PROJECT – PART 2

Abstract: The report describes the process of performing chi-square analysis on the datasets from project part 1 i.e. youtuber subscriber count dataset and time interval between cars dataset. This data is used to verify or reject Null hypothesis

COURSE: DASC 5302-002 Intro to Probability & Statistics
INSTRUCTOR: Obiageli Lawrentia Ngwu

"I Chirag Hebbal Rudresh did not give or receive any assistance on this project, and the report submitted is wholly my own."

Chirag Hebbal Rudresh
X
Chirag Hebbal Rudresh

Name: Chirag Hebbal Rudresh

ID: 1002160960
Email: cxr0960@mavs.uta.edu

Date: 11/24/2023

INDEX

DATASET 1

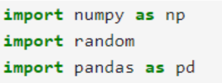
DATA:

Set 1 consists of 125 randomly picked YouTube channels and their subscriber count data. The data was collected, and the following analysis was done according to the data on September 2023.

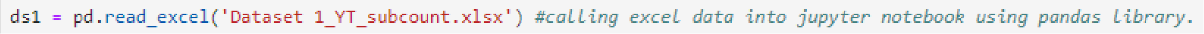
PROCESS OF DATA COLLECTION:

The YouTube channel and subscriber count data was collected from social blade website (socialblade.com). The website's home page has a drop-down menu labelled "TOP LISTS" and under YouTube popular countries, 5 countries lists were used. Top 100 list of the following countries were used in the data: United states, United Kingdom, Australia, Canada, and Germany. The data consisted of 125 YouTube channels which were picked randomly from the list of top 100 YouTube channels in these 5 countries. From a list of each country's top 100 channels, 25 were randomly picked which total to 125 channels.

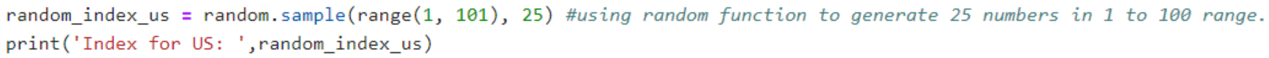
To pick 25 random indexes from each country's top 100 list, following python code was used to generate 25 random values within a range of 1-100:



```
import numpy as np
import random
import pandas as pd
```

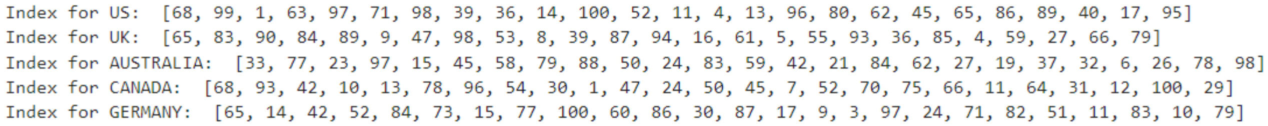


```
ds1 = pd.read_excel('Dataset 1_YT_subcount.xlsx') #calling excel data into jupyter notebook using pandas library.
```



```
random_index_us = random.sample(range(1, 101), 25) #using random function to generate 25 numbers in 1 to 100 range.
print('Index for US: ',random_index_us)
```

The same code is run 5 times to get 25 random indexes for the 5 different lists as shown below:



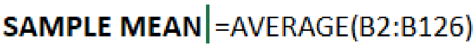
```
Index for US:  [68, 99, 1, 63, 97, 71, 98, 39, 36, 14, 100, 52, 11, 4, 13, 96, 80, 62, 45, 65, 86, 89, 40, 17, 95]
Index for UK:  [65, 83, 90, 84, 89, 9, 47, 98, 53, 8, 39, 87, 94, 16, 61, 5, 55, 93, 36, 85, 4, 59, 27, 66, 79]
Index for AUSTRALIA:  [33, 77, 23, 97, 15, 45, 58, 79, 88, 50, 24, 83, 59, 42, 21, 84, 62, 27, 19, 37, 32, 6, 26, 78, 98]
Index for CANADA:  [68, 93, 42, 10, 13, 78, 96, 54, 30, 1, 47, 24, 50, 45, 7, 52, 70, 75, 66, 11, 64, 31, 12, 100, 29]
Index for GERMANY:  [65, 14, 42, 52, 84, 73, 15, 77, 100, 60, 86, 30, 87, 17, 9, 3, 97, 24, 71, 82, 51, 11, 83, 10, 79]
```

Now, the YouTube Channels were picked based on these indexes from the top 100 lists of 5 countries to make a total 125 channels. The corresponding subscriber counts for each channel were also collected by visiting their channels on the YouTube website. The channel name and subscriber count data were stored in an excel sheet.

Note: The method of using python code to generate 25 random indexes was used to ensure the data was truly selected at random without bias. Instead, the observer can pick 25 YouTube channels at random from each list and collect the data.

DESCRIPTIVE STATISTICS:

1. **Sample mean**: The mean of the data was calculated using the inbuilt average formula in excel. An empty cell was selected in the excel sheet and AVERAGE(range of data) formula was used to get the mean of the range of values mentioned in the braces of the formula. For example, the subscriber count values ranged from cell B2 to cell B126 which are 125 values. So the formula will be AVERAGE(C2:C110).

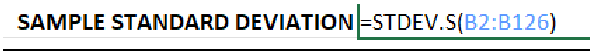


SAMPLE MEAN =AVERAGE(B2:B126)

The sample mean of 125 YouTube channel's subscriber count was found to be 60,65,632. This means that majority of channels in the dataset have 60,65,632 subscribers.

SAMPLE MEAN 60,65,632

2. **<u>Sample Standard Deviation</u>**: The standard deviation of the data was calculated using the inbuilt STDEV formula in excel. An empty cell was selected in the excel sheet and STDEV.S(range of data) formula was used to get the standard deviation of the range of values mentioned in the braces of the formula. For example, the subscriber count values ranged from cell B2 to cell B126 which are 125 values. So, the formula will be STDEV.S(C2:C110).



SAMPLE STANDARD DEVIATION =STDEV.S(B2:B126)

The standard deviation of 125 YouTube channel's subscriber count was found to be 99,42,564. This gives information about how spread the data is. From this data it can be implied that on an average that the subscriber counts of each channel deviate from the mean subscriber count by 99,42,564.

SAMPLE STANDARD DEVIATION 99,42,564

3. **<u>Quartiles (Q1, Q2, Q3)</u>**:
Q1 (First Quartile): This is the 25th percentile. It is the value below which 25% of the data falls. To calculate the first quartile, the excel formula- QUARTILE(range of values, quartile number) to get the first quartile of the range of values mentioned in the braces of the formula. For the data, the following formula was used: QUARTILE(C2:C110,1). The first quartile value was found to be 7,47,000.

Q1 =QUARTILE(B2:B126,1)

Q1 7,47,000

Q2 (Second Quartile): This is the 50th percentile and is also the median of the dataset. It divides the data into two equal halves, with 50% of the data falling below it and 50% above it. Similarly, for the data the following formula was used: QUARTILE(C2:C110,2). The first second quartile value was found to be 22,70,000.

Q2 =QUARTILE(B2:B126,2)

Q2 22,70,000

Q3 (Third Quartile): This is the 75th percentile. It is the value below which 75% of the data falls. For the data, the following formula was used QUARTILE(C2:C110,3). The third quartile value was found to be 67,10,000.

Q3 =QUARTILE(B2:B126,3)

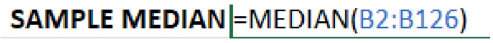
Q3 67,10,000

4. **<u>Geometric mean</u>**: The geometric mean is a measure of central tendency that is used to find the average of a set of positive numbers. Unlike the arithmetic mean, which sums the values and divides by the number of values, the geometric mean calculates the nth root of the product of n values. Here, the excel formula: GEOMEAN(C2:C110) is used and the geometric mean was found to be 22,65,214.

GEOMETRIC MEAN 22,65,214

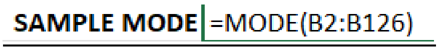
5. **<u>Sample median</u>**: The median is a statistical measure of central tendency used to find the middle value in a dataset when the data is ordered (sorted) from smallest to largest. It is the

value that separates the higher half from the lower half of the data. Here, the excel formula: MEDIAN(C2:C110) is used and the median was found to be 22,70,000.



SAMPLE MEDIAN =MEDIAN(B2:B126)

6. **Sample mode**: The mode is measure of central tendency that represents the most frequently occurring value (or values) in a dataset. The excel formula: MODE(C2:C110) was used to calculate mode and it was found to be 28,00,000.



SAMPLE MODE =MODE(B2:B126)

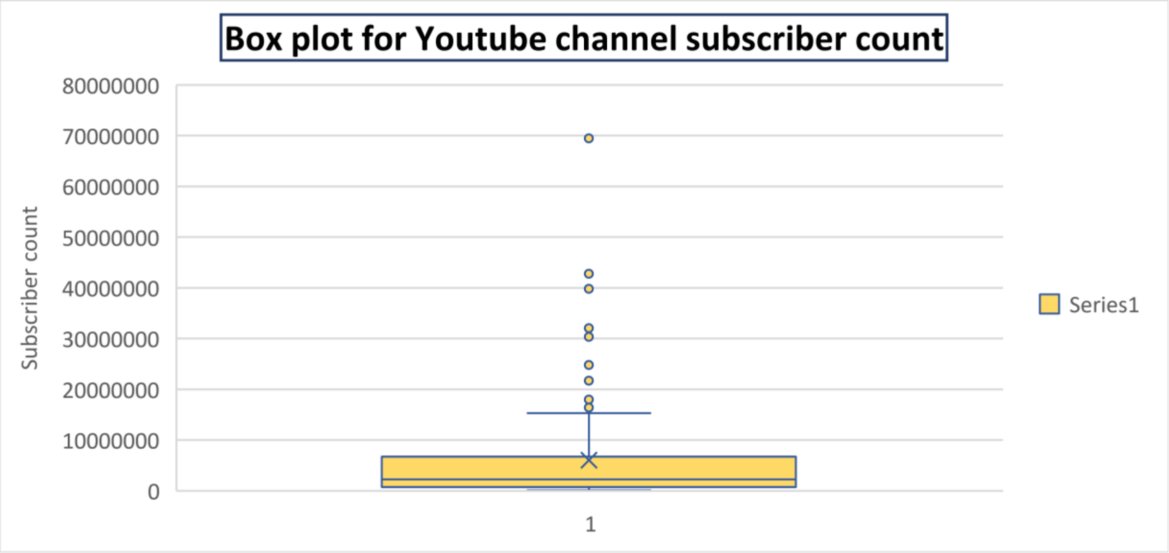
7. **Sample range**: The sample range is a measure that describes the spread or variability within a dataset. It can be calculated by taking the difference of the maximum and minimum value in the dataset. In this case, the range was found to be 6,93,77,000(6,95,00,000- 1,23,000).

8. **Sample variance**: It provides a numerical value that indicates how much individual data points deviate from the sample mean (average). The excel formula used to calculate it is: VAR(C2:C110). The variance of the dataset is 9,88,54,58,28,47,354.80.

9. **Coefficient of variation**: It is a measure that assesses the relative variability or dispersion of data points in relation to their mean (average)Coefficient of Variation (CV)= (Standard Deviation/Mean). In this case the coefficient of Variation was found to be 1.64. The data can be considered as having less variability and most of the data lies around the mean of the dataset.

COEFFICIENT OF VARIATION	1.64

10. **Box-Whisker plot**:
To gain insight on the gathered data, a Box & Whisker plot was used. To get the plot in excel first the subscriber count data was selected and then Insert->Charts->Box & Whisker plot. The following plot is obtained, and it shows the distribution of the subscriber count data.



Plot 1.

From the above plot we can see there are many outliers present in the data. The data points above the whisker lines are abnormal values when compared with remaining data. For perfect analysis these outliers need to be dealt with either by removing those values or replacing them with the mean of the sub_count column. To keep the analysis simple, cleaning steps haven't been considered.

Median(Q2): The rectangular box is where most of the central values lie. A line in the middle of this rectangle depicts the median(Q2) i.e., the middle value of the entire dataset.

Whiskers: The lines extending from the rectangular box are called whiskers. These typically extend to the minimum and maximum values within a certain range, excluding outliers. Whiskers depict the full data range.

Outliers: Values represented beyond the whiskers are called outliers, they indicate values that are significantly different from most of the data.

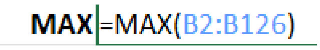
Skewness: Since the median is closer to Q1, it can be assumed that the distribution is right-skewed.

Type of distribution: The two whiskers in the plot are not equal or nearly equal also. So, it can be said that the dataset's distribution is not normal.

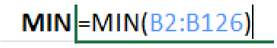
In summary, a box and whisker plot is a versatile tool for understanding and visualizing data distributions. By examining the minimum, maximum, quartiles, and outliers, we can gain valuable insights into the central tendency, spread, and shape of the data.

11. **<u>Frequency table and Frequency histogram</u>**:

To create a frequency table, first take the subscriber count data and find out the maximum and minimum values. The formulas MAX(range of data) and MIN(range of data) are used here.



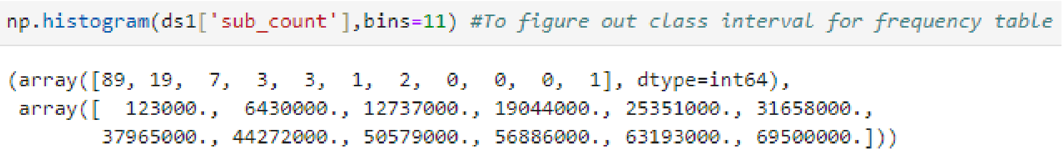
MAX =MAX(B2:B126)



MIN =MIN(B2:B126)

It was found that the maximum subscriber count was 6,95,00,000 and the minimum count was 1,23,000. For the frequency table, the data was divided into 11 class intervals where each interval was separated by a count of 63,07,000 and the number of channels within those subscriber counts were counted.

For this process, python code was used to make this process error free and efficient. The following code was used to find the class intervals and frequency of channels that had a subscriber count within those ranges.

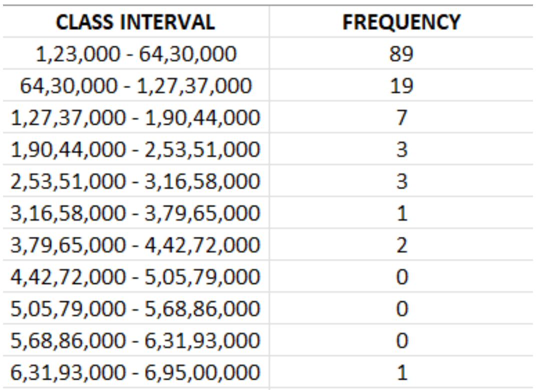


```
np.histogram(ds1['sub_count'],bins=11) #To figure out class interval for frequency table
```

```
(array([89, 19,  7,  3,  3,  1,  2,  0,  0,  0,  1], dtype=int64),
 array([  123000.,  6430000., 12737000., 19044000., 25351000., 31658000.,
        37965000., 44272000., 50579000., 56886000., 63193000., 69500000.]))
```

In the above code, 'sub_count' refers to the column name that is given in the excel sheet for the subscriber count numbers and 'bins=11' refers to the number of class intervals needed from the data. Below the code, in the output, the first line has 11 numbers which represent the frequency of channel occurrences, and the next line shows 12 numbers which are the class intervals required.

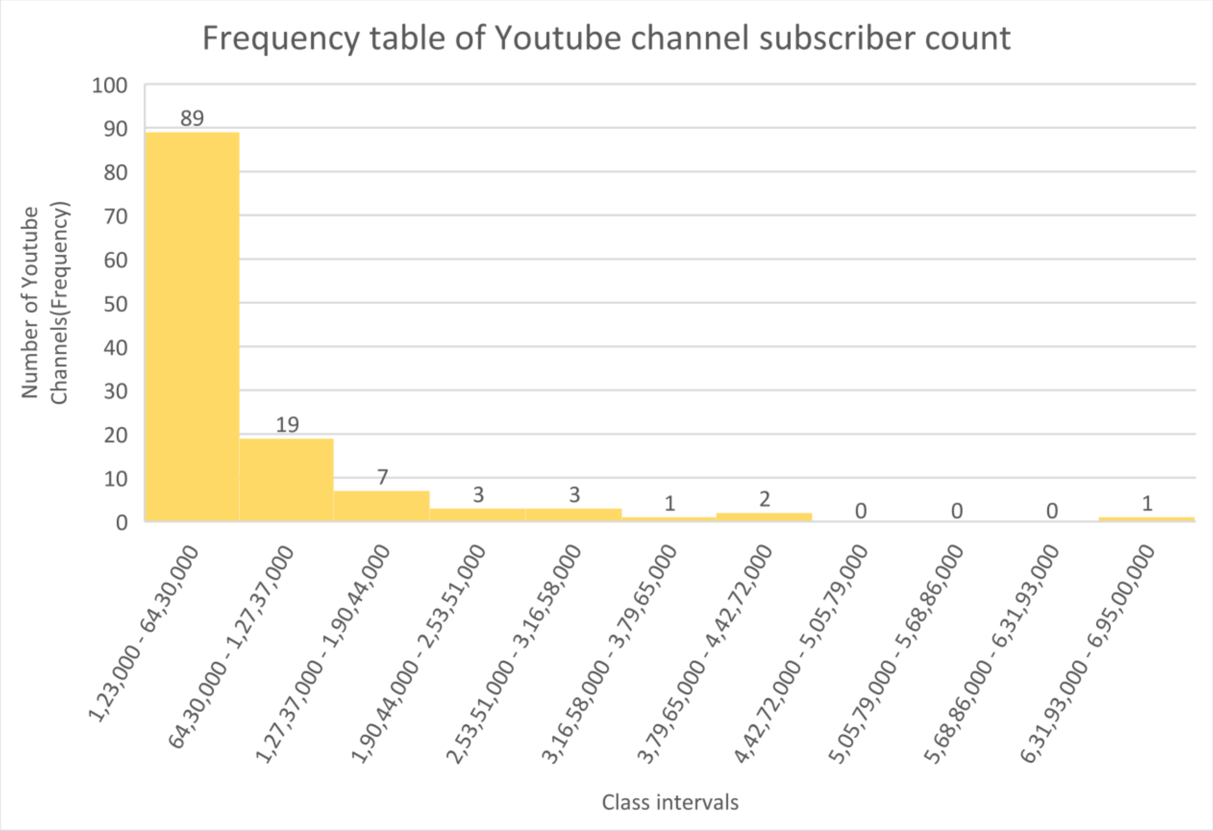
For example, the first number 89 is the frequency of channels between the class interval 1,23,000 and 64,30,000. Similarly, 19 is the frequency of channels between the class interval 64,30,000 and 1,27,37,000. The same logic is used to get all 11 class intervals and their respective frequencies. The following table was made from the above data.



CLASS INTERVAL	FREQUENCY
1,23,000 - 64,30,000	89
64,30,000 - 1,27,37,000	19
1,27,37,000 - 1,90,44,000	7
1,90,44,000 - 2,53,51,000	3
2,53,51,000 - 3,16,58,000	3
3,16,58,000 - 3,79,65,000	1
3,79,65,000 - 4,42,72,000	2
4,42,72,000 - 5,05,79,000	0
5,05,79,000 - 5,68,86,000	0
5,68,86,000 - 6,31,93,000	0
6,31,93,000 - 6,95,00,000	1

Table 1.

The above data was used to plot a frequency histogram graph. The class interval and frequency column data were selected, then Insert->Charts->Histogram. The following chart will be generated.



Plot 2.

The above histogram depicts frequency of YouTube channels between different ranges of subscriber count. The x-axis is the range of subscribers with minimum being 1,23,000 and maximum being 6,95,00,000 and they are the interval ranges which were calculated in Table 1. The y-axis depicts the number of YouTube channels.

Skewed: In plot 2., we can see that majority of the YouTube channels from the dataset have subscribe count that lie between 1,23,000 and 64,30,000 with 89 channels. From the plot it can be concluded that the dataset has a right-skewed distribution since the tail of the plot is leading to the right side. From the above analysis, it can be concluded that dataset 1 **doesn't have a normal distribution** of data.

CHI-SQUARE ANALYSIS

DATA:

Set 1 consists of 125 randomly picked YouTube channels and their subscriber count data. The data was collected, and the following analysis was done according to the data in September 2023.

PROCEDURE: The dataset was analyzed, and the sample mean of the dataset was calculated to be 60,65,632 subscribers (6065.632 thousand subscribers). The sample standard deviation was also calculated to be 99,42,564.2 (9942.5642 thousand subscribers). Dataset 1 is assumed to have a normal distribution.

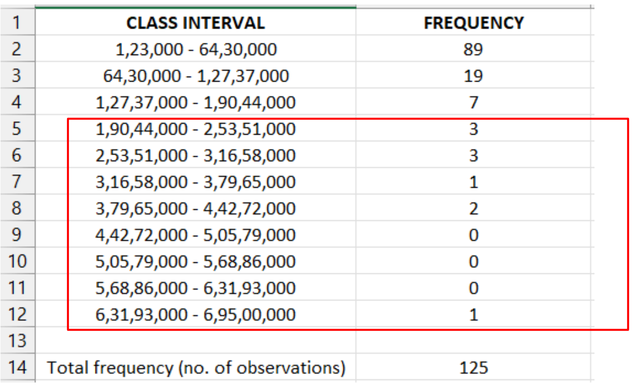
HYPOTHESIS:

H_0 (Null hypothesis) = Dataset 1 is normally distributed.

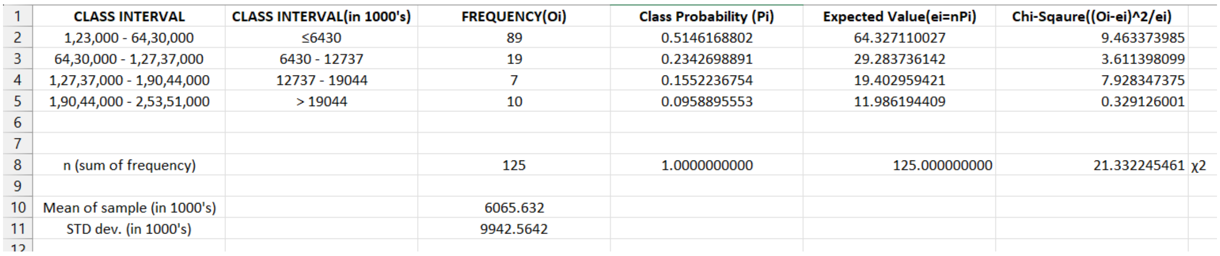
H_1 (Alternate Hypothesis) = Dataset 1 is not normally distributed.

COMBINING INTERVALS:

Originally, the dataset had 11 class intervals. During calculation of chi-squared values, the expected value(e_i) values of intervals after 2,53,51,000 subscribers were coming out to be less than 5. So, these intervals were combined, and their respective frequencies were added which resulted in a new interval.



1	CLASS INTERVAL	FREQUENCY
2	1,23,000 - 64,30,000	89
3	64,30,000 - 1,27,37,000	19
4	1,27,37,000 - 1,90,44,000	7
5	1,90,44,000 - 2,53,51,000	3
6	2,53,51,000 - 3,16,58,000	3
7	3,16,58,000 - 3,79,65,000	1
8	3,79,65,000 - 4,42,72,000	2
9	4,42,72,000 - 5,05,79,000	0
10	5,05,79,000 - 5,68,86,000	0
11	5,68,86,000 - 6,31,93,000	0
12	6,31,93,000 - 6,95,00,000	1
13		
14	Total frequency (no. of observations)	125



1	CLASS INTERVAL	CLASS INTERVAL(in 1000's)	FREQUENCY(Oi)	Class Probability (Pi)	Expected Value(ei=nPi)	Chi-Sqaure((Oi-ei)^2/ei)	
2	1,23,000 - 64,30,000	≤6430	89	0.5146168802	64.327110027	9.463373985	
3	64,30,000 - 1,27,37,000	6430 - 12737	19	0.2342698891	29.283736142	3.611398099	
4	1,27,37,000 - 1,90,44,000	12737 - 19044	7	0.1552236754	19.402959421	7.928347375	
5	1,90,44,000 - 2,53,51,000	> 19044	10	0.0958895553	11.986194409	0.329126001	
6							
7							
8	n (sum of frequency)		125	1.0000000000	125.000000000	21.332245461	χ2
9							
10	Mean of sample (in 1000's)		6065.632				
11	STD dev. (in 1000's)		9942.5642				

The class probability (P_i) was calculated using the formula:

NORMDIST (x, mean(μ), standard deviation(σ),1)

The following is the formula used for first class interval: =**NORMDIST (6430,6065.632,9942.5642,1)**

For the second interval and upcoming intervals the formula is as follows:

=NORMDIST (upper limit, μ, σ,1) - NORMDIST (lower limit, μ, σ,1)

=NORMDIST (12737,6065.632,9942.5642,1) - NORMDIST (6430,6065.632,9942.5642,1) (2nd interval)

For the last interval, the formula is as follows: = **1 - NORMDIST (19044,6065.632,9942.5642,1)**

The expected value of the class intervals (ei) was calculated using the formula: **ei=n*Pi** where n is the number of observations(n=125).

For calculating Chi-square, sum of [(Oi – ei) ^2]/2 values for all intervals were calculated. The summation of these values was done to find the chi-square value of the dataset. **χ^2= Σ [(Oi – ei) ^2]/2.**

In the above chi-square formula Oi is the frequency of observations for that class interval.

From the table above, the Chi-square **χ^2** value was computed to be 21.332. When chi-square value was calculated using critical values of chi-squared distributions (Table A5), for **α=0.05** and degrees of freedom v= [(number of classes(k) – 1], i.e., 3, the value **$\chi^2_{\alpha, k-1}$** was found to be 7.815.

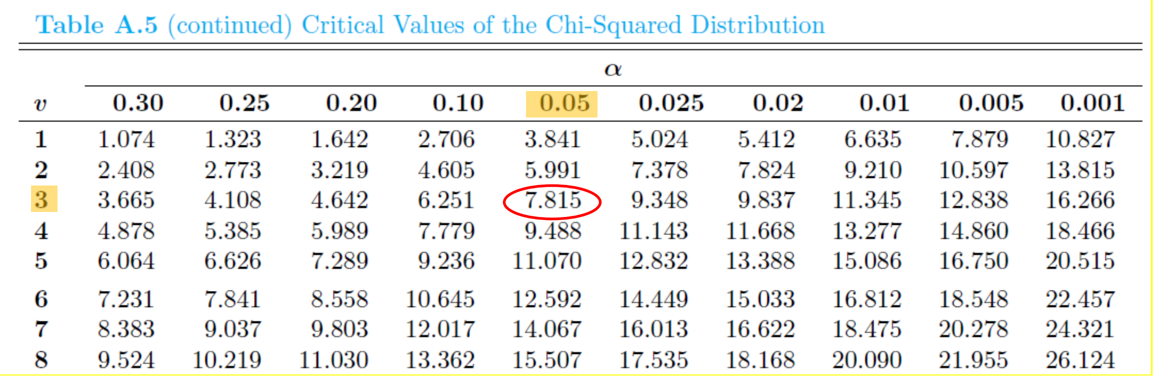
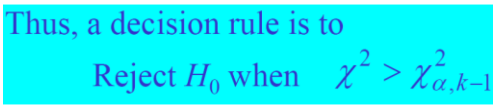


Table A.5 (continued) Critical Values of the Chi-Squared Distribution

	α									
v	0.30	0.25	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.001
1	1.074	1.323	1.642	2.706	3.841	5.024	5.412	6.635	7.879	10.827
2	2.408	2.773	3.219	4.605	5.991	7.378	7.824	9.210	10.597	13.815
3	3.665	4.108	4.642	6.251	7.815	9.348	9.837	11.345	12.838	16.266
4	4.878	5.385	5.989	7.779	9.488	11.143	11.668	13.277	14.860	18.466
5	6.064	6.626	7.289	9.236	11.070	12.832	13.388	15.086	16.750	20.515
6	7.231	7.841	8.558	10.645	12.592	14.449	15.033	16.812	18.548	22.457
7	8.383	9.037	9.803	12.017	14.067	16.013	16.622	18.475	20.278	24.321
8	9.524	10.219	11.030	13.362	15.507	17.535	18.168	20.090	21.955	26.124

<u>Decision Rule</u>:



Thus, a decision rule is to

Reject H_0 when $\chi^2 > \chi^2_{\alpha,k-1}$

Here, H_0 (Null hypothesis) = Dataset 1 is normally distributed.

H_1 (Alternate Hypothesis) = Dataset 1 is not normally distributed.

Therefore, **χ^2 is greater than $\chi^2_{\alpha, k-1}$, We reject H_0**.

- We are 95% confident that dataset 1 does not follow normal distribution.

DATASET 2

DATA:

Set 2 consists of data regarding clock times of 110 cars when crossing an apartment. Data for both data sets were collected and stored as excel files. The data was collected, and the following analysis was done on a Saturday in the month of September 2023.

PROCESS OF DATA COLLECTION:

Readings for dataset 2 were collected by observing cars cross the apartment at the 404 Border apartments, opposite White Rhino Cafe on East border street, Arlington. The readings were taken on a Saturday between 9:27:00 am and 10:15:00 am. The dataset consists of clock times when a car passes by the apartment. The clock times of 110 cars were collected and saved in an excel file. The excel file consists of 2 columns: First column are the original clock times, and the second column is time interval between each observation i.e., each car crossing the apartment. The data for the second column is obtained by subtracting a particular clocked time with the previous clock time. For example, if the first car crossed the apartment at 9:27:05 am and the second car crossed at 9:27:15 am (hours: minutes: seconds), then the time interval between them would be [(9:27:15) - (9:27:05)] which would be 10 seconds. Similarly, after following this process for all the clock times, we get 109 time intervals. Further analysis is done on these time interval data points.

Descriptive Statistics:

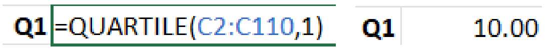
1. **Sample mean**: The mean of the data was calculated using the inbuilt average formula in excel. An empty cell was selected in the excel sheet and AVERAGE(range of data) formula was used to get the mean of the range of values mentioned in the braces of the formula. For example, the time interval values ranged from cell C2 to cell C110 which are 109 values. So the formula will be AVERAGE(C2:C110).

   The sample mean of 109 time intervals was found to be 25.93 seconds. This means that majority of the cars in the dataset have an average of 25.93 seconds of time interval between each instance of car crossing.

2. **Sample Standard Deviation**: The standard deviation of the data was calculated using the inbuilt STDEV formula in excel. An empty cell was selected in the excel sheet and STDEV.S(range of data) formula was used to get the standard deviation of the range of values mentioned in the braces of the formula. For example, the subscriber count values ranged from cell C2 to cell C110 which are 109 values. So, the formula will be STDEV.S(C2:C110).

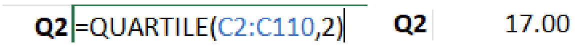
   The standard deviation of 109 time intervals was found to be 23.44 seconds. This gives information about how spread the data is. From this data it can be implied that on an average that the crossing clock time of each car deviates from the mean interval time by 23.44 seconds.

3. **Quartiles (Q1, Q2, Q3)**:
   Q1 (First Quartile): This is the 25th percentile. It is the value below which 25% of the data falls. To calculate the first quartile, the excel formula - QUARTILE(range of values, quartile number) to get the first quartile of the range of values mentioned in the braces of the formula. For the data, the following formula was used: QUARTILE(C2:C110,1). The first quartile value was found to be 10.



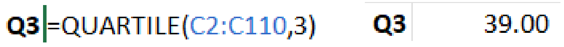
Q1 =QUARTILE(C2:C110,1) Q1 10.00

Q2 (Second Quartile): This is the 50th percentile and is also the median of the dataset. It divides the data into two equal halves, with 50% of the data falling below it and 50% above it. Similarly, for the data the following formula was used: QUARTILE(C2:C110,2). The first second quartile value was found to be 17.



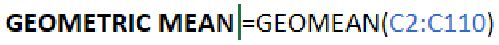
Q2 =QUARTILE(C2:C110,2) Q2 17.00

Q3 (Third Quartile): This is the 75th percentile. It is the value below which 75% of the data falls. For the data, the following formula was used QUARTILE(C2:C110,3). The third quartile value was found to be 39.



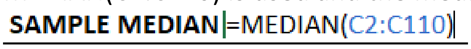
Q3 =QUARTILE(C2:C110,3) Q3 39.00

4. **Geometric mean**: The geometric mean is a measure of central tendency that is used to find the average of a set of positive numbers. Unlike the arithmetic mean, which sums the values and divides by the number of values, the geometric mean calculates the nth root of the product of n values. Here, the excel formula: GEOMEAN(C2:C110) is used and the geometric mean was found to be 17.25.



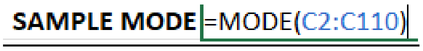
GEOMETRIC MEAN =GEOMEAN(C2:C110)

5. **Sample median**: The median is a statistical measure of central tendency used to find the middle value in a dataset when the data is ordered (sorted) from smallest to largest. It is the value that separates the higher half from the lower half of the data. Here, the excel formula: MEDIAN(C2:C110) is used and the median was found to be 17.00.



SAMPLE MEDIAN =MEDIAN(C2:C110)

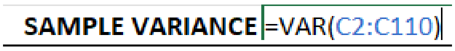
6. **Sample mode**: The mode is measure of central tendency that represents the most frequently occurring value (or values) in a dataset. The excel formula: MODE(C2:C110) was used to calculate mode and it was found to be 12.00.



SAMPLE MODE =MODE(C2:C110)

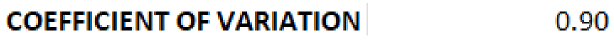
7. **Sample range**: The sample range is a measure that describes the spread or variability within a dataset. It can be calculated by taking the difference of the maximum and minimum value in the dataset. In this case, the range was found to be 99.00 seconds (100- 1).

8. **Sample variance**: It provides a numerical value that indicates how much individual data points deviate from the sample mean (average). The excel formula used to calculate it is: VAR(C2:C110). The variance of the dataset is 549.55.



SAMPLE VARIANCE =VAR(C2:C110)

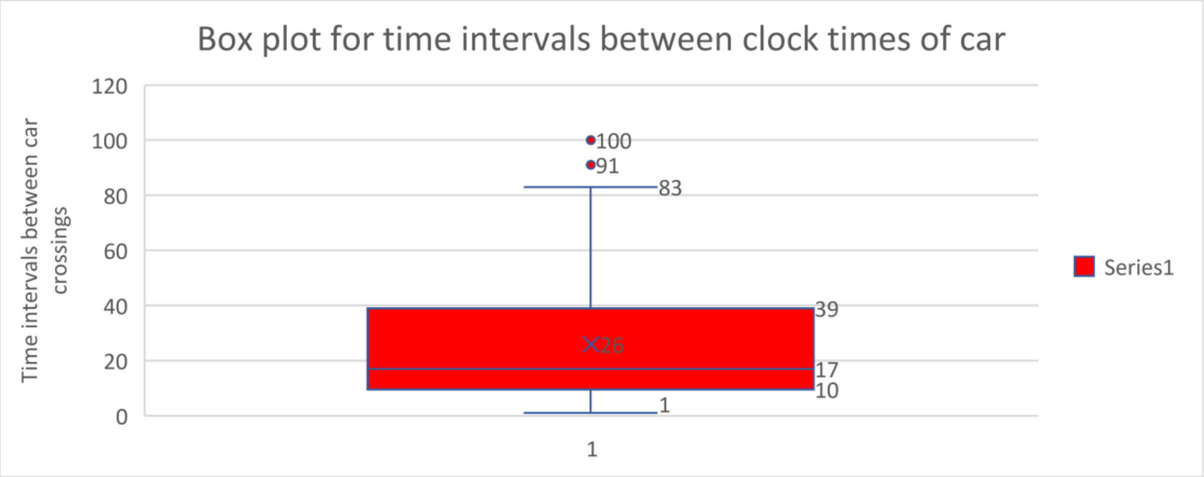
9. **Coefficient of variation**: It is a measure that assesses the relative variability or dispersion of data points in relation to their mean (average)Coefficient of Variation (CV)= (Standard Deviation/Mean). In this case the coefficient of Variation was found to be 0.90. The data can be considered as having less variability and most of the data lies around the mean of the dataset.



COEFFICIENT OF VARIATION 0.90

10. **Box-Whisker plot**:

To gain insight on the gathered data, a Box & Whisker plot was used. To get the plot in excel first the subscriber count data was selected and then Insert->Charts->Box & Whisker plot. The following plot is obtained, and it shows the distribution of time intervals between clocked time data.



Plot 3.

Here, we can spot the outliers which are labelled as 100 and 91. These are the values that deviate very much from the rest of the data.

Median(Q2): The rectangular box is where most of the central values lie. A line in the middle of this rectangle depicts the median(Q2) i.e., the middle value of the entire dataset.

Whiskers: The lines extending from the rectangular box are called whiskers. These typically extend to the minimum and maximum values within a certain range, excluding outliers. Whiskers depict the full data range.

Outliers: Values represented beyond the whiskers are called outliers, they indicate values that are significantly different from most of the data.

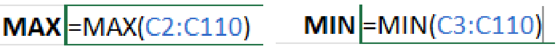
Skewness: Since the median is closer to Q1, it can be assumed that the distribution is right-skewed.

Type of distribution: The two whiskers in the plot are not equal or nearly equal also. So, it can be said that the dataset's distribution is not normal.

In summary, a box and whisker plot is a versatile tool for understanding and visualizing data distributions. By examining the minimum, maximum, quartiles, and outliers, we can gain valuable insights into the central tendency, spread, and shape of the data.

12. **Frequency table and Frequency histogram**:

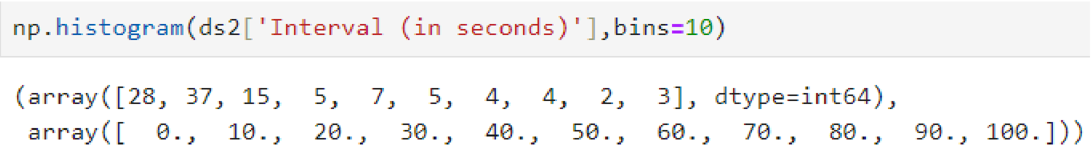
To create a frequency table, first take the time interval data and find out the maximum and minimum values. The formulas MAX(range of data) and MIN(range of data) are used here.



MAX	=MAX(C2:C110)	MIN	=MIN(C3:C110)

It was found that the maximum time interval was 100 seconds, and the minimum time interval was 1 second. For the frequency table, the data was divided into 10 class intervals where each interval was separated by 10 seconds and the number of car crossings within those time intervals were counted.

For this process, python code was used to make this process error free and efficient. The following code was used to find the class intervals and frequency of time intervals that had a car crossing within those ranges.

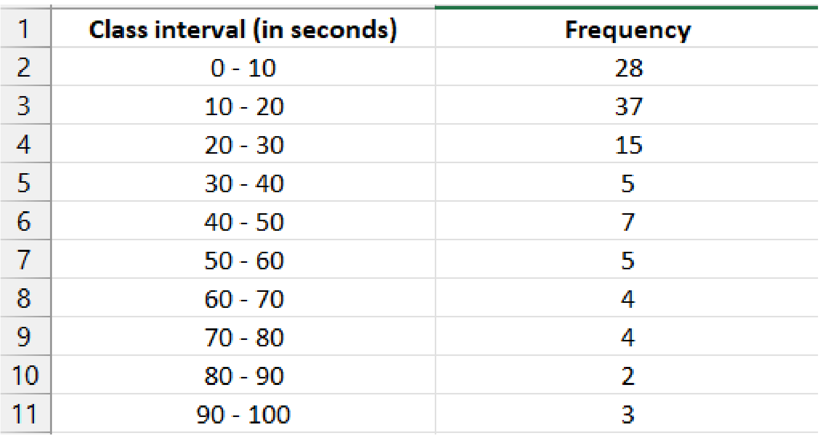


```
np.histogram(ds2['Interval (in seconds)'],bins=10)
```

```
(array([28, 37, 15,  5,  7,  5,  4,  4,  2,  3], dtype=int64),
 array([  0.,  10.,  20.,  30.,  40.,  50.,  60.,  70.,  80.,  90., 100.]))
```

In the above code, ‘Interval (in seconds) refers to the column name that is given in the excel sheet for the time intervals and ‘bins=10’ refers to the number of class intervals needed from the data. In the code, in the output, the first line has 10 numbers which represent the frequency of car crossing occurrences, and the next line shows 11 numbers which are the class intervals required.

