

# RETAIL RELAY

MAR653 – Homework #2

Team Y - TeKeya Crenshaw, Jennifer Gibbins, Danuel James, John Fields

November 14, 2019



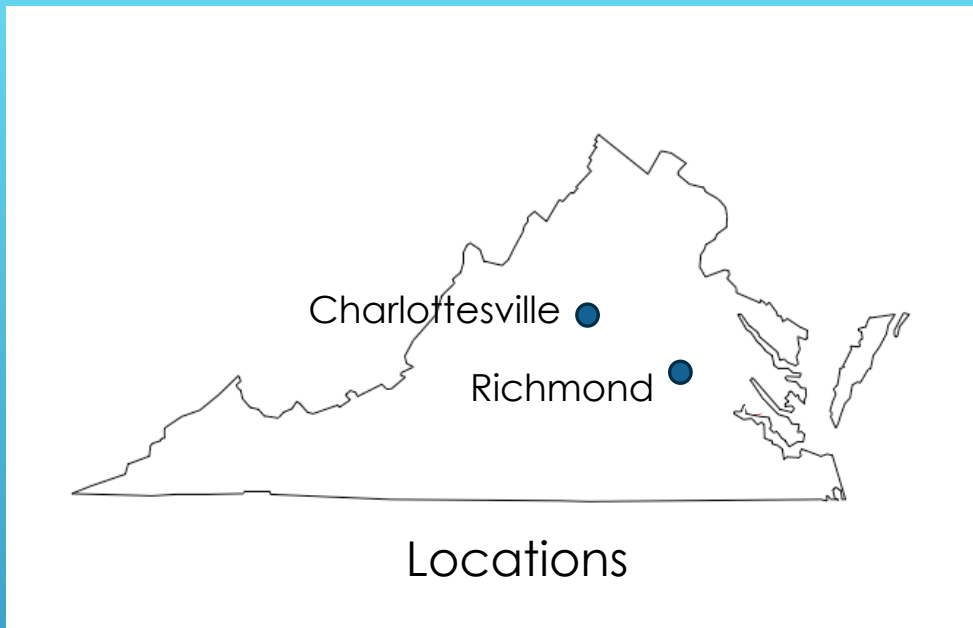
STARTED IN 2007

**2010:** DELIVERY SERVICE WITH TWO LOCATIONS

**2016:** ANNOUNCED MERGER WITH DOOR TO DOOR ORGANICS

**2017:** SHUTTERED THE RE-BRANDED RELAY FOODS IN VIRGINIA STATING DOOR TO DOOR WOULD CONTINUE THE AREA SERVICES

**2017:** DOOR TO DOOR ORGANICS SHUTTERS ABRUPTLY



Locations



- <https://richmondbizsense.com/2017/01/05/online-grocer-to-close-scotts-addition-warehouse-lay-off-virginia-workers/>
- <https://progressivegrocer.com/door-door-organics-shutters-abruptly>

# FIRST THINGS FIRST, WE CLEANED THE DATA

- ❑ Original - 24,579 rows and 16 columns
- ❑ Cleansing through deletion
  - ❑ Data without dates and customer IDs (17 rows)
  - ❑ Data with invalid dates for orders (12 rows)
  - ❑ Data with invalid dates for last order (11 rows)
- ❑ Creation of new features
  - ❑ Days between account creation and first order
  - ❑ Days since last order based on January 20,2014
  - ❑ Favorite day conversion to number
  - ❑ City conversion to number
  - ❑ Created columns for favorite city and day of week to make them binary
- ❑ **Final Dataset** – 24,539 rows and 20 columns

## CHECKING FOR MULTI-COLLINEARITY

	retained	Days.between.creation.and.first.order	Days.since.last.order	esent	eopenrate	eclickrate	avgorder	ordfreq	paperless	refill	doorstep	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	BWI	CHO	DCX	RIC	
retained	1	-0.0136744	-0.3061547	0.65424238	0.22641171	0.27154204	0.00326066	0.02448814	0.17716572	0.10280088	0.06880409	0.01626172	-0.010114	-0.0115532	-0.0041813	0.01458463	0.02409256	-0.0500025	0.04754149	-0.0893459	0.05757701	0.00722163	
Days.between.creation.and.first.order	-0.0136744	1	0.09628705	0.17784853	-0.0311014	0.0364308	0.2363255	0.13333478	-0.2205856	-0.0734856	-0.0411124	0.0174267	0.03191501	0.02667866	-0.0047364	-0.0213925	8.01E-05	-0.1439395	0.01189322	0.10732328	-0.0651524	-0.0450574	
Days.since.last.order	-0.3061547	0.09628705	1	-0.0587267	-0.5096255	-0.5299321	0.12769493	-0.070482	-0.5784385	-0.3032702	-0.2483174	-0.0631625	-0.0072657	0.04460094	0.03802079	-0.0248531	-0.0488665	0.12534322	-0.0989549	0.19435277	-0.217767	0.06825905	
esent	0.65424238	0.17784853	-0.0587267	1	0.01727821	0.08736542	0.1952127	0.20716768	-0.0494032	0.02205234	0.03005405	0.02538854	0.00058857	-0.012483	0.01113924	-0.0177642	0.01935263	-0.0535405	0.01414509	-0.0220171	-0.0253338	0.03740059	
eopenrate	0.22641171	-0.0311014	-0.5096255	0.01727821	1	0.64062933	-0.0448647	0.07082715	0.3221832	0.15992608	0.12137939	0.03323553	0.01605104	-0.0184908	-0.0113597	-0.0135735	0.03055452	-0.0732942	0.06540027	-0.1231067	0.14800717	-0.0538923	
eclickrate	0.27154204	0.0364308	-0.5299321	0.08736542	0.64062933	1	-0.0370853	0.11913423	0.29530348	0.17082807	0.12936248	0.03151756	0.00923892	-0.0177646	-0.0098637	-0.0066619	0.03080703	-0.0729813	0.06264425	-0.1047614	0.17727966	-0.0969241	
avgorder	0.00326066	0.2363255	0.12769493	0.1952127	-0.0448647	-0.0370853	1	0.26400284	-0.2192793	-0.0781957	0.04031802	0.03752084	-0.0113031	0.00965512	-0.0029816	-0.0533701	0.00089312	0.04734784	0.02548201	0.09487269	-0.0998652	-0.0075853	
ordfreq	0.02448814	0.13333478	-0.070482	0.20716768	0.07082715	0.11913423	0.26400284	1	-0.1469134	0.07961534	0.17724667	0.09830709	0.03676868	0.01737748	-0.0372962	-0.0294009	0.0271021	0.05635591	-0.1358587	0.09678748	-0.2306491	0.21390635	-0.0298065
paperless	0.17716572	-0.2205856	-0.5784385	-0.0494032	0.3221832	0.29530348	-0.2192793	-0.1469134	1	0.17724667	0.09830709	0.03676868	0.01737748	-0.0372962	-0.0294009	0.0271021	0.05635591	-0.1358587	0.09678748	-0.2306491	0.21390635	-0.0298065	
refill	0.10280088	-0.0734856	-0.3032702	0.02205234	0.15992608	0.17082807	-0.0781957	0.07961534	0.17724667	1	0.203284	0.02416181	-0.0129467	-0.0109535	-0.0059264	0.02714109	-0.0115603	-0.045238	-0.0147143	-0.0199068	-0.0464085	0.06869749	
doorstep	0.06880409	-0.0411124	-0.2483174	0.03005405	0.12137939	0.12936248	0.04031802	0.13148667	0.09830709	0.203284	1	0.03409021	-0.0286317	-0.0071726	-0.0141982	0.00665892	0.02601052	-0.0141811	0.03159561	-0.1034028	0.20242653	-0.1068941	
Monday	0.01626172	0.0174267	-0.0631625	0.02538854	0.03323553	0.03151756	0.03752084	0.04254854	0.03676868	0.02416181	0.03409021	1	-0.2810406	-0.2296243	-0.2412329	-0.241753	-0.1098371	-0.0764588	-0.0907362	0.07357877	-0.0463495	0.01747088	
Tuesday	-0.010114	0.03191501	-0.0072657	0.00058857	0.01605104	0.00923892	-0.0113031	-0.0018713	0.01737748	-0.0129467	-0.0129467	-0.0286317	1	-0.2286701	-0.2402305	-0.2407484	-0.1093807	-0.076141	0.03291174	-0.016654	-0.0057822	0.00533008	
Wednesday	-0.0115532	0.02667866	0.04460094	-0.012483	-0.0184908	-0.0177646	0.00965512	-0.0020543	-0.0372962	-0.0109535	-0.0071726	-0.2296243	-0.2286701	1	-0.1962804	-0.1967035	-0.0893695	-0.062211	-0.0800067	0.0474227	-0.0155345	0.00810627	
Thursday	-0.0041813	-0.0047364	0.03802079	0.01113924	-0.0113597	-0.0098637	-0.0029816	-0.0157496	-0.0294009	-0.0059264	-0.0141982	-0.2412329	-0.2402305	-0.1962804	1	-0.2066478	-0.0938876	-0.0653561	0.07020546	-0.0048728	-0.0169669	-0.0128941	
Friday	0.01458463	-0.0213925	-0.0248531	-0.0177642	-0.0135735	-0.0066619	-0.0533701	-0.0332979	0.0271021	0.02714109	0.00665892	-0.241753	-0.2407484	-0.1967035	-0.2066478	1	-0.09409	-0.065497	-0.0962972	-0.0054112	-0.0587415	0.10526548	
Saturday	0.02409256	8.01E-05	-0.0488665	0.01935263	0.03055452	0.03080703	0.00089312	-0.0338336	0.05635591	-0.0115603	0.02601052	-0.1098371	-0.1093807	-0.0893695	-0.0938876	-0.09409	1	-0.0297577	0.28250022	-0.1283171	0.1573302	-0.1604441	
Sunday	-0.0500025	-0.1439395	0.12534322	-0.0535405	-0.0732942	-0.0729813	0.04734784	0.0644426	-0.1358587	-0.045238	-0.0141811	-0.0764588	-0.076141	-0.062211	-0.0653561	-0.065497	-0.0297577	1	0.04955344	-0.0819457	0.17437985	-0.1093	
BWI	0.04754149	0.01189322	-0.0989549	0.01414509	0.06540027	0.06264425	0.02548201	-0.0138733	0.09678748	-0.0147143	0.03159561	-0.0907362	0.03291174	-0.0800067	0.07020546	-0.0962972	0.28250022	0.04955344	1	-0.1532942	-0.1526049	-0.1883342	
CHO	-0.0893459	0.10732328	0.19435277	-0.0220171	-0.1231067	-0.1047614	0.09487269	0.10068712	-0.2306491	-0.0199068	-0.1034028	0.07357877	-0.016654	0.0474227	-0.0048728	-0.0054112	-0.1283171	-0.0819457	-0.1532942	1	-0.3971571	-0.4901435	
DCX	0.05757701	-0.0651524	-0.217767	-0.0253338	0.14800717	0.17727966	-0.0998652	-0.0523605	0.21390635	-0.0464085	0.20242653	-0.0463495	-0.0057822	-0.0155345	-0.0169669	-0.0587415	0.1573302	0.17437985	-0.1526049	-0.3971571	1	-0.4879395	
RIC	0.00722163	-0.0450574	0.06825905	0.03740059	-0.0538923	-0.0969241	-0.0075853	-0.0385795	-0.0298065	0.06869749	-0.1068941	0.01747088	0.00533008	0.00810627	-0.0128941	0.10526548	-0.1604441	-0.1093	-0.1883342	-0.4901435	-0.4879395	1	

NO CORRELATION &gt; 70%

# LOGISTIC REGRESSION - TRAINING

Dependent Variable (y) = retained

## First regression

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.8775239  0.1973918  -9.512 < 2e-16 ***
esent        0.2131455  0.0033232  64.138 < 2e-16 ***
Days.between.creation.and.first.order -0.0015497  0.0002756  -5.623 1.88e-08 ***
Days.since.last.order  0.0002078  0.0001243   1.671 0.094657 .
eopenrate    0.0097459  0.0009433  10.332 < 2e-16 ***
avgorder    -0.0031019  0.0007520  -4.125 3.71e-05 ***
ordfreq     -0.5545270  0.2313585  -2.397 0.016538 *
paperless1  0.4026440  0.1018634   3.953 7.72e-05 ***
refill1     0.7939441  0.1071259   7.411 1.25e-13 ***
doorstep1   0.8218779  0.1657724   4.958 7.13e-07 ***
favdayMonday -0.3901440  0.0900264  -4.334 1.47e-05 ***
favdaySaturday -0.3008229  0.1624990  -1.851 0.064137 .
favdaySunday -1.1346789  0.2622520  -4.327 1.51e-05 ***
favdayThursday -0.3358937  0.0953105  -3.524 0.000425 ***
favdayTuesday -0.4581017  0.0895081  -5.118 3.09e-07 ***
favdayWednesday -0.2567542  0.0958297  -2.679 0.007378 **
cityCHO     -0.4539297  0.1480720  -3.066 0.002172 **
cityDCX    -0.5358201  0.1385058  -3.869 0.000109 ***
cityRIC    -0.6084798  0.1429976  -4.255 2.09e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

## Second regression (removed variables with high p-values + city + day)

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.5332408  0.0792574 -31.962 < 2e-16 ***
esent        0.2108716  0.0032737  64.414 < 2e-16 ***
Days.between.creation.and.first.order -0.0015413  0.0002700  -5.709 1.13e-08 ***
eopenrate    0.0091655  0.0008847  10.361 < 2e-16 ***
avgorder    -0.0034284  0.0007553  -4.539 5.65e-06 ***
paperless1  0.3263693  0.0705521   4.626 3.73e-06 ***
refill1     0.7835126  0.1045612   7.493 6.71e-14 ***
doorstep1   0.8259425  0.1601697   5.157 2.51e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

R-Squared  
(Cox-Snell)  
0.4885567

# LOGISTIC REGRESSION - TEST

Dependent Variable (y) = retained

PREDICTED	HIT RATE	OBSERVED	
		0	1
0	1090	197	
1	174	4761	

$$\text{ACCURACY} = (1090 + 4761) / (1090 + 197 + 174 + 4761)$$

**ACCURACY = 94.0%**

# RANDOM FOREST

```
Call:
  randomForest(x = retailnumericRF[, -1], y = retailnumericRF[, 1])
```

```
  Type of random forest: classification
```

```
  Number of trees: 500
```

```
No. of variables tried at each split: 2
```

```
  OOB estimate of error rate: 4.47%
```

```
Confusion matrix:
```

```

  0  1 class.error
0 4116  917 0.182197497
1  179 19327 0.009176664
```

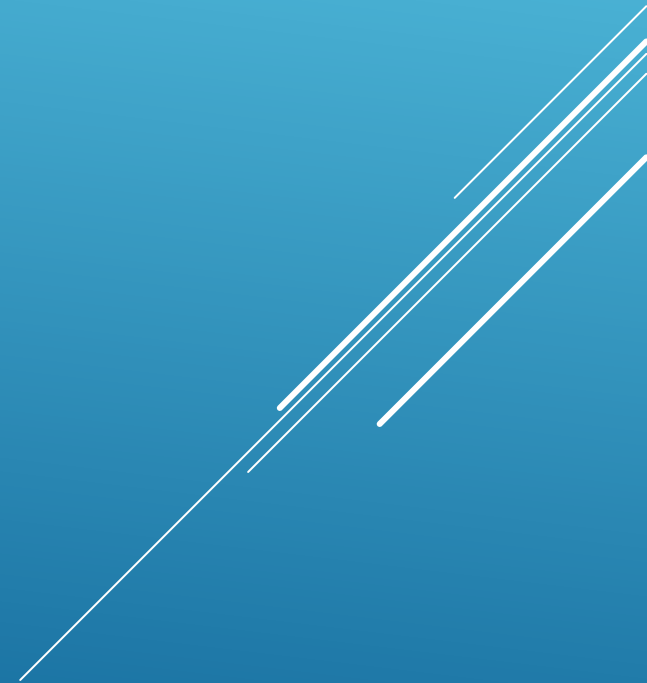
	MeanDecreaseGini
Days.between.creation.and.first.order	202.94564
esent	4875.70160
eopenrate	914.05265
avgorder	276.83291
paperless	126.12651
refill	44.34797
doorstep	23.09119

95.5% ACCURACY ON TRAINING DATA

# CONCLUSION

- ❑ Most important independent variables
  - ❑  $E_{sent}$  = Number of e-mails sent
  - ❑  $E_{openrate}$  = Number of e-mails opened divided by number of emails sent
- ❑ Recommendations:
  - ❑ Retail Relay should invest to increase their e-mail marketing program
  - ❑ Track e-mail campaign effectiveness
    - ❑ Constant Contact, MailChimp, etc.

BACKUP SLIDES



## Assignment Questions:

1. Use the Relay train data to develop a model to predict customer retention. You may use logistic regression to predict the variable "retained." You can use any combination of the independent variables available in the data to obtain a model with the best predictive ability and usability. You are free to use different transformations and combinations of the independent variables. Be aware that there is no "magic bullet" to finding the ideal model. You will have to go through multiple iterations.
2. Once you obtain the best model that you can find, predict retention in the test data. You will use the coefficients obtained from the model estimated using the train data to do this. Name this predicted value "pretrain."
3. Calculate the hit rate. This can be calculated as % of matches between the value of pre-train and retained in the train data.
4. Be prepared to present your results in class. The team with the best model, as judged by your peers, will win "bragging rights"