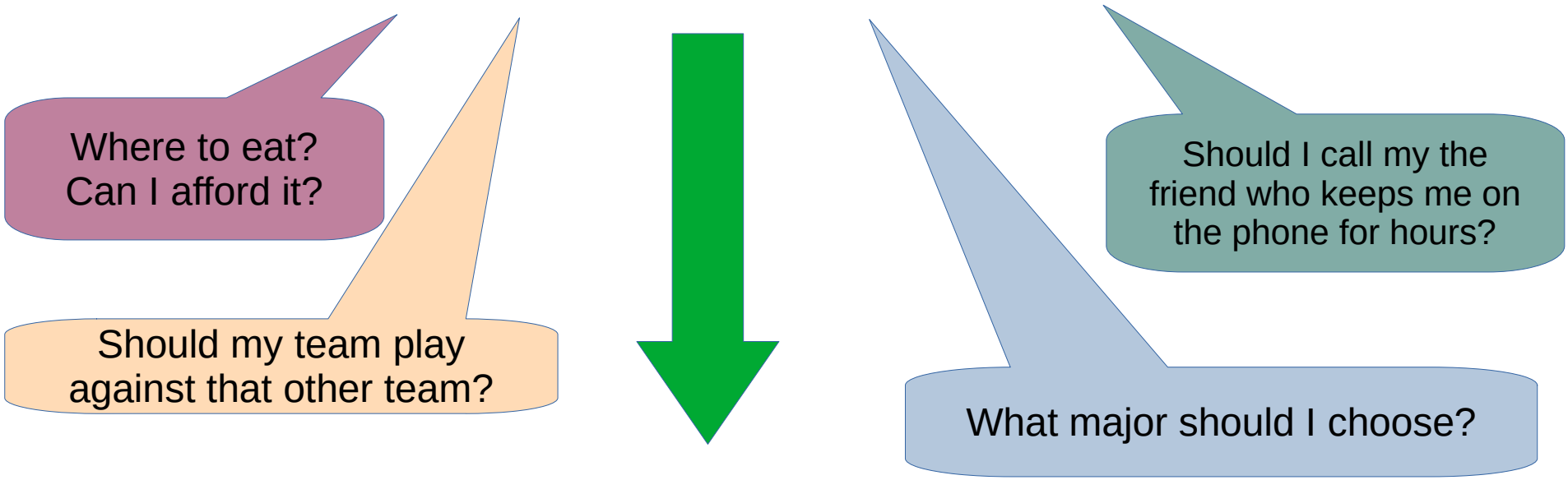
Data Science
CS301
Let’s Have A Group Project!
Activity 08

Week 11
Fall 2024
Oliver BONHAM-CARTER

Overview

Building Decision Trees

- Break into groups of two or three people to complete the project
- Build a decision tree to make some involved decision from everyday life
- Three decision nodes to split data

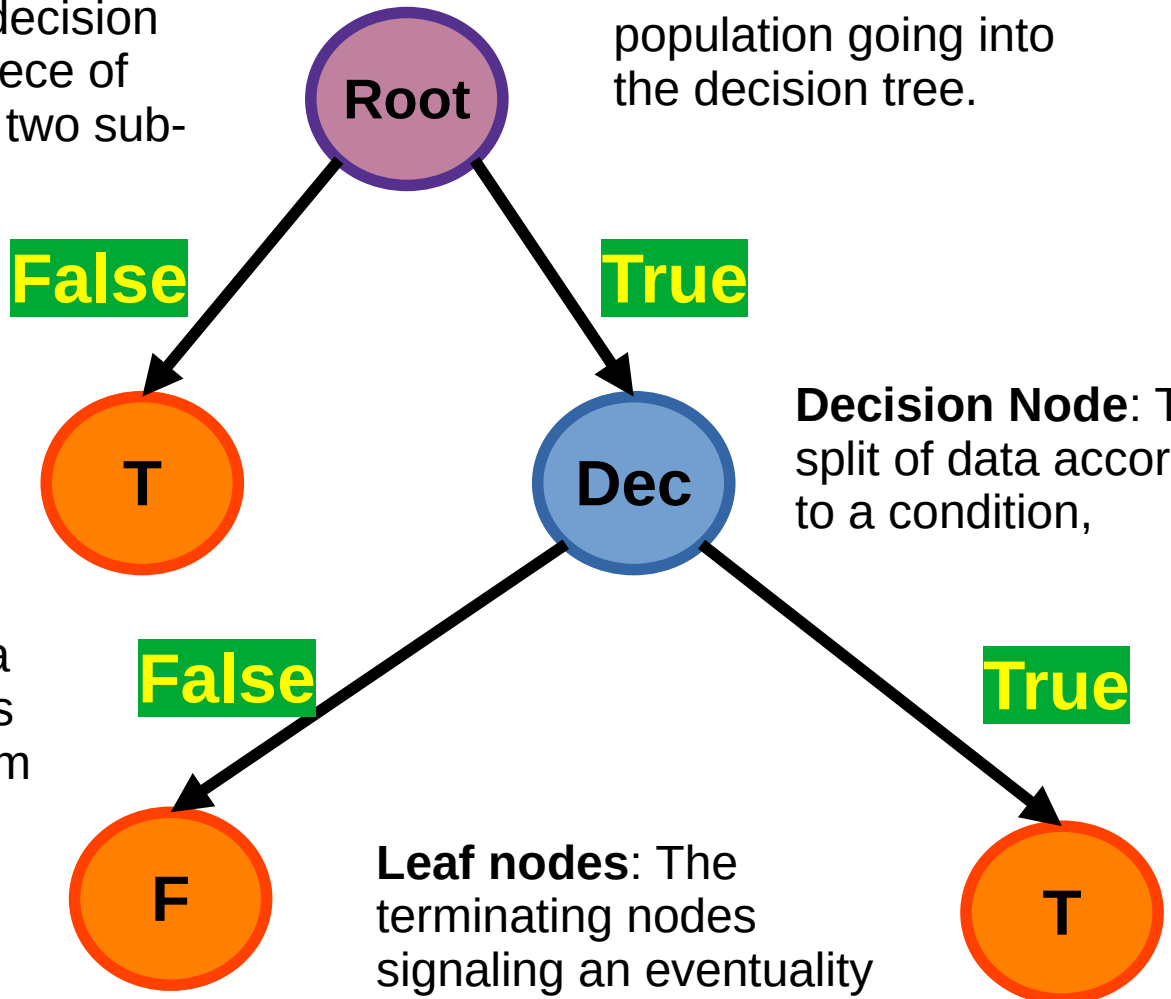


Decision Tree

Your tree must have all these parts

Node Splitting: After a decision of True or False, the a piece of data must go into one of two sub-nodes,

Root Node: The entire population going into the decision tree.



Decision Node: The split of data according to a condition,

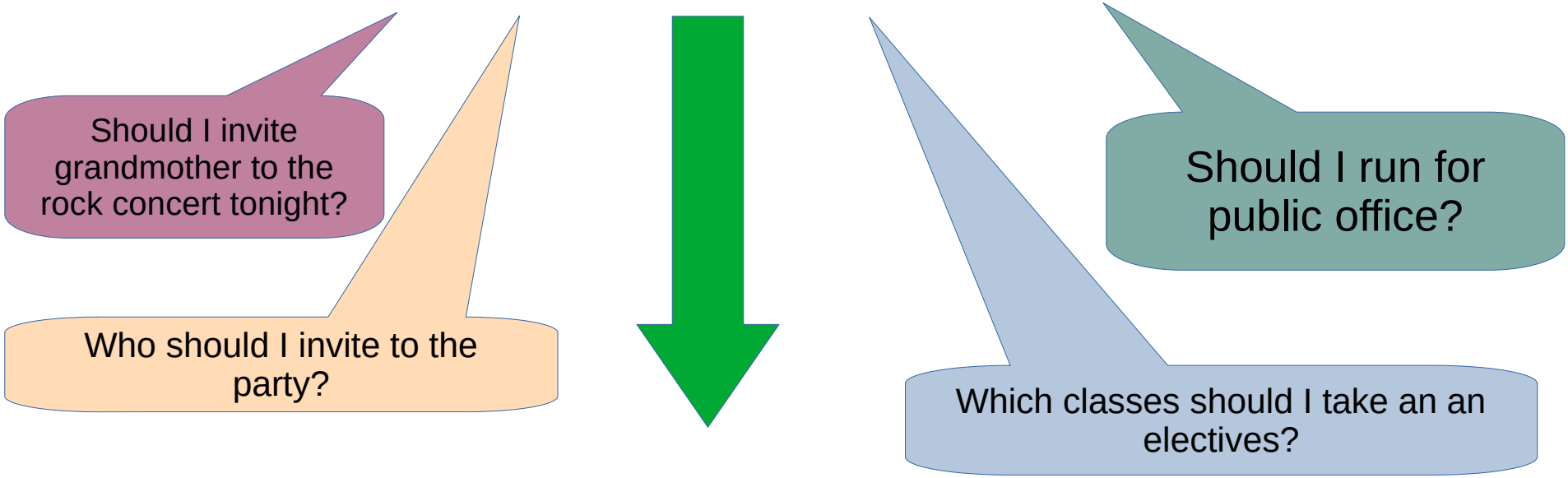
Terminal Node: A final category of data after a condition. The data has nowhere else to go from here.

Leaf nodes: The terminating nodes signaling an eventuality or taken decision.

Group Work

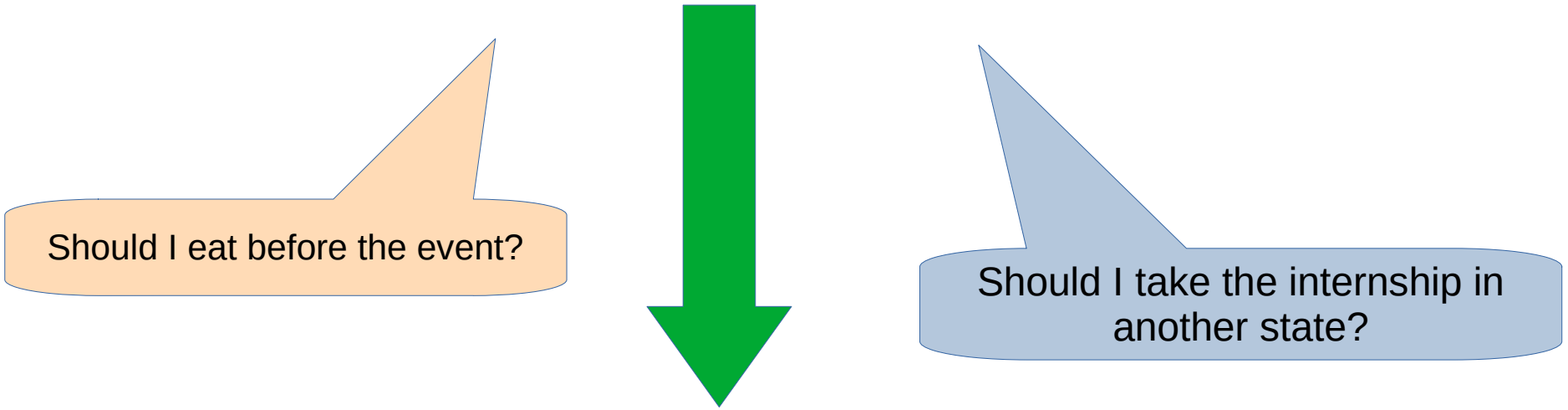
Create

- Must have the following in your work
  - A general question that will make categories of outcomes
  - Data: Create at least five (5) rows
  - Three (3) decision nodes to split the data
  - An example output of the categories made after splitting
  - Try different orders of the decision nodes to check where the categories are the most *pure* (same outcomes at each leaf). Which decision node seems better at splitting data?



Lightning Talk:
Preparation

- Work together to draw the tree for slides
- Add and label the decision nodes
- Perform the analysis of data split by the tree
- Compare orders of decision nodes in the tree
- Describe which question seems to be best at determining potential decisions. Note: not all decision nodes have same value in the tree for splitting data. Which question is best? Does your group accept the results?



Presentations on Friday

Lightning Talks of five minutes

- Come prepared on Friday to discuss the following using about five slides for a five minute lightning talk.
- *Discuss*
- *The research question to address by the decision tree*
  - *Data, Decision nodes*
  - *The output from data splitting*
  - *Comparison of different orders of the decision nodes to check for purity in categories.*
  - *General performance of the tree to make decisions*
- **No submission, your presentation is the activity.**