EDA PROJECT

INT – 353

CONTINUOUS ASSESSMENT – 3

On

SHARK TANK INDIA DATASET ANALYSIS

By

Aman Vishwakarma

12113404

Submitted to

Anjana Maam

Introduction

This data set was selected to work on the “Shark Tank India data. I have taken the data from “Kaggle.com” to create this dataset. In this dataset the data consists from20 December 2021, to 4 February 2022.

"Shark Tank" is a popular international television show format that features aspiring entrepreneurs pitching their business ideas to a panel of wealthy investors in the hope of securing investment deals.

Key Concept-

“Shark Tank India” features a panel of potential investors, termed as "Sharks", who listen to entrepreneurs' pitch ideas for a business or product they wish to develop. These self-made multi-millionaires judge the business concepts and products pitched and then decide whether to invest their money to help market and mentor each contestant.

Terminology-

Due Diligence	This refers to the general responsibility of a business to audit and investigate details for the matter at hand. While this could be in response to a rising issue, performing due diligence is common practice before a transaction with an outside party.

Franchise **Valuation**	This refers to the general responsibility of a business to audit and investigate details for the matter at hand. While this could be in response to a rising issue, performing due diligence is common practice before a transaction with an outside party. Valuation is the calculative process to assess a company’s/asset’s worth. To do this, an analyst will take a look at the way the business is composed and managed.

Franchise **Valuation**	This refers to the general responsibility of a business to audit and investigate details for the matter at hand. While this could be in response to a rising issue, performing due diligence is common practice before a transaction with an outside party. Valuation is the calculative process to assess a company’s/asset’s worth. To do this, an analyst will take a look at the way the business is composed and managed.

Cash Flow	Cash flow refers to the net amount of monetary capital transferred to and from a business. **Inflow** refers to the cash businesses receive while **outflow** refers to the cash businesses spend.
Gross	This term refers to the total that is made before calculating any **gross deductions**. These deductions include tax liability, paychecks, union dues, retirement plans, etc.
Net	“Net” is also known as “net income” or “net earnings”. Either way, it’s determined as the sales amount minus the cost of operating expenses, interest, depreciation, etc.
Margin	Margin refers to the amount of cash a broker borrows in order to buy an investment. Mathematically, it’s determined by the amount of the total value of investment minus the loan amount.

Capital	Capital is a very general term that doesn't only apply to finances, at least not in a direct sense. "Capital" can refer to anything that offers value and can point to both tangible and intangible capital.
Return	Also known as a financial return, a "return" refers to the cash made/lost on an investment over time. Percentage changes or price changes usually represent these value shifts.

Data Understanding-

Data Understanding-

Episode number	Number of the episode
Pitch number	Number of the Pitch
Brandname Idea	Name of the brand Idea behind the brand building

Deal	Deal done or not ; 1 - YES, 0 - NO
Pitcheraskamount	Amount asked by the Pitchers
Ask equity	Equity offered by the pitchers
Askvaluation	Valuation asked by pitchers
Dealamount	Final Deal Amount
Deal equity	Final Deal equity percentage
Shark Name	Name of the Sharks in the show

Reason-

Reason-

1. Entrepreneurial Strategies: You can examine the various strategies employed by entrepreneurs to persuade the sharks to invest in their businesses. This can include pitch tactics, valuation methods, and negotiation skills.

2. Predictive Analytics: If you have access to data on which businesses succeeded or failed post-investment, you could explore predictive analytics to identify factors that contribute to a business's long-term success or failure.

3. Diverse Entrepreneurship: The show features a wide range of entrepreneurs from various industries, including technology, fashion, food, and more. This diversity can provide a rich dataset with a broad scope.

4. Real-world Business Cases: "Shark Tank" presents real entrepreneurs seeking real investments from experienced investors. Analysing the show's data can provide insights into the challenges, successes, and failures that entrepreneurs face when pitching their businesses.

5. Investment and Valuation Data: The heart of "Shark Tank" is the negotiation between entrepreneurs and the "sharks" (investors) over equity and valuation. Analysing these negotiations can provide valuable insights into investment trends and business valuation practices.

6. Education and Learning: Using "Shark Tank" data for educational purposes can be highly engaging for students studying entrepreneurship, finance, marketing, or business strategy. It can serve as a real-world case study.

7. Entertainment Value: Aside from its educational and analytical potential, "Shark Tank" data can simply be fun to work with due to its entertainment value. It allows you to dive into the world of innovative products, charismatic entrepreneurs, and sometimes dramatic negotiations.

Questions for Analysis-

A dataset related to "Shark Tank" can yield a wide range of interesting and insightful queries. Here are some queries and questions you can explore using this dataset:

- **1.** Success Rate Analysis:

- What is the overall success rate of entrepreneurs securing investments from the sharks?

- How does the success rate vary by industry or product category?

- Do entrepreneurs with specific backgrounds or demographics have a higher success rate?

- **2**.Investment Patterns:

- Which shark invests the most frequently, and in which types of businesses?

- Are there any trends in the types of deals (equity percentage, valuation) that the sharks prefer?

- Do sharks collaborate on deals, and if so, under what circumstances?

- **3**.Valuation Analysis:

- What is the average valuation of businesses that appear on "Shark Tank"?

- How do entrepreneurs' initial valuations compare to the final valuations after negotiations?

- Are there any correlations between valuation and post-investment success?

- **4**.Pitch Performance:

- How do entrepreneurs' pitch quality, charisma, and presentation skills affect their chances of securing an investment?

- Is there a correlation between the length of a pitch and its success?

- Which types of pitches are more likely to result in offers from the sharks?

- **5**.Post-Investment Outcomes:

- What is the long-term success rate of businesses that secure investments on the show?

- How much revenue or growth do businesses experience after receiving investments?

- Are there common challenges or issues that businesses face post-investment?

