

PREDICTING BRIBING PATTERNS USING LAND RECORDS...

Hussain Boltwala (61210213)

Karthik Vemparala (61210505)

Naveen Kumar HS (61210144)

Salman Siddiqui (61210626)

Smita Chakravorty (61210558)

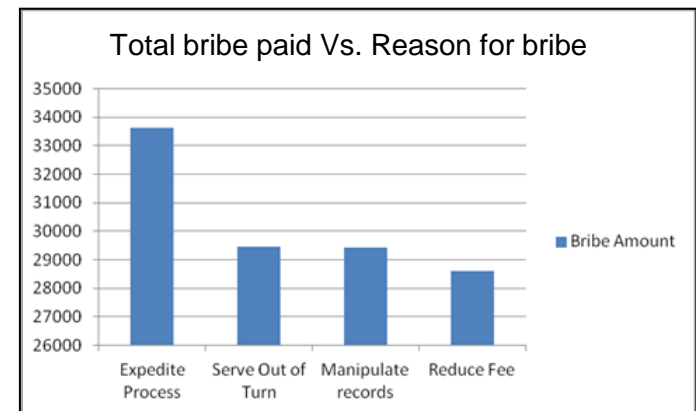
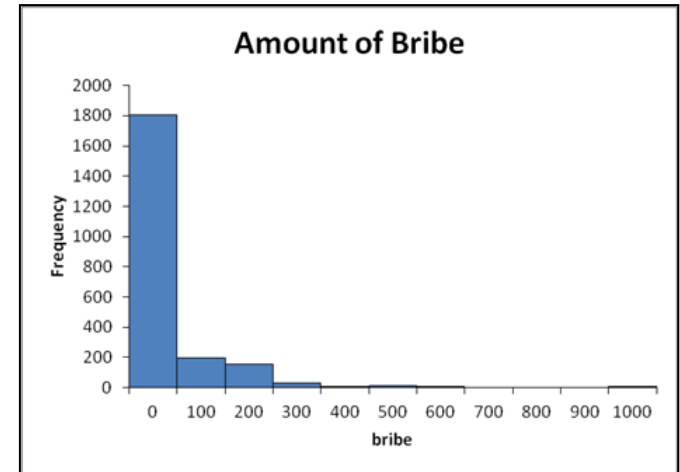
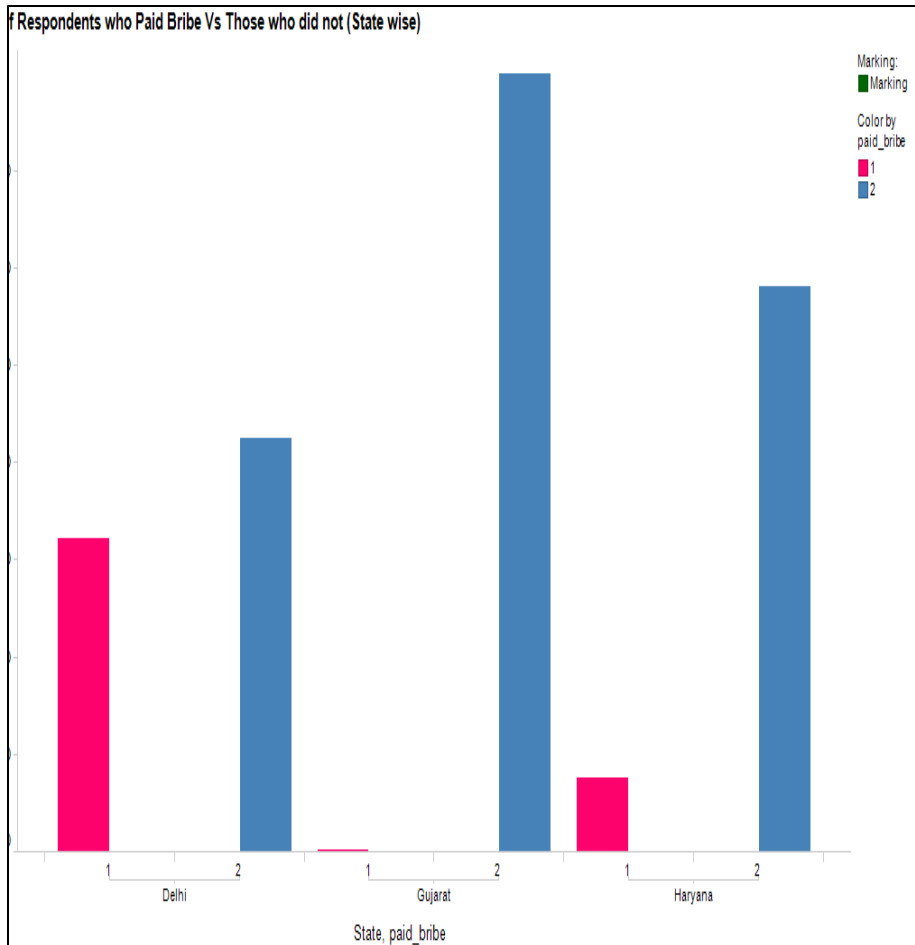
Brisk Business for Brokers

- Predict whether a person availing the e-Governance Service will pay a bribe more INR 100.
- Use it as a predictive tool to approach new clients for “brokerage” services.
- Study of land record data using e-Governance.

It collected the following:

- Demography (Age, Income Level, Education Level, Gender)
- Type of services availed using e-Governance
- Feedback on service experience
- Land and housing data

Visualizing the Data



Classification Trees – Finding Predictors

Main Prediction Variables using Classification Trees

Wage loss	Service charges
Wait time	Total payment
Age	Level of education
Occupation	Mode of travel
No. of trips made to the kiosk	Travel time
Expediting process	

K-NN

&

Naïve Bayes

Training Data scoring - Summary Report (for k=1)			
Cut off Prob.Val. for Success (Updatable)		0.5	
Classification Confusion Matrix			
	Predicted Class		
Actual Class	1	0	
1	168	0	
0	0	932	
Error Report			
Class	# Cases	# Errors	% Error
1	168	0	0.00
0	932	0	0.00
Overall	1100	0	0.00
Validation Data scoring - Summary Report (for k=1)			
Cut off Prob.Val. for Success (Updatable)		0.5	
Classification Confusion Matrix			
	Predicted Class		
Actual Class	1	0	
1	80	35	
0	47	498	
Error Report			
Class	# Cases	# Errors	% Error
1	115	35	30.43
0	545	47	8.62
Overall	660	82	12.42
Test Data scoring - Summary Report (for k=1)			
Cut off Prob.Val. for Success (Updatable)		0.5	
Classification Confusion Matrix			
	Predicted Class		
Actual Class	1	0	
1	52	23	
0	24	341	
Error Report			
Class	# Cases	# Errors	% Error
1	75	23	30.67
0	365	24	6.58
Overall	440	47	10.68

Training Data scoring - Summary Report			
Cut off Prob.Val. for Success (Updatable)		0.5	
Classification Confusion Matrix			
	Predicted Class		
Actual Class	1	0	
1	136	32	
0	57	875	
Error Report			
Class	# Cases	# Errors	% Error
1	168	32	19.05
0	932	57	6.12
Overall	1100	89	8.09
Validation Data scoring - Summary Report			
Cut off Prob.Val. for Success (Updatable)		0.5	
Classification Confusion Matrix			
	Predicted Class		
Actual Class	1	0	
1	75	40	
0	44	501	
Error Report			
Class	# Cases	# Errors	% Error
1	115	40	34.78
0	545	44	8.07
Overall	660	84	12.73
Test Data scoring - Summary Report			
Cut off Prob.Val. for Success (Updatable)		0.5	
Classification Confusion Matrix			
	Predicted Class		
Actual Class	1	0	
1	48	27	
0	27	338	
Error Report			
Class	# Cases	# Errors	% Error
1	75	27	36.00
0	365	27	7.40
Overall	440	54	12.27

Results...

- Knowing the age, level of education, occupation wage loss (income), charges for the services availed, total payment to be made, mode of travel and travel time will help us predict whether or not the individual will pay a bribe greater than Rs 100 with an accuracy of 12% using K-NN and 12 % using Naïve Bayes.