Section 3 - Experimental Design

Questions:

1. The original version of an e-commerce website is quite basic, featuring only text on the site. They plan to add some product images, and intend to run an A/B test on their website to see if it helps sales. You are the analyst for this test.

a. What would your hypothesis be for running this test?

To check the significant difference if change in site by adding some product images brings more conversions than with only text on the site. Randomly split visitors into two equally sized groups; variant and the control group.

H0: There is no statistically significant difference between control and variant groups with respect to the average of number of purchases.

H1: There is statistically significant differences between control and variant groups with respect to the average of number of purchases.

b. What would be your primary metric to measure the success and why?

The most important metric for measure the success is the number of purchases. More purchases helps increase in the revenue then **profits** are also likely to increase and help in increase its margin of safety by selling more products and help in more visitors.

c. Tell us what secondary metrics you might look at to help you make a call on if a test is performing correctly and why?

Secondary metric: Revenue to look at; to check test is performing correctly because small and unaffordable customers have lot of purchases that generate insignificant revenue, and focus on highest value providers.

Revenue = Active visitor count * Order count * Average revenue per order
Other metrics**:**

* Amount spend / Number of purchases
* Amount spend / Number of visitors
* Number of Website Clicks / Number of Impressions; more engagement and active users on the website

d. You are asked to filter down the results further to see if the primary metric is performing well on different breakdowns. What potential problems could occur with your analysis if you layer different filters on top of the other?

Filtering down further below top-level KPIs by using multi-dimensional segmentation; analyzing performance of each group exactly who is reacting well to the new changes and not reacting well. It could have problems like segments can end up having a very small size. Thus, insights drawn out by comparing different segments of the variations may not hold statistical significance and comparing too many segments, the greater the likelihood of getting a false positive.

e. Below are the results for an A/B test that you have analysed. Based off the results displayed below, what would be your recommendation to your product manager? Why did you make this recommendation

Cohort	Visitors	Converters	Traffic split
Variant 1	8000	3000	10%
Control	72000	25000	90%

Cohort	Converters/ Visitors
Variant 1	37.5%
Control	34.72%
Overall	2.78%

~3% improvement has been observed in visitors to convertors in variant 1
Variant 1 group has better rate; if this difference is statistically significant.

H0: There is no statistically significant difference between two cohort rates
H1: There is statistically significant differences between two cohort rates

```
count = np.array([25000, 3000])
nobs = np.array([72000, 8000])
stat, pval = proportions_ztest(count, nobs)
print(pval)
```

```
7.745937202529398e-07
```

p-value found smaller than 0.05; reject the null hypothesis. There is a statistically significant difference between conversion rates. Variant 1 is better.

f. In what situations would you choose not to run an experiment, favouring rolling out a new feature immediately? When this occurs, how should you aim to handle this as an analyst

When not to choose run an experiment if there's low risk to rolling out a new feature right away, if there is no informed hypothesis: no success metric defined or if there is no meaningful traffic.
Handle this by analysing time period based before and after change in the metrics.