SFO CUSTOMERS SATISFACTION

We help improve your travelers’ satisfaction

Contents

Meet Our Cool Team

We are "among "the coolest people in MSBA class of 2019

Tiffany Sung
宋耘嬋

Joey Chen
陳竑諺

Jess Lee
이지현

Pooshan Shah
浦上 沙

01. Our Goal

Help our clients figure out which aspect of the airport can they improve on?

01. Our Goal

Help our clients figure out which aspect of the airport can they improve on?

In order to reach the goal, we first need to know what do the customers (travelers) think.

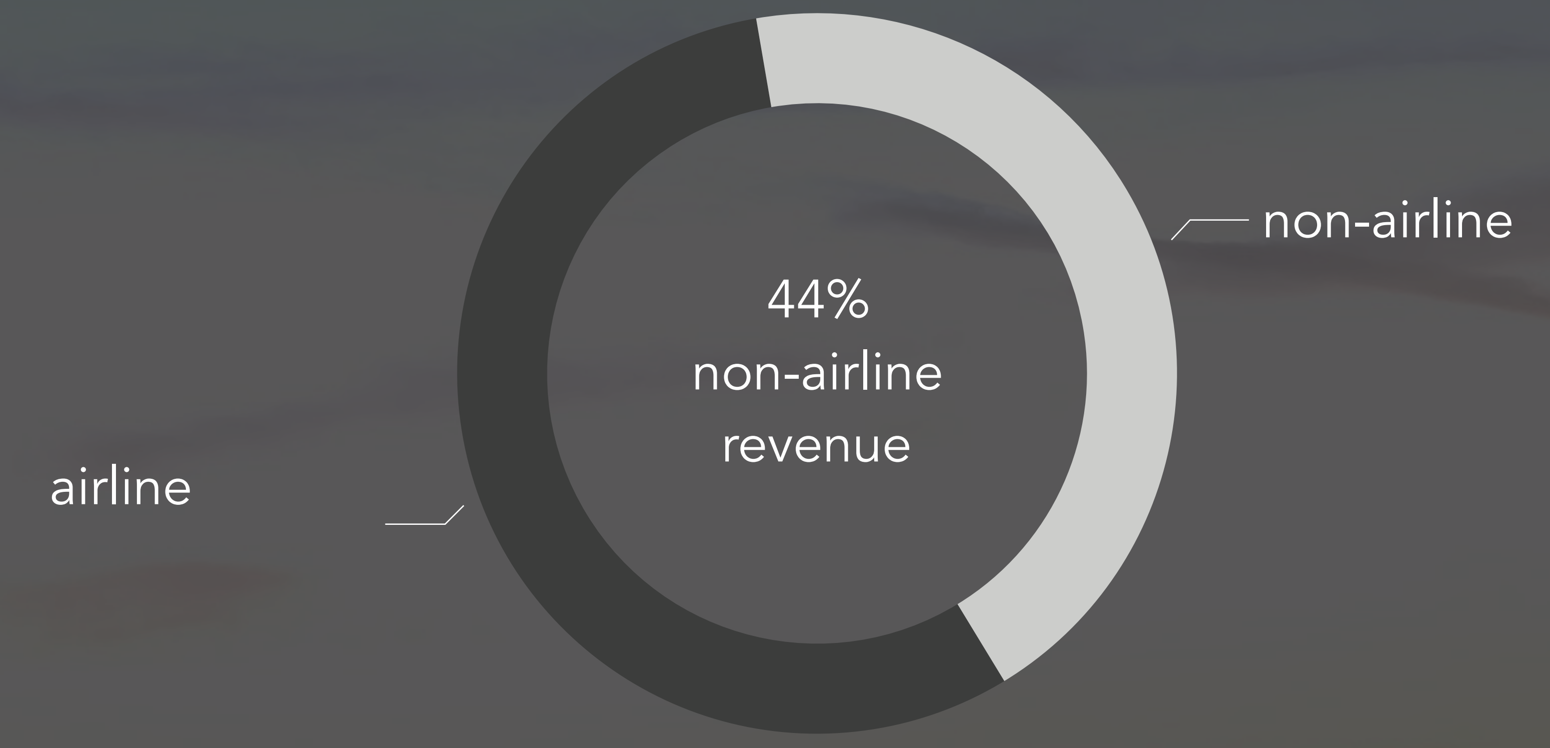
According to 2018 passenger satisfaction study from J.D. Power, SFO turns out to be 10th most satisfying airport in the United States. Also, customer satisfaction level directly impacts their consumption at airport facilities.

02 Background

“

Decrease in customer satisfaction deteriorates not only its brand but also its revenue

44% of SFO’s revenue comes from non-airline sources
Duty free sales takes 62.7% of terminal concession sales(the largest proportion compared to other international airports)



