San Francisco International Airport (SFO)

Samantha Riedener | Funmi Esuruoso | Parin Patel | Joseph Azar

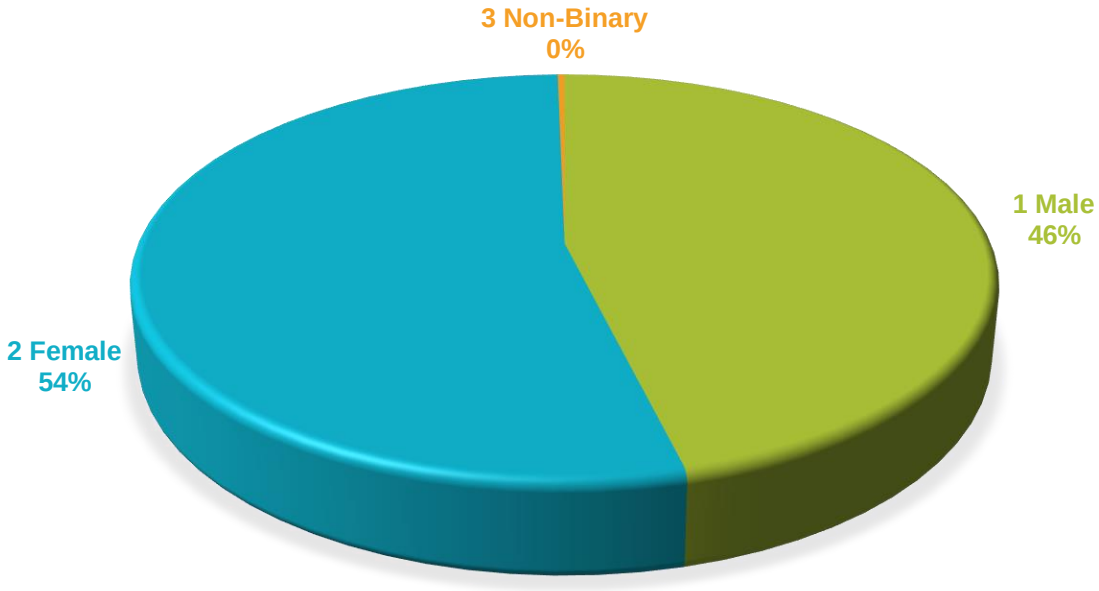
Executive Summary

San Francisco International Airport (SFO) is located in California and is a major gateway to various San Francisco attractions and internationally, Europe and Asia. SFO is the largest airport in Northern California and the second busiest in California and West Coast, after Los Angeles. In 2017, it was listed as the seventh-busiest airport in the United States and the twenty-fourth busiest in the world (by passenger count). Serving as the fifth-largest airline hub for United Airlines and the secondary hub for Alaska airlines, SFO houses restaurants, cafes, stores, and many more.

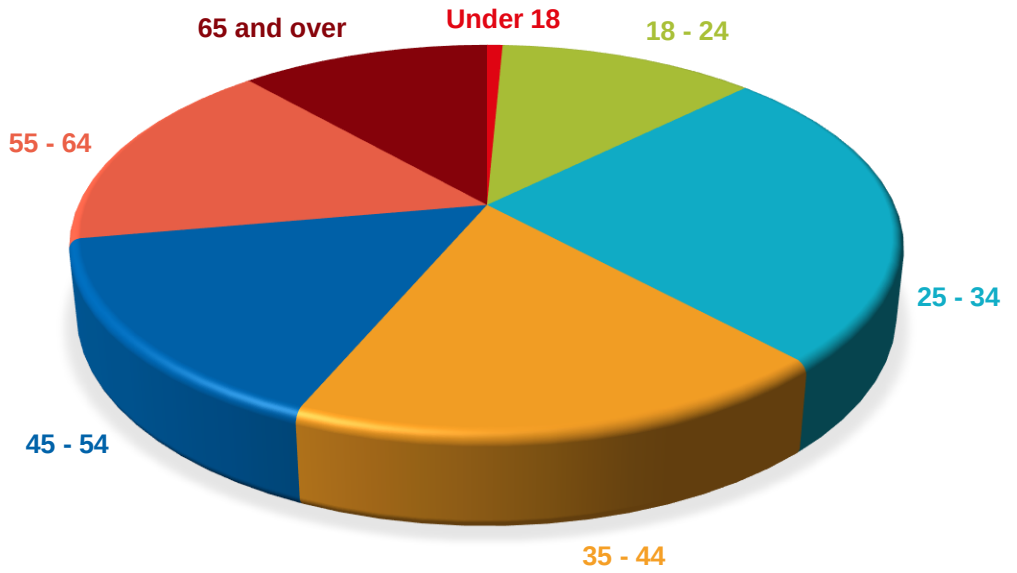
Following analysis of SFO's survey data where customers were asked to rate their experience at the airport with nineteen (19) different categories from cleanliness to travel style and spending behavior at the airport, we identified lapses in restaurant and store spending. By identifying where customer satisfaction were the lowest, we could identify and provide recommendations for improvement. Some of the analysis techniques included k-means clustering, ordinal logit regression, and logistic regression.

