ABSTRACT

This research was conducted for the purposes of my master's degree thesis and aims to create a machine learning model that can make funding predictions based on the data collected from entrepreneurs' funding pitches. Additionally, the project aims to unwrap and explain the main factors that induce investors trust and invest in entrepreneurs.

The methodology required the data collection by viewing 12 seasons of the TV series Shark Tank, performing audio analysis and emotion classification. The data were collected in a spreadsheet file which was used to export descriptive statistics charts and to conduct the statistical analysis.

The statistically significant results of this project are the following. As the amount asked, or the evaluation increase it is less likely that the team will make a deal. Additionally, having male speakers and presenting a product in fashion or beauty industry had a negative impact on the result. Teams that own a patent or don't have depts or loans are more likely to make a deal. Presenters with previous business success are more likely to have a deal. Also, as the number of previous sales or the number of presenters or number of syllables of the pitch increases there are beneficial results to the deal outcome. According to the findings the emotions angry and sad had a small statistically significant correlation.

The findings of this research reveal helpful information for entrepreneurs and provide a practical and important use of machine learning techniques and tools and the ways the above can be used to predict the success of an entrepreneur's pitch.

“VIDEO ANALYSIS OF ENTREPRENEURS’ PITCHES AND FUNDING PREDICTION”

GIORGOS MOLESKIS

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science at the University of Cyprus

Recommended for Acceptance
by the Department of Computer Science
June, 2022

APPROVAL PAGE

Master of Science Thesis

VIDEO ANALYSIS OF ENTREPRENEURS' PITCHES AND FUNDING PREDICTION

Presented by

GIORGOS MOLESKIS

Research Supervisor ______________________________

Dr. George Pallis

Committee Member ______________________________

Dr. Marios Dikaiakos

Committee Member ______________________________

Dr. Andreas Aristidou

CREDITS

The most challenging part of my academic career so far has come to an end. This experience provided me with new knowledge and challenged me as much as a student and as a person. Undoubtedly, I could never do this without the guidance of my professor Dr. Pallis and my supervisor Mr. Stefanidis whom I thank for generously sharing their knowledge. Last but not least, I would like to thank my wife for supporting me emotionally through this hard period of my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Motivation

In the competitive times we now live it is highly important to be able to identify and deeply understand the characteristics that provide us, as entrepreneurs, as students or as just members of a society, with advantages that can help achieve our goals. The computer science field opens countless doors when it comes to predicting if a characteristic or an aspect is eventually helpful for and will benefit someone's goals or if there are improvements and changes a person can make in order to have better results. I have always found the ability to predict an outcome very important as it can be the best way to begin any task. Prediction is the source of our success as species. With the help of advanced science and statistics we can predict the weather, the changes in economics, even the movements of planets around the earth. It is only logical that we should use predicting algorithms in order to benefit in a lot of the aspects of our lives. When a speaker or presenter tries to communicate with others and persuade them to share the same believes as them, they use technics, language and even specific words or phrases that the listener will find appealing. Also, the speaker sends a message to the listener not only with their words but with their appearance as well. Another important part is the way a presenter talks and moves as the presentation goes on.

A lot of people find it hard to communicate and send the right message to the listener and most of the time they pass unnoticed or even get misunderstood. Therefor it is important if they could predict their possibility of success and be educated on how to change in order to achieve. These are the reasons why I chose this topic for my thesis.

1.2 Goal

The goal of this project is to help the reader better understand the characteristics and features of a good presenter and to create an algorithm that will analyze data as gender, age, previous business success, other commitments, sales etc. in order to predict the potential success of the presenter. This goal is important because through the achievement of the goal, an advantage is given to every entrepreneur who can now predict their success or failure and understand deeper the aspects of their work that need improvement. In addition, an algorithm like this can be used on big data to separate the entrepreneurs that can potentially succeed in very limited time. Furthermore, this work can benefit other researchers who can built on this existing algorithm.

1.3 Contributions

This study is one of a few studies that use vocal data in order to predict success in entrepreneurs' funding pitches. The importance of such a model in the entrepreneurial field is utmost. Entrepreneurs can benefit from the results of this thesis by studying all the aspects of the pitch that are beneficial according to the findings. Also, it is possible to use the algorithm again and again to test small changes in the pitch, testing the importance they appear to have on the outcome. In addition, it is important to state that this prediction model can be used in several cases and fields such as medical research, psychology examination, even in education to test how vocal elements affect the learning process.

1.4 Hypotheses

Hypothesis 1: "When delivering a pitch, the entrepreneur's company characteristics affects the deal making."

Hypothesis 2: "When delivering a pitch, pronunciation posterior score of the entrepreneur's voice has a positive effect on the deal making."

Hypothesis 3: "When delivering a pitch, the display of strong emotion on the entrepreneur's voice has a positive effect on the deal making."

1.5 Document Structure

This thesis consists of five chapters. Chapter 1, the introduction, has four subsections (1.1 Motivation, 1.2 Contributions, 1.3 Document Structure, 1.4 Goal). Chapter 2 - Related work consists of three subsections (2.1 Investors' Requirements, 2.2 Characteristics of a good entrepreneur funding pitch, 2.3 Voice). Following the previous two chapters is Chapter 3, the Methodology, that consists of three subsections (3.1 Data Source, 3.2 Data Annotation and 3.3 Variables description). Chapter 4 - Data analysis has seven subsections (4.1 Dataset Preparation for statistical analysis, 4.2 Demographic Data of the sample, 4.3 Findings of the research regarding all the participants, 4.4 Findings of the research regarding participants that managed to make a deal, 4.5 Statistical Analysis, 4.6 Training Models and 4.7 Performance Evaluation). Last, Chapter 5 is the Conclusion that is followed by the Bibliography.

Chapter 2

Related Work

Emotions are discrete affective states that allow individuals to respond to evolutionarily important threats and opportunities[1] Although as the current knowledge about emotions indicates that we feel the way we do because of internal experiences. The synthesis of people emotion is always unique and it is clear that sometimes we use behavioral patterns to successfully express our emotions to others[2] These patterns are referred to as emotional expressions or emotional expressions because they manifest an emotion to others. Emotional expressions convey information to others from which observers infer information about the expresser, such as their environment assessment and behavioral intentions[3]. According to the sources there are three types of benefits associated with investments: utilitarian, expressive, and emotional [4].

In entrepreneurship it is important that the person giving a funding pitch is able to express their passion using not only positive but also negative emotions. Entrepreneurs express anger in order to communicate an emotion like determination, but they also use fear, sadness and happy facial expressions that help them tell their story or emphasize on the aspect they value as important.

Various studies were contacted around the Shark Tank tv show and most of them are on an entrepreneurial perspective, seeking important and decisive elements that form a successful funding pitch. In their project "Blood in the water: An abductive approach to startup valuation on ABC's Shark Tank", Using the Heckman twostep procedure, Lavanchy et al. (2022) conducted a two-stage analysis. The objective of the first step was to estimate the likelihood that an entrepreneur will receive an offer. In the second step, the ratio of best-to-initial offer is regressed against the categorical bidding dynamics variables [5].

An additional study, "SharkTank Deal Prediction: Dataset and Computational Model" from T. Sherk, M. -T. Tran and T. V. Nguyen,2019 intended to predict whether a startup will make a deal with "sharks." The study revealed that entrepreneurs with lower initial equity contributions are more likely to receive an offer [6].

SharkTank was also studied by Naimah Ahmed S. Al-Ghamdi in his study "Cross-Cultural Linguistic Analysis of Persuasive Techniques in Shark Tank". This research makes use of William McGuire's Model of Persuasion as its theoretical basis, and it analyzes two episodes of the show "Shark Tank" (American and Saudi versions). The methodology utilized in this research is known as qualitative content analysis. The purpose of the study was to investigate the steps involved in the process of persuasion and to investigate the various strategies utilized by the contestants on the show "Shark Tank" as they related to each stage of the theory. The process of persuasion can be broken down into a series of six sequential processes, as outlined by McGuire's Model. These steps are referred to as exposure, attention, comprehension, acceptance, retention, and action. [7]

Viceisza and Smith in their published work in 2017 "Bite me! ABC's Shark Tank as a path to entrepreneurship" aimed to identify the impact of an intention-to-fund on three post-Shark Tank outcomes utilizing all the pitches, entrepreneur-contestants, and firms that had appeared on the show. These outcomes included existence one year after broadcast and patent applications.[8]

Wang et al (2021) indicate that the rate of speech, pitch, volume, and tone of a speaker can all have an impact on how their personality traits and character are perceived by listeners. Also, the researchers found that lower-pitched voices are interpreted as indicating that the speaker is more powerful/strong, competent, truthful, empathic, and trustworthy. Giving an explanation to their previous observation, Wang et. al, state that lower-pitched voices are associated with a greater capacity for empathy. In a related vein, those who speak at a quicker pace are typically perceived as being more fluent, competent, socially attractive, truthful, and persuasive.[9] The research showed important information regarding how being focused affects the fundraising outcome. It seems that the ability to maintain focus is positively associated with successful fundraising, whereas extreme emotion is negatively associated with the same outcome. On the other hand, it does not appear that stress is associated with success. According to the source, using data sets from two Kickstarter categories, gives a consistent impact of information in the speaker vocal tones on funding outcomes. It is also pointed out that there is not a way to observe receivers' perception of persuaders' competence, and that is the reason why experimental setups are needed to show that perceptions of competence mediate the relationship between vocal tone and funding success.

In their published piece "Can you hear me now? Alison et al., in "Encouraging views of passion and preparedness with vocal expressions in crowdfunding pitches", identified the effect of vocal expressions on perceptions of passion and preparedness, and by extension, funding. Valence-arousal congruence in vocal expressions enhanced financing through the perception of preparation. According to the research, high-arousal expressions result in increased perceptions of enthusiasm and greater funding, regardless of whether they have a positive or negative valence. This may come as a surprise, considering the popular belief that entrepreneurs' pitches are or should be positive. Regarding the duration of vocal emotions, pitches are predominantly favorable (87%). Findings imply that brief negative, high-arousal expressions may have a disproportionate effect on views that the expressing entrepreneur is passionate, suggesting a disproportionate influence of negative expressions on attitudes and decision-making compared to positive expressions. Thus, entrepreneurs who exhibit rage have the opportunity to demonstrate how passionate they are about their firm, and potentially how passionate they are about the chance to overcome whatever is causing their fury. Findings indicate that the extent to which an audience views a presenter's zeal and preparedness influences the amount of cash offered.[10]

2.1 Investors' Requirements

The study of Clark et al.,2008 discovered a clear and statistically significant correlation between business investors' perceptions of the quality and content of entrepreneurs' presentations and their decisions to pursue an entrepreneur's investment opportunity. Presentational (rather than non-presentational) factors tended to have the greatest influence on the presentation's overall score[11]. While at least the 'Interested' business investors did not appear to require entrepreneurs to deliver a flawless presentation in order to convince them of the investment opportunity's merits, their presentational comments consistently revealed that the clarity, understandability, and structure of a presentation, the level and type of investment-related information provided, the personal attributes of the entrepreneur, and whether the entrepreneurs had a track record of success were the most important factors.

Collectively, these findings indicate that the business investors’ evaluations of what constituted a pursuable investment opportunity were based not only on the investment-related content of the entrepreneurs' presentations or the traditional "human capital" elements of an investment opportunity, but also on the manner in which the investment-related content was presented and the perceived attributes of the entrepreneurs who presented it. At the very least, the findings presented in this paper provide additional evidence that the manner in which an investment opportunity is presented and the means by which entrepreneurs communicate themselves, their companies, and their investment opportunities is a topic that warrants much more research[11]. According to the research of Huang and Pearce (2015) there are similarities but also differences in the way investors make decisions based on their gender. The findings of their research are presented in Table 2.1.

Table 2.1 [12]

Actions and thoughts on investments	Female investors	Male investors
Investors are concerned with past dividends paid by corporations when purchasing equity shares. Those companies with a high dividend yield are deemed desirable for investment purposes;	✓	✓
Investors analyze financial ratios such as the P/E ratio, D/P ratio, and other liquidity ratios, whereas female investors, due to a lack of financial literacy, are not as familiar with financial ratios.		✓

Investors analyze the company's current financial position in terms of profitability, liquidity, and performance in terms of productivity and innovation.		✓
Investors examine daily reports published by stock exchanges on gainers and losers before investing in equity shares.		✓
Investors seek past bonuses paid to shareholders by the company.	✓	✓
Investors are interested in the effectiveness and capability of the company's management and Board of Directors structure. When investing in a company's shares, the qualifications, experience, and professional expertise of the CEO, MD, and directors are considered.	✓	✓
When investing in equity shares, investors accept the recommendations of reputable and trusted stockbrokers or experts.	✓	✓
Investors give more weight to the advice of their friends and family, and on their advice, they invest in the stock market.	✓	
Investors are more concerned with the safety of their investments; they invest more cautiously.	✓	
Investors are inspired by those who have achieved success in share investing.	✓	

According to Tyebjee and Bruno (1984), in order to make a capital investment five steps have to be taken[13]. The five-step procedure is presented in the table below (Table 2.2)

Table 2.2 [13]

STEP 1	The process by which potential investments are presented to a venture capitalist for consideration.
STEP 2	Screening, the phase in which the venture capitalist decides whether or not to investigate the investment further.
STEP 3	Evaluation, a phase (or phases) in which the venture capitalist conducts a comprehensive analysis of the venture.
STEP 4	The step of structuring a transaction, in which the framework of an investment is established. If accepted by both parties, the transaction is finalized.
STEP 5	Post-investment activity in which the venture capitalist monitors and is more or less involved with the company.

Manson and Stark in their published work in 2004 describe an investment criteria description. According to the source the background, experience, and track record of the Team entrepreneur, their personal qualities, and the range of skills possessed by the management team are essential[14]. In addition, the business's overall concept and strategy must be taken into consideration. Also, the way the business is structured to produce and deliver product and the market's potential and growth, demonstrated market need, level of competition and entry barriers are equally important. Another factor that appears to be prioritized is the concept, uniqueness, distinction, and originality of the product or service. It also includes the product or service's quality, standards, performance, appearance, styling, and aesthetic appeal, as well as its ergonomics, functionality, and adaptability. One of the most important parts of the decision making is the financial considerations. This includes the financial structure of the business, the value of the equity/worth of the business, and the expected rate of return and potential exit routes. Additionally, the relationship between the investor's background, skills, and knowledge of the industry, market, technology, etc. and the investment opportunity as well as the investor's preferences, always play an important role in the outcome of investing or not.[14]

Table 2.3 [15]

Criteria	Number of articles which mentioned the criteria.
Potential market growth rate	37
Track-record of the entrepreneur	31
Market/industry familiarity of the entrepreneurs	30
Investor’s required rate of return	30
Exit opportunity and liquidation	28
Stage of product development	21
General management expertise of the entrepreneur	20
Investors familiarity with the industry	20
Ability of the entrepreneur to evaluate and react well to risk	19
Entrepreneur and investor personality compatibility	19

According to the bibliography there are multiple criteria that investors take into consideration before investing in a company or product. According to the research of Ferrati and Muffatto (2021) the main criteria that are described in the articles they have studied are divided into ten categories[15]. The ten categories and the number of bibliographical references that support each category as essential in the investors’ decision making are shown in Table 2.3.

2.2 Characteristics of a good entrepreneur funding pitch

Entrepreneurship is a process of social, physical, and mental transformation. In contemplating entrepreneurial action, aspiring entrepreneurs transition into roles that may emphasize the necessity of adopting an entrepreneurially oriented occupational. As social and task involvement for entrepreneurial activities requires more commitment and interaction with members who already possess the requisite role identity, these transmission activities reactivate cycles of iteration [16] Nonetheless, it can be difficult to acquire and maintain a new entrepreneurial identity. Typically, well-organized presentations adhere to a simple three-part structure. According to psychologists, we remember most what is highlighted effectively and presented last. Experts recommend that we create "the residual message," repeat it multiple times, and conclude with it. This should be the primary takeaway you wish your audience to have from your presentation.[17]

Numerous significant decisions are made following human interactions. Frequently, these interpersonal interactions at the face-to-face level are a crucial step in reaching pertinent decisions[18].Researchers from various disciplines concur on the significance of such human interactions. According to psychologists and sociologists, humans are "social animals." Through face-to-face interactions, humans form social perceptions of the other party, which can influence their behaviors and decisions[18]. Economists, who frequently model humans as rational, also value interactions because the decision maker can glean additional information about the other party, such as communication skills or confidence in a business idea, from human interactions.

Initially, human interactions are complicated. In terms of data, they consist of unstructured, high-dimensional information with a low signal-to-noise ratio. Parsing and representing human interactions, therefore, requires reasonable economic and methodological structures. Even with a structure to which human interactions can be projected, capturing and measuring them is difficult[18]. Due to conceptual, methodological, and data obstacles, systematic studies on the relationship between human interactions and economic decisions are still uncommon.[18]

Entrepreneurs create startup narratives to organize information about the new venture into a coherent whole and to reduce uncertainty[19]. According to Pentland (2010), success prediction must be based on the decisions of the judges and not on the quality of the ideas. Pentland (2010) discovered that the manner in which the plan was presented was related to their ratings and concluded that they were not listening to the facts, but rather the presenter's enthusiasm, passion, or "how excited the presenter was about the plan."[20] Narrative theory posits that audiences will accept a story that relies on performance, persuasion, and symbolic representation, and therefore the logic of good reasons trumps the logic of scientific reasoning and empirical evidence[21]. Founders use narratives to persuade their audience to base their decisions on narrative logic (good reasons). Co-founders interpret, evaluate, and determine whether to contribute resources based on the two narrative rationality criteria of coherence and resonance. Coherence indicates whether a story is believable and whether its characters behave consistently and without contradictions[19].

According to previous research, an entrepreneur's enthusiasm can elicit favorable responses and evaluations from prospective investors[22]. However, a number of studies have found that the entrepreneur's enthusiasm has no effect on the funding intentions of investors [23]. Recent research has even revealed a negative effect of expressing positive emotions on funding outcomes, prompting speculation that, in certain situations, positive emotions may be perceived as nothing more than a persuasion technique[24].

The research of Jiang et al.,2022 indicates that the interpersonal impact of entrepreneur emotions is mediated by both affective (automatic) and cognitive (deliberate) processes[25]. This effort is similar to a recent study that sought to advance our understanding of how the entrepreneur's passion may influence employees through affective and cognitive processes [26]. According to the theory of emotion in interpersonal contexts [27], both processes are significant because potential funders' decisions are influenced not only by the emotions evoked automatically while watching an entrepreneur's pitch, but also by how the funders interpret the displayed emotions.
Entrepreneurial activity plays a crucial role in the creation of new businesses and innovations. Therefore, entrepreneurial scholars have a long history of interest in identifying the variables that influence entrepreneurial outcomes.

In another study, Olguin and Pentland (2010) found that members of the winning teams, as determined by the judges, spoke more, were more energetic and consistent in their physical activity levels, spoke with less energy and spent more time in close proximity to others[28]. According to Olguin and Pentland (2010), "the best predictor was the average percentage of speaking time (activity) among team members," followed by "number of people met, physical activity level, consistency or variation in physical activity, speech energy, and time spent near others."[28] Prior research on entrepreneurship has examined personality traits, sociocognitive, organizational, and managerial variables to explain entrepreneurial success and outcomes. However, there is no entrepreneurship without the entrepreneur, as entrepreneurs are the drivers of entrepreneurial activity and processes. Additionally, the personal characteristics of entrepreneurs have been found to be the most significant predictors of entrepreneurial success[29] .

