Introduction

In this assignment we will import data from the San Francisco airport and create a dataset that allows us to perform some summary statistics on the customers' experience with the airport and generate a list of customers that we will target for a 2017 follow up focus group.

The code below assumes that you are running this Notebook from the folder you unzipped the .zip package containing the datafiles and other assets as needed to run thes scripts.

Part 1

Part 1: Import the data and create a single data frame containing key fields on survey responses for further analysis. (a) List the the variables from each year that you used to create this data set using the original names they had in the data set they appeared in.

(b) Document any variable name changes so that it's clear what the original variables are whose names you changed. (Otherwise, a user of your data wouldn't be able to know what these variables are, or how to use them.)

(c) Describe your DataFrame in terms of its size, the variables in it, and how the data types of the variables. How many missing values do you have on the ratings variables?

(d) Write your new DataFrame to a csv file with an initial header record that includes the variable names. Verify that you wrote this file without errors.

In [2]:
```
import os
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import pandas_profiling
import pickle

sfo_2016 = pd.read_csv("SFO_2016.csv", parse_dates=True)
sfo_2015 = pd.read_csv("SFO_2015.csv", parse_dates=True)
sfo_2014 = pd.read_csv("SFO_2014.csv", parse_dates=True)
```

```
C:\Users\Jeff\Anaconda3\lib\site-packages\matplotlib\__init__.py:140
5: UserWarning:
This call to matplotlib.use() has no effect because the backend has
already
been chosen; matplotlib.use() must be called *before* pylab, matplot
lib.pyplot,
or matplotlib.backends is imported for the first time.

  warnings.warn(_use_error_msg)
```