Question 1

1. Forecasting

   Forecasting can be used to collect this week's banana purchases, predict the next week's banana purchases.

2. Anomaly Detection

   The system can find abnormal purchase records, stolen credit cards, abnormal signals through the card purchase history.

3. Recommendation systems

   The recommendation system can find out Netflix users' preferences based on various factors such as the category and type of shows in their viewing history. When a Netflix user's behavior pattern is discovered by the system, Netflix can find and push the shows and movies that the user is interested in according to the user's preference.

4. Risk Prediction

   Risk prediction can be used to collect all of the patients with breast cancer and the people are the body of data, analysis of breast cancer patient's lifestyle and genome with this person place to predict a person suffering from breast cancer next year, such as breast cancer patients with long-term drinking the people also like drinking and some early symptoms of breast cancer, That puts him at higher risk for breast cancer in the next year.

Question 2

1.

H \ R	R = 0	R = 1	R = 2	Total
H = 0	2	8	9	19
H = 1	5	2	12	19
Total	7	19	21	38

Win in home: probabilities of a win = $\frac{12}{38}$=0.316 as P(H=1, R=2)

P(H=1, R=0)= $\frac{5}{38}$=0.132

P(H=1, R=1)= $\frac{2}{38}$=0.053

P(H=0, R=0)= $\frac{2}{38}$=0. 053

P(H=0, R=1)= $\frac{8}{38}$=0.211

P(H=0, R=2)= $\frac{9}{38}$=0.237

Here the joint probabilities of a win/draw/loss for home/away games will be,

	R = 0	R = 1	R = 2
H = 0	0. 053	0.211	0.237
H = 1	0.132	0.053	0.316

2. P(R = 2) = P(H=1, R=2) + P(H=0, R=2) = 0.237+0.316=0.553

3. By conditional probability formula:

P(R=2| H=1) = $\frac{\mathrm{P(H=1,R=2)}}{\mathrm{P(H=1)}} = \frac{\frac{12}{38}}{\frac{12}{38}+\frac{5}{38}+\frac{2}{38}} = \frac{0.316}{0.5} = 0.632$

4. By conditional probability formula:

P(R=2| H=0) = $\frac{\mathrm{P(H=0,R=2)}}{\mathrm{P(H=1)}} = \frac{\frac{9}{38}}{\frac{12}{38}+\frac{5}{38}+\frac{2}{38}} = \frac{0.237}{0.5} = 0.474$

5. Yes. To compare P(R=2| H=1) and P(R=2| H=0), P(R=2| H=1) =0.632 and P(R=2| H=0) =0.473, P(R=2| H=1) > P(R=2| H=0) By comparing the results we can see that Barcelona win more games at home than away

6.

P(not lose 2 out of 3) =P(win at least 2)
=P(win>=2)
= 1-P(win = 1) – P(win = 0)

P(win = 1) = 2*(P(R=2 | H=0) * (1-P(R=2 | H=1))* (1-P(R=2 | H=0)))
+ ((1-P(R=2 | H = 0))*(1-P(R=2 | H = 0)) * P(R=2|H=1))
= 2*((0.474) * (1-0.632)*(1-0.474)) +((1-0.474)*(1-0.474)*0.632)
= 2*(0.0918)+ 0.1748
= 0.184 +0.1748
= 0.359

P(win = 0) = (1-P(R=2 | H=0)) * (1-P(R=2 | H=1)) * (1-P(R=2 | H=0))
= (1-0.474)*(1-0.632)*(1-0.474)
= 0.102

P(not lose 2 out of 3) = 1 – 0.359 – 0.102 = 0.539

Question 3

1.

$$P(X_1 = 1) = \frac{1}{6}$$

$$P(X_1 = 2) = \frac{1}{6}$$

$$P(X_1 = 3) = \frac{1}{6}$$

$$P(X_1 = 4) = \frac{1}{6}$$

$$P(X_1 = 5) = \frac{1}{6}$$

$$P(X_1 = 6) = \frac{1}{6}$$

$$P(Y_1 = 1) = \frac{1}{4}$$

$$P(Y_1 = 2) = \frac{1}{4}$$

$$P(Y_1 = 3) = \frac{1}{4}$$

$$P(Y_1 = 4) = \frac{1}{4}$$

$E[X_1] = 1\times(\frac{1}{6})+2\times(\frac{1}{6})+3\times(\frac{1}{6})+4\times(\frac{1}{6})+5\times(\frac{1}{6})+6\times(\frac{1}{6}) = \frac{7}{2}$

$E[Y_1] = 1\times(\frac{1}{4})+2\times(\frac{1}{4})+3\times(\frac{1}{4})+4\times(\frac{1}{4}) = \frac{5}{2}$

$E[{X_1}^2] = 1^2\times(\frac{1}{6})+2^2\times(\frac{1}{6})+3^2\times(\frac{1}{6})+4^2\times(\frac{1}{6})+5^2\times(\frac{1}{6})+6^2\times(\frac{1}{6}) = \frac{91}{6}$

$E[{Y_1}^2] = 1^2\times(\frac{1}{4})+2^2\times(\frac{1}{4})+3^2\times(\frac{1}{4})+4^2\times(\frac{1}{4}) = \frac{15}{2}$

$Var[X_1] = E[{X_1}^2]-(E[X_1])^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$

$Var[Y_1] = E[{Y_1}^2]-(E[Y_1])^2 = \frac{15}{2} - \left(\frac{5}{2}\right)^2 = \frac{5}{4}$

S = $X_1 + 3Y_1$

$V[S] = V[X_1 + 3Y_1]$

$V[S] = V[X_1] + 3^2V[Y_1]$

$V[S] = \frac{35}{12} + 9(\frac{5}{4}) = \frac{170}{12}$

$V[S] = \frac{85}{6}$

2.

3Y1 \ X1	1	2	3	4	5	6
3x1	4	5	6	7	8	9
3x2	7	8	9	10	11	12
3x3	10	11	12	13	14	15
3x4	13	14	15	16	17	18

P(S = x)= $\frac{1}{24}$, x=4, 5, 6, 16, 17, 18

= $\frac{2}{24}$, x=7, 8, 9, 10, 11, 12, 13, 14, 15

3. E($\sqrt{S}$)= $\frac{\sqrt{4}+\sqrt{5}+\sqrt{6}+\sqrt{16}+\sqrt{17}+\sqrt{18}}{24}+\frac{\sqrt{7}+\sqrt{8}+\sqrt{9}+\sqrt{10}+\sqrt{11}+\sqrt{12}+\sqrt{13}+\sqrt{14}+\sqrt{15}}{12}$

$=\frac{19.0513}{24}+\frac{29.6374}{12}$

$=\frac{78.3261}{24}$ = 3.264

4.

E[x] = $\frac{a+b}{2}$ E[s] = E[$X_1 + 3Y_1$] = $E[X_1] + 3E[Y_1]$ $= \frac{7}{2}+\frac{15}{2} = 11$

E[f(x)] = f(μx) +$\frac{\sigma^2}{2}\iint$ f(μx)

f(x) = $\sqrt{S}$ $\iint$ f(μx) = $\int \frac{1}{2}(E(s))^{\frac{-1}{2}} = \frac{-1}{4}(E(s))^{\frac{-3}{2}}$

$\sigma^2 = V[S]=\frac{85}{6}$ f(μx) = $\sqrt{\mathrm{E[x]}} = \sqrt{11}$

E[f(x)] = f(μx) +$\frac{\sigma^2}{2}\iint$ f(μx) = $\sqrt{11} + \frac{\frac{85}{6}}{2}\left(\frac{-1}{4}(11)^{\frac{-3}{2}}\right) = 3.268$

5. $E[Y_2] = E[Y_1] = \frac{5}{2}$

$E[{Y_2}^2] = E[{Y_1}^2]= \frac{15}{2}$

$Var[Y_2] = Var[Y_1]=\frac{5}{4}$

$V[X_1 + 3Y_1 - 2Y_2] = E[(X_1 + 3Y_1 - 2Y_2)^2] - E[X_1 + 3Y_1 - 2Y_2]^2$

$E(X_1 + 3Y_1 - 2Y_2)^2 = V[X_1 + 3Y_1 - 2Y_2] + E[X_1 + 3Y_1 - 2Y_2]^2$

$= V[X_1] + 9V[Y_1] + 4V[Y_2] + (E[X_1 + 3Y_1] - 2E[Y_2])^2$

$= 2.917 + 9 * 1.25 + 4 * 1.25 + (11 - 2 * 2.25)^2$

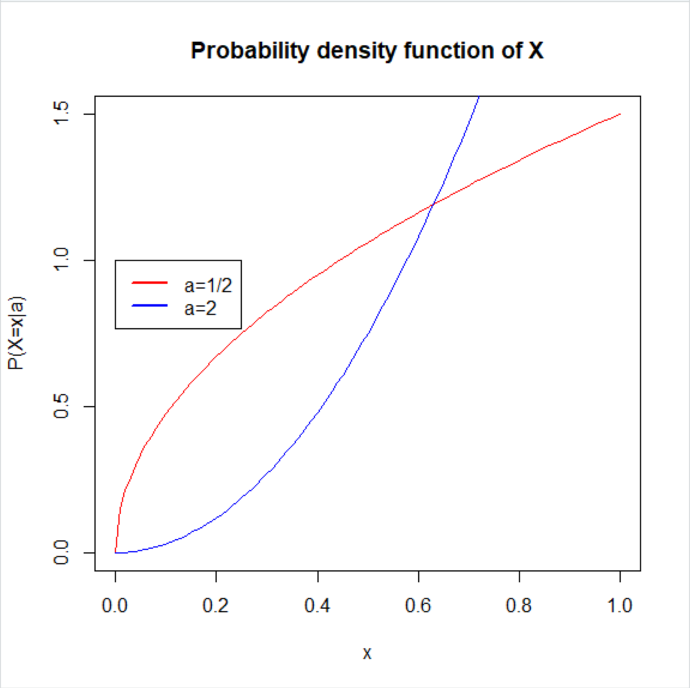
$= 19.167 + 36$

$= 55.167$

Question 4

1. The cod to plot the probability density function of X:

```
f1 <- function(x){((1/2)+1)*x^(1/2)}
f2 <- function(x){(2+1)*(x^2)}
curve(f1, from=0, to=1, xlab="x", ylab= "P(X=x|a)" col="blue", add=TRUE)
curve(f2, from=0, to=1, col="blue", add=TRUE)
legend(x=0, y=1, legend=c("a=1/2","a=2"), col=c("red","blue"), lwd=2, lty=c(1, 1))
```



2. $E[X] = \int_{-\infty}^{\infty} xf(x)dx$

$$=\int_0^1 x\big((a+1)x^a\big)dx$$

$$=\int_0^1 x(a+1)x^a dx$$

$$= (a+1)\int_0^1 xx^a dx$$

$= \ (a+1)\int_0^1 x^{1+a}dx$

$=(a+1)[\frac{x^{a+2}}{a+2}]_0^1$

$= \ (a+1)\frac{1}{a+2} \ =\frac{a+1}{a+2}$

3. $E\left[\frac{1}{X}\right] \ = \ \int_{-\infty}^{\infty}\frac{1}{x}f(x)dx$

$=\int_0^1\frac{1}{x}\left((a+1)x^a\right)dx$

$=\int_0^1\frac{1}{x}(a+1)x^a dx$

$= \ (a+1)\int_0^1\frac{1}{x}x^a dx$

$= \ (a+1)\int_0^1 x^{a-1}dx$

$=(a+1)[\frac{x^a}{a}]_0^1$

$= \ (a+1)\frac{1}{a} \ =\frac{a+1}{a}$

4. $V[X] \ = \ E[X^2]-\mu^2 = \int_{-\infty}^{\infty}x^2 f(x)dx - \mu^2$

$=\int_0^1 x^2(a+1)x^a dx - \ \mu^2$

$= \ (a+1)\int_0^1 x^2 x^a dx - \ \mu^2$

$= \ (a+1)\int_0^1 x^{a+2}dx - \ \mu^2$

$= (a+1)[\frac{x^{a+3}}{a+3}]_0^1 - \ (\frac{a+1}{a+2})^2$

$= \ (a+1)\frac{1}{a+3} \ - \ \frac{(a+1)^2}{(a+2)^2}$

$=\frac{a+1}{a+3} - \ \frac{(a+1)^2}{(a+2)^2}$

$=\frac{a+1}{a+3} - \ \frac{(a+1)^2}{(a+2)^2}$

5. As x$\in$ [0,1]

E[x] = $\int_0^x (a+1)x^a dx$

$$= [\frac{a+1}{a+1}x^{a+1}]_0^x$$

$$= x^{a+1}$$

X = 0 for x∈ $(-\infty, 0)$

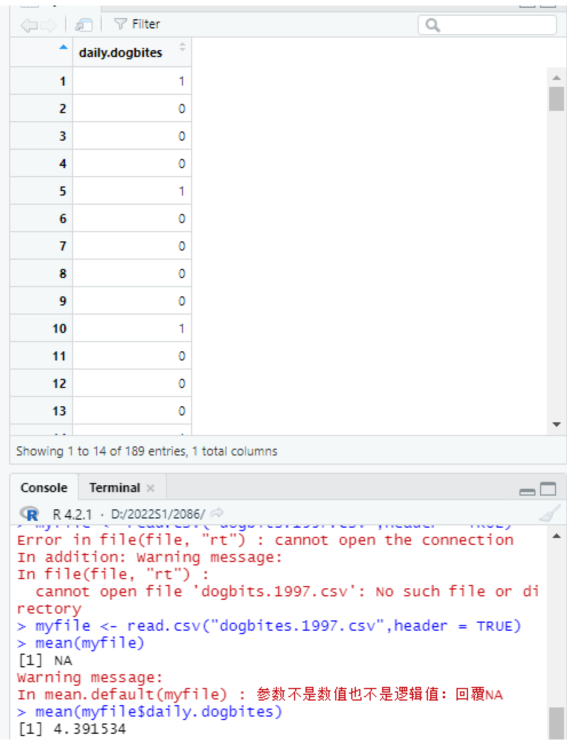
X = x^{a+1} for x∈ $[0,1]$

X = 1 for x∈ $(1, +\infty)$

Find median X when x = 1/2

x = $\sqrt[a+1]{\frac{1}{2}}$

Q5



1

Code:

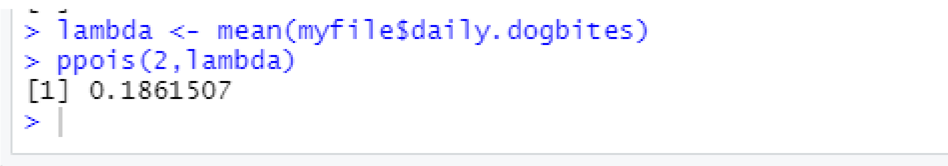
> myfile <- read.csv("dogbites.1997.csv",header = TRUE)

> mean(myfile$daily.dogbites)

Output:

[1] 4.391534

2.a



Code:

> lambda <- mean(myfile$daily.dogbites)

> ppois(2,lambda)

Output:

[1] 0.1861507

b.

```
> table(myfile)
daily.dogbites
 0  1  2  3  4  5  6  7  8  9 10 12 13 15 16 18 22
16 14 25 33 23 24 23  9  4  6  3  3  1  1  1  2  1
>
```

2 or 3 /day

The number of bites per day showed a normal distribution over six months, so the odds of getting bitten two and three times a day were highest.

c. 0.6346646

```
> ppois(32,lambda*7)
[1] 0.6346646
```

code:

> ppois(32,lambda*7)

output:

[1] 0.6346646

d. 0.5012691

```
> y <-ppois(2,lambda)
> y <- 1-ppois(2,lambda)
> pbinom(11,14,y)
[1] 0.4987309
> 1-pbinom(11,14,y)
[1] 0.5012691
>
```

code:

> y <- 1-ppois(2,lambda)

> pbinom(11,14,y)

[1] 0.4987309

> 1-pbinom(11,14,y)

[1] 0.5012691

3.