PYTHON PROJECT TCBDA14

TALYA DAVID 066092370

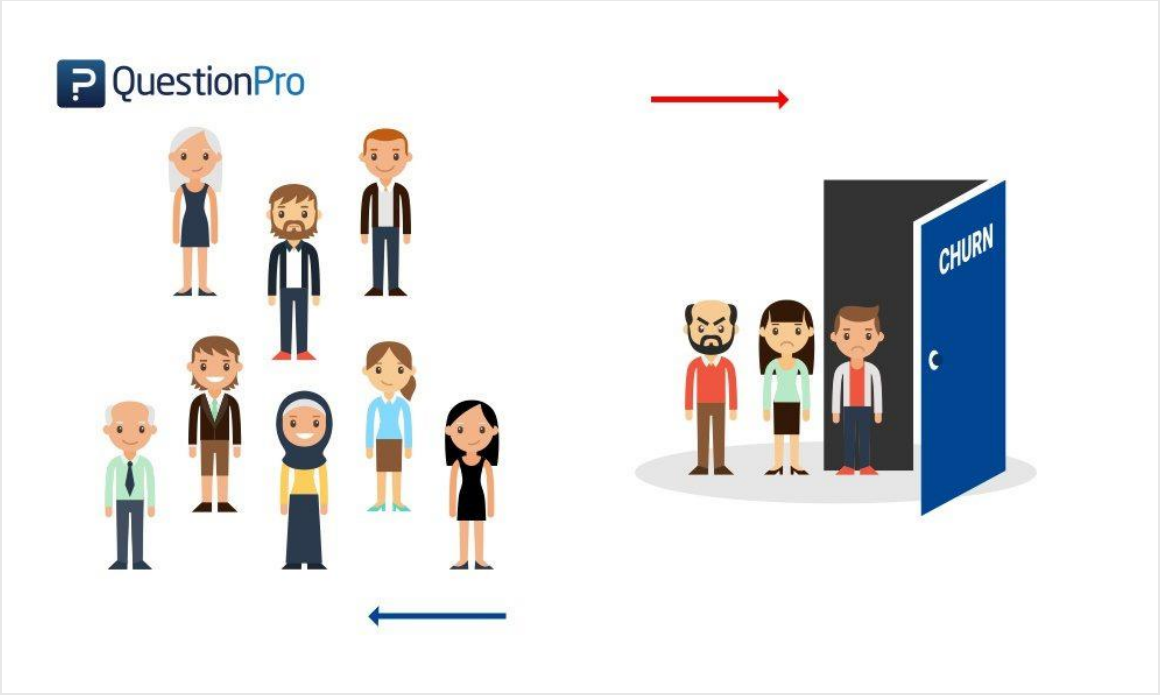
EINAT TEPPER 032291999

SIVAN LITVAK 065976029

The problem

Churn prediction

In this project, we analyzed a set of data of a telecommunication services company, in order to predict whether a client will churn the company or retain.



The motivation to solve the problem

Churn prediction is critical for many businesses because **acquiring new clients often costs more than retaining existing ones**, and if limited in resources - where to invest the available funds in order to retain the majority of clients in the most efficient way

EDA

The process

- Data preparation
- Data cleaning
- Feature engineering

The Dataset

- 7043 examples (rows)
- 21 columns
- 19 features
- 1 label (‘churn’)
- 11 missing values

The Data

- **Demographic**- Gender, senior citizen, Partner, dependents.
- **Account info**- contract type, Payment method, paperless billing, monthly charges, tenure.
- **The services provided info**-
  - Phone services
  - Multiple lines
  - Internet Services
  - Online Security
  - Online Backup
  - Device Protection
  - Tech Support
  - Streaming TV
  - Streaming Movie

- **Churn info** - Yes\ No.

Data preparation and cleaning

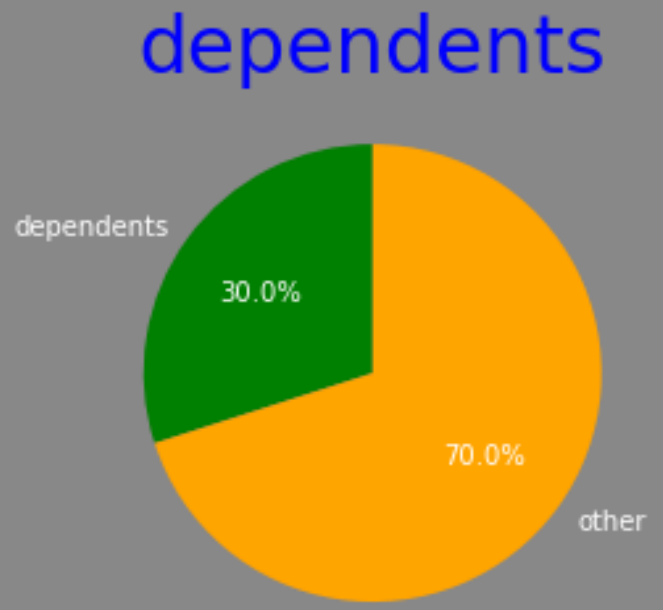
- One-Hot-Encoding for categorical features.
- Considering 'no internet service' etc as 'no'.
- Imputing-
  - 11 missing values of total charges replaced with new value: (monthly charge) x (tenure).
- Adding new features-
  - Has triple : Yes\ No
  - Tenure group
    - new customers 0-3
    - regular customers 4-10
    - solid customers 11-72
  - Charges group
    - small charge <35
    - medium charger 36-90
    - expensive charge> 90

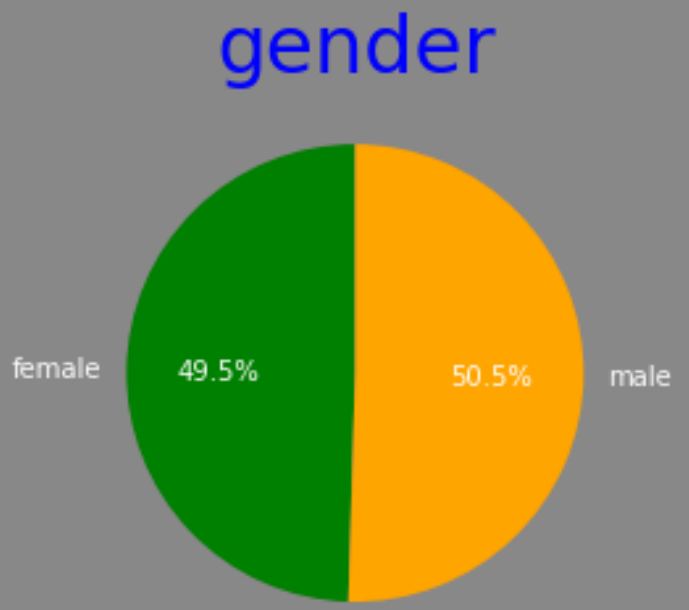
Our customers:

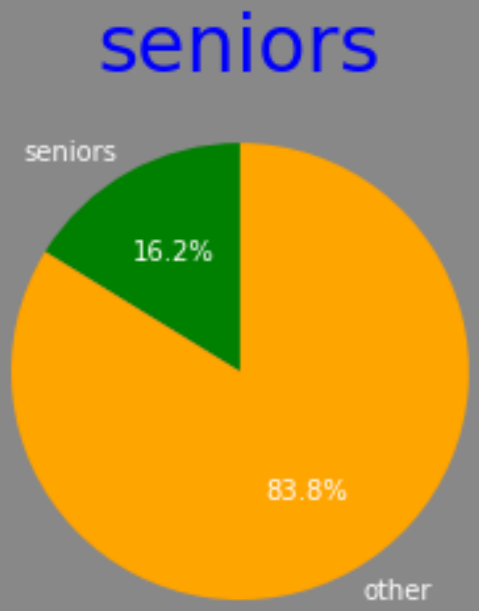
- 50% male

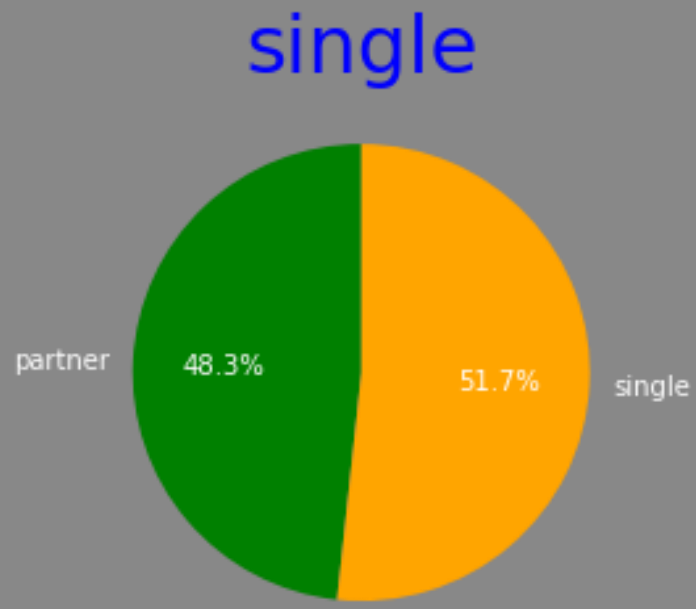
  50% female
- 16% seniors
- 30% dependents
- 52% singles

  48% partner



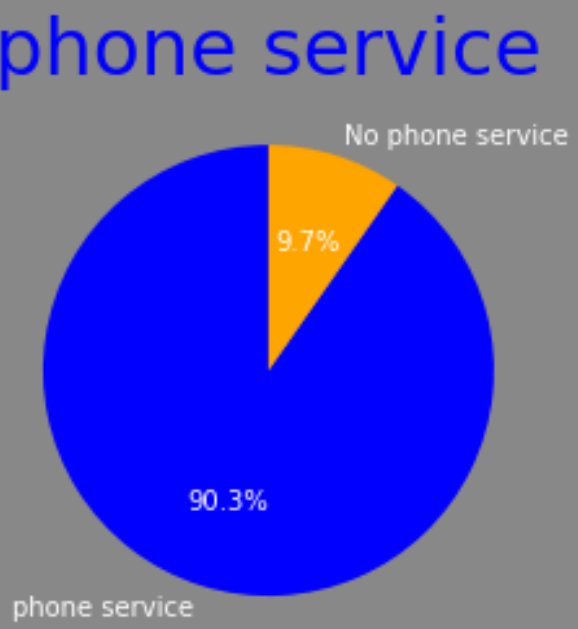


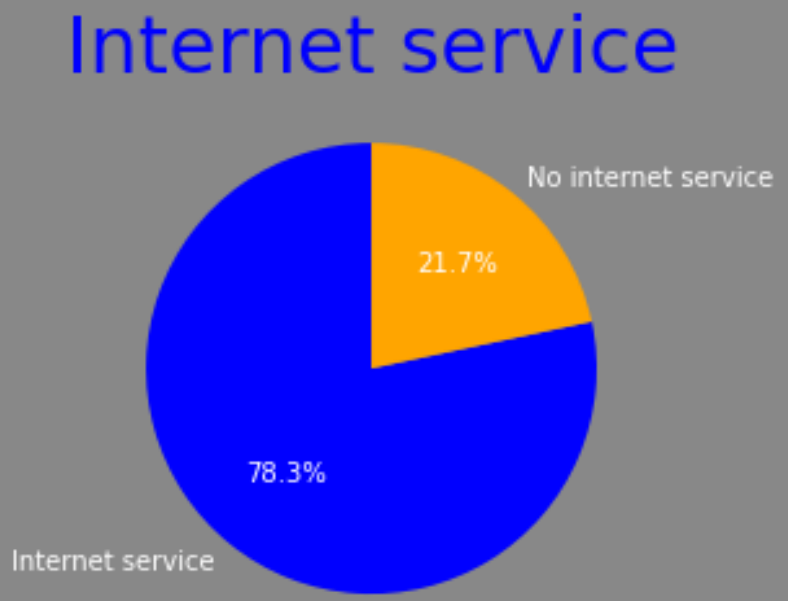


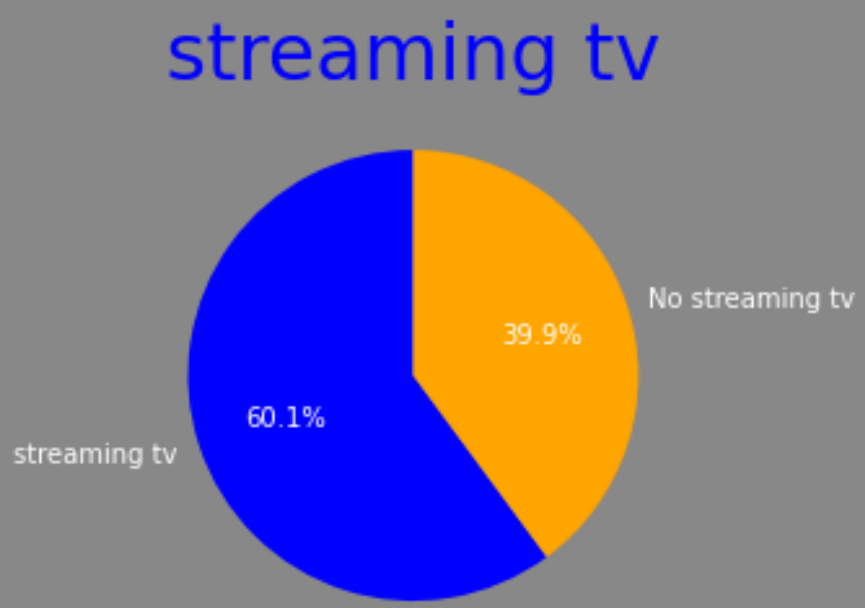


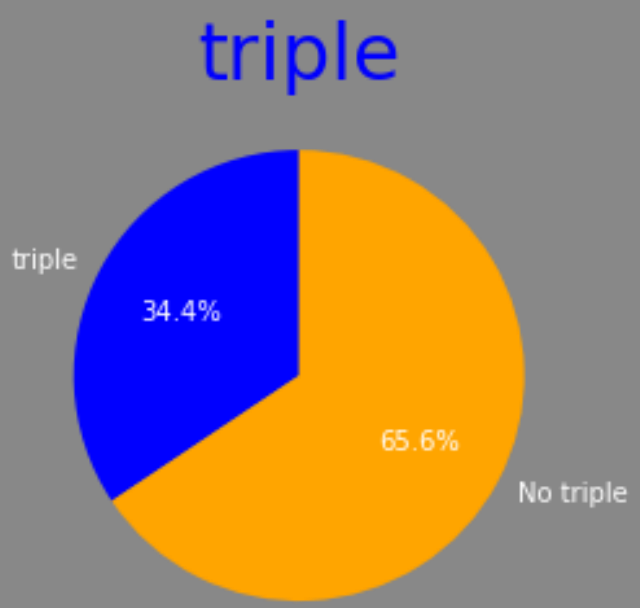
Our customers by services:

- 90% phones service
- 78% internet service
- 60% streaming tv
- 34% triple



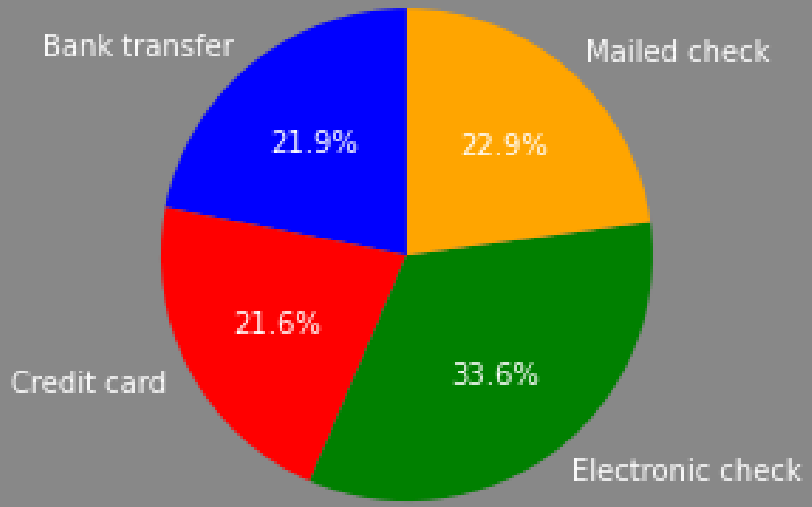






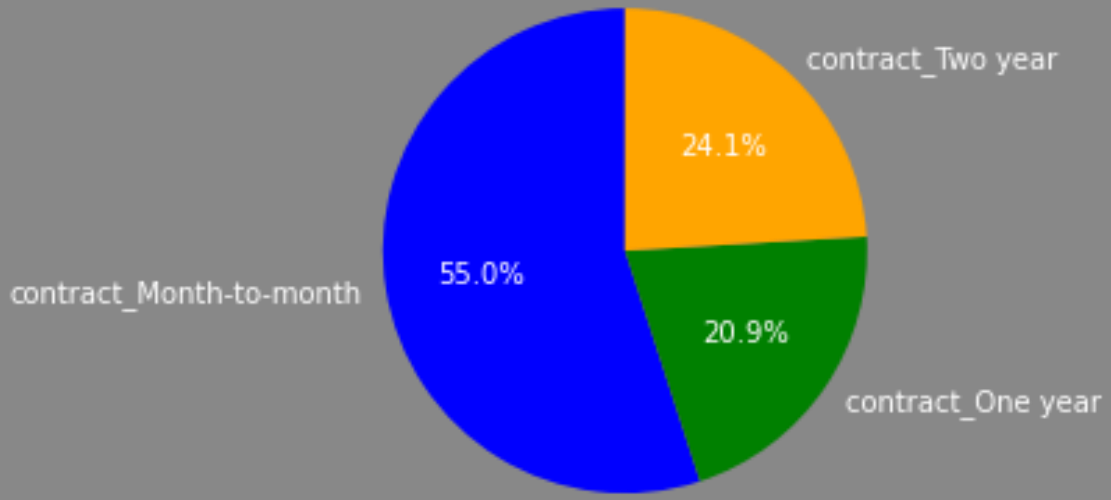
Payment method

33% of the customers are paying with electronic check.

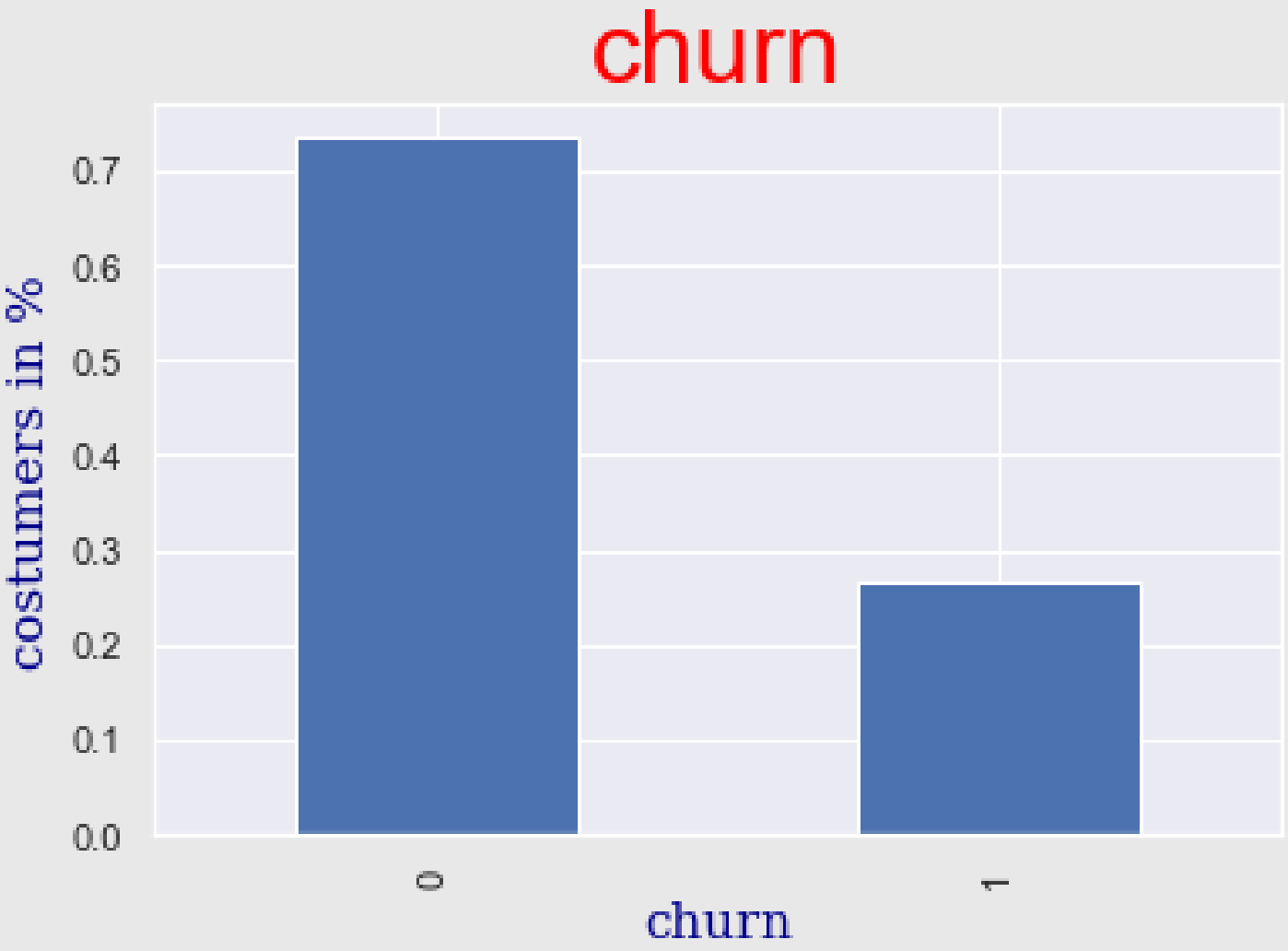


Contract

55% of the customers have month to month contracts.



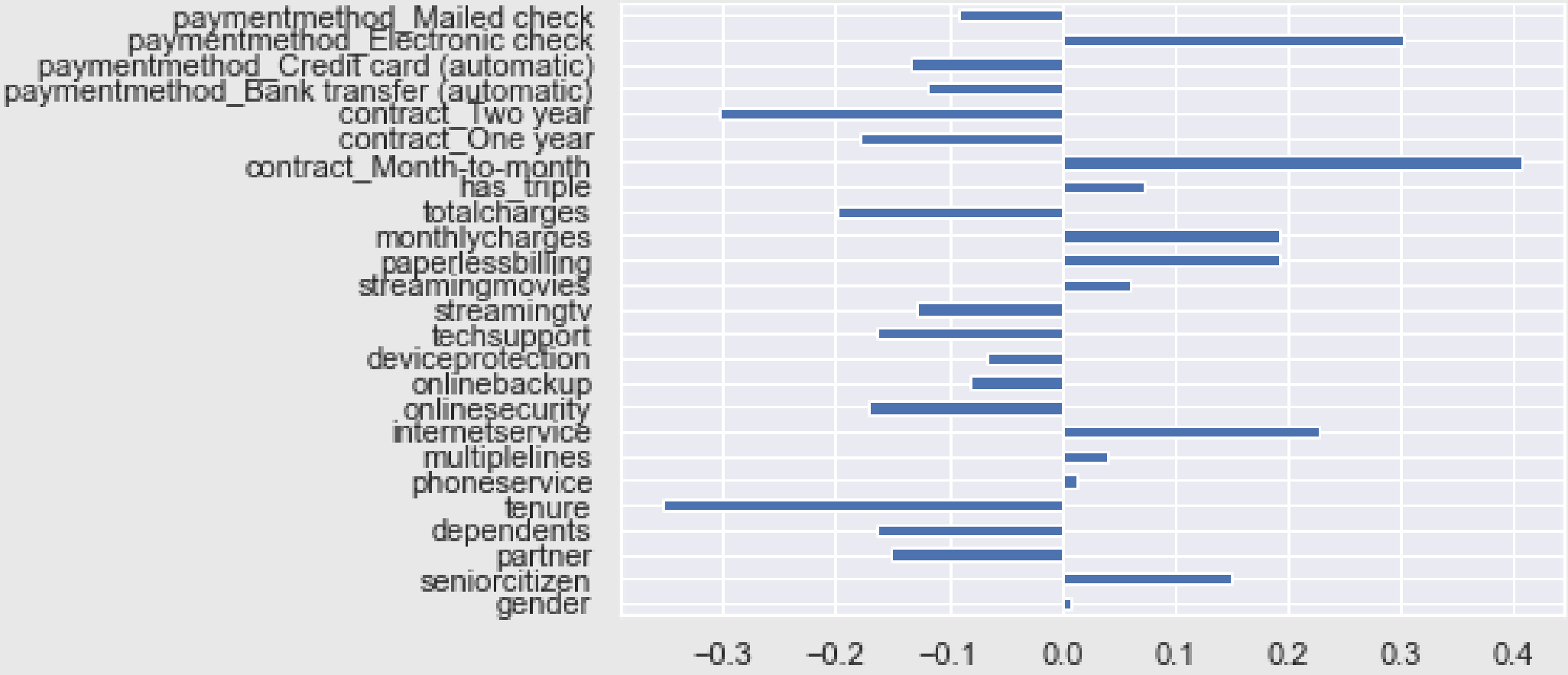
How many customers has churned—> 26.5%



Correlation-

we checked the correlation between churn and our features(including the new features we have created).

The features with high correlation to churn are: tenure , contract, payment method ,total charges and monthly charges.



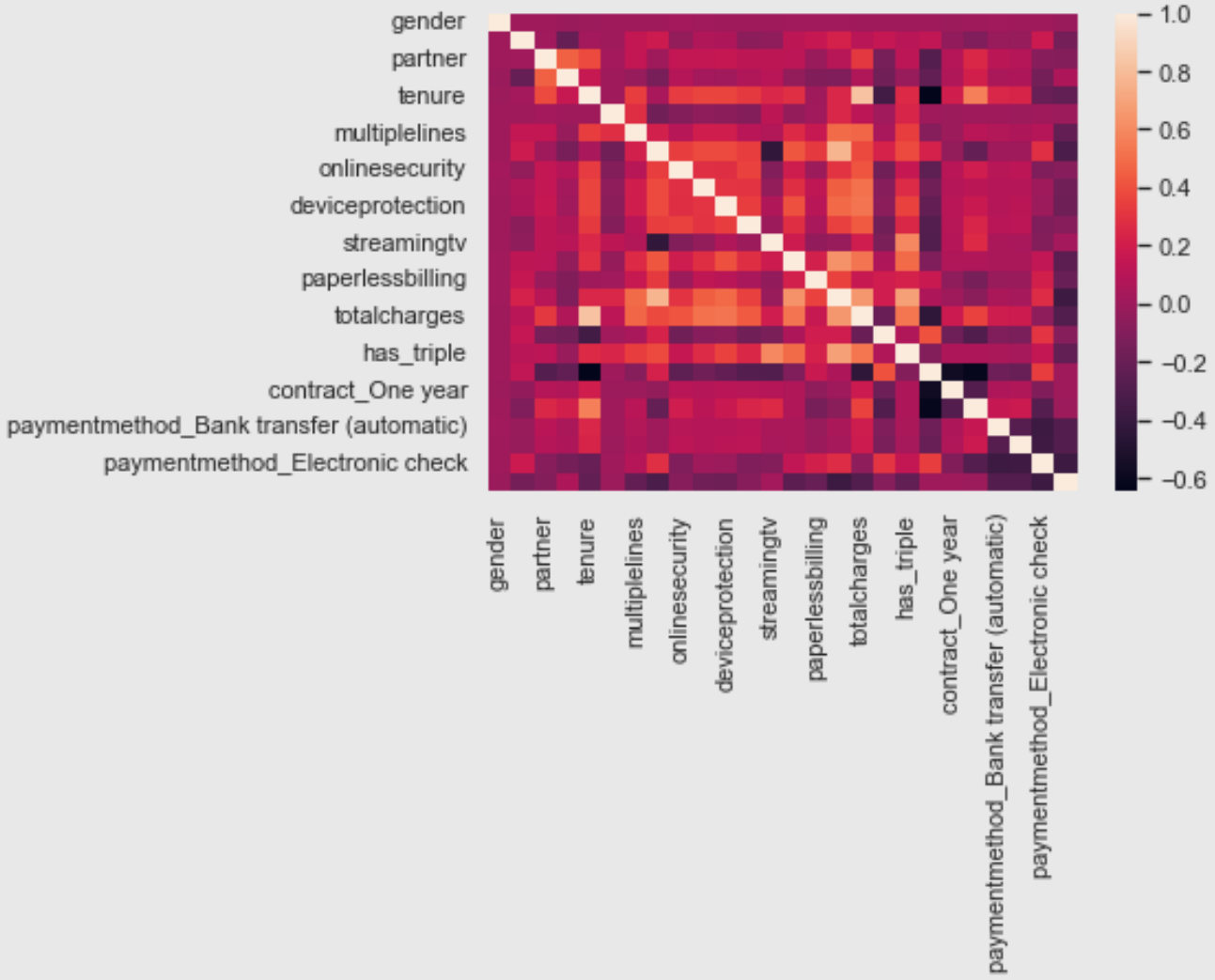
Correlation with Heatmap

The correlation Heatmap display uses different shades of red to visualized the strength of the correlation .

The shades are used to distinguish between positive/negative correlation.

brighter shades = positive correlation.

darker shades = negative correlation.

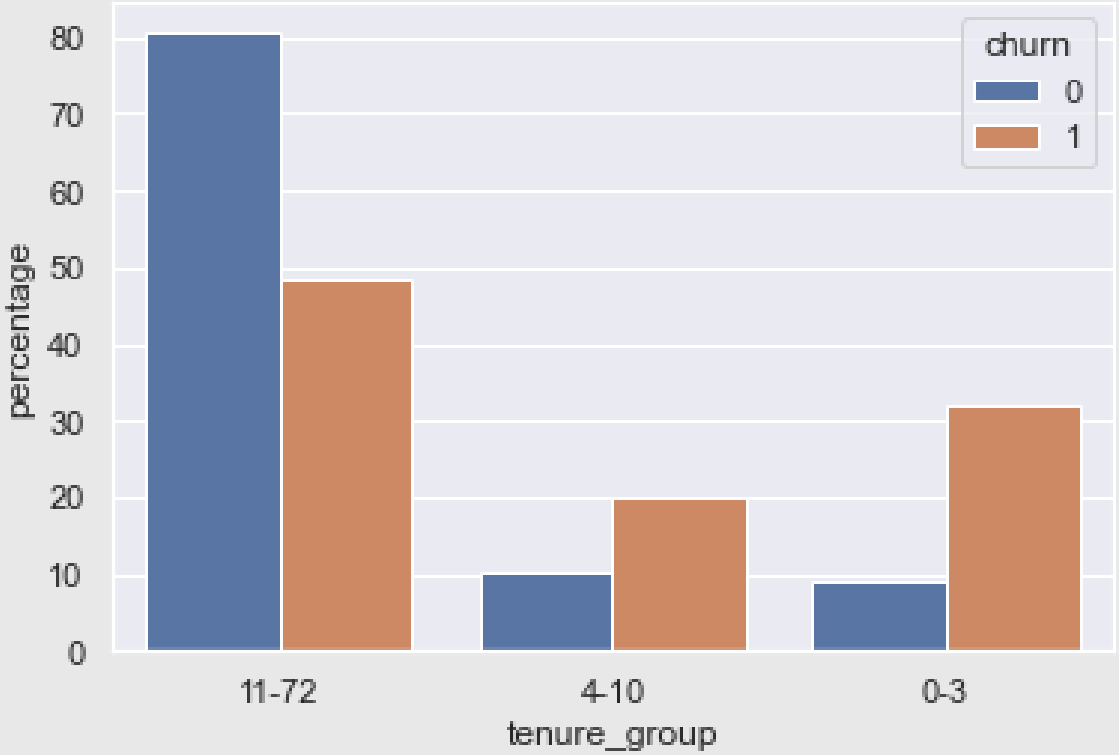


TENURE

There is a correlation between the feature “tenure” to the label “churn”:

New customers tend to churn more then others.

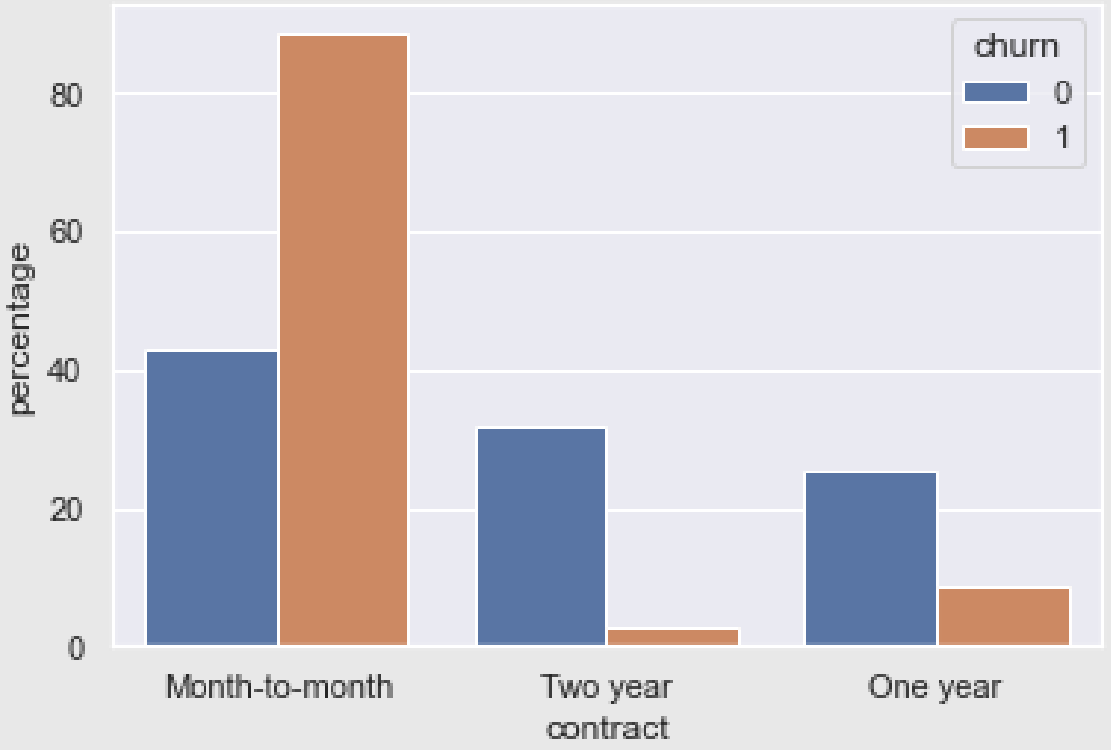
- Tenurity greater than 10 years → less chance to churn.
- High tenure = less chance to churn
- Low tenure = higher chance to churn



CONTRACT

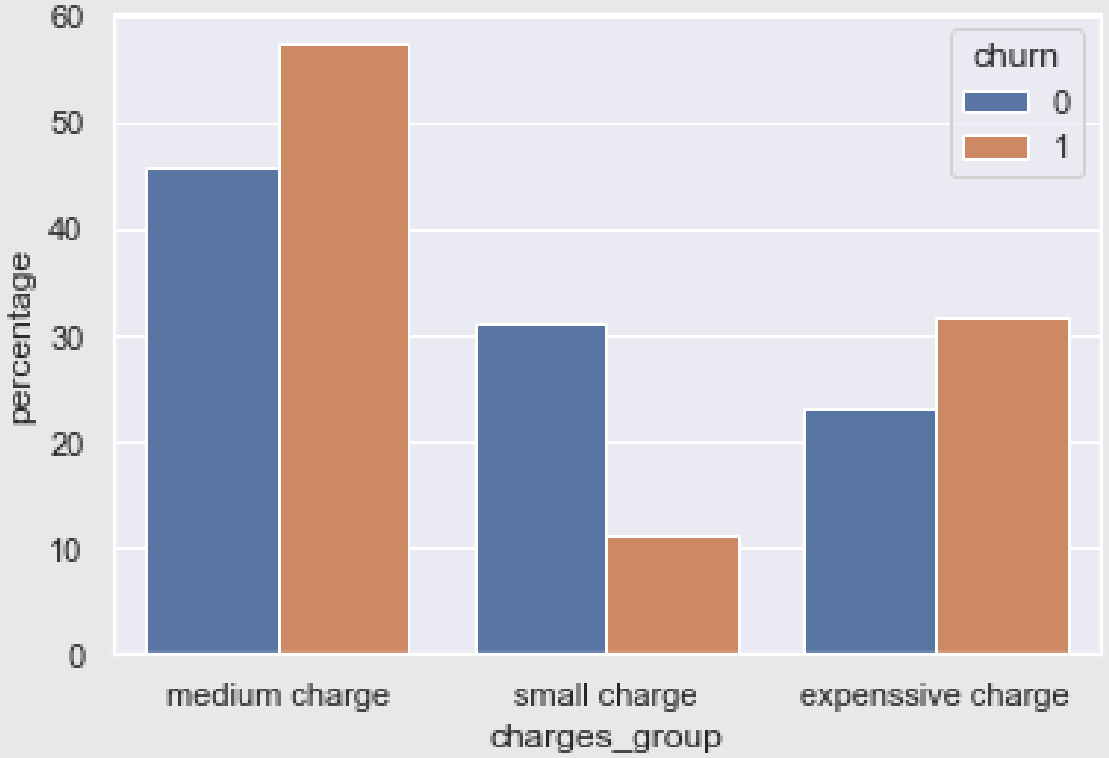
There are 3 types of contracts. we examined the correlation between each type and churn → the longer the contract is the less chance to churn.

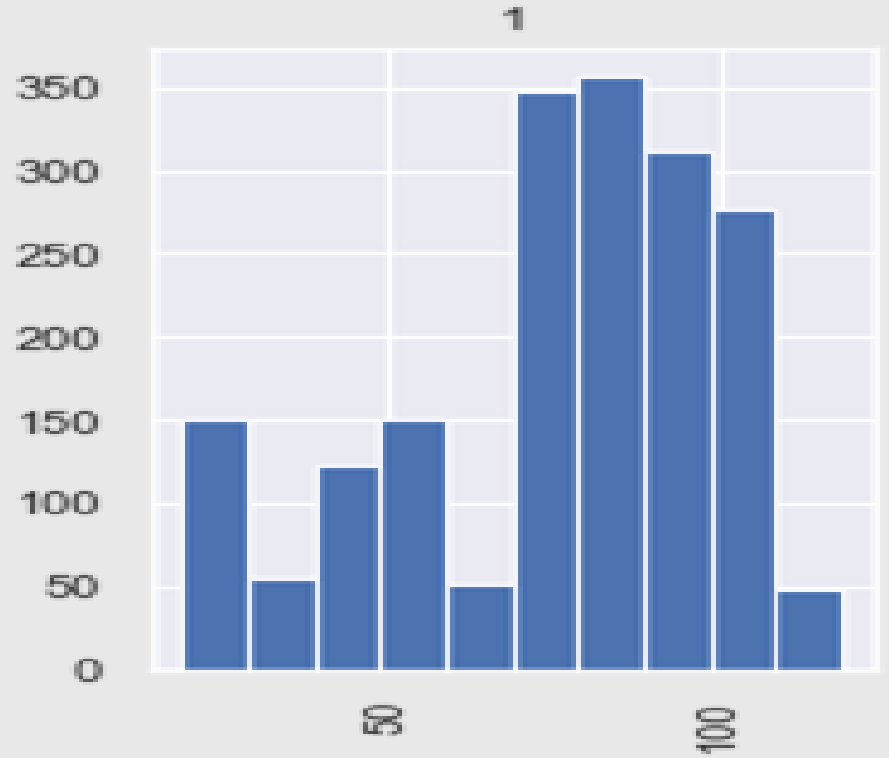
- month to month contract → higher chance to churn.
- long contract → less chance to churn



CHARGES

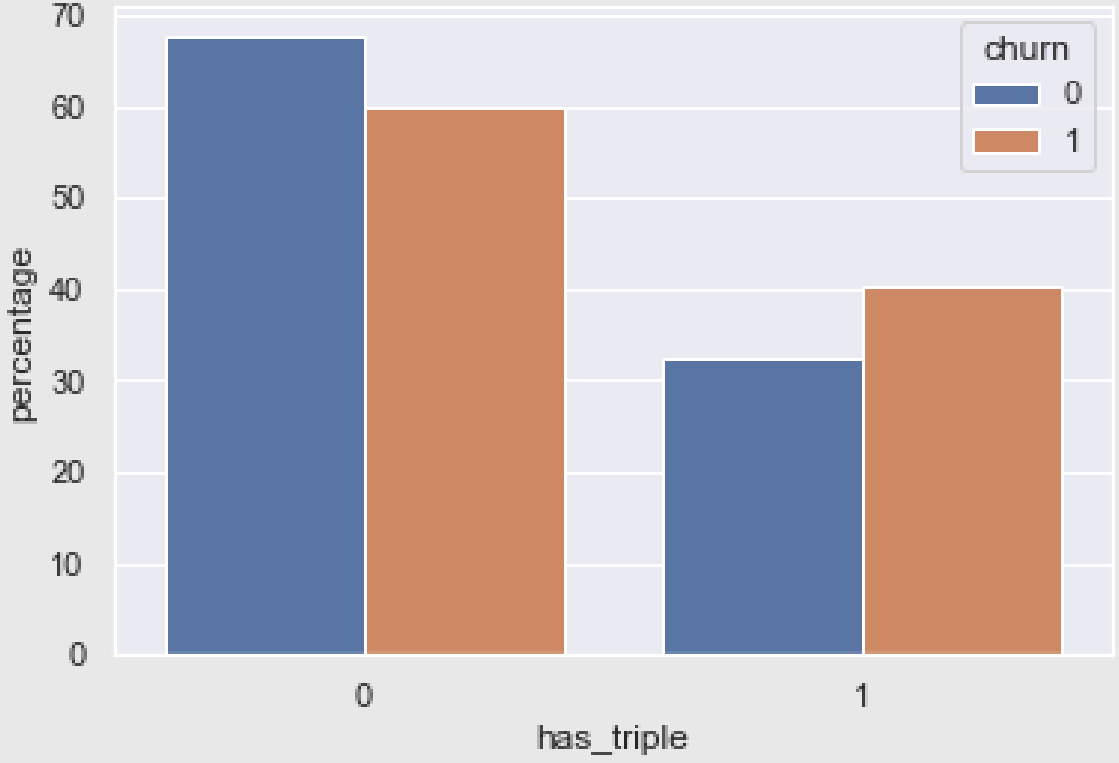
- We created 3 groups, divided by monthly charges rate.
- Small monthly charges → smallest chance to churn.
- Expansive monthly charges → higher chance to churn.





TRIPLE- Phone service+Internet service+Streaming TV

We created a new feature to examine if there is a correlation between customers who required triple to churn.



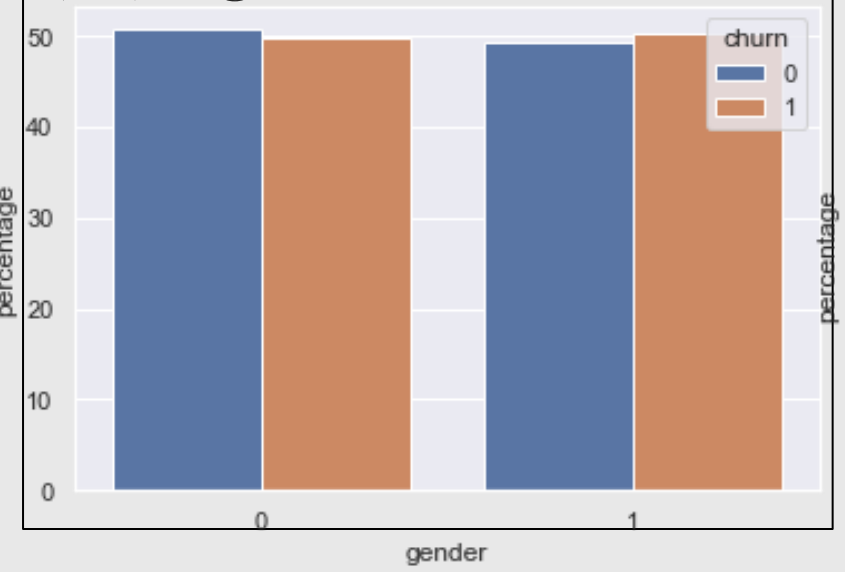
34% of the customers required triple.

26.5% of those customers has churned.

→ Customers who consume triple are more likely to churn.

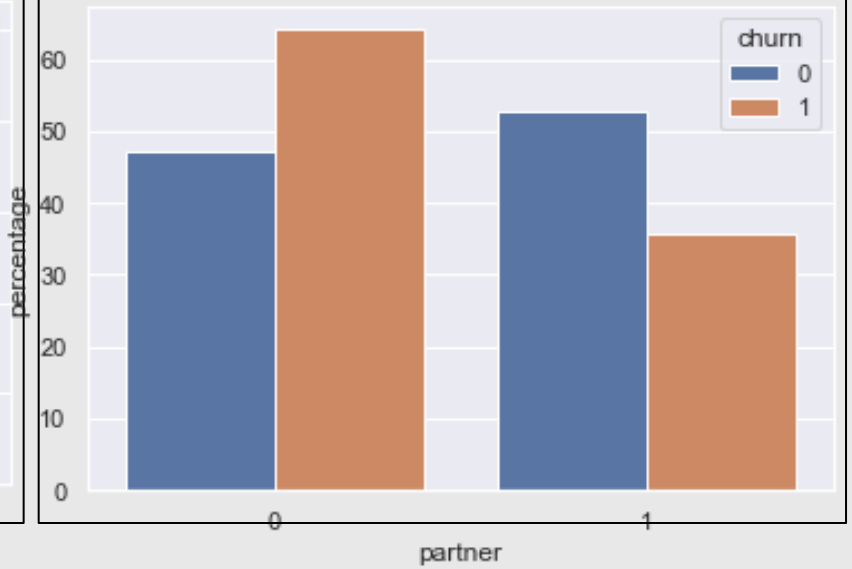
Gender:

Gender has no correlation with churn. This is why we are ignoring this feature in our machine learning (ml) algorithms.



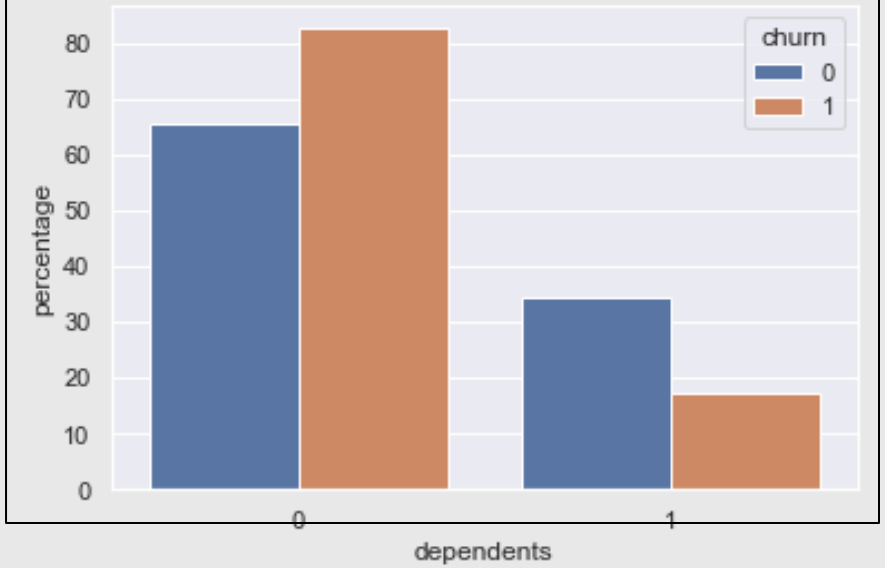
Clients has partners:

Customers without partners (single) are more likely to churn then customers with partners



Clients has dependents:

Customers with no dependents are more likely to churn.



MACHINE LEARNING ALGORITHMS

In order to predict if a customer would churn or not, we splitted the data to 80% train model and 20% test model.There enough examples in the dataset in relation to the number of features, therefore we can not expect overfitting.

Train		Test
80%	percentage	20%
5634	examples	1409

The algorithms we tried are:

- Decision tree
- Random forest algorithm.
- KNN algorithm.

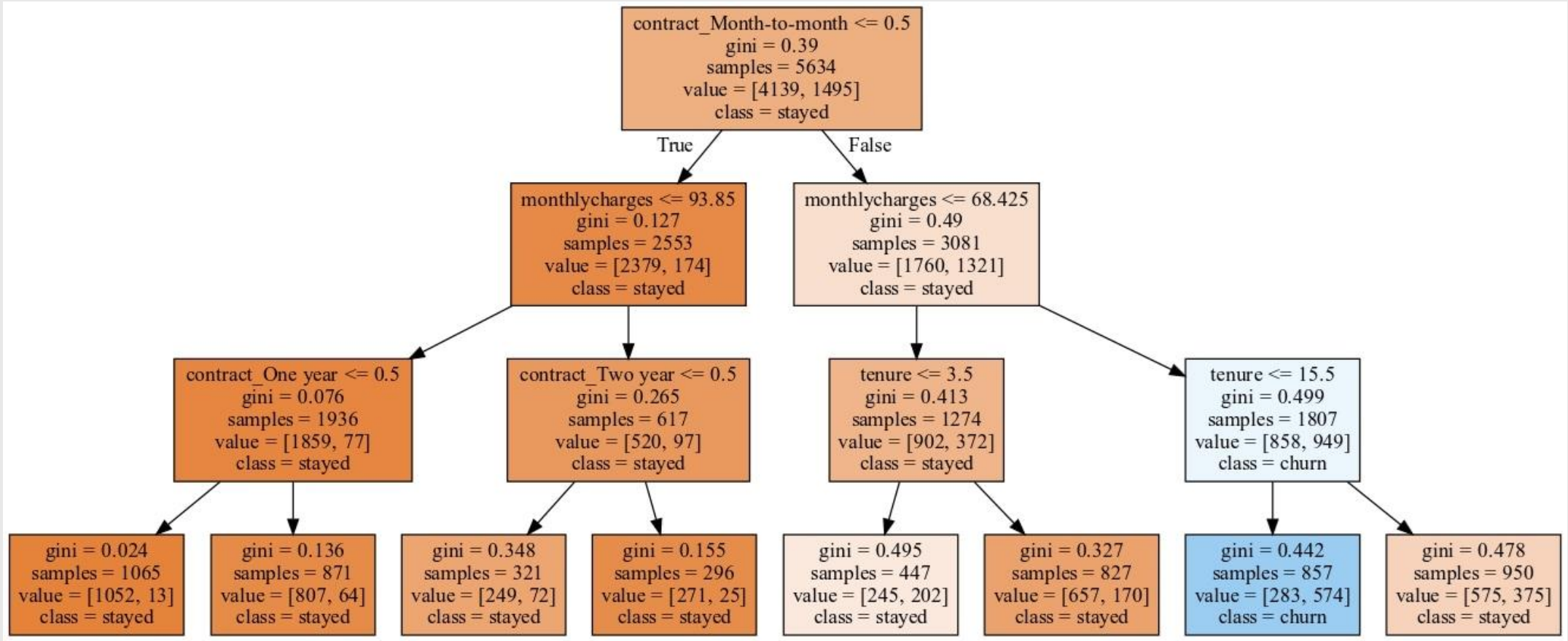
Decision tree

max depth=3 accuracy= 79.20%

max depth=4 accuracy= 79.34%

max depth=6 accuracy= 79.48%

Max depth of 6 is the best accuracy for decision tree!