Demographics

GENDER

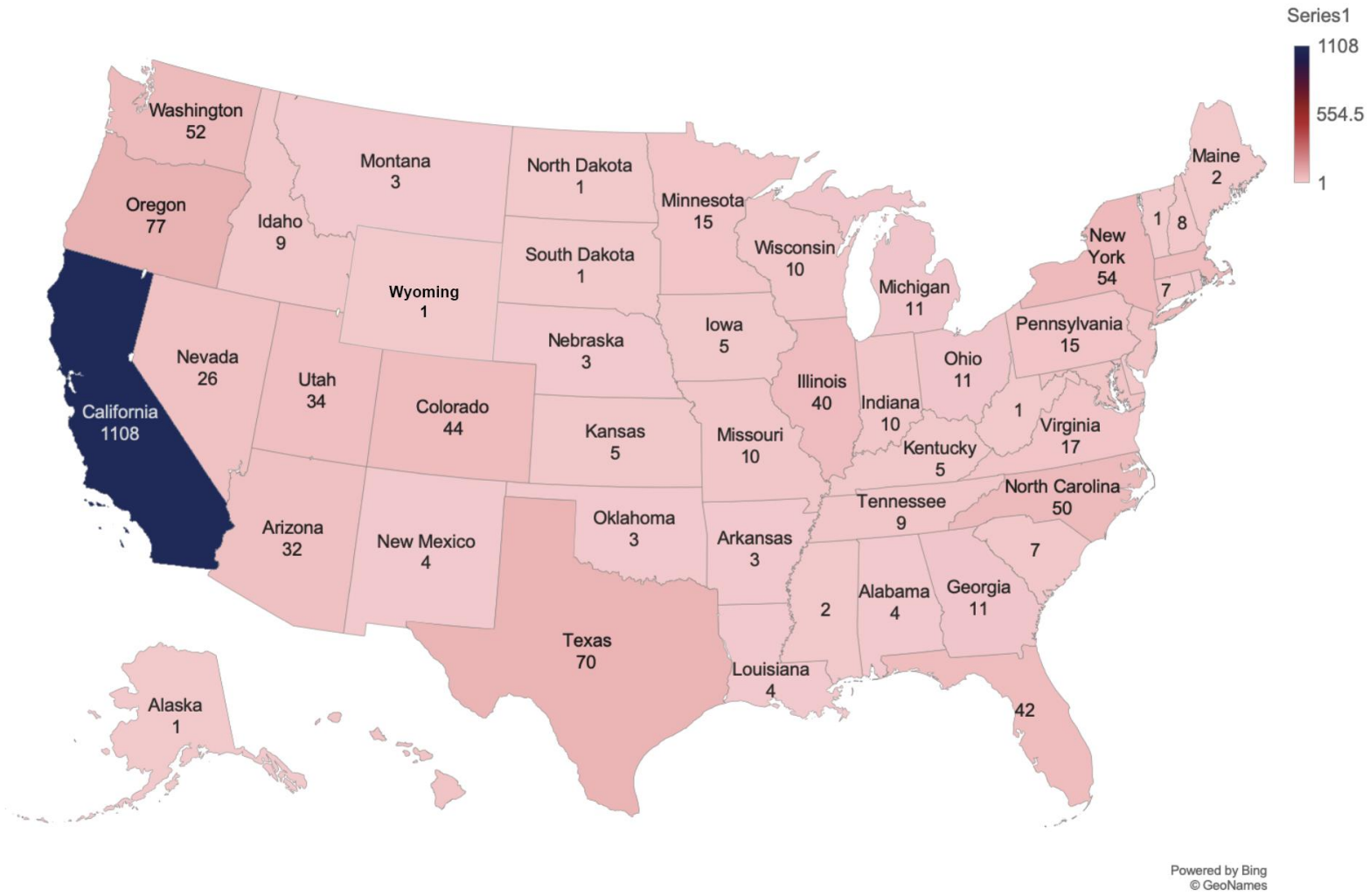


AGE



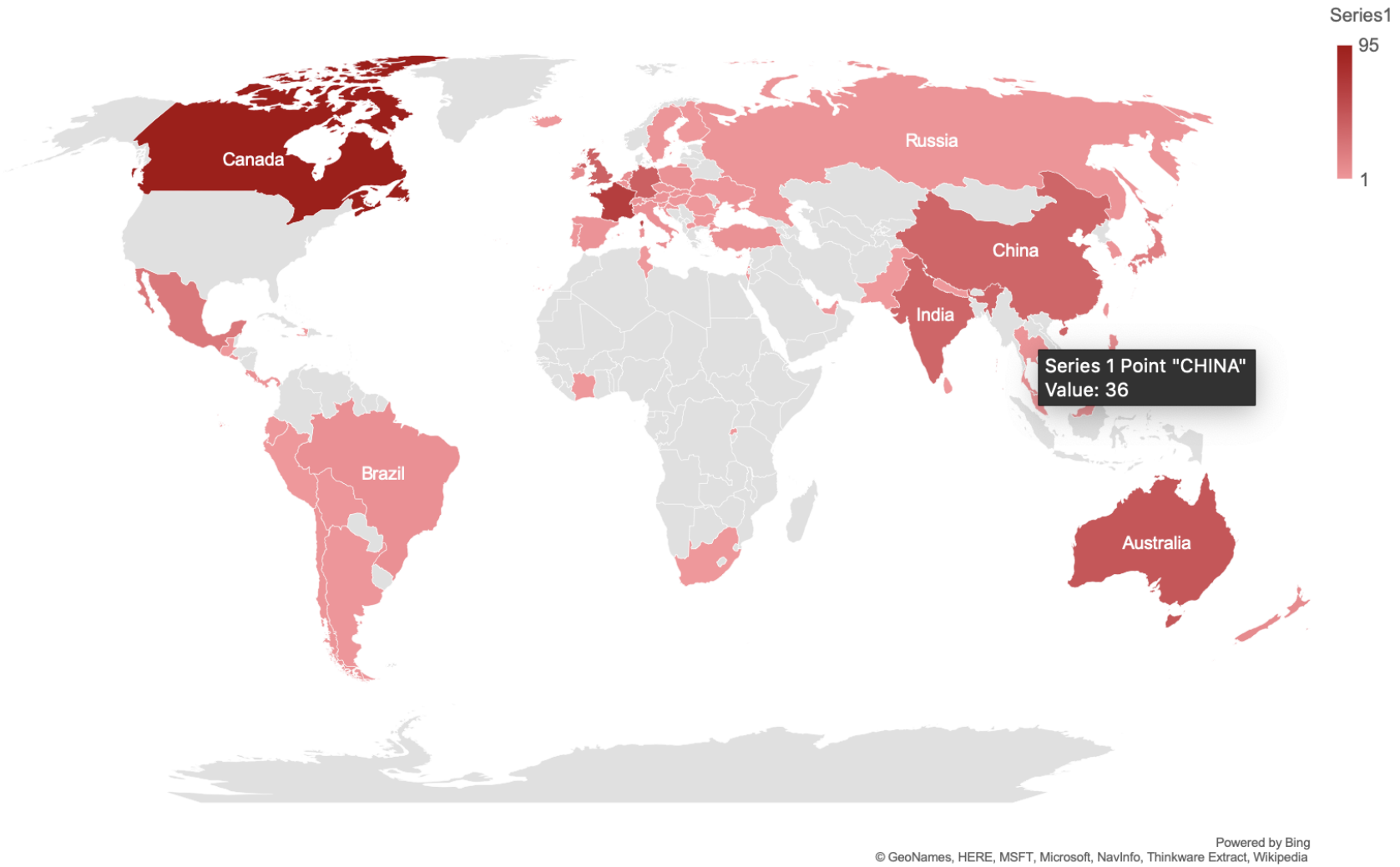
United States

CA	1108	39.44%
OR	77	2.74%
TX	70	2.49%
NY	54	1.92%
WA	52	1.85%
NC	50	1.78%
MA	45	1.60%



World

US	2048	72.91%
CANADA	95	3.38%
FRANCE	66	2.35%
AUSTRALIA	47	1.67%
GERMANY	40	1.42%
UNITED KINGDOM	39	1.39%
INDIA	36	1.28%
CHINA	36	1.28%
MEXICO	23	0.82%
NEW ZEALAND	10	0.36%



Elbow Plot

Evolution of Variance:

Variance\Classes	2	3	4	5	6	7	8	9	10
Within-class	5831.149	2921.962	1840.306	1440.581	1132.088	739.305	633.150	582.689	497.705
Between-classes	7439.018	10348.205	11429.861	11829.586	12138.079	12530.862	12637.017	12687.478	12772.462
Total	13270.167	13270.167	13270.167	13270.167	13270.167	13270.167	13270.167	13270.167	13270.167
Ratio	0.783860093	0.282364163	0.161008627	0.12177778	0.093267438	0.058998732	0.05010278	0.04592628	0.03896706

K-means clustering

Survey Questions	Flyer 1 Profile	Flyer 2 Profile	Flyer 3 Profile
Flight time	PM (Flights departing after 5 pm)	MID (Flights departing 11 am to 5 pm)	PM (Flights departing after 5 pm)
Length from arrival to departure	4 hours	2 hours	12 hours
Made store purchase	Yes	No	Yes
Made food purchase	Yes	Close split (yes & no)	Yes
Used free Wi-fi	Yes	Yes	Yes
Overall perception of SFO	Good view of SFO as a while.	Good view of SFO as a while.	Good view of SFO as a while.
Perception of restaurants (cleanliness)	(4) Slightly above average	(4) Slightly above average	(4) Slightly above average
Perception of restroom (cleanliness)	(4) Slightly above average	(4) Slightly above average	(4) Slightly above average
Navigating through the airport	(5) Easy	(5) Easy	(5) Easy
Age	B/w 25-34 & 55-64	B/w 25-34 & 35-44	35-44
Gender	Closely male & female	Mostly female	Mostly male
Recommendation	Likely	Extremely likely	Extremely likely
Income	$50,000 - $100,000	Over $150,000	Over $150,000

The K-means clustering method was used to profile flyers in order to gain a deeper understanding of the types of people visiting and not visiting stores and restaurants at SFO.

Based on our model, we identified three (3) flyer profiles. This would not only serve as a means to understand SFO flyers, but would become essential for our recommendations to store and restaurant owners.

For example, the Flyer 3 profile are older people who are wealthy. They are more likely to spend more on stores and restaurants/bars due to their disposable income. Additionally, their longer stay at the airport makes them susceptible to impulse buys.

Ordinal Logit Regression

- Created two (2) different ordinal logit models to assess:
  - Does time at the airport affect passenger referral rating?
  - Does restaurant rating have an effect on passenger referral rating?
- Note: Chose ordinal logit regression in this case because the survey asked respondents to give answers on an ordinal scale

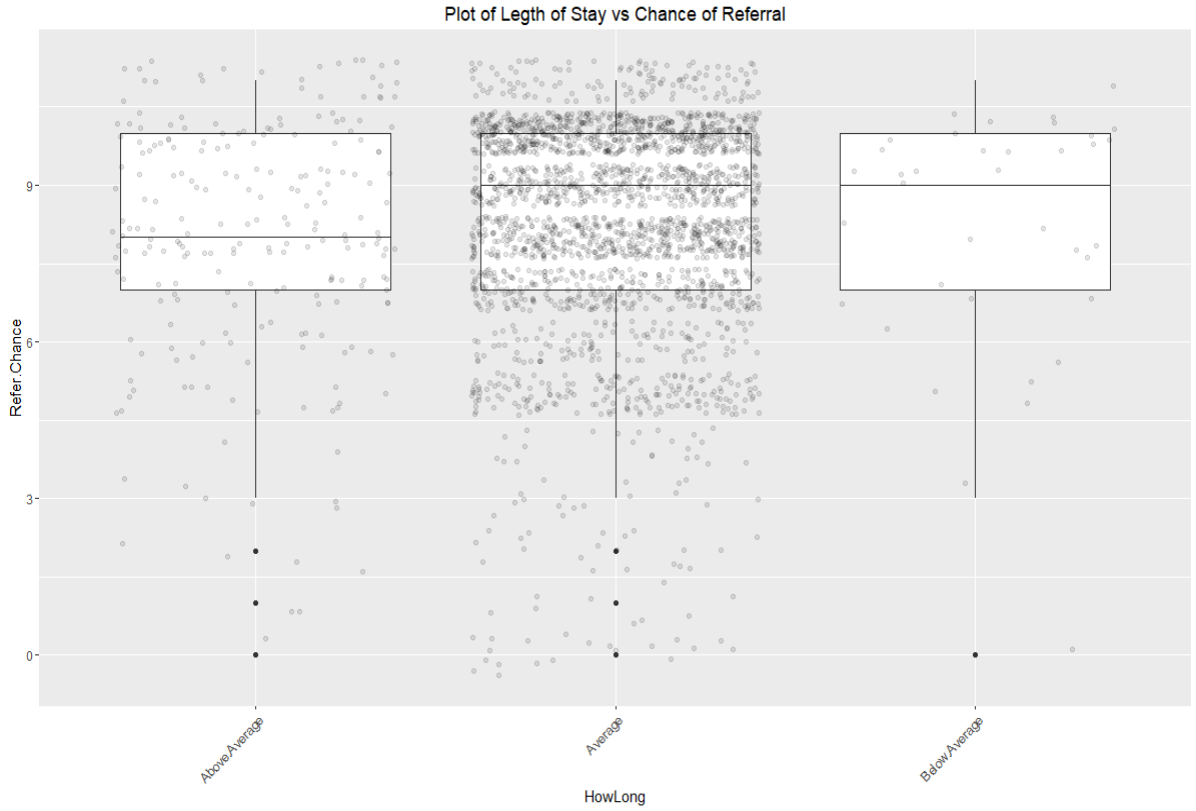
First model:
Time At Airport vs Referral

Summary Statistics - HowLong	
Mean	168.7001
Standard Error	2.197973
Median	145
Mode	120
Standard Deviatic	115.0739
Sample Variance	13242
Kurtosis	25.70159
Skewness	4.054929
Range	1380
Minimum	15
Maximum	1395
Sum	462407
Count	2741
Confidence Level	4.309852

- Using the Standard Deviation and Mean, we were able to create an ordinal code for passenger time at the airport.
  - Noted in Minutes
  - Removed all rows with no responses.

- Outcome variable: Time at airport.

New Code	
Below 52 Minutes	Below Average
53-283 Minutes	Average
Above 284 Minutes	Above Average



