

**NAME:** Sarvesh Patil

**CLASS:** D15A

**ROLL NO:** 46

**LAB 01**

**LAB 01:** To understand the process of Breaking the Mono-alphabetic Substitution Cipher using the Frequency analysis method

<b>ROLL NO</b>	46
<b>NAME</b>	Sarvesh Patil
<b>CLASS</b>	D15A
<b>SUBJECT</b>	Internet Security Lab
<b>LO MAPPED</b>	LO1: To apply the knowledge of symmetric cryptography to implement classical ciphers

**AIM:**

To understand the process of Breaking the Mono-alphabetic Substitution Cipher using Frequency analysis method

**INTRODUCTION:****Mono-Alphabetic Substitution Cipher**

A mono-alphabetic cipher (aka simple substitution cipher) is a substitution cipher where each letter of the plain text is replaced with another letter of the alphabet. It uses a fixed key which consists of the 26 letters of a "shuffled alphabet".

Plain text alphabet:	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
Cipher text alphabet (key):	M	U	A	L	V	O	Z	K	R	N	J	X	Q	D	F	S	H	P	E	B	C

With the above key, all "A" letters in the plain text will be encoded to an "M".

This type of cipher is a form of symmetric encryption as the same key can be used to both encrypt and decrypt a message.

A monoalphabetic substitution is a cipher in which each occurrence of a plaintext symbol is replaced by a corresponding ciphertext symbol to generate the ciphertext. The key for such a cipher is a table of the correspondence or a function from which the correspondence is computed.

*Example:* An affine cipher  $E(x) = (ax + b) \text{ MOD } 26$  is an example of a monoalphabetic substitution.

There are other ways to "generate" a monoalphabetic substitution.

***Alphabet Mixing via a Keyword***

A keyword or keyphrase can be used to mix the letters to generate the cipher alphabet.

*Example:* If the keyword is ANDREW DICKSON WHITE, then the cipher alphabet is given by

plain A B C D E F G H I J K L M N O P Q R S T U V W X Y Z  
 cipher A N D R E W I C K S O H T B F G J L M P Q U V X Y Z

Perhaps a better keyword is EZRA CORNELL:

plain A B C D E F G H I J K L M N O P Q R S T U V W X Y Z  
 cipher E Z R A C O N L B D F G H I J K M P Q S T U V W X Y

***Alphabet Mixing via a Columnar Transposition***

The letters from the keyword form the headings of the columns, and the remaining letters of the alphabet fill in order in the rows below. Mixing is achieved by transcribing columns.

Example: If the keyword is CORNELL, then write

C O R N E L

A B D F G H

I J K M P Q

S T U V W X

Y Z

so that transcribing columns left-to-right gives the substitution

plain A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

cipher C A I S Y O B J T Z R D K U N F M V E G P W L H Q X

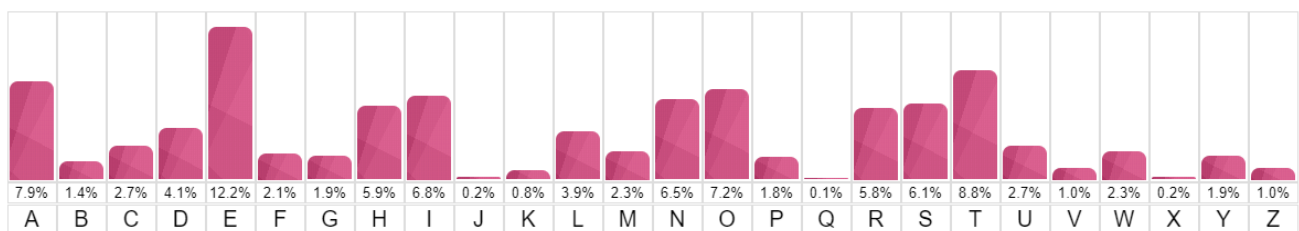
For instance, FAR ABOVE CAYUGA'S WATERS is enciphered as OCVCA NWYIC QPBCE LCGYE.

## Frequency Analysis

In cryptography, frequency analysis is the study of the frequency of letters or groups of letters in a ciphertext. The method is used as an aid in breaking substitution ciphers (e.g. mono-alphabetic substitution cipher, Caesar shift cipher, Vatsyayana cipher).

Frequency analysis consists of counting the occurrence of each letter in a text. Frequency analysis is based on the fact that, in any given piece of text, certain letters and combinations of letters occur with varying frequencies. For instance, given a section of the English language, letters E, T, A, and O are the most common, while letters Z, Q, and X are not as frequently used.

The following chart shows the frequency of each letter of the alphabet for the English language:



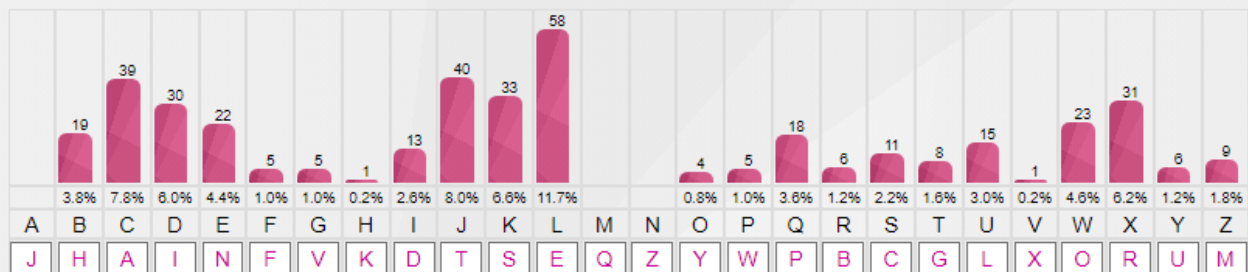
We can assume that most samples of text written in English would have a similar distribution of letters. However, this is only true if the sample of the text is long enough. A very short text may lead to a significantly different distribution.

When trying to decrypt a cipher text based on a substitution cipher, we can use frequency analysis to help identify the most recurring letters in a cipher text and hence make the hypothesis of what these letters have been encoded as (e.g. E, T, A, O, etc). This will help us decrypt some of the letters in the text. We can then recognize patterns/words in the partly decoded text to identify more substitutions.

**RESULTS:****1. CIPHER #1****KEY: RANDOM****Frequency Analysis**

Text:

DJ DK C QLXDWI WF SDGDU PCX. XLRLU KQCSLKBDQK, KJXDHDET FXWZ C BDIILE RCKL, BCGL PWE  
 JBLDX FDXKJ GDSJWXO CTCDEKJ JBL LGDU TCUCSJDS LZQDXL. IYXDET JBL RCJJUL, XLRLU KQDLK  
 ZCECTLI JW KJLCU KLSXLJ QUCEK JW JBL LZQDXL'K YUJDZCJL PLCQWE, JBL ILCJB KJCK, CE  
 CXZWXLI KQCSL KJCJDWE PDJB LEWYTB QWPLX JW ILKJXWO CE LEJDXL QUCELJ. QYXKYLI RO JBL  
 LZQDXL'K KDEDKJLX CTLEJK, QXDESLKK ULDC XCSLK BWZL CRWCXI BLX KJCXKBDQ, SYKJWIDCE WF JBL  
 KJWULE QUCEK JBCJ SCE KCGL BLX QLWQUL CEI XLKJWL FXLLIWZ JW JBL TCUCVO...

**1. Start Frequency Analysis****2. Start Substitution**

Text After Substitution:

IT IS A PERIOD OF CIVIL WAR. REBEL SPACESHIPS, STRIKING FROM A HIDDEN BASE, HAVE WON  
 THEIR FIRST VICTORY AGAINST THE EVIL GALACTIC EMPIRE. DURING THE BATTLE, REBEL SPIES  
 MANAGED TO STEAL SECRET PLANS TO THE EMPIRE'S ULTIMATE WEAPON, THE DEATH STAR, AN  
 ARMORED SPACE STATION WITH ENOUGH POWER TO DESTROY AN ENTIRE PLANET. PURSUED BY THE  
 EMPIRE'S SINISTER AGENTS, PRINCESS LEIA RACES HOME ABOARD HER STARSHIP, CUSTODIAN OF  
 THE STOLEN PLANS THAT CAN SAVE HER PEOPLE AND RESTORE FREEDOM TO THE GALAXY...



3. CIPHER #3

KEY: RANDOM

Frequency Analysis

Text:

"OK OH R WRFD KOIS QPF KNS FSTISJJOPX. RJKNPAGN KNS WSRKN HKRF NRH TSSX WSHKFEPCSW, OIBSFORJ KEPEBH NRYS WFOYSX KNS FSTISJ QPFMSH QFPI KNSOF NOWWSX TRHS RXW BAFHASW KNSI RMFFPHH KNS GRJREC. SYRWXG KNS WFSRWSW OIBSFORJ HKRFQJSSK, R GFPAB PQ QFSSWPI QOGNKSEH JSW TC JADS HDCVRJDSF NRH SHKRIJOHNSW R XSV HSMFSK TRHS PX KNS FSIPKS OMS VFFJW PQ NPKN. KNS SYOJ JPEW WRFKN YRWSE, PTHSHSW VOKN QOXWGX CPXG HDCVRJDSF, NRH WOHERKMNWSW KNPAHRXWH PQ FSIPKS BFPTSH OXKP KNS QRF FSRMNSH PQ HBRMS..."

1. Start Frequency Analysis

7	8	6	6	1	32	7	29	8	16	30		7	27	21	26	12	32	60	9		5	25	13	5	
1.4%	1.7%	1.2%	1.2%	0.2%	6.6%	1.4%	6.0%	1.7%	3.3%	6.2%		1.4%	5.6%	4.3%	5.4%	2.5%	6.6%	12.4%	1.9%		1.0%	5.2%	2.7%	1.0%	
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
U	P	Y	K	X	R	G	S	M	L	T		C	H	I	O	F	A	E	B		W	D	N	V	

2. Start Substitution

Text After Substitution:

"IT IS A DARK TIME FOR THE REBELLION. ALTHOUGH THE DEATH STAR HAS BEEN DESTROYED, IMPERIAL TROOPS HAVE DRIVEN THE REBEL FORCES FROM THEIR HIDDEN BASE AND PURSUED THEM ACROSS THE GALAXY. EVADING THE DREADED IMPERIAL STARFLEET, A GROUP OF FREEDOM FIGHTERS LED BY LUKE SKYWALKER HAS ESTABLISHED A NEW SECRET BASE ON THE REMOTE ICE WORLD OF HOTH. THE EVIL LORD DARTH VADER, OBSESSED WITH FINDING YOUNG SKYWALKER, HAS DISPATCHED THOUSANDS OF REMOTE PROBES INTO THE FAR REACHES OF SPACE..."

4. CIPHER #4  
KEY: RANDOM

Frequency Analysis

Text:

"ZRIFT IH PQFTHZ IQ ZRT XBGBOZIO HTQBZI. HTWIFBG ZRLPHBQV HLGBF HYHZTSH RBWT VIOGBFIV ZRTIF IQZIQZILQH ZL GIBWT ZRT FTEPKGIO.  
ZRIH HTEBFBZINH SLWISTIQZ, EQVTF ZRT GTBVFHRIE LD ZRT SYHZTFILPH OLPQZ VLLAP, RBH SEVT IZ VIDDIOFGZ DLF ZRT GISIZIV QPSKTF LD CTVI AQIXRZH ZL SBIQZBIQ EIBOT BQV LEVIF IQ ZRT XGBBJY.  
HTQBZLF BSIVBGB, ZRT DLFSTF NPITQ LD QBKLL, IH FTZPFQIQX ZL ZRT XBGBOZIO HTQBZI ZL WLZI LQ ZRT OFIZIOBG IHHPF LD OFTBZIQX BQ BFSY LD ZRT FTEPKGIO ZL BHHIHZ ZRT LWFMRGTSTV CIVI..."

1. Start Frequency Analysis

2

33

0.4%

6.6%

A

K

2

9

0.4%

1.8%

C

A

5

25

1.0%

5.1%

E

J

15

26

3.1%

5.3%

G

P

32

27

6.6%

5.5%

I

R

1

4

0.2%

0.8%

J

L

1

1

0.2%

0.2%

M

O

13

13

2.7%

2.7%

O

W

25

20

5.1%

4.1%

Q

Q

12

56

2.5%

11.5%

S

C

15

6

3.1%

1.2%

T

U

6

6

1.2%

1.2%

V

N

4

44

0.8%

9.0%

X

H

Y

M

Z

E

2. Start Substitution

Text After Substitution:

"THERE IS UNREST IN THE GALACTIC SENATE. SEVERAL THOUSAND SOLAR SYSTEMS HAVE DECLARED THEIR INTENTIONS TO LEAVE THE REPUBLIC.  
THIS SEPARATIST MOVEMENT, UNDER THE LEADERSHIP OF THE MYSTERIOUS COUNT DOOKU, HAS MADE IT DIFFICULT FOR THE LIMITED NUMBER OF JEDI KNIGHTS TO MAINTAIN PEACE AND ORDER IN THE GALAXY.  
SENATOR AMIDALA, THE FORMER QUEEN OF NABOO, IS RETURNING TO THE GALACTIC SENATE TO VOTE ON THE CRITICAL ISSUE OF CREATING AN ARMY OF THE REPUBLIC TO ASSIST THE OVERWHELMED JEDI..."



## 5. CIPHER #5

### KEY: RANDOM

### Frequency Analysis

Text:

"FX IWBBJX PB NB PB PWX GBBD. VSP FWO, JBGX JRO, PWX GBBD? FWO IWBBJX PWUJ RJ BSA NBRK? RDL PWXO GRO FXKK RJM FWO IKUGV PWX WUNWXP GBSDPRUD? FWO, 35 OXRAJ RNB, EKO PWX RPKRDPUI? FWO LBKJ AUIX CKRO FXQRJ? FX IWBBJX PB NB PB PWX GBBD UD PWUJ LXIRLX RDL LB PWX BPWXA PWUDNJ, DBP VXIRSJX PWXO RAX XRJO, VSP VXIRSJX PWXO RAX WRAL, VXIRSJX PWRP NBRK FUKK JXATX PB BANRDUZX RDL GXRJSAX PWX VXJP BE BSA XDXANUXJ RDL JMUKKJ, VXIRSJX PWRP IWRKKXDNX UJ BDX PWRP FX RAX FUKKUDN PB RIIXCP, BDX FX RAX SDFUKKUDN PB CBJPCBDX, RDL BDX FWUIW FX UDPXDL PB FUD, RDL PWX BPWXA, PBB."

1. Start Frequency Analysis

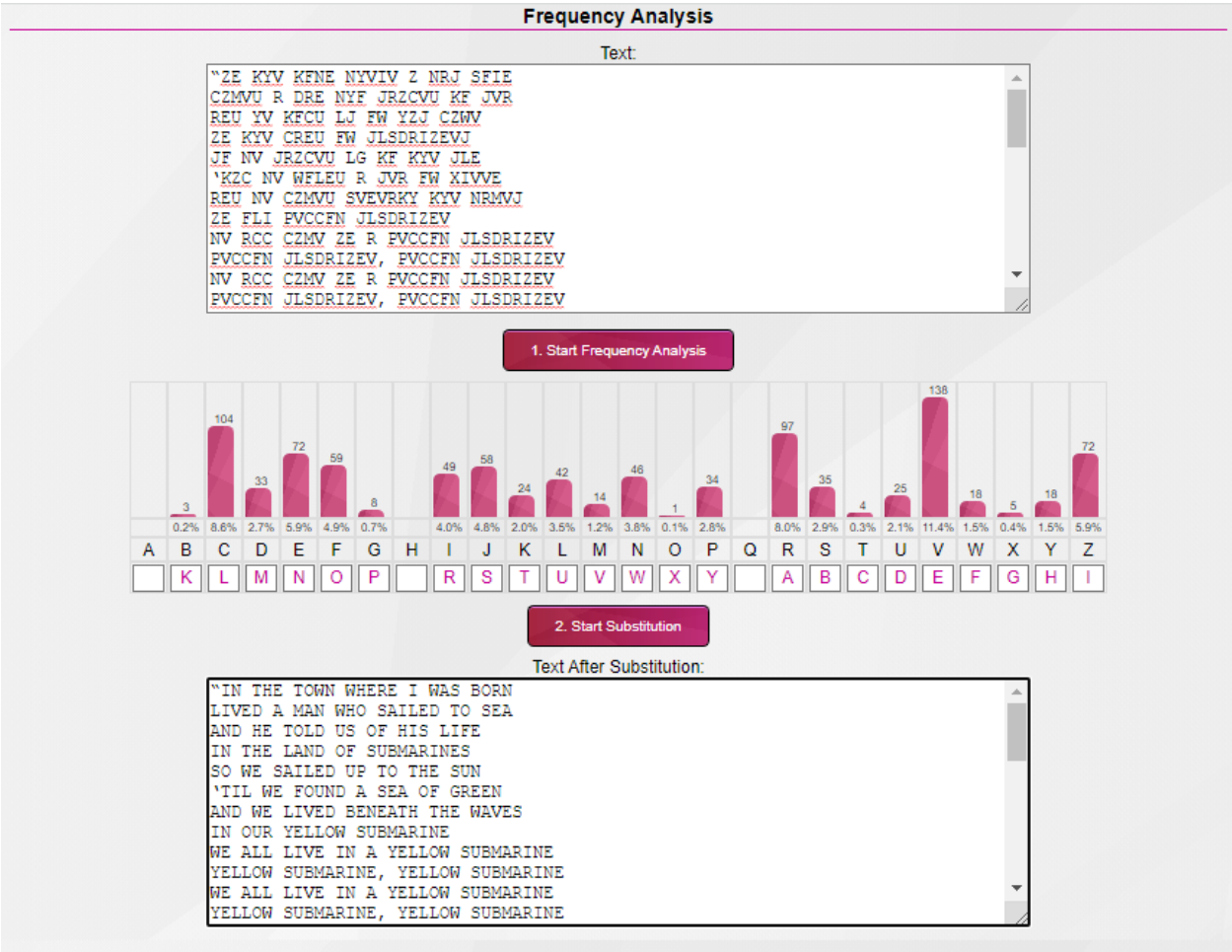
Letter	Count	Percentage
A	15	2.7%
B	43	7.6%
C	4	0.7%
D	28	5.0%
E	2	0.4%
F	16	2.8%
G	8	1.4%
H	0	0.0%
I	15	2.7%
J	28	5.0%
K	18	3.2%
L	12	2.1%
M	2	0.4%
N	12	2.1%
O	14	2.5%
P	43	7.6%
Q	1	0.2%
R	37	6.5%
S	11	1.9%
T	1	0.2%
U	21	3.7%
V	8	1.4%
W	33	5.8%
X	60	10.6%
Y	0	0.0%
Z	1	0.2%

2. Start Substitution

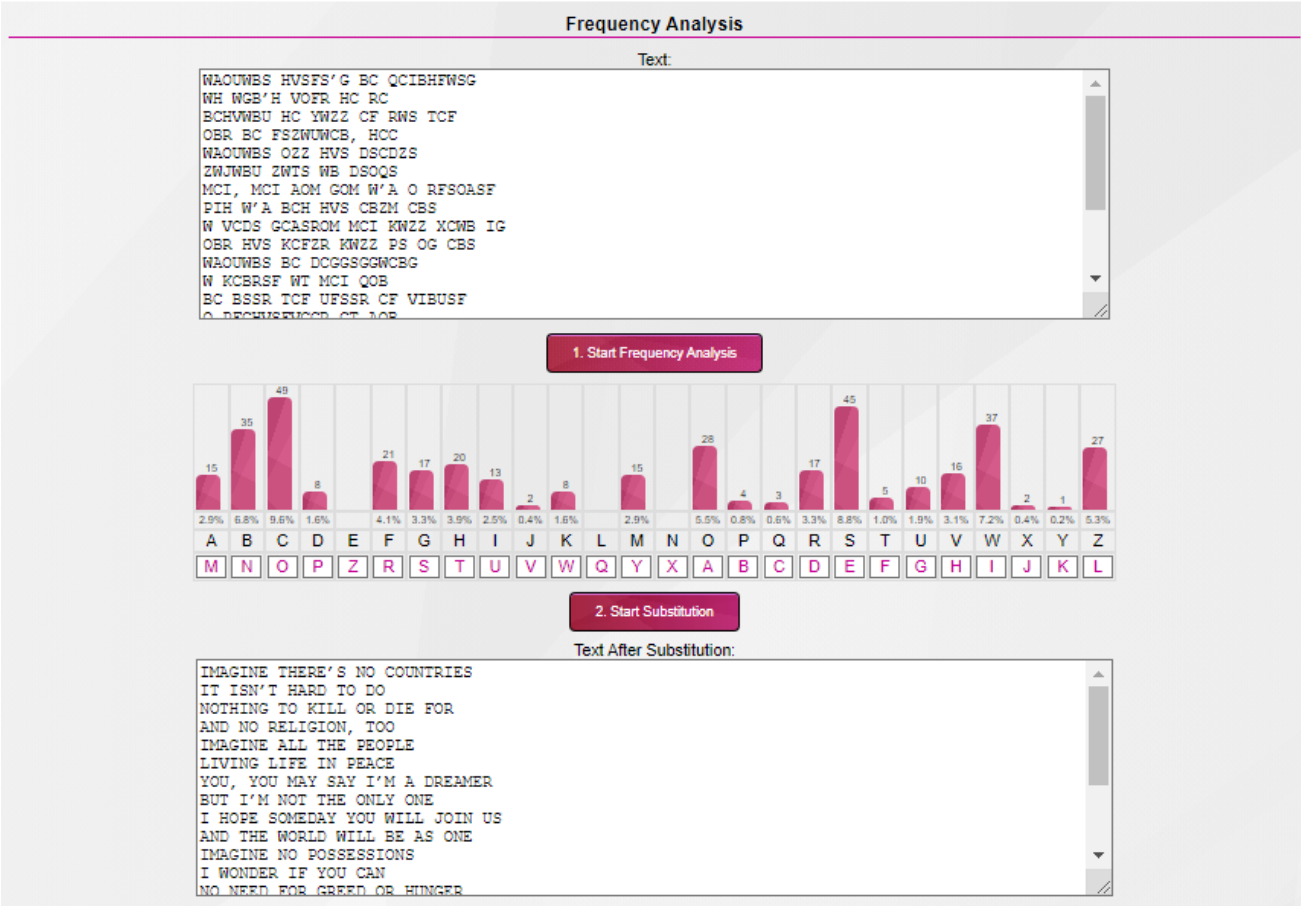
Text After Substitution:

"WE CHOOSE TO GO TO THE MOON. BUT WHY, SOME SAY, THE MOON? WHY CHOOSE THIS AS OUR GOAL? AND THEY MAY WELL ASK WHY CLIMB THE HIGHEST MOUNTAIN? WHY, 35 YEARS AGO, FLY THE ATLANTIC? WHY DOES RICE PLAY TEXAS? WE CHOOSE TO GO TO THE MOON IN THIS DECADE AND DO THE OTHER THINGS, NOT BECAUSE THEY ARE EASY, BUT BECAUSE THEY ARE HARD, BECAUSE THAT GOAL WILL SERVE TO ORGANISE AND MEASURE THE BEST OF OUR ENERGIES AND SKILLS, BECAUSE THAT CHALLENGE IS ONE THAT WE ARE WILLING TO ACCEPT, ONE WE ARE UNWILLING TO POSTPONE, AND ONE WHICH WE INTEND TO WIN, AND THE OTHERS, TOO."

6. CIPHER #6  
KEY: 10



7. CIPHER #7  
KEY: 12



8. CIPHER #8

KEY: 12

Frequency Analysis

Text:

SBRL ZRFDHSRLY OHZ YLABYULK AV OPZ OVTL WSHULA VM AHAVVPUL PU HU HAALTWA AV YLZJBL OPZ MYPLUK OHU ZVSV MYVT AOL JSBAJOLZ VM AOL CPSL NHUNZALY QHIIH AOL OBAA.  
SPAASL KVLZ SBRL RUVD ACHA AOL NHSHJAPJ LTWPYL OHZ ZLOJLASF ILNBU JVUZAYBJAPVU VU H ULD HYTVYLK ZWHJL ZAHAPVU LCLU TVYL WVDLYMBS AOHU AOL MPYZA KYLHKLK KLHAO ZAHY.  
  
DOLU JVTWLSLALK, AOPZ BSAPTHAL DLHWVU DPSS ZWLSS JLYAHPU KVVV MUY AOL ZTHSS IHUK VM YLILSZ ZAYBNNSPUN AV YLZAVYL MYLLKVT AV AOL NHSHEF...

1. Start Frequency Analysis

43

11

2

7

1

3

29

5

11

12

52

9

8

20

17

1

5

23

11

22

29

8

24

22

9.4%

2.4%

0.4%

1.5%

0.2%

0.7%

6.3%

1.1%

2.4%

2.6%

11.4%

2.0%

1.8%

4.4%

3.7%

0.2%

1.1%

5.0%

2.4%

4.8%

6.3%

1.8%

5.3%

4.8%

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

T

U

V

W

X

Y

Q

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Z

R

S

2. Start Substitution

Text After Substitution:

LUKE SKYWALKER HAS RETURNED TO HIS HOME PLANET OF TATOOINE IN AN ATTEMPT TO RESCUE HIS FRIEND HAN SOLO FROM THE CLUTCHES OF THE VILE GANGSTER JABBA THE HUTT.  
LITTLE DOES LUKE KNOW THAT THE GALACTIC EMPIRE HAS SECRETLY BEGUN CONSTRUCTION ON A NEW ARMORED SPACE STATION EVEN MORE POWERFUL THAN THE FIRST DREADED DEATH STAR.  
  
WHEN COMPLETED, THIS ULTIMATE WEAPON WILL SPELL CERTAIN DOOM FOR THE SMALL BAND OF REBELS STRUGGLING TO RESTORE FREEDOM TO THE GALAXY...

**NAME:** Sarvesh Patil

**CLASS:** D15A

**ROLL NO:** 46

**ANALYSIS:**

In most of the problems, it is observed that "e" is being repeated the most number of times. and "a" is repeated the 2nd most number of times. Many problems had random keys but some had keys like 10, 12, and 19

**CONCLUSION:**

We have successfully understood the process of Breaking the Mono-alphabetic Substitution Cipher using the Frequency analysis method