# PYTHON FOR DATA ANALYSIS

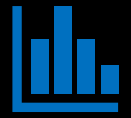
## DIABETES IN 130-US HOSPITALS


Alexandre BARRAT

Théo JAOUDET

# SUMMARY

Presentation of the subject and the dataset



First analysis of the data

Deeper analysis on specific criterias

Machine learning

# Presentation of the subject and the dataset

Diabetes in 130-US hospitals
between 1999 and 2008

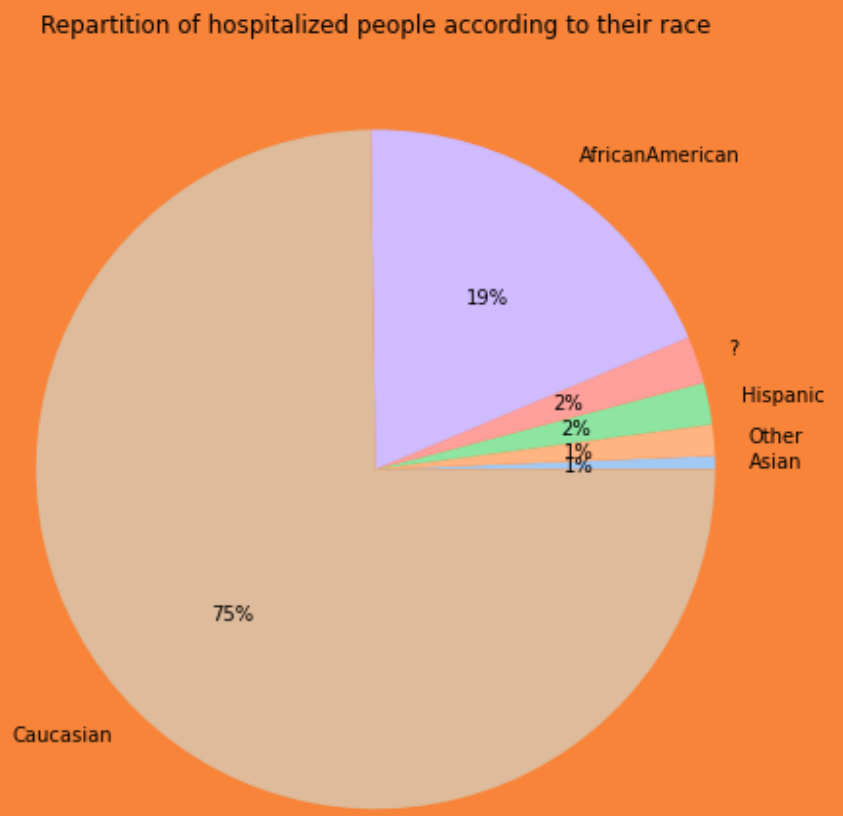
Polish Companies Bankruptcy

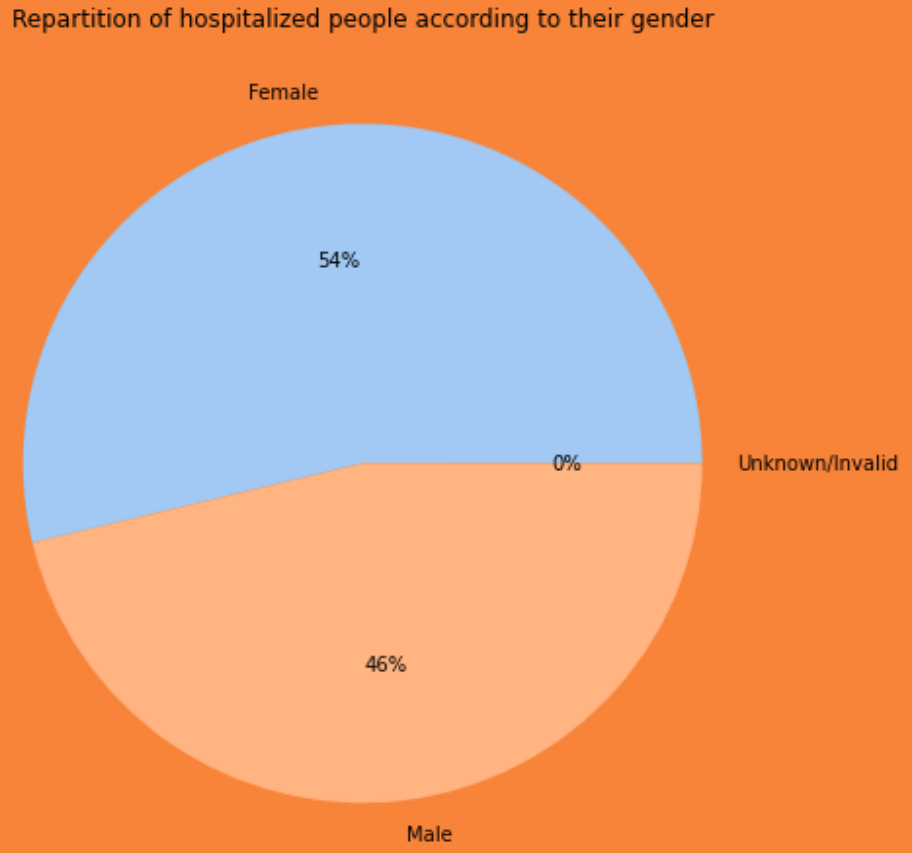
# Presentation of the subject and the dataset

Diabetes in 130-US hospitals between 1999 and 2008

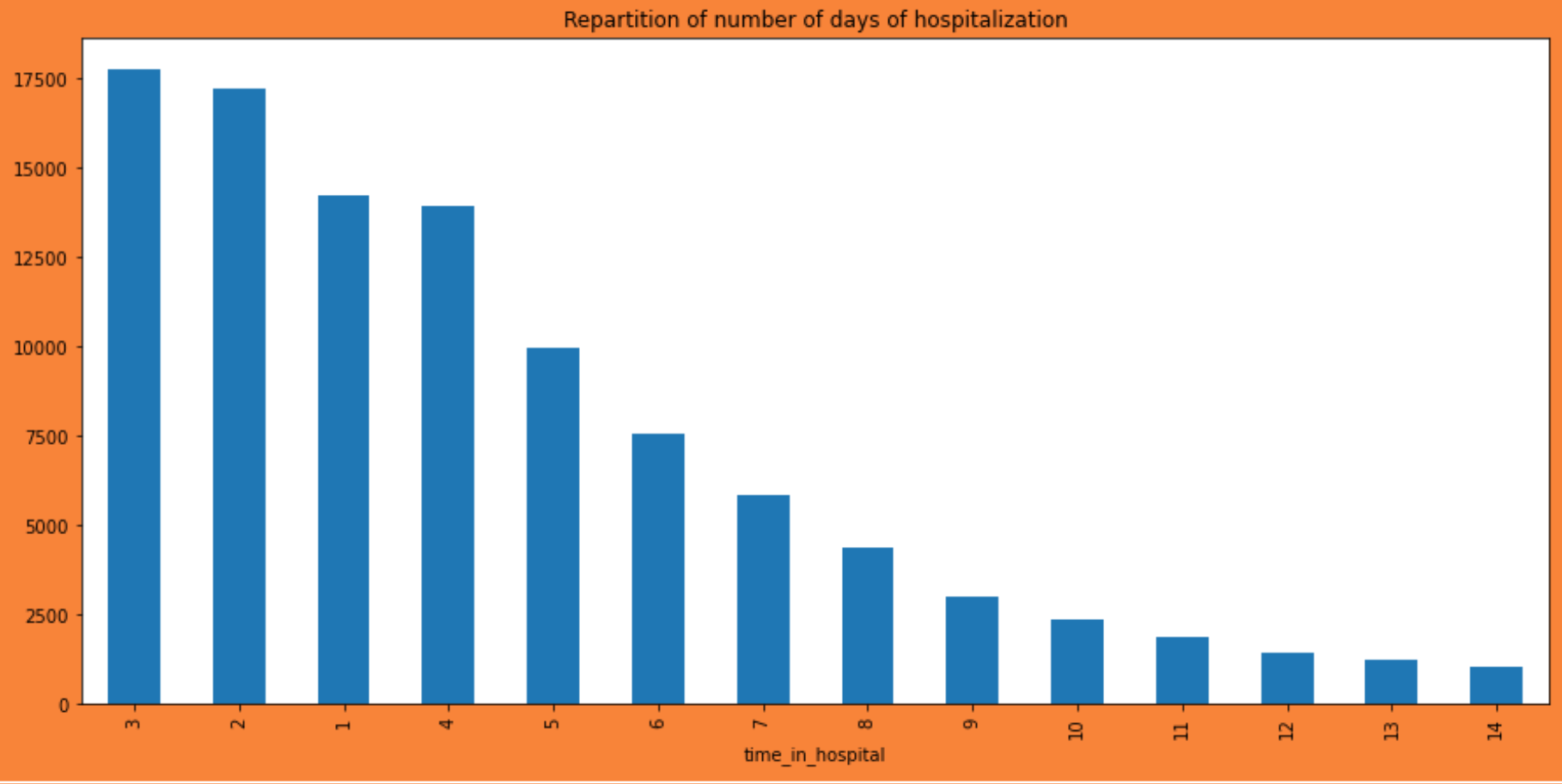
- Inpatient encounter (a hospital admission)
- Diabetic encounter (any kind of diabetes)
- The length of stay was at least 1 day and at most 14 days
- Laboratory tests were performed during the encounter
- Medications were administered during the encounter

# First analysis of the data

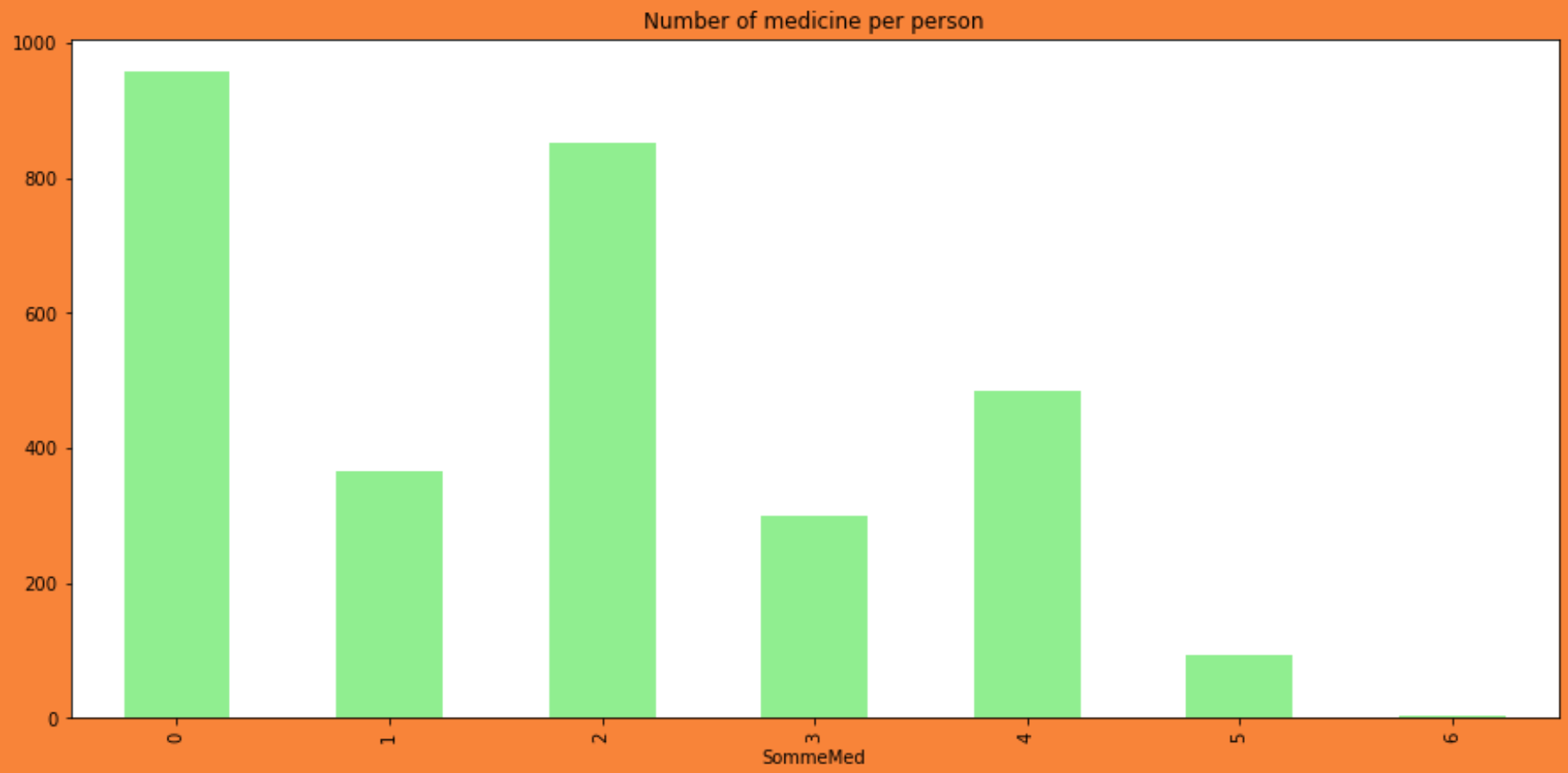


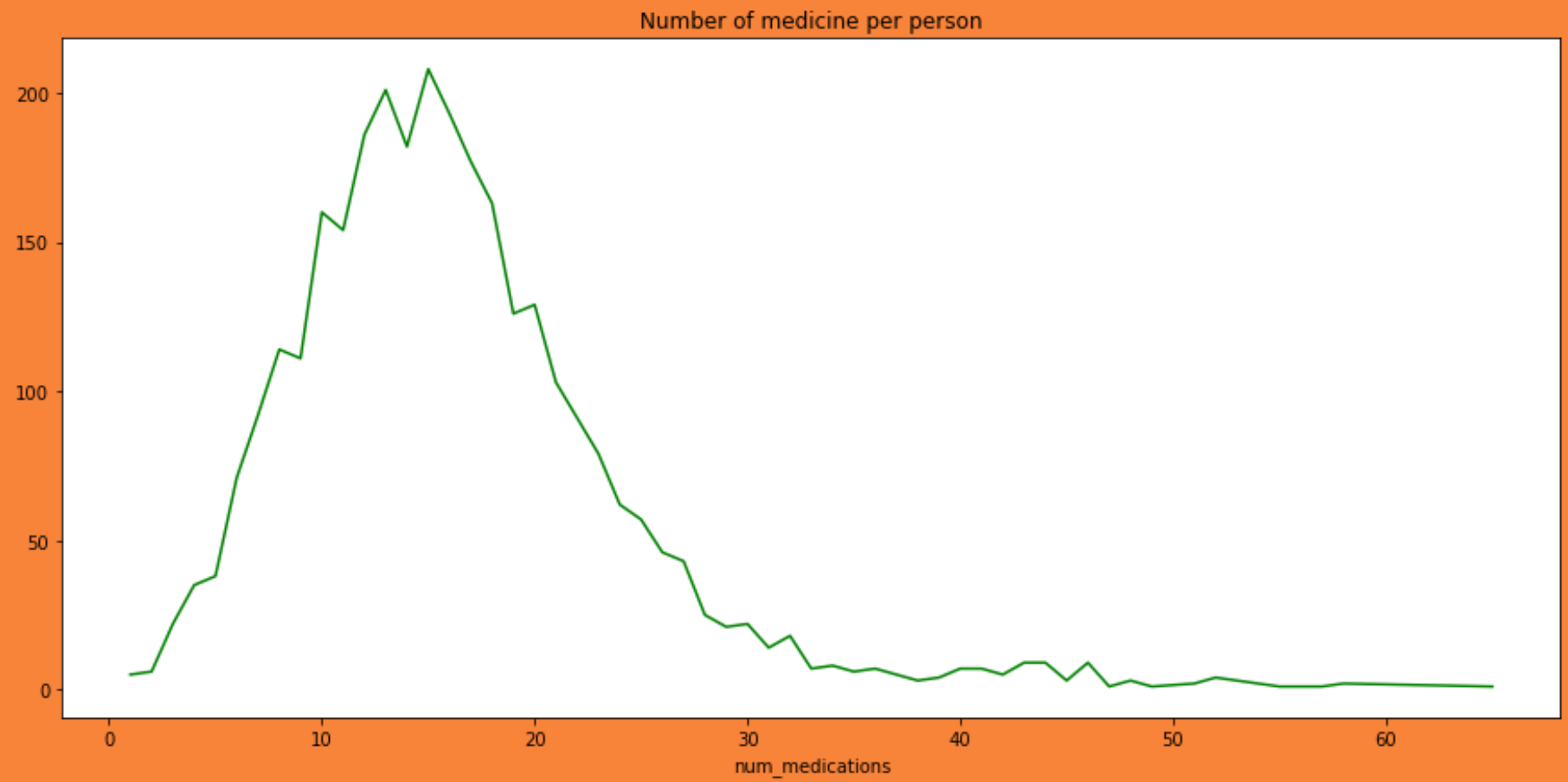
# First analysis of the data



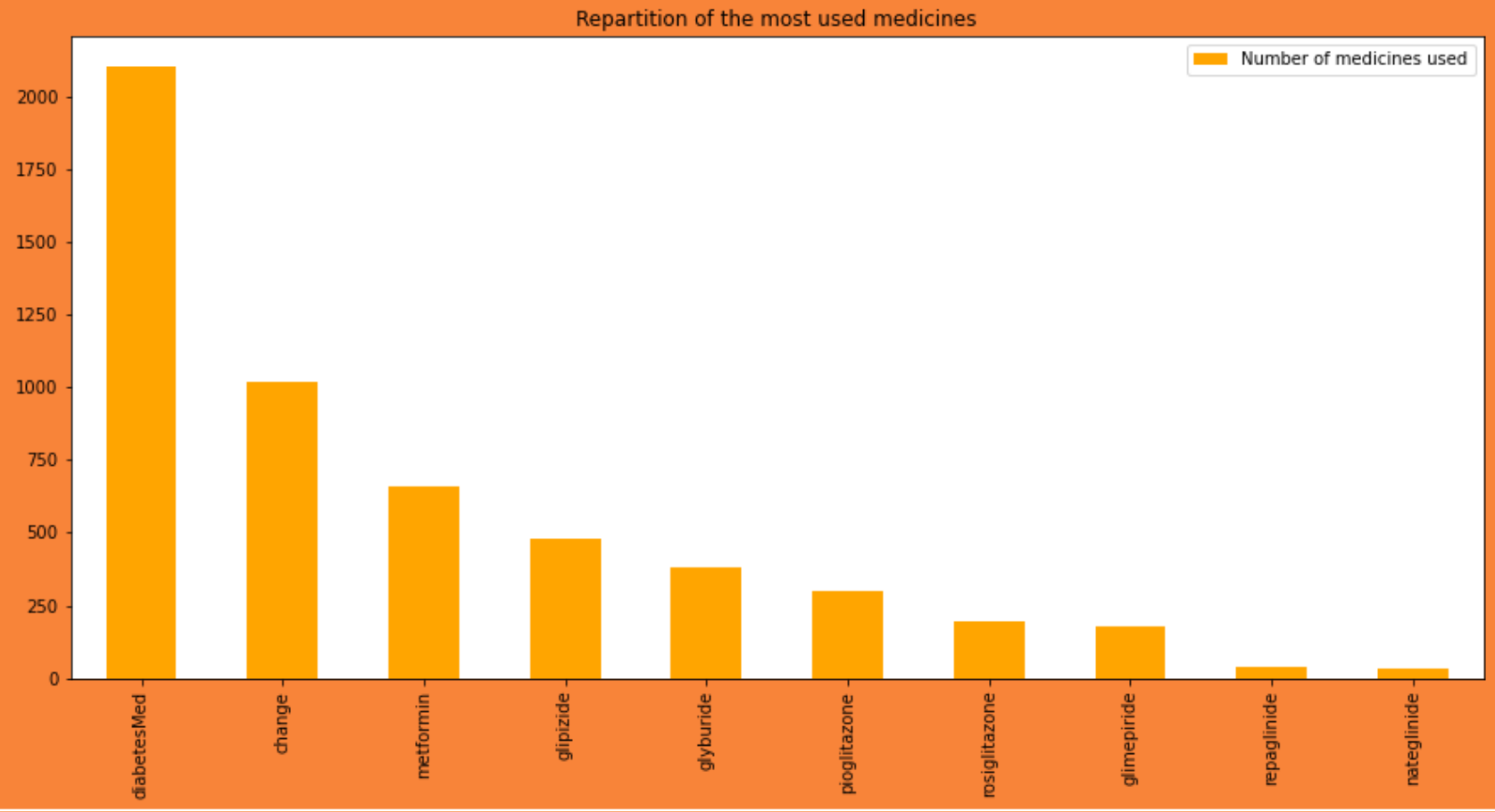
# First analysis of the data



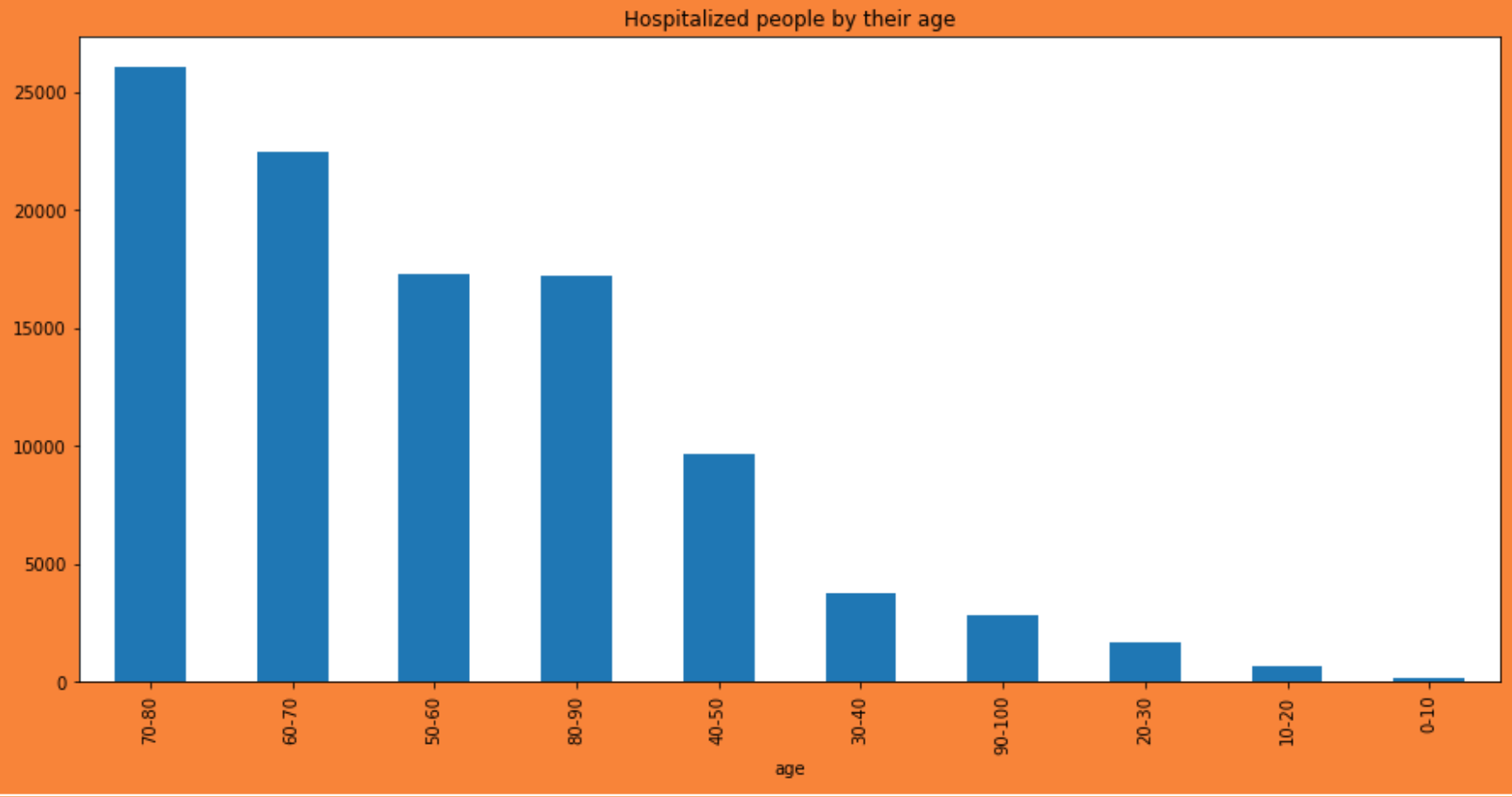
# First analysis of the data



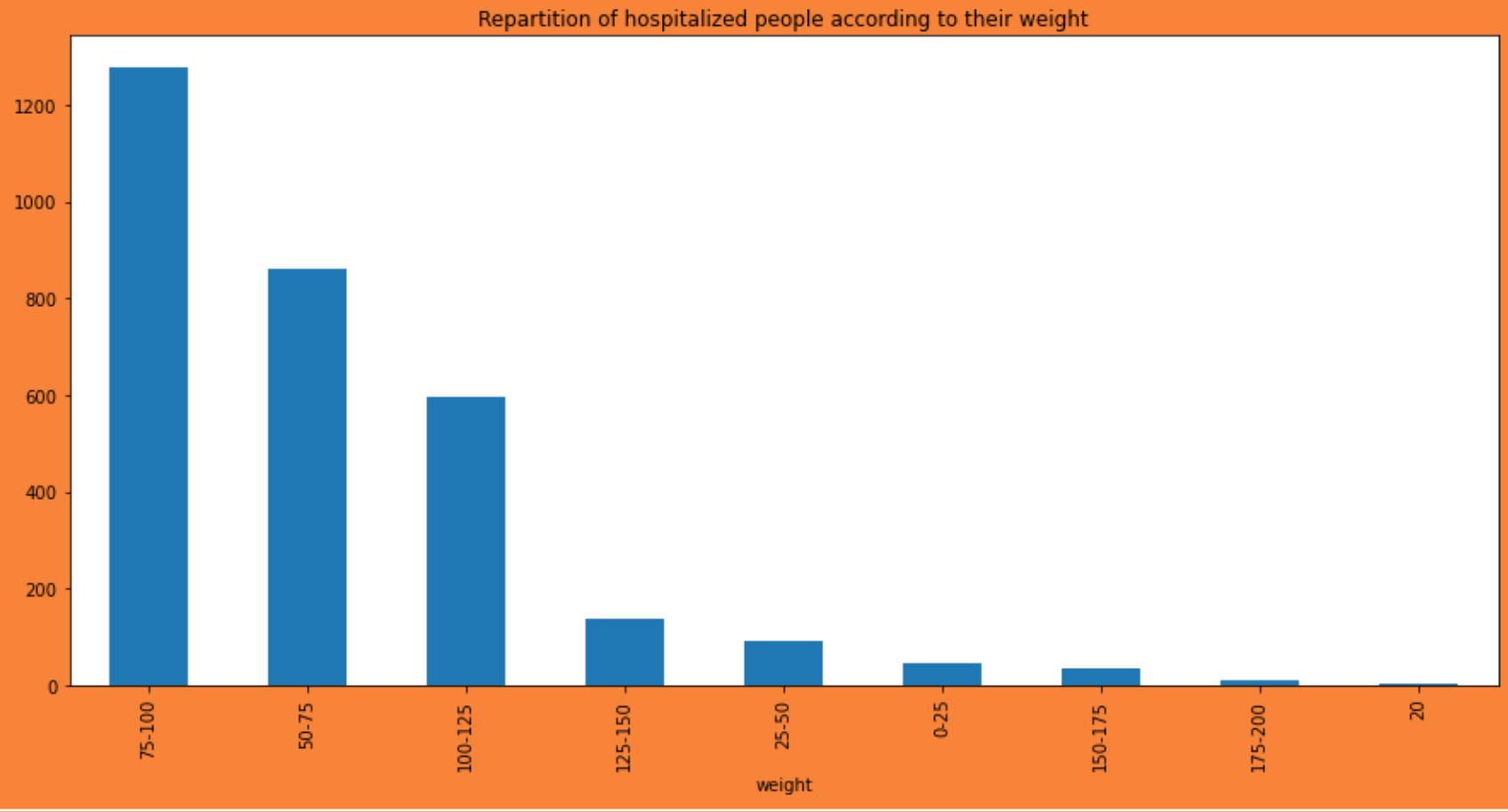
# First analysis of the data



# First analysis of the data

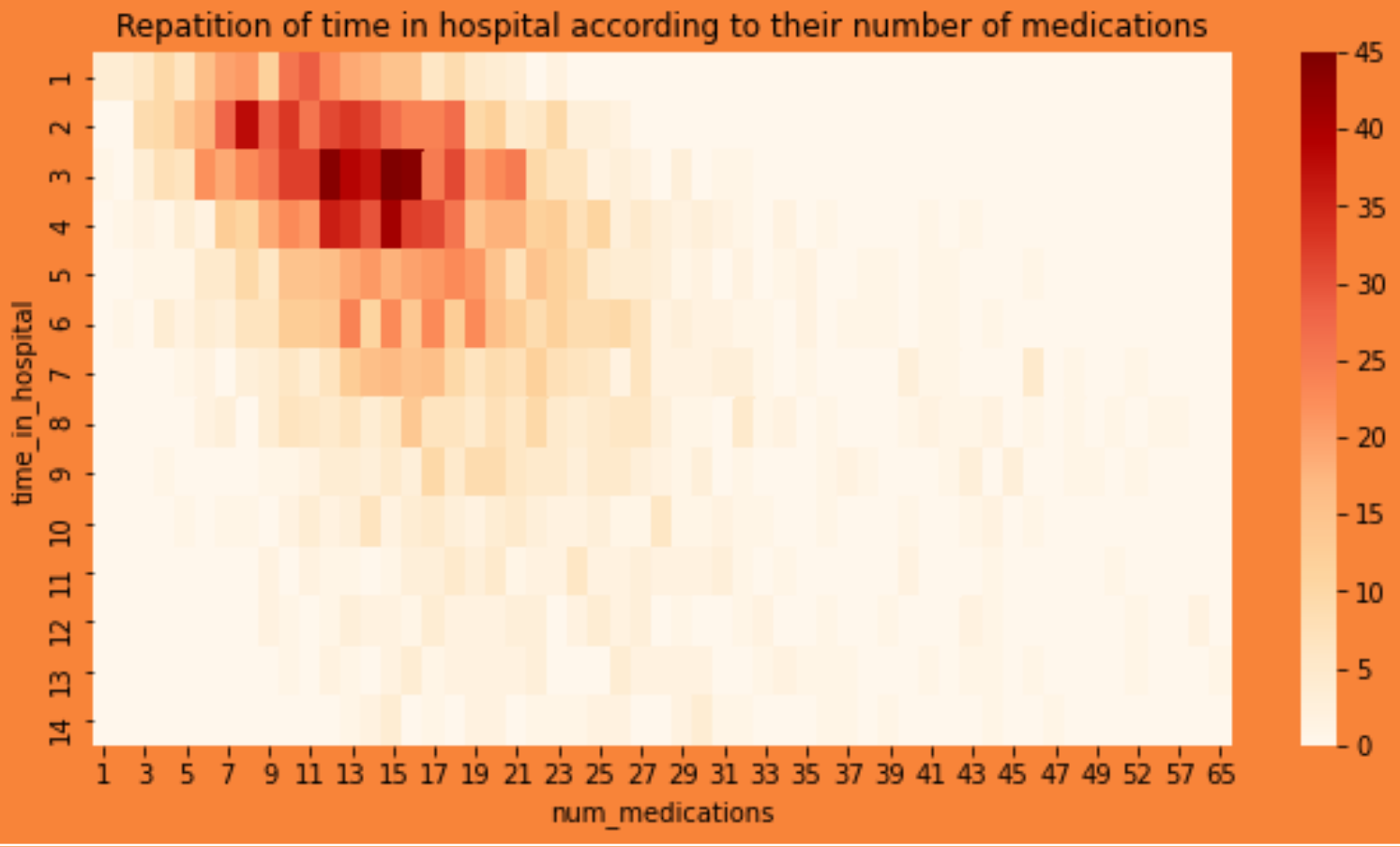


# First analysis of the data

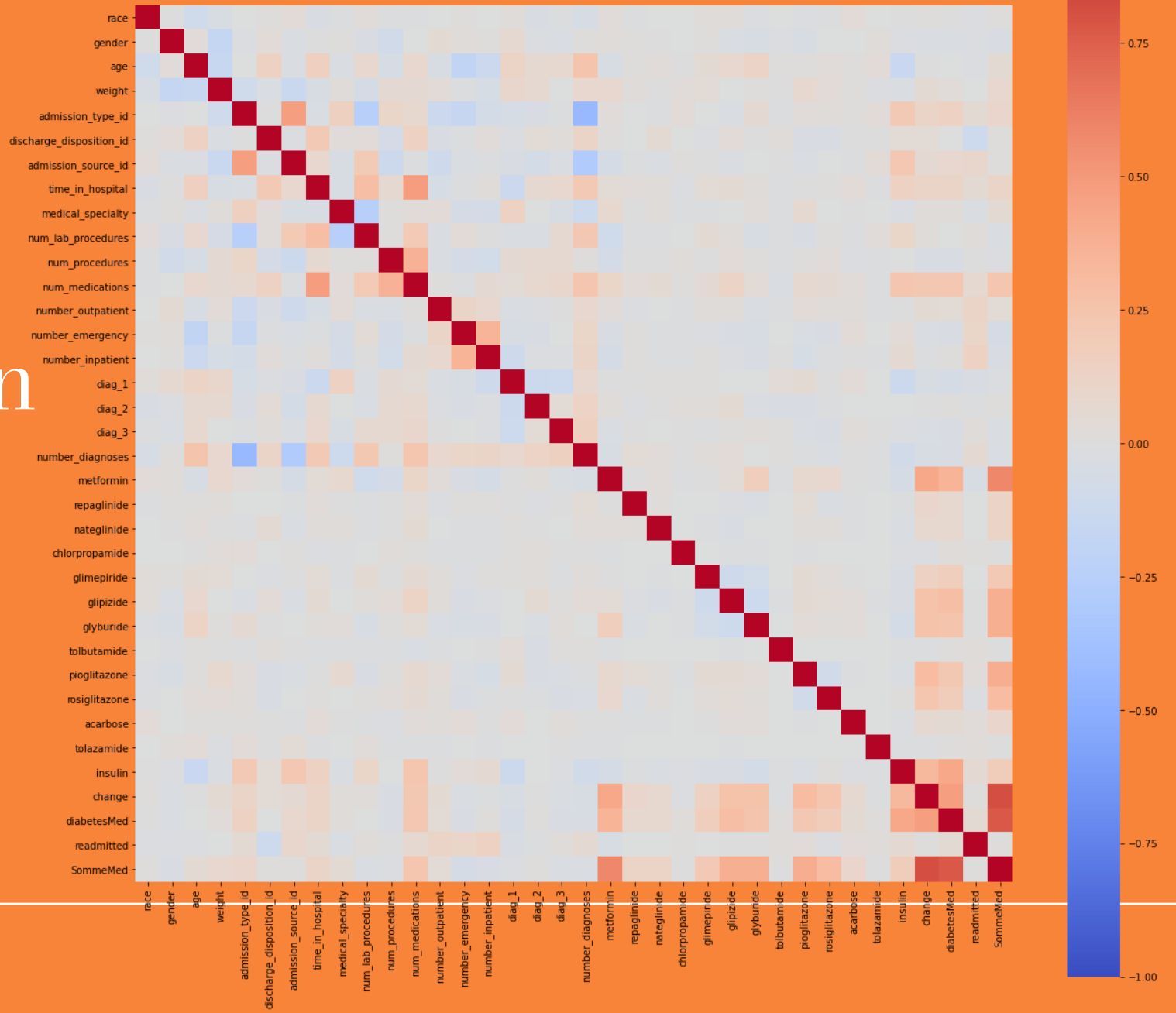


# Deeper analysis on specific criterias

# Deeper analysis on specific criterias

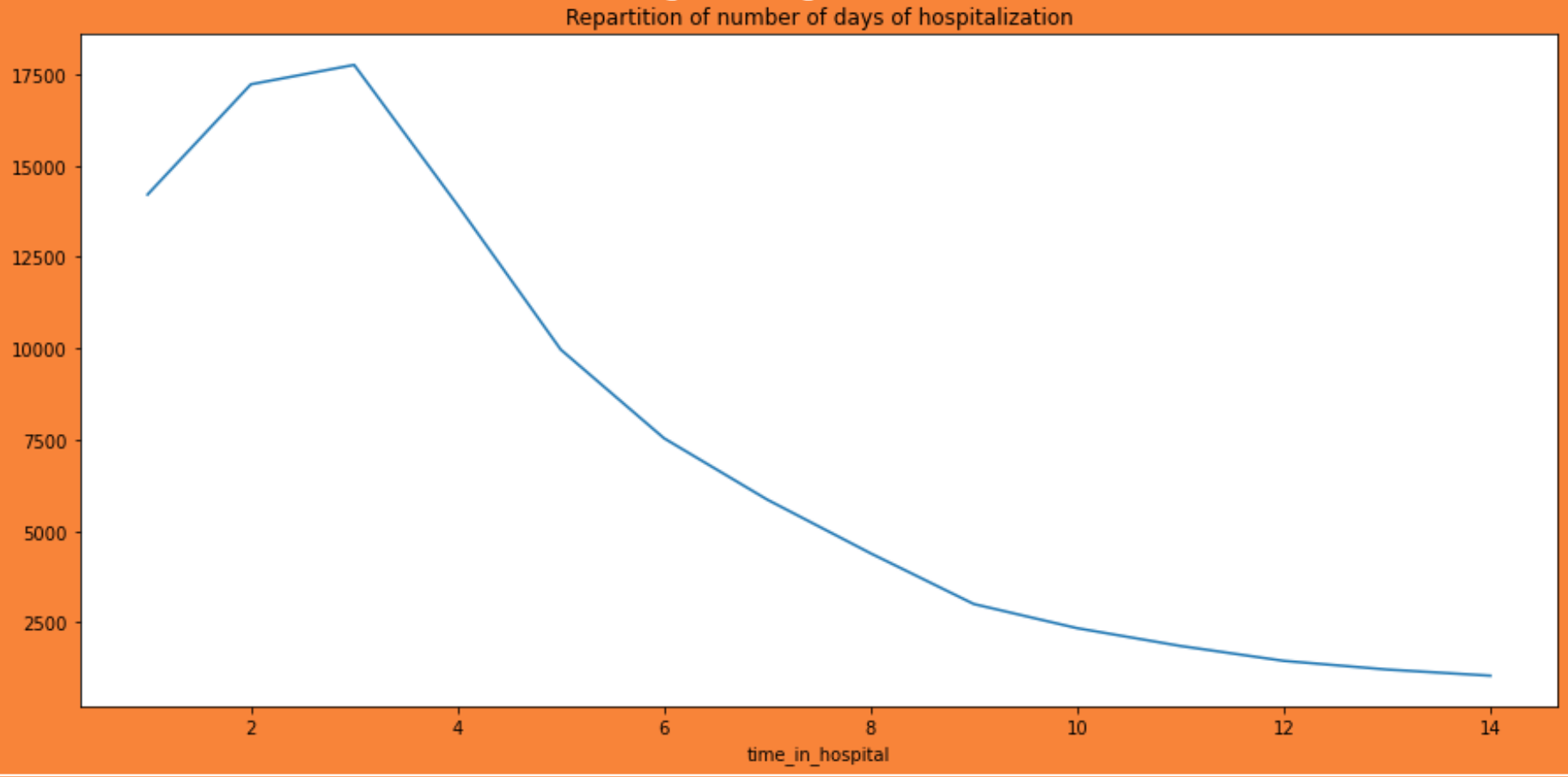


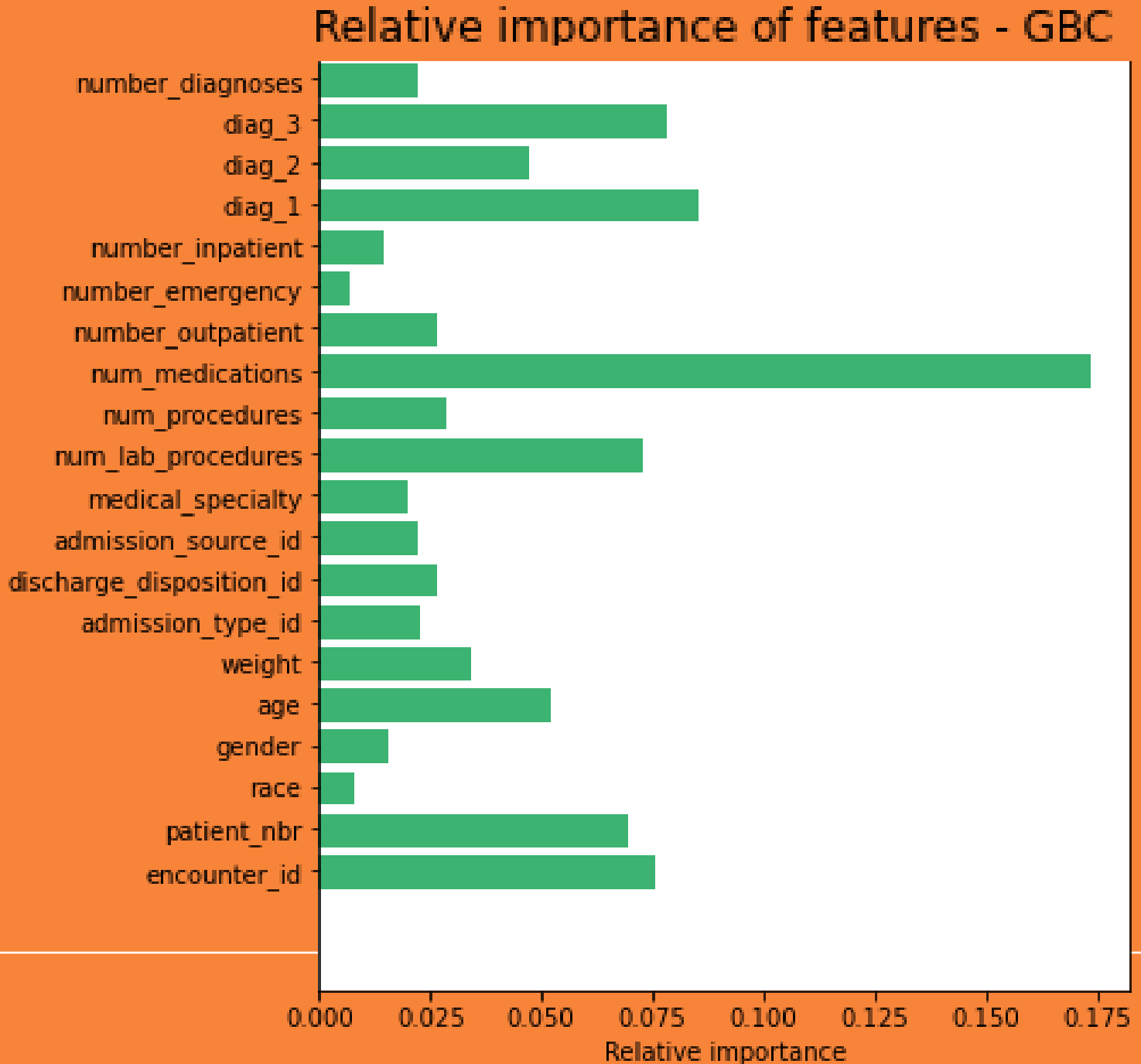
# Correlation Matrix



# Machine Learning

## Which variable are we going to study ?




Relative importance of features - GBC
number_diagnoses
diag_3
diag_2
diag_1
number_inpatient
number_emergency
number_outpatient
num_medications
num_procedures
num_lab_procedures
medical_specialty
admission_source_id
discharge_disposition_id
admission_type_id
weight
age
gender
race
patient_nbr
encounter_id
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
Relative importance

# Machine Learning

## Best models

| Model | Accuracy Score |
|---|---|
| Support Vector Machine | 0.679739 |
| Naive Bayes | 0.679739 |
| Logistic Regression | 0.679739 |
| K-Neighbours | 0.573856 |
| Ridge | 0.186569 |

# Machine Learning

## Best model

| Model | Best Score |
|---|---|
| Support Vector Machine | 0.745415 |

## With

```
SVC(gamma=0.01)
```

# Conclusion

- Important correlation between number of meds and time of hospital
- Optimisation of the rooms distribution
- Better stay for the patient
- Prioritazing some kind of patients that are more fragile
- Other variables could be useful : number of meds, number of lab procedures or the type of medicines used




# Thank you for your attention !