



## PROJECT – PART 1

**Abstract:** The report describes the process of collection and descriptive analysis of two sets of data sets and arrive at insights which are used to summarize trends in the data sets.

**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."*

*x Chirag Hebbal Rudresh*

Chirag Hebbal Rudresh

**Name:** Chirag Hebbal Rudresh

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

**Date:** 9/23/2023

## **INDEX**

1. DATASET 1	Pg 2
I. Data	Pg 2
II. Process of Data collection	Pg 2
III. Descriptive Statistics	Pg 2
 2. DATASET 2	 Pg 7
I. Data	Pg 7
II. Process of Data collection	Pg 7
III. Descriptive Statistics	Pg 7
 3. Appendix 1	 Pg 12
 4. Appendix 2	 Pg 14

## 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.

<b>SAMPLE MEAN</b>	60,65,632
--------------------	-----------

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 B2 to cell B126 which are 125 values. So, the formula will be STDEV.S(C2:C110).

<b>SAMPLE STANDARD DEVIATION</b>	=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.

<b>SAMPLE STANDARD DEVIATION</b>	99,42,564
----------------------------------	-----------

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 7,47,000.

<b>Q1</b>	=QUARTILE(B2:B126,1)	<b>Q1</b>	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.

<b>Q2</b>	=QUARTILE(B2:B126,2)	<b>Q2</b>	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.

<b>Q3</b>	=QUARTILE(B2:B126,3)	<b>Q3</b>	67,10,000
-----------	----------------------	-----------	-----------

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 22,65,214.

<b>GEOMETRIC MEAN</b>	22,65,214
-----------------------	-----------

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 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).

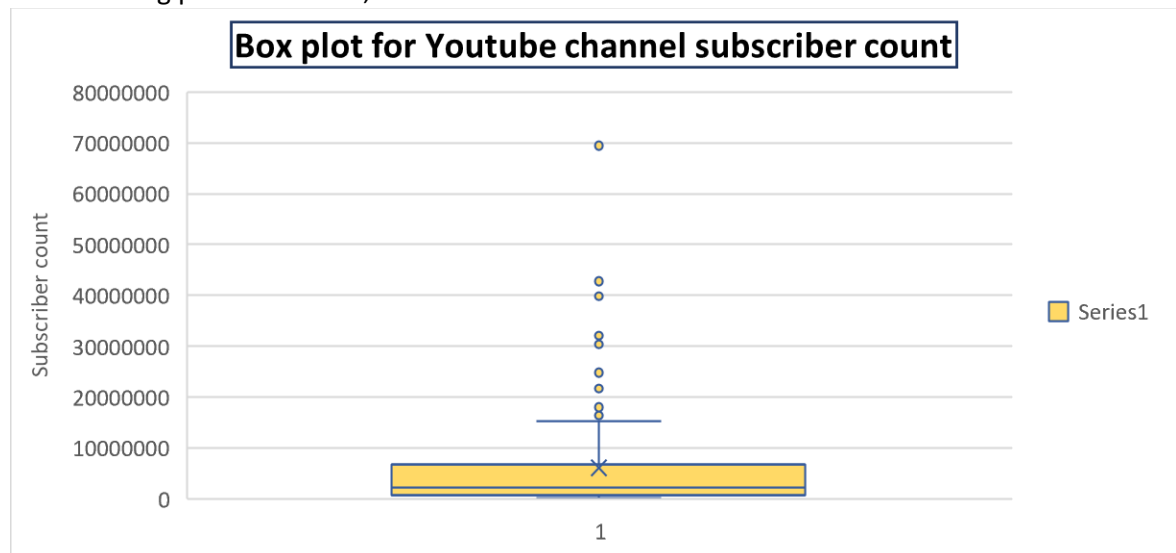
8. **Sample variance:** It provides a numerical value that indicates how much individual data points deviate from the sample mean (average). The excel formula used to calculate it is: VAR(C2:C110). The variance of the dataset is 9,88,54,58,28,47,354.80.

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.

<b>COEFFICIENT OF VARIATION</b>	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. **Frequency table and Frequency histogram:**

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.]))
```

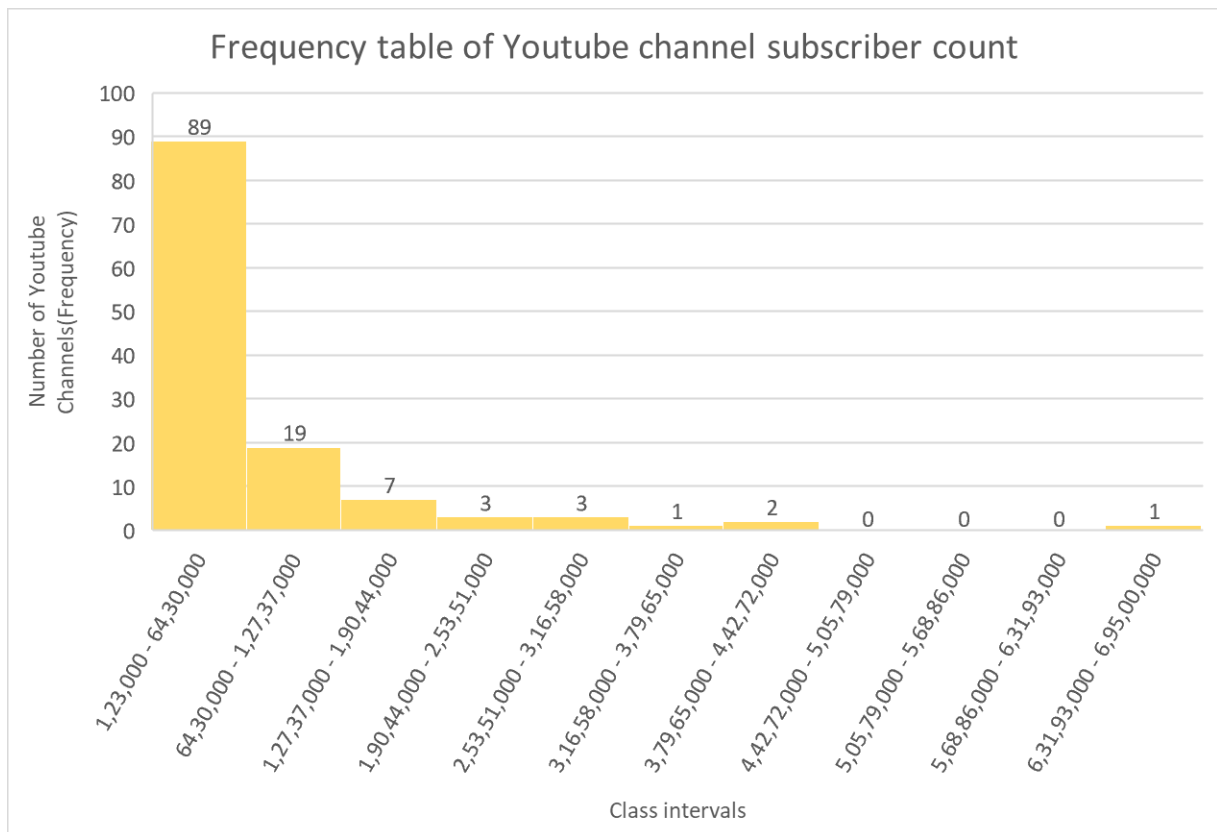
In the above code, 'sub\_count' refers to the column name that is given in the excel sheet for the subscriber count numbers and 'bins=11' refers to the number of class intervals needed from the data. Below the code, in the output, the first line has 11 numbers which represent the frequency of channel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

## **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.



<b>Q1</b>	<b>=QUARTILE(C2:C110,1)</b>	<b>Q1</b>	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.

<b>Q2</b>	<b>=QUARTILE(C2:C110,2)</b>	<b>Q2</b>	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.

<b>Q3</b>	<b>=QUARTILE(C2:C110,3)</b>	<b>Q3</b>	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.

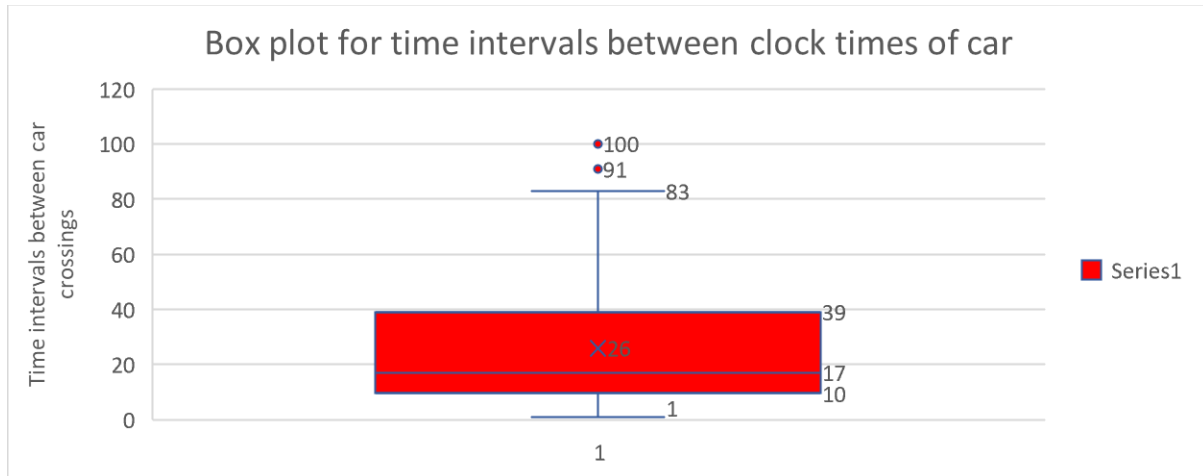
**SAMPLE VARIANCE**=VAR(C2:C110)

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.

<b>COEFFICIENT OF VARIATION</b>	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.]))
```

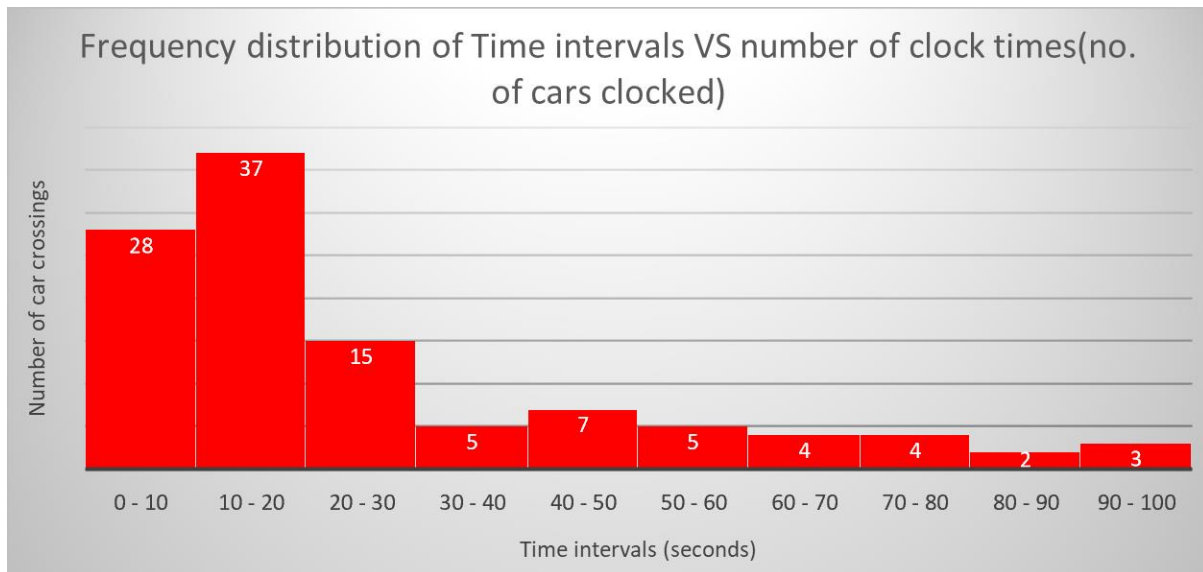
In the above code, 'Interval (in seconds)' refers to the column name that is given in the excel sheet for the time intervals and 'bins=10' refers to the number of class intervals needed from the data. In the code, in the output, the first line has 10 numbers which represent the frequency of car crossing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

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

60	The Brandon Vu	1080000	<a href="https://www.youtube.com/@thebrandonvu">https://www.youtube.com/@thebrandonvu</a>
61	Scary Teacher Joker	986000	<a href="https://www.youtube.com/@scaryteacherjoker">https://www.youtube.com/@scaryteacherjoker</a>
62	Lachlan	14800000	<a href="https://www.youtube.com/@lachlan">https://www.youtube.com/@lachlan</a>
63	Ben Lionel Scott 3	404000	<a href="https://www.youtube.com/@benlionelscottthree">https://www.youtube.com/@benlionelscottthree</a>
64	Nadeem Sarwar	5450000	<a href="https://www.youtube.com/@syednadeemsarwar">https://www.youtube.com/@syednadeemsarwar</a>
65	AmosPoop Music	235000	<a href="https://www.youtube.com/@amospoop">https://www.youtube.com/@amospoop</a>
66	Bundun	354000	<a href="https://www.youtube.com/@bundun">https://www.youtube.com/@bundun</a>
67	The Rybka Twins	7870000	<a href="https://www.youtube.com/@theyrybkatwins">https://www.youtube.com/@theyrybkatwins</a>
68	Effective Spaces	747000	<a href="https://www.youtube.com/@effectivespaces">https://www.youtube.com/@effectivespaces</a>
69	Sky News Australia	3540000	<a href="https://www.youtube.com/@skynewsaustralia">https://www.youtube.com/@skynewsaustralia</a>
70	AC/DC	10100000	<a href="https://www.youtube.com/@acdc">https://www.youtube.com/@acdc</a>
71	Lion Dance Culture	273000	<a href="https://www.youtube.com/@liondanceculture">https://www.youtube.com/@liondanceculture</a>
72	brockfit__	152000	<a href="https://www.youtube.com/@brockfit__">https://www.youtube.com/@brockfit__</a>
73	Ellie Eleanor	1240000	<a href="https://www.youtube.com/@ellieeleanor">https://www.youtube.com/@ellieeleanor</a>
74	Nicolas Grant	228000	<a href="https://www.youtube.com/@nicolasgrant">https://www.youtube.com/@nicolasgrant</a>
75	YÊU LU	201000	<a href="https://www.youtube.com/@yeulushort">https://www.youtube.com/@yeulushort</a>
76	We Got The Chocolates	450000	<a href="https://www.youtube.com/@wegothethechocolates">https://www.youtube.com/@wegothethechocolates</a>
77	Kids Fun House	792000	<a href="https://www.youtube.com/@kidsfunhouse5522">https://www.youtube.com/@kidsfunhouse5522</a>
78	El Michelle	255000	<a href="https://www.youtube.com/@elmichelle1">https://www.youtube.com/@elmichelle1</a>
79	Ali Koca	2040000	<a href="https://www.youtube.com/@thealikoca">https://www.youtube.com/@thealikoca</a>
80	Brennan Rogers	1820000	<a href="https://www.youtube.com/@brennan.rogers">https://www.youtube.com/@brennan.rogers</a>
81	SaifShawaf	3180000	<a href="https://www.youtube.com/@saifshawaf">https://www.youtube.com/@saifshawaf</a>
82	Celine Dion	7780000	<a href="https://www.youtube.com/@celinedion">https://www.youtube.com/@celinedion</a>
83	Klip King	254000	<a href="https://www.youtube.com/@klip-king">https://www.youtube.com/@klip-king</a>
84	Not What You Think	2660000	<a href="https://www.youtube.com/@notwhatyouthink">https://www.youtube.com/@notwhatyouthink</a>
85	Nick Eh 30 Shorts	1420000	<a href="https://www.youtube.com/@nickeh30shorts">https://www.youtube.com/@nickeh30shorts</a>
86	andpacker	1050000	<a href="https://www.youtube.com/@andpacker">https://www.youtube.com/@andpacker</a>
87	Super Simple Songs - Kids Songs	39800000	<a href="https://www.youtube.com/@supersimplesongs">https://www.youtube.com/@supersimplesongs</a>
88	PumToons	1740000	<a href="https://www.youtube.com/@pumtoons">https://www.youtube.com/@pumtoons</a>
89	VEXR	428000	<a href="https://www.youtube.com/@vexrmedia">https://www.youtube.com/@vexrmedia</a>
90	PB The Prince	1840000	<a href="https://www.youtube.com/@pb_the_prince">https://www.youtube.com/@pb_the_prince</a>
91	Mr. Lee ASMR	3300000	<a href="https://www.youtube.com/@mrleeasmr">https://www.youtube.com/@mrleeasmr</a>
92	MDMotivator	6450000	<a href="https://www.youtube.com/@rmdmotivator">https://www.youtube.com/@rmdmotivator</a>
