# Main Question

*What are the characteristics of a customer who is likely to follow through with a purchase?*

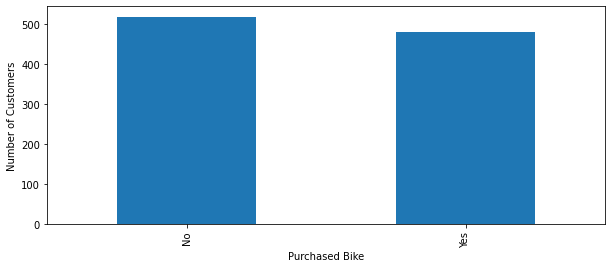
# Tasks

1. Do some EDA
2. Data cleaning and feature engineering
3. Model Building
4. Present findings

# EDA

## Target

Shape is (1000,13)



No 519

Yes 481

## Variables

**Categorical**

['Marital Status', 'Gender', 'Education', 'Occupation', 'Home Owner', 'Commute Distance', 'Region']

* ['Marital Status',

Married 539

Single 461

* 'Gender',

Male 509

Female 490

* 'Education',

Bachelors 305

Partial College 265

High School 179

Graduate Degree 174

Partial High School 76

* 'Occupation',

Professional 276

Skilled Manual 255

Clerical 177

Management 173

Manual 119

* 'Home Owner'

Yes 685

No 315

* 'Commute Distance',

0-1 Miles 366

5-10 Miles 192

1-2 Miles 169

2-5 Miles 162

10+ Miles 111

* 'Region',

North America 508

Europe 300

Pacific 191

* 'Purchased Bike'

No 519

Yes 481

**Numerical**

['Income', 'Children', 'Cars', 'Age']

ID Income Children Cars Age

count 1000.000000 1000.000000 1000.00000 1000.00000 1000.000000

mean 19965.992000 56220.000000 1.91500 1.45100 44.190000

std 5347.333948 31475.423783 1.65424 1.12644 11.353537

min 11000.000000 10000.000000 0.00000 -1.00000 25.000000

25% 15290.750000 30000.000000 0.00000 1.00000 35.000000

50% 19744.000000 60000.000000 2.00000 1.00000 43.000000

75% 24470.750000 70000.000000 3.00000 2.00000 52.000000

max 29447.000000 250000.000000 12.00000 4.00000 89.000000

\* Can see that the min for cars is -1, thus need to ensure all columns are positive

ID Income Children Cars Age

count 1000.000000 1000.000000 1000.00000 1000.000000 1000.000000

mean 19965.992000 56220.000000 1.91500 1.452000 44.190000

std 5347.333948 31475.423783 1.65424 1.124705 11.353537

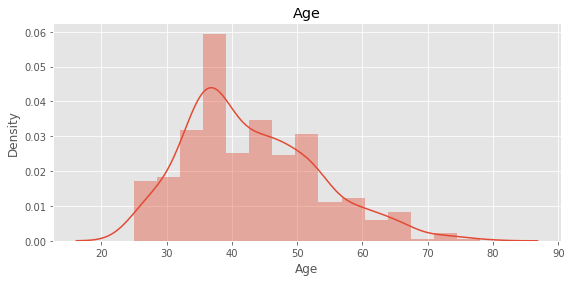
min 11000.000000 10000.000000 0.00000 0.000000 25.000000

25% 15290.750000 30000.000000 0.00000 1.000000 35.000000

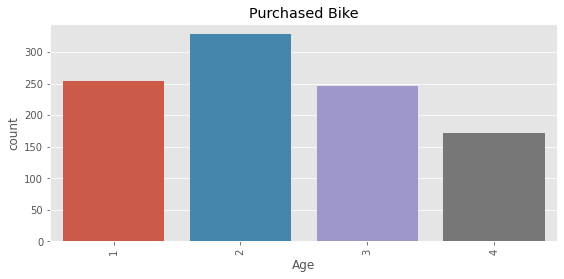
50% 19744.000000 60000.000000 2.00000 1.000000 43.000000

75% 24470.750000 70000.000000 3.00000 2.000000 52.000000

max 29447.000000 250000.000000 12.00000 4.000000 89.000000



Tried converting to categorical



Tried with income

Chart, histogram

Description automatically generated

Chart, bar chart

Description automatically generated

## Missing Data

Three variables with missing data with one of each missing

['Gender', 'Education', 'Region']

Replaced with mode