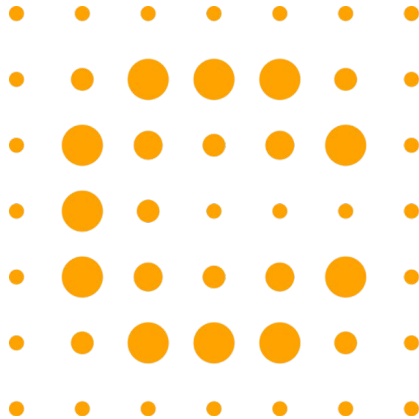


# DATA ANALYSIS USING PYTHON

*Submitted To:*



**cognino**

**By:** Rishabh Samra  
*MS Scholar, IIT-Madras*

# DATA

## Interested Users

phone	email_id	school	grade	source	Submitted At
911100221000	s2m_test_20001@tests2m.com	St.paul's school	Grade 2	lastchance	2020-04-16 08:46:39
911100221001	s2m_test_20002@tests2m.com	K L E School haveri	Grade 4	lastchance	2020-04-16 08:47:55
911100221002	s2m_test_20003@tests2m.com	Delhu Public School Howrah	Grade 2	lastchance	2020-04-16 08:47:59
911100221003	s2m_test_20004@tests2m.com	Bharathi vidya bhavans	Grade 2	lastchance	2020-04-16 08:48:27
911100221004	s2m_test_20005@tests2m.com	Army public school Delhi cantt	Grade 4	lastchance	2020-04-16 08:52:06

## Booked Users

Batch	Enrolment Type	Grade	School	Phone	Parent Email	Booked Time	Attendance
fREADom Trial - Grades 1, 2 [I] 11:15am - Thur...	Trial	Grade 2	DPS	911100222445	s2m_test_21459@tests2m.com	16-04-2020 11:15:00	Present
fREADom Trial - Grades 1, 2 [I] 11:15am - Thur...	Trial	Grade 2	DPS	911100222446	s2m_test_21460@tests2m.com	16-04-2020 11:15:00	Present
fREADom Trial - Grades 1, 2 [I] 11:15am - Thur...	Trial	Grade 2	City Montessori school lucknow	911100222447	s2m_test_21461@tests2m.com	16-04-2020 11:15:00	absent
fREADom Trial - Grades 3, 4 [I] 11:15am - Thur...	Trial	Grade 4	Reliance Foundation	911100222448	s2m_test_21462@tests2m.com	16-04-2020 11:15:00	absent
fREADom Trial - Grades 3, 4 [I] 11:15am - Thur...	Trial	Grade 3	KRS	911100222449	s2m_test_21463@tests2m.com	16-04-2020 11:15:00	Present

# DATA

## Trainers

Trainer Name	Date	Timing	Grades	Batch	Topics
FAC_5001	16 April Thursday	11.15 am - 12 noon	Grades 1, 2	trial I	Rocking with rhymes
FAC_5002	16 April Thursday	11.15 am - 12 noon	Grades 3, 4	trial I	In my mind's eye
FAC_5003	16 April Thursday	11.15 am - 12 noon	Grades 1, 2	trial II	Word Whiz
FAC_5006	16 April Thursday	11.15 am - 12 noon	Grades 3, 4	trial II	Rocking with rhymes
FAC_5007	16 April Thursday	4:00-4:45 pm	Grades 3, 4	trial I	In my head or real

## Subscription Bought

Registered Number	Registered Email ID	Payment date	Sale by	Grade	Pricing	Sessions preferred	Live session details
9.111002e+11	s2m_test_22161@tests2m.com	2020-04-27 00:00:00	SP_1005	2	1999	Assigned	NaN
9.111002e+11	s2m_test_22162@tests2m.com	2020-04-27 00:00:00	SP_1012	3	1999	Assigned	NaN
9.111002e+11	s2m_test_21876@tests2m.com	2020-04-27 00:00:00	SP_1005	3	1999	Assigned	NaN
9.111002e+11	s2m_test_22158@tests2m.com	2020-04-27 00:00:00	SP_1005	2	1999	NaN	NaN
9.111002e+11	s2m_test_21404@tests2m.com	2020-04-27 00:00:00	SP_1012	4	1999	Morning	NaN

# 1. Sales Conversion with respect to Trainers

- Firstly, we combined *Booked Users* Data with Trainers Data based on 4 attributes: Grades, Trial, Date and Time(information present in batch column of *Booked Users* data) to assign Trainer to each user.
- Then we combined this data on *Subscription Bought* Data while doing Inner Join based on *Phone* key.

	0	1	2	3	4
phone	9.111e+11	9.111e+11	9.111e+11	9.111e+11	9.111e+11
Registered Email ID	s2m_test_21964@tests2m.com	s2m_test_21807@tests2m.com	s2m_test_21807@tests2m.com	s2m_test_21486@tests2m.com	s2m_test_21753@tests2m.com
Payment date	26-03-2020 12:00:00 AM	2020-04-04 13:26:05	2020-04-04 13:26:05	2020-06-04 23:22:12	2020-08-04 16:44:32
Sale by	SP_1001	SP_1005	SP_1005	SP_1005	SP_1007
Grade_x	4	2	2	3	4
Pricing	1999	1999	1999	999	1999
Sessions preferred	NaN	2020-04-07 10:00:00	2020-04-07 10:00:00	2020-04-07 16:00:00	SCS/13042020/Batch 52
Live session details	Live - Upsell (Larger plan)	Live - Upsell (Larger plan)	Live - Upsell (Larger plan)	Live - Upsell (Larger plan)	Live - Upsell (Larger plan)
Batch_x	fREADom Trial - Grades 3, 4 [II] 06:15pm - Thu...	fREADom Trial - Grades 1, 2 [I] 11:45am - Frid...	fREADom Trial - Grades 1, 2 [II] 06:00pm - Fri...	fREADom Trial - Grades 3, 4 [II] 11:15am - Thu...	fREADom Trial - Grades 3, 4 [I] 11:45am - Tues...
Enrolment Type	Trial	Trial	Duplicate - Trial	Duplicate - Trial	Trial
Grade_y	Grade 4	Grade 2	Grade 2	Grade 3	Grade 4
School	Don bosco	Bhavaris	Bhavans	DPSH	Rajmata krishna kumari
Parent Email	s2m_test_21571@tests2m.com	s2m_test_21807@tests2m.com	s2m_test_21807@tests2m.com	s2m_test_21486@tests2m.com	s2m_test_21753@tests2m.com
Booked Time	16-04-2020 18:15:00	24-04-2020 11:45:00	24-04-2020 18:00:00	16-04-2020 11:15:00	21-04-2020 11:45:00
Attendance	Present	Present	Cancelled	Present	Present
Batch_Key	Grades 3, 4_II_3_16	Grades 1, 2_I_4_24	Grades 1, 2_II_5_24	Grades 3, 4_II_1_16	Grades 3, 4_I_4_21
Trainer Name	FAC_5015	FAC_5043	FAC_5006	FAC_5006	FAC_5015
Date	16 April Thursday	24 April Friday	24 April Friday	16 April Thursday	21 April Tuesday
Timing	6:15 - 7:00 pm	11.45 am -12.30 pm	6.00 pm - 6.45 pm	11.15 am - 12 noon	11.45 am -12.30 pm
Grades	Grades 3, 4	Grades 1, 2	Grades 1, 2	Grades 3, 4	Grades 3, 4
Batch_y	trial II	Trial I	Trial II	trial II	Trial I
Topics	In my head or real	Rocking with Rhymes	Word Whiz	Rocking with rhymes	Storm in my brain

## Note:

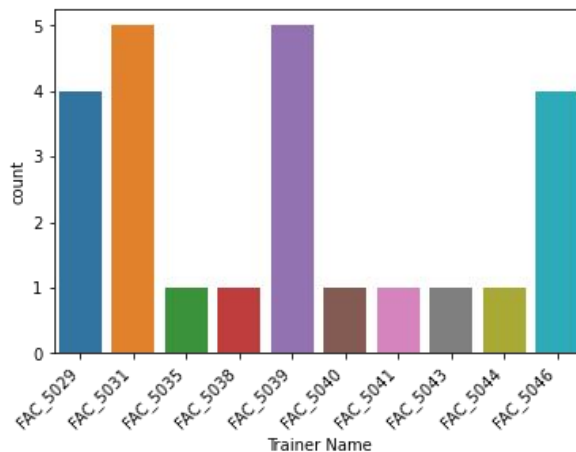
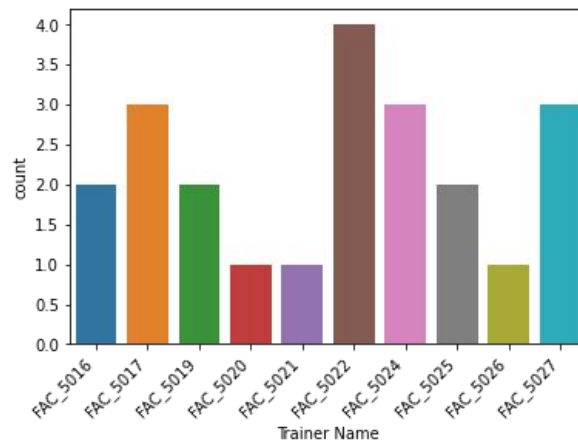
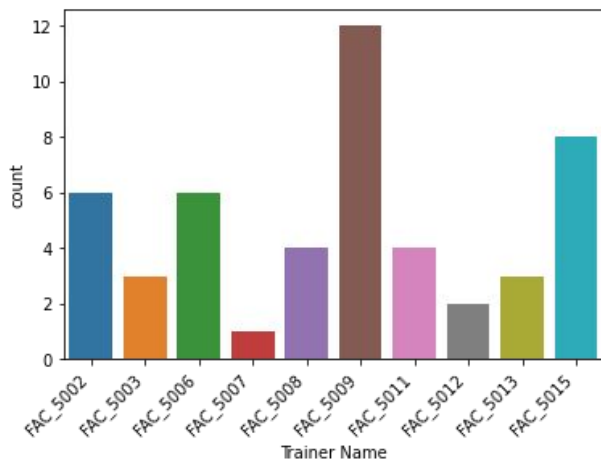
Here the data is shown which is transposed.

