

# Open Data Tourisme en train

*> Comment faciliter et encourager le tourisme en train en France ?*

Réalisé par:  
Elsa Catteau  
Safia Zarri Jabri  
Charlotte Prouzet

# I. Présentation de nos dataset

## 1) Liste-des-gares.csv

- |              |               |           |             |
|--------------|---------------|-----------|-------------|
| • Code_uic   | • Rg_troncon  | • IdGaia  | • C_GEO     |
| • Libelle    | • Pk          | • X_L93   | • Geo_Point |
| • Fret       | • Commune     | • Y_L93   | • Geo_Shape |
| • Voyageurs  | • Département | • X_WGS84 |             |
| • Code_ligne | • IdRéseau    | • Y_WGS84 |             |

> Format csv

> 3884 lignes (gares étudiées)

> 18 colonnes (features)

> Types de données : variables numériques    catégorielles    binaires    identifiants    géométriques

# I. Présentation de nos dataset

## 1) Liste-des-gares.csv

- Code\_uic
- Libelle
- Fret
- Voyageurs
- Code\_ligne
- Rg\_troncon
- Pk
- Commune
- Département
- IdRéseau
- IdGaia
- X\_L93
- Y\_L93
- X\_WGS84
- Y\_WGS84
- C\_GEO
- Geo\_Point
- Geo\_Shape

> Format csv

> 3884 lignes (gares étudiées)

> 18 colonnes (features)

> Types de données : variables numériques    catégorielles    binaires    identifiants    géométriques

# I. Présentation de nos dataset

## 1) Liste-des-gares.csv

- Code\_uic
- Libelle
- Fret
- Voyageurs
- Code\_ligne
- Rg\_troncon
- Pk
- Commune
- Département
- IdRéseau
- IdGaia
- X\_L93
- Y\_L93
- X\_WGS84
- Y\_WGS84
- C\_GEO
- Geo\_Point
- Geo\_Shape

> Format csv

> 3884 lignes (gares étudiées)

> 18 colonnes (features)

> Types de données : variables numériques    catégorielles    binaires    identifiants    géométriques

# I. Présentation de nos dataset

## 1) Liste-des-gares.csv

- Code\_uic
- Libelle
- Fret
- Voyageurs
- Code\_ligne
- Rg\_troncon
- Pk
- Commune
- Département
- IdRéseau
- IdGaia
- X\_L93
- Y\_L93
- X\_WGS84
- Y\_WGS84
- C\_GEO
- Geo\_Point
- Geo\_Shape

> Format csv

> 3884 lignes (gares étudiées)

> 18 colonnes (features)

> Types de données : variables numériques catégorielles binaires identifiants géométriques

# I. Présentation de nos dataset

## 1) Liste-des-gares.csv

- Code\_uic
- Libelle
- Fret
- Voyageurs
- Code\_ligne
- Rg\_troncon
- Pk
- Commune
- Département
- IdRéseau
- IdGaia
- X\_L93
- Y\_L93
- X\_WGS84
- Y\_WGS84
- C\_GEO
- Geo\_Point
- Geo\_Shape

> Format csv

> 3884 lignes (gares étudiées)

> 18 colonnes (features)

> Types de données : variables numériques catégorielles binaires identifiants géométriques

# I. Présentation de nos dataset

## 1) Liste-des-gares.csv

- Code\_uic
- Libelle
- Fret
- Voyageurs
- Code\_ligne
- Rg\_troncon
- Pk
- Commune
- Département
- IdRéseau
- IdGaia
- X\_L93
- Y\_L93
- X\_WGS84
- Y\_WGS84
- C\_GEO
- Geo\_Point
- Geo\_Shape

> Format csv

> 3884 lignes (gares étudiées)

> 18 colonnes (features)

> Types de données : variables numériques catégorielles binaires identifiants géométriques

# I. Présentation de nos dataset

## 1) Liste-des-gares.csv

- |                         |                          |                      |                        |
|-------------------------|--------------------------|----------------------|------------------------|
| • <del>Code_uic</del>   | • <del>Rg_troncon</del>  | • <del>IdGaia</del>  | • <del>C_GEO</del>     |
| • <del>Libelle</del>    | • <del>Pk</del>          | • <del>X_L93</del>   | • <del>Geo_Point</del> |
| • <del>Fret</del>       | • <del>Commune</del>     | • <del>Y_L93</del>   | • <del>Geo_Shape</del> |
| • <del>Voyageurs</del>  | • <del>Département</del> | • <del>X_WGS84</del> |                        |
| • <del>Code_ligne</del> | • <del>IdRéseau</del>    | • <del>Y_WGS84</del> |                        |

> Format csv

> 3884 lignes (gares étudiées)

> 18 colonnes (features)

> Types de données : variables numériques    catégorielles    binaires    identifiants    géométriques

**FILTRAGE :** 1. Sélection des features



# I. Présentation de nos dataset

## 1) Liste-des-gares.csv

- |                         |                         |                     |                        |
|-------------------------|-------------------------|---------------------|------------------------|
| • Code_uic              | • <del>Rg_troncon</del> | • <del>IdGaia</del> | • <del>C_GEO</del>     |
| • Libelle               | • <del>Pk</del>         | • <del>X_L93</del>  | • <del>Geo_Point</del> |
| • <del>Fret</del>       | • Commune               | • <del>Y_L93</del>  | • <del>Geo_Shape</del> |
| • <del>Voyageurs</del>  | • Département           | • X_WGS84           |                        |
| • <del>Code_ligne</del> | • <del>IdRéseau</del>   | • Y_WGS84           |                        |

> Format csv

> 3884 lignes (gares étudiées)

> 18 colonnes (features)

> Types de données : variables numériques   catégorielles   binaires   identifiants   géométriques

**FILTRAGE :** 1. Sélection des features  
2. Suppression des doublons

# I. Présentation de nos dataset

## 1) Liste-des-gares.csv

- |                         |                         |                     |                        |
|-------------------------|-------------------------|---------------------|------------------------|
| • Code_uic              | • <del>Rg_troncon</del> | • <del>IdGaia</del> | • <del>C_GEO</del>     |
| • Libelle               | • <del>Pk</del>         | • <del>X_L93</del>  | • <del>Geo_Point</del> |
| • <del>Fret</del>       | • Commune               | • <del>Y_L93</del>  | • <del>Geo_Shape</del> |
| • <del>Voyageurs</del>  | • Département           | • X_WGS84           |                        |
| • <del>Code_ligne</del> | • <del>IdRéseau</del>   | • Y_WGS84           |                        |

> Format csv

> 3884 lignes (gares étudiées)

> 18 colonnes (features)

> Types de données : variables numériques    catégorielles    binaires    identifiants    géométriques

**FILTRAGE :**

1. Sélection des features
2. Suppression des doublons
3. Filtre sur “Libelle” : (suppression des tirets, accents et changement de “Charles” par “CDG”)

# I. Présentation de nos dataset

## 2) Tgvmax.csv

- Date
- Train\_no
- Entity
- Axe
- Origine\_IATA
- Destination\_IATA
- Origine
- Destination
- Heure\_depart
- Heure\_arrivee
- Disponibilité de places MAX JEUNE et MAX SENIOR

> Format csv

> 377621 lignes (trajets étudiées)

> 11 colonnes (features)

> Types de données : variables temporelles    catégorielles    binaires    identifiants

# I. Présentation de nos dataset

## 2) Tgvmax.csv

- Date
- Train\_no
- Entity
- Axe
- Origine\_IATA
- Destination\_IATA
- Origine
- Destination
- Heure\_depart
- Heure\_arrivee
- Disponibilité de places MAX JEUNE et MAX SENIOR

> Format csv

> 377621 lignes (trajets étudiées)

> 11 colonnes (features)

> Types de données : variables temporelles    catégorielles    binaires    identifiants

# I. Présentation de nos dataset

## 2) Tgvmax.csv

- Date
- Train\_no
- Entity
- Axe
- Origine\_IATA
- Destination\_IATA
- Origine
- Destination
- Heure\_depart
- Heure\_arrivee
- Disponibilité de places MAX JEUNE et MAX SENIOR

> Format csv

> 377621 lignes (trajets étudiées)

> 11 colonnes (features)

> Types de données : variables temporelles   catégories   binaires   identifiants

# I. Présentation de nos dataset

## 2) Tgvmax.csv

- Date
- Train\_no
- Entity
- Axe
- Origine\_IATA
- Destination\_IATA
- Origine
- Destination
- Heure\_depart
- Heure\_arrivee
- Disponibilité de places MAX JEUNE et MAX SENIOR

> Format csv

> 377621 lignes (trajets étudiées)

> 11 colonnes (features)

> Types de données : variables temporelles catégorielles binaires identifiants

# I. Présentation de nos dataset

## 2) Tgvmax.csv

- Date
- Train\_no
- Entity
- Axe
- Origine\_IATA
- Destination\_IATA
- Origine
- Destination
- Heure\_depart
- Heure\_arrivee
- Disponibilité de places MAX JEUNE et MAX SENIOR

> Format csv

> 377621 lignes (trajets étudiées)

> 11 colonnes (features)

> Types de données : variables temporelles catégorielles binaires identifiants

# I. Présentation de nos dataset

## 2) Tgvmax.csv

- Date
- Train\_no
- Entity
- Axe
- Origine\_IATA
- Destination\_IATA
- Origine
- Destination
- Heure\_depart
- Heure\_arrivee
- Disponibilité de places MAX JEUNE et MAX SENIOR

> Format csv

> 377621 lignes (trajets étudiées)

> 11 colonnes (features)

> Types de données : variables temporelles catégorielles binaires identifiants

**FILTRAGE :** 1. Suppression des trajets internationaux



# I. Présentation de nos dataset

## 2) Tgvmax.csv

- ~~Date~~
- ~~Train\_no~~
- ~~Entity~~
- ~~Axe~~
- ~~Origine\_IATA~~
- ~~Destination\_IATA~~
- ~~Origine~~
- ~~Destination~~
- ~~Heure\_depart~~
- ~~Heure\_arrivee~~
- ~~Disponibilité de places MAX JEUNE et MAX SENIOR~~

- > Format csv
- > 377621 lignes (trajets étudiées)
- > 11 colonnes (features)
- > Types de données : variables temporelles    catégorielles    binaires    identifiants

**FILTRAGE :** 1. Suppression des trajets internationaux  
2. Sélection des features

# I. Présentation de nos dataset

## 2) Tgvmax.csv

- ~~Date~~
- ~~Train\_no~~
- ~~Entity~~
- ~~Axe~~
- ~~Origine\_IATA~~
- ~~Destination\_IATA~~
- ~~Origine~~
- ~~Destination~~
- ~~Heure\_depart~~
- ~~Heure\_arrivee~~
- ~~Disponibilité de places MAX JEUNE et MAX SENIOR~~

- > Format csv
- > 377621 lignes (trajets étudiées)
- > 11 colonnes (features)
- > Types de données : variables temporelles    catégorielles    binaires    identifiants

**FILTRAGE :**

1. Suppression des trajets internationaux
2. Sélection des features
3. Filtre sur Origine : (suppression des parenthèses, tirets et accents)
  - > exemple : “Paris (intramuros)” devient “Paris”

# I. Présentation de nos dataset

## 3) emission-co2-perimetre-complet.csv

- Transporteur
- Origine
- Origine\_uic
- Destination
- Destination\_uic
- Distance entre les gares
- Train - Empreinte carbone (kgCO2e)
- Autocar longue distance - Empreinte carbone (kgCO2e)
- Avion - Empreinte carbone (kgCO2e)
- Voiture électrique (2,2 pers.) - Empreinte carbone (kgCO2e)
- Voiture thermique (2,2 pers.) - Empreinte carbone (kgCO2e)

> Format csv

> 119 lignes (trajets étudiées)

> 11 colonnes (features)

> Types de données : variables numériques      catégorielles      identifiants

# I. Présentation de nos dataset

## 3) emission-co2-perimetre-complet.csv

- Transporteur
- Origine
- Origine\_uic
- Destination
- Destination\_uic
- Distance entre les gares
- Train - Empreinte carbone (kgCO2e)
- Autocar longue distance - Empreinte carbone (kgCO2e)
- Avion - Empreinte carbone (kgCO2e)
- Voiture électrique (2,2 pers.) - Empreinte carbone (kgCO2e)
- Voiture thermique (2,2 pers.) - Empreinte carbone (kgCO2e)

> Format csv

> 119 lignes (trajets étudiées)

> 11 colonnes (features)

> Types de données : variables numériques    catégorielles    identifiants

# I. Présentation de nos dataset

## 3) emission-co2-perimetre-complet.csv

- Transporteur
- Origine
- Origine\_uic
- Destination
- Destination\_uic
- Distance entre les gares
- Train - Empreinte carbone (kgCO2e)
- Autocar longue distance - Empreinte carbone (kgCO2e)
- Avion - Empreinte carbone (kgCO2e)
- Voiture électrique (2,2 pers.) - Empreinte carbone (kgCO2e)
- Voiture thermique (2,2 pers.) - Empreinte carbone (kgCO2e)

> Format csv

> 119 lignes (trajets étudiées)

> 11 colonnes (features)

> Types de données : variables numériques    catégorielles    identifiants

# I. Présentation de nos dataset

## 3) emission-co2-perimetre-complet.csv

- Transporteur
- Origine
- Origine\_uic
- Destination
- Destination\_uic
- Distance entre les gares
- Train - Empreinte carbone (kgCO2e)
- Autocar longue distance - Empreinte carbone (kgCO2e)
- Avion - Empreinte carbone (kgCO2e)
- Voiture électrique (2,2 pers.) - Empreinte carbone (kgCO2e)
- Voiture thermique (2,2 pers.) - Empreinte carbone (kgCO2e)

> Format csv

> 119 lignes (trajets étudiées)

> 11 colonnes (features)

