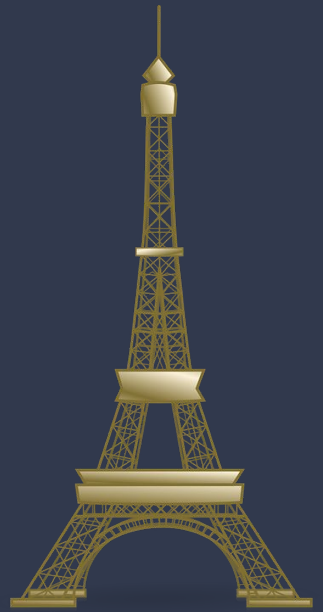
French Salaries
[Aka companies in french cities]

Angela Salgado and Jaime Sastre
Ironhack Paris 11/2022
Project week 4



Index.

1. Dataset Description

```
salary.head()
✓ 0.2s
```

	CODGEO	LIBGEO	REG	DEP	E14TST	E14TS0ND	E14TS1	E14TS6	E14TS10	E14TS20	E14TS50	E14TS100	E14TS200	E14TS500
0	01001	L'Abergement-Clémenciat	82	01	25	22	1	2	0	0	0	0	0	0
1	01002	L'Abergement-de-Varey	82	01	10	9	1	0	0	0	0	0	0	0

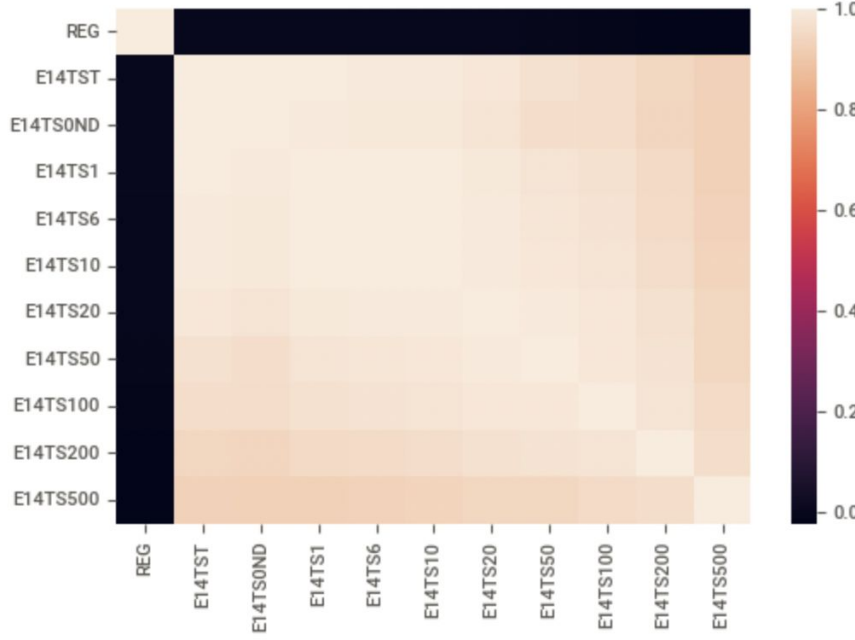
```
salary.isnull().sum()
✓ 0.3s

CODGEO      0
LIBGEO      0
REG         0
DEP         0
E14TST      0
E14TS0ND    0
E14TS1      0
E14TS6      0
E14TS10     0
E14TS20     0
E14TS50     0
E14TS100    0
E14TS200    0
E14TS500    0
dtype: int64
```

```
salary.shape
✓ 0.2s

(36681, 14)
```

	REG	E14TST	E14TS0ND	E14TS1	E14TS6	E14TS10	E14TS20	E14TS50	E14TS100	E14TS200	E14TS500
REG	1.000000	-0.002858	-0.002906	-0.001119	-0.005153	-0.005035	-0.007237	-0.008729	-0.015250	-0.020460	-0.023357
E14TST	-0.002858	1.000000	0.999489	0.996461	0.995297	0.993956	0.984406	0.969626	0.962369	0.943975	0.924913
E14TS0ND	-0.002906	0.999489	1.000000	0.993396	0.991979	0.990579	0.979331	0.963099	0.956575	0.938243	0.921855
E14TS1	-0.001119	0.996461	0.993396	1.000000	0.998088	0.996207	0.990042	0.978388	0.968586	0.949235	0.923146
E14TS6	-0.005153	0.995297	0.991979	0.998088	1.000000	0.998597	0.994095	0.983196	0.973983	0.955951	0.929608
E14TS10	-0.005035	0.993956	0.990579	0.996207	0.998597	1.000000	0.995787	0.985659	0.977239	0.959891	0.934023
E14TS20	-0.007237	0.984406	0.979331	0.990042	0.994095	0.995787	1.000000	0.992583	0.985338	0.970295	0.940543
E14TS50	-0.008729	0.969626	0.963099	0.978388	0.983196	0.985659	0.992583	1.000000	0.987325	0.974962	0.943851
E14TS100	-0.015250	0.962369	0.956575	0.968586	0.973983	0.977239	0.985338	0.987325	1.000000	0.979842	0.955842
E14TS200	-0.020460	0.943975	0.938243	0.949235	0.955951	0.959891	0.970295	0.974962	0.979842	1.000000	0.961836
E14TS500	-0.023357	0.924913	0.921855	0.923146	0.929608	0.934023	0.940543	0.943851	0.955842	0.961836	1.000000



2. Challenges

Find what to analyse in our charts

Look for the proper charts to display what we want

Create a column of Latitude and Longitude from the name of the town

3. Project process

1. Planning on Trello
2. Python Code
    a. Download dataset
    b. Examine the dataset for the possible issues. (Too beauty to be true)
    c. Pivot tables, groupby
    d. Charts to display insights (tough, because we didn't have a lot of info)
3. Conclusions in the Readme file

4. Learnings

-We tried to draw as many inferences as possible even though the data was quite sterile.

Work with the different Python libraries for data visualization

5. Improvements

Improve data visualization. It takes a long time to build the charts, and we didn't manage to customize them to everything we wanted: colors, shapes, legend etc.

We would like to be able to incorporate data on salaries, which was initially the name of the dataset

6. Comparison of the initial dataset and the final one

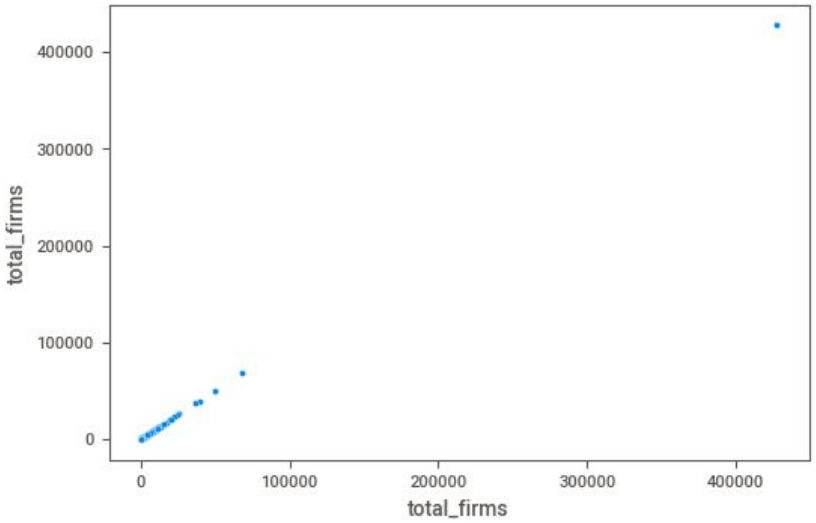
We add three columns in order to group the number of employees of the companies into three groups: small, medium and large companies

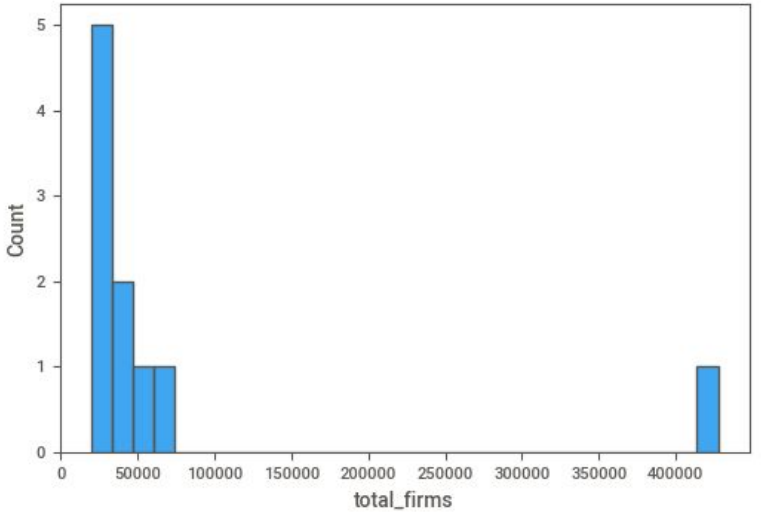
7. Highlights

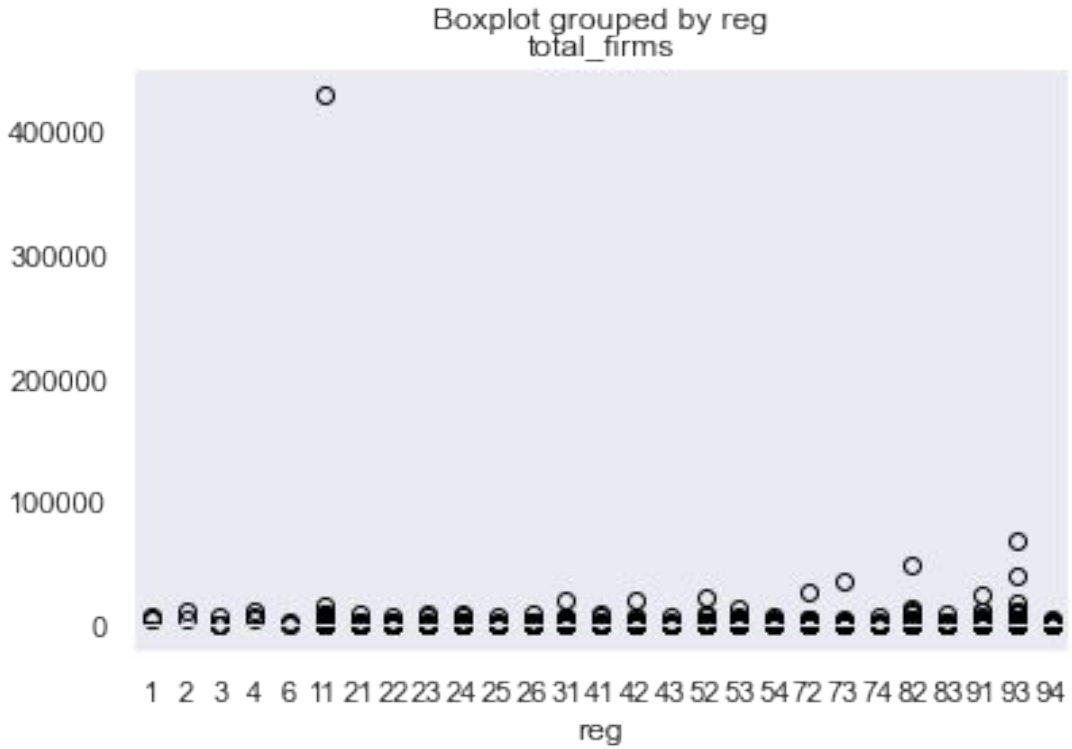
-Number of companies

-Number of employees

Number of companies in centralized

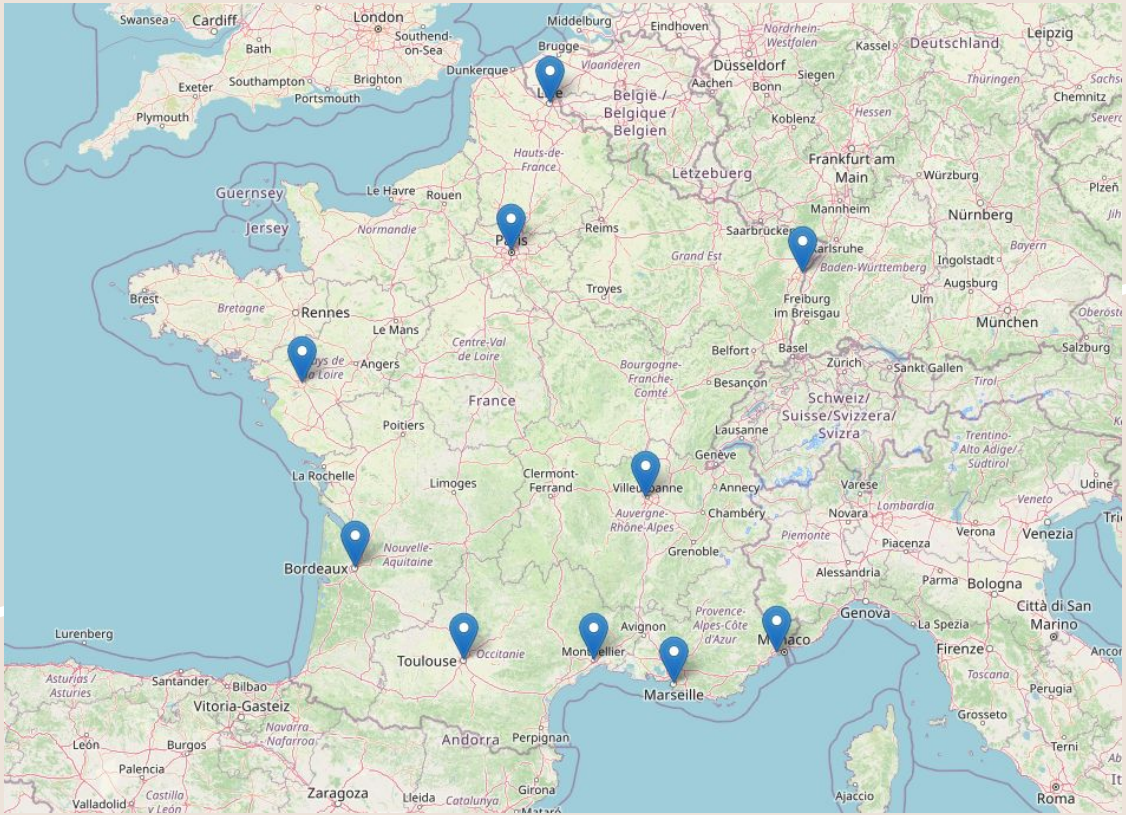




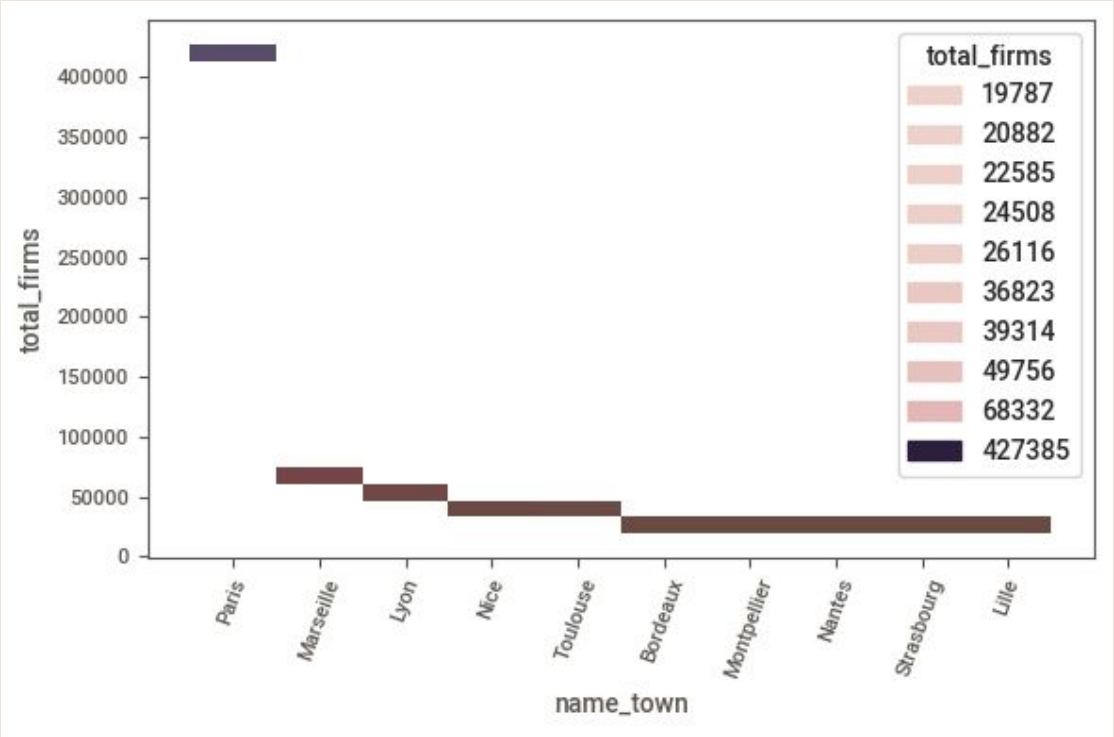

Boxplot grouped by reg
total_firms
400000
300000
200000
100000
0
1 2 3 4 6 11 21 22 23 24 25 26 31 41 42 43 52 53 54 72 73 74 82 83 91 93 94
reg

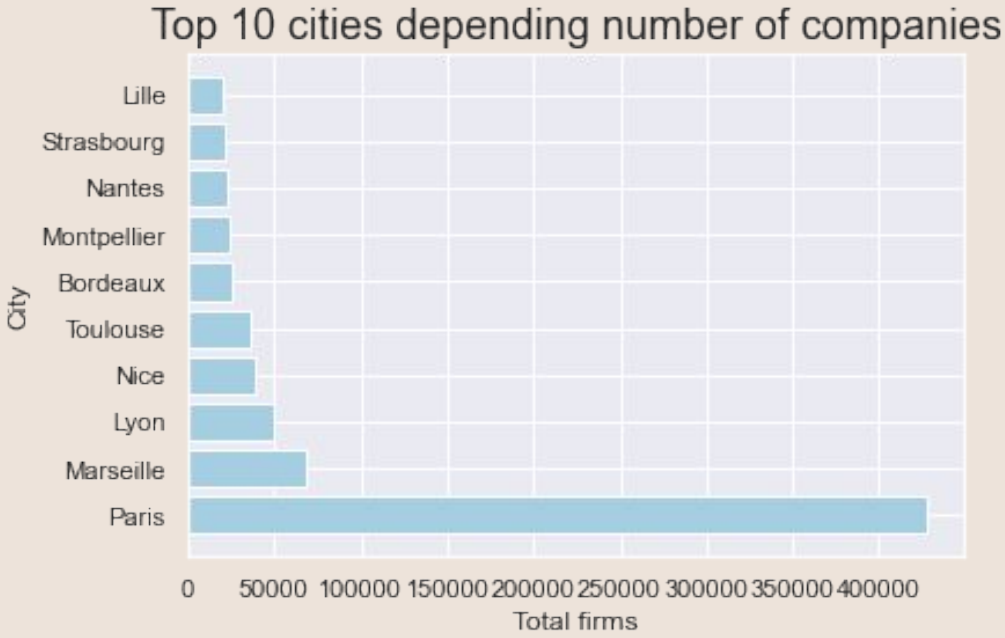
The location of companies is dramatically centralized in Paris.

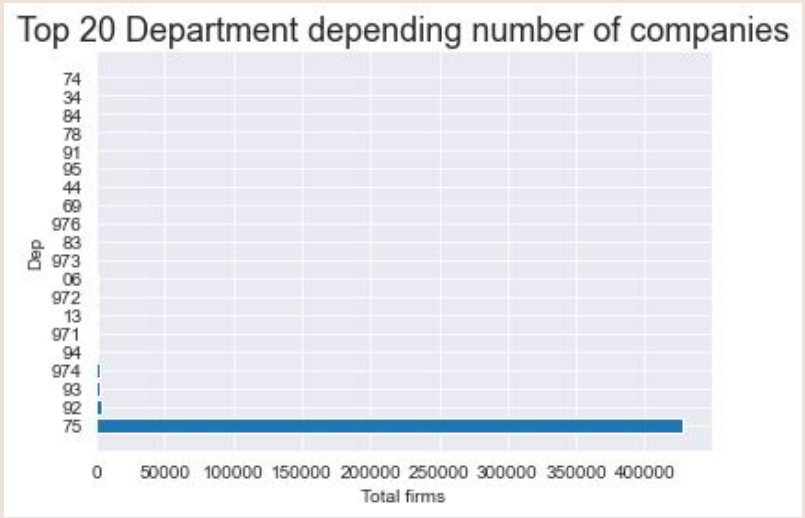
Top 10 cities depending of numbers of companies

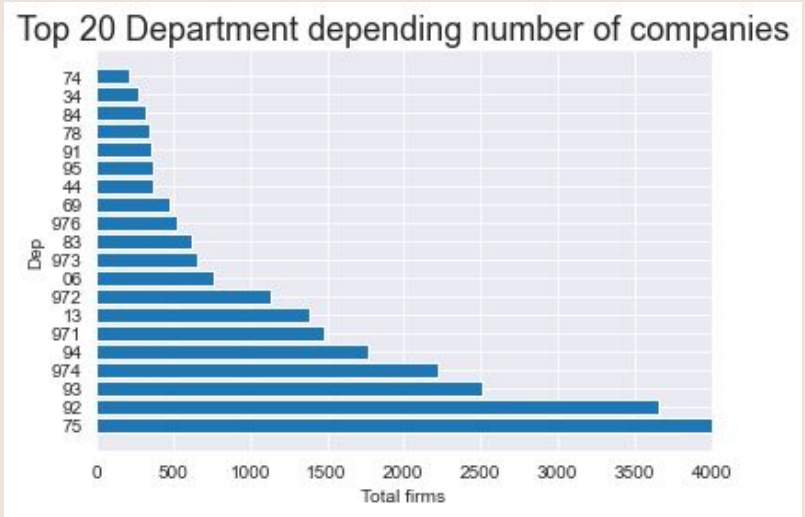


Paris has 7 times more companies than the second city with the most companies in France




Top 10 cities depending number of companies
Lille
Strasbourg
Nantes
Montpellier
Bordeaux
Toulouse
Nice
Lyon
Marseille
Paris
City
0
50000
100000
150000
200000
250000
300000
350000
400000
Total firms





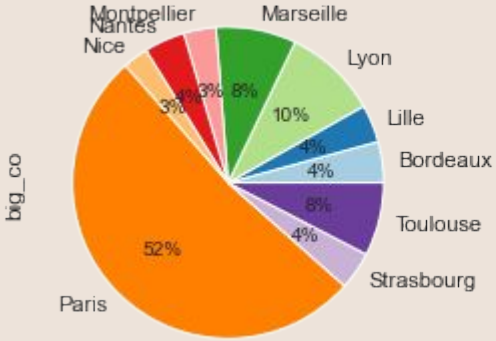
To the centralization of the number of companies in Paris, we have also added the dynamics of 93 and 92 departments, which concentrate high numbers of companies.

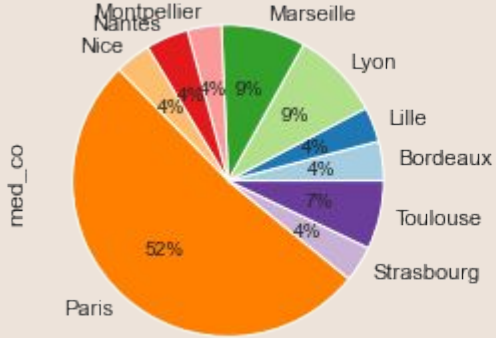
Number of employees

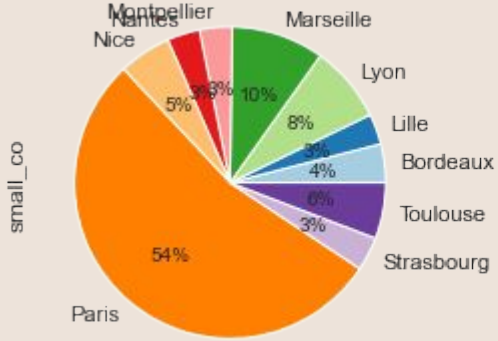
Small companies = 1-19 employees

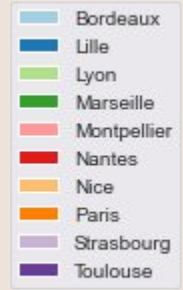
Medium companies = 20 - 199

Big companies =200 ….

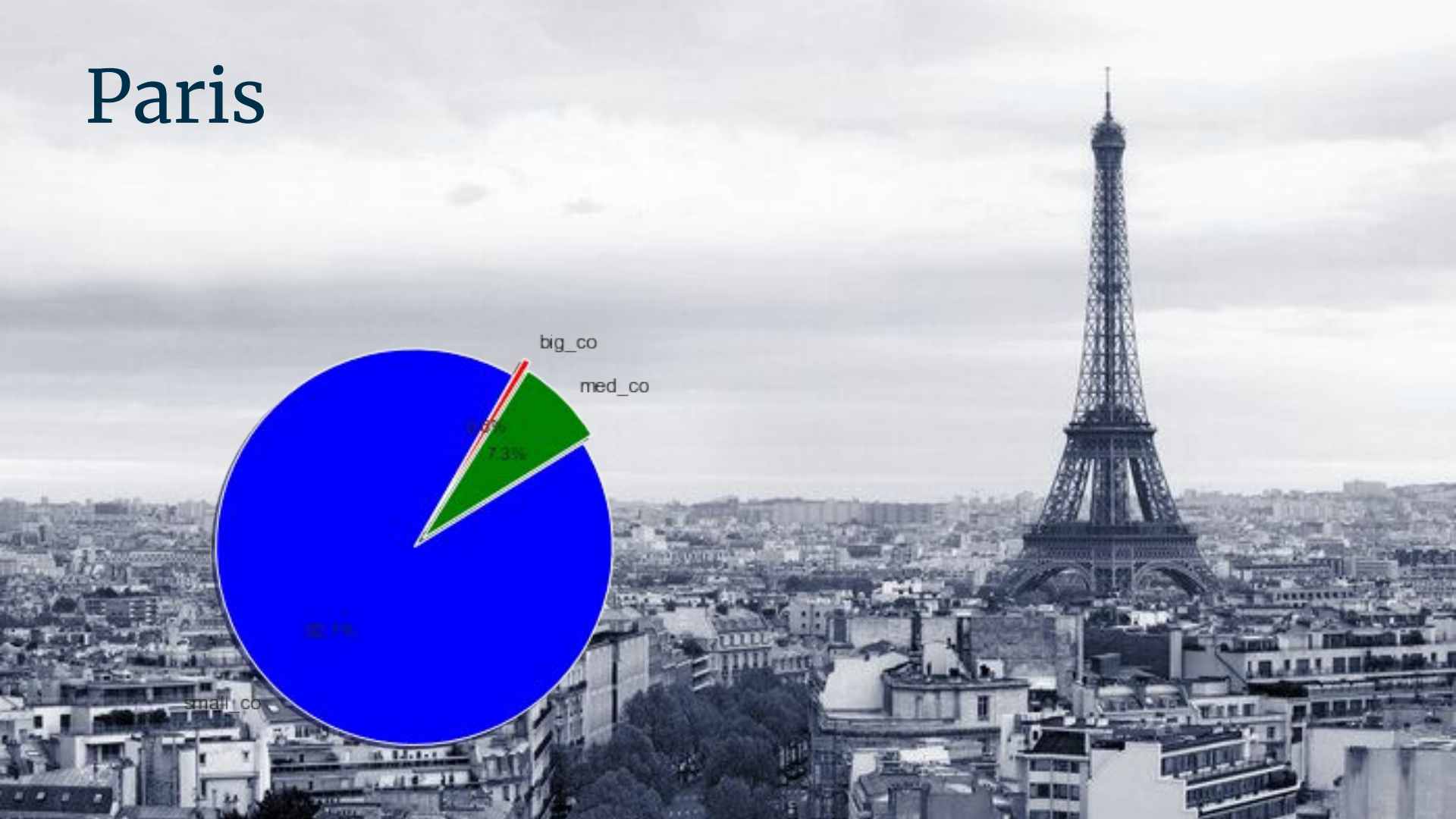

big_co
Montpellier
Nantes
Nice
Marseille
Lyon
Lille
Bordeaux
Toulouse
Strasbourg
Paris
3%
4%
3%
8%
10%
4%
4%
6%
4%
52%


med_co
Montpellier
Nantes
Nice
Marseille
Lyon
Lille
Bordeaux
Toulouse
Strasbourg
Paris
4%
4%
4%
9%
9%
4%
4%
7%
4%
52%

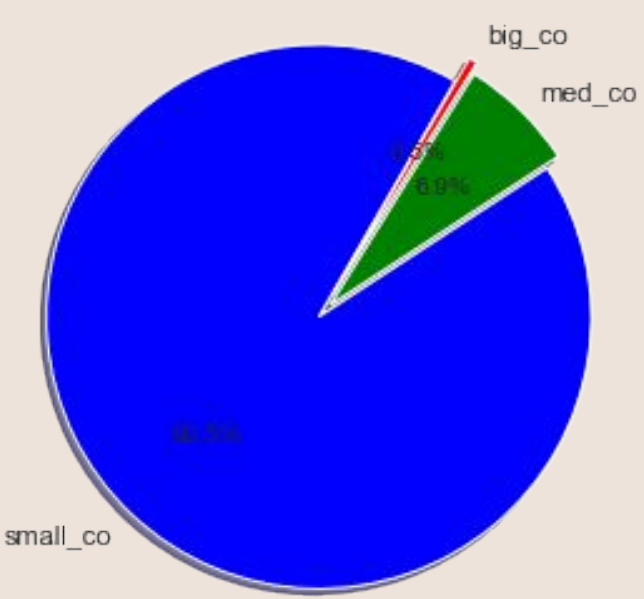

small_co
Montpellier
Nantes
Nice
Marseille
Lyon
Lille
Bordeaux
Toulouse
Strasbourg
Paris
3%
3%
5%
10%
8%
3%
4%
6%
3%
54%


