



Data Analytics

Carpooling in France
which uses? for which territories?

Clémence Legrand

March, 2023

Plan

- A. Introduction
- B. Data collection and data sources
- C. Data cleaning and Exploratory data analysis
- D. Data base type selection
- E. Entities. ERD
- F. MySQL queries
- G. Some more insights
- H. Conclusion
- I. Links

A. Introduction

In France 73% of people travel to work by car, according to the latest INSEE home-work travel survey of 2017. At the creation of Blablacar (formerly called covoiturage.fr) in 2008, carpooling was intended for tourist use, more economical solution than the train. Now, it is the home-work carpooling that is encouraged. Numerous operators, such as Klaxit, have developed on this niche and work with companies and local authorities to encourage this ecological and economical practice.

Carpooling can be defined as the common use of a vehicle by a non-professional driver with at least one passenger to make all or part of a trip initially planned by the driver. The travel costs are shared between the passengers.

These last years, the public authorities (local authorities and government) have put the subject on the political agenda and propose financial aids to develop the practice. In what interest? Reduce the harmful effects of individual car use, such as traffic jams, parking problems, and air pollution, by encouraging a practice that requires little investment in infrastructures. For the user, it is an economical practice that can offer more possibilities to move for some trip where public transports are not available or affordable.

The French government has set an ambitious goal: to triple the number of daily carpooling trips by 2027 to 3 million. In 2021, the Ministry of Transport and the Ministry of Transformation and Public Service wished to co-finance, within the framework of the France Relance plan, a new extension of the carpooling plan, with an ambitious plan to exploit the data of carpooling.

The coronavirus epidemic, the strikes for the pensions reform in 2019, or the fluctuations of gasoline prices in 2022 are so many elements which could impact the practice of carpooling. That is why I wanted to study the global evolution of carpooling, to characterize the practice over these years, and to understand in which territories and with which incentives this practice is more likely to develop.

B. Data collection and data sources

a. My main data set: carpooling journeys since 2019

The main data source I use is the Carpooling Proof Registry which is a digital service provided by the General Directorate for Infrastructure, Transport and the Sea (DGITM). It is a service set up by the State which collects data from carpooling operators, anonymizes the operators, and gives this information back to the local authorities to help them set up carpooling in their territory. Part of the data in this register is in open data.

Indeed, without this anonymization by a trusted third party, the operators do not give their data on the journeys made in order not to disclose their market share. Also, I tried to use the Blablacar and Klaxit API's but they don't give more information than a simple search to book a trip.

Field	Description
journey_id	Unique identifier for one row
trip_id	Unique identifier for a trip (can be duplicates, because one row correspond to a couple driver-passenger)
journey_start_datetime	Date and time of departure in ISO 8601 format (YYYY-MM-DDThh:mm:ssZ). The time is expressed in UTC (Coordinated Universal Time). UTC is not adjusted to summer and winter time
journey_start_date	Departure date in YYYY-MM-DD format
journey_start_time	Departure time in hh:mm:ss format
journey_start_lat	Latitude of the departure point
journey_start_lon	Longitude of the departure point
journey_start_insee	INSEE code of the departure city
journey_start_department	Department of the starting point
journey_start_towngroup	EPCI or AOM of departure. Determined via the geographic coordinates.
journey_end_datetime	Date and time of arrival in ISO 8601 format (YYYY-MM-DDThh:mm:ssZ). The time is expressed in UTC (Coordinated Universal Time).
journey_end_date	Arrival date in YYYY-MM-DD format
journey_end_time	Arrival time in hh:mm:ss format
journey_end_lat	Latitude of the arrival point
journey_end_lon	Longitude of the arrival point
journey_end_insee	INSEE code for the municipality of the arrival point

journey_end_department	Department of the arrival point
journey_end_towngroup	EPCI or AOM of arrival. Determined via the geographical coordinates.
number_of_passengers	Number of passenger for each trip.
journey_distance	Distance traveled in meters. Information sent by the operator
journey_duration	Duration of the journey in minutes. Information sent by the operator
has_incentive	If the trip is incited or not by an operator or a campaign set on the register.

b. Dimensional datasets

1. Cities information

Since my main dataset aggregates geographically localized data, I have aggregated several other territorial information from several INSEE datasets :

- The list of cities with their names and codes, their department and region pertenency
- The last population census at the municipal level from 2019
- The density of population by cities, with a 7 level classification established by INSEE, also based on the last population census at the municipal level from 2019
- And finally the commuting database of INSEE which counts for each municipality the number of people working in a municipality other than their place of residence. The data is also based on 2019. This allowed me to compute the percentage of people commuting in each city.

Field	Description
id_city	INSEE code for the municipality
id_reg	Code for region
id_dep	Code for department
city_name	Name of city
dens_pop	Population divided by the city surface
density_degree	Level of density according to INSEE
density_type	Name of the level of density
total_pop_2019	Municipal population on last census
no_commute_pop	Number of people working in the same city where they live
work_commute_perc	$(total_pop_2019 - no_commute_pop) / total_pop_2019$

Field	Description
id_reg	Code for region
region_name	Name of region

2. Carpool areas dataset

Carpooling does not require any infrastructure but reserving parking areas should logically encourage this practice. That's why I wanted to compare the territories of high carpooling with the existing carpooling parking lots in France. I downloaded the national database of carpooling places which is a state database managed by Etalab and regularly updated.

Field	Description
id	Unique id for the park area
name	Name
city_name	Name of city where the parking is located
id_city	INSEE code for the city
type	Type of parking
date	Date of creation
long	Longitude
lat	Latitude

3. Schedule information

In order to analyze the seasonality, the impact of vacation and weekdays, I aggregated a table the scholar vacation schedule of each year from 2019. I found these calendars in OpenDataSoft. I also create a new column that indicate for each datetime the day of the week.

Field	Description
date	Date in YYYY-MM-DD format
holidays_name	Name of the holidays
weekday	Day of the week
vacation_level	Number of geographical areas in vacation (0,1,2 or 3)

4. Gasoline price information

One of the main motivations for carpooling is to save expenses since the costs are shared. France having known particularly important fluctuations of the price of gasoline in 2022, I sought a database of the prices of fuels with the finest temporal granularity possible to see if there is a correlation with the decision to carpool. The Ministry of Energy Transition has made available a dataset on petroleum products that gives the price for each type of fuel for each week since January 2020.

Field	Description
date	Date in YYYY-MM-DD format for the first day of the week
gazole	Week price of gazole
Super SP95	Week price of SP95
Super SP95-E10	Week price of SP95-E10
Super SP98	Week price of SP98
mean_price	Average price of all fuels

All the files above were collecting in open sources in csv or excel files. Then the data was parsed and concatenated as shown in the next section, using Python.

C. Data cleaning and Exploratory data analysis

Cleaning the main dataset

My original main dataset counted 10,182,966 rows and 30 columns. I drop some columns:

- Unnamed column
- Operator_class which is not needed in the analysis
- Postalcode for cities is less precise than the INSEE code already in the dataset.
- The name of town column is not useful here because it will appear in another table.

I use the column trip_id to count the number of passengers per trip. I add this column and I delete the column passenger_seats that gives only information about the number of seats reserved at once, which is not a very interesting value.

```
#We want now to transform the number of frequencies in number of passengers

trip_id_counts = global_df['trip_id'].value_counts()

# create a new column with the frequencies of trip_id and the number of passenger_seats (we should retire 1 but we w
global_df['nb_passengers'] = global_df['trip_id'].map(trip_id_counts)+global_df["passenger_seats"]#-1
global_df.head()
```

I decided also to drop the outliers in the trip_id column, I couldn't use the "IQR*1,5" method because of the distribution, so I decided to drop all the trip_id with more than 8 people. The maximum was 58 people for a trip, so I consider it as a mistake.

I decided to ignore the trip between France and other countries. I know it could be interesting to analyse the cross-border commuting but I want my analysis to focus on how to encourage the carpooling in metropolitan France, so I decided to drop all the extern trips, and that allowed me to drop the two columns "country".

For the missing values, I fill the missing values in town_group column with "NA" in order to keep the rows, but to exclude, if necessary, these unknown data from the analysis. On another hand, I drop the rows with missing values in department because there was a small number.

For the missing values in duration, I calculate it from the time columns when it was possible. Otherwise, I drop the rows with missing values in duration.

```
#calculate between two times columns
from pandas import datetime
test['journey_start_time'] = pd.to_datetime(test['journey_start_time'])
test['journey_end_time'] = pd.to_datetime(test['journey_end_time'])

test['journey_duration'] = test['journey_end_time'] - test['journey_start_time']

# convert in minutes
test['journey_duration'] = test['journey_duration'].dt.total_seconds() / 60
```

I drop also one negative value in duration and the outliers regarding to the distribution with a boxplot: I drop the duration above 11 hours.

As journey_distance and journey_duration is correlated, cleaning duration also cleaned the distance column. I just drop the incorrect values, i.e. the zeros, that remained.

Finally, I created a column with only the year of each row, in order to filter easily the dataset.

After this cleaning, the dataset shape was 9,607,669 rows and 23 columns.

Cleaning other datasets

For the other datasets, I mostly check the types of the data. I drop the useless columns for my analysis. I change the column in English.

Export to MySQL

For reasons that I will explain in the next section, after cleaning the data I created the database in MySQL Workbench. To export all the dataset to MySQL workbench, I used SQLAlchemy library. This library allows to connect our Python file with the MySQL, create a database and push tables.

```
Entrée [112]: ! pip install pymysql
              from sqlalchemy import create_engine
              import pymysql.cursors
              import os
              import getpass

Requirement already satisfied: pymysql in /Users/clemence/opt/anaconda3/lib/python3.9/site-packages (1.0.2)
```

```
Entrée [122]: pw = getpass.getpass()
              .....
              
```

```
Entrée [123]: connection_string = 'mysql+pymysql://root:' + pw + '@localhost:3306/'
              engine = create_engine(connection_string)
```

```
Entrée [124]: with engine.connect() as conn:
              conn.execute(f"CREATE DATABASE IF NOT EXISTS Final_Project")
```

```
Entrée [127]: density_2019.to_sql('density', engine, 'Final_Project', if_exists='replace', index = True)
              Out[127]: 34955
```

For the main dataset, it needed to be split on 5 datasets: one per year, and two for the 2022 year. Otherwise, the dataset was too heavy. I exported maximum 2,5 millions rows at a time and I recreated the main dataset after on SQL with INSERT INTO.

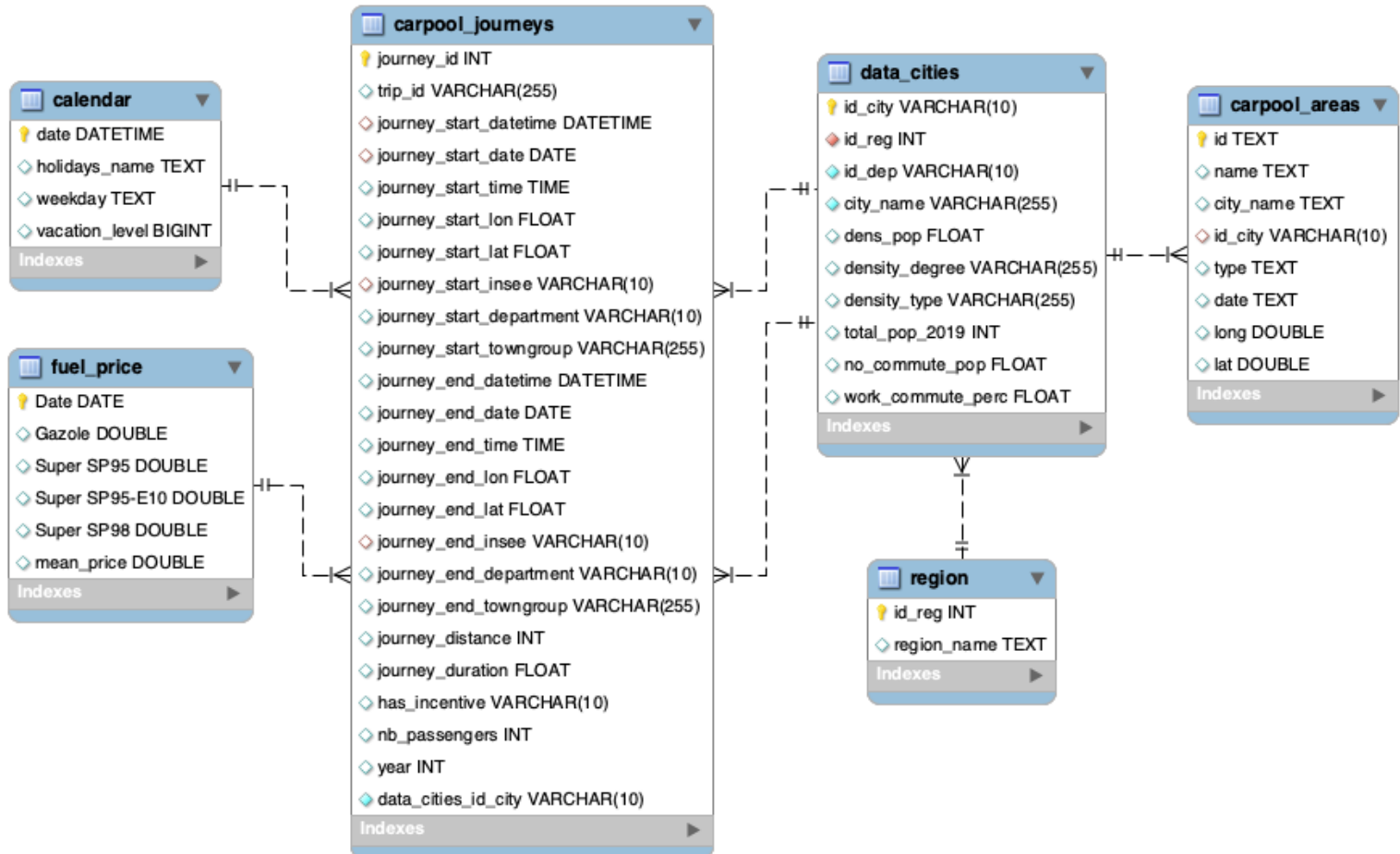
D. Data base type selection

Here are the differences between SQL and NoSQL databases and the reasons why I chose SQL :

1. SQL databases are relational, NoSQL databases are non-relational : that means we can link the data by foreign keys in a SQL database.
2. SQL databases use structured query language and have a predefined schema NoSQL databases have dynamic schemas for unstructured data. SQL tables impose a strict data model, so it is difficult to make mistakes. NoSQL is more flexible but the ability to store data anywhere can lead to consistency issues.
3. SQL databases are vertically scalable, while NoSQL databases are horizontally scalable: horizontal scalability consists in adding additional machines in parallel. This simplifies the architecture while increasing storage capacity tenfold.
4. SQL databases are table-based, while NoSQL databases are document, key-value, graph, or wide-column stores.
5. SQL databases are better for multi-row transactions, while NoSQL is better for unstructured data like documents or JSON.

For these reasons, it is obvious that SQL is the right database for this final project. The database is structured in tables, with predefined schemas and formatted data. SQL allows to quickly make queries and join all the six tables thanks to the links in this relational database. Besides, the multi-rows transaction's ability of SQL databases was needed.

E. Entities. ERD



F. MySQL queries

Here first, it is an example of SQL queries to create the aggregated data_cities table:

```
INSERT INTO data_cities (id_city, id_reg, id_dep, city_name, dens_pop, density_degree, density_type, total_pop_2019, no_commute_pop)
SELECT
    ct.id_city,
    ct.id_reg,
    ct.id_dep,
    ct.city_name,
    d.dens_pop,
    dl.density_degree,
    dl.density_type,
    dl.total_pop_2019,
    wc.no_commute_pop
FROM code_territories ct
JOIN density d ON ct.id_city = d.id_city
JOIN density_labels dl ON ct.id_city = dl.id_city
LEFT JOIN work_commute wc ON ct.id_city = wc.id_city;

## I want to create an indicator in data_cities to have the proportion of people movinf to another city to work
ALTER TABLE data_cities
ADD COLUMN work_commute_perc FLOAT;

UPDATE data_cities
SET work_commute_perc=(total_pop_2019-no_commute_pop)/total_pop_2019*100;
```

The relational database on SQL allowed to show interesting insights by joining my main table with the dimensional tables.

We can first look at the evolution of number of carpool journeys, and among these, at the evolution of incentivized journeys.

```
#Carpooling journeys per year and per incentives
SELECT year, count(*) AS total_journeys, count(CASE WHEN has_incentive="OUT" THEN 1 END) AS with_incentives
FROM carpool_journeys
GROUP BY year
ORDER BY year;
```

year	total_journeys	with_incentives
2019	751622	570696
2020	1712378	1083626
2021	1477834	1398116
2022	4911173	4681458
2023	754662	733718

We can examine the regions' ranking based on the overall number of carpooling journeys.

```
#List of the region where there is the most carpool journeys per capita
SELECT r.region_name,
count(cj.journey_id)/sum(total_pop_2019)*1000 as total_journeys_per_1000inhab
FROM carpool_journeys cj
JOIN data_cities dc ON dc.id_city=cj.journey_start_insee
JOIN region r ON dc.id_reg=r.id_reg
GROUP BY r.region_name
ORDER BY total_journeys_per_1000inhab DESC;
```

region_name	total_journeys_per_1000inhab
Corse	0.1438
Guadeloupe	0.0521
Bourgogne-Franche-Comté	0.0284
Normandie	0.0267
Nouvelle-Aquitaine	0.0251
Bretagne	0.0247
Centre-Val de Loire	0.0243
Hauts-de-France	0.0221
Grand Est	0.0197
Auvergne-Rhône-Alpes	0.0179
Pays de la Loire	0.0145
La Réunion	0.0115
Provence-Alpes-Côte d'Azur	0.0079
Île-de-France	0.0071
Occitanie	0.0061

Then, the query can be specified to only count financially incentivized trips.

```
#The list of the regions where there is the most financially incentivized journey per capita
SELECT r.region_name,
count(cj.journey_id)/sum(total_pop_2019)*1000 as total_incentivized_journeys_1000inhab
FROM carpool_journeys cj
JOIN data_cities dc ON dc.id_city=cj.journey_start_insee
JOIN region r ON dc.id_reg=r.id_reg
WHERE cj.has_incentive="OUI"
GROUP BY r.region_name
ORDER BY total_incentivized_journeys_1000inhab DESC;
```

region_name	total_incentivized_journeys_1000inhab
Corse	0.0824
Guadeloupe	0.0357
Bourgogne-Franche-Comté	0.0308
Nouvelle-Aquitaine	0.0302
Bretagne	0.0278
Normandie	0.0268
Hauts-de-France	0.0258
Centre-Val de Loire	0.0245
Grand Est	0.0210
Auvergne-Rhône-Alpes	0.0203
Pays de la Loire	0.0147
La Réunion	0.0116
Provence-Alpes-Côte d'Azur	0.0093
Occitanie	0.0074
Île-de-France	0.0071

To be more specific about the location, we can observe the town clusters with a higher number of carpool trips in absolute terms.

journey_start_towngroup	total_journeys
Ile-De-France Mobilités	3303739
Métropole du Grand Paris	599461
Metropole Rouen Normandie	489663
Montpellier Méditerranée Métropole	239269
Nantes Métropole	172359
Syndicat Mixte Des Transports En Commun De...	125243
CA Communauté Paris-Saclay	124964
Metz Métropole	107374
CA Grand Paris Sud Seine Essonne Sénart	104695
CA du Beauvaisis	96691
CU Angers Loire Métropole	93143
Metropole Européenne De Lille	80687
Metropole Aix-Marseille-Provence	79628
CA Cœur d'Essonne Agglomération	78786
Angers Loire Métropole	74970

These absolute figures can be significant since they reveal that Métropole Aix-Marseille-Provence, the second largest metropolis in France, lags behind many other more carpooling-savvy metropolises.

Here, is a query to classify the different types of territory densities.

```
#Look at the number of carpool journeys per capita depending of the type of density territories
SELECT dc.density_type, dc.density_degree,
count(cj.journey_id)/sum(total_pop_2019)*1000 as total_journeys_1000inhab
FROM carpool_journeys cj
JOIN data_cities dc ON dc.id_city=cj.journey_start_insee
GROUP BY dc.density_type, dc.density_degree
ORDER BY total_journeys_1000inhab DESC;
```

density_type	density_degree	total_journeys_1000inhab
Rural à habitat très dispersé	7	1.5413
Rural à habitat dispersé	6	0.5978
Bourgs ruraux	5	0.2578
Ceintures urbaines	4	0.1647
Petites villes	3	0.0959
Centres urbains intermédiaires	2	0.0346
Grands centres urbains	1	0.0063

It is interesting to note that the number of trips per 1,000 inhabitants follow the degree of density: the more dispersed the housing, the more carpooling trips per person. This can be explained by the lack of public transports in rural areas.

The following two queries compare the departments from which the most carpooling trips start per capita, and the departments where there are the most trips to work per capita (by measuring the share of the population who work outside their city of residence).

```

#Top 15 of the departments with most carpool journeys per capita
SELECT pd.id_dep as department,
       pd.total_pop_dep,
       count(cj.journey_id),
       count(cj.journey_id)/pd.total_pop_dep*1000 as journeys_per_1000inhab
FROM population_department pd
INNER JOIN carpool_journeys cj ON cj.journey_start_department=pd.id_dep
-- WHERE cj.journey_id<936378
GROUP BY department, total_pop_dep
ORDER BY journeys_per_1000inhab DESC
LIMIT 15;

```

#Top 15 of the departments with high work-commute population

```

SELECT id_dep as department,
       sum(no_commute_pop) as no_commute_pop,
       sum(total_pop_2019) as total_pop_dep,
       (sum(total_pop_2019)-sum(no_commute_pop))/sum(total_pop_2019)*1000 as commute_people_per_1000_inhab
FROM data_cities
GROUP BY department
ORDER BY commute_people_per_1000_inhab DESC
LIMIT 15;

```

departme...	total_pop_d...	count(cj.journey_...	journeys_per_1000inh...
91	1301659	974008	748.2820
78	1448207	707442	488.4951
76	1255633	560902	446.7086
77	1421197	611286	430.1205
94	1407124	571798	406.3594
92	1624357	609775	375.3947
95	1249674	441881	353.5970
49	818273	278038	339.7864
53	307062	93563	304.7039
34	1175623	326885	278.0526
93	1644903	413701	251.5048
60	829419	195136	235.2683
44	1429272	306737	214.6107
85	685442	129608	189.0867
75	2165423	401764	185.5360

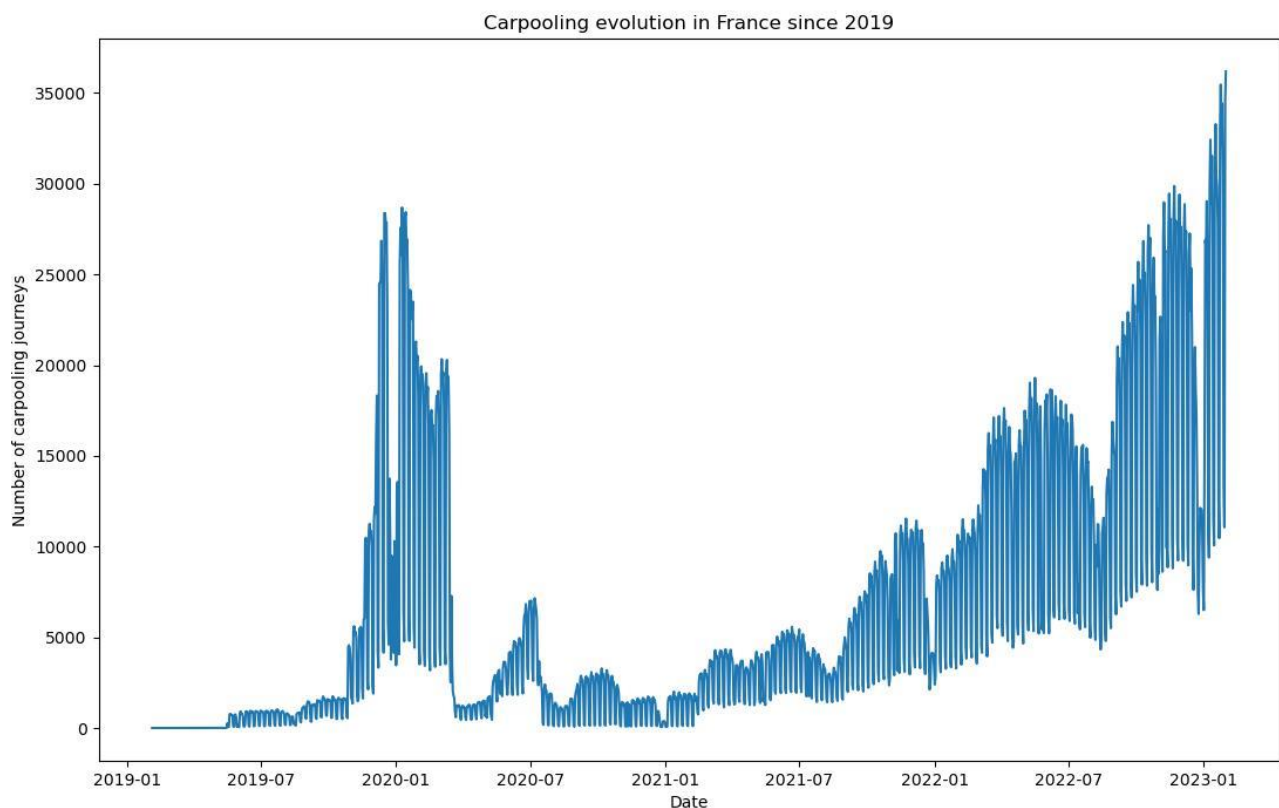
department	no_commute_pop	total_pop_dep	commute_people_per_1000_habitant
69	127856.70292425156	1875747	931.8369146136171
95	97319.38430213928	1249674	922.1241825450963
91	108999.2395772934	1301659	916.261294565402
77	119161.27219343185	1421197	916.154289522542
93	140418.6276550293	1644903	914.6340983905864
78	127067.91999721527	1448207	912.258454767022
94	123477.57485961914	1407124	912.2482632236965
13	181778.7587928772	2043110	911.0284033689438
62	131353.9445066452	1465278	910.3556154486417
60	76781.72447490692	829419	907.4270971910374
54	70550.85018777847	733760	903.8502368788453
57	102233.73611211777	1046543	902.3129139346231
27	59754.1141834259	599507	900.3279124623634
59	263542.6309623718	2608346	898.9617823086462
92	166278.53475952148	1624357	897.6342424974796

Comparing these two variables is useful as it helps identify areas with a high number of people working outside their city but not engaging in carpooling. For instance, the department 69 "Rhône" has the highest percentage of commuting individuals, but its carpooling rate per inhabitant is not the best.

G. SOME MORE INSIGHTS

Evolution

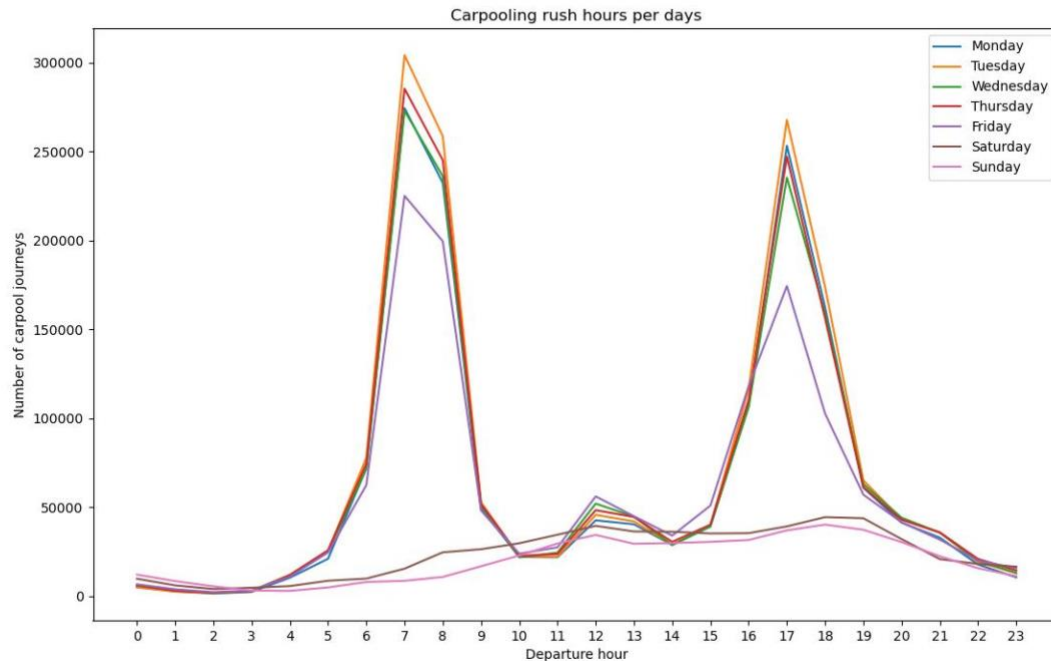
The first insight by analyzing the main dataset is to see the evolution of carpooling since 2019. We can clearly see the net stop given to the growth of carpooling during the first containment in France and also during the second. It took time for carpooling to return to the same level, which can be explained by the health recommendations that did not encourage carpooling.



We can also see that there are less trips during August and December, which correspond to the most popular holidays periods.

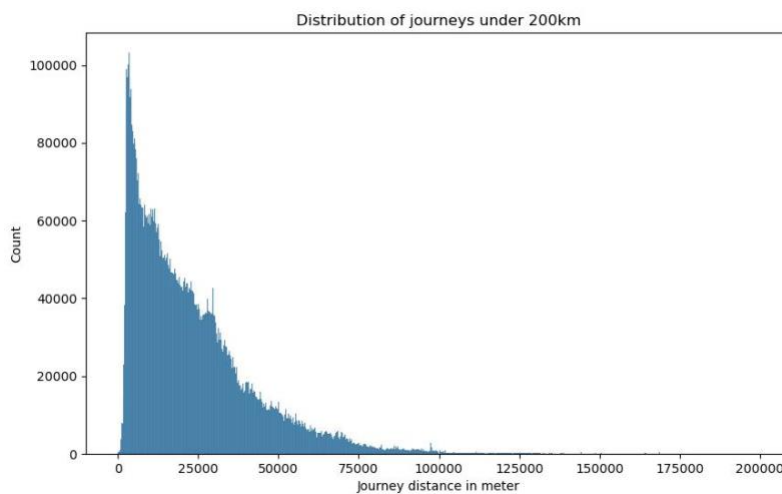
Commute journeys

Indeed, a learning is also that carpooling has mainly a short distance use to go to work. Here are some graphs to attest to this hypothesis:



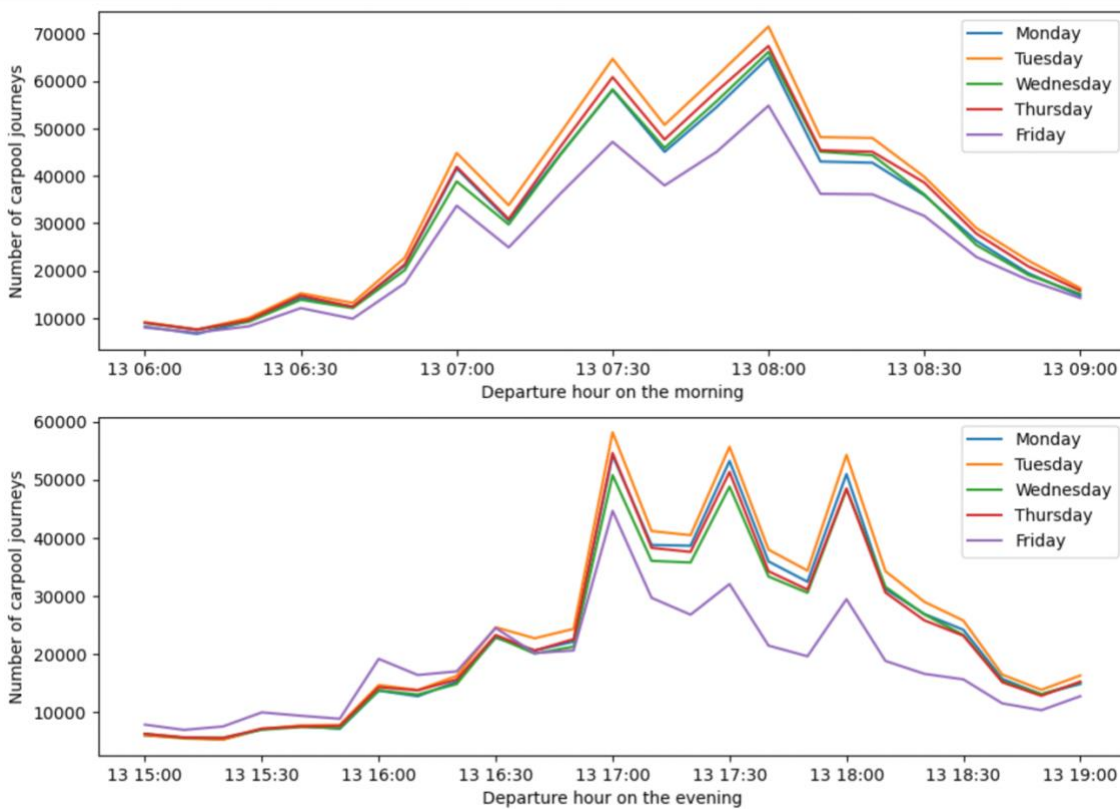
The analysis shows that there are more trips in weekdays than in weekends, and that the peak hours correspond to the commute trip hours.

According to the Ministry of Ecological Transition, short-distance carpooling is defined as a trip under 80 kilometers. We can see here with this right-skewed plot that most of the trips are short distance. Indeed, 75% of the carpool trip since 2019 are under 32km.



journey_distance	
count	9607669.00
mean	23576.18
std	22103.07
min	1.00
25%	8678.00
50%	18372.00
75%	31902.00
max	1132379.00

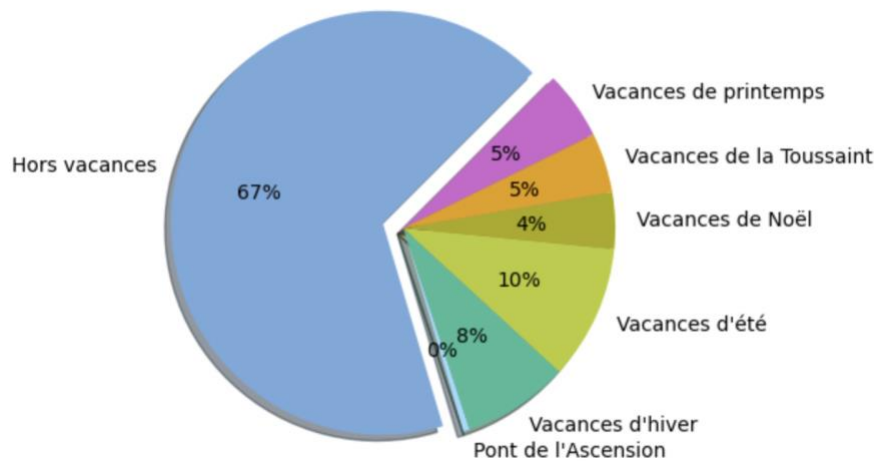
We can zoom in on the peak hours to see what time the carpoolers leave in the morning and evening. This analysis can be carried out at a finer territorial scale to anticipate the demands and the traffic.



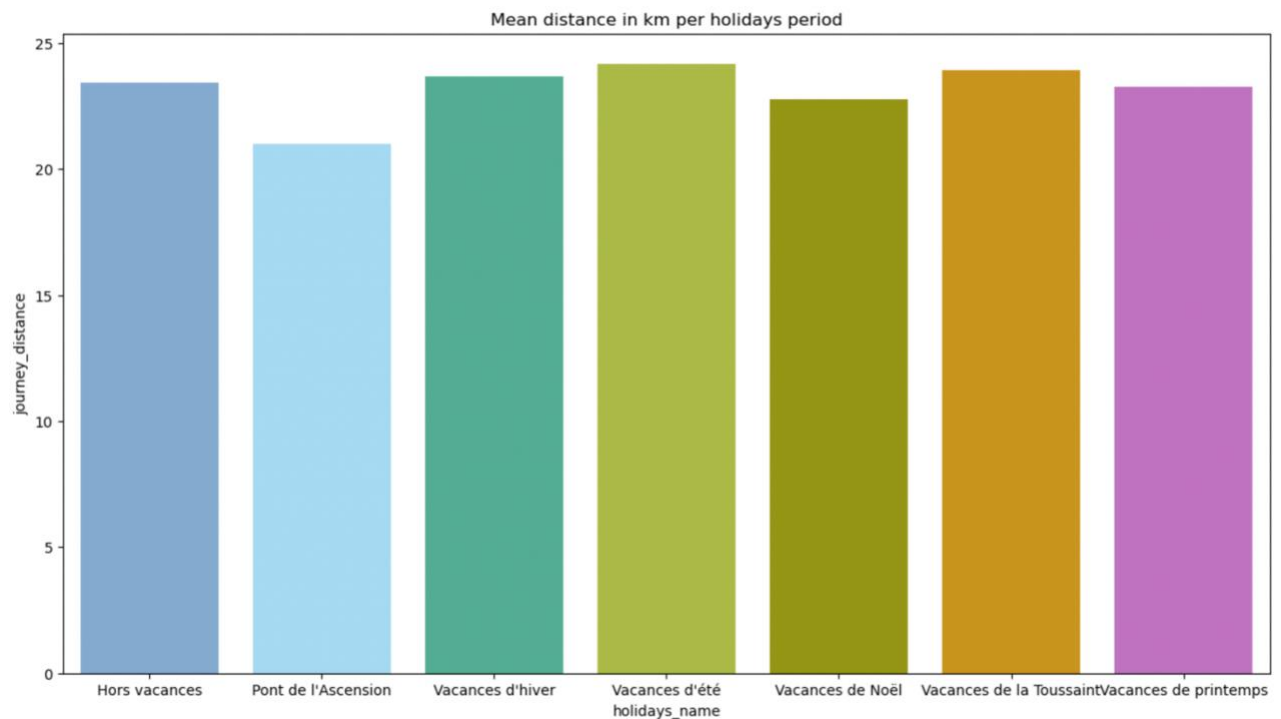
Vacation schedule

The calendar dataset allowed us to see the importance of weekday trips. It also allows us to compare trip volumes between work and vacation periods. As you can see from this graph, 67% of carpool trips occur during school hours.

Repartition of journeys according to holidays periods



In addition, I looked to see if the distances changed depending on the vacation period. There is no impact of holidays on the mean distance. It confirms that the proportion of commute trips is too big compared to the touristic use of carpooling.

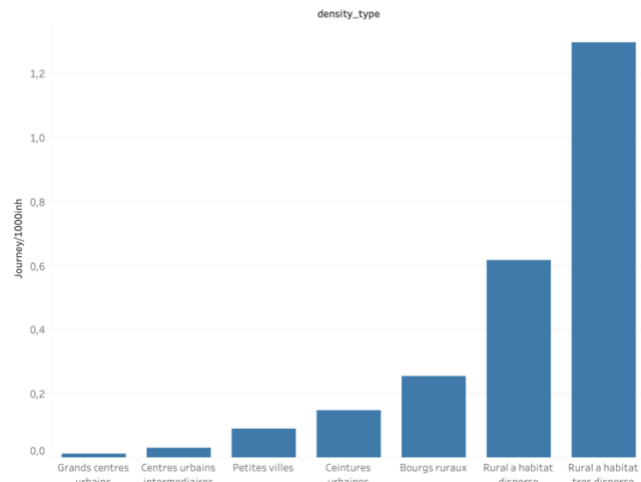


Territories

I choose to analyze also territorial data as density. We can see that in rural areas, there is more trips per inhabitants.

key_0	journeys_per_1000inh	density_type
1	0.009847	Grands centres urbains
2	0.033707	Centres urbains intermédiaires
3	0.096292	Petites villes
4	0.157009	Ceintures urbaines
5	0.264898	Bourgs ruraux
6	0.611251	Rural à habitat dispersé
7	1.750829	Rural à habitat très dispersé

Mean duration according to the density degree (jan 2023)



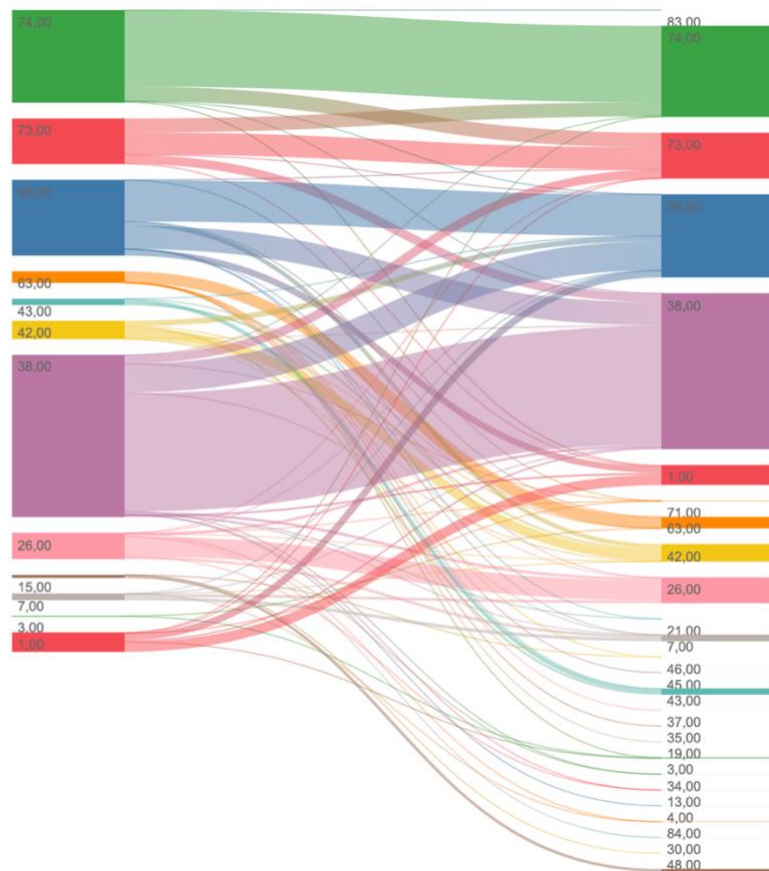
The objective is to be able to analyze the different territories according to their needs in terms of carpooling. The graph below allows us to distinguish several types of territories according to their out-of-town work population and their propensity to carpool.

Other visualization tools can help us to see on each territory how the carpool journeys are organized as this Sankey diagram for example.

The labels are the number of the administrative French departments.

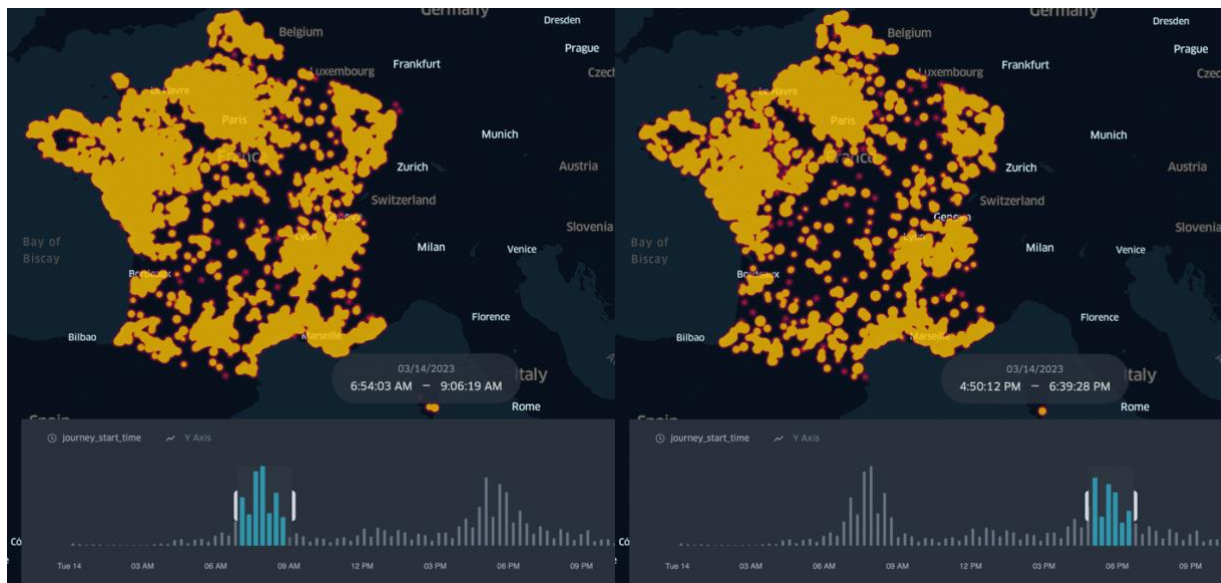
Sankey diagram of start and end department in the Auvergne-Rhone-Alpes region

journey_en..	1	3	7	15
1	540			
3	2	36		
4				
7			178	
13				
15				
19				
21				
26			147	
30				
34				
35				
37				
38	84		19	
42			7	
43				
45				
46				
48				
63		2		
69	371		3	
71	8			
73		11		
74	29			
83				
84			2	

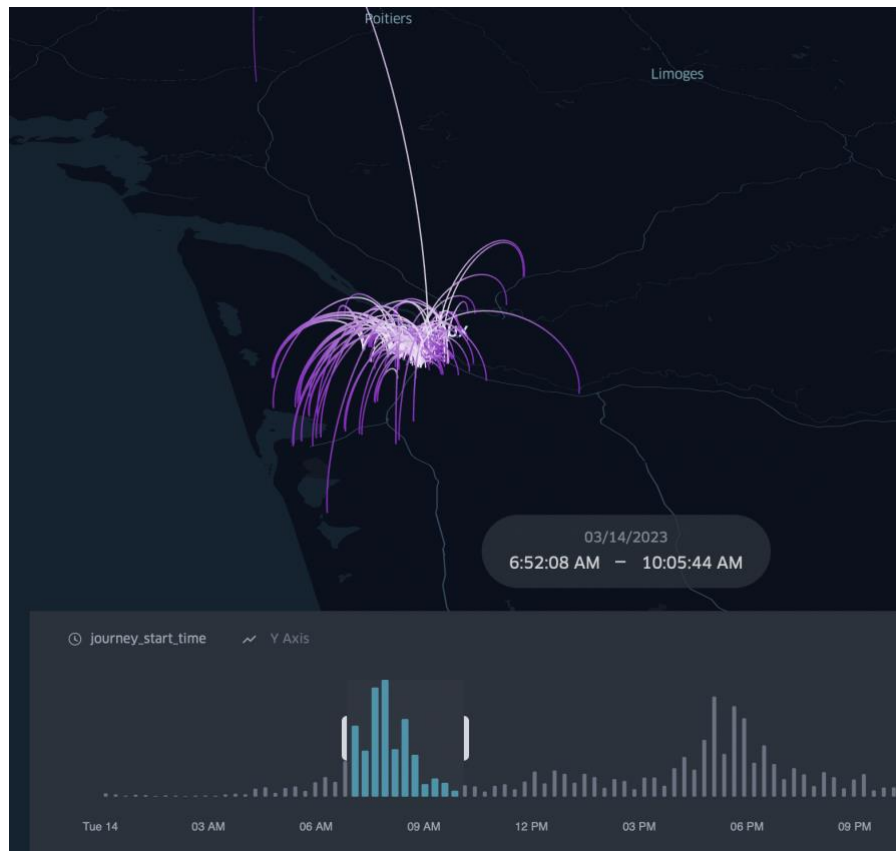


Also the application Kepler.gl helped me to visualize geographically my dataset, and we can also filter by region or department. This can be usefull to construct a public transport policy, helping to decide where to create alternative transports to cars, or where to reserve parking places for carpooling for example.

First with these two maps, we can see that departure places in the morning are dispersed. Departure place on the evening seems to be more concentrated in urbans areas. The other insight that jumps out on these maps is that we see the "diagonal of the void", famous concept to characterize French geography.



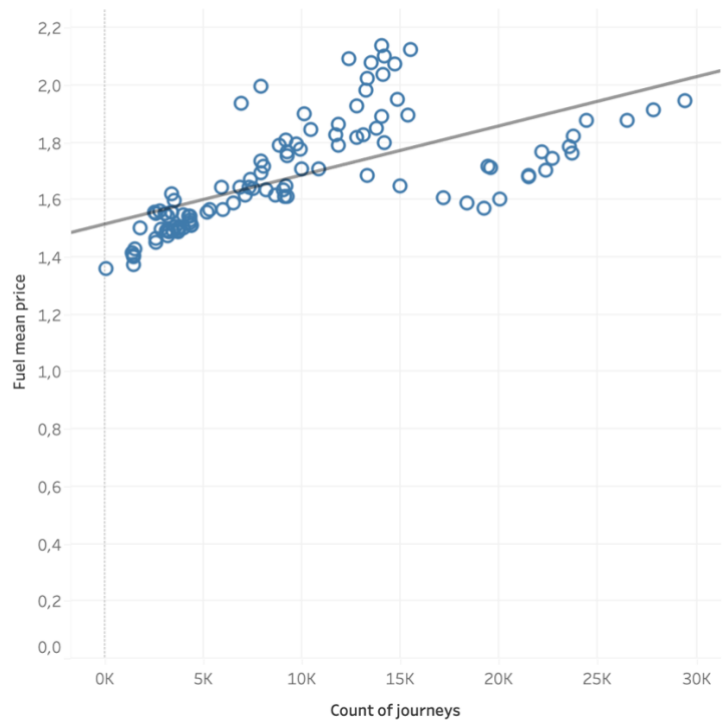
With this application, we can make a detailed territorial analysis of carpooling flows. For example, on the map below (which is interactive), we can see where the carpooling trips that arrive in the morning between 7am and 10am in the Bordeaux metropolis come from.



Fuel price

The comparison with the price of gasoline shows a very slight positive correlation, however, as we have seen that carpooling is mainly a practice for home-work trips, we can imagine that the habit and the organization have more impact than the price. More than the price itself, the growing uncertainty about the price of gasoline could have an impact on converting people to this practice, but this impact is not visible from one day to another.

Correlation between fuel price and number of trips



H. CONCLUSION

In conclusion, we have seen that on the one hand carpooling had suffered greatly from the covid epidemic, but that its use was picking up very strongly in 2022 and January 2023.

Most of the trips are short-distance carpools, since 75% of the trips are less than 32 km long. These short-distance trips correspond to home-work trips. Indeed, carpooling trips are more important during the week than at weekends, and outside of school vacations. Moreover, there are clearly peak hours between 6am and 9am in the morning and between 5pm and 7pm in the evening.

We can see a tendency to go to the urban centers in the morning and leave in the evening, but this needs to be studied at a finer territorial level. This database allows for more precise studies to observe a particular territory. For example, by comparing the percentage of commute person in the population with the percentage of trips per person, we can identify territories that represent a development potential for carpooling. By using this data, I wanted to show the relevance of observing a territory in order to propose relevant travel solutions to the inhabitants.

In this database, I linked the carpooling data to other dimensional data to examine the factors exogenous to carpooling. Uncertainty about gasoline prices may convert to carpooling even if there is no immediate correlation. The rankings of the different types of jurisdictions show that there is not a simple correlation between population and number of trips. The practice of carpooling seems to depend on the willingness of politicians and companies to set up platforms and incentives to encourage carpooling, and to make this practice a habit.

I. LINKS

Link to Github repository :

https://github.com/ClemenceLegrand/Final_project_carpooling.git

Link to Jira Project :

<https://clemencelegrand.atlassian.net/jira/core/projects/FP/board?atlOrigin=eyJpIjoiOWVhYzA1Y2E3YW5mNDdlYmJiYTc2ZmM4YmU0NTU1YzciLCJwIjoiaSJ9>