## Note:

Only data where Pricing>=499 is used.

	Trainer Name	count
0	FAC_5002	6
1	FAC_5003	3
2	FAC_5006	6
3	FAC_5007	1
4	FAC_5008	4
5	FAC_5009	12
6	FAC_5011	4
7	FAC_5012	2
8	FAC_5013	3
9	FAC_5015	8
10	FAC_5016	2
11	FAC_5017	3
12	FAC_5019	2
13	FAC_5020	1
14	FAC_5021	1
15	FAC_5022	4
16	FAC_5024	3
17	FAC_5025	2
18	FAC_5026	1
19	FAC_5027	3
20	FAC_5029	4
21	FAC_5031	5
22	FAC_5035	1
23	FAC_5038	1
24	FAC_5039	5
25	FAC_5040	1
26	FAC_5041	1
27	FAC_5043	1
28	FAC_5044	1
29	FAC_5046	4
30	FAC_5052	2

## Count of each trainer who has contributed to Sales Conversion



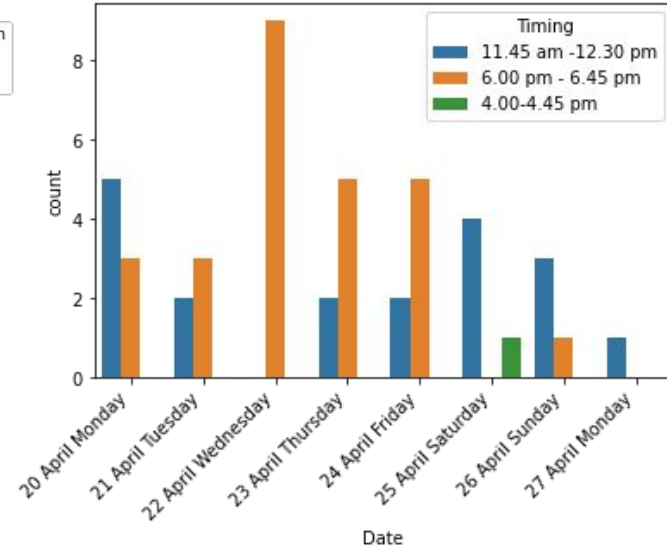
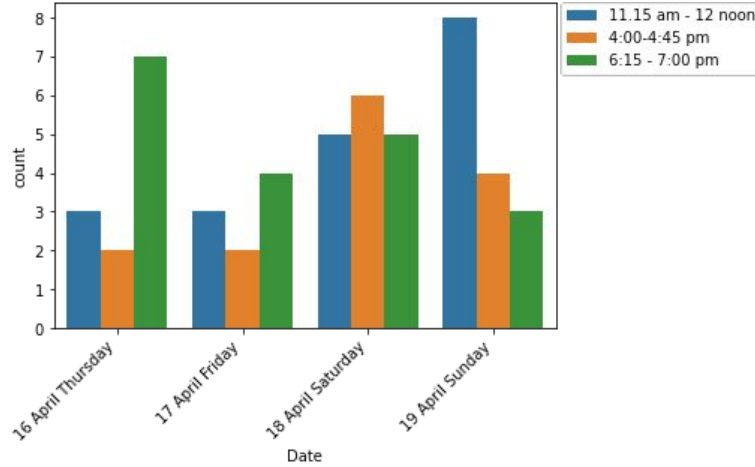
### Inference:

Trainer FAC\_5009 has trained maximum 12 users.

## 2. Sales Conversion with respect to Time Slots

*Count of Users Registered*

	Date	Timing	count
0	16 April Thursday	11.15 am - 12 noon	3
1	16 April Thursday	4:00-4:45 pm	2
2	16 April Thursday	6:15 - 7:00 pm	7
3	17 April Friday	11.15 am - 12 noon	3
4	17 April Friday	4:00-4:45 pm	2
5	17 April Friday	6:15 - 7:00 pm	4
6	18 April Saturday	11.15 am - 12 noon	5
7	18 April Saturday	4:00-4:45 pm	6
8	18 April Saturday	6:15 - 7:00 pm	5
9	19 April Sunday	11.15 am - 12 noon	8
10	19 April Sunday	4:00-4:45 pm	4
11	19 April Sunday	6:15 - 7:00 pm	3
12	20 April Monday	11.45 am -12.30 pm	5
13	20 April Monday	6.00 pm - 6.45 pm	3
14	21 April Tuesday	11.45 am -12.30 pm	2
15	21 April Tuesday	6.00 pm - 6.45 pm	3
16	22 April Wednesday	6.00 pm - 6.45 pm	9
17	23 April Thursday	11.45 am -12.30 pm	2
18	23 April Thursday	6.00 pm - 6.45 pm	5
19	24 April Friday	11.45 am -12.30 pm	2
20	24 April Friday	6.00 pm - 6.45 pm	5
21	25 April Saturday	11.45 am -12.30 pm	4
22	25 April Saturday	4.00-4.45 pm	1
23	26 April Sunday	11.45 am -12.30 pm	3
24	26 April Sunday	6.00 pm - 6.45 pm	1
25	27 April Monday	11.45 am -12.30 pm	1



### Inference:

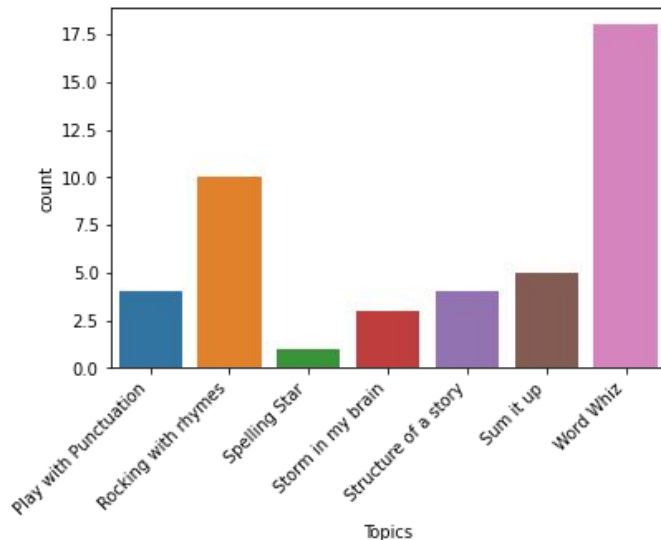
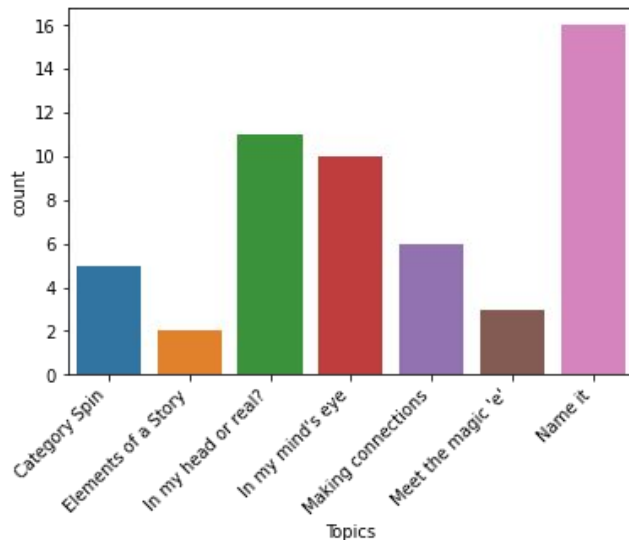
On 22 April,2020 at 6-6:45 pm slot, maximum users registered which are 9.



### 3. Sales Conversion with respect to Topics

Count of registered users w.r.t each topic

	Topics	count
0	Category Spin	5
1	Elements of a Story	2
2	In my head or real?	11
3	In my mind's eye	10
4	Making connections	6
5	Meet the magic 'e'	3
6	Name it	16
7	Play with Punctuation	4
8	Rocking with rhymes	10
9	Spelling Star	1
10	Storm in my brain	3
11	Structure of a story	4
12	Sum it up	5
13	Word Whiz	18



#### Inference:

Maximum 18 registered users are having interests to study the topic *Word Whiz*.

