ADS505_Final_Project_v2

October 17, 2022

1 Final Project

1.0.1 ADS 505 Applied Data Science for Business

Team_6

1.1 Import of Packages and Libraries

[ ]:
```
pip install dmba
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
wheels/public/simple/
Collecting dmba
  Downloading dmba-0.1.0-py3-none-any.whl (11.8 MB)
     |                  | 11.8 MB 5.1 MB/s
Installing collected packages: dmba
Successfully installed dmba-0.1.0
```

[ ]:
```
%matplotlib inline
from pathlib import Path
from sklearn import preprocessing
import numpy as np
import pandas as pd
from sklearn.linear_model import LogisticRegression, LogisticRegressionCV
from sklearn.model_selection import train_test_split
from sklearn.neighbors import NearestNeighbors, KNeighborsClassifier
import statsmodels.api as sm
import matplotlib.pylab as plt
import seaborn as sns
from dmba import classificationSummary, gainsChart, liftChart
from dmba.metric import AIC_score
from sklearn.metrics import accuracy_score
from sklearn.linear_model import LinearRegression, Lasso, Ridge, LassoCV,␣
↪BayesianRidge
import statsmodels.formula.api as sm
from sklearn.tree import DecisionTreeClassifier, DecisionTreeRegressor
```

```
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from sklearn.model_selection import train_test_split, cross_val_score,␣
↪GridSearchCV
import matplotlib.pylab as plt
from dmba import plotDecisionTree, classificationSummary, regressionSummary
from dmba import regressionSummary, exhaustive_search
from dmba import backward_elimination, forward_selection, stepwise_selection
from dmba import adjusted_r2_score, AIC_score, BIC_score
from sklearn.neural_network import MLPClassifier
from dmba import classificationSummary
from sklearn.preprocessing import MinMaxScaler
from imblearn.under_sampling import NearMiss
from sklearn.impute import SimpleImputer
from sklearn.metrics import classification_report
from sklearn import metrics
from sklearn.metrics import recall_score
from sklearn.metrics import classification_report, confusion_matrix

import warnings
warnings.filterwarnings('ignore')
```

```
no display found. Using non-interactive Agg backend
```

1.2 Load Data

[ ]:
```
df = pd.DataFrame(pd.read_csv('/content/cross_sell_dataset (1).tab', sep='\t'))
df
```

[ ]:

	cross_buy	acad_title	age	calls	complaints	customer_tenure_months	\
0	0	0	60	0	0	221	
1	0	0	55	0	0	227	
2	0	0	61	0	0	221	
3	0	0	70	0	0	222	
4	0	1	61	0	0	227	
…	…	…	…	…	…	…	
99995	1	0	53	0	0	206	
99996	1	0	25	0	0	206	
99997	1	0	19	0	0	205	
99998	1	0	58	0	0	204	
99999	1	0	55	0	0	205	

	directmails	gender	joint_account	inflows	…	volume_debit	\
0	0	0.0	0.0	0.0	…	0.00	
1	0	0.0	1.0	0.0	…	3.28	
2	0	1.0	0.0	3000.0	…	31963.13	
3	0	0.0	0.0	6000.0	…	54048.40	

```
4                0     1.0            0.0      0.0  …    1374743.09
…              …      …              …       …   …           …
99995            0     0.0            0.0   3300.0  …      28839.32
99996            1     0.0            0.0    150.0  …       2133.01
99997            1     1.0            0.0   2350.0  …      47476.22
99998            0     1.0            0.0      0.0  …       1573.06
99999            2     0.0            1.0  20000.0  …      39503.93

       volume_debit_6months  ext_city_size ext_house_size  ext_purchase_power  \
0                      0.00            7.0            4.0                 5.0
1                      3.28            7.0            1.0                 7.0
2                  28963.13            8.0            4.0                 4.0
3                  48048.40            2.0            1.0                 7.0
4                 1341722.91            4.0            1.0                 7.0
…                      …              …              …                   …
99995              25611.08            5.0            1.0                 5.0
99996               1983.01            1.0            1.0                 3.0
99997              41939.13            8.0            3.0                 5.0
99998               1373.06            8.0            1.0                 7.0
99999              14823.28            7.0            1.0                 3.0

       ext_share_new_houses  ext_share_new_cars ext_car_power  \
0                       4.0                 4.0           3.0
1                       1.0                 3.0           1.0
2                       1.0                 7.0           3.0
3                       3.0                 5.0           4.0
4                       3.0                 6.0           5.0
…                      …                   …             …
99995                   1.0                 7.0           5.0
99996                   5.0                 2.0           1.0
99997                   1.0                 6.0           1.0
99998                   1.0                 NaN           NaN
99999                   4.0                 1.0           4.0

       ext_living_duration  giro_mailing
0                      9.0             0
1                      NaN             0
2                      9.0             0
3                      9.0             0
4                      7.0             0
…                     …              …
99995                  9.0             0
99996                  9.0             0
99997                  9.0             0
99998                  7.0             0
99999                  7.0             0
```

```
[100000 rows x 35 columns]
```

```
[ ]: # Summary Information about the variables and their types in the data:
data_desc = pd.DataFrame(pd.read_csv('/content/Data Set Description.tab - Data␣
 ↪Set Description.tab.tsv', sep='\t'))
data_desc
```

[ ]:

	Variable	VariableName	\
0	Dependent Variable	NaN	
1	cross_buy	Cross-buy	
2	Transaction Data	NaN	
3	calls	Calls	
4	complaints	Complaints	
5	customer_tenure_months	Customer Tenure	
6	inflows	Inflows	
7	last_acc_opening_days	Last Account	
8	logins_desktop	Desktop Logins	
9	logins_mobile	Mobile Logins	
10	nr_products	Number of Products	
11	outflows	Outflows	
12	prod_loan	Loans	
13	prod_mortgages	Mortgages	
14	prod_brokerage	Brokerage	
15	prod_pensionplan	Pension Plan	
16	prod_savings	Savings	
17	relocations	Relocations	
18	volume_debit	Total Debit	
19	volume_debit_6months	Total Debit Six Months	
20	Marketing Efforts	NaN	
21	directmails	Direct Mailing	
22	giro_mailing	Giro Mailing	
23	Customer Characteristics	NaN	
24	acad_title	Academic Title	
25	age	Age	
26	joint_account	Joint Account	
27	gender	Gender	
28	marital_status	Marital Status	
29	occupation	Occupation	
30	member_get_member_active	Get Member Active	
31	member_get_member_passive	Get Member Passive	
32	ext_city_size	City Size	
33	ext_house_size	House Size	
34	ext_purchase_power	Purchase Power	
35	ext_share_new_houses	Share of New Houses	
36	ext_share_new_cars	Share of New Cars	
37	ext_car_power	Car Power	
38	ext_living_duration	Living Duration	

	Operationalization
0	NaN
1	Customer opened a checking account: 1 (yes), 0…
2	NaN
3	Number of calls in last 180 days
4	Number of complaints in last year
5	Number of months since customer onboarding
6	Total volume of inflows on savings account fro…
7	Number of days since last account opening
8	Number of logins in the last 180 days
9	Number of mobile sessions in the last 180 days
10	Total number of products (accounts)
11	Total volume of outflows from savings account …
12	Number of consumer loan accounts
13	Number of mortgage accounts
14	Number of investment accounts
15	Number of long term savings plans
16	Number of savings accounts
17	Number of relocations/address changes in the l…
18	Total balances of all debit (savings) accounts…
19	Credit balance all of products from 6 months a…
20	NaN
21	Total number of mailing in the last year
22	Received an email about opening a checking acc…
23	NaN
24	Does the customer have an academic title: 1 (y…
25	Customer's age in years
26	Customer has a joint bank account: 1 (yes), 0 …
27	Customer's gender: 1 (male), 0 (female)
28	Customer's marital status: divorced, married, …
29	Customer's occupation: white-collar worker, se…
30	Customer recommended a customer: 1 (yes), 0 (no)
31	Customer was recommended by a customer: 1 (yes…
32	City size: 1 (< 5.000 inhabitants), 2 (5.000-1…
33	Average number of households per building in t…
34	Average purchase power in the residential bloc…
35	Share of new buildings in the residential bloc…
36	Share of new vehicle registrations in the resi…
37	Predominant vehicle category in the neighborho…
38	Average duration of residence in the customer'…

1.3 Exploratory Data Analysis

1.3.1 Initial Investigation into the Dataset and the Response Variable

[ ]:
```
# View columns, dimensions, and data types
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100000 entries, 0 to 99999
Data columns (total 35 columns):
 #   Column                      Non-Null Count   Dtype
---  ------                      --------------   -----
 0   cross_buy                   100000 non-null  int64
 1   acad_title                  100000 non-null  int64
 2   age                         100000 non-null  int64
 3   calls                       100000 non-null  int64
 4   complaints                  100000 non-null  int64
 5   customer_tenure_months      100000 non-null  int64
 6   directmails                 100000 non-null  int64
 7   gender                      99998 non-null   float64
 8   joint_account               99998 non-null   float64
 9   inflows                     99527 non-null   float64
 10  last_acc_opening_days       100000 non-null  int64
 11  logins_desktop              100000 non-null  int64
 12  logins_mobile               100000 non-null  int64
 13  marital_status              100000 non-null  object
 14  member_get_member_active    100000 non-null  int64
 15  member_get_member_passive   100000 non-null  int64
 16  nr_products                 100000 non-null  int64
 17  occupation                  48725 non-null   object
 18  outflows                    99527 non-null   float64
 19  prod_loan                   100000 non-null  int64
 20  prod_mortgages              100000 non-null  int64
 21  prod_brokerage              100000 non-null  int64
 22  prod_pensionplan            100000 non-null  int64
 23  prod_savings                100000 non-null  int64
 24  relocations                 100000 non-null  int64
 25  volume_debit                100000 non-null  float64
 26  volume_debit_6months        97002 non-null   float64
 27  ext_city_size               97338 non-null   float64
 28  ext_house_size              96953 non-null   float64
 29  ext_purchase_power          95435 non-null   float64
 30  ext_share_new_houses        97338 non-null   float64
 31  ext_share_new_cars          84845 non-null   float64
 32  ext_car_power               89866 non-null   float64
 33  ext_living_duration         90935 non-null   float64
 34  giro_mailing                100000 non-null  int64
```

```
dtypes: float64(13), int64(20), object(2)
memory usage: 26.7+ MB
```

[ ]:
```
# The dimension of the dataset
print('Number of Rows:', df.shape[0])
print('Number of Columns:', df.shape[1])
```

```
Number of Rows: 100000
Number of Columns: 35
```

[ ]:
```
# check for missing values
df.isnull().sum()
```

[ ]:
```
cross_buy                       0
acad_title                      0
age                             0
calls                           0
complaints                      0
customer_tenure_months          0
directmails                     0
gender                          2
joint_account                   2
inflows                       473
last_acc_opening_days           0
logins_desktop                  0
logins_mobile                   0
marital_status                  0
member_get_member_active        0
member_get_member_passive       0
nr_products                     0
occupation                  51275
outflows                      473
prod_loan                       0
prod_mortgages                  0
prod_brokerage                  0
prod_pensionplan                0
prod_savings                    0
relocations                     0
volume_debit                    0
volume_debit_6months         2998
ext_city_size                2662
ext_house_size               3047
ext_purchase_power           4565
ext_share_new_houses         2662
ext_share_new_cars          15155
ext_car_power               10134
ext_living_duration          9065
```

```
giro_mailing                    0
dtype: int64
```

Response variable - cross_buy

```
[ ]: # Check Target Class Distribution
     # 0 = Customers didn't open a checking account
     # 1 = Customers opened a checking account
     df['cross_buy'].value_counts()
```

```
[ ]: 0    90000
     1    10000
     Name: cross_buy, dtype: int64
```

```
[ ]: # Plot Class Distribution
     sns.countplot(data = df, x = 'cross_buy')
     plt.xlabel('cross_buy')
     plt.ylabel('Frequency')
     plt.title('cross_buy Distribution')
     plt.show()
```

"cross_buy" tells us if an existing customer opened a checking account. From here we see that 10,000 customers out of the 10,000 customers did.

For this problem, the class imbalancement might have an effect since 90% of the data are in one class and there aren't enough negative and positive classes for training. We should resample the data so that more customers would seemed to open the bank account.

Note, the cross buy rate for the existing customers is 10%. So our baseline suggests randomly selecting customers to contact for checking accounts will result in a 10% success rate.

1.3.2 Investigation in other Independent Variables

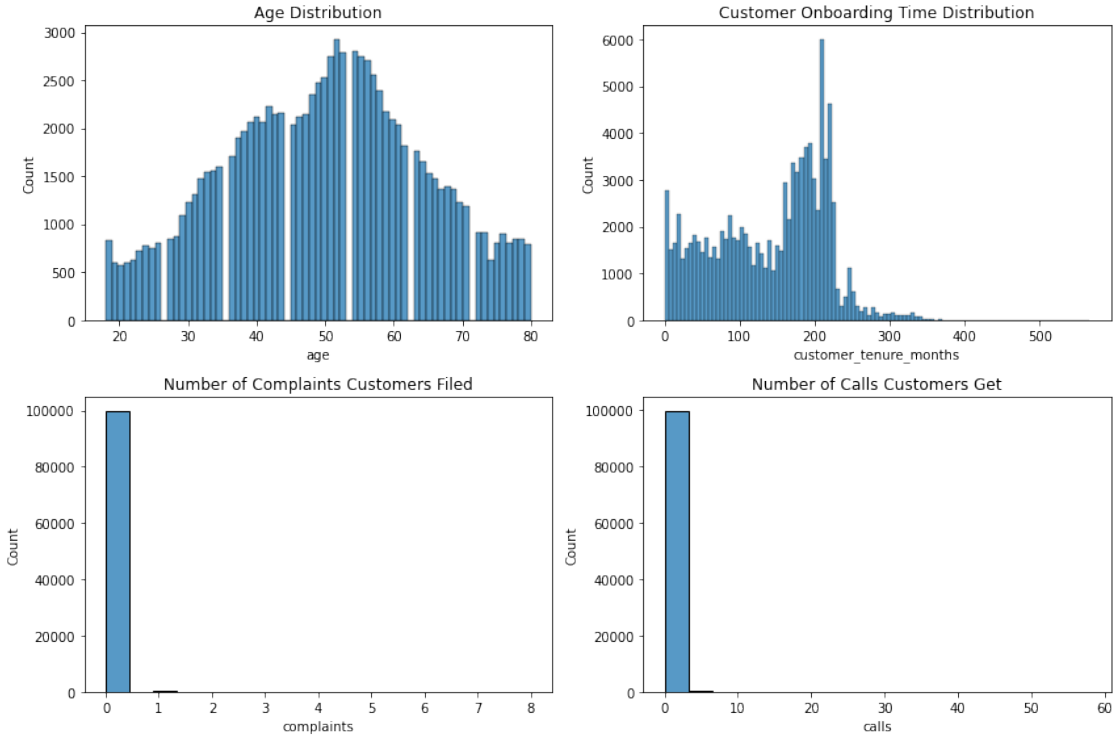
Selected Numerical features

```
[ ]: # Subsetting the selected numerical features into a dataset
     num_features = df[['age','calls', 'complaints', 'customer_tenure_months',␣
      ↪'nr_products', 'directmails']]

     # Retunr decription of the numerical features
     num_features.describe()
```

```
[ ]:                 age          calls     complaints  customer_tenure_months  \
     count  100000.000000  100000.000000  100000.000000           100000.000000
     mean       49.892260       0.104600       0.003530              140.181700
     std        14.534085       0.564395       0.078598               74.901654
```

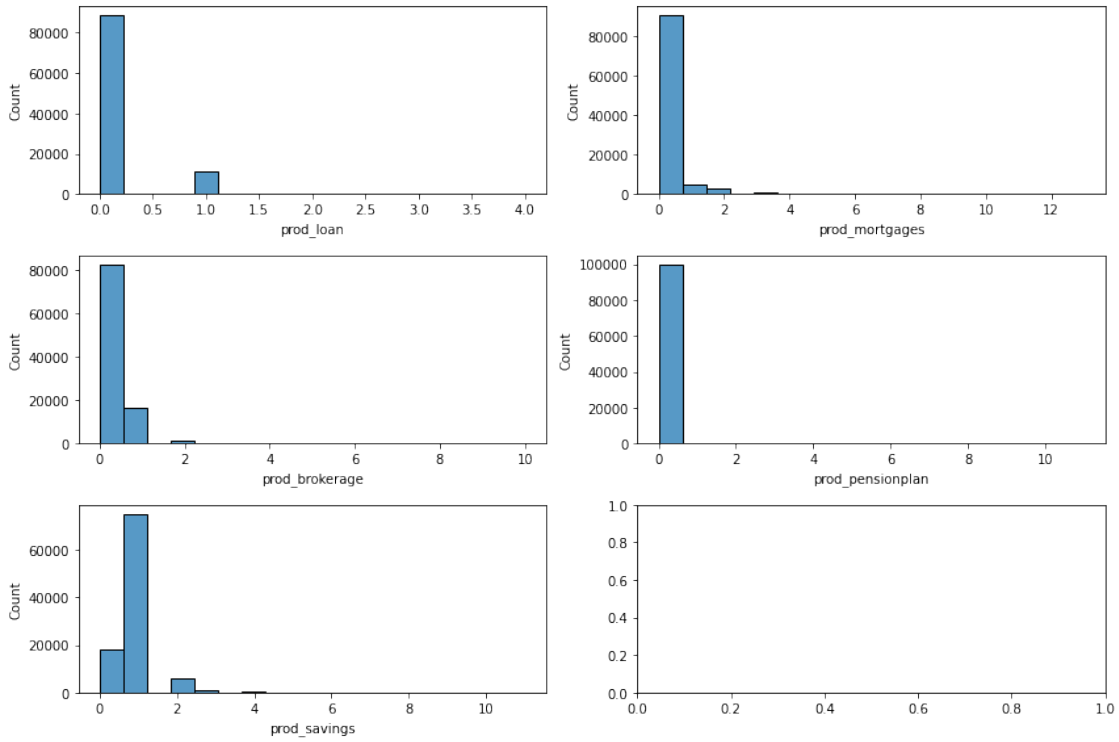
```
min        18.000000       0.000000       0.000000                0.000000
25%        39.000000       0.000000       0.000000               78.000000
50%        51.000000       0.000000       0.000000              160.000000
75%        60.000000       0.000000       0.000000              201.000000
max        80.000000      58.000000       8.000000              567.000000

         nr_products    directmails
count  100000.000000  100000.000000
mean        1.433380       0.470660
std         0.798579       0.915952
min         1.000000       0.000000
25%         1.000000       0.000000
50%         1.000000       0.000000
75%         2.000000       1.000000
max        17.000000       9.000000
```

[ ]:
```
# Plot Features
%matplotlib inline
f, axs = plt.subplots(2, 2, figsize = (12, 8))

# Customer's age in years
sns.histplot(data = df, x = 'age', ax = axs[0,0])

# Number of months since customer onboarding
sns.histplot(data = df, x = 'customer_tenure_months', ax = axs[0,1])

# Number of complaints in the last year
sns.histplot(data = df, x = 'complaints', ax = axs[1,0])

# Number of calls customers get in the last 180 days
sns.histplot(data = df, x = 'calls', ax = axs[1,1])

axs[0, 0].title.set_text("Age Distribution")
axs[0, 1].title.set_text("Customer Onboarding Time Distribution")
axs[1, 0].title.set_text("Number of Complaints Customers Filed")
axs[1, 1].title.set_text("Number of Calls Customers Get")

plt.tight_layout()
```



Financial Products that customers own:

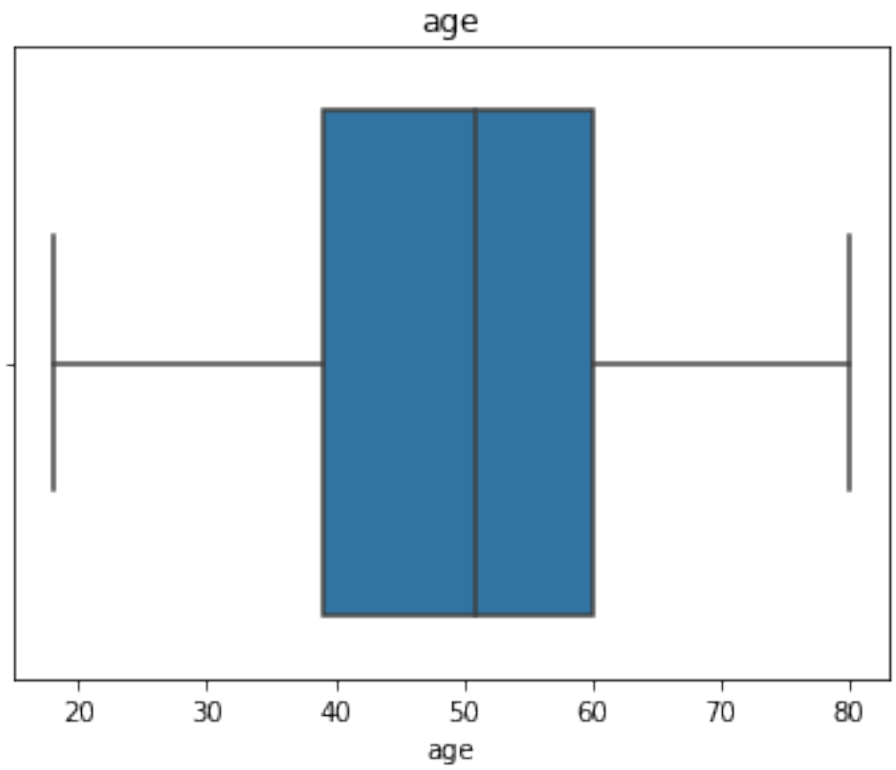
[ ]:
```
# Plot the financial products that consumers have
%matplotlib inline

f, axs = plt.subplots(3, 2, figsize = (12, 8))

# Distribution of number of loan accounts consumers have
sns.histplot(data = df, x = 'prod_loan', ax = axs[0,0])

# Distribution of number of mortgage accounts consumers have
sns.histplot(data = df, x = 'prod_mortgages', ax = axs[0,1])

# Distribution of number of investment accounts consumers have
sns.histplot(data = df, x = 'prod_brokerage', ax = axs[1,0])

# Distribution of number of long-term savings accounts consumers have
sns.histplot(data = df, x = 'prod_pensionplan', ax = axs[1,1])

# Distribution of number of savings accounts consumers have
sns.histplot(data = df, x = 'prod_savings', ax = axs[2,0])

plt.tight_layout()
```

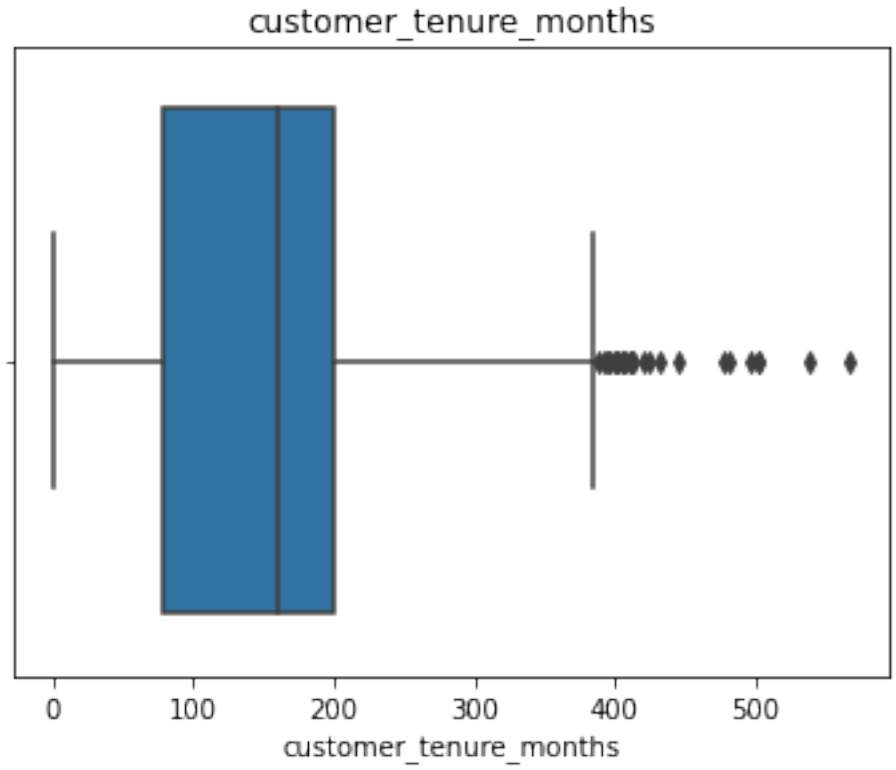


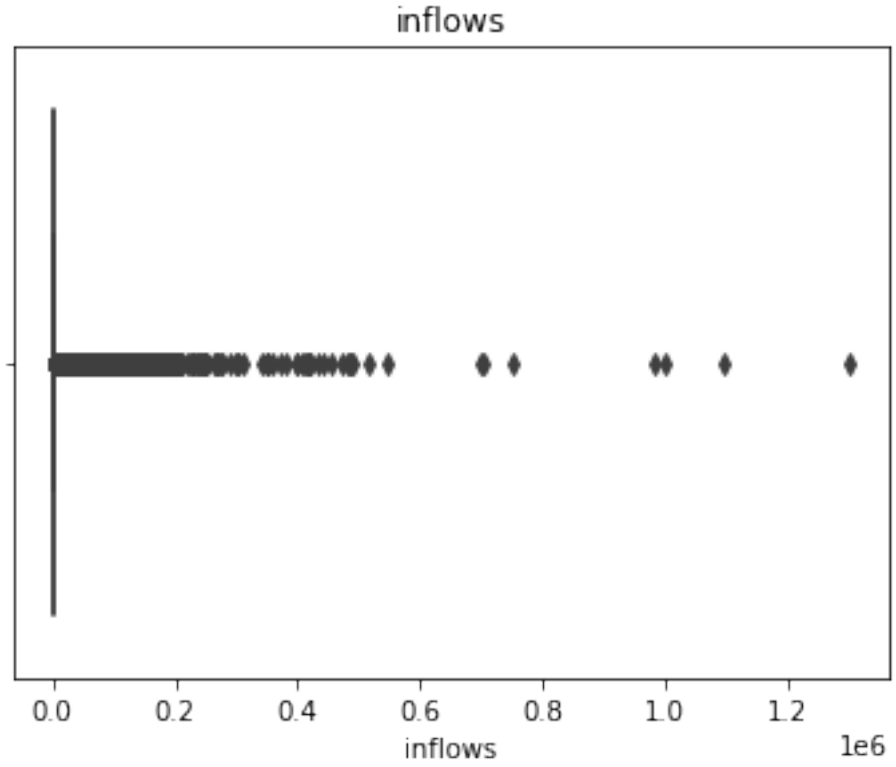
We can see that most customers own a savings account but not the other financial accounts.

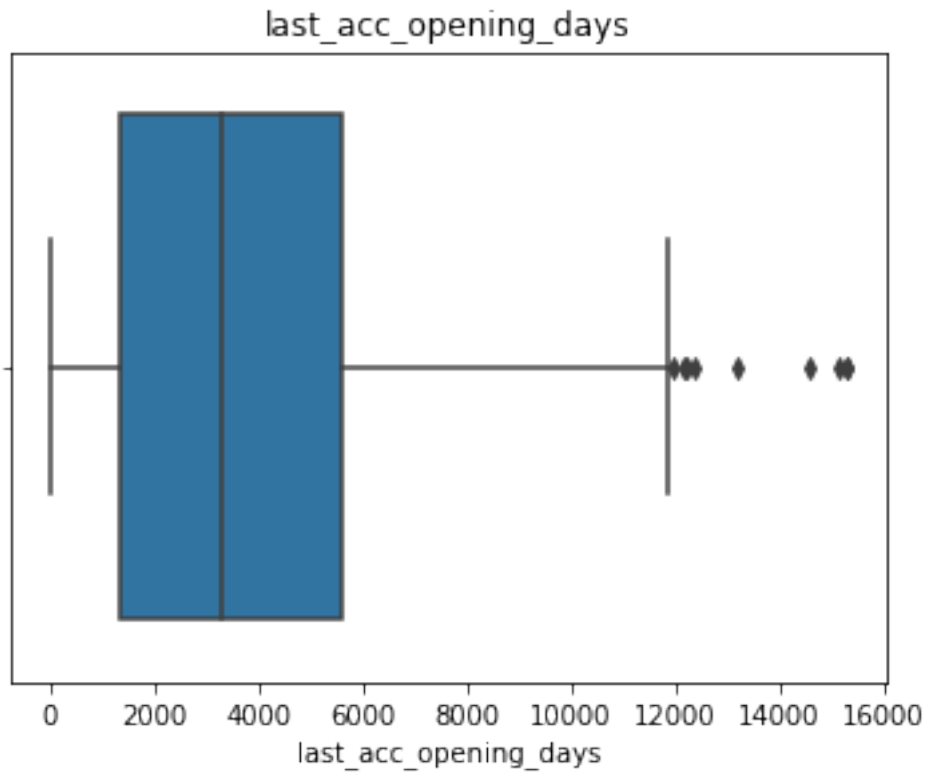
1.3.3 Visualizing Outliers

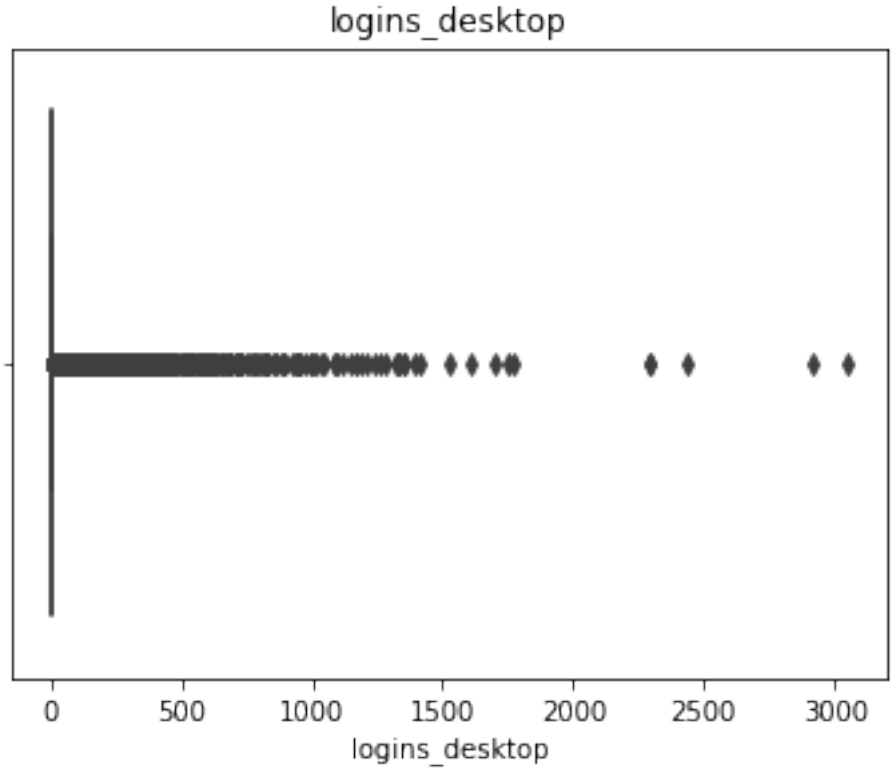
[ ]:
```
# Visualizing each variables outlier
col = df[['age','customer_tenure_months','inflows',
                'last_acc_opening_days','logins_desktop',
                'outflows','volume_debit','volume_debit_6months']]

for i in col:
  n=1
  plt.figure(figsize=(20,20))
  plt.subplot(4,3,1)
  sns.boxplot(df[i])
  plt.title(i)
  plt.show()
  n=n+1
```

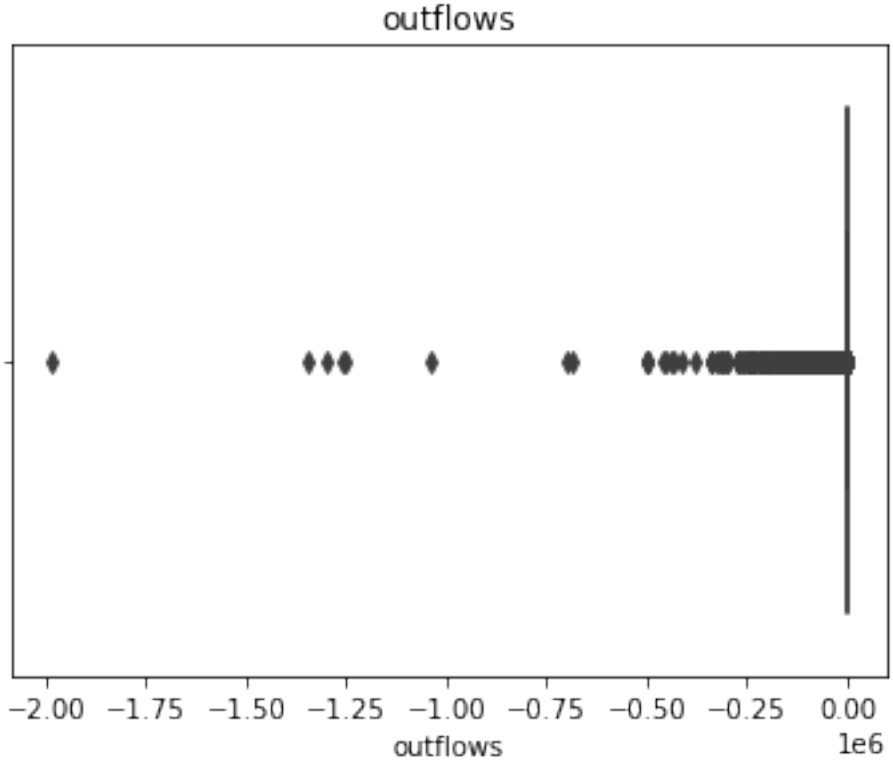

age
20
30
40
50
60
70
80
age

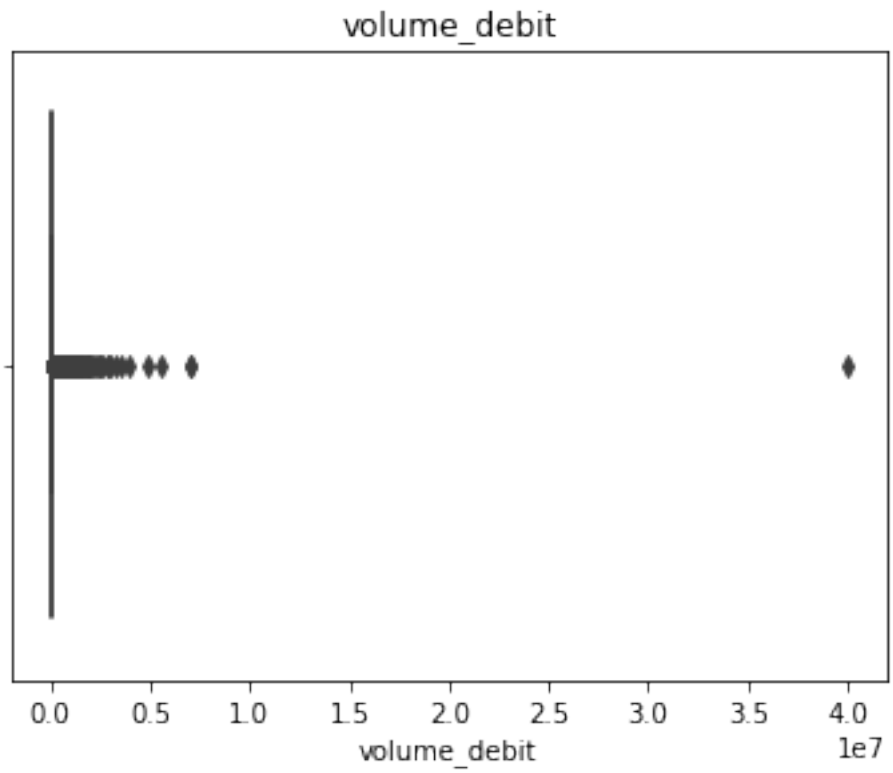

customer_tenure_months
0
100
200
300
400
500
customer_tenure_months

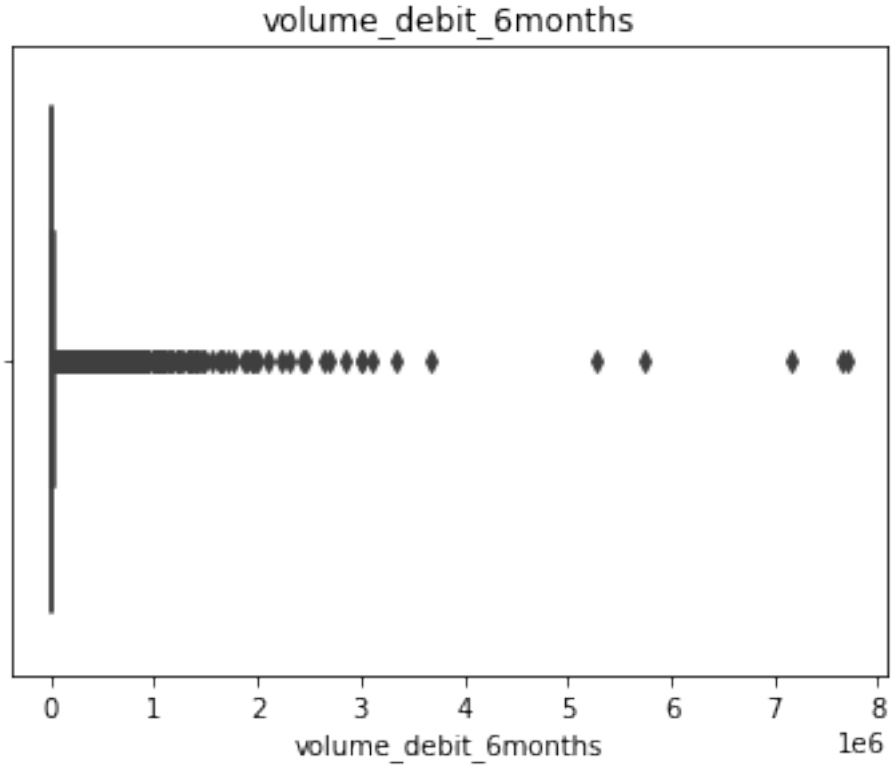

inflows
0.0
0.2
0.4
0.6
0.8
1.0
1.2
inflows
1e6


last_acc_opening_days
0
2000
4000
6000
8000
10000
12000
14000
16000
last_acc_opening_days


logins_desktop
0
500
1000
1500
2000
2500
3000
logins_desktop

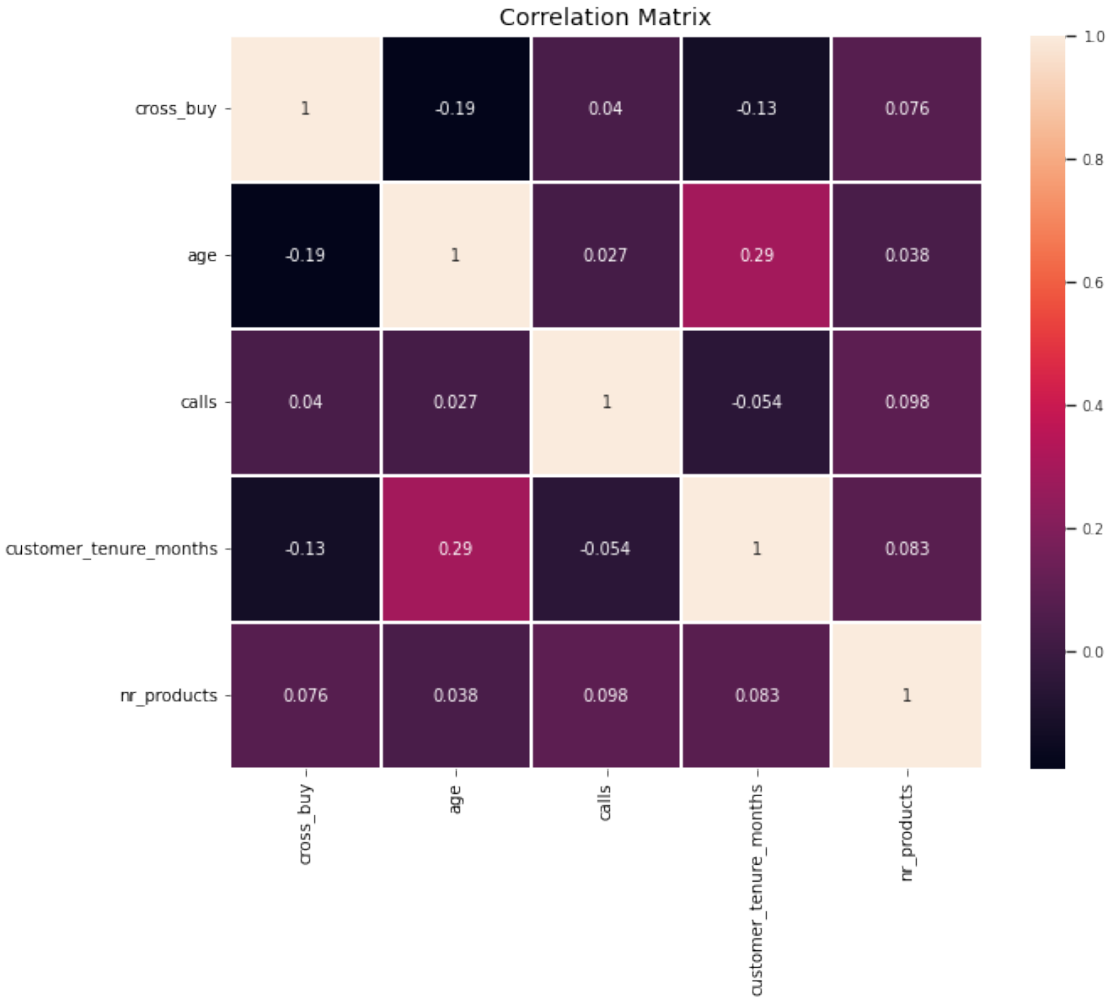

outflows
−2.00
−1.75
−1.50
−1.25
−1.00
−0.75
−0.50
−0.25
0.00
outflows
1e6


volume_debit
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
volume_debit
1e7


volume_debit_6months
0
1
2
3
4
5
6
7
8
volume_debit_6months
1e6

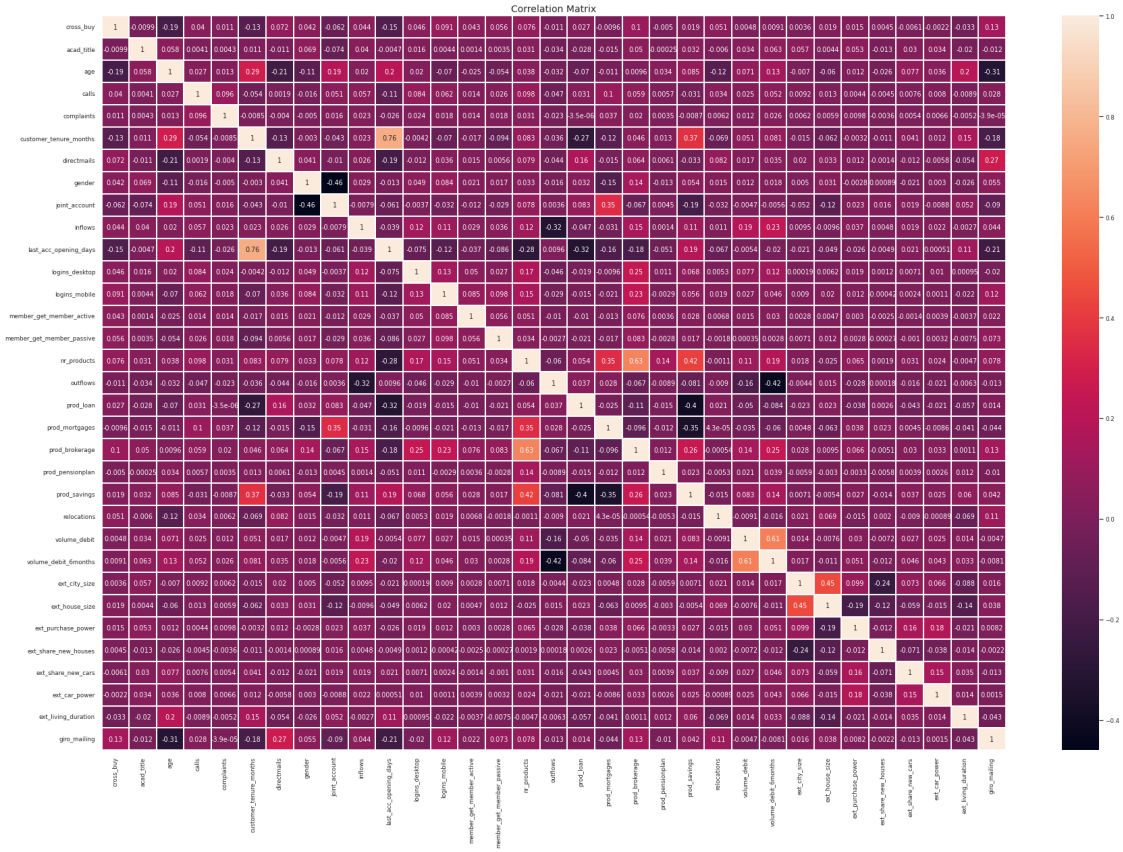
1.3.4 Correlation Matrix

```
[ ]: # Selecting more relevant variables to compute the correlation matrix
df_corr = df[['cross_buy','age', 'calls', 'customer_tenure_months',␣
 ↪'nr_products']]

corrMatrix = df_corr.corr()
plt.subplots(figsize = (10,8))
sns.set(font_scale = .8)
sns.heatmap(corrMatrix, annot = True, linewidths = 1)
plt.title(r'Correlation Matrix', fontsize = 14)
plt.savefig('Correlation Matrix')
```



[ ]:
```
corrMatrix = df.corr()
plt.subplots(figsize = (30,20))
sns.set(font_scale = .8)
sns.heatmap(corrMatrix, annot = True, linewidths = 1)
plt.title(r'Correlation Matrix', fontsize = 14)
```

[ ]:
```
Text(0.5, 1.0, 'Correlation Matrix')
```



1.4 Pre-processing Data

[ ]:
```
# Check for missing values
df.isnull().sum()
```

[ ]:
```
cross_buy                    0
acad_title                   0
age                          0
calls                        0
complaints                   0
customer_tenure_months       0
```

```
directmails                     0
gender                          2
joint_account                   2
inflows                       473
last_acc_opening_days           0
logins_desktop                  0
logins_mobile                   0
marital_status                  0
member_get_member_active        0
member_get_member_passive       0
nr_products                     0
occupation                  51275
outflows                      473
prod_loan                       0
prod_mortgages                  0
prod_brokerage                  0
prod_pensionplan                0
prod_savings                    0
relocations                     0
volume_debit                    0
volume_debit_6months         2998
ext_city_size                2662
ext_house_size               3047
ext_purchase_power           4565
ext_share_new_houses         2662
ext_share_new_cars          15155
ext_car_power               10134
ext_living_duration          9065
giro_mailing                    0
dtype: int64
```

[ ]:
```
# Columns including occupationshould be dropped since it contains mostly NaN␣
 ↪values and
# won't provide a lot of meaningful information.
df = df.drop(columns=['occupation'])
```

The curse of dimensionality means that the error increases with the increase in the number of features. Since we have 35 features, there is a need to select the important features and remove the irrelevant ones for better performance of the model.

Regarding account balances, we have four features that describes the account balances: * inflows - It describes the total volume of inflows on savings account * outflows - It describes the total volume of outflows on savings account * volume_debit - It describes the total balances of all debit (savings) accounts * volume_debit_6months - It describes the total balances of all debit (savings) accounts in six months

Since we are interested in knowing the account balances that customers have which will have a relationship with whether a customer will open a checking account, we will use the most relevant

feature which is volume_debit, and drop the other features.

Regarding the days since the account opened, we have features that described it: * customer_tenure_months - Number of months since customer onboarding * last_acc_opening_days - Number of days since last account opening

We selected the customer_tenure_months since that gives us a better idea of the time customers are with the bank. Thus, we will remove the other feature.

[ ]:
```
# Since we will be using the total balances of all accounts, we can drop the␣
↪redundant features
df = df.drop(columns=['inflows','outflows','volume_debit_6months'])

# Also dropping duplicate feature last_acc_opening_days
df = df.drop(columns=['last_acc_opening_days'])
```

As mentioned above, most customers own 0 financial products except savings account, we would drop the irrelevant features and keep the prod_savings feature.

[ ]:
```
# Dropping the features of financial products that majority of consumers do not␣
↪have since they are irrelevant
df = df.
↪drop(columns=['prod_loan','prod_mortgages','prod_brokerage','prod_pensionplan'])
```

Looking atthe external features including the following: * ext_city_size - City size * ext_house_size - Average number of households per building in the residential block * ext_purchase_power - Average purchase power in the residential block * ext_share_new_houses - Share of new buildings in the residential block * ext_share_new_cars - Share of new vehicle registrations in the residential block * ext_car_power - Predominant vehicle category in the neighborhood * ext_living_duration - Average duration of residence in the customer's building

The only features that are more relevant features to whether a customer will open a checking account will be the purchase power pf clients. The other features including city size, average number of households, etc. is a lot less relevant in impacting someone to open a checking account. Thus, we will remove the irrelevant features.

[ ]:
```
# Dropping irrelevant external factors that are less likely to affect if the␣
↪consumer will open a checking account
df = df.
↪drop(columns=['ext_city_size','ext_house_size','ext_share_new_houses','ext_share_new_cars',
```

Lastly, there are some other less irrelevant features that we consider removing prior to the training of model: * relocations - Number of relocations/address changes in the last year * acad_title - Whether the customer have an academic title * logins_desktop - Number of logins in the last 180 days * logins_mobile - Number of mobile sessions in the last 180 days

[ ]:
```
# Dropping other less irrelevant factors to finalize the dataset for modeling
df = df.
↪drop(columns=['relocations','acad_title','logins_desktop','logins_mobile','directmails'])
```

[ ]:
```
df
```

[ ]:
```
       cross_buy  age  calls  complaints  customer_tenure_months  gender  \
0              0   60      0           0                     221     0.0
1              0   55      0           0                     227     0.0
2              0   61      0           0                     221     1.0
3              0   70      0           0                     222     0.0
4              0   61      0           0                     227     1.0
…            …  …    …         …                     …      …
99995          1   53      0           0                     206     0.0
99996          1   25      0           0                     206     0.0
99997          1   19      0           0                     205     1.0
99998          1   58      0           0                     204     1.0
99999          1   55      0           0                     205     0.0

       joint_account marital_status  member_get_member_active  \
0                0.0       divorced                         0
1                1.0        married                         0
2                0.0        married                         0
3                0.0        married                         0
4                0.0        married                         0
…              …            …                      …
99995            0.0        married                         0
99996            0.0         single                         0
99997            0.0         single                         0
99998            0.0        married                         0
99999            1.0        married                         1

       member_get_member_passive  nr_products  prod_savings  volume_debit  \
0                              0            2             0          0.00
1                              0            1             1          3.28
2                              0            1             1      31963.13
3                              0            1             1      54048.40
4                              0            2             1    1374743.09
…                            …          …           …           …
99995                          1            2             1      28839.32
99996                          0            1             1       2133.01
99997                          0            2             1      47476.22
99998                          0            1             0       1573.06
99999                          0            3             1      39503.93

       ext_purchase_power  giro_mailing
0                     5.0             0
1                     7.0             0
2                     4.0             0
3                     7.0             0
4                     7.0             0
```

```
…                    …            …
99995                5.0          0
99996                3.0          0
99997                5.0          0
99998                7.0          0
99999                3.0          0

[100000 rows x 15 columns]
```

[ ]:
```
# Converting categorical variables to dummies

# Treat marital_status and ext_purchase_power as categorical, then convert to␣
 ↪dummy variables
df['marital_status'] = df['marital_status'].astype('category')

df = pd.get_dummies(df, columns=['marital_status'], drop_first=True)
df.head(20)
```

[ ]:

	cross_buy	age	calls	complaints	customer_tenure_months	gender	\
0	0	60	0	0	221	0.0	
1	0	55	0	0	227	0.0	
2	0	61	0	0	221	1.0	
3	0	70	0	0	222	0.0	
4	0	61	0	0	227	1.0	
5	0	67	0	0	227	0.0	
6	0	55	0	0	227	0.0	
7	0	78	0	0	227	0.0	
8	0	58	0	0	227	1.0	
9	0	47	0	0	221	1.0	
10	0	65	0	0	221	0.0	
11	0	56	0	0	227	1.0	
12	0	63	0	0	222	0.0	
13	0	54	0	0	227	1.0	
14	0	52	0	0	227	1.0	
15	0	70	0	0	134	0.0	
16	0	52	0	0	221	1.0	
17	0	59	0	0	227	1.0	
18	0	70	0	0	227	0.0	
19	0	47	0	0	225	1.0	

	joint_account	member_get_member_active	member_get_member_passive	\
0	0.0	0	0	
1	1.0	0	0	
2	0.0	0	0	
3	0.0	0	0	
4	0.0	0	0	
5	1.0	0	0	

6	1.0	0	0
7	1.0	0	0
8	0.0	0	0
9	0.0	0	0
10	1.0	0	0
11	0.0	0	0
12	0.0	0	0
13	0.0	0	0
14	0.0	0	0
15	0.0	0	0
16	0.0	0	0
17	0.0	0	0
18	1.0	0	0
19	0.0	0	0

	nr_products	prod_savings	volume_debit	ext_purchase_power	giro_mailing	\
0	2	0	0.00	5.0	0	
1	1	1	3.28	7.0	0	
2	1	1	31963.13	4.0	0	
3	1	1	54048.40	7.0	0	
4	2	1	1374743.09	7.0	0	
5	3	2	34673.54	1.0	0	
6	1	1	60209.87	2.0	0	
7	3	2	84357.51	6.0	0	
8	1	1	13868.59	7.0	0	
9	1	1	12.26	5.0	0	
10	3	2	127494.49	2.0	0	
11	7	6	8169.64	NaN	0	
12	1	1	20194.16	1.0	0	
13	2	1	147749.83	7.0	0	
14	2	1	27578.02	6.0	0	
15	1	0	0.00	3.0	0	
16	1	1	549.13	7.0	0	
17	1	1	77.68	7.0	0	
18	2	2	225259.12	6.0	0	
19	1	1	70180.59	NaN	0	

	marital_status_divorced	marital_status_married	marital_status_separated	\
0	1	0	0	
1	0	1	0	
2	0	1	0	
3	0	1	0	
4	0	1	0	
5	0	1	0	
6	0	1	0	
7	0	1	0	
8	0	0	0	

```
9                          0                          1                          0
10                         0                          1                          0
11                         0                          1                          0
12                         0                          0                          0
13                         0                          0                          0
14                         0                          0                          0
15                         0                          0                          0
16                         0                          1                          0
17                         0                          0                          0
18                         0                          1                          0
19                         0                          0                          0

    marital_status_single  marital_status_unmarried  marital_status_widowed
0                          0                          0                          0
1                          0                          0                          0
2                          0                          0                          0
3                          0                          0                          0
4                          0                          0                          0
5                          0                          0                          0
6                          0                          0                          0
7                          0                          0                          0
8                          1                          0                          0
9                          0                          0                          0
10                         0                          0                          0
11                         0                          0                          0
12                         1                          0                          0
13                         1                          0                          0
14                         1                          0                          0
15                         0                          1                          0
16                         0                          0                          0
17                         1                          0                          0
18                         0                          0                          0
19                         1                          0                          0
```

[ ]:
```
df['ext_purchase_power'].value_counts()
```

[ ]:
```
7.0    22197
6.0    16750
5.0    14529
4.0    12370
3.0    10802
1.0     9491
2.0     9296
Name: ext_purchase_power, dtype: int64
```

[ ]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100000 entries, 0 to 99999
Data columns (total 20 columns):
 #   Column                      Non-Null Count   Dtype
---  ------                      --------------   -----
 0   cross_buy                   100000 non-null  int64
 1   age                         100000 non-null  int64
 2   calls                       100000 non-null  int64
 3   complaints                  100000 non-null  int64
 4   customer_tenure_months      100000 non-null  int64
 5   gender                      99998 non-null   float64
 6   joint_account               99998 non-null   float64
 7   member_get_member_active    100000 non-null  int64
 8   member_get_member_passive   100000 non-null  int64
 9   nr_products                 100000 non-null  int64
 10  prod_savings                100000 non-null  int64
 11  volume_debit                100000 non-null  float64
 12  ext_purchase_power          95435 non-null   float64
 13  giro_mailing                100000 non-null  int64
 14  marital_status_divorced     100000 non-null  uint8
 15  marital_status_married      100000 non-null  uint8
 16  marital_status_separated    100000 non-null  uint8
 17  marital_status_single       100000 non-null  uint8
 18  marital_status_unmarried    100000 non-null  uint8
 19  marital_status_widowed      100000 non-null  uint8
dtypes: float64(4), int64(10), uint8(6)
memory usage: 11.3 MB
```

[ ]:
```
# Check for missing values again
df.isnull().sum()
```

[ ]:
```
cross_buy                     0
age                           0
calls                         0
complaints                    0
customer_tenure_months        0
gender                        2
joint_account                 2
member_get_member_active      0
member_get_member_passive     0
nr_products                   0
prod_savings                  0
volume_debit                  0
ext_purchase_power         4565
giro_mailing                  0
marital_status_divorced       0
marital_status_married        0
```

```
marital_status_separated        0
marital_status_single           0
marital_status_unmarried        0
marital_status_widowed          0
dtype: int64
```

[ ]:
```
# imputation of missing data with mean imputer
mean_imputer = SimpleImputer(strategy='mean')
df[['gender','joint_account']] = pd.DataFrame(mean_imputer.
 ↪fit_transform(df[['gender','joint_account']]))
```

[ ]:
```
# imputation of missing data with mode imputer
mode_imputer = SimpleImputer(strategy='most_frequent')
df[['ext_purchase_power']] = pd.DataFrame(mode_imputer.
 ↪fit_transform(df[['ext_purchase_power']]))
```

[ ]:
```
df.isnull().sum()
```

[ ]:
```
cross_buy                   0
age                         0
calls                       0
complaints                  0
customer_tenure_months      0
gender                      0
joint_account               0
member_get_member_active    0
member_get_member_passive   0
nr_products                 0
prod_savings                0
volume_debit                0
ext_purchase_power          0
giro_mailing                0
marital_status_divorced     0
marital_status_married      0
marital_status_separated    0
marital_status_single       0
marital_status_unmarried    0
marital_status_widowed      0
dtype: int64
```

[ ]:
```
# Partition the data into training (60%) and validation (40%). Use seed = 1.
y = df['cross_buy']
X = df.drop(columns=['cross_buy'])

# Resampling the cross_buy variables using over-sampling techniques
nm = NearMiss()
X_res, y_res = nm.fit_resample(X, y)
```

[ ]:
```
# Split the data in training and valid data
X_train, X_valid, y_train, y_valid = train_test_split(X_res, y_res,␣
 ↪train_size=0.4, random_state=1)
```

[ ]:
```
# Scaling the variables
norm = MinMaxScaler().fit(X_train)
X_train = norm.transform(X_train)
X_valid = norm.transform(X_valid)
```

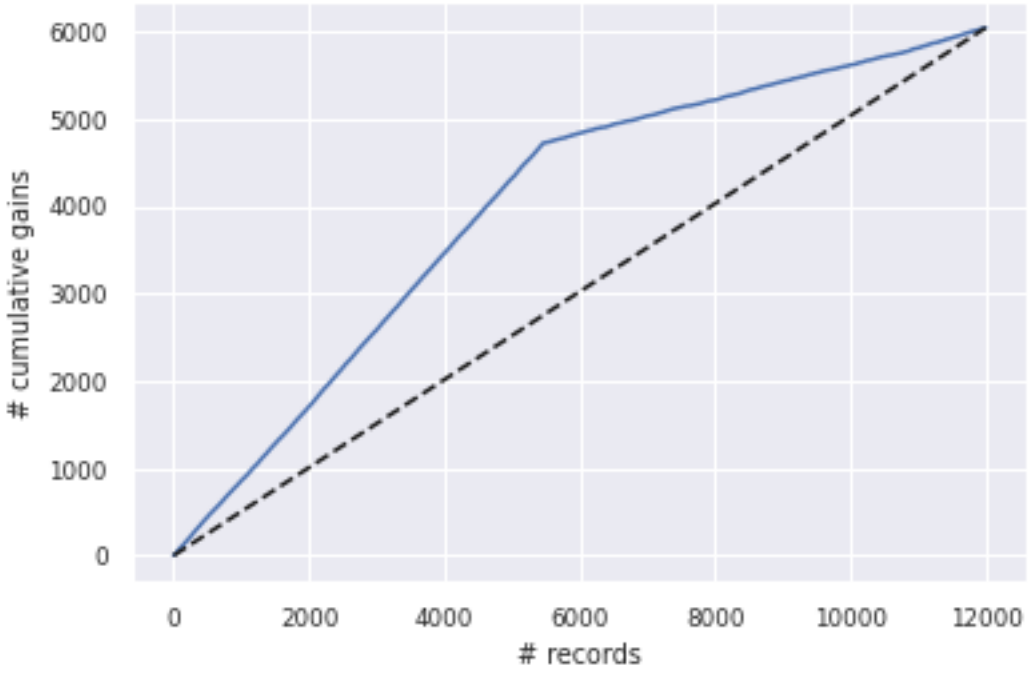
1.5 Modeling

1.5.1 Logistic Regression Model

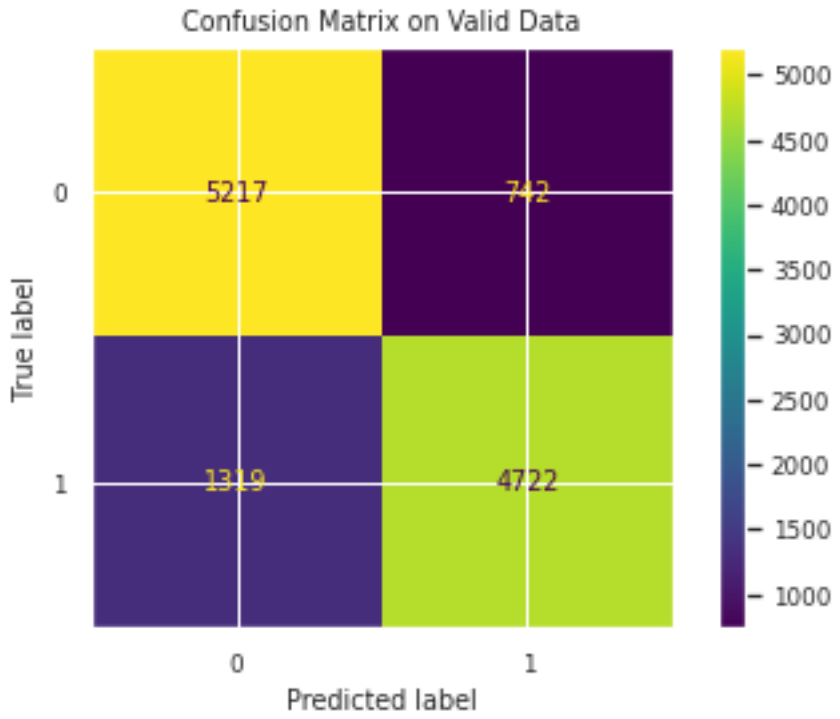
[ ]:
```
# Training logistic regression model
model = LogisticRegression()
grid = {"C": np.logspace(-3,3,7),
      "penalty":["l1", "l2"],# l1 lasso l2 ridge
      "solver": ['liblinear']}

logreg_cv = GridSearchCV(model, grid, cv=10)
logreg_cv.fit(X_train, y_train)

print("tuned hpyerparameters :(best parameters) ", logreg_cv.best_params_)
print("accuracy :",logreg_cv.best_score_)
log_pred =  logreg_cv.predict(X_valid)

# Evaluation
print(classification_report(y_valid,log_pred))

# Plot the cumulative gains chart of the expected spending
gains_df = pd.DataFrame({'actual': y_valid,
                         'prob': log_pred})
gains_df = gains_df.sort_values(by=['prob'], ascending=False).
 ↪reset_index(drop=True)
gainsChart(gains_df.actual)
plt.show()
```

```
tuned hpyerparameters :(best parameters)  {'C': 10.0, 'penalty': 'l1', 'solver':
'liblinear'}
accuracy : 0.82125
              precision    recall  f1-score   support

           0       0.80      0.88      0.84      5959
           1       0.86      0.78      0.82      6041

    accuracy                           0.83     12000
```

```
   macro avg       0.83      0.83      0.83     12000
weighted avg       0.83      0.83      0.83     12000
```



[ ]:
```
# Plot Confusion Matrix
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(logreg_cv, X_valid, y_valid)
plt.title('Confusion Matrix on Valid Data')
plt.show()
X_train2, X_valid2, y_train2, y_valid2 = train_test_split(X_res, y_res,␣
↪train_size=0.4, random_state=1)
```

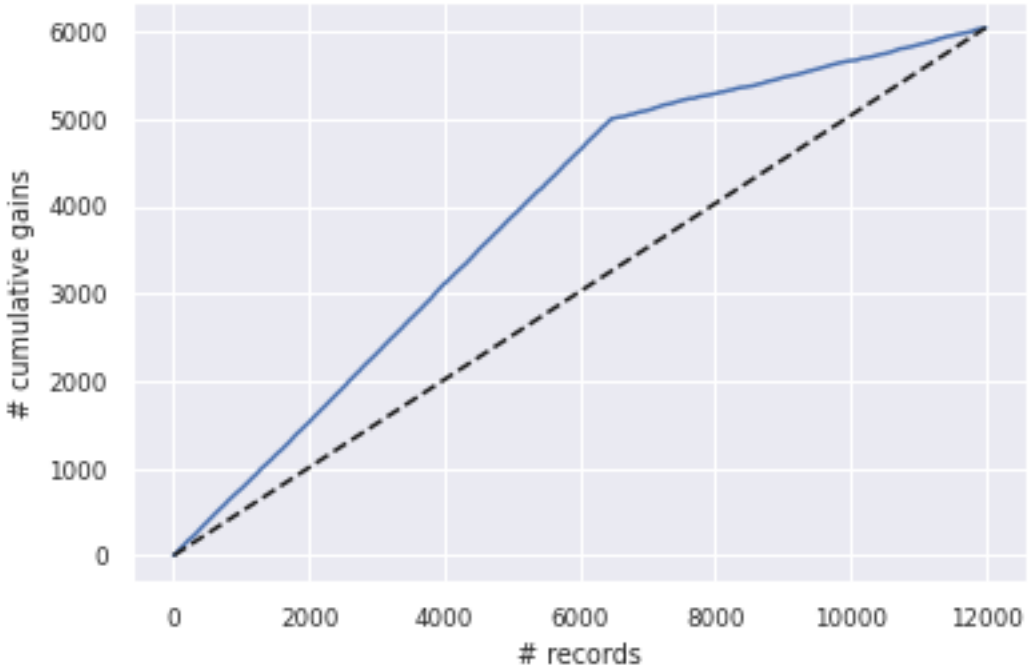


Support Vector Machine

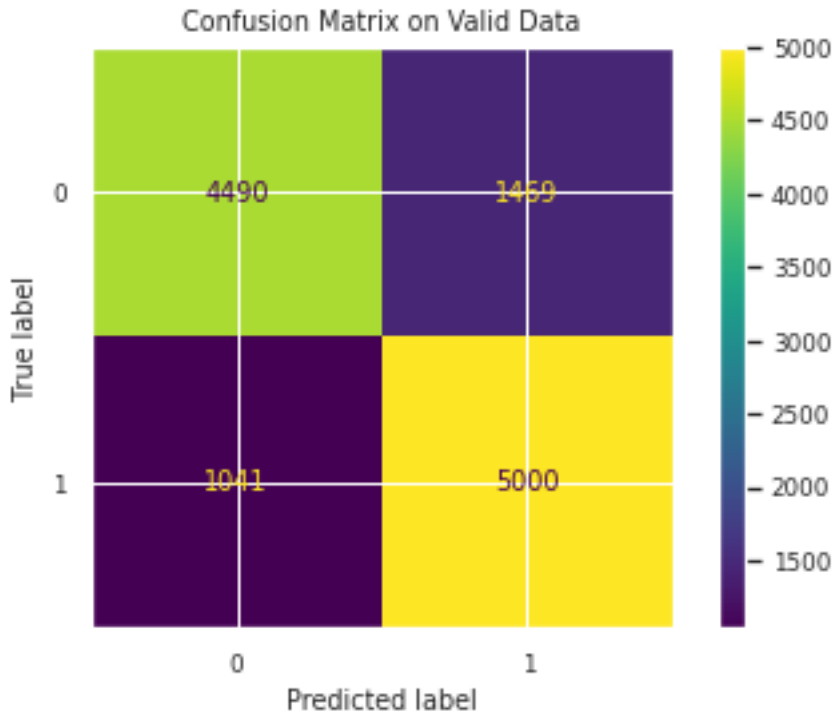
[ ]:
```
from sklearn.svm import SVC
svclassifier = SVC(kernel='linear')
svclassifier.fit(X_train, y_train)

#Making Predictions
Svm_pred = svclassifier.predict(X_valid)

#Evaluating the Algorithm
classificationSummary(y_valid, svclassifier.predict(X_valid))
print(classification_report(y_valid,Svm_pred))

# Plot the cumulative gains chart of the expected spending
gains_df = pd.DataFrame({'actual': y_valid,
                         'prob': Svm_pred})
gains_df = gains_df.sort_values(by=['prob'], ascending=False).
 ↪reset_index(drop=True)
gainsChart(gains_df.actual)
plt.show()
```

```
Confusion Matrix (Accuracy 0.7908)

       Prediction
```

```
Actual    0    1
     0 4490 1469
     1 1041 5000
              precision    recall  f1-score   support

           0       0.81      0.75      0.78      5959
           1       0.77      0.83      0.80      6041

    accuracy                           0.79     12000
   macro avg       0.79      0.79      0.79     12000
weighted avg       0.79      0.79      0.79     12000
```



[ ]:
```
# Plot Confusion Matrix
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(svclassifier, X_valid, y_valid)
plt.title('Confusion Matrix on Valid Data')
plt.show()
```

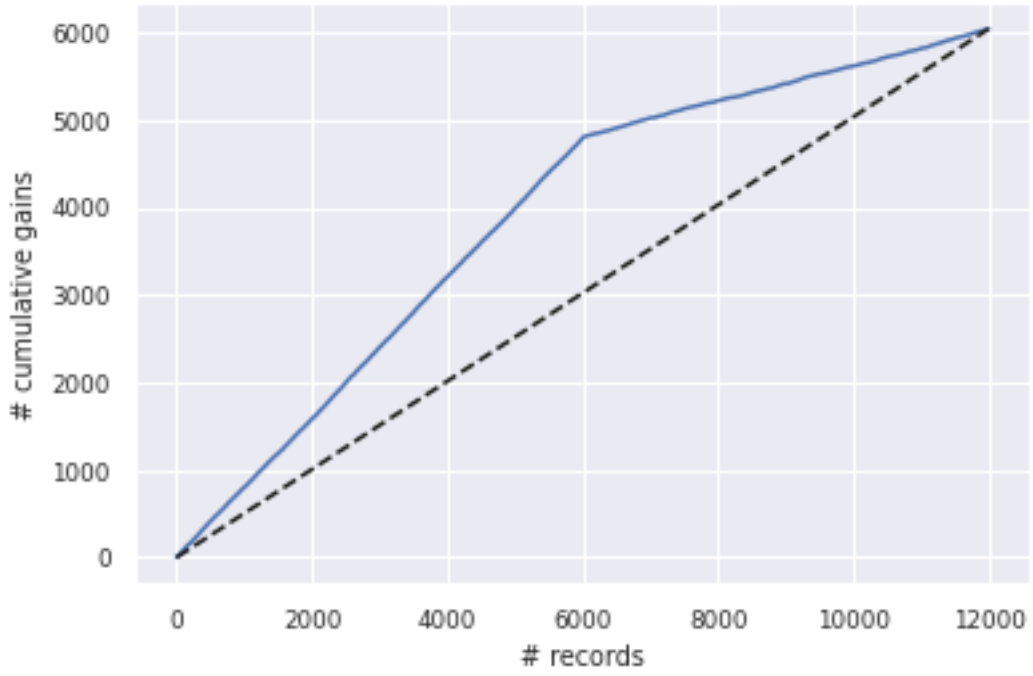


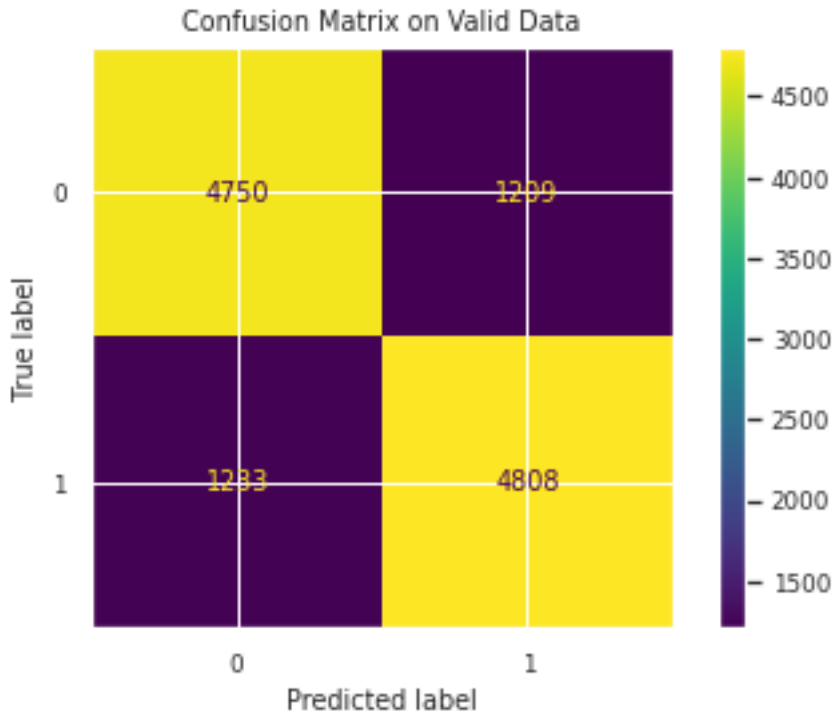
Linear Discriminant Analysis

[ ]:
```
from sklearn.model_selection import RepeatedStratifiedKFold
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis

model = LinearDiscriminantAnalysis()
model.fit(X_train, y_train)

classificationSummary(y_valid, model.predict(X_valid))
print(classification_report(y_valid, model.predict(X_valid)))


# Plot the cumulative gains chart of the expected spending
gains_df = pd.DataFrame({'actual': y_valid,
                         'prob': model.predict(X_valid)})
gains_df = gains_df.sort_values(by=['prob'], ascending=False).
 ↪reset_index(drop=True)
gainsChart(gains_df.actual)
plt.show()

# Plot Confusion Matrix
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(model, X_valid, y_valid)
plt.title('Confusion Matrix on Valid Data')
```

```
plt.show()
```

```
Confusion Matrix (Accuracy 0.7965)

       Prediction
Actual    0    1
     0 4750 1209
     1 1233 4808
              precision    recall  f1-score   support

           0       0.79      0.80      0.80      5959
           1       0.80      0.80      0.80      6041

    accuracy                           0.80     12000
   macro avg       0.80      0.80      0.80     12000
weighted avg       0.80      0.80      0.80     12000
```



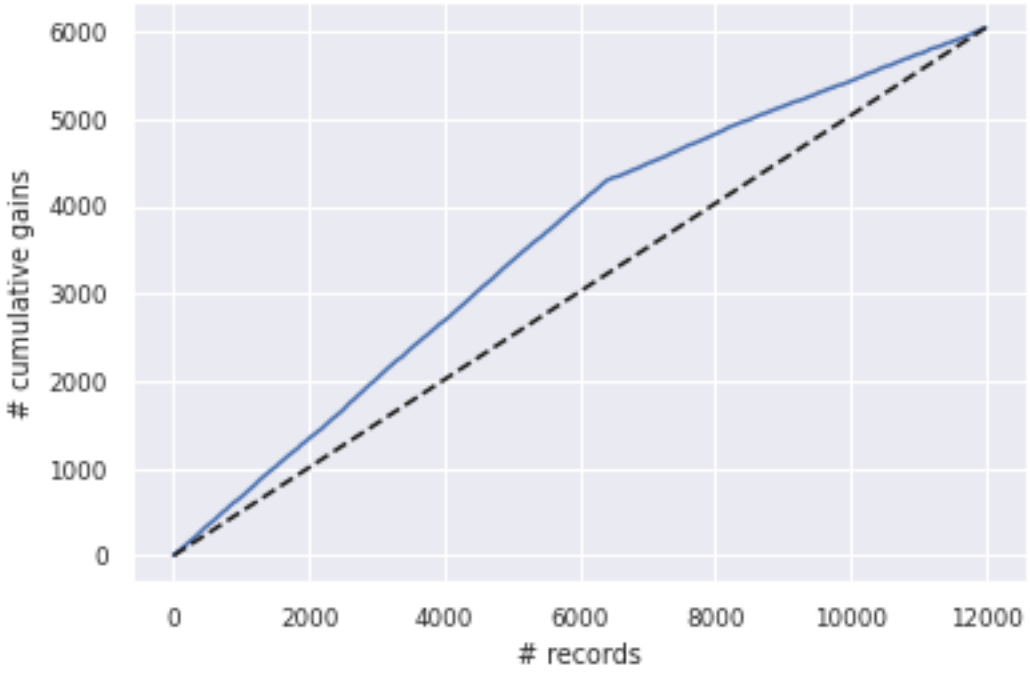


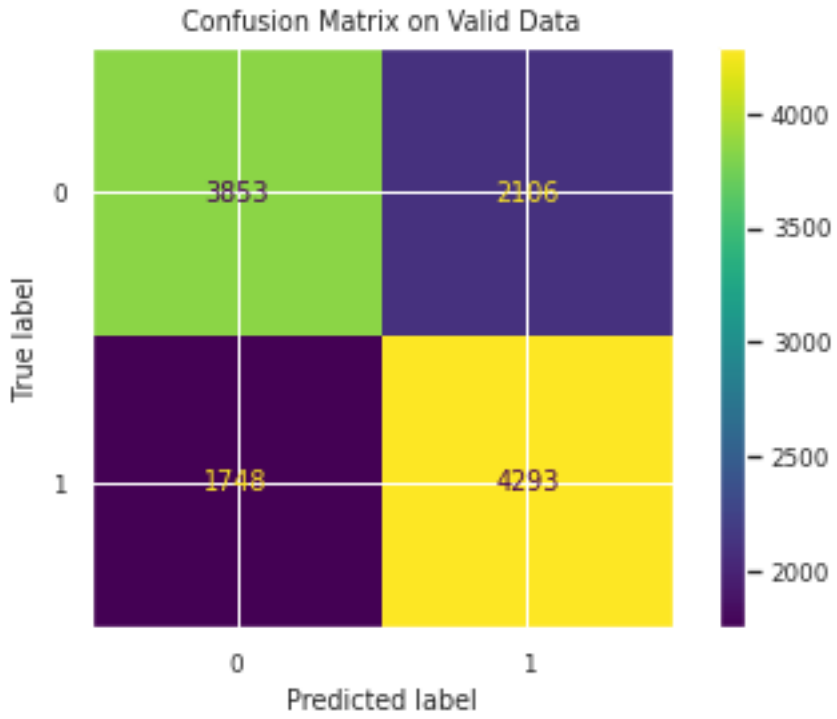
Naive Bayes Classifiers

[ ]:
```
from sklearn.naive_bayes import MultinomialNB
NB_bayes = MultinomialNB(alpha=0.001)
NB_bayes.fit(X_train, y_train)
classificationSummary(y_valid, NB_bayes.predict(X_valid))
print(classification_report(y_valid, NB_bayes.predict(X_valid)))
# Plot the cumulative gains chart of the expected spending
gains_df = pd.DataFrame({'actual': y_valid,
                         'prob': NB_bayes.predict(X_valid)})
gains_df = gains_df.sort_values(by=['prob'], ascending=False).
 ↪reset_index(drop=True)
gainsChart(gains_df.actual)
plt.show()

#
# Plot Confusion Matrix
from sklearn.metrics import plot_confusion_matrix
plot_confusion_matrix(NB_bayes, X_valid, y_valid)
plt.title('Confusion Matrix on Valid Data')
plt.show()
```

```
Confusion Matrix (Accuracy 0.6788)
```

```
       Prediction
Actual    0    1
     0 3853 2106
     1 1748 4293
              precision    recall  f1-score   support

           0       0.69      0.65      0.67      5959
           1       0.67      0.71      0.69      6041

    accuracy                           0.68     12000
   macro avg       0.68      0.68      0.68     12000
weighted avg       0.68      0.68      0.68     12000
```





[ ]:
```
!apt-get install texlive texlive-xetex texlive-latex-extra pandoc
!pip install pypandoc
```

```
Reading package lists… Done
Building dependency tree
Reading state information… Done
pandoc is already the newest version (1.19.2.4~dfsg-1build4).
pandoc set to manually installed.
The following package was automatically installed and is no longer required:
  libnvidia-common-460
Use 'apt autoremove' to remove it.
The following additional packages will be installed:
  fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono fonts-texgyre
  javascript-common libcupsfilters1 libcupsimage2 libgs9 libgs9-common
  libijs-0.35 libjbig2dec0 libjs-jquery libkpathsea6 libpotrace0 libptexenc1
  libruby2.5 libsynctex1 libtexlua52 libtexluajit2 libzzip-0-13 lmodern
  poppler-data preview-latex-style rake ruby ruby-did-you-mean ruby-minitest
  ruby-net-telnet ruby-power-assert ruby-test-unit ruby2.5
  rubygems-integration t1utils tex-common tex-gyre texlive-base
  texlive-binaries texlive-fonts-recommended texlive-latex-base
  texlive-latex-recommended texlive-pictures texlive-plain-generic tipa
Suggested packages:
  fonts-noto apache2 | lighttpd | httpd poppler-utils ghostscript
  fonts-japanese-mincho | fonts-ipafont-mincho fonts-japanese-gothic
```

```
  | fonts-ipafont-gothic fonts-arphic-ukai fonts-arphic-uming fonts-nanum ri
  ruby-dev bundler debhelper gv | postscript-viewer perl-tk xpdf-reader
  | pdf-viewer texlive-fonts-recommended-doc texlive-latex-base-doc
  python-pygments icc-profiles libfile-which-perl
  libspreadsheet-parseexcel-perl texlive-latex-extra-doc
  texlive-latex-recommended-doc texlive-pstricks dot2tex prerex ruby-tcltk
  | libtcltk-ruby texlive-pictures-doc vprerex
The following NEW packages will be installed:
  fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono fonts-texgyre
  javascript-common libcupsfilters1 libcupsimage2 libgs9 libgs9-common
  libijs-0.35 libjbig2dec0 libjs-jquery libkpathsea6 libpotrace0 libptexenc1
  libruby2.5 libsynctex1 libtexlua52 libtexluajit2 libzzip-0-13 lmodern
  poppler-data preview-latex-style rake ruby ruby-did-you-mean ruby-minitest
  ruby-net-telnet ruby-power-assert ruby-test-unit ruby2.5
  rubygems-integration t1utils tex-common tex-gyre texlive texlive-base
  texlive-binaries texlive-fonts-recommended texlive-latex-base
  texlive-latex-extra texlive-latex-recommended texlive-pictures
  texlive-plain-generic texlive-xetex tipa
0 upgraded, 47 newly installed, 0 to remove and 12 not upgraded.
Need to get 146 MB of archives.
After this operation, 460 MB of additional disk space will be used.
Get:1 http://archive.ubuntu.com/ubuntu bionic/main amd64 fonts-droid-fallback
all 1:6.0.1r16-1.1 [1,805 kB]
Get:2 http://archive.ubuntu.com/ubuntu bionic/main amd64 fonts-lato all 2.0-2
[2,698 kB]
Get:3 http://archive.ubuntu.com/ubuntu bionic/main amd64 poppler-data all
0.4.8-2 [1,479 kB]
Get:4 http://archive.ubuntu.com/ubuntu bionic/main amd64 tex-common all 6.09
[33.0 kB]
Get:5 http://archive.ubuntu.com/ubuntu bionic/main amd64 fonts-lmodern all
2.004.5-3 [4,551 kB]
Get:6 http://archive.ubuntu.com/ubuntu bionic/main amd64 fonts-noto-mono all
20171026-2 [75.5 kB]
Get:7 http://archive.ubuntu.com/ubuntu bionic/universe amd64 fonts-texgyre all
20160520-1 [8,761 kB]
Get:8 http://archive.ubuntu.com/ubuntu bionic/main amd64 javascript-common all
11 [6,066 B]
Get:9 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libcupsfilters1
amd64 1.20.2-0ubuntu3.1 [108 kB]
Get:10 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libcupsimage2
amd64 2.2.7-1ubuntu2.9 [18.6 kB]
Get:11 http://archive.ubuntu.com/ubuntu bionic/main amd64 libijs-0.35 amd64
0.35-13 [15.5 kB]
Get:12 http://archive.ubuntu.com/ubuntu bionic/main amd64 libjbig2dec0 amd64
0.13-6 [55.9 kB]
Get:13 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libgs9-common
all 9.26~dfsg+0-0ubuntu0.18.04.17 [5,092 kB]
Get:14 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libgs9 amd64
```

```
9.26~dfsg+0-0ubuntu0.18.04.17 [2,267 kB]
Get:15 http://archive.ubuntu.com/ubuntu bionic/main amd64 libjs-jquery all
3.2.1-1 [152 kB]
Get:16 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libkpathsea6
amd64 2017.20170613.44572-8ubuntu0.1 [54.9 kB]
Get:17 http://archive.ubuntu.com/ubuntu bionic/main amd64 libpotrace0 amd64
1.14-2 [17.4 kB]
Get:18 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libptexenc1
amd64 2017.20170613.44572-8ubuntu0.1 [34.5 kB]
Get:19 http://archive.ubuntu.com/ubuntu bionic/main amd64 rubygems-integration
all 1.11 [4,994 B]
Get:20 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 ruby2.5 amd64
2.5.1-1ubuntu1.12 [48.6 kB]
Get:21 http://archive.ubuntu.com/ubuntu bionic/main amd64 ruby amd64 1:2.5.1
[5,712 B]
Get:22 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 rake all
12.3.1-1ubuntu0.1 [44.9 kB]
Get:23 http://archive.ubuntu.com/ubuntu bionic/main amd64 ruby-did-you-mean all
1.2.0-2 [9,700 B]
Get:24 http://archive.ubuntu.com/ubuntu bionic/main amd64 ruby-minitest all
5.10.3-1 [38.6 kB]
Get:25 http://archive.ubuntu.com/ubuntu bionic/main amd64 ruby-net-telnet all
0.1.1-2 [12.6 kB]
Get:26 http://archive.ubuntu.com/ubuntu bionic/main amd64 ruby-power-assert all
0.3.0-1 [7,952 B]
Get:27 http://archive.ubuntu.com/ubuntu bionic/main amd64 ruby-test-unit all
3.2.5-1 [61.1 kB]
Get:28 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libruby2.5
amd64 2.5.1-1ubuntu1.12 [3,073 kB]
Get:29 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libsynctex1
amd64 2017.20170613.44572-8ubuntu0.1 [41.4 kB]
Get:30 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libtexlua52
amd64 2017.20170613.44572-8ubuntu0.1 [91.2 kB]
Get:31 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libtexluajit2
amd64 2017.20170613.44572-8ubuntu0.1 [230 kB]
Get:32 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 libzzip-0-13
amd64 0.13.62-3.1ubuntu0.18.04.1 [26.0 kB]
Get:33 http://archive.ubuntu.com/ubuntu bionic/main amd64 lmodern all 2.004.5-3
[9,631 kB]
Get:34 http://archive.ubuntu.com/ubuntu bionic/main amd64 preview-latex-style
all 11.91-1ubuntu1 [185 kB]
Get:35 http://archive.ubuntu.com/ubuntu bionic/main amd64 t1utils amd64 1.41-2
[56.0 kB]
Get:36 http://archive.ubuntu.com/ubuntu bionic/universe amd64 tex-gyre all
20160520-1 [4,998 kB]
Get:37 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 texlive-
binaries amd64 2017.20170613.44572-8ubuntu0.1 [8,179 kB]
Get:38 http://archive.ubuntu.com/ubuntu bionic/main amd64 texlive-base all
```

```
2017.20180305-1 [18.7 MB]
Get:39 http://archive.ubuntu.com/ubuntu bionic/universe amd64 texlive-fonts-
recommended all 2017.20180305-1 [5,262 kB]
Get:40 http://archive.ubuntu.com/ubuntu bionic/main amd64 texlive-latex-base all
2017.20180305-1 [951 kB]
Get:41 http://archive.ubuntu.com/ubuntu bionic/main amd64 texlive-latex-
recommended all 2017.20180305-1 [14.9 MB]
Get:42 http://archive.ubuntu.com/ubuntu bionic/universe amd64 texlive all
2017.20180305-1 [14.4 kB]
Get:43 http://archive.ubuntu.com/ubuntu bionic/universe amd64 texlive-pictures
all 2017.20180305-1 [4,026 kB]
Get:44 http://archive.ubuntu.com/ubuntu bionic/universe amd64 texlive-latex-
extra all 2017.20180305-2 [10.6 MB]
Get:45 http://archive.ubuntu.com/ubuntu bionic/universe amd64 texlive-plain-
generic all 2017.20180305-2 [23.6 MB]
Get:46 http://archive.ubuntu.com/ubuntu bionic/universe amd64 tipa all 2:1.3-20
[2,978 kB]
Get:47 http://archive.ubuntu.com/ubuntu bionic/universe amd64 texlive-xetex all
2017.20180305-1 [10.7 MB]
Fetched 146 MB in 7s (21.8 MB/s)
Extracting templates from packages: 100%
Preconfiguring packages …
Selecting previously unselected package fonts-droid-fallback.
(Reading database … 123934 files and directories currently installed.)
Preparing to unpack …/00-fonts-droid-fallback_1%3a6.0.1r16-1.1_all.deb …
Unpacking fonts-droid-fallback (1:6.0.1r16-1.1) …
Selecting previously unselected package fonts-lato.
Preparing to unpack …/01-fonts-lato_2.0-2_all.deb …
Unpacking fonts-lato (2.0-2) …
Selecting previously unselected package poppler-data.
Preparing to unpack …/02-poppler-data_0.4.8-2_all.deb …
Unpacking poppler-data (0.4.8-2) …
Selecting previously unselected package tex-common.
Preparing to unpack …/03-tex-common_6.09_all.deb …
Unpacking tex-common (6.09) …
Selecting previously unselected package fonts-lmodern.
Preparing to unpack …/04-fonts-lmodern_2.004.5-3_all.deb …
Unpacking fonts-lmodern (2.004.5-3) …
Selecting previously unselected package fonts-noto-mono.
Preparing to unpack …/05-fonts-noto-mono_20171026-2_all.deb …
Unpacking fonts-noto-mono (20171026-2) …
Selecting previously unselected package fonts-texgyre.
Preparing to unpack …/06-fonts-texgyre_20160520-1_all.deb …
Unpacking fonts-texgyre (20160520-1) …
Selecting previously unselected package javascript-common.
Preparing to unpack …/07-javascript-common_11_all.deb …
Unpacking javascript-common (11) …
Selecting previously unselected package libcupsfilters1:amd64.
```

```
Preparing to unpack …/08-libcupsfilters1_1.20.2-0ubuntu3.1_amd64.deb …
Unpacking libcupsfilters1:amd64 (1.20.2-0ubuntu3.1) …
Selecting previously unselected package libcupsimage2:amd64.
Preparing to unpack …/09-libcupsimage2_2.2.7-1ubuntu2.9_amd64.deb …
Unpacking libcupsimage2:amd64 (2.2.7-1ubuntu2.9) …
Selecting previously unselected package libijs-0.35:amd64.
Preparing to unpack …/10-libijs-0.35_0.35-13_amd64.deb …
Unpacking libijs-0.35:amd64 (0.35-13) …
Selecting previously unselected package libjbig2dec0:amd64.
Preparing to unpack …/11-libjbig2dec0_0.13-6_amd64.deb …
Unpacking libjbig2dec0:amd64 (0.13-6) …
Selecting previously unselected package libgs9-common.
Preparing to unpack …/12-libgs9-common_9.26~dfsg+0-0ubuntu0.18.04.17_all.deb
…
Unpacking libgs9-common (9.26~dfsg+0-0ubuntu0.18.04.17) …
Selecting previously unselected package libgs9:amd64.
Preparing to unpack …/13-libgs9_9.26~dfsg+0-0ubuntu0.18.04.17_amd64.deb …
Unpacking libgs9:amd64 (9.26~dfsg+0-0ubuntu0.18.04.17) …
Selecting previously unselected package libjs-jquery.
Preparing to unpack …/14-libjs-jquery_3.2.1-1_all.deb …
Unpacking libjs-jquery (3.2.1-1) …
Selecting previously unselected package libkpathsea6:amd64.
Preparing to unpack …/15-libkpathsea6_2017.20170613.44572-8ubuntu0.1_amd64.deb
…
Unpacking libkpathsea6:amd64 (2017.20170613.44572-8ubuntu0.1) …
Selecting previously unselected package libpotrace0.
Preparing to unpack …/16-libpotrace0_1.14-2_amd64.deb …
Unpacking libpotrace0 (1.14-2) …
Selecting previously unselected package libptexenc1:amd64.
Preparing to unpack …/17-libptexenc1_2017.20170613.44572-8ubuntu0.1_amd64.deb
…
Unpacking libptexenc1:amd64 (2017.20170613.44572-8ubuntu0.1) …
Selecting previously unselected package rubygems-integration.
Preparing to unpack …/18-rubygems-integration_1.11_all.deb …
Unpacking rubygems-integration (1.11) …
Selecting previously unselected package ruby2.5.
Preparing to unpack …/19-ruby2.5_2.5.1-1ubuntu1.12_amd64.deb …
Unpacking ruby2.5 (2.5.1-1ubuntu1.12) …
Selecting previously unselected package ruby.
Preparing to unpack …/20-ruby_1%3a2.5.1_amd64.deb …
Unpacking ruby (1:2.5.1) …
Selecting previously unselected package rake.
Preparing to unpack …/21-rake_12.3.1-1ubuntu0.1_all.deb …
Unpacking rake (12.3.1-1ubuntu0.1) …
Selecting previously unselected package ruby-did-you-mean.
Preparing to unpack …/22-ruby-did-you-mean_1.2.0-2_all.deb …
Unpacking ruby-did-you-mean (1.2.0-2) …
Selecting previously unselected package ruby-minitest.
```

```
Preparing to unpack …/23-ruby-minitest_5.10.3-1_all.deb …
Unpacking ruby-minitest (5.10.3-1) …
Selecting previously unselected package ruby-net-telnet.
Preparing to unpack …/24-ruby-net-telnet_0.1.1-2_all.deb …
Unpacking ruby-net-telnet (0.1.1-2) …
Selecting previously unselected package ruby-power-assert.
Preparing to unpack …/25-ruby-power-assert_0.3.0-1_all.deb …
Unpacking ruby-power-assert (0.3.0-1) …
Selecting previously unselected package ruby-test-unit.
Preparing to unpack …/26-ruby-test-unit_3.2.5-1_all.deb …
Unpacking ruby-test-unit (3.2.5-1) …
Selecting previously unselected package libruby2.5:amd64.
Preparing to unpack …/27-libruby2.5_2.5.1-1ubuntu1.12_amd64.deb …
Unpacking libruby2.5:amd64 (2.5.1-1ubuntu1.12) …
Selecting previously unselected package libsynctex1:amd64.
Preparing to unpack …/28-libsynctex1_2017.20170613.44572-8ubuntu0.1_amd64.deb
…
Unpacking libsynctex1:amd64 (2017.20170613.44572-8ubuntu0.1) …
Selecting previously unselected package libtexlua52:amd64.
Preparing to unpack …/29-libtexlua52_2017.20170613.44572-8ubuntu0.1_amd64.deb
…
Unpacking libtexlua52:amd64 (2017.20170613.44572-8ubuntu0.1) …
Selecting previously unselected package libtexluajit2:amd64.
Preparing to unpack
…/30-libtexluajit2_2017.20170613.44572-8ubuntu0.1_amd64.deb …
Unpacking libtexluajit2:amd64 (2017.20170613.44572-8ubuntu0.1) …
Selecting previously unselected package libzzip-0-13:amd64.
Preparing to unpack …/31-libzzip-0-13_0.13.62-3.1ubuntu0.18.04.1_amd64.deb …
Unpacking libzzip-0-13:amd64 (0.13.62-3.1ubuntu0.18.04.1) …
Selecting previously unselected package lmodern.
Preparing to unpack …/32-lmodern_2.004.5-3_all.deb …
Unpacking lmodern (2.004.5-3) …
Selecting previously unselected package preview-latex-style.
Preparing to unpack …/33-preview-latex-style_11.91-1ubuntu1_all.deb …
Unpacking preview-latex-style (11.91-1ubuntu1) …
Selecting previously unselected package t1utils.
Preparing to unpack …/34-t1utils_1.41-2_amd64.deb …
Unpacking t1utils (1.41-2) …
Selecting previously unselected package tex-gyre.
Preparing to unpack …/35-tex-gyre_20160520-1_all.deb …
Unpacking tex-gyre (20160520-1) …
Selecting previously unselected package texlive-binaries.
Preparing to unpack …/36-texlive-
binaries_2017.20170613.44572-8ubuntu0.1_amd64.deb …
Unpacking texlive-binaries (2017.20170613.44572-8ubuntu0.1) …
Selecting previously unselected package texlive-base.
Preparing to unpack …/37-texlive-base_2017.20180305-1_all.deb …
Unpacking texlive-base (2017.20180305-1) …
```

```
Selecting previously unselected package texlive-fonts-recommended.
Preparing to unpack …/38-texlive-fonts-recommended_2017.20180305-1_all.deb …
Unpacking texlive-fonts-recommended (2017.20180305-1) …
Selecting previously unselected package texlive-latex-base.
Preparing to unpack …/39-texlive-latex-base_2017.20180305-1_all.deb …
Unpacking texlive-latex-base (2017.20180305-1) …
Selecting previously unselected package texlive-latex-recommended.
Preparing to unpack …/40-texlive-latex-recommended_2017.20180305-1_all.deb …
Unpacking texlive-latex-recommended (2017.20180305-1) …
Selecting previously unselected package texlive.
Preparing to unpack …/41-texlive_2017.20180305-1_all.deb …
Unpacking texlive (2017.20180305-1) …
Selecting previously unselected package texlive-pictures.
Preparing to unpack …/42-texlive-pictures_2017.20180305-1_all.deb …
Unpacking texlive-pictures (2017.20180305-1) …
Selecting previously unselected package texlive-latex-extra.
Preparing to unpack …/43-texlive-latex-extra_2017.20180305-2_all.deb …
Unpacking texlive-latex-extra (2017.20180305-2) …
Selecting previously unselected package texlive-plain-generic.
Preparing to unpack …/44-texlive-plain-generic_2017.20180305-2_all.deb …
Unpacking texlive-plain-generic (2017.20180305-2) …
Selecting previously unselected package tipa.
Preparing to unpack …/45-tipa_2%3a1.3-20_all.deb …
Unpacking tipa (2:1.3-20) …
Selecting previously unselected package texlive-xetex.
Preparing to unpack …/46-texlive-xetex_2017.20180305-1_all.deb …
Unpacking texlive-xetex (2017.20180305-1) …
Setting up libgs9-common (9.26~dfsg+0-0ubuntu0.18.04.17) …
Setting up libkpathsea6:amd64 (2017.20170613.44572-8ubuntu0.1) …
Setting up libjs-jquery (3.2.1-1) …
Setting up libtexlua52:amd64 (2017.20170613.44572-8ubuntu0.1) …
Setting up fonts-droid-fallback (1:6.0.1r16-1.1) …
Setting up libsynctex1:amd64 (2017.20170613.44572-8ubuntu0.1) …
Setting up libptexenc1:amd64 (2017.20170613.44572-8ubuntu0.1) …
Setting up tex-common (6.09) …
update-language: texlive-base not installed and configured, doing nothing!
Setting up poppler-data (0.4.8-2) …
Setting up tex-gyre (20160520-1) …
Setting up preview-latex-style (11.91-1ubuntu1) …
Setting up fonts-texgyre (20160520-1) …
Setting up fonts-noto-mono (20171026-2) …
Setting up fonts-lato (2.0-2) …
Setting up libcupsfilters1:amd64 (1.20.2-0ubuntu3.1) …
Setting up libcupsimage2:amd64 (2.2.7-1ubuntu2.9) …
Setting up libjbig2dec0:amd64 (0.13-6) …
Setting up ruby-did-you-mean (1.2.0-2) …
Setting up t1utils (1.41-2) …
Setting up ruby-net-telnet (0.1.1-2) …
```

```
Setting up libijs-0.35:amd64 (0.35-13) …
Setting up rubygems-integration (1.11) …
Setting up libpotrace0 (1.14-2) …
Setting up javascript-common (11) …
Setting up ruby-minitest (5.10.3-1) …
Setting up libzzip-0-13:amd64 (0.13.62-3.1ubuntu0.18.04.1) …
Setting up libgs9:amd64 (9.26~dfsg+0-0ubuntu0.18.04.17) …
Setting up libtexluajit2:amd64 (2017.20170613.44572-8ubuntu0.1) …
Setting up fonts-lmodern (2.004.5-3) …
Setting up ruby-power-assert (0.3.0-1) …
Setting up texlive-binaries (2017.20170613.44572-8ubuntu0.1) …
update-alternatives: using /usr/bin/xdvi-xaw to provide /usr/bin/xdvi.bin
(xdvi.bin) in auto mode
update-alternatives: using /usr/bin/bibtex.original to provide /usr/bin/bibtex
(bibtex) in auto mode
Setting up texlive-base (2017.20180305-1) …
mktexlsr: Updating /var/lib/texmf/ls-R-TEXLIVEDIST…
mktexlsr: Updating /var/lib/texmf/ls-R-TEXMFMAIN…
mktexlsr: Updating /var/lib/texmf/ls-R…
mktexlsr: Done.
tl-paper: setting paper size for dvips to a4:
/var/lib/texmf/dvips/config/config-paper.ps
tl-paper: setting paper size for dvipdfmx to a4:
/var/lib/texmf/dvipdfmx/dvipdfmx-paper.cfg
tl-paper: setting paper size for xdvi to a4: /var/lib/texmf/xdvi/XDvi-paper
tl-paper: setting paper size for pdftex to a4:
/var/lib/texmf/tex/generic/config/pdftexconfig.tex
Setting up texlive-fonts-recommended (2017.20180305-1) …
Setting up texlive-plain-generic (2017.20180305-2) …
Setting up texlive-latex-base (2017.20180305-1) …
Setting up lmodern (2.004.5-3) …
Setting up texlive-latex-recommended (2017.20180305-1) …
Setting up texlive-pictures (2017.20180305-1) …
Setting up tipa (2:1.3-20) …
Regenerating '/var/lib/texmf/fmtutil.cnf-DEBIAN'… done.
Regenerating '/var/lib/texmf/fmtutil.cnf-TEXLIVEDIST'… done.
update-fmtutil has updated the following file(s):
        /var/lib/texmf/fmtutil.cnf-DEBIAN
        /var/lib/texmf/fmtutil.cnf-TEXLIVEDIST
If you want to activate the changes in the above file(s),
you should run fmtutil-sys or fmtutil.
Setting up texlive (2017.20180305-1) …
Setting up texlive-latex-extra (2017.20180305-2) …
Setting up texlive-xetex (2017.20180305-1) …
Setting up ruby2.5 (2.5.1-1ubuntu1.12) …
Setting up ruby (1:2.5.1) …
Setting up ruby-test-unit (3.2.5-1) …
Setting up rake (12.3.1-1ubuntu0.1) …
```

```
Setting up libruby2.5:amd64 (2.5.1-1ubuntu1.12) …
Processing triggers for mime-support (3.60ubuntu1) …
Processing triggers for libc-bin (2.27-3ubuntu1.6) …
Processing triggers for man-db (2.8.3-2ubuntu0.1) …
Processing triggers for fontconfig (2.12.6-0ubuntu2) …
Processing triggers for tex-common (6.09) …
Running updmap-sys. This may take some time… done.
Running mktexlsr /var/lib/texmf … done.
Building format(s) --all.
        This may take some time… done.
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
wheels/public/simple/
Collecting pypandoc
  Downloading pypandoc-1.9-py3-none-any.whl (20 kB)
Installing collected packages: pypandoc
Successfully installed pypandoc-1.9
```

[ ]:
```
%shell jupyter nbconvert --to pdf '/content/ADS505_Final_Project_v2.ipynb'
```

```
[NbConvertApp] Converting notebook /content/ADS505_Final_Project_v2.ipynb to pdf
[NbConvertApp] Support files will be in ADS505_Final_Project_v2_files/
[NbConvertApp] Making directory ./ADS505_Final_Project_v2_files
[NbConvertApp] Making directory ./ADS505_Final_Project_v2_files
[NbConvertApp] Making directory ./ADS505_Final_Project_v2_files
[NbConvertApp] Making directory ./ADS505_Final_Project_v2_files
[NbConvertApp] Making directory ./ADS505_Final_Project_v2_files
[NbConvertApp] Making directory ./ADS505_Final_Project_v2_files
[NbConvertApp] Making directory ./ADS505_Final_Project_v2_files
[NbConvertApp] Making directory ./ADS505_Final_Project_v2_files
[NbConvertApp] Making directory ./ADS505_Final_Project_v2_files
[NbConvertApp] Making directory ./ADS505_Final_Project_v2_files
[NbConvertApp] Making directory ./ADS505_Final_Project_v2_files
[NbConvertApp] Making directory ./ADS505_Final_Project_v2_files
[NbConvertApp] Making directory ./ADS505_Final_Project_v2_files
[NbConvertApp] Making directory ./ADS505_Final_Project_v2_files
[NbConvertApp] Making directory ./ADS505_Final_Project_v2_files
[NbConvertApp] Making directory ./ADS505_Final_Project_v2_files
[NbConvertApp] Making directory ./ADS505_Final_Project_v2_files
[NbConvertApp] Making directory ./ADS505_Final_Project_v2_files
[NbConvertApp] Writing 104043 bytes to ./notebook.tex
[NbConvertApp] Building PDF
[NbConvertApp] Running xelatex 3 times: ['xelatex', './notebook.tex', '-quiet']
[NbConvertApp] Running bibtex 1 time: ['bibtex', './notebook']
[NbConvertApp] WARNING | bibtex had problems, most likely because there were no
citations
[NbConvertApp] PDF successfully created
[NbConvertApp] Writing 692721 bytes to /content/ADS505_Final_Project_v2.pdf
```

[ ]: