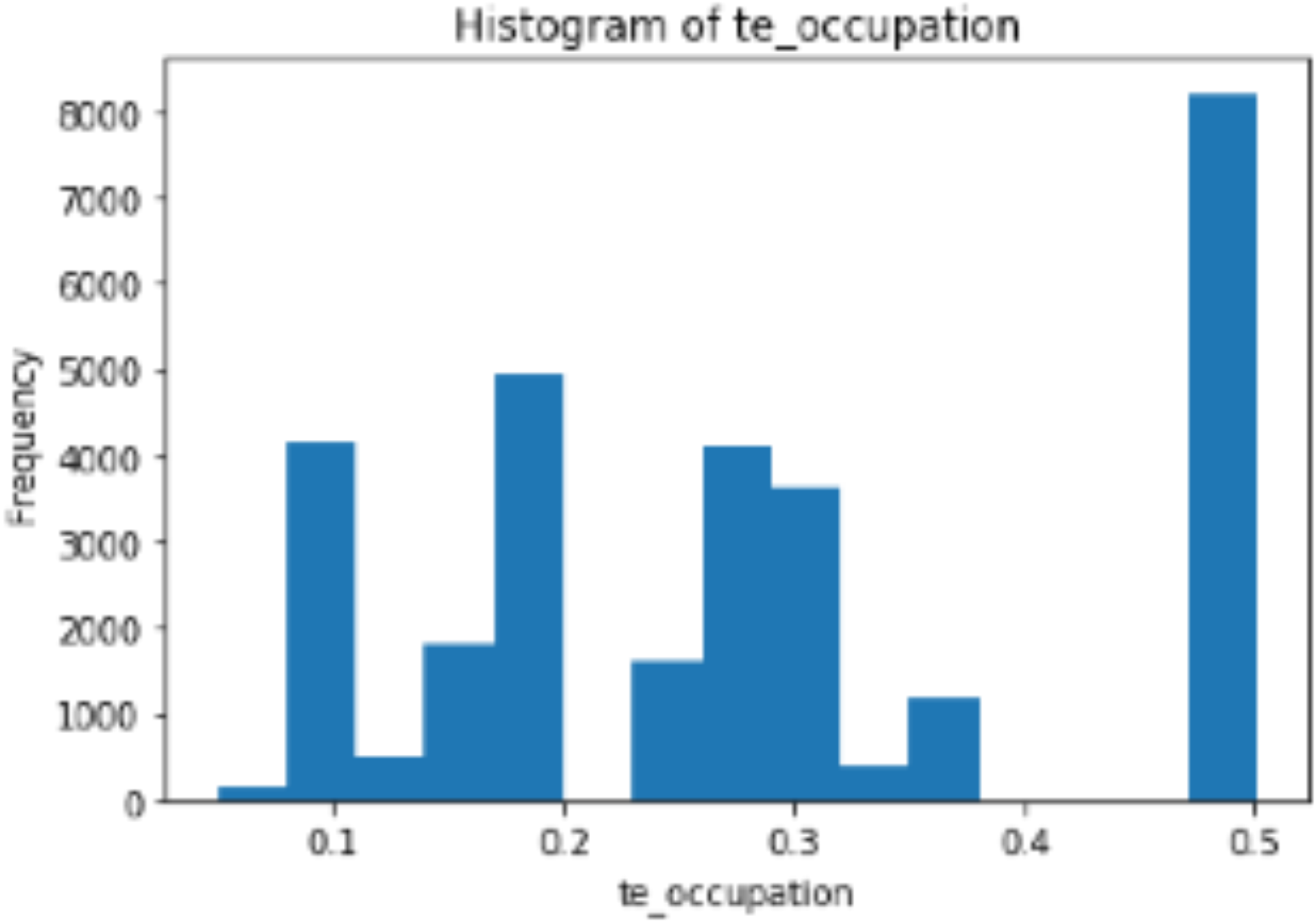

H2O.ai

Histogram of Numeric Columns

```
import matplotlib.pyplot as plt
%matplotlib inline

import warnings
import matplotlib.cbook
warnings.filterwarnings("ignore", category = matplotlib.cbook.mplDeprecation)

census_data["te_occupation"].hist()
```


Histogram of te_occupation
Frequency
8000
7000
6000
5000
4000
3000
2000
1000
0
0.1
0.2
0.3
0.4
0.5
te_occupation

Case Study: Lending Club Dataset

- Loan data from 2007 up until 2015 including rejected applications and accepted applications.
- Of the 500k accepted applicants about 160k loans have either been completely paid off or defaulted.
- There are about 4 million applicants in the rejected loans dataset.
- **Use Case 1**: Predict the likelihood of a user defaulting based on the information supplied when applying for a loan.
- **Use Case 2**: Determine the interest rate Lending Club would have offered the user based on the information supplied when applying for a loan.
- Full Data: https://www.kaggle.com/wendykan/lending-club-loan-data
- H2O Subset: https://s3.amazonaws.com/h2o-public-test-data/bigdata/laptop/lending-club/loan.csv

Histogram of Numeric Columns

```
1 import matplotlib.pyplot as plt
2 %matplotlib inline
3
4 import warnings
5 import matplotlib.cbook
6 warnings.filterwarnings("ignore", category = matplotlib.cbook.mplDeprecation)
7
8 census_data["te_occupation"].hist()
```