### **Shark Tank**

Shark Tank is a reality TV show. Contestants present their idea for a company to a panel of 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!

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

# YOUR CODE HERE df = pd.read\_csv("sharktank.csv")

29.0

29.0

df Greiner Details / Notes No. in series Out[2]: Company Deal **Industry Entrepreneur Gender** Equity Corcoran Cuban Herjavec John O'Leary Harrington Guest Season Amount 0 1.0 1.0 Ava the Elephant Yes Healthcare \$50,000 55% 1.0 NaN NaN NaN NaN Female NaN NaN NaN NaN 1.0 1.0 Mr. Tod's Pie Factory Yes Food and Beverage \$460,000 50% 1.0 NaN NaN NaN 1.0 NaN NaN NaN NaN 1.0 1.0 Wispots **Business Services** Male NaN 1.0 College Foxes Packing Boxes 1.0 Lifestyle / Home Male NaN 1.0 1.0 4 Ionic Ear No Uncertain / Other Male 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 6.0 29.0 1.0 10% royalty until \$500K; then converts to 5% e... 491 Shark Wheel Fitness / Sports \$225,000 NaN 1.0 NaN 1.0 NaN NaN

NaN

20%

20%

NaN

\$300,000

Male \$500,000

Female

494 6.0 29.0

495 rows × 17 columns

6.0

6.0

492

493

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

NaN

1.0

NaN

NaN

NaN

NaN

NaN

NaN

NaN

NaN

NaN

1.0

NaN

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

Gato Cafe

Spikeball

Sway Motorsports

No

Yes

Uncertain / Other

Green/CleanTech

Fitness / Sports

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.

Out[3]: Season No. in series Company Deal Industry Entrepreneur Gender Equity Corcoran Cuban Greiner Herjavec John O'Leary Harrington Guest Details / Notes Amount 0 1.0 1.0 Ava the Elephant Yes Healthcare 0.0 0.0 NaN \$50,000 55% 1.0 0.0 0.0 0.0 0.0 0.0 Mr. Tod's Pie Factory Yes Food and Beverage 1.0 1.0 \$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 No Business Services NaN 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 NaN Wispots NaN 1.0 College Foxes Packing Boxes 4 1.0 1.0 No NaN 0.0 0.0 0.0 0.0 0.0 0.0 0.0 NaN Ionic Ear Uncertain / Other NaN 0.0 Female \$100.000 0.0 490 6.0 28.0 You Kick Ass Yes Children / Education 10% 0.0 1.0 0.0 0.0 0.0 0.0 0.0 NaN 6.0 Fitness / Sports Male \$225,000 1.0 10% royalty until \$500K; then converts to 5% e... 491 29.0 Shark Wheel 8% 0.0 1.0 0.0 1.0 0.0 0.0 Uncertain / Other 492 6.0 29.0 Gato Cafe No NaN NaN 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 NaN Male \$300,000 NaN 493 6.0 29.0 Sway Motorsports Yes Green/CleanTech 20% 0.0 1.0 0.0 0.0 0.0 0.0 0.0 0.0 Male \$500,000 494 6.0 29.0 20% 0.0 0.0 0.0 0.0 0.0 0.0 NaN Spikeball Yes Fitness / Sports 0.0 1.0

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)

Company Deal Season No. in series **Industry Entrepreneur Gender** Amount Equity Corcoran Cuban Greiner Herjavec John O'Leary Harrington Guest Details / Notes Out[4]: 0 1.0 1.0 Ava the Elephant Yes Healthcare Female 50000.0 1.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 NaN 1.0 1.0 50.0 NaN Mr. Tod's Pie Factory Food and Beverage 460000.0 1.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0 Yes Male 1.0 1.0 **Business Services** 0.0 0.0 0.0 0.0 0.0 0.0 NaN Wispots NaN 0.0 0.0 1.0 College Foxes Packing Boxes 1.0 3 0.0 0.0 0.0 0.0 0.0 0.0 0.0 NaN Lifestyle / Home Male NaN NaN 0.0 4 1.0 1.0 NaN 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 NaN Ionic Ear No Uncertain / Other Male NaN 490 6.0 28.0 Yes Children / Education 100000.0 10.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0 NaN You Kick Ass 1.0 10% royalty until \$500K; then converts to 5% e... 6.0 29.0 Fitness / Sports 491 Shark Wheel 225000.0 8.0 0.0 1.0 0.0 1.0 0.0 0.0 Yes 0.0 492 6.0 29.0 0.0 0.0 0.0 Gato Cafe No Uncertain / Other NaN NaN 0.0 0.0 0.0 0.0 6.0 493 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 6.0 Spikeball Yes Fitness / Sports Male 500000.0 0.0 0.0 1.0 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 valution of inf print(df.loc[[tempVal.idxmax()]]) print(df.loc[[df["Amount"].idxmax()]]) df Season No. in series Company Deal

421 11.0 Zipz Yes Food and Beverage Entrepreneur Gender Amount Equity Corcoran Cuban Greiner \ 421 Male 2500000.0 10.0

Herjavec John O'Leary Harrington Guest \ 0.0 0.0 421 0.0 1.0

Details / Notes with an option for another \$2.5 Million for an... 25000000.0 Season No. in series Company Deal Industry Entrepreneur Gender \ 27.0 AirCar Yes Green/CleanTech

Amount Equity Corcoran Cuban Greiner Herjavec John O'Leary \ 5000000.0 483 50.0 0.0 0.0 0.0 1.0 0.0

Harrington Guest Details / Notes \ Contingent on getting deal to bring to contine... 483

483 10000000.0

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

495 rows × 18 columns

249 rows × 20 columns

In [8]: # ENTER CODE HERE.

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. # 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"].multiply(tempDF["InvestPerShark"],0).sum(0).plot.barh() tempDF

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

Guest Harrington O'Leary Herjavec Greiner Cuban Corcorar 0.50 0.75 1.00 1.25 0.00 0.25 1.50

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.

tempDF = df[df["Deal"] == "Yes"] tempDF.groupby("Industry").size().plot.barh() Out[8]: <AxesSubplot:ylabel='Industry'>

Uncertain / Other Software / Tech Pet Products Media / Entertainment Lifestyle / Home Healthcare Green/CleanTech Food and Beverage Fitness / Sports Fashion / Beauty Consumer Products Children / Education Business Services 50

YOUR EXPLANATION HERE The sharks have invested the most money into Food and Beverage.

## **Submission Instructions** Once you are finished, follow these steps:

- 1. Restart the kernel and re-run this notebook from beginning to end by going to Kernel > Restart Kernel and Run All Cells. 2. If this process stops halfway through, that means there was an error. Correct the error and repeat Step 1 until the notebook runs from beginning to end. 3. Double check that there is a number next to each code cell and that these numbers are in order.
- Then, submit your lab as follows:
- 1. Go to File > Export Notebook As > PDF.

2. Double check that the entire notebook, from beginning to end, is in this PDF file. (If the notebook is cut off, try first exporting the notebook to HTML and printing to PDF.) 3. Upload the PDF to Gradescope and Notebook (ipynb) to iLearn

4. Demo your lab.