TASK 1

Harshath Muruganantham

QUESTION 1

1. Generative AI Systems
2. Recommendation Systems
3. Forecasting
4. Risk Prediction

QUESTION 2

1. $\Pr(W=1) = \Pr(W=1, H=0, P=0) + \Pr(W=1, H=0, P=1) + \Pr(W=1, H=1, P=1)$

$$= 0.235 + 0.117 + 0.058 + 0.178$$
$$= 0.588$$

2. $\Pr(W=1|P=0) = \frac{\Pr(W=1,P=0,H=0)+\Pr(W=1,P=0,H=1)}{\Pr(P=0,W=0,H=0)+\Pr(P=0,W=1,H=0)+\Pr(P=0,W=0,H=1)+\Pr(P=0,W=1,H=1)}$

$$= \frac{0.235 + 0.058}{0.176 + 0.235 + 0.117 + 0.058}$$
$$= 0.500$$

3. $\Pr(W=1|\, P=1) = \frac{\Pr(W=1,P=1,H=0)+\Pr(W=1,P=1,H=1)}{\Pr(P=1,W=0,H=0)+\Pr(P=1,W=1,H=0)+\Pr(P=1,W=0,H=1)+\Pr(P=1,W=1,H=1)}$

$$= \frac{0.117 + 0.178}{0.060 + 0.117 + 0.059 + 0.178}$$
$$= 0.7126$$

4. Yes, a team is more likely to win after winning their previous game than if they lost their previous game as $\Pr(W=1|\, P=1) > \Pr(W=1|P=0)$ as seen in part 2 and 3.
5. Let A be the probability that the team will win one or more of their next to games where they play at home and then away, given that the team has won their previous game.

$\Pr(A \geq 1) = 1 - \Pr(A=0)$

$$= 1 - \left(\frac{\Pr(W=0, P=1, H=1)}{\Pr(W=0, P=1, H=1) + \Pr(W=1, P=1, H=1)} * \frac{\Pr(W=0, P=0, H=0)}{\Pr(W=0, P=0, H=0) + \Pr(W=1, P=0, H=0)}\right)$$

$$= 1 - \left(\frac{0.059}{0.059 + 0.178} * \frac{0.176}{0.176 + 0.235}\right) = 0.8934$$

QUESTION 3

1. $\mathbb{E}[S] = \mathbb{E}[2X_1 - Y_1]$

$$= \mathbb{E}[2X_1] - \mathbb{E}[Y_1]$$
$$= 2 * \mathbb{E}[X_1] - \mathbb{E}[Y_1]$$
$$= 2 * \left(1 * \frac{1}{6} + 2 * \frac{1}{6} + 3 * \frac{1}{6} + 4 * \frac{1}{6} + 5 * \frac{1}{6} + 6 * \frac{1}{6}\right) - \left(1 * \frac{1}{4} + 2 * \frac{1}{4} + 3 * \frac{1}{4} + 4 * \frac{1}{4}\right)$$
$$= 2 * 3.5 - 2.5 = 4.5$$

2. $\mathbb{V}[S] = \mathbb{V}[2X_1 - Y_1]$
$= \mathbb{V}[2X_1] + \mathbb{V}[Y_1]$
$= 2^2 * \mathbb{V}[X_1] + \mathbb{V}[Y_1]$
$= 4 * \left((1-3.5)^2 * \frac{1}{6} + (2-3.5)^2 * \frac{1}{6} + (3-3.5)^2 * \frac{1}{6} + (4-3.5)^2 * \frac{1}{6} + (5-3.5)^2 * \frac{1}{6} + (6-3.5)^2 * \frac{1}{6}\right) + \left((1-2.5)^2 * \frac{1}{4} + (2-2.5)^2 * \frac{1}{4} + (3-2.5)^2 * \frac{1}{4} + (4-2.5)^2 * \frac{1}{4}\right)$
$= 12.9167$

3.

$\Pr(S = s)$	-2	-1	0	1	2	3	4
$\Pr(s)$	$\frac{1}{6} * \frac{1}{4} = \frac{1}{24}$	$\frac{1}{6} * \frac{1}{4} = \frac{1}{24}$	$\frac{1}{6} * \frac{1}{4} = \frac{1}{24}$	$(\frac{1}{6} * \frac{1}{4}) + (\frac{1}{6} + \frac{1}{4}) = \frac{1}{12}$	$(\frac{1}{6} * \frac{1}{4}) + (\frac{1}{6} + \frac{1}{4}) = \frac{1}{12}$	$(\frac{1}{6} * \frac{1}{4}) + (\frac{1}{6} + \frac{1}{4}) = \frac{1}{12}$	$(\frac{1}{6} * \frac{1}{4}) + (\frac{1}{6} + \frac{1}{4}) = \frac{1}{12}$
$\Pr(S = s)$	5	6	7	8	9	10	11
$\Pr(s)$	$(\frac{1}{6} * \frac{1}{4}) + (\frac{1}{6} + \frac{1}{4}) = \frac{1}{12}$	$(\frac{1}{6} * \frac{1}{4}) + (\frac{1}{6} + \frac{1}{4}) = \frac{1}{12}$	$(\frac{1}{6} * \frac{1}{4}) + (\frac{1}{6} + \frac{1}{4}) = \frac{1}{12}$	$(\frac{1}{6} * \frac{1}{4}) + (\frac{1}{6} + \frac{1}{4}) = \frac{1}{12}$	$(\frac{1}{6} * \frac{1}{4}) + (\frac{1}{6} + \frac{1}{4}) = \frac{1}{12}$	$\frac{1}{6} * \frac{1}{4} = \frac{1}{24}$	$\frac{1}{6} * \frac{1}{4} = \frac{1}{24}$

4.

$$\mathbb{E}[S^3] = \sum_{i=-2}^{11} \left(s_i^3\right) * \Pr(s_i)$$
$$= (-2)^3 * \frac{1}{24} + (-1)^3 * \frac{1}{24} + (0)^3 * \frac{1}{24} + (1)^3 * \frac{1}{12} + (2)^3 * \frac{1}{12} + (3)^3 * \frac{1}{12} + (4)^3 * \frac{1}{12} + (5)^3 * \frac{1}{12} + (6)^3 * \frac{1}{12} + (7)^3 * \frac{1}{12} + (8)^3 * \frac{1}{12} + (9)^3 * \frac{1}{12} + (10)^3 * \frac{1}{24} + (11)^3 * \frac{1}{24}$$
$$= 265.5$$

5. Let $f(S) = S^3$, where $S = 2X_1 - Y_1$

$$\mathbb{E}[f(S)] \approx f(\mu_S) + \frac{\sigma^2}{2} * f''(\mu_S)$$

$$f(S) = S^3$$
$$f'(S) = 3S^2$$
$$f''(S) = 6S$$

$\mu_S = \mathbb{E}[S] = 4.5$ (From Part 1)
$\sigma^2 = \mathbb{V}[S] = 12.9167$ (From Part 2)

$$\mathbb{E}[S^3] = \mathbb{E}[f(S)] \approx f(4.5) + \frac{12.9167}{2} * f''(4.5)$$
$$= (4.5)^3 + \frac{12.9167}{2} * 6 * (4.5) = 265.5$$

$\mathbb{E}[S^3] \approx 265.5005$ (to 4d.p.)

6. Let $D = 2X_1 - Y_1 + 2Y_2$

$\Pr(D = d)$	0	1	2	3	4	5	6
$\Pr(d)$	$\frac{1}{6}*\frac{1}{4}*\frac{1}{4} = \frac{1}{96}$	$\frac{1}{6}*\frac{1}{4}*\frac{1}{4} = \frac{1}{96}$	$3(\frac{1}{6}*\frac{1}{4}*\frac{1}{4}) = \frac{1}{32}$	$3(\frac{1}{6}*\frac{1}{4}*\frac{1}{4}) = \frac{1}{32}$	$5(\frac{1}{6}*\frac{1}{4}*\frac{1}{4}) = \frac{5}{96}$	$5(\frac{1}{6}*\frac{1}{4}*\frac{1}{4}) = \frac{5}{96}$	$7(\frac{1}{6}*\frac{1}{4}*\frac{1}{4}) = \frac{7}{96}$

$\Pr(D = d)$	7	8	9	10	11	12	13
$\Pr(d)$	$7\left(\frac{1}{6}*\frac{1}{4}*\frac{1}{4}\right) = \frac{7}{96}$	$8\left(\frac{1}{6}*\frac{1}{4}*\frac{1}{4}\right) = \frac{1}{12}$	$8\left(\frac{1}{6}*\frac{1}{4}*\frac{1}{4}\right) = \frac{1}{12}$	$8\left(\frac{1}{6}*\frac{1}{4}*\frac{1}{4}\right) = \frac{1}{12}$	$8\left(\frac{1}{6}*\frac{1}{4}*\frac{1}{4}\right) = \frac{1}{12}$	$7(\frac{1}{6}*\frac{1}{4}*\frac{1}{4}) = \frac{7}{96}$	$7(\frac{1}{6}*\frac{1}{4}*\frac{1}{4}) = \frac{7}{96}$

$\Pr(D = d)$	14	15	16	17	18	19
$\Pr(d)$	$5(\frac{1}{6}*\frac{1}{4}*\frac{1}{4}) = \frac{5}{96}$	$5(\frac{1}{6}*\frac{1}{4}*\frac{1}{4}) = \frac{5}{96}$	$3(\frac{1}{6}*\frac{1}{4}*\frac{1}{4}) = \frac{1}{32}$	$3(\frac{1}{6}*\frac{1}{4}*\frac{1}{4}) = \frac{1}{32}$	$\frac{1}{6}*\frac{1}{4}*\frac{1}{4} = \frac{1}{96}$	$\frac{1}{6}*\frac{1}{4}*\frac{1}{4} = \frac{1}{96}$

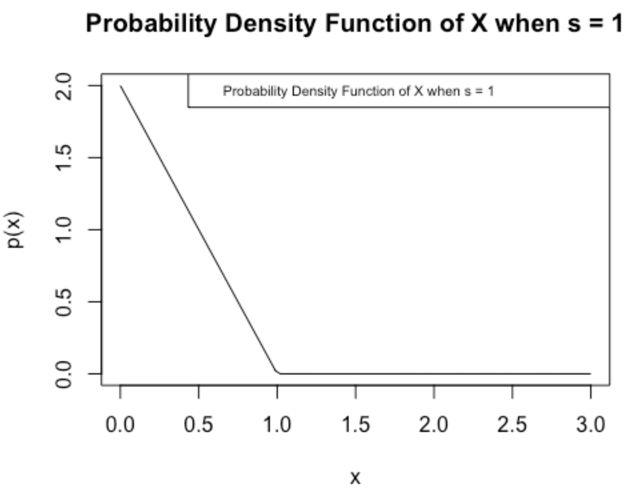
$$\mathbb{E}[(2X_1 - Y_1 + 2Y_2)^2] = \mathbb{E}[D^2] = \sum_{i=0}^{19} (d_i^2) * \Pr(d_i)$$

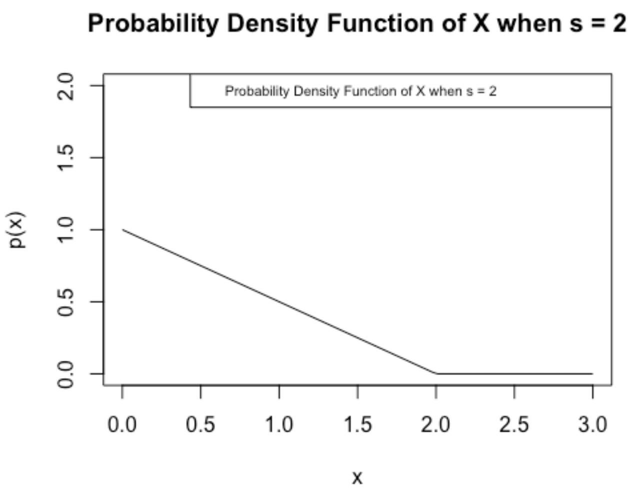
$$= (0)^2 * \frac{1}{96} + (1)^2 * \frac{1}{96} + (2)^2 * \frac{1}{32} + (3)^2 * \frac{1}{32} + (4)^2 * \frac{5}{96} + (5)^2 * \frac{5}{96} + (6)^2 * \frac{7}{96} + (7)^2 * \frac{7}{96} + (8)^2 * \frac{1}{12} + (9)^2 * \frac{1}{12} + (10)^2 * \frac{1}{12} + (11)^2 * \frac{1}{12} + (12)^2 * \frac{7}{96} + (13)^2 * \frac{7}{96} + (14)^2 * \frac{5}{96} + (15)^2 * \frac{5}{96} + (16)^2 * \frac{1}{32} + (17)^2 * \frac{1}{32} + (18)^2 * \frac{1}{96} + (19)^2 * \frac{1}{96} + (2)^2$$
$$= 108.1667$$

$$\mathbb{E}[(2X_1 - Y_1 + 2Y_2)^2] = 108.1667$$

QUESTION 4

1.





2.

$$\mathbb{E}[X] = \int_0^s x * p(x)\ dx$$
$$= \int_0^s x * \frac{2(s-x)}{s^2}\ dx$$
$$= \left[\frac{-2x^3}{3s^2} + \frac{x^2}{s}\right]_0^s$$
$$\mathbb{E}[X] = \frac{s}{3}$$

3.

$$\mathbb{E}\left[\sqrt{X}\right] = \int_0^s \sqrt{x} * p(x)\ dx$$
$$= \int_0^s \sqrt{x} * \frac{2(s-x)}{s^2}\ dx$$
$$= \left[\frac{-12 * x^{\frac{5}{2}} + 20 * s * x^{\frac{3}{2}}}{15 * s^2}\right]_0^s$$
$$\mathbb{E}\left[\sqrt{X}\right] = \frac{8\sqrt{s}}{15}$$

4.

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$
$$= \int_0^s x^2 * p(x)\ dx - \left(\int_0^s x * p(x)\ dx\right)^2$$
$$= \int_0^s x^2 * \frac{2(s-x)}{s^2}\ dx - \left(\int_0^s x * \frac{2(s-x)}{s^2}\ dx\right)^2$$
$$= \frac{s^2}{6} - \frac{s^2}{9}$$
$$\mathbb{V}[X] = \frac{s^2}{18}$$

5. Let m = median of X

$$\int_0^m \frac{2(s-x)}{s^2}\ dx = \frac{1}{2}$$
$$\int_0^m \frac{2s-2x}{s^2}\ dx = \frac{1}{2}$$
$$\int_0^m \frac{2}{s}\ dx - \int_0^m \frac{2x}{s^2}\ dx = \frac{1}{2}$$
$$\frac{2}{s}\int_0^m 1\, dx - \frac{2}{s^2}\int_0^m x\ dx = \frac{1}{2}$$
$$\frac{2}{s}[x]_0^m - \frac{2}{s^2}\left[\frac{1}{2}x^2\right]_0^m = \frac{1}{2}$$
$$\frac{2}{s}(m) - \frac{2}{s^2}\left(\frac{m^2}{2}\right) = \frac{1}{2}$$
$$\frac{2m}{s} - \frac{m^2}{s^2} = \frac{1}{2}$$
$$\frac{s^2(2m) - s(m^2)}{s^3} = \frac{1}{2}$$
$$\frac{s(2ms - m^2)}{s^3} = \frac{1}{2}$$
$$\frac{(2ms - m^2)}{s^2} = \frac{1}{2}$$

$$\frac{m(2s-m)}{s^2} = \frac{1}{2}$$
$$m(2s-m) = \frac{s^2}{2}$$
$$m = s \pm \frac{\sqrt{2}s}{2}$$

When $s = 2$, test m. m should be within the bounds $(0, s)$ as per requirement.

$$m = 2 + \frac{\sqrt{2} * 2}{2} \approx 3.4142$$
$$m = 2 - \frac{\sqrt{2} * 2}{2} \approx 0.5858$$

Therefore, reject $s + \frac{\sqrt{2}s}{2}$, as m is not within the acceptable bounds ($m \in (0,2)$) when $s = 2$.

Hence, the median of X is $s - \frac{\sqrt{2}s}{2}$.

QUESTION 5

1. For maximum likelihood in a Poisson distribution, $\lambda =$ sample mean.

```
covid = read.csv("covid.2023.csv", header = TRUE)
ml = sum(covid$Days)/length(covid$Days)
print(ml)
```

```
> covid = read.csv("covid.2023.csv", header = TRUE)
> ml = sum(covid$Days)/length(covid$Days)
> print(ml)
[1] 15.556
```

$$\lambda^{\wedge} = 15.556$$

2.

a.

```
prob52 = ppois(10,ml)
print(prob52)
```

```
> prob52 = ppois(10,ml)
> print(prob52)
[1] 0.0938464
```

$$\Pr(recover\ in\ 10\ or\ less\ days) = 0.0938$$

b.

```
x = min(covid$Days):max(covid$Days)
y = dpois(x,ml)
likelydays = data.frame(days = x, likelihood = y)
likelydays = likelydays[order(likelydays$likelihood,decreasing = TRUE),]
likelydays[1:3,]
```

```
> x = min(covid$Days):max(covid$Days)
> y = dpois(x,ml)
> likelydays = data.frame(days = x, likelihood = y)
> likelydays = likelydays[order(likelydays$likelihood,decreasing = TRUE),]
> likelydays[1:3,]
   days likelihood
15   15 0.10141086
16   16 0.09859670
14   14 0.09778625
```

The three most likely number of days it takes a patient to recover are 15, 16 and 14 from most likely to least likely.

c.

```
#Recall that ml is the variable created in part 2a.
#The value of ml is 15.556
ppois(80,ml*5) - ppois(59,ml*5)
```

```
> #Recall that ml is the variable created in part 2a.
> #The value of ml is 15.556
> ppois(80,ml*5) - ppois(59,ml*5)
[1] 0.6115397
```

Therefore, the probability that five individuals (who have COVIS) take a combined total of between 60 – 80 days (inclusive) between them to recover is 0.6115.

d.

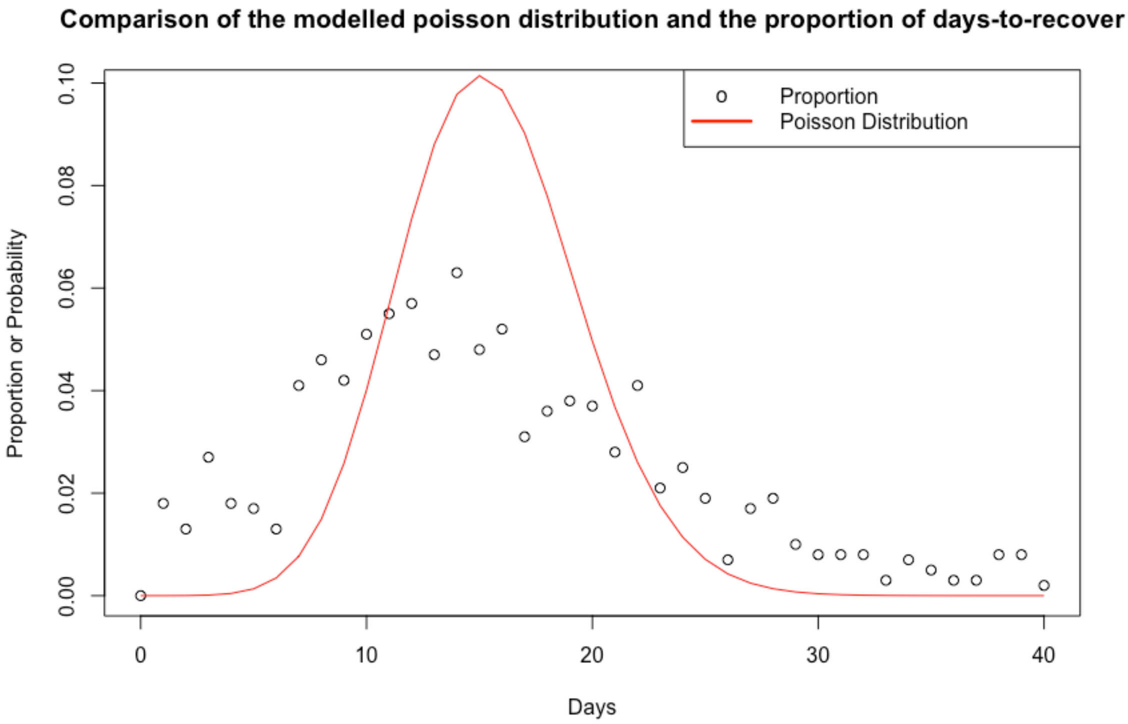
```
#Model this question as a binomial distribution question where the
#probability of success is that 'One Person recovers on or after day 14'.
#Then find the probability for the probability of 3, 4 and 5 successes in 5 trials.

#Finding the probability for one person to recover on or after day 14.
PronePersonToRecover = 1- ppois(13, ml)

#Finding the probability that 3 or more of 5 patients will recover on or after day 14
sum(dbinom(c(3,4,5), 5, PronePersonToRecover))
```

```
> #Model this question as a binomial distribution question where the
> #probability of success is that 'One Person recovers on or after day 14'.
> #Then find the probability for the probability of 3, 4 and 5 successes in 5 trials.
>
> #Finding the probability for one person to recover on or after day 14.
> PronePersonToRecover = 1- ppois(13, ml)
>
> #Finding the probability that 3 or more of 5 patients will recover on or after day 14
> sum(dbinom(c(3,4,5), 5, PronePersonToRecover))
[1] 0.8205065
```

Therefore, the probability that 3 or more of 5 patients will recover on or after day 14 is 0.8205.

3.

```
proportions = c()
for (days in 0:40){
 proportions = append(proportions, (length(which(covid$Days == days)))/length(covid$Days))
}
dataSet_Proportions = data.frame(Days = 0:40, Proportions = proportions)
plot(dataSet_Proportions$Days, dataSet_Proportions$Proportion, xlim = c(0,40), ylim = c(0,dpois(round(ml),ml)), xlab = "Days", ylab = "Proportion or Probability")
lines(dataSet_Proportions$Days, dpois(dataSet_Proportions$Days,ml), col = "red")
title("Comparison of the modelled poisson distribution and the proportion of days-to-recover")
legend("topright",c("Proportion","Poisson Distribution"), lty=c(0,1), pch=c("o",""), col=c("black","red"), lwd=c(1,2.5))
```



Based on the above graph, I do not believe this model is a good fit for the data. This is primarily because, the shape of the proportion plots, do not seem correlate or lie on the model largely. One reason for this could be due to the fact that Poisson distribution is modelled through the key characteristic of a constant rate, which does not seem to exist in the COVID data.