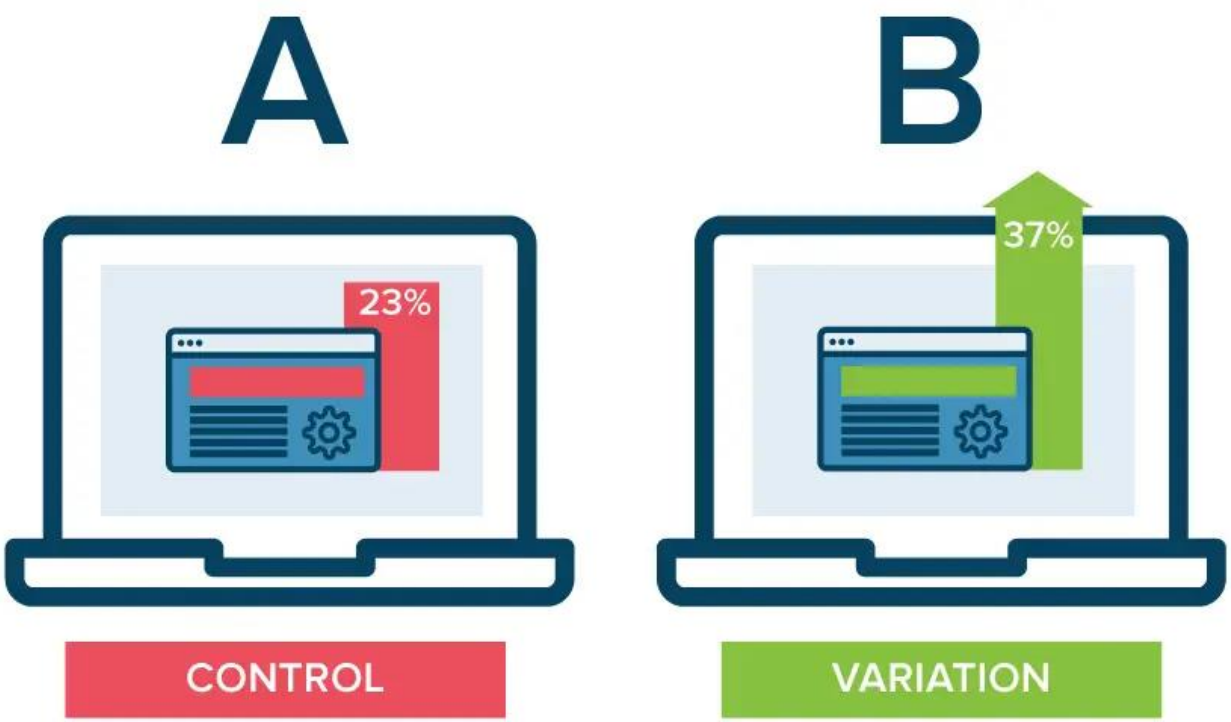

A
B
23%
37%
CONTROL
VARIATION

A/B Testing Using Python

Overview:

Problem Scenario:

- A company recently introduced a new bidding type, **“average bidding”**, as an alternative to its existing bidding type, called **“maximum bidding”**. One of their clients, has decided to test this new feature and wants to conduct an **A/B test** to understand if **"average bidding"** brings more **Purchase** than **"maximum bidding".**

- In this A/B test, the company randomly splits its audience into two equally sized groups;
  1. the **Test Group**
  2. the **Control Group**
- Company ad campaign with “maximum bidding” is served to “control group” and another campaign with “average bidding” is served to the “test group”

- The A/B test has run for 40 days and the company now expects to analyze and present the results of this A/B test.

Objectives:

1. How would you define the hypothesis of this A/B test?
2. Which statistical test did you use, and why?
3. Can we conclude statistically significant results?
4. Based on your answer to Question 3, what would be your recommendation to client?

Aim:

1. Presentation should last about 15 minutes, and should be presented in English.
2. The ultimate success metric for the company is `Number of Purchases`. Therefore, we should focus on `Purchase metrics` for statistical testing.
3. Explain the concept of statistical testing for a non-technical audience.
4. Use visualizations to compare test and control group metrics, such as Website Click Through Rate, Cost per Action, and Conversion Rates in addition to Purchase numbers.
5. If any trends, anomalies or other patterns, discuss these in presentation.
6. We can make assumptions if needed.

