**Graded by Gus Ostow**

Reviewed by **David Yerrington**

Project 1 Evaluation Rubric

Jordan Fritz – Denver

10/08/17

Instructions:

1. Rename this file with your city and name last name first i.e. Project 1 Evaluation Tool - LA - Cook\_Joshua.docx
2. For each of the selected questions, compare your code with the given solution.
3. It's OK if the path you took to get there was different, but the result should be the same.
4. Evaluate yourself based on:
   1. your code
   2. conceptual understanding
   3. what you could've improved

Keep in mind:

* Self-evaluation is important, and learning how to identify issues yourself is critical.
* The goal here is for both you AND an instructor to evaluate your work. You should be thinking critically about your answers and compare your self-evaluation with instructor feedback.
* Your comments on your work are in an integral part of how we will be providing feedback. The more detailed that you are with your discussion of your work, the more feedback we will be able to give you to help you grow.
* Our answers should not be thought of as gospel. According to the Zen of Python, “There should be one-- and preferably only one --obvious way to do it. Although that way may not be obvious at first unless you're Dutch.” Keep in mind that none of your instructors are Dutch and while we do our best to adhere to [PEP 8](https://www.python.org/dev/peps/pep-0008/), we may do things that are less than Pythonic.
* You should use the [Skitch](https://evernote.com/products/skitch) tool or some other similar tool to take screenshots of your work that you paste into this document as we have done with our answers.

Submit this doc on Sunday by 11:59 PM PST.

You will receive instructor evaluation by:

[Project 1 Evaluation Rubric](#2s8eyo1)

[**3. Using a for loop, load your data into a dict called data\_dict.**](#17dp8vu)

[**5. Transform your data\_dict to be oriented by column and call it data\_dict\_columns**](#3rdcrjn)

[**9. Summary Statistics of Numerical Columns**](#26in1rg)

[**10. Splitting the Data to Predicting Survival**](#lnxbz9)

[**14. Hypothesis Testing**](#35nkun2)

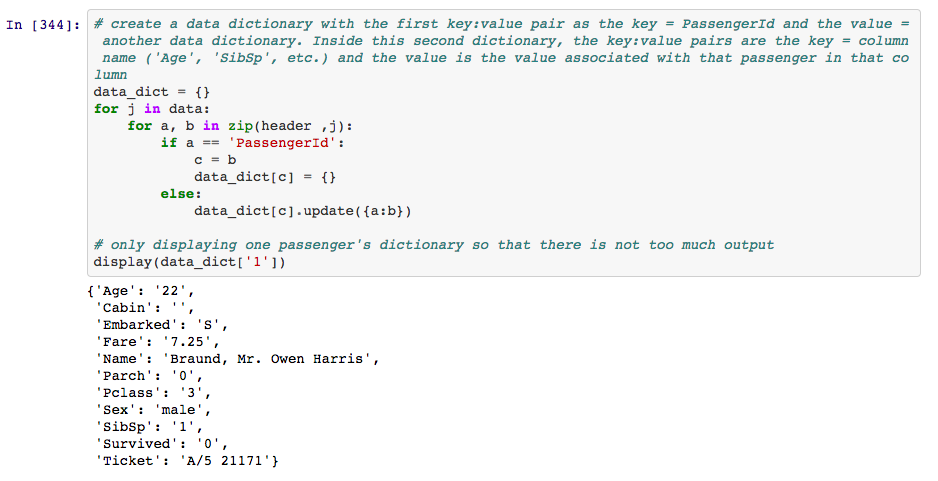
[**High-Level Evaluation**](#1ksv4uv)

**3. Using a for loop, load your data into a dict called data\_dict.**

***Our Answer***



***Your*** ***Response***

**

***Your Rating for your response: (3)***

( 3 - I killed it, 2 - I did pretty good, 1 - I struggled)

***Why you selected this rating***

My code works the same as the provided code, i.e. the dictionary is the exact same. For example, in the outside for loop, I use the variable “j” in the place the solution code uses the variable “row”. Instead of initiating the variable “zipped”, my code uses the zip function in place in the second for loop and assigns the variable to “a” and “b” rather than “col” and “element”. The solution code creates an extra dictionary (row\_dict) that assigns key value pairs to the “col” and “element” as “col” : ”element”. It then assigns all of the key : value pairs in the row\_dict to the data\_dict where the key : value pairs are “row[0]” (‘PassengerId’) : “row\_dict”. My code uses an if statement to find the ‘PassengerId’ and uses the value of the ‘PassengerId’ to assign a key : value pair equal to the value of the ‘PassengerId’ : an empty dictionary that is updated in the second if statement. In the second if statement, if the ‘PassengerId’ is not the value being of a, then the empty dictionary from the first if statement is updated with the key : value pair of each column name (key) and the corresponding value (value). I feel comfortable using for loops and manipulating dictionaries and lists. I had a solid command of this problem, and I can articulate why each code works for this question.

***Our Rating for your response: (3/2/1)***

**2**

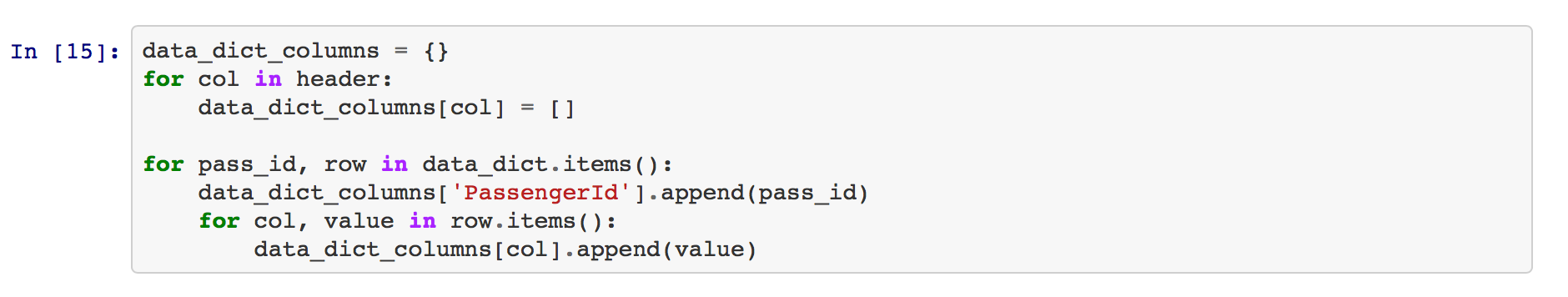
***Why we selected this rating***

**Your answer is algorithmically the same as our solution. I gave you a two on this question because** a**,** b**,** c**, and** j **are not okay names for variables, especially when you know somebody else is going to be reading your code.** b = c **?come on!!!**

## 

**5. Transform your data\_dict to be oriented by column and call it data\_dict\_columns**

***Our Answer***



***Your*** ***Response*****

***Your Rating for your response: (3)***

( 3 - I killed it, 2 - I did pretty good, 1 - I struggled)

***Why you selected this rating***

Both the provided code and my code have the same output. My code uses two for loops the same way in which the provided code does. The for loops loop through all values of all dictionaries in the original data dictionary (instantiated in number 3) and appends these values to the value of the key : value pair where the keys are the column labels. This creates a dictionary of lists where the lists contain all of the values of each column. The solution code uses similar concepts to loop through and append values of each column to a new dictionary. I could improve my readability by creating more descriptive variable names in the for loops rather than just ‘i’ or ‘j’.

