**Shark**

**Tank**

Name: Saurabh Dharmadhikari

PGP-DSBA Online January’ 21

Date: 26/10/2022

Table of Contents:

Executive Summary……………………...................................................................................................................................3

Introduction ...................................................................................................................................................................3

Data Description ............................................................................................................................................................3

Sample of the dataset ...................................................................................................................................................4

1. Pick out the Deal (Dependent Variable) and Description columns into a separate data frame. …….........................4

2. Create two corpora, one with those who secured a Deal, the other with those who did not secure a deal. ..........5

3. The following exercise is to be done for both the corpora: …………………………………………………………………………………5

a) Find the number of characters for both the corpuses. ……………………………………………………………………………………...5

b) Remove Stop Words from the corpora. (Words like ‘also’, ‘made’, ‘makes’, ‘like’, ‘this’,

‘even’ and ‘company’ are to be removed) ………………………………………………………………………………………………………….….7

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

stop words)? ………………………………………………………………………………………………………………………………………………………...7

d) Plot the Word Cloud for both the corpora. ............................................................................................................8

4. Refer to both the word clouds. What do you infer? …………………………………………………………………….......................9

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? .........................................................................................................9

Executive Summary

A dataset of Shark Tank episodes is made available. It contains 495 entrepreneurs making their pitch

to the VC sharks.

Introduction

We are going to analyse the pitches made by the entrepreneurs at shark tank and we also have the data of them securing the deal or not. We can thus analyse the words that they used in their pitch and the key words that were in the pitches that secured the deal and also key words in the pitches that did not. So we will know what kind of products or services did not secure deal.

Data Description

1. deal: If the deal was secured.

2. description: Description of the products.

Sample data set:

|  | **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... |

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

We can see that the data with us has two possibilities, one that the deal was true and the other that the deal is false. True means that the deal was done and false means deal was not done. We have sorted and separated the two into separate data frames as shown below. Also number jof stop words are being shown here which need to be treated.

For False:

|  | **deal** | **description** | **stopwords** |
| --- | --- | --- | --- |
| **0** | False | bluetooth device implant ear | 2 |
| **3** | False | organizing packing moving services delivered c... | 2 |
| **4** | False | interactive media centers healthcare waiting r... | 2 |
| **6** | False | mixed martial arts clothing line looking becom... | 4 |
| **7** | False | attach noted detachable arm holds postit notes... | 7 |

For True:

|  | **deal** | **description** | **stopwords** |
| --- | --- | --- | --- |
| **1** | True | retail wholesale pie factory two retail locati... | 3 |
| **2** | True | ava elephant godsend frazzled parents young ch... | 14 |
| **5** | True | one first entrepreneurs pitch shark tank susan... | 14 |
| **9** | True | educational record label publishing house desi... | 6 |
| **10** | True | batteryoperated cooking device siphons juice s... | 7 |

2. Create two corpora, one with those who secured a Deal, the other with those who did not secure

a deal.

Word corpora is now made for both the data frames as shown above.

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.

1. Find the number of characters for both the corpuses.

For False data frame:

|  | **deal** | **description** | **no.ofcharacters** |
| --- | --- | --- | --- |
| **0** | False | bluetooth device implant for your ear | 37 |
| **3** | False | organizing packing and moving services deliver... | 65 |
| **4** | False | interactive media centers for healthcare waiti... | 111 |
| **6** | False | a mixed martial arts clothing line looking to ... | 107 |
| **7** | False | attach noted is a detachable arm that holds po... | 87 |
| **...** | ... | ... | ... |
| **482** | False | buck mason makes highquality mens clothing in ... | 53 |
| **484** | False | frameri answers the question why arent your gl... | 388 |
| **485** | False | the paleo diet bar is a nutrition bar that is ... | 252 |
| **488** | False | sunscreen mist adds another point of access fo... | 415 |
| **493** | False | adriana montano wants to open the first cat ca... | 417 |

244 rows × 3 columns

For True data frame:

|  | **deal** | **description** | **no.ofcharacters** |
| --- | --- | --- | --- |
| **1** | True | retail and wholesale pie factory with two reta... | 72 |
| **2** | True | ava the elephant is a godsend for frazzled par... | 242 |
| **5** | True | one of the first entrepreneurs to pitch on sha... | 352 |
| **9** | True | an educational record label and publishing hou... | 121 |
| **10** | True | a batteryoperated cooking device that siphons ... | 114 |
| **...** | ... | ... | ... |
| **489** | True | syndaver labs makes synthetic body parts for u... | 427 |
| **490** | True | zoom interiors is a virtual service for interi... | 319 |
| **491** | True | spikeball started out as a casual outdoors gam... | 492 |
| **492** | True | shark wheel is out to literally reinvent the w... | 486 |
| **494** | True | sway motorsports makes a threewheeled allelect... | 494 |

251 rows × 3 columns

b) Remove Stop Words from the corpora. (Words like ‘also’, ‘made’, ‘makes’, ‘like’, ‘this’,

‘even’ and ‘company’ are to be removed)

For False data frame after removing stop words:

0 bluetooth device implant ear

3 organizing packing moving services delivered c...

4 interactive media centers healthcare waiting r...

6 mixed martial arts clothing line looking becom...

7 attach noted detachable arm holds postit notes...

Name: description, dtype: object

For True data frame after removing stop words:

1 retail wholesale pie factory two retail locati...

2 ava elephant godsend frazzled parents young ch...

5 one first entrepreneurs pitch shark tank susan...

9 educational record label publishing house desi...

10 batteryoperated cooking device siphons juice s...

Name: description, dtype: object

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

stop words)?

**Top 3 most frequently occurring words in False datasheet:**

make 19

use 17

water 17

system 16

product 15

online 15

designed 15

bottle 14

without 14

products 13

dtype: int64

We can observe that top 3 most occurring words in False data frame are make, use and water.

**Top 3 most frequently occurring words in False datasheet:**

make 25

designed 19

products 19

line 18

easy 18

way 18

without 18

use 17

kids 17

new 15

dtype: int64

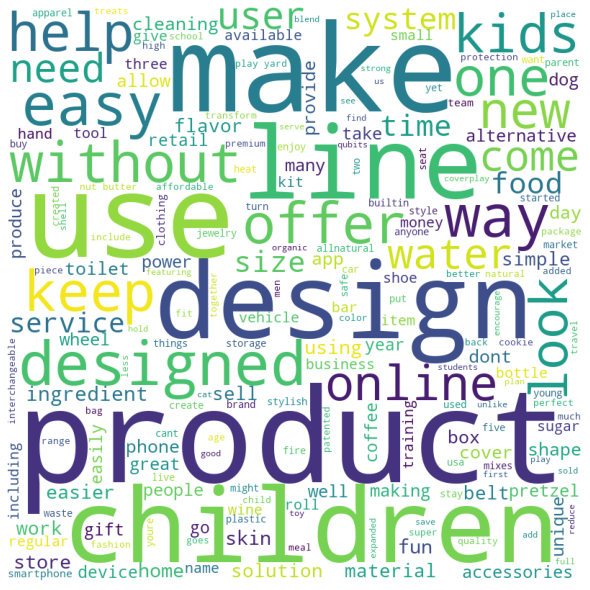
We can observe that top 3 most occurring words in True data frame are make, designed and products.

1. Plot the Word Cloud for both the corpora.

For False data frame:



For True data frame:



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

When we observe the word cloud for True, we see that deals that were encouraged in shark tank were more related to products that were related to children, design. Services and products that could help people more like skincare.

On the other hand, when we observe for False, we find that sharks did not prefer investing in devices, electronics, online systems and services, or traditional premium products like accessories and rooms. Also, products related to water, drinks or bar were also not a hit among them.

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?

Yes, it is we can say that the entrepreneurs who introduced devices are less likely to secure a deal as it appears a lot of times in the word cloud of False deal data.