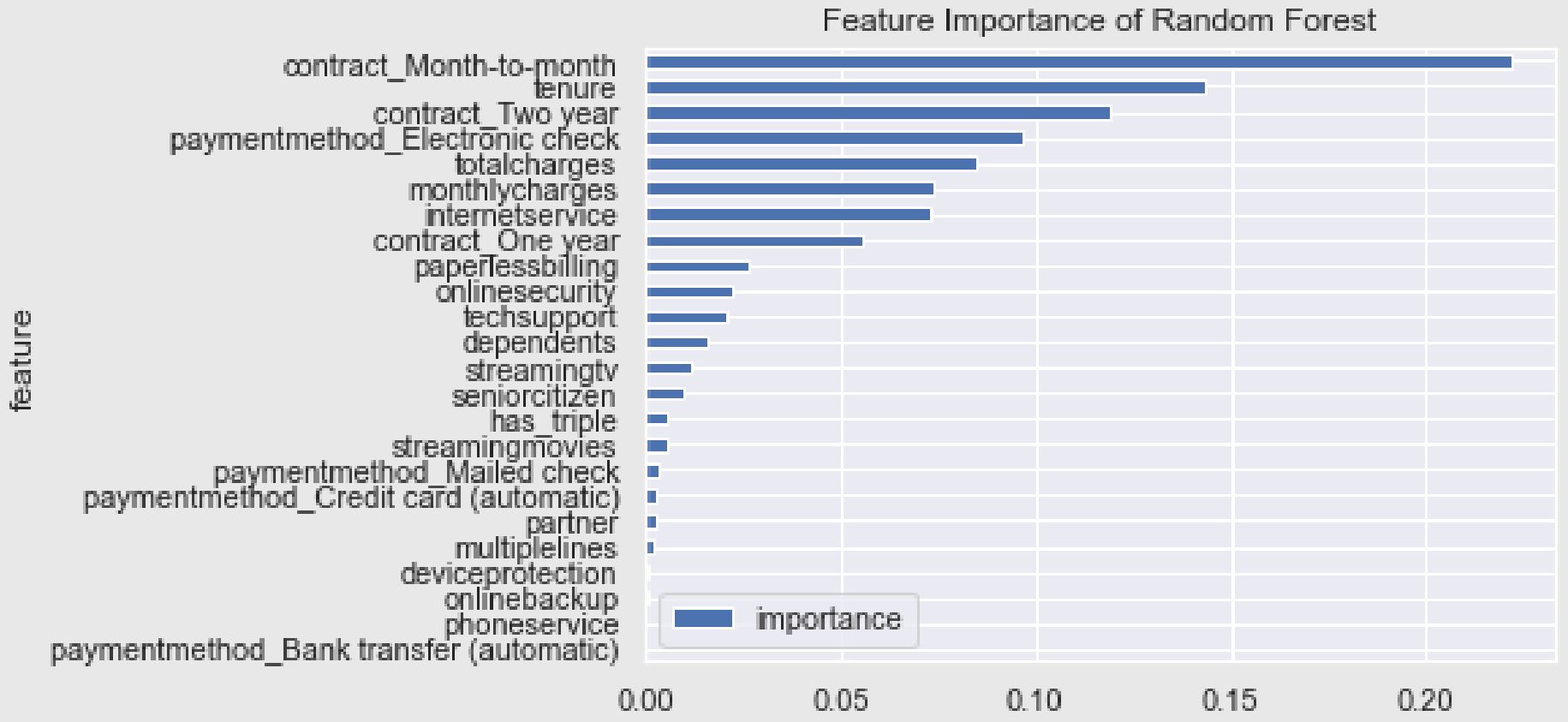
Random forest

max depth=4 accuracy= 78.99%

max depth=5 accuracy= 79.48%

max depth=6 accuracy= 81.40%

Max depth of 6 is the best accuracy for Random Forest!



KNN

k nearest neighbor

accuracy=74.44%

Benchmark

benchmark means to insert churn=0 (stayed) in the train and to see what the accuracy for this model.

accuracy=73.45%

Best performance model

RANDOM FOREST

Accuracy= 81.4%

CONFUSION MATRIX

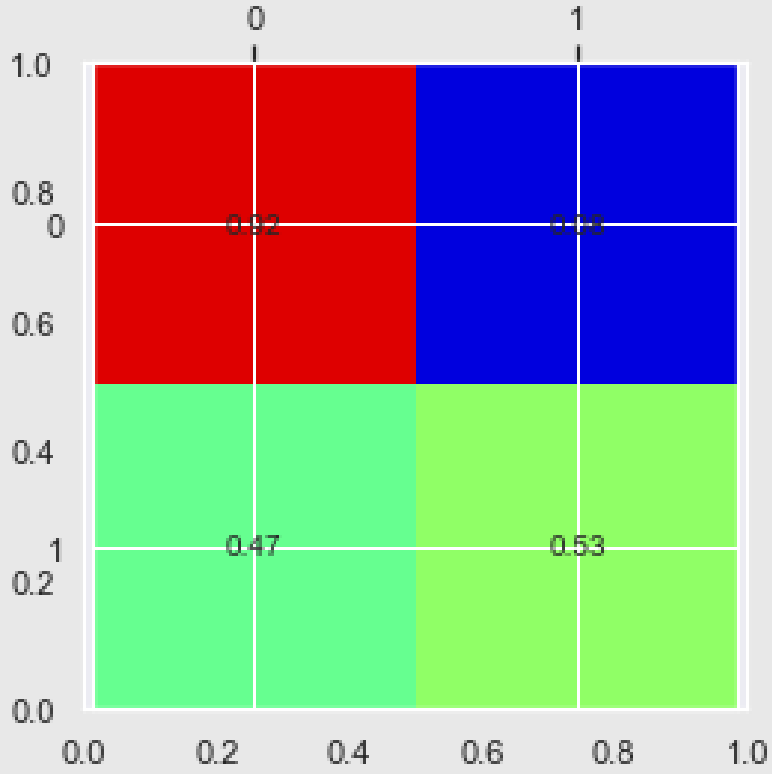
Confusion matrix demonstrates the model's accuracy:

The model predicted:

- 92% of the staying customers

(it failed with 8% that were predicted as existing customers although they churned), and -

- 53% of the churning customers (it failed with 47% that were predicted as churn while they didn't).



For small companies which don't have the resources to invest in customer retention - confusion matrix can help aiming the resources to the most profitable customers - in this case - to focus on the 53% that the model has correctly predicted as churn.

Conclusions

Conclusions-

The company can now estimate whether a customer will churn or not, and concentrate its efforts on the predicted churning customers in order to preserve them (or not, as a function of the required resources in preserving them and the net gain from from such efforts.

Based on those predictions the company can build its strategic business plan diverting resources in the most efficient way

Thank you