

SHARK TANK US ANALYSIS

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

```
df=pd.read_csv("C:/Users/lekhs/OneDrive/Documents/startup
analysis/Shark Tank US.csv")
```

```
df.head(10)
```

	Season Number	Episode Number	Pitch Number	Original Air Date	\
0	1	1	1	09-08-2009	
1	1	1	2	09-08-2009	
2	1	1	3	09-08-2009	
3	1	1	4	09-08-2009	
4	1	1	5	09-08-2009	
5	1	2	6	16-08-2009	
6	1	2	7	16-08-2009	
7	1	2	8	16-08-2009	
8	1	2	9	16-08-2009	
9	1	2	10	16-08-2009	

	Startup Name	Industry	\
0	AvaTheElephant	Health/Wellness	
1	Mr.Tod'sPieFactory	Food and Beverage	
2	Wispots	Business Services	
3	CollegeFoxesPackingBoxes	Lifestyle/Home	
4	IonicEar	Software/Tech	
5	APerfectPear	Food and Beverage	
6	ClassroomJams	Children/Education	
7	Lifebelt	Automotive	
8	CrookedJaw	Fashion/Beauty	
9	StickyNoteHolder	Lifestyle/Home	

	Business Description	Pitchers	Gender	\
0	Ava The Elephant - Baby and Child Care		Female	
1	Mr. Tod's Pie Factory - Specialty Food		Male	
2	Wispots - Consumer Services		Male	
3	College Foxes Packing Boxes - Consumer Services		Male	
4	Ionic Ear - Novelties		Male	
5	A Perfect Pear - Specialty Food		Female	
6	Classroom Jams - Education		Male	
7	Lifebelt - Automotive		Male	
8	Crooked Jaw - Men and Women's Apparel		Male	
9	Sticky Note Holder - Productivity Tools		Female	

	Pitchers City	Pitchers State	... Kevin O Leary	Investment Equity	\
0	Atlanta	GA	...	0.0	

1	Somerset	NJ	...	0.0
2	Cary	NC	...	0.0
3	Tampa	FL	...	0.0
4	St. Paul	MN	...	0.0
5	Napa Valley	CA	...	25.0
6	Chicago	IL	...	20.0
7	Las Vegas	NV	...	0.0
8	Hollywood	CA	...	0.0
9	Gardena	CA	...	0.0

	Guest Investment Amount	Guest Investment Equity	Guest Name \
0	0.0	0.0	0
1	0.0	0.0	0
2	0.0	0.0	0
3	0.0	0.0	0
4	0.0	0.0	0
5	0.0	0.0	0
6	0.0	0.0	0
7	0.0	0.0	0
8	0.0	0.0	0
9	0.0	0.0	0

	Barbara Corcoran Present	Mark Cuban Present	Lori Greiner Present
\			
0	1	0	0
1	1	0	0
2	1	0	0
3	1	0	0
4	1	0	0
5	1	0	0
6	1	0	0
7	1	0	0
8	1	0	0
9	1	0	0

	Robert Herjavec Present	Daymond John Present	Kevin O Leary Present
0	1	1	
1			
1	1	1	

1		
2	1	1
1		
3	1	1
1		
4	1	1
1		
5	1	1
1		
6	1	1
1		
7	1	1
1		
8	1	1
1		
9	1	1
1		

[10 rows x 47 columns]

DATA CLEANING

df.drop_duplicates()

	Season Number	Episode Number	Pitch Number	Original Air Date	\
0	1	1	1	09-08-2009	
1	1	1	2	09-08-2009	
2	1	1	3	09-08-2009	
3	1	1	4	09-08-2009	
4	1	1	5	09-08-2009	
...	
1269	14	21	1270	05-05-2023	
1270	14	22	1271	19-05-2023	
1271	14	22	1272	19-05-2023	
1272	14	22	1273	19-05-2023	
1273	14	22	1274	19-05-2023	

	Startup Name	Industry	\
0	AvaTheElephant	Health/Wellness	
1	Mr.Tod'sPieFactory	Food and Beverage	
2	Wispots	Business Services	
3	CollegeFoxesPackingBoxes	Lifestyle/Home	
4	IonicEar	Software/Tech	
...	
1269	NoshiFood	Food and Beverage	
1270	Tucky	Fashion/Beauty	
1271	InfluencersInTheWild	Lifestyle/Home	
1272	Eyewris	Lifestyle/Home	
1273	TonesOfMelanin	Fashion/Beauty	

Business Description Pitchers				
Gender \				
0	Ava The Elephant - Baby and Child Care			
Female				
1	Mr. Tod's Pie Factory - Specialty Food			
Male				
2	Wispots - Consumer Services			
Male				
3	College Foxes Packing Boxes - Consumer Services			
Male				
4	Ionic Ear - Novelties			
Male				
...	
.				
1269	Food paint			
Male				
1270	Turn any shirt into a crop top			
Female				
1271	Board game that requires you to act as a socia...			
Male				
1272	Reading glasses			
Male				
1273	Collegiate fashion for minority communities			
Female				
	Pitchers City Pitchers State	... Kevin 0 Leary Investment Equity		
\				
0	Atlanta	GA	...	0.0
1	Somerset	NJ	...	0.0
2	Cary	NC	...	0.0
3	Tampa	FL	...	0.0
4	St. Paul	MN	...	0.0
...
1269	0	0	...	0.0
1270	0	0	...	0.0
1271	0	0	...	0.0
1272	0	0	...	4.0
1273	0	0	...	0.0

	Guest Investment Amount	Guest Investment Equity	Guest Name \
0	0.0	0.0	0
1	0.0	0.0	0
2	0.0	0.0	0
3	0.0	0.0	0
4	0.0	0.0	0
...
1269	0.0	0.0	0
1270	0.0	0.0	0
1271	0.0	0.0	0
1272	0.0	0.0	0
1273	0.0	0.0	0

	Barbara Corcoran Present	Mark Cuban Present	Lori Greiner Present \
0	1	0	
0			
1	1	0	
0			
2	1	0	
0			
3	1	0	
0			
4	1	0	
0			
...
..			
1269	0	1	
0			
1270	0	0	
0			
1271	0	0	
0			
1272	1	1	
1			
1273	0	1	
0			

	Robert Herjavec Present	Daymond John Present	Kevin O Leary Present
0	1	1	
1			
1	1	1	
1			
2	1	1	
1			
3	1	1	
1			
4	1	1	

```

1
...
...
1269      0      0
0
1270      0      1
0
1271      0      0
0
1272      0      1
1
1273      0      0
0

```

```
[1274 rows x 47 columns]
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 1274 entries, 0 to 1273
```

```
Data columns (total 47 columns):
```

#	Column	Non-Null Count	Dtype
0	Season Number	1274 non-null	int64
1	Episode Number	1274 non-null	int64
2	Pitch Number	1274 non-null	int64
3	Original Air Date	1274 non-null	object
4	Startup Name	1274 non-null	object
5	Industry	1274 non-null	object
6	Business Description	1274 non-null	object
7	Pitchers Gender	1274 non-null	object
8	Pitchers City	1274 non-null	object
9	Pitchers State	1274 non-null	object
10	Entrepreneur Names	1274 non-null	object
11	Company Website	1274 non-null	object
12	Multiple Entrepreneurs	1274 non-null	int64
13	US Viewership	1274 non-null	float64
14	Original Ask Amount	1274 non-null	int64
15	Original Offered Equity	1274 non-null	float64
16	Valuation Requested	1274 non-null	int64
17	Got Deal	1274 non-null	int64
18	Total Deal Amount	1274 non-null	int64
19	Total Deal Equity	1274 non-null	float64
20	Deal Valuation	1274 non-null	int64
21	Number of sharks in deal	1274 non-null	int64
22	Investment Amount Per Shark	1274 non-null	float64
23	Equity Per Shark	1274 non-null	float64
24	Royalty Deal	1274 non-null	int64
25	Loan	1274 non-null	int64
26	Barbara Corcoran Investment Amount	1274 non-null	int64

27	Barbara Corcoran Investment Equity	1274	non-null	float64
28	Mark Cuban Investment Amount	1274	non-null	float64
29	Mark Cuban Investment Equity	1274	non-null	float64
30	Lori Greiner Investment Amount	1274	non-null	float64
31	Lori Greiner Investment Equity	1274	non-null	float64
32	Robert Herjavec Investment Amount	1274	non-null	float64
33	Robert Herjavec Investment Equity	1274	non-null	float64
34	Daymond John Investment Amount	1274	non-null	float64
35	Daymond John Investment Equity	1274	non-null	float64
36	Kevin O Leary Investment Amount	1274	non-null	float64
37	Kevin O Leary Investment Equity	1274	non-null	float64
38	Guest Investment Amount	1274	non-null	float64
39	Guest Investment Equity	1274	non-null	float64
40	Guest Name	1274	non-null	object
41	Barbara Corcoran Present	1274	non-null	int64
42	Mark Cuban Present	1274	non-null	int64
43	Lori Greiner Present	1274	non-null	int64
44	Robert Herjavec Present	1274	non-null	int64
45	Daymond John Present	1274	non-null	int64
46	Kevin O Leary Present	1274	non-null	int64

dtypes: float64(18), int64(19), object(10)

memory usage: 467.9+ KB

df.isnull()

Date	Season	Number	Episode	Number	Pitch	Number	Original	Air
0		False		False		False		False
1		False		False		False		False
2		False		False		False		False
3		False		False		False		False
4		False		False		False		False
...	
1269		False		False		False		False
1270		False		False		False		False
1271		False		False		False		False
1272		False		False		False		False
1273		False		False		False		False

Startup	Name	Industry	Business	Description	Pitchers	Gender	\
---------	------	----------	----------	-------------	----------	--------	---

0	False	False	False	False
1	False	False	False	False
2	False	False	False	False
3	False	False	False	False
4	False	False	False	False
...
1269	False	False	False	False
1270	False	False	False	False
1271	False	False	False	False
1272	False	False	False	False
1273	False	False	False	False

Pitchers City Pitchers State ... Kevin 0 Leary Investment Equity \

0	False	False	...
1	False	False	...
2	False	False	...
3	False	False	...
4	False	False	...
...
...			
1269	False	False	...
1270	False	False	...
1271	False	False	...
1272	False	False	...
1273	False	False	...

	Guest Investment Amount	Guest Investment Equity	Guest Name \
0	False	False	False
1	False	False	False
2	False	False	False
3	False	False	False
4	False	False	False
...
1269	False	False	False
1270	False	False	False
1271	False	False	False
1272	False	False	False
1273	False	False	False

Barbara Corcoran Present		Mark Cuban Present		Lori Greiner	
Present \					
0	False		False		
False					
1	False		False		
False					
2	False		False		
False					
3	False		False		
False					
4	False		False		
False					
...
..					
1269	False		False		
False					
1270	False		False		
False					
1271	False		False		
False					
1272	False		False		
False					
1273	False		False		
False					
Robert Herjavec Present		Daymond John Present		Kevin O Leary	
Present					
0	False		False		
False					
1	False		False		
False					
2	False		False		
False					
3	False		False		
False					
4	False		False		
False					
...		
...					
1269	False		False		
False					
1270	False		False		
False					
1271	False		False		
False					
1272	False		False		
False					
1273	False		False		
False					

```
[1274 rows x 47 columns]
```

```
print(df.isna().sum())
```