- **6**.Shark Preferences:

- Which industries or product categories do specific sharks prefer to invest in?

- Do certain sharks have distinct investment strategies or criteria?

- How has the investment behavior of individual sharks evolved over different seasons?

- **7**. Entrepreneur Demographics:

- What are the demographic characteristics (age, gender, location) of the entrepreneurs who appear on the show?

- Do certain demographics tend to perform better or worse in terms of securing investments?

- **8**. Consumer Trends and Preferences:

- Are there trends in the types of products or services that receive the most attention from the sharks?

- How do consumer preferences reflected on the show compare to broader market trends?

- **9**.Predictive Modeling:

- Can you develop a predictive model to forecast which businesses are likely to secure investments based on various factors from their pitch?

- What factors are the most influential in predicting investment success?

- **10**. Ethical and Social Impact:

- Are there any ethical or social impact considerations related to the businesses that secure investments on "Shark Tank"?

- How do the sharks evaluate businesses with a social or environmental mission?

Questions and Answers

- Question1-What are the names and data types of the columns?
- Ans- The Shark Tank TV show dataset typically consists of information related to the entrepreneurs and their pitches to the investors . The specific columns and their data types can vary depending on the source and version of the dataset. However, I can provide you with a general idea of the common columns and their data types that you might find in a Shark Tank dataset:
-
- 1. Entrepreneur Information:
-   - Entrepreneur Name
-   - Entrepreneur Gender
-   - Entrepreneur Age
-   - Entrepreneur Location
-
- 2. Pitch Details:
-   -Product Name
-   -Product Description
-   -Product Category
-   - Investment Amount Requested
-   - Equity Offered
-
- 3. Shark Information:
-   - Shark Name (String)
-   - Shark Gender
-   - Shark Net Worth
-   - Shark Deal Status
-
- 4. Pitch Outcome:
-   - Deal Outcome
-   - Amount Invested
-   - Equity Given
-   - Valuation

- Question2-What are the basic summary statistics?

  Ans-Basic summary statistics are fundamental metrics that provide a concise and informative overview of a dataset.

  1. Mean :The mean is the sum of all values in a dataset divided by the number of values. It represents the central value of the data distribution.

  2. Median :The median is the middle value of an ordered dataset. It is less sensitive to outliers than the mean and represents the central tendency.

  3. Mode: The mode is the value that appears most frequently in a dataset. dataset can have one mode , more than one mode , or no mode at all.

  4. Range: The range is the difference between the maximum and minimum values in a dataset.

- Question3-Are there any categorical variables and missing values ? t .

  Ans-import pandas as pd

```
# Load your dataset (replace 'your_dataset.csv' with the actual dataset file path)

df = pd.read_csv('your_dataset.csv')


# Check for categorical variables

categorical_columns = df.select_dtypes(include=['object'])

if not categorical_columns.empty:

    print("Categorical Variables:")

    print(categorical_columns)
```

```
# Check for missing values
missing_values = df.isnull().sum()
if not missing_values.empty:
    print("\nMissing Values:")
    print(missing_values)
```

- Question4-Are there any outliers in the data? If so use box plots, histograms and visualize .

Ans-import pandas as pd

```
import matplotlib.pyplot as plt
import seaborn as sns

# Load your dataset (replace 'your_dataset.csv' with the actual dataset file path)
df = pd.read_csv('your_dataset.csv')

# Specify the numeric columns you want to analyze for outliers
numeric_columns = df.select_dtypes(include=['int64', 'float64'])

# Create box plots for numeric columns
plt.figure(figsize=(12, 6))
sns.boxplot(data=numeric_columns)
plt.title('Box Plot of Numeric Columns (Outliers)')
plt.xticks(rotation=90)
```

plt.show()

Create histograms for numeric columns

plt.figure(figsize=(12, 6))

numeric_columns.hist(bins=20, figsize=(12, 6))

plt.title('Histograms of Numeric Columns')

plt.tight_layout()

plt.show()

- Question5-Is the data balanced or imbalanced? Visualize .

  Ans-import pandas as pd

  import matplotlib.pyplot as plt

  import seaborn as sns

  # Load your dataset (replace 'your_dataset.csv' with the actual dataset file path)

  df = pd.read_csv('your_dataset.csv')

  # Specify the categorical variable you want to analyze (replace 'target_column' with the actual column name)

  categorical_column = 'target_column'

  # Create a count plot to visualize the distribution

  plt.figure(figsize=(8, 6))

  sns.countplot(data=df, x=categorical_column)

```
plt.title(f'Distribution of {categorical_column}')

plt.xlabel(categorical_column)

plt.ylabel('Count')

plt.xticks(rotation=90)

plt.show()
```

- Question6-What is the target variable ?

  Ans-The target variable, also known as the dependent variable, is the variable in a dataset that you are trying to predict or explain. It is the variable whose values you want to model or analyze based on the values of other variables

  1. Regression Analysis:In regression tasks, the target variable is a continuous or numeric variable that you want to predict..

  2. Classification Analysis: In classification tasks, the target variable is a categorical variable that you want to classify or predict. For example, in spam email detection, the target variable could be "spam" or "not spam."

  3. Time Series Analysis: In time series analysis, the target variable is typically a variable that varies over time, such as stock prices, temperature, or sales revenue.

- Question7-What are the units of measurement for numerical columns?

Ans-The units of measurement for numerical columns in a dataset depend on the specific dataset and the nature of the variables. Numerical columns can represent a wide range of measurements, and their units can vary accordingly. Here are some common examples of numerical columns and their potential units of measurement:

1. **Age:** Numeric variable representing a person's age. The unit of measurement is typically "years

2. **Weight:** Numeric variable representing a person's weight. The unit of measurement is typically "kilograms" (kg) or "pounds" (lbs).

3. **Height:** Numeric variable representing a person's height. The unit of measurement is typically "centimeters" (cm) or "inches" (in).

4. **Time:** Numeric variable representing time. The unit of measurement can be "seconds" (s), "minutes" (min), "hours" (hr), or other time units.

- Question8-Do you have domain clarification? Brief it .

  Ans-Certainly! In the context of data analysis and data science, "domain clarification" refers to the process of gaining a deep understanding of the specific field or industry to which the data belongs.

  1. Understanding the Domain:

  - Domain clarification starts with acquiring a solid understanding of the field or industry associated with the dataset. This could be healthcare, finance, retail, education, or any other domain.

  2. Industry Jargon

- Familiarize yourself with the industry-specific jargon, terminology, and concepts. Different domains have their own set of terms and definitions.

  3. Data Relevance:

  - Determine which aspects of the domain are most relevant to the dataset and analysis. Not all information may be pertinent to your specific goals.

  5. Data Sources:

  - Understand the sources of the data and how it was collected within the domain. Know if there are any data quality or bias issues specific to the domain.

- Question9-Are there any time-based trends or pattern?

Ans-To identify time-based trends or patterns in a dataset, you can perform time series analysis or visualizations. Time-based trends are common in many domains, and they can provide valuable insights into how data changes over time. Here are steps you can follow to assess time-based trends:

1. Data Preparation:

- Ensure that your dataset includes a time-related column, such as a timestamp, date, or period, that indicates when each observation was recorded or measured.

2. Data Visualization:

- Create time series plots or line plots to visualize the data over time. The x-axis should represent time, and the y-axis represents the variable of interest. Use different colors or styles to distinguish between different categories or groups if applicable.

3. Seasonal Decomposition:

- Use seasonal decomposition techniques, such as decomposition into trend, seasonality, and residuals (e.g., using the decomposition function in Python's statsmodels library), to separate the data into its underlying components. This helps identify recurring patterns and trends.

4. Time Series Analysis:

- Perform time series analysis techniques like autocorrelation and partial autocorrelation analysis to detect any temporal dependencies or patterns in the data.

5. Statistical Tests:

- Apply statistical tests, such as the Augmented Dickey-Fuller test for stationarity or other relevant tests, to check for trends and seasonality in the data.

6. Visualizations:

- Consider using heatmaps, calendar plots, or other specialized visualizations to highlight time-based patterns, especially if your data involves multiple dimensions or categories.

- Question10-Are there any correlations between variables? Calculate correlations

  Ans-import pandas as pd

  import numpy as np

  # Load your dataset (replace 'your_dataset.csv' with the actual dataset file path)

  df = pd.read_csv('your_dataset.csv')

  # Calculate Pearson's correlation matrix for all numeric columns

  correlation_matrix = df.corr(method='pearson')

  # Display the correlation matrix

  print("Pearson's Correlation Matrix:")

  print(correlation_matrix)

Code based questions are done on jupyter notebook in the form of pdf file below.

Thank You.

eda-3-1

November 10, 2023

0.1 Shark Tank India

0.2 "Unlocking India's entrepreneurial potential one pitch at a time on Shark Tank India!"

0.3 Question1-What are the names and data types of the columns?

[72]:
```python
import pandas as pd

# Create a sample DataFrame
data = {
    'Column1': [1, 2, 3, 4, 5],
    'Column2': ['A', 'B', 'C', 'D', 'E'],
    'Column3': [1.1, 2.2, 3.3, 4.4, 5.5]
}

df = pd.DataFrame(data)

# Get column names
column_names = df.columns

# Get data types of columns
column_data_types = df.dtypes

# Print column names and data types
print("Column Names:")
print(column_names)

print("\nData Types of Columns:")
print(column_data_types)
```

```
Column Names:
Index(['Column1', 'Column2', 'Column3'], dtype='object')

Data Types of Columns:
Column1      int64
Column2     object
Column3    float64
dtype: object
```

0.4 Question2-What are the basic summary statistics?

[73]:
```
import pandas as pd

# Create a sample DataFrame
data = {
    'Age': [25, 30, 35, 40, 45],
    'Income (in thousands)': [50, 60, 70, 80, 90],
    'Score': [85, 90, 78, 92, 88]
}

df = pd.DataFrame(data)

# Generate basic summary statistics
summary_statistics = df.describe()

# Print the summary statistics
print("Basic Summary Statistics for Numerical Columns:")
print(summary_statistics)
```

```
Basic Summary Statistics for Numerical Columns:
             Age  Income (in thousands)      Score
count   5.000000               5.000000   5.000000
mean   35.000000              70.000000  86.600000
std     7.905694              15.811388   5.458938
min    25.000000              50.000000  78.000000
25%    30.000000              60.000000  85.000000
50%    35.000000              70.000000  88.000000
75%    40.000000              80.000000  90.000000
max    45.000000              90.000000  92.000000
```

Question-3:Are there any categorical variables and missing values

[11]:
```
import pandas as pd

# Load your dataset (replace 'your_dataset.csv' with the actual dataset file␣
↪path)
df = pd.read_csv("C:\\Users\\admin\\Shark Tank India Dataset.csv")

# Check for categorical variables
categorical_columns = df.select_dtypes(include=['object'])
if not categorical_columns.empty:
    print("Categorical Variables:")
    print(categorical_columns)

# Check for missing values
missing_values = df.isnull().sum()
if not missing_values.empty:
```

```
    print("\nMissing Values:")
    print(missing_values)

df.head(10)
```

```
Categorical Variables:
              brand_name                                         idea
0    BluePine Industries                                  Frozen Momos
1          Booz scooters  Renting e-bike for mobility in private spaces
2    Heart up my Sleeves                           Detachable Sleeves
3             Tagz Foods                          Healthy Potato Chips
4         Head and Heart                       Brain Development Course
..                   …                                              …
112        Green Protein                          Plant-Based Protein
113              On2Cook                        Fastest Cooking Device
114        Jain Shikanji                                     Lemonade
115                Woloo                               Washroom Finder
116         Elcare India                          Carenting for Elders

[117 rows x 2 columns]

Missing Values:
episode_number         0
pitch_number           0
brand_name             0
idea                   0
deal                   0
pitcher_ask_amount     0
ask_equity             0
ask_valuation          0
deal_amount            0
deal_equity            0
deal_valuation         0
ashneer_present        0
anupam_present         0
aman_present           0
namita_present         0
vineeta_present        0
peyush_present         0
ghazal_present         0
ashneer_deal           0
anupam_deal            0
aman_deal              0
namita_deal            0
vineeta_deal           0
peyush_deal            0
ghazal_deal            0
```

```
total_sharks_invested    0
amount_per_shark         0
equity_per_shark         0
dtype: int64
```

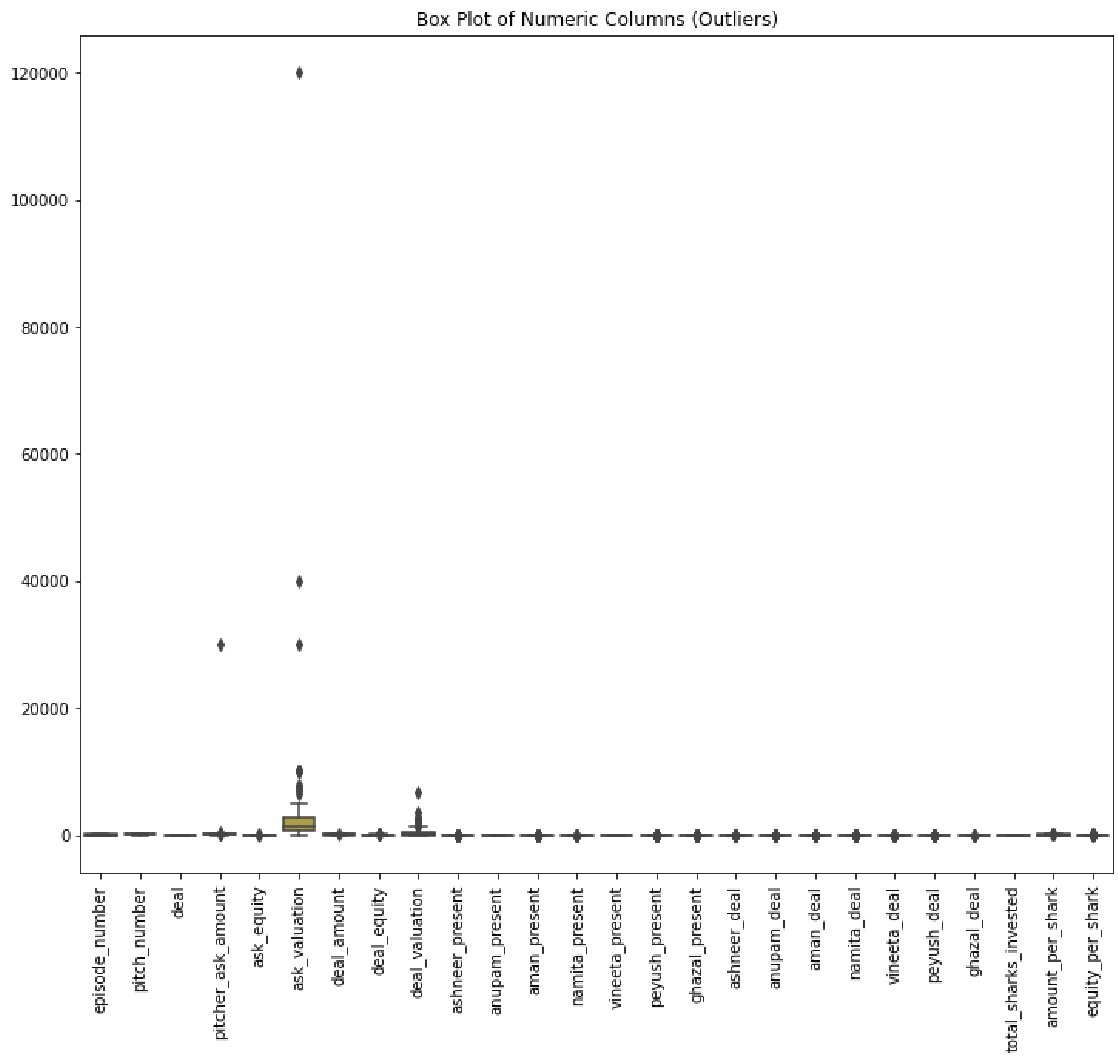
[11]:

```
   episode_number  pitch_number           brand_name  \
0               1             1  BluePine Industries
1               1             2        Booz scooters
2               1             3  Heart up my Sleeves
3               2             4           Tagz Foods
4               2             5       Head and Heart
5               2             6        Agro tourism
6               3             7          Qzense Labs
7               3             8            Peeschute
8               3             9                 NOCD
9               4            10                Cosiq

                                     idea  deal  pitcher_ask_amount  \
0                            Frozen Momos     1                50.0
1  Renting e-bike for mobility in private spaces     1            40.0
2                      Detachable Sleeves     1                25.0
3                    Healthy Potato Chips     1                70.0
4                Brain Development Course     0                50.0
5                                 Tourism     0                50.0
6                  Food Freshness Detector     0               100.0
7                    Disposable Urine Bag     1                75.0
8                            Energy Drink     1                50.0
9                     Intelligent Skincare     1                50.0

   ask_equity  ask_valuation  deal_amount  deal_equity  …  ashneer_deal  \
0        5.00        1000.00         75.0        16.00  …             1
1       15.00         266.67         40.0        50.00  …             1
2       10.00         250.00         25.0        30.00  …             0
3        1.00        7000.00         70.0         2.75  …             1
4        5.00        1000.00          0.0         0.00  …             0
5        5.00        1000.00          0.0         0.00  …             0
6        0.25       40000.00          0.0         0.00  …             0
7        4.00        1875.00         75.0         6.00  …             0
8        2.00        2500.00         20.0        15.00  …             0
9        7.50         666.67         50.0        25.00  …             0

   anupam_deal  aman_deal  namita_deal  vineeta_deal  peyush_deal  \
0            0          1            0             1            0
1            0          0            0             1            0
2            1          0            0             1            0
3            0          0            0             0            0
4            0          0            0             0            0
```

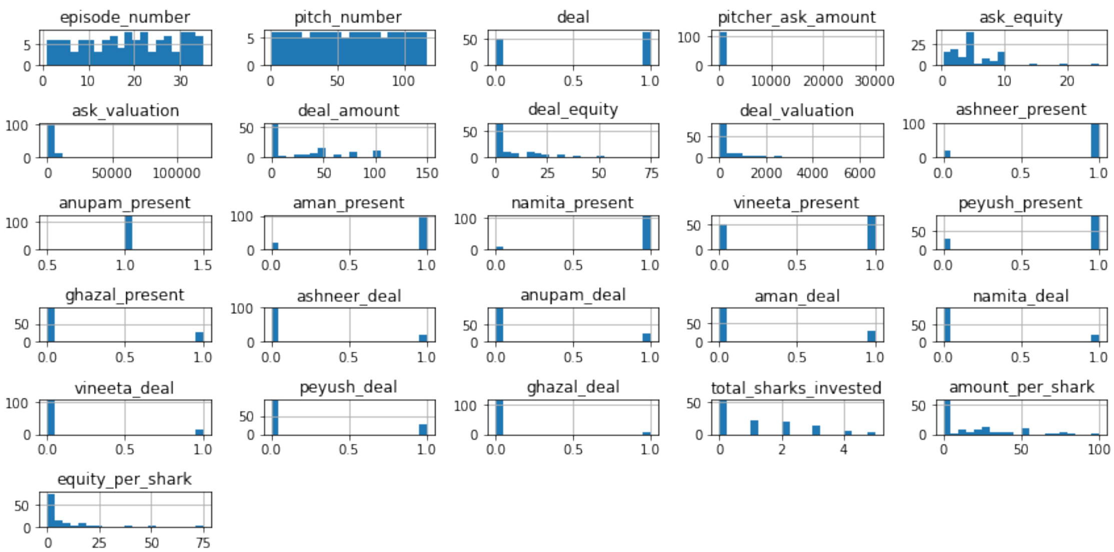
```
5             0          0            0             0            0
6             0          0            0             0            0
7             0          1            0             0            0
8             0          0            0             1            0
9             1          0            0             1            0

   ghazal_deal  total_sharks_invested  amount_per_shark  equity_per_shark
0            0                      3              25.0          5.333333
1            0                      2              20.0         25.000000
2            0                      2              12.5         15.000000
3            0                      1              70.0          2.750000
4            0                      0               0.0          0.000000
5            0                      0               0.0          0.000000
6            0                      0               0.0          0.000000
7            0                      1              75.0          6.000000
8            0                      1              20.0         15.000000
9            0                      2              25.0         12.500000

[10 rows x 28 columns]
```

Question 4: Are there any outliers in the data? If so use box plots, histograms and visualize

[9]:
```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load your dataset (replace 'your_dataset.csv' with the actual dataset file␣
↪path)
df = pd.read_csv("C:\\Users\\admin\\Shark Tank India Dataset.csv")

# Specify the numeric columns you want to analyze for outliers
numeric_columns = df.select_dtypes(include=['int64', 'float64'])

# Create box plots for numeric columns
plt.figure(figsize=(12, 6))
sns.boxplot(data=numeric_columns)
plt.title('Box Plot of Numeric Columns (Outliers)')
plt.xticks(rotation=90)
plt.show()

# Create histograms for numeric columns
plt.figure(figsize=(12, 6))
numeric_columns.hist(bins=20, figsize=(12, 6))
plt.title('Histograms of Numeric Columns')
plt.tight_layout()
plt.show()
```




```
<Figure size 864x432 with 0 Axes>
```



0.5 Question5-Is the data balanced or imbalanced? Visualize

[20]:
```
import pandas as pd
import numpy as np

# Load your dataset (replace 'your_dataset.csv' with the actual dataset file␣
↪path)
# df = pd.read_csv("C:\\Users\\admin\\Shark Tank India Dataset.csv")

# Calculate Pearson's correlation matrix for all numeric columns
correlation_matrix = df.corr(method='pearson')

# Display the correlation matrix
print("Pearson's Correlation Matrix:")
print(correlation_matrix)
```

```
Pearson's Correlation Matrix:
                      episode_number  pitch_number      deal  \
episode_number              1.000000      0.998850 -0.214033
pitch_number                0.998850      1.000000 -0.223068
deal                       -0.214033     -0.223068  1.000000
pitcher_ask_amount         -0.069500     -0.074881 -0.106927
ask_equity                 -0.077674     -0.080792 -0.076438
ask_valuation              -0.070829     -0.073517 -0.151695
deal_amount                -0.187023     -0.199542  0.736002
deal_equity                -0.114719     -0.119973  0.609043
deal_valuation             -0.146336     -0.151003  0.409138
ashneer_present            -0.342116     -0.338967 -0.020728
```

anupam_present	NaN	NaN	NaN
aman_present	-0.342116	-0.338967	-0.020728
namita_present	-0.107808	-0.104920	0.065484
vineeta_present	0.136628	0.157186	-0.161874
peyush_present	0.425047	0.405589	0.084100
ghazal_present	0.528721	0.535060	-0.059761
ashneer_deal	-0.116462	-0.117380	0.373509
anupam_deal	-0.121671	-0.129104	0.454369
aman_deal	-0.264855	-0.272245	0.461369
namita_deal	-0.041819	-0.053756	0.430422
vineeta_deal	-0.035758	-0.028765	0.342997
peyush_deal	0.022564	0.013214	0.489898
ghazal_deal	0.229482	0.226210	0.225630
total_sharks_invested	-0.121894	-0.131541	0.759342
amount_per_shark	-0.157896	-0.166871	0.653882
equity_per_shark	-0.073187	-0.079109	0.461046

	pitcher_ask_amount	ask_equity	ask_valuation \
episode_number	-0.069500	-0.077674	-0.070829
pitch_number	-0.074881	-0.080792	-0.073517
deal	-0.106927	-0.076438	-0.151695
pitcher_ask_amount	1.000000	0.470198	0.911174
ask_equity	0.470198	1.000000	0.260603
ask_valuation	0.911174	0.260603	1.000000
deal_amount	-0.080719	-0.177955	-0.090163
deal_equity	-0.067233	0.288341	-0.155565
deal_valuation	-0.045988	-0.310331	0.017869
ashneer_present	0.040759	-0.047404	0.053817
anupam_present	NaN	NaN	NaN
aman_present	0.040759	-0.047404	0.053817
namita_present	0.029396	0.030743	0.042539
vineeta_present	-0.104309	0.068224	-0.032090
peyush_present	0.052044	-0.065365	-0.050228
ghazal_present	-0.049727	0.035071	-0.060577
ashneer_deal	-0.044099	-0.088782	-0.052650
anupam_deal	-0.050348	0.013584	-0.095488
aman_deal	-0.053685	-0.140932	-0.085926
namita_deal	-0.045850	0.027449	-0.075664
vineeta_deal	-0.036575	0.096837	-0.068814
peyush_deal	-0.052946	-0.026575	-0.089131
ghazal_deal	-0.024308	0.071451	-0.050633
total_sharks_invested	-0.084288	-0.028387	-0.140668
amount_per_shark	-0.071281	-0.165141	-0.066631
equity_per_shark	-0.050873	0.238074	-0.118541

	deal_amount	deal_equity	deal_valuation \
episode_number	-0.187023	-0.114719	-0.146336
pitch_number	-0.199542	-0.119973	-0.151003

deal	0.736002	0.609043	0.409138
pitcher_ask_amount	-0.080719	-0.067233	-0.045988
ask_equity	-0.177955	0.288341	-0.310331
ask_valuation	-0.090163	-0.155565	0.017869
deal_amount	1.000000	0.370487	0.636411
deal_equity	0.370487	1.000000	-0.082694
deal_valuation	0.636411	-0.082694	1.000000
ashneer_present	-0.026851	-0.093570	0.018973
anupam_present	NaN	NaN	NaN
aman_present	-0.026851	-0.093570	0.018973
namita_present	0.065587	-0.065976	0.073284
vineeta_present	-0.082982	0.014315	0.008492
peyush_present	-0.004062	0.088750	-0.138948
ghazal_present	-0.057153	0.062931	-0.082732
ashneer_deal	0.392853	0.127167	0.206179
anupam_deal	0.335148	0.267897	0.219510
aman_deal	0.482279	0.150735	0.360149
namita_deal	0.443929	0.136270	0.178616
vineeta_deal	0.315109	0.359560	-0.001066
peyush_deal	0.384539	0.399927	0.138179
ghazal_deal	0.188542	0.188478	-0.002742
total_sharks_invested	0.695950	0.432880	0.319298
amount_per_shark	0.793392	0.359031	0.581751
equity_per_shark	0.239212	0.897922	-0.080662

	ashneer_present	…	ashneer_deal	anupam_deal	\
episode_number	-0.342116	…	-0.116462	-0.121671	
pitch_number	-0.338967	…	-0.117380	-0.129104	
deal	-0.020728	…	0.373509	0.454369	
pitcher_ask_amount	0.040759	…	-0.044099	-0.050348	
ask_equity	-0.047404	…	-0.088782	0.013584	
ask_valuation	0.053817	…	-0.052650	-0.095488	
deal_amount	-0.026851	…	0.392853	0.335148	
deal_equity	-0.093570	…	0.127167	0.267897	
deal_valuation	0.018973	…	0.206179	0.219510	
ashneer_present	1.000000	…	0.205939	-0.005886	
anupam_present	NaN	…	NaN	NaN	
aman_present	1.000000	…	0.205939	-0.005886	
namita_present	-0.141843	…	-0.001957	0.018596	
vineeta_present	-0.387059	…	-0.127833	-0.108356	
peyush_present	-0.252768	…	0.010581	0.095532	
ghazal_present	-0.823754	…	-0.142857	-0.016971	
ashneer_deal	0.205939	…	1.000000	0.203653	
anupam_deal	-0.005886	…	0.203653	1.000000	
aman_deal	0.246972	…	0.311859	0.211158	
namita_deal	-0.025346	…	0.116921	0.134733	
vineeta_deal	-0.108422	…	0.220354	0.185069	
peyush_deal	-0.143856	…	0.219578	0.324617	

ghazal_deal	-0.475201	…	-0.024079	0.139591
total_sharks_invested	-0.026656	…	0.573341	0.609817
amount_per_shark	-0.049810	…	0.107673	0.088905
equity_per_shark	-0.102568	…	-0.049699	0.063640

	aman_deal	namita_deal	vineeta_deal	peyush_deal \
episode_number	-0.264855	-0.041819	-0.035758	0.022564
pitch_number	-0.272245	-0.053756	-0.028765	0.013214
deal	0.461369	0.430422	0.342997	0.489898
pitcher_ask_amount	-0.053685	-0.045850	-0.036575	-0.052946
ask_equity	-0.140932	0.027449	0.096837	-0.026575
ask_valuation	-0.085926	-0.075664	-0.068814	-0.089131
deal_amount	0.482279	0.443929	0.315109	0.384539
deal_equity	0.150735	0.136270	0.359560	0.399927
deal_valuation	0.360149	0.178616	-0.001066	0.138179
ashneer_present	0.246972	-0.025346	-0.108422	-0.143856
anupam_present	NaN	NaN	NaN	NaN
aman_present	0.246972	-0.025346	-0.108422	-0.143856
namita_present	0.043412	0.155022	-0.051660	0.037427
vineeta_present	-0.234098	-0.150433	0.337100	-0.254899
peyush_present	-0.002776	0.073610	-0.194332	0.314426
ghazal_present	-0.251628	0.005846	0.163984	0.048795
ashneer_deal	0.311859	0.116921	0.220354	0.219578
anupam_deal	0.211158	0.134733	0.185069	0.324617
aman_deal	1.000000	0.294019	0.024582	0.120692
namita_deal	0.294019	1.000000	0.142601	0.151757
vineeta_deal	0.024582	0.142601	1.000000	-0.028006
peyush_deal	0.120692	0.151757	-0.028006	1.000000
ghazal_deal	-0.141494	0.247520	0.442232	0.203965
total_sharks_invested	0.549698	0.559506	0.464893	0.560000
amount_per_shark	0.323914	0.232283	0.061331	0.198726
equity_per_shark	0.007305	0.024287	0.113549	0.311324

	ghazal_deal	total_sharks_invested	amount_per_shark \
episode_number	0.229482	-0.121894	-0.157896
pitch_number	0.226210	-0.131541	-0.166871
deal	0.225630	0.759342	0.653882
pitcher_ask_amount	-0.024308	-0.084288	-0.071281
ask_equity	0.071451	-0.028387	-0.165141
ask_valuation	-0.050633	-0.140668	-0.066631
deal_amount	0.188542	0.695950	0.793392
deal_equity	0.188478	0.432880	0.359031
deal_valuation	-0.002742	0.319298	0.581751
ashneer_present	-0.475201	-0.026656	-0.049810
anupam_present	NaN	NaN	NaN
aman_present	-0.475201	-0.026656	-0.049810
namita_present	0.081264	0.073785	0.066861
vineeta_present	0.221751	-0.137834	-0.048368

peyush_present	0.144814	0.122518	-0.037216
ghazal_present	0.471940	0.014639	-0.049405
ashneer_deal	-0.024079	0.573341	0.107673
anupam_deal	0.139591	0.609817	0.088905
aman_deal	-0.141494	0.549698	0.323914
namita_deal	0.247520	0.559506	0.232283
vineeta_deal	0.442232	0.464893	0.061331
peyush_deal	0.203965	0.560000	0.198726
ghazal_deal	1.000000	0.394771	0.004715
total_sharks_invested	0.394771	1.000000	0.293030
amount_per_shark	0.004715	0.293030	1.000000
equity_per_shark	0.025510	0.138429	0.407074

	equity_per_shark
episode_number	-0.073187
pitch_number	-0.079109
deal	0.461046
pitcher_ask_amount	-0.050873
ask_equity	0.238074
ask_valuation	-0.118541
deal_amount	0.239212
deal_equity	0.897922
deal_valuation	-0.080662
ashneer_present	-0.102568
anupam_present	NaN
aman_present	-0.102568
namita_present	-0.114103
vineeta_present	0.028488
peyush_present	0.083101
ghazal_present	0.040184
ashneer_deal	-0.049699
anupam_deal	0.063640
aman_deal	0.007305
namita_deal	0.024287
vineeta_deal	0.113549
peyush_deal	0.311324
ghazal_deal	0.025510
total_sharks_invested	0.138429
amount_per_shark	0.407074
equity_per_shark	1.000000

[26 rows x 26 columns]

0.6 Question6-What is the target variable ?

[70]:
```python
import pandas as pd
from sklearn.linear_model import LinearRegression

# Create a sample DataFrame
data = {
    'X1': [1, 2, 3, 4, 5],
    'X2': [2, 4, 5, 4, 5],
    'y': [3, 6, 8, 8, 10]
}

df = pd.DataFrame(data)

# Define features (X) and the target variable (y)
X = df[['X1', 'X2']]
y = df['y']

# Create a linear regression model
model = LinearRegression()

# Fit the model to the data
model.fit(X, y)

# Make predictions
predictions = model.predict(X)

# Print the predictions
print("Predictions:", predictions)
```

```
Predictions: [ 3.  6.  8.  8. 10.]
```

