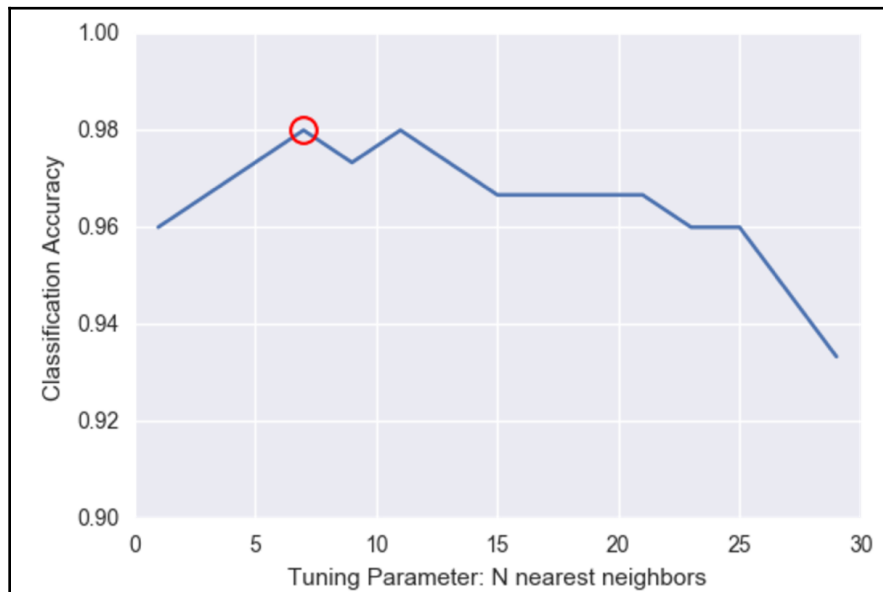
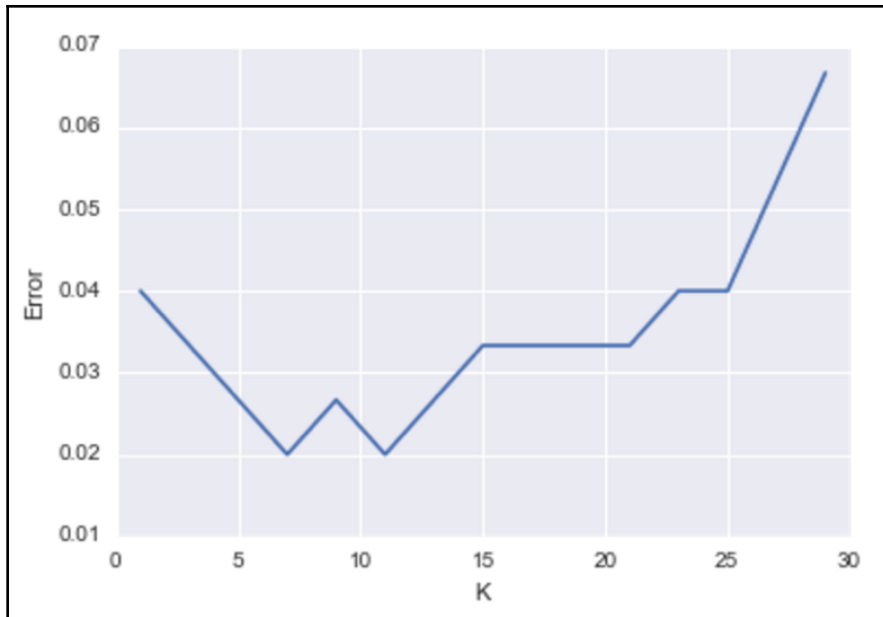
	brain	body
sample		
1	18.113778	52.068889
2	13.323364	34.669091