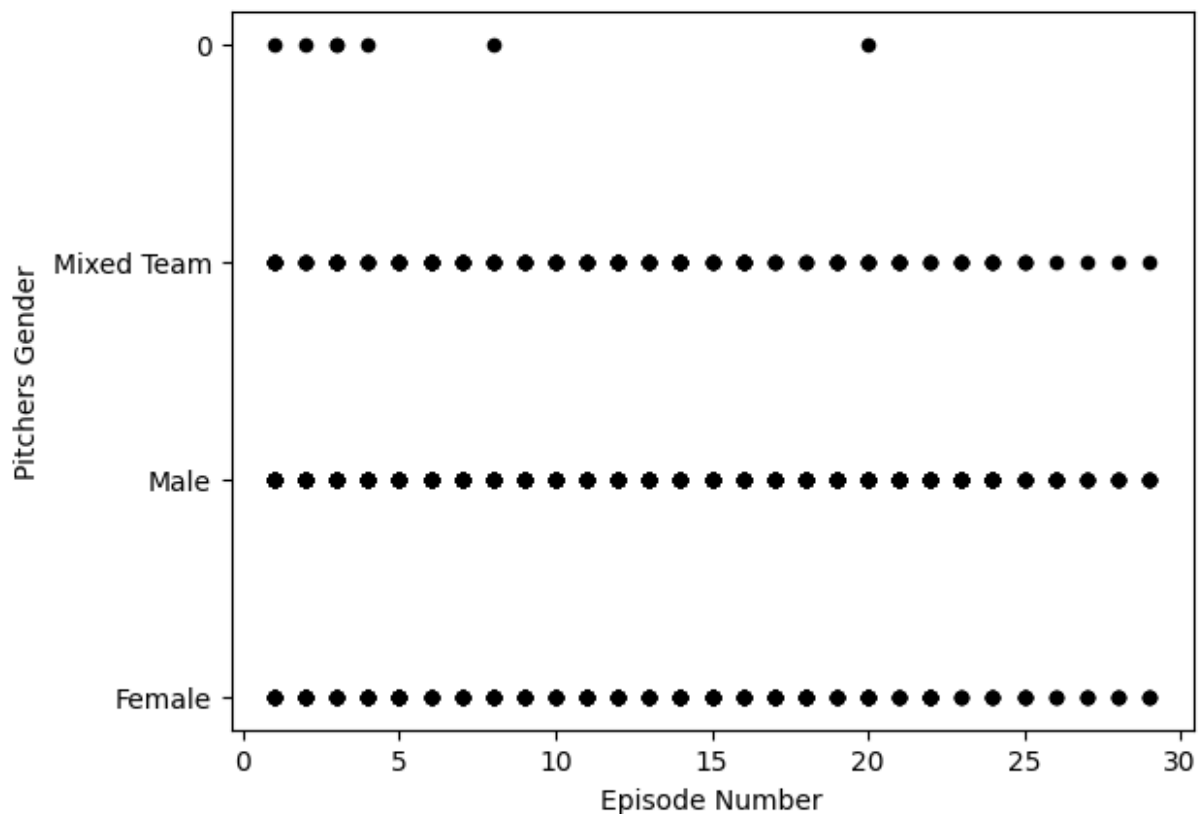
Season Number	0
Episode Number	0
Pitch Number	0
Original Air Date	0
Startup Name	0
Industry	0
Business Description	0
Pitchers Gender	0
Pitchers City	0
Pitchers State	0
Entrepreneur Names	0
Company Website	0
Multiple Entrepreneurs	0
US Viewership	0
Original Ask Amount	0
Original Offered Equity	0
Valuation Requested	0
Got Deal	0
Total Deal Amount	0
Total Deal Equity	0
Deal Valuation	0
Number of sharks in deal	0
Investment Amount Per Shark	0
Equity Per Shark	0
Royalty Deal	0
Loan	0
Barbara Corcoran Investment Amount	0
Barbara Corcoran Investment Equity	0
Mark Cuban Investment Amount	0
Mark Cuban Investment Equity	0
Lori Greiner Investment Amount	0
Lori Greiner Investment Equity	0
Robert Herjavec Investment Amount	0
Robert Herjavec Investment Equity	0
Daymond John Investment Amount	0
Daymond John Investment Equity	0
Kevin O Leary Investment Amount	0
Kevin O Leary Investment Equity	0
Guest Investment Amount	0
Guest Investment Equity	0
Guest Name	0
Barbara Corcoran Present	0
Mark Cuban Present	0
Lori Greiner Present	0
Robert Herjavec Present	0

```
Daymond John Present      0
Kevin O Leary Present     0
dtype: int64
```

EDA

```
df=pd.read_csv("C:/Users/lekhs/OneDrive/Documents/startup
analysis/Shark Tank US.csv")

df.plot(kind='scatter', x="Episode Number", y="Pitchers Gender",
color="black")
plt.xlabel("Episode Number")
plt.ylabel("Pitchers Gender")
plt.show()
```



```
df["Industry"].hist(bins=20,figsize=(7,6), xrot=68, grid=True)
plt.xlabel("Industry"),
plt.ylabel("Got Deal"),
plt.show()
```

