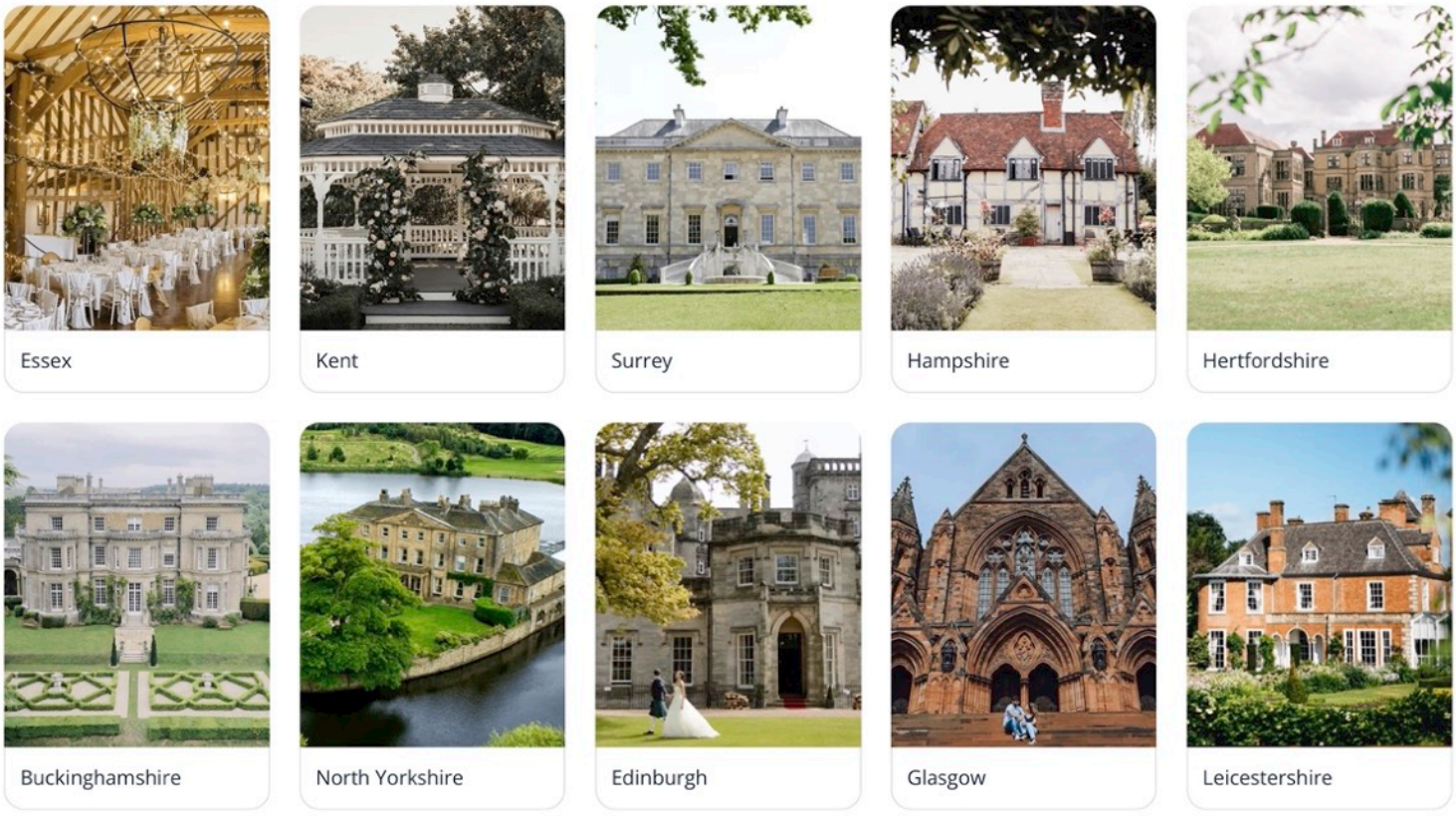
BRIDEBOOK
The No.1 Wedding Planning App

Author Name:
Ahmed Adel
Task Report
25 July 2023

BrideBook A/B testing and Analysis Report

Our top wedding venue locations



Bridebook is the world's no1 Wedding Planning app! We have all the tools you need to plan your wedding, all in one place.

Abstract :

In this report I will be cover the steps o the implementation of A/B testing and results with visualized graphs and make the business interpretation and recommendation based on the results coming from A/B testing and EDA analysis with business insights.
Aiming to make the most adequate for upcoming business model plan and the efficient decision for the current and the short term plan.

Table of Contents:

1-Introduction:

1.1 Background:

It may be unnecessary to provide a background of Bridebook since I am portraying this report to the Bridebook team, but from my point of view, you wouldn't mind reading again some of the history of Bridebook.

Staring by the Creation of BrideBook by Hamish Shephard to streamline and simplify the procedure of planning weddings for engaged couples. The website's free utilities include a customizable checklist, venue and supplier search, guestlist, and budget planner [1]. In addition, offering pieces of advice on organizing the perfect wedding, Shephard wrote a new columns for the events Standards [1] i

Bridebook's primary goal has always been to help couples enjoy wedding planning rather than dread it. In its first year of operation, it aimed to attract 100 banner ads and 400 dealer listings [3]. Bridebook.com is one of seven competitors in the wedding website industry, but it stands out because of its original and engaging content. This approach has allowed it to recruit respectable suppliers and advertising from various demographics [3].

When it first launched, Bridebook.com faced the formidable task of cracking the $3.8 billion Metro Vancouver wedding industry. Bridebook.com, on the other hand, planned to grab 40% of the market share in its region with its unique offers and complete services before extending to neighboring regions [3].

In the future, Bridebook.com hopes to increase its user base and the number of businesses with whom it collaborates. BrideBook also strives to keep its services ahead of the competition by constantly developing and enhancing them. Plans for the firm include turning a profit of $1.2 million after expenses in the first year of business [3]. Bridebook plans to become a lucrative and successful wedding planning company [3] with an estimated $86,040 in initial start-up capital.

1.2 Motivation:

I am writing to convey my strong desire to join Bridebook's innovative team in the role of Data Scientist. The opportunity to contribute my skills to BrideBpook whose mission is to simplify the wedding planning process and encouraging along with spreading love is very appealing to me.

By joining Bridebook, I'd be able to have a positive impact on the lives of others via my work. I'm very like the concept of employing cutting-edge tools to help others relax and revel in life's unforgettable moments. I find the breadth and depth of this industry's challenges fascinating, and I can't wait to put my analytical insights and data science development skills, abilities and talent to use to solve and kind of challenge and provide effective insight that will help to achieve our goals and plan for a distinguishable innovative future with BrideBook.

Personally, I like that Bridebook places a premium on original and engaging material. I'm certain that my innovative spirit and knack for solving problems would help protect and grow this competitive advantage. In addition, I am thrilled to have the chance to work in a challenging and fast-paced setting that will allow me to develop professionally.

Particularly appealing to me as a Data Scientist is the opportunity to integrate data from several sources, including that provided by users and marketers. It is my hope that I will be able make a contribution to the growth of the user base and the realization of financial prosperity through analysing the data and providing predictive and recommender systems to help BrideBook get more insight regarding the current state and the upcoming periods whether next month or year.

Finally, the opportunity to work with Bridebook excites me since I know that my background and expertise will be a great fit for the company's current initiatives and future plans. I'm excited by the prospect of playing a role in Bridebook's future growth.

I really appreciate your time and consideration of my profile and Cv.

1.3 Aim of the tasks :

The purpose of this project is to use A/B testing techniques on our data, determine the results, and then use those findings to generate strategic suggestions for our company. We want to unearth fundamental trends and useful insights that might inform our future strategic choices. Our focus also includes investigating any further insights or recommendations that the data may provide. If we had more resources, we'd ask more in-depth questions to understand better how users interact with the site. In addition, we will use principal component analysis (PCA) and exploratory data analysis (EDA) to comprehend the structure and variables of the data fully, so bettering our capacity to anticipate and enhance it. Finding what works is only part of the challenge; we also need to use the information we collect to inform our future decisions.

2- Tools and software :

The Scripts has been wrote by using Pycharm and Visual Studio Code:

But why using these? as a Data scientist code in PyCharm and Visual Studio Code. A/B testing requires PyCharm's clear interface, easy navigation, and strong Python programming tools. Data visualisation tools enhance A/B testing analysis. Visual Studio Code supports various languages and has many extensions. It can handle several tasks, including data processing and statistical modelling. Both IDEs include debugging and version control, simplifying code tracking and error management during A/B testing.

Language: has been used is Python, and runner Jupyter kernel

Libraries: the have been used

Data manipulation and analysis library pandas (pd).

numpy: Python's numerical computing library.

os: Allows operating system-dependent functionality.

matplotlib: A popular Python library for creating static, interactive, and animated visualisations.

geopandas: Python library for geospatial data.

scipy: Python's scientific computing and statistical library.

sys (import sys): A module that accesses interpreter variables and functions.

subprocess (import subprocess): Create new processes, connect to their input/output/ error pipes, and get their return codes.

platform (from platform import python_version): Common platform-specific functions.

random: A module for random sampling and number generation.

seaborn (import seaborn as sns): A high-level interface for producing attractive and useful statistical visuals based on matplotlib.

statsmodels: A library for estimating statistical models and hypothesis testing.statsmodels.stats: A submodule for percentage and count data tests.

.

3- Methodology

The task procedure start as follows, I will be implementing A/B testing which have a certain steps to properly implemented and tested as we aim from theA/B testing to determine if the options are different in a statistical sense “hypothesis testing” to be able to determine which option is better corresponding o which variant is better to enhance Bridebook traffic and user experience.

3.1 - A/B testing

steps that will be implemented for performing the test

Subject: which refers to the individuals participating in the study.

Randomization process: is a method used to assign participants to different groups in a way that ensures fairness and reduces bias. It involves randomly assigning participants to one of the three groups in this task we have control group “variant 0” and two treatments group “variant 1” and “variant 2”

Results: by exposing the subjects “finalize the operation” to the three groups that we have as variant from 0 to 2.

Hypothesis testing: which we will determine if the observed difference is staticicaly significant; can be done with permutation, t-test, chi square, z-test and Anova for multiple variant groups

Lastly, Decision/Action; which is based on the results.

Initially; we will start by setting our hypothesis

Hypothesis
H0: There is not statistically significant difference between the old page and new pages

H1: There is statistically significant difference between the old and new pages regarding control group and 2 treament groups variant 1 and 2 of finalizing operation.

Secondly, I have started with initialization of my environment the adding the packages and installing libraries that will be used

Then, I have adjusted Some plot styling preference

```
# packages that will be used and extra import that might be

%matplotlib inline

import pandas as pd
import numpy as np
import os
import matplotlib.pyplot as plt
import geopandas as gpd
import scipy.io
import sys
import subprocess
from platform import python_version
from random import sample
from random import seed
import scipy.stats as stats
import matplotlib as mpl
import seaborn as sns
from math import ceil
import statsmodels.api as sm
#from statsmodels.stats import some_function_from_sms_module
from statsmodels.stats import proportion
```

```
import scipy.stats as stats
import statsmodels.stats.api as sms
import matplotlib as mpl
import matplotlib.pyplot as plt
import seaborn as sns
from math import ceil


#Some plot styling preferences
%matplotlib inline

# Some plot styling preferences
plt.style.use('seaborn-whitegrid')
font = {'family' : 'Helvetica',
        'weight' : 'bold',
        'size'   : 14}

mpl.rc('font', **font)
effect_size = sms.proportion_effectsize(0.13, 0.15)    # Calculating effect size based on our exp

required_n = sms.NormalIndPower().solve_power(
    effect_size,
    power=0.1,
    alpha=0.05,
    ratio=10
    )                                                  # Calculating sample size needed

required_n = ceil(required_n)                          # Rounding up to next whole number

print("required_n value is " + str(required_n))
```

finOnb	0	1
testVariant		
0	51	536
1	27	270
2	11	156

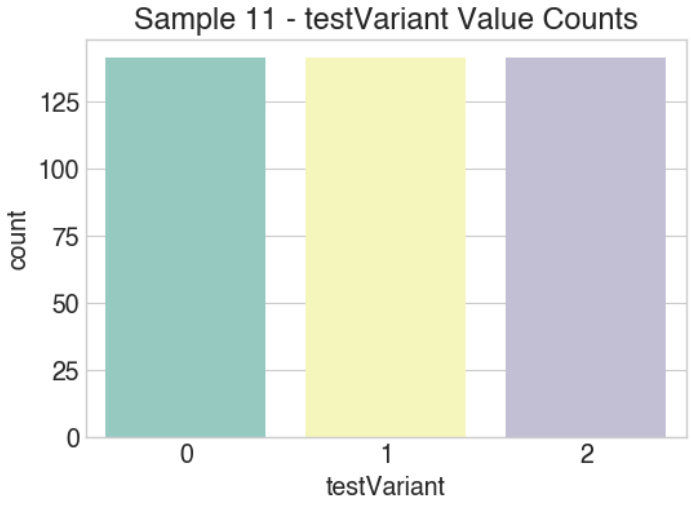
Then I have proceeded with loading the data the coking for duplicate and nan values and apply cross tab for checking contingency table by crosstab function which values as shown users whom finalized op from control group are 536 and treatment group one with testVarint 1 are 270, and users whom finalized op from control group are 536 and treatment group two with testVarint 2 are 156

Following with making sure e there are no users that have been sampled multiple times.

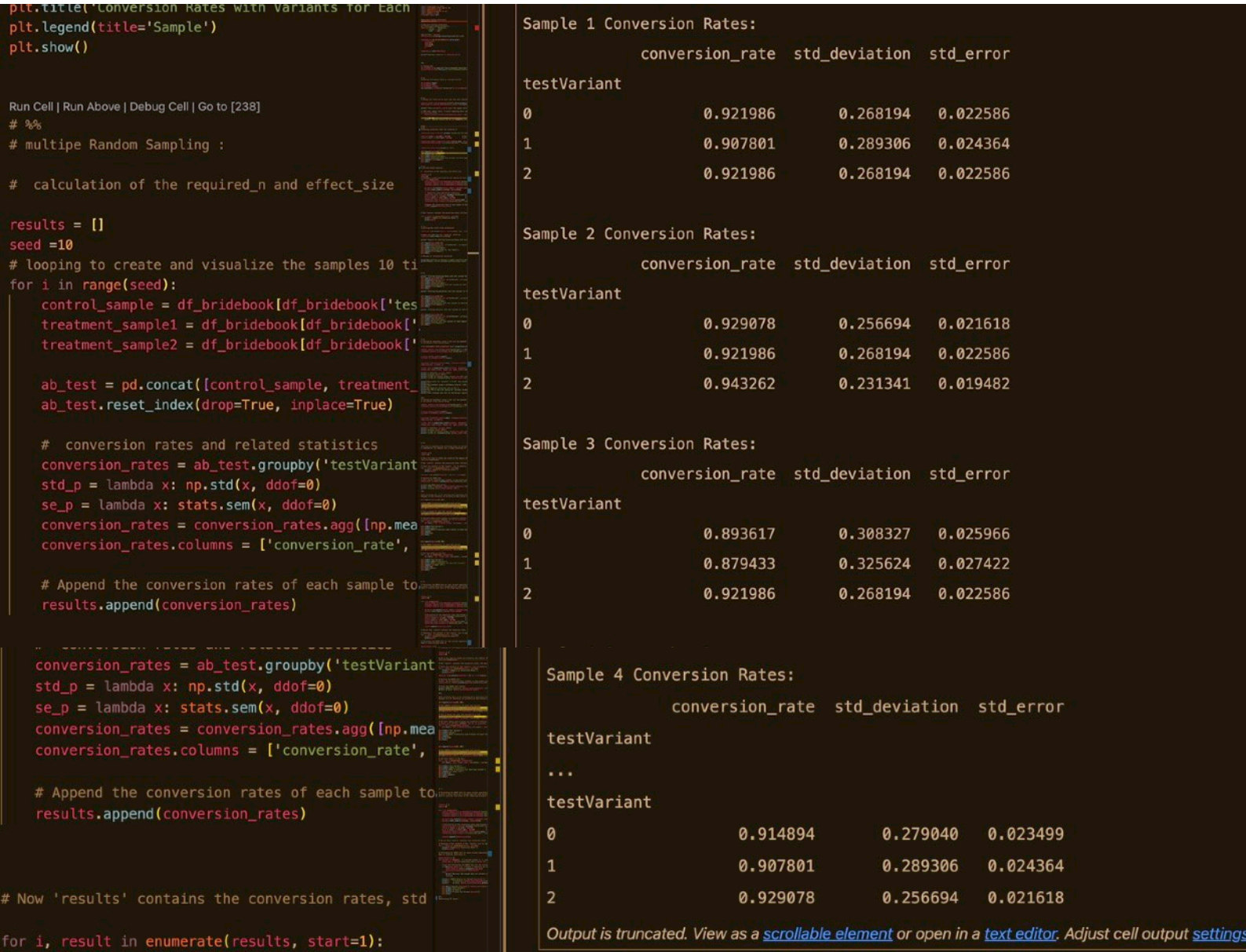
And after running the python cell it printed There are 0 users that appear multiple times in the dataset.

Proceeding with random sampling which is a method used in research to select a sample from a larger population. It ensures that each member of the population has an equal chance of being included in the sample, reducing bias and increasing the generalizability of the findings. In my case the data that have give is not large as it's a sample of the kind of the data of Bridebook, eventhogh I performed the random sampling and followed by the rest of procedure in the file scrip.py in the actual large data in BrideBook

I have separated the sample into 3 groups control_sample0, treatment_sample1, treatment_sample2 the concatenated the in "ab_test" and checked the info and by checking the info I have found some null values in location country and browser version which will not impact our AB test the counted the values of the samples and plated them as shown in the next graph



Following with Checking conversion rates and plot it with multipe Random Sampling and calculating the required_n and effect_size as shown in the following pictures which indicates that the conversion rate in the 4 sample is higher with variant 2 than varian one and control group as in simple 1 and 2 is slightly higher but in sample 3 and 4 is little bit more higher than the rest with 92% corresponding to 87% in sample 3.



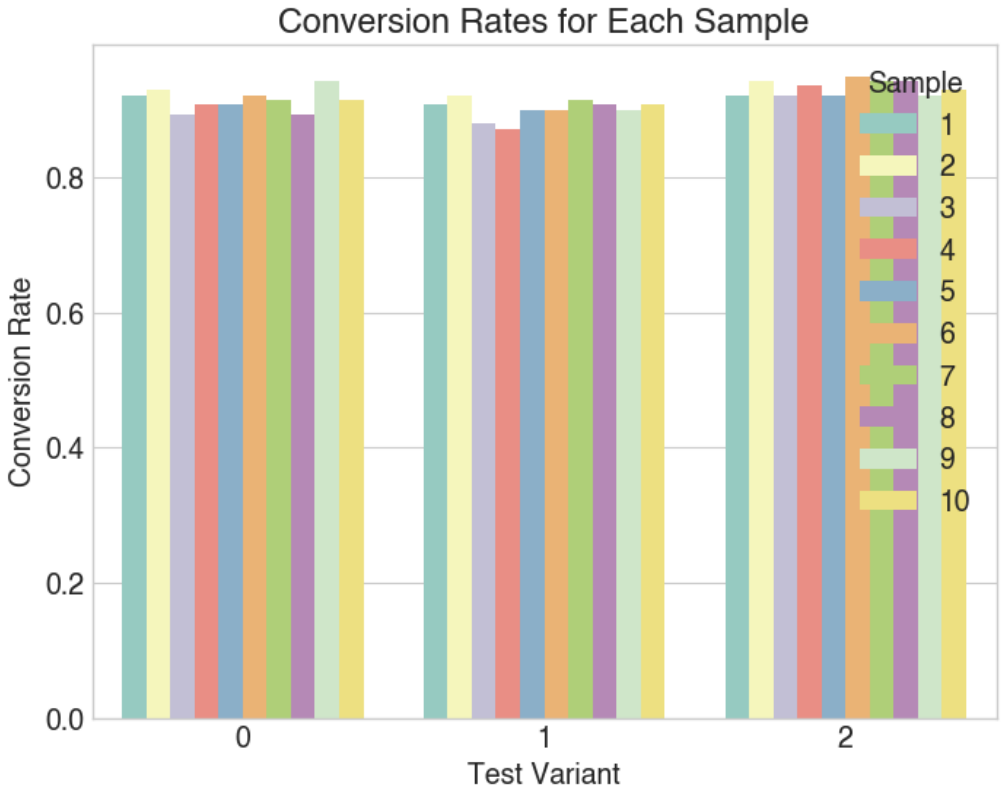
```
plt.title('Conversion Rates with Variants for Each
plt.legend(title='Sample')
plt.show()

Run Cell | Run Above | Debug Cell | Go to [238]
# %%
# multipe Random Sampling :

#  calculation of the required_n and effect_size

results = []
seed =10
# looping to create and visualize the samples 10 ti
for i in range(seed):
    control_sample = df_bridebook[df_bridebook['tes
    treatment_sample1 = df_bridebook[df_bridebook['
    treatment_sample2 = df_bridebook[df_bridebook['

    ab_test = pd.concat([control_sample, treatment_
    ab_test.reset_index(drop=True, inplace=True)

    #  conversion rates and related statistics
    conversion_rates = ab_test.groupby('testVariant
    std_p = lambda x: np.std(x, ddof=0)
    se_p = lambda x: stats.sem(x, ddof=0)
    conversion_rates = conversion_rates.agg([np.mea
    conversion_rates.columns = ['conversion_rate',

    # Append the conversion rates of each sample to
    results.append(conversion_rates)
```

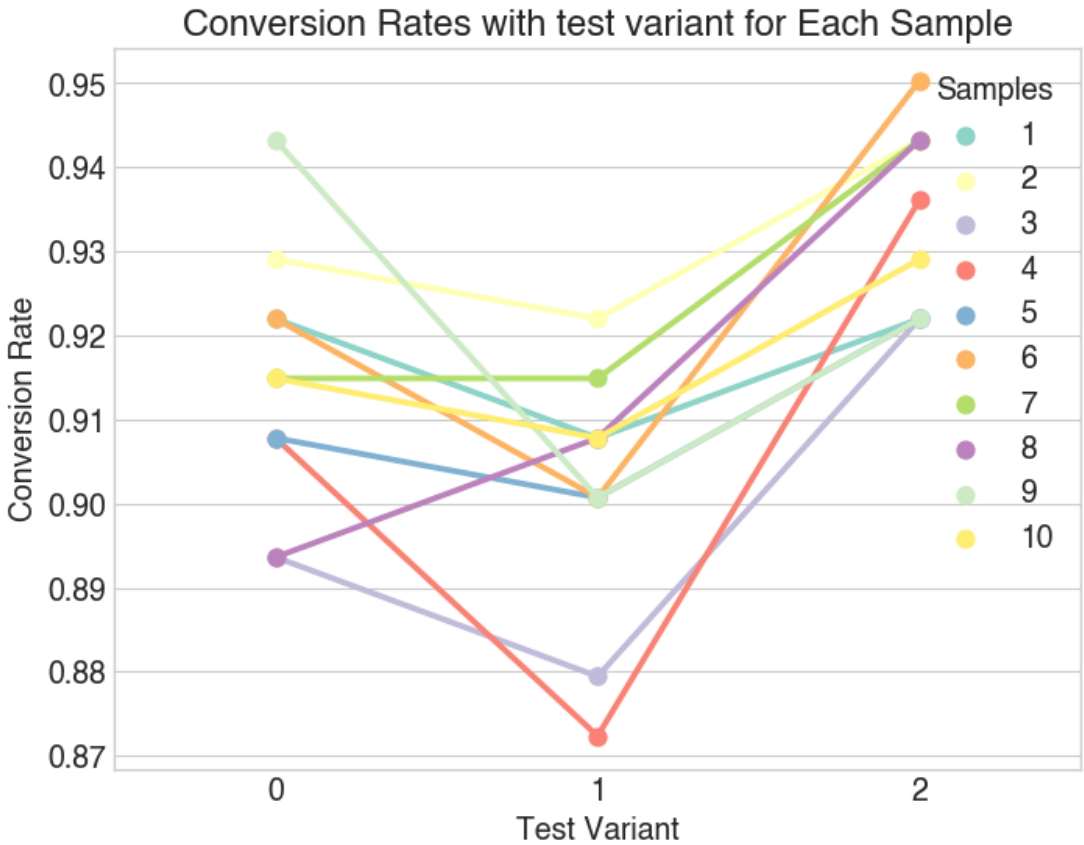
```
Sample 1 Conversion Rates:
             conversion_rate  std_deviation  std_error
testVariant
0                   0.921986       0.268194   0.022586
1                   0.907801       0.289306   0.024364
2                   0.921986       0.268194   0.022586

Sample 2 Conversion Rates:
             conversion_rate  std_deviation  std_error
testVariant
0                   0.929078       0.256694   0.021618
1                   0.921986       0.268194   0.022586
2                   0.943262       0.231341   0.019482

Sample 3 Conversion Rates:
             conversion_rate  std_deviation  std_error
testVariant
0                   0.893617       0.308327   0.025966
1                   0.879433       0.325624   0.027422
2                   0.921986       0.268194   0.022586
```

```
    conversion_rates = ab_test.groupby('testVariant
    std_p = lambda x: np.std(x, ddof=0)
    se_p = lambda x: stats.sem(x, ddof=0)
    conversion_rates = conversion_rates.agg([np.mea
    conversion_rates.columns = ['conversion_rate',

    # Append the conversion rates of each sample to
    results.append(conversion_rates)



# Now 'results' contains the conversion rates, std

for i, result in enumerate(results, start=1):
    print(f"\nSample {i} Conversion Rates:")
```

```
Sample 4 Conversion Rates:
             conversion_rate  std_deviation  std_error
testVariant
...
testVariant
0                   0.914894       0.279040   0.023499
1                   0.907801       0.289306   0.024364
2                   0.929078       0.256694   0.021618
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings
```

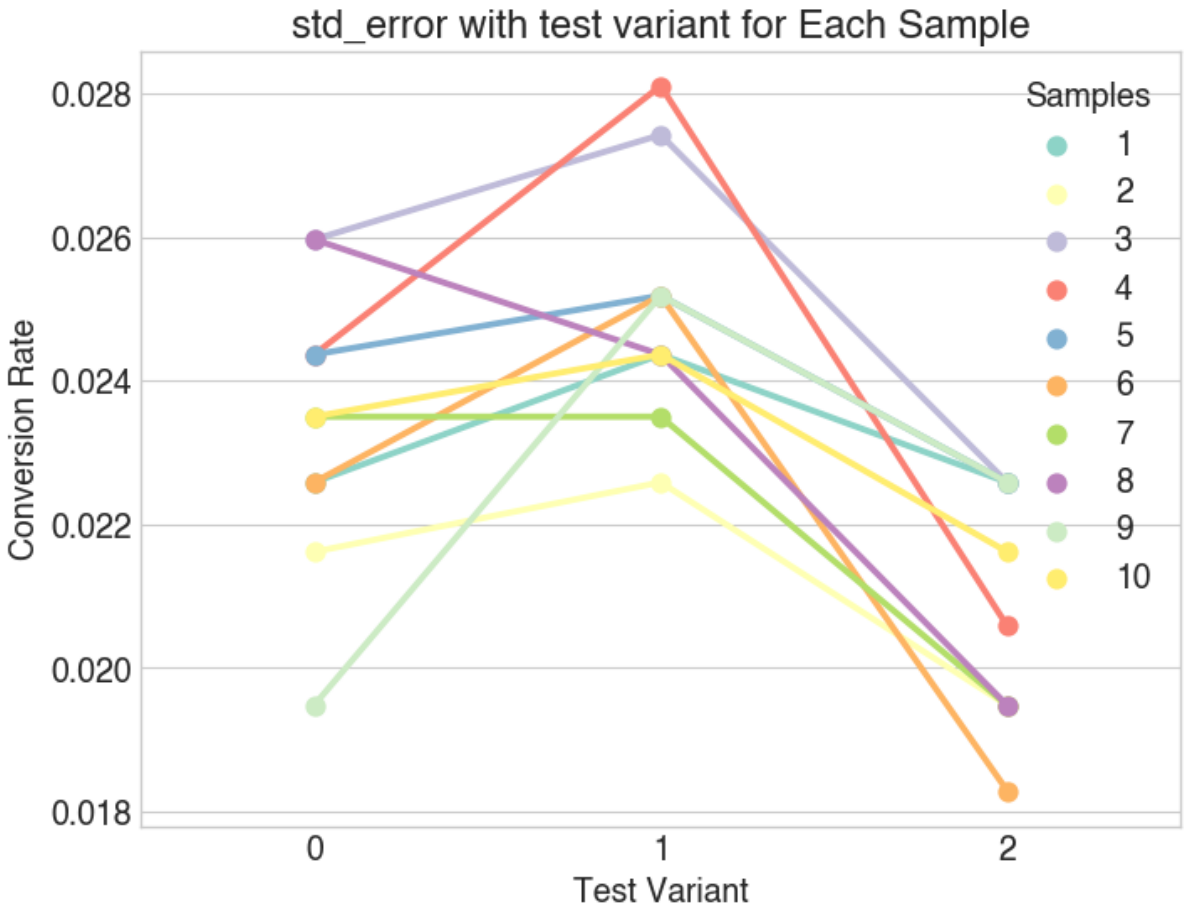
Let's visualize our plotting to see the difference between conversion rates as Bar plotting is showing a slightly positive indication towards a particular variant.However, we will proceed after by plotting using a point plot for better visualization.



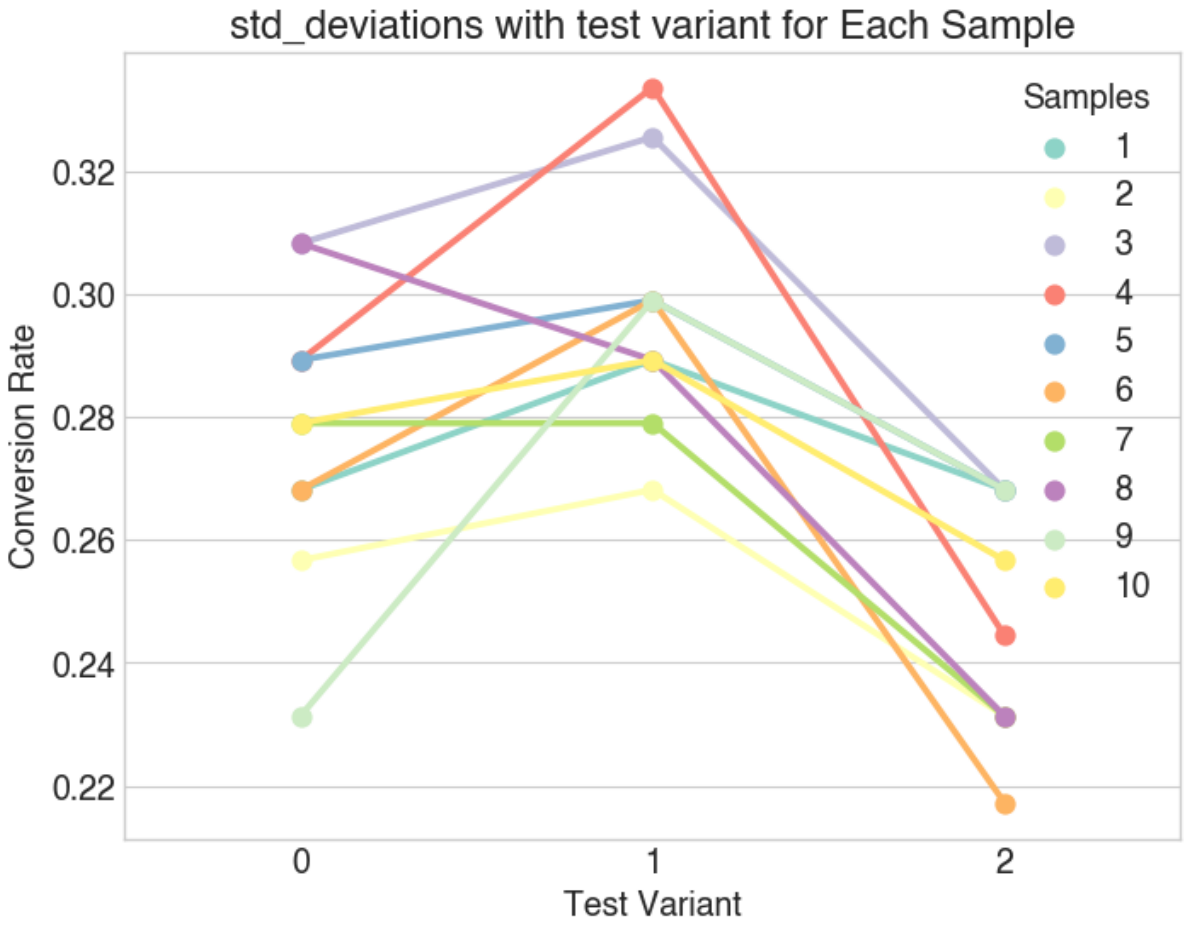


As shown for the standard error (std error) evaluates how well a sample represents a population. A standard error between 0.022 and 0.017 shows little dispersion from the mean. Because the observations are clustered around the mean, the sample data may represent the population well. The sample is more representative of the population when the standard error is less.

For the STD In context determines whether a bigger standard deviation is beneficial. A high standard deviation, say 0.4 in sample 4 of variant 1, may indicate a wide data set dispersion and substantial value variability. For example; In ecological biodiversity studies, this may be beneficial [4]. In manufacturing and quality control, a reduced standard deviation means outcomes are closer to the mean and more consistent [4]. Thus, a standard deviation of 0.4 depends on your data set and research setting [4][5]. In our Case the distribution of the samples regarding variant 2 is properly consistent with proper variation values of deviations on the other side variant 1 show

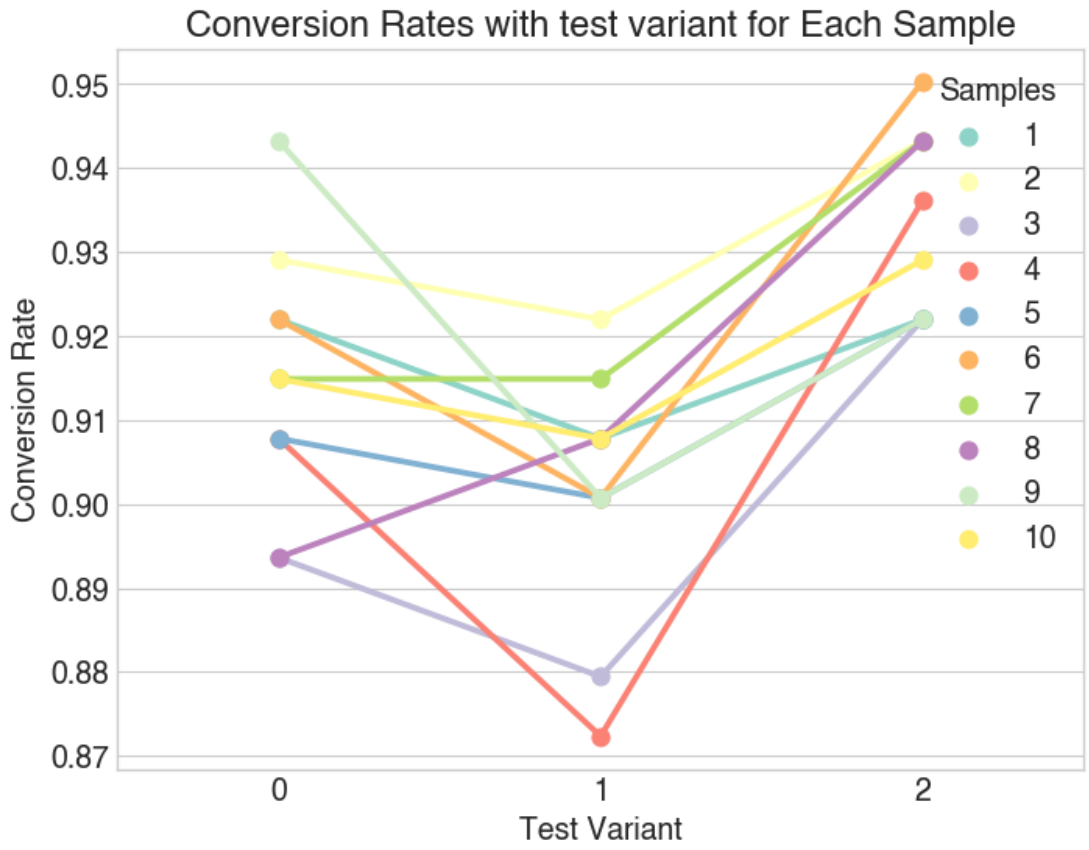


inconsistencies in deviation values variations.



The conversion rates of 0.95 and 0.92 indicate the relative success of our online marketing campaigns, as a higher conversion rate is generally desired [5]. These rates imply that 95% and 92% of total visits on your digital assets resulted in the intended user actions, as variant 2 in our case hid the text of variant 2 ""we have some venues for you" is scoring a high rate on CTA which indicates better-predetermined objective on "finob". However, the slight decline from 0.95 to 0.92 may indicate a slight decrease in the efficacy of your campaign or a slight shift in user behavior. which could be tested by A/B by taking samples with limiting each group by a certain duration. This modest variation could be the result of a number of factors, such as shifting market trends, competition, or consumer preferences.[6] To optimize

our marketing strategies, it is crucial to continually monitor and analyze these rates.

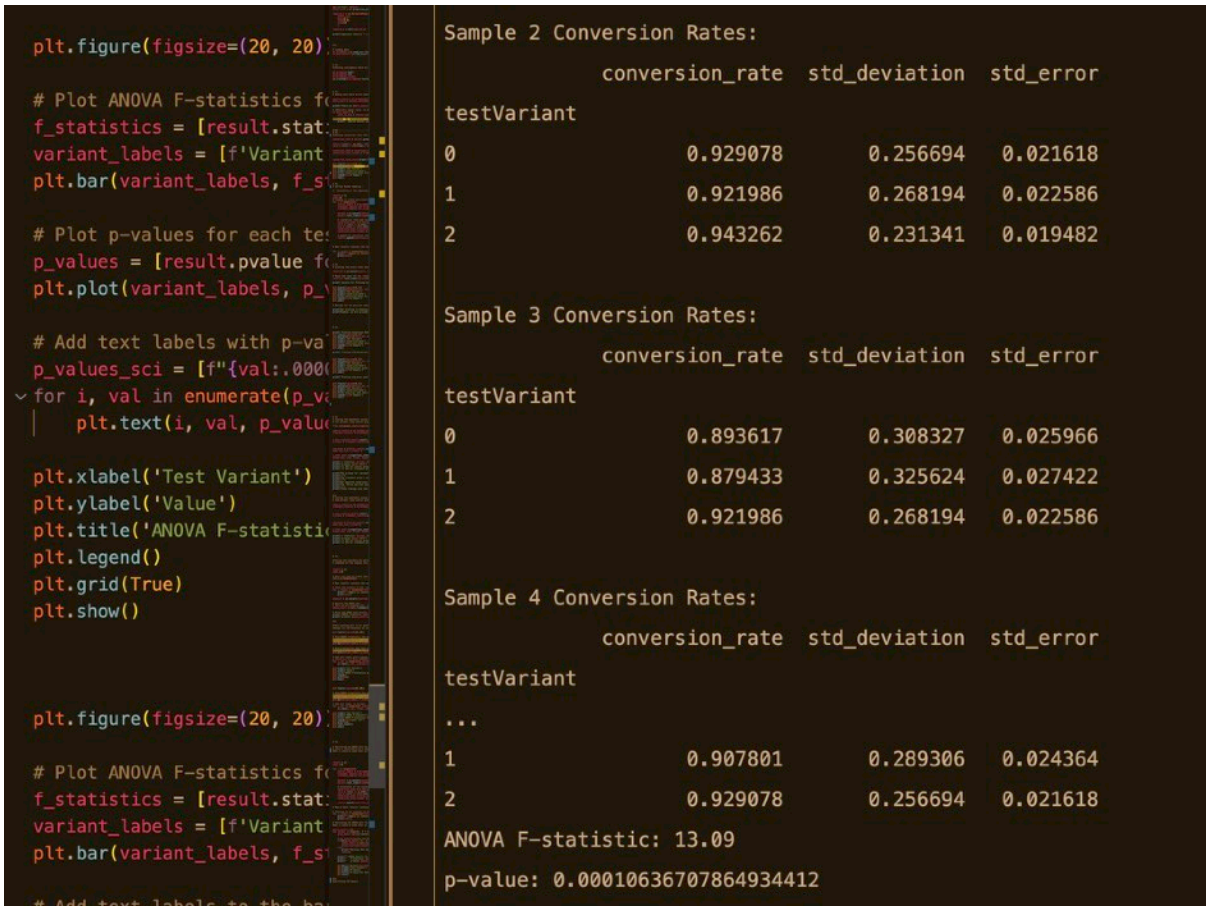


Hypothesis Testing:

By testing the hypothesis using Anova, the result of a statistical test comparing the means of three or more groups to discover whether there is a significant difference. Results interpretation:

ANOVA F-statistic: The F-statistic compares group means to group variability. Our F-statistic is 13.09. A higher F-statistic suggests a bigger difference between group means than within the group variance.

P-value: represents the chance of getting the observed findings (or more extreme) on the premise that there is no real difference between the group means. our p-value is 0.00010636707864934412. This p-value is exceptionally low (less than 0.05), suggesting that group means vary statistically.



```
plt.figure(figsize=(20, 20)

# Plot ANOVA F-statistics f
f_statistics = [result.stat
variant_labels = [f'Variant
plt.bar(variant_labels, f_s

# Plot p-values for each te
p_values = [result.pvalue f
plt.plot(variant_labels, p_

# Add text labels with p-va
p_values_sci = [f"{val:.000
for i, val in enumerate(p_v
    plt.text(i, val, p_valu

plt.xlabel('Test Variant')
plt.ylabel('Value')
plt.title('ANOVA F-statisti
plt.legend()
plt.grid(True)
plt.show()



plt.figure(figsize=(20, 20)

# Plot ANOVA F-statistics f
f_statistics = [result.stat
variant_labels = [f'Variant
plt.bar(variant_labels, f_s
```

```
Sample 2 Conversion Rates:
             conversion_rate  std_deviation  std_error
testVariant
0                   0.929078       0.256694   0.021618
1                   0.921986       0.268194   0.022586
2                   0.943262       0.231341   0.019482

Sample 3 Conversion Rates:
             conversion_rate  std_deviation  std_error
testVariant
0                   0.893617       0.308327   0.025966
1                   0.879433       0.325624   0.027422
2                   0.921986       0.268194   0.022586

Sample 4 Conversion Rates:
             conversion_rate  std_deviation  std_error
testVariant
...
1                   0.907801       0.289306   0.024364
2                   0.929078       0.256694   0.021618
ANOVA F-statistic: 13.09
p-value: 0.00010636707864934412
```

Based on the previously mentioned ANOVA shows a significant difference in group means. Thus, one group has a different mean. The tiny p-value

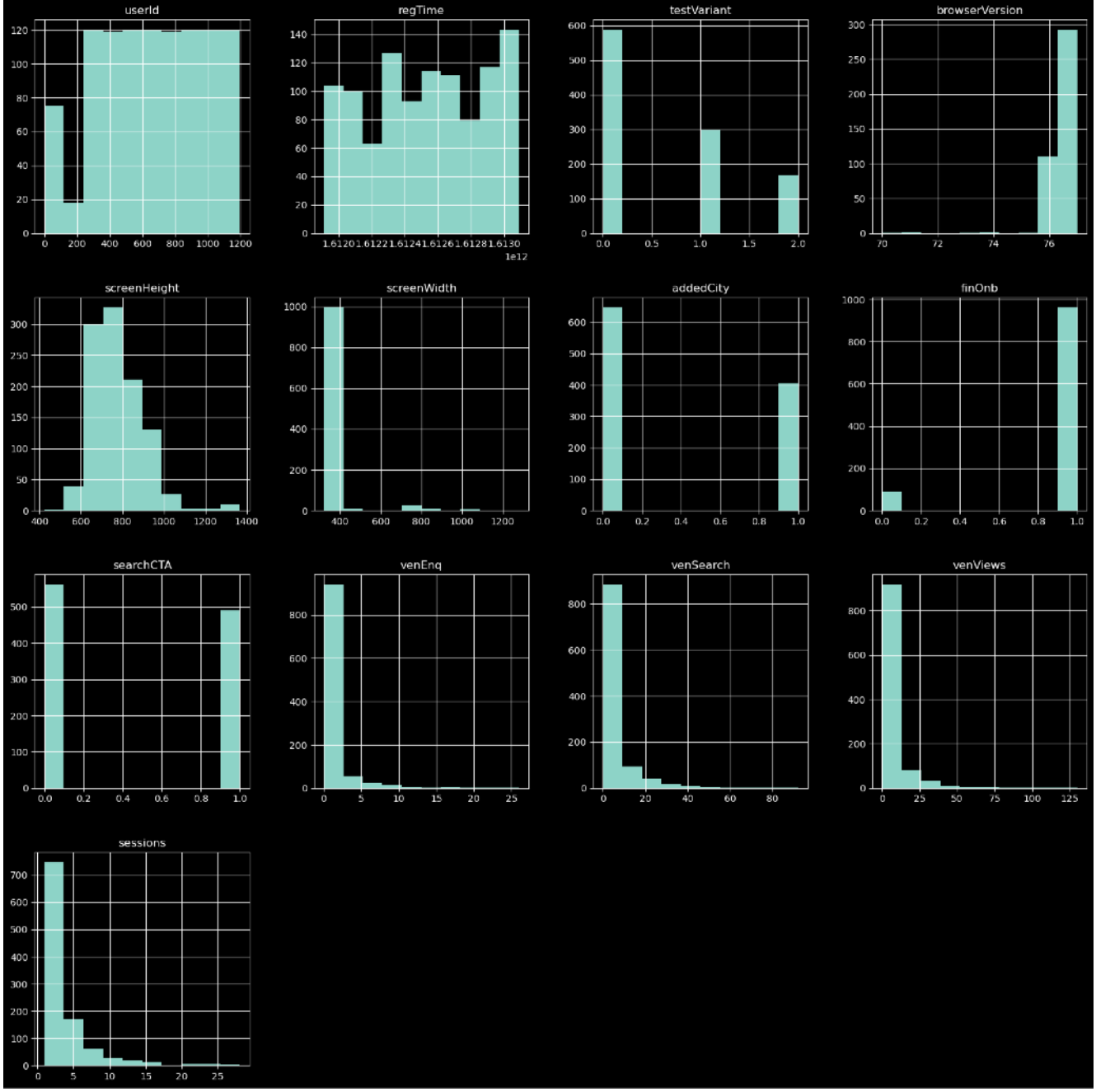
suggests that the observed difference is unlikely to be random chance and supports rejecting the null hypothesis.

In a nutshell the ANOVA F-test demonstrates a statistically significant difference in the means of the groups being compared, and post hoc tests may be performed to establish which groups have quite different means.

3.2- EDA

Exploratory Data Analysis (EDA) is a crucial step in data analytics that emphasises the significant characteristics of a dataset in order to identify its patterns and trends [9]. EDA provides insights into consumer behaviour, market trends, and business performance through the detection of data anomalies [9]. It enables data analysts to identify voids or incomplete data[8], outliers, and determine the crucial variables required for analysis [8]. EDA employs methods such as data visualisation and correlation analysis as the initial phase in the development of a data model [7]. This process of deciphering data through EDA is crucial because it provides a comprehensive comprehension of the data, resulting in more accurate forecasts and optimised strategies [9].

I started exploring the data to comprehend the provided information and features. I performed some steps like info of the data and it's types and checking for duplicated values and checking the null values and check if it's related to the aim to drop it, it was unnecessary to drop. Then visualized the features by using histogram plot to be able to see the features values distribution, as shown in the next graph output.



By checking the means of the multiple features by grouping it with corresponding features so we can get indications of our data statistics and distribution. I've found the following; By grouping and aggregating [testVariant , finOnb] numbers shows that the average user lean to finalize the operation on variant 2 by slightly higher value of 0.02 with 0.93 compared to the control group with 0.913.

```
df.groupby(['testVariant']).agg({'finOnb': 'mean'})
```

testVariant	finOnb
0	0.913118
1	0.909091
2	0.934132

And By grouping and aggregating [testVariant , venSearch] numbers shows that the mean value of users on variant2 is 50.9 compared to control group and treatment group1 sequentially 5.2 and 4.1 which show is less than the distribution of group 1. But on Venue enquiry control group is likely higher than the rest but we are keeping in mind that number of users in control is almost 4 times the size so later we will be checking on ratios.

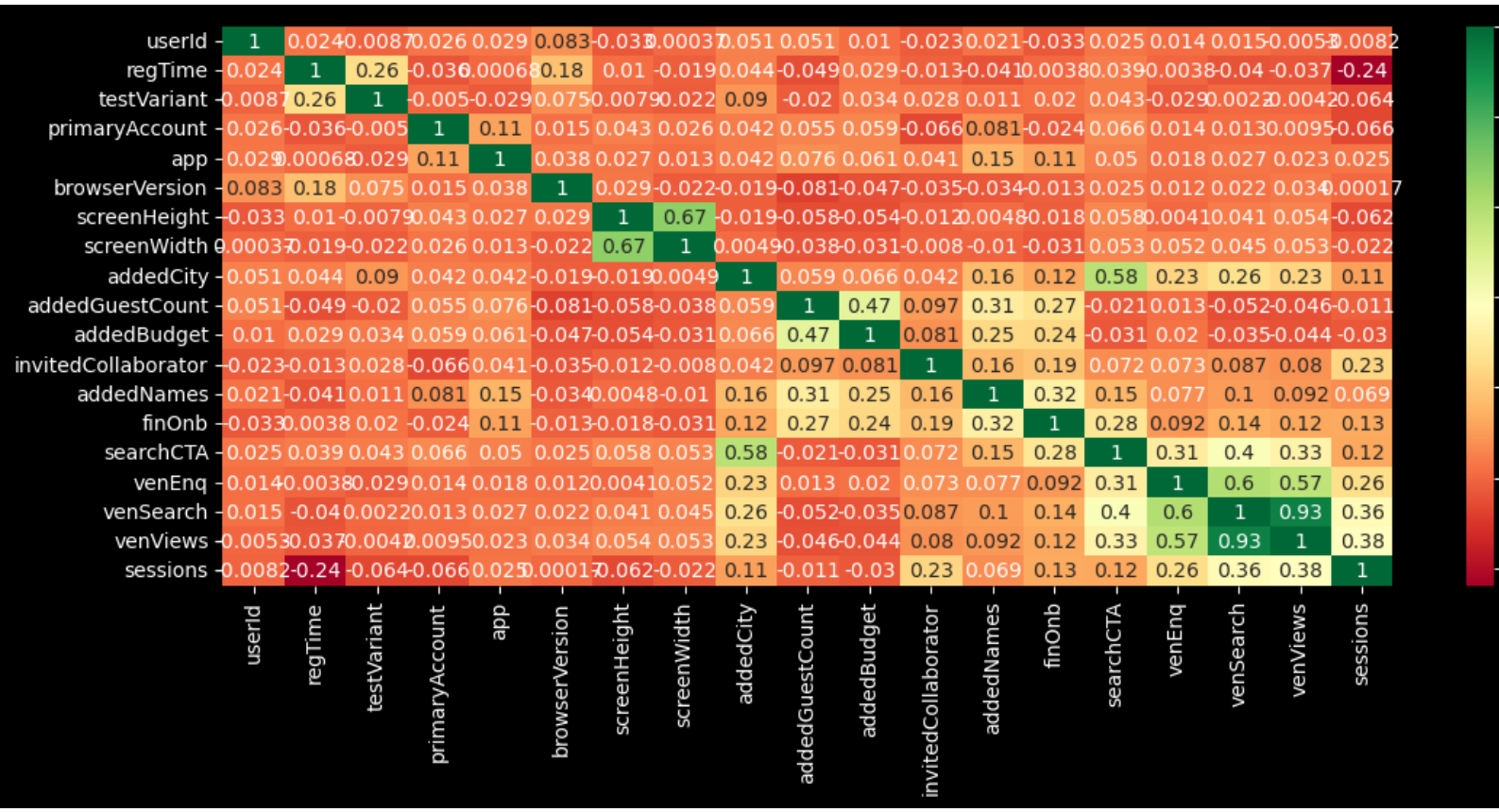
```
df.groupby(['testVariant']).agg({'venSearch': 'mean'})
```

