

# Анализа, визуелизација и модели за предвидување на хотелски резервации

Андреј Станојковиќ 186039

# Податочно множество

36 275 онлајн хотелски резервации

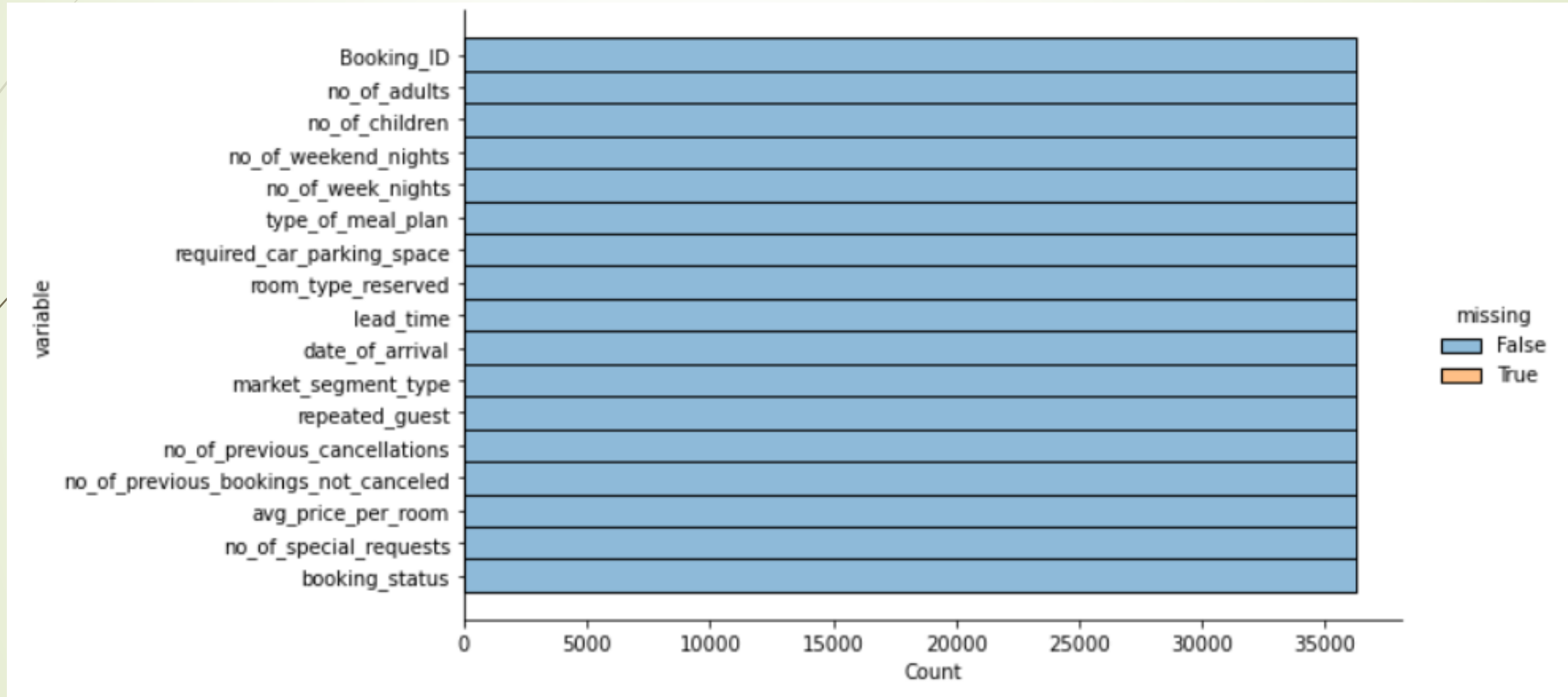
## ► 7 категориски атрибути

- Booking ID
- Type of meal plan
- Required car parking space
- Room type
- Market segment type
- Repeated guest
- Booking status