Dataset & Variables

San Francisco International Airport (SFO) 2017 Customer Survey Dataset

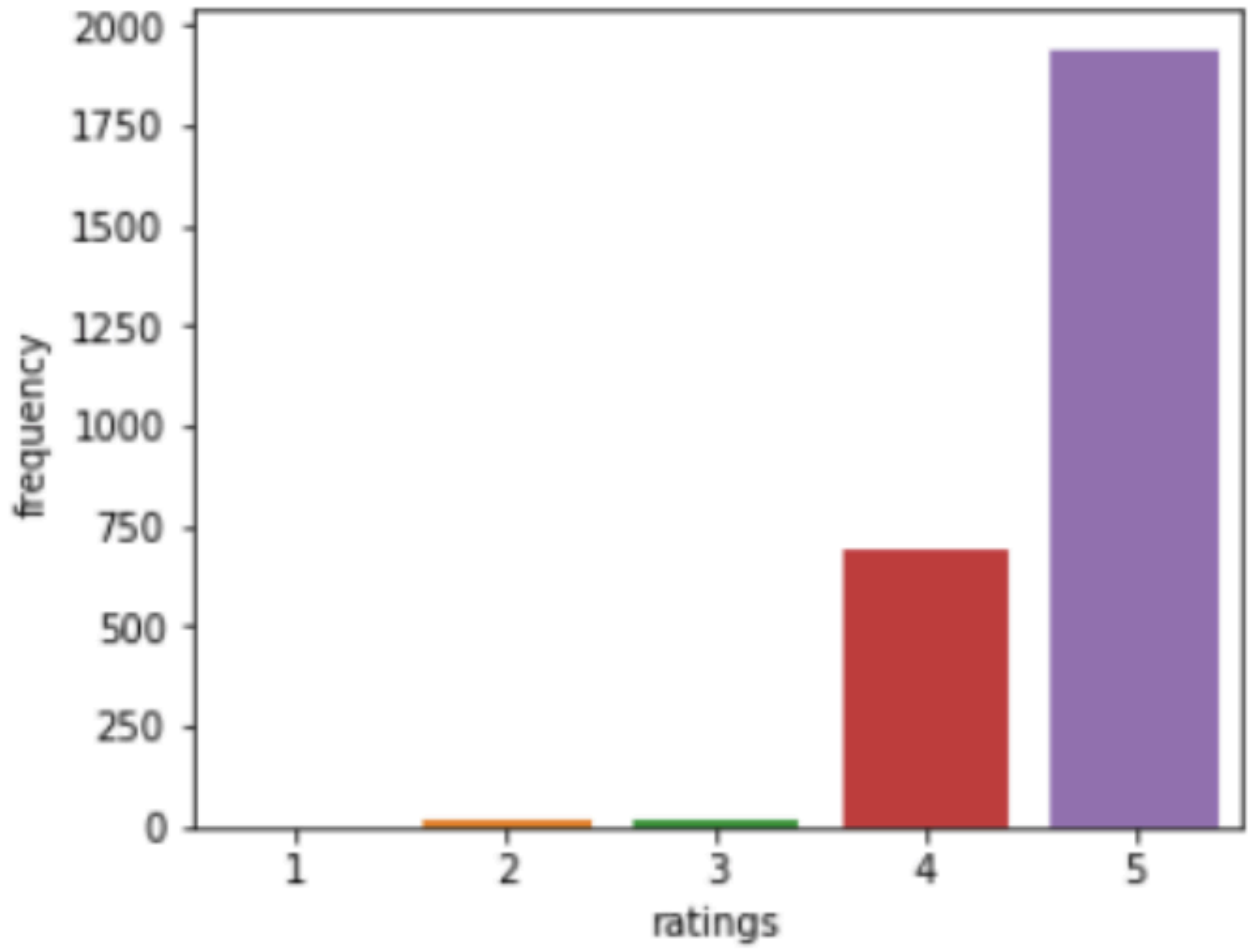
- Timeline: 2017 & 2018
- Collected through 5,640 customers surveys in each of SFO's four terminals and seven boarding areas.
- Data reflect customers' experiences with 19 individual categories, including 6 cleanliness measures, and their experience overall.
- Some demographic and behavioral attributes about the customers including their age, incomes, spending behavior, travel style…etc.

03. Analysis

We will first do explanatory data analysis on the cleanliness and safety issues survey questions. Then, we will use regression on the overall satisfaction ratings to see which facility customers care the most. During the analysis journey, we also tried by using k-means clustering to identify the customers; PCA to optimize variables…etc.

03 | Analysis— EDA

Safety



Rating 1,2

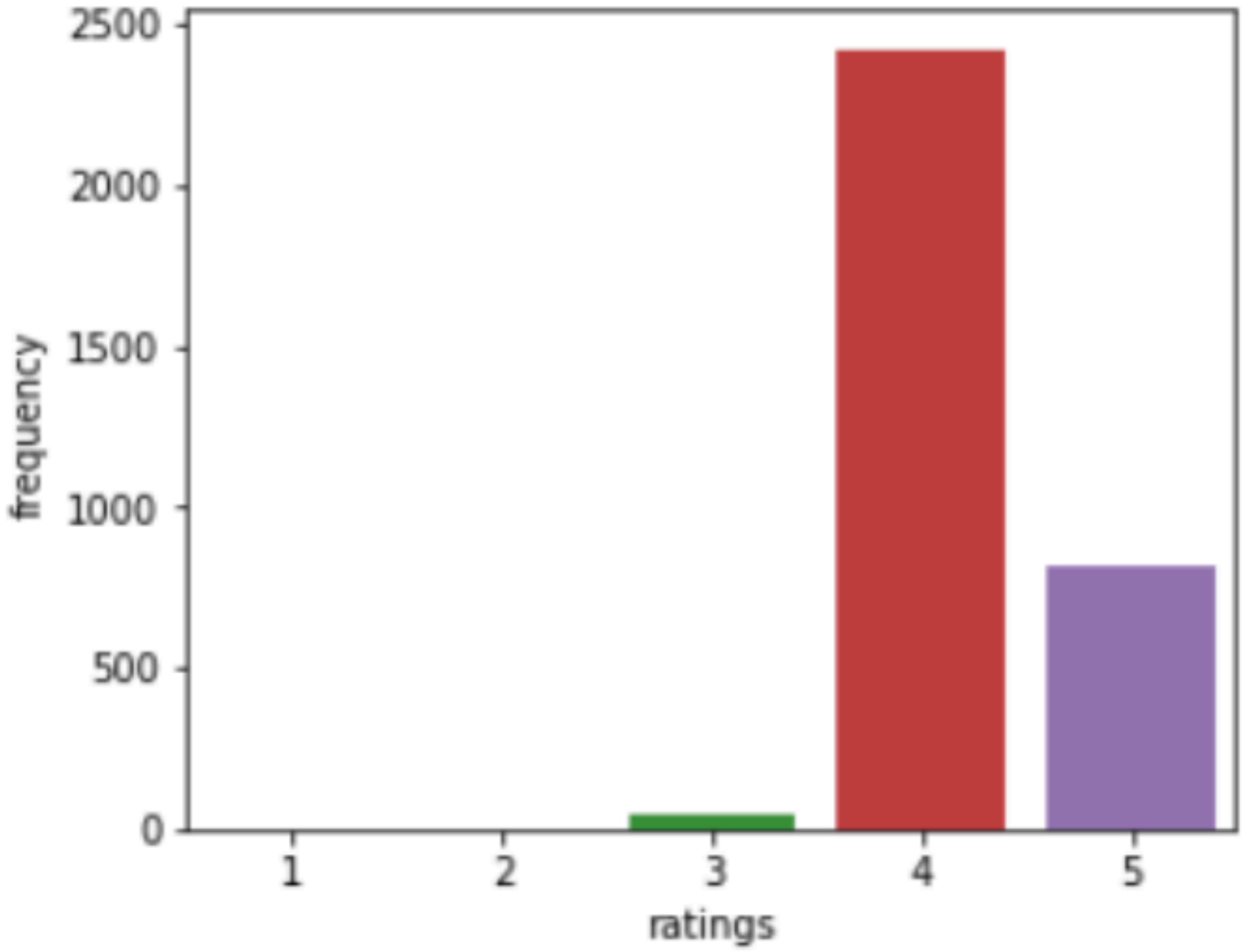
20: TSA/security did not check thoroughly/were chatting and waved people through/let things slip through they should have checked/did not care
21: Told of/saw incident/had incident occur/too many people
7: Security presence is a sham/doesn't make us any safer/just security theatre /if someone wants to do something they will
17: (negative) Haven't seen enough/very many security/police

Rating 4,5

1. There are a lot of security/officers/airport staff who are walking around/alert/effective/staff is friendly/professional
2. Security procedures/equipment/cameras are visible/effective
3. Airport is open/brightly lit/well-maintained/calm/clean/good environment
16. Nothing's happened to me personally/have not been attacked/haven't had items taken/fell asleep and nothing was taken

03 Analysis— EDA

Cleanliness



Rating 1,2

13. Security area/checkpoint dirty
11. Would like to see fellow passengers do more to clean up after themselves
9. Pet areas/other areas of airport are dirty/not maintained
7. Restaurants/eating areas dirty/littered/not cleaned

Rating 4,5

1. Airport is very clean/well maintained/keep up the good work/ other positive comment related to cleanliness
5. Carpeting is dirty/not maintained
6. Boarding areas/seating dirty/stained/litter on floor
4 Bathrooms are dirty/not maintained
7. Restaurants/eating areas dirty/littered/not cleaned

Besides cleanliness and safety, how can we continue to improve the airport in the sense of increasing customers' satisfaction and as a result generate more revenue?

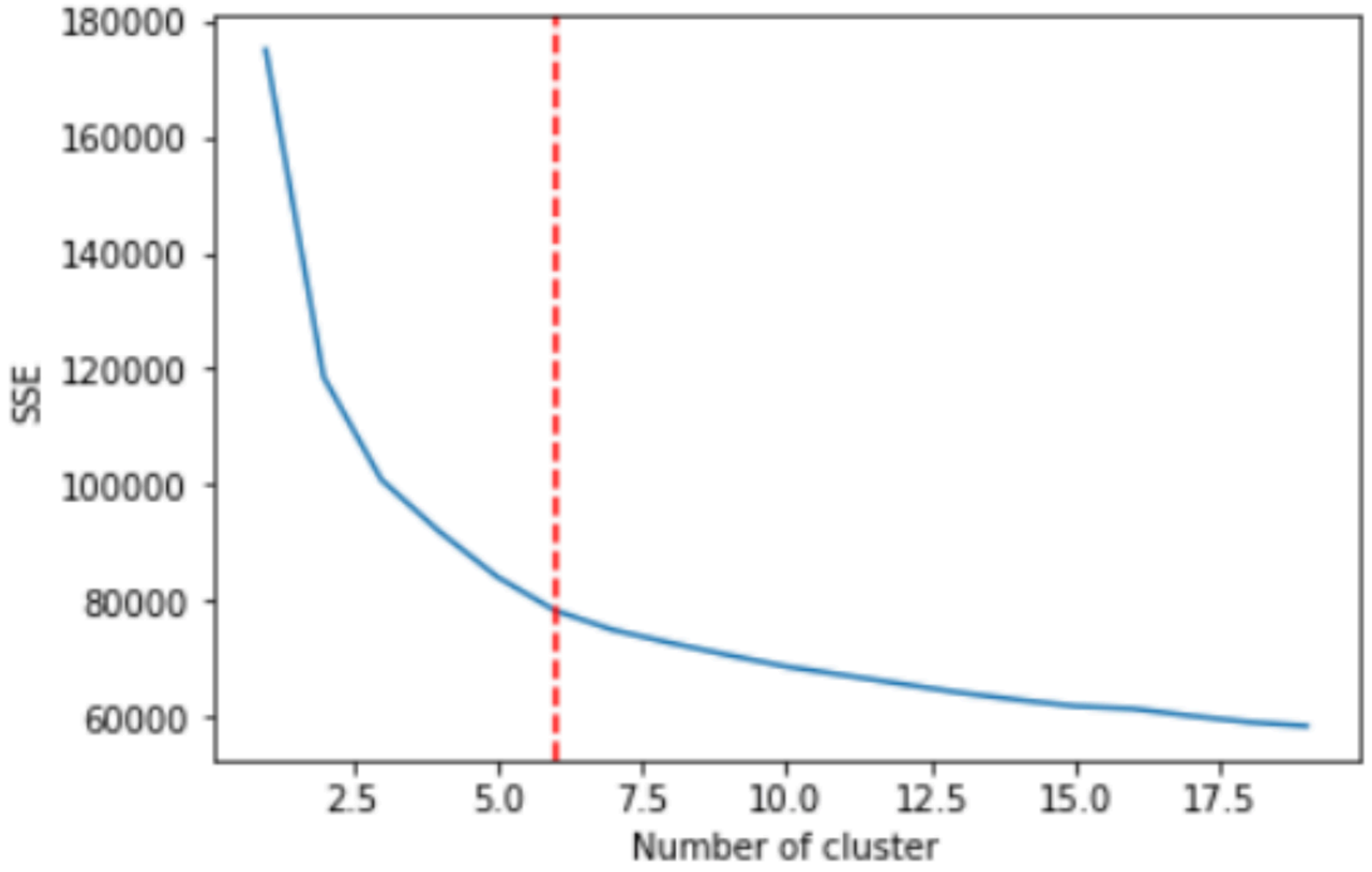
03 | Analysis —Regression

“

Regression on all variables to the overall ratings of facility…

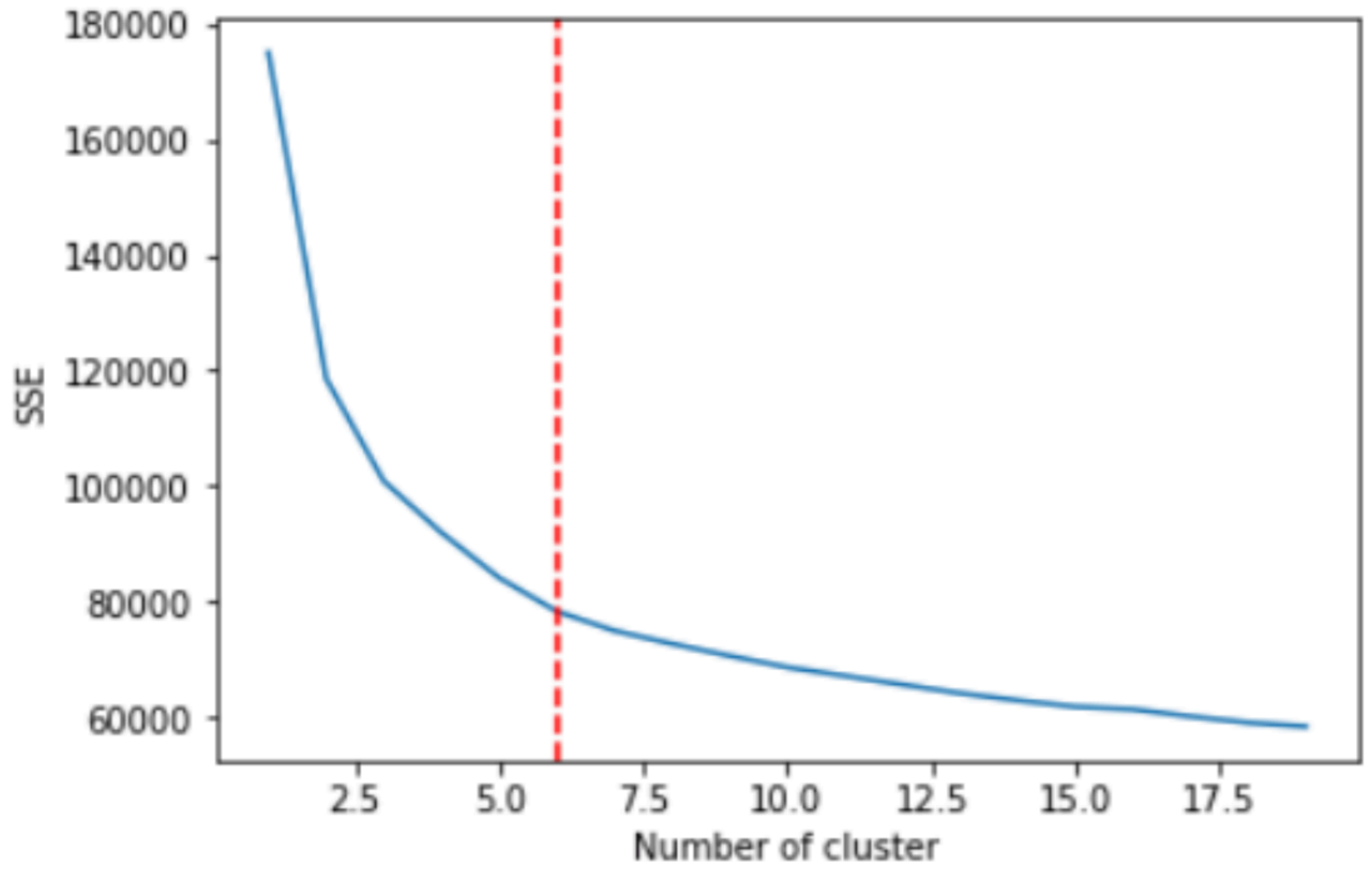
However, we should see how different types of travelers think → First, we do clustering…

03 | Analysis — K-means Clustering



We decided on 6 clusters and for each cluster, we then run the regression

03 | Analysis — K-means Clustering



Population in each Cluster

Cluster 1: 1185
Cluster 2: 242
Cluster 3: 1718
Cluster 4: 162
Cluster 5: 1029
CLuster 6: 1304

03 | Analysis — K-means Clustering

Cluster: 4

	Name	Coefficients	Standard Errors	t values	Probabilites
0	intercept	6.461450	0.150708	42.874015	0.000
1	Q7ART	0.147683	0.083748	1.763411	0.080
2	Q7FOOD	-0.203663	0.106830	-1.906430	0.058
3	Q7STORE	-0.033564	0.127879	-0.262465	0.793
4	Q7SIGN	-0.025717	0.099268	-0.259069	0.796
5	Q7WALKWAY	0.111418	0.098491	1.131251	0.260
6	Q7SCREENS	-0.278389	0.104634	-2.660601	0.009
7	Q7INFODOWN	-0.535763	0.205072	-2.612565	0.010
8	Q7INFOUP	-0.011828	0.179939	-0.065732	0.948
9	Q7WIFI	-0.129965	0.092829	-1.400045	0.163
10	Q7ROADS	-0.251713	0.095032	-2.648727	0.009
11	Q7PARK	0.056886	0.122736	0.463484	0.644
12	Q7AIRTRAIN	-0.308020	0.140867	-2.186602	0.030
13	Q7LTPARKING	-0.013718	0.137756	-0.099582	0.921
14	Q7RENTAL	-0.048784	0.154953	-0.314829	0.753

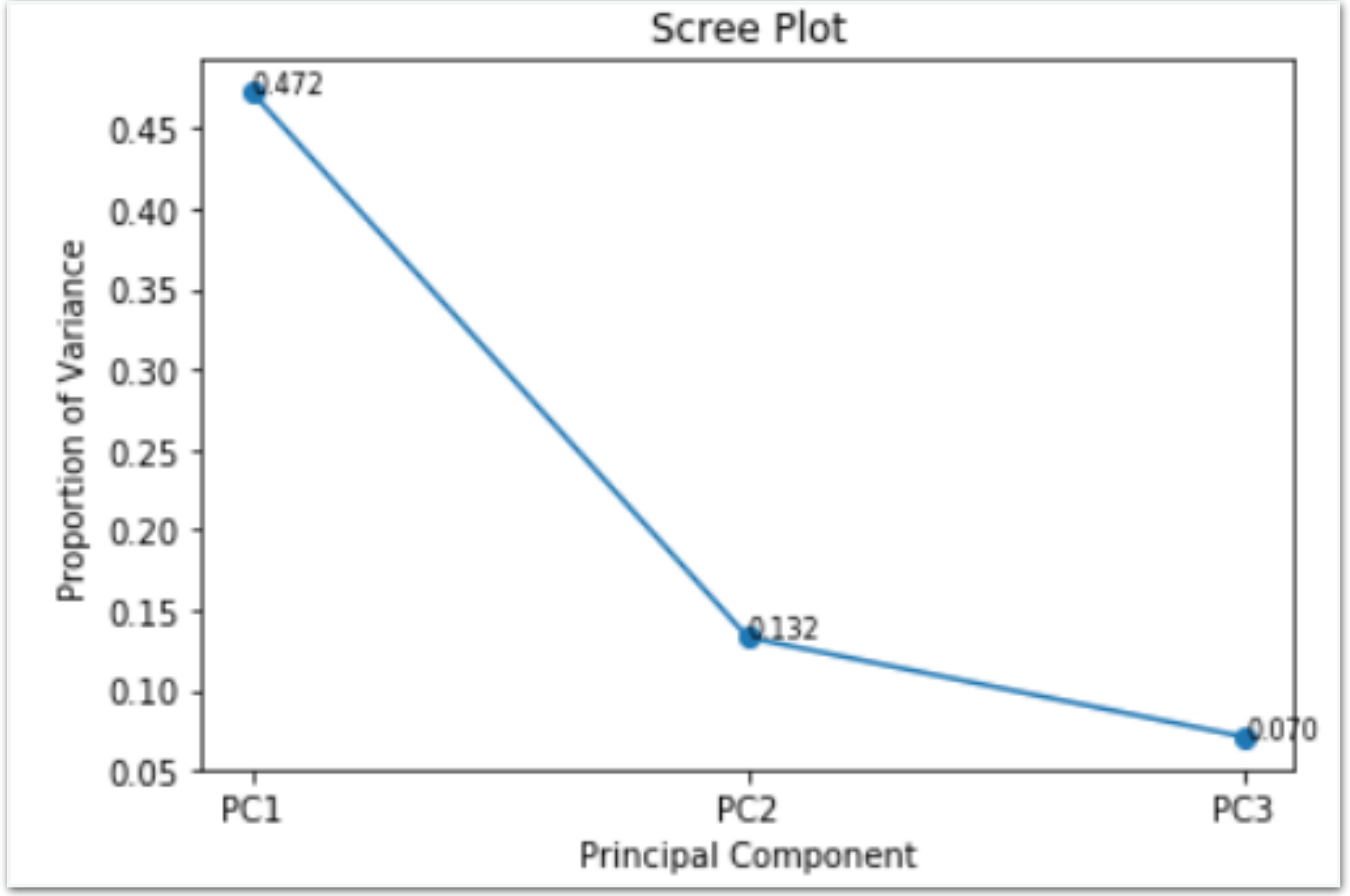
Cluster: 6

	Name	Coefficients	Standard Errors	t values	Probabilites
0	intercept	2.850935	0.403641	7.063049	0.000
1	Q7ART	0.058919	0.015873	3.711938	0.000
2	Q7FOOD	0.175303	0.020396	8.595145	0.000
3	Q7STORE	0.039293	0.019363	2.029340	0.043
4	Q7SIGN	0.168847	0.023248	7.262893	0.000
5	Q7WALKWAY	0.021741	0.022923	0.948422	0.343
6	Q7SCREENS	0.124615	0.024654	5.054502	0.000
7	Q7INFODOWN	-0.081997	0.048443	-1.692642	0.091
8	Q7INFOUP	-0.156061	0.047421	-3.290983	0.001
9	Q7WIFI	0.014817	0.014418	1.027663	0.304
10	Q7ROADS	0.064070	0.017675	3.624942	0.000
11	Q7PARK	0.006190	0.018519	0.334251	0.738
12	Q7AIRTRAIN	0.009447	0.018874	0.500565	0.617
13	Q7LTPARKING	-0.015388	0.023642	-0.650861	0.515
14	Q7RENTAL	-0.009588	0.017585	-0.545245	0.586

Looks like there're problems with collinearity among variables

→ Solution: PCA

03 | Analysis — PCA as Regression Predictors



	Q7ART	Q7FOOD	Q7STORE	Q7SIGN	Q7WALKWAY	Q7SCREENS	Q7INFODOWN
PC1	-0.230746	-0.229860	-0.241125	-0.248286	-0.266436	-0.266854	-0.287512
PC2	0.203776	0.263105	0.260874	0.282864	0.256012	0.270360	0.116452
PC3	0.179040	0.352513	0.287164	0.141468	0.121802	0.026985	-0.561549

	Q7INFOUP	Q7WIFI	Q7ROADS	Q7PARK	Q7AIRTRAIN	Q7LTPARKING	Q7RENTAL
PC1	-0.283746	-0.245896	-0.277137	-0.290853	-0.290272	-0.287377	-0.282932
PC2	0.128693	0.106901	-0.287738	-0.346219	-0.319151	-0.368324	-0.347586
PC3	-0.572188	-0.207354	0.157909	0.055229	0.082751	0.017605	0.069023

→ We choose to use 3 principal components

03 Analysis — PCA as Regression Predictors

Cluster: 1

	Name	Coefficients	Standard Errors	t values	Probabilites
0	intercept	4.172376	0.045744	91.211177	0.000
1	PC1	-0.147259	0.028750	-5.122099	0.000
2	PC2	0.188822	0.028935	6.525837	0.000
3	PC3	0.041296	0.028400	1.454103	0.146

Cluster: 2

	Name	Coefficients	Standard Errors	t values	Probabilites
0	intercept	3.327820	0.486117	6.845723	0.000
1	PC1	0.026631	0.071750	0.371166	0.711
2	PC2	0.294909	0.071714	4.112300	0.000
3	PC3	0.104463	0.060234	1.734284	0.084

Cluster: 3

	Name	Coefficients	Standard Errors	t values	Probabilites
0	intercept	4.107192	0.059715	68.779919	0.0
1	PC1	-0.154548	0.025779	-5.995068	0.0
2	PC2	0.260184	0.027474	9.470235	0.0
3	PC3	0.122887	0.032137	3.823900	0.0

Cluster: 4

	Name	Coefficients	Standard Errors	t values	Probabilites
0	intercept	0.507266	0.869971	0.583083	0.561
1	PC1	0.558992	0.089211	6.265980	0.000
2	PC2	0.103177	0.083546	1.234975	0.219
3	PC3	0.217617	0.152308	1.428794	0.155

Cluster: 5

	Name	Coefficients	Standard Errors	t values	Probabilites
0	intercept	4.266674	0.055169	77.337727	0.000
1	PC1	-0.284062	0.022562	-12.590334	0.000
2	PC2	0.057987	0.030407	1.907023	0.057
3	PC3	0.096114	0.026188	3.670113	0.000

Cluster: 6

	Name	Coefficients	Standard Errors	t values	Probabilites
0	intercept	4.284625	0.056438	75.917666	0.0
1	PC1	-0.157322	0.028313	-5.556569	0.0
2	PC2	0.130976	0.023092	5.671833	0.0
3	PC3	0.313859	0.038876	8.073336	0.0

Problems: Hard to interpret the features

→ Solution: Combine predictors together to solve collinearity

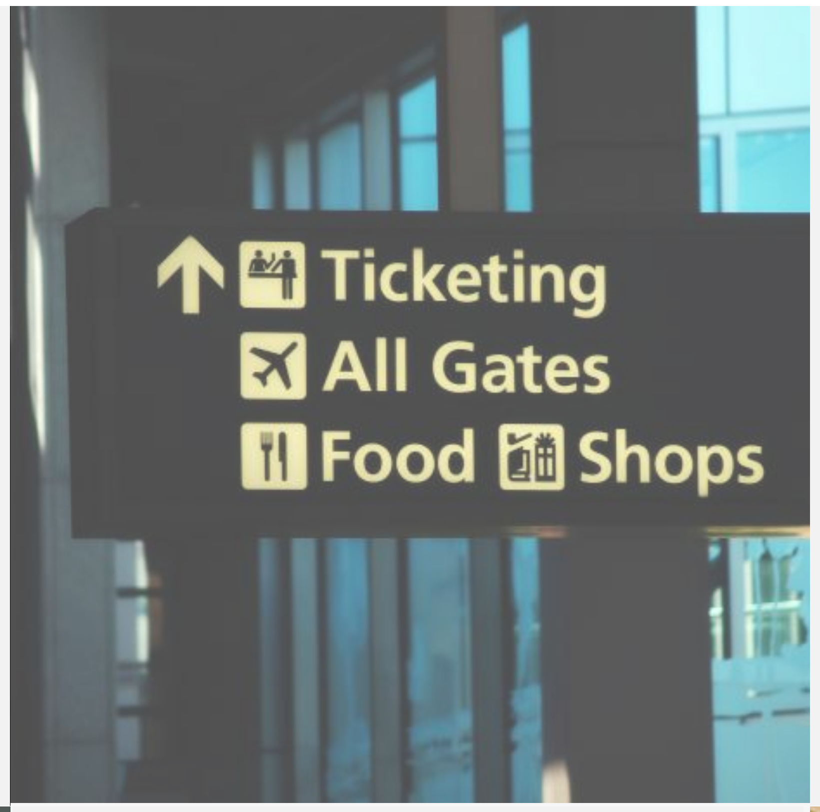
03 | Analysis —Binned Variables

03 | Analysis —Binned Variables

Based on the correlation between original variables, we binned them into 6 variables.

