

## 2.1 Pick out the Deal (Dependent Variable) and Description columns into a separate data frame.

In this dataset have 495 rows and 2 columns.

	deal	description
0	False	Bluetooth device implant for your ear.
1	True	Retail and wholesale pie factory with two reta...
2	True	Ava the Elephant is a godsend for frazzled par...
3	False	Organizing, packing, and moving services deliv...
4	False	Interactive media centers for healthcare waiti...
...	...	...
490	True	Zoom Interiors is a virtual service for interi...
491	True	Spikeball started out as a casual outdoors gam...
492	True	Shark Wheel is out to literally reinvent the w...
493	False	Adriana Montano wants to open the first Cat Ca...
494	True	Sway Motorsports makes a three-wheeled, all-el...

495 rows × 2 columns

## 2.2 Create two corpora, one for those who secured a Deal, the other for those who did not secure a deal.

Here the two corpora has been created one is the secured deal and another one is deal not secured.

### Deal Secured:

```
1      Retail and wholesale pie factory with two reta...
2      Ava the Elephant is a godsend for frazzled par...
5      One of the first entrepreneurs to pitch on Sha...
9      An educational record label and publishing hou...
10     A battery-operated cooking device that siphons...
...
489    SynDaver Labs makes synthetic body parts for u...
490    Zoom Interiors is a virtual service for interi...
491    Spikeball started out as a casual outdoors gam...
492    Shark Wheel is out to literally reinvent the w...
494    Sway Motorsports makes a three-wheeled, all-el...
Name: description, Length: 251, dtype: object
```

### Deal Not Secured:

```
0      Bluetooth device implant for your ear.
3      Organizing, packing, and moving services deliv...
4      Interactive media centers for healthcare waiti...
6      A mixed martial arts clothing line looking to ...
7      Attach Noted is a detachable "arm" that holds ...
...
482    Buck Mason makes high-quality men's clothing i...
484    Frameri answers the question, "Why aren't your...
485    The Paleo Diet Bar is a nutrition bar that is ...
488    Sunscreen Mist adds another point of access fo...
493    Adriana Montano wants to open the first Cat Ca...
Name: description, Length: 244, dtype: object
```

**2.3 The following exercise is to be done for both the corpora:**

- a) Find the number of characters for both the corpuses.**
- b) Remove Stop Words from the corpora. (Words like ‘also’, ‘made’, ‘makes’, ‘like’, ‘this’, ‘even’ and ‘company’ are to be removed)**
- c) What were the top 3 most frequently occurring words in both corpuses (after removing stop words)?**
- d) Plot the Word Cloud for both the corpora.**

**a) Find the number of characters for both the corpuses.**

Characters of deal are **64060** and characters of no secure deal are **47184**.

Number of characters in Secured Deal corpus: 64060

Number of characters in No Secure Deal corpus: 47184

**b) Remove Stop Words from the corpora. (Words like ‘also’, ‘made’, ‘makes’, ‘like’, ‘this’, ‘even’ and ‘company’ are to be removed).**

Here you can see the stop word has been removed.

Word count after removal of stopwords: 244

```
Sample sentence of corpus words:
0          Bluetooth device implant ear.
3  Organizing, packing, moving services delivered...
4  Interactive media centers healthcare waiting r...
6  mixed martial arts clothing line looking becom...
7  Attach Noted detachable "arm" holds Post-It no...
...
482 Buck Mason makes high-quality men's clothing USA.
484 Frameri answers question, "Why glasses flexibl...
485 Paleo Diet Bar nutrition bar gluten, soy, dair...
488 Sunscreen Mist adds another point access sunsc...
493 Adriana Montano wants open first Cat Cafe Flor...
Name: description, Length: 244, dtype: object
```

**c) What were the top 3 most frequently occurring words in both corpuses (after removing stop words)?**

The common words is made is 37, 2<sup>nd</sup> common word is also 19 and 3rd common word is 18.

Corpuses	
Most Common word	(made, 37)
2nd common word	(also, 19)
3rd common word	(company, 18)

**d) Plot the Word Cloud for both the corpora.**

In this plot water, make, product, company, system has most repeatable words are make, made, company, product etc.



#### 2.4 Refer to both the word clouds. What do you infer?

In this word cloud has provide more frequent words in “Description column. We can infer the common topics that businesses pitched on the show **“SHARK TANK”**.This can help in understanding the popular trends among pitched featured in the dataset.

## 2.5 Looking at the word clouds, is it true that the entrepreneurs who introduced devices are less likely to secure a deal based on your analysis?

The word clouds provide a visual representation of the most frequent words in the "**Description**" column of the dataset. Here the most frequent words is **made, make, product, company, etc.,**

By analyzing word cloud, we did **not directly indicate the secured a deal** based based on the type of product or device being pitched. We need a more detailed analysis and **statistical modeling would be required.**