According to the article entrepreneurs tend to be tolerant of ambiguity, prefer autonomy (autonomy can be described as self-reliance, dominance, and independence), resist conformity, be interpersonally aloof yet socially adept, enjoy taking risks, adapt easily to change, and have a low need for support [25]. These factors can result in significant problems with delegation and communication, two factors of utmost importance to a growing concern. They may also cause the entrepreneur to experience intense stress or isolation.[30]

Variations in the social skills of entrepreneurs likely influence the efficacy of their interactions with key stakeholders. Some entrepreneurs, for instance, lack the expressiveness required to effectively communicate their emotions and present themselves favorably in accordance with contextual display rules, which has implications for their ability to influence others. [31] This may, for instance, affect subjective ratings of project quality. Similarly, some individuals are more adept at social perception and attuned to the emotional expressions of others (e.g., emotional intelligence). Consequently, such individuals may be able to infer the emotions of others with greater accuracy, especially in terms of drawing conclusions about motivations, intentions, and confidence [31]

Adaptability does not negate the need to acquire the skills and techniques required of a capable executive. In another article, the authors hypothesize that an entrepreneur must be a competent executive and must also possess a variety of psychological traits to a greater or lesser degree than their corporate counterparts.[30] This does not suggest that all entrepreneurs are identical, nor does it imply that all managers or executives are identical. Neither is it true that a higher or lower level of a psychological trait or characteristic is sufficient for success. There appear to be a variety of sociological, psychological, demographic, and economic factors that influence the decision to enter entrepreneurial occupations. Although neither the absolute impact of a psychological trait nor the interrelationship of the combined factors on the final decision-making process is known, research indicates that there are significant differences in the intensity level of psychological traits or characteristics between entrepreneurs and managers or executives.[30]

2.3 Voice

Due to the fact that voice is a distinctive trait that conveys socially relevant information about traits and personality[32], it has significant implications in the context of entrepreneurial pitching and may explain why some entrepreneurs are more effective than others at convincing potential investors. Additionally, entrepreneurs can engage in distinct cognitive processes that can result in the identification of entrepreneurial opportunities and their success [33].

As investors have gendered expectations for how entrepreneurs should act, behave, speak, and respond during a venture pitch, entrepreneurs often attempt to persuade investors by displaying gender-appropriate [34]. As it requires agentic qualities, leadership skills, dominance, and assertiveness, entrepreneurship has frequently been viewed as a male-identified occupation [34]. These conceptions of entrepreneurship have frequently disadvantaged female entrepreneurs[35].

As speakers frequently express their masculinity or femininity through their voice, entrepreneurs' use of gendered voice during pitching can provide subtle cues to convince investors. While previous literature has suggested that women should portray masculinity during pitch presentations to convince investors, such wholesale adoption of masculinity and aggressiveness can result in negative outcomes for female entrepreneurs, as it conflicts with investors' gender role expectations and can lead to prejudice and backlash [35]. As the manner in which entrepreneurs communicate with and seek resources from potential investors is crucial for persuading them and securing funding from them [36], the ability to deliver successful venture pitches is essential for entrepreneurs.

Vocally attractive individuals are evaluated more favorably than vocally unattractive individuals. As human voice is a distinctive trait that conveys socially significant information about individual traits and personality, gender identity is frequently indexed in voice, and speakers frequently express their masculinity or femininity through voice [37].

Chapter 3

METHODOLOGY

The goal of this thesis is to predict whether an entrepreneur can secure funding based on their pitch. In Figure 3.1: Proposed Framework , the necessary steps to accomplish this task are shown. The steps are: collect data source, annotate data, preprocessing, feature extraction and prediction.

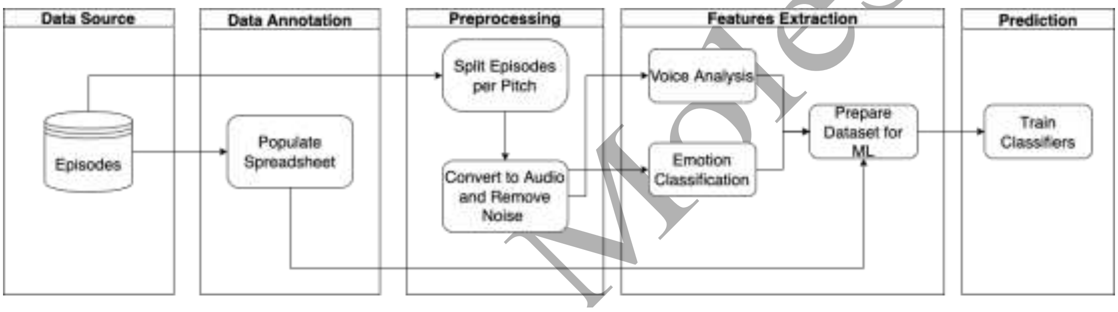

Figure 3.1: Proposed Framework

3.1 Data Source

The data source used consists of video data of 12 seasons of the TV show Shark Tank US. The first episode of the American business reality television program Shark Tank aired on ABC on August 9, 2009 and had a duration of 13 seasons. The program is the American adaptation of the global format Dragons' Den, which made its debut in Japan in 2001 under the name Money Tigers. It depicts business owners presenting their plans to a group of five investors, or “sharks” who will ultimately decide whether to fund their venture. The sharks are compensated for their roles as cast members, but the money they invest is their own. If a panel member is interested, the entrepreneur can make a handshake transaction (gentleman's agreement) on air. The sharks can compete with one another or make joint offers. They may also withdraw offers at any time if they change their minds. The entrepreneur can counteroffer or decline an offer anytime. If every panel member declines to participate, the entrepreneur leaves empty-handed.

Shark Tank accepts applications online, in 12 open casting calls across the country, and through "proactive" casting, in which producers visit trade events or contact companies directly to cherry pick entrepreneurs to apply. The applicants upload a five-minute video, and producers assign two-person teams to vet goods they like and conduct in-depth interviews with entrepreneurs after limiting the field. Each season, the program draws an average of 35,000 to 40,000 candidates, some of whom reapply after previous rejections. Approximately 1,000 candidates pass to the second phase of screening. Once the applicates pass the final screening and get the chance to pitch their idea they only get one change once the camera start recording. Pitches on average last about forty-five minutes each and they are edited into eleven minutes on average. The video editors exclude the footage that is unappealing to the audience, such as financial data, but include all essential elements that affect the final outcome [8].

The Table 3.1: Shark Tank Episodes shows the number of episodes for each season that are available. Most of the episodes are high definition with resolution at 720p and the video format is .mp4.

Table 3.1: Shark Tank Episodes

Season	Episodes
1	1 – 15
2	1 – 8
3	1 - 15
4	1 – 26
5	1 – 29
6	1 – 29
7	1 – 29
8	1 – 24
9	1 – 24
10	1 – 23
11	1 – 24
12	1 – 25

3.2 Data Source Pre-Processing

The tool used for the data source pre-processing is anaconda python. Anaconda is an open-source distribution of the python language for data science that aims to simplify package management and deployment. Anaconda's package versions are handled by the package management system conda, which examines the current environment prior to conducting an installation in order to prevent conflicts with other frameworks and packages. With anaconda environments can be created when there is a need for the same packages but different versions. Each pre-processing step is implemented in a different anaconda virtual environment.

3.2.1 Split Episodes

Before the pitches can be annotated with data, they must first be separated into different video files. There are four to five pitches in each episode. To split them, it is important to know the ending time of a pitch as well as the starting time of the next pitch. After watching a few episodes, it was noticed that Shark Tank plays the same five seconds sound each time a new pitch starts, and this sound appears the same for all of twelve seasons. This sound was extracted and used to detect when it starts playing in the video file. Using moviepy [38], a python module for video editing, the video file with type .mp4 is converted to an audio file with type .wav. In Figure 3.2 we see an example of the code that uses episode 1 of season 1 to extract the sound we hear each time a new pitch begins.

```
import moviepy.editor as mp
clip = mp.VideoFileClip("shark-tank-season-1-episode-01.mp4").subclip(160,165)
clip.audio.write_audiofile("Start.wav")
```

Figure 3.2 : Convert a video to an audio file

In Figure 3.3: Start Pitch Sound we can hear the audio file that plays before each pitch starts.

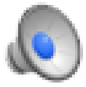

Figure 3.3: Start Pitch Sound

Using the librosa a python library, the start pitch sound and all episodes were loaded as audio time series. Then for each episode cross correlation was performed using Fast Fourier Transform. The output is an array of numbers, each of which represents the degree of similarity between the start pitch sound and the episode at each position of the episode file. In Figure 3.4: Cross Correlation for Season 10 Episode 1 we see the results of the correlation. The points marked with “x” show the top five positions in the episode with the highest correlation, which means these are the points in time where each pitch starts. These results are converted into minutes and seconds format and are saved in a text file. Knowing when each pitch begins then the time they finish can be calculated by checking when the following pitch begins. For example, we have the following times for Episode 1 of Season 10:

['0:04:53', '0:14:23', '0:22:40', '0:31:47', '0:42:33']

The final results are shown in Table 3.2: Season 10 Episode 1 Pitch Times which shows the start and end time for each pitch. Using these results, the episodes are split, and each pitch is now a different video file.

Table 3.2: Season 10 Episode 1 Pitch Times

Episode	Start Time	End Time
S10E01P01	0:04:53	0:14:22
S10E01P02	0:14:23	0:22:39
S10E01P03	0:22:40	0:31:46
S10E01P04	0:31:47	End of Episode

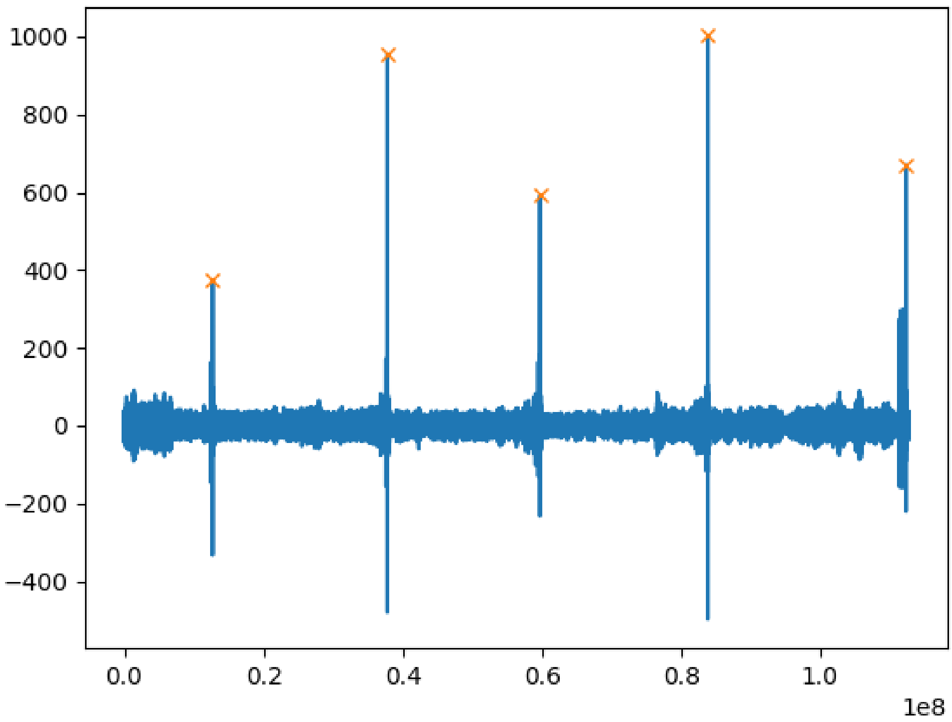

Figure 3.4: Cross Correlation for Season 10 Episode 1

```
def find_offset(within_file, find_file, window):
    #Load main video file
    y_within, sr_within = librosa.load(within_file, sr=None)
    #Load audio to search it in video file
    y_find, _ = librosa.load(find_file, sr=sr_within)

    #Perform cross correlation using Fast Fourier Transform method
    c = signal.correlate(y_within, y_find[:sr_within*window], mode='valid', method='fft')

    #find highest similarity
    peak = np.argmax(c)

    #find highest similarities and create a plot
    peaks, _ = find_peaks(c, distance=15000000, height=np.average(c))
    np.diff(peaks)
    plt.plot(c)
    plt.plot(peaks, c[peaks], "x")
    plt.savefig("cross-correlation.png")

    #keep only the 5 top similarities
    offsets = []
    while len(peaks) > 5:
        minValue = np.amin(c[peaks])
        for p in peaks:
            if c[p] == minValue:
                peaks = peaks[peaks != p]


    #calculate the time of similarities and return the result
    times = []
    for idx, p in enumerate(peaks):
        offset = round(p / sr_within, 2)
        offset = offset + librosa.get_duration(filename=find_file) + 1
        time = str(datetime.timedelta(seconds=offset))
        times.append(str(time).split(".")[0])
        offsets.append(offset)

    return times
```
Figure 3.5: Find Pitch Start Times in an Episode

3.2.2 Remove Noise from video file

Each video file pitch was converted to an audio file .wav. Using the python package spleeter [39] we are able to extract only the speaker's audio from the pitch audio file. Spleeter is a tool for music source separation with pre-trained models and it is based on TensorFlow. Figure 3.7: Season 10 Episode 1 Audio and Figure 3.6: Season 10 Episode 1 Vocals Only is an example of the spleeter results.

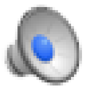

Figure 3.7: Season 10 Episode 1 Audio

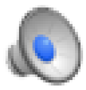

Figure 3.6: Season 10 Episode 1 Vocals Only

3.2.3 Isolate Entrepreneurs' Voices

In each pitch video file, numerous speakers are present. There are at least six investors, one entrepreneur, and the host of the show which makes isolating the entrepreneurs' voices a difficult task. To accomplish this task, we need to identify who spoke when, this process is called speaker diarization. There are many speaker diarization tools but for our data the python library "Neural speaker diarization with pyannote.audio" [40] [41] had the best results. In Table 3.3: Speaker Diarization Output there are the results of pyannote.audio for pitch S10E01P01.

Table 3.3: Speaker Diarization Output

Speaker Id	Time Start (s)	Time End (s)
SPEAKER_04	1.3884375	7.9696875
SPEAKER_04	8.3071875	49.4821875
SPEAKER_04	50.4609375	55.5571875
SPEAKER_04	55.7765625	59.5903125
SPEAKER_04	59.7253125	63.1003125
SPEAKER_04	63.7415625	76.8028125
SPEAKER_04	77.2415625	81.4940625
SPEAKER_04	82.2028125	89.0878125

SPEAKER_05	87.3665625	87.7546875
SPEAKER_01	93.0196875	95.2134375
SPEAKER_02	94.8421875	95.2471875
SPEAKER_06	95.2471875	96.8334375
SPEAKER_06	99.6178125	110.9409375
SPEAKER_06	111.5653125	116.1721875

By calculating the total sum of seconds each speaker is active and checking which speaker was the first to speak the entrepreneur is identified. Each pitch starts with the entrepreneur greeting the investors and the person which speaks most of the time in the presentation is the entrepreneur and main presenter. A new audio file is created based on the speaker id and the diarization results which contains the parts the entrepreneur is speaking only.

3.3 Data Annotation

3.3.1 Manual Annotation

The manual annotation is done by watching the entrepreneurs' pitches video files. The pitches were watched in a randomized order. This way we have a pitch from all seasons and most episodes. The total number of pitches that have been watched are four hundred sixty-seven and for each one of them the important information is saved in a spreadsheet.

3.3.2 Audio Analysis

To be able to perform any audio feature analysis the audio files must be converted to .wav format, recorded at 44kHz sample frame and 16 bits of resolution. In Figure 3.8: Prepare Audio for Analysis the audio file is loaded using target sample rate of 44kHz and is saved using PCM_16 (pulse-code modulation) which is a standard encoding scheme used in the wav file format.

```
y, s = librosa.load(pathWav, sr=44000)  # Downsample to 44.0kHz
sf.write(pathWav, y, s, subtype='PCM_16')
```

Figure 3.8: Prepare Audio for Analysis

Using the python library my-voice-analysis the voice features are extracted from entrepreneur's pitch. A python program was developed to load each pitch and perform audio features analysis. The results were then saved in a spreadsheet.

3.3.3 Emotion Classification

To detect the entrepreneur's emotion the python library Audio Emotion Classification from Multiple Datasets [42] was used. The library is able to identify 8 emotions which are neutral, calm, happy, sad, angry, fearful, disgust and surprised. For our dataset most of pitches was classified as disgust and surprised incorrectly. To try and increase the accuracy the emotions fearful, disgust and surprised were removed and the model has been trained again using the datasets RAVDESS and TESS which are the same the library used to train the original model. The RAVDESS is a multimodal database of validated emotional speech and music. The database is gender-balanced, with 24 professional actors reading lexically matched lines with a neutral North American accent. [43]. Is consists of 1440 speech files and 1012 song files. Speech includes the emotions calm, happy, sad, angry, fearful, surprise, and disgust and song includes the emotions calm, happy, sad, angry, and fearful. The TESS dataset consists of 2800 files and is a set of 200 target words that were spoken by two actresses and recordings were made of the set portraying the emotions anger, disgust, fear, happiness, pleasant surprise, sadness, and neutral.

To train the new model the library Audio Emotion Classification from Multiple Datasets has been modified. First, the data set must be prepared by removing all audio files that do not correspond to the new emotions we wish to predict. This procedure generates two files, X.joblib and y.joblib, which contain the data and labels, allowing us to train new machine learning models with ease. Utilizing the two files we create the new model using the default model which is the sequential. The new model predicts the emotions neutral, calm, happy, sad and angry. Using the new model, a python program was developed to load each pitch and predict the emotion. The results were then saved in a spreadsheet.

3.4 Variables Description

To prepare the final dataset all the spreadsheets in section Data Annotation are combined into one using the unique identifier for each pitch.

Pitch ID is the unique identifier for each presentation. S stands for season, E for episode and P for the number of presentations in the episode. For example, the pitch id “S01E01P01” is referring to the first presenter of the first episode in the first season.

The first variable about the entrepreneurs is the number of entrepreneurs that pitch their idea excluding the people there to help them. To check if stereotypes affect the decision of the sharks the entrepreneur’s ethnicity [6] and entrepreneur’s gender [44] were collected. To examine an entrepreneur's financial situation, information such as debt, previous business success [5], and other job commitments are gathered.

Next the basic information about the entrepreneur’s company were collected. The company name just for reference, industry [5] (Automotive, Business Services, Children / Education, Fashion / Beauty, Fitness / Sports / Outdoors, Food and Beverage, Green/Cleantech, Health / Wellness / Cleaning, Lifestyle / Home, Media / Entertainment, Pet Products, Software / Tech, Travel, Other), if the product or service gets more sales on specific seasons, the revenue model (Production/Transactional model, Rental or leasing model, Advertising model, Licensing model, Freemium models, Subscription model), if the entrepreneur’s company provides a product or service and whether the product/service is sold in physical stores or online stores.

Another important variable about the company is if the entrepreneur’s product or service is patented [5]. A patent is a form of intellectual property that grants its owner the legal right to prevent others from creating, using, or selling an invention for a limited time in exchange for the publication of an enabling disclosure of the invention. In most nations, patent rights are governed by private law, and the patent holder must sue an infringer to enforce their rights. In some industries, patents are a crucial source of competitive advantage.

In addition, data about the economic state of the entrepreneur's company were gathered. The last year's number of sales [5], the amount of funding the entrepreneur asks for and the percentage of equity the entrepreneur is willing to give in exchange. With this information we can calculate the valuation of the company using the below formula.

$$Valuation = \frac{100}{\text{Equity (\%) (ASK)}} * \text{ Amount (\$) (ASK)}$$

Moreover, from audio analysis the features number of syllables, number of fillers and pauses, rate of speech which is the number of syllables per second, articulation (clarity of sounds and words), speaking time (excl. fillers and pause), total speaking duration (inc. fillers and pauses), ratio between speaking duration and total speaking duration. Also, the fundamental frequency statistics are extracted which can provide us insight about the perceived pitch of a speaker's voice. Moreover, the pronunciation posteriori probability score percentage is extracted which is a score based on phonemic errors like phoneme mispronunciation, syllable-level coarticulation errors, phoneme insertion, phoneme substitution, phoneme deletion and prosodic errors like stress rhythm and intonation.

Next from entrepreneur's pitch emotion classification the emotion was extracted. The possible values are neutral, calm, happy, sad and angry.

Finally, the information about the deal were collected like if there was an offer, if there was a deal[6], the deal amount, deal equity, if it was a royalty deal, if it was a loan deal and the number of sharks that closed the deal. Using this information, we can calculate the valuation from the sharks' side using the formula below:

$$Valuation = \frac{100}{\text{Equity (\%) (DEAL)}} * \text{ Amount (\$) (DEAL)}$$

Based on our hypothesis the depended variable is Deal (Y/N) and all other variables are our independent variables.

3.5 Dataset Preparation for Statistical Analysis

3.5.1 Remove records with missing values

The number of pitches for which all required information is available is less than 1091, the total number of available pitches. To remove these records, we check which ones do not have an answer if a deal has been made and are removed from the dataset. The total number of pitches remaining after cleaning the data set is four hundred sixty-seven.

3.5.2 Manage null values

Figure 3.9: Columns With Empty Values depicts the columns with empty values. These are the columns which were populated only when there was a deal between the entrepreneur and investors. The columns Amount (Deal), Equity (Deal) are numbers and are replaces with zero. The columns Royalty Deal and Loan accept answers Yes and No and we replace the empty values with No.

```
Amount ($) (DEAL)          173
Equity (%) (DEAL)          177
Royalty Deal (Y/N)         189
Loan (Y/N)                 189
```

Figure 3.9: Columns With Empty Values

3.5.3 Manage Variables Column Types

The columns that are of type object must be modified in order for the data analysis and machine learning algorithms to work. These columns contain categorical data. Categorical features have a limited and typically fixed set of possible values. There are 2 types:

- Nominal: Features where the categories are only labeled without any order. For example, Gender with options Male or Female

- Ordinal: Features where the order matters. For example, education level with options High School, BS, MS, PhD.

We can handle categorical data using Label Encoding and One-Hot Encoding.

Label Encoding assigns a unique integer based on alphabetical ordering. Columns with values Yes and No are converted to numerical values, zero for No and one for Yes. These columns are:

- Has Patent?
- Any Debt?
- Previous Business Success
- Other Job Commitments
- Seasonality
- Offer
- Royalty Deal
- Loan
- Deal

For more complicated categorical data Label Encoding cannot be used because there is a high probability that the machine learning model captures a relationship between these values, $3 > 2 > 1$.

One-Hot Encoding generates additional features based on the number of distinct values contained in the categorical feature. Each unique category value will be added as a feature. One-Hot Encoding has been performed for the following columns:

1. Emotion From Audio
2. Presenters Ethnicity
3. Industry
4. Product or Service
5. Presenters Gender
6. Revenue Model

7. Retail/E-Commerce/Both

$$VIFj = \frac{1}{1 - Rj^2}$$

The minimum value of VIF is one which means the testing variable is not correlated with any other variable. The higher the VIF is it means there is high correlation with other variables and it won't be evaluated as statistically significant. Using the python library Statsmodels' implementation of VIF the variables that have been identified with extreme multicollinearity are:

- Presenters Ethnicity White
- Retail/E-Commerce/Both(Retail)
- Product or Service(Product)
- Original duration
- f0 mean
- f0 median
- f0 quan75
- Speaking duration
- Industry (Lifestyle / Home)
- Presenters Gender(Female)
- Revenue Model(Advertising)
- Revenue Model(Production/Transactional)
- Emotion neutral
- Emotion calm
- Emotion happy
- Emotion sad

3.5.4 Data Normalization

Normalization is the process of translating data into the range [0,1]. Min-Max normalization has been used for the columns:

- Number of Sales ($) (Last Year)
- Amount ($) (ASK)
- Amount ($) (DEAL)
- Valuation
- Shark Valuation

CHAPTER 4

Data Analysis and Predictions

In this chapter the data analysis of the research done for the purposes of this thesis, will be presented. To perform data analysis Python and Jupyter Notebook was used. The Jupyter Notebook is an incredibly powerful tool for developing and presenting data science projects in an interactive manner. A notebook combines code and its output with visualizations, narrative text, mathematical equations, and other rich media in a single document. It's a single document where you can execute code, view the output, and add explanations, formulas, and charts to make your work more transparent, understandable, repeatable, and shareable. All the graphs were generated using the python library Matplotlib.

4.1 Demographic data of the sample

The demographic data of the sample are shown in the diagrams below. The total of the entrepreneurs' teams that will be discussed below is 457 teams. The Figure 4.1: Presenters' Gender Frequency shows the frequency of the participants' genders. According to the figure 259 teams out of 457 were consisted of men (56,67%) , 115 teams were consisted of women (25,16%) and 83 of the teams presenting their companies or products were mixed (18,16%).

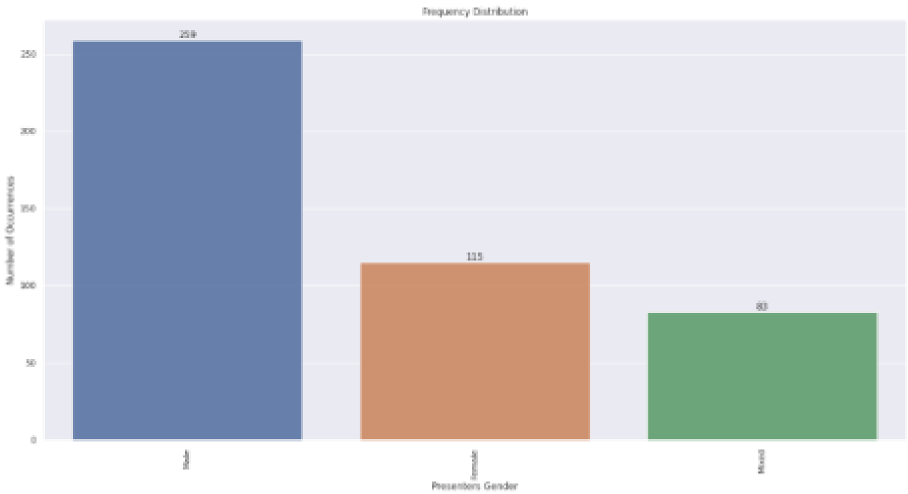

Figure 4.1: Presenters' Gender Frequency

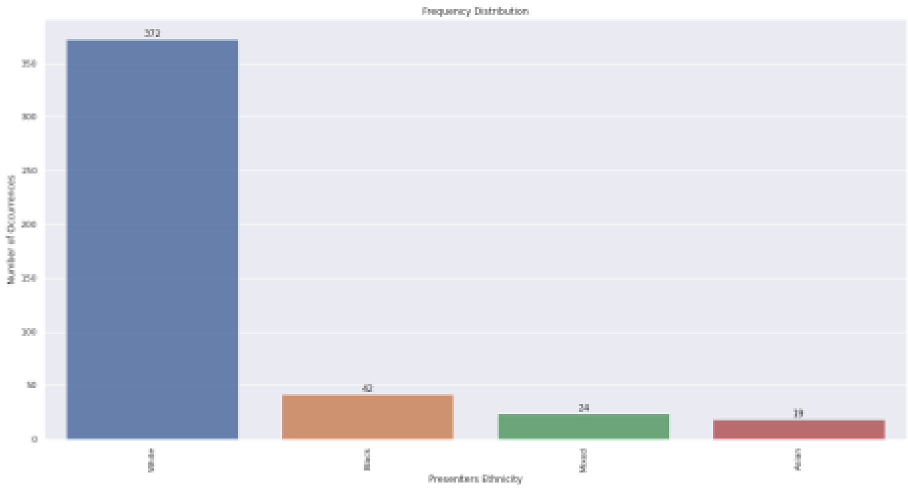

Figure 4.2:Presenters' Ethnicity Frequency

In the next figure (Figure 4.2:Presenters' Ethnicity Frequency) the ethnicity of the 457 entrepreneurs' teams is presented. As shown in the figure 372 out of 457 teams of presenters consisted of only white presenters men or women (81.40%) , 42 teams were consisted of only black entrepreneurs (9,19%), 24 of the teams were mixed (5,25%) and 19 of the teams included Asian entrepreneurs (4,16%).

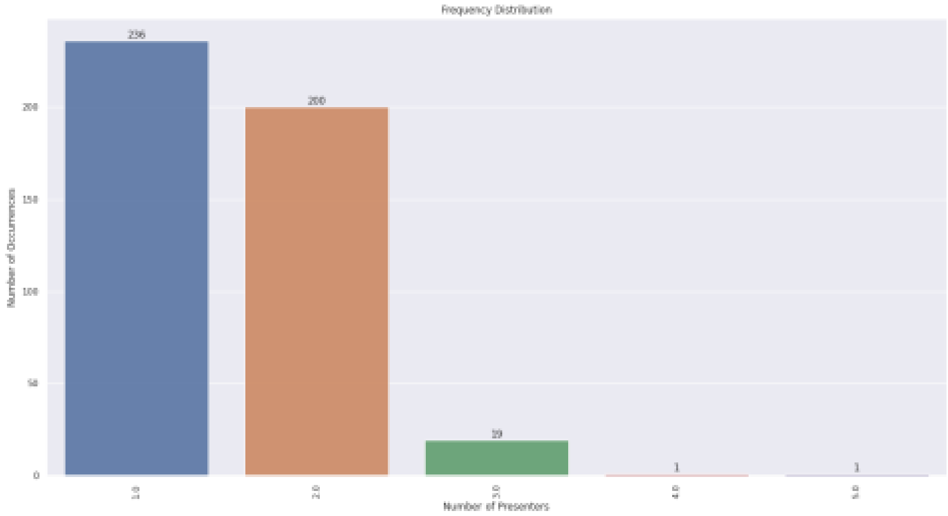

Figure 4.3:Number of Presenters per Team Frequency

As shown in the Figure 4.3:Number of Presenters per Team Frequency, the 457 teams had multiple numbers of members. 236 of the teams were consisted of one individual (51,64%), there were 200 teams of two presenters (43,76%), 19 teams of 3 presenters (4,16%), 1 team of four entrepreneurs (0,22%) and 1 team of 5 presenters (0,22%).

4.2 Findings of the research regarding all the participants

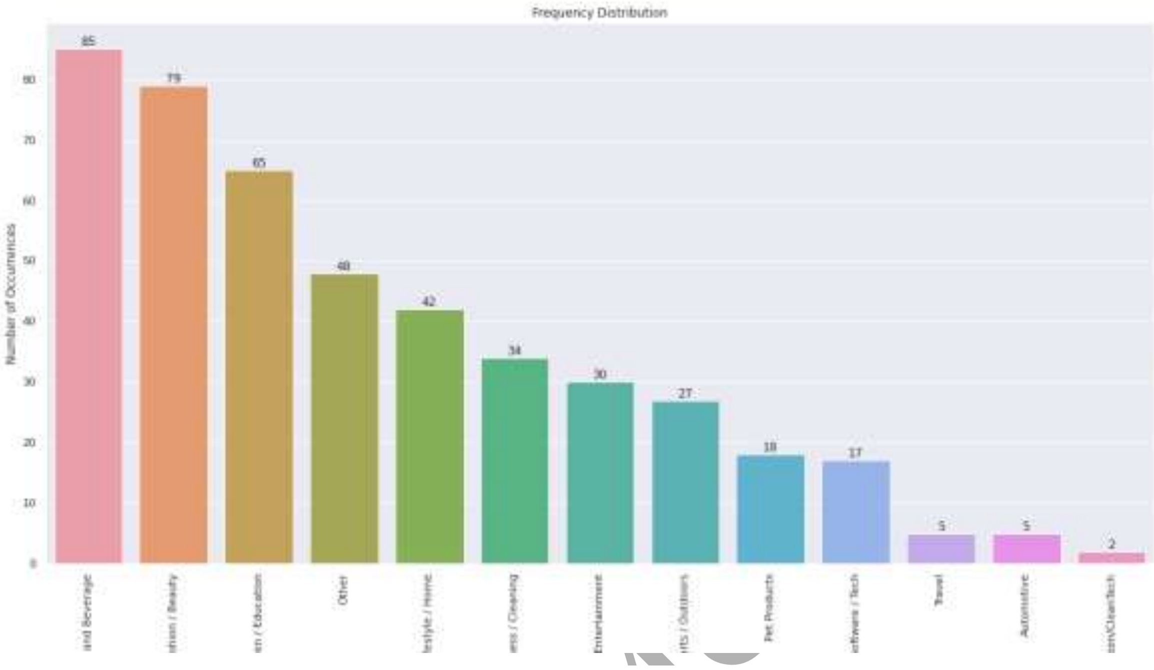

Figure 4.4:Industry Category Frequency

A shown in the Figure 4.4:Industry Category Frequency, there are 13 different industry categories appearing among the participants of the show Shark Tank. 85 out of 457 teams were in the food and beverage industry (18,60%), 79 in the beauty or fashion (17,29%), 65 of the teams were about children or education (14,22%),42 lifestyle or home (9,19%), 34 about cleaning (7,44%), 30 were in the entertainment industry(6,56%), 27 had to do with sports or outdoors activities (5,91%), 18 teams were in the pet products industry (3,94%), 17 were in the software or tech industry (3,72%), 5 were in the travel industry(1.10%), 5 were in the automotive industry (1.10%), 2 were in the green or cleantech industry (0,44%) and 48 teams were in other industries (10,50%).

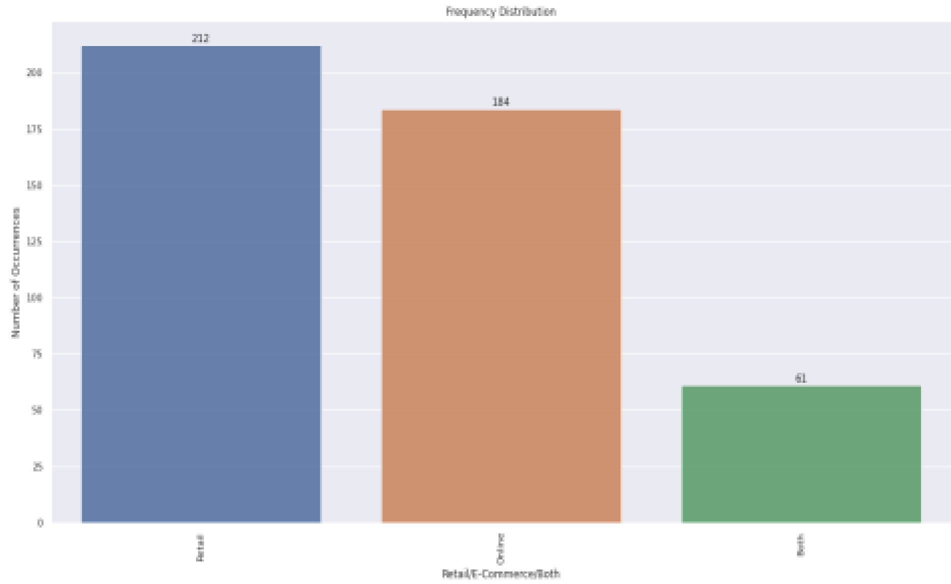

Figure 4.5:Retail/E-Commerce Frequency

Furthermore, in Figure 4.5:Retail/E-Commerce Frequency, it is shown that 212 of the 457 teams sold their product through retail (46,39%), 184 teams sold exclusively online (40,26%) and 61 teams combined both retail and online sales (13,35%).

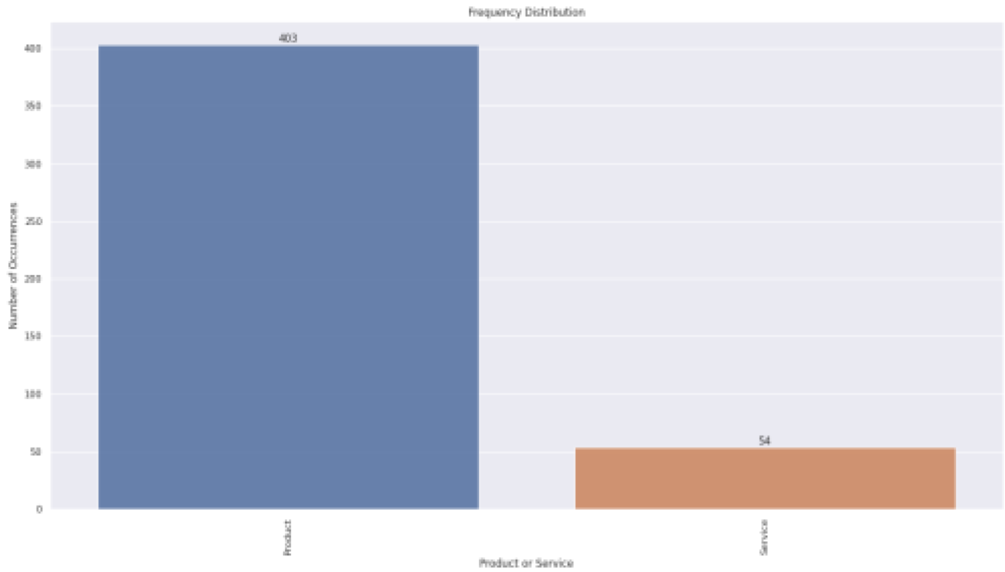

Figure 4.6:Product or Service Frequency

In the Figure 4.6:Product or Service Frequency, the number of teams that sold a product appears to be 403 teams (88,18%) and the number of teams that offered a service to the public were 54 (11,82%).

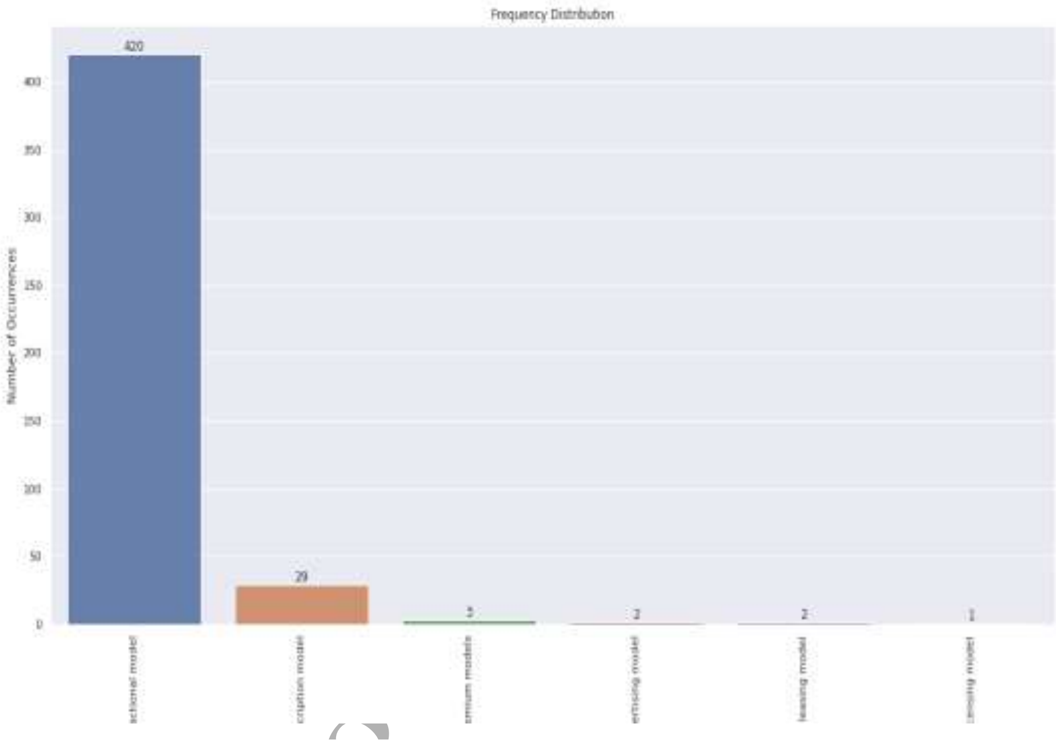

Figure 4.7: Revenue Model Frequency

The Figure 4.7: Revenue Model Frequency, shows that the majority of the teams (420 teams out of 457 that is 91,90% of the teams) use the transactional model, 29 teams use the subscription model (6,35%), 3 teams use freemium models (0,66%), 2 teams use licensing model (0,44%) and 1 team uses the advertising model (0,22%).

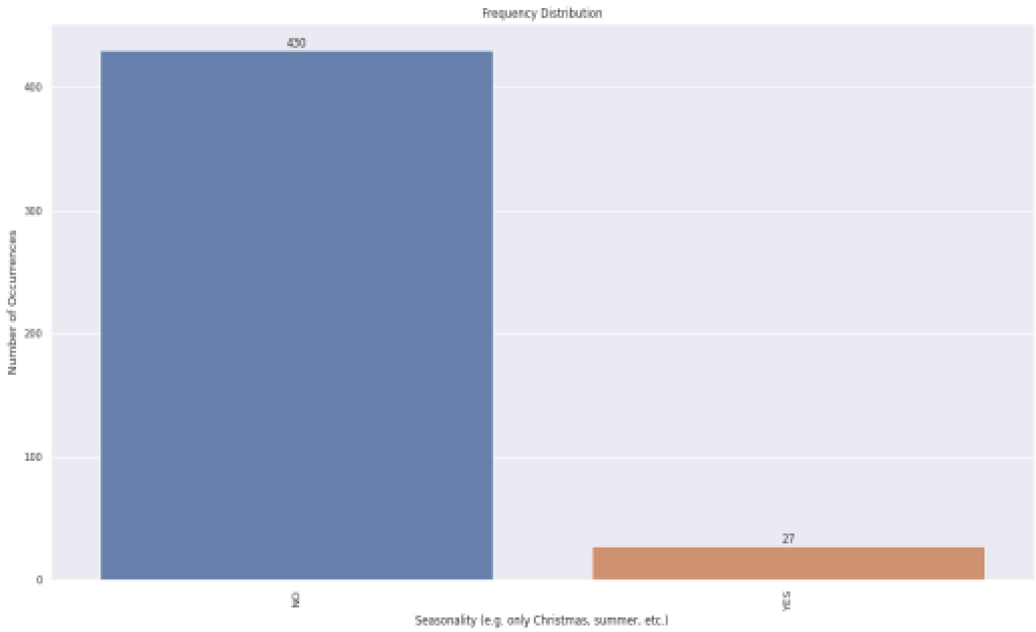

Figure 4.8: Seasonality of product Frequency

In the Figure 4.8: Seasonality of product Frequency, we can see that 430 teams (94,09%) did not sell a seasonal product and 27 teams (5,91%) chose seasonality regarding to the product or service they offer.

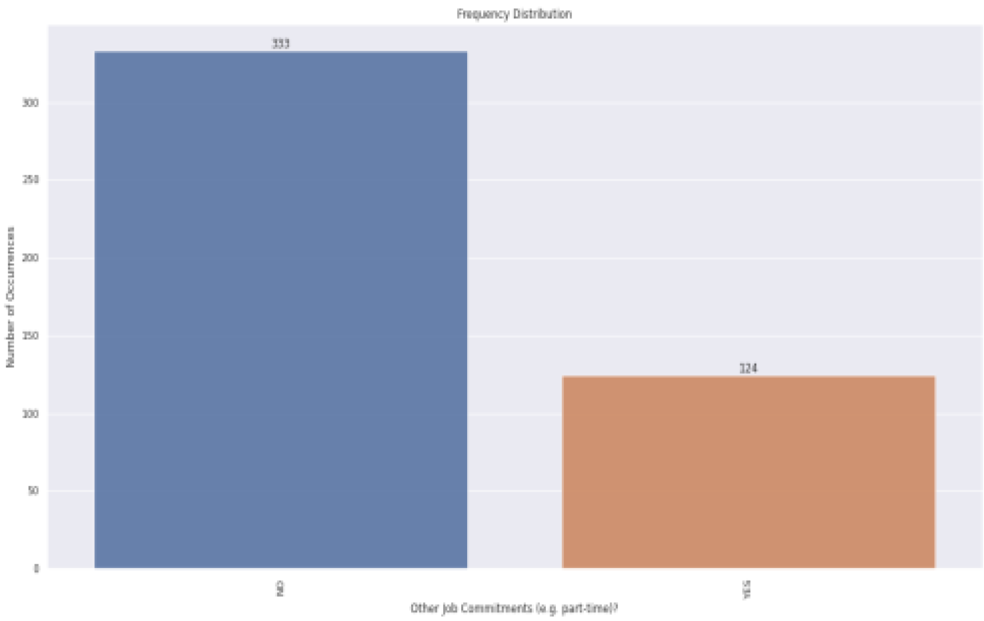

Figure 4.9:Frequency of having other job commitments

As shown in the Figure 4.9:Frequency of having other job commitments, 333 of the teams (72,87%) stated that they do not have other job or commitments and are dedicated exclusively to the company or product they presented. 124 teams out of 457 (27,13%) stated that there are other commitments in their schedules.

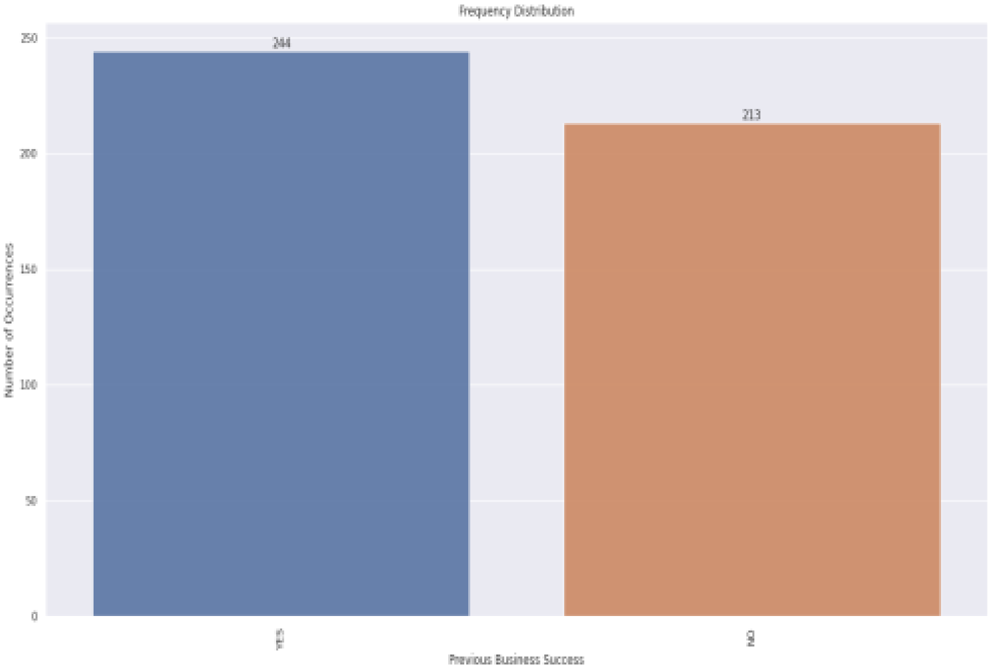

Figure 4.10:Previous Business Success Frequency

The Figure 4.10:Previous Business Success Frequency, presents that 244 of a total of 457 teams (53,39%) had a previous business success but 213 of the teams (46,61%) had no previous business success.

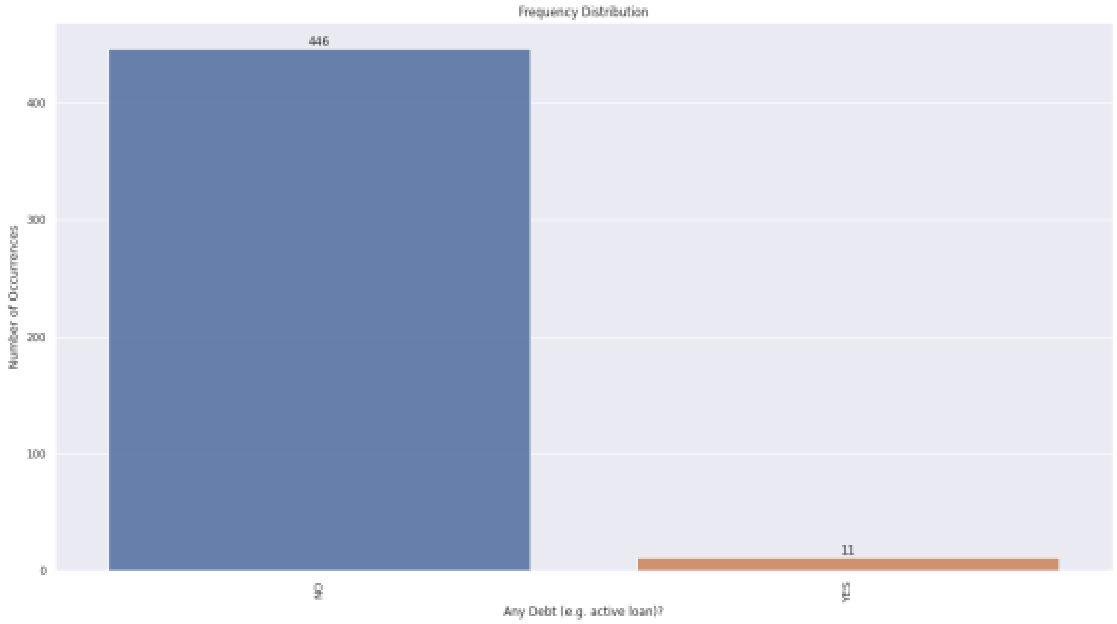

Figure 4.11:Dept or Loan Frequency

In the Figure 4.11:Dept or Loan Frequency, we can see that 446 out of 457 teams (97,59%) that presented their work on Shark Tank did not have ant depts or loans. On the other hand, 11 of the teams (2,40%) stated that they have a dept or active loan.

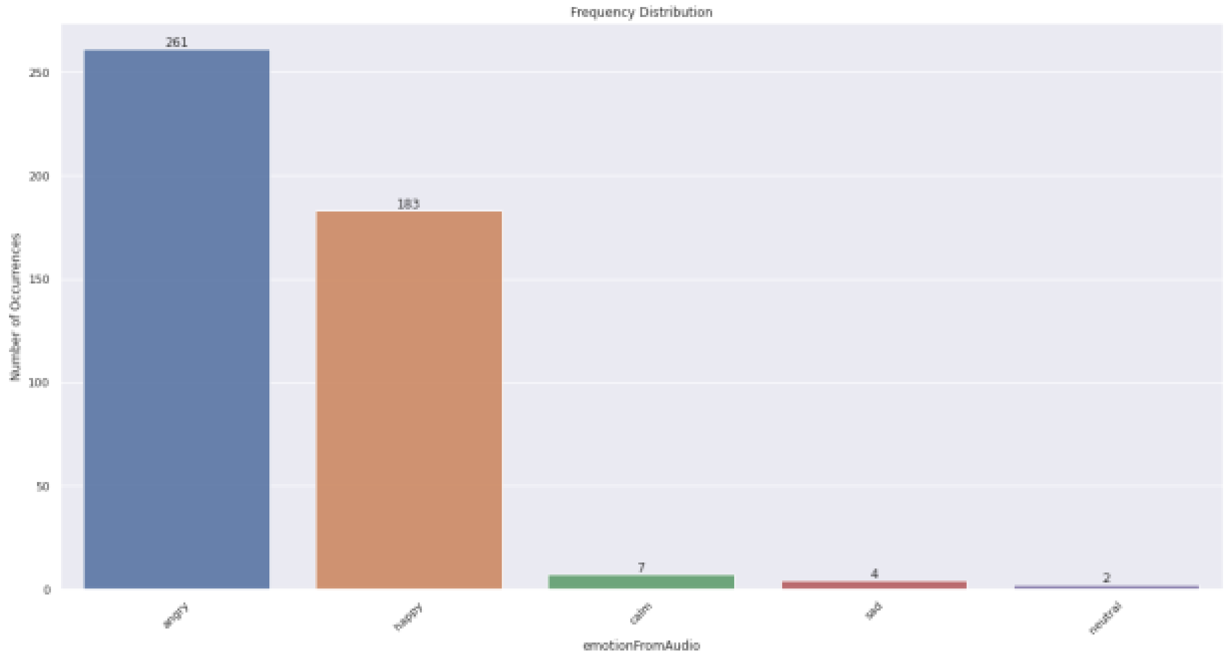

Figure 4.12:Emotion Type Frequency

After running a voice analyzing algorithm on the voices of the entrepreneurs the Figure 4.12:Emotion Type Frequency, was created. According to the figure 349 of the 457 teams (76,37%) expressed the emotion of anger throughout their presentation, 92 teams (20,13%) expressed happy emotions,14 teams expressed disgust (3,06%), 1 sadness (0,22%) and 1 team (0,22%) was neutral.

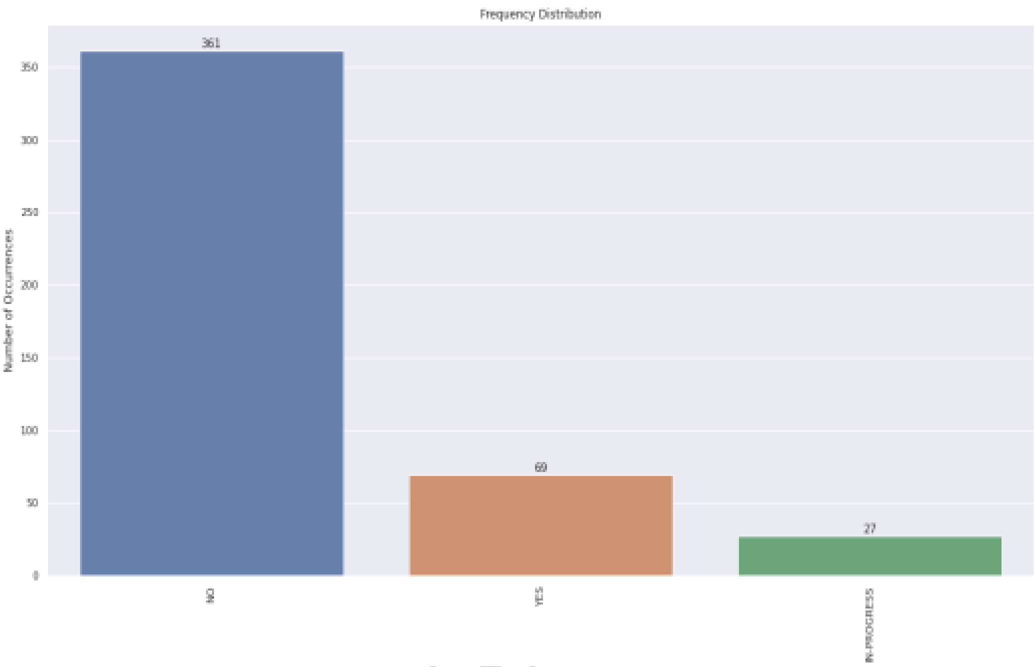

Figure 4.13:Patent Frequency

The Figure 4.13:Patent Frequency, shows that 361 out of 457 teams (78,99%) did not own a patent, 69 teams owned a patent (15,10%), and 27 teams (5,91%) had a patent or patents pending.

4.3 Findings of the research regarding participants that managed to make a deal

In this section of the thesis results the findings that refer to the teams that manage to make a deal with the sharks are presented. It is important to find out and also understand the ways these facts helped each team to accomplish their goals.

As shown in the Figure 4.14:Persentage of deal teams that had dept or loan, 58,74% of the teams that had no dept managed to make a deal with the Sharks and a percentage of 27,27% of the teams that appeared to have an active loan or dept also managed to make a deal. In the Figure 4.14:Persentage of deal teams that had dept or loan, 100% of the teams displayed the emotion of sadness managed to make a deal, 64,13% of the teams that expressed happy emotions made a deal, 56,73% of the teams that displayed anger also were successful, and 50% of the teams that expressed disgusts succeeded to make a deal. Not one of the teams that remained neutral made a deal.

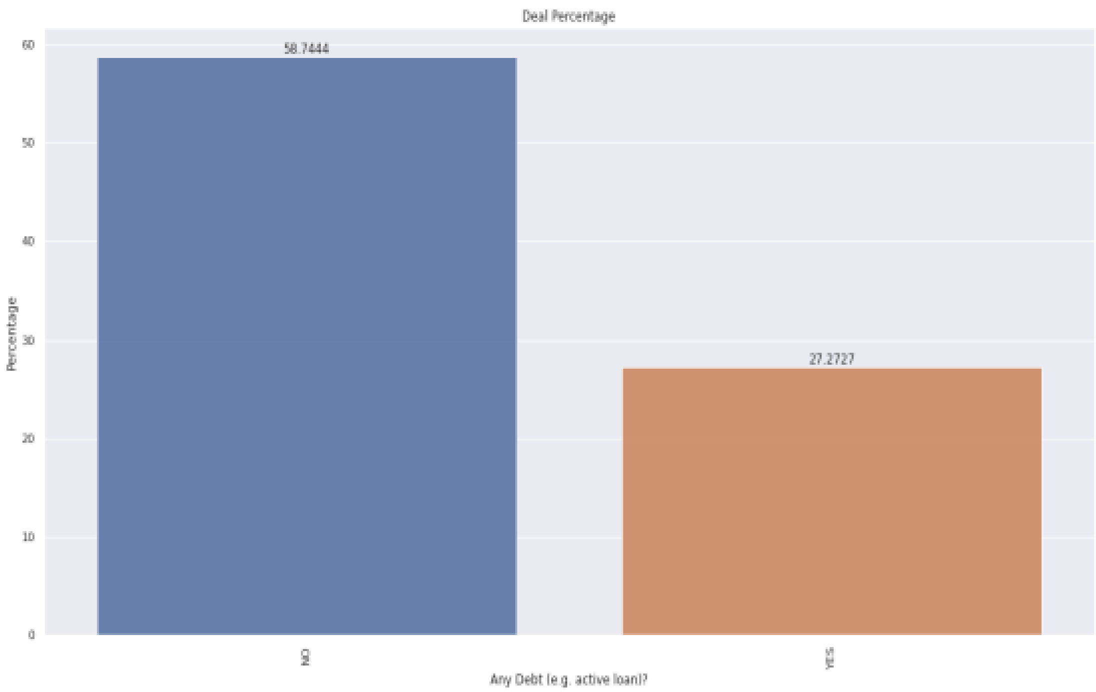

Figure 4.14:Persentage of deal teams that had dept or loan

Figure 4.14:Persentage of deal teams that had dept or loan, shows that 50,85% of the teams that had one presenter made a deal, 64,5% of the teams with 2 presenters, 73,68% of the teams that had 3 presenters were also successful. Also, all of the teams that had 4 or 5 members succeeded to secure a deal with the sharks.

