Business Analytics using Data Mining

Project Presentation

Your Travel Time

Travel Time Prediction

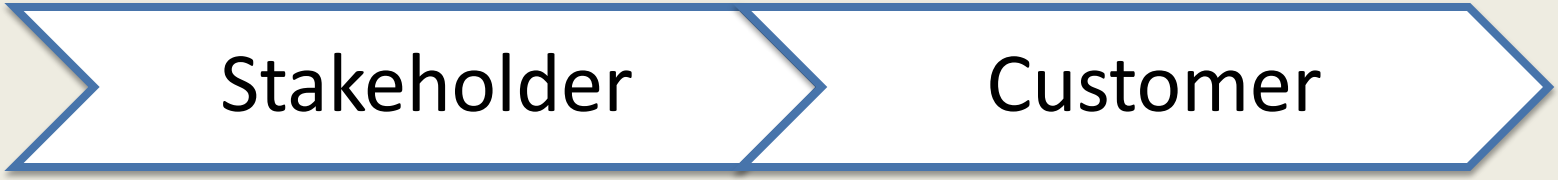
Group – A6

Bhushan Khandelwal – 61410182 | 61410806 - Mahabaleshwar Bhat

Mayank Gupta – 61410659 | 61410697 - Shikhar Angra

Sujay Koparde - 61410812

Business Problem



Opportunity

- **Scenario 1**: A customer books a cab for travelling to airport. Not sure about the travel time she books cab with booking time well in advance of her flight time. She reaches airport 2 hours before the flight and now has to somehow spend those 2 hours.
- **Scenario 2**: A businessman goes to a new city for business meeting. Books a cab from hotel to client office for an early morning meeting. Not aware of the slow travel time on the route, he books the cab closer to meeting time and reaches late for the meeting

Value

- Better Customer Value: A customer booking a cab would be able to estimate the travel time which would allow her plan the travel in a better way.
- This would ideally lead to time savings for the customer.
- For the company this could translate into higher customer acquisition and retention rates.

Data mining Problem

Objective : Predict the time required to reach destination from the source

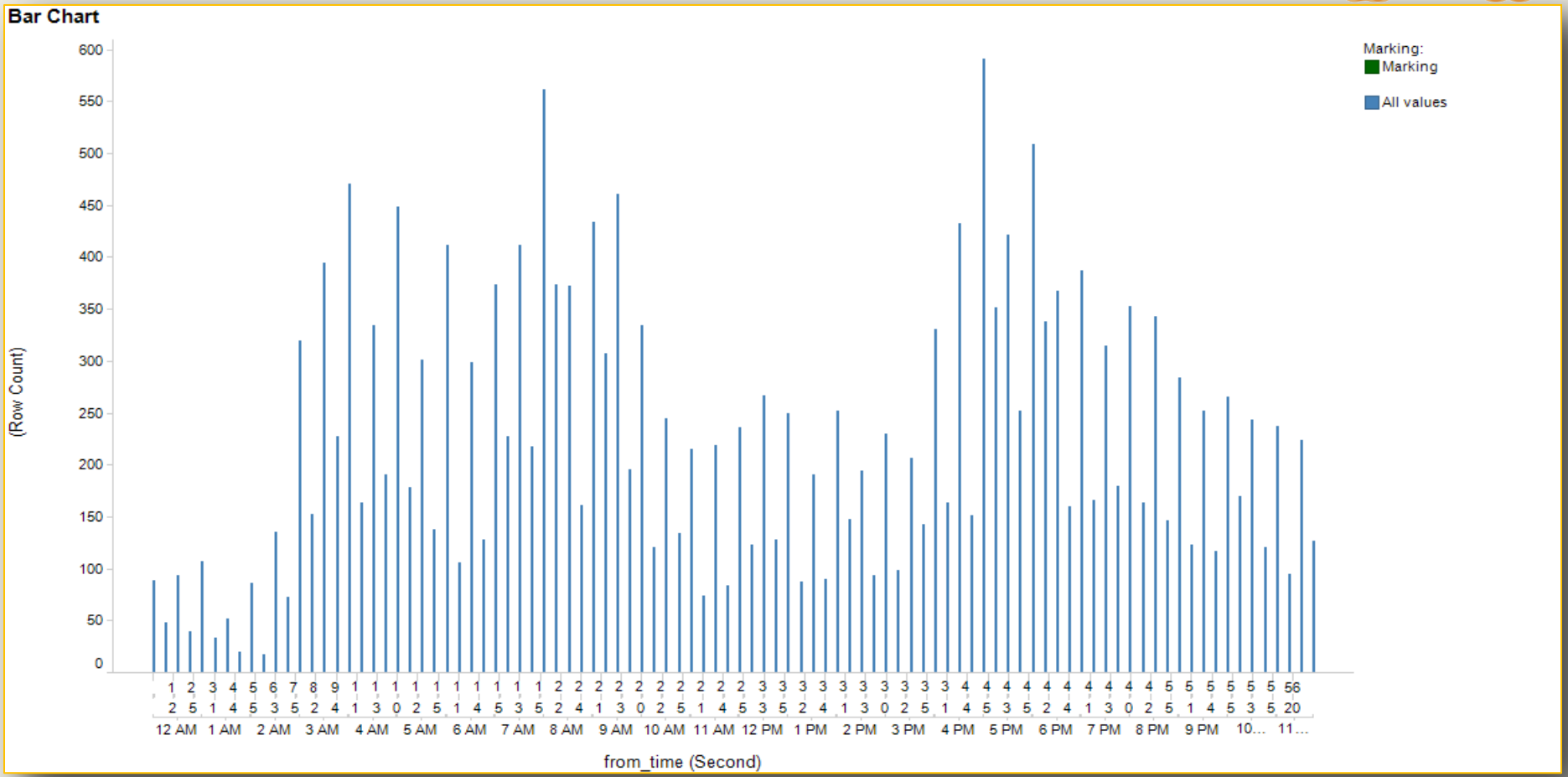
- Supervised data mining technique with TIME as output variable

- Predictive analytics
  - Time (in minutes) is a continuous variable
  - Predictors are available at time of prediction

Why predictive?

- Business problem is about estimation of time

Data Description



Methods

We used two most popular methods for prediction:

Multiple Linear regression

The Regression Model

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	27.63209534	2.50517654	0	24986870
dSmallVehicle	-8.19571781	1.9801811	0.00004101	12670.71484
dSedan	-7.69910574	2.1125114	0.00029614	15585.70313
dWeekday	1.21291506	0.88571805	0.17148946	296.6911926
dMorning	32.16956711	1.35973084	0	183611.8125
dAfternoon	23.90462112	1.31217837	0	58820.14063
dEvening	41.53961945	1.15913796	0	565248.5625
Distance	1.77110898	0.05189084	0	833813.3125

Residual df	3658
R-squared	0.389447201
Std. Dev. estimate	26.7534523
Residual SS	2618203

K-Nearest Neighbors (K-NN)

Validation error log for different k

Value of k	Training RMS Error	Validation RMS Error	
1	10.75977374	26.43684402	
2	11.20844488	25.37828334	
3	11.2711506	25.03700329	
4	11.27911672	24.84265342	
5	11.28233938	24.73272635	
6	11.28268382	24.66610855	
7	11.28266233	24.62589106	
8	11.28557082	24.59279435	
9	11.28724586	24.56662964	
10	11.28730541	24.54182235	<--- Best k

Evaluation

Multiple Linear Regression

Training Data scoring - Summary Report

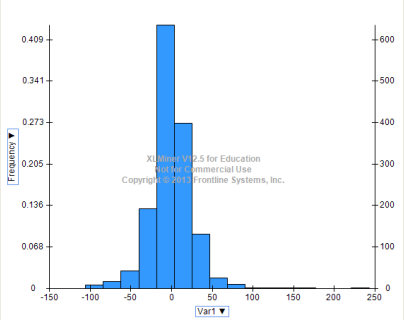
Total sum of squared errors	RMS Error	Average Error
2618203.149	26.72424473	1.31665E-06

Validation Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
1311029.462	24.41707353	-0.12607358

Test Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
962957.7471	25.62929327	-0.53262161



K-Nearest Neighbors (K-NN)

Training Data scoring - Summary Report (for k=10)

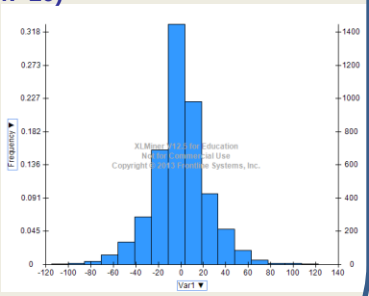
Total sum of squared errors	RMS Error	Average Error
1274032.635	11.28730541	0.001412209

Validation Data scoring - Summary Report (for k=10)

Total sum of squared errors	RMS Error	Average Error
3972175.385	24.54182235	-0.12621997

Test Data scoring - Summary Report (for k=10)

Total sum of squared errors	RMS Error	Average Error
2584652.566	24.24779459	-0.1369072



Evaluation: Both the models show similar results and give a prediction accuracy of 70% as compared to the benchmark figure. We have chosen linear regression model considering the size of data set, which could be comparatively smaller in case we would want to model for specific routes.

Recommendations

yourcabs.com

yourcabs.com | BANGALORE | Phone Bookings + Helpline: (080) 30 600 600 | EMAIL US

Majestic Railway Station → **Koramangala 1st Block**

Thu, 26 December 2013 at 6:00 PM | Modify Search | Drop-off

SMALL 4 | SEDAN 4 | SUV / MUV 6 - 7 | AC CAB | NON AC | MAX Rs 1,300

TATA INDICA

Rs 334 approx.

Rs 13 / km ~~Rs 13.5 / km~~

16 - 24 mins. approx.

Non AC | 4

TATA INDICA

Rs 356 approx.

Rs 14.5 / km ~~Rs 15 / km~~

16 - 24 mins. approx.

AC Cab | 4

REGULAR SEDAN

Rs 392 approx.

Rs 17 / km ~~Rs 17.5 / km~~

14 - 20 mins. approx.

Non AC | 4


Thank You

Appendix

yourcabs.com

Erroneous Data examples

Predicted Value	Actual Value	Residual	dSmallVehicle	dSedan	dWeekday	dMorning	dAfternoon	dEvening	Distance
89.376999	21	-68.376999	0	1	0	0	0	1	24.377842
89.839048	17	-72.839048	1	0	1	0	0	1	32.53963
88.899518	149	60.100482	1	0	0	0	0	1	7.9122705
88.704777	133	44.295223	1	0	1	0	0	1	10.169565
85.57141	139	53.42859	0	1	0	0	1	0	1.5644999
98.182346	149	50.817654	0	0	1	0	0	1	12.797752
85.397725	142	56.602275	0	1	1	0	1	0	4.2370328
87.876512	39	-48.876512	1	0	0	0	1	0	39.976991
87.536647	16	-71.536647	1	0	0	0	1	0	33.274181
65.616117	94	28.383883	0	1	0	0	0	0	2.7318935