Forecasts and predictions

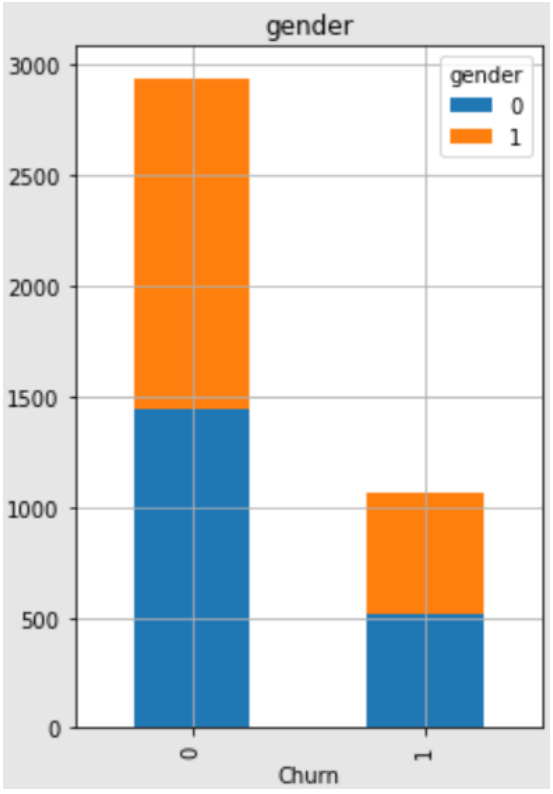
Charikov Andrey

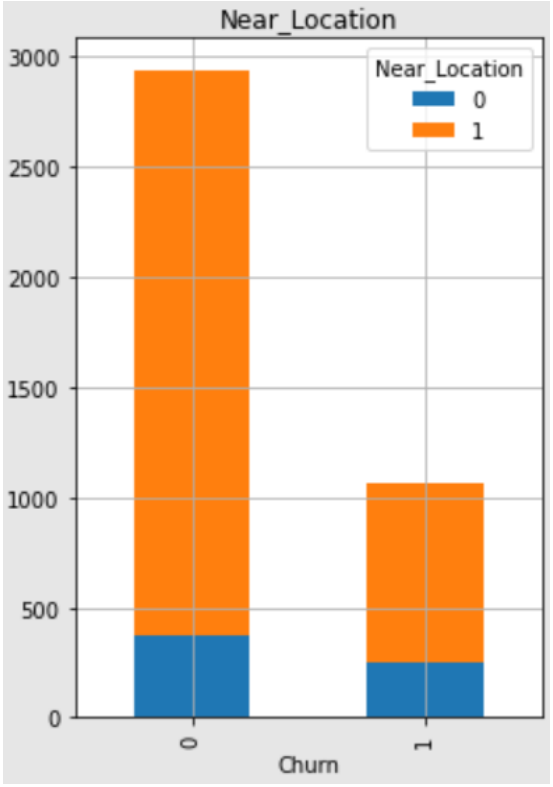
Goal of the project:

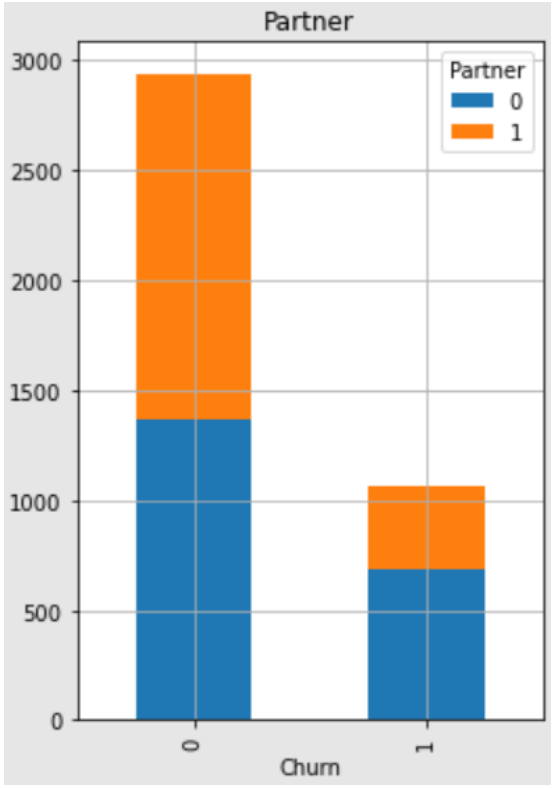
Main purpose of this project is to analyze Gym customers' data in order to find out the main reason for churn.

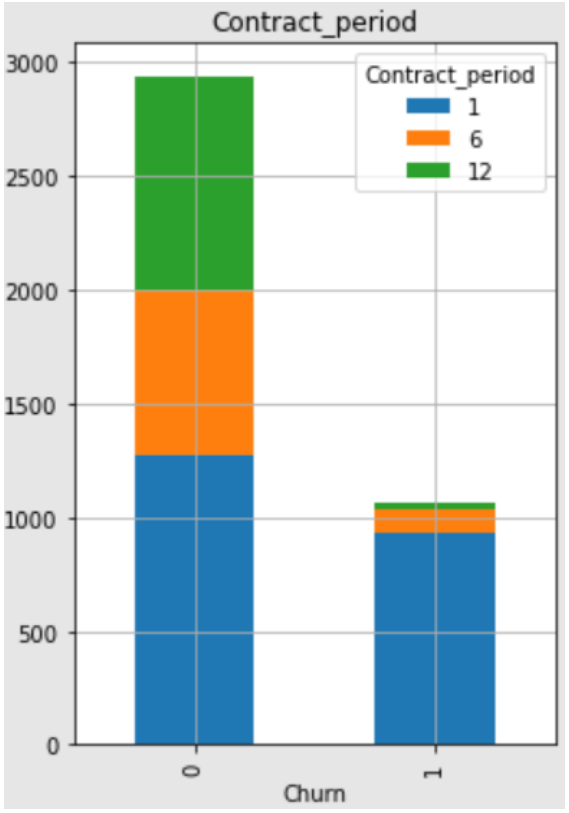
In addition, I need to divide all customers to groups and predict which one is more vulnerable and what can be improved in order to lower churn

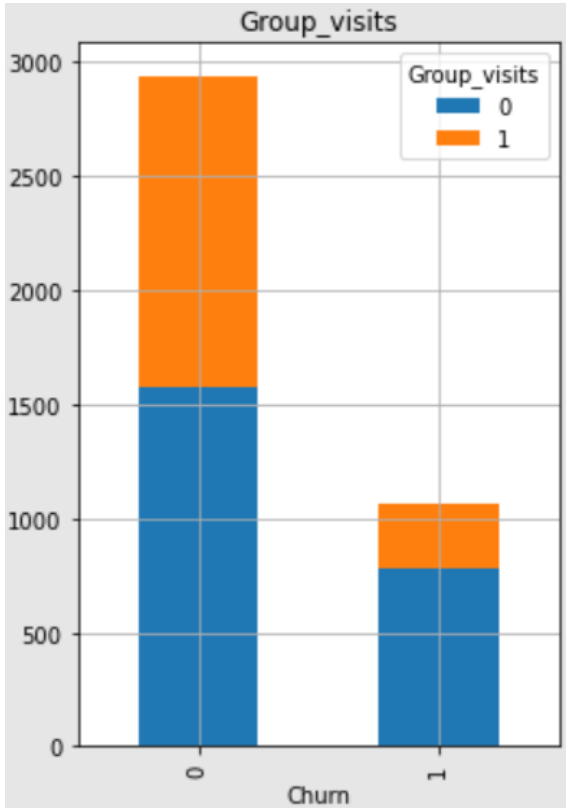
Feature distributions for those who left (churn) and those who stayed.

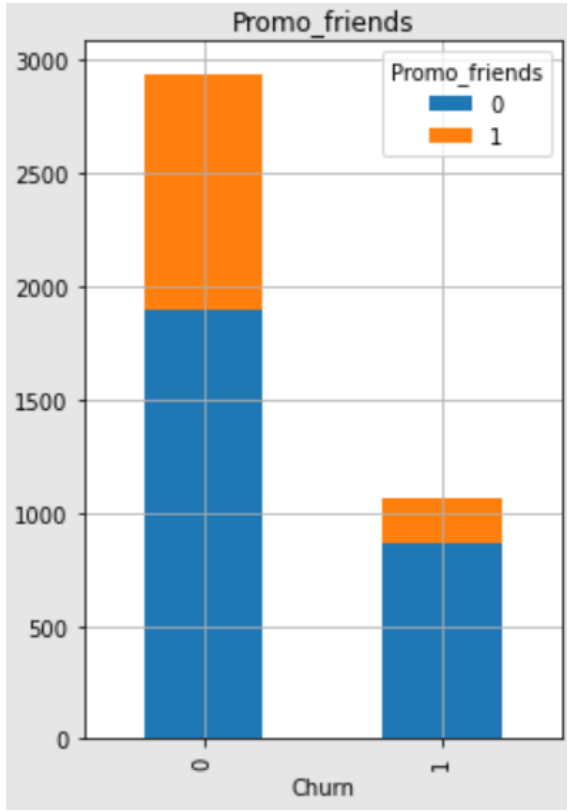




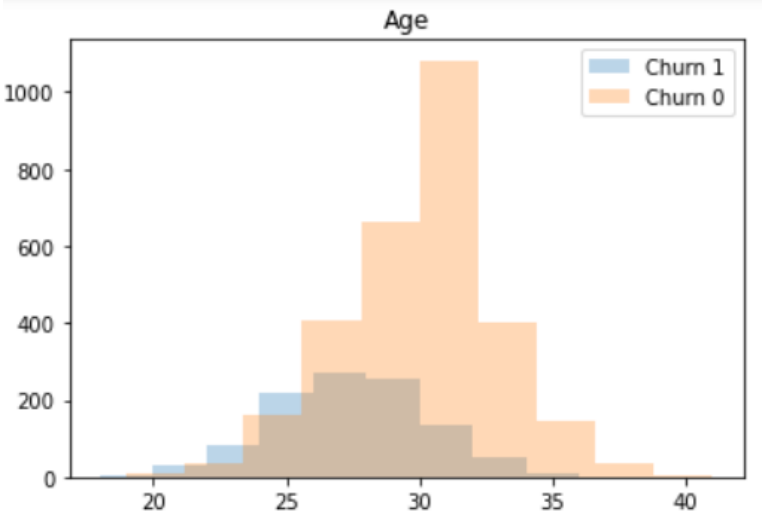


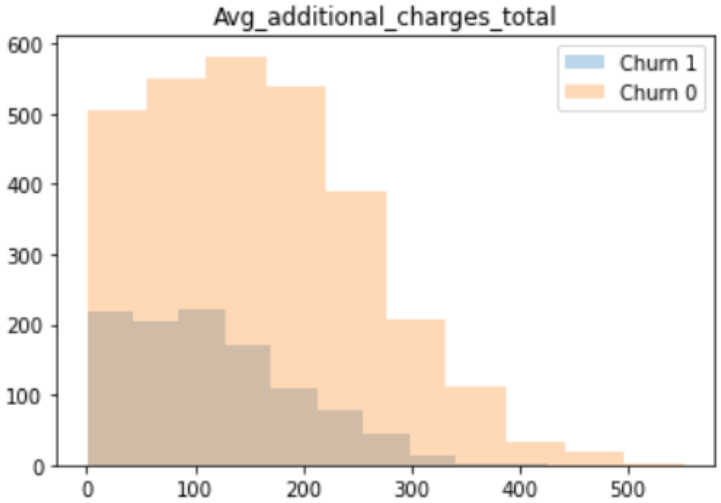

Contract_period
Contract_period
1
6
12
3000
2500
2000
1500
1000
500
0
0
1
Churn

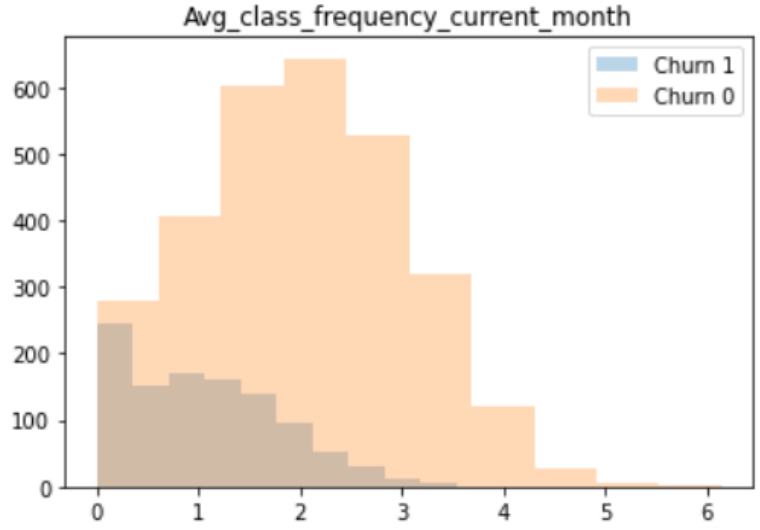

Group_visits
Group_visits
0
1
3000
2500
2000
1500
1000
500
0
0
1
Churn

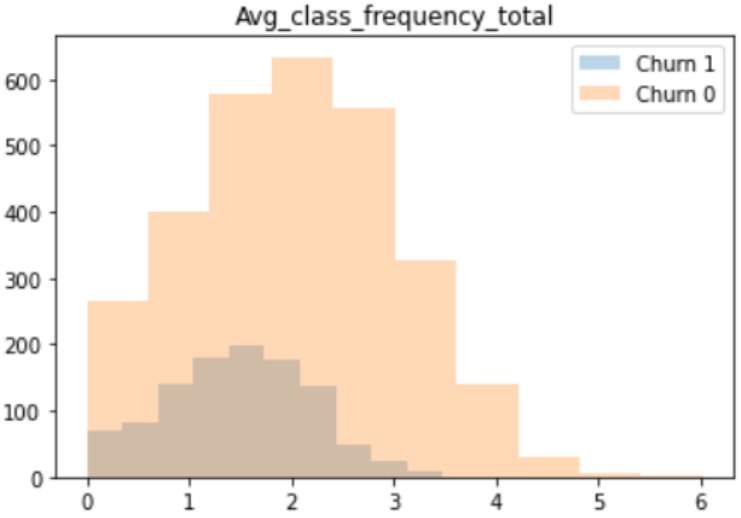

Promo_friends
Promo_friends
0
1
3000
2500
2000
1500
1000
500
0
0
1
Churn

Distributions for numerical values.

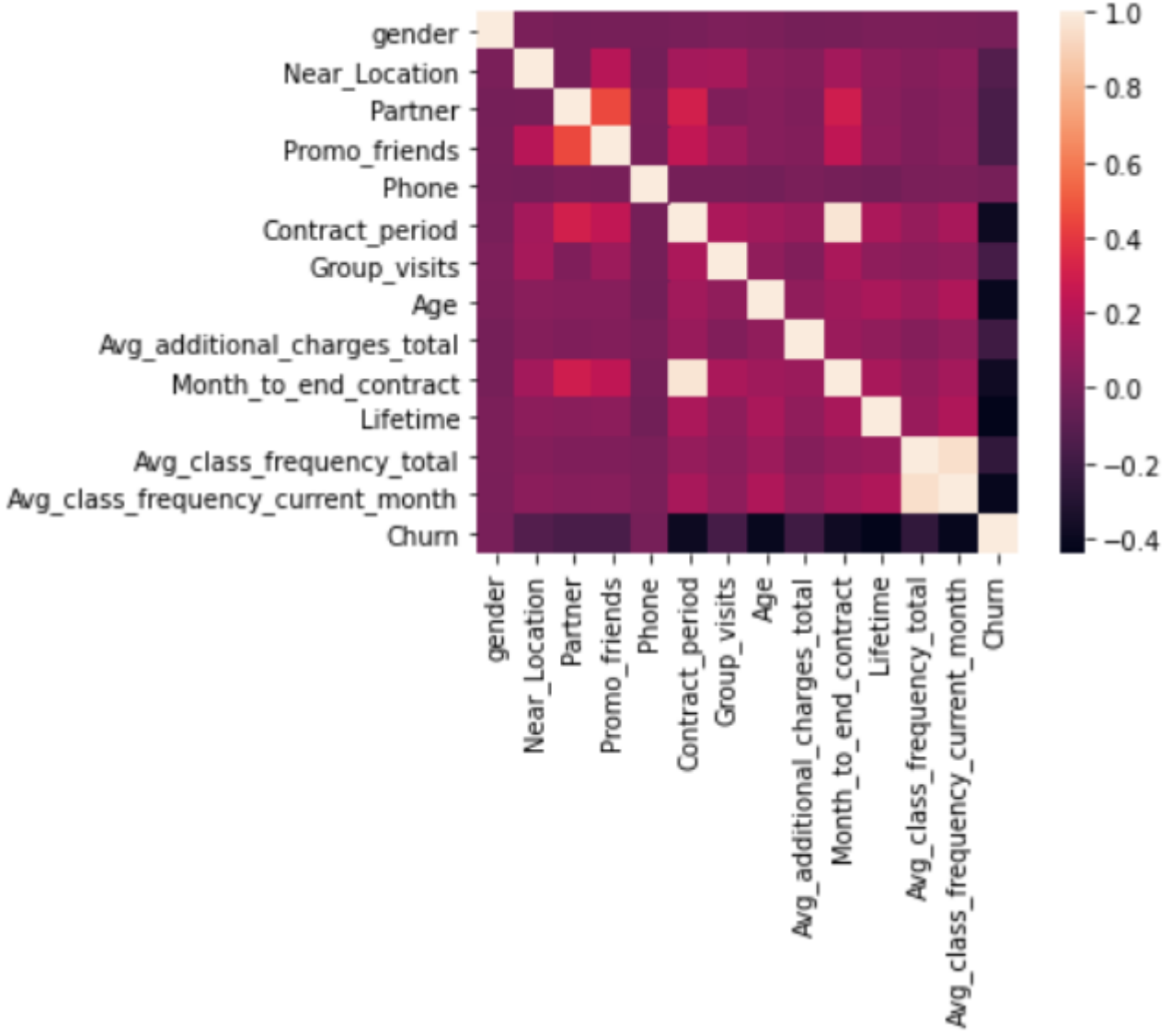








Heatmap


gender
Near_Location
Partner
Promo_friends
Phone
Contract_period
Group_visits
Age
Avg_additional_charges_total
Month_to_end_contract
Lifetime
Avg_class_frequency_total
Avg_class_frequency_current_month
Churn
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

Building and testing models.

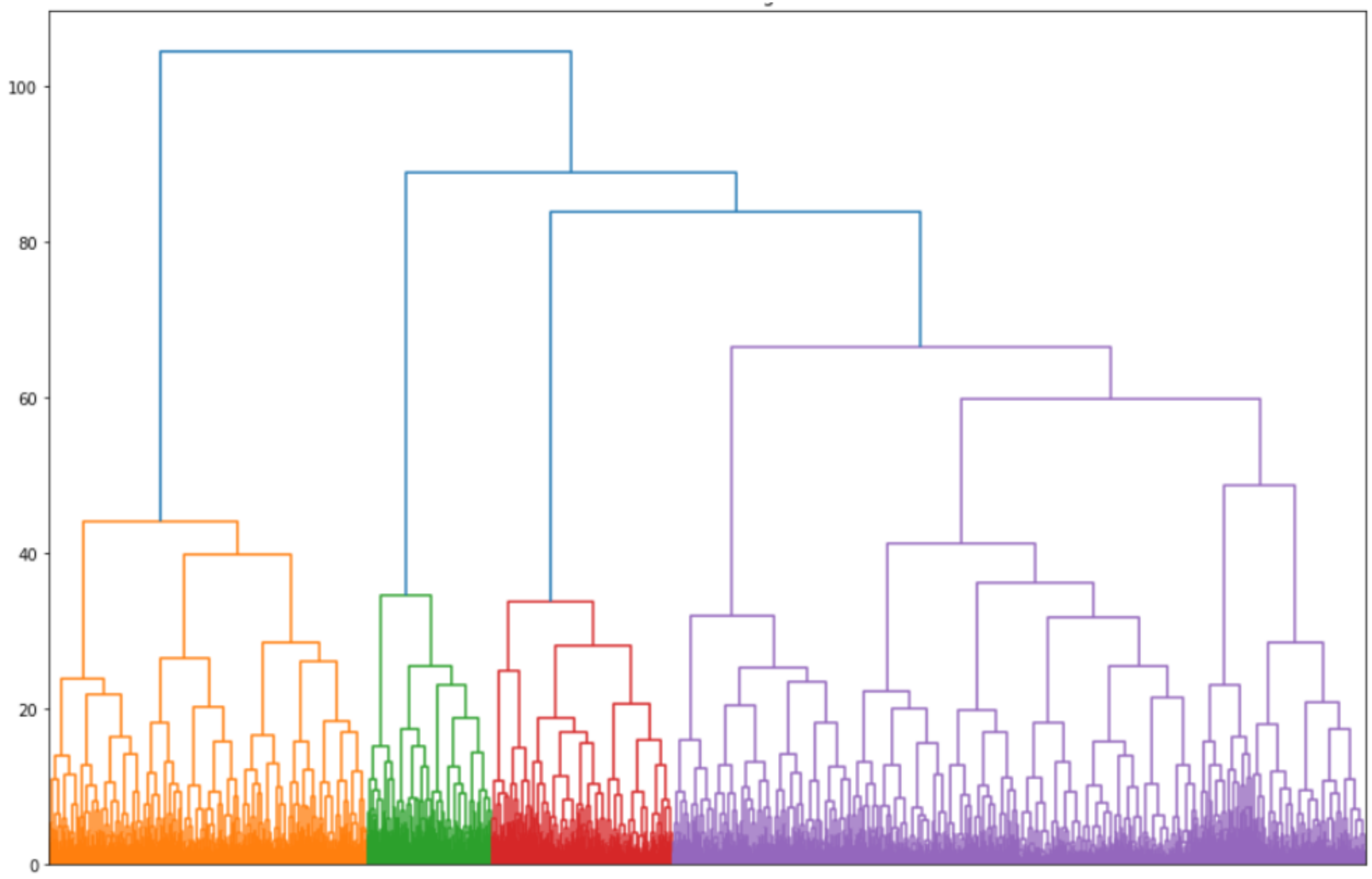
Logistic regression.

```
Accuracy: 0.92
Precision: 0.85
Recall: 0.83
F1: 0.84
```

Random forest.

```
Accuracy: 0.92
Precision: 0.84
Recall: 0.82
F1: 0.83
```

User clusters (Dendrogram)



K-means with 5 predefined clusters.

As an output we are getting cluster 4, which has lowest average churn value, which indicates that customers from this cluster:

- **90% of them live next to our gym.**
- **Were invited by friends.**
- **Visit the gym more frequently than customers from other clusters.**
- **Bought a 12 month contract.**
- **Spend more money inside of the gym.**