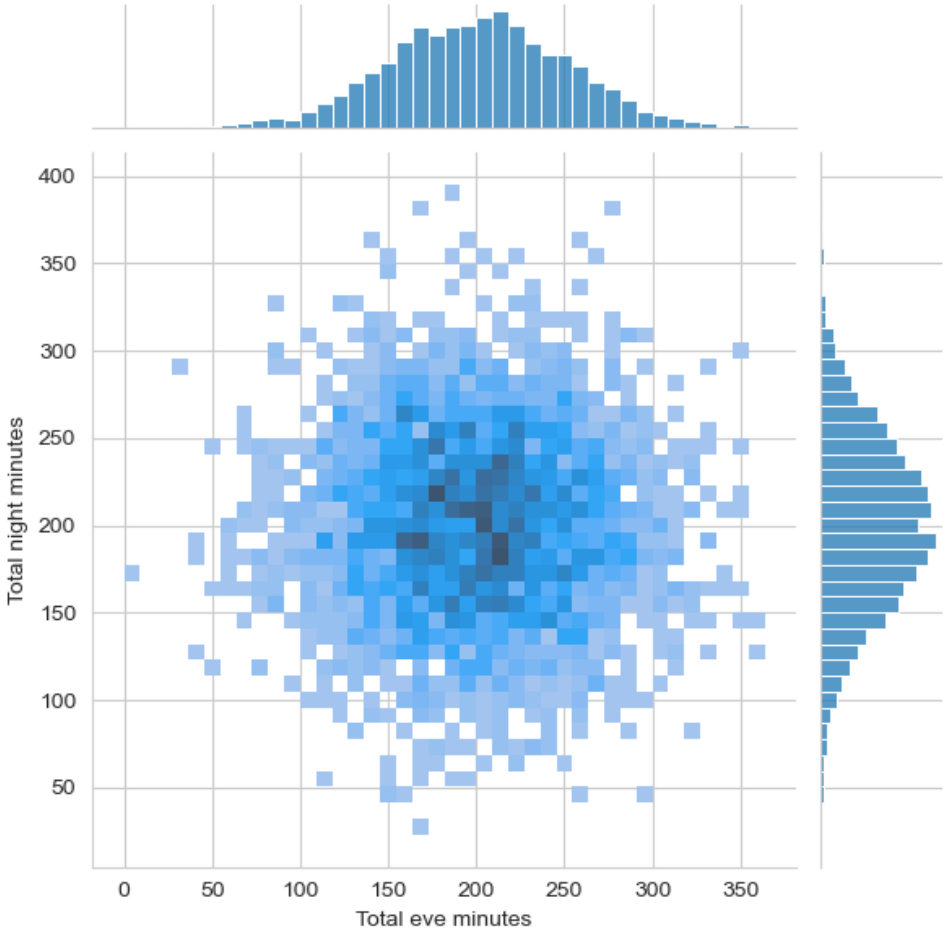
PA - Mid-semester answer sheet

		Predicted condition	
	Total population = P + N	Positive (PP)	Negative (PN)
Actual condition	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

1	Prevalence $\frac{P}{P+N}$
2	$\mathrm{MCC} = \sqrt{PPV \times TPR \times TNR \times NPV} - \sqrt{FDR \times FNR \times FPR \times FOR}$
3	$\frac{2TP}{2TP + FP + FN}$
4	Positive likelihood ratio (LR+) $= \frac{TPR}{FPR}$ Negative likelihood ratio (LR−) $= \frac{FNR}{TNR}$ Diagnostic odds ratio (DOR) $= \frac{LR+}{LR-}$
5	False discovery rate (FDR) $= \frac{FP}{PP} = 1 - PPV$
6	2x2
7	False
8	-1 -0.8728716 0 -0.2182179 1 1.0910895

9	True
10	1.0000000 0.9819805 0.9819805 1.0000000
11	Both are same after scaling. X1 shows more variance before scaling as compared to X2
12	Eigen values = 1.98198051 0.01801949 Eigen vectors 0.7071068 -0.7071068 0.7071068 0.7071068
13	0.990990253 0.009009747
14	False
15	False
16	from sklearn.preprocessing import StandardScaler
17	1. PCA methods i. Eigen-value criteria ii. Scree plot iii. Explained variance/covariance 2. PCA applications i. Dimensionality reduction ii. Model building
18	1. Manhattan distance 2. Euclidean distance
19	Merit: It scales the distribution to visualize the proportions Demerit: Scaling transforms the data and changes the original values
20	False - To remove a nearly unary variable is not a mistake
21	True - It is a mistake to not remove (omit)
22	True sns.jointplot(data=data, x='Total eve minutes', y='Total night minutes', kind='hist') plt.show()

	
23	True
24	False
25	False