

Analysis on how to spend our Marketing Dollars based on Sales Data

blue  of california

BSC's Insurance Business

BlueShield of California is in the healthcare insurance business. We sell insurance in 4 main markets.

- Individual & Family Plan (IFP)
- Employer based
- Medicare
- Small & medium businesses

Our IFP Sales Data

- Blue Shield of California(BSC) sells Insurance in the individual & family line of business in California
- BSC sells insurance on & off the state exchange.
- The plans fall into 6 different categories
 - Platinum – Silver – Catastrophic
 - Gold – Bronze – Nonmetal

Organization of Data

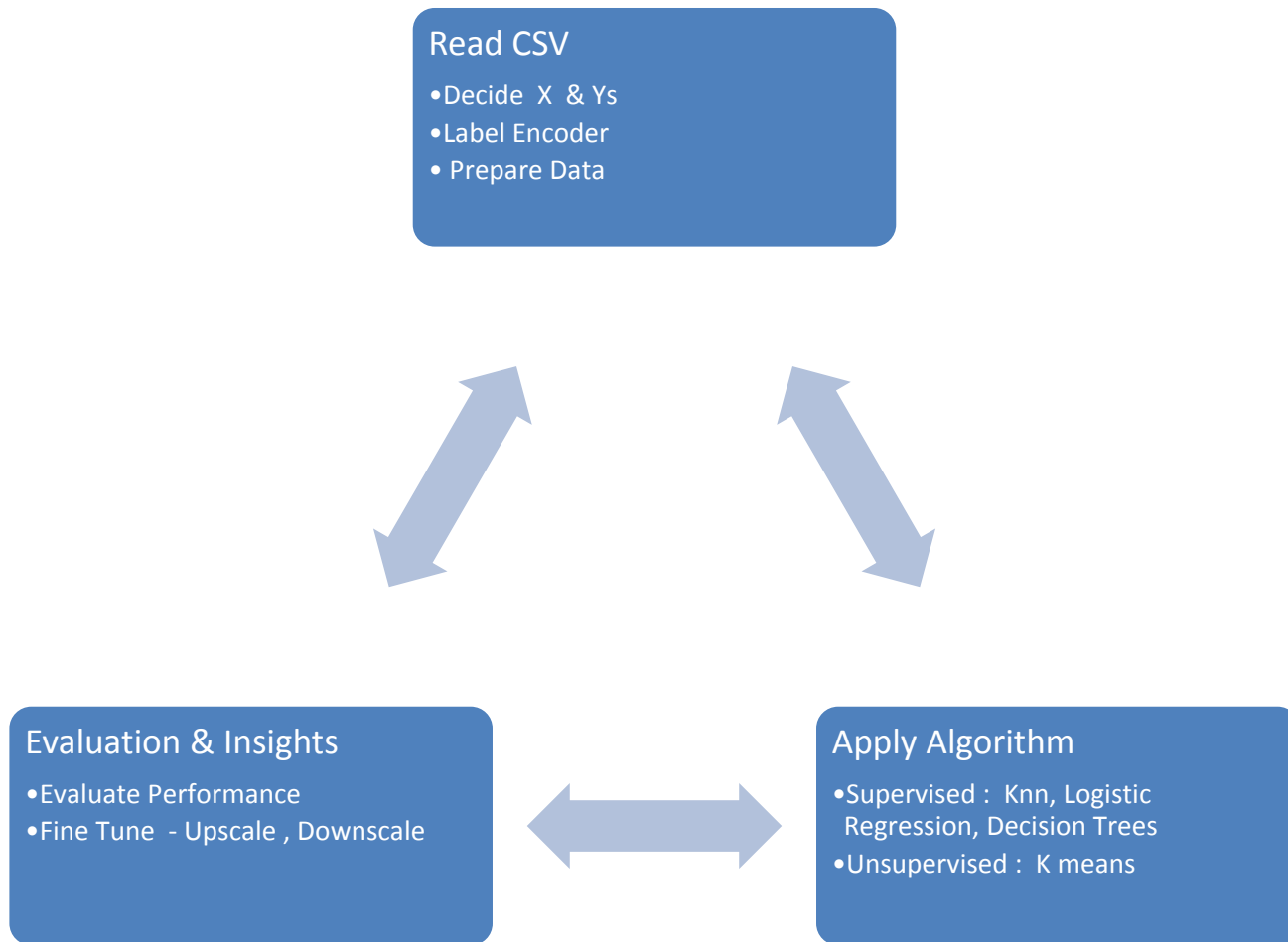
The data has been obtained in excel format.
Data is split between 2 tabs.

- Tab1: Data from 3/16 - 200 records
- Tab2: Nov 15- 2014 – March 1st - ~ 80K records

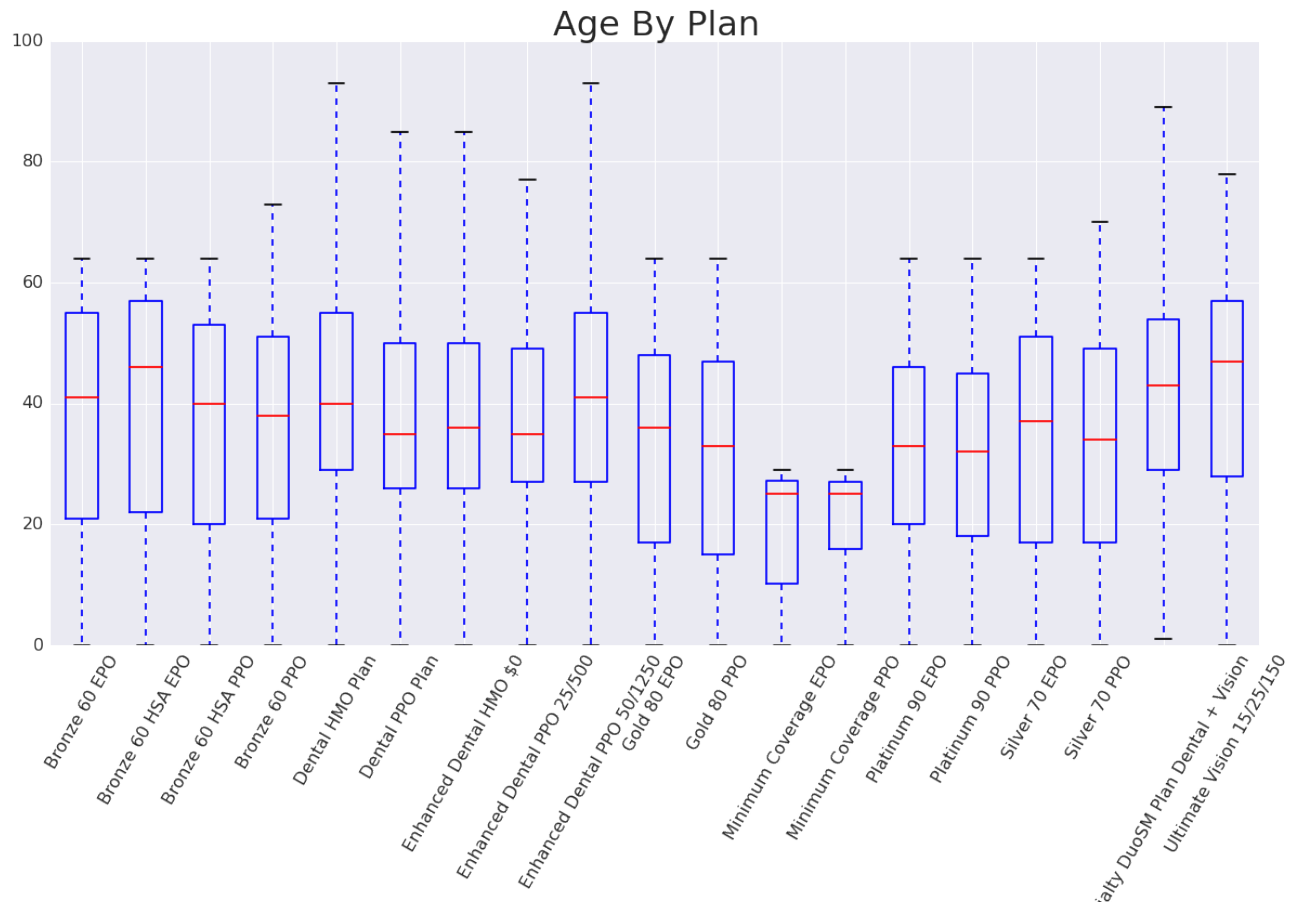
Data Elements

Gender	Age	Zip
Region	Plan Name	Language
Channel	App Submit Date	Purchase
Consumer Id	Application Type	

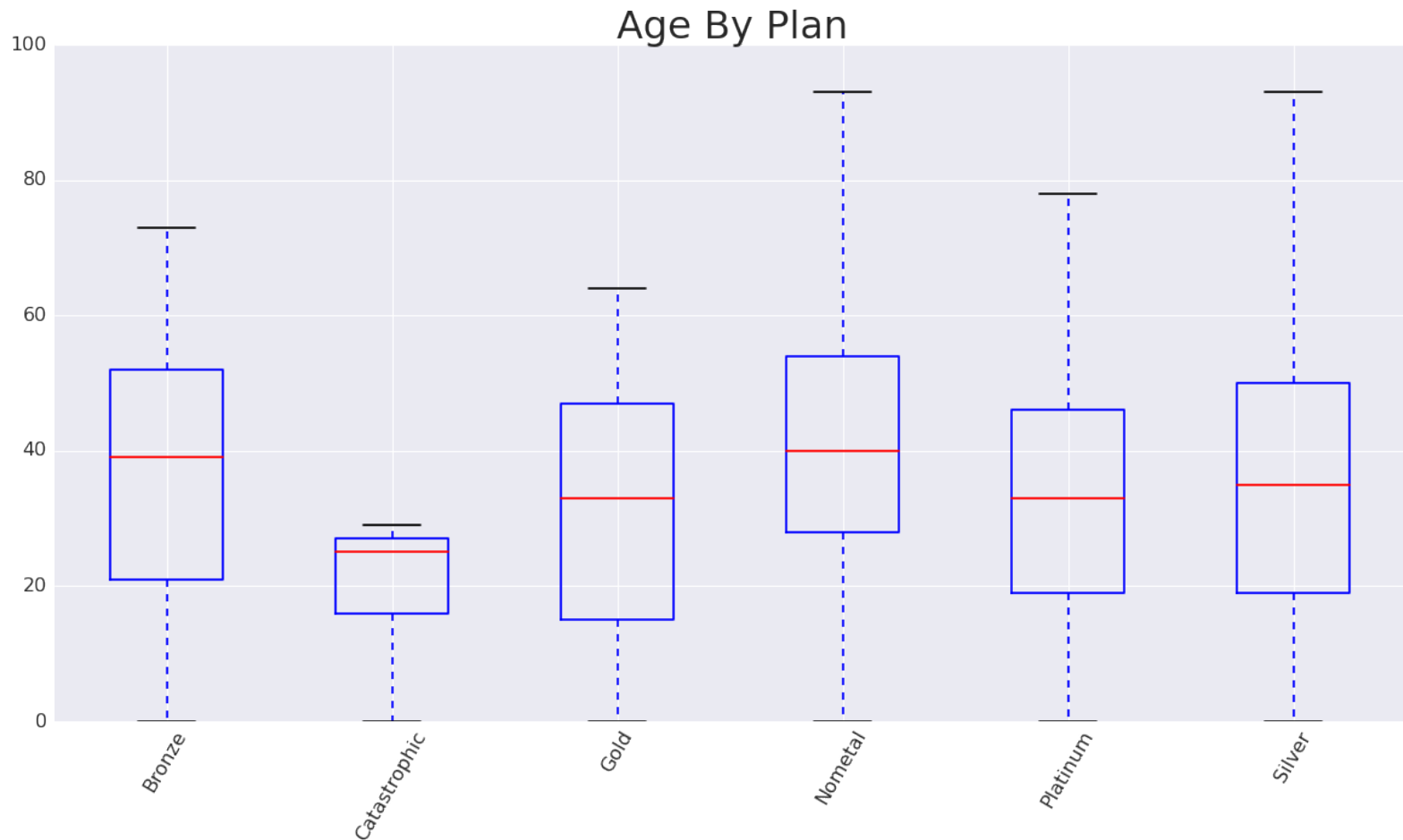
The Process



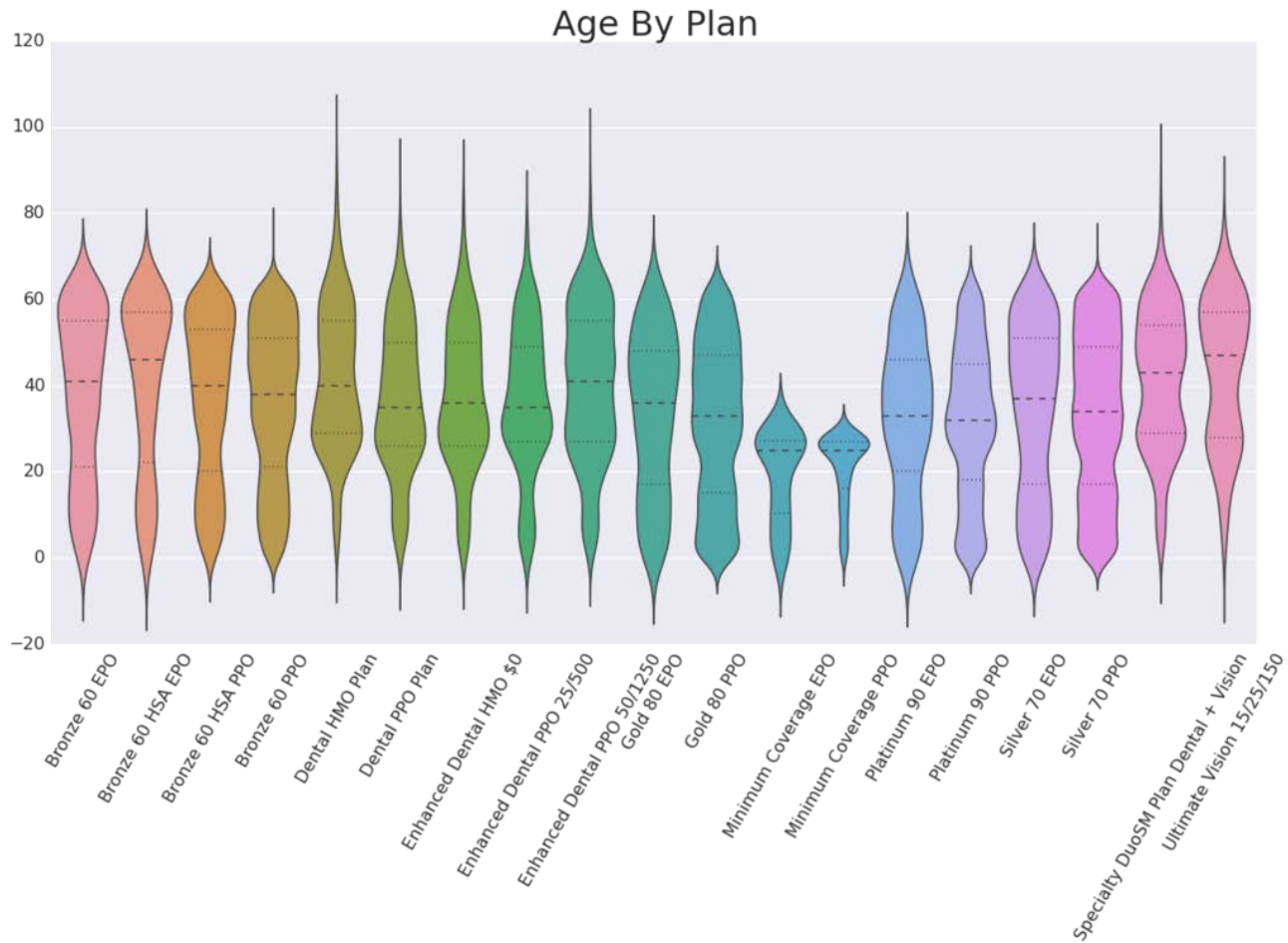
BoxPlots – Age v/s Plans



BoxPlots – Age v/s Metal Level



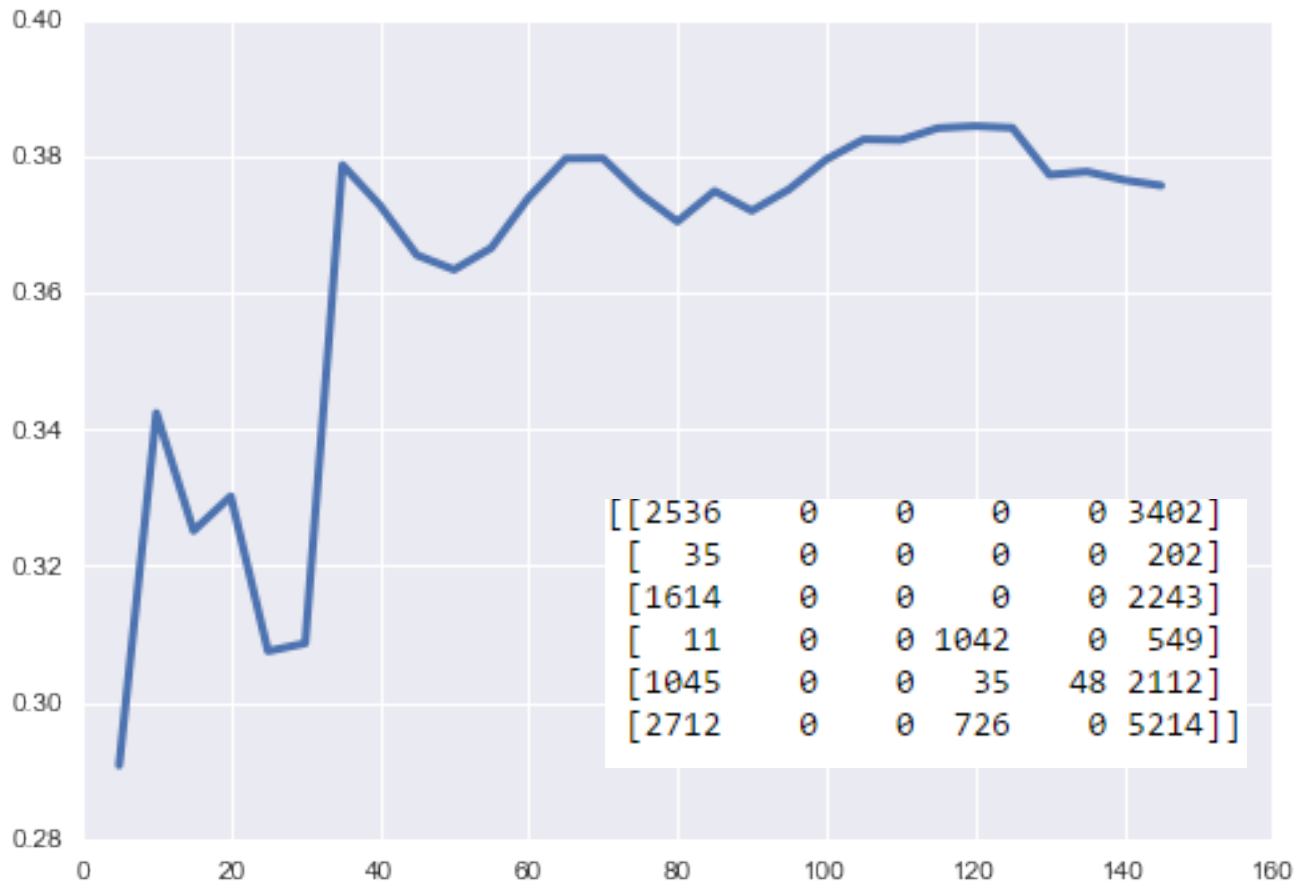
Violin Plots – Age v/s Plans



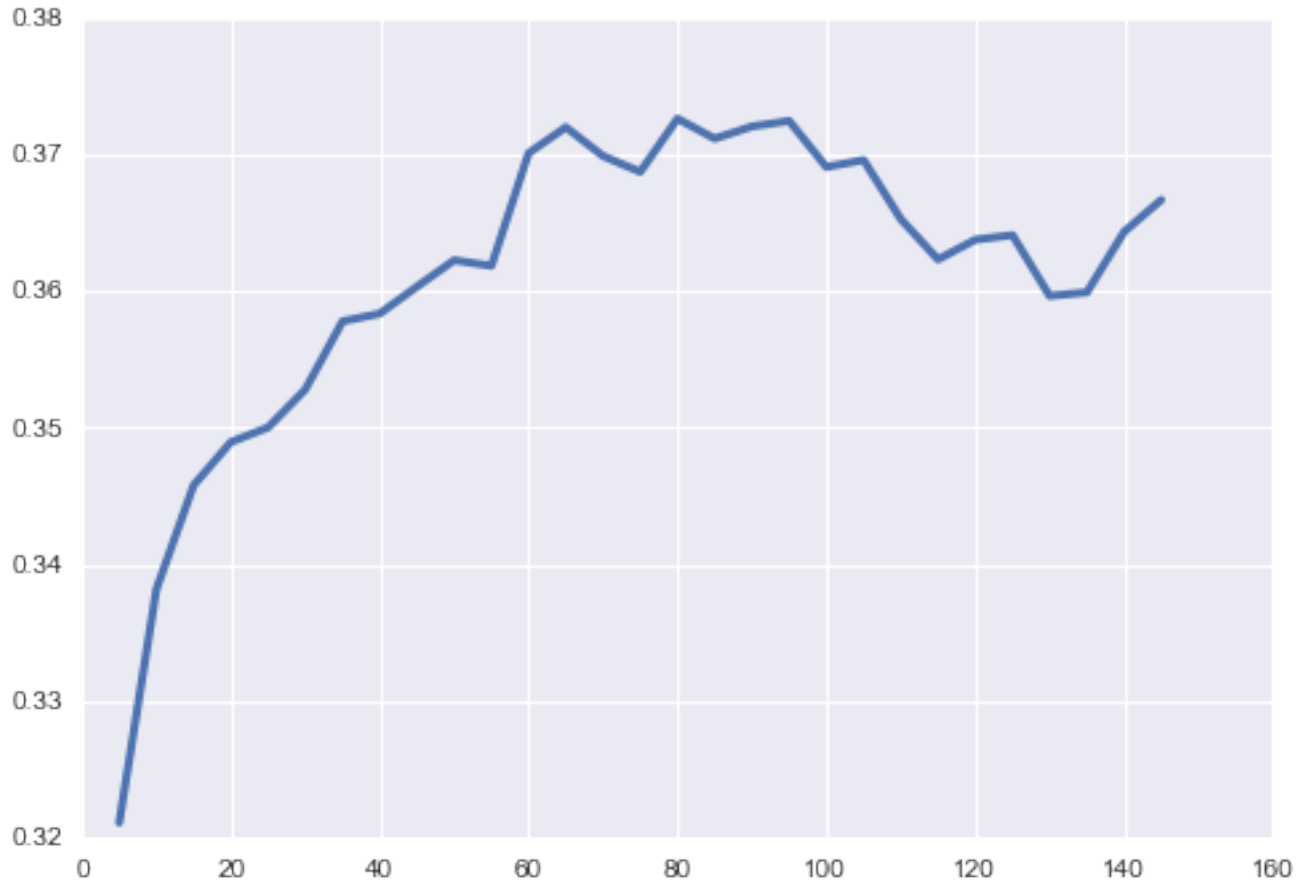
The Algorithm Matrix

	Continuous	Categorical
Supervised		X – Knn, Logistic Regression, Decision Trees,
Unsupervised		X – K Mean

Knn Classifier All v/s Metal Level



Knn Classifier with Cross Val- All v/s Metal Level



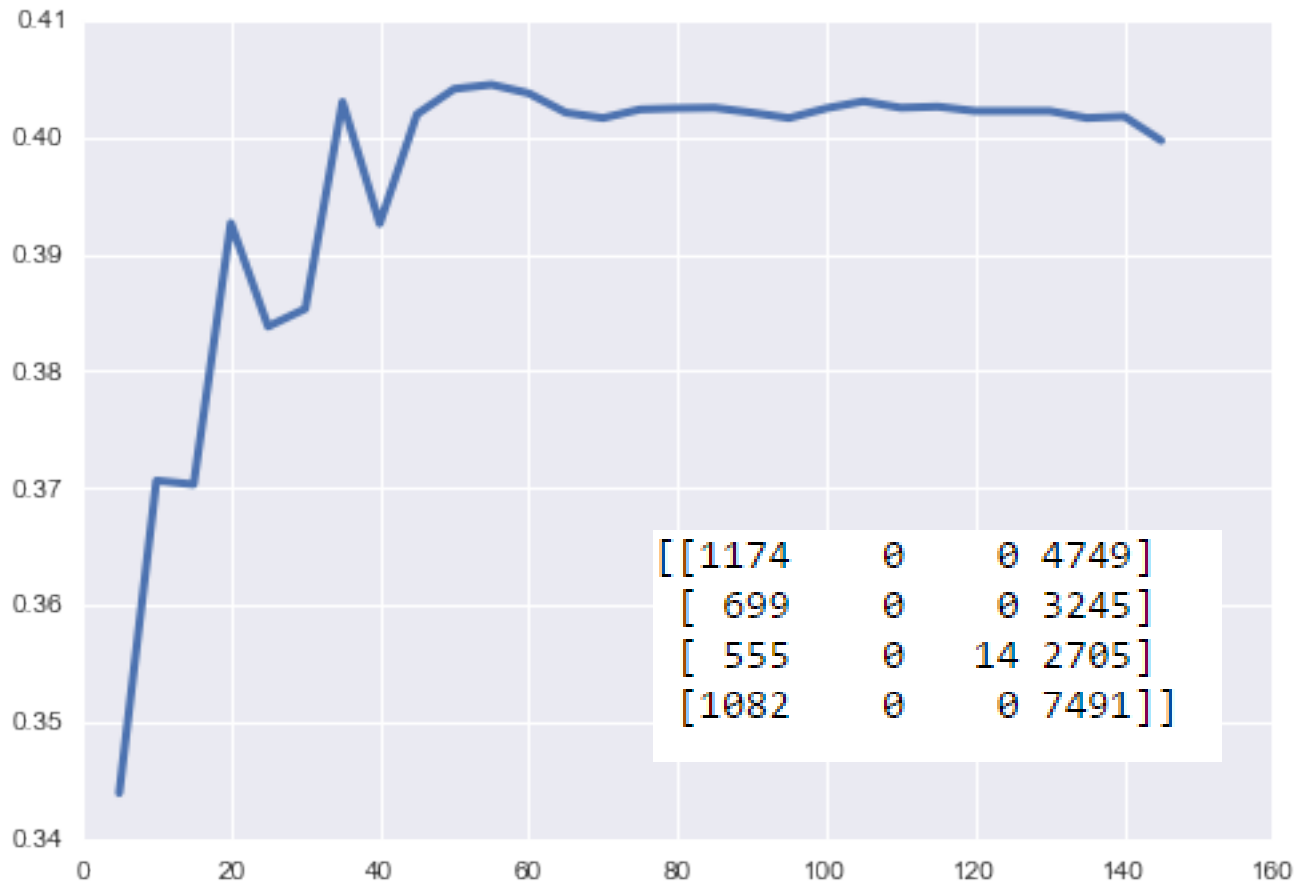
Decision Tree Classifier : All v/s Metal

```
[ [1137    0    5    0    9 3859]
  [  50    0    0    0    0  164]
  [ 591    0    5    0    2 2642]
  [   0    0    0 1089    0  224]
  [ 538    0    1    0   117 2088]
  [ 936    0    4   815    5 5324] ]
```

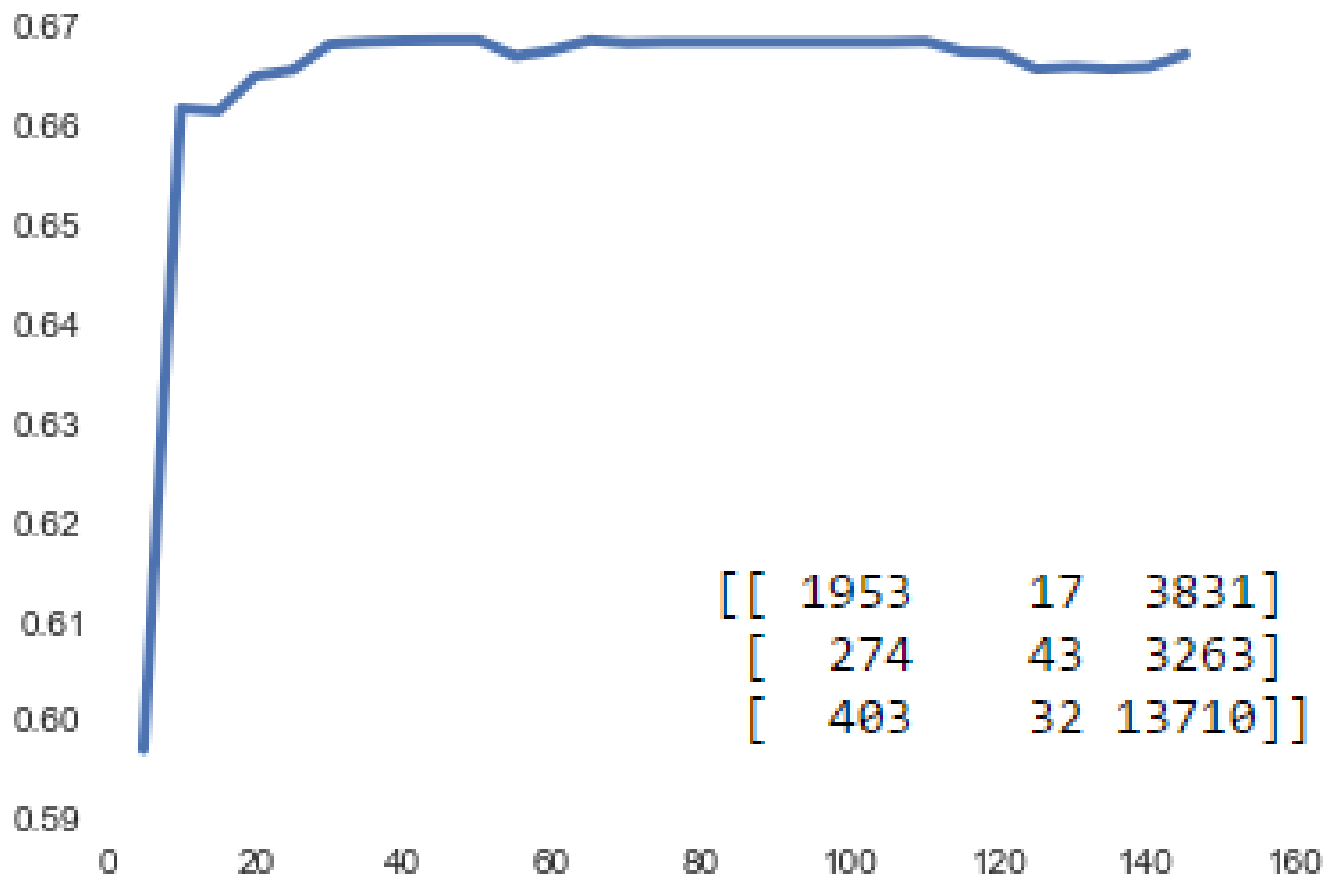
Insights

- The Data could possibly explain the poor performance
- Catastrophic has 900 Records, Non Metal 1500
- Compare it with 29K Bronze records, 32K Silver records.
- Downscale – Drop Catastrophic & Non Metal

Knn Classifier- Down Scaling



Knn Classifier – All v/s Channel

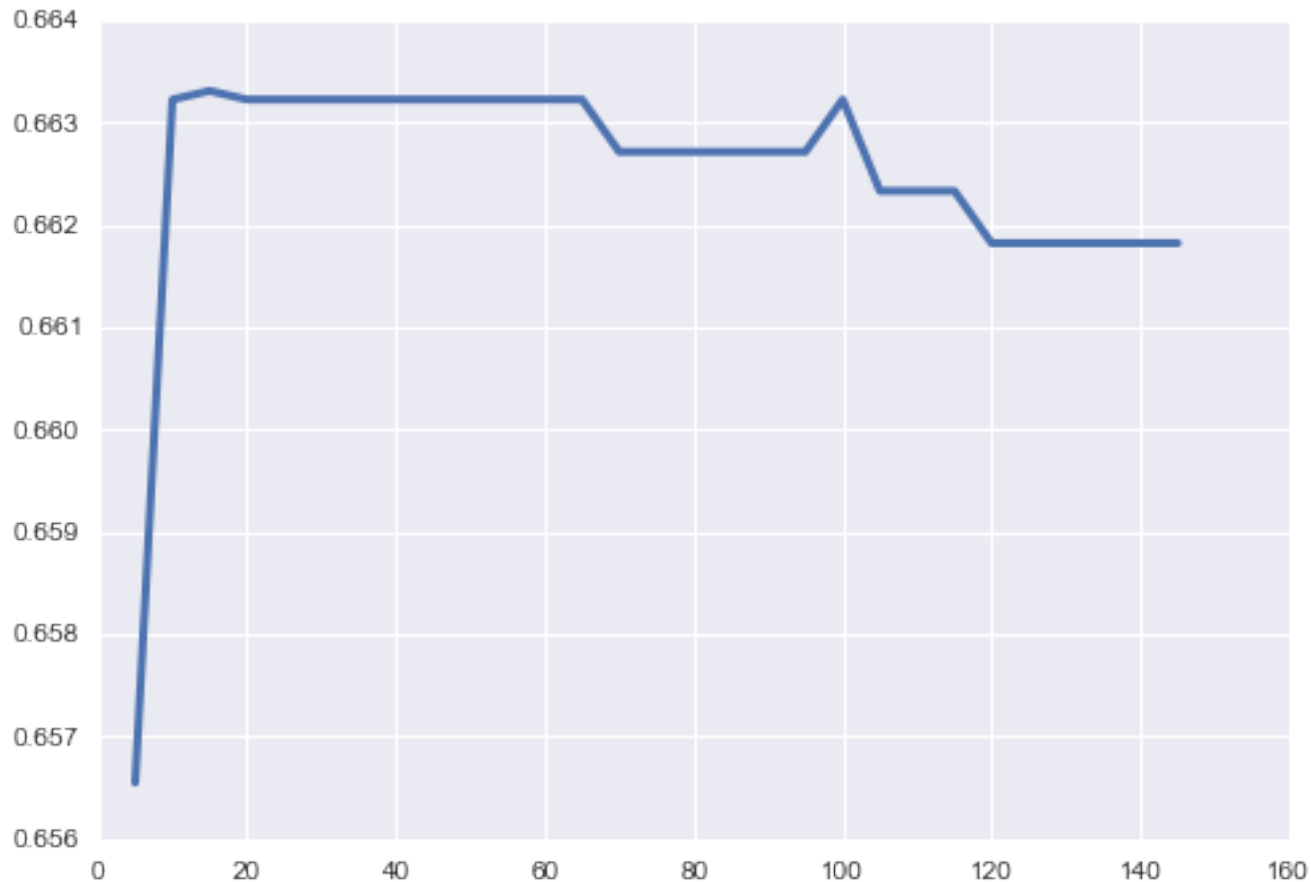


Decision Tree Classifier : All v/s Channel

```
[[2355  658 1785]  
 [ 613  754 1481]  
[2061 1788 8110]]
```

	feature	importance
2	Zip	0.389663
1	Age	0.247807
8	Application Type	0.147449
0	Gender	0.057055
4	PlanName	0.051625
3	Region	0.034961
7	consumer_id	0.034488
6	Purchase	0.025786
5	languageDesc	0.011167

