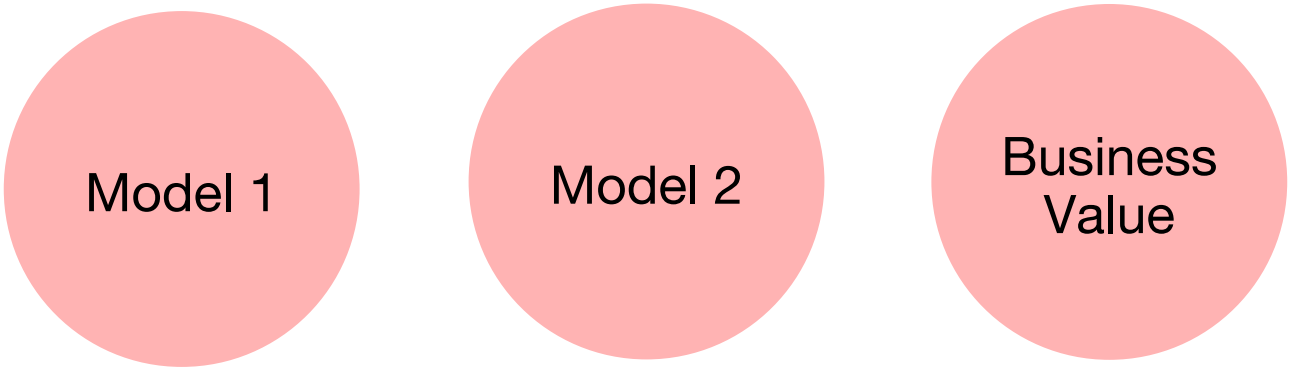
Travelers Case Competition

A Kangaroo Auto Insurance Company Modeling Problem

Xin (Amy) Ni
Jun Sun
Zhong (Verse) He
Julie Shih

© Uconn School of Business

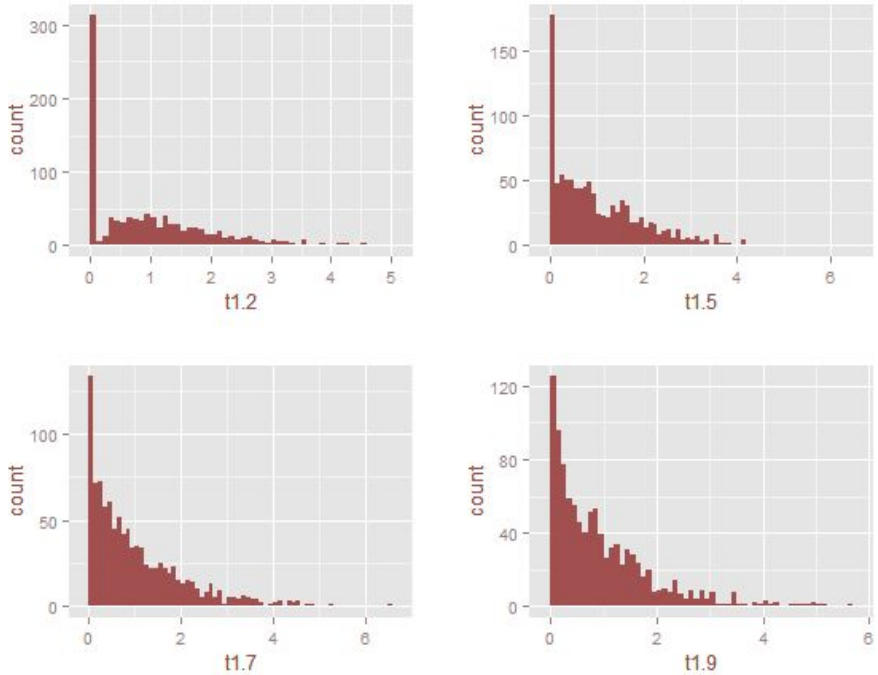
Objective : build a high accuracy predictive model and forecast the claim cost



© Uconn School of Business

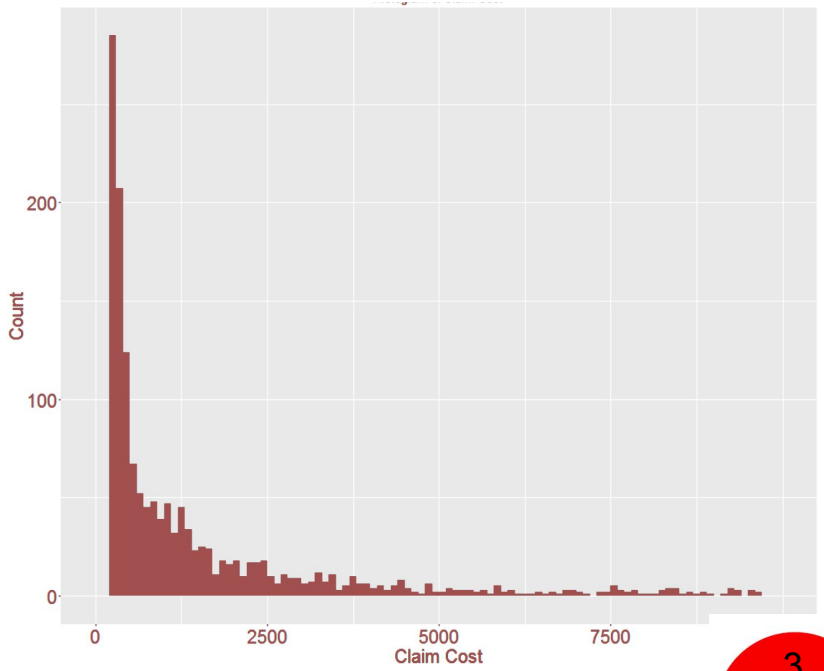
Tweedie Distribution

Power: 1-2



Our Distribution

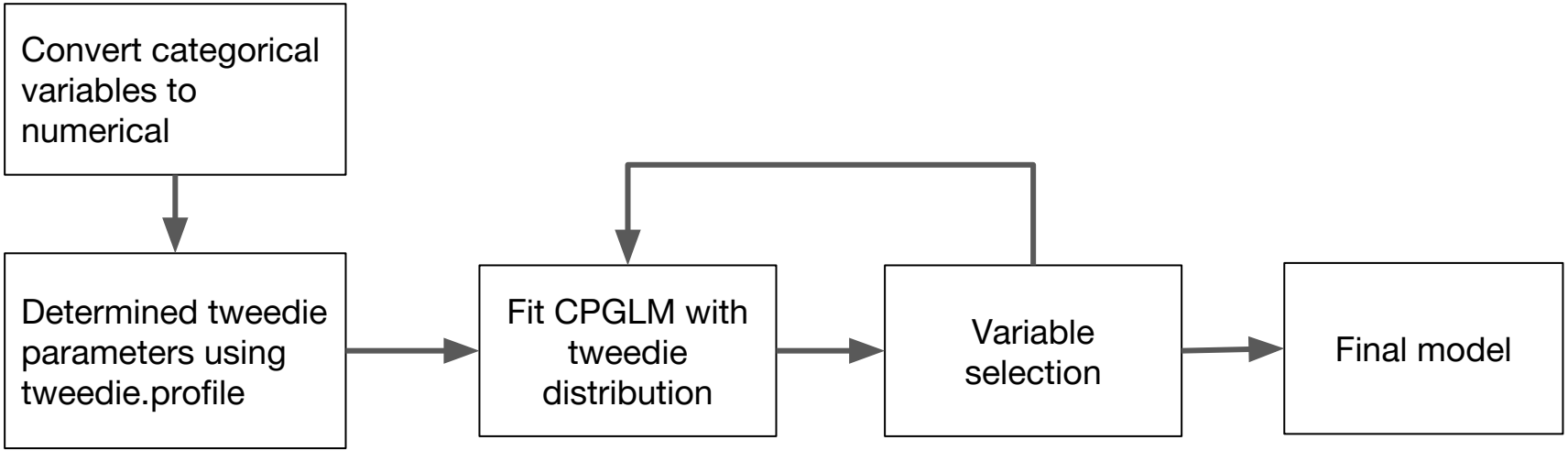
Exclude 0 and extreme costs



© Uconn School of Business

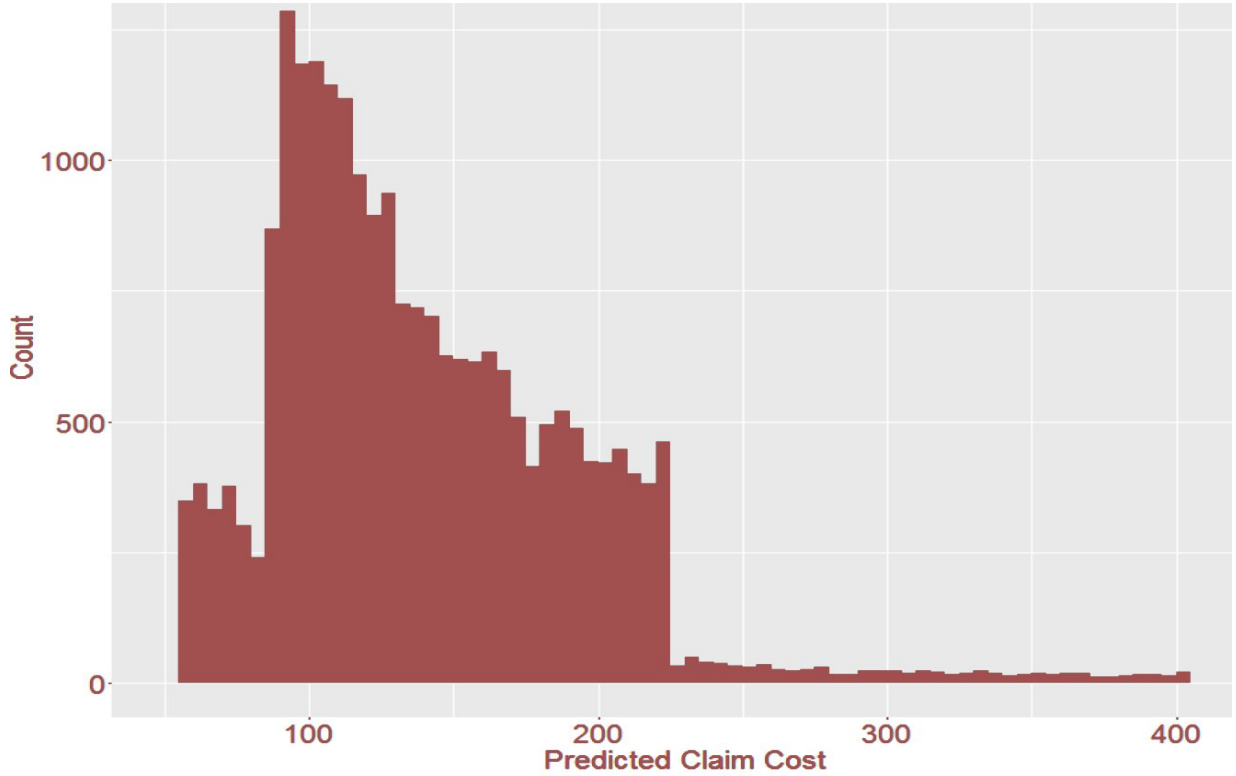
TRAVELERS

Compound Poisson Generalized Linear Model (CPGLM) with Tweedie



© Uconn School of Business

Our key variables are exposure, vehicle age.1, and agecat.1



There are four other variables we think could be useful.

Annual Mileage

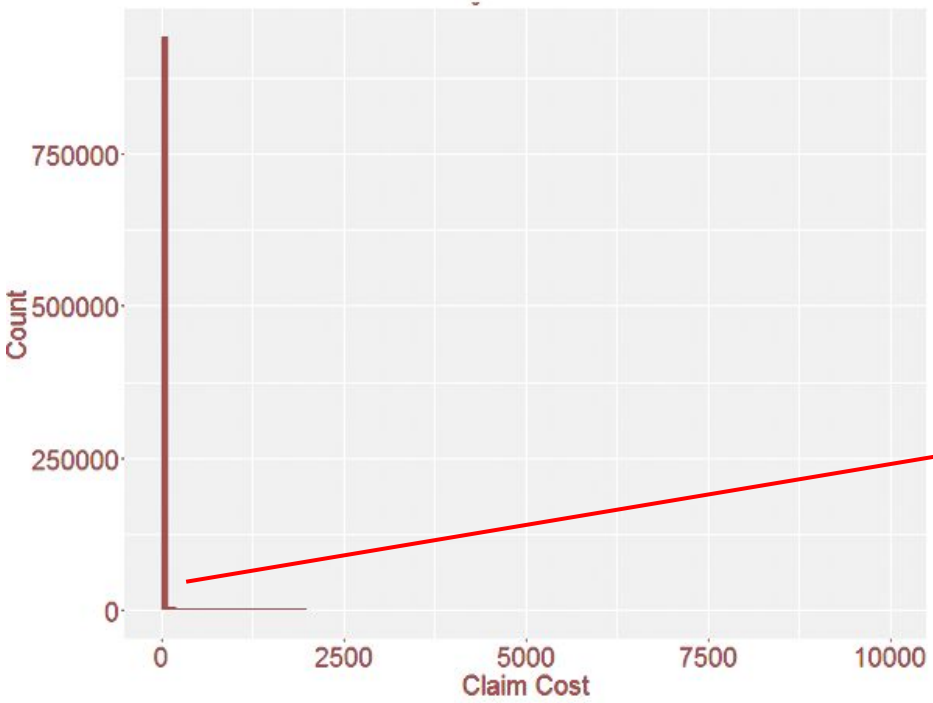
Maximum Coverage
&
Deductible

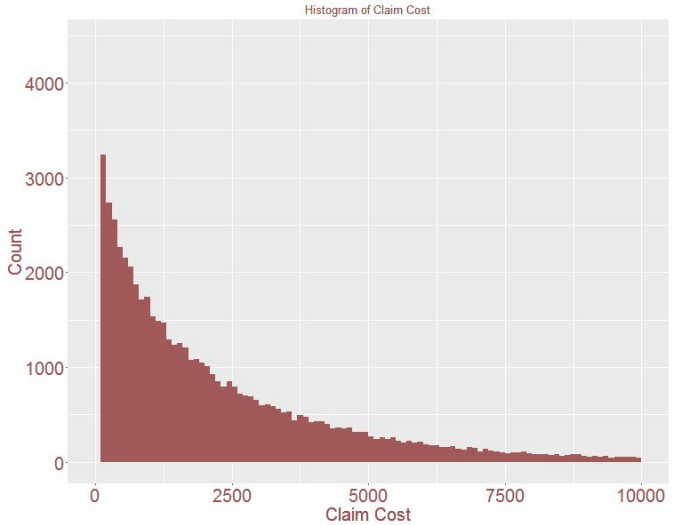
Claim History

Driving History
(speeding
tickets, etc)

© Uconn School of Business

We reconstructed a tweedie distribution.

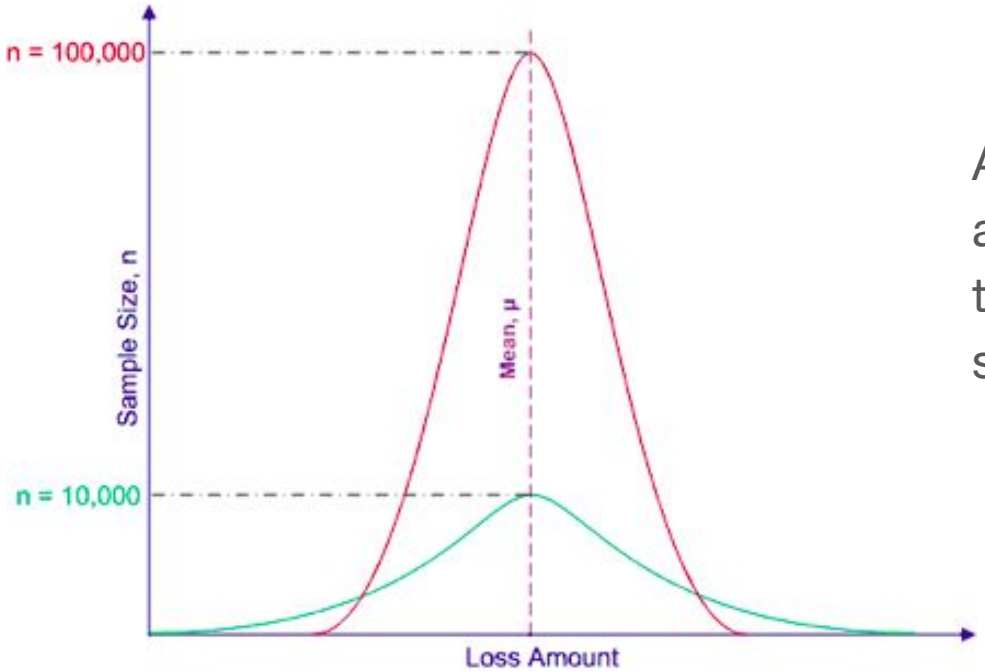




Mean = 141
Power = 1.56
Dispersion = 313

© Uconn School of Business

Our theoretical basis: the Law of Large Numbers



Although single events are random and unpredictable,
the average outcome of many similar events can be predicted.

© Uconn School of Business

Our assumption: three groups of drivers

Three Groups of Divers	Safe Driver	Normal Driver	Dangerous Driver
Claim Possibility	4%	6%	8%

© Uconn School of Business

Choose Segmentation Variables

veh_value_bin	veh_body	veh_age	gender	area	agecat	Avg_monthly_cost	Count
0.6-0.9	HBACK	3	F	C	4	8.934837231	47
0.9-1.2	SEDAN	3	F	C	3	7.645400264	46
1.2-1.5	HBACK	2	F	C	4	153.1620101	46
0.6-0.9	SEDAN	4	F	C	4	117.1959183	45
1.2-1.5	HBACK	1	F	C	4	140.4875732	45
0.9-1.2	SEDAN	3	F	C	4	30.01457345	44
0.3-0.6	HBACK	4	F	C	4	9.328165426	43
0.3-0.6	HBACK	4	F	A	4	30.02837979	42
1.2-1.5	SEDAN	3	F	C	4	83.70160116	41
0.3-0.6	SEDAN	4	F	C	4	14.67230575	37
0.9-1.2	HBACK	3	F	C	2	71.1220102	36
1.2-1.5	HBACK	2	F	C	2	8.458277035	36
1.2-1.5	SEDAN	3	F	B	4	30.24051351	36
0.9-1.2	HBACK	3	F	C	3	42.70048489	35

45239 observations in total.

Key variables from our model:

- Exposure
- Vehicle Age
- Agecat
- Area - not that significant

Our target variable:

$$monthly\ cost = \frac{claimcst0}{(1 - exposure) * 12}$$

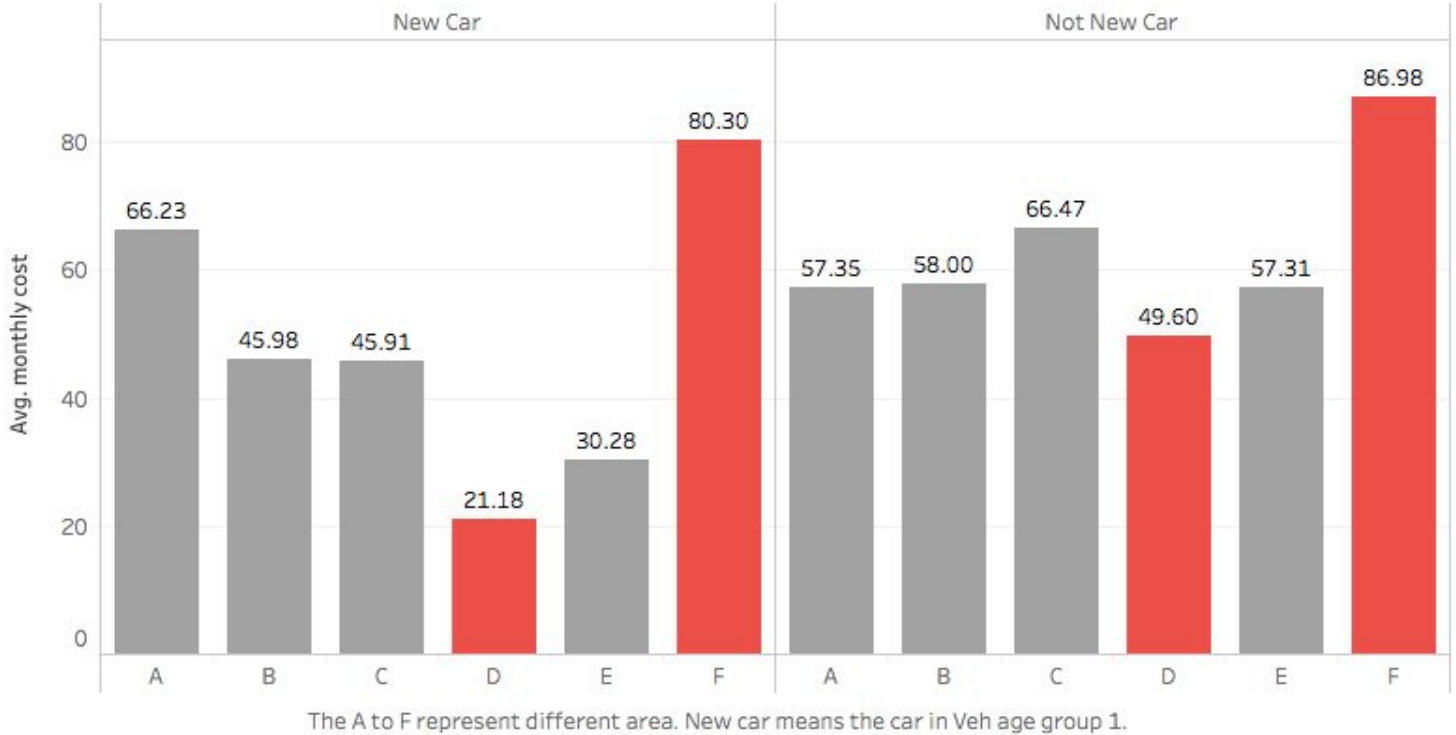
Assumption: the duration of each policy is one year

© Uconn School of Business

Area D has lowest Avg. claim cost while area F has the highest.

Key Variables

- Exposure
- Vehicle Age
- Agecat
- Area

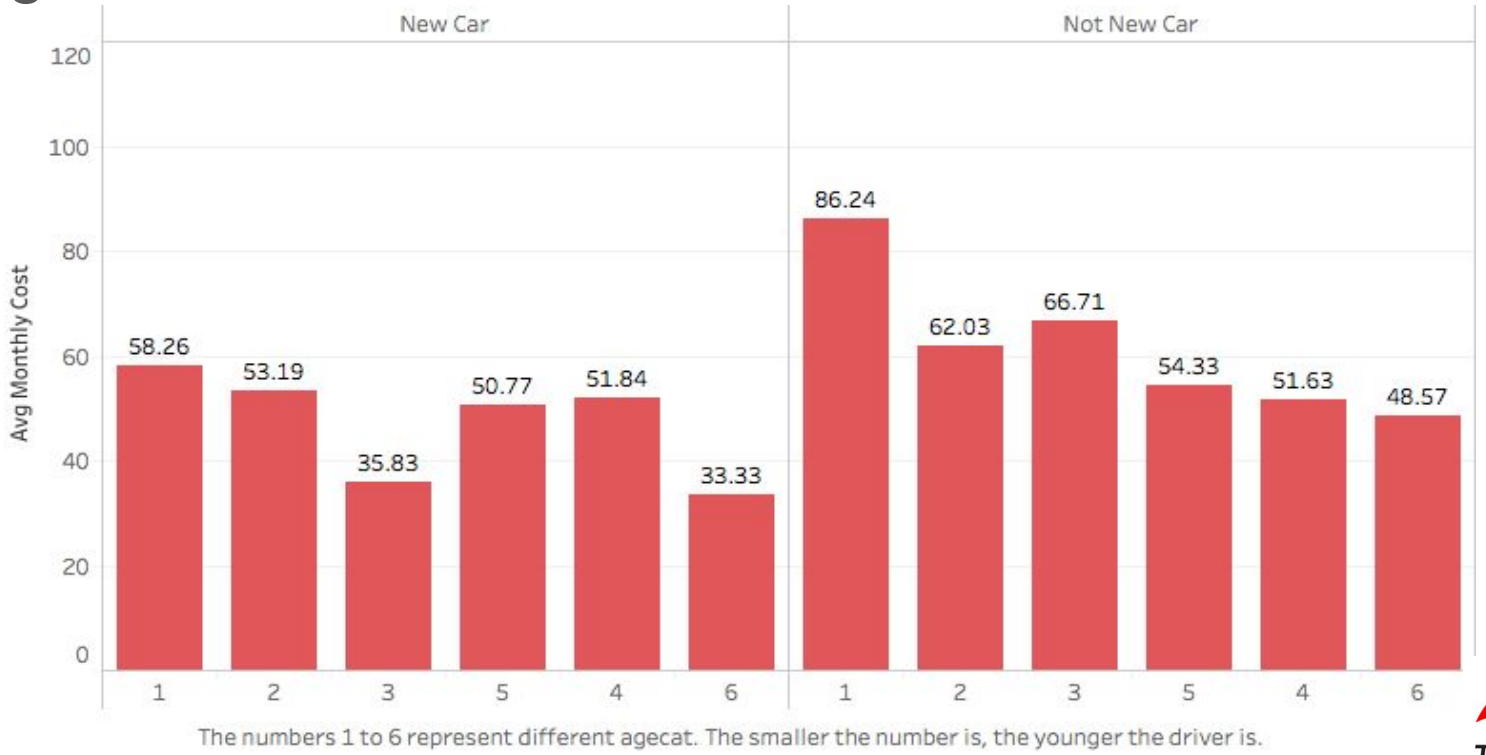


© Uconn School of Business

The Avg. claim cost of new cars is lower than that of others.

Key Variables

- Exposure
- Vehicle Age
- Agecat
- Area



© Uconn School of Business

Highest Avg.claim cost vs. lowest Avg.claim cost

© Uconn School of Business

© Uconn School of Business

Appendix: Initial variables analysis

(Intercept)	4.550838	0.378296	12.030	< 2e-16	***
veh_value	0.061599	0.064654	0.953	0.34073	
exposure	1.040241	0.191415	5.434	5.55e-08	***
veh_body.BUS	-0.274148	1.938056	-0.141	0.88751	
veh_body.CONVT	-25.203052	270.264715	-0.093	0.92570	
veh_body.COUPE	0.617282	0.512558	1.204	0.22848	
veh_body.HBACK	0.062010	0.251287	0.247	0.80509	
veh_body.HDTOP	0.582772	0.372184	1.566	0.11741	
veh_body.MCARA	-0.949523	1.675582	-0.567	0.57094	
veh_body.MIBUS	-0.032788	0.566677	-0.058	0.95386	
veh_body.PANVN	-0.356301	0.600310	-0.594	0.55283	
veh_body.RDSTR	0.077100	2.798439	0.028	0.97802	
veh_body.SEDAN	0.150129	0.239812	0.626	0.53130	
veh_body.STNWG	0.001138	0.242296	0.005	0.99625	
veh_body.TRUCK	0.203210	0.375793	0.541	0.58869	
veh_body.UTE	NA	0.223168	NA	NA	
veh_age.1	-0.585104	0.178980	-3.269	0.00108	**
veh_age.2	-0.136351	0.151462	-0.900	0.36801	
veh_age.3	-0.123507	0.115131	-1.073	0.28339	
veh_age.4	NA	0.247040	NA	NA	
gender.F	-0.184545	0.251485	-0.734	0.46307	
gender.M	NA	0.241584	NA	NA	
area.A	-0.425126	0.269297	-1.579	0.11443	
area.B	-0.319246	0.286641	-1.114	0.26540	
area.C	-0.428310	0.254831	-1.681	0.09282	.
area.D	-0.491874	0.222485	-2.211	0.02706	*
area.E	-0.542081	0.219602	-2.468	0.01358	*
area.F	NA	0.220130	NA	NA	
agecat.1	0.807305	0.240498	3.357	0.00079	***
agecat.2	0.586692	0.378296	1.551	0.12094	
agecat.3	0.304782	0.064654	4.714	2.44e-06	***
agecat.4	0.124916	0.191415	0.653	0.51403	
agecat.5	-0.185002	1.938056	-0.095	0.92395	