**Name : Bhushan S Kanade**

**NPV MINI PROJECT**

**REPORT**

**TOPIC**

**FINANCE**

**PROBLEM STATEMENT:**

**The problem is that the Bank Marketing campaigns of a Portuguese banking institution need to identify the factors that cause the customers to tend to take the subscription, as well as Bank Marketing campaigns of a Portuguese banking institution need to identify the reasons behind the customer which make them not take the subscription.**

**Technologies Used :**

**Python , Jupyter Notebook**

**Data Dictionary:**

**DATA DICTIONARY:**

1)age(numeric)

2) job : type of job (categorical:

"admin.","unknown","unemployed","management","housemaid","entrepreneur", "student","bluecollar","self-employed","retired","technician","services")

3 - marital : marital status (categorical: "married","divorced","single"; note: "divorced" means

divorced or widowed)

4 - education (categorical: "unknown","secondary","primary","tertiary")

5 - default: has credit in default? (binary: "yes","no")

6 - balance: average yearly balance, in euros (numeric)

7 - housing: has a housing loan? (binary: "yes","no")

8 - loan: has personal loan? (binary: "yes","no")

- related to the last contact of the current campaign:

9 - contact: contact communication type (categorical: "unknown","telephone","cellular")

10 - day: last contact day of the month (numeric)

11 - month: last contact month of year (categorical: "jan", "feb", "mar", ..., "nov", "dec")

12 - duration: last contact duration, in seconds (numeric)

- other attributes:

13 - campaign: number of contacts performed during this campaign and for this client (numeric,

includes the last contact

14 - P-days: number of days that passed by after the client was last contacted from a previous

campaign (numeric, -1 means client was not previously contacted)

15 - previous: number of contacts performed before this campaign and for this client (numeric)

16 – poutcome : outcome of the previous marketing campaign (categorical:

" unknown "," other "," failure ","success")

- output variable (desired target):

17 - y - has the client subscribed to a term deposit? (binary: "yes","no")

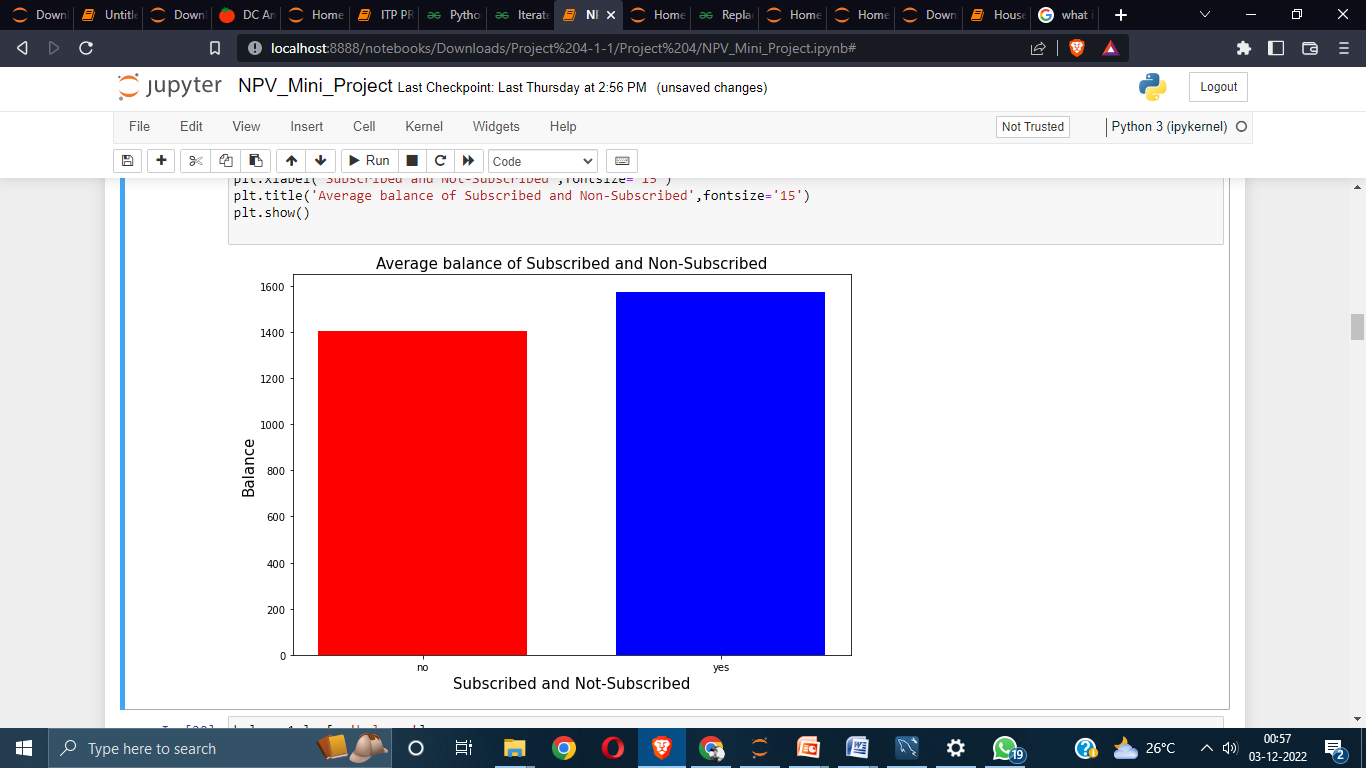
**Data Collection:**

* So I used python programming language and tool used is Jupyter notebook.
* I retrieved the data through with statement and created the function to load the dataset.

**Data Cleaning:**

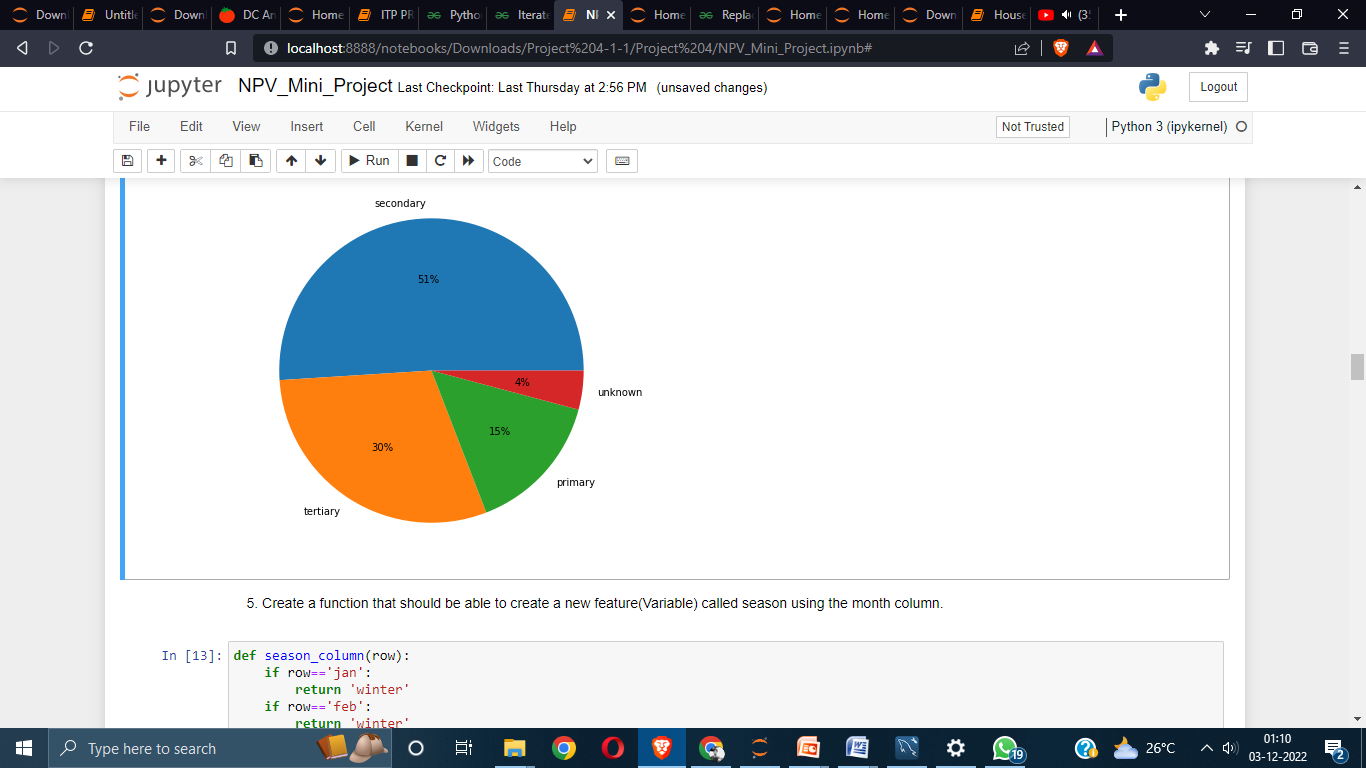
* So I identified the Features and Data types of each column
* I converted the Data types of numerical columns from object to integer.
* I checked for the missing values but there weren’t any.
* I checked for duplicate values and there were 0 duplicate rows

**Average Balance of the Customers:**

****

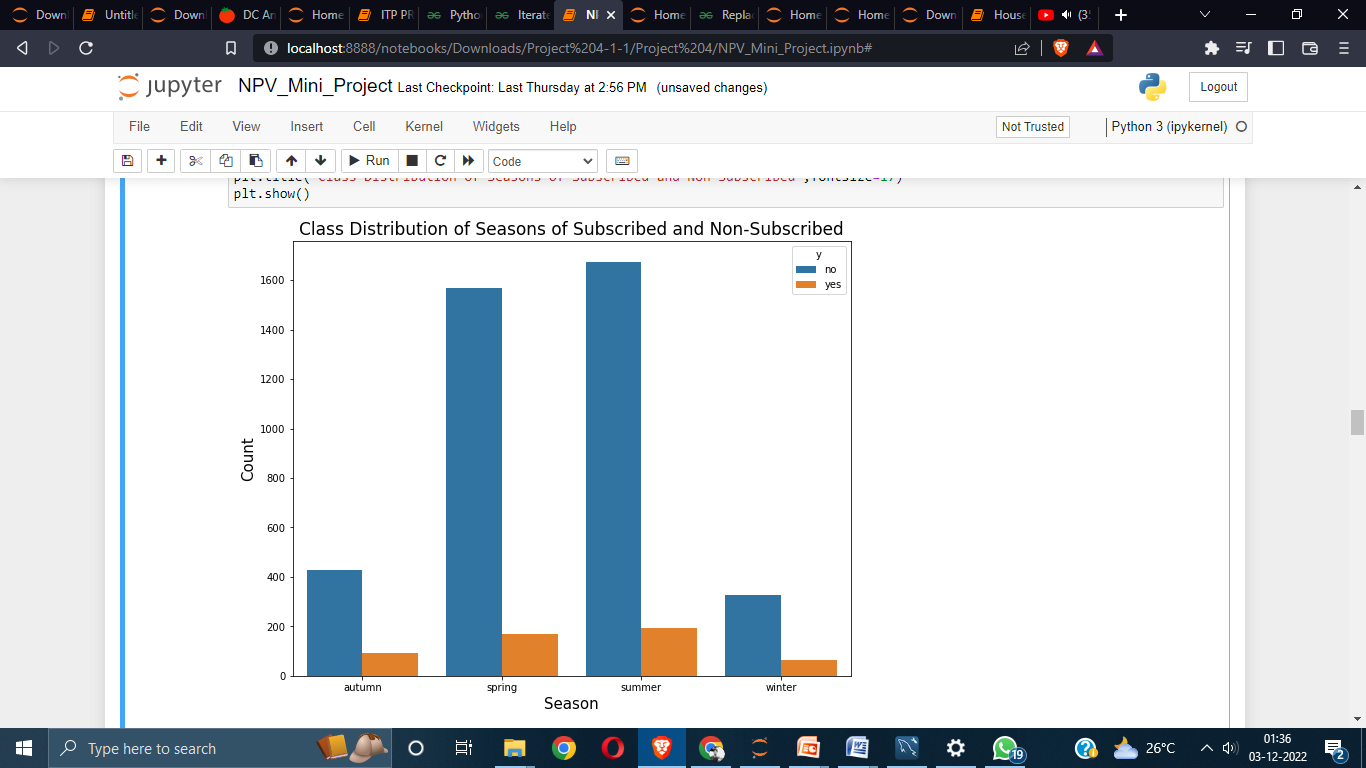
This is quite obvious that average balance of the customers who are subscribed to term deposit is higher than those who are not-subscribed which is around 1570 but as we can see in the above bar graph that average balance of the customers who are not subscribed is not much lower than the subscribed one which is around 1400.

**Education of Customers:**

****

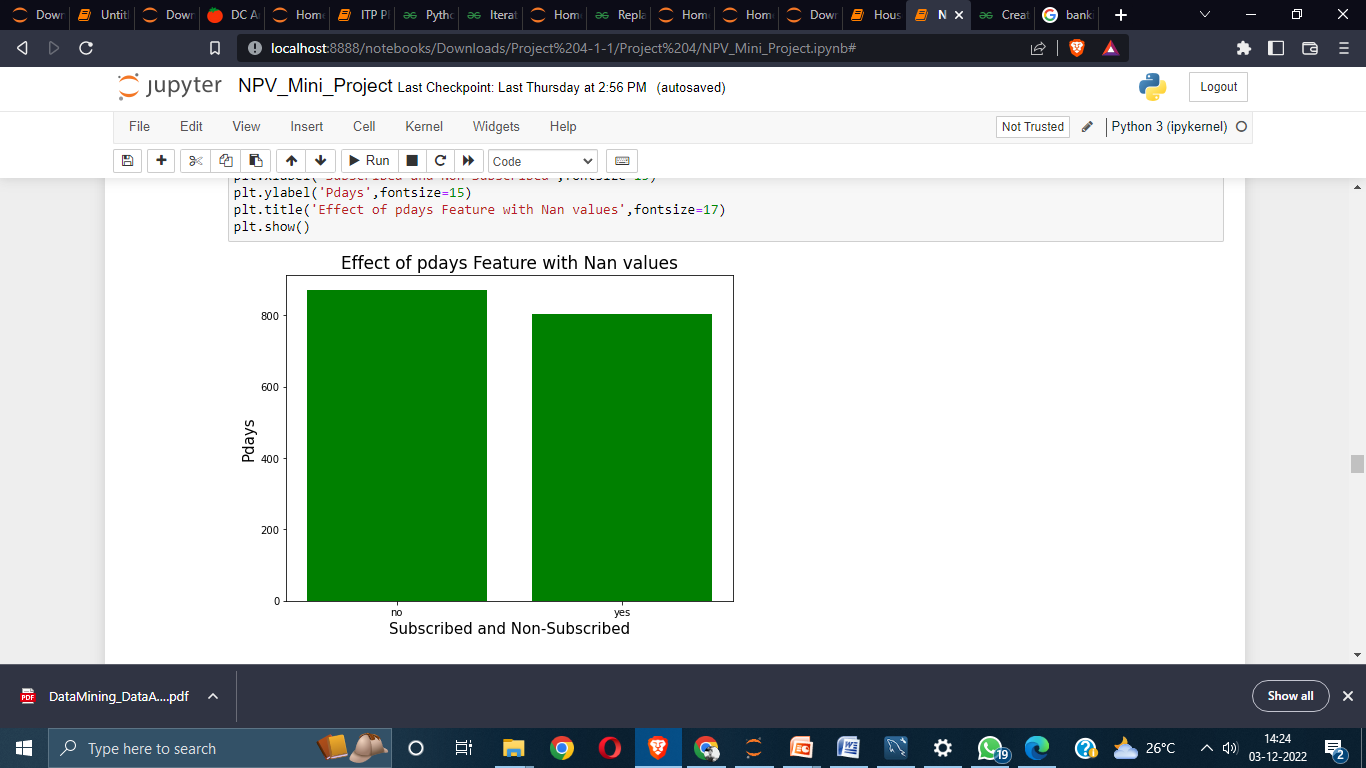
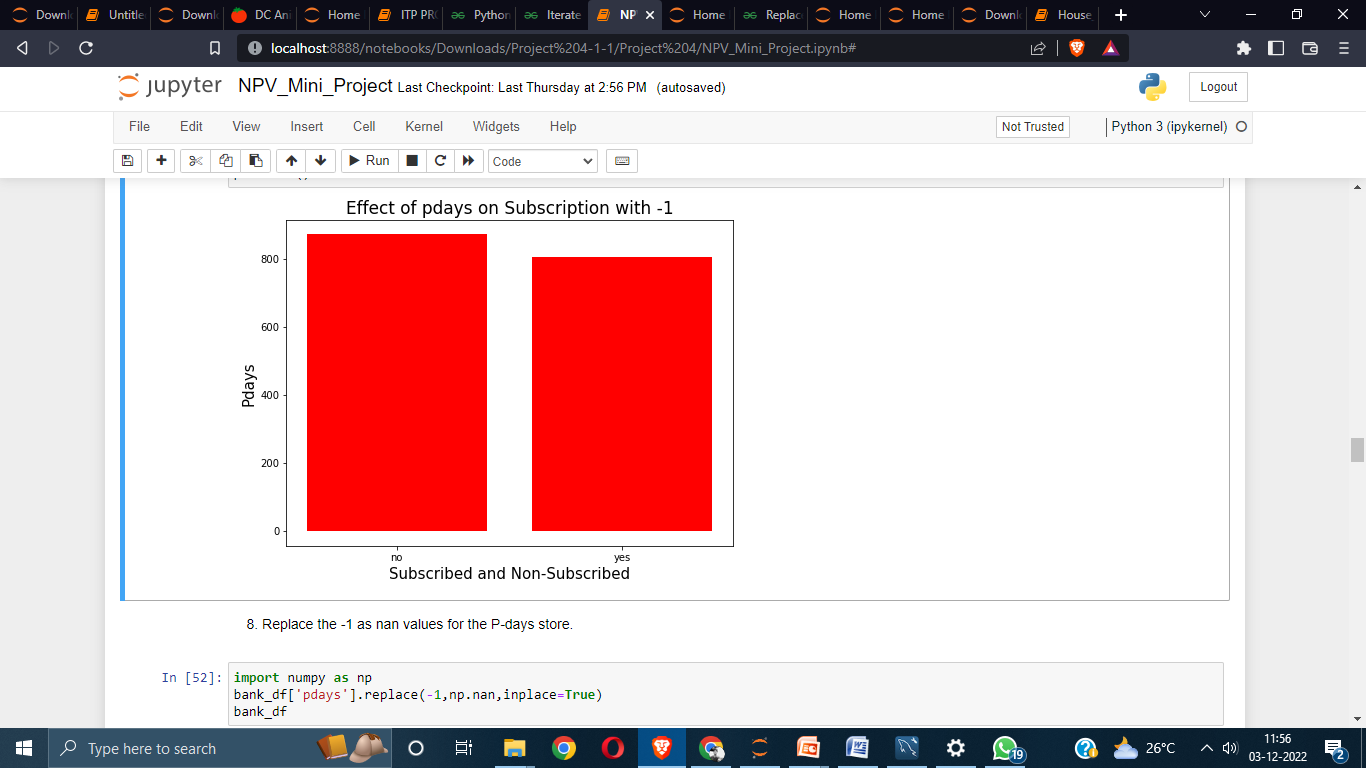
**As we can see in the above pie chart out of all the customers around 51% of customers comes from the background of secondary education followed by tertiary education with 30% then primary education with 15% and unknown with 4%.**

**Class Distribution of Seasons**

****

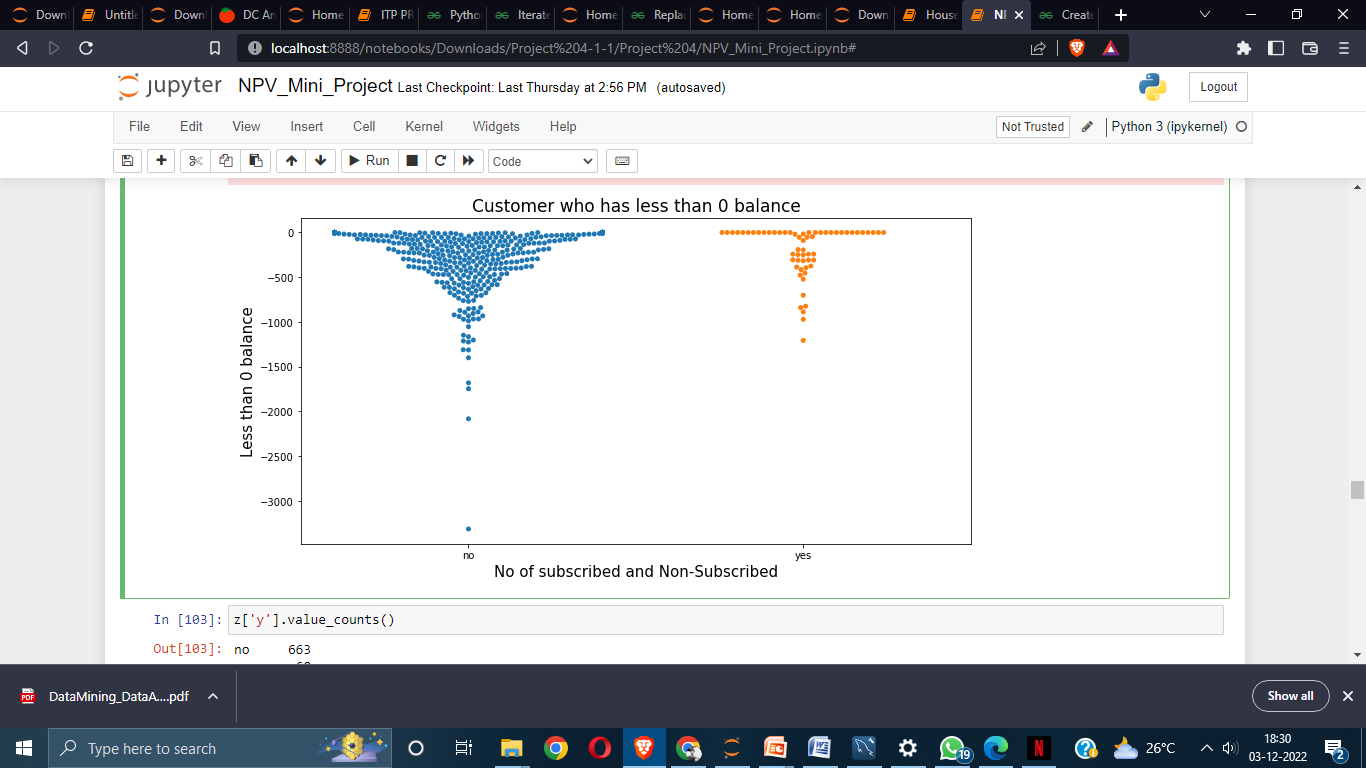
**The above graph shows class distribution of seasons with respect to the number of people who are subscribed and not subscribed, so this shows that large number of people who were subscribed as well as not-subscribed to the term deposit were in the seasons of spring and summer. In the season of autumn the numbers were lower than spring and autumn and in winter the number were lowest.**

**Effect of Pdays:**

****

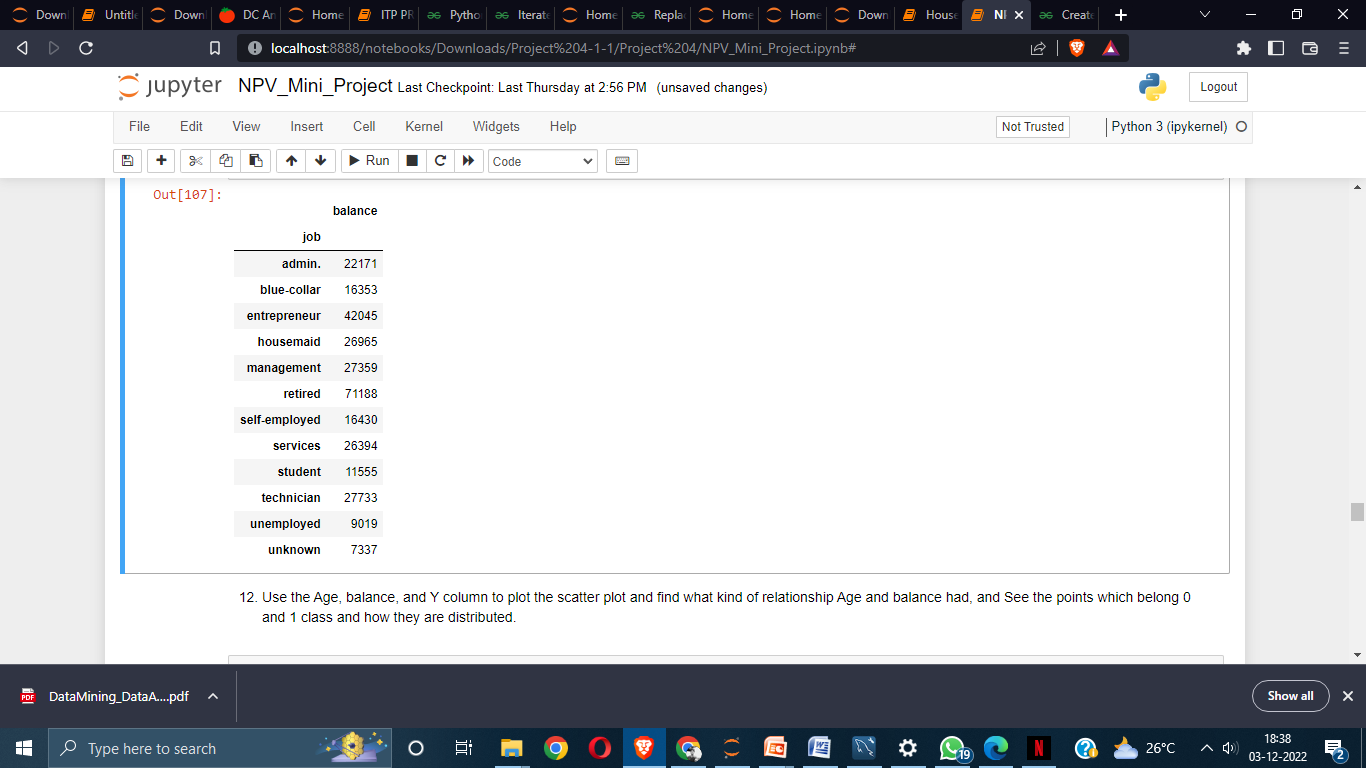
**The first picture shows the bar graph of pdays including -1 which means the number of customers who were not contacted previously**

**Customers with less than 0 balance:**

****

In the above stripplot we can see the no customers who have less than 0 balance and not-subscribed are more than the customers who have subscribed and the number varies in a large amount where the customers who have subscribed is 60 whereas customer who are not subscribed are 663.

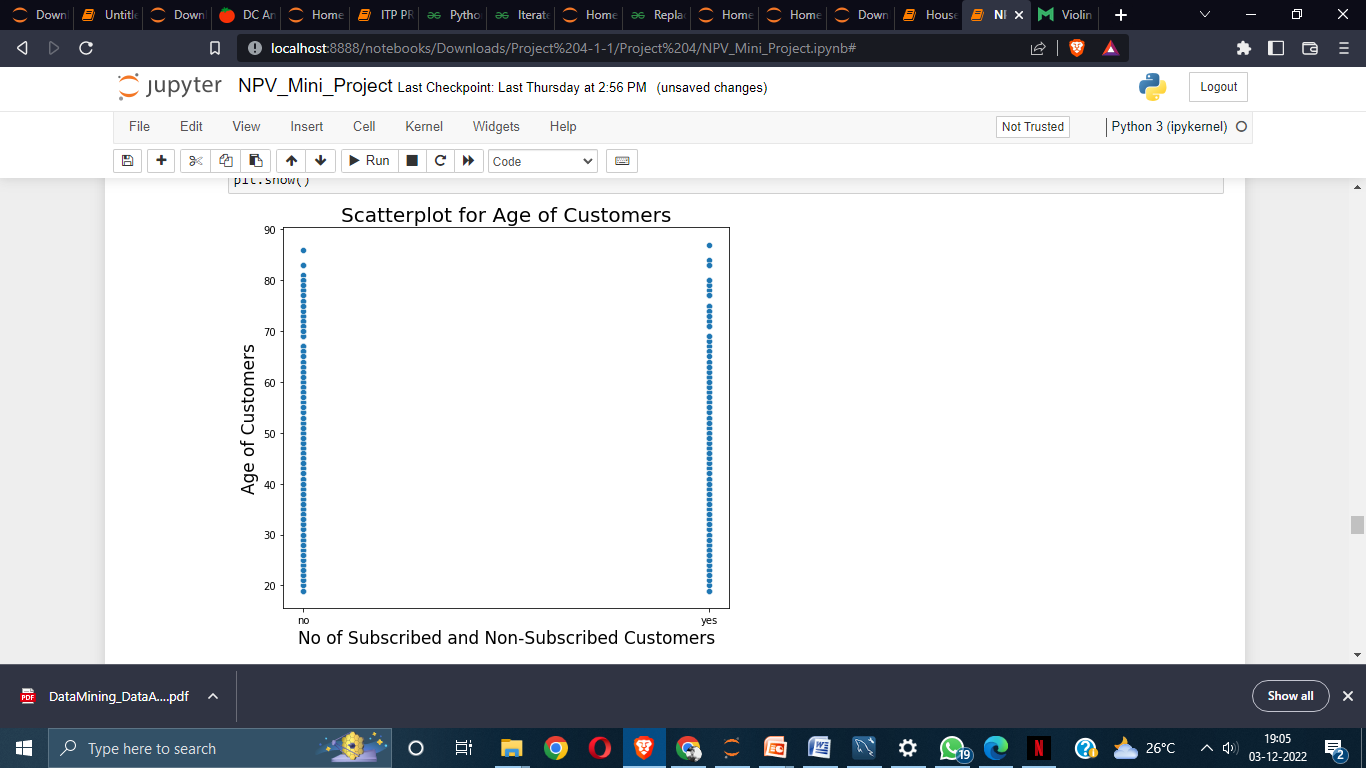
**Maximum Balance for Each Job:**

****

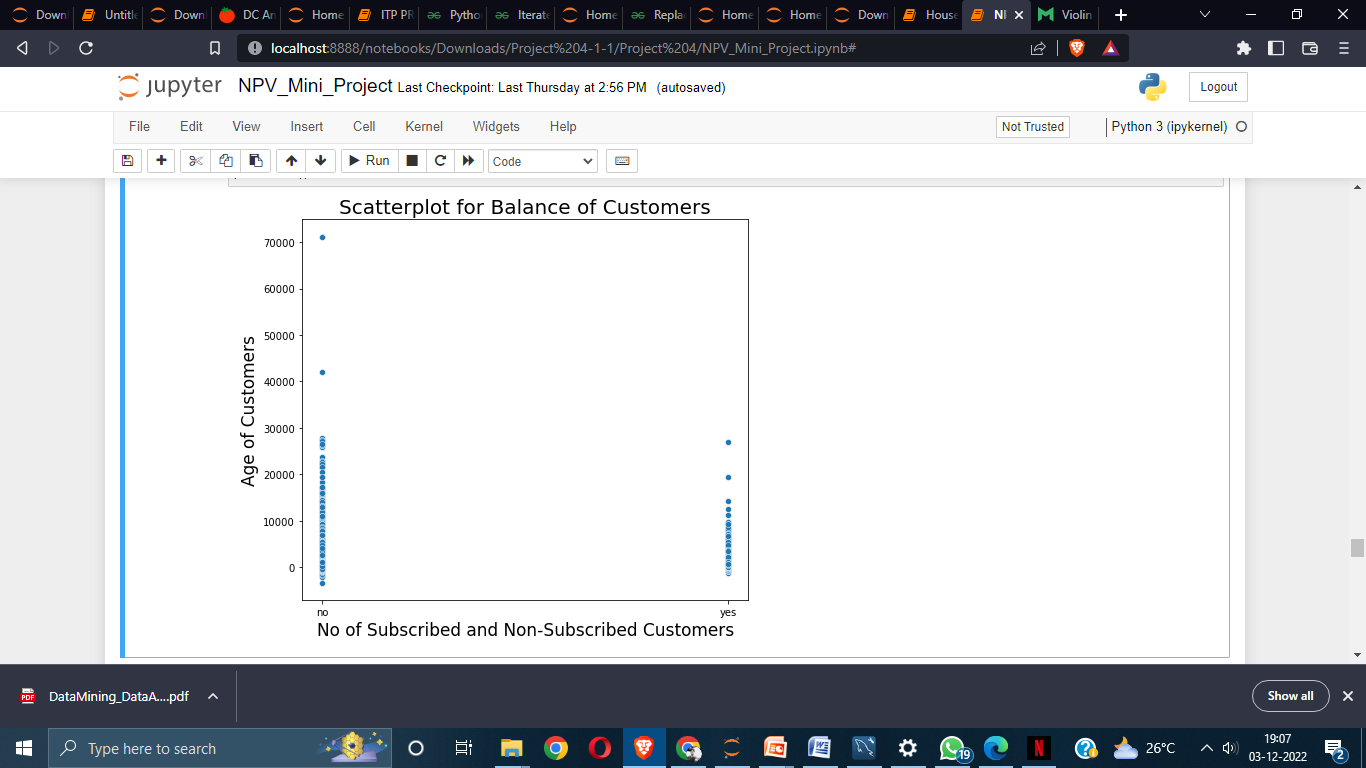
**In this table, we can see the maximum account balance of a customer for each type of jobs like admin, entrepreneur, housemaid, etc. We can also see maximum balance of the customers who are retired and unemployed.**

**Scatterplot of Age and Balance**

**Here we can see the scatter plot of balance of the customers with respect to the number of people who are subscribed or not**

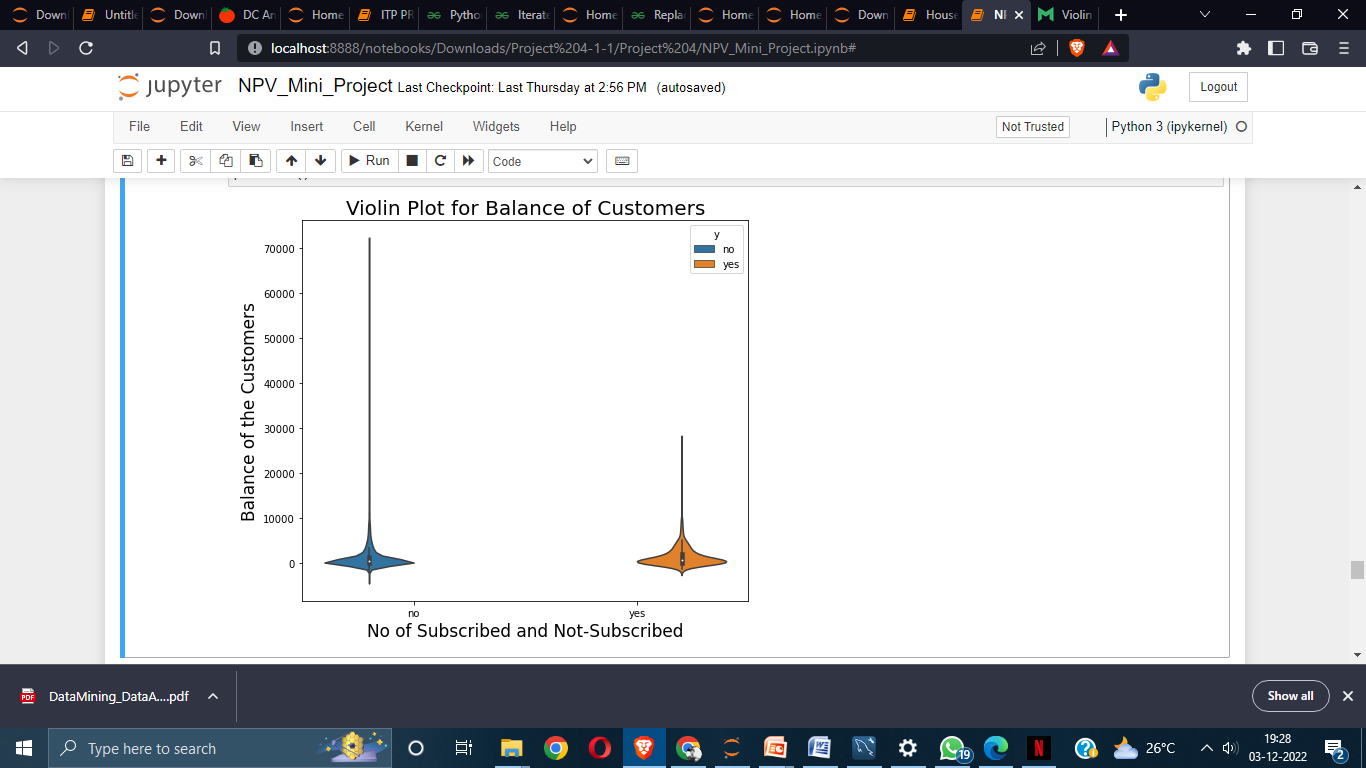
****

**Here we can see the scatter plot of balance of the customers with respect to the number of people who are subscribed or not**

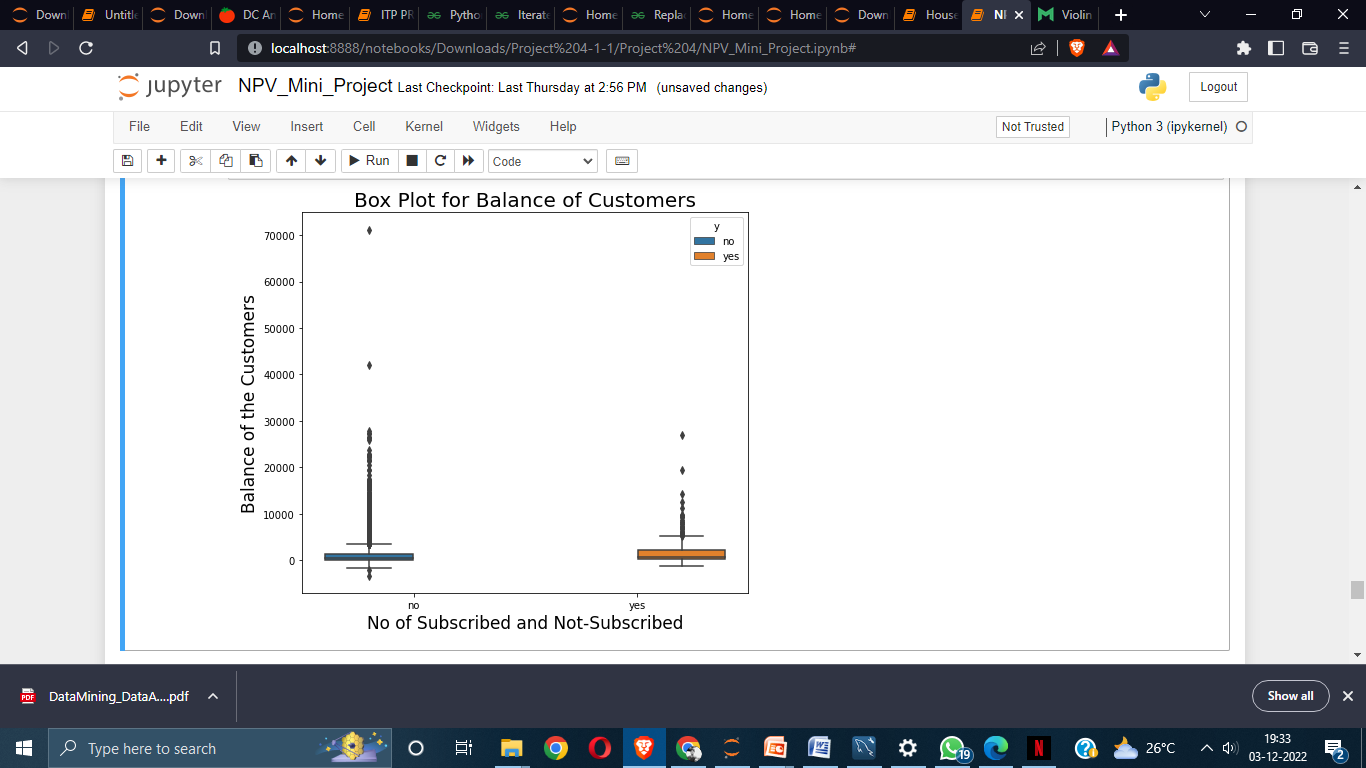
****

**Here we can see the scatter plot of balance of the customers with respect to the number of people who are subscribed or not**

Box plot and violin plot of Balance of Customers:

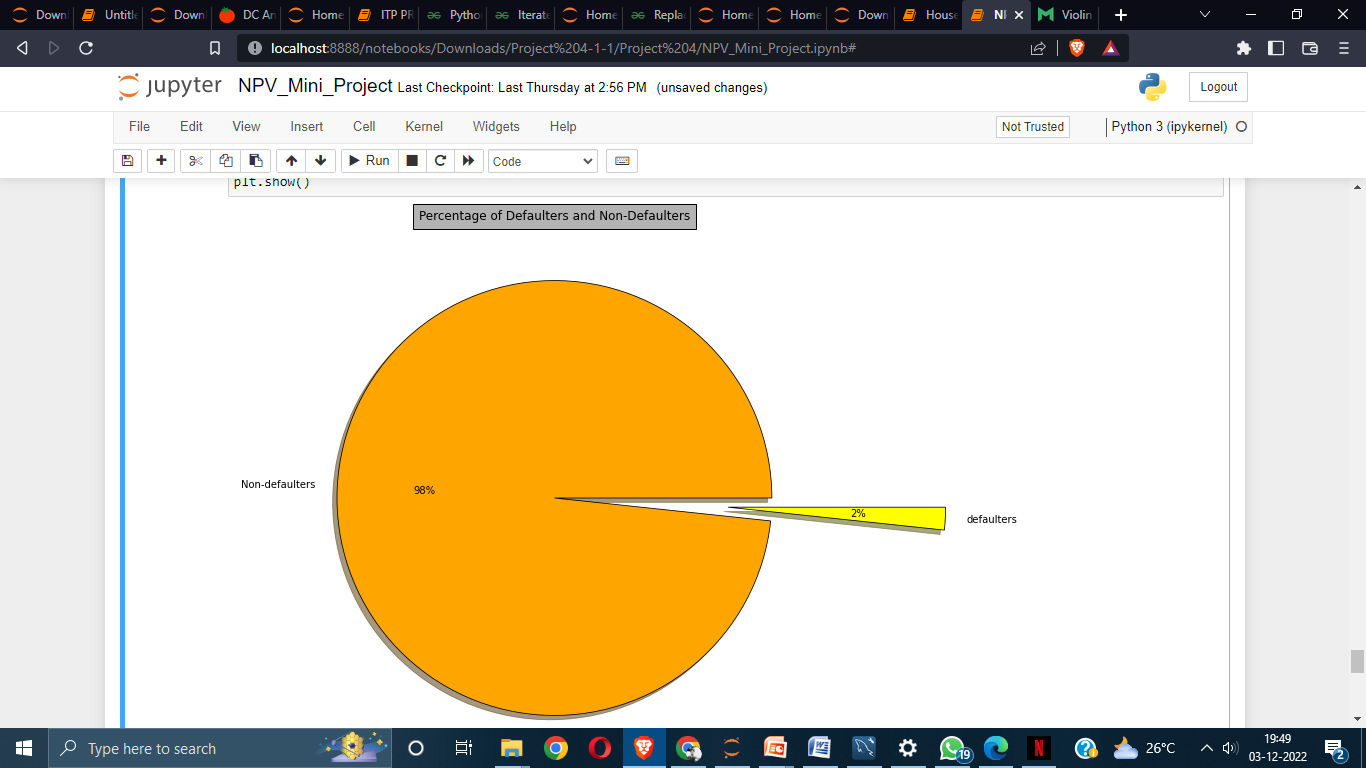
****

Here we can see violin plot of no of subscribed and non-subscribed customers with respect to subscribed and non-subscribed and we can see here the large no of customers have 0 or less than 0 balance and most number of outliers are also in the non-subscribed and who are subscribed most also have more no of customers who have 0 balance but the outliers are very less in the subscribed part.

****

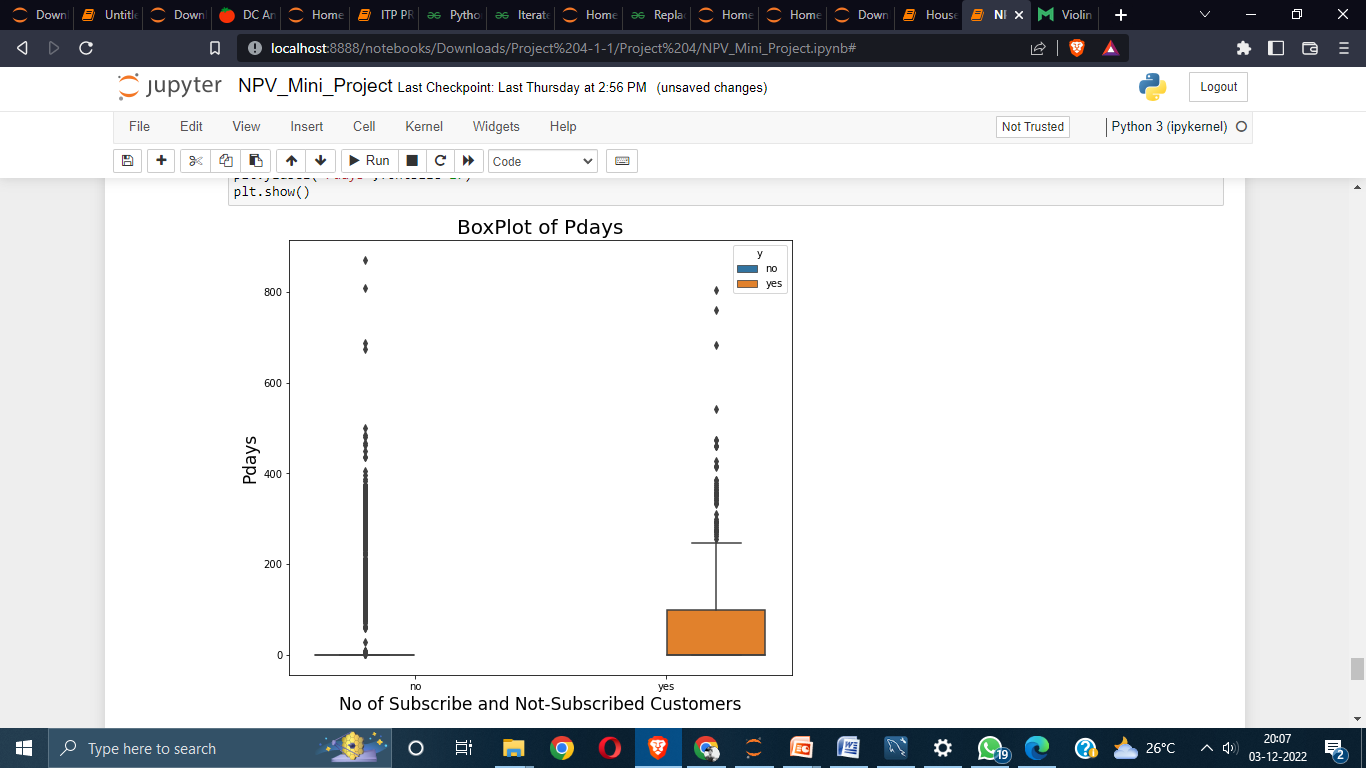
Here we can see violin plot of no of subscribed and non-subscribed customers with respect to subscribed and non-subscribed and we can see the midpoint inside the box is on 0 that means that there are more number of customers who has 0 balance or less than 0 balance who have not subscribed and the outliers of balance majority of customers are not more than 30000 and some have balance around 42000.In the other box we can see the midpoint of box is 0 and the outliers of balance doesn’t go above 30000

**Distribution of Defaulters and Non-Defaulters**

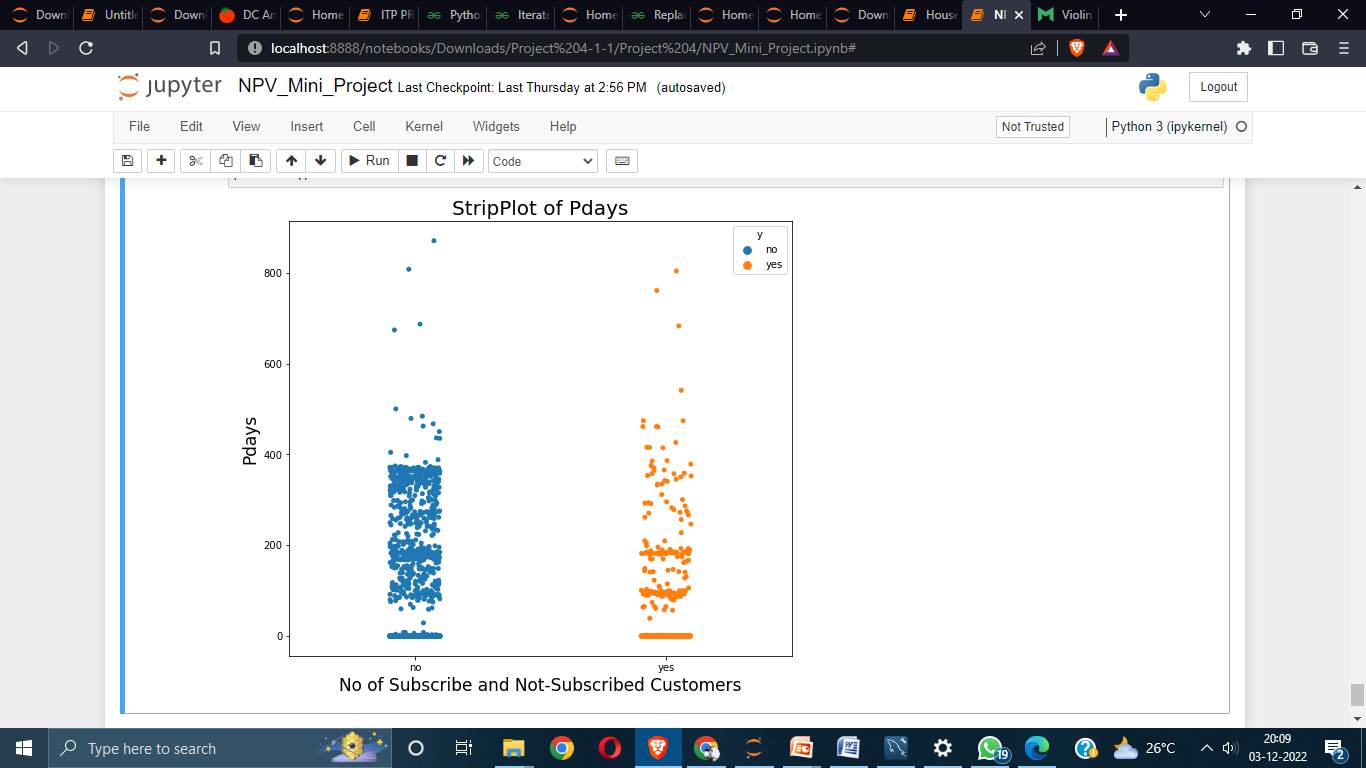


Above we can see distribution percentage of defaulters and Non-Defaulters of the customers where we can see there are 98% Non-Defaulters and 2% defaulters

**Box Plot and Strip Plot of Pdays of Customers**

****

Above we can see the box plot of customers who are subscribed and non-subscribed with respect to pdays

****

Above we can see the box plot of customers who are subscribed and non-subscribed with respect to pdays

**Summary**

1. **So in this dataset we analysed that how different factors like Pdays , age , balance of customers affect the customers choice of subscribing to term deposit or not.**
2. **We also noticed that majority of the customers are non-defaulters around 98% and only 2% customers are defaulters**
3. **The majority of the customers who have less than 0 balance amount haven’t subscribed to term deposit**
4. **The large amount of customers comes under the secondary education , followed by tertiary and secondary.**
5. **Average balance of the customers who are subscribed to a term deposit is around 1570/- and non-subscribed is just bit lesser than subscribed which is close to 1400**
6. **The job which has maximum balance is entrepreneur with 42045/- , and if we look for just highest maximum balance then the customers who are retired have maximum balance with 71188.If we exclude the unknown, unemployed and student column then the account with least balance are the type of customers who have job as blue collar.**