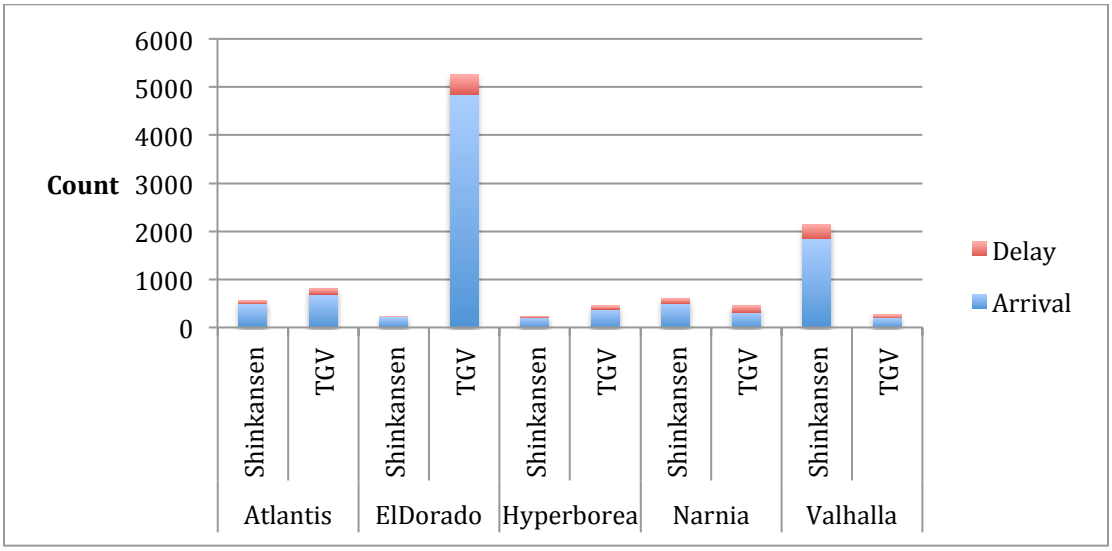
Question 1. The chart below shows the on time and delayed arrivals for two competing railway companies to five cities. What would you conclude about the relative performance of these two railway companies?

		Atlantis	El Dorado	Hyperborea	Narnia	Valhalla
SHINKANSEN	on time	497	221	212	503	1841
	delayed	62	12	20	102	305
TGV	on time	694	4840	383	320	201
	delayed	117	415	65	129	61

Given data set is discrete. We use bar graph to visualize. A bar graph provides better opportunity to comprehend the data and analyze.



Let us analyze the data by each station

Station	TGV Total Services	TGV Total On time services %	Shinkansen Total Services	Shinkansen Total On time services %
Atlantis	811	85.5	559	88.9
ElDorado	5255	92.1	233	94.8
Hyperborea	448	85.4	232	91.3
Narnia	449	71.2	605	83.1
Valhalla	262	76.1	2146	85.7

From the above table we can understand that TGV makes most trips to all stations except Narnia and Valhalla compared to Shinkansen. TGV's on time service record is almost as good as Shinkansen for all stations except Narnia.

Question 2. If you roll four dice at once, what is the probability that the same value appears on all four dice? Generalize the formula that you used for n dice.

The probability of an event is, P (E)= favorable outcomes/total outcomes
So, Let us find the total possible outcomes when throwing four dices.
(Number of outcomes per dice) $^{\text{Number of dices}}$
When throwing a single dice there are 6 possible out comes {1}{2}{3}{4}{5}{6}. When applying the above formula we get (6)^4. i.e. 6x6x6x6=1296.

Expected event is same value on all four dice. The following are the same value possibilities on all 4 dices.
{1,1,1,1},{2,2,2,2},{3,3,3,3},{4,4,4,4},{5,5,5,5},{6,6,6,6}

So, P (all same value)=6/1296

Question 3 The table below shows four small datasets, each with an x variable and y variable. What can you say about these four datasets?

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

We can say that x is independent variable and y is dependent variable. This mean that “y” value changes with respect to x value. We can plot these data points (x,y) on a graph and find out a model which closely fits in the form y=mx+b. By using this model we can predict the next possible value of y given x.

y is dependent variable, m is slope, x is independent variable and b is constant or y intercept.

Let us take the first data set and find out the regression line for prediction.

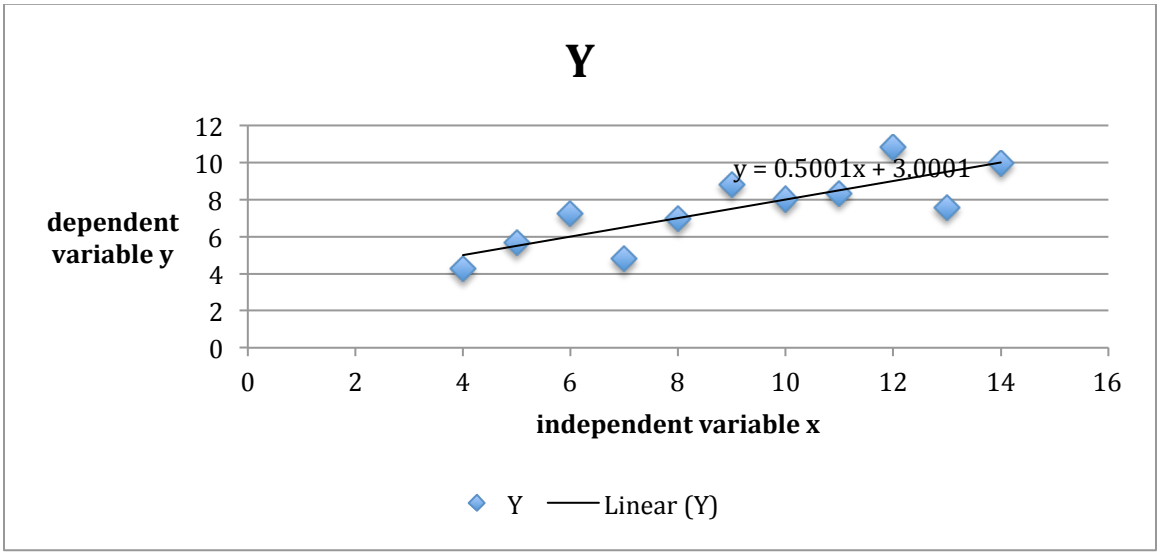
$X-\bar{X}$ is the difference between x value and the x mean
$Y-\bar{Y}$ is the difference between y value and y mean.
To find the regress line model of y= mx+b we have to find the values of b and m.

	X	Y	X-$\bar{X}$	Y-$\bar{Y}$	(X-$\bar{X}$)^2	(X-$\bar{X}$) * (Y-$\bar{Y}$)
	10	8.04	1	0.539090909	1	0.539090909
	8	6.95	-1	-0.550909091	1	0.550909091
	13	7.58	4	0.079090909	16	0.316363636
	9	8.81	0	1.309090909	0	0
	11	8.33	2	0.829090909	4	1.658181818
	14	9.96	5	2.459090909	25	12.29545455
	6	7.24	-3	-0.260909091	9	0.782727273
	4	4.26	-5	-3.240909091	25	16.20454545
	12	10.84	3	3.339090909	9	10.01727273
	7	4.82	-2	-2.680909091	4	5.361818182
	5	5.68	-4	-1.820909091	16	7.283636364
Mean	9	7.500909091			110	*55.01*

M(slope)= $\sum(X-\bar{X}) * (Y-\bar{Y})/ \sum(X-\bar{X})^2=55/110=.5$

Find b value. The regression line has to cross the point (x mean, y mean). i.e. (9,7.5). We use these values to find b
Y=7.5 x=9 , m=.5 then b is
*7.5=.5*9+b*
b=7.5-4.5=3

The best file line model is y=.5x+3. We can use this model to predict the y value for given x value.



We can find the model for the other data sets in the similar way.

Question 4 Write a program in a language of your choice to determine how many numbers from 1 to 1000 are not divisible by any of 3, 7, and 11.

```
/**
 * Find numbers from 1 to 1000 are not divisible by any of 3, 7, and 11.
 *
 * Logic:
 *		Find the numbers divisible by 3, 7 and 11 and remove form the whole list.
 *		Print the whole list.
 *
 */

package com.math;

import java.util.ArrayList;
import java.util.Hashtable;
import java.util.Iterator;
import java.util.Set;

public class FindNumbers {

	public static void main(String[] args) {
		ArrayList<Integer> divisiblebyThree = new ArrayList<Integer>();
		ArrayList<Integer> divisiblebySeven = new ArrayList<Integer>();
		ArrayList<Integer> divisiblebyEleven = new ArrayList<Integer>();

		Hashtable<Integer, Integer> wholeSet = new Hashtable<Integer, Integer>();

		// first fill the whole Set
		for (int i = 1; i <= 1000; i++)
			wholeSet.put(i, i);

		// Find the numbers divisible 3
		for (int i = 1; i <= 1000; i++) {
			if (i % 3 == 0) {
				divisiblebyThree.add(i);
				// Remove the number form wholeset
				if (wholeSet.containsKey(i)) {
					wholeSet.remove(i);
				}
			}
		}

		// Find the numbers divisible 7
		for (int i = 1; i <= 1000; i++) {
			if (i % 7 == 0) {
				divisiblebySeven.add(i);
				// Remove the number form wholeset
				if (wholeSet.containsKey(i)) {
					wholeSet.remove(i);
				}
			}
		}

		// Find the numbers divisible 11
		for (int i = 1; i <= 100; i++) {
```

```
                        if (i % 11 == 0) {
                                divisiblebyEleven.add(i);
                                // remove the number form wholeset
                                if (wholeSet.containsKey(i)) {
                                        wholeSet.remove(i);
                                }
                        }
                }

                // print the wholeSet
                Set<Integer> keyset = wholeSet.keySet();
                Iterator<Integer> itr = keyset.iterator();
                while (itr.hasNext()) {
                        Integer key = itr.next();
                        System.out.println("Numbers not divisible by 3,7,11:" +key);
                }
        }

}
```

Numbers not divisible by 3,7,11
1000,998,997,995,992,991,989,988,986,985,983,982,979,977,976,974,971,970,968,967,965,964
,962,961,958,956,955,953,950,949,947,946,944,943,941,940,937,935,934,932,929,928,926,925,
923,922,920,919,916,914,913,911,908,907,905,904,902,901,899,898,895,893,892,890,887,886,
884,883,881,880,878,877,874,872,871,869,866,865,863,862,860,859,857,856,853,851,850,848,
845,844,842,841,839,838,836,835,832,830,829,827,824,823,821,820,818,817,815,814,811,809,
808,806,803,802,800,799,797,796,794,793,790,788,787,785,782,781,779,778,776,775,773,772,
769,767,766,764,761,760,758,757,755,754,752,751,748,746,745,743,740,739,737,736,734,733,
731,730,727,725,724,722,719,718,716,715,713,712,710,709,706,704,703,701,698,697,695,694,
692,691,689,688,685,683,682,680,677,676,674,673,671,670,668,667,664,662,661,659,656,655,
653,652,650,649,647,646,643,641,640,638,635,634,632,631,629,628,626,625,622,620,619,617,
614,613,611,610,608,607,605,604,601,599,598,596,593,592,590,589,587,586,584,583,580,578,
577,575,572,571,569,568,566,565,563,562,559,557,556,554,551,550,548,547,545,544,542,541,
538,536,535,533,530,529,527,526,524,523,521,520,517,515,514,512,509,508,506,505,503,502,
500,499,496,494,493,491,488,487,485,484,482,481,479,478,475,473,472,470,467,466,464,463,
461,460,458,457,454,452,451,449,446,445,443,442,440,439,437,436,433,431,430,428,425,424,
422,421,419,418,416,415,412,410,409,407,404,403,401,400,398,397,395,394,391,389,388,386,
383,382,380,379,377,376,374,373,370,368,367,365,362,361,359,358,356,355,353,352,349,347,
346,344,341,340,338,337,335,334,332,331,328,326,325,323,320,319,317,316,314,313,311,310,
307,305,304,302,299,298,296,295,293,292,290,289,286,284,283,281,278,277,275,274,272,271,
269,268,265,263,262,260,257,256,254,253,251,250,248,247,244,242,241,239,236,235,233,232,
230,229,227,226,223,221,220,218,215,214,212,211,209,208,206,205,202,200,199,197,194,193,
191,190,188,187,185,184,181,179,178,176,173,172,170,169,167,166,164,163,160,158,157,155,
152,151,149,148,146,145,143,142,139,137,136,134,131,130,128,127,125,124,122,121,118,116,
115,113,110,109,107,106,104,103,101,100,97,95,94,92,89,86,85,83,82,80,79,76,74,73,71,68,67,
65,64,62,61,59,58,53,52,50,47,46,43,41,40,38,37,34,32,31,29,26,25,23,20,19,17,16,13,10,8,5,4,
2,1.

Question 5 Create a normalized SQL database that shows a many to many relationship between information in two tables. Combine information from both tables into a single result set, and export its result to a CSV file. So that your work is reproducible, please provide scripts to show how you created the tables, populated the tables with a few sample records, combined information from the two tables into a result set, and exported the result set into a .CSV file.

```
/*Orders table*/
create table orders (ORDERID INT  auto_increment primary key, TIMEOFORDER  TIMESTAMP
DEFAULT CURRENT_TIMESTAMP, CUSTOMERNAME varchar(12));

/*items table*/
create table items (ITEMID INT  primary key , ITEMNAME VARCHAR(250) NOT NULL);
/* Many to many. We can associate many orders with many items or many items with many
orders*/

create table order_item (ORDERID INT NOT NULL,  ITEMID INT NOT NULL);

/**insert value into orders*/
insert into orders(Customername) values( 'Sridhar');
insert into orders(Customername) values( 'Max');
insert into orders(Customername) values('Mini');

/**insert value into items*/
insert into items(ITEMID,ITEMNAME) values(1,'TV');
insert into items(ITEMID,ITEMNAME) values(2,'Radio');
insert into items(ITEMID,ITEMNAME) values(3,'Ipad');

/**Track customer orders to items. Many orders may have many items. Many items associated
with many orders */
insert into order_item(ORDERID, ITEMID) values(1,1);
insert into order_item(ORDERID, ITEMID) values(1,3);
insert into order_item(ORDERID, ITEMID) values(2,2);
insert into order_item(ORDERID, ITEMID) values(3,1);
insert into order_item(ORDERID, ITEMID) values(3,2);
insert into order_item(ORDERID, ITEMID) values(3,3);

/** List all orders with the items*/
SELECT customername, itemname FROM orders o
JOIN order_item  ON o.ORDERID =order_item.ORDERID
JOIN items ON order_item.itemid = items.itemid
 INTO OUTFILE '/Users/sridhar/sridhar/temp/orderitems.csv'
FIELDS TERMINATED BY ',' ENCLOSED BY '"' LINES TERMINATED BY '\n';

/**now find out what sridhar purchased*/
SELECT customername, itemname FROM orders o
JOIN order_item  ON o.ORDERID =order_item.ORDERID
JOIN items ON order_item.itemid = items.itemid where customername='sridhar';
```

Question 6: Write code that uses latitude and longitude to calculate the distance between New York and Mumbai. Briefly explain why the concept of “distance” is important in data mining

```
package com.math;

public class FindDistance {

        public static void main (String[] args) throws java.lang.Exception
        {
                //Negative Longitude is West.  //Negative Latitude is south. Distance between
New york and Mumbai
                System.out.println("distance between Newyork and  Mumbai is "+
calculateDistance(40.7128, -74.0059, 19.0760, 72.8777) + " Miles\n");
```

```
        }

        private static double calculateDistance(double latitude1, double longitude1, double
latitude2, double longitude2) {
                double theta = longitude1 - longitude2;
                //formula to find the distance
                //dist = arccos(sin(lat1) · sin(lat2) + cos(lat1) · cos(lat2) · cos(lon1 - lon2)) · R
                double dist = Math.sin(degress2radians(latitude1)) *
Math.sin(degress2radians(latitude2)) + Math.cos(degress2radians(latitude1)) *
Math.cos(degress2radians(latitude2)) * Math.cos(degress2radians(theta));
                dist = Math.acos(dist);
                dist = radians2degrees(dist);
                dist = dist * 60 * 1.1515;
                return (dist);
        }


        //2PI radians is 360 degrees so, PI radian is 180 degree
        private static double degress2radians(double deg) {
                return (deg * Math.PI / 180.0);
        }


        private static double radians2degrees(double rad) {
                return (rad * 180 / Math.PI);
        }

}
```

Distance between New York and Mumbai is 7790 Miles

Question 7. Box A contains one white ball and two red balls. Box B contains one white ball and three red balls. A ball is picked at random from box A and put in box B. A ball is then picked at random from box B. What is the probability that the final ball picked is white?

Conditional Probability:

1.
Let us consider the first case of successfully drawing a while ball form box A then placed on box B.

Box A: 1 while + 2 red =3 balls
Box B: 1 whilte+ 3 red=4 balls

P(While from box A): 1/3

P(White from Box B | White ball from Box A) is = 2/5

2. Second case what if the ball from Box A is Red.
P(White from Box B | White ball from Box A) is = 1/5

Question 8. Find $\int t^2 e^t dt$

Let us solve this by parts

Formula fg-∫fg' and
let $f=e^t$ $g=t^2$
$f'= e^t$ $g'=2t$
$e^t t^2-\int e^t 2tdt+c$
$-\int 2t e^t dt+ e^t t^2$
Again we solve $\int 2t e^t dt$ by parts

Solve
$2\int t e^t dt$
$f=e^t$ $g=t$
$f'=et$ $g'=1$

$t e^{t} - \int e^{t\, dt+ c}$

$\int e^{t=} e^t$

$\int t e^t dt = t e^{t} - e^t$

Substitute this value in $- 2\int t e^t dt+ e^t t^2$
$-2(t e^{t} - e^t)+ e^t t^2$
$-2 t e^{t} + 2e^{t} + e^t t^2+ c$

Question 9. Find A^{-1} if A= $\begin{matrix} 5 & -1 \\ 2 & 3 \end{matrix}$

1. First find the Matrix of Minors
M(1,1)=3 M(1,2)=2 M(2,1)=-1, M(2,2)=5
$\begin{matrix} 3 & 2 \\ -1 & 5 \end{matrix}$

2. Find the cofactor matrix (Change sign for add number subscripts)
$\begin{matrix} 3 & -2 \\ 1 & 5 \end{matrix}$

3. Find Adjoint Matrix
$\begin{matrix} 3 & 1 \\ -2 & 5 \end{matrix}$

4. Find the determinant of the original matrix
Which is ad-bc= 5*3 –(-1*2)=15+2=17

5. A^{-1} = 1/determinant * Adjoint Matrix
1/17 * $\begin{matrix} 3 & 1 \\ -2 & 5 \end{matrix}$ = $\begin{matrix} 3/17 & 1/17 \\ -2/17 & 5/17 \end{matrix}$

Question **10.Find the eigenvalues and corresponding eigenvectors of**
A = $\begin{matrix} 4 & -6 \\ 3 & -7 \end{matrix}$
Formula to find the Eigenvalues and Eigenvectors
$A \cdot v = \lambda \cdot v$

A·v-**λ**·v**=0**
A·v-**λ**·I·v**=0**
(A-**λ**·I)·v**=0**
|A-**λ**·I**|=0**
I = $\begin{matrix}1 & 0\\ 0 & 1\end{matrix}$
|A-**λ**·I**|=** **det of** $\begin{matrix}\mathbf{4} & \mathbf{-6}\\ \mathbf{3} & \mathbf{-7}\end{matrix}$ **-** $\begin{matrix}\boldsymbol{\lambda} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\lambda}\end{matrix}$ **= 0**

deter of $\begin{matrix}4-\boldsymbol{\lambda} & -6\\ 3 & -7-\boldsymbol{\lambda}\end{matrix}$ = $\boldsymbol{\lambda}^2+3\boldsymbol{\lambda}-10$

This is a quadratic equation we solve this by factoring
$\boldsymbol{\lambda}^2+3\boldsymbol{\lambda}-10$
$\boldsymbol{\lambda}^2+5\boldsymbol{\lambda}-2\boldsymbol{\lambda}-\mathbf{10}=\mathbf{0}$

$\boldsymbol{\lambda}(\boldsymbol{\lambda}+5)-2(\boldsymbol{\lambda}+5)=0$

$(\boldsymbol{\lambda}+5)(\boldsymbol{\lambda}-\mathbf{2})\mathbf{=0}$
$(\boldsymbol{\lambda}+5)=0$ or $(\boldsymbol{\lambda}-\mathbf{2})\mathbf{=0}$

The possible $eigenvalues$ values are 2 and -5

Now, Find the eigenvector for **λ** =2
A·v1=λ1·v1
(A- λ) v1=0
$\begin{matrix}\mathbf{4}-\boldsymbol{\lambda} & \mathbf{-6}\\ \mathbf{3} & \mathbf{-7}-\boldsymbol{\lambda}\end{matrix}$ X $\begin{matrix}\boldsymbol{v1,1}\\ \boldsymbol{v1,2}\end{matrix}$ = 0
$\begin{matrix}\mathbf{2} & \mathbf{-6}\\ \mathbf{3} & \mathbf{-9}\end{matrix}$ X $\begin{matrix}\boldsymbol{v1,1}\\ \boldsymbol{v1,2}\end{matrix}$=0

2v1,1-6v1,2=0

v1,1=3v1,2

Vector V1= $\begin{matrix}\mathbf{3}\\ \mathbf{1}\end{matrix}$
Now, Find the eigenvector for **λ** =-5

A·v1=λ1·v1

$\begin{matrix}\mathbf{4}-\boldsymbol{\lambda} & \mathbf{-6}\\ \mathbf{3} & \mathbf{-7}-\boldsymbol{\lambda}\end{matrix}$ X $\begin{matrix}\boldsymbol{v2,1}\\ \boldsymbol{v2,2}\end{matrix}$=0

$\begin{matrix}\mathbf{9} & \mathbf{-6}\\ \mathbf{3} & \mathbf{-2}\end{matrix}$ X $\begin{matrix}\boldsymbol{v2,1}\\ \boldsymbol{v2,2}\end{matrix}$=0

9v2,1-6v2,2=0

3v2,1=2v2,2

v2,1=2/3v2,2
Vector v2= $\begin{matrix}\mathbf{2/3}\\ \mathbf{1}\end{matrix}$

Question 11 At a move theater, customers wait in line to buy tickets for an average of 5 minutes, and they wait to buy popcorn for an average of 8 minutes. Assuming that the wait times are independent, find the probability that a customer waits a total of less than 20 minutes before taking his or her seat.

The probability density function of an exponential distribution is

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x}, & x < 0 \\ 0, & x \geq 0 \end{cases}$$

Where λ is rate of event occurrence.

Associate the events to variables X and Y
X is buying tickets and Y is buying popcorn.
λ of X is 5 minutes (average wait time to buy ticket)
λ of Y is 8 minutes (average wait time to buy popcorn)
Both if these events are independent. So we have to multiply their probabilities.

$$f(x) = \begin{cases} \frac{1}{5}e^{-x/5}, & x < 0 \\ 0, & x \geq 0 \end{cases} \qquad f(y) = \begin{cases} \frac{1}{8}e^{-y/8}, & y < 0 \\ 0, & y \geq 0 \end{cases}$$

$$f(x,y) = f(x)f(y) = \begin{cases} \frac{1}{40}e^{-\frac{x}{5}}e^{-\frac{y}{8}}, & x > 0 \\ 0, & y > 0 \end{cases}$$

Less than 20 minutes before taking seat, so

$$P(A + B < 20) = \int_0^{20} \int_0^{20-x} \frac{1}{40} e^{-\frac{x}{5}} e^{-\frac{y}{8}} \, dy \, dx$$

Double integral. So the first one becomes constant. i.e. $e^{-\frac{x}{5}}$ remains same. Integrate by u substitution and find $e^{-\frac{y}{8}}$. Which is $-8\, e^{-\frac{y}{8}}$

$$1/40 \int_0^{20} \left[e^{-\frac{x}{5}} (-8)\, e^{-\frac{y}{8}} \right] \text{ apply y limit values of 0 to 20-x}$$

$$1/40 \int_0^{20} \left[e^{-\frac{x}{5}} \left((-8)\, e^{-\frac{x-20}{8}} \right) - \left((-8) e^{-\frac{0}{8}} \right) \right]$$

$$1/40 \int_0^{20} \left[e^{-\frac{x}{5}} \left(-8\, e^{-\frac{x-20}{8}} \right) + 8 \right]$$

$$1/40 \int_0^{20} \left[e^{-\frac{x}{5}}\, 8\, (1 - e^{(x-20)/8}) \right]$$

$$8/40 \int_0^{20} \left[e^{-\frac{x}{5}} \ (1 - e^{(x-20)/8}) \right]$$

Question 12.Briefly (in 200 words or less) compare business intelligence and predictive analytics.

Business intelligence and Predictive analytics are used to gain more insights from data so that business can benefit.

Business intelligence

Business intelligence is an umbrella term, which includes data mining, processing, querying and reporting. Variety of software applications used in business intelligence. Businesses use BI to improve decision-making, cut costs and identify new business opportunities. Business intelligence is of descriptive analytics. . i.e. "what happened in the past with the business?" It is kind of reporting on what happened and what is currently happening. It helps to understand the relationship between customers and products and the objective is to gain an understanding of what approach to take in the future: learn from past behavior to influence future outcomes.

Predictive Analytics

In todays big data world we entered the new area of predictive analytics, which focuses on answering the question: "what is probably going to happen in the future?" It is a forward-looking analysis: providing future-looking insights on the business—predicting what is likely to happen. It provides organizations with actionable insights based on data. It provides an estimation regarding the likelihood of a future outcome. In order to do this, a variety of techniques are used, such as machine learning, data mining, modeling. Predictive analytics can help to identify risks or opportunities in the future.

We need business intelligence to know what really happened in the past and need predictive analytics to optimize resources, make decisions and take actions for the future.

Question 13. If you were to play a match of 100 "rock-paper-scissors" games against a well-written computer program, who do you expect would win? Briefly explain your reasoning**.**

"rock-paper-scissors" is a random game. It requires understanding the pattern of the opponent to win over him. So definitely computer will win over time, because computer learns from human actions. Computers use the previously learned outcomes to play.

Question 14 https://xkcd.com/882/. Please read and explain

This is comedy cartoon about the scientific experiment and the media reporting on it. Scientist trying to find probability of jellybeans causing acne. They are announcing various results by conducting a statistical survey. Trying to find out probability of various colored jellybeans causing acne and announcing their results every time. Almost 19 colored jellybeans are ruled out each with a probability of .05.

Which comes with a total of 19 X .05=.95. Found a link between green jellybeans causing acne of .05 probability. So finally the media reports the results, which is misleading. It says 95% of confidence that jellybeans are causing acne, but the actual result is that 95% chance that jellybeans are not causing acne. Media reports gives the idea to the reader that green jellybeans causing acne is of 95%. Which is wrong. It is a humor about how media understand the science.

Question 15 Have you ever used data to solve an interesting problem? If so, please describe.

Yes. To tune performance of java applications I used “jvm” garbage collection data. Data tells the number of objects created, long-lived objects, memory used etc. Normally this will help to benchmark the minimum required operation conditions of the software.

Question 16 What do you see as the most important skills in the day-to-day work of a successful data scientist?

A data scientist does not simply collect and report on data, but also looks at it from many angles, determines what it means, then recommends ways to apply the data. So a data scientist must be a talented individual in multiple fields such as statistics, software and math. **Data scientist should pick the right problems that have the most value to the organization.**

Question 17 Describe what you see as the biggest obstacle for you to become a successful data scientist, and your plan to overcome this obstacle.

Worked as software engineer and architect for many years. My job is done with onboarding data, processing, storing and supporting software applications. Usually business intelligence team will analyze and prepare the reports after wards. Currently working with big data to find user sentiment, data classification and rating. This requires knowledge in machine learning, which is a field of predictive analysis. I learned some of the machine learning algorithms and associated math with that. Initially confused with the terms used such as vector, feature vector, models and training etc in this field. But now I am getting comfortable and working towards this goal. The biggest challenge is to link various fields such software development, applying math models to solve the real world problems. I am working towards this and hoping to join “MS Data Analytics” programs from CUNY.

Question 18 If you were to ask one additional question for this challenge exam, what would that question be?

I feel the current level of questions is almost testing the applicant in most of the math fields such as algebra, geometry, statistics, probability and matrix. I feel it is good enough. May be there is room to ask few questions about the software development process and cycles.