Bar plot of MAE of the different model

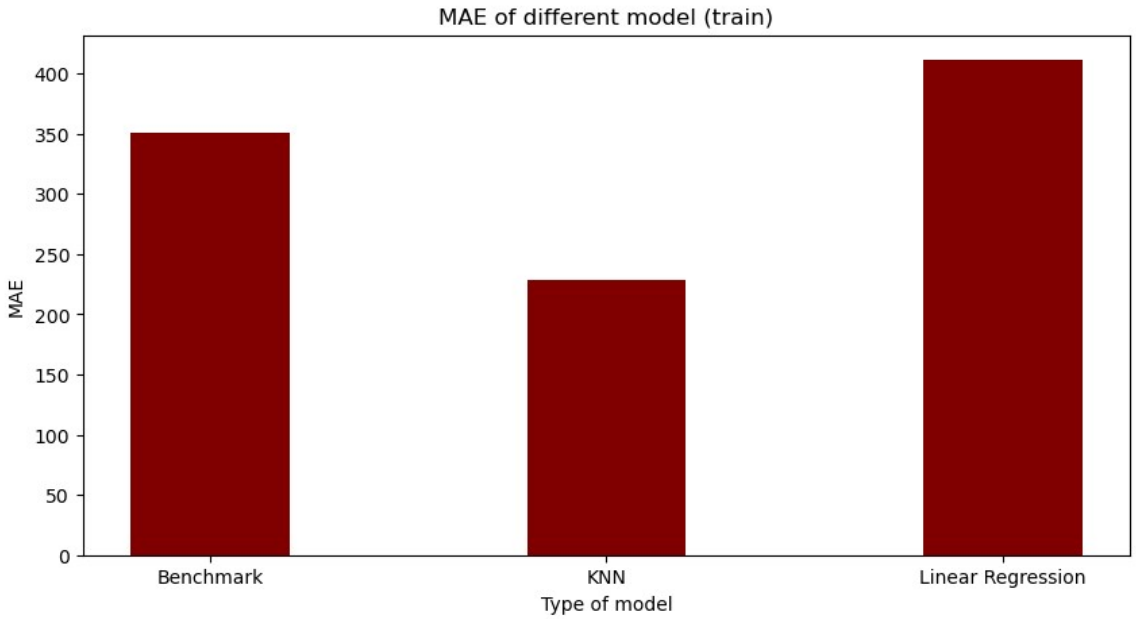
In [10]:
```
import numpy as np
import matplotlib.pyplot as plt


# creating the dataset
data = {'Benchmark':351.26564067758295, 'KNN':228.93482320977935, 'Linear R
courses = list(data.keys())
values = list(data.values())

fig = plt.figure(figsize = (10, 5))

# creating the bar plot
plt.bar(courses, values, color ='maroon',
        width = 0.4)

plt.xlabel("Type of model")
plt.ylabel("MAE")
plt.title("MAE of different model (train)")
plt.show()
```



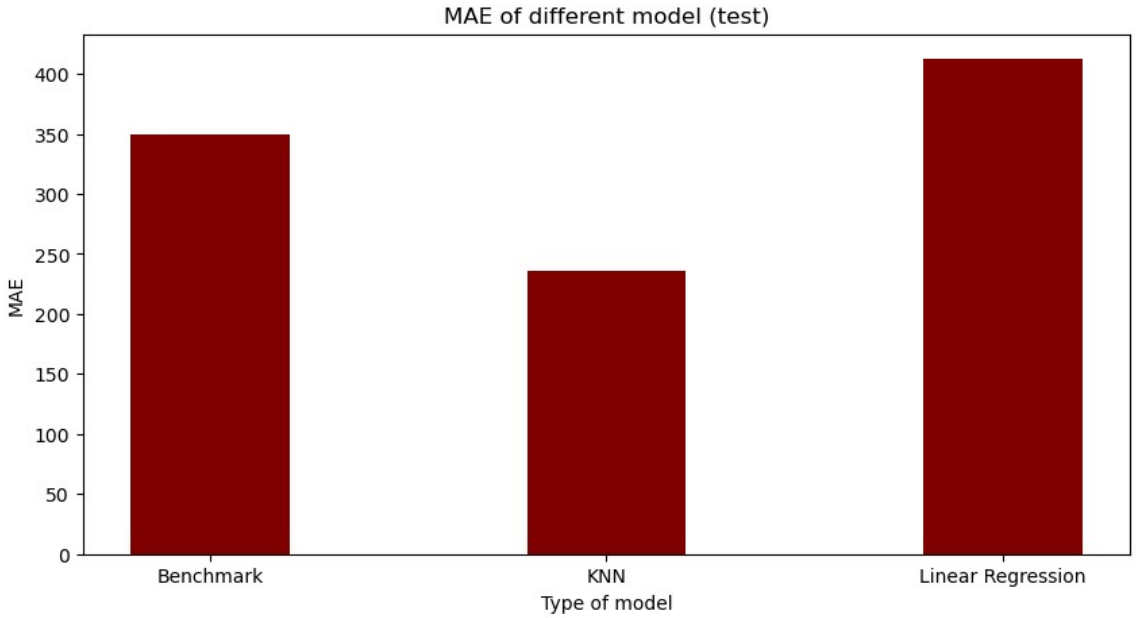
Linear Regression Model has the highest MAE with least MAE of KNN (train).

In [11]:
```
# creating the dataset
data = {'Benchmark':349.949874273212, 'KNN':235.4292275718515, 'Linear Regr
courses = list(data.keys())
values = list(data.values())

fig = plt.figure(figsize = (10, 5))

# creating the bar plot
plt.bar(courses, values, color ='maroon',
        width = 0.4)

plt.xlabel("Type of model")
plt.ylabel("MAE")
plt.title("MAE of different model (test)")
plt.show()
```



Linear Regression Model has the highest MAE with least MAE of KNN (test).

Evaluation metric: Since our target variable is continous (trip_duration in seconds) we have used MAE, MSE for KNN; MAE for Benchmark; MAE for Linear Regression.

Considering analysis from EDA provided: 1>Some outiliers are removed. trip_duration>6000 are removed(Considering Box plot for passenger count in EDA analysis) 2>'Short distance' is calculated using features- pickup and dropoff- (lattitude and longitude) repectively.

Observations and Conclusions: 1>KNN vs Benchmark vs Linear: Since KNN has the least MAE it is the best model out of these.

In [ ]: