DRAFT

MOBILE APP LOGIN FAILURE ANALYSIS & MACHINE LEARNING PREDICTION

Springboard.com Capstone Project for Foundations of Data Science

By: Donald Gennetten

May 2017

DRAFT

Problem & Business Value

PROBLEM:

With increasing speed to market pressures, technology delivery teams are continually prioritizing feature development against technical debt. This results in problems detecting, measuring and resolving issues impacting end users.

With more than 125 million opportunities to fail each month, a 1% failure rate would equate to over 1.24 million negative customer experiences. This significantly increases customer dissatisfaction and may reduce market share.

BUSINESS VALUE:

The goal of this project is to identity iOS Mobile App login issues that may not be clearly evident to the business, establish correlations with problematic devices, and offer machine learning prediction on statistically significant predictors.

- **Business/Product Owners** will have improved visibility of issues allowing them to refine their delivery roadmap and drive prioritization of technical debt and other fixes which impact end users.

- **Platform/Technology/DevOps** teams will be able to identity production support, capacity and infrastructure needs.

Approach

DRAFT

1. Collected hourly login volumes for April 2017.
2. Obtained manufacturer device data for lookup and join purposes.
3. Imported, wrangled and joined the data in R for iOS Devices.
4. Explored the data, comparing available dimensions with login results (Success, Policy and Failure).
5. Investigated patterns, disproportionate rates, and any other notable observations.
6. Developed Logistical Machine Learning Model to predict failures.
7. Provided summary of results with proposed next steps.

DRAFT

Data

Data was extracted from APIs, login activity logs, and publicly available device manufacturer lists.

LIMITATIONS:

- Steps were taken to ensure sensitive and proprietary data was not included in raw data files. Impact on the final output for this project was minimal.
- Login attempts resulting in fatal device level failures, and/or where there was no connection to the API, will not be reflected in the data. These failures will therefore not be reflected in final results.
- Login volumes were aggregated hourly. This is believed to have low impacting on the final output as weighting was applied where necessary.

CLEANING & WRANGLING:

- Login data was collected and aggregated using Splunk and internal data warehouse sources.
- Device detail was collected from available online sources.
- Wrangling, joining, analysis, summarization and visualization was conducted in Rstudio.

Data: Important Categorical Fields*

DRAFT

Important categorical fields were identified in the data collection plan.

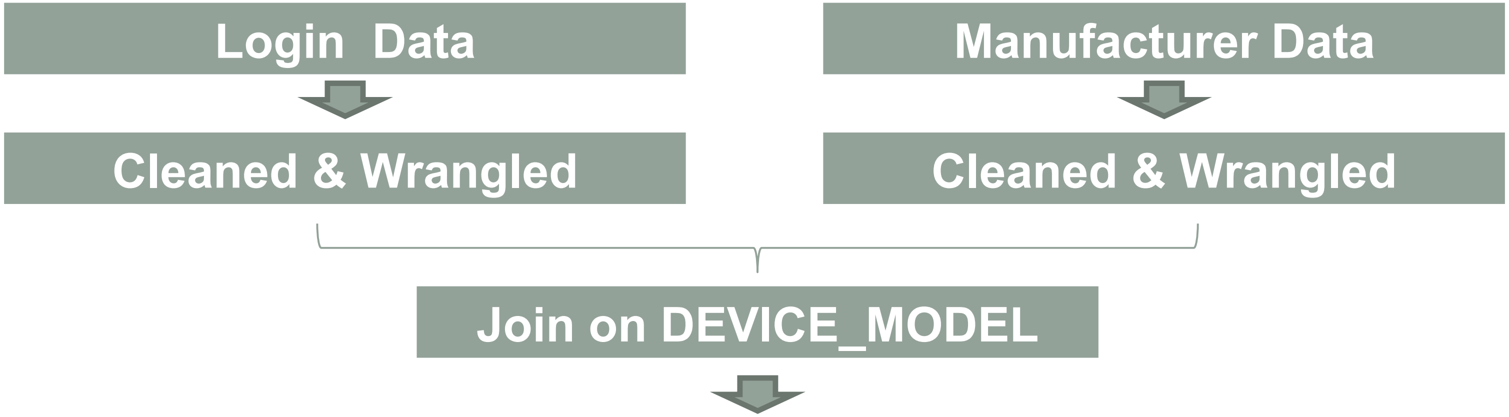
Source	Field Name	Sample Values	Definition
Internal API & Activity Logs	APP_VERSION	8.28.1, 9.16.0, 9.15.0	Code version for the installed mobile application
	AUTH_METHOD	Password, Finger Print, Pattern	Method used by the user to authenticate
	CHANNEL__TYPE	MOBILE, WEB	Channel used by the customer during Login. Always expected to be “MOBILE”.
	DEVICE_OPERATING_SYSTEM	iOS, iPhone OS	Operating system installed on the mobile device
	DEVICE_OPERATING_SYSTEM_VERSION	10.2.1, 6.0.1, 9.3	Operating system version installed on the mobile device
	APP_TYPE	iPhone, iPad	App type installed on the device
	RESULT_DISPOSITION	SUCCESS, POLICY, DEFECT	General business result from a login attempt. SUCCESS = Successful login, DEFECT = Failed login due to technical issue, POLICY = Failed login due to business rule (Ex: Invalid Password)
Both	DEVICE_MODEL	iPhone5,3, iPhone8,1	Unique device model identifier. Used as lookup to get friendly product names
Manufacturer Information	FRIENDLY_PRODUCT_NAME	iPhone 6, iPhone 6s Plus, iPad mini	Commonly recognized marketing device names established by the Manufacturer

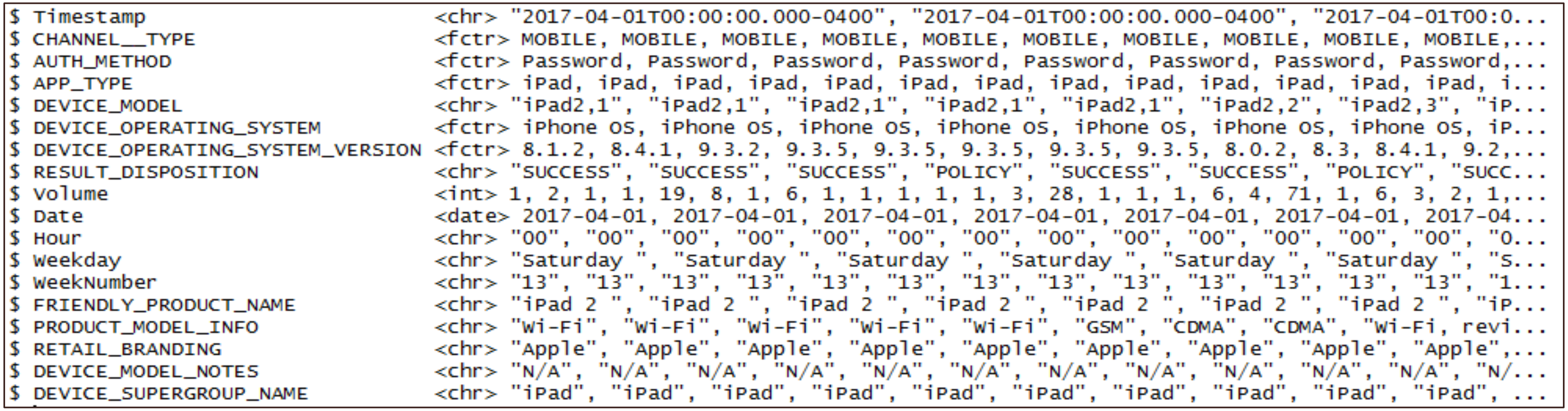
* This is a subset of the total data and represents the top categorical values used in analysis and prediction.

DRAFT

Collection, Wrangling & Joining

Data from 82MM iOS login attempts was wrangled and joined with device information into a single data frame.



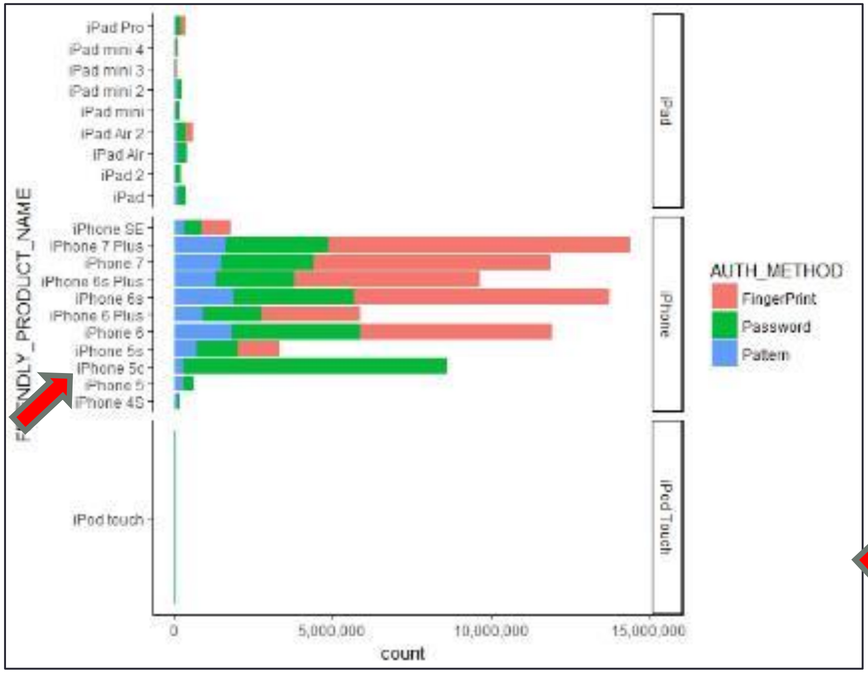


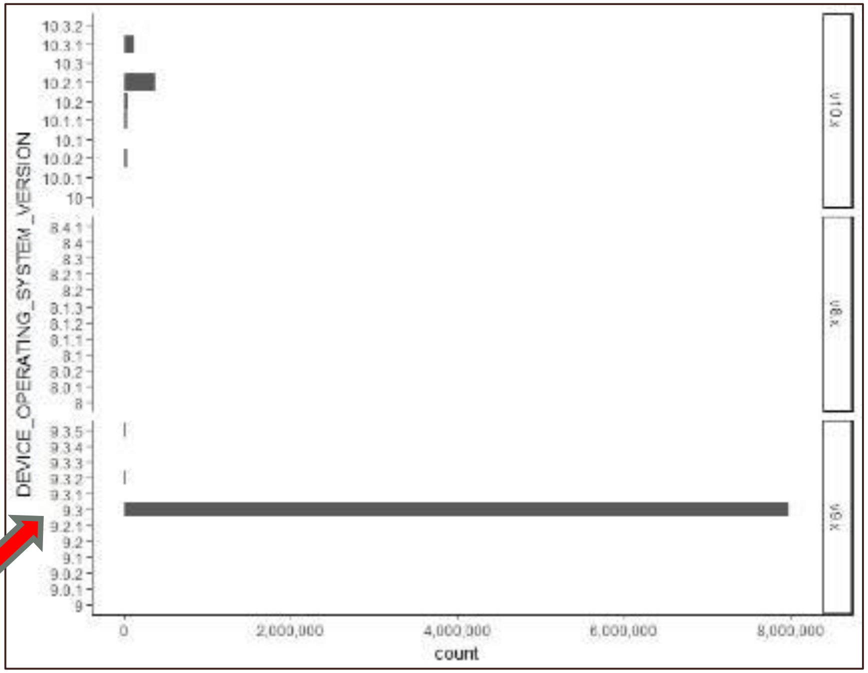
```
$ Timestamp                       <chr> "2017-04-01T00:00:00.000-0400", "2017-04-01T00:00:00.000-0400", "2017-04-01T00:0...
$ CHANNEL__TYPE                   <fctr> MOBILE, MOBILE, MOBILE, MOBILE, MOBILE, MOBILE, MOBILE, MOBILE, MOBILE, MOBILE,...
$ AUTH_METHOD                     <fctr> Password, Password, Password, Password, Password, Password, Password, Password,...
$ APP_TYPE                        <fctr> iPad, iPad, iPad, iPad, iPad, iPad, iPad, iPad, iPad, iPad, iPad, iPad, iPad, i...
$ DEVICE_MODEL                    <chr> "iPad2,1", "iPad2,1", "iPad2,1", "iPad2,1", "iPad2,1", "iPad2,2", "iPad2,3", "iP...
$ DEVICE_OPERATING_SYSTEM         <fctr> iPhone OS, iPhone OS, iPhone OS, iPhone OS, iPhone OS, iPhone OS, iPhone OS, iP...
$ DEVICE_OPERATING_SYSTEM_VERSION <fctr> 8.1.2, 8.4.1, 9.3.2, 9.3.5, 9.3.5, 9.3.5, 9.3.5, 9.3.5, 8.0.2, 8.3, 8.4.1, 9.2,...
$ RESULT_DISPOSITION              <chr> "SUCCESS", "SUCCESS", "SUCCESS", "POLICY", "SUCCESS", "SUCCESS", "POLICY", "SUCC...
$ Volume                          <int> 1, 2, 1, 1, 19, 8, 1, 6, 1, 1, 1, 1, 1, 3, 28, 1, 1, 1, 6, 4, 71, 1, 6, 3, 2, 1,...
$ Date                            <date> 2017-04-01, 2017-04-01, 2017-04-01, 2017-04-01, 2017-04-01, 2017-04-01, 2017-04...
$ Hour                            <chr> "00", "00", "00", "00", "00", "00", "00", "00", "00", "00", "00", "00", "00", "0...
$ Weekday                         <chr> "Saturday ", "Saturday ", "Saturday ", "Saturday ", "Saturday ", "Saturday ", "S...
$ WeekNumber                      <chr> "13", "13", "13", "13", "13", "13", "13", "13", "13", "13", "13", "13", "13", "1...
$ FRIENDLY_PRODUCT_NAME           <chr> "iPad 2 ", "iPad 2 ", "iPad 2 ", "iPad 2 ", "iPad 2 ", "iPad 2 ", "iPad 2 ", "iP...
$ PRODUCT_MODEL_INFO              <chr> "Wi-Fi", "Wi-Fi", "Wi-Fi", "Wi-Fi", "Wi-Fi", "GSM", "CDMA", "CDMA", "Wi-Fi, revi...
$ RETAIL_BRANDING                 <chr> "Apple", "Apple", "Apple", "Apple", "Apple", "Apple", "Apple", "Apple", "Apple",...
$ DEVICE_MODEL_NOTES              <chr> "N/A", "N/A", "N/A", "N/A", "N/A", "N/A", "N/A", "N/A", "N/A", "N/A", "N/A", "N/...
$ DEVICE_SUPERGROUP_NAME          <chr> "iPad", "iPad", "iPad", "iPad", "iPad", "iPad", "iPad", "iPad", "iPad", "iPad", ...
```

DRAFT

Preliminary Findings

~10% of all iOS logins are from iPhone 5c devices; ~93% of these are using an old iOS version (v9.3)



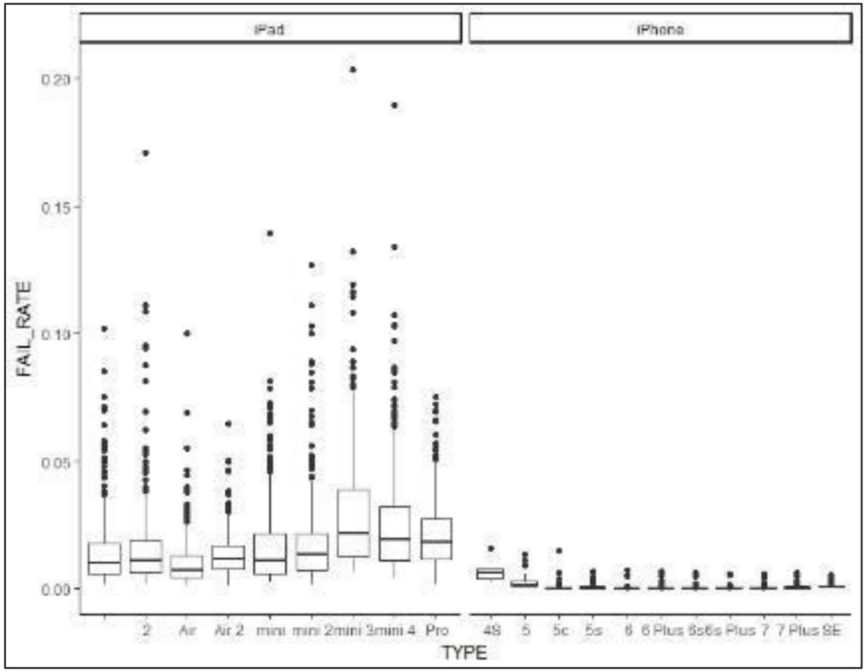


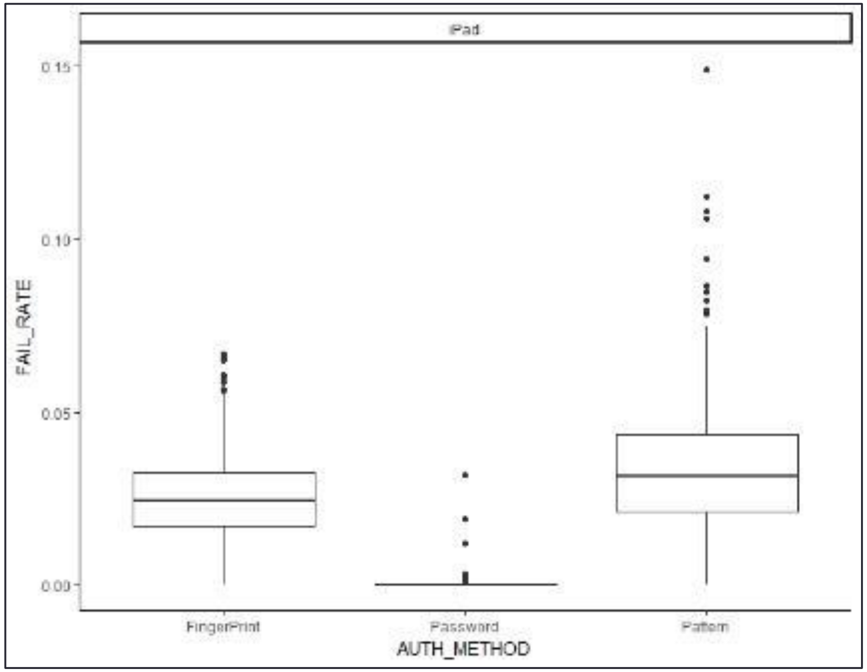
This may pose security risks to customers and the business. A risk assessment is suggested.

DRAFT

Preliminary Findings

iPad failure rates are significantly higher; This is predominant in Touch-Id and Pattern login methods.





Evaluation of iPad app code is suggested to determine root cause of disproportionate failure rates.

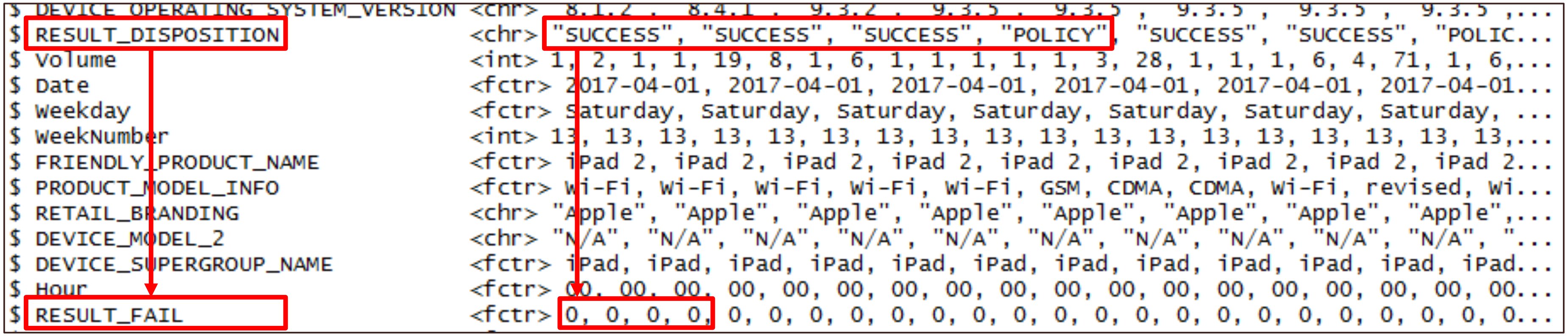
DRAFT

Logistic Regression

Data was prepped for Logistic Regression by converting the dependent variable to binomial

…Join on DEVICE_MODEL

Binomial Classification

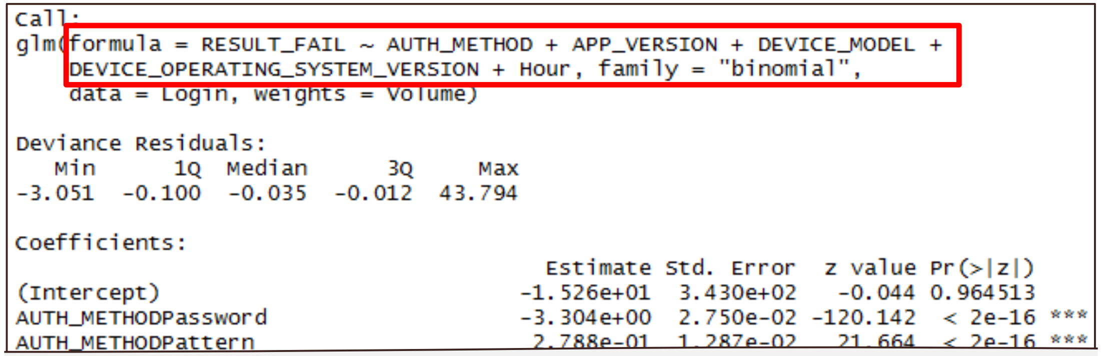


```
$ DEVICE_OPERATING_SYSTEM_VERSION <chr> "8.1.2", "8.4.1", "9.3.2", "9.3.5", "9.3.5", "9.3.5", "9.3.5", "9.3.5",...
$ RESULT_DISPOSITION              <chr> "SUCCESS", "SUCCESS", "SUCCESS", "POLICY", "SUCCESS", "SUCCESS", "POLIC...
$ Volume                          <int> 1, 2, 1, 1, 19, 8, 1, 6, 1, 1, 1, 1, 1, 3, 28, 1, 1, 1, 6, 4, 71, 1, 6,...
$ Date                            <fctr> 2017-04-01, 2017-04-01, 2017-04-01, 2017-04-01, 2017-04-01, 2017-04-01...
$ Weekday                         <fctr> Saturday, Saturday, Saturday, Saturday, Saturday, Saturday, Saturday, ...
$ WeekNumber                      <int> 13, 13, 13, 13, 13, 13, 13, 13, 13, 13, 13, 13, 13, 13, 13, 13, 13, 13,...
$ FRIENDLY_PRODUCT_NAME           <fctr> iPad 2, iPad 2, iPad 2, iPad 2, iPad 2, iPad 2, iPad 2, iPad 2, iPad 2...
$ PRODUCT_MODEL_INFO              <fctr> Wi-Fi, Wi-Fi, Wi-Fi, Wi-Fi, Wi-Fi, GSM, CDMA, CDMA, Wi-Fi, revised, Wi...
$ RETAIL_BRANDING                 <chr> "Apple", "Apple", "Apple", "Apple", "Apple", "Apple", "Apple", "Apple",...
$ DEVICE_MODEL_2                  <chr> "N/A", "N/A", "N/A", "N/A", "N/A", "N/A", "N/A", "N/A", "N/A", "N/A", "...
$ DEVICE_SUPERGROUP_NAME          <fctr> iPad, iPad, iPad, iPad, iPad, iPad, iPad, iPad, iPad, iPad, iPad, iPad...
$ Hour                            <fctr> 00, 00, 00, 00, 00, 00, 00, 00, 00, 00, 00, 00, 00, 00, 00, 00, 00, 00...
$ RESULT_FAIL                     <fctr> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...
```

Logistic Regression

Dependent variables were selected based on the lowest possible AIC in the coefficients summary for the full dataset

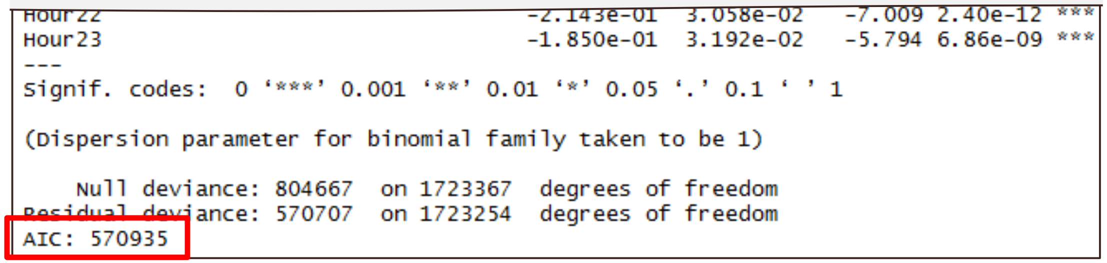
AIC	Dependent Variables
570,935	AUTH_METHOD + APP_VERSION + DEVICE_MODEL + DEVICE_OPERATING_SYSTEM_VERSION + Hour
571,821	AUTH_METHOD + FRIENDLY_PRODUCT_NAME + DEVICE_SUPERGROUP_NAME + APP_VERSION + DEVICE_MODEL + DEVICE_OPERATING_SYSTEM_VERSION
572,488	AUTH_METHOD + FRIENDLY_PRODUCT_NAME + DEVICE_SUPERGROUP_NAME + APP_VERSION + DEVICE_MODEL + Weekday
572,715	AUTH_METHOD + FRIENDLY_PRODUCT_NAME + DEVICE_SUPERGROUP_NAME + APP_VERSION + Weekday + DEV_MOD_FAIL,
635,627	DEVICE_MODEL
636,522	FRIENDLY_PRODUCT_NAME
641,277	DEVICE_SUPERGROUP_NAME
761,253	AUTH_METHOD
775,591	APP_VERSION
784,778	DEVICE_OPERATING_SYSTEM_VERSION
803,314	Hour



```
Call:
glm(formula = RESULT_FAIL ~ AUTH_METHOD + APP_VERSION + DEVICE_MODEL +
    DEVICE_OPERATING_SYSTEM_VERSION + Hour, family = "binomial",
    data = Login, weights = Volume)

Deviance Residuals:
   Min       1Q   Median       3Q      Max
-3.051   -0.100   -0.035   -0.012   43.794

Coefficients:
                                        Estimate Std. Error  z value Pr(>|z|)
(Intercept)                           -1.526e+01  3.430e+02   -0.044 0.964513
AUTH_METHODPassword                   -3.304e+00  2.750e-02 -120.142  < 2e-16 ***
AUTH_METHODPattern                     2.788e-01  1.287e-02   21.664  < 2e-16 ***
```



```
Hour22                                -2.143e-01  3.058e-02   -7.009 2.40e-12 ***
Hour23                                -1.850e-01  3.192e-02   -5.794 6.86e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 804667  on 1723367  degrees of freedom
Residual deviance: 570707  on 1723254  degrees of freedom
AIC: 570935
```

Converting statistically significant independent variables to binomial was not found to improve AIC or GLM

DRAFT

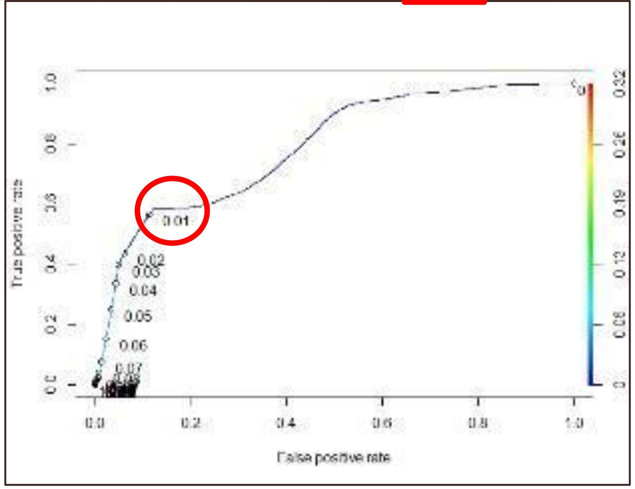
Logistic Regression: Training

Data split 65/35 for training. Threshold chosen based on ROC curve & Confusion Matrix results

Rates	Explanation	ROC Threshold						
		0.2	**0.1**	**0.03**	**0.02**	**0.01**	**0.009**	**0.003**
Accuracy	Overall, how often is the classifier correct?	98.66%	98.39%	94.33%	93.20%	88.30%	87.91%	86.69%
Misclassification	Overall, how often is it wrong?	1.34%	1.61%	5.67%	6.80%	11.70%	12.09%	13.31%
True Positive / **Sensitivity**	When it's actually login failure, how often does it predict failure?	0.02%	1.68%	39.89%	43.68%	56.48%	57.43%	58.45%
False Positive	When it's actually a login success or policy, how often does it predict failure?	0.01%	0.30%	4.93%	6.13%	11.27%	11.68%	12.92%
Specificity	When it's actually a login success or policy, how often does it predict success or policy?	99.99%	99.70%	95.07%	93.87%	88.73%	88.32%	87.08%
Precision	When it predicts login failure, how often is it correct?	3.37%	7.02%	9.85%	8.77%	6.34%	6.23%	5.76%
Prevalence	How often does the login failure actually occur in the sample?	1.33%	1.33%	1.33%	1.33%	1.33%	1.33%	1.33%

AUC = 0.7887759

n=	1,120,189	Predicted		
		Non-Failure	Failure	
Actual	Non-Failure	980,674	124,585	1,105,259
	Failure	6,498	8,432	14,930
		987,172	133,017	



DRAFT

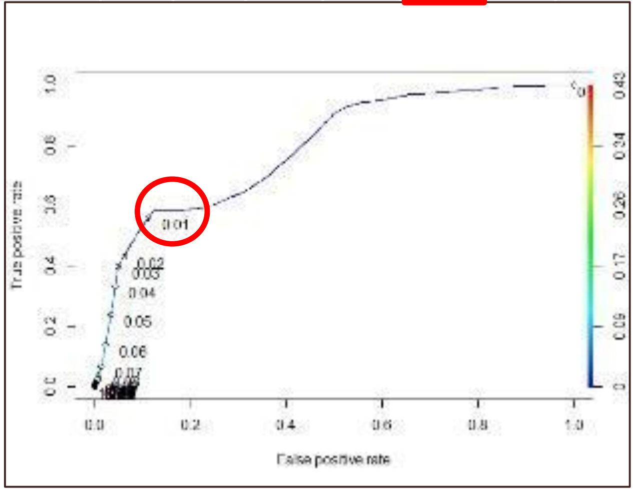
Logistic Regression: Testing

Running the Logistic Regression Model on testing data provided nearly identical results.

Rates	Explanation	ROC Threshold						
		0.2	**0.1**	**0.03**	**0.02**	**0.01**	**0.009**	**0.003**
Accuracy	Overall, how often is the classifier correct?	98.66%	98.39%	94.36%	93.23%	88.36%	87.97%	86.78%
Misclassification	Overall, how often is it wrong?	1.34%	1.61%	5.64%	6.77%	11.64%	12.03%	13.22%
True Positive / **Sensitivity**	When it's actually login failure, how often does it predict failure?	0.00%	1.60%	39.67%	43.16%	55.90%	56.91%	58.33%
False Positive	When it's actually a login success or policy, how often does it predict failure?	0.01%	0.30%	4.90%	6.09%	11.20%	11.61%	12.83%
Specificity	When it's actually a login success or policy, how often does it predict success or policy?	99.99%	99.70%	95.10%	93.91%	88.80%	88.39%	87.17%
Precision	When it predicts login failure, how often is it correct?	0.00%	6.77%	9.86%	8.73%	6.32%	6.21%	5.79%
Prevalence	How often does the login failure actually occur in the sample?	1.33%	1.33%	1.33%	1.33%	1.33%	1.33%	1.33%

AUC = 0.7892382

n=	603,179	Predicted		
		Non-Failure	Failure	
Actual	Non-Failure	528,504	66,636	595,140
	Failure	3,545	4,494	8,039
		532,049	71,130	



DRAFT

Summary & Proposed Next Steps

SUMMARY:

This model has been developed and tested to provide the following results:

- An overall accuracy rate of 88% for any login result.
- A true positive rate (login failure predicted as failure) of 56%.
- A false positive rate (login success or policy predicted as failure) of 11%.
- Increasing the true positive rate by 2% is possible but false positive rates will increase by 1.5%

NEXT STEPS:

- Improve prediction capabilities thru Feature Engineering production incident and user level data.
- Predict fatal app crash errors thru measuring device mix proportions to determine what is missing.
- Improve true positive rates by excluding failures associated with special cause incidents then re-run Logistic Model training.
- Expand analysis to Android Devices.