***Our Rating for your response: (3/2/1)***

**3**

***Why we selected this rating***

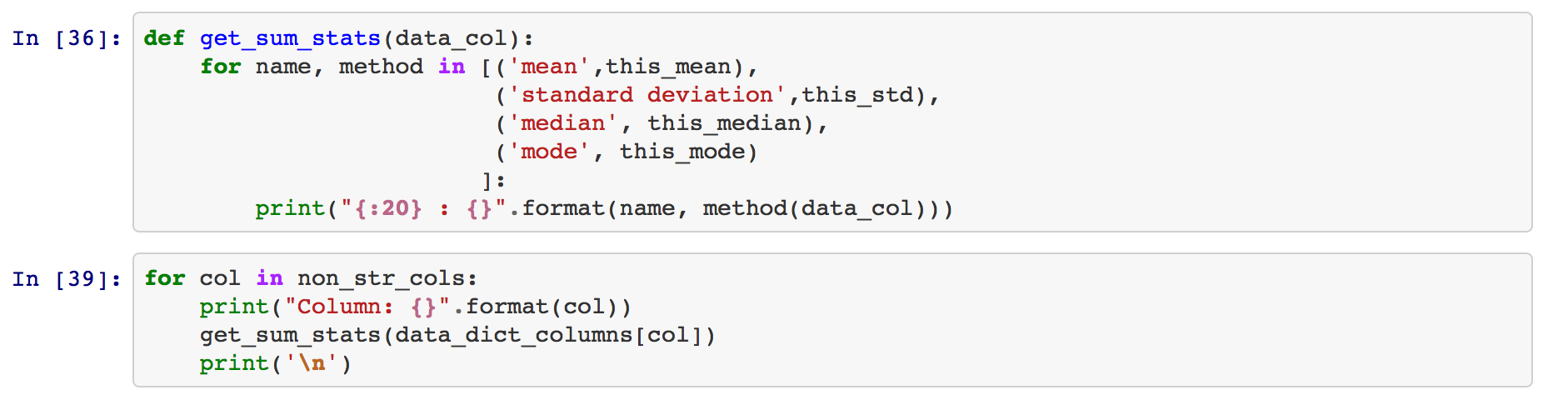
**I’m going to stop dinging you for readability because it seems like you’re aware of the problem