testVariant	venSearch
0	5.287905
1	4.198653
2	5.916168

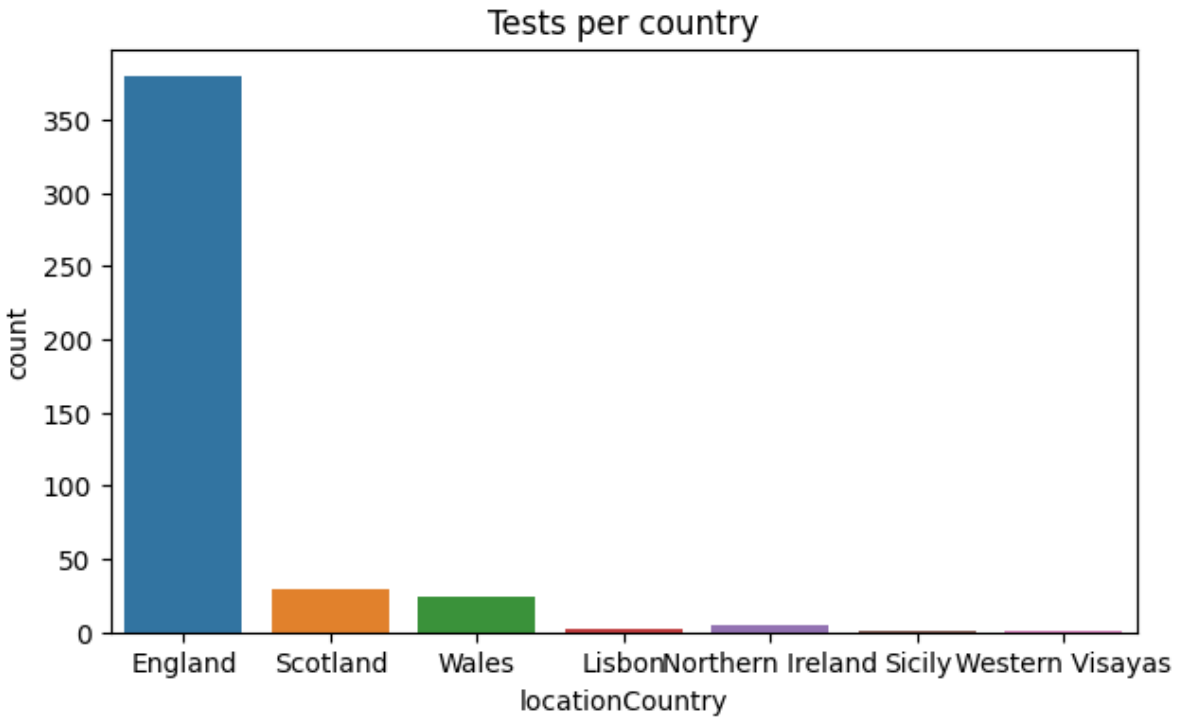
```
df.groupby(['testVariant']).agg({'venEnq': 'mean'})
```

testVariant	venEnq
0	0.921635
1	0.875421
2	0.688623

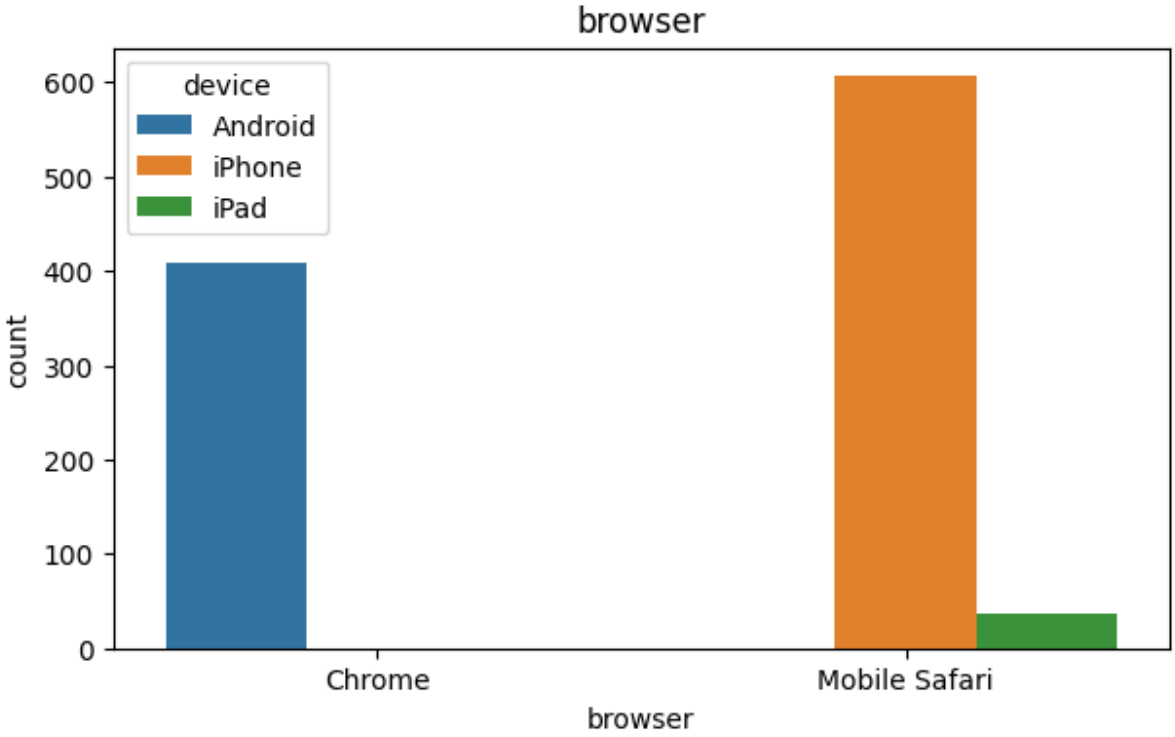
Then proceeded by implementing the correlation function and plotting the heat map of the correlation between the features. As, firms cannot effectively foresee market trends and client behaviour without data-driven decision making enabled by correlation analysis [10]. and the results showed some high correlations between and “finob” and features = ["addedNames", "addedGuestCount", "addedCity", “searchCTA"] with the following values [0.32, 0.27, 0.16,.028]
These feature are vital to work on as it will provide us better insights to work on and performance of upcoming models. The heat map graph is portrayed below.



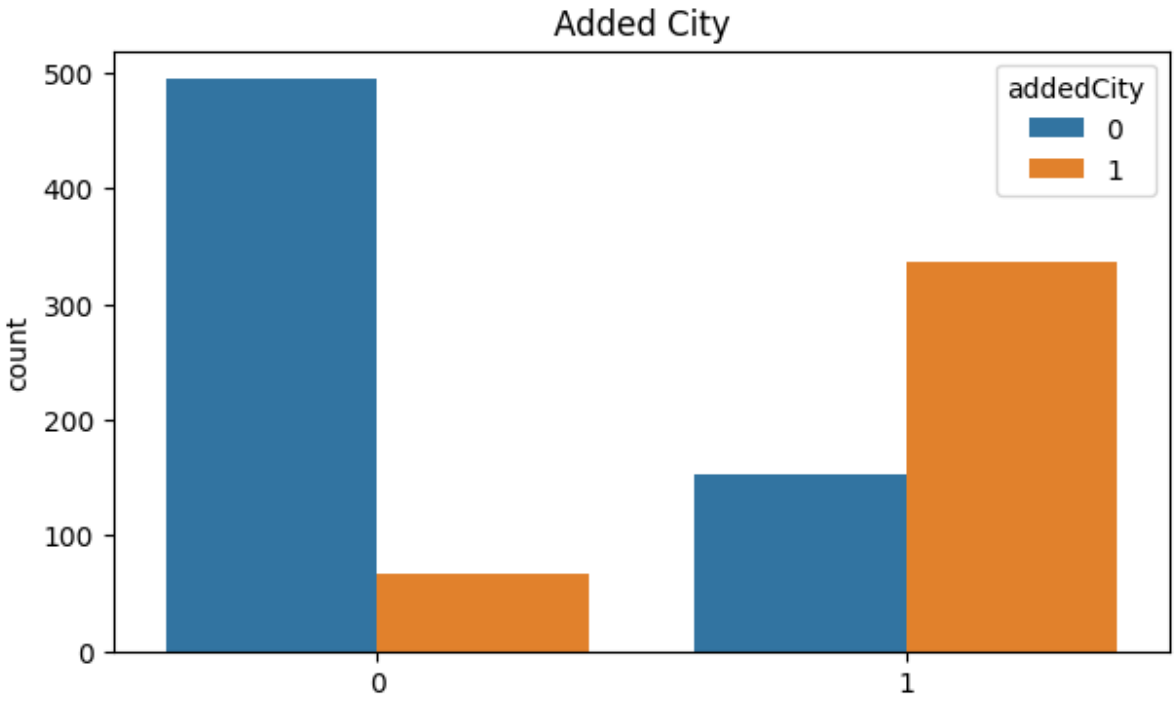
In the following I will provide some insights on Bridebook's dataMost of the users are located in England. Scotland and Wales But England has the highest con density of users, which later on could consider potnetially other countries for expansion plans!

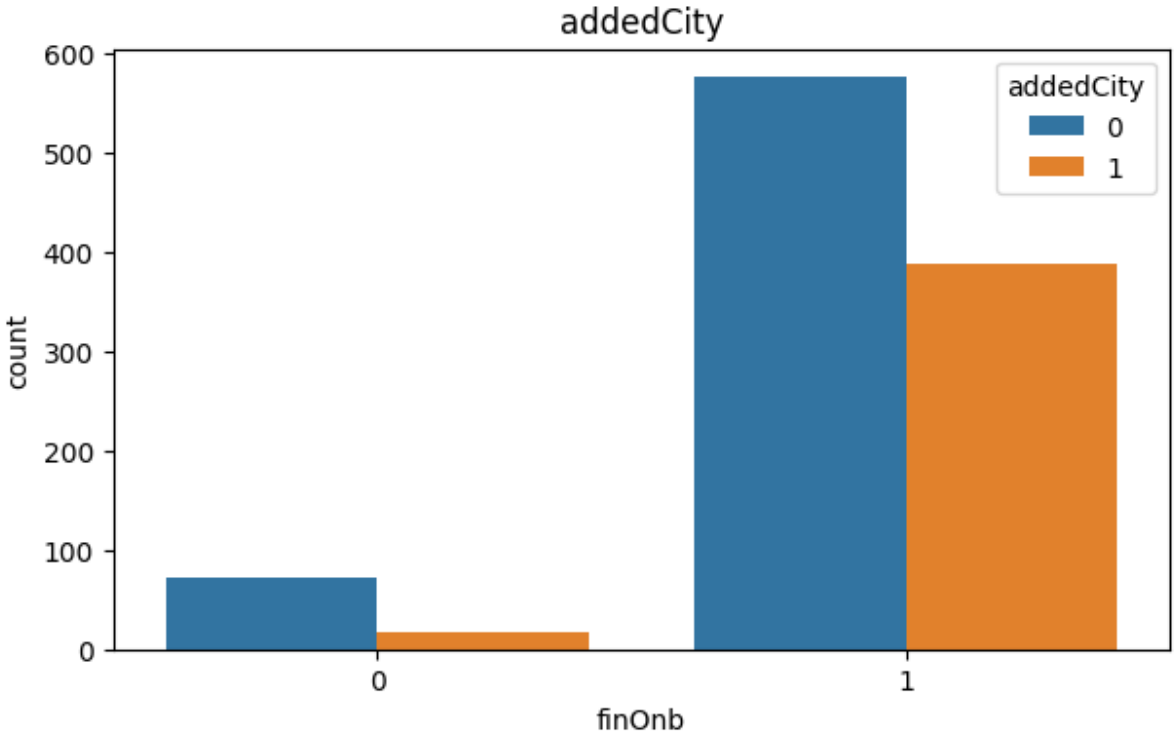


Iphone(with safari) is the most common device used with around 600 users compared to around 400 users of Android(with chrome)

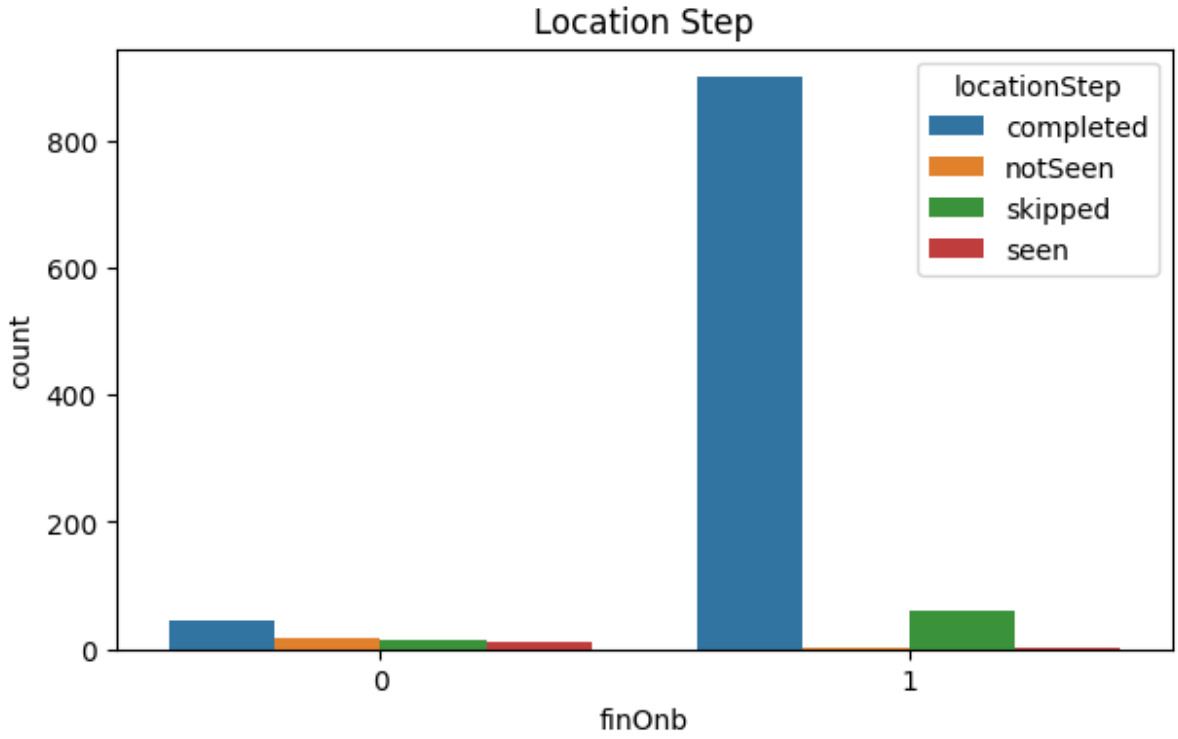


For the following insights, we will go through how the features are correlated with Finishing on Boarding and Search CTA. For each feature, we will discover the relation and then check the user ratios who used the feature. The next graph it indicates that user who add the city are most likely to continue search "CTA" as shown in the orange bar and user who didn't add city are most likely to finalize operation on boarding.

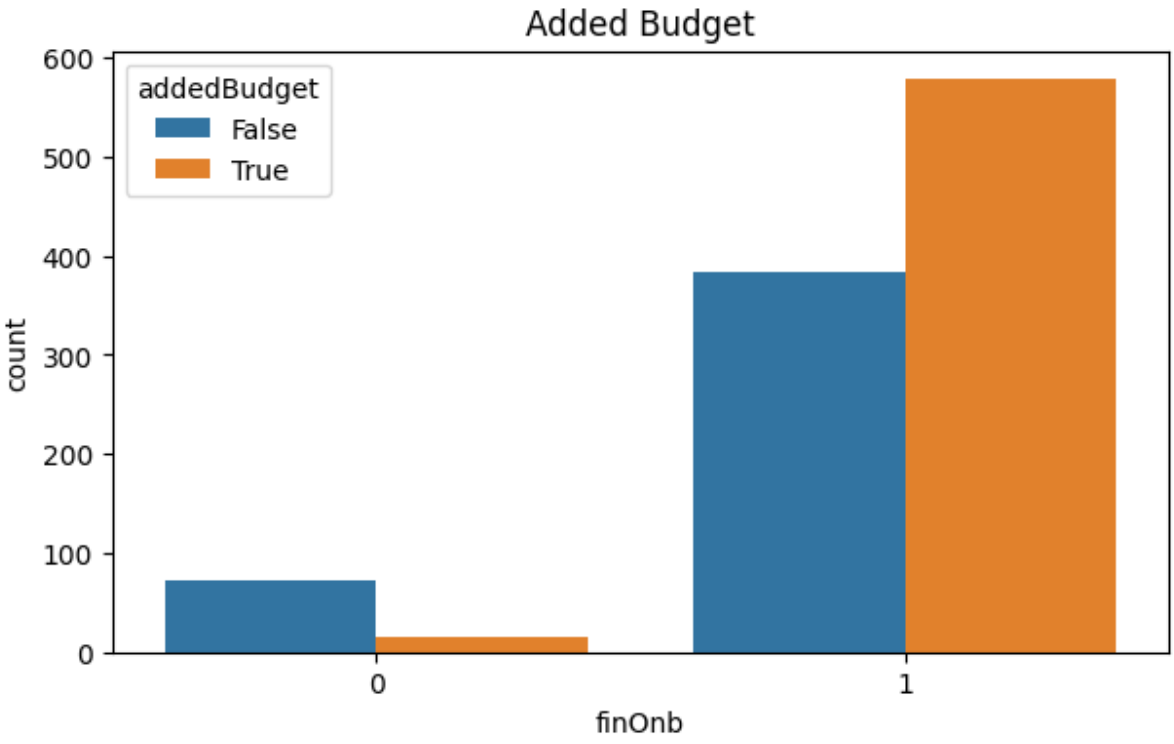


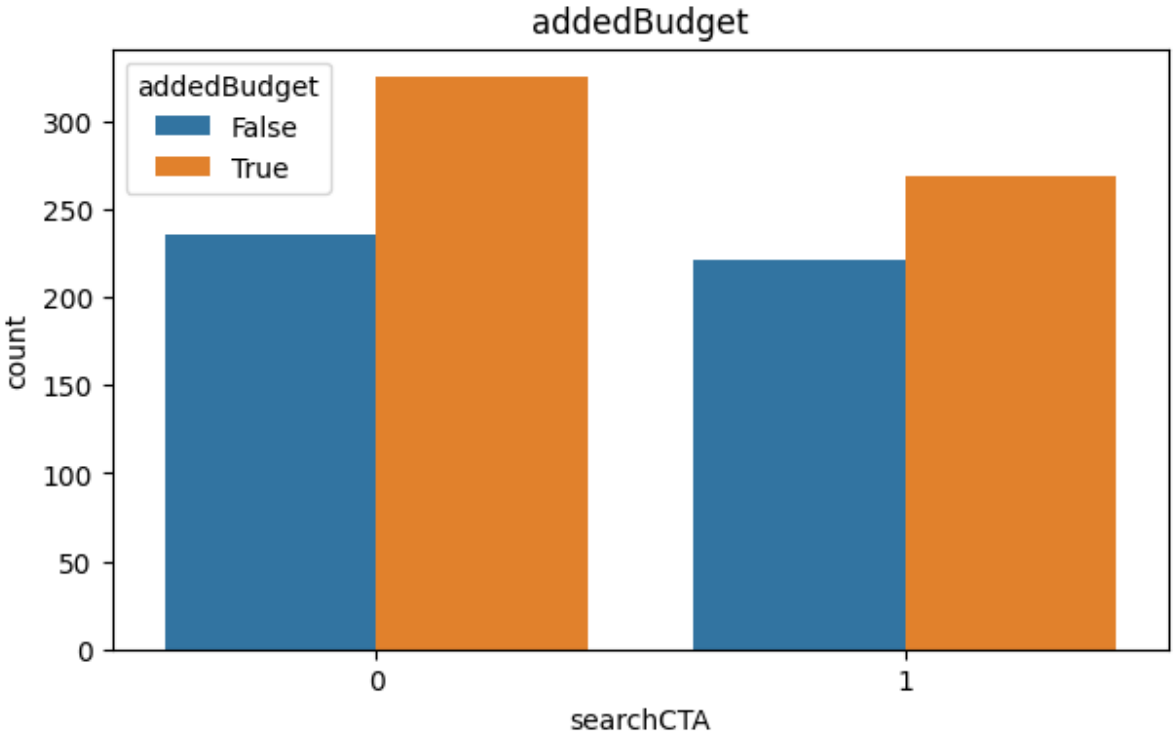


Users who completed the location step city are more likely to use finish on boarding! with some percentage of people who skipped the location step.

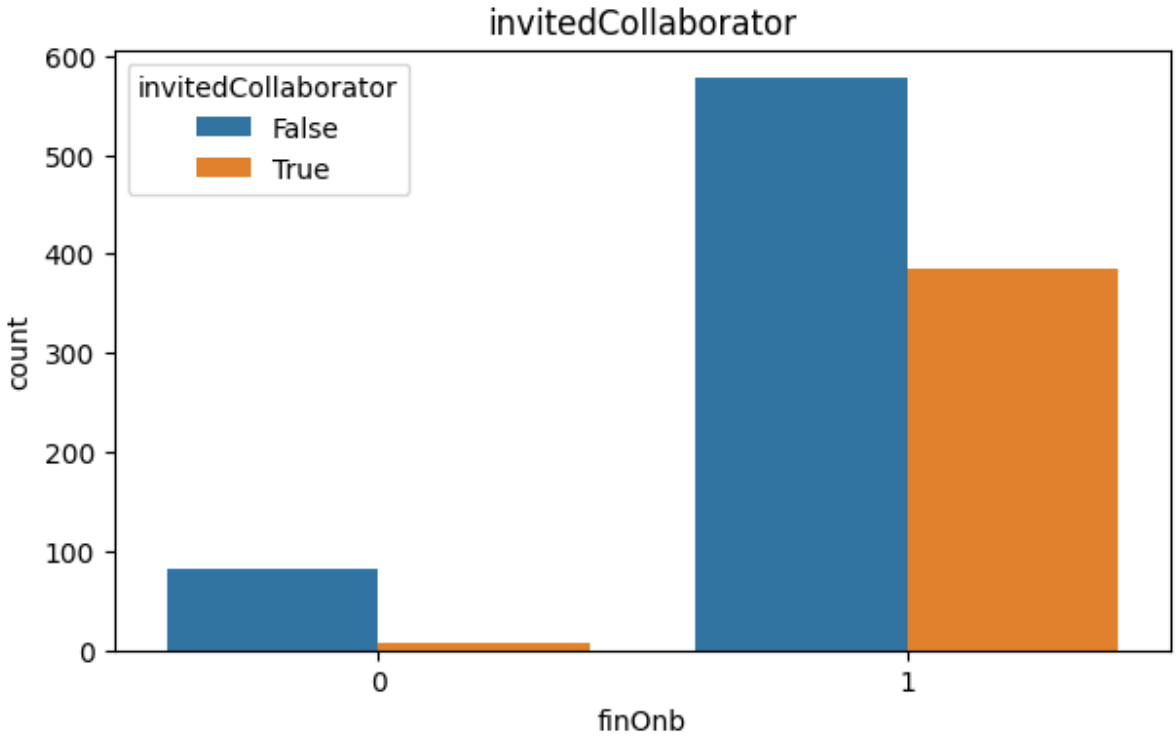


Users who added the budget will proceed to be on board and Search CTA as shown in the grapghs this shows that the feature of adding the budget has an impact and is important.