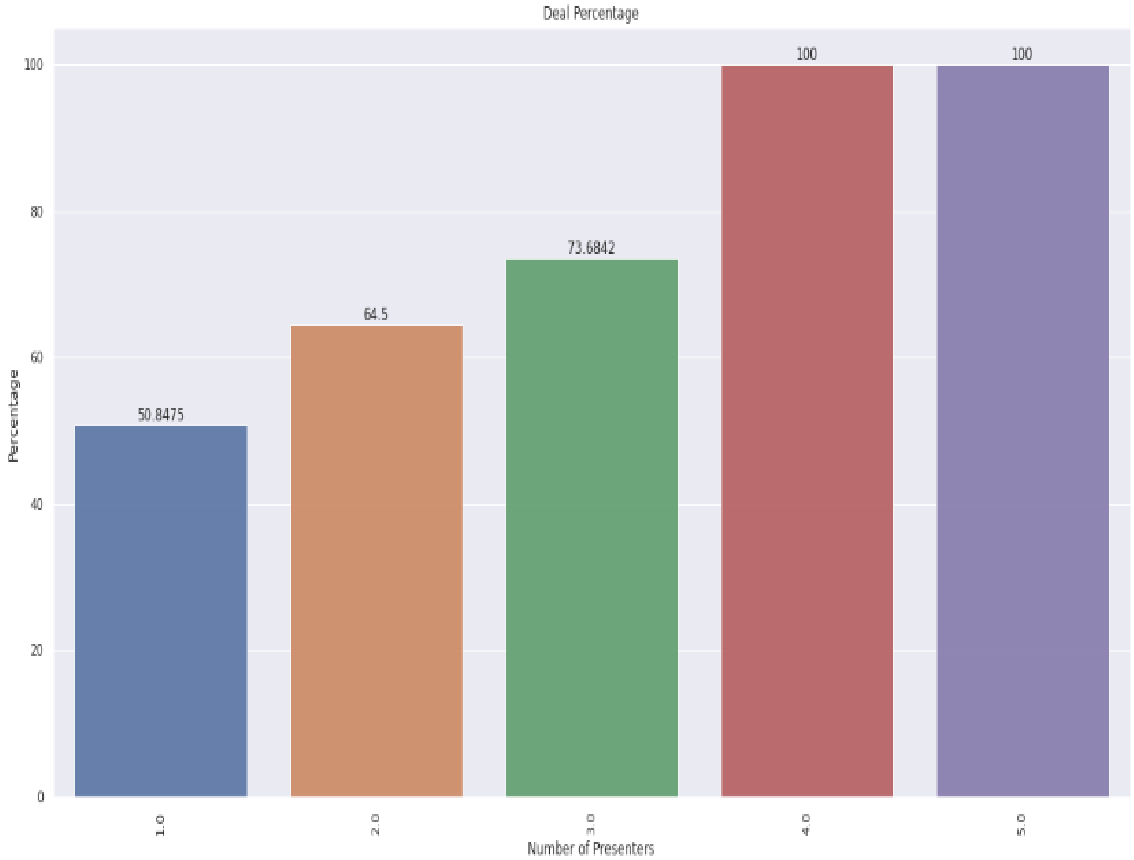

Figure 4.15:Deal percentage of Number of presenters

In Figure 4.15:Deal percentage of Number of presenters, it is shown that 50,85% of the teams that had only one presenter, 64,5% of the teams that had two presenters and 73,68% of the teams that had 3 presenters managed to make a deal with the Sharks. The teams with four or five presenters had a 100% success although this outcome needs further investigation due to the very small number of teams with four or five presenters.

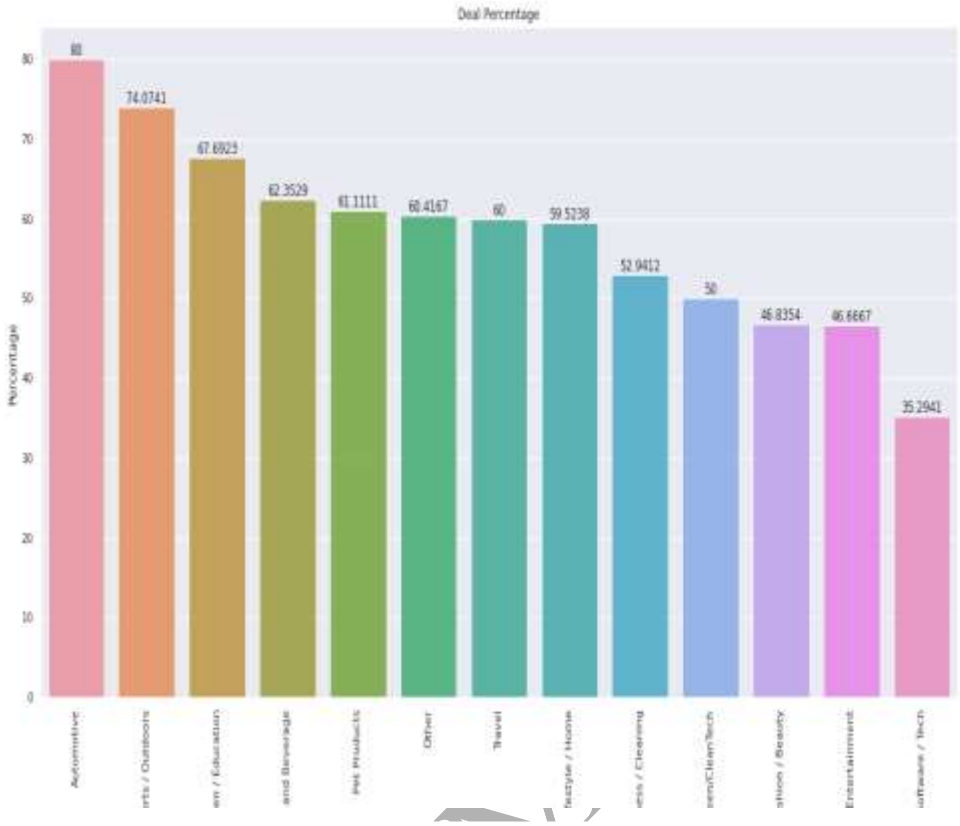

Figure 4.16:Percendage of deals of industry categories

In the Figure 4.16:Percendage of deals of industry categories, it is shown that 80% of the team in the automotive industry, 74,07% of the teams in sports or outdoors, 67.69% of teams in the children or education industry, 62,35% of the teams in the food and beverage industry, 61,11% of the teams in pet products industry, 60% of the teams in travel industry, 59,52% if the teams in lifestyle or home industry, 52,94% of the teams in cleaning industry, 50% of the teams in cleantech industry, 46,84% of the teams in the fashion and beauty industry, 46,67% of the teams in the entertainment industry and 35,29% of the teams in the software or tech industry, successfully made a deal with the sharks. Also, a percentage of 60,42% of other industry field had successfully made a deal.

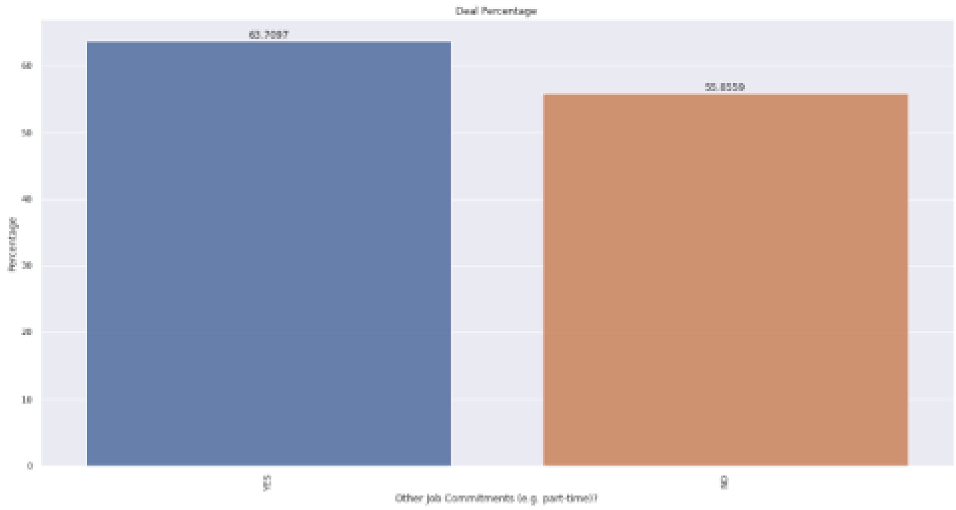

Figure 4.17:Percentage of deal teams with other commitments

The 63,71% of the teams that appeared to have other commitments successfully sealed a deal and the 55,86% of the teams that had no other commitments made a deal as well, as shown in the Figure 4.17:Percentage of deal teams with other commitments.

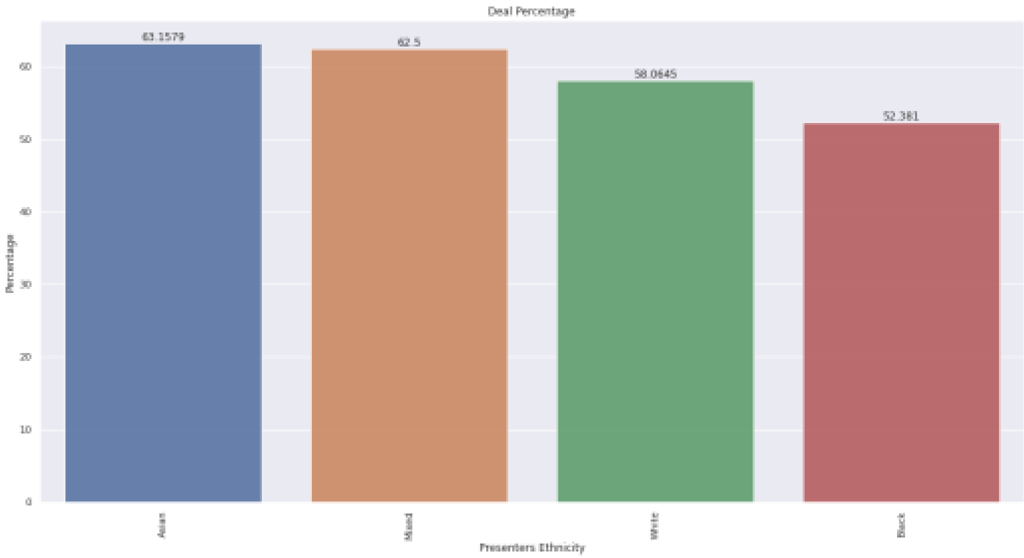

Figure 4.18: Percentage of deals for each ethnicity group

In the Figure 4.18: Percentage of deals for each ethnicity group, it is presented that 63,16% of Asian teams, 62,5% of mixed teams, 58,10% of white and 52,38% of black entrepreneurs' teams managed to make a deal.

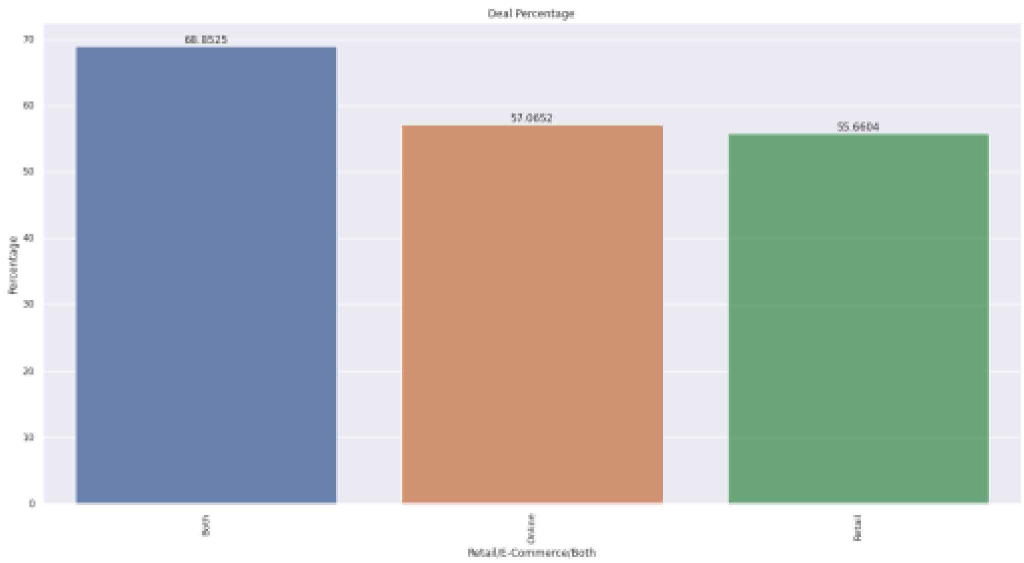

Figure 4.19:Percentage of retail/e-commerce/both that made a deal

The percentage of teams that used both retail and online sales that successfully made a deal was 68,85%. Also, 57.07% of the teams that only sell online managed to make a deal and 55,66% of the teams that only sell on retail made a deal as shown in the Figure 4.19:Percentage of retail/e-commerce/both that made a deal.

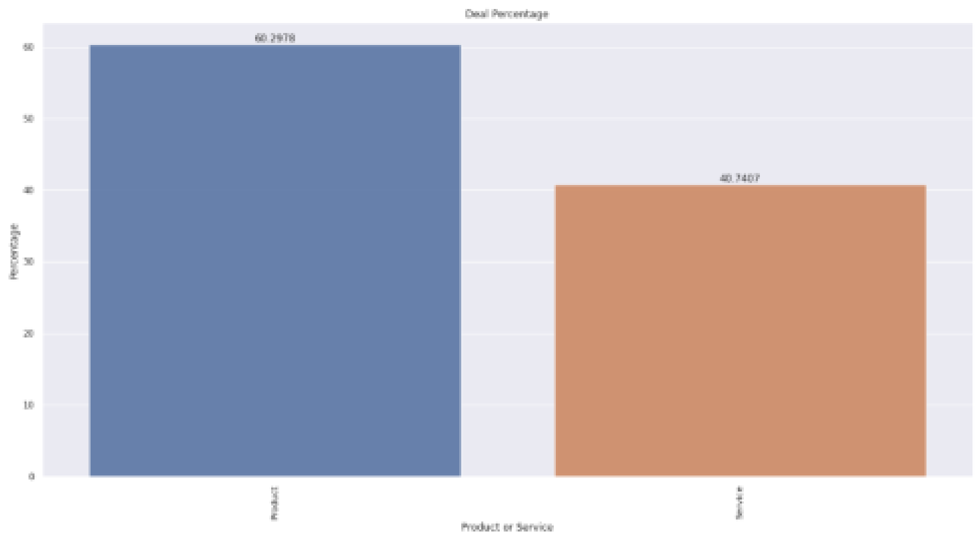

Figure 4.20:Percentage of deals with product and service

In the Figure 4.20:Percentage of deals with product and service, it is shown that 60,30% of the teams that sell a product and 40,74% of the teams that sell a service managed to make a deal.

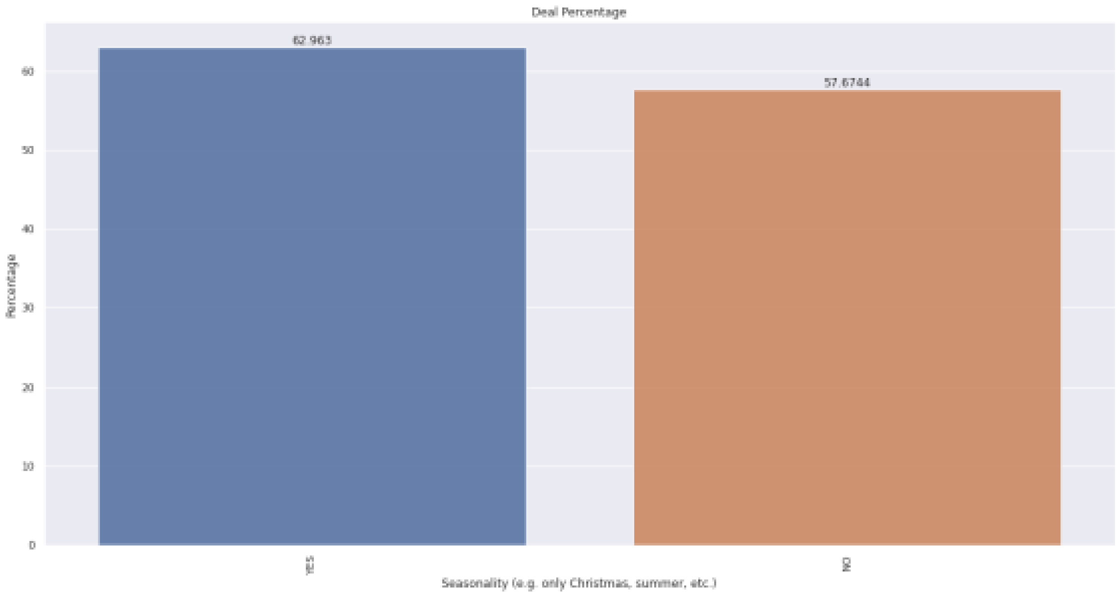

Figure 4.21:Percentage of deal of seasonal and nonseasonal products

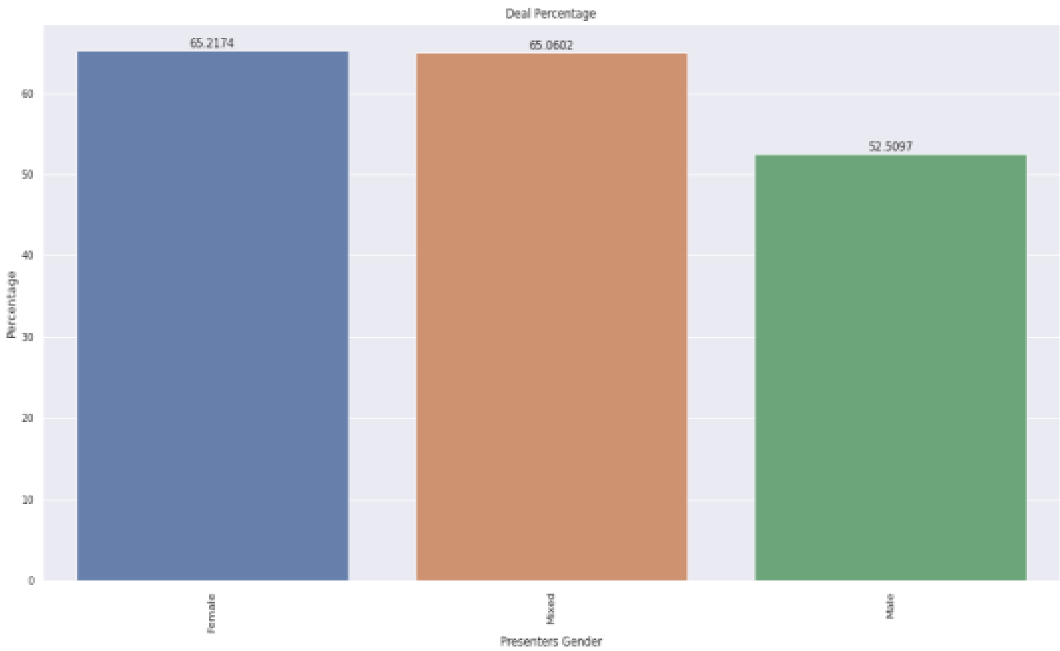

Figure 4.22:Percentage of deals per gender

According to the Figure 4.22:Percentage of deals per gender, 62,96% of the teams that sell seasonal products or services and 57,67% of the teams that sell non seasonal products or services have made a deal with the sharks.

As the Figure 3.23 presents 65,22% of female teams, 65.06% of mixed teams and 52.51% of men teams managed to make a deal.

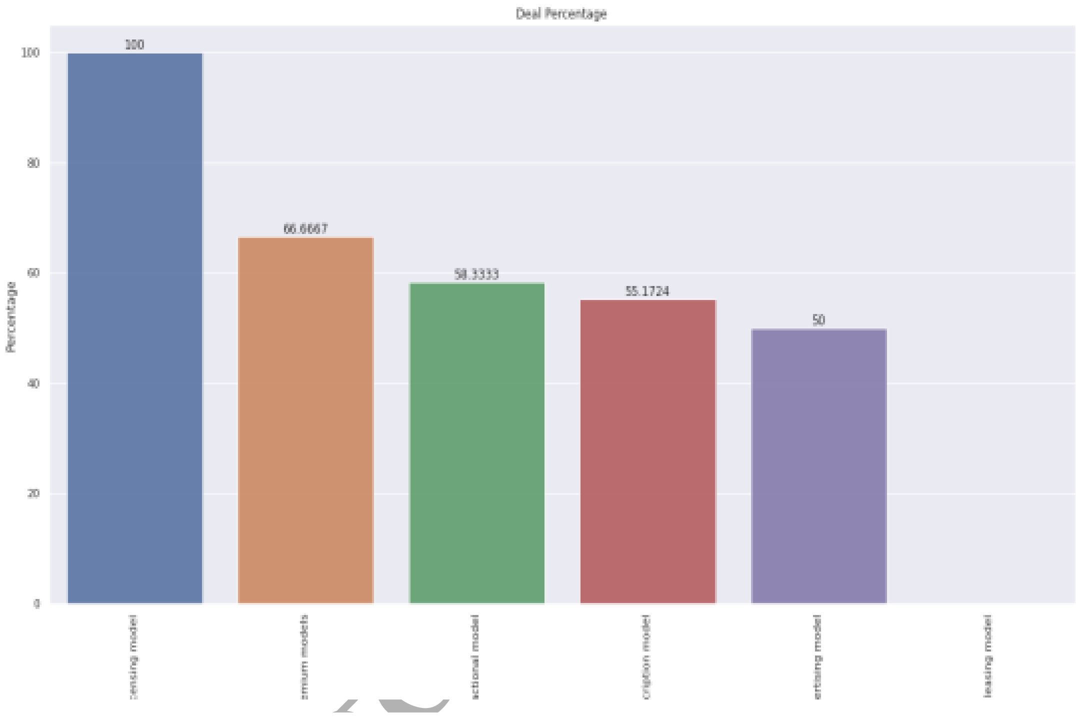

Figure 4.23:Percentage of deal per revenue model

As we can see on the Figure 4.23:Percentage of deal per revenue model, 100% of the teams that use a licensing model, 66,67% of the teams that use a freemium model, 58,33% of the teams that use transactional model, 55,17% of the teams that use subscription model and 50% of the teams that use advertising model were successful in making a deal with the sharks. None of the teams that used rental, or leasing model was successful.

4.4 Statistical Analysis

The statistical analysis of the results will be presented in this part of the thesis

Table 4.1:Variables Statistics

	Number of Presenters	Number of Sales ($) (Last Year)	Amount ($) (ASK)	Equity (%) (ASK)	Valuation
mean	1.54	534,731.04	300,376.37	14.53	13,870,744.93
std	0.61	1,093,770.01	550,566.19	9.75	109,360,491.31
min	1.00	0.00	10,000.00	0.03	40,000.00
0.25	1.00	39,000.00	100,000.00	10.00	500,000.00
0.50	1.00	150,000.00	175,000.00	10.00	1,500,000.00
0.75	2.00	500,000.00	300,000.00	20.00	3,500,000.00
max	5.00	9,000,000.00	8,700,000.00	100.00	2,000,000,000.00
	pronunciation	**number_of_syllables**	**rate_of_speech**	**articulation_rate**	**balance**
mean	98.96	906.72	4.21	4.69	0.92
std	2.11	295.11	0.41	0.46	0.05
min	85.06	284.00	3.00	4.00	0.80
0.25	100.00	671.00	4.00	4.00	0.90
0.50	100.00	879.00	4.00	5.00	0.90
0.75	100.00	1,115.00	4.00	5.00	0.90
max	100.00	1,750.00	5.00	5.00	1.00
	f0_mean	**f0_std**	**f0_median**	**f0_min**	**f0_max**
mean	181.56	46.03	174.43	74.22	409.44
std	41.58	9.63	43.31	4.08	14.89
min	112.73	22.43	104.90	61.00	334.00
0.25	147.39	38.86	138.10	71.00	398.00
0.50	166.88	45.28	158.20	73.00	415.00
0.75	222.59	52.63	216.10	78.00	422.00
max	281.01	84.74	286.70	89.00	430.00

In Table 4.1:Variables Statistics shows the calculations for the mean, standard deviation, minimum value in dataset, 0.25 Quantile, 0.5 Quantile, 0.75 Quantile, and maximum value in dataset. For variable Number of Presenters, is observed that in most pitches there is one entrepreneur and the maximum entrepreneurs there are in a pitch is five. Moreover, for variables Number of Sales, Amount (ASK) and Valuation is observed that there is high deviation within the dataset based on the standard deviation value. Most entrepreneurs offer a small number of their company based on the Equity (ASK) with a few to offer to sell their entire company. Furthermore, on audio analysis variables the pronunciation, rate of speech, articulation rate and balance have low standard deviation which means most values are close to the mean. In column number of syllables there is a higher standard deviation which might be due to the different length of each pitch.

4.4.1 Correlation Test – Pearson's Correlation

In order to test how the different values, correlate to the value "Deal" (yes) the Pearson's Correlation was used. In the diagram below the values that have a strong correlation to the Deal value appear to have a longer column on the diagram and the values with a small correlation appear with a shorter column. The column that represents positive numbers indicate that the values are positively correlated. The columns that represent negative numbers indicate that the values are negatively correlated. The values that have a zero correlation to the Deal value have no statistically proven impact on the deal outcome according to the Pearson's correlation analysis.

According to the Pearson's Correlation Hypothesis 2 is rejected. Negative correlation was found between pronunciation posterior score and the funding outcome.

The findings of the Pearson's Correlation support Hypothesis 3. There is a small positive correlation between anger and the funding outcome.

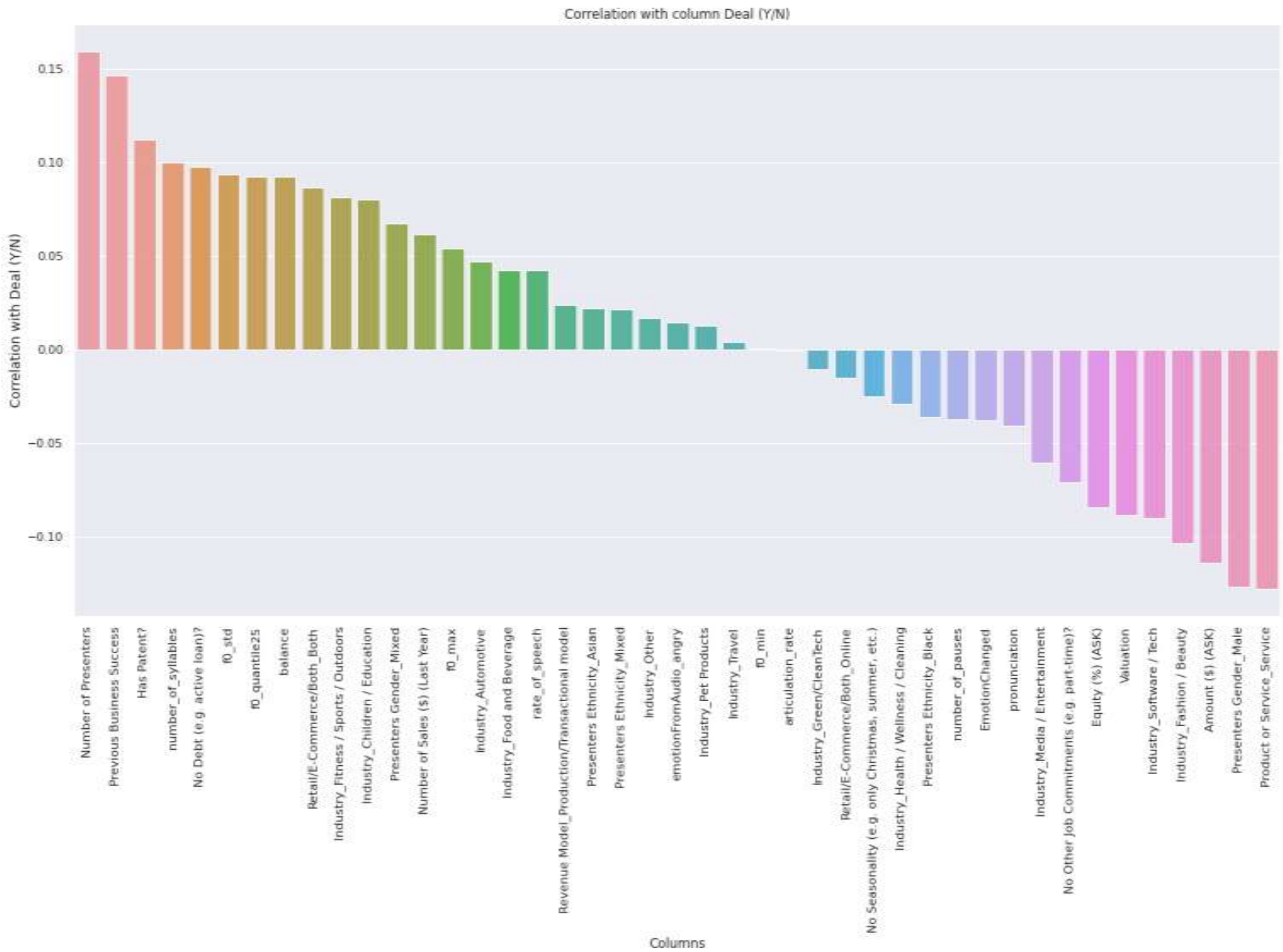

Figure 4.24: Correlation for Variable Deal

4.4.2 Logistic regression

The method of logistic regression was chosen to correlate the relationship between the depended on and the independent variables, in order to find out which of the findings were statistically significant. The logistic regression estimates the probability of an event occurring based on a given dataset of independent variables. Our goal is to determine which of the variables that seem to be affecting the outcome (deal or no deal) are statistically significant.

To perform statistical analysis stats models was used. Stats models is a Python module that provides classes and functions for estimating numerous statistical models, conducting statistical tests, and exploring statistical data. For each estimator, an extensive list of result statistics is available, and the results are tested against existing statistical packages to ensure that they are correct.

```
                           Logit Regression Results
==============================================================================
Dep. Variable:             Deal (Y/N)   No. Observations:                  457
Model:                          Logit   Df Residuals:                      414
Method:                           MLE   Df Model:                           42
Date:                Thu, 28 Jul 2022   Pseudo R-squ.:                  0.1724
Time:                        01:09:41   Log-Likelihood:                -257.31
converged:                       True   LL-Null:                       -310.91
Covariance Type:            nonrobust   LLR p-value:                 1.296e-07
========================================================================================================================
                                                  coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------------------------------------
Number of Presenters                            1.1059      0.267      4.146      0.000       0.583       1.629
Has Patent?                                     0.5659      0.295      1.921      0.055      -0.012       1.143
Previous Business Success                       0.4353      0.247      1.763      0.078      -0.049       0.919
Number of Sales ($) (Last Year)                62.7687     30.466      2.060      0.039       3.056     122.481
Amount ($) (ASK)                               -8.6374      3.619     -2.387      0.017     -15.730      -1.545
Equity (%) (ASK)                               -0.0081      0.013     -0.642      0.521      -0.033       0.017
pronunciation                                  -0.0304      0.040     -0.763      0.446      -0.109       0.048
number_of_syllables                             0.0019      0.001      3.095      0.002       0.001       0.003
number_of_pauses                                0.0030      0.010      0.290      0.772      -0.018       0.024
rate_of_speech                                  0.0106      0.334      0.032      0.975      -0.644       0.666
articulation_rate                              -0.2568      0.264     -0.974      0.330      -0.774       0.260
balance                                         2.8624      3.221      0.889      0.374      -3.450       9.175
f0_std                                         -0.0240      0.020     -1.212      0.226      -0.063       0.015
f0_min                                          0.0319      0.027      1.200      0.230      -0.020       0.084
f0_max                                         -0.0114      0.009     -1.207      0.228      -0.030       0.007
f0_quantile25                                   0.0038      0.006      0.650      0.516      -0.008       0.015
Valuation                                     -10.3360      4.881     -2.118      0.034     -19.902      -0.770
EmotionChanged                                 -0.2347      0.301     -0.781      0.435      -0.824       0.354
No Debt (e.g. active loan)?                     2.4360      0.855      2.848      0.004       0.760       4.112
No Seasonality (e.g. only Christmas, summer, etc.)  -0.4589  0.492   -0.933      0.351      -1.422       0.505
No Other Job Commitments (e.g. part-time)?     -0.1438      0.268     -0.537      0.591      -0.669       0.381
emotionFromAudio_angry                          0.0959      0.254      0.377      0.706      -0.402       0.594
Presenters Ethnicity_Asian                     -0.0912      0.553     -0.165      0.869      -1.174       0.992
Presenters Ethnicity_Black                     -0.0417      0.397     -0.105      0.916      -0.820       0.736
Presenters Ethnicity_Mixed                      0.0236      0.535      0.044      0.965      -1.025       1.072
Industry_Automotive                             1.0694      1.261      0.848      0.397      -1.403       3.542
Industry_Children / Education                   0.3459      0.490      0.706      0.480      -0.614       1.306
Industry_Fashion / Beauty                      -1.1114      0.454     -2.449      0.014      -2.001      -0.222
Industry_Fitness / Sports / Outdoors            0.5314      0.651      0.816      0.414      -0.744       1.807
Industry_Food and Beverage                      0.1847      0.446      0.414      0.679      -0.689       1.059
Industry_Green/CleanTech                       -0.9296      1.557     -0.597      0.551      -3.982       2.123
Industry_Health / Wellness / Cleaning          -0.5504      0.543     -1.014      0.311      -1.615       0.514
Industry_Media / Entertainment                 -0.0544      0.583     -0.093      0.926      -1.198       1.089
Industry_Other                                  0.0352      0.488      0.072      0.943      -0.921       0.992
Industry_Pet Products                          -0.4395      0.661     -0.665      0.506      -1.734       0.855
Industry_Software / Tech                       -0.0753      0.706     -0.107      0.915      -1.460       1.309
Industry_Travel                                 0.0784      1.052      0.074      0.941      -1.984       2.141
Product or Service_Service                     -0.6779      0.413     -1.641      0.101      -1.488       0.132
Presenters Gender_Male                         -1.0362      0.417     -2.485      0.013      -1.853      -0.219
Presenters Gender_Mixed                        -0.6839      0.418     -1.635      0.102      -1.504       0.136
Revenue Model_Production/Transactional model    0.1861      0.455      0.409      0.682      -0.705       1.078
Retail/E-Commerce/Both_Both                     0.3669      0.381      0.962      0.336      -0.380       1.114
Retail/E-Commerce/Both_Online                   0.1210      0.258      0.469      0.639      -0.385       0.627
========================================================================================================================
```

Figure 4.25: Logistic Regression Results

In Figure 4.25: Logistic Regression Results the dependent variable is Deal (Y/N) and the model chosen is logistic regression.

4.4.3 Statistically Significant Parameters:

Number of presenters:

P-value<0.05, 0.000 is less than a=0.05. Therefor the number of presenters effect on the deal outcome is statistically significant. The coefficient value, 1,11 is positive and that indicated that as the number of presenters increases it is more likely that the team will make a deal.

Number of sales:

P-value<0.05,0.039 is less than a=0.05. Therefor the amount of sales effect on the deal outcome is statistically significant. The coefficient value, 62.78 is positive than one and that indicates that as the teams' sales increase, they are more likely to make a deal. The findings about the number of sales supports Hypothesis 1 proving that the characteristics of the company, in this case number of sales is an important factor to the funding outcome.

Amount Ask:

P-value<0.05, 0.017 is less than a=0.05. Therefor the amount the presenters asked affected the deal outcome in a statistically significant way. The coefficient value, -8.63 is less than 0 and that indicated that as the amount the entrepreneurs ask increases, it is less likely that they will have a deal. The findings about the Amount Ask supports Hypothesis 1 proving that the characteristics of the company, in this case Amount Ask is an important factor to the funding outcome.

Number of syllables:

P-value<0.05, 0.002 is less than a=0.05. Therefore, the number of syllables affects the deal outcome in a statistically significant way. The coefficient value, 0.0019 is positive, therefore as the number of syllables increases it is more likely for the team to make a deal.

Valuation:

P-value<0.05, 0.034 is less than a=0.05. Therefor the valuation that the entrepreneurs give to their company affects the deal outcome in a statistically significant way. The coefficient value, -10.33 is less than 0 and that indicates that as the valuation increases it is less likely for the teams to make a deal. The findings about the Valuation supports Hypothesis 1 proving that the characteristics of the company, in this case Valuation is an important factor to the funding outcome.

No Dept:

P-Value<0.05, 0.004 is less than a=0.05. Therefore, the absence of dept effect on the deal outcome is statistically significant. The coefficient value, 2.44 is greater than 0 and that indicates that as we move towards teams without dept it is more likely to make a deal. The findings about the No Dept supports Hypothesis 1 proving that the characteristics of the company, in this case No Dept is an important factor to the funding outcome.

Industry Fashion or Beauty:

P-Value<0.05, 0.014 is less than a=0.05. The variable is statistically significant. The coefficient value, -1.11 is less than 0 therefore the products in fashion and beauty industries seem to have a negative correlation to the deal outcome. The findings about the Industry Fashion or Beauty supports Hypothesis 1 proving that the characteristics of the company, in this case Industry Fashion or Beauty is an important factor to the funding outcome.

Presenter Male:

P-Value<0.05, 0.013 is less than a=0.05. The variable effects the deal outcome in a statistically significant way. The coefficient value, -1.04 indicates that as we move towards male presenters it is less likely to secure a deal.

According to the logistic regression analysis Hypothesis 2 is rejected. No correlation was found between pronunciation posterior score and the funding outcome.

According to the logistic regression analysis Hypothesis 3 is rejected. No correlation was found between emotion and the funding outcome.

4.4.4 Odds Ratios

In Figure 4.26: Significant Variables the Odds rations have been calculated. Odds ratios describe the probability of an outcome occurring in one group compared to the probability of that outcome occurring in another group. A value greater than one (>1) indicates that as the variable tested increases, the event (deal) is more likely to occur. In the same way, when the odds ratio has a value smaller than one (<1), then the event tested is less likely to occur. In case the odds ratio takes a value of 1 (=1) then the variable does not affect the events outcome.

	2.5%	97.5%	Odds Ratio	pvalues	significant?
Number of Sales ($) (Last Year)	2.124905e+01	1.559067e+53	1.820129e+27	0.039372	significant
balance	3.174790e-02	9.650200e+03	1.750353e+01	0.374127	not significant
No Debt (e.g. active loan)?	2.137767e+00	6.107875e+01	1.142682e+01	0.004395	significant
Number of Presenters	1.791714e+00	5.097136e+00	3.022021e+00	0.000034	significant
Industry_Automotive	2.459023e-01	3.452079e+01	2.913544e+00	0.396549	not significant
Has Patent?	9.884487e-01	3.137623e+00	1.761073e+00	0.054790	not significant
Industry_Fitness / Sports / Outdoors	4.750728e-01	6.092616e+00	1.701304e+00	0.414250	not significant
Previous Business Success	9.525374e-01	2.507257e+00	1.545399e+00	0.077898	not significant
Retail/E-Commerce/Both_Both	6.836693e-01	3.046634e+00	1.443222e+00	0.335847	not significant
Industry_Children / Education	5.413146e-01	3.689750e+00	1.413264e+00	0.479904	not significant
Revenue Model_Production/Transactional model	4.938721e-01	2.938017e+00	1.204577e+00	0.682427	not significant
Industry_Food and Beverage	5.018440e-01	2.883191e+00	1.202877e+00	0.678769	not significant
Retail/E-Commerce/Both_Online	6.803750e-01	1.872259e+00	1.128644e+00	0.639331	not significant
emotionFromAudio_angry	6.686740e-01	1.811610e+00	1.100626e+00	0.706106	not significant
Industry_Travel	1.374617e-01	8.509825e+00	1.081561e+00	0.940615	not significant
Industry_Other	3.980097e-01	2.695634e+00	1.035803e+00	0.942534	not significant
f0_min	9.799764e-01	1.087679e+00	1.032424e+00	0.230304	not significant
Presenters Ethnicity_Mixed	3.586967e-01	2.922549e+00	1.023869e+00	0.964841	not significant
rate_of_speech	5.249485e-01	1.945637e+00	1.010623e+00	0.974776	not significant
f0_quantile25	9.923160e-01	1.015481e+00	1.003832e+00	0.515926	not significant
number_of_pauses	9.826220e-01	1.023900e+00	1.003048e+00	0.771849	not significant
number_of_syllables	1.000692e+00	1.003086e+00	1.001888e+00	0.001966	significant
Equity (%) (ASK)	9.675950e-01	1.016829e+00	9.919065e-01	0.520977	not significant
f0_max	9.704227e-01	1.007168e+00	9.886244e-01	0.227554	not significant
f0_std	9.391927e-01	1.014905e+00	9.763154e-01	0.225549	not significant
pronunciation	8.971442e-01	1.048872e+00	9.700462e-01	0.445504	not significant
Presenters Ethnicity_Black	4.405983e-01	2.087940e+00	9.591364e-01	0.916280	not significant
Industry_Media / Entertainment	3.018607e-01	2.971488e+00	9.470879e-01	0.925757	not significant
Industry_Software / Tech	2.322578e-01	3.703274e+00	9.274235e-01	0.915061	not significant
Presenters Ethnicity_Asian	3.090779e-01	2.696237e+00	9.128786e-01	0.868974	not significant
No Other Job Commitments (e.g. part-time)?	5.124446e-01	1.463806e+00	8.660944e-01	0.591335	not significant
EmotionChanged	4.387576e-01	1.425272e+00	7.907901e-01	0.434830	not significant
articulation_rate	4.613260e-01	1.297057e+00	7.735411e-01	0.330213	not significant
Industry_Pet Products	1.765234e-01	2.351942e+00	6.443391e-01	0.505833	not significant
No Seasonality (e.g. only Christmas, summer, etc.)	2.411270e-01	1.656424e+00	6.319879e-01	0.350601	not significant
Industry_Health / Wellness / Cleaning	1.989831e-01	1.671395e+00	5.766970e-01	0.310652	not significant
Product or Service_Service	2.259223e-01	1.140829e+00	5.076798e-01	0.100793	not significant
Presenters Gender_Mixed	2.222769e-01	1.145660e+00	5.046323e-01	0.102069	not significant
Industry_Green/CleanTech	1.865163e-02	8.352920e+00	3.947095e-01	0.550547	not significant
Presenters Gender_Male	1.567066e-01	8.033255e-01	3.548047e-01	0.012948	significant
Industry_Fashion / Beauty	1.352256e-01	8.009818e-01	3.291098e-01	0.014326	significant
Amount ($) (ASK)	1.474029e-07	2.133908e-01	1.773540e-04	0.016996	significant
Valuation	2.272743e-09	4.631538e-01	3.244425e-05	0.034204	significant

Figure 4.26: Significant Variables

4.5 Training Models

The training and evaluation of classifiers was performed using k-fold cross validation. The k-fold cross validation divides the dataset into k numbers of non-overlapping folds. One of the folds is used for test set while the others are used for training. The k number of folds used is five.

Machine Learning Algorithms:

- Logistic Regression
- Support Vector Machines
- Decision Trees
- Random Forest
- Naive Bayes
- K-Nearest Neighbor
- XGB Classifier

To train the models the significand features that were found in statistical analysis were used which are the Number of Presenters, Number of Sales ($) (Last Year), Amount ($) (ASK), Number of Syllables, Valuation, No Debt (e.g. active loan)?, Industry Fashion / Beauty, Presenters Gender Male. Also, the variables Has Patent?, No Other Job Commitments (e.g. part-time)?, No Debt (e.g. active loan)?, Emotion From Audio Angry, Industry Automotive, Industry Fitness / Sports / Outdoors, Emotion From Audio Sad were used because of positive correlation with the funding outcome.

4.6 Performance Evaluation

To evaluate the models the metrics accuracy, precision, recall and f1 score is used.

4.6.1 Mean Accuracy

Accuracy is the number of correct predictions to the total number of predictions made.

$$Accuracy = \frac{NumberOfCorrectPredictions}{NumberOfTotalPredictions}$$

In Figure 4.27: Mean Accuracy Score the best performing models are Logistic Regression with 68% accuracy and Support Vector Machines with 67% accuracy.

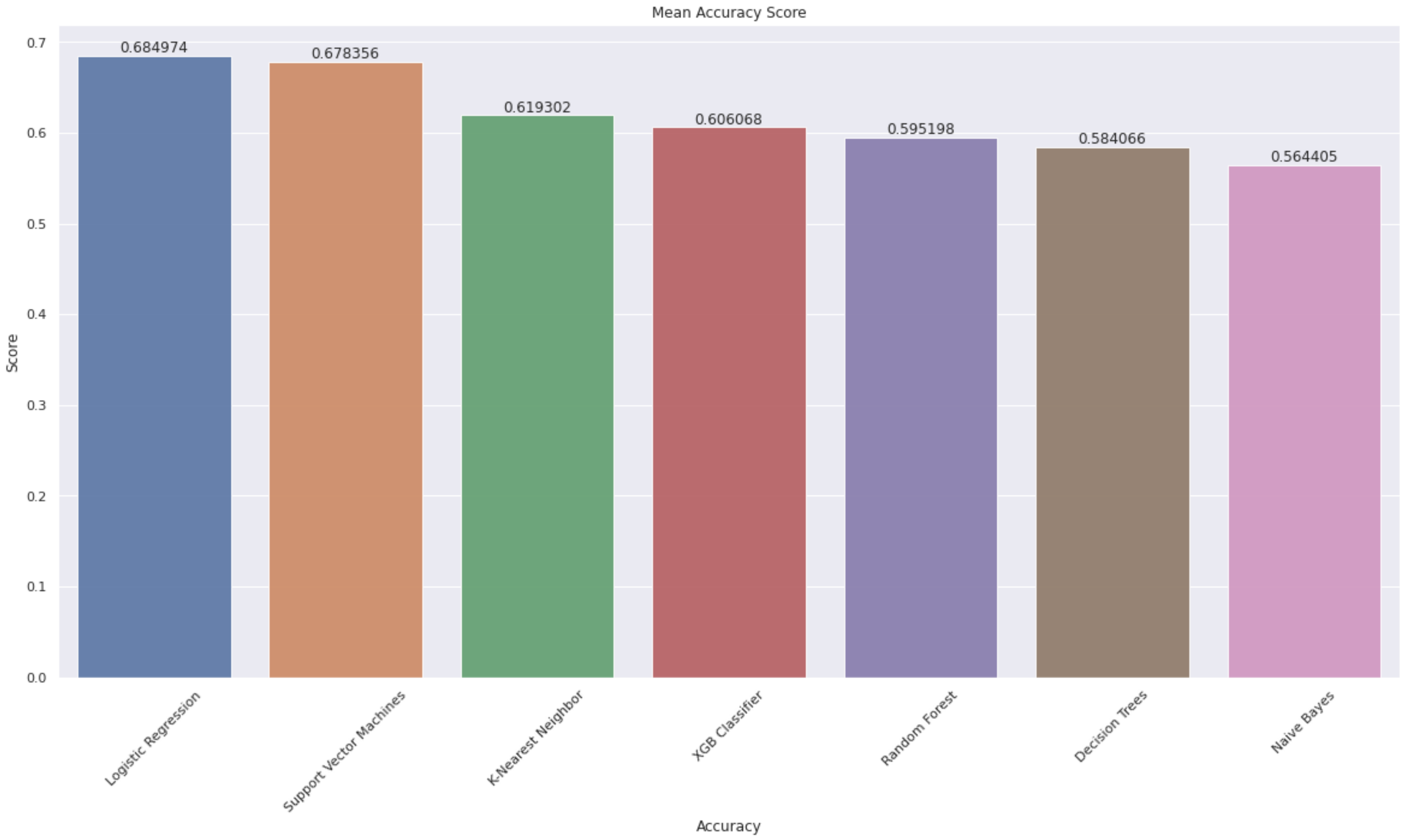

Figure 4.27: Mean Accuracy Score

4.6.2 Mean Precision

Precision is the ratio of correctly classified positive samples to a total number of classified positive samples.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives}$$

In Figure 4.28: Mean Precision Score the models with the highest mean precision are Logistic Regression with 69% precision and Support Vector Machines with 69% precision.

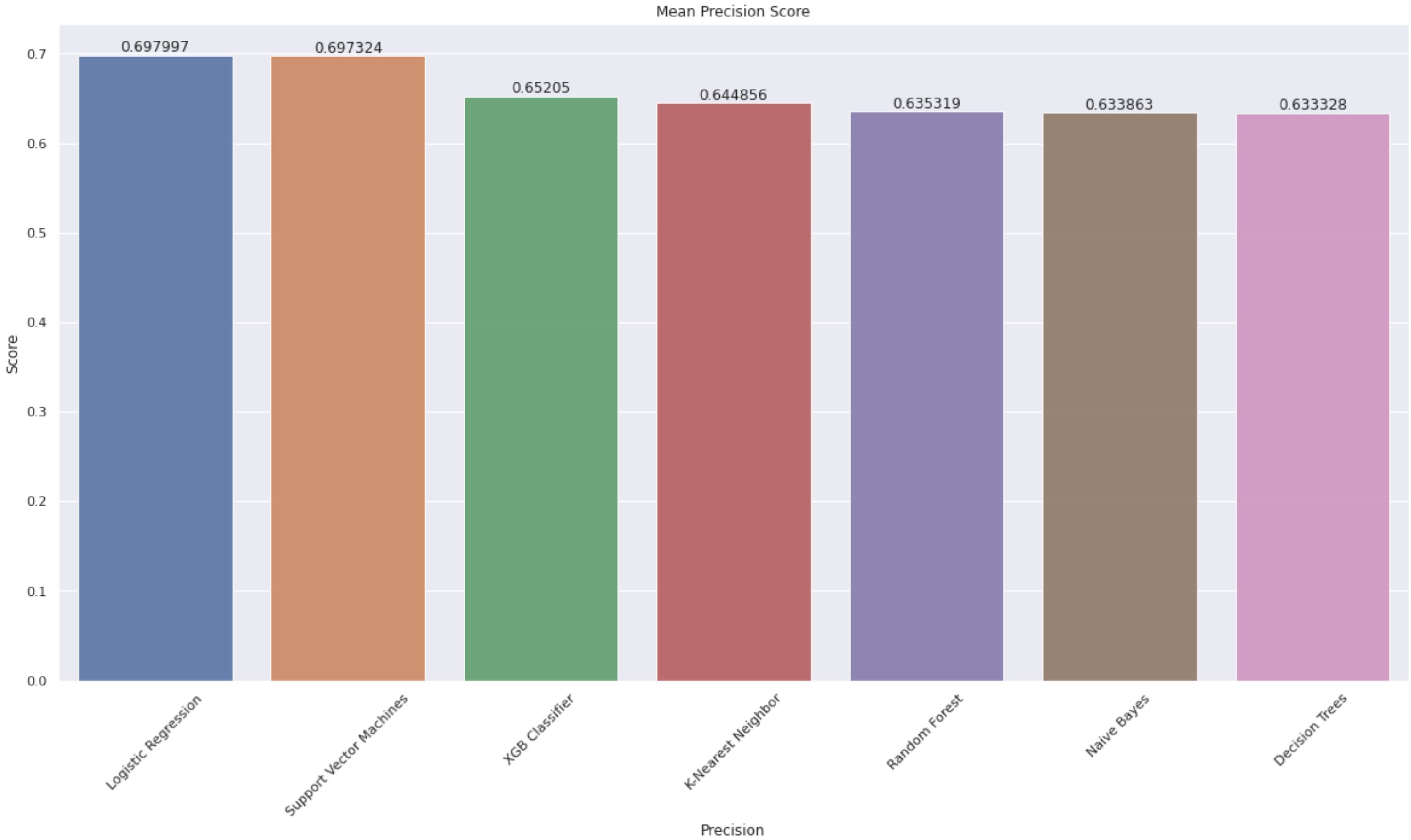

Figure 4.28: Mean Precision Score

4.6.3 Mean Recall

Recall is the ratio of positive samples correctly classified as positive to the total number of positive samples.

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives}$$

In Figure 4.29: Mean Recall Score the models with the highest mean precision are Logistic Regression with 80% precision and Support Vector Machines with 78% precision.

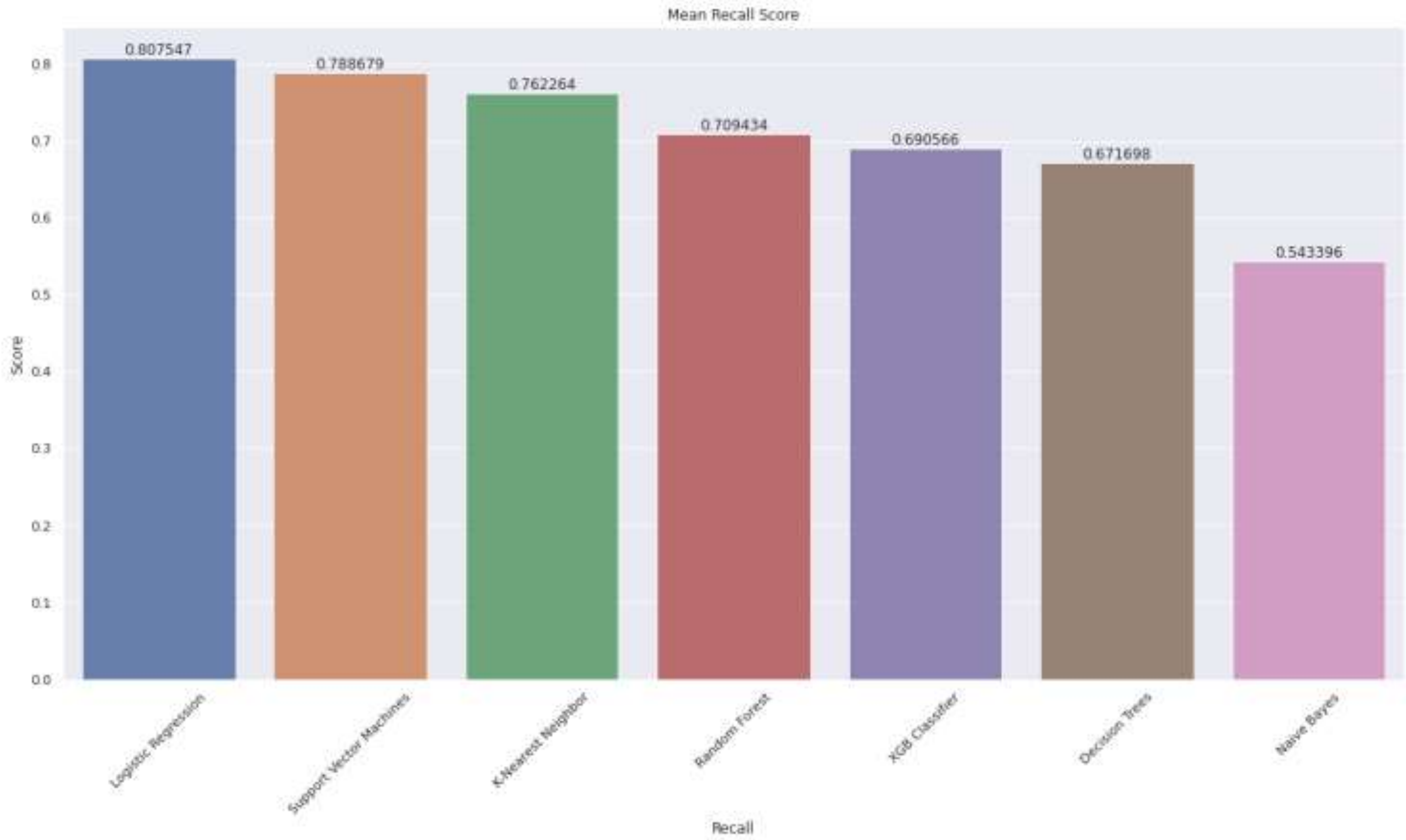

Figure 4.29: Mean Recall Score

4.6.4 Mean F1 Score

F1 Score combines the Precision and Recall metrics to one.

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

In Figure 4.30: Mean F1 Score the models with the highest score are the Logistic Regression with 74% score and Support Vector Machines with 73%.

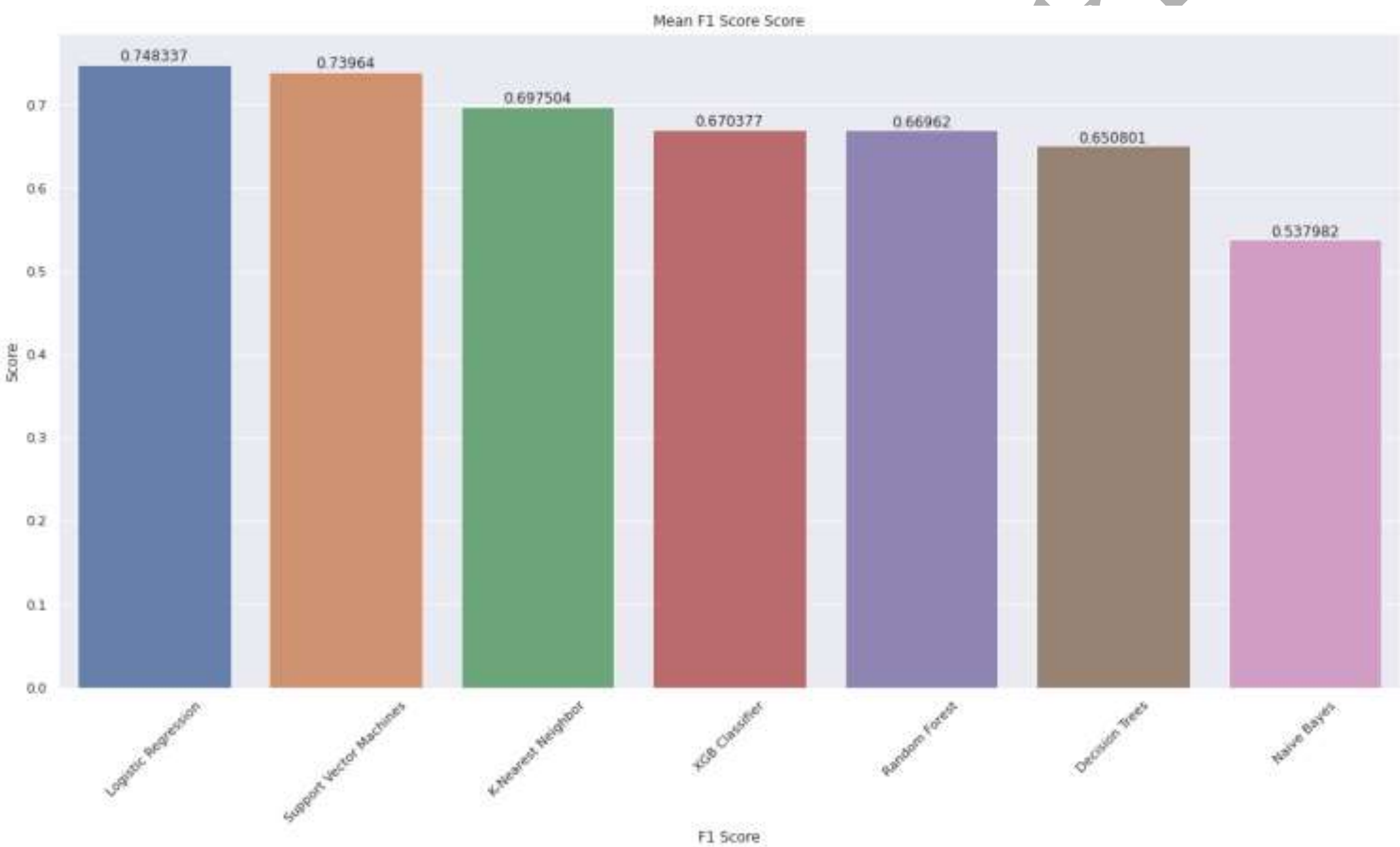

Figure 4.30: Mean F1 Score

4.6.5 Evaluation Metrics

Table 4.2: Performance Evaluation shows all the performance evaluations scores. The colour green marks the highest score for each evaluation. Logistic regression has the best scores in all our tests and a close second is the Support Vector Machines model.

Table 4.2: Performance Evaluation

	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.68	0.6979	0.80	0.74
Support Vector Machines	0.67	0.6973	0.78	0.73
Decision Trees	0.58	0.6333	0.67	0.65
Random Forest	0.59	0.6353	0.70	0.66
Naive Bayes	0.56	0.6338	0.54	0.53
K-Nearest Neighbor	0.61	0.6448	0.76	0.69
XGB Classifier	0.60	0.6520	0.69	0.67

Chapter 5

CONCLUSIONS

It appears that having a debt or an active loan has a negative impact on the insurance of a deal as 58,74% of the teams that were free of debts or loans were successful but the percentage of success for the teams with active loan or debt that secured a deal was only 27,27%.

Regarding the expression of emotions during their funding pitch it seems that expressing sadness gave the entrepreneurs a significant precedence (100% success). It is important to underline that expressing any emotion appeared to have a positive impact on the result of the funding pitch (64,13% of the teams that expressed happy emotions made a deal, 56,73% of the teams that displayed anger also were successful, and 50% of the teams that expressed disgusts succeeded to make a deal). It is equally important to state that according to our findings remaining neutral had a negative impact on the deal making once none of the teams that remained neutral managed to make a deal.

The number of the team members seem to positively correlate with the chance of making a deal with the Sharks. To simplify, as the number of team members increased, the percentage of successful deals increased as well (one presenter-50,85% success, two presenters-64,5% success, three presenters-73,68%, 4 or 5 presenters-100%). These results need further investigation because the teams that had 4 or 5 members were only one of each.

The industry of the product or service that each team presented did not appear to have significant fluctuations regarding the deal making. Most of the industry categories had a success percentage around 50-70% (74,07% of the teams in sports or outdoors, 67.69% - children or education industry, 62,35% - food and beverage industry, 61,11% - pet products industry, 60% - travel industry, 59,52% - lifestyle or home industry, 52,94% - cleaning industry, 50% - cleantech industry, 46,84% - fashion and beauty industry, 46,67% - entertainment industry). The most successful industry category appeared to be the automotive industry (80% deal) and the least successful were teams in the software or tech industry (35,29%).

Another factor that was tested through this project was the correlation between having other commitments or another job that had an impact on the deal making. According to the results a higher percentage of people that had other commitments managed to make a deal (63,71%) and less of the teams that had no other commitments were able to succeed. At a first glance these results could seem odd. On the other hand, the factors that conclude to a decision are so complex and multiple that the results require more research. In addition, this outcome can be the result of other factors such as the fact that maybe the people that had no other responsibilities were young and at the beginning of their entrepreneur careers and therefore, they lack experience in pursuing and communicating through a funding pitch.

The research of potential influence of race on the final deal results showed that there was no significant difference among races (63,16% of Asian teams, 62,5% of mixed teams, 58,10% of white teams and 52,38% of black teams managed to make a deal). Although the success percentage of black teams seems lower comparatively to other categories, this could have happened because of the lower number of teams with only black entrepreneurs (9,19%). On the other hand, it must be addressed that Asian entrepreneurs who appear to have the best percentage of success were also represented by an even smaller number of teams (4,16%) and the mixed teams that came second in the success list with a percentage of 62.5% were also represented by 5,25% of the teams.

Another factor that this thesis examined was the percentage of the successful teams that sold through retail in comparison to the teams that sold online and those who combined both retail and online sales. Viewing the data gives a clear image that using both paths of selling was in fact a deal maker. The teams that sold both through retail and online had a higher percentage of success (68.85%) compared to online sales (57.07%) and retail sales (55,66%). The investors on Shark tank seemed to be more excited about online sales because selling direct to consumers can benefit the company by being able to take a larger margin (no middle sellers) and deliver faster to the consumer. On the other hand, the Sharks stated that through retail a company can expand their market share. This means that retail could help find more customers. These facts explain why the sharks selected to invest in companies that used both online sales and sold their products through retail.

Continuing with the results of this project the teams that appeared to sell a product had a higher percentage of success (60,30%) in comparison to the teams that provide a service (40,74%). The reason why there is a difference between product and service can be the fact that these types of selling goods are inevitably linked to different revenue models which could be the problem hiding behind services. In addition, it is a fact that services require more time, effort and money to be provided online or the require a physical store, place, office etc. that also requires prior funding. Therefore, it seems that products are more attractive to investors.

Regarding the products' seasonality it appeared that seasonal products had a higher percentage of success (62,96%) than non-seasonal products (57,67%) but this result can be since a small number of seasonal products were presented and could change if that number increased.

The gender of the contestants seemed to have a small impact on the success of the team once 65,22% of female teams, 65.06% of mixed teams and 52.51% of men teams managed to make a deal. The only men's team appears to have less success percentage than the other two teams.

Viewing the results on percentage of deal per revenue model it firstly appears that having a licensing model gives the company a very good chance to make a deal (100% success), but this cannot be a reliable fact because the sample was only 2 out of 457 teams that used this specific revenue model. The same happens for the freemium model teams (3 teams out of 457) which seem to have a very high percentage of success and teams that use advertising model (1 team) had a 50% chance of success, but the cases are also very few and this fact forbids us from jumping to conclusions. In addition, none of the teams that used rental, or leasing model was successful. This might be due to the very small number of teams that used these models (just 1 team used the leasing model and none of the teams used the rental model). The most teams used a transactional model (420/457, 91.90%) and had a very good percentage of success at 58.33%. Also, the 29 teams (6,35% of the teams' total) had a 55,17% deal percentage.

Based on the findings of this thesis statistical analysis only the Number of Syllables from the audio analysis features is significand and has a positive affect to securing funding.

Examining how entrepreneur emotions during funding pitch affect the deal outcome, it was found that expressing anger or sadness appears to affect the deal.

The machine learning models with the best results overall to predict if an entrepreneur will secure funding were Logistic Regression and Support Vector Machines with F1 Score 74% and 73% respectively. The variables used to train the models are the Number of Presenters, Has Patent?, Number of Sales ($) (Last Year), Amount ($) (ASK), Number of Syllables, Valuation, No Debt (e.g. active loan)?, Industry Fashion / Beauty, Presenters Gender Male, No Other Job Commitments (e.g. part-time)?, Emotion From Audio Angry, Industry Automotive, Industry Fitness / Sports / Outdoors, Emotion From Audio Sad.

Bibliography

[1] D. Keltner, D. Sauter, J. Tracy, and A. Cowen, ‘Emotional Expression: Advances in Basic Emotion Theory’, *J. Nonverbal Behav.*, vol. 43, no. 2, pp. 133–160, Jun. 2019, doi: 10.1007/s10919-019-00293-3.

[2] J. E. Stellar *et al.*, ‘Self-Transcendent Emotions and Their Social Functions: Compassion, Gratitude, and Awe Bind Us to Others Through Prosociality’, *Emot. Rev.*, vol. 9, no. 3, pp. 200–207, Jul. 2017, doi: 10.1177/1754073916684557.

[3] G. A. van Kleef, *The Interpersonal Dynamics of Emotion*. Cambridge University Press, 2016.

[4] M. Statman, ‘What Investors *Really* Want’, *Financ. Anal. J.*, vol. 66, no. 2, pp. 8–10, Mar. 2010, doi: 10.2469/faj.v66.n2.5.

[5] ‘Blood in the water: An abductive approach to startup valuation on ABC’s Shark Tank | Elsevier Enhanced Reader’. https://reader.elsevier.com/reader/sd/pii/S2352673422000038?token=D2A844D1CE0FD151B4583CD02E77D92EAD8219F30B2EC0CC55563712ADCE93E756C929977CE2D9F05D9C915190851E36&originRegion=eu-west-1&originCreation=20220907072622 (accessed Sep. 07, 2022).

[6] T. Sherk, M.-T. Tran, and T. V. Nguyen, ‘SharkTank Deal Prediction: Dataset and Computational Model’, in *2019 11th International Conference on Knowledge and Systems Engineering (KSE)*, Oct. 2019, pp. 1–6. doi: 10.1109/KSE.2019.8919477.

[7] N. Ahmed S. Al-Ghamdi and R. Alghofaily, ‘Cross-Cultural Linguistic Analysis of Persuasive Techniques in Shark Tank’, *Int. J. Engl. Lang. Educ.*, vol. 7, no. 2, p. 82, Sep. 2019, doi: 10.5296/ijele.v7i2.15416.

[8] B. Smith and A. Viceisza, ‘Bite me! ABC’s Shark Tank as a path to entrepreneurship’, *Small Bus. Econ.*, vol. 50, no. 3, pp. 463–479, Mar. 2018, doi: 10.1007/s11187-017-9880-8.

[9] X. (Shane) Wang, S. Lu, X. I. Li, M. Khamitov, and N. Bendle, ‘Audio Mining: The Role of Vocal Tone in Persuasion’, *J. Consum. Res.*, vol. 48, no. 2, pp. 189–211, Aug. 2021, doi: 10.1093/jcr/ucab012.

[10] T. H. Allison, B. J. Warnick, B. C. Davis, and M. S. Cardon, ‘Can you hear me now? Engendering passion and preparedness perceptions with vocal expressions in crowdfunding pitches’, *J. Bus. Ventur.*, vol. 37, no. 3, p. 106193, May 2022, doi: 10.1016/j.jbusvent.2022.106193.

[11] C. Clark, ‘The impact of entrepreneurs’ oral “pitch” presentation skills on business angels’ initial screening investment decisions’, *Venture Cap.*, vol. 10, no. 3, pp. 257–279, Jul. 2008, doi: 10.1080/13691060802151945.

[12] L. Huang and J. L. Pearce, ‘Managing the Unknowable: The Effectiveness of Early-stage Investor Gut Feel in Entrepreneurial Investment Decisions’, *Adm. Sci. Q.*, vol. 60, no. 4, pp. 634–670, Dec. 2015, doi: 10.1177/0001839215597270.

[13] L. Feeney, G. H. Haines, and A. L. Riding, ‘Private investors’ investment criteria: Insights from qualitative data’, *Venture Cap.*, vol. 1, no. 2, pp. 121–145, Apr. 1999, doi: 10.1080/136910699295938.

[14] C. Mason and M. Stark, ‘What do Investors Look for in a Business Plan?: A Comparison of the Investment Criteria of Bankers, Venture Capitalists and Business Angels’, *Int. Small Bus. J.*, vol. 22, no. 3, pp. 227–248, Jun. 2004, doi: 10.1177/0266242604042377.

[15] F. Ferrati and M. Muffatto, ‘Reviewing equity investors’ funding criteria: a comprehensive classification and research agenda’, *Venture Cap.*, vol. 23, no. 2, pp. 157–178, Apr. 2021, doi: 10.1080/13691066.2021.1883211.
[16] K. V. Lewis, M. Ho, C. Harris, and R. Morrison, ‘Becoming an entrepreneur: opportunities and identity transitions’, *Int. J. Gend. Entrep.*, vol. 8, no. 2, pp. 98–116, Jan. 2016, doi: 10.1108/IJGE-02-2015-0006.
[17] E. Bulska, ‘Good oral presentation of scientific work’, *Anal. Bioanal. Chem.*, vol. 385, no. 3, pp. 403–405, Jun. 2006, doi: 10.1007/s00216-006-0441-2.
[18] A. Hu and S. Ma, ‘Human Interactions and Financial Investment: A Video-Based Approach’, *SSRN Electron. J.*, 2020, doi: 10.2139/ssrn.3583898.
[19] R. van Werven, O. Bouwmeester, and J. P. Cornelissen, ‘Pitching a business idea to investors: How new venture founders use micro-level rhetoric to achieve narrative plausibility and resonance’, *Int. Small Bus. J.*, vol. 37, no. 3, pp. 193–214, May 2019, doi: 10.1177/0266242618818249.
[20] A. Pentland, *Honest Signals: How They Shape Our World*. MIT Press, 2010.
[21] R. W. Griffin, *Management*. Cengage Learning, 2012.
[22] J. H. Marshel *et al.*, ‘Cortical layer–specific critical dynamics triggering perception’, *Science*, vol. 365, no. 6453, p. eaaw5202, Aug. 2019, doi: 10.1126/science.aaw5202.
[23] M. S. Cardon, M. Glauser, and C. Y. Murnieks, ‘Passion for what? Expanding the domains of entrepreneurial passion’, *J. Bus. Ventur. Insights*, vol. 8, pp. 24–32, Nov. 2017, doi: 10.1016/j.jbvi.2017.05.004.
[24] B. J. Warnick, B. C. Davis, T. H. Allison, and A. H. Anglin, ‘Express yourself: Facial expression of happiness, anger, fear, and sadness in funding pitches’, *J. Bus. Ventur.*, vol. 36, no. 4, p. 106109, Jul. 2021, doi: 10.1016/j.jbusvent.2021.106109.
[25] L. Jiang, D. Yin, D. Liu, and R. Johnson, ‘The More Enthusiastic, the Better? Unveiling a Negative Pathway From Entrepreneurs’ Displayed Enthusiasm to Funders’ Funding Intentions’, *Entrep. Theory Pract.*, p. 10422587221076392, May 2022, doi: 10.1177/10422587221076391.
[26] S. Hubner, M. Baum, and M. Frese, ‘Contagion of Entrepreneurial Passion: Effects on Employee Outcomes’, *Entrep. Theory Pract.*, vol. 44, no. 6, pp. 1112–1140, Nov. 2020, doi: 10.1177/1042258719883995.
[27] G. A. Van Kleef, A. C. Homan, and A. Cheshin, ‘Emotional influence at work: Take it EASI’, *Organ. Psychol. Rev.*, vol. 2, no. 4, pp. 311–339, Nov. 2012, doi: 10.1177/2041386612454911.
[28] D. O. Olguín, ‘Assessing Group Performance from Collective Behavior’, p. 6.
[29] M. A. Hitt, R. D. Ireland, S. M. Camp, and D. L. Sexton, ‘Strategic entrepreneurship: entrepreneurial strategies for wealth creation’, *Strateg. Manag. J.*, vol. 22, no. 6–7, pp. 479–491, 2001, doi: 10.1002/smj.196.
[30] D. L. Sexton and N. Bowman, ‘The entrepreneur: A capable executive and more’, *J. Bus. Ventur.*, vol. 1, no. 1, pp. 129–140, Dec. 1985, doi: 10.1016/0883-9026(85)90012-6.
[31] R. A. Baron and J. Tang, ‘The role of entrepreneurs in firm-level innovation: Joint effects of positive affect, creativity, and environmental dynamism’, *J. Bus. Ventur.*, vol. 26, no. 1, pp. 49–60, Jan. 2011, doi: 10.1016/j.jbusvent.2009.06.002.
[32] ‘Human voice perception | Elsevier Enhanced Reader’. https://reader.elsevier.com/reader/sd/pii/S096098221001701X?token=E63DA863B0E3857BBE695E211828AAD0692D9E3AF36CA95C8AD61FB6C1E5D55FA9

B861F9EBD13B06A10952C96BF9F796&originRegion=eu-west-1&originCreation=20220721094619 (accessed Jul. 21, 2022).

[33] R. K. Mitchell, L. Busenitz, T. Lant, P. P. McDougall, E. A. Morse, and J. B. Smith, 'Toward a Theory of Entrepreneurial Cognition: Rethinking the People Side of Entrepreneurship Research', *Entrep. Theory Pract.*, vol. 27, no. 2, pp. 93–104, Apr. 2002, doi: 10.1111/1540-8520.00001.

[34] J. J. Skelly and J. B. Johnson, 'GLASS CEILINGS AND GREAT EXPECTATIONS: GENDER STEREOTYPE IMPACT ON FEMALE PROFESSIONALS', p. 12.

[35] C. Brush, P. Greene, L. Balachandra, and A. Davis, 'The gender gap in venture capital- progress, problems, and perspectives', *Venture Cap.*, vol. 20, no. 2, pp. 115–136, Apr. 2018, doi: 10.1080/13691066.2017.1349266.

[36] M. Brännback and A. L. Carsrud, Eds., *Revisiting the Entrepreneurial Mind: Inside the Black Box: An Expanded Edition*, vol. 35. Cham: Springer International Publishing, 2017. doi: 10.1007/978-3-319-45544-0.

[37] M. Weirich and A. P. Simpson, 'Gender identity is indexed and perceived in speech', p. 18.

[38] 'User Guide — MoviePy 1.0.2 documentation'. https://zulko.github.io/moviepy/ (accessed Jul. 24, 2022).

[39] R. Hennequin, A. Khlif, F. Voituret, and M. Moussallam, 'Spleeter: a fast and efficient music source separation tool with pre-trained models', *J. Open Source Softw.*, vol. 5, no. 50, p. 2154, Jun. 2020, doi: 10.21105/joss.02154.

[40] H. Bredin *et al.*, 'Pyannote.Audio: Neural Building Blocks for Speaker Diarization', in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2020, pp. 7124–7128. doi: 10.1109/ICASSP40776.2020.9052974.

[41] H. Bredin and A. Laurent, 'End-to-end speaker segmentation for overlap-aware resegmentation', 2021.

[42] M. G. de Pinto, M. Polignano, P. Lops, and G. Semeraro, 'Emotions Understanding Model from Spoken Language using Deep Neural Networks and Mel-Frequency Cepstral Coefficients', in *2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, May 2020, pp. 1–5. doi: 10.1109/EAIS48028.2020.9122698.

[43] S. R. Livingstone and F. A. Russo, 'The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English', *PLOS ONE*, vol. 13, no. 5, p. e0196391, May 2018, doi: 10.1371/journal.pone.0196391.

[44] L. Balachandra, T. Briggs, K. Eddleston, and C. Brush, 'Don't Pitch Like a Girl!: How Gender Stereotypes Influence Investor Decisions', *Entrep. Theory Pract.*, vol. 43, no. 1, pp. 116–137, Jan. 2019, doi: 10.1177/1042258717728028.