Customer Lifetime Value Analysis

Auto Insurance Company

OUR TEAM CREW

greatlearning
Learning for Life

Mr. Srikar Muppidi

Harish Ch

Shaik Shahul

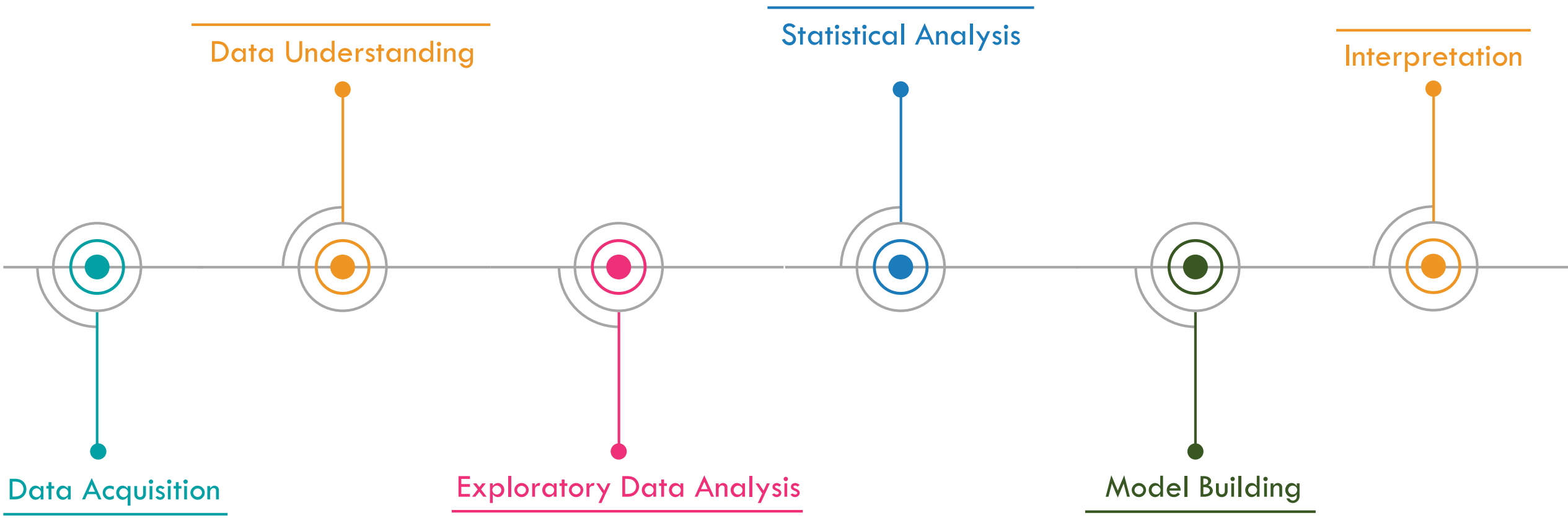
Sirisha T

Uday Kumar S

Ved Kumar Verma

PROCESS TIMELINE

greatlearning
Learning for Life



PROBLEM STATEMENT

We are expected to create an analytical and modelling framework to predict the lifetime value of each customer based on the quantitative and qualitative features.

Company's Approach and Insight

- A major non-life insurance company wants to evaluate customer lifetime value based on each customer's demographics and policy information including claim details.
- The CLV is a profitability metric in terms of a value placed by the company on each customer and can be conceived in two dimensions: the customer`s present Value and potential future Value.

DATA DESCRIPTION

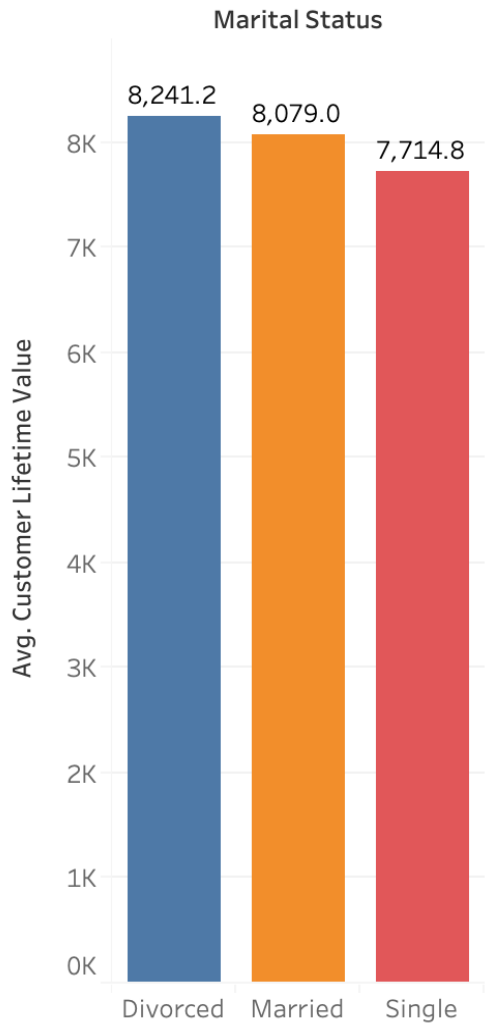
greatlearning
Learning for Life

- CLTV
- Total Claim Amount
- Employment Status
- No. of Policies
- State
- Income
- Gender
- Policy/Policy Type
- Months Since Last Claim
- Response
- Location Code
- Renew Offer Type
- Monthly Premium Auto
- Coverage
- Marital Status
- Sales Channel
- Months Since Last Policy Inception
- Education
- Number of open complaints
- Vehicle Class/Vehicle Size

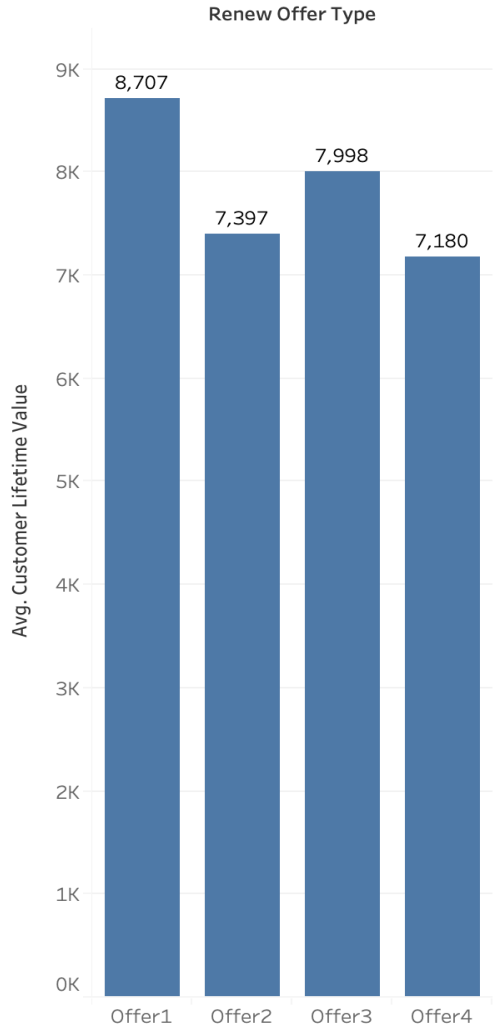
Exploratory Data Analysis

greatlearning
Learning for Life

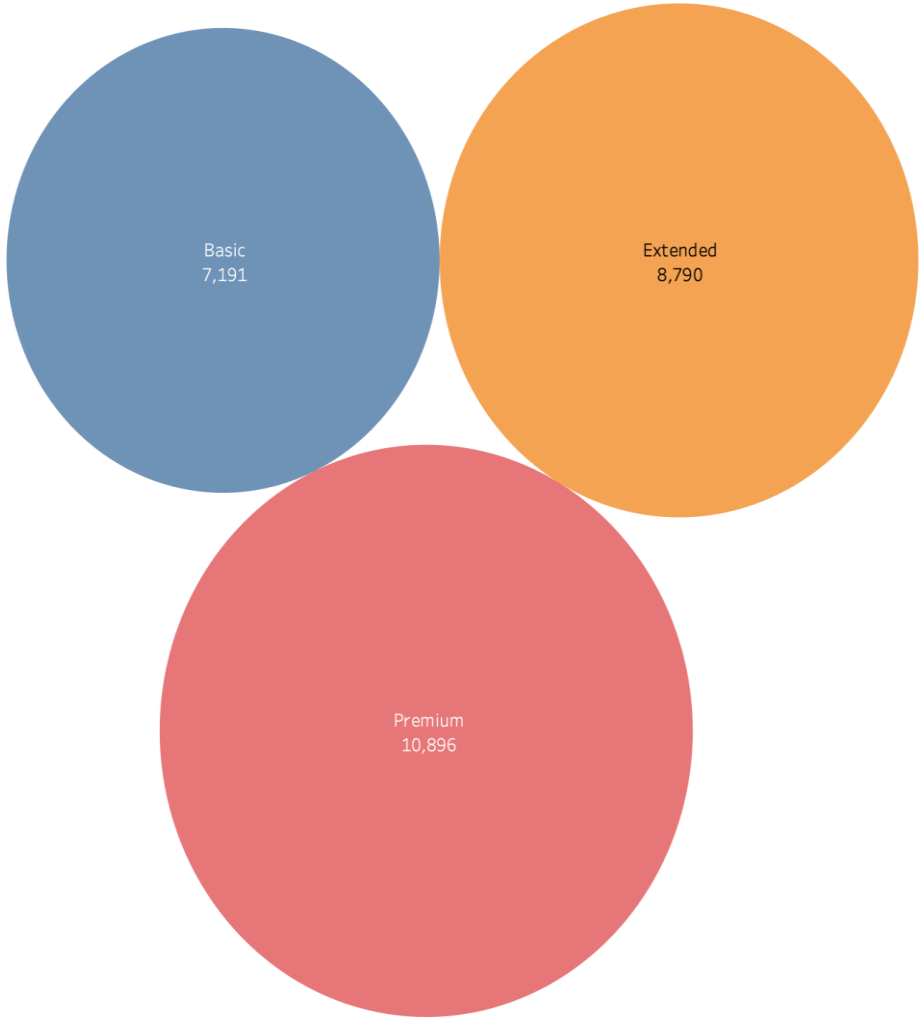
CLTV vs Maritial Status



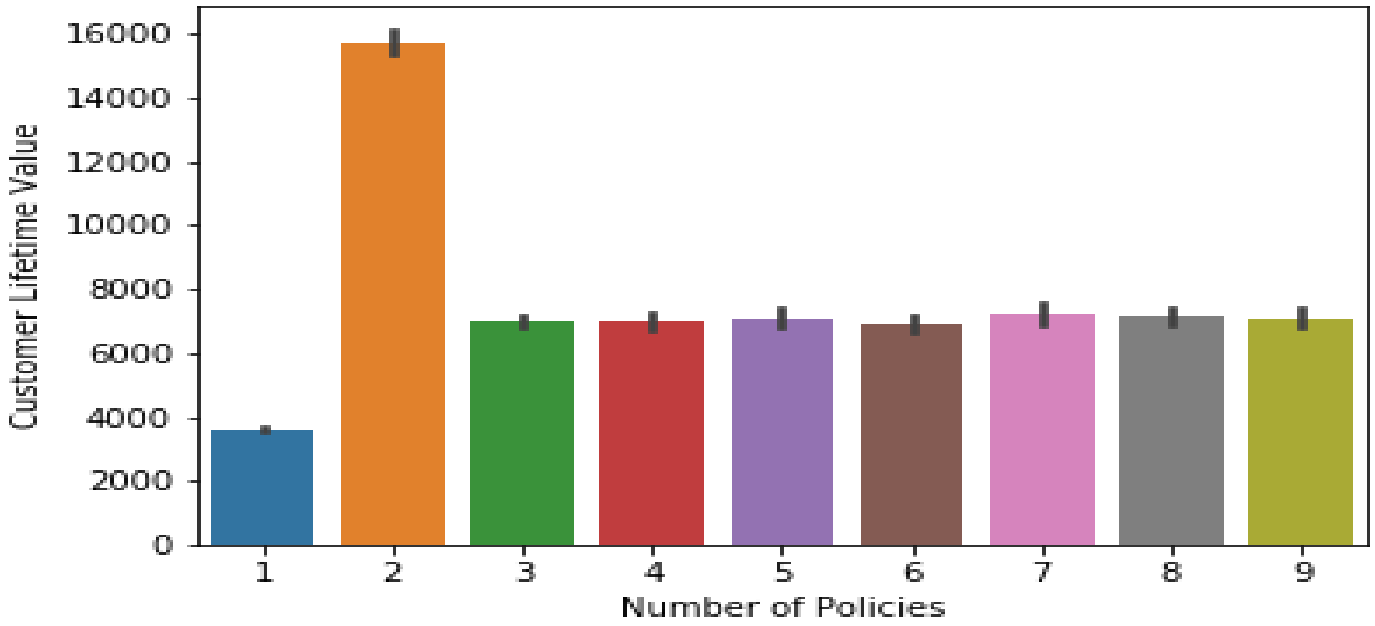
CLTV vs Renew offer type

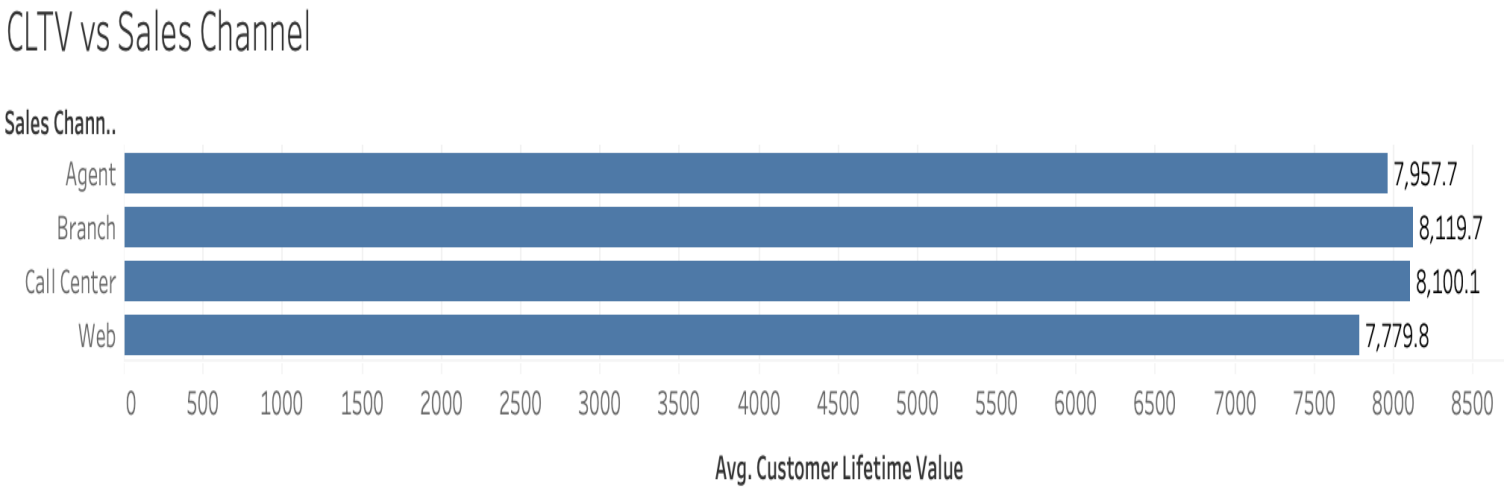


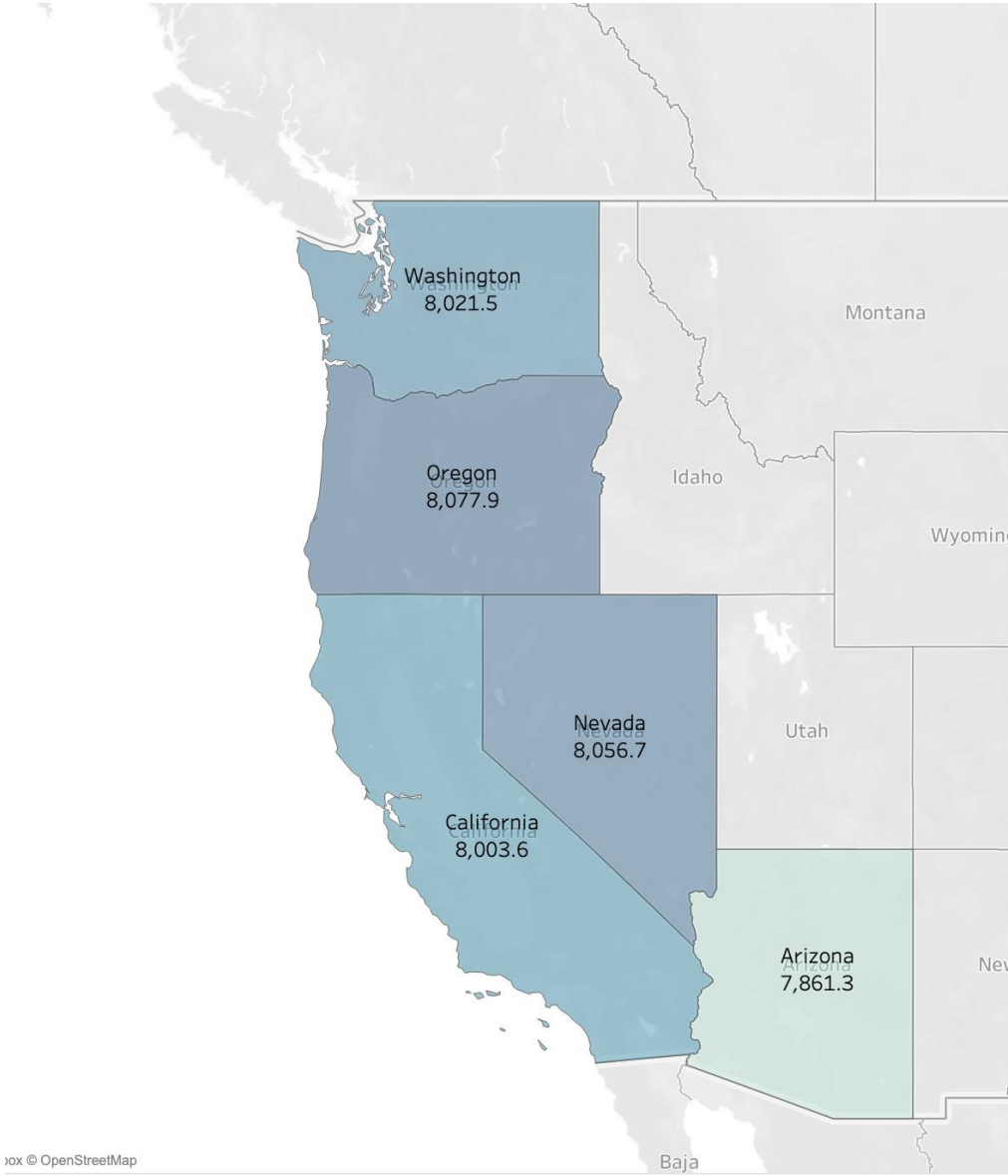
CLTV vs coverage



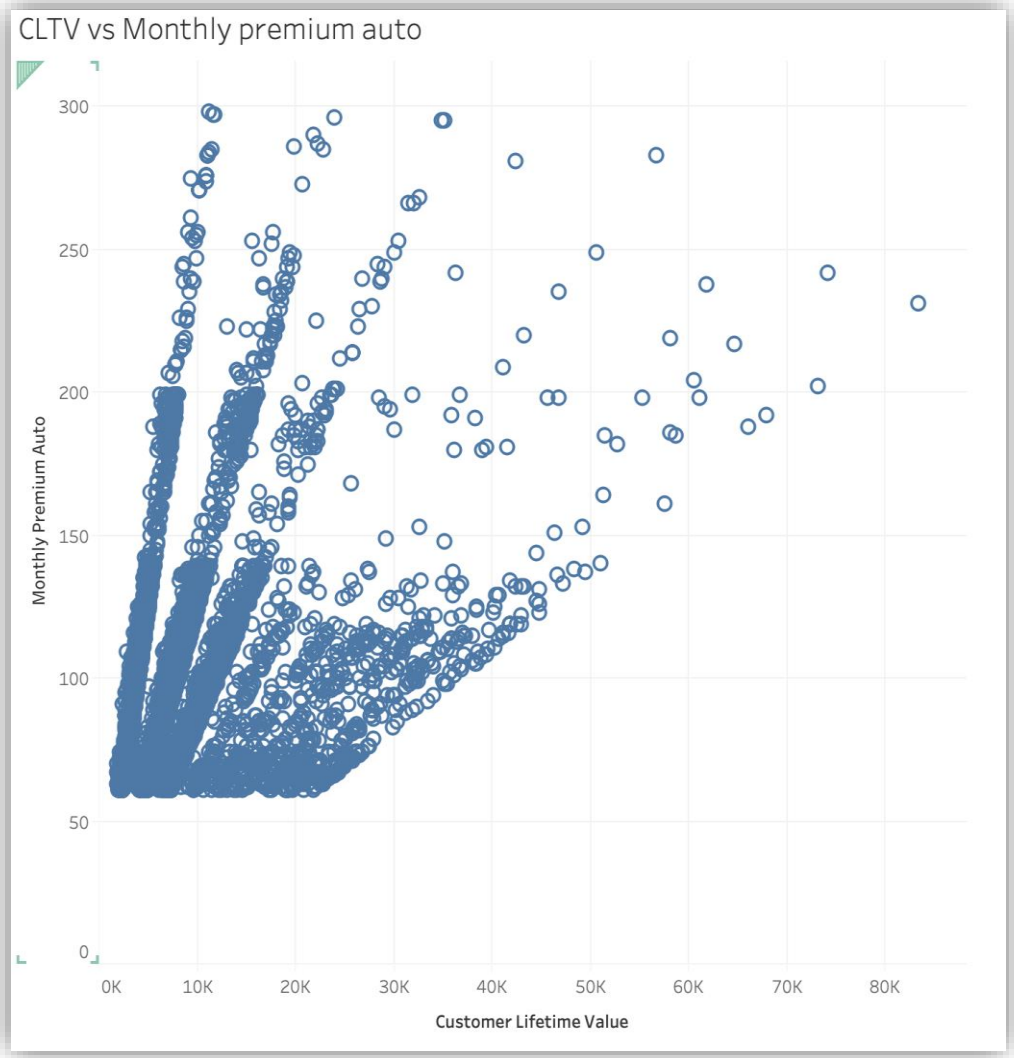
Exploratory Data Analysis

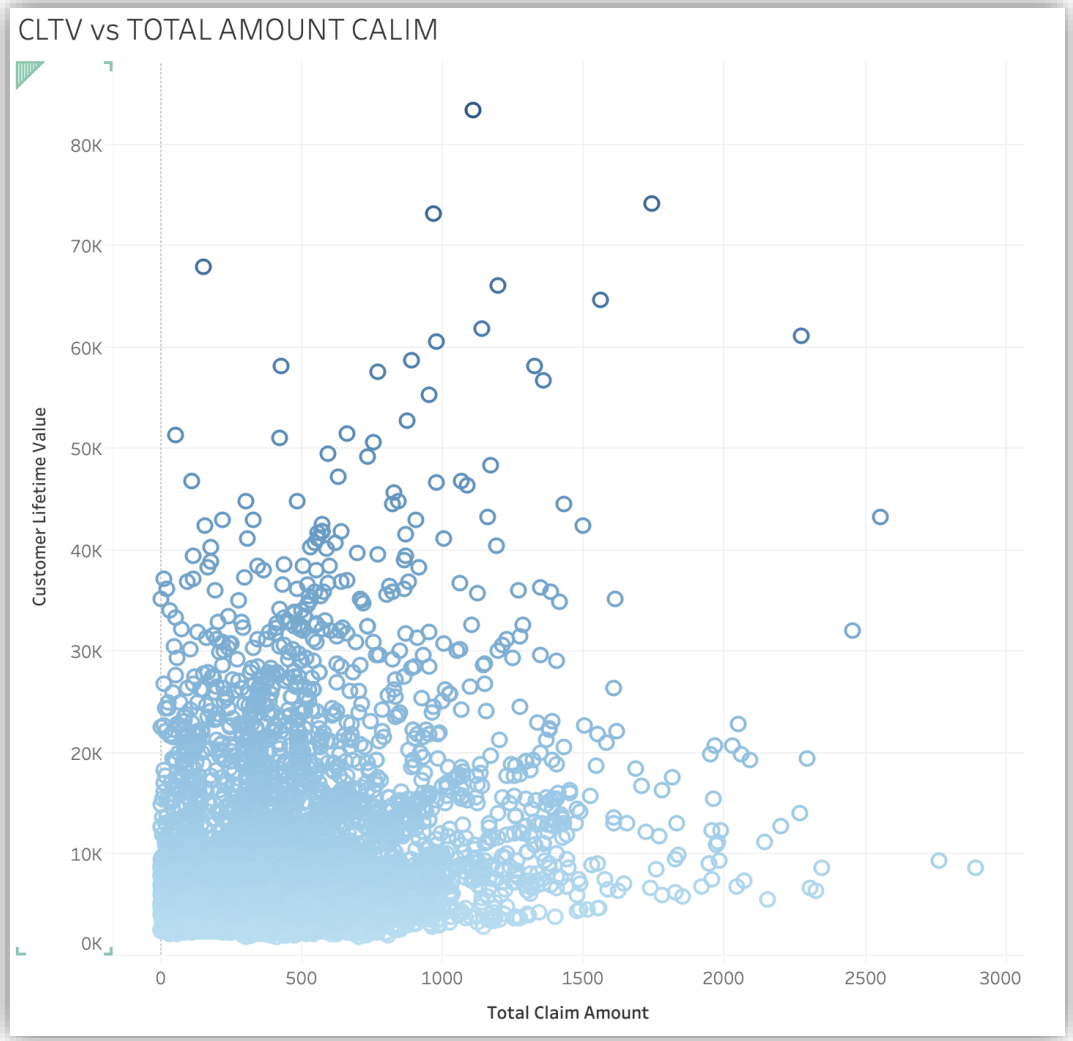




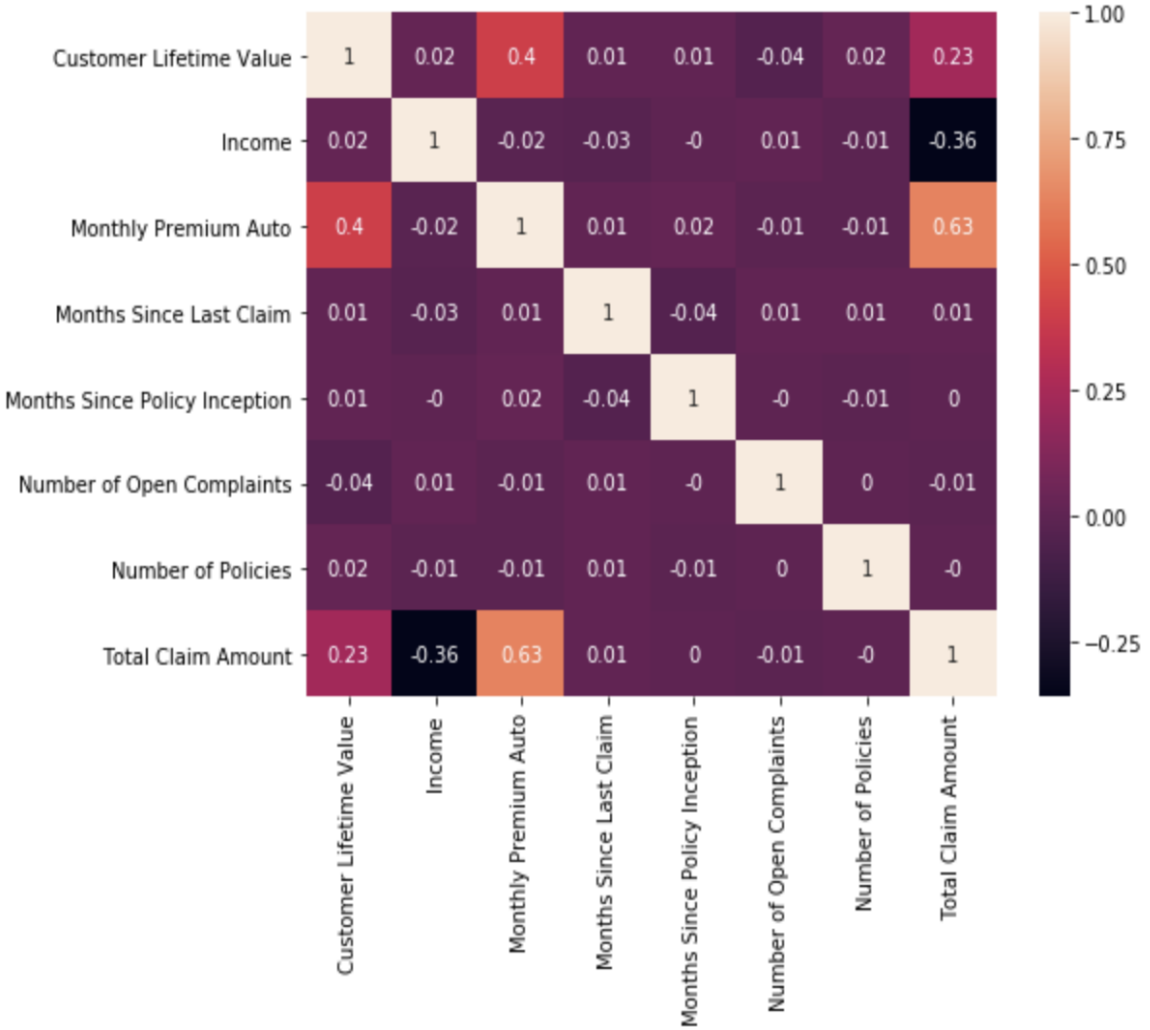


Exploratory Data Analysis





Exploratory Data Analysis



Heatmap

- With respect to our target variable, only Monthly Premium Auto & Total Claim Amount are highly correlated.
- Other than these two variables, there is no correlation amongst the independent Numerical variables

Base Model: OLS

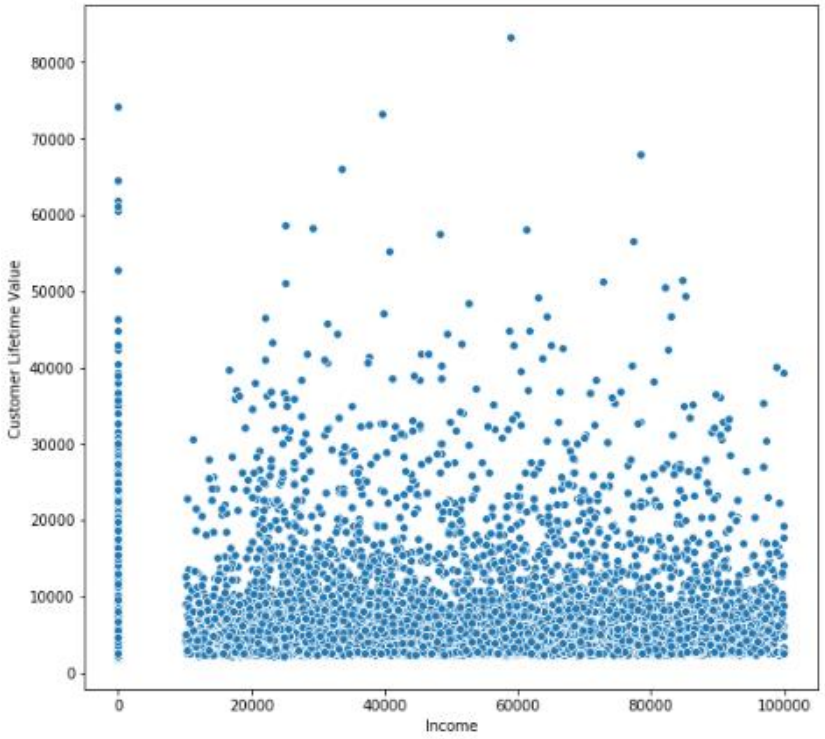
OLS Regression Results

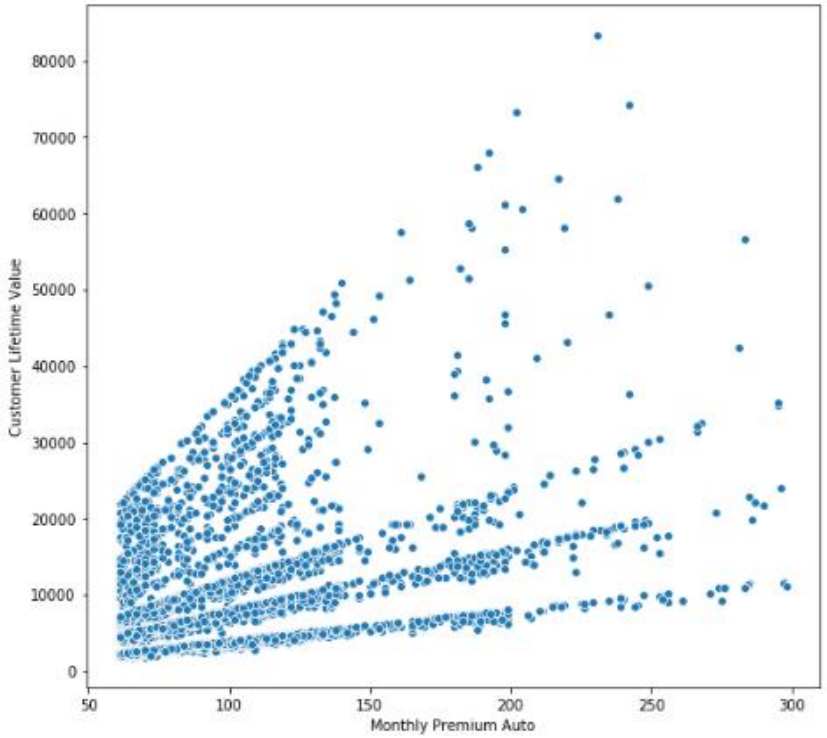
Dep. Variable:	Customer Lifetime Value	**R-squared:**	0.166
Model:	OLS	**Adj. R-squared:**	0.164
Method:	Least Squares	**F-statistic:**	78.82
Date:	Wed, 13 Nov 2019	**Prob (F-statistic):**	0.00
Time:	12:16:53	**Log-Likelihood:**	-92831.
No. Observations:	9134	**AIC:**	1.857e+05
Df Residuals:	9110	**BIC:**	1.859e+05
Df Model:	23		
Covariance Type:	nonrobust		

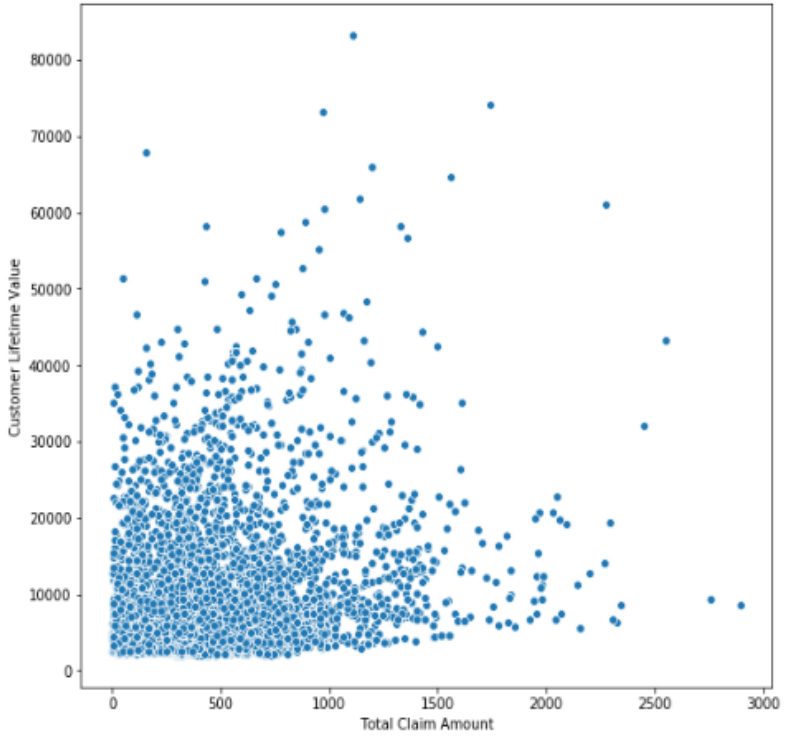
	coef	std err	t	P>|t|	[0.025	0.975]
const	872.5455	499.361	1.747	0.081	-106.315	1851.406
Customer	-0.0767	0.025	-3.073	0.002	-0.126	-0.028
State	30.3958	51.107	0.595	0.552	-69.786	130.578
Response	-434.8693	192.448	-2.260	0.024	-812.111	-57.627
Coverage	-186.4730	113.665	-1.641	0.101	-409.281	36.335
Education	93.7398	47.771	1.962	0.050	0.097	187.382
Effective To Date	-0.1302	3.883	-0.034	0.973	-7.742	7.481
EmploymentStatus	-104.9096	73.576	-1.426	0.154	-249.134	39.315
Gender	-137.3192	132.608	-1.036	0.300	-397.261	122.623
Income	0.0022	0.003	0.662	0.508	-0.004	0.009
Location Code	109.7248	116.551	0.941	0.347	-118.742	338.191
Marital Status	-237.4657	110.640	-2.146	0.032	-454.344	-20.587
Monthly Premium Auto	82.7472	2.923	28.304	0.000	77.017	88.478
Months Since Last Claim	7.0304	6.544	1.074	0.283	-5.798	19.859
Months Since Policy Inception	-0.8386	2.373	-0.353	0.724	-5.489	3.812
Number of Open Complaints	-250.4702	72.355	-3.462	0.001	-392.303	-108.638
Number of Policies	63.8915	27.655	2.310	0.021	9.681	118.102
Policy Type	441.5321	290.952	1.518	0.129	-128.798	1011.863
Policy	-48.2763	86.230	-0.560	0.576	-217.306	120.753
Renew Offer Type	-348.8194	68.467	-5.095	0.000	-483.030	-214.609
Sales Channel	22.4563	62.272	0.361	0.718	-99.610	144.523
Total Claim Amount	-0.5882	0.362	-1.626	0.104	-1.298	0.121
Vehicle Class	45.9054	32.409	1.416	0.157	-17.623	109.434
Vehicle Size	187.3131	123.322	1.519	0.129	-54.426	429.052

Assumptions for Linear Regression

1. Linear relationship exists between the dependent and predictor variables

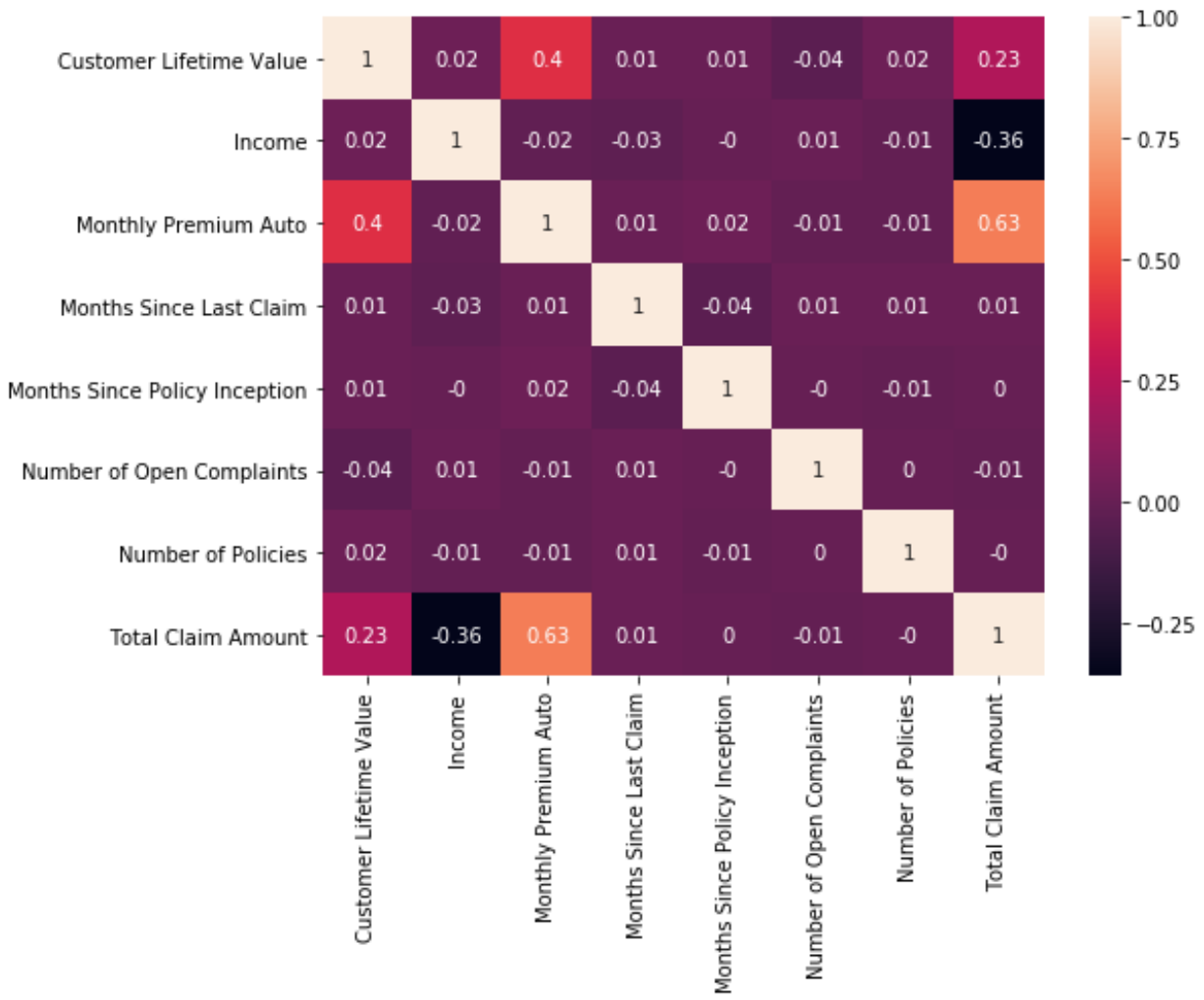






Assumptions for Linear Regression

2. Multicollinearity : Independent variables should not be correlated



Iteration : 1

VIF Factor	features
12.406466	Monthly Premium Auto
6.664016	Total Claim Amount
3.267690	Months Since Policy Inception
2.829624	Income
2.784210	Months Since Last Claim
2.296918	Number of Policies
1.165267	Number of Open Complaints

Iteration : 2

VIF Factor	features
3.065106	Months Since Policy Inception
2.667965	Months Since Last Claim
2.571419	Total Claim Amount
2.250677	Number of Policies
2.128979	Income
1.163048	Number of Open Complaints

Assumptions for Linear Regression

3. Auto Correlation : There should be no correlation between the residuals

Durbin-Watson Test :

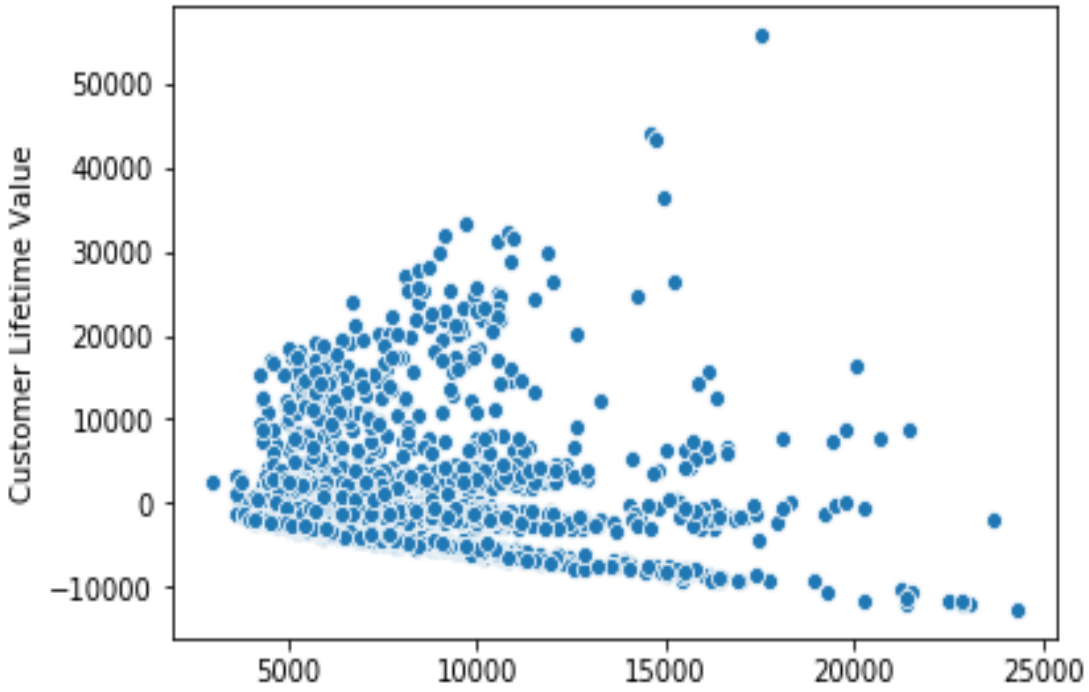
- Measure the relevance of features by their correlation with dependent variable
- Uses statistical methods for evaluation

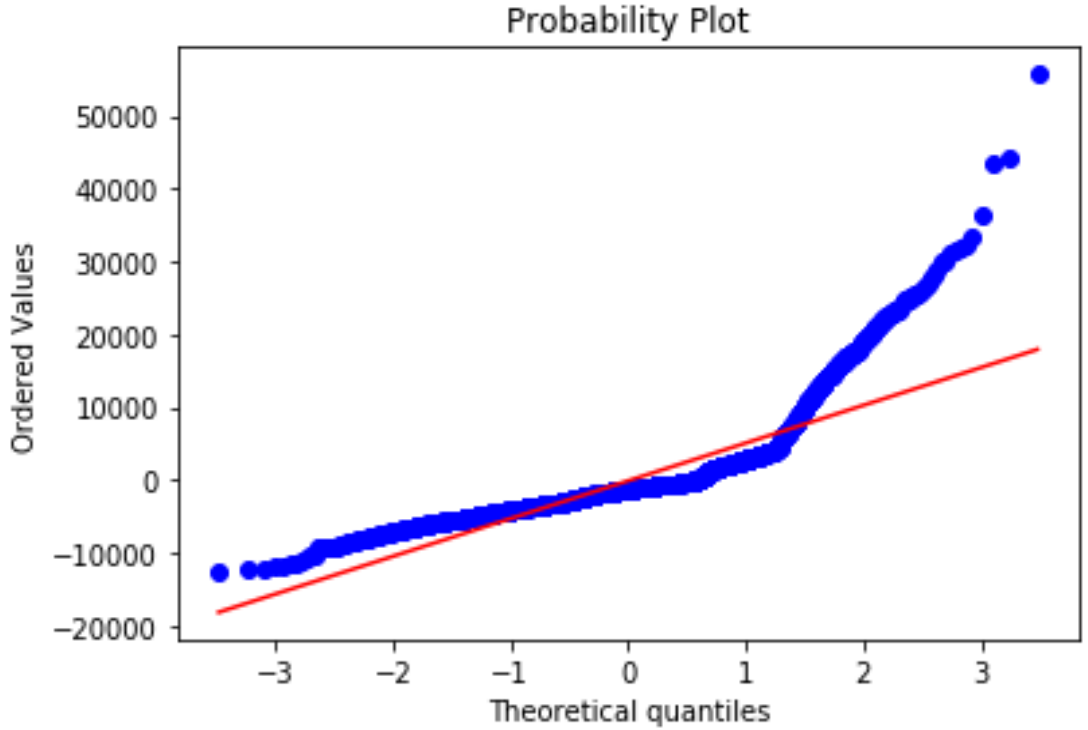
Ljung Box Test :

- Measure the relevance of features by their correlation with dependent variable
- Uses statistical methods for evaluation

Assumptions for Linear Regression

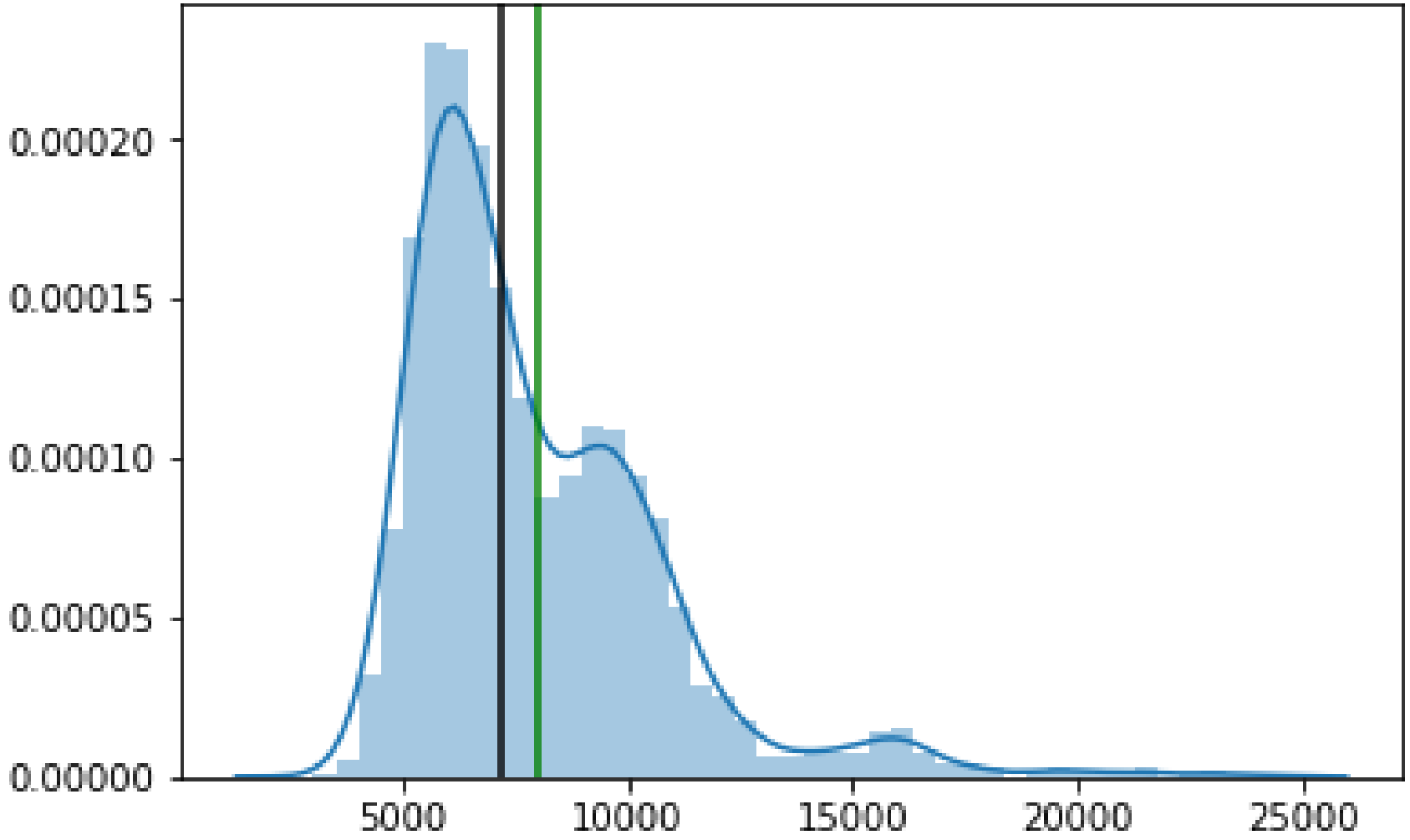
4. Homoskedasticity : Residuals must have constant variance