## 4. Sales conversion with respect to Source

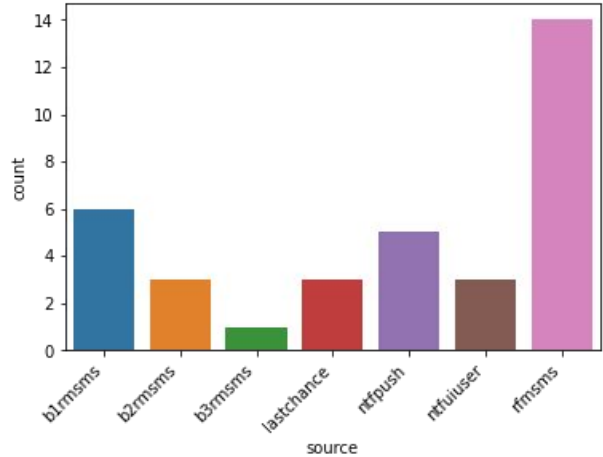
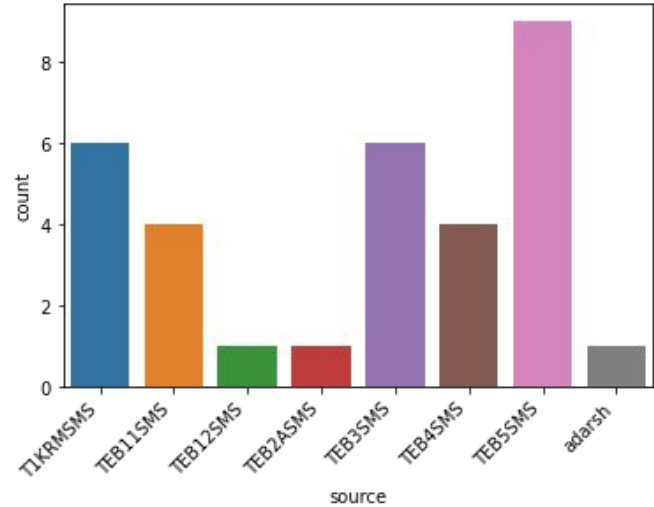
Combining **Interested Users** data with **Subscription Bought Data** while doing Inner Join based on Phone key

	0	1	2	3
phone	9.111e+11	9.111e+11	9.111e+11	9.111e+11
Registered Email ID	s2m_test_20297@tests2m.com	s2m_test_20016@tests2m.com	s2m_test_20164@tests2m.com	s2m_test_20265@tests2m.com
Payment date	27-03-2020 12:00:00 AM	2020-12-04 01:22:58	13-04-2020 9:52:27 AM	18-04-2020 11:12:15 AM
Sale by	SP_1005	SP_1005	SP_1005	SP_1012
Grade	4	2	3	3
Pricing	999	1999	1999	1999
Sessions preferred	NaN	SCJ/20042020/Batch53	SCS/13042020/ Batch 51	SCS/20042020/Batch 3
Live session details	Live - Upsell (Larger plan)	Live - Upsell (Larger plan)	Live - Upsell (Larger plan)	Live - Upsell (Larger plan)
email_id	s2m_test_20297@tests2m.com	s2m_test_20016@tests2m.com	s2m_test_20164@tests2m.com	s2m_test_20265@tests2m.com
school	Samashti	Indus Valley World School	Delhi public school south bangalore	SSRVM Bangalore south
grade	Grade 4	Grade 2	Grade 3	Grade 3
source	lastchance	lastchance	b1rmsms	ntfpush
Submitted At	2020-04-17 06:56:59	2020-04-16 09:06:17	2020-04-22 05:32:58	2020-04-17 05:45:31



Count of registered users w.r.t each topic

	source	count
0	T1KRMSMS	6
1	TEB11SMS	4
2	TEB12SMS	1
3	TEB2ASMS	1
4	TEB3SMS	6
5	TEB4SMS	4
6	TEB5SMS	9
7	adarsh	1
8	b1rmsms	6
9	b2rmsms	3
10	b3rmsms	1
11	lastchance	3
12	ntfpush	5
13	ntfuiuser	3
14	rfmsms	14



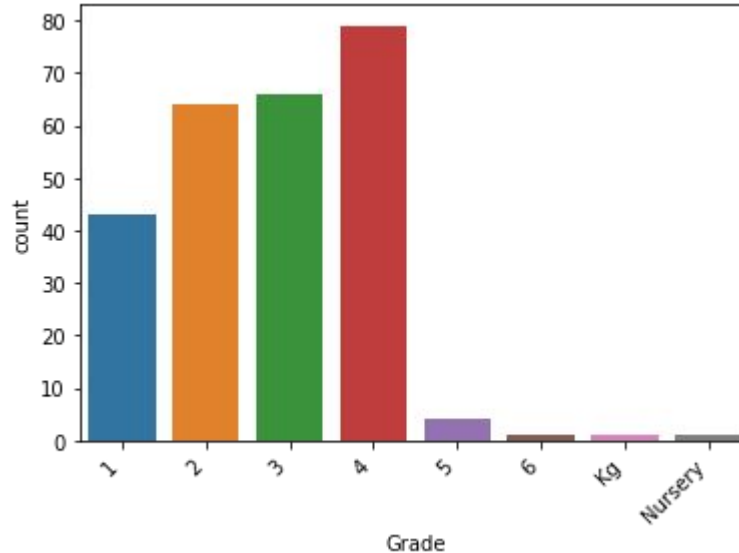
**Inference:**

Maximum 14 registered users came from source rfmsms.

## 5. Sales Conversion with respect to Grades

We will be using Subscription Bought Data for users having Pricing $\geq$ 499 for getting count of users for each grade.

	Grade	count
0	1	43
1	2	64
2	3	66
3	4	79
4	5	4
5	6	1
6	Kg	1
7	Nursery	1



### **Inference:**

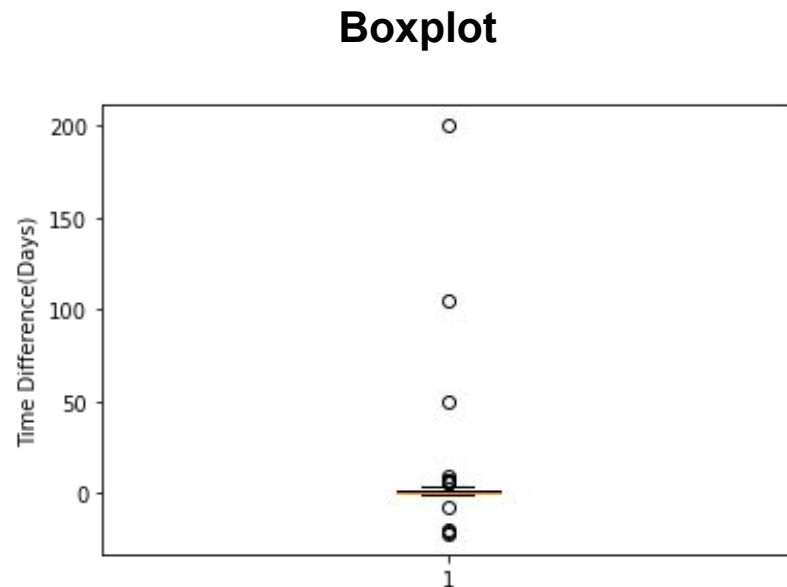
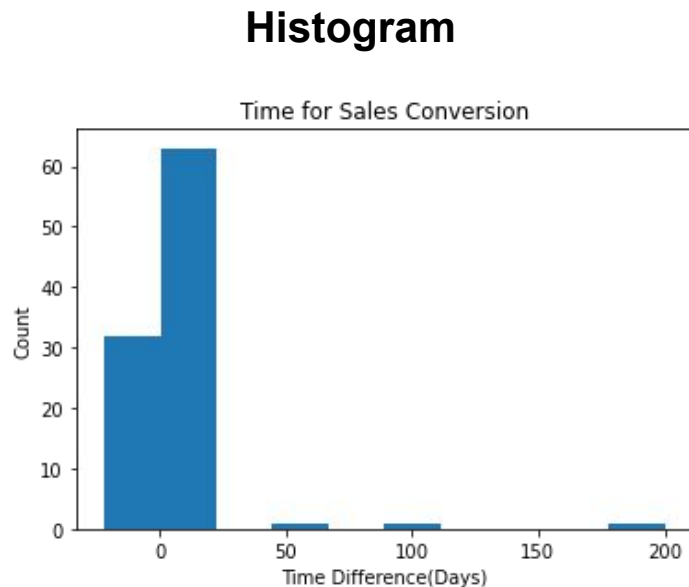
There are highest registered users of Grade 4 with 79 Users.

## 6. Time Taken for Sales Conversion

Using combined Booked Users Data with Subscription Brought Data to get the difference between timing of Payment Date and Booked Timing.

	0	1	2	3	4
phone	9.111e+11	9.111e+11	9.111e+11	9.111e+11	9.111e+11
Registered Email ID	s2m_test_21964@tests2m.com	s2m_test_21807@tests2m.com	s2m_test_21807@tests2m.com	s2m_test_21486@tests2m.com	s2m_test_21753@tests2m.com
Payment date	26-03-2020 12:00:00 AM	2020-04-04 13:26:05	2020-04-04 13:26:05	2020-06-04 23:22:12	2020-08-04 16:44:32
Sale by	SP_1001	SP_1005	SP_1005	SP_1005	SP_1007
Grade_x	4	2	2	3	4
Pricing	1999	1999	1999	999	1999
Sessions preferred	NaN	2020-04-07 10:00:00	2020-04-07 10:00:00	2020-04-07 16:00:00	SCS/13042020/Batch 52
Live session details	Live - Upsell (Larger plan)	Live - Upsell (Larger plan)	Live - Upsell (Larger plan)	Live - Upsell (Larger plan)	Live - Upsell (Larger plan)
Batch	fREADom Trial - Grades 3, 4 [II] 06:15pm - Thu...	fREADom Trial - Grades 1, 2 [I] 11:45am - Fri...	fREADom Trial - Grades 1, 2 [II] 06:00pm - Fri...	fREADom Trial - Grades 3, 4 [II] 11:15am - Thu...	fREADom Trial - Grades 3, 4 [I] 11:45am - Tues...
Enrolment Type	Trial	Trial	Duplicate - Trial	Duplicate - Trial	Trial
Grade_y	Grade 4	Grade 2	Grade 2	Grade 3	Grade 4
School	Don bosco	Bhavans	Bhavans	DPSH	Rajmata krishna kumari
Parent Email	s2m_test_21571@tests2m.com	s2m_test_21807@tests2m.com	s2m_test_21807@tests2m.com	s2m_test_21486@tests2m.com	s2m_test_21753@tests2m.com
Booked Time	16-04-2020 18:15:00	24-04-2020 11:45:00	24-04-2020 18:00:00	16-04-2020 11:15:00	21-04-2020 11:45:00
Attendance	Present	Present	Cancelled	Present	Present
Batch_Key	Grades 3, 4_II_3_16	Grades 1, 2_I_4_24	Grades 1, 2_II_5_24	Grades 3, 4_II_1_16	Grades 3, 4_I_4_21

## Data With Outliers



The smallest 5 values are : [-21.792, -20.208, -19.958, -7.833, -1.042]

The largest 5 values are : [7.417, 9.333, 49.5, 105.167, 200.0]

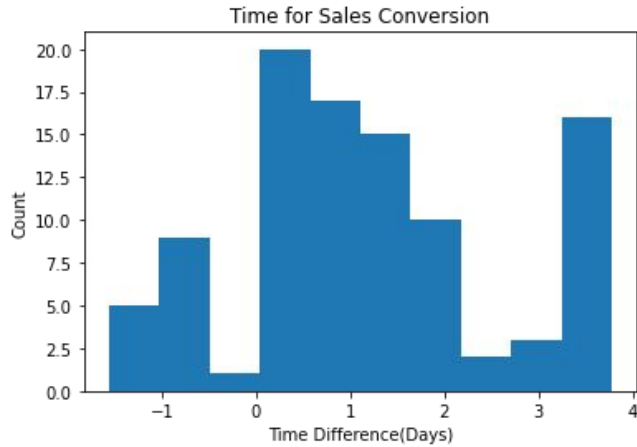
Mean: 4.321    1<sup>st</sup> Quartile, Q1: 0.2185    Median, Q2 : 1.083    3<sup>rd</sup> Quartile, Q3 : 2.0

IQR = Q3 - Q1

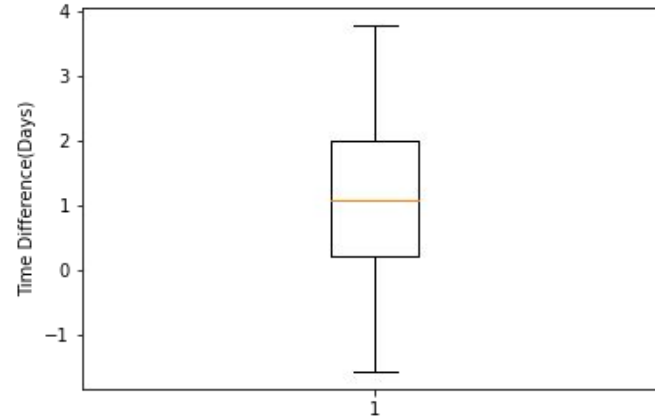
Range of Outlier : [ Q1 - 1.5\*IQR , Q3 + 1.5\*IQR ] = [ -1.563 , 3.782 ]