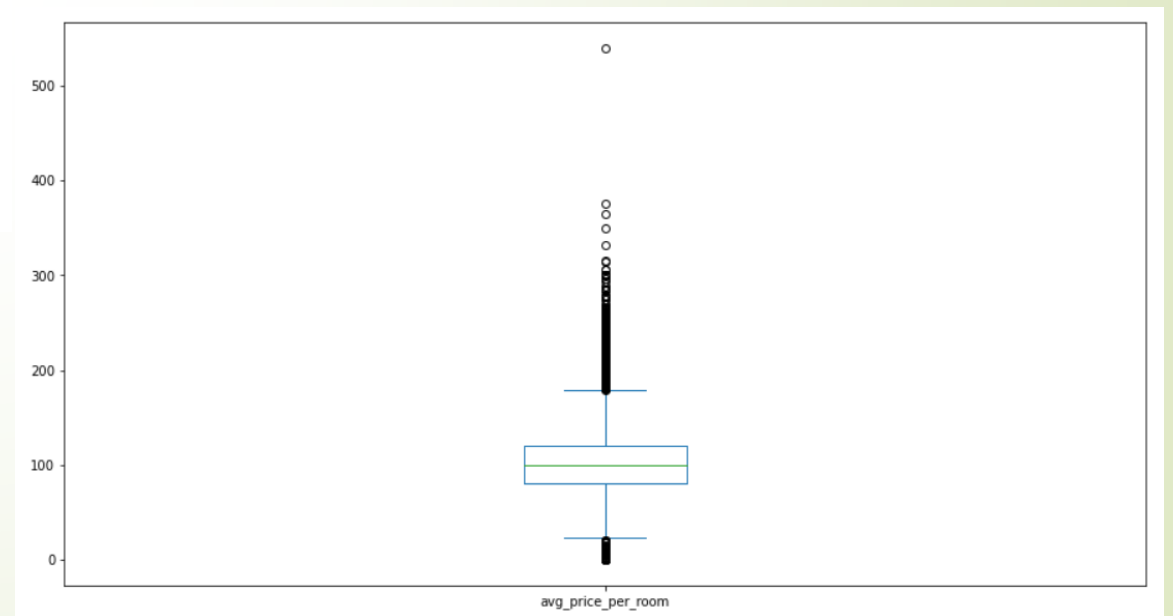
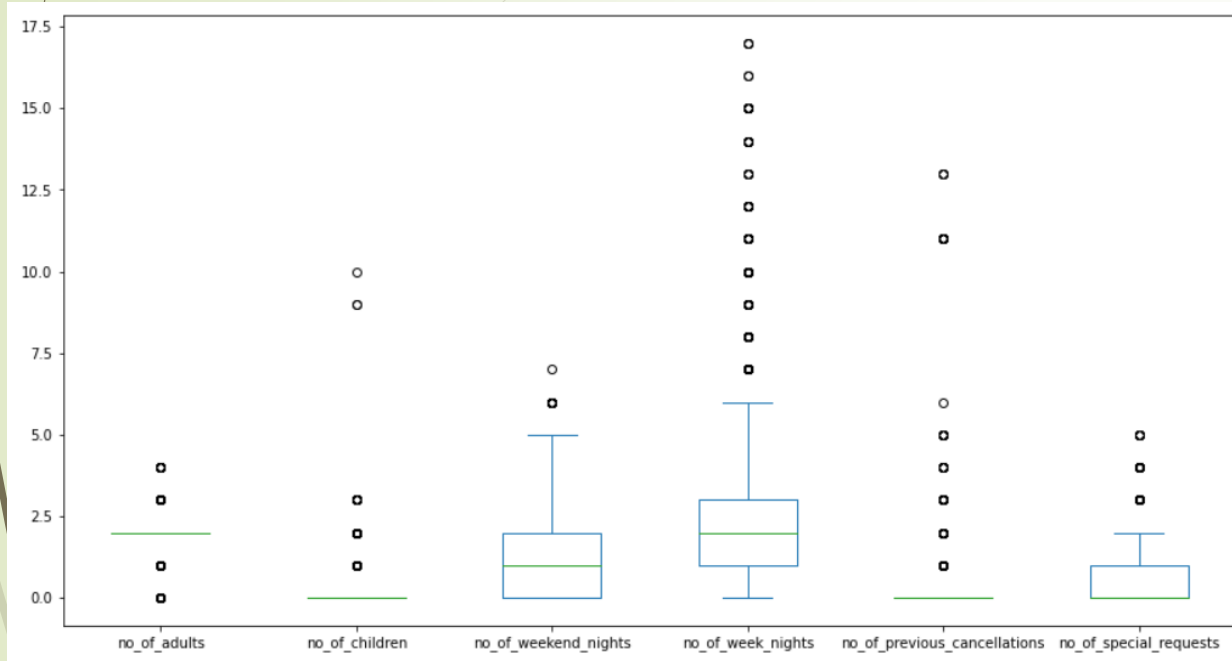
## ► 9 нумерички атрибути

- Number of adults
- Number of children
- Number of weekend nights
- Number of week nights
- Lead time
- Number of previous cancellations
- Number of previous booking not canceled
- Average price per room
- Number of special requests

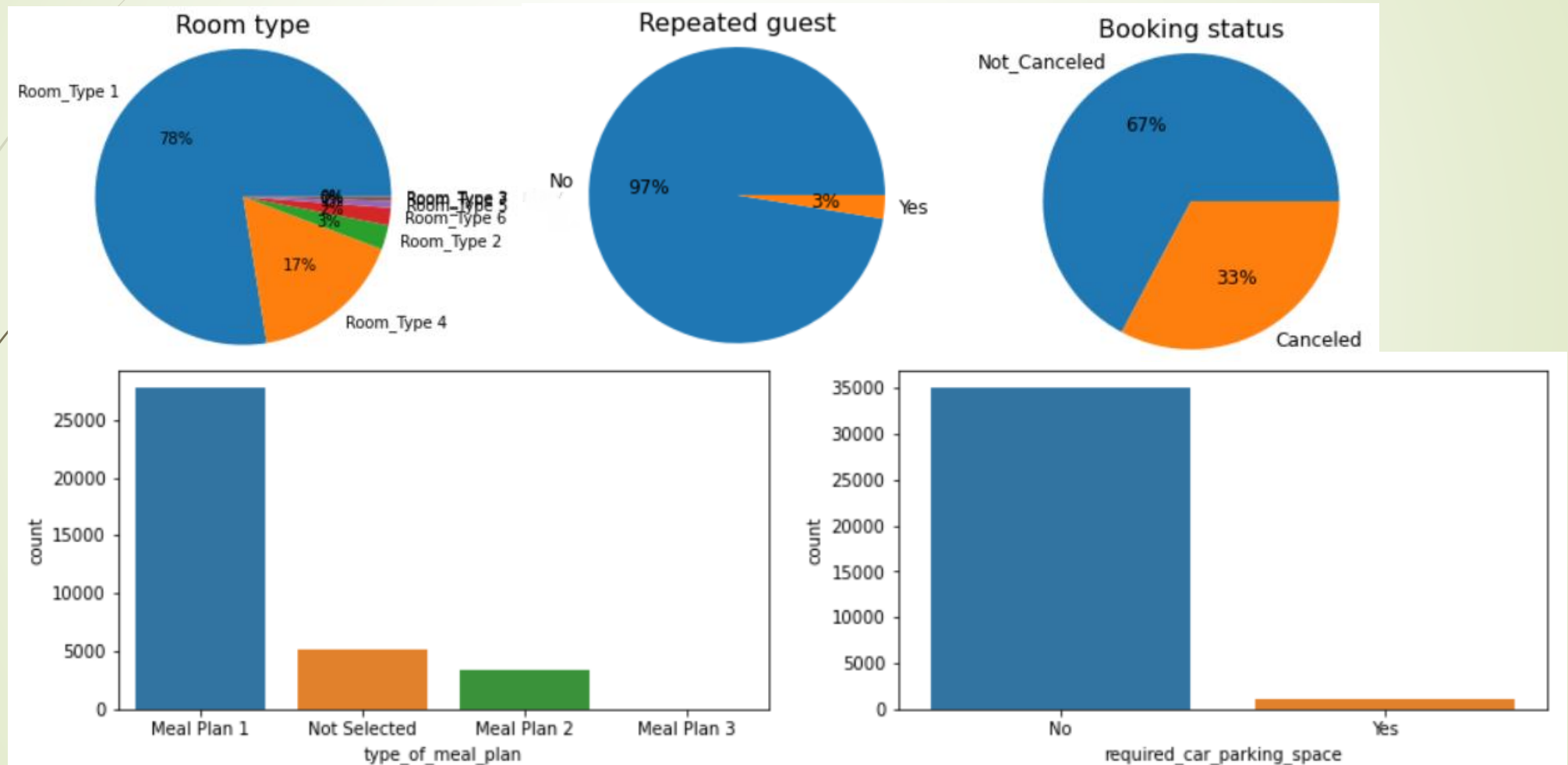
# Missing values

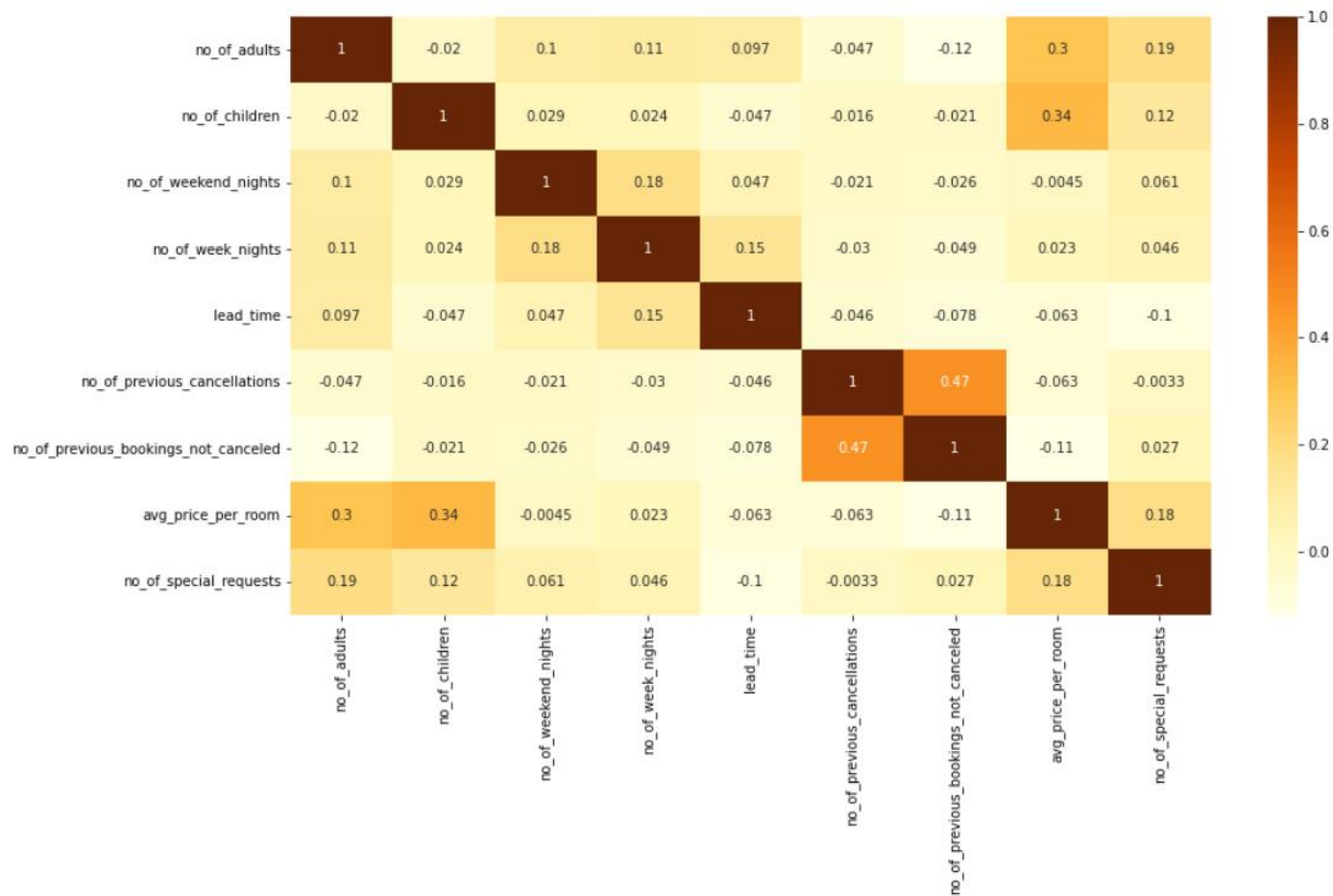
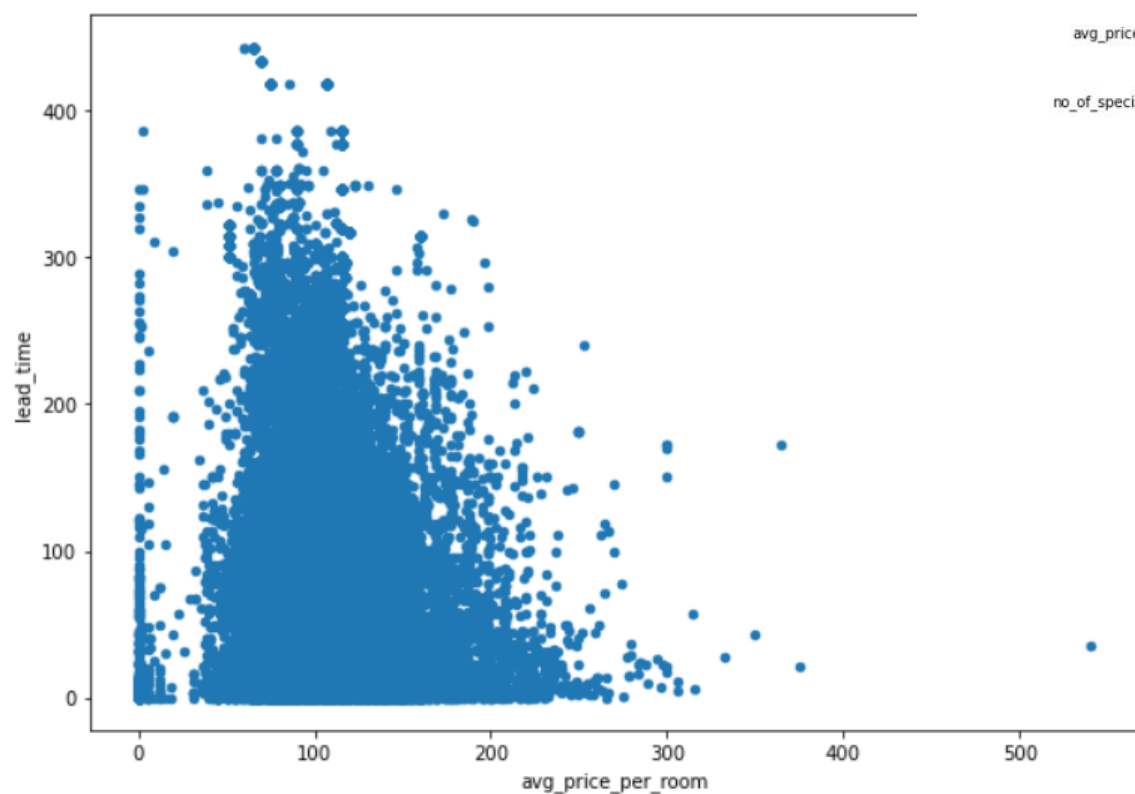


# Outliers



# Визуелизации

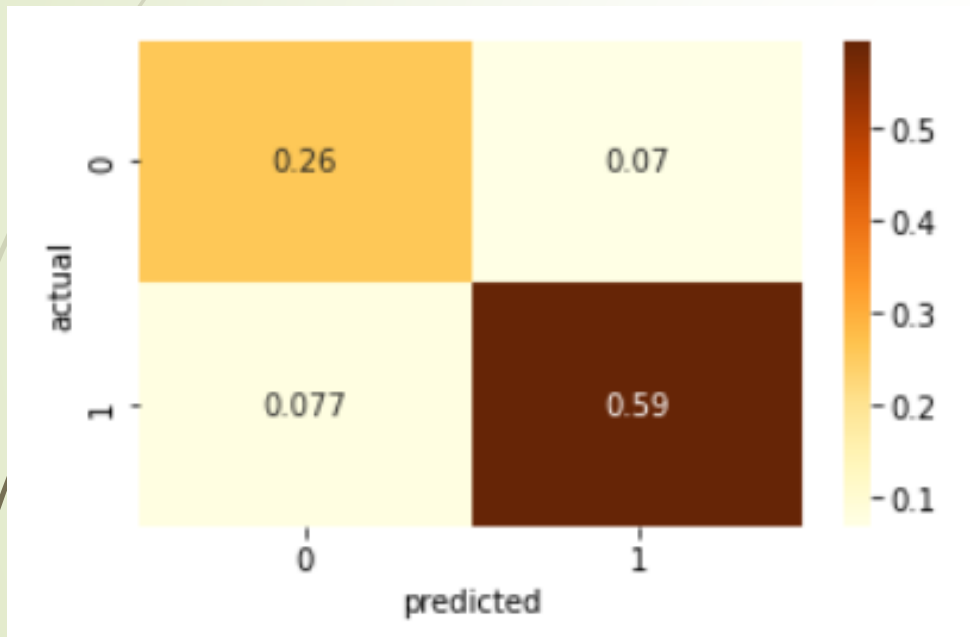




# Класификација

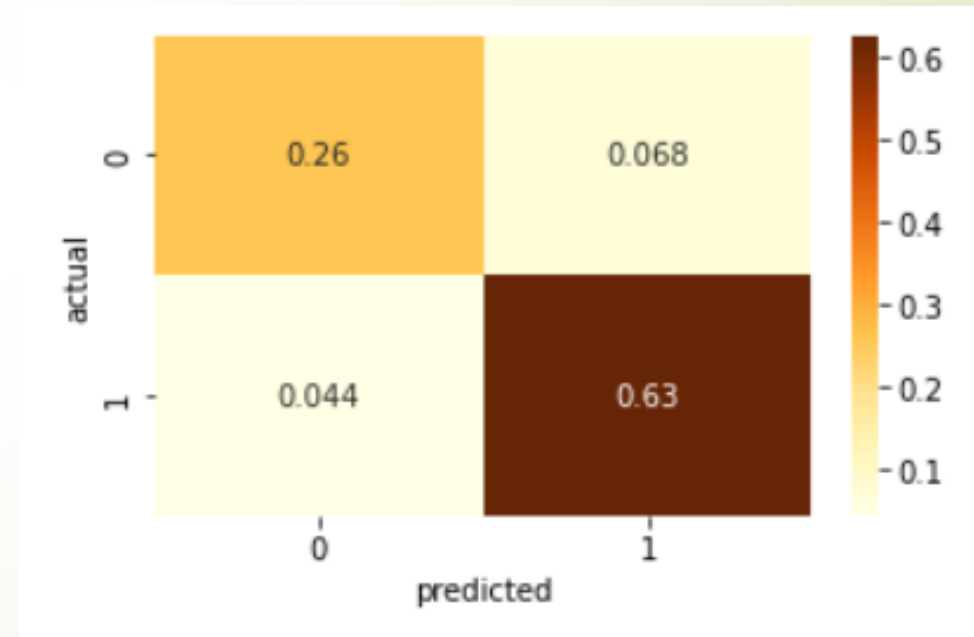
Целта е да се предвиди дали одредена резервација е откажана или не

Decision trees



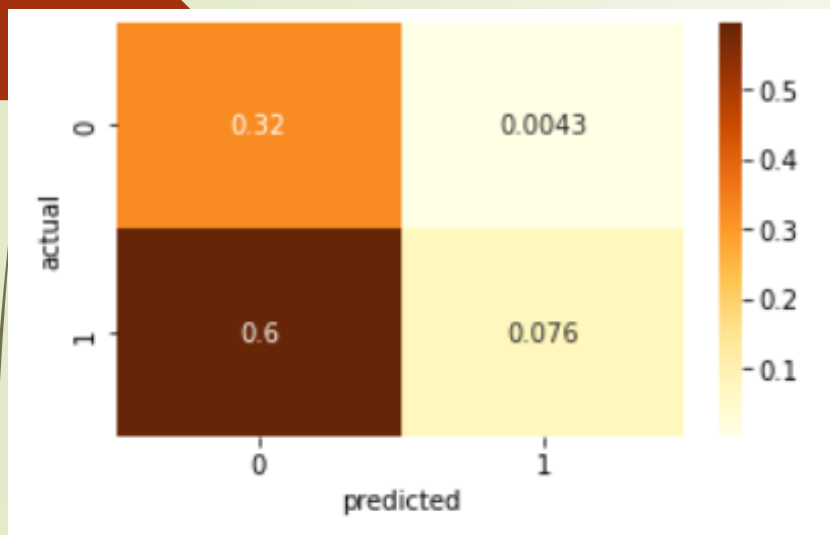
Accuracy: 0.85

Random Forest



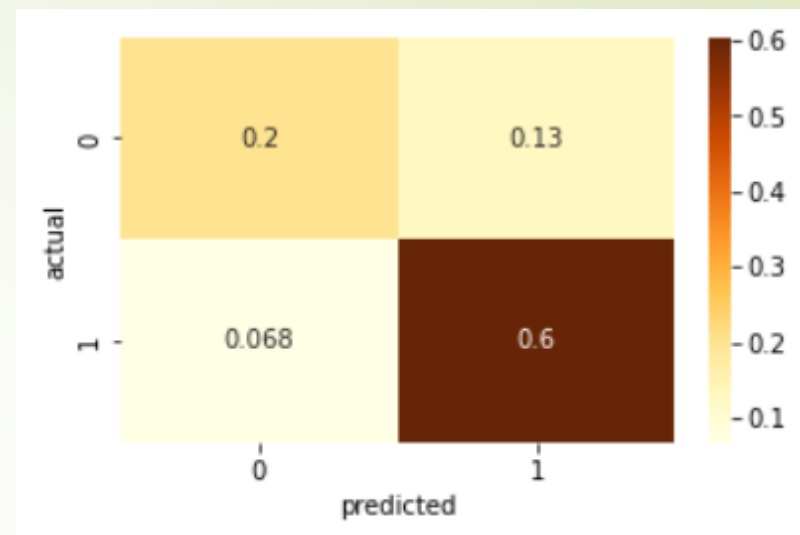
Accuracy: 0.88

Naïve Bayes



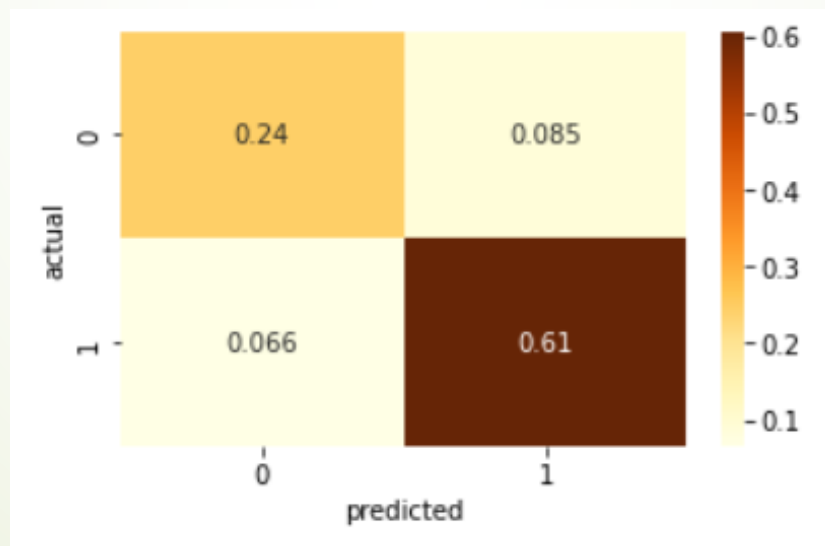
Accuracy: 0.4

Logistic Regression



Accuracy: 0.8

K-Nearest Neighbors

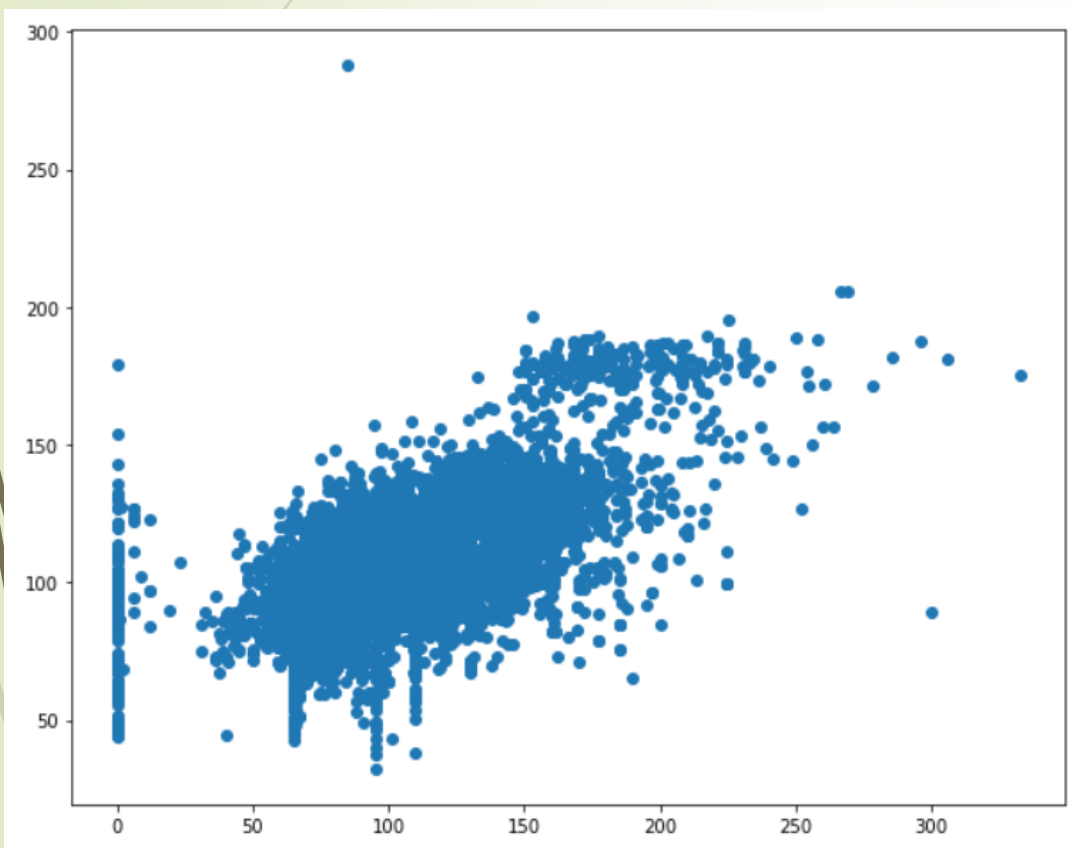


Accuracy: 0.85

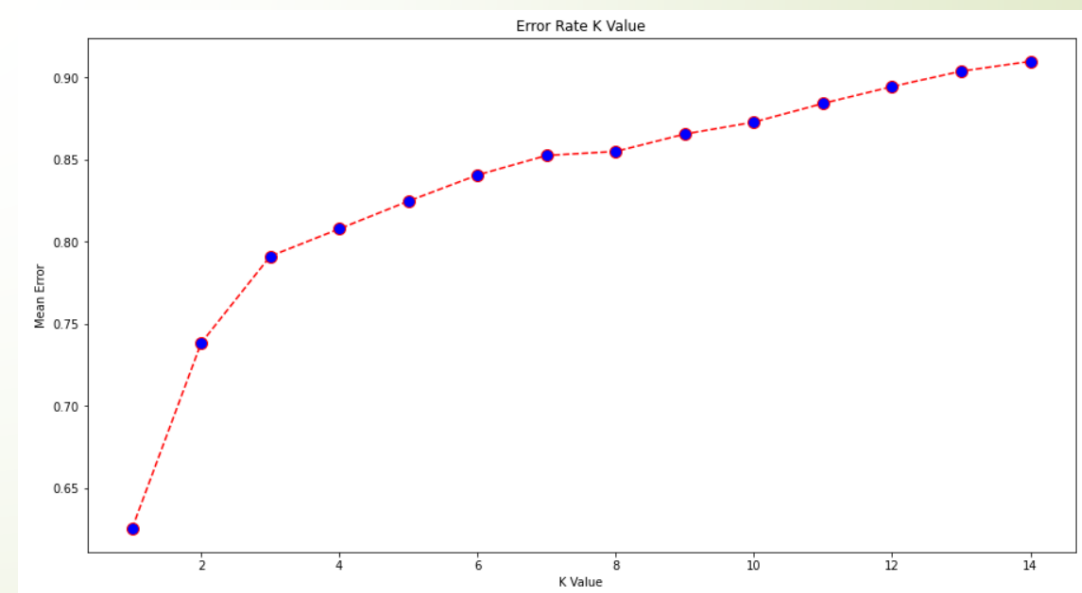


# Линеарна регресија

Целта е да се предвиди просечната цена по соба

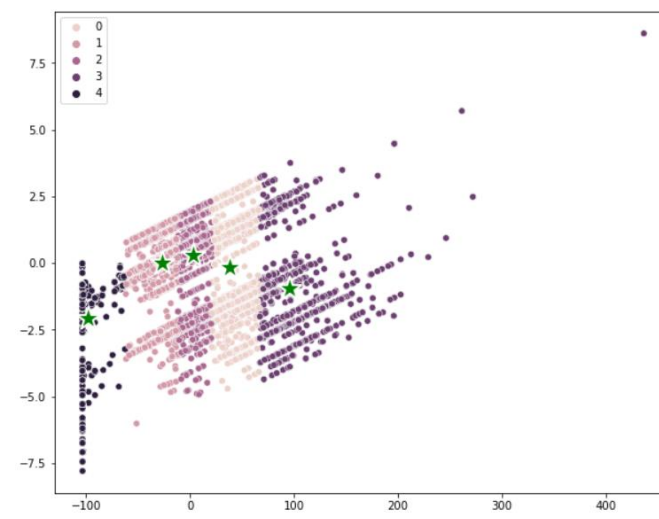
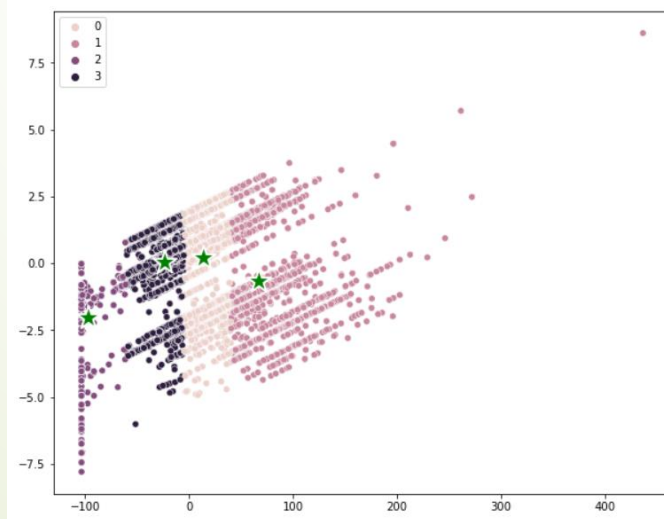
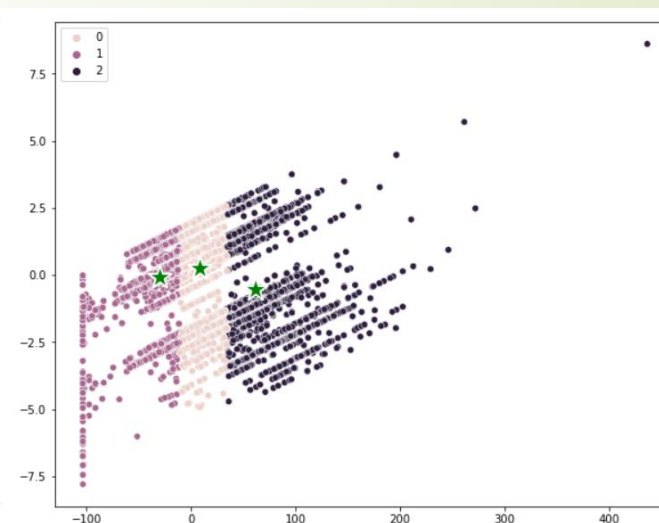
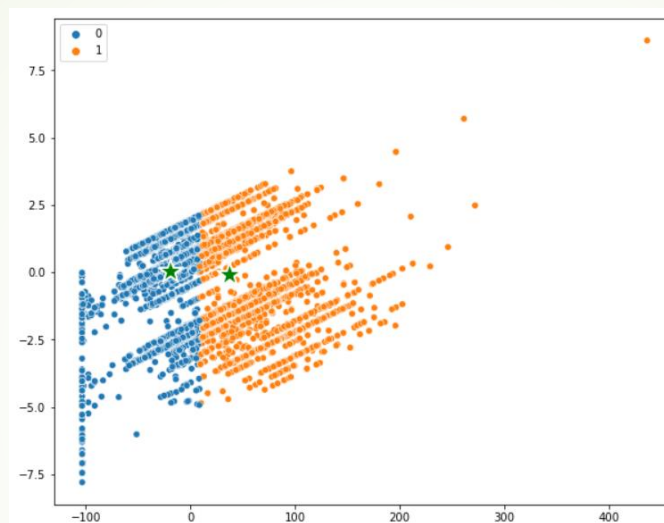


	MAE	MSE	RMSE
Linear	20.53	771.23	27.77
Lasso	20.53	771.23	27.77
Ridge	20.53	771.23	27.77
KNN	14.4	474.78	21.79

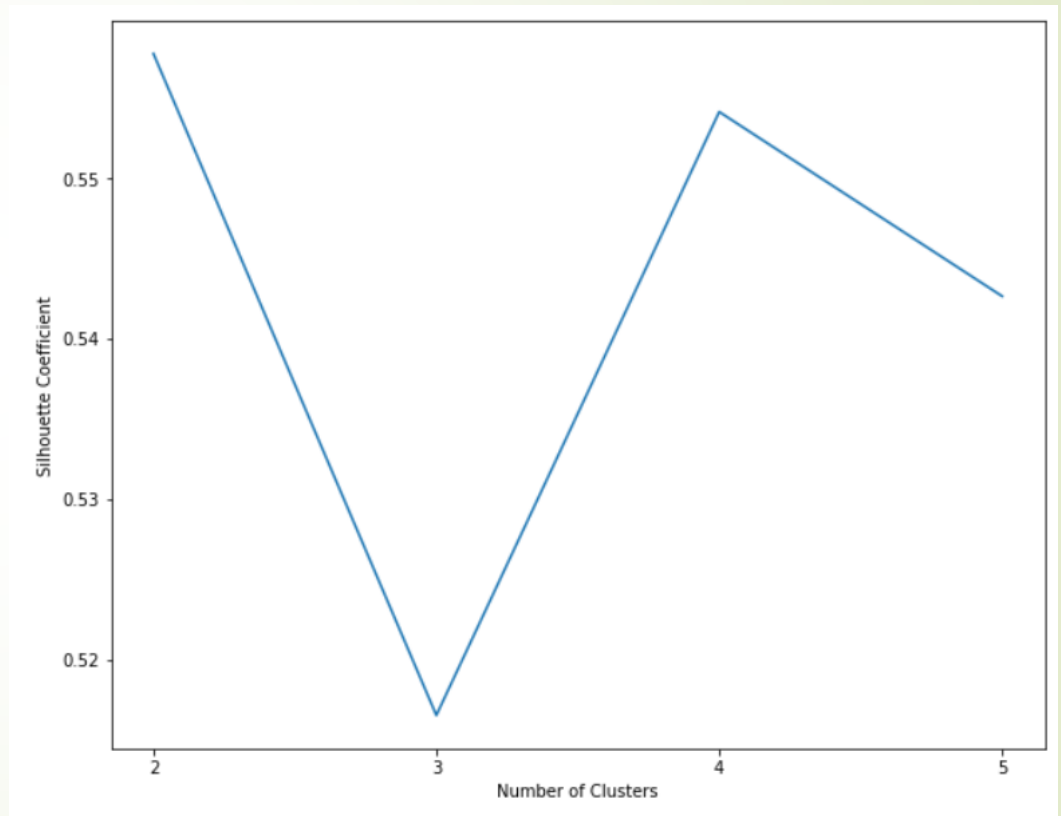
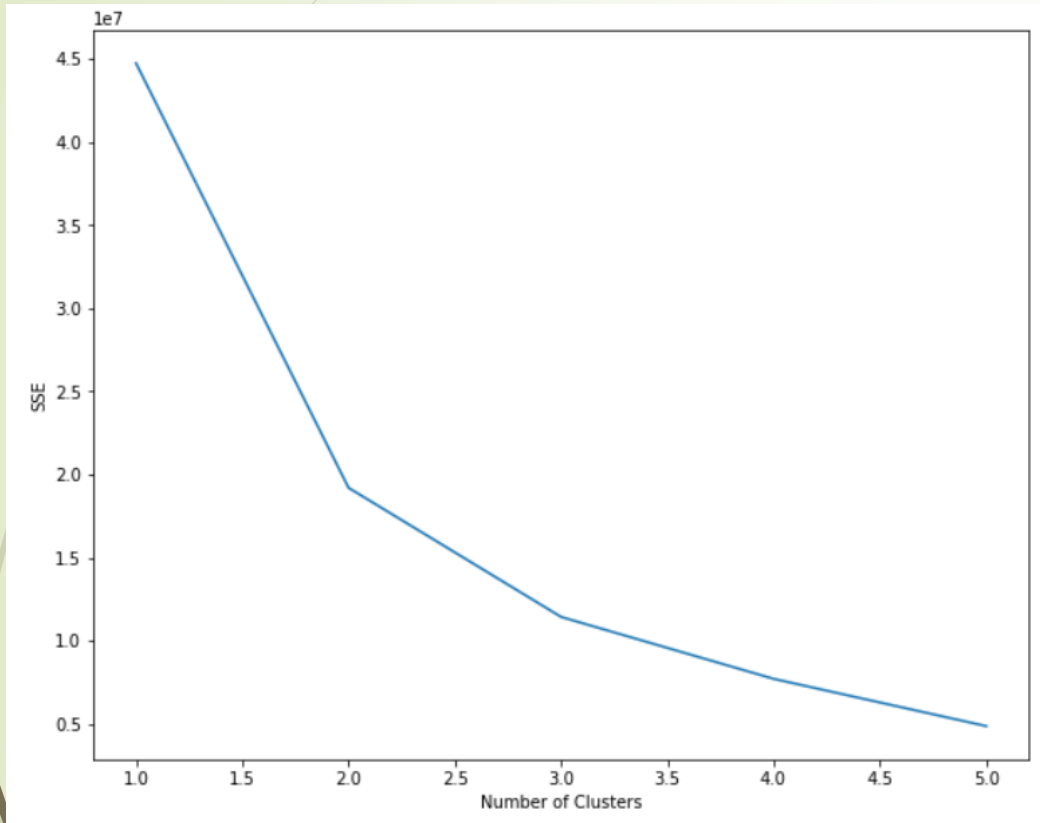


# Кластерирање

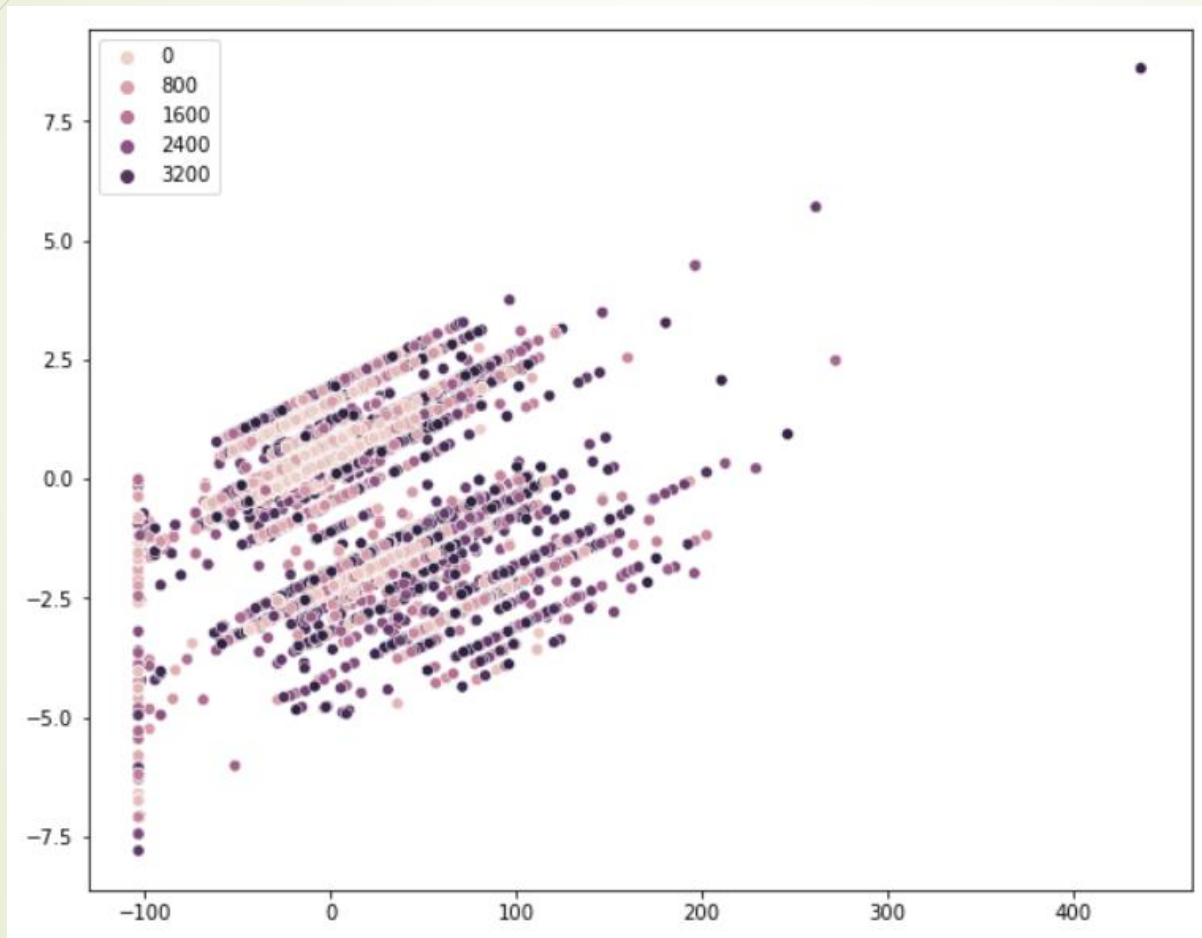
K-Means



# K-Means



# DBSCAN



Silhouette coefficient: 0.64



# Заклучок

- Просечната цена по соба зависи правопрпорционално од бројот на возрасни и деца.
- Најголем дел од гостите (97%) немаат потреба од паркинг место.
- Не постои зависност помеѓу времето од резервација до пристигнување и просечната цена по соба.
- Скоро сите модели за класификација даваат добри резултати, освен моделот на Naïve Bayes.
- Моделите за регресија не даваат добри резултати.
- Со зголемување на бројот на кластери кај K-Means се намалува SSE, додека silhouette coefficient не се менува премногу.
- И K-Means и DBSCAN имаат прилично добри silhouette coefficients.