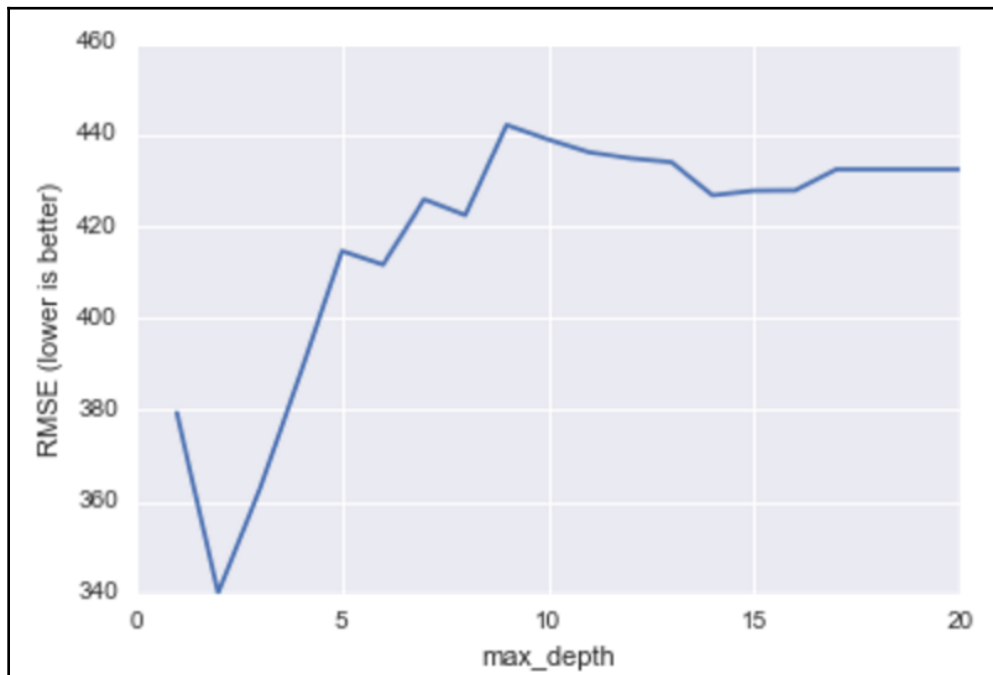
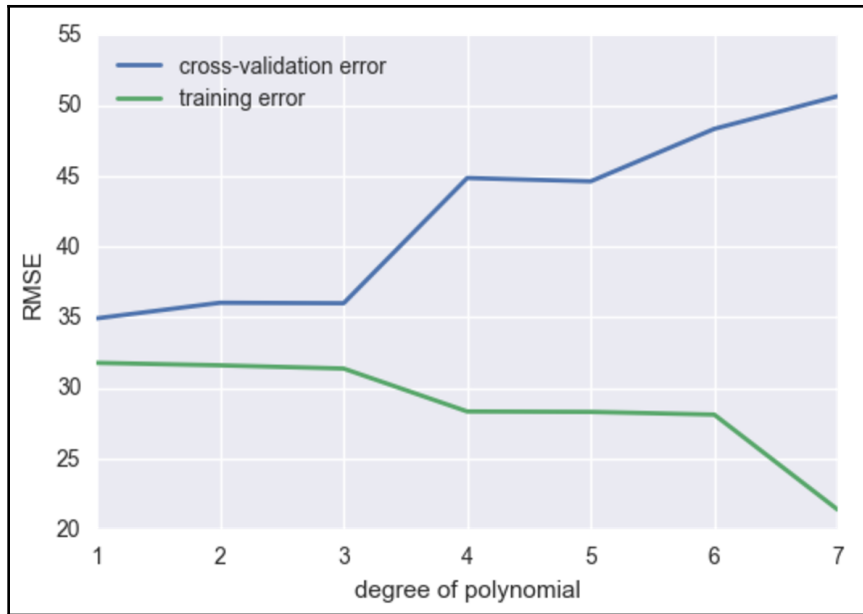


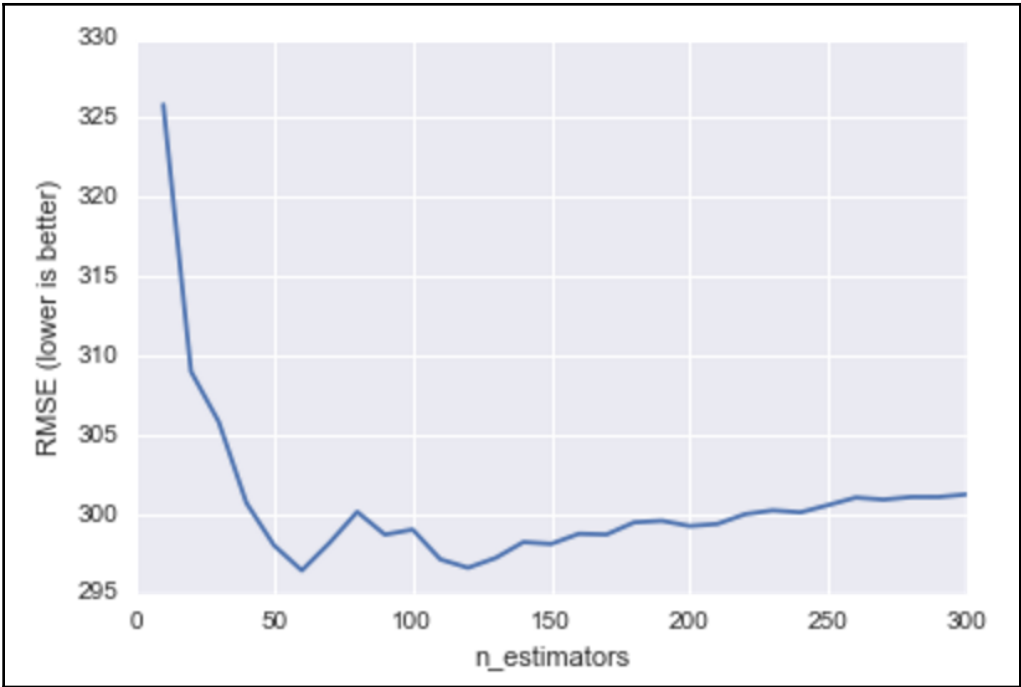




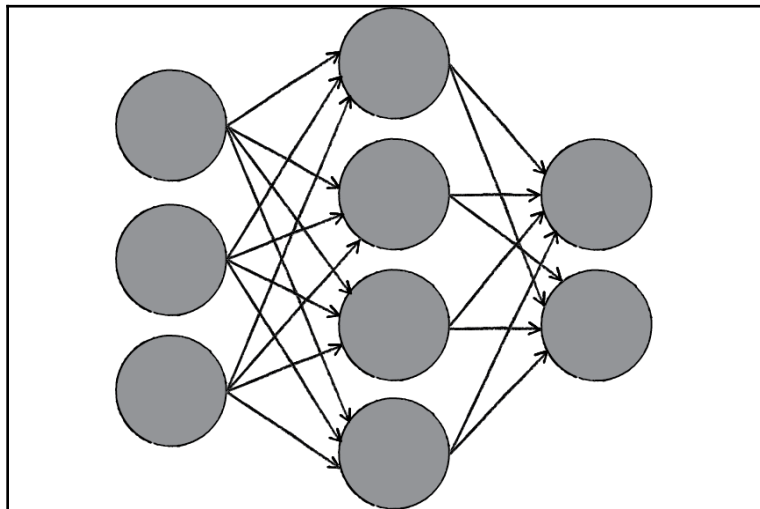


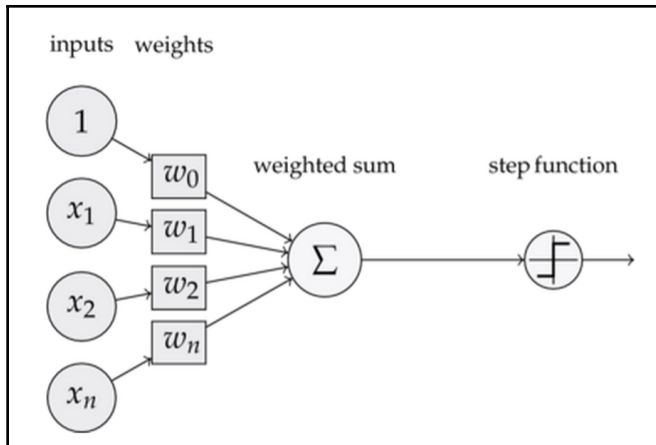






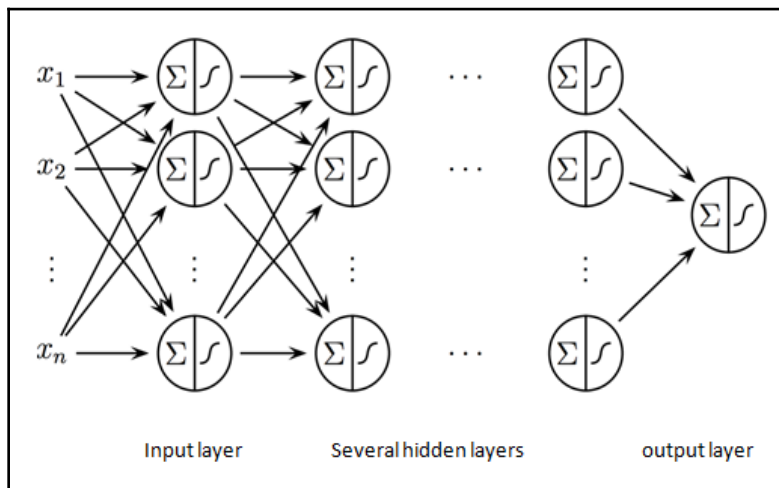
	feature	importance
6	Years	0.263990
5	Walks	0.146786
1	Hits	0.139801
4	RBI	0.136265
0	AtBat	0.091551
9	PutOuts	0.060647
3	Runs	0.057460
2	HmRun	0.040183
11	Errors	0.024711
10	Assists	0.023367
8	Division	0.007628
12	NewLeague	0.004545
7	League	0.003067

