```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
```

```
df=pd.read_csv('ShartankIndiaAllPitches.csv')
df
```

| | Episode Number | Pitch Number | Brand | Idea | Investment Amount (In Lakhs INR) | Debt (In lakhs INR) | Equity | Anupam | Ashneer | Namita | Aman | Peyush | Vineeta | Ghazal | Season |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 1 | BluePine Industries | Frozen Momos | 75 | 0 | 18% | N | Y | N | Y | N | Y | N | 1 |
| **1** | 1 | 2 | Booz scooters | Renting e-bike for mobility in private spaces | 40 | 0 | 50% | N | Y | N | N | N | Y | N | 1 |
| **2** | 1 | 3 | Heart up my Sleeves | Detachable Sleeves | 25 | 0 | 30% | Y | N | N | N | N | Y | N | 1 |
| **3** | 2 | 4 | Tagz Foods | Healthy Potato Chips | 70 | 0 | 2.75% | N | Y | N | N | N | N | N | 1 |
| **4** | 2 | 5 | Head and Heart | Brain Development Course | 0 | 0 | 0 | N | N | N | N | N | N | N | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **112** | 34 | 113 | Green Protein | Plant-Based Protein | 0 | 0 | 0 | N | N | N | N | N | N | N | 1 |
| **113** | 34 | 114 | On2Cook | Fastest Cooking Device | 0 | 0 | 0 | N | N | N | N | N | N | N | 1 |
| **114** | 35 | 115 | Jain Shikanji | Lemonade | 40 | 0 | 30.00% | Y | Y | N | Y | N | Y | N | 1 |
| **115** | 35 | 116 | Woloo | Washroom Finder | 0 | 0 | 0 | N | N | N | N | N | N | N | 1 |
| | | | | Floore | | | Carpeting for | | | | | | | | |

```
df.columns
```

```
Index(['Episode Number', 'Pitch Number', 'Brand', 'Idea',
       'Investment Amount (In Lakhs INR) ', 'Debt (In lakhs INR)', 'Equity',
       'Anupam', 'Ashneer', 'Namita', 'Aman', 'Peyush', 'Vineeta', 'Ghazal',
       'Season'],
      dtype='object')
```

```
df.index
```

```
RangeIndex(start=0, stop=117, step=1)
```

```
df.shape
```

```
(117, 15)
```

```
df.size
```

```
1755
```

```
df.dtypes
```

| | 0 |
|---|---|
| **Episode Number** | int64 |
| **Pitch Number** | int64 |
| **Brand** | object |
| **Idea** | object |
| **Investment Amount (In Lakhs INR)** | int64 |
| **Debt (In lakhs INR)** | int64 |
| **Equity** | object |
| **Anupam** | object |
| **Ashneer** | object |
| **Namita** | object |
| **Aman** | object |
| **Peyush** | object |
| **Vineeta** | object |
| **Ghazal** | object |
| **Season** | int64 |

```
df.head(5)
```

| | Episode Number | Pitch Number | Brand | Idea | Investment Amount (In Lakhs INR) | Debt (In lakhs INR) | Equity | Anupam | Ashneer | Namita | Aman | Peyush | Vineeta | Ghazal | Season |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 1 | BluePine Industries | Frozen Momos | 75 | 0 | 18% | N | Y | N | Y | N | Y | N | 1 |
| **1** | 1 | 2 | Booz scooters | Renting e-bike for mobility in private spaces | 40 | 0 | 50% | N | Y | N | N | N | Y | N | 1 |
| **2** | 1 | 3 | Heart up my Sleeves | Detachable Sleeves | 25 | 0 | 30% | Y | N | N | N | N | Y | N | 1 |
| **3** | 2 | 4 | Tagz Foods | Healthy Potato Chips | 70 | 0 | 2.75% | N | Y | N | N | N | N | N | 1 |
| | | | | Brain | | | | | | | | | | | |

```
df.tail(5)
```

| | Episode Number | Pitch Number | Brand | Idea | Investment Amount (In Lakhs INR) | Debt (In lakhs INR) | Equity | Anupam | Ashneer | Namita | Aman | Peyush | Vineeta | Ghazal | Season |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **112** | 34 | 113 | Green Protein | Plant-Based Protein | 0 | 0 | 0 | N | N | N | N | N | N | N | 1 |
| **113** | 34 | 114 | On2Cook | Fastest Cooking Device | 0 | 0 | 0 | N | N | N | N | N | N | N | 1 |
| **114** | 35 | 115 | Jain Shikanji | Lemonade | 40 | 0 | 30.00% | Y | Y | N | Y | N | Y | N | 1 |
| **115** | 35 | 116 | Woloo | Washroom Finder | 0 | 0 | 0 | N | N | N | N | N | N | N | 1 |
| | | | Elcare | Carenting | | | | | | | | | | | |

```
df.empty
```

```
False
```

```
df.any()
```

| | 0 |
|---|---|
| **Episode Number** | True |
| **Pitch Number** | True |
| **Brand** | True |
| **Idea** | True |
| **Investment Amount (In Lakhs INR)** | True |
| **Debt (In lakhs INR)** | True |
| **Equity** | True |
| **Anupam** | True |
| **Ashneer** | True |
| **Namita** | True |
| **Aman** | True |
| **Peyush** | True |
| **Vineeta** | True |
| **Ghazal** | True |
| **Season** | True |

```
df.all()
```

| | 0 |
|---|---|
| **Episode Number** | True |
| **Pitch Number** | True |
| **Brand** | True |
| **Idea** | True |
| **Investment Amount (In Lakhs INR)** | False |
| **Debt (In lakhs INR)** | False |
| **Equity** | True |
| **Anupam** | True |
| **Ashneer** | True |
| **Namita** | True |
| **Aman** | True |
| **Peyush** | True |
| **Vineeta** | True |
| **Ghazal** | True |
| **Season** | True |

```
df.dropna()
```

| | Episode Number | Pitch Number | Brand | Idea | Investment Amount (In Lakhs INR) | Debt (In lakhs INR) | Equity | Anupam | Ashneer | Namita | Aman | Peyush | Vineeta | Ghazal | Season |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 1 | BluePine Industries | Frozen Momos | 75 | 0 | 18% | N | Y | N | Y | N | Y | N | 1 |
| **1** | 1 | 2 | Booz scooters | Renting e-bike for mobility in private spaces | 40 | 0 | 50% | N | Y | N | N | N | Y | N | 1 |
| **2** | 1 | 3 | Heart up my Sleeves | Detachable Sleeves | 25 | 0 | 30% | Y | N | N | N | N | Y | N | 1 |
| **3** | 2 | 4 | Tagz Foods | Healthy Potato Chips | 70 | 0 | 2.75% | N | Y | N | N | N | N | N | 1 |
| **4** | 2 | 5 | Head and Heart | Brain Development Course | 0 | 0 | 0 | N | N | N | N | N | N | N | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **112** | 34 | 113 | Green Protein | Plant-Based Protein | 0 | 0 | 0 | N | N | N | N | N | N | N | 1 |
| **113** | 34 | 114 | On2Cook | Fastest Cooking Device | 0 | 0 | 0 | N | N | N | N | N | N | N | 1 |
| **114** | 35 | 115 | Jain Shikanji | Lemonade | 40 | 0 | 30.00% | Y | Y | N | Y | N | Y | N | 1 |
| **115** | 35 | 116 | Woloo | Washroom Finder | 0 | 0 | 0 | N | N | N | N | N | N | N | 1 |

```
df.duplicated()
```

| | 0 |
|---|---|
| **0** | False |
| **1** | False |
| **2** | False |
| **3** | False |
| **4** | False |
| **...** | ... |
| **112** | False |
| **113** | False |
| **114** | False |
| **115** | False |
| **116** | False |

117 rows × 1 columns

```
df.describe()
```

| | Episode Number | Pitch Number | Investment Amount (In Lakhs INR) | Debt (In lakhs INR) | Season |
|---|---|---|---|---|---|
| **count** | 117.000000 | 117.000000 | 117.000000 | 117.000000 | 117.0 |
| **mean** | 18.735043 | 59.000000 | 21.683761 | 2.572650 | 1.0 |
| **std** | 10.070778 | 33.919021 | 26.067766 | 11.544753 | 0.0 |
| **min** | 1.000000 | 1.000000 | 0.000000 | 0.000000 | 1.0 |
| **25%** | 10.000000 | 30.000000 | 0.000000 | 0.000000 | 1.0 |
| **50%** | 19.000000 | 59.000000 | 10.000000 | 0.000000 | 1.0 |
| **75%** | 27.000000 | 88.000000 | 50.000000 | 0.000000 | 1.0 |
| **max** | 35.000000 | 117.000000 | 80.000000 | 99.000000 | 1.0 |

```
df.columns = df.columns.str.strip()
```

```
investor_columns = ['Anupam', 'Ashneer', 'Namita', 'Aman', 'Peyush', 'Vineeta', 'Ghazal']
for col in investor_columns:
    df[col] = df[col].apply(lambda x: 1 if x == 'Y' else 0)
```

```
df['Total Investors'] = df[investor_columns].sum(axis=1)
```

```
features = ['Equity', 'Debt (In lakhs INR)', 'Season', 'Total Investors']
target = 'Investment Amount (In Lakhs INR)'
```

```
df['Equity'] = df['Equity'].str.rstrip('%').astype(float)
```

```
X = df[features]
Y = df[target]
```

```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=42)
```

```
model = LinearRegression()
```

```
model.fit(X_train, Y_train)
```

```
LinearRegression()
```

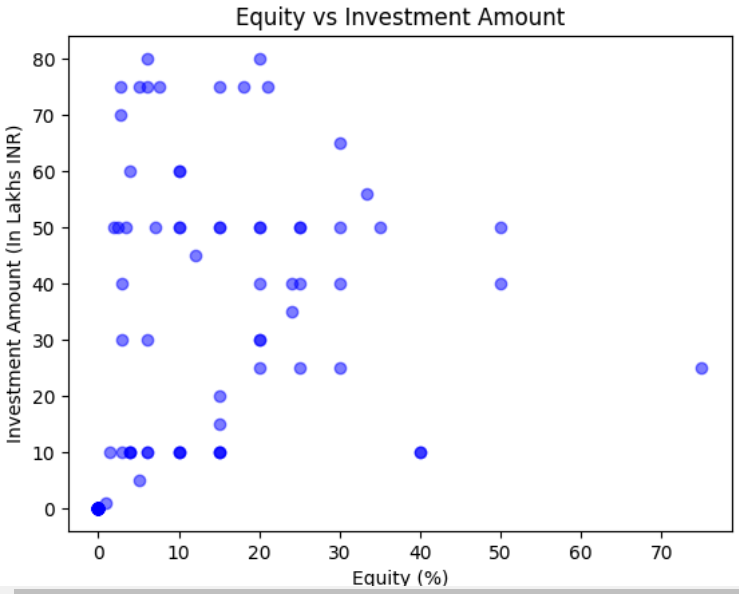
```
y_pred = model.predict(X_test)
```

```
mse = mean_squared_error(Y_test, y_pred)
r2 = r2_score(Y_test, y_pred)
```

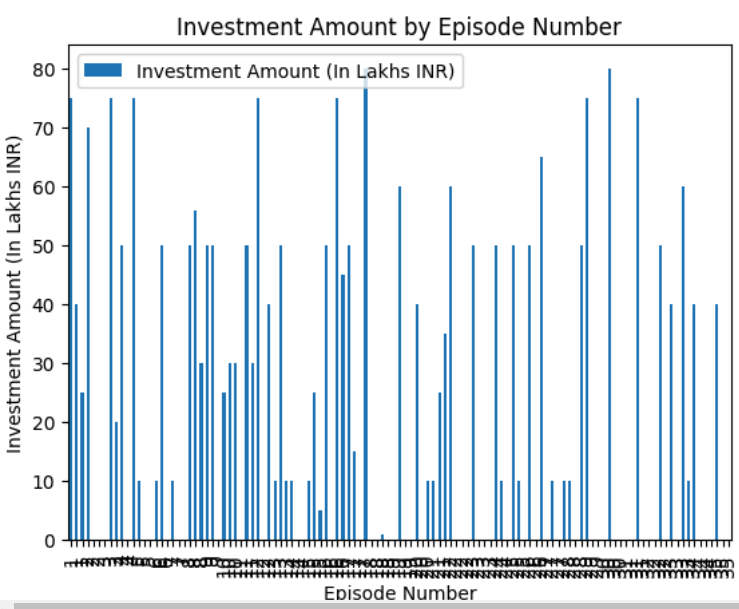
```
print(f"Mean Squared Error: {mse}")
print(f"R-squared: {r2}")
```

```
Mean Squared Error: 545.2947471245842
R-squared: 0.1725224939899611
```

```
plt.scatter(df['Equity'], df[target], alpha=0.5, color='blue')
plt.title('Equity vs Investment Amount')
plt.xlabel('Equity (%)')
plt.ylabel('Investment Amount (In Lakhs INR)')
plt.show()
```




```
df.plot(kind='bar',x='Episode Number', y='Investment Amount (In Lakhs INR)' )
plt.title('Investment Amount by Episode Number'),
plt.xlabel('Episode Number'),
plt.ylabel('Investment Amount (In Lakhs INR)')
plt.show()
```




```
df.plot(kind='hist', y='Investment Amount (In Lakhs INR)', width=0.50, color='r', x='Aman')
```

```
<Axes: ylabel='Frequency'>
```