Externship_Webacy_Frequency_and_Correlation_Analysis

June 28, 2024

1 Smart Contract Frequency and Correlation Analysis of Security Audits By Alessandra Adina

Objectives: 1. **Identify and classify security risks:** Learn to spot common and critical vulnerabilities in smart contracts. 2. **Analyze data in detail:** Use statistical methods to explore the relationships and correlations between different risk factors. 3. **Visualize data clearly:** Use tools like Excel and Python to make complex information easy to understand and interesting. 4. **Create actionable security insights:** Turn your analysis into recommendations for improving smart contract security.

1.1 Analysis of Risk Tag Frequencies in Smart Contract Security Audits

```
[42]: # Import necessary libraries
      import pandas as pd
      import seaborn as sns
      import matplotlib.pyplot as plt

      # Display plots inline
      %matplotlib inline

      print("Libraries imported successfully!")
```

Libraries imported successfully!

```
[43]: !gdown 15Y1WnWIhaa5CKvjNR2S8SfbX61Z1PzuK
```

Downloading…
From: https://drive.google.com/uc?id=15Y1WnWIhaa5CKvjNR2S8SfbX61Z1PzuK
To: /content/compiled_risk_data.xlsx
100% 310k/310k [00:00<00:00, 5.13MB/s]

```
[48]: # Loading the dataset

      df = pd.read_excel('compiled_risk_data.xlsx')

      # Display the first five rows of the dataframe
      df.head()
```

```
[48]:       project_name                    Smart contract address  \
      0  Data Analytics    384571416209d08623c6ace9422613fc8970475d
      1  Data Analytics    0xAb5801a7D398351b8bE11C439e05C5B3259ae9B
      2  Data Analytics    0x4B20993Bc481177ec7E8f571ceCaE8A9e22C02db
      3  Data Analytics    0x78731D3Ca6b7E34aC0F824c42a7cC18A495cabaB
      4  Data Analytics    0x617F2E2fD72FD9D5503197092aC168c91465E7f2


                                           Blog post link  \
      0  https://chainsecurity.com/security-audit/circl…
      1  https://stackoverflow.com/questions/75030483/w…
      2  https://stackoverflow.com/questions/71115106/s…
      3  https://stackoverflow.com/questions/75030483/w…
      4  https://stackoverflow.com/questions/69466137/h…


                                           Audit website     Chain  \
      0                       https://chainsecurity.com   Ethereum
      1  https://studygroup.moralis.io/t/compilation-er…   Ethereum
      2  https://ethereum.stackexchange.com/questions/1…   Ethereum
      3  https://studygroup.moralis.io/t/compilation-er…   Ethereum
      4  https://ethereum.stackexchange.com/questions/1…   Ethereum


         Is_closed_source  hidden_owner  anti_whale_modifiable  Is_anti_whale  \
      0             False         False                  False          False
      1             False         False                   True           True
      2              True         False                  False           True
      3              True         False                  False          False
      4              True         False                   True          False


         Is_honeypot  …  centralized_risk_high  centralized_risk_low  \
      0        False  …                  False                 False
      1        False  …                  False                  True
      2        False  …                  False                  True
      3         True  …                  False                  True
      4         True  …                  False                 False


         event_setter  external_dependencies  immutable_states  \
      0          True                   True              True
      1         False                   True             False
      2         False                   True             False
      3         False                  False              True
      4          True                   True             False


         reentrancy_without_eth_transfer  incorrect_inheritance_order  \
      0                             True                        False
      1                            False                         True
      2                            False                        False
      3                             True                        False
```

```
4                          True                         False

    shadowing_local  events_maths  \
0            False         False
1            False          True
2             True          True
3            False          True
4            False          True


        Summary/rationale of risk tags marked true
0  Bad Contract: Assigned for flaws that indicate…
1  Summary/rationale of risk tags marked true: Th…
2                                              NaN
3                                              NaN
4  The smart contract is marked with risks such a…

[5 rows x 38 columns]
```

Since our goal is to "document the occurrence rates of each risk tag across the dataset, providing a clear quantification of the most prevalent vulnerabilities", the first 4 columns are irrelevant, so we can drop them.

Let us also go ahead and set the first row as index, to make our analysis easier.

```python
[49]: # Drop the first 5 columns
      df = df.iloc[:, 5:]

      # Drop the summary column (last column)
      df = df.iloc[:, :-1]

      # Set the first row as index
      df = df.set_index(df.columns[0])

      # Reset columns index
      df.reset_index(inplace=True)

      # Display the first five rows of the dataframe
      df.head()
```

```
[49]:    Is_closed_source  hidden_owner  anti_whale_modifiable  Is_anti_whale  \
      0             False         False                  False          False
      1             False         False                   True           True
      2              True         False                  False           True
      3              True         False                  False          False
      4              True         False                   True          False

         Is_honeypot  buy_tax  sell_tax  slippage_modifiable  Is_blacklisted  \
      0        False    False     False                False           False
```

```
    1       False     True     False                    False          False
    2       False     False    False                    True           False
    3       True      False    True                     False          False
    4       True      False    False                    False          True

       can_take_back_ownership  …  centralized_risk_medium  \
    0                    False  …                     True
    1                    False  …                     False
    2                     True  …                     False
    3                    False  …                     True
    4                    False  …                     False

       centralized_risk_high  centralized_risk_low  event_setter  \
    0                  False                 False          True
    1                  False                  True         False
    2                  False                  True         False
    3                  False                  True         False
    4                  False                 False          True

       external_dependencies  immutable_states  reentrancy_without_eth_transfer  \
    0                   True              True                             True
    1                   True             False                            False
    2                   True             False                            False
    3                  False              True                             True
    4                   True             False                             True

       incorrect_inheritance_order  shadowing_local  events_maths
    0                        False            False         False
    1                         True            False          True
    2                        False             True          True
    3                        False            False          True
    4                        False            False          True

    [5 rows x 32 columns]
```

Now we can do a full frequency analysis on the columns.

```python
[51]: # Calculating the frequency of 'True' in each column
      frequencies = df.apply(lambda x: x.value_counts()).loc[True]
      frequencies = frequencies.fillna(0)  # Replace NaN with 0 for any column that
       ↪may not have True values
      frequencies
```

```
[51]: Is_closed_source          146
      hidden_owner              164
      anti_whale_modifiable     122
      Is_anti_whale             155
```

```
Is_honeypot                        94
buy_tax                           128
sell_tax                          126
slippage_modifiable               149
Is_blacklisted                     81
can_take_back_ownership           194
owner_change_balance              222
is_airdrop_scam                    69
selfdestruct                      116
trust_list                        149
is_whitelisted                    109
is_fake_token                      90
illegal_unicode                    62
exploitation                      468
bad_contract                      373
reusing_state_variable            124
encode_packed_collision            81
encode_packed_parameters           87
centralized_risk_medium           283
centralized_risk_high             205
centralized_risk_low              190
event_setter                      149
external_dependencies             316
immutable_states                  154
reentrancy_without_eth_transfer   199
incorrect_inheritance_order       100
shadowing_local                    88
events_maths                      149
Name: True, dtype: int64
```

Now that we have the frequencies. We can also visualise this using a barchart.

```
[52]:  # Visualizing the frequencies using a bar chart
       sns.set_style("whitegrid")
       plt.figure(figsize=(12, 8))
       sns.barplot(x=frequencies.index, y=frequencies.values, palette='viridis')
       plt.title('Frequency of True Values for Each Risk Tag')
       plt.xlabel('Risk Tags')
       plt.ylabel('Frequency of True')
       plt.xticks(rotation=90)
       plt.show()
```

```
<ipython-input-52-5a383aa58995>:4: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.
```

```
sns.barplot(x=frequencies.index, y=frequencies.values, palette='viridis')
```