Assumptions for Linear Regression

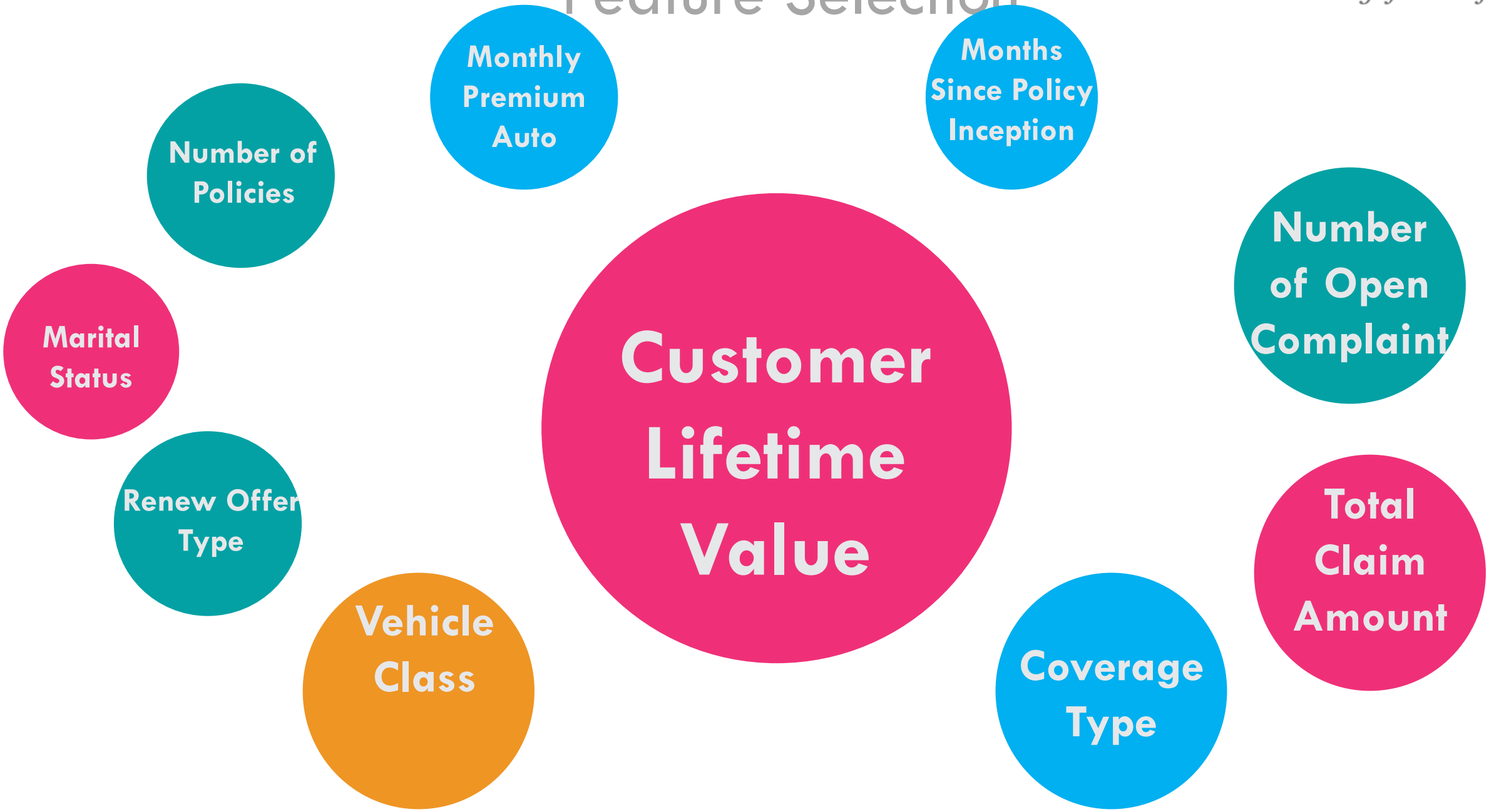
5. Residuals must be normally distributed



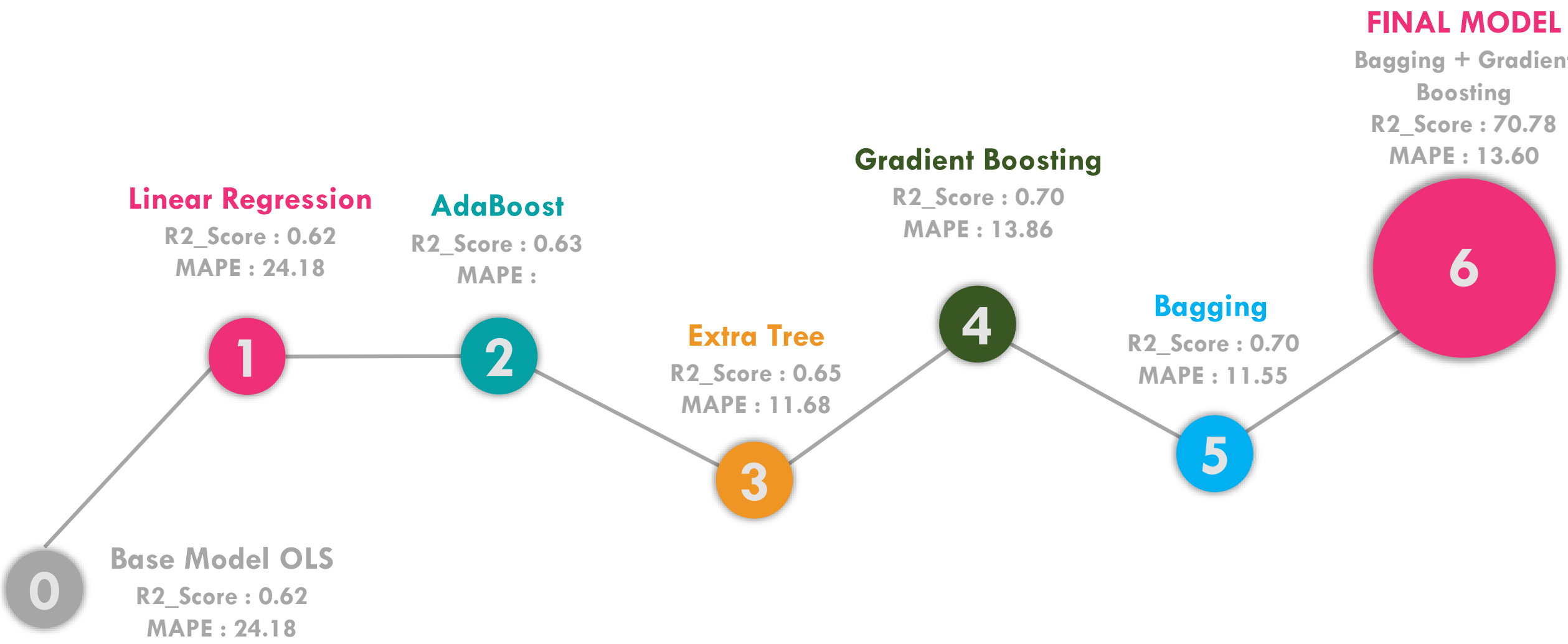
Statistical Analysis

Statistical Test	Feature	T-Statistic	P-Value
ANOVA	State	0.273	0.896
ANOVA/ttest_ind	Customer Response	0.728	0.393
ANOVA	Coverage	133.675	6.01e-58
ANOVA	Education	2.422	0.046
ANOVA	Employment Status	3.809	0.004
ANOVA/ttest_ind	Gender	1.691	0.193
ANOVA	Location Code	0.108	0.897
ANOVA	Marital Status	3.317	0.036
ANOVA	Policy	1.183	0.304
ANOVA	Renew Offer Type	25.832	1.23e-16
ANOVA	Sales Channel	0.880	0.450
ANOVA	Vehicle Class	267.158	2.08e-267
ANOVA	Vehicle Size	2.382	0.092

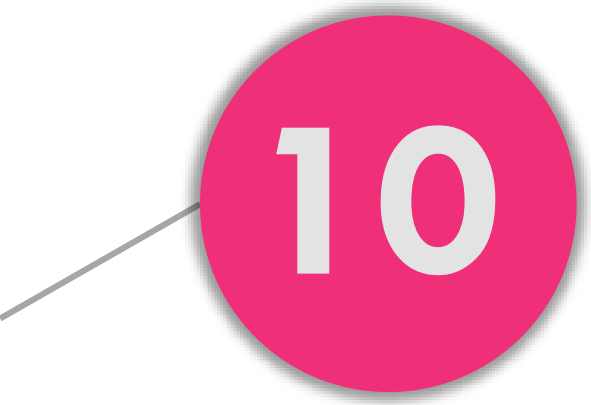
Feature Selection



Model Comparison



Magic Model



KFold Validation using Bagging regressor with base estimator Gradient Boosting.

R2_Score : 0.86
MAPE :

THANK YOU FOR YOUR TIME