in your self-assessment responses.**

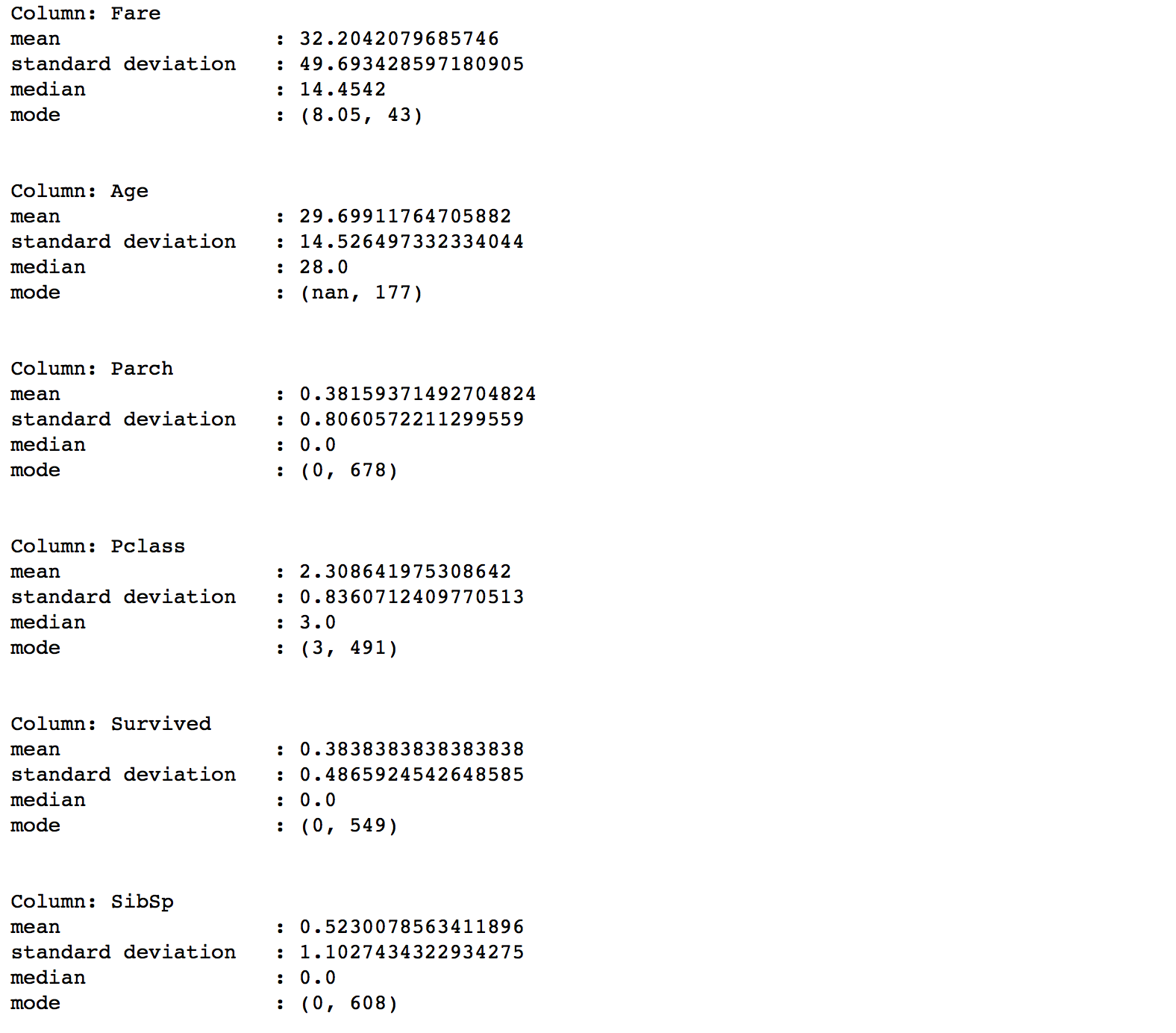
**Love the defaultdict. Perfect for this problem. That way you didn’t have to initialize the keys in the dictionary like the our solution did.**

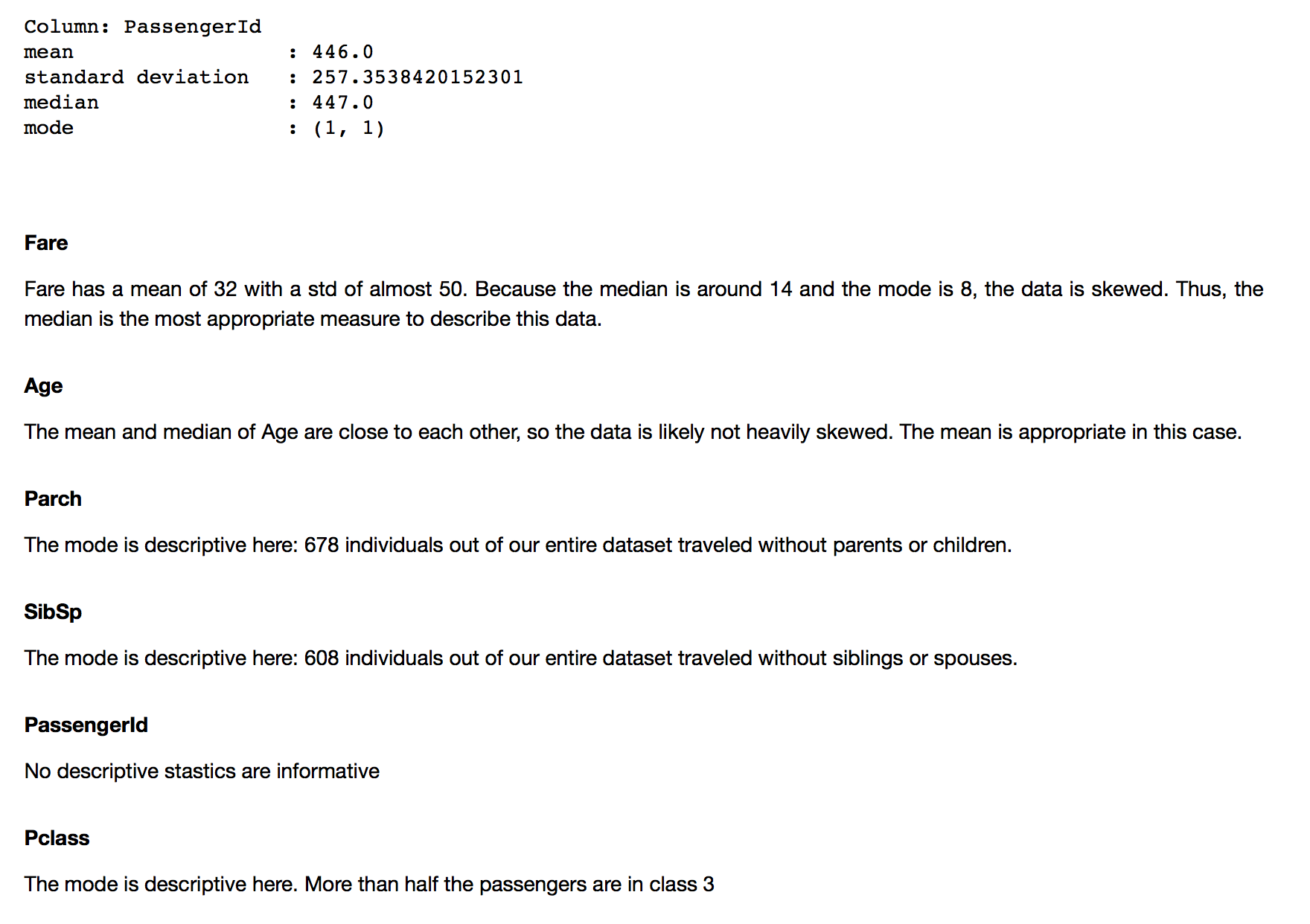
## 

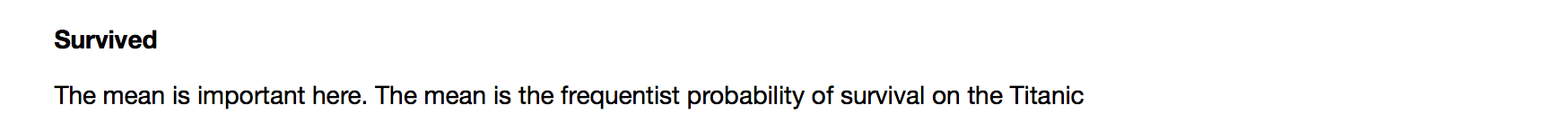
**9. Summary Statistics of Numerical Columns**

***Our Answer***

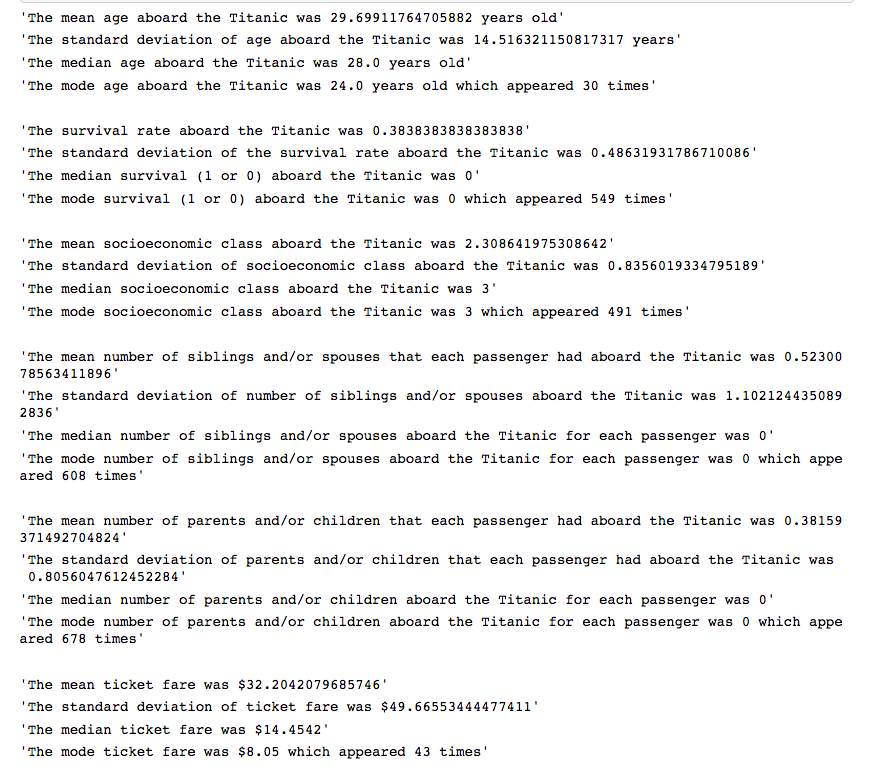
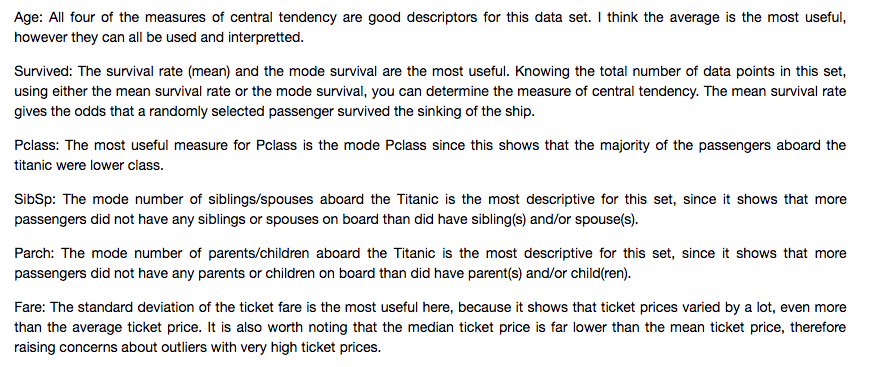








***Your*** ***Response***

**

***Your Rating for your response: (2)***

( 3 - I killed it, 2 - I did pretty good, 1 - I struggled)

***Why you selected this rating***

All of my final metrics are calculated correctly except for the mode age, because I did not include the nan values in my count (-1). I also did not run any of the functions on my ‘PassengerId’ metric (-4). I gave myself a 2 because I completed 23/28 of the metrics correctly, and I also completed 5/6 of the discussion questions correctly. The only one that is incorrect is Fare, since I thought the high standard deviation (greater than the mean) proved that the data had many outliers since the data is very spread out. My lab mentions how the median may be a good metric to use. I understand the problem, however missed on a few minor answers.

***Our Rating for your response: (3/2/1)***

**2**

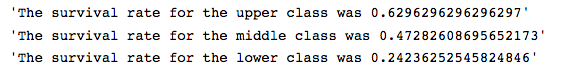
***Why we selected this rating***

**I’m not giving you a two because of tiny problems with your metric calculations and interpretations. I’m giving you a two because you didn’t generate the output programmatically. That looks like it was so tedious.**

**10. Splitting the Data to Predicting Survival**

***Our Answer***



***Your*** ***Response***

***Your Rating for your response: (3)***

( 3 - I killed it, 2 - I did pretty good, 1 - I struggled)

***Why you selected this rating***

All of my outcomes are the same as the solution code. I have similar code to the answer key, however the answer key uses list comprehensions, which make the code a little more concise. The for loop goes through all values in the ‘Pclass’ column and adds up the number of people aboard the titanic in each socioeconomic class. Each if statement also has a nested if statement which checks to see whether or not that particular passenger survived. If the passenger survived, the survived count is incremented. By the end of the code, there are 6 incremented counters, one for the total number of passengers from each ‘Pclass’ and one counting the number of passengers from each ‘Pclass’ who survived the sinking of the ship. The survival rate is calculated using the number who survived and dividing that by the total passengers from that socioeconomic class. The comments in my code show what each individual if statement is used for.

I have struggled learning the correct syntax for dictionary and list comprehensions. These comprehensions make the code much more concise and are something am trying to better understand.

***Our Rating for your response: (3/2/1)***

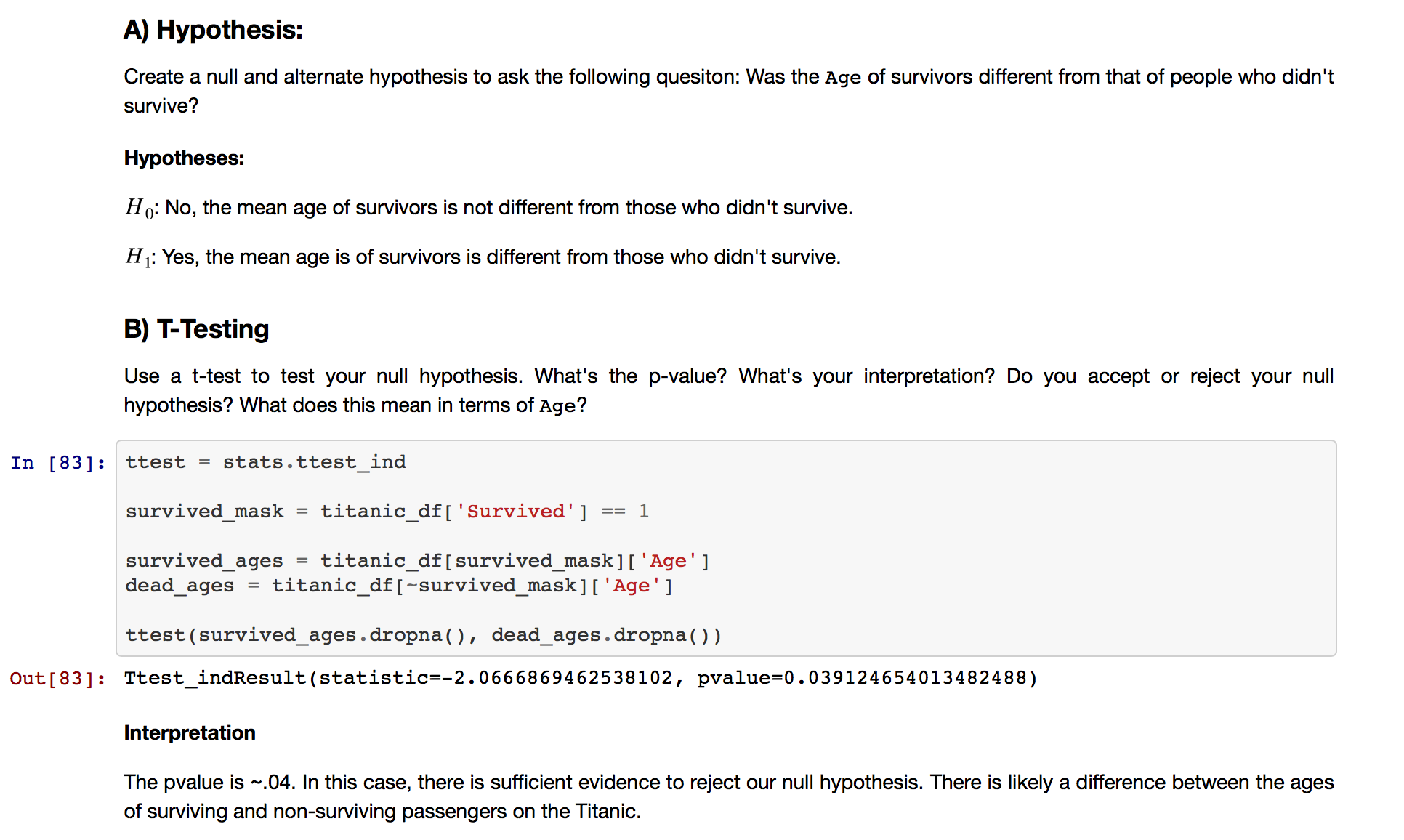
**2**

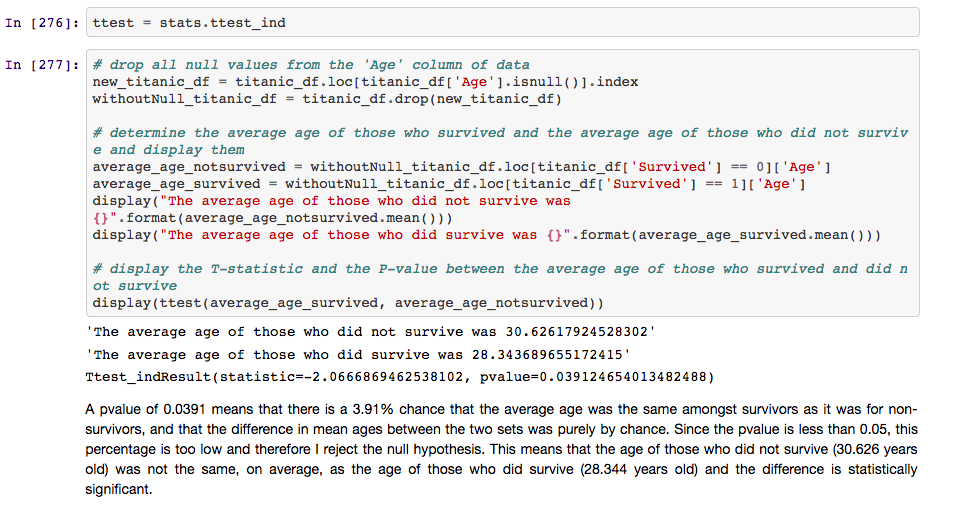
***Why we selected this rating***

**I think your self-assessing your code in a very clear way. You’re right that comprehensions really would have improved your solution. I’m not willing to give out a three without them.**

**14. Hypothesis Testing**

***Our Answer***



***Your*** ***Response*** **

***Your Rating for your response: (3)***

( 3 - I killed it, 2 - I did pretty good, 1 - I struggled)

***Why you selected this rating***

My code first finds all nan values included in the ‘Age’ column of the dataframe and assigns these to a new dataframe (new\_titanic\_df). This new dataframe is dropped from the original titanic\_df and assigned to withoutNull\_titanic\_df. Using the .loc() function, two variables are created (average\_age\_notsurvived and average\_age\_survived) where the ages of all passengers who survived (or not survived) are appended. Then, using the .mean() function, the mean age for each cohort of individuals is calculated and displayed. The solution code uses a survived\_mask and an inverse survived\_mask to create similar cohorts. Running the ttest function with each average returns the pvalue. Using the same thought process as the solution code (pvalue < 0.05), the null hypothesis is rejected.

***Our Rating for your response: (3/2/1)***

**2**

***Why we selected this rating***

**The setup of the hypotheses and the interpretation of the p-value is flawless. However, your implementation in python has problems. Use the** .dropna() **dataframe method to more efficiently handle nulls. Also,** average\_age\_survived **and** average\_age\_notsurvived **are misleading variable names. “Average” suggests a scalar when these are really pandas series.**

**High-Level Evaluation**

***One thing I did very well on this project was ...***

I feel comfortable using basic python functions and items, and feel my past experience in other coding languages has been an asset establishing a good understanding of python. This assignment was relatively easy.

***O******ne thing that I’m not clear on and would like to know more about is***

I am struggling with working with dictionary and list comprehensions. I do not understand the syntax, and I really struggled getting number 4. I think knowing this would help me to become a better python user.

***If I was starting this project over from the beginning knowing what I know now, one thing I would do differently is ...***

Try to make my code more efficient and readable. Most of the project, my variables are just arbitrarily named instead of being descriptive. I think adding descriptive variables would help me to be less confused after I put my code down, and then revisiting my work later to check over my work. Double checking my work would also help others read my code and follow what is happening.