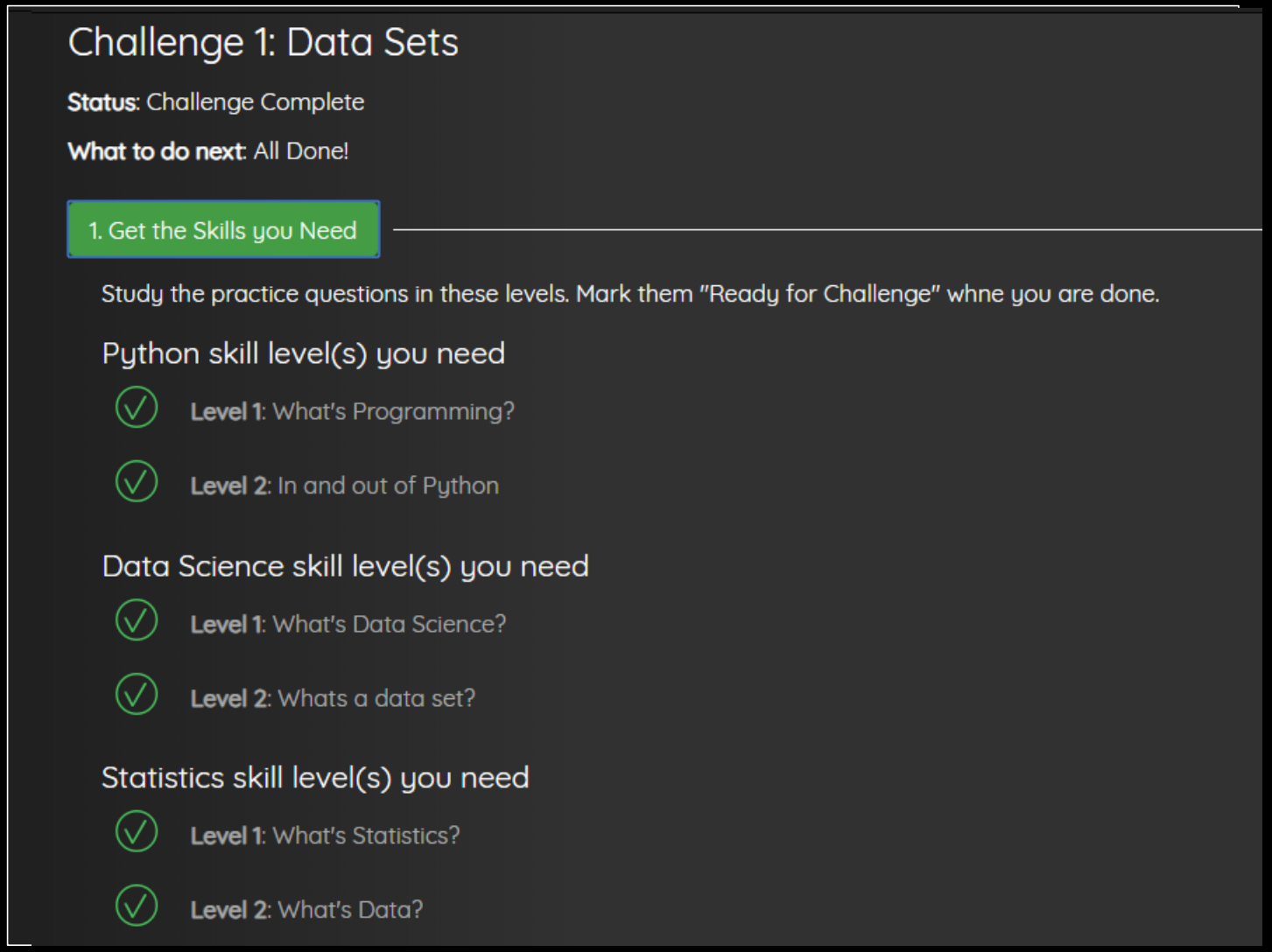
Intro to Challenge 1

Data Sets

What you need to know



What you do

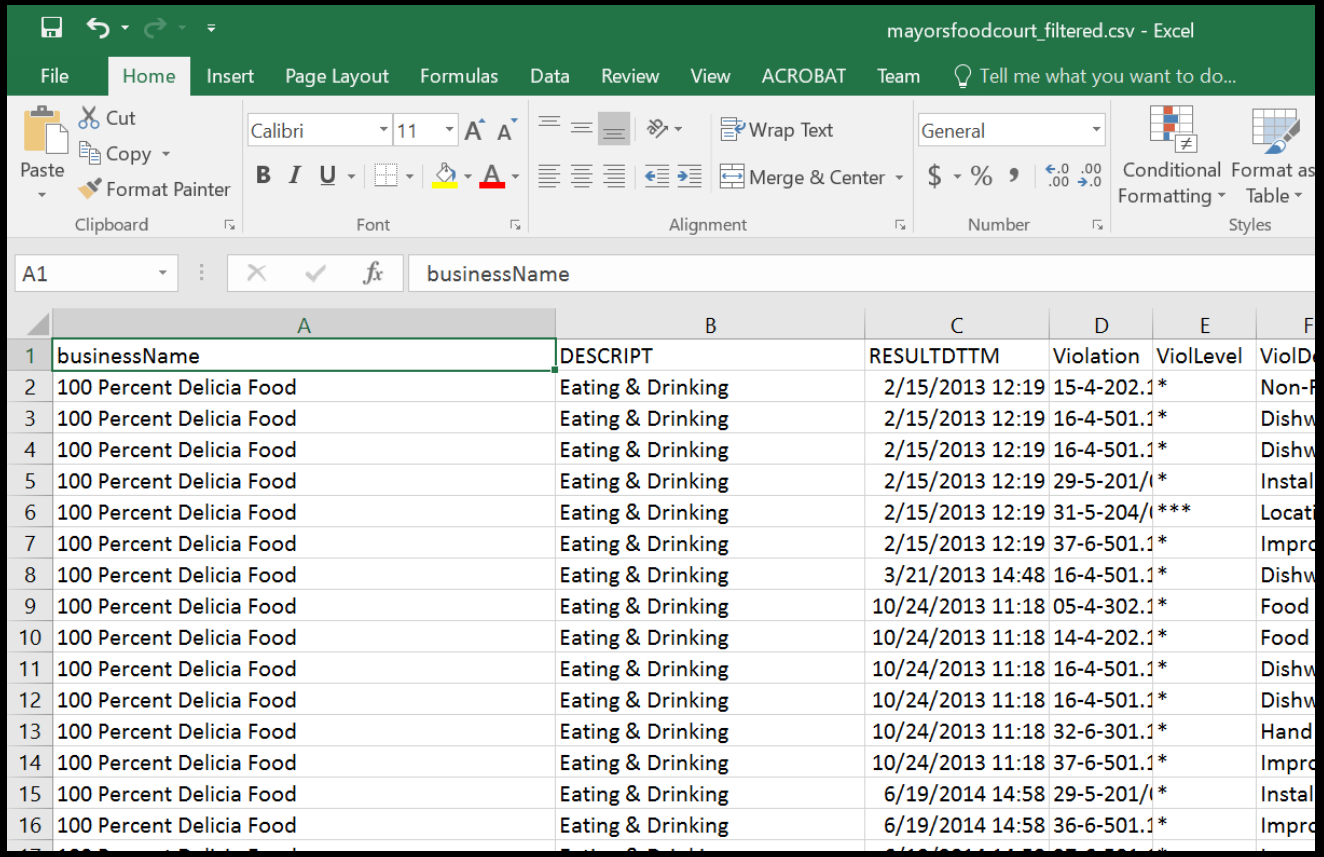
1. Download and open a data set
2. Talk about the social context of the data
3. Talk about the data itself
4. Do a data analysis to find answers

Key Data Science Points

- What is in data science that is not in stats and python?
- What is a data context?
- What is a data set?

Key Stats Points

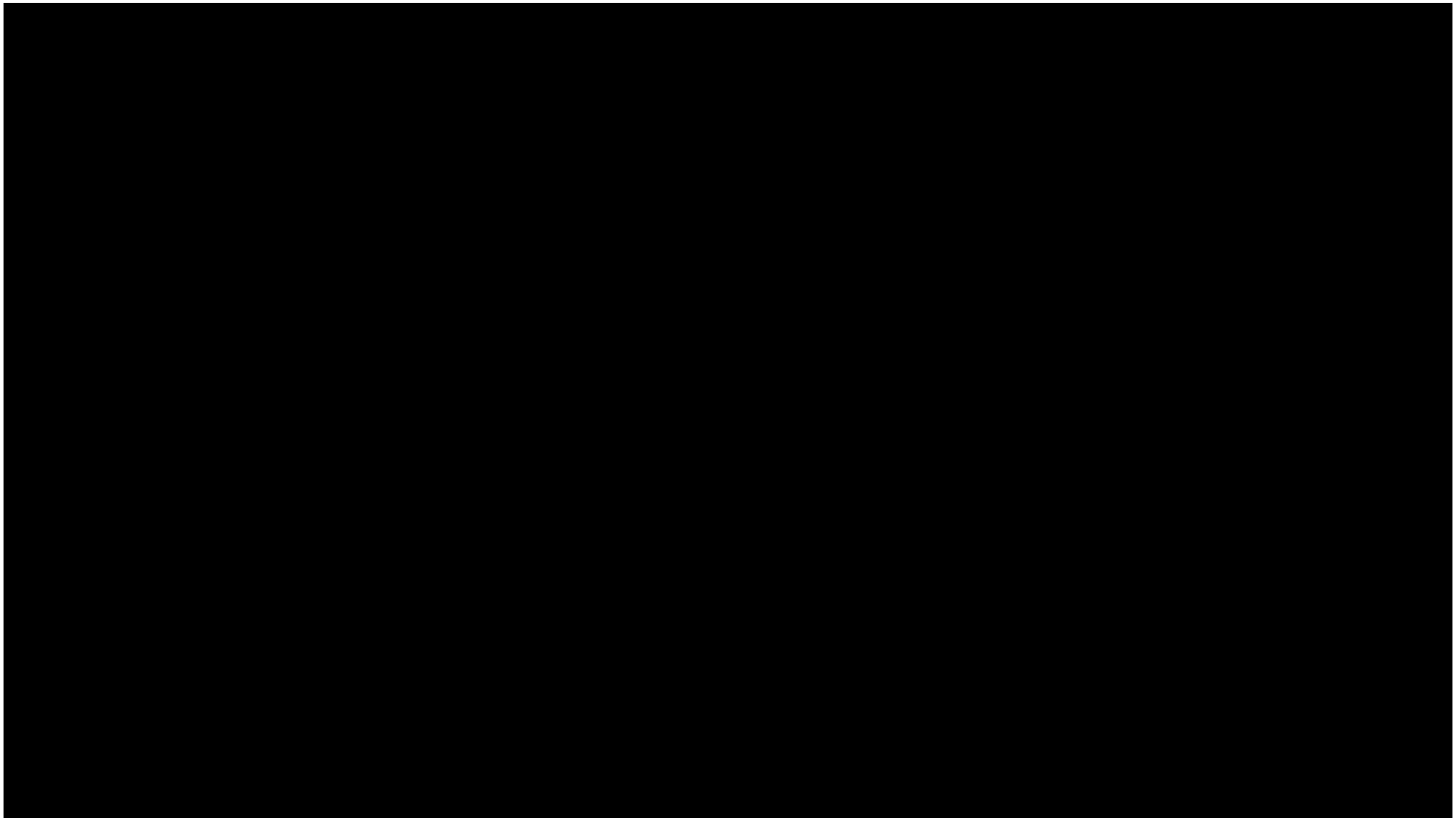
- What is Data?
- What's a data type?
- What is a table
- Row
- Column



	A	B	C	D	E	F
1	businessName	DESCRIPT	RESULTDTTM	Violation	ViolLevel	ViolD
2	100 Percent Delicia Food	Eating & Drinking	2/15/2013 12:19	15-4-202.1	*	Non-F
3	100 Percent Delicia Food	Eating & Drinking	2/15/2013 12:19	16-4-501.1	*	Dishw
4	100 Percent Delicia Food	Eating & Drinking	2/15/2013 12:19	16-4-501.1	*	Dishw
5	100 Percent Delicia Food	Eating & Drinking	2/15/2013 12:19	29-5-201/(	*	Instal
6	100 Percent Delicia Food	Eating & Drinking	2/15/2013 12:19	31-5-204/(	***	Locati
7	100 Percent Delicia Food	Eating & Drinking	2/15/2013 12:19	37-6-501.1	*	Impro
8	100 Percent Delicia Food	Eating & Drinking	3/21/2013 14:48	16-4-501.1	*	Dishw
9	100 Percent Delicia Food	Eating & Drinking	10/24/2013 11:18	05-4-302.1	*	Food
10	100 Percent Delicia Food	Eating & Drinking	10/24/2013 11:18	14-4-202.1	*	Food
11	100 Percent Delicia Food	Eating & Drinking	10/24/2013 11:18	16-4-501.1	*	Dishw
12	100 Percent Delicia Food	Eating & Drinking	10/24/2013 11:18	16-4-501.1	*	Dishw
13	100 Percent Delicia Food	Eating & Drinking	10/24/2013 11:18	32-6-301.1	*	Hand
14	100 Percent Delicia Food	Eating & Drinking	10/24/2013 11:18	37-6-501.1	*	Impro
15	100 Percent Delicia Food	Eating & Drinking	6/19/2014 14:58	29-5-201/(	*	Instal
16	100 Percent Delicia Food	Eating & Drinking	6/19/2014 14:58	36-6-501.1	*	Impro