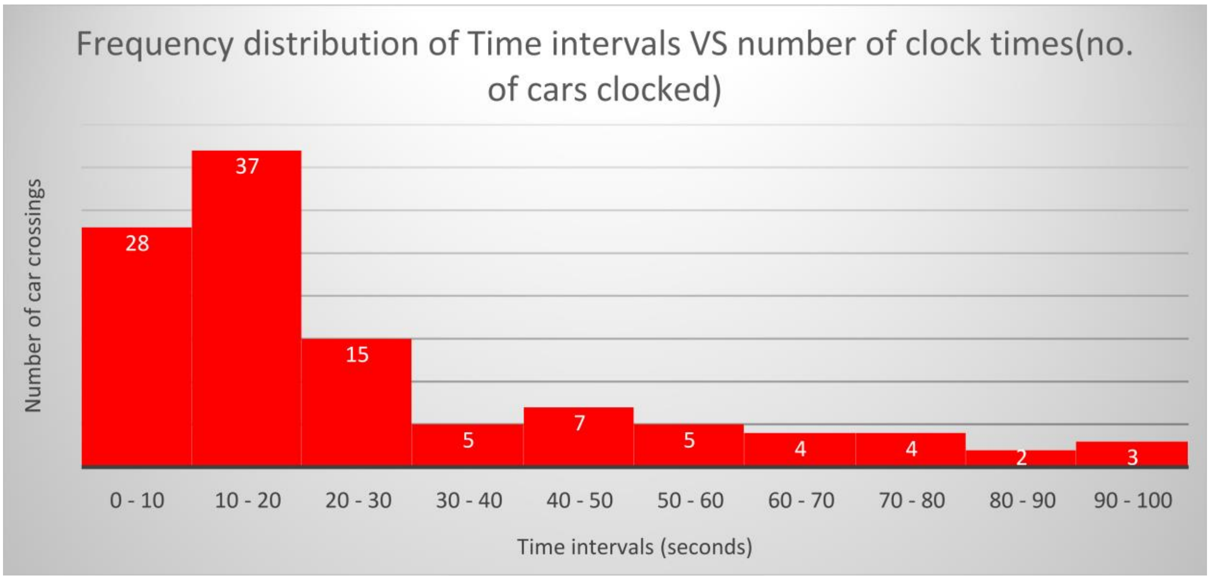
For example, the first number 28 is the frequency of car crossings between the class interval 0 seconds and 10 seconds. Similarly, 37 is the frequency of car crossings between the class interval 10 seconds and 20 seconds. The same logic is used to get all 10 class intervals and their respective frequencies. The following table was made from the above data.



1	**Class interval (in seconds)**	**Frequency**
2	0 - 10	28
3	10 - 20	37
4	20 - 30	15
5	30 - 40	5
6	40 - 50	7
7	50 - 60	5
8	60 - 70	4
9	70 - 80	4
10	80 - 90	2
11	90 - 100	3

Table 2.

The above data was used to plot a frequency histogram graph. The class interval and frequency column data were selected, then Insert->Charts->Histogram. The following chart will be generated.



Plot 4.

The above histogram depicts frequency of car crossings between different ranges of time intervals. The x-axis is the range of time intervals with minimum being 1 second and maximum being 100 seconds and they are the interval ranges which were calculated in Table 1. The y-axis depicts the number of car crossings.

Skewed: In plot 4., we can see that majority of the car crossings from the dataset have time interval that lie between 10 seconds and 20 seconds with 37 car crossings. From the plot it can be concluded that the dataset has a right-skewed distribution since the tail of the plot is leading to the right side.

From the above analysis, it can be concluded that dataset 2 **doesn't have a normal distribution** of data and **it is not an exponentially decreasing distribution**.

CHI-SQUARE ANALYSIS

DATA:

Set 2 consists of data regarding clock times of 110 cars when crossing an apartment. Data for both data sets were collected and stored as excel files. The data was collected, and the following analysis was done on a Saturday in the month of September 2023.

PROCEDURE: The dataset was analyzed, and the sample mean of the dataset was calculated to be 25.93 seconds. The sample standard deviation was also calculated to be 23.44 seconds. Dataset 2 is assumed to have an exponential distribution.

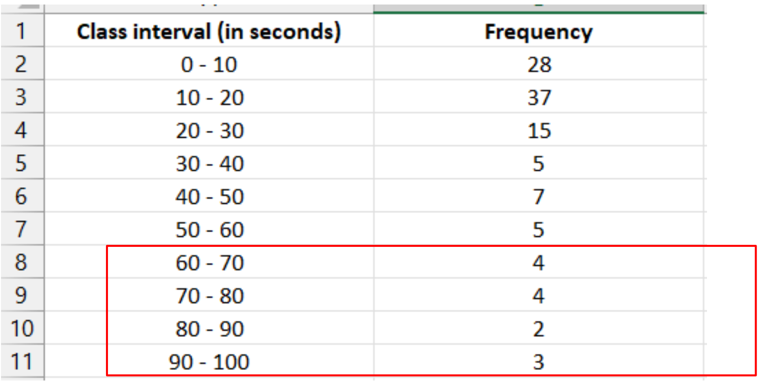
HYPOTHESIS:

H_0 (Null hypothesis) = Dataset 2 is exponentially distributed.

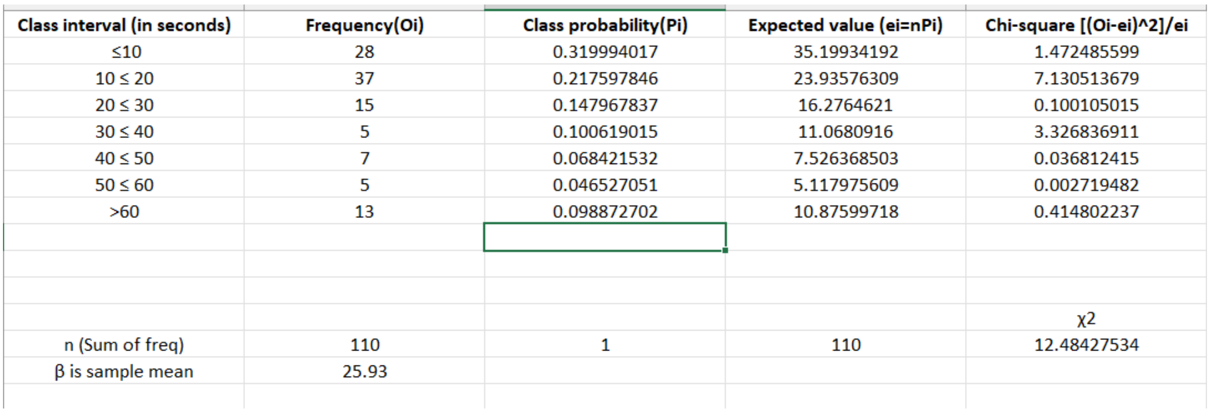
H_1 (Alternate Hypothesis) = Dataset 2 is not exponentially distributed.

COMBINING INTERVALS:

Originally, the dataset had 11 class intervals. During calculation of chi-squared values, the expected value(ei) values of intervals after 2,53,51,000 subscribers were coming out to be less than 5. So, these intervals were combined, and their respective frequencies were added which resulted in a new interval.



1	Class interval (in seconds)	Frequency
2	0 - 10	28
3	10 - 20	37
4	20 - 30	15
5	30 - 40	5
6	40 - 50	7
7	50 - 60	5
8	60 - 70	4
9	70 - 80	4
10	80 - 90	2
11	90 - 100	3



Class interval (in seconds)	Frequency(Oi)	Class probability(Pi)	Expected value (ei=nPi)	Chi-square [(Oi-ei)^2]/ei
≤10	28	0.319994017	35.19934192	1.472485599
10 ≤ 20	37	0.217597846	23.93576309	7.130513679
20 ≤ 30	15	0.147967837	16.2764621	0.100105015
30 ≤ 40	5	0.100619015	11.0680916	3.326836911
40 ≤ 50	7	0.068421532	7.526368503	0.036812415
50 ≤ 60	5	0.046527051	5.117975609	0.002719482
>60	13	0.098872702	10.87599718	0.414802237
				χ2
n (Sum of freq)	110	1	110	12.48427534
β is sample mean	25.93			

The class probability (P_i) was calculated using the formula:

GAMMADIST (x, α=1, mean(β), 1)

The following is the formula used for first class interval: **= GAMMADIST (10,1,25.93,1)**

For the second interval and upcoming intervals, the formula is as follows:

= GAMMADIST (upper limit, α, mean(β), 1) - GAMMADIST (lower limit, α, mean(β), 1)

=GAMMADIST (20,1,25.93,1) - GAMMADIST (10,1,25.93,1) (for 2nd interval)

For the last interval, the formula is as follows: = **1 - GAMMADIST (60,1,25.93,1)**

The expected value of the class intervals (ei) was calculated using the formula: **ei=n*Pi** where n is the number of observations(n=110).

For calculating Chi-square, sum of [(Oi – ei)^2]/2 values for all intervals were calculated. The summation of these values was done to find the chi-square value of the dataset. **χ2= Σ [(Oi – ei)^2]/2.**

In the above chi-square formula Oi is the frequency of observations for that class interval.

From the table above, the Chi-square χ^2 value was computed to be 12.484. When chi-square value was calculated using critical values of chi-squared distributions (Table A5), for **α=0.05** and degrees of freedom v= [(number of classes(k) – 1], i.e., 6, the value $\chi^2_{\alpha, k-1}$ was found to be 12.592.

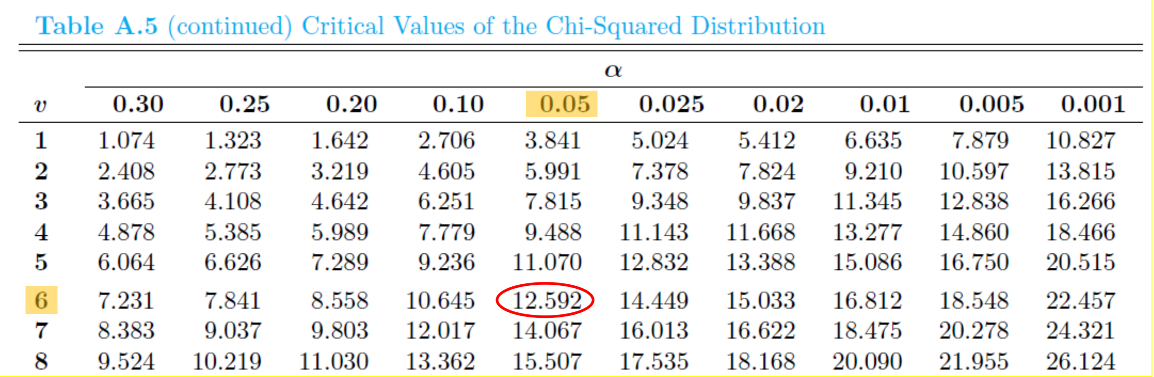
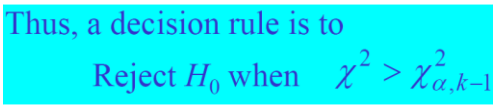


Table A.5 (continued) Critical Values of the Chi-Squared Distribution

	α									
v	0.30	0.25	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.001
1	1.074	1.323	1.642	2.706	3.841	5.024	5.412	6.635	7.879	10.827
2	2.408	2.773	3.219	4.605	5.991	7.378	7.824	9.210	10.597	13.815
3	3.665	4.108	4.642	6.251	7.815	9.348	9.837	11.345	12.838	16.266
4	4.878	5.385	5.989	7.779	9.488	11.143	11.668	13.277	14.860	18.466
5	6.064	6.626	7.289	9.236	11.070	12.832	13.388	15.086	16.750	20.515
6	7.231	7.841	8.558	10.645	12.592	14.449	15.033	16.812	18.548	22.457
7	8.383	9.037	9.803	12.017	14.067	16.013	16.622	18.475	20.278	24.321
8	9.524	10.219	11.030	13.362	15.507	17.535	18.168	20.090	21.955	26.124

<u>Decision Rule</u>:



Thus, a decision rule is to

Reject H_0 when $\chi^2 > \chi^2_{\alpha,k-1}$

Here, H_0 (Null hypothesis) = Dataset 2 has is exponentially distributed.

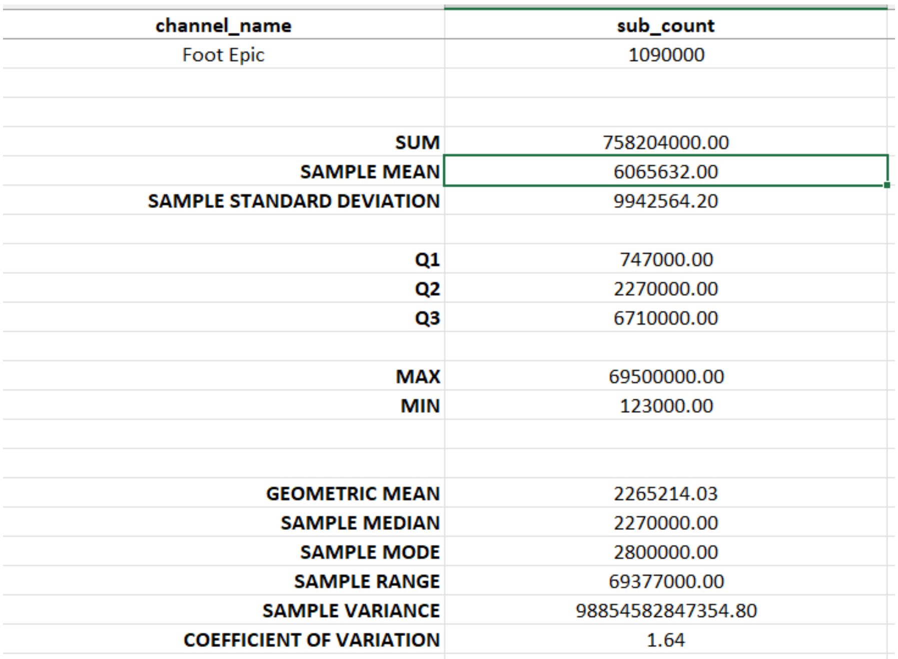
H_1 (Alternate Hypothesis) = Dataset 2 is not exponentially distributed.

Therefore, **χ^2 is not greater than $\chi^2_{\alpha, k-1}$, We fail reject H_0**.

- We are 95% confident that dataset 2 follows exponential distribution.

APPENDIX 1

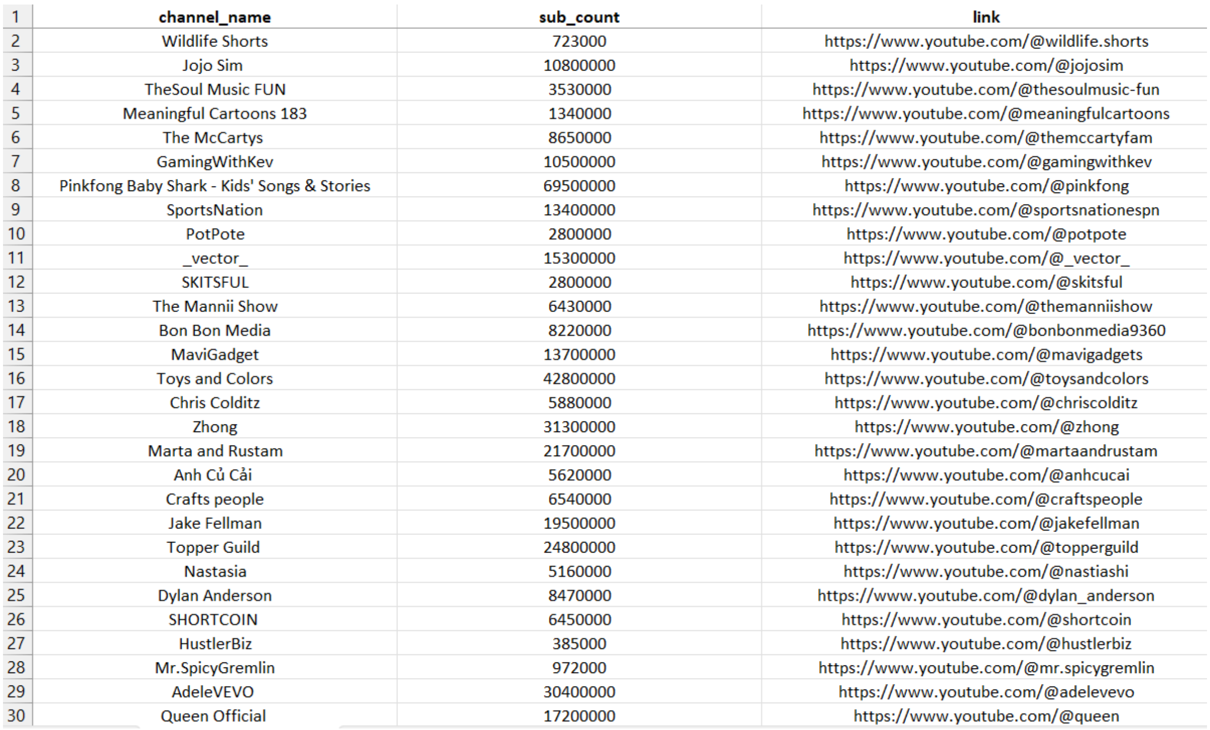
- The data for project part 2 was derived from project part 1.

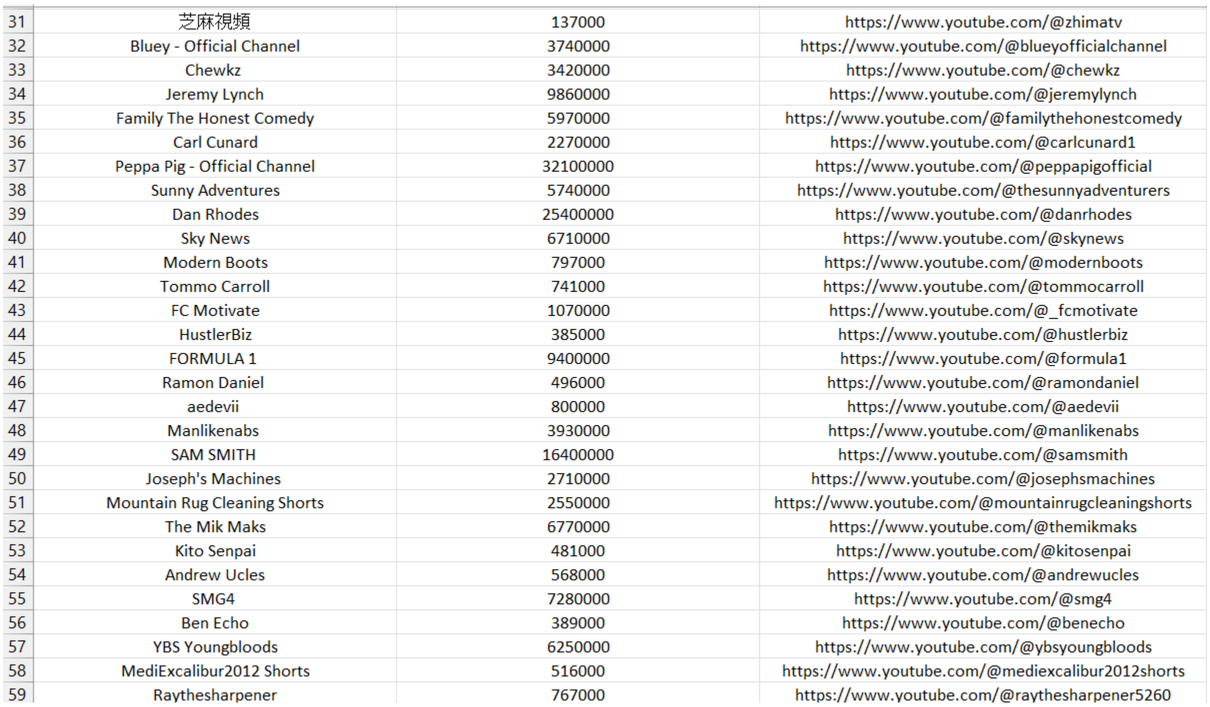


channel_name	sub_count
Foot Epic	1090000
SUM	758204000.00
SAMPLE MEAN	6065632.00
SAMPLE STANDARD DEVIATION	9942564.20
Q1	747000.00
Q2	2270000.00
Q3	6710000.00
MAX	69500000.00
MIN	123000.00
GEOMETRIC MEAN	2265214.03
SAMPLE MEDIAN	2270000.00
SAMPLE MODE	2800000.00
SAMPLE RANGE	69377000.00
SAMPLE VARIANCE	98854582847354.80
COEFFICIENT OF VARIATION	1.64

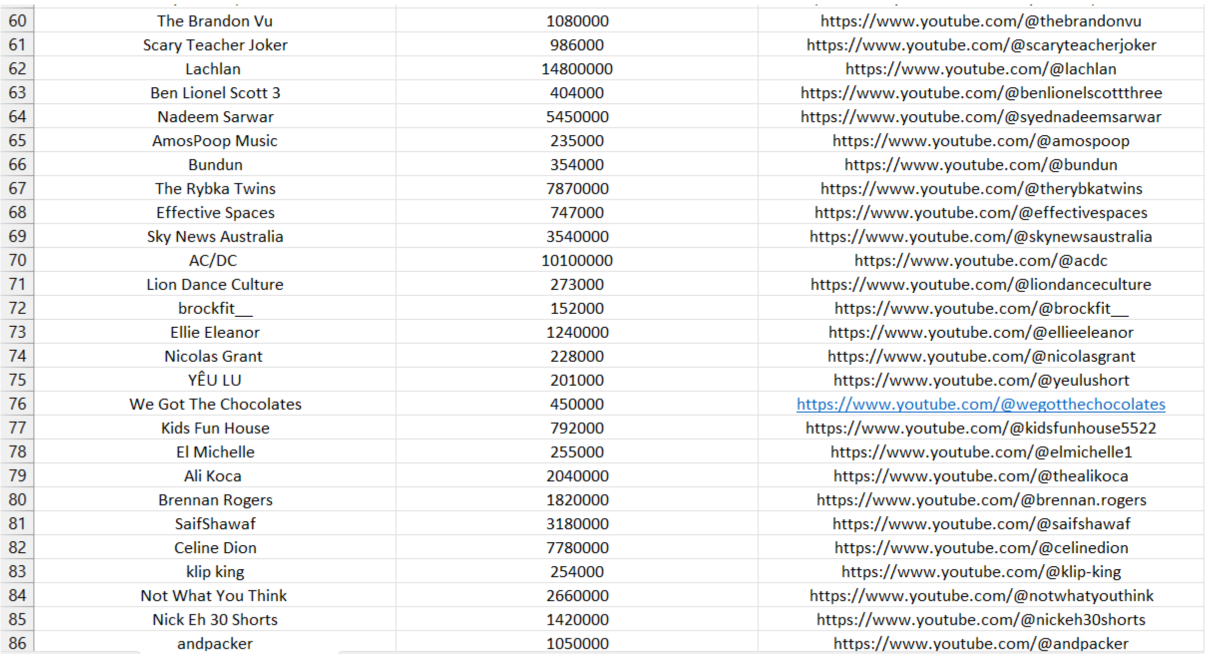
- Excel formulas for calculating Pi in dataset 1:
  =NORMDIST(6430,6065.632,9942.5642,1)
  =NORMDIST(12737,6065.632,9942.5642,1)-NORMDIST(6430,6065.632,9942.5642,1)
  =NORMDIST(19044,6065.632,9942.5642,1)-NORMDIST(12737,6065.632,9942.5642,1)
  =1-NORMDIST(19044,6065.632,9942.5642,1)

- Raw dataset of YouTube channels and their subscriber count. The list of top 100 YouTube channels was sourced from https://socialblade.com/ and the correct subscriber count was updated using https://www.youtube.com/ .

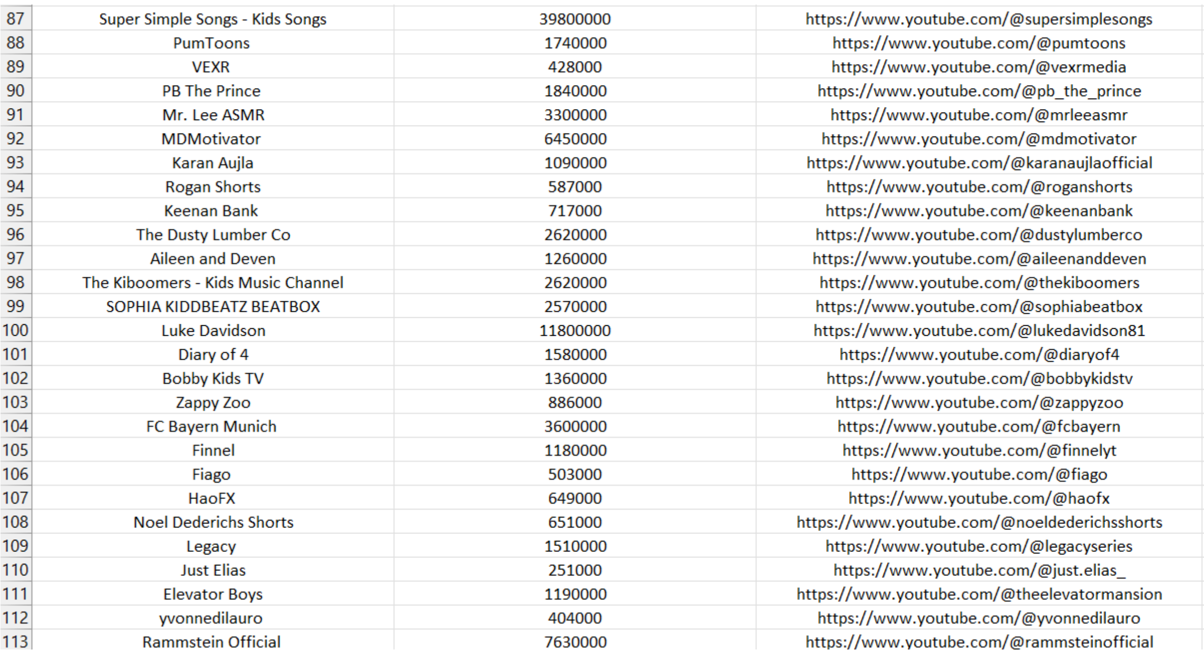




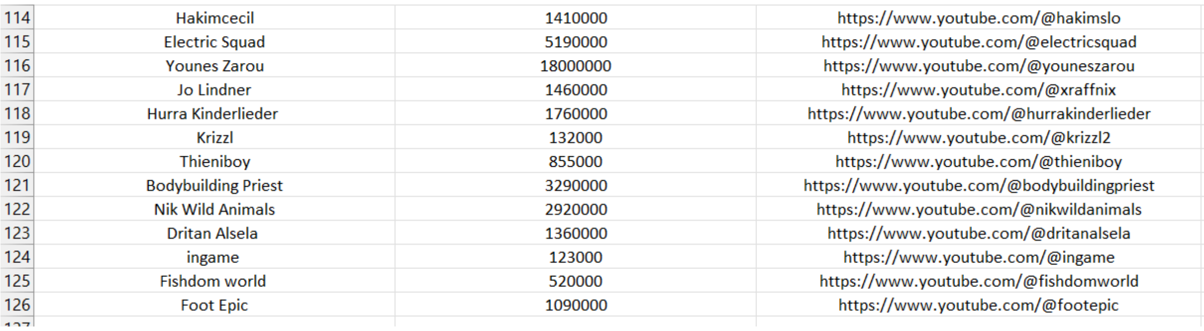
1	channel_name	sub_count	link
2	Wildlife Shorts	723000	https://www.youtube.com/@wildlife.shorts
3	Jojo Sim	10800000	https://www.youtube.com/@jojosim
4	TheSoul Music FUN	3530000	https://www.youtube.com/@thesoulmusic-fun
5	Meaningful Cartoons 183	1340000	https://www.youtube.com/@meaningfulcartoons
6	The McCartys	8650000	https://www.youtube.com/@themccartyfam
7	GamingWithKev	10500000	https://www.youtube.com/@gamingwithkev
8	Pinkfong Baby Shark - Kids' Songs & Stories	69500000	https://www.youtube.com/@pinkfong
9	SportsNation	13400000	https://www.youtube.com/@sportsnationespn
10	PotPote	2800000	https://www.youtube.com/@potpote
11	_vector_	15300000	https://www.youtube.com/@_vector_
12	SKITSFUL	2800000	https://www.youtube.com/@skitsful
13	The Mannii Show	6430000	https://www.youtube.com/@themanniishow
14	Bon Bon Media	8220000	https://www.youtube.com/@bonbonmedia9360
15	MaviGadget	13700000	https://www.youtube.com/@mavigadgets
16	Toys and Colors	42800000	https://www.youtube.com/@toysandcolors
17	Chris Colditz	5880000	https://www.youtube.com/@chriscolditz
18	Zhong	31300000	https://www.youtube.com/@zhong
19	Marta and Rustam	21700000	https://www.youtube.com/@martaandrustam
20	Anh Củ Cải	5620000	https://www.youtube.com/@anhcucai
21	Crafts people	6540000	https://www.youtube.com/@craftspeople
22	Jake Fellman	19500000	https://www.youtube.com/@jakefellman
23	Topper Guild	24800000	https://www.youtube.com/@topperguild
24	Nastasia	5160000	https://www.youtube.com/@nastiashi
25	Dylan Anderson	8470000	https://www.youtube.com/@dylan_anderson
26	SHORTCOIN	6450000	https://www.youtube.com/@shortcoin
27	HustlerBiz	385000	https://www.youtube.com/@hustlerbiz
28	Mr.SpicyGremlin	972000	https://www.youtube.com/@mr.spicygremlin
29	AdeleVEVO	30400000	https://www.youtube.com/@adelevevo
30	Queen Official	17200000	https://www.youtube.com/@queen
31	芝麻視頻	137000	https://www.youtube.com/@zhimatv
32	Bluey - Official Channel	3740000	https://www.youtube.com/@blueyofficialchannel
33	Chewkz	3420000	https://www.youtube.com/@chewkz
34	Jeremy Lynch	9860000	https://www.youtube.com/@jeremylynch
35	Family The Honest Comedy	5970000	https://www.youtube.com/@familythehonestcomedy
36	Carl Cunard	2270000	https://www.youtube.com/@carlcunard1
37	Peppa Pig - Official Channel	32100000	https://www.youtube.com/@peppapigofficial
38	Sunny Adventures	5740000	https://www.youtube.com/@thesunnyadventurers
39	Dan Rhodes	25400000	https://www.youtube.com/@danrhodes
40	Sky News	6710000	https://www.youtube.com/@skynews
41	Modern Boots	797000	https://www.youtube.com/@modernboots
42	Tommo Carroll	741000	https://www.youtube.com/@tommocarroll
43	FC Motivate	1070000	https://www.youtube.com/@_fcmotivate
44	HustlerBiz	385000	https://www.youtube.com/@hustlerbiz
45	FORMULA 1	9400000	https://www.youtube.com/@formula1
46	Ramon Daniel	496000	https://www.youtube.com/@ramondaniel
47	aedevii	800000	https://www.youtube.com/@aedevii
48	Manlikenabs	3930000	https://www.youtube.com/@manlikenabs
49	SAM SMITH	16400000	https://www.youtube.com/@samsmith
50	Joseph's Machines	2710000	https://www.youtube.com/@josephsmachines
51	Mountain Rug Cleaning Shorts	2550000	https://www.youtube.com/@mountainrugcleaningshorts
52	The Mik Maks	6770000	https://www.youtube.com/@themikmaks
53	Kito Senpai	481000	https://www.youtube.com/@kitosenpai
54	Andrew Ucles	568000	https://www.youtube.com/@andrewucles
55	SMG4	7280000	https://www.youtube.com/@smg4
56	Ben Echo	389000	https://www.youtube.com/@benecho
57	YBS Youngbloods	6250000	https://www.youtube.com/@ybsyoungbloods
58	MediExcalibur2012 Shorts	516000	https://www.youtube.com/@mediexcalibur2012shorts
59	Raythesharpener	767000	https://www.youtube.com/@raythesharpener5260



60	The Brandon Vu	1080000	https://www.youtube.com/@thebrandonvu
61	Scary Teacher Joker	986000	https://www.youtube.com/@scaryteacherjoker
62	Lachlan	14800000	https://www.youtube.com/@lachlan
63	Ben Lionel Scott 3	404000	https://www.youtube.com/@benlionelscottthree
64	Nadeem Sarwar	5450000	https://www.youtube.com/@syednadeemsarwar
65	AmosPoop Music	235000	https://www.youtube.com/@amospoop
66	Bundun	354000	https://www.youtube.com/@bundun
67	The Rybka Twins	7870000	https://www.youtube.com/@therybkatwins
68	Effective Spaces	747000	https://www.youtube.com/@effectivespaces
69	Sky News Australia	3540000	https://www.youtube.com/@skynewsaustralia
70	AC/DC	10100000	https://www.youtube.com/@acdc
71	Lion Dance Culture	273000	https://www.youtube.com/@liondanceculture
72	brockfit__	152000	https://www.youtube.com/@brockfit__
73	Ellie Eleanor	1240000	https://www.youtube.com/@ellieeleanor
74	Nicolas Grant	228000	https://www.youtube.com/@nicolasgrant
75	YÊU LU	201000	https://www.youtube.com/@yeulushort
76	We Got The Chocolates	450000	https://www.youtube.com/@wegotthechocolates
77	Kids Fun House	792000	https://www.youtube.com/@kidsfunhouse5522
78	El Michelle	255000	https://www.youtube.com/@elmichelle1
79	Ali Koca	2040000	https://www.youtube.com/@thealikoca
80	Brennan Rogers	1820000	https://www.youtube.com/@brennan.rogers
81	SaifShawaf	3180000	https://www.youtube.com/@saifshawaf
82	Celine Dion	7780000	https://www.youtube.com/@celinedion
83	klip king	254000	https://www.youtube.com/@klip-king
84	Not What You Think	2660000	https://www.youtube.com/@notwhatyouthink
85	Nick Eh 30 Shorts	1420000	https://www.youtube.com/@nickeh30shorts
86	andpacker	1050000	https://www.youtube.com/@andpacker



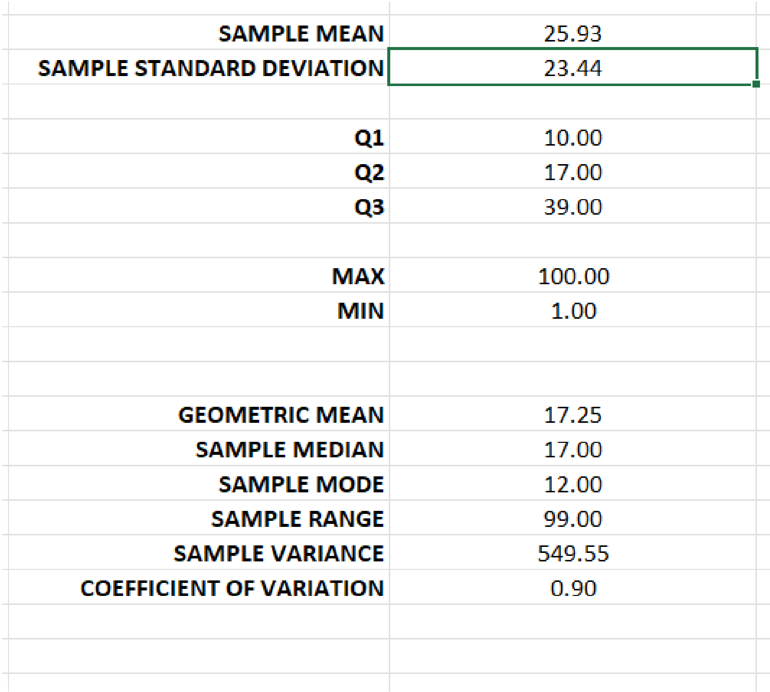
87	Super Simple Songs - Kids Songs	39800000	https://www.youtube.com/@supersimplesongs
88	PumToons	1740000	https://www.youtube.com/@pumtoons
89	VEXR	428000	https://www.youtube.com/@vexrmedia
90	PB The Prince	1840000	https://www.youtube.com/@pb_the_prince
91	Mr. Lee ASMR	3300000	https://www.youtube.com/@mrleeasmr
92	MDMotivator	6450000	https://www.youtube.com/@mdmotivator
93	Karan Aujla	1090000	https://www.youtube.com/@karanaujlaofficial
94	Rogan Shorts	587000	https://www.youtube.com/@roganshorts
95	Keenan Bank	717000	https://www.youtube.com/@keenanbank
96	The Dusty Lumber Co	2620000	https://www.youtube.com/@dustylumberco
97	Aileen and Deven	1260000	https://www.youtube.com/@aileenanddeven
98	The Kiboomers - Kids Music Channel	2620000	https://www.youtube.com/@thekiboomers
99	SOPHIA KIDDBEATZ BEATBOX	2570000	https://www.youtube.com/@sophiabeatbox
100	Luke Davidson	11800000	https://www.youtube.com/@lukedavidson81
101	Diary of 4	1580000	https://www.youtube.com/@diaryof4
102	Bobby Kids TV	1360000	https://www.youtube.com/@bobbykidstv
103	Zappy Zoo	886000	https://www.youtube.com/@zappyzoo
104	FC Bayern Munich	3600000	https://www.youtube.com/@fcbayern
105	Finnel	1180000	https://www.youtube.com/@finnelyt
106	Fiago	503000	https://www.youtube.com/@fiago
107	HaoFX	649000	https://www.youtube.com/@haofx
108	Noel Dederichs Shorts	651000	https://www.youtube.com/@noeldederichsshorts
109	Legacy	1510000	https://www.youtube.com/@legacyseries
110	Just Elias	251000	https://www.youtube.com/@just.elias_
111	Elevator Boys	1190000	https://www.youtube.com/@theelevatormansion
112	yvonnedilauro	404000	https://www.youtube.com/@yvonnedilauro
113	Rammstein Official	7630000	https://www.youtube.com/@rammsteinofficial



114	Hakimcecil	1410000	https://www.youtube.com/@hakimslo
115	Electric Squad	5190000	https://www.youtube.com/@electricsquad
116	Younes Zarou	18000000	https://www.youtube.com/@youneszarou
117	Jo Lindner	1460000	https://www.youtube.com/@xraffnix
118	Hurra Kinderlieder	1760000	https://www.youtube.com/@hurrakinderlieder
119	Krizzl	132000	https://www.youtube.com/@krizzl2
120	Thieniboy	855000	https://www.youtube.com/@thieniboy
121	Bodybuilding Priest	3290000	https://www.youtube.com/@bodybuildingpriest
122	Nik Wild Animals	2920000	https://www.youtube.com/@nikwildanimals
123	Dritan Alsela	1360000	https://www.youtube.com/@dritanalsela
124	ingame	123000	https://www.youtube.com/@ingame
125	Fishdom world	520000	https://www.youtube.com/@fishdomworld
126	Foot Epic	1090000	https://www.youtube.com/@footepic

APPENDIX 2

- The data for project part 2 was derived from project part 1.



SAMPLE MEAN	25.93
SAMPLE STANDARD DEVIATION	23.44
Q1	10.00
Q2	17.00
Q3	39.00
MAX	100.00
MIN	1.00
GEOMETRIC MEAN	17.25
SAMPLE MEDIAN	17.00
SAMPLE MODE	12.00
SAMPLE RANGE	99.00
SAMPLE VARIANCE	549.55
COEFFICIENT OF VARIATION	0.90

- Excel formulas for calculating Pi:

=GAMMADIST(10,1,25.93,1)

=GAMMADIST(20,1,25.93,1)-GAMMADIST(10,1,25.93,1)

=GAMMADIST(30,1,25.93,1)-GAMMADIST(20,1,25.93,1)

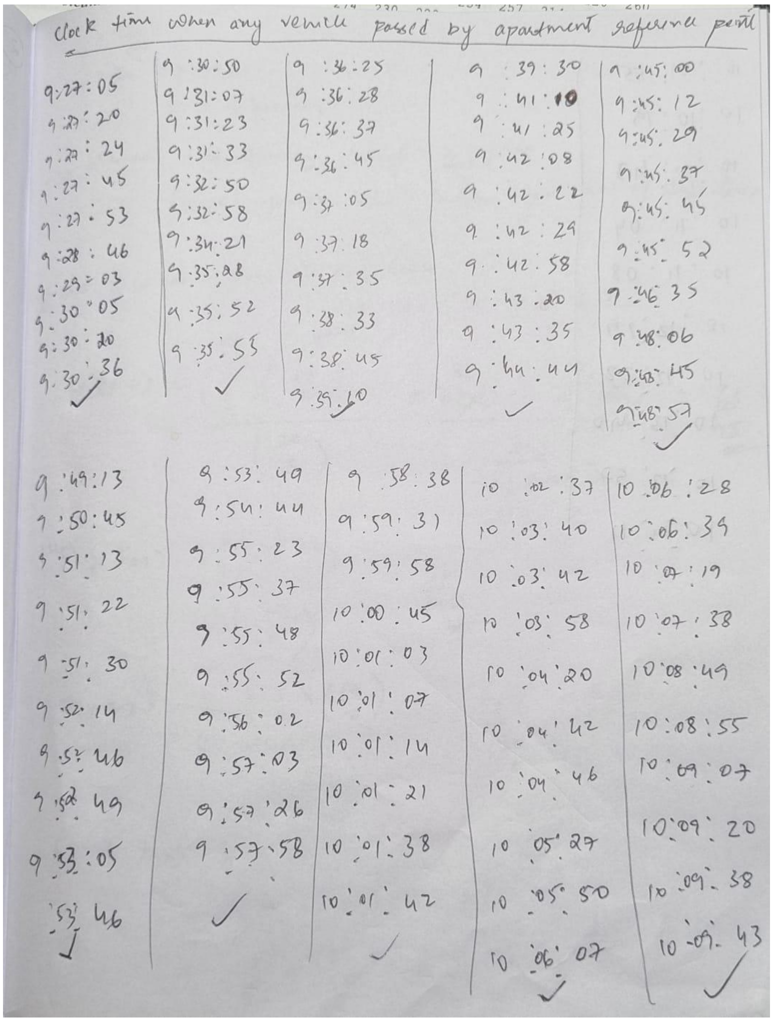
=GAMMADIST(40,1,25.93,1)-GAMMADIST(30,1,25.93,1)

=GAMMADIST(50,1,25.93,1)-GAMMADIST(40,1,25.93,1)

=GAMMADIST(60,1,25.93,1)-GAMMADIST(50,1,25.93,1)

=1-GAMMADIST(60,1,25.93,1)

- The raw dataset for clock times of cars crossing an apartment was recorded by the observer in real-time.



clock time when any vehicle passed by apartment reference point

9:27:05
9:27:20
9:27:24
9:27:45
9:27:53
9:28:46
9:29:03
9:30:05
9:30:20
9:30:36
✓

9:30:50
9:31:07
9:31:23
9:31:33
9:32:50
9:32:58
9:34:21
9:35:28
9:35:52
9:35:53
✓

9:36:25
9:36:28
9:36:37
9:36:45
9:37:05
9:37:18
9:37:35
9:38:33
9:38:45
9:39:10
✓

9:39:30
9:41:10
9:41:25
9:42:08
9:42:22
9:42:29
9:42:58
9:43:20
9:43:35
9:44:44
✓

9:45:00
9:45:12
9:45:29
9:45:37
9:45:45
9:45:52
9:46:35
9:48:06
9:48:45
9:48:57
✓

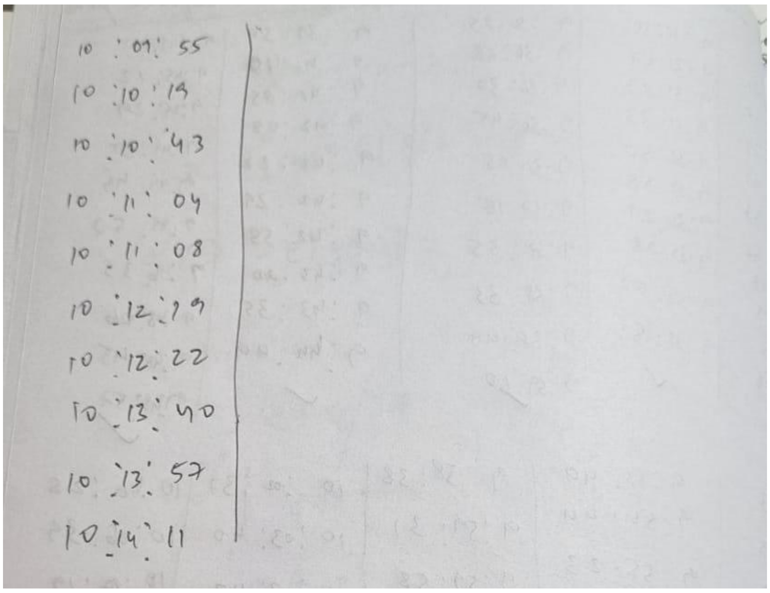
9:49:13
9:50:45
9:51:13
9:51:22
9:51:30
9:52:14
9:52:46
9:52:49
9:53:05
:53:46
✓

9:53:49
9:54:44
9:55:23
9:55:37
9:55:48
9:55:52
9:56:02
9:57:03
9:57:26
9:57:58
✓

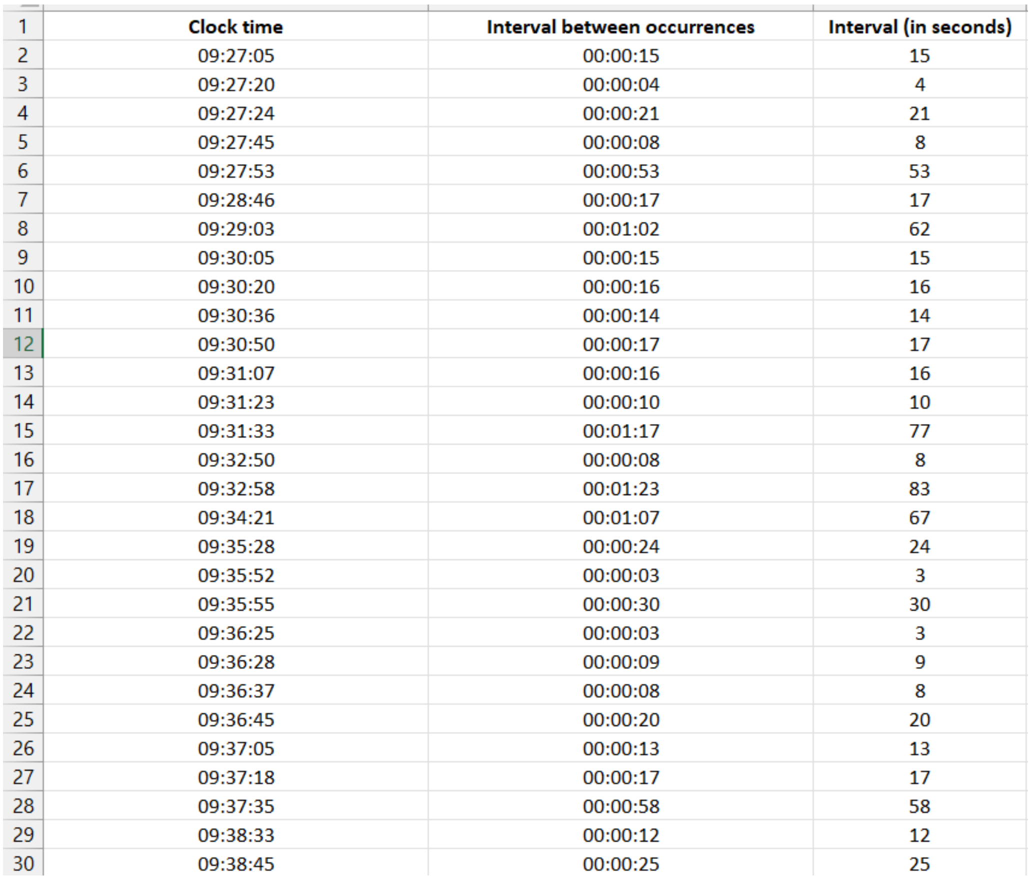
9:58:38
9:59:31
9:59:58
10:00:45
10:01:03
10:01:07
10:01:14
10:01:21
10:01:38
10:01:42
✓

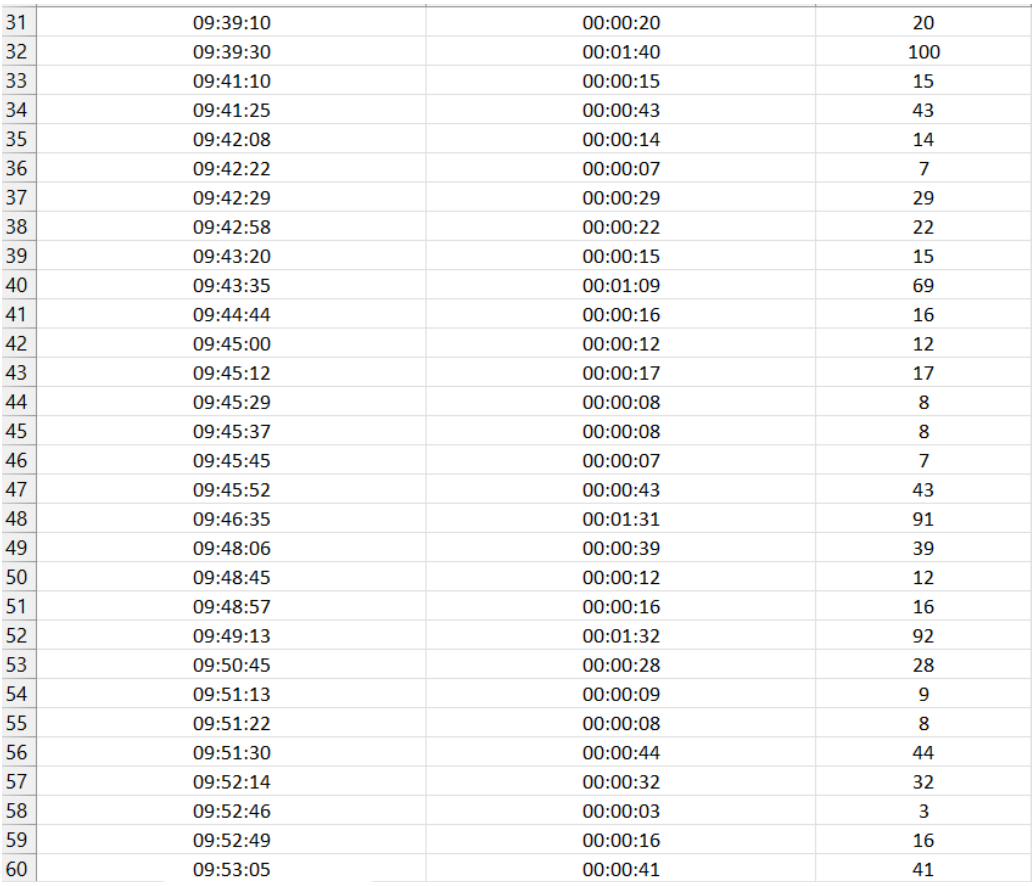
10:02:37
10:03:40
10:03:42
10:03:58
10:04:20
10:04:42
10:04:46
10:05:27
10:05:50
10:06:07
✓

10:06:28
10:06:39
10:07:19
10:07:38
10:08:49
10:08:55
10:09:07
10:09:20
10:09:38
10:09:43
✓

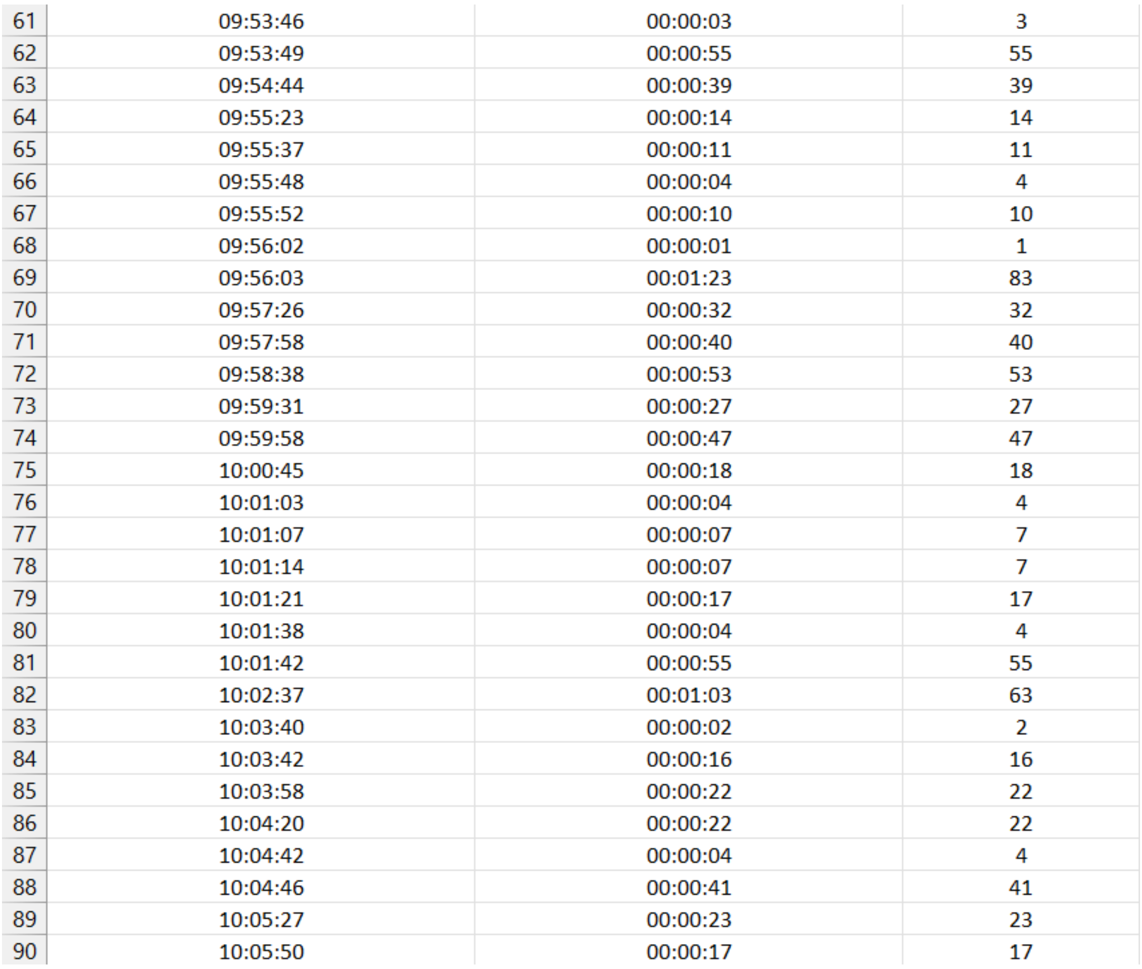


10:09:55
10:10:19
10:10:43
10:11:04
10:11:08
10:12:19
10:12:22
10:13:40
10:13:57
10:14:11

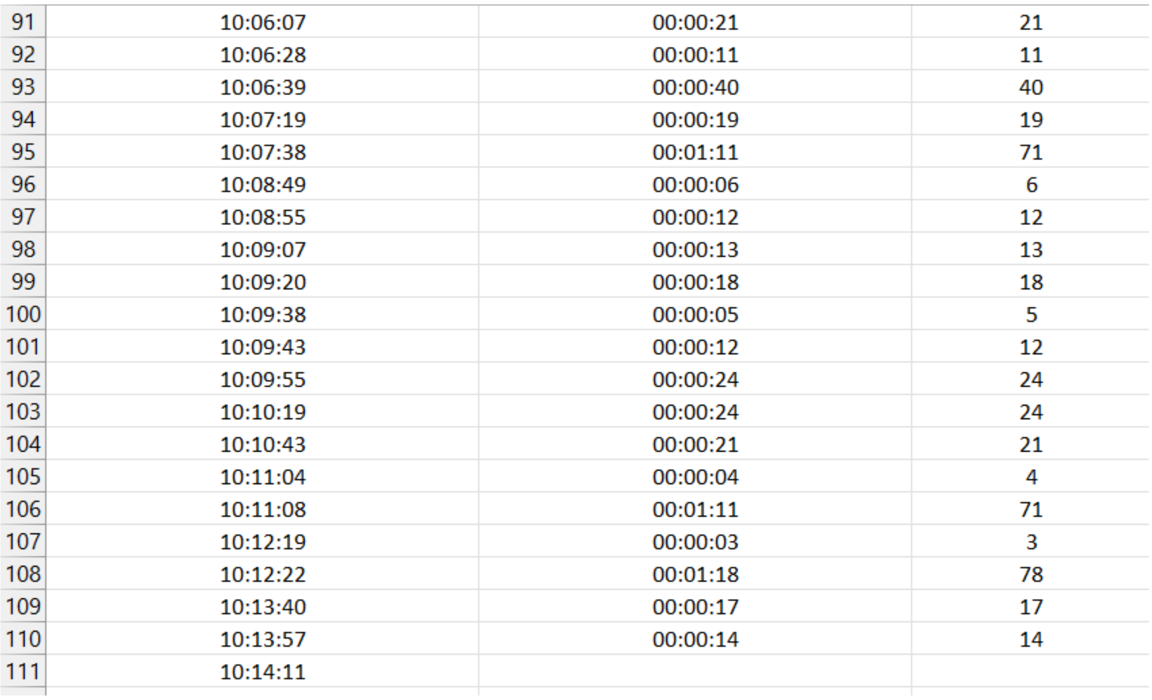




1	Clock time	Interval between occurrences	Interval (in seconds)
2	09:27:05	00:00:15	15
3	09:27:20	00:00:04	4
4	09:27:24	00:00:21	21
5	09:27:45	00:00:08	8
6	09:27:53	00:00:53	53
7	09:28:46	00:00:17	17
8	09:29:03	00:01:02	62
9	09:30:05	00:00:15	15
10	09:30:20	00:00:16	16
11	09:30:36	00:00:14	14
12	09:30:50	00:00:17	17
13	09:31:07	00:00:16	16
14	09:31:23	00:00:10	10
15	09:31:33	00:01:17	77
16	09:32:50	00:00:08	8
17	09:32:58	00:01:23	83
18	09:34:21	00:01:07	67
19	09:35:28	00:00:24	24
20	09:35:52	00:00:03	3
21	09:35:55	00:00:30	30
22	09:36:25	00:00:03	3
23	09:36:28	00:00:09	9
24	09:36:37	00:00:08	8
25	09:36:45	00:00:20	20
26	09:37:05	00:00:13	13
27	09:37:18	00:00:17	17
28	09:37:35	00:00:58	58
29	09:38:33	00:00:12	12
30	09:38:45	00:00:25	25
31	09:39:10	00:00:20	20
32	09:39:30	00:01:40	100
33	09:41:10	00:00:15	15
34	09:41:25	00:00:43	43
35	09:42:08	00:00:14	14
36	09:42:22	00:00:07	7
37	09:42:29	00:00:29	29
38	09:42:58	00:00:22	22
39	09:43:20	00:00:15	15
40	09:43:35	00:01:09	69
41	09:44:44	00:00:16	16
42	09:45:00	00:00:12	12
43	09:45:12	00:00:17	17
44	09:45:29	00:00:08	8
45	09:45:37	00:00:08	8
46	09:45:45	00:00:07	7
47	09:45:52	00:00:43	43
48	09:46:35	00:01:31	91
49	09:48:06	00:00:39	39
50	09:48:45	00:00:12	12
51	09:48:57	00:00:16	16
52	09:49:13	00:01:32	92
53	09:50:45	00:00:28	28
54	09:51:13	00:00:09	9
55	09:51:22	00:00:08	8
56	09:51:30	00:00:44	44
57	09:52:14	00:00:32	32
58	09:52:46	00:00:03	3
59	09:52:49	00:00:16	16
60	09:53:05	00:00:41	41



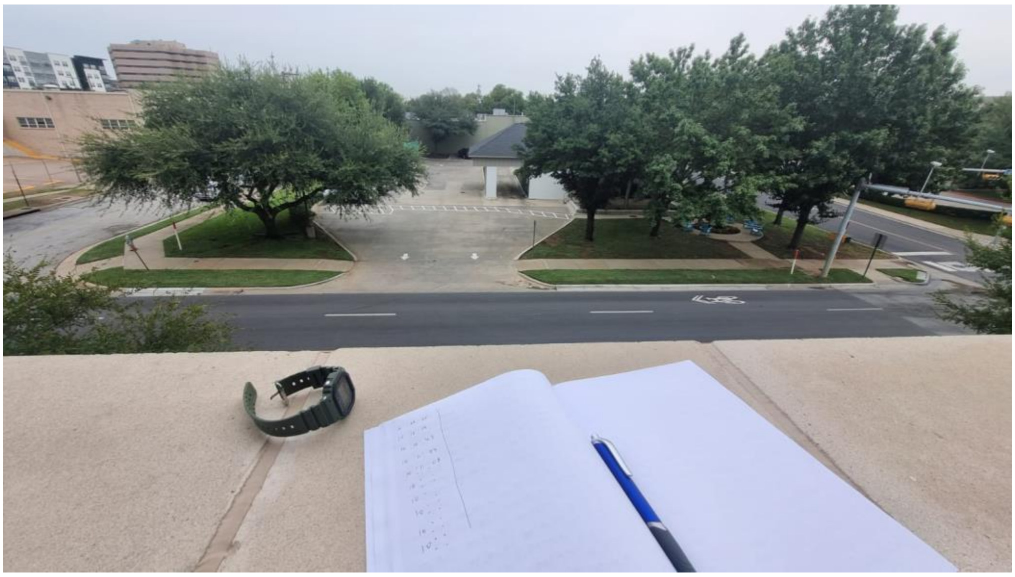
61	09:53:46	00:00:03	3
62	09:53:49	00:00:55	55
63	09:54:44	00:00:39	39
64	09:55:23	00:00:14	14
65	09:55:37	00:00:11	11
66	09:55:48	00:00:04	4
67	09:55:52	00:00:10	10
68	09:56:02	00:00:01	1
69	09:56:03	00:01:23	83
70	09:57:26	00:00:32	32
71	09:57:58	00:00:40	40
72	09:58:38	00:00:53	53
73	09:59:31	00:00:27	27
74	09:59:58	00:00:47	47
75	10:00:45	00:00:18	18
76	10:01:03	00:00:04	4
77	10:01:07	00:00:07	7
78	10:01:14	00:00:07	7
79	10:01:21	00:00:17	17
80	10:01:38	00:00:04	4
81	10:01:42	00:00:55	55
82	10:02:37	00:01:03	63
83	10:03:40	00:00:02	2
84	10:03:42	00:00:16	16
85	10:03:58	00:00:22	22
86	10:04:20	00:00:22	22
87	10:04:42	00:00:04	4
88	10:04:46	00:00:41	41
89	10:05:27	00:00:23	23
90	10:05:50	00:00:17	17

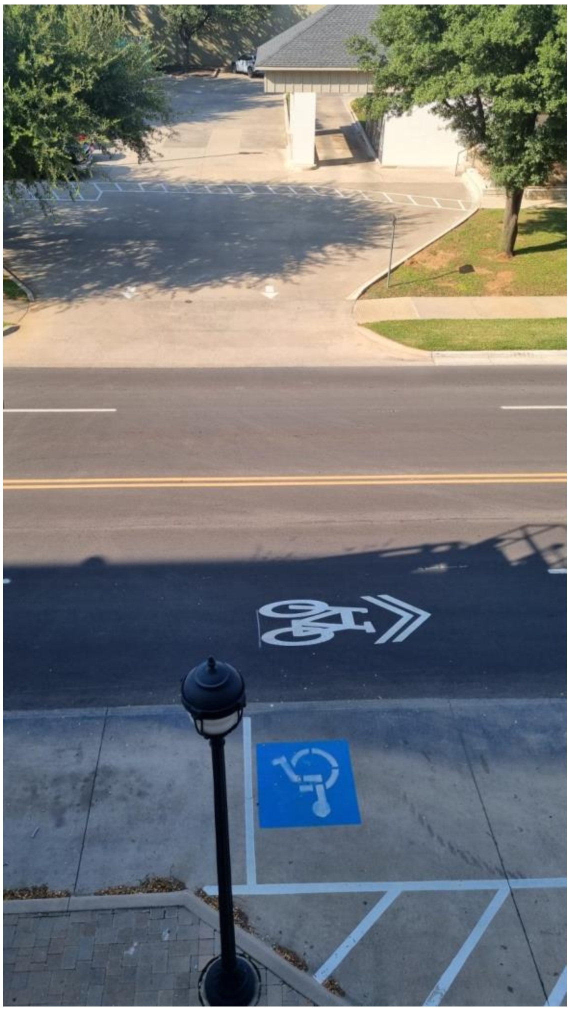


91	10:06:07	00:00:21	21
92	10:06:28	00:00:11	11
93	10:06:39	00:00:40	40
94	10:07:19	00:00:19	19
95	10:07:38	00:01:11	71
96	10:08:49	00:00:06	6
97	10:08:55	00:00:12	12
98	10:09:07	00:00:13	13
99	10:09:20	00:00:18	18
100	10:09:38	00:00:05	5
101	10:09:43	00:00:12	12
102	10:09:55	00:00:24	24
103	10:10:19	00:00:24	24
104	10:10:43	00:00:21	21
105	10:11:04	00:00:04	4
106	10:11:08	00:01:11	71
107	10:12:19	00:00:03	3
108	10:12:22	00:01:18	78
109	10:13:40	00:00:17	17
110	10:13:57	00:00:14	14
111	10:14:11		

LOCATION FOR DATA COLLECTION/ OBSERVATION:

The reference point for the observation is 404 apartments, East Border Street, opposite White Rhino café.(Location tag - https://maps.app.goo.gl/Nedr8kQezbuVmH3R6)





The lamppost in the above image was taken as reference for crossing point for the cars while collecting data.