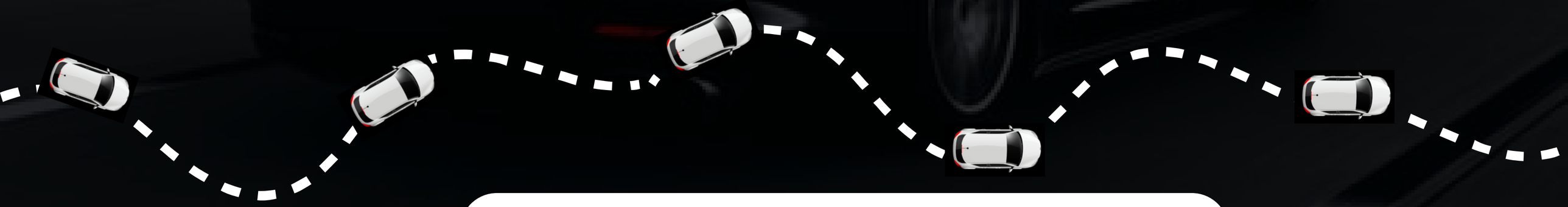
StrataScratch

Uber - Will New Drivers Start Driving?



Team: Kento Morita, Ariel Liang, Phuong Duong, Vincent Perez

Goals

Predict whether or not a driver signup will start driving

Help Uber leverage the insights gained from the model to generate more first trips

The Result Could Generate:

- Driver supply
- Trip availability
- Revenue and service quality

Problem

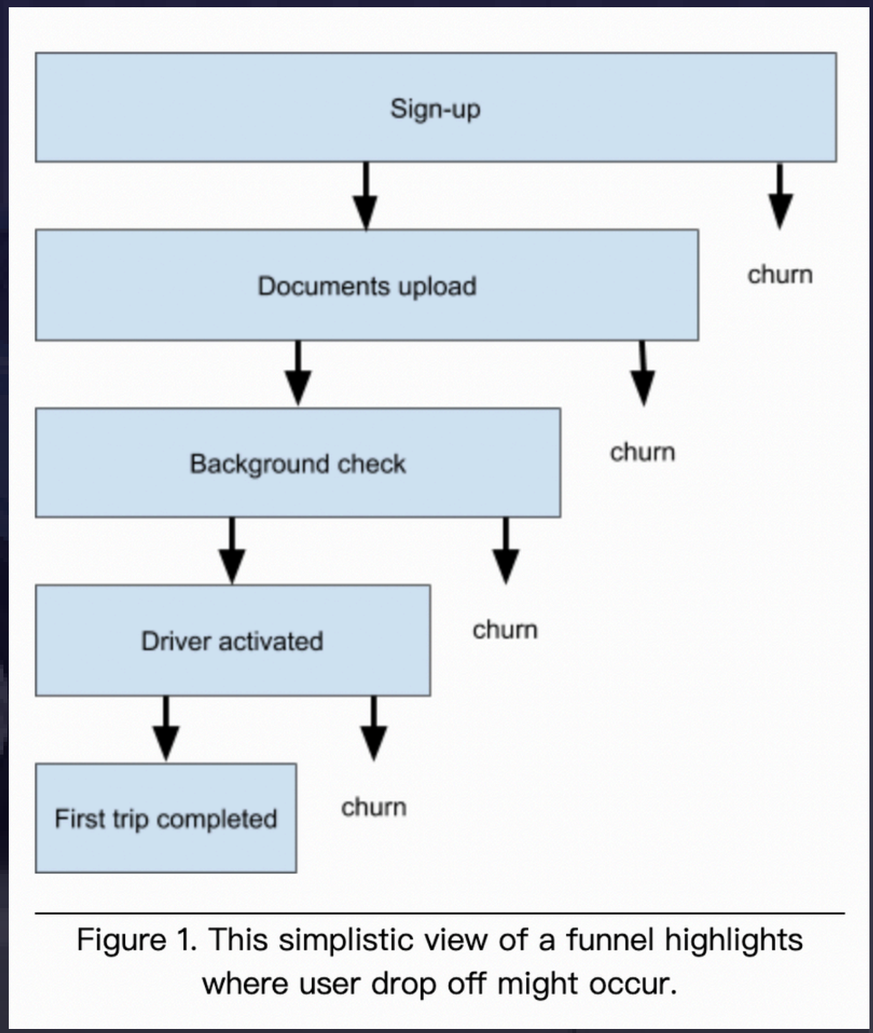


Figure 1. This simplistic view of a funnel highlights where user drop off might occur.

Original Data

Y

first_completed_date
date

X

id	city_name	signup_os	signup_channel	signup_date
nominal	categorical	categorical	categorical	date
bgc_date	**vehicle_added_date**	**vehicle_make**	**vehicle_model**	**vehicle_year**
date	date	categorical	categorical	numerical

Derived Data

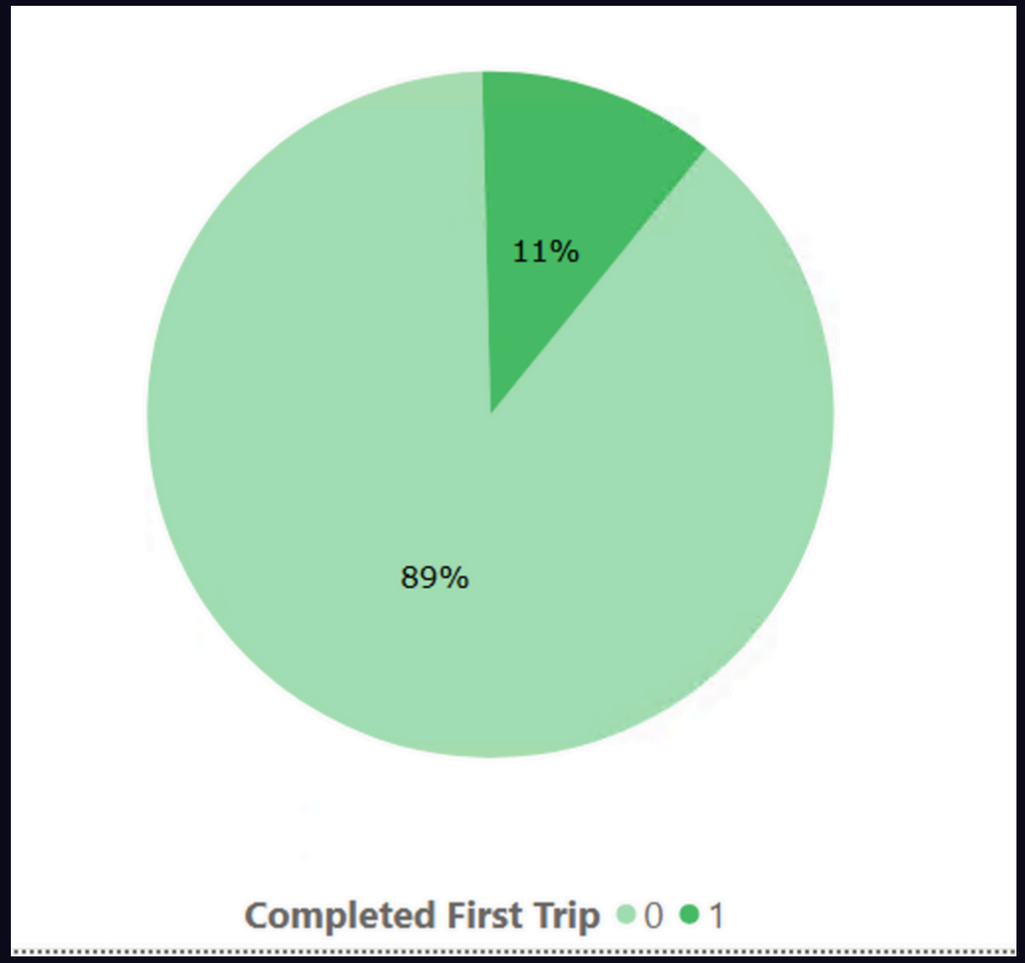
replace **Y** with

started_driving
binary
(0 = no first completed date 1 = have first completed date)

add to **X**

days_to_bgc	days_to_vehicle_add
numeric	**numeric**
(= bgc_date - signup_date)	(= vehicle_added_date - bgc_date)

EDA



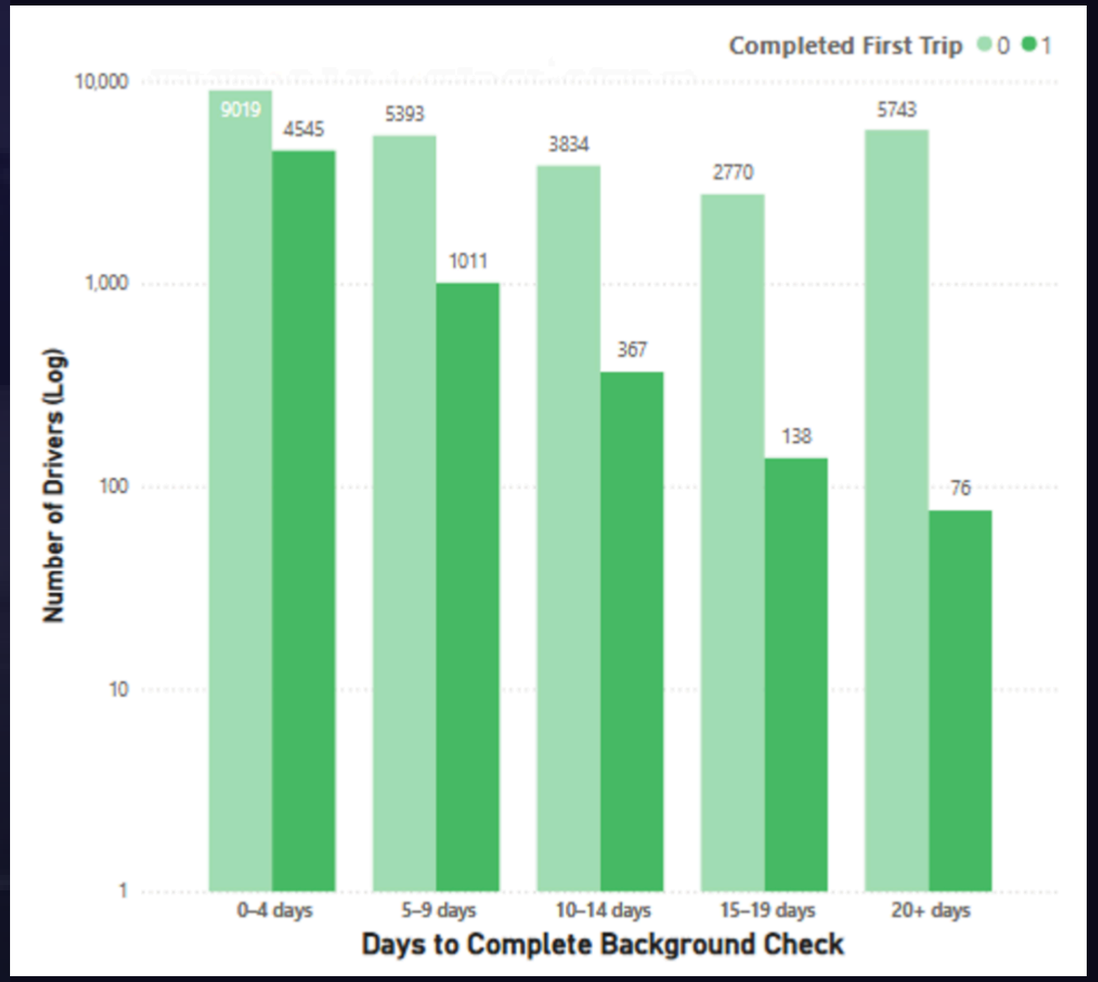
First Trip Completion Rate

Total Sign-ups: 54k+

11% of sign-ups completed their first trip

89% Dropped did not take their first trip

EDA



Faster Backgroud Checks, More Trips?

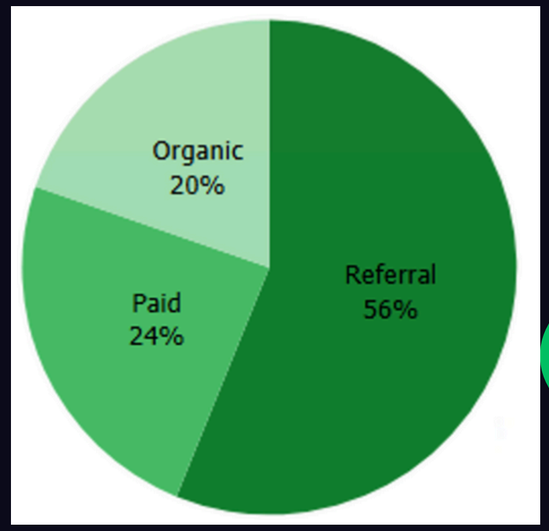
For **drivers who didn't complete first trip**, the distribution is more evenly spread across the background check time range

For **drivers who took first trip,** the distribution is skewed toward faster background checks

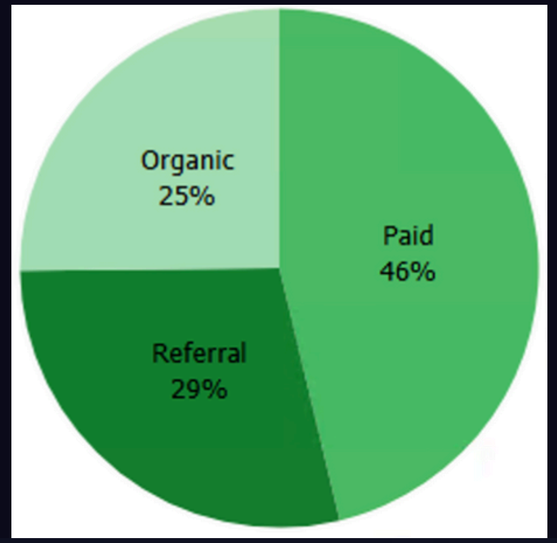
Majority of first trips occur when background checks finish within **10 days**

EDA

DRIVERS WHO TOOK FIRST TRIP



DRIVERS WHO DIDN'T TAKE FIRST TRIP



Which Sign-up Channels Drive Action?

- Half of drivers who took their first trip (**56%**) signed up through the **"Referral"** channel
- Among inactive drivers, referral only makes up a small portion (29%)
- Referral sign-ups are significantly convert better than other channels

Model Comparison

Methodology

- Data Imputation
- Oversampling for Imbalance
- Hyperparameter tuning
- Cross Validation

Varriables

- X => days_to_bgc, days_to_vehicle_add, city_name, signup_os, signup_channel
- y => started_driving

Logistic Regression

Accuracy : 74%

- Struggles due to non-linear relationship
- Good baseline, but unsuited for this model

Moderate

Gradient Boosting

Accuracy : 94%

- Top-performing model
- Strong for weak prediction label
- Capture complex interactions

BEST

Random Forest

Accuracy : 92%

- Robust & interpretable
- Resistant to overfitting
- High balanced tree model

Solid

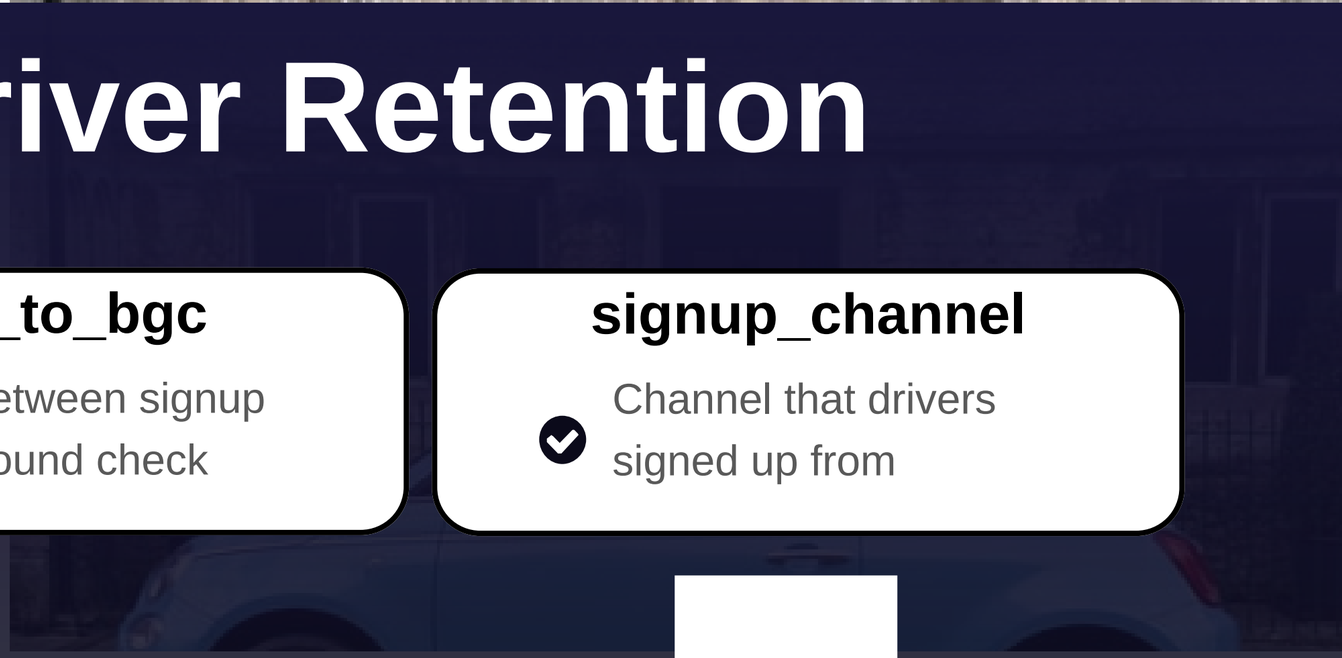
Neural Network

Accuracy : 94%

- Deep learning ready
- Flexible architecture adaptable
- Tuned to avoid overfitting

BEST

Key Features for Driver Retention



days_to_vehicle_added

- The days between signup and vehicle registration

days_to_bgc

- The days between signup and background check

signup_channel

- Channel that drivers signed up from

Insights

The **longer** it takes to complete both of the processes, the **less likely** the drivers **complete first drive**

Referral signup significantly boosts driver's retention rate

Business Strategy Suggestions

Background Check

- Optimize BCG vendors/processes: Ensure the fastest turnaround possible
- Transparent tracking: Give drivers real-time updates on BGC to reduce drop-odd

Speed of Vehicle Addition

- Create a Nudge System: Send reminders or provide guided steps to complete vehicle onboarding quickly.
- Incentivize Fast Action: Offer a small bonus or priority if the vehicle is added within the first 5 days

Referral Channel SignUp

- Double down on referrals: promote referral programs more aggressively with better bonuses or tiered rewards
- Promote the referrals using mobile devices

Uber

Thank You!
Q&A