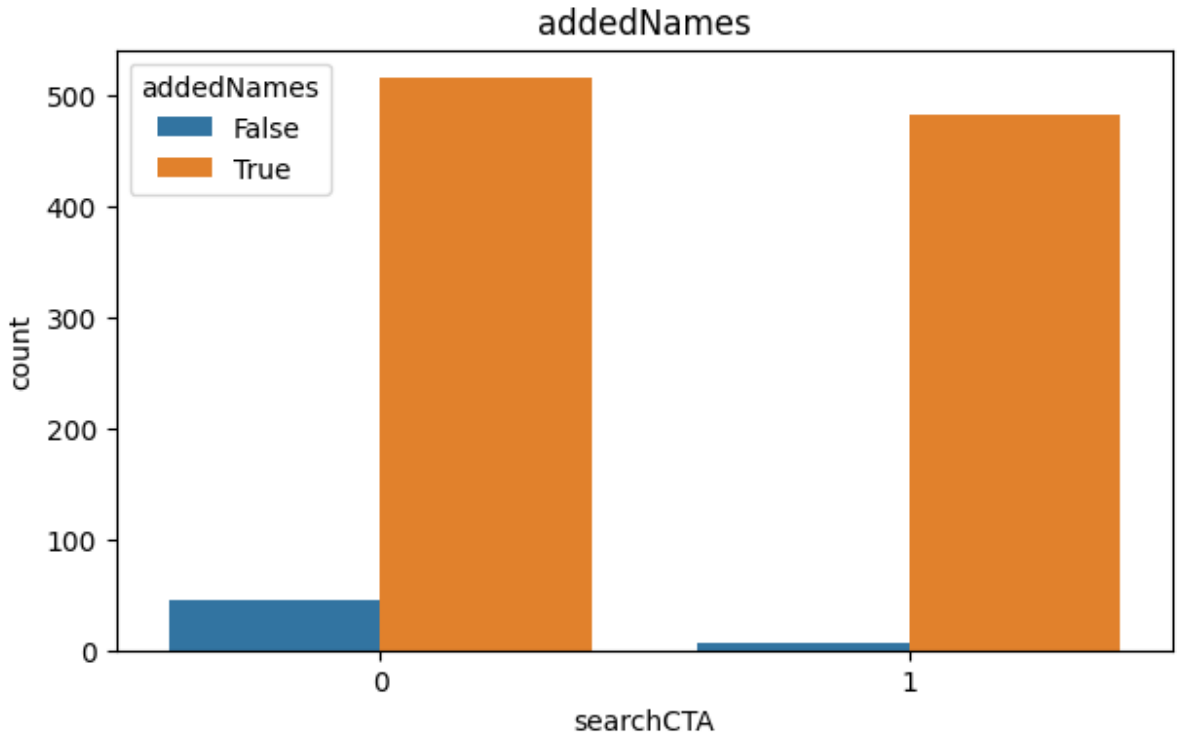


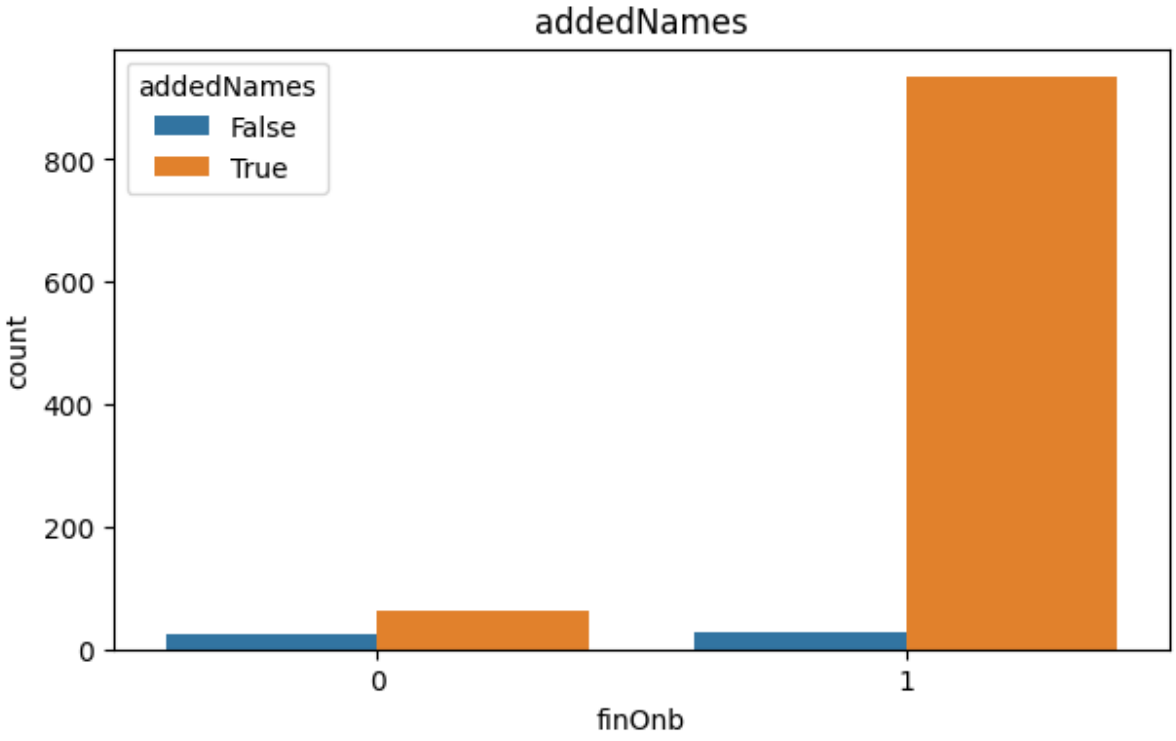


Regarding the invite collaborators feature results shows that users who didn't invite collaborators are highly proceeding to be on board and continue search.

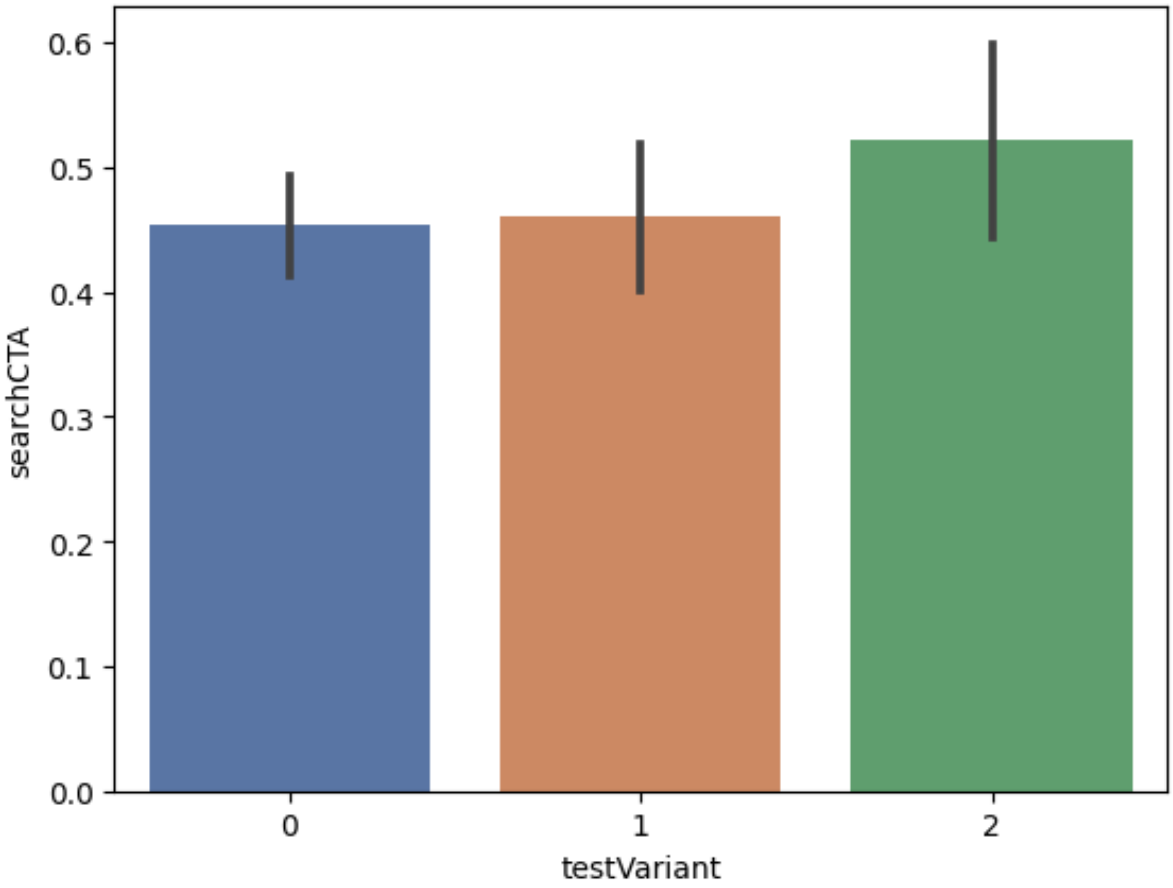


Last feature that consider is add names, results provided us with some indications that user who didn't add names the proceed with CTA with slightly higher difference. On the other side users who addnames are way likely to proceed to be on board by huge difference





The following graph is portraying that control to action performs better with variant 2.



3.3-PCA

Defined a list of feature names that include the column names "addedNames", "addedGuestCount", "addedCity", and "searchCTA". These are the characteristics you should employ in your analysis. Create and store another instance of the DataFrame df in the variable X.

The target variable "finOnb" was extracted from X and stored in the variable y. This is the variable that requires estimation. Only the columns specified in the features list are extracted from X. This creates a new DataFrame with only the specified features. Standardize the chosen DataFrame X_scaled features. Standardization is a preprocessing phase in which the features are scaled to have a mean of zero and a variance of one. This is done so that all characteristics are on the same scale and so that larger features don't skew the results of the study.

I have started with defining a list of feature names with the column names "addedNames", "addedGuestCount", "addedCity", and "searchCTA". Based on the results we got the correlation map. then Copy the DataFrame df to X.

and Stored "finOnb" in y from X. Predict this variable. extracted X's features list columns. This creates a DataFrame with only the selected features.

Standardise selected DataFrame X_scaled features. Standardisation preprocesses features to zero mean and unit

variance. This ensures that all characteristics are the same scale and prevents larger features from dominating the study.

```
features = ["addedNames", "addedGuestCount", "addedCity", "searchCTA"]
X = df.copy()
y = X.pop('finOnb')
X = X.loc[:, features]

# Standardize
X_scaled = (X - X.mean(axis=0)) / X.std(axis=0)
```

By Creating a default-parameter PCA class instance in the variable then Fit_transform() the PCA model to standardised data X_scaled. This stage calculates principle components and converts the data into principal component space. DataFrame X_pca holds altered data. List the principal components in component_names. The names are "PC1", "PC2", etc., for the first, second, and so on principal components.Transform X_pca into a DataFrame with primary component names as column names. And Displaying the first few rows of DataFrame X_pca to view results.

This following code creates a DataFrame X_pca with principle components as columns and data points translated into principal component space as rows. These principle components are orthogonal linear combinations of the original features that capture the most data variance. And we can see the result

```python
from sklearn.decomposition import PCA

# Create principal components
pca = PCA()
X_pca = pca.fit_transform(X_scaled)

# Convert to dataframe
component_names = [f"PC{i+1}" for i in range(X_pca.shape[1])]
X_pca = pd.DataFrame(X_pca, columns=component_names)

X_pca.head()
```

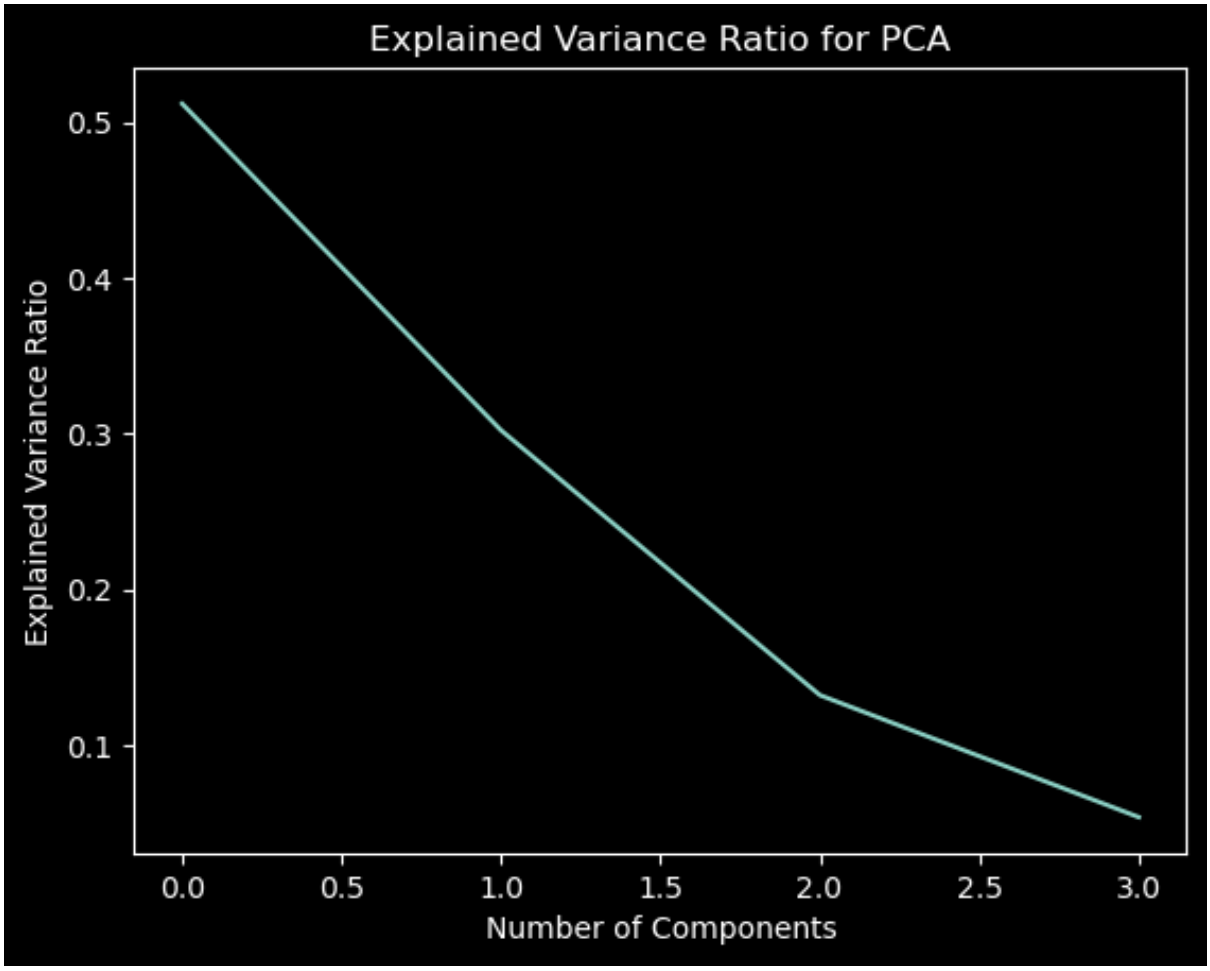
	PC1	PC2	PC3	PC4
0	0.854717	-1.141355	0.107836	0.036177
1	0.854717	-1.141355	0.107836	0.036177
2	-0.399820	-0.480905	0.159850	-1.378656
3	-1.719825	0.037058	0.576143	0.048889
4	1.300528	0.380438	-1.277789	0.300314

