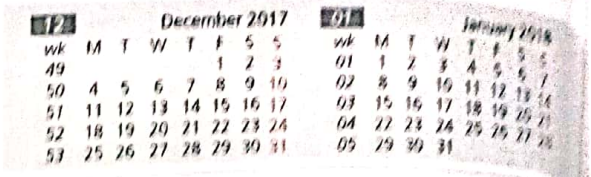

Q.1) We can apply Naive Bayesian classifier to map input tuples into accurate classes

Hence, with the help of the dataset we need to find all the prior & posterior probabilities.

For prior probabilities →

P(On time) = 14/20 ... (From dataset)

P(Late) = 2/20

P(Very late) = 3/20

P(Cancelled) = 1/20

Calculating posterior probabilities → for attribute 'Day'.

P(weekday | On Time) = 9/14

P(weekday | Very late) = 3/3

P(weekday | Late) = 1/2

P(Weekday | Cancelled) = 0/1

Similarly calculating and tabulating posterior probabilities for all attributes

Day	Class: On Time	Late	Very Late	Cancelled
Weekday	9/14	1/2	3/3	0/1
Saturday	2/14	0/2	0/3	1/1
Sunday	1/14	0/2	0/3	0/1
Holiday	2/14	1/2	0/3	0/1

For attribute 'Season' -

Season	Class: On Time	Late	Very Late	Cancelled
Spring	4/14	0/2	0/3	1/1
Summer	6/14	0/2	0/3	0/1
Autumn	2/14	0/2	1/3	0/1
Winter	2/14	2/2	2/3	0/1

For attribute 'Fog' -

Fog	Class: On Time	Late	Very Late	Cancelled
None	5/14	0/2	0/3	0/1
High	4/14	1/2	1/3	1/1
Normal	5/14	1/2	2/3	0/1

For attribute 'Rain' -

Rain	Class: On Time	Late	Very Late	Cancelled
None	6/14	1/2	1/3	0/1
Slight	6/14	1/2	0/3	0/1
Heavy	2/14	0/2	2/3	1/1

For the instance <weekday, winter, High, None>

$P_{NB}(\text{onTime}) = P(\text{onTime}) \times P(\text{weekday} \mid \text{onTime}) \times P(\text{winter} \mid \text{onTime}) \times P(\text{High} \mid \text{onTime}) \times P(\text{None} \mid \text{onTime})$

$= \frac{14}{20} \times \frac{9}{14} \times \frac{2}{14} \times \frac{4}{14} \times \frac{6}{14} = 0.0079$

Similarly,

$P_{NB}(\text{Late}) = \frac{2}{20} \times \frac{1}{2} \times \frac{2}{2} \times \frac{1}{2} \times \frac{1}{2} = 0.0125$

$P_{NB}(\text{Very Late}) = \frac{3}{20} \times \frac{3}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} = 0.0111$

$P_{NB}(\text{Cancelled}) = \frac{1}{20} \times \frac{0}{1} \times \frac{0}{1} \times \frac{1}{1} \times \frac{0}{1} = 0$

$P_{NB}(\text{Late})$ is highest, hence correct classification is 'Late'.

Similarly, we can classify any input tuple into accurate class

Q.2) Here, we have to test hypothesis that genders and preferred reading are independent, that is there is no correlation between them.

We can use, χ^2 correlation test with contingency table of size 2 x 2 (given) and $(2-1) \times (2-1)$ degrees of freedom.

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(O_{ij} - e_{ij})^2}{e_{ij}}$$

O_{ij} → observed frequency
e_{ij} → expected frequency.

$$\therefore \chi^2 = \frac{(250-90)^2}{90} + \frac{(50-210)^2}{210} + \frac{(200-360)^2}{360} + \frac{(1000-840)^2}{840} = 507.9365$$

For 1 degree of freedom, the χ^2 value needed to reject the hypothesis at 0.01 significance level is 6.635.
Since our computed value is above this, we can reject that gender and preferred reading are independent and conclude that the two attributes are correlated for the given group of people.