



Self-driving Ridesharing

San Francisco Bay Area

Gaurav Dembla

Oct 21, 2020



Problem Statement

Context

- Massive Dynamic (MD), a start-up in San Francisco
- Manufacturing all-electric self-driving vehicles
- Safe for passengers and friendly for environment

Problem

1. **Market Selection** – Which market to get into?
 - a) Residents of SF, operating within the boundaries of the city, or
 - b) Residents of adjoining and nearby counties to commute in to and out of SF
2. **Marketing Segmentation** – Which groups of customers to target and why?

Source Data

1) [Travel Decision Survey Data 2017](#) from San Francisco Municipal Transportation

Agency (SFMTA). Size: 804 x 184

- Travel behavior and opinions, including trip purpose for each trip
- Number of trips to, from, and within San Francisco by Bay Area residents (does not include trips by visitors to the Bay Area and for commercial purposes)
- Basic demographic data for the sample of Bay Area residents (n=804) who travel to, from, and within San Francisco

2) Average distance of other counties to SF county from [Google maps](#).

- For SF residents, average commute distance within SF set as 5 miles (refer to assumption v).

County	Avg Distance
Alameda	55
Contra Costa	40
Marin	40
Napa	65
San Mateo	30
Santa Clara	50
Solano	60
Sonoma	80
San Francisco	5

Methodology

- ✓ Calculate potential *value* of each *market* based on travel behavior of (~800) people for a month

$$\text{Market Value} = \sum \text{Potential Value}_{\text{commuter}}$$
$$\text{Potential Value}_{\text{commuter}} = \# \text{Targetable Trips} \times \text{Average Distance of a trip}$$
$$\# \text{Targetable Trips} = \# \text{Trips} \times \text{TTF}$$

Refer to # Trips slide *Refer to TTF slide*

County	Avg Distance
Alameda	55
Contra Costa	40
Marin	40
Napa	65
San Mateo	30
Santa Clara	50
Solano	60
Sonoma	80
San Francisco	5

Trips

Trips of a commuter for a month:

- a) Within-SF: Number of trips made by an SF resident within city boundaries in last 2 days, multiplied by 11 to get an estimate for last month (refer to assumptions *vi* and *x*).
- b) To-and-from SF: Number of trips by a non-SF resident to SF and back to home county, excluding within-SF trips (refer to assumption *vii*).

$$TTF = f(\text{transportation mode/category, driving reason type, residence county})$$

			SF	non-SF	
Category of Transport	Mode of Transport	Type of reason for "Drive own vehicle"	TTF	TTF	Rationale
Car	1=Drove my vehicle alone	Soft (defined below)	0.75		Though such commuters did not cite any of the hard reasons for driving own car, it is unreasonable to assume that all their trips would qualify for ridesharing service
	2=Drove my vehicle with others 3=Drove car share	Hard (defined below)	0.25		Though such commuters cited atleast one of the hard reasons for driving by themselves, it is reasonable to assume that some of their trips could still qualify for ridesharing service
Taxi	4=Uber, Lyft, etc. 5=Regular taxi		0.5		Since Uber / Lyft / taxis will definitely respond to MD offerings, we assume that MD shall be able to capture only 50% of this segment easily (hence, TTF as 0.5) - refer to assumption ix.
Public	6=Public transportation (e.g. BART, VTA, Amtrak)		0.25	0.05	Public transport is usually cheaper and faster than driving own car or taking Uber / taxi; more so, for longer distances (non-SF residents) than for shorter distances (SF residents). For some non-SF residents who live far from the nearest public station/stand, it might be an inconvenient and only option to choose from. Since survey data does not indicate reasons for commuters to choose public transport over other options, it is safe to assume that few commuters would still be willing to use ridesharing service; more of SF residents than of non-SF residents.
Mass Private	7=Private bus or van		0		No incentive for commuter to ditch this facility because it is mostly free (provided complimentary by employers) or economical (as part of local incentives, such as SRP initiatives) and convenient for precise pickups / dropoffs
Legs	8=Bicycle		0		We assume that these options must be getting used for short distances, and hence, commuter would most likely not have a strong motive to switch to ridesharing service
	9=Walk				
Others	10=Scooter / Motorcy		0		Ignore these options as there is no tangible information available about them
	11=Other (specify)				
	12=Don't know / Don't remember				

TTF (Targetable Trips Factor) has been introduced to consider only those trips that could potentially be taken up with MD ridesharing, instead of all the trips that the commuter makes on a regular basis

Reason for "Driving car"	Type
Driving and parking is faster than other modes of travel (transit, biking, and walking)	Soft
Parking was available close to my destination	Soft
I needed to carry something	Hard
Parking at my destination was free	Soft
I needed to make multiple stops before returning home	Hard
Driving and parking is safer than other modes of travel (transit, biking, and walking)	Soft
I was traveling with children	Hard
Parking at my destination was cheap	Soft

Calculation Demo

Trips

respnum	county	sf_resident	car_reason_hard	car_trips	taxi_trips	public_trips
2584	San Mateo	0	0	0	8	4
167	Contra Costa	0	0	4	0	36
6218	Marin	0	0	20	0	30
3509	San Francisco	1	0	0	22	22
1677	Solano	0	1	30	0	0
1143	San Francisco	1	1	55	0	11

TTF Values

respnum	county	car_reason_hard	car_trips	car_ttf	taxi_trips	taxi_ttf	sf_resident	public_trips	public_ttf
2584	San Mateo	0	0	0.75	8	0.5	0	4	0.05
167	Contra Costa	0	4	0.75	0	0.5	0	36	0.05
6218	Marin	0	20	0.75	0	0.5	0	30	0.05
3509	San Francisco	0	0	0.75	22	0.5	1	22	0.25
1677	Solano	1	30	0.25	0	0.5	0	0	0.05
1143	San Francisco	1	55	0.25	0	0.5	1	11	0.25

Car TTF is dependent
on commuter's
driving reason

Taxi TTF is
same for all

Public TTF is different
for SF vs. non-SF
residence county

taxi_trips	taxi_ttf	targ_taxi_trips
8	0.5	4
0	0.5	0
0	0.5	0
22	0.5	11
0	0.5	0
0	0.5	0

Targetable Trips

respnum	county	car_trips	taxi_trips	public_trips	targ_car_trips	targ_taxi_trips	targ_public_trips	targ_trips
2584	San Mateo	0	8	4	0	4	0.2	4.2
167	Contra Costa	4	0	36	3	0	1.8	4.8
6218	Marin	20	0	30	15	0	1.5	16.5
3509	San Francisco	0	22	22	0	11	5.5	16.5
1677	Solano	30	0	0	7.5	0	0	7.5
1143	San Francisco	55	0	11	13.75	0	2.75	16.5

Potential Value_{commuter}

respnum	county	targ_trips	avg_travel_dist	value
2584	San Mateo	4.2	30	126
167	Contra Costa	4.8	40	192
6218	Marin	16.5	40	660
3509	San Francisco	16.5	5	82.5
1677	Solano	7.5	60	450
1143	San Francisco	16.5	5	82.5

Market Value

respnum	sf_resident	targ_trips	avg_travel_dist	value
2584	0	4.2	30	126
167	0	4.8	40	192
6218	0	16.5	40	660
3509	1	16.5	5	82.5
1677	0	7.5	60	450
1143	1	16.5	5	82.5

Potential Value of "non-SF
residents" market = 1428
(=126+192+660+450)

Potential Value of "SF
residents" market = 165
(=82.5+82.5)

Conclusion

Market Selection

Observations

- “SF residents” market witnesses roughly 4 times as many trips as “non-SF residents” market
- Since non-SF commuters experience much longer distances than SF residents, potential value of the former market is twice as that of the latter one

Market	# Commuters	# Trips	# Targetable Trips	Potential Value
SF residents	401	25,674	6,578	32,890
non-SF residents	403	6,204	1,615	70,000

Recommendation

Operate ridesharing service in “non-SF residents” market

Conclusion

Marketing Segmentation

1 Dive into demographics of non-SF residents with regards to potential market value

Gender	Value	Age	Value
Male	39,576	25-34	16,824
Female	29,850	35-44	14,960
Non-Binary	574	45-54	14,249
		55-64	12,578
		65+	5,967
		Unknown	3,128
		18-24	2,291

Race	Value	Income	Value
White	33,615	100k-200k	19,855
Asian	12,426	75k-100k	12,723
Hispanic/Latino	8,870	35k-75k	12,646
African American	8,116	More than 200k	10,418
Refused	6,058	Unknown	9,856
Mixed (Unspecified)	71	Less than 15k	2,018
Other	5	15k-25k	1,485
		25k-35k	997

2 Figure out *high-contrast dimensions*, showing up some segments much more valuable than others

3 Visualize heatmap of potential market value for the high-contrast dimensions

Income	Race							
	Asian	African American	Hispanic/Lati..	White	Native American	Other	Refused	Mixed (Unspecified)
Less than 15k	916	28	757	318				
15k-25k	966		396	124				
25k-35k	231	138	300	270				60
35k-75k	1,029	2,450	2,390	5,925	838	6		11
75k-100k	2,703	2,664	2,783	4,412			162	
100k-200k	3,682	2,772	1,481	11,548			372	
More than 200k	1,432	61	450	7,277			1,200	
Unknown	1,469	6	314	3,744			4,324	

4 Choose the highest value segments

Recommendation

Target white people earning \$100k-200k per year

Assumptions

- i. Actual number of SF residents moving within SF boundaries is roughly equal to actual number of non-SF residents making trips to and from SF, mimicking the (1:1) ratio of number of SF and number of non-SF residents available in the survey data.
However, we will need to check upon this assumption and most likely relax it when the project makes it way to real-life product decision.
- ii. Purpose of a trip (work, school, social, etc.) does not influence selection of MD ridesharing service over an alternate option, and hence, has not been considered for setting up TTF values.
- iii. Commuters' alternative mode of transport (Q20) has not been considered in the identification of targetable trip factor, as MD offerings are still not in the market; it is possible that if MD offerings were in the market and were quite compelling, more commuters would lean towards the ridesharing service.
- iv. TTF values have been set up based on best guesses and can be appropriately changed based on more facts when they pop up.
- v. Average distance of commute within SF has been estimated to be 5 miles, which is just half of the farthest two points in SF county (calculated from Google maps). Though there is a potential to set it to a more appropriate value, it would not change my recommendation unless it was set to at least 10.7, which is more than the longest possible trip within the city boundaries, and hence, not feasible.

Assumptions

- vi. We are willfully ignoring within-SF trips made by non-SF residents, since the question statement specifically mentioned to consider residents of SF for this market. However, we can relax this assumption when scope of markets is reconsidered, as the survey data has information about non-SF residents making within-SF trips.
- vii. We are knowingly ignoring inter-county trips originating from SF by SF residents because the problem statement specifically mentioned to consider non-SF residents for this market.
- viii. Assuming MD launches offerings that are not costlier by a substantial amount than driving own car, we can reasonably persuade drivers to switch to ridesharing service to avoid hassles of driving (especially, during rush hours) and parking own car. The higher cost of ridesharing must be balanced by more convenience associated with the service as perceived by the drivers.
- ix. It is safe to assume that MD can make ridesharing more economical than regular taxis and modern commuting options such as Uber/Lyft, easily luring commuters from the incumbent services to MD ridesharing service. However, the competitors in this segment will inevitably respond to MD offerings in some manner or the other to prevent complete loss of their market share. Hence, we assume that MD shall be able to capture only 50% of this segment easily, thereby setting TTF as 0.5.
- x. What SF residents experienced in last 2 days is a good reflection of what they must be experiencing over weekdays. As a result, we need to multiply the number of rides in last 2 days with 11 to get an estimate of total trips over 22 weekdays of the last month.

Useful Links

[Project Repository](#)

[Detailed Report](#)

[Code](#) (Jupyter Notebook)