

Projet data :

Offres promotionnelles

Contexte:

Nous avons été embauchés par une entreprise pour analyser cet ensemble de données, il s'agit d'un échantillon d'une centaine de personnes qui ont reçu un e-mail contenant un lien promotionnel. L'entreprise souhaite orienter ces futurs liens promotionnels et augmenter son nombre d'ouverture.

Dans ce jeu de données, il y a trois colonnes:

- Age
- Sexe
- Tookaction (c'est-à-dire si la personne a cliqué sur le lien promotionnel)

Déroulement du projet :

1) Question:

Combien est le nombre de personnes qui cliquent ou non sur le lien ?

Total nombre personnes ayant cliqué ou pas

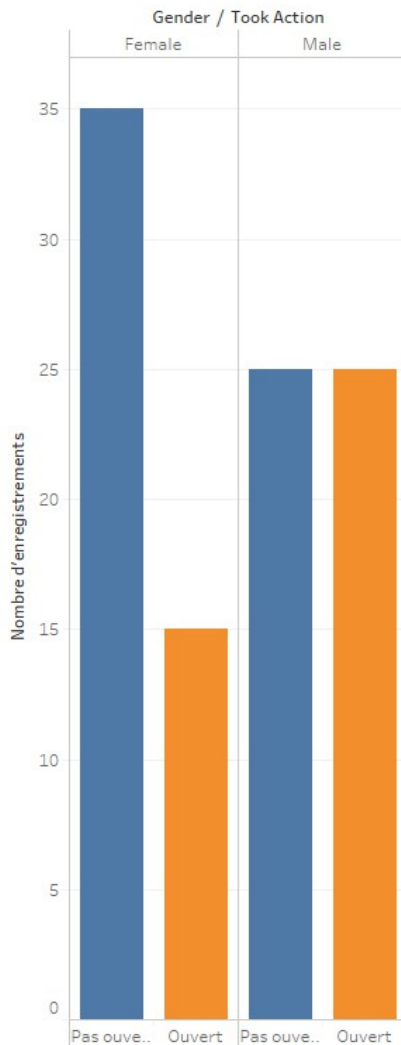
Took Action	
Pas ouvert	60
Ouvert	40

Ce premier diagramme nous montre que parmi les cents personnes ayant reçu ce mail seulement quarante personnes l'ont ouvert.

2) Question:

Combien, parmi ces cents personnes, sont des femmes ou des hommes et la répartition d'ouverture de ces mails?

Diagramme de Ouverture du lien promo/par genre



Ce diagramme révèle que sur les 100 personnes qui composent cet ensemble de données, 50 sont des hommes et 50 sont des femmes, on peut aussi dire que 25 hommes ont ouvert le lien et que les 25 autres ne l'ont pas fait alors que chez les femmes, 15 l'ont ouvert et 35 ne l'ont pas fait. Nous avons donc pu avoir notre premier aperçu, c'est-à-dire que le lien promotionnel a eu un franc succès avec les hommes.

3) Question

Comment pouvons nous augmenter le clic vers le lien promotionnel chez les femmes?

Prenons le temps de développer :

Pour essayer de trouver des réponses à nos questions, je vais utiliser le logiciel Gretl. Grâce à ce processus, j'essaierai de savoir s'il existe des relations entre chaque entité (par exemple dans notre cas où l'on cherche à savoir si le fait d'être une femme plus âgée influence le fait d'ouvrir ou non un lien promotionnel, par exemple).

Modèle 2: Logit, utilisant les observations 1-100

Variable dépendante: TookAction

Écart type basés sur la matrice hessienne

	coefficient	erreur std.	z	p. critique	
const	-33,7146	8,83304	-3,817	0,0001	***
Age	0,887226	0,231787	3,828	0,0001	***
Female	-4,43738	1,49193	-2,974	0,0029	***
Moy. var. dép.	0,400000	Éc. type var. dép.	0,492366		
R2 de McFadden	0,779993	R2 ajusté	0,735417		
Log de vraisemblance	-14,80675	Critère d'Akaike	35,61349		
Critère de Schwarz	43,42901	Hannan-Quinn	38,77657		

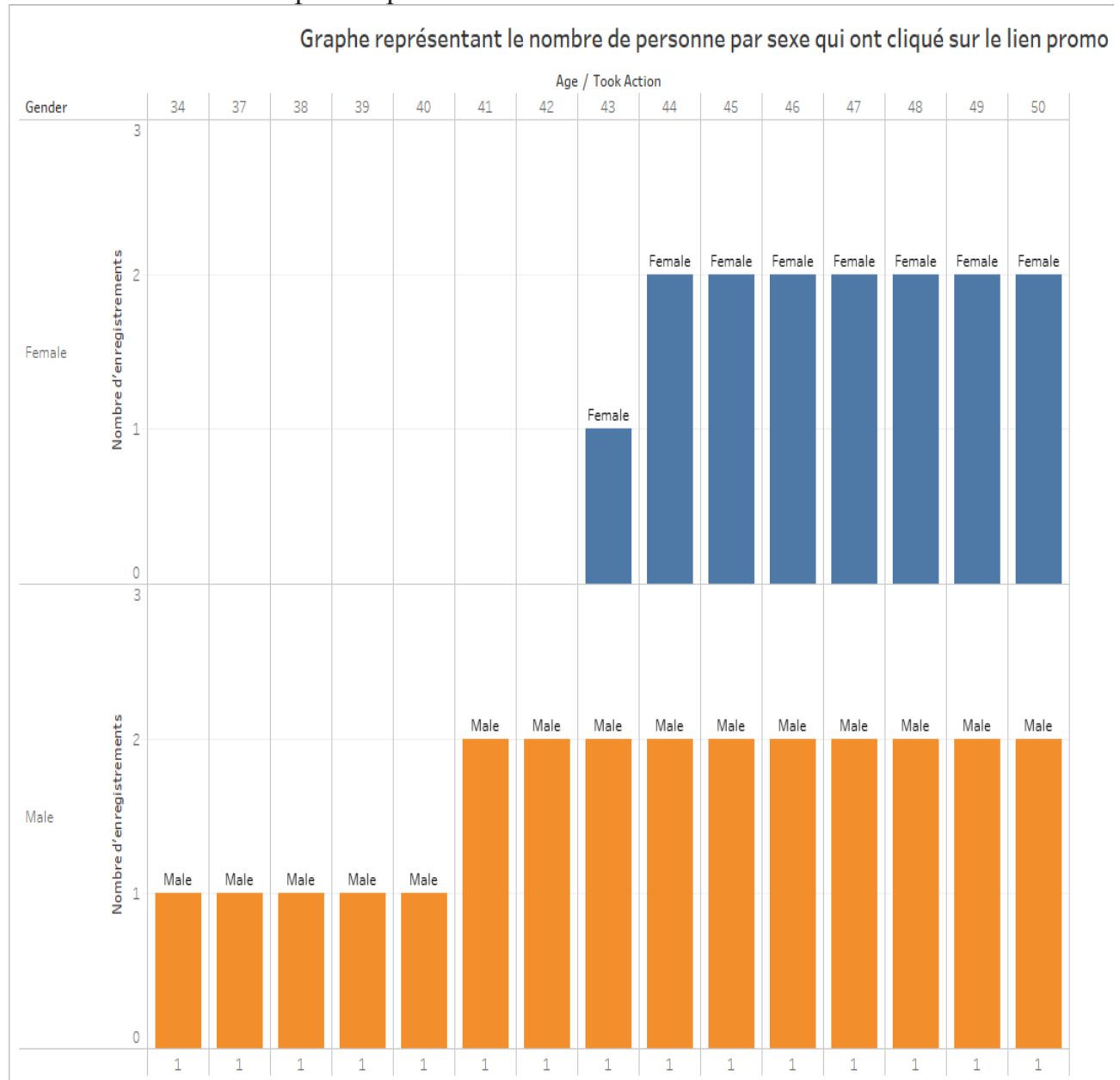
Nombre de cas 'correctement prédis' = 94 (94,0%)

f(beta'x) à la moyenne des variables indépendantes = 0,088

Test du ratio de vraisemblance: Chi-deux(2) = 104,989 [0,0000]

Mon modèle Gretl me permet de savoir si l'âge et le sexe jouait un rôle dans ce sujet. Nous pouvons interpréter ces données de cette manière: si nous prenons le variable âge en gardant toutes les autres variables constantes, **nous pouvons dire** que pour deux femmes à un an d'intervalle, **la femme la plus âgée sera plus susceptible d'ouvrir ce lien** (car le chiffre coéefficient est positif). De même pour la variable "female", on peut dire qu'**une femme est moins susceptible de cliquer** sur le lien promotionnel **que l'homme** (Chiffre coéefficient négatif).

En effet, grâce à ce dernier graphe nous pouvons dire que les femmes âgées de moins de 43 ans n'ont pas cliqué sur le lien.



Conclusion

Grâce à notre travail en amont, nous pouvons **prévoir, donner des indications** à l'entreprise, tirer des enseignements afin d'**augmenter le nombre de clics sur l'offre promotionnelle**:

- **Envoyez ce lien à plus d'hommes**, car nous avons vu que le genre a un impact sur le clic.
- **Envoyez ce lien aux femmes plus âgées(43 ans et +)**.
- **Retravailler cette promotion pour plaire aux femmes de moins de 43 ans**