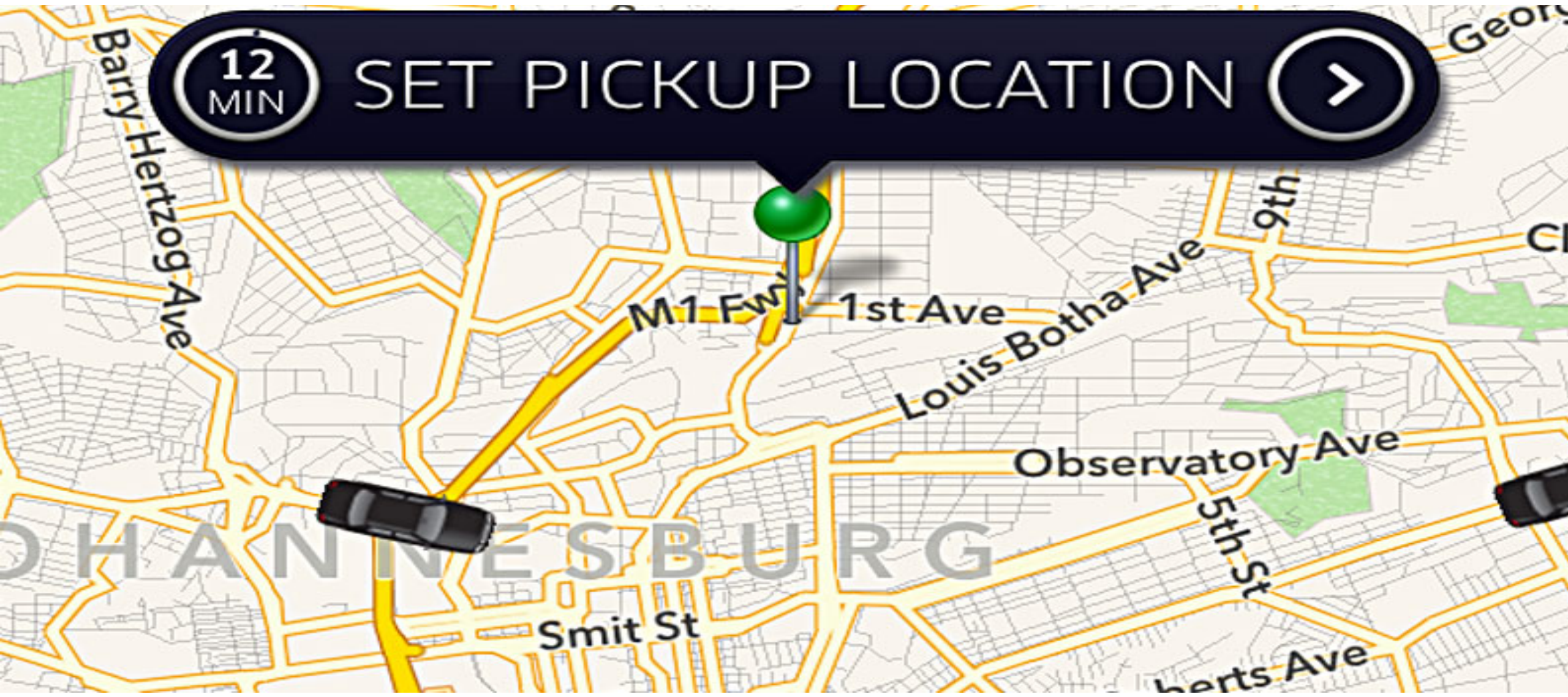


Predicting Cab Booking Cancellations

by Devesh Khandelwal

Problem Statement

Customers can cancel the booking up to the **last minute** of pick up at **no cost** to them

Cancelled booking dents the revenue of the company and adds operational overheads

Use the Data collected over time to predict the probability of booking cancellation

Problem Statement

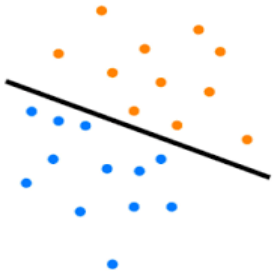
- Customers can cancel the booking up to the **last minute** of pick up at **no cost** to them
- Cancelled booking dents the revenue of the company and adds operational overheads

- Use the Data collected over time to predict the probability of booking cancellation

- The problem is of significant importance to any company in the ride sharing business as:

✓ It could help to streamline the operations
✓ It could help reduce the overheads and costs
✓ It could help the company make the best and most efficient use of its resources

Problem Analysis



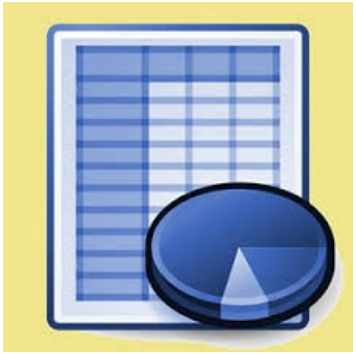
Classification Task – Classify the Cancellation feature into :

- ✓ ‘0’ (Not Cancelled)

  or

- ✓ ‘1’ (Cancelled)

Dataset



Training Data-

- ✓ 43 K records
- ✓ 18 Features

Uneven Classes

- ✓ Approx 7% of the total bookings are actually Cancelled(Training Data)

Source:- https://inclass.kaggle.com/c/predicting-cab-booking-cancellations/data

Features at a Glance

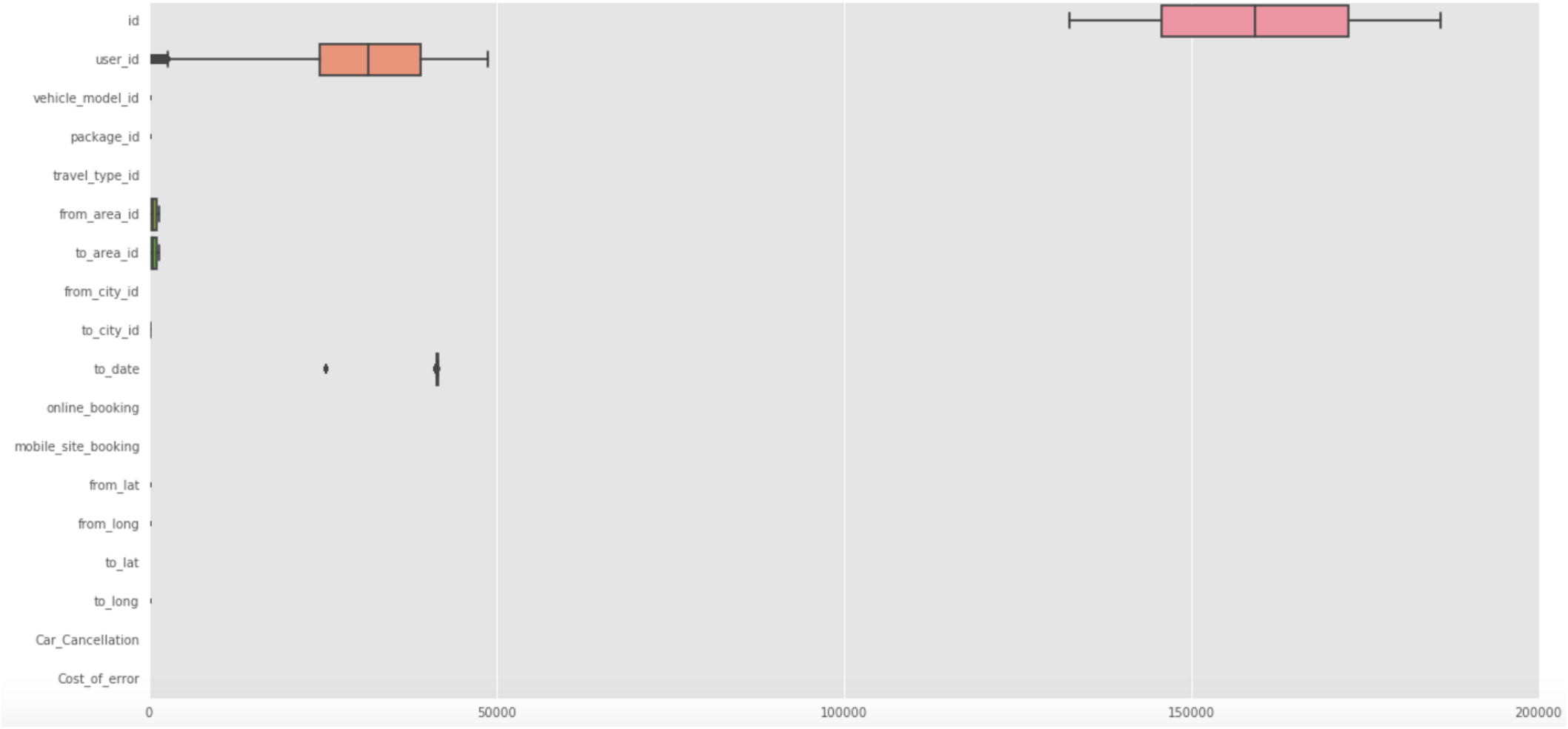
Features set includes:

✓ Vehicle attributes

✓Booking attributes including-

- Online
- GPS data
- Mobile
- Travel Type
- Source
- Destination

Features at a Glance(Contd..)



Hypothesis

The solution could be to make a Machine learning model that could achieve the following:

- ✓ A classification accuracy in the range of 60-90%
- ✓ To explain what are the most important features of the data set that drive this relationship