After an extensive examination of the vuln_data.csv dataset, it has become apparent that certain Common Vulnerability Scoring System (CVSS) factors demonstrate a more pronounced correlation with malware occurrences than others. Interestingly, despite conventional wisdom suggesting that access vector, authentication, and access complexity are pivotal in determining a vulnerability's exploitability, our analysis suggests a more nuanced reality

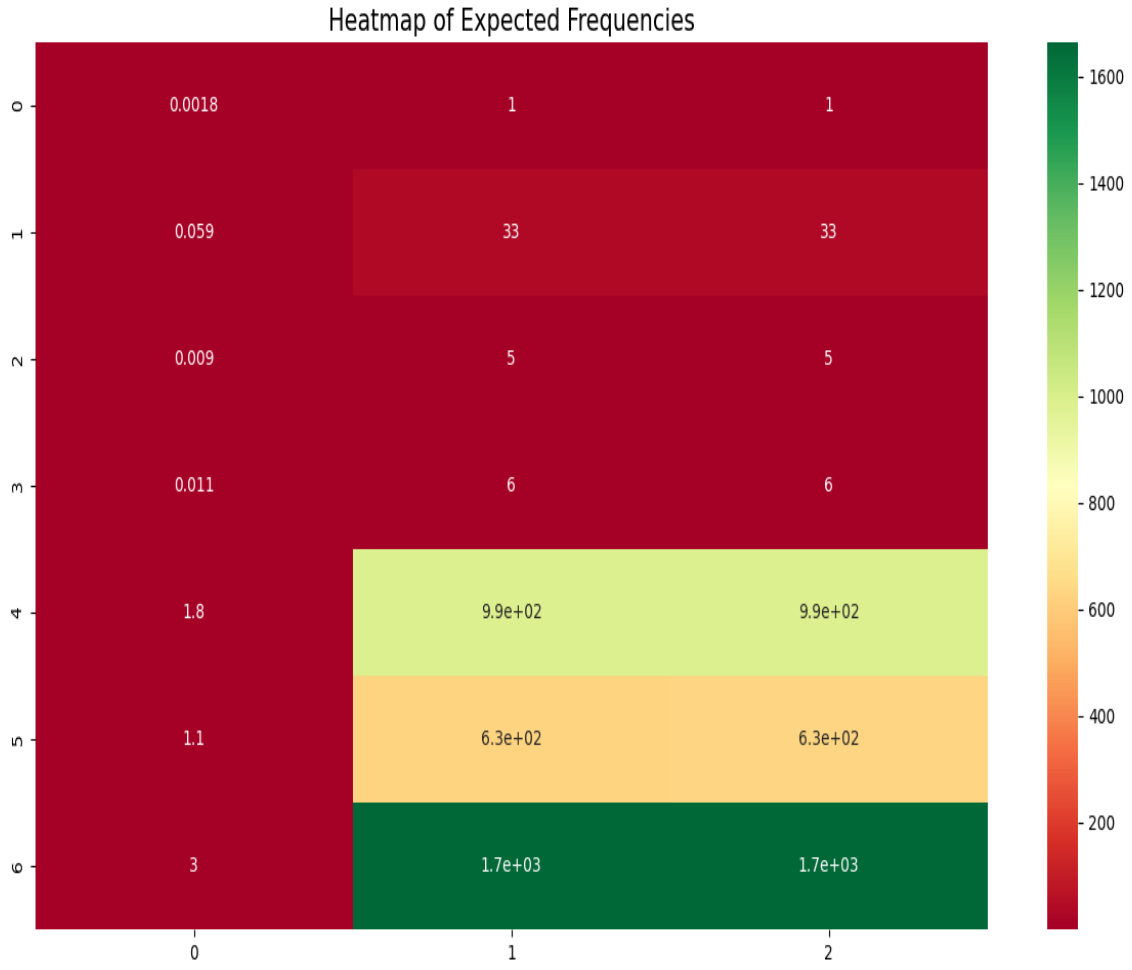
From the count plots generated for different CVSS factors, it is clear that the majority of vulnerabilities have a network-related access vector (N), suggesting that remotely exploitable vulnerabilities are more common. This aligns with expectations, as remote access would naturally increase the potential for malware exploitation

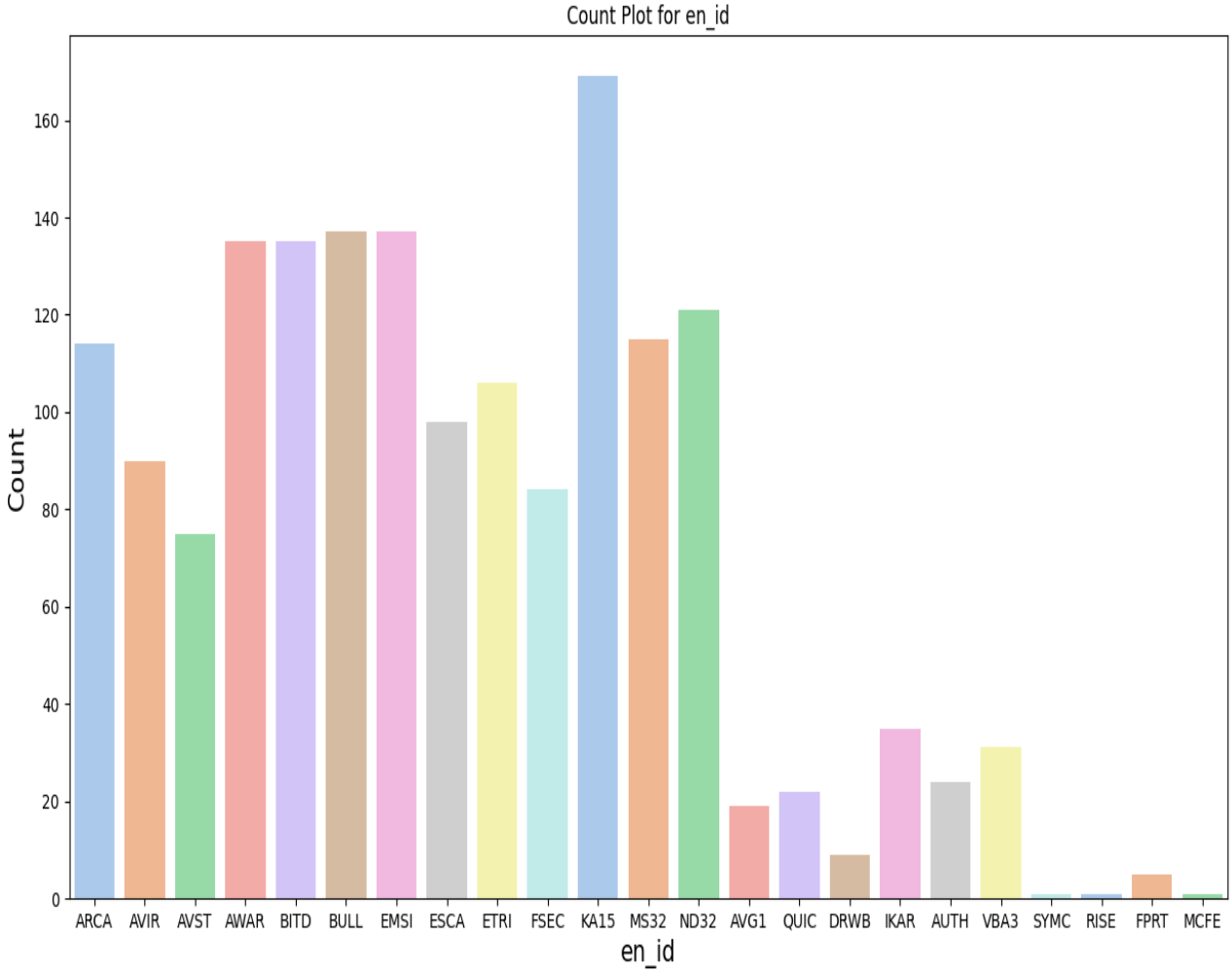
However, the complexity of access does not show a definitive trend correlating with malware occurrences. The distribution across low (L), medium (M), and high (H) access complexity is relatively balanced, with a slight predominance of the medium complexity. This is somewhat surprising as it may imply that the difficulty of exploiting a vulnerability does not significantly deter attackers, contrary to what might be expected.

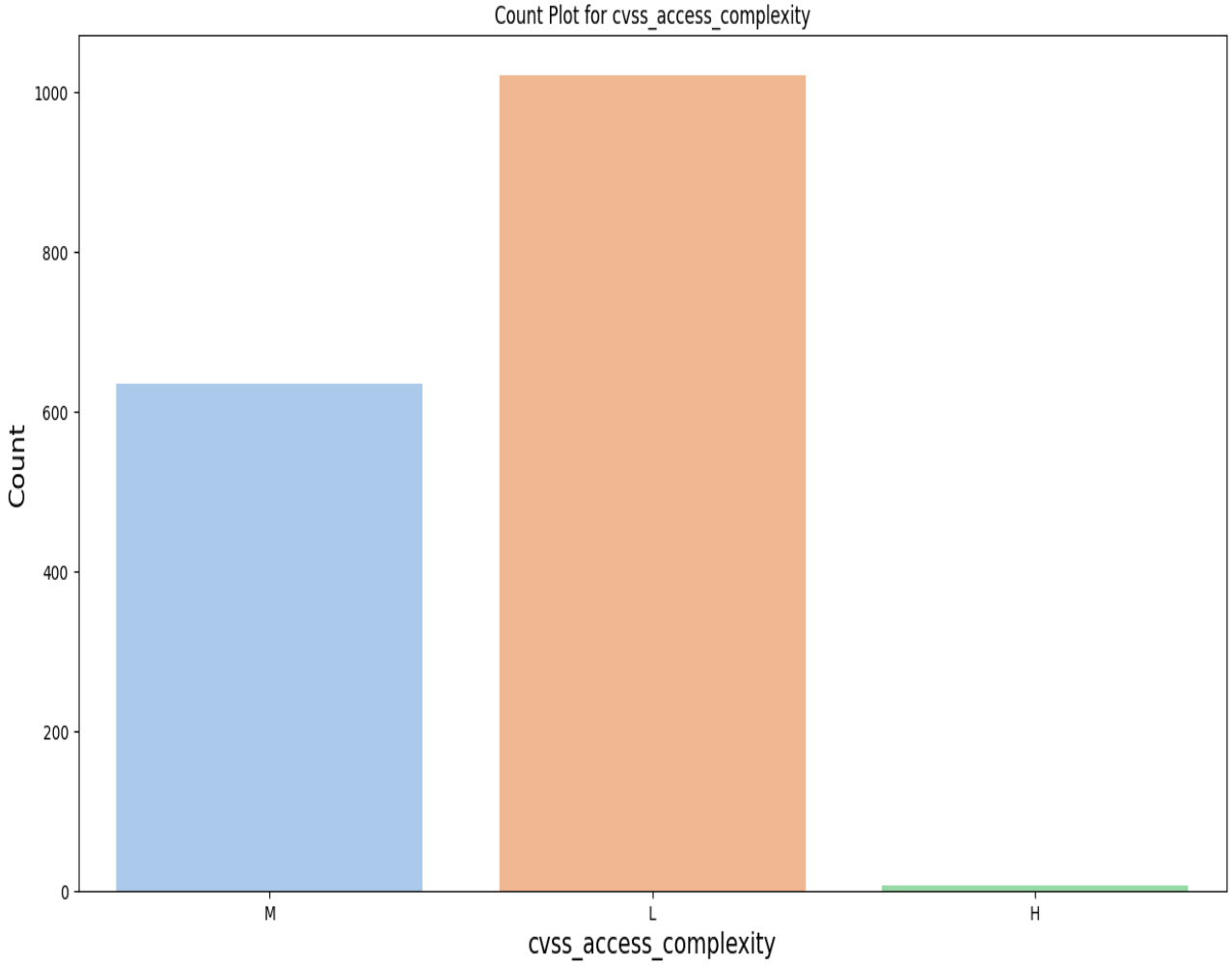
As for authentication (M for multiple, N for none), the vast majority of vulnerabilities require no authentication to exploit. This finding is in line with the assumption that vulnerabilities that are easier to exploit (i.e., requiring no authentication) would be more commonly leveraged in malware attacks.

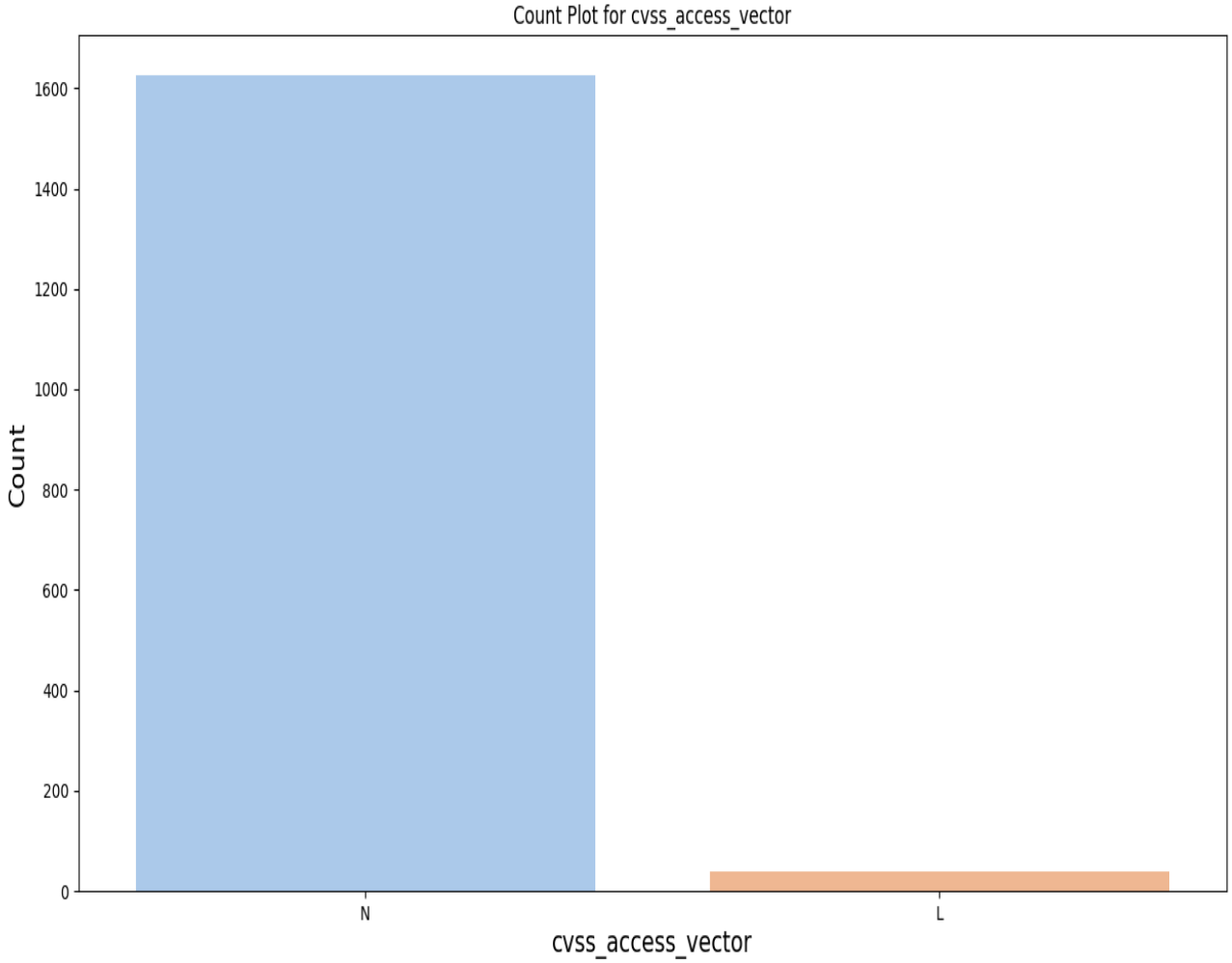
The chi-square tests and Cramér's V statistics further reveal that while there is an association between some CVSS factors and malware presence, the strength of this association varies. Notably, the Cramér's V values for certain pairs of variables indicate a moderate association, such as between confidentiality impact and availability impact, as well as confidentiality impact and integrity impact, both scoring a Cramér's V of 0.7075. These findings suggest that vulnerabilities with a higher impact on confidentiality, integrity, or availability are more likely to be associated with malware.

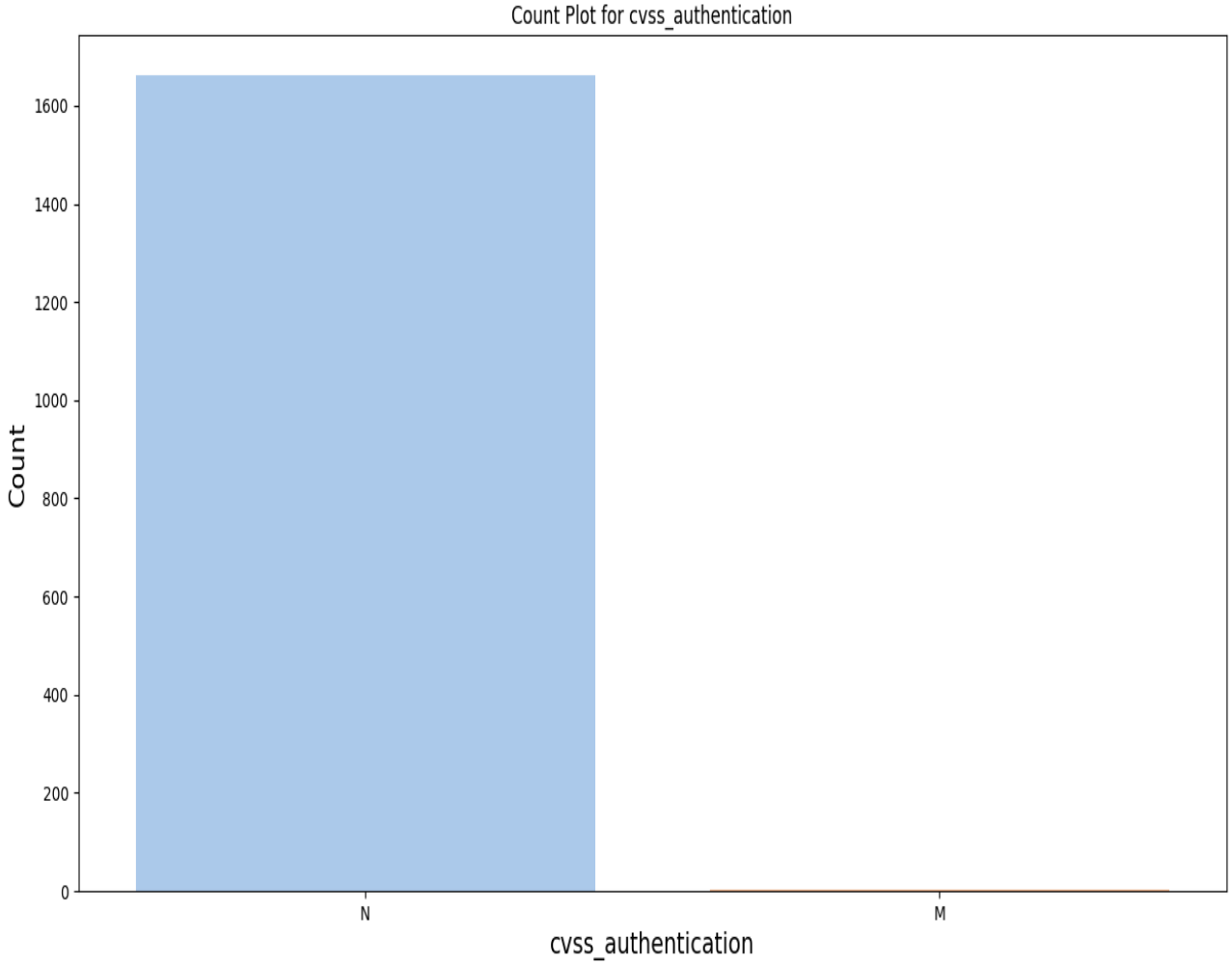
In conclusion, while some CVSS factors such as the need for authentication and the impact on confidentiality, integrity, or availability do appear related to malware prevalence, the complexity of access does not show a strong correlation. This analysis challenges some aspects of conventional wisdom and highlights the multifaceted nature of cybersecurity threats. It is crucial for cybersecurity professionals to consider a broader range of factors when assessing the risk and potential for exploitation of vulnerabilities.

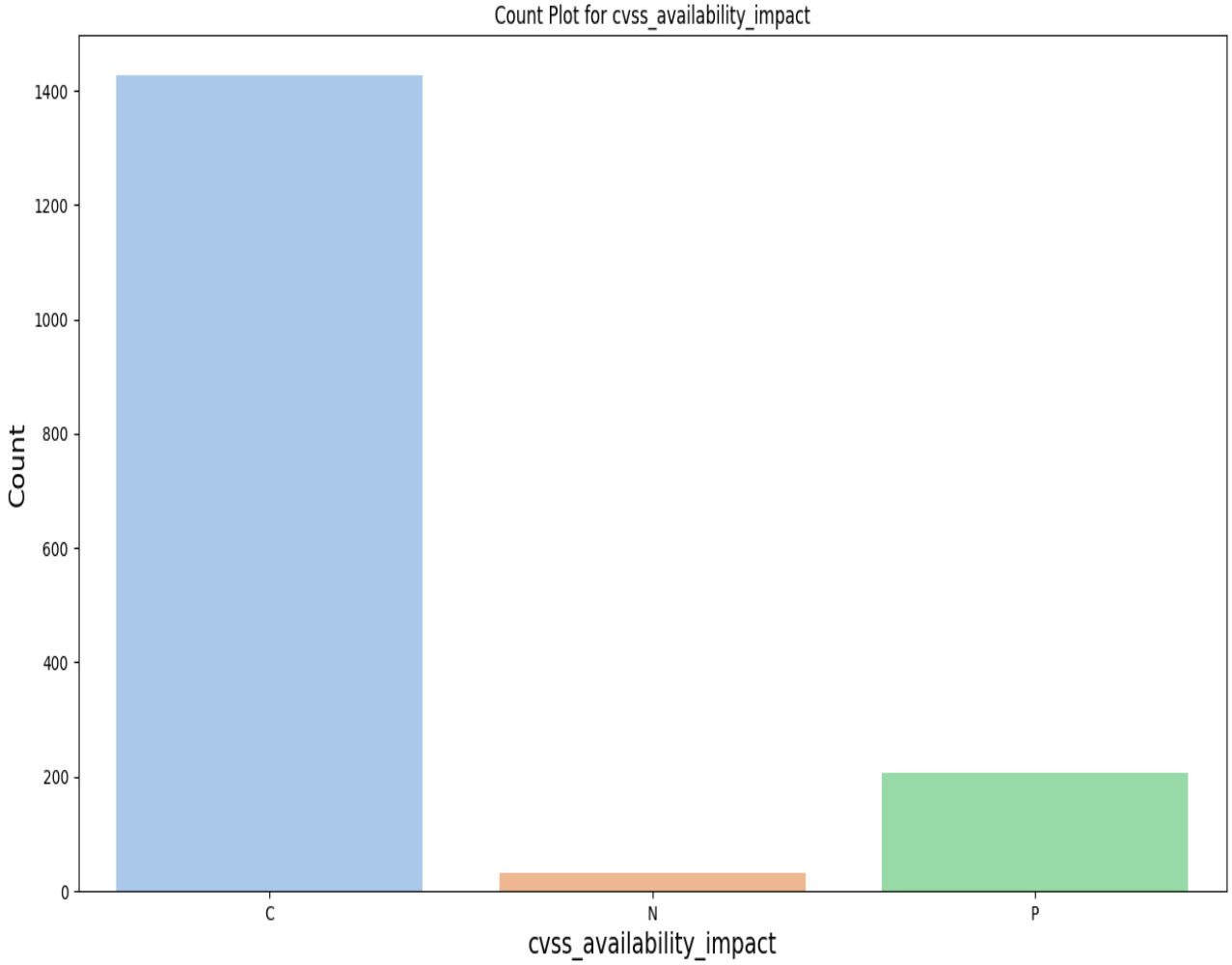

Heatmap of Expected Frequencies
0.0018
1
1
0.059
33
33
0.009
5
5
0.011
6
6
1.8
9.9e+02
9.9e+02
1.1
6.3e+02
6.3e+02
3
1.7e+03
1.7e+03
0
1
2
0
1
2
3
4
5
6
1600
1400
1200
1000
800
600
400
200

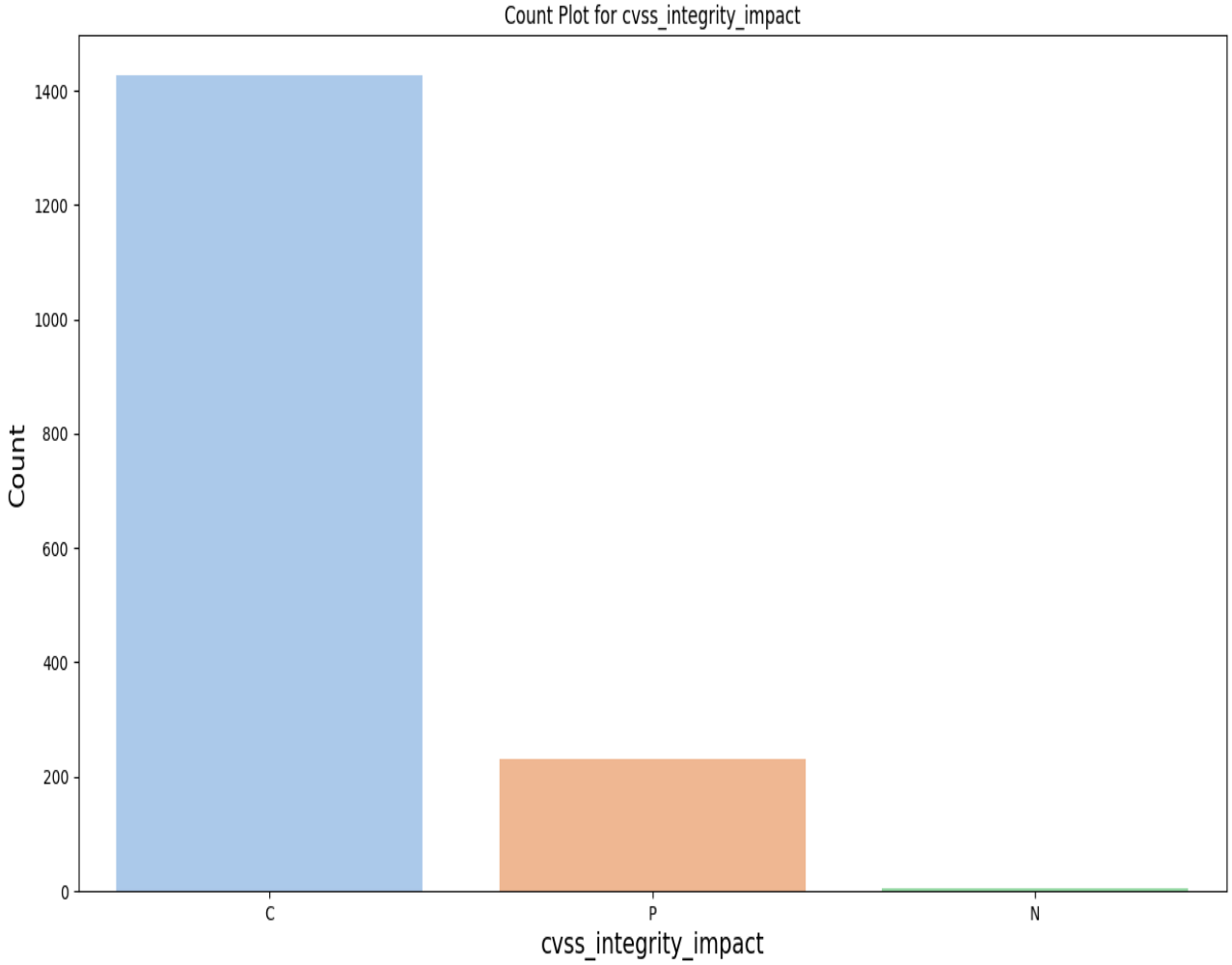

Count Plot for en_id
Count
0
20
40
60
80
100
120
140
160
ARCA
AVIR
AVST
AWAR
BITD
BULL
EMSI
ESCA
ETRI
FSEC
KA15
MS32
ND32
AVG1
QUIC
DRWB
IKAR
AUTH
VBA3
SYMC
RISE
FPRT
MCFE
en_id


Count Plot for cvss_access_complexity
Count
1000
800
600
400
200
0
M
L
H
cvss_access_complexity


Count Plot for cvss_access_vector
Count
1600
1400
1200
1000
800
600
400
200
0
N
L
cvss_access_vector


Count Plot for cvss_authentication
Count
1600
1400
1200
1000
800
600
400
200
0
N
M
cvss_authentication


Count Plot for cvss_availability_impact
Count
1400
1200
1000
800
600
400
200
0
C
N
P
cvss_availability_impact


Count Plot for cvss_integrity_impact
Count
1400
1200
1000
800
600
400
200
0
C
P
N
cvss_integrity_impact