DATA 101 – DATA LITERACY

BAYESIAN INFERENCE

~DHRUVAL PATEL

LET'S SAY WE WANT TO CALL THOSE STUDENTS WHO LIKED BLACK AND BLUE DRESS FOR EXHIBITION AND DISTRIBUTE ICE CREAM TO THEM, BUT WE ARE CONFUSED IF WE SHOULD BRING CHOCOLATE TO VANILLA.

PRIOR

WHAT IS THE PROBABILITY THAT A DATA 101 PARTICIPANT LIKES CHOCOLATE OVER VANILLA ICE CREAM.

P(Chocolate) =0.51417

P(Vanilla)= 0.48583

Without any other information, there is a 51.42% chance that the participant will prefer chocolate, and a 48.58%chance that they'll prefer vanilla ice cream.

WHAT IF WE KNOW WHAT TYPE OF DRESS COLOR DO THEY PREFER.

P(LIKESBLUE&BLACK DRESS) = 129/247 = 0.5223

P(BLUE AND BLACK)/ VANILLA)= 67/120 =0.5583

P(BLUE AND BLACK/ CHOCOLATE)= 62/127 = 0.4882

P(VANILLA / BLUE AND BLACK) = P(BLUE AND BLACK)/ VANILLA) *P(VANILLA)/ P(LIKESBLUE&BLACKDRESS)
              = 0.5583*((0.48583)/(0.5223)) = 0.5193

P(CHOCOLATE /BLUE AND BLACK) = P(BLUE AND BLACK)/CHOCOLATE) *P(VANILLA)/ P(LIKESBLUE&BLACKDRESS)
              = 0.4882 * ((0.51417)/(0.5223)) = 0.4807

SUPPORT P(VANILLA/BLUE AND BLACK)= Percentage of participants who like blue and balck and like vanilla = 67/247 =0.2712 = 27.1%

Support P(Chocolate|BLUE AND BLACK) = Percentage of participants who like blue and balck and like Chocolate =82/247=33.1%

IF I HAVE A DATA101 CLASS STUDENTS WHO LIKES BLUE AND BLACK DRESS MORE, THEY ARE MORE LIKELY TO PREFER VANILLA OVER CHOCOLATE.

PATTERN 2

LET'S SAY THAT WE HAVE ALL THE STUDENTS OF DATA 101 GATHERED, AND WE WANT TO FIX THE THERMOSTATS BEFOREHAD BEACUSE A LOT OF TIME GETS WASTED WHEN IT IS DONE DURING THE LECTURE. BUT WE ARE UNSURE ON WHETHER SHOULD WE SET IT TO COOLER OR WARMER. IN THE SURVEY, WERE STUDENTS CHOSE OPTION THAT THEY CHANGE THE THERMOSTATS WHEN THE TEMPERATURE IS COOLER MEANS THAT THEY ARE MORE COMFORTABLE IN WARMER WEATHER. STUDENTS WHO CHOOSE WARMER PREFERED MORE COOLER TEMPERATURE.

PRIOR

WHAT IS THE PROBABILITY THAT A DATA 101 PARTICIPANT CHANGES TEMPERATURE WHEN IT IS COOLER VS WARMER.

P(COOLER) = 58.3%
P(WARMER) = 41.3%

WITHOUT ANY OTHER INFORMATION ONE CAN TELL THAT THE STUDENT FROM DATA 101 IS MORE LIKELY TO CHANGE THE THERMOSTAT WHEN IT IS COOLER(58.3%).

WHAT IF WE KNOW IF THEY CAN ROLL THEIR TONGUE.

P(CAN ROLL THEIR TONGUE) = 168/247 = 0.68
P(CAN ROLL | COOLER) = 94/144 = 0.6528
P(CAN ROLL | WARMER) = 74/102 = 0.7255
P(COOLER| CAN ROLL) = P(CAN ROLL | COOLER) *((P(COOLER) /P(CAN ROLL TONGUE)
                = 0.6528 * ((0.5830/0.68)) = 0.5597=56%

P(WARMER| CAN ROLL) = P(CAN ROLL |WARMER) *((P(WARMER) / P(CAN ROLL TONGUE)
                = 0.7255 * (0.4130/0.68)= 0.4406 = 44%

SUPPORT = P(COOLER | CAN ROLL) = Percentage of participant who can roll their tongue and changes thermostat when it is cooler = 94/246= 0.3821= 38%

SUPPORT = P(WARMER | CAN ROLL) = Percentage of participant who can roll their tongue and change thermostat when it is warmer = 74/246 = 0.3008= 30%

IF I KNEW THE DATA101 PARTICIPANT WHO CAN ROLL THEIR TONGUE, THEY ARE MOST LIKELY PREFER A COOLER ATMOSPHERE IN THE CLASS AS WELL.

3RD PATTERN (2 PIECES OF EVIDENCE)

WHAT IS I KNOW THAT THE PARTICIPANT IS IN EVEN NUMBER SECTION THAT MEETS ON THURSDAY

P(THURSDAY EVEN SECTION) = 65/247 = 0.2632
P(THURSDAY EVEN |COOLER) = 41/144 = 0.2847
P(THURSDAY EVEN| WARMER) = 24/102 = 0.2353

P(COOLER | THURSDAY EVEN) = **P(THURSDAY EVEN | COOLER)** * ((P(COOLER) / **P(THURSDAY EVEN SECTION)**
= 0.2847* (0.5830 / 0.2632) =0.6306

P(WARMER | THURSDAY EVEN) = **P(THURSDAY EVEN | WARMER)** * ((P(WARMER) / **P(THURSDAY EVEN SECTION)**
= 0.2353 * (0.4130 / 0.2632) = 0.3692

SUPPORT P (COOLER | THURSDAY EVEN) = percent of participants who are in Thursday even section and adjust the thermostats when the temperature when it is cooler = 41/ 246 = 0.1667 = 16.7 %

SUPPORT P (WARMER| THURSDAY EVEN) = percent of participants who are in Thursday even section and adjust the thermostats when the temperature when it is warmer = 24/ 246 = 0.0976 = 9.7%

IF WE JUST HAVE STUDENTS FROM EVEN SECTION THAT COMES ON THURSDAYS. THEY ARE MORE LIKELY (63 %) PREFER TO CHANGE THE THERMOSTATS WHEN IT IS COOLER. SO WE CAN INCREASE THE TEMPERATURE OF THE THERMOSTATS BEFOREHAND.

HOW CAN WE USE THIS INFORMATION

If I know the information about the Data 101 participant, I can use that for selecting the most like ice cream flavor.
Without prior information I assumed it was chocolate (51.42%). But when I subsetted the students and only selected those students who like blue and black dress. They are most likely gonna prefer Vanilla flavored ice cream then the chocolate one.

I further wanted to adjust the thermostats of the class full of Data 101 students. Without any prior knowledge, I thought that most students preferred to changed their household thermostats when it is cold.

If I had students out of all data 101 students that are able to roll their tongue, I figured out that in the class of students which can roll tongue will also also like to change the thermostats when it is cooler. (56%)

I further tested students choice when they belonged from even section on Thursdays, those student will also most likely adjust the thermostats when it is cooler and thus during their class I can set the temperature before hand.

THANKS