Objective

Analyzing the data of DNA products the customers bought and the cross-selling customers the ACOM subscriptions.

Data Description

The data consists of 10 variables with 251942 observations.

Variable	Description
prospectid	Uniqueid for each customer
ordernumber	Unique order id of DNA
ordercreatedate	DNA product was ordered(date)
regtenure	Customer registered email id with ancestry (days)
customer_type_group	Customer type (at DNA product order)
dnatestactivationdayid	DNA test is activated (date)
daystogetresult_grp	Test to get ready(days)
dna_visittrafficsubtype	Traffic channel (DNA product)
xsell_gsa	New ACOM subscriber Addition
xsell_day_exact	From DNA product to ACOM Subscription (days

Cleaning and imputation

- Dropped 73097 rows where either ordercreatedate or daystogetresult_grp missing, if one of them is not null can be imputed
- daystogetresult_grp imputed with the

Exploratory Data Analysis

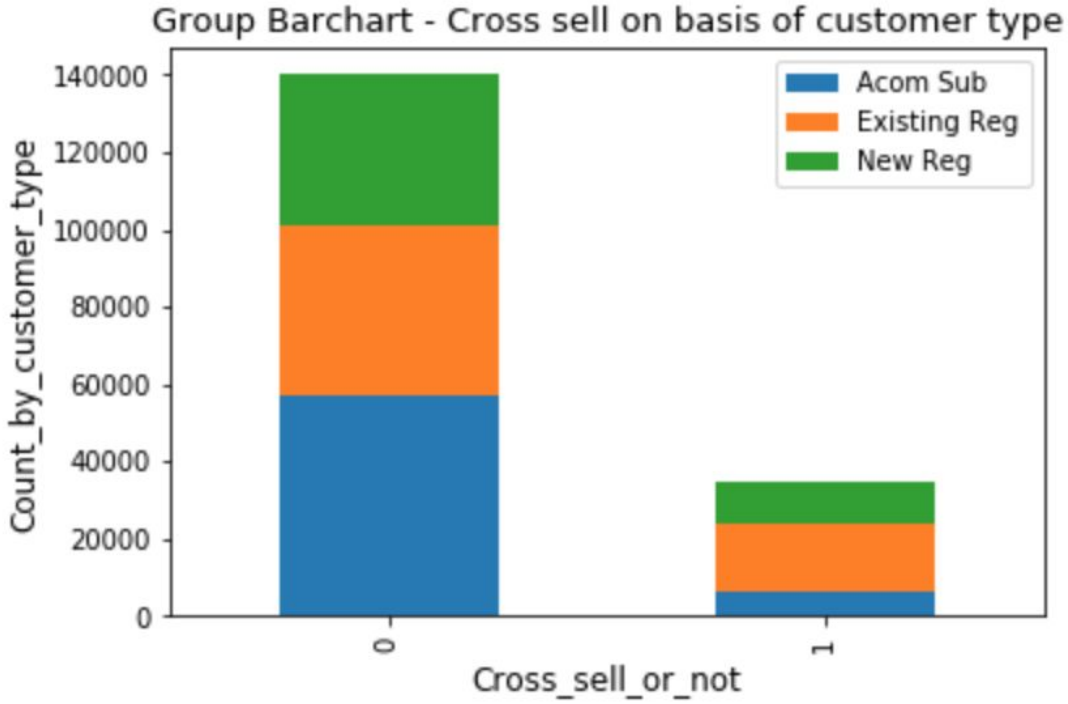
Post cleaning and imputation data set looked like:

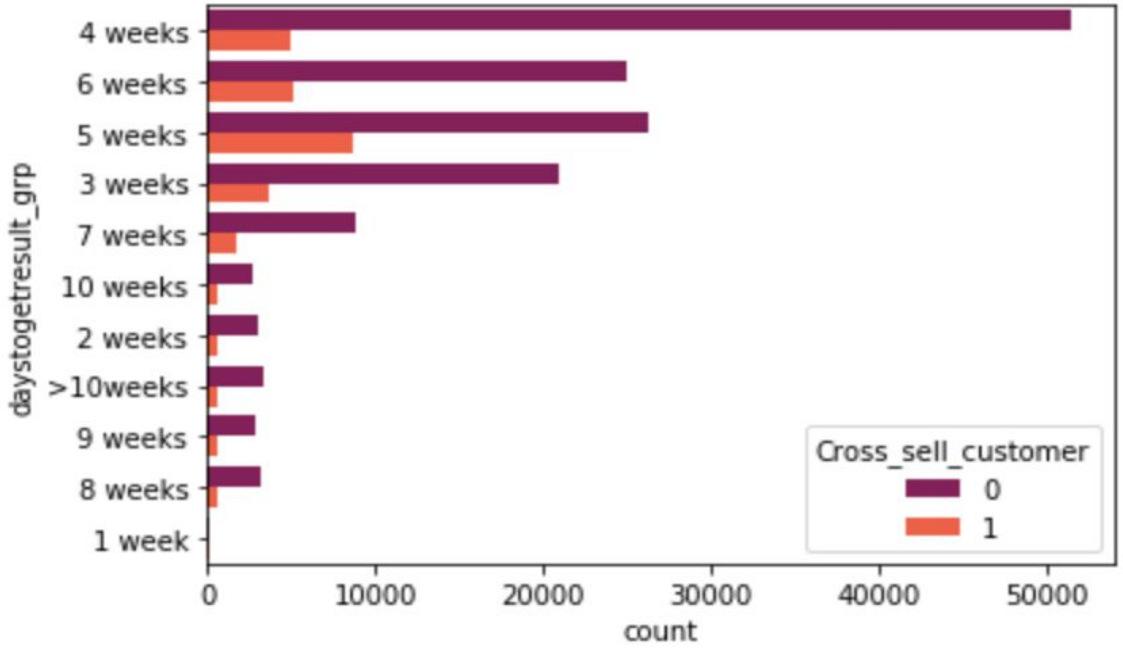
```
data_file_clean_imp["prospectid"].count(
```

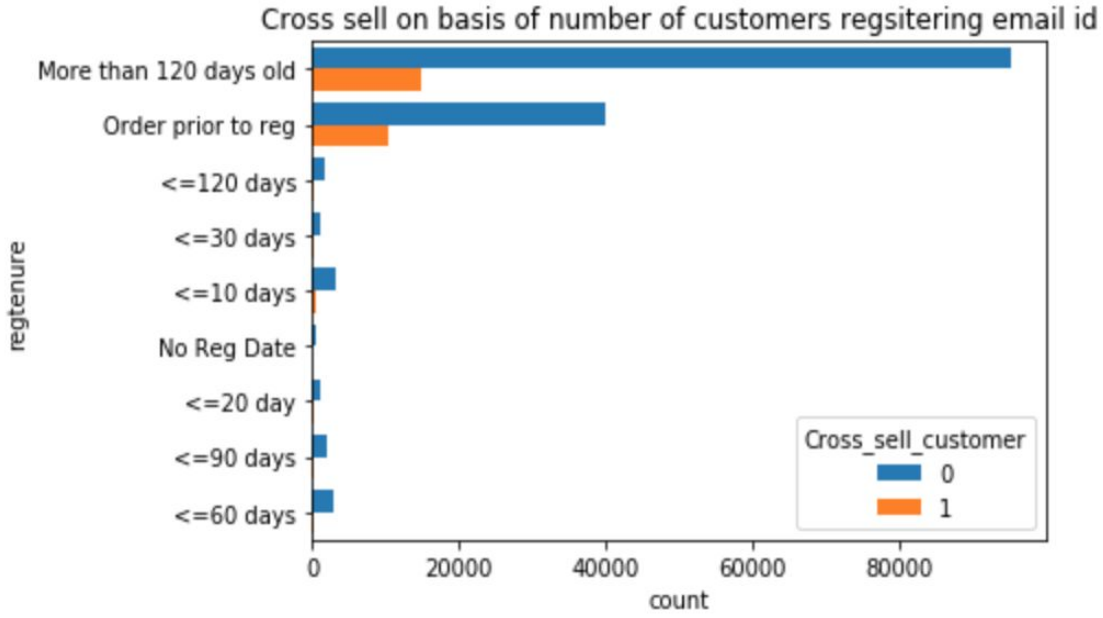
```
174982
```

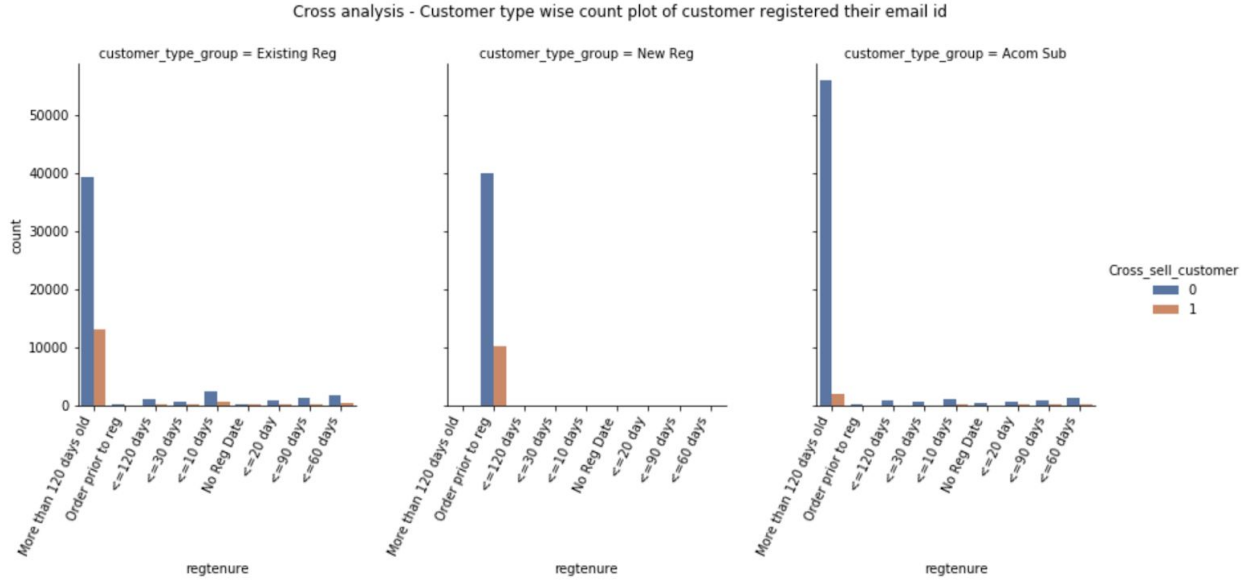
The data variables and their respective datatype were such:

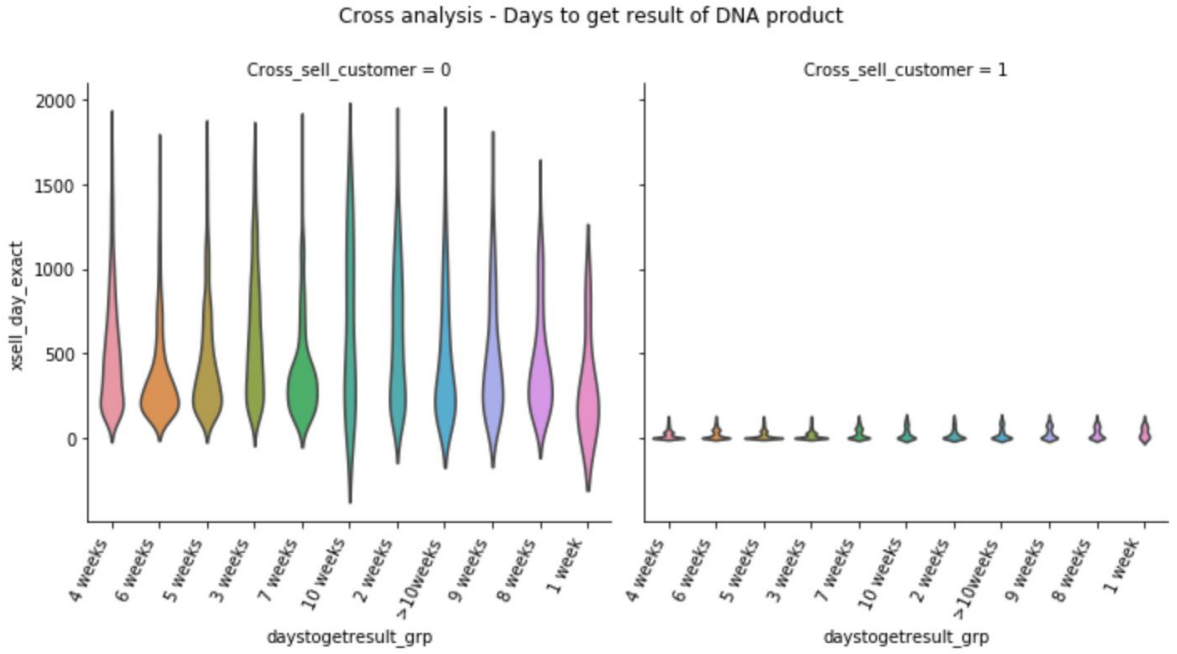
```
prospectid                int64
ordernumber               int64
ordercreatedate          object
regtenure                object
customer_type_group      object
dnatestactivationdayid   object
daystogetresult_grp      object
dna_visittrafficsubtype  object
xsell_gsa                 int64
xsell_day_exact         float64
dtype: object
```

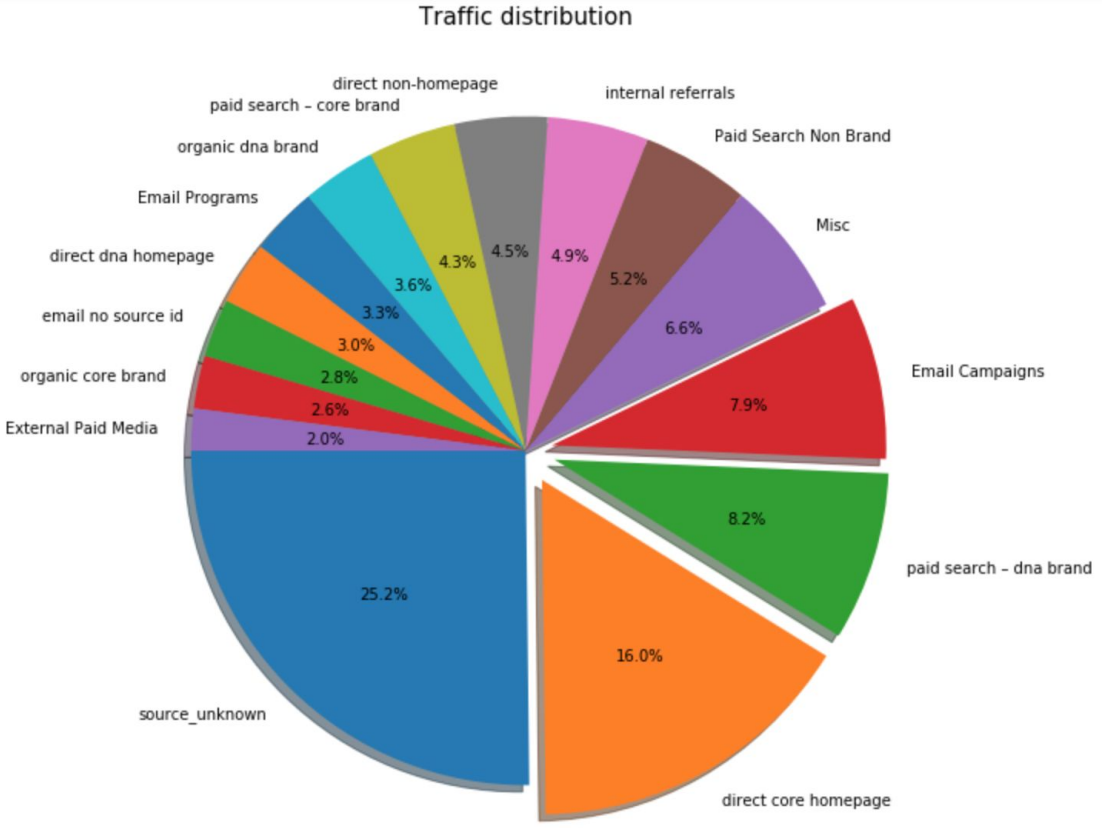

Group Barchart - Cross sell on basis of customer type
Count_by_customer_type
140000
120000
100000
80000
60000
40000
20000
0
Acom Sub
Existing Reg
New Reg
0
1
Cross_sell_or_not

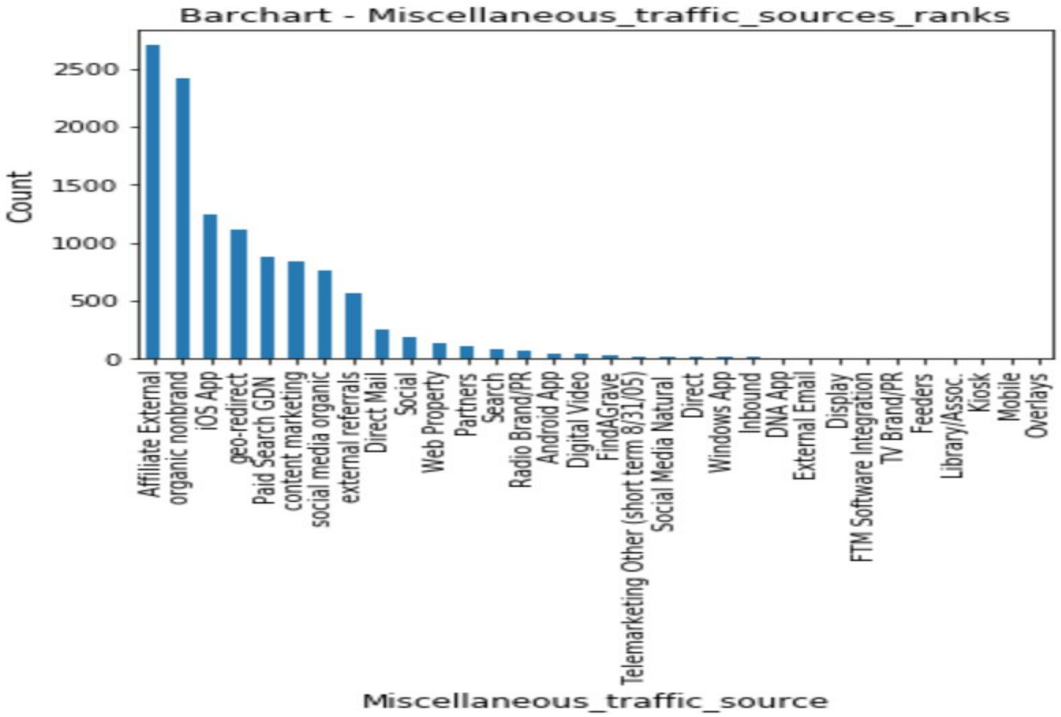

daystogetresult_grp
4 weeks
6 weeks
5 weeks
3 weeks
7 weeks
10 weeks
2 weeks
>10weeks
9 weeks
8 weeks
1 week
0
10000
20000
30000
40000
50000
count
Cross_sell_customer
0
1


Cross sell on basis of number of customers regsitering email id
More than 120 days old
Order prior to reg
<=120 days
<=30 days
<=10 days
No Reg Date
<=20 day
<=90 days
<=60 days
regtenure
0
20000
40000
60000
80000
count
Cross_sell_customer
0
1


Cross analysis - Customer type wise count plot of customer registered their email id
customer_type_group = Existing Reg
customer_type_group = New Reg
customer_type_group = Acom Sub
count
50000
40000
30000
20000
10000
0
More than 120 days old
Order prior to reg
<=120 days
<=30 days
<=10 days
No Reg Date
<=20 day
<=90 days
<=60 days
regtenure
Cross_sell_customer
0
1


Cross analysis - Days to get result of DNA product
Cross_sell_customer = 0
Cross_sell_customer = 1
xsell_day_exact
2000
1500
1000
500
0
4 weeks
6 weeks
5 weeks
3 weeks
7 weeks
10 weeks
2 weeks
>10weeks
9 weeks
8 weeks
1 week
daystogetresult_grp


Traffic distribution
direct non-homepage
paid search – core brand
organic dna brand
Email Programs
direct dna homepage
email no source id
organic core brand
External Paid Media
internal referrals
Paid Search Non Brand
Misc
Email Campaigns
paid search – dna brand
direct core homepage
source_unknown
4.5%
4.3%
3.6%
3.3%
3.0%
2.8%
2.6%
2.0%
4.9%
5.2%
6.6%
7.9%
8.2%
16.0%
25.2%


Barchart - Miscellaneous_traffic_sources_ranks
Count
2500
2000
1500
1000
500
0
Affiliate External
organic nonbrand
iOS App
geo-redirect
Paid Search GDN
content marketing
social media organic
external referrals
Direct Mail
Social
Web Property
Partners
Search
Radio Brand/PR
Android App
Digital Video
FindAGrave
Telemarketing Other (short term 8/31/05)
Social Media Natural
Direct
Windows App
Inbound
DNA App
External Email
Display
FTM Software Integration
TV Brand/PR
Feeders
Library/Assoc.
Kiosk
Mobile
Overlays
Miscellaneous_traffic_source