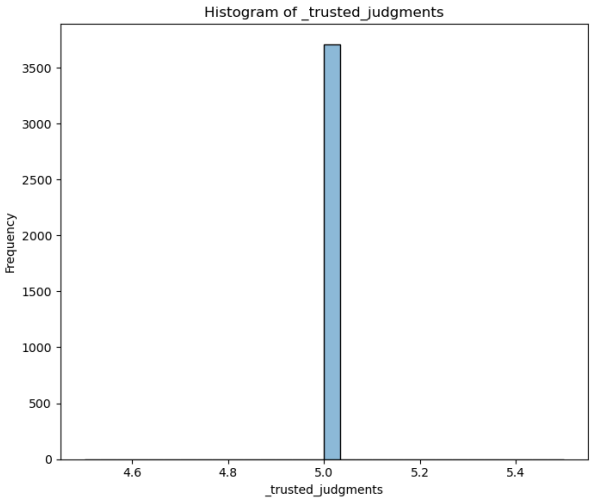
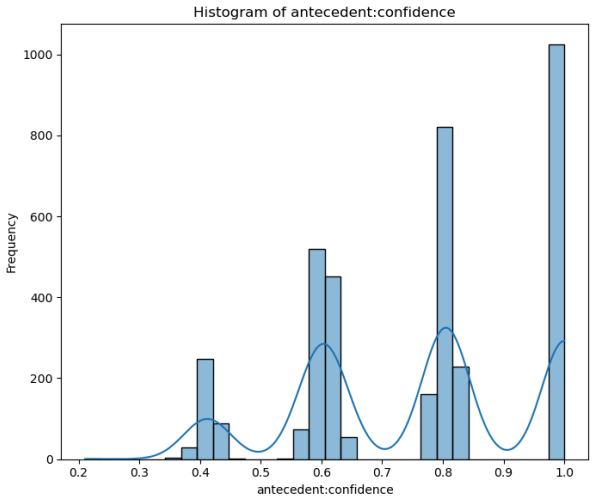
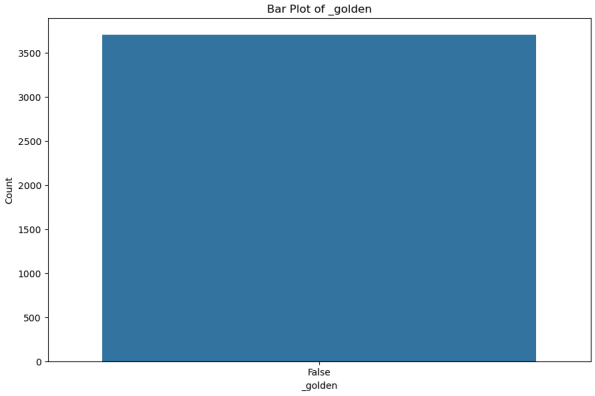
In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
file_path = '/Users/pavanijain/Desktop/Football-Scenarios-DFE-832307.csv'
data = pd.read_csv(file_path)

# Display sample dates to determine the format
print("Sample dates:")
print(data['_last_judgment_at'].head(10))

# Convert '_last_judgment_at' to datetime with a specific format if known
try:
    data['_last_judgment_at'] = pd.to_datetime(data['_last_judgment_at'], fo
except ValueError:
    data['_last_judgment_at'] = pd.to_datetime(data['_last_judgment_at'], er

# Handle missing values: Drop rows where '_last_judgment_at' is missing
data.dropna(subset=['_last_judgment_at'], inplace=True)

# Drop 'antecedent_gold' column if not needed
if 'antecedent_gold' in data.columns:
    data.drop(columns=['antecedent_gold'], inplace=True)

# Display basic information
print("Data Info:")
data_info = data.info()

print("\nSummary Statistics:")
summary_statistics = data.describe(include='all')

print("\nUnique Values:")
unique_values = {column: data[column].nunique() for column in data.columns}
print(unique_values)

# Plot histograms for numerical columns
numerical_columns = ['_trusted_judgments', 'antecedent:confidence']

plt.figure(figsize=(14, 6))
for i, column in enumerate(numerical_columns, 1):
    plt.subplot(1, 2, i)
    sns.histplot(data[column], kde=True, bins=30)
    plt.title(f'Histogram of {column}')
    plt.xlabel(column)
    plt.ylabel('Frequency')
plt.tight_layout()
plt.show()

# Plot bar plots for categorical columns
categorical_columns = ['_golden', '_unit_state', 'antecedent']

plt.figure(figsize=(18, 12))
for i, column in enumerate(categorical_columns, 1):
    plt.subplot(2, 2, i)
    sns.countplot(x=column, data=data)
    plt.title(f'Bar Plot of {column}')
    plt.xlabel(column)
    plt.ylabel('Count')
plt.tight_layout()
plt.show()
```

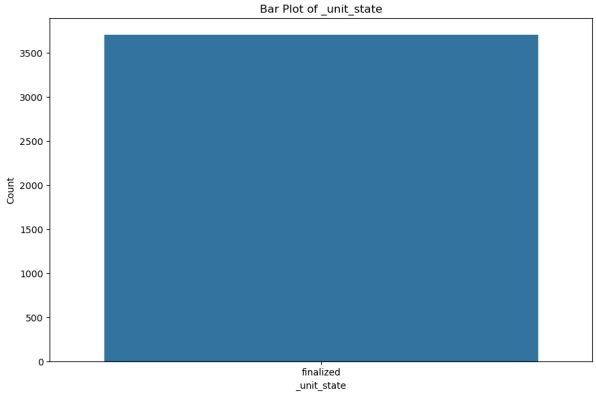
```
# Select only numeric columns for the correlation matrix
numeric_data = data[numerical_columns]

# Plot correlation heatmap
plt.figure(figsize=(10, 6))
correlation_matrix = numeric_data.corr()
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', vmin=-1, vmax=1
plt.title('Correlation Heatmap')
plt.show()

# Additional Analysis: Box plots for numerical features grouped by categori
plt.figure(figsize=(14, 12))
for i, column in enumerate(numerical_columns, 1):
    plt.subplot(2, 2, i)
    sns.boxplot(x='_unit_state', y=column, data=data)
    plt.title(f'Box Plot of {column} by _unit_state')
    plt.xlabel('_unit_state')
    plt.ylabel(column)

    # Adjusting limits if necessary to avoid identical limits
    y_min, y_max = data[column].min(), data[column].max()
    if y_min == y_max:
        y_min -= 0.1  # Adding a small margin
        y_max += 0.1
    plt.ylim(y_min, y_max)
plt.tight_layout()
plt.show()

# Time series analysis
data.set_index('_last_judgment_at', inplace=True)
plt.figure(figsize=(14, 6))
data['_trusted_judgments'].resample('M').mean().plot()
plt.title('Average Trusted Judgments Over Time')
plt.xlabel('Date')
plt.ylabel('Average Trusted Judgments')
plt.show()

# Reset index to original
data.reset_index(inplace=True)

# Print out the results for information, summary statistics, and unique valu
print(data_info)
print(summary_statistics)
print(unique_values)
```

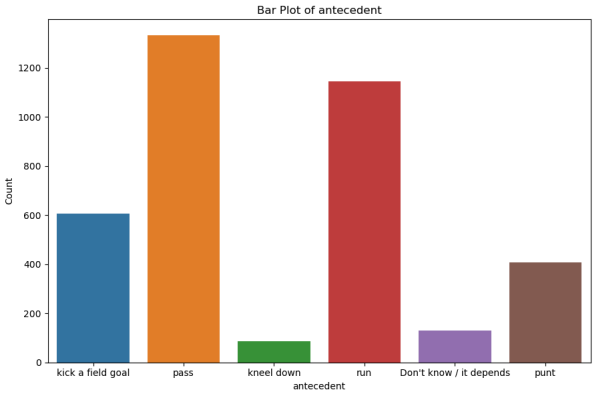
```
Sample dates:
0    11/20/15 20:20
1    11/18/15 21:59
2    11/20/15 22:43
3     11/19/15 7:41
4     11/21/15 8:01
5    11/21/15 18:25
6    11/20/15 20:31
7    11/20/15 19:10
8    11/20/15 18:20
9     11/20/15 9:58
Name: _last_judgment_at, dtype: object
Data Info:
<class 'pandas.core.frame.DataFrame'>
Index: 3706 entries, 0 to 3705
Data columns (total 13 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   _unit_id               3706 non-null   int64
 1   _golden                3706 non-null   bool
 2   _unit_state            3706 non-null   object
 3   _trusted_judgments     3706 non-null   int64
 4   _last_judgment_at      3706 non-null   datetime64[ns]
 5   antecedent             3706 non-null   object
 6   antecedent:confidence  3706 non-null   float64
 7   orig_antecedent        3706 non-null   object
 8   option1                3706 non-null   object
 9   option2                3706 non-null   object
 10  option3                3706 non-null   object
 11  option4                3706 non-null   object
 12  option5                3706 non-null   object
dtypes: bool(1), datetime64[ns](1), float64(1), int64(2), object(8)
memory usage: 380.0+ KB

Summary Statistics:

Unique Values:
{'_unit_id': 3706, '_golden': 1, '_unit_state': 1, '_trusted_judgments': 1,
'_last_judgment_at': 410, 'antecedent': 6, 'antecedent:confidence': 954, 'o
rig_antecedent': 3634, 'option1': 1, 'option2': 1, 'option3': 1, 'option4':
1, 'option5': 1}
```

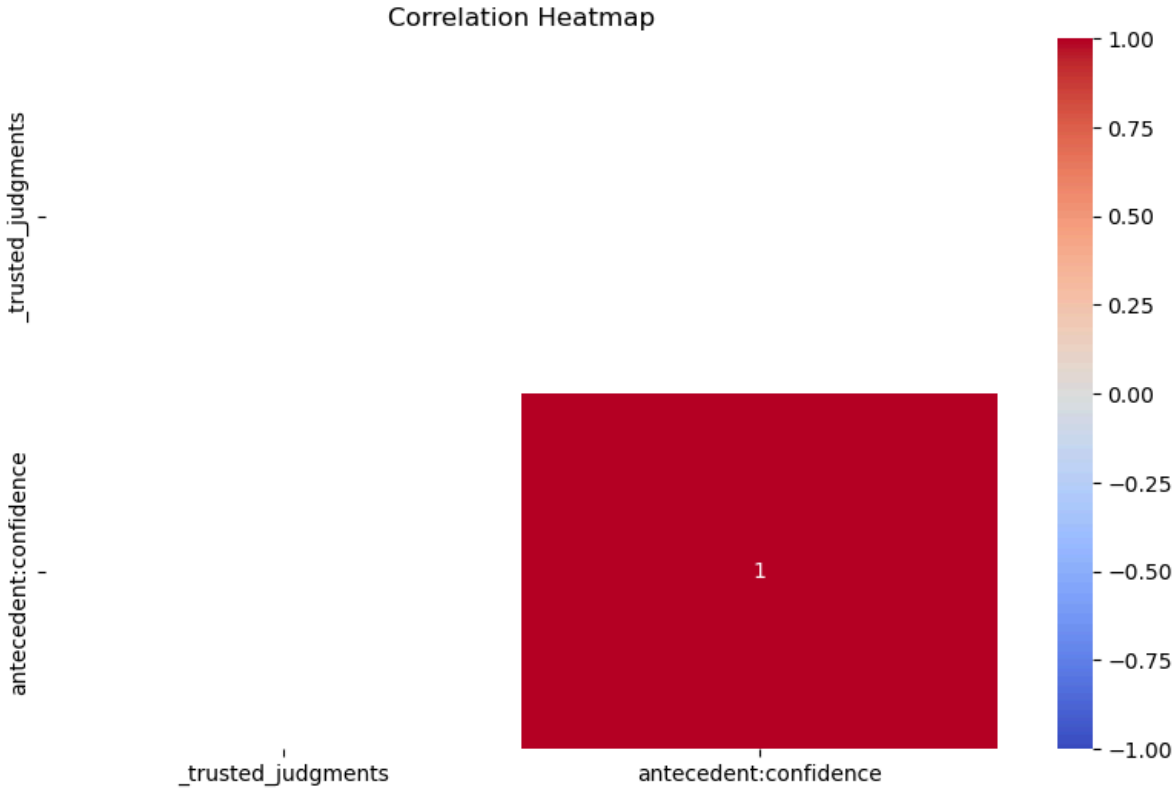


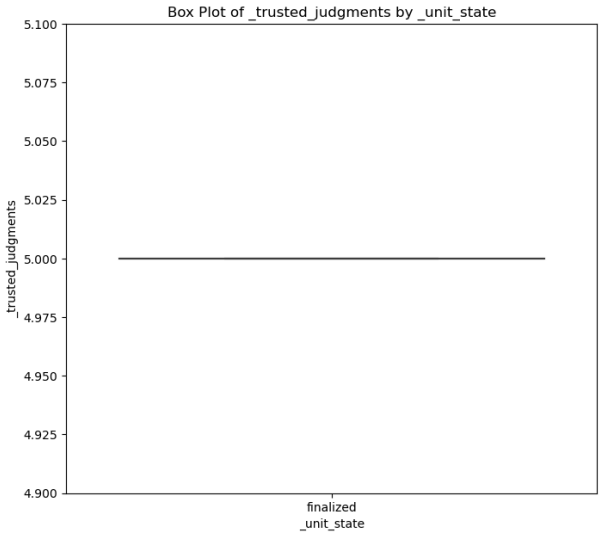



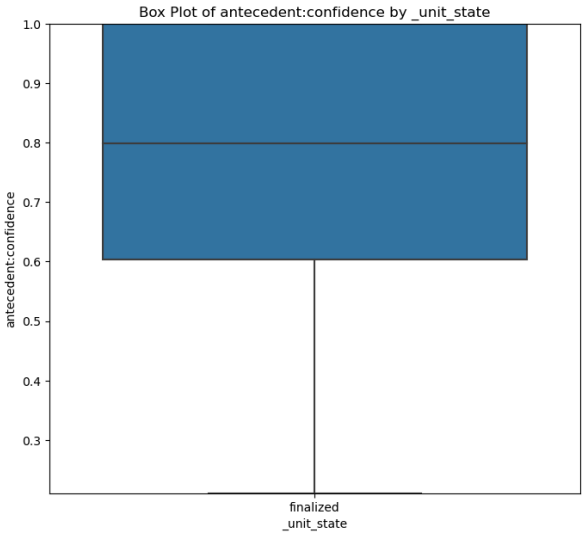

Bar Plot of _golden
Count
3500
3000
2500
2000
1500
1000
500
0
False
_golden

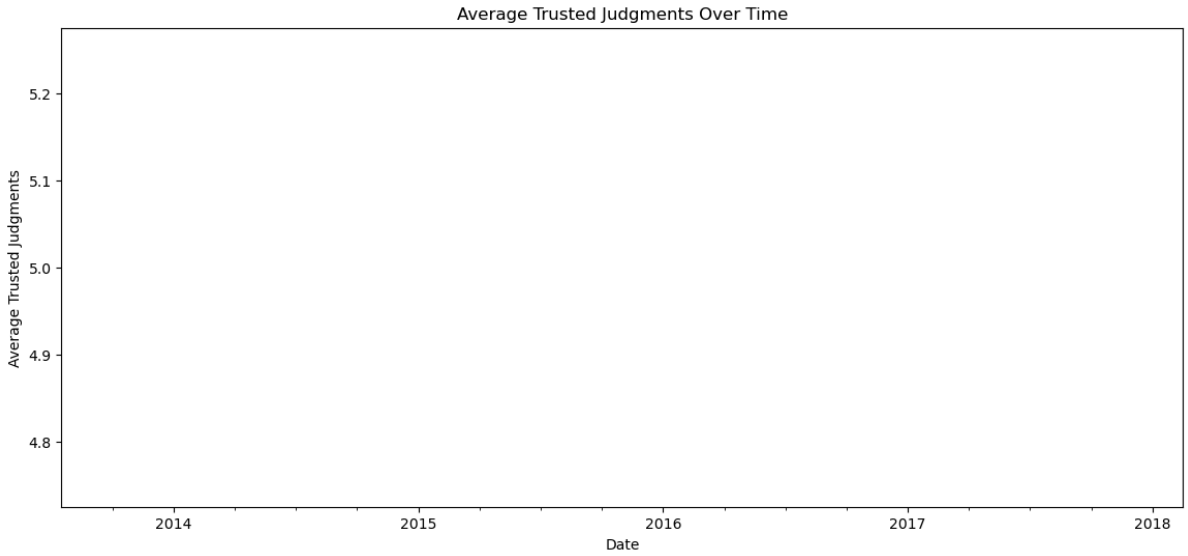

Bar Plot of _unit_state
Count
3500
3000
2500
2000
1500
1000
500
0
finalized
_unit_state


Bar Plot of antecedent
Count
1200
1000
800
600
400
200
0
kick a field goal
pass
kneel down
run
Don't know / it depends
punt
antecedent


Correlation Heatmap
_trusted_judgments
antecedent:confidence
1
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
_trusted_judgments
antecedent:confidence


Box Plot of _trusted_judgments by _unit_state
_trusted_judgments
5.100
5.075
5.050
5.025
5.000
4.975
4.950
4.925
4.900
finalized
_unit_state


Box Plot of antecedent:confidence by _unit_state
antecedent:confidence
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
finalized
_unit_state


Average Trusted Judgments Over Time
Average Trusted Judgments
5.2
5.1
5.0
4.9
4.8
2014
2015
2016
2017
2018
Date

