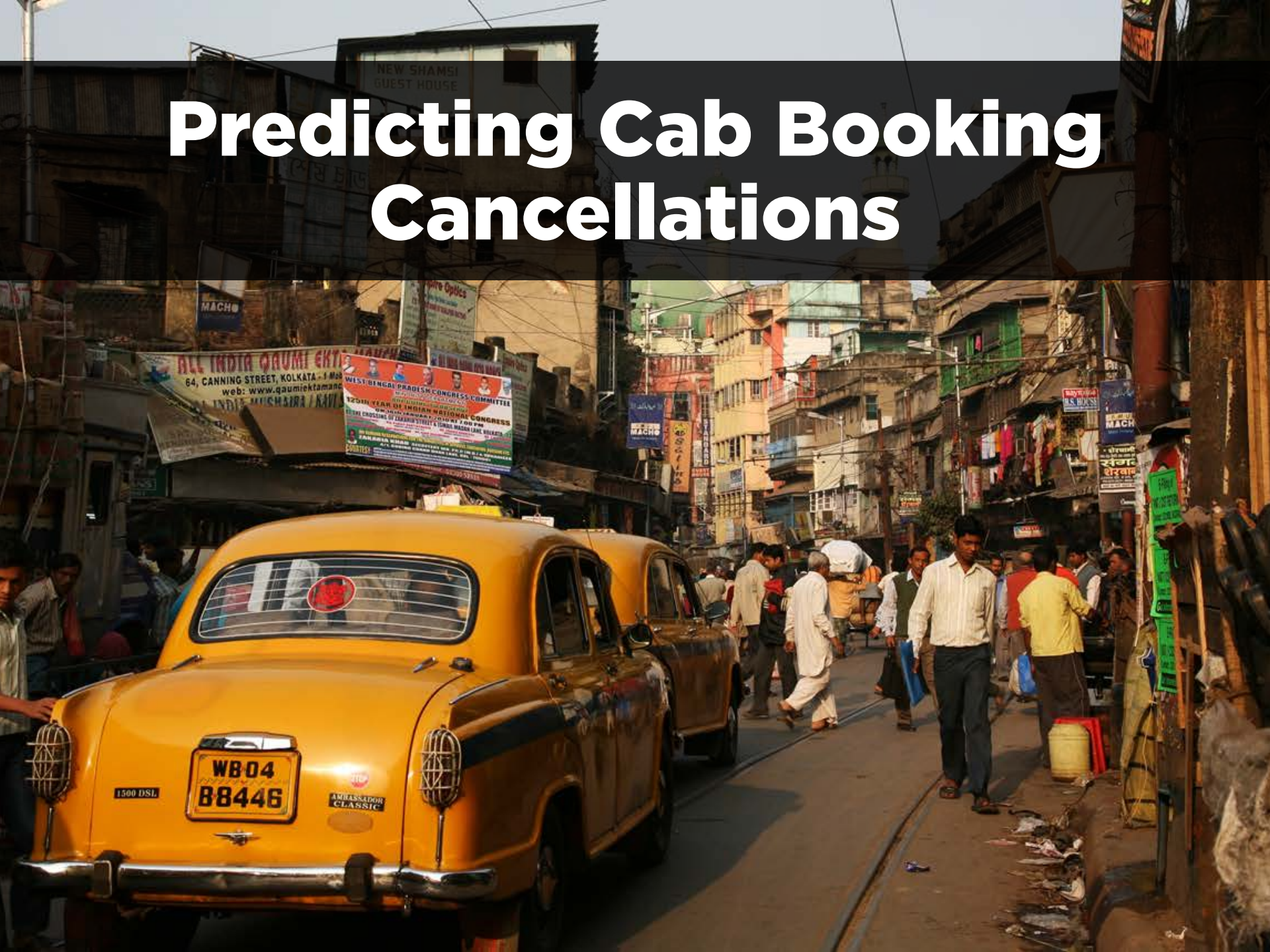
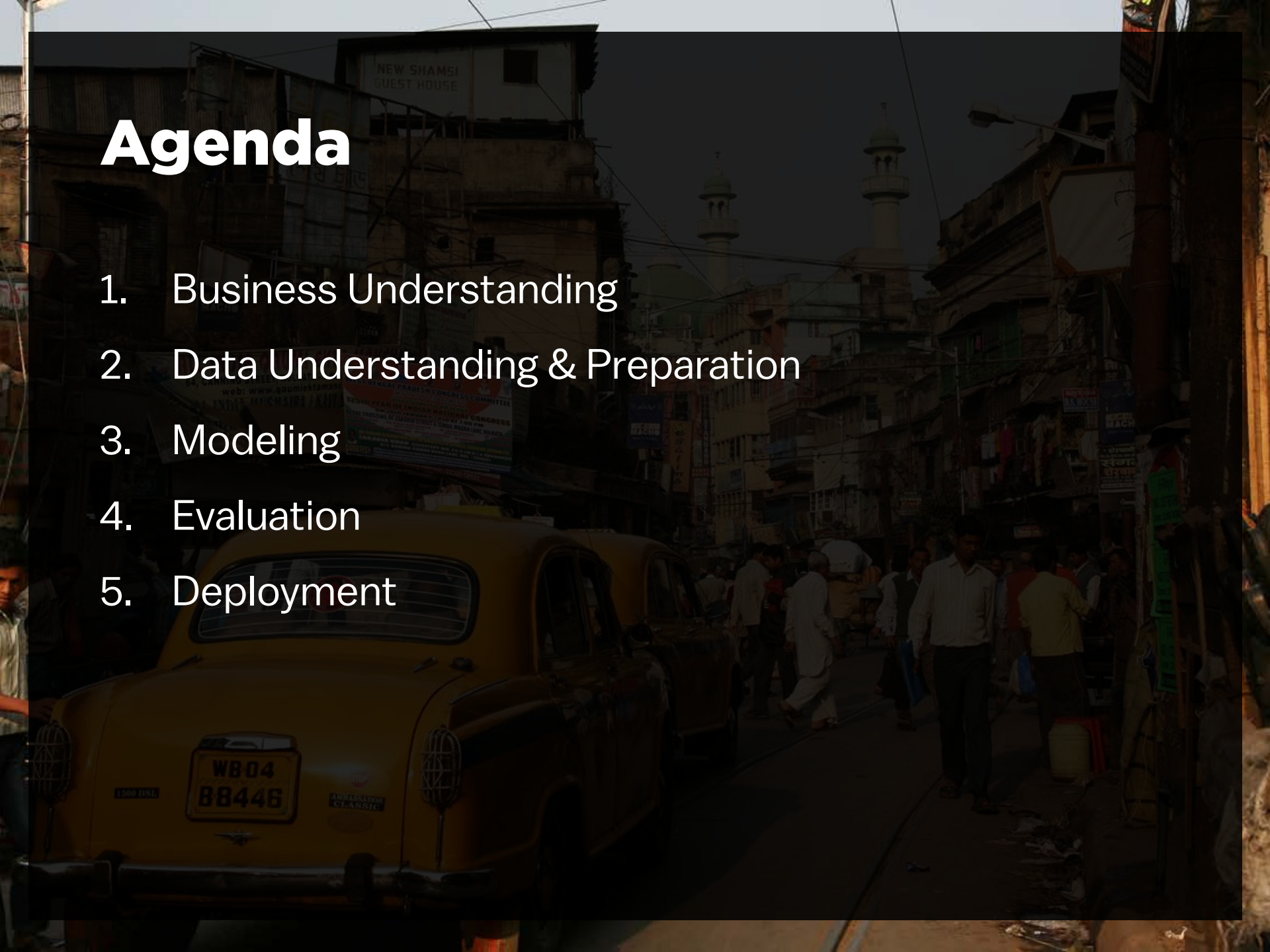
Predicting Cab Booking Cancellations



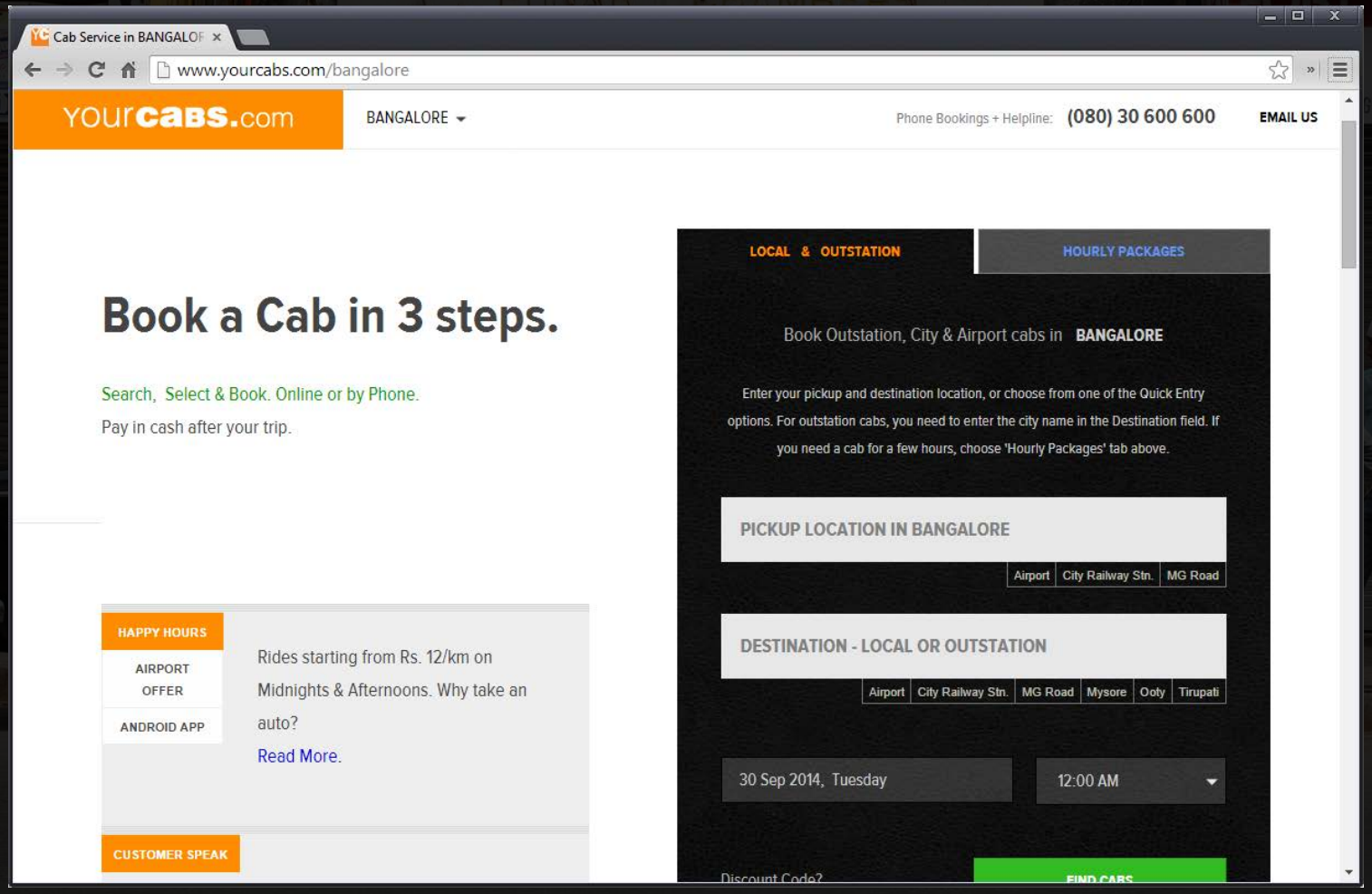


Agenda

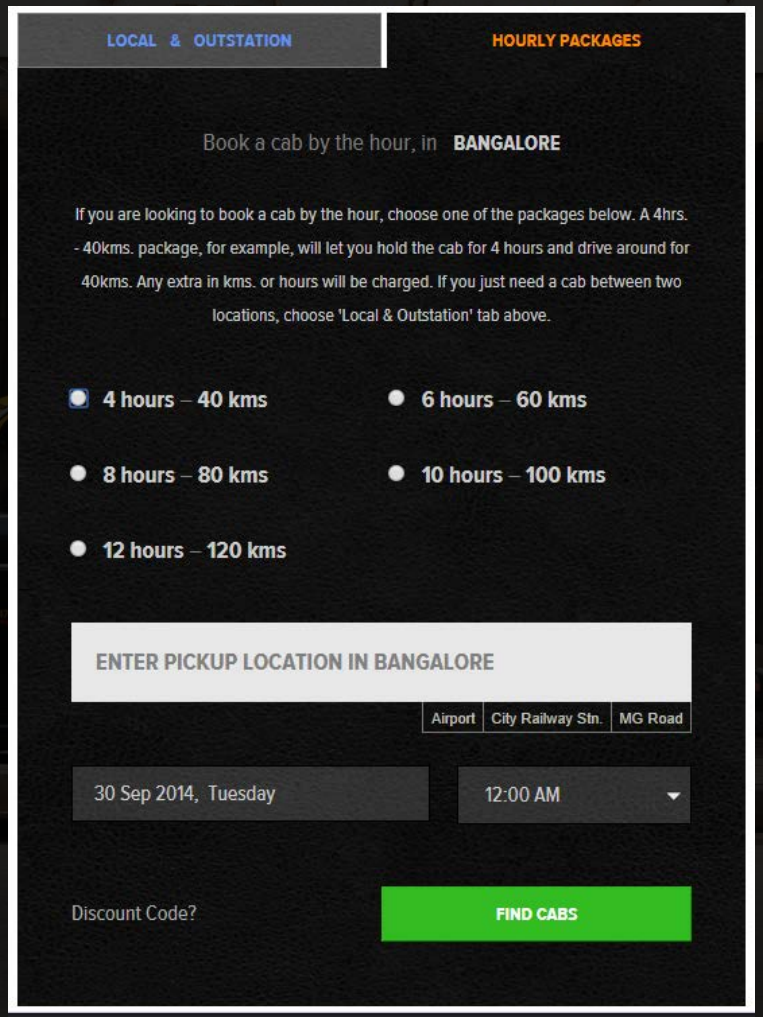
1. Business Understanding
2. Data Understanding & Preparation
3. Modeling
4. Evaluation
5. Deployment

1. Business Understanding

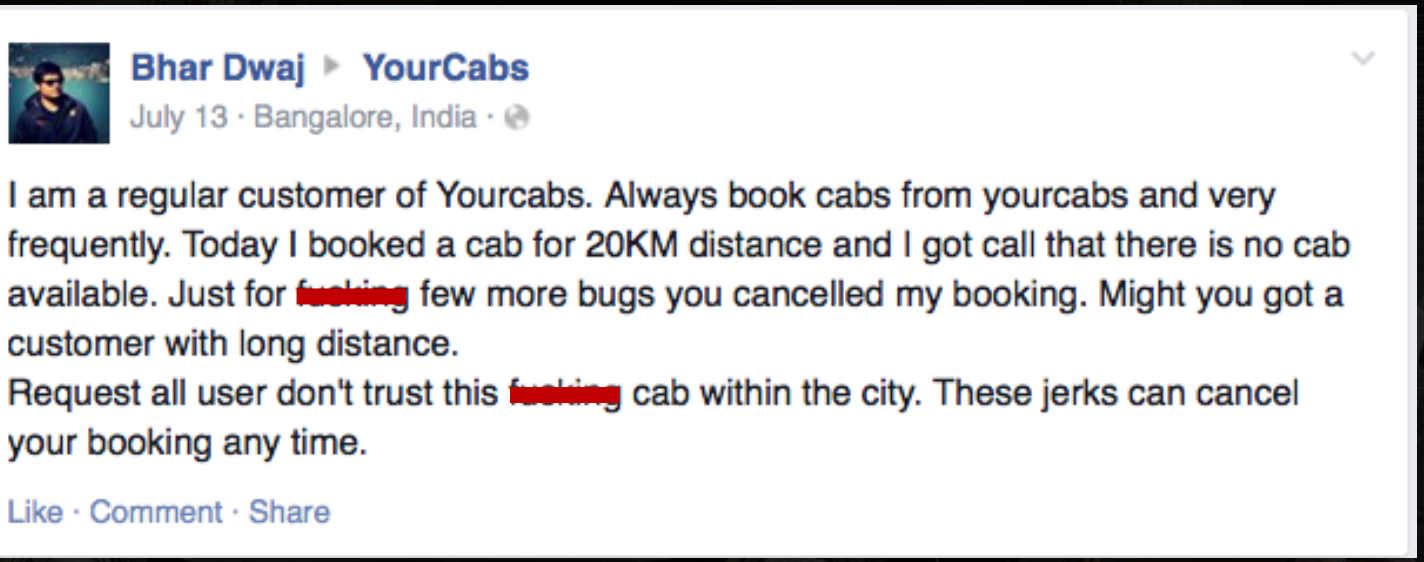
1. Business Understanding



1. Business Understanding



1. Business Understanding



Bhar Dwaj ▸ **YourCabs**
July 13 · Bangalore, India ·

I am a regular customer of Yourcabs. Always book cabs from yourcabs and very frequently. Today I booked a cab for 20KM distance and I got call that there is no cab available. Just for [redacted] few more bugs you cancelled my booking. Might you got a customer with long distance.
Request all user don't trust this [redacted] cab within the city. These jerks can cancel your booking any time.

Like · Comment · Share

2. Data Understanding & Preparation

2. Data Understanding & Preparation

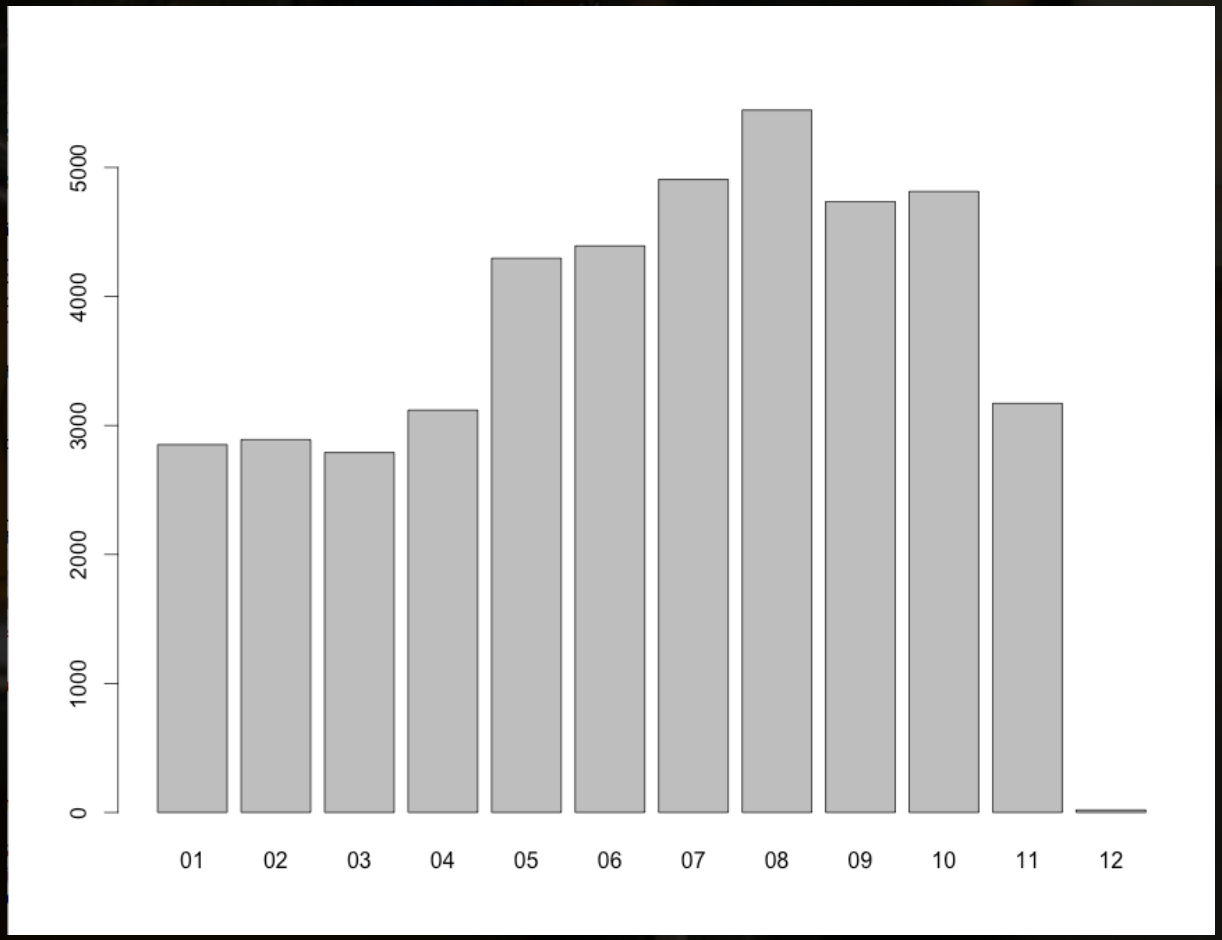
- Kaggle.com
- Contained 43,431 transactions and 20 variables
- Target variable = “Car_Cancellation”
  - If car was cancelled – 1
  - If car not cancelled – 0
- “Null” values & ID’s

2. Data Understanding & Preparation

- Existing Variables:
  - vehicle_model_id
  - package_id
  - travel_type_id
  - from_date
  - online_booking
  - mobile_site_booking
  - booking_created
- New Variables:
  - difftime
  - month_booked_for
  - dayofweek_booked_for
  - hour_booked_for

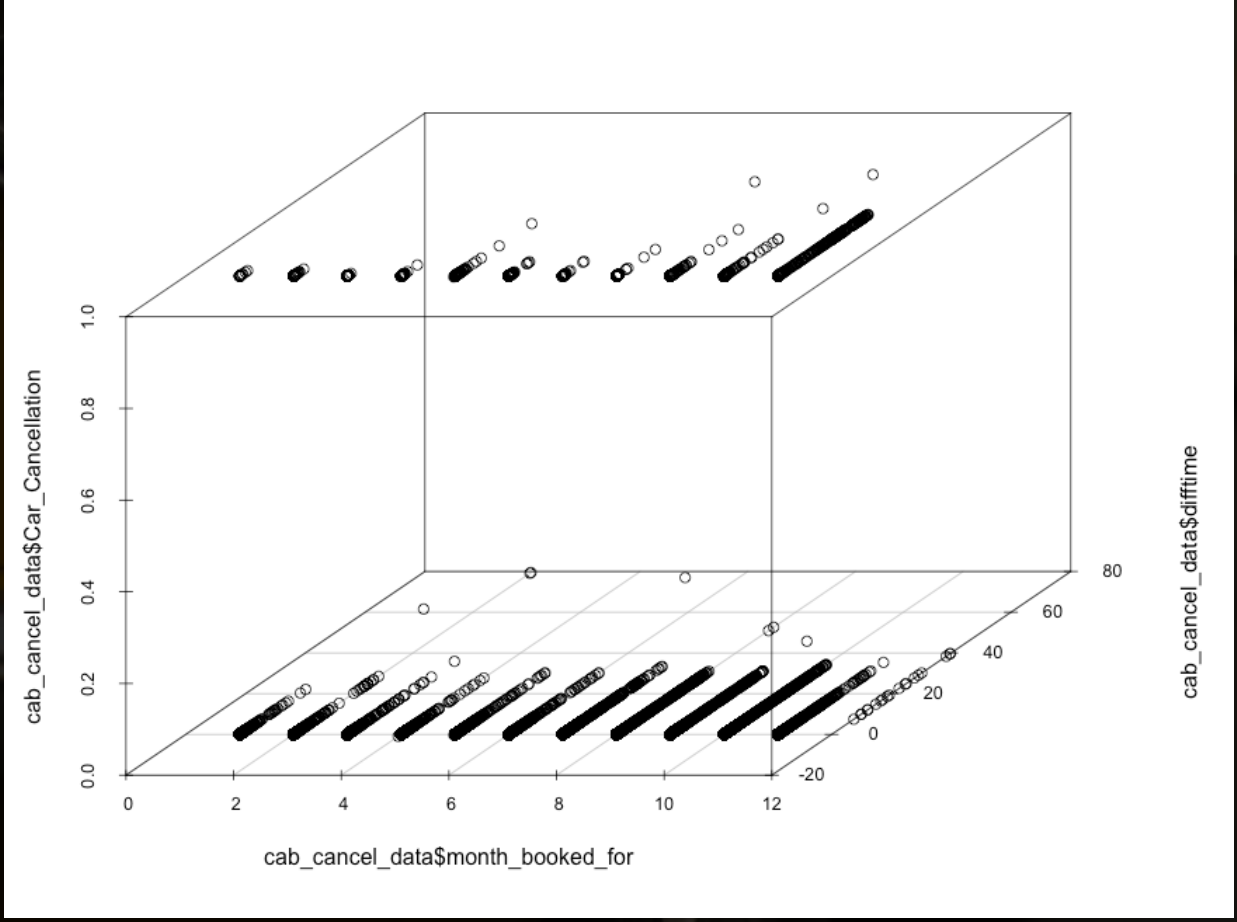
2. Data Understanding & Preparation

Frequency of Months when Bookings are Created For



2. Data Understanding & Preparation

How does DiffTime and Month Booked For Effect Car Cancellation?

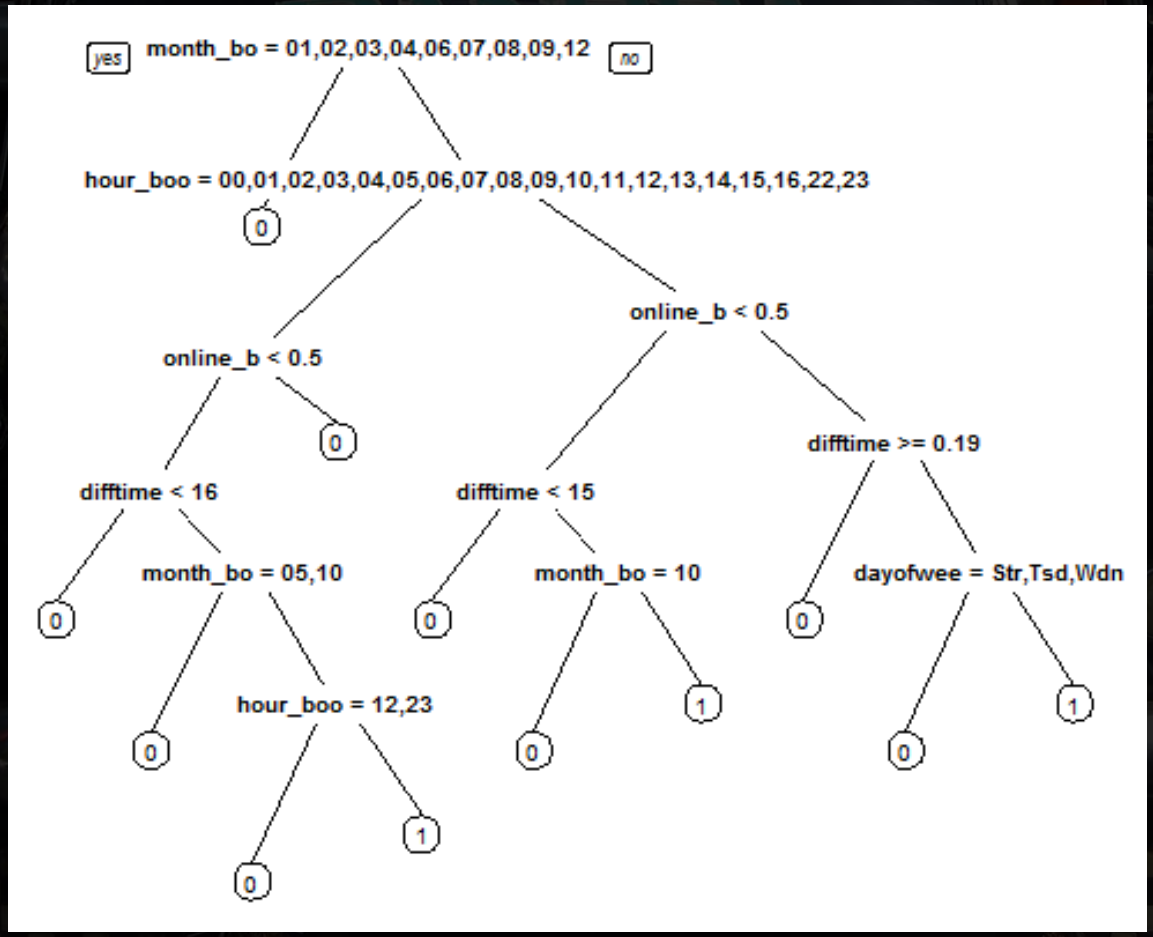



NewMumbai
Thane
Bhayander
BANDRA
MH·01
G·1558
MH.01.G.1558.C.97
DELUXE

3. Modeling

3. Modeling

Decision Tree for Car Cancellation Dataset



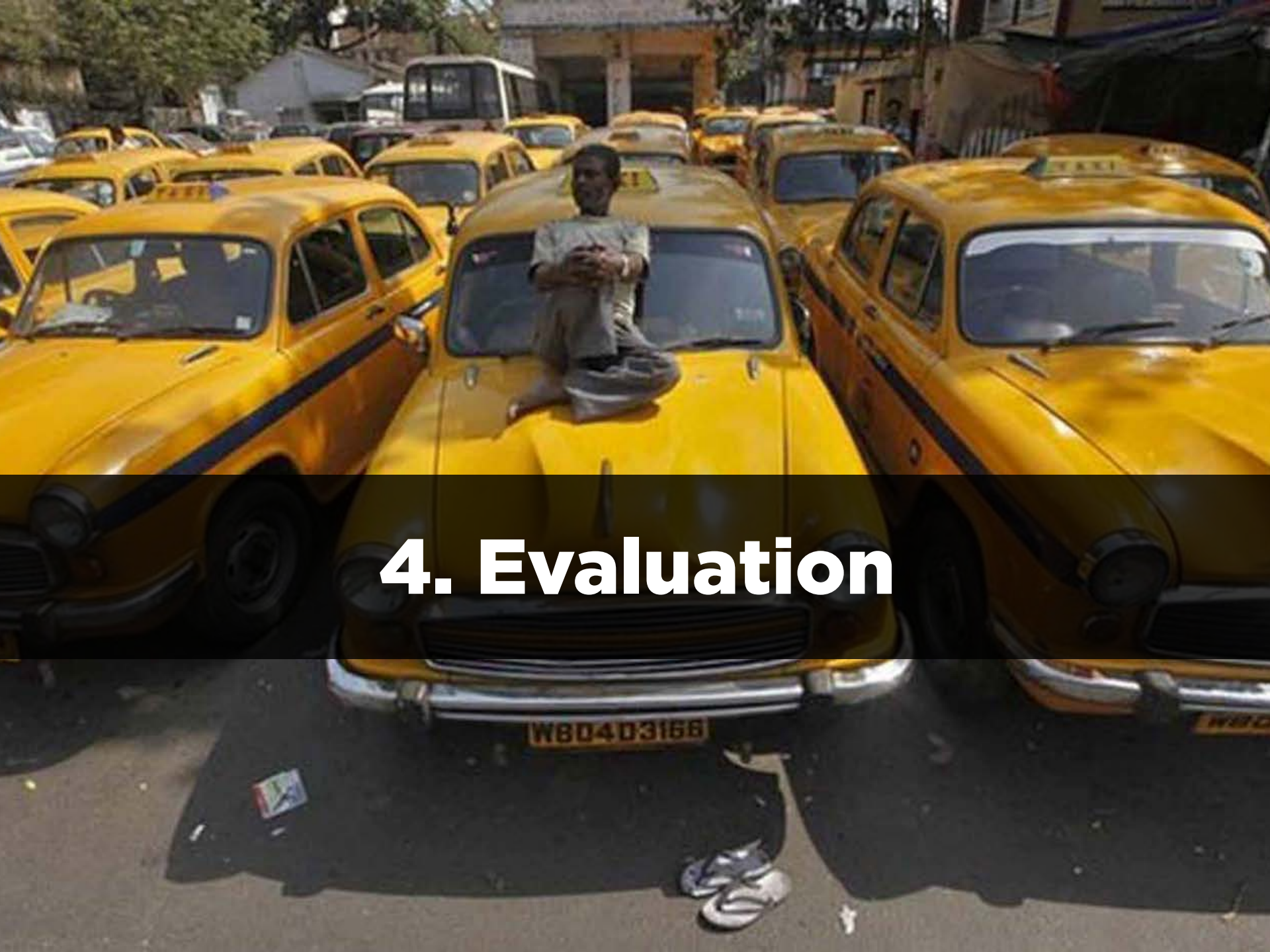
3. Modeling

- Decision Tree for Car Cancellation Dataset

```
> print(training_dt_model)        # model results
n= 43431

node), split, n, loss, yval, (yprob)
      * denotes terminal node

  1) root 43431 3132 0 (0.92788561 0.07211439)
    2) month_booked_for=01,02,03,04,06,07,08,09,12 31150 1368 0
(0.95608347 0.04391653) *
    3) month_booked_for=05,10,11 12281 1764 0 (0.85636349 0.14363651)
      6)
hour_booked_for=00,01,02,03,04,05,06,07,08,09,10,11,12,13,14,15,16,22,23
9523 1085 0 (0.88606532 0.11393468)
       12) online_booking< 0.5 5674  442 0 (0.92210081 0.07789919)
         24) difftime< 15.85868 5446  350 0 (0.93573265 0.06426735) *
         25) difftime>=15.85868 228   92 0 (0.59649123 0.40350877)
```



4. Evaluation

4. Evaluation

- "Cost_of_error" – cost incurred if booking is misclassified
  - Un-cancelled booking, cost = 1
  - Cancelled booking, cost = function of the cancellation time relative to the trip start time
- Cost / Benefit Information:

	Actual		
Predicted		Cancel	No Cancel
	Cancel	8	-1
	No Cancel	0	0

- $p_R(x) > 0.1111$
  - Expected benefit calculation: we should target bookings as long as the estimated probability of cancellation is greater than 11%.

4. Evaluation

```
> table(test$dt_pred_class_error)
FALSE  TRUE
 7270  1416
```

- Accuracy = 7270 / (1416 + 7270) = 83.70%
- Confusion Matrix:

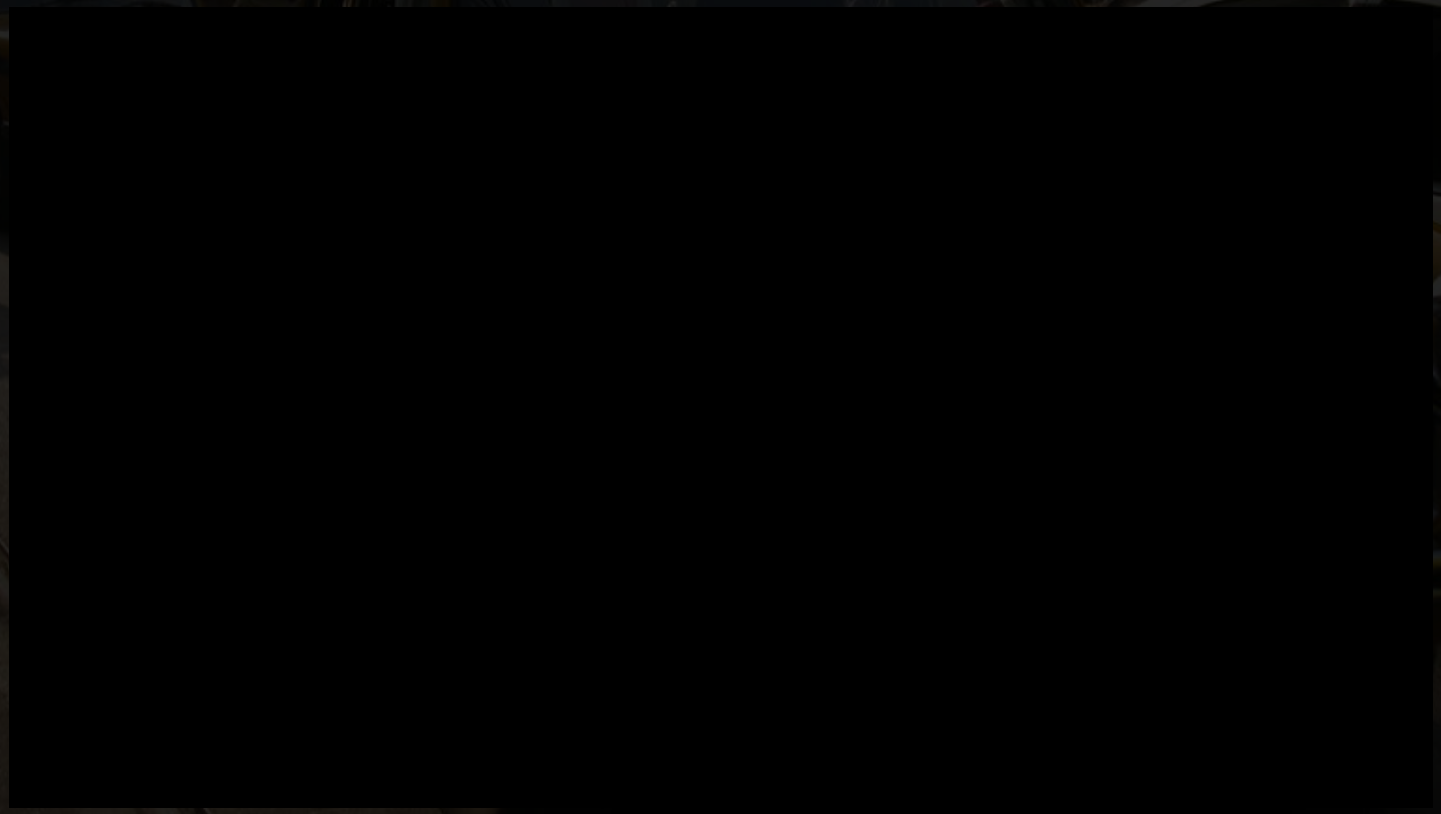
	Actual		
Predicted		0	1
	False	6987	366
	True	1050	283

- Expected Profit = 0.80*8 + 0.04*(-1) = $6.39

5. Deployment

5. Deployment

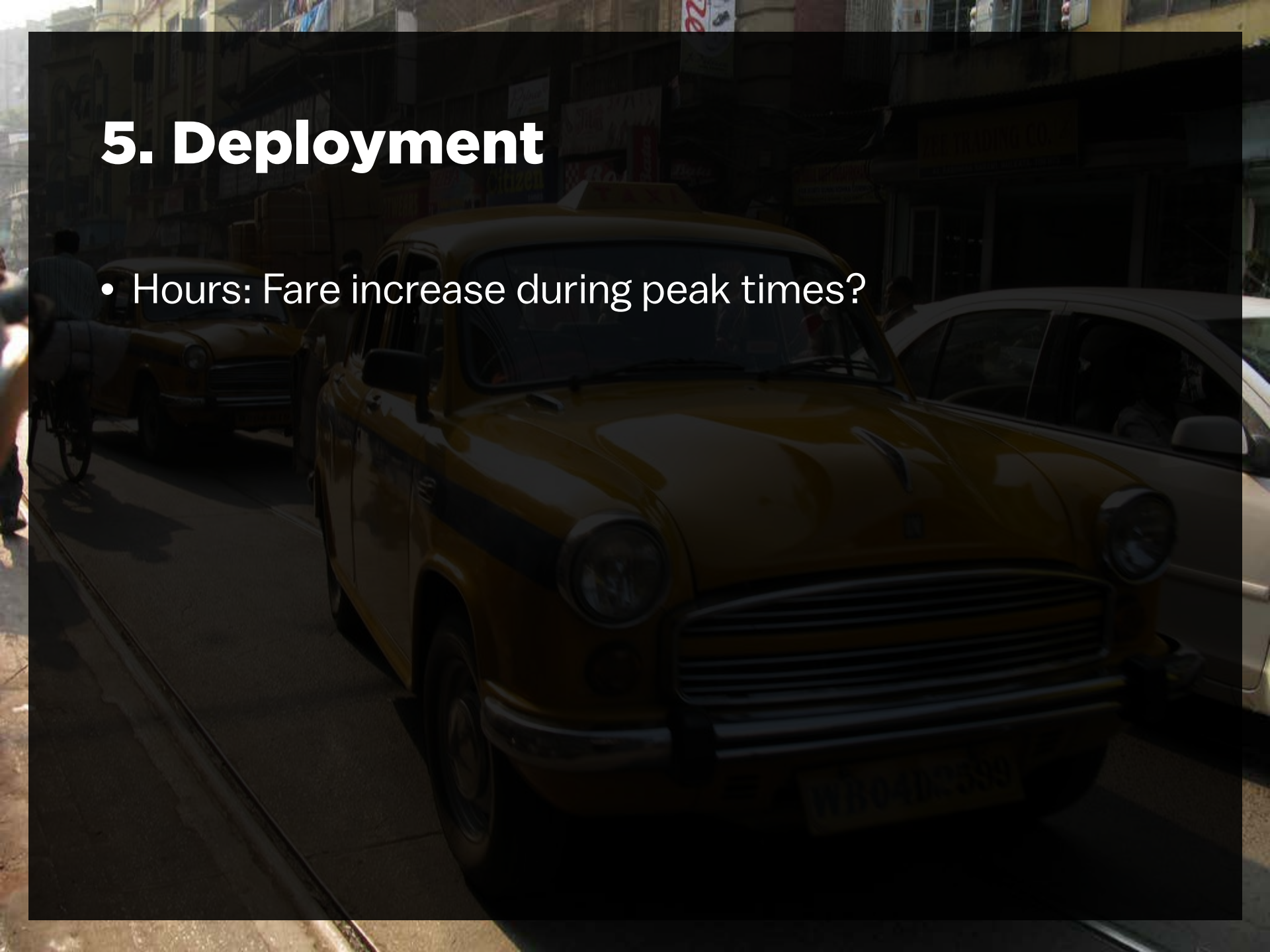
- Months: new traffic patterns needed?

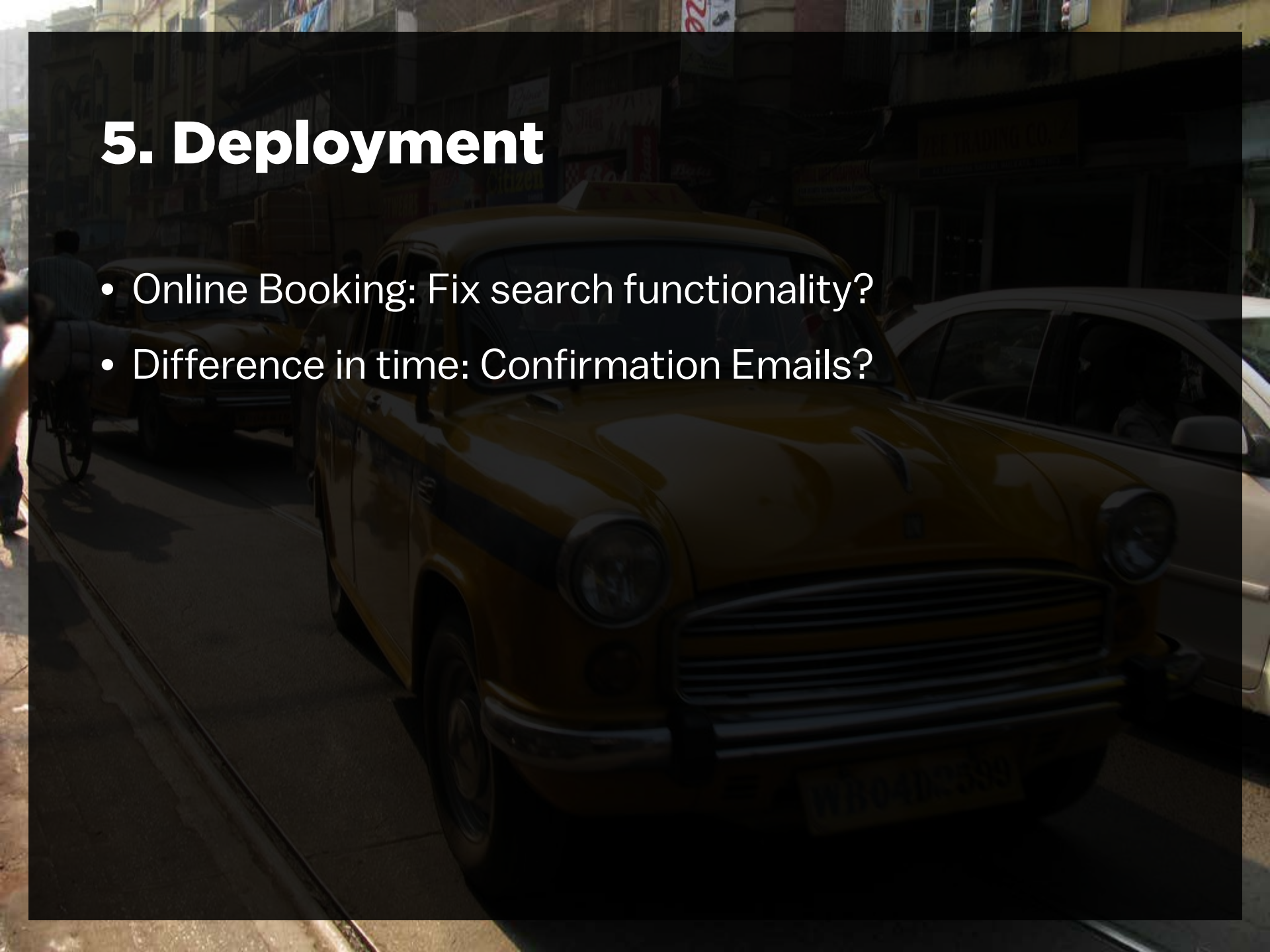


https://www.youtube.com/watch?v=VxhuXHH8-i8

5. Deployment

- Hours: Fare increase during peak times?



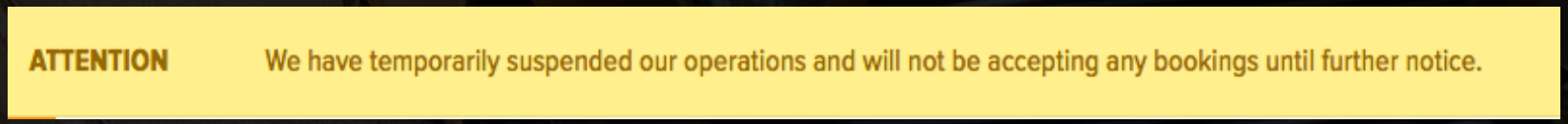


5. Deployment

- Online Booking: Fix search functionality?
- Difference in time: Confirmation Emails?

5. Deployment

- Problems:
  - Manipulating Customer does not address the entire problem
  - Privacy concerns
  - Supply side?



Questions?