Persona Verification

Aaron Childress

Agenda

Assess Overall Model Effectiveness

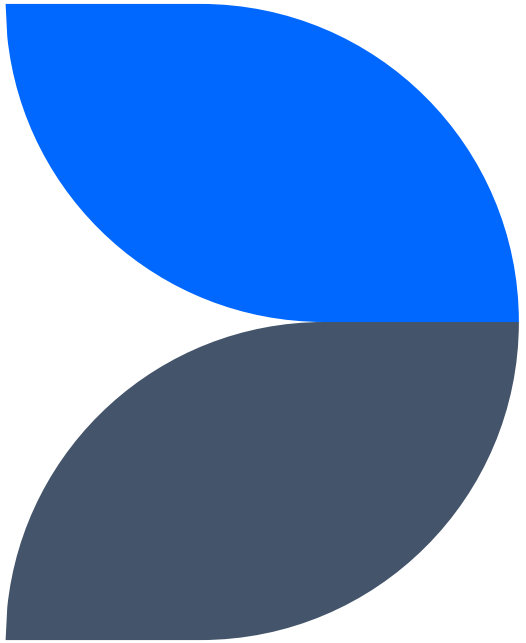
Assess Individual Check Effectiveness

Highlight Meaningful End User Trends

Summary

Model Effectiveness

Overall & Individual Checks



Approach

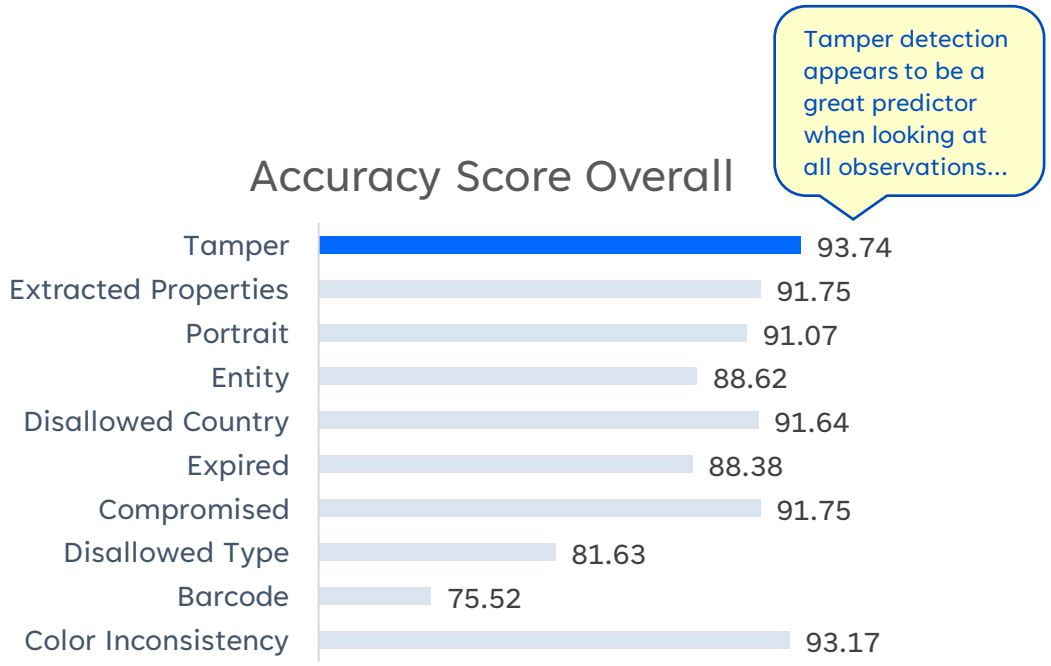
I decided to focus on records where the ground truth was observable (no image quality issues). This represents ~86% of the data. Further, this is an imbalanced classification problem, so I used relevant measures to score the model – precision and recall.

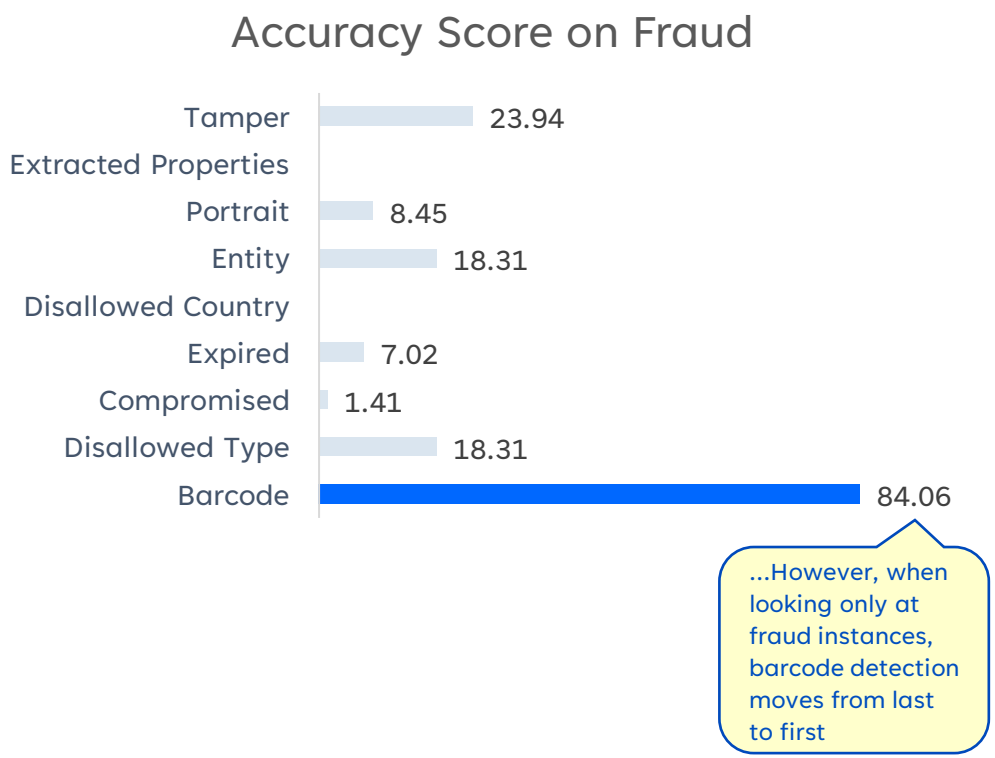
Model Goal: Minimize False Negatives

	Precision	Recall	F1-score	Support
No Fraud	0.98	0.60	0.74	793
Fraud	0.16	0.87	0.27	71
Accuracy			0.62	864
Macro Avg	0.57	0.74	0.51	864
Weighted Avg	0.91	0.62	0.70	864

- The least desirable outcome is an incidence of fraud that is predicted as not fraud
- In this respect the model does a better job than the accuracy score would suggest
- 87% of the positive fraud observations were correctly classified

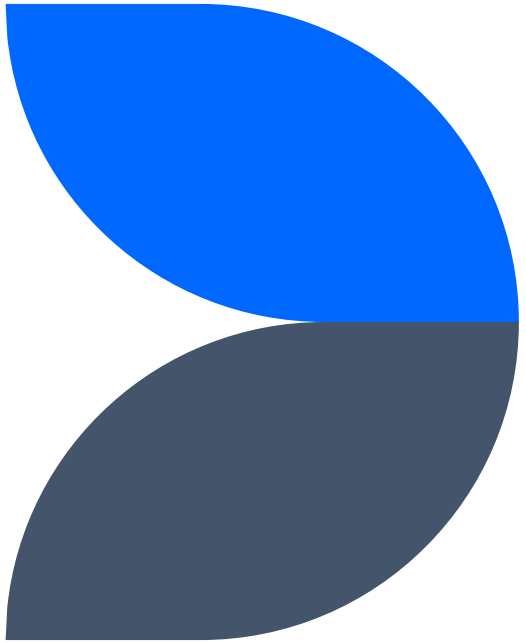
Best Individual Check Surprises



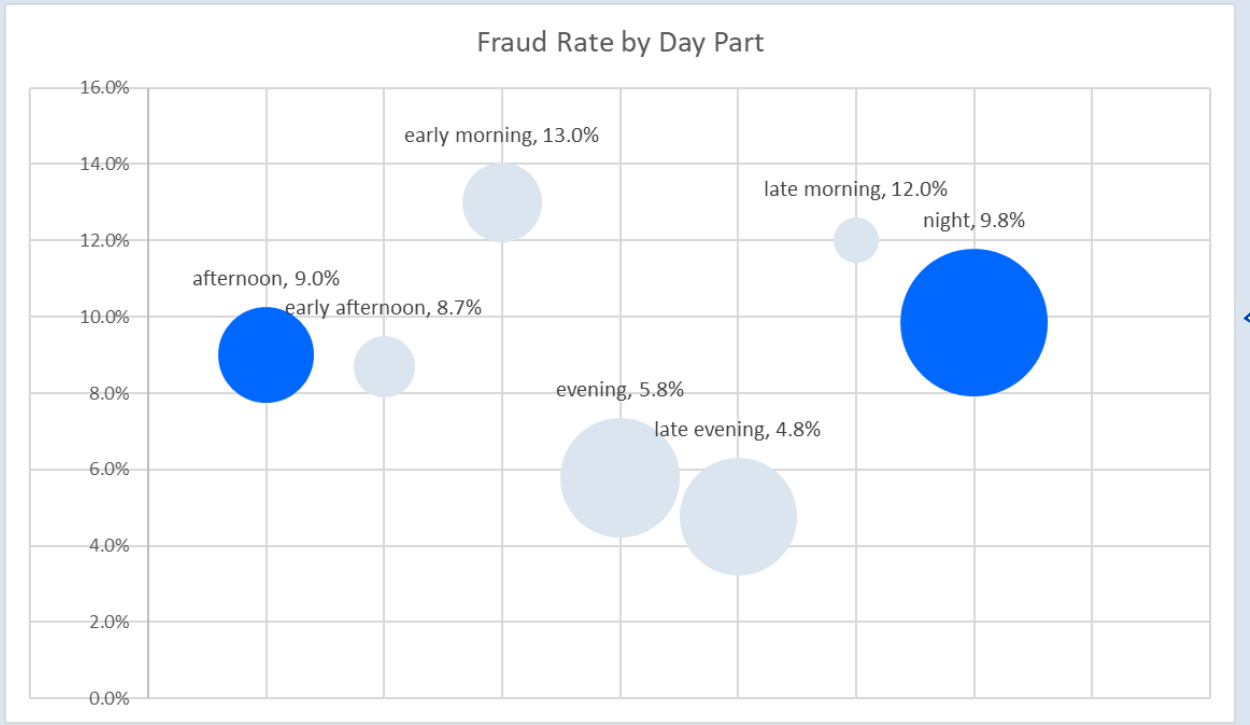


End User Analysis

Device, Calendar & Location

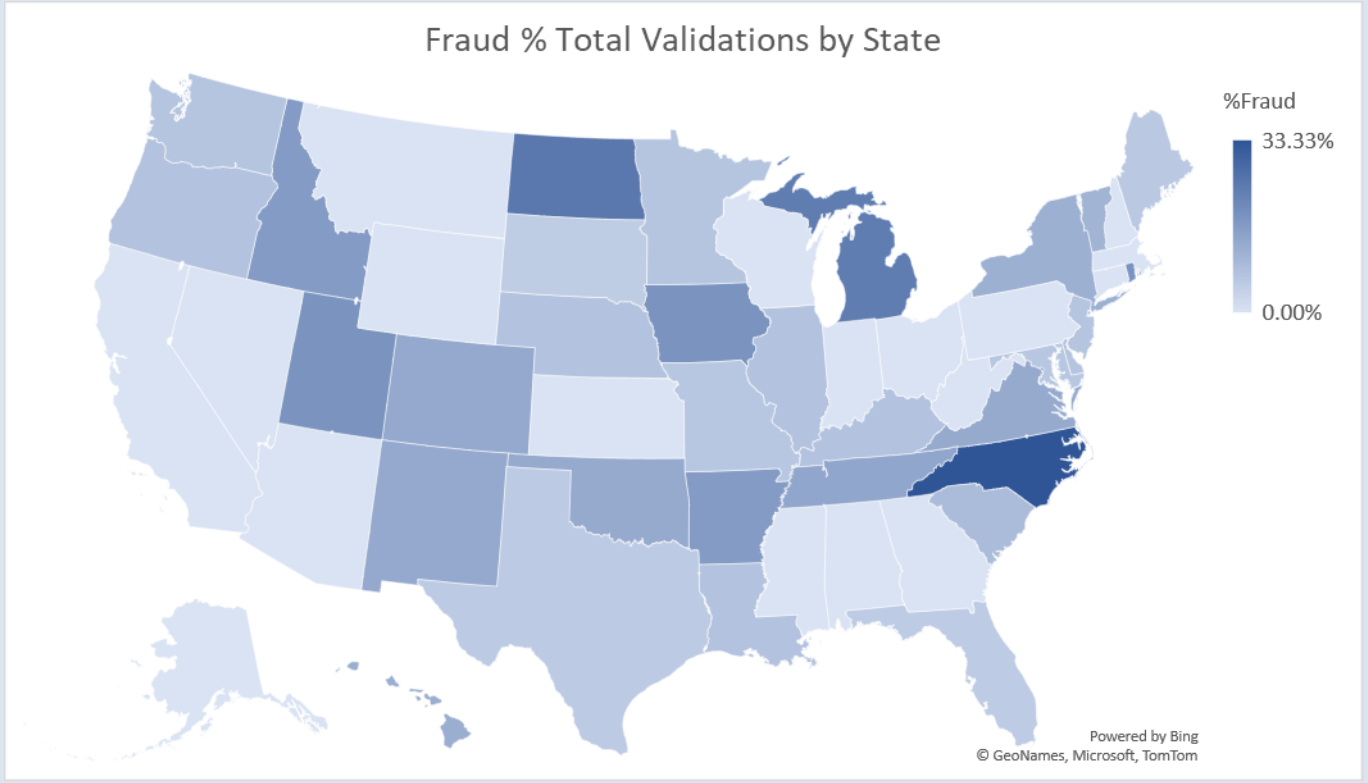


Isolate Day Parts to Address Spikes

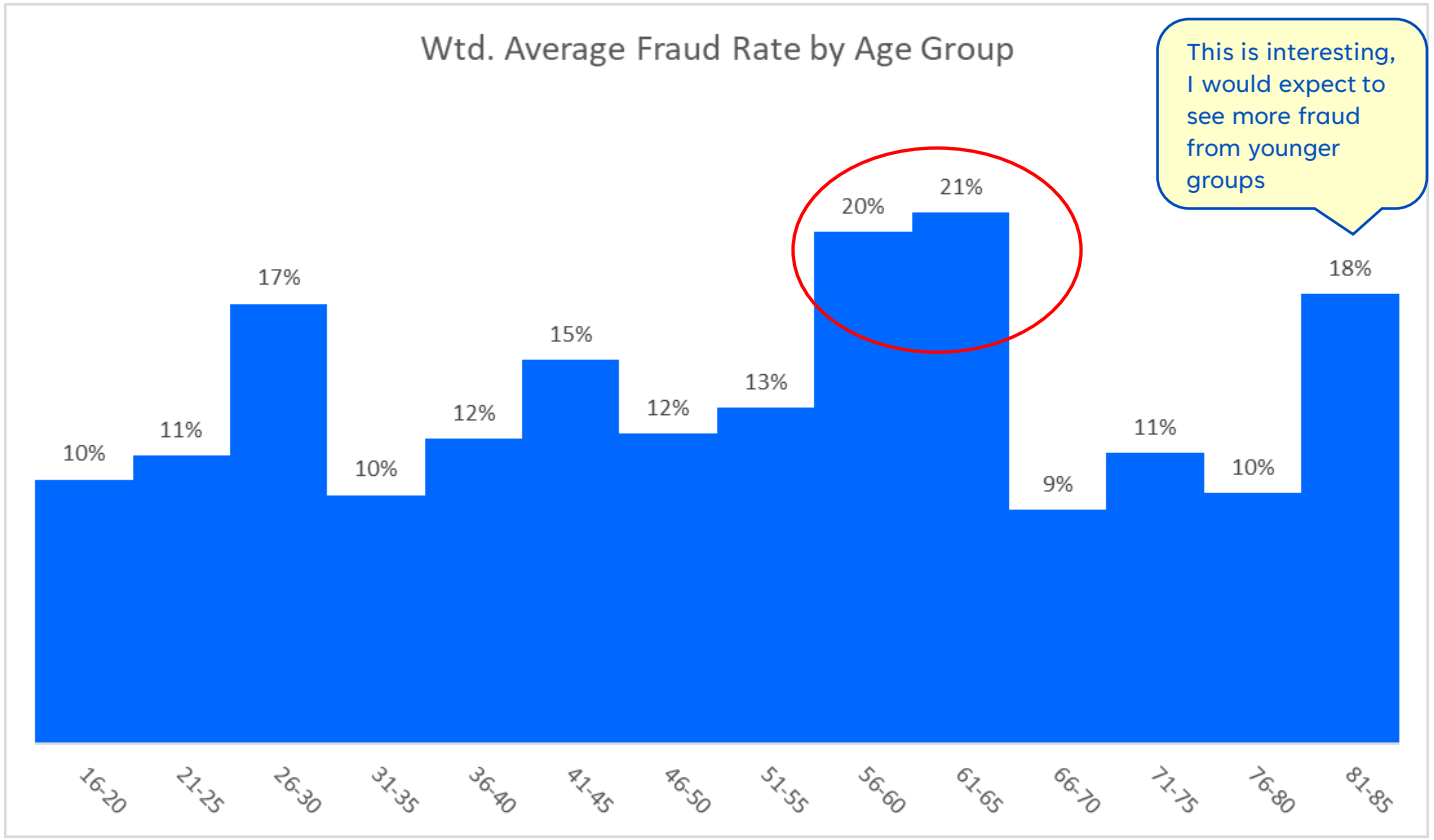


While early and late morning have the highest fraud rates, it could be more impactful to focus on night and afternoon as there's larger volume.

DC and NC have the highest fraud rates at 33%; ND and MI come in at ~25%



Looking into a few age groups with above average rates could be worthwhile



Device and Calendar Based Spikes

1

Device

73% of all traffic comes from mobile, but it accounts for %83 of fraud – a 10ppt over-index

2

OS

39% of all fraud comes from Android vs 33% of traffic overall. Moreover, 10% of Android traffic is fraudulent

3

Browser

19% of Safari traffic on mobile is fraudulent. Safari's pct composition of fraud is 2x its pct composition of non-fraud

4

Month

There's a spike in fraud during Dec – 11.5% vs 8% average for other months. However, Dec sample is small (~3% of data). This might normalize with larger sample.

5

Weekday

Sunday fraud rate is an outlier – 13.8% vs 7.5% average for other weekdays. And volume is sufficient to warrant a closer look

Key Take Aways

In summary, Persona's automated solution is reasonably effective at catching fraud as 87% of the positive fraud observations were correctly classified. Barcodes are the best individual check. Regarding End Users, isolating day parts and unpacking device information is the likeliest means to effectively address fraud.