

# Data Visualization

Teacher: **Marco Guerzoni**

SEDDIK Mohamed   16/11/2021 11H   Software: Rstudio   M2   DS2E

# PLAN

1→ descriptive analysais : Price Distribution , Churners Distribution ,

Correlation Matrix

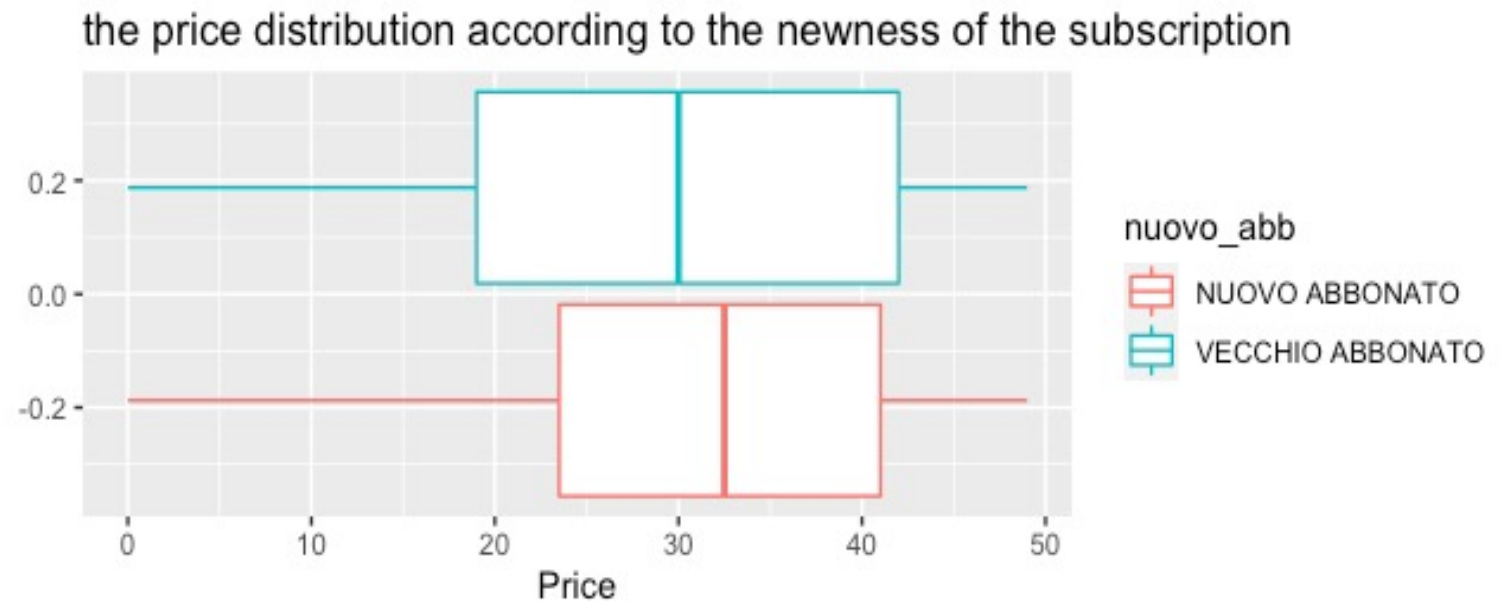
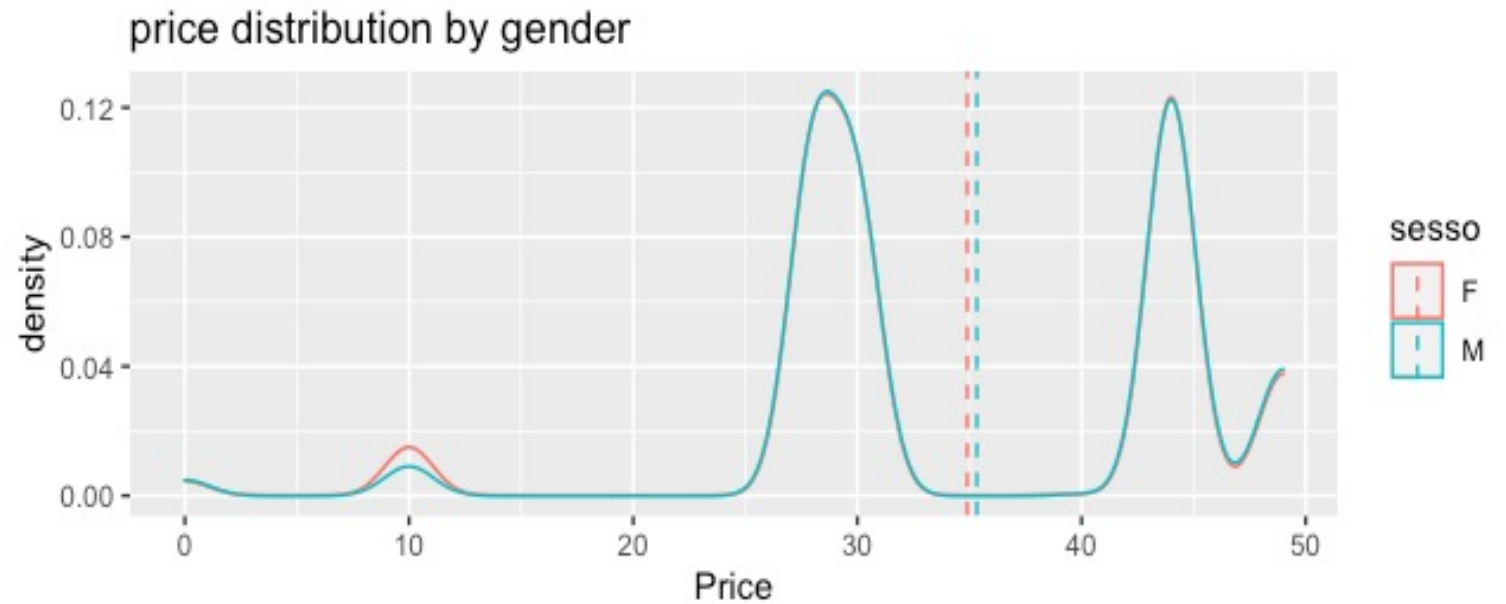
2→ Cluster and geographical analysais.

3→ Prediction analysis "Beta version"

# descriptive analysis : Price Distribution

→ Men and Women  
payed same price

→ The new Card are  
mainly more  
expensive

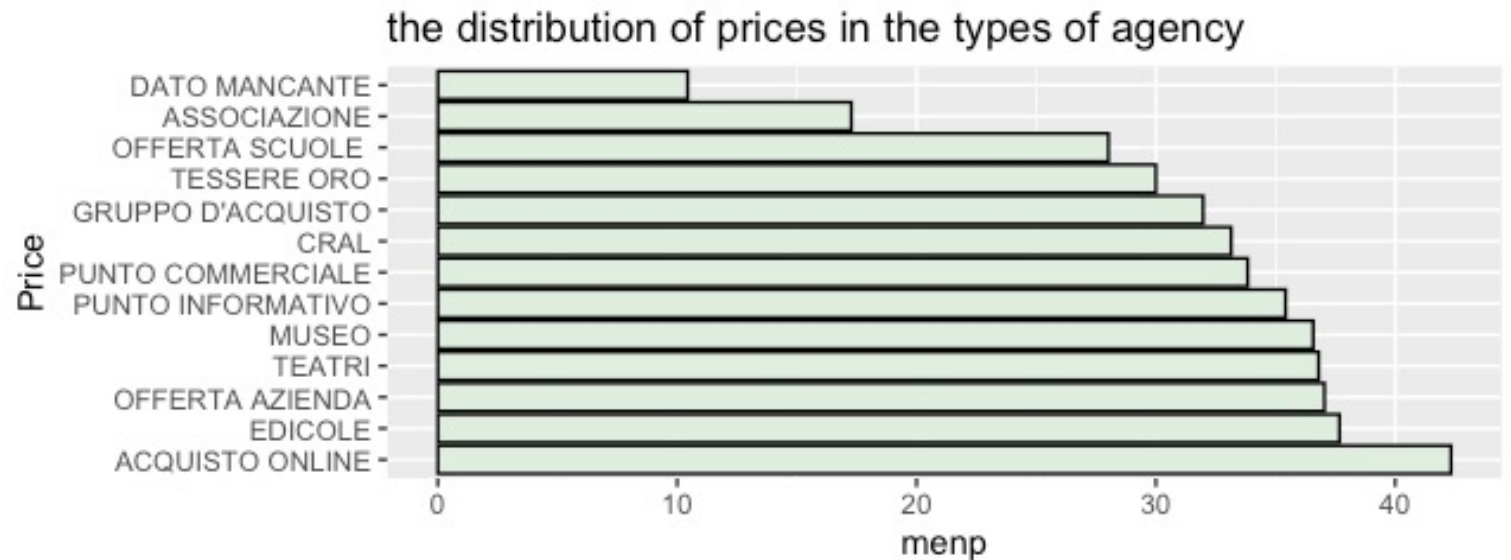
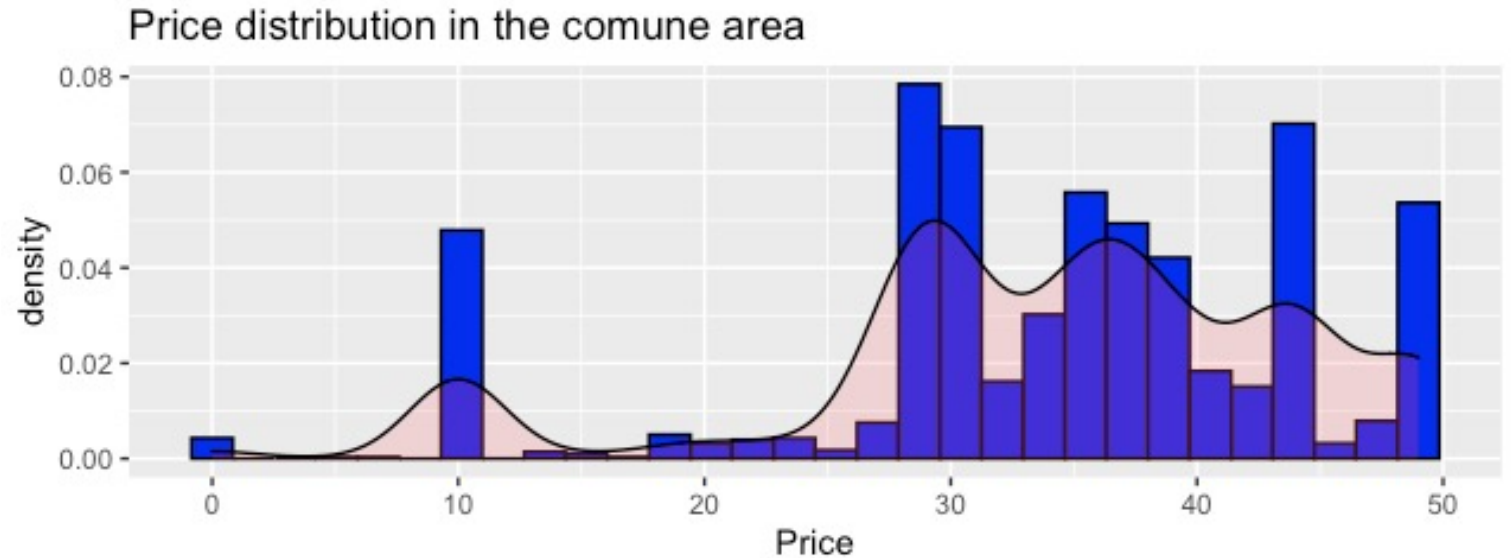


## descriptive analysis : Price Distribution

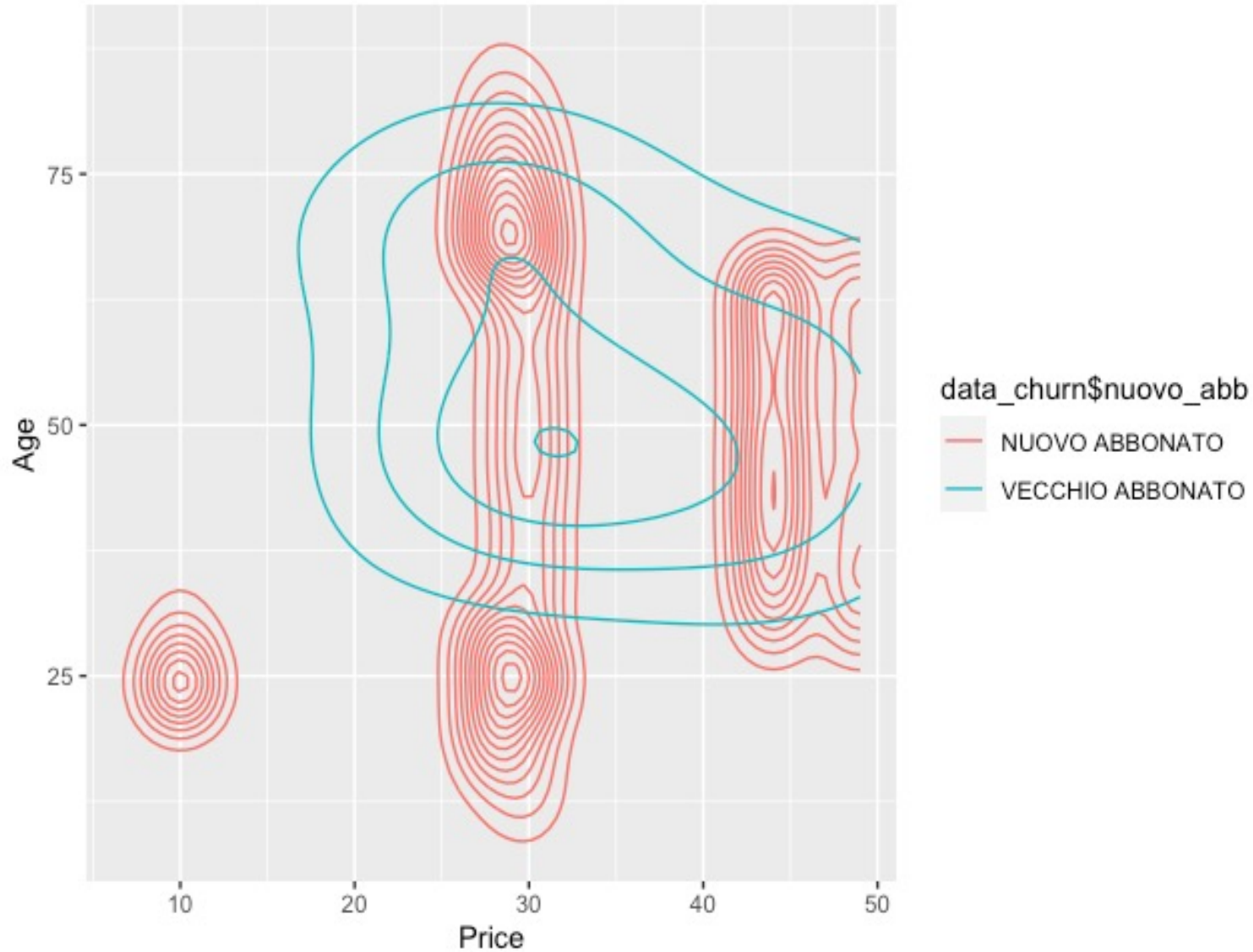
→ Most of commune price card are between 30 and 54

→ the online price is the most expensive

→ NA and association the less expensive



concentration of churners by age and card price



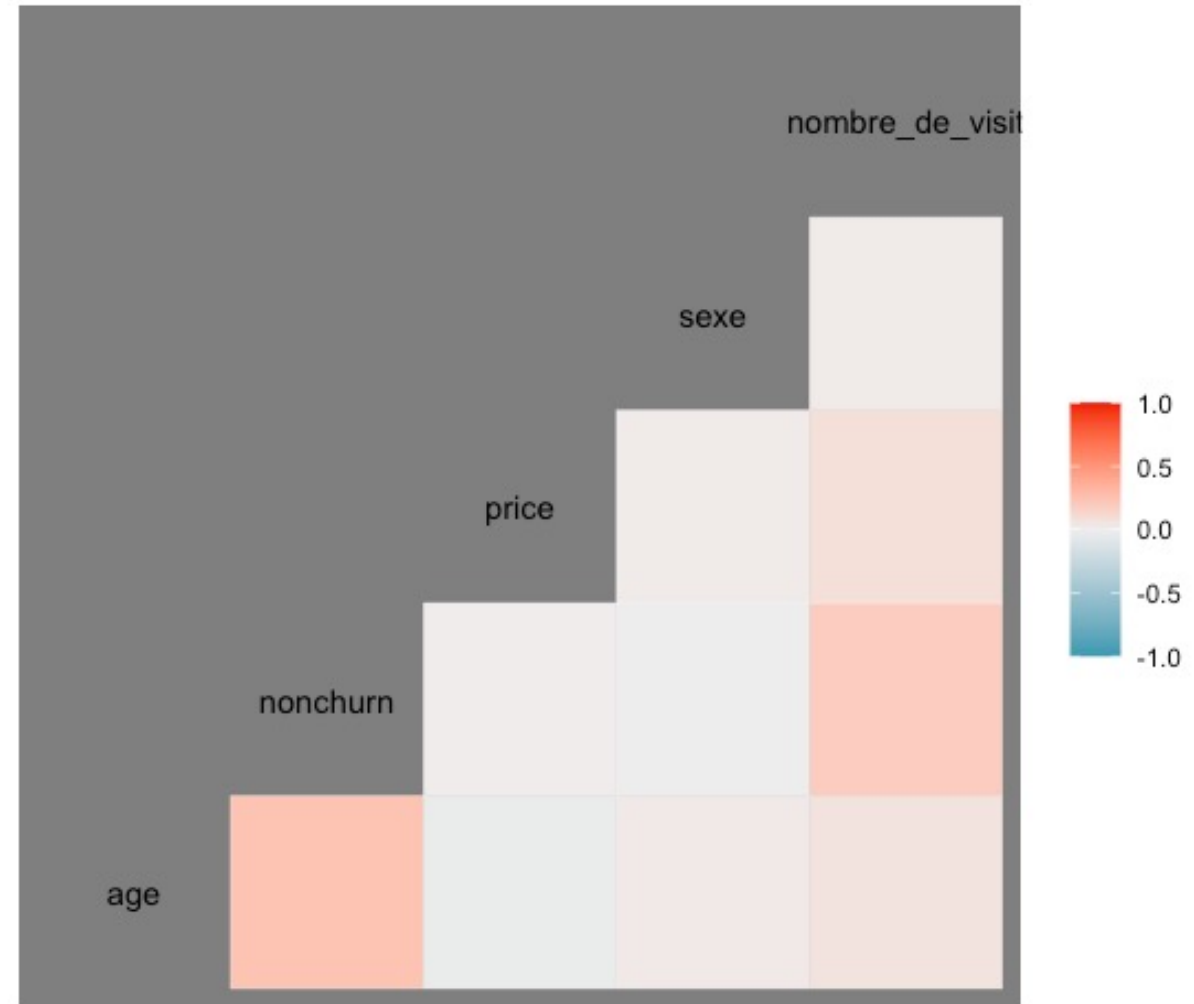
## Churners Distribution

- ➔ there are many concentration of new subscribers churners
- ➔ One little concentration Of old subscribers churners

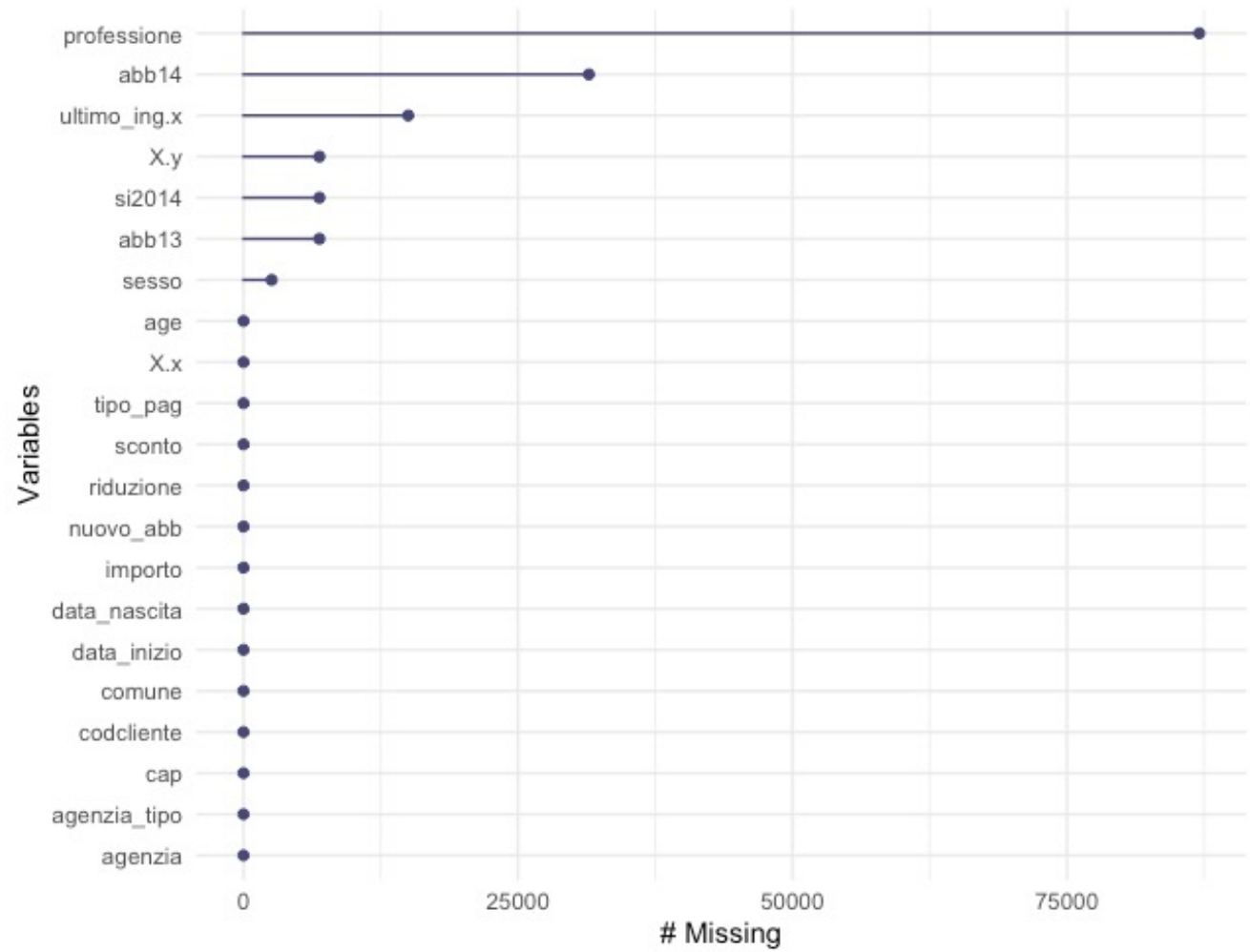
## descriptive analysais : Correlation Matrix

- → many variables non correlated
  - positive relationship between age and no-churn , number of visits and no-Churn

Correlation Matrix



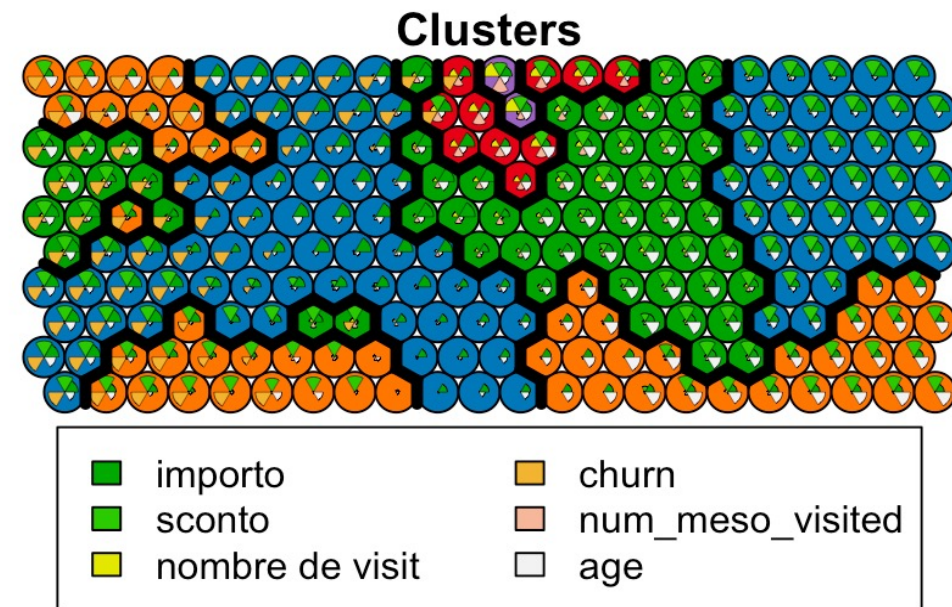
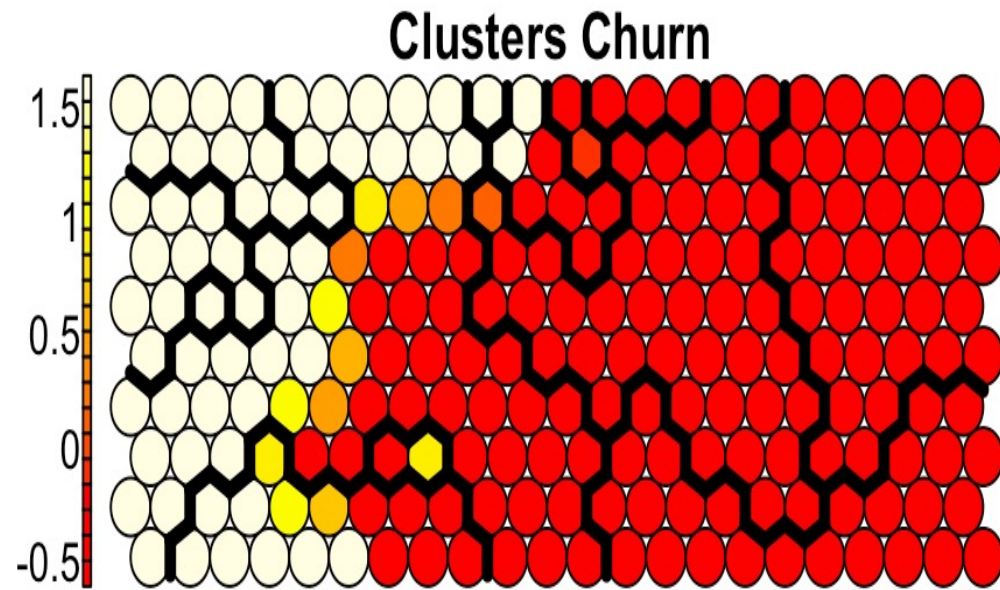
# NA Analysis





# Cluster and geographical analysais.

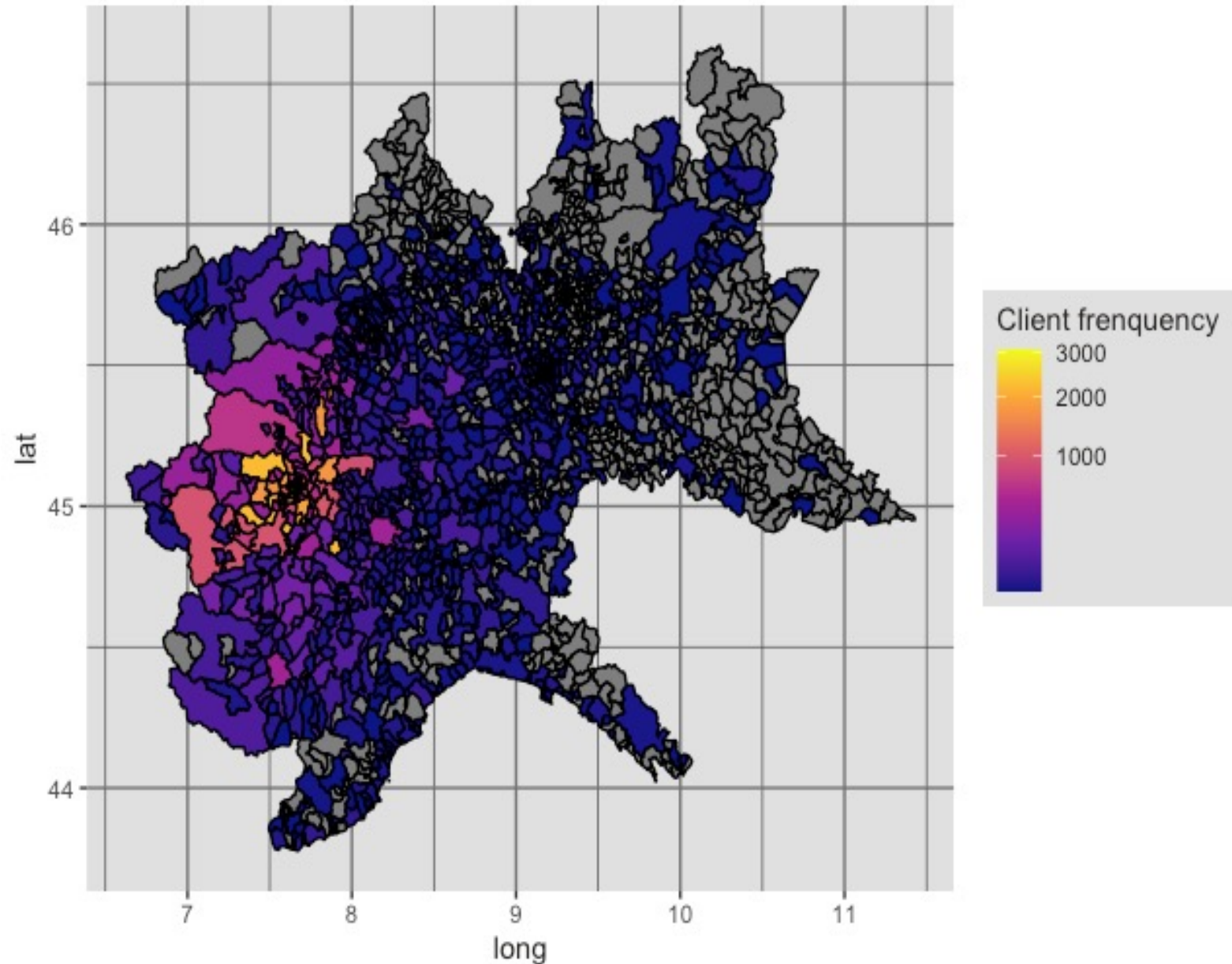
K=5





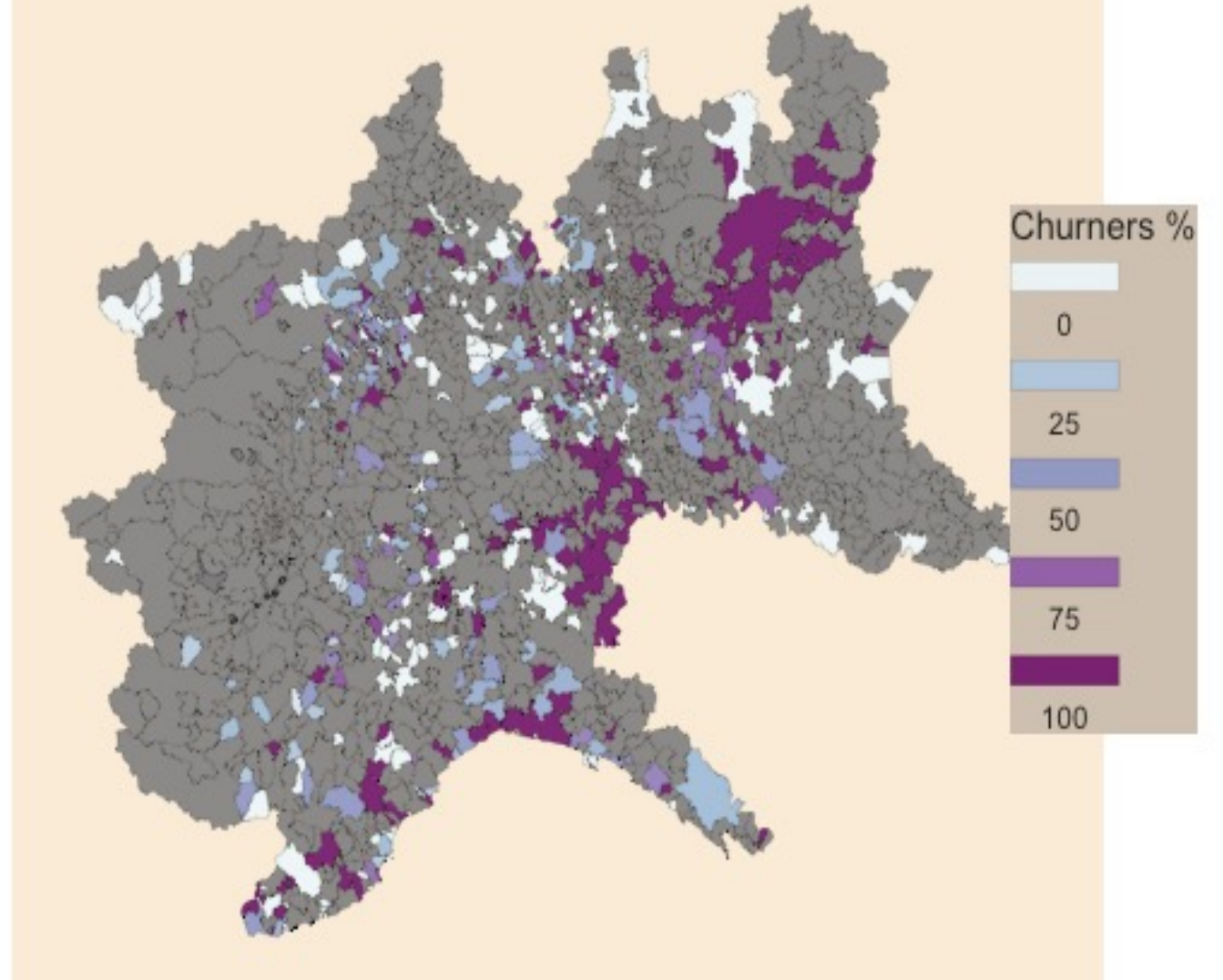
# Cluster and geographical analysis.

Geographical distribution of card holders in Turin



Cluster and  
geographical  
analysis

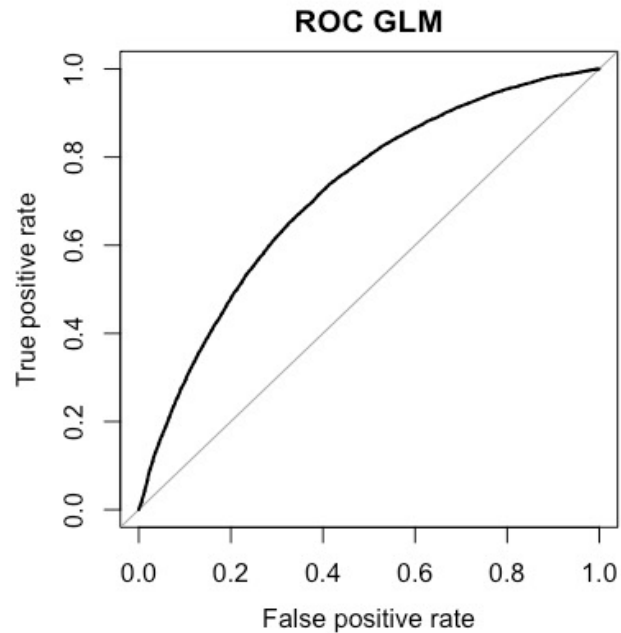
geographical distribution of churners by sector



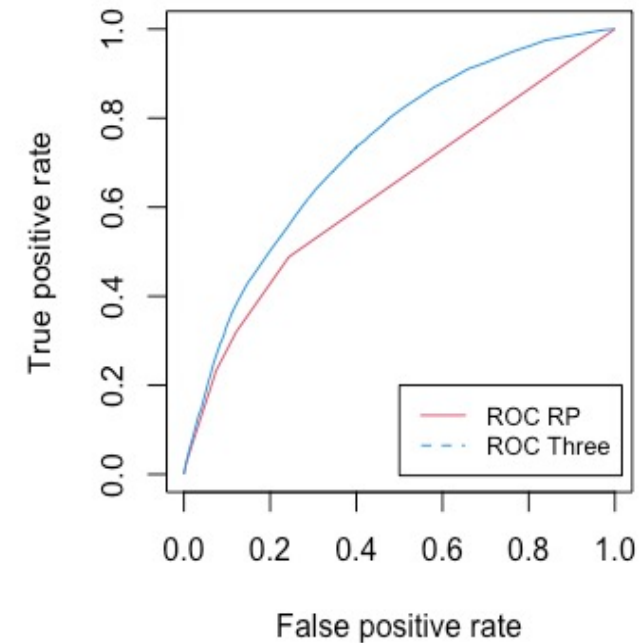
# Prediction analysis

## "Beta version"

- ROC curves comparing "Logit vs Tree vs recursive partitioning
- Features: Age , gendre , num\_museo\_visited , num\_visit , new\_abb , price , discount



**ROC curves comparing classification performance**



# Prediction analysis "Beta version"

---

- → logit and tree prediction of profit are pretty similar then real prediction
- → RP lose in similarty versus logit and tree.

