Review

AI and Machine Learning

Hult International Business School

Michael de la Maza

Version 1.0

ACME Telephone Case Study

Business understanding

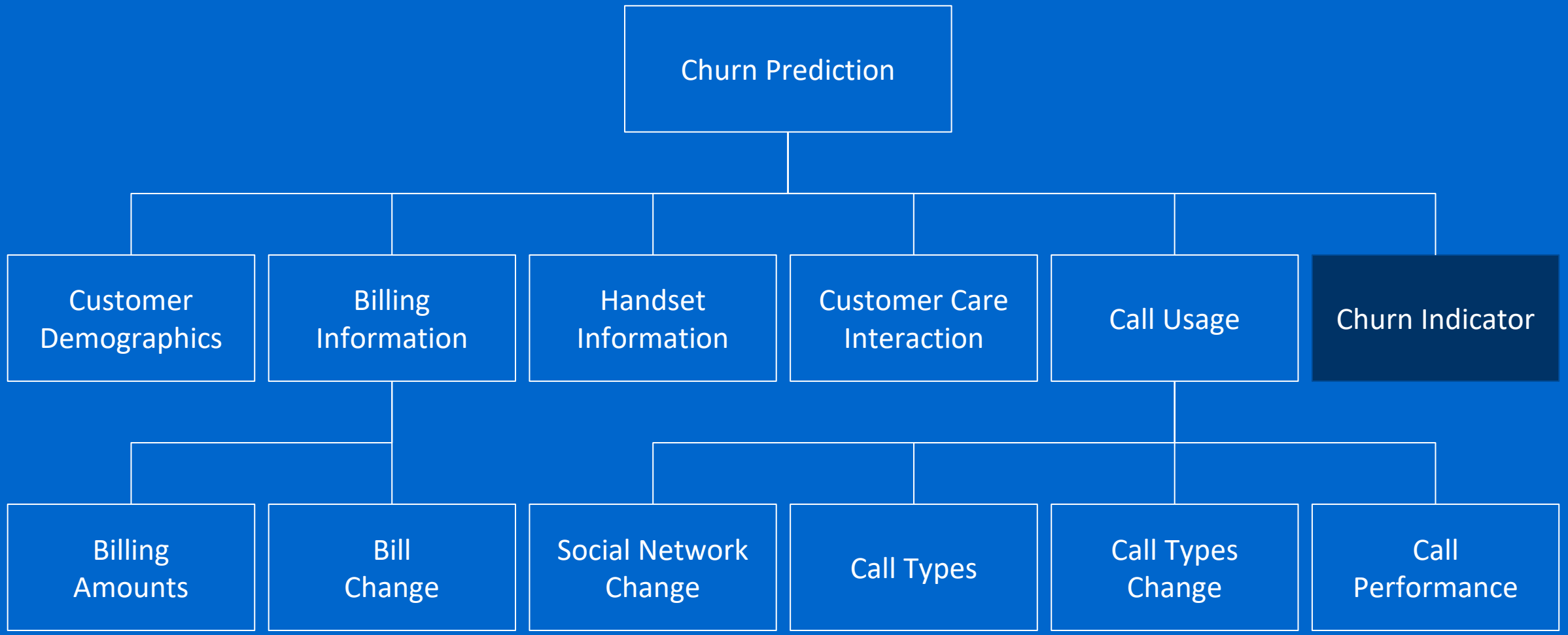
Acme sells cell phone plans

Acme has approached you to reduce customer churn

They have not proposed any particular predictive analytics solution

They have provided you with a dataset with 10K customer records and approximately 30 attributes

Domain understanding



Source: Fundamentals of Machine Learning for Predictive Analytics by Kelleher, et al

Attribute descriptions

Feature	Description
BILLAMOUNTCHANGEPCT	The percent by which the customer's bill has changed from last month to this month
CALLMINUTESCHANGEPCT	The percent by which the call minutes used by the customer has changed from last month to this month
AVGBILL	The average monthly bill amount
AVGRECURRINGCHARGE	The average monthly recurring charge paid by the customer
AVGDROPPEDCALLS	The average number of customer calls dropped each month
PEAKRATIOCHANGEPCT	The percent by which the customer's peak calls to off-peak calls ratio has changed from last month to this month
AVGRECEIVEDMINS	The average number of calls received each month by the customer
AVGMINS	The average number of call minutes used by the customer each month
AVGOVERBUNDLEMINS	The average number of out-of-bundle minutes used by the customer each month
AVGROAMCALLS	The average number of roaming calls made by the customer each month
PEAKOFFPEAKRATIO	The ratio between peak and off peak calls made by the customer this month
NEWFREQUENTNUMBERS	How many new numbers the customer is frequently calling this month?

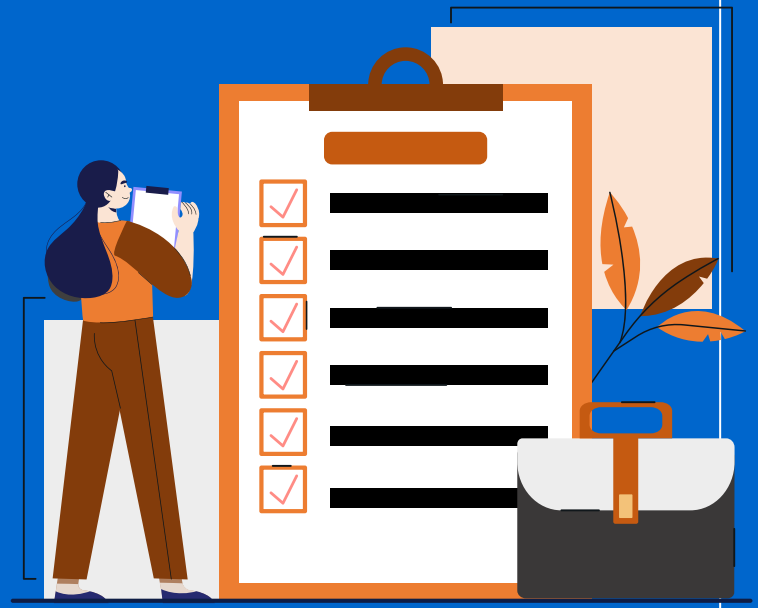
Source: Fundamentals of Machine Learning for Predictive Analytics by Kelleher, et al

Attribute descriptions (continued)

Feature	Description
CUSTOMERCARECALLS	The number of customer care calls made by the customer last month
NUMRETENTIONCALLS	The number of times the customer has been called by the retention team
NUMRETENTIONOFFERS	The number of retention offers the customer has accepted
AGE	The customer's age
CREDITRATING	The customer's credit rating
INCOME	The customer's income level
LIFETIME	The number of months the customer has been with AT
OCCUPATION	The customer's occupation
REGIONTYPE	The type of region the customer lives in
HANDSETPRICE	The price of the customer’s current handset
HANDSETAGE	The age of the customer’s current handset
NUMHANDSETS	The number of handsets the customer has had in the past 3 years
SMARTPHONE	Is the customer’s current handset a smart phone?
CHURN	The target feature

Exercise - 5 minutes

- What attributes do you think will be important in predicting churn? Why?
- For each attribute, indicate whether a high value for the attribute will increase or decrease churn?



