A/B Testing   Draft saved

File   Edit   View   Run   Add-ons   Help

+ | ▷   ▷▷   Run All   •••   Save Version

[1]:
```python
import numpy as np
import pandas as pd
import scipy.stats as stats
import statsmodels.stats.api as sms
import seaborn as sns
from math import ceil

# Load dataset
df = pd.read_csv('/kaggle/input/ab-testing/ab_data.csv')
df.head()
```

[1]:

	user_id	timestamp	group	landing_page	converted
0	851104	2017-01-21 22:11:48.556739	control	old_page	0
1	804228	2017-01-12 08:01:45.159739	control	old_page	0
2	661590	2017-01-11 16:55:06.154213	treatment	new_page	0
3	853541	2017-01-08 18:28:03.143765	treatment	new_page	0
4	864975	2017-01-21 01:52:26.210827	control	old_page	1

[2]:
```python
# Users with multiple session counts
session_counts = df['user_id'].value_counts(ascending = Fa
multi_users = session_counts[session_counts > 1].count()

# dropping with multiple session counts
users_to_drop = session_counts[session_counts > 1].index

df = df[~ df['user_id'].isin(users_to_drop)]
```

[3]:
```python
# Sample control and treatment group
control_sample = df[df['group'] == 'control'].sample(n = 5
treatment_sample = df[df['group'] == 'treatment'].sample(n
ab_test = pd.concat([control_sample, treatment_sample], ax
ab_test.reset_index(drop=True, inplace=True)
# Define functions for standard deviation and standard erro
std_dev = lambda x : np.std(x, ddof = 0)
```

>_   ● Draft Session (0m)