

ELEVATE LABS DATA ANALYSIS PROJECT

- **ABSTRACT**

The Startup Investment Analysis of the Shark Tank Data (USA) is an analysis project done on a Shark Tank dataset which holds the data of the individual sharks, their contributions, the prospective entrepreneurs and their respective valuation amounts. This data was collected over 16 seasons of the reality TV show. By analyzing the datasets, trends can be extrapolated which would give insights on the pattern of the show. By critical analysis and visualizations, recommendations and solutions can be offered which would increase the likelihood of entrepreneurs gaining higher number of investments in the future.

- **INTRODUCTION**

The Shark Tank USA is an ABC reality television series which brings real-world businessmen and women face-to-face with prospective entrepreneurs who desire their investments. The prospectors range from a variety of sectors such as healthcare, green tech and food industries etc. The dataset provided is a record of 16 seasons which highlight the number of years the show has been running. Amongst the sharks on Shark Tank USA, includes Mark Cuban who currently serves as the minority owner of the Dallas Mavericks of the National Basketball Association. The dataset also contains the demographics of the entrepreneurs along with their respective business descriptions. Summarily, the objective of this analysis is to provide a seamless overview of all the key metrics involved in the Shark Tank series.

- **TOOLS USED**

Microsoft Excel 2021 was the tool used to complete this project. No other tool was used as Excel provided adequate data cleaning, transformation and visualization options. The Power Query Editor was used to transform the original dataset as it was unstructured and Pivot Charts were now used to visualize the secondary KPIs on a dashboard.

- **PROJECT BUILDING STEPS**

Step 1: Downloaded the ZIP folder of the Shark Tank US dataset from Kaggle online.

Step 2: Extracted the Shark Tank US dataset into a folder and then opened it on Excel 2021.

Step 3: Highlighted the entire dataset with the CTRL + A shortcut and then proceeded to the Data Tab where I selected the From Table/Range command on the Get & Transform Data Group. This brought up the Power Query Editor where I could proceed with data cleaning.

Step 4: On the Power Query Editor, I removed the Company Email Address, Royalty Deal, Advisory Shares Equity, Loan, Deal Has Conditions and the Investment Equity columns of all six sharks. This process was carried out as the above-mentioned columns were not needed for my analysis.

Step 5: Then I proceeded to change the data types of some columns. For example, for columns that contained binary values (0 and 1) to represent Yes and No, I changed their data types to Boolean (True & False). This made it easier for me to visualize any KPI as a pie chart. The columns that contained amounts were also changed from Whole Number to Currency while the Equity columns remained as whole numbers as equity is not a monetary value but a number.

Step 6: I changed the values of columns that contained null to 0. This was because the columns that contained null values were all numerical columns. 0 in this case would mean that that particular shark did not invest in a certain pitch and hence contributed no investment amount or investment equity. Other columns such as Pitchers Average Age that did not contained logical values were changed to N/A from null.

Step 7: I also added columns such as No of Entrepreneurs which was a conditional column based on the Entrepreneur Names column. Basically, if a name in the Entrepreneur Names column contained 'and', the output will be Multiple but if no record in the Entrepreneur Names column did not have 'and', the output will be Single. I did this so as to visualize the investment amount based on the number of pitchers on a team.

Step 8: I also reordered some columns so as to make the cleaned dataset appear better.

Step 9: I closed and loaded the cleaned dataset from my Power Query Editor back into Excel.

Step 10: I then proceeded to create Pivot Tables and Pivot Charts answering the questions stipulated for this task in the pdf. These tables and charts are located on individual worksheets in the Shark Tank USA Analysis workbook.

- **CONCLUSION**

The analysis provides a comprehensive overview of the Shark Tank USA reality show. Despite the analysis, there are still some drawbacks due to the inadequate information given by the dataset which are impossible to determine such as the average age of the pitchers. However, this issue can be mitigated with better data collection techniques. The dashboard created successfully reflected the investment behavior of the sharks along with the yearly trend of investments.