## Data Without Outliers

### Histogram



### Boxplot



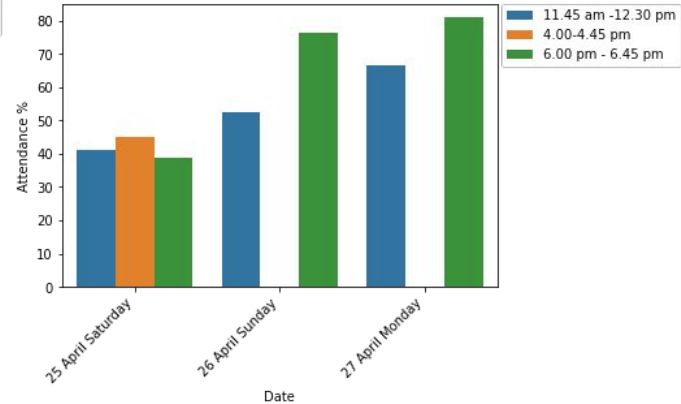
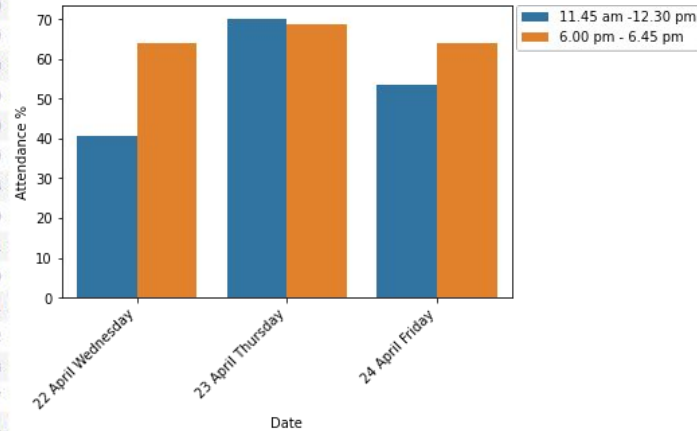
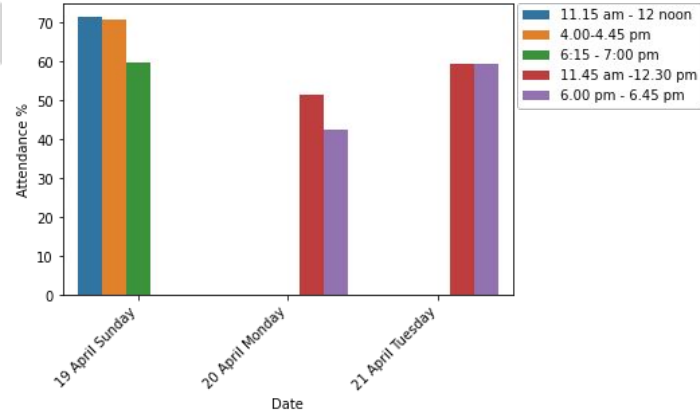
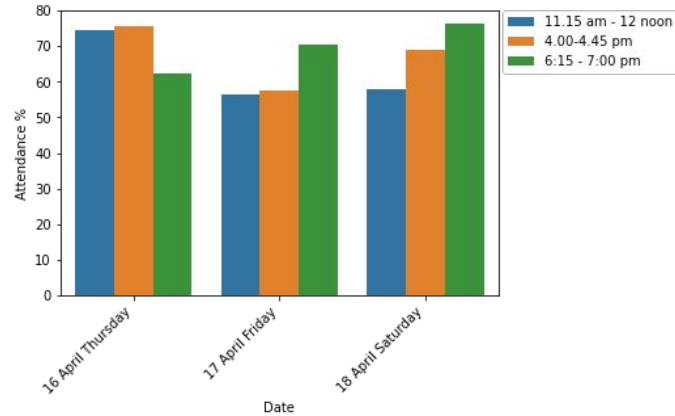
### Inference:

On an average, it takes one day for a user to make payment after taking the trial.

## 7. Time Slot- Attendance Percent of Users

Booked Users and Trainers Data when merged has information of both time slots and attendance.

	Date	Timing	count	Attendance %
0	16 April Thursday	11.15 am - 12 noon	35	74.285714
1	16 April Thursday	4.00-4.45 pm	41	75.609756
2	16 April Thursday	6:15 - 7:00 pm	61	62.295082
3	17 April Friday	11.15 am - 12 noon	23	56.521739
4	17 April Friday	4.00-4.45 pm	40	57.500000
5	17 April Friday	6:15 - 7:00 pm	54	70.370370
6	18 April Saturday	11.15 am - 12 noon	50	58.000000
7	18 April Saturday	4.00-4.45 pm	64	68.750000
8	18 April Saturday	6:15 - 7:00 pm	59	76.271186
9	19 April Sunday	11.15 am - 12 noon	63	71.428571
10	19 April Sunday	4.00-4.45 pm	62	70.967742
11	19 April Sunday	6:15 - 7:00 pm	52	59.615385
12	20 April Monday	11.45 am -12.30 pm	74	51.351351
13	20 April Monday	6.00 pm - 6.45 pm	66	42.424242
14	21 April Tuesday	11.45 am -12.30 pm	64	59.375000
15	21 April Tuesday	6.00 pm - 6.45 pm	54	59.259259
16	22 April Wednesday	11.45 am -12.30 pm	64	40.625000
17	22 April Wednesday	6.00 pm - 6.45 pm	75	64.000000
18	23 April Thursday	11.45 am -12.30 pm	50	70.000000
19	23 April Thursday	6.00 pm - 6.45 pm	70	68.571429
20	24 April Friday	11.45 am -12.30 pm	45	53.333333
21	24 April Friday	6.00 pm - 6.45 pm	64	64.062500
22	25 April Saturday	11.45 am -12.30 pm	68	41.176471
23	25 April Saturday	4.00-4.45 pm	31	45.161290
24	25 April Saturday	6.00 pm - 6.45 pm	57	38.596491
25	26 April Sunday	11.45 am -12.30 pm	42	52.380952
26	26 April Sunday	6.00 pm - 6.45 pm	17	76.470588
27	27 April Monday	11.45 am -12.30 pm	36	66.666667
28	27 April Monday	6.00 pm - 6.45 pm	21	80.952381





**THANK YOU**