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 (http://abc.go.com/shows/shark-tank) 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]:

```
%matplotlib inline
import numpy as np
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]:

```
shark_tank_df = pd.read_csv("sharktank.csv")
shark_tank_df
```

Out[2]:

	Season	No. in series	Company	Deal	Industry	Entrepreneur Gender	Amount	Equity	Corcora
0	1.0	1.0	Ava the Elephant	Yes	Healthcare	Female	$50,000	55%	1
1	1.0	1.0	Mr. Tod's Pie Factory	Yes	Food and Beverage	Male	$460,000	50%	1
2	1.0	1.0	Wispots	No	Business Services	Male	NaN	NaN	Na
3	1.0	1.0	College Foxes Packing Boxes	No	Lifestyle / Home	Male	NaN	NaN	Na
4	1.0	1.0	Ionic Ear	No	Uncertain / Other	Male	NaN	NaN	Na
...	...	...	...	...	...	...	...	...	
490	6.0	28.0	You Kick Ass	Yes	Children / Education	Female	$100,000	10%	Na
491	6.0	29.0	Shark Wheel	Yes	Fitness / Sports	Male	$225,000	8%	Na
492	6.0	29.0	Gato Cafe	No	Uncertain / Other	Female	NaN	NaN	Na
493	6.0	29.0	Sway Motorsports	Yes	Green/CleanTech	Male	$300,000	20%	Na
494	6.0	29.0	Spikeball	Yes	Fitness / Sports	Male	$500,000	20%	Na

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]:

```
shark_tank_df.loc[:, "Corcoran":"Guest"] = shark_tank_df.loc[:,"Corcoran":"Guest"].fillna(
0)
shark_tank_df
```

Out[3]:

	Season	No. in series	Company	Deal	Industry	Entrepreneur Gender	Amount	Equity	Corcora
0	1.0	1.0	Ava the Elephant	Yes	Healthcare	Female	$50,000	55%	1
1	1.0	1.0	Mr. Tod's Pie Factory	Yes	Food and Beverage	Male	$460,000	50%	1
2	1.0	1.0	Wispots	No	Business Services	Male	NaN	NaN	0
3	1.0	1.0	College Foxes Packing Boxes	No	Lifestyle / Home	Male	NaN	NaN	0
4	1.0	1.0	Ionic Ear	No	Uncertain / Other	Male	NaN	NaN	0
...	...	...	...	...	...	...	...	...	
490	6.0	28.0	You Kick Ass	Yes	Children / Education	Female	$100,000	10%	0
491	6.0	29.0	Shark Wheel	Yes	Fitness / Sports	Male	$225,000	8%	0
492	6.0	29.0	Gato Cafe	No	Uncertain / Other	Female	NaN	NaN	0
493	6.0	29.0	Sway Motorsports	Yes	Green/CleanTech	Male	$300,000	20%	0
494	6.0	29.0	Spikeball	Yes	Fitness / Sports	Male	$500,000	20%	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]:

```
shark_tank_df["Amount"] = shark_tank_df["Amount"].str[1:]
shark_tank_df["Equity"] = shark_tank_df["Equity"].str[:-1]

shark_tank_df["Amount"] = shark_tank_df["Amount"].str.replace(',', '').astype(float)
shark_tank_df["Equity"] = shark_tank_df["Equity"].astype(float)
```

In [5]:

```
shark_tank_df
```

Out[5]:

	Season	No. in series	Company	Deal	Industry	Entrepreneur Gender	Amount	Equity	Corcora
0	1.0	1.0	Ava the Elephant	Yes	Healthcare	Female	50000.0	55.0	1
1	1.0	1.0	Mr. Tod's Pie Factory	Yes	Food and Beverage	Male	460000.0	50.0	1
2	1.0	1.0	Wispots	No	Business Services	Male	NaN	NaN	0
3	1.0	1.0	College Foxes Packing Boxes	No	Lifestyle / Home	Male	NaN	NaN	0
4	1.0	1.0	Ionic Ear	No	Uncertain / Other	Male	NaN	NaN	0
...	...	...	...	...	...	...	...	...	
490	6.0	28.0	You Kick Ass	Yes	Children / Education	Female	100000.0	10.0	0
491	6.0	29.0	Shark Wheel	Yes	Fitness / Sports	Male	225000.0	8.0	0
492	6.0	29.0	Gato Cafe	No	Uncertain / Other	Female	NaN	NaN	0
493	6.0	29.0	Sway Motorsports	Yes	Green/CleanTech	Male	300000.0	20.0	0
494	6.0	29.0	Spikeball	Yes	Fitness / Sports	Male	500000.0	20.0	0

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?

Create new dataframe which has only companies with investments

In [6]:

```
temp_shark = shark_tank_df[shark_tank_df["Amount"].notnull()].copy()
temp_shark.reset_index(drop = True, inplace = True)
```

Let's find which line has Equity 0 and drop it

In [7]:

```
temp_shark[temp_shark["Equity"] == 0]
```

Out[7]:

	Season	No. in series	Company	Deal	Industry	Entrepreneur Gender	Amount	Equity	Corcoran	Cubar
150	5.0	13.0	The Wall DoctoRX	Yes	Lifestyle / Home	Male	150000.0	0.0	0.0	0.(

In [8]:

```
# Drop index 150 because equity is zero .
worth_shark = temp_shark.drop(150)
worth_shark
```

Out[8]:

	Season	No. in series	Company	Deal	Industry	Entrepreneur Gender	Amount	Equity	Corco
0	1.0	1.0	Ava the Elephant	Yes	Healthcare	Female	50000.0	55.0	
1	1.0	1.0	Mr. Tod's Pie Factory	Yes	Food and Beverage	Male	460000.0	50.0	
2	1.0	2.0	A Perfect Pear	Yes	Food and Beverage	Female	500000.0	50.0	
3	1.0	2.0	Classroom Jams	Yes	Children / Education	Male	250000.0	10.0	
4	1.0	3.0	Turbobaster	Yes	Food and Beverage	Female	35000.0	100.0	
...	...	...	...	...	...	...	...	...	
244	6.0	28.0	SynDaver Labs	Yes	Healthcare	Male	3000000.0	25.0	
245	6.0	28.0	You Kick Ass	Yes	Children / Education	Female	100000.0	10.0	
246	6.0	29.0	Shark Wheel	Yes	Fitness / Sports	Male	225000.0	8.0	
247	6.0	29.0	Sway Motorsports	Yes	Green/CleanTech	Male	300000.0	20.0	
248	6.0	29.0	Spikeball	Yes	Fitness / Sports	Male	500000.0	20.0	

248 rows × 17 columns

Calculate which one has biggest value:

In [9]:

```
valuable = worth_shark["Amount"] / worth_shark["Equity"] * 100
worth_shark.loc[[valuable.idxmax()]]
```

Out[9]:

	Season	No. in series	Company	Deal	Industry	Entrepreneur Gender	Amount	Equity	Corcoran	Cub
206	6.0	11.0	Zipz	Yes	Food and Beverage	Male	2500000.0	10.0	0.0	

Zipz is the most valuable company.

See which company was funded more that others:

In [10]:

```
shark_tank_df.loc[[shark_tank_df["Amount"].idxmax()]]
```

Out[10]:

	Season	No. in series	Company	Deal	Industry	Entrepreneur Gender	Amount	Equity	Corcora
483	6.0	27.0	AirCar	Yes	Green/CleanTech	Male	5000000.0	50.0	0.

AirCar is the most amount invested.
Total cost = amount / equity * 100
Biggest amount will not necessary point to the biggest total cost

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.

Create new column with amount of sharks invested in each company

In [11]:

```
temp_shark["Sharks Quantity"] = temp_shark.loc[:,'Corcoran':'Guest'].sum(axis=1)

#Update table so only lines with more than 0 investors will be present
temp_shark = temp_shark[temp_shark["Sharks Quantity"] != 0].reset_index(drop=True)
```

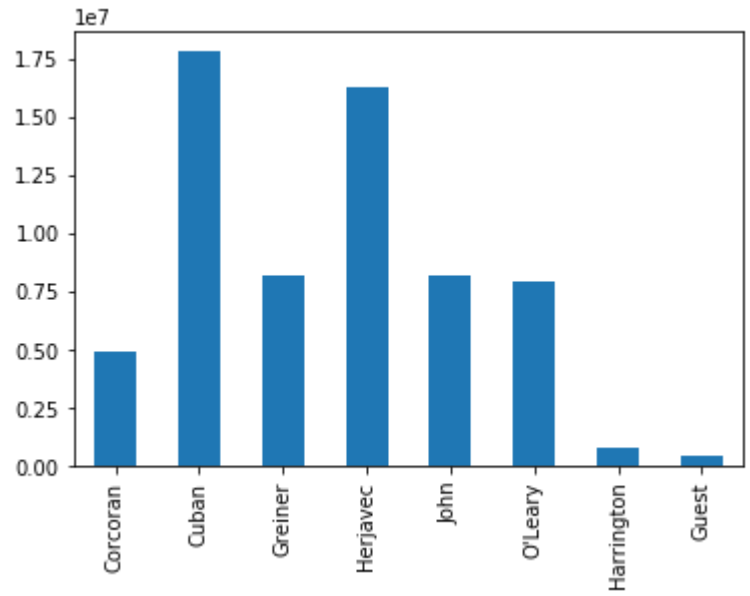
Multiply all sharks columns by 'invested_amount' column to have amount each shark invested in each case and sum those amounts

In [12]:

```
temp_shark["Invested Amount"] = temp_shark["Amount"] / temp_shark["Sharks Quantity"]
temp_shark.loc[:, "Corcoran":"Guest"].multiply(temp_shark["Invested Amount"],axis=0).sum(a
xis=0).plot.bar()
```

Out[12]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1dd23d52448>
```



Cuban invested more than others ; $17,817,500.00

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.

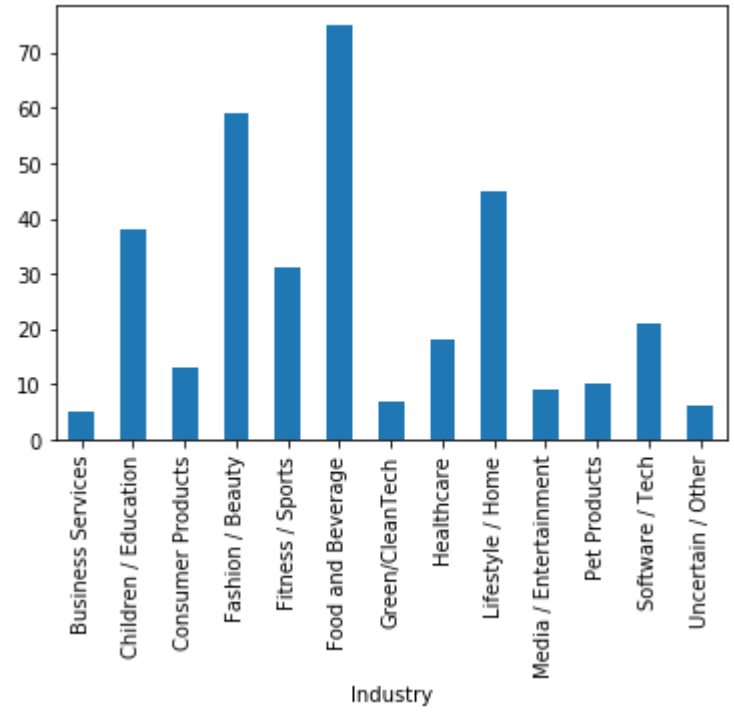
Here is a bar graph representing amount of sharks invested in each industry

In [13]:

```
shark_tank_df["Number of Sharks"] = shark_tank_df.loc[:, "Corcoran":"Guest"].sum(axis=1)
shark_tank_df.groupby("Industry")["Number of Sharks"].sum().plot.bar()
```

Out[13]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1dd24fee848>
```



Food in Beverage has highest amount of sharks invested in.

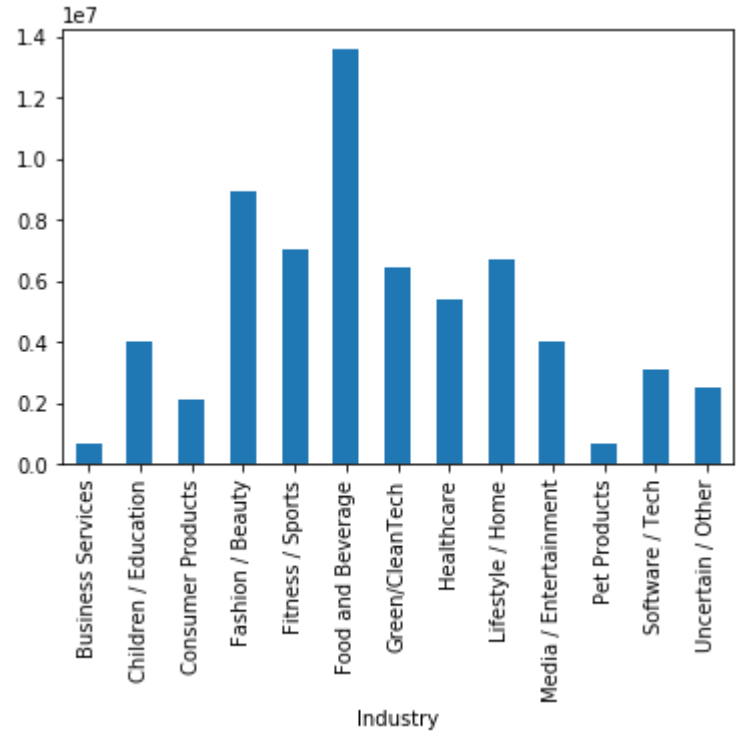
Here is a bar graph representing total amount of money invested in each industry

In [14]:

```
shark_tank_df.groupby("Industry")["Amount"].sum().plot.bar()
```

Out[14]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1dd250853c8>
```



Sharks invested more money in Food and Beverage than other industries.
Sharks prefer Food and Beverage industry.

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 and Notebook (ipynb) to iLearn