

## HW3 A/B testi

1. Dosyadan verileri okuyun ve A/B testinin sonuçları hakkında bilgileri çıkarın:

```
[11]: df[(df['test_group']=='a')].count()
```

```
[11]: user_id      10013
      timestamp    10013
      test_group    10013
      conversion    10013
      dtype: int64
```

```
[12]: df[(df['test_group']=='b')].count()
```

```
[12]: user_id      9985
      timestamp    9985
      test_group    9985
      conversion    9985
      dtype: int64
```

A ve B gruplarındaki kullanıcı sayısı;

A= 10 013 B= 9985

```
: df[(df['test_group']=='a') & (df['conversion']== 0)].count()
```

```
: user_id      9402
   timestamp    9402
   test_group    9402
   conversion    9402
   dtype: int64
```

```
: df[(df['test_group']=='a') & (df['conversion']== 1)].count()
```

```
: user_id      611
   timestamp    611
   test_group    611
   conversion    611
   dtype: int64
```

---

```
[35]: df[(df['test_group']=='b') & (df['conversion']== 0)].count()
```

```
[35]: user_id      9096
      timestamp    9096
      test_group   9096
      conversion   9096
      dtype: int64
```

```
[36]: df[(df['test_group']=='b') & (df['conversion']== 1)].count()
```

```
[36]: user_id      889
      timestamp    889
      test_group   889
      conversion   889
      dtype: int64
```

A ve B gruplarındaki dönüşüm sayısı;

A= 611 B=889

A ve B gruplarındaki dönüşüm oranı;

A= 611/10013                      B= 889/9985=  
=0.061                              B= 0.089  
= %6.1                              = 8.9

Testin başlangıç tarihi, bitiş tarihi ve testin süresi (gün olarak).

```
] : df["timestamp"] = pd.to_datetime(df["timestamp"])

dates = df.groupby("test_group").agg(start = ("timestamp", "min"), finish= ("timestamp", "max"))

dates["total_days"] = (dates["finish"] - dates["start"]).dt.days

print(dates)
```

	start	finish	total_days
test_group			
a	2023-07-03 01:46:15	2023-07-25 01:41:19	21
b	2023-07-03 01:42:34	2023-07-25 01:35:59	21

```
] :
```



## 2) Student t test -

```
[59]: import pandas as pd

# Test verisini oluşturun
test_data = pd.DataFrame({
    'test_group': ['a'] * 10013 + ['b'] * 9985,
    'conversion': [1] * 9402 + [0] * (10013 - 9402) +
                  [1] * 9096 + [0] * (9985 - 9096)
})

# Her şeyin doğru bir şekilde oluşturulup oluşturulmadığını kontrol edelim:
test_data.groupby('test_group').describe()
```

```
[59]:
```

								conversion
	count	mean	std	min	25%	50%	75%	max
test_group								
a	10013.0	0.938979	0.239380	0.0	1.0	1.0	1.0	1.0
b	9985.0	0.910966	0.284806	0.0	1.0	1.0	1.0	1.0

```
[69]: from scipy import stats

alpha = 0.05

statistic, pvalue = stats.ttest_ind(test_data[test_data['test_group'] == 'a']['conversion'],
                                   test_data[test_data['test_group'] == 'b']['conversion'],
                                   alternative='less')

print(f't-statistic: {round(statistic, 2)}, p-value: {round(pvalue, 2)}')

if pvalue < alpha:
    print('The difference is statistically significant, Null Hypothesis is rejected.')
else:
    print('The difference is insignificant, Null Hypothesis cannot be rejected.')

t-statistic: 7.53, p-value: 1.0
The difference is insignificant, Null Hypothesis cannot be rejected.
```

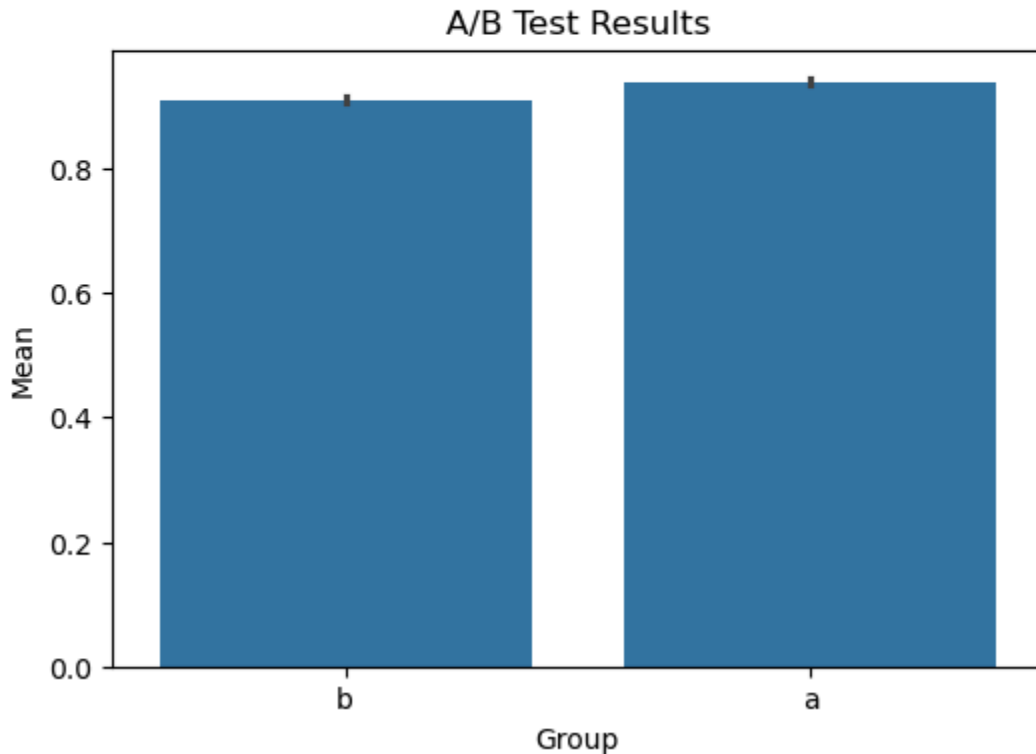
Null hipotezini kabul ediyoruz (p-value= 1.0)

3)

```
errorbar=('ci', 95)) # Confidence Intervals

plt.title('A/B Test Results')
plt.xlabel('Group')
plt.ylabel('Mean')

plt.show()
```



4) **Null hipotez (H0)** : A ve B grubun ortalamaları birbirinden farklı değildir

**Alternatif hipotez (H1)** : A ve B grubunun ortalamaları arasında önemli bir fark vardır.

Bunu test etmek için Student t testi kullanıldı. Yaptığımız analizde p-value= 1.0 olarak bulundu p değerimiz alpha= 0.05 değerinden büyük olduğundan önemli bir fark görülmedi.

Zamana göre baktığımızda conversion rate'de A'da başlangıçta azalma görülürken B' de ise başlangıçta hızlı bir yükselme görülmektedir. Bu da gösteriyor ki yardımcı metrik olarak, süreyi 21 günden daha az bir süreye indirirsek A ve B grubu arasında istatistiksel olarak anlamlı bir fark görülebilir.

Bonus Görev:

