

Ford GoBike Business Analysis (MIDS w205 Query Project)

August 3, 2018

0.1 Purpose

The Data Science team at Ford GoBike has been analyzing the numbers from San Francisco bike share use. During our work, we discovered an opportunity for increased revenue. Currently, our packages and offerings do not adequately address our customers' use models. Based on our assessment, we believe that visitors to SF are likely frustrated by GoBike's per-trip time limits. On the reverse side, the data indicates that SF locals and commuters are given a per-trip time allocation *well* above their desired amount. This report presents an overview of the our markets and customers' use models, analyzes the fit of our current packages, and finally offers suggestions to management to better match our customers' desired bikeshare usage.

```
In [69]: import pandas as pd
import numpy as np
%matplotlib inline
```

0.2 Data Source

Data from Google Big Query's public dataset "san_francisco.bikeshare_trips" was used throughout this analysis. This data set contains 983,648 unique trips and gives information on the start & end stations, start & end times, trip duration, customer home location, and customer payment plan type. The full definition of the data set scheme can be found on the Google Big Query website.

For this report, we define "commuting hours" as 0600-1000 (6:00am to 10:00am) and 1600-1900 (4:00pm to 7:00 pm). To be considered a commuter trip, a trips must start AND end within one of those time windows and the start and end stations must be different. Additionally, we assume that commuters that use GoBike as part of their commute do so regularly, and therefore purchase an annual pass, as opposed to a 1-day or 3-day bike share pass. San Francisco locals are defined by a billing zipcode starting with "94xxx".

0.3 Characterizing our Users

When we break down the data set by subscriber type, we see that only 14% of GoBike customers make a one-time purchase (for a 1- or 3-day pass). Of 983,648 total trips, 846,839 (an 86% vast majority) of bike trips are taken by annual-pass holders ("subscriber" pass type). We can further restrict the search to SF local subscribers and see that 77% of bike trips are taken by *local* annual-pass holders. It is critical for management to understand that our strongest customer base is made up of local, regular bike-share users who hold annual passes.

```
In [70]: ! bq query --use_legacy_sql=FALSE 'SELECT count(distinct trip_id) AS Purchases, subscriber_type AS LocalSubscriberPurchases'
! bq query --use_legacy_sql=FALSE 'SELECT count(distinct trip_id) AS LocalSubscriberPurchases'
```

Waiting on bqjob_r41ec586ffd295751_00000161a61cd06d_1 ... (0s) Current status: DONE

```
+-----+-----+
| Purchases | PassType |
+-----+-----+
|      846839 | Subscriber |
|      136809 | Customer   |
+-----+-----+
```

Waiting on bqjob_r3e605913d943ea2_00000161a61cdfcc_1 ... (0s) Current status: DONE

```
+-----+
| LocalSubscriberPurchases |
+-----+
|                          762980 |
+-----+
```

Below, we categorize these local annual-pass holders further by pulling out commuters. As stated above, we characterize “commuter trips” as one-way trips that occur during morning rush hour (6:00am-10:00am) or evening rush hour (4:00pm-7:00pm). Given these criteria, there are 410,537 trips that are defined as “commuter trips”. This represents 42% of ALL GoBike trips!

The set of commuter trips are saved in a data table “CommuterTrips” for easy reference¹ and the five most popular commuter trips are shown. As one might expect, they are focused at stations for other transportation modes – for example, the Ferry Building or the Caltrain Station.

```
In [71]: ! bq query --use_legacy_sql=FALSE 'SELECT count(trip_id) AS freq, CONCAT(start_station_id, end_station_id) AS TripPath'
```

Waiting on bqjob_r6985cf5b4a4b16a5_00000161a61cee94_1 ... (0s) Current status: DONE

```
+-----+-----+
| freq | TripPath |
+-----+-----+
| 4760 | Harry Bridges Plaza (Ferry Building) --T0-- 2nd at Townsend |
| 4225 | Steuart at Market --T0-- 2nd at Townsend |
| 4072 | San Francisco Caltrain 2 (330 Townsend) --T0-- Townsend at 7th |
| 3570 | San Francisco Caltrain (Townsend at 4th) --T0-- Harry Bridges Plaza (Ferry Building) |
| 3545 | Embarcadero at Sansome --T0-- Steuart at Market |
+-----+-----+
```

¹Due to the limitations of the BigQuery CLI, this subsetting operation was performed in the web interface using the following query:

```
#standardSQL SELECT * FROM bigquery-public-data.san_francisco.bikeshare_trips
WHERE ((EXTRACT(HOUR FROM start_date) > 6 AND EXTRACT(HOUR FROM end_date) < 10)
OR (EXTRACT(HOUR FROM start_date) > 16 AND EXTRACT(HOUR FROM end_date) < 19))
AND (zip_code LIKE "94%")
AND (subscriber_type="Subscriber")
AND (start_station_id != end_station_id)
```

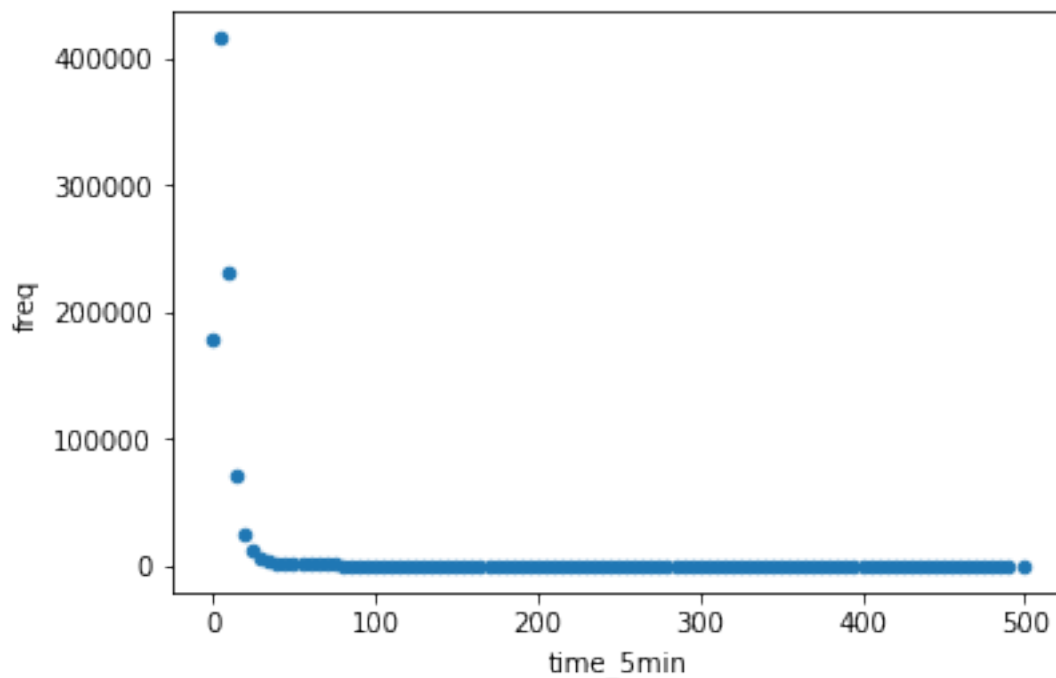
Finally, we present a picture of the true GoBike use model by plotting trip durations. As shown below, the average GoBike trip is actually quite short! The frequency of trips of each duration decreases exponentially.²

```
In [72]: ! bq query --use_legacy_sql=FALSE --format=csv 'SELECT count(floor((duration_sec/60)/5))
```

Waiting on bqjob_r4e5ed0e382bae9a3_00000161a61cfceb_1 ... (0s) Current status: DONE

```
In [73]: TripDurations = pd.read_csv("results.csv")
TripDurations.plot.scatter(x="time_5min", y="freq")
TripDurations["logfreq"] = np.log(TripDurations["freq"])
TripDurations.plot.scatter(x="time_5min", y="logfreq")
```

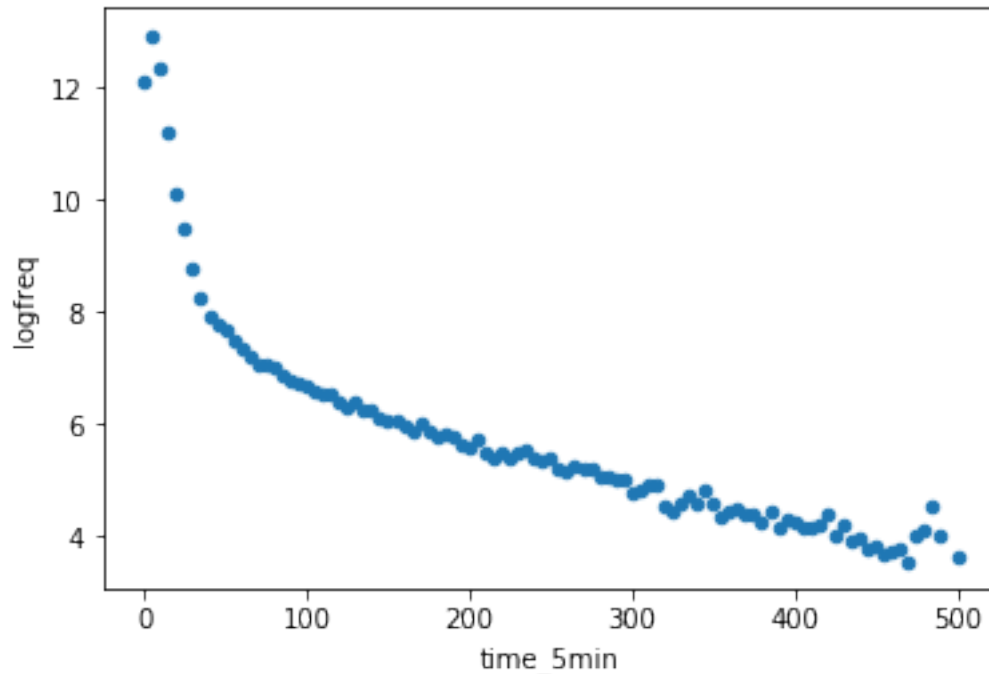
```
Out[73]: <matplotlib.axes._subplots.AxesSubplot at 0x7f0b4cc8ea90>
```



²For this duration query, the trip duration was rounded down to 5 minute intervals. (The one-second resolution offered in the data set is simply not necessary here, and makes grouping patterns less informative.) The formula applied was:

$$\text{RoundedDuration} = \text{floor}\left(\frac{\text{DurationSec}}{60} * \frac{1}{5}\right) * 5$$

This formula divides the duration (in minutes) by 5, rounds down to a whole number, and then brings the value back to minutes by multiplying by 5.



0.4 Our Current Offerings

Currently, GoBike offers two types of packages.

* Customers can choose to purchase a 1- or 3-day pass. They are allotted 30 minutes per trip *
 Subscribers can choose to purchase an annual pass. They are allotted 45 minutes per trip

In each case, the customer is charged \$3 for each 15-minute extension after they pass the allotted trip duration.

0.4.1 1- and 3-day Customers

14% of the bike trips in our data set were taken by these customers, who are likely visitors from out of town. Unfortunately, the 30-minute per-trip allocation is not serving these customers well. 42,013 of the 136,809 total customer trips were over the allowed 30 minutes. This represents a substantial 30.7%! We imagine that these customers are having poor customer experiences, since they are charged a \$3 fee for each 15 minute extension over the 30 minute trip allocation.

In [74]: ! bq query --use_legacy_sql=FALSE 'SELECT count(distinct trip_id) AS Ntrips, (duration

Waiting on bqjob_r66adc25b321909f7_00000161a61d0d70_1 ... (0s) Current status: DONE

```
+-----+-----+
| Ntrips | over_30m |
+-----+-----+
| 42013 | true |
| 94796 | false |
+-----+-----+
```

Of those customers who go over their allowed 30 minutes, 10,502 (25%) ultimately take less than 45 minutes to return the bike.

```
In [75]: ! bq query --use_legacy_sql=FALSE 'SELECT count(distinct trip_id) AS Ntrips, (duration
```

Waiting on bqjob_r2a5145631af10e22_00000161a61d1ba8_1 ... (0s) Current status: DONE

Ntrips	btwn_30m_45m
10502	true
126307	false

0.4.2 Annual Pass Holders

86% of the bike trips in our data set were taken by annual pass holders. These customers are offered 45 minute trip allowances, and yet only 2,926 of the 843,913 total subscriber trips were over 45 minutes (just 0.3%). In fact, most trips are very short. Sorting by trip length (with 15 minute resolution), we see that 772,263 subscriber trips (91%) are between 0 and 15 minutes long. (Because on the time resolution equation in footnote 2, the time values shown in the table correspond to the **bottom** of the 15-minute bucket. That is, “time_15min=0.0” marks the set of trips between 0 and 15 minutes.)

```
In [76]: ! bq query --use_legacy_sql=FALSE 'SELECT count(distinct trip_id) AS Ntrips, (duration
```

```
! bq query --use_legacy_sql=FALSE 'SELECT count(floor((duration_sec/60)/15)*15) AS fr
```

Waiting on bqjob_r66c35190c3706360_00000161a61d299c_1 ... (0s) Current status: DONE

Ntrips	over_45m
843913	false
2926	true

Waiting on bqjob_r49df7e99fcf2c62c_00000161a61d36b2_1 ... (0s) Current status: DONE

freq	time_15min
772263	0.0
69240	15.0
2410	30.0
656	45.0
383	60.0
254	75.0
184	90.0
146	105.0
132	480.0
104	120.0

```
+-----+-----+
```

Performing this same operation on the CommuterTrips sub-table reveals the probable cause of this skew towards short trips. Commuters, who represent 77% of GoBike usage, take short trips. Most are between 5 and 10 minutes. (Again recall that the time values shown in the table correspond to the **bottom** of the 5-minute bucket, so “time_5min=5.0” marks the set of trips between 5 and 10 minutes.)

```
In [77]: ! bq query --use_legacy_sql=FALSE 'SELECT count(floor((duration_sec/60)/5)*5) AS freq
```

```
Waiting on bqjob_r7cd0f06f8aa283e8_00000161a61d4584_1 ... (0s) Current status: DONE
```

```
+-----+-----+
|  freq  | time_5min |
+-----+-----+
| 199455 |      5.0 |
| 102102 |     10.0 |
|   77781 |      0.0 |
|   24812 |     15.0 |
|    4123 |     20.0 |
|    1141 |     25.0 |
|     345 |     30.0 |
|     154 |     35.0 |
|      86 |     40.0 |
|      52 |     45.0 |
+-----+-----+
```

0.5 Our Proposed Strategy

It is clear from the above analysis that the current GoBike packages are simply not in sync with our customers’ needs. 1- and 3-day pass customers want to take bikes out for long trips, and yet we limit their per-trip allocation to 30 minutes. On the other hand, annual pass holders want to use GoBike as a utility, with 10-minute trips to get around the city, and yet we offer them 45 minute allowances. With this mind, we hope management will consider the following offering suggestions:

- * Ford GoBike could increase goodwill with 1- and 3-day customers by offering just 15 minutes more per trip. This represents a small increase for the business, but give our customers the desired freedom and peace-of-mind to explore the city. Since these people are likely visitors from outside of SF, their satisfaction is of utmost importance. Positive word-of-mouth press may be a powerful asset as GoBike expands its geographic reach.

- * Ford GoBike could better serve annual pass holders, and specifically commuters, by offering a reduced-cost annual pass with a 15-minute per-trip allocation. This offering would certainly be a better match for the SF local user model, and would show our local customers that we value their loyal business.