# **Type I**

A test is administered annually. The test has a **mean score of 100** and a **standard deviation of 40**.

If Rohit's z-score is 1.5, what was his score on the test?

From the z-score equation, we know z = (X - μ) / σ where z is the z-score, X is the value of the element, is the mean of the population, and is μ σ the standard deviation.

Solving for Jane's test score (X), we get X = (z \* σ) + μ = (1.5 \* 40) + 100 = 60 + 100 = **160**

# Type III

Build a Logistic Regression model with the following parameters to predict whether a customer will purchase the product or not.

Parameters:

* random\_state for train and test split= 2012,
* test\_size = 0.30.

Leave all other parameters to default values.

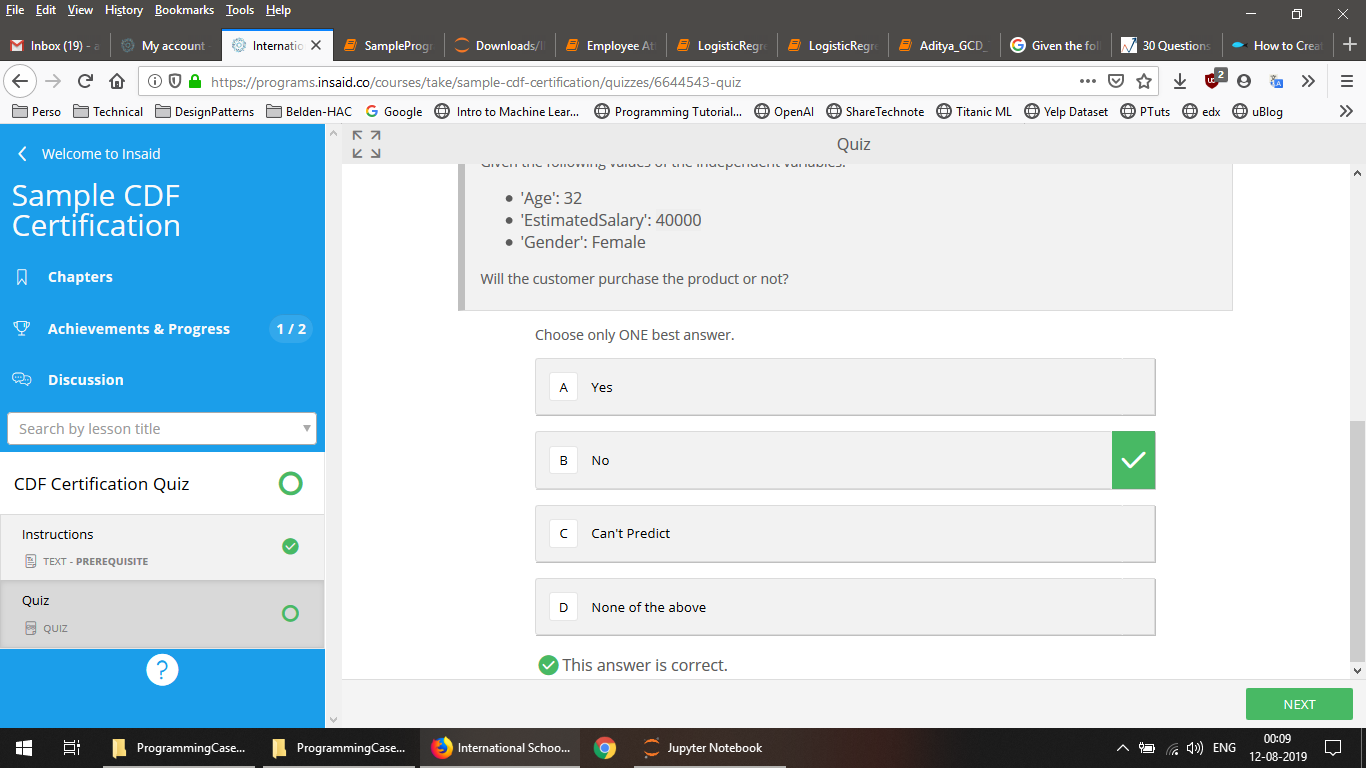
Dataset: [Purchase-Report](https://github.com/insaid2018/Sample-CDF-Certification/tree/master/Purchase-Report)

Kindly go through data description available in the link.

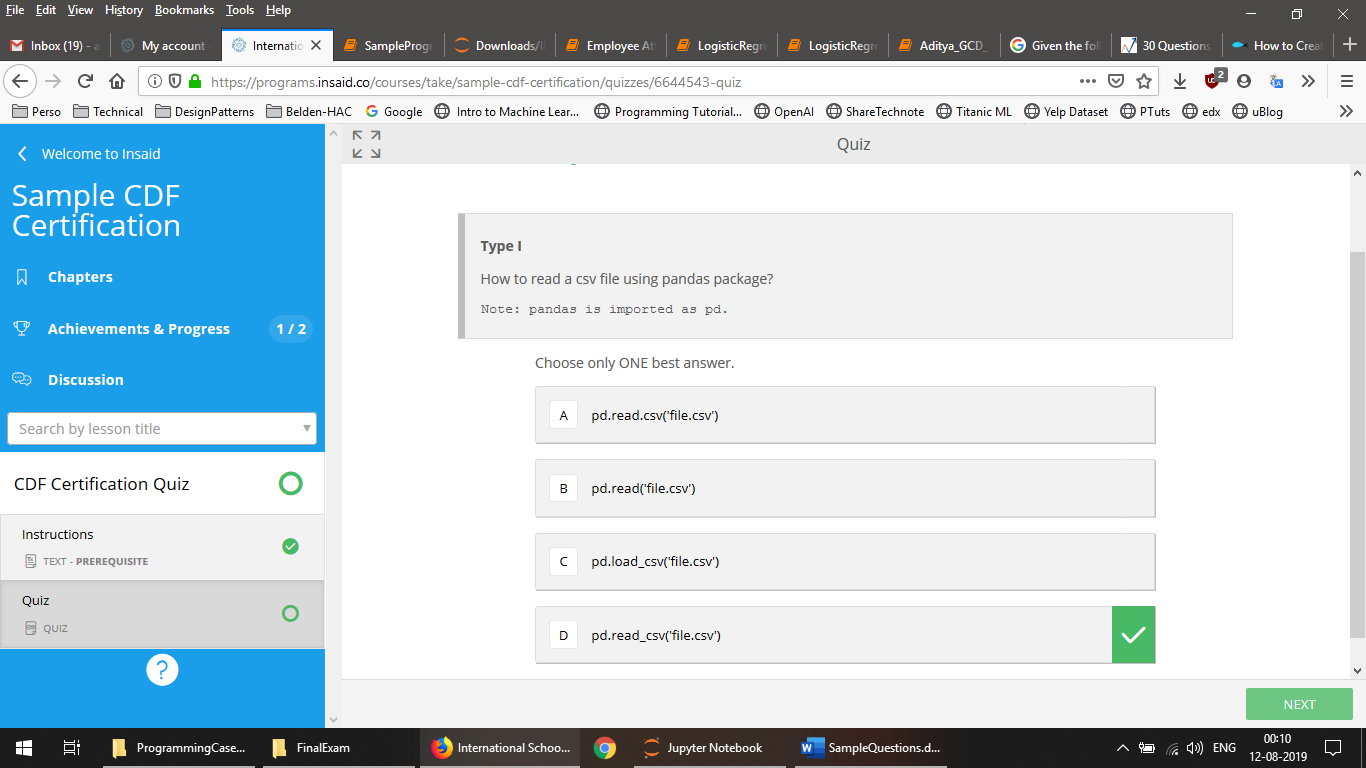
Given the following values of the independent variables:

* 'Age': 32
* 'EstimatedSalary': 40000
* 'Gender': Female

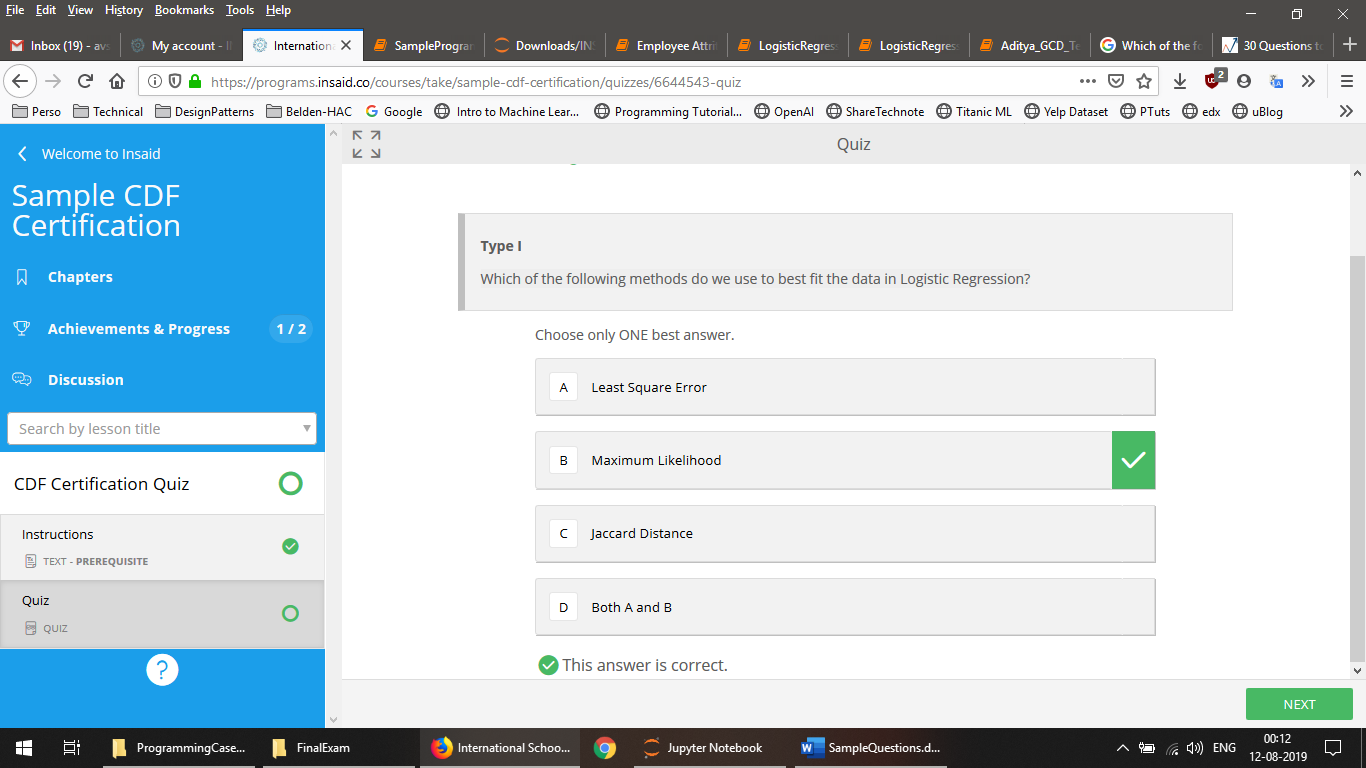
Will the customer purchase the product or not?



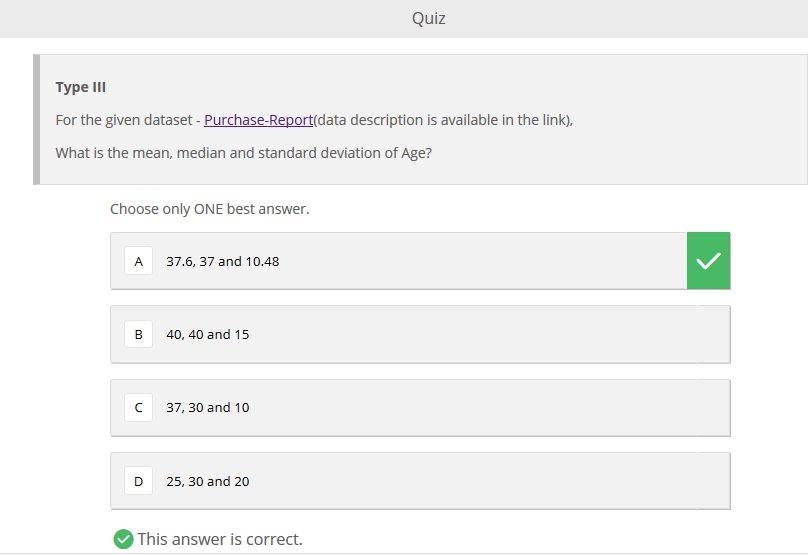
# **Type I**



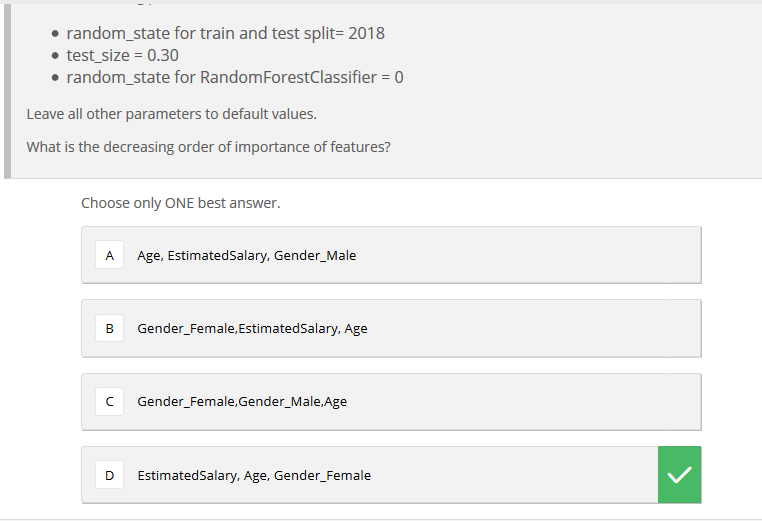
# **Type I**



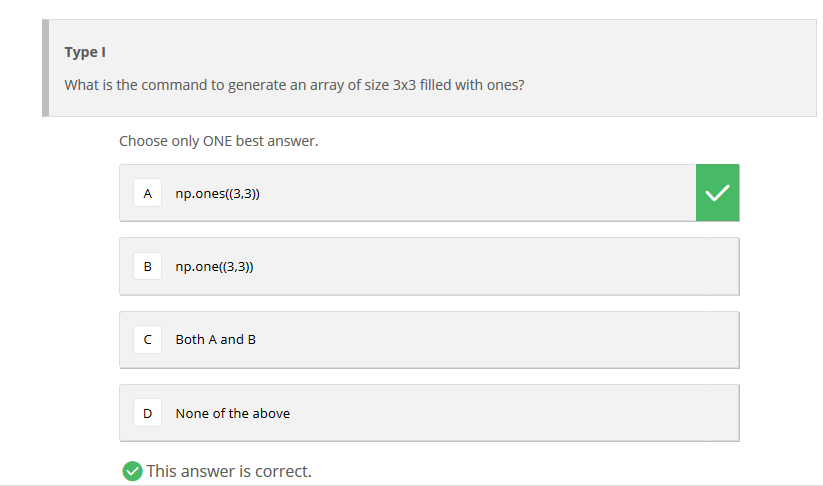
# Type III



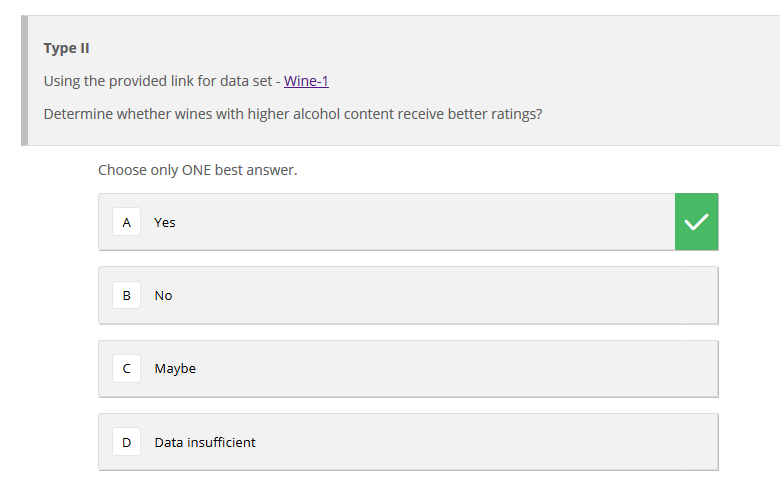
# Type III



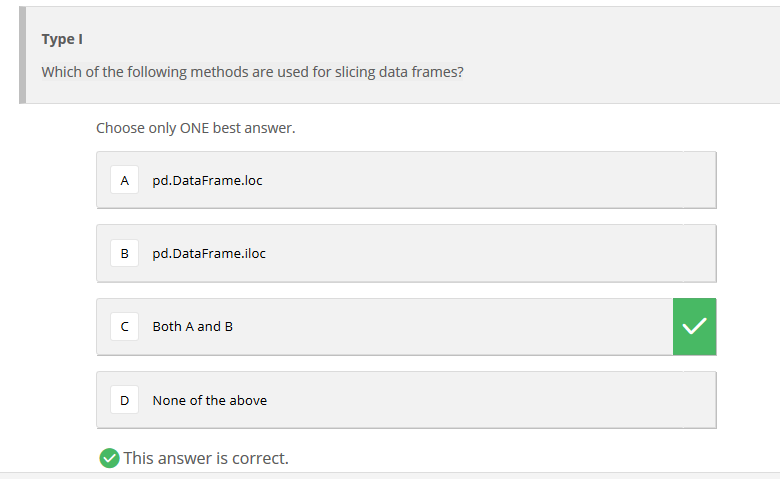
# Type I



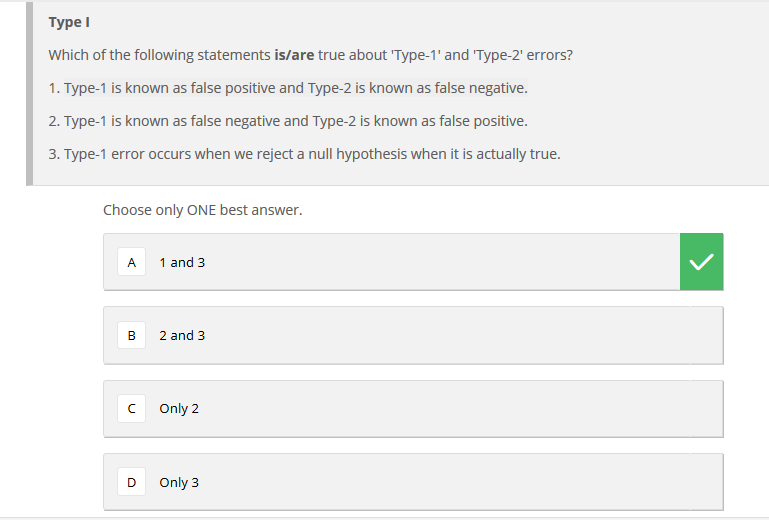
# Type II



# Type I



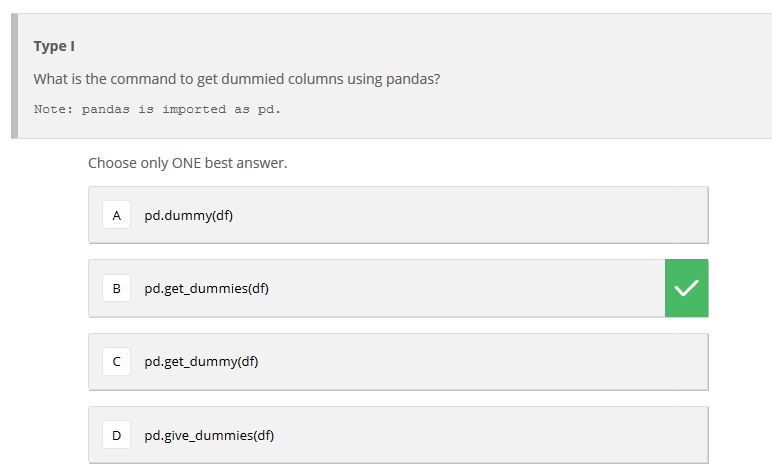
# Type I



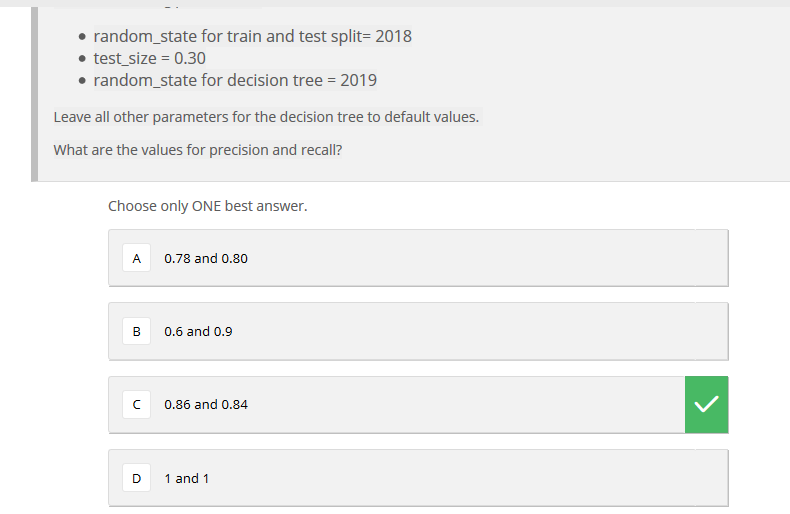
# Type II

# 

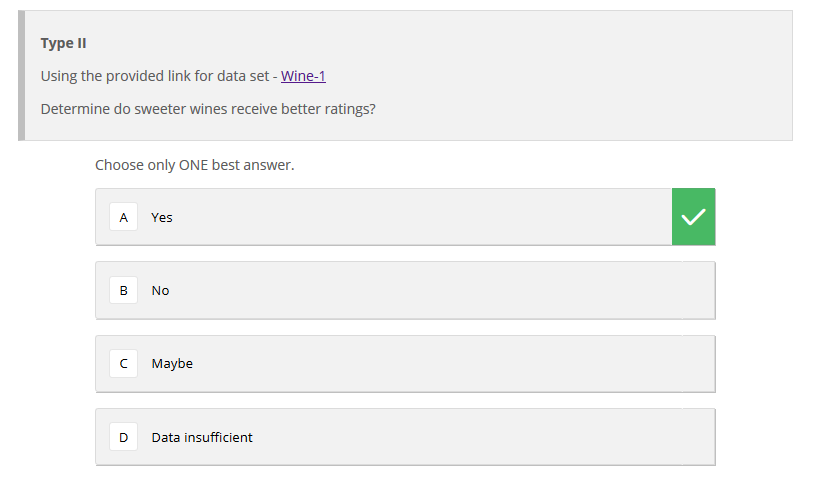
# Type I



# Type III



# Type II



# Type I

Suppose, you are working with **categorical features** and **NOT** looked at distribution of the categorical variable in the test data.You want to apply One hot encoding on the categorical variables.

What challenges you may face if applied One Hot Encoding on a categorical feature of train data set?

A) All categories of categorical variable are not present in the test dataset.

B) Frequency distribution of categories is different in train as compared to the test dataset.

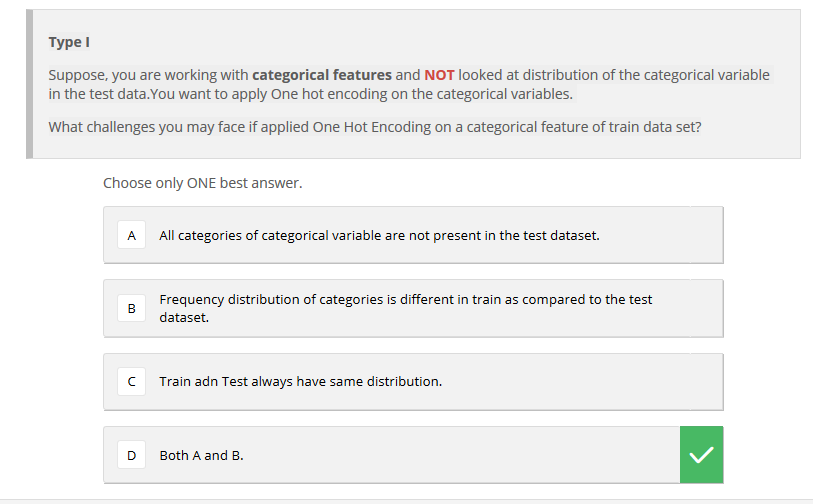
C) Train and Test always have same distribution.

D) Both A and B

E) None of these

**Solution: (D)**

Both are true, The OHE will fail to encode the categories which is present in test but not in train so it could be one of the main challenges while applying OHE. The challenge given in option B is also true you need to more careful while applying OHE if frequency distribution doesn’t same in train and test.



# References

<https://www.analyticsvidhya.com/blog/2017/08/skilltest-logistic-regression/>

<https://tomreads.com/2018/02/21/data-analysis-process-case-study-1-udacity/>

<https://www.analyticsvidhya.com/blog/2017/04/40-questions-test-data-scientist-machine-learning-solution-skillpower-machine-learning-datafest-2017/>

<https://towardsdatascience.com/explaining-feature-importance-by-example-of-a-random-forest-d9166011959e>