

Zamano est un site commercial qui a 2 types d'utilisateur : les utilisateurs « *basiques* » ne payent pas d'abonnements et les utilisateurs « *premiums* » payent un abonnement pour bénéficier de fonctionnalités supplémentaires.

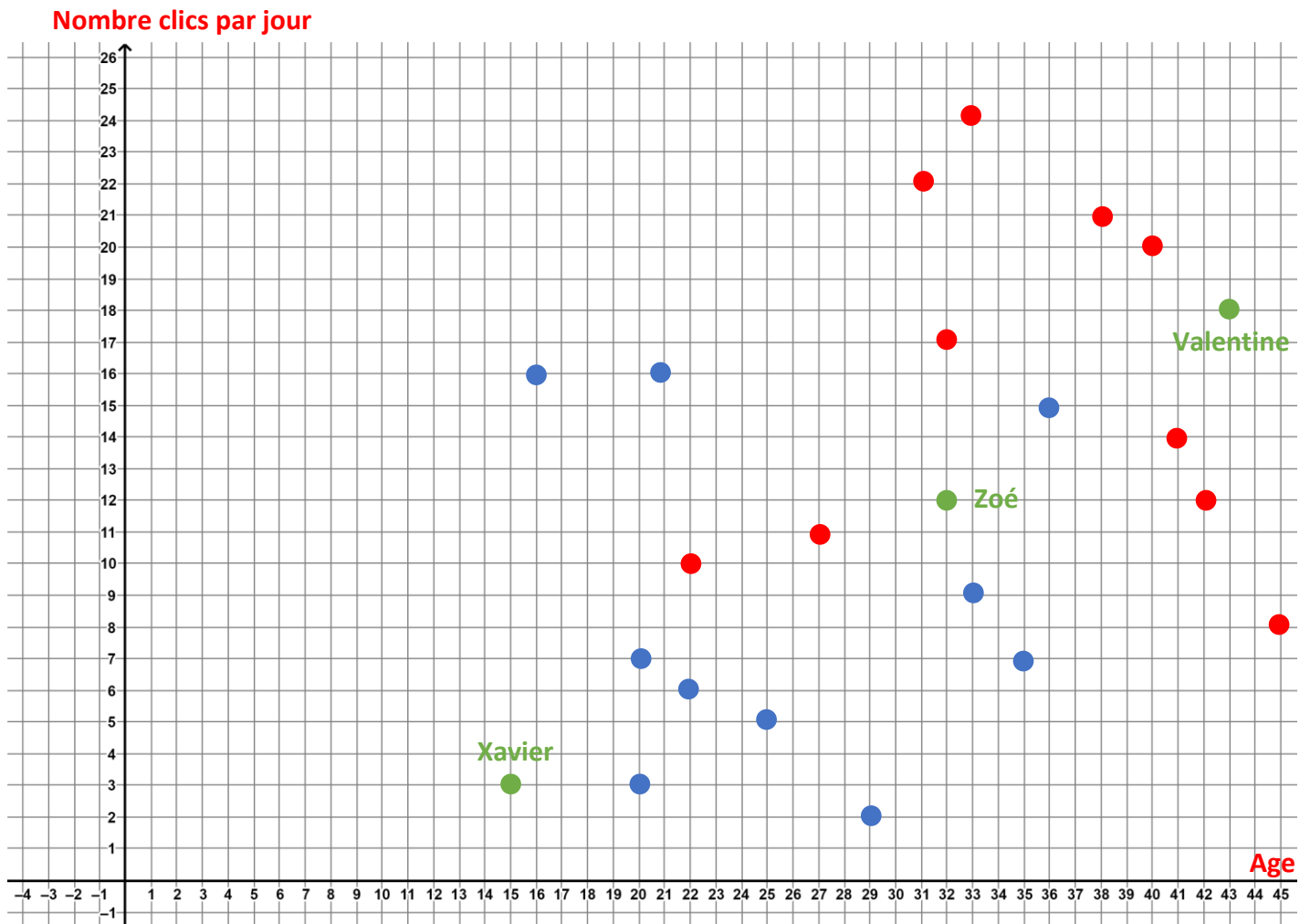
Ce site souhaite anticiper les recettes du mois suivant en effectuant des prédictions sur le type d'abonnement que vont choisir les utilisateurs.

Voici les renseignements recueillis sur un échantillon :

<i>Nom</i>	<i>Age</i>	<i>Nombre de clics par jour</i>	<i>Type</i>
Anaïs	16	16	Basique
Bilal	20	3	Basique
Chloé	20	7	Basique
David	21	16	Basique
Etienne	22	10	Premium
Farida	22	6	Basique
Gaël	25	5	Basique
Hermione	27	11	Premium
Isidore	29	2	Basique
Janine	31	22	Premium
Karim	32	17	Premium
Léa	33	9	Basique
Mathieu	33	24	Premium
Noé	35	7	Basique
Olivier	36	15	Basique
Phong	38	21	Premium
Quentin	40	20	Premium
Riad	41	14	Premium
Sophie	42	12	Premium
Thomas	45	8	Premium

1. Travail préparatoire sur papier

- 1- Représenter les données de l'échantillon sur le graphique qui suit avec en abscisse, l'âge et en ordonnée, le nombre de clics : utiliser des points bleus pour les utilisateurs *Basiques* et des points rouges pour les utilisateurs *Premiums*.



- 2- Trois nouveaux utilisateurs s'inscrivent sur le site : Valentine, 43 ans qui effectue en moyenne 18 clics par jour ; Xavier, 15 ans, qui effectue en moyenne 3 clics par jour et Zoé, 32 ans, qui effectue en moyenne 12 clics par jour.
- A l'aide du graphique, prédire le type d'abonnement que choisira Valentine
D'après le graphe, on peut dire que Valentine prendra un abonnement « premium »
 - Prédire le type d'abonnement que choisira Xavier
D'après le graphe, on peut dire que Xavier prendra un abonnement « basique »
 - Expliquer pourquoi le graphique seul ne permet pas de prédire le type d'abonnement que choisira Zoé
 - Le point correspondant à Zoé se situe à la frontière de la zone « basique » et « premium ». Il est donc difficile de prévoir le type d'abonnement que choisira Zoé
- 3- La société Zamano décide d'utiliser la formule suivante pour calculer la distance entre 2 personnes :
 $d(\text{personne1}, \text{personne2}) = |\text{âge1} - \text{âge2}| + |\text{clics1} - \text{clics2}|$.
 Par exemple, la distance entre Zoé et David sera : $d = |32 - 21| + |12 - 16| = 11 + 4 = 15$
 Avec cette distance, quels sont les 5 plus proches voisins de Zoé ? En se basant sur les abonnements choisis par ces 5 plus proches voisins, prédire le type d'abonnement que Zoé choisira.

Nom	$ Age1-Age2 $	$ Click1-Click2 $	d	Type
Anaïs	$ 16-32 =16$	$ 16-12 =4$	20	Basique
Bilal	$ 20-32 =12$	$ 3-12 =9$	21	Basique
Chloé	$ 20-32 =12$	$ 7-12 =5$	17	Basique
David	$ 21-32 =11$	$ 16-12 =4$	15	Basique
Etienne	$ 22-32 =10$	$ 10-12 =2$	12	Premium
Farida	$ 22-32 =10$	$ 6-12 =6$	16	Basique
Gaël	$ 25-32 =7$	$ 5-12 =7$	14	Basique
Hermione	$ 27-32 =5$	$ 11-12 =1$	6	Premium
Isidore	$ 29-32 =3$	$ 2-12 =10$	13	Basique
Janine	$ 31-32 =1$	$ 22-12 =10$	11	Premium
Karim	$ 32-32 =1$	$ 17-12 =5$	6	Premium
Léa	$ 33-32 =1$	$ 9-12 =3$	4	Basique
Mathieu	$ 33-32 =1$	$ 24-12 =12$	13	Premium
Noé	$ 35-32 =3$	$ 7-12 =5$	8	Basique
Olivier	$ 36-32 =4$	$ 15-12 =3$	7	Basique
Phong	$ 38-32 =6$	$ 21-12 =9$	15	Premium
Quentin	$ 40-32 =8$	$ 20-12 =8$	16	Premium
Riad	$ 41-32 =9$	$ 14-12 =2$	11	Premium
Sophie	$ 42-32 =10$	$ 12-12 =0$	10	Premium
Thomas	$ 45-32 =13$	$ 8-12 =4$	11	Premium

En se basant sur les abonnements choisis par ces 5 plus proches voisins, on constate que l'abonnement *Basique* est majoritaire. On pourrait ainsi prédire que Zoé choisira un abonnement *Basique*.

2. Implémentation de l'algorithme

```

def distance(connu,mystere):
    """
    Arguments :
        connu (liste) et mystere (liste)
    Renvoie la distance (float) entre "connu" et "mystère" : distance = |age1-age2| + |click1-click2|
    """
    d = abs(connu[1]-mystere[1]) + abs(connu[2]-mystere[2])
    return d

def tri_selection(liste,mystere,k) :
    """
    Arguments :
        liste : (liste) dont les éléments sont des listes du type
                [prenom (string) , Age (int) , Nb_click (int)]
        mystere : (liste) de type [prenom (string) , Age (int) , Nb_click (int)].
                caractéristiques de la personne dont on veut prédire le type d'abonnement
        k (int) : nombre de voisins souhaités
    Modifie l'ordre de liste, on mettant en début de liste et dans l'ordre, les k éléments pour lesquels la distance
    avec mystere est la plus petite. On utilise un algo du type "tri par sélection".
    """
    if k >= len(liste)-1 : k = len(liste)-1
    for i in range(k) :
        jMin = i
        for j in range(i+1,len(liste)-1) :
            if distance(liste[j],mystere) < distance(liste[jMin],mystere) :
                jMin = j
        liste[i],liste[jMin] = liste[jMin],liste[i]
    return k

def kProchesVoisins(echantillon , mystere) :
    """
    Arguments :
        echantillon : liste dont les éléments sont des listes du type
                [prenom (string) , Age (int) , Nb_click (int)]
        mystere (liste) de type [prenom (string) , Age (int) , Nb_click (int)].
                caractéristiques de la personne dont on veut prédire
                le type d'abonnement
    Ce code retourne le type d'abonnement majoritaire autour de "mystere".
    (algorithme des k plus proches voisins)
    """

    k = int(input('nombre de voisins : '))
    tri_selection(echantillon,mystere,k)
    n = 0
    for i in range(k) :
        if echantillon[i][3]=="Basique" : n += 1
    if n%2 == 0 :
        if n < k/2 : return "Premium"
        elif n == k/2 : return "Indéterminé"
        else : return "Basique"
    else :
        if n <= k//2 : return "Premium"
        else : return "Basique"

# Main
echantillon = [
["Anaïs", 16, 16, "Basique"],
["Bilal", 20, 3, "Basique"],
["Chloé", 20, 7, "Basique"],
["David", 21, 16, "Basique"],
["Etienne", 22, 10, "Premium"],
["Farida", 22, 6, "Basique"],
["Gaël", 25, 5, "Basique"],
["Hermione",27, 11, "Premium"],
["Isidore", 29, 2, "Basique"],
["Janine", 31, 22, "Premium"],
["Karim", 32, 17, "Premium"],
["Léa", 33, 9, "Basique"],
["Mathieu", 33, 24, "Premium"],
["Noé", 35, 7, "Basique"],
["Olivier", 36, 15, "Basique"],
["Phong", 38, 21, "Premium"],
["Quentin", 40, 20, "Premium"],
["Riad", 41, 14, "Premium"],
["Sophie", 42, 12, "Premium"],
["Thomas", 45, 8, "Premium"],
]
mystere = ["Zoé",32,12]

prediction = kProchesVoisins(echantillon , mystere)
print(f"Prévision pour {mystere[0]} : {prediction}")

```

A l'exécution, on obtient :

```

>>> (executing file "zamano_voisins.py")
nombre de voisins : 5
Prévision pour Zoé : Basique

```