

Homework 3 (MA 412)	Stark, Cameron
----------------------------	-----------------------

This homework pertains to your final project. Answer the following questions as best you can. If you have any questions or run into any problems, feel free to talk to me.

Question	Response
State your project topic (one sentence)	Create a program for users (those wanting to travel somewhere on an airline), that based on the users input outputs the probability the flight will encounter some issue (being a delay or cancellation) and the estimated price of an airfare on that day.
Explain what you plan to do for your project (in 2 paragraphs or less).	<p>I plan to take data from the Bureau of Transportation Statistics site which contains the delay and cancel data for U.S airlines for the past couple of years broken up by airport, and the cost of airfares and some expenses the airlines must deal with for the time period which can affect the price of the tickets.</p> <p>With this data I plan to create a linear regression of the data and based on the user input will return the probability based on prior conditions and the cost of an airfare based on the regression of passed data. This will be done with data languages like R or Python and will either be a console application or implement a Front-End into the application.</p>
What is the purpose of your study? (In other words why is this of interest to you or should be of interest to others)	The purpose of this study and the creation of the program are important to me and others that travel a lot, because the price of a ticket and the likelihood of delay and/or cancel of flight could hinder plans for a vacation or business trip
What is the population of interest in this study (go back and read section 2 if you are not sure what this means)	The population in this study is the data acquired for the airlines operating in the U.S., where the flight, revenue and expense data is publicly made available on the Bureau of Transportation website
What variable(s) will you study?	The variables that will be studied during this study and will be used to determine the probability of a delay or cancel will be summation or average of several on-time performance datasets and for the airfare prediction a similar summation or average of several financial datasets will be used for a linear regression
Give the data type of each variable	Both are quantitative values with the probabilities coming from ratios of total flights and flights that got delayed or cancelled and the predictive airfares coming from expenses and past ticket prices
What sources do you think might be good to use for getting data?	Will be using multiple data tables from the Bureau of Transportation site for the past flight data and also the google flights application has an api that contains flights that could be used for reference