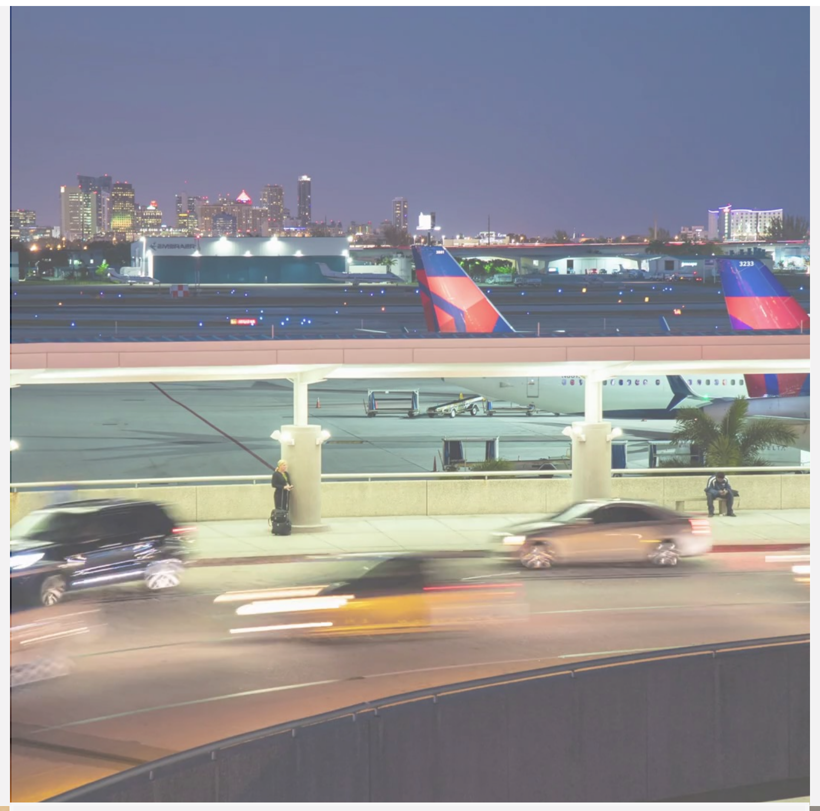
FOOD&STORE

Restaurants, retail shops and concessions



INFO BOOTH

Information booths



WIFI

Accessing and using free WiFi at SFO

ART

Artwork and exhibitions inside of the airport

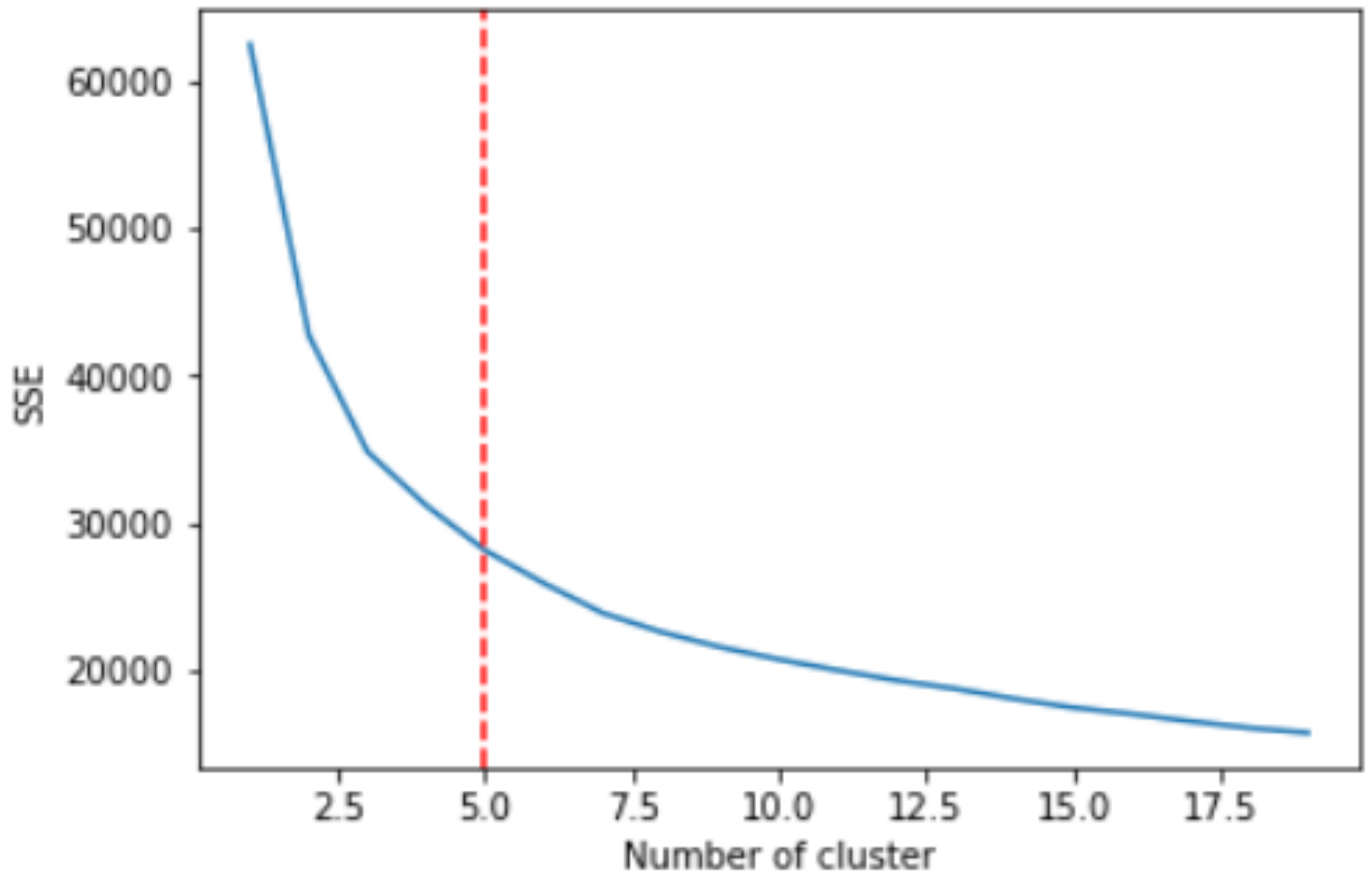
DIRECTION

Signs, directions, screens that shows information and moving walkways inside SFO

GROUND TRANSPORTATION

SFO parking facilities, airtrain, rental car center… etc.

03 | Analysis — Clustering on Binned Variables



→ Looks like 5 clusters is the best choice

Population in each Cluster

Cluster 1: 2016
Cluster 2: 1540
Cluster 3: 160
Cluster 4: 734
Cluster 5: 1190

03 | Analysis —Results

Cluster: 1

	Name	Coefficients	Standard Errors	t values	Probabilites
0	Art	0.037064	0.020499	1.808098	0.071
1	Food&Store	0.162442	0.019453	8.350465	0.000
2	Direction	0.571718	0.025152	22.730438	0.000
3	InfoBooth	0.021603	0.023404	0.923026	0.356
4	Transportation	0.045865	0.025250	1.816488	0.069
5	Wifi	0.106358	0.020668	5.146034	0.000

Cluster: 2

	Name	Coefficients	Standard Errors	t values	Probabilites
0	Art	0.070767	0.017318	4.086297	0.0
1	Food&Store	0.194932	0.023187	8.406861	0.0
2	Direction	0.400922	0.029481	13.599490	0.0
3	InfoBooth	0.143714	0.024406	5.888527	0.0
4	Transportation	0.123644	0.018481	6.690380	0.0
5	Wifi	0.079015	0.017132	4.612180	0.0

Cluster: 3

	Name	Coefficients	Standard Errors	t values	Probabilites
0	Art	1.089257	0.292514	3.723777	0.000
1	Food&Store	0.414397	0.386413	1.072420	0.285
2	Direction	0.727743	0.362852	2.005617	0.047
3	InfoBooth	0.298957	0.730252	0.409389	0.683
4	Transportation	0.650649	0.324655	2.004126	0.047
5	Wifi	0.716038	0.373667	1.916248	0.057

Cluster: 4

	Name	Coefficients	Standard Errors	t values	Probabilites
0	Art	0.115788	0.034833	3.324089	0.001
1	Food&Store	0.350722	0.048257	7.267764	0.000
2	Direction	0.546864	0.049243	11.105452	0.000
3	InfoBooth	-0.082414	0.035899	-2.295713	0.022
4	Transportation	-0.013339	0.025695	-0.519114	0.604
5	Wifi	0.195545	0.031516	6.204559	0.000

Cluster: 5

	Name	Coefficients	Standard Errors	t values	Probabilites
0	Art	0.082426	0.019003	4.337579	0.000
1	Food&Store	0.200021	0.024473	8.173148	0.000
2	Direction	0.412996	0.028782	14.348876	0.000
3	InfoBooth	0.088140	0.030327	2.906523	0.004
4	Transportation	0.120062	0.024775	4.846060	0.000
5	Wifi	0.044219	0.015684	2.819331	0.005

Then, we got to figure out who are the customers in each clusters?

04 | Target Customers

We got 5 Clusters

In each clusters, we look at their traveling purposes, gender, income and main language…etc. We then counted the percentage of those features in order to figure out who are the main composite in that particular cluster.

2016

Their main traveling purposes are educational student events like conferences. Besides Direction, they care about food and wifi. They are the least income group.

1540

This group of people travel for vacation. It's not hard to imagine they are relatively high income people and some use foreign currency. In general, they care about every facilities.

160

This is the least representative clusters which have the smallest population and they care about art facility in the airport. They travel for other purposes instead of those main stream purposes.

734

They travel for special occasion like friends wedding, graduation and funeral. They have average income and they care about food, stores and wifi the most.

1190

These are the business guys. They are the highest income group among all clusters. They're not only the richest but they care about everything in the airport. So don't mess with them.

04 | Target Customers — Cluster 2 and 5

Top 5 Suggestions

- Offer a wider variety of restaurants
- Need more selections of shops
- Need more seating at boarding areas
- Rental car center and parking are too congested

Top 5 Complaints

- Security waiting time too long
- Difficulty finding gates and terminals
- Signs/info boards are confusing and incomplete
- WiFi could not connect

05 | Recommendation

“

All in all, we’ve concluded with some main concerns for SFO…

- Direction issue: confusing signs
- Operational/ services issue: more seat numbers, less waiting time
- Food/stores issue: need more variety
- Wifi issue: better connection

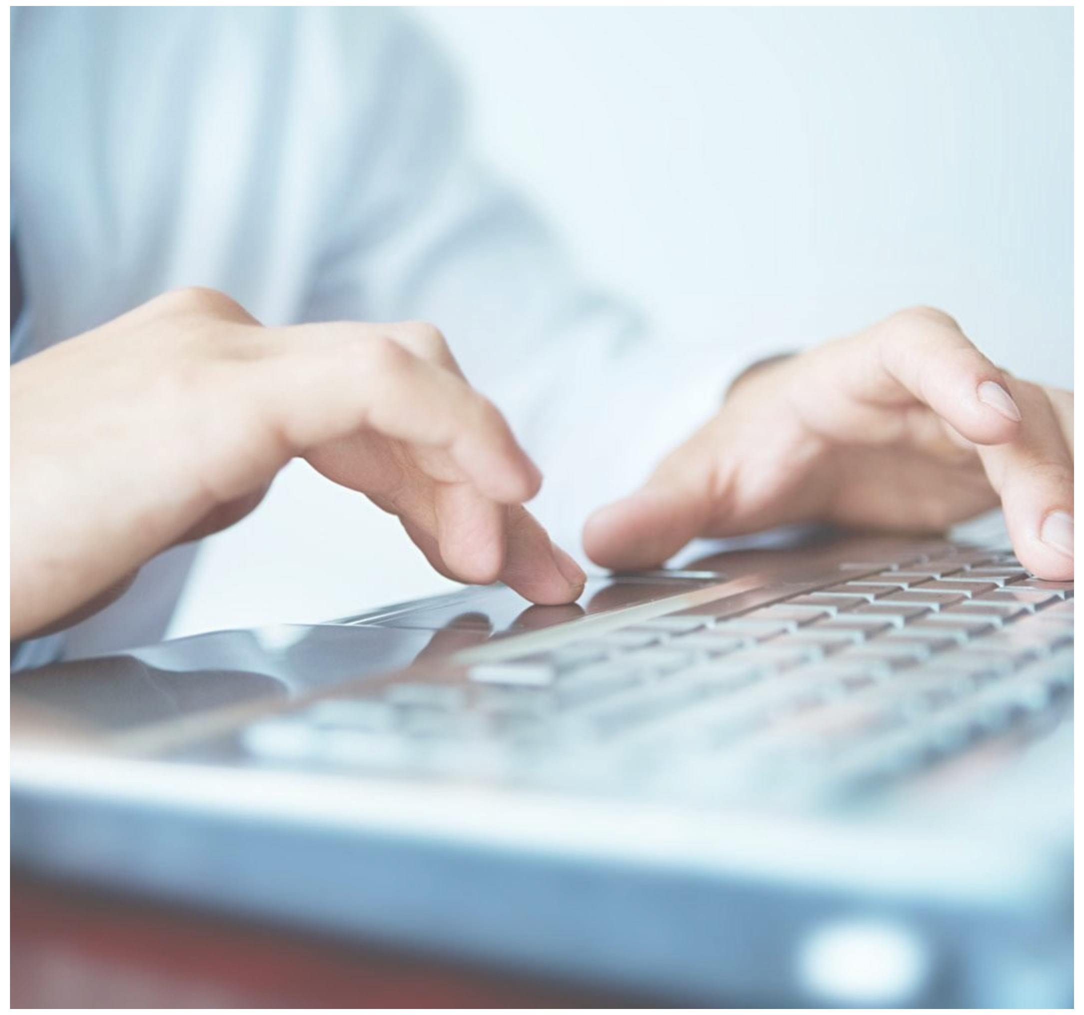
05 | Recommendation

Benchmark to Solve Direction Issue

Use bench marking and compare to similar airport as reference. Figure out which specific part of airport need to fix direction issues

Customers Demographic Analysis

Understand the race, gender, cultural background of customers to figure out which types of cuisine, stores to included in the airport



WiFi Business Model

Build the wifi business model for segregating different wifi needs of customers to solve the connection problems and can even bring more revenue.

Thank You

Q&A

* # Appendix — Our Assumptions

1. We didn’t include Cleanliness, Safety but only rating on facilities because we expect customers needs of them not to be different by their background
2. Target customers are cluster 2 and 5 based on income. It was important to focus on their suggestions and complaints because they were also caring about all facilities.
3. Among recommendations, we decide to focus on few issues which is directly related to marketing
4. We found that enhancing variety of food/restaurants is especially import because customers are suffering from lack of seats and long waiting time. Their longer stay at the airport is a good opportunity for shops/restaurants.