Svetlana
Charu
YiChen
Harish

Data Analysis and Preparation

- Data sets 1 and 3 had structured data, while sets 2 and 4 had transactional-format data
- Data has lots of missing, some errors
- Highly unbalanced for Success (96% and 4%)
- Derive target variables for classification (moved in?) and regression models for days before move-in, days in the facility, and expected revenue:
  - Success = Status in (MovedIn, Future)
  - DaysToMoveIn = DateMovedIn - Inquiry
  - DaysIn = DateMovedOut - DateMovedIn
  - Revenue: aggregate data4 on patient ID
  - Join data3 and aggregated data4 into data1

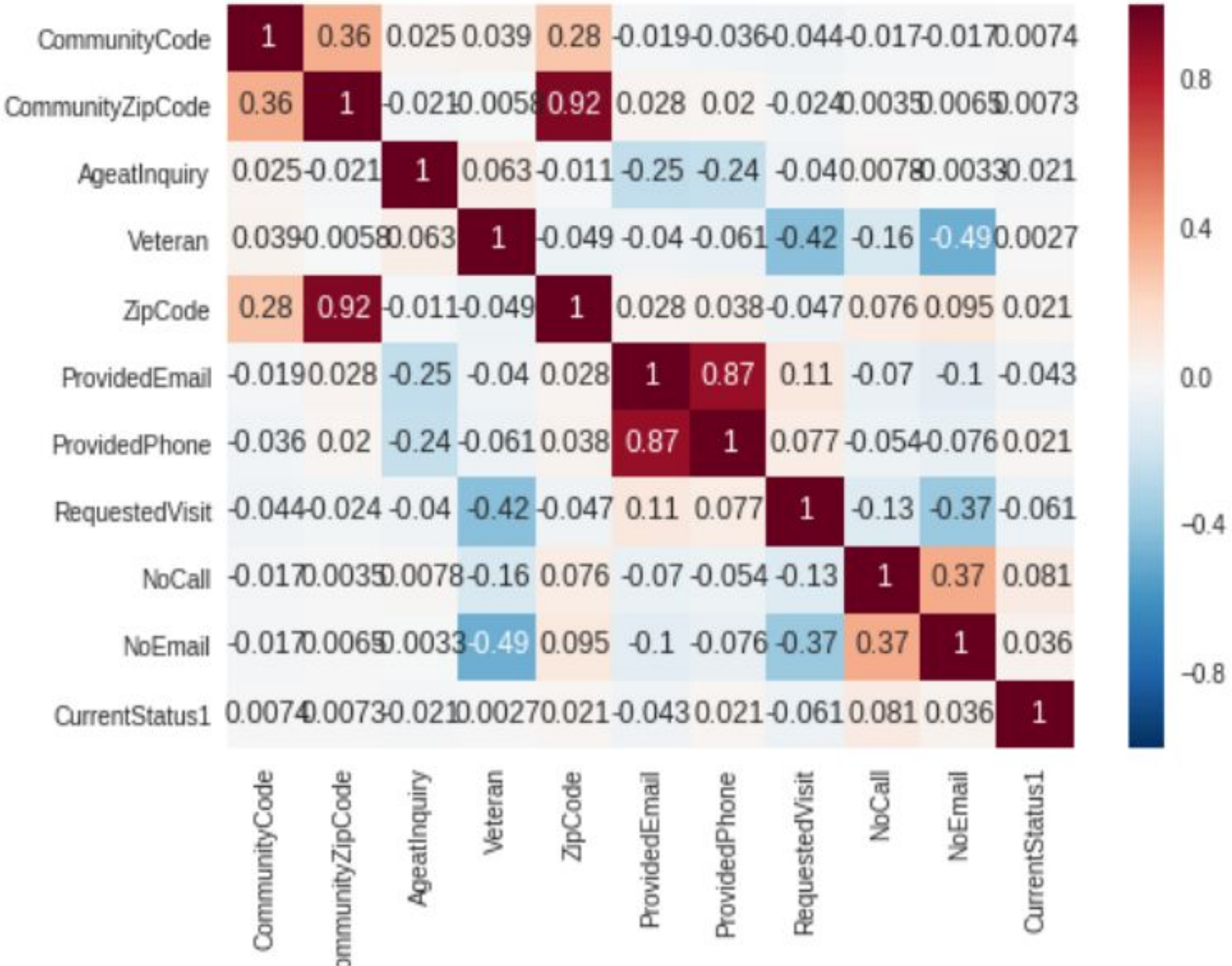
Combining Dataset

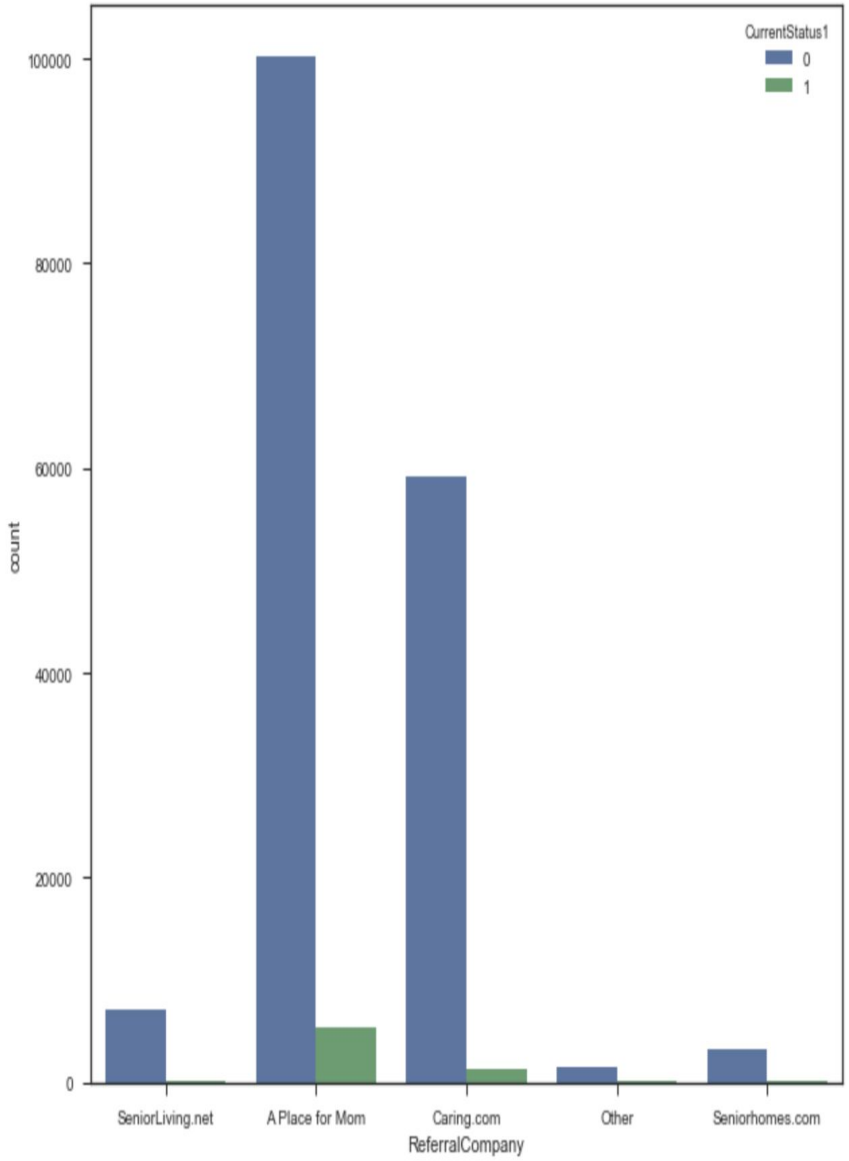
```
NewData=data3.merge(data4, left_on='ResidentID', right_on='ResidentID', how='outer')
NewData.head()
```

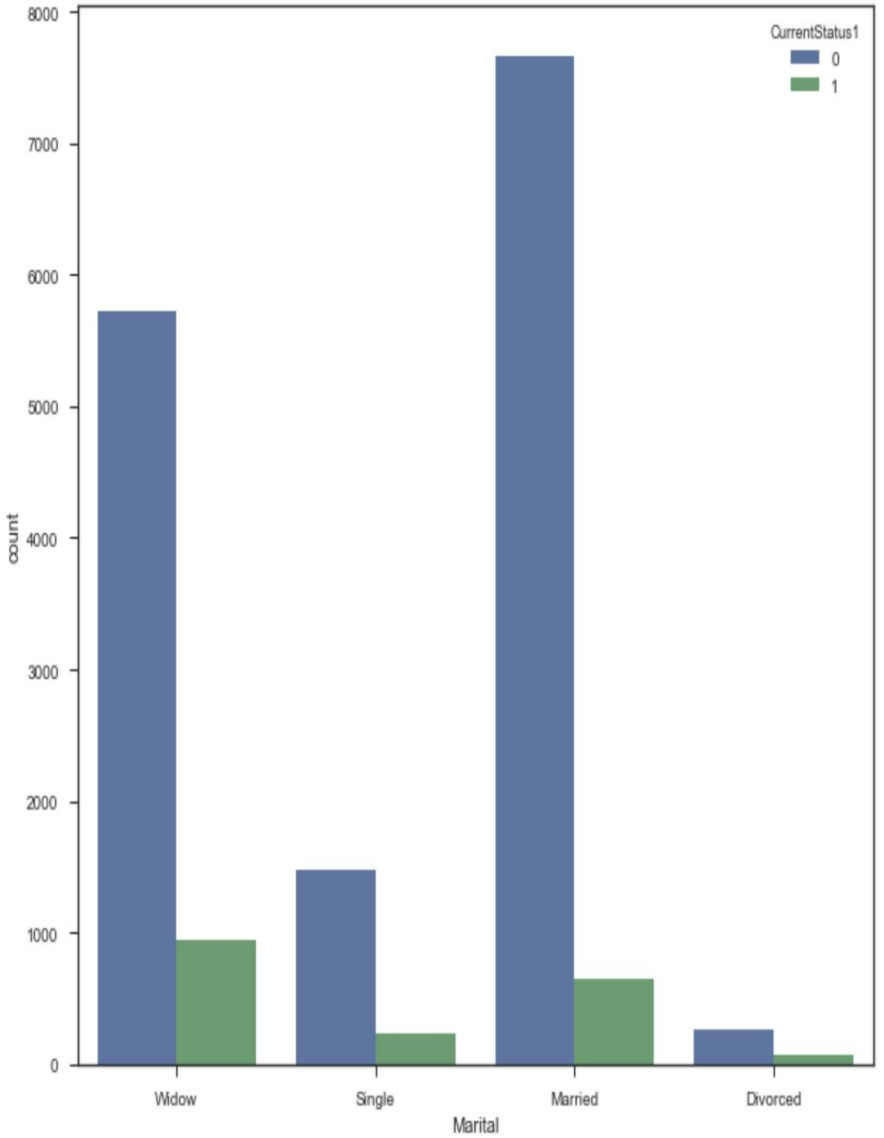
executed in 47ms, finished 15:30:18 2018-04-28

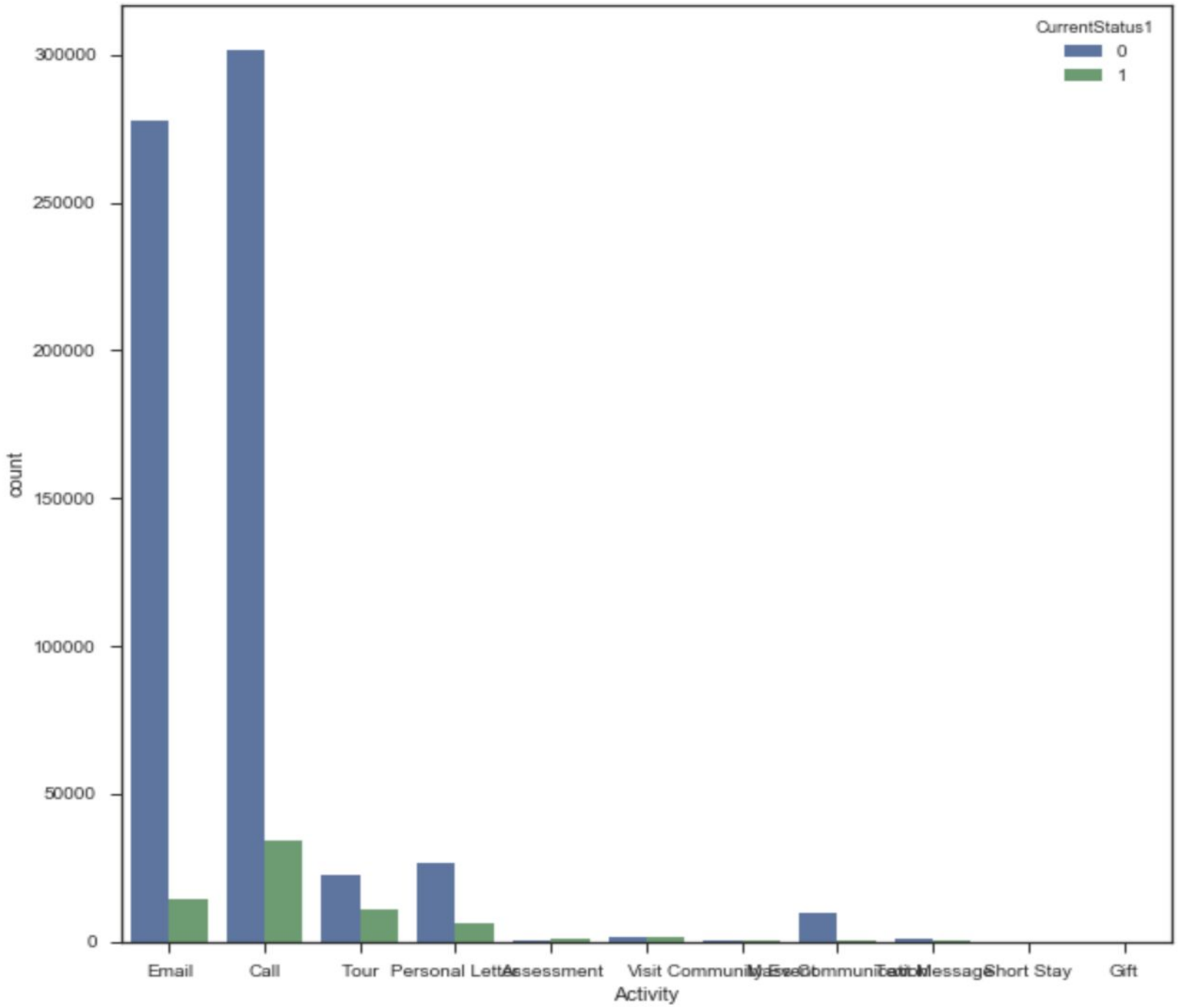
	ProspectID	ResidentID	dtmovein	dtmoveout	istatus	ireason	ServiceType	RateType	FromDate	ToDate	Amount
0	a8e02699-058a-40ff-a9a3-e86e78b9607b	1022063	06/10/2016	NaN	Current	NaN	RB	MLY	06/10/2016	12/31/2016	2095.0
1	a8e02699-058a-40ff-a9a3-e86e78b9607b	1022063	06/10/2016	NaN	Current	NaN	RB	MLY	01/01/2017	12/31/2017	2095.0
2	a8e02699-058a-40ff-a9a3-e86e78b9607b	1022063	06/10/2016	NaN	Current	NaN	RB	MLY	01/01/2018	NaN	2200.0
3	2cdbeff6-6206-e611-80f6-3863bb2eb148	1022071	04/26/2016	04/19/2018	Moved Out	Death	CS	DLY	04/26/2016	07/31/2016	175.0
4	2cdbeff6-6206-e611-80f6-3863bb2eb148	1022071	04/26/2016	04/19/2018	Moved Out	Death	RB	MLY	08/01/2016	12/31/2016	5625.0

Feature Engineering




CurrentStatus1
0
1
100000
80000
60000
40000
20000
0
count
SeniorLiving.net
A Place for Mom
Caring.com
Other
Seniorhomes.com
ReferralCompany


CurrentStatus1
0
1
8000
7000
6000
5000
4000
3000
2000
1000
0
count
Widow
Single
Married
Divorced
Marital


CurrentStatus1
0
1
300000
250000
200000
150000
100000
50000
0
count
Email
Call
Tour
Personal Letter
Assessment
Visit
Short Stay
Gift
Activity

Outcome/ Preprocessing stage:

1 . Zipcode and Communitycode and providedPhone and ProvidedEmail are highly correlated

2. However is the prospect is a veteran has a negative correlation with emails and requested visit

3. **Place for Mom** and **Caring.com** has converted more prospect ID to ResidentID.

4. **Divorced Martial** Status has lesser contribution to the more Resident move In

5. It seems **Email and Call** has more success **but** there is a **higher proportion of conversion** occurs in **TOUR, ASSESMENT, VISIT**

Modeling

1. Classification for MoveIn:

Increase weights for those moved in
Use Chaid tree and Naive Bayes model, exported PMML
2. Regression models for DaysToMoveIn for each
care level. Using regression trees or automatic linear regression.
3. Predicting revenue amount by a regression model
4. Predict number of days in the facility by a survival model (e.g. Cox regression or Weibull models)
5. Analyze activities by aggregating and modeling or by a Sequence model

Future Directions

- More of Iterating preprocessing and analysis step to build more accurate model.
- Combining the ZIP code and the location area Information which is preferable.(say: based on the rent, weather)
- XGBoost models!!!