

Final Project | Part One

EAST 2: Jonathan Gragg, William Johnson, Douglas Wiley

1 San Francisco International Airport Survey

1.1 Introduction

The San Francisco International Airport conducts a yearly comprehensive survey [the survey] of airport guests to rate their satisfaction of facilities, services and amenities. The goal of the survey is to compare the results to previous years and look for areas of and discover new opportunities for improvement. The survey is comprised of sections:

- Flight Information: choice of airline, the destination, reason for traveling.
- Passenger Experience: airport aesthetics, security and safety.
- SFO Website - access and overall usefulness
- Residence - Bay Area, state, country
- Demographic Information - age, gender, income

The survey version under study was taken from 2015.

1.1.1 Part A

1.1.1.1 Research Questions

This research centers around developing an in-depth view of who are the satisfied and unsatisfied customers. The approach is to explore the data without any explicit hypothesis, through the application of data science essentials: collecting, cleansing, exploring, and visualizing the data. Specifically, the research will focus on:

- **Who is satisfied or unsatisfied?** This research will identify if a customer satisfaction proxy can be created from the survey questions.
- **Who are the customers?** This research will explore the customer’s demographic data in the survey.
- **What about their flight habits?** Using the flight information in the survey, show the ways customers are intersecting with the airport.

In total this research seeks to narrate a story - in a most literal sense tell the customer journey. This insight will be useful in creating opportunities for improvement at the facility as well as developing future surveys.

1.1.1.2 Exploratory Data Analysis

The survey data is cross-sectional, wide format, and attitudinal in nature, implemented using dichotomous (Yes/No), multiple-choice rating questions along with open-ended text. There are 3,234 total observations, with 101 columns (also called features in this document).

Further inspection shows there are a significant number of features with missing values. Overall, this sparsity is not problematic, but most likely by design as these represent question categories without responses as well as comment fields without any entered text.

Of primary significance are the results from Question 6. This survey item asks ‘How does SFO rate on each of the following attributes?’ on 14 categories identified as a, b, c...n. Responses denote a level of acceptability ranging from 1-Unacceptable to 5-Outstanding, with 0 representing a ‘blank’, and 6 meaning N/A. These ratings will be useful in deriving a sentiment score for each observation. This table summarizes the items for Question 6.

Item	Topic
6a	Artwork and exhibitions
6b	Restaurants
6c	Retail shops and concessions
6d	Signs and directions inside SFO
6e	Escalators/elevators/moving walkways
6f	Information on screens/monitors
6g	Information booths (lower level near baggage claim)
6h	Information booths (upper level – departure area)
6i	Signs and directions on SFO airport roadways
6j	Airport parking facilities
6k	AirTrain
6l	Long term parking lot shuttle
6m	Airport rental car center
6n	SFO Airport as a whole

For example, shown here are two items from Question 6. These graphics summarize the responses for all of the observations for 6a. Artwork & Exhibitions and 6n. SFO Airport (considering the whole airport).

1.1.1.3 Analysis Plan

In answering **who is satisfied or unsatisfied**, the analysis will create a proxy variable ‘satisfied’ by isolating Questions 6a through 6n. For each observation, use the statistical mode (the value that appears most often) for the question:

- A mode corresponding with either Neutral, Good or Outstanding will result in a positive value.
- A mode corresponding with Poor or Unacceptable will result in a negative value.

Using the ‘satisfied’ proxy variable from the previous research and survey demographic data, **who are the customers** can be better understood by:

- Plotting the ages (survey question 17) and gender (survey question 18) of the customer and their level of satisfaction.
- Identifying the customer’s country of origin (survey question 16) and their level of satisfaction.

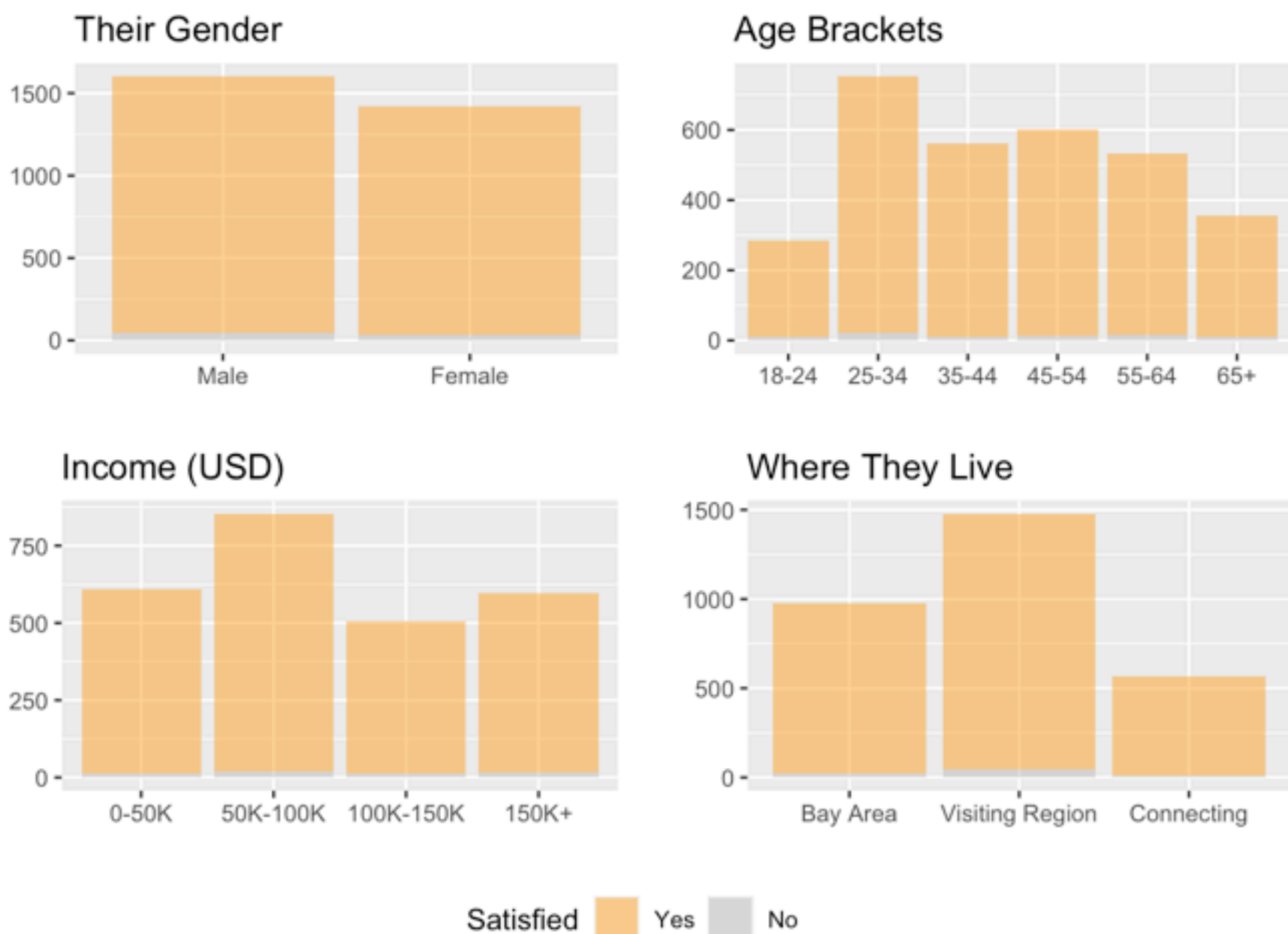
Better understanding the **customers flight habits** can be achieved by:

- Using the ‘satisfied’ proxy variable for the previous research.
- Viewing the customer’s yearly number of flights (survey question 5).
- Look at the details of their flight: connection (survey question 1) and destination (survey item destgeo).

1.1.1.4 Results

Survey question 6 was utilized to answer **who is satisfied or unsatisfied**. A look at all responses suggest a normal distribution, thankfully skewed towards more satisfied customers.

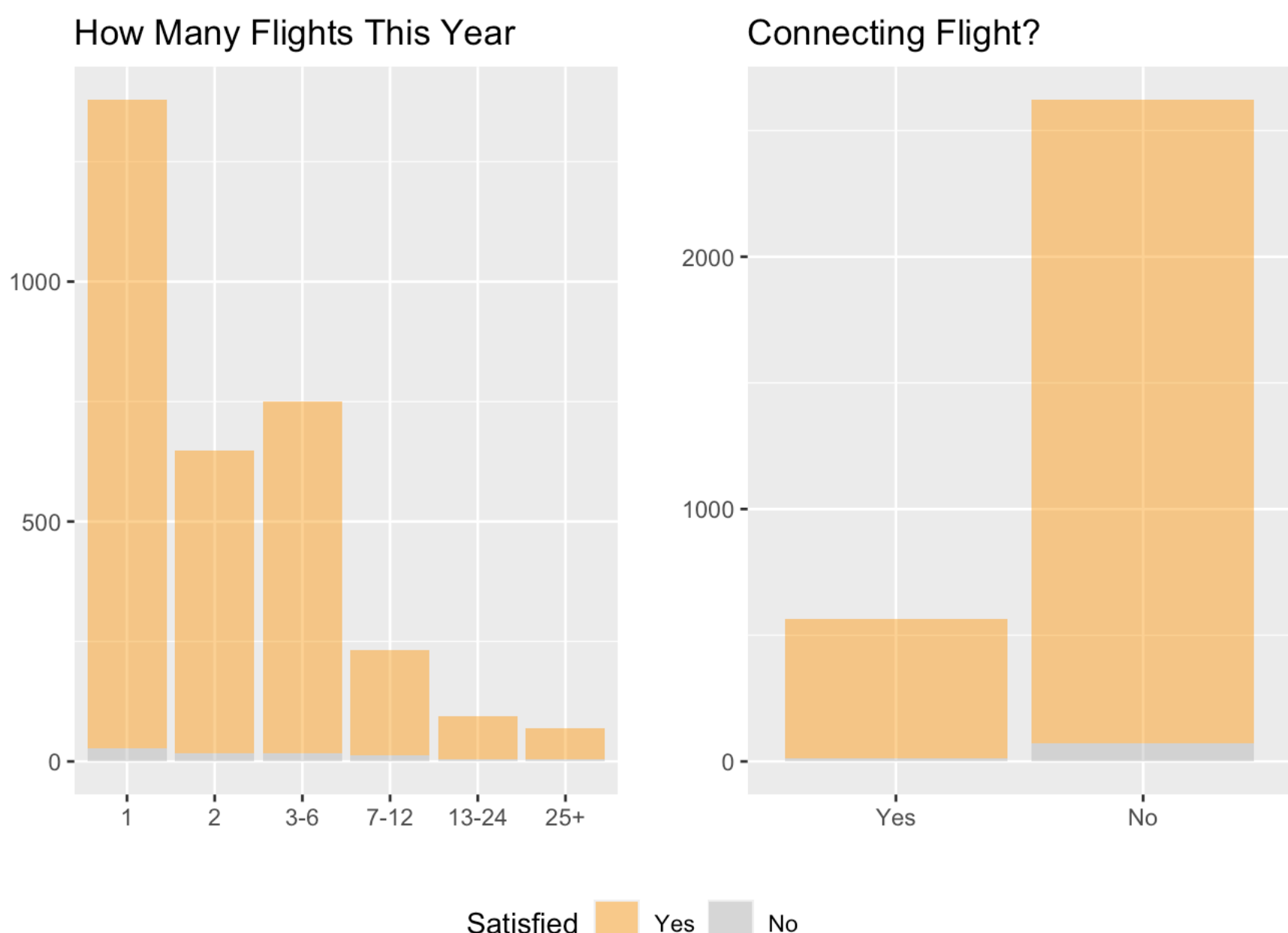
Using the satisfaction proxy created in the previous research, **who are the customers** becomes clear. Survey respondents across all genders, age brackets and levels of income are satisfied with the SFO airport as show here.



Survey respondents come from many different countries. Of the entire survey population, here are the countries that had greater than 10 surveys completed. The good news continues, across all countries:

Country	Surveys	Satisfied
USA	2297	97%
Canada	106	98%
Germany	66	98%
India	61	96%
Japan	57	96%
Australia	56	100%
UK	42	97%
Mexico	41	100%
China	36	100%
New Zealand	23	95%
Netherlands	19	100%
Brazil	18	100%
Taiwan	15	100%
Korea	12	100%
Italy	10	90%
Peru	10	100%
Philippines	10	100%

In viewing **customer flight habits**, for a large majority of survey respondents this was their only flight of the year at the time the survey was recorded. And concurring with the ‘Where They Live’ chart above, SFO is their final destination - San Francisco is either home or their place to visit.



1.1.1.5 Discussion

The story that is told here is that survey respondents are greatly satisfied with SFO Airport operations. This result is across all genders, ages and income levels. Regardless if the Bay Area is their home or their vacation spot, respondents are giving the airport the highest marks.

In terms of extracting actionable information from this story, there are some issues. First, the research based the satisfied proxy variable on a single set of questions. Though the questions covered a wide and important topics, it didn’t factor in other aspects of airport operations such as security. There are other opportunities to derive a satisfaction score for the survey responses:

- It would make interesting research to base the satisfaction score on sentiment analysis done only on the textual comments, and compare those results with the proxy variable created here.
- Given the imbalances in this data, a critical review of the survey itself may be required. It may need to be restructured in order to gain more actionable information for airport leadership.

1.1.1.6 Appendix A: Code