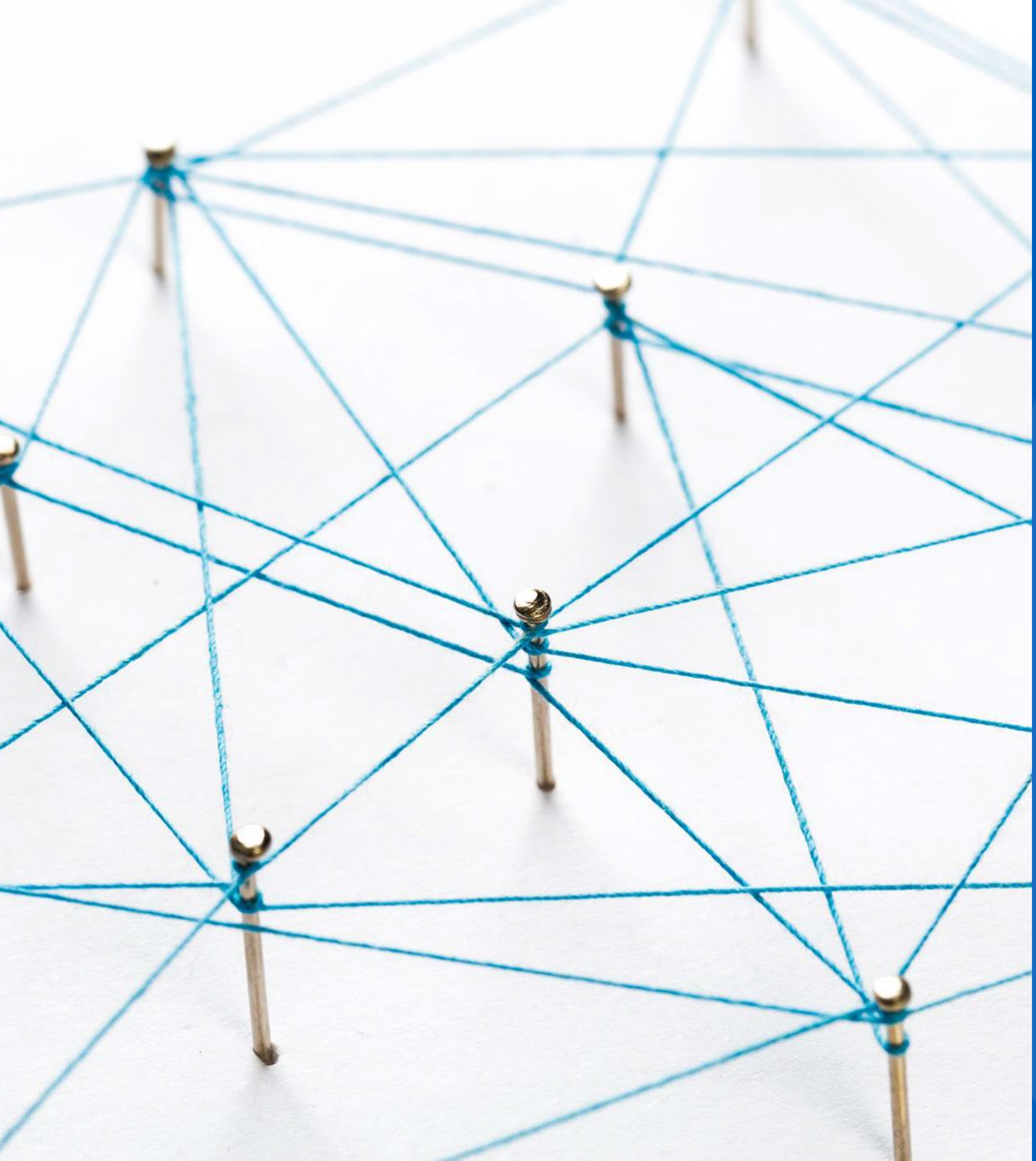


Increasing Speed to Market

NAIC Model Review Team

April 25, 2022

NAIC NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

NAIC Rate Model Review Team

- Kris DeFrain, FCAS, MAAA, CPCU
  - Director of Research and Actuarial Services
- Dorothy Andrews, ASA, CSPA, MAAA
  - Senior Behavioral Data Scientist and Actuary
- Sam Kloese, ACAS, CSPA, MAAA
  - P/C Rate Modeling Actuary and Project Coordinator
- Roberto Perez, FCAS, MAAA
  - P/C Rate Modeling Actuary
- Eric King, FSA, MAAA
  - Senior Health Actuary
- Lia Thomas
  - Senior Administrative Assistant

NAIC Rate Model Support

The NAIC **DOES:**

- Write a Technical Report;
- Maintain consistency with *Regulatory Review of Predictive Models;*
- Verify the model is documented so it can be evaluated and understood;
- Highlight areas for regulatory review.

The NAIC **DOES NOT:**

- Assume regulatory authority;
- Make recommendations for state action or send objections to the company;
- Recommend acceptance/ rejection of the model or rating variable;
- Directly work with an insurer (without state lead).

Typical Regulatory Issues with Filed Rate Models

- Rating Variables
- Data – Sources, Consumers' Data Correction Recourse, "Actuarial" Adjustments
- Use of the Model
- Technical Issues
- Black Box
- Rate or Underwriting Manual

Rate Filing Suggestions to Improve Speed to Market

The Basics:

- Document
  - Start with an actuarial and/or overview rate model memo.
  - Include a complete data dictionary
  - Communicate so a technical regulator can understand
  - Include information that supports the choices; not a whole bunch of charts that don’t help tell the story.
- Use Actuarial ASOPs or, if not an actuary, ASOPs are a useful guide to understand what info. is expected.

ASOPs

ACTUARIAL STANDARDS BOARD

- AAA Applicability Guidelines*
  - All Assignments in all practice areas
    - ASOP 1 – Introductory ASOP
    - ASOP 23 – Data Quality
    - ASOP 41 – Actuarial Communications
  - 6.1 Creation and review of risk classification plan
    - ASOP 12 – Risk Classification
    - ASOP 13 – Trending Procedures in P&C
    - ASOP 25 – Credibility Procedures
    - ASOP 38 – Catastrophe Modeling
    - ASOP 39 – Cat Losses in P&C Ratemaking
    - ASOP 53 – Estimating Future Costs for Prospective P&C
    - ASOP 56 – Modeling

*https://www.actuary.org/content/applicability-guidelines-actuarial-standards-practice-0

Dorothy L. Andrews

The Good, Bad & Ugly of Data

- The Good

  Good data is not problematic. It shows clear relationship to the risk being insured with no connection to unfair discrimination. Statistically sound.

- The Bad

  Bad is somewhat problematic. May not be a source of unfair discrimination but may reflect an unclear relationship to risk. Not statistically significant

- The Ugly

  Ugly data is highly problematic. Clearly source of unfair discrimination. May proxy for a protected class attribute. May or may not be statistically significant.

The Risks of Third-Party Data

- Unregulated
- Nearly Un-Auditable
- Redundant Encodings
- Design Constraints
- Survey Based Data
- May Lack Veracity
- Mismatched Time Period
- Growing Reliance Upon

Feature Engineering

The Plastic Surgery of Data Science

- Pros
  - Easy to implement to obtain complete datasets
  - Preserves variance of original variable
  - Captures the importance of missingness
  - Nullifies the negative effects of outliers
- Cons
  - It only works if you have good knowledge of the domain
  - Consumers may have difficulty recognizing data errors*
  - Distortion of data metrics (e.g., ρ) and distribution
  - May mask or create outliers; Loss of interpretation
  - May lead to overrepresentation skewing distribution
  - May mask predictive power of original variable

* May have regulatory implications

Feature Engineering

Best Practices

- Clearly identify all variables subject to feature engineered
- Provide business and risk-related rationales to support
- Provide the methodology and formulas to support all features engineered variables
- Provide academic references to support their use
- Include clear examples that show the calculation. The example should be detailed enough so that an independent reviewer can replicate the values for other observation in the data set.

Predictor Variable Rationales

NAIC CASTF Regulatory Review of Predictive Models White Paper:

"A <u>rational explanation</u> refers to a plausible narrative connecting the variable and/or treatment in question with real-world circumstances or behaviors that contribute to the risk of insurance loss in a manner that is readily understandable to a consumer or other educated layperson."

NAIC
NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

Predictor Variable Rationales

Operationalizing the NAIC Definition:

- ❑ Create a plausible narrative connecting the predictor to the risk being insured
- ❑ Support the narrative with academic or empirical research
- ❑ Avoid “fuzzy” connections that are difficult to validate empirically
- ❑ Avoid statistical validation as a rationale
- ❑ Strive for explainability to laypersons

Predictor Variable Rationales

[Philadelphia - Explain This To Me Like I'm a Two Year Old ...]

Where to Look for Unfairly Discriminatory Variables

Variables of the type below would give rise to concerns over potential unfair discrimination:

- ❑ Socio-economic
- ❑ Behavioral
- ❑ Demographic, such as zip code
- ❑ Consumer related data
- ❑ GPS related
- ❑ Geo-Spatial
- ❑ Discriminatory Data Generators
- ❑ Medical related data

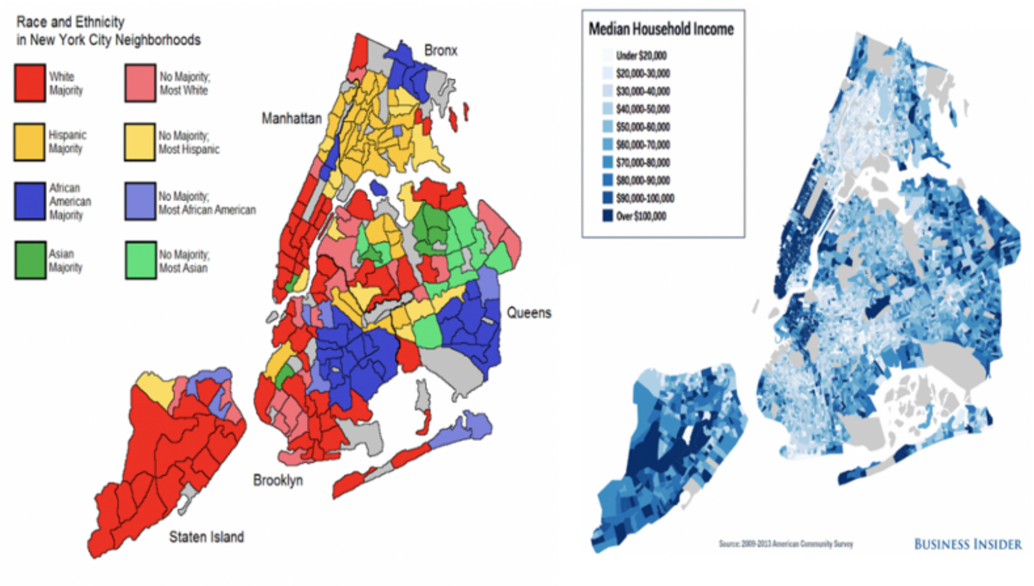
Identifying Bias by Proxy

Redundant Encodings: The protected attribute is encoded across one or multiple features in a dataset, making the removal of the protected attribute useless.

A study by Samuel Yeom, Anupam Datta, and Matt Fredrikson designed to predict the crime rate per community based on 1990 Census data and 1995 FBI Crime Reporting Data.

Findings:

- Removed the 32 out of the 122 features explicitly linked to race.
- They found a proxy for race consisting of a combination of ***58 features out of the 90 remaining features***.
- This proxy had an association with race of 0.85, while the single feature with the strongest association in the dataset only had an association of 0.73.



Race and ethnicity vs. median household income in New York City.

Consumer Dispute Process

Attributes of a good consumer dispute process:

- ❑ It must be known to the insured
- ❑ It must be easily accessible
- ❑ Insured data must be available for inspection, including featured engineered data
- ❑ Consumers must have a clear understanding of how feature engineered data was constructed in order to dispute it.
- ❑ Transparency into vendor models

ASOP 23: Data Quality

- ❑ Definitions; Data sources; Reliance statements
- ❑ Adjustments or assumptions applied
- ❑ Disclosure of limitations, defects, biases, and conflicts with laws & regulations
- ❑ Discussion original design and use of the data
- ❑ Identify questionable values
- ❑ Provide data reasonability tests preformed
- ❑ Must disclose, disclose, disclose!!!

"If in doubt, disclose!" – Quote from a wise ASOP

Sam Kloese

ASOPs

ACTUARIAL STANDARDS BOARD

- ASOP 41
  - The regulator should be considered an “intended user”
  - The rate filing should be considered an “Actuarial Report”
  - The needs of the intended users should be considered
  - Document methods, procedures, assumptions and data
  - Requires sufficient clarity for another actuary to make an objective assessment

Factor Selection Exhibits

- Filing memo should:
  - Identify where you've deviated (if applicable)
  - Explain why you've deviated
- Minimum requirements:
  - Document current, indicated, and proposed
- Efficient Exhibits
  - All 3 side by side for easy comparison
  - Include visualization

Factor Selection Exhibits

GOOD

Variable	BI Current	BI Indicated	BI Proposed
A	0.900	0.704	0.800
B	0.950	0.856	0.900
C	1.000	1.000	1.000
D	1.050	1.157	1.100
E	1.100	1.309	1.200

Variable	PD Current	PD Indicated	PD Proposed
A	0.900	0.700	0.800
B	0.950	0.850	0.900
C	1.000	1.000	1.000
D	1.050	1.150	1.100
E	1.100	1.300	1.200

Variable	COLL Current	COLL Indicated	COLL Proposed
A	0.900	0.709	0.800
B	0.950	0.860	0.900
C	1.000	1.000	1.000
D	1.050	1.153	1.100
E	1.100	1.307	1.200

Variable	COMP Current	COMP Indicated	COMP Proposed
A	0.900	0.718	0.800
B	0.950	0.864	0.900
C	1.000	1.000	1.000
D	1.050	1.160	1.100
E	1.100	1.317	1.200

Factor Selection Exhibits

Current

Variable	BI	PD	COLL	COMP
A	0.900	0.900	0.900	0.900
B	0.950	0.950	0.950	0.950
C	1.000	1.000	1.000	1.000
D	1.050	1.050	1.050	1.050
E	1.100	1.100	1.100	1.100

Indicated

Variable	BI	PD	COLL	COMP
A	0.704	0.700	0.709	0.718
B	0.856	0.850	0.860	0.864
C	1.000	1.000	1.000	1.000
D	1.157	1.150	1.153	1.160
E	1.309	1.300	1.307	1.317

Proposed

Variable	BI	PD	COLL	COMP
A	0.800	0.800	0.800	0.800
B	0.900	0.900	0.900	0.900
C	1.000	1.000	1.000	1.000
D	1.100	1.100	1.100	1.100
E	1.200	1.200	1.200	1.200

BAD

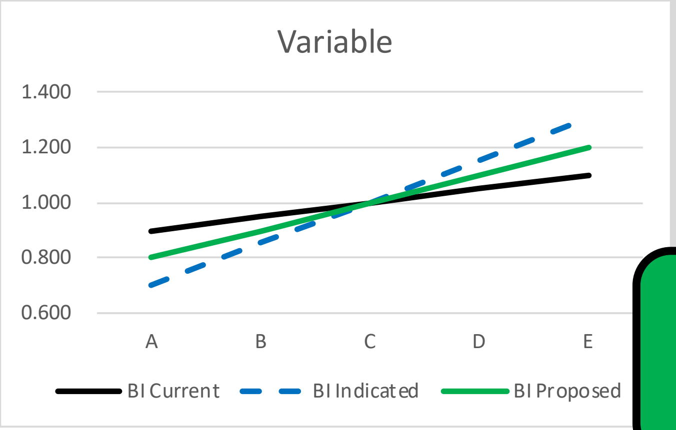
Factor Selection Exhibits

UGLY

Variable	BI Current	BI Indicated	BI Proposed	Variable	PD Current	PD Indicated	PD Proposed
A	0.900	0.704	0.800	A	0.900	0.700	0.800
B	0.950	0.856	0.900	B	0.950	0.850	0.900
C	1.000	1.000	1.000	C	1.000	1.000	1.000
D	1.050	1.157	1.100	D	1.050	1.150	1.100
E	1.100	1.309	1.200	E	1.100	1.300	1.200
Variable	**COLL Current**	**COLL Indicated**	**COLL Proposed**	**Variable**	**COMP Current**	**COMP Indicated**	**COMP Proposed**
A	0.900	0.709	0.800	A	0.900	0.718	0.800
B	0.950	0.860	0.900	B	0.950	0.864	0.900
C	1.000	1.000	1.000	C	1.000	1.000	1.000
D	1.050	1.153	1.100	D	1.050	1.160	1.100
E	1.100	1.307	1.200	E	1.100	1.317	1.200

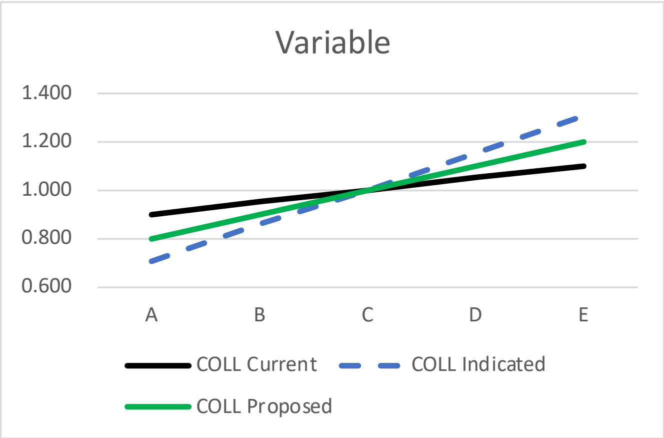
Factor Selection Exhibits

Variable	BI Current	BI Indicated	BI Proposed
A	0.900	0.704	0.800
B	0.950	0.856	0.900
C	1.000	1.000	1.000
D	1.050	1.157	1.100
E	1.100	1.309	1.200



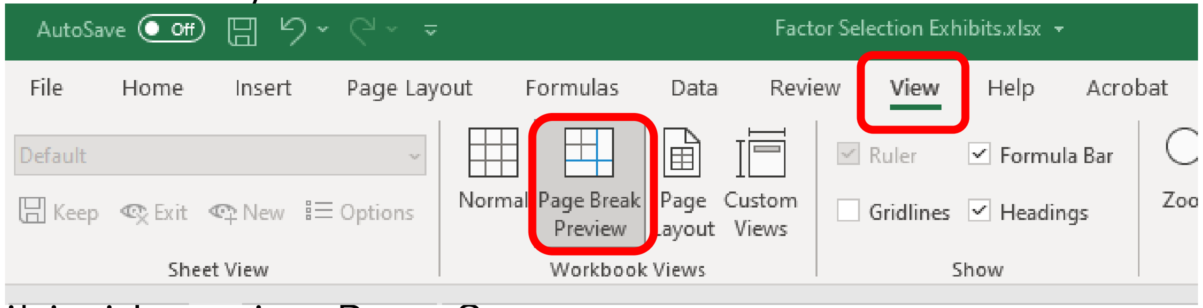
BEST

Variable	COLL Current	COLL Indicated	COLL Proposed
A	0.900	0.709	0.800
B	0.950	0.860	0.900
C	1.000	1.000	1.000
D	1.050	1.153	1.100
E	1.100	1.307	1.200

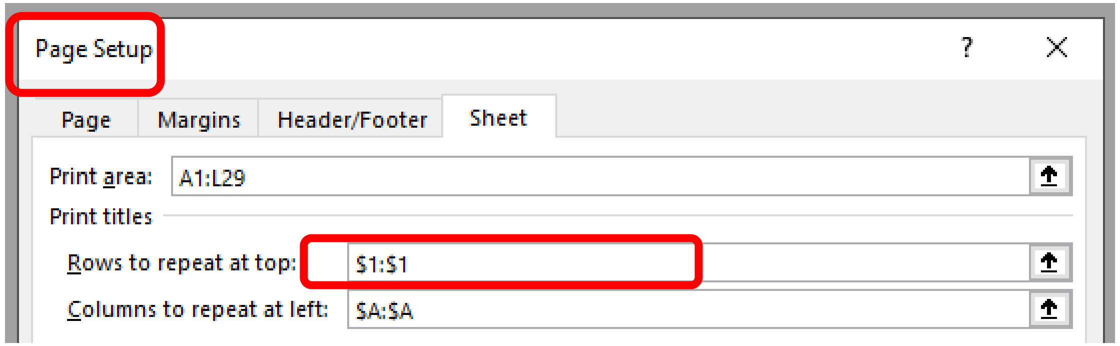


Factor Selection Exhibits

- Excel Tips
  - Page Break Preview is your friend



  - Repeat Exhibit titles using Page Setup
    - Right click in a cell, click on "Page Setup"

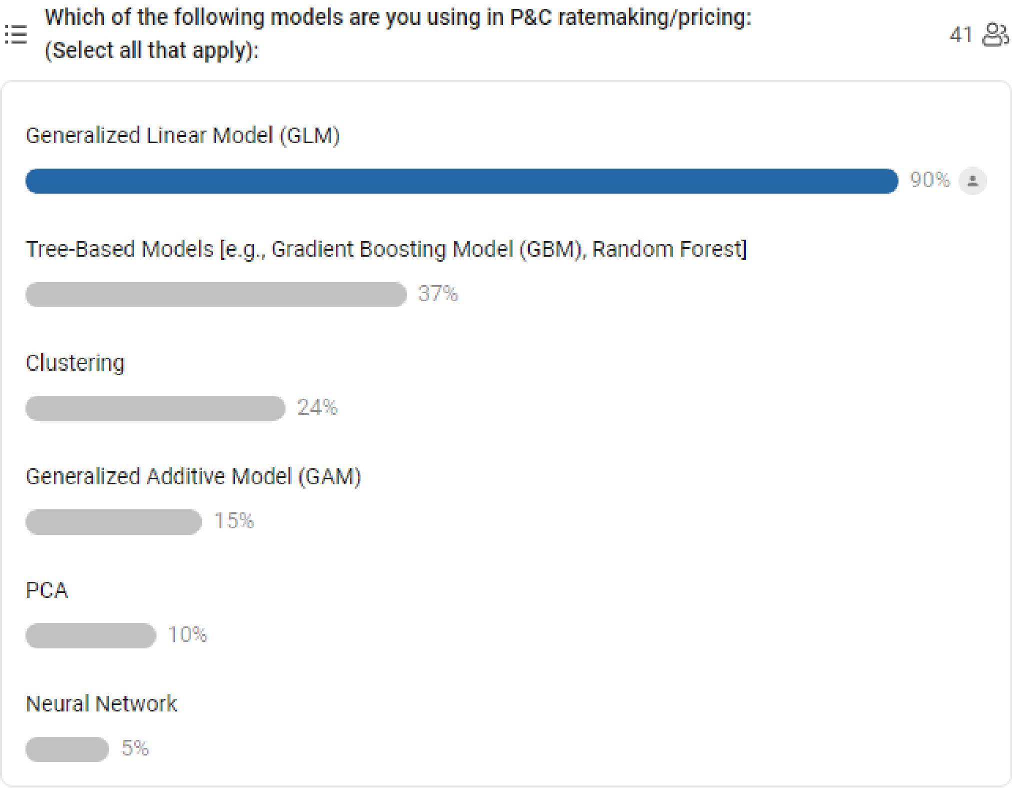


Factor Selection Exhibits

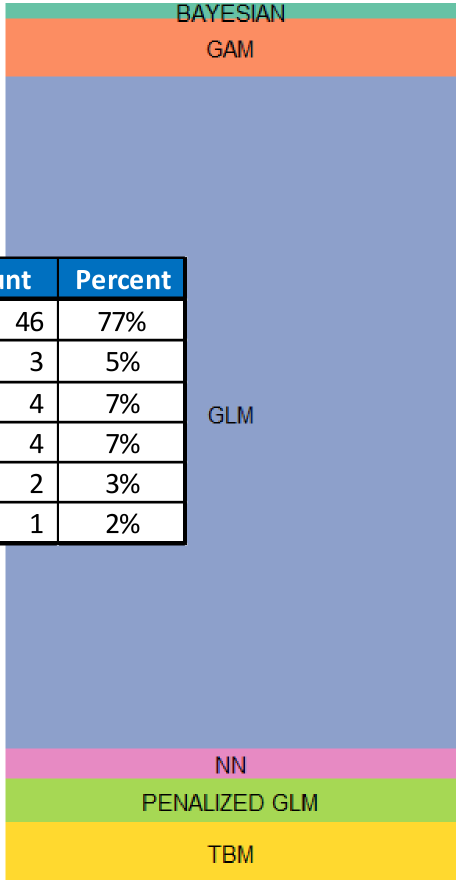
- General Tips
  - Explain every acronym (“LR Rel”, “Freq”, “NAF” vs. “AF”)
    - Even if we know what they mean, we have to ask
  - Use the same variable names throughout filing OR provide a mapping
    - “pol_age” vs. “Renewal Number”
    - “married_ind” vs. “Marital Status”
    - “af_above_threshold_count” vs. “At Fault Accident Count”
- Graph Tips
  - Label every axis
  - Include a legend
  - Put the numbers and the graph on the same page

Non-GLM

2022 CAS RPM Poll



2021-22 NAIC Model Reviews



2021-2022 Reviews	Count	Percent
Traditional GLM	46	77%
Penalized GLM	3	5%
Tree Based Models	4	7%
GAM	4	7%
Nueral Networks	2	3%
Bayesian	1	2%

Non-GLM Support

Complexity

GLM

Penalized GLM
- Elastic Net
- Lasso
- Ridge
- AGLM

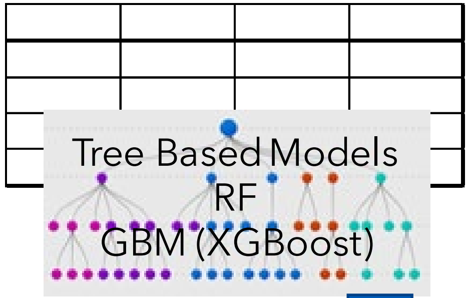
GAM

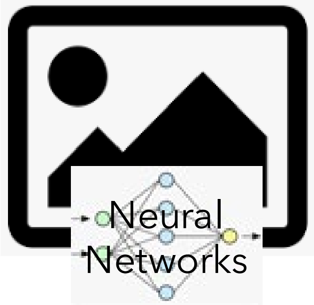
Tree Based Models
- Random Forest
- GBM

Neural Networks

Documentation

Non-GLM Support





PP/Freq/Sev	V1	TBM Score	V2	V3	NN Score

Regression:
- Traditional GLM
- Penalized GLM
  - Elastic Net
  - Ridge
  - Lasso
- GAM

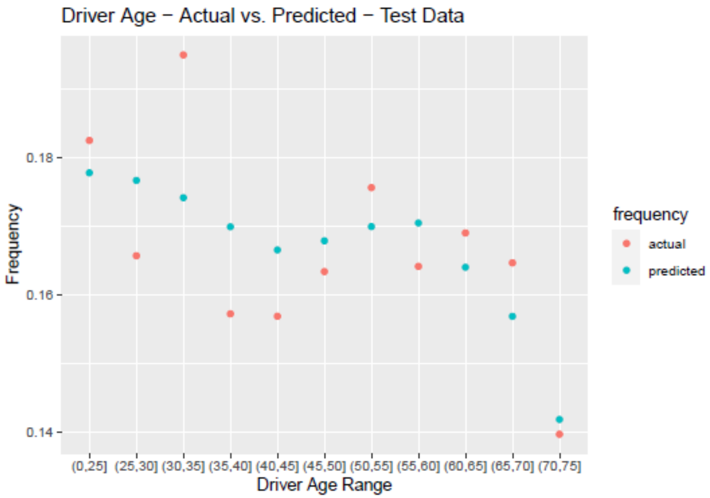
$$g(\mu_i) = \eta_i = \alpha + \sum_{k=1}^{n_\beta} \beta_k z_{ki} + \sum_{j=1}^{n_f} f^{(j)}(w_{ji}) + \varepsilon_i$$

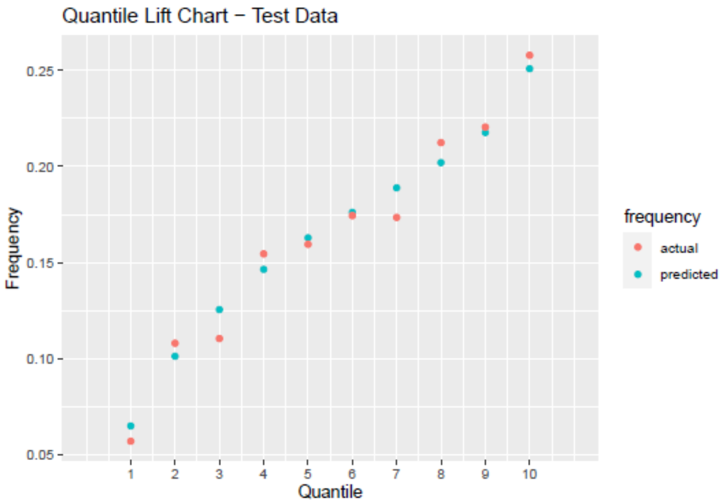
Rating Algorithm

Non-GLM Support

- Applicable to GAM / GBM / RF
  - Example Rate Derivation
  - Correlation Matrix
  - Actual vs. Expected Exhibits
  - Quantile Plots
  - Lorenz Curves & Gini Coefficients

	pol_bonus	drv_age1	vh_age	vh_speed	vh_weight
pol_bonus	1.000	-0.392	-0.031	0.005	-0.036
drv_age1	-0.392	1.000	0.086	-0.047	-0.032
vh_age	-0.031	0.086	1.000	-0.370	-0.276
vh_speed	0.005	-0.047	-0.370	1.000	0.527
vh_weight	-0.036	-0.032	-0.276	0.527	1.000





Non-GLM Support

- General Comments (All other model types)
  - Provide academic reference for background
  - What are the assumptions of the modeling type?
    - Provide commentary on why the assumptions are appropriate
    - Provide metrics/plots that demonstrate appropriateness
  - What are the strengths of the modeling type?
    - Discuss how your model takes advantage of the strengths
  - What are the weaknesses of the modeling type?
    - How did you mitigate the weaknesses of this modeling type?
    - Provide metrics that demonstrate the weaknesses are not material
  - How did you tune the hyperparameters?

Non-GLM Support

- General Comments (All other model types)
  - How have you tested each variable for significance?
    - For the least important, why are they still included?
  - What impact does each variable have?
    - Provide visualizations to interpret what is going on
    - What combination of variables result in highest indications
  - Can you demonstrate how to go from input data to final predictions?
    - Provide all intermediate steps
    - Provide at least 10 sample calculations

Roberto Perez

ASOPs

ACTUARIAL STANDARDS BOARD

- ASOP 56 - Modeling
  - ❑ An actuary using a model developed by others in which the actuary is responsible for the model output is subject to this standard
  - ❑ Hold-out data = NOT used to develop the model
  - ❑ Model Output Validation
    - ❑ Testing against historical actual results
    - ❑ Consistency between model output on holdout data vs. training data
    - ❑ Statistical or analytical tests
    - ❑ Sensitivity testing
    - ❑ Comparing to benchmarks
  - ❑ Consider review by another professional based on model complexity
  - ❑ Documentation

Summary Metrics

- Reviewing the data summary metrics can help with the following:
  - Verifying the training/test data split was performed correctly
  - Evaluating the credibility of the datasets
  - Understanding whether most data was relatively recent (when year is provided)

Summary Metrics

GOOD

	Train				Test			
Year	Losses	Exposures	Claim Counts	Row Count	Losses	Exposures	Claim Counts	Row Count
2015	4,996,424	333,838	9,181	355,801	1,665,475	111,279	3,060	118,600
2016	5,029,107	414,722	9,174	376,375	1,676,369	138,241	3,058	125,458
2017	4,928,700	470,254	9,471	464,323	1,642,900	156,751	3,157	154,774
2018	4,685,700	349,448	10,146	332,891	1,561,900	116,483	3,382	110,964
2019	5,394,450	361,198	10,276	347,581	1,798,150	120,399	3,425	115,860
2020	4,958,234	386,619	9,907	403,331	1,652,745	128,873	3,302	134,444
2021	4,740,089	393,588	9,337	381,988	1,580,030	131,196	3,112	127,329
Total	**34,732,705**	**2,709,667**	**67,492**	**2,662,290**	**11,577,568**	**903,222**	**22,497**	**887,430**

- A data table containing the following information for both train/test datasets
  - Year
  - Losses
  - Exposures
  - Claim Counts (if frequency and severity were modeled separately)
  - Row Count

Summary Metrics

Year	Adjusted Losses	Exposures	Row Count
2015	4,616,058	404,572	434,361
2016	5,053,743	399,550	406,165
2017	4,586,116	340,508	353,109
2018	5,098,512	369,154	374,012
2019	5,175,657	355,404	363,641
2020	4,558,328	336,585	327,547
2021	5,414,183	448,695	418,343
Total	**34,502,596**	**2,655,468**	**2,677,178**

BAD

- An incomplete or hard to read table
  - Showing only information pertaining to the training dataset
  - Not providing enough information to allow for a clear picture of the data
  - Hard to read (inconsistent formatting, no column names, etc.)

Summary Metrics

UGLY

Summary Metrics

		Train				Test
Year	State	Losses	Exposures	Claim Count	Row Count	...
2015	MI	4,759,954	342,033	7,323	328,650	
2016	MI	5,122,883	419,113	10,292	406,410	
2017	MI	4,846,406	439,624	6,339	425,934	
2018	MI	4,783,321	428,416	12,554	408,628	
2019	MI	5,132,176	402,666	4,647	414,051	
2020	MI	5,241,278	488,316	9,060	485,131	
2021	MI	4,660,550	308,538	3,805	311,524	
MI Total		**34,546,568**	**2,828,705**	**54,019**	**2,780,330**	

		Train				Test
Year	State	Losses	Exposures	Claim Count	Row Count	...
2015	CT	5,417,169	428,381	7,993	419,929	
2016	CT	4,832,764	316,436	6,581	307,793	
2017	CT	4,833,872	364,744	7,683	367,797	
2018	CT	5,413,355	448,759	12,221	442,905	
2019	CT	4,609,043	344,833	9,326	340,075	
2020	CT	5,403,028	345,171	9,246	338,213	
2021	CT	4,549,212	433,040	12,253	444,468	
CT Total		**35,058,442**	**2,681,364**	**65,305**	**2,661,180**	

		Train				Test
Year	State	Losses	Exposures	Claim Count	Row Count	...
2015	CW	244,577,587	23,298,373	22,547,305	1,167,953,074	
2016	CW	240,389,125	21,186,998	18,471,543	1,041,722,571	
2017	CW	258,253,985	15,858,340	23,768,631	790,859,343	
2018	CW	239,632,879	21,297,424	14,495,844	1,018,814,019	
2019	CW	272,807,546	21,310,259	22,553,742	1,084,565,878	
2020	CW	270,259,292	20,347,223	17,589,830	1,005,423,748	
2021	CW	266,089,547	20,822,453	16,506,365	1,016,104,287	
CW Total		**1,792,009,961**	**144,121,070**	**135,933,259**	**7,125,442,920**	

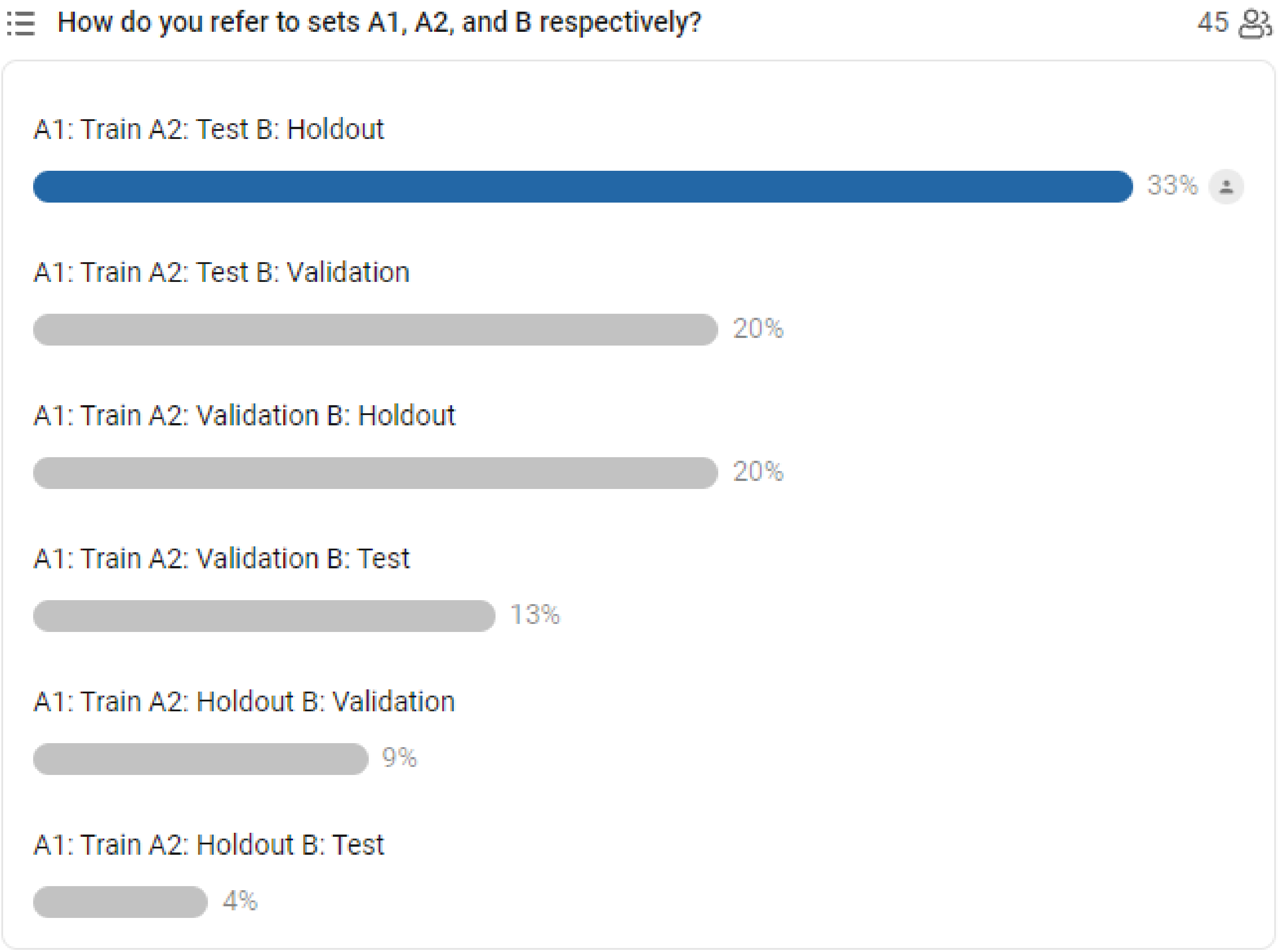
BEST

- It is also useful to see a breakdown of this data by state to ensure the data is appropriate for the state the model is being filed in.

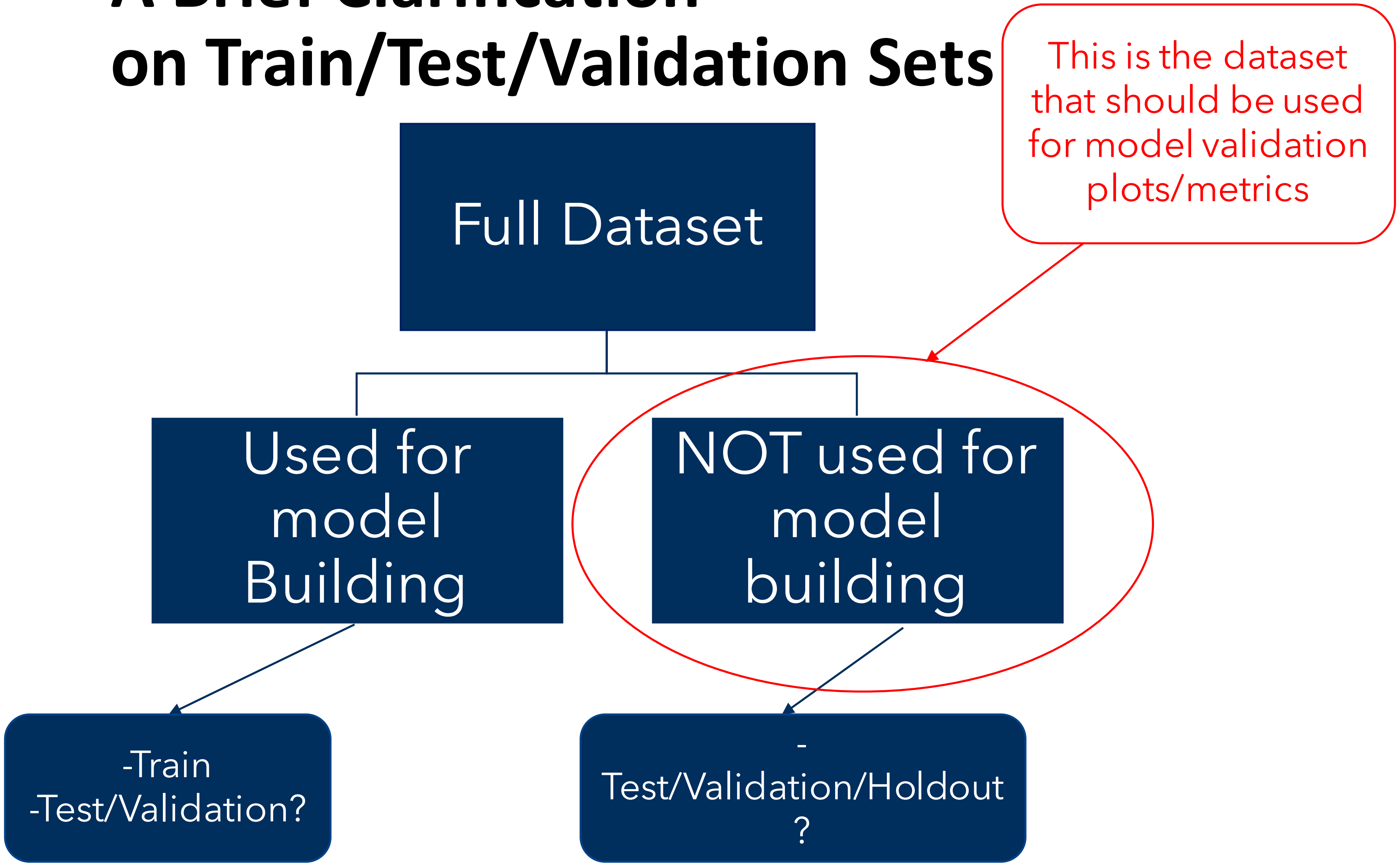
Model Fit

Poll Question

- Question: Suppose you are building a predictive model. You split the data into 2 sets, A (to build the model) and B (to evaluate the model at the end).
- You further split set A into A1 (to fit the model) and A2 (to tune the model).
- How do you refer to sets A1, A2, and B respectively?



A Brief Clarification on Train/Test/Validation Sets



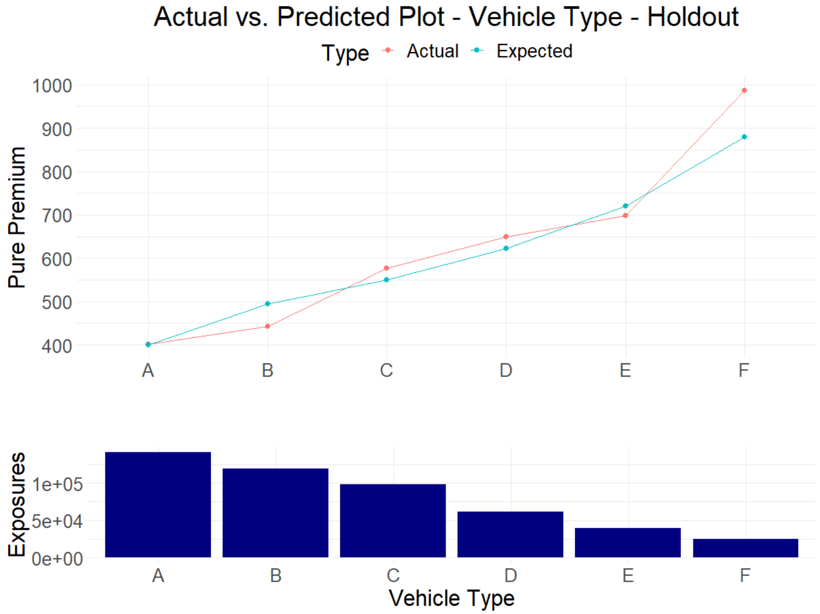
A Brief Clarification on Train/Test/Validation Sets - Summary

- Data used for model building
  - This includes the train dataset
  - It may also include an additional data subset used for model refinement (i.e., used to tune parameters/test variables etc.). This dataset is usually referred to as the validation set. Clarity regarding the dataset usage is more important than terminology. (Note that this additional subset may not be necessary if doing cross-validation).
- Data not used for model building
  - Model validation plots/metrics should be based on this dataset since otherwise they would be overly optimistic.
  - This dataset is usually referred to as either test or holdout. Clarity regarding the dataset usage is more important than terminology.

Actual vs. Expected Plots

- Highlight how well the model fits each individual segment of business. They also help the regulator understand the impact of the model as a whole on each segment.
- Actual vs. Expected by variable summaries look at one variable at a time, yet the predicted values are based on the entire multivariate model. These exhibits can help validate the model across every segment of business after final model parameterization.

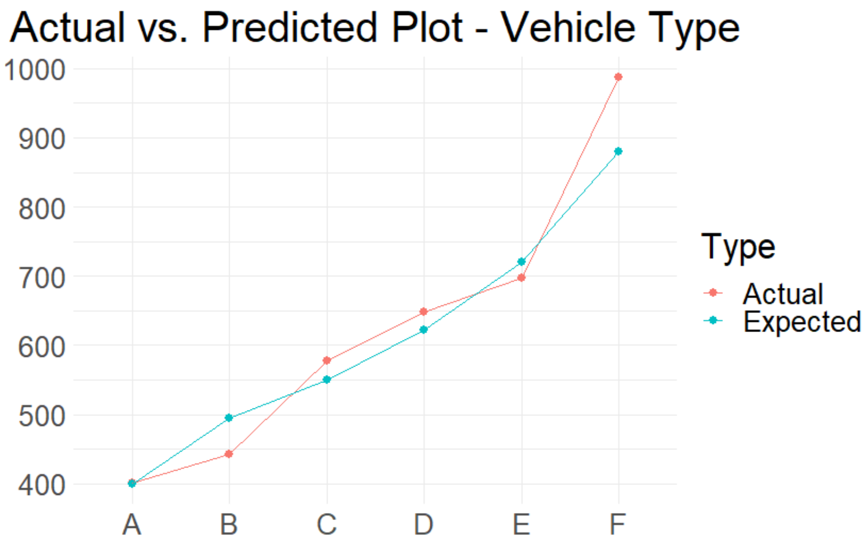
Actual vs. Expected Plots



GOOD

- Gives a visual representation of how well each rating variable contributes to explaining the variation in the modeling data set.
- Plots observed and fitted pure premium for every level of a rating variable over an exposure histogram reflecting the levels of the rating variable.
- The primary (top) y-axis is the pure premium values. The secondary (bottom) y-axis is the percent of exposures.

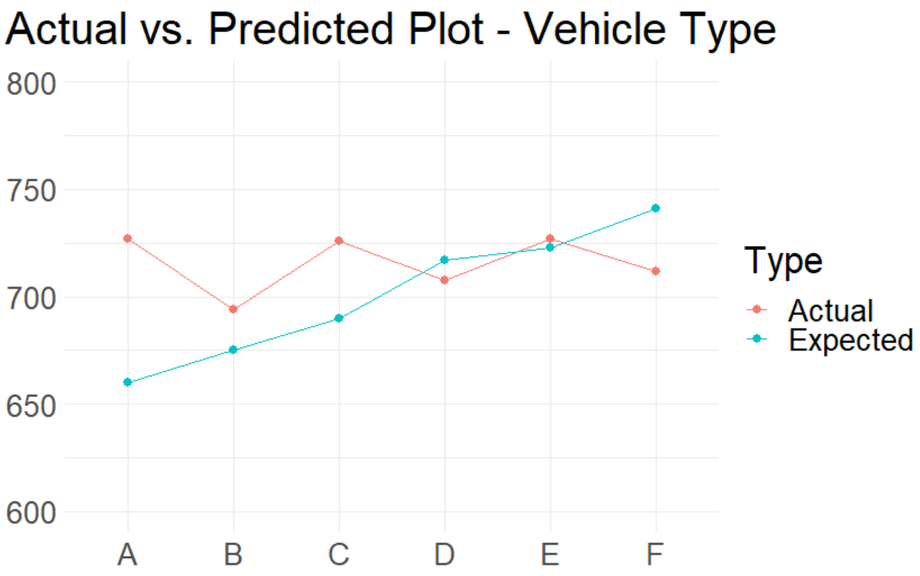
Actual vs. Expected Plots



BAD

- Providing an Actual vs. Expected that is hard to read/unclear
  - Not clear what data it's being built on
  - Not providing exposure information

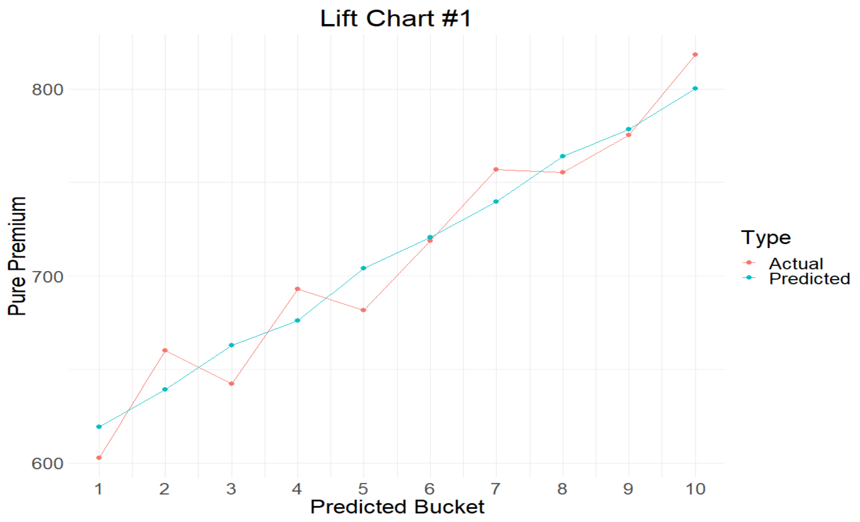
Actual vs. Expected Plots

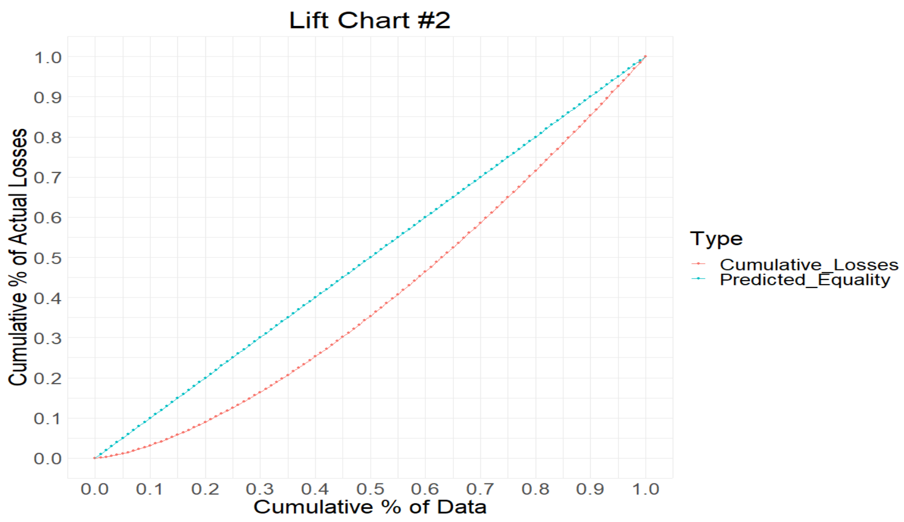


UGLY

- Variable shows no clear relationship to outcome
- Chart built on data used for model building

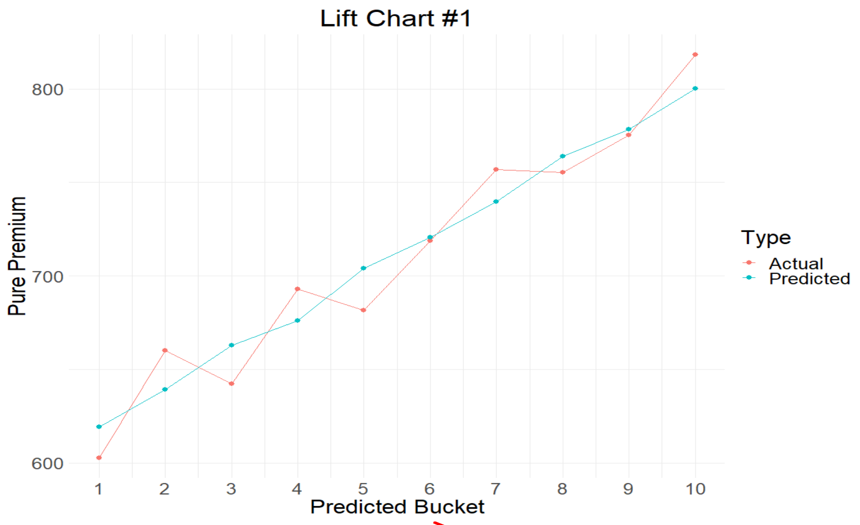
What is a Lift Chart?

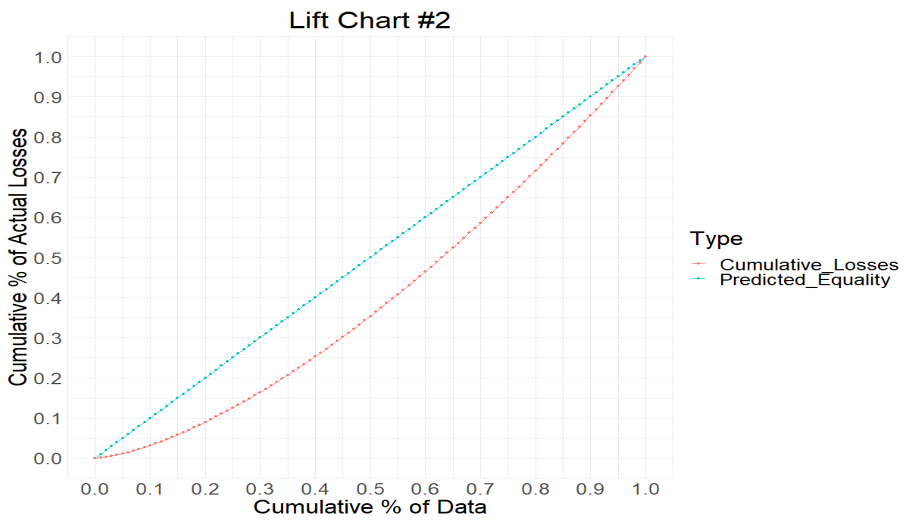




Could refer to two different chart types

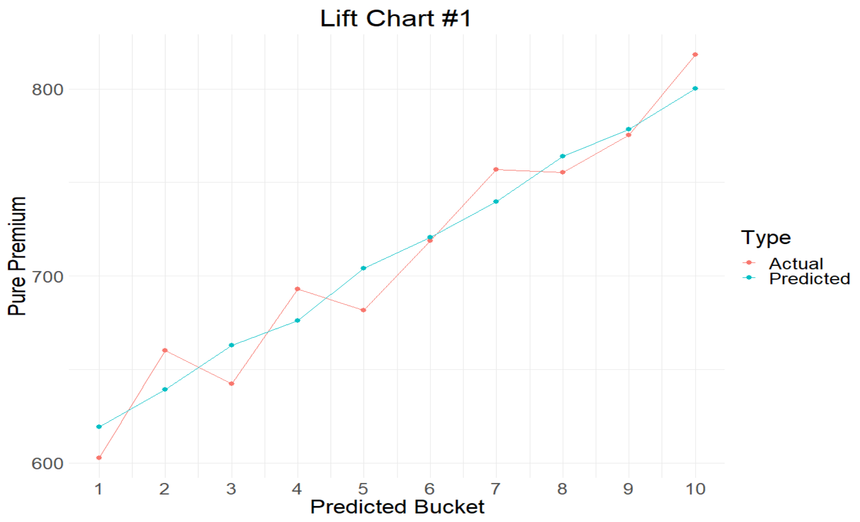
What is a Lift Chart?

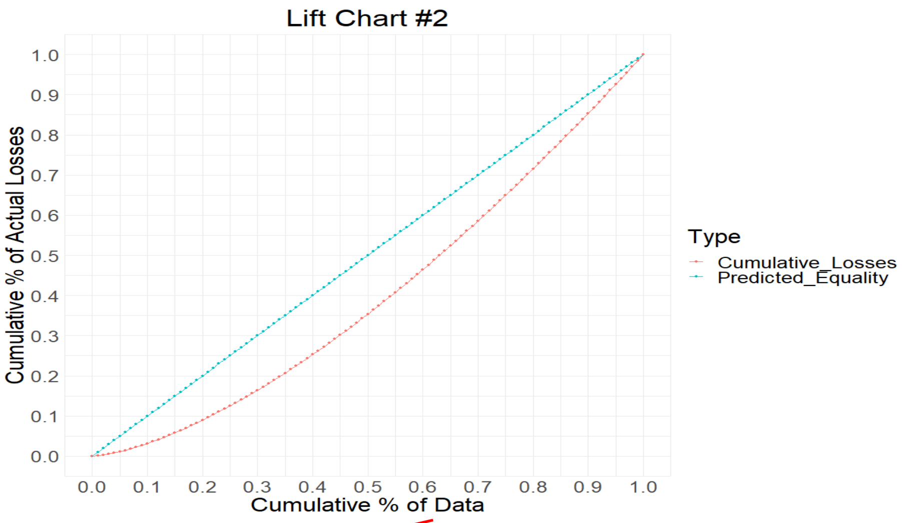




- Sort data by model predictions and create groupings
-Plot the average actual losses by grouping on the y axis

What is a Lift Chart?



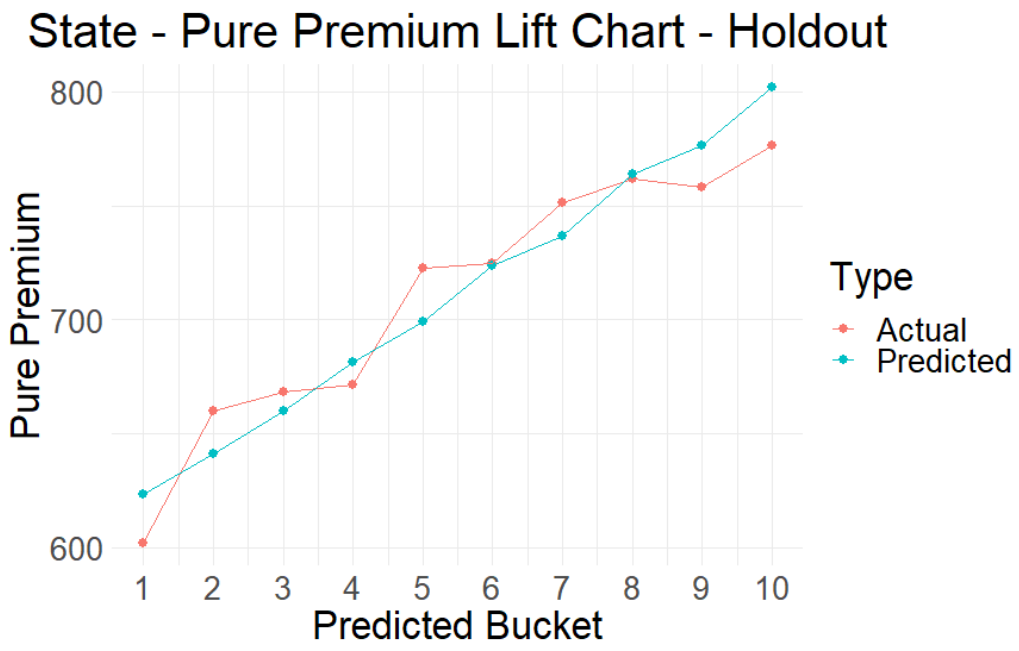


Lorenz Curve
-Sort the data by predicted losses
-Plot the cumulative percentage of actual losses on the y axis
- The Gini statistic is two times the area between the plotted curve and the line of equality

State Lift Charts

- A model could work well in aggregate but fail to produce appropriate results for a subset of the data (i.e., the state the model is being filed in). In addition to providing state level summary metrics, state lift charts are useful to evaluate the appropriateness of the model.

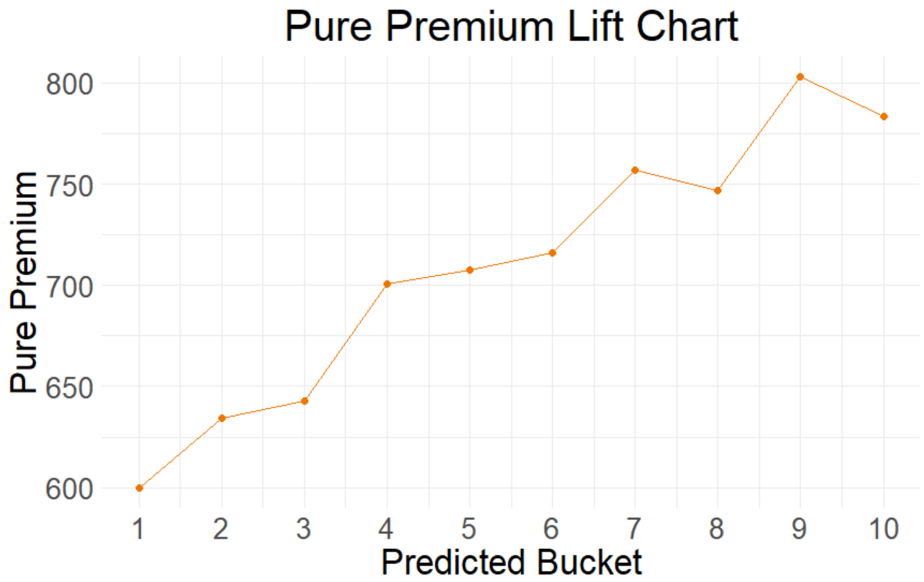
State Lift Charts - Good



GOOD

- Built with a state specific holdout dataset

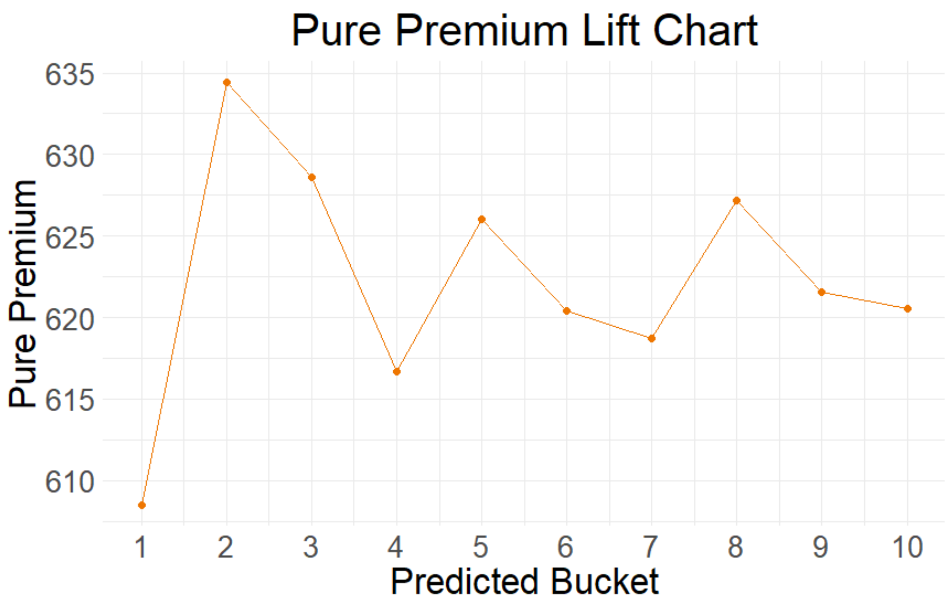
State Lift Charts - Bad



BAD

- Providing a lift chart that is hard to read/unclear
- Not clear what data it's being built on
  - Not providing the predicted losses by grouping
  - Mismatch between actual and predicted groupings (i.e., one corresponding to CW and the other to a State)

State Lift Charts - Ugly

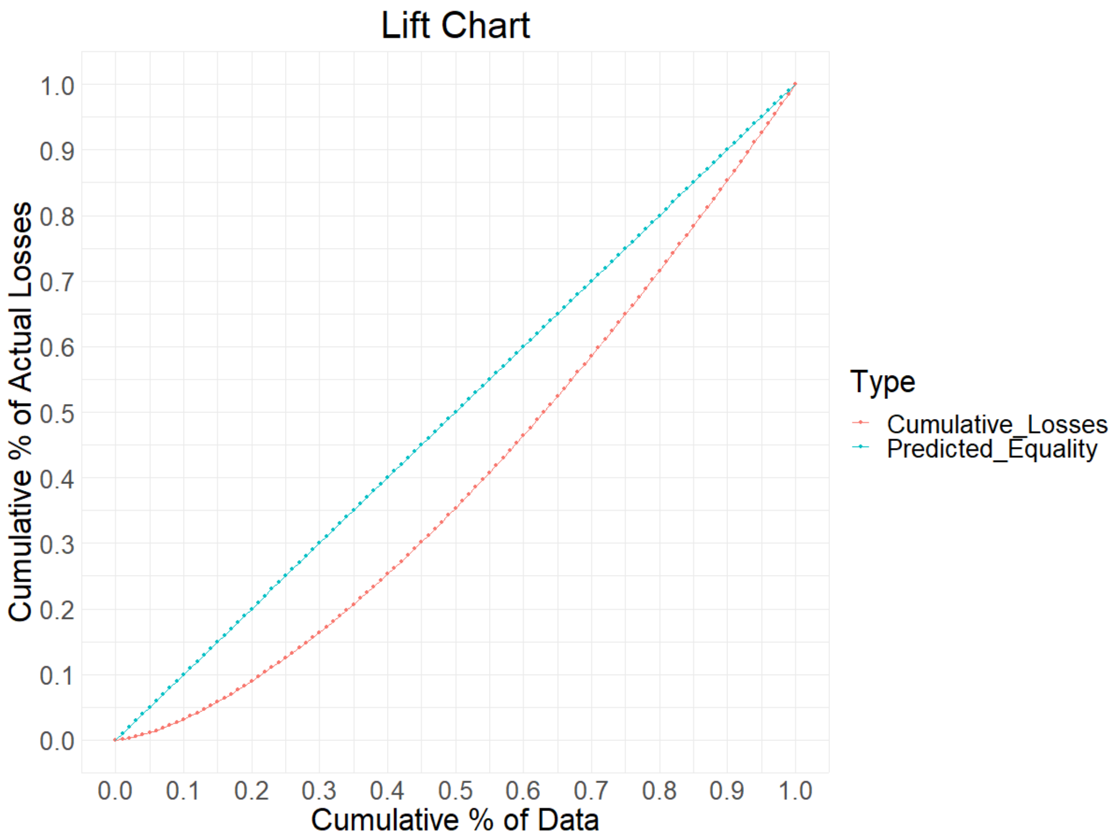


UGLY

- Lift chart does not support model appropriateness
- Providing a lift chart on data used for model building
- Not providing a lift chart or any commentary on the model appropriateness for state specific data

Gini Statistic

- Represents a measure of how well the model is ordering the data
- Is appropriate to provide as a complement to other model validation metrics
- Important to provide a benchmark



Answer in Chat:

Balancing the regulatory review of predictive models with speed to market, what change would you recommend to regulators?