## *ITERITEKA Axel GCYS-1*

*DUSHIMIRIMANA Christ Leycène*

## ***Introduction***

Les attaques de phishing représentent une menace croissante pour les entreprises. Ce projet vise à analyser les données d'attaques de phishing recueillies au sein d'une entreprise pour identifier les départements les plus vulnérables et les types d'attaques les plus fréquents. En comprenant mieux ces attaques, nous pourrons proposer des recommandations pour renforcer la cybersécurité de l'entreprise.

***Analyse statistique descriptive***

| Canal de communication | Taux de réussite moyen | Taux de réussite médian | Fréquence d'attaques |
| --- | --- | --- | --- |
| Email | 27.9% | 0 | 59.5% |
| SMS | 28.6% | 0 | 20.6% |
| Social Media | 33.0% | 0 | 9.7% |
| Website | 21.6% | 0 | 10.2% |

· ***Taux de Réussite :***

**Media** a le taux de réussite moyen le plus élevé à 33.0%, ce qui indique que les attaques via ce canal sont relativement plus efficaces par rapport aux autres.

**Website** a le taux de réussite moyen le plus bas à 21.6%, ce qui signifie que les attaques via les sites web sont les moins efficaces.

**Email** et **SMS** se situent entre ces deux extrêmes avec des taux de réussite moyens de 27.9% et 28.6% respectivement.

***Fréquence des Attaques :***

**Email** est de loin le canal le plus utilisé pour les attaques de phishing, avec 595 occurrences. Cela peut s'expliquer par la facilité d'envoi massif de courriels.

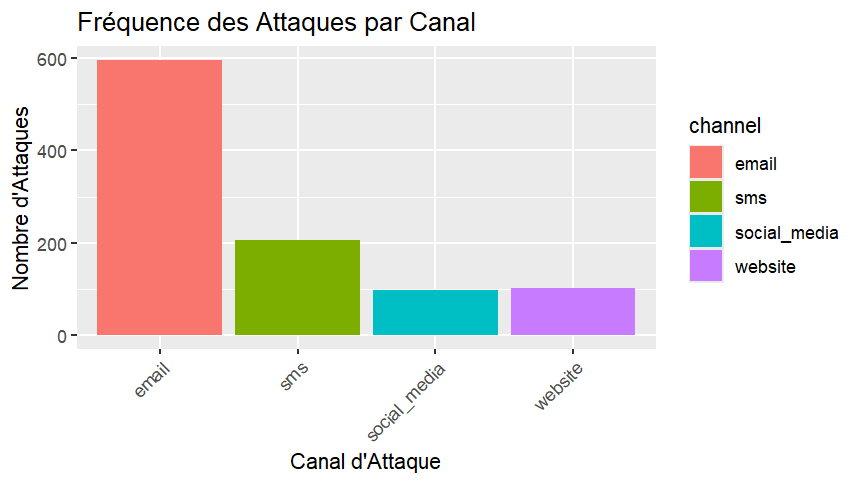
**SMS** le deuxième avec 206 attaques, ce qui montre que c'est aussi un canal populaire, bien que moins fréquenté que l’e-mail.

**Social Media** et **Website** ont des fréquences beaucoup plus faibles, ce qui pourrait être dû à la nature plus ciblée et personnalisée des attaques via ces canaux.

***Médiane à 0 :***

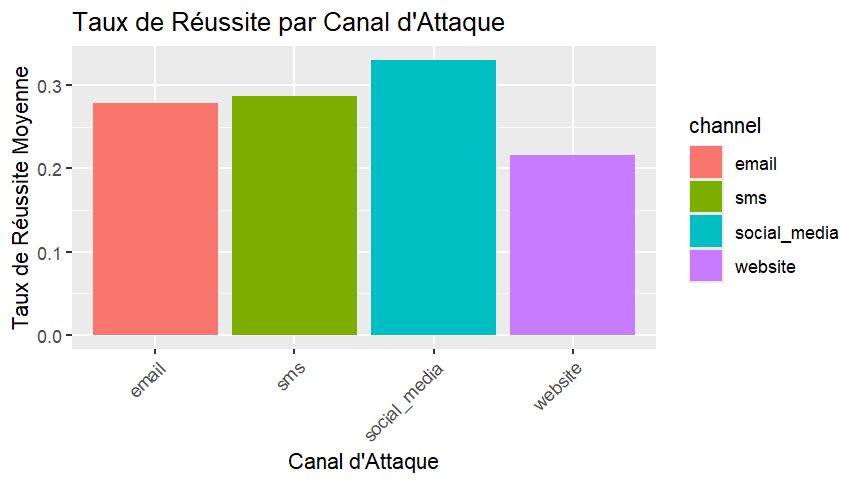
Le fait que le taux de réussite médian soit 0 pour tous les canaux suggère que la majorité des attaques échouent, avec quelques exceptions qui parviennent à réussir, ce qui tire la moyenne vers le haut.

*Ci dessous le graphe qui représente la fréquence des attaques*



***Analyse statistique bivariée***

*Ci dessous le graphique représentant le taux de réussite par Canal d’attaque*



**Réseaux sociaux** sont le canal le plus efficace pour les attaques de phishing selon le taux de réussite moyen. **Sites web** sont le moins efficace, malgré une utilisation notable. **SMS** et **Email** ont des taux de réussite moyens similaires, mais **SMS** est légèrement plus efficace.

***Paramètres de population et Intervalles de confiance***

| Canal de communication | Taux de réussite moyen | Intervalle de Confiance |
| --- | --- | --- |
| Email | 0.279 | [0.243, 0.315] |
| SMS | 0.286 | [0.225, 0.348] |
| Social Media | 0.330 | [0.236, 0.423] |
| Website | 0.216 | [0.136, 0.296] |

*Les intervalles de confiance sont calculées à 95%.*

Les réseaux sociaux ont le taux de réussite moyen le plus élevé (33.0%), tandis que les sites web ont le plus bas (21.6%). Les attaques par SMS et email sont similaires, avec environ 28% de réussite. Ces résultats montrent que les attaques via les réseaux sociaux sont les plus efficaces, mais moins fréquentes que les attaques par email, qui sont les plus courantes.

***Tests d’hypothèses***

Nous avons testé l’hypothèse selon laquelle plus de 25% des attaques par email sont réussies.

Après avoir filtré les données pour ne sélectionner que les attaques par email, puis calculé la proportion de succès observée parmi ces attaques. En fixant l'hypothèse nulle à **25% de réussite**, nous avons ensuite calculé une statistique de test **z** pour comparer la proportion observée (27.9%) à cette hypothèse. Enfin, nous avons comparé cette statistique **z** à une valeur critique prédéfinie pour déterminer s'il existe suffisamment de preuves pour rejeter l'hypothèse nulle.

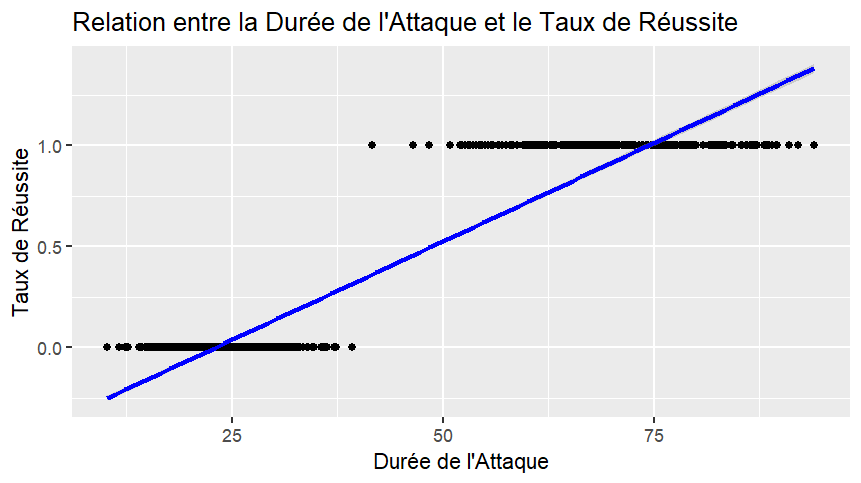
La statistique **z** calculée est de **2.12**, ce qui est supérieur à la valeur critique de **1.645**. Cela signifie qu'il y a suffisamment de preuves pour rejeter l'hypothèse nulle. En d'autres termes, nous pouvons conclure que le taux de réussite des attaques par email est significativement supérieur à **25%**, ce qui suggère une certaine efficacité des attaques par email.

***Modélisation par régression Linéaire***

On a filtré les données pour enlever celles où il manquait des valeurs pour le temps de détection et le succès des attaques. Ensuite, on a utilisé un modèle de régression linéaire pour voir comment le temps de détection affecte le taux de réussite des attaques. Le modèle permet de comprendre si plus de temps pris pour détecter une attaque impacte la probabilité que l'attaque soit réussie.

On a analysé les données et trouvé que chaque heure supplémentaire de détection augmente le taux de réussite des attaques de 1.95%. Les résultats montrent une forte relation entre le temps de détection et le succès des attaques. Le modèle est statistiquement significatif et explique 91.35% des variations de succès.

*Ci dessous le graphique de la Relation entre la Durée de Détection et le Taux de Réussite des Attaques*



Le graphique montre que plus le temps de détection est long, plus le taux de réussite des attaques augmente, avec une tendance positive représentée par la ligne bleue de régression. Les points de données montrent les résultats individuels des attaques.