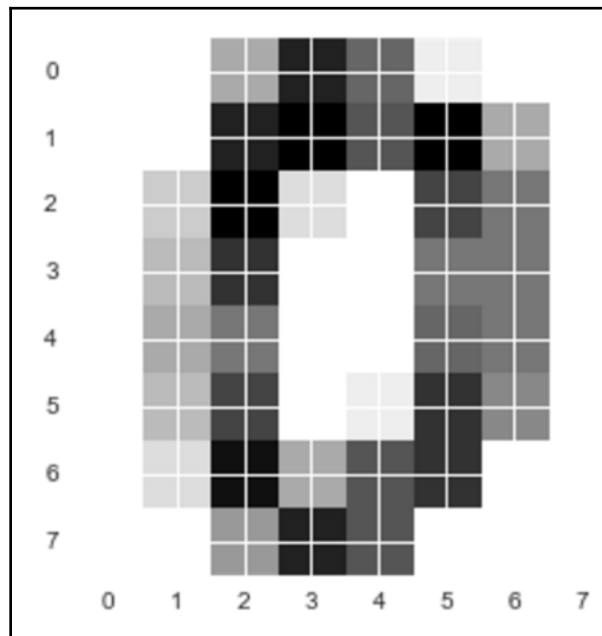
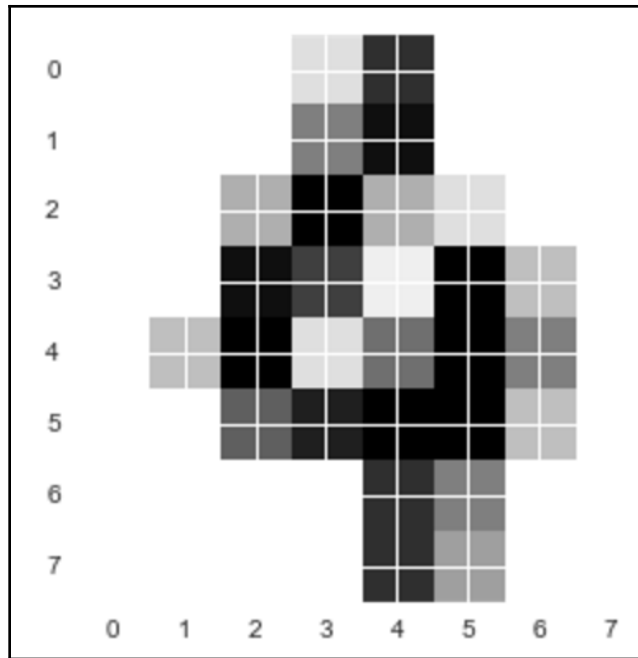




$$f_{log}(z) = \frac{1}{1 + e^{-z}}$$

f_{log} is called **logistic function**





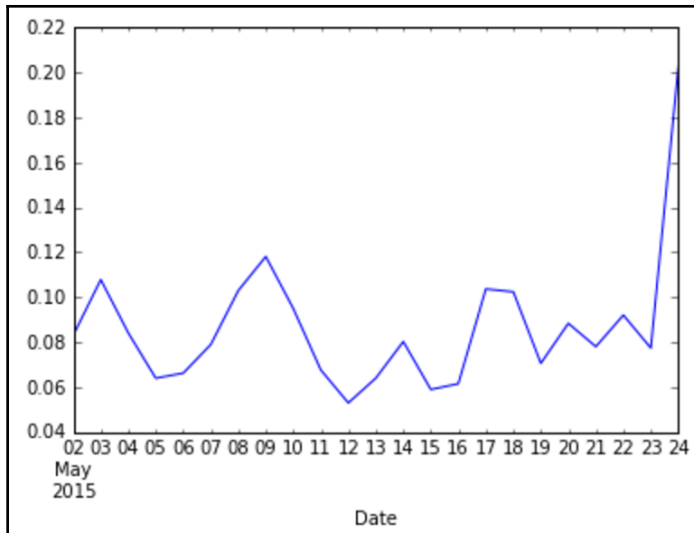
Chapter 13: Case Studies

	Text	Date	Status	Retweet
0	RT @j_o_h_n_danger: \$TWTR now top holding for ...	2015-05-24 03:46:08	602319644234395648	6.022899e+17
1	RT diggingplatinum RT WWalkerWW: iOS 9 vs. And...	2015-05-24 04:17:42	602327586983796737	NaN
2	RT bosocial RT insidermonkey RT j_o_h_n_danger...	2015-05-24 04:13:22	602326499534966784	NaN
3	RT @WWalkerWW: iOS 9 vs. Android M â The New...	2015-05-24 04:08:34	602325288740114432	6.023104e+17
4	RT @seeitmarket: Apple Chart Update: Big Test ...	2015-05-24 04:04:42	602324318903771136	6.023215e+17

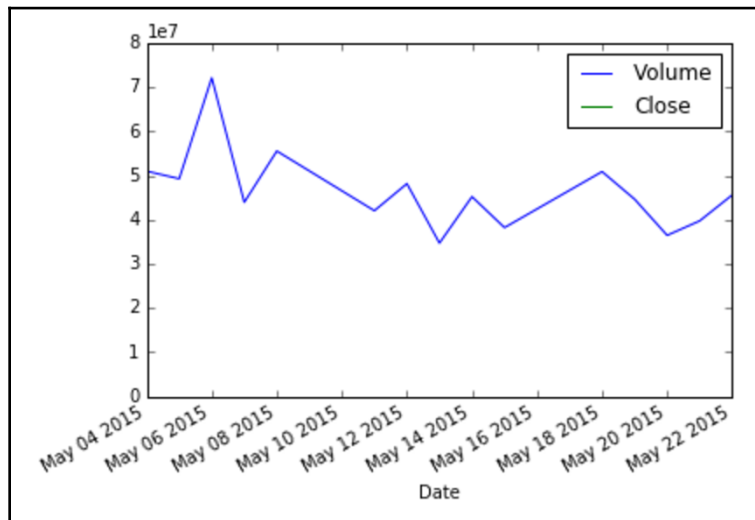
	Text	Date	Status	Retweet	sentiment
0	RT @j_o_h_n_danger: \$TWTR now top holding for ...	2015-05-24 03:46:08	602319644234395648	6.022899e+17	0.500000
1	RT diggingplatinum RT WWalkerWW: iOS 9 vs. And...	2015-05-24 04:17:42	602327586983796737	NaN	0.136364
2	RT bosocial RT insidermonkey RT j_o_h_n_danger...	2015-05-24 04:13:22	602326499534966784	NaN	0.500000
3	RT @WWalkerWW: iOS 9 vs. Android M â The New...	2015-05-24 04:08:34	602325288740114432	6.023104e+17	0.136364
4	RT @seeitmarket: Apple Chart Update: Big Test ...	2015-05-24 04:04:42	602324318903771136	6.023215e+17	0.000000