93	Karan Aujla	1090000	<a href="https://www.youtube.com/@karanaujlaofficial">https://www.youtube.com/@karanaujlaofficial</a>
94	Rogan Shorts	587000	<a href="https://www.youtube.com/@roganshorts">https://www.youtube.com/@roganshorts</a>
95	Keenan Bank	717000	<a href="https://www.youtube.com/@keenanbank">https://www.youtube.com/@keenanbank</a>
96	The Dusty Lumber Co	2620000	<a href="https://www.youtube.com/@dustylumberco">https://www.youtube.com/@dustylumberco</a>
97	Aileen and Deven	1260000	<a href="https://www.youtube.com/@aileenanddeven">https://www.youtube.com/@aileenanddeven</a>
98	The Kiboomers - Kids Music Channel	2620000	<a href="https://www.youtube.com/@thekiboomers">https://www.youtube.com/@thekiboomers</a>
99	SOPHIA KIDDBEATZ BEATBOX	2570000	<a href="https://www.youtube.com/@sophiabeatbox">https://www.youtube.com/@sophiabeatbox</a>
100	Luke Davidson	11800000	<a href="https://www.youtube.com/@lukedavidson81">https://www.youtube.com/@lukedavidson81</a>
101	Diary of 4	1580000	<a href="https://www.youtube.com/@diaryof4">https://www.youtube.com/@diaryof4</a>
102	Bobby Kids TV	1360000	<a href="https://www.youtube.com/@bobbykidstv">https://www.youtube.com/@bobbykidstv</a>
103	Zappy Zoo	886000	<a href="https://www.youtube.com/@zappyzoo">https://www.youtube.com/@zappyzoo</a>
104	FC Bayern Munich	3600000	<a href="https://www.youtube.com/@fcbayern">https://www.youtube.com/@fcbayern</a>
105	Finnel	1180000	<a href="https://www.youtube.com/@finnelyt">https://www.youtube.com/@finnelyt</a>
106	Fiago	503000	<a href="https://www.youtube.com/@fiago">https://www.youtube.com/@fiago</a>
107	HaoFX	649000	<a href="https://www.youtube.com/@haoFX">https://www.youtube.com/@haoFX</a>
108	Noel Dederichs Shorts	651000	<a href="https://www.youtube.com/@noeldderichsshorts">https://www.youtube.com/@noeldderichsshorts</a>
109	Legacy	1510000	<a href="https://www.youtube.com/@legacyseries">https://www.youtube.com/@legacyseries</a>
110	Just Elias	251000	<a href="https://www.youtube.com/@just.elias_">https://www.youtube.com/@just.elias_</a>
111	Elevator Boys	1190000	<a href="https://www.youtube.com/@theelevatormansion">https://www.youtube.com/@theelevatormansion</a>
112	yvonedilaura	404000	<a href="https://www.youtube.com/@yvonedilaura">https://www.youtube.com/@yvonedilaura</a>
113	Rammstein Official	7630000	<a href="https://www.youtube.com/@rammsteinofficial">https://www.youtube.com/@rammsteinofficial</a>
114	Hakimcecil	1410000	<a href="https://www.youtube.com/@hakimslo">https://www.youtube.com/@hakimslo</a>
115	Electric Squad	5190000	<a href="https://www.youtube.com/@electricsquad">https://www.youtube.com/@electricsquad</a>
116	Younes Zarou	18000000	<a href="https://www.youtube.com/@youneszarou">https://www.youtube.com/@youneszarou</a>
117	Jo Lindner	1460000	<a href="https://www.youtube.com/@xraffnix">https://www.youtube.com/@xraffnix</a>
118	Hurra Kinderlieder	1760000	<a href="https://www.youtube.com/@hurrakinderlieder">https://www.youtube.com/@hurrakinderlieder</a>
119	Krizzl	132000	<a href="https://www.youtube.com/@krizzl2">https://www.youtube.com/@krizzl2</a>
120	Thieniboy	855000	<a href="https://www.youtube.com/@thieniboy">https://www.youtube.com/@thieniboy</a>
121	Bodybuilding Priest	3290000	<a href="https://www.youtube.com/@bodybuildingpriest">https://www.youtube.com/@bodybuildingpriest</a>
122	Nik Wild Animals	2920000	<a href="https://www.youtube.com/@nikwildanimals">https://www.youtube.com/@nikwildanimals</a>
123	Dritan Alsela	1360000	<a href="https://www.youtube.com/@dritanalsela">https://www.youtube.com/@dritanalsela</a>
124	ingame	123000	<a href="https://www.youtube.com/@ingame">https://www.youtube.com/@ingame</a>
125	Fishdom world	520000	<a href="https://www.youtube.com/@fishdomworld">https://www.youtube.com/@fishdomworld</a>
126	Foot Epic	1090000	<a href="https://www.youtube.com/@footepic">https://www.youtube.com/@footepic</a>

## APPENDIX 2

- 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	9:30:50	9:36:25	9:39:30	9:45:00
9:27:20	9:31:07	9:36:28	9:41:10	9:45:12
9:27:24	9:31:23	9:36:37	9:41:25	9:45:29
9:27:45	9:31:33	9:36:45	9:42:08	9:45:37
9:27:53	9:32:50	9:37:05	9:42:22	9:45:45
9:28:46	9:32:58	9:37:18	9:42:29	9:45:52
9:29:03	9:34:21	9:37:35	9:42:58	9:46:35
9:30:05	9:35:22	9:38:33	9:43:20	9:48:06
9:30:20	9:35:55	9:38:45	9:43:35	9:48:45
9:30:36	✓	9:39:60	9:44:44	9:48:57
9:49:13	9:53:49	9:58:38	10:02:37	10:06:28
9:50:45	9:54:44	9:59:31	10:03:40	10:06:39
9:51:13	9:55:23	9:59:58	10:03:42	10:07:19
9:51:22	9:55:37	10:00:45	10:03:58	10:07:38
9:51:30	9:55:48	10:01:03	10:04:20	10:08:49
9:52:14	9:55:52	10:01:07	10:04:42	10:08:55
9:52:46	9:56:02	10:01:14	10:04:46	10:09:07
9:52:49	9:57:03	10:01:21	10:05:27	10:09:20
9:53:05	9:57:26	10:01:38	10:05:50	10:09:38
9:53:46	✓	10:01:42	10:06:07	10:09:43

10:01:55
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



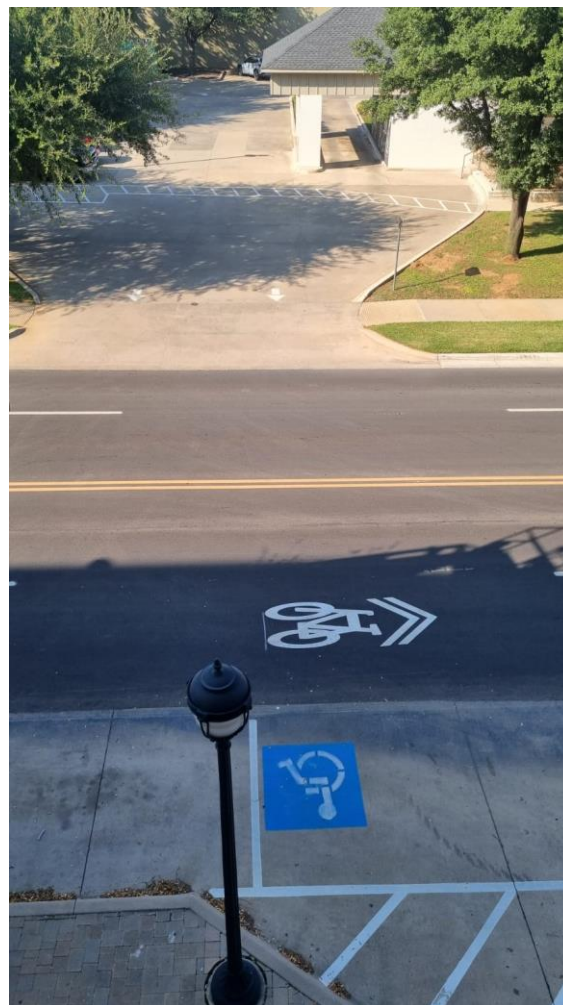
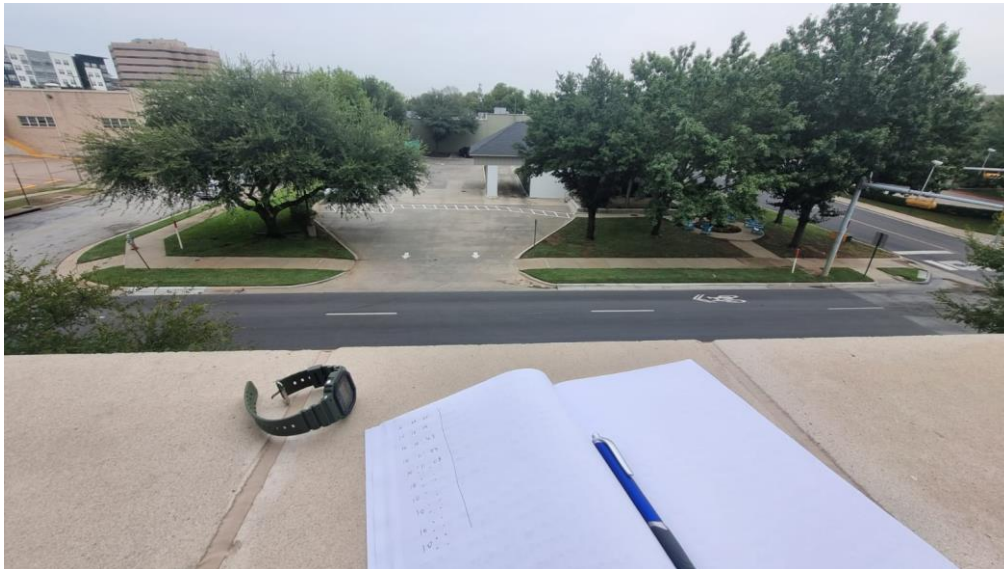
1	Clock time	Interval between occurrences	Interval (in seconds)
2	09:27:05	00:00:15	15
3	09:27:20	00:00:04	4
4	09:27:24	00:00:21	21
5	09:27:45	00:00:08	8
6	09:27:53	00:00:53	53
7	09:28:46	00:00:17	17
8	09:29:03	00:01:02	62
9	09:30:05	00:00:15	15
10	09:30:20	00:00:16	16
11	09:30:36	00:00:14	14
12	09:30:50	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 collected data.