Experimental Design

by Krishna Devabhaktuni¶

https://www.kaggle.com/henriqueyamahata/bank-marketing

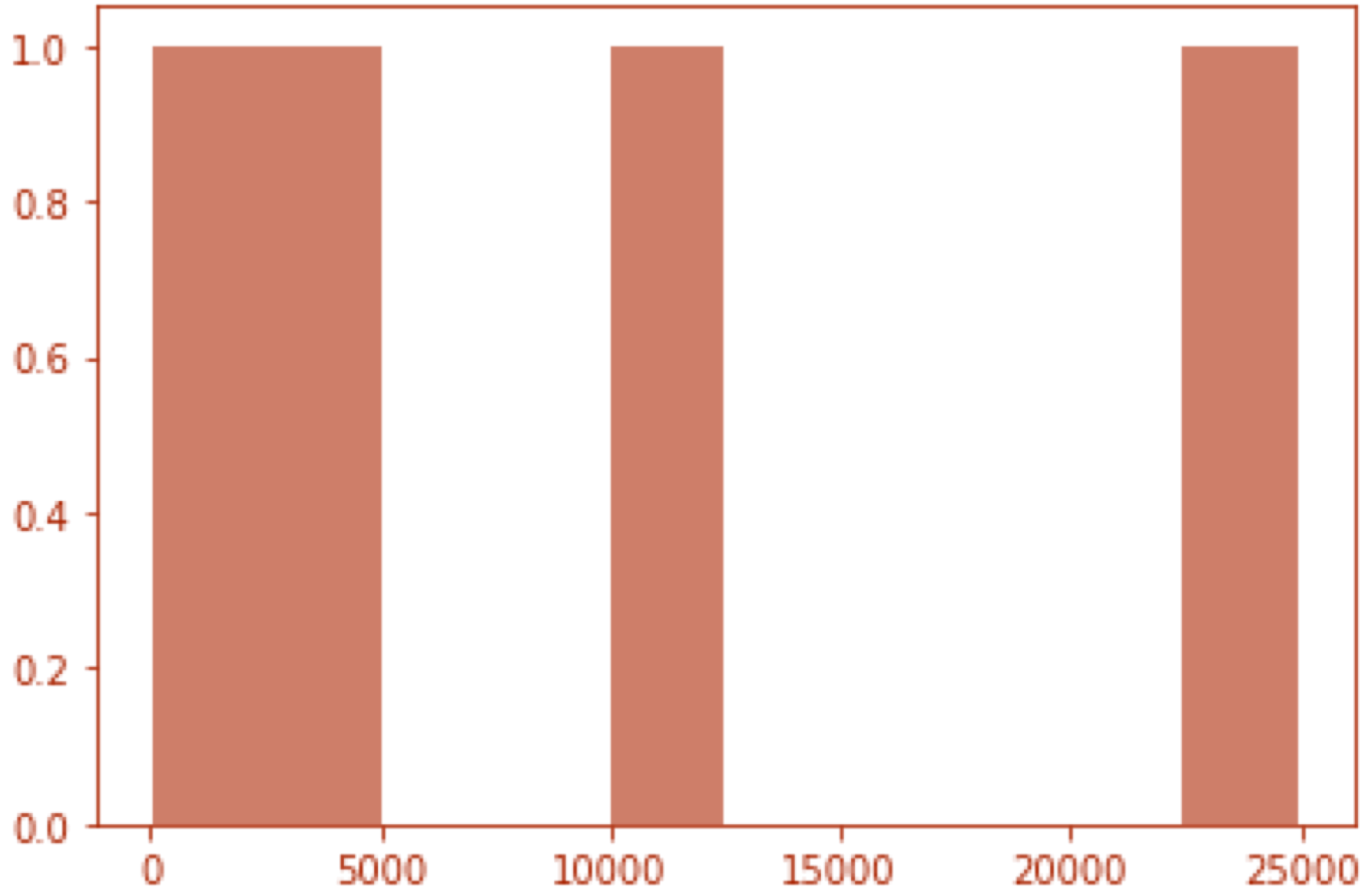
Data Set Information: The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls.

We will be performing an A|B TEST and looking into how married and single people are subscribing based on the phone calls. By the end of this presentation, we are going to look at aspects like subscription rates of married and single people, whose subscription rates are better, what can we do to improve the subscription rates. We will be looking at mean subscription rates under various conditions, visualizing the subscription rates, we will also check if they are significantly different from one another

Data Dictionary

Age (numeric)
Job : type of job (categorical: 'admin.', 'blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'self-employed', 'services', 'student', 'technician', 'unemployed', 'unknown')
Marital : marital status (categorical: 'divorced', 'married', 'single', 'unknown' ; note: 'divorced' means divorced or widowed)
Education (categorical: 'basic.4y', 'basic.6y', 'basic.9y', 'high.school', 'illiterate', 'professional.course', 'university.degree', 'unknown')
Default: has credit in default? (categorical: 'no', 'yes', 'unknown')
Housing: has housing loan? (categorical: 'no', 'yes', 'unknown')
Loan: has personal loan? (categorical: 'no', 'yes', 'unknown')

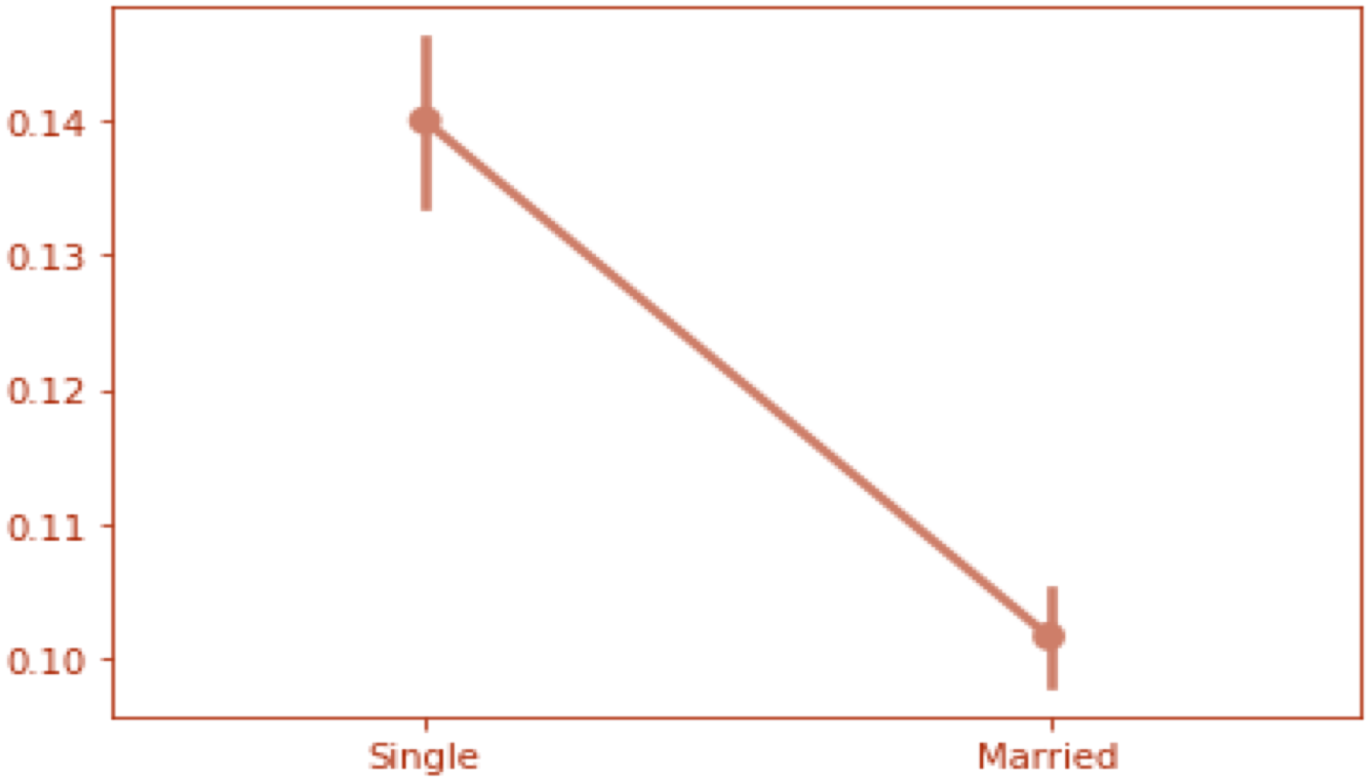
Histograms of Categories in Marital Variable(Married, Divorced, Single)



A difference in sample size is not an issue for an independent samples t-test. Each sample is large enough and should give reliable estimates, given the data meets the rest of our assumptions

- The Average Subscription rate of Married people is 0.10
- The Average Subscription rate of Single people is 0.14
- We can see that the subscription rate is higher in Single's, Let's perform a t-test and make sure if there is a significant difference in our groups

Test indResult(statistic=-10.785981914415991, pvalue=4.4050099238429465e-27)



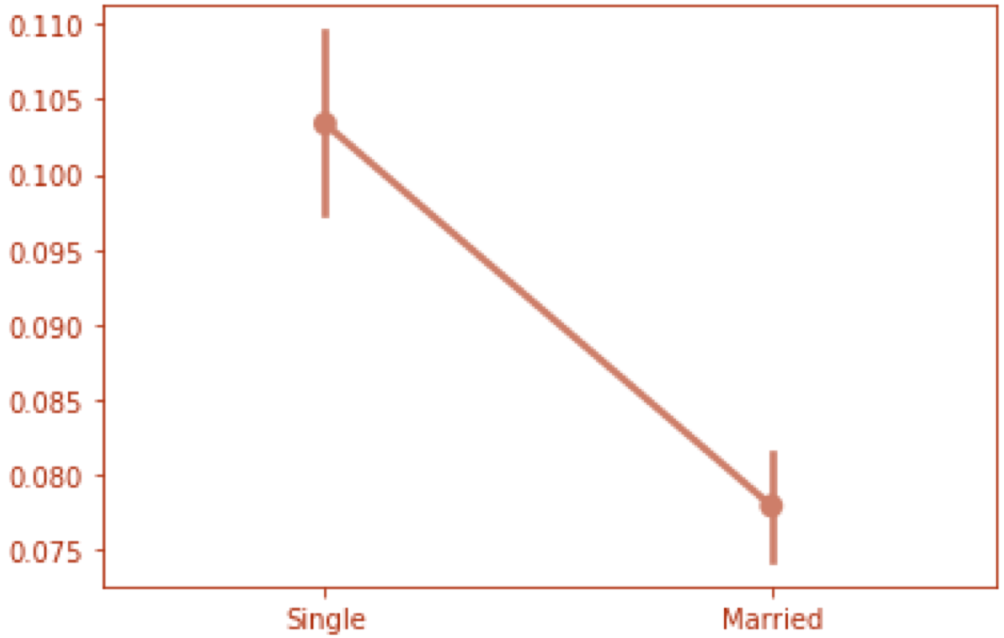
There is a significant difference between our groups


BIAS!
BIAS EVERYWHERE!
LIGHTYEAR
imgflip.com

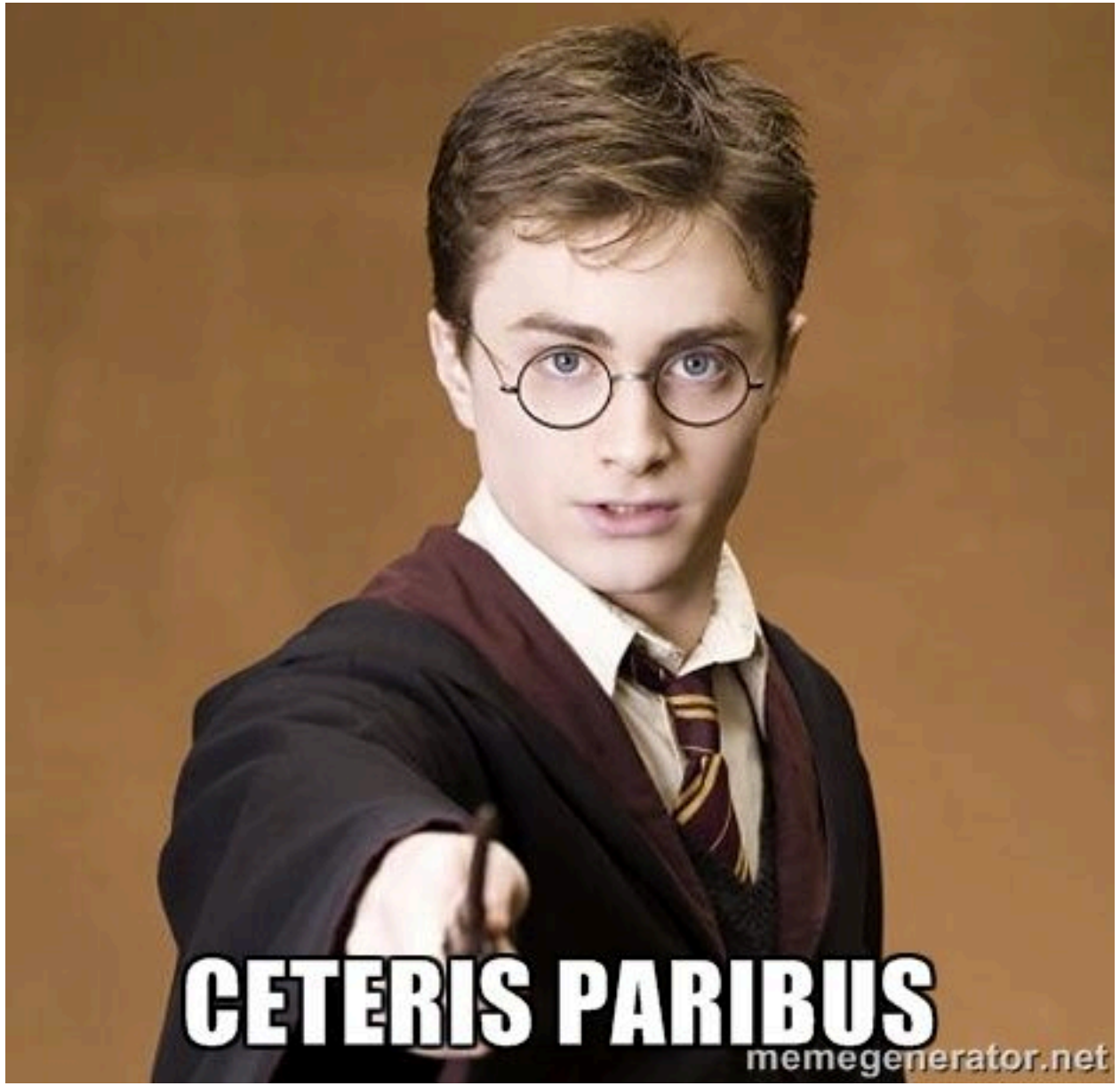
The Overall Subscription rate is definitely higher, however, we need to think about seasonality. May be we were not able to contact Married people during Winter months due to their busy schedules. We will look into subscription rates during summer months only

The Average Subscription rate of Married people in Summer is 0.08
The Average Subscription rate of Single people in Summer is 0.10

Ttest_indResult(statistic=7.165024230811873, pvalue=7.963954305898505e-13)



Although the subscription rate is higher in singles, we cannot see a significant difference between our groups

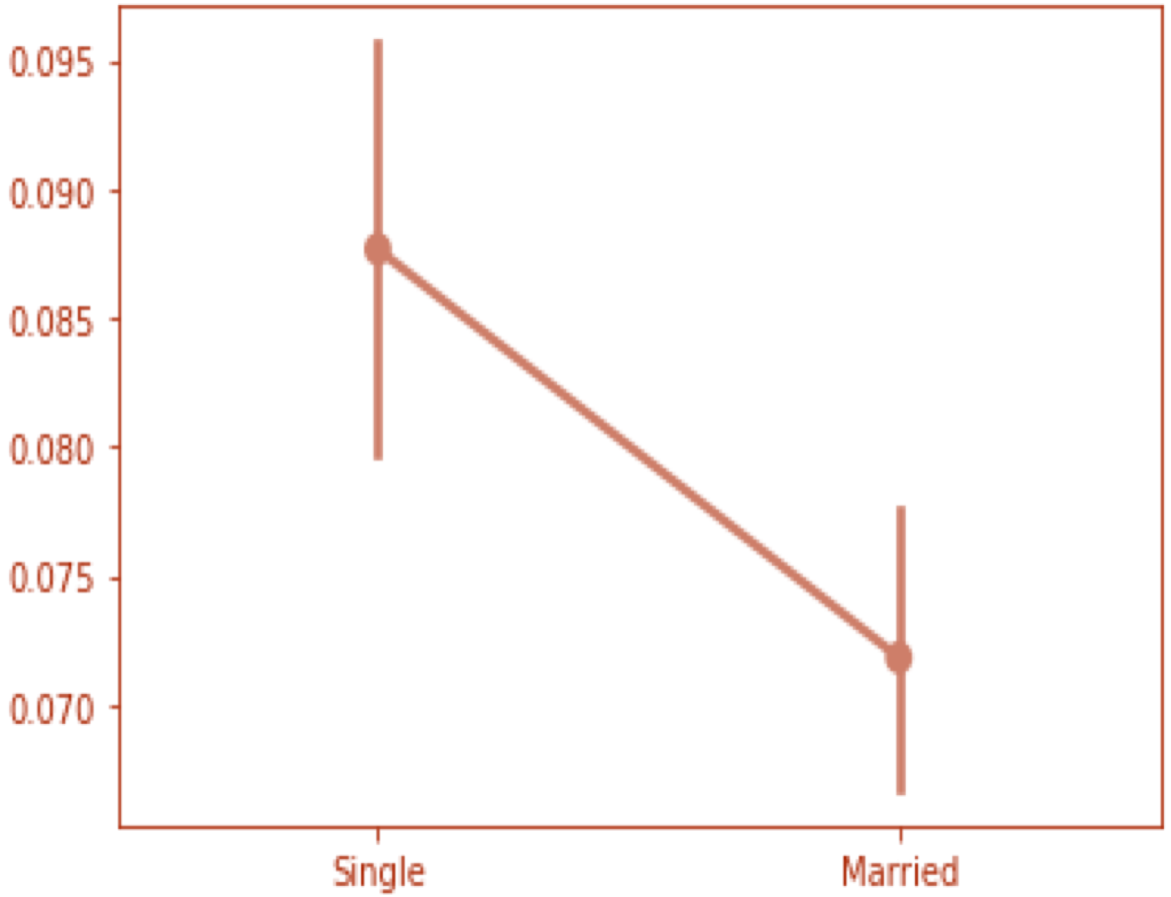

CETERIS PARIBUS
memegenerator.net

We need to hold everything equal in order to compare 2 groups. We reduced seasonality by comparing summer months only, however, we need to compare similar age groups as well. In order to do that we compare ages 30 to 40 between the two groups

- The Average Subscription rate of **Married people** in Summer with age **30-40 is 0.07**
- The Average Subscription rate of **Single people** in Summer with age **30-40 is 0.09**
- Test_indResult(statistic=3.234645895027084, pvalue=0.0012209880773163208)

We can now see the significant difference between the groups Married and Single.

Singles have higher subscription rate than Married people



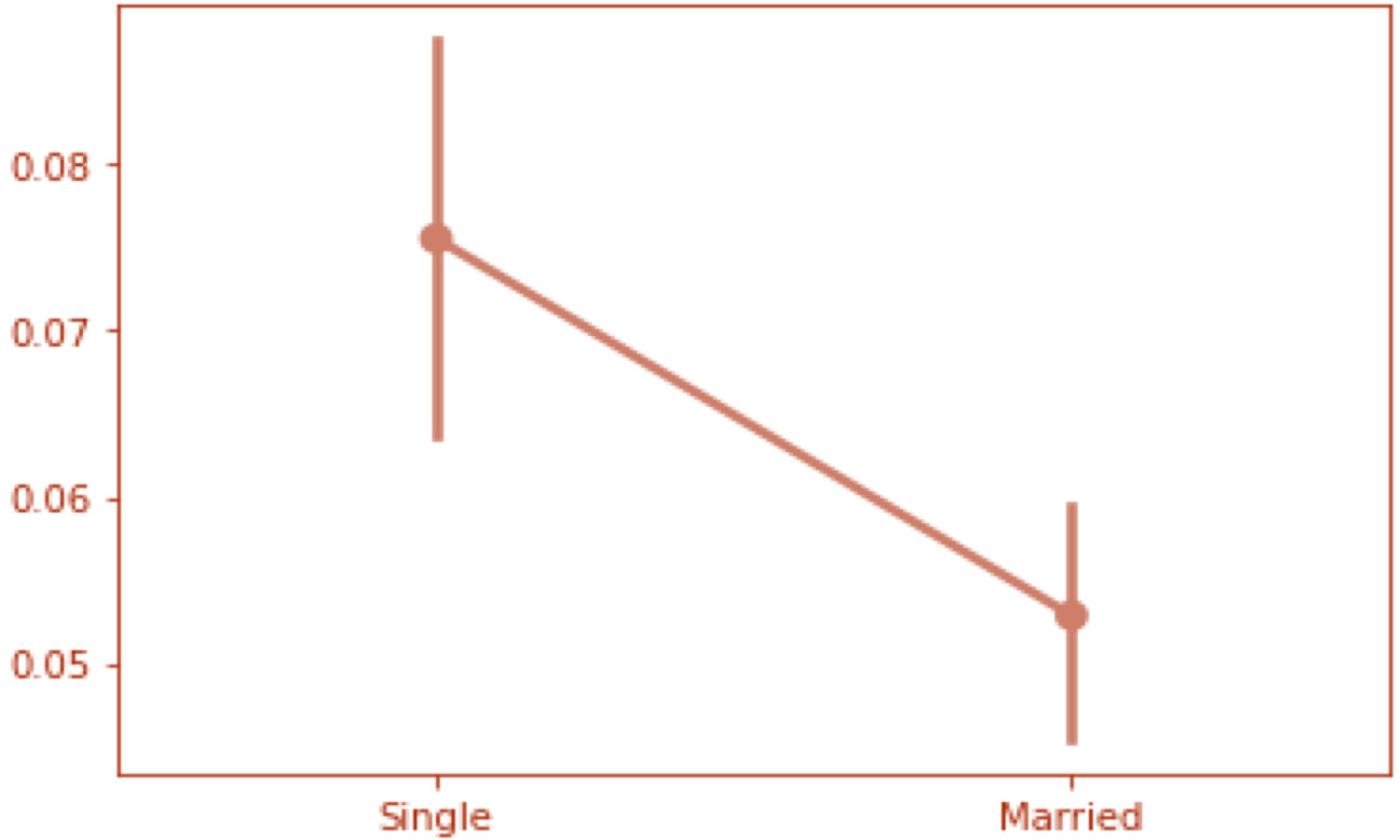
Comparing our Groups in November

- The Average Subscription rate of **Married people in November is 0.07**
- The Average Subscription rate of **Single people in November is 0.10**
- Ttest_indResult(statistic=1.8205867426070397, pvalue=0.06883375951315605)

Singles have high subscription rate in November as well, however, the p-value is a little higher than 0.05. Now lets compare the month of May

The Average Subscription rate of **Married people in May is 0.05**
The Average Subscription rate of **Single people in May is 0.08**

Ttest_indResult(statistic=3.4311223741725865, pvalue=0.0006052369855814975)



We have compared 2 major categories in our dataset with a response variable in various settings and compared the subscription rates and we can clearly see that the subscription rate of singles category is higher than subscription rate of the married people. In most of the conditions, the subscription rate is significantly different from other group. We have also visualized the differences using a point plot.

We can suggest the Portugese Banking institution that our results conclude higher subscription rate in singles than Married people. In order to increase the subscription rate of the Married people we need to come up with different plans that are tailored for Married people. We may conduct some surveys with Married customers in order to create better plans that suits Married Customers. Also we need to find other ways to contact Married customers such as "Texting" and "Emailing" as Married customers may have limited time due to their busy schedules. We also suggest the bank customer service to be open during the weekends so that the Married customers can be contacted and it is also convenient for customers to resolve their queries during the weekend