Bordeaux
Lille
Lyon
Marseille
Montpellier
Nantes
Nice
Paris
Strasbourg
Toulouse

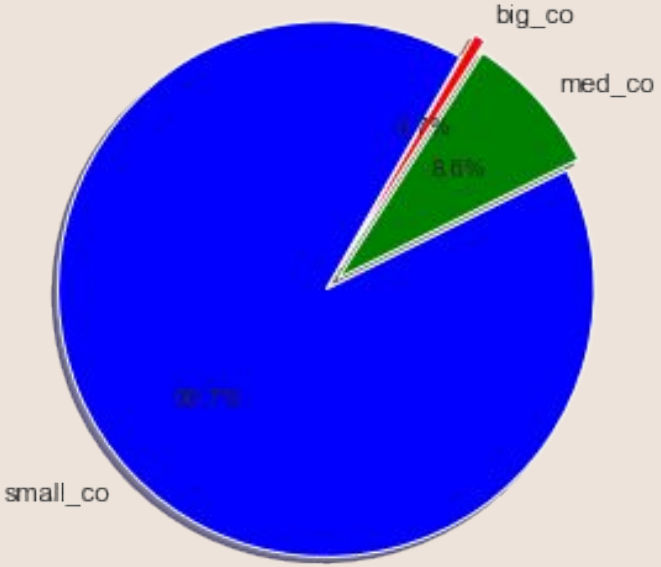
Paris



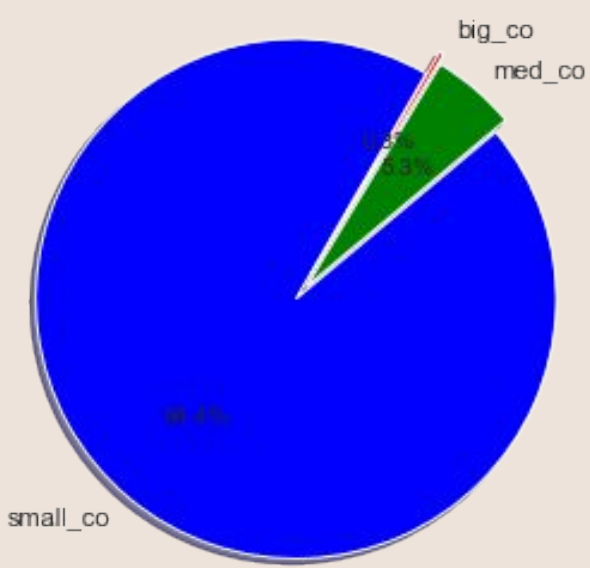
Marseille



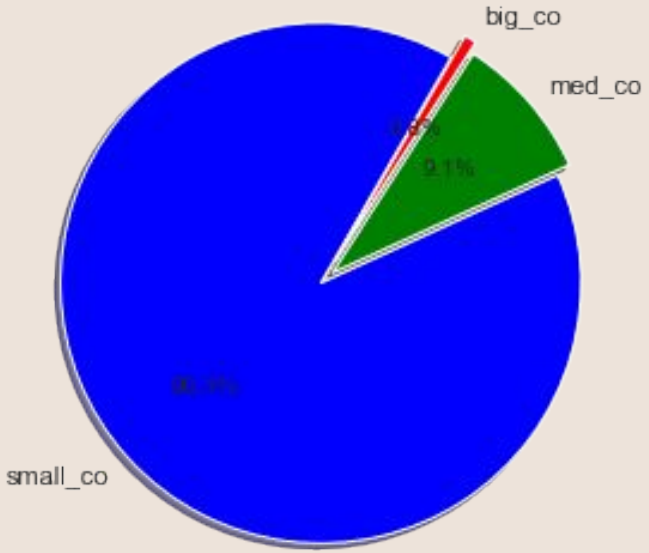
Lyon

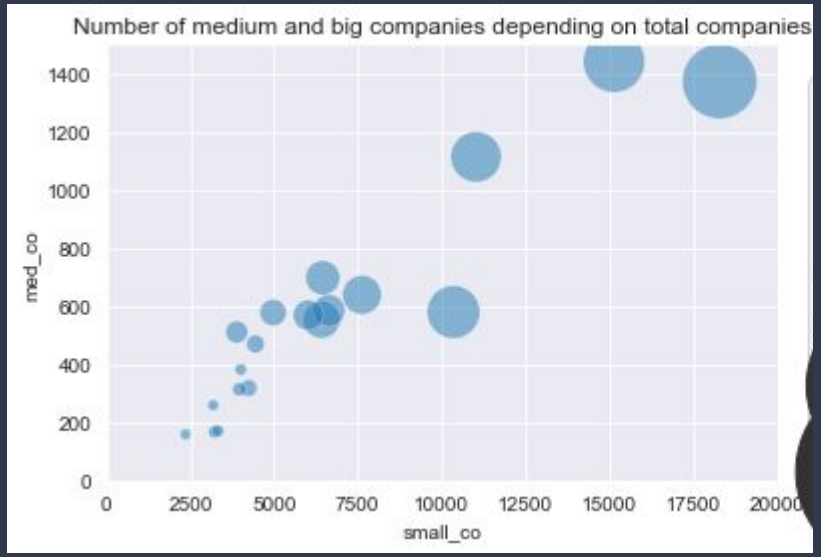


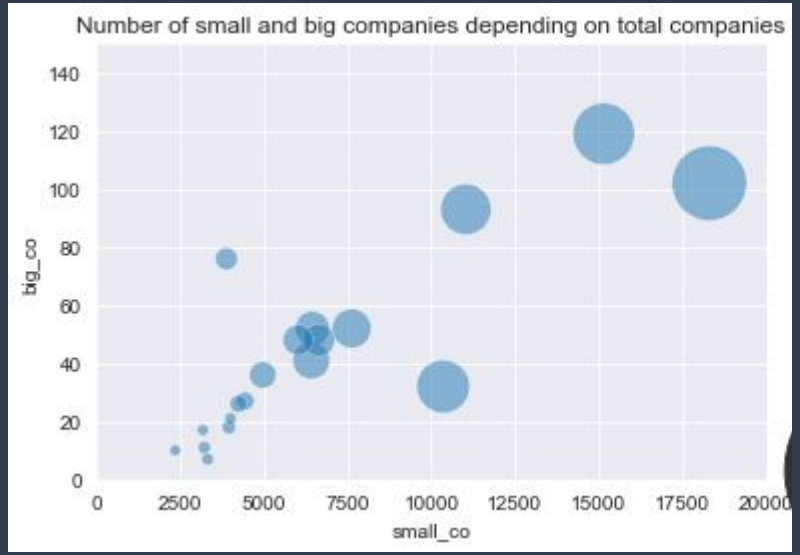
Nice



Toulouse




Number of medium and big companies depending on total companies
med_co
small_co
0
200
400
600
800
1000
1200
1400
0
2500
5000
7500
10000
12500
15000
17500
20000


Number of small and big companies depending on total companies
big_co
small_co
0
20
40
60
80
100
120
140
0
2500
5000
7500
10000
12500
15000
17500
20000

Q&A :)