Dataset Overview:

- Dataset contains 4 different metric variables observed during one month:
  - User sees an ad (**Impression**)
  - User clicks on the website link on the ad (**Click**)
  - User purchases the product (**Purchase**)
  - Value from the purchase (**Earning**)

Conclusion Of the Analysis:

1. Hypothesis of this A/B test ?

Hypothesis:

- **H0**: **Average Bidding** ≤ **Maximum Bidding**, for the **purchase amounts** of the two groups.
- **H1**: **Average Bidding** > **Maximum Bidding**, for the **purchase amounts** of the two groups.

2. Which statistical test did used, and why?

- I used the independent sample t-test statistical method.
- Because the return of a new system is wanted to be tested.
- For this purpose, user behaviors are examined and statistically compared with each other with the Control Group, which uses the old system with equal number of observations, and the Test Group, which uses the new system.

3. Can we conclude statistically significant results?

- From above analysis we observed that there was no statistically significant mean difference between the number of **Purchases** made with the Maximum Bidding and Average Bidding methods.
- Since the p-value for the **Parametric Independent One Sample t-Test** was performed using the Test and Control group data set was greater than 0.05.

4. Based on answer to Question 3, what would be the recommendation to client?

- This test was conducted on datasets containing 40 observations.
- In accordance with the Law of Large Numbers, as the number of observations increases, the tests performed will produce the most realistic results
- Therefore, since there is no statistically significant mean difference between the returns of the old and new systems, it would be better to examine the systems for a while by increasing the number of observations in the data set.
- In other words, repeating this hypothesis test with new data sets by increasing the number of purchases in both systems and continuing to collect data can give us a statistically significant mean difference and this situation should be tested.

Evaluations:

Explain the concept of statistical testing for a non-technical audience.

- When we develop a system technically by investing time and resources, we want to measure whether the new system is a better system than the old one, whether the work we do is worth the result we have,
- in short, whether the stone we throw is worth the frog we frightened. It is possible to talk about this situation with numbers, net values and statistics.
- For example, here we divide our data into 2 equal parts. We measured the performance of the old system with the first of these parts and the performance of the new system with the second, and these measurements produced some mathematical values for us.
- By comparing these values with some statistical assumptions (such as comparing the p-value with 0.05), we were able to determine whether there was a statistical difference between the performance of these systems and performed a statistical test.

Recommendation:

- When the new system is compared with the old system, it is seen that the number of purchases does not change, the cost increases and the clicks decrease.
- When the Conversion Rate is based on the purchasing status among those who add to the basket and those who look at the product details, a comparison is made in favor of the test group.
- **Therefore, retesting is recommended by collecting more data. If there is no such opportunity or time, it is recommended to continue with the existing bidding system.**

Let's Analyze the data

Let's explore DataGroupA

	Impression	Click	Purchase	Earning
0	82529.459271	6090.077317	665.211255	2311.277143
1	98050.451926	3382.861786	315.084895	1742.806855
2	82696.023549	4167.965750	458.083738	1797.827447
3	109914.400398	4910.882240	487.090773	1696.229178
4	108457.762630	5987.655811	441.034050	1543.720179

DataGroupA Info

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 40 entries, 0 to 39
Data columns (total 4 columns):
 #   Column      Non-Null Count  Dtype  
---  ------      --------------  -----  
 0   Impression  40 non-null     float64
 1   Click       40 non-null     float64
 2   Purchase    40 non-null     float64
 3   Earning     40 non-null     float64
dtypes: float64(4)
memory usage: 1.4 KB
```

Summary Statistics of DataGroupA

	Impression	Click	Purchase	Earning
count	40.0	40.0	40.00	40.00
mean	101711.44	5100.7	550.89	1908.56
std	20302.15	1329.9	134.10	302.91
min	45475.0	2189.8	267.02	1253.98
25%	85726.9	4124.3	470.09	1685.84
50%	99790.8	5001.2	531.20	1975.16
75%	115212.8	5923.8	637.95	2119.80
max	147539.0	7959.0	801.79	2497.29

Let's explore DataGroupB

	Impression	Click	Purchase	Earning
0	120103.50	3216.54	702.16	1939.61
1	134775.94	3635.08	834.05	2929.40
2	107806.62	3057.14	422.93	2526.24
3	116445.27	4650.47	429.03	2281.42
4	145082.51	5201.38	749.86	2781.69

Dataset Info:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 40 entries, 0 to 39
Data columns (total 4 columns):
```

#	Column	Non-Null Count	Dtype
0	Impression	40 non-null	float64
1	Click	40 non-null	float64
2	Purchase	40 non-null	float64
3	Earning	40 non-null	float64

```
dtypes: float64(4)
memory usage: 1.4 KB
```

Summary Statistics:

	Impression	Click	Purchase	Earning
count	40.000000	40.000000	40.000000	40.000000
mean	120512.411758	3967.549761	582.106097	2514.890733
std	18807.448712	923.095073	161.152513	282.730852
min	79033.834921	1836.629861	311.629515	1939.611243
25%	112691.970770	3376.819024	444.626828	2280.537426
50%	119291.300775	3931.359804	551.355732	2544.666107
75%	132050.578933	4660.497911	699.862360	2761.545405
max	158605.920483	6019.695079	889.910460	3171.489708

Feature Engineering

Feature Engineering: Conversion Rate

For GroupA

	Impression	Click	Purchase	Earning	Conversion Rate
0	82529.459271	6090.077317	665.211255	2311.277143	10.922870
1	98050.451926	3382.861786	315.084895	1742.806855	9.314152
2	82696.023549	4167.965750	458.083738	1797.827447	10.990583
3	109914.400398	4910.882240	487.090773	1696.229178	9.918600
4	108457.762630	5987.655811	441.034050	1543.720179	7.365721

For GroupB

	Impression	Click	Purchase	Earning	Conversion Rate
0	120103.503796	3216.547958	702.160346	1939.611243	21.829625
1	134775.943363	3635.082422	834.054286	2929.405820	22.944577
2	107806.620788	3057.143560	422.934258	2526.244877	13.834295
3	116445.275526	4650.473911	429.033535	2281.428574	9.225587
4	145082.516838	5201.387724	749.860442	2781.697521	14.416546

Feature Engineering: Earning per Purchase

For GroupA

	Impression	Click	Purchase	Earning	Conversion Rate	\
0	82529.459271	6090.077317	665.211255	2311.277143	10.922870	
1	98050.451926	3382.861786	315.084895	1742.806855	9.314152	
2	82696.023549	4167.965750	458.083738	1797.827447	10.990583	
3	109914.400398	4910.882240	487.090773	1696.229178	9.918600	
4	108457.762630	5987.655811	441.034050	1543.720179	7.365721	

	Earning per Purchase
0	347.450096
1	553.122947
2	392.466987
3	348.236771
4	350.022902

For GroupB

	Impression	Click	Purchase	Earning	Conversion Rate	\
0	120103.503796	3216.547958	702.160346	1939.611243	21.829625	
1	134775.943363	3635.082422	834.054286	2929.405820	22.944577	
2	107806.620788	3057.143560	422.934258	2526.244877	13.834295	
3	116445.275526	4650.473911	429.033535	2281.428574	9.225587	
4	145082.516838	5201.387724	749.860442	2781.697521	14.416546	

	Earning per Purchase
0	276.234802
1	351.224839
2	597.313845
3	531.759965
4	370.962030

Data Manipulation

Data Manipulation: Indication of Groups in A New Variable

For GroupA

	Impression	Click	Purchase	Earning	Conversion Rate	\
0	82529.459271	6090.077317	665.211255	2311.277143	10.922870	
1	98050.451926	3382.861786	315.084895	1742.806855	9.314152	
2	82696.023549	4167.965750	458.083738	1797.827447	10.990583	
3	109914.400398	4910.882240	487.090773	1696.229178	9.918600	
4	108457.762630	5987.655811	441.034050	1543.720179	7.365721	

	Earning per Purchase	Group
0	347.450096	GroupA
1	553.122947	GroupA
2	392.466987	GroupA
3	348.236771	GroupA
4	350.022902	GroupA

For GroupB

```
      Impression        Click    Purchase      Earning  Conversion Rate  \
0  120103.503796  3216.547958  702.160346  1939.611243        21.829625
1  134775.943363  3635.082422  834.054286  2929.405820        22.944577
2  107806.620788  3057.143560  422.934258  2526.244877        13.834295
3  116445.275526  4650.473911  429.033535  2281.428574         9.225587
4  145082.516838  5201.387724  749.860442  2781.697521        14.416546

   Earning per Purchase   Group
0            276.234802  GroupB
1            351.224839  GroupB
2            597.313845  GroupB
3            531.759965  GroupB
4            370.962030  GroupB
```

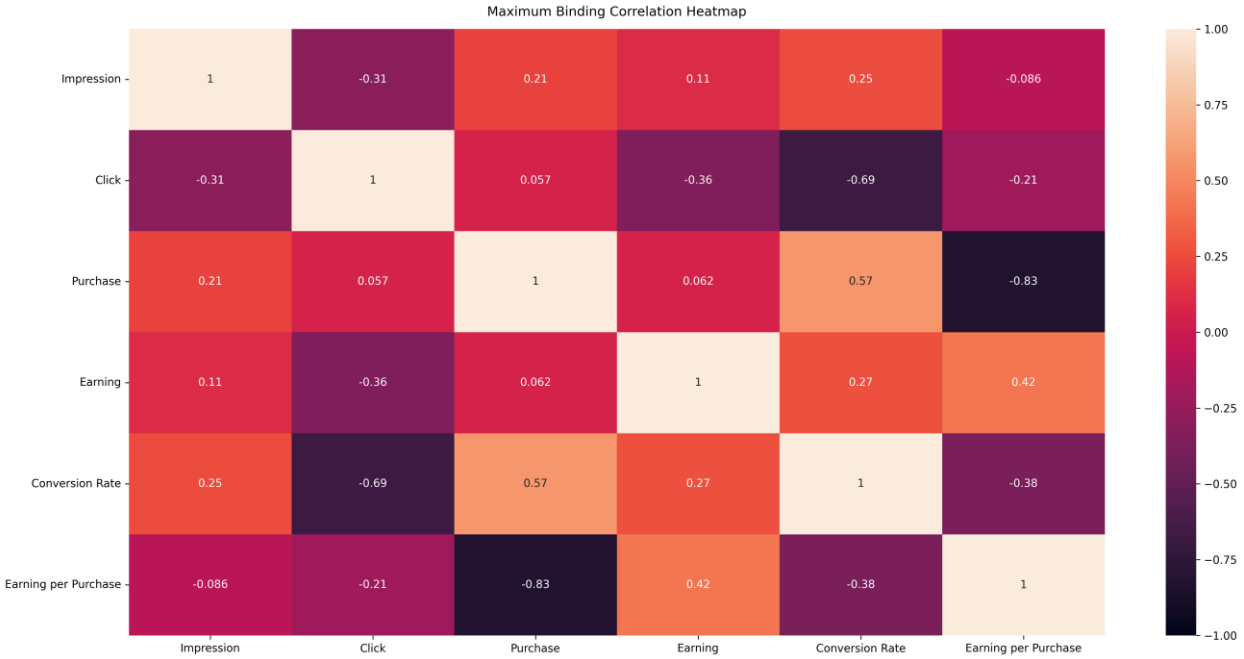
Let's Combine the Dataset

	Impression	Click	Purchase	Earning	Conversion Rate	Earning per Purchase	Group
0	82529.459271	6090.077317	665.211255	2311.277143	10.922870	347.450096	GroupA
1	98050.451926	3382.861786	315.084895	1742.806855	9.314152	553.122947	GroupA
2	82696.023549	4167.965750	458.083738	1797.827447	10.990583	392.466987	GroupA
3	109914.400398	4910.882240	487.090773	1696.229178	9.918600	348.236771	GroupA
4	108457.762630	5987.655811	441.034050	1543.720179	7.365721	350.022902	GroupA
5	77773.633900	4462.206586	519.669656	2081.851850	11.646024	400.610624	GroupA
40	120103.503796	3216.547958	702.160346	1939.611243	21.829625	276.234802	GroupB
41	134775.943363	3635.082422	834.054286	2929.405820	22.944577	351.224839	GroupB
42	107806.620788	3057.143560	422.934258	2526.244877	13.834295	597.313845	GroupB
43	116445.275526	4650.473911	429.033535	2281.428574	9.225587	531.759965	GroupB
44	145082.516838	5201.387724	749.860442	2781.697521	14.416546	370.962030	GroupB
45	115923.006949	4213.868620	778.373161	2157.408552	18.471700	277.168929	GroupB
46	106116.436642	3279.472973	491.614531	2560.411202	14.990657	520.816827	GroupB
47	125957.116104	4690.569911	855.719803	2563.579756	18.243408	299.581679	GroupB

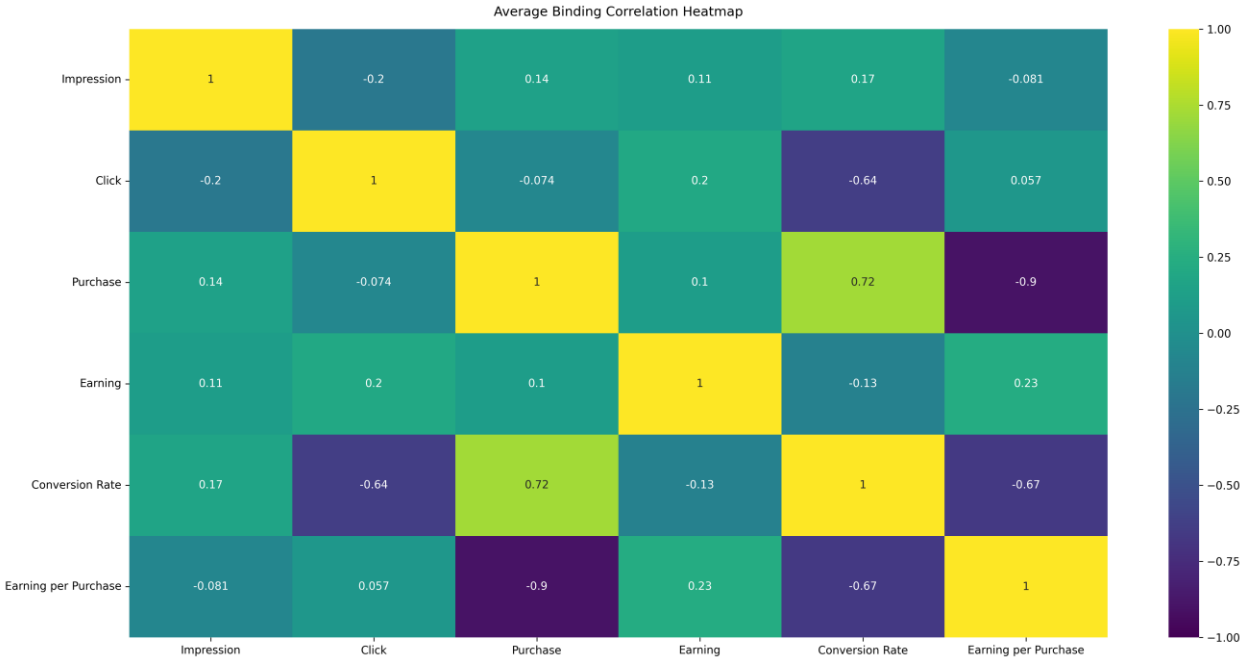
A/B Testing for Purchase Matrix:

Correlation between feature

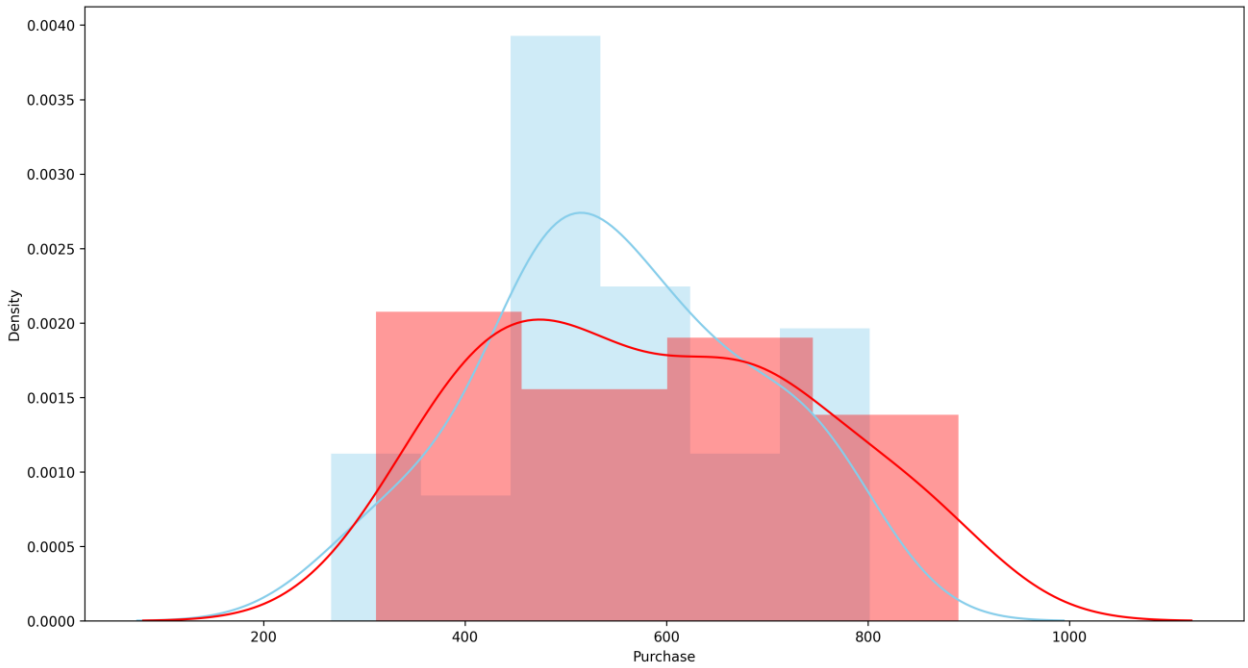
GroupA



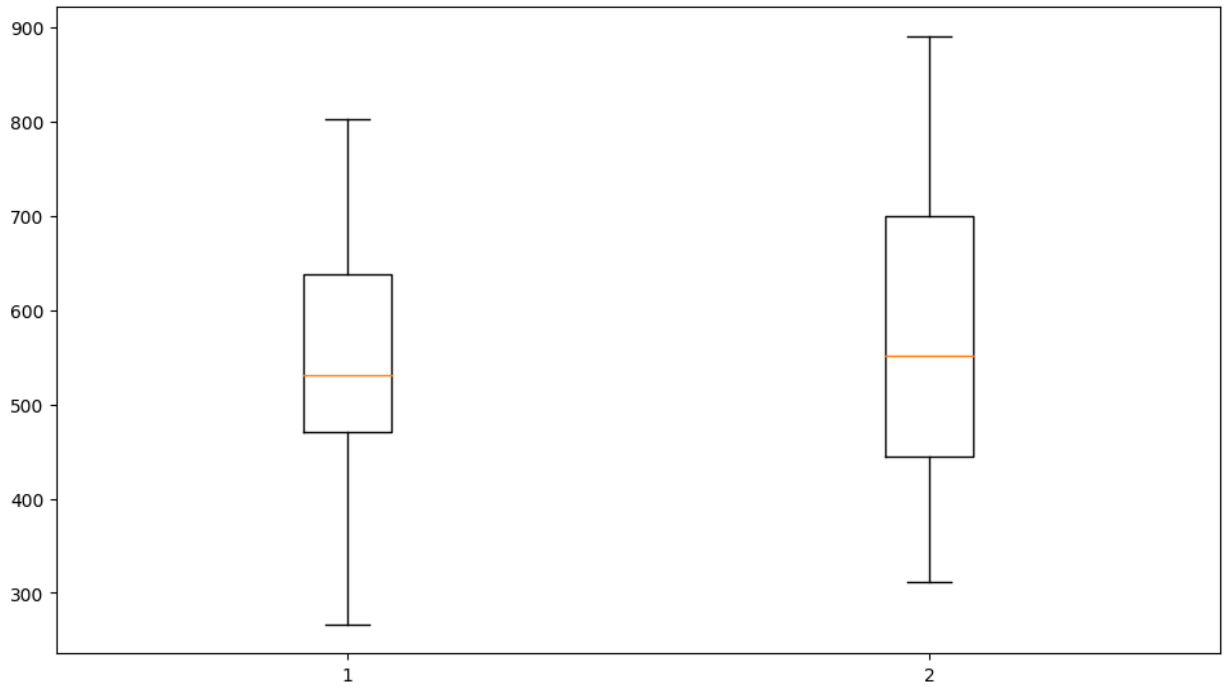
GroupB



Distribution of Purchase Matrix



Outlier Checking



Step-1: Testing the Normality Assumption

Hypothesis:

- **H0:** the distribution of the sample is not significantly different from a normal distribution
- **H1:** the distribution in question is significantly different from a normal distribution.

The confidence level will be set as 95% or 0.95. Hence, $\alpha = 1 - 0.95 = 0.05$

Results:

```
GroupA_shaphiro_p = 0.589, GroupB_shaphiro_p = 0.154
```

Interpretation:

- As both of the Group's p-values are bigger than o.o5, (p> .05)
- therefore, we cannot reject the null hypothesis,
- thus, we can say the distribution of the sample is not significantly different from a normal distribution.
- Means, normality assumption is fulfilled.

Step-2: Testing Homogeneity Levene's Test:

Hypothesis:

- **H0:** the variance among groups is equal.
- **H1:** are not equal.

Results:

```
levene_F = 2.639, levene_p = 0.108
```

Interpretation:

- The p-value for the Levene test is greater than 0.05,
- therefore, we cannot reject the null hypothesis.
- thus, we can say that the variances are not significantly different from each other
- Means the homogeneity assumption of the variance is met.

Info:

- **if levene_p > 0.05:**
    - Alternative 1--> Independent Samples t-Test
    - equal_var = True
- **if levene_p < 0.05:**
    - Alternative 1--> Independent Samples t-Test
    - equal_var = False

Decision:

- As our Homogeneity assumption is fullfilled and
- levene_p> 0.05
- we will select Independent Sample t-Test where,
- equal_var = True

Final Hypothesis:

- **H0**: Average Bidding ≤ Maximum Bidding, for the purchase amounts of the two groups.
- **H1**: Average Bidding > Maximum Bidding, for the purchase amounts of the two groups.

Results:

```
t_value = -0.942, t_test_p = 0.349
```

Interpretation:

- Independent Samples t-Test resulted as p > .05
- Which indicates that H0 cannot be rejected.
- Thus, we can say that, **there is no significant mean difference between average binding and maximum binding for Purchase values.**
- Means, Purchase doesn't differ much from one another.
- One can select either **bidding system** if they wish to make a decision based on just one measure.

Conclusion

1. Define the hypothesis of this A/B test?

Hypothesis:

- **H0**: Average Bidding ≤ Maximum Bidding, for the purchase amounts of the two groups.
- **H1**: Average Bidding > Maximum Bidding, for the purchase amounts of the two groups.

2. Which statistical test did used, and why?

- I used the independent sample t-test statistical method.
- Because the return of a new system is wanted to be tested.
- For this purpose, user behaviors are examined and statistically compared with each other with the Control Group, which uses the old system with equal number of observations, and the Test Group, which uses the new system.

3. Can we conclude statistically significant results?

- From above analysis we observed that there was no statistically significant mean difference between the number of **Purchases** made with the Maximum Bidding and Average Bidding methods.
- Since the p-value for the **Parametric Independent One Sample t-Test** was performed using the Test and Control group data set was greater than 0.05.

4. Based on answer to Question 3, what would be the recommendation to client?

- This test was conducted on datasets containing 40 observations.
- In accordance with the Law of Large Numbers, as the number of observations increases, the tests performed will produce the most realistic results
- Therefore, since there is no statistically significant mean difference between the returns of the old and new systems, it would be better to examine the systems for a while by increasing the number of observations in the data set.
- In other words, repeating this hypothesis test with new data sets by increasing the number of purchases in both systems and continuing to collect data can give us a statistically significant mean difference and this situation should be tested.

Evaluations:

Explain the concept of statistical testing for a non-technical audience.

- When we develop a system technically by investing time and resources, we want to measure whether the new system is a better system than the old one, whether the work we do is worth the result we have,
- in short, whether the stone we throw is worth the frog we frightened. It is possible to talk about this situation with numbers, net values and statistics.
- For example, here we divide our data into 2 equal parts. We measured the performance of the old system with the first of these parts and the performance of the new system with the second, and these measurements produced some mathematical values for us.
- By comparing these values with some statistical assumptions (such as comparing the p-value with 0.05), we were able to determine whether there was a statistical mean difference between the performance of these systems and performed a statistical test.

Recommendation:

- When the new system is compared with the old system, it is seen that the number of purchases does not change, the cost increases and the clicks decrease.
- When the Conversion Rate is based on the purchasing status among those who add to the basket and those who look at the product details, a comparison is made in favor of the test group.
- **Therefore, retesting is recommended by collecting more data. If there is no such opportunity or time, it is recommended to continue with the existing bidding system.**