First model:
Time At Airport vs Referral

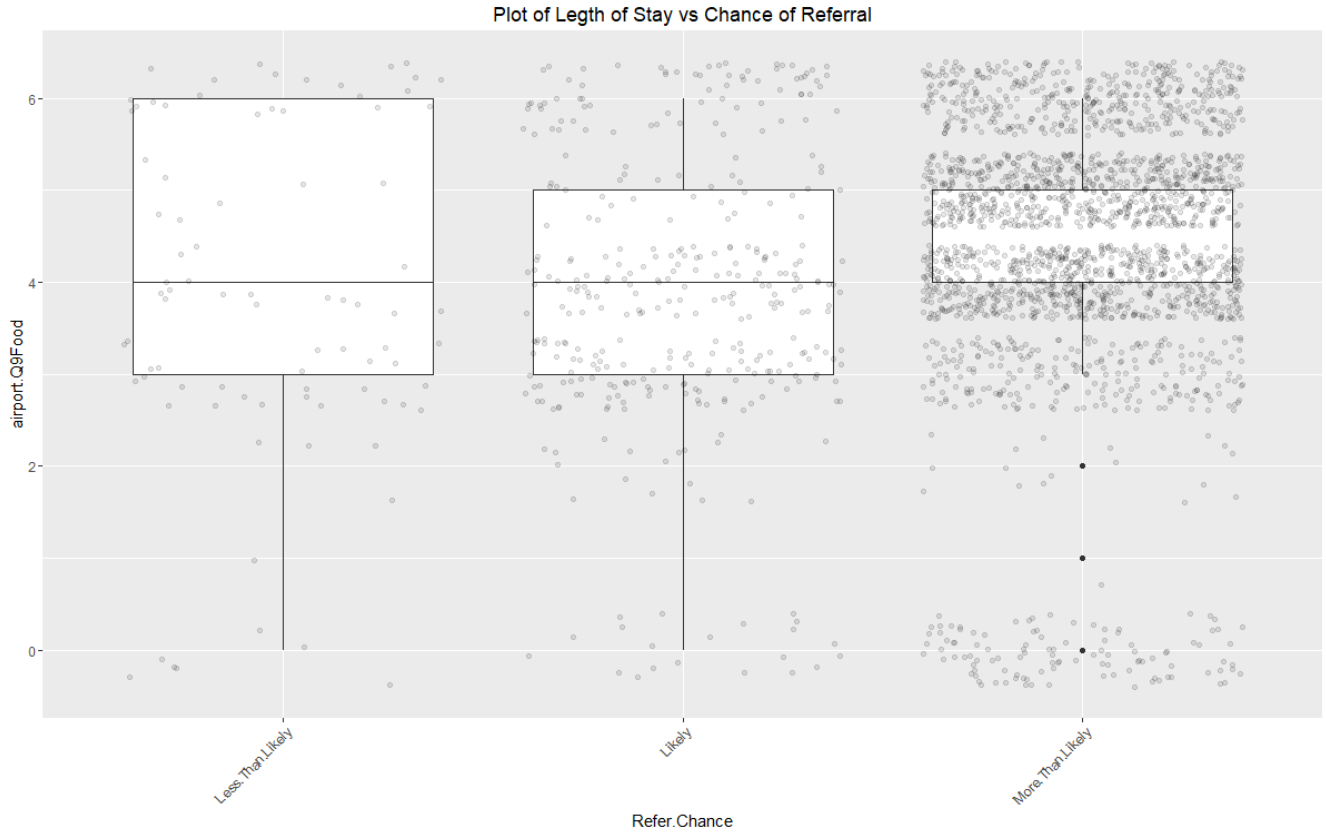
```
> stargazer(m,type = "text", out = "m.htm")

=============================================
                      Dependent variable:
                  ---------------------------
                            HowLong
---------------------------------------------
Refer.Chance                0.056*
                            (0.031)

---------------------------------------------
Observations                 2,741
=============================================
Note:              *p<0.1; **p<0.05; ***p<0.01

>
```

- When Referral Chance increases one unit, it is 0.056 times more likely to be in a higher category. The odds of moving to a higher category in the outcome variable is -94% more likely when Referral Chance moves one unit.
- Referral Chance is not statistically significant of time at the airport.
- The longer a customer stays at the airport is not related to a better approval rating. We can take this to the restaurants and shops to help tune marketing strategies that are not focused on taking the time to attract customers.

Secondly:
Passenger Referral vs Restaurant Rating

- For the second model we re-coded passenger referral ratings to narrow our spread.
- Similar to the first, we grouped the referral into three categories.

New Code Referral	
Below Rating of 4	Less Than Likely to Refer
Rating 4-6	Likely to Refer
Rating Over 6	More Than Likely To Refer



Secondly:
Passenger Referral vs Restaurant Rating

```
> (ctable <- cbind(ctable, "p value" = p))
                        Value Std. Error    t value      p value
airport.Q9Food          0.2365578  0.0334059   7.081318 1.427900e-12
4|Less.Than.Likely     -10.5045882  5.5492792  -1.892964 5.836260e-02
Less.Than.Likely|Likely -2.4470705  0.1701598 -14.381018 6.808602e-47
Likely|More.Than.Likely -0.6180365  0.1430515  -4.320377 1.557631e-05

>
➢ (ci <- confint(m)) # default method gives profiled Cis
➢ Waiting for profiling to be done...

➢  2.5 %    97.5 %
➢ 0.1703717 0.3010730
```

```
> stargazer(m, type = "text", out = "m.htm")

================================================
                         Dependent variable:
                     ---------------------------
                              Refer.Chance
------------------------------------------------
airport.Q9Food                  0.237***
                                (0.033)

------------------------------------------------
Observations                     2,649
================================================
Note:                *p<0.1; **p<0.05; ***p<0.01
```

- When passengers chance of referral increases by one unit, it is 0.237 times more likely to be in a higher category. The odds of moving to a higher category in the outcome variable is 76.3% more likely when Referral Chance moves one unit.
- Referral Chance is statistically significant to a passengers rating of the restaurant.
- A passenger experience at SFO's restaurants does affect the outcome of their referral.

Logistic Regression

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.5408251	0.4357312	1.241	0.2145	
airport.DAY_Monday1	-0.3159634	0.3656865	-0.864	0.3876	
airport.DAY_Tuesday1	-0.0046261	0.2186808	-0.021	0.9831	
airport.DAY_Wednesday1	-0.1694702	0.2197414	-0.771	0.4406	
airport.DAY_Thursday1	-0.1206198	0.2235084	-0.540	0.5894	
airport.DAY_Friday1	-0.1478772	0.2357684	-0.627	0.5305	
airport.DAY_Saturday1	-0.4940928	0.2717188	-1.818	0.0690	.
airport.STRATA_AM1	-0.0832948	0.1522747	-0.547	0.5844	
airport.STRATA_MID1	-0.0541082	0.1458825	-0.371	0.7107	
airport.Q21Gender_Male1	-0.0640320	0.1170864	-0.547	0.5845	
airport.HOWLONG	-0.0026514	0.0006112	-4.338	1.44e-05	***
airport.Q20Age	0.0882543	0.0396712	2.225	0.0261	*
airport.Q22Income	-0.0049168	0.0560148	-0.088	0.9301	
airport.Q7FOOD	-0.0376460	0.0707198	-0.532	0.5945	
airport.Q9Food	0.0884613	0.0815956	1.084	0.2783	

Logistic regression was done to try and isolate which variables have an effect on the likelihood of making a purchase at the restaurants.

Dummy variables to examine day of the week, time of flight and gender were introduced to measure if they had an effect.

Only the total time spent in the airport and the age showed statistical significance. As you spend more time in the airport, you're less likely to make a purchase. As your age increases, you're more likely to make a purchase.

People who are older most likely have more disposable income to spend in restaurants.

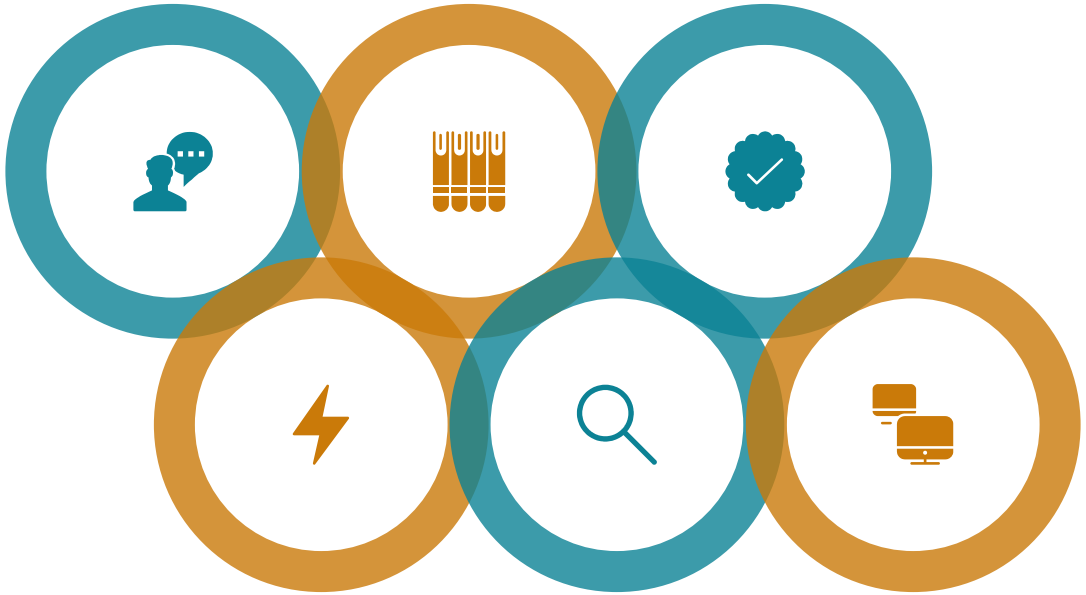
People spending more time in the airport most likely planned out their long stay and ate beforehand or packed their own food.

Recommendations

For future surveys, include information about grab-and-go food (premade food or vending machines)

Have smaller food options available for people in the airport a long time

Rerun survey to gather more info about people in SFO on layovers



Create a loyalty program where flyers can gain points by purchasing food or products from SFO stores & restaurants

Increase impulse purchases by redesigning SFO layouts and design leading up to stores & restaurants

Utilize Google Adwords for target advertising to show the great variety of restaurants

Thank You