
Livrable

Python Project

Auteurs:
Alec DIAMIDIA

Enseignant:
Matthieu LARBOULLET

Sommaire

1. Introduction.....	3
1.1. Contexte.....	3
1.2. Problématique.....	3
2. Analyse du site et des données.....	4
2.1. Description du site.....	4
2.2. Structure des données ciblées.....	6
3. Code principal.....	7
3.1. Découverte de l'API.....	7
3.2. Les fonctions de récupération.....	8
3.3. Fonctions de nettoyage.....	8
3.4. Fonction de nettoyage principal.....	10
3.5. Les autres fonctions.....	11
4. Problématiques rencontrées.....	12
5. Analyse des données & Visualisations.....	13
5.1. Qui court ? , Profil démographique.....	13
5.2. Comment courent-ils ? : Performance.....	17
5.3 Analyse continentale.....	19
6. Conclusion.....	20
6.1. Réponse à la problématique.....	20
6.2. Limites et améliorations possibles.....	22
7. Annexes.....	23

1. Introduction

1.1. Contexte

Le marathon est aujourd'hui l'une des disciplines sportives les plus pratiquées au monde. Le Schneider Electric Marathon de Paris 2025 représente l'un des événements les plus emblématiques du calendrier international, rassemblant chaque année des dizaines de milliers de coureurs de tous horizons.

Au-delà du sport, ces événements génèrent une masse de données démographiques et de performance inexploitées publiquement, pourtant précieuses pour de nombreux acteurs : marques d'équipement sportif, organisateurs d'événements, fédérations, applications de coaching ou encore acteurs de la nutrition sportive.

Ce projet consiste à collecter, nettoyer et analyser les données des 50 001 finishers du Marathon de Paris 2025, via scraping de l'API officielle du site de résultats results.timeto.com. Il simule la démarche qu'une organisation pourrait adopter pour mieux comprendre les profils de ses coureurs ou de sa cible.

Pourquoi ce site ?

- Données officielles, structurées et homogènes (même format pour chaque coureur)
- Échantillon massif : 50 001 coureurs → analyses statistiques robustes
- Variables riches : identité, nationalité, catégorie FFA, temps, allure, vitesse
- Données récentes (avril 2025) et représentatives du running européen
- Possibilité d'utiliser les données récoltées afin d'alimenter des enquêtes ou datasets d'organismes orientés running afin d'avoir des insights actionnables.

Type de données disponibles :

Pour chaque coureur :

- classement général
- nom & prénom
- sexe
- catégorie FFA
- nationalité
- temps officiel
- allure (min/km) & vitesse moyenne (km/h).

1.2. Problématique

La question centrale

Le Marathon de Paris génère chaque année une masse considérable de données publiques sur ses participants. Ces données, bien qu'accessibles, ne

font l'objet d'aucune analyse à destination des acteurs du secteur du running. C'est en ce sens que je me suis dit qu'il fallait les utiliser pour les rendre plus parlantes et utiles pour des organismes de running ou autres.

Notre projet part de ce constat pour formuler la problématique suivante :

"Quel est le profil démographique et le niveau de performance des finishers du Marathon de Paris 2025, et quels enseignements ces données offrent-elles aux acteurs du secteur du running ?"

Cette question se décomposera en 5 sous-questions analytiques qui structurent notre analyse :

- Quelle est la répartition hommes/femmes parmi les finishers ?
- Quelles tranches d'âge sont les plus représentées ?
- Quelles nationalités et quels continents dominent la participation ?
- Comment les caractéristiques démographiques influencent-elles la performance ?
- Quels grands segments de coureurs se dégagent de l'analyse ?

Hypothèses de départ

Avant de collecter et d'analyser les données, nous avons formulé 4 hypothèses basées sur notre connaissance générale du marathon de masse et de la culture running en France.

H1 : les hommes représentent une large majorité des finishers

Le marathon reste historiquement une discipline masculine. Malgré la progression constante de la participation féminine ces dernières années, nous supposons que les hommes représentent encore environ les trois quarts des finishers.

H2 : la tranche 18-34 ans serait la plus représentée

Courir un marathon nécessite un investissement physique et temporel important. Nous supposons que les coureurs jeunes et actifs, dont le mode de vie permet un entraînement soutenu, constituent la part la plus importante des participants.

H3 : les coureurs africains dominent le haut du classement

Les nations est-africaines, notamment le Kenya et l'Éthiopie sont mondialement reconnues pour leur domination en athlétisme de fond. Nous supposons que cette supériorité se reflètera dans les premières places du classement parisien.

H4 : la performance décline progressivement avec l'âge

La physiologie humaine implique une dégradation des capacités cardio-respiratoires et musculaires avec l'âge. Nous supposons que cette réalité biologique se traduit par des temps de course systématiquement plus longs dans les tranches d'âge avancées.

Ces hypothèses seront confrontées aux données réelles dans la section 5, et leur validation ou invalidation alimentera les conclusions du rapport.

2. Analyse du site et des données

2.1. Description du site

Le site results.timeto.com est la plateforme officielle de publication des résultats du Marathon de Paris. Il présente les données sous forme d'un tableau dynamique : chaque coureur est listé par classement général, avec des filtres disponibles par catégorie, nationalité ou numéro de dossard.

Structure générale : À première vue, le tableau semble paginé (bouton "page suivante" visible). Cependant, l'URL de la page ne change pas lors de la navigation : le contenu est chargé dynamiquement via JavaScript (AJAX), sans rechargement de page.

Pagination : Il n'existe pas de pagination classique au sens URL. Le scraping HTML traditionnel (via requests + BeautifulSoup) s'est révélé inutilisable directement, car le contenu est rendu côté client. La solution a été d'identifier l'API sous-jacente utilisée par le site, via l'outil Inspecteur réseau du navigateur (F12 → Network).

2.2. Structure des données ciblées

Les variables issues du json tiré de la page du site sur lequel le scraping a été effectué :

Variable	Description	Type	Exemple
<code>place</code>	Classement général du coureur	int	1, 342, 12 847
<code>nom</code>	Nom de famille	string	"Kipchoge"
<code>prenom</code>	Prénom	string	"Eliud"
<code>sexe</code>	Sexe brut (code API)	string	"M" ou "F"
<code>categorie</code>	Catégorie FFA officielle	string	"SEH", "M1F"
<code>nationalite</code>	Code pays ISO (2 lettres)	string	"FR", "KE"
<code>temps</code>	Temps officiel de course	string	"03:45:22"
<code>allure</code>	Allure moyenne en min/km	string	"05:21"
<code>vitesse_moy</code>	Vitesse moyenne en km/h	float	11.2

Les variables custom créées par nos fonctions :

Variable	Description	Type	Construit depuis
<code>continent</code>	Continent du coureur	string	<code>nationalite</code> via <code>search_continent()</code>
<code>sexe_custom</code>	Sexe lisible	string	<code>categorie</code> via <code>runner_cat()</code>
<code>tranche_age</code>	Tranche d'âge regroupée	string	<code>categorie</code> via <code>runner_cat()</code>
<code>cust_category</code>	Tranche + Sexe combinés	string	<code>categorie</code> via <code>runner_cat()</code>

Les 4 variables calculées n'existent pas dans les données brutes de l'API. Elles ont été construites lors de l'étape de nettoyage pour rendre les données directement exploitables dans les visualisations.

3. Code principal

Le projet est structuré en 3 étapes : récupération → nettoyage → analyse.

3.1. Découverte de l'API

Plutôt que de scraper le HTML (impossible ici), l'API du site a été identifiée via l'onglet Network des outils de développement Chrome (F12). En filtrant les requêtes XHR lors du chargement du tableau, une requête GET vers l'endpoint suivant a été découverte :

GET <https://results.timeto.com/frontend/results/?act=GetRaceData&rquest=658>

Cette API retourne directement 50 001 coureurs en une seule requête JSON, sans pagination à gérer.

3.2. Les fonctions de récupération

- La fonction **recuperer_donnees_api()**

```
def recuperer_donnees_api():
    cookies = {
        'PHPSESSID': 'ccc23f67e61ecf906bda9f736837f673',
        '_ga': 'GA1.1.884142582.1770825792',
        '_ga_VGV5FC38P2': 'GS2.1.s1770828057$02$g0$t1770828057$j60$l0$h0',
    }

    headers = {
        'referer': 'https://results.timeto.com/schneider_electric_marathon_de_paris_2025/',
        'user-agent': 'Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36',
    }

    params = {
        'act': 'GetRaceData',
        'rquest': '658',
    }

    response = requests.get(
        'https://results.timeto.com/frontend/results/',
        params=params,
        cookies=cookies,
        headers=headers
    )

    data = json.loads(response.text)

    print(f"Status: {response.status_code}")

    print(f"{len(data):,} coureurs disponibles")

    # Sauvegarder les données brutes complètes
    with open('data_brute_complete.json', 'w', encoding='utf-8') as f:
        json.dump(data, f, indent=2, ensure_ascii=False)

    with open('50k_coureurs.json', 'r', encoding='utf-8') as f:
        data = json.load(f)

    return data
```

Cette fonction nous a permis de contacter l'API du site et récupérer les données brutes de tous les coureurs. Elle retourne une liste de dictionnaires JSON, un par coureur.

3.3. Fonctions de nettoyage

- La fonction **search_continent()**

```
def search_continent(nationalite):
    if nationalite in ['KE', 'ET', 'DJ', 'MA', 'TZ', 'UG', 'ER', 'ZA', 'NG', 'CI', 'SN']:
        return 'Afrique'
    elif nationalite in ['FR', 'ES', 'IT', 'GB', 'DE', 'BE', 'NL', 'PT', 'PL', 'CH', 'AT', 'SE', 'NO']:
        return 'Europe'
    elif nationalite in ['JP', 'CN', 'IN', 'KR', 'TH', 'VN', 'MY', 'ID', 'PH']:
        return 'Asie'
    elif nationalite in ['US', 'CA', 'BR', 'MX', 'AR', 'CO', 'CL', 'PE']:
        return 'Amérique'
    elif nationalite in ['AU', 'NZ']:
        return 'Océanie'
    else:
        return 'Autre'
    return nationalite
```


Cette fonction nous a permis de récupérer et catégoriser un code pays (ex: "FR", "KE") et retourner le continent correspondant. Elle est appelée pour chaque coureur pendant le nettoyage des données.

Exemple :

nationalite = "KE"

↓

"KE" in liste Afrique ? → OUI → return "Afrique"

nationalite = "FR"

↓

"FR" in liste Afrique ? → NON

"FR" in liste Europe ? → OUI → return "Europe"

nationalite = "IS"

↓

Aucune liste ne correspond → return "Autre"

Pourquoi ? Car le continent n'est pas fourni par l'API. Cette fonction (avec des conditions en son sein) permet d'éviter toute dépendance à une API externe et d'avoir un regroupement macro pour les visualisations.

- La fonction : **runner_cat()**

Cette fonction prend une catégorie FFA brute (ex: "M1H", "SEF") et la transforme en 3 informations lisibles : le sexe, la tranche d'âge, et une combinaison des deux. Elle sert à rendre les données exploitables pour les visualisations.

```
def runner_cat(categorie):
    # je veux me baser sur le sexe et les catégories pour avoir les tranches d'âges => avoir une visu plus macro (large)
    # Déterminer le sexe
    if categorie.endswith('H'):
        sexe = "Homme"
    elif categorie.endswith('F'):
        sexe = "Femme"
    else:
        sexe = "Inconnu"

    # Déterminer la caté via et segmenter par tranche d'âge
    if categorie in ['JUH', 'ESH', 'SEH', 'JUF', 'ESF', 'SEF']:
        tranche = "18-34 ans - [Jeune]"
    elif categorie in ['M0H', 'M1H', 'M0F', 'M1F']:
        tranche = "35-44 ans - [Jeune +]"
    elif categorie in ['M2H', 'M3H', 'M2F', 'M3F']:
        tranche = "45-54 ans - [Mid - Senior]"
    elif categorie in ['M4H', 'M5H', 'M4F', 'M5F']:
        tranche = "55-64 ans - [Senior]"
    elif categorie in ['M6H', 'M7H', 'M8H', 'M9H', 'M10H', 'M6F', 'M7F', 'M8F', 'M9F']:
        tranche = "65+ ans - [Senior +]"
    else:
        tranche = "Autre"

    return {
        'sexe_custom': sexe,
        'tranche_age': tranche,
        'cust_category': f"{tranche} ____ ({sexe})"
    }
```

Pourquoi ? Les catégories FFA brutes (SEH, M1H, M2F...) sont trop granulaires pour des visualisations lisibles. Le regroupement en 5 tranches permet des graphiques clairs et comparables. Et pour avoir une vision plus macro et avoir une étude extrapolée pour chaque échantillon.

3.4. Fonction de nettoyage principal

```

def clean_data(nombre_max=52000):
    coureurs_top50K = data[:52000]
    runners_clean = []
    for coureur in coureurs_top50K:
        # Extraire les infos de base
        coureur_clean = {
            'place': coureur['computedGeneralRanking'],
            'nom': coureur['lastName'],
            'prenom': coureur['firstName'],
            'sexe': coureur['sex'],
            'categorie': coureur['category'],
            'nationalite': coureur['nationality'],
            'temps': coureur['computedRealTime'],
            'allure': coureur['computedPace'],
            'vitesse_moy': coureur['computedAverageSpeed'],
        }

        # Ajouter continent
        coureur_clean['continent'] = search_continent(coureur['nationality'])

        # Ajouter catégorisation personnalisée
        cat_info = runner_cat(coureur['category'])
        coureur_clean['cust_category'] = cat_info['cust_category']
        coureur_clean['sexe_custom'] = cat_info['sexe_custom']
        coureur_clean['tranche_age'] = cat_info['tranche_age']

        runners_clean.append(coureur_clean)

    print(f"{len(runners_clean):,} coureurs nettoyés\n")

    return runners_clean

```

C'est la **fonction centrale du projet** : elle prend les données brutes de l'API (50 001 dictionnaires complexes) et les transforme en données propres, structurées et enrichies , prêtes pour l'analyse.

Résumé visuel

data (brut, 50 001 coureurs, ~40 champs chacun)



Découpage → data[:52000]



BOUCLE pour chaque coureur :

- └─ Extraction de 9 champs utiles + renommage
- └─ search_continent() → +1 colonne
- └─ runner_cat() → +3 colonnes
- └─ append() → ajout à runners_clean



runners_clean (propre, 50 001 coureurs, 13 champs chacun)



return → transmis à sauvegarder_json() et sauvegarder_csv()

3.5. Les autres fonctions

Trois fonctions utilitaires complètent le script : sauvegarder_json() et sauvegarder_csv() exportent les données nettoyées dans leurs formats respectifs pour les rendre exploitables dans d'autres outils (Google Sheets) pour l'analyse présentent des lignes plus bas. afficher_aperçu() permet de vérifier visuellement les premières lignes dans le terminal pendant le développement.

Sans oublier la fonction principale :

```
if __name__ == "__main__":
    cookies = {
        'PHPSESSID': 'ccc23f67e61ecf906bda9f736837f673',
        '_ga': 'GA1.1.884142582.1770825792',
        '_ga_VGV5FC38P2': 'GS2.1.s1770828057$02$g0$t1770828057$j60$l0$h0',
    }

    headers = {
        'referer': 'https://results.timeto.com/schneider_electric_marathon_de_paris_2025/',
        'user-agent': 'Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36',
    }

    params = {
        'act': 'GetRaceData',
        'rquest': '658'
    }
    response = requests.get('https://results.timeto.com/frontend/results/', params=params, cookies=cookies, headers=headers)
    data = json.loads(response.text)

    # Les 50 000 nettoyés
    coureurs = clean_data(nombre_max=57000)

    # Sauvegarder
    sauvegarder_json("50k_coureurs_propres.json")

    # Afficher aperçu
    afficher_aperçu(nombre=101)

    #pour avoir le csv
    sauvegarder_csv("50k_coureurs_propres.csv")
```

4. Problématiques rencontrées

Problème 1 , Contenu chargé dynamiquement (JavaScript)

Problème : requests + BeautifulSoup ne récupèrent aucune donnée

Cause : le tableau est rendu en JavaScript côté client, le HTML brut est vide

Solution : Identification de l'API via l'onglet Network (F12) → scraping direct du JSON

Problème 2 : Pas de pagination à gérer

Problème : l'URL ne change pas lors de la navigation entre pages

Cause : chargement dynamique sans rechargement de page

Solution : l'API retourne tous les coureurs en une seule requête → pas de boucle de pagination nécessaire

Problème 3 : Variable nombre_max ignorée

Problème : la fonction clean_data découpait toujours à [:52000] quelle que soit la valeur passée

Cause : le paramètre nombre_max était déclaré mais non utilisé dans le slicing

Solution : remplacement de data[:52000] par data[:nombre_max]

Problème 4 : Ligne dupliquée dans la boucle

Problème : certains coureurs étaient comptés deux fois dans la liste finale

Cause : runners_clean.append(coureur_clean) appelé deux fois (mauvaise indentation)

Solution : suppression de la ligne dupliquée

Problème 5 : le sexe toujours "Homme" dans runner_cat

Problème : toutes les femmes étaient catégorisées "Homme"

Cause : la variable sexe était hardcodée à "Homme" sans condition

Solution : détection automatique via categorie.endswith('H') / endswith('F')

Problème 6 , sauvegarder_csv ignorait le paramètre nom_fichier

Problème : le fichier CSV était toujours sauvegardé sous le même nom fixe

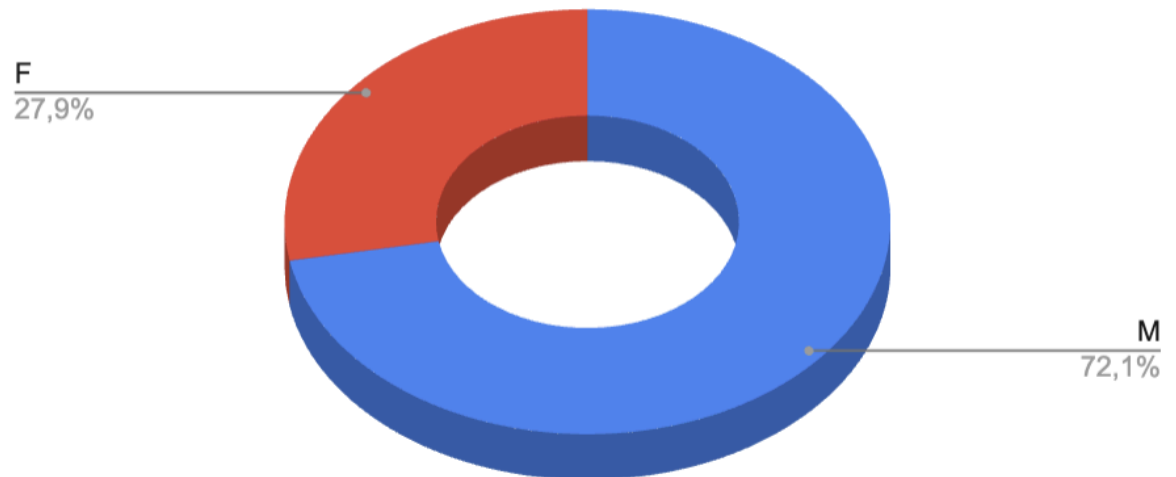
Cause : le nom était hardcodé dans la fonction malgré un paramètre déclaré

Solution : utilisation de nom_fichier comme argument de df.to_csv()

5. Analyse des données & Visualisations

5.1. Qui court ? , Profil démographique

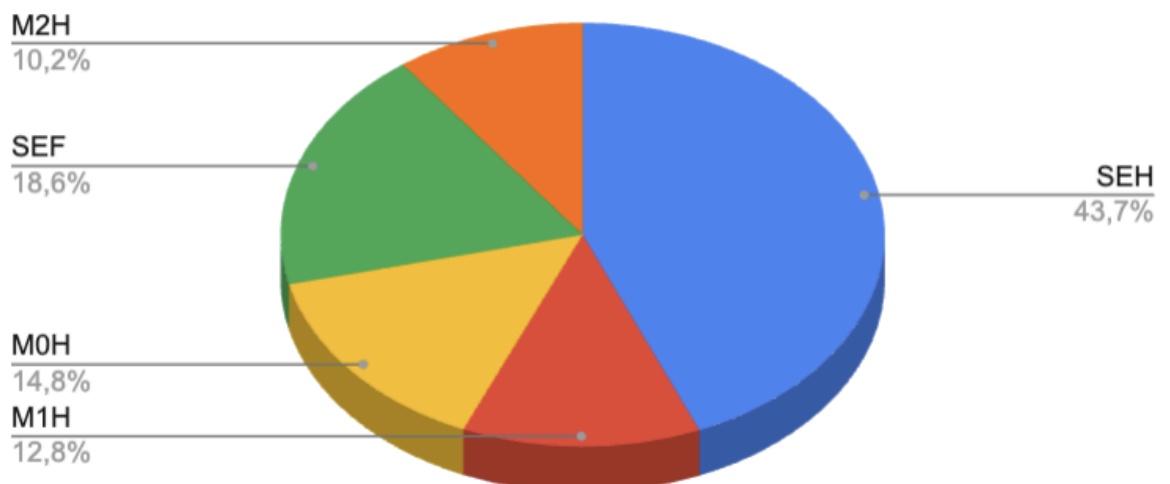
La répartition par Sexe (F/H)



Graphique 1 : Répartition par sexe

Le Marathon de Paris 2025 est largement dominé par les hommes : 72,1% des 50 001 finishers sont masculins, contre 27,9% de femmes. Cette répartition confirme notre hypothèse H1 et reflète une tendance structurelle du marathon de masse, où les femmes restent historiquement sous-représentées malgré une progression constante de leur participation ces dernières années. Pour les acteurs du secteur du running, ce déséquilibre représente un levier de croissance clair : le segment féminin constitue un marché sous-exploité à fort potentiel.

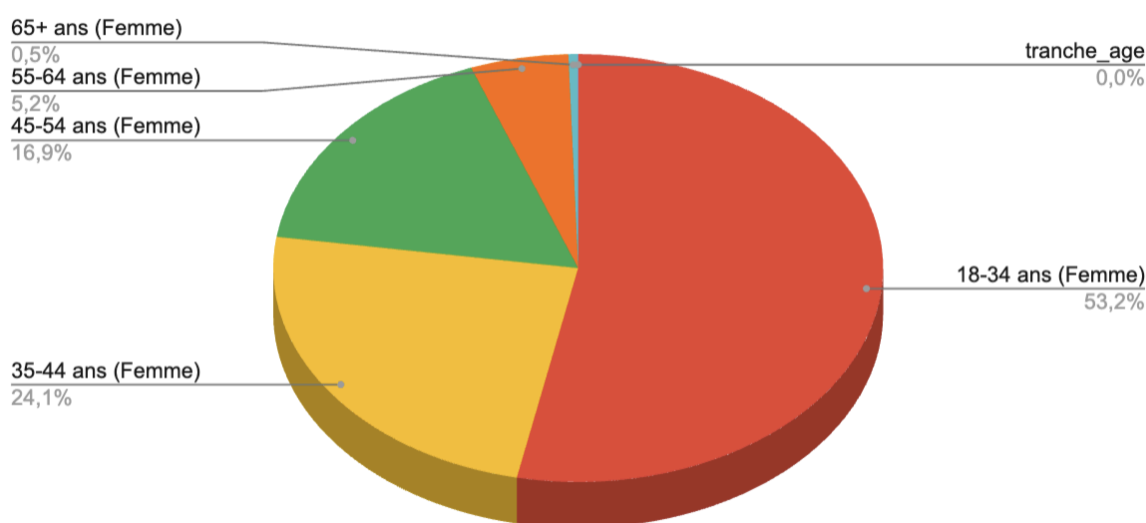
Top 5 des catégories



Graphique 2 : le top 5 des catégories FFA

L'analyse des catégories officielles FFA révèle que la SEH (Senior Homme, 18-39 ans) concentre à elle seule 43,7% des finishers du top 5 des catégories, suivie de la SEF (Senior Femme) à 18,6%, puis des catégories Masters masculins (M0H : 14,8% ; M1H : 12,8% ; M2H : 10,2%). Ce classement confirme que les coureurs jeunes à intermédiaires dominent massivement l'événement. La présence de trois catégories Masters masculines dans le top 5 souligne également l'ancrage du marathon chez les hommes de 35 à 55 ans.

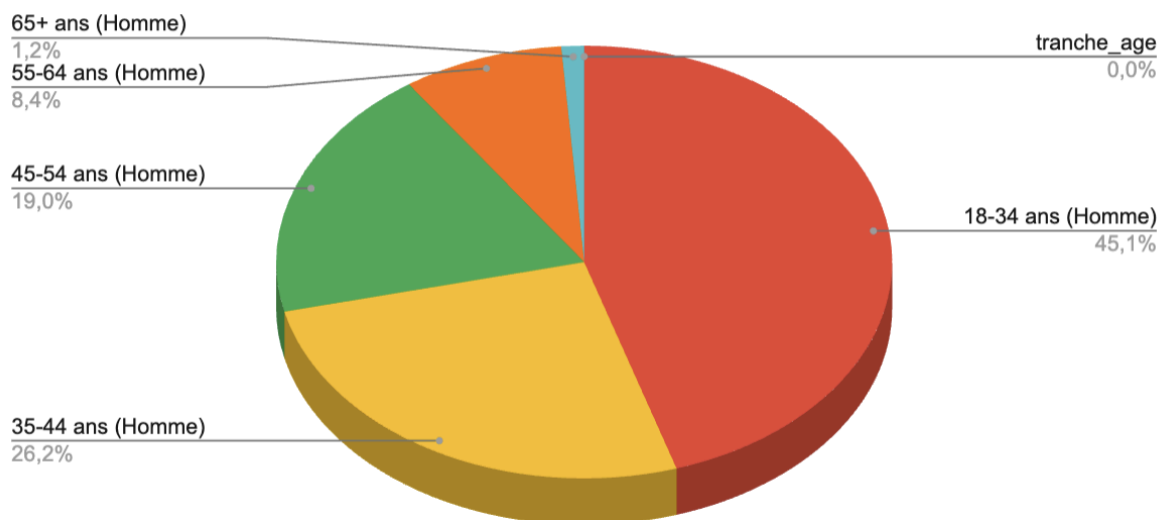
Répart. tranches d'âge (Femme)



Graphique 3 : la répartition par tranche d'âge (Femmes)

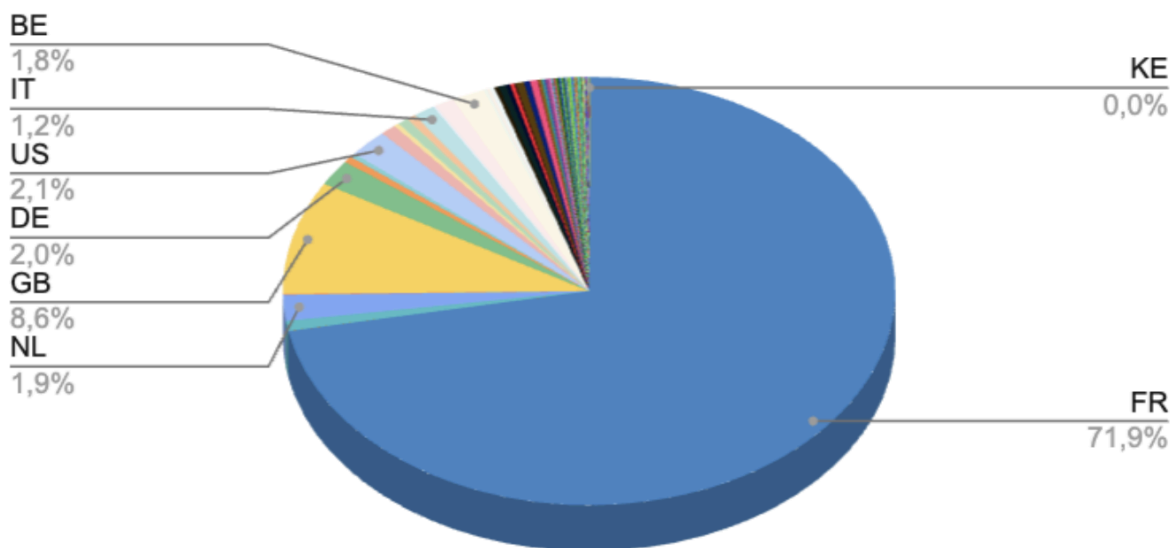
Le profil féminin est sensiblement plus jeune que le profil masculin : 53,2% des finisseuses ont entre 18 et 34 ans, contre 45,1% chez les hommes. Les 35-44 ans représentent 24,1%, les 45-54 ans 16,9%, les 55-64 ans 5,2% et les 65+ seulement 0,5%. Cette structure d'âge indique que la pratique féminine du marathon est portée par une génération plus jeune, avec une présence plus faible dans les tranches seniors, ce qui peut s'expliquer par une entrée plus récente des femmes dans la course à pied longue distance.

Répart. tranches d'âge (Homme)



Graphique 4 : la répartition par tranche d'âge (Hommes)

Chez les hommes, la tranche 18-34 ans représente 45,1% des finishers masculins, devant les 35-44 ans (26,2%), les 45-54 ans (19,0%), les 55-64 ans (8,4%) et les 65+ (1,2%). La participation décroît logiquement avec l'âge, mais la solidité de la tranche 35-44 ans (plus d'un coureur sur quatre) confirme que les hommes actifs de la quarantaine constituent un segment central et fidèle de l'événement.



Graphique 5 : la répartition des pays représentés

La France domine très largement avec 71,9% des finishers, faisant du Marathon de Paris un événement avant tout domestique malgré son rayonnement international. Le Royaume-Uni arrive en deuxième position avec 8,6% des participants, une présence remarquable qui s'explique par la proximité géographique et la forte culture running britannique. Suivent l'Allemagne (2,0%), les États-Unis (2,1%), les Pays-Bas (1,9%), la Belgique (1,8%) et l'Italie (1,2%). Le Kenya (KE), bien que quasi-invisible en volume (<0,1%), se retrouve en tête du classement de performance, une illustration parfaite de la distinction entre volume de participation et niveau de performance.

L'analyse démographique des finishers du Marathon de Paris 2025 offre des enseignements directement exploitables pour tout acteur du marché du running.

- Sur la cible principale :
Le profil dominant : homme français de 18 à 44 ans constitue le cœur de marché à adresser en priorité. Les catégories SEH et M0H/M1H représentent à elles seules plus de 70% des catégories FFA les plus représentées. Une marque d'équipement, une application de coaching ou un organisateur d'événements a tout intérêt à concentrer ses efforts de communication sur cette tranche, avec un message axé sur la performance et la progression.
- Sur le segment féminin :
Avec seulement 27,9% de finisseuses, la femme reste sous-représentée mais son profil est particulièrement jeune (53,2% ont entre 18 et 34 ans). Ce segment est en croissance et constitue une opportunité de conquête claire : des produits et contenus spécifiquement pensés pour les femmes débutantes ou intermédiaires pourraient accélérer leur entrée dans la discipline.
- Sur la géographie :
La domination française (71,9%) combinée à la forte présence britannique (8,6%) dessine une carte de priorités géographiques nette : France d'abord, puis Royaume-Uni, Allemagne et Benelux pour une expansion européenne ciblée.

En résumé :

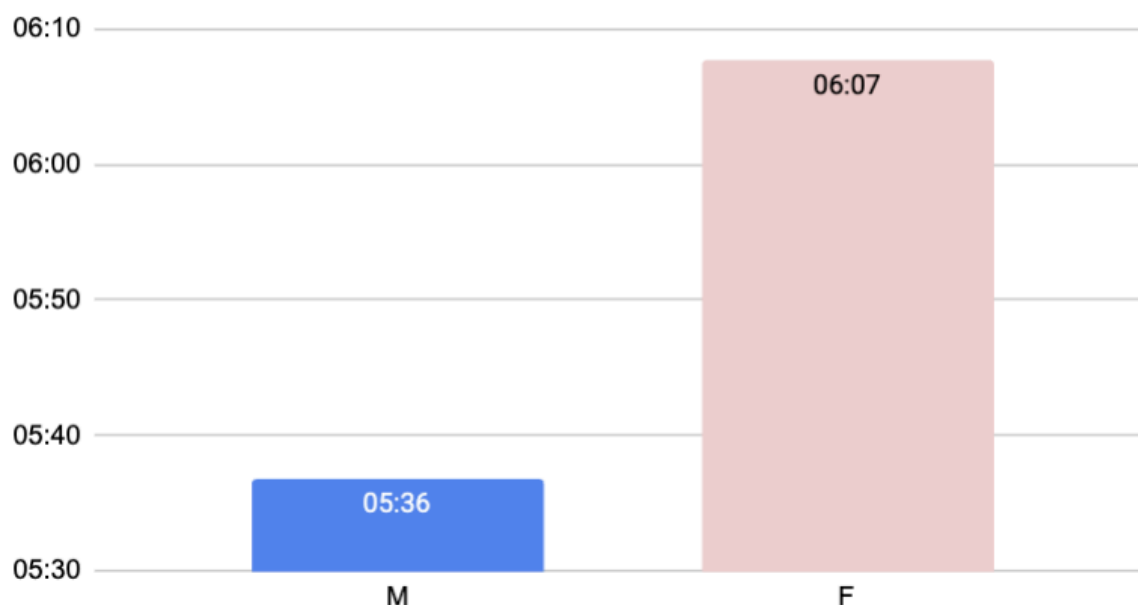
<u>Opportunité</u>	<u>Segment concerné</u>	<u>Action possible</u>
Cœur de marché	Homme FR, 18-44 ans	Communication performance & progression
Croissance	Femme, 18-34 ans	Offres d'entrée, contenu inclusif
Expansion	GB, DE, BE, NL	Campagnes digitales ciblées par pays

Fidélisation	55-64 ans, homme	Produits récupération & longévité
--------------	------------------	-----------------------------------

5.2. Comment courent-ils ? : Performance

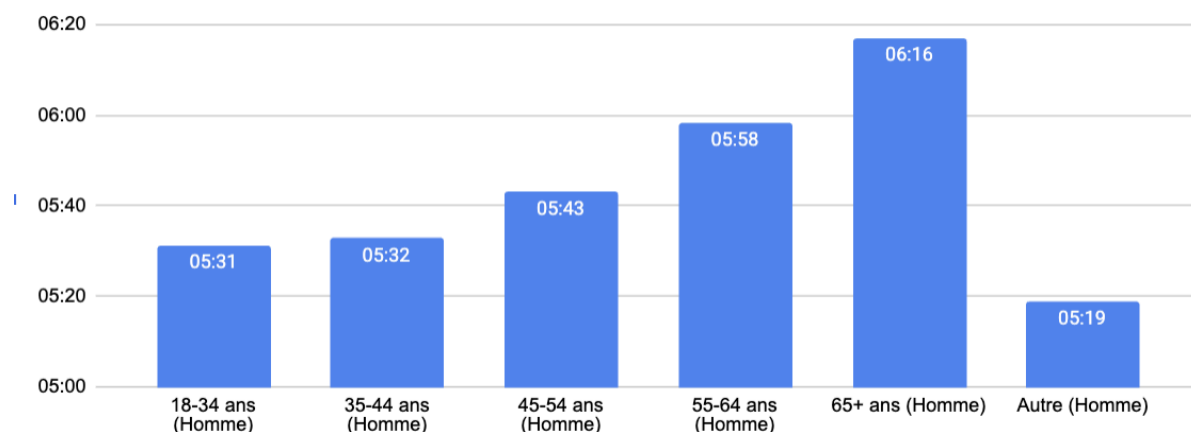
Graphique 1 : le temps moyen Hommes vs Femmes

Temps moyen H vs F



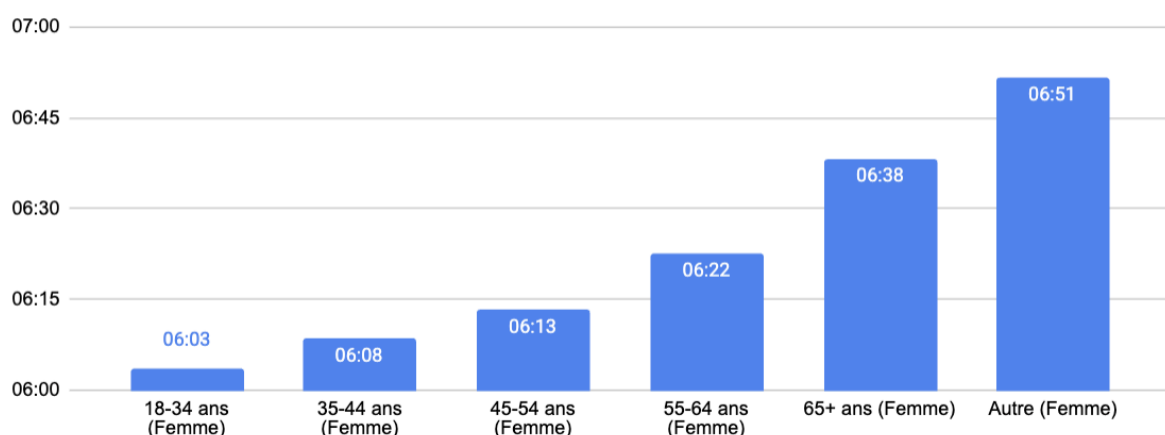
L'analyse de l'allure moyenne révèle un écart significatif entre les genres : les hommes courent en moyenne à 5:36 min/km contre 6:07 min/km pour les femmes, soit un écart de 31 secondes par kilomètre. Rapporté aux 42,195 km du marathon, cela représente environ 22 minutes d'écart sur le temps total, les hommes finissant en moyenne vers 3h57 contre 4h19 pour les femmes. Cet écart, constant et cohérent avec les données observées sur les autres grands marathons mondiaux, s'explique principalement par des facteurs physiologiques (masse musculaire, capacité cardiorespiratoire) et une moindre expérience moyenne des femmes dans la discipline.

Graphique 2 : la distribution des allures par tranche d'âge (Hommes)



Chez les hommes, les données révèlent un résultat surprenant : les 18-34 ans (5:31/km) et les 35-44 ans (5:32/km) affichent des allures quasi identiques, avec seulement 1 seconde d'écart par kilomètre. Ce plateau de performance entre les deux premières tranches indique que l'expérience et la régularité d'entraînement des coureurs de 35-44 ans compensent très efficacement l'avantage physiologique de la jeunesse. La dégradation devient ensuite plus nette : 45-54 ans à 5:43/km (+12 sec), 55-64 ans à 5:58/km (+27 sec), puis 65+ à 6:16/km (+45 sec). La catégorie "Autre" (5:19/km) correspond probablement à des coureurs hors-catégorie ou élite, ce qui explique leur allure plus rapide.

Graphique 3 : la distribution des allures par tranche d'âge (Femmes)



Chez les femmes, la dégradation avec l'âge est plus progressive et régulière : de 6:03/km pour les 18-34 ans à 6:38/km pour les 65+ ans, soit un écart de 35

secondes par kilomètre entre la tranche la plus jeune et la plus âgée. Contrairement aux hommes, il n'y a pas de plateau entre les deux premières tranches , chaque décennie entraîne une baisse visible de performance (+5 à +9 sec/km par tranche). La catégorie "Autre Femme" à 6:51/km est ici la plus lente, ce qui contraste avec les hommes et suggère des profils de coureuses moins expérimentées dans cette catégorie non standard.

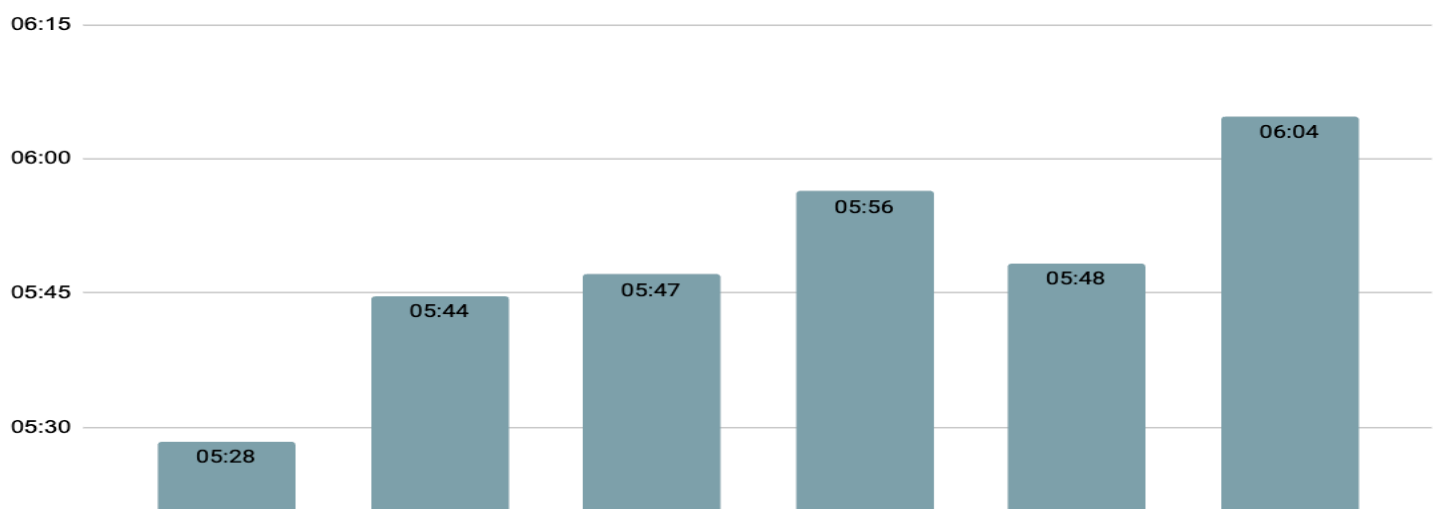
Ce que ces données révèlent pour les acteurs du secteur
La performance décroît avec l'âge, mais de façon différente selon le genre , et ces nuances ont des implications concrètes :

<u>Opportunité</u>	<u>Données</u>	<u>Action possible</u>
Produit "plateau 35-44H"	18-34 et 35-44 ans courent à la même allure	Cibler les 35-44 ans avec des produits performance, pas seulement des produits "confort"
Progression féminine régulière	Chaque tranche féminine perd ~7 sec/km	Des plans d'entraînement progressifs par tranche d'âge représentent une opportunité de service
Seniors masculins (55-65+)	Allure +45 sec/km vs jeunes	Produits orientés endurance, récupération et maintien de la performance dans la durée
Réduction de l'écart H/F avec l'âge	Écart passe de 36 à 22 sec/km	Contenu & produits mixtes pertinents pour les 55+

5.3 Analyse continentale

Graphique : l'allure moyenne par continent

Allure moyenne par continent



Le classement des continents par allure moyenne révèle une hiérarchie nette et sans surprise : l'Afrique domine massivement avec une allure de 5:28 min/km, soit 16 secondes de moins par kilomètre que l'Europe (5:44/km), le continent le plus représenté en volume. À l'autre extrémité, l'Asie affiche l'allure la plus lente avec 6:04 min/km.

Ce que ces données montrent :

L'Afrique court en moyenne 25 minutes plus vite que l'Asie et 11 minutes plus vite que l'Europe sur la distance complète du marathon. Cet écart reflète la domination structurelle des coureurs est-africains (Kenya, Éthiopie) qui, même peu nombreux en volume, tirent l'allure moyenne du continent vers le bas de manière significative.