Frequency of True Values for Each Risk Tag
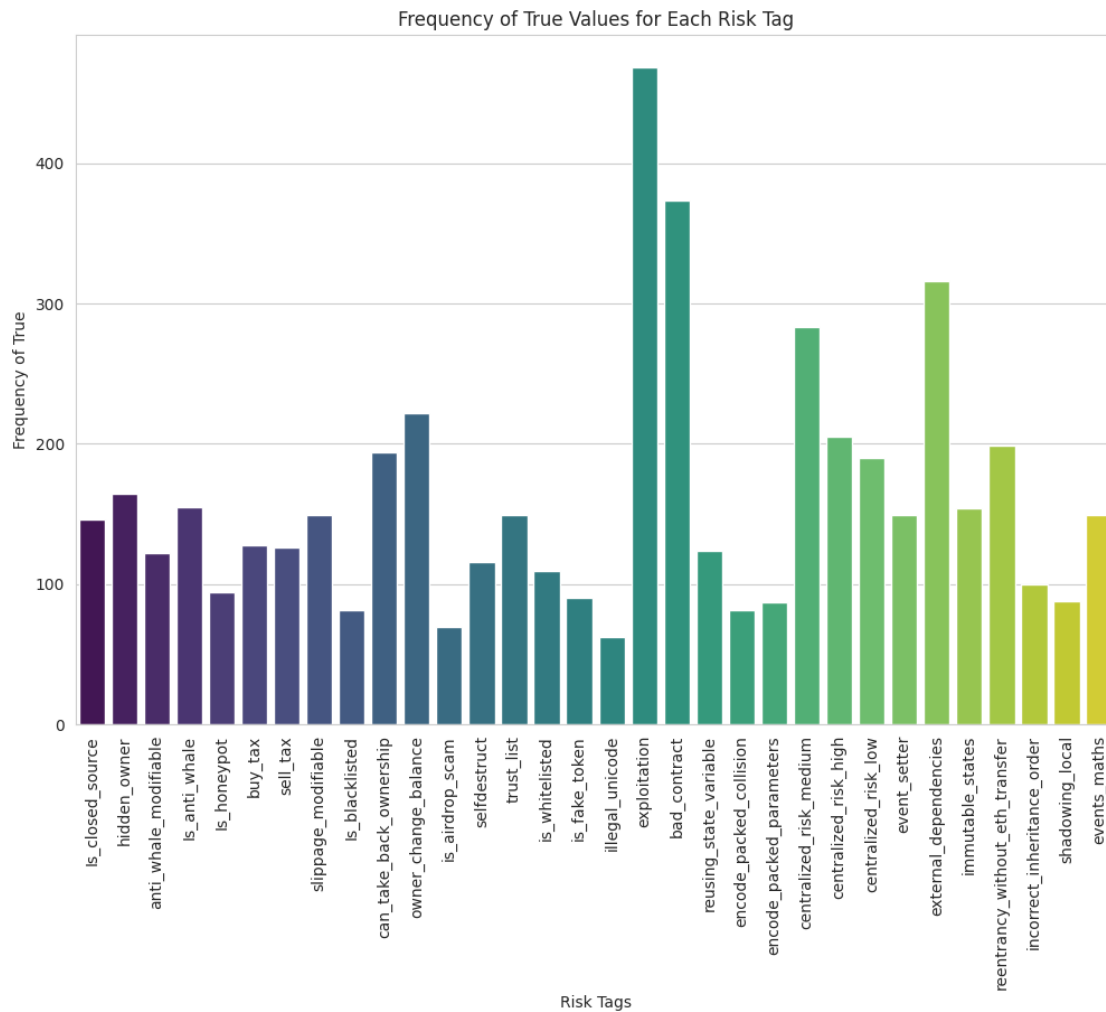
1.2 Correlation Analysis

```
[53]:  # Import necessary libraries that weren't imported earlier
       import numpy as np
       import scipy.stats

       print("Libraries imported successfully!")
```

Libraries imported successfully!

To calculate the Phi coefficient, which is suitable for pairs of binary variables, we first need to establish a function that can handle this calculation:

```
[54]: def phi_coefficient(x, y):
          """Calculate the Phi coefficient for two binary variables."""
          # Create a contingency table
          contingency_table = pd.crosstab(x, y)
          # Calculate the phi coefficient from the contingency table
          chi2 = scipy.stats.chi2_contingency(contingency_table, correction=False)[0]
          n = np.sum(np.sum(contingency_table))
          phi = np.sqrt(chi2 / n)
          return phi


      # Example calculation between two risk tags
      phi = phi_coefficient(df['Is_honeypot'], df['anti_whale_modifiable'])
      print(f"Phi Coefficient between 'Is_honeypot' and 'anti_whale_modifiable':␣
       ↪{phi}")
```

Phi Coefficient between 'Is_honeypot' and 'anti_whale_modifiable':
0.43014356785902874

Phi value close to 0 indicates no correlation between the two columns.

Note: Phi values range from -1 to 1. A negative value of Phi indicates that the variables are inversely related, or when one variable increases, the other decreases. On the other hand, positive values indicate that when one variable increases, so does the other.

```
[55]: # Create a DataFrame to store Phi coefficients
      phi_matrix = pd.DataFrame(index=df.columns, columns=df.columns)

      # Calculate Phi coefficient for each pair of binary variables
      for var1 in df.columns:
          for var2 in df.columns:
              phi_matrix.loc[var1, var2] = phi_coefficient(df[var1], df[var2])

      print("Phi coefficients calculated for all pairs of variables:")
      phi_matrix
```

Phi coefficients calculated for all pairs of variables:

[55]:

	Is_closed_source	hidden_owner
Is_closed_source	1.0	0.377241
hidden_owner	0.377241	1.0
anti_whale_modifiable	0.330545	0.396169
Is_anti_whale	0.295222	0.35796
Is_honeypot	0.378356	0.318871
buy_tax	0.325367	0.476409
sell_tax	0.380376	0.441975
slippage_modifiable	0.392646	0.400521
Is_blacklisted	0.350922	0.30169
can_take_back_ownership	0.268073	0.361349
owner_change_balance	0.28307	0.316426

is_airdrop_scam	0.417823	0.333385
selfdestruct	0.327478	0.287755
trust_list	0.353457	0.273593
is_whitelisted	0.398871	0.330425
is_fake_token	0.361777	0.312184
illegal_unicode	0.403646	0.295686
exploitation	0.105907	0.123116
bad_contract	0.171161	0.194702
reusing_state_variable	0.308981	0.342472
encode_packed_collision	0.453591	0.340811
encode_packed_parameters	0.38133	0.340238
centralized_risk_medium	0.154701	0.207843
centralized_risk_high	0.210932	0.32317
centralized_risk_low	0.188913	0.179091
event_setter	0.353457	0.295992
external_dependencies	0.141124	0.138936
immutable_states	0.258405	0.234851
reentrancy_without_eth_transfer	0.267738	0.259852
incorrect_inheritance_order	0.341803	0.328749
shadowing_local	0.368141	0.299391
events_maths	0.282918	0.206395

	anti_whale_modifiable	Is_anti_whale \
Is_closed_source	0.330545	0.295222
hidden_owner	0.396169	0.35796
anti_whale_modifiable	1.0	0.522184
Is_anti_whale	0.522184	1.0
Is_honeypot	0.430144	0.361724
buy_tax	0.467261	0.504441
sell_tax	0.508889	0.485596
slippage_modifiable	0.604309	0.480527
Is_blacklisted	0.343386	0.345546
can_take_back_ownership	0.337212	0.271046
owner_change_balance	0.391569	0.290203
is_airdrop_scam	0.385865	0.390541
selfdestruct	0.406115	0.285672
trust_list	0.282547	0.335307
is_whitelisted	0.347463	0.346104
is_fake_token	0.358883	0.38389
illegal_unicode	0.365253	0.365163
exploitation	0.139525	0.104054
bad_contract	0.192207	0.144392
reusing_state_variable	0.413718	0.408641
encode_packed_collision	0.465394	0.395606
encode_packed_parameters	0.432474	0.403681
centralized_risk_medium	0.234878	0.214742
centralized_risk_high	0.343256	0.274977

```
centralized_risk_low                                0.151742      0.214919
event_setter                                        0.367221      0.327664
external_dependencies                               0.216158      0.163063
immutable_states                                    0.290703      0.363041
reentrancy_without_eth_transfer                      0.30709      0.270344
incorrect_inheritance_order                         0.391393      0.344022
shadowing_local                                     0.396996      0.380936
events_maths                                        0.324884      0.304734


                                 Is_honeypot   buy_tax  sell_tax  \
Is_closed_source                    0.378356  0.325367  0.380376
hidden_owner                        0.318871  0.476409  0.441975
anti_whale_modifiable               0.430144  0.467261  0.508889
Is_anti_whale                       0.361724  0.504441  0.485596
Is_honeypot                              1.0  0.385562   0.42062
buy_tax                             0.385562       1.0  0.706026
sell_tax                             0.42062  0.706026       1.0
slippage_modifiable                 0.382178  0.468933  0.466078
Is_blacklisted                      0.473817  0.342314  0.379094
can_take_back_ownership             0.275998  0.344645  0.274693
owner_change_balance                0.323701  0.332415  0.358952
is_airdrop_scam                     0.470526  0.431949  0.459961
selfdestruct                        0.487618  0.308689  0.377891
trust_list                          0.372668  0.369426  0.340841
is_whitelisted                      0.464232  0.401065  0.443969
is_fake_token                       0.525426  0.387748  0.402548
illegal_unicode                     0.503183  0.378111  0.369726
exploitation                        0.136852  0.121866  0.087132
bad_contract                        0.164641  0.145937  0.181274
reusing_state_variable               0.43562  0.399051  0.367706
encode_packed_collision             0.498733  0.396625  0.455652
encode_packed_parameters             0.47644  0.418506  0.380726
centralized_risk_medium              0.19862  0.271892  0.309845
centralized_risk_high               0.220442  0.306085  0.325616
centralized_risk_low                0.229566  0.178313  0.204843
event_setter                        0.353648  0.303088  0.332492
external_dependencies               0.200276  0.188246  0.199511
immutable_states                    0.401028  0.359557  0.339642
reentrancy_without_eth_transfer     0.328827  0.277935  0.334513
incorrect_inheritance_order         0.457276  0.377879   0.35248
shadowing_local                     0.436948  0.425561  0.430177
events_maths                        0.382178  0.327965  0.307444


                                 slippage_modifiable  Is_blacklisted  \
Is_closed_source                            0.392646        0.350922
hidden_owner                                0.400521         0.30169
anti_whale_modifiable                       0.604309        0.343386
```

	slippage_modifiable	Is_blacklisted
Is_anti_whale	0.480527	0.345546
Is_honeypot	0.382178	0.473817
buy_tax	0.468933	0.342314
sell_tax	0.466078	0.379094
slippage_modifiable	1.0	0.315123
Is_blacklisted	0.315123	1.0
can_take_back_ownership	0.352786	0.25251
owner_change_balance	0.356143	0.299934
is_airdrop_scam	0.434104	0.457952
selfdestruct	0.356551	0.401474
trust_list	0.370571	0.315123
is_whitelisted	0.366104	0.477019
is_fake_token	0.346589	0.46162
illegal_unicode	0.352127	0.489304
exploitation	0.211241	0.129297
bad_contract	0.225801	0.2089
reusing_state_variable	0.387554	0.416417
encode_packed_collision	0.416914	0.479981
encode_packed_parameters	0.415123	0.484584
centralized_risk_medium	0.233869	0.151716
centralized_risk_high	0.307745	0.32934
centralized_risk_low	0.204737	0.165155
event_setter	0.378341	0.355839
external_dependencies	0.217158	0.204795
immutable_states	0.321931	0.317108
reentrancy_without_eth_transfer	0.351499	0.273813
incorrect_inheritance_order	0.354841	0.394809
shadowing_local	0.431227	0.455437
events_maths	0.316176	0.325302

	can_take_back_ownership	… \
Is_closed_source	0.268073	…
hidden_owner	0.361349	…
anti_whale_modifiable	0.337212	…
Is_anti_whale	0.271046	…
Is_honeypot	0.275998	…
buy_tax	0.344645	…
sell_tax	0.274693	…
slippage_modifiable	0.352786	…
Is_blacklisted	0.25251	…
can_take_back_ownership	1.0	…
owner_change_balance	0.467797	…
is_airdrop_scam	0.263413	…
selfdestruct	0.376362	…
trust_list	0.283002	…
is_whitelisted	0.26894	…
is_fake_token	0.339952	…

illegal_unicode	0.289802	…
exploitation	0.202925	…
bad_contract	0.200978	…
reusing_state_variable	0.317045	…
encode_packed_collision	0.32564	…
encode_packed_parameters	0.332265	…
centralized_risk_medium	0.250195	…
centralized_risk_high	0.353414	…
centralized_risk_low	0.166029	…
event_setter	0.269045	…
external_dependencies	0.163273	…
immutable_states	0.314297	…
reentrancy_without_eth_transfer	0.233799	…
incorrect_inheritance_order	0.367523	…
shadowing_local	0.311373	…
events_maths	0.234152	…

	centralized_risk_medium	centralized_risk_high \
Is_closed_source	0.154701	0.210932
hidden_owner	0.207843	0.32317
anti_whale_modifiable	0.234878	0.343256
Is_anti_whale	0.214742	0.274977
Is_honeypot	0.19862	0.220442
buy_tax	0.271892	0.306085
sell_tax	0.309845	0.325616
slippage_modifiable	0.233869	0.307745
Is_blacklisted	0.151716	0.32934
can_take_back_ownership	0.250195	0.353414
owner_change_balance	0.288234	0.340004
is_airdrop_scam	0.258806	0.3186
selfdestruct	0.243888	0.283406
trust_list	0.25213	0.314575
is_whitelisted	0.221511	0.301593
is_fake_token	0.240842	0.299445
illegal_unicode	0.234285	0.28747
exploitation	0.235642	0.228418
bad_contract	0.191285	0.247368
reusing_state_variable	0.183705	0.278933
encode_packed_collision	0.207531	0.320393
encode_packed_parameters	0.212005	0.308982
centralized_risk_medium	1.0	0.256407
centralized_risk_high	0.256407	1.0
centralized_risk_low	0.164269	0.113576
event_setter	0.179086	0.259931
external_dependencies	0.268048	0.22609
immutable_states	0.192864	0.243302
reentrancy_without_eth_transfer	0.240715	0.192386

incorrect_inheritance_order	0.225362	0.327165
shadowing_local	0.254973	0.297145
events_maths	0.130389	0.212117

	centralized_risk_low	event_setter \
Is_closed_source	0.188913	0.353457
hidden_owner	0.179091	0.295992
anti_whale_modifiable	0.151742	0.367221
Is_anti_whale	0.214919	0.327664
Is_honeypot	0.229566	0.353648
buy_tax	0.178313	0.303088
sell_tax	0.204843	0.332492
slippage_modifiable	0.204737	0.378341
Is_blacklisted	0.165155	0.355839
can_take_back_ownership	0.166029	0.269045
owner_change_balance	0.092467	0.349516
is_airdrop_scam	0.238306	0.379284
selfdestruct	0.202506	0.339237
trust_list	0.204737	0.300634
is_whitelisted	0.217967	0.348306
is_fake_token	0.196337	0.365989
illegal_unicode	0.231934	0.352127
exploitation	0.031002	0.173521
bad_contract	0.029736	0.225801
reusing_state_variable	0.186085	0.353924
encode_packed_collision	0.248104	0.406735
encode_packed_parameters	0.212966	0.395421
centralized_risk_medium	0.164269	0.179086
centralized_risk_high	0.113576	0.259931
centralized_risk_low	1.0	0.183629
event_setter	0.183629	1.0
external_dependencies	0.133498	0.205395
immutable_states	0.216707	0.421562
reentrancy_without_eth_transfer	0.127669	0.261675
incorrect_inheritance_order	0.21451	0.382586
shadowing_local	0.219189	0.411627
events_maths	0.155485	0.424966

	external_dependencies	immutable_states \
Is_closed_source	0.141124	0.258405
hidden_owner	0.138936	0.234851
anti_whale_modifiable	0.216158	0.290703
Is_anti_whale	0.163063	0.363041
Is_honeypot	0.200276	0.401028
buy_tax	0.188246	0.359557
sell_tax	0.199511	0.339642
slippage_modifiable	0.217158	0.321931

Is_blacklisted	0.204795	0.317108
can_take_back_ownership	0.163273	0.314297
owner_change_balance	0.22983	0.253097
is_airdrop_scam	0.199585	0.435534
selfdestruct	0.199601	0.364192
trust_list	0.181869	0.344923
is_whitelisted	0.265954	0.33914
is_fake_token	0.190735	0.37615
illegal_unicode	0.218767	0.389534
exploitation	0.272019	0.085353
bad_contract	0.217763	0.157769
reusing_state_variable	0.261857	0.410688
encode_packed_collision	0.251021	0.397421
encode_packed_parameters	0.28219	0.386116
centralized_risk_medium	0.268048	0.192864
centralized_risk_high	0.22609	0.243302
centralized_risk_low	0.133498	0.216707
event_setter	0.205395	0.421562
external_dependencies	1.0	0.257991
immutable_states	0.257991	1.0
reentrancy_without_eth_transfer	0.269151	0.258688
incorrect_inheritance_order	0.259617	0.400515
shadowing_local	0.234074	0.37308
events_maths	0.223039	0.360251

	reentrancy_without_eth_transfer \
Is_closed_source	0.267738
hidden_owner	0.259852
anti_whale_modifiable	0.30709
Is_anti_whale	0.270344
Is_honeypot	0.328827
buy_tax	0.277935
sell_tax	0.334513
slippage_modifiable	0.351499
Is_blacklisted	0.273813
can_take_back_ownership	0.233799
owner_change_balance	0.274503
is_airdrop_scam	0.345471
selfdestruct	0.276187
trust_list	0.240947
is_whitelisted	0.254424
is_fake_token	0.324415
illegal_unicode	0.314778
exploitation	0.286489
bad_contract	0.235462
reusing_state_variable	0.354525
encode_packed_collision	0.355271

```
encode_packed_parameters                                    0.334253
centralized_risk_medium                                     0.240715
centralized_risk_high                                       0.192386
centralized_risk_low                                        0.127669
event_setter                                                0.261675
external_dependencies                                       0.269151
immutable_states                                            0.258688
reentrancy_without_eth_transfer                                  1.0
incorrect_inheritance_order                                 0.343682
shadowing_local                                             0.357069
events_maths                                                0.268585
```

	incorrect_inheritance_order	shadowing_local \
Is_closed_source	0.341803	0.368141
hidden_owner	0.328749	0.299391
anti_whale_modifiable	0.391393	0.396996
Is_anti_whale	0.344022	0.380936
Is_honeypot	0.457276	0.436948
buy_tax	0.377879	0.425561
sell_tax	0.35248	0.430177
slippage_modifiable	0.354841	0.431227
Is_blacklisted	0.394809	0.455437
can_take_back_ownership	0.367523	0.311373
owner_change_balance	0.352501	0.335381
is_airdrop_scam	0.478723	0.545393
selfdestruct	0.426416	0.465743
trust_list	0.336344	0.401827
is_whitelisted	0.392156	0.428966
is_fake_token	0.493708	0.461834
illegal_unicode	0.498375	0.494373
exploitation	0.187154	0.185659
bad_contract	0.226719	0.198358
reusing_state_variable	0.47685	0.509112
encode_packed_collision	0.479611	0.57097
encode_packed_parameters	0.469444	0.521722
centralized_risk_medium	0.225362	0.254973
centralized_risk_high	0.327165	0.297145
centralized_risk_low	0.21451	0.219189
event_setter	0.382586	0.411627
external_dependencies	0.259617	0.234074
immutable_states	0.400515	0.37308
reentrancy_without_eth_transfer	0.343682	0.357069
incorrect_inheritance_order	1.0	0.465934
shadowing_local	0.465934	1.0
events_maths	0.410331	0.401827

```
                         events_maths
```

```
Is_closed_source              0.282918
hidden_owner                  0.206395
anti_whale_modifiable         0.324884
Is_anti_whale                 0.304734
Is_honeypot                   0.382178
buy_tax                       0.327965
sell_tax                      0.307444
slippage_modifiable           0.316176
Is_blacklisted                0.325302
can_take_back_ownership       0.234152
owner_change_balance          0.289868
is_airdrop_scam               0.357356
selfdestruct                  0.313266
trust_list                    0.300634
is_whitelisted                0.366104
is_fake_token                 0.356289
illegal_unicode               0.329073
exploitation                  0.184298
bad_contract                  0.186435
reusing_state_variable        0.345516
encode_packed_collision       0.396555
encode_packed_parameters      0.395421
centralized_risk_medium       0.130389
centralized_risk_high         0.212117
centralized_risk_low          0.155485
event_setter                  0.424966
external_dependencies         0.223039
immutable_states              0.360251
reentrancy_without_eth_transfer  0.268585
incorrect_inheritance_order   0.410331
shadowing_local               0.401827
events_maths                       1.0

[32 rows x 32 columns]
```
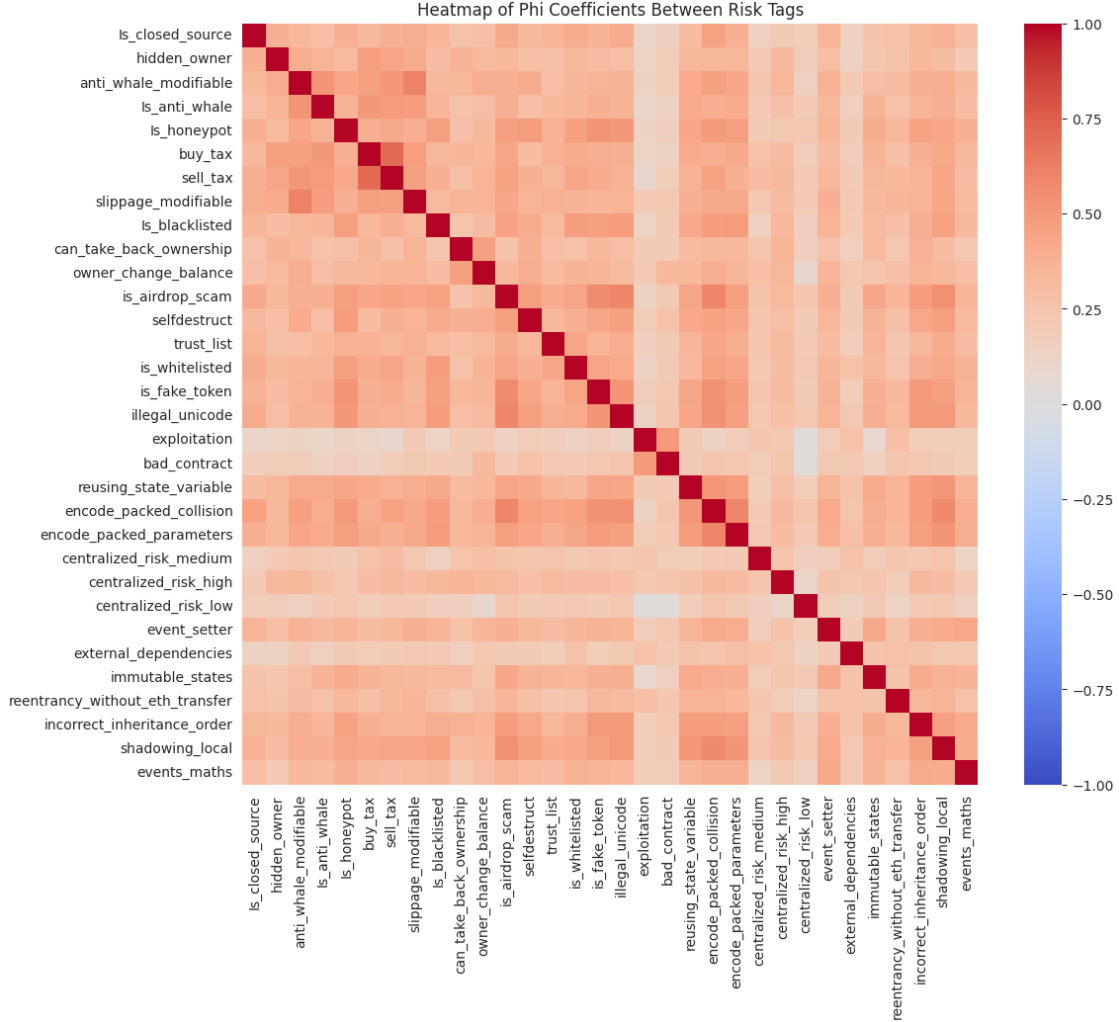
Now we can visualise this matrix as a heatmap

```python
[58]:  # Setting the size of the plot
       plt.figure(figsize=(12, 10))

       # Creating a heatmap
       sns.heatmap(phi_matrix.astype(float), annot=False, fmt=".2f", cmap='coolwarm',
       ↪vmin=-1, vmax=1)
       plt.title('Heatmap of Phi Coefficients Between Risk Tags')
       plt.show()
```

Heatmap of Phi Coefficients Between Risk Tags

1.3 Conclusion

- The most frequently occurring risk tag in the dataset is `exploitation` with 468 occurrences. This makes sense as `exploitation` refers to vulnerabilities or weaknesses in smart contracts that can be exploited by malicious actors, to steal funds, manipulate transactions, or cause other unintended behaviours. Companies typically handle this risk tag by undergoing regular and thorough security audits by third-party experts help identify potential vulnerabilities. These audits involve a detailed review of the contract's code and logic. Other ways to mitigate the effects of vulnerabilities include automated testing, bug bounty programs, and following best practices and standards.

- A few risk tags stood out as surprising due to their high or low frequencies. These included:

  1. **Bad_contract (373 occurrences):** The frequency of this tag is surprisingly high. This suggests a significant number of contracts are fundamentally flawed or poorly written, leading to potential security and operational risks.

  2. **External_dependencies (316 occurrences):** The high frequency of external dependencies highlights the reliance on external systems and contracts, which can introduce

risks if those dependencies are compromised or malfunction.

In existing literature, the focus is often on high-impact vulnerabilities like reentrancy, integer overflows, and access control issues. While `exploitation` is frequently discussed, tags like `bad_contract` and `external_dependencies` might not receive as much attention in literature.

- An example of two closely related smart contract risks are `Anti_whale_modifiable` and `Slippage_modifiable`. These two have a phi-coefficient value of 0.604309. It makes sense that these two risks are closely related because both involve modifications to parameters that can affect transaction dynamics. Anti-whale mechanisms typically limit the amount one can buy or sell to prevent manipulation by large holders, while slippage modification impacts the acceptable price variation during transactions. Both aim to prevent market manipulation and protect smaller investors.

- An example of two smart contract risks that are *not* closely related are `Exploitation` and `External_dependencies`. These two have a phi-coefficient value of 0.272019 It makes sense that these risks are not related because they are inherently different in nature; one is internal to the contract's logic, while the other is external.