Only Important Ones v/s Channel



Logistic Regression: All v/s Channel

	precision	recall	f1-score	support
DirectProspect	0.92	0.19	0.32	3832
DirectSales	0.00	0.00	0.00	2295
Producer	0.64	1.00	0.78	9557
avg / total	0.61	0.65	0.55	15684

[[745	0 3087]
[34	0 2261]
[35	0 9522]]

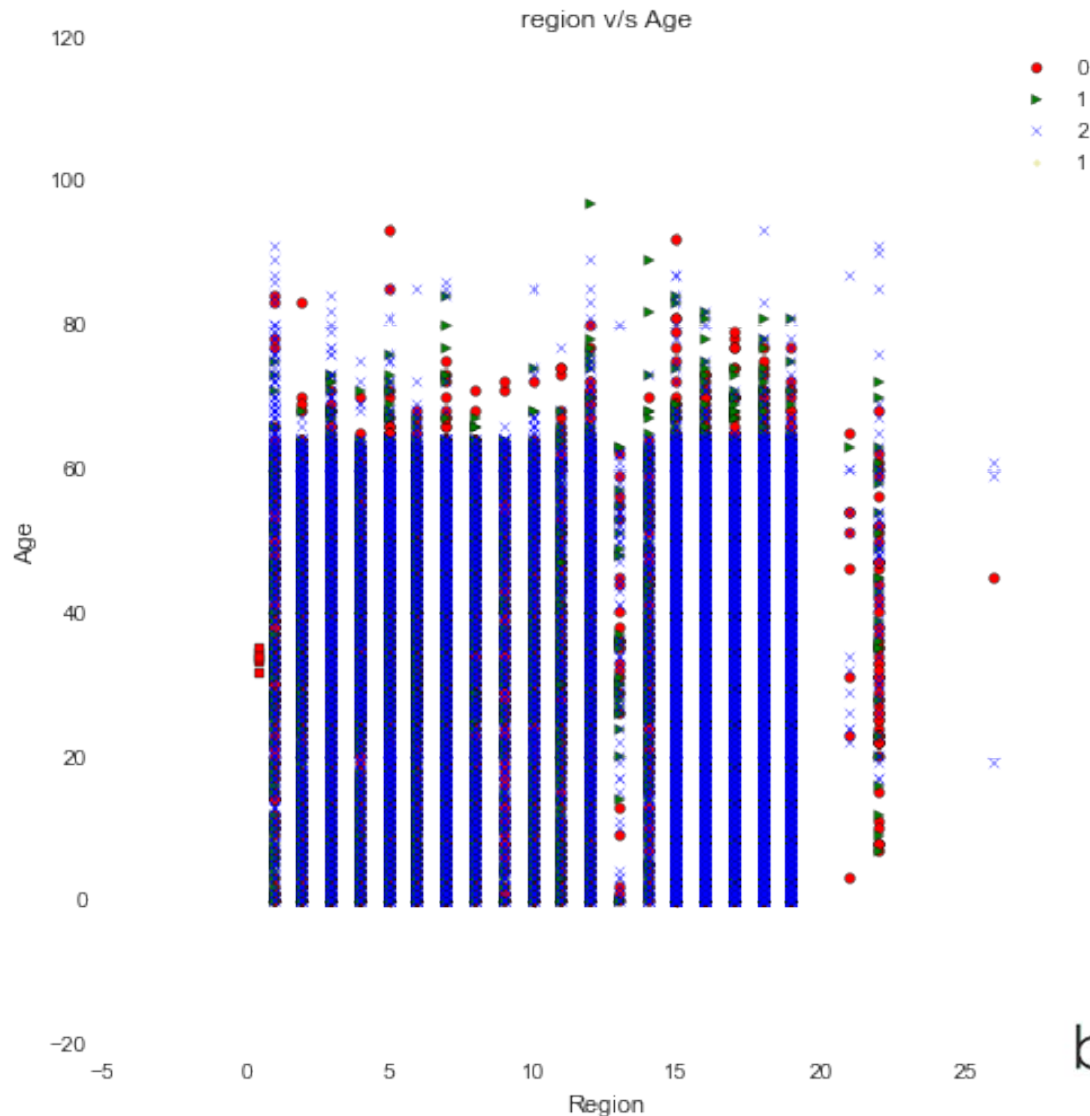
Decision Tree Classifier : All v/s family

```
[[9423 3000]
```

```
[3139 4043]]
```

```
Accuracy: 0.69 (+/- 0.00)
```

Unsupervised Learning- Kmeans(6 Clusters, Age v/s Region)



Insights

- Silver is our most popular plan – Stating the obvious 😊
- We sell a lot of Bronze plans in Region 22(Needs more analysis). We can fine tune our marketing campaigns in this region.
- We are not selling in some of the regions – 21, 22, 24,25 – (Needs more analysis)
- Senior Citizens are buying bronze plans.

Project Takeaways

- The term “Machine Learning” could be misleading.
 - . You have to do all the work
- Columns are better than Rows
- Keep an open mind
- Vect Dictionary way better than Label Encoder

APPENDIX

Knn Classifier

- X –['languageDesc', 'Channel', 'Purchase', 'consumer_id', 'Application Type', 'Gender', 'consumer_id']
- Y –[Purchase]