

Identifying individual usage patterns for an effective promotion strategy



Team

Sheel Jaitley – 61410271

Sunny Sapra – 61410579

Amit Phatak – 61410158

Biswajit Pattnaik – 61410498

Sourav Paul – 61410391

Business problem

- Business problems analyzed
 - Benchmark future revenue growth
 - Identify key customers to develop more effective promotional campaigns
- Boundary conditions
 - Business segment analyzed – Point to point travel type
 - Time period under analysis – July 2013 to November 2013

Forecasting problem

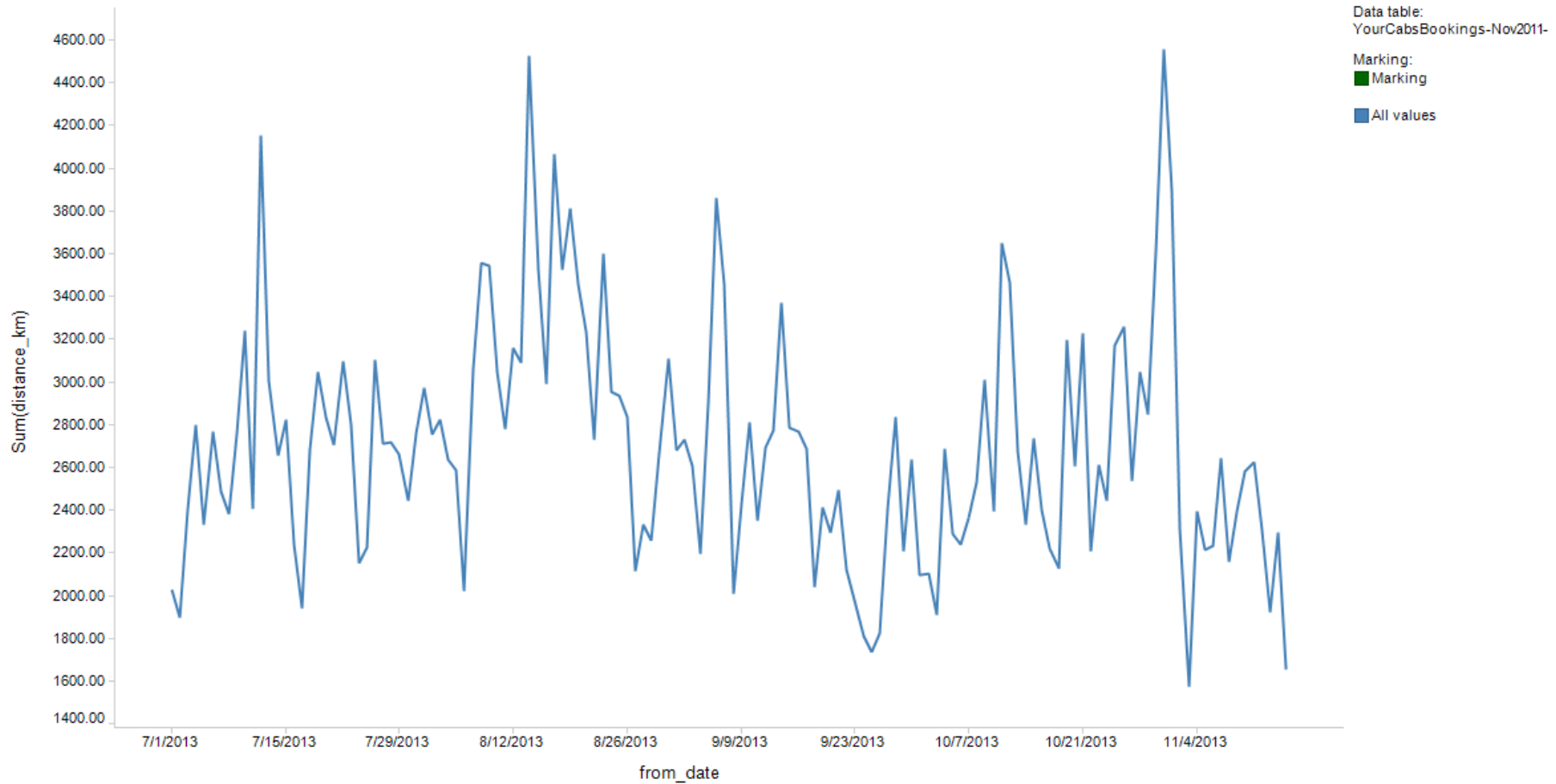
- Forecast usage pattern for each user using logistic regression on a weekly basis
- Each forecast indicates the user's likelihood of booking a cab the next week
- Data Set used
 - Number of weekly booking for each user
 - Latitude and longitude position
 - User Id

Forecasting methodology

- Data visualization
- Data preparation
- Model selection
- Developing forecasting model
- Performance evaluation

Visualizing the Data Set

Line Chart



Data preparation and selecting forecasting model

Data preparation

- Filter data with correct to and from timing
- Remove incorrect data

Model for revenue forecast

- Linear regression

Model for Usage forecast

- Logistic regression

Model for revenue forecast

- Predictors:
 - Lag 1 to lag7 for number of daily booking

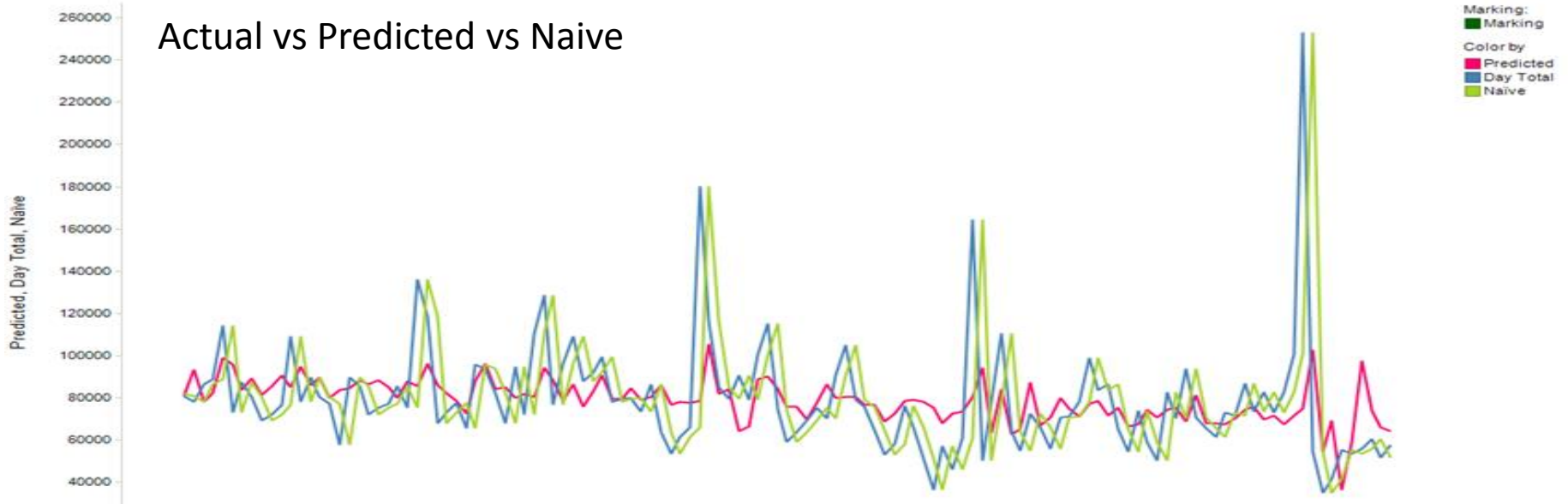
Forecasting model

- $$Y = \beta_0 + \beta_1 * W_{t-1} + \beta_2 * W_{t-2} + \beta_3 * W_{t-3} + \dots + \beta_7 * W_{t-7}$$

Performance Evaluation

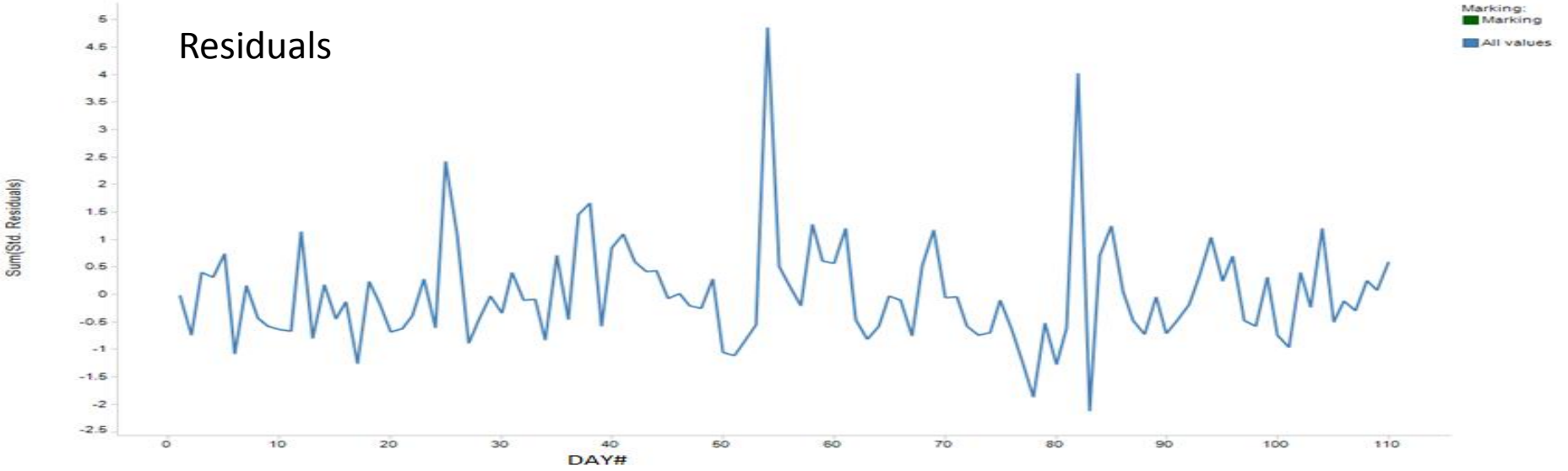
Line Chart

Actual vs Predicted vs Naive



Line Chart

Residuals



Forecasting weekly usage pattern for each user – single user model

- Predictors:
 - Number of booking per user in the week (W)
 - Lag 1 and lag 2 of weekly bookings of the particular user

Forecasting model

- $\text{Logit}(\text{week}=1) = \beta_0 + \beta_1 * W_{t-1} + \beta_2 * W_{t-2}$

Performance evaluation

Output for logistic regression

Classification Confusion Matrix		
	Predicted Class	
Actual Class	1	0
1	1	2
0	2	1

Error Report			
Class	# Cases	# Errors	% Error
1	3	2	66.67
0	3	2	66.67
Overall	6	4	66.67

Output for naïve forecast

Classification Confusion Matrix		
	Predicted Class	
Actual Class	1	0
1	3	0
0	3	0

Error Report			
Class	# Cases	# Errors	% Error
1	3	0	0
0	3	3	100
Overall	6	3	50

Forecasting weekly usage pattern for each user

- single model for multiple users

- Predictors:
 - Number of booking per user in the week (W_n)
 - Lag 1 and lag 2 of weekly bookings per user
- Interaction variable
 - $D_1, D_2 \dots D_n$ for n users

Forecasting model

- $$\text{Logit}(\text{week}=1) = \beta_0 + \beta_1 * D_1 * W_{t-1} + \beta_2 * D_1 * W_{t-2} + \beta_3 * D_2 * W_{t-1} \dots \beta_{2n} * D_n * W_{t-1}$$

Performance evaluation

Output for logistic regression

Classification Confusion Matrix		
	Predicted Class	
Actual Class	1	0
1	11	8
0	0	8

Error Report			
Class	# Cases	# Errors	% Error
1	19	8	42.11
0	8	0	0.00
Overall	27	8	29.63

Output for naïve forecast

Classification Confusion Matrix		
	Predicted Class	
Actual Class	1	0
1	6	4
0	13	4

Error Report			
Class	# Cases	# Errors	% Error
1	10	4	40.00
0	17	13	76.47
Overall	27	17	62.96

Performance evaluation

Line Chart