The PCA instance's components_ attribute contains loadings after fitting. Unfortunately, PCA terminology varies. The transformed columns in X_pca are called components by convention. We'll dataframe the loadings. Then we can see results as following;

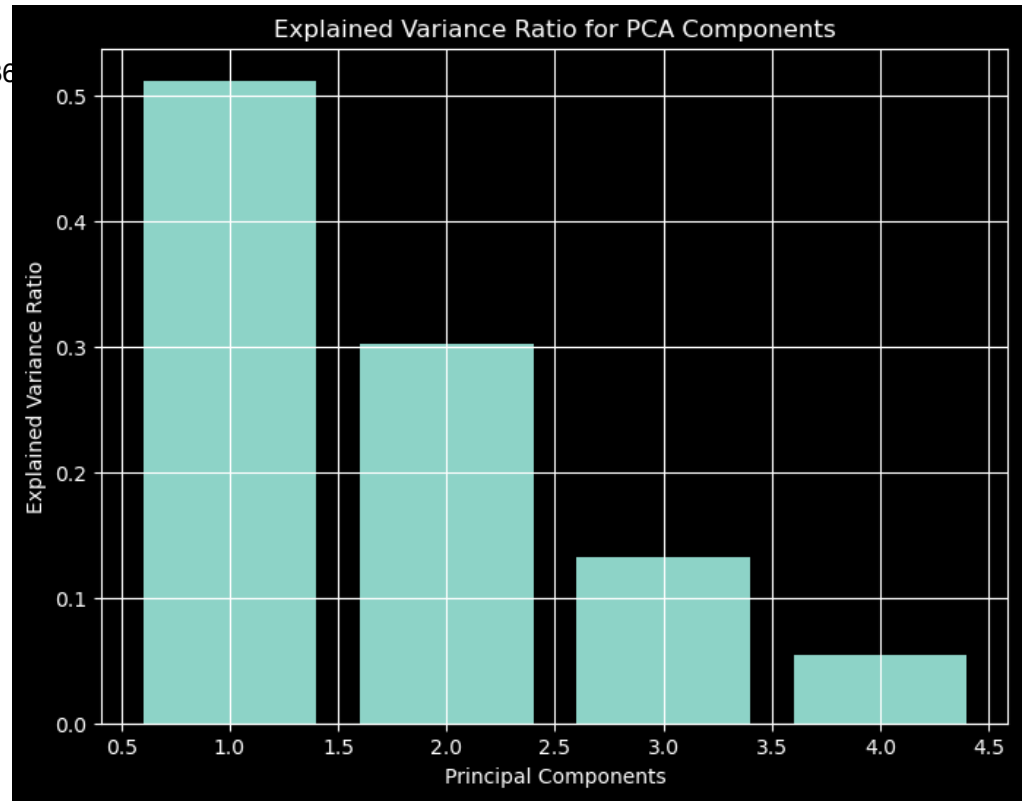
```python
loadings = pd.DataFrame(
    pca.components_.T,
    columns=component_names,
    index=X.columns,
)
loadings
```

	PC1	PC2	PC3	PC4
addedNames	-0.388745	-0.560053	-0.729404	0.056451
addedGuestCount	-0.210056	-0.717035	0.652876	-0.124456
addedCity	-0.642425	0.252084	0.202603	0.694763
searchCTA	-0.626134	0.329627	0.025960	-0.706136

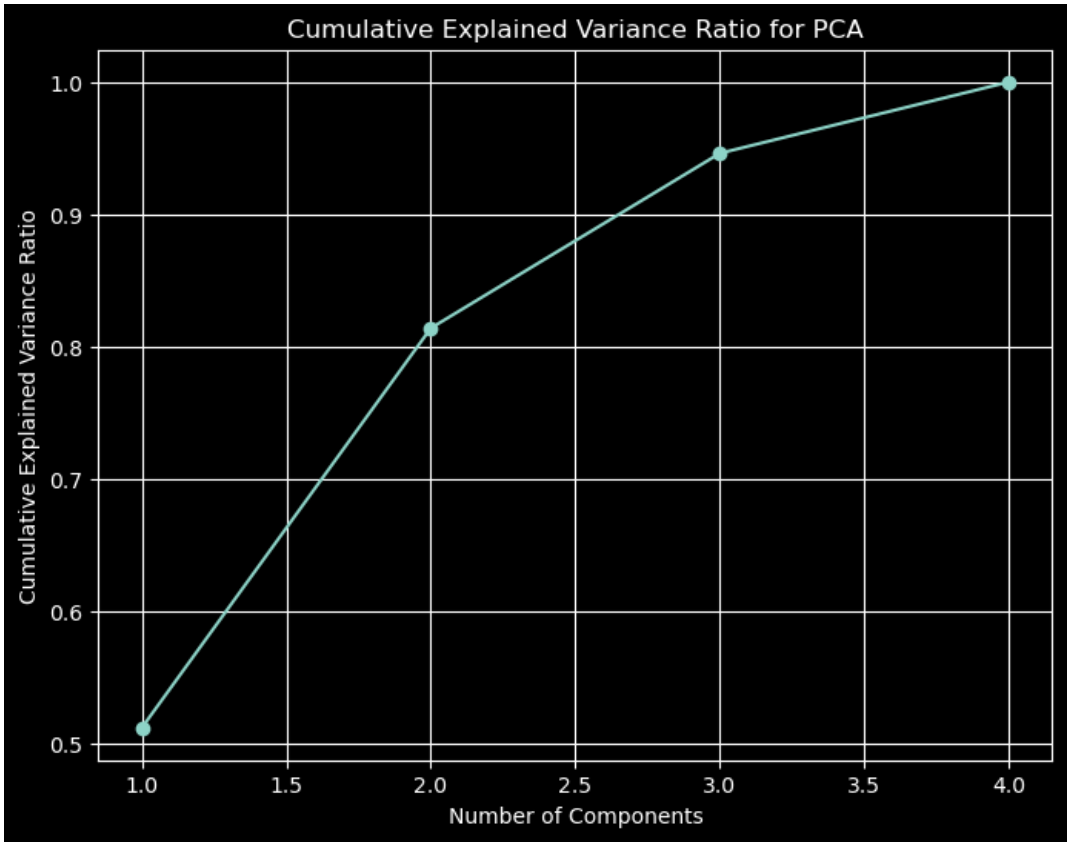
Now I am going to present the results by plotting it, after fitting the data to PCA then Plotting the explained variance ratio.



The previous graph is showing the ratio of the variance and we can deduce that the higher ration is for the first component the graph is shifted one step but it's the same indicationa, and the next following bar graph prove and indicate that



Following with Cumulative Explained Variance Ratio for PCA. Although PC1 is quite informative, the remaining components still have a substantial link with finalizing the operation despite their little fluctuation. It may be beneficial to investigate these subfactors later in order to identify associations missed.



4- Discussion

After performing the A/B tests and taking samples to perform the test and checking the hypothesis, we can deduce that variant 2 which is hid the text around the counts of venues in your location on the CTA, leaving only “We found some perfect venues for you!” We will be more efficient during CTA and for on boarding stage, some of the samples where slightly higher other few samples values had higher significant. Our web marketing initiatives' 0.95 and 0.92 conversion rates are good, These rates reflect that 95% and 92% of total visits to your digital assets resulted in the anticipated user actions, while variation 2 in our case buried the language of variant 2 ""we have several venues for you" is scoring a high rate on CTA which shows better-predetermined objective on “finob”.

For STD Depends on context. Sample 4 of variant 1's 0.4 standard deviation suggests a large data set dispersion and high value variability. Variant 2's sample distribution is consistent with deviation values, but variant 1's is inconsistent.

Std error showed how well a sample represents a population. 0.022–0.017 standard errors demonstrate modest deviation from the mean. Sample data may represent the population since observations are grouped around the mean. Lower standard errors make samples more typical of the population.

Using the Anova by comparing the means of three or more groups to see if there is a significant difference. Interpreting results:

ANOVA Group means and variability are compared by the F-statistic.

F-statistic 13.09. A higher F-statistic indicates a greater difference between group means than group variance.

P-value: the probability of receiving the results (or worse) assuming no genuine difference between group means. It's 0.00010636707864934412. This p-value (less than 0.05) indicates significant group mean variation.

. Thus, one group's mean differs. Rejecting the null hypothesis is supported by the low p-value. In summary, the ANOVA F-test shows a statistically significant difference in group means, and post hoc testing can determine which groups have very different means.

5 - Conclusion and Future Recommendations.

In conclusion, the data derived from the A/B testing, hypothesis checking, and sample testing reveal significant insights. Variant 2 of our web marketing initiative outperformed variant 1 in terms of conversion rates, demonstrating a more effective CTA and onboarding strategy. However, the standard deviation analysis revealed a larger dispersion and variability of data for variant 1. The standard error results demonstrate a modest deviation from the mean, suggesting that the sample data fairly represents the population. The results of the ANOVA F-test further solidify these findings, indicating a statistically significant difference in group means. The low p-value supports the rejection of the null hypothesis, implying a significant variation in group mean. These findings will guide our future marketing initiatives and strategies, to ensure optimized user engagement and conversion rates.

5.1 Future Recommendations

AI personalizes user experience, recommendations and services by analysing user data and behaviours.

AI-based predictive maintenance prevents system faults. This reduces downtime and improves product reliability, by using the some recommendations takes time like venues and getting quotas we can reduce the loading time by gathering and training more data on other ML model and fetch the features output to the current model.

User interface could be enhanced to adequately fit with users with Who are dyslexic also adding an option in planning for

personalized personal assistance for individuals with mobility disabilities

Prediction of wedding rates in England since our users are mostly in England, if we studied the divorce and wedding rates for the upcoming 6-12 months that could helps us arrange our services to our current user and attract more. How? By studying the reasons as description and launch a campaigns to enlightened guide people through articles and videos to reach relationship stability

So we can drift a little bit on our business model with focus on the psychology of relationship but that will widen our market and expose our main goal of gathering people together and spreading with the knowledge to maintain and secure our relations with love

References:

1-BrideBook (2021). *About Us | Bridebook*. [online] bridebook.com. Available at: https://bridebook.com/uk/about-us [Accessed 25 Jul. 2023].

2- https://vintonmessenger.com/author/debbie-adams (2017). 'bridebook' sparks romance. [online] Vinton Messenger. Available at: https://vintonmessenger.com/bridebook-sparks-romance/ [Accessed 25 Jul. 2023].

3-Sangha, S. (2009). BUSINESS PLAN FOR BRIDEBOOK.COM. [online] https://summit.sfu.ca. Available at: https://summit.sfu.ca/_flysystem/fedora/sfu_migrate/659/MBA%202009%20Sangha%2C%20S..pdf [Accessed 24 Jul. 2023]. PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF BUSINESS ADMINISTRATION.

4- DataStar, Inc. (2013). How to Interpret Standard Deviation and Standard Error in Survey Research | Article | GreenBook.org. [online] Greenbook.org. Available at: https://www.greenbook.org/marketing-research/how-to-interpret-standard-deviation-and-standard-error-in-survey-research-03377.

5- Rumsey, D. (2021). How to Interpret Standard Deviation in a Statistical Data Set. [online] dummies. Available at: https://www.dummies.com/article/academics-the-arts/math/statistics/how-to-interpret-standard-deviation-in-a-statistical-data-set-169772/.

6- Mailchimp. (n.d.). Conversion Rate: What is Conversion Rate? - Mailchimp. [online] Available at: https://mailchimp.com/marketing-glossary/conversion-rates/.

7- www.itl.nist.gov. (2022). *1.4.3. References For Chapter 1: Exploratory Data Analysis*. [online] Available at: https://www.itl.nist.gov/div898/handbook/eda/section4/eda43.htm [Accessed 25 Jul. 2023].

8- Sarker, I.H. (2021). Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective. *SN Computer Science*, [online] 2(5). doi:https://doi.org/10.1007/s42979-021-00765-8.

9- Takyar, A. (2023). *Exploratory Data Analysis - A Comprehensive Guide to EDA*. [online] LeewayHertz - AI Development Company. Available at: https://www.leewayhertz.com/what-is-exploratory-data-analysis/ [Accessed 25 Jul. 2023].

10- Stobierski, T. (2019). *The advantages of data-driven decision-making | HBS online*. [online] Harvard Business School Online. Available at: https://online.hbs.edu/blog/post/data-driven-decision-making.

.

7- AppendixAll the scripts is available on Bridebook task on Github; you can access it with this link