Key Python Points

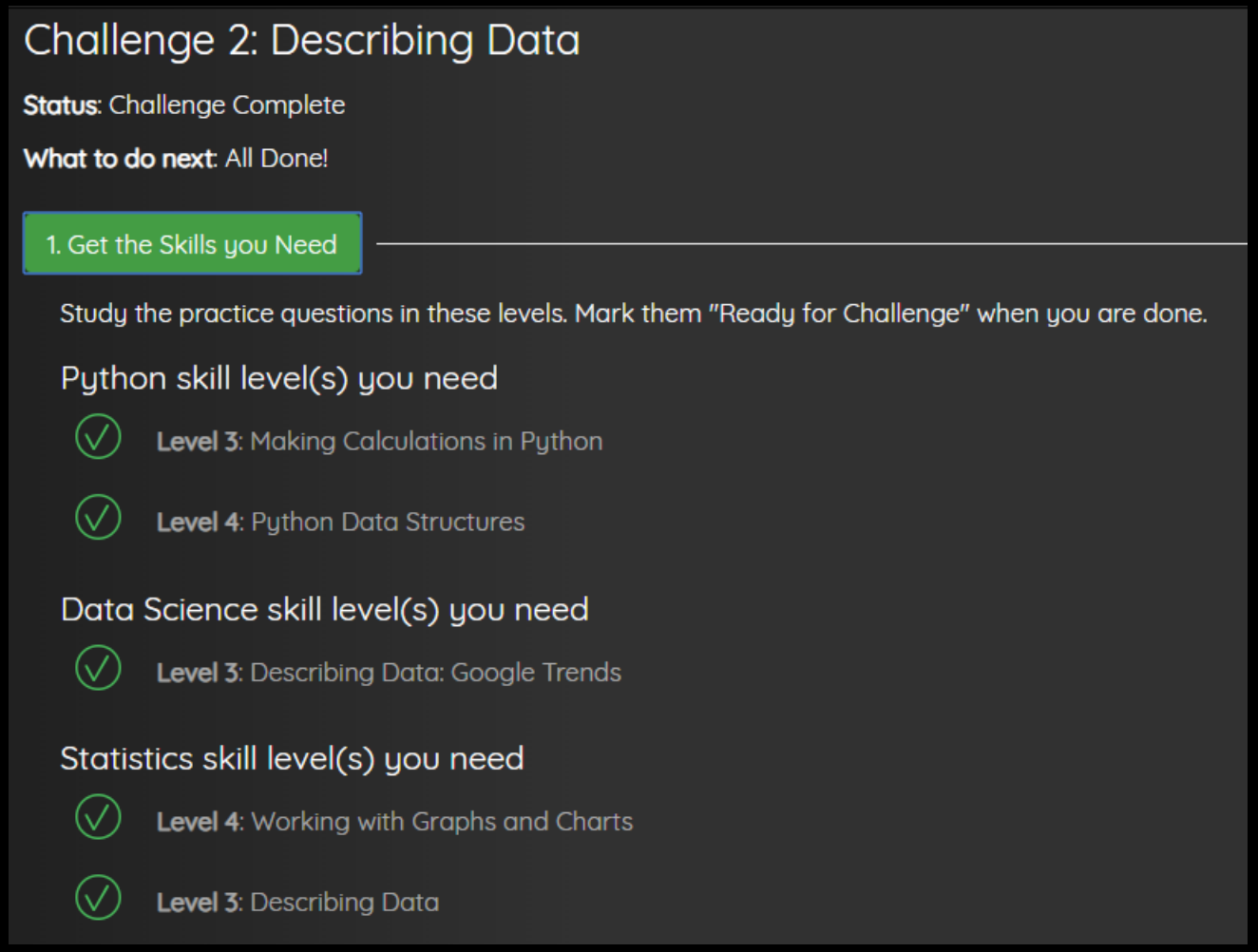
- import pandas as pd
- dataFrame = pd.read_csv('', sep="")
- print()
- dataFrame.head()



Intro to Challenge 2

Describing Data

What you need to know



What you do

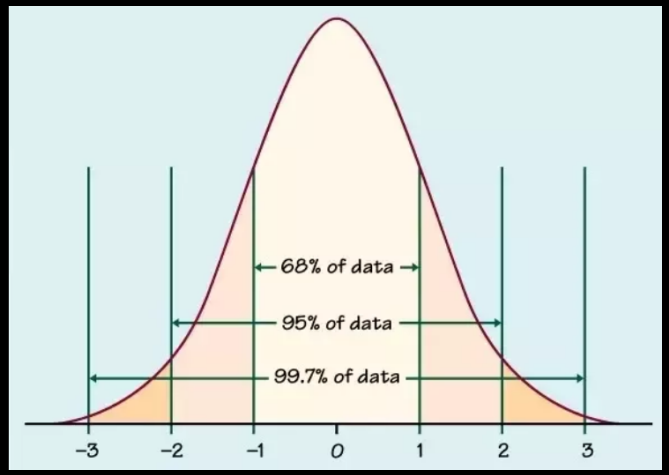
1. Download and open a data set
2. Talk about the social context of the data
3. Talk about the data itself
4. Do a data analysis to find answers

Key Data Science Points

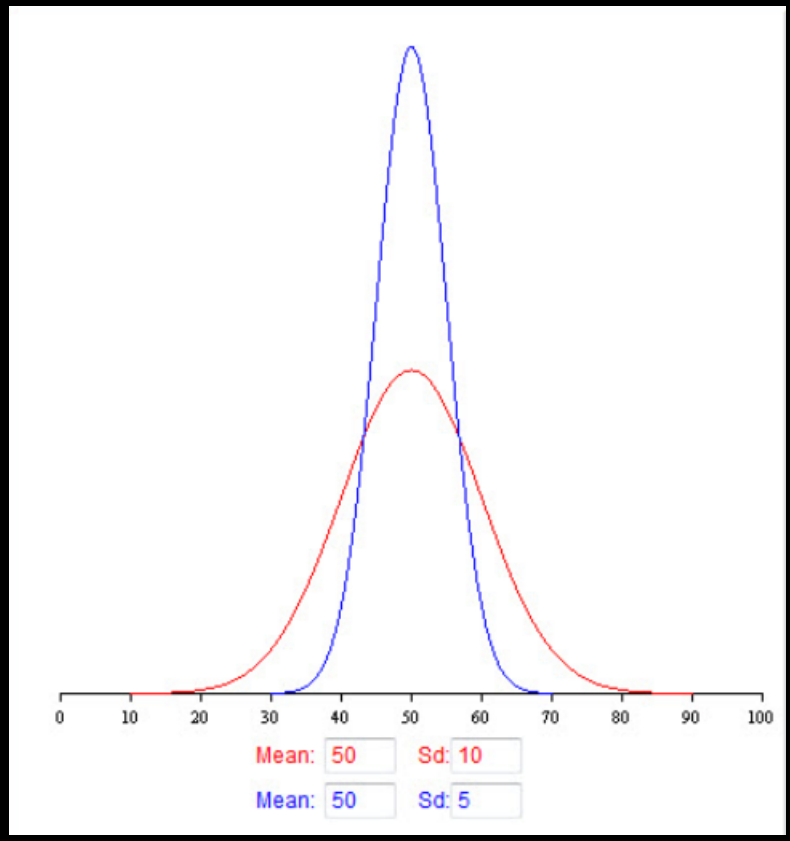
- Bias
- Normalized
- Anonymized
- Missing values
- Google Trends

Key Stats Points

- Central tendency
- Mean
- Standard deviation



Quora



OnlineStatBook.com

Key Python Points

- df = pd.read_csv('')
- meanAsNumber = df[''].mean()
- meanAsMoney = "${???}".format(meanAsNumber)