```
None
           _unit_id _golden _unit_state  _trusted_judgments  \
count   3.706000e+03    3706        3706               3706.0
unique           NaN       1           1                  NaN
top              NaN   False   finalized                  NaN
freq             NaN    3706        3706                  NaN
mean    8.310075e+08     NaN         NaN                  5.0
min     8.310057e+08     NaN         NaN                  5.0
25%     8.310066e+08     NaN         NaN                  5.0
50%     8.310075e+08     NaN         NaN                  5.0
75%     8.310085e+08     NaN         NaN                  5.0
max     8.310094e+08     NaN         NaN                  5.0
std     1.075271e+03     NaN         NaN                  0.0

                    _last_judgment_at antecedent  antecedent:confidence  \
count                            3706       3706            3706.000000
unique                            NaN          6                    NaN
top                               NaN       pass                    NaN
freq                              NaN       1332                    NaN
mean    2015-11-20 09:13:37.026443520        NaN               0.759625
min               2015-11-18 20:40:00        NaN               0.210700
25%               2015-11-19 10:22:00        NaN               0.603325
50%               2015-11-20 16:03:00        NaN               0.798700
75%               2015-11-21 02:56:00        NaN               1.000000
max               2015-11-21 21:57:00        NaN               1.000000
std                               NaN        NaN               0.190261

                             orig_antecedent option1             option2
\
count                                   3706    3706                3706
unique                                  3634       1                   1
top      You are down by 3 points. Would you:    punt   kick a field goal
freq                                      13    3706                3706
mean                                     NaN     NaN                 NaN
min                                      NaN     NaN                 NaN
25%                                      NaN     NaN                 NaN
50%                                      NaN     NaN                 NaN
75%                                      NaN     NaN                 NaN
max                                      NaN     NaN                 NaN
std                                      NaN     NaN                 NaN

       option3 option4      option5
count     3706    3706         3706
unique       1       1            1
top        run    pass   kneel down
freq      3706    3706         3706
mean       NaN     NaN          NaN
min        NaN     NaN          NaN
25%        NaN     NaN          NaN
50%        NaN     NaN          NaN
75%        NaN     NaN          NaN
max        NaN     NaN          NaN
std        NaN     NaN          NaN
{'_unit_id': 3706, '_golden': 1, '_unit_state': 1, '_trusted_judgments': 1,
'_last_judgment_at': 410, 'antecedent': 6, 'antecedent:confidence': 954, 'o
rig_antecedent': 3634, 'option1': 1, 'option2': 1, 'option3': 1, 'option4':
1, 'option5': 1}
```

In [ ]: