Car Insurance Claim Predictor Machine Learning Model & Dashboard

Car Insurance Claim Predictor

On The Road Car Insurance

The Car Insurance Claim Predictor dashboard provides insight for managers on the underwriter's risk assessing abilities and the profitability of the business. It aims to give an overview of a simple model (one feature) to predict whether a customer will make a claim on their insurance during the policy period.

Industry
Insurance

Persona/Audience
- Pricing, UW Managers

Data Sources:
- Encoded client dataset in csv
- Data from Datacamp

Key Business Problem & Mandate

- Which customer characteristic in the current database (aka feature) can be used to predict accurately the likelihood of making a claim during the policy period?

- **Mandate:** To build a simple model with a single predictor feature, due to constraints in machine learning expertise and deployment capabilities.

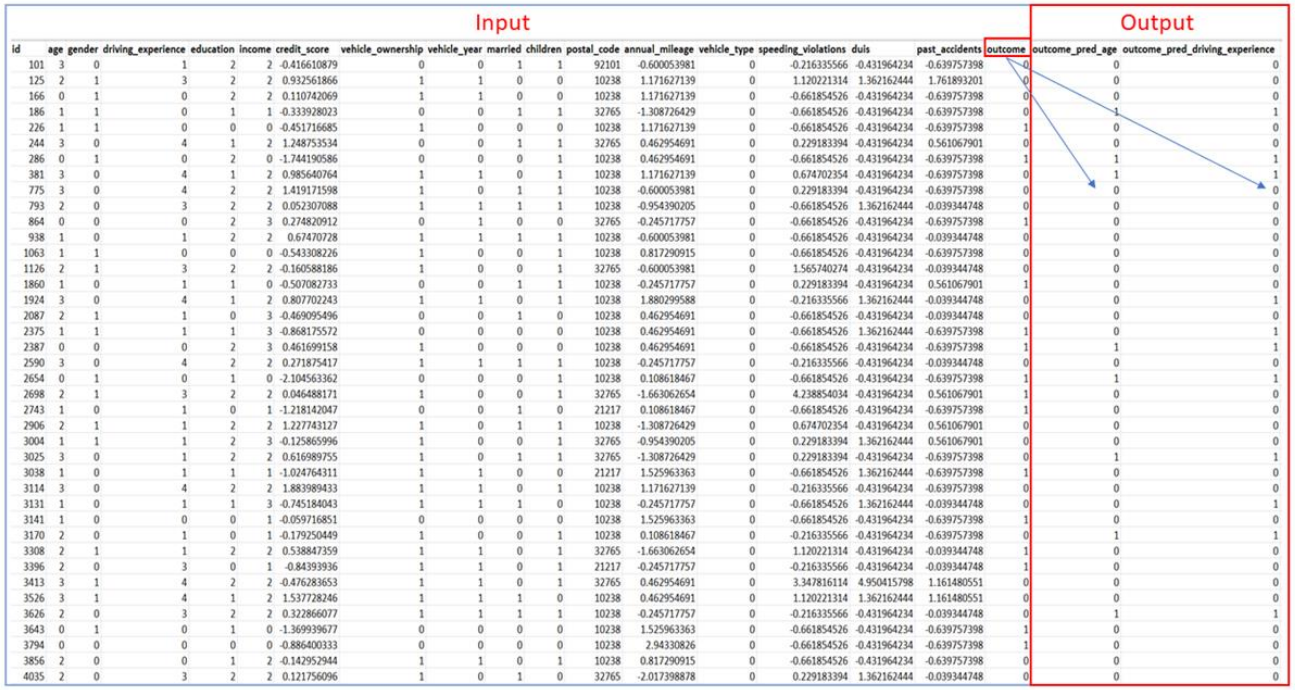
- **Key metrics for model evaluation:** Accuracy

- **Business impacts:** A crucial role in the Pricing and Underwriting strategies of the company

Technical

Version: 2023.1
Supported Layouts: Desktop
Contact: www.ontheroadinsurance.com

Claim Predictor Model Approach



Input / Output

id	age	gender	driving_experience	education	income	credit_score	vehicle_ownership	vehicle_year	married	children	postal_code	annual_mileage	vehicle_type	speeding_violations	duis	past_accidents	outcome	outcome_pred_age	outcome_pred_driving_experience
101	3	0	1	2	2	-0.416610879	0	0	1	1	92101	-0.600053981	0	-0.216335566	-0.431964234	-0.639757398	0	0	0
125	2	1	3	2	2	0.932561866	1	1	0	0	10238	1.171627139	0	1.120221314	1.362162444	1.761893201	0	0	0
166	0	1	0	2	2	0.110742069	1	1	0	0	10238	1.171627139	0	-0.661854526	-0.431964234	-0.639757398	0	0	0
186	1	1	0	1	1	-0.333928023	0	0	1	1	32765	-1.308726429	0	-0.661854526	-0.431964234	-0.639757398	0	1	1
226	1	1	0	0	0	-0.451716685	1	0	0	0	10238	1.171627139	0	-0.661854526	-0.431964234	-0.639757398	1	0	0
244	3	0	4	1	2	1.248753534	0	0	1	1	32765	0.462954691	0	0.229183394	-0.431964234	0.561067901	0	0	0
286	0	1	0	2	0	-1.744190586	0	0	0	1	10238	0.462954691	0	-0.661854526	-0.431964234	-0.639757398	1	1	1
381	3	0	4	1	2	0.985640764	1	1	0	1	10238	1.171627139	0	0.674702354	-0.431964234	-0.639757398	0	1	1
775	3	0	4	2	2	1.419171598	1	0	1	1	10238	-0.600053981	0	0.229183394	-0.431964234	-0.639757398	0	0	0
793	2	0	3	2	2	0.052307088	1	1	1	1	10238	-0.954390205	0	-0.661854526	1.362162444	-0.039344748	0	0	0
864	0	0	0	2	3	0.274820912	0	1	0	0	32765	-0.245717757	0	-0.661854526	-0.431964234	-0.639757398	1	0	0
938	1	0	1	2	2	0.67470728	1	1	1	1	10238	-0.600053981	0	-0.661854526	-0.431964234	-0.039344748	0	0	0
1063	1	1	0	0	0	-0.543308226	1	0	0	1	10238	0.817290915	0	-0.661854526	-0.431964234	-0.639757398	0	0	0
1126	2	1	3	2	2	-0.160588186	1	0	0	1	32765	-0.600053981	0	1.565740274	-0.431964234	-0.039344748	0	0	0
1860	1	0	1	1	0	-0.507082733	0	0	1	1	10238	-0.245717757	0	0.229183394	-0.431964234	0.561067901	1	0	0
1924	3	0	4	1	2	0.807702243	1	1	0	1	10238	1.880299588	0	-0.216335566	1.362162444	-0.039344748	0	0	1
2087	2	1	1	0	3	-0.469095496	0	0	1	0	10238	0.462954691	0	-0.661854526	-0.431964234	-0.039344748	0	0	0
2375	1	1	1	1	3	-0.868175572	0	0	0	0	10238	0.462954691	0	-0.661854526	1.362162444	-0.639757398	1	0	1
2387	0	0	0	2	3	0.461699158	1	0	0	0	10238	0.462954691	0	-0.661854526	-0.431964234	-0.639757398	1	1	1
2590	3	0	4	2	2	0.271875417	1	1	1	1	10238	-0.245717757	0	-0.216335566	-0.431964234	-0.039344748	0	0	0
2654	0	1	0	1	0	-2.104563362	0	0	0	1	10238	0.108618467	0	-0.661854526	-0.431964234	-0.639757398	1	1	1
2698	2	1	3	2	2	0.046488171	1	0	0	1	32765	-1.663062654	0	4.238854034	-0.431964234	0.561067901	1	0	0
2743	1	0	1	0	1	-1.218142047	0	0	1	0	21217	0.108618467	0	-0.661854526	-0.431964234	-0.639757398	1	0	0
2906	2	1	1	2	2	1.227743127	1	0	1	1	10238	-1.308726429	0	0.674702354	-0.431964234	0.561067901	0	0	0
3004	1	1	1	2	3	-0.125865996	1	0	0	1	32765	-0.954390205	0	0.229183394	1.362162444	0.561067901	0	0	0
3025	3	0	1	2	2	0.616989755	1	0	1	1	32765	-1.308726429	0	0.229183394	-0.431964234	-0.639757398	0	1	1
3038	1	0	1	1	1	-1.024764311	1	1	0	0	21217	1.525963363	0	-0.661854526	1.362162444	-0.639757398	1	0	0
3114	3	0	4	2	2	1.883989433	1	1	0	1	10238	1.171627139	0	-0.216335566	-0.431964234	-0.639757398	0	0	0
3131	1	0	1	1	3	-0.745184043	1	1	1	0	10238	-0.245717757	0	-0.661854526	1.362162444	-0.039344748	0	0	1
3141	1	0	0	0	1	-0.059716851	0	0	0	0	10238	1.525963363	0	-0.661854526	-0.431964234	-0.639757398	1	0	0
3170	2	0	1	0	1	-0.179250449	1	0	0	1	10238	0.108618467	0	-0.216335566	-0.431964234	-0.639757398	0	1	1
3308	2	1	1	2	2	0.538847359	1	1	0	1	32765	-1.663062654	0	1.120221314	-0.431964234	-0.039344748	1	0	0
3396	2	0	3	0	1	-0.84393936	1	1	0	1	21217	-0.245717757	0	-0.216335566	-0.431964234	-0.039344748	1	0	0
3413	3	1	4	2	2	-0.476283653	1	1	0	1	32765	0.462954691	0	3.347816114	4.950415798	1.161480551	0	0	0
3526	3	1	4	1	2	1.537728246	1	1	1	0	10238	0.462954691	0	1.120221314	1.362162444	1.161480551	0	0	0
3626	2	0	3	2	2	0.322866077	1	1	1	1	10238	-0.245717757	0	-0.216335566	-0.431964234	-0.039344748	0	1	1
3643	0	1	0	1	0	-1.369939677	0	0	0	0	10238	1.525963363	0	-0.661854526	-0.431964234	-0.639757398	1	0	0
3794	0	0	0	0	0	-0.886400333	0	0	0	0	10238	2.94330826	0	-0.661854526	-0.431964234	-0.639757398	1	0	0
3856	2	0	0	1	2	-0.142952944	1	1	0	1	10238	0.817290915	0	-0.661854526	-0.431964234	-0.639757398	0	0	0
4035	2	0	3	2	2	0.121756096	1	0	1	0	32765	-2.017398878	0	0.229183394	1.362162444	-0.039344748	0	0	0

Two candidate models using Logistic Regression:

$$p_{outcome} = \frac{1}{1 + e^{-(0.6438193 - 1.11448209 \times (age))}}$$

$$p_{outcome} = \frac{1}{1 + e^{-(0.38132064 - 1.2841026 \times (driving\_experience))}}$$

Tools used for machine learning and dashboard:

python

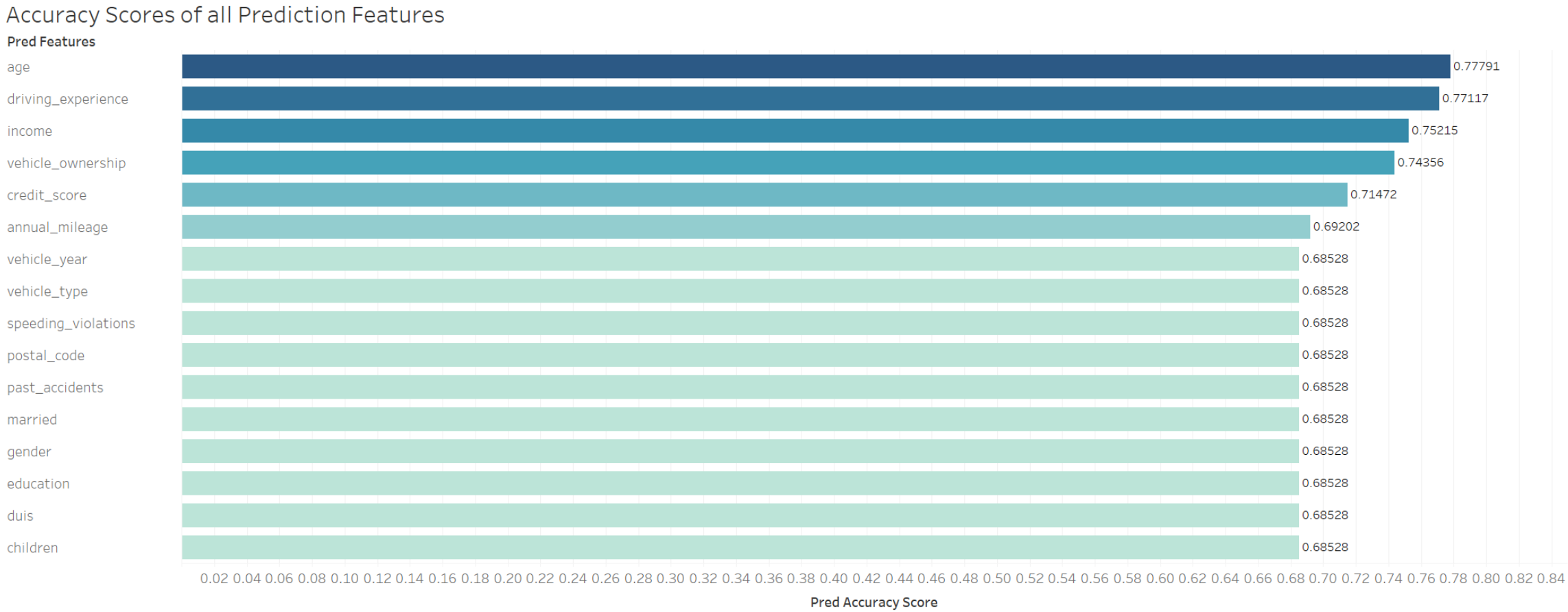
scikit learn

+ a b | e a u

Claim Predictor Model Evaluation Metric

Some features are highly correlated with each other, thus the team built a simple logistic regression model for each feature to choose the best candidates for business stakeholders.

As the company has very little expertise and infrastructure for deploying and and monitoring machine learning models, **accuracy score** is used to measure the performance of each model.

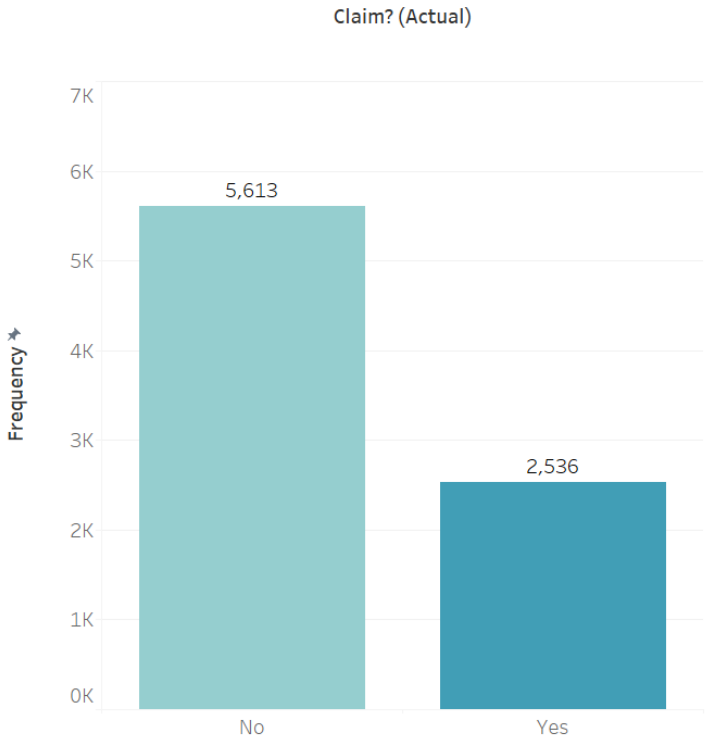


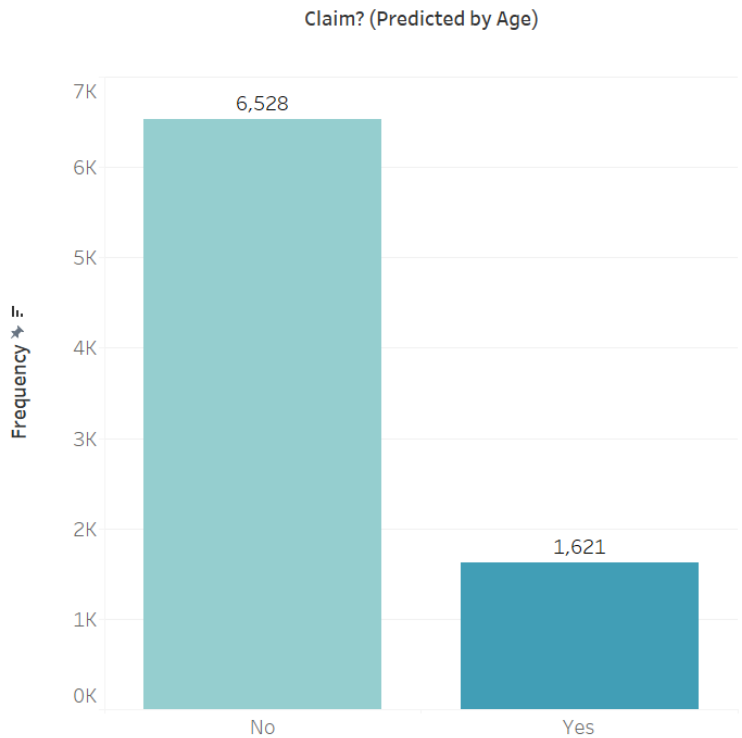
Claim Predictor Model Candidates

Which candidate model to choose from?

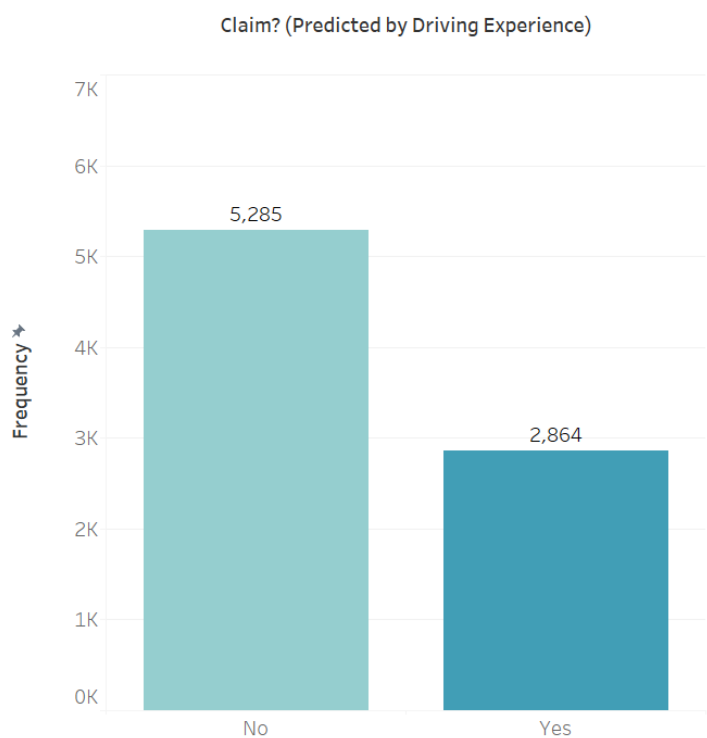
The underlying theme of car insurance contracts are based on **risk assumptions** of how likely the insured will claim on the insurance to determine the rates charged. Predicting more people claiming the auto insurance is better and less risky than predicting less than the actual number.

With business in mind, using 'driving_experience' as the main predictor for claims is recommmended.





Accuracy Score: 0.77791



Accuracy Score: 0.77117

Claim Predictor Model Monitor & Control

Model monitor and control per client id for the claim prediction model using **"Driving Experience"**

Claim? (Actual)
- [ ] No
- [x] Yes

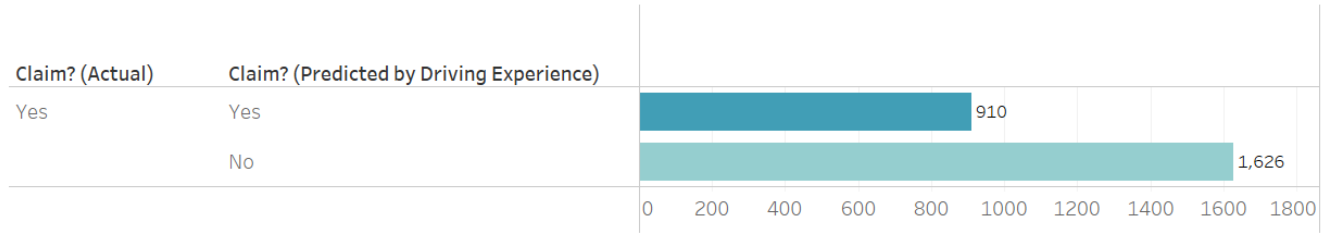
Claim? (Predicted by Drivi..
- [x] No
- [x] Yes

Claim?
- No
- Yes

Predicted Diff. from Actual Claims (using Driving Experience)

78.68%

Actual Claims vs. Predicted Claims using Driving Experience



Predicted Claim Id (using Driving Experience)

Id	Claim? (Predicted by Driving Experience)
101	No
125	No
166	No
186	Yes
226	No
244	No
286	Yes
381	Yes
775	No
793	No
864	No
938	No
1063	No
1126	No
1860	No
1924	Yes
2087	No
2375	Yes
2387	Yes
2590	No
2654	Yes
2698	No
2743	No
2906	No
3004	No
3025	Yes
3038	No
3114	No
3131	Yes
3141	No
3170	Yes

Actual Claim Id

Id	Claim? (Actual)
101	No
125	No
166	No
186	No
226	Yes
244	No
286	Yes
381	No
775	No
793	No
864	Yes
938	No
1063	No
1126	No
1860	Yes
1924	No
2087	No
2375	Yes
2387	Yes
2590	No
2654	Yes
2698	Yes
2743	Yes
2906	No
3004	No
3025	No
3038	Yes
3114	No
3131	No
3141	Yes
3170	No
3308	Yes