0.7 Question7-What are the units of measurement for numerical columns?

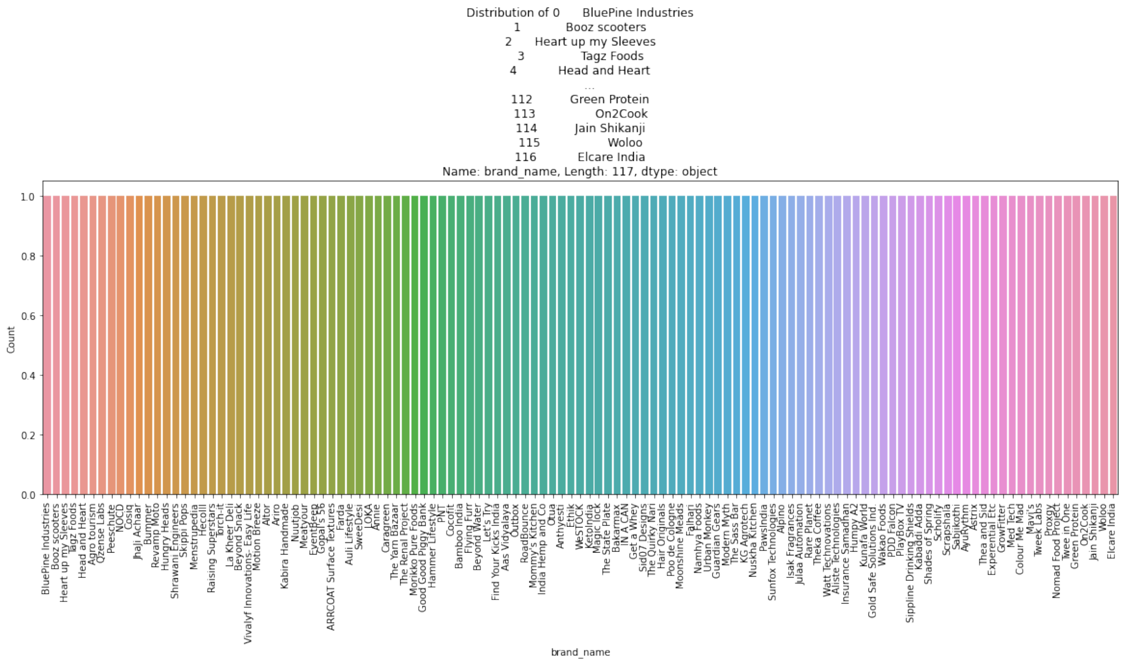
[71]:
```python
import pandas as pd

# Create a sample DataFrame
data = {
    'Length (inches)': [10.2, 8.5, 12.7, 9.8, 11.0],
    'Weight (pounds)': [0.5, 0.4, 0.7, 0.6, 0.8],
    'Price (dollars)': [25, 30, 35, 28, 40],
    'Product Name': ['Widget A', 'Widget B', 'Widget C', 'Widget D', 'Widget E']
}

df = pd.DataFrame(data)

# Display summary statistics for numerical columns
numerical_summary = df.describe()
```

```
# Identify unique values in non-numeric columns
unique_product_names = df['Product Name'].unique()

# Print summary statistics and unique product names
print("Summary Statistics for Numerical Columns:")
print(numerical_summary)

print("\nUnique Product Names:")
print(unique_product_names)
```

```
Summary Statistics for Numerical Columns:
       Length (inches)  Weight (pounds)  Price (dollars)
count         5.000000         5.000000          5.00000
mean         10.440000         0.600000         31.60000
std           1.553383         0.158114          5.94138
min           8.500000         0.400000         25.00000
25%           9.800000         0.500000         28.00000
50%          10.200000         0.600000         30.00000
75%          11.000000         0.700000         35.00000
max          12.700000         0.800000         40.00000

Unique Product Names:
['Widget A' 'Widget B' 'Widget C' 'Widget D' 'Widget E']
```

0.8 Question -8

[18]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load your dataset (replace 'your_dataset.csv' with the actual dataset file␣
↪path)
df = pd.read_csv("C:\\Users\\admin\\Shark Tank India Dataset.csv")

# Specify the categorical variable you want to analyze (replace 'target_column'␣
↪with the actual column name)
categorical_column = df['brand_name']

# Create a count plot to visualize the distribution
plt.figure(figsize=(20, 6))
sns.countplot(data=df, x=categorical_column)
plt.title(f'Distribution of {categorical_column}')
# plt.xlabel(categorical_column)
plt.ylabel('Count')
plt.xticks(rotation=90)
plt.show()
```



0.9 Question 9: Do you have domain clarification? Brief it

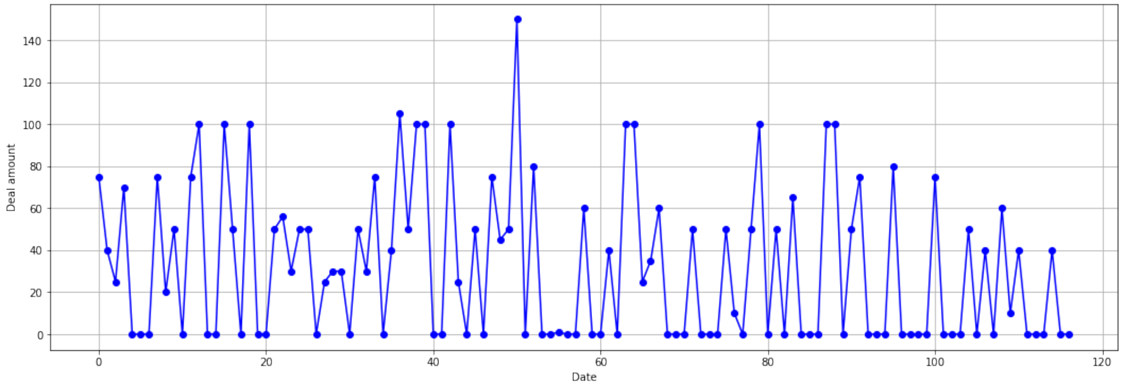
[67]:
```python
print("Total deal amount : ", df['Total Deal Amount'].sum())
print("Total Successful deals : ", (df['Got Deal']==1).sum())
print("Average deal amount (ADA, in dollars):", (df['Total Deal Amount'].sum()/
↪(df['Got Deal']==1).sum()))
```

```
Total deal amount :  226488166.0
Total Successful deals :  765
Average deal amount (ADA, in dollars): 296062.96209150326
```

0.10 Question10-Are there any correlations between variables? Calculate correlations

[34]:
```python
# Resample data to analyze trends over time (e.g., monthly or weekly)
# You can change 'M' to 'W' for weekly or use other frequencies
resampled_data = df['deal_amount']

# Create a line plot to visualize the viewership trends
plt.figure(figsize=(18, 6))
plt.plot(resampled_data.index, resampled_data.values, marker='o',␣
↪linestyle='-', color='b')
plt.xlabel('Date')
plt.ylabel('Deal amount')
plt.grid(True)
```

```
plt.show()
```




```
[ ]:
```