Examine the data: Descriptive statistics

0s

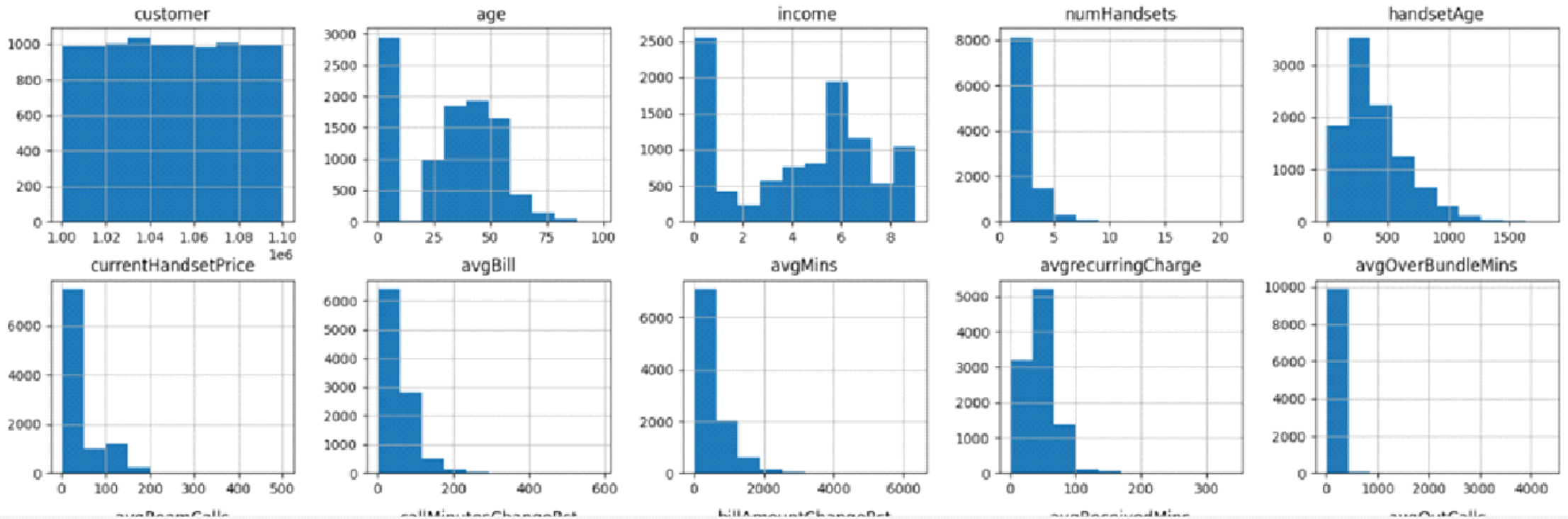
```
# Descriptive statistics
descriptive_stats = df.describe(include='all')
print(descriptive_stats)
```

```
                customer            age  occupation  regionType marriageStatus  \
count    1.000000e+04  10000.000000       10000       10000          10000
unique            NaN           NaN           8           8              3
top               NaN           NaN                                  unknown
freq              NaN           NaN        7400        4776           3920
mean     1.049974e+06     30.318400         NaN         NaN            NaN
std      2.879841e+04     22.158676         NaN         NaN            NaN
min      1.000001e+06      0.000000         NaN         NaN            NaN
25%      1.025200e+06      0.000000         NaN         NaN            NaN
50%      1.049833e+06     34.000000         NaN         NaN            NaN
75%      1.074990e+06     48.000000         NaN         NaN            NaN
max      1.099988e+06     98.000000         NaN         NaN            NaN
```

Examine the data: Histograms

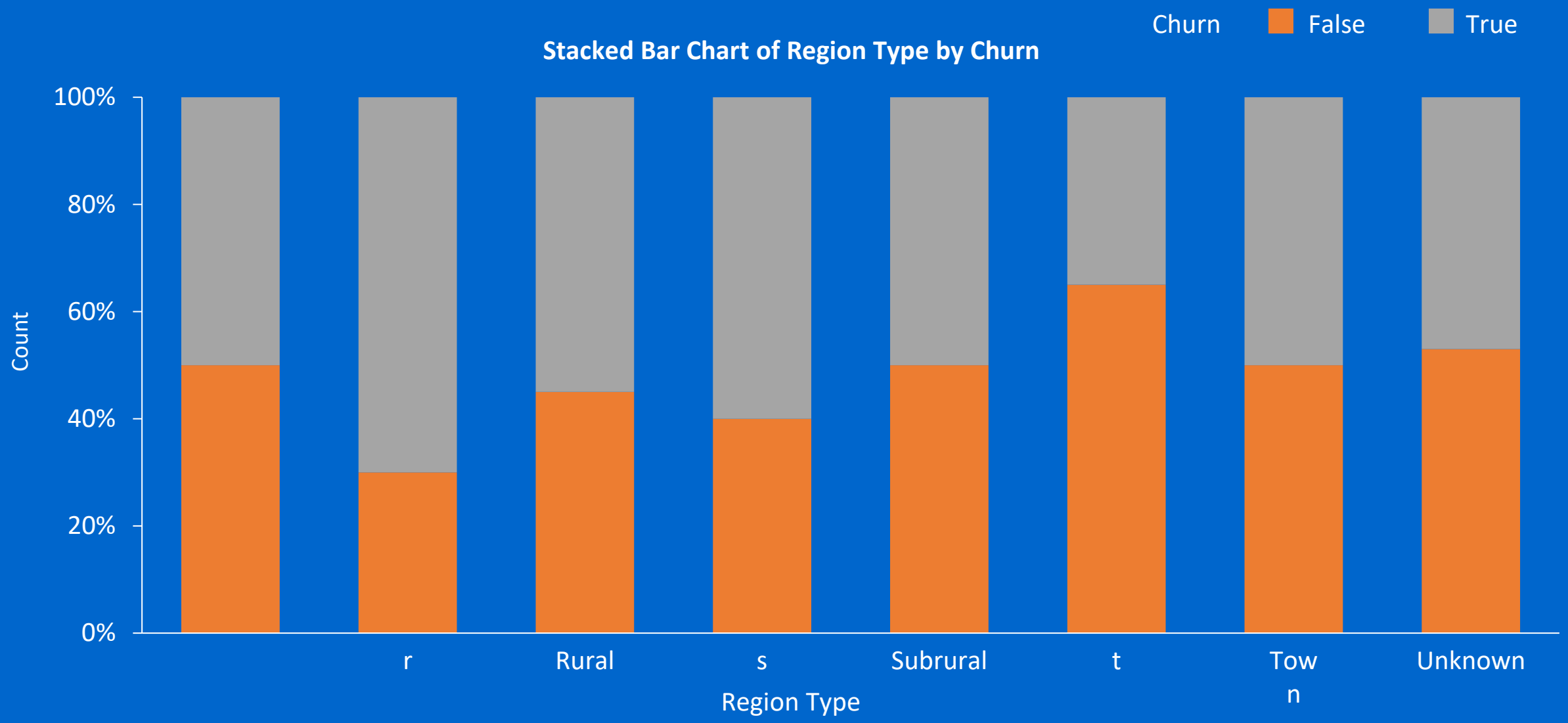
```
# Histograms for numerical attributes

import matplotlib.pyplot as plt
df.hist(figsize=(20, 16), bins=10, grid=True)
plt.show()
```

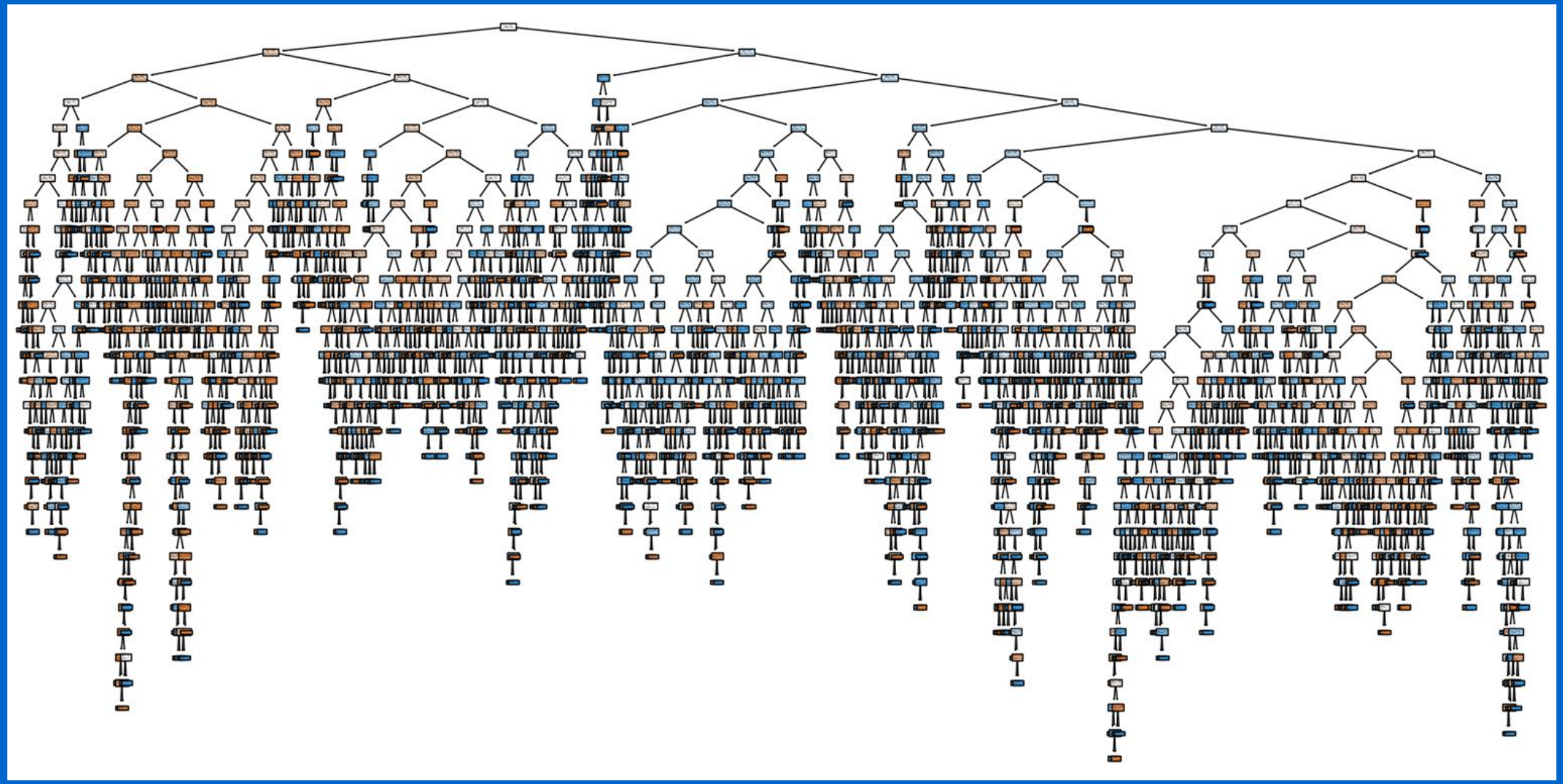


✓ 7s completed at 9:14 AM

Examine the data: Stacked RegionType bar chart by Churn

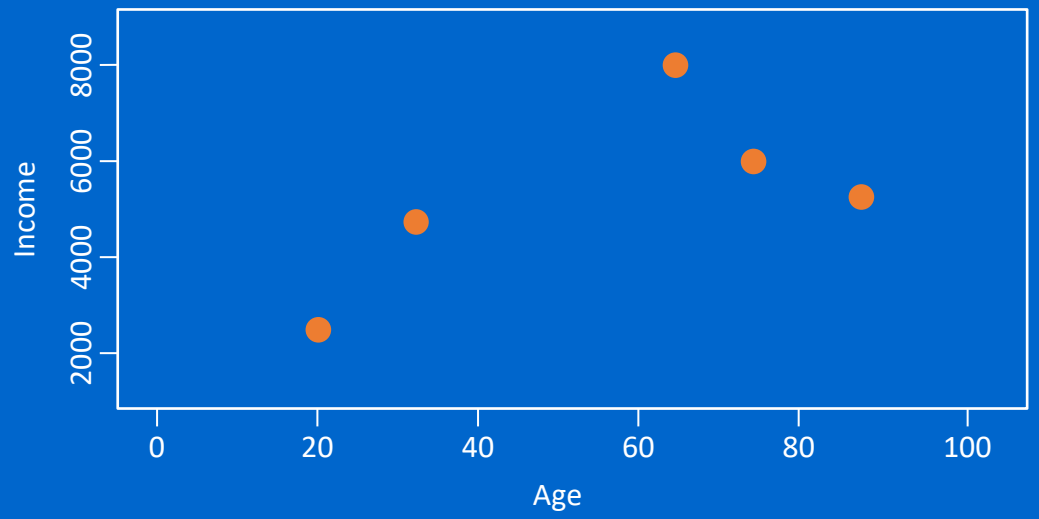


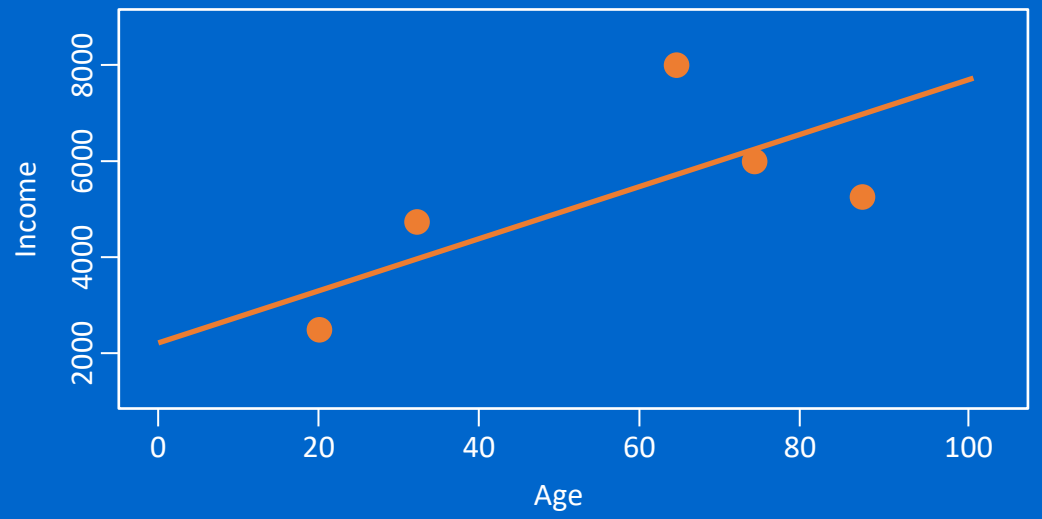
Creating a simple decision tree

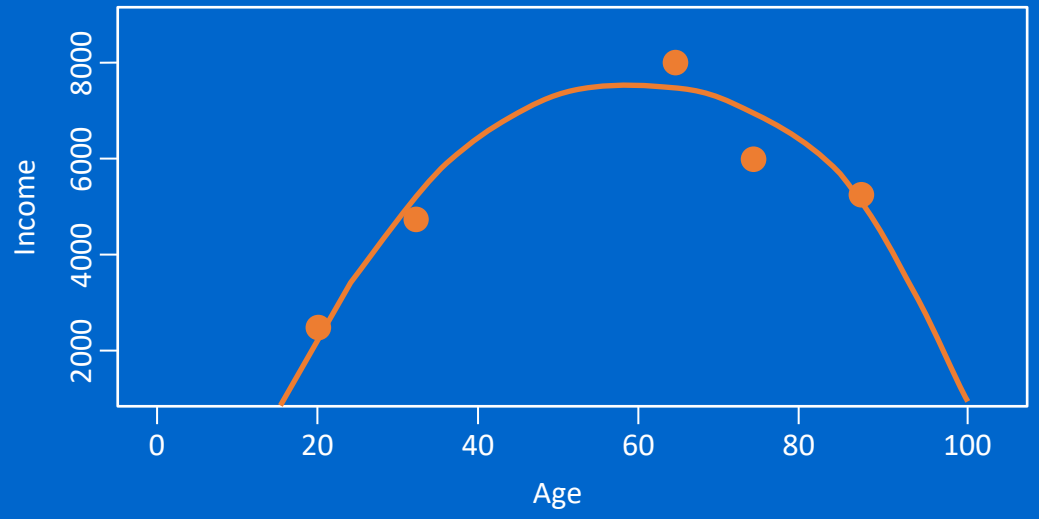


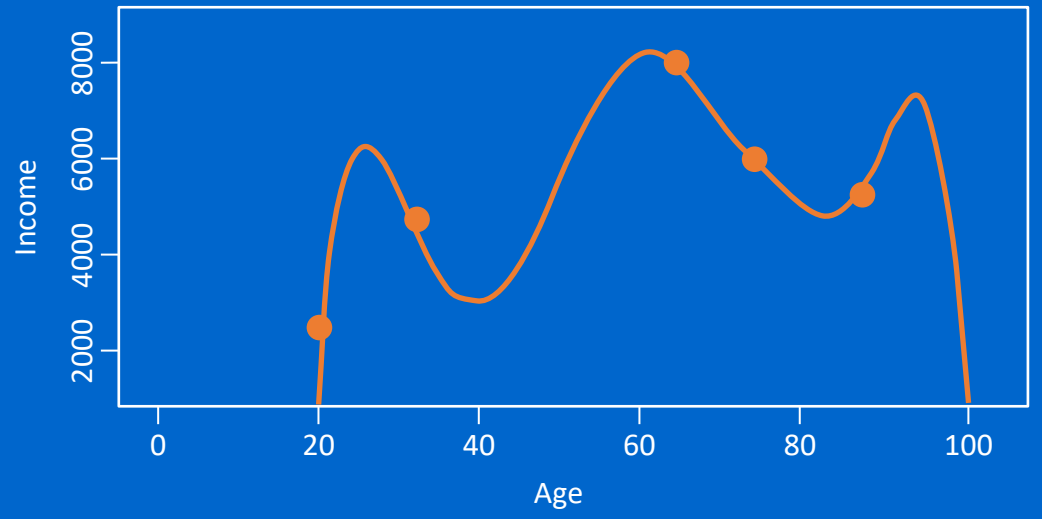
Review of key concepts

What is the best model?









Source: "Fundamentals of Machine Learning for Predictive Data Analytics" by Kelleher, et a