L'Europe, malgré sa place de premier continent en volume (71,9% des finishers), se classe seconde en performance, ce qui confirme que participation massive et niveau de performance sont deux réalités distinctes. L'Amérique et l'Asie, bien que représentant des marchés du running en forte croissance, affichent les allures les plus lentes, ce qui peut s'expliquer par une pratique du marathon plus récente et une culture de la compétition longue distance moins ancrée.

Ce que ces données révèlent pour les acteurs du secteur :

<u>Continent</u>	<u>Profil</u>	<u>Opportunité</u>
Afrique	Petit volume, excellence performative	Sponsoring élite, visibilité course
Europe	Grand volume, performance intermédiaire	Marché principal , produits polyvalents
Amérique	Volume moyen, allure lente	Marché en croissance , entrée de gamme & coaching
Asie	Petit volume, allure la plus lente	Marché émergent , fort potentiel de progression

En d'autres termes, un acteur du running qui veut maximiser son impact doit distinguer sa stratégie "performance" (Afrique → sponsoring élite) de sa stratégie "volume" (Europe → masse market). Ce sont deux marchés différents avec des besoins produits différents.

6. Conclusion

6.1. Réponse à la problématique

L'analyse des 50 001 finishers du Marathon de Paris 2025 permet de répondre à [nos questions initiales](#) et de valider/corriger nos hypothèses de départ.

En croisant, démographie, performance et géographie, cette étude permet d'avoir un portrait complet et nuancé des coureurs du marathon. Elle révèle des résultats parfois attendus ou non. Lesquels enrichissent notre compréhension du running comme phénomène social autant que sportif.

Validation des hypothèses

✓ H1 validée : les hommes représentent 72,1% des finishers

Avec 72,1% d'hommes contre 27,9% de femmes, soit près de 3 coureurs sur 4, notre première hypothèse est confirmée. Cette répartition traduit une réalité structurelle du marathon de masse : la discipline reste historiquement masculine, même si la progression de la participation féminine ces dernières années est réelle et documentée. Ce déséquilibre, loin d'être une simple statistique, dessine une opportunité de croissance claire pour les acteurs du secteur souhaitant élargir leur audience.

✓ H2 validée : la tranche 18-34 ans qui domine

Notre deuxième hypothèse s'est révélée correcte. La tranche 18-34 ans qui est la plus représentée : elle regroupe 45,1% des hommes et 53,2% des femmes. La tranche 35-44 ans arrive en deuxième position (26,2% H / 24,1% F). Cette différence s'explique en partie par la largeur de la catégorie FFA SEH (Senior Homme, 18 à 39 ans), qui est la plus étendue du système de classification.

✓ H3 validée : l'Afrique domine massivement le classement de performance

Avec une allure moyenne de 5:28 min/km, les coureurs africains, principalement kenyans et éthiopiens, dominent sans partage le haut du classement, soit 16 secondes par kilomètre de moins que les coureurs européens et 36 secondes de moins que les coureurs asiatiques. Cette supériorité représente un écart de plus de 25 minutes sur la distance complète du marathon, confirmant la domination structurelle des nations est-africaines en athlétisme de fond.

✓ H4 validée avec nuance, La performance décline avec l'âge, mais pas de façon linéaire

Notre quatrième hypothèse est globalement confirmée, mais les données révèlent une nuance importante. Chez les hommes, les tranches 18-34 ans et 35-44 ans affichent des allures quasi identiques, respectivement 5:31 et 5:32 min/km, soit seulement 1 seconde d'écart. Ce plateau de performance entre les deux premières tranches suggère que l'expérience et la régularité d'entraînement des coureurs de 35-44 ans compensent parfaitement leur désavantage physiologique naturel. La dégradation devient ensuite progressive : 5:43 pour les 45-54 ans, 5:58 pour les 55-64 ans, et 6:16 pour les 65+.

Profil-type du finisher du Marathon de Paris 2025

L'ensemble de ces analyses convergent vers un profil dominant :

Homme français de 18 à 34 ans, courant à une allure moyenne de 5:36 min/km (soit un temps total estimé autour de 3h57). La France représente à elle seule 71,9% des finishers, devant le Royaume-Uni (8,6%) et l'Allemagne (2,0%).

Ce profil-type, bien qu'utile pour synthétiser les tendances, ne doit pas masquer la diversité réelle du peloton : 100+ nationalités représentées, des tranches d'âge allant de moins de 20 ans à plus de 80 ans, et des allures s'étalant de 2h05 (élite mondiale) à plus de 6 heures pour les finishers les plus endurants.

Ce que ces données révèlent pour les acteurs du secteur du running :

Acteur	Enseignement clé
Marques d'équipement	Le cœur du marché est le coureur 18-44 ans, pas l'élite africaine qui représente une infime minorité en volume. Aller chercher ce marché serait stratégiquement plus lucratif au regard de la masse.
Applications de coaching	La quasi-égalité de performance entre 18-34 et 35-44 ans pourrait suggérer une demande forte pour le maintien du niveau, pas seulement la progression
Nutrition sportive	Les seniors actifs (55-65 ans) représentent un segment à fort pouvoir d'achat et en croissance
Fédérations sportives	L'écart H/F se réduit avec l'âge (de 36 sec/km chez les jeunes à 22 sec/km chez les seniors), la pratique féminine mature tend vers la parité

6.2. Limites et améliorations possibles

Limites :

- Première édition de ce Marathon avec Schneider Electric (pas d'historique sur le même site) : mono-événement résultats propres à Paris 2025, à valider sur d'autres courses
- Fragilité face aux changements du site : l'API utilisée est une API interne au site, non documentée publiquement. Si les développeurs modifient son architecture, le script devient inutilisable sans préavis.
- Absence de gestion des erreurs : le script ne gère pas les cas d'échec (cookies expirés, timeout, serveur indisponible). Tout incident provoquerait un crash immédiat sans message d'erreur explicite.

Améliorations possibles :

- Comparer les éditions 2023, 2024, 2025 pour analyser les tendances sur 3 ans

- Élargir à d'autres marathons (Berlin, Londres, New York) pour valider la représentativité et comparer des coureurs réguliers
- Intégrer les données météo du jour J pour corréler conditions et performance
- Automatiser le renouvellement des cookies via Selenium pour un scraping plus robuste
- Ajouter une segmentation par allure (min/km) pour affiner les profils

7. Annexes

Code source complet dans GitHub :

https://github.com/venom6-13/scraping_marathonParis