** A/B test**

```python
import pandas as pd
from scipy import stats
```

```python
taxi_data = pd.read_csv("2017_Yellow_Taxi_Trip_Data.csv", index_col = 0)
```

```python
# descriptive stats code for EDA
taxi_data.describe(include='all')
```

	VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	RatecodeID	store_and_fwd_flag	PULocationID	DOLocationID	payment_type	fare_amount	extra	mta_tax	tip_amount	tolls_amount	improvement_surcharge	total_amo
count	22699.000000	22699	22699	22699.000000	22699.000000	22699.000000	22699	22699.000000	22699.000000	22699.000000	22699.000000	22699.000000	22699.000000	22699.000000	22699.000000	22699.000000	22699.000
unique	NaN	22687	22688	NaN	NaN	NaN	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
top	NaN	07/03/2017 3:45:19 PM	10/18/2017 8:07:45 PM	NaN	NaN	NaN	N	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
freq	NaN	2	2	NaN	NaN	NaN	22600	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
mean	1.556236	NaN	NaN	1.642319	2.913313	1.043394	NaN	162.412353	161.527997	1.336887	13.026629	0.333275	0.497445	1.835781	0.312542	0.299551	16.310
std	0.496838	NaN	NaN	1.285231	3.653171	0.708391	NaN	66.633373	70.139691	0.496211	13.243791	0.463097	0.039465	2.800626	1.399212	0.015673	16.097
min	1.000000	NaN	NaN	0.000000	0.000000	1.000000	NaN	1.000000	1.000000	1.000000	-120.000000	-1.000000	-0.500000	0.000000	0.000000	-0.300000	-120.300
25%	1.000000	NaN	NaN	1.000000	0.990000	1.000000	NaN	114.000000	112.000000	1.000000	6.500000	0.000000	0.500000	0.000000	0.000000	0.300000	8.750
50%	2.000000	NaN	NaN	1.000000	1.610000	1.000000	NaN	162.000000	162.000000	1.000000	9.500000	0.000000	0.500000	1.350000	0.000000	0.300000	11.800
75%	2.000000	NaN	NaN	2.000000	3.060000	1.000000	NaN	233.000000	233.000000	2.000000	14.500000	0.500000	0.500000	2.450000	0.000000	0.300000	17.800
max	2.000000	NaN	NaN	6.000000	33.960000	99.000000	NaN	265.000000	265.000000	4.000000	999.990000	4.500000	0.500000	200.000000	19.100000	0.300000	1200.290

```python
taxi_data.groupby('payment_type')['total_amount'].mean()
```

```
payment_type
1    17.663577
2    13.545821
3    13.579669
4    11.238261
Name: total_amount, dtype: float64
```

```python
#hypothesis test, A/B test
#significance level

credit_card = taxi_data[taxi_data['payment_type'] == 1]['total_amount']
```

```python
cash = taxi_data[taxi_data['payment_type'] == 2]['total_amount']
stats.ttest_ind(a=credit_card, b=cash, equal_var=False)
```

Out[5]: Ttest_indResult(statistic=20.34644022783838, pvalue=4.5301445359736376e-91)