	yesterday_sentiment	Close	yesterday_close	percent_change_in_price	change_close_big_deal
Date					
2015-05-05	0.084062	125.800003	128.699997	-0.022533	True
2015-05-06	0.063882	125.010002	125.800003	-0.006280	False
2015-05-07	0.066166	125.260002	125.010002	0.002000	False
2015-05-08	0.078892	127.620003	125.260002	0.018841	True
2015-05-11	0.102898	126.320000	127.620003	-0.010187	True

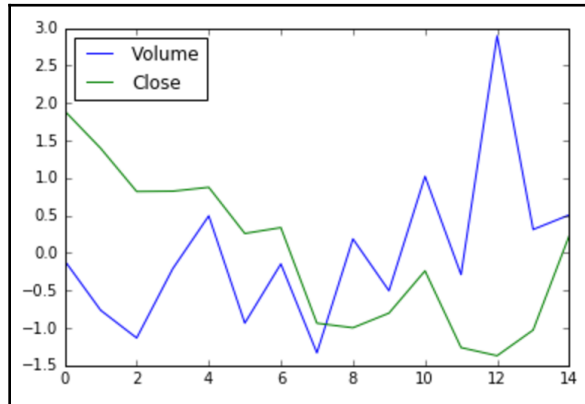
	sentiment
Date	
2015-05-02	0.083031
2015-05-03	0.107789
2015-05-04	0.084062
2015-05-05	0.063882
2015-05-06	0.066166



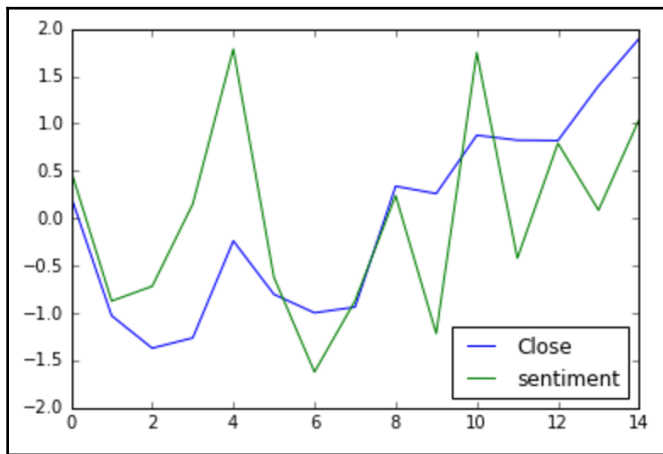
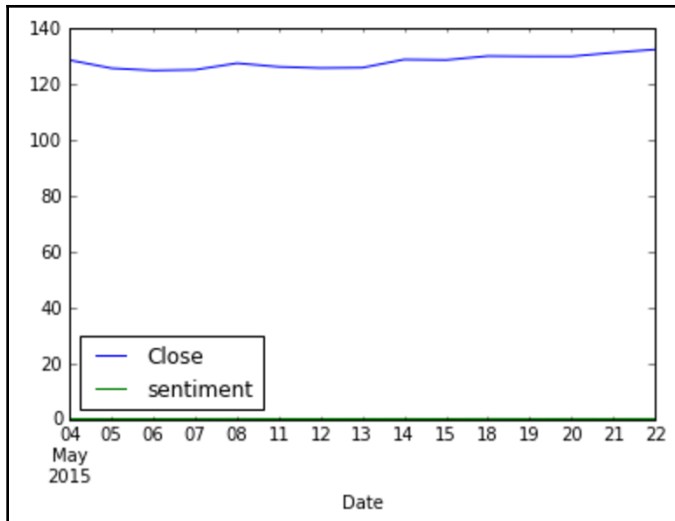
	Adj_Close	Close	Date	High	Low	Open	Symbol	Volume
0	129.180748	132.539993	2015-05-22	132.970001	131.399994	131.600006	AAPL	45596000
1	128.059901	131.389999	2015-05-21	131.630005	129.830002	130.070007	AAPL	39730400
2	126.763608	130.059998	2015-05-20	130.979996	129.339996	130.00	AAPL	36454900
3	126.773364	130.070007	2015-05-19	130.880005	129.639999	130.690002	AAPL	44633200
4	126.890318	130.190002	2015-05-18	130.720001	128.360001	128.380005	AAPL	50882900



	Volume	Close
count	1.500000e+01	15.000000
mean	4.649939e+07	128.170667
std	9.167054e+06	2.386393
min	3.469420e+07	125.010002
25%	4.088310e+07	125.940002
50%	4.520350e+07	128.699997
75%	5.007715e+07	130.065002
max	7.214100e+07	132.539993



	sentiment
Date	
2015-05-02	0.083031
2015-05-03	0.107789
2015-05-04	0.084062
2015-05-05	0.063882
2015-05-06	0.066166



	Close	sentiment	yesterday_sentiment
Date			
2015-05-04	128.699997	0.084062	NaN
2015-05-05	125.800003	0.063882	0.084062
2015-05-06	125.010002	0.066166	0.063882
2015-05-07	125.260002	0.078892	0.066166
2015-05-08	127.620003	0.102898	0.078892

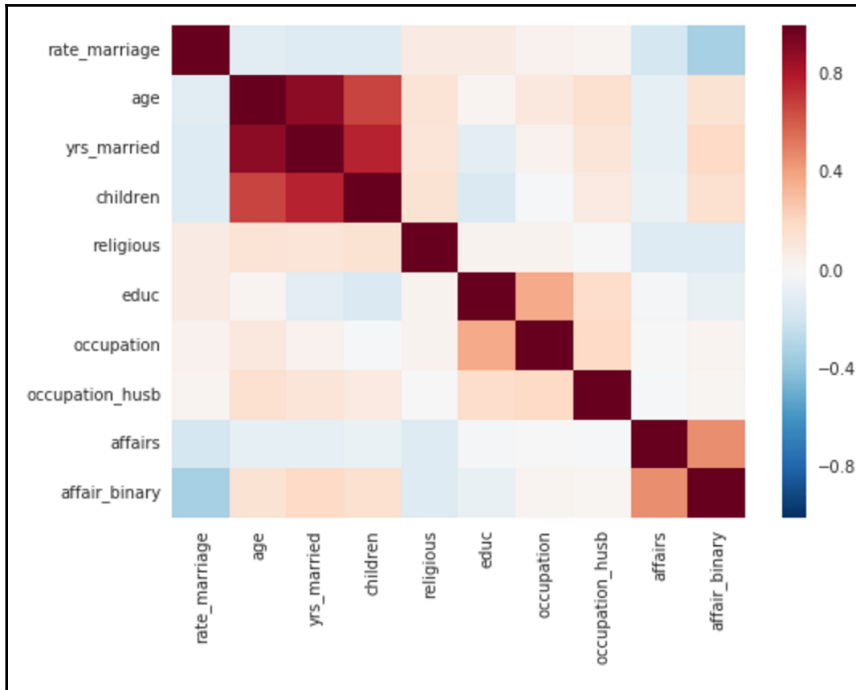
	yesterday_sentiment	Close
Date		
2015-05-05	0.084062	125.800003
2015-05-06	0.063882	125.010002
2015-05-07	0.066166	125.260002
2015-05-08	0.078892	127.620003
2015-05-11	0.102898	126.320000

$$\text{change_close_big_deal} = \begin{cases} 1, & \text{the change in stock price was } > 1\% \text{ or } < -1\% \\ 0, & \text{else} \end{cases}$$

	yesterday_sentiment	Close	yesterday_close	percent_change_in_price	change_close_big_deal
Date					
2015-05-05	0.084062	125.800003	128.699997	-0.022533	True
2015-05-06	0.063882	125.010002	125.800003	-0.006280	False
2015-05-07	0.066166	125.260002	125.010002	0.002000	False
2015-05-08	0.078892	127.620003	125.260002	0.018841	True
2015-05-11	0.102898	126.320000	127.620003	-0.010187	True

	rate_marriage	age	yrs_married	children	religious	educ	occupation	occupation_husb	affairs
0	3.0	32.0	9.0	3.0	3.0	17.0	2.0	5.0	0.111111
1	3.0	27.0	13.0	3.0	1.0	14.0	3.0	4.0	3.230769
2	4.0	22.0	2.5	0.0	1.0	16.0	3.0	5.0	1.400000
3	4.0	37.0	16.5	4.0	3.0	16.0	5.0	5.0	0.727273
4	5.0	27.0	9.0	1.0	1.0	14.0	3.0	4.0	4.666666

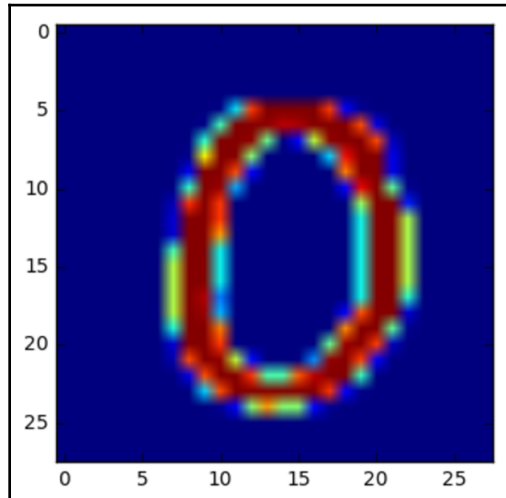
	rate_marriage	age	yrs_married	children	religious	educ	occupation	occupation_husb	affairs	affair_binary
rate_marriage	1.000000	-0.111127	-0.128978	-0.129161	0.078794	0.079869	0.039528	0.027745	-0.178068	-0.331776
age	-0.111127	1.000000	0.894082	0.673902	0.136598	0.027960	0.106127	0.162567	-0.089964	0.146519
yrs_married	-0.128978	0.894082	1.000000	0.772806	0.132683	-0.109058	0.041782	0.128135	-0.087737	0.203109
children	-0.129161	0.673902	0.772806	1.000000	0.141845	-0.141918	-0.015068	0.086660	-0.070278	0.159833
religious	0.078794	0.136598	0.132683	0.141845	1.000000	0.032245	0.035746	0.004061	-0.125933	-0.129299
educ	0.079869	0.027960	-0.109058	-0.141918	0.032245	1.000000	0.382286	0.183932	-0.017740	-0.075280
occupation	0.039528	0.106127	0.041782	-0.015068	0.035746	0.382286	1.000000	0.201156	0.004469	0.028981
occupation_husb	0.027745	0.162567	0.128135	0.086660	0.004061	0.183932	0.201156	1.000000	-0.015614	0.017637
affairs	-0.178068	-0.089964	-0.087737	-0.070278	-0.125933	-0.017740	0.004469	-0.015614	1.000000	0.464046
affair_binary	-0.331776	0.146519	0.203109	0.159833	-0.129299	-0.075280	0.028981	0.017637	0.464046	1.000000

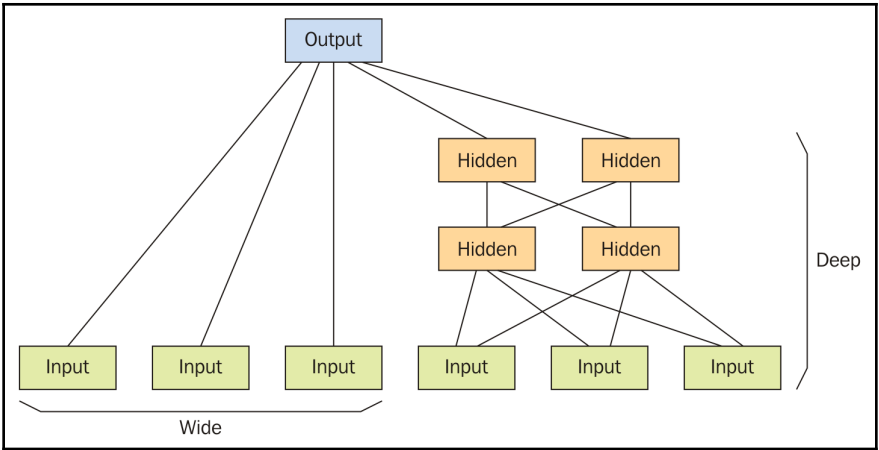


	feature	importance
2	yrs_married	0.136953
0	rate_marriage	0.142588
7	occupation_husb	0.173304

educ	occupation	occupation_husb	affairs	affair_binary	occ_2.0	occ_3.0	occ_4.0	occ_5.0	occ_6.0
17.0	2.0	5.0	0.111111	True	1.0	0.0	0.0	0.0	0.0
14.0	3.0	4.0	3.230769	True	0.0	1.0	0.0	0.0	0.0
16.0	3.0	5.0	1.400000	True	0.0	1.0	0.0	0.0	0.0
16.0	5.0	5.0	0.727273	True	0.0	0.0	0.0	1.0	0.0
14.0	3.0	4.0	4.666666	True	0.0	1.0	0.0	0.0	0.0

	feature	importance
15	occ_husb_6.0	0.024299
11	occ_husb_2.0	0.030418
14	occ_husb_5.0	0.042021
13	occ_husb_4.0	0.047874
4	religious	0.098630
1	age	0.111628
5	educ	0.131468
2	yrs_married	0.132034
3	children	0.134374
0	rate_marriage	0.139502







Chapter 14: Microsoft Azure Databricks

Create Cluster

New Cluster | **0 Workers:** 0.0 GB Memory, 0 Cores, 0 DBU
1 Driver: 6.0 GB Memory, 0.88 Cores, 1 DBU

Cluster Name

Databricks Runtime Version 

Python Version 

Create Notebook













Name




Language

Cluster

1	dteday	season	yr	mnth	hr	holiday	weekday	workingday	weathersi	temp	atemp	hum	windspeed	casual	registere	cnt
2	1/1/2011	1	0	1	0	0	6	0	1	0.24	0.2879	0.81	0	3	13	16
3	1/1/2011	1	0	1	1	0	6	0	1	0.22	0.2727	0.8	0	8	32	40
4	1/1/2011	1	0	1	2	0	6	0	1	0.22	0.2727	0.8	0	5	27	32
5	1/1/2011	1	0	1	3	0	6	0	1	0.24	0.2879	0.75	0	3	10	13
6	1/1/2011	1	0	1	4	0	6	0	1	0.24	0.2879	0.75	0	0	1	1
7	1/1/2011	1	0	1	5	0	6	0	2	0.24	0.2576	0.75	0.0896	0	1	1
8	1/1/2011	1	0	1	6	0	6	0	1	0.22	0.2727	0.8	0	2	0	2
9	1/1/2011	1	0	1	7	0	6	0	1	0.2	0.2576	0.86	0	1	2	3
10	1/1/2011	1	0	1	8	0	6	0	1	0.24	0.2879	0.75	0	1	7	8
11	1/1/2011	1	0	1	9	0	6	0	1	0.32	0.3485	0.76	0	8	6	14
12	1/1/2011	1	0	1	10	0	6	0	1	0.38	0.3939	0.76	0.2537	12	24	36
13	1/1/2011	1	0	1	11	0	6	0	1	0.36	0.3333	0.81	0.2836	26	30	56
14	1/1/2011	1	0	1	12	0	6	0	1	0.42	0.4242	0.77	0.2836	29	55	84
15	1/1/2011	1	0	1	13	0	6	0	2	0.46	0.4545	0.72	0.2985	47	47	94
16	1/1/2011	1	0	1	14	0	6	0	2	0.46	0.4545	0.72	0.2836	35	71	106
17	1/1/2011	1	0	1	15	0	6	0	2	0.44	0.4394	0.77	0.2985	40	70	110
18	1/1/2011	1	0	1	16	0	6	0	2	0.42	0.4242	0.82	0.2985	41	52	93

2011-01-01	1	0	1	2	0
2011-01-01	1	0	1	3	0
2011-					0
2011-					0
2011-					0
2011-					0
2011-					0

-  Bar
-  Scatter
-  Map
-  Line
-  Area
-  Pie
-  Quantile
-  Histogram
-  Box plot
-  Q-Q plot
-  Pivot
-  Line (deprecated)

Show




Customize Plot

All fields:

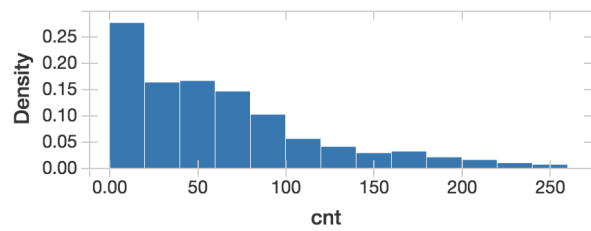
- weekday
- workingday
- weathersit
- temp
- atemp
- hum
- windspeed
- casual
- registered
- cnt
- <id>

Keys:

Series groupings:

Values:

cnt ✕



Density

cnt

Showing sample based on the first 1000 rows. Apply to aggregate over all results. ?

Number of bins

20

Aggregation: SUM

Display type: Histogram plot

Cancel
Apply

Customize Plot

All fields:

- dteday
- season
- yr
- mnth
- hr
- holiday
- weekday
- workingday
- weathersit
- temp
- atemp

Keys:

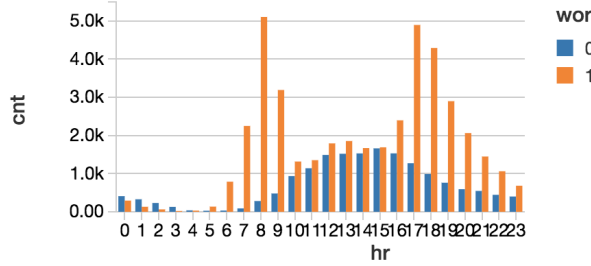
hr ✕

Series groupings:

workin... ✕

Values:

cnt ✕



cnt

hr

workin

- 0
- 1

Showing sample based on the first 1000 rows. Apply to aggregate over all results. ?

Grouped

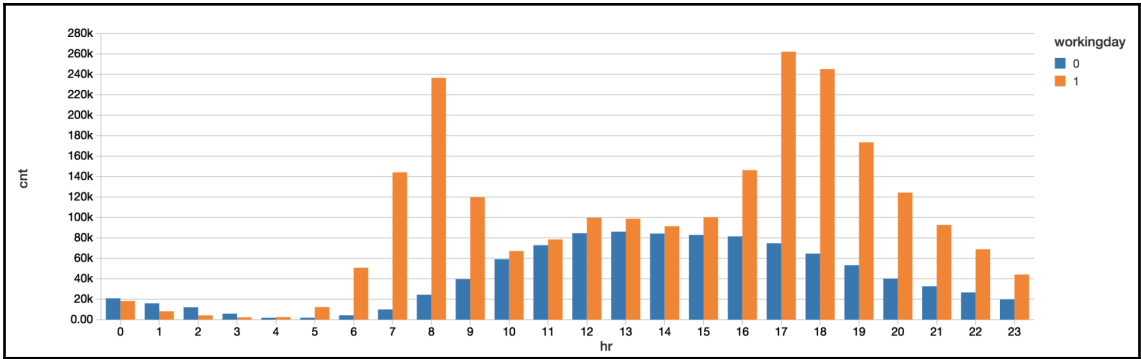
Stacked

100% Stacked

Aggregation: SUM

Display type: Bar chart

Cancel
Apply



New Library

Language

Install PyPi Package

You can specify a package name with an optional [version specification](#)

PyPi Name