

Shark Tank

Shark Tank is a reality TV show. Contestants present their idea for a company to a panel of investors (a.k.a. "sharks"), who then decide whether or not to invest in that company. The investors give a certain amount of money in exchange for a percentage stake in the company ("equity"). If you are not familiar with the show, you may want to watch part of an episode [here](#) to get a sense of how it works. You can also search for a clip on YouTube.

The data that you will examine in this lab contains data about all contestants from the first 6 seasons of the show, including:

- the name and industry of the proposed company
- whether or not it was funded (i.e., the "Deal" column)
- which sharks chose to invest in the venture (N.B. There are 7 regular sharks, not including "Guest". Each shark has a column in the data set, labeled by their last name.)
- if funded, the amount of money the sharks put in and the percentage equity they got in return

To earn full credit on this lab, you should:

- use built-in `pandas` methods (like `.sum()` and `.max()`) instead of writing a for loop over a `DataFrame` or `Series`
- use the split-apply-combine pattern wherever possible

Of course, if you can't think of a vectorized solution, a `for` loop is still better than no solution at all!

```
In [1]: import pandas as pd
```

Question 0. Getting and Cleaning the Data

The data is stored in the CSV file `sharktank.csv`. Read in the data into a `Pandas DataFrame`.

```
In [2]: # YOUR CODE HERE
df = pd.read_csv("sharktank.csv")
df
```

	Season	No. in series	Company	Deal	Industry	Entrepreneur	Gender	Amount	Equity	Corcoran	Cuban	Greiner	Herjavec	John	O'Leary	Harrington	Guest	Details / Notes
0	1.0	1.0	Ava the Elephant	Yes	Healthcare		Female	\$50,000	55%	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	1.0	1.0	Mr. Tod's Pie Factory	Yes	Food and Beverage		Male	\$460,000	50%	1.0	NaN	NaN	NaN	NaN	1.0	NaN	NaN	NaN
2	1.0	1.0	Wispos	No	Business Services		Male	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	1.0	1.0	College Foxes Packing Boxes	No	Lifestyle / Home		Male	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	1.0	1.0	Ionic Ear	No	Uncertain / Other		Male	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...
490	6.0	28.0	You Kick Ass	Yes	Children / Education		Female	\$100,000	10%	NaN	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
491	6.0	29.0	Shark Wheel	Yes	Fitness / Sports		Male	\$225,000	8%	NaN	1.0	NaN	NaN	NaN	NaN	NaN	1.0	10% royalty until \$500K; then converts to 5% e...
492	6.0	29.0	Gato Cafe	No	Uncertain / Other		Female	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
493	6.0	29.0	Sway Motorsports	Yes	Green/CleanTech		Male	\$300,000	20%	NaN	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
494	6.0	29.0	Spikeball	Yes	Fitness / Sports		Male	\$500,000	20%	NaN	NaN	NaN	NaN	NaN	1.0	NaN	NaN	NaN

495 rows × 17 columns

There is one column for each of the sharks. A 1 indicates that they chose to invest in that company, while a missing value indicates that they did not choose to invest in that company. Notice that these missing values show up as NaNs when we read in the data. Fill in these missing values with zeros. Other columns may also contain NaNs; be careful not to fill those columns with zeros, or you may end up with strange results down the line.

```
In [3]: # YOUR CODE HERE
df.loc[:, "Corcoran":"Guest"] = df.loc[:, "Corcoran":"Guest"].fillna(0)
```

	Season	No. in series	Company	Deal	Industry	Entrepreneur	Gender	Amount	Equity	Corcoran	Cuban	Greiner	Herjavec	John	O'Leary	Harrington	Guest	Details / Notes
0	1.0	1.0	Ava the Elephant	Yes	Healthcare		Female	\$50,000	55%	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN
1	1.0	1.0	Mr. Tod's Pie Factory	Yes	Food and Beverage		Male	\$460,000	50%	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	NaN
2	1.0	1.0	Wispos	No	Business Services		Male	NaN	NaN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN
3	1.0	1.0	College Foxes Packing Boxes	No	Lifestyle / Home		Male	NaN	NaN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN
4	1.0	1.0	Ionic Ear	No	Uncertain / Other		Male	NaN	NaN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN
...
490	6.0	28.0	You Kick Ass	Yes	Children / Education		Female	\$100,000	10%	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN
491	6.0	29.0	Shark Wheel	Yes	Fitness / Sports		Male	\$225,000	8%	0.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	10% royalty until \$500K; then converts to 5% e...
492	6.0	29.0	Gato Cafe	No	Uncertain / Other		Female	NaN	NaN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN
493	6.0	29.0	Sway Motorsports	Yes	Green/CleanTech		Male	\$300,000	20%	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN
494	6.0	29.0	Spikeball	Yes	Fitness / Sports		Male	\$500,000	20%	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	NaN

495 rows × 17 columns

Notice that Amount and Equity are currently being treated as categorical variables (`dtype: object`). Can you figure out why this is? Clean up these columns and cast them to numeric types (i.e., a `dtype` of `int` or `float`) because we'll need to perform mathematical operations on these columns.

```
In [4]: # YOUR CODE HERE
df['Amount'] = df['Amount'].str[1:].replace(',','', regex=True).astype(float)
df['Equity'] = df['Equity'].str[:-1].astype(float)
df
```

	Season	No. in series	Company	Deal	Industry	Entrepreneur	Gender	Amount	Equity	Corcoran	Cuban	Greiner	Herjavec	John	O'Leary	Harrington	Guest	Details / Notes
0	1.0	1.0	Ava the Elephant	Yes	Healthcare		Female	50000.0	55.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN
1	1.0	1.0	Mr. Tod's Pie Factory	Yes	Food and Beverage		Male	460000.0	50.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	NaN
2	1.0	1.0	Wispos	No	Business Services		Male	NaN	NaN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN
3	1.0	1.0	College Foxes Packing Boxes	No	Lifestyle / Home		Male	NaN	NaN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN
4	1.0	1.0	Ionic Ear	No	Uncertain / Other		Male	NaN	NaN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN
...
490	6.0	28.0	You Kick Ass	Yes	Children / Education		Female	100000.0	10.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN
491	6.0	29.0	Shark Wheel	Yes	Fitness / Sports		Male	225000.0	8.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	10% royalty until \$500K; then converts to 5% e...
492	6.0	29.0	Gato Cafe	No	Uncertain / Other		Female	NaN	NaN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN
493	6.0	29.0	Sway Motorsports	Yes	Green/CleanTech		Male	300000.0	20.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN
494	6.0	29.0	Spikeball	Yes	Fitness / Sports		Male	500000.0	20.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	NaN

495 rows × 17 columns

Question 1. Which Company was Worth the Most?

The valuation of a company is how much it is worth. If someone invests \$10,000 for a 40% equity stake in the company, then this means the company must be valued at \$25,000, since 40% of \$25,000 is \$10,000.

Calculate the valuation of each company that was funded. Which company was most valuable? Is it the same as the company that received the largest total investment from the sharks?

```
In [5]: # YOUR CODE HERE
df["Value"] = df["Amount"] * (1 / (df["Equity"]*0.01))
tempVal = df["Value"].drop(312) # This company has an equity of zero resulting in a valuation of inf
print(df.loc[tempVal.idxmax()])
print(df.loc[df["Amount"].idxmax()])
df
```

421	Season	No. in series	Company	Deal	Industry	Entrepreneur	Gender	Amount	Equity	Corcoran	Cuban	Greiner	Herjavec	John	O'Leary	Harrington	Guest	Details / Notes	Value
421	6.0	11.0	Zipz	Yes	Food and Beverage		Male	2500000.0	10.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	with an option for another \$2.5 Million for an...	25000000.0
483	Season	No. in series	Company	Deal	Industry	Entrepreneur	Gender	Amount	Equity	Corcoran	Cuban	Greiner	Herjavec	John	O'Leary	Harrington	Guest	Details / Notes	Value
483	6.0	27.0	AirCar	Yes	Green/CleanTech		Male	5000000.0	50.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	Contingent on getting deal to bring to continue...	10000000.0

	Season	No. in series	Company	Deal	Industry	Entrepreneur	Gender	Amount	Equity	Corcoran	Cuban	Greiner	Herjavec	John	O'Leary	Harrington	Guest	Details / Notes	Value
0	1.0	1.0	Ava the Elephant	Yes	Healthcare		Female	50000.0	55.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN	9.090909e+04
1	1.0	1.0	Mr. Tod's Pie Factory	Yes	Food and Beverage		Male	460000.0	50.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	NaN	9.200000e+05
2	1.0	1.0	Wispos	No	Business Services		Male	NaN	NaN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN	NaN
3	1.0	1.0	College Foxes Packing Boxes	No	Lifestyle / Home		Male	NaN	NaN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN	NaN
4	1.0	1.0	Ionic Ear	No	Uncertain / Other		Male	NaN	NaN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN	NaN
...
490	6.0	28.0	You Kick Ass	Yes	Children / Education		Female	100000.0	10.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN	1.000000e+06
491	6.0	29.0	Shark Wheel	Yes	Fitness / Sports		Male	225000.0	8.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	10% royalty until \$500K; then converts to 5% e...	2.812500e+06
492	6.0	29.0	Gato Cafe	No	Uncertain / Other		Female	NaN	NaN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN	NaN
493	6.0	29.0	Sway Motorsports	Yes	Green/CleanTech		Male	300000.0	20.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN	1.500000e+06
494	6.0	29.0	Spikeball	Yes	Fitness / Sports		Male	500000.0	20.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	NaN	2.500000e+06

495 rows × 18 columns

YOUR EXPLANATION HERE The most valuable company was Zipz at 25,000,000 USD. The company that received the largest investment was AirCar for 5,000,000 USD.

Question 2. Which Shark Invested the Most?

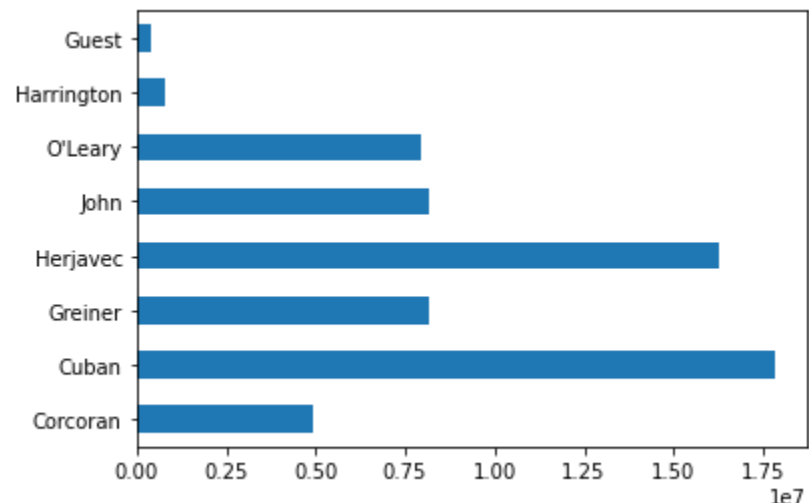
Calculate the total amount of money that each shark invested over the 6 seasons. Which shark invested the most total money over the 6 seasons?

Hint: If n sharks funded a given venture, then the amount that each shark invested is the total amount divided by n .

```
In [6]: # ENTER CODE HERE.
df["NumShar"] = df.loc[:, "Corcoran":"Guest"].sum(1)
df["InvestPerShark"] = df["Amount"] / df["NumShar"]
tempDF = df[df["Deal"] == "Yes"] # only use rows that have sharks investing in
tempDF.loc[:, "Corcoran":"Guest"] = tempDF["InvestPerShark"].sum(0).plot.barh()
```

	Season	No. in series	Company	Deal	Industry	Entrepreneur	Gender	Amount	Equity	Corcoran	Cuban	Greiner	Herjavec	John	O'Leary	Harrington	Guest	Details / Notes	Value	NumShar	InvestPerShark
0	1.0	1.0	Ava the Elephant	Yes	Healthcare		Female	50000.0	55.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN	9.090909e+04	1.0	50000.0
1	1.0	1.0	Mr. Tod's Pie Factory	Yes	Food and Beverage		Male	460000.0	50.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	NaN	9.200000e+05	2.0	230000.0
5	1.0	2.0	A Perfect Pear	Yes	Food and Beverage		Female	500000.0	50.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	NaN	1.000000e+06	2.0	250000.0
6	1.0	2.0	Classroom Jams	Yes	Children / Education		Male	250000.0	10.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0	0.0	NaN	2.500000e+06	5.0	50000.0
10	1.0	3.0	Turbobaster	Yes	Food and Beverage		Female	35000.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	2% royalty	3.500000e+04	1.0	35000.0
...
489	6.0	28.0	SynDaver Labs	Yes	Healthcare		Male	3000000.0	25.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	NaN	1.200000e+07	1.0	3000000.0
490	6.0	28.0	You Kick Ass	Yes	Children / Education		Female	100000.0	10.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN	1.000000e+06	1.0	100000.0
491	6.0	29.0	Shark Wheel	Yes	Fitness / Sports		Male	225000.0	8.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	10% royalty until \$500K; then converts to 5% e...	2.812500e+06	3.0	75000.0
493	6.0	29.0	Sway Motorsports	Yes	Green/CleanTech		Male	300000.0	20.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN	1.500000e+06	1.0	300000.0
494	6.0	29.0	Spikeball	Yes	Fitness / Sports		Male	500000.0	20.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	NaN	2.500000e+06	1.0	500000.0

249 rows × 20 columns



YOUR EXPLANATION HERE Mark Cuban invested the most money, at roughly 17,500,000 USD.

Question 3. Do the Sharks Prefer Certain Industries?

Calculate the funding rate (the proportion of companies that were funded) for each industry. Make a visualization showing this information.

```
In [8]: # ENTER CODE HERE.
tempDF = df[df["Deal"] == "Yes"]
tempDF.groupby("Industry").size().plot.barh()
```

