Data Loading and Preprocessing

First, we'll connect to the SQLite database and load the data into a pandas DataFrame.

```
   meeting_id  duration  attendees    salary
0      755788      0.75          1  250000.0
1      755788      0.75          1  163808.0
2      653454      0.25          1  250000.0
3      653454      0.25          1  135000.0
4      653454      0.25          1  163808.0
```

Calculate Total Cost and Total Cost per Meeting

Now, we'll add columns for total cost and total cost per meeting.

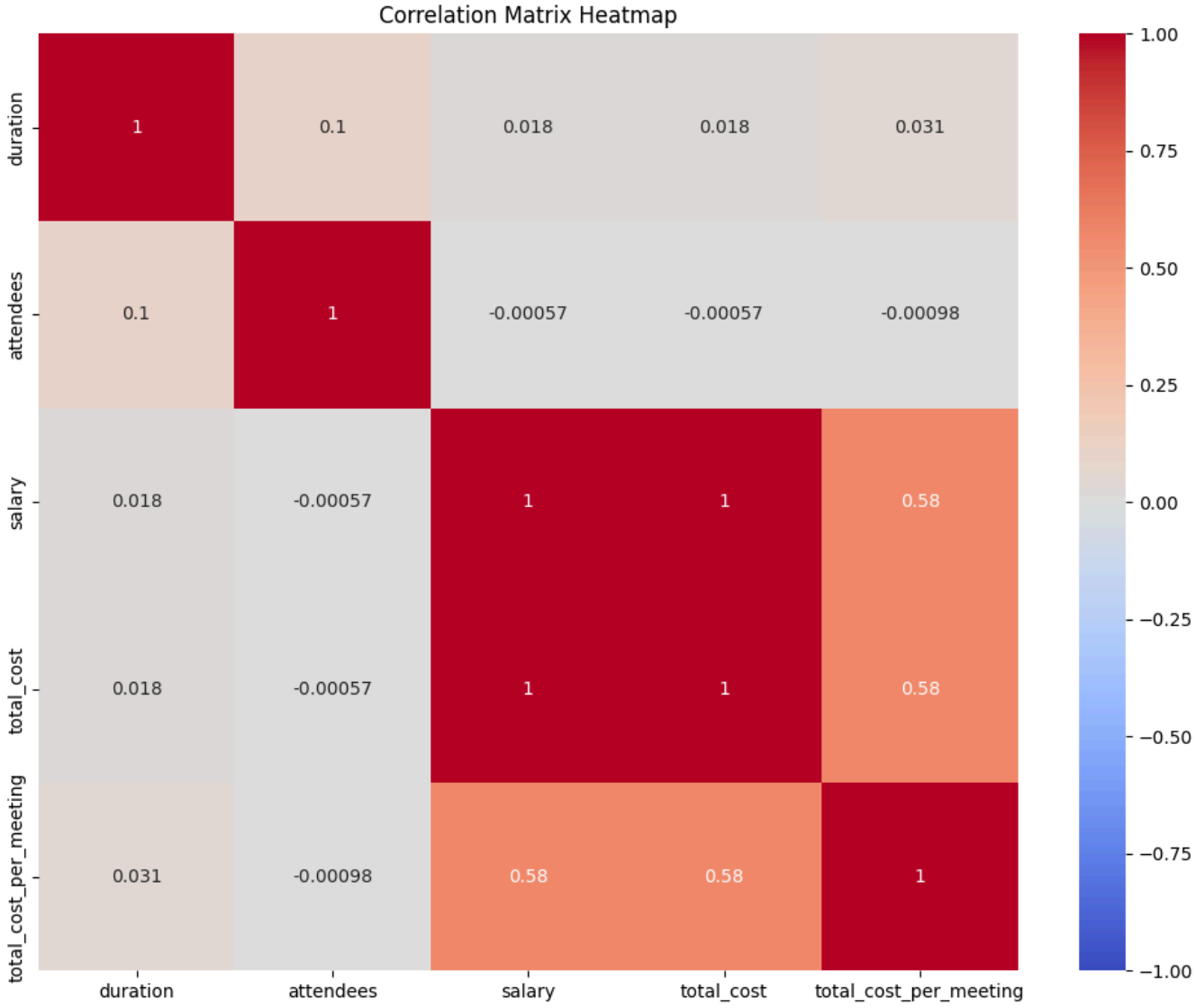
```
   meeting_id  duration  attendees    salary  total_cost  total_cost_per_meeting
0      755788      0.75          1  250000.0    187500.0                310356.0
1      755788      0.75          1  163808.0    122856.0                310356.0
2      653454      0.25          1  250000.0     62500.0                168452.0
3      653454      0.25          1  135000.0     33750.0                168452.0
4      653454      0.25          1  163808.0     40952.0                168452.0
```

	meeting_id	duration	attendees	salary	total_cost	total_cost_per_meeting
0	755788	0.75	1	250000.0	187500.0	310356.0
1	755788	0.75	1	163808.0	122856.0	310356.0
2	653454	0.25	1	250000.0	62500.0	168452.0
3	653454	0.25	1	135000.0	33750.0	168452.0
4	653454	0.25	1	163808.0	40952.0	168452.0

Correlation Analysis

Let's calculate and visualize the correlation matrix for our numeric columns.

