Analyzing Uber’s Ride-sharing Economy

Farshad Kooti  
USC / Facebook

Mihajlo Grbovic  
Yahoo Research

Luca Maria Aiello  
Bell Labs

Nemanja Djuric  
Yahoo Research

Vladan Radosavljevic  
Yahoo Research

Kristina Lerman  
USC - ISI
Popularity of ride-sharing

Number of riders and drivers is increasing exponentially [Hall and Krueger, 2015].
Related Work

• Impact of sharing economy on traditional markets, in particular Airbnb on hotel industry [Zervas et al’15].

• Survey-based studies:
  • 15% of population use ride-sharing apps [Pew’16].
  • Ease of payment and shorter wait time are main reasons people prefer Uber over taxis [Rayle et al’16].

• Uber drivers are more similar to US workforce than taxi drivers [Hall and Kreuger’15]
Dataset

- All activity on Uber
  - # of rides, length, duration, surge, etc.
- Extracted from confirmation emails
- Oct 2015–May 2016 (8 months)
- 59M rides, 4.1M riders, 222K drivers
- Demographics: age, gender, zip code
Riders
Rider Demographics

• Men and younger riders are more active

• By age:
  • Average age of 34yrs, younger than US population

• By gender:
  • 51% women, 49% men

• By race:
  • White: 80.5% vs 72.4% (US)
  • Hispanic: 8.5% vs 16.3%
  • Afr.-Amr.: 8.2% vs 12.6%
  • Asian: 2.8% vs 4.8%
Surge Pricing

Riders charged surged price are slightly more affluent and younger.

Riders charged surged price are slightly more affluent and younger.
Rider Attrition

- Know rider’s first ride (received welcome email).
- Created monthly activity vector for riders with 4+ months data and 1+ rides (30k riders).
- Clustered using k-means, with k=3 (best choice based on mean square error).

<table>
<thead>
<tr>
<th>Clusters</th>
<th>% riders</th>
<th>Month 1</th>
<th>Month 2</th>
<th>Month 3</th>
<th>Month 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inactive</td>
<td>90.9%</td>
<td>2.1</td>
<td>0.4</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>Low activity</td>
<td>8.0%</td>
<td>8.5</td>
<td>5.8</td>
<td>5.6</td>
<td>5.6</td>
</tr>
<tr>
<td>High activity</td>
<td>1.1%</td>
<td>18.0</td>
<td>21.6</td>
<td>23.3</td>
<td>22.1</td>
</tr>
</tbody>
</table>
Driver Demographics: Not an all-serve-all economy

Compared to riders:

- Drivers are less likely to be female
  - 51% (riders) vs 24% (drivers)
- Drivers are less likely to be white
  - 81% (riders) vs 60% (drivers)
  - 52% (US taxi drivers), 75% (US workforce)
- Drivers are less likely to be affluent
  - median income $62K (riders) vs $53K (drivers)
Men and older drivers work more, on average. Highly correlated with the earnings.

Only 19% of drivers work 40hrs/week or more.
Drivers’ Ratings and Surge

Surge pricing negatively affects ratings.
Drivers vs Riders: Matching

• Match riders and drivers using location, time, and cost of rides and first names.

• Consider drivers with at least 10 rides and 75% of rides matched.

• Age: average age difference of 11.4yrs (±0.47) for above-average ratings and 13.1yrs (±0.56) for below-average ratings.

• Role of gender:

<table>
<thead>
<tr>
<th>Women Drivers</th>
<th>Men drivers</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Women Riders</td>
<td>% Above Avg. Weeks</td>
</tr>
<tr>
<td>0%-45%</td>
<td>62.6%</td>
</tr>
<tr>
<td>45%-55%</td>
<td>53.4%</td>
</tr>
<tr>
<td>55%-100%</td>
<td>50.2%</td>
</tr>
</tbody>
</table>
Driver Attrition

• Know driver’s sign up day (welcome email).
• Cluster based on **monthly hours worked**.
• Drivers have higher retention than riders.

<table>
<thead>
<tr>
<th>Clusters</th>
<th>% drivers</th>
<th>Month 1</th>
<th>Month 2</th>
<th>Month 3</th>
<th>Month 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inactive</td>
<td>73.3%</td>
<td>20.1</td>
<td>4.7</td>
<td>3.0</td>
<td>2.2</td>
</tr>
<tr>
<td>Low activity</td>
<td>21.0%</td>
<td>89.9</td>
<td>45.1</td>
<td>16.4</td>
<td>16.4</td>
</tr>
<tr>
<td>High activity</td>
<td>5.7%</td>
<td>150.3</td>
<td>133.8</td>
<td>126.8</td>
<td>94.1</td>
</tr>
</tbody>
</table>
Riders

Drivers

Activity Prediction
Rider Activity

• Predict whether a rider will remain active, based on early activity.

• Features:
  • Rider features: age, gender, etc.
  • Ride features: # of rides, avg. distance, time of rides, etc.
  • Driver features: age, gender, rating, etc.
  • Social features: Number of contacts who are riders or drivers

• Training on 80% of users and balanced classes.
Rider Activity Results

• Used C5.0 algorithm for classification.
• Used Logistic Regression on independent features for feature ranking.

<table>
<thead>
<tr>
<th>Prediction Results</th>
<th>Top Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>Total # of rides</td>
</tr>
<tr>
<td>Relative Improvement</td>
<td># cities with rides</td>
</tr>
<tr>
<td>Precision</td>
<td>Gender (men)</td>
</tr>
<tr>
<td>Recall</td>
<td>Average fair</td>
</tr>
<tr>
<td></td>
<td>Age</td>
</tr>
</tbody>
</table>

• Used C5.0 algorithm for classification.
• Used Logistic Regression on independent features for feature ranking.
Driver Activity Prediction

• Use first two weeks to predict if the driver would become active driver in weeks 3-8.
• Similar features and approach as riders

<table>
<thead>
<tr>
<th>Prediction Results</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>83.1%</td>
</tr>
<tr>
<td>Relative Improvement</td>
<td>66.2%</td>
</tr>
<tr>
<td>Precision</td>
<td>0.775</td>
</tr>
<tr>
<td>Recall</td>
<td>0.689</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Top Features</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (men)</td>
<td>0.371***</td>
</tr>
<tr>
<td>Hours drove</td>
<td>0.157***</td>
</tr>
<tr>
<td>Age</td>
<td>0.037***</td>
</tr>
<tr>
<td>Earning rate</td>
<td>0.029***</td>
</tr>
<tr>
<td>Acceptance rate</td>
<td>-0.015*</td>
</tr>
</tbody>
</table>
Implications and Summary

• Overall understanding of the fast-growing ecosystem
  • Not an “all-serve-all” economy

• Age, gender, and (maybe) racial bias in ratings
  • People’s awareness
  • Improving the rider-driver matching

• Predictions and analyses help us find users more likely to stop using Uber
  • Promotions
  • Changing the experience
Thanks
Drivers’ Ratings

Younger drivers get better ratings compared to older drivers.
Ride Dynamics

Clear daily and weekly patterns.
Summary

• Uber is not an “all-serve-all” market.
• Rider and driver attrition is very high, but the influx of new-comers leads to an overall growth in the number of rides.
• Matching affects rating results.
• Drivers are “punished” for surge rides with lower ratings.
• Based on a rider’s or driver’s initial activity, we can better predict whether she or he will remain active or quit Uber.
Driver Demographics: Not an all-serve-all economy

Compared to riders:

- Drivers are less likely to be female
  - US: 24.0%, Malaysia: 10.1%, Singapore: 9.9%, Canada: 9.4%, …, UK: 4.3% are women

- Drivers are less likely to be white
  - 81% (riders) vs 60% (drivers)

- Drivers are less likely to be affluent
  - median income $62K (riders) vs $53K (drivers)

<table>
<thead>
<tr>
<th>Race</th>
<th>% Drivers</th>
<th>US Taxi Drivers</th>
<th>US Workforce</th>
<th>% Women</th>
<th>Avg. Age</th>
<th>Avg. hrs Worked / week</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>60.0%</td>
<td>52.3%</td>
<td>75.2%</td>
<td>21.9%</td>
<td>41.9yrs</td>
<td>15.4hrs</td>
</tr>
<tr>
<td>Afr.-Amr.</td>
<td>21.6%</td>
<td>25.2%</td>
<td>11.6%</td>
<td>36.5%</td>
<td>40.8yrs</td>
<td>14.8hrs</td>
</tr>
<tr>
<td>Hispanic</td>
<td>13.7%</td>
<td>10.3%</td>
<td>7.6%</td>
<td>23.9%</td>
<td>38.5yrs</td>
<td>15.2hrs</td>
</tr>
<tr>
<td>Asian-Amr.</td>
<td>4.7%</td>
<td>12.2%</td>
<td>5.6%</td>
<td>16.4%</td>
<td>41.6yrs</td>
<td>18.2hrs</td>
</tr>
</tbody>
</table>
Driver Demographics: Not an all-serve-all economy

Compared to riders:

- Drivers are less likely to be female
  - 51% (riders) vs 24% (drivers)

- Drivers are less likely to be affluent
  - median income $62K (riders) vs $53K (drivers)

- Drivers are less likely to be white
  - 81% (riders) vs 60% (drivers)
  - 52% (US taxi drivers), 75% (US workforce)

<table>
<thead>
<tr>
<th>Race</th>
<th>% Drivers</th>
<th>US Taxi Drivers</th>
<th>US Workforce</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>60.0%</td>
<td>52.3%</td>
<td>75.2%</td>
</tr>
<tr>
<td>Afr.-Amr.</td>
<td>21.6%</td>
<td>25.2%</td>
<td>11.6%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>13.7%</td>
<td>10.3%</td>
<td>7.6%</td>
</tr>
<tr>
<td>Asian-Amr.</td>
<td>4.7%</td>
<td>12.2%</td>
<td>5.6%</td>
</tr>
</tbody>
</table>
Car Type

People with income of $100k are 84% relatively more likely to take an Uber Black compared to users with annual income of $50k.
Drivers