

**PREDICTION OF EXACT NICHE USING BANK DATA**



Team Members:

Maria john antony

Pragadesh

Shivang Sunil Singh

**CONTENTS**

1. Title of the project
   1. Introduction
   2. Objectives of research
   3. Problem Statement
   4. Industry Profile
2. Review of literature
3. Data Collection
4. Methodology
   1. Exploratory Data Analysis

4.1.1 Figures and tables

* 1. Statistical Techniques and Data Visualization
  2. Data Modeling using Supervised ML techniques

1. Findings and Suggestion
2. Conclusion
3. References

BANK MARKETING

# INTRODUCTION:

Nowadays customer is becoming more sophisticated about their choice and quality of service being served to them. Throughout the life, customers try their best to fulfill their demands. The place where things are swapped or say common needs are fulfilled is called the market. The efforts to make people aware of your offerings inspiring them to deal with you and let them believe that in doing so, they are fulfilling their needs at its best is called marketing. After the banking sector improvements, marketing has established as a more combined function within financial service. Financial institutions as banks have done rapid changes in the operational environment. The marketing of bank products has become a very difficult subject as it involves the knowledge of economics, sociology, psychology and also essential marketing idea. In marketing, it is the customer who has the personal choice and the intonation of effective marketing of banking products lies in the regular and professional tactic towards fulfilling customer’s needs.

# ABSTARCT:

Marketing of bank products is the aggregate function absorbed at providing facility to satisfy customer’s monetary needs and wants, more than the rivalry keeping in view the organizational objectives. Banking is a personalized service oriented industry and hence should provide services which satisfy the customers’ needs. The marketing tactic includes forestalling, classifying, responding and satisfying the customers’ needs and wants effectually, professionally, and beneficially. It can be said that the presence of the bank has miniature value without the presence of the customer. The main role of the bank is not only to attain and win more and more customers but also to preserve them through operative customer facility. Marketing as associated to banking is to explain a suitable promise to a customer through a variety of products and services and also to confirm operative distribution through satisfaction. The actual contentment delivered to a customer on how the customer is cooperated with. It goes on to prominence that every employee from the highest executive to the most junior employee of the bank should be concerned with marketing.

# PROBLEM STATEMENT:

This project utilizes different types of Machine Learning algorithms, using the Bank

Marketing dataset, to check if the client has subscribed for a term deposit depending on various bank marketing attributes like age, type of job, education level, if the client has a housing loan or not, last date of contact, etc.

Decision tree has been chosen as the benchmark model for this project and will be compared to the following classifiers:

* CART algorithm.
* Regression.

# INDUSTRY PROFILE

FICO (Fair Isaac Corporation) as a system had (and has) its strengths, but there were weaknesses, too.

Applicants who lacked credit history couldn’t prove their ability to pay, although those without credit history could be working toward a future that would certainly support paying back loans.

Because of reasons like this, a new form of credit checking, banking, loans and mortgages needed to be developed, and that’s where machine learning begin to enter the scene  as detailed in a recent Financial Stability Board summary. From my observation, a lot of new fin tech firms or bank tech firms have erupted in recent times to target newer groups of consumers who prefer to do things virtually. Some banks have also developed solutions to target thin-file customers. For example, technology can help monitor various alternative sources of information on creditworthiness, like ensuring they are paying rent and utilities on time.

This might seem like a complicated concept, but there are several machine learning factors that can benefit the lending industry.

Following are some of the potential benefits machine learning could provide:

* Detect Frauds And Threats to Financial Systems.
* Personalize Customer Service In A Digital World.

**DATA COLLECTION**

The data collection consists of

* **Data Set Information**
* **Attribute Information**

**Data Set Information**

The data is related to bank marketing. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

The dataset is Bank additional-full.csv with all examples (41188) and 21 inputs.

The classification goal is to predict if the client will subscribe (yes/no) a term deposit (variable y).

**Attribute Information:**

### **Bank client data:**

* Age (numeric)
* Job : type of job (categorical: 'admin.', 'blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'self-employed', 'services', 'student', 'technician', 'unemployed', 'unknown')
* 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')

### **Related with the last contact of the current campaign:**

* Contact: contact communication type (categorical: 'cellular','telephone')
* Month: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
* Day\_of\_week: last contact day of the week (categorical: 'mon','tue','wed','thu','fri')
* Duration: last contact duration, in seconds (numeric). Important note: this attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.

### **Other attributes:**

* Campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)
* Pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
* Previous: number of contacts performed before this campaign and for this client (numeric)
* Poutcome: outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')

### **Social and economic context attributes**

* Emp\_va\_\_rate: employment variation rate - quarterly indicator (numeric)
* Cons\_price\_idx: consumer price index - monthly indicator (numeric)
* Cons\_conf\_idx: consumer confidence index - monthly indicator (numeric)
* Euribor3m: euribor 3 month rate - daily indicator (numeric)
* Nr\_employed: number of employees - quarterly indicator (numeric)

### **Output variable (desired target):**

* y - Has the client subscribed a term deposit? (binary: 'yes', 'no')

Out of the 21 attributes in the dataset some of the attributes are dropped from the dataset as they are independent of the target attribute.

The dropped attributes are pdays ,default, contact, day\_of\_week, month.

# 4 Methodology

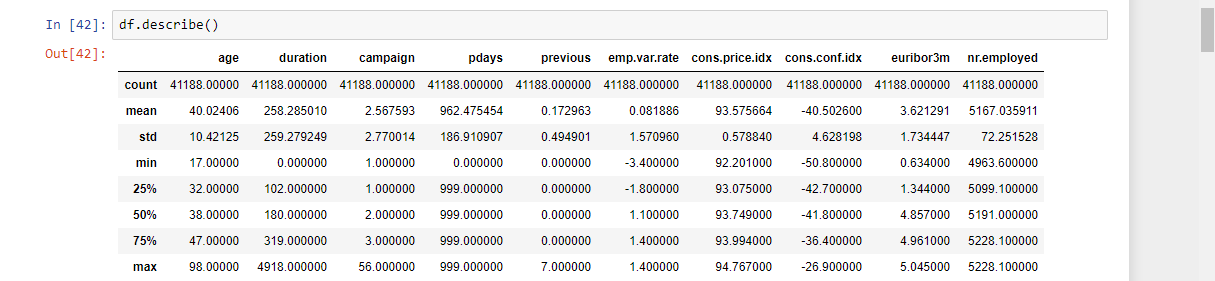
Methodology consists of

* Exploratory Data Analysis
* Figures and tables
* Statistical Techniques and Data Visualization
* Data Modeling using Supervised ML techniques

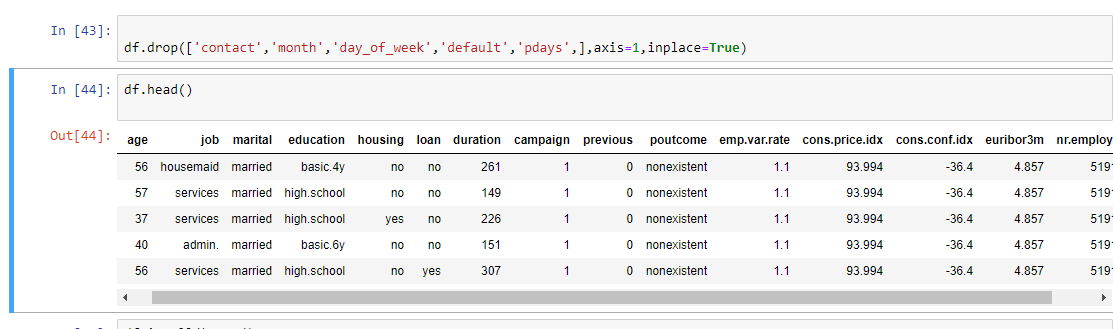
# 4.1 EXPLORATORY DATA ANALYSIS

To obtain a better understanding of the dataset, the distribution of key variables and the relationships among them were plotted.

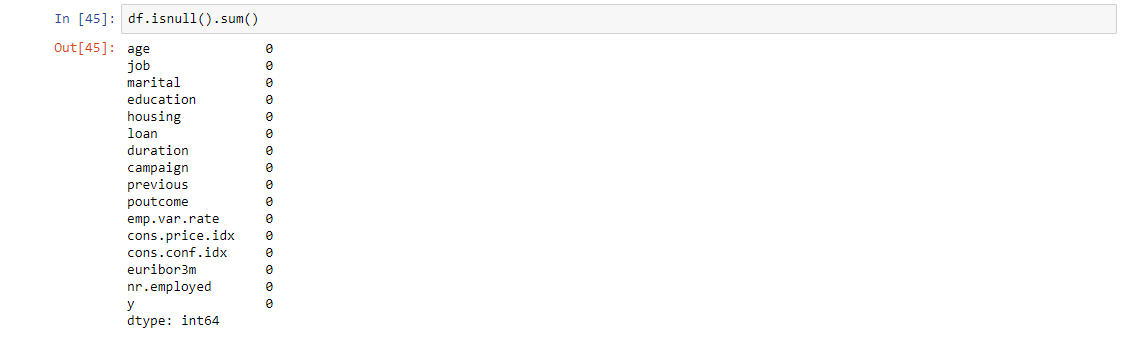
# 4.1.1 FIGURES AND TABLES



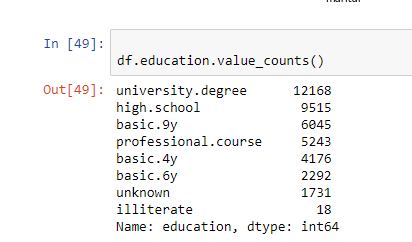
Pandas **describe()** is used to view some basic statistical details like percentile, mean, std etc. of a data frame or a series of numeric values. When this method is applied to a series of string, it returns a different output

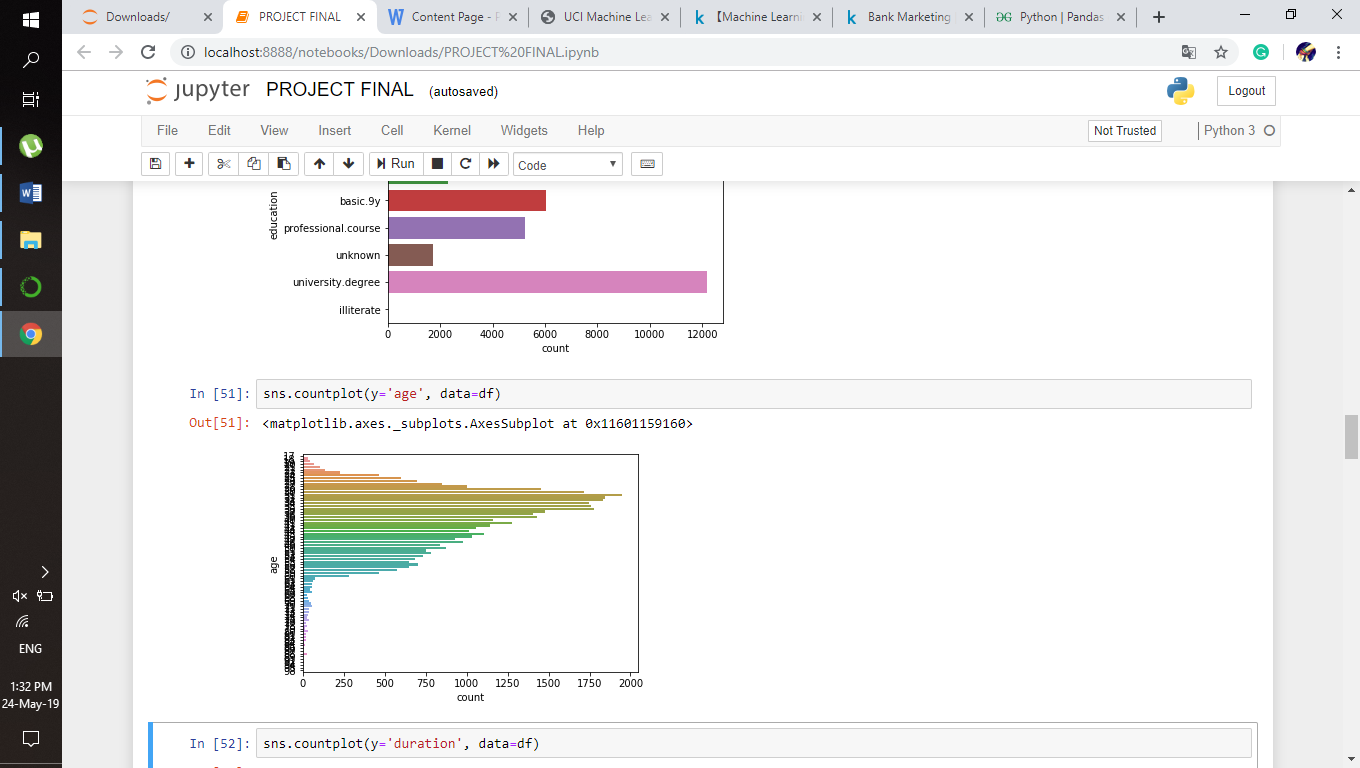


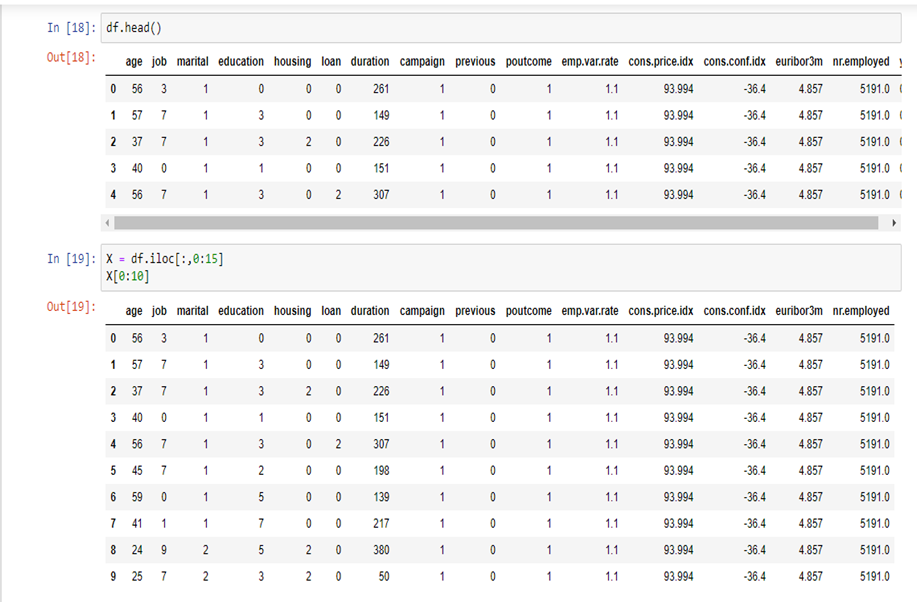
Pandas provide data analysts a way to delete and filter data frame using **drop()**method. Rows or columns can be removed using index label or column name using this method.

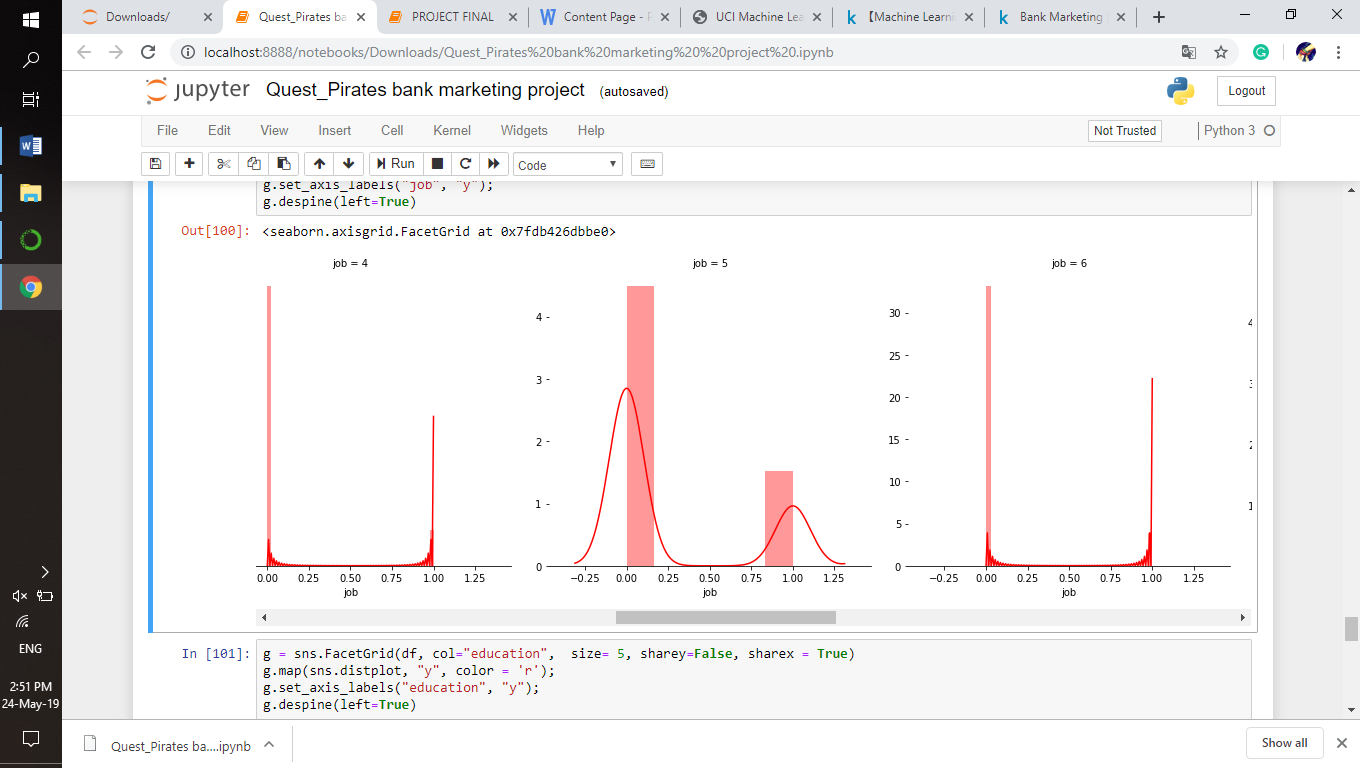


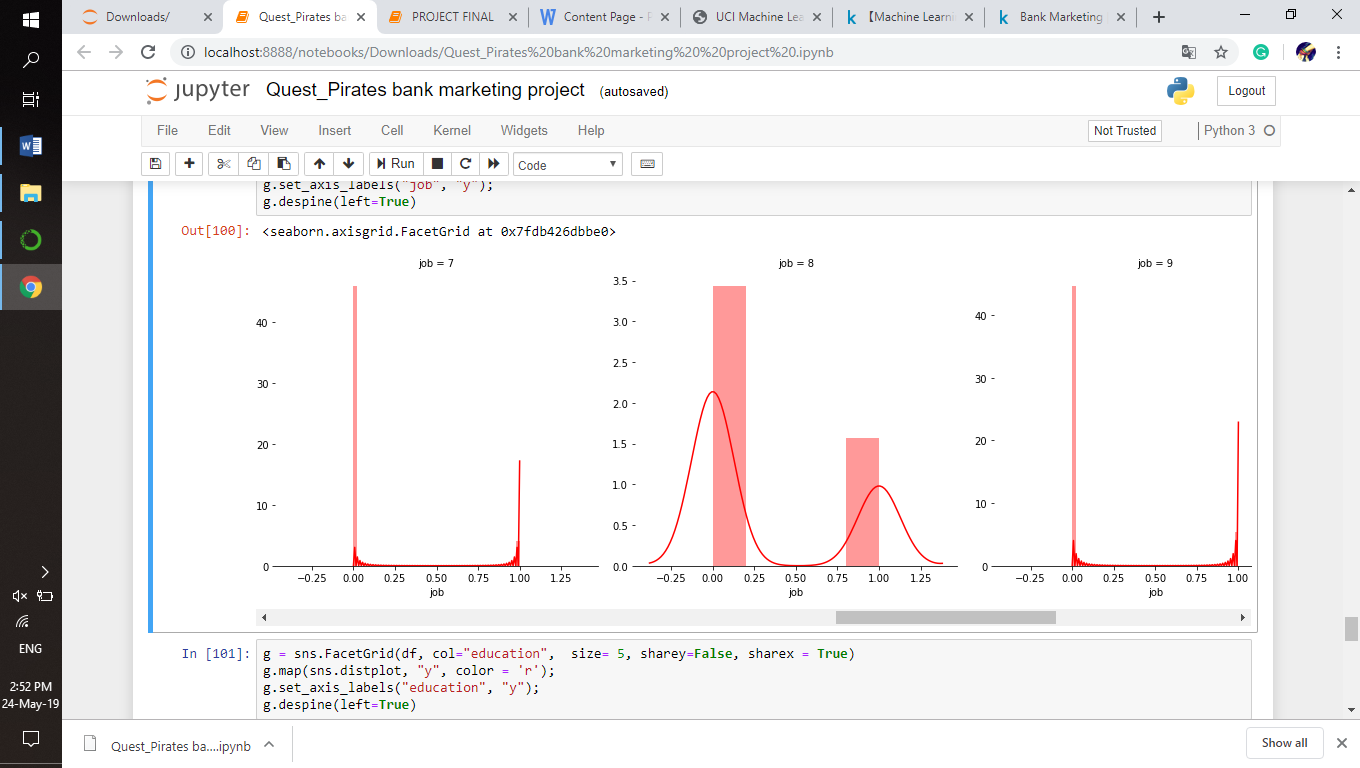
While making a Data Frame from a csv file, many blank columns are imported as null value into the Data Frame which later creates problems while operating that data frame. Pandas isnull() and notnull() methods are used to check and manage NULL values in a data frame. Pandas**dataframe.sum()** function return the sum of the values for the requested axis.







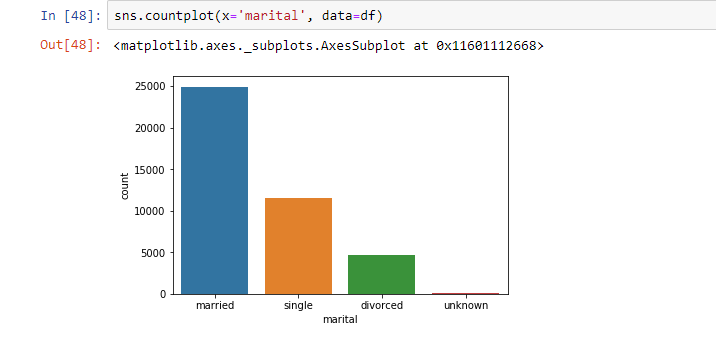


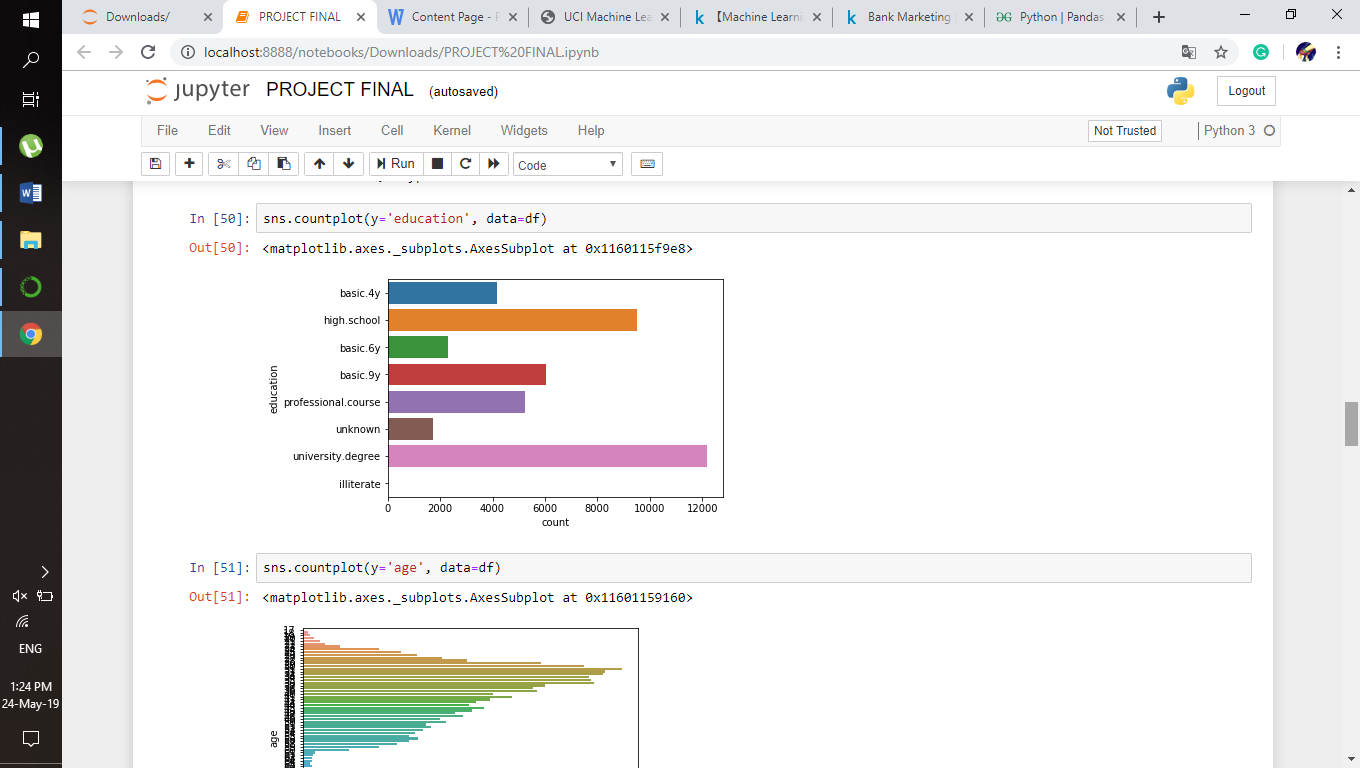


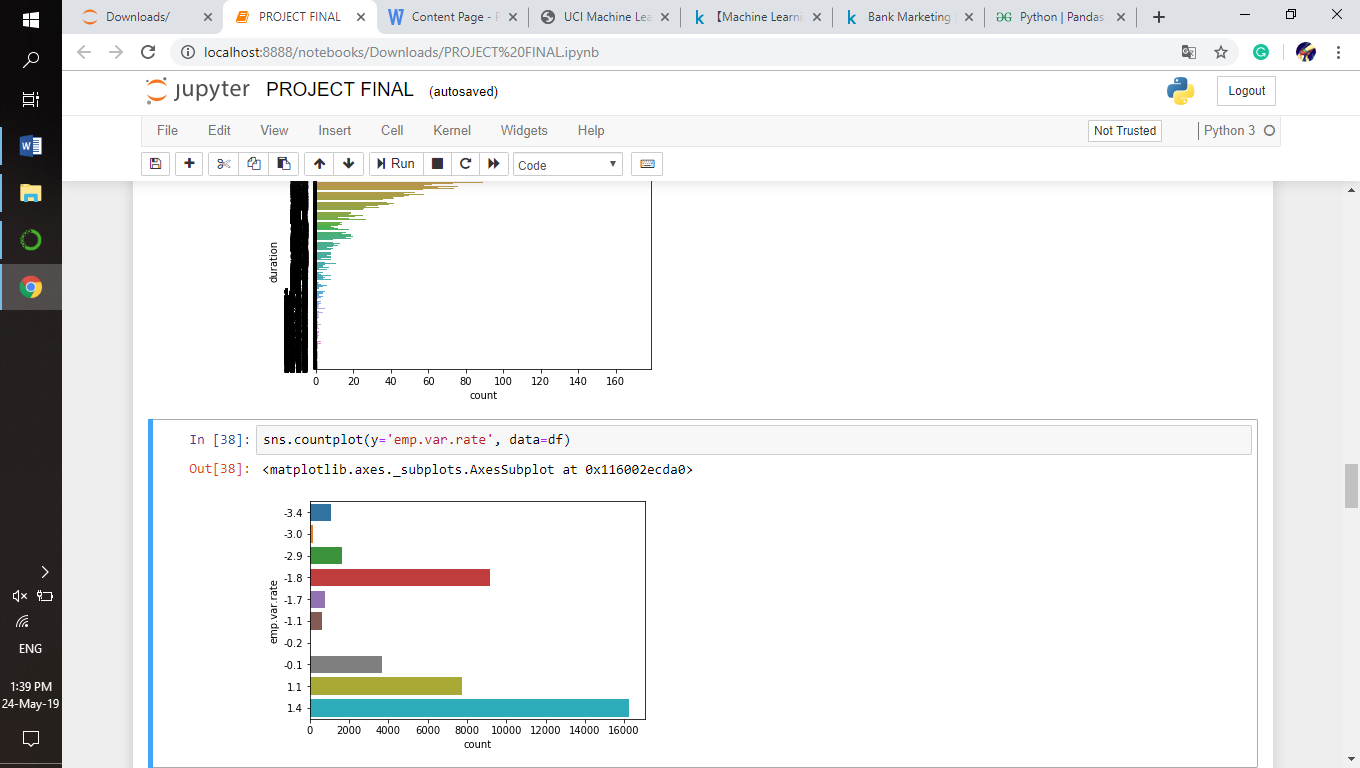
# 4.2 STATISTICAL TECHNIQUES AND DATA VISUALIZATION

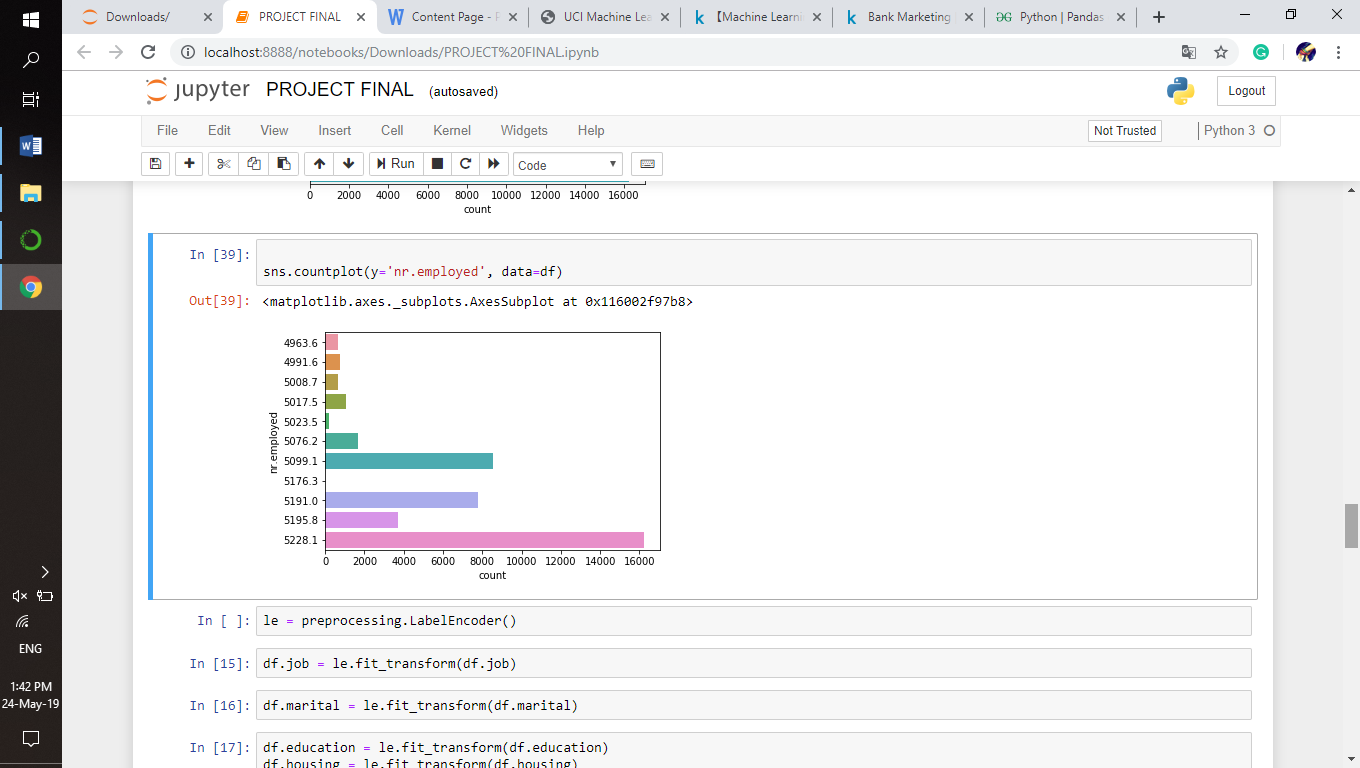
# C:\Users\hp\OneDrive\Pictures\Screenshots\2019-05-24 (5).png

# C:\Users\hp\OneDrive\Pictures\Screenshots\2019-05-24 (6).png



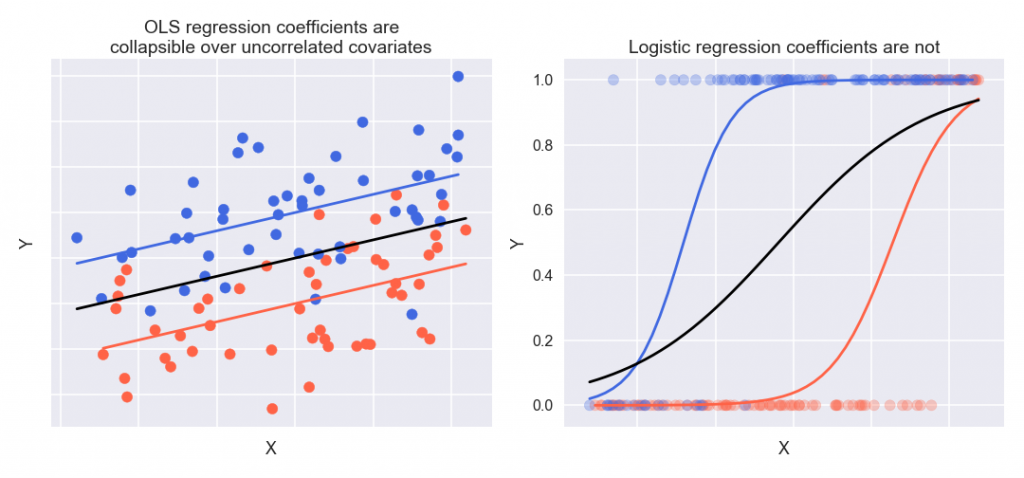






# 4.3 DATA MODELING USING SUPERVISED ML TECHNIQUES

[**Logistic regression**](http://www.statisticssolutions.com/academic-solutions/membership-resources/member-profile/data-analysis-plan-templates/data-analysis-plan-logistic-regression/) is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary).  Like all regression analyses, the logistic regression is a predictive analysis.  Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.



# 5 FINDING AND SUGGESTIONS

* Here duration attribute plays an important role in prediction. Duration attribute indicates the call duration between the customer and bank employee. The more the duration lasts, more likely the customer is interested in crediting (subscribing) for the deposit.
* From the count plots we can infer that the people who are married are more likely to subscribe for the deposit.
* And for the education attribute, the people with university degree holders are more likely to subscribe for the deposit.
* When implementing a marketing strategy, external factors, such as the time of calling, should also be carefully considered.

# 6 CONCLUSION

* Applying logistic regression algorithms, classification and estimation model were successfully built. With this model, the bank will be able to predict a customer's response to its telemarketing campaign before calling this customer.
* In this way, the bank can allocate more marketing efforts to the clients who are classified as highly likely to accept term deposits, and call less to those who are unlikely to make term deposits.
* Predicting the duration before calling and adjusting marketing plan benefit both the bank and its clients.

# 7 REFERENCE

* <https://www.kaggle.com/yufengsui/machine-learning-project-bank-marketing-analytics/notebook>
* <https://archive.ics.uci.edu/ml/datasets/bank+marketing>