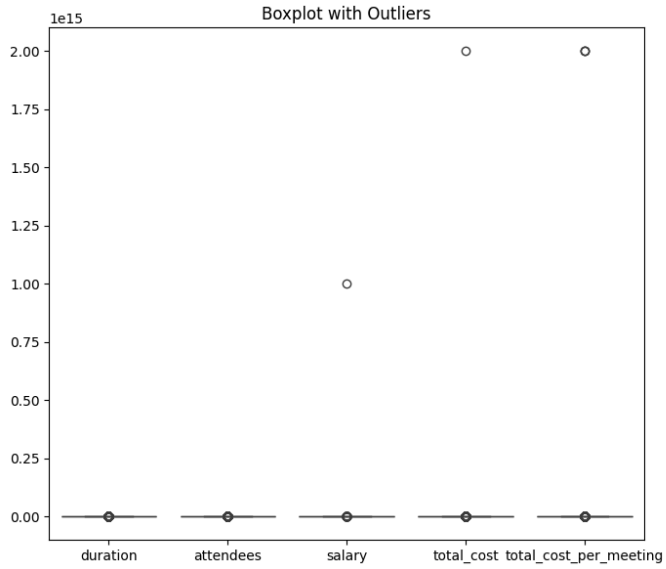
```
Correlation Matrix:
                        duration  attendees    salary  total_cost  total_cost_per_meeting
duration                1.000000   0.100116  0.018141    0.018141                0.031437
attendees               0.100116   1.000000 -0.000566   -0.000566               -0.000981
salary                  0.018141  -0.000566  1.000000    1.000000                0.577059
total_cost              0.018141  -0.000566  1.000000    1.000000                0.577059
total_cost_per_meeting  0.031437  -0.000981  0.577059    0.577059                1.000000
```

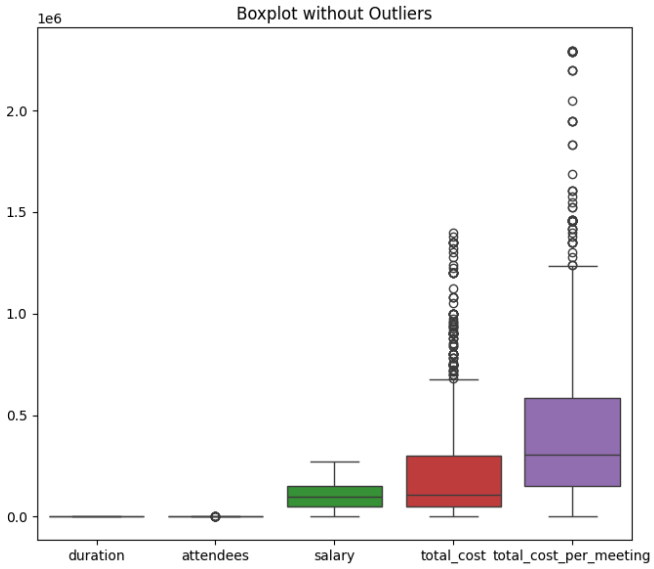


Outlier Detection and Removal

Now, let's identify and remove outliers using the Interquartile Range (IQR) method.

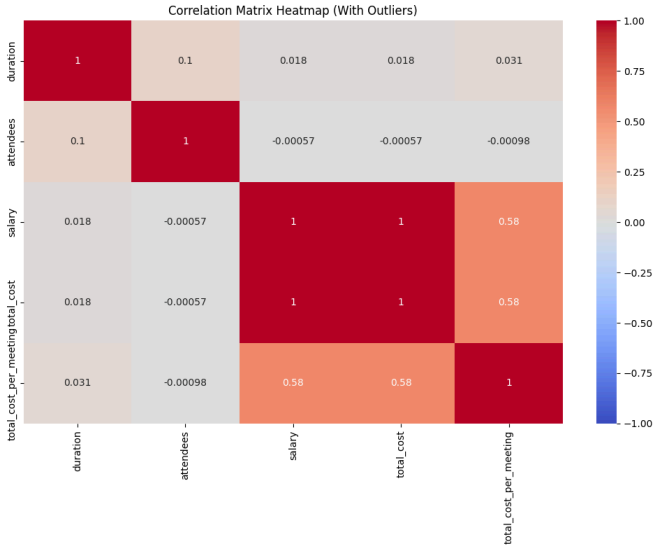
```
Q1 values:
duration                       0.5
attendees                      1.0
salary                     60000.0
total_cost                 60000.0
total_cost_per_meeting    195000.0
Name: 0.25, dtype: float64

Q3 values:
duration                       1.0
attendees                      6.0
salary                    150000.0
total_cost                600000.0
total_cost_per_meeting   1060000.0
Name: 0.75, dtype: float64

IQR values:
duration                       0.5
attendees                      5.0
salary                     90000.0
total_cost                540000.0
total_cost_per_meeting    865000.0
dtype: float64

Outliers detected:
0       False
1       False
2       False
3       False
4       False
        ...
1801    False
1802    False
1803    False
1804     True
1805    False
Length: 1806, dtype: bool

DataFrame after removing outliers:
   meeting_id  duration  attendees    salary  total_cost  total_cost_per_meeting
0      755788      0.75          1  250000.0    187500.0                310356.0
1      755788      0.75          1  163808.0    122856.0                310356.0
2      653454      0.25          1  250000.0     62500.0                168452.0
3      653454      0.25          1  135000.0     33750.0                168452.0
4      653454      0.25          1  163808.0     40952.0                168452.0

Original dataset size: 1806
Dataset size after removing outliers: 1339

Min/Max values before outlier handling:
     duration  attendees        salary    total_cost  total_cost_per_meeting
min      0.25          0  0.000000e+00  0.000000e+00            0.000000e+00
max      8.00    1000000  1.000000e+15  2.000000e+15            2.000100e+15

Min/Max values after outlier handling:
     duration  attendees    salary  total_cost  total_cost_per_meeting
min      0.25          0       0.0         0.0                     0.0
max      1.75         13  270000.0   1400000.0               2293750.0
```

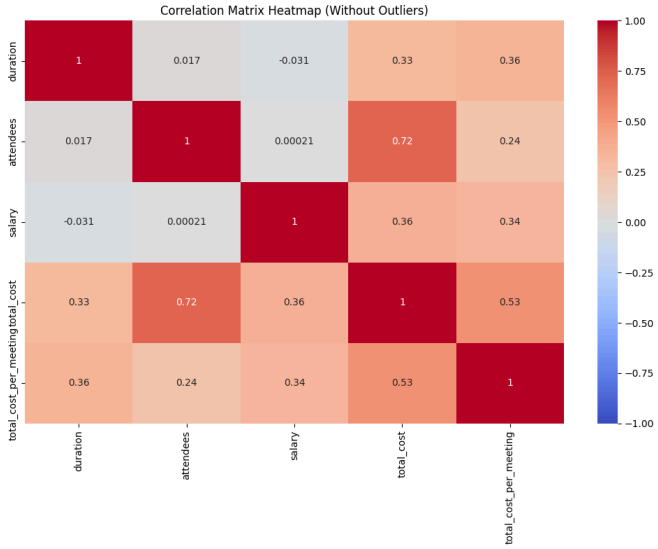




Correlation Analysis Without Outliers

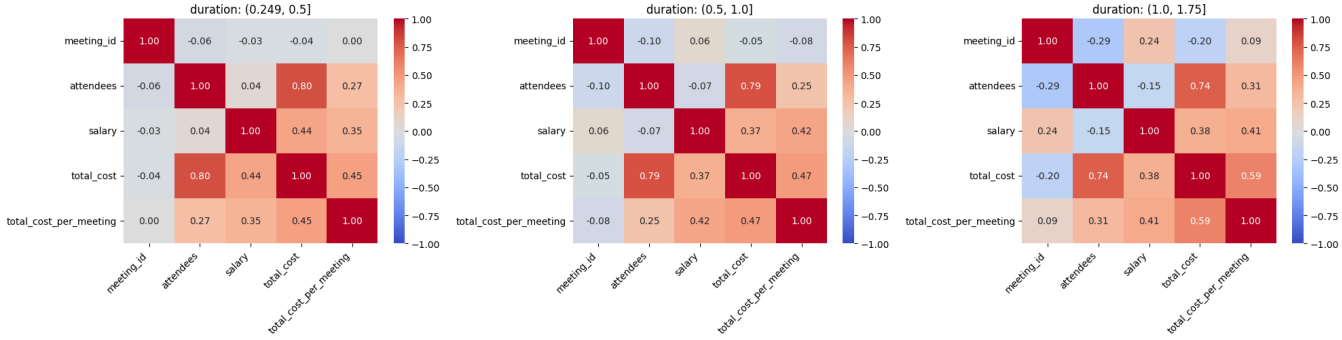
Let's recalculate the correlation matrix after removing outliers and compare it with the original.





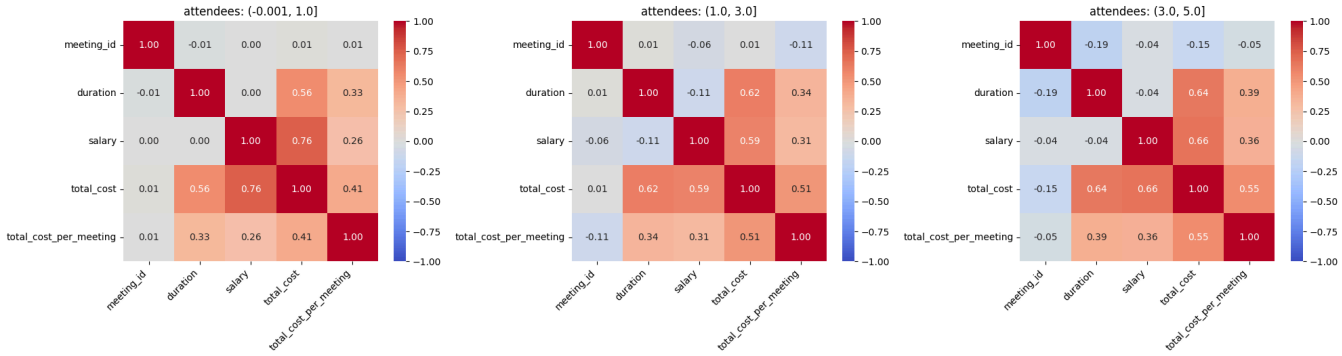

```
Correlation Matrix (With Outliers):
                         duration  attendees    salary  total_cost  total_cost_per_meeting
duration                 1.000000   0.100116  0.018141    0.018141                0.031437
attendees                0.100116   1.000000 -0.000566   -0.000566               -0.000981
salary                   0.018141  -0.000566  1.000000    1.000000                0.577059
total_cost               0.018141  -0.000566  1.000000    1.000000                0.577059
total_cost_per_meeting   0.031437  -0.000981  0.577059    0.577059                1.000000

Correlation Matrix (Without Outliers):
                         duration  attendees    salary  total_cost  total_cost_per_meeting
duration                 1.000000   0.016571 -0.031110    0.326703                0.356491
attendees                0.016571   1.000000  0.000209    0.724696                0.242574
salary                  -0.031110   0.000209  1.000000    0.361457                0.337871
total_cost               0.326703   0.724696  0.361457    1.000000                0.529158
total_cost_per_meeting   0.356491   0.242574  0.337871    0.529158                1.000000

Analyzing correlations by duration
```