> Types de données : variables numériques    catégorielles    identifiants

# I. Présentation de nos dataset

## 3) emission-co2-perimetre-complet.csv

- Transporteur
- Origine
- Origine\_uic
- Destination
- Destination\_uic
- Distance entre les gares
- Train - Empreinte carbone (kgCO2e)
- Autocar longue distance - Empreinte carbone (kgCO2e)
- Avion - Empreinte carbone (kgCO2e)
- Voiture électrique (2,2 pers.) - Empreinte carbone (kgCO2e)
- Voiture thermique (2,2 pers.) - Empreinte carbone (kgCO2e)

> Format csv

> 119 lignes (trajets étudiées)

> 11 colonnes (features)

> Types de données : variables numériques    catégorielles    identifiants

**FILTRAGE :** 1. Suppression des transporteurs différents de “TGV”

# I. Présentation de nos dataset

## 3) emission-co2-perimetre-complet.csv

- ~~Transporteur~~
- Origine
- ~~Origine\_uic~~
- Destination
- ~~Destination\_uic~~
- Distance entre les gares
- Train - Empreinte carbone (kgCO2e)
- Autocar longue distance - Empreinte carbone (kgCO2e)
- Avion - Empreinte carbone (kgCO2e)
- ~~Voiture électrique (2,2 pers.) - Empreinte carbone (kgCO2e)~~
- Voiture thermique (2,2 pers.) - Empreinte carbone (kgCO2e)

> Format csv

> 119 lignes (trajets étudiées)

> 11 colonnes (features)

> Types de données : variables numériques    catégorielles    identifiants

**FILTRAGE :** 1. Suppression des transporteurs différents de “TGV”  
2. Sélection des features



# I. Présentation de nos dataset

## 3) emission-co2-perimetre-complet.csv

- ~~Transporteur~~
- Origine
- ~~Origine\_uic~~
- Destination
- ~~Destination\_uic~~
- Distance entre les gares
- Train - Empreinte carbone (kgCO2e)
- Autocar longue distance - Empreinte carbone (kgCO2e)
- Avion - Empreinte carbone (kgCO2e)
- ~~Voiture électrique (2,2 pers.) - Empreinte carbone (kgCO2e)~~
- Voiture thermique (2,2 pers.) - Empreinte carbone (kgCO2e)

> Format csv

> 119 lignes (trajets étudiées)

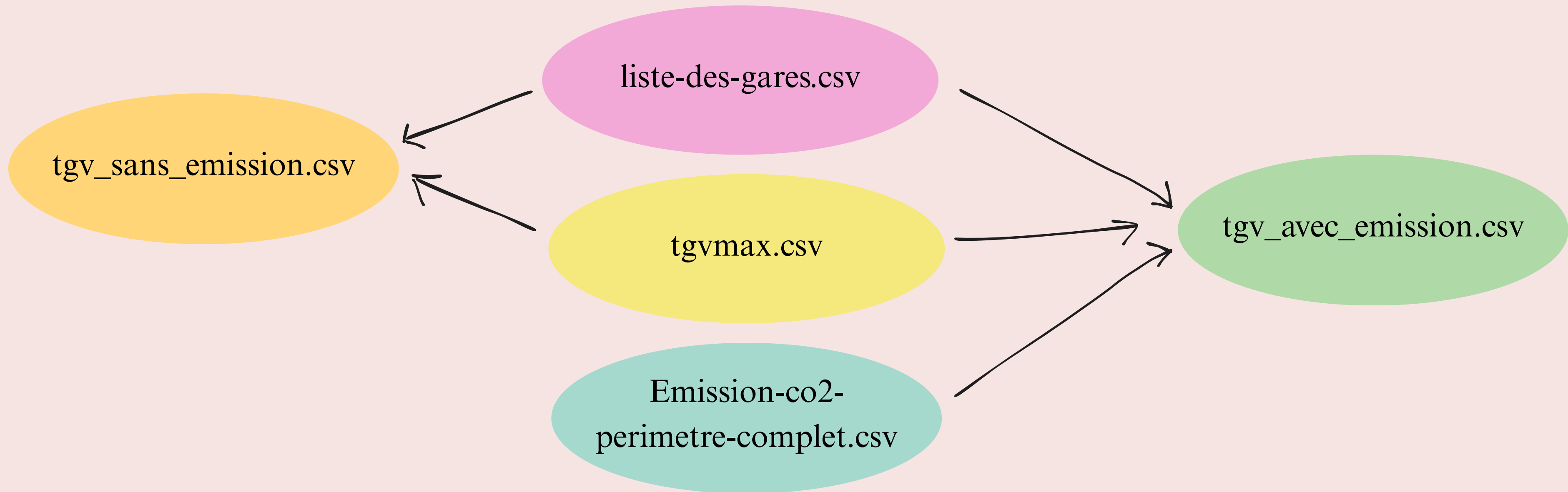
> 11 colonnes (features)

> Types de données : variables numériques    catégorielles    identifiants

**FILTRAGE :**

1. Suppression des transporteurs différents de “TGV”
2. Sélection des features
3. Filtre sur “Origine” : (suppression des tirets, accents et changement de “Charles” par “CDG”)

## II. Fusion de nos dataset filtrés



**FILTRAGE :** Sélection des features et suppression des doublons

>> fusion\_avec\_emission.csv

Origine	longitude_origine	latitude_origine	Destination	longitude_destination	latitude_destination	Heure_depart	Heure_arrivee	Distance entre les gares	Train - Empreinte carbone (kgCO	Autocar longue distance	Avion	Voiture thermique (2,2
PARIS	2.2806439225378	48.8530378277834	REIMS	4.022706051740367	49.23341625486098	18:28	19:14	146.156	0.42385239999999999	5.2025		15.631
PARIS	2.2806439225378	48.8530378277834	CHAMBERY	5.919644744721905	45.57190136351876	20:37	23:33	531.951	1.5426579	19.6995		61.5538
PARIS	2.2806439225378	48.8530378277834	ANNECY	6.121832970715078	45.90221094284461	20:37	00:26	545.0	1.5805	19.32		60.368
PARIS	2.2806439225378	48.8530378277834	NANCY	6.174101796891322	48.68970240439984	16:07	17:44	367.0	1.0643	13.317		41.6108
PARIS	2.2806439225378	48.8530378277834	CHAMPAGNE	3.993655995605394	49.21455924201552	07:58	08:38	175.0	0.5075	5.106		15.9544
PARIS	2.2806439225378	48.8530378277834	LENS	2.792160590965308	50.43781911422396	16:45	18:02	198.0	0.57419999999999999	7.383		23.0692
PARIS	2.2806439225378	48.8530378277834	ANGOULEME	0.16414524722233	45.65364279345685	16:06	18:05	428.0	1.2412	15.525		48.51
PARIS	2.2806439225378	48.8530378277834	BREST	-4.479243011743669	48.3879297952412	10:57	14:50	613.0	1.7777	20.424	72.07	63.8176
PARIS	2.2806439225378	48.8530378277834	POITIERS	0.332786866293225	46.58236426463122	06:52	08:11	320.0	0.928	11.7645		36.7598
PARIS	2.2806439225378	48.8530378277834	RENNES	-1.691801073204263	48.11826355636143	10:57	12:25	363.952	1.0554608	20.2515	71.36	63.2786
PARIS	2.2806439225378	48.8530378277834	STRASBOUR	7.731594953278417	48.58287025491014	07:18	09:02	450.58	1.306682	16.836		52.6064
PARIS	2.2806439225378	48.8530378277834	GRENOBLE	5.714516819858535	45.19083025228243	07:13	10:13	556.09	1.612661	19.872		62.0928
PARIS	2.2806439225378	48.8530378277834	LE MANS	0.194090806361883	47.99501903643902	07:44	08:38	202.0	0.5858	7.2105		22.5302
PARIS	2.2806439225378	48.8530378277834	BELFORT MC	6.898445295065638	47.58602332147397	20:20	22:53	449.0	1.30209999999999998	17.112		53.4688
PARIS	2.2806439225378	48.8530378277834	TOULON	5.929717733657508	43.12858683541123	08:21	12:10	812.0	2.3548	28.9455	181.1	90.4442
MARNE LA VAL	2.7826949499979	48.8700341139662	AVIGNON TG	4.785988680479314	43.92185220386844	08:44	11:52	646.0	1.8734	24.1845		75.5678
PARIS	2.2806439225378	48.8530378277834	NANTES	-1.544230274778459	47.21645350148569	17:59	22:19	385.841	1.11893889999999998	13.317	48.86	41.6108
PARIS	2.2806439225378	48.8530378277834	MACON LOC	4.778443228894927	46.282781818898	20:59	22:40	369.0	1.07009999999999998	13.9035		43.4434
MARNE LA VAL	2.7826949499979	48.8700341139662	AIX EN PROV	5.317306755615249	43.45523657055668	16:49	19:58	720.0	2.088	26.565		83.006
PARIS	2.2806439225378	48.8530378277834	VANNES	-2.752403881216251	47.66520979265519	16:57	19:25	489.0	1.4181	16.008		50.0192
PARIS	2.2806439225378	48.8530378277834	TOULOUSE M	1.454396139877268	43.61069941718024	18:11	22:47	793.993	2.30257970000000003	23.4255	84.40	73.19619999999999
CDG	2.5700355070970	49.0085079715308	RENNES	-1.691801073204263	48.11826355636143	12:13	14:47	421.0	1.2209	13.179	47.14	41.1796
PARIS	2.2806439225378	48.8530378277834	BIARRITZ	-1.546701196505992	43.45903906706433	16:11	20:16	744.496	2.1590384	27.0135	95.86	84.4074
PARIS	2.2806439225378	48.8530378277834	BAYONNE	-1.47066304089878	43.49694782534664	16:11	20:03	734.758	2.1307982	26.565		83.006
PARIS	2.2806439225378	48.8530378277834	DUNKERQUE	2.369189693600089	51.02926760373617	18:51	21:04	305.0	0.8845	10.3845		32.4478
PARIS	2.2806439225378	48.8530378277834	ARRAS	2.781877763520992	50.28670060943111	12:25	13:13	179.0	0.5191	6.555		20.482
PARIS	2.2806439225378	48.8530378277834	ARCACHON	-1.165311959015783	44.65884069382592	18:47	21:37	595.0	1.7255	19.5		70.07
PARIS	2.2806439225378	48.8530378277834	DOUAI	3.090430914770062	50.37158212907979	17:45	18:55	204.0	0.5916	6.831		21.3444
PARIS	2.2806439225378	48.8530378277834	VALENCIENN	3.518515366298679	50.3659654003862	10:46	12:35	237.0	0.68729999999999999	7.314		22.8536
STRASBOURG	7.7315949532784	48.5828702549101	NANTES	-1.544230274778459	47.21645350148569	17:01	22:23	834.0	2.4186	29.8425	184.8	93.247
PARIS	2.2806439225378	48.8530378277834	CALAIS VILLI	1.850475991847154	50.95341588718877	20:18	22:24	335.0	0.9715	10.2465		32.0166



## II. Création d'une carte interactive

Utilisation de la librairie folium de python

Création de 2 cartes :

- Avec prise en compte de l'empreinte carbone :

Point de départ : **Paris**

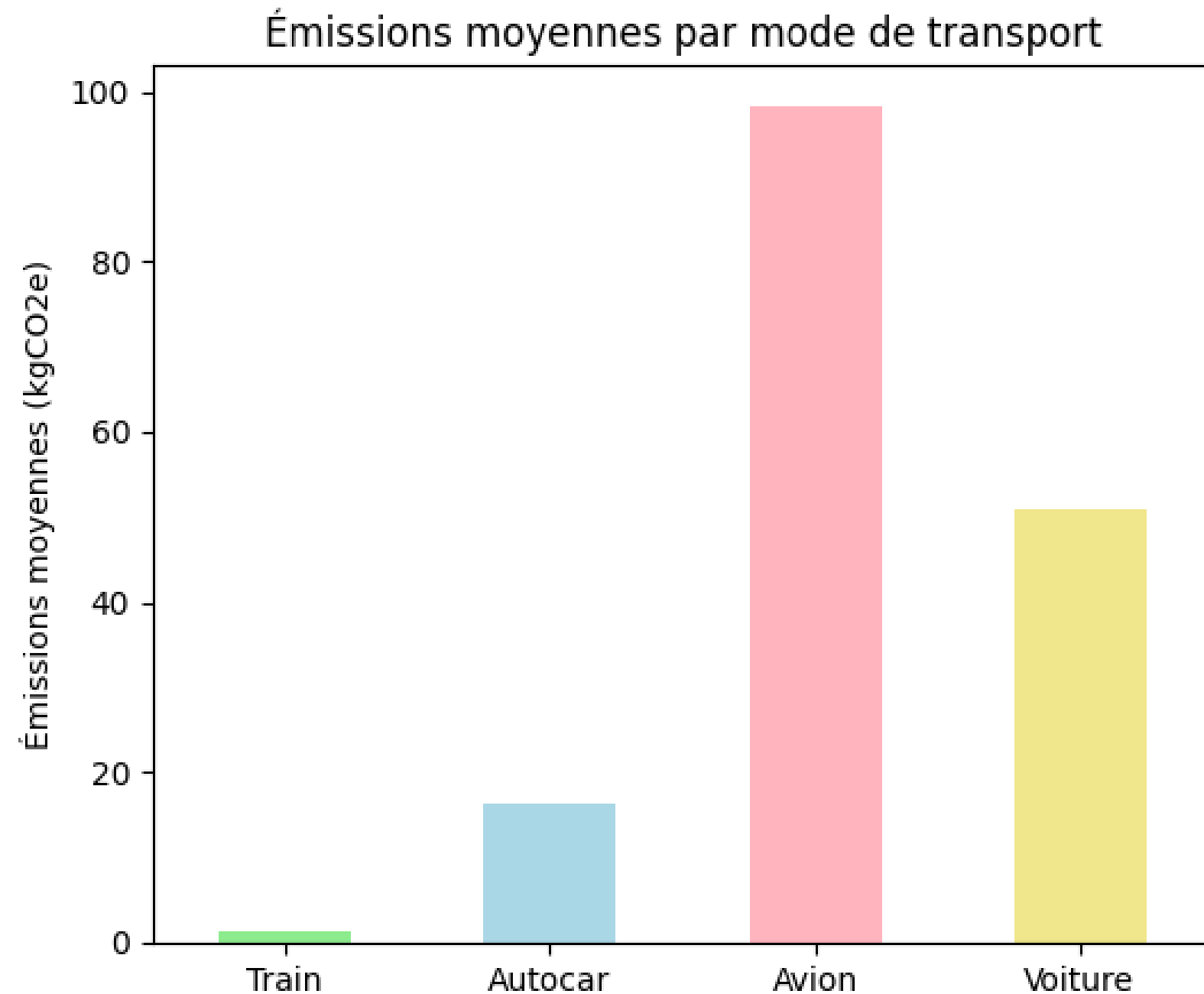
*file:///C:/Users/elsac/Downloads/carte\_emission.html*

- Sans prise en compte de l'empreinte carbone :

Point de départ multiple, ici on a choisi **Nice** :

*file:///C:/Users/elsac/Downloads/carte\_sans\_emission.html*

# Comparaison des émissions par rapport au mode de transport

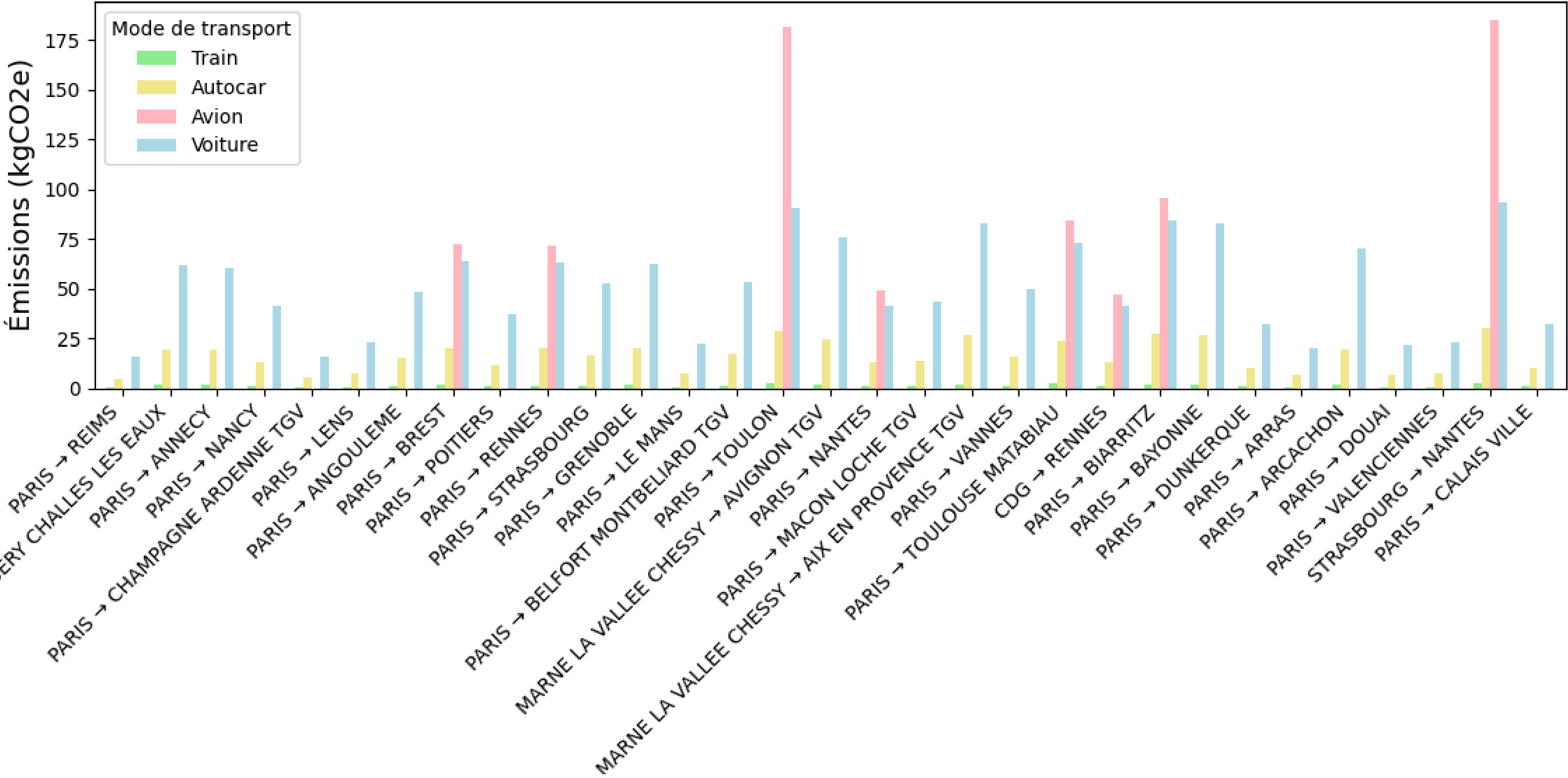


En moyenne, le **Train** est **38.4 fois** moins polluant que la **Voiture**, il émet **97.4%** de CO2 de moins

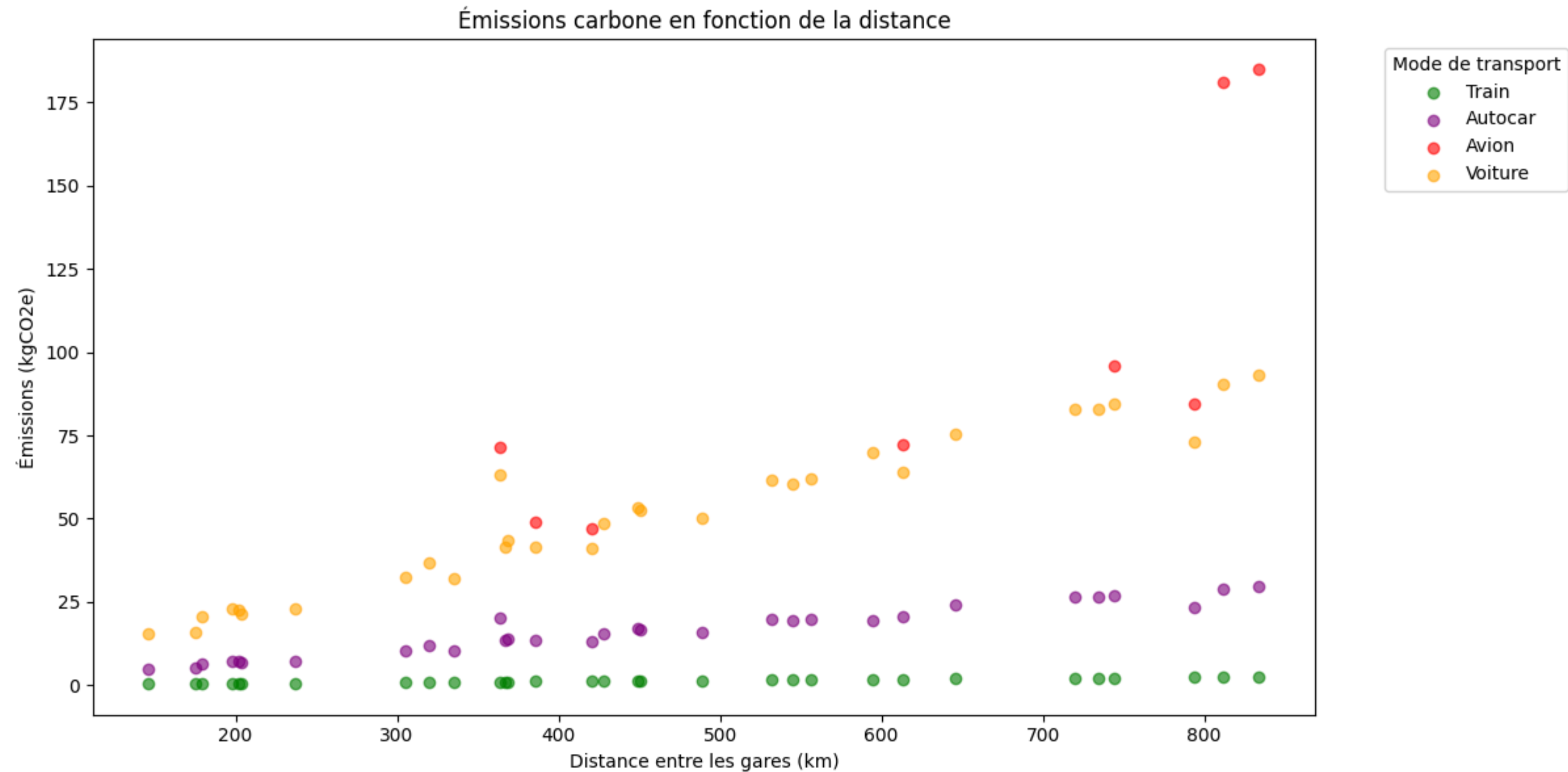
En moyenne, le **Train** est **53.1 fois** moins polluant que l'**Avion**, il émet **97.9%** de CO2 de moins

En moyenne, le **Train** est **12.2 fois** moins polluant que l'**Autocar**, il émet **91.7%** de CO2 de moins

# Comparaison des émissions carbone par mode de transport




# Comparaison des émissions carbone en fonction de la distance



# Conclusion

- Sensibiliser les voyageurs à l'impact écologique de leurs choix de transport.
- Encourager des déplacements touristiques plus respectueux du climat, surtout pour les trajets nationaux.





**Merci pour votre attention !**