Correlation Heatmaps by duration



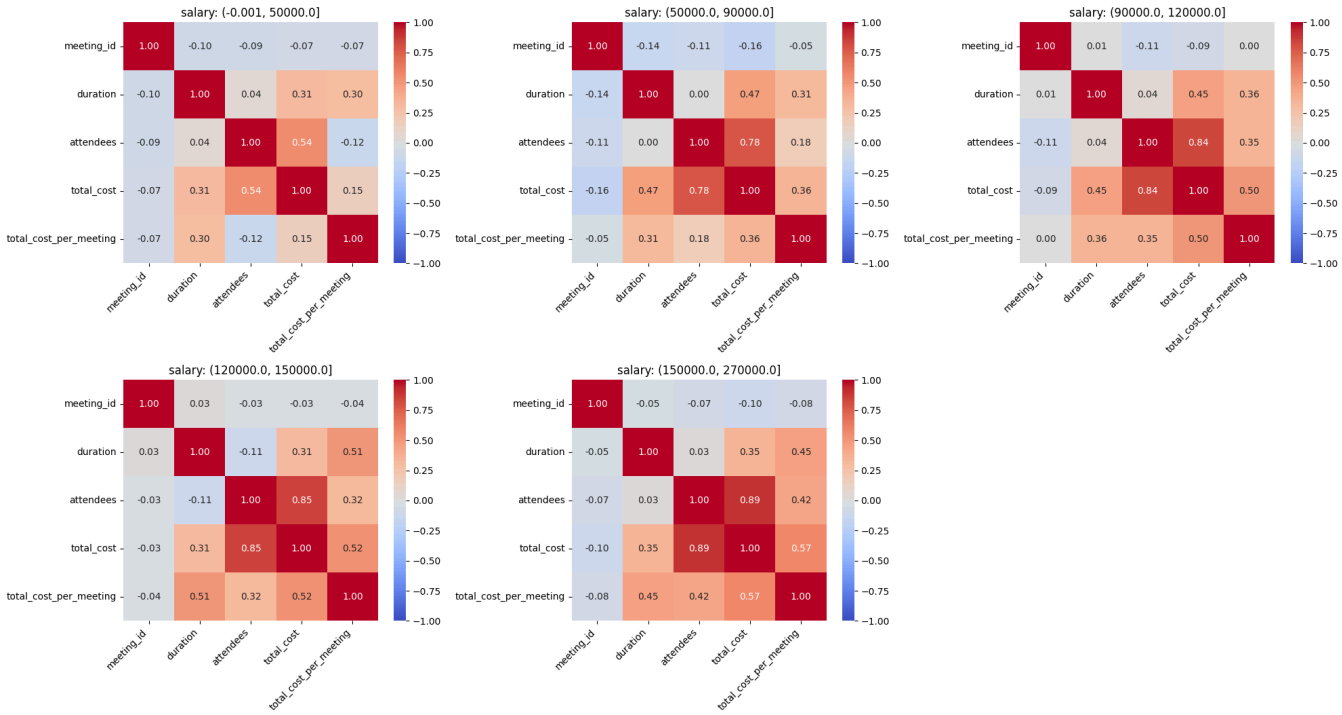
Analyzing correlations by attendees

Correlation Heatmaps by attendees

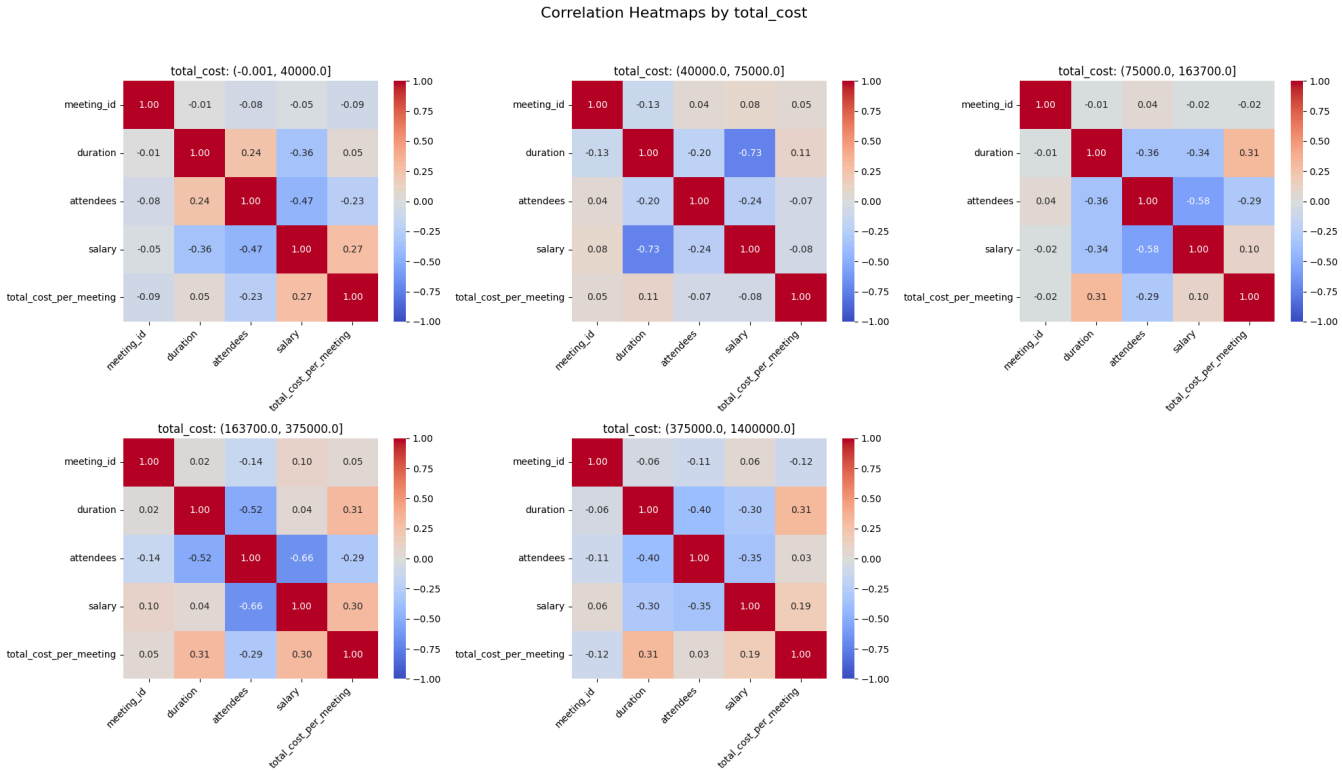


Analyzing correlations by salary

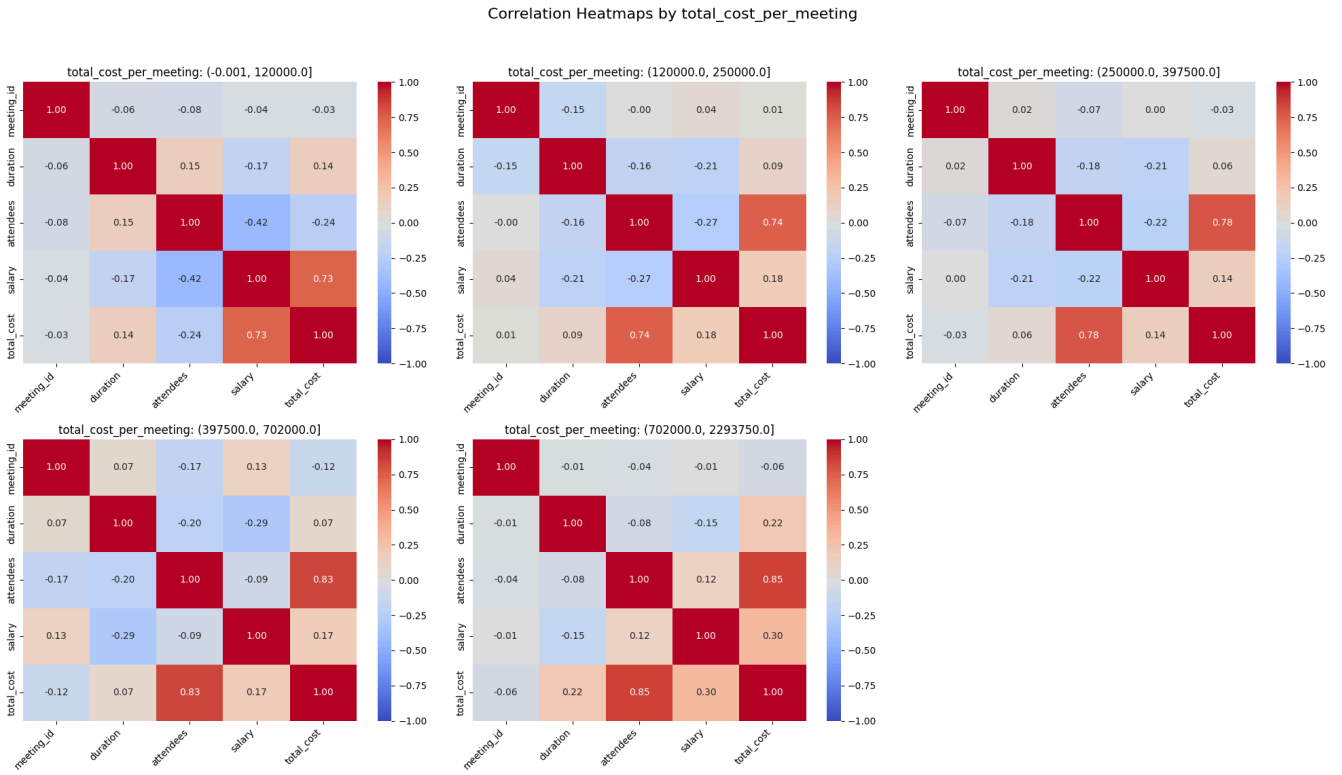
Correlation Heatmaps by salary



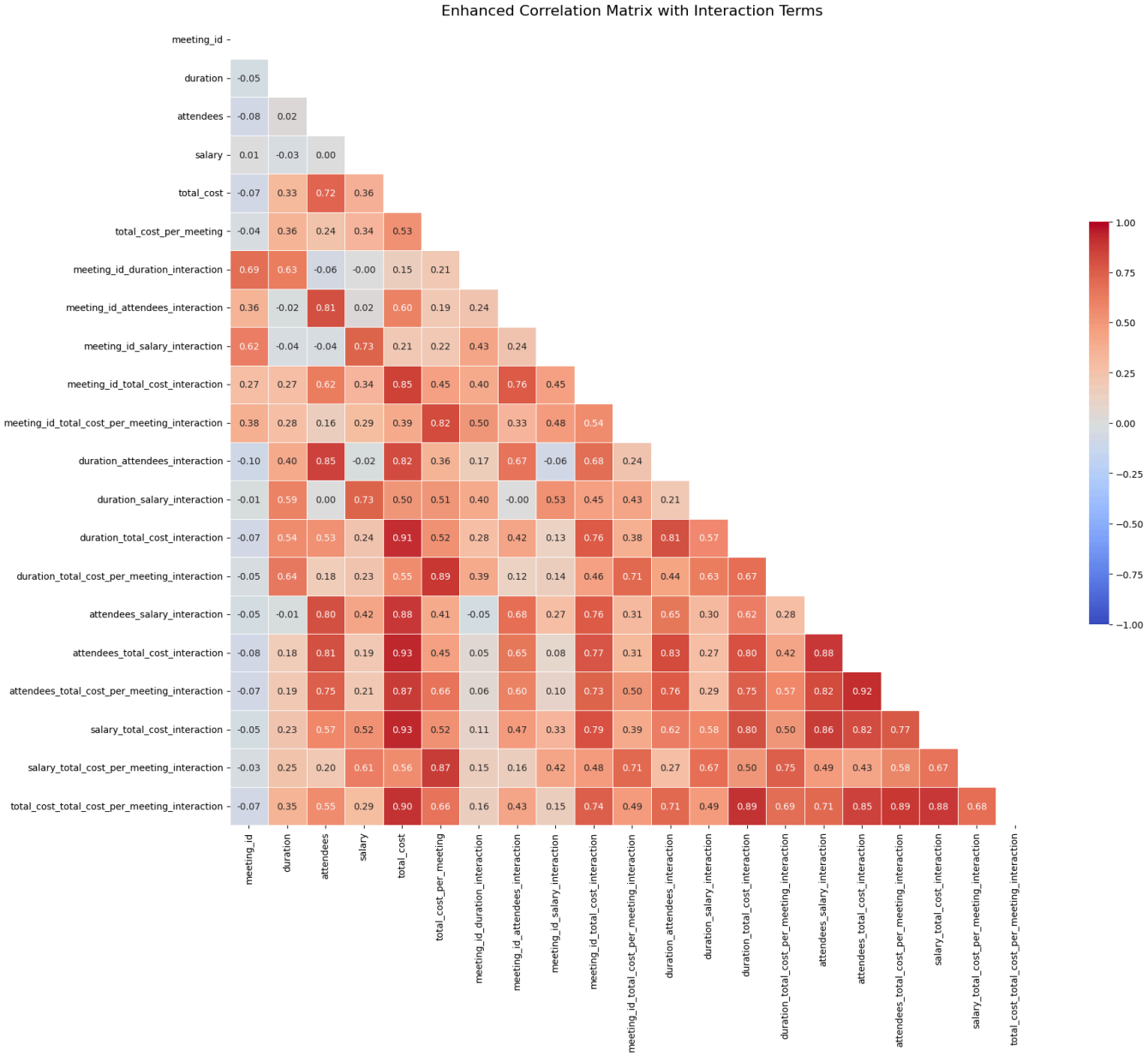
Analyzing correlations by total_cost



Analyzing correlations by total_cost_per_meeting

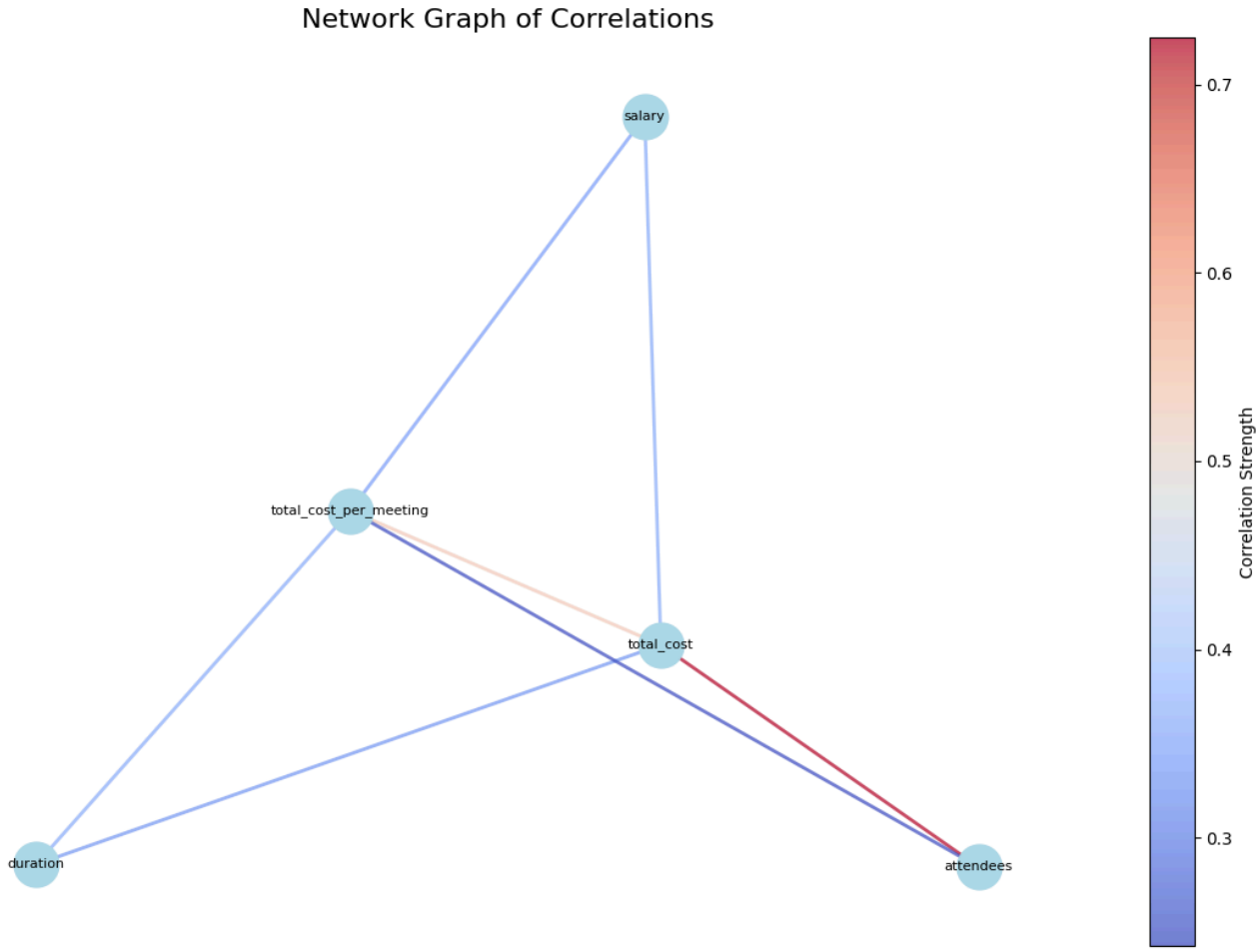


Enhanced Correlation Matrix with Interaction Terms


Enhanced Correlation Matrix with Interaction Terms
meeting_id
duration
attendees
salary
total_cost
total_cost_per_meeting
meeting_id_duration_interaction
meeting_id_attendees_interaction
meeting_id_salary_interaction
meeting_id_total_cost_interaction
meeting_id_total_cost_per_meeting_interaction
duration_attendees_interaction
duration_salary_interaction
duration_total_cost_interaction
duration_total_cost_per_meeting_interaction
attendees_salary_interaction
attendees_total_cost_interaction
attendees_total_cost_per_meeting_interaction
salary_total_cost_interaction
salary_total_cost_per_meeting_interaction
total_cost_total_cost_per_meeting_interaction
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

```
Top 10 Highest Correlations:
total_cost                                        attendees_total_cost_interaction
0.928775
                                                  salary_total_cost_interaction
0.928191
attendees_total_cost_per_meeting_interaction      attendees_total_cost_interaction
0.919330
total_cost                                        duration_total_cost_interaction
0.913106
                                                  total_cost_total_cost_per_meeting_interaction
0.900974
total_cost_total_cost_per_meeting_interaction     duration_total_cost_interaction
0.890676
total_cost_per_meeting                            duration_total_cost_per_meeting_interaction
0.886660
total_cost_total_cost_per_meeting_interaction     attendees_total_cost_per_meeting_interaction
0.885984
attendees_total_cost_interaction                  attendees_salary_interaction
0.879603
attendees_salary_interaction                      total_cost
0.876116
dtype: float64

Top 10 Lowest Correlations:
duration                           meeting_id_attendees_interaction   -0.019141
meeting_id_attendees_interaction   salary                              0.016692
attendees                          duration                            0.016571
duration                           attendees_salary_interaction       -0.014578
meeting_id                         duration_salary_interaction        -0.011840
salary                             meeting_id                          0.005372
                                   meeting_id_duration_interaction    -0.003347
meeting_id_attendees_interaction   duration_salary_interaction        -0.003261
salary                             attendees                           0.000209
duration_salary_interaction        attendees                           0.000030
dtype: float64
```

Network Graph of Correlations



Top 5 Most Connected Features:
total_cost: 4 connections
total_cost_per_meeting: 4 connections
duration: 